



**HAL**  
open science

# Analyse de sensibilité fiabiliste en présence d'incertitudes épistémiques introduites par les données d'apprentissage

Gabriel Sarazin

► **To cite this version:**

Gabriel Sarazin. Analyse de sensibilité fiabiliste en présence d'incertitudes épistémiques introduites par les données d'apprentissage. Probabilités [math.PR]. Institut Supérieur de l'Aéronautique et de l'Espace Toulouse, 2021. Français. NNT: . tel-03361835

**HAL Id: tel-03361835**

**<https://hal.science/tel-03361835>**

Submitted on 1 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# THÈSE

En vue de l'obtention du  
**DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE**  
Délivré par l'Institut Supérieur de l'Aéronautique et de l'Espace

---

Présentée et soutenue par  
**Gabriel SARAZIN**

Le 27 mai 2021

**Analyse de sensibilité fiabiliste en présence d'incertitudes  
épistémiques introduites par les données d'apprentissage**

---

Ecole doctorale : **EDMITT - Ecole Doctorale Mathématiques, Informatique et  
Télécommunications de Toulouse**

Spécialité : **Mathématiques et Applications**

Unité de recherche :  
**ISAE-ONERA MOIS MODélisation et Ingénierie des Systèmes**

Thèse dirigée par  
**Jérôme MORIO et Agnes LAGNOUX**

Jury

M. Bruno TUFFIN, Rapporteur  
M. Arnaud GUYADER, Rapporteur  
M. Jean-Marc BOURINET, Examineur  
M. Merlin KELLER, Examineur  
M. Mathieu BALESDENT, Examineur  
M. Jérôme MORIO, Directeur de thèse  
Mme Agnès LAGNOUX, Co-directrice de thèse  
Mme Véronique MAUME-DESCHAMPS, Présidente



**À mon frère,  
À mes parents,  
À mes grands-parents.**



# Table des matières

<b>Table des matières</b>	<b>v</b>
<b>Liste des figures</b>	<b>vii</b>
<b>Liste des tableaux</b>	<b>ix</b>
<b>Remerciements</b>	<b>1</b>
<b>Introduction</b>	<b>5</b>
<b>1 Apprentissage d'une distribution de probabilité multidimensionnelle</b>	<b>15</b>
1.1 Introduction . . . . .	16
1.2 Caractérisation de la distribution d'un vecteur aléatoire . . . . .	17
1.3 Délimitation du problème d'apprentissage statistique . . . . .	29
1.4 Méthodes d'apprentissage statistique pour les lois marginales . . . . .	31
1.5 Méthodes d'apprentissage statistique pour la copule . . . . .	40
1.6 Conclusion . . . . .	65
<b>2 Estimation de probabilités d'évènements rares pour l'analyse de fiabilité</b>	<b>77</b>
2.1 Introduction . . . . .	78
2.2 Analyse de fiabilité déterministe . . . . .	79
2.3 Analyse de fiabilité stochastique . . . . .	85
2.4 Conclusion . . . . .	99
<b>3 Analyse de sensibilité fiabiliste</b>	<b>105</b>
3.1 Introduction . . . . .	106
3.2 Analyse de sensibilité de la sortie d'un code de calcul . . . . .	108
3.3 Application à l'analyse de sensibilité d'une probabilité de défaillance . . . . .	121
3.4 Conclusion . . . . .	128
<b>4 Analyse de la variabilité d'un risque de défaillance suite à l'apprentissage de la distribution conjointe des entrées</b>	<b>139</b>
4.1 Introduction . . . . .	140
4.2 Développement d'une procédure d'ASF dans un cadre simplifié . . . . .	141
4.3 Validation de la procédure d'ASF . . . . .	163
4.4 Renforcement de la procédure d'ASF pour le cadre général . . . . .	183
4.5 Validation de la procédure d'ASF renforcée . . . . .	191
4.6 Application au flambage d'une plaque composite . . . . .	201
4.7 Conclusion . . . . .	212
<b>5 Analyse de la variabilité d'un risque de défaillance suite à l'apprentissage de la structure de corrélation des entrées</b>	<b>217</b>
5.1 Introduction . . . . .	218
5.2 Formulation d'un problème d'ASF . . . . .	220
5.3 Développement d'une procédure d'ASF pour le sous-problème . . . . .	233

5.4 Développement d'une stratégie d'assimilation de nouvelles données . . . . .	249
5.5 Validation des deux procédures d'ASF . . . . .	257
5.6 Application au flambage d'une plaque composite . . . . .	275
5.7 Conclusion . . . . .	278
<b>Conclusion</b>	<b>283</b>

# Liste des figures

1	Schématisation de la méthodologie de propagation et de quantification des incertitudes. . . . .	7
2	Schématisation de la première méthodologie d'analyse de sensibilité fiabiliste. . . .	10
3	Schématisation de la seconde méthodologie d'analyse de sensibilité fiabiliste. . . . .	11
1.1	Représentation schématique de l'ensemble $\mathcal{C}_2$ . . . . .	24
1.2	Trois exemples de R-vignes : une C-vigne, une D-vigne et une R-vigne quelconque. . .	55
1.3	Exemple d'un modèle HAC et de ses extensions. . . . .	60
1.4	Graphe <b>G</b> pour un modèle PBC. . . . .	62
1.5	Graphe <b>D</b> pour un modèle CBN. . . . .	62
3.1	Illustration de l'intuition ayant motivé la création des $\delta$ -indices de Borgonovo. . . .	117
4.1	Décomposition des incertitudes épistémiques au cours de l'apprentissage statistique. .	144
4.2	Propagation de la variabilité d'échantillonnage. . . . .	145
4.3	Partitionnement de l'échantillon d'apprentissage en plusieurs blocs. . . . .	150
4.4	Propagation de la variabilité d'échantillonnage après séparation des sources d'apprentissage. . . . .	151
4.5	Mécanisme de ré-échantillonnage $\mathcal{B}$ permettant d'imiter les générateurs $G_Z$ et $G_V$ . .	154
4.6	Illustration des différentes briques de calcul intervenant dans l'algorithme d'ASF. . .	156
4.7	Schématisation du principe de <i>Pick-Freeze</i> pour l'application $\Psi$ . . . . .	157
4.8	Mécanisme de ré-échantillonnage $\mathcal{B}_1$ en présence d'observations supplémentaires de $X_1$ . . . . .	160
4.9	Mécanisme de ré-échantillonnage $\mathcal{B}_{\text{cop}}$ en présence d'observations supplémentaires de $\mathbf{U}$ . . . . .	162
4.10	Visualisation de l'impact du contexte " <i>small data</i> " sur l'estimation du risque de défaillance. . . . .	166
4.11	Convergence des indices $\hat{S}_j$ et $\hat{S}_{\text{cop}}$ en fonction du nombre $M$ d'itérations. . . . .	169
4.12	Illustration du processus de raffinement des mesures empiriques. . . . .	173
4.13	Visualisation du problème d'analyse de fiabilité. . . . .	177
4.14	Évolution de la variabilité affectant l'apprentissage de la densité marginale $f_{X_1}$ . . . .	178
4.15	Influence du modèle d'apprentissage $\mathcal{M}$ sur la distribution de l'estimateur $P_f(\tilde{\mathbf{Z}}, \tilde{\mathbf{V}})$ . . .	181
4.16	Mise en évidence de la charge de calcul induite par $\mathcal{R}_w$ . . . . .	184
4.17	Différentes stratégies pour conduire l'ASF avec $\mathcal{R}_w$ . . . . .	187
4.18	Illustration des différentes briques de calcul lorsque l'ASF est exécutée avec $\mathcal{R}_w$ . . . .	190
4.19	Visualisation des techniques d'échantillonnage dans l'espace normal standard. . . .	192
4.20	Visualisation du domaine de défaillance $D_f$ . . . . .	194
4.21	Visualisation de l'impact du contexte " <i>small data</i> " sur l'estimation du risque de défaillance. . . . .	196
4.22	Flambage d'une plaque en appui simple sous un effort $L_x$ de compression uniaxiale. . .	201
4.23	Plan de coupe d'un stratifié à huit couches respectant la symétrie "miroir". . . . .	202
4.24	Contexte de propagation des incertitudes pour le flambage d'une plaque composite. . .	206
4.25	Visualisation de la sensibilité de la loi $\mu_D$ à la modélisation de l'épaisseur $e$ . . . . .	209

5.1	Propagation de la variabilité d'échantillonnage. . . . .	221
5.2	Propagation de la variabilité d'échantillonnage via les coefficients de corrélation. . . . .	229
5.3	Décomposition de la stochasticité intervenant dans l'application $\tilde{\Psi}_{\tau}$ . . . . .	231
5.4	Illustration des différentes briques de calcul intervenant dans l'algorithme d'ASF. . . . .	239
5.5	Illustration des différentes briques de calcul lorsque l'ASF est exécutée avec $\mathcal{R}_w$ . . . . .	241
5.6	Transformation de $\tilde{\Psi}_{\tau}$ en une application déterministe $\check{\Psi}_{\tau}$ incomplète. . . . .	245
5.7	Idée d'une méthode pour la simulation de $P_f(\mathbf{U}_{\text{obs}})$ conditionnellement à $\{\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}}) = \mathbf{t}_0\}$ . . . . .	247
5.8	Mécanisme de ré-échantillonnage $\mathcal{B}_{ij}$ en présence d'observations supplémentaires de $\mathbf{U}_{ij}$ . . . . .	251
5.9	Transformation d'une copule gaussienne en une copule D-vigne. . . . .	254
5.10	Incorporation des données supplémentaires lors de l'estimation d'une copule gaussienne. . . . .	256
5.11	Convergence des indices $\hat{\delta}_{ij}^F$ en fonction du nombre $M$ d'itérations. . . . .	259
5.12	Décomposition schématique des applications $\Psi_{\tau}$ et $\tilde{\Psi}_{\tau}$ . . . . .	263
5.13	Visualisation des lois marginales et bidimensionnelles de $\mathbf{Z}$ à partir de $\mathbf{z}_{\text{tot}}$ . . . . .	266
5.14	Visualisation de l'influence du modèle $\mathcal{M}_{\text{cop}}$ et de l'échantillon $\tilde{\mathbf{U}}$ sur l'approximation $\hat{D}_f^{\mathcal{N}}$ . . . . .	269
5.15	Adaptation de la chaîne de propagation des incertitudes. . . . .	277
4	Synthèse de l'imbrication des deux procédures d'ASF. . . . .	289

# Liste des tableaux

4.1	Estimation des indices de Sobol pour trois échantillons distincts. . . . .	169
4.2	Convergence des indices de Sobol après raffinement des mesures empiriques. . . . .	172
4.3	Impact de l'assimilation de données supplémentaires sur le classement des indices de Sobol. . . . .	175
4.4	Résultats de l'ASF pour différentes versions du modèle d'apprentissage $\mathcal{M}$ . . . . .	182
4.5	Résultats de l'ASF avec $\mathcal{R}$ et avec différentes versions de $\mathcal{R}_w$ . . . . .	193
4.6	Résultats de l'ASF avec $\mathcal{R}$ et avec différentes versions de $\mathcal{R}_w$ . . . . .	198
4.7	Impact de l'assimilation de données supplémentaires sur le classement des indices de Sobol. . . . .	200
4.8	Modèle probabiliste affecté aux variables macroscopiques décrivant la plaque composite. . . . .	207
4.9	Analyse de la sensibilité du facteur de flambage $\Lambda$ aux entrées $\mathbf{M}$ , $\Theta$ et $e$ . . . . .	208
4.10	Résultats de l'ASF initiale et des scénarios d'assimilation de données. . . . .	211
5.1	Lien entre les coefficients $\hat{\tau}_{ij}$ et le paramètre $\hat{\theta}$ pour différentes copules archimédiennes. . . . .	228
5.2	Estimation des $\delta$ -indices de Borgonovo pour trois échantillons distincts. . . . .	260
5.3	Impact de l'assimilation de données supplémentaires sur le classement des $\delta$ -indices. . . . .	262
5.4	Comparaison des résultats obtenus pour les deux sous-problèmes d'ASF . . . . .	264
5.5	Comparaison des résultats obtenus pour les deux sous-problèmes d'ASF . . . . .	272
5.6	Bilan des incertitudes indésirables introduites au cours de l'ASF. . . . .	273
5.7	Impact de l'assimilation de données supplémentaires sur le classement des $\delta$ -indices. . . . .	274
5.8	Résultats de l'ASF initiale et du scénario d'assimilation de données. . . . .	278



# Remerciements

Par ces quelques lignes, je souhaite remercier toutes les personnes qui ont participé de près ou de loin au bon déroulement et à la réussite de mon parcours doctoral, cet ovni présenté par certains comme l'ultime étape d'un long périple universitaire et par d'autres comme l'indispensable première étape d'une carrière d'ingénieur de recherche.

Tout d'abord, j'adresse de sincères remerciements aux différents membres du jury, à commencer par Messieurs **Arnaud GUYADER** et **Bruno TUFFIN** qui ont accepté avec beaucoup d'enthousiasme de rapporter les travaux conduits au cours de ma thèse. Je leur suis gré d'avoir examiné avec autant de minutie et de bienveillance le corps de mon manuscrit. Dans le même élan, je tiens à remercier Messieurs **Jean-Marc BOURINET** et **Merlin KELLER**, tous deux examinateurs de ce travail. Le partage de leur expertise lors de la séance de questions-réponses a fait de ce dialogue un précieux moment d'échanges entre spécialistes, dont la subtilité se trouvait quelque part au croisement entre l'ingénierie, l'algorithmie, les statistiques et la modélisation probabiliste. De même, je remercie avec la plus grande courtoisie Madame **Véronique MAUME-DESCHAMPS** qui m'a fait l'honneur de présider le jury et d'assurer la direction de la soutenance en distanciel.

Mes pensées se tournent ensuite vers ceux qui, quatre années durant, m'ont aidé à accomplir la fameuse métamorphose de stagiaire à docteur. Je parle évidemment du quatuor d'encadrants composé d'Agnès, Jérôme, Mathieu et Loïc.

- À mon directeur de thèse **Jérôme MORIO**, je souhaite témoigner toute ma gratitude. Après un stage de fin d'études qui avait convaincu les deux partis, il m'a offert l'opportunité d'entreprendre une thèse de doctorat financée par l'Office National d'Études et de Recherches Aérospatiales (ONERA) et la Région Occitanie. Les années passées sous sa direction m'ont permis d'acquérir de solides connaissances en ingénierie des incertitudes et de gagner en épaisseur et en assurance tant sur le plan intellectuel que professionnel. Je garderai un souvenir agréable des longues discussions techniques que nous avons pu avoir, que ce soit pour statuer sur le bienfondé d'une intuition, pour dessiner les contours d'un nouvel édifice méthodologique, ou pour débattre de la qualité d'un résultat numérique.
- Je remercie très chaleureusement ma co-directrice de thèse **Agnès LAGNOUX** dont la présence au sein de cet encadrement m'a ouvert les portes de l'Institut Mathématique de Toulouse (IMT). Ce contact avec le monde académique m'a été d'autant plus profitable qu'il m'a incité à consacrer une plus grande énergie à la justification rigoureuse des solutions proposées dans cette thèse. Le soutien d'Agnès a été indéfectible tout au long de mon expérience doctorale et je lui en suis très reconnaissant.
- Je tiens également à exprimer mes plus vifs remerciements à **Mathieu BALESDENT** et **Loïc BREVAULT** qui ont suivi le déroulement de cette thèse depuis Palaiseau. Au cours des différentes réunions d'avancement qui ont accompagné la progression de mes recherches, j'ai pu apprécier leur ouverture d'esprit et leur curiosité mais aussi leur sens du détail et leur opiniâtreté. Leur regard sur mon travail, contrepoids indispensable aux influences académiques évoquées précédemment, m'a poussé à

chercher inlassablement à améliorer des solutions que je pensais satisfaisantes mais qui étaient trop souvent en désaccord avec la réalité des applications aérospatiales et de leurs contraintes. Je remercie Mathieu qui m’a directement aidé à “élaguer” l’article principal et à identifier les idées forces de la méthodologie proposée. Enfin, j’ai une pensée émue à l’attention de Loïc qui a eu l’immense bravoure de continuer à me superviser en dépit des épreuves qui l’ont frappé. Je me réjouis de son rétablissement et je le remercie pour ses sages conseils.

Le cas test présenté dans cette thèse a été proposé par **François-Xavier IRISARRI** et **Antoine HURMANE** du Département Matériaux et Structures (DMAS) de Châtillon. N’étant au départ pas initié au comportement mécanique des matériaux composites, je les remercie d’avoir patiemment répondu à mes questions et de m’avoir aidé à acquérir un bon niveau de compréhension de la physique sous-jacente. Cette balade au pays des stratifiés orthotropes fut très enrichissante et a indéniablement donné une plus grande coloration applicative à des travaux qui, sans cela, auraient été essentiellement d’ordre méthodologique. J’en profite pour remercier **Sylvain DUBREUIL** qui m’a prêté main forte à plusieurs reprises lorsque le réglage du cas test exigeait un peu plus que les quelques éléments que j’avais glané ça et là.

Pendant ces trois années de thèse, j’ai eu l’occasion d’enseigner auprès d’un public d’étudiants de premier et de second cycle. À ce titre, je souhaite remercier ceux qui m’ont impliqué dans ces projets pédagogiques et qui m’ont confié pendant quelques heures une partie de leur contingent d’étudiants : **Agnès LAGNOUX** et **Marc PERRET** à l’Université Jean Jaurès, **Sylvain DUBREUIL** à l’INSA de Toulouse et **Xavier GENDRE** à l’ISAE Supaéro. Je remercie par la même occasion **James LEDOUX**, **Jérémy OMER** et **Teddy FURRON** pour leur invitation au « Séminaire de l’Entreprise » du département Génie Mathématique de l’INSA de Rennes.

Au sein de l’ONERA, je souhaite remercier les membres permanents du Département Traitement de l’Information et Systèmes (DTIS) que j’ai eu la chance de côtoyer. Je pense en particulier aux toulousains des équipes RIME, M2CI et MIDL que sont **Pierre BIEBER**, **Christel SEGUIN**, **Xavier OLIVE**, **Thomas DUBOT**, **Luis BASORA**, **Antoine JOULIA**, **Rémi LAFAGE**, **Thomas CHABOUD**, **Claire SARRAT**, **Brieuc DANET**, **Nathalie BARTOLI**, **Jean HERMETZ**, **Peter SCHMOLLGRUBER**, **Thierry LEFEBVRE**, **Sébastien DEFOORT**, **Frank MOREL**, **Armand OYZEL** et **Christophe DAVID**. Je suis conscient d’avoir eu le privilège d’évoluer dans une ambiance de travail à la fois saine, conviviale, stimulante et épanouissante. En plus du très bon accueil qui m’a été réservé dès mon arrivée, j’ai toujours été sensible à l’intérêt que les différents collègues ont pu porter à mes travaux et à mon évolution professionnelle. Je décerne une mention spéciale à **Sébastien AUBRY** qui a volé à mon secours pour surmonter la compilation de la bibliothèque VineCopulaCPP et la recherche de fuites de mémoire dans OpenTURNS. Enfin, je suis infiniment reconnaissant envers **Claire DABIN** pour son incroyable efficacité et sa grande sollicitude.

Si j’ai été suffisamment armé pour affronter les obstacles qui jalonnent l’aventure doctorale, c’est en grande partie grâce à la solide formation en mathématiques appliquées qui m’a été dispensée au sein du département Génie Mathématique de l’INSA Rennes, inauguré à la rentrée 2014, et dont je crois être le premier docteur. Je remercie l’ensemble de l’équipe pédagogique, en particulier les Professeurs **Loïc HERVÉ**, **Mohamed CAMAR-EDDINE**, **Jean-François DUPUY**, **Pierrette CHAGNEAU** et **Mounir HADDOU** qui ont façonné l’ingénieur-statisticien que je pense être devenu. Je souhaite également témoigner toute ma gratitude à **Marc BRIANE** et **James LEDOUX**, deux mentors aux styles différents mais qui ont œuvré conjointement à l’aiguillage de ma trajectoire professionnelle. De plus, je ne laisserai pas passer une si belle occasion de réitérer mes remerciements à l’intention du Professeur **Yves CAPDEBOSCQ** qui m’a accueilli à l’Université d’Oxford à

l'été 2016. Ce stage d'initiation à la recherche s'est révélé être une heureuse expérience et une formidable première étincelle.

Le moment est désormais venu de saluer mes camarades doctorants, en commençant par **Morgane MENZ** et **Rémy PRIEM**, mes plus proches compagnons d'armes. Morgane et moi avons cohabité en parfaite harmonie et avec une certaine complicité au sein du très spacieux bureau R241. La discrétion et la gentillesse que nous semblons partager en est sûrement la première raison. . . Par ailleurs, je me rappelle au bon souvenir de mon prédécesseur **Pierre DERENNES**. Je le remercie d'avoir accepté de me laisser prendre une place importante dans la construction de l'ultime chapitre de sa thèse consacrée à l'estimation des indices de Borgonovo. Par ailleurs, je souhaite la plus franche réussite à mes amis thésards, notamment à **Maxime EL MASRI** (avec qui j'ai passé tant d'agréables moments, à la pause méridienne ou sur les terrains de sport), mais aussi à **Gaspard BERTHELIN**, **Marco SAPORITO**, **Vincenzo PALLADINO** et **Marie CHIRON**. Je n'oublie pas l'intrépide **Julien MARINÉ**, que j'ai souvent entraîné dans des expéditions hasardeuses, qui resteront je l'espère des souvenirs mémorables. Enfin, je remercie **Tiphaine SABATIER** et **Vincent DROUET**, post-doctorants au moment où je les ai fréquentés, pour leur retour d'expérience et pour la richesse des discussions qui ont pu être les nôtres.

Pour supporter les longues heures de travail qu'il s'impose souvent de son propre chef, le doctorant a un impérieux devoir de savourer les quelques moments de liberté qui s'offrent à lui. Je n'ai pas dérogé à la règle et je souhaite pour cela remercier ceux qui m'ont aidé à garder le sourire même dans les moments les plus compliqués. Un grand merci à mon camarade de promotion Anthony, au charisme si inspirant, pour m'avoir recruté dans l'équipe de football à 7 d'Expleo, anciennement Assytem. À l'heure de quitter Toulouse, je dis tendrement adieu à la bande de foot à 5 du jeudi soir (mon Éric, mon Philippe, mon Thibault et mon Rodolphe sans oublier Oriol, Alejo, Sebas, Michael, Dorian, Laurent, Erwan, Farid et Valéry). J'ai également une pensée pour Mathieu, ami de toujours et fidèle compère du Comité des Fêtes de Gouaux de Luchon. Chacune de nos retrouvailles, qu'elle soit estivale ou hivernale, constitue un moment irremplaçable de gaieté et d'insouciance.

Bien évidemment, je n'aurais pas pu mener à bien cette thèse sans le soutien à toute épreuve de mes proches. Je remercie ma grand-mère toulousaine pour tous les bons moments passés en sa compagnie. J'embrasse affectueusement mes grands-parents bordelais dont l'amour est incommensurable et pour qui, je le sais, ma réussite est une intarissable source de fierté. Je remercie également du fond du cœur mes parents pour m'avoir toujours exhorté à donner le meilleur de moi-même, à viser haut et à rester intègre. Pour finir, je remercie mon frère, si loin, et en même temps si près. Qu'il sache que son sang-froid, sa clairvoyance et son érudition forcent jour après jour mon admiration.

À Aix-en-Provence, le 23 août 2021.



# Introduction

## Cadre

Ce travail de thèse a pu voir le jour grâce à un cofinancement de l'Office National d'Études et de Recherches Aérospatiales (ONERA) et de la Région Occitanie. **Gabriel Sarazin** est inscrit à l'Institut Supérieur de l'Aéronautique et de l'Espace (ISAE) et son parcours de thèse est supervisé par l'école doctorale Mathématiques, Informatique, Télécommunications de Toulouse (MITT). Il est rattaché au Département Traitement de l'Information et Systèmes (DTIS) de l'ONERA mais aussi à l'équipe Statistique et Optimisation (SO) de l'Institut de Mathématiques de Toulouse (IMT) et dispose à ce titre d'un bureau dans chacun des deux laboratoires. La direction de thèse est assurée par Jérôme Morio, maître de recherche dans l'unité Risque - Identification, Modélisation, Évaluation (RIME) du DTIS sur le centre ONERA de Toulouse, ainsi que par Agnès Lagnoux, maître de conférence à l'Université Toulouse 2 Jean Jaurès. Mathieu Balesdent et Loïc Brevault, tous les deux ingénieurs de recherche dans l'unité Méthodes Multidisciplinaires et Concepts Intégrés (M2CI) du DTIS sur le centre ONERA de Palaiseau ont également participé activement à l'encadrement de cette thèse.

## Contexte général

### Simulation numérique du comportement d'un système complexe

Dans le milieu industriel, l'ingénieur sûreté de fonctionnement doit veiller au respect strict des performances (fiabilité, maintenabilité, disponibilité et sécurité) du système complexe dont il est responsable. La question de la fiabilité est une problématique transversale dans tous les pans de l'industrie mais elle prend une autre dimension dans des contextes tel que l'aérospatial ou le nucléaire où les questions de pure rentabilité financière sont dépassées par l'enjeu crucial de la protection des vies humaines et de l'environnement. L'évaluation de la fiabilité ne peut pas être réalisée à partir de mesures instrumentales ou d'expérimentations sur les installations concernées. Au contraire, il est obligatoire d'utiliser des outils de simulation numérique complexes soutenus par des infrastructures de calcul adaptées. De cette manière, une maquette numérique du système est créée et on peut alors étudier par anticipation la réponse du système à différents scénarios redoutés par l'ingénieur car pouvant conduire à une défaillance partielle ou totale.

Le plus souvent, l'expertise scientifique sur le comportement physique du système est réunie au sein d'un code de simulation (code elsA<sup>1</sup> dans l'industrie aéronautique, code CATHARE<sup>2</sup> dans l'industrie nucléaire) qui calcule à partir de variables d'entrée (dimensions d'une structure, charges appliquées, propriétés élastiques des matériaux, paramètres climatiques, météorologiques ou atmosphériques) une variable de sortie synthétisant l'état du système (valeur maximale de la température, de la pression, des déformations). Si la sortie dépasse un seuil limite, le système est considéré comme défaillant. Comme il permet de mettre en place un critère de défaillance, le code de simulation est aussi appelé fonction de performance. Dans la plupart des cas, l'obtention de la sortie passe par la résolution numérique d'équations aux dérivées partielles issues de la modéli-

---

1. Plateforme multidisciplinaire dédiée à l'aéroélasticité, l'aérothermie et l'aéroacoustique.  
2. Code de thermohydraulique pour les accidents de réacteurs nucléaires à eau.

sation physique (analyse par éléments finis, analyse de dynamique des fluides). La manipulation du code de simulation présente deux difficultés majeures.

- La première difficulté vient du fait que l'ingénieur ne sait pas exprimer analytiquement la sortie en fonction des entrées. Par conséquent, il est contraint d'adopter une approche de type "boîte noire", c'est-à-dire qu'il doit faire abstraction de la théorie physique sous-jacente et se focaliser sur l'étude extérieure de la fonction mathématique reliant les entrées et la sortie. Il ne peut acquérir de l'information sur le comportement du code de simulation qu'à travers des appels successifs en différents points de l'espace d'entrée.
- La seconde difficulté s'explique par le fort coût de calcul inhérent à la résolution numérique d'équations aux dérivées partielles. En effet, pour garantir un niveau de précision suffisant, les schémas numériques sont construits à partir d'une discrétisation fine des équations. Chaque appel au code de simulation est donc onéreux et un budget de simulation exprimé en nombre d'appels est fixé en amont et imposé à l'ingénieur.

Cette double complexité (mathématique et numérique) du code de simulation met en jeu des contraintes antagonistes. D'un côté, l'opacité de la modélisation physique place l'ingénieur dans une situation où sa marge de manœuvre se limite à la compréhension d'une correspondance entre les valeurs fournies en entrée et celles obtenues en sortie. De l'autre, toute velléité exploratoire est bridée par un budget de simulation n'autorisant qu'un nombre restreint d'appels au code.

Une méthode très répandue pour s'affranchir de la problématique du coût de calcul consiste à utiliser un modèle de substitution (polynômes de chaos, processus gaussien, réseau de neurones, machine à vecteurs de support) que l'on désigne aussi sous le nom de métamodèle. L'idée est de dépenser l'intégralité du budget de simulation afin de constituer une base d'apprentissage qui est ensuite utilisée pour calibrer un modèle prédictif. Comme le coût de calcul d'une évaluation de ce modèle est insignifiant, il devient possible d'estimer un nombre illimité de sorties du code de simulation. Toutefois, il faut rester conscient que le remplacement du code initial s'accompagne toujours de l'introduction d'une erreur de prédiction. Dans un souci de clarté, il n'est pas envisagé dans ce travail d'entreprendre une métamodélisation du code de simulation.

## **Propagation d'incertitudes aléatoires à travers un code de simulation**

En règle générale, les valeurs nominales qu'il faut affecter aux entrées du code de calcul sont inconnues ou entachées d'incertitudes (variabilités spatiale et temporelle, défauts géométriques, dérives de fabrication, erreurs de mesure). Ces incertitudes peuvent provenir de phénomènes naturels et leur maîtrise dépasse largement le champ d'action de l'ingénieur. En ce sens, elles sont incompressibles et une façon d'en tenir compte est de les considérer comme aléatoires. L'aléa sur les entrées est modélisé dans un cadre probabiliste par une distribution multidimensionnelle qui se propage à travers le code de simulation. Comme la fonction de performance est déterministe, la distribution de la sortie est simplement l'image de la distribution des entrées suite à une transformation fortement non-linéaire induite par le code de simulation. Le fait de considérer la sortie comme une variable aléatoire permet de concevoir la défaillance comme un évènement dont la probabilité peut être calculée, ou du moins estimée. L'objectif principal de l'analyse de fiabilité est alors de quantifier la probabilité de défaillance du système. Ce n'est pas une tâche évidente car elle revient à estimer la queue de distribution supérieure d'une loi de sortie dont on ignore tout. On y parvient grâce à une méthodologie de propagation d'incertitudes dont les étapes clés sont schématisées à la Figure 1, inspirée des illustrations proposées par SUDRET [2007] puis IOOSS [2009].

La première étape est consacrée à la spécification du modèle probabiliste en entrée du code de simulation. En première approche, il a longtemps été supposé que les variables d'entrée étaient indépendantes. Sous cette hypothèse, la spécification se limite à l'examen successif et cloisonné de chaque entrée et une pratique courante consiste à solliciter un avis d'expert. Toutefois, si l'hypothèse d'indépendance est retenue à tort, il y a fort à craindre que la probabilité de défaillance soit sous-estimée ou sur-estimée. Depuis l'énoncé du théorème de SKLAR [1959], la caractérisation des

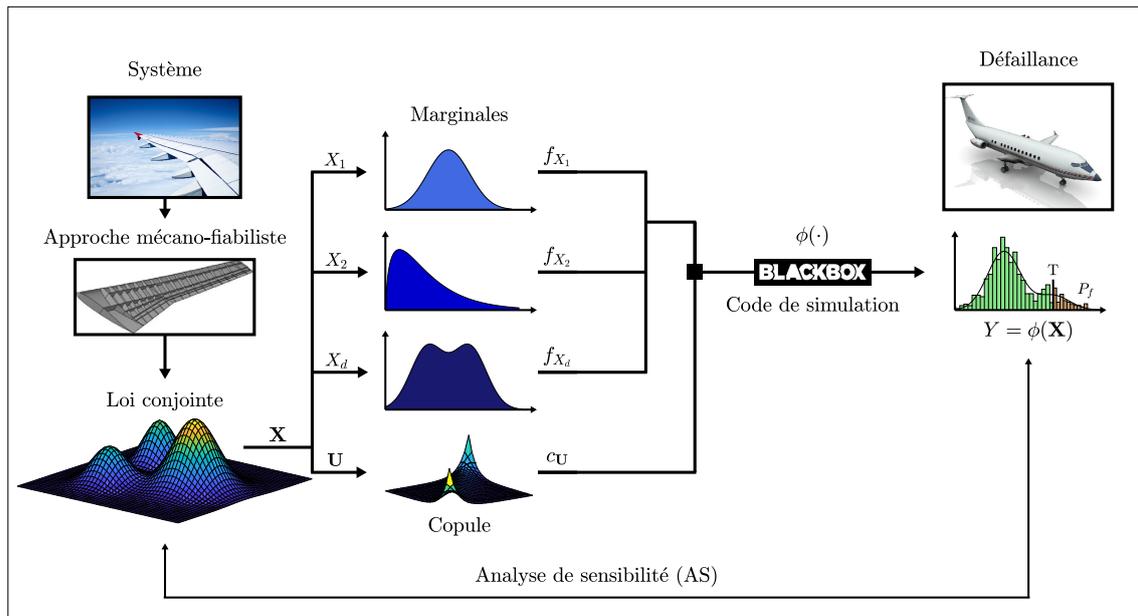


FIGURE 1 – Schématisation de la méthodologie de propagation et de quantification des incertitudes.

lois multidimensionnelles peut s'envisager sous l'angle de la décomposition copule-marginales. D'un côté, les spécificités inhérentes à chaque entrée (dispersion, symétrie, multimodalité ou comportement des valeurs extrêmes) sont traitées dans un cadre unidimensionnel tandis que, de l'autre, la copule concentre toute l'information sur la structure de dépendance intrinsèque. L'avènement de la théorie des copules avec [JOE \[1997\]](#) et [NELSEN \[2007\]](#) et l'intérêt qu'elle a suscité dans de nombreux domaines d'application ont permis la genèse d'une gamme de modèles dédiés exclusivement à la description de la dépendance (copules elliptiques, copules archimédiennes, copules aux valeurs extrêmes). Au cours de la dernière décennie, les familles usuelles de copules paramétriques ont été dépassées par des modèles de copules capables de résister à la montée en dimension (copules vignes régulières, copules archimédiennes emboîtées, réseaux bayésiens de copules). En pratique, il est rare qu'un expert soit en mesure d'apporter une réponse catégorique à la question du choix de la copule. Ce problème existe également au moment de la spécification des lois marginales mais il est bien plus criant dans le cas de la copule. Pour compenser cette déficience, une approche avancée par [TORRE et collab. \[2019\]](#) repose sur l'apprentissage de la copule à l'aide d'un jeu de données opérationnelles.

Une fois que la loi des entrées est fixée, la deuxième étape est la propagation des incertitudes à proprement parler. Son ressort principal est la méthode de simulation de Monte-Carlo. Ainsi, une version discrétisée de la loi de sortie est obtenue en simulant des échantillons dans l'espace d'entrée puis en calculant les sorties associées avec la fonction de performance. En présence d'un système très fiable et donc d'une probabilité de défaillance très faible, cette technique devient inopérante car aucun des échantillons simulés ne conduit à une défaillance. Pour y remédier, il faut utiliser des algorithmes de fiabilité qui exploitent différemment le budget alloué. En effet, ces algorithmes facilitent la simulation de l'évènement rare redouté, tout en introduisant après coup des mécanismes de pénalisation adéquats.

La troisième et dernière étape consiste à analyser la sensibilité de la sortie aux incertitudes véhiculées par les entrées. Comme le montre la Figure 1, le principe de l'analyse de sensibilité (AS) est la quantification de l'influence des entrées sur la distribution de sortie. La réponse apportée par l'estimation des indices de sensibilité permet un rebouclage de l'expertise sur le système et présente plusieurs intérêts.

- On dispose d'un classement de l'influence des différentes entrées. Si on cherche ensuite à réduire la variabilité sur la sortie, la nature de la variable la plus influente peut fournir

des indications sur les départements d'ingénierie ou les corps de métiers à contacter pour intervenir sur cette variable pilote.

- On affine la compréhension du code de simulation, notamment en détectant d'éventuelles interactions entre des variables.
- On identifie les variables de très faible influence sur la sortie. Elles peuvent être éliminées du processus de propagation des incertitudes et fixées à des valeurs nominales. Cette action permet notamment de réduire la dimensionnalité du problème.

### **Incertitudes épistémiques issues de la méconnaissance de la loi d'entrée**

Dans certaines situations, l'étape initiale de la méthodologie de propagation des incertitudes ne peut pas être menée à bien car on est incapable de spécifier avec certitude la loi des entrées. Ces incertitudes supplémentaires qui sont liées à la méconnaissance de la loi théorique sous-jacente sont qualifiées d'incertitudes épistémiques et sont traditionnellement opposées aux incertitudes aléatoires évoquées précédemment. Ce clivage décrit par **HOFFMAN et HAMMONDS [1994]** ou par **HELTON et BURMASTER [1996]** se fonde sur le caractère réductible des incertitudes épistémiques dont l'amplitude diminue à chaque apport supplémentaire d'information sur la loi des entrées. Même si **RIESCH [2013]** propose des systèmes de classification bien plus sophistiqués, la grille de lecture aléatoire/épistémique permet à l'ingénieur d'identifier les incertitudes qu'il peut espérer atténuer et celles dont il va devoir s'accommoder. L'existence d'incertitudes épistémiques vient perturber l'analyse de fiabilité dont les résultats sont conditionnés par la spécification du modèle probabiliste affecté aux variables d'entrée. En présence d'incertitudes épistémiques, la loi de la sortie et donc la probabilité de défaillance sont non seulement inconnues, mais elles sont aussi et surtout sujettes à des fluctuations car elles s'expriment comme les images d'un modèle probabiliste qui n'est plus complètement figé.

On précise que l'on entend par "méconnaissance" une information incomplète ou bien une information qui n'est pas jugée assez fiable pour mériter de fixer définitivement la loi des entrées sans introduire des outils permettant d'encadrer une éventuelle erreur de spécification. Dans un cadre probabiliste, il est assez fréquent de diviser les incertitudes épistémiques en deux catégories, celles qui proviennent de la méconnaissance du modèle de distributions (familles paramétriques usuelles, familles multidimensionnelles, assemblage paramétrique par morceaux à l'aide d'une copule) et celles qui proviennent de la méconnaissance des paramètres (moyennes, coefficients de variation, corrélations, degrés de liberté). Dans le second cas, l'existence d'incertitudes épistémiques peut correspondre à différentes situations rencontrées en pratique :

- des avis d'experts divergents quant à la valeur nominale à affecter à un paramètre,
- un expert qui fournit un intervalle de confiance pour chaque paramètre,
- un expert qui étalonne une famille paramétrique à partir d'une base de données susceptible d'évoluer en fonction des mises à jour du retour d'expérience.

D'un point de vue mathématique, il existe plusieurs méthodes pour construire une extension du cadre probabiliste initial prévu pour les seules incertitudes aléatoires. Pour accoucher d'un cadre probabiliste parfaitement unifié, il faut adopter le point de vue bayésien. En effet, l'essence même de la statistique bayésienne réside dans l'amalgame d'une incertitude sur la loi sous-jacente. En outre, le recours à une modélisation bayésienne hiérarchique à deux niveaux permet de prendre en compte en même temps l'incertitude liée à la sélection d'une famille et l'incertitude liée à la détermination des paramètres. Toutefois, le paradigme bayésien est exigeant car un pré-requis à son utilisation est la connaissance des lois *a priori*. La théorie des "probabilités imprécises" offre l'accès à d'autres outils de description des incertitudes épistémiques. Ils rendent parfois mieux compte de la réalité des méconnaissances mais leur incorporation, leur manipulation et leur interprétation dans un environnement purement probabiliste sont beaucoup moins aisées.

Quel que soit le formalisme privilégié, la loi des entrées n'est plus un invariant du problème et elle doit être vue comme une mesure de probabilité aléatoire. L'analyse de fiabilité est directement

impactée étant donné que la probabilité de défaillance devient une variable aléatoire fonction des paramètres incertains de la distribution d'entrée. Dans ce contexte, [DER KIUREGHIAN \[1989\]](#) a défini une mesure de fiabilité globale qui peut être estimée en propageant successivement les incertitudes aléatoires et épistémiques. En raison du coût de calcul induit par l'estimation d'une probabilité de défaillance, cette double propagation ne peut pas être effectuée en se contentant d'imbriquer deux procédures de Monte-Carlo. Pour limiter la charge de calcul liée à la présence de deux couches d'incertitudes, de nombreuses méthodes ont été développées et continuent d'être améliorées comme en témoignent les travaux de [SCHÖBI et SUDRET \[2017\]](#), [CHABRIDON \[2018\]](#) et [MEYNAOUI \[2019\]](#). Pour l'ingénieur, il peut être intéressant d'identifier les paramètres de la loi des entrées qui sont responsables de la plus grande part de variabilité dans la prédiction du risque de défaillance. Si la connaissance de ces paramètres est raffinée (par exemple en réclamant une expertise à leur sujet), une réduction significative du volume des incertitudes épistémiques peut être espérée. Cette analyse de sensibilité sur la probabilité de défaillance porte le nom d'analyse de sensibilité fiabiliste (ASF). Des algorithmes d'ASF compatibles avec un budget de simulation réduit ont été proposés par [MORIO \[2011\]](#), [WANG et collab. \[2013\]](#) ou encore [SONG et collab. \[2019\]](#).

## Contexte spécifique de cette thèse

### Incertitudes épistémiques issues d'un échantillon d'apprentissage

Dans cette thèse, on choisit de construire la distribution d'entrée directement à partir des données du retour d'expérience. Dans un problème où l'on sait que les entrées sont dépendantes et donc pour lequel il est nécessaire d'incorporer une copule au modèle probabiliste, il est vain de se mettre à la recherche d'un avis d'expert tranché et définitif à ce sujet. Au mieux, on peut espérer récupérer des valeurs nominales pour les coefficients de corrélation linéaire. De ce fait, l'approche proposée ici est pragmatique et suppose que l'ingénieur est capable d'assurer le traitement statistique des données. On se place dans une situation où la seule connaissance de la loi des entrées se résume à un unique échantillon ne comportant qu'un petit nombre d'observations.

Contrairement aux géants du Web, les départements d'étude et de conception de l'industrie traditionnelle ne sont pas toujours capables de rassembler un volume important de données. Cela peut s'expliquer par des contraintes techniques (mesures trop chères et trop chronophages) ou par l'inexistence d'un circuit de valorisation des données. Par opposition au "*big data*", la situation examinée ici correspond à un contexte de "*small data*" où le nombre d'observations est limité à quelques centaines. La distribution d'entrée est ensuite construite par apprentissage statistique, c'est-à-dire en utilisant des outils de statistique inférentielle pour extraire le plus d'information possible de l'échantillon fourni. Le résultat de cette phase d'apprentissage est le plus souvent une estimation de la densité conjointe des variables d'entrée. Pour conduire l'analyse de fiabilité, il est désormais nécessaire d'exécuter successivement l'algorithme d'apprentissage puis l'algorithme de fiabilité. On peut alors se demander comment évolue la probabilité de défaillance si un autre petit échantillon de la même loi sous-jacente sert à l'apprentissage. Cette question de l'impact de la variabilité d'échantillonnage est incontournable pour pouvoir donner du crédit au résultat de l'analyse de fiabilité. La Figure 2 insiste sur le fait que la variabilité d'échantillonnage est convertie d'abord en variabilité d'apprentissage puis en variabilité de prédiction du risque de défaillance.

Cette variabilité d'apprentissage s'explique par le fait que les outils de statistique inférentielle ne sont pas adaptés pour le "*small data*". On comprend facilement que l'information apportée par l'histogramme d'un petit échantillon est assez pauvre. De même, dans le cadre d'un apprentissage paramétrique, la robustesse de l'estimation des paramètres diminue avec la rareté des données utilisées. Enfin, en ce qui concerne les techniques de sélection de modèle, étant donné que les critères d'information et les tests d'adéquation s'appuient sur des considérations asymptotiques, les réponses qu'ils apportent ne sont pertinentes que pour des échantillons de taille conséquente. Du point de vue de l'analyse de fiabilité, on peut craindre que cette variabilité d'apprentissage engendre une variabilité trop importante de la prédiction du risque de défaillance.

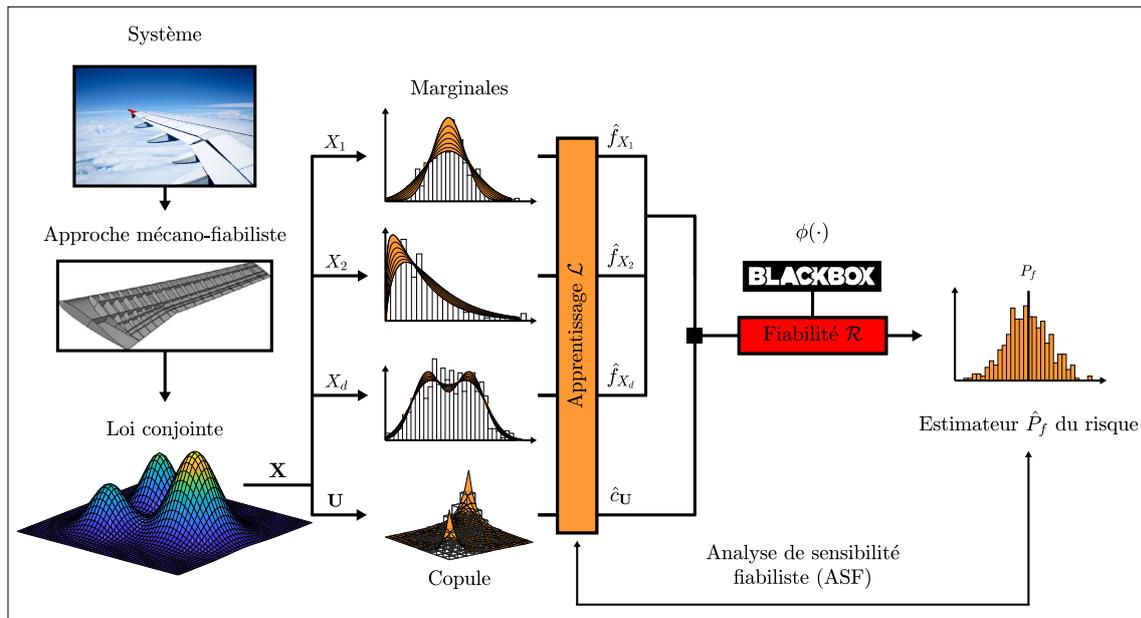


FIGURE 2 – Schématisation de la première méthodologie d'analyse de sensibilité fiabiliste.

## Énoncé des problématiques

Comme on ne dispose que d'un seul échantillon d'apprentissage et qu'il n'est pas raisonnable de le subdiviser en plusieurs sous-échantillons, on ne peut mener qu'une seule analyse de fiabilité. Or, pour pouvoir évaluer l'impact de la variabilité d'échantillonnage, il faut être capable d'estimer les probabilités de défaillance associées à toute une séquence d'échantillons. Cette impasse conduit à formuler une première problématique.

### (P<sub>1</sub>) Comment mesurer l'impact de la variabilité d'échantillonnage?

Dans un contexte de "small data", on peut s'attendre à ce que la prédiction du risque de défaillance soit sujette à une forte variabilité. Comme cette variabilité est imputable à l'action d'incertitudes épistémiques, l'ingénieur peut espérer la réduire en récoltant des informations supplémentaires sur la distribution d'entrée. Les règles universelles de la statistique inférentielle incitent à collecter des données supplémentaires pour toutes les entrées afin de réduire la variabilité d'apprentissage. Toutefois, cette intuition se heurte au fait que le processus de collecte des données est souvent fastidieux et qu'il ne peut pas être réenclenché rétrospectivement. Par contre, il est beaucoup plus facile de se procurer de nouvelles données pour une seule variable ou pour un petit nombre de variables de même nature. On parle d'acquisition ciblée pour désigner la stratégie qui consiste à récupérer de nouvelles données pour une ou deux variables et à les utiliser pour raffiner une partie de la distribution d'entrée (une moyenne, une corrélation, une loi marginale). La mise en œuvre d'une telle approche constitue la seconde problématique de ce travail.

### (P<sub>2</sub>) Est-il possible de réduire l'impact de la variabilité d'échantillonnage avec une procédure d'acquisition ciblée de données supplémentaires?

Cette problématique se divise en fait en deux sous-problématiques. D'une part, il faut être capable de décomposer la distribution d'entrée en plusieurs entités susceptibles d'être déconnectées les unes des autres au moment de l'apprentissage. D'autre part, il faut mettre au point un algorithme d'ASF capable de quantifier l'incertitude transférée sur la probabilité de défaillance par chaque entité au moment de son apprentissage. Pour envisager l'implémentation d'une ASF, il est nécessaire de propager l'incertitude d'échantillonnage, ce qui implique de répondre aux problématiques (P<sub>1</sub>) et (P<sub>2</sub>) en même temps. Comme on a déjà pu le dire, une ASF passe par la propagation de deux couches d'incertitudes et la maîtrise de la charge de calcul correspondante nécessite le recours à des solutions spécifiques.

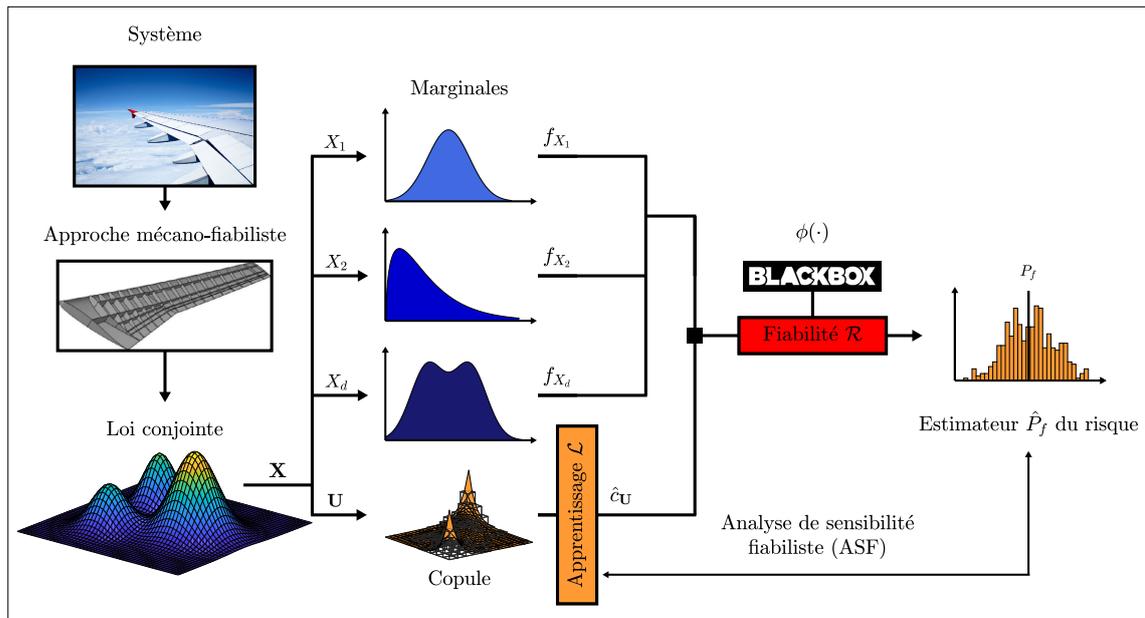


FIGURE 3 – Schématisation de la seconde méthodologie d'analyse de sensibilité fiabiliste.

## Organisation du document

Ce manuscrit est divisé en cinq chapitres qu'il est préférable de lire les uns à la suite des autres. Les Chapitres 1 à 3 constituent un état de l'art des méthodes existantes pour réaliser les phases d'apprentissage statistique, d'analyse de fiabilité et d'analyse de sensibilité (fiabiliste). Viennent ensuite les Chapitres 4 et 5 qui apportent la contribution de ce travail de thèse en cherchant à répondre aux problématiques ( $P_1$ ) et ( $P_2$ ).

- **Chapitre 1.** L'objectif principal de ce chapitre est de mettre en avant le rôle crucial joué par la séparation copule-marginales dans la construction de modèles d'apprentissage flexibles et robustes. Après de brefs rappels sur la théorie des probabilités, on mentionne plusieurs caractérisations possibles de la loi conjointe d'un vecteur aléatoire. On présente ensuite les techniques d'estimation (paramétrique et non-paramétrique) des lois marginales et de la copule. La dernière partie du chapitre est réservée à un inventaire des modèles de copules développés pour résister à la montée en dimension.
- **Chapitre 2.** Il s'inscrit dans la continuité directe du précédent chapitre. L'objectif est alors de comprendre comment estimer la probabilité de défaillance associée à une distribution d'entrée, éventuellement fournie par une phase préalable d'apprentissage. On insiste sur la complémentarité des algorithmes déterministes (FORM/SORM) et stochastiques (basés sur des techniques d'échantillonnage de Monte-Carlo). L'accent est par ailleurs mis sur la façon dont il faut conduire une analyse de fiabilité en présence d'une copule.
- **Chapitre 3.** Dans la perspective de construire des algorithmes d'ASF, la première partie du chapitre présente les principaux outils permettant d'effectuer une AS sur la sortie du code de simulation. On examine certains indices de sensibilité et leurs schémas d'estimation. Une attention particulière est réservée au cas épineux de l'AS pour des entrées dépendantes. La seconde partie du chapitre se focalise sur les procédures d'ASF (locales et globales) qui sont proposées dans le cadre bayésien.
- **Chapitre 4.** Dans ce premier chapitre de contribution, on cherche à analyser la sensibilité de la probabilité de défaillance à l'estimation des différentes composantes fonctionnelles de la décomposition copule-marginales. Cette situation est illustrée à la Figure 2. Si l'une des lois marginales est identifiée comme la composante dominante, cela indique à l'ingénieur la variable sur laquelle il doit chercher à collecter des données supplémentaires. Dans un

premier temps, un algorithme d'ASF est développé en supposant que la fonction de performance est peu coûteuse. Le point de départ de la solution proposée est un mécanisme de ré-échantillonnage *bootstrap* permettant de simuler la variabilité d'apprentissage sur les composantes fonctionnelles tout en garantissant leur indépendance mutuelle. Il est ensuite possible de quantifier l'influence des composantes fonctionnelles en utilisant la définition étendue des indices de Sobol et en déployant les schémas d'estimation associés. Dans un second temps, le recours à l'échantillonnage préférentiel multiple permet de mettre en place une stratégie d'économie du budget de simulation qui rend l'ASF possible même lorsque l'évaluation de la fonction de performance est plus coûteuse.

- **Chapitre 5.** Ce second chapitre de contribution vient compléter la réponse apportée par le premier dans le cas particulier où la copule est la composante fonctionnelle de plus grande influence. Pour cela, on fixe les lois marginales et on ne s'intéresse plus qu'à l'incertitude provenant de l'apprentissage de la copule. Cette nouvelle configuration est représentée à la Figure 3. L'objectif poursuivi est ensuite l'identification de la copule bidimensionnelle dont l'estimation a le plus d'impact sur la probabilité de défaillance. L'ingénieur peut alors savoir quelle est la paire de variables dont il doit en priorité améliorer l'état de connaissance. Une étude préliminaire montre qu'il est difficilement envisageable de répondre à ce problème d'ASF en quantifiant directement l'influence des copules bidimensionnelles. Le parti pris est alors de remplacer chaque copule bidimensionnelle estimée (implicitement ou explicitement) au cours de l'apprentissage par une unique variable scalaire, à savoir le coefficient de corrélation de Kendall. L'idée est ensuite d'analyser la sensibilité de la probabilité de défaillance aux différents coefficients de ce résumé matriciel de la structure de dépendance. Le nouveau problème d'ASF est donc en dimension finie, mais il porte sur une application stochastique dont les entrées sont dépendantes. La solution mise en avant consiste à estimer des indices de Borgonovo à partir d'un schéma "orienté copule" afin de quantifier l'influence de chaque coefficient de corrélation et de disposer ainsi d'une première idée de l'influence de chaque copule bidimensionnelle. De plus, on mesure la part d'information perdue au moment du résumé matriciel grâce à un indicateur de stochasticité pour lequel on propose un schéma d'estimation qui fait appel aux copules R-vignes non-paramétriques.

Les algorithmes d'ASF développés aux Chapitres 4 et 5 sont appliqués sur des cas tests analytiques de complexité croissante puis sur un cas test physique proposé par le Département MATériaux et Structures (DMAS) du centre ONERA de Châtillon. En fin de manuscrit, une conclusion récapitule les deux volets de la méthodologie proposée ainsi que les résultats obtenus.

## Liste des contributions

- **Publications dans des revues internationales**
  - ◇ [G. Sarazin](#), P. Derennes et J. Morio, 2020. "Estimation of high-order moment-independent importance measures for Shapley value analysis". *Applied Mathematical Modelling*, vol. 88, p. 396-417.
  - ◇ [G. Sarazin](#), J. Morio, A. Lagnoux, M. Balesdent, L. Brevault, 2020. "Reliability-oriented sensitivity analysis in presence of data-driven epistemic uncertainty". En révision mineure dans *Reliability Engineering and System Safety*.
- **Publication dans une conférence internationale à comité de lecture**
  - ◇ [G. Sarazin](#), J. Morio, A. Lagnoux, M. Balesdent, L. Brevault. "Sensitivity analysis of risk assessment with data-driven dependence modelling". *29th European Safety and Reliability Conference*, Hannovre, Septembre 2019.

## Références

CHABRIDON, V. 2018, *Reliability-oriented sensitivity analysis under probabilistic model uncertainty—Application to aerospace systems*, thèse de doctorat, Université Clermont Auvergne.

- DER KIUREGHIAN, A. 1989, «Measures of structural safety under imperfect states of knowledge», *Journal of Structural Engineering*, vol. 115, n° 5, p. 1119–1140.
- HELTON, J. C. et D. E. BURMASTER. 1996, «Guest editorial : treatment of aleatory and epistemic uncertainty in performance assessments for complex systems», *Reliability Engineering & System Safety*, vol. 54, n° 2-3, p. 91–94.
- HOFFMAN, F. O. et J. S. HAMMONDS. 1994, «Propagation of uncertainty in risk assessments : the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability», *Risk analysis*, vol. 14, n° 5, p. 707–712.
- IOOSS, B. 2009, *Contributions au traitement des incertitudes en modélisation numérique*, Habilitation à diriger des recherches (HDR), Université Toulouse Paul Sabatier, Toulouse.
- JOE, H. 1997, *Multivariate models and multivariate dependence concepts*, CRC Press.
- MEYNAOUI, A. 2019, *New developments around dependence measures for sensitivity analysis : application to severe accident studies for generation IV reactors*, thèse de doctorat, INSA de Toulouse, Toulouse.
- MORIO, J. 2011, «Influence of input PDF parameters of a model on a failure probability estimation», *Simulation Modelling Practice and Theory*, vol. 19, n° 10, p. 2244–2255.
- NELSEN, R. B. 2007, *An introduction to copulas*, Springer Science & Business Media.
- RIESCH, H. 2013, «Levels of uncertainty», dans *Essentials of risk theory*, Springer, p. 29–56.
- SCHÖBI, R. et B. SUDRET. 2017, «Structural reliability analysis for p-boxes using multi-level meta-models», *Probabilistic Engineering Mechanics*, vol. 48, p. 27–38.
- SKLAR, A. 1959, «Fonctions de répartition à  $n$  dimensions et leurs marges», *Publ. inst. statist. univ. Paris*, vol. 8, p. 229–231.
- SONG, J., P. WEI, M. VALDEBENITO, S. BI, M. BROGGI, M. BEER et Z. LEI. 2019, «Generalization of non-intrusive imprecise stochastic simulation for mixed uncertain variables», *Mechanical Systems and Signal Processing*, vol. 134, p. 106316.
- SUDRET, B. 2007, *Uncertainty propagation and sensitivity analysis in mechanical models : contributions to structural reliability and stochastic spectral methods*, Habilitation à diriger des recherches (HDR), Université Blaise Pascal, Clermont-Ferrand.
- TORRE, E., S. MARELLI, P. EMBRECHTS et B. SUDRET. 2019, «A general framework for data-driven uncertainty quantification under complex input dependencies using vine copulas», *Probabilistic Engineering Mechanics*, vol. 55, p. 1–16.
- WANG, P., Z. LU et Z. TANG. 2013, «An application of the kriging method in global sensitivity analysis with parameter uncertainty», *Applied Mathematical Modelling*, vol. 37, n° 9, p. 6543–6555.



# Chapitre 1

## Apprentissage d'une distribution de probabilité multidimensionnelle

### Sommaire

---

<b>1.1 Introduction</b> . . . . .	<b>16</b>
<b>1.2 Caractérisation de la distribution d'un vecteur aléatoire</b> . . . . .	<b>17</b>
1.2.1 Éléments de théorie des probabilités . . . . .	17
1.2.2 Fonction de répartition . . . . .	18
1.2.3 Densité de probabilité . . . . .	18
1.2.4 Loi conjointe et lois marginales . . . . .	19
1.2.5 Structure de dépendance . . . . .	20
1.2.6 Fonction caractéristique . . . . .	24
1.2.7 Statistiques issues d'une distribution multidimensionnelle . . . . .	24
<b>1.3 Délimitation du problème d'apprentissage statistique</b> . . . . .	<b>29</b>
1.3.1 Hypothèses sur la distribution sous-jacente . . . . .	29
1.3.2 Description du problème d'apprentissage statistique . . . . .	30
<b>1.4 Méthodes d'apprentissage statistique pour les lois marginales</b> . . . . .	<b>31</b>
1.4.1 Méthodes paramétriques pour l'apprentissage des lois marginales . . . . .	32
1.4.2 Méthodes non-paramétriques pour l'apprentissage des lois marginales . . . . .	37
<b>1.5 Méthodes d'apprentissage statistique pour la copule</b> . . . . .	<b>40</b>
1.5.1 Apprentissage conjoint des lois marginales et de la copule . . . . .	40
1.5.2 Méthodes paramétriques pour l'apprentissage des copules . . . . .	42
1.5.3 Méthodes non-paramétriques pour l'apprentissage des copules . . . . .	49
1.5.4 Modèles de copules adaptés à la grande dimension . . . . .	51
<b>1.6 Conclusion</b> . . . . .	<b>65</b>

---

## 1.1 Introduction

En ingénierie des incertitudes, le cadre méthodologique le plus habituel consiste à spécifier l'incertitude des entrées du problème, à propager ces incertitudes à travers un code de calcul puis à quantifier l'incertitude véhiculée sur la sortie ou sur une quantité d'intérêt obtenue à partir de la sortie. Dans le cadre de cette thèse, on est confronté à une méconnaissance du comportement aléatoire simultané des entrées. En effet, on ne dispose que d'un échantillon formé par un petit nombre d'observations conjointes des entrées. En pratique, cette situation est assez fréquente et apparaît comme une problématique de première importance dans [BEER et collab. \[2013\]](#) ou [ZHANG et collab. \[2014\]](#).

On adopte une approche probabiliste et on considère que chaque observation conjointe est en fait la réalisation d'un vecteur aléatoire obéissant à une certaine distribution de probabilité multidimensionnelle, sous-jacente mais inconnue. La statistique inférentielle vise à extraire une information pertinente de l'échantillon afin de construire une représentation mathématique précise de la distribution responsable de la variabilité des entrées. Dans la suite, on parle d'apprentissage statistique lorsqu'il faut désigner l'action d'un algorithme qui estime la distribution de probabilité des entrées à partir des données disponibles. Ce terme peut avoir une acceptation plus large, voire différente, en particulier dans [FRIEDMAN et collab. \[2001\]](#) où il désigne la construction d'un modèle prédictif dans un contexte d'apprentissage supervisé. Dans ce travail, il n'est pas question de prédire la sortie du code de calcul mais de mettre en place un modèle probabiliste qui restitue fidèlement l'incertitude affectant les entrées du problème. Ici, l'apprentissage statistique est une tâche qui se mène en amont de la phase de propagation des incertitudes et qui est complètement déconnectée du fonctionnement du code de simulation. Le modèle issu de l'apprentissage est ensuite couplé avec un algorithme de fiabilité. Celui-ci propage des incertitudes qui dépendent par conséquent du déroulement de la phase d'apprentissage. Ainsi, la prédiction fiabiliste est tributaire des informations qui sont extraites de l'échantillon. La qualité du modèle probabiliste construit détermine ainsi la qualité de l'estimation de la probabilité de défaillance.

Une distribution est un concept mathématique en lien avec la théorie de la mesure. Si on veut s'appropriier une distribution, il faut parvenir à estimer l'un des outils fonctionnels permettant de la caractériser. Depuis les travaux historiques de [BERNOULLI \[1711\]](#) sur la loi uniforme et ceux de [DE MOIVRE \[1756\]](#) sur la loi normale, une multitude de modèles paramétriques ont été inventés pour satisfaire les besoins des différents domaines où il est apparu important de se doter d'une modélisation fine des incertitudes. On renvoie aux ouvrages de [KOTZ et collab. \[2004\]](#) et [FANG \[2018\]](#) pour des monographies à ce sujet. La littérature oppose généralement ces modèles paramétriques aux techniques d'apprentissage dites non-paramétriques. Lorsque les entrées sont indépendantes, la dimension du problème n'est pas une contrainte car la distribution conjointe s'obtient directement à partir des distributions marginales. Le problème multidimensionnel se décompose alors en problèmes unidimensionnels pouvant être traités en parallèle. En raison du faible budget d'observations, il est toutefois assez difficile d'acquérir de l'information sur les régions extrêmes des distributions marginales. Pour des entrées dépendantes, l'apprentissage doit en revanche s'envisager dans un cadre multidimensionnel. Par ailleurs, la montée en dimension fait naître des motifs de dépendance polymorphes que la matrice de corrélation d'une loi gaussienne n'est plus en mesure d'appréhender. Face à ce problème, l'apparition de la théorie des copules avec [SKLAR \[1959\]](#) a offert un nouveau moyen de caractériser les distributions multidimensionnelles et de mettre en place leur apprentissage. La structure de dépendance intrinsèque peut être isolée des distributions marginales et décrite à l'aide d'une distribution de copule. La nature mathématique de l'objet copule permet notamment de construire des modèles plus flexibles affichant une meilleure résistance à la dimension.

Ce chapitre a pour objectif de présenter les principaux outils d'apprentissage statistique, en mettant principalement l'accent sur ceux qui sont utilisés dans la suite des contributions. La Section 1.2 pose les bases du cadre probabiliste en insistant notamment sur les différentes caractérisations d'une distribution multidimensionnelle. On y présente également divers statistiques

pouvant être calculées à partir de l'échantillon initial et délivrant une information sur la structure de dépendance. La Section 1.3 clarifie le périmètre et énonce les objectifs de l'apprentissage statistique dans le contexte de ce travail. La Section 1.4 décrit les méthodes d'apprentissage des distributions marginales avant que la Section 1.5 ne fasse de même avec la copule. Cette dernière section est l'occasion de présenter certains modèles de copules adaptés à la grande dimension, notamment les copules R-vignes.

## 1.2 Caractérisation de la distribution d'un vecteur aléatoire

### 1.2.1 Éléments de théorie des probabilités

On se place dans le cadre classique d'un espace probabilisé défini par un triplet  $(\Omega, \mathcal{A}, \mathbb{P})$ . On désigne par  $\Omega$  l'univers, c'est-à-dire l'ensemble des résultats possibles  $\omega$  de l'expérience aléatoire examinée. Tout sous-ensemble  $A$  de  $\Omega$  est alors appelé un évènement. Dans de nombreuses situations, l'ensemble des évènements issus de  $\Omega$  n'est pas dénombrable et on doit se munir d'une tribu  $\mathcal{A}$  (aussi appelée  $\sigma$ -algèbre) pour établir de façon rigoureuse les parties de l'univers à prendre ou non en considération. Toute tribu  $\mathcal{A}$  se définit alors comme une collection d'évènements, qui contient l'ensemble vide  $\emptyset$ , et qui est stable à la fois par passage au complémentaire et par union dénombrable. Pour une tribu  $\mathcal{A}$  donnée, une mesure de probabilité est une application de la forme :

$$\mathbb{P}: \begin{array}{ccc} \mathcal{A} & \longrightarrow & [0, 1] \\ A & \longmapsto & \mathbb{P}(A) \end{array}, \quad (1.1)$$

qui est  $\sigma$ -additive et qui vérifie  $\mathbb{P}(\Omega) = 1$ . Dans l'espace probabilisé ainsi constitué, on considère l'application  $\mathbf{X}$  définie par :

$$\mathbf{X}: \begin{array}{ccc} \Omega & \longrightarrow & \mathbb{R}^d \\ \omega & \longmapsto & \mathbf{x} := \mathbf{X}(\omega) \text{ avec } \mathbf{X} = [X_j]_{1 \leq j \leq d} \end{array}, \quad (1.2)$$

où chaque application coordonnée  $X_j$  est scalaire. Pour toute collection  $\mathcal{C}$  de parties de  $\mathbb{R}^d$  qui ne serait pas une tribu, on appelle tribu engendrée par  $\mathcal{C}$ , et on note  $\sigma(\mathcal{C})$ , la plus petite tribu (au sens de l'inclusion) contenant tous les évènements de  $\mathcal{C}$ . Pour munir  $\mathbb{R}^d$  d'une tribu adéquate, on a l'habitude d'utiliser la tribu borélienne  $\mathcal{B}(\mathbb{R}^d)$  qui est engendrée à partir des ensembles ouverts de  $\mathbb{R}^d$ . Elle peut aussi être générée par l'ensemble des pavés ouverts :

$$\mathcal{P}_1 := \{P = ]a_1, b_1[ \times \dots \times ]a_d, b_d[ \text{ avec } (a_j, b_j) \in \mathbb{R}^2 \quad \forall 1 \leq j \leq d\}, \quad (1.3)$$

ou par l'ensemble des produits d'intervalles non bornés à gauche et fermés à droite :

$$\mathcal{P}_2 := \{P = ]-\infty, x_1] \times \dots \times ]-\infty, x_d] \text{ avec } x_j \in \mathbb{R} \quad \forall 1 \leq j \leq d\}, \quad (1.4)$$

et on a donc :

$$\mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{P}_1) = \sigma(\mathcal{P}_2). \quad (1.5)$$

L'application  $\mathbf{X}$  est un vecteur aléatoire si elle est mesurable, c'est-à-dire si :

$$\forall B \in \mathcal{B}(\mathbb{R}^d), \mathbf{X}^{-1}(B) := \{\omega : \mathbf{X}(\omega) \in B\} \in \mathcal{A}. \quad (1.6)$$

Le couple  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  peut alors être équipé d'une mesure de probabilité induite  $\mu_{\mathbf{X}}$  :

$$\mu_{\mathbf{X}}: \begin{array}{ccc} \mathcal{B}(\mathbb{R}^d) & \longrightarrow & [0, 1] \\ B & \longmapsto & \mu_{\mathbf{X}}(B) := \mathbb{P}(\mathbf{X}^{-1}(B)) \end{array}, \quad (1.7)$$

qui est souvent appelée loi (de probabilité) ou distribution (de probabilité) de  $\mathbf{X}$  et décrit parfaitement son comportement aléatoire.

### **Théorème 1.2.1 (Loi d'un vecteur aléatoire)**

Soient  $\mathbf{X}$  et  $\tilde{\mathbf{X}}$  deux vecteurs aléatoires à valeurs dans  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ .

Les assertions suivantes sont alors équivalentes :

- (i)  $\mathbf{X}$  et  $\tilde{\mathbf{X}}$  ont la même loi.
- (ii)  $\forall B \in \mathcal{B}(\mathbb{R}^d), \mu_{\mathbf{X}}(B) = \mu_{\tilde{\mathbf{X}}}(B)$ .
- (iii) Pour tout pavé  $P = ]a_1, b_1[ \times \dots \times ]a_d, b_d[, \mu_{\mathbf{X}}(P) = \mu_{\tilde{\mathbf{X}}}(P)$ .
- (iv) Pour tout produit de la forme  $P = ]-\infty, x_1] \times \dots \times ]-\infty, x_d], \mu_{\mathbf{X}}(P) = \mu_{\tilde{\mathbf{X}}}(P)$ .

Si l'une des assertions est vérifiée, on est en droit d'écrire que  $\mathbf{X} \stackrel{\text{loi}}{=} \tilde{\mathbf{X}}$ .

Les passages de (iii) à (ii) et de (iv) à (ii) reposent sur le lemme d'unicité des mesures de probabilité. Il stipule que, pour toute collection  $\mathcal{C}$  de parties de  $\mathbb{R}^d$  qui est stable par intersection finie (ce qui est par exemple le cas de  $\mathcal{P}_1$  et  $\mathcal{P}_2$ ), on a :

$$\{\forall B \in \mathcal{C} : \mu_{\mathbf{X}}(B) = \mu_{\tilde{\mathbf{X}}}(B)\} \implies \{\forall B \in \sigma(\mathcal{C}) : \mu_{\mathbf{X}}(B) = \mu_{\tilde{\mathbf{X}}}(B)\}. \quad (1.8)$$

Ainsi, pour montrer que deux vecteurs aléatoires  $\mathbf{X}$  et  $\tilde{\mathbf{X}}$  ont la même loi, il suffit de montrer que les mesures de probabilité associées coïncident sur l'ensemble  $\mathcal{C}$  ayant permis d'engendrer  $\mathcal{B}(\mathbb{R}^d)$ .

### **1.2.2 Fonction de répartition**

En vertu du Théorème 1.2.1, toute l'information sur la loi d'un vecteur aléatoire  $\mathbf{X}$  peut être décrite à l'aide d'un outil appelé fonction de répartition et qui se définit en tout point de  $\mathbb{R}^d$  par :

$$\forall \mathbf{x} \in \mathbb{R}^d, F_{\mathbf{X}}(\mathbf{x}) := \mu_{\mathbf{X}}(] -\infty, x_1] \times \dots \times ] -\infty, x_d]) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d) \quad (1.9)$$

En anglais, on parle de *Cumulative Distribution Function* (CDF). Une égalité de type  $F_{\mathbf{X}} = F_{\tilde{\mathbf{X}}}$  est parfaitement équivalente à (i), ce qui offre une caractérisation plus synthétique et moins abstraite de la loi de  $\mathbf{X}$ . Les propriétés mathématiques de  $F_{\mathbf{X}}$  sont explicitées ci-après.

### **Théorème 1.2.2 (Propriétés de la fonction de répartition)**

Soit  $\mathbf{X}$  un vecteur aléatoire à valeurs dans  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ . Sa fonction de répartition  $F_{\mathbf{X}}$  vérifie :

- (F1)  $F_{\mathbf{X}}(x_1, \dots, x_d) = 0$  dès lors que  $\exists j : x_j = -\infty$ .
- (F2)  $F_{\mathbf{X}}(+\infty, \dots, +\infty) = 1$ .
- (F3)  $F_{\mathbf{X}}$  est isotonique :  $\forall \mathbf{x}, \tilde{\mathbf{x}} \in \mathbb{R}^d$  tels que  $\mathbf{x} \leq \tilde{\mathbf{x}}$ ,  $F_{\mathbf{X}}(\mathbf{x}) \leq F_{\mathbf{X}}(\tilde{\mathbf{x}})$ .
- (F4) La fonction  $t \mapsto F_{\mathbf{X}}(x_1, \dots, x_{j-1}, t, x_{j+1}, \dots, x_d)$  est continue à droite.
- (F5)  $F_{\mathbf{X}}$  est  $d$ -croissante.

La notion de  $d$ -croissance mentionnée dans la propriété (F5) du précédent théorème nécessite de définir l'ensemble  $C_P := \{a_1, b_1\} \times \dots \times \{a_d, b_d\}$  des coins d'un pavé  $P = [a_1, b_1] \times \dots \times [a_d, b_d]$ .  $F_{\mathbf{X}}$  est alors  $d$ -croissante si et seulement si :

$$\forall P \in \mathcal{P}_1, \sum_{\mathbf{c} \in C_P} (-1)^{v(\mathbf{c})} F_{\mathbf{X}}(\mathbf{c}) \geq 0 \quad \text{avec} \quad v(\mathbf{c}) := \text{Card}(\{j : c_j = a_j\}). \quad (1.10)$$

En dimension  $d = 2$ , cela se comprend très bien en remarquant que :

$$0 \leq \mu_{\mathbf{X}}([a_1, b_1] \times [a_2, b_2]) = F_{\mathbf{X}}(a_2, b_2) - F_{\mathbf{X}}(a_1, b_2) - F_{\mathbf{X}}(a_2, b_1) + F_{\mathbf{X}}(a_1, b_1). \quad (1.11)$$

Réciproquement, les propriétés énumérées dans le Théorème 1.2.2 suffisent à identifier une fonction de répartition. Ainsi, à toute fonction  $F : \mathbb{R}^d \rightarrow [0, 1]$  vérifiant les conditions (F1) à (F5) correspond un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$  et un vecteur aléatoire  $\mathbf{X}$  dont  $F$  est la fonction de répartition.

### 1.2.3 Densité de probabilité

Un élément essentiel de la théorie de l'intégration est la mesure de Lebesgue  $\lambda$  définie pour tout pavé  $P = [a_1, b_1] \times \dots \times [a_d, b_d]$  par :

$$\lambda(P) := \prod_{j=1}^d (b_j - a_j), \quad (1.12)$$

le théorème d'extension de Carathéodory lui donnant un sens pour tout borélien de  $\mathcal{B}(\mathbb{R}^d)$ .  $\lambda$  est une mesure mais ce n'est pas une mesure de probabilité car elle n'est pas finie. Une mesure de probabilité  $\mu_{\mathbf{X}}$  est dite absolument continue par rapport à  $\lambda$  si elle est vérifie la propriété suivante :

$$\forall B \in \mathcal{B}(\mathbb{R}^d), \lambda(B) = 0 \implies \mu_{\mathbf{X}}(B) = 0. \quad (1.13)$$

Cela exclut les lois discrètes et, plus généralement, toutes les lois possédant un atome. Si  $\mu_{\mathbf{X}}$  est absolument continue, alors il existe une fonction  $f_{\mathbf{X}} : \mathbb{R}^d \rightarrow [0, +\infty[$  mesurable telle que :

$$\forall B \in \mathcal{B}(\mathbb{R}^d), \mu_{\mathbf{X}}(B) = \int_B f_{\mathbf{X}} d\lambda = \int_B f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}. \quad (1.14)$$

$f_{\mathbf{X}}$  est la densité de probabilité de  $\mathbf{X}$ , appelée *Probability Density Function* (PDF) en anglais. De par la définition ci-dessus,  $f_{\mathbf{X}}$  intègre à 1 sur  $\mathbb{R}^d$ . Par ailleurs, un lien direct peut alors être établi entre  $F_{\mathbf{X}}$  et  $f_{\mathbf{X}}$  :

$$\forall \mathbf{x} \in \mathbb{R}^d, F_{\mathbf{X}} = \mu_{\mathbf{X}}([-\infty, x_1] \times \dots \times [-\infty, x_d]) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} f_{\mathbf{X}}(z_1, \dots, z_d) dz_1 \dots dz_d. \quad (1.15)$$

La dérivée partielle mixte d'ordre  $d$  de  $F_{\mathbf{X}}$  existe presque partout dans  $\mathbb{R}^d$  et on a l'égalité suivante :

$$f_{\mathbf{X}}(x_1, \dots, x_d) = \frac{\partial^d F_{\mathbf{X}}}{\partial x_1 \dots \partial x_d}(x_1, \dots, x_d). \quad (1.16)$$

On peut ainsi remarquer que  $f_{\mathbf{X}}$ , sous réserve de son existence, caractérise complètement la loi de  $\mathbf{X}$ . Elle est d'autant plus intéressante qu'elle fournit une expression intégrale de l'espérance de toute quantité s'écrivant comme une fonction de  $\mathbf{X}$ .

#### **Théorème 1.2.3 (Théorème de transfert)**

Soit  $\mathbf{X}$  un vecteur aléatoire admettant une densité de probabilité  $f_{\mathbf{X}}$ .

Soit  $h : \mathbb{R}^d \rightarrow \mathbb{R}$  une fonction mesurable.

L'espérance de la variable  $h(\mathbf{X})$  est alors définie par :

$$\mathbb{E}[h(\mathbf{X})] = \int_{\mathbb{R}^d} h(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}. \quad (1.17)$$

En particulier, si  $h$  est la fonction indicatrice d'un borélien  $B \in \mathcal{B}(\mathbb{R}^d)$ , on retrouve la définition initiale de  $f_{\mathbf{X}}$  donnée dans (1.14) :

$$\mu_{\mathbf{X}}(B) = \mathbb{E}[\mathbb{1}_B(\mathbf{X})] = \int_{\mathbb{R}^d} \mathbb{1}_B(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \int_B f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}. \quad (1.18)$$

### 1.2.4 Loi conjointe et lois marginales

On cherche souvent à déduire de  $\mu_{\mathbf{X}}$  des informations sur le comportement aléatoire d'une application coordonnée  $X_j$  prise séparément des autres. Pour cela, appliquons au vecteur  $\mathbf{X}$  une fonction de projection  $p_j$  permettant d'accéder à sa  $j$ -ème coordonnée :

$$\begin{array}{ccccc} (\Omega, \mathcal{A}, \mathbb{P}) & \longrightarrow & (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) & \longrightarrow & (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ \omega & \xrightarrow{\mathbf{X}} & \mathbf{x} := \mathbf{X}(\omega) & \xrightarrow{p_j} & x_j := p_j(\mathbf{x}) \end{array}. \quad (1.19)$$

La plus petite tribu de  $\mathbb{R}^d$  qui rend toutes les projections  $p_j$  mesurables est appelée la tribu produit des tribus boréliennes. Elle est notée  $\mathcal{B}(\mathbb{R}^d) \otimes \dots \otimes \mathcal{B}(\mathbb{R}^d)$  et il se trouve qu'elle coïncide exactement avec  $\mathcal{B}(\mathbb{R}^d)$ . Les applications coordonnées définies par  $X_j := p_j \circ \mathbf{X}$  sont donc bien mesurables et sont des variables aléatoires au sens de la tribu  $\mathcal{A}$ . On extrait de  $\mu_{\mathbf{X}}$  une mesure de probabilité pour chaque variable  $X_j$  :

$$\begin{aligned} \forall B \in \mathcal{B}(\mathbb{R}), \mu_{X_j}(B) &:= \mathbb{P}(X_j \in B) = \mathbb{P}(\mathbf{X} \in \mathbb{R} \times \dots \times \mathbb{R} \times B \times \mathbb{R} \times \dots \times \mathbb{R}) \\ &= \mu_{\mathbf{X}}(\mathbb{R} \times \dots \times \mathbb{R} \times B \times \mathbb{R} \times \dots \times \mathbb{R}) . \end{aligned} \quad (1.20)$$

La mesure  $\mu_{X_j}$  représente la loi marginale de la variable aléatoire  $X_j$ . Par opposition, la loi de  $\mathbf{X}$  est souvent qualifiée de loi conjointe car elle régit le comportement simultanée des variables. Les fonctions de répartition marginales  $F_{X_j}$  sont déduites de la fonction de répartition conjointe  $F_{\mathbf{X}}$  comme suit :

$$\begin{aligned} \forall x_j \in \mathbb{R}, F_{X_j}(x_j) &:= \mu_{X_j}([\!-\infty, x_j]) = \mu_{\mathbf{X}}(\mathbb{R} \times \dots \times \mathbb{R} \times [\!-\infty, x_j] \times \mathbb{R} \times \dots \times \mathbb{R}) \\ &= F_{\mathbf{X}}(+\infty, \dots, +\infty, x_j, +\infty, \dots, +\infty) . \end{aligned} \quad (1.21)$$

En appliquant l'expression (1.15) dans l'équation (1.21), il vient que :

$$\begin{aligned} \forall x_j \in \mathbb{R}, F_{X_j}(x_j) &= \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \int_{-\infty}^{x_j} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} f_{\mathbf{X}}(z_1, \dots, z_d) dz_1 \dots dz_d \\ &= \int_{-\infty}^{x_j} \left( \int_{\mathbb{R}^{d-1}} f_{\mathbf{X}}(\mathbf{z}) dz_1 \dots dz_{j-1} dz_{j+1} \dots dz_d \right) dz_j = \int_{-\infty}^{x_j} f_{X_j}(z_j) dz_j , \end{aligned} \quad (1.22)$$

ce qui nous permet d'identifier les densités de probabilité marginales  $f_{X_j}$  :

$$\forall x_j \in \mathbb{R}, f_{X_j}(x_j) := \int_{\mathbb{R}^{d-1}} f_{\mathbf{X}}(z_1, \dots, z_{j-1}, x_j, z_{j+1}, \dots, z_d) dz_1 \dots dz_{j-1} dz_{j+1} \dots dz_d . \quad (1.23)$$

Chaque variable aléatoire  $X_j$  dispose donc de sa propre loi marginale  $\mu_{X_j}$  dont une caractérisation est offerte par la fonction de répartition marginale  $F_{X_j}$  ou la densité de probabilité marginale  $f_{X_j}$ . En général, la seule connaissance des lois marginales ne suffit pas à reconstituer la loi conjointe. En effet, celle-ci peut incorporer une structure de dépendance. C'est une information supplémentaire qui s'ajoute aux tendances marginales et qui doit faire l'objet d'une caractérisation spécifique.

## 1.2.5 Structure de dépendance

Un point important dans la caractérisation de la loi  $\mu_{\mathbf{X}}$  est la description du rapport que les variables  $X_j$  entretiennent entre elles. Si elles sont mutuellement indépendantes, la juxtaposition des lois marginales suffit à décrire la loi conjointe.

### Définition 1.2.1 (Variables indépendantes)

Soit  $\mathbf{X} = [X_j]_{1 \leq j \leq d}$  un vecteur aléatoire à valeurs dans  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ . Les variables aléatoires  $X_1, \dots, X_d$  sont dites (mutuellement) indépendantes si et seulement si :

$$\forall B := B_1 \times \dots \times B_d \in \mathcal{B}(\mathbb{R}^d), \mu_{\mathbf{X}}(B) = \mathbb{P}(\mathbf{X} \in B) = \prod_{j=1}^d \mathbb{P}(X_j \in B_j) = \prod_{j=1}^d \mu_{X_j}(B_j) . \quad (1.24)$$

Dans le cas particulier où  $B \in \mathcal{P}_2$ , l'indépendance des variables permet d'écrire que :

$$\forall \mathbf{x} \in \mathbb{R}^d, F_{\mathbf{X}}(\mathbf{x}) = \mu_{\mathbf{X}}([\!-\infty, x_1] \times \dots \times [\!-\infty, x_d]) = \prod_{j=1}^d \mu_{X_j}([\!-\infty, x_j]) = \prod_{j=1}^d F_{X_j}(x_j) . \quad (1.25)$$

En utilisant les expressions intégrales des fonctions de répartition puis en appliquant le théorème de Fubini, on obtient :

$$\forall \mathbf{x} \in \mathbb{R}^d, \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} f_{\mathbf{X}}(\mathbf{z}) d\mathbf{z} = \prod_{j=1}^d \left( \int_{-\infty}^{x_j} f_{X_j}(z_j) dz_j \right) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} \left( \prod_{i=1}^d f_{X_i}(z_i) \right) d\mathbf{z} , \quad (1.26)$$

ce qui finit par donner :

$$\forall \mathbf{x} \in \mathbb{R}^d, f_{\mathbf{X}}(\mathbf{x}) = \prod_{j=1}^d f_{X_j}(x_j). \quad (1.27)$$

Les équations (1.25) et (1.27) montrent que la loi conjointe s'écrit à partir du produit des seules lois marginales dès lors qu'on fait une hypothèse d'indépendance sur les coordonnées de  $\mathbf{X}$ . Sans cette hypothèse, il faut introduire un outil de caractérisation de la structure de dépendance. Pour un ensemble de lois marginales données, on peut se demander dans quel espace mathématique vivent toutes les lois conjointes pouvant les engendrer. Du point de vue des fonctions de répartition, un tel espace est appelé une classe de Fréchet. À toute collection  $F_{X_1}, \dots, F_{X_d}$  de fonctions de répartition marginales, on peut associer sa classe de Fréchet. Elle est toujours non vide car elle contient au moins le produit  $F_{X_1} \times \dots \times F_{X_d}$  obtenu lorsque les variables sont indépendantes. Les premiers travaux consacrés à l'étude des structures de dépendance datent de la seconde moitié des années 1930 avec notamment FRÉCHET [1935] et Hoeffding [1940]. Une approche unifiée voit ensuite le jour grâce au développement de la théorie des copules. Quelques résultats importants sont rappelés dans la suite de cette section. Pour disposer d'un aperçu plus complet, on pourra se référer à DURANTE et SEMPI [2010] ou aux ouvrages de JOE [1997] et NELSEN [2007].

### Définition 1.2.2 (Copule)

Soit  $\mathbf{U} = [U_j]_{1 \leq j \leq d}$  un vecteur aléatoire à valeurs dans  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ .  
On suppose que toutes ses lois marginales  $\mu_{U_j}$  sont uniformes :

$$U_j \sim \mathcal{U}([0, 1]) \quad \text{i.e.} \quad \forall u_j \in \mathbb{R}, f_{U_j}(u_j) = \mathbb{1}_{[0, 1]}(u_j). \quad (1.28)$$

La loi conjointe  $\mu_{\mathbf{U}}$  est une  $d$ -copule.

Si  $\mu_{\mathbf{U}}$  est une  $d$ -copule, sa fonction de répartition et sa densité de probabilité sont respectivement notées  $C_{\mathbf{U}}$  et  $c_{\mathbf{U}}$ . Par abus de langage, le terme "copule" désigne souvent  $C_{\mathbf{U}}$ . Dans la suite, on note  $\mathcal{C}_d$  l'ensemble des fonctions de répartition des  $d$ -copules. Par ailleurs, on appelle  $d$ -cube ou hypercube  $d$ -dimensionnel le pavé de  $\mathbb{R}^d$  défini par :

$$\mathcal{H}_d := [0, 1] \times \dots \times [0, 1] = [0, 1]^d. \quad (1.29)$$

Pour tout pavé  $P \in \mathcal{P}_1$ , on peut remarquer que :

$$\mu_{\mathbf{U}}(P \setminus \mathcal{H}_d) = 0 \quad \text{et} \quad \mu_{\mathbf{U}}(P) = \mu_{\mathbf{U}}(P \cap \mathcal{H}_d). \quad (1.30)$$

Ainsi, connaître  $\mu_{\mathbf{U}}$  pour les seuls pavés inclus dans  $\mathcal{H}_d$  suffit à complètement caractériser la loi de  $\mathbf{U}$ . On dit que  $\mathcal{H}_d$  est le support de  $\mathbf{U}$  et les domaines de définition de  $C_{\mathbf{U}}$  et  $c_{\mathbf{U}}$  y sont restreints. Les propriétés (F1) et (F2) deviennent :

- $C_{\mathbf{U}}(u_1, \dots, u_d) = 0$  dès lors que  $\exists j : u_j = 0$ .
- $C_{\mathbf{U}}(1, \dots, 1) = 1$ .

D'un point de vue purement fonctionnel, et non plus probabiliste, la caractérisation d'une copule est moins fastidieuse que celle d'une fonction de répartition quelconque.

### Théorème 1.2.4 (Caractérisation d'une copule)

La fonction  $C : \mathcal{H}_d \rightarrow [0, 1]$  est une copule si et seulement si :

- (C1)  $\forall u_j \in [0, 1], C(1, \dots, 1, u_j, 1, \dots, 1) = u_j$ .
- (C2)  $C$  est isotonique.
- (C3)  $C$  est  $d$ -croissante.

En particulier, le caractère isotonique (C2) donne :

$$\forall \mathbf{u} \in \mathcal{H}_d, C(u_1, \dots, u_{j-1}, 0, u_{j+1}, \dots, u_d) \leq C(1, \dots, 1, 0, 1, \dots, 1) = 0. \quad (1.31)$$

Par ailleurs, en combinant la  $d$ -croissance (C3) et les conditions aux bords (C1), on montre que  $C$  est une application lipschitzienne :

$$\forall \mathbf{u}, \mathbf{v} \in \mathcal{H}_d, |C(\mathbf{u}) - C(\mathbf{v})| \leq \|\mathbf{u} - \mathbf{v}\|_1 := \sum_{j=1}^d |u_j - v_j|. \quad (1.32)$$

Par conséquent,  $C$  est continue, ce qui est plus fort que (F4) qui n'impose que la continuité à droite dans chaque direction et en tout point. L'utilité des copules devient fondamentale dans l'étude des distributions multidimensionnelles grâce à un résultat de SKLAR [1959].

### **Théorème 1.2.5 (Sklar)**

Soit  $\mathbf{X}$  un vecteur aléatoire à valeurs dans  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ .  
On note  $F_{\mathbf{X}}$  sa fonction de répartition et on suppose qu'il existe une densité de probabilité  $f_{\mathbf{X}}$ .  
Il existe une unique copule  $C \in \mathcal{C}_d$  telle que :

$$\forall \mathbf{x} \in \mathbb{R}^d, F_{\mathbf{X}}(\mathbf{x}) = C(F_1(x_1), \dots, F_d(x_d)). \quad (1.33)$$

Sous l'hypothèse d'existence de  $f_{\mathbf{X}}$ , toutes les fonctions de répartition marginales sont inversibles, car continues et strictement croissantes. En remarquant que  $u_j = F_{X_j}[F_{X_j}^{-1}(u_j)]$ , on peut voir que :

$$\forall \mathbf{u} \in \mathcal{H}_d, C(\mathbf{u}) = \mathbb{P}(X_1 \leq F_{X_1}^{-1}(u_1), \dots, X_d \leq F_{X_d}^{-1}(u_d)) = \mathbb{P}(F_{X_1}(X_1) \leq u_1, \dots, F_{X_d}(X_d) \leq u_d). \quad (1.34)$$

$C$  est donc la copule associée au vecteur aléatoire  $\mathbf{U}$  constitué des variables  $U_j = F_{X_j}(X_j)$  qui sont toutes uniformes mais corrélées entre elles. Pour cette raison, on notera désormais  $C_{\mathbf{U}}$  la copule  $C$  utilisée ci-dessus pour exprimer la fonction de répartition conjointe  $F_{\mathbf{X}}$ . Le Théorème 1.2.5 est primordial dans l'étude des distributions multidimensionnelles car il permet une stricte séparation des tendances marginales et de la structure de dépendance. La copule  $C_{\mathbf{U}}$  peut être vue comme une version simplifiée de la fonction de répartition  $F_{\mathbf{X}}$  une fois que celle-ci est délestée de toute l'information propre aux lois marginales. Une conséquence importante réside alors dans le fait que les copules présentent certaines particularités mathématiques facilitant leur modélisation en grande dimension. Certains développements dans cette direction sont par exemple présentés dans KUROWICKA et COOKE [2006]. L'avènement de la théorie des copules a permis une démultiplication de la palette des modèles de distributions. Ainsi, la réunion d'une copule  $C : \mathcal{H}_d \rightarrow [0, 1]$  et d'une collection de fonctions de répartition inversibles  $F_j : \mathbb{R} \rightarrow [0, 1]$  donne immédiatement naissance à une fonction de répartition conjointe :

$$\begin{aligned} \forall \mathbf{x} \in \mathbb{R}^d, C(F_1(x_1), \dots, F_d(x_d)) &= \mathbb{P}(U_1 \leq F_1(x_1), \dots, U_d \leq F_d(x_d)) \\ &= \mathbb{P}(F_1^{-1}(U_1) \leq x_1, \dots, F_d^{-1}(U_d) \leq x_d) = F_{\mathbf{X}}(\mathbf{x}), \end{aligned} \quad (1.35)$$

où  $\mathbf{X}$  est obtenu en appliquant les fonctions  $F_j^{-1}$  aux coordonnées du vecteur  $\mathbf{U}$  émanant de  $C$ . Réciproquement, pour toute fonction de répartition conjointe  $F_{\mathbf{X}}$  dont les marginales sont inversibles, on peut extraire la copule à l'aide de la formule suivante :

$$\forall \mathbf{u} \in \mathcal{H}_d, C_{\mathbf{U}}(\mathbf{u}) = F_{\mathbf{X}}\left(F_{X_1}^{-1}(u_1), \dots, F_{X_d}^{-1}(u_d)\right). \quad (1.36)$$

### **Théorème 1.2.6 (Invariance)**

Soit  $\mathbf{X}$  un vecteur aléatoire à valeurs dans  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ .  
On définit  $\tilde{\mathbf{X}} = [T_j(X_j)]_{1 \leq j \leq d}$  où les applications  $T_j$  sont continues et strictement croissantes.  
La copule  $\mu_{\mathbf{U}}$  du vecteur  $\mathbf{X}$  est également la copule du vecteur  $\tilde{\mathbf{X}}$ .

Ainsi, toute transformation bijective d'une variable  $X_j$  de  $\mathbf{X}$  affecte la loi marginale  $\mu_{X_j}$  mais laisse la structure de dépendance inchangée. La preuve vient immédiatement en remarquant que :

$$\begin{aligned} \forall \mathbf{x} \in \mathbb{R}^d, F_{\tilde{\mathbf{X}}}(\mathbf{x}) &= F_{\mathbf{X}}(T_1^{-1}(x_1), \dots, T_d^{-1}(x_d)) \\ &= C_{\mathbf{U}}(F_{X_1} \circ T_1^{-1}(x_1), \dots, F_{X_d} \circ T_d^{-1}(x_d)) \\ &= C_{\mathbf{U}}(F_{\tilde{X}_1}(x_1), \dots, F_{\tilde{X}_d}(x_d)). \end{aligned} \quad (1.37)$$

Un autre résultat important est obtenu en dérivant (1.33) par rapport à chacune de ses variables :

$$\forall \mathbf{x} \in \mathbb{R}^d, f_{\mathbf{X}}(\mathbf{x}) := c_{\mathbf{U}}(F_{X_1}(x_1), \dots, F_{X_d}(x_d)) \times \prod_{j=1}^d f_{X_j}(x_j) \quad \text{avec} \quad c_{\mathbf{U}} := \frac{\partial^d C_{\mathbf{U}}}{\partial u_1 \dots \partial u_d}. \quad (1.38)$$

La densité conjointe  $f_{\mathbf{X}}$  s'exprime donc comme le produit des densités marginales  $f_{X_j}$  auquel s'ajoute un terme dépendant de la densité de copule  $c_{\mathbf{U}}$  et devant être compris comme une pondération qui permet de prendre en compte la nature de la structure de dépendance.

**Théorème 1.2.7 (Bornes de Fréchet-Hoeffding)**

Soit  $C \in \mathcal{C}_d$ . On a alors les inégalités suivantes :

$$\forall \mathbf{u} \in \mathcal{H}_d, W_d(\mathbf{u}) \leq C(\mathbf{u}) \leq M_d(\mathbf{u}) \quad \text{avec} \quad \begin{cases} W_d(\mathbf{u}) := \max\left(0, \sum_{j=1}^d u_j - d + 1\right) \\ M_d(\mathbf{u}) := \min_{1 \leq j \leq d} u_j \end{cases}. \quad (1.39)$$

Attribué à FRÉCHET [1935] et Hoeffding [1940], ce résultat d'encadrement découle de calculs probabilistes élémentaires. Pour l'inégalité de droite, on peut voir que :

$$C(\mathbf{u}) = \mathbb{P}(U_1 \leq u_1, \dots, U_d \leq u_d) \leq \mathbb{P}(U_j \leq u_j) \implies C(\mathbf{u}) \leq \min_{1 \leq j \leq d} u_j. \quad (1.40)$$

Pour l'inégalité de gauche, on utilise un argument classique de majoration :

$$C(\mathbf{u}) = 1 - \mathbb{P}(\{U_1 > u_1\} \cup \dots \cup \{U_d > u_d\}) \implies C(\mathbf{u}) \geq 1 - \sum_{j=1}^d \mathbb{P}(\{U_j > u_j\}) = \sum_{j=1}^d u_j - d + 1. \quad (1.41)$$

- $M_d \in \mathcal{C}_d$  est la copule *comonotone*. Elle est obtenue dans le cas où  $U_1 = \dots = U_d$ . C'est une situation de dépendance totale par égalité des variables. Il existe donc une copule  $M_d \in \mathcal{C}_d$  faisant office de borne supérieure uniformément sur  $\mathcal{H}_d$  :

$$\forall \mathbf{u} \in \mathcal{H}_d, \sup_{C \in \mathcal{C}_d} C(\mathbf{u}) = \max_{C \in \mathcal{C}_d} C(\mathbf{u}) = M_d(\mathbf{u}). \quad (1.42)$$

- $W_2 \in \mathcal{C}_2$  est la copule *anti-monotone*. Elle est obtenue dans le cas où  $U_2 = 1 - U_1$ . C'est une situation de dépendance totale par complémentarité des variables. Par contre, dès que l'on a  $d \geq 3$ ,  $W_d \notin \mathcal{C}_d$  et la situation ne peut plus être interprétée à partir du vecteur  $\mathbf{U}$ . Cependant, FRÉCHET [1935] a montré que la borne  $W_d$  reste optimale au sens où :

$$\forall \mathbf{u} \in \mathcal{H}_d, \inf_{C \in \mathcal{C}_d} C(\mathbf{u}) = W_d(\mathbf{u}). \quad (1.43)$$

Les bornes de Fréchet-Hoeffding définissent une enveloppe fonctionnelle au sein de l'espace  $\mathcal{C}_d$ . En plus de  $M_d$  et  $W_d$ , une troisième copule d'intérêt est celle issue du cas où les variables de  $\mathbf{U}$  sont mutuellement indépendantes. On a alors :

$$\Pi_d(\mathbf{u}) = \mathbb{P}(U_1 \leq u_1, \dots, U_d \leq u_d) = \prod_{j=1}^d \mathbb{P}(U_j \leq u_j) = \prod_{j=1}^d u_j. \quad (1.44)$$

Dans la suite, on note  $\pi_d$  la densité de copule associée à  $\Pi_d$ . Elle est constamment égale à 1 sur  $\mathcal{H}_d$ . L'existence des bornes de Fréchet-Hoeffding ne doit pas nous laisser croire qu'il existe une relation d'ordre total sur  $\mathcal{C}_d$ . Il est vrai qu'on peut trouver des copules  $C_1, C_2 \in \mathcal{C}_d$  telles que l'inégalité  $C_1(\mathbf{u}) \leq C_2(\mathbf{u})$  reste vraie partout sur  $\mathcal{H}_d$ . On note alors  $C_1 < C_2$ . Par exemple,  $\Pi_d < M_d$ . On se place maintenant dans le cas où  $d = 2$  pour exhiber un contre-exemple. Si on regarde la fonction  $\tilde{C} := (W_2 + M_2)/2$ , qui est une copule par combinaison convexe de deux autres copules, il n'est pas

possible de la comparer à  $\Pi_2$  uniformément sur  $\mathcal{H}_2$ . On peut en effet trouver deux points distincts  $\mathbf{u} \neq \mathbf{v} \in \mathcal{H}_2$  qui vont être tels que :

$$\tilde{C}(\mathbf{u}) < \Pi(\mathbf{u}) \quad \text{et} \quad \tilde{C}(\mathbf{v}) > \Pi(\mathbf{v}) . \quad (1.45)$$

Pour pouvoir établir une relation d'ordre total, il faudrait se restreindre à un sous-espace de  $\mathcal{C}_d$ . Il est ainsi montré dans NELSEN [1997] que certaines familles paramétriques peuvent le permettre. Même sans possibilité de définir une relation d'ordre total, l'espace  $\mathcal{C}_2$  est articulé autour de trois tendances : égalité des variables ( $M_2$ ), indépendance des variables ( $\Pi_2$ ) et antagonisme des variables ( $W_2$ ). Certaines copules s'écartent de l'indépendance par accroissement de la dépendance "positive", c'est-à-dire par augmentation des chances que la variable  $U_2$  prenne une valeur proche de 1 lorsqu'on sait déjà que l'observation  $u_1$  de  $U_1$  est proche de 1. Ces copules sont dites à dépendance positive par quadrant (DPQ) et vérifient :

$$\forall \mathbf{u} \in \mathcal{H}_d, \Pi_2(\mathbf{u}) \leq C(\mathbf{u}) \leq M_2(\mathbf{u}) . \quad (1.46)$$

On les oppose aux copules à dépendance négative par quadrant (DNQ) pour lesquelles on a :

$$\forall \mathbf{u} \in \mathcal{H}_d, W_2(\mathbf{u}) \leq C(\mathbf{u}) \leq \Pi_2(\mathbf{u}) . \quad (1.47)$$

Divers exemples illustrant ces concepts de dépendance sont examinés dans LAI et XIE [2003]. La situation peut être résumée par le schéma de la Figure 1.1 qui met en perspective le Théorème 1.2.7 et les équations (1.46) et (1.47). Les mêmes idées peuvent être étendues en plus grande dimension avec la notion de dépendance positive (ou négative) par orthant, détaillée dans JOE [1997].

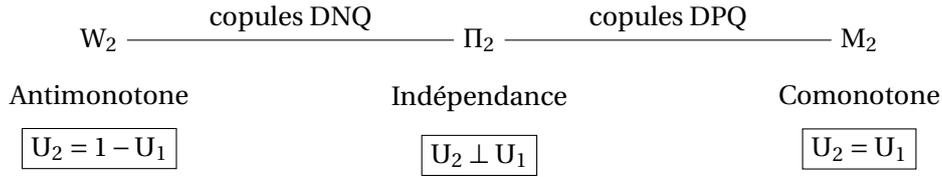


FIGURE 1.1 – Représentation schématique de l'ensemble  $\mathcal{C}_2$  .

## 1.2.6 Fonction caractéristique

Une autre caractérisation de la loi de  $\mathbf{X}$  est fournie par la fonction caractéristique  $\phi_{\mathbf{X}}$  :

$$\begin{aligned} \phi_{\mathbf{X}}: \mathbb{R}^d &\longrightarrow \mathbb{R} \\ \mathbf{t} &\longmapsto \mathbb{E} \left[ \exp(i \mathbf{t}^T \mathbf{X}) \right] . \end{aligned} \quad (1.48)$$

Cette fonction est parfaitement définie sur  $\mathbb{R}^d$  car  $|\exp(i \mathbf{t}^T \mathbf{X})| \leq 1$ . Calculer  $\phi_{\mathbf{X}}$  revient à prendre la transformée de Fourier de la densité conjointe  $f_{\mathbf{X}}$ . De même que précédemment, si deux fonctions caractéristiques  $\phi_{\mathbf{X}}$  et  $\phi_{\tilde{\mathbf{X}}}$  coïncident sur  $\mathbb{R}^d$ , alors  $\mathbf{X}$  et  $\tilde{\mathbf{X}}$  sont de même loi. L'utilisation de cette transformée de  $f_{\mathbf{X}}$  est d'une grande aide dans de nombreux contextes probabilistes (calcul de la loi d'une somme de variables, calcul de moments, justification de convergences, inégalités de concentration, manipulation de chaînes de Markov). Un large éventail d'applications est présenté dans FLORESCU et TUDOR [2013].

## 1.2.7 Statistiques issues d'une distribution multidimensionnelle

Les paragraphes précédents ont montré que toute caractérisation opérationnelle de la loi du vecteur  $\mathbf{X}$  prend la forme d'une fonction (réelle ou complexe) définie sur  $\mathbb{R}^d$ . Grâce au calcul de statistiques, on peut extraire un résumé en dimension finie de la loi  $\mu_{\mathbf{X}}$ . Les quantités calculées peuvent cibler une tendance marginale (valeur moyenne, dispersion, niveau d'asymétrie), viser la structure de dépendance (corrélations, concordances, dépendances extrémales) mais également impliquer ces deux entités.

### 1.2.7.1 Moments

On suppose que  $\mathbf{X}$  est intégrable, c'est-à-dire que  $\mathbb{E}[\|\mathbf{X}\|_1] < \infty$  avec  $\|\cdot\|_1$  la norme 1 pour les vecteurs de  $\mathbb{R}^d$ . Une idée de la position moyenne de  $\mathbf{X}$  est alors donnée par le vecteur moyen  $\mathbf{m}_{\mathbf{X}}$  qui regroupe les espérances des variables  $X_j$  :

$$\mathbf{m}_{\mathbf{X}} := \mathbb{E}[\mathbf{X}] = [m_j]_{1 \leq j \leq d} \in \mathbb{R}^d \quad \text{avec} \quad m_j := \mathbb{E}[X_j] = \int_{\mathbb{R}} x_j f_{X_j}(x_j) dx_j. \quad (1.49)$$

C'est le principal indicateur de tendance centrale. En revanche, la variabilité simultanée des composantes du vecteur  $\mathbf{X}$  ne peut pas être résumée par le seul vecteur des variances. Celui-ci se contente de juxtaposer les dispersions marginales et occulte le rôle de la structure de dépendance. La densité conjointe  $f_{\mathbf{X}}$  résulte en réalité de l'agrégation de toutes les sous-structures de dépendance pouvant être constituées en agrégeant un minimum de deux variables. Une simplification assez naturelle consiste à ne quantifier que les dépendances entre paires de variables. À ce titre, la matrice de covariance  $\Sigma_{\mathbf{X}}$  propose un premier aperçu de la structure de dépendance pour lequel chaque coefficient non diagonal mesure l'intensité du lien de dépendance au sein d'une paire de variables. En supposant que  $\mathbb{E}[\|\mathbf{X}\|_2^2] < \infty$  avec  $\|\cdot\|_2$  la norme euclidienne,  $\Sigma_{\mathbf{X}}$  est définie par :

$$\Sigma_{\mathbf{X}} := \mathbb{E} \left[ (\mathbf{X} - \mathbf{m}_{\mathbf{X}})(\mathbf{X} - \mathbf{m}_{\mathbf{X}})^{\top} \right] = [\Gamma_{kl}]_{\substack{1 \leq k \leq d \\ 1 \leq l \leq d}} \in \mathbb{R}^{d \times d},$$

avec :

$$\begin{cases} \forall j, & \Gamma_{jj} = \mathbb{V}(X_j) = \mathbb{E}[(X_j - m_j)^2] = \mathbb{E}[X_j^2] - m_j^2 \\ \forall k \neq l, & \Gamma_{kl} = \text{Cov}(X_k, X_l) = \mathbb{E}[(X_k - m_k)(X_l - m_l)] = \mathbb{E}[X_k X_l] - m_k m_l \end{cases}. \quad (1.50)$$

$\Sigma_{\mathbf{X}}$  est une matrice symétrique semi-définie positive. Connaître le vecteur moyen  $\mathbf{m}_{\mathbf{X}}$  et la matrice de covariance  $\Sigma_{\mathbf{X}}$  est généralement insuffisant pour caractériser la loi  $\mu_{\mathbf{X}}$  sauf si  $\mathbf{X}$  est un vecteur gaussien. Dans ce cas, sous réserve que  $\Sigma_{\mathbf{X}}$  soit définie positive,  $\mathbf{X}$  admet la densité suivante :

$$f_{\mathcal{N}}(\mathbf{x} | \mathbf{m}_{\mathbf{X}}, \Sigma_{\mathbf{X}}) = \left( \frac{1}{\sqrt{2\pi}} \right)^d \left( \frac{1}{\sqrt{\det \Sigma_{\mathbf{X}}}} \right) \exp \left[ -\frac{1}{2} (\mathbf{x} - \mathbf{m}_{\mathbf{X}})^{\top} \Sigma_{\mathbf{X}}^{-1} (\mathbf{x} - \mathbf{m}_{\mathbf{X}}) \right]. \quad (1.51)$$

### 1.2.7.2 Coefficients de corrélation linéaire

On vient de voir que la matrice  $\Sigma_{\mathbf{X}}$  propose de mesurer les dépendances au sein des paires de variables à l'aide de l'opérateur de covariance. D'autres mesures existent et remplissent mieux cette fonction. On considère ici une paire  $(X_1, X_2)$  de loi résumée comme suit :

$$\mathbb{V}(x_1, x_2) \in \mathbb{R}^2, F_{12}(x_1, x_2) := \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) = C_{12}(F_1(x_1), F_2(x_2)). \quad (1.52)$$

L'inégalité de Cauchy-Schwarz appliquée à la covariance donne :

$$|\text{Cov}(X_1, X_2)| \leq \sigma_1 \sigma_2 \quad \text{avec} \quad \sigma_i = \sqrt{\mathbb{V}(X_i)} \quad \forall i \in \{1, 2\}. \quad (1.53)$$

Le cas d'égalité au produit des écarts-types est obtenu dans les cas limites de comonotonie et d'antimonotonie. Pour disposer d'un indicateur dont les bornes ne dépendent pas de la loi sous-jacente, on préfère utiliser le coefficient de corrélation linéaire de **PEARSON [1920]** :

$$\rho_{12} := \rho(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sigma_1 \sigma_2}. \quad (1.54)$$

Ainsi le coefficient  $\rho_{12}$  reste confiné dans l'intervalle  $[-1, 1]$ . Néanmoins, la covariance présente de nombreux autres défauts auxquels la renormalisation par le produit des écarts-types ne change rien. Ils sont énumérés dans **MEISSNER [2013]** et on choisit d'en mentionner quelques-uns.

- Il est indispensable d'avoir  $\mathbb{E}[\|\mathbf{X}\|_2^2] < \infty$ .
- Une information de type  $\rho_{12} = 0$ , synonyme de décorrélation, n'induit pas l'indépendance entre  $X_1$  et  $X_2$ . On peut en effet observer une parfaite décorrélation dans une situation où l'une des variables est une fonction quadratique de l'autre.

- La valeur prise par  $\rho_{12}$  dépend des lois marginales alors qu'on souhaite accéder à une information caractérisant la seule copule. **HOEFFDING** [1940] fait notamment remarquer que :

$$\rho_{12} = \frac{1}{\sigma_1 \sigma_2} \int_{\mathcal{H}_2} F_1^{-1}(u_1) F_2^{-1}(u_2) \times (c_{12}(u_1, u_2) - \pi_2(u_1, u_2)) du_1 du_2. \quad (1.55)$$

D'une certaine manière, on mesure l'écart entre la dépendance réelle et une situation fictive d'indépendance des deux variables. La différence entre la densité de copule  $c_{12}$  et la densité  $\pi_2$  de la copule indépendance bidimensionnelle est intégrée sur  $\mathcal{H}_2$  après pondération par une fonction positive liée aux lois marginales. À copule fixée, la valeur prise par  $\rho_{12}$  change si l'une des marginales est modifiée.

- Une conséquence du point précédent est la sensibilité de  $\rho_{12}$  à tout changement d'échelle opéré sur l'une des variables. Le seul cas d'invariance concerne les transformations affines :

$$\tilde{X}_1 = aX_1 + b \text{ avec } a > 0 \implies \rho(\tilde{X}_1, X_2) = \rho(X_1, X_2). \quad (1.56)$$

On note  $\boldsymbol{\rho}_{\mathbf{X}}$  la matrice de corrélation obtenue à partir de  $\boldsymbol{\Sigma}_{\mathbf{X}}$ . Dans certains modèles de distributions, le système de paramétrage s'appuie sur la corrélation de  $X_1$  et  $X_2$  conditionnellement aux valeurs prises par les variables  $X_3, \dots, X_d$  que l'on appelle alors des covariables et qui sont rassemblées dans un vecteur  $\tilde{\mathbf{X}}$ . Pour tout  $\tilde{\mathbf{x}} \in \mathbb{R}^{d-2}$ , le coefficient de corrélation conditionnel vaut :

$$\rho_{12|3\dots d}(\tilde{\mathbf{x}}) := \rho(X_1, X_2 | \tilde{\mathbf{X}} = \tilde{\mathbf{x}}) = \frac{\text{Cov}(X_1, X_2 | \tilde{\mathbf{X}} = \tilde{\mathbf{x}})}{\sigma(X_1 | \tilde{\mathbf{X}} = \tilde{\mathbf{x}}) \sigma(X_2 | \tilde{\mathbf{X}} = \tilde{\mathbf{x}})}. \quad (1.57)$$

Évidemment, ce coefficient ne peut pas être estimé à partir de la connaissance d'un échantillon d'observations conjointes car les covariables prennent des valeurs différentes à chaque nouvelle observation. Le cas gaussien fait exception car la corrélation conditionnelle est constamment égale à une valeur  $\rho_{12|3\dots d}$  ne dépendant pas de  $\tilde{\mathbf{x}}$ . Toute corrélation conditionnelle peut alors être calculée à partir des seuls coefficients de la matrice  $\boldsymbol{\rho}_{\mathbf{X}}$  en utilisant des formules disponibles dans **JOE** [2014]. Dans les autres cas, le coefficient de corrélation partielle de **FISHER** [1924] a été pensé pour mesurer l'influence des variables  $X_3, \dots, X_d$  sur la corrélation de la paire  $(X_1, X_2)$ . Étant donné que les liens de corrélation sont invariants par translation, on peut supposer sans perte de généralité que  $\mathbf{X}$  est un vecteur centré. Ensuite, il faut remarquer que le coefficient  $\rho_{12}$  donné à l'équation (1.54) vérifie l'égalité :

$$\rho_{12} = \text{sgn}(b_{12}) \sqrt{b_{12} b_{21}} \text{ avec } \begin{cases} b_{12} := \underset{b \in \mathbb{R}}{\text{argmin}} \mathbb{E}[(X_1 - bX_2)^2] \\ b_{21} := \underset{b \in \mathbb{R}}{\text{argmin}} \mathbb{E}[(X_2 - bX_1)^2] \end{cases}. \quad (1.58)$$

On précise que  $\text{sgn}(\cdot)$  est la fonction "signe". La droite  $Y = b_{12}X_2$  est la meilleur approximation linéaire de la variable  $X_1$  au sens des moindres carrés. En généralisant l'expression (1.58) de  $\rho_{12}$ , on peut définir la corrélation partielle. Pour cela, on commence par résoudre deux problèmes de régression linéaire :

$$\begin{aligned} (b_{12;34\dots d} \quad b_{13;24\dots d} \quad \dots \quad b_{1d;23\dots d-1}) &= \underset{\mathbf{b} \in \mathbb{R}^{d-1}}{\text{argmin}} \mathbb{E}[(X_1 - b_2X_2 - b_3X_3 - \dots - b_dX_d)^2] \\ (b_{21;34\dots d} \quad b_{23;14\dots d} \quad \dots \quad b_{2d;13\dots d-1}) &= \underset{\mathbf{b} \in \mathbb{R}^{d-1}}{\text{argmin}} \mathbb{E}[(X_2 - b_1X_1 - b_3X_3 - \dots - b_dX_d)^2] \end{aligned}, \quad (1.59)$$

puis le coefficient de corrélation partielle est donné par :

$$\rho_{12;34\dots d} := \text{sgn}(b_{12;34\dots d}) \sqrt{b_{12;34\dots d} b_{21;34\dots d}}. \quad (1.60)$$

Contrairement à la corrélation conditionnelle, la corrélation partielle ne dépend pas d'une valeur prise par les covariables et elle peut être estimée à partir d'un échantillon d'observations conjointes. En outre, les corrélations partielles vérifient la formule de récurrence suivante :

$$\rho_{12;3\dots d} = \frac{\rho_{12;3\dots d-1} - \rho_{1d;3\dots d-1} \rho_{2d;3\dots d-1}}{\sqrt{1 - \rho_{1d;3\dots d-1}^2} \sqrt{1 - \rho_{2d;3\dots d-1}^2}}. \quad (1.61)$$

Il semble évident que les corrélations partielle et conditionnelle poursuivent un même objectif mais sont différentes. Elles peuvent néanmoins coïncider sous certaines conditions énoncées dans [LAWRANCE \[1976\]](#) et dans [BABA et collab. \[2004\]](#).

### 1.2.7.3 Mesures de dépendance

De par ses nombreux défauts, le coefficient de corrélation linéaire ne semble pas être l'outil le plus adapté pour mesurer la dépendance au sens large d'une paire de variables. On aimerait disposer d'un opérateur  $\delta$  qui mesure l'intensité du motif de dépendance intrinsèque d'une paire  $(X_k, X_l)$  de  $\mathbf{X}$ . Un tel opérateur doit être capable d'extraire de l'information sur la seule copule, sans être impacté par la nature des lois marginales. [SCARSINI \[1984\]](#) propose une liste des propriétés à exiger d'un opérateur pour qu'il puisse être qualifié de mesure de dépendance.

#### Définition 1.2.3 (Mesure de dépendance)

Un opérateur bivarié  $\delta$  est une mesure de dépendance s'il vérifie les propriétés suivantes :

(M1) **Domaine de définition** :  $\delta(X_k, X_l)$  existe dès que  $\mu_{X_{kl}}$  est absolument continue.

(M2) **Symétrie** :  $\delta(X_k, X_l) = \delta(X_l, X_k)$ .

(M3) **Cohérence** : Notons  $C_{ij}$  et  $C_{kl}$  les copules des paires  $(X_i, X_j)$  et  $(X_k, X_l)$ .

$$C_{ij} < C_{kl} \implies \delta(X_i, X_j) \leq \delta(X_k, X_l) \quad (1.62)$$

(M4) **Bornes** :  $-1 \leq \delta(X_k, X_l) \leq 1$  avec égalité à 1 (resp.  $-1$ ) si  $C_{kl} = M_2$  (resp.  $C_{kl} = W_2$ ).

(M5) **Indépendance** :  $X_k \perp X_l \implies \delta(X_k, X_l) = 0$ .

(M6) **Changement de signe** :  $\delta(-X_k, X_l) = \delta(X_k, -X_l) = -\delta(X_k, X_l)$ .

(M7) **Continuité** : Notons  $\left\{ (X_k^{(n)}, X_l^{(n)}) \right\}_{n \geq 1}$  une suite de paires de variables.

$$\left( X_k^{(n)}, X_l^{(n)} \right) \xrightarrow[n \rightarrow \infty]{\text{loi}} (X_k, X_l) \implies \lim_{n \rightarrow \infty} \delta \left( X_k^{(n)}, X_l^{(n)} \right) = \delta(X_k, X_l) \quad (1.63)$$

(M8) **Invariance** : Notons  $T_k$  et  $T_l$  des applications continues et strictement croissantes.

$$\delta(T_k(X_k), T_l(X_l)) = \delta(X_k, X_l) \quad (1.64)$$

Au vu des remarques faites précédemment, le coefficient de corrélation linéaire de Pearson ne vérifie pas les axiomes (M1), (M3) et (M8). On introduit ensuite deux exemples de mesure de dépendance, à savoir le tau de [KENDALL \[1948\]](#) et le rho de [SPEARMAN \[1961\]](#).

#### Définition 1.2.4 (Tau de Kendall)

Soient  $(X_k, X_l)$  et  $(\tilde{X}_k, \tilde{X}_l)$  deux copies indépendantes d'une même paire de variables.

Le "tau" de Kendall de cette paire est alors défini par :

$$\tau_{kl} = \tau_K(X_k, X_l) := \mathbb{P} \left( (X_k - \tilde{X}_k)(X_l - \tilde{X}_l) > 0 \right) - \mathbb{P} \left( (X_k - \tilde{X}_k)(X_l - \tilde{X}_l) < 0 \right) . \quad (1.65)$$

L'opérateur  $\tau_K$  s'exprime donc comme la différence entre une première probabilité dite de "concordance" et une seconde dite de "discordance". La concordance désigne une situation de correspondance entre les classements observés pour l'une et l'autre des variables de la paire. Au contraire, la discordance implique l'inversion des classements. Après quelques manipulations élémentaires, on obtient une réécriture intégrale de (1.65) :

$$\begin{aligned} \tau_{kl} &= 2 \times \mathbb{P} \left( (X_k - \tilde{X}_k)(X_l - \tilde{X}_l) > 0 \right) - 1 = 4 \times \mathbb{P} \left( X_k < \tilde{X}_k, X_l < \tilde{X}_l \right) - 1 \\ &= 4 \times \mathbb{E} \left[ F_{kl}(\tilde{X}_k, \tilde{X}_l) \right] - 1 = 4 \int_{\mathcal{H}_2} C_{kl}(u_k, u_l) c_{kl}(u_k, u_l) du_k du_l - 1 . \end{aligned} \quad (1.66)$$

On remarque que  $\tau_{kl}$  ne dépend plus que du produit  $C_{kl} \times c_{kl}$  et donc d'une information propre à la structure de dépendance de la paire  $(X_k, X_l)$ . Par ailleurs, on peut voir que :

$$\tau_{kl} = 4 \int_{\mathcal{H}_2} (C_{kl}(u_k, u_l) c_{kl}(u_k, u_l) - \Pi_2(u_k, u_l) \pi_2(u_k, u_l)) du_k du_l, \quad (1.67)$$

ce qui nous amène, à nouveau, à envisager  $\tau_{kl}$  comme une certaine mesure de l'écart entre le motif de dépendance sous-jacent et l'indépendance. Une discussion plus approfondie à ce sujet est proposée dans [PLISCHKE et BORGONOVO \[2015\]](#). L'opérateur  $\tau_K$  vérifie tous les axiomes de la Définition 1.2.3. Les points (M1), (M2), (M4) et (M6) sont triviaux au regard de l'expression (1.65) tandis que les points (M5) et (M8) se justifient facilement à partir de la réécriture intégrale (1.66). Pour une preuve des points (M3) et (M7), on peut se reporter à [JOE \[2014\]](#).

### Définition 1.2.5 (Rho de Spearman)

Le "rho" de Spearman de la paire  $(X_k, X_l)$  est défini par :

$$s_{kl} := \rho_S(X_k, X_l) = \rho(F_k(X_k), F_l(X_l)) = \rho(U_k, U_l). \quad (1.68)$$

L'opérateur  $\rho_S$  doit donc être compris comme le coefficient de corrélation linéaire des variables initiales rendues uniformes. Ce changement d'échelle opéré sur chaque composante de la paire permet d'éliminer l'influence des lois marginales. Il vient alors immédiatement que :

$$\begin{aligned} s_{kl} &= \frac{\mathbb{E}[U_k U_l] - \frac{1}{4}}{\frac{1}{12}} = 12 \int_{\mathcal{H}_2} u_k u_l c_{kl}(u_k, u_l) du_k du_l - 3 \\ &= 12 \int_{\mathcal{H}_2} C_{kl}(u_k, u_l) du_k du_l - 3 \\ &= 12 \int_{\mathcal{H}_d} (C_{kl}(u_k, u_l) - \Pi_2(u_k, u_l)) du_k du_l. \end{aligned} \quad (1.69)$$

La majorité des remarques faites au sujet de  $\tau_K$  restent valables pour  $\rho_S$ . D'une part, on constate que  $\rho_S$  capture une information caractérisant la seule structure de dépendance. D'autre part, l'interprétation en terme d'écart moyen à l'indépendance est à nouveau vérifiée. Sans surprise,  $\rho_S$  vérifie tous les axiomes. Les points (M1), (M2), (M4), (M5) et (M6) sont évidents en repartant de la définition (1.68) et l'expression (1.69) donne (M3), (M7) et (M8).

Contrairement au coefficient de corrélation linéaire  $\rho$ , le tau de Kendall  $\tau_K$  et le rho de Spearman  $\rho_S$  ne font pas intervenir les lois marginales. Pour cette raison, ces deux mesures de dépendance sont parfois mentionnées comme des mesures de concordance ou des coefficients de corrélation basés sur les rangs. Les relations qu'entretiennent  $\tau_K$  et  $\rho_S$  ont fait l'objet de certains travaux. L'inégalité universelle de [DANIELS \[1950\]](#), revisitée bien plus tard par [GENEST et NEŠLEHOVÁ \[2009\]](#), indique que l'on a toujours  $|3\tau_K - 2\rho_S| \leq 1$ . Par ailleurs, [CAPÉRAÀ et GENEST \[1993\]](#) précisent les hypothèses qu'il faut imposer sur la copule pour avoir  $\rho_S \geq \tau_K \geq 0$  ou au contraire  $\rho_S \leq \tau_K \leq 0$ . Pour certaines applications, la littérature propose des alternatives à  $\tau_K$  et  $\rho_S$ . On peut citer le "gamma" de [GINI \[1914\]](#) ou le "beta" de [BLOMQUIST \[1950\]](#).

Parfois, la dépendance que l'on cherche à mesurer est d'une nature différente. En finance, par exemple, on a souvent besoin d'évaluer la propension d'une paire de variables à engendrer fréquemment des valeurs extrêmes simultanées. On s'intéresse toujours à la structure de dépendance mais seulement dans des régions spécifiques : les queues (inférieure et supérieure) de la distribution conjointe. Cela correspond au voisinage des points  $(0, 0)$  et  $(1, 1)$  de  $\mathcal{H}_2$ . Ces dépendances extrémales sont quantifiées par les coefficients de dépendance de queue. Ces indicateurs sont d'abord apparus en théorie des valeurs extrêmes avec [SIBUYA \[1960\]](#) avant que [JOE \[1993\]](#) ne popularise leur utilisation en théorie des copules.

### Définition 1.2.6 (Coefficients de dépendance de queue)

Soit  $(X_k, X_l)$  une paire de variables de loi conjointe absolument continue.

Les fonctions de répartition  $F_k$  et  $F_l$  sont supposées inversibles.

Sous réserve que les limites existent, on définit les quantités suivantes :

$$\begin{aligned} \bullet \lambda_L &:= \lim_{u \rightarrow 0^+} \mathbb{P}\left(X_k \leq F_k^{-1}(u) \mid X_l \leq F_l^{-1}(u)\right) = \lim_{u \rightarrow 0^+} \frac{C_{kl}(u, u)}{u}, \\ \bullet \lambda_U &:= \lim_{u \rightarrow 1^-} \mathbb{P}\left(X_k > F_k^{-1}(u) \mid X_l > F_l^{-1}(u)\right) = \lim_{u \rightarrow 1^-} \frac{1 - 2u + C_{kl}(u, u)}{1 - u}. \end{aligned}$$

$\lambda_L$  (resp.  $\lambda_U$ ) est le coefficient de dépendance de queue inférieure (resp. supérieure).

Parfois, on adopte les conventions  $\lambda_L = \lambda_{SW}$  (*South West*) et  $\lambda_U = \lambda_{NE}$  (*North East*).

Les notations  $\lambda_L$  et  $\lambda_U$  renvoient aux terminologies anglo-saxonnes, à savoir *lower tail dependence coefficient* et *upper tail dependence coefficient*. Par construction, les coefficients de queue sont positifs ou nuls. Certains des axiomes de SCARSINI [1984] ne sont donc pas vérifiés. C'est notamment le cas des points (M4) et (M6). Si aucune hypothèse de régularité n'est faite sur la copule, rien ne garantit l'existence des limites. Le point (M1) est également sujet à discussion. Il existe des cas pathologiques, comme celui mis en évidence par KORTSCHAK et ALBRECHER [2009], où la limite n'existe pas. Cependant, EMBRECHTS et HOFERT [2013] établissent un résultat d'existence sous des hypothèses assez raisonnables. Ainsi, les coefficients  $\lambda_L$  et  $\lambda_U$  peuvent être calculés pour la plupart des familles paramétriques existantes. Dans le même esprit que les coefficients de dépendance de queue, d'autres mesures susceptibles de privilégier les dépendances extrémales ont été développées par KRUPSKII et JOE [2015] et LEE et collab. [2018].

Pour toute paire  $(X_k, X_l)$ , il existe donc une large gamme de statistiques permettant de mesurer l'intensité de la dépendance entre les deux variables (le coefficient de corrélation linéaire, les coefficients basés sur les rangs, les coefficients de dépendance de queue). Toute matrice de type :

$$\mathbf{S} := [S_{kl}]_{\substack{1 \leq k \leq d \\ 1 \leq l \leq d}} \in \mathbb{R}^{d \times d} \quad \text{avec} \quad \begin{cases} \forall k, & S_{kk} = 1 \\ \forall k \neq l, & S_{kl} = \delta(X_k, X_l) \end{cases}, \quad (1.70)$$

constitue ainsi un résumé de la structure de dépendance au sein de  $\mu_{\mathbf{X}}$ . Chacun des  $C_d^2 = d(d-1)/2$  coefficients de la partie triangulaire inférieure de  $\mathbf{S}$  correspond à une paire de variables et donc à un motif bidimensionnel au sein de la structure de dépendance globale. On pourrait aussi vouloir quantifier la dépendance inhérente à chacun des  $C_d^3$  triplets de variables. À ce titre, WOLFF [1980] et NELSEN [2002] proposent des extensions multidimensionnelles de  $\tau_K$  et  $\rho_S$ . TAYLOR [2007] a adapté le travail de SCARSINI [1984] afin d'énoncer des axiomes pour un nombre quelconque de variables. Par ailleurs, les coefficients de dépendance de queue ont été généralisés en plus grande dimension par ALINK et collab. [2007] puis LI [2008].

## 1.3 Délimitation du problème d'apprentissage statistique

### 1.3.1 Hypothèses sur la distribution sous-jacente

On choisit de se placer dans un contexte favorable où la mesure  $\mu_{\mathbf{X}}$  est absolument continue et admet donc une densité de probabilité conjointe  $f_{\mathbf{X}}$ . On exclut les situations évoquées dans GUNAWAN et PAPALAMBROS [2007] et YANG et HSIEH [2011] où  $\mathbf{X}$  rassemble des variables discrètes et continues. Par ailleurs, le vecteur  $\mathbf{X}$  est supposé de carré intégrable, c'est-à-dire que toutes les variables  $X_j$  admettent un moment d'ordre 2 et qu'il est alors possible de calculer  $\mathbf{m}_{\mathbf{X}}$  et  $\Sigma_{\mathbf{X}}$ . On note  $\mathbb{L}^2(\mathbb{R}^d)$  l'ensemble des vecteurs aléatoires de carré intégrable. Par abus de notation,  $\mathbb{L}^2(\mathbb{R}^d)$  désigne aussi l'espace fonctionnel rassemblant les densités de probabilité associées :

$$f_{\mathbf{X}} \in \mathbb{L}^2(\mathbb{R}^d) \iff \sum_{j=1}^d \int_{\mathbb{R}} x_j^2 f_{X_j}(x_j) dx_j < \infty \iff \mathbb{E}[\|\mathbf{X}\|_2^2] < \infty. \quad (1.71)$$

L'apprentissage statistique survient lorsque la mesure de probabilité théorique  $\mu_{\mathbf{X}}$  est inconnue ou lorsqu'elle n'est pas complètement spécifiée. Le matériel initial est un lot de  $N$  réalisations

indépendantes et identiquement distribuées du vecteur  $\mathbf{X}$  :

$$\forall 1 \leq i \leq N, \quad \mathbf{X}^{(i)} = (X_1^{(i)}, \dots, X_d^{(i)}) \sim \mu_{\mathbf{X}}. \quad (1.72)$$

Cette information issue du terrain est d'un point de vue statistique un N-échantillon. Il est souvent réarrangé sous la forme d'une matrice de données après empilement des différentes observations :

$$\mathbf{X}_{\text{obs}} = \left[ X_j^{(i)} \right]_{\substack{1 \leq i \leq N \\ 1 \leq j \leq d}} = \begin{pmatrix} X_1^{(1)} & X_2^{(1)} & \dots & X_j^{(1)} & \dots & X_d^{(1)} \\ X_1^{(2)} & X_2^{(2)} & \dots & X_j^{(2)} & \dots & X_d^{(2)} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_1^{(i)} & X_2^{(i)} & \dots & X_j^{(i)} & \dots & X_d^{(i)} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_1^{(N)} & X_2^{(N)} & \dots & X_j^{(N)} & \dots & X_d^{(N)} \end{pmatrix} \sim (\mu_{\mathbf{X}})^{\otimes N}. \quad (1.73)$$

La  $i$ -ème ligne est notée  $\mathbf{X}_{\text{obs}}^{[i \bullet]}$  et n'est autre que la  $i$ -ème réalisation  $\mathbf{X}^{(i)}$  de  $\mathbf{X}$  introduite plus haut. Extraire la  $j$ -ème colonne donne en revanche accès à un N-échantillon de la variable  $X_j$  :

$$\mathbf{X}_{\text{obs}}^{[\bullet j]} = \left[ X_j^{(i)} \right]_{1 \leq i \leq N} = \begin{pmatrix} X_j^{(1)} \\ \vdots \\ X_j^{(N)} \end{pmatrix} \sim (\mu_{X_j})^{\otimes N}. \quad (1.74)$$

On part du principe que le processus de collecte des données s'opère à la perfection. Ainsi, bien que ce soit rarement le cas en pratique selon RUBIN [1976], on suppose qu'il n'y a pas de valeur manquante. On exclut aussi le cas où les modalités expérimentales d'observation des variables conduisent au recensement d'intervalles et non de valeurs ponctuelles. Ce contexte particulier est étudié dans BILLARD et DIDAY [2000] ou LAURO et PALUMBO [2000].

### 1.3.2 Description du problème d'apprentissage statistique

À partir du matériel contenu dans  $\mathbf{X}_{\text{obs}}$ , le rôle de l'apprentissage statistique est de construire une mesure de probabilité  $\hat{\mu}_{\mathbf{X}}$  qui approche suffisamment bien la mesure théorique  $\mu_{\mathbf{X}}$ . Cette tâche est d'autant plus facile que le nombre  $N$  d'observations fournies est grand. Dans ce travail, on est confronté à une situation de "small data" étant donné que  $N$  ne dépasse pas quelques centaines dans les cas traités. Pour ces tailles d'échantillon, il existe un fort risque de surapprentissage. En effet, lorsque  $N$  devient trop petit,  $\hat{\mu}_N$  capture une information qui est trop spécifique à l'échantillon  $\mathbf{X}_{\text{obs}}$  et ne peut alors plus être considérée comme une représentation satisfaisante de  $\mu_{\mathbf{X}}$ . Cette préoccupation est au cœur du travail de ZHANG et SHIELDS [2018].

En fiabilité, quantifier la probabilité de défaillance  $P_f$  revient à quantifier la probabilité que le vecteur  $\mathbf{X}$  atteigne le domaine de défaillance  $D_f \subset \mathbb{R}^d$ . Pour y parvenir, il faut intégrer la densité conjointe  $f_{\mathbf{X}}$  sur ce domaine  $D_f$ . Comme  $f_{\mathbf{X}}$  semble être la caractérisation la plus adaptée à ce contexte, la construction de  $\hat{\mu}_N$  passe par la construction d'une estimation  $\hat{f}_{\mathbf{X}}$  de  $f_{\mathbf{X}}$ . On a alors :

$$P_f = \underbrace{\mathbb{P}(\mathbf{X} \in D_f)}_{\mu_{\mathbf{X}}(D_f)} = \int_{D_f} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \approx \int_{D_f} \hat{f}_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \underbrace{\hat{\mathbb{P}}(\mathbf{X} \in D_f)}_{\hat{\mu}_N(D_f)} = \hat{P}_f. \quad (1.75)$$

Dans la suite, tout algorithme exécutant la tâche d'apprentissage statistique est noté  $\mathcal{L}$  et peut être systématiquement formalisé de la manière suivante :

$$\mathcal{L}: \mathbb{R}^{N \times d} \longrightarrow \mathbb{L}^2(\mathbb{R}^d) \\ \mathbf{x}_{\text{obs}} \longmapsto \hat{f}_{\mathbf{X}}(\cdot | \mathbf{x}_{\text{obs}}). \quad (1.76)$$

En plus d'appartenir à  $\mathbb{L}^2(\mathbb{R}^d)$ , la fonction  $\hat{f}_{\mathbf{X}}(\cdot | \mathbf{x}_{\text{obs}})$  doit intégrer à 1 car il s'agit d'une densité. La construction de  $\hat{f}_{\mathbf{X}}(\cdot | \mathbf{x}_{\text{obs}})$  impose le choix préalable d'un modèle d'apprentissage noté  $\mathcal{M}$ . Dans ce travail, on ne s'intéresse qu'aux modèles s'articulant autour de la décomposition copule-marginales. Tout modèle de ce type se subdivise en des sous-modèles  $\mathcal{M}_1, \dots, \mathcal{M}_d$  et  $\mathcal{M}_{\text{cop}}$  qui sont respectivement destinés à l'apprentissage de chacune des lois marginales et à l'apprentissage de la copule. En vertu de la version (1.38) du théorème de SKLAR [1959], la densité qui est produite par l'algorithme  $\mathcal{L}$  s'écrit alors :

$$\forall \mathbf{x} \in \mathbb{R}^d, \quad \hat{f}_{\mathbf{X}}(\mathbf{x} | \mathcal{M}, \mathbf{x}_{\text{obs}}) = \left[ \prod_{j=1}^d \hat{f}_{X_j}(x_j | \mathcal{M}_j, \mathbf{x}_{\text{obs}}) \right] \times \hat{c}_{\mathbf{U}}(\hat{\mathbf{u}} | \mathcal{M}_{\text{cop}}, \mathbf{x}_{\text{obs}}) \quad (1.77)$$

avec  $\hat{\mathbf{u}} = [\hat{u}_j]_{1 \leq j \leq d}$  où  $\hat{u}_j = \hat{F}_{X_j}(x_j | \mathcal{M}_j, \mathbf{x}_{\text{obs}})$  ,

et les différentes composantes fonctionnelles peuvent être estimées séparément. La seule étude de ces modèles peut sembler restrictive mais il est justifié dans KUROWICKA et COOKE [2006] et il est désormais communément admis que l'utilisation de la disjonction copule-marginales est une stratégie incontournable pour estimer une densité en grande dimension. Ce point de vue est corroboré par les arguments de BEDFORD et collab. [2016] ou GENEST et SCHERER [2019]. Le cadre d'apprentissage établi par les équations (1.76) et (1.77) est étudié en détail au Chapitre 4.

Dans certains cas, le statisticien peut s'appuyer sur la connaissance parfaite des différentes lois marginales  $\mu_{X_j}$  et il suffit d'estimer la copule  $c_{\mathbf{U}}$  pour disposer d'une estimation de la densité conjointe  $f_{\mathbf{X}}$ . Ce cadre de travail est par exemple adopté dans CUADRAS et collab. [2002] ou BENES et STEPÁN [2012]. La connaissance des fonctions de répartition marginales  $F_{X_j}$  permet de transformer les observations du vecteur  $\mathbf{X}$  en des observations du vecteur  $\mathbf{U}$  :

$$\mathbf{U}_{\text{obs}} = \left[ U_j^{(i)} \right]_{\substack{1 \leq i \leq N \\ 1 \leq j \leq d}} = \begin{pmatrix} U_1^{(1)} & \dots & U_d^{(1)} \\ \vdots & & \vdots \\ U_1^{(N)} & \dots & U_d^{(N)} \end{pmatrix} \sim (\mu_{\mathbf{X}})^{\otimes N} \quad \text{avec} \quad U_j^{(i)} := F_{X_j}(X_j^{(i)}) \quad . \quad (1.78)$$

L'action de l'algorithme d'apprentissage se limite alors à :

$$\mathcal{L}: \begin{array}{ccc} [0, 1]^{N \times d} & \longrightarrow & \mathbb{L}^2(\mathcal{H}_d) \\ \mathbf{u}_{\text{obs}} & \longmapsto & \hat{c}_{\mathbf{U}}(\cdot | \mathbf{u}_{\text{obs}}) \end{array} \quad . \quad (1.79)$$

La fonction  $\hat{c}_{\mathbf{U}}(\cdot | \mathbf{u}_{\text{obs}})$  produite par  $\mathcal{L}$  doit intégrer à 1 et ses densités marginales doivent être uniformes. Le modèle  $\mathcal{M}$  se limite à la composante  $\mathcal{M}_{\text{cop}}$  et la densité estimée s'écrit :

$$\hat{f}_{\mathbf{X}}(\mathbf{x} | \mathcal{M}_{\text{cop}}, \mathbf{u}_{\text{obs}}) = \left[ \prod_{j=1}^d f_{X_j}(x_j) \right] \times \hat{c}_{\mathbf{U}}(\mathbf{u} | \mathcal{M}_{\text{cop}}, \mathbf{u}_{\text{obs}}) \quad \text{avec} \quad u_j = F_{X_j}(x_j) \quad . \quad (1.80)$$

Cette situation est notamment rencontrée au Chapitre 5. De plus, LIU et DER KIUREGHIAN [1986] expliquent qu'il n'est pas rare que la connaissance de la matrice de corrélation  $\rho_{\mathbf{X}}$  s'ajoute à la connaissance des lois marginales. Afin de prendre en compte les contraintes supplémentaires imposées par les corrélations, GHOSH et HENDERSON [2002] ainsi que FARIDAFSHIN et NAESS [2017] ont développé des modèles de distributions spécifiques. Certains fondements des copules R-vignes présentés dans KUROWICKA et COOKE [2001] s'inscrivent aussi dans ce contexte. Dans la suite de ce chapitre, on étudie les techniques inférentielles les plus courantes pour réaliser l'apprentissage des densités marginales ou de la densité de copule.

## 1.4 Méthodes d'apprentissage statistique pour les lois marginales

Pour toute variable  $X_j$  de  $\mathbf{X}$ , on s'intéresse au modèle  $\mathcal{M}_j$  guidant l'estimation de la densité marginale  $f_{X_j}$ . Ce modèle donne à  $f_{X_j}$  une expression particulière. Lors de la phase inférentielle, c'est-à-dire au moment de l'exécution de l'algorithme  $\mathcal{L}$ , cette expression est calibrée aux données de l'échantillon  $\mathbf{X}_{\text{obs}}^{[\cdot, j]}$ . Comme la démarche est la même pour toutes les variables  $X_j$ , on se focalise désormais sur une variable  $Z$  de loi  $\mu_Z$  et pour laquelle on dispose d'un  $N$ -échantillon noté  $\mathbf{Z}_{\text{obs}}$ .

### 1.4.1 Méthodes paramétriques pour l'apprentissage des lois marginales

Toute approche paramétrique consiste à supposer, puis ensuite à éventuellement remettre en cause, le fait que la densité sous-jacente  $f_Z$  appartienne à une certaine famille  $\mathcal{F}_\theta$  de distributions :

$$\mathcal{F}_\theta := \{h_\theta : z \mapsto h_\theta(z) = h(z | \theta) \text{ avec } \theta \in D_\theta\} \subset \mathbb{L}^2(\mathbb{R}). \quad (1.81)$$

Elle correspond à une certaine expression analytique de  $f_Z$  dépendant d'un vecteur  $\theta$  formé par  $p$  paramètres qui prennent leurs valeurs dans un domaine  $D_\theta \subseteq \mathbb{R}^p$ . La procédure inférentielle vise à utiliser les données de  $\mathbf{Z}_{\text{obs}}$  pour estimer le paramètre vectoriel  $\theta_0$  tel que  $f_Z = h(\cdot | \theta_0)$  sur  $D_\theta$ . Une multitude de disciplines ont recours à une description paramétrique des incertitudes mais les distributions rencontrées sont souvent de la même nature. Un petit nombre de familles suffit à appréhender correctement la plupart des phénomènes, bien que plusieurs dizaines d'autres modèles soient répertoriés dans la littérature, comme par exemple dans [JOHNSON et collab. \[1995\]](#) ou dans [KOTZ et collab. \[2004\]](#). Pour une famille paramétrique  $\mathcal{F}_\theta$  donnée, il est souvent utile de savoir si elle appartient à la classe exponentielle.

#### Définition 1.4.1 (Famille exponentielle)

Soit  $\mathcal{F}_\theta$  une famille paramétrique. Elle est dite exponentielle si tout élément de  $\mathcal{F}_\theta$  s'écrit :

$$h_\theta(z) = a(\theta) \times b(z) \times \exp[\theta^\top t(z)], \quad (1.82)$$

avec  $a : D_\theta \rightarrow \mathbb{R}^+$ ,  $b : \mathbb{R} \rightarrow \mathbb{R}^+$  et  $t : \mathbb{R} \rightarrow \mathbb{R}^p$  des applications usuelles.

En posant  $A(\theta) = -\log a(\theta)$ , on obtient une expression appelée la forme canonique de  $\mathcal{F}_\theta$  :

$$h_\theta(z) = b(z) \times \exp[\theta^\top t(z) - A(\theta)]. \quad (1.83)$$

La classe exponentielle comprend par exemple la loi normale, les lois bêta et Gamma ainsi que les lois de Pareto, Weibull et Laplace. La Définition 1.4.1 se généralise aux familles paramétriques décrivant des distributions multidimensionnelles. Qu'elle soit exponentielle ou non, une famille  $\mathcal{F}_\theta$  se caractérise par ses moments, son support, le poids de ses queues de distribution, son niveau d'asymétrie ou son rapport à l'intégrabilité. En règle générale, il y a deux compétences principales à acquérir dans la manipulation des familles paramétriques :

1. Choisir une famille  $\mathcal{F}_\theta$  appropriée à l'estimation de  $f_Z$ .
2. Pour la famille  $\mathcal{F}_\theta$  choisie, estimer le paramètre  $\theta_0$  à partir des données de  $\mathbf{Z}_{\text{obs}}$ .

Ces deux tâches sont souvent imbriquées car les outils aidant à la prise de décision au sujet de  $\mathcal{F}_\theta$  nécessitent au préalable de calibrer les différentes familles candidates à  $\mathbf{Z}_{\text{obs}}$ .

#### 1.4.1.1 Méthode du maximum de vraisemblance

Le point de départ de cette méthode consiste à supposer que les observations rassemblées dans  $\mathbf{Z}_{\text{obs}}$  sont identiquement distribuées suivant  $h_{\theta_0}$  où le paramètre  $\theta_0$  est inconnu. Pour une densité quelconque  $h_\theta$  de  $\mathcal{F}_\theta$ , il faut exprimer la probabilité d'observer  $Z^{(1)}, \dots, Z^{(N)}$  et chercher à la maximiser par rapport à  $\theta$ . À cet égard, les fonctions de vraisemblance et de log-vraisemblance sont définies comme suit :

$$L_N(\theta) := \prod_{i=1}^N h_\theta(Z^{(i)}) \quad \text{et} \quad \ell_N(\theta) := \sum_{i=1}^N \log h_\theta(Z^{(i)}). \quad (1.84)$$

Une estimation de  $\theta_0$  est fournie par toute valeur  $\hat{\theta}^{\text{MV}} \in D_\theta$  permettant de maximiser  $L_N$ . En pratique, on préfère souvent utiliser la formulation équivalente visant à minimiser l'opposé de  $\ell_N$  :

$$\hat{\theta}^{\text{MV}} := \arg \max_{\theta \in D_\theta} L_N(\theta) = \arg \min_{\theta \in D_\theta} (-\ell_N(\theta)). \quad (1.85)$$

Dans le cas général, on ne peut garantir ni l'existence ni l'unicité de l'estimateur du maximum de vraisemblance (EMV). Si on suppose que  $\ell_N$  est continue par rapport à  $\boldsymbol{\theta}$  et que  $D_{\boldsymbol{\theta}}$  est un compact de  $\mathbb{R}^p$ , alors on est assuré de l'existence d'un optimum global. Toutefois, ces hypothèses s'accordent souvent assez mal avec la pratique standard qui est fondée sur la recherche de points stationnaires dans un ouvert à l'aide de la condition nécessaire d'optimalité du premier ordre :

$$\nabla_{\boldsymbol{\theta}} \ell_N(\hat{\boldsymbol{\theta}}^{\text{MV}}) = \sum_{i=1}^N \nabla_{\boldsymbol{\theta}} \log h(Z^{(i)} | \hat{\boldsymbol{\theta}}^{\text{MV}}) = \mathbf{0}. \quad (1.86)$$

La variable  $\nabla_{\boldsymbol{\theta}} \log h(Z | \boldsymbol{\theta}_0)$  est appelée le score et elle est centrée. Dans certains cas très particuliers (comme le cas gaussien), on peut résoudre cette équation et en déduire une expression analytique de l'EMV. Lorsque ce n'est pas possible, il faut avoir recours à une résolution numérique avec un algorithme d'optimisation. Lorsque l'EMV est bien défini, sous certaines conditions de régularité et d'intégrabilité rappelées dans [WASSERMAN \[2004\]](#), il présente des propriétés statistiques attrayantes. Tout d'abord, il est consistant :

$$\hat{\boldsymbol{\theta}}^{\text{MV}} \xrightarrow[N \rightarrow \infty]{\mathbb{P}} \boldsymbol{\theta}_0. \quad (1.87)$$

De plus,  $\hat{\boldsymbol{\theta}}^{\text{MV}}$  est équivariant. Ainsi, pour construire un EMV de la quantité  $\boldsymbol{\eta} := g(\boldsymbol{\theta})$  où la fonction  $g : \mathbb{R}^p \rightarrow \mathbb{R}^q$  est supposée suffisamment lisse, il suffit de prendre  $\hat{\boldsymbol{\eta}}^{\text{MV}} = g(\hat{\boldsymbol{\theta}}^{\text{MV}})$ . L'estimateur  $\hat{\boldsymbol{\theta}}^{\text{MV}}$  est également asymptotiquement normal, au sens du théorème central limite (TCL) suivant :

$$\sqrt{N} \times (\hat{\boldsymbol{\theta}}^{\text{MV}} - \boldsymbol{\theta}_0) \xrightarrow[N \rightarrow \infty]{\text{loi}} \mathcal{N}\left(\mathbf{0}, [\mathcal{I}(\boldsymbol{\theta}_0)]^{-1}\right), \quad (1.88)$$

avec  $\mathcal{I}(\boldsymbol{\theta}_0) \in \mathbb{R}^{p \times p}$  la matrice d'information de Fisher définie comme la matrice de covariance du score. Sachant le score centré et après quelques calculs, on peut montrer que :

$$\mathcal{I}(\boldsymbol{\theta}_0) = \mathbb{E} \left[ \nabla_{\boldsymbol{\theta}} \log h(Z | \boldsymbol{\theta}_0) \left( \nabla_{\boldsymbol{\theta}} \log h(Z | \boldsymbol{\theta}_0) \right)^\top \right] = -\mathbb{E} \left[ \nabla_{\boldsymbol{\theta}}^2 \log h(Z | \boldsymbol{\theta}_0) \right]. \quad (1.89)$$

où  $\nabla_{\boldsymbol{\theta}}^2 \log h(Z | \boldsymbol{\theta}_0)$  est la matrice hessienne de l'application  $\boldsymbol{\theta} \mapsto \log h(Z | \boldsymbol{\theta})$  évaluée en  $\boldsymbol{\theta}_0$ . Des manipulations qu'on choisit de ne pas détailler ici permettent alors de construire un intervalle de confiance pour tout paramètre  $\theta_j$  de  $\boldsymbol{\theta}_0$ . Par ailleurs, le TCL (1.88) se réduit à :

$$\sqrt{N} \times (\hat{\theta}_j^{\text{MV}} - \theta_j) \xrightarrow[N \rightarrow \infty]{\text{loi}} \mathcal{N}\left(0, [\mathcal{I}_j(\theta_j)]^{-1}\right). \quad (1.90)$$

Comme la variance asymptotique est égale à la borne de Cramér-Rao, définie comme le meilleur minorant de la variance de tout estimateur non biaisé,  $\hat{\theta}_j^{\text{MV}}$  est asymptotiquement efficace. L'étude des propriétés asymptotiques de l'EMV occupe une place importante au sein de l'ouvrage que [LEHMANN et CASELLA \[2006\]](#) consacrent à l'estimation paramétrique.

#### 1.4.1.2 Méthode des moments

La méthode des moments est basée sur l'égalité supposée entre les moments empiriques qui sont calculés à partir de  $\mathbf{Z}_{\text{obs}}$  et les moments théoriques qui sont des fonctions du paramètre  $\boldsymbol{\theta}_0$ . Supposons que l'on dispose de  $p$  fonctions  $g_j : \mathbb{R} \rightarrow \mathbb{R}$  telles que toutes les variables  $g_j(Z)$  soient de carré intégrable. Pour chaque fonction, en vertu du théorème de transfert, on a :

$$\mathbb{E} [g_j(Z)] = \int_{\mathbb{R}} g_j(z) h(z | \boldsymbol{\theta}_0) dz = \alpha_j(\boldsymbol{\theta}_0). \quad (1.91)$$

Chaque espérance mise en jeu peut être approchée par son équivalent empirique :

$$\hat{\alpha}_j^{\text{M}} := \frac{1}{N} \sum_{i=1}^N g_j(Z^{(i)}). \quad (1.92)$$

Dans la suite, on adopte les notations vectorielles suivantes :

$$g(Z) := [g_j(Z)]_{1 \leq j \leq p} \quad , \quad \alpha(\boldsymbol{\theta}) := [\alpha_j(\boldsymbol{\theta})]_{1 \leq j \leq p} \quad \text{et} \quad \hat{\boldsymbol{\alpha}}^M := [\hat{\alpha}_j^M]_{1 \leq j \leq p} \quad . \quad (1.93)$$

On a alors deux résultats classiques. D'une part, la loi faible des grands nombres assure que :

$$\hat{\boldsymbol{\alpha}}^M = \frac{1}{N} \sum_{i=1}^N g(Z^{(i)}) \xrightarrow[N \rightarrow \infty]{\mathbb{P}} \mathbb{E}[g(Z)] = \alpha(\boldsymbol{\theta}_0) \quad . \quad (1.94)$$

D'autre part, un TCL vient immédiatement :

$$\sqrt{N} \times (\hat{\boldsymbol{\alpha}}^M - \alpha(\boldsymbol{\theta}_0)) \xrightarrow[N \rightarrow \infty]{\text{loi}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \quad \text{avec} \quad \boldsymbol{\Sigma} = \mathbb{E}[g(Z)g(Z)^\top] - \alpha(\boldsymbol{\theta}_0)\alpha(\boldsymbol{\theta}_0)^\top \quad . \quad (1.95)$$

La correspondance entre les moments empiriques et théoriques peut s'écrire :

$$\frac{1}{N} \sum_{i=1}^N g(Z^{(i)}) = \hat{\boldsymbol{\alpha}}^M \approx \alpha(\hat{\boldsymbol{\theta}}^M) = \mathbb{E}[g(Z)] \quad . \quad (1.96)$$

L'estimateur  $\hat{\boldsymbol{\theta}}^M$  défini par la méthode des moments est la solution du système d'équations décrit par la relation  $\alpha(\hat{\boldsymbol{\theta}}^M) = \hat{\boldsymbol{\alpha}}^M$ . Supposons que  $\alpha : D_{\boldsymbol{\theta}} \rightarrow \mathbb{R}^p$  soit injective. Pour que  $\hat{\boldsymbol{\theta}}^M$  existe, il faut que  $\hat{\boldsymbol{\alpha}}^M$  appartienne à l'image de l'application  $\alpha$ . Au vu de (1.94), cela se produit avec une probabilité qui tend vers 1 lorsque  $N \rightarrow \infty$ . La fonction réciproque  $\alpha^{-1}$  a alors un sens et on est libre d'écrire que  $\hat{\boldsymbol{\theta}}^M = \alpha^{-1}(\hat{\boldsymbol{\alpha}}^M)$ . L'accès à cette réciproque dépend des fonctions  $g_j$  sélectionnées. Un choix par défaut est l'utilisation des fonctions  $g_j(z) = z^j$ . Lorsqu'une résolution analytique n'est pas possible, GELMAN [1995] propose une méthode numérique pour trouver une racine de l'équation  $\alpha(\boldsymbol{\theta}) - \hat{\boldsymbol{\alpha}}^M = \mathbf{0}$  et en déduire  $\hat{\boldsymbol{\theta}}^M$ .

Supposons désormais que  $\alpha : D_{\boldsymbol{\theta}} \rightarrow \mathbb{R}^p$  soit continûment différentiable sur  $D_{\boldsymbol{\theta}}$  et que la matrice jacobienne  $J_{\alpha}(\boldsymbol{\theta}_0)$  soit inversible. Dans ces conditions, le théorème d'inversion locale indique que la réciproque  $\alpha^{-1}$  est différentiable dans un voisinage de  $\alpha(\boldsymbol{\theta}_0)$ . Par continuité de  $\alpha^{-1}$  en  $\alpha(\boldsymbol{\theta}_0)$ , le résultat de convergence (1.94) se transforme en :

$$\hat{\boldsymbol{\theta}}^M \xrightarrow[N \rightarrow \infty]{\mathbb{P}} \boldsymbol{\theta}_0 \quad . \quad (1.97)$$

L'estimateur obtenu avec la méthode des moments est donc lui aussi consistant. Par ailleurs, la delta méthode présentée dans VAN DER VAART [2000] permet de déduire un nouveau TCL à partir de celui de l'équation (1.95) :

$$\sqrt{N} \times (\hat{\boldsymbol{\theta}}^M - \boldsymbol{\theta}_0) \xrightarrow[N \rightarrow \infty]{\text{loi}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}) \quad \text{avec} \quad \boldsymbol{\Gamma} = \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^\top \quad \text{où} \quad \mathbf{B} := J_{\alpha^{-1}}(\alpha(\boldsymbol{\theta}_0)) = [J_{\alpha}(\boldsymbol{\theta}_0)]^{-1} \quad . \quad (1.98)$$

L'estimateur est donc asymptotiquement normal mais sa variance asymptotique dépend de  $\alpha$ , c'est-à-dire du choix de la fonction  $g$ . Dans de nombreux cas, la méthode des moments est facile à mettre en œuvre d'un point de vue pratique car une résolution analytique est possible. Cependant, elle présente aussi plusieurs désavantages. D'une part, l'estimateur  $\hat{\boldsymbol{\theta}}^M$  n'est pas asymptotiquement efficace comme l'était l'estimateur  $\hat{\boldsymbol{\theta}}^{MV}$ . D'autre part, lorsque la taille  $N$  de l'échantillon est petite, la convergence (1.94) n'est pas encore suffisante pour que l'égalité  $\alpha(\boldsymbol{\theta}_0) \approx \hat{\boldsymbol{\alpha}}^M$  puisse être considérée comme raisonnable. L'estimation rendue par la méthode des moments sert souvent de point de départ (facile à obtenir) à l'algorithme utilisé pour maximiser la log-vraisemblance. Une version plus sophistiquée de la méthode des moments consiste à mettre en place un système de  $q > p$  équations (non-linéaires) à  $p$  inconnues. Dès lors,  $\alpha(\boldsymbol{\theta}) = \hat{\boldsymbol{\alpha}}^M$  est surdéterminée et il n'existe plus de solution exacte. On parle alors de la méthode des moments généralisée (MG) et on renvoie au papier précurseur de HANSEN et SINGLETON [1982] à ce sujet. Sous certaines conditions rappelées dans ZSOHAR [2012], l'estimateur  $\hat{\boldsymbol{\theta}}^{MG}$  est consistant et asymptotiquement normal.

### 1.4.1.3 Méthode du maximum *a posteriori*

Passer dans le cadre bayésien implique un changement de point de vue quant à la nature du paramètre  $\theta_0$ . Dans les deux précédents paragraphes, il s'agit d'une valeur immuable commune à tous les  $N$ -échantillons de la loi  $\mu_Z$ . On considère maintenant que la réalisation  $\mathbf{z}_{\text{obs}}$  de  $\mathbf{Z}_{\text{obs}}$  est constituée de  $N$  observations i.i.d. d'une loi de densité  $h_{\theta}$  avec  $\theta$  la réalisation d'un vecteur aléatoire  $\Theta$ . Un premier espace probabilisé modélise l'incertitude sur les paramètres :

$$\begin{array}{ccc} (\Omega_1, \mathcal{A}_1, \mathbb{P}_1) & \longrightarrow & (\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), \mu_{\Theta}) \\ \omega_1 & \xrightarrow{\Theta} & \theta := \Theta(\omega_1) \end{array} . \quad (1.99)$$

$\mu_{\Theta}$  est appelée la loi *a priori* et elle est fournie en amont (par un avis d'expert ou à l'aide des résultats de précédentes calibrations). Sa densité conjointe  $f_{\Theta}$  fait office de "croyance" au sujet de la distribution de  $\Theta$ . Pour une valeur  $\theta$  fixée, un second espace probabilisé modélise l'incertitude affectant la variable  $Z$  :

$$\begin{array}{ccc} (\Omega_2, \mathcal{A}_2, \mathbb{P}_2) & \longrightarrow & (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_{Z|\Theta=\theta}) \\ \omega_2 & \xrightarrow{Z} & z := Z(\omega_2) \end{array} . \quad (1.100)$$

La clef de voûte du paradigme bayésien consiste à utiliser les données de  $\mathbf{Z}_{\text{obs}}$ , considérées comme les seules "évidences", pour mettre à jour la loi  $f_{\Theta}$ . Pour cela, on cherche à exprimer la distribution de  $\Theta$  conditionnellement à l'évènement  $\{\mathbf{Z}_{\text{obs}} = \mathbf{z}_{\text{obs}}\} = \{Z^{(1)} = z^{(1)}, \dots, Z^{(N)} = z^{(N)}\}$  :

$$f_{\Theta|\mathbf{Z}_{\text{obs}}}(\theta | \mathbf{z}_{\text{obs}}) = \frac{f_{\Theta}(\theta) \times \prod_{i=1}^N f_{Z_{\text{obs}}|\Theta}(\mathbf{z}_{\text{obs}} | \theta)}{f_{\mathbf{Z}_{\text{obs}}}(\mathbf{z}_{\text{obs}})} \propto f_{\Theta}(\theta) \times L_N(\theta) . \quad (1.101)$$

Il s'agit de la densité *a posteriori*, c'est-à-dire après la prise en compte des données disponibles. Le dénominateur (qui est appelé vraisemblance prédictive ou vraisemblance marginale) n'est rien d'autre qu'une constante de normalisation du produit de la loi *a priori* et de la vraisemblance pour avoir une densité. La question est alors de savoir comment estimer la valeur  $\theta_0$  prise par  $\Theta$  dans le cas précis où des observations  $\mathbf{Z}_{\text{obs}}$  sont fournies. Pour cela, on se munit d'une fonction de coût  $\gamma$  et on cherche l'estimateur  $\hat{\theta}^*$  qui minimise le critère suivant :

$$\hat{\theta}^* = \underset{\theta \in \mathbb{L}^2(\mathbb{R}^p)}{\text{argmin}} \mathbb{E} \left[ \gamma(\Theta, \hat{\theta}(\mathbf{Z}_{\text{obs}})) \right] . \quad (1.102)$$

Si  $\gamma$  est la fonction de coût 0 – 1 définie par  $\gamma(\theta_1, \theta_2) = \mathbb{1}_{\{\|\theta_1 - \theta_2\| > \epsilon\}}$ ,  $\hat{\theta}^*$  est appelé l'estimateur du maximum *a posteriori* (MAP) et se note  $\hat{\theta}^{\text{MAP}}$ . Un résultat rappelé dans [BASSETT et DERIDE \[2019\]](#) indique que cet estimateur est le mode de la densité *a posteriori* :

$$\hat{\theta}^{\text{MAP}} = \underset{\theta \in D_{\Theta}}{\text{argmax}} f_{\Theta|\mathbf{Z}_{\text{obs}}}(\theta | \mathbf{z}_{\text{obs}}) , \quad (1.103)$$

On peut alors voir que :

$$\hat{\theta}^{\text{MAP}} = \underset{\theta \in D_{\Theta}}{\text{argmax}} f_{\Theta}(\theta) L_N(\theta) = \underset{\theta \in D_{\Theta}}{\text{argmin}} (- [\log f_{\Theta}(\theta) + \ell_N(\theta)]) . \quad (1.104)$$

On doit résoudre un problème d'optimisation très proche de celui qui gouverne l'estimation par maximum de vraisemblance. La seule différence réside dans l'ajout du logarithme de  $f_{\Theta}$  dans la fonction objectif. [ROBERT \[2007\]](#) souligne que l'estimateur MAP est souvent appréhendé comme un EMV "pénalisé". L'estimateur  $\hat{\theta}^{\text{MAP}}$  conserve les propriétés asymptotiques de  $\hat{\theta}^{\text{MV}}$  sous certaines conditions sur  $f_{Z|\Theta}$  et  $f_{\Theta}$  que l'on retrouve dans [LEHMANN et CASELLA \[2006\]](#). Lorsque  $N$  devient grand, l'information apportée par les observations de  $\mathbf{Z}_{\text{obs}}$  devient prédominante et l'information provenant de la loi *a priori* n'a plus d'importance au niveau asymptotique. En guise de remarque, on signale que d'autres fonctions de coût peuvent être utilisées pour définir  $\hat{\theta}^*$ . La fonction de coût quadratique  $\gamma(\theta_1, \theta_2) = \|\theta_1 - \theta_2\|_2^2$  engendre ainsi l'estimateur MMSE (*Minimum Mean Square Error*) présenté, entre autres, dans [GELMAN et collab. \[2013\]](#).

#### 1.4.1.4 Sélection d'une famille paramétrique

L'un des principaux soucis inhérents à l'utilisation d'une approche paramétrique réside dans le choix et la validation de la famille paramétrique  $\mathcal{F}_\theta$  au sein de laquelle la densité sous-jacente doit être recherchée. En pratique, on distingue typiquement deux situations :

- (a) S'il est indiqué qu'il faut utiliser une famille  $\mathcal{F}_\theta$ , on veut pouvoir valider cette indication à partir de l'information disponible dans le N-échantillon  $\mathbf{Z}_{\text{obs}}$ .
- (b) Si le choix de  $\mathcal{F}_\theta$  est laissé à l'utilisateur et que plusieurs familles semblent cohérentes avec l'histogramme de  $\mathbf{Z}_{\text{obs}}$ , on veut disposer d'un critère pour sélectionner la famille permettant d'atteindre le plus haut degré d'ajustement aux données.

Dans la littérature, la situation (a) est majoritairement résolue à l'aide d'un test d'hypothèse. Il en existe plusieurs mais les principaux sont les tests de Kolmogorov-Smirnov, Anderson-Darling et Cramér-Von Mises. Si les prémices de cette méthodologie remontent aux travaux historiques de CRAMÉR [1928], KOLMOGOROV [1933] ou SMIRNOFF [1939], leur formalisme actuel est établi un peu plus tard dans MASSEY JR [1951] et DARLING [1957]. Une étape préliminaire consiste à utiliser l'une des techniques d'estimation paramétrique présentées dans les précédents paragraphes pour estimer  $\theta_0$  à partir de  $\mathbf{Z}_{\text{obs}}$ . On compare ensuite la fonction de répartition empirique  $\hat{F}_N$  qui est construite indépendamment de  $\mathcal{F}_\theta$  et la fonction de répartition  $F_{\hat{\theta}}$  obtenue après le calcul de  $\hat{\theta}$ . Quel que soit le test, l'hypothèse nulle reste la même :

$$H_0 : F_{\hat{\theta}} \text{ est la fonction de répartition de la variable } Z. \quad (1.105)$$

En revanche, la statistique de test dépend naturellement du test concerné. Les trois tests sont basés sur le calcul d'une distance entre  $\hat{F}_N$  et  $F_{\hat{\theta}}$  mais ils n'utilisent pas la même métrique. Ainsi, les tests d'Anderson-Darling et de Cramér-Von Mises apparaissent comme des versions plus puissantes du test original de Kolmogorov-Smirnov. BROWNE et CUDECK [1992] font néanmoins remarquer qu'on ne connaît que la loi asymptotique des statistiques de test et qu'il est par conséquent risqué d'appliquer ces tests d'hypothèse pour un échantillon  $\mathbf{Z}_{\text{obs}}$  de petite taille.

En ce qui concerne la situation (b), commençons par noter  $\mathcal{F}_1, \dots, \mathcal{F}_M$  les différentes familles paramétriques candidates. On cherche alors à calculer un indicateur de la qualité de l'adéquation du modèle aux données. On pourrait penser à la valeur maximale  $\ell_N(\hat{\theta}^{\text{MV}})$  de la log-vraisemblance. Toutefois, cette quantité est croissante avec le nombre de paramètres  $p$  au sein de  $\theta$  et son emploi conduirait à privilégier systématiquement les familles incorporant le plus de paramètres. Pour éviter le phénomène de surapprentissage, d'autres critères ont été développés pour valoriser les familles offrant un compromis entre le nombre de paramètres et la qualité de l'ajustement aux données. AKAIKE [1973] propose d'ajouter un terme de pénalisation relatif à la taille de  $\theta$  et définit alors le critère AIC (*Akaike Information Criterion*) :

$$\text{AIC}(\mathcal{F}_\theta | \mathbf{Z}_{\text{obs}}) := -2\ell_N(\hat{\theta}^{\text{MV}}) + 2p. \quad (1.106)$$

On retient alors la famille  $\mathcal{F}_i$  pour laquelle AIC est minimal. AKAIKE [1974] explique que le choix de ce critère n'est pas anodin car il s'agit de la divergence de Kullback-Leibler entre la vraie densité  $f_Z$  et son plus proche représentant  $f_{\theta^*}$  au sein de la famille  $\mathcal{F}_\theta$ . Le lien qu'entretiennent l'EMV et AIC est présenté succinctement dans PORTET [2020]. Pour un exposé exhaustif, on conseille l'ouvrage de ANDERSON et BURNHAM [2004]. Un autre critère fréquemment utilisé a été introduit par SCHWARZ et collab. [1978] et porte le nom de BIC (*Bayesian Information Criterion*). Il accentue la pénalisation vis-à-vis du nombre de paramètres, notamment pour de grands échantillons :

$$\text{BIC}(\mathcal{F}_\theta | \mathbf{Z}_{\text{obs}}) := -2\ell_N(\hat{\theta}^{\text{MV}}) + p \log N. \quad (1.107)$$

Dans le paradigme bayésien où la famille paramétrique est une entité aléatoire obéissant à une loi discrète sur l'ensemble des familles candidates, il est rappelé dans SAPORTA [2006] qu'il existe une correspondance entre la probabilité *a posteriori* de la famille  $\mathcal{F}_\theta$  et la valeur prise par BIC. Si on ne sait pas calculer la log-vraisemblance, le critère DIC (*Deviance Information Criterion*) introduit par SPIEGELHALTER et collab. [2002] peut être obtenu plus facilement tout en gardant des propriétés analogues à celles de AIC.

## 1.4.2 Méthodes non-paramétriques pour l'apprentissage des lois marginales

### 1.4.2.1 Fonction de répartition empirique

L'approche non-paramétrique la plus simple repose sur la fonction de répartition empirique. Dans ce cadre, on envisage les observations comme les réalisations d'une loi discrète uniforme. Chaque observation  $Z^{(i)}$  est appelée un atome et se voit affecter une probabilité d'occurrence égale à  $1/N$ . Sachant que la fonction de répartition est l'application  $F_Z : z \mapsto \mathbb{P}(Z \leq z)$ , elle peut être estimée comme suit :

$$\hat{F}_N(z) := \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\{Z^{(i)} \leq z\}}. \quad (1.108)$$

Cette fonction en escalier caractérise la mesure empirique sous-jacente :

$$\hat{\mu}_N := \hat{\mu}[\mathbf{Z}_{\text{obs}}] = \frac{1}{N} \sum_{i=1}^N \delta_{Z^{(i)}}, \quad (1.109)$$

où  $\delta_z$  est la mesure de Dirac en un point  $z \in \mathbb{R}$ . Comme l'objectif principal de l'apprentissage est la construction d'une mesure  $\hat{\mu}_Z$  absolument continue, la mesure empirique  $\hat{\mu}_N$  n'est pas conforme aux attentes. C'est un outil qui présente néanmoins une grande utilité en pratique si on a besoin de fabriquer de nouveaux échantillons d'une loi inconnue. Pour un jeu de données fixé, si on veut simuler des échantillons suivant  $\hat{\mu}_N$ , il suffit de faire des tirages avec remise au sein du lot initial d'observations. Cette procédure dite de *bootstrap* s'est répandue dans la communauté statistique suite au travail initié par [EFRON \[1982\]](#). Elle sert notamment à évaluer la robustesse d'un indicateur calculé à partir d'un échantillon unique.

La fonction empirique  $\hat{F}_N$  permet aussi de construire un intervalle de confiance pour toute valeur  $F_Z(z)$ . On peut commencer par remarquer que  $\hat{F}_N(z)$  est non biaisé, consistant et asymptotiquement normal :

$$\mathbb{E}[\hat{F}_N(z)] = F_Z(z), \quad \hat{F}_N(z) \xrightarrow[N \rightarrow \infty]{\text{p.s.}} F_Z(z) \quad \text{et} \quad \sqrt{N} \frac{\hat{F}_N(z) - F_Z(z)}{\sqrt{F_Z(z)(1 - F_Z(z))}} \xrightarrow[N \rightarrow \infty]{\text{loi}} \mathcal{N}(0, 1). \quad (1.110)$$

La combinaison des deux dernières propriétés donne naissance à un intervalle de confiance. De plus, il existe des résultats asymptotiques très puissants sur la fonction de répartition empirique. On se réfère aux énoncés trouvés dans [VAN DER VAART \[2000\]](#) et [WASSERMAN \[2006\]](#). Tout d'abord, le théorème de Glivenko-Cantelli assure la convergence uniforme presque sûre de  $\hat{F}_N$  vers  $F_Z$  :

$$\|\hat{F}_N - F_Z\|_{\infty} := \sup_{z \in \mathbb{R}} |\hat{F}_N(z) - F_Z(z)| \xrightarrow[N \rightarrow \infty]{\text{p.s.}} 0. \quad (1.111)$$

Par ailleurs, en considérant  $\sqrt{N}(\hat{F}_N - F_Z)$  comme une fonction aléatoire, le théorème de Donsker énonce un résultat de convergence dans l'espace de Skorokhod<sup>1</sup> :

$$\sqrt{N}(\hat{F}_N - F_Z) \xrightarrow[N \rightarrow \infty]{\text{loi}} \mathbb{B}_{F_Z}. \quad (1.112)$$

L'objet limite est un  $F_Z$ -pont brownien, c'est-à-dire un processus gaussien tel que :

$$\mathbb{E}[\mathbb{B}_{F_Z}(z)] = 0 \quad \text{et} \quad \text{Cov}(\mathbb{B}_{F_Z}(z_1), \mathbb{B}_{F_Z}(z_2)) = F_Z(z_1 \wedge z_2) - F_Z(z_1)F_Z(z_2), \quad (1.113)$$

où  $z_1 \wedge z_2$  désigne le minimum des deux valeurs. Par ailleurs, la loi du supremum indique que :

$$\forall z_0 \in \mathbb{R}, \quad \mathbb{P}\left(\|\mathbb{B}_{F_Z}\|_{\infty} \leq z_0\right) = 1 - 2 \sum_{k=1}^{\infty} (-1)^{k+1} \exp(-2k^2 z_0^2). \quad (1.114)$$

Sous l'hypothèse  $H_0$  posée par (1.105), la variable  $B_N := \|\sqrt{N}(\hat{F}_N - F_{\hat{\theta}})\|_{\infty}$  devient  $\|\sqrt{N}(\hat{F}_N - F_Z)\|_{\infty}$  et sa distribution asymptotique est donnée par (1.114).  $B_N$  est en fait la statistique du test de

1. Espace des fonctions càdlàg (continuité à droite et existence d'une limite à gauche).

Kolmogorov-Smirnov. Sa loi asymptotique ne dépendant pas de la fonction  $F_Z$ , on peut calculer la p-valeur sans difficulté. Ainsi, bien qu'elle ne puisse pas engendrer une densité marginale  $f_Z$ , la fonction de répartition empirique  $\hat{F}_N$  a un rôle incontestable dans l'exécution de certaines tâches d'inférence et de ré-échantillonnage. La suite du chapitre viendra aussi souligner que  $\hat{F}_N$  présente un intérêt majeur au moment de transformer les observations initiales de la variable  $Z$  en des pseudo-observations de la variable  $F_Z(Z)$ .

### 1.4.2.2 Méthode de l'estimation par noyaux

La méthode des noyaux vient formaliser et généraliser la multitude de techniques répertoriées dans SILVERMAN [1986] pour construire un histogramme. D'une manière générale, l'élaboration de ce type de représentation graphique passe par la délimitation d'un nombre fini de classes sur la droite réelle puis par la comptabilisation de leurs effectifs. Pour un histogramme, la valeur  $\hat{f}_Z(z)$  dépend uniquement du nombre d'observations  $Z^{(i)}$  positionnées dans la même classe que  $z$ . On cherche à affiner la description du lien entre la position du point  $z$  au sein de  $\mathbf{Z}_{\text{obs}}$  et la valeur de  $\hat{f}_Z(z)$ . Chaque observation  $Z^{(i)}$  est supposée apporter une contribution à la densité finale  $\hat{f}_Z$ . Cette contribution doit prendre la forme d'une densité  $\hat{g}_h^{(i)}$  susceptible d'avoir généré  $Z^{(i)}$ . Pour cela, il semble raisonnable de poser :

$$\hat{g}_h^{(i)}(z) := K_h(z - Z^{(i)}) = \frac{1}{h} K\left(\frac{z - Z^{(i)}}{h}\right), \quad (1.115)$$

où  $K : \mathbb{R} \rightarrow \mathbb{R}^+$  est un noyau, c'est-à-dire une fonction qui vérifie :

$$\forall z \in \mathbb{R}, \quad K(-z) = K(z) \quad \text{mais aussi} \quad \int_{\mathbb{R}} K(z) dz = 1 \quad \text{et} \quad \int_{\mathbb{R}} z^2 K(z) dz = \kappa_2 < \infty. \quad (1.116)$$

La propriété d'existence d'un moment d'ordre 2 n'est pas nécessaire pour que  $K$  soit un noyau mais elle est très souvent requise dans les applications. Les propriétés (1.116) impliquent que  $\hat{g}_h^{(i)}$  est la densité d'une loi d'espérance  $Z^{(i)}$  et de variance  $h^2 \kappa_2$ . Le paramètre  $h$ , appelé la fenêtre de lissage, renforce ou atténue le caractère local de la densité  $\hat{g}_h^{(i)}$ . TSYBAKOV [2008] propose une liste des noyaux couramment utilisés dans la littérature. On choisit de présenter le noyau gaussien  $K_{\mathcal{N}}$  et le noyau d'Epanechnikov  $K_E$  :

$$\forall z \in \mathbb{R}, \quad K_{\mathcal{N}}(z) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \quad \text{et} \quad K_E(z) := \frac{3}{4} (1 - z^2) \mathbb{1}_{\{|z| \leq 1\}}. \quad (1.117)$$

Si on exclut le noyau de Silverman, le noyau gaussien est l'un des seuls noyaux dont le support est la droite réelle. Pour la plupart des autres noyaux rencontrés dans les formulaires, à l'instar du noyau d'Epanechnikov, le support de  $\hat{g}_h^{(i)}$  est un compact centré sur  $Z^{(i)}$ . La contribution d'une observation  $Z^{(i)}$  est ainsi limitée dans l'espace. En supposant que la densité  $f_Z$  est au moins deux fois continûment différentiable, et en s'appuyant sur les propriétés (1.116) du noyau  $K$ , le recours à un développement limité à l'ordre 2 conduit à :

$$\mathbb{E}[\hat{g}_h^{(i)}(z)] = \int_{\mathbb{R}} K_h(z - t) f_Z(t) dt = f_Z(z) + \frac{1}{2} \kappa_2 f_Z''(z) h^2 + o(h^2) \quad \text{et} \quad \lim_{h \rightarrow 0} \mathbb{E}[\hat{g}_h^{(i)}(z)] = f_Z(z). \quad (1.118)$$

$\hat{g}_h^{(i)}(z)$  est donc un estimateur de  $f_Z(z)$  dont le biais disparaît asymptotiquement. Cependant, si  $h$  est très petit, la variance explose car on a :

$$\mathbb{V}(\hat{g}_h^{(i)}(z)) \underset{h \rightarrow 0}{\sim} \frac{1}{h} f_Z(z) \mathcal{R}(K) \quad \text{avec} \quad \mathcal{R} : g \longmapsto \int_{\mathbb{R}} g^2(z) dz. \quad (1.119)$$

Ce problème est tout à fait normal étant donné qu'on essaye d'estimer  $f_Z(z)$  à partir de la seule observation  $Z^{(i)}$ . Étudions désormais la fonction obtenue en faisant la moyenne des contributions :

$$\hat{f}_h(z) := \frac{1}{N} \sum_{i=1}^N \hat{g}_h^{(i)}(z) = \frac{1}{N} \sum_{i=1}^N K_h(z - Z^{(i)}). \quad (1.120)$$

La fonction  $\hat{f}_h$  est l'estimateur KDE (*Kernel Density Estimator*) de  $f_Z$ . On parle aussi d'estimateur de Parzen-Rosenblatt. Les propriétés de l'estimateur  $\hat{f}_h(z)$  sont directement obtenues à partir des résultats (1.118) et (1.119) :

$$\mathbb{E}[\hat{f}_h(z) - f_Z(z)] \underset{h \rightarrow 0}{\sim} \frac{1}{2} \kappa_2 f_Z''(z) h^2 \quad \text{et} \quad \mathbb{V}(\hat{f}_h(z)) \underset{h \rightarrow 0}{\sim} \frac{1}{N h} f_Z(z) \mathcal{R}(K) . \quad (1.121)$$

On a besoin d'avoir  $h \rightarrow 0$  pour éliminer le biais mais aussi  $N h \rightarrow \infty$  pour réduire la variance. Le problème de la sélection d'une fenêtre de lissage est examiné dans [PARK et MARRON \[1990\]](#) ainsi que dans [SHEATHER et JONES \[1991\]](#). Un état de l'art des techniques existantes est proposé par [HEIDENREICH et collab. \[2013\]](#). Si  $h$  est trop petit, la courbe représentative de  $\hat{f}_h$  est fortement oscillante et témoigne d'un surapprentissage de  $\mathbf{Z}_{\text{obs}}$ . Au contraire, si  $h$  est trop grand, la courbe est unimodale et peu informative. Un compromis biais-variance est atteint en cherchant à minimiser le critère MISE (*Mean Integrated Square Error*) :

$$h^* = \underset{h > 0}{\operatorname{argmin}} \operatorname{MISE}(h) \quad \text{avec} \quad \operatorname{MISE}(h) := \mathbb{E} \left[ \int_{\mathbb{R}} (\hat{f}_h(z) - f_Z(z))^2 dz \right] . \quad (1.122)$$

En appliquant le théorème de Fubini puis en réinjectant les équivalents donnés à l'équation (1.121), on obtient un critère asymptotique AMISE (*Asymptotic Mean Integrated Square Error*) :

$$\operatorname{MISE}(h) \underset{h \rightarrow 0}{\sim} \frac{\mathcal{R}(K)}{N h} + \frac{1}{4} \kappa_2^2 \mathcal{R}(f_Z'') h^4 =: \operatorname{AMISE}(h) . \quad (1.123)$$

Lors de la minimisation du critère AMISE, on obtient :

$$\frac{d}{dh} \operatorname{AMISE}(h^*) = 0 \quad \iff \quad h^* = \left( \frac{\mathcal{R}(K)}{N \kappa_2^4 \mathcal{R}(f_Z'')} \right)^{1/5} . \quad (1.124)$$

L'expression de  $h^*$  dépend d'un terme de rugosité  $\mathcal{R}(f_Z'')$  inconnu. Pour contourner cet écueil, une règle empirique proposée par [SILVERMAN \[1986\]](#) encourage à supposer que  $Z \sim \mathcal{N}(0, \hat{\sigma}_Z^2)$  avec  $\hat{\sigma}_Z^2$  la variance empirique et à en déduire la valeur de  $\mathcal{R}(f_Z'')$ .

La grande popularité de l'estimateur KDE de la densité  $f_Z$  provient de sa flexibilité. L'emploi d'une approche non-paramétrique offre un bon degré de robustesse dans l'estimation de  $f_Z$ , tout en évitant l'épineux problème du choix d'une famille paramétrique. Le principe d'estimation par noyaux peut être généralisé à l'apprentissage d'une densité conjointe  $f_Z$  de copule inconnue. Dès [SILVERMAN \[1986\]](#), il est question d'utiliser des noyaux multidimensionnels de type  $K : \mathbb{R}^d \rightarrow \mathbb{R}^+$  au sein de l'équation (1.120) et de remplacer la fenêtre de lissage  $h$  par une matrice de lissage  $\mathbf{H}$ . [SCOTT \[1991\]](#) souligne toutefois que la recherche des paramètres de  $\mathbf{H}^*$  devient vite irréalisable lorsque  $d$  augmente. Par ailleurs, il faut être conscient que la méthode KDE, que ce soit dans sa version unidimensionnelle ou multidimensionnelle, n'est précise qu'au cœur de la distribution. L'utilisation du noyau gaussien engendre la formation d'une densité  $\hat{f}_h$  avec des queues légères. Pour les autres noyaux, dont les supports sont compacts, la situation est encore plus problématique avec la mise en place de queues courtes. Si  $\mathbf{Z}_{\text{obs}}$  est l'échantillon d'une loi à queues lourdes, l'estimateur KDE peut conduire à une compréhension erronée de la décroissance réelle de  $f_Z$  dans ses régions extrêmes. Des modèles de distributions issus de la théorie des valeurs extrêmes sont plus adaptés pour estimer  $f_Z$  au niveau de ses queues. On peut notamment citer les modèles :

- GEV (*Generalized Extreme Value*) issus du théorème de Fisher-Tippett-Gnedenko,
- GPD (*Generalized Pareto Distribution*) issus du théorème de Pickands-Balkema-De Haan.

Si la méthode KDE apparaît comme un incontournable de la statistique non-paramétrique, il existe des alternatives qui ne sont pas détaillées ici par souci de concision. On mentionne seulement la méthode du maximum d'entropie de [JAYNES \[1957\]](#) qui peut être facilement mise en œuvre depuis [KAPUR et KESAVAN \[1992\]](#) et qui conduit à une densité dont le niveau de régularité dépend de la nature des contraintes imposées.

## 1.5 Méthodes d'apprentissage statistique pour la copule

Comme on vient de le voir, une profusion de modèles d'apprentissage peut être déployée pour estimer la densité  $f_{X_j}$  d'une loi marginale. Si les variables sont indépendantes, l'élaboration de la densité conjointe  $\hat{f}_{\mathbf{X}}$  se limite au produit des densités marginales estimées. Il est tout à fait possible de combiner des densités de différentes natures. Pour  $d = 3$ , on peut très bien imaginer que  $\hat{f}_{X_1}$  appartienne à une famille paramétrique, que  $\hat{f}_{X_2}$  soit issue d'un apprentissage par maximum d'entropie et que  $\hat{f}_{X_3}$  provienne d'un mélange d'outils KDE et GPD. Pendant longtemps, il a été systématiquement supposé que les variables étaient indépendantes, même lorsqu'il était évident que ce n'était pas le cas. Actuellement, au regard des avancées en termes de modélisation de la dépendance statistique, la non prise en compte arbitraire de la copule sous-jacente n'est plus une pratique tolérable lors de la phase d'apprentissage.

Comme indiqué à la Définition 1.2.2, la copule est une distribution d'un type particulier car elle modélise une loi de support  $\mathcal{H}_d$  dont toutes les lois marginales sont uniformes. Toutefois, la copule demeure une distribution donc la plupart des techniques déjà présentées pour l'estimation des densités marginales  $f_{X_j}$  peuvent être appliquées pour son apprentissage. Trois points essentiels méritent néanmoins d'être abordés :

1. La copule est la loi du vecteur  $\mathbf{U}$  des variables  $U_j := F_{X_j}(X_j)$ . Son apprentissage requiert un N-échantillon  $\mathbf{U}_{\text{obs}}$  obéissant à la mesure d'échantillonnage  $(\mu_{\mathbf{U}})^{\otimes N}$ . Si cette mesure a une signification du point de vue mathématique, elle n'en a aucun du point de vue physique car on ne pourra jamais collecter des observations de  $\mathbf{U}$ . L'obtention de  $\mathbf{U}_{\text{obs}}$  à partir de  $\mathbf{X}_{\text{obs}}$  est donc à la charge de l'utilisateur et fait l'objet de la Section 1.5.1 où sont présentées différentes procédures permettant d'estimer simultanément les densités marginales  $f_{X_j}$  et la densité de copule  $c_{\mathbf{U}}$ .
2. Comme la copule est un objet mathématique conçu pour modéliser la structure de dépendance, les modèles de copules sont d'un genre différent. En effet, ce ne sont pas de simples généralisations des modèles de distributions marginales. Si on choisit d'utiliser une famille paramétrique, le paramétrage doit rendre possible la description puis l'interprétation des motifs de dépendance. De même, si on adopte une approche non-paramétrique, certains aménagements sont nécessaires pour s'adapter à la compacité du support. On retrouve ces idées dans les Sections 1.5.2 et 1.5.3 qui visent à présenter certaines spécificités des modèles gouvernant l'apprentissage de la copule.
3. Séparer l'apprentissage des lois marginales et de la copule permet d'aborder le fléau de la dimension sous le prisme de la théorie des copules. Les familles de copules développées en dimension  $d = 2$  ou  $d = 3$  ne sont pas assez flexibles pour résister à la montée en dimension. Toutefois, d'autres modèles, présentés à la Section 1.5.4, impliquent un ciblage préalable des sous-motifs de dépendance les plus pertinents afin de permettre la mise en place d'une structure de dépendance polymorphe et suffisamment parcimonieuse.

### 1.5.1 Apprentissage conjoint des lois marginales et de la copule

Dans le cadre de ce travail, on rappelle que tout algorithme d'apprentissage statistique  $\mathcal{L}$  se sert d'un N-échantillon  $\mathbf{X}_{\text{obs}}$  pour fabriquer une densité conjointe  $\hat{f}_{\mathbf{X}}(\cdot | \mathcal{M}, \mathbf{X}_{\text{obs}})$  formée par les  $d + 1$  composantes fonctionnelles : les densités marginales  $\hat{f}_{X_j}(\cdot | \mathcal{M}_j, \mathbf{X}_{\text{obs}})$  et la densité de copule  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ . Si l'estimation des densités marginales n'est plus un problème au vu des éléments déjà introduits, la question de l'estimation simultanée des  $d + 1$  composantes est plus épineuse. Elle est abordée dans [CHOROŚ et collab. \[2010\]](#) et [MCNEIL et collab. \[2015\]](#) où plusieurs procédures d'estimation sont comparées.

Tout d'abord, examinons ce qui peut être mis en œuvre lorsque les modèles  $\mathcal{M}_1, \dots, \mathcal{M}_d$  et  $\mathcal{M}_{\text{cop}}$  sont en fait des familles paramétriques  $\mathcal{F}_1, \dots, \mathcal{F}_d$  et  $\mathcal{F}_{\text{cop}}$ . Les paramètres associés à ces familles sont notés  $\alpha_1, \dots, \alpha_d$  et  $\beta$ . Ils vivent respectivement dans des sous-domaines de  $\mathbb{R}^{p_1}, \dots, \mathbb{R}^{p_d}$

et  $\mathbb{R}^{p_C}$ . La distribution issue de l'assemblage des différentes familles paramétriques peut être vue comme un famille multidimensionnelle  $\mathcal{F}$  de paramètres  $\boldsymbol{\theta} := [\boldsymbol{\alpha}_1 \ \dots \ \boldsymbol{\alpha}_d \ \boldsymbol{\beta}]^\top$  :

$$f_{\mathbf{X}}(\mathbf{x} \mid \mathcal{F}, \boldsymbol{\theta}) := \left[ \prod_{j=1}^d f_{X_j}(x_j \mid \mathcal{F}_j, \boldsymbol{\alpha}_j) \right] \times c_{\mathbf{U}}\left(F_{X_1}(x_1 \mid \mathcal{F}_1, \boldsymbol{\alpha}_1), \dots, F_{X_d}(x_d \mid \mathcal{F}_d, \boldsymbol{\alpha}_d) \mid \mathcal{F}_{\text{cop}}, \boldsymbol{\beta}\right). \quad (1.125)$$

Tout se passe comme si la disjonction copule-marginales n'existait pas et  $f_{\mathbf{X}}(\mathbf{x} \mid \mathcal{F}, \boldsymbol{\theta})$  est vue comme la densité conjointe d'une famille multidimensionnelle à  $p_T := p_1 + \dots + p_d + p_C$  paramètres. Les méthodes d'estimation paramétrique qui ont été présentées à la Section 1.4.1 peuvent alors fournir  $\hat{\boldsymbol{\theta}}$ . Dans le cas de la méthode du maximum de vraisemblance, JOE [1997] rappelle que la log-vraisemblance  $\ell_N(\boldsymbol{\theta})$  se décompose comme suit :

$$\begin{aligned} \ell_N(\boldsymbol{\theta} \mid \mathbf{X}_{\text{obs}}) &= \sum_{i=1}^N \log f_{\mathbf{X}}(\mathbf{X}^{(i)} \mid \mathcal{F}, \boldsymbol{\theta}) \\ &= \sum_{j=1}^d \left( \sum_{i=1}^N \log f_{X_j}(X_j^{(i)} \mid \mathcal{F}_j, \boldsymbol{\alpha}_j) \right) + \sum_{i=1}^N \log c_{\mathbf{U}}\left(F_{X_1}(X_1^{(i)} \mid \mathcal{F}_1, \boldsymbol{\alpha}_1), \dots, F_{X_d}(X_d^{(i)} \mid \mathcal{F}_d, \boldsymbol{\alpha}_d) \mid \mathcal{F}_{\text{cop}}, \boldsymbol{\beta}\right) \\ &= \sum_{j=1}^d \ell_j(\boldsymbol{\alpha}_j \mid \mathbf{X}_{\text{obs}}^{[\cdot, j]}) + \ell_C(\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_d, \boldsymbol{\beta} \mid \mathbf{X}_{\text{obs}}). \end{aligned} \quad (1.126)$$

L'estimation  $\hat{\boldsymbol{\theta}}^{\text{MV}}$  découle de la maximisation de  $\ell_N(\boldsymbol{\theta})$  sur  $D_{\boldsymbol{\theta}} := D_{\boldsymbol{\alpha}_1} \times \dots \times D_{\boldsymbol{\alpha}_d} \times D_{\boldsymbol{\beta}} \subseteq \mathbb{R}^{p_T}$ . Cette fonction objectif est la somme de  $d + 1$  termes. Chacun des termes  $\ell_j(\boldsymbol{\alpha}_j \mid \mathbf{X}_{\text{obs}}^{[\cdot, j]})$  correspond à la log-vraisemblance des observations marginales  $\mathbf{X}_{\text{obs}}^{[\cdot, j]}$  pour la famille  $\mathcal{F}_j$  dont les paramètres sont rassemblés dans le vecteur  $\boldsymbol{\alpha}_j$ . Le terme  $\ell_C(\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_d, \boldsymbol{\beta} \mid \mathbf{X}_{\text{obs}})$  quantifie la log-vraisemblance des données  $\mathbf{X}_{\text{obs}}$  pour une famille de copules  $\mathcal{F}_{\text{cop}}$  de paramètres  $\boldsymbol{\beta}$  après l'envoi des observations  $\mathbf{X}^{(i)}$  dans  $\mathcal{H}_d$  à l'aide des fonctions de répartition  $F_{X_j}(\cdot \mid \mathcal{F}_j, \boldsymbol{\alpha}_j)$ . L'estimateur final  $\hat{\boldsymbol{\theta}}^{\text{MV}}$  présente toutes les propriétés mentionnées à la Section 1.4.1.1 mais il n'a pas d'expression analytique en général, sauf dans le cas très particulier où  $\mathcal{F}$  désigne la famille gaussienne. Dans le cas général, l'optimisation dans un espace de dimension  $p_T$  peut s'avérer périlleuse, et c'est d'autant plus vrai lorsque  $d$  augmente. Pour réduire la taille de l'espace de recherche, une heuristique à deux temps appelée IFM (*Inference Functions for Margins*) a été proposée dans MACLEISH et SMALL [1988]. L'idée est simplement d'estimer séparément chaque bloc de paramètres. Il faut procéder de la manière suivante :

1. Maximiser chaque log-vraisemblance marginale :

$$\hat{\boldsymbol{\alpha}}_j^{\text{IFM}} := \arg \max_{\boldsymbol{\alpha}_j} \ell_j(\boldsymbol{\alpha}_j \mid \mathbf{X}_{\text{obs}}^{[\cdot, j]}). \quad (1.127)$$

2. Maximiser la log-vraisemblance de la copule après avoir figé  $\hat{\boldsymbol{\alpha}}_1^{\text{IFM}}, \dots, \hat{\boldsymbol{\alpha}}_d^{\text{IFM}}$  :

$$\hat{\boldsymbol{\beta}}^{\text{IFM}} := \arg \max_{\boldsymbol{\beta}} \ell_C(\hat{\boldsymbol{\alpha}}_1^{\text{IFM}}, \dots, \hat{\boldsymbol{\alpha}}_d^{\text{IFM}}, \boldsymbol{\beta} \mid \mathbf{X}_{\text{obs}}). \quad (1.128)$$

La première étape revient à estimer par maximum de vraisemblance les paramètres  $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_d$ . Par ailleurs, on peut remarquer que la seconde étape correspond à la résolution du problème :

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{\text{IFM}} &:= \arg \max_{\boldsymbol{\beta}} \sum_{i=1}^N \log c_{\mathbf{U}}(\hat{\mathbf{U}}^{(i)} \mid \mathcal{F}_{\text{cop}}, \boldsymbol{\beta}) \\ \text{avec } \hat{\mathbf{U}}_{\text{obs}} &= [\hat{\mathbf{U}}^{(i)}]_{1 \leq i \leq N} = \left[ \hat{\mathbf{U}}_j^{(i)} \right]_{\substack{1 \leq i \leq N \\ 1 \leq j \leq d}} \quad \text{où } \hat{\mathbf{U}}_j^{(i)} = \hat{F}_{X_j}(X_j^{(i)} \mid \mathcal{F}_j, \hat{\boldsymbol{\alpha}}_j). \end{aligned} \quad (1.129)$$

Le calcul de  $\hat{\boldsymbol{\beta}}^{\text{IFM}}$  se ramène à une estimation par maximum de vraisemblance du paramètre  $\boldsymbol{\beta}$  à partir d'observations  $\hat{\mathbf{U}}^{(i)}$  de l'échantillon  $\hat{\mathbf{U}}_{\text{obs}}$  qui sont appelées les pseudo-observations de  $\mathbf{U}$ .

Pour des paramètres estimés  $\hat{\alpha}_1, \dots, \hat{\alpha}_d$  fixés à l'avance, on note  $\hat{\mathbf{U}}$  le vecteur qui regroupe toutes les variables  $\hat{U}_j := F_{X_j}(X_j | \mathcal{F}_j, \hat{\alpha}_j)$ . Tout comme  $\hat{\boldsymbol{\theta}}^{\text{MV}}$ , l'estimateur  $\hat{\boldsymbol{\theta}}^{\text{IFM}}$  est à la fois consistant et asymptotiquement normal. Néanmoins, il est plus difficile de déduire sa matrice de covariance asymptotique, même numériquement. Les estimateurs  $\hat{\boldsymbol{\theta}}^{\text{MV}}$  et  $\hat{\boldsymbol{\theta}}^{\text{IFM}}$  sont équivalents lorsque  $\mathcal{F}$  est la famille gaussienne. L'efficacité de la procédure IFM comparativement à une maximisation globale de  $\ell_N(\boldsymbol{\theta} | \mathbf{X}_{\text{obs}})$  est démontrée numériquement dans [XU \[1996\]](#) où une étude comparative est menée en s'appuyant sur les matrices de covariance asymptotiques et les erreurs quadratiques moyennes. Le principal problème de la méthode IFM réside dans les potentielles nuisances que peut engendrer le choix des familles  $\mathcal{F}_1, \dots, \mathcal{F}_d$ . Comme elles interviennent dans la fabrication des pseudo-observations de  $\hat{\mathbf{U}}_{\text{obs}}$ , elles sont susceptibles de biaiser l'apprentissage de la copule. Selon [SILVAPULLE et collab. \[2004\]](#), la mauvaise spécification d'une famille  $\mathcal{F}_j$  peut entraîner une surestimation du niveau de dépendance existant au sein de  $\mathbf{X}_{\text{obs}}$ . D'après le Théorème 1.2.6 d'invariance, on comprend que les vecteurs  $\mathbf{X}$ ,  $\mathbf{U}$  et  $\hat{\mathbf{U}}$  partagent la même copule  $c_{\mathbf{U}}$ . Toutefois, les marginales de  $\mathbf{U}$  sont exactement uniformes alors que celles de  $\hat{\mathbf{U}}$  ne le sont qu'à condition que  $\hat{F}_{X_j} = F_{X_j}$  pour toutes les variables, ce qui n'est jamais le cas en pratique. Le paramètre  $\boldsymbol{\beta}$  est ainsi estimé en maximisant la vraisemblance de la famille  $\mathcal{F}_{\text{cop}}$  pour des données  $\hat{\mathbf{U}}^{(i)}$  suivant une loi  $\mu_{\hat{\mathbf{U}}}$  qui n'est pas exactement celle d'une copule.

Pour éviter ce problème, les pseudo-observations peuvent être fabriquées avec la technique semi-paramétrique (SP) mise au goût du jour par [GENEST et collab. \[1995\]](#) dont les travaux ont recoupé ceux de [SHIH et LOUIS \[1995\]](#) :

$$\hat{U}_j^{(i)} = \frac{N}{N+1} \hat{F}_N(X_j^{(i)} | \mathbf{X}_{\text{obs}}^{[\bullet, j]}) = \frac{1}{N+1} \sum_{k=1}^N \mathbb{1}_{\{X_j^{(k)} \leq X_j^{(i)}\}}. \quad (1.130)$$

$\hat{F}_N(\cdot | \mathbf{X}_{\text{obs}}^{[\bullet, j]})$  est la fonction de répartition empirique construite à partir des données de  $\mathbf{X}_{\text{obs}}^{[\bullet, j]}$ . Grâce au coefficient multiplicatif  $\frac{N}{N+1}$  pondérant  $\hat{F}_N$ , toutes les pseudo-observations sont situées strictement à l'intérieur de  $\mathcal{H}_d$ . Les valeurs prises par  $\hat{U}_j^{(i)}$  sont alors les ratios  $\frac{i}{N+1}$  pour  $i \in \{1, \dots, N\}$ . Ils correspondent aux rangs (normalisés) des observations  $X_j^{(i)}$  au sein de  $\mathbf{X}_{\text{obs}}^{[\bullet, j]}$ . Avec la procédure SP, les  $N$ -échantillons  $\hat{\mathbf{U}}_{\text{obs}}^{[\bullet, j]}$  s'apparentent parfaitement à des échantillons issus d'une loi uniforme. Certains auteurs comme [GENEST et FAVRE \[2007\]](#) mettent en avant le fait que le calcul des rangs est le meilleur résumé possible de la structure de dépendance. Selon eux, comme la copule est censée être déconnectée des lois marginales, son estimation ne devrait pas être pénalisée par le choix erroné d'une famille  $\mathcal{F}_j$ . En ce sens, la méthode SP semble plus fidèle à l'intuition initiale du théorème de [SKLAR \[1959\]](#). Dans la littérature, elle est aussi connue sous le nom de méthode du maximum de pseudo-vraisemblance (MPV) et [TSUKAHARA \[2005\]](#) relie la copule empirique et les propriétés asymptotiques de  $\hat{\boldsymbol{\theta}}^{\text{SP}}$ . S'il existe un débat entre IFM et SP, il n'en demeure pas moins que ces deux techniques ouvrent la voie à des procédures d'estimation pour tous types de modèles  $\mathcal{M}_1, \dots, \mathcal{M}_d$  et  $\mathcal{M}_{\text{cop}}$ . Il suffit d'appliquer les étapes suivantes :

- (a) Estimer les densités marginales  $f_{X_j}$  avec le modèle  $\mathcal{M}_j$  et les observations  $\mathbf{X}_{\text{obs}}^{[\bullet, j]}$ .
- (b) Utiliser les densités  $\hat{f}_{X_j}(\cdot | \mathcal{M}_j, \mathbf{X}_{\text{obs}}^{[\bullet, j]})$  pour fabriquer les fonctions de répartition  $\hat{F}_{X_j}(\cdot | \mathcal{M}_j, \mathbf{X}_{\text{obs}}^{[\bullet, j]})$ .
- (c) Fabriquer  $\hat{\mathbf{U}}_{\text{obs}}$  avec (1.129) ou avec (1.130) selon la préférence de l'utilisateur.
- (d) Estimer la densité de copule  $c_{\mathbf{U}}$  avec le modèle  $\mathcal{M}_{\text{cop}}$  et les observations  $\hat{\mathbf{U}}_{\text{obs}}$ .
- (e) Réunir toutes les composantes selon (1.77) pour former  $\hat{f}_{\mathbf{X}}(\cdot | \mathcal{M}, \mathbf{X}_{\text{obs}})$ .

## 1.5.2 Méthodes paramétriques pour l'apprentissage des copules

Pour une famille de copules  $\mathcal{F}_{\text{cop}}$ , l'estimation du paramètre  $\boldsymbol{\beta}$  peut être envisagée avec toutes les techniques de la Section 1.4.1 bien qu'il faille à chaque fois s'interroger sur celle qui est la plus adaptée au contexte. La méthode du maximum de vraisemblance reste de très loin la plus répandue et celle qui est utilisée par défaut. [JOE \[1997\]](#) insiste sur le fait qu'il ne faut pas espérer trouver

une expression analytique pour  $\hat{\boldsymbol{\beta}}^{MV}$  mais il loue les mérites des différentes méthodes de résolution numérique. Par ailleurs, [GENEST et RIVEST \[1993\]](#) proposent une variante de la méthode des moments dans le cas bidimensionnel lorsque  $\boldsymbol{\beta}$  se réduit à un unique paramètre scalaire  $\beta_{12}$ . Dans cette situation, il existe souvent une relation de type  $\tau_{12} = g(\beta_{12})$  entre le tau de Kendall  $\tau_{12}$  de la paire de variables et le paramètre  $\beta_{12}$  de la famille de copules. L'idée est alors d'estimer  $\tau_{12}$  directement à partir de  $\mathbf{X}_{\text{obs}}$  et d'en déduire une estimation de  $\hat{\beta}_{12}$ . Comme l'estimateur  $\hat{\tau}_{12}$  est uniquement calculé à partir des rangs au sein des colonnes  $\mathbf{X}_{\text{obs}}^{[\bullet 1]}$  et  $\mathbf{X}_{\text{obs}}^{[\bullet 2]}$ ,  $\hat{\beta}_{12}^M$  n'est pas détérioré par d'éventuelles valeurs aberrantes, ce qui est un avantage important par rapport à la méthode du maximum de vraisemblance. Il est parfois dit que  $\hat{\beta}_{12}$  est calculé par "inversion" du tau de Kendall. Dans un autre registre, [HUARD et collab. \[2006\]](#) et [DOS SANTOS SILVA et LOPES \[2008\]](#) abordent l'estimation des copules du point de vue bayésien mais cette approche est assez peu reprise par la suite. Lorsqu'il faut sélectionner ou valider une famille, la plupart des outils de la Section 1.4.1.4 peuvent être réutilisés. Les critères AIC et BIC définis aux équations (1.106) et (1.107) permettent de mesurer le niveau d'adéquation d'une famille  $\mathcal{F}_{\text{cop}}$  à un N-échantillon  $\mathbf{U}_{\text{obs}}$  de la copule. Il est souligné dans [JORDANGER et TJØSTHEIM \[2014\]](#) que AIC demeure le critère de prédilection malgré le développement de plusieurs versions d'un critère CIC (*Copula Information Criterion*) dans [GRØNNEBERG et HJORT \[2008\]](#) et dans [CLAESKENS et collab. \[2008\]](#). En ce qui concerne les tests d'adéquation, [FERMANIAN et collab. \[2005\]](#) puis [GENEST et collab. \[2009\]](#) ont recensé les méthodes existantes. Il y est fait mention du test développé par [BREYMAN et collab. \[2003\]](#) où une statistique du  $\chi^2$  est calculée après avoir appliqué la transformation de Rosenblatt. De leur côté, [GENEST et collab. \[2006\]](#) ont construit une statistique de test grâce à la transformation de Kendall étudiée dans [BARBE et collab. \[1996\]](#). De même, [GENEST et RÉMILLARD \[2008\]](#) ont étendu l'idée du test de Kolmogorov-Smirnov en comparant la copule estimée  $\hat{C}_{\mathbf{U}}(\cdot | \mathcal{F}_{\text{cop}}, \hat{\boldsymbol{\beta}})$  et la copule empirique  $\hat{C}_N$ . Lorsque  $d = 2$ , [GENEST et FAVRE \[2007\]](#) rappellent que des techniques de diagnostic graphique peuvent servir à la validation ou à la remise en cause du choix initial d'une famille. À cet égard, des procédures de diagnostic ont été spécifiquement conçues pour les copules archimédiennes dans [GENEST et RIVEST \[1993\]](#) et pour les copules aux valeurs extrêmes dans [CAPÉRAÀ et collab. \[1997\]](#).

Compte tenu de ces éléments, et pour éviter les répétitions, le but de la présente section n'est pas de redétailler le fonctionnement des techniques d'estimation mais plutôt d'étudier certaines familles de copules à travers les motifs de dépendance qu'elles permettent de modéliser. On se contente ici de présenter les copules elliptiques et archimédiennes. Un recueil plus exhaustif des familles inventées à ce jour est proposé par [NADARAJAH et collab. \[2018\]](#). D'une manière générale, dès lors qu'on dispose de la densité conjointe ou de la fonction de répartition conjointe d'une famille paramétrique, on peut en extraire la copule avec l'une des deux formules suivantes :

$$\forall \mathbf{u} \in \mathcal{H}_d, \quad C_{\mathbf{U}}(\mathbf{u}) = F_{\mathbf{X}}(F_{X_1}^{-1}(u_1), \dots, F_{X_d}^{-1}(u_d)) \quad \text{et} \quad c_{\mathbf{U}}(\mathbf{u}) = \frac{f_{\mathbf{X}}(F_{X_1}^{-1}(u_1), \dots, F_{X_d}^{-1}(u_d))}{\prod_{j=1}^d f_{X_j}(F_{X_j}^{-1}(u_j))}. \quad (1.131)$$

Dans le cas de la densité (1.51) d'un vecteur gaussien  $\mathbf{X} \sim \mathcal{N}(\mathbf{m}_{\mathbf{X}}, \boldsymbol{\Sigma}_{\mathbf{X}})$ , on obtient ainsi :

$$\begin{aligned} \bullet \quad c_{\mathcal{N}}(\mathbf{u} | \boldsymbol{\rho}_{\mathbf{X}}) &= \frac{1}{\sqrt{\det \boldsymbol{\rho}_{\mathbf{X}}}} \exp \left[ -\frac{1}{2} \mathbf{w}^{\top} (\boldsymbol{\rho}_{\mathbf{X}}^{-1} - \mathbf{I}_d) \mathbf{w} \right] \quad \text{avec} \quad \mathbf{w} := [\Phi^{-1}(u_j)]_{1 \leq j \leq d} \\ \bullet \quad C_{\mathcal{N}}(\mathbf{u} | \boldsymbol{\rho}_{\mathbf{X}}) &= \int_0^{u_1} \dots \int_0^{u_d} c_{\mathcal{N}}(\mathbf{v} | \boldsymbol{\rho}_{\mathbf{X}}) d\mathbf{v} = \Phi_{\boldsymbol{\rho}_{\mathbf{X}}}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)), \end{aligned} \quad (1.132)$$

avec  $\Phi$  et  $\Phi_{\boldsymbol{\rho}_{\mathbf{X}}}$  les fonctions de répartition des lois  $\mathcal{N}(0, 1)$  et  $\mathcal{N}(\mathbf{0}, \boldsymbol{\rho}_{\mathbf{X}})$ . On remarque que les fonctions  $c_{\mathcal{N}}$  et  $C_{\mathcal{N}}$  ne dépendent plus que de la matrice  $\boldsymbol{\rho}_{\mathbf{X}}$  des corrélations linéaires. La loi de Student multidimensionnelle présente des similarités importantes avec la loi gaussienne multidimensionnelle au sens où elle laisse aussi la part belle à la matrice de covariance  $\boldsymbol{\Sigma}_{\mathbf{X}}$ .

#### **Définition 1.5.1 (Distribution de Student multidimensionnelle)**

Soient  $\mathbf{Z}_1$  et  $\mathbf{Z}_2$  des vecteurs aléatoires indépendants à valeurs dans  $\mathbb{R}^d$  et  $\mathbb{R}^{\nu}$  ( $d, \nu \geq 2$ ).

Si  $\mathbf{Z}_1 \sim \mathcal{N}(\mathbf{m}, \Sigma)$  et  $\mathbf{Z}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_v)$  alors le vecteur  $\mathbf{X}$  défini par :

$$\mathbf{X} := \mathbf{Z}_1 / \sqrt{Y/v} \quad \text{avec} \quad Y = \|\mathbf{Z}_2\|_2^2 \sim \chi_v^2, \quad (1.133)$$

suit une loi de Student à  $v$  degrés de libertés qui est notée  $t_v(\mathbf{m}_X, \Sigma_X)$ .

La densité conjointe d'un vecteur  $\mathbf{X} \sim t_v(\mathbf{m}_X, \Sigma_X)$  est alors fournie par :

$$f_t(\mathbf{x} | \mathbf{m}_X, \Sigma_X, v) = \frac{\Gamma\left(\frac{v+d}{2}\right)}{\Gamma\left(\frac{v}{2}\right)(v\pi)^{d/2}} \times \left(\frac{1}{\sqrt{\det \Sigma_X}}\right) \times \left[1 + (\mathbf{x} - \mathbf{m}_X)^\top \Sigma_X^{-1} (\mathbf{x} - \mathbf{m}_X)\right]^{-\left(\frac{v+d}{2}\right)}, \quad (1.134)$$

où  $\Gamma : ]0, +\infty[ \rightarrow \mathbb{R}$  est la fonction gamma définie par :

$$\Gamma(z) := \int_0^{+\infty} t^{z-1} e^{-t} dt. \quad (1.135)$$

Contrairement à ce qu'on a pu voir à l'équation (1.132) pour la loi gaussienne, l'application de la formule (1.131) ne conduit pas à une expression analytique plus simple. Néanmoins, la copule résultante ne dépend plus que de  $v$  et  $\rho_X$ . Elle se note alors  $c_t(\cdot | v, \rho_X)$ . La loi gaussienne et la loi de Student appartiennent à une plus large famille de distributions multidimensionnelles appelées les distributions elliptiques. Les copules elliptiques en sont directement extraites.

### 1.5.2.1 Copules elliptiques

On commence par rappeler brièvement ce que sont les distributions sphériques et elliptiques. Pour plus de détails à ce sujet, on conseille de se reporter à l'ouvrage de FANG [2018].

#### Définition 1.5.2 (Distribution sphérique)

Soit  $\mathbf{X}$  un vecteur aléatoire à valeurs dans  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ .

Sa loi  $\mu_X$  est dite sphérique si et seulement si l'une des assertions suivantes est vérifiée :

- (i) Pour toute matrice orthogonale  $\mathbf{P}$ , on a :  $\mathbf{P}\mathbf{X} \stackrel{\text{loi}}{=} \mathbf{X}$ .
- (ii)  $\phi_X(\mathbf{t}) = \mathbb{E}\left[\exp(i\mathbf{t}^\top \mathbf{X})\right] = \psi(\mathbf{t}^\top \mathbf{t})$  avec  $\psi : \mathbb{R}^+ \rightarrow \mathbb{R}$ .
- (iii)  $\forall \boldsymbol{\lambda} \in \mathbb{R}^d, \boldsymbol{\lambda}^\top \mathbf{X} \stackrel{\text{loi}}{=} \|\boldsymbol{\lambda}\|_2 X_1$ .

La fonction  $\psi$  est appelée le générateur et on note  $\mathbf{X} \sim \mathcal{S}_d(\psi)$ .

Le point (i) témoigne de l'invariance par rotation d'un vecteur à loi sphérique. À partir de (ii), on peut voir que toutes les fonctions caractéristiques marginales valent  $\phi_{X_j}(t_j) = \psi(t_j^2)$  et donc que toutes les lois marginales sont identiques. De même, on peut vérifier que  $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$  est bien une loi sphérique. Par ailleurs, on a  $\mathbb{E}[\mathbf{X}] = \mathbf{0}$  par conséquence directe du point (iii).

#### Théorème 1.5.1 (Représentation stochastique d'une distribution sphérique)

Soit  $\mathbf{X}$  un vecteur aléatoire à valeurs dans  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ .

La loi  $\mu_X$  est sphérique si et seulement si on a :

$$\mathbf{X} \stackrel{\text{loi}}{=} \mathbf{R}\mathbf{S} \quad \text{avec} \quad \begin{cases} \mathbf{R} \text{ une variable positive indépendante de } \mathbf{S}, \\ \mathbf{S} \text{ un vecteur uniformément distribué sur la sphère unité } \mathbb{S}_d. \end{cases} \quad (1.136)$$

La sphère unité est définie par  $\mathbb{S}_d := \{\mathbf{x} \in \mathbb{R}^d : x_1^2 + \dots + x_d^2 = 1\}$ . Pour simuler  $\mathbf{S}$  uniformément sur  $\mathbb{S}_d$ , il est rappelé dans CHAFAÏ et MALRIEU [2016] qu'il suffit de simuler un vecteur  $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$  et de prendre  $\mathbf{X}/\|\mathbf{X}\|_2$ . Dans ce format de représentation, l'information du générateur  $\psi$  est reportée sur la loi de  $\mathbf{R}$ . Comme il est supposé que  $\mathbf{R} \geq 0$ , s'il existe une densité de probabilité  $f_R$ , celle-ci peut toujours être mise sous la forme suivante :

$$\forall r \in \mathbb{R}^+, \quad f_R(r) = s_d r^{d-1} g(r^2) \quad \text{avec} \quad s_d := \frac{2\pi^{d/2}}{\Gamma(d/2)}. \quad (1.137)$$

On fait remarquer que la constante de renormalisation  $s_d$  est égale à l'aire en surface de  $\mathbb{S}_d$ . La densité conjointe vaut alors  $f_{\mathbf{X}}(\mathbf{x}) = g(\mathbf{x}^\top \mathbf{x})$  et ses lignes d'isovaleurs sont des cercles concentriques centrés sur l'origine du repère. Les distributions elliptiques s'obtiennent alors par distorsion des lois sphériques.

**Définition 1.5.3 (Distribution elliptique)**

Soit  $\mathbf{m} \in \mathbb{R}^d$  et soit  $\mathbf{A} \in \mathbb{R}^{d \times d}$  telle que  $\Sigma = \mathbf{A}\mathbf{A}^\top > 0$ .  
 Si  $\mathbf{Z} \sim \mathcal{S}_d(\psi)$  alors le vecteur  $\mathbf{X} = \mathbf{m} + \mathbf{A}\mathbf{Z}$  suit la loi elliptique notée  $\mathcal{E}_d(\mathbf{m}, \Sigma, \psi)$ .

De par le Théorème 1.5.1, toute loi elliptique est représentée par un vecteur de type  $\mathbf{X} \stackrel{\text{loi}}{=} \mathbf{m} + \mathbf{R}\mathbf{A}\mathbf{S}$ . De plus, la fonction caractéristique prend la forme suivante :

$$\phi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E} \left[ \exp(i\mathbf{t}^\top \mathbf{X}) \right] = \exp(i\mathbf{t}^\top \mathbf{m}) \phi_{\mathbf{Z}}(\mathbf{A}^\top \mathbf{t}) = \exp(i\mathbf{t}^\top \mathbf{m}) \psi(\mathbf{t}^\top \Sigma \mathbf{t}). \quad (1.138)$$

Comme précédemment, les fonctions caractéristiques marginales  $\phi_{X_j}$  sont égales, ce qui indique que les lois marginales le sont aussi. L'expression de  $\phi_{\mathbf{X}}$  permet d'ailleurs de démontrer que :

$$\forall \mathbf{b} \in \mathbb{R}^d, \quad \forall \mathbf{B} \in \mathbb{R}^{d \times d}, \quad \mathbf{X} \sim \mathcal{E}_d(\mathbf{m}, \Sigma, \psi) \implies \mathbf{B}\mathbf{X} + \mathbf{b} \sim \mathcal{E}_d(\mathbf{B}\mathbf{m} + \mathbf{b}, \mathbf{B}\Sigma\mathbf{B}^\top, \psi). \quad (1.139)$$

Ce résultat de stabilité de la classe elliptique par transformation affine est d'ailleurs bien connu pour les lois gaussiennes. En utilisant le même système de paramétrage qu'à l'équation (1.137), la densité conjointe peut alors s'écrire :

$$\forall \mathbf{x} \in \mathbb{R}^d, \quad f_{\mathbf{X}}(\mathbf{x}) := \frac{1}{\sqrt{\det \Sigma}} g\left(\frac{(\mathbf{x} - \mathbf{m})^\top \Sigma^{-1} (\mathbf{x} - \mathbf{m})}{\det \Sigma}\right). \quad (1.140)$$

Les densités (1.51) et (1.134) rentrent évidemment dans ce cadre. Les lignes d'isovaleurs de  $f_{\mathbf{X}}$  sont alors des ellipses concentriques (autour de  $\mathbf{m}$ ) et toutes de même excentricité. La théorie des distributions elliptiques permet ainsi de rassembler les distributions gaussiennes et de Student au sein d'une même classe. Toutefois, de façon assez étonnante, aucun autre modèle n'a réellement émergé alors même qu'on peut construire autant de familles qu'il y a de fonctions  $\psi : \mathbb{R}^+ \rightarrow \mathbb{R}$ . Examinons maintenant ce qu'est une copule elliptique.

**Définition 1.5.4 (Copule elliptique)**

Soit  $\mathbf{X} \sim \mathcal{E}_d(\mathbf{m}, \Sigma, \psi)$  et  $F_\psi$  la fonction de répartition commune des variables  $X_j$ .  
 La copule elliptique associée à  $\mathcal{E}_d(\mathbf{m}, \Sigma, \psi)$  est la loi du vecteur :

$$\mathbf{U} := \left[ F_\psi \left( \frac{X_j - m_j}{\sigma_j} \right) \right]_{1 \leq j \leq d}. \quad (1.141)$$

Cette copule est alors notée  $\mathcal{E}_d^C(\boldsymbol{\rho}, \psi)$ .

La construction des copules elliptiques revient simplement à extraire la structure de dépendance intrinsèque des distributions elliptiques. Toute copule elliptique ne dépend que du générateur  $\psi$  et de la matrice de corrélation  $\boldsymbol{\rho}$ . On souligne que le générateur  $\psi_\nu$  de la copule de Student inclut l'information relative au nombre  $\nu$  de degrés de liberté. Le fait que les copules elliptiques soient articulées autour de la matrice de corrélation présente deux intérêts majeurs :

1. Une copule elliptique incorpore au moins  $C_d^2 = \frac{1}{2}d(d-1)$  paramètres. Ce nombre augmente donc avec la dimension  $d$  du problème et cela confère une certaine flexibilité.
2. L'interprétation est facile car il y a autant de coefficients  $\rho_{ij}$  que de paires de variables  $(U_i, U_j)$ .

Ces raisons expliquent l'attrait exercé par les copules elliptiques. Souvent, on choisit de combiner une copule elliptique avec des lois marginales d'une tout autre nature. La distribution résultante est dite méta-elliptique dans FANG et collab. [2002]. Le principal défaut des copules elliptiques se

trouve dans leur incapacité à proposer une modélisation asymétrique des queues de distribution. En effet, on a :

Copule	$c_{\mathcal{N}}(\cdot   \boldsymbol{\rho})$	$c_t(\cdot   \boldsymbol{\rho}, \nu)$
$\lambda_L = \lambda_U$	0	$2T_{\nu+1} \left( -\sqrt{\frac{(\nu+1)(1-\rho_{ij})}{1+\rho_{ij}}} \right)$

(1.142)

avec  $T_{\nu+1}$  la fonction de répartition d'une loi de Student centrée réduite à  $(\nu + 1)$  degrés de liberté. Lorsque  $\nu \rightarrow \infty$ , les coefficients  $\lambda_L$  et  $\lambda_U$  tendent vers 0 et on retrouve le résultat bien connu selon lequel  $t_\nu(\mathbf{m}, \boldsymbol{\Sigma}) \xrightarrow[\nu \rightarrow \infty]{\text{loi}} \mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma})$ . Les résultats (1.142) indiquent que :

- (a) La copule gaussienne ne permet pas de capturer la dépendance asymptotique.
- (b) Les queues NE et SW d'une paire  $(U_i, U_j)$  sont modélisées symétriquement ( $\lambda_L = \lambda_U$ ).

Des résultats généraux sont fournis dans HULT et LINDSKOG [2002] ainsi que SCHMIDT [2002] pour les éventuelles autres copules elliptiques et ils vont dans le même sens. En dimension  $d = 2$ , les copules archimédiennes permettent de remédier au défaut (b).

### 1.5.2.2 Copules archimédiennes

Comme pour les copules elliptiques, la structure de dépendance d'une copule archimédienne repose sur une fonction appelée générateur. Si toute fonction  $\psi : \mathbb{R}^+ \rightarrow \mathbb{R}$  peut être le générateur d'une copule elliptique, il n'en va pas de même pour le générateur d'une copule archimédienne qui doit vérifier des conditions très strictes.

#### Définition 1.5.5 (Générateur archimédien)

On appelle générateur archimédien toute fonction  $\psi : [0, +\infty[ \rightarrow ]0, 1]$  telle que :

1.  $\psi$  est décroissante et continue sur  $[0, +\infty[$ .
2.  $\psi(0) = 1$  et  $\lim_{t \rightarrow \infty} \psi(t) = 0$ .
3.  $\psi$  est strictement décroissante sur  $[0, T_0[$  avec  $T_0 := \inf\{t \geq 0 : \psi(t) = 0\}$ .

Les fonctions  $t \mapsto e^{-t}$  et  $t \mapsto 1 - t^\alpha$  sont des exemples de générateurs archimédiens. Sous les conditions mentionnées ci-dessus,  $\psi$  est inversible et la réciproque se note  $\psi^{-1} : ]0, 1] \rightarrow [0, T_0[$ . Si le formalisme du générateur archimédien remonte à GENEST et MACKAY [1986], la plupart des copules archimédiennes lui sont antérieures. Par ailleurs, pour disposer de résultats intéressants sur les copules archimédiennes, il faut que  $\psi$  soit un peu plus qu'un générateur archimédien. Cela conduit à définir quelques notions supplémentaires. Dans la suite, on se contente d'étudier des générateurs stricts, c'est-à-dire des générateurs pour lesquels  $T_0 = \infty$ .

#### Définition 1.5.6 (Monotonie)

Un générateur archimédien  $\psi$  est dit  $d$ -monotone si :

1.  $\psi$  est au moins  $d - 2$  fois continûment dérivable sur  $[0, +\infty[$ .
2.  $\forall 1 \leq k \leq d - 2, \forall t \in [0, +\infty[, (-1)^k \frac{d^k}{dt^k} \psi(t) \geq 0$ .
3.  $\frac{d^{d-2}}{dt^{d-2}} \psi$  est une fonction décroissante et convexe sur  $[0, +\infty[$ .

Un générateur archimédien  $\psi$  est dit totalement monotone si :

1.  $\psi$  est infiniment dérivable sur  $[0, +\infty[$ .
2.  $\forall k \geq 0, \forall t \in [0, +\infty[, (-1)^k \frac{d^k}{dt^k} \psi(t) \geq 0$ .

On note  $\Psi_\infty$  l'ensemble des générateurs archimédiens totalement monotones.

MCNEIL et collab. [2009] ont montré que l'hypothèse de  $d$ -monotonie est suffisante pour définir une copule archimédienne.

### Théorème 1.5.2 (Copule archimédienne)

Si  $\psi$  un générateur archimédien  $d$ -monotone, alors l'application :

$$\begin{aligned} C_\psi : \mathcal{H}_d &\longrightarrow [0, 1] \\ \mathbf{u} &\longrightarrow \psi \left( \sum_{j=1}^d \psi^{-1}(u_j) \right) , \end{aligned} \quad (1.143)$$

est la fonction de répartition d'une  $d$ -copule. Cette loi est notée  $\mathcal{A}_d(\psi)$ .

Dans le travail initial de **GENEST et MACKEY [1986]**, les copules archimédiennes sont examinées dans un cadre bidimensionnel et il est supposé que  $\psi$  est deux fois dérivable. La  $d$ -monotonie est une hypothèse plus faible qui suffit à vérifier les conditions du Théorème 1.2.4 et à justifier que  $C_\psi$  est bien la fonction de répartition d'une copule. On fait remarquer que la copule  $C_\psi$  peut être engendrée par tout générateur  $\psi_k : t \mapsto \psi(kt)$  avec  $k > 0$ . Parmi les nombreux modèles de copules archimédiennes, on se contente de présenter les trois plus célèbres. **GUMBEL [1960]** et **HOUGAARD [1984]** se sont intéressés à une copule qui a gardé le nom du premier :

$$\forall \theta \geq 1, \quad \psi(t) = \exp\left(-t^{\frac{1}{\theta}}\right) \quad \text{et} \quad C_\psi(\mathbf{u}) = \exp\left(-\left[\sum_{j=1}^d (-\log u_j)^\theta\right]^{\frac{1}{\theta}}\right). \quad (1.144)$$

Les travaux de **FRANK [1979]** ont contribué à populariser la copule du même nom :

$$\forall \theta > 0, \quad \psi(t) = -\frac{1}{\theta} \log\left(1 - (1 - e^{-\theta}) e^{-t}\right) \quad \text{et} \quad C_\psi(\mathbf{u}) = -\frac{1}{\theta} \log\left(1 + \frac{\prod_{j=1}^d (e^{-\theta} u_j - 1)}{(e^{-\theta} - 1)^{d-1}}\right). \quad (1.145)$$

De même, **CLAYTON [1978]** a développé un modèle :

$$\forall \theta > 0, \quad \psi(t) = (1 + t)^{-\frac{1}{\theta}} \quad \text{et} \quad C_\psi(\mathbf{u}) = \left(1 + \sum_{j=1}^d u_j^{-\theta} - d\right)^{-\frac{1}{\theta}}. \quad (1.146)$$

qui a ensuite été repris par **COOK et JOHNSON [1981]** et **OAKES [1982]**. Pour un inventaire plus complet des générateurs archimédiens et des copules associées, on peut consulter celui proposé dans **JOE [2014]**. Contrairement à la copule gaussienne pour laquelle l'équation (1.132) fournit une expression analytique de  $c_{\mathcal{N}}$  mais une expression implicite de  $C_{\mathcal{N}}$ , toute copule archimédienne est construite à partir de la connaissance explicite de  $C_\psi$ . La densité de copule  $c_\psi = \frac{\partial^d C_\psi}{\partial u_1 \dots \partial u_d}$  est obtenue par différentiation mais les calculs ne peuvent plus être conduits dès lors que  $d \geq 3$ . Il faut alors se tourner vers des outils de différentiation automatique ou numérique.

S'il est facile de simuler des copules elliptiques en vertu de la Définition 1.5.4, la question de la simulation des copules archimédiennes est plus délicate à appréhender. Pour tout générateur archimédien totalement monotone, **HOFERT [2008]** rappelle que le théorème de Bernstein assure qu'il existe une variable  $W \geq 0$  de transformée de Laplace  $\psi$  :

$$\forall t \in [0, +\infty], \quad \psi(t) = \mathbb{E}[e^{tW}] = \int_0^\infty e^{tw} f_W(w) dw. \quad (1.147)$$

La densité  $f_W$  est alors la transformée de Laplace inverse du générateur archimédien  $\psi$ . Si on est capable d'identifier la densité  $f_W$ , l'algorithme de **MARSHALL et OLKIN [1988]** permet de simuler des échantillons suivant la copule archimédienne de générateur  $\psi$ . Dans la suite, on se contente d'étudier les copules archimédiennes pour  $d = 2$ . Comme le Théorème 1.5.2 donne :

$$\mathbf{U} \sim \mathcal{A}_d(\psi) \implies \forall 1 \leq i < j \leq d, \quad (U_i, U_j) \sim \mathcal{A}_2(\psi), \quad (1.148)$$

chaque copule bidimensionnelle résume parfaitement le motif de dépendance induit par  $\psi$ . Pour les trois familles archimédiennes présentées plus haut, les coefficients de dépendance de queue valent respectivement :

Copule	Clayton (C)	Gumbel (G)	Frank (F)
$\lambda_L = \lambda_{SW}$	$2^{\frac{1}{\theta}}$	0	0
$\lambda_U = \lambda_{NE}$	0	$2 - 2^{\frac{1}{\theta}}$	0

(1.149)

Au niveau de la dépendance asymptotique, la copule de Frank se comporte exactement comme la copule gaussienne avec  $\lambda_L = \lambda_U = 0$ . Pour la copule de Clayton, un phénomène de dépendance extrême est présent dans la queue SW mais pas dans la queue NE. De son côté, la copule de Gumbel est marquée par une tendance antagoniste. Ces deux copules permettent d'introduire de l'asymétrie dans la modélisation des queues de distribution. Elles ne sont par contre pas adaptées à la description des motifs de dépendance discordants. **GENEST et MACKEY [1986]** ont démontré que le tau de Kendall d'une copule  $\mathcal{A}_2(\psi)$  peut s'exprimer sous la forme :

$$\tau = 1 - 4 \int_0^{\infty} t (\psi'(t))^2 dt . \quad (1.150)$$

Pour la copule de Gumbel, on obtient  $\tau = \frac{\theta-1}{\theta}$  avec  $\theta \in [1, +\infty[$  et on a donc toujours  $\tau \geq 0$ . Pour résorber ce problème, il est possible d'opérer des rotations de la copule. En effet, si la loi du vecteur  $\mathbf{U} = (U_1, U_2)$  est une copule, celles des trois vecteurs  $\mathbf{U}_{90} := (1 - U_1, U_2)$ ,  $\mathbf{U}_{180} := (1 - U_1, 1 - U_2)$  et  $\mathbf{U}_{270} := (U_1, 1 - U_2)$  sont aussi des copules. Elles correspondent à des rotations de respectivement  $90^\circ$ ,  $180^\circ$  et  $270^\circ$  autour du point  $(1/2, 1/2)$  de  $\mathcal{H}_2$ . Trois nouveaux motifs de dépendance peuvent ainsi être créés. Dans le cas d'une rotation de  $90^\circ$  de la copule de Gumbel  $C_G$ , on obtient :

$$\forall \mathbf{u} \in \mathcal{H}_2, \quad C_G^{90}(\mathbf{u}) = \mathbb{P}(\mathbf{U}_{90} \leq \mathbf{u}) = \mathbb{P}(1 - U_1 \leq u_1, U_2 \leq u_2) = u_2 - C_G(1 - u_1, u_2) . \quad (1.151)$$

Le tau de Kendall vaut alors  $\tau_K(1 - U_1, U_2) = -\tau_K(U_1, U_2) = \frac{1-\theta}{\theta}$  et il devient compatible avec la modélisation d'un motif de dépendance discordant. De même, la dépendance asymptotique est désormais présente dans la queue NW (*North West*) étant donné que  $\lambda_{NW}(C_G^{90}) = \lambda_{NE}(C_G) = 2 - 2^{\frac{1}{\theta}}$ . Si on souhaite que la dépendance asymptotique affecte simultanément les queues NE et NW, on peut construire un mélange  $\tilde{C} = \frac{1}{2}(C_G + C_G^{90})$  pour lequel  $\lambda_{NE}(\tilde{C}) = \lambda_{NW}(\tilde{C}) = \frac{1}{2}(2 - 2^{\frac{1}{\theta}})$ . Ainsi, même à partir d'un catalogue réduit de familles archimédiennes, on peut créer une multitude de familles paramétriques en s'appuyant sur des rotations et des mélanges.

En dépit de leur aptitude à générer de l'asymétrie au sein des structures de dépendance, les copules archimédiennes souffrent d'un cruel manque de flexibilité. Lorsque  $d$  augmente, le système de paramétrage reste fondé sur le seul paramètre scalaire  $\theta$  du générateur archimédien  $\psi$ . Par ailleurs, le fait que toutes les paires  $(U_i, U_j)$  d'un vecteur  $\mathbf{U}$  suivent la même copule  $\mathcal{A}_2(\psi)$  est évidemment une limite importante de ces modèles. On souhaite attirer l'attention sur le fait que la copule de Gumbel définie à l'équation (1.144) vérifie :

$$\forall n \geq 1, \quad \forall \mathbf{u} \in \mathcal{H}_d, \quad C_G(\mathbf{u}) = \left( C_G(u_1^{1/n}, \dots, u_d^{1/n}) \right)^n . \quad (1.152)$$

Il s'agit de la propriété de max-stabilité et elle détermine l'appartenance à la classe des copules aux valeurs extrêmes. Ces copules ont été introduites par **GALAMBOS [1978]** et **DEHEUVELS [1984]** au moment du développement de la théorie multidimensionnelle des valeurs extrêmes. Pour un échantillon  $\mathbf{X}_{obs}$ , on pourrait s'intéresser au comportement conjoint des maxima, c'est-à-dire à :

$$\mathbf{M}^N := \left[ M_j^N \right]_{1 \leq j \leq d} \quad \text{avec} \quad M_j^N := \max_{1 \leq i \leq N} X_j^{(i)} . \quad (1.153)$$

Le théorème de Fisher-Tippett-Gnedenko indique qu'il existe des suites renormalisantes assurant la convergence de  $\mathbf{M}^N$  vers une loi d'extrema généralisée. Les copules aux valeurs extrêmes sont

exactement les copules issues de ces lois d'extrema généralisées. À l'instar de la copule de Gumbel, celles des copules archimédiennes qui sont aussi des copules aux valeurs extrêmes appartiennent à la classe Archimax qui est étudiée dans [CHARPENTIER et collab. \[2014\]](#).

Trois familles centrales dans la théorie des copules viennent d'être présentées : les copules elliptiques, les copules archimédiennes et les copules aux valeurs extrêmes. Suite aux lacunes constatées pour chacune de ces familles, on s'intéresse ensuite aux techniques d'apprentissage non-paramétriques.

### 1.5.3 Méthodes non-paramétriques pour l'apprentissage des copules

On présente ici les principales méthodes d'estimation non-paramétrique d'une copule. D'une manière générale, il s'agit de généralisations des techniques décrites à la Section 1.4.2 dans le cadre de l'apprentissage non-paramétrique des lois marginales. Cependant, en raison de la compacité du support  $\mathcal{H}_d$ , certains aménagements sont nécessaires et on en explique les principaux ressorts. Par souci de simplicité, il est supposé qu'on dispose d'un N-échantillon  $\mathbf{U}_{\text{obs}}$ . En pratique, comme il faut travailler à partir de pseudo-observations, certains des résultats théoriques énoncés ci-après peuvent être altérés.

#### 1.5.3.1 Copule empirique

En suivant l'idée de [DEHEUVELS \[1979\]](#) et [STUTE \[1984\]](#), les fonctions de répartition associées aux lois  $\mu_{\mathbf{X}}$  et  $\mu_{\mathbf{U}}$  peuvent être estimées empiriquement :

$$\forall \mathbf{x} \in \mathbb{R}^d, \hat{F}_{\mathbf{N}}(\mathbf{x}) := \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\{X_1^{(i)} \leq x_1, \dots, X_d^{(i)} \leq x_d\}} \text{ et } \forall \mathbf{u} \in \mathcal{H}_d, \hat{C}_{\mathbf{N}}(\mathbf{u}) := \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\{U_1^{(i)} \leq u_1, \dots, U_d^{(i)} \leq u_d\}}. \quad (1.154)$$

La convergence en loi (1.112) du théorème de Donsker peut être généralisée à  $\sqrt{N}(\hat{F}_{\mathbf{N}} - F_{\mathbf{X}})$  où  $\hat{F}_{\mathbf{N}}$  est la fonction de répartition empirique d'un vecteur aléatoire  $\mathbf{X}$  à valeurs dans  $\mathbb{R}^d$ . Dans ce contexte multidimensionnel, le processus limite est un  $F_{\mathbf{X}}$ -pont brownien. Grâce à ce résultat, des tests d'adéquation de type Kolmogorov-Smirnov sont construits par [DEHEUVELS \[1981\]](#) pour valider le choix d'une famille paramétrique. Comme l'hypercube  $\mathcal{H}_d$  est un domaine borné, il n'est pas possible de conclure directement au sujet de la convergence du processus  $\sqrt{N}(\hat{C}_{\mathbf{N}} - C_{\mathbf{U}})$ . Après avoir plongé  $C_{\mathbf{U}}$  dans l'espace  $\ell^\infty(\mathcal{H}_d)$  des fonctions bornées sur  $\mathcal{H}_d$ , [FERMANIAN et collab. \[2004\]](#) puis [TSUKAHARA \[2005\]](#) ont démontré qu'il est nécessaire que  $C_{\mathbf{U}}$  soit continûment différentiable. Si c'est effectivement le cas, alors on a :

$$\sqrt{N}(\hat{C}_{\mathbf{N}} - C_{\mathbf{U}}) \xrightarrow[N \rightarrow \infty]{\text{loi}} \mathbb{G}_{C_{\mathbf{U}}} \text{ avec } \mathbb{G}_{C_{\mathbf{U}}}(\mathbf{u}) = \mathbb{B}_{C_{\mathbf{U}}}(\mathbf{u}) - \sum_{j=1}^d \mathbb{B}_{C_{\mathbf{U}}}(1, \dots, 1, u_j, 1, \dots, 1) \frac{\partial C_{\mathbf{U}}}{\partial u_j}(\mathbf{u}), \quad (1.155)$$

où  $\mathbb{B}_{C_{\mathbf{U}}}$  est un  $C_{\mathbf{U}}$ -pont brownien. On constate que le processus  $\mathbb{G}_{C_{\mathbf{U}}}$  est un  $C_{\mathbf{U}}$ -pont brownien auquel s'ajoutent des effets de bords. [SEGGERS et collab. \[2012\]](#) précisent que le résultat (1.155) tient même avec des hypothèses légèrement moins fortes. En dépit de cette propriété de convergence, la copule  $\hat{C}_{\mathbf{N}}$  présente les défauts inhérents à toute mesure empirique. Au niveau de la visualisation graphique, son aspect pyramidal peut être dérangeant. De plus, il n'est pas possible d'en déduire une estimation de la densité de copule  $c_{\mathbf{U}}$  car il s'agit d'une loi discrète sur les observations  $\mathbf{U}^{(i)}$ .

#### 1.5.3.2 Adaptations de la méthode d'estimation par noyaux

L'idée d'utiliser une approche KDE multidimensionnelle pour estimer la densité de copule  $c_{\mathbf{U}}$  est apparue dès [DEHEUVELS \[1979\]](#) et [BEHNEN et collab. \[1985\]](#). Une application directe donne :

$$\forall \mathbf{u} \in \mathcal{H}_d, \hat{c}_{\mathbf{H}}^{\text{KDE}}(\mathbf{u}) := \frac{1}{N} \sum_{i=1}^N K_{\mathbf{H}}(\mathbf{u} - \mathbf{U}^{(i)}) \text{ avec } K_{\mathbf{H}}(\mathbf{v}) = \frac{1}{\det \mathbf{B}} K_d(\mathbf{B}^{-1} \mathbf{v}) \text{ et } \mathbf{H} = \mathbf{B} \mathbf{B}^{\top}. \quad (1.156)$$

On signale que rien ne garantit que la fonction  $\hat{c}_{\mathbf{H}}^{\text{KDE}}$  soit la densité d'une  $d$ -copule. S'il est choisi de faire simple en posant  $\mathbf{H} = h\mathbf{I}_d$  et en fabriquant le noyau multidimensionnel  $K_d$  par produit tensoriel d'un noyau unidimensionnel  $K_1$ , on obtient finalement :

$$\forall \mathbf{u} \in \mathcal{H}_d, \quad \hat{c}_h(\mathbf{u}) := \frac{1}{h^d N} \sum_{i=1}^N K_1 \left( \frac{u_1 - \mathbf{U}_1^{(i)}}{h} \right) \dots K_1 \left( \frac{u_d - \mathbf{U}_d^{(i)}}{h} \right). \quad (1.157)$$

Comme l'expliquent [CHARPENTIER et collab. \[2007\]](#), l'utilisation de cet estimateur naïf conduit à une forte sous-estimation des valeurs de la densité de copule  $c_{\mathbf{U}}$ . Dans le cas de l'estimation d'une densité marginale par KDE, il a été vu à l'équation (1.121) que  $\hat{f}_h(z_0)$  constitue un estimateur biaisé de la valeur  $f_Z(z_0)$  mais que ce biais disparaît asymptotiquement lorsque  $h \rightarrow 0$ . Dans le cas de l'estimation d'une densité de copule par KDE, on distingue deux situations selon la position du point  $\mathbf{u}_0$  au sein de l'hypercube. Si  $\mathbf{u}_0$  est situé strictement à l'intérieur de  $\mathcal{H}_d$ , l'estimateur  $\hat{c}_h(\mathbf{u}_0)$  reste asymptotiquement non biaisé mais le biais est d'ordre  $\mathcal{O}(h)$  alors qu'il était d'ordre  $\mathcal{O}(h^2)$  pour  $\hat{f}_h(z_0)$ . En revanche, pour un point  $\mathbf{u}_0$  placé sur un bord de  $\mathcal{H}_d$ , on a  $\mathbb{E}[\hat{c}_h(\mathbf{u}_0)] = k c_{\mathbf{U}}(\mathbf{u}_0) + \mathcal{O}(h)$  avec  $k < 1$  une constante multiplicative indésirable. L'estimateur est donc biaisé, même asymptotiquement. L'explication de ce phénomène réside dans la manière dont la densité  $\hat{c}_h$  est construite. Les observations  $\mathbf{U}^{(i)}$  positionnées au voisinage des bords de  $\mathcal{H}_d$  créent une contribution  $\mathbf{u} \mapsto K_{\mathbf{H}}(\mathbf{u} - \mathbf{U}^{(i)})$  dont une partie du support, et donc de la "masse", débordent à l'extérieur de  $\mathcal{H}_d$ . L'intégrale de la densité  $\hat{c}_h$  n'est alors plus égale à 1. Si ce problème est caricatural lorsque  $K_1$  est le noyau gaussien, il affecte aussi les noyaux à support compact tel que le noyau d'Epanechnikov. Plusieurs solutions ont été développées pour éliminer les nuisances causées par les effets de bords.

La méthode dite de l'image miroir (IM) est proposée par [DEHEUVELS \[1979\]](#) avant d'être étudiée plus en détail par [GIJBELS et MIELNICZUK \[1990\]](#). Elle consiste à compenser la perte de masse en insérant à l'extérieur de  $\mathcal{H}_d$  des observations supplémentaires dont le noyau débordent à l'intérieur de  $\mathcal{H}_d$ . Dans le cas de l'estimation d'une copule en dimension  $d = 2$ , pour chaque observation  $\mathbf{U}^{(i)} = (\mathbf{U}_1^{(i)}, \mathbf{U}_2^{(i)})$ , il faut fabriquer huit observations factices notées  $\mathbf{U}_{[k]}^{(i)}$  avec  $1 \leq k \leq 8$ . Il s'agit de l'image de  $\mathbf{U}^{(i)}$  pour les symétries axiales par rapport aux quatre bords et pour les symétries centrales par rapport aux quatre coins. Supposons qu'une observation  $\mathbf{U}^{(i)}$  soit située à proximité d'un bord mais suffisamment loin des deux coins de ce bord. La perte de masse est unilatérale et elle est compensée par l'action du vis-à-vis de  $\mathbf{U}^{(i)}$  pour ce bord. Tous les autres points sont alors placés très loin à l'extérieur de  $\mathcal{H}_d$  et les noyaux associés ne débordent pas à l'intérieur de  $\mathcal{H}_d$ . Supposons maintenant que  $\mathbf{U}^{(i)}$  soit situé à proximité d'un coin. La perte de masse est à la fois bilatérale et diagonale. Trois points factices sont sollicités et l'action simultanée de leurs noyaux réintroduit la masse perdue. Finalement, si  $\mathbf{U}_{\text{obs}}$  comptabilise  $N$  observations  $\mathbf{U}^{(i)}$ , la construction d'une densité de copule par la méthode IM met en jeu  $9N$  noyaux :

$$\forall \mathbf{u} \in \mathcal{H}_2, \quad \hat{c}_{\mathbf{H}}^{\text{IM}}(\mathbf{u}) := \frac{1}{h^2 N} \sum_{i=1}^N \left( K_{\mathbf{H}}(\mathbf{u} - \mathbf{U}^{(i)}) + \sum_{k=1}^8 K_{\mathbf{H}}(\mathbf{u} - \mathbf{U}_{[k]}^{(i)}) \right) \quad \text{avec}$$

$$\begin{aligned} \mathbf{U}^{(i)} &= (\mathbf{U}_1^{(i)}, \mathbf{U}_2^{(i)}) & ; & \quad \mathbf{U}_{[1]}^{(i)} = (\mathbf{U}_1^{(i)}, -\mathbf{U}_2^{(i)}) & \quad ; & \quad \mathbf{U}_{[2]}^{(i)} = (\mathbf{U}_1^{(i)}, 2 - \mathbf{U}_2^{(i)}) & \quad ; & \quad (1.158) \\ \mathbf{U}_{[3]}^{(i)} &= (-\mathbf{U}_1^{(i)}, \mathbf{U}_2^{(i)}) & ; & \quad \mathbf{U}_{[4]}^{(i)} = (2 - \mathbf{U}_1^{(i)}, \mathbf{U}_2^{(i)}) & \quad ; & \quad \mathbf{U}_{[5]}^{(i)} = (-\mathbf{U}_1^{(i)}, -\mathbf{U}_2^{(i)}) & \quad ; & \\ \mathbf{U}_{[6]}^{(i)} &= (-\mathbf{U}_1^{(i)}, 2 - \mathbf{U}_2^{(i)}) & ; & \quad \mathbf{U}_{[7]}^{(i)} = (2 - \mathbf{U}_1^{(i)}, -\mathbf{U}_2^{(i)}) & \quad ; & \quad \mathbf{U}_{[8]}^{(i)} = (2 - \mathbf{U}_1^{(i)}, 2 - \mathbf{U}_2^{(i)}) & \quad . & \end{aligned}$$

Grâce à la méthode IM, on peut supprimer l'impact du coefficient  $k$  et l'estimateur  $\hat{c}_h^{\text{IM}}(\mathbf{u}_0)$  est asymptotiquement sans biais pour tout point  $\mathbf{u}_0 \in \mathcal{H}_d$ . Par contre, la vitesse de convergence reste d'ordre  $\mathcal{O}(h)$  sur les bords de l'hypercube.

Une autre méthode est introduite par [CHARPENTIER et collab. \[2007\]](#) et cherche à disperser dans  $\mathbb{R}^d$  les observations initiales  $\mathbf{U}^{(i)}$ . Pour cela, il est proposé de faire appel à des fonctions de répartition  $G_1, \dots, G_d$  strictement croissantes de  $\mathbb{R}$  dans  $]0, 1[$  et continûment dérivables. On note alors  $g_j$  la densité de probabilité découlant de  $G_j$ . Ensuite, l'idée est naturellement de définir

les variables  $\tilde{X}_j := G_j^{-1}(U_j)$  et d'estimer avec une méthode KDE multidimensionnelle la densité conjointe  $f_{\tilde{\mathbf{X}}}$ . De par le Théorème 1.2.6,  $\tilde{\mathbf{X}}$  et  $\mathbf{U}$  partagent la même copule  $c_{\mathbf{U}}$  mais les valeurs de  $\tilde{\mathbf{X}}$  peuvent exister dans  $\mathbb{R}^d$  alors que celles de  $\mathbf{U}$  sont confinées dans  $\mathcal{H}_d$ . Pour cette méthode parfois appelée TKE (*Transformed-Kernel Estimation*), la formule (1.131) offre un estimateur de  $c_{\mathbf{U}}$  :

$$\begin{aligned} \forall \mathbf{u} \in \mathcal{H}_d, \quad \hat{c}_{\mathbf{H}}^{\text{TKE}}(\mathbf{u}) &:= \hat{f}_{\mathbf{H}}(G_1^{-1}(u_1), \dots, G_d^{-1}(u_d) \mid \tilde{\mathbf{X}}_{\text{obs}}) \Big/ \prod_{j=1}^d g_j(G_j^{-1}(u_j)) \\ &= \frac{1}{N} \sum_{i=1}^N K_{\mathbf{H}}\left(G_1^{-1}(u_1) - \tilde{X}_1^{(i)}, \dots, G_d^{-1}(u_d) - \tilde{X}_d^{(i)}\right) \Big/ D(\mathbf{u}) \end{aligned} \quad (1.159)$$

Sans originalité aucune, il est souvent conseillé de prendre  $G_1 = \dots = G_d = \Phi$ . WEN et WU [2015a] mettent en garde contre le dénominateur  $D(\mathbf{u})$  qui peut devenir potentiellement infini lorsque  $u_j \rightarrow 0$  ou  $u_j \rightarrow 1$ . L'effet amplificateur de ce terme  $D(\mathbf{u})$  est parfois désiré, notamment lorsque la densité de copule  $c_{\mathbf{U}}$  n'est pas bornée. Cette situation est très courante, un exemple trivial étant la copule gaussienne bidimensionnelle  $c_{\mathcal{N}}(\cdot \mid \rho)$  qui admet des pics infinis au niveau des coins  $(0, 0)$  et  $(1, 1)$  de  $\mathcal{H}_2$  si  $\rho > 0$ . L'action de  $D(\mathbf{u})$  s'avère par contre être un problème lorsqu'elle engendre un comportement erratique dans les queues de distribution de la copule estimée  $\hat{c}_{\mathbf{H}}^{\text{TKE}}$ . Pour éviter ce dysfonctionnement, WEN et WU [2015b] suggèrent de remplacer les densités  $g_j$  par des fonctions  $\tilde{g}_j$  pour lesquelles la décroissance vers 0 des termes  $\tilde{g}_j(G_j^{-1}(u_j))$  est beaucoup plus lente. En vue de parvenir au même objectif, GEENENS [2014] a montré qu'il est particulièrement efficace d'estimer  $f_{\tilde{\mathbf{X}}}$  en utilisant la méthode de vraisemblance locale de LOADER et collab. [1996]. Selon GEENENS et collab. [2017], l'estimateur  $\hat{c}_{\mathbf{H}}^{\text{TKE}}$  qui en résulte n'affiche aucun biais asymptotique sur les bords, se révèle d'une grande précision même pour des copules hérissées, et garde la flexibilité inhérente à sa nature non-paramétrique.

### 1.5.3.3 Autres méthodes d'estimation non-paramétrique

Dans l'esprit de la méthode KDE, BROWN et CHEN [1999] ont introduit l'usage du noyau bêta. Sa manipulation et son interprétation diffèrent légèrement de celles des noyaux étudiés jusqu'ici. Ce noyau a été initialement conçu pour estimer une densité marginale de support  $[0, 1]$ . La densité  $\hat{f}_h$  s'écrit toujours comme la combinaison linéaire de  $N$  noyaux mais ces noyaux sont tous définis sur  $[0, 1]$  et chacun d'eux adopte une forme qui dépend de la position de  $Z^{(i)}$  au sein de l'intervalle. De ce fait, il n'y a pas de biais asymptotique sur les bords. CHARPENTIER et collab. [2007] montrent par ailleurs que la méthode se généralise facilement à l'estimation des copules bidimensionnelles. Outre les méthodes basées sur les noyaux, d'autres initiatives peuvent être rencontrées. GENEST et NEŠLEHOVÁ [2009] proposent de lisser la copule empirique à l'aide d'un filtre à ondelettes tandis que KAUERMANN et collab. [2013] déploient une base tensorisée de B-splines. La méthode du maximum d'entropie a été appliquée avec succès par DERENNES et collab. [2018] en dimension  $d = 2$  mais se révèle hautement imprécise lorsque  $d \geq 3$ . On souhaite également mentionner les avancées permises par SANCETTA et SATCHELL [2004] qui ont développé le modèle des copules de Bernstein et démontré son utilité en pratique.

### 1.5.4 Modèles de copules adaptés à la grande dimension

Dès que  $d \geq 4$ , les modèles d'apprentissage présentés jusqu'ici montrent certaines limites et il est imprudent de les utiliser sans précaution. Les copules elliptiques disposent d'une aptitude à reproduire la structure de corrélation mais elles imposent une modélisation strictement symétrique des queues de distribution. À l'inverse, les copules archimédiennes permettent une plus grande asymétrie dans la modélisation des dépendances extrémales mais elles reposent sur un trop petit nombre de paramètres pour être en mesure d'appréhender des motifs de dépendance polymorphes. En ce qui concerne les techniques non-paramétriques, elles ont été adaptées au contexte spécifique des copules afin d'éliminer les effets de bord mais les modifications apportées ne font évidemment pas disparaître le fléau de la dimension. Ainsi, un catalogue d'outils très

performants est disponible pour l'estimation des copules bidimensionnelles mais ces mêmes outils perdent en flexibilité et en précision lorsqu'on les déploie pour estimer des  $d$ -copules de plus grande dimension. Face à ce constat, différents modèles de copules ont été inventés en cherchant à capitaliser sur le savoir-faire acquis sur les copules bidimensionnelles. Dans la suite de cette section, plusieurs modèles sont introduits, en particulier le modèle des copules R-vignes qui est utilisé dans le cas d'application commun des Chapitres 4 et 5.

#### 1.5.4.1 Copules vignes

L'idée originelle remonte au papier de JOE [1996] qui cherche à construire des modèles de copules incorporant autant de paramètres que la copule gaussienne tout en ayant une structure polymorphe. Pour un vecteur  $\mathbf{X}$  à valeurs dans  $\mathbb{R}^3$ , il suggère d'écrire la fonction de répartition conjointe  $F_{\mathbf{X}} = F_{123}$  de la manière suivante :

$$\begin{aligned} F_{123}(\mathbf{x}) &= \int_{-\infty}^{x_3} \frac{\partial F_{\mathbf{X}}}{\partial x_3}(x_1, x_2, z_3) dz_3 = \int_{-\infty}^{x_3} F_{12|3}(x_1, x_2 | z_3) f_3(z_3) dz_3 \\ &= \int_{-\infty}^{x_3} C_{12|3}\left(F_{1|3}(x_1 | z_3), F_{2|3}(x_2 | z_3) \mid z_3\right) f_3(z_3) dz_3 \\ &= \int_{-\infty}^{x_3} C_{12|3}\left(\frac{\partial C_{13}}{\partial u_3}(F_1(x_1), F_3(z_3)), \frac{\partial C_{23}}{\partial u_3}(F_2(x_2), F_3(z_3)) \mid z_3\right) f_3(z_3) dz_3. \end{aligned} \quad (1.160)$$

Dans cette expression intégrale, trois sous-motifs de dépendance semblent être séparables. La présence des copules  $C_{13}$ ,  $C_{23}$  et  $C_{12|3}$  indique qu'il s'agit de la dépendance entre  $X_1$  et  $X_3$ , de celle entre  $X_2$  et  $X_3$  et de celle entre  $X_1$  et  $X_2$  conditionnellement à  $X_3$ . Si on affecte trois familles  $\mathcal{F}_A$ ,  $\mathcal{F}_B$  et  $\mathcal{F}_C$  (de paramètres scalaires respectifs  $\alpha$ ,  $\beta$  et  $\gamma$ ) à ces copules bidimensionnelles, la fonction  $F_{\mathbf{X}}$  peut alors être vue comme un modèle de distribution dont les lois marginales sont connues mais dont la structure de dépendance est pilotée par trois paramètres. C'est autant de paramètres que la copule gaussienne mais en ayant la possibilité d'introduire de la polymorphie. L'idée d'une décomposition en copules bidimensionnelles est reprise par BEDFORD et COOKE [2001, 2002] qui préfèrent prendre comme point de départ la densité conjointe  $f_{\mathbf{X}}$ . Tout démarre en appliquant le théorème de Bayes à  $f_{\mathbf{X}}$  :

$$\forall \mathbf{x} \in \mathbb{R}^d, \quad f_{\mathbf{X}}(\mathbf{x}) = \prod_{j=1}^d f_{X_j | X_{j-1} \dots X_1}(x_j | x_{j-1}, \dots, x_1). \quad (1.161)$$

Il faut ensuite réécrire les densités conditionnelles  $f_{X_j | X_{j-1} \dots X_1}$  pour faire intervenir des copules bidimensionnelles. Soit  $f_{X_i | \mathbf{X}_{\mathbf{a}}}$  l'une de ces densités avec  $\mathbf{a} := \{j_1, \dots, j_r\} \subseteq \{1, \dots, d\} \setminus \{i\}$ . On pioche  $j \in \mathbf{a}$  et on pose  $\mathbf{a}_j := \mathbf{a} \setminus \{j\}$ . Des manipulations sur les densités donnent alors :

$$f_{X_i | \mathbf{X}_{\mathbf{a}}} = \frac{f_{X_i | \mathbf{X}_{\mathbf{a}}}}{f_{\mathbf{X}_{\mathbf{a}}}} = \frac{f_{X_i | \mathbf{X}_{\mathbf{a}_j}} f_{\mathbf{X}_{\mathbf{a}_j}}}{f_{\mathbf{X}_{\mathbf{a}}}} = \frac{f_{X_i | \mathbf{X}_{\mathbf{a}_j}}}{f_{X_j | \mathbf{X}_{\mathbf{a}_j}}}, \quad (1.162)$$

et une version conditionnelle de l'égalité (1.38) conduit à :

$$f_{X_i | \mathbf{X}_{\mathbf{a}}}(x_i | \mathbf{x}_{\mathbf{a}}) = c_{ij | \mathbf{a}_j}\left(F_{X_i | \mathbf{X}_{\mathbf{a}_j}}(x_i | \mathbf{x}_{\mathbf{a}_j}), F_{X_j | \mathbf{X}_{\mathbf{a}_j}}(x_j | \mathbf{x}_{\mathbf{a}_j}) \mid \mathbf{x}_{\mathbf{a}_j}\right) f_{X_j | \mathbf{X}_{\mathbf{a}_j}}(x_j | \mathbf{x}_{\mathbf{a}_j}). \quad (1.163)$$

Pour aérer les équations, on note  $u_{i | \mathbf{a}_j} := F_{X_i | \mathbf{X}_{\mathbf{a}_j}}(x_i | \mathbf{x}_{\mathbf{a}_j})$  et  $u_{j | \mathbf{a}_j} := F_{X_j | \mathbf{X}_{\mathbf{a}_j}}(x_j | \mathbf{x}_{\mathbf{a}_j})$ . La fonction  $(u_{i | \mathbf{a}_j}, u_{j | \mathbf{a}_j}) \mapsto c_{ij | \mathbf{a}_j}(u_{i | \mathbf{a}_j}, u_{j | \mathbf{a}_j | \mathbf{x}_{\mathbf{a}_j}})$  est une copule bidimensionnelle dans l'espace probabilisé de mesure  $\mathbb{P}(\cdot | \mathbf{X}_{\mathbf{a}_j} = \mathbf{x}_{\mathbf{a}_j})$ . Cette copule représente la loi conjointe des variables  $F_{X_i | \mathbf{X}_{\mathbf{a}_j}}(X_i | \mathbf{x}_{\mathbf{a}_j})$  et  $F_{X_j | \mathbf{X}_{\mathbf{a}_j}}(X_j | \mathbf{x}_{\mathbf{a}_j})$  et non celle des variables  $U_i$  et  $U_j$  conditionnellement à  $\{\mathbf{X}_{\mathbf{a}_j} = \mathbf{x}_{\mathbf{a}_j}\}$ . De ce fait, l'expression de cette copule dépend théoriquement des valeurs  $\mathbf{x}_{\mathbf{a}_j}$  prises par les covariables, c'est-à-dire les variables conditionnantes  $\mathbf{X}_{\mathbf{a}_j}$ . Lorsqu'on applique la formule (1.163), la densité  $f_{X_i | \mathbf{X}_{\mathbf{a}}}$  à  $r$  variables conditionnantes est transformée en le produit de deux termes :

- une copule conditionnelle  $c_{ij|\mathbf{a}_j}$  à  $r - 1$  variables conditionnantes,
- une densité conditionnelle  $f_{X_j|\mathbf{X}_{a_j}}$  à  $r - 1$  variables conditionnantes.

Suite à l'apparition de la copule bidimensionnelle  $c_{ij|\mathbf{a}_j}$ ,  $X_j$  n'est plus une covariable et tous les ensembles de conditionnement en sont délestés. L'application récursive de cette astuce élimine toutes les densités conditionnelles pour ne plus obtenir que des densités marginales et des copules conditionnelles. En dimension  $d = 4$ , la densité conjointe  $f_{\mathbf{X}}$  peut alors être factorisée comme suit :

$$\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= f_{X_1}(x_1) f_{X_2}(x_2) f_{X_3}(x_3) f_{X_4}(x_4) \dots && \rightsquigarrow \text{Densités marginales} \\
&\times c_{12}(u_1, u_2) c_{13}(u_1, u_3) c_{14}(u_1, u_4) \dots && \rightsquigarrow \text{Copules non-conditionnelles} \\
&\times c_{23|1}(u_{2|1}, u_{3|1} | x_1) c_{24|1}(u_{2|1}, u_{4|1} | x_1) \dots && \rightsquigarrow \text{Copules conditionnelles d'ordre 1} \\
&\times c_{34|12}(u_{3|12}, u_{4|12} | \mathbf{x}_{12}) . && \rightsquigarrow \text{Copules conditionnelles d'ordre 2}
\end{aligned} \tag{1.164}$$

Comme il est possible d'isoler les densités marginales, la factorisation de  $f_{\mathbf{X}}$  met surtout en exergue celle de  $c_{\mathbf{U}}$ . Les trois dernières lignes de la factorisation (1.164) sont donc égales à  $c_{\mathbf{U}}(\mathbf{u})$  et doivent pouvoir être calculées à partir du seul point  $\mathbf{u} \in \mathcal{H}_d$ . On peut le vérifier après avoir remarqué que pour tout  $r$ -uplet  $\mathbf{a} := \{j_1, \dots, j_r\}$  et pour tout  $k \notin \mathbf{a}$  :

$$u_{k|\mathbf{a}} = F_{X_k|\mathbf{X}_{\mathbf{a}}}(x_k | \mathbf{x}_{\mathbf{a}}) = \frac{\partial^r C_{\mathbf{U}}}{\partial u_{j_1} \dots \partial u_{j_r}}(v_1, \dots, v_d) \quad \text{avec} \quad v_l := \begin{cases} u_l & \text{si } l \in k \cup \mathbf{a} \\ 1 & \text{sinon} \end{cases} . \tag{1.165}$$

Ainsi, on peut expliciter l'imbrication des calculs ayant conduit de  $\mathbf{u}$  à  $c_{\mathbf{U}}(\mathbf{u})$  :

$$\begin{aligned}
c_{\mathbf{U}}(\mathbf{u}) &= c_{12}(u_1, u_2) c_{13}(u_1, u_3) c_{14}(u_1, u_4) \dots \\
&\times c_{23|1} \left( \frac{\partial C_{12}}{\partial u_1}(u_1, u_2), \frac{\partial C_{13}}{\partial u_1}(u_1, u_3) \mid x_1 \right) c_{24|1} \left( \frac{\partial C_{12}}{\partial u_1}(u_1, u_2), \frac{\partial C_{14}}{\partial u_1}(u_1, u_4) \mid x_1 \right) \dots \\
&\times c_{34|12} \left( \frac{\partial^2 C_{123}}{\partial u_1 \partial u_2}(u_1, u_2), \frac{\partial^2 C_{124}}{\partial u_1 \partial u_2}(u_1, u_3) \mid \mathbf{x}_{12} \right) .
\end{aligned} \tag{1.166}$$

La factorisation (1.164) n'est pas unique. Un séquençage différent dans l'application récursive de la formule (1.163) aurait pu donner :

$$\begin{aligned}
c_{\mathbf{U}}(\mathbf{u}) &= c_{12}(u_1, u_2) c_{23}(u_2, u_3) c_{24}(u_2, u_4) \dots \\
&\times c_{13|2}(u_{1|2}, u_{3|2} | x_2) c_{14|2}(u_{1|2}, u_{4|2} | x_2) c_{34|12}(u_{3|12}, u_{4|12} | \mathbf{x}_{12}) ,
\end{aligned} \tag{1.167}$$

ou encore :

$$\begin{aligned}
c_{\mathbf{U}}(\mathbf{u}) &= c_{12}(u_1, u_2) c_{23}(u_2, u_3) c_{34}(u_3, u_4) \dots \\
&\times c_{13|2}(u_{1|2}, u_{3|2} | x_2) c_{24|3}(u_{2|3}, u_{4|3} | x_3) c_{14|23}(u_{1|23}, u_{4|23} | \mathbf{x}_{23}) .
\end{aligned} \tag{1.168}$$

En fait, de nombreuses factorisations de  $c_{\mathbf{U}}$  découlent de la factorisation initiale (1.161) de  $f_{\mathbf{X}}$ . Par ailleurs, les variables peuvent être permutées dans (1.161) et donnent naissance à autant de nouvelles versions équivalentes du théorème de Bayes. **MORALES-NÁPOLES [2010]** démontre que le nombre total de factorisations vaut :

$$N_{\text{facto}} = \frac{d!}{2} \times 2^{C_{d-2}^2} \quad \forall d \geq 4 . \tag{1.169}$$

**BEDFORD et COOKE [2001, 2002]** ont remarqué que toute factorisation peut être représentée à l'aide de graphes, et plus précisément à l'aide d'arbres. On rappelle qu'un arbre est un graphe connexe et acyclique. Cette schématisation sous forme d'arbres permet de mieux comprendre les règles assurant la cohérence de l'indexation des termes d'une factorisation. Avant d'aller plus loin, on a besoin d'introduire la notion de vigne régulière.

### Définition 1.5.7 (Vigne régulière)

Soit  $\mathbf{T}_{\text{vg}} = \{\mathbf{T}_1, \dots, \mathbf{T}_d\}$  une séquence d'arbres.

Chaque arbre  $\mathbf{T}_i$  est formé d'un ensemble  $\mathbf{N}_i$  de nœuds et d'un ensemble  $\mathbf{E}_i$  d'arêtes.

- Tout nœud  $\mathbf{n}$  de  $\mathbf{N}_i$  est un sous-ensemble de  $\{1, \dots, d\}$ .
- Toute arête  $\mathbf{e}$  de  $\mathbf{E}_i$  est une paire  $(\mathbf{n}_1, \mathbf{n}_2)$  de nœuds pris dans  $\mathbf{N}_i$ .

On dit que  $\mathbf{T}_{\text{vg}}$  est une vigne régulière (ou une R-vigne) si et seulement si :

(VR1)  $\mathbf{T}_1$  est un arbre couvrant avec  $\mathbf{N}_1 := \{1, \dots, d\}$ .

(VR2) Pour tout  $2 \leq j \leq d$ ,  $\mathbf{N}_j$  est directement hérité de  $\mathbf{E}_{j-1}$  :

$$\mathbf{N}_j := \{\mathbf{n}_1 \cup \mathbf{n}_2 : (\mathbf{n}_1, \mathbf{n}_2) \in \mathbf{E}_{j-1}\}. \quad (1.170)$$

(VR3) Pour tout  $2 \leq j \leq d$ ,  $\text{Card}(\mathbf{E}_j) = d - j$  et  $\mathbf{E}_j$  vérifie la "condition de proximité" :

$$\mathbf{e} = (\mathbf{n}_1, \mathbf{n}_2) \in \mathbf{E}_j \implies \exists \mathbf{n}_a, \mathbf{n}_b, \mathbf{n}_c \in \mathbf{N}_{j-1} \text{ tels que } \begin{cases} \mathbf{n}_1 = (\mathbf{n}_a, \mathbf{n}_b) \\ \mathbf{n}_2 = (\mathbf{n}_a, \mathbf{n}_c) \end{cases}. \quad (1.171)$$

Concernant le point (VR1), on précise qu'un arbre est dit couvrant s'il connecte tous les sommets entre eux. Sachant qu'un arbre est un graphe sans cycle, on a  $\text{Card}(\mathbf{E}_1) = d - 1$ . Si (VR2) détermine complètement  $\mathbf{N}_j$  à partir de  $\mathbf{E}_{j-1}$ , il n'en va pas de même pour (VR3) qui laisse la possibilité de construire plusieurs versions de  $\mathbf{E}_j$  à partir de  $\mathbf{N}_j$ . La condition de proximité (VR3) garantit la compatibilité entre la séquence  $\mathbf{T}_{\text{vg}}$  et une factorisation de  $c_{\mathbf{U}}$ . Pour  $d = 4$ , un exemple de R-vigne est donné par :

$$\begin{aligned} \mathbf{T}_1 : \quad \mathbf{N}_1 &= \{\{1\}; \{2\}; \{3\}; \{4\}\} & \mathbf{E}_1 &= \{(\{1\}, \{2\}); (\{1\}, \{3\}); (\{1\}, \{4\})\} \\ \mathbf{T}_2 : \quad \mathbf{N}_2 &= \{\{1, 2\}; \{1, 3\}; \{1, 4\}\} & \mathbf{E}_2 &= \{(\{1, 2\}, \{1, 3\}); (\{1, 2\}, \{1, 4\})\} \\ \mathbf{T}_3 : \quad \mathbf{N}_3 &= \{\{1, 2, 3\}; \{1, 2, 4\}\} & \mathbf{E}_3 &= \{(\{1, 2, 3\}, \{1, 2, 4\})\} \\ \mathbf{T}_4 : \quad \mathbf{N}_4 &= \{\{1, 2, 3, 4\}\} & \mathbf{E}_4 &= \emptyset \end{aligned} \quad (1.172)$$

La dernière ligne est souvent omise car on a toujours  $\mathbf{T}_d = (\mathbf{N}_d, \mathbf{E}_d)$  avec  $\mathbf{N}_d = \{1, \dots, d\}$  et  $\mathbf{E}_d = \emptyset$ . Au vu de l'exemple qui précède, on conjecture que :

- $\text{Card}(\mathbf{E}_j) = \text{Card}(\mathbf{N}_j) - 1 = d - j$ ,
  - $\forall 2 \leq j \leq d, \forall \mathbf{n} \in \mathbf{N}_j, \text{Card}(\mathbf{n}) = j$  et  $\forall \mathbf{e} = (\mathbf{n}_1, \mathbf{n}_2) \in \mathbf{E}_j, \text{Card}(\mathbf{n}_1 \cap \mathbf{n}_2) = j - 1$ .
- (1.173)

Les preuves sont élémentaires par récurrence. Le lien entre une factorisation de  $c_{\mathbf{U}}$  et une R-vigne se tisse au niveau de la séquence  $\{\mathbf{E}_1, \dots, \mathbf{E}_{d-1}\}$ . D'une part, le nombre total d'arêtes est égal au nombre de copules bidimensionnelles. D'autre part, compte tenu des résultats (1.173), on est en droit de définir :

$$\forall 1 \leq j \leq d - 1, \quad \forall \mathbf{e} = (\mathbf{n}_1, \mathbf{n}_2) \in \mathbf{E}_j, \quad \mathbf{D}(\mathbf{e}) := \mathbf{n}_1 \cap \mathbf{n}_2 \quad \text{et} \quad \begin{cases} k(\mathbf{e}) := \mathbf{n}_1 \setminus \mathbf{D}(\mathbf{e}) \\ l(\mathbf{e}) := \mathbf{n}_2 \setminus \mathbf{D}(\mathbf{e}) \end{cases}. \quad (1.174)$$

Ainsi, pour toute arête  $\mathbf{e} = (\mathbf{n}_1, \mathbf{n}_2)$ ,  $\mathbf{n}_1 \cup \mathbf{n}_2 = \{k(\mathbf{e}), l(\mathbf{e}), \mathbf{D}(\mathbf{e})\}$ . Chaque arête  $\mathbf{e}$  doit être mise en relation avec la copule bidimensionnelle entre  $X_{k(\mathbf{e})}$  et  $X_{l(\mathbf{e})}$  conditionnellement à  $\{\mathbf{X}_{\mathbf{D}(\mathbf{e})} = \mathbf{x}_{\mathbf{D}(\mathbf{e})}\}$ . Pour la R-vigne (1.172), on obtient les copules  $c_{12}$ ,  $c_{13}$ ,  $c_{14}$ ,  $c_{23|1}$ ,  $c_{24|1}$  et  $c_{34|12}$ . S'il s'agit exactement des copules de la factorisation (1.164), c'est parce que la R-vigne (1.172) a été exhibée dans ce but.

**BEDFORD et COOKE [2001, 2002]** ont prouvé que l'indexation des copules conditionnelles d'une factorisation est équivalente à la séquence d'arbres d'une R-vigne. Toute factorisation peut donc être représentée par une R-vigne. Réciproquement, si une séquence d'arbres  $\mathbf{T}_{\text{vg}}$  vérifie les critères énoncés à la Définition 1.5.7, alors elle correspond à une factorisation existante. La Figure 1.2a

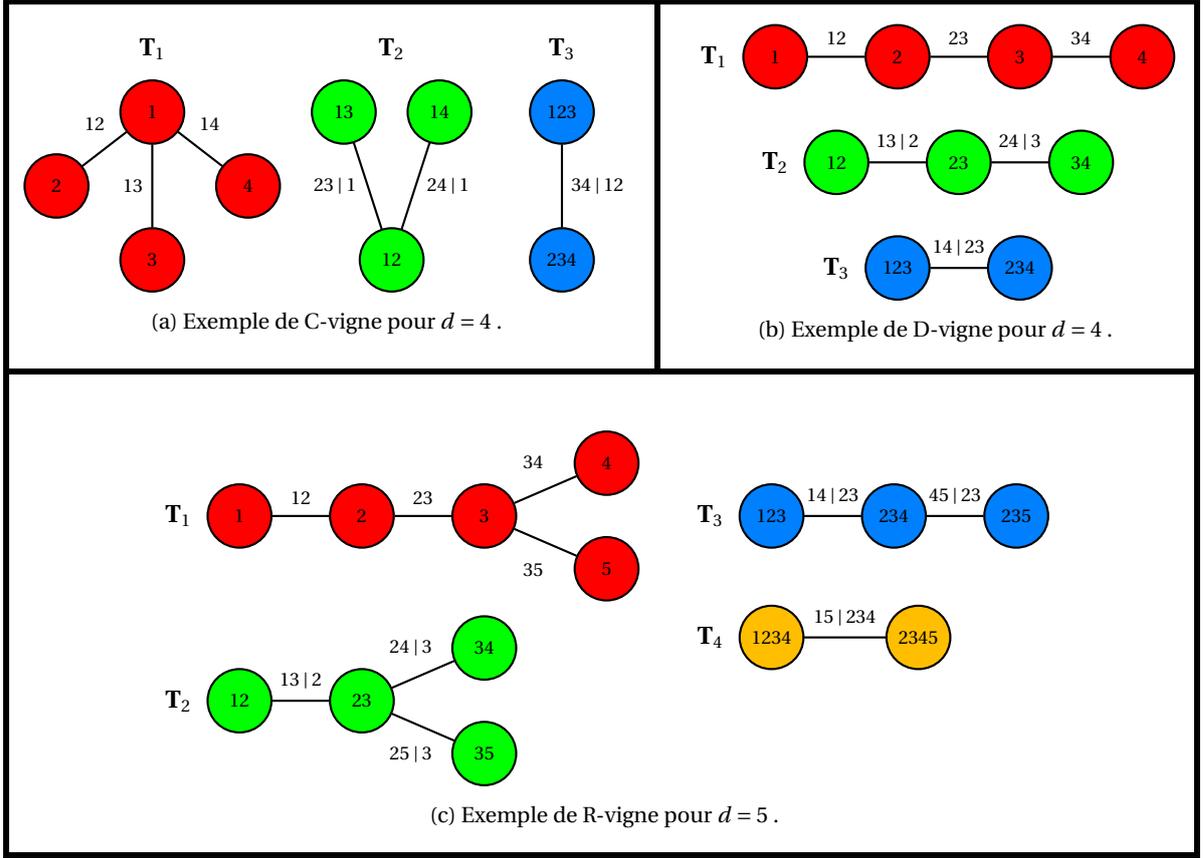


FIGURE 1.2 – Trois exemples de R-vignes : une C-vigne, une D-vigne et une R-vigne quelconque.

montre la R-vigne associée à la factorisation (1.164). On y remarque que les arbres affichent une configuration particulière où l'un des sommets est commun à toutes les arêtes. Il s'agit du sommet  $\{1\}$  dans  $T_1$  puis du sommet  $\{1, 2\}$  dans  $T_2$  et enfin du sommet  $\{1, 2, 3\}$  de  $T_3$ . Ce type de R-vigne, avec des arbres en forme d'étoiles, est appelée une vigne canonique (ou C-vigne). On souligne que la factorisation (1.167) correspond aussi à une C-vigne. Pour le voir, il suffit d'inverser les rôles des indices 1 et 2 pour retrouver (1.164). La R-vigne associée à la factorisation (1.168) est en revanche d'une nature différente. On peut se reporter à la Figure 1.2b et y voir que chaque arbre adopte une configuration en forme de chaîne. Une R-vigne respectant cette disposition est appelée une vigne étirable (ou D-vigne).

Si  $d = 3$ , toute R-vigne est à la fois une C-vigne ou une D-vigne. Si  $d = 4$ , une R-vigne est soit une C-vigne soit une D-vigne. Pour  $d \geq 5$ , il existe des R-vignes qui ne sont ni des C-vignes ni des D-vignes. On en veut pour preuve l'exemple proposé à la Figure 1.2c où l'arbre  $T_1$  présente une forme hybride. D'une manière générale, quelle que soit la factorisation retenue, le lien avec une R-vigne permet de réécrire la densité de copule de la manière suivante :

$$\forall \mathbf{u} \in \mathcal{H}_d, \quad c_{\text{vg}}(\mathbf{u} | \mathbf{T}_{\text{vg}}) = \prod_{j=1}^{d-1} \prod_{e \in E_j} c_{k(e)l(e)|D(e)}(u_{j(e)|D(e)}, u_{k(e)|D(e)} | \mathbf{x}_{D(e)}) . \quad (1.175)$$

Si toutes les copules bidimensionnelles sont décrites par des familles paramétriques, on aboutit à un modèle paramétrique d'un genre nouveau : les copules R-vignes. Il faut toutefois être capable de paramétrer des copules conditionnelles. AAS et collab. [2009] prennent le parti de supposer qu'une copule conditionnelle ne dépend des covariables  $\mathbf{X}_{D(e)}$  qu'au travers des valeurs  $u_{k(e)|D(e)}$  et  $u_{l(e)|D(e)}$  prises par les fonctions de répartition conditionnelles passées en argument. Cela revient à dire que la fonction  $(u_1, u_2, \mathbf{v}) \mapsto c_{k(e)l(e)|D(e)}(u_1, u_2 | \mathbf{v})$  est constante par rapport à  $\mathbf{v}$ . On a alors :

$$\forall \mathbf{u} \in \mathcal{H}_d, \quad c_{\text{vg}}(\mathbf{u} | \mathbf{T}_{\text{vg}}) = \prod_{j=1}^{d-1} \prod_{e \in E_j} c_{k(e)l(e)|D(e)}(u_{j(e)|D(e)}, u_{k(e)|D(e)}) , \quad (1.176)$$

et ce modèle porte le nom de copule R-vigne simplifiée. Dans la suite, par pure commodité, le terme R-vigne désigne toujours une R-vigne simplifiée. Pour parler du modèle le plus général, on choisit l'appellation de copule R-vigne non simplifiée. La validité de l'hypothèse "simplificatrice" suscite un vif débat chez les spécialistes du domaine. [HAFF et collab. \[2010\]](#) expliquent que les copules R-vignes offrent une flexibilité suffisante pour approcher tout type de copule sous-jacente. [STOEBER et collab. \[2013\]](#) tempèrent néanmoins en soulignant que seul un très petit nombre de copules elliptiques et archimédiennes vérifient cette hypothèse. De leur côté, [ACAR et collab. \[2012\]](#) puis [SPANHEL et KURZ \[2015\]](#) mettent en garde contre les erreurs que peut entraîner l'application intempestive des copules R-vignes. Pour chaque arête  $e$  présente dans  $\mathbf{T}_{\text{vg}}$ , on note  $\mathcal{F}_e$  une famille de copules bidimensionnelles et  $\theta_e$  ses différents paramètres. Le modèle (1.176) se transforme en :

$$\forall \mathbf{u} \in \mathcal{H}_d, \quad c_{\text{vg}}(\mathbf{u} \mid \mathbf{T}_{\text{vg}}, \mathbf{F}_{\text{vg}}, \boldsymbol{\theta}_{\text{vg}}) = \prod_{j=1}^{d-1} \prod_{e \in \mathbf{E}_j} c_{k(e)l(e)|D(e)} \left( u_{j(e)|D(e)}, u_{k(e)|D(e)} \mid \mathcal{F}_e, \theta_e \right) \quad (1.177)$$

avec  $\mathbf{F}_j := \{\mathcal{F}_e : e \in \mathbf{E}_j\}$ ,  $\boldsymbol{\theta}_j := \{\theta_e : e \in \mathbf{E}_j\}$ ,  $\mathbf{F}_{\text{vg}} := \bigcup_{j=1}^{d-1} \mathbf{F}_j$  et  $\boldsymbol{\theta}_{\text{vg}} := \bigcup_{j=1}^{d-1} \boldsymbol{\theta}_j$ .

Ainsi, l'apprentissage d'une copule R-vigne passe par la sélection d'une séquence d'arbres, le choix d'un assortiment de familles et l'estimation de tous les paramètres associés. [AAS et collab. \[2009\]](#) ont développé un premier algorithme d'apprentissage dans le cas des C-vignes et des D-vignes. La procédure est séquentielle au sens où il faut commencer par estimer les copules de l'arbre  $\mathbf{T}_1$ , puis celles de l'arbre  $\mathbf{T}_2$  et ainsi de suite jusqu'à l'arbre  $\mathbf{T}_{d-1}$ . Cet algorithme de descente dans les arbres  $\mathbf{T}_1, \dots, \mathbf{T}_{d-1}$  a ensuite été complété par [DISSMANN et collab. \[2013\]](#) qui a également réussi à l'étendre aux R-vignes. Dans la suite de cette section, on détaille la procédure inférentielle qui permet de construire une estimation  $\hat{c}_{\text{vg}}(\cdot \mid \hat{\mathbf{T}}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}}, \hat{\boldsymbol{\theta}}_{\text{vg}})$  de  $c_{\mathbf{U}}$ .

### Étape 1 Apprentissage des copules bidimensionnelles de l'arbre $\mathbf{T}_1$

Les copules de  $\mathbf{T}_1$  sont non-conditionnelles. On les note  $c_{ij}$  avec  $1 \leq i < j \leq d$ .

#### Étape 1.1 Choix de l'ensemble $\mathbf{E}_1$ des nœuds de $\mathbf{T}_1$

Pour respecter le critère (VR1), il faut que  $\mathbf{T}_1$  soit un arbre couvrant. Dans [AAS et collab. \[2009\]](#), la structure de l'arbre  $\mathbf{T}_1$  est choisie arbitrairement. L'intuition de [DISSMANN et collab. \[2013\]](#) est de construire  $\mathbf{E}_1$  pour que les arrêtes  $e = (\{i\}, \{j\})$  correspondent à des paires présentant un fort niveau de concordance ou de discordance. Il faut donc choisir une mesure de dépendance  $\delta$  et calculer à partir de  $\mathbf{U}_{\text{obs}}$  les mesures empiriques des différentes paires  $(U_i, U_j)$ . Par défaut,  $\delta$  est le tau de Kendall mais on peut aussi prendre le rho de Spearman ou des coefficients de dépendance de queue. Au moment de constituer  $\mathbf{E}_1$ , chaque paire se voit affecter un poids égal à  $|\hat{\tau}_{ij}|$  et il faut trouver l'arbre couvrant dont la somme des poids des arêtes est maximale :

$$\mathbf{E}_1^* = \underset{\mathbf{E}_1}{\text{argmax}} \sum_{e=(\{i\}, \{j\})} |\hat{\tau}_{ij}| \quad \text{sous la contrainte : } \mathbf{E}_1 \text{ est un arbre couvrant.} \quad (1.178)$$

Dans le vocabulaire de l'optimisation discrète,  $\mathbf{E}_1^*$  est dit être un arbre couvrant de poids maximal. Pour trouver  $\mathbf{E}_1^*$ , on peut utiliser l'un des algorithmes présentés dans [GABOW et collab. \[1986\]](#). Entre autres, on peut mentionner l'algorithme de Kruskal et celui de Prim. Si on veut que  $\mathbf{T}_{\text{vg}}$  soit une C-vigne, la résolution est immédiate. En revanche, si on préfère que  $\mathbf{T}_{\text{vg}}$  soit une D-vigne, le problème (1.178) est équivalent à celui du voyageur de commerce.

#### Étape 1.2 Sélection des familles $\mathbf{F}_1$

On suppose qu'on dispose d'un catalogue de  $r$  familles paramétriques  $\mathcal{F}_1, \dots, \mathcal{F}_r$ . Dans le package `RVineCopula`<sup>2</sup>, le catalogue inclut par exemple  $r = 36$  familles paramétriques. Pour chaque arête  $e = (i, j)$  de  $\mathbf{E}_1$ , la famille  $\mathcal{F}_{ij}^*$  offrant la meilleure adéquation aux données est sélectionnée en minimisant un critère d'information de type AIC ou BIC. [CZADO et collab. \[2013\]](#) conseillent de

2. <https://cran.r-project.org/package=VineCopula>

valider le choix final en s'appuyant sur l'un des tests proposés par [GENEST et collab. \[2009\]](#). Une autre façon de construire  $\mathbf{E}_1^*$  peut être de maximiser l'adéquation aux données :

$$\begin{aligned} \mathbf{E}_1^* &= \operatorname{argmin}_{\mathbf{E}_1} \sum_{e=\{i,j\}} \alpha_{ij}^* \quad \text{sous la contrainte : } \mathbf{E}_1 \text{ est un arbre couvrant ,} \\ \text{avec } \alpha_{ij}^* &= \min_{1 \leq k \leq r} \operatorname{AIC}(\mathcal{F}_k \mid \mathbf{U}_{\text{obs}}^{[i,j]}) \quad \text{où } \mathbf{U}_{\text{obs}}^{[i,j]} := \left[ \left( \mathbf{U}_i^{(n)}, \mathbf{U}_j^{(n)} \right) \right]_{1 \leq n \leq N} . \end{aligned} \quad (1.179)$$

Cette approche est beaucoup plus coûteuse car le calcul de chacun des  $C_d^2$  poids  $\alpha_{ij}^*$  impose de calibrer toutes les familles du catalogue aux observations de la paire  $(U_i, U_j)$ .

### Étape 1.3 Estimation des paramètres $\theta_1$

Pour chaque copule bidimensionnelle  $c_{ij}(\cdot \mid \mathcal{F}_{ij}^*, \theta_{ij})$ , la méthode du maximum de vraisemblance ou une stratégie d'inversion du tau de Kendall permet de calculer  $\hat{\theta}_{ij}$ . Dans le cas du maximum de vraisemblance,  $\hat{\theta}_{ij}$  est déjà calculé au cours du processus de sélection de  $\mathcal{F}_{ij}^*$  avec AIC.

### Étape 2 Apprentissage des copules bidimensionnelles de l'arbre $\mathbf{T}_2$

Les copules de  $\mathbf{T}_2$  sont conditionnées par une seule covariable. On les note  $c_{i|k}$  avec  $k \notin \{i, j\}$ .

#### Étape 2.0 Calcul des observations conditionnelles d'ordre 1

Pour estimer une copule de type  $c_{i|k}$ , on a besoin d'observations des vecteurs  $U_{i|k}$  et  $U_{j|k}$ . On rappelle que  $U_{i|k}$  peut s'exprimer de la manière suivante :

$$U_{i|k} = F_{X_i|X_k}(X_i \mid X_k) = \frac{\partial C_{ik}}{\partial u_k}(U_i, U_k) . \quad (1.180)$$

De par la condition de proximité, s'il y a besoin d'estimer la copule  $c_{i|k}$ , cela signifie qu'il existe une arête  $(i, k)$  dans  $\mathbf{E}_1^*$  et que la copule  $c_{ik}$  appartient à la factorisation. Ainsi, les observations conditionnelles sont uniquement dépendantes des copules bidimensionnelles de l'arbre  $\mathbf{T}_1$ . Pour toute arête  $(i, k)$  de  $\mathbf{E}_1^*$ , on définit :

- la h-fonction à gauche :  $h_{ik}^{(G)}(u_i, u_k) = \frac{\partial C_{ik}}{\partial u_i}(u_i, u_k) = \int_{-\infty}^{u_k} c_{ik}(u_i, v_k) dv_k ,$
- la h-fonction à droite :  $h_{ik}^{(D)}(u_i, u_k) = \frac{\partial C_{ik}}{\partial u_k}(u_i, u_k) = \int_{-\infty}^{u_i} c_{ik}(v_i, u_k) dv_i .$

Pour chaque copule bidimensionnelle  $\hat{c}_{ij}(\cdot \mid \mathcal{F}_{ij}^*, \hat{\theta}_{ij})$  estimée à l'arbre  $\mathbf{T}_1$ , on fabrique deux types d'observations conditionnelles :

$$\forall 1 \leq n \leq N, \quad \hat{U}_{j|i}^{(n)} = \hat{h}_{ij}^{(G)}\left(\mathbf{U}_i^{(n)}, \mathbf{U}_j^{(n)} \mid \mathcal{F}_{ij}^*, \hat{\theta}_{ij}\right) \quad \text{et} \quad \hat{U}_{i|j}^{(n)} = \hat{h}_{ij}^{(D)}\left(\mathbf{U}_i^{(n)}, \mathbf{U}_j^{(n)} \mid \mathcal{F}_{ij}^*, \hat{\theta}_{ij}\right) . \quad (1.182)$$

Une grande partie de ces observations conditionnelles est destinée à l'apprentissage des copules de l'arbre  $\mathbf{T}_2$ . Comme leur fabrication dépend des informations  $(\hat{\mathbf{F}}_1, \hat{\theta}_1)$  issues de l'apprentissage de  $\mathbf{T}_1$ , elles ne suivent pas exactement la loi théorique des observations conditionnelles. Si des erreurs sont commises au moment de l'estimation des copules de  $\mathbf{T}_1$ , celles-ci se répercutent sur l'apprentissage des copules de  $\mathbf{T}_2$ .

### Étape 2.1 Choix de l'ensemble $\mathbf{E}_2$ des nœuds de $\mathbf{T}_2$

L'application de la condition de proximité (VR3) limite les possibilités pour  $\mathbf{E}_2$  car seules des arêtes de type  $e = (\{i, k\}, \{j, k\})$  peuvent être tracées. Deux cas particuliers sont notables :

- Si  $\mathbf{T}_1$  arbore une disposition en forme de chaîne,  $\mathbf{T}_2$  est entièrement déterminée.
- Si  $\mathbf{T}_1$  arbore une disposition en forme d'étoile, les sommets de  $\mathbf{N}_2$  ont en commun l'indice  $k$  du sommet central de  $\mathbf{T}_1$  et toutes les arêtes sont de type  $e = (\{i, k\}, \{j, k\})$ . Le problème est alors exactement celui d'un arbre couvrant de poids maximal.

Lorsque que  $\mathbf{T}_2$  n'est pas complètement spécifié par l'application de la condition de proximité, on procède comme pour  $\mathbf{T}_1$ . La mesure de dépendance utilisée par le calcul des poids est le tau de Kendall conditionnel  $\hat{\tau}_{i|j|k}$ , c'est-à-dire le tau de Kendall des variables  $\hat{U}_{i|k}$  et  $\hat{U}_{j|k}$ . Le problème d'optimisation est une version légèrement surcontrainte du problème (1.178) :

$$\mathbf{E}_2^* := \operatorname{argmax}_{\mathbf{E}_2} \sum_{e=(i,k),(j,k)} |\hat{\tau}_{i|j|k}| \quad \text{sous la contrainte : } \mathbf{E}_2 \text{ est un arbre couvrant .} \quad (1.183)$$

**Étape 2.2** Sélection des familles  $\mathbf{F}_2 \Rightarrow$  Voir [Étape 1.2](#)

**Étape 2.3** Estimation des paramètres  $\theta_2 \Rightarrow$  Voir [Étape 1.3](#)

**Étape 3** Apprentissage des copules bidimensionnelles dans les autres arbres

Pour estimer les copules des arbres inférieurs, on répète la procédure inférentielle menée pour estimer les copules de l'arbre  $\mathbf{T}_2$ .

**Étape 4** Estimation globale des paramètres  $\theta_{\text{vg}}$

À l'issue de la descente dans les arbres, l'information fournie par  $(\hat{\mathbf{T}}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}})$  permet de réduire le modèle (1.177) à une famille fixée de paramètres  $\theta_{\text{vg}}$ . On cherche alors à maximiser la log-vraisemblance globale de cette famille pour  $\mathbf{U}_{\text{obs}}$  :

$$\hat{\theta}_{\text{vg}} = \operatorname{argmax}_{\theta_{\text{vg}}} \ell_{\mathbf{N}}(\theta_{\text{vg}}) = \sum_{i=1}^{\mathbf{N}} \log \left( c_{\text{vg}}(\mathbf{U}^{(i)} \mid \hat{\mathbf{T}}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}}, \theta_{\text{vg}}) \right). \quad (1.184)$$

Pour ce recalage final des paramètres, l'algorithme d'optimisation peut être initialisé en prenant les résultats de la descente séquentielle dans les arbres. Cette phase termine l'apprentissage de  $c_{\mathbf{U}}$  à l'aide du modèle des copules R-vignes. [DISSMANN et collab. \[2013\]](#) ont mis au point des routines numériques permettant de :

- (a) Calculer la valeur  $\hat{c}_{\text{vg}}(\mathbf{u}_0 \mid \hat{\mathbf{T}}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}}, \hat{\theta}_{\text{vg}})$  en tout point  $\mathbf{u}_0$  de  $\mathcal{H}_d$ .
- (b) Simuler des échantillons suivant  $\hat{c}_{\text{vg}}(\cdot \mid \hat{\mathbf{T}}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}}, \hat{\theta}_{\text{vg}})$ .

Par contre, il n'est pas possible de reconstituer l'expression exacte de la fonction de répartition associée à la densité de copule  $\hat{c}_{\text{vg}}(\cdot \mid \hat{\mathbf{T}}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}}, \hat{\theta}_{\text{vg}})$ . Par ailleurs, on signale qu'il existe plusieurs variantes de l'algorithme d'apprentissage :

- [MIN et CZADO \[2010\]](#) adoptent une approche bayésienne dans un cas très particulier où la factorisation est imposée et où toutes les paires sont modélisées par des copules de Student. Un algorithme bayésien plus ambitieux, incluant la sélection automatique de  $\mathbf{T}_{\text{vg}}$  et de  $\mathbf{F}_{\text{vg}}$ , est ensuite développé par [GRUBER et collab. \[2015\]](#).
- Pour que le modèle reste relativement parcimonieux, l'idée d'une troncature est examinée par [KUROWICKA \[2010\]](#) puis [BRECHMANN et collab. \[2012\]](#). Il s'agit d'arrêter la descente à l'arbre  $\mathbf{T}_r$  avec  $r < d - 1$  en partant du principe que les erreurs qui s'accumulent lors du calcul des observations conditionnelles rendent peu fiables les copules estimées dans les arbres inférieurs.
- Au lieu d'estimer les copules bidimensionnelles dans un cadre paramétrique, [NAGLER et CZADO \[2016\]](#) proposent d'implémenter des techniques non-paramétriques afin d'échapper au fléau de la dimension. [NAGLER et collab. \[2017\]](#) comparent les résultats du couplage des R-vignes avec différentes approches non-paramétriques d'estimation des copules bidimensionnelles. La méthode TKE, dans sa version améliorée par [GEENENS et collab. \[2017\]](#), semble faire office de référence en la matière.

### 1.5.4.2 Modèles factoriels pour les copules

Les modèles factoriels reposent sur l'idée selon laquelle la structure de dépendance reliant les variables  $X_1, \dots, X_d$  peut être expliquée en grande partie par l'action de variables latentes  $Z_1, \dots, Z_r$ . Celles-ci sont inobservables mais exercent un fort impact sur les valeurs prises par les variables observables du problème. Cette modélisation peut être pertinente en finance de marché où une forte corrélation entre les prix des titres boursiers s'explique souvent par l'état de la conjoncture. Si on a accès à une collection d'indicateurs décrivant de manière exhaustive l'état du marché, l'aléa restant sur le prix d'un titre boursier ne dépend alors plus des autres prix. On renvoie vers l'ouvrage de [JOHNSON et collab. \[2002\]](#) pour de plus amples explications à ce sujet.

Cette démarche peut être directement appliquée aux variables  $U_j = F_{X_j}(X_j)$  du vecteur  $\mathbf{U}$ . On introduit un vecteur  $\mathbf{V}$  de variables qui sont supposées uniformes et indépendantes<sup>3</sup>. L'hypothèse centrale du modèle factoriel est l'indépendance des variables  $U_j$  conditionnellement à  $\mathbf{V}$ . Après avoir incorporé les variables latentes avec le théorème de Bayes, la densité  $c_{\mathbf{U}}$  s'écrit :

$$c_{\mathbf{U}}(\mathbf{u}) = \int_{\mathcal{H}_r} c_{\mathbf{U}|\mathbf{V}}(\mathbf{u} | \mathbf{v}) \pi_r(\mathbf{v}) d\mathbf{v} = \int_{\mathcal{H}_r} \prod_{j=1}^d c_{U_j|\mathbf{V}}(u_j | \mathbf{v}) \pi_r(\mathbf{v}) d\mathbf{v} = \int_{\mathcal{H}_r} \prod_{j=1}^d c_{U_j, \mathbf{V}}(u_j, \mathbf{v}) d\mathbf{v}. \quad (1.185)$$

Dans les premières avancées de [HULL et WHITE \[2004\]](#) puis de [LAURENT et GREGORY \[2005\]](#), une structure de corrélation linéaire est jugée suffisante pour relier les vecteurs  $\mathbf{U}$  et  $\mathbf{V}$ . Les copules  $c_{\mathbf{U}, \mathbf{V}}$  sont estimées en conséquence au sein de la famille gaussienne. Les distributions de type CFM (*Copula Factor Models*) ont ensuite évolué vers plus d'asymétrie avec la prise en compte de motifs de dépendance plus subtils. En pratique, le nombre de variables latentes est souvent limité à  $r \in \{1, 2\}$ . Pour  $r = 1$ , on obtient une copule CFM du premier ordre :

$$c_{\text{CFM}}(\mathbf{u} | \mathbf{F}, \boldsymbol{\theta}) = \int_0^1 \prod_{j=1}^d c_{U_j, V_1}(u_j, v_1 | \mathcal{F}_j, \boldsymbol{\theta}_j) dv_1 \quad \text{avec} \quad \mathbf{F} = \bigcup_{j=1}^d \mathcal{F}_j \quad \text{et} \quad \boldsymbol{\theta} = \bigcup_{j=1}^d \boldsymbol{\theta}_j. \quad (1.186)$$

Pour un ensemble de familles paramétriques et un N-échantillon  $\mathbf{U}_{\text{obs}}$ , la log-vraisemblance vaut :

$$\ell_N(\boldsymbol{\theta}) = \sum_{i=1}^N \log \left( \int_0^1 \prod_{j=1}^d c_{U_j, V_1}(U_j^{(i)}, v_1 | \mathcal{F}_j, \boldsymbol{\theta}_j) dv_1 \right). \quad (1.187)$$

[KRUPSKII et JOE \[2013\]](#) indiquent que les N tâches d'intégration sur l'intervalle  $[0, 1]$  ne constituent pas un obstacle majeur et peuvent être exécutées facilement avec des méthodes de quadrature comme celles de Gauss-Legendre ou de Gauss-Hermite. L'accès à la log-vraisemblance permet alors de calculer le score d'un critère d'information (AIC/BIC) et ainsi d'automatiser la sélection des familles paramétriques  $\mathcal{F}_j^*$  offrant la meilleure adéquation aux données. La démarche est la même pour  $r = 2$  avec une intégration sur  $\mathcal{H}_2$ . Pour assurer un maximum de flexibilité dans l'apprentissage des copules tridimensionnelles  $c_{U_j, V_1, V_2}$ , [KRUPSKII et JOE \[2013\]](#) proposent d'utiliser des copules R-vignes. À ce propos, comme  $V_1$  et  $V_2$  sont indépendantes, chaque R-vigne se limite au produit des copules  $c_{U_j, V_1}$  et  $c_{U_j, V_2|V_1}$  :

$$c_{\text{vg}}(u_j, \mathbf{v} | \mathcal{F}_j, \boldsymbol{\theta}_j) = c_{U_j, V_1}(u_j, v_1 | \mathcal{F}_{1j}, \boldsymbol{\theta}_{1j}) c_{U_j, V_2|V_1}(h^{(D)}(u_j, v_1 | \mathcal{F}_{1j}, \boldsymbol{\theta}_{1j}), v_2 | \mathcal{F}_{2j}, \boldsymbol{\theta}_{2j}). \quad (1.188)$$

On retrouve alors l'hypothèse "simplificatrice" évoquée à la Section 1.5.4.1 lors du passage de l'équation (1.175) à l'équation (1.176). En effet, la copule  $c_{U_j, V_2|V_1}$  ne dépend de la variable conditionnante  $V_1$  qu'au travers de son premier argument. En réinjectant dans l'équation (1.185), on obtient l'expression de la densité d'une copule CFM du second ordre :

$$c_{\text{CFM}}(\mathbf{u} | \mathbf{F}, \boldsymbol{\theta}) = \int_0^1 \int_0^1 \prod_{j=1}^d c_{\text{vg}}(u_j, \mathbf{v} | \mathcal{F}_j, \boldsymbol{\theta}_j) d\mathbf{v} \quad \text{avec} \quad \mathbf{F} = \bigcup_{i=1}^2 \bigcup_{j=1}^d \mathcal{F}_{ij} \quad \text{et} \quad \boldsymbol{\theta} = \bigcup_{i=1}^2 \bigcup_{j=1}^d \boldsymbol{\theta}_{ij}. \quad (1.189)$$

3. Il n'y pas de perte de généralité. En effet, les variables  $V_1, \dots, V_r$  n'ont aucune signification d'un point de vue physique. Pour des variables latentes de loi  $\mu_{\mathbf{V}}$  quelconque, on peut toujours appliquer la transformée de Rosenblatt pour fabriquer d'autres variables latentes  $W_1, \dots, W_r$  uniformes et indépendantes.

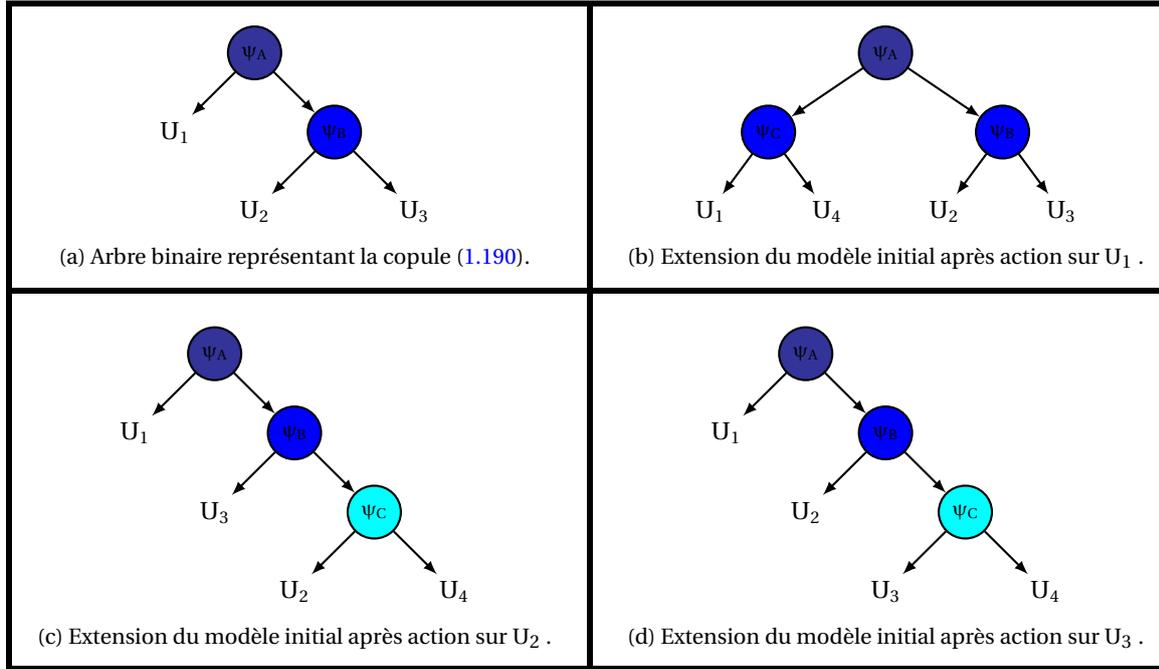


FIGURE 1.3 – Exemple d'un modèle HAC et de ses extensions.

Pour les modèles (1.186) et (1.189), en supposant que toutes les familles sont monoparamétriques,  $\theta$  comptabilise respectivement  $d$  et  $2d$  paramètres. Le nombre total de paramètres est donc d'ordre  $\mathcal{O}(d)$ . À titre de comparaison, pour les copules R-vignes, si la descente séquentielle est conduite jusqu'à l'arbre  $T_{d-1}$ , le nombre de paramètres est d'ordre  $\mathcal{O}(d^2)$ . Il peut néanmoins être ramené à  $\mathcal{O}(d)$  si une stratégie de troncature est envisagée.

Les distributions CFM permettent de construire des modèles de copules plus parcimonieux tout en maintenant un fort niveau de polymorphie et d'asymétrie. Il faut néanmoins garder à l'esprit que l'expression (1.189) de  $c_{CFM}$  découle d'une indépendance supposée des variables initiales conditionnellement à la connaissance d'une collection de variables latentes dont on ignore le nombre. De ce point de vue, on peut se demander s'il est vraiment raisonnable d'espérer décrire l'intégralité de la structure latente avec au plus  $r = 2$  variables.

### 1.5.4.3 Copules archimédiennes emboîtées

Les copules archimédiennes introduites à la Section 1.5.2.2 fournissent un large éventail de motifs bidimensionnels dont l'intérêt repose principalement sur l'asymétrie entre les queues de distribution. Toutefois, dès que  $d \geq 3$ , le résultat (1.148) est très limitant en termes de flexibilité car il impose que toutes les paires  $(U_i, U_j)$  soient modélisées par la même copule bidimensionnelle. Par conséquent, l'intérêt de ces copules devient très relatif, et surtout théorique. Les modèles HAC (*Hierarchical Archimedean Copulas*) sont apparus avec JOE [1996] qui souligne dès le départ que leur simplicité apparente dissimule de nombreux problèmes. Pour construire une copule HAC de dimension  $d = 3$ , il suffit de partir d'une copule archimédienne bidimensionnelle de générateur  $\psi_A$  puis de remplacer son second argument par la fonction de répartition d'une copule archimédienne de générateur  $\psi_B$  :

$$C_{\mathbf{U}}(\mathbf{u} \mid \psi_A, \psi_B) = C_{\psi_A}(u_1, C_{\psi_B}(u_2, u_3)) = \psi_A\left(\psi_A^{-1}(u_1) + \psi_A^{-1}\left(\psi_B(\psi_B^{-1}(u_2) + \psi_B^{-1}(u_3))\right)\right). \quad (1.190)$$

Le squelette de la structure de dépendance de cette copule peut être représenté graphiquement à l'aide d'un arbre binaire<sup>4</sup> noté  $\mathbf{A}$ . Pour la copule (1.190), la Figure 1.3a représente son arbre

4. Avant de définir ce qu'est un arbre binaire, on a besoin de préciser que le degré d'un nœud désigne le nombre d'arêtes qui sont connectées à ce nœud. Dans le cas d'un graphe orienté, il faut comptabiliser les arêtes partant de ce

binaire. Le premier générateur est à la racine, le second occupe un nœud intermédiaire tandis que les feuilles sont réservées aux différentes variables  $U_j$ . En suivant le même principe, on peut choisir d'imbriquer un troisième générateur archimédien  $\psi_C$  et de remplacer au choix l'une des variables  $u_1$ ,  $u_2$  ou  $u_3$ . On obtient alors trois extensions différentes (voir Figure 1.3) en dimension  $d = 4$ . Lorsqu'il existe une feuille de profondeur  $d - 1$ , la copule est dite totalement emboîtée. Le cas échéant, on parle d'emboîtement partiel. On peut itérer ce processus pour construire un modèle HAC en dimension  $d$  quelconque. Une généralisation consiste à combiner des copules  $\mathcal{A}_{r_1}(\psi_1), \dots, \mathcal{A}_{r_q}(\psi_q)$  de dimensions  $r_1, \dots, r_q$  libres d'être supérieures à 2. Dans cette situation, l'arbre  $\mathbf{A}$  n'est alors plus binaire mais enraciné.

Les modèles HAC ont été appliqués en ingénierie financière pour la gestion des portefeuilles d'actifs où des structures hiérarchiques émergent souvent des données opérationnelles. Des cas d'étude sont disponibles dans SAVU et TREDE [2010] et HOFERT et SCHERER [2011]. Si l'imbrication de plusieurs générateurs au sein d'une même copule permet indéniablement aux modèles HAC de gagner en flexibilité, plusieurs défauts majeurs viennent limiter leur utilisation en pratique. D'une part, le problème du manque de flexibilité n'est que partiellement résolu. On constate par exemple que les paires  $(U_1, U_2)$  et  $(U_1, U_3)$  de la copule (1.190) ont exactement la même loi. D'autre part, l'application  $C_U(\cdot | \psi_A, \psi_B)$  n'est pas obligatoirement la fonction de répartition d'une copule. MCNEIL [2008] énonce des conditions suffisantes pour fabriquer une  $d$ -copule.

### **Théorème 1.5.3 (Condition d'emboîtement)**

Soient  $\psi_1, \dots, \psi_{d-1} \in \Psi_\infty$ .

On définit récursivement l'application :

$$\begin{aligned} C(\cdot | \psi_1 \dots \psi_{d-1}) : \mathcal{H}_d &\longrightarrow [0, 1] \\ \mathbf{u} &\longmapsto C_{\psi_1} \left( u_1, C(u_2, \dots, u_d | \psi_2, \dots, \psi_{d-1}) \right) \end{aligned} \quad (1.191)$$

Si toutes les fonctions  $\psi_k^{-1} \circ \psi_{k+1}$  sont totalement monotones au sens de la Définition 1.5.6, alors  $C(\cdot | \psi_1 \dots \psi_{d-1})$  est la fonction de répartition d'une  $d$ -copule.

Ce théorème est formulé pour les copules archimédiennes complètement emboîtées mais il vaut aussi pour celles qui ne sont que partiellement emboîtées. Lorsque les générateurs  $\psi_k$  sont de la même famille, il est montré dans NELSEN [2007] que la condition d'emboîtement impose des règles simples sur la suite des paramètres  $\theta_k$ . Il suffit en fait que les paramètres diminuent au fur et à mesure qu'on remonte des feuilles à la racine dans l'arbre  $\mathbf{A}$ .

Un modèle HAC est complètement caractérisé par son arbre binaire  $\mathbf{A}$ , la collection  $\boldsymbol{\psi}$  de ses générateurs et le vecteur  $\boldsymbol{\theta}$  rassemblant les paramètres qui leur sont associés. En s'appuyant sur un catalogue de générateurs archimédiens, OKHRIN et collab. [2013] a développé un algorithme destiné à l'apprentissage du triplet  $(\mathbf{A}, \boldsymbol{\psi}, \boldsymbol{\theta})$ . Les principales idées de la procédure inférentielle sont esquissées ci-dessous :

- Pour chaque paire  $(U_i, U_j)$ , on estime par maximum de vraisemblance les paramètres de tous les générateurs du catalogue et on calcule la statistique du test d'adéquation de CHEN et collab. [2004]. La meilleure paire et le meilleur générateur sont identifiés grâce à la plus petite valeur de la statistique de test.
- À une renumérotation près, on suppose que la première étape a permis de cibler la paire  $(U_1, U_2)$  et le générateur  $\psi_A$ . Il faut alors remplacer dans  $\mathbf{U}_{\text{obs}}$  les observations de  $U_1$  et  $U_2$  par celles de la variable  $Z_{12} := \hat{C}_{\psi_A}(U_1, U_2 | \hat{\theta}_A)$ .
- On répète les deux premières étapes. Si  $Z_{12}$  est sollicitée, l'arbre binaire temporaire est celui d'une copule archimédienne totalement emboîtée. En revanche, si ce sont d'autres variables

---

nœud et celles arrivant à ce nœud. Un arbre est dit binaire si le degré de chaque nœud est au plus égal à 3. Pour chaque nœud, on peut définir un père ainsi qu'un ou deux fils. Le seul nœud sans père est appelé la racine. Les nœuds sans fils sont appelés les feuilles. La profondeur d'un nœud désigne sa distance (en nombre d'arêtes) à la racine.

$U_i$  et  $U_j$  (avec  $j > i \geq 3$ ) qui sont appariées, un emboîtement partiel se met en place. La procédure se poursuit ensuite jusqu'à ce qu'il ne reste plus qu'une seule paire de variables.

#### 1.5.4.4 Produit de copules bidimensionnelles

Pour comprendre l'intuition ayant motivé la construction de cette nouvelle classe de copules, il faut remarquer que le produit des fonctions de répartition de plusieurs copules ne conduit pas en général à la fonction de répartition d'une autre copule. Néanmoins, il est facile de s'y ramener. En reprenant l'exemple introductif de MAZO [2014], la copule indépendance  $\Pi_2$  révèle que :

$$(\Pi_2(u_1, u_2))^2 = (u_1 u_2)^2 \text{ donc } (\Pi_2)^2 \notin \mathcal{C}_2 \text{ mais } \left( \Pi_2(\sqrt{u_1}, \sqrt{u_2}) \right)^2 = u_1 u_2 = \Pi_2(u_1, u_2). \quad (1.192)$$

La clé de voûte est un théorème proposé par LIEBSCHER [2008].

#### Théorème 1.5.4 (Liebscher)

Soit  $\mathbf{C} := \{C_k\}_{k=1}^K$  une collection de fonctions de répartition de  $d$ -copules.  
 Pour chaque copule  $C_k$ , soit  $\mathbf{g}_k := \{g_{kl}\}_{l=1}^d$  une collection de fonctions de  $[0, 1]$  dans  $[0, 1]$ .  
 Toute fonction  $g_{kl}$  est au choix :

- strictement croissante avec  $\lim_{v \rightarrow 0} g_{kl}(v) = 0$ ,
- constamment égale à 1.

Si  $\prod_{k=1}^K g_{kl}(v) = v$ , alors l'application définie par :

$$\begin{aligned} C_{\text{pr}} : \mathcal{H}_d &\longrightarrow [0, 1] \\ \mathbf{u} &\longmapsto \prod_{k=1}^K C_k(g_{k1}(u_1), \dots, g_{kd}(u_d)) \end{aligned} \quad (1.193)$$

est la fonction de répartition d'une  $d$ -copule.

MAZO et collab. [2015] ont créé les modèles PBC (*Product of Bivariate Copulas*) à partir de ce résultat. Les liens de dépendance entre les variables sont représentés au sein d'un graphe  $\mathbf{G} = (\mathbf{N}, \mathbf{E})$  où  $\mathbf{N} := \{1, \dots, d\}$ . On peut tracer autant d'arêtes qu'on le souhaite entre des sommets car  $\mathbf{G}$  n'est pas nécessairement un arbre couvrant, ni même un arbre<sup>5</sup>. La présence d'une arête entre deux sommets indique simplement que l'utilisateur souhaite une modélisation explicite de la copule bidimensionnelle correspondante. Un exemple en dimension  $d = 5$  est proposé à la Figure 1.4 où l'on voit que les arêtes entre les sommets 2, 3 et 4 forment un cycle.

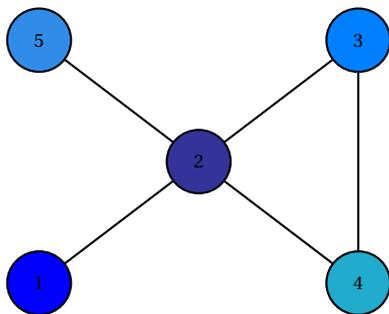


FIGURE 1.4 – Graphe  $\mathbf{G}$  pour un modèle PBC.

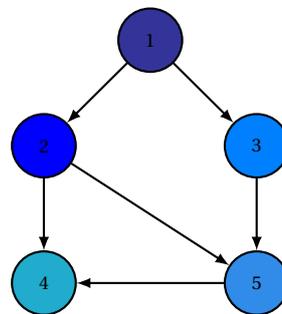


FIGURE 1.5 – Graphe  $\mathbf{D}$  pour un modèle CBN.

Pour chaque sommet  $j$  de  $\mathbf{N}$ , on note  $n_j$  le nombre de voisins auquel il est connecté. À chaque arête  $(i, j)$  de  $\mathbf{E}$  est affectée une copule bidimensionnelle  $\tilde{C}_{ij}$  et MAZO et collab. [2015] proposent

5. Un arbre est simplement connexe et acyclique. Un sommet peut très bien y être isolé. Au contraire, les arêtes d'un arbre couvrant connectent tous les sommets entre eux.

alors de définir une  $d$ -copule de la manière suivante :

$$\forall \mathbf{u} \in \mathcal{H}_d, \quad C_{\mathbf{U}}(\mathbf{u}) = \prod_{e \in \mathbf{E}} \tilde{C}_{ij} \left( u_i^{1/n_i}, u_j^{1/n_j} \right) . \quad (1.194)$$

Le Théorème 1.5.4 justifie que l'application ainsi créée est une  $d$ -copule. En effet, il suffit de voir que  $K = \frac{1}{2}(n_1 + \dots + n_d)$  et de poser :

$$\forall \mathbf{u} \in \mathcal{H}_d, \quad C_k(\mathbf{u}) := \tilde{C}_{ij}(u_i, u_j) \left( \prod_{l \notin \{i, j\}} u_l \right) \text{ et } \forall 1 \leq l \leq d, \quad g_{kl}(v) := \begin{cases} v^{1/n_l} & \text{si } l \in \{i, j\} \\ 1 & \text{sinon} \end{cases} , \quad (1.195)$$

pour que l'expression (1.194) devienne (1.193). Comme toutes les hypothèses du Théorème 1.5.4 sont vérifiées, la classe PBC est uniquement constituée de  $d$ -copules. Dans le cas particulier de la copule schématisée à la Figure 1.4, on obtient :

$$C_{\mathbf{U}}(\mathbf{u}) = C_{12}(u_1, u_2^{1/4}) C_{23}(u_2^{1/4}, u_3^{1/2}) C_{24}(u_2^{1/4}, u_4^{1/2}) C_{25}(u_2^{1/4}, u_5) C_{34}(u_3^{1/2}, u_4^{1/2}) . \quad (1.196)$$

On souhaite attirer l'attention sur le fait que la copule  $\tilde{C}_{ij}$  impliquée dans la définition (1.194) n'est pas exactement la copule de la paire de variables  $(U_i, U_j)$  dans le modèle final :

$$C_{ij}(u_i, u_j) = C_{\mathbf{U}}(1, \dots, 1, u_i, 1, \dots, 1, u_j, 1, \dots, 1) = \tilde{C}_{ij} \left( u_i^{1/n_i}, u_j^{1/n_j} \right) u_i^{(n_i-1)/n_i} u_j^{(n_j-1)/n_j} . \quad (1.197)$$

Après avoir mis en place un paramétrage des copules  $\tilde{C}_{ij}$ , on aboutit à une distribution PBC :

$$C_{\text{PBC}}(\mathbf{u} \mid \mathbf{G}, \mathbf{F}, \boldsymbol{\theta}) = \prod_{e \in \mathbf{E}} \tilde{C}_{ij} \left( u_i^{1/n_i}, u_j^{1/n_j} \mid \mathcal{F}_e, \boldsymbol{\theta}_e \right) \text{ avec } \mathbf{F} = \bigcup_{e \in \mathbf{E}} \mathcal{F}_e \text{ et } \boldsymbol{\theta} = \bigcup_{e \in \mathbf{E}} \boldsymbol{\theta}_e . \quad (1.198)$$

À supposer qu'un graphe  $\mathbf{G}$  soit mis en place et que des familles  $\mathbf{F}$  soient affectées aux différentes copules bidimensionnelles, il reste à examiner les modalités d'estimation des paramètres  $\boldsymbol{\theta}$ . La méthode du maximum de vraisemblance n'est ici pas immédiate car il faut être capable d'évaluer la densité  $c_{\text{PBC}}$ , c'est-à-dire d'évaluer une dérivée partielle d'ordre  $d$  de la fonction de répartition  $C_{\text{PBC}}$  qui est un produit de  $K = \text{Card}(\mathbf{E})$  termes. On peut y parvenir avec l'algorithme "message-passing" de HUANG et JOJIC [2010] à condition que  $\mathbf{G}$  soit un arbre. Pour cette raison, il est toujours supposé en pratique que  $\mathbf{G}$  est un arbre (par exemple l'arbre couvrant de poids maximal pour une certaine mesure de dépendance). De la capacité à calculer la log-vraisemblance découle la possibilité d'utiliser un critère d'information (AIC/BIC) afin de sélectionner les familles  $\mathcal{F}_e^*$  offrant le meilleur niveau d'adéquation aux données de  $\mathbf{U}_{\text{obs}}$ .

Les modèles PBC souffrent d'une incapacité à estimer les motifs de dépendance pour lesquels une forte structure de corrélation est à l'œuvre. Pour un modèle PBC, MAZO et collab. [2015] montrent que le tau de Kendall  $\tau_{ij}$  d'une paire  $(U_i, U_j)$  ne peut exister qu'au sein d'un intervalle  $[a_{ij}, b_{ij}]$  avec  $a_{ij} > -1$  et  $b_{ij} < 1$ . De plus, l'amplitude diminue lorsque  $n_i$  et/ou  $n_j$  augmentent. Une parade peut consister à retirer des arêtes mais cela revient à occulter la dépendance s'exerçant entre certaines paires. Ce dilemme est l'un des principaux défauts de la classe PBC.

#### 1.5.4.5 Réseaux bayésiens de copules

La théorie des réseaux bayésiens a été développée pour aller au-delà de la factorisation de  $f_{\mathbf{X}}$  dictée par le théorème de Bayes. L'idée est d'exploiter les relations d'indépendance conditionnelle entre les variables  $X_j$  pour éliminer les variables conditionnantes inutiles. Pour une introduction sur ces thèmes, on renvoie aux ouvrages de LAURITZEN [1996] ou de COWELL et collab. [2006]. La modélisation repose sur un graphe orienté acyclique  $\mathbf{D} = (\mathbf{N}, \mathbf{E})$  avec  $\mathbf{N} := \{1, \dots, d\}$  l'ensemble des nœuds et  $\mathbf{E}$  un ensemble d'arcs orientés de type  $(i \rightarrow j)$  avec  $1 \leq i \neq j \leq d$ . Dans la littérature, l'acronyme anglais est DAG (*Direct Acyclic Graph*). La Figure 1.5 présente un exemple pour  $d = 5$ . Le fait que les arcs soient orientés permet de définir certains concepts généalogiques au sein de  $\mathbf{D}$ .

### Définition 1.5.8 (Parents, descendants et non-descendants)

Soit  $\mathbf{D} := (\mathbf{N}, \mathbf{E})$  un graphe orienté acyclique.

Pour chaque sommet  $i$ , on définit :

- ses parents :  $\text{Pa}(i) := \{j \in \mathbf{N} : (j \rightarrow i) \in \mathbf{E}\}$ ,
- ses descendants :  $\text{De}(i) := \{j \in \mathbf{N} : (i \rightarrow i_1 \rightarrow \dots \rightarrow i_r \rightarrow j) \in \mathbf{E}\}$ ,
- ses non-descendants :  $\text{Nd}(i) := \mathbf{N} \setminus (\text{De}(i) \cup \{i\})$ .

Pour la Figure 1.5, les définitions précédentes appliquées au sommet 3 donnent :

$$\text{Pa}(3) = \{1\}, \quad \text{De}(3) = \{4, 5\}, \quad \text{et} \quad \text{Nd}(3) = \{1, 2\} . \quad (1.199)$$

Le lien entre  $\mathbf{D}$  et la loi  $\mu_{\mathbf{X}}$  est indiqué ci-dessous :

### Définition 1.5.9 (Loi $\mathbf{D}$ -markovienne)

Soit  $\mathbf{D} = (\mathbf{N}, \mathbf{E})$  un graphe orienté acyclique.

Soit  $\mathbf{X}$  un vecteur aléatoire à valeurs dans  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ .

La loi  $\mu_{\mathbf{X}}$  est dite  $\mathbf{D}$ -markovienne si :

$$\forall 1 \leq j \leq d, \quad X_j \perp \mathbf{X}_{\text{Nd}(j) \setminus \text{Pa}(j)} \mid \mathbf{X}_{\text{Pa}(j)} . \quad (1.200)$$

Dans le cas de la Figure 1.5, on obtient les relations suivantes<sup>6</sup> :

$$\begin{aligned} \text{(IC2)} \quad X_2 \perp X_3 \mid X_1, & & \text{(IC4)} \quad X_4 \perp (X_1, X_3) \mid (X_2, X_5), \\ \text{(IC3)} \quad X_3 \perp X_2 \mid X_1, & & \text{(IC5)} \quad X_5 \perp X_1 \mid (X_2, X_3). \end{aligned}$$

Contrairement aux copules R-vignes où les arêtes des arbres  $\mathbf{T}_i$  indiquent les paires pour lesquelles un lien de dépendance est identifié, les arcs permettent ici d'établir des relations d'indépendance conditionnelle. Le fait que toutes les assertions (1.200) soient vérifiées par une loi  $\mu_{\mathbf{X}}$  implique que la densité  $f_{\mathbf{X}}$  est factorisable comme suit :

$$\forall \mathbf{x} \in \mathbb{R}^d, \quad f_{\mathbf{X}}(\mathbf{x}) = \prod_{j=1}^d f_{X_j \mid \mathbf{X}_{\text{Pa}(j)}}(x_j \mid \mathbf{x}_{\text{Pa}(j)}) . \quad (1.201)$$

Pour le graphe de la Figure 1.5, si  $\mu_{\mathbf{X}}$  est  $\mathbf{D}$ -markovienne, la densité  $f_{\mathbf{X}}$  s'écrit :

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1}(x_1) f_{X_2 \mid X_1}(x_2 \mid x_1) f_{X_3 \mid X_1}(x_3 \mid x_1) f_{X_4 \mid X_{25}}(x_4 \mid \mathbf{x}_{25}) f_{X_5 \mid X_{23}}(x_5 \mid \mathbf{x}_{23}) . \quad (1.202)$$

Pour s'en persuader, on peut écrire la version suivante du théorème de Bayes :

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1}(x_1) f_{X_2 \mid X_1}(x_2 \mid x_1) \underbrace{f_{X_3 \mid X_{12}}(x_3 \mid \mathbf{x}_{12})}_{\text{(IC2)} \Rightarrow f_{X_3 \mid X_1}} \underbrace{f_{X_5 \mid X_{123}}(x_5 \mid \mathbf{x}_{123})}_{\text{(IC5)} \Rightarrow f_{X_5 \mid X_{23}}} \underbrace{f_{X_4 \mid X_{1235}}(x_4 \mid \mathbf{x}_{1235})}_{\text{(IC4)} \Rightarrow f_{X_4 \mid X_{25}}} , \quad (1.203)$$

et utiliser les relations d'indépendance conditionnelle fournies par  $\mathbf{D}$  pour retrouver (1.202). La réciproque est aussi vraie. Toute factorisation de type (1.201) résume parfaitement l'information du réseau bayésien spécifié par  $\mathbf{D}$ . Pour un échantillon  $\mathbf{X}_{\text{obs}}$  d'une loi  $\mu_{\mathbf{X}}$  inconnue, trouver un graphe  $\mathbf{D}$  tel que  $\mu_{\mathbf{X}}$  soit  $\mathbf{D}$ -markovienne est une tâche ardue qui nécessite le recours à des algorithmes de "structure learning" tel que l'algorithme PC<sup>7</sup> de SPIRITES et GLYMOUR [1991]. La procédure est initialisée avec un réseau vide (sans arc) puis le graphe  $\mathbf{D}$  est construit par répétition d'un tryptique d'opérations Ajouter/Effacer/Retourner. Pour des explications détaillées, on renvoie à l'ouvrage

6. On constate une redondance d'information entre les relations (IC2) et (IC3).

7. En l'honneur de Peter Spirtes et Clark Glymour.

de **KOLLER et FRIEDMAN [2009]**. L'idée des réseaux bayésiens est appliquée directement à la densité de copule  $c_{\mathbf{U}}$  par **ELIDAN [2010, 2012]**. En notant  $\mathbf{P}_j := \mathbf{U}_{\{j\} \cup \text{Pa}(j)}$  et  $\text{Pa}(j) = \{j_1, \dots, j_r\}$ , il fait remarquer qu'on peut écrire :

$$f_{X_j | \mathbf{X}_{\text{Pa}(j)}}(x_j | \mathbf{x}_{\text{Pa}(j)}) = R_j(F_{X_j}(u_j), F_{X_{j_1}}(x_{j_1}), \dots, F_{X_{j_r}}(x_{j_r})) f_{X_j}(x_j)$$

$$\text{avec } R_j(u_j, \mathbf{u}_{\text{Pa}(j)}) := \frac{c_{\mathbf{P}_j}(u_j, u_{j_1}, \dots, u_{j_r})}{\frac{\partial^r c_{\mathbf{P}_j}}{\partial u_{j_1}, \dots, \partial u_{j_r}}(1, u_{j_1}, \dots, u_{j_r})}, \quad (1.204)$$

et que la factorisation (1.201) devient alors :

$$\forall \mathbf{x} \in \mathbb{R}^d, \quad f_{\mathbf{X}}(\mathbf{x}) = \left( \prod_{j=1}^d f_{X_j}(x_j) \right) \times \underbrace{\left( \prod_{j=1}^d R_j(u_j, \mathbf{u}_{\text{Pa}(j)}) \right)}_{c_{\mathbf{U}}(\mathbf{u})}. \quad (1.205)$$

Cela permet de reporter l'information du réseau bayésien sur la copule et de donner naissance au modèle CBN (*Copula Bayesian Network*). La phase d'apprentissage passe alors par l'estimation paramétrique des copules associées aux vecteurs  $\mathbf{P}_j$  :

$$c_{\text{CBN}}(\mathbf{u} | \mathbf{D}, \mathbf{F}, \boldsymbol{\theta}) = \prod_{j=1}^d R_j(u_j, \mathbf{u}_{\text{Pa}(j)} | \mathcal{F}_j, \boldsymbol{\theta}_j) \text{ avec } \mathbf{F} = \bigcup_{j=1}^d \mathcal{F}_j \text{ et } \boldsymbol{\theta} = \bigcup_{j=1}^d \boldsymbol{\theta}_j. \quad (1.206)$$

Si  $\mathbf{D}$  et  $\mathbf{F}$  sont connus, on peut calculer la log-vraisemblance du modèle CBN pour l'échantillon  $\mathbf{U}_{\text{obs}}$ . À l'instar de ce qui est fait pour les autres modèles, l'utilisation d'un critère d'information (AIC/BIC) permet d'identifier les familles paramétriques  $\mathcal{F}_1^*, \dots, \mathcal{F}_d^*$  qui sont les plus à même de représenter les données disponibles.

Un modèle permettant de n'impliquer que des copules bidimensionnelles est développé dans **BAUER et collab. [2012]**. Pour chaque sommet  $j$  de  $\mathbf{N}$ , il faut commencer par définir arbitrairement une relation d'ordre total  $<_j$  entre les indices de  $\text{Pa}(j)$ . On est ensuite en mesure de poser :

$$\forall 1 \leq j \leq d, \quad \forall k \in \text{Pa}(j), \quad \text{Pa}(j, k) := \{l \in \text{Pa}(j) : l <_j k\}. \quad (1.207)$$

Ensuite, au lieu de faire intervenir la fonction  $R_j$  de l'équation (1.204), il est démontré que  $f_{X_j | \mathbf{X}_{\text{Pa}(j)}}$  peut être décomposée uniquement à partir de copules bidimensionnelles :

$$f_{X_j | \mathbf{X}_{\text{Pa}(j)}}(x_j | \mathbf{x}_{\text{Pa}(j)}) = f_{X_j}(x_j) \prod_{k \in \text{Pa}(j)} c_{jk | \text{Pa}(j, k)}(u_j | \text{Pa}(j, k), u_k | \text{Pa}(j, k)). \quad (1.208)$$

Comme pour les R-vignes, toutes les copules conditionnelles sont supposées vérifier l'hypothèse "simplificatrice". Le modèle PBCN (*Pair Copula Bayesian Networks*) prend alors la forme suivante :

$$c_{\text{PBCN}}(\mathbf{u} | \mathbf{D}, \mathbf{F}, \boldsymbol{\theta}) = \prod_{j=1}^d \prod_{k \in \text{Pa}(j)} c_{jk | \text{Pa}(j, k)}(u_j | \text{Pa}(j, k), u_k | \text{Pa}(j, k) | \mathcal{F}_{jk}, \boldsymbol{\theta}_{jk})$$

$$\text{avec } \mathbf{F} = \bigcup_{j=1}^d \bigcup_{k \in \text{Pa}(j)} \mathcal{F}_{jk} \text{ et } \boldsymbol{\theta} = \bigcup_{j=1}^d \bigcup_{k \in \text{Pa}(j)} \boldsymbol{\theta}_{jk}. \quad (1.209)$$

L'utilisation de l'algorithme PC pour un modèle PBCN est détaillée dans **BAUER et CZADO [2016]**.

## 1.6 Conclusion

Dans ce chapitre, on a présenté une très large palette d'outils de statistique inférentielle. Tous peuvent être utilisés lors de l'apprentissage d'une distribution multidimensionnelle à partir d'un échantillon d'une loi sous-jacente inconnue. Bien qu'il existe plusieurs caractérisations possibles

de la loi d'un vecteur aléatoire, c'est l'estimation de la densité de probabilité conjointe qui est l'objectif central de tout algorithme d'apprentissage statistique. Ici, on a fait le choix de se limiter à l'étude des modèles de distributions basés sur la séparation copule-marginales car ils offrent suffisamment de flexibilité et d'asymétrie pour estimer avec précision la structure de dépendance et les queues de distribution.

Parmi les nombreuses techniques permettant de calibrer un modèle de densité marginale, une frontière claire est apparue entre les méthodes paramétriques, et tout particulièrement celle du maximum de vraisemblance, et les méthodes non-paramétriques telles que KDE et ses variantes. Les différents cas tests des Chapitres 4 et 5 utilisent très largement ces deux approches.

En ce qui concerne l'estimation de la densité de copule, véritable enjeu de l'apprentissage en plus grande dimension, plusieurs idées importantes ont été dégagées. Dans un premier temps, il est important de souligner qu'il existe un large savoir-faire paramétrique et non-paramétrique pour estimer avec précision les copules bidimensionnelles. Dans un second temps, il faut bien comprendre que la généralisation en plus grande dimension des classes paramétriques de copules bidimensionnelles n'est pas assez sophistiquée pour que les modèles résultants puissent s'approprier la complexité des nouveaux motifs rencontrés. Les techniques non-paramétriques d'estimation d'une densité de copule ne font guère mieux car elles se heurtent à l'incurable malédiction de la dimension. Des modèles ont été développés pour gagner en polymorphie (R-vignes, CFM, HAC, PBC, PCBN) et reposent tous sur un mécanisme d'imbrication de copules bidimensionnelles.

Les copules R-vignes et les réseaux bayésiens sont les approches les plus largement diffusées dans la communauté scientifique. D'ailleurs, il a été récemment montré qu'ils entretiennent un lien étroit, les seconds pouvant être vus comme une version parcimonieuse des premiers. Pour le cas d'étude commun des Chapitres 4 et 5, on privilégie l'utilisation des copules R-vignes pour deux raisons principales. D'une part leur procédure inférentielle est plus simple. En effet, la construction du graphe orienté acyclique d'un réseau bayésien est bien plus ardue que celle de la séquence d'arbres d'une copule R-vigne. Par ailleurs, ces deux modèles se focalisent sur des informations antagonistes (relations d'indépendance conditionnelle pour les réseaux bayésiens, relations de dépendance conditionnelle pour les copules R-vignes) et certaines des sous-problématiques de cette thèse pointent clairement en direction du second point de vue.

## Références

- AAS, K., C. CZADO, A. FRIGESSI et H. BAKKEN. 2009, «Pair-copula constructions of multiple dependence», *Insurance : Mathematics and economics*, vol. 44, n° 2, p. 182–198.
- ACAR, E. F., C. GENEST et J. NEŠLEHOVÁ. 2012, «Beyond simplified pair-copula constructions», *Journal of Multivariate Analysis*, vol. 110, p. 74–90.
- AKAIKE, H. 1973, «Information theory and an extension of maximum likelihood principle», dans *Proc. 2nd Int. Symp. on Information Theory*, p. 267–281.
- AKAIKE, H. 1974, «A new look at the statistical model identification», *IEEE transactions on automatic control*, vol. 19, n° 6, p. 716–723.
- ALINK, S., M. LÖWE et M. V. WÜTHRICH. 2007, «Diversification for general copula dependence», *Statistica Neerlandica*, vol. 61, n° 4, p. 446–465.
- ANDERSON, D. et K. BURNHAM. 2004, «Model selection and multi-model inference», *Second. NY : Springer-Verlag*, vol. 63, n° 2020, p. 10.
- BABA, K., R. SHIBATA et M. SIBUYA. 2004, «Partial correlation and conditional correlation as measures of conditional independence», *Australian & New Zealand Journal of Statistics*, vol. 46, n° 4, p. 657–664.

- BARBE, P., C. GENEST, K. GHOUDI et B. REMILLARD. 1996, «On Kendall's process», *Journal of Multivariate Analysis*, vol. 58, n° 2, p. 197–229.
- BASSETT, R. et J. DERIDE. 2019, «Maximum a posteriori estimators as a limit of Bayes estimators», *Mathematical Programming*, vol. 174, n° 1-2, p. 129–144.
- BAUER, A. et C. CZADO. 2016, «Pair-copula Bayesian networks», *Journal of Computational and Graphical Statistics*, vol. 25, n° 4, p. 1248–1271.
- BAUER, A., C. CZADO et T. KLEIN. 2012, «Pair-copula constructions for non-Gaussian DAG models», *Canadian Journal of Statistics*, vol. 40, n° 1, p. 86–109.
- BEDFORD, T. et R. M. COOKE. 2001, «Probability density decomposition for conditionally dependent random variables modeled by vines», *Annals of Mathematics and Artificial Intelligence*, vol. 32, n° 1, p. 245–268.
- BEDFORD, T. et R. M. COOKE. 2002, «Vines : A new graphical model for dependent random variables», *Annals of Statistics*, p. 1031–1068.
- BEDFORD, T., A. DANESHKHAH et K. J. WILSON. 2016, «Approximate uncertainty modeling in risk analysis with vine copulas», *Risk Analysis*, vol. 36, n° 4, p. 792–815.
- BEER, M., Y. ZHANG, S. T. QUEK et K. K. PHOON. 2013, «Reliability analysis with scarce information : comparing alternative approaches in a geotechnical engineering context», *Structural Safety*, vol. 41, p. 1–10.
- BEHNEN, K., M. HUŠKOVÁ et G. NEUHAUS. 1985, «Rank estimators of scores for testing independence», *Statistics & Risk Modeling*, vol. 3, n° 3-4, p. 239–262.
- BENES, V. et J. STEPÁN. 2012, *Distributions with given marginals and moment problems*, Springer Science & Business Media.
- BERNOULLI, J. 1711, «Lettre écrite de Bâle le 10 janvier 1711, touchant la manière de trouver les forces centrales dans des milieux résistants en raisons composées de leurs densités et des puissances quelconque des vitesses du mobile», *Mémoires de mathématique et de physique de l'Académie royale des sciences*.
- BILLARD, L. et E. DIDAY. 2000, «Regression analysis for interval-valued data», dans *Data Analysis, Classification, and Related Methods*, Springer, p. 369–374.
- BLOMQUIST, N. 1950, «On a measure of dependence between two random variables», *The Annals of Mathematical Statistics*, p. 593–600.
- BRECHMANN, E. C., C. CZADO et K. AAS. 2012, «Truncated regular vines in high dimensions with application to financial data», *Canadian Journal of Statistics*, vol. 40, n° 1, p. 68–85.
- BREYMAN, W., A. DIAS et P. EMBRECHTS. 2003, «Dependence structures for multivariate high-frequency data in finance», *Quantitative Finance*, vol. 3, n° 1, p. 1–14.
- BROWN, B. M. et S. X. CHEN. 1999, «Beta-Bernstein smoothing for regression curves with compact support», *Scandinavian Journal of Statistics*, vol. 26, n° 1, p. 47–59.
- BROWNE, M. W. et R. CUDECK. 1992, «Alternative ways of assessing model fit», *Sociological methods & research*, vol. 21, n° 2, p. 230–258.
- CAPÉRAA, P., A.-L. FOUGÈRES et C. GENEST. 1997, «A nonparametric estimation procedure for bivariate extreme value copulas», *Biometrika*, vol. 84, n° 3, p. 567–577.

- CAPÉRAÀ, P. et C. GENEST. 1993, «Spearman's  $\rho$  is larger than Kendall's  $\tau$  for positively dependent random variables», *Journal of Nonparametric Statistics*, vol. 2, n° 2, p. 183–194.
- CHAFAY, D. et F. MALRIEU. 2016, *Recueil de modèles aléatoires*, vol. 78, Springer.
- CHARPENTIER, A., J.-D. FERMANIAN et O. SCAILLET. 2007, «The estimation of copulas : theory and practice», *Copulas : From theory to application in finance*, p. 35–64.
- CHARPENTIER, A., A.-L. FOUGÈRES, C. GENEST et J. NEŠLEHOVÁ. 2014, «Multivariate Archimax copulas», *Journal of Multivariate Analysis*, vol. 126, p. 118–136.
- CHEN, X., Y. FAN et A. J. PATTON. 2004, «Simple tests for models of dependence between multiple financial time series, with applications to US equity returns and exchange rates», .
- CHOROŚ, B., R. IBRAGIMOV et E. PERMIAKOVA. 2010, «Copula estimation», dans *Copula theory and its applications*, Springer, p. 77–91.
- CLAESKENS, G., N. L. HJORT et collab.. 2008, «Model selection and model averaging», *Cambridge Books*.
- CLAYTON, D. G. 1978, «A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence», *Biometrika*, vol. 65, n° 1, p. 141–151.
- COOK, R. D. et M. E. JOHNSON. 1981, «A family of distributions for modelling non-elliptically symmetric multivariate data», *Journal of the Royal Statistical Society : Series B (Methodological)*, vol. 43, n° 2, p. 210–218.
- COWELL, R. G., P. DAWID, S. L. LAURITZEN et D. J. SPIEGELHALTER. 2006, *Probabilistic networks and expert systems : exact computational methods for Bayesian networks*, Springer Science & Business Media.
- CRAMÉR, H. 1928, «On the composition of elementary errors», *Scandinavian Actuarial Journal*, vol. 1928, n° 1, p. 13–74.
- CUADRAS, C. M., J. FORTIANA et J. A. RODRIGUEZ-LALLENA. 2002, *Distributions with given marginals and statistical modelling*, Springer.
- CZADO, C., S. JESKE et M. HOFMANN. 2013, «Selection strategies for regular vine copulae», *Journal de la Société Française de Statistique*, vol. 154, n° 1, p. 174–191.
- DANIELS, H. 1950, «Rank correlation and population models», *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 12, n° 2, p. 171–191.
- DARLING, D. A. 1957, «The Kolmogorov-Smirnov, Cramér-Von Mises tests», *The Annals of Mathematical Statistics*, vol. 28, n° 4, p. 823–838.
- DE MOIVRE, A. 1756, *The doctrine of chances : or, A method of calculating the probabilities of events in play*, vol. 200, Chelsea Publishing Company, Incorporated.
- DEHEUVELS, P. 1979, «La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance», *Bulletins de l'Académie Royale de Belgique*, vol. 65, n° 1, p. 274–292.
- DEHEUVELS, P. 1981, «An asymptotic decomposition for multivariate distribution-free tests of independence», *Journal of Multivariate Analysis*, vol. 11, n° 1, p. 102–113.
- DEHEUVELS, P. 1984, «Probabilistic aspects of multivariate extremes», dans *Statistical extremes and applications*, Springer, p. 117–130.

- DERENNES, P., J. MORIO et F. SIMATOS. 2018, «Estimation of moment-independent importance measures using a copula and maximum entropy framework», dans *2018 Winter Simulation Conference (WSC)*, IEEE, p. 1623–1634.
- DISSMANN, J., E. C. BRECHMANN, C. CZADO et D. KUROWICKA. 2013, «Selecting and estimating regular vine copulae and application to financial returns», *Computational Statistics & Data Analysis*, vol. 59, p. 52–69.
- DURANTE, F. et C. SEMPI. 2010, «Copula theory : an introduction», dans *Copula theory and its applications*, Springer, p. 3–31.
- EFRON, B. 1982, *The jackknife, the bootstrap and other resampling plans*, SIAM.
- ELIDAN, G. 2010, «Copula Bayesian networks.», dans *NIPS*, p. 559–567.
- ELIDAN, G. 2012, «Inference-less density estimation using copula Bayesian networks», *arXiv preprint arXiv :1203.3476*.
- EMBRECHTS, P. et M. HOFERT. 2013, «Risk measures and dependence modeling», dans *Handbook of Insurance*, Springer, p. 135–165.
- FANG, H.-B., K.-T. FANG et S. KOTZ. 2002, «The meta-elliptical distributions with given marginals», *Journal of Multivariate Analysis*, vol. 82, n° 1, p. 1–16.
- FANG, K. W. 2018, *Symmetric multivariate and related distributions*, CRC Press.
- FARIDAFSHIN, F. et A. NAESS. 2017, «Multivariate log-concave probability density class for structural reliability applications», *Structural Safety*, vol. 69, p. 57–67.
- FERMANIAN, J.-D., D. RADULOVIC, M. WEGKAMP et collab.. 2004, «Weak convergence of empirical copula processes», *Bernoulli*, vol. 10, n° 5, p. 847–860.
- FERMANIAN, J.-D. et collab.. 2005, «Goodness-of-fit tests for copulas», *Journal of Multivariate Analysis*, vol. 95, n° 1, p. 119–152.
- FISHER, R. A. 1924, «The distribution of the partial correlation coefficient», *Metron*, vol. 3, p. 329–332.
- FLORESCU, I. et C. A. TUDOR. 2013, *Handbook of probability*, John Wiley & Sons.
- FRANK, M. J. 1979, «On the simultaneous associativity of  $(x, y)$  and  $x + y - F(x, y)$ », *Aequationes mathematicae*, vol. 19, n° 1, p. 194–226.
- FRÉCHET, M. 1935, «Généralisation du théoreme des probabilités totales», *Fundamenta mathematicae*, vol. 1, n° 25, p. 379–387.
- FRIEDMAN, J., T. HASTIE, R. TIBSHIRANI et collab.. 2001, *The elements of statistical learning*, vol. 1, Springer Series in Statistics, New York.
- GABOW, H. N., Z. GALIL, T. SPENCER et R. E. TARJAN. 1986, «Efficient algorithms for finding minimum spanning trees in undirected and directed graphs», *Combinatorica*, vol. 6, n° 2, p. 109–122.
- GALAMBOS, J. 1978, *The asymptotic theory of extreme order statistics*, Wiley Series in Probability and Mathematical Statistics, New York, Chichester, Brisbane.
- GEENENS, G. 2014, «Probit transformation for kernel density estimation on the unit interval», *Journal of the American Statistical Association*, vol. 109, n° 505, p. 346–358.

- GEENENS, G., A. CHARPENTIER, D. PAINDAVEINE et collab.. 2017, «Probit transformation for non-parametric kernel estimation of the copula density», *Bernoulli*, vol. 23, n° 3, p. 1848–1873.
- GELMAN, A. 1995, «Method of moments using Monte-Carlo simulation», *Journal of Computational and Graphical Statistics*, vol. 4, n° 1, p. 36–54.
- GELMAN, A., J. B. CARLIN, H. S. STERN, D. B. DUNSON, A. VEHTARI et D. B. RUBIN. 2013, *Bayesian data analysis*, CRC press.
- GENEST, C. et A.-C. FAVRE. 2007, «Everything you always wanted to know about copula modeling but were afraid to ask», *Journal of hydrologic engineering*, vol. 12, n° 4, p. 347–368.
- GENEST, C., K. GHOUDI et L.-P. RIVEST. 1995, «A semiparametric estimation procedure of dependence parameters in multivariate families of distributions», *Biometrika*, vol. 82, n° 3, p. 543–552.
- GENEST, C. et J. MACKAY. 1986, «The joy of copulas : Bivariate distributions with uniform marginals», *The American Statistician*, vol. 40, n° 4, p. 280–283.
- GENEST, C. et J. NEŠLEHOVÁ. 2009, «Analytical proofs of classical inequalities between Spearman's  $\rho$  and Kendall's  $\tau$ », *Journal of statistical planning and inference*, vol. 139, n° 11, p. 3795–3798.
- GENEST, C., J.-F. QUESSY et B. RÉMILLARD. 2006, «Goodness-of-fit procedures for copula models based on the probability integral transformation», *Scandinavian Journal of Statistics*, vol. 33, n° 2, p. 337–366.
- GENEST, C. et B. RÉMILLARD. 2008, «Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models», dans *Annales de l'IHP Probabilités et statistiques*, vol. 44, p. 1096–1127.
- GENEST, C., B. RÉMILLARD et D. BEAUDOIN. 2009, «Goodness-of-fit tests for copulas : a review and a power study», *Insurance : Mathematics and economics*, vol. 44, n° 2, p. 199–213.
- GENEST, C. et L.-P. RIVEST. 1993, «Statistical inference procedures for bivariate Archimedean copulas», *Journal of the American statistical Association*, vol. 88, n° 423, p. 1034–1043.
- GENEST, C. et M. SCHERER. 2019, «The world of vines», *Dependence Modeling*, vol. 7, n° 1, p. 169–180.
- GHOSH, S. et S. G. HENDERSON. 2002, «Chessboard distributions and random vectors with specified marginals and covariance matrix», *Operations Research*, vol. 50, n° 5, p. 820–834.
- GIJBELS, I. et J. MIELNICZUK. 1990, «Estimating the density of a copula function», *Communications in Statistics-Theory and Methods*, vol. 19, n° 2, p. 445–464.
- GINI, C. 1914, *L'ammontare e la composizione della ricchezza delle nazioni*, vol. 62, Fratelli Bocca.
- GRØNNEBERG, S. et N. L. HJORT. 2008, «The copula information criterion», *Preprint series. Statistical Research Report*.
- GRUBER, L., C. CZADO et collab.. 2015, «Sequential Bayesian model selection of regular vine copulas», *Bayesian Analysis*, vol. 10, n° 4, p. 937–963.
- GUMBEL, E. J. 1960, «Bivariate exponential distributions», *Journal of the American Statistical Association*, vol. 55, n° 292, p. 698–707.
- GUNAWAN, S. et P. Y. PAPALAMBROS. 2007, «Reliability optimization with mixed continuous-discrete random variables and parameters», .

- HAFF, I. H., K. AAS et A. FRIGESSI. 2010, «On the simplified pair-copula construction : simply useful or too simplistic?», *Journal of Multivariate Analysis*, vol. 101, n° 5, p. 1296–1310.
- HANSEN, L. P. et K. J. SINGLETON. 1982, «Generalized instrumental variables estimation of nonlinear rational expectations models», *Econometrica : Journal of the Econometric Society*, p. 1269–1286.
- HEIDENREICH, N.-B., A. SCHINDLER et S. SPERLICH. 2013, «Bandwidth selection for kernel density estimation : a review of fully automatic selectors», *AStA Advances in Statistical Analysis*, vol. 97, n° 4, p. 403–433.
- HOEFFDING, W. 1940, «Masstabinvariante korrelationstheorie», *Schriften des Mathematischen Instituts und Instituts für Angewandte Mathematik der Universität Berlin*, vol. 5, p. 181–233.
- HOFERT, M. 2008, «Sampling Archimedean copulas», *Computational Statistics & Data Analysis*, vol. 52, n° 12, p. 5163–5174.
- HOFERT, M. et M. SCHERER. 2011, «CDO pricing with nested Archimedean copulas», *Quantitative Finance*, vol. 11, n° 5, p. 775–787.
- HOUGAARD, P. 1984, «Life table methods for heterogeneous populations : distributions describing the heterogeneity», *Biometrika*, vol. 71, n° 1, p. 75–83.
- HUANG, J. et N. JOJIC. 2010, «Maximum-likelihood learning of cumulative distribution functions on graphs», dans *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, JMLR Workshop and Conference Proceedings, p. 342–349.
- HUARD, D., G. EVIN et A.-C. FAVRE. 2006, «Bayesian copula selection», *Computational Statistics & Data Analysis*, vol. 51, n° 2, p. 809–822.
- HULL, J. C. et A. D. WHITE. 2004, «Valuation of a CDO and an  $n$ -th to default CDS without Monte Carlo simulation», *The Journal of Derivatives*, vol. 12, n° 2, p. 8–23.
- HULT, H. et F. LINDSKOG. 2002, «Multivariate extremes, aggregation and dependence in elliptical distributions», *Advances in Applied probability*, p. 587–608.
- JAYNES, E. T. 1957, «Information theory and statistical mechanics», *Physical review*, vol. 106, n° 4, p. 620.
- JOE, H. 1993, «Parametric families of multivariate distributions with given margins», *Journal of Multivariate analysis*, vol. 46, p. 262–282.
- JOE, H. 1996, «Families of  $m$ -variate distributions with given margins and  $m(m - 1)/2$  bivariate dependence parameters», *Lecture Notes-Monograph Series*, p. 120–141.
- JOE, H. 1997, *Multivariate models and multivariate dependence concepts*, CRC Press.
- JOE, H. 2014, *Dependence modeling with copulas*, CRC press.
- JOHNSON, N. L., S. KOTZ et N. BALAKRISHNAN. 1995, *Continuous univariate distributions, volume 2*, vol. 289, John Wiley & Sons.
- JOHNSON, R. A., D. W. WICHERN et collab.. 2002, *Applied multivariate statistical analysis*, vol. 5, Prentice Hall, Upper Saddle River, NJ.
- JORDANGER, L. A. et D. TJØSTHEIM. 2014, «Model selection of copulas : AIC versus a cross validation copula information criterion», *Statistics & Probability Letters*, vol. 92, p. 249–255.
- KAPUR, J. N. et H. K. KESAVAN. 1992, «Entropy optimization principles and their applications», dans *Entropy and energy dissipation in water resources*, Springer, p. 3–20.

- KAUERMANN, G., C. SCHELLHASE et D. RUPPERT. 2013, «Flexible copula density estimation with penalized hierarchical B-splines», *Scandinavian Journal of Statistics*, vol. 40, n° 4, p. 685–705.
- KENDALL, M. G. 1948, «Rank correlation methods», *Griffin*.
- KOLLER, D. et N. FRIEDMAN. 2009, *Probabilistic graphical models : principles and techniques*, MIT press.
- KOLMOGOROV, A. 1933, «Sulla determinazione empirica di una legge di distribuzione», *Inst. Ital. Attuari, Giorn.*, vol. 4, p. 83–91.
- KORTSCHAK, D. et H. ALBRECHER. 2009, «Asymptotic results for the sum of dependent non-identically distributed random variables», *Methodology and Computing in Applied Probability*, vol. 11, n° 3, p. 279–306.
- KOTZ, S., N. BALAKRISHNAN et N. L. JOHNSON. 2004, *Continuous multivariate distributions, Volume 1 : Models and applications*, John Wiley & Sons.
- KRUPSKII, P. et H. JOE. 2013, «Factor copula models for multivariate data», *Journal of Multivariate Analysis*, vol. 120, p. 85–101.
- KRUPSKII, P. et H. JOE. 2015, «Tail-weighted measures of dependence», *Journal of Applied Statistics*, vol. 42, n° 3, p. 614–629.
- KUROWICKA, D. 2010, «Optimal truncation of vines», dans *Dependence modeling : Vine copula handbook*, World Scientific, p. 233–247.
- KUROWICKA, D. et R. COOKE. 2001, «Conditional, partial and rank correlation for the elliptical copula; dependence modelling in uncertainty analysis», dans *Proceedings ESREL*, vol. 1, Politecnico di Torino Torino, p. 1795–1802.
- KUROWICKA, D. et R. M. COOKE. 2006, *Uncertainty analysis with high dimensional dependence modelling*, John Wiley & Sons.
- LAI, C. et M. XIE. 2003, «Concepts of stochastic dependence in reliability analysis», dans *Handbook of reliability engineering*, Springer, p. 141–156.
- LAURENT, J.-P. et J. GREGORY. 2005, «Basket default swaps, CDOs and factor copulas», *Journal of risk*, vol. 7, n° 4, p. 103–122.
- LAURITZEN, S. L. 1996, *Graphical models*, vol. 17, Clarendon Press.
- LAURO, C. N. et F. PALUMBO. 2000, «Principal component analysis of interval data : a symbolic data analysis approach», *Computational statistics*, vol. 15, n° 1, p. 73–87.
- LAWRANCE, A. 1976, «On conditional and partial correlation», *The American Statistician*, vol. 30, n° 3, p. 146–149.
- LEE, D., H. JOE et P. KRUPSKII. 2018, «Tail-weighted dependence measures with limit being the tail dependence coefficient», *Journal of Nonparametric Statistics*, vol. 30, n° 2, p. 262–290.
- LEHMANN, E. L. et G. CASELLA. 2006, *Theory of point estimation*, Springer Science & Business Media.
- LI, H. 2008, «Tail dependence comparison of survival Marshall–Olkin copulas», *Methodology and Computing in Applied Probability*, vol. 10, n° 1, p. 39–54.
- LIEBSCHER, E. 2008, «Construction of asymmetric multivariate copulas», *Journal of Multivariate Analysis*, vol. 99, n° 10, p. 2234–2250.

- LIU, P.-L. et A. DER KIUREGHIAN. 1986, «Multivariate distribution models with prescribed marginals and covariances», *Probabilistic Engineering Mechanics*, vol. 1, n° 2, p. 105–112.
- LOADER, C. R. et collab.. 1996, «Local likelihood density estimation», *The Annals of Statistics*, vol. 24, n° 4, p. 1602–1618.
- MACLEISH, D. L. et C. G. SMALL. 1988, *The theory and applications of statistical inference functions*, Springer.
- MARSHALL, A. W. et I. OLKIN. 1988, «Families of multivariate distributions», *Journal of the American statistical association*, vol. 83, n° 403, p. 834–841.
- MASSEY JR, F. J. 1951, «The Kolmogorov-Smirnov test for goodness of fit», *Journal of the American statistical Association*, vol. 46, n° 253, p. 68–78.
- MAZO, G. 2014, *Construction et estimation de copules en grande dimension*, thèse de doctorat, Université de Grenoble.
- MAZO, G., S. GIRARD et F. FORBES. 2015, «A class of multivariate copulas based on products of bivariate copulas», *Journal of Multivariate Analysis*, vol. 140, p. 363–376.
- MCNEIL, A. J. 2008, «Sampling nested Archimedean copulas», *Journal of Statistical Computation and Simulation*, vol. 78, n° 6, p. 567–581.
- MCNEIL, A. J., R. FREY et P. EMBRECHTS. 2015, *Quantitative risk management : concepts, techniques and tools-revised edition*, Princeton University Press.
- MCNEIL, A. J., J. NEŠLEHOVÁ et collab.. 2009, «Multivariate Archimedean copulas,  $d$ -monotone functions and  $\ell_1$ -norm symmetric distributions», *The Annals of Statistics*, vol. 37, n° 5B, p. 3059–3097.
- MEISSNER, G. 2013, *Correlation Risk Modeling and Management : An Applied Guide including the Basel III Correlation Framework-With Interactive Models in Excel/VBA*, John Wiley & Sons.
- MIN, A. et C. CZADO. 2010, «Bayesian inference for multivariate copulas using pair-copula constructions», *Journal of Financial Econometrics*, vol. 8, n° 4, p. 511–546.
- MORALES-NÁPOLES, O. 2010, «Counting vines», dans *Dependence modeling : Vine copula handbook*, World Scientific, p. 189–218.
- NADARAJAH, S., E. AFUECHETA et S. CHAN. 2018, «A compendium of copulas», *Statistica*, vol. 77, n° 4, p. 279–328.
- NAGLER, T. et C. CZADO. 2016, «Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas», *Journal of Multivariate Analysis*, vol. 151, p. 69–89.
- NAGLER, T., C. SCHELLHASE et C. CZADO. 2017, «Nonparametric estimation of simplified vine copula models : comparison of methods», *Dependence Modeling*, vol. 5, n° 1, p. 99–120.
- NELSEN, R. B. 1997, «Dependence and order in families of Archimedean copulas», *Journal of Multivariate Analysis*, vol. 60, n° 1, p. 111–122.
- NELSEN, R. B. 2002, «Concordance and copulas : A survey», dans *Distributions with given marginals and statistical modelling*, Springer, p. 169–177.
- NELSEN, R. B. 2007, *An introduction to copulas*, Springer Science & Business Media.
- OAKES, D. 1982, «A model for association in bivariate survival data», *Journal of the Royal Statistical Society : Series B (Methodological)*, vol. 44, n° 3, p. 414–422.

- OKHRIN, O., Y. OKHRIN et W. SCHMID. 2013, «On the structure and estimation of hierarchical Archimedean copulas», *Journal of Econometrics*, vol. 173, n° 2, p. 189–204.
- PARK, B. U. et J. S. MARRON. 1990, «Comparison of data-driven bandwidth selectors», *Journal of the American Statistical Association*, vol. 85, n° 409, p. 66–72.
- PEARSON, K. 1920, «Notes on the history of correlation», *Biometrika*, vol. 13, n° 1, p. 25–45.
- PLISCHKE, E. et E. BORGONOVO. 2015, «Copula-based sensitivity measures of computer experiments», *Safety and Reliability of Complex Engineered Systems*.
- PORTET, S. 2020, «A primer on model selection using the Akaike information criterion», *Infectious Disease Modelling*, vol. 5, p. 111–128.
- ROBERT, C. 2007, *The Bayesian choice : from decision-theoretic foundations to computational implementation*, Springer Science & Business Media.
- RUBIN, D. B. 1976, «Inference and missing data», *Biometrika*, vol. 63, n° 3, p. 581–592.
- SANCETTA, A. et S. SATCHELL. 2004, «The Bernstein copula and its applications to modeling and approximations of multivariate distributions», *Econometric theory*, vol. 20, n° 3, p. 535–562.
- DOS SANTOS SILVA, R. et H. F. LOPES. 2008, «Copula, marginal distributions and model selection : a Bayesian note», *Statistics and Computing*, vol. 18, n° 3, p. 313–320.
- SAPORTA, G. 2006, *Probabilités, analyse des données et statistique*, Editions Technip.
- SAVU, C. et M. TREDE. 2010, «Hierarchies of Archimedean copulas», *Quantitative Finance*, vol. 10, n° 3, p. 295–304.
- SCARSINI, M. 1984, «On measures of concordance.», *Stochastica*, vol. 8, n° 3, p. 201–218.
- SCHMIDT, R. 2002, «Tail dependence for elliptically contoured distributions», *Mathematical Methods of Operations Research*, vol. 55, n° 2, p. 301–327.
- SCHWARZ, G. et collab.. 1978, «Estimating the dimension of a model», *The annals of statistics*, vol. 6, n° 2, p. 461–464.
- SCOTT, D. W. 1991, «Feasibility of multivariate density estimates», *Biometrika*, vol. 78, n° 1, p. 197–205.
- SEGERS, J. et collab.. 2012, «Asymptotics of empirical copula processes under non-restrictive smoothness assumptions», *Bernoulli*, vol. 18, n° 3, p. 764–782.
- SHEATHER, S. J. et M. C. JONES. 1991, «A reliable data-based bandwidth selection method for kernel density estimation», *Journal of the Royal Statistical Society : Series B (Methodological)*, vol. 53, n° 3, p. 683–690.
- SHIH, J. H. et T. A. LOUIS. 1995, «Inferences on the association parameter in copula models for bivariate survival data», *Biometrics*, p. 1384–1399.
- SIBUYA, M. 1960, «Bivariate extreme statistics, I», *Annals of the Institute of Statistical Mathematics*, vol. 11, n° 3, p. 195–210.
- SILVAPULLE, P., G. KIM, M. J. SILVAPULLE et collab.. 2004, «Robustness of a semiparametric estimator of a copula», dans *Econometric Society 2004 Australasian Meetings*, 317, Econometric Society.
- SILVERMAN, B. W. 1986, *Density estimation for statistics and data analysis*, vol. 26, CRC press.

- SKLAR, A. 1959, «Fonctions de répartition à  $n$  dimensions et leurs marges», *Publ. inst. statist. univ. Paris*, vol. 8, p. 229–231.
- SMIRNOFF, N. 1939, «Sur les écarts de la courbe de distribution empirique», *Matematicheskii Sbornik*, vol. 48, n° 1, p. 3–26.
- SPANHEL, F. et M. S. KURZ. 2015, «Simplified vine copula models : Approximations based on the simplifying assumption», *arXiv preprint arXiv :1510.06971*, vol. 715.
- SPEARMAN, C. 1961, «The proof and measurement of association between two things.».
- SPIEGELHALTER, D. J., N. G. BEST, B. P. CARLIN et A. VAN DER LINDE. 2002, «Bayesian measures of model complexity and fit», *Journal of the royal statistical society : Series b (statistical methodology)*, vol. 64, n° 4, p. 583–639.
- SPIRITES, P. et C. GLYMOUR. 1991, «An algorithm for fast recovery of sparse causal graphs», *Social science computer review*, vol. 9, n° 1, p. 62–72.
- STOEBER, J., H. JOE et C. CZADO. 2013, «Simplified pair-copula constructions : limitations and extensions», *Journal of Multivariate Analysis*, vol. 119, p. 101–118.
- STUTE, W. 1984, «The oscillation behavior of empirical processes : the multivariate case», *The Annals of Probability*, p. 361–379.
- TAYLOR, M. 2007, «Multivariate measures of concordance», *Annals of the Institute of Statistical Mathematics*, vol. 59, n° 4, p. 789–806.
- TSUKAHARA, H. 2005, «Semiparametric estimation in copula models», *Canadian Journal of Statistics*, vol. 33, n° 3, p. 357–375.
- TSYBAKOV, A. B. 2008, *Introduction to nonparametric estimation*, Springer Science & Business Media.
- VAN DER VAART, A. W. 2000, *Asymptotic statistics*, vol. 3, Cambridge university press.
- WASSERMAN, L. 2004, «Parametric inference», dans *All of Statistics*, Springer, p. 119–148.
- WASSERMAN, L. 2006, *All of nonparametric statistics*, Springer Science & Business Media.
- WEN, K. et X. WU. 2015a, «An improved transformation-based kernel estimator of densities on the unit interval», *Journal of the American Statistical Association*, vol. 110, n° 510, p. 773–783.
- WEN, K. et X. WU. 2015b, «Transformation-kernel estimation of the copula density», *Preprint*: <http://agecon2.tamu.edu/people/faculty/wu-ximing/agecon2/public/copula.pdf>.
- WOLFF, E. F. 1980, « $n$ -dimensional measures of dependence.», *Stochastica*, vol. 4, n° 3, p. 175–188.
- XU, J. J. 1996, *Statistical modelling and inference for multivariate and longitudinal discrete response data*, thèse de doctorat, University of British Columbia.
- YANG, I.-T. et Y.-H. HSIEH. 2011, «Reliability-based design optimization with discrete design variables and non-smooth performance functions : AB-PSO algorithm», *Automation in construction*, vol. 20, n° 5, p. 610–619.
- ZHANG, J., H. HUANG, C. JUANG et W. SU. 2014, «Geotechnical reliability analysis with limited data : consideration of model selection uncertainty», *Engineering geology*, vol. 181, p. 27–37.

ZHANG, J. et M. D. SHIELDS. 2018, «The effect of prior probabilities on quantification and propagation of imprecise probabilities resulting from small datasets», *Computer Methods in Applied Mechanics and Engineering*, vol. 334, p. 483–506.

ZSOHAR, P. 2012, «Short introduction to the generalized method of moments», *Hungarian Statistical Review*, vol. 16, p. 150–170.

## Chapitre 2

# Estimation de probabilités d'évènements rares pour l'analyse de fiabilité

### Sommaire

---

<b>2.1 Introduction</b> . . . . .	<b>78</b>
<b>2.2 Analyse de fiabilité déterministe</b> . . . . .	<b>79</b>
2.2.1 Transformations isoprobabilistes . . . . .	79
2.2.2 Techniques d'approximation de l'état limite . . . . .	82
<b>2.3 Analyse de fiabilité stochastique</b> . . . . .	<b>85</b>
2.3.1 Algorithme de Monte-Carlo naïf . . . . .	85
2.3.2 Échantillonnage préférentiel . . . . .	86
2.3.3 Améliorations et variantes de l'échantillonnage préférentiel . . . . .	93
2.3.4 Autres méthodes d'analyse de fiabilité stochastique . . . . .	97
<b>2.4 Conclusion</b> . . . . .	<b>99</b>

---

## 2.1 Introduction

La défaillance d'un système ou d'une installation est un évènement dont on cherche à quantifier le risque d'occurrence grâce à des outils probabilistes. L'analyse de fiabilité est la démarche scientifique qui conduit à l'estimation d'une probabilité de défaillance. L'évènement redouté est décrit par un critère de défaillance qui coïncide souvent avec le dépassement d'un seuil limite  $T$  par une variable d'intérêt  $Y$  de loi  $\mu_Y$  inconnue. Pour appréhender le comportement aléatoire de  $Y$ , on utilise un code de calcul déterministe  $\phi$  dont les entrées  $\mathbf{X}$  sont des paramètres incertains mais mesurables. La probabilité de défaillance est la quantité d'intérêt du problème :

$$P_f := \mathbb{P}(Y > T) \quad \text{avec} \quad Y = \phi(\mathbf{X}) . \quad (2.1)$$

En fiabilité,  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$  est appelée la fonction de performance, ou plus prosaïquement la “boîte noire” pour insister sur l'opacité des opérations qu'elle exécute. Le domaine de défaillance est le sous-ensemble  $D_f \subseteq \mathbb{R}^d$  qui engendre des sorties  $y$  strictement supérieures au seuil  $T$  :

$$D_f := \{\mathbf{x} \in \mathbb{R}^d : \phi(\mathbf{x}) > T\} = \phi^{-1}(\{y \in \mathbb{R} : y > T\}) . \quad (2.2)$$

Le complémentaire de  $D_f$  est appelé le domaine de sûreté. La courbe  $\phi(\mathbf{x}) = T$  qui sépare les deux domaines est connue sous le nom d'état limite. Par ailleurs, il est supposé que la loi  $\mu_{\mathbf{X}}$  des entrées du problème admet une densité  $f_{\mathbf{X}}$ . Dans le contexte de cette thèse, on a expliqué qu'elle était inconnue et qu'elle devait être estimée à partir d'un échantillon  $\mathbf{x}_{\text{obs}} \in \mathbb{R}^{N \times d}$  et à l'aide de l'un des algorithmes d'apprentissage statistique présentés au Chapitre 1. Dans le double but d'alléger les notations et de rester aussi proche que possible de la littérature, les différents algorithmes de fiabilité sont étudiés dans un cadre où  $f_{\mathbf{X}}$  est connue. Tout peut évidemment être adapté à une densité de type  $\hat{f}_{\mathbf{X}}(\cdot | \mathcal{M}, \mathbf{x}_{\text{obs}})$  à condition que celle-ci respecte les mêmes hypothèses que  $f_{\mathbf{X}}$ . En vertu du Théorème 1.2.3, on peut écrire que :

$$P_f = \mathbb{P}(Y > T) = \int_T^{+\infty} f_Y(y) dy \quad (2.3)$$

$$= \mathbb{P}(\phi(\mathbf{X}) > T) = \mathbb{E}[\mathbb{1}_{D_f}(\mathbf{X})] = \int_{\mathbb{R}^d} \mathbb{1}_{D_f}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} . \quad (2.4)$$

Le calcul d'une probabilité de défaillance  $P_f$  consiste donc à intégrer la queue de distribution supérieure de la densité  $f_Y$  inconnue. Il peut aussi être vu comme l'intégration de la densité conjointe  $f_{\mathbf{X}}$  sur le domaine de défaillance  $D_f$  dont la délimitation est inconnue et potentiellement hautement non-linéaire. Dans le premier cas, la méconnaissance de  $f_Y$  empêche toute velléité d'intégration numérique avec les méthodes de quadrature décrites dans [DAVIS et RABINOWITZ \[2007\]](#). Dans le second cas, comme l'indicatrice  $\mathbb{1}_{D_f}$  et la densité  $f_{\mathbf{X}}$  sont évaluables en tout point, rien n'empêche à première vue le déploiement d'une technique d'intégration multidimensionnelle. Toutefois, comme le soulignent [HINRICHS et collab. \[2014\]](#), le nombre de points requis pour atteindre un niveau de précision donné est une fonction exponentielle du nombre  $d$  d'entrées. Ce phénomène est connu sous le nom de fléau (ou malédiction) de la dimension. De plus, sachant que chaque appel à  $\phi$  est coûteux et que les points choisis à l'extérieur de  $D_f$  n'apportent rien, on comprend que cette approche est condamnée.

L'analyse de fiabilité est une sous-branche assez spécifique du domaine de la propagation des incertitudes. Pour un budget limité d'appels à  $\phi$ , il est illusoire de penser pouvoir construire une estimation très précise de  $f_Y$  sur tout l'intervalle couvrant les variations de  $Y$ . La fiabilité a ceci de spécifique qu'elle cherche à estimer une quantité qui ne dépend que d'une portion restreinte de la densité  $f_Y$ . Le point commun des différents algorithmes de fiabilité réside dans le fait qu'ils cherchent à mettre en place des stratégies pour que le budget de calcul serve prioritairement à enrichir la connaissance de la queue de distribution supérieure de  $f_Y$ . Dans l'idéal, tout calcul de type  $y = \phi(\mathbf{x})$  doit impliquer un point  $\mathbf{x} \in \mathbb{R}^d$  situé au sein du domaine de défaillance  $D_f$  ou dans son voisinage proche. Comme on ignore tout de  $D_f$  au démarrage de la procédure d'analyse

de fiabilité, la plupart des algorithmes sont itératifs au sens où ils cherchent à explorer l'espace d'échantillonnage  $\mathbb{R}^d$  afin de se rapprocher progressivement de  $D_f$ .

Dans la Section 2.2, on présente succinctement les principales méthodes déterministes qui ont été développées dans le cadre de la fiabilité des structures. Pour un exposé plus complet à ce sujet, on renvoie aux ouvrages de [DITLEVSEN et MADSEN \[1996\]](#), [LEMAIRE et collab. \[2009\]](#) ou [MELCHERS et BECK \[2018\]](#). Après avoir transformé le problème initial pour que les entrées soient gaussiennes et décorréelées, ces méthodes cherchent à identifier certains points caractéristiques de l'état limite puis émettent des hypothèses sur la nature géométrique de la frontière séparant les domaines de sûreté et de défaillance. Des formules analytiques permettent ensuite d'obtenir une estimation de  $P_f$ . Face aux insuffisances constatées lors de l'emploi des méthodes déterministes, il est souvent préférable de se tourner vers des méthodes stochastiques. Il s'agit d'algorithmes de simulation qui reposent sur des tirages aléatoires et qui sont d'autant plus performants que les particules simulées sont concentrées à la frontière de  $D_f$ . Si la méthode naïve de Monte-Carlo ne répond pas aux exigences du cahier des charges d'une analyse de fiabilité, diverses techniques de réduction de la variance (échantillonnages dits préférentiel, séquentiel, stratifié, multiple, en ligne ou directionnel) permettent de construire des estimateurs précis et robustes même lorsque la dimension  $d$  du problème est grande ou que la probabilité de défaillance  $P_f$  est très petite au regard du budget de simulation accordé. On précise que la Section 2.3 s'attache uniquement à présenter les algorithmes d'échantillonnage préférentiel qui seront utilisés aux Chapitres 4 et 5 au sein des procédures d'analyse de sensibilité fiabiliste (ASF). Pour disposer d'un plus large panorama des méthodes d'analyse de fiabilité stochastiques, on peut se reporter aux ouvrages de [RUBINO et TUFFIN \[2009\]](#) ou de [RUBINSTEIN et KROESE \[2016\]](#).

## 2.2 Analyse de fiabilité déterministe

Les méthodes présentées dans cette section sont toutes articulées autour de deux étapes fondamentales. La première étape repose sur l'utilisation d'une transformation isoprobabiliste qui agit sur le vecteur initial  $\mathbf{X}$  pour fabriquer un vecteur gaussien centré réduit  $\mathbf{W}$ . Ce changement de variable n'affecte pas la valeur de  $P_f$  mais induit une modification du paramétrage de l'état limite dont l'équation cartésienne s'écrit désormais sous une forme standard. La Section 2.2.1 explique comment choisir la transformation isoprobabiliste en fonction du modèle de distribution auquel appartient  $f_{\mathbf{X}}$ . La seconde étape consiste à construire une approximation de l'état limite obéissant à une configuration géométrique élémentaire. La probabilité de défaillance associée à cette nouvelle frontière obéit alors à une formule analytique et sert d'estimation de la probabilité de défaillance théorique. La Section 2.2.2 présente plusieurs méthodes d'approximation de l'état limite et précise les formules qui en découlent.

### 2.2.1 Transformations isoprobabilistes

Soit  $\mathcal{T} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  un difféomorphisme, c'est-à-dire une bijection continûment différentiable. Pour que  $\mathcal{T}$  soit considérée comme la transformation isoprobabiliste associée à  $\mathbf{X}$ , son action doit pouvoir être résumée de la manière suivante :

$$\begin{aligned} \mathcal{T} : \quad \mathbf{x} &\longmapsto \mathbf{w} \\ \mathbf{X} \sim \mu_{\mathbf{X}} &\longmapsto \mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d) \end{aligned} \quad (2.5)$$

On dit que  $\mathbf{X}$  vit dans l'espace physique par opposition à  $\mathbf{W}$  qui vit dans l'espace normal standard. La transformation isoprobabiliste inverse  $\mathcal{T}^{-1}$  permet le retour de  $\mathbf{W}$  à  $\mathbf{X}$ . Dans la suite, la densité et la fonction de répartition de la loi unidimensionnelle  $\mathcal{N}(0, 1)$  sont notées  $\varphi$  et  $\Phi$  tandis que la densité  $f_{\mathbf{W}}$  du vecteur gaussien  $d$ -dimensionnel centré réduit  $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$  est notée  $\varphi_d$ . Une

expression intégrale de  $P_f$  dans l'espace normal standard s'obtient avec le Théorème 1.2.3 :

$$P_f = \mathbb{P}(\phi(\mathbf{X}) > T) = \mathbb{P}([\phi \circ \mathcal{T}^{-1}](\mathbf{W}) > T) = \int_{\mathbb{R}^d} \mathbb{1}_{D_f^{\mathcal{N}}}(\mathbf{w}) \varphi_d(\mathbf{w}) d\mathbf{w} \quad (2.6)$$

avec  $D_f^{\mathcal{N}} := \{\mathbf{w} \in \mathbb{R}^d : H(\mathbf{w}) > T\}$  et  $H := \phi \circ \mathcal{T}^{-1}$

Lorsqu'on passe dans l'espace normal standard,  $D_f^{\mathcal{N}}$  et  $H$  se substituent à  $D_f$  et  $\phi$ . Reste à savoir comment construire la transformation  $\mathcal{T}$  à partir de la connaissance de  $\mu_{\mathbf{X}}$ .

### 2.2.1.1 Cas d'un vecteur à variables indépendantes

On suppose ici que le vecteur  $\mathbf{X}$  est constitué de variables indépendantes. Pour davantage de commodité, on ne considère que des situations où toutes les fonctions de répartition  $F_{X_j}$  sont inversibles. Sous cette hypothèse, il suffit alors de poser :

$$\forall 1 \leq j \leq d, \quad W_j = \Phi^{-1}(F_{X_j}(X_j)) \quad \text{ou} \quad X_j = F_{X_j}^{-1}(\Phi(W_j)) \quad , \quad (2.7)$$

pour définir les transformations  $\mathcal{T}$  et  $\mathcal{T}^{-1}$ . Cette transformation élémentaire connue sous le nom de *Probability Integral Transform* (PIT) repose sur le fait que les variables  $F_{X_j}(X_j)$  et  $\Phi(W_j)$  sont identiquement distribuées suivant une loi uniforme sur l'intervalle  $[0, 1]$ .

### 2.2.1.2 Cas d'un vecteur gaussien

On suppose désormais que  $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . On choisit d'exprimer la matrice de covariance sous la forme  $\boldsymbol{\Sigma} = \boldsymbol{\Delta} \boldsymbol{\rho} \boldsymbol{\Delta}$  avec  $\boldsymbol{\rho}$  la matrice de corrélation et  $\boldsymbol{\Delta}$  la matrice diagonale des écarts-types. La factorisation de Cholesky de la matrice de corrélation donne  $\boldsymbol{\rho} = \mathbf{L} \mathbf{L}^T$  avec  $\mathbf{L}$  une matrice triangulaire inférieure. On peut alors obtenir  $\mathbf{W}$  avec :

$$\mathbf{W} = \mathbf{L}^{-1} \boldsymbol{\Delta}^{-1} (\mathbf{X} - \boldsymbol{\mu}) = \mathbf{L}^{-1} \tilde{\mathbf{X}} \quad \text{où} \quad \tilde{\mathbf{X}} := \boldsymbol{\Delta}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \quad . \quad (2.8)$$

La transformation linéaire  $\boldsymbol{\Delta}^{-1} (\mathbf{X} - \boldsymbol{\mu})$  permet de passer des variables  $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$  aux variables  $\tilde{X}_j \sim \mathcal{N}(0, 1)$ . Ce n'est rien d'autre qu'un cas particulier de l'équation (2.7) qui établit PIT dans le cas général. Le vecteur  $\tilde{\mathbf{X}}$  suit une loi  $\mathcal{N}(\mathbf{0}, \boldsymbol{\rho})$  et le produit matriciel  $\mathbf{L}^{-1} \tilde{\mathbf{X}}$  élimine la structure de corrélation restante.

### 2.2.1.3 Transformation de Nataf

La transformation de **NATAF** [1962] a été introduite par **LIU et DER KIUREGHIAN** [1986] en fiabilité des structures dans un contexte où  $f_{\mathbf{X}}$  n'est pas complètement spécifiée. Les lois marginales  $\mu_{X_j}$  sont fournies mais l'information sur la structure de dépendance se limite à la connaissance de la matrice de corrélation  $\boldsymbol{\rho}$ . Une première étape consiste à définir un vecteur  $\tilde{\mathbf{X}}$  dont les lois marginales sont des gaussiennes centrées réduites :

$$\tilde{\mathbf{X}} := \mathcal{T}_A(\mathbf{X}) = [\tilde{X}_j]_{1 \leq j \leq d} \quad \text{avec} \quad \tilde{X}_j = T_j(X_j) := \Phi^{-1}(F_{X_j}(X_j)) \quad . \quad (2.9)$$

La pierre angulaire de cette méthode est l'hypothèse selon laquelle  $\tilde{\mathbf{X}}$  suit une loi gaussienne  $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_0)$  où  $\boldsymbol{\Sigma}_0 = \boldsymbol{\rho}_0$  est une matrice de corrélation. Dans le cas général de lois  $\mu_{X_j}$  non-gaussiennes, les transformations  $T_j$  ne sont pas linéaires. Par ailleurs, on rappelle que les coefficients de corrélation de **PEARSON** [1920] ne sont pas invariants lorsque des transformations  $T_j$  continues, strictement croissantes mais non-linéaires sont appliquées aux variables initiales. De ce fait, la matrice de corrélation  $\boldsymbol{\rho}$  initialement fournie ne coïncide pas avec la matrice  $\boldsymbol{\rho}_0$  du vecteur gaussien  $\tilde{\mathbf{X}}$ . Pour accéder au coefficient  $r_{ij}$  de  $\boldsymbol{\rho}_0$ , on le fait intervenir dans la définition du coefficient  $\rho_{ij}$  de  $\boldsymbol{\rho}$  :

$$\begin{aligned} \rho_{ij} &= \mathbb{E} \left[ \left( \frac{X_i - \mu_j}{\sigma_i} \right) \left( \frac{X_j - \mu_j}{\sigma_j} \right) \right] = \frac{1}{\sigma_i \sigma_j} \left( \mathbb{E} \left[ T_i^{-1}(W_i) T_j^{-1}(W_j) \right] - \mu_i \mu_j \right) \\ &= \frac{1}{\sigma_i \sigma_j} \int_{\mathbb{R}^2} T_i^{-1}(w_i) T_j^{-1}(w_j) \varphi_2(\mathbf{w} | r_{ij}) d\mathbf{w} = \mathcal{V}_{ij}(r_{ij}) \quad . \quad (2.10) \end{aligned}$$

En résolvant l'équation intégrale  $\rho_{ij} = \mathcal{V}_{ij}(r_{ij})$  pour chacun des coefficients non diagonaux, on accède à  $\boldsymbol{\rho}_0$ . **LEBRUN [2013]** mentionne deux principales difficultés dans l'obtention de  $\boldsymbol{\rho}_0$ .

- (a) L'une des équations intégrales peut ne pas avoir de solution, en particulier lorsque  $\rho_{ij}$  est proche des bornes de l'intervalle  $[-1, 1]$ .
- (b) Comme la résolution numérique des  $p = \frac{1}{2}d(d-1)$  équations intégrales est effectuée en parallèle, rien ne garantit que  $\boldsymbol{\rho}_0$  soit finalement une matrice définie positive.

Une alternative est proposée dans **LIU et DER KIUREGHIAN [1986]** où des formules permettant d'estimer  $\boldsymbol{\rho}_0$  sont établies pour plusieurs familles paramétriques  $f_{\mathbf{X}}(\cdot | \mathcal{F}, \boldsymbol{\theta})$ . Une fois que  $\boldsymbol{\rho}_0$  est déterminée, il suffit de prendre  $\mathbf{W} = \mathcal{T}_{\mathbf{B}}(\tilde{\mathbf{X}}) = \mathbf{L}_0^{-1}\tilde{\mathbf{X}}$  avec  $\mathbf{L}_0$  la matrice triangulaire inférieure issue de la factorisation de Cholesky de  $\boldsymbol{\rho}_0$ . À partir de ce qui vient d'être fait, la transformation inverse  $\mathcal{T}^{-1}$  peut être décomposée comme suit :

$$\mathcal{T}^{-1}: \mathbf{W} \sim \varphi_d \xrightarrow{\mathcal{T}_{\mathbf{B}}^{-1}} \tilde{\mathbf{X}} \sim \varphi_d(\cdot | \boldsymbol{\rho}_0) \xrightarrow{\mathcal{T}_{\mathbf{A}}^{-1}} \mathbf{X} . \quad (2.11)$$

**LEBRUN et DUTFOY [2009c]** font remarquer que la copule de  $\tilde{\mathbf{X}}$  est de la forme  $c_{\mathcal{N}}(\cdot | \boldsymbol{\rho}_0)$ . De par le résultat d'invariance du Théorème 1.2.6, la transformation  $\mathcal{T}_{\mathbf{A}}^{-1}$  laisse cette copule inchangée. Par conséquent, appliquer sans précaution la transformation de Nataf revient à supposer que la copule  $c_{\mathbf{U}}$  de  $\mathbf{X}$  est gaussienne. Ce choix peut avoir des conséquences importantes sur l'estimation de  $P_f$  car la copule gaussienne implique une indépendance asymptotique des queues de distribution qui est mise en lumière par le fait que  $\lambda_{\mathbf{L}} = \lambda_{\mathbf{U}} = 0$ . Pour éviter d'encourir le risque d'une sous-estimation de la probabilité de défaillance, **LEBRUN et DUTFOY [2009b]** généralisent la transformation pour qu'elle soit exacte dès lors que  $\tilde{\mathbf{X}}$  suit une distribution elliptique, ce qui inclut par exemple la loi de Student multidimensionnelle dont les queues sont plus lourdes.

#### 2.2.1.4 Transformation de Rosenblatt

L'idée de la transformation de **ROSENBLATT [1952]** vient d'un résultat élémentaire qui indique que les variables  $V_1 := F_{X_1}(X_1)$  et  $V_2 := F_{X_2|X_1}(X_2 | X_1)$  sont uniformes et indépendantes quelle que soit la loi de la paire  $(X_1, X_2)$ . La généralisation à un vecteur  $\mathbf{X}$  de dimension quelconque repose sur l'application successive des fonctions de répartition conditionnelles :

$$\mathcal{T}: \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \mapsto \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{bmatrix} \quad \text{avec} \quad \begin{cases} v_1 = F_{X_1}(x_1) \\ v_2 = F_{X_2|X_1}(x_2 | x_1) \\ \vdots \\ v_d = F_{X_d|X_{d-1}, \dots, X_1}(x_d | x_{d-1}, \dots, x_1) \end{cases} . \quad (2.12)$$

$$\mathbf{X} \sim f_{\mathbf{X}} \quad \mapsto \quad \mathbf{V} \sim \mathcal{U}([0, 1])^{\otimes d}$$

Le vecteur  $\mathbf{V}$  aurait pu être fabriqué en appliquant la transformation  $\mathcal{T}$  ci-dessus à un vecteur  $\tilde{\mathbf{X}}$  obtenu par permutation des composantes de  $\mathbf{X}$ . Ainsi, il existe  $d!$  façons différentes de construire  $\mathcal{T}$  et chacune implique son propre séquençage des opérations de conditionnement. L'idée est de choisir un vecteur  $\tilde{\mathbf{X}}$  pour lequel les fonctions de répartition conditionnelles s'expriment le plus simplement. Dans certains cas, par exemple lorsque la densité  $f_{\mathbf{X}}$  inclut une copule gaussienne, l'ordre de conditionnement n'a aucune importance. Pour ce cas précis, il est d'ailleurs montré dans **LEBRUN et DUTFOY [2009a]** que les transformations de Nataf et Rosenblatt sont équivalentes. En revanche, pour des modèles de distribution plus sophistiqués, l'ordre de conditionnement peut être primordial dans la mise en œuvre numérique de  $\mathcal{T}$ . Lorsque la structure de dépendance de  $\mathbf{X}$  est décrite par une copule R-vigne  $c_{\text{vg}}(\cdot | \mathbf{T}_{\text{vg}}, \mathbf{F}_{\text{vg}}, \boldsymbol{\theta}_{\text{vg}})$ , **AAS et collab. [2009]** ont démontré qu'il fallait s'adapter à la séquence d'arbres  $\mathbf{T}_{\text{vg}}$  pour que les fonctions de répartition conditionnelles puissent s'écrire à partir des h-fonctions des copules bidimensionnelles impliquées dans la factorisation. De manière plus générale, les fonctions de répartition conditionnelles  $F_{X_k|X_{k-1}, \dots, X_1}$  peuvent être obtenues à partir de la fonction de répartition  $C_{\mathbf{U}}$  de la copule par différentiation.

En notant  $C_{1\dots j}$  la fonction de répartition de la copule modélisant la dépendance des  $j$  premières composantes de  $\mathbf{X}$ , il vient :

$$F_{X_k|X_{k-1},\dots,X_1}(x_k | x_{k-1}, \dots, x_1) = \frac{\frac{\partial^{k-1} C_{1\dots k}}{\partial u_1 \dots \partial u_{k-1}}(F_{X_1}(x_1), \dots, F_{X_k}(x_k))}{\frac{\partial^{k-1} C_{1\dots k-1}}{\partial u_1 \dots \partial u_{k-1}}(F_{X_1}(x_1), \dots, F_{X_{k-1}}(x_{k-1}))} . \quad (2.13)$$

Le calcul formel ou la différentiation numérique permettent en général d'évaluer facilement les fonctions de répartition conditionnelles. La construction de la transformation inverse  $\mathcal{T}^{-1}$  n'est par contre pas immédiate comme le souligne BOURINET [2018]. Dans le cas d'une copule R-vigne  $c_{\text{vg}}(\cdot | \mathbf{T}_{\text{vg}}, \mathbf{F}_{\text{vg}}, \boldsymbol{\theta}_{\text{vg}})$ , l'imbrication des  $h$ -fonctions les unes dans les autres permet une inversion analytique de  $\mathcal{T}$ .

## 2.2.2 Techniques d'approximation de l'état limite

Il est désormais considéré qu'une transformation isoprobabiliste  $\mathcal{T}$  assure le passage entre l'espace physique et l'espace normal standard. Les méthodes visant à estimer la probabilité de défaillance  $P_f$  en s'appuyant sur le changement de variable  $\mathbf{W} = \mathcal{T}(\mathbf{X})$  sont apparues dans les travaux de CORNELL [1969] puis se sont largement diffusées en fiabilité des structures. Les algorithmes FORM et SORM rendent une estimation de  $P_f$  après avoir intégré la densité  $\varphi_d$  sur un domaine  $\mathring{D}_f^{\mathcal{N}}$ . Celui-ci doit être suffisamment proche de  $D_f^{\mathcal{N}}$  pour que l'estimation soit précise et en même temps suffisamment simple pour que l'intégrale (2.6) puisse être calculée.

### 2.2.2.1 Algorithme FORM

La méthode connue sous le nom de FORM (*First-Order Reliability Method*) consiste à faire une approximation linéaire de l'état limite  $H(\mathbf{w}) = T$  dans l'espace normal standard. Pour trouver un hyperplan affine susceptible de séparer les domaines de défaillance et de sûreté, on commence par chercher un point d'ancrage situé à la frontière des deux domaines et pour lequel la valeur de  $\varphi_d$  est maximale. Il s'agit du point de l'état limite qui est le plus proche de l'origine :

$$\mathbf{w}^* = \underset{H(\mathbf{w})=T}{\operatorname{argmin}} \|\mathbf{w}\|_2 . \quad (2.14)$$

On appelle  $\mathbf{w}^*$  le point de conception. La plus petite distance entre l'origine  $\mathbf{0}$  de l'espace normal standard et un point  $\mathbf{w}$  appartenant à l'état limite vaut  $\beta = \|\mathbf{w}^*\|_2$ . Elle sert d'indice de fiabilité depuis HASOFER et LIND [1974]. En ce qui concerne la résolution du problème d'optimisation, différentes techniques sont comparées par LIU et DER KIUREGHIAN [1991]. Pour construire l'hyperplan, on a aussi besoin du vecteur unitaire  $\boldsymbol{\alpha}$  indiquant la normale à l'état limite en  $\mathbf{w}^*$  :

$$\boldsymbol{\alpha} = \frac{\nabla_{\mathbf{w}} H(\mathbf{w}^*)}{\|\nabla_{\mathbf{w}} H(\mathbf{w}^*)\|_2} . \quad (2.15)$$

Comme le domaine de défaillance correspond aux points  $\mathbf{w}$  pour lesquels  $H(\mathbf{w}) > T$ , le vecteur  $\boldsymbol{\alpha}$  pointe vers  $D_f^{\mathcal{N}}$  et on a  $\beta = \boldsymbol{\alpha}^\top \mathbf{w}^*$ . La connaissance de  $\mathbf{w}^*$  et  $\boldsymbol{\alpha}$  permet alors de construire une approximation linéaire de l'état limite en utilisant l'hyperplan d'équation  $\boldsymbol{\alpha}^\top \mathbf{w} = \beta$ . Le domaine  $\mathring{D}_f^{\mathcal{N}}$  sur lequel il faut intégrer  $\varphi_d$  s'écrit alors :

$$\mathring{D}_f^{\mathcal{N}} := \left\{ \mathbf{w} \in \mathbb{R}^d : \boldsymbol{\alpha}^\top \mathbf{w} > \beta \right\} . \quad (2.16)$$

Par ailleurs, si on réalise un développement de Taylor d'ordre 1 de  $H$  au voisinage de  $\mathbf{w}^*$ , il vient :

$$H(\mathbf{w}) = H(\mathbf{w}^*) + \nabla_{\mathbf{w}} H(\mathbf{w}^*)^\top (\mathbf{w} - \mathbf{w}^*) + o(\|\mathbf{w} - \mathbf{w}^*\|_2) . \quad (2.17)$$

L'omission du reste donne l'équation de l'hyperplan tangent à l'état limite au point  $\mathbf{w}^*$  :

$$\begin{aligned} H(\mathbf{w}) &\approx \hat{H}_1(\mathbf{w}) := H(\mathbf{w}^*) + \nabla_{\mathbf{w}} H(\mathbf{w}^*)^\top (\mathbf{w} - \mathbf{w}^*) \\ &= T + \|\nabla_{\mathbf{w}} H(\mathbf{w}^*)\|_2 \boldsymbol{\alpha}^\top (\mathbf{w} - \mathbf{w}^*) = T + \|\nabla_{\mathbf{w}} H(\mathbf{w}^*)\|_2 (\boldsymbol{\alpha}^\top \mathbf{w} - \beta). \end{aligned} \quad (2.18)$$

Avec l'équation (2.18), on retrouve le fait que l'approximation linéaire  $\hat{H}_1(\mathbf{w}) = T$  de l'état limite  $H(\mathbf{w}) = T$  équivaut à  $\boldsymbol{\alpha}^\top \mathbf{w} = \beta$ . Une estimation  $\hat{P}_f^{\text{FORM}}$  de  $P_f$  s'obtient alors très facilement :

$$P_f = \mathbb{P}(H(\mathbf{W}) > T) \approx \mathbb{P}(\boldsymbol{\alpha}^\top \mathbf{W} > \beta) = 1 - \mathbb{P}(\boldsymbol{\alpha}^\top \mathbf{W} \leq \beta) = \Phi(-\beta) =: \hat{P}_f^{\text{FORM}}. \quad (2.19)$$

La dernière égalité s'obtient en remarquant que  $\|\boldsymbol{\alpha}\|_2 = 1$  et donc que  $\boldsymbol{\alpha}^\top \mathbf{W} \sim \mathcal{N}(0, 1)$ . La formule  $\hat{P}_f^{\text{FORM}} = \Phi(-\beta)$  illustre la grande simplicité de la méthode. L'hypothèse de linéarité peut sembler grossière mais STRAUB [2014] souligne qu'elle donne néanmoins des résultats satisfaisants dans de nombreux cas d'étude industriels. De surcroît, les seuls appels à  $\Phi$  sont ceux exigés par la résolution du problème d'optimisation (2.14) et leur nombre est souvent limité.

Sans surprise, les limites de FORM surviennent lorsque l'état limite  $H(\mathbf{w}) = T$  est marqué par une forte non-linéarité ou lorsque que le domaine  $D_f^{\mathcal{N}}$  n'est pas connexe. Dans cette situation, il faudrait utiliser plusieurs points de conceptions  $\mathbf{w}_1^*, \dots, \mathbf{w}_r^*$  et intégrer  $\varphi_d$  sur les sous-domaines associés. Un algorithme a été développé par DER KIUREGHIAN et DAKESSIAN [1998] en suivant cette intuition. Notons  $\mathbf{w}_1^*$  le point fourni par l'exécution initiale de FORM. La fonction  $H$  est ensuite modifiée localement pour que  $\mathbf{w}_1^*$  soit exclue du nouveau domaine de défaillance et que l'algorithme d'optimisation puisse éventuellement conduire à un point  $\mathbf{w}_2^*$  situé dans une autre région de  $\mathbb{R}^d$ . Pour cela, il faut définir un voisinage  $\mathcal{V}_1$  de  $\mathbf{w}_1^*$  et poser :

$$\forall \mathbf{w} \in \mathbb{R}^d, \quad \tilde{K}_1(\mathbf{w}) = H(\mathbf{w}) - \tilde{k}_1(\mathbf{w}) \quad \text{avec} \quad \begin{cases} 0 < \tilde{k}_1(\mathbf{w}) \leq \tilde{k}_1(\mathbf{w}^*) & \text{si } \mathbf{w} \in \mathcal{V}_1 \\ \tilde{k}_1(\mathbf{w}) = 0 & \text{si } \mathbf{w} \notin \mathcal{V}_1 \end{cases}. \quad (2.20)$$

Il est également supposé que  $\tilde{k}_1$  est continue à la frontière  $\partial\mathcal{V}_1$ . Il est alors évident que  $\tilde{K}_1(\mathbf{w}_1^*) < T$  et donc que le point  $\mathbf{w}_1^*$  se situe à l'extérieur du domaine de défaillance associé à la nouvelle fonction de performance  $\tilde{K}_1$ . La prise en compte de  $\tilde{k}_1$  provoque un bombement de l'état limite suite à l'absorption d'une partie de  $\mathcal{V}_1 \cap D_f^{\mathcal{N}}$  dans le domaine de sûreté. La position du point  $\mathbf{w}_2^*$  rendu par FORM pour  $\tilde{K}_1$  fournit une indication quant à la nature de l'état limite initial  $H(\mathbf{w}) = T$ .

- (a) Si  $\mathbf{w}_2^*$  est situé aux extrémités du renflement apparu lors du passage de  $H$  à  $\tilde{K}_1$ , on peut considérer que  $\mathbf{w}_1^*$  est l'unique point de conception du problème. Reste ensuite à savoir s'il est raisonnable de supposer que  $H(\mathbf{w}) = T$  est linéaire au voisinage de  $\mathbf{w}_1^*$ . En effet, si  $D_f^{\mathcal{N}}$  est connexe mais fortement convexe, l'approximation linéaire peut conduire à une forte surestimation de  $P_f$ .
- (b) Si  $\mathbf{w}_2^*$  est situé dans une tout autre région, l'existence concurrentielle de plusieurs points de conception indique de la surface  $H(\mathbf{w}) = T$  est fortement non-linéaire ou que le domaine  $D_f$  est constituée d'au moins deux sous-domaines disjoints. Une estimation de  $P_f$  peut être alors obtenue à partir des deux indices de fiabilité calculés :

$$\hat{P}_f = \sum_{i=1}^2 \Phi(-\beta_i) \quad \text{avec} \quad \beta_i = \|\mathbf{w}_i^*\|_2 \quad \forall i \in \{1, 2\}. \quad (2.21)$$

Cette formule n'est valable que si le recouvrement des deux demi-espaces est négligeable. Cette situation correspond à  $\mathbb{P}(\boldsymbol{\alpha}_1^\top \mathbf{W} > \beta_1, \boldsymbol{\alpha}_2^\top \mathbf{W} > \beta_2) \approx 0$ .

Tant que la situation (a) n'est pas rencontrée, la construction d'une série de fonctions  $\tilde{K}_i$  avec  $i \geq 2$  conduit au calcul d'une série de points de conception et au raffinement progressif de l'estimation de  $P_f$ . La principale difficulté de l'algorithme réside dans la mise en place d'un critère d'arrêt et dans les réglages des hyperparamètres responsables des changements de géométrie de l'état limite. Une description des détails d'implémentation est fournie dans BOURINET [2018]. Selon le nombre et la disposition des différents points de conception, il peut être dangereux de généraliser la formule (2.21) à cause des problèmes de recouvrements. Une idée naturelle est alors de construire une approximation non-linéaire de l'état limite  $H(\mathbf{w}) = T$ .

### 2.2.2.2 Algorithme SORM

Pour améliorer l'estimation de  $P_f$  dans une situation où il y a de fortes suspicions de non-linéarité de l'état limite, **FISSLER et collab. [1979]** proposent de construire une approximation quadratique. On parle alors de l'algorithme SORM (*Second Order Reliability Method*). En reprenant l'équation (2.17), le développement de Taylor d'ordre 2 de  $H$  au voisinage de  $\mathbf{w}^*$  donne :

$$H(\mathbf{w}) \approx \hat{H}_2(\mathbf{w}) := T + \|\nabla_{\mathbf{w}} H(\mathbf{w}^*)\|_2 (\boldsymbol{\alpha}^\top \mathbf{w} - \beta) + \frac{1}{2} (\mathbf{w} - \mathbf{w}^*)^\top \left( \nabla_{\mathbf{w}}^2 H(\mathbf{w}^*) \right) (\mathbf{w} - \mathbf{w}^*), \quad (2.22)$$

avec  $\nabla_{\mathbf{w}}^2 H(\mathbf{w}^*)$  la matrice hessienne de  $H$  au point  $\mathbf{w}^*$ . La décomposition spectrale de  $\nabla_{\mathbf{w}}^2 H(\mathbf{w}^*)$  inclut  $d - 1$  valeurs propres non nulles, notées  $\kappa_1, \dots, \kappa_{d-1}$  et appelées les courbures principales. Les vecteurs propres  $\mathbf{e}_1, \dots, \mathbf{e}_{d-1}$  qui leur sont associés forment les directions principales. Il est montré dans **BOURINET [2018]** qu'une simplification de l'équation (2.22) est possible lorsqu'on se place dans la base  $(\mathbf{e}_1, \dots, \mathbf{e}_{d-1}, \boldsymbol{\alpha})$ . En notant  $\mathbf{P}$  la matrice de passage et  $\mathbf{w}'$  les coordonnées dans la nouvelle base, on a :

$$\frac{\hat{H}_2(\mathbf{P}\mathbf{w}') - T}{\|\nabla_{\mathbf{w}} H(\mathbf{w}^*)\|_2} = w'_d - \beta + \frac{1}{2} \sum_{j=1}^{d-1} \kappa_j (w'_j)^2. \quad (2.23)$$

De par l'orthogonalité de  $\mathbf{P}$ , on a  $\mathbf{W}' = \mathbf{P}^\top \mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$  et il s'agit désormais d'intégrer  $\varphi_d$  sur :

$$\mathring{D}_f^{\mathcal{N}} := \left\{ \mathbf{w}' \in \mathbb{R}^d : w'_d - \beta + \frac{1}{2} \sum_{j=1}^{d-1} \kappa_j (w'_j)^2 > 0 \right\}. \quad (2.24)$$

Contrairement à ce qui a pu être fait à l'équation (2.19) où  $P_f$  a une expression analytique exacte, on ne peut obtenir ici qu'une expression analytique approchée. Celle de **BREITUNG [1984]** :

$$\hat{P}_f^{\text{SORM}} := \Phi(-\beta) \prod_{j=1}^{d-1} \frac{1}{\sqrt{1 - \beta \kappa_j}}, \quad (2.25)$$

devient asymptotiquement exacte lorsque  $\beta$  tend vers l'infini et  $P_f$  vers 0. On fait remarquer que  $\hat{P}_f^{\text{SORM}}$  est définie différemment dans **HOHENBICHLER et RACKWITZ [1988]**. Indépendamment de la formule retenue pour l'ultime étape, l'algorithme SORM appelle le code de calcul  $\phi$  pour exécuter deux tâches bien spécifiques :

- (a) Le calcul du point de conception  $\mathbf{w}^*$ .
- (b) L'estimation de  $\nabla_{\mathbf{w}}^2 H(\mathbf{w}^*)$  par différences finies pour déduire les courbures.

L'algorithme SORM tel qu'il vient d'être décrit est souvent désigné par le qualificatif *curve-fitting*. Lorsqu'il n'est pas possible d'estimer numériquement les courbures, la version *point-fitting* de **DER KIUREGHIAN et collab. [1987]** peut servir d'alternative. À partir d'une collection de points de l'état limite, l'idée est de construire une surface paraboloidale par morceaux qui est tangente à  $\mathbf{w}^*$  tout en interpolant les points choisis.

Dans cette section, on a présenté une gamme d'algorithmes de fiabilité dont le point de départ est le calcul d'un point de conception dans l'espace normal standard. Les méthodes FORM et SORM sont déterministes et elles peuvent être mises en œuvre même lorsqu'on ne dispose que d'un faible budget de simulation. L'engouement suscité par ces atouts est néanmoins terni par des insuffisances qui interviennent à deux niveaux :

- (a) La construction de la transformation isoprobabiliste inverse  $\mathcal{T}^{-1}$  est une première étape obligatoire. Pour les familles paramétriques traditionnelles, cette tâche ne présente aucune difficulté particulière. Elle peut s'avérer bien plus complexe si une structure de dépendance sophistiquée relie les variables de  $\mathbf{X}$ .
- (b) La qualité de l'estimation de  $P_f$  dépend d'une hypothèse qu'on ne peut pas vérifier sur la géométrie de l'état limite au voisinage du point de conception. En cas de forte non-linéarité de l'état limite ou en présence de plusieurs sous-domaines de défaillance, les valeurs rendues par FORM ou même SORM peuvent être fortement biaisées. Par ailleurs, on est dans l'incapacité d'évaluer ou de majorer l'erreur commise.

## 2.3 Analyse de fiabilité stochastique

Dans la cette section, on montre que les méthodes stochastiques apportent d'autres outils pour estimer la probabilité de défaillance  $P_f$ . Leur fonctionnement repose principalement sur des tirages aléatoires dans l'esprit de la méthode de Monte-Carlo. Toute estimation peut alors être vue comme la réalisation d'un estimateur, c'est-à-dire d'une variable aléatoire construite à partir des particules simulées au moment de l'échantillonnage. La Section 2.3.1 est consacrée à de succincts rappels sur la méthode naïve de Monte-Carlo. Les Sections 2.3.2 et 2.3.3 proposent un aperçu des différentes techniques d'échantillonnage préférentiel utilisées dans la suite de ce travail. Enfin la Section 2.3.4 présente certaines autres techniques d'analyse de fiabilité stochastique.

### 2.3.1 Algorithme de Monte-Carlo naïf

Comme  $P_f$  est l'espérance de la fonction indicatrice  $\mathbb{1}_{D_f}(\mathbf{X})$ , son estimateur de Monte-Carlo naïf ou *Crude Monte Carlo* (CMC) s'obtient avec un comptage empirique des défaillances :

$$\hat{P}_f^{\text{CMC}} := \frac{1}{N_s} \sum_{k=1}^{N_s} \mathbb{1}_{D_f}(\mathbf{X}^{(k)}) = \sum_{k=1}^{N_s} \mathbb{1}_{\{Y^{(k)} > T\}} \quad \text{avec} \quad \begin{cases} \mathbf{X}^{(k)} \stackrel{\text{i.i.d.}}{\sim} f_{\mathbf{X}} \\ Y^{(k)} = \phi(\mathbf{X}^{(k)}) \end{cases} . \quad (2.26)$$

De par la linéarité de l'espérance, il est trivial de voir que l'estimateur construit avec la méthode CMC est sans biais. Sa consistance est garantie par la loi forte des grands nombres. Ainsi, on a :

$$\mathbb{E} \left[ \hat{P}_f^{\text{CMC}} \right] = P_f \quad \text{et} \quad \hat{P}_f^{\text{CMC}} \xrightarrow[N_s \rightarrow \infty]{\text{p.s.}} P_f . \quad (2.27)$$

En remarquant que  $\mathbb{1}_{D_f}(\mathbf{X})$  suit une loi de Bernoulli de paramètre  $P_f$ , un TCL vient facilement :

$$\sqrt{N_s} \left( \hat{P}_f^{\text{CMC}} - P_f \right) \xrightarrow[N_s \rightarrow \infty]{\text{loi}} \mathcal{N} \left( 0, P_f(1 - P_f) \right) . \quad (2.28)$$

Le lemme de Slutsky permet de combiner les convergences (2.27) et (2.28) pour obtenir :

$$\frac{\sqrt{N_s} \left( \hat{P}_f^{\text{CMC}} - P_f \right)}{\sqrt{\hat{P}_f^{\text{CMC}}(1 - \hat{P}_f^{\text{CMC}})}} \xrightarrow[N_s \rightarrow \infty]{\text{loi}} \mathcal{N}(0, 1) . \quad (2.29)$$

Ces différents résultats de convergence ne dépendent pas de la régularité de  $\phi$ . Par ailleurs, ces propriétés sont conservées indépendamment de la nature géométrique de  $D_f$ ,  $y$  compris en cas de multimodalité. Par ailleurs, le TCL (2.29) permet d'établir un intervalle de confiance asymptotique de  $P_f$  de niveau  $1 - \alpha$  :

$$\text{IC}_{1-\alpha}(P_f) := \left[ \hat{P}_f^{\text{CMC}} \pm \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \sqrt{\frac{\hat{P}_f^{\text{CMC}}(1 - \hat{P}_f^{\text{CMC}})}{N_s}} \right] = \hat{P}_f^{\text{CMC}} \left[ 1 \pm \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \epsilon_{\text{rel}} \right] , \quad (2.30)$$

où  $\epsilon_{\text{rel}}$  est l'erreur relative tolérée au sein de l'intervalle de confiance  $\text{IC}_{1-\alpha}(P_f)$ . Lorsque  $P_f$  est la probabilité d'un évènement rare, il est facile de voir que :

$$\hat{P}_f^{\text{CMC}}(1 - \hat{P}_f^{\text{CMC}}) \approx \hat{P}_f^{\text{CMC}} \quad \text{et} \quad N_s \propto \frac{1}{\epsilon_{\text{rel}}^2 \times \hat{P}_f^{\text{CMC}}} . \quad (2.31)$$

Pour viser une erreur relative  $\epsilon_{\text{rel}}$  de 10% sur une probabilité  $P_f \approx 10^{-r}$ , il faut donc simuler au moins  $N_s = 10^{r+2}$  entrées  $\mathbf{X}^{(i)}$  et calculer les sorties  $Y^{(i)}$  associées. Pour un code de calcul  $\phi$  qui est réellement coûteux, le nombre  $N_s$  de particules nécessaires à une estimation de  $P_f$  est donc beaucoup trop important. Réciproquement, pour un budget  $N_s$  fixé à des limites réalisables, la variance de l'estimateur CMC vaut  $P_f(1 - P_f)/N_s$ . Le coefficient de variation est alors donné par :

$$\text{CV}(\hat{P}_f^{\text{CMC}}) := \frac{\sqrt{\mathbb{V}(\hat{P}_f^{\text{CMC}})}}{\mathbb{E}[\hat{P}_f^{\text{CMC}}]} = \frac{\sqrt{P_f(1 - P_f)/N_s}}{P_f} = \sqrt{\frac{1 - P_f}{P_f \times N_s}} \approx \frac{1}{\sqrt{P_f \times N_s}} . \quad (2.32)$$

Ce coefficient devient excessivement grand lorsque  $P_f$  est la probabilité d'un évènement rare. Dans cette situation, l'utilisation de la méthode CMC ne peut plus être envisagée.

La famille des méthodes de quasi-Monte-Carlo (QMC) est une première alternative car elle offre des vitesses de convergence d'ordre  $(\log N_s)^d / N_s$  au lieu de  $1/\sqrt{N_s}$ . En posant  $u_j = F_{X_j}(x_j)$  pour toutes les variables, on obtient une expression intégrale de  $P_f$  dans l'hypercube  $\mathcal{H}_d$  :

$$P_f = \int_{\mathcal{H}_d} \mathbb{1}_{D_f^U}(\mathbf{u}) c_{\mathbf{U}}(\mathbf{u}) d\mathbf{u} \quad \text{avec} \quad D_f^U := \left\{ \mathbf{u} \in \mathcal{H}_d : \phi(F_{X_1}^{-1}(u_1), \dots, F_{X_d}^{-1}(u_d)) > T \right\} . \quad (2.33)$$

L'utilisation d'une suite à discrédance faible comme celles de HALTON [1960], de SOBOL [1967] ou de NIEDERREITER [1992] permet un bien meilleur remplissage de  $\mathcal{H}_d$  que des tirages aléatoires suivant  $c_{\mathbf{U}}$ . Avec les  $N_s$  premiers points d'une suite  $\{\mathbf{u}_k\}_{k \geq 1}$ , l'estimation vaut :

$$\hat{P}_f^{\text{QMC}} := \sum_{k=1}^{N_s} \mathbb{1}_{D_f^U}(\mathbf{u}_k) c_{\mathbf{U}}(\mathbf{u}_k) . \quad (2.34)$$

Il est rappelé dans KUIPERS et NIEDERREITER [2012] que c'est l'inégalité de Koksma-Hlawka qui garantit la convergence de  $\hat{P}_f^{\text{QMC}}$  vers  $P_f$ . Les méthodes QMC se positionnent ainsi à la frontière entre simulation de Monte-Carlo et intégration numérique. Toutefois, étant donné qu'elles ne sont pas conçues spécifiquement pour échantillonner au voisinage du domaine de défaillance  $D_f^U$ , elles sont peu adaptées à la réduction de la variance de l'estimateur CMC. On renvoie à l'ouvrage de LEMIEUX [2009] pour des compléments à ce sujet.

## 2.3.2 Échantillonnage préférentiel

### 2.3.2.1 Principe de la méthode

L'échantillonnage préférentiel, ou *Importance Sampling* (IS), est l'une des méthodes les plus efficaces pour réduire la variance de l'estimateur de la probabilité de défaillance  $P_f$ . Son invention est attribuée à KAHN et HARRIS [1951] et motivée par le besoin d'estimer l'espérance d'une variable aléatoire scalaire  $\psi(\mathbf{X})$ . Étant donné que  $P_f = \mathbb{E}[\mathbb{1}_{D_f}(\mathbf{X})]$ , cette approche est utilisée comme un outil d'analyse de fiabilité dès SHINOZUKA [1983] et HARBITZ [1986]. Tout repose sur la réécriture de  $P_f$  comme l'espérance d'une autre variable aléatoire :

$$P_f = \mathbb{E}_{f_{\mathbf{X}}}[\mathbb{1}_{D_f}(\mathbf{X})] = \mathbb{E}_h[\mathbb{1}_{D_f}(\mathbf{X}) w(\mathbf{X})] \quad \text{avec} \quad w(\mathbf{X}) = \frac{f_{\mathbf{X}}(\mathbf{X})}{h(\mathbf{X})} \\ = \int_{\mathbb{R}^d} \mathbb{1}_{D_f}(\mathbf{x}) w(\mathbf{x}) h(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}^d} \left( \mathbb{1}_{D_f}(\mathbf{x}) \frac{f_{\mathbf{X}}(\mathbf{x})}{h(\mathbf{x})} \right) h(\mathbf{x}) d\mathbf{x} . \quad (2.35)$$

L'opérateur espérance est lié à une autre densité conjointe notée  $h$ . En contrepartie, la variable aléatoire initiale  $\mathbb{1}_{D_f}(\mathbf{X})$  est pénalisée par l'introduction d'un facteur  $w(\mathbf{X})$  appelé le rapport des vraisemblances. Cette nouvelle écriture de  $P_f$  suggère d'utiliser l'estimateur suivant :

$$\hat{P}_f^{\text{IS}} := \frac{1}{N_s} \sum_{k=1}^{N_s} \mathbb{1}_{\{Y^{(k)} > T\}} \frac{f_{\mathbf{X}}(\mathbf{X}^{(k)})}{h(\mathbf{X}^{(k)})} \quad \text{avec} \quad \begin{cases} \mathbf{X}^{(k)} \stackrel{\text{i.i.d.}}{\sim} h \\ Y^{(k)} = \phi(\mathbf{X}^{(k)}) \end{cases} . \quad (2.36)$$

Le nouvel estimateur est obtenu dans le même esprit que  $\hat{P}_f^{\text{CMC}}$  après avoir remplacé la densité d'échantillonnage initiale  $f_{\mathbf{X}}$  par une densité d'échantillonnage auxiliaire  $h$ . Indépendamment de la question de la variance de  $\hat{P}_f^{\text{IS}}$ , on fait remarquer qu'on ne peut pas choisir n'importe quelle densité  $h$  pour mettre en place un échantillonnage préférentiel. Pour écrire (2.35), il faut que :

$$\mathcal{H} := \text{supp}(h) \supseteq \left( \text{supp}(f_{\mathbf{X}}) \cap D_f \right) \quad \text{où} \quad \text{supp}(g) := \{ \mathbf{x} \in \mathbb{R}^d : g(\mathbf{x}) > 0 \} . \quad (2.37)$$

Sous cette condition,  $\hat{P}_f^{\text{IS}}$  est non biaisé et des calculs élémentaires donnent :

$$V(h) := \mathbb{V}_h(\hat{P}_f^{\text{IS}}) = \frac{\mathbb{V}_h(\mathbb{1}_{D_f}(\mathbf{X}) w(\mathbf{X}))}{N_s} = \frac{1}{N_s} \int_{\mathcal{H}} \frac{(\mathbb{1}_{D_f}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) - P_f h(\mathbf{x}))^2}{h(\mathbf{x})} d\mathbf{x} . \quad (2.38)$$

La variance dépend de la densité auxiliaire  $h$  utilisée pour l'échantillonnage. Avec la première égalité, on peut voir qu'elle s'annule lorsque  $\mathbb{1}_{D_f}(\mathbf{X}) w(\mathbf{X})$  est une constante. Cela correspond à l'utilisation de la densité auxiliaire optimale :

$$\forall \mathbf{x} \in \mathbb{R}^d, \quad h_{\text{opt}}(\mathbf{x}) := \frac{\mathbb{1}_{D_f}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{P_f} = f_{\mathbf{X}|\mathbf{X} \in D_f}(\mathbf{x}) \quad . \quad (2.39)$$

$h_{\text{opt}}$  est la densité de  $\mathbf{X}$  conditionnellement à la réalisation de l'évènement  $\{\mathbf{X} \in D_f\}$ . Elle ne peut pas être utilisée directement car elle implique la quantité d'intérêt  $P_f$  que l'on cherche déjà à estimer. Par ailleurs, le membre de droite de l'équation (2.38) va dans le même sens en montrant que  $V(h)$  est d'autant plus faible que les valeurs  $h(\mathbf{x})$  sont proportionnelles aux valeurs  $\mathbb{1}_{D_f}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})$ . Si la densité auxiliaire  $h$  n'est pas choisie soigneusement,  $V(h)$  peut être supérieure à la variance de l'estimateur CMC, voire infinie. En effet, l'intégrale (2.38) peut aussi donner :

$$N_s V(h) + P_f^2 = \mathbb{E}_h [\mathbb{1}_{D_f}(\mathbf{X}) w^2(\mathbf{X})] = \int_{\mathcal{H}} \mathbb{1}_{D_f}(\mathbf{x}) \left( \frac{f_{\mathbf{X}}^2(\mathbf{x})}{h(\mathbf{x})} \right) d\mathbf{x}, \quad (2.40)$$

et rien ne garantit que l'intégrale du membre de droite est finie. Si  $f_{\mathbf{X}}$  est une loi à queues lourdes, il est par exemple dangereux de choisir  $h$  au sein d'une famille à queues légères. Comme l'expliquent OWEN et ZHOU [2000], si  $D_f$  est une région où  $f_{\mathbf{X}}$  décroît à une vitesse polynomiale alors que  $h$  décroît à une vitesse exponentielle, le rapport des vraisemblances explose et la variance devient infinie. Toutefois, si  $h$  est choisie astucieusement, la variance finale peut théoriquement être nulle. On souhaite souligner que ce résultat n'est valable que lorsque  $\psi(\mathbf{X})$  est une variable positive. Dans le cas contraire, une borne strictement positive est fournie par l'inégalité de Cauchy-Schwarz :

$$\mathbb{V}_h(\psi(\mathbf{X}) w(\mathbf{X})) \geq \mathbb{E}[|\psi(\mathbf{X})|]^2 - \mathbb{E}[\psi(\mathbf{X})]^2 \quad . \quad (2.41)$$

Même si  $h_{\text{opt}}$  n'est pas accessible en pratique, on peut espérer d'importantes réductions de la variance en utilisant une densité sous-optimale  $\hat{h}_{\text{opt}}$ . Cette densité n'est pas nécessairement une estimation de  $h_{\text{opt}}$  mais elle doit se comporter globalement comme  $h_{\text{opt}}$  dans les régions où la défaillance est la plus probable. Une grande diversité de méthodes a vu le jour dans le but de trouver une densité auxiliaire  $\hat{h}_{\text{opt}}$  pouvant être considérée comme sous-optimale. Elle diffèrent par le budget de simulation qu'elles requièrent ainsi que par les hypothèses qu'elles font quant à la nature de  $h_{\text{opt}}$ . D'une manière générale, plus le budget de simulation est grand, plus la connaissance de  $D_f$  s'affine et plus  $\hat{h}_{\text{opt}}$  permet une réduction importante de la variance. Un compromis doit être trouvé en utilisant une densité auxiliaire dont le coût de construction est faible mais dont la réduction de variance est conséquente.

Une stratégie très simple présentée dans HARBITZ [1983] consiste à se placer dans l'espace normal standard et à utiliser l'expression intégrale (2.6) de  $P_f$ . Il semble alors naturel de choisir comme loi d'échantillonnage une gaussienne centrée sur le point de conception  $\mathbf{w}^*$  rendu par une exécution de FORM. Pour cette méthode souvent appelée *Gaussian Importance Sampling* (GIS), l'estimateur de  $P_f$  s'écrit :

$$\hat{P}_f^{\text{GIS}} := \frac{1}{N_s} \sum_{k=1}^{N_s} \mathbb{1}_{\{Z^{(k)} > T\}} \frac{\varphi_d(\mathbf{W}^{(k)})}{\varphi_d(\mathbf{W}^{(k)} - \mathbf{w}^*)} \quad \text{avec} \quad \begin{cases} \mathbf{W}^{(k)} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{w}^*, \mathbf{I}_d) \\ Z^{(k)} = [\phi \circ \mathcal{T}^{-1}](\mathbf{W}^{(k)}) \end{cases} \quad . \quad (2.42)$$

Par défaut, selon ce que préconisent SCHUËLLER et STIX [1987], la loi gaussienne est réduite. Dans le même esprit, on peut construire l'estimateur GIS dans l'espace physique :

$$\hat{P}_f^{\text{GIS}} := \frac{1}{N_s} \times \frac{1}{\det(\mathbf{L})} \times \sum_{k=1}^{N_s} \mathbb{1}_{\{Y^{(k)} > T\}} \frac{f_{\mathbf{X}}(\mathbf{X}^{(k)})}{\varphi_d(\mathbf{L}^{-1}(\mathbf{X}^{(k)} - \mathbf{x}^*))} \quad \text{avec} \quad \begin{cases} \mathbf{X}^{(k)} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{x}^*, \mathbf{\Sigma}) \\ Y^{(k)} = \phi(\mathbf{X}^{(k)}) \end{cases} \quad , \quad (2.43)$$

où  $\mathbf{x}^*$  est le point de conception dans l'espace physique tandis que  $\mathbf{L}$  est la racine carré de  $\mathbf{\Sigma}$  au sens de Cholesky. Dans cette situation, MELCHERS [1989] conseille de prendre une loi gaussienne

dont les composantes sont indépendantes mais dont les dispersions marginales sont calquées sur celles de  $f_{\mathbf{X}}$ . Cela revient à considérer une matrice  $\mathbf{L}$  diagonale dont le  $j$ -ème coefficient est égal à  $\sigma_j = w_j^* \times c_j$  avec  $c_j$  le coefficient de variation de la variable  $X_j$ . L'emploi de cette méthode fait resurgir les insuffisances rencontrées à la Section 2.2.2.1 avec l'algorithme FORM. Si  $h_{\text{opt}}$  est multimodale, la méthode GIS ne peut générer des échantillons qu'au voisinage du mode principal. Pour remédier à ce problème, **AU et BECK [2003]** proposent de mélanger les lois gaussiennes associées aux points de conception principaux et de définir  $\hat{h}_{\text{opt}}$  comme la densité résultante. La popularité de la méthode GIS s'explique par la simplicité avec laquelle s'obtient la densité auxiliaire. Si une analyse de fiabilité a déjà été réalisée avec FORM, la connaissance de  $\mathbf{w}^*$  fournit directement  $\hat{h}_{\text{opt}}$  et les seuls appels à  $\phi$  sont ceux requis par l'estimateur (2.36) au moment de calculer les sorties  $Y^{(k)}$ . Confronter les résultats de GIS et de FORM permet d'ailleurs d'apprécier le niveau de linéarité de l'état limite au voisinage du point de conception.

### 2.3.2.2 Apprentissage paramétrique de la densité optimale

La plupart des méthodes d'échantillonnage préférentiel ne cherchent pas à exploiter l'information apportée par une exécution préliminaire de FORM. Une démarche classique consiste plutôt à trouver le meilleur représentant  $\hat{h}_{\text{opt}} := h(\cdot | \boldsymbol{\lambda}^*)$  de  $h_{\text{opt}}$  au sein d'une famille paramétrique :

$$\mathcal{F}_{\boldsymbol{\lambda}} := \{ \mathbf{x} \mapsto h(\mathbf{x} | \boldsymbol{\lambda}) = h_{\boldsymbol{\lambda}}(\mathbf{x}) \quad \text{avec} \quad \boldsymbol{\lambda} \in \mathcal{D}_{\boldsymbol{\lambda}} \subseteq \mathbb{R}^p \} . \quad (2.44)$$

Une approche décrite dans **KROESE et collab. [2013]** invite à concevoir la variance (2.38) comme une fonction des paramètres  $\boldsymbol{\lambda}$  que l'on doit minimiser. Sachant que l'espérance de  $\mathbb{1}_{D_f}(\mathbf{X}) f_{\mathbf{X}}(\mathbf{X})$  reste égale à  $P_f$  quelle que soit  $h$ , le problème d'optimisation devient :

$$\boldsymbol{\lambda}^* := \underset{\boldsymbol{\lambda} \in \mathcal{D}_{\boldsymbol{\lambda}}}{\text{argmin}} G_A(\boldsymbol{\lambda}) \quad \text{avec} \quad G_A(\boldsymbol{\lambda}) := \mathbb{E}_{h_{\boldsymbol{\lambda}}} [\mathbb{1}_{D_f}(\mathbf{X}) w^2(\mathbf{X})] = \int_{\mathbb{R}^d} \mathbb{1}_{D_f}(\mathbf{x}) \frac{f_{\mathbf{X}}^2(\mathbf{x})}{h_{\boldsymbol{\lambda}}(\mathbf{x})} d\mathbf{x} . \quad (2.45)$$

Une résolution numérique classique est envisageable à condition de pouvoir évaluer  $G_A$ . Pour tout point  $\boldsymbol{\lambda} \in \mathcal{D}_{\boldsymbol{\lambda}}$ ,  $G_A(\boldsymbol{\lambda})$  peut être remplacé par son estimation  $\hat{G}_A(\boldsymbol{\lambda})$  obtenue par Monte-Carlo en simulant des échantillons suivant  $h_{\boldsymbol{\lambda}}$ . Toutefois, le coût d'un appel à  $\phi$  rend impossible l'idée d'un couplage entre un algorithme d'optimisation non convexe et l'estimateur CMC de la fonction objectif. Le principe de fonctionnement d'un tel algorithme est toutefois intéressant car il suggère le recours à une approche itérative pour converger vers  $\boldsymbol{\lambda}^*$ . Chaque itération implique deux étapes :

- (a) La simulation d'échantillons  $\mathbf{X}_k^{(1)}, \dots, \mathbf{X}_k^{(N_s)}$  suivant la densité auxiliaire  $h(\cdot | \boldsymbol{\lambda}_{k-1})$ .
- (b) Le calcul de  $\boldsymbol{\lambda}_k$  à partir des échantillons  $\mathbf{X}_k^{(i)}$  pour mettre à jour  $\boldsymbol{\lambda}_{k-1}$ .

On parle d'échantillonnage préférentiel adaptatif et il existe une multitude de stratégies pour exécuter la phase (b). Dans le travail précurseur de **BUCHER [1988]** où la densité auxiliaire est supposée gaussienne, l'actualisation du vecteur moyen  $\boldsymbol{\mu}$  et de la matrice de covariance  $\boldsymbol{\Sigma}$  est effectuée seulement à partir des échantillons conduisant à la défaillance. **OH et BERGER [1992]** montrent qu'on peut voir les paramètres  $\boldsymbol{\lambda}$  comme l'espérance de  $k(\mathbf{X})$  pour une certaine fonction  $k : \mathbb{R}^d \rightarrow \mathcal{D}_{\boldsymbol{\lambda}}$  et construire la procédure itérative à partir de cette idée. D'autres méthodes ont également été développées par **KARAMCHANDANI et collab. [1989]** et **MELCHERS [1990]**. En outre, **BUGALLO et collab. [2017]** dressent un bilan récent des contributions et des perspectives en matière d'échantillonnage préférentiel adaptatif.

La méthode ayant engendré les recherches les plus poussées est celle de *Cross-Entropy* (CE). On renvoie à **RUBINSTEIN et KROESE [2004]** pour une présentation exhaustive sur ce thème. Un outil essentiel de la théorie sur l'entropie croisée est la divergence de Kullback-Leibler :

$$\mathcal{D}_{\text{KL}}(h_1 | h_2) := \int_{\mathbb{R}^d} \log \left( \frac{h_1(\mathbf{x})}{h_2(\mathbf{x})} \right) h_1(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}^d} \log(h_1(\mathbf{x})) h_1(\mathbf{x}) d\mathbf{x} - \int_{\mathbb{R}^d} \log(h_2(\mathbf{x})) h_1(\mathbf{x}) d\mathbf{x} . \quad (2.46)$$

Contrairement au problème (2.45) où l'objectif est directement de minimiser la variance, l'idée de la méthode CE est de trouver le représentant de  $\mathcal{F}_\lambda$  qui se rapproche le plus de la densité optimale. La valeur optimale  $\lambda^*$  du vecteur des paramètres est donc solution de :

$$\lambda^* = \operatorname{argmin}_{\lambda \in D_\lambda} \mathcal{D}_{\text{KL}}(h_{\text{opt}} | h_\lambda) . \quad (2.47)$$

L'utilisation de  $\mathcal{D}_{\text{KL}}$  permet de construire une fonction objectif qui est plus souvent concave et différentiable. En remarquant que le premier terme de  $\mathcal{D}_{\text{KL}}(h_{\text{opt}} | h_\lambda)$  ne dépend pas de  $\lambda$  et en explicitant  $h_{\text{opt}}$ , le problème d'optimisation se transforme de la manière suivante :

$$\lambda^* = \operatorname{argmax}_{\lambda \in D_\lambda} \int_{\mathbb{R}^d} \log(h_\lambda(\mathbf{x})) h_{\text{opt}}(\mathbf{x}) d\mathbf{x} = \operatorname{argmax}_{\lambda \in D_\lambda} \mathbb{E}_{f_{\mathbf{X}}} \left[ \mathbb{1}_{D_f}(\mathbf{X}) \log(h_\lambda(\mathbf{X})) \right] . \quad (2.48)$$

et on note désormais  $G_B(\lambda)$  l'espérance à maximiser sur  $D_\lambda$ . Soient  $\mathbf{X}_1^{(1)}, \dots, \mathbf{X}_1^{(N_s)}$  des observations de la loi initiale  $f_{\mathbf{X}}$  avec  $N_s \leq 10^3$ . En remplaçant l'espérance par son estimateur empirique, on obtient une version dégradée de la fonction objectif  $G_B$  :

$$\lambda^* = \operatorname{argmax}_{\lambda \in D_\lambda} \hat{G}_B(\lambda) \quad \text{avec} \quad \hat{G}_B(\lambda) := \sum_{i=1}^{N_s} \mathbb{1}_{D_f}(\mathbf{X}_1^{(i)}) \log(h_\lambda(\mathbf{X}_1^{(i)})) \quad \text{et} \quad \mathbf{X}_1^{(i)} \stackrel{\text{i.i.d.}}{\sim} f_{\mathbf{X}} . \quad (2.49)$$

Si  $P_f$  est la probabilité d'un évènement assez fréquent, de nombreuses variables  $\mathbb{1}_{D_f}(\mathbf{X}_1^{(i)})$  sont égales à 1 et la condition nécessaire d'optimalité du premier ordre s'écrit :

$$\nabla_\lambda \hat{G}_B(\lambda^*) = \sum_{i=1}^{N_s} \mathbb{1}_{D_f}(\mathbf{X}_1^{(i)}) \left[ \nabla_\lambda \log h_{\lambda^*} \right] (\mathbf{X}_1^{(i)}) = \mathbf{0} . \quad (2.50)$$

Elle fait intervenir le score  $\mathbf{x} \mapsto \nabla_\lambda \log h_{\lambda^*}(\mathbf{x})$  de la densité sous-optimale recherchée dans  $\mathcal{F}_\lambda$ . Il est précisé dans RUBINSTEIN et KROESE [2004] que le système d'équations (2.50) admet une unique solution  $\lambda^*$  si  $\mathcal{F}_\lambda$  est une famille de la classe exponentielle. En effet, lorsqu'on exprime  $h_\lambda$  sous la forme canonique (1.82), il vient :

$$h_\lambda(\mathbf{x}) = b(\mathbf{x}) \exp[\lambda^\top t(\mathbf{x}) - A(\lambda)] \quad \text{et} \quad \nabla_\lambda \log h_\lambda(\mathbf{x}) = t(\mathbf{x}) - \nabla_\lambda A(\lambda) . \quad (2.51)$$

L'équation (2.50) conduit alors à :

$$\nabla_\lambda A(\lambda^*) = \sum_{i=1}^{N_s} \mathbb{1}_{D_f}(\mathbf{X}_1^{(i)}) t(\mathbf{X}_1^{(i)}) / \sum_{i=1}^{N_s} \mathbb{1}_{D_f}(\mathbf{X}_1^{(i)}) . \quad (2.52)$$

Sous réserve que  $\nabla_\lambda A$  soit inversible,  $\lambda^*$  peut être calculé grâce à une formule analytique. Si  $P_f$  est la probabilité d'un évènement rare, les sorties  $Y_1^{(1)}, \dots, Y_1^{(N_s)}$  sont (presque) toutes inférieures au seuil  $T$ . Dans cette situation, la somme (2.49) est nulle sur l'ensemble du domaine  $D_\lambda$  et le problème devient aberrant. Une idée consiste à fixer un seuil intermédiaire  $T_1$  franchi par une certaine proportion  $\alpha$  des sorties  $Y_1^{(i)}$  et à l'utiliser pour définir un domaine de défaillance  $D_1$  de plus forte probabilité :

$$T_1 = q_\alpha \left( \left\{ Y_1^{(i)} \right\}_{i=1}^{N_s} \right) \quad \text{et} \quad D_1 := \{ \mathbf{x} \in \mathbb{R}^d : \phi(\mathbf{x}) > T_1 \} \supseteq D_f . \quad (2.53)$$

Comme la plupart des fonctions indicatrices  $\mathbb{1}_{D_1}(\mathbf{X}_1^{(i)})$  sont activées, on peut chercher le paramètre  $\lambda_1^*$  minimisant  $\mathcal{D}_{\text{KL}}(\mathbb{1}_{D_1} f_{\mathbf{X}} | h_\lambda)$  en posant :

$$\lambda_1^* = \operatorname{argmax}_{\lambda \in D_\lambda} \hat{G}_1(\lambda) \quad \text{avec} \quad \hat{G}_1(\lambda) := \sum_{i=1}^{N_s} \mathbb{1}_{D_1}(\mathbf{X}_1^{(i)}) \log(h_\lambda(\mathbf{X}_1^{(i)})) \quad \text{et} \quad \mathbf{X}_1^{(i)} \stackrel{\text{i.i.d.}}{\sim} f_{\mathbf{X}} . \quad (2.54)$$

Il suffit d'adapter les équations (2.50) et (2.52) en remplaçant  $D_f$  par  $D_1$ . On génère ensuite des échantillons  $\mathbf{X}_2^{(1)}, \dots, \mathbf{X}_2^{(N_s)}$  selon  $h_{\lambda_1^*}$  et les sorties correspondantes  $Y_2^{(1)}, \dots, Y_2^{(N_s)}$  sont obtenues en

faisant appel à  $\phi$ . Le ré-échantillonnage à l'aide d'une densité  $h_{\lambda^*}$  proche de la densité  $f_{\mathbf{X}|\mathbf{X}\in D_1}$  provoque une légère migration des échantillons en direction de  $D_f$ . Le calcul des paramètres optimaux  $\lambda_k^*$  pour une suite croissante de seuils intermédiaires  $T_k$  aboutit à la mise en place d'un algorithme d'échantillonnage préférentiel adaptatif appelé *multilevel Cross-Entropy* (m-CE). Pour mieux comprendre le mécanisme itératif, supposons que la  $k$ -ème étape soit initiée après avoir simulé des échantillons  $\mathbf{X}_k^{(1)}, \dots, \mathbf{X}_k^{(N_s)}$  avec  $h_{\lambda_{k-1}^*}$ . En s'inspirant de ce qui est fait à l'équation (2.53), le seuil  $T_k$  est défini comme le quantile d'ordre  $\alpha$  des sorties  $Y_k^{(1)}, \dots, Y_k^{(N_s)}$  et on note  $D_k$  le domaine de défaillance associé à ce nouveau seuil intermédiaire. La fonction objectif  $G_k$  est alors :

$$\begin{aligned} G_k(\lambda) &= \mathbb{E}_{f_{\mathbf{X}}} \left[ \mathbb{1}_{D_k}(\mathbf{X}) \log(h_{\lambda}(\mathbf{X})) \right] \\ &= \mathbb{E}_{h_{\lambda_{k-1}^*}} \left[ \mathbb{1}_{D_k}(\mathbf{X}) w_{k-1}(\mathbf{X}) \log(h_{\lambda}(\mathbf{X})) \right] \quad \text{avec} \quad w_{k-1}(\mathbf{X}) = \frac{f_{\mathbf{X}}(\mathbf{X})}{h_{\lambda_{k-1}^*}(\mathbf{X})}. \end{aligned} \quad (2.55)$$

La seconde expression de  $G_k$  laisse la possibilité de construire  $\hat{G}_k$  à partir des derniers échantillons simulés. Le problème d'optimisation guidant la mise à jour des paramètres s'écrit alors :

$$\lambda_k^* = \underset{\lambda \in D_{\lambda}}{\operatorname{argmax}} \hat{G}_k(\lambda) \quad \text{avec} \quad \hat{G}_k(\lambda) := \sum_{i=1}^{N_s} \mathbb{1}_{D_k}(\mathbf{X}_k^{(i)}) w_{k-1}(\mathbf{X}_k^{(i)}) \log(h_{\lambda}(\mathbf{X}_k^{(i)})) \quad \text{et} \quad \mathbf{X}_k^{(i)} \stackrel{\text{i.i.d.}}{\sim} h_{\lambda_{k-1}^*}. \quad (2.56)$$

Si  $\mathcal{F}_{\lambda}$  appartient à la classe exponentielle, la  $k$ -ème condition d'optimalité donne :

$$\nabla_{\lambda} A(\lambda_k^*) = \sum_{i=1}^{N_s} \mathbb{1}_{D_f}(\mathbf{X}_k^{(i)}) w_{k-1}(\mathbf{X}_k^{(i)}) t(\mathbf{X}_k^{(i)}) \Big/ \sum_{i=1}^{N_s} \mathbb{1}_{D_f}(\mathbf{X}_k^{(i)}) w_{k-1}(\mathbf{X}_k^{(i)}) . \quad (2.57)$$

L'accès à  $\lambda_k^*$  permet de construire la densité auxiliaire  $h_{\lambda_k^*}$  et de simuler les échantillons  $\mathbf{X}_{k+1}^{(i)}$  pour alimenter l'itération suivante. La mise à jour des paramètres est répétée jusqu'à l'itération finale où  $T_n > T$ . Pour obtenir une estimation de  $P_f$ , il suffit de réutiliser les échantillons ayant déclenché l'activation du critère d'arrêt :

$$\hat{p}_f^{\text{CE}} := \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbb{1}_{\{Y_n^{(i)} > T\}} \frac{f_{\mathbf{X}}(\mathbf{X}_n^{(i)})}{h_{\lambda_{n-1}^*}(\mathbf{X}_n^{(i)})} \quad \text{avec} \quad \begin{cases} \mathbf{X}_n^{(i)} \stackrel{\text{i.i.d.}}{\sim} h_{\lambda_{n-1}^*} \\ Y_n^{(i)} = \phi(\mathbf{X}_n^{(i)}) \end{cases} . \quad (2.58)$$

Une question récurrente est alors de savoir si la suite des densités auxiliaires  $h(\cdot | \lambda_k^*)$  converge vers la densité optimale  $h_{\text{opt}}$ . Un résultat de convergence formulé par RUBINSTEIN et KROESE [2004] indique que l'algorithme m-CE s'achève après un nombre fini d'itérations sous réserve que  $\mathcal{F}_{\lambda}$  appartienne à la classe exponentielle et qu'une propriété supplémentaire sur les fonctions de répartition soit vérifiée. MARGOLIN [2005] précise que la convergence ne peut pas se produire sans la réunion de trois facteurs :

- La densité optimale  $h_{\text{opt}}$  doit appartenir à la classe  $\mathcal{F}_{\lambda}$ .
- Les paramètres  $\lambda_k^*$  doivent maximiser les fonctions objectifs  $G_k$  déterministes.
- Les hyperparamètres ( $\alpha$  et  $N_s$ ) doivent faire l'objet d'un réglage minutieux.

Dans l'algorithme qui vient d'être présenté et qui est utilisé en pratique, la condition (b) n'est pas vérifiée. Par ailleurs, la condition (a) soulève à nouveau le problème de la possible multimodalité de  $h_{\text{opt}}$ . Si on sait que  $D_f$  rassemble plusieurs sous-domaines, KURTZ et SONG [2013] suggèrent de chercher  $h_{\lambda^*}$  au sein de la classe des mélanges gaussiens. Pour augmenter les capacités d'apprentissage, WANG et SONG [2016] ainsi que PAPAIOANNOU et collab. [2019] proposent de construire des mélanges plus sophistiqués afin d'approcher la densité optimale dans l'espace normal standard. GEYER et collab. [2019] soulignent toutefois que ce gain en termes de flexibilité fait naître une difficulté supplémentaire. En effet, dès lors que  $h_{\lambda}$  est écrite comme la densité d'un mélange, les paramètres ne peuvent plus être mis à jour avec une formule de type (2.57) et m-CE doit alors être couplé avec l'algorithme *Expectation-Maximization* (EM) décrit par CHEN et GUPTA [2010]. Parmi les autres tentatives d'hybridation de la méthode CE, BOTEV et collab. [2007] établissent

un périmètre d'investigation plus large en cherchant une densité auxiliaire au sein d'un espace fonctionnel respectant seulement certaines contraintes sur les moments.

On souhaite faire remarquer que les performances de m-CE sont détériorées par la montée en dimension. Une discussion plus large sur la dégénérescence de l'échantillonnage préférentiel en grande dimension est proposée dans RUBINSTEIN et GLYNN [2009]. Pour lutter contre ce fléau, des outils de réduction de la dimension sont combinés avec l'algorithme m-CE dans les travaux de URIBE et collab. [2020] et EL MASRI et collab. [2020]. Le vecteur  $\lambda$  des paramètres est projeté dans un sous-espace de petite dimension et seul ce résumé s'actualise au fil des itérations de m-CE.

### 2.3.2.3 Apprentissage non-paramétrique de la densité optimale

Pour éviter les écueils posés par le cadre paramétrique de la méthode CE, il a aussi été envisagé de construire une densité sous-optimale  $\hat{h}_{\text{opt}}$  à l'aide d'outils non-paramétriques. Les premières avancées dans cette direction remontent aux recherches conduites par ANG et collab. [1989]. Soient  $\mathbf{X}_1^{(1)}, \dots, \mathbf{X}_1^{(N_s)}$  des échantillons de la densité conditionnelle  $f_{\mathbf{X}|\mathbf{X} \in D_f}$  qui sont simulés avec la méthode d'acceptation/rejet dans ANG et collab. [1992]. La densité optimale  $h_{\text{opt}}$  est alors estimée avec une version multidimensionnelle de la procédure non-paramétrique KDE :

$$h_1(\mathbf{x}) := \hat{h}_{\text{KDE}}(\mathbf{x} | \mathcal{X}_1, \mathbf{H}_1) = \frac{1}{N_s} \sum_{i=1}^{N_s} K_{\mathbf{H}_1}(\mathbf{x} - \mathbf{X}_1^{(i)}) \quad \text{avec} \quad \mathcal{X}_1 := \{\mathbf{X}_1^{(i)}\}_{i=1}^{N_s}. \quad (2.59)$$

Le noyau gaussien  $K_{\mathbf{H}_1}$  désigne la densité d'une loi  $\mathcal{N}(\mathbf{0}, \mathbf{H}_1)$  et s'écrit donc :

$$K_{\mathbf{H}_1}(\mathbf{z}) = \frac{1}{\det(\mathbf{B}_1)} \varphi_d(\mathbf{B}_1^{-1} \mathbf{z}) \quad \text{avec} \quad \mathbf{H}_1 = \mathbf{B}_1 \mathbf{B}_1^\top. \quad (2.60)$$

La matrice de lissage  $\mathbf{H}_1$  est calibrée pour minimiser la variance de l'estimateur  $\hat{P}_f^{\text{IS}}$ . Au regard de l'équation (2.40), cela signifie que :

$$\mathbf{H}_1 = \underset{\mathbf{H} > \mathbf{0}}{\text{argmin}} \int_{\mathbb{R}^d} \mathbb{1}_{D_f}(\mathbf{x}) \left( \frac{f_{\mathbf{X}}^2(\mathbf{x})}{\hat{h}_{\text{KDE}}(\mathbf{x} | \mathcal{X}_1, \mathbf{H})} \right) d\mathbf{x}. \quad (2.61)$$

Au moment d'évaluer les rapports de vraisemblance  $w(\mathbf{X}_1^{(i)})$ , l'utilisation de la densité auxiliaire  $\hat{h}_{\text{KDE}}(\cdot | \mathcal{X}_1, \mathbf{H}_1)$  implique cependant un plus fort coût de calcul car il faut réaliser  $N_s$  évaluations de  $K_{\mathbf{H}_1}$ . Pour disposer d'une densité auxiliaire moins coûteuse, WEST [1993] propose une technique pour transformer  $\hat{h}_{\text{KDE}}(\cdot | \mathcal{X}_1, \mathbf{H}_1)$  en la densité d'un mélange n'incluant qu'un petit nombre de composantes. Lorsque les échantillons  $\mathbf{X}_1^{(1)}, \dots, \mathbf{X}_1^{(N_s)}$  ne peuvent pas être générés avec  $f_{\mathbf{X}|\mathbf{X} \in D_f}$ , il faut utiliser une densité instrumentale  $h_0$ . ZHANG [1996] fait remarquer qu'une première estimation de  $P_f$  est donnée par :

$$\hat{P}_1 := \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbb{1}_{\{Y_1^{(i)} > T\}} \frac{f_{\mathbf{X}}(\mathbf{X}_1^{(i)})}{h_0(\mathbf{X}_1^{(i)})} = \frac{1}{N_s} \sum_{i=1}^{N_s} w_1^{(i)} \quad \text{avec} \quad \begin{cases} \mathbf{X}_1^{(i)} \stackrel{\text{i.i.d.}}{\sim} h_0 \\ Y_1^{(i)} = \phi(\mathbf{X}_1^{(i)}) \end{cases}, \quad (2.62)$$

où chaque poids  $w_1^{(i)}$  peut être vu comme la contribution de  $\mathbf{X}_1^{(i)}$  à l'estimation de  $\hat{P}_1$ . Ces poids permettent de construire différemment la densité auxiliaire non-paramétrique :

$$h_1(\mathbf{x}) := \hat{h}_{\text{KDE}}(\mathbf{x} | \mathcal{X}_1, \mathcal{W}_1, \mathbf{H}_1) = \frac{1}{N_s \hat{P}_1} \sum_{i=1}^{N_s} w_1^{(i)} K_{\mathbf{H}_1}(\mathbf{x} - \mathbf{X}_1^{(i)}) \quad \text{avec} \quad \mathcal{W}_1 := \{w_1^{(i)}\}_{i=1}^{N_s}. \quad (2.63)$$

$\mathbf{H}_1$  est ici obtenue en minimisant le critère AMISE (*Asymptotic Mean Integrated Square Error*) qui est détaillé dans WAND et JONES [1994]. La présence de  $\hat{P}_1$  au dénominateur garantit que  $h_1$  soit une densité de probabilité. Quand l'estimation finale de  $P_f$  est réalisée avec  $h_1$ , l'algorithme est appelé NIS (*Non-parametric Importance Sampling*). Si on choisit de simuler des échantillons

$\mathbf{X}_2^{(1)}, \dots, \mathbf{X}_2^{(N_s)}$  à partir de  $h_1$  puis de répéter l'opération, NIS devient NAIS en raison du caractère adaptatif. À la  $k$ -ième itération,  $P_f$  peut être estimée comme suit :

$$\hat{P}_k := \frac{1}{kN_s} \sum_{j=1}^k \sum_{i=1}^{N_s} \mathbb{1}_{\{Y_j^{(i)} > T\}} \frac{f_{\mathbf{X}}(\mathbf{X}_j^{(i)})}{h_0(\mathbf{X}_j^{(i)})} = \frac{1}{N_s} \sum_{j=1}^k \sum_{i=1}^{N_s} w_j^{(i)} \quad \text{avec} \quad \begin{cases} \mathbf{X}_j^{(i)} \stackrel{\text{i.i.d.}}{\sim} h_{j-1} \\ Y_j^{(i)} = \phi(\mathbf{X}_j^{(i)}) \end{cases}, \quad (2.64)$$

et la densité auxiliaire  $h_k$  se construit de la manière suivante :

$$h_k(\mathbf{x}) := \hat{h}_{\text{KDE}}\left(\mathbf{x} \mid \bigcap_{j=1}^k \mathcal{X}_j, \bigcap_{j=1}^k \mathcal{W}_j, \mathbf{H}_k\right) = \frac{1}{kN_s \hat{P}_k} \sum_{j=1}^k \sum_{i=1}^{N_s} w_j^{(i)} \mathbf{K}_{\mathbf{H}_k}(\mathbf{x} - \mathbf{X}_j^{(i)}). \quad (2.65)$$

Lorsque  $N_s$  augmente, [ZHANG \[1996\]](#) a prouvé que  $h_1$  converge plus rapidement vers  $h_{\text{opt}}$  que les densités paramétriques. En revanche, il déplore que la densité produite par NAIS ne soit pas meilleure que celle produite par NIS. [GIVENS et RAFTERY \[1996\]](#) adaptent NIS pour qu'une matrice de lissage  $\mathbf{H}_k^{(i)}$  soit calibrée pour chaque échantillon  $\mathbf{X}_k^{(i)}$ . Cette flexibilité supplémentaire permet à chaque noyau  $\mathbf{K}_{\mathbf{H}_k^{(i)}}$  de reproduire avec une plus grande fidélité la structure de covariance locale de  $f_{\mathbf{X}}$  autour de  $\mathbf{X}_k^{(i)}$ . Pour réduire le temps d'exécution, [NEDDERMEYER \[2009\]](#) montre que les noyaux gaussiens peuvent être remplacés par des noyaux LBFP (*Linear Blend Frequency Polygon*) créés par [TERRELL et SCOTT \[1985\]](#). Au niveau applicatif, NAIS a été notamment utilisé pour la maintenance des réseaux de télécommunication dans [KIM et collab. \[2000\]](#) puis pour l'évaluation des produits financiers dérivés dans [NEDDERMEYER \[2011\]](#).

Comme pour la méthode CE, si  $P_f$  est la probabilité d'un évènement rare, aucune des sorties ne dépasse le seuil  $T$ , tous les poids sont nuls,  $\hat{P}_1 = 0$  et la densité (2.63) ne peut pas être construite. Une stratégie plus fine doit être mise en place pour se doter d'une densité sous-optimale  $\hat{h}_{\text{opt}}$ . Pour cela, [AU et BECK \[1999\]](#) construisent un processus de Markov de distribution stationnaire la distribution d'échantillonnage optimale. Ce processus est une suite  $\{\mathbf{Z}_l\}_{l \geq 0}$  telle que :

$$f_{\mathbf{Z}_l | \mathbf{Z}_{l-1} \mathbf{Z}_{l-2} \dots \mathbf{Z}_0}(\mathbf{z}_l | \mathbf{z}_{l-1}, \mathbf{z}_{l-2}, \dots, \mathbf{z}_0) = f_{\mathbf{Z}_l | \mathbf{Z}_{l-1}}(\mathbf{z}_l | \mathbf{z}_{l-1}) = f_{\mathbf{Z}_1 | \mathbf{Z}_0}(\mathbf{z}_1 | \mathbf{z}_0), \quad (2.66)$$

où la densité conditionnelle  $f_{\mathbf{Z}_1 | \mathbf{Z}_0}$  est appelée la densité de transition d'un pas et vérifie :

$$\forall (\mathbf{z}_0, \mathbf{z}_1) \in (\mathbb{R}^d)^2, \quad h_{\text{opt}}(\mathbf{z}_1) = \int_{\mathbb{R}^d} f_{\mathbf{Z}_1 | \mathbf{Z}_0}(\mathbf{z}_1 | \mathbf{z}_0) h_{\text{opt}}(\mathbf{z}_0) d\mathbf{z}_0. \quad (2.67)$$

Après une période de rodage de  $N_b$  pas, on peut considérer  $\mathbf{Z}_{N_b}$  est approximativement distribué suivant  $h_{\text{opt}}$ . Il faut cependant garder à l'esprit que l'initialisation du processus de Markov nécessite la connaissance d'au moins un échantillon défaillant, ce qui n'est pas toujours possible en pratique. Par ailleurs, lorsque la zone de défaillance  $D_f$  n'est pas connexe, les étapes successives de la vie du processus de Markov  $\{\mathbf{Z}_l\}_{l \geq 0}$  sont marquées par des sauts d'un sous-domaine de  $D_f$  à l'autre. Théoriquement, on sait que les sous-domaines constituant  $D_f$  seront tous atteints mais on ne sait pas au bout de combien de temps. Pour remédier à ces différents problèmes, [MORIO \[2012\]](#) propose d'améliorer l'algorithme NAIS de [ZHANG \[1996\]](#) en s'inspirant du mécanisme itératif de m-CE. À partir des échantillons initiaux  $\mathbf{X}_1^{(1)}, \dots, \mathbf{X}_1^{(N_s)}$  simulés à partir de  $h_0$ , un seuil intermédiaire  $T_1$  et un domaine de défaillance  $D_1$  sont définis exactement comme à l'équation (2.53) puis la probabilité  $\mathbb{P}(\mathbf{X} \in D_1)$  est estimée par :

$$\hat{P}_{D_1} := \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbb{1}_{\{Y_1^{(i)} > T_1\}} \frac{f_{\mathbf{X}}(\mathbf{X}_1^{(i)})}{h_0(\mathbf{X}_1^{(i)})} = \frac{1}{N_s} \sum_{i=1}^{N_s} w_{1, T_1}^{(i)} \quad \text{avec} \quad \begin{cases} \mathbf{X}_1^{(i)} \stackrel{\text{i.i.d.}}{\sim} h_0 \\ Y_1^{(i)} = \phi(\mathbf{X}_1^{(i)}) \end{cases}. \quad (2.68)$$

La densité  $h_1$  est ensuite construite dans l'esprit de l'équation (2.63) en adaptant les poids et la constante de renormalisation. Cette nouvelle version de  $h_1$  permet d'activer un nombre suffisant de fonctions indicatrices. On est alors en mesure de simuler des échantillons  $\mathbf{X}_2^{(1)}, \dots, \mathbf{X}_2^{(N_s)}$  puis

d'enclencher le mécanisme itératif. À la  $k$ -ième itération,  $\mathbb{P}(\mathbf{X} \in D_k)$  est estimée en utilisant tous les échantillons simulés au cours des itérations précédentes :

$$\hat{\mathbb{P}}_{D_k} := \frac{1}{k N_s} \sum_{j=1}^k \sum_{i=1}^{N_s} \mathbb{1}_{\{Y_j^{(i)} > T_k\}} \frac{f_{\mathbf{X}}(\mathbf{X}_j^{(i)})}{h_0(\mathbf{X}_j^{(i)})} = \frac{1}{k N_s} \sum_{j=1}^k \sum_{i=1}^{N_s} w_{j,T_k}^{(i)} \quad \text{avec} \quad \begin{cases} \mathbf{X}_j^{(i)} \text{ i.i.d. } h_{j-1} \\ Y_j^{(i)} = \phi(\mathbf{X}_j^{(i)}) \end{cases} \quad (2.69)$$

Contrairement à l'algorithme NAIS de ZHANG [1996] où le système de pondération utilisé pour construire la densité (2.65) est obtenu en rassemblant les poids calculés au fil des itérations, les poids sont ici remis à jour à chaque itération en fonction du seuil  $T_k$  devant être dépassé pour activer les fonctions indicatrices. La densité auxiliaire  $h_k$  prend alors la forme suivante :

$$\begin{aligned} h_k(\mathbf{x}) &:= \hat{h}_{\text{KDE}}\left(\mathbf{x} \mid \bigcap_{j=1}^k \mathcal{X}_j, \overline{\mathcal{W}}_k, \mathbf{H}_k\right) \quad \text{avec} \quad \overline{\mathcal{W}}_k = \left\{ w_{j,T_k}^{(i)} \right\}_{\substack{1 \leq i \leq N_s \\ 1 \leq j \leq k}} \\ &= \frac{1}{k N_s \hat{\mathbb{P}}_k} \sum_{j=1}^k \sum_{i=1}^{N_s} w_{j,T_k}^{(i)} \mathbb{K}_{\mathbf{H}_k}(\mathbf{x} - \mathbf{X}_j^{(i)}) . \end{aligned} \quad (2.70)$$

Les itérations s'enchaînent jusqu'à la  $n$ -ième itération où  $T_n > T$ . La dernière densité auxiliaire construite est considérée comme sous-optimale et détermine la variance de l'estimateur final :

$$\hat{\mathbb{P}}_f^{\text{NAIS}} := \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbb{1}_{\{Y_n^{(i)} > T\}} \frac{f_{\mathbf{X}}(\mathbf{X}_n^{(i)})}{h_{n-1}(\mathbf{X}_n^{(i)})} \quad \text{avec} \quad \begin{cases} \mathbf{X}_n^{(i)} \text{ i.i.d. } h_{n-1} \\ Y_n^{(i)} = \phi(\mathbf{X}_n^{(i)}) \end{cases} \quad (2.71)$$

Cette version de NAIS est appliquée à l'estimation de la probabilité de retombée hors-zone d'un étage de lanceur de satellite dans MORIO [2011] puis dans MORIO et BALESDENT [2016]. À l'instar de m-CE, l'exploration de  $D_f$  via NAIS impose  $n \times N_s$  appels au code de calcul  $\phi$ . Si cette charge de calcul ne peut pas être supportée, BALESDENT et collab. [2013] montrent que NAIS peut être combiné avec une métamodélisation de  $\phi$  à base de krigeage. Le processus gaussien introduit une incertitude supplémentaire dont l'impact sur  $\hat{\mathbb{P}}_f^{\text{NAIS}}$  est étudié dans MENZ et collab. [2020].

D'une manière générale, le déploiement d'une méthode KDE multidimensionnelle accélère et facilite l'apprentissage du support et des niveaux de vraisemblance de la densité optimale  $h_{\text{opt}}$  même lorsque l'état limite est fortement non-linéaire. Les problèmes inhérents à la montée en dimension, déjà évoqués pour l'algorithme m-CE, ne sont pas pour autant éradiqués et restent prégnants. En plus grande dimension, si on souhaite maintenir un certain niveau de précision dans l'estimation de  $h_{\text{opt}}$ , SCOTT [1991] indique que le nombre d'échantillons à fournir augmente à un rythme exponentiel. Pour cette raison, il est fortement déconseillé d'utiliser NAIS dès lors que  $d > 7$ . Toutefois, l'avènement des R-vignes non-paramétriques dans NAGLER et CZADO [2016] porte la promesse de futures améliorations de cet algorithme.

### 2.3.3 Améliorations et variantes de l'échantillonnage préférentiel

Dans certaines situations, on dispose de plusieurs densités auxiliaires  $\hat{h}_1, \dots, \hat{h}_r$  pour une même densité initiale  $f_{\mathbf{X}}$ . Il peut s'agir des densités auxiliaires obtenues après avoir exécuté à  $r$  reprises un algorithme d'échantillonnage préférentiel adaptatif. On peut aussi imaginer la situation où les densités auxiliaires produites par  $r$  techniques d'échantillonnage préférentiel sont rassemblées et comparées. Les supports peuvent être disjoints mais ils peuvent aussi se superposer. On part du principe que toutes les densités apportent une information à prendre en considération et on cherche à toutes les incorporer au sein d'un même estimateur. Pour y parvenir, il existe plusieurs stratégies que l'on veille à présenter succinctement dans les prochains paragraphes.

#### 2.3.3.1 Échantillonnage préférentiel stratifié

Cette méthode est née du mariage entre l'échantillonnage préférentiel et une autre technique de réduction de la variance appelée l'échantillonnage stratifié. Le principe de la stratification est

un partitionnement de l'espace d'échantillonnage. Toutefois, le cadre originel qui est décrit dans **GLASSERMAN [2013]** peut être très largement généralisé. Ici, on se place dans un contexte où les supports des densités  $\hat{h}_j$  ne forment pas nécessairement une partition de  $\mathbb{R}^d$ . Pour chaque densité auxiliaire  $\hat{h}_j$ , on suppose qu'on connaît  $n_j$  appariements entre des entrées et des sorties :

$$\forall 1 \leq j \leq r, \quad \mathcal{X}_j := \left\{ (\mathbf{X}_j^{(i)}, Y_j^{(i)}) \right\}_{i=1}^{n_j} \quad \text{avec} \quad \mathbf{X}_j^{(i)} \stackrel{\text{i.i.d.}}{\sim} \hat{h}_j \quad \text{et} \quad Y_j^{(i)} = \phi(\mathbf{X}_j^{(i)}) . \quad (2.72)$$

Le matériel  $\mathcal{X}_j$  provient de l'ultime étape de l'algorithme adaptatif ayant produit  $\hat{h}_j$ . On écrit  $\hat{h}_{\text{opt}}$  comme une combinaison convexe des densités initiales :

$$\forall \mathbf{x} \in \mathbb{R}^d, \quad \hat{h}_{\text{strat}}(\mathbf{x}) := \sum_{j=1}^r \epsilon_j \hat{h}_j(\mathbf{x}) \quad \text{avec} \quad \sum_{j=1}^r \epsilon_j = 1 . \quad (2.73)$$

En utilisant les densités  $\hat{h}_j$  et les coefficients  $\epsilon_j$ ,  $P_f$  peut s'exprimer comme suit :

$$P_f = \sum_{j=1}^r \epsilon_j \mathbb{E}_{\hat{h}_j} [\mathbb{1}_{D_f}(\mathbf{X}) w_j(\mathbf{X})] \bigg/ \sum_{j=1}^r \epsilon_j \mathbb{E}_{\hat{h}_j} [w_j(\mathbf{X})] \quad \text{avec} \quad w_j = \frac{f_{\mathbf{X}}}{\hat{h}_j} . \quad (2.74)$$

Un estimateur de  $P_f$  en découle immédiatement en réutilisant tous les échantillons disponibles :

$$\hat{P}_f^{\text{strat}} := \sum_{j=1}^r \epsilon_j \left( \frac{1}{n_j} \sum_{i=1}^{n_j} \mathbb{1}_{\{Y_j^{(i)} > T\}} w_j(\mathbf{X}_j^{(i)}) \right) \bigg/ \sum_{j=1}^r \epsilon_j \left( \frac{1}{n_j} \sum_{i=1}^{n_j} w_j(\mathbf{X}_j^{(i)}) \right) . \quad (2.75)$$

La consistance de cet estimateur est démontrée dans **OH et BERGER [1993]**. Par ailleurs, sa variance peut être approchée de la manière suivante :

$$\mathbb{V}(\hat{P}_f^{\text{strat}}) \approx \sum_{j=1}^r \frac{\epsilon_j^2 \sigma_j^2}{n_j} \quad \text{où} \quad \sigma_j^2 = \mathbb{V}_{\hat{h}_j} (\mathbb{1}_{D_f}(\mathbf{X}) w_j(\mathbf{X})) . \quad (2.76)$$

Au regard de la contrainte de convexité, la variance minimale s'obtient en prenant la combinaison  $\epsilon^*$  telle que  $\epsilon_j^* \propto n_j / \sigma_j^2$ . Cela favorise les densités fournissant le plus d'échantillons ainsi que les densités conduisant à des estimateurs à faible variance. En pratique, chaque variance  $\sigma_j^2$  peut être estimée empiriquement à partir du matériel fourni par  $\mathcal{X}_j$ . Ainsi, l'échantillonnage préférentiel stratifié s'apparente à un simple post-traitement des précédentes explorations de  $D_f$ .

### 2.3.3.2 Échantillonnage préférentiel après un mélange

Un autre point de vue consiste à supposer que les échantillons  $\mathcal{X} := \mathcal{X}_1 \cup \dots \cup \mathcal{X}_r$  sont issus d'une distribution de mélange dont la densité est donnée par :

$$\forall \mathbf{x} \in \mathbb{R}^d, \quad \hat{h}_{\text{mix}} := \sum_{j=1}^r \alpha_j \hat{h}_j(\mathbf{x}) \quad \text{avec} \quad \alpha_j = \frac{n_j}{n_1 + \dots + n_r} = \frac{n_j}{n_{\text{tot}}} . \quad (2.77)$$

L'estimateur de  $P_f$  qui est associé à cette méthode s'obtient comme suit :

$$\hat{P}_f^{\text{mix}} := \frac{1}{n_{\text{tot}}} \sum_{j=1}^r \sum_{i=1}^{n_j} \mathbb{1}_{\{Y_j^{(i)} > T\}} \frac{f_{\mathbf{X}}(\mathbf{X}_j^{(i)})}{\sum_{k=1}^r \alpha_k \hat{h}_k(\mathbf{X}_j^{(i)})} . \quad (2.78)$$

C'est un estimateur de type  $\hat{P}_f^{\text{IS}}$  dans le cas particulier où  $N_s = n_{\text{tot}}$  et  $h = \hat{h}_{\text{mix}}$ . Pour un échantillon  $\mathbf{X}_j^{(i)}$  donné, on doit calculer la valeur de toutes les densités  $\hat{h}_k$  et pas uniquement la valeur de la densité  $\hat{h}_j$  qui a contribué à le générer. C'est une différence importante par rapport à l'estimateur produit par l'échantillonnage préférentiel stratifié. Les stratégies de mélange pour l'échantillonnage préférentiel ont été développées par **HESTERBERG [1988]**. Entre autres, elles sont appliquées

à l'estimation de la fiabilité des mémoires vives dans [KANJ et collab. \[2006\]](#). L'emploi de la densité auxiliaire  $\hat{h}_{\text{mix}}$  est très naturel lorsque les échantillons  $\mathcal{X}_j$  forment des îlots distants les uns des autres. Toutefois, il existe d'autres situations pour lesquelles l'emploi d'un mélange est salutaire. Lorsqu'on dispose d'une densité instrumentale  $\hat{h}_0$  et qu'il existe des suspicions quant à la faculté de  $\hat{h}_0$  à réduire la variance, le recours au mélange de  $\hat{h}_0$  et  $f_{\mathbf{X}}$  offre une solution de sécurité. Cette méthode est appelée *Defensive Importance Sampling* (DIS) dans [HESTERBERG \[1995\]](#). On a alors  $r = 2$  et la densité sous-optimale s'écrit :

$$\forall \mathbf{x} \in \mathbb{R}^d, \quad \hat{h}_{\alpha}(\mathbf{x}) = \alpha_1 f_{\mathbf{X}}(\mathbf{x}) + \alpha_2 \hat{h}_0(\mathbf{x}) \quad \text{avec} \quad \alpha_1 + \alpha_2 = 1, \quad (2.79)$$

et on peut montrer l'inégalité suivante sur la variance de l'estimateur :

$$\mathbb{V}(\hat{P}_f^{\alpha}) = \frac{1}{N_s} \mathbb{V}_{\hat{h}_{\alpha}}(\mathbb{1}_{D_f}(\mathbf{X}) w_{\alpha}(\mathbf{X})) \leq \frac{1}{\alpha_1 N_s} \left( \mathbb{V}_{f_{\mathbf{X}}}(\mathbb{1}_{D_f}(\mathbf{X})) + \alpha_2 P_f^2 \right). \quad (2.80)$$

Quelle que soit la combinaison convexe  $\alpha$ , la variance reste finie et sa valeur est bridée par celle de la variance de l'estimateur CMC. [HESTERBERG \[1995\]](#) montre que la méthode DIS est encore plus intéressante lorsque  $f_{\mathbf{X}}$  n'est connue qu'à une constante près et qu'il faut se tourner vers l'échantillonnage préférentiel auto-normalisant.

### 2.3.3.3 Échantillonnage préférentiel multiple

Dans ce paragraphe, on note  $\hat{P}_f^{[j]}$  l'estimateur de  $P_f$  qui est obtenu à partir de  $\hat{h}_j$  et de  $\mathcal{X}_j$  :

$$\forall 1 \leq j \leq r, \quad \hat{P}_f^{[j]} := \frac{1}{n_j} \sum_{i=1}^{n_j} \mathbb{1}_{\{Y_j^{(i)} > T\}} w_j(\mathbf{X}_j^{(i)}) \quad \text{avec} \quad \begin{cases} \mathbf{X}_j^{(i)} \text{ i.i.d. } \hat{h}_j \\ Y_j^{(i)} = \phi(\mathbf{X}_j^{(i)}) \end{cases}. \quad (2.81)$$

Une possibilité qui n'a pas encore été examinée est celle offerte par l'estimateur moyen :

$$\hat{P}_f^{[\bullet]} = \sum_{j=1}^r \frac{n_j}{n_{\text{tot}}} \hat{P}_f^{[j]} = \frac{1}{n_{\text{tot}}} \sum_{j=1}^r \sum_{i=1}^{n_j} \mathbb{1}_{\{Y_j^{(i)} > T\}} w_j(\mathbf{X}_j^{(i)}) = \sum_{j=1}^r \frac{1}{n_j} \sum_{i=1}^{n_j} \left( \frac{n_j}{n_{\text{tot}}} \right) \mathbb{1}_{\{Y_j^{(i)} > T\}} w_j(\mathbf{X}_j^{(i)}). \quad (2.82)$$

Il est alors facile de calculer la variance puis de la mettre sous la forme suivante :

$$\mathbb{V}(\hat{P}_f^{[\bullet]}) = \frac{1}{n_{\text{tot}}^2} \sum_{j=1}^r n_j \sigma_j^2 = \sum_{j=1}^r \frac{n_j}{n_{\text{tot}}} \left( \frac{\sigma_j^2}{n_{\text{tot}}} \right) = \sum_{j=1}^r \frac{n_j}{n_{\text{tot}}} \mathbb{V}(\hat{P}_f^{[j]}) = \frac{\sigma_m^2}{n_{\text{tot}}}. \quad (2.83)$$

La variance finale est alors la moyenne pondérée des variances initiales. La méthode de *Multiple Importance Sampling* (MIS) peut être vue comme une généralisation de l'estimateur moyen pour moduler la variance à l'objectif souhaité. Pour cela, on affecte à chaque échantillon  $\mathbf{X}_j^{(i)}$  un poids  $p_j^{(i)} := p_j(\mathbf{X}_j^{(i)})$  qui dépend de la fonction de pondération  $p_j$  affectée à  $\hat{h}_j$  ainsi que de la valeur prise par cette fonction au point  $\mathbf{X}_j^{(i)}$ . Ce poids remplace alors le ratio  $n_j/n_{\text{tot}}$  présent dans le membre de droite de l'équation (2.82) et cela conduit à l'estimateur suivant :

$$\hat{P}_f^{\text{MIS}} = \sum_{j=1}^r \frac{1}{n_j} \sum_{i=1}^{n_j} p_j(\mathbf{X}_j^{(i)}) \mathbb{1}_{\{Y_j^{(i)} > T\}} w_j(\mathbf{X}_j^{(i)}) \quad \text{avec} \quad \begin{cases} \mathbf{X}_j^{(i)} \text{ i.i.d. } \hat{h}_j \\ Y_j^{(i)} = \phi(\mathbf{X}_j^{(i)}) \end{cases}. \quad (2.84)$$

Pour que l'estimateur  $\hat{P}_f^{\text{MIS}}$  soit non biaisé, il faut que les fonctions de pondération constituent une partition de l'unité :

$$\forall \mathbf{x} \in \mathbb{R}^d, \quad \sum_{j=1}^r p_j(\mathbf{x}) = 1. \quad (2.85)$$

Sous cette condition, on peut en effet vérifier que :

$$\mathbb{E} \left[ \hat{P}_f^{\text{MIS}} \right] = \sum_{j=1}^r \frac{1}{n_j} \sum_{i=1}^{n_j} \mathbb{E}_{\hat{h}_j} \left[ p_j(\mathbf{X}_j^{(i)}) \mathbb{1}_{D_f}(\mathbf{X}_j^{(i)}) w_j(\mathbf{X}_j^{(i)}) \right] \quad (2.86)$$

$$= \sum_{j=1}^r \frac{1}{n_j} \sum_{i=1}^{n_j} \int_{D_f} p_j(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \sum_{j=1}^r \frac{1}{n_j} \int_{D_f} \left( \sum_{j=1}^r p_j(\mathbf{x}) \right) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = P_f. \quad (2.87)$$

Cette approche a été développée par **VEACH et GUIBAS [1995a]** et permet de créer une gamme d'estimateurs aux propriétés très hétérogènes. Examinons une situation impliquant  $r = 3$  densités auxiliaires. Elles sont renumérotées de manière à avoir  $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$ . De plus, on suppose que les échantillons sont répartis au sein d'îlots disjoints :

$$\forall 1 \leq j \leq 3, \quad \forall 1 \leq i \leq n_j, \quad \mathbf{X}_j^{(i)} \in A_j \quad \text{avec} \quad \begin{cases} A_1 \cup A_2 \cup A_3 = \mathbb{R}^d \\ A_1 \cap A_2 \cap A_3 = \emptyset \end{cases}. \quad (2.88)$$

Si on pose  $p_1 := \mathbb{1}_{A_1}$ ,  $p_2 := \mathbb{1}_{A_3}$  et  $p_3 := \mathbb{1}_{A_2}$ , on isole les échantillons de  $\mathcal{X}_1$  et on se ramène à un estimateur de variance  $\sigma_1^2/n_1$ . Avec une astuce du même genre, on peut aussi fabriquer un estimateur de variance  $\sigma_2^2/n_2$  ou  $\sigma_3^2/n_3$ . Évidemment, ce cas jouet n'a que peu d'intérêt en pratique où l'objectif n'est pas de se ramener aux estimateurs initiaux. Parmi les fonctions de pondération les plus fréquemment utilisées, on peut citer l'heuristique d'équilibrage, ou *Balance Heuristic* (BH). On choisit des fonctions  $w_j \propto n_j \hat{h}_j$  ce qui conduit finalement à :

$$p_j^{\text{BH}}(\mathbf{x}) = n_j \hat{h}_j(\mathbf{x}) / \sum_{k=1}^r n_k \hat{h}_k(\mathbf{x}). \quad (2.89)$$

Un échantillon  $\mathbf{X}_j^{(i)}$  est muni d'un poids  $p_j^{\text{BH}}(\mathbf{X}_j^{(i)})$  qui est alors d'autant plus grand que  $\mathbf{X}_j^{(i)}$  se situe dans une zone de forte vraisemblance de  $\hat{h}_j$  non partagée par les autres densités  $\hat{h}_k$ . En injectant l'expression (2.89) des fonctions  $p_j^{\text{BH}}$  au sein de l'estimateur (2.84), il vient :

$$\hat{P}_f^{\text{BH}} := \sum_{j=1}^r \sum_{i=1}^{n_j} \mathbb{1}_{\{Y_j^{(i)} > T\}} \frac{f_{\mathbf{X}}(\mathbf{X}_j^{(i)})}{\sum_{k=1}^r n_k \hat{h}_k(\mathbf{X}_j^{(i)})} \quad \text{avec} \quad \begin{cases} \mathbf{X}_j^{(i)} \stackrel{\text{i.i.d.}}{\sim} \hat{h}_j \\ Y_j^{(i)} = \phi(\mathbf{X}_j^{(i)}) \end{cases}. \quad (2.90)$$

$\hat{P}_f^{\text{BH}}$  peut être vu comme le résultat d'un mélange des densités  $\hat{h}_j$  car il est équivalent à :

$$\hat{P}_f^{\text{BH}} = \frac{1}{n_{\text{tot}}} \sum_{k=1}^{n_{\text{tot}}} \mathbb{1}_{\{Y^{(k)} > T\}} \frac{f_{\mathbf{X}}(\mathbf{X}^{(k)})}{\hat{h}_{\text{mix}}(\mathbf{X}^{(k)})} \quad \text{avec} \quad \begin{cases} \mathbf{X}^{(k)} \stackrel{\text{i.i.d.}}{\sim} \hat{h}_{\text{mix}} = \sum_{j=1}^r \alpha_j \hat{h}_j \quad \text{où} \quad \alpha_j = \frac{n_j}{n_{\text{tot}}} \\ Y^{(k)} = \phi(\mathbf{X}^{(k)}) \end{cases}. \quad (2.91)$$

Il est montré dans **VEACH et GUIBAS [1995b]** que la méthode BH est presque optimale. En effet, il existe une inégalité entre la variance de  $\hat{P}_f^{\text{BH}}$  et celle de tout autre estimateur construit par MIS :

$$\mathbb{V}(\hat{P}_f^{\text{BH}}) - \mathbb{V}(\hat{P}_f^{\text{MIS}}) \leq \left( \frac{1}{n_{\min}} - \frac{1}{n_{\text{tot}}} \right) P_f^2. \quad (2.92)$$

L'écart entre la variance du meilleur estimateur  $\hat{P}_f^{\text{MIS}}$  et celle de  $\hat{P}_f^{\text{BH}}$  est au plus égal au terme de droite qui est d'autant plus faible que  $P_f$  est la probabilité d'un évènement rare. D'autres fonctions de pondération comme les heuristiques *cutoff* et *power* sont testées dans **VEACH [1997]**. En notant  $\mathcal{H}$  l'une de ses deux heuristiques et  $c_{\mathcal{H}}$  une constante qui s'y rapporte, un résultat analogue à l'inégalité (2.92) peut être énoncé :

$$\mathbb{V}(\hat{P}_f^{\mathcal{H}}) - c_{\mathcal{H}} \mathbb{V}(\hat{P}_f^{\text{MIS}}) \leq \left( \frac{1}{n_{\min}} - \frac{1}{n_{\text{tot}}} \right) P_f^2. \quad (2.93)$$

Dans cette section, on a mis en avant les possibilités offertes par l'échantillonnage préférentiel pour réduire la variance de l'estimateur CMC. Tout repose sur l'utilisation d'une densité auxiliaire sous-optimale, c'est-à-dire d'une densité auxiliaire suffisamment proche de la densité optimale qui est inaccessible. Ainsi, la capacité de l'échantillonnage préférentiel à réduire la variance dépend avant tout de la densité auxiliaire mise au service de l'utilisateur ou construite par ce dernier. Parmi la multitude de techniques disponibles, on a examiné trois algorithmes de fiabilité. Ils reposent respectivement sur le point de conception rendu par FORM, l'optimisation par entropie croisée au sein d'une famille paramétrique et l'apprentissage non-paramétrique de la densité optimale. Lorsque la défaillance est un évènement très rare, des versions adaptatives de ces algorithmes peuvent être implémentées et des mécanismes de ré-échantillonnage font migrer les échantillons initiaux en direction des régions d'intérêt. Si ces méthodes réagissent différemment en présence d'une densité optimale multimodale, la montée en grande dimension est un point bloquant qui les freinent toutes. Par ailleurs, lorsqu'on met en commun les résultats de plusieurs algorithmes d'échantillonnage préférentiel ou les résultats de plusieurs exécutions d'un de ces algorithmes, différentes stratégies peuvent être suivies pour aboutir à un estimateur à faible variance qui ne requiert aucun appel supplémentaire au code de calcul.

### 2.3.4 Autres méthodes d'analyse de fiabilité stochastique

Après un tour d'horizon des techniques d'échantillonnage préférentiel, cette dernière section mentionne l'existence de certaines autres méthodes de fiabilité, soit parce qu'elles sont de premier plan soit parce qu'elles suscitent actuellement un fort intérêt des spécialistes.

Tout d'abord, on souhaite s'attarder sur l'algorithme de *Subset Simulation* (SS) qui occupe une place centrale dans la littérature dédiée à la simulation d'évènements rares. Ses balbutiements sont concomitants avec ceux de l'échantillonnage préférentiel et sont également attribuables à [KAHN et HARRIS \[1951\]](#). On adopte ici le formalisme proposé par [AU et BECK \[2001\]](#) dans le contexte de la fiabilité des structures. Il existe également d'autres manières d'aborder cet algorithme mais elles impliquent un plus haut niveau d'abstraction mathématique. En particulier, on peut évoquer le lien avec les processus de branchement dans [GLASSERMAN et collab. \[1999\]](#), avec les processus de Markov dans [CÉROU et GUYADER \[2007\]](#) ou avec les modèles particuliers de Feynmann-Kac dans [CÉROU et collab. \[2012\]](#). Toutefois, quel que soit le point de vue adopté, l'idée reste la même. Il faut fractionner la probabilité de défaillance  $P_f = \mathbb{P}(Y > T)$  après avoir défini une suite croissante de seuils  $T_1 < T_2 < \dots < T_n = T$ . En effet, comme les ensembles  $D_k := \{\mathbf{x} \in \mathbb{R}^d : \phi(\mathbf{x}) > T_k\} \subset \mathbb{R}^d$  sont inclus les uns dans les autres, le théorème de Bayes donne immédiatement :

$$P_f = \mathbb{P}(\mathbf{X} \in D_f) = \underbrace{\mathbb{P}(\mathbf{X} \in D_1)}_{:=p_1} \prod_{k=2}^n \underbrace{\mathbb{P}(\mathbf{X} \in D_k \mid \mathbf{X} \in D_{k-1})}_{:=p_k}. \quad (2.94)$$

L'idée est simplement d'estimer les probabilités  $p_k$  les unes après autres grâce à un mécanisme de ré-échantillonnage qui fait converger les échantillons simulés de  $D_1$  vers  $D_n = D_f$ . En pratique, on ne connaît pas les seuils  $T_1, \dots, T_n$  et leur nombre  $n$  dépend de la vitesse d'exploration que l'on souhaite imposer. Dans ce contexte, l'emploi des notations  $\hat{T}_1, \dots, \hat{T}_n$  permet d'insister sur le fait que la collection des seuils est une entité aléatoire. Le principe de fonctionnement de l'algorithme SS peut alors être résumé comme suit :

1. **Initialisation.** On simule des échantillons  $\mathcal{X}_1 := \{\mathbf{X}_1^{(i)}\}_{i=1}^{N_s}$  avec  $f_{\mathbf{X}}$  et on calcule toutes les sorties associées  $\mathcal{Y}_1 := \{Y_1^{(i)}\}_{i=1}^{N_s}$  avec  $\phi$ . On pose ensuite  $\hat{T}_1 := q_{\alpha}(\mathcal{Y}_1)$ .
2. **Étapes de ré-échantillonnage.** Les échantillons de  $\mathcal{X}_1$  sont scindés en deux blocs  $\mathcal{X}_1^+$  et  $\mathcal{X}_1^-$  selon qu'ils dépassent ou non le seuil  $\hat{T}_1$ . Au vu de la définition de  $\hat{T}_1$ , on a alors :

$$\text{Card}(\mathcal{X}_1^+) = N_s^+ := (1 - \alpha) N_s \quad \text{et} \quad \text{Card}(\mathcal{X}_1^-) = N_s^- := \alpha N_s \quad . \quad (2.95)$$

Les  $N_s^+$  échantillons de  $\mathcal{X}_1^+$  alimentent l'algorithme de Metropolis-Hastings (MH) afin de produire  $N_s^-$  échantillons distribués selon la densité conditionnelle  $f_{\mathbf{X}|\mathbf{X}\in D_1}$  et rassemblés dans  $\mathcal{X}_2^{\text{MH}}$ . Après avoir utilisé  $\phi$  pour calculer les sorties  $\mathcal{Y}_2^{\text{MH}}$  associées à  $\mathcal{X}_2^{\text{MH}}$ , on définit :

$$\mathcal{X}_2 := \mathcal{X}_1^+ \cup \mathcal{X}_2^{\text{MH}} \quad \text{et} \quad \mathcal{Y}_2 := \mathcal{Y}_1^+ \cup \mathcal{Y}_2^{\text{MH}}. \quad (2.96)$$

On peut ensuite passer à la deuxième itération et répéter la procédure.

3. **Arrêt.** Dès que  $\hat{T}_{\hat{n}} > T$ , on interrompt le ré-échantillonnage et on estime  $P_f$  avec :

$$\hat{P}_f^{\text{SS}} := (1 - \alpha)^{\hat{n}-1} \left( \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbb{1}_{\{Y_n^{(i)} > T\}} \right). \quad (2.97)$$

Les principales difficultés rencontrées lors de l'implémentation de l'algorithme SS sont le choix d'un quantile optimal et la mise en œuvre de l'algorithme MH, surtout en grande dimension.

Parmi les techniques qui n'ont pas été abordées, il y a aussi celles qui visent une synergie des méthodes déterministes et stochastiques. La méthode de *Line Sampline* (LS) présentée dans [KOUTSOURELAKIS et collab. \[2004\]](#) et [SCHUËLLER et collab. \[2004\]](#) en est un exemple. L'idée est de renforcer l'algorithme FORM pour qu'il reste précis dans des situations où l'état limite dans l'espace normal standard est légèrement non-linéaire. Supposons qu'une première exécution de FORM ait fourni un point de conception  $\mathbf{w}^*$  et un vecteur normal unitaire  $\boldsymbol{\alpha}$  tels qu'ils sont définis aux équations (2.14) et (2.15). On rappelle qu'ils donnent accès à l'indice de fiabilité  $\beta = \boldsymbol{\alpha}^\top \mathbf{w}^*$  et à la probabilité de défaillance  $P_f = \Phi(-\beta)$ . Dans la méthode d'échantillonnage en ligne, le vecteur  $\boldsymbol{\alpha}$  est figé mais on cherche à recalculer  $\beta$  en s'appuyant sur la connaissances d'autres points de l'état limite. La démarche pourrait être résumée de la manière suivante :

1. On commence par simuler des échantillons  $\mathbf{W}^{(1)}, \dots, \mathbf{W}^{(N_s)}$  dans l'espace normal standard.
2. Pour chaque échantillon  $\mathbf{W}^{(i)}$ , il faut trouver un point d'intersection entre la droite  $\mathcal{D}_i$  de direction  $\boldsymbol{\alpha}$  passant par  $\mathbf{W}^{(i)}$  et l'état limite  $H(\mathbf{w}) = T$ . Comme il est supposé que l'état limite est un hyperplan de direction  $\boldsymbol{\alpha}$ , on en déduit facilement de nouvelles valeurs  $\beta^{(i)}$  et  $\hat{P}_f^{(i)}$ .
3. On estime  $P_f$  avec  $\hat{P}_f^{\text{LS}} := \frac{1}{N_s} \sum_{i=1}^{N_s} \hat{P}_f^{(i)}$ .

Contrairement au cas de l'échantillonnage préférentiel où la réduction de la variance dépend de la densité auxiliaire, la variance de  $\hat{P}_f^{\text{LS}}$  est toujours plus petite que celle de  $\hat{P}_f^{\text{CMC}}$ . La méthode de *Directional Sampling* (DS) apparue dans [BJERAGER \[1988\]](#) est construite sur le même principe, mais en supposant que l'état limite est une hypersphère.

Actuellement, un grande partie des efforts de recherche dans le domaine de la fiabilité des structures porte sur le couplage entre les méthodes stochastiques de fiabilité et les méthodes de métamodélisation du code de simulation. Sachant que l'objectif final d'une analyse de fiabilité est l'estimation d'une probabilité de défaillance, il faut adapter la construction du métamodèle dans ce but. Sous la contrainte d'un budget de simulation réduit, le métamodèle ne peut pas être précis partout et il doit impérativement l'être dans les zones explorées par l'algorithme de fiabilité, en particulier la zone de défaillance. L'apprentissage actif est une stratégie qui consiste à mettre à jour le métamodèle à chaque nouvelle itération de l'algorithme de fiabilité de manière à ce qu'il reste toujours précis dans les régions où il est sollicité. Lorsque le modèle de substitution est un processus gaussien, [ECHARD et collab. \[2011\]](#) ont développé la méthode AK-MCS (*Active learning reliability method combining Kriging and Monte Carlo Simulation*) et de nombreuses variantes ont suivi comme AK-IS dans [ECHARD et collab. \[2013\]](#), AK-SS dans [Lv et collab. \[2015\]](#) ou AK-LS dans [HUANG et collab. \[2016\]](#) pour n'en citer que quelques unes. [BOURINET et collab. \[2011\]](#) ont montré qu'il était également possible de combiner l'algorithme SS et un apprentissage du code de simulation avec des machines à vecteurs de support. Des stratégies d'apprentissage actif sont également explorées dans [BOURINET \[2016\]](#) pour ce type de métamodèles.

Pour finir, les méthodes basées sur la théorie des valeurs extrêmes s'appuient uniquement sur des échantillons de sortie  $\mathcal{Y} := \{Y^{(i)}\}_{i=1}^{N_s}$  pour estimer la probabilité de défaillance, souvent réécrite

$P_f = \mathbb{P}(Y > T) = 1 - F_Y(T)$  à l'occasion. Ce contexte survient notamment lorsque la variable d'intérêt  $Y$  est directement observable. Il faut alors quantifier la probabilité de dépassement d'un seuil très supérieur à la plus grande des observations de  $\mathcal{Y}$ . Si on s'en tient à l'information apportée par la fonction de répartition empirique de la sortie, la probabilité de défaillance est nulle. Pour estimer  $P_f$ , l'idée est alors de calibrer à partir de  $\mathcal{Y}$  un modèle de distribution paramétrique susceptible d'être suffisamment précis au niveau de la queue de distribution supérieure. Il existe deux grandes approches en pratique :

- L'approche *Block Maxima* (BM) s'intéresse à la distribution de la variable  $Y_{(N_s)} := \max_i Y^{(i)}$ . Il faut partitionner  $\mathcal{Y}$  en plusieurs sous-échantillons et calculer la valeur maximale dans chacun d'eux. L'échantillon des maxima permet ensuite de calibrer un modèle *Generalized Extreme Value* (GEV). L'un des paramètres estimés est appelé l'indice aux valeurs extrêmes et sert à déterminer si la queue de distribution est courte, légère ou lourde.
- L'approche *Peak Over Threshold* (POT) s'intéresse à la distribution de la variable  $Y | Y > T_0$  pour  $T_0$  un seuil défini comme le quantile d'ordre  $\alpha$  de  $\mathcal{Y}$ . À partir des  $(1 - \alpha)N_s$  échantillons situés au-dessus de  $T_0$ , on calibre un modèle *Generalized Pareto Distribution* (GPD).

Dans les deux cas, des formules analytique permettent ensuite d'estimer  $P_f$ . La mise en œuvre des deux méthodes implique un compromis biais-variance (partitionnement de  $\mathcal{Y}$  dans BM, choix du seuil  $T_0$  dans POT). Il a été démontré dans [PICKANDS et collab. \[1975\]](#) que les deux approches sont en fait équivalentes. Pour un exposé plus détaillé sur la théorie des valeurs extrêmes, on renvoie aux ouvrages de [RESNICK \[2013\]](#) ou de [EMBRECHTS et collab. \[2013\]](#).

## 2.4 Conclusion

Ce chapitre a été l'occasion de présenter les techniques de fiabilité utilisées dans la suite de ce travail. Une frontière très nette a été établie entre les méthodes déterministes qui reposent sur l'existence d'une transformation isoprobabiliste et les méthodes stochastiques qui requièrent une capacité d'échantillonnage au voisinage du domaine de défaillance. Dans la première situation, l'estimation est invariante mais on ne contrôle pas l'erreur commise. Dans la seconde situation, on construit un estimateur c'est-à-dire une variable aléatoire dont la dispersion dépend de la robustesse de la technique d'échantillonnage mise en œuvre. Ceci n'est valable qu'à condition que la loi des entrées soit figée. Dans le cadre de ce travail, la loi des entrées porte au contraire une incertitude épistémique émanant de la variabilité d'échantillonnage. Une analogie peut être faite avec le cadre bayésien où le paramétrage de la loi des entrées est incertain. Que l'algorithme soit déterministe ou stochastique, la valeur prise par les paramètres de la densité d'entrée conditionne l'estimation de la probabilité de défaillance qui doit alors être vue comme la réalisation d'une variable aléatoire. Dans ce contexte, on peut chercher les facteurs exerçant une influence importante sur la distribution de l'estimateur du risque. Cette méthodologie est appelée analyse de sensibilité fiabiliste et fait l'objet du prochain chapitre.

## Références

- AAS, K., C. CZADO, A. FRIGESSI et H. BAKKEN. 2009, «Pair-copula constructions of multiple dependence», *Insurance : Mathematics and economics*, vol. 44, n° 2, p. 182–198.
- ANG, G. L., A. H. S. ANG et W. H. TANG. 1989, «Kernel method in importance sampling density function estimation», dans *Proceedings of ICOSSAR'89, the 5th International Conference on Structural Safety and Reliability*, vol. 89, p. 193–200.
- ANG, G. L., A. H.-S. ANG et W. H. TANG. 1992, «Optimal importance-sampling density estimator», *Journal of engineering mechanics*, vol. 118, n° 6, p. 1146–1163.
- AU, S.-K. et J. BECK. 2003, «Important sampling in high dimensions», *Structural safety*, vol. 25, n° 2, p. 139–163.

- AU, S.-K. et J. L. BECK. 1999, «A new adaptive importance sampling scheme for reliability calculations», *Structural safety*, vol. 21, n° 2, p. 135–158.
- AU, S.-K. et J. L. BECK. 2001, «Estimation of small failure probabilities in high dimensions by subset simulation», *Probabilistic engineering mechanics*, vol. 16, n° 4, p. 263–277.
- BALESDENT, M., J. MORIO et J. MARZAT. 2013, «Kriging-based adaptive importance sampling algorithms for rare event estimation», *Structural Safety*, vol. 44, p. 1–10.
- BJERAGER, P. 1988, «Probability integration by directional simulation», *Journal of Engineering Mechanics*, vol. 114, n° 8, p. 1285–1302.
- BOTEV, Z. I., D. P. KROESE et T. TAIMRE. 2007, «Generalized cross-entropy methods with applications to rare-event simulation and optimization», *Simulation*, vol. 83, n° 11, p. 785–806.
- BOURINET, J.-M. 2016, «Rare-event probability estimation with adaptive support vector regression surrogates», *Reliability Engineering & System Safety*, vol. 150, p. 210–221.
- BOURINET, J.-M. 2018, *Reliability analysis and optimal design under uncertainty : focus on adaptive surrogate-based approaches*, Habilitation à diriger des recherches (HDR), Université Clermont Auvergne.
- BOURINET, J.-M., F. DEHEEGER et M. LEMAIRE. 2011, «Assessing small failure probabilities by combined subset simulation and support vector machines», *Structural Safety*, vol. 33, n° 6, p. 343–353.
- BREITUNG, K. 1984, «Asymptotic approximations for multinormal integrals», *Journal of Engineering Mechanics*, vol. 110, n° 3, p. 357–366.
- BUCHER, C. G. 1988, «Adaptive sampling : an iterative fast Monte Carlo procedure», *Structural safety*, vol. 5, n° 2, p. 119–126.
- BUGALLO, M. F., V. ELVIRA, L. MARTINO, D. LUENGO, J. MIGUEZ et P. M. DJURIC. 2017, «Adaptive importance sampling : the past, the present, and the future», *IEEE Signal Processing Magazine*, vol. 34, n° 4, p. 60–79.
- CÉROU, F., P. DEL MORAL, T. FURON et A. GUYADER. 2012, «Sequential Monte Carlo for rare event estimation», *Statistics and computing*, vol. 22, n° 3, p. 795–808.
- CÉROU, F. et A. GUYADER. 2007, «Adaptive multilevel splitting for rare event analysis», *Stochastic Analysis and Applications*, vol. 25, n° 2, p. 417–443.
- CHEN, Y. et M. R. GUPTA. 2010, «EM demystified : an expectation-maximization tutorial», dans *Electrical Engineering*, Citeseer.
- CORNELL, C. A. 1969, «A probability-based structural code», dans *Journal Proceedings*, vol. 66, p. 974–985.
- DAVIS, P. J. et P. RABINOWITZ. 2007, *Methods of numerical integration*, Courier Corporation.
- DER KIUREGHIAN, A. et T. DAKESSIAN. 1998, «Multiple design points in first and second-order reliability», *Structural Safety*, vol. 20, n° 1, p. 37–49.
- DER KIUREGHIAN, A., H.-Z. LIN et S.-J. HWANG. 1987, «Second-order reliability approximations», *Journal of Engineering mechanics*, vol. 113, n° 8, p. 1208–1225.
- DITLEVSEN, O. et H. O. MADSEN. 1996, *Structural reliability methods*, vol. 178, Wiley, New York.
- ECHARD, B., N. GAYTON et M. LEMAIRE. 2011, «AK-MCS : an active learning reliability method combining kriging and Monte Carlo simulation», *Structural Safety*, vol. 33, n° 2, p. 145–154.

- ECHARD, B., N. GAYTON, M. LEMAIRE et N. RELUN. 2013, «A combined importance sampling and kriging reliability method for small failure probabilities with time-demanding numerical models», *Reliability Engineering & System Safety*, vol. 111, p. 232–240.
- EL MASRI, M., J. MORIO et F. SIMATOS. 2020, «Improvement of the cross-entropy method in high dimension through a one-dimensional projection without gradient estimation», *arXiv preprint arXiv :2012.11241*.
- EMBRECHTS, P., C. KLÜPPELBERG et T. MIKOSCH. 2013, *Modelling extremal events : for insurance and finance*, vol. 33, Springer Science & Business Media.
- FISSLER, B., R. RACKWITZ et H.-J. NEUMANN. 1979, «Quadratic limit states in structural reliability», *Journal of the Engineering Mechanics Division*, vol. 105, n° 4, p. 661–676.
- GEYER, S., I. PAPAIOANNOU et D. STRAUB. 2019, «Cross entropy-based importance sampling using Gaussian densities revisited», *Structural Safety*, vol. 76, p. 15–27.
- GIVENS, G. H. et A. E. RAFTERY. 1996, «Local adaptive importance sampling for multivariate densities with strong nonlinear relationships», *Journal of the American Statistical Association*, vol. 91, n° 433, p. 132–141.
- GLASSERMAN, P. 2013, *Monte Carlo methods in financial engineering*, vol. 53, Springer Science & Business Media.
- GLASSERMAN, P., P. HEIDELBERGER, P. SHAHABUDDIN et T. ZAJIC. 1999, «Multilevel splitting for estimating rare event probabilities», *Operations Research*, vol. 47, n° 4, p. 585–600.
- HALTON, J. H. 1960, «On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals», *Numerische Mathematik*, vol. 2, n° 1, p. 84–90.
- HARBITZ, A. 1983, «Efficient and accurate probability of failure calculation by the use of importance sampling technique», dans *Proc. of ICASP*, vol. 4, p. 825–836.
- HARBITZ, A. 1986, «An efficient sampling method for probability of failure calculation», *Structural safety*, vol. 3, n° 2, p. 109–115.
- HASOFER, A. M. et N. C. LIND. 1974, «Exact and invariant second-moment code format», *Journal of the Engineering Mechanics division*, vol. 100, n° 1, p. 111–121.
- HESTERBERG, T. 1995, «Weighted average importance sampling and defensive mixture distributions», *Technometrics*, vol. 37, n° 2, p. 185–194.
- HESTERBERG, T. C. 1988, *Advances in importance sampling*, thèse de doctorat, Stanford University.
- HINRICHS, A., E. NOVAK, M. ULLRICH et H. WOŹNIAKOWSKI. 2014, «The curse of dimensionality for numerical integration of smooth functions», *Mathematics of Computation*, vol. 83, n° 290, p. 2853–2863.
- HOHENBICHLER, M. et R. RACKWITZ. 1988, «Improvement of second-order reliability estimates by importance sampling», *Journal of Engineering Mechanics*, vol. 114, n° 12, p. 2195–2199.
- HUANG, X., J. CHEN et H. ZHU. 2016, «Assessing small failure probabilities by AK-SS : an active learning method combining kriging and subset simulation», *Structural Safety*, vol. 59, p. 86–95.
- KAHN, H. et T. E. HARRIS. 1951, «Estimation of particle transmission by random sampling», *National Bureau of Standards applied mathematics series*, vol. 12, p. 27–30.
- KANJ, R., R. JOSHI et S. NASSIF. 2006, «Mixture importance sampling and its application to the analysis of SRAM designs in the presence of rare failure events», dans *2006 43rd ACM/IEEE Design Automation Conference*, IEEE, p. 69–72.

- KARAMCHANDANI, A., P. BJERAGER et C. CORNELL. 1989, «Adaptive importance sampling», dans *Structural Safety and Reliability*, ASCE, p. 855–862.
- KIM, Y. B., D. S. ROH et M. Y. LEE. 2000, «Nonparametric adaptive importance sampling for rare event simulation», dans *2000 Winter Simulation Conference Proceedings (Cat. No. 00CH37165)*, vol. 1, IEEE, p. 767–772.
- KOUTSOURELAKIS, P.-S., H. PRADLWARTER et G. SCHUËLLER. 2004, «Reliability of structures in high dimensions, part I : algorithms and applications», *Probabilistic Engineering Mechanics*, vol. 19, n° 4, p. 409–417.
- KROESE, D. P., T. TAIMRE et Z. I. BOTEV. 2013, *Handbook of Monte Carlo methods*, vol. 706, John Wiley & Sons.
- KUIPERS, L. et H. NIEDERREITER. 2012, *Uniform distribution of sequences*, Courier Corporation.
- KURTZ, N. et J. SONG. 2013, «Cross-entropy-based adaptive importance sampling using Gaussian mixture», *Structural Safety*, vol. 42, p. 35–44.
- LEBRUN, R. 2013, *Contributions à la modélisation de la dépendance stochastique*, thèse de doctorat, Université Paris VII - Denis Diderot.
- LEBRUN, R. et A. DUTFOY. 2009a, «Do Rosenblatt and Nataf isoprobabilistic transformations really differ?», *Probabilistic Engineering Mechanics*, vol. 24, n° 4, p. 577–584.
- LEBRUN, R. et A. DUTFOY. 2009b, «A generalization of the Nataf transformation to distributions with elliptical copula», *Probabilistic Engineering Mechanics*, vol. 24, n° 2, p. 172–178.
- LEBRUN, R. et A. DUTFOY. 2009c, «An innovating analysis of the Nataf transformation from the copula viewpoint», *Probabilistic Engineering Mechanics*, vol. 24, n° 3, p. 312–320.
- LEMAIRE, M., A. CHATEAUNEUF et J.-C. MITTEAU. 2009, *Structural reliability*, Wiley Online Library.
- LEMIEUX, C. 2009, *Monte Carlo and quasi-Monte Carlo sampling*, Springer Science & Business Media.
- LIU, P.-L. et A. DER KIUREGHIAN. 1986, «Multivariate distribution models with prescribed marginals and covariances», *Probabilistic Engineering Mechanics*, vol. 1, n° 2, p. 105–112.
- LIU, P.-L. et A. DER KIUREGHIAN. 1991, «Optimization algorithms for structural reliability», *Structural safety*, vol. 9, n° 3, p. 161–177.
- LV, Z., Z. LU et P. WANG. 2015, «A new learning function for kriging and its applications to solve reliability problems in engineering», *Computers & Mathematics with Applications*, vol. 70, n° 5, p. 1182–1197.
- MARGOLIN, L. 2005, «On the convergence of the cross-entropy method», *Annals of Operations Research*, vol. 134, n° 1, p. 201–214.
- MELCHERS, R. 1989, «Importance sampling in structural systems», *Structural safety*, vol. 6, n° 1, p. 3–10.
- MELCHERS, R. 1990, «Search-based importance sampling», *Structural safety*, vol. 9, n° 2, p. 117–128.
- MELCHERS, R. E. et A. T. BECK. 2018, *Structural reliability analysis and prediction*, John Wiley & Sons.

- MENZ, M., S. DUBREUIL, J. MORIO, C. GOGU, N. BARTOLI et M. CHIRON. 2020, «Variance based sensitivity analysis for Monte Carlo and importance sampling reliability assessment with Gaussian processes», *arXiv preprint arXiv :2011.15001*.
- MORIO, J. 2011, «Non-parametric adaptive importance sampling for the probability estimation of a launcher impact position», *Reliability engineering & system safety*, vol. 96, n° 1, p. 178–183.
- MORIO, J. 2012, «Extreme quantile estimation with nonparametric adaptive importance sampling», *Simulation Modelling Practice and Theory*, vol. 27, p. 76–89.
- MORIO, J. et M. BALESSENT. 2016, «Estimation of a launch vehicle stage fallout zone with parametric and non-parametric importance sampling algorithms in presence of uncertain input distributions», *Aerospace Science and Technology*, vol. 52, p. 95–101.
- NAGLER, T. et C. CZADO. 2016, «Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas», *Journal of Multivariate Analysis*, vol. 151, p. 69–89.
- NATAF, A. 1962, «Détermination des distribution dont les marges sont données», *Comptes Rendus de l'Academie des Sciences*, vol. 225, p. 42–43.
- NEDDERMEYER, J. C. 2009, «Computationally efficient nonparametric importance sampling», *Journal of the American Statistical Association*, vol. 104, n° 486, p. 788–802.
- NEDDERMEYER, J. C. 2011, «Non-parametric partial importance sampling for financial derivative pricing», *Quantitative Finance*, vol. 11, n° 8, p. 1193–1206.
- NIEDERREITER, H. 1992, *Random number generation and quasi-Monte Carlo methods*, SIAM.
- OH, M.-S. et J. O. BERGER. 1992, «Adaptive importance sampling in Monte Carlo integration», *Journal of Statistical Computation and Simulation*, vol. 41, n° 3-4, p. 143–168.
- OH, M.-S. et J. O. BERGER. 1993, «Integration of multimodal functions by Monte Carlo importance sampling», *Journal of the American Statistical Association*, vol. 88, n° 422, p. 450–456.
- OWEN, A. et Y. ZHOU. 2000, «Safe and effective importance sampling», *Journal of the American Statistical Association*, vol. 95, n° 449, p. 135–143.
- PAPAIOANNOU, I., S. GEYER et D. STRAUB. 2019, «Improved cross entropy-based importance sampling with a flexible mixture model», *Reliability Engineering & System Safety*, vol. 191, p. 106–116.
- PEARSON, K. 1920, «Notes on the history of correlation», *Biometrika*, vol. 13, n° 1, p. 25–45.
- PICKANDS, J. et collab.. 1975, «Statistical inference using extreme order statistics», *Annals of statistics*, vol. 3, n° 1, p. 119–131.
- RESNICK, S. I. 2013, *Extreme values, regular variation and point processes*, Springer.
- ROSENBLATT, M. 1952, «Remarks on a multivariate transformation», *The annals of mathematical statistics*, vol. 23, n° 3, p. 470–472.
- RUBINO, G. et B. TUFFIN. 2009, *Rare event simulation using Monte Carlo methods*, John Wiley & Sons.
- RUBINSTEIN, R. Y. et P. W. GLYNN. 2009, «How to deal with the curse of dimensionality of likelihood ratios in Monte Carlo simulation», *Stochastic Models*, vol. 25, n° 4, p. 547–568.
- RUBINSTEIN, R. Y. et D. P. KROESE. 2004, *The Cross-Entropy Method : a Unified Approach to Combinatorial Optimization, Monte Carlo Simulation and Machine Learning*, Springer Science & Business Media.

- RUBINSTEIN, R. Y. et D. P. KROESE. 2016, *Simulation and the Monte Carlo method*, vol. 10, John Wiley & Sons.
- SCHUËLLER, G., H. PRADLWARTER et P.-S. KOUTSOURELAKIS. 2004, «A critical appraisal of reliability estimation procedures for high dimensions», *Probabilistic engineering mechanics*, vol. 19, n° 4, p. 463–474.
- SCHUËLLER, G. I. et R. STIX. 1987, «A critical appraisal of methods to determine failure probabilities», *Structural Safety*, vol. 4, n° 4, p. 293–309.
- SCOTT, D. W. 1991, «Feasibility of multivariate density estimates», *Biometrika*, vol. 78, n° 1, p. 197–205.
- SHINOZUKA, M. 1983, «Basic analysis of structural safety», *Journal of Structural Engineering*, vol. 109, n° 3, p. 721–740.
- SOBOL, I. M. 1967, «On the distribution of points in a cube and the approximate evaluation of integrals», *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, vol. 7, n° 4, p. 784–802.
- STRAUB, D. 2014, «Engineering risk assessment», dans *Risk-A multidisciplinary introduction*, Springer, p. 333–362.
- TERRELL, G. R. et D. W. SCOTT. 1985, «Oversmoothed nonparametric density estimates», *Journal of the American Statistical Association*, vol. 80, n° 389, p. 209–214.
- URIBE, F., I. PAPAIOANNOU, Y. M. MARZOUK et D. STRAUB. 2020, «Cross-entropy-based importance sampling with failure-informed dimension reduction for rare event simulation», *arXiv preprint arXiv:2006.05496*.
- VEACH, E. 1997, *Robust Monte Carlo methods for light transport simulation*, thèse de doctorat, Stanford University.
- VEACH, E. et L. GUIBAS. 1995a, «Bidirectional estimators for light transport», dans *Photorealistic Rendering Techniques*, Springer, p. 145–167.
- VEACH, E. et L. J. GUIBAS. 1995b, «Optimally combining sampling techniques for Monte Carlo rendering», dans *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, p. 419–428.
- WAND, M. P. et M. C. JONES. 1994, *Kernel smoothing*, CRC press.
- WANG, Z. et J. SONG. 2016, «Cross-entropy-based adaptive importance sampling using Von Mises-Fisher mixture for high dimensional reliability analysis», *Structural Safety*, vol. 59, p. 42–52.
- WEST, M. 1993, «Approximating posterior distributions by mixtures», *Journal of the Royal Statistical Society : Series B (Methodological)*, vol. 55, n° 2, p. 409–422.
- ZHANG, P. 1996, «Nonparametric importance sampling», *Journal of the American Statistical Association*, vol. 91, n° 435, p. 1245–1253.

# Chapitre 3

## Analyse de sensibilité fiabiliste

### Sommaire

---

<b>3.1 Introduction</b> . . . . .	<b>106</b>
<b>3.2 Analyse de sensibilité de la sortie d'un code de calcul</b> . . . . .	<b>108</b>
3.2.1 Analyse de sensibilité locale et criblage . . . . .	108
3.2.2 Analyse de sensibilité globale pour les variables indépendantes . . . . .	109
3.2.3 Analyse de sensibilité globale pour les variables dépendantes . . . . .	113
3.2.4 Bilan . . . . .	120
<b>3.3 Application à l'analyse de sensibilité d'une probabilité de défaillance</b> . . . . .	<b>121</b>
3.3.1 Modélisation des incertitudes épistémiques . . . . .	121
3.3.2 Méthodes locales pour l'analyse de sensibilité fiabiliste . . . . .	122
3.3.3 Méthodes globales pour l'analyse de sensibilité fiabiliste . . . . .	124
3.3.4 Bilan . . . . .	127
<b>3.4 Conclusion</b> . . . . .	<b>128</b>

---

### 3.1 Introduction

Lorsqu'on souhaite approcher la distribution de probabilité conjointe gouvernant le comportement simultané des entrées aléatoires d'un problème, le Chapitre 1 a mis en avant le rôle des algorithmes d'apprentissage statistique qui permettent de construire un modèle probabiliste à partir de la seule connaissance d'un échantillon  $\mathbf{X}_{\text{obs}}$  constitué d'un petit nombre d'observations. Les algorithmes de fiabilité présentés au Chapitre 2 sont ensuite couplés à la densité  $\hat{f}_{\mathbf{X}}(\cdot | \mathcal{M}, \mathbf{X}_{\text{obs}})$  issue de l'apprentissage et produisent une estimation  $\hat{P}_f(\mathbf{X}_{\text{obs}})$  de la probabilité de défaillance du système étudié. La spécificité de ce travail de thèse se trouve dans l'étude du transfert d'incertitude entre  $\mathbf{X}_{\text{obs}}$  et  $\hat{P}_f(\mathbf{X}_{\text{obs}})$ . La variabilité initiale d'échantillonnage se transforme en une variabilité sur la prédiction du risque du défaillance. Dans un contexte de "small data", deux échantillons  $\mathbf{x}_{\text{obs}}^{(1)}$  et  $\mathbf{x}_{\text{obs}}^{(2)}$  peuvent alors mener à des estimations  $\hat{P}_f(\mathbf{x}_{\text{obs}}^{(1)})$  et  $\hat{P}_f(\mathbf{x}_{\text{obs}}^{(2)})$  sensiblement différentes. Ce manque de robustesse est évidemment préjudiciable car l'utilisateur est dans l'impossibilité de se fier au résultat final de la chaîne de prédiction. Si la variabilité induite par  $\mathbf{X}_{\text{obs}}$  est d'une grande amplitude, on peut s'interroger sur la pertinence de l'information apportée par une unique estimation. La solution explorée dans cette thèse consiste à déterminer une stratégie pour *in fine* collecter des observations supplémentaires pour la variable  $X_k$  ou la paire de variables  $(X_i, X_j)$  dont la variabilité d'échantillonnage a le plus d'impact sur la dispersion de  $\hat{P}_f(\mathbf{X}_{\text{obs}})$ . Dans l'un ou l'autre des cas, il faut résoudre un problème d'analyse de sensibilité fiabiliste pour identifier le sous-ensemble de  $\mathbf{X}_{\text{obs}}$  dont la base d'observations doit être enrichie en priorité.

L'analyse de sensibilité (AS) est une thématique qui a éveillé un intérêt croissant au cours des trente dernières années. Dans de nombreuses disciplines, c'est un outil essentiel pour la conception et la compréhension des modèles physiques. Les grands principes et les bonnes pratiques de cette vaste discipline sont éclaircis dans SALTELLI et collab. [2004]. D'une manière générale, l'analyse de sensibilité peut se prêter à tout contexte de type :

$$\Psi: \begin{array}{ccccccc} E_1 & \times \dots \times & E_d & \longrightarrow & S_1 & \times \dots \times & S_r \\ e_1 & \dots & e_d & \longmapsto & s_1 & \dots & s_r \end{array}, \quad (3.1)$$

avec  $\Psi$  une application mesurable pour les tribus accompagnant les espaces  $E := E_1 \times \dots \times E_d$  et  $S := S_1 \times \dots \times S_r$ . Quel que soit le contexte, l'AS intervient comme la troisième et dernière étape d'une méthodologie globale qui cherche à :

1. Spécifier la loi  $\mu_e$  des entrées  $\mathbf{e} := (e_1, \dots, e_d)$ .
2. Propager  $\mu_e$  à travers  $\Psi$  pour engendrer la loi  $\mu_s$  des sorties  $\mathbf{s} := (s_1, \dots, s_r)$ .
3. Quantifier l'influence de chaque entrée  $e_j$  sur une ou plusieurs sorties.

Naturellement, les premières idées ont été développées, expérimentées et validées dans le contexte mathématique le plus favorable, c'est-à-dire pour  $E \subseteq \mathbb{R}^d$  et  $S \subseteq \mathbb{R}$ . Ce format correspond à celui de la fonction de performance  $\phi$  du Chapitre 2. De plus, le problème d'AS est bien plus simple à traiter lorsque les entrées sont indépendantes et que la fonction  $\phi$  est à la fois déterministe et à faible coût de calcul. L'approche fondatrice de SOBOL [1993] consiste à concevoir la variance de la sortie comme la somme des contributions des différents sous-groupes pouvant être formés à partir des  $d$  variables d'entrée. Comme le soulignent HELTON et collab. [2006b], les méthodes d'AS reposent principalement sur des stratégies d'échantillonnage et suivent un plan d'exécution quasiment identique. Tout d'abord, il faut commencer par choisir (ou par définir) des indices de sensibilité compatibles avec la loi  $\mu_{\mathbf{X}}$  et le budget de simulation. Des échantillons sont ensuite simulés dans l'espace d'entrée et les sorties associées sont calculées avec  $\phi$ . Une phase finale de post-traitement fournit l'estimation des indices. IOOSS et SALTELLI [2017] rappellent les principales motivations de la construction d'une procédure d'AS :

- **Classement de l'influence des entrées.** Le calcul des indices de sensibilité permet d'établir une hiérarchie entre les entrées  $X_j$  à partir d'une mesure de leur impact sur la sortie  $Y$ .

- **Réduction de la variance de la sortie.** Si on souhaite que  $\mathbb{V}(Y)$  soit inférieure à une valeur prescrite, les plus grands indices de sensibilité indiquent les variables  $X_j$  sur lesquelles il faut agir en priorité (en cherchant à réduire leur niveau de dispersion).
- **Réduction de la dimensionnalité.** Un très petit indice de sensibilité signifie que la variabilité de l'entrée considérée n'a pratiquement aucun effet sur celle de la sortie. On peut alors envisager de fixer cette entrée à une valeur nominale. Le modèle physique est ainsi délesté d'une entrée aléatoire qui devient constante mais qui continue d'intervenir dans  $\phi$ .
- **Compréhension du modèle physique.** De nombreuses techniques d'AS permettent non seulement de quantifier l'effet d'une entrée seule sur la sortie mais aussi de quantifier l'effet conjoint d'un groupe d'entrées. L'AS permet ainsi de mieux appréhender les interactions au sein du modèle  $\phi$  et donc d'affiner les connaissances qu'on peut en avoir.

Effectuer une AS en présence d'entrées indépendantes et d'une fonction  $\phi$  analytique est devenu une tâche assez élémentaire pour laquelle on dispose d'outils performants. Toutefois, ce cadre initial est en fait assez éloigné des défis rencontrés en pratique lors de la modélisation de systèmes complexes. De ce fait, les efforts de recherche les plus récents en AS ont pour objectif d'améliorer les méthodologies existantes afin de se départir des hypothèses les moins réalistes.

- **Entrées dépendantes.** Cette situation a été examinée dès [SALTELLI et TARANTOLA \[2002\]](#) et de nombreuses méthodes ont vu le jour par la suite pour se substituer à la méthode originale de [SOBOL \[1993\]](#) qui devient inopérante lorsqu'il existe des corrélations entre les entrées. Des stratégies alternatives pour décomposer la variance de la sortie ont été proposées par [MARA et TARANTOLA \[2012\]](#), [CHASTAING et collab. \[2012\]](#) ou [SUDRET et CANIOU \[2013\]](#) mais les réponses les plus convaincantes sont apparues grâce à la définition d'indices basés sur des mesures de dissimilarité entre la loi de la sortie  $\mu_Y$  et les différentes lois conditionnelles  $\mu_{Y|X_j}$ . Bien que les indices de [BORGONOVO \[2007\]](#) soient les plus emblématiques de cette gamme, d'autres indices peuvent être créés de manière analogue et un aperçu en est donné dans [DA VEIGA \[2015\]](#) ou [BORGONOVO et collab. \[2016\]](#).
- **Code de calcul excessivement coûteux.** Le nombre d'appels à  $\phi$  autorisé par le budget de simulation devient trop petit pour que les résultats de l'AS soient précis. Il faut commencer par construire un métamodèle  $\hat{\phi}$  à partir d'une base d'apprentissage rassemblant un petit nombre d'entrées et de sorties de  $\phi$ . On peut trouver dans [LE GRATIET et collab. \[2017\]](#) une confrontation des deux familles de métamodèles les plus fréquemment utilisés : les polynômes de chaos et les processus gaussiens. Ils ont été respectivement introduits au service de l'AS par [SUDRET \[2008, 2015\]](#) et [MARREL et collab. \[2008\]](#). Souvent, l'application  $\hat{\phi}$  est stochastique. Cela signifie qu'il existe une incertitude sur la valeur prise par  $\hat{\phi}(\mathbf{x}_0)$  pour une valeur  $\mathbf{x}_0 \in \mathbb{R}^d$  fixée. La maîtrise de la variabilité liée à la prédiction de la sortie est le point de départ des travaux de [MARREL et collab. \[2012\]](#) et de [HART et collab. \[2017\]](#). De plus, les contributions de [ZHU et SUDRET \[2020\]](#) et de [FORT et collab. \[2020\]](#) témoignent de l'intérêt que ce sujet continue de susciter.
- **Entrées ou sorties fonctionnelles.** Souvent, la modélisation impose de considérer des entrées ou des sorties qui ne sont pas simplement des scalaires. [IOOSS et RIBATET \[2009\]](#) de même que [NANTY et collab. \[2016\]](#) se sont intéressés à la manipulation d'entrées fonctionnelles (éventuellement dépendantes) tandis que [GAMBOA et collab. \[2013, 2014\]](#) ont généralisé l'emploi des indices de Sobol à des sorties vectorielles ou fonctionnelles.

Cette thèse va au-delà de la propagation à travers  $\phi$  des incertitudes aléatoires du vecteur  $\mathbf{X}$ . Comme la loi  $\mu_{\mathbf{X}}$  est inconnue, une phase d'apprentissage de  $\mu_{\mathbf{X}}$  est nécessaire avant de pouvoir propager les incertitudes aléatoires à travers  $\phi$ . L'AS survient dans le cadre de ce second niveau d'incertitude (qualifié d'épistémique) qui est lié au manque de connaissances sur  $\mu_{\mathbf{X}}$  et à la variabilité de l'échantillon  $\mathbf{X}_{\text{obs}}$  qui permet d'en acquérir. La chaîne de propagation des incertitudes épistémiques peut être résumée par :

$$\mathbf{X}_{\text{obs}} \longmapsto \hat{f}_{\mathbf{X}}(\cdot | \mathcal{M}, \mathbf{X}_{\text{obs}}) \longmapsto \hat{P}_f(\mathbf{X}_{\text{obs}}) . \quad (3.2)$$

Comme la variable d'intérêt en fin de chaîne est une probabilité de défaillance, on parle d'analyse de sensibilité fiabiliste (ASF). Lorsque la densité  $f_{\mathbf{x}}$  est connue,  $\hat{P}_f$  n'est rendue aléatoire que par le bruit engendré par un éventuel échantillonnage de particules au sein de l'algorithme de fiabilité. D'ailleurs, si celui-ci est déterministe (FORM/SORM),  $\hat{P}_f$  l'est tout autant. Dans cette thèse,  $\hat{P}_f(\mathbf{X}_{\text{obs}})$  provient d'un algorithme de fiabilité stochastique et dépend du résultat d'un processus de collecte de données. Dans l'intérêt de cette étude, on suppose que les incertitudes épistémiques véhiculées par  $\mathbf{X}_{\text{obs}}$  sont très supérieures au bruit intrinsèque de l'algorithme de fiabilité. La question de la propagation d'incertitudes épistémiques sur une probabilité de défaillance a déjà été étudiée dans le cas d'une densité  $f_{\mathbf{x}}(\cdot | \mathcal{F}, \Theta)$  dont les paramètres  $\Theta$  sont incertains. La variable d'intérêt est alors  $\hat{P}_f(\Theta)$  et il existe différentes approches pour caractériser l'aléa sur  $\Theta$ . Mener une AS de l'application  $\Psi : \Theta \rightarrow P_f(\Theta)$  permet d'identifier le paramètre  $\Theta_j$  dont la dispersion est la plus préjudiciable à la robustesse de la prédiction fiabiliste. D'une manière générale, la principale difficulté d'une ASF vient du fait qu'elle exige d'estimer des centaines voire des milliers de probabilités de défaillance, chaque estimation exigeant à son tour des milliers d'appels à la fonction de performance  $\phi$ . La gestion du budget de calcul est ainsi une préoccupation centrale de l'ASF.

Dans un premier temps, la Section 3.2 est consacrée à la présentation des méthodes incontournables en AS. On détaille avec plus d'exhaustivité les outils utilisés dans les Chapitres 4 et 5. Dans un second temps, on introduit à la Section 3.3 les contributions les plus notables en ASF. Certaines découlent de l'application directe des méthodes d'AS à l'estimateur d'une probabilité de défaillance tandis que d'autres sont plus spécifiques au point de vue fiabiliste.

## 3.2 Analyse de sensibilité de la sortie d'un code de calcul

### 3.2.1 Analyse de sensibilité locale et criblage

L'AS locale consiste à étudier les variations de la sortie lorsqu'une perturbation est appliquée en un point  $\mathbf{x}^*$ . Dans la suite, on note  $\mathcal{V}_{\mathbf{x}^*}$  un voisinage immédiat de ce point. La perturbation doit être comprise comme un petit déplacement de  $\mathbf{x}^*$  vers  $\mathbf{x} \in \mathcal{V}_{\mathbf{x}^*}$  et le déplacement de chaque coordonnée  $x_j^*$  entraîne une modification de la valeur en sortie. L'objectif est d'identifier l'entrée dont la perturbation engendre la variation de plus grande amplitude en sortie. Une première appréciation de la sensibilité de la sortie aux perturbations affectant les entrées est obtenue à partir du gradient  $\nabla_{\mathbf{x}}\phi(\mathbf{x}^*)$ . En effet, un développement de Taylor d'ordre 1 de  $\phi$  donne :

$$\forall \mathbf{x} \in \mathcal{V}_{\mathbf{x}^*}, \quad \phi(\mathbf{x}) - \phi(\mathbf{x}^*) = \sum_{j=1}^d \delta_j \frac{\partial \phi}{\partial x_j}(\mathbf{x}^*) + o(\|\mathbf{x} - \mathbf{x}^*\|_2) \quad \text{avec} \quad \delta_j = x_j - x_j^*. \quad (3.3)$$

On suppose que  $\mathcal{V}_{\mathbf{x}^*}$  est suffisamment restreint pour que le reste de Taylor contenant les termes d'ordre supérieur ou égal à 2 puisse être négligé. Dans le formalisme de HELTON [1993], des indices de sensibilité locaux sont définis à partir des dérivées partielles au point  $\mathbf{x}^*$  :

$$\forall 1 \leq j \leq d, \quad \Delta_j(\mathbf{x}^*) = \frac{\partial \phi}{\partial x_j}(\mathbf{x}^*). \quad (3.4)$$

Lorsque  $\phi$  est connue analytiquement, les dérivées partielles s'obtiennent avec une différentiation manuelle ou automatique. Par contre, lorsque  $\phi$  est une véritable boîte noire, il est expliqué dans IOTT et collab. [1985] qu'il faut avoir recours à des schémas aux différences finies. On peut par exemple construire un plan OAT (*One-variable-At-a-Time*) pour  $\mathbf{x}^*$ . Soit  $(\mathbf{e}_1, \dots, \mathbf{e}_d)$  la base canonique de  $\mathbb{R}^d$ . On rassemble dans  $\mathcal{X}_{\text{OAT}}(\mathbf{x}^*)$  l'ensemble des points  $\mathbf{x}^* + \delta_j \mathbf{e}_j$  pour  $j \in \{1, \dots, d\}$ . Chacun de ces points est le résultat d'un déplacement unidirectionnel d'amplitude  $\delta_j$ . Après avoir calculé les sorties associées à  $\mathbf{x}^*$  et  $\mathcal{X}_{\text{OAT}}(\mathbf{x}^*)$ , on peut approcher chaque indice  $\Delta_j(\mathbf{x}^*)$  avec :

$$\forall 1 \leq j \leq d, \quad \hat{\Delta}_j(\mathbf{x}^*) = \frac{\phi(\mathbf{x}^* + \delta_j \mathbf{e}_j) - \phi(\mathbf{x}^*)}{\delta_j}, \quad (3.5)$$

et l'erreur commise dépend de la régularité de  $\phi$  et du caractère "infinitésimal" du pas  $\delta_j$ . En général, les indices locaux ne peuvent pas être comparés car ils ne sont pas exprimés dans la même unité physique. Dans [BORGONOVO et APOSTOLAKIS \[2001\]](#), cet écueil est évité en considérant les quantités  $\Delta_j(\mathbf{x}^*) \delta_j$  après qu'elles ont été renormalisées. De même, [BORGONOVO et PECCATI \[2004\]](#) suggèrent de calculer l'élasticité de la sortie par rapport à sa  $j$ -ème entrée pour se ramener à un indice local sans unité.

L'AS locale perd de son intérêt dès lors que le domaine de variation des entrées ne se limite plus au voisinage  $\mathcal{V}_{\mathbf{x}^*}$  mais qu'il s'étend à tout le support de la densité conjointe  $f_{\mathbf{X}}$ . En s'appuyant sur l'approche locale tout en cherchant à l'améliorer, les méthodes de criblage consistent à calculer des indices locaux pour toute une collection de points du support. Il est alors supposé<sup>1</sup> que les différentes lois marginales  $\mu_{X_j}$  sont uniformes. Le support de  $f_{\mathbf{X}}$  est donc  $\mathcal{H}_d$  et les stratégies permettant de mettre en place un plan d'expérience  $\mathcal{X}$  de taille  $N_s$  dans ce domaine compact sont passées en revue dans [WOODS et LEWIS \[2017\]](#). La méthode proposée par [MORRIS \[1991\]](#) cherche à calculer la collection des indices locaux en tout point  $\mathbf{x}^{(i)}$  de  $\mathcal{X}$ . L'indice  $\hat{\Delta}_j(\mathbf{x}^{(i)})$  est appelé l'effet élémentaire de  $\mathbf{x}^{(i)}$  dans la direction  $\mathbf{e}_j$ . Appliquer la procédure de Morris impose le recours à un plan OAT en chaque point  $\mathbf{x}^{(i)}$  avec notamment  $\delta_0 := \delta_1 = \dots = \delta_d$ . Pour une entrée  $X_j$  donnée, afin de résumer l'information apportée par les effets élémentaires, [CAMPOLONGO et collab. \[2007\]](#) se sont intéressés à deux moments empiriques :

$$\forall 1 \leq j \leq d, \quad \hat{\mu}_{|\hat{\Delta}_j|} := \frac{1}{N_s} \sum_{i=1}^{N_s} \left| \hat{\Delta}_j(\mathbf{x}^{(i)}) \right| \quad \text{et} \quad \hat{\sigma}_{\hat{\Delta}_j}^2 := \frac{1}{N_s} \sum_{i=1}^{N_s} \left( \hat{\Delta}_j(\mathbf{x}^{(i)}) \right)^2 - \hat{\mu}_{\hat{\Delta}_j}^2 . \quad (3.6)$$

Le premier sert à mesurer l'influence de la  $j$ -ème entrée tandis que le second vise à mesurer le degré de non-linéarité de  $\phi$  sur  $\mathcal{H}_d$ . En effet, une forte variabilité spatiale des indices  $\hat{\Delta}_j(\mathbf{x}^{(i)})$  signale qu'il n'existe pas d'approximation linéaire convenable de  $\phi$  sur l'ensemble du domaine. Dans ce contexte, l'utilisation d'approximations linéaires locales pour fabriquer un indice global peut être discutable. Cette première méthode de criblage peut être généralisée à l'aide des DGSM (*Derivative-based Global Sensitivity Measures*) apparues dans [SOBOL et GRESHAM \[1995\]](#) et définies sous cette étiquette dans [SOBOL et KUCHERENKO \[2010\]](#) :

$$\forall 1 \leq j \leq d, \quad v_j = \mathbb{E} \left[ (\Delta_j(\mathbf{X}))^2 \right] = \int_{\mathcal{H}_d} \left( \frac{\partial \phi}{\partial x_j}(\mathbf{x}) \right)^2 f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x} . \quad (3.7)$$

D'une manière assez naturelle, [KUCHERENKO et collab. \[2009\]](#) proposent d'estimer les DGSM à l'aide d'une approche de Monte-Carlo utilisant le schéma (3.5) pour le calcul des dérivées partielles :

$$\forall 1 \leq j \leq d, \quad \hat{v}_j = \frac{1}{N_s} \sum_{i=1}^{N_s} \left( \frac{\phi(\mathbf{X}^{(i)} + \delta_j \mathbf{e}_j) - \phi(\mathbf{X}^{(i)})}{\delta_j} \right)^2 \quad \text{avec} \quad \mathbf{X}^{(i)} \stackrel{\text{i.i.d.}}{\sim} f_{\mathbf{X}} . \quad (3.8)$$

D'un point de vue théorique, l'inégalité de [POINCARÉ \[1890\]](#) permet de majorer l'indice de Sobol d'ordre total  $S_j^T$  grâce à une borne  $Cv_j$  où  $C$  est une constante indépendante de  $\phi$ . Les relations étroites qu'entretiennent les procédures d'AS basées sur le calcul du gradient de la fonction de performance et celles basées sur la décomposition de la variance de la sortie sont décortiquées dans [LAMBONI et collab. \[2013\]](#), [ROUSTANT et collab. \[2014\]](#) et [KUCHERENKO et IOOSS \[2017\]](#). Par ailleurs, on trouve dans [KUCHERENKO et collab. \[2010\]](#) une adaptation des DGSM qui est destinée à la quantification de l'influence conjointe de plusieurs entrées.

La méthode de Morris ainsi que les DGSM s'inscrivent dans une volonté d'impliquer la totalité de la plage de variation des entrées au sein de la procédure d'AS tout en préservant le rôle central des indices locaux. Ce compromis est atteint en définissant de nouveaux indices qui se veulent être la synthèse d'indices locaux calculés en différents points de l'espace. Comme le coût de l'estimation d'un indice local est extrêmement réduit, ces deux méthodes de criblage permettent d'obtenir à moindre coût une réponse qualitative au problème d'AS, c'est-à-dire une réponse dont on peut

1. On peut toujours s'y ramener en posant  $Y = \phi(\mathbf{X}) = \phi(F_{X_1}^{-1}(U_1), \dots, F_{X_d}^{-1}(U_d)) = \phi_0(\mathbf{U})$  avec  $U_j \sim \mathcal{U}([0, 1])$ .

se servir pour discriminer les entrées selon qu'elles aient une petite, une moyenne ou une grande influence sur la sortie. Dans les deux prochains paragraphes, on se penche sur les méthodes d'AS globale en distinguant celles qui sont spécifiques aux entrées indépendantes et celles qui peuvent être appliquées en toutes circonstances.

### 3.2.2 Analyse de sensibilité globale pour les variables indépendantes

Dans ce paragraphe, on reste dans un cadre où  $\phi : \mathcal{H}_d \rightarrow \mathbb{R}$  et où les lois marginales  $\mu_{X_j}$  sont uniformes. Par ailleurs, il est supposé que les variables  $X_j$  sont indépendantes. L'approche la plus répandue repose sur un partage de la variance de la sortie  $\sigma^2 = \mathbb{V}(Y)$ . Soit  $\mathbf{u} := \{i_1, \dots, i_r\}$  une liste d'indices telle que  $r < d$ . Son ensemble complémentaire dans  $[d] := \{1, \dots, d\}$  est noté  $-\mathbf{u} := \{j_1, \dots, j_{d-r}\}$ . On introduit alors les notations suivantes :

$$|\mathbf{u}| := \text{Card}(\mathbf{u}) = r \quad \text{tandis que} \quad \mathbf{X}_{\mathbf{u}} := [X_{i_k}]_{1 \leq k \leq r} \quad \text{et} \quad \mathbf{X}_{-\mathbf{u}} := [X_{j_k}]_{1 \leq k \leq d-r} . \quad (3.9)$$

Les prémices de la théorie de l'AS globale remontent aux différents indices proposés par :

- **HORA et IMAN [1986]** :  $\sqrt{\mathbb{V}(Y) - \mathbb{E}[\mathbb{V}(Y | X_j)]}$  ;
- **IMAN et HORA [1990]** :  $\mathbb{V}(\mathbb{E}[\log Y | X_j]) / \mathbb{V}(\log Y)$  ;
- **MCKAY [1997]** :  $\mathbb{V}(\mathbb{E}[Y | X_j]) / \mathbb{V}(Y)$  .

Pour clarifier le rôle des différents groupes d'entrées dans le calcul de la sortie, **HOEFFDING [1948]** suggère d'écrire  $\phi$  comme la somme de leurs apports respectifs :

$$Y = \phi(\mathbf{X}) = \sum_{\mathbf{u} \subseteq [d]} \eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}) = \eta_0 + \sum_{j=1}^d \eta_j(X_j) + \sum_{j < k} \eta_{jk}(X_j, X_k) + \dots + \eta_{1\dots d}(X_1, \dots, X_d) . \quad (3.10)$$

Cette décomposition HDMR (*High-Dimensional Model Representation*) comporte autant de termes  $\eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}})$  que de sous-ensembles  $\mathbf{u} \subseteq [d]$ , à savoir  $2^d$ . Si aucune contrainte supplémentaire n'est imposée, il existe une infinité de façons d'identifier les fonctions  $\eta_{\mathbf{u}}$  à partir de  $\phi$ . Une condition d'orthogonalité est énoncée par **SOBOL [1993]** :

$$\forall \mathbf{u} := \{i_1, \dots, i_r\}, \quad \forall 1 \leq s \leq r, \quad \int_0^1 \eta_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}}) dx_{i_s} = 0 . \quad (3.11)$$

Il existe alors une unique décomposition (3.10) vérifiant cette condition. Il faut prendre  $\eta_0 = \mathbb{E}[Y]$  et les fonctions associées aux singletons et aux paires d'indices s'obtiennent en posant :

$$\eta_j(X_j) = \mathbb{E}[Y | X_j] - \eta_0 \quad \text{et} \quad \eta_{jk}(X_j, X_k) = \mathbb{E}[Y | X_j, X_k] - \eta_j(X_j) - \eta_k(X_k) - \eta_0 . \quad (3.12)$$

D'une manière générale, la fonction  $\eta_{\mathbf{u}}$  s'identifie de la manière suivante :

$$\eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}) = \mathbb{E}[Y | \mathbf{X}_{\mathbf{u}}] - \sum_{\mathbf{v} \subset \mathbf{u}} \eta_{\mathbf{v}}(\mathbf{X}_{\mathbf{v}}) = \mathbb{E}[Y | \mathbf{X}_{\mathbf{u}}] - \sum_{\mathbf{v} \subset \mathbf{u}} (-1)^{|\mathbf{u}| - |\mathbf{v}|} \mathbb{E}[Y | \mathbf{X}_{\mathbf{v}}] . \quad (3.13)$$

Ici, comme  $\mathbf{X}$  est composé de variables indépendantes,  $\mathbb{E}[Y | \mathbf{X}_{\mathbf{u}}]$  est également l'espérance de  $Y$  par rapport aux seules variables  $\mathbf{X}_{-\mathbf{u}}$ . Si on utilise les termes (3.13), la décomposition de Hoeffding est appelée décomposition ANOVA (*ANalysis Of VAriance*) ou décomposition RS-HDMR<sup>2</sup>. Si on relâche la condition d'orthogonalité, une alternative à ANOVA est la décomposition *Cut*-HDMR présentée notamment dans **RABITZ et collab. [1999]**. On la construit en projetant la fonction  $\phi$  dans des sous-espaces affines partageant un point d'ancrage commun. Dans le cas d'ANOVA, on parle d'orthogonalité car la condition (3.11) implique la décorrélation des variables  $\eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}})$ . Un passage à la variance dans la décomposition de Hoeffding donne alors :

$$\sigma^2 = \sum_{\mathbf{u} \subseteq [d]} \sigma_{\mathbf{u}}^2 \quad \text{avec} \quad \sigma_{\mathbf{u}}^2 := \mathbb{V}(\eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}})) . \quad (3.14)$$

---

2. (*Random Sampling*)-HDMR

Les indices de Sobol “partiels” et “fermés” sont alors définis par :

$$S_{\mathbf{u}} := \frac{\mathbb{V}(\eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}))}{\mathbb{V}(Y)} = \frac{\sigma_{\mathbf{u}}^2}{\sigma^2} \quad \text{et} \quad S_{\mathbf{u}}^c := \frac{\mathbb{V}(\mathbb{E}[Y | \mathbf{X}_{\mathbf{u}}])}{\mathbb{V}(Y)} = \frac{1}{\sigma^2} \sum_{\mathbf{v} \subseteq \mathbf{u}} \sigma_{\mathbf{v}}^2. \quad (3.15)$$

Dans la littérature, il demeure souvent une ambiguïté quant à la définition à donner aux “indices de Sobol” en l’absence d’un qualificatif supplémentaire. Cela vient notamment du fait que les indices partiels et fermés coïncident au premier ordre, avec  $S_j = S_j^c$  pour tout  $j \in \{1, \dots, d\}$ . L’indice  $S_{\mathbf{u}}^c$  est la somme de toutes les variances  $\sigma_{\mathbf{v}}^2$  avec  $\mathbf{v} \subseteq \mathbf{u}$ . Si on décidait de “geler” le sous-vecteur  $\mathbf{X}_{-\mathbf{u}}$ , la variance résiduelle de la sortie serait précisément égale à  $\mathbb{V}(\mathbb{E}[Y | \mathbf{X}_{\mathbf{u}}])$ . On dit que  $S_{\mathbf{u}}^c$  quantifie l’effet principal des entrées  $\mathbf{X}_{\mathbf{u}}$  sur  $Y$ . Dans la définition de cet indice fermé, on ignore les parts de variance que peuvent créer les variables de  $\mathbf{X}_{\mathbf{u}}$  lorsqu’elles interagissent avec les variables de  $\mathbf{X}_{-\mathbf{u}}$ . Ces termes d’interaction correspondent aux variables  $\eta_{\mathbf{v}}(\mathbf{X}_{\mathbf{v}})$  pour lesquelles  $\mathbf{v} \not\subseteq \mathbf{u}$  mais  $\mathbf{v} \cap \mathbf{u} \neq \emptyset$ . Pour les inclure dans la mesure de l’influence de  $\mathbf{X}_{\mathbf{u}}$ , **HOMMA et SALTELLI [1996]** ont défini les indices de Sobol “totaux” :

$$S_{\mathbf{u}}^T := 1 - \frac{\mathbb{V}(\mathbb{E}[Y | \mathbf{X}_{-\mathbf{u}}])}{\mathbb{V}(Y)} = \frac{\mathbb{E}[\mathbb{V}(Y | \mathbf{X}_{-\mathbf{u}})]}{\mathbb{V}(Y)} = \frac{1}{\sigma^2} \sum_{\mathbf{v} \cap \mathbf{u} \neq \emptyset} \sigma_{\mathbf{v}}^2. \quad (3.16)$$

On a évidemment  $S_{\mathbf{u}}^T \geq S_{\mathbf{u}}^c$  car l’indice total s’obtient simplement en ajoutant à l’indice fermé les parts de variance engendrées par les interactions. Dans la suite, on ne s’intéresse qu’à la question de l’estimation des indices fermés car ils suffisent à estimer les indices partiels et totaux :

$$S_{\mathbf{u}} := S_{\mathbf{u}}^c + \sum_{\mathbf{v} \subseteq \mathbf{u}} (-1)^{|\mathbf{u}| - |\mathbf{v}|} S_{\mathbf{v}}^c \quad \text{et} \quad S_{\mathbf{u}}^T := 1 - S_{-\mathbf{u}}^c. \quad (3.17)$$

Si l’estimation de  $\sigma^2 = \mathbb{V}(Y)$  n’est pas une difficulté, il est moins évident d’estimer la variance de l’espérance conditionnelle. Dans la suite, on détaille les différentes stratégies pouvant être mises en œuvre pour estimer  $V_{\mathbf{u}} := \mathbb{V}(\mathbb{E}[Y | \mathbf{X}_{\mathbf{u}}])$ . Tout d’abord, on peut évoquer l’approche par force brute qui est suggérée par **McKAY [1996]**. Elle consiste à calculer la variance empirique des espérances conditionnelles  $m_k := \mathbb{E}[Y | \mathbf{X}_{\mathbf{u}} = \mathbf{x}_{\mathbf{u}}^{(k)}]$  issues d’une séquence  $\{\mathbf{x}_{\mathbf{u}}^{(k)}\}_{k=1}^{N_s}$  d’observations du sous-vecteur  $\mathbf{X}_{\mathbf{u}}$ . Pour estimer  $m_k$ , il suffit de suivre les étapes suivantes :

- Simuler une séquence d’observations  $\{\mathbf{x}_{-\mathbf{u}}^{(l)}\}_{l=1}^{N_s}$  de la loi  $\mathbf{X}_{-\mathbf{u}} | \mathbf{X}_{\mathbf{u}} = \mathbf{x}_{\mathbf{u}}^{(k)}$ .
- Calculer les sorties conditionnelles  $y_k^{(l)} = \phi(\mathbf{x}_{\mathbf{u}}^{(k)}, \mathbf{x}_{-\mathbf{u}}^{(l)})$  pour  $l \in \{1, \dots, N_s\}$ .
- Calculer la moyenne empirique des sorties conditionnelles :  $\hat{m}_k = \frac{1}{N_s} \sum_{l=1}^{N_s} y_k^{(l)}$ .

Comme les variables  $X_j$  sont indépendantes, l’étape (a) revient à simuler des observations de  $\mathbf{X}_{-\mathbf{u}}$ . D’ailleurs, la même séquence  $\{\mathbf{x}_{-\mathbf{u}}^{(l)}\}_{l=1}^{N_s}$  peut être utilisée pour estimer toutes les espérances conditionnelles. Cette procédure est connue sous le nom de CRM (*Correlation Ratio Method*). Bien qu’elle garantisse la convergence vers la valeur théorique de  $V_{\mathbf{u}}$ , elle consomme un budget bien trop conséquent d’appels à  $\phi$ . Dès les travaux de **SOBOL [1993]**, il est indiqué qu’on peut éradiquer la “double boucle” en remarquant que :

$$V_{\mathbf{u}} = \mathbb{E}[\phi(\mathbf{X}) \phi(\mathbf{X}_{\mathbf{u}}, \tilde{\mathbf{X}}_{-\mathbf{u}})] - \eta_0^2 \quad \text{avec} \quad \begin{cases} \mathbf{X}_{-\mathbf{u}} \perp \tilde{\mathbf{X}}_{-\mathbf{u}} \\ \mathbf{X}_{-\mathbf{u}} \stackrel{\text{loi}}{=} \tilde{\mathbf{X}}_{-\mathbf{u}} \end{cases}. \quad (3.18)$$

On parle de méthode “Pick-Freeze” (PF) car on passe de  $\phi(\mathbf{X})$  à  $\phi(\mathbf{X}_{\mathbf{u}}, \tilde{\mathbf{X}}_{-\mathbf{u}})$  en “gelant” le sous-vecteur  $\mathbf{X}_{\mathbf{u}}$  mais en “piochant” une copie indépendante  $\tilde{\mathbf{X}}_{-\mathbf{u}}$  du sous-vecteur  $\mathbf{X}_{-\mathbf{u}}$ . Supposons que l’on dispose de deux blocs  $\mathbf{X}_{\text{sim}}$  et  $\tilde{\mathbf{X}}_{\text{sim}}$  de  $N_s$  échantillons de la loi  $\mu_{\mathbf{X}}$ . Les estimateurs de  $V_{\mathbf{u}}$  et  $S_{\mathbf{u}}^c$  associés à la formule (3.18) sont donnés par :

$$\hat{V}_{\mathbf{u}}^{\text{PF}} = \frac{1}{N_s} \sum_{i=1}^{N_s} \phi(\mathbf{X}^{(i)}) \phi(\mathbf{X}_{\mathbf{u}}^{(i)}, \tilde{\mathbf{X}}_{-\mathbf{u}}^{(i)}) - \hat{\eta}_0^2 \quad \text{et} \quad \hat{S}_{\mathbf{u}}^{\text{PF}} = \frac{\hat{V}_{\mathbf{u}}^{\text{PF}}}{\hat{\sigma}^2}. \quad (3.19)$$

$\hat{V}_u^{\text{PF}}$  s'obtient au prix de  $2N_s$  évaluations de  $\phi$  alors que  $\hat{V}_u^{\text{CRM}}$  en requiert  $N_s^2$ . Les propriétés de  $\hat{S}_u^{\text{PF}}$  sont étudiées dans [JANON et collab. \[2014\]](#) où il est démontré que  $\hat{S}_u^{\text{PF}}$  est à la fois consistant et asymptotiquement normal. Par ailleurs, [GAMBOA et collab. \[2016\]](#) proposent un TCL gouvernant la convergence conjointe d'une collection d'indices de Sobol  $\hat{S}_{u_1}^{\text{PF}}, \dots, \hat{S}_{u_r}^{\text{PF}}$ . Pour calculer  $\hat{\eta}_0$  et  $\hat{\sigma}^2$ , on peut recourir à leurs estimateurs les plus traditionnels mais [JANON et collab. \[2014\]](#) conseillent d'exploiter tous les échantillons disponibles en posant :

$$\hat{\eta}_0 = \frac{1}{2N_s} \sum_{i=1}^{N_s} (Y^{(i)} + Y_u^{(i)}) \quad \text{et} \quad \hat{\sigma}^2 = \frac{1}{2N_s} \sum_{i=1}^{N_s} \left( (Y^{(i)})^2 + (Y_u^{(i)})^2 \right) - \hat{\eta}_0^2 \quad (3.20)$$

avec  $Y^{(i)} = \phi(\mathbf{X}^{(i)})$  et  $Y_u^{(i)} = \phi(\mathbf{X}_u^{(i)}, \tilde{\mathbf{X}}_{-u}^{(i)})$  .

En utilisant le *bootstrap*, il est montré dans [ARCHER et collab. \[1997\]](#) qu'on peut construire un intervalle de confiance de  $S_u^c$ . Les sorties  $Y^{(i)}$  et  $Y_u^{(i)}$  calculées lors de l'estimation préliminaire de  $S_u^c$  sont réutilisées et aucun appel supplémentaire à  $\phi$  n'est nécessaire. Le principal désavantage de l'estimateur  $\hat{S}_u^{\text{PF}}$  est sa forte variance dans les situations où  $\eta_0^2 \gg V_u$ . Ce problème est d'autant plus prégnant lors de l'estimation d'un petit indice. D'ailleurs, dans cette situation, rien n'empêche  $\hat{S}_u^{\text{PF}}$  de rendre un résultat négatif alors que  $S_u \geq 0$  par définition. Plusieurs estimateurs ont été développés pour éviter cet écueil. Une formule alternative pour les indices totaux est proposée dans [JANSEN et collab. \[1994\]](#) puis [JANSEN \[1999\]](#) :

$$S_u^T = \mathbb{E} \left[ \left( \phi(\mathbf{X}) - \phi(\tilde{\mathbf{X}}_u, \mathbf{X}_{-u}) \right)^2 \right] / 2\mathbb{V}(Y) , \quad (3.21)$$

et en utilisant la seconde relation de l'équation (3.17), on en déduit une nouvelle expression de  $S_u^c$ . Cette approche garantit la positivité de  $S_u^T$  car l'estimateur dérivant de la formule (3.21) s'écrit comme la somme de carrés. Toutefois, l'obtention d'un résultat négatif reste possible pour l'indice fermé  $S_u^c$ . [SOBOL et MYSHETSKAYA \[2008\]](#) passent en revue les astuces permettant de garantir la positivité des estimateurs associés aux plus petits indices. Comme les indices de Sobol sont invariants par translation, il est conseillé dès [SOBOL \[1993\]](#) de les calculer pour une application  $\phi_c = \phi - c$  avec  $c$  une constante aussi proche que possible de  $\eta_0$ . [SALTELLI \[2002\]](#) exhorte à ne pas estimer  $\eta_0$  avec (3.20) mais incite en revanche à estimer directement  $\eta_0^2$  grâce à une utilisation différente des données de  $\mathbf{X}_{\text{sim}}$  et  $\tilde{\mathbf{X}}_{\text{sim}}$  :

$$\hat{\eta}_0^2 = \frac{1}{N_s} \sum_{i=1}^{N_s} \phi(\mathbf{X}^{(i)}) \phi(\tilde{\mathbf{X}}^{(i)}) \quad \text{car} \quad \eta_0^2 = \left( \mathbb{E}[\phi(\mathbf{X})] \right)^2 = \mathbb{E}[\phi(\mathbf{X})] \mathbb{E}[\phi(\tilde{\mathbf{X}})] = \mathbb{E}[\phi(\mathbf{X}) \phi(\tilde{\mathbf{X}})] . \quad (3.22)$$

En combinant la formule (3.18) du *Pick-Freeze* et l'estimateur (3.22) de  $\eta_0^2$ , on obtient :

$$V_u = \mathbb{E} \left[ \phi(\mathbf{X}) \left( \phi(\mathbf{X}_u, \tilde{\mathbf{X}}_{-u}) - \phi(\tilde{\mathbf{X}}) \right) \right] \quad \text{et} \quad \hat{V}_u^{C_1} = \frac{1}{N_s} \sum_{i=1}^{N_s} \phi(\mathbf{X}^{(i)}) \left( \phi(\mathbf{X}_u^{(i)}, \tilde{\mathbf{X}}_{-u}^{(i)}) - \phi(\tilde{\mathbf{X}}_u^{(i)}) \right) . \quad (3.23)$$

Apparu dans [TARANTOLA et collab. \[2007\]](#), l'estimateur  $\hat{V}_u^{C_1}$  porte le nom de *Correlation 1* ( $C_1$ ). Avec ses  $3N_s$  appels à  $\phi$ , il est légèrement plus onéreux que  $\hat{V}_u^{\text{PF}}$  mais il est plus précis pour estimer les petits indices. Imaginons par exemple que  $S_u^c$  soit un très petit indice. Cela signifie que la variabilité de  $Y$  s'explique principalement par l'incertitude véhiculée par  $\mathbf{X}_{-u}$ . De ce fait, chacune des différences  $\phi(\mathbf{X}_u^{(i)}, \tilde{\mathbf{X}}_{-u}^{(i)}) - \phi(\tilde{\mathbf{X}}^{(i)})$  est petite en valeur absolue étant donné que les deux termes partagent le même second argument. Numériquement, on peut comprendre assez intuitivement que le schéma  $C_1$  qui calcule la moyenne d'une séquence de petits nombres confère une plus grande précision que l'unique soustraction (3.19) du schéma PF. On fait toutefois remarquer que les termes dont on calcule la moyenne ne sont petits en valeur absolue qu'à condition que chacune des quantités  $\phi(\mathbf{X}^{(i)})$  le soit aussi, ce qui n'est pas acquis. Pour y parvenir, [OWEN \[2013\]](#) a développé le schéma *Correlation 2* ( $C_2$ ) en introduisant une seconde copie indépendante  $\hat{\mathbf{X}}$  de  $\mathbf{X}$  :

$$V_u = \mathbb{E} \left[ \left( \phi(\mathbf{X}) - \phi(\hat{\mathbf{X}}_u, \mathbf{X}_{-u}) \right) \left( \phi(\mathbf{X}_u, \tilde{\mathbf{X}}_{-u}) - \phi(\tilde{\mathbf{X}}) \right) \right] \quad \text{avec} \quad \begin{cases} \mathbf{X}_{-u} \perp \tilde{\mathbf{X}}_{-u} \\ \mathbf{X}_{-u} \stackrel{\text{loi}}{=} \tilde{\mathbf{X}}_{-u} \end{cases} \quad \text{et} \quad \begin{cases} \mathbf{X}_u \perp \hat{\mathbf{X}}_u \\ \mathbf{X}_u \stackrel{\text{loi}}{=} \hat{\mathbf{X}}_u \end{cases} . \quad (3.24)$$

L'estimateur  $\hat{V}_u^{C_2}$  s'obtient en remplaçant les sorties  $\phi(\mathbf{X}^{(i)})$  par les différences  $\phi(\mathbf{X}^{(i)}) - \phi(\hat{\mathbf{X}}_u^{(i)}, \mathbf{X}_{-u}^{(i)})$  dans l'expression de  $\hat{V}_u^{C_1}$ . On fait alors la moyenne de termes s'exprimant comme le produit de deux petits nombres. Ce schéma est le plus adapté pour estimer les petits indices mais il faut concéder un budget de  $4N_s$  appels à  $\phi$ .

Il existe aussi des schémas d'estimation fondés sur des outils mathématique d'une tout autre nature. [SALTELLI et collab. \[1999\]](#) montrent qu'il est possible d'utiliser la méthode FAST (*Fast Amplitude Sensitivity Test*) introduite initialement par [CUKIER et collab. \[1973\]](#). Il faut commencer par décomposer  $\phi$  dans une base de Fourier puis une formule permet d'exprimer les indices  $S_u^c$  en fonction des coefficients  $\{c_k\}_{k \in \mathbb{Z}}$  de la projection. Une version améliorée de FAST appelée RBD (*Random Balance Design*) est mise au point dans [TARANTOLA et collab. \[2006\]](#) pour contourner certaines difficultés liées au choix des fréquences. Cette méthode est étendue aux indices de Sobol totaux  $S_u^T$  par [MARA \[2009\]](#). En outre, les estimateurs construits avec la méthode RBD présentent un biais que [TISSOT et PRIEUR \[2012\]](#) sont parvenus à corriger. D'un point de vue conceptuel, la méthode FAST est assez proche de celle développée par [SUDRET \[2008\]](#) lorsque qu'un métamodèle est construit par chaos polynomial. Il est en effet démontré qu'il existe une expression analytique simple des indices de Sobol à partir des coefficients de la projection dans la base de polynômes.

Dans un style tout à fait original, [GAMBOA et collab. \[2020\]](#) sont aussi parvenus à estimer les indices de Sobol du premier ordre à partir de la notion de corrélation de [CHATTERJEE \[2020\]](#). Le principal intérêt de ce schéma réside dans le fait que tous les indices du premier ordre peuvent être estimés à partir de seulement  $N_s$  appels à  $\phi$ . Parfois, les indices de Sobol sont vus comme trop réducteurs étant donné qu'ils décrivent la variabilité de la sortie  $Y$  à partir de son seul moment d'ordre 2. Si on cherche à focaliser l'AS sur une autre statistique de  $Y$ , [FORT et collab. \[2016\]](#) ont développé des indices basés sur des fonctions de contraste. De même, si on veut utiliser des indices qui rendent compte de l'impact des entrées sur l'ensemble de la loi  $\mu_Y$ , on peut utiliser les indices construits par [GAMBOA et collab. \[2018\]](#) à l'aide de la distance de Cramér-Von Mises. Toutefois, à l'instar des indices de Sobol, ces indices ne sont rigoureusement définis que pour des entrées indépendantes. Pour aller au-delà de cette hypothèse trop souvent limitante, de nombreuses pistes ont été explorées et sont détaillées dans la prochaine section.

### 3.2.3 Analyse de sensibilité globale pour les variables dépendantes

La question de l'AS en présence d'entrées dépendantes a été abordée de plusieurs manières différentes. Très naturellement, il y a eu une volonté de la communauté statistique d'étendre les indices de Sobol pour continuer à travailler à partir d'une décomposition de la variance. D'autres indices ont aussi été construits pour mesurer l'impact des entrées sur l'ensemble de la loi en sortie du code de calcul. Cette approche de l'AS est dite indépendante des moments et s'appuie sur la quantification de l'écart séparant la distribution de sortie non-conditionnelle et les différentes distribution de sortie conditionnelles. Enfin, la théorie des jeux, et notamment la notion de valeur de Shapley, peut être mise au service de l'AS et permettre une manipulation efficace de la structure de dépendance entre les variables.

#### 3.2.3.1 Méthodes basées sur une décomposition de la variance

Au vu de ce qui a été présenté dans la Section 3.2.2, on rappelle que la suppression de l'hypothèse d'indépendance entraîne l'apparition de deux problèmes majeurs :

- (a) Les fonctions  $\eta_u(\mathbf{X}_u)$  telles qu'elles sont définies à l'équation (3.13) ne vérifient plus la condition d'orthogonalité (3.11) et il n'est donc plus possible d'écrire la décomposition (3.14) de  $\mathbb{V}(Y)$ . Par conséquent, la somme des indices partiels  $S_u$  n'est plus égale à 1 et l'interprétation en termes de pourcentage de  $\mathbb{V}(Y)$  ne tient plus.
- (b) Lorsque les variables  $\mathbf{X}_u$  et  $\mathbf{X}_{-u}$  sont dépendantes,  $V_u$  ne peut plus s'exprimer à l'aide des formules (3.18), (3.23) et (3.24) qui sont respectivement à l'origine des schémas PF,  $C_1$  et  $C_2$ .

Une première solution peut être d'ignorer le point (a) et de considérer que les indices  $S_u^c = V_u/\sigma^2$  restent pertinents, tout en étant conscient que la somme des indices partiels  $S_u$  qui en découlent ne vaut plus 1. Pour adapter la méthode CRM, il faut être en mesure de simuler des échantillons de  $\mathbf{X}_{-u} | \mathbf{X}_u = \mathbf{x}_u^{(k)}$ , ce qui n'est pas forcément facile lorsque la copule sous-jacente  $c_U$  est autre que gaussienne. Cette méthode est par exemple utilisée dans les travaux de [SALTELLI et collab. \[2001\]](#) et [RATTO et collab. \[2001\]](#). De plus, [KUCHERENKO et collab. \[2012\]](#) montrent que la formule (3.18) du *Pick-Freeze* peut être adaptée au cas dépendant :

$$V_u = \mathbb{E} \left[ \phi(\mathbf{X}) \phi(\mathbf{X}_u, \tilde{\mathbf{X}}_{-u}) \right] - \eta_0^2 \quad \text{avec} \quad \begin{cases} \mathbf{X}_{-u} | \mathbf{X}_u \perp \tilde{\mathbf{X}}_{-u} | \mathbf{X}_u \\ \mathbf{X}_{-u} | \mathbf{X}_u \stackrel{\text{loi}}{=} \tilde{\mathbf{X}}_{-u} | \mathbf{X}_u \end{cases} . \quad (3.25)$$

On en déduit un estimateur de type PF dans le cas dépendant :

$$\hat{V}_u^{\text{PF}} = \frac{1}{N_s} \sum_{i=1}^{N_s} \phi(\mathbf{X}^{(i)}) \phi(\mathbf{X}_u^{(i)}, \tilde{\mathbf{X}}_{-u}^{(i)}) - \hat{\eta}_0^2 \quad \text{avec} \quad \begin{cases} \mathbf{X}_u^{(i)} \sim f_{\mathbf{X}_u} \\ \mathbf{X}_{-u}^{(i)} \perp \tilde{\mathbf{X}}_{-u}^{(i)} \sim f_{\mathbf{X}_{-u} | \mathbf{X}_u}(\cdot | \mathbf{X}_u^{(i)}) \end{cases} . \quad (3.26)$$

Les estimateurs (3.19) et (3.26) semblent identiques mais le processus d'échantillonnage de  $\mathbf{X}_{\text{sim}}$  et  $\tilde{\mathbf{X}}_{\text{sim}}$  est différent. Dans le second cas, il impose la connaissance des lois conditionnelles  $\mu_{\mathbf{X}_{-u} | \mathbf{X}_u}$ . On signale par ailleurs que la méthode RBD a été rendue applicable à des entrées dépendantes grâce au travail de [XU et GERTNER \[2007\]](#). Une autre contribution notable pour estimer  $V_j$  est celle proposée par [DA VEIGA et collab. \[2009\]](#) qui mettent en place une régression linéaire hétéroscédastique de  $Y$  en  $X_j$  :

$$Y = m_j(X_j) + \sigma_j(X_j) \epsilon_j \quad \text{avec} \quad \begin{cases} \mathbb{E}[\epsilon_j | X_j] = 0 \\ \mathbb{V}(\epsilon_j | X_j) = 1 \end{cases} . \quad (3.27)$$

On utilise  $\mathbf{X}_{\text{sim}}$  pour estimer les fonctions  $m_j(z) = \mathbb{E}[Y | X_j = z]$  et  $\sigma^2(z) = \mathbb{V}(Y | X_j = z)$  avec une technique d'ajustement polynomial local décrite dans [FAN et GIJBELS \[1996\]](#). Les deux quantités  $V_j = \mathbb{V}(\mathbb{E}[Y | X_j])$  et  $E_j = \mathbb{E}[\mathbb{V}(Y | X_j)]$  sont ensuite estimées à partir des données de  $\tilde{\mathbf{X}}_{\text{sim}}$  :

$$\hat{V}_j = \frac{1}{N_s} \sum_{i=1}^{N_s} \left( \hat{m}_j(\tilde{X}_j^{(i)}) \right)^2 - \left( \frac{1}{N_s} \sum_{i=1}^{N_s} \hat{m}_j(\tilde{X}_j^{(i)}) \right)^2 \quad \text{et} \quad \hat{E}_j = \frac{1}{N_s} \sum_{i=1}^{N_s} \hat{\sigma}_j^2(\tilde{X}_j^{(i)}) . \quad (3.28)$$

Toutes les méthodes présentées jusqu'ici se contentent de traiter la difficulté (b). Certains auteurs ont cherché des parades à la difficulté (a) en modifiant la définition des indices de Sobol. Les différentes initiatives rencontrées dans la littérature imposent des hypothèses supplémentaires sur la loi  $\mu_X$  ou sur la fonction  $\phi$ , ce qui réduit le cadre d'application des méthodologies développées.

L'approche de [JACQUES et collab. \[2006\]](#) et [GILQUIN et collab. \[2015\]](#) intervient dans un contexte où il est possible de séparer  $\mathbf{X}$  en plusieurs sous-vecteurs indépendants :

$$\exists r \in \{2, \dots, d-1\}, \exists \mathbf{u}_1, \dots, \mathbf{u}_r \subset [d] \quad \text{tels que} \quad \begin{cases} \bigcup_{k=1}^r \mathbf{u}_k = [d] \\ \bigcap_{k=1}^r \mathbf{u}_k = \emptyset \end{cases} \quad \text{et} \quad \mathbf{X}_{\mathbf{u}_1} \perp \dots \perp \mathbf{X}_{\mathbf{u}_r} . \quad (3.29)$$

Pour chaque sous-vecteur  $\mathbf{X}_{\mathbf{u}_k}$ , l'idée est ensuite de calculer l'indice fermé  $S_{\mathbf{u}_k}^c$  exactement comme dans le cas indépendant. Dans des situations pratiques où  $\mathbf{X}$  peut être désassemblé en un grand nombre de sous-vecteurs indépendants de petites tailles, cette approche peut avoir une grande utilité en dépit de son apparente simplicité.

Lorsque la boîte noire  $\phi$  peut être assimilée à une application linéaire, [XU et GERTNER \[2008\]](#) rappellent que  $V_j = \mathbb{V}(\mathbb{E}[Y | X_j])$  peut être estimée à l'aide de la variance empirique des résidus de la régression de  $Y$  en  $X_j$ . Les auteurs posent alors  $V_j = V_j^U + V_j^C$  où :

- $V_j^U$  est la part de variance expliquée par le comportement marginal de  $X_j$ ,
- $V_j^C$  est la part de variance expliquée par les dépendances entre  $X_j$  et  $\mathbf{X}_{-j}$ .

Les exposants U et C font référence aux adjectifs “*uncorrelated*” et “*correlated*”. Le terme  $V_j^U$  peut être estimé à partir de la variance empirique des résidus de la régression de  $Y$  en  $\hat{Z}_j$  avec  $\hat{Z}_j$  l’erreur commise lors que la régression de  $X_j$  en  $\mathbf{X}_{-j}$ . Le terme  $V_j^C$  s’obtient ensuite par soustraction. Cette méthode est reprise et étendue par **Li et collab. [2011]**.

La méthode ANCOVA (*ANalysis of COVariance*) de **Li et collab. [2010]** cherche à renouer avec la propriété de sommation à 1 des indices de sensibilité. Supposons qu’on dispose d’une certaine décomposition de Hoeffding (3.10) de  $Y = \phi(\mathbf{X})$ . On précise bien qu’il ne peut pas s’agir d’ANOVA car les variables sont dépendantes. Si le passage direct à la variance comme à l’équation (3.14) n’est plus possible, on peut toujours utiliser la formule de la variance d’entrées dépendantes :

$$\mathbb{V}(Y) = \text{Cov}(Y, Y) = \sum_{\mathbf{u} \subseteq [d]} \text{Cov}(\eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}), Y) = \sum_{\mathbf{u} \subseteq [d]} \sum_{\mathbf{v} \subseteq [d]} \text{Cov}(\eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}), \eta_{\mathbf{v}}(\mathbf{X}_{\mathbf{v}})) \quad (3.30)$$

$$= \sum_{\mathbf{u} \subseteq [d]} \left( \mathbb{V}(\eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}})) + \sum_{\mathbf{v} \neq \mathbf{u}} \text{Cov}(\eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}), \eta_{\mathbf{v}}(\mathbf{X}_{\mathbf{v}})) \right). \quad (3.31)$$

Cela suggère la définition de trois indices pour mesurer l’influence de  $\mathbf{X}_{\mathbf{u}}$  :

$$S_{\mathbf{u}}^{\text{Tot}} := \frac{\text{Cov}(\eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}), Y)}{\mathbb{V}(Y)} = S_{\mathbf{u}}^{\text{Str}} + S_{\mathbf{u}}^{\text{Corr}} \quad (3.32)$$

avec  $S_{\mathbf{u}}^{\text{Str}} := \frac{\mathbb{V}(\eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}))}{\mathbb{V}(Y)}$  et  $S_{\mathbf{u}}^{\text{Corr}} = \frac{1}{\mathbb{V}(Y)} \sum_{\mathbf{v} \neq \mathbf{u}} \text{Cov}(\eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}), \eta_{\mathbf{v}}(\mathbf{X}_{\mathbf{v}}))$ .

L’indice  $S_{\mathbf{u}}^{\text{Tot}}$  est la contribution totale de  $\mathbf{X}_{\mathbf{u}}$ . Elle s’écrit comme la somme d’une contribution structurelle  $S_{\mathbf{u}}^{\text{Str}}$  et d’une contribution corrélée  $S_{\mathbf{u}}^{\text{Corr}}$ . La somme des contributions totales vaut 1. La principale difficulté accompagnant la mise en œuvre de cette méthode est l’identification d’une décomposition de Hoeffding. Pour y parvenir, **Li et collab. [2010]** conseillent le recours à un métamodèle pour lequel cette identification est naturelle. **SUDRET et CANIOU [2013]** appliquent la méthode ANCOVA dans le cas d’une métamodélisation par chaos polynomial. Si la mise en place d’un métamodèle est un mécanisme essentiel et salutaire pour l’implémentation, c’est aussi un défaut majeur de cette méthode dont la réponse finale dépend de la qualité du métamodèle. En outre, l’interprétation d’une collection de triplets  $(S_{\mathbf{u}}^{\text{Tot}}, S_{\mathbf{u}}^{\text{Str}}, S_{\mathbf{u}}^{\text{Corr}})$  n’est pas toujours évidente et implique parfois des indices négatifs.

La méthode de **CHASTAING et collab. [2012]** présente certaines similarités avec ANCOVA. Elle s’inspire du travail initial de **HOOKER [2007]** et s’applique seulement aux vecteurs aléatoires dont la densité  $f_{\mathbf{X}}$  vérifie la propriété suivante :

$$\exists 0 < M \leq 1, \quad \forall \mathbf{u} \subseteq [d], \quad \forall \mathbf{x} \in \mathbb{R}^d, \quad f_{\mathbf{X}}(\mathbf{x}) \geq M f_{\mathbf{X}_{\mathbf{u}}}(\mathbf{x}_{\mathbf{u}}) f_{\mathbf{X}_{-\mathbf{u}}}(\mathbf{x}_{-\mathbf{u}}). \quad (3.33)$$

Dans le cas particulier d’entrées indépendantes, l’inégalité précédente est vérifiée et correspond au cas d’égalité avec  $M = 1$ . Sous la condition (3.33), il existe une unique décomposition que l’on peut qualifier d’orthogonale hiérarchique :

$$Y = \phi(\mathbf{X}) = \sum_{\mathbf{u} \subseteq [d]} \eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}) \quad \text{telle que} \quad \mathbf{u} \subset \mathbf{v} \implies \text{Cov}(\eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}), \eta_{\mathbf{v}}(\mathbf{X}_{\mathbf{v}})) = 0. \quad (3.34)$$

En reprenant l’équation (3.30), il vient :

$$\mathbb{V}(Y) = \sum_{\mathbf{u} \subseteq [d]} \sum_{\mathbf{v} \subseteq [d]} \text{Cov}(\eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}), \eta_{\mathbf{v}}(\mathbf{X}_{\mathbf{v}})) \quad (3.35)$$

$$= \sum_{\mathbf{u} \subseteq [d]} \left( \underbrace{\mathbb{V}(\eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}})) + \sum_{\substack{\mathbf{u} \subset \mathbf{v} \\ \mathbf{v} \subset \mathbf{u}}} \text{Cov}(\eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}), \eta_{\mathbf{v}}(\mathbf{X}_{\mathbf{v}}))}_{=0} + \sum_{\mathbf{u} \cap \mathbf{v} \neq \mathbf{u}, \mathbf{v}} \text{Cov}(\eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}), \eta_{\mathbf{v}}(\mathbf{X}_{\mathbf{v}})) \right), \quad (3.36)$$

et on déduit de nouveaux indices de sensibilité :

$$S_{\mathbf{u}}^{\text{H}} := \frac{1}{\mathbb{V}(Y)} \left( \mathbb{V}(\eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}})) + \sum_{\mathbf{u} \cap \mathbf{v} \neq \mathbf{u}, \mathbf{v}} \text{Cov}(\eta_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}), \eta_{\mathbf{v}}(\mathbf{X}_{\mathbf{v}})) \right), \quad (3.37)$$

dont la somme est bien égale à 1. Pour leur estimation, une procédure numérique d'identification des fonctions  $\eta_{\mathbf{u}}$  de la décomposition (3.34) est disponible dans [CHASTAING et collab. \[2015\]](#) ainsi que dans [CHAMPION et collab. \[2015\]](#). Toutefois, l'expression (3.37) ne garantit pas que les indices  $S_{\mathbf{u}}^{\text{H}}$  soient tous dans l'intervalle  $[0, 1]$ . En ce sens, cette procédure d'AS n'est pas complètement satisfaisante car il est difficile de statuer sur le rôle joué par  $\mathbf{X}_{\mathbf{u}}$  si  $S_{\mathbf{u}}^{\text{H}} \leq 0$  ou si  $S_{\mathbf{u}}^{\text{H}} \geq 1$ .

Avant de clore cet examen des techniques basées sur la décomposition de la variance, on souhaite mentionner les travaux menés par [MARA et TARANTOLA \[2012\]](#) puis [MARA et collab. \[2015\]](#). La méthodologie se veut plus générique, en ce sens qu'elle est applicable sans hypothèse aucune sur  $f_{\mathbf{X}}$  ou sur  $\phi$ . L'outil principal est la transformation de Rosenblatt détaillée au Chapitre 2 et dont on sait qu'il existe  $d!$  versions différentes en fonction du séquençage des variables conditionnantes dans l'équation (2.12). Ici, seules les versions obtenues par permutation cyclique de  $\mathbf{X}$  sont requises. On note  $\mathcal{T}_k$  la transformation résultant du séquençage  $X_k, X_{k+1}, \dots, X_d, X_1, \dots, X_{k-1}$  et  $\mathbf{U}^k$  le vecteur à variables indépendantes qui en résulte. On peut alors calculer rigoureusement des indices de Sobol pour toutes les applications  $\phi_k = \phi \circ \mathcal{T}_k^{-1}$ . Supposons que l'on veuille mesurer l'influence de la variable  $X_j$  sur  $Y$ .

- Si on applique  $\mathcal{T}_j$ , on fabrique une variable  $U_1^j$  qui est corrélée à toutes les variables de  $\mathbf{X}_{-j}$ . Les deux indices de Sobol calculés pour  $U_1^j$  rendent compte de l'influence de  $X_j$  sur  $\mathbb{V}(Y)$  lorsqu'on inclut les dépendances entre  $X_j$  et  $\mathbf{X}_{-j}$ .
- De façon complémentaire, l'utilisation de  $\mathcal{T}_{j+1}$  met en jeu la variable  $U_d^{j+1}$  qui est indépendante de toutes les variables de  $\mathbf{X}_{-j}$ . Les deux indices de Sobol calculés pour  $U_d^{j+1}$  excluent alors les dépendances entre  $X_j$  et  $\mathbf{X}_{-j}$ .

Cette méthode se solde par le calcul de quatre indices pour chaque entrée  $X_j$  et l'interprétation peut s'avérer à nouveau assez complexe.

Dans cette section, on a constaté qu'il existe des velléités de se ramener à une décomposition intuitive de la variance même en présence d'entrées dépendantes. On peut déplorer que la plupart des méthodes développées conduisent à des indices qui sont difficiles à interpréter. Dans certains cas, cela tient au fait que les indices puissent prendre des valeurs en dehors de  $[0, 1]$ . Dans d'autres cas, le problème vient du fait que plusieurs indices doivent être confrontés simultanément pour appréhender le rôle joué par la variable examinée.

### 3.2.3.2 Méthodes basées sur l'emploi d'une mesure de dissimilarité

On s'intéresse désormais aux indices de sensibilité qui peuvent quantifier l'impact d'entrées dépendantes sur l'ensemble de la loi image  $\mu_Y$ . Contrairement au formalisme HDMR, on a ici  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$  et  $\mu_{\mathbf{X}}$  une loi multidimensionnelle à support dans  $\mathbb{R}^d$ . Une première méthode d'AS est esquissée dans [BORGONOVO \[2006\]](#). Elle repose sur l'intuition qui consiste à penser qu'un groupe de plusieurs entrées  $\mathbf{X}_{\mathbf{u}}$  est fortement influent sur la sortie  $Y$  lorsque la loi de probabilité  $\mu_Y$  est fortement perturbée par l'action visant à rendre  $\mathbf{X}_{\mathbf{u}}$  constamment égal à une valeur  $\mathbf{x}_{\mathbf{u}} \in \mathbb{R}^r$ . Selon ce point de vue, l'amplitude de la perturbation est alors mesurée par le "shift" :

$$\text{shift}(\mathbf{x}_{\mathbf{u}}) := \|f_Y - f_{Y|\mathbf{X}_{\mathbf{u}}}(\cdot | \mathbf{x}_{\mathbf{u}})\|_{L^1} = \int_{\mathbb{R}} |f_Y(y) - f_{Y|\mathbf{X}_{\mathbf{u}}}(y | \mathbf{x}_{\mathbf{u}})| dy = 2 d_{\text{VT}}(\mu_Y, \mu_{Y|\mathbf{X}_{\mathbf{u}}=\mathbf{x}_{\mathbf{u}}}). \quad (3.38)$$

La dernière égalité vient rappeler que la norme  $L^1$  de la différence entre deux densités est égale au double de la distance en variation totale<sup>3</sup> entre les deux lois associées. La Figure 3.1 propose une

3. Soient  $\mu_1$  et  $\mu_2$  deux mesures de probabilité sur  $(\Omega, \mathcal{A})$ . On a alors  $d_{\text{VT}}(\mu_1, \mu_2) := \sup_{A \in \mathcal{A}} |\mu_1(A) - \mu_2(A)|$ .

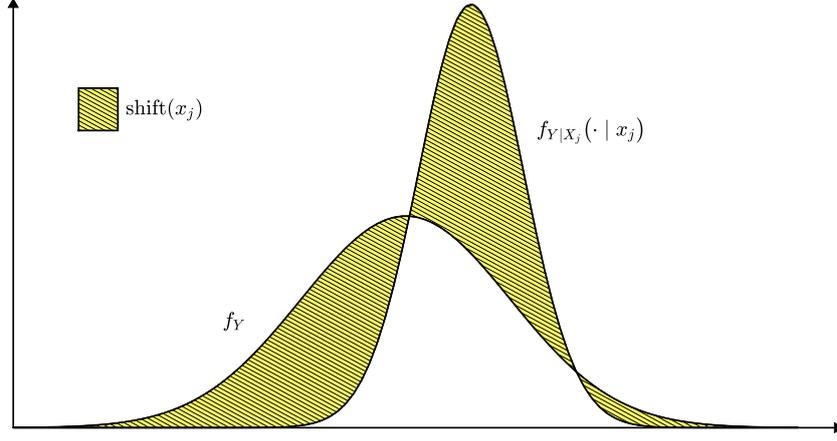


FIGURE 3.1 – Illustration de l'intuition ayant motivé la création des  $\delta$ -indices de Borgonovo.

illustration de ce que représente  $\text{shift}(\mathbf{x}_u)$  pour les densités  $f_Y$  et  $f_{Y|\mathbf{X}_u}(\cdot | \mathbf{x}_u)$ . Pour tout  $\mathbf{u} \subseteq [d]$ , l'indice de Borgonovo de  $\mathbf{X}_u$  par rapport à  $Y$  est défini par :

$$\delta_{\mathbf{u}}^Y = \delta(Y | \mathbf{X}_u) = \frac{1}{2} \mathbb{E} \left[ \text{shift}(\mathbf{X}_u) \right] = \mathbb{E} \left[ d_{\text{VT}}(\mu_Y, \mu_{Y|\mathbf{X}_u}) \right]. \quad (3.39)$$

De par la notation choisie,  $\delta_{\mathbf{u}}^Y$  est aussi appelé le  $\delta$ -indice de  $\mathbf{X}_u$  par rapport à  $Y$ . Lorsqu'il n'y a pas d'ambiguïté, on note simplement  $\delta_{\mathbf{u}}$ . Contrairement aux indices de Sobol, la définition donnée aux indices de Borgonovo convient aussi bien aux entrées dépendantes et indépendantes. De plus, les  $\delta$ -indices ont des propriétés assez différentes.

### **Théorème 3.2.1 (Propriétés des $\delta$ -indices)**

- ( $\delta 1$ )  $\forall \mathbf{u} \subseteq [d], 0 \leq \delta_{\mathbf{u}} \leq 1$       ( $\delta 2$ )  $\forall \mathbf{u} \subset \mathbf{v} \subseteq [d], \delta_{\mathbf{u}} \leq \delta_{\mathbf{v}}$       ( $\delta 3$ )  $\delta_{[d]} = 1$
- ( $\delta 4$ ) Soient  $g_1 : \mathbb{R} \rightarrow \mathbb{R}$  et  $g_2 : \mathbb{R}^d \rightarrow \mathbb{R}^d$  des difféomorphismes.  
Pour  $Y_1 := g_1 \circ \phi(\mathbf{X})$  et  $Y_2 := \phi \circ g_2(\mathbf{X})$ , on a  $\delta_{\mathbf{u}}^{Y_1} = \delta_{\mathbf{u}}^{Y_2} = \delta_{\mathbf{u}}^Y$ .

Les points ( $\delta 1$ ) à ( $\delta 3$ ) sont démontrés dans [BORGONOVO \[2007\]](#). De même, on trouve une preuve de ( $\delta 4$ ) dans [BORGONOVO et collab. \[2011\]](#). Le problème de l'estimation des  $\delta$ -indices a motivé bon nombre de travaux de recherche au cours de la dernière décennie. [DERENNES \[2019\]](#) est l'auteur d'un état de l'art à ce sujet. Les schémas d'estimation s'appuient sur les différentes réécritures intégrales de la définition (3.39) :

$$\delta_{\mathbf{u}} = \frac{1}{2} \int_{\mathbb{R}^r} \left( \int_{\mathbb{R}} |f_Y(y) - f_{Y|\mathbf{X}_u}(y | \mathbf{x}_u)| dy \right) f_{\mathbf{X}_u}(\mathbf{x}_u) d\mathbf{x}_u, \quad (3.40)$$

$$= \frac{1}{2} \int_{\mathbb{R}^{r+1}} |f_{\mathbf{X}_u Y}(\mathbf{x}_u, y) - f_{\mathbf{X}_u}(\mathbf{x}_u) f_Y(y)| d\mathbf{x}_u dy, \quad (3.41)$$

$$= \frac{1}{2} \int_{\mathcal{H}_{r+1}} |c_{\mathbf{R}_u S}(\mathbf{r}_u, s) - c_{\mathbf{R}_u}(\mathbf{r}_u)| d\mathbf{r}_u ds \quad \text{avec } \mathbf{R}_u := F_{\mathbf{X}_u}(\mathbf{X}_u) \quad \text{et } S := F_Y(Y). \quad (3.42)$$

L'équation (3.40) conduit à un estimateur basé sur une double boucle d'échantillonnage :

$$\hat{\delta}_{\mathbf{u}} = \frac{1}{2N_s} \sum_{i=1}^{N_s} \left\| \hat{f}_Y(\cdot) - \hat{f}_{Y|\mathbf{X}_u}(\cdot | \mathbf{X}_u^{(i)}) \right\|_{L^1} \quad \text{avec } \mathbf{X}_u^{(i)} \stackrel{\text{i.i.d.}}{\sim} f_{\mathbf{X}_u}. \quad (3.43)$$

Ce schéma d'estimation implique d'être capable d'estimer la densité de sortie  $f_Y$  ainsi que tout une séquence de densités conditionnelles  $f_{Y|\mathbf{X}_u}(\cdot | \mathbf{X}_u^{(i)})$ . Une variante de cet estimateur naïf est proposée dans [PLISCHKE et collab. \[2013\]](#) et garantit la convergence de  $\hat{\delta}_{\mathbf{u}}$  vers l'indice théorique

$\delta_{\mathbf{u}}$ . Les densités sont toujours estimées avec une technique non-paramétrique, le plus souvent par KDE comme le font [LUO et collab. \[2014\]](#). Toutefois, [YUN et collab. \[2019\]](#) montrent qu'il est également possible de combiner cet estimateur avec la méthode d'apprentissage par maximum d'entropie. Enfin, [CASTAINGS et collab. \[2012\]](#) ont développé une stratégie d'échantillonnage qui permet d'obtenir tous les  $\delta$ -indices du premier ordre à partir du budget de simulation d'un unique indice  $\delta_j$ . Pour éliminer la double boucle d'échantillonnage, il faut utiliser l'expression (3.41) qui mène à l'estimateur :

$$\hat{\delta}_{\mathbf{u}} := \|\hat{f}_{\mathbf{X}_{\mathbf{u}}Y} - f_{\mathbf{X}_{\mathbf{u}}} \hat{f}_Y\|_{L^1}. \quad (3.44)$$

Une unique boucle d'échantillonnage permet de disposer du matériel nécessaire à l'estimation des densités  $f_Y$  et  $f_{\mathbf{X}_{\mathbf{u}}Y}$ . La norme  $L^1$  est ensuite estimée par Monte-Carlo. Pour exécuter cette tâche, [WEI et collab. \[2013\]](#) choisissent de simuler des échantillons suivant  $f_{\mathbf{X}_{\mathbf{u}}} \hat{f}_Y$ . Il est toutefois montré dans [DERENNES et collab. \[2019a\]](#) que l'emploi de  $\hat{f}_{\mathbf{X}_{\mathbf{u}}Y}$  permet de fabriquer un estimateur final plus robuste. Dans certaines situations, notamment lorsqu'on veut estimer des indices d'ordre supérieur (c'est-à-dire  $r = |\mathbf{u}| \geq 2$ ), il est préférable de se tourner vers l'expression (3.42) qui est à l'origine de l'estimateur :

$$\hat{\delta}_{\mathbf{u}} := \frac{1}{2N_S} \sum_{i=1}^{N_S} \left| \hat{c}_{\mathbf{R}_{\mathbf{u}}S}(\mathbf{R}_{\mathbf{u}}^{(i)}, S^{(i)}) - c_{\mathbf{R}_{\mathbf{u}}}(\mathbf{R}_{\mathbf{u}}^{(i)}) \right| \quad \text{avec} \quad (\mathbf{R}_{\mathbf{u}}^{(i)}, S^{(i)}) \stackrel{\text{i.i.d.}}{\sim} \mathcal{U}([0, 1])^{\otimes r+1}. \quad (3.45)$$

La précision de  $\hat{\delta}_{\mathbf{u}}$  dépend alors de la qualité de l'estimation  $\hat{c}_{\mathbf{R}_{\mathbf{u}}S}$ . Pour les  $\delta$ -indices du premier ordre qui impliquent seulement l'estimation de copules bidimensionnelles, [WEI et collab. \[2014\]](#) utilisent la méthode KDE (1.156) tandis que [DERENNES et collab. \[2018\]](#) privilégient une technique d'apprentissage par maximum d'entropie. Pour des  $\delta$ -indices d'ordre supérieur, les copules  $c_{\mathbf{R}_{\mathbf{u}}S}$  sont estimées avec des R-vignes paramétriques dans [ZHOU et collab. \[2019\]](#) et non-paramétriques dans [SARAZIN et collab. \[2020\]](#). Les estimateurs (3.43), (3.44) et (3.45) sont respectivement appelés l'estimateur "double boucle", l'estimateur "simple boucle" et l'estimateur "orienté copule". Leurs schémas numériques sont réexaminés au Chapitre 5 à l'aune d'une problématique d'ASF. D'autres techniques d'estimation peuvent être rencontrées dans la littérature à l'image de celle développée par [LIU et HOMMA \[2009\]](#). Elle repose sur la recherche des points d'intersection entre les densités  $f_Y$  et  $f_{Y|\mathbf{X}_{\mathbf{u}}}$  puis l'évaluation des fonctions de répartition  $F_Y$  et  $F_{Y|\mathbf{X}_{\mathbf{u}}}$  en ces points.

Au vu de la définition initiale (3.39), [BORGONOVO et collab. \[2014\]](#) soulignent qu'on pourrait créer de nouveaux indices en remplaçant la distance en variation totale  $d_{VT}$  par une autre mesure de dissimilarité. On rappelle qu'il existe plusieurs objets mathématiques permettant d'apprécier la "distance" entre deux mesures de probabilité.

### Définition 3.2.1 (Distance, divergence et mesure de dissimilarité)

Soit  $\mathcal{M}$  l'ensemble des mesures de probabilité sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

On considère une application  $d_0 : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}^+$ .

- $d_0$  est une mesure de dissimilarité si :  $\forall \mu \in \mathcal{M}, d_0(\mu, \mu) = 0$ .
- $d_0$  est une divergence si :  $\forall \mu_1, \mu_2 \in \mathcal{M}, d_0(\mu_1, \mu_2) = 0 \iff \mu_1 = \mu_2$ .
- $d_0$  est une distance si c'est une divergence symétrique vérifiant l'inégalité triangulaire.

$d_{VT}$  est une distance, donc une divergence et plus précisément une divergence de [CSISZÁR \[1967\]](#). En effet, elle peut s'écrire sous la forme :

$$d_g(\mu_1, \mu_2) = \int_{\mathbb{R}^d} g\left(\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})}\right) f_2(\mathbf{x}) d\mathbf{x} \quad \text{avec} \quad g : \mathbb{R} \rightarrow \mathbb{R} \text{ convexe et } g(0) = 1, \quad (3.46)$$

en prenant  $g(t) = \frac{1}{2}|t - 1|$ . [DA VEIGA \[2015\]](#) s'est intéressé aux indices  $\mathbb{E}[d_g(\mu_Y, \mu_{Y|\mathbf{X}_{\mathbf{u}}})]$  issus de l'emploi des autres membres de cette classe de divergences. Une étude plus approfondie est ensuite venue avec [RAHMAN \[2016\]](#). Dans le même esprit, [BORGONOVO et IOOSS \[2016\]](#) étudient les indices de type  $\mathbb{E}[h(F_Y, F_{Y|\mathbf{X}_{\mathbf{u}}})]$  avec  $h$  une distance (ou une divergence) sur les fonctions de répartition. Le cas particulier où  $h$  est la distance de Kolmogorov-Smirnov aboutit aux  $\beta$ -indices apparus à l'origine dans [BAUCELLS et BORGONOVO \[2013\]](#).

L'expression (3.42) de  $\delta_{\mathbf{u}}$  révèle le rôle central de la copule  $c_{\mathbf{R}_{\mathbf{u}}}$ . Pour élargir l'idée d'une capture de la dépendance statistique reliant  $\mathbf{X}_{\mathbf{u}}$  et  $Y$ , DA VEIGA [2015] définit de nouveaux indices à l'aide de la distance en corrélation de SZÉKELY et collab. [2007] et de la mesure HSIC (*Hilbert-Schmidt Independence Criterion*) de GRETTON et collab. [2005].

### 3.2.3.3 Méthodes basées sur l'analogie avec la théorie des jeux

Dans la théorie des jeux, les valeurs de SHAPLEY [1953] permettent de répartir équitablement les gains entre des joueurs sachant que certains d'entre eux ont pu coopérer pour augmenter leurs chances de succès. Dans un jeu coopératif, l'ensemble  $[d] := \{1, \dots, d\}$  réunit tous les participants et chaque sous-ensemble  $\mathbf{u} \subseteq [d]$  est vu comme un groupe de joueurs choisissant de nouer une alliance. Dans la suite, on note  $\mathcal{P}_d$  l'ensemble des parties de  $[d]$ . Il réunit les  $2^d$  sous-ensembles  $\mathbf{u} \subseteq [d]$ . Le bilan pécunier est fixé par la fonction caractéristique  $c_0$  qui attribue à chaque alliance  $\mathbf{u}$  le montant de ses gains. D'un point de vue mathématique, on a :

$$\begin{aligned} c_0 : \mathcal{P}_d &\longrightarrow \mathbb{R}^+ \\ \mathbf{u} &\longmapsto c_0(\mathbf{u}) \end{aligned} \quad \text{avec } c_0(\emptyset) = 0. \quad (3.47)$$

La valeur de Shapley  $s_j$  associée au  $j$ -ème joueur est alors donnée par :

$$\forall 1 \leq j \leq d, \quad s_j = s_j(c_0) := \frac{1}{d} \sum_{j \notin \mathbf{u}} \frac{1}{C_{d-1}^{|\mathbf{u}|}} \left( c_0(\mathbf{u} \cup \{j\}) - c_0(\mathbf{u}) \right). \quad (3.48)$$

Les fonctions  $s_1, \dots, s_d$  vérifient des propriétés assez naturelles dans un contexte du partage des gains obtenus à la fin d'un jeu collaboratif.

#### **Théorème 3.2.2 (Propriétés des valeurs de Shapley)**

(S1) **Partage de la totalité des gains :**

$$\sum_{j=1}^d s_j(c_0) = c_0([d]). \quad (3.49)$$

(S2) **Symétrie de la répartition des gains :**

$$\forall 1 \leq j, k \leq d, \left( \forall \mathbf{u} \subseteq [d] \setminus \{j, k\}, c_0(\mathbf{u} \cup \{j\}) = c_0(\mathbf{u} \cup \{k\}) \right) \implies s_j(c_0) = s_k(c_0). \quad (3.50)$$

(S3) **Non-rétribution des collaborations infructueuses :**

$$\forall 1 \leq j \leq d, \left( \forall \mathbf{u} \subseteq [d] \setminus \{j\}, c_0(\mathbf{u} \cup \{j\}) = c_0(\mathbf{u}) \right) \implies s_j(c_0) = 0. \quad (3.51)$$

(S4) **Linéarité par rapport à la fonction caractéristique :**

$$\begin{cases} c_1 : \mathcal{P}_d \longrightarrow \mathbb{R}^+ \\ c_2 : \mathcal{P}_d \longrightarrow \mathbb{R}^+ \end{cases} \implies \forall 1 \leq j \leq d, s_j(c_1 + c_2) = s_j(c_1) + s_j(c_2). \quad (3.52)$$

Historiquement, SHAPLEY [1953] a suivi la démarche inverse. Après avoir posé les axiomes (S1) à (S4), il a démontré que la collection des fonctions  $s_1, \dots, s_d$  données par la formule (3.48) est l'unique solution du problème. Examinons l'expérience aléatoire qui amène le  $j$ -ème joueur à établir une unique alliance avec des partenaires  $\mathbf{a} \subseteq [d] \setminus \{j\}$ . Le gain futur  $G(\mathbf{a})$  que cette collaboration rapportera au  $j$ -ème joueur lui est inconnu au moment où il prend la décision. Ce gain s'exprime comme la différence  $c_0(\mathbf{a} \cup \{j\}) - c_0(\mathbf{a})$  entre les gains obtenus en sa présence et en son absence. On fait l'hypothèse que le choix du nombre de partenaires  $r = |\mathbf{a}| \in \{0, \dots, d-1\}$  suit une loi discrète uniforme. De même, on suppose que le choix du sous-ensemble  $\mathbf{a}$  de taille  $r$  fixée suit

une loi discrète uniforme. Dans ce cadre, l'espérance du gain vaut :

$$\mathbb{E}[G(\mathbf{A})] = \sum_{j \neq \mathbf{a}} G(\mathbf{a}) \mathbb{P}(\mathbf{A} = \mathbf{a}) = \sum_{j \neq \mathbf{a}} G(\mathbf{a}) \underbrace{\mathbb{P}(\mathbf{A} = \mathbf{a} \mid \mathbf{R} = r)}_{1/C_{d-1}^{|\mathbf{a}|}} \underbrace{\mathbb{P}(\mathbf{R} = r)}_{1/d} = s_j(c_0), \quad (3.53)$$

et on retrouve l'expression (3.48) de la valeur de Shapley  $s_j$ . OWEN [2014] est le premier à mettre en avant l'analogie entre la théorie des jeux et l'AS. Les entrées  $X_1, \dots, X_d$  remplacent les joueurs et la mesure de sensibilité fait office de fonction caractéristique. Dans le cas d'entrées indépendantes, on peut alors poser :

$$\forall \mathbf{u} \subseteq [d], \quad c_{\text{Sb}}(\mathbf{u}) = \frac{\mathbb{V}(\mathbb{E}[Y \mid \mathbf{X}_{\mathbf{u}}])}{\mathbb{V}(Y)} = S_{\mathbf{u}}^c \quad \text{et} \quad \forall 1 \leq j \leq d, \quad \text{Sh}_j := \frac{1}{d} \sum_{j \notin \mathbf{u}} \frac{1}{C_{d-1}^{|\mathbf{u}|}} (S_{\mathbf{u} \cup \{j\}}^c - S_{\mathbf{u}}^c). \quad (3.54)$$

Comme  $c_{\text{Sb}}(\emptyset) = 0$  et  $c_{\text{Sb}}([d]) = 1$ , la propriété (S1) indique que  $\sum_j \text{Sh}_j = 1$ . En repartant de la décomposition ANOVA (3.14), OWEN [2014] démontre que l'on a :

$$\forall 1 \leq j \leq d, \quad \text{Sh}_j = \frac{1}{\sigma^2} \sum_{j \in \mathbf{u}} \frac{\sigma_{\mathbf{u}}^2}{|\mathbf{u}|} \quad \text{et donc} \quad S_j^c \leq \text{Sh}_j \leq S_j^T. \quad (3.55)$$

Les valeurs de Shapley  $\text{Sh}_j$  présentent donc un double avantage :

- (a) Comme elles sont positives et somment à 1, elles permettent de répartir la variance  $\mathbb{V}(Y)$  uniquement entre les variables  $X_j$ . Ce n'était pas le cas pour les indices de Sobol car  $\sum_j S_j \leq 1$  avec égalité uniquement si  $\phi$  est linéaire.
- (b) L'encadrement (3.55) fait apparaître  $\text{Sh}_j$  comme un compromis entre l'effet principal représenté par  $S_j^c$  et la prise en compte de toutes les interactions via  $S_j^T$ .

Dans le cas où les variables sont dépendantes, les indices de Sobol ne peuvent plus être interprétés comme des pourcentages de la variance de sortie. Toutefois,  $c_{\text{Sb}}$  reste une fonction caractéristique compatible avec la théorie des valeurs de Shapley. D'ailleurs, OWEN et PRIEUR [2017] soulignent que ce cadre permet d'éliminer sans effort les problèmes méthodologiques posés jusqu'alors par la structure de dépendance des entrées. Par contre, comme la décomposition ANOVA n'est plus valable, il en est de même pour l'inégalité (3.55) dont l'évolution en présence de dépendances est examinée par IOOSS et PRIEUR [2019]. Afin de calculer des valeurs de Shapley à partir d'une mesure de sensibilité sur l'intégralité de la loi  $\mu_Y$ , l'idée originelle de OWEN [2014] est transposée aux indices de Borgonovo dans SARAZIN et collab. [2020]. La fonction caractéristique  $c_{\text{Sb}}$  est alors remplacée par  $c_{\text{Bg}} : \mathbf{u} \mapsto \delta_{\mathbf{u}} = \delta(Y \mid \mathbf{X}_{\mathbf{u}})$ . Étant donné que  $\delta_{[d]} = 1$ , on retrouve à nouveau la propriété de sommation à 1.

Si les valeurs de Shapley présentent des propriétés particulièrement attrayantes, y compris lorsque les entrées sont dépendantes, leur schéma d'estimation est excessivement coûteux car il impose le calcul des  $2^d$  indices de sensibilité. Pour remédier à ce problème, plusieurs approches ont été explorées. SONG et collab. [2016] exploitent le résultat de CASTRO et collab. [2009] qui consiste à réécrire les valeurs de Shapley de la manière suivante :

$$\forall 1 \leq j \leq d, \quad s_j(c_0) = \frac{1}{d!} \sum_{\pi \in \mathcal{P}_d} c_0(\mathbf{P}_j(\pi) \cup \{j\}) - c_0(\mathbf{P}_j(\pi)), \quad (3.56)$$

avec  $\mathcal{P}_d$  l'ensemble des permutations de  $[d] := \{1, \dots, d\}$ . De plus, pour une permutation  $\pi \in \mathcal{P}_d$  donnée,  $\mathbf{P}_j(\pi)$  est le sous-ensemble de  $\pi$  rassemblant les prédécesseurs de  $j$  dans  $\pi$ . Cela suggère que chaque valeur de Shapley  $s_j(c_0)$  peut être vue comme l'espérance d'une image de la loi uniforme sur  $\mathcal{P}_d$  :

$$s_j(c_0) = \mathbb{E} \left[ c_0(\mathbf{P}_j(\Pi) \cup \{j\}) - c_0(\mathbf{P}_j(\Pi)) \right] \quad \text{avec} \quad \Pi \sim \mathcal{U}(\mathcal{P}_d). \quad (3.57)$$

Une estimation de cette espérance peut alors être obtenue à partir d’un petit nombre  $m \ll d!$  de permutations aléatoires. Dans le but de réduire encore davantage le budget de simulation requis pour estimer les valeurs de Shapley, [BENOUMECHIARA et ELIE-DIT-COSAQUE \[2019\]](#) combinent l’algorithme de [SONG et collab. \[2016\]](#) avec la construction d’un métamodèle de  $\phi$ . Par ailleurs, il est démontré dans [PLISCHKE et collab. \[2020\]](#) que le coût de calcul peut également être abaissé en utilisant la formule d’inversion de Möbius.

### 3.2.4 Bilan

Dans cette première partie consacrée à l’AS sur un code de calcul  $\phi : \mathbf{X} \longrightarrow Y$ , on a distingué les méthodes locales des méthodes globales. Les premières reposent essentiellement sur la capacité à calculer  $\nabla_{\mathbf{x}}\phi$  et évaluent la sensibilité de la sortie à des variations infinitésimales et unidirectionnelles à partir d’une position de référence  $\mathbf{x}^*$ . Pour hiérarchiser l’influence des entrées  $X_j$  en tenant compte de l’ensemble de l’information portée par leur distribution conjointe  $\mu_{\mathbf{X}}$ , il faut se tourner vers les méthodes globales. Un clivage apparaît alors entre les méthodes réservées aux variables indépendantes et les méthodes développées pour s’accommoder de la présence d’une éventuelle structure de dépendance entre les entrées. Dans le premier cas, la décomposition ANOVA justifie la définition des indices de Sobol et permet d’attribuer à chaque entrée une part de la variance de la sortie. Dans le second cas, l’utilisation d’indices basés sur une mesure de la dissimilarité entre la loi non-conditionnelle  $\mu_Y$  et les lois conditionnelles  $\mu_{Y|X_j}$  apparaît comme le cadre le plus souple et le plus à même à digérer la notion de dépendance.

Dans la seconde partie de ce chapitre, on s’intéresse aux procédures qui peuvent être mises en œuvre pour analyser la sensibilité de l’application  $\Psi : \Theta \longrightarrow P_f(\Theta)$ . Les incertitudes propagées ne sont plus aléatoires mais épistémiques. Par ailleurs, chaque appel à  $\Psi$  implique de nombreux appels à  $\phi$ . Pour cela, la gestion du budget de simulation ne peut pas être appréhendée de la même manière et on ne peut pas simplement se contenter de décliner les méthodes locales et globales d’AS pour traiter le problème d’ASF.

## 3.3 Application à l’analyse de sensibilité d’une probabilité de défaillance

### 3.3.1 Modélisation des incertitudes épistémiques

Des incertitudes épistémiques s’ajoutent aux incertitudes aléatoires dès lors qu’il n’est pas possible de spécifier complètement la loi  $\mu_{\mathbf{X}}$  des entrées du système. La présence ou l’absence d’incertitudes épistémiques est ainsi déterminée par l’état de connaissance du comportement aléatoire conjoint des entrées. Supposons que l’on souhaite caractériser  $\mu_{\mathbf{X}}$  à partir de la densité conjointe  $f_{\mathbf{X}}(\cdot | \mathcal{F}, \theta)$  d’une certaine famille paramétrique. Deux types d’incertitudes épistémiques coexistent dans ce contexte. D’une part, à supposer qu’il existe une famille paramétrique usuelle  $\mathcal{F}$  susceptible de caractériser parfaitement la loi  $\mu_{\mathbf{X}}$ , la méconnaissance de cette famille engendre une première forme d’incertitudes épistémiques. D’autre part, la méconnaissance des paramètres sous-jacents fait naître une seconde forme d’incertitudes épistémiques. Ici, on se place dans un cadre simplifié où la famille  $\mathcal{F}$  est connue mais où les paramètres  $\theta$  sont inconnus.

Le cadre bayésien évoqué à la Section 1.4.1.3 n’est pas la seule façon de décrire les incertitudes épistémiques sur les paramètres. Il existe aussi un certain nombre d’approches extra-probabilistes émanant de la théorie des “probabilités imprécises” qui est formalisée dans [WALLEY \[1991, 2000\]](#). Au-delà de la théorie des probabilités présentée à la Section 1.2.1, les “probabilités imprécises” englobent la théorie des évidences, la théorie des possibilités et la théorie des *probability boxes* (ou *p-boxes*). Une mise en perspective de ces différentes théories est proposée dans [BREVAULT et collab. \[2020\]](#). En outre, [BEER et collab. \[2014\]](#) insistent sur la relation de complémentarité qui existe entre le point de vue bayésien et les approches extra-probabilistes.

D’une manière générale, la question du couplage de l’AS et des “probabilités imprécises” est peu abordée dans la littérature. [FERSON et TUCKER \[2006\]](#) s’y intéressent dans le cas des *p-boxes*

tandis que **HELTON et collab. [2006a]** font de même avec la théorie des évidences. Quelle que soit l'approche retenue pour encadrer la méconnaissance des paramètres, la propagation des incertitudes épistémiques sur la probabilité de défaillance implique systématiquement le recours à des simulations de Monte-Carlo et l'implémentation d'algorithmes à forte charge de calcul. En présence de *p-boxes*, **SCHÖBI et SUDRET [2019]** montrent que l'ASF peut être conduite après avoir construit un métamodèle dans l'espace augmenté.

Dans cette section, on se limite au cadre probabiliste bayésien qui est le plus répandu. Cela s'explique par le fait qu'il est plus simple à interpréter, qu'il offre une grande richesse de concepts mathématiques et qu'il se fonde naturellement dans le formalisme déjà utilisé pour les incertitudes aléatoires. On trouve dans **ZHANG et collab. [2009]** et **CHIU et collab. [2012]** des cas d'application de la modélisation bayésienne à des problèmes d'ingénierie géotechnique. L'aléa sur les paramètres est représenté par un vecteur aléatoire  $\Theta$  de loi *a priori*  $\mu_\Theta$  qui est fournie en amont et qui dépend d'un vecteur d'hyperparamètres  $\xi$ . La densité conjointe des entrées, notée jusqu'alors  $f_{\mathbf{X}}(\cdot | \mathcal{F}, \Theta)$  devient désormais  $f_{\mathbf{X}|\Theta}(\cdot | \Theta)$  pour insister sur le passage du point de vue fréquentiste au point de vue bayésien. Une hypothèse assez classique consiste à supposer que les différentes variables  $\Theta_1, \dots, \Theta_p$  sont indépendantes. Les fondements de la théorie bayésienne sont présentés avec plus de détails dans les ouvrages de **ROBERT [2007]** et **GELMAN et collab. [2013]**. Pour une réalisation  $\theta$  de  $\Theta$ , la probabilité de défaillance conditionnelle vaut :

$$P_f(\theta) = \int_{\mathbb{R}^d} \mathbb{1}_{D_f}(\mathbf{x}) f_{\mathbf{X}|\theta}(\mathbf{x} | \theta) d\mathbf{x}, \quad (3.58)$$

et l'estimateur CMC correspondant vient naturellement :

$$\hat{P}_f^{\text{CMC}}(\theta) = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbb{1}_{\{Y^{(i)} > T\}} \quad \text{avec} \quad \begin{cases} \mathbf{X}^{(i)} \stackrel{\text{i.i.d.}}{\sim} f_{\mathbf{X}|\theta}(\cdot | \theta) \\ Y^{(i)} = \phi(\mathbf{X}^{(i)}) \end{cases}. \quad (3.59)$$

Comme le résultat de l'analyse de fiabilité dépend de la valeur prise par  $\Theta$ , **DITLEVSEN [1979]** et **DER KIUREGHIAN [1989]** ont souhaité construire une mesure de fiabilité incorporant l'intégralité de l'information apportée par  $\mu_\Theta$ . Ces réflexions ont abouti à la définition de la probabilité de défaillance prédictive  $\tilde{P}_f$  comme l'espérance de  $P_f(\Theta)$ . Cette quantité peut alors s'exprimer de trois façons différentes :

$$\tilde{P}_f(\xi) = \mathbb{E}_{f_\Theta} [P_f(\Theta)] = \int_{D_\Theta} P_f(\theta) f_\Theta(\theta | \xi) d\theta = \int_{D_\Theta} \left( \int_{\mathbb{R}^d} \mathbb{1}_{D_f}(\mathbf{x}) f_{\mathbf{X}|\theta}(\mathbf{x} | \theta) d\mathbf{x} \right) f_\Theta(\theta | \xi) d\theta \quad (3.60)$$

$$= \int_{\mathbb{R}^d} \mathbb{1}_{D_f}(\mathbf{x}) \underbrace{\left( \int_{D_\Theta} f_{\mathbf{X}\Theta}(\mathbf{x}, \theta) d\theta \right)}_{\tilde{f}_{\mathbf{X}}(\mathbf{x})} d\mathbf{x} = \mathbb{E}_{\tilde{f}_{\mathbf{X}}} [\mathbb{1}_{D_f}(\mathbf{X})] \quad (3.61)$$

$$= \int_{\mathbb{R}^d \times D_\Theta} \mathbb{1}_{D_f}(\mathbf{x}) \underbrace{f_{\mathbf{X}|\theta}(\mathbf{x} | \theta) f_\Theta(\theta)}_{f_{\mathbf{X}\Theta}(\mathbf{x}, \theta)} d\mathbf{x} d\theta. \quad (3.62)$$

**DER KIUREGHIAN [2008]** dresse un état des lieux des techniques permettant d'estimer  $\tilde{P}_f$ . Elles correspondent aux différentes expressions intégrales présentées ci-dessus. La définition initiale de  $\tilde{P}_f$  conduit à l'expression (3.60) qui suggère l'utilisation d'une double boucle d'échantillonnage pour estimer  $\tilde{P}_f$ . On appelle NRA<sup>4</sup> (*Nested Reliability Approach*) l'estimateur obtenu en imbriquant les deux boucles de simulation :

$$\hat{P}_f^{\text{NRA}} = \frac{1}{N'_s} \sum_{k=1}^{N'_s} \left( \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbb{1}_{\{Y_k^{(i)} > T\}} \right) \quad \text{avec} \quad \Theta^{(k)} \stackrel{\text{i.i.d.}}{\sim} f_\Theta \quad \text{puis} \quad \begin{cases} \mathbf{X}_k^{(i)} \stackrel{\text{i.i.d.}}{\sim} f_{\mathbf{X}|\Theta}(\cdot | \Theta^{(k)}) \\ Y_k^{(i)} = \phi(\mathbf{X}_k^{(i)}) \end{cases}. \quad (3.63)$$

Par ailleurs, l'expression (3.61) fait intervenir la densité prédictive  $\tilde{f}_{\mathbf{X}}$  qui se déduit de la densité conjointe  $f_{\mathbf{X}\Theta} = f_{\mathbf{X}|\Theta} f_\Theta$  après intégration sur  $D_\Theta$ . Si un calcul analytique donne un accès explicite

4. Le sigle NRA (resp. ARA) est proposé dans **CHABRIDON [2018]**.

à  $\tilde{f}_{\mathbf{X}}$ , tout se ramène à une analyse de fiabilité classique mais cette situation n'est jamais rencontrée en pratique. Enfin, une alternative à NRA mise en lumière par l'expression (3.62) est de passer dans l'espace augmenté, c'est-à-dire de considérer  $\tilde{P}_f$  comme l'espérance d'une certaine fonction du vecteur  $\mathbf{Z} = [\mathbf{X}, \boldsymbol{\Theta}]$  qui agrège les entrées  $\mathbf{X}$  du code de calcul  $\phi$  et les paramètres  $\boldsymbol{\Theta}$  de la distribution. L'algorithme ARA (*Augmented Reliability Approach*) qui en résulte fait appel à une unique boucle d'échantillonnage sur  $\mathbf{Z}$  tout en gardant le conditionnement :

$$\hat{P}_f^{\text{ARA}} = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbb{1}_{\{Y^{(i)} > T\}} \quad \text{avec} \quad \begin{cases} \mathbf{Z}^{(i)} = (\mathbf{X}^{(i)}, \boldsymbol{\Theta}^{(i)}) \stackrel{\text{i.i.d.}}{\sim} f_{\mathbf{Z}} = f_{\mathbf{X}|\boldsymbol{\Theta}} f_{\boldsymbol{\Theta}} \\ Y^{(i)} = \phi(\mathbf{X}^{(i)}) \end{cases} . \quad (3.64)$$

ARA peut être combiné avec des algorithmes de fiabilité autres que CMC. Les possibilités offertes par l'exécution de la méthode FORM dans l'espace augmenté ont été étudiées par HONG [1996] puis PENDOLA [2000]. Le couplage avec des algorithmes adaptatifs (échantillonnage préférentiel ou *Subset Simulation*) est réalisé dans CHABRIDON et collab. [2017].

Si la probabilité de défaillance prédictive  $\tilde{P}_f$  est un indicateur pertinent, elle ne permet pas d'avoir une compréhension fine du rôle joué par chacun des paramètres incertains. Les prochains paragraphes s'attachent à présenter les méthodes développées pour répondre au problème d'ASF posé par  $\Psi : \boldsymbol{\Theta} \mapsto P_f(\boldsymbol{\Theta})$ .

### 3.3.2 Méthodes locales pour l'analyse de sensibilité fiabiliste

Comme on a pu le voir à la Section 3.2.1, l'AS locale pour une application  $\phi : \mathbf{X} \mapsto Y$  se fonde sur le calcul des indices (3.4) c'est-à-dire sur l'évaluation du gradient  $\nabla_{\mathbf{x}}\phi$  en un point  $\mathbf{x}^*$ . Si on s'intéresse à  $\Psi : \boldsymbol{\Theta} \mapsto P_f(\boldsymbol{\Theta})$ , une étude de la sensibilité locale consiste alors à calculer  $\nabla_{\boldsymbol{\theta}} P_f(\boldsymbol{\theta}^*)$  pour une valeur nominale  $\boldsymbol{\theta}^*$  donnée aux paramètres, par exemple  $\boldsymbol{\theta}^* = \mathbb{E}[\boldsymbol{\Theta}]$ . La variable d'intérêt  $P_f(\boldsymbol{\Theta})$  étant la probabilité de défaillance d'un évènement rare, on ne peut pas espérer estimer avec précision les indices locaux en utilisant un schéma aux différences finies. De par l'expression analytique (3.58), il existe des techniques spécifiques pour l'estimation des dérivées partielles de  $P_f(\boldsymbol{\Theta})$  par rapport aux paramètres  $\theta_j$ . La méthode à appliquer dépend de l'algorithme de fiabilité utilisé. Commençons par examiner le cas d'une analyse de fiabilité déterministe avec FORM. En reprenant les équations de la Section 2.2.2.1 et en explicitant le rôle de  $\boldsymbol{\theta}$ , il vient :

$$\beta(\boldsymbol{\theta}) := \min_{\mathbf{H}(\mathbf{w}, \boldsymbol{\theta})=T} \|\mathbf{w}\|_2 \quad \text{avec} \quad \mathbf{H}(\mathbf{w}, \boldsymbol{\theta}) = \phi(\mathcal{T}^{-1}(\mathbf{w}, \boldsymbol{\theta})) \quad \text{et} \quad P_f(\boldsymbol{\theta}) = 1 - \Phi(\beta(\boldsymbol{\theta})) . \quad (3.65)$$

Les transformations isoprobabilistes directe  $\mathbf{x} \mapsto \mathcal{T}(\mathbf{x}, \boldsymbol{\theta})$  et inverse  $\mathbf{w} \mapsto \mathcal{T}^{-1}(\mathbf{w}, \boldsymbol{\theta})$  dépendent de  $\boldsymbol{\theta}$  car elles sont construites à partir de  $f_{\mathbf{X}|\boldsymbol{\Theta}}(\cdot | \boldsymbol{\theta})$ . Par conséquent, la fonction de performance  $H$  dans l'espace normal standard est impactée par l'incertitude portée par  $\boldsymbol{\theta}$ . Le calcul de  $\nabla_{\boldsymbol{\theta}} P_f$  est entrepris dans HOHENBICHLER et RACKWITZ [1986] puis dans BJERAGER et KRENK [1987, 1989]. Il est facile de voir que :

$$\nabla_{\boldsymbol{\theta}} \beta(\boldsymbol{\theta}) = -\varphi(\beta(\boldsymbol{\theta})) \nabla_{\boldsymbol{\theta}} \beta(\boldsymbol{\theta}) , \quad (3.66)$$

et il reste à calculer  $\nabla_{\boldsymbol{\theta}} \beta$ . Le résultat analytique fait intervenir la matrice jacobienne de  $\mathcal{T}^{-1}$  ainsi que le vecteur normal unitaire  $\boldsymbol{\alpha}$  défini à l'équation (2.15) :

$$\nabla_{\boldsymbol{\theta}} \beta(\boldsymbol{\theta}) = \left( \mathbf{J}_{\boldsymbol{\theta}}[\mathcal{T}^{-1}](\mathbf{w}^*, \boldsymbol{\theta}) \right)^{\top} \boldsymbol{\alpha} \quad \text{avec} \quad \mathbf{J}_{\boldsymbol{\theta}}[\mathcal{T}^{-1}] := \begin{bmatrix} \frac{\partial \mathcal{T}_j^{-1}}{\partial \theta_k} \end{bmatrix}_{\substack{1 \leq j \leq d \\ 1 \leq k \leq p}} . \quad (3.67)$$

Pour un détail des calculs, on renvoie vers DITLEVSEN et MADSEN [1996]. Ainsi, pour calculer  $\nabla_{\boldsymbol{\theta}} \beta$ , il faut utiliser le point de conception  $\mathbf{w}^*$  rendu par FORM et connaître l'expression des dérivées partielles des fonctions coordonnées  $\mathcal{T}_j^{-1}$  de la transformation isoprobabiliste inverse. Dans les premiers travaux à ce sujet, les variables  $\mathbf{X}_j$  sont supposées indépendantes et le vecteur  $\boldsymbol{\theta}$  agrège les paramètres  $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_d$  des lois marginales. Lorsque la matrice de corrélation  $\boldsymbol{\rho}$  est fournie et que l'utilisateur décide de l'incorporer à l'analyse de fiabilité en appliquant la transformation de

Nataf décrite à la Section 2.2.1.3, la probabilité de défaillance peut alors s'écrire  $P_f(\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_d, \boldsymbol{\rho})$ . Le calcul de  $\nabla_{\boldsymbol{\rho}} P_f$  est proposé dans ŽANIC et ŽIHA [1998, 2001] puis dans BOURINET [2017].

On a pu voir au Chapitre 2 que FORM ne convient qu'à de trop rares situations et que ce sont les algorithmes basés sur l'échantillonnage de particules au voisinage du domaine de défaillance  $D_f$  qui assurent le plus de précision. L'équation (3.59) rappelle l'expression de l'estimateur CMC dans un contexte bayésien. KARAMCHANDANI [1990] puis WU [1994b] ont mis en avant le rôle central que peut jouer le score  $\mathbf{x} \mapsto \nabla_{\boldsymbol{\theta}} \log f_{\mathbf{X}|\boldsymbol{\theta}}(\mathbf{x} | \boldsymbol{\theta})$  que l'on a déjà rencontré<sup>5</sup> à plusieurs reprises. En effet, des calculs élémentaires donnent :

$$\Delta_j^F(\boldsymbol{\theta}) := \frac{\partial P_f}{\partial \theta_j}(\boldsymbol{\theta}) = \int_{\mathbb{R}^d} \mathbb{1}_{D_f}(\mathbf{x}) \frac{\partial f_{\mathbf{X}|\boldsymbol{\theta}}}{\partial \theta_k}(\mathbf{x} | \boldsymbol{\theta}) d\mathbf{x} = \int_{\mathbb{R}^d} \mathbb{1}_{D_f}(\mathbf{x}) \frac{\partial \log f_{\mathbf{X}|\boldsymbol{\theta}}}{\partial \theta_k}(\mathbf{x} | \boldsymbol{\theta}) f_{\mathbf{X}|\boldsymbol{\theta}}(\mathbf{x} | \boldsymbol{\theta}) d\mathbf{x}. \quad (3.68)$$

En réunissant toutes les équations de ce type, on peut écrire que :

$$\nabla_{\boldsymbol{\theta}} P_f(\boldsymbol{\theta}) = \mathbb{E} \left[ \mathbb{1}_{D_f}(\mathbf{X}) \nabla_{\boldsymbol{\theta}} \log f_{\mathbf{X}|\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}) \right]. \quad (3.69)$$

Les indices locaux définis à l'équation (3.68) peuvent alors être estimés comme suit :

$$\hat{\Delta}_j^F(\boldsymbol{\theta}^*) = \frac{1}{N_s} \sum_{j=1}^{N_s} \mathbb{1}_{\{Y^{(i)} > T\}} \frac{\partial \log f_{\mathbf{X}|\boldsymbol{\theta}}}{\partial \theta_j}(\mathbf{X}^{(i)} | \boldsymbol{\theta}^*) \quad \text{avec} \quad \begin{cases} \mathbf{X}^{(i)} \stackrel{\text{i.i.d.}}{\sim} f_{\mathbf{X}|\boldsymbol{\theta}}(\cdot | \boldsymbol{\theta}^*) \\ Y^{(i)} = \phi(\mathbf{X}^{(i)}) \end{cases}. \quad (3.70)$$

Les échantillons  $\mathbf{X}^{(i)}$  et  $Y^{(i)}$  n'ont pas besoin d'être simulés. Ils sont déjà disponibles car ils ont été fabriqués au cours de la phase préliminaire d'estimation de  $\hat{P}_f(\boldsymbol{\theta}^*)$ . Les indices  $\hat{\Delta}_j^F(\boldsymbol{\theta}^*)$  sont donc des produits dérivés de l'analyse de fiabilité. Toutefois, il faut être en mesure d'évaluer le score  $\nabla_{\boldsymbol{\theta}} \log f_{\mathbf{X}|\boldsymbol{\theta}}$  pour chaque échantillon d'entrée  $\mathbf{X}^{(i)}$ . MILLWATER [2009] parle des "kernel functions" pour désigner les différentes dérivées partielles  $\frac{\partial}{\partial \theta_j} \log f_{\mathbf{X}|\boldsymbol{\theta}}$  et celles-ci sont notées  $\kappa_{\theta_j}$  selon une convention souvent adoptée dans la littérature. Lorsque les variables sont indépendantes, il est assez facile d'obtenir analytiquement les fonctions  $\kappa_{\theta_j}$  car la situation est la suivante :

$$\log f_{\mathbf{X}|\boldsymbol{\theta}} = \sum_{j=1}^d \log f_{X_j|A_j} \quad \text{et} \quad \forall 1 \leq j \leq d, \quad \nabla_{\boldsymbol{\alpha}_j} \log f_{\mathbf{X}|\boldsymbol{\theta}} = \nabla_{\boldsymbol{\alpha}_j} \log f_{X_j|A_j}. \quad (3.71)$$

Tout se ramène en fait au calcul du score des différentes densités marginales. Les formules des fonctions  $\kappa_{\theta_j}$  sont précisées dans RAHMAN [2009] pour les lois gaussienne et log-normale. Lorsque les entrées sont dépendantes et que la structure de dépendance est modélisée par une copule, les calculs sont moins immédiats. Pour un paramètre  $\theta_j$  se rapportant à la  $j$ -ème entrée, on a :

$$\begin{aligned} \log f_{\mathbf{X}|\boldsymbol{\theta}} &= \sum_{j=1}^d \log f_{X_j|A_j} + \log c_{\mathbf{U}|\mathbf{B}}(F_{X_1|A_1}, \dots, F_{X_d|A_d}), \\ \text{et donc : } \kappa_{\theta_j} = \frac{\partial \log f_{\mathbf{X}|\boldsymbol{\theta}}}{\partial \theta_j} &= \frac{\partial \log f_{X_j|A_j}}{\partial \theta_j} + f_{X_j|A_j} \frac{\partial \log c_{\mathbf{U}|\mathbf{B}}}{\partial u_j}(F_{X_1|A_1}, \dots, F_{X_d|A_d}). \end{aligned} \quad (3.72)$$

En plus des scores marginaux  $\nabla_{\boldsymbol{\alpha}_j} \log f_{X_j|A_j}$ , il faut donc aussi avoir accès au gradient  $\nabla_{\mathbf{u}} \log c_{\mathbf{U}|\mathbf{B}}$ . Pour les principales familles de copules paramétriques, un formulaire est notamment disponible dans LEE et collab. [2011]. De façon assez étonnante, les indices locaux ne sont définis que pour les paramètres marginaux  $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_d$  et il n'est nulle part fait mention d'indices locaux pour les paramètres  $\boldsymbol{\beta}$  de la copule.

Souvent, l'approche CMC est trop grossière pour permettre de conduire l'analyse de fiabilité et il faut recourir à des algorithmes adaptatifs plus sophistiqués pour exécuter la tâche  $\boldsymbol{\theta}^* \mapsto \hat{P}_f(\boldsymbol{\theta}^*)$ . Néanmoins, en manipulant légèrement la formule (3.69), il reste possible d'estimer "gratuitement"

5. Dans le Chapitre 1, le score  $\nabla_{\boldsymbol{\theta}} \log f_{\mathbf{X}|\boldsymbol{\theta}}$  apparaît dans l'expression du gradient de la log-vraisemblance. Dans le Chapitre 2, le score  $\nabla_{\boldsymbol{\lambda}} \log h_{\boldsymbol{\lambda}}$  intervient dans le processus de mise à jour des paramètres de l'algorithme m-CE.

les indices locaux  $\Delta_j^F(\boldsymbol{\theta}^*)$  à l'aide des particules recueillies à la fin de l'exploration du domaine de défaillance. Cette démarche est adoptée dans [WU \[1994a\]](#) pour l'échantillonnage préférentiel adaptatif, dans [ZHANG et collab. \[2010\]](#) pour l'échantillonnage préférentiel stratifié et également dans [SONG et collab. \[2009\]](#) pour l'algorithme de *Subset Simulation*.

Si on s'intéresse à la probabilité de défaillance prédictive  $\tilde{P}_f$  de l'équation (3.60), on constate qu'elle dépend des hyperparamètres de la densité *a priori*  $f_{\boldsymbol{\xi}}(\cdot | \boldsymbol{\xi})$ . [CHABRIDON et collab. \[2018\]](#) proposent d'étudier la sensibilité à ces hyperparamètres via le gradient  $\nabla_{\boldsymbol{\xi}} \tilde{P}_f$  et démontrent que les dérivées partielles  $\frac{\partial}{\partial \xi_j} \tilde{P}_f$  peuvent être obtenues à partir des échantillons  $(\mathbf{X}^{(i)}, \boldsymbol{\Theta}^{(i)})$  simulés au moment d'estimer  $\tilde{P}_f$  avec ARA dans le cas d'une loi *a priori*  $\mu_{\boldsymbol{\theta}}$  à support non borné.

### 3.3.3 Méthodes globales pour l'analyse de sensibilité fiabiliste

Dans un cadre bayésien où la méconnaissance de la distribution d'entrée  $f_{\mathbf{X}|\boldsymbol{\theta}}$  est décrite par une loi *a priori*  $\mu_{\boldsymbol{\theta}}$  qui régit le comportement aléatoire des paramètres, les méthodes locales d'ASF sont assez peu pertinentes. En effet, on souhaite acquérir une information sur l'évolution de  $P_f(\boldsymbol{\theta})$  suite à des variations de  $\boldsymbol{\theta}$  qui ne se limitent pas à un voisinage de  $\boldsymbol{\theta}^*$  mais qui peuvent couvrir la totalité du domaine  $D_{\boldsymbol{\theta}}$ . Pour apporter une réponse globale au problème d'ASF, on peut s'appuyer sur les outils d'AS globale présentés à la Section 3.2.2 mais ils doivent obligatoirement être intégrés à des schémas d'estimation compatibles avec le budget de calcul qui est accordé à l'ASF.

Tout d'abord, les méthodes de criblage de la Section 3.2.1 peuvent être adaptées à ce contexte d'ASF. [XIAO et collab. \[2016\]](#) reprennent l'algorithme de Morris dans cet objectif. Pour cela, il faut se munir d'un plan d'expérience  $\mathcal{X}_{\boldsymbol{\theta}} := \{\boldsymbol{\theta}^{(i)}\}_{i=1}^{N_s}$  et estimer en chaque point  $\boldsymbol{\theta}^{(i)}$  tous les indices  $\Delta_j^F(\boldsymbol{\theta}^{(i)}) := \frac{\partial}{\partial \theta_j} P_f(\boldsymbol{\theta}^{(i)})$ . On ne peut pas utiliser les estimateurs (3.70) spécifiques à l'ASF locale car cela impose d'estimer avec un algorithme de fiabilité stochastique (échantillonnage préférentiel ou *Subset Simulation*) toutes les probabilités de défaillance  $P_f(\boldsymbol{\theta}^{(i)})$  et on ne peut pas l'envisager. Il faut se résoudre à utiliser un schéma aux différences finies :

$$\forall 1 \leq i \leq N_s, \quad \forall 1 \leq j \leq p, \quad \hat{\Delta}_j^F(\boldsymbol{\theta}^{(i)}) = \frac{\hat{P}_f(\boldsymbol{\theta}^{(i)} + \delta_0 \mathbf{e}_j) - \hat{P}_f(\boldsymbol{\theta}^{(i)})}{\delta_0} . \quad (3.73)$$

Toutes les probabilités de défaillance sont estimées avec la méthode de [ZHANG et PANDEY \[2013\]](#) dont le coût de calcul est extrêmement réduit. Cette technique de fiabilité ne donne toutefois pas accès à des échantillons proches de la zone de défaillance  $D_f$  et ne peut donc pas être couplée à la formule (3.70) pour estimer plus subtilement les dérivées partielles. En se calquant sur ce qui est fait pour  $\phi$  à l'équation (3.6), la réponse du problème d'ASF est donnée par :

$$\forall 1 \leq j \leq d, \quad \hat{\mu}_{|\hat{\Delta}_j^F|} := \frac{1}{N_s} \sum_{i=1}^{N_s} \left| \hat{\Delta}_j^F(\boldsymbol{\theta}^{(i)}) \right| \quad \text{et} \quad \hat{\sigma}_{\hat{\Delta}_j^F}^2 := \frac{1}{N_s} \sum_{i=1}^{N_s} \left( \hat{\Delta}_j^F(\boldsymbol{\theta}^{(i)}) \right)^2 - \hat{\mu}_{\hat{\Delta}_j^F}^2 . \quad (3.74)$$

Dans l'esprit des DGSM définies à l'équation (3.7), [WANG et collab. \[2013b\]](#) posent :

$$\forall 1 \leq j \leq d, \quad SP_j(\boldsymbol{\theta}^*) := \int_{D_{\theta_j}} \left| \frac{\partial P_f}{\partial \theta_j}(\theta_j, \boldsymbol{\theta}_{-j}^*) \right| d\theta_j = \mathbb{E} \left[ \frac{1}{f_{\Theta_j}(\Theta_j)} \left| \frac{\partial P_f}{\partial \theta_j}(\Theta_j, \boldsymbol{\theta}_{-j}^*) \right| \right] . \quad (3.75)$$

avec  $D_{\theta_j}$  la plage de variation du paramètre  $\Theta_j$ . On souhaite insister sur le fait que les indices  $SP_j(\boldsymbol{\theta}^*)$  ne sont pas exactement les DGSM associés à l'application  $\Psi$ . En effet, ils dépendent d'un point de référence  $\boldsymbol{\theta}^*$  et l'intégration des dérivées partielles est unidimensionnelle. À partir de l'expression sous forme d'espérance dans le membre de droite de l'équation (3.75), l'idée d'une double boucle d'échantillonnage vient assez naturellement. Pour que le budget de calcul ne soit pas exorbitant, une approche QMC est mise en œuvre dans chaque boucle d'échantillonnage :

- Il faut commencer par générer une suite à discrétion faible  $\{\theta_j^{(i)}\}_{i=1}^{N_1}$  dans  $D_{\theta_j}$ .
- Chaque dérivée partielle  $\frac{\partial P_f}{\partial \theta_j}(\theta_j, \boldsymbol{\theta}_{-j}^*)$  est réécrite sous forme intégrale avec la formule (3.68) avant d'effectuer un changement de variables qui envoie  $\mathbb{R}^d$  dans  $\mathcal{H}_d$ . L'estimation repose alors sur la fabrication d'une suite à discrétion faible  $\{\mathbf{u}^{(k)}\}_{k=1}^{N_2}$  dans  $\mathcal{H}_d$ .

Au final, l'estimation de tous les indices  $SP_j$  implique  $d \times N_1 \times N_2$  appels à  $\phi$ . Pour atteindre le même niveau de précision avec la méthode de simulation de Monte-Carlo, il faudrait prendre  $N'_1 \gg N_1$  et  $N'_2 \gg N_2$ . Au-delà des techniques de criblage, l'AS de la fonction  $\Psi$  peut évidemment viser à attribuer à chaque paramètre  $\theta_j$  la part de la variance de  $P_f(\boldsymbol{\theta})$  qu'il a permis d'engendrer. En raison de l'indépendance supposée entre les variables de  $\boldsymbol{\theta}$ , l'obtention d'une décomposition ANOVA de  $\Psi$  ne pose aucune difficulté théorique :

$$P_f(\boldsymbol{\theta}) := \sum_{\mathbf{u} \subseteq [d]} \Psi_{\mathbf{u}}(\boldsymbol{\theta}_{\mathbf{u}}) \quad \text{avec} \quad \begin{cases} \Psi_0 = \mathbb{E}[P_f(\boldsymbol{\theta})] = \tilde{P}_f, \\ \Psi_{\mathbf{u}}(\boldsymbol{\theta}_{\mathbf{u}}) = \mathbb{E}[P_f(\boldsymbol{\theta}) | \boldsymbol{\theta}_{\mathbf{u}}] - \sum_{\mathbf{v} \subset \mathbf{u}} \Psi_{\mathbf{v}}(\boldsymbol{\theta}_{\mathbf{v}}). \end{cases} \quad (3.76)$$

Les indices de Sobol fermés et totaux sont alors définis comme suit :

$$\forall \mathbf{u} \subseteq [d], \quad S_{\mathbf{u}}^c := \frac{\mathbb{V}(\mathbb{E}[P_f(\boldsymbol{\theta}) | \boldsymbol{\theta}_{\mathbf{u}}])}{\mathbb{V}(P_f(\boldsymbol{\theta}))} = \frac{V_{\mathbf{u}}^F}{\mathbb{V}(P_f(\boldsymbol{\theta}))} \quad \text{et} \quad S_{\mathbf{u}}^T := 1 - \frac{\mathbb{V}(\mathbb{E}[P_f(\boldsymbol{\theta}) | \boldsymbol{\theta}_{-\mathbf{u}}])}{\mathbb{V}(P_f(\boldsymbol{\theta}))}. \quad (3.77)$$

Parmi les nombreux schémas d'estimation présentés à la Section 3.2.2, la procédure d'ASF mise au point par MORIO [2011] fait appel au *Pick-Freeze*. Les numérateurs  $V_{\mathbf{u}}^F$  des indices  $S_{\mathbf{u}}^c$  sont alors estimés comme suit :

$$\hat{V}_{\mathbf{u}}^F = \frac{1}{N_s} \sum_{i=1}^{N_s} \Psi(\boldsymbol{\theta}^{(i)}) \Psi(\boldsymbol{\theta}_{\mathbf{u}}^{(i)}, \tilde{\boldsymbol{\theta}}_{-\mathbf{u}}^{(i)}) - \Psi_0^2 \approx \frac{1}{N_s} \sum_{i=1}^{N_s} \hat{P}_f(\boldsymbol{\theta}^{(i)}) \hat{P}_f(\boldsymbol{\theta}_{\mathbf{u}}^{(i)}, \tilde{\boldsymbol{\theta}}_{-\mathbf{u}}^{(i)}) - (\hat{P}_f^{\text{NRA}})^2. \quad (3.78)$$

Une première étape consiste à simuler deux blocs  $\boldsymbol{\theta}_{\text{sim}}$  et  $\tilde{\boldsymbol{\theta}}_{\text{sim}}$  de  $N_s$  échantillons de la loi *a priori*  $\mu_{\boldsymbol{\theta}}$ . La seconde étape est celle de l'estimation de toutes les probabilités de défaillance. Si on veut obtenir les indices de Sobol du premier ordre, il faut exécuter l'algorithme de fiabilité à  $(d+1) \times N_s$  reprises. Comme cette charge de calcul n'est pas supportable, MORIO [2011] propose d'utiliser l'astuce "*inverse-reverse*"<sup>6</sup> qui repose sur une unique exploration du domaine de défaillance par échantillonnage préférentiel. On rappelle que la densité auxiliaire optimale pour l'estimation de  $P_f(\boldsymbol{\theta})$  est donnée par :

$$\forall \mathbf{x} \in \mathbb{R}^d, \quad h_{\text{opt}}^{\boldsymbol{\theta}}(\mathbf{x}) := \frac{\mathbb{1}_{D_f}(\mathbf{x}) f_{\mathbf{X}|\boldsymbol{\theta}}(\mathbf{x} | \boldsymbol{\theta})}{P_f(\boldsymbol{\theta})}. \quad (3.79)$$

L'idée de l'astuce "*inverse-reverse*" est de construire une unique densité auxiliaire  $h_1$  et de l'utiliser pour estimer la probabilité  $P_f(\boldsymbol{\theta}_0)$  quelle que soit la valeur de  $\boldsymbol{\theta}_0 \in D_{\boldsymbol{\theta}}$  :

$$\hat{P}_f^{\text{IS}}(\boldsymbol{\theta}_0) = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbb{1}_{\{Y^{(i)} > T\}} \frac{f_{\mathbf{X}|\boldsymbol{\theta}}(\mathbf{X}^{(i)} | \boldsymbol{\theta}_0)}{h_1(\mathbf{X}^{(i)})} \quad \text{avec} \quad \begin{cases} \mathbf{X}^{(i)} \stackrel{\text{i.i.d.}}{\sim} h_1 \\ Y^{(i)} = \phi(\mathbf{X}^{(i)}) \end{cases}. \quad (3.80)$$

L'estimateur obtenu avec  $h_1$  est non-biaisé mais sa variance dépend de la proximité entre  $h_1$  et  $h_{\text{opt}}^{\boldsymbol{\theta}_0}$ . Par conséquent, une bonne option peut être de choisir  $h_1 = \hat{h}_{\text{opt}}^{\boldsymbol{\theta}^*}$  avec  $\boldsymbol{\theta}^* = \mathbb{E}[\boldsymbol{\theta}]$ . Il suffit de lancer un algorithme d'échantillonnage préférentiel à partir de  $f_{\mathbf{X}|\boldsymbol{\theta}}(\cdot | \boldsymbol{\theta}^*)$  et de réutiliser  $\hat{h}_{\text{opt}}^{\boldsymbol{\theta}^*}$  ainsi que les échantillons  $\mathbf{X}^{(i)}$  et  $Y^{(i)}$  fabriqués à la dernière itération. Toute probabilité  $P_f(\boldsymbol{\theta}_0)$  peut alors être estimée sans aucun appel supplémentaire à  $\phi$ . L'astuce "*inverse-reverse*" est d'abord apparue dans BECKMAN et MCKAY [1987] pour mettre à jour une quantité  $\eta(\boldsymbol{\theta})$  estimée à partir de  $f_{\mathbf{X}|\boldsymbol{\theta}}(\cdot | \boldsymbol{\theta})$  grâce à une technique d'échantillonnage préférentiel. Si une légère perturbation  $\delta_{\boldsymbol{\theta}}$  affecte les paramètres, il est alors possible de recalculer "gratuitement"  $\eta(\boldsymbol{\theta} + \delta_{\boldsymbol{\theta}})$  en recyclant les échantillons impliqués dans le calcul initial de  $\eta(\boldsymbol{\theta})$ . Dans le même esprit, HESTERBERG [1996] utilise l'astuce "*inverse-reverse*" pour construire à moindre coût un intervalle de confiance de  $\eta(\boldsymbol{\theta})$ .

Sous la contrainte d'un budget de simulation restreint, une autre stratégie pour l'estimation des indices de Sobol est la mise en place d'un métamodèle  $\hat{\Psi}$  de l'application  $\Psi : \boldsymbol{\theta} \mapsto P_f(\boldsymbol{\theta})$ . Si la métamodélisation de la fonction de performance  $\phi$  est une pratique désormais très courante,

6. On retrouve cette appellation dans CHABRIDON [2018]. On peut en attribuer la paternité à LEMAÎTRE [2014].

la métamodélisation globale d'une analyse de fiabilité est une démarche bien moins répandue. Il faut commencer par constituer un ensemble d'apprentissage  $\mathcal{X} := \{\boldsymbol{\theta}^{(i)}, P_f(\boldsymbol{\theta}^{(i)})\}_{i=1}^n$  en exécutant à  $n$  reprises l'algorithme de fiabilité. C'est au cours de cette étape qu'est dépensé le budget de calcul alloué à l'estimation des indices de Sobol. Le krigeage est une technique de métamodélisation très courante qui permet de construire un prédicteur linéaire de la sortie dont on sait encadrer l'erreur. Elle a l'avantage de pouvoir être mise en œuvre même à partir d'une base d'apprentissage assez restreinte. **WANG et collab. [2013a]** proposent de calibrer un processus gaussien  $\hat{\Psi}_{\text{PG}}$  à partir de  $\mathcal{X}$ . Ils expliquent que l'intérêt majeur du krigeage dans ce contexte vient du fait qu'on dispose d'une expression analytique de la fonction  $\mu_{\mathbf{u}} : \boldsymbol{\theta}_{\mathbf{u}} \mapsto \mathbb{E}[\hat{\Psi}_{\text{PG}}(\boldsymbol{\theta}) \mid \boldsymbol{\theta}_{\mathbf{u}} = \boldsymbol{\theta}_{\mathbf{u}}]$ . On a alors :

$$V_{\mathbf{u}}^F = \mathbb{V}\left(\mathbb{E}[P_f(\boldsymbol{\theta}) \mid \boldsymbol{\theta}_{\mathbf{u}}]\right) \approx \mathbb{V}\left(\mathbb{E}[\hat{\Psi}_{\text{PG}}(\boldsymbol{\theta}) \mid \boldsymbol{\theta}_{\mathbf{u}}]\right) = \mathbb{V}(\mu_{\mathbf{u}}(\boldsymbol{\theta}_{\mathbf{u}})). \quad (3.81)$$

Il suffit alors de calculer la variance empirique des espérances conditionnelles  $\mu_{\mathbf{u}}(\boldsymbol{\theta}_{\mathbf{u}}^{(k)})$  obtenues à partir d'une séquence  $\{\boldsymbol{\theta}_{\mathbf{u}}^{(k)}\}_{k=1}^{N_s}$  d'observations du vecteur  $\boldsymbol{\theta}_{\mathbf{u}}$ . **WANG et collab. [2013c]** montrent qu'un schéma d'estimation très similaire peut être implémenté après avoir utilisé la méthode des moindres carrés mobiles de **LANCASTER et SALKAUSKAS [1981]** pour construire un métamodèle  $\hat{\Psi}$ . Par ailleurs, **EHRE et collab. [2018]** proposent la métamodélisation successive de la fonction de performance  $\phi$  puis de l'analyse de fiabilité  $\Psi$ . Les métamodèles  $\hat{\phi}$  et  $\hat{\Psi}$  sont respectivement construits à l'aide du chaos polynomial et d'une approximation de rang faible.

Si la métamodélisation permet de soulager la charge de calcul, la réponse de l'ASF devient dépendante de la qualité du métamodèle  $\hat{\Psi}$  et de son ensemble d'apprentissage  $\mathcal{X}$ . Si on ne peut exécuter qu'un très petit nombre de fois l'algorithme de fiabilité, l'erreur commise en remplaçant  $\Psi$  par  $\hat{\Psi}$  peut être conséquente et conduire à des résultats d'ASF erronés. Pour contourner cet écueil, une méthode appelée NISS (*Non-Intrusive Stochastic Simulation*) est mise à l'honneur dans **WEI et collab. [2019a,b]** ainsi que dans **SONG et collab. [2019]**. On suppose que la probabilité de défaillance prédictive  $\tilde{P}_f$  a été estimée avec ARA et que les différents échantillons  $(\boldsymbol{\theta}^{(i)}, \mathbf{X}^{(i)}, Y^{(i)})$  ont été récupérés. Il est alors possible d'en déduire une estimation de tous les indices de Sobol sans aucun appel supplémentaire à  $\phi$ . Prenons l'exemple d'un indice du premier ordre  $S_j$  pour lequel il faut estimer  $V_j^F = \mathbb{V}\left(\mathbb{E}[P_f(\boldsymbol{\theta}) \mid \Theta_j]\right)$ . Au départ, NISS se contente de reprendre la méthode CRM. Il est simplement question de calculer la variance empirique des espérances conditionnelles  $\mu_j(\theta_j^{(k)}) = \mathbb{E}[P_f(\boldsymbol{\theta}) \mid \Theta_j = \theta_j^{(k)}]$  pour une séquence  $\{\theta_j^{(k)}\}_{k=1}^{N_s}$  d'observations de  $\Theta_j$ . Reste à savoir comment estimer les espérances conditionnelles  $\mu_j(\theta_j^{(k)})$ . Pour une valeur  $\theta_j^*$  fixée, on a :

$$\begin{aligned} \mu_j(\theta_j^*) &= \mathbb{E}[P_f(\boldsymbol{\theta}) \mid \Theta_j = \theta_j^*] = \mathbb{E}_{f_{\Theta_{-j}}} [P_f(\theta_j^*, \boldsymbol{\theta}_{-j})] = \int_{D_{\Theta_{-j}}} P_f(\theta_j^*, \boldsymbol{\theta}_{-j}) f_{\Theta_{-j}}(\boldsymbol{\theta}_{-j}) d\boldsymbol{\theta}_{-j} \\ &= \left( \int_{\mathbb{R}^d \times D_{\Theta_{-j}}} \mathbb{1}_{D_f}(\mathbf{x}) f_{\mathbf{X}|\boldsymbol{\theta}}(\mathbf{x} \mid \theta_j^*, \boldsymbol{\theta}_{-j}) f_{\Theta_{-j}}(\boldsymbol{\theta}_{-j}) d\boldsymbol{\theta}_{-j} d\mathbf{x} \right) \times \underbrace{\left( \int_{D_{\Theta_j}} f_{\Theta_j}(\theta_j) d\theta_j \right)}_{=1} \\ &= \int_{\mathbb{R}^d \times D_{\boldsymbol{\theta}}} \mathbb{1}_{D_f}(\mathbf{x}) \frac{f_{\mathbf{X}|\boldsymbol{\theta}}(\mathbf{x} \mid \theta_j^*, \boldsymbol{\theta}_{-j})}{f_{\mathbf{X}|\boldsymbol{\theta}}(\mathbf{x} \mid \boldsymbol{\theta})} \underbrace{f_{\mathbf{X}|\boldsymbol{\theta}}(\mathbf{x} \mid \boldsymbol{\theta}) f_{\boldsymbol{\theta}}(\boldsymbol{\theta})}_{=f_{\mathbf{X}\boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta})} d\mathbf{x} d\boldsymbol{\theta} \\ &= \mathbb{E}_{f_{\mathbf{X}\boldsymbol{\theta}}} \left[ \mathbb{1}_{D_f}(\mathbf{X}) \frac{f_{\mathbf{X}|\boldsymbol{\theta}}(\mathbf{X} \mid \theta_j^*, \boldsymbol{\theta}_{-j})}{f_{\mathbf{X}|\boldsymbol{\theta}}(\mathbf{X} \mid \boldsymbol{\theta})} \right]. \end{aligned} \quad (3.82)$$

Il est ainsi possible d'estimer  $\mu_j(\theta_j^*)$  avec :

$$\hat{\mu}_j(\theta_j^*) = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbb{1}_{\{Y^{(i)} > T\}}(\mathbf{X}^{(i)}) \frac{f_{\mathbf{X}|\boldsymbol{\theta}}(\mathbf{X}^{(i)} \mid \theta_j^*, \boldsymbol{\theta}_{-j}^{(i)})}{f_{\mathbf{X}|\boldsymbol{\theta}}(\mathbf{X}^{(i)} \mid \boldsymbol{\theta}^{(i)})} \quad \text{avec} \quad \begin{cases} (\mathbf{X}^{(i)}, \boldsymbol{\theta}^{(i)}) \stackrel{\text{i.i.d.}}{\sim} f_{\mathbf{X}\boldsymbol{\theta}} \\ Y^{(i)} = \phi(\mathbf{X}^{(i)}) \end{cases}. \quad (3.83)$$

Ainsi, des estimations de toutes les espérances conditionnelles  $\mu_j(\theta_j^{(k)})$  peuvent être déduites des échantillons produits initialement par ARA. Un calcul de variance empirique termine la procédure

et donne accès à une estimation de  $V_j^F$ . Les indices de Sobol du premier ordre peuvent donc être vus comme des produits dérivés de l'estimation de la probabilité de défaillance prédictive  $\tilde{P}_f$ . Il en est de même pour les indices d'ordre supérieur. Pour les indices du second ordre, la formule qui permet de réécrire  $\mu_{ij}(\boldsymbol{\theta}_{ij}^*)$  avec l'opérateur  $\mathbb{E}_{f_{x\theta}}$  est donnée dans [SONG et collab. \[2019\]](#).

### 3.3.4 Bilan

Les différentes méthodes présentées dans cette section montrent que les outils classiques d'AS ne peuvent pas être directement déployés dans un contexte d'ASF. En effet, leur implémentation la plus naïve engendre un coût de calcul qui dépasse largement la charge maximale fixée par le budget de simulation. Dans le cas des méthodes locales d'ASF, différentes astuces permettent de soulager la charge de calcul (*kernel functions*, simulation QMC, techniques fiabilistes à bas coût). Dans le cas des méthodes globales d'ASF, différentes stratégies ont été adoptées afin de mettre en place des procédures viables d'un point de vue calculatoire. Ainsi, le budget de simulation peut être dépensé pour :

- Construire une densité auxiliaire  $h_1$  qui permet par la suite d'estimer "gratuitement" toutes les probabilités de défaillance par échantillonnage préférentiel.
- Constituer un ensemble d'apprentissage  $\mathcal{X}$  destiné à la métamodélisation de  $\Psi$ .
- Estimer la probabilité de défaillance prédictive  $\tilde{P}_f$  avec ARA et faire ensuite appel à NISS pour le post-traitement.

On peut alors faire deux remarques. D'une part, on constate qu'il est nécessaire de dégrader la qualité de l'analyse de fiabilité (utilisation d'une unique densité auxiliaire ou d'un métamodèle) pour être en mesure de calculer les probabilités de défaillance associées à un très grand nombre de valeurs différentes prises par  $\boldsymbol{\theta}$ . D'autre part, on constate que seules des techniques d'AS pour entrées indépendantes ont été mises en application. Cela vient du fait que les variables  $\Theta_1, \dots, \Theta_p$  sont supposées indépendantes dans le cadre bayésien le plus courant. La situation est différente si on considère des paramètres  $\hat{\boldsymbol{\theta}}$  issus d'un apprentissage statistique :

$$\mathbf{X}_{\text{obs}} \sim (\mu_{\mathbf{X}})^{\otimes N} \xrightarrow{\mathcal{L}} f_{\mathbf{X}}(\cdot | \hat{\boldsymbol{\theta}}, \mathcal{F}) \xrightarrow{\mathcal{R}} P_f(\hat{\boldsymbol{\theta}}). \quad (3.84)$$

Comme ils sont tous estimés à partir d'un même échantillon  $\mathbf{X}_{\text{obs}}$ , les paramètres  $\hat{\boldsymbol{\theta}}$  ne sont plus indépendants. L'utilisation de méthodes spécifiques aux entrées dépendantes est alors justifiée et même nécessaire.

## 3.4 Conclusion

Dans ce chapitre, on s'est intéressé à une large palette d'outils d'AS. Une attention particulière a été accordée aux méthodes permettant de propager et quantifier les incertitudes épistémiques liées à la méconnaissance des paramètres de la loi des entrées. Le point de vue bayésien fournit un cadre mathématique confortable pour manipuler les incertitudes épistémiques. Un pré-requis est néanmoins la connaissance de la loi *a priori* des paramètres et l'obtention de cette information peut s'avérer compliquée en pratique.

Les incertitudes épistémiques peuvent également être décrites en s'émancipant de tout cadre paramétrique. Dans [LEMAÎTRE et collab. \[2015\]](#), ce sont des perturbations unidirectionnelles qui sont appliquées successivement aux entrées du système. Les indices PLI (*Perturbed-Law Indices*) permettent alors de mesurer la sensibilité de la probabilité de défaillance à cette nouvelle forme d'incertitudes épistémiques. Pour une réduction maximale des coûts de calcul, l'estimation des indices PLI peut s'appuyer sur l'échantillonnage préférentiel et l'astuce "*inverse-reverse*".

Un autre problème d'AS qui touche de près la communauté fiabiliste est celui de l'analyse de sensibilité conditionnelle. Il s'agit d'identifier l'entrée  $X_j$  qui a le plus d'impact sur l'occurrence

d'une défaillance, c'est-à-dire sur la fonction indicatrice du domaine de défaillance  $\mathbb{1}_{D_f}(\mathbf{X})$ . Étant donné que l'énoncé initial de ce problème ne fait pas mention d'incertitudes épistémiques sur la distribution des entrées, il a été décidé de ne pas y consacrer une section spécifique dans ce chapitre. On souligne que c'est un thème de recherche qui suscite actuellement un fort intérêt. En témoignent les contributions de WEI et collab. [2012] ou de PERRIN et DEFAUX [2019] avec des indices de Sobol, celle de DERENNES et collab. [2019b] avec des indices de Borgonovo, celle de MARREL et CHABRIDON [2020] avec des mesures HSIC ou celle de IDRISSE et collab. [2021] avec des valeurs de Shapley.

La fin du présent chapitre coïncide avec la fin de l'état de l'art entrepris depuis le début du Chapitre 1. On a mis en place un cadre méthodologique, introduit plusieurs formalismes mathématiques et présenté les atouts et les limites des principaux algorithmes existants pour l'apprentissage d'une distribution multidimensionnelle (voir Chapitre 1), l'estimation d'une probabilité de défaillance (voir Chapitre 2) et l'analyse de sensibilité, éventuellement fiabiliste (voir Chapitre 3). Les deux prochains chapitres constituent les contributions de ce travail de thèse et constituent une étude de l'incertitude entourant la prédiction fiabiliste lorsque la connaissance de la distribution d'entrée se limite à la possession d'un N-échantillon  $\mathbf{X}_{\text{obs}}$ . Sous de nombreux aspects, notamment pour ce qui est de la gestion du budget de simulation, la construction d'une procédure d'ASF pour  $\mathbf{X}_{\text{obs}} \mapsto P_f(\mathbf{X}_{\text{obs}})$  peut s'inspirer des procédures d'ASF construites pour  $\Theta \mapsto P_f(\Theta)$ . Toutefois, de nombreuses autres difficultés apparaissent et leur traitement fait l'objet des Chapitres 4 et 5.

## Références

- ARCHER, G., A. SALTELLI et I. SOBOL. 1997, «Sensitivity measures, ANOVA-like techniques and the use of bootstrap», *Journal of Statistical Computation and Simulation*, vol. 58, n° 2, p. 99–120.
- BAUCELLS, M. et E. BORGONOVO. 2013, «Invariant probabilistic sensitivity analysis», *Management Science*, vol. 59, n° 11, p. 2536–2549.
- BECKMAN, R. J. et M. D. MCKAY. 1987, «Monte Carlo estimation under different distributions using the same simulation», *Technometrics*, vol. 29, n° 2, p. 153–160.
- BEER, M., F. A. DIAZ DE LA O, E. PATELLI et S. K. AU. 2014, «Conceptual comparison of Bayesian approaches and imprecise probabilities», *Computational Technology Reviews*, vol. 9, p. 1–29.
- BENOUMECHIARA, N. et K. ELIE-DIT-COSAQUE. 2019, «Shapley effects for sensitivity analysis with dependent inputs : bootstrap and kriging-based algorithms», *ESAIM : Proceedings and Surveys*, vol. 65, p. 266–293.
- BJERAGER, P. et S. KRENK. 1987, «Sensitivity measures in structural reliability analysis», dans *Reliability and optimization of structural systems*, Springer, p. 459–470.
- BJERAGER, P. et S. KRENK. 1989, «Parametric sensitivity in first order reliability theory», *Journal of Engineering Mechanics*, vol. 115, n° 7, p. 1577–1582.
- BORGONOVO, E. 2006, «Measuring uncertainty importance : investigation and comparison of alternative approaches», *Risk Analysis*, vol. 26, n° 5, p. 1349–1361.
- BORGONOVO, E. 2007, «A new uncertainty importance measure», *Reliability Engineering & System Safety*, vol. 92, n° 6, p. 771–784.
- BORGONOVO, E. et G. E. APOSTOLAKIS. 2001, «A new importance measure for risk-informed decision making», *Reliability Engineering & System Safety*, vol. 72, n° 2, p. 193–212.
- BORGONOVO, E., W. CASTAINGS et S. TARANTOLA. 2011, «Moment independent importance measures : new results and analytical test cases», *Risk Analysis*, vol. 31, n° 3, p. 404–428.

- BORGONOVO, E., G. B. HAZEN et E. PLISCHKE. 2016, «A common rationale for global sensitivity measures and their estimation», *Risk Analysis*, vol. 36, n° 10, p. 1871–1895.
- BORGONOVO, E. et B. IOOSS. 2016, «Moment independent and reliability-based importance measures», .
- BORGONOVO, E. et L. PECCATI. 2004, «Sensitivity analysis in investment project evaluation», *International Journal of Production Economics*, vol. 90, n° 1, p. 17–25.
- BORGONOVO, E., S. TARANTOLA, E. PLISCHKE et M. MORRIS. 2014, «Transformations and invariance in the sensitivity analysis of computer experiments», *Journal of the Royal Statistical Society : Series B : Statistical Methodology*, p. 925–947.
- BOURINET, J.-M. 2017, «FORM sensitivities to distribution parameters with the Nataf transformation», dans *Risk and Reliability Analysis : Theory and Applications*, Springer, p. 277–302.
- BREVAULT, L., M. BALESSENT et J. MORIO. 2020, «Uncertainty propagation and sensitivity analysis», dans *Aerospace System Analysis and Optimization in Uncertainty*, Springer, p. 69–117.
- CAMPOLONGO, F., J. CARIBONI et A. SALTELLI. 2007, «An effective screening design for sensitivity analysis of large models», *Environmental modelling & software*, vol. 22, n° 10, p. 1509–1518.
- CASTAINGS, W., E. BORGONOVO, M. MORRIS et S. TARANTOLA. 2012, «Sampling strategies in density-based sensitivity analysis», *Environmental Modelling & Software*, vol. 38, p. 13–26.
- CASTRO, J., D. GÓMEZ et J. TEJADA. 2009, «Polynomial calculation of the Shapley value based on sampling», *Computers & Operations Research*, vol. 36, n° 5, p. 1726–1730.
- CHABRIDON, V. 2018, *Reliability-oriented sensitivity analysis under probabilistic model uncertainty—Application to aerospace systems*, thèse de doctorat, Université Clermont Auvergne.
- CHABRIDON, V., M. BALESSENT, J.-M. BOURINET, J. MORIO et N. GAYTON. 2017, «Evaluation of failure probability under parameter epistemic uncertainty : application to aerospace system reliability assessment», *Aerospace Science and Technology*, vol. 69, p. 526–537.
- CHABRIDON, V., M. BALESSENT, J.-M. BOURINET, J. MORIO et N. GAYTON. 2018, «Reliability-based sensitivity estimators of rare event probability in the presence of distribution parameter uncertainty», *Reliability Engineering & System Safety*, vol. 178, p. 164–178.
- CHAMPION, M., G. CHASTAING, S. GADAT et C. PRIEUR. 2015, «L<sub>2</sub>-boosting for sensitivity analysis with dependent inputs», *Statistica sinica*, p. 1477–1502.
- CHASTAING, G., F. GAMBOA et C. PRIEUR. 2015, «Generalized Sobol sensitivity indices for dependent variables : numerical methods», *Journal of Statistical Computation and Simulation*, vol. 85, n° 7, p. 1306–1333.
- CHASTAING, G., F. GAMBOA, C. PRIEUR et collab.. 2012, «Generalized Hoeffding-Sobol decomposition for dependent variables : application to sensitivity analysis», *Electronic Journal of Statistics*, vol. 6, p. 2420–2448.
- CHATTERJEE, S. 2020, «A new coefficient of correlation», *Journal of the American Statistical Association*, p. 1–21.
- CHIU, C., W. YAN et K.-V. YUEN. 2012, «Reliability analysis of soil-water characteristics curve and its application to slope stability analysis», *Engineering Geology*, vol. 135, p. 83–91.
- CSISZÁR, I. 1967, «Information-type measures of difference of probability distributions and indirect observation», *Studia Scientiarum Mathematicarum Hungarica*, vol. 2, p. 229–318.

- CUKIER, R., C. FORTUIN, K. E. SHULER, A. PETSCHKE et J. SCHAIBLY. 1973, «Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory», *The Journal of chemical physics*, vol. 59, n° 8, p. 3873–3878.
- DA VEIGA, S. 2015, «Global sensitivity analysis with dependence measures», *Journal of Statistical Computation and Simulation*, vol. 85, n° 7, p. 1283–1305.
- DA VEIGA, S., F. WAHL et F. GAMBOA. 2009, «Local polynomial estimation for sensitivity analysis on models with correlated inputs», *Technometrics*, vol. 51, n° 4, p. 452–463.
- DER KIUREGHIAN, A. 1989, «Measures of structural safety under imperfect states of knowledge», *Journal of Structural Engineering*, vol. 115, n° 5, p. 1119–1140.
- DER KIUREGHIAN, A. 2008, «Analysis of structural reliability under parameter uncertainties», *Probabilistic engineering mechanics*, vol. 23, n° 4, p. 351–358.
- DERENNES, P. 2019, *Mesures de sensibilité de Borgonovo : estimation des indices d'ordre un et supérieur, et application à l'analyse de fiabilité*, thèse de doctorat, Université Paul Sabatier, Toulouse.
- DERENNES, P., J. MORIO et F. SIMATOS. 2018, «Estimation of moment independent importance measures using a copula and maximum entropy framework», dans *2018 Winter Simulation Conference (WSC)*, IEEE, p. 1623–1634.
- DERENNES, P., J. MORIO et F. SIMATOS. 2019a, «A nonparametric importance sampling estimator for moment independent importance measures», *Reliability Engineering & System Safety*, vol. 187, p. 3–16.
- DERENNES, P., J. MORIO et F. SIMATOS. 2019b, «Simultaneous estimation of complementary moment independent sensitivity measures for reliability analysis», *arXiv preprint arXiv :1911.02488*.
- DITLEVSEN, O. 1979, «Generalized second moment reliability index», *Journal of Structural Mechanics*, vol. 7, n° 4, p. 435–451.
- DITLEVSEN, O. et H. O. MADSEN. 1996, *Structural reliability methods*, vol. 178, Wiley New York.
- EHRE, M., I. PAPAIOANNOU et D. STRAUB. 2018, «Efficient estimation of variance-based reliability sensitivities in the presence of multi-uncertainty», dans *IFIP*.
- FAN, J. et I. GIJBELS. 1996, *Local polynomial modelling and its applications : monographs on statistics and applied probability 66*, vol. 66, CRC Press.
- FERSON, S. et W. T. TUCKER. 2006, «Sensitivity analysis using probability bounding», *Reliability Engineering & System Safety*, vol. 91, n° 10-11, p. 1435–1442.
- FORT, J.-C., T. KLEIN et A. LAGNOUX. 2020, «Global sensitivity analysis and Wasserstein spaces», *arXiv preprint arXiv :2007.12378*.
- FORT, J.-C., T. KLEIN et N. RACHDI. 2016, «New sensitivity analysis subordinated to a contrast», *Communications in Statistics-Theory and Methods*, vol. 45, n° 15, p. 4349–4364.
- GAMBOA, F., P. GREMAUD, T. KLEIN et A. LAGNOUX. 2020, «Global sensitivity analysis : a new generation of mighty estimators based on rank statistics», *arXiv preprint arXiv :2003.01772*.
- GAMBOA, F., A. JANON, T. KLEIN et A. LAGNOUX. 2013, «Sensitivity indices for multivariate outputs», *Comptes Rendus Mathématique*, vol. 351, n° 7-8, p. 307–310.
- GAMBOA, F., A. JANON, T. KLEIN, A. LAGNOUX et C. PRIEUR. 2016, «Statistical inference for Sobol pick-freeze Monte Carlo method», *Statistics*, vol. 50, n° 4, p. 881–902.

- GAMBOA, F., A. JANON, T. KLEIN, A. LAGNOUX et collab.. 2014, «Sensitivity analysis for multidimensional and functional outputs», *Electronic Journal of Statistics*, vol. 8, n° 1, p. 575–603.
- GAMBOA, F., T. KLEIN et A. LAGNOUX. 2018, «Sensitivity analysis based on Cramér-Von Mises distance», *SIAM/ASA Journal on Uncertainty Quantification*, vol. 6, n° 2, p. 522–548.
- GELMAN, A., J. B. CARLIN, H. S. STERN, D. B. DUNSON, A. VEHTARI et D. B. RUBIN. 2013, *Bayesian data analysis*, CRC press.
- GILQUIN, L., C. PRIEUR et E. ARNAUD. 2015, «Replication procedure for grouped Sobol'indices estimation in dependent uncertainty spaces», *Information and Inference : A Journal of the IMA*, vol. 4, n° 4, p. 354–379.
- GRETTON, A., O. BOUSQUET, A. SMOLA et B. SCHÖLKOPF. 2005, «Measuring statistical dependence with Hilbert-Schmidt norms», dans *International conference on algorithmic learning theory*, Springer, p. 63–77.
- HART, J. L., A. ALEXANDERIAN et P. A. GREMAUD. 2017, «Efficient computation of Sobol'indices for stochastic models», *SIAM Journal on Scientific Computing*, vol. 39, n° 4, p. A1514–A1530.
- HELTON, J. C. 1993, «Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal», *Reliability Engineering & System Safety*, vol. 42, n° 2-3, p. 327–367.
- HELTON, J. C., J. D. JOHNSON, W. OBERKAMPF et C. J. SALLABERRY. 2006a, «Sensitivity analysis in conjunction with evidence theory representations of epistemic uncertainty», *Reliability Engineering & System Safety*, vol. 91, n° 10-11, p. 1414–1434.
- HELTON, J. C., J. D. JOHNSON, C. J. SALLABERRY et C. B. STORLIE. 2006b, «Survey of sampling-based methods for uncertainty and sensitivity analysis», *Reliability Engineering & System Safety*, vol. 91, n° 10-11, p. 1175–1209.
- HESTERBERG, T. C. 1996, «Estimates and confidence intervals for importance sampling sensitivity analysis», *Mathematical and computer modelling*, vol. 23, n° 8-9, p. 79–85.
- HOEFFDING, W. 1948, «A class of statistics with asymptotically normal distribution», *Annals of Mathematical Statistics*, vol. 19, p. 273–325.
- HOHENBICHLER, M. et R. RACKWITZ. 1986, «Sensitivity and importance measures in structural reliability», *Civil engineering systems*, vol. 3, n° 4, p. 203–209.
- HOMMA, T. et A. SALTELLI. 1996, «Importance measures in global sensitivity analysis of nonlinear models», *Reliability Engineering & System Safety*, vol. 52, n° 1, p. 1–17.
- HONG, H. P. 1996, «Evaluation of the probability of failure with uncertain distribution parameters», *Civil Engineering Systems*, vol. 13, n° 2, p. 157–168.
- HOOKE, G. 2007, «Generalized functional ANOVA diagnostics for high-dimensional functions of dependent variables», *Journal of Computational and Graphical Statistics*, vol. 16, n° 3, p. 709–732.
- HORA, S. et R. IMAN. 1986, «A comparison of maximum/bounding and the Bayes/Monte Carlo for fault tree uncertainty analysis», cahier de recherche, Hawaii Univ., Hilo (USA) ; Sandia National Labs., Albuquerque, NM (USA).
- IDRISSI, M. I., V. CHABRIDON et B. IOOSS. 2021, «Developments and applications of Shapley effects to reliability-oriented sensitivity analysis with correlated inputs», *arXiv preprint arXiv :2101.08083*.

- IMAN, R. L. et S. C. HORA. 1990, «A robust measure of uncertainty importance for use in fault tree system analysis», *Risk analysis*, vol. 10, n° 3, p. 401–406.
- IOOSS, B. et C. PRIEUR. 2019, «Shapley effects for sensitivity analysis with correlated inputs : comparisons with Sobol'indices, numerical estimation and applications», *International Journal for Uncertainty Quantification*, vol. 9, n° 5.
- IOOSS, B. et M. RIBATET. 2009, «Global sensitivity analysis of computer models with functional inputs», *Reliability Engineering & System Safety*, vol. 94, n° 7, p. 1194–1204.
- IOOSS, B. et A. SALTELLI. 2017, «Introduction to sensitivity analysis», *Handbook of uncertainty quantification*, p. 1103–1122.
- IOTT, J., R. T. HAFTKA et H. M. ADELMAN. 1985, *Selecting step sizes in sensitivity analysis by finite differences*, vol. 86382, National Aeronautics and Space Administration, Scientific and Technical Information Branch.
- JACQUES, J., C. LAVERGNE et N. DEVICTOR. 2006, «Sensitivity analysis in presence of model uncertainty and correlated inputs», *Reliability Engineering & System Safety*, vol. 91, n° 10-11, p. 1126–1134.
- JANON, A., T. KLEIN, A. LAGNOUX, M. NODET et C. PRIEUR. 2014, «Asymptotic normality and efficiency of two Sobol index estimators», *ESAIM : Probability and Statistics*, vol. 18, p. 342–364.
- JANSEN, M. J. 1999, «Analysis of variance designs for model output», *Computer Physics Communications*, vol. 117, n° 1-2, p. 35–43.
- JANSEN, M. J., W. A. ROSSING et R. A. DAAMEN. 1994, «Monte Carlo estimation of uncertainty contributions from several independent multivariate sources», dans *Predictability and nonlinear modelling in natural sciences and economics*, Springer, p. 334–343.
- KARAMCHANDANI, A. 1990, *New approaches to structural system reliability*, thèse de doctorat, Dept. of Civil Engineering, Stanford University.
- KUCHERENKO, S. et B. IOOSS. 2017, «Derivative-based global sensitivity measures», dans *Handbook of Uncertainty Quantification*, p. 1241–1263.
- KUCHERENKO, S., M. RODRIGUEZ-FERNANDEZ, C. PANTELIDES et N. SHAH. 2009, «Monte Carlo evaluation of derivative-based global sensitivity measures», *Reliability Engineering & System Safety*, vol. 94, n° 7, p. 1135–1148.
- KUCHERENKO, S., S. TARANTOLA et P. ANNONI. 2012, «Estimation of global sensitivity indices for models with dependent variables», *Computer physics communications*, vol. 183, n° 4, p. 937–946.
- KUCHERENKO, S. et collab.. 2010, «A new derivative based importance criterion for groups of variables and its link with the global sensitivity indices», *Computer Physics Communications*, vol. 181, n° 7, p. 1212–1217.
- LAMBONI, M., B. IOOSS, A.-L. POPELIN et F. GAMBOA. 2013, «Derivative-based global sensitivity measures : general links with Sobol'indices and numerical tests», *Mathematics and Computers in Simulation*, vol. 87, p. 45–54.
- LANCASTER, P. et K. SALKASKAS. 1981, «Surfaces generated by moving least squares methods», *Mathematics of computation*, vol. 37, n° 155, p. 141–158.
- LE GRATIET, L., S. MARELLI et B. SUDRET. 2017, «Metamodel-based sensitivity analysis : polynomial chaos expansions and Gaussian processes», dans *Handbook of Uncertainty Quantification*, Springer International Publishing.

- LEE, I., K. K. CHOI, Y. NOH, L. ZHAO et D. GORSICH. 2011, «Sampling-based stochastic sensitivity analysis using score functions for RBDO problems with correlated random variables», *Journal of Mechanical Design*, vol. 133, n° 2.
- LEMAÎTRE, P. 2014, *Analyse de sensibilité en fiabilité des structures*, thèse de doctorat, Université de Bordeaux.
- LEMAÎTRE, P., E. SERGIENKO, A. ARNAUD, N. BOUSQUET, F. GAMBOA et B. IOOSS. 2015, «Density modification-based reliability sensitivity analysis», *Journal of Statistical Computation and Simulation*, vol. 85, n° 6, p. 1200–1223.
- LI, G., H. RABITZ, P. E. YELVINGTON, O. O. OLUWOLE, F. BACON, C. E. KOLB et J. SCHOENDORF. 2010, «Global sensitivity analysis for systems with independent and/or correlated inputs», *The journal of physical chemistry A*, vol. 114, n° 19, p. 6022–6032.
- LI, L., Z. LU et C. ZHOU. 2011, «Importance analysis for models with correlated input variables by the state dependent parameters method», *Computers & Mathematics with Applications*, vol. 62, n° 12, p. 4547–4556.
- LIU, Q. et T. HOMMA. 2009, «A new computational method of a moment-independent uncertainty importance measure», *Reliability Engineering & System Safety*, vol. 94, n° 7, p. 1205–1211.
- LUO, X., Z. LU et X. XU. 2014, «A fast computational method for moment-independent uncertainty importance measure», *Computer Physics Communications*, vol. 185, n° 1, p. 19–27.
- MARA, T. A. 2009, «Extension of the RBD-FAST method to the computation of global sensitivity indices», *Reliability Engineering & System Safety*, vol. 94, n° 8, p. 1274–1281.
- MARA, T. A. et S. TARANTOLA. 2012, «Variance-based sensitivity indices for models with dependent inputs», *Reliability Engineering & System Safety*, vol. 107, p. 115–121.
- MARA, T. A., S. TARANTOLA et P. ANNONI. 2015, «Non-parametric methods for global sensitivity analysis of model output with dependent inputs», *Environmental modelling & software*, vol. 72, p. 173–183.
- MARREL, A. et V. CHABRIDON. 2020, «Statistical developments for target and conditional sensitivity analysis : application on safety studies for nuclear reactor», .
- MARREL, A., B. IOOSS, S. DA VEIGA et M. RIBATET. 2012, «Global sensitivity analysis of stochastic computer models with joint metamodels», *Statistics and Computing*, vol. 22, n° 3, p. 833–847.
- MARREL, A., B. IOOSS, F. VAN DORPE et E. VOLKOVA. 2008, «An efficient methodology for modeling complex computer codes with Gaussian processes», *Computational Statistics & Data Analysis*, vol. 52, n° 10, p. 4731–4744.
- MCKAY, M. 1996, «Variance-based methods for assessing uncertainty importance in nureg-1150 analyses», *Los Alamos National Laboratory LA-UR-96-2695*.
- MCKAY, M. D. 1997, «Nonparametric variance-based methods of assessing uncertainty importance», *Reliability engineering & system safety*, vol. 57, n° 3, p. 267–279.
- MILLWATER, H. 2009, «Universal properties of kernel functions for probabilistic sensitivity analysis», *Probabilistic Engineering Mechanics*, vol. 24, n° 1, p. 89–99.
- MORIO, J. 2011, «Influence of input PDF parameters of a model on a failure probability estimation», *Simulation Modelling Practice and Theory*, vol. 19, n° 10, p. 2244–2255.
- MORRIS, M. D. 1991, «Factorial sampling plans for preliminary computational experiments», *Technometrics*, vol. 33, n° 2, p. 161–174.

- NANTY, S., C. HELBERT, A. MARREL, N. PÉROT et C. PRIEUR. 2016, «Sampling, metamodeling, and sensitivity analysis of numerical simulators with functional stochastic inputs», *SIAM/ASA Journal on Uncertainty Quantification*, vol. 4, n° 1, p. 636–659.
- OWEN, A. B. 2013, «Better estimation of small Sobol'sensitivity indices», *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, vol. 23, n° 2, p. 1–17.
- OWEN, A. B. 2014, «Sobol'indices and Shapley value», *SIAM/ASA Journal on Uncertainty Quantification*, vol. 2, n° 1, p. 245–251.
- OWEN, A. B. et C. PRIEUR. 2017, «On Shapley value for measuring importance of dependent inputs», *SIAM/ASA Journal on Uncertainty Quantification*, vol. 5, n° 1, p. 986–1002.
- PENDOLA, M. 2000, *Fiabilité des structures en contexte d'incertitudes statistiques et d'écarts de modélisation*, thèse de doctorat, Clermont-Ferrand 2.
- PERRIN, G. et G. DEFAUX. 2019, «Efficient evaluation of reliability-oriented sensitivity indices», *Journal of Scientific Computing*, vol. 79, n° 3, p. 1433–1455.
- PLISCHKE, E., E. BORGONOVO et C. L. SMITH. 2013, «Global sensitivity measures from given data», *European Journal of Operational Research*, vol. 226, n° 3, p. 536–550.
- PLISCHKE, E., G. RABITTI et E. BORGONOVO. 2020, «Computing Shapley effects for sensitivity analysis», *arXiv preprint arXiv :2002.12024*.
- POINCARÉ, H. 1890, «Sur les équations aux dérivées partielles de la physique mathématique», *American Journal of Mathematics*, p. 211–294.
- RABITZ, H., Ö. F. ALIŞ, J. SHORTER et K. SHIM. 1999, «Efficient input-output model representations», *Computer physics communications*, vol. 117, n° 1-2, p. 11–20.
- RAHMAN, S. 2009, «Stochastic sensitivity analysis by dimensional decomposition and score functions», *Probabilistic Engineering Mechanics*, vol. 24, n° 3, p. 278–287.
- RAHMAN, S. 2016, «The f-sensitivity index», *SIAM/ASA Journal on Uncertainty Quantification*, vol. 4, n° 1, p. 130–162.
- RATTO, M., S. TARANTOLA et A. SALTELLI. 2001, «Estimation of importance indicators for correlated inputs», *Proceedings of ESREL2001*, p. 157–164.
- ROBERT, C. 2007, *The Bayesian choice : from decision-theoretic foundations to computational implementation*, Springer Science & Business Media.
- ROUSTANT, O., J. FRUTH, B. IOOSS et S. KUHN. 2014, «Crossed-derivative based sensitivity measures for interaction screening», *Mathematics and Computers in Simulation*, vol. 105, p. 105–118.
- SALTELLI, A. 2002, «Making best use of model evaluations to compute sensitivity indices», *Computer physics communications*, vol. 145, n° 2, p. 280–297.
- SALTELLI, A., M. RATTO et S. TARANTOLA. 2001, «Model-free importance indicators for dependent input», dans *SAMO 2001 : third international symposium on sensitivity analysis of model output (Madrid, 18-20 June 2001)*, p. 21–25.
- SALTELLI, A. et S. TARANTOLA. 2002, «On the relative importance of input factors in mathematical models : safety assessment for nuclear waste disposal», *Journal of the American Statistical Association*, vol. 97, n° 459, p. 702–709.

- SALTELLI, A., S. TARANTOLA, F. CAMPOLONGO et M. RATTO. 2004, *Sensitivity analysis in practice : a guide to assessing scientific models*, vol. 1, Wiley Online Library.
- SALTELLI, A., S. TARANTOLA et K.-S. CHAN. 1999, «A quantitative model-independent method for global sensitivity analysis of model output», *Technometrics*, vol. 41, n° 1, p. 39–56.
- SARAZIN, G., P. DERENNES et J. MORIO. 2020, «Estimation of high-order moment-independent importance measures for Shapley value analysis», *Applied Mathematical Modelling*, vol. 88, p. 396–417.
- SCHÖBI, R. et B. SUDRET. 2019, «Global sensitivity analysis in the context of imprecise probabilities (p-boxes) using sparse polynomial chaos expansions», *Reliability Engineering & System Safety*, vol. 187, p. 129–141.
- SHAPLEY, L. S. 1953, «A value for  $n$ -person games», *Contributions to the Theory of Games*, vol. 2, n° 28, p. 307–317.
- SOBOL, I. et A. GRESHAM. 1995, «On an alternative global sensitivity estimators», *Proceedings of SAMO*, p. 40–42.
- SOBOL, I. et S. KUCHERENKO. 2010, «Derivative based global sensitivity measures», *Procedia-Social and Behavioral Sciences*, vol. 2, n° 6, p. 7745–7746.
- SOBOL, I. M. 1993, «Sensitivity analysis for non-linear mathematical models», *Mathematical modelling and computational experiment*, vol. 1, p. 407–414.
- SOBOL, I. M. et E. MYSHETSKAYA. 2008, «Monte Carlo estimators for small sensitivity indices.», *Monte Carlo Methods & Applications*, vol. 13.
- SONG, E., B. L. NELSON et J. STAUM. 2016, «Shapley effects for global sensitivity analysis : Theory and computation», *SIAM/ASA Journal on Uncertainty Quantification*, vol. 4, n° 1, p. 1060–1083.
- SONG, J., P. WEI, M. VALDEBENITO, S. BI, M. BROGGI, M. BEER et Z. LEI. 2019, «Generalization of non-intrusive imprecise stochastic simulation for mixed uncertain variables», *Mechanical Systems and Signal Processing*, vol. 134, p. 106316.
- SONG, S., Z. LU et H. QIAO. 2009, «Subset simulation for structural reliability sensitivity analysis», *Reliability Engineering & System Safety*, vol. 94, n° 2, p. 658–665.
- SUDRET, B. 2008, «Global sensitivity analysis using polynomial chaos expansions», *Reliability engineering & system safety*, vol. 93, n° 7, p. 964–979.
- SUDRET, B. 2015, «Polynomial chaos expansions and stochastic finite element methods», *Risk and reliability in geotechnical engineering*, p. 265–300.
- SUDRET, B. et Y. CANIOU. 2013, *Analysis of covariance (ANCOVA) using polynomial chaos expansions*, CRC Press/Balkema.
- SZÉKELY, G. J., M. L. RIZZO, N. K. BAKIROV et collab.. 2007, «Measuring and testing dependence by correlation of distances», *The annals of statistics*, vol. 35, n° 6, p. 2769–2794.
- TARANTOLA, S., D. GATELLI, S. KUCHERENKO, W. MAUNTZ et collab.. 2007, «Estimating the approximation error when fixing unessential factors in global sensitivity analysis», *Reliability Engineering & System Safety*, vol. 92, n° 7, p. 957–960.
- TARANTOLA, S., D. GATELLI et T. A. MARA. 2006, «Random balance designs for the estimation of first order global sensitivity indices», *Reliability Engineering & System Safety*, vol. 91, n° 6, p. 717–727.

- TISSOT, J.-Y. et C. PRIEUR. 2012, «Bias correction for the estimation of sensitivity indices based on random balance designs», *Reliability Engineering & System Safety*, vol. 107, p. 205–213.
- WALLEY, P. 1991, «Statistical reasoning with imprecise probabilities», .
- WALLEY, P. 2000, «Towards a unified theory of imprecise probability», *International Journal of Approximate Reasoning*, vol. 24, n° 2-3, p. 125–148.
- WANG, P., Z. LU et Z. TANG. 2013a, «An application of the Kriging method in global sensitivity analysis with parameter uncertainty», *Applied Mathematical Modelling*, vol. 37, n° 9, p. 6543–6555.
- WANG, P., Z. LU et Z. TANG. 2013b, «A derivative based sensitivity measure of failure probability in the presence of epistemic and aleatory uncertainties», *Computers & Mathematics with Applications*, vol. 65, n° 1, p. 89–101.
- WANG, P., Z. LU et Z. TANG. 2013c, «Importance measure analysis with epistemic uncertainty and its moving least squares solution», *Computers & Mathematics with Applications*, vol. 66, n° 4, p. 460–471.
- WEI, P., Z. LU, W. HAO, J. FENG et B. WANG. 2012, «Efficient sampling methods for global reliability sensitivity analysis», *Computer Physics Communications*, vol. 183, n° 8, p. 1728–1743.
- WEI, P., Z. LU et J. SONG. 2014, «Moment-independent sensitivity analysis using copula», *Risk Analysis*, vol. 34, n° 2, p. 210–222.
- WEI, P., Z. LU et X. YUAN. 2013, «Monte carlo simulation for moment-independent sensitivity analysis», *Reliability Engineering & System Safety*, vol. 110, p. 60–67.
- WEI, P., J. SONG, S. BI, M. BROGGI, M. BEER, Z. LU et Z. YUE. 2019a, «Non-intrusive stochastic analysis with parameterized imprecise probability models : 1. Performance estimation», *Mechanical Systems and Signal Processing*, vol. 124, p. 349–368.
- WEI, P., J. SONG, S. BI, M. BROGGI, M. BEER, Z. LU et Z. YUE. 2019b, «Non-intrusive stochastic analysis with parameterized imprecise probability models : 2. Reliability and rare events analysis», *Mechanical Systems and Signal Processing*, vol. 126, p. 227–247.
- WOODS, D. et S. LEWIS. 2017, «Design of experiments for screening», dans *Handbook of Uncertainty Quantification*, p. 1143–1185.
- WU, Y.-T. 1994a, «Adaptive importance sampling (AIS)-based system reliability sensitivity analysis method», dans *Probabilistic Structural Mechanics : Advances in Structural Reliability Methods*, Springer, p. 550–564.
- WU, Y.-T. 1994b, «Computational methods for efficient structural reliability and reliability sensitivity analysis», *AIAA journal*, vol. 32, n° 8, p. 1717–1723.
- XIAO, S., Z. LU et L. XU. 2016, «A new effective screening design for structural sensitivity analysis of failure probability with the epistemic uncertainty», *Reliability Engineering & System Safety*, vol. 156, p. 1–14.
- XU, C. et G. GERTNER. 2007, «Extending a global sensitivity analysis technique to models with correlated parameters», *Computational Statistics & Data Analysis*, vol. 51, n° 12, p. 5579–5590.
- XU, C. et G. Z. GERTNER. 2008, «Uncertainty and sensitivity analysis for models with correlated parameters», *Reliability Engineering & System Safety*, vol. 93, n° 10, p. 1563–1573.

- YUN, W., Z. LU et X. JIANG. 2019, «An efficient method for moment-independent global sensitivity analysis by dimensional reduction technique and principle of maximum entropy», *Reliability Engineering & System Safety*, vol. 187, p. 174–182.
- ŽANIĆ, V. et K. ŽIHA. 1998, «Sensitivity to correlation in multivariate models», *Computer assisted mechanics and engineering sciences*, vol. 5, p. 75.
- ŽANIĆ, V. et K. ŽIHA. 2001, «Sensitivity to correlation in structural reliability problems», *Transaction of FAMENA*, vol. 25, n° 2, p. 1.
- ZHANG, F., Z. LU, L. CUI et S. SONG. 2010, «Reliability sensitivity algorithm based on stratified importance sampling method for multiple failure modes systems», *Chinese Journal of Aeronautics*, vol. 23, n° 6, p. 660–669.
- ZHANG, J., L. ZHANG et W. H. TANG. 2009, «Bayesian framework for characterizing geotechnical model uncertainty», *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 135, n° 7, p. 932–940.
- ZHANG, X. et M. D. PANDEY. 2013, «Structural reliability analysis based on the concepts of entropy, fractional moment and dimensional reduction method», *Structural Safety*, vol. 43, p. 28–40.
- ZHOU, Y., Z. LU, Y. SHI et K. CHENG. 2019, «A vine copula-based method for analyzing the moment-independent importance measure of the multivariate output», *Proceedings of the Institution of Mechanical Engineers, Part O : Journal of Risk and Reliability*, vol. 233, n° 3, p. 338–354.
- ZHU, X. et B. SUDRET. 2020, «Global sensitivity analysis for stochastic simulators based on generalized lambda surrogate models», *arXiv preprint arXiv :2005.01309*.

## Chapitre 4

# Analyse de la variabilité d'un risque de défaillance suite à l'apprentissage de la distribution conjointe des entrées

### Sommaire

---

<b>4.1 Introduction</b> . . . . .	<b>140</b>
<b>4.2 Développement d'une procédure d'ASF dans un cadre simplifié</b> . . . . .	<b>141</b>
4.2.1 Description du problème d'ASF . . . . .	142
4.2.2 Mise en place d'un mécanisme de ré-échantillonnage . . . . .	145
4.2.3 Implémentation de l'algorithme d'ASF . . . . .	153
4.2.4 Adaptation du mécanisme de ré-échantillonnage pour conduire une ASF après assimilation de données supplémentaires . . . . .	157
<b>4.3 Validation de la procédure d'ASF</b> . . . . .	<b>163</b>
4.3.1 Étude d'un cas test linéaire . . . . .	163
4.3.2 Étude d'un cas test non-linéaire . . . . .	176
<b>4.4 Renforcement de la procédure d'ASF pour le cadre général</b> . . . . .	<b>183</b>
4.4.1 Étude des difficultés introduites par ce nouveau cadre . . . . .	183
4.4.2 Utilisation de l'échantillonnage préférentiel dans un contexte d'ASF . . . . .	185
4.4.3 Mise en place d'une stratégie de diminution de la charge de calcul . . . . .	186
<b>4.5 Validation de la procédure d'ASF renforcée</b> . . . . .	<b>191</b>
4.5.1 Retour sur le cas test linéaire . . . . .	191
4.5.2 Étude d'un cas test non-linéaire . . . . .	193
<b>4.6 Application au flambage d'une plaque composite</b> . . . . .	<b>201</b>
4.6.1 Description du cas test . . . . .	201
4.6.2 Description du problème d'analyse de fiabilité . . . . .	204
4.6.3 Implémentation de la méthodologie développée . . . . .	210
<b>4.7 Conclusion</b> . . . . .	<b>212</b>

---

## 4.1 Introduction

Dans le Chapitre 3, on a présenté les méthodes d'analyse de sensibilité fiabiliste (ASF) locales et globales existant dans la littérature. Il s'agissait d'étudier et de quantifier la sensibilité de l'estimation d'une probabilité de défaillance à l'incertitude portée par les paramètres de la distribution d'entrée. Dans la plupart des travaux consultés, ce second niveau d'incertitude est décrit par une loi *a priori*  $f_{\Theta}$  qui résume toute l'information sur le comportement aléatoire du vecteur  $\Theta$  des paramètres incertains d'une famille  $\mathcal{F}$  fixée au préalable.

$$\Theta \sim \mu_{\Theta} \longrightarrow f_{\mathbf{X}}(\cdot | \Theta, \mathcal{F}) \longrightarrow P_f(\Theta). \quad (4.1)$$

Il a pu être constaté que ces algorithmes d'ASF ne prévoient pas que la densité d'entrée  $f_{\mathbf{X}}(\cdot | \Theta, \mathcal{F})$  incorpore une copule. Cela s'explique par le fait qu'il est difficile d'obtenir des connaissances *a priori* sur le modèle et les paramètres de la copule. Ici, afin de maîtriser pleinement la prédiction du risque de défaillance, on aimerait être capable de mesurer l'impact d'une caractérisation incertaine de la copule. La Section 1.5.4 a montré que les modèles de copules adaptés à la grande dimension reposent sur une spécification agile, résultant du mariage de la théorie des graphes et des copules bidimensionnelles. Dans le cas des copules R-vignes, en supposant les lois marginales parfaitement connues, la chaîne de propagation (4.1) deviendrait :

$$(\mathbf{T}_{vg}, \mathbf{F}_{vg}, \Theta_{vg}) \sim \mu_{vg} \longrightarrow c_{vg}(\cdot | \mathbf{T}_{vg}, \mathbf{F}_{vg}, \Theta_{vg}) \longrightarrow P_f(\mathbf{T}_{vg}, \mathbf{F}_{vg}, \Theta_{vg}). \quad (4.2)$$

Cependant, il est difficilement envisageable d'acquérir une connaissance *a priori* de la mesure  $\mu_{vg}$  gouvernant le comportement conjoint de la séquence d'arbres  $\mathbf{T}_{vg}$ , de l'assortiment de familles  $\mathbf{F}_{vg}$  et des paramètres  $\Theta_{vg}$ . MIN et CZADO [2010, 2011] puis GRUBER et collab. [2015] ont apporté certains éléments de réponse mais ils ne sont pas suffisants pour mettre en place une démarche bayésienne globale. Par ailleurs, la séparation des incertitudes au sein du triplet  $(\mathbf{T}_{vg}, \mathbf{F}_{vg}, \Theta_{vg})$  est une question excessivement complexe. Pour ces raisons, on choisit de se détourner de l'approche bayésienne qui ne semble pas encore assez mature.

L'idée est plutôt de travailler directement à partir des données disponibles. Elles sont rassemblées dans un échantillon d'apprentissage  $\mathbf{X}_{obs}$  représenté comme une matrice aléatoire de  $\mathbb{R}^{N \times d}$ . La probabilité de défaillance s'exprime alors comme une fonction de l'échantillon  $\mathbf{X}_{obs}$  :

$$\mathbf{X}_{obs} \sim (\mu_{\mathbf{X}})^{\otimes N} \longrightarrow \hat{f}_{\mathbf{X}}(\cdot | \mathbf{X}_{obs}) \longrightarrow P_f(\mathbf{X}_{obs}). \quad (4.3)$$

L'amplitude de la variabilité de  $P_f(\mathbf{X}_{obs})$  est d'autant plus grande que le nombre  $N$  d'observations dans  $\mathbf{X}_{obs}$  est petit. Le "*small data*", de par la variabilité qu'il engendre sur la prédiction du risque, peut être vu comme la source d'incertitude motivant ce nouveau schéma d'ASF. On se place dans le cadre d'un apprentissage de la distribution d'entrée via une disjonction de la distribution de copule et des distributions marginales. Toutes les techniques (paramétriques ou non-paramétriques) d'apprentissage présentées au Chapitre 1 peuvent donc être envisagées. Sachant qu'on ne dispose que d'une seule réalisation  $\mathbf{x}_{obs}$  de la matrice aléatoire  $\mathbf{X}_{obs}$ , cette variabilité d'échantillonnage n'est en fait pas observable. Dès lors, appréhender la distribution de la quantité d'intérêt  $P_f(\mathbf{X}_{obs})$  est une première difficulté.

### Problème 4.1: Évaluation de la variabilité d'échantillonnage

Comment mesurer la variabilité affectant  $P_f(\mathbf{X}_{obs})$  ?

L'objectif est de reconstituer à partir de  $\mathbf{x}_{obs}$  la chaîne de propagation (4.3) et notamment d'être capable de simuler de nouvelles observations de la variable  $P_f(\mathbf{X}_{obs})$ . Dans ce contexte, la densité estimée  $\hat{f}_{\mathbf{X}}$  doit être vue comme une fonction aléatoire. Ses densités marginales  $\hat{f}_{X_j}$  et sa densité de copule  $\hat{c}_{\mathbf{U}}$  sont sujettes à une variabilité d'apprentissage. On peut alors se demander quelle est

l'entité fonctionnelle qui véhicule le plus d'incertitude sur  $P_f(\mathbf{X}_{\text{obs}})$  au cours de son apprentissage.

**Problème 4.2: Décomposition de la variabilité d'échantillonnage**

Quelle partie de la distribution conjointe engendre le plus de variabilité sur  $P_f(\mathbf{X}_{\text{obs}})$  ?

Il faut mettre en place une procédure d'ASF dans laquelle les composantes fonctionnelles de la décomposition copule-marginales jouent le rôle des sources d'incertitude. Disposer d'une réponse au Problème 4.2 peut être utile en pratique si l'on cherche à affiner le niveau de connaissance sur la probabilité de défaillance théorique  $P_f$ . Imaginons que l'on ait identifié la densité marginale  $\hat{f}_{X_1}$  comme l'entité la plus influente. Une stratégie pour réduire la variabilité de  $P_f(\mathbf{X}_{\text{obs}})$  peut consister à améliorer l'état de connaissance de  $f_{X_1}$ , par exemple en réunissant des observations supplémentaires de la variable  $X_1$ . Reste à savoir quel volume de nouvelles données est nécessaire à une telle action et quelles sont les modalités d'intégration à la distribution conjointe déjà estimée.

**Problème 4.3: Réduction de la variabilité d'échantillonnage**

L'acquisition ciblée de nouvelles données permet-elle de réduire la variabilité sur  $P_f(\mathbf{X}_{\text{obs}})$  ?

Tout au long de ce chapitre, et ce sera également le cas au Chapitre 5, on utilise les notations suivantes pour alléger la syntaxe et mieux conceptualiser les aspects algorithmiques du discours :

- $\mathcal{L} : \mathbf{x}_{\text{obs}} \mapsto \hat{f}_{\mathbf{X}}(\cdot | \mathbf{x}_{\text{obs}}) \implies$  Algorithme d'apprentissage  $\implies$  Chapitre 1
- $\mathcal{R} : f_{\mathbf{X}} \mapsto P_f \implies$  Algorithme de fiabilité déterministe  $\implies$  Section 2.2
- $\mathcal{R}_w : f_{\mathbf{X}} \mapsto \hat{P}_f \implies$  Algorithme de fiabilité stochastique  $\implies$  Section 2.3

Le présent chapitre vise à apporter des réponses aux trois problèmes énoncés ci-dessus et il est organisé de la manière suivante. La Section 4.2 présente les difficultés conceptuelles inhérentes à l'identification de la composante fonctionnelle de plus grande influence sur la probabilité de défaillance. Une procédure d'ASF est proposée pour répondre au problème. Pour des questions de simplicité, cette première section n'étudie que le cas d'un algorithme de fiabilité déterministe  $\mathcal{R}$ . La variabilité sur  $P_f(\mathbf{X}_{\text{obs}})$  est alors entièrement expliquée par la variabilité d'échantillonnage initiée par  $\mathbf{X}_{\text{obs}}$ . Dans la Section 4.3, l'algorithme d'ASF est expérimenté sur différents cas tests analytiques pour justifier la pertinence des indices proposés et la robustesse de leur schéma d'estimation. Dans la Section 4.4, on explique comment adapter la procédure d'ASF aux situations où il est obligatoire d'utiliser un algorithme de fiabilité stochastique  $\mathcal{R}_w$ . Les améliorations apportées à l'algorithme initial sont validées dans la Section 4.5 sur des exemples analytiques et dans la Section 4.6 sur un cas test physique issu d'une situation métier.

## 4.2 Développement d'une procédure d'ASF dans un cadre simplifié

Dans cette section, on émet l'hypothèse qu'il est possible de calculer avec une formule analytique la probabilité de défaillance  $P_f(\mathbf{X}_{\text{obs}})$  associée à une densité apprise  $\hat{f}_{\mathbf{X}}(\cdot | \mathbf{X}_{\text{obs}})$ . De plus, il faut supposer qu'un appel à la fonction de performance  $\phi$  est gratuit. Ces hypothèses ne sont évidemment pas réalistes mais elles permettent de se placer dans un cadre simplifié facilitant le traitement des problèmes évoqués. Pour la phase de validation, les cas tests sont construits pour que l'algorithme d'apprentissage statistique  $\mathcal{L}$  et la fonction de performance  $\phi$  soient compatibles avec une évaluation exacte du risque de défaillance. La gestion de la charge de calcul induite par  $\phi$  ainsi que la maîtrise de la stochasticité de  $\mathcal{R}_w$  seront traitées plus tard.

## 4.2.1 Description du problème d'ASF

### 4.2.1.1 État de connaissance basé sur un échantillon

On rappelle que la matière première se limite à une unique réalisation  $\mathbf{x}_{\text{obs}}$  de  $\mathbf{X}_{\text{obs}}$ . On reprend la plupart des notations introduites dans le Chapitre 1. En particulier, le vecteur  $\mathbf{U}$  est obtenu en posant  $U_j := F_{X_j}(X_j)$ , c'est-à-dire en renormalisant toutes les lois marginales de  $\mathbf{X}$  pour qu'elles soient uniformes. La distribution de  $\mathbf{U}$  est la copule de  $\mathbf{X}$ . Cette loi  $\mu_{\mathbf{U}}$  est à support dans l'hypercube  $\mathcal{H}_d := [0, 1]^d$ . De plus, lorsque la renormalisation implique des estimations  $\hat{F}_{X_j}$  des fonctions de répartition marginales, on adopte la notation  $\hat{\mathbf{U}}$ . En ce qui concerne la notation des différents échantillons manipulés, le formalisme suivant est employé dans le reste du chapitre :

- $\mathbf{X}_{\text{obs}} = \left[ X_j^{(i)} \right]_{\substack{1 \leq i \leq N \\ 1 \leq j \leq d}} \sim (\mu_{\mathbf{X}})^{\otimes N} \implies$  La matrice des  $N$  observations du vecteur  $\mathbf{X}$ .
- $\mathbf{X}_{\text{obs}}^{[\bullet j]} = \left[ X_j^{(i)} \right]_{1 \leq i \leq N} \sim (\mu_{X_j})^{\otimes N} \implies$  Le vecteur des  $N$  observations de la variable  $X_j$ .
- $\mathbf{X}_{\text{obs}}^{[i \bullet]} = \left[ X_j^{(i)} \right]_{1 \leq j \leq d} \sim \mu_{\mathbf{X}} \implies$  La  $i$ -ème observation du vecteur  $\mathbf{X}$ .
- $\mathbf{U}_{\text{obs}} = \left[ U_j^{(i)} \right]_{\substack{1 \leq i \leq N \\ 1 \leq j \leq d}} \sim (\mu_{\mathbf{U}})^{\otimes N} \implies$  La matrice des  $N$  observations du vecteur  $\mathbf{U}$ .
- $\hat{\mathbf{U}}_{\text{obs}} = \left[ \hat{U}_j^{(i)} \right]_{\substack{1 \leq i \leq N \\ 1 \leq j \leq d}} \implies$  La matrice des  $N$  pseudo-observations du vecteur  $\mathbf{U}$ .

Soit  $\mathcal{M}$  un modèle d'apprentissage sélectionné pour estimer la densité  $f_{\mathbf{X}}$ . On suppose que  $\mathcal{M}$  est explicitement décomposable en une collection de  $d+1$  sous-modèles  $\mathcal{M}_1, \dots, \mathcal{M}_d, \mathcal{M}_{\text{cop}}$  que l'on peut respectivement affecter aux distributions marginales et à la distribution de copule. Cette hypothèse exclut les modèles multidimensionnels n'intégrant pas directement la notion de copule. Par exemple, l'estimation d'une loi conjointe avec des noyaux multidimensionnels ne se prête pas à la spécification d'un sous-modèle  $\mathcal{M}_{\text{cop}}$ . On peut évidemment extraire de  $\hat{f}_{\mathbf{X}}$  une densité de copule  $\hat{c}_{\mathbf{U}}$  mais celle-ci n'est pas régie par un modèle  $\mathcal{M}_{\text{cop}}$  au sens où l'accès à  $\hat{\mathbf{U}}_{\text{obs}}$  ne permet pas de reconstruire  $\hat{c}_{\mathbf{U}}$ . Pour tout échantillon d'apprentissage  $\mathbf{x}_{\text{obs}}$ , la calibration du modèle  $\mathcal{M}$  peut être décrite par l'application suivante :

$$\mathcal{C}_{\mathcal{M}} : \begin{array}{ccc} \mathbb{R}^{N \times d} & \longrightarrow & \mathbb{L}^2(\mathbb{R}^d) \\ \mathbf{x}_{\text{obs}} & \longmapsto & \hat{f}_{\mathbf{X}}(\cdot | \mathcal{M}, \mathbf{x}_{\text{obs}}) \end{array} , \quad (4.4)$$

avec :

$$\hat{f}_{\mathbf{X}}(\mathbf{x} | \mathcal{M}, \mathbf{x}_{\text{obs}}) = \left[ \prod_{j=1}^d \hat{f}_{X_j}(x_j | \mathcal{M}_j, \mathbf{x}_{\text{obs}}) \right] \times \hat{c}_{\mathbf{U}}(\hat{\mathbf{u}} | \mathcal{M}_{\text{cop}}, \mathbf{x}_{\text{obs}}) , \quad (4.5)$$

où le vecteur  $\hat{\mathbf{u}} = (\hat{u}_1, \dots, \hat{u}_d)$  s'obtient à partir des densités marginales estimées :

$$\hat{u}_j = \hat{F}_{X_j}(x_j | \mathcal{M}_j, \mathbf{x}_{\text{obs}}) = \int_{-\infty}^{x_j} \hat{f}_{X_j}(t | \mathcal{M}_j, \mathbf{x}_{\text{obs}}) dt . \quad (4.6)$$

Imaginons que nous sommes dans un cadre complètement paramétrique dans lequel  $\mathcal{M}$  est assimilable à une famille multidimensionnelle  $\mathcal{F}$  divisible en  $d+1$  sous-familles  $\mathcal{F}_1, \dots, \mathcal{F}_d, \mathcal{F}_{\text{cop}}$ . Si la loi sous-jacente  $\mu_{\mathbf{X}}$  appartient effectivement à cette classe de distribution  $\mathcal{F}$ , la densité apprise  $\hat{f}_{\mathbf{X}} = \mathcal{C}_{\mathcal{M}}(\mathbf{X}_{\text{obs}})$  n'est affectée que par une incertitude liée au manque de données ("*data uncertainty*"). Elle rend compte de la capacité à reconstruire  $f_{\mathbf{X}}$  à partir de  $\mathbf{x}_{\text{obs}}$  une fois que le modèle de distribution est correctement spécifié. Elle diminue avec l'augmentation de la taille de l'échantillon d'apprentissage. Pour les différentes méthodes d'inférence paramétrique présentées dans le Chapitre 1, des résultats de convergence nous renseignent sur la vitesse de la réduction de cette incertitude au fur et à mesure de l'accroissement de  $N$ . Cependant, dès lors que la dimension  $d$  augmente ou que l'on incorpore des outils semi-paramétriques, voire non-paramétriques, il n'est pas raisonnable de supposer connaître le modèle de distribution  $\mathcal{M}$  auquel appartient  $\mu_{\mathbf{X}}$ . S'ajoute

ainsi un second type d'incertitudes épistémiques, introduit par la méconnaissance du modèle  $\mathcal{M}$  (“*model uncertainty*”). Une étude de l'effet combiné de ces deux types d'incertitude est notamment proposée par [SANKARARAMAN et MAHADEVAN \[2013\]](#).

Quand le modèle  $\mathcal{M}$  est inconnu, il faut choisir un modèle d'apprentissage parmi une collection de  $m$  candidats  $\mathcal{M}^{(1)}, \dots, \mathcal{M}^{(m)}$  à l'aide des techniques de sélection de modèle décrites dans le Chapitre 1. Par exemple, pour comparer la qualité de l'ajustement à  $\mathbf{x}_{\text{obs}}$  pour deux modèles  $\mathcal{M}^{(i)}$  et  $\mathcal{M}^{(j)}$  s'apparentant à des familles paramétriques  $\mathcal{F}^{(i)}$  et  $\mathcal{F}^{(j)}$ , le processus de sélection consiste souvent à maximiser un critère d'information tel que celui de [AKAIKE \[1974\]](#). Pour valider le choix d'une famille paramétrique  $\mathcal{F}$  que l'on pense suffisamment adaptée, il est rappelé dans [WOODRUFF et MOORE \[1988\]](#) que l'on peut utiliser le très classique test du  $\chi^2$  ou l'un des nombreux tests d'adéquation utilisant la fonction de répartition empirique. Récemment, [KANG et collab. \[2019\]](#) ont proposé un outil de décision pour départager les techniques d'apprentissage paramétriques et non-paramétriques d'une distribution marginale. Quelle que soit la démarche de sélection utilisée, le processus d'élection d'un modèle optimal  $\mathcal{M}^*$  peut être décrit par l'application suivante :

$$\begin{aligned} \mathcal{S}: \mathbb{R}^{N \times d} &\longrightarrow \{\mathcal{M}^{(1)}, \dots, \mathcal{M}^{(m)}\} \\ \mathbf{x}_{\text{obs}} &\longmapsto \mathcal{M}^* \end{aligned} \quad (4.7)$$

Évidemment, notamment dans le contexte du “*small data*”, le modèle retenu  $\mathcal{M}^* = \mathcal{S}(\mathbf{x}_{\text{obs}})$  peut différer d'un échantillon à l'autre. La phase d'apprentissage statistique  $\mathcal{L}$  s'écrit donc :

$$\begin{aligned} \mathcal{L}: \mathbb{R}^{N \times d} &\longrightarrow \mathbb{L}^2(\mathbb{R}^d) \\ \mathbf{x}_{\text{obs}} &\longmapsto \hat{f}_{\mathbf{X}}(\cdot | \mathcal{M}^*, \mathbf{x}_{\text{obs}}) = \mathcal{C}_{\mathcal{M}^*}(\mathbf{x}_{\text{obs}}) \quad \text{où} \quad \mathcal{M}^* = \mathcal{S}(\mathbf{x}_{\text{obs}}). \end{aligned} \quad (4.8)$$

L'incertitude liée au manque de données impacte donc la phase de sélection d'un meilleur modèle  $\mathcal{M}^*$  ainsi que la phase de calibration de ce modèle. Lorsque l'état de connaissance se limite à un échantillon  $\mathbf{x}_{\text{obs}}$ , il n'est pas aisé d'opérer une stricte délimitation du périmètre d'existence de l'incertitude liée au manque de données et de celle liée à la méconnaissance du modèle de distribution sous-jacent. Comme on peut le voir sur la Figure 4.1, elles sont intimement liées au cœur du mécanisme inférentiel. Dans ce travail, on ne cherchera pas à quantifier séparément l'impact de ces deux types d'incertitude. Au contraire, elles seront traitées conjointement comme un faisceau indissociable d'incertitudes épistémiques émanant du fait que l'état de connaissance sur  $\mu_{\mathbf{X}}$  est restreint au seul  $\mathbf{x}_{\text{obs}}$ . Par ailleurs, dans la plupart des exemples qui sont traités, un modèle d'apprentissage  $\mathcal{M}$  est choisi en amont et l'apprentissage statistique  $\mathcal{L}$  se limite à  $\mathcal{C}_{\mathcal{M}}$ .

S'il a été dit que tout modèle d'apprentissage  $\mathcal{M}$  issu de la séparation copule-marginales convient à notre étude, il est évidemment recommandé de pré-sélectionner des modèles permettant d'estimer  $\hat{f}_{\mathbf{X}}$  avec précision. Les performances d'un modèle de densité marginale  $\mathcal{M}_j$  sont jugées à l'aune de la résistance à la multimodalité et de la finesse de la description des queues de distribution. De même, un modèle de copule  $\mathcal{M}_{\text{cop}}$  est décrété satisfaisant s'il peut prendre en compte des motifs de dépendance asymétriques et polymorphes tout en restant assez parcimonieux pour que son utilisation postérieure n'implique pas un coût de calcul déraisonnable.

#### 4.2.1.2 De la variabilité d'échantillonnage à la variabilité de prédiction du risque de défaillance

Mener une analyse de fiabilité lorsqu'on ne dispose que d'un échantillon  $\mathbf{x}_{\text{obs}}$  des entrées nécessite d'exécuter successivement un algorithme  $\mathcal{L}$  d'apprentissage statistique et un algorithme  $\mathcal{R}$  de calcul du risque de défaillance. L'algorithme final, résultant de l'enchaînement de  $\mathcal{R}$  et  $\mathcal{L}$ , est noté  $\mathcal{D}$ . Il englobe tout le savoir-faire déployé pour l'estimation puis l'intégration de la densité

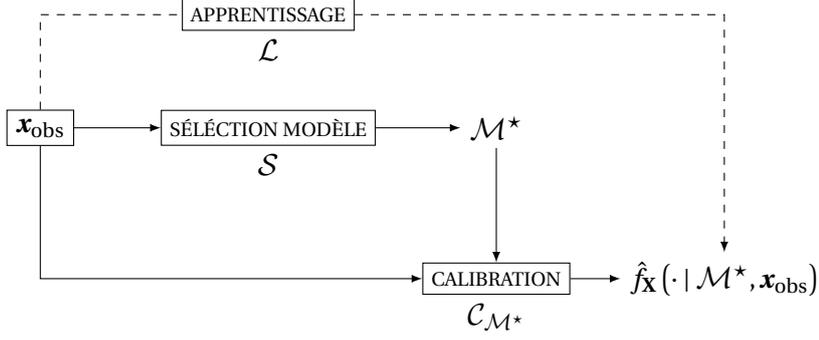


FIGURE 4.1 – Décomposition des incertitudes épistémiques au cours de l'apprentissage statistique.

multidimensionnelle des entrées du problème :

$$P_f(\mathbf{x}_{\text{obs}}) = \mathbb{P}(\mathbf{X} \in D_f \mid \mathbf{X}_{\text{obs}} = \mathbf{x}_{\text{obs}}) \quad \text{avec} \quad D_f := \{\mathbf{x} \in \mathbb{R}^d : \phi(\mathbf{x}) > T\} \quad (4.9)$$

$$= \int_{\mathbb{R}^d} \mathbb{1}_{D_f}(\mathbf{x}) \hat{f}_{\mathbf{X}}(\mathbf{x} \mid \mathbf{x}_{\text{obs}}) d\mathbf{x} = \int_{\mathbb{R}^d} \mathbb{1}_{D_f} \mathcal{L}(\mathbf{x}_{\text{obs}}) d\lambda \quad (4.10)$$

$$= [\mathcal{R} \circ \mathcal{L}](\mathbf{x}_{\text{obs}}) := \mathcal{D}(\mathbf{x}_{\text{obs}}). \quad (4.11)$$

Traiter le Problème 4.1 consiste à se demander comment auraient évolué la densité estimée  $\mathcal{L}(\mathbf{x}_{\text{obs}})$  et la probabilité calculée  $\mathcal{D}(\mathbf{x}_{\text{obs}})$  si un second échantillon  $\tilde{\mathbf{x}}_{\text{obs}}$  s'était substitué à  $\mathbf{x}_{\text{obs}}$ . Plus largement, l'idée est de parvenir à appréhender le mécanisme de propagation de la variabilité d'échantillonnage à travers les algorithmes  $\mathcal{L}$  et  $\mathcal{D}$ . Sachant que  $\mathbf{X}_{\text{obs}}$  est constitué d'observations indépendantes et identiquement distribuées, la loi de  $\mathbf{X}_{\text{obs}}$  s'écrit  $(\mu_{\mathbf{X}})^{\otimes N}$  et porte le nom de mesure d'échantillonnage. Comme les opérateurs  $\mathcal{L}$  et  $\mathcal{D}$  sont déterministes, le transfert de l'incertitude portée par  $\mathbf{X}_{\text{obs}}$  donne naissance à des mesures images qui caractérisent le comportement aléatoire des différents objets créés. D'un point de vue probabiliste, si on souhaitait être parfaitement rigoureux, il faudrait supposer que l'on est capable de munir l'espace fonctionnel  $\mathbb{L}^2(\mathbb{R}^d)$  (resp. le segment  $[0, 1]$ ) d'une tribu rendant l'application  $\mathcal{L}$  (resp.  $\mathcal{D}$ ) mesurable par rapport à la mesure d'échantillonnage. De plus amples détails à ce sujet sont fournis dans BOGACHEV [2007]. Les différentes étapes du mécanisme apparaissent sur le schéma de la Figure 4.2 avec notamment :

- $(\mu_{\mathbf{X}})^{\otimes N} \implies$  La mesure d'échantillonnage,
- $\mu_{\hat{f}_j} \implies$  L'image de  $(\mu_{\mathbf{X}})^{\otimes N}$  après apprentissage de la  $j$ -ème marginale,
- $\mu_{\hat{c}} \implies$  L'image de  $(\mu_{\mathbf{X}})^{\otimes N}$  après apprentissage de la copule,
- $\mu_{\hat{f}} \implies$  L'image de  $(\mu_{\mathbf{X}})^{\otimes N}$  après ré-assemblage des composantes fonctionnelles,
- $\mu_{P_f} \implies$  L'image de  $\mu_{\hat{f}}$  après calcul de la probabilité de défaillance.

La définition de ces mesures permet de clarifier le lien existant entre la variabilité d'échantillonnage, la variabilité d'apprentissage et la variabilité dans l'estimation du risque de défaillance. Le Problème 4.2 de séparation des incertitudes conformément à la décomposition copule-marginales revient donc à examiner la sensibilité de l'application  $\Psi$  définie par :

$$\Psi : \quad \mathbb{L}^2(\mathbb{R}) \quad \times \dots \times \quad \mathbb{L}^2(\mathbb{R}) \quad \times \quad \mathbb{L}^2(\mathcal{H}_d) \quad \longrightarrow \quad [0, 1] \quad (4.12)$$

$$\hat{f}_{X_1}(\cdot \mid \mathcal{M}_1, \mathbf{X}_{\text{obs}}) \quad \dots \quad \hat{f}_{X_d}(\cdot \mid \mathcal{M}_d, \mathbf{X}_{\text{obs}}) \quad \hat{c}_{\mathbf{U}}(\cdot \mid \mathcal{M}_{\text{cop}}, \mathbf{X}_{\text{obs}}) \quad \longmapsto \quad P_f(\mathbf{X}_{\text{obs}})$$

$\Psi$  calcule une probabilité de défaillance à partir d'une version désassemblée de la densité estimée au cours de l'apprentissage statistique. Exécuter  $\Psi$  consiste à :

- (a) Dédire les fonctions de répartition marginales  $\hat{F}_{X_j}(\cdot \mid \mathcal{M}_j, \mathbf{X}_{\text{obs}})$  en utilisant la formule (4.6).
- (b) Reconstruire la densité conjointe  $\hat{f}_{\mathbf{X}}(\cdot \mid \mathcal{M}, \mathbf{X}_{\text{obs}})$  avec la formule (4.5).

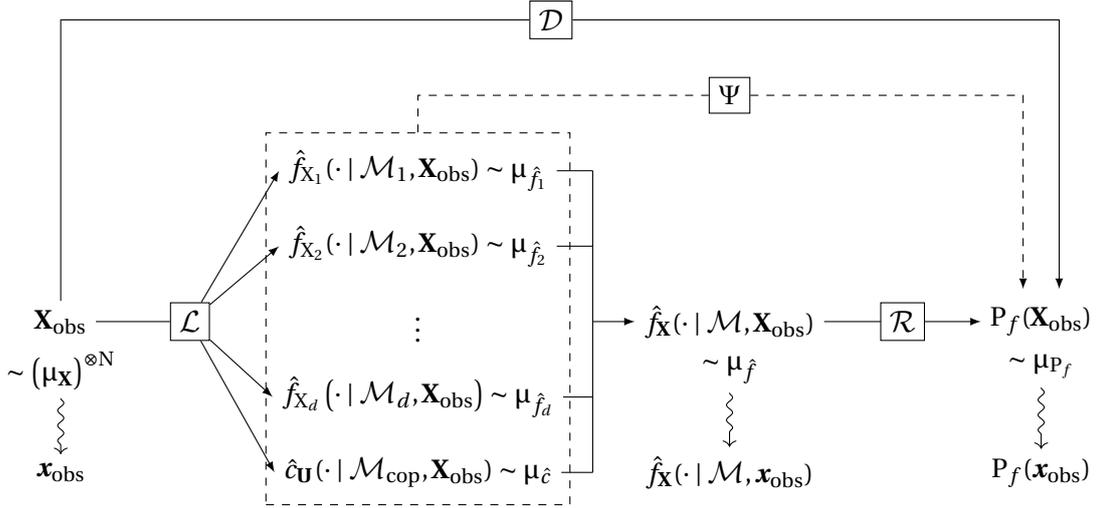


FIGURE 4.2 – Propagation de la variabilité d'échantillonnage.

(c) Appliquer l'algorithme de fiabilité  $\mathcal{R}$ .

Sachant que la sortie de  $\Psi$  est une probabilité de défaillance, on est dans un contexte d'ASF. Le coût de calcul de la fonction de performance  $\phi$  est supposé négligeable dans cette section, il en est de même pour celui d'une évaluation de  $\Psi$ . Le principal obstacle à la conduite de l'ASF est l'ignorance de la loi des entrées de  $\Psi$ . La connaissance de  $\mathbf{x}_{\text{obs}}$  donne seulement accès à  $\hat{f}_{\mathbf{X}}(\cdot | \mathbf{x}_{\text{obs}})$  puis  $P_f(\mathbf{x}_{\text{obs}})$ . Au départ, la connaissance du fonctionnement de  $\Psi$  se limite donc aux résultats fournis par une seule et unique exécution de la chaîne de propagation des incertitudes. Pour pallier ce problème, on développe dans la prochaine section un mécanisme de ré-échantillonnage permettant de simuler de nouvelles réalisations des objets  $\hat{f}_{\mathbf{X}}(\cdot | \mathbf{X}_{\text{obs}})$  et  $P_f(\mathbf{X}_{\text{obs}})$  en ne s'appuyant que sur le jeu de données  $\mathbf{x}_{\text{obs}}$  fourni initialement.

## 4.2.2 Mise en place d'un mécanisme de ré-échantillonnage

### 4.2.2.1 Étude de la dépendance fonctionnelle entre les composantes

Dans la Section 1.5.1, on a présenté les différentes méthodes d'inférence statistique utilisées pour l'estimation d'une densité multidimensionnelle suivant un modèle d'apprentissage  $\mathcal{M}$  émanant de la séparation copule-marginales. On cherche ici à analyser la nature de la relation existant entre les densités marginales  $\hat{f}_{X_j}(\cdot | \mathcal{M}_j, \mathbf{X}_{\text{obs}})$  et la densité de copule  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{X}_{\text{obs}})$ .

La situation la plus simple est celle où les entrées  $X_j$  sont indépendantes. Les différentes colonnes de  $\mathbf{X}_{\text{obs}}^{[ \cdot j ]}$  de  $\mathbf{X}_{\text{obs}}$  sont alors indépendantes. Il en résulte une indépendance mutuelle des densités marginales estimées puisque l'apprentissage de chaque marginale  $f_{X_j}$  ne fait appel qu'aux observations marginales associées, qui sont toutes rangées dans la colonne  $\mathbf{X}_{\text{obs}}^{[ \cdot j ]}$ . La distribution de copule peut être ignorée. Éventuellement, on peut considérer qu'il s'agit d'une fonction aléatoire de l'espace  $\mathbb{L}^2(\mathcal{H}_d)$ , constamment égale à la densité  $\pi_d$  de la copule indépendance.

Regardons maintenant ce qui se produit dans le cas général où les variables de  $\mathbf{X}$  sont dépendantes. Commençons par examiner le cas d'un modèle  $\mathcal{M}$  complètement paramétrique assimilable à une famille  $\mathcal{F}$ . On note  $\alpha_1, \dots, \alpha_d$  et  $\beta$  les paramétrages respectivement associés aux familles  $\mathcal{F}_1, \dots, \mathcal{F}_d$  et  $\mathcal{F}_{\text{cop}}$ . Le vecteur augmenté  $\theta$  est obtenu par concaténation des  $d + 1$  sous-paramétrages et représente le paramétrage global de la famille considérée. En supposant que la distribution sous-jacente est effectivement membre de la classe  $\mathcal{F}$  et en notant  $\theta$  le paramètre sous-jacent, il a été vu au Chapitre 1 que l'estimateur du maximum de vraisemblance (EMV) de  $\theta$

est asymptotiquement normal :

$$\sqrt{N} \times (\hat{\boldsymbol{\theta}}^{\text{MV}} - \bar{\boldsymbol{\theta}}) \xrightarrow[N \rightarrow \infty]{\text{loi}} \mathcal{N}(\mathbf{0}, [\mathcal{I}(\bar{\boldsymbol{\theta}})]^{-1}), \quad (4.13)$$

La matrice d'information de Fisher  $\mathcal{I}(\bar{\boldsymbol{\theta}})$  peut être décomposée en blocs de la manière suivante :

$$\mathcal{I}(\bar{\boldsymbol{\theta}}) = \begin{bmatrix} \mathcal{I}(\bar{\boldsymbol{\alpha}}_1) & \mathcal{J}_{\bar{\boldsymbol{\alpha}}_1 \bar{\boldsymbol{\alpha}}_2} & \cdots & \mathcal{J}_{\bar{\boldsymbol{\alpha}}_1 \bar{\boldsymbol{\alpha}}_d} & \mathcal{J}_{\bar{\boldsymbol{\alpha}}_1 \bar{\boldsymbol{\beta}}} \\ \mathcal{J}_{\bar{\boldsymbol{\alpha}}_1 \bar{\boldsymbol{\alpha}}_2} & \mathcal{I}(\bar{\boldsymbol{\alpha}}_2) & \cdots & \mathcal{J}_{\bar{\boldsymbol{\alpha}}_2 \bar{\boldsymbol{\alpha}}_d} & \mathcal{J}_{\bar{\boldsymbol{\alpha}}_2 \bar{\boldsymbol{\beta}}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{J}_{\bar{\boldsymbol{\alpha}}_1 \bar{\boldsymbol{\alpha}}_d} & \mathcal{J}_{\bar{\boldsymbol{\alpha}}_2 \bar{\boldsymbol{\alpha}}_d} & \cdots & \mathcal{I}(\bar{\boldsymbol{\alpha}}_d) & \mathcal{J}_{\bar{\boldsymbol{\alpha}}_d \bar{\boldsymbol{\beta}}} \\ \mathcal{J}_{\bar{\boldsymbol{\alpha}}_1 \bar{\boldsymbol{\beta}}} & \mathcal{J}_{\bar{\boldsymbol{\alpha}}_2 \bar{\boldsymbol{\beta}}} & \cdots & \mathcal{J}_{\bar{\boldsymbol{\alpha}}_d \bar{\boldsymbol{\beta}}} & \mathcal{I}(\bar{\boldsymbol{\beta}}) \end{bmatrix} \quad (4.14)$$

Chaque bloc  $\mathcal{I}(\bar{\boldsymbol{\alpha}}_j)$  correspond à la matrice de Fisher associée à la densité marginale  $f_{X_j}(\cdot | \mathcal{F}_j, \bar{\boldsymbol{\alpha}}_j)$ . Le bloc  $\mathcal{I}(\bar{\boldsymbol{\beta}})$  se rapporte quant à lui à la densité de copule  $c_{\mathbf{U}}(\cdot | \mathcal{F}_{\text{cop}}, \bar{\boldsymbol{\beta}})$ . Tous les blocs non diagonaux rendent compte des dépendances asymptotiques qui sont à l'œuvre entre les paramètres qui n'appartiennent pas à la même composante fonctionnelle. Dans le cas général d'une densité multidimensionnelle  $f_{\mathbf{X}}(\cdot | \mathcal{F}, \bar{\boldsymbol{\theta}})$ , il n'y a aucune raison que tous les blocs non-diagonaux soient nuls et que  $\mathcal{I}(\bar{\boldsymbol{\theta}})$  soit diagonale par blocs. Par conséquent, la matrice de covariance asymptotique de l'EMV  $\hat{\boldsymbol{\theta}}^{\text{MV}}$  ne sera pas non plus diagonale par blocs. On comprend donc que les estimateurs  $\hat{\boldsymbol{\alpha}}_1^{\text{MV}}, \dots, \hat{\boldsymbol{\alpha}}_d^{\text{MV}}$  et  $\hat{\boldsymbol{\beta}}^{\text{MV}}$  ne sont pas asymptotiquement indépendants.

Dans le contexte "small data" de ce travail, la notion de dépendance asymptotique pourrait ne pas être considérée comme pertinente. Cependant des mécanismes de dépendance existent entre  $\hat{\boldsymbol{\alpha}}_1^{\text{MV}}, \dots, \hat{\boldsymbol{\alpha}}_d^{\text{MV}}$  et  $\hat{\boldsymbol{\beta}}^{\text{MV}}$  bien avant ce stade. Il n'y a pas de résultat théorique permettant de décrire ce phénomène finement. Toutefois, on se propose d'examiner un cas jouet en dimension  $d = 2$  :

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{où } \boldsymbol{\mu} \text{ est inconnu et } \boldsymbol{\Sigma} = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} \quad \text{avec } r \in ]-1, 1[. \quad (4.15)$$

On dispose d'un échantillon d'apprentissage  $\mathbf{X}_{\text{obs}}$  constitué de  $N$  observations conjointes de  $\mathbf{X}$ . Dans ce contexte, l'EMV (noté ici  $\hat{\boldsymbol{\mu}}$  pour alléger les notations) est donné par :

$$\hat{\boldsymbol{\mu}}^\top = [\hat{\mu}_1 \quad \hat{\mu}_2] = \left[ \frac{1}{N} \sum_{i=1}^N X_1^{(i)} \quad \frac{1}{N} \sum_{i=1}^N X_2^{(i)} \right] = \frac{1}{N} \left[ \mathbf{1}^\top \mathbf{X}_{\text{obs}}^{[\cdot 1]} \quad \mathbf{1}^\top \mathbf{X}_{\text{obs}}^{[\cdot 2]} \right] \quad \text{avec } \mathbf{1} := \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^d. \quad (4.16)$$

et peut être réécrit comme suit :

$$\hat{\boldsymbol{\mu}} = \frac{1}{N} \begin{bmatrix} \mathbf{1}^\top & \mathbf{0}_{1 \times N} \\ \mathbf{0}_{1 \times N} & \mathbf{1}^\top \end{bmatrix} \begin{bmatrix} \mathbf{X}_{\text{obs}}^{[\cdot 1]} \\ \mathbf{X}_{\text{obs}}^{[\cdot 2]} \end{bmatrix}. \quad (4.17)$$

On peut alors remarquer que :

$$\begin{bmatrix} \mathbf{X}_{\text{obs}}^{[\cdot 1]} \\ \mathbf{X}_{\text{obs}}^{[\cdot 2]} \end{bmatrix} \sim \mathcal{N}(\boldsymbol{\mu}_{\text{obs}}, \boldsymbol{\Sigma}_{\text{obs}}) \quad \text{avec } \boldsymbol{\mu}_{\text{obs}} = \begin{bmatrix} \mu_1 \mathbf{1} \\ \mu_2 \mathbf{1} \end{bmatrix} \quad \text{et } \boldsymbol{\Sigma}_{\text{obs}} = \begin{bmatrix} I_N & r I_N \\ r I_N & I_N \end{bmatrix}. \quad (4.18)$$

Des calculs matriciels élémentaires permettent alors d'avoir les résultats suivants :

$$\mathbb{E}[\hat{\boldsymbol{\mu}}] = \boldsymbol{\mu} \quad \text{et} \quad \mathbb{V}(\hat{\boldsymbol{\mu}}) = \mathbb{E} \left[ (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^\top \right] = \frac{1}{N} \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} =: \boldsymbol{\Sigma}_N. \quad (4.19)$$

Comme  $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , la copule  $c_{\mathbf{U}}$  est gaussienne et paramétrée par la matrice de corrélation  $\boldsymbol{\rho}$  du vecteur  $\mathbf{X}$ . Pour cet exemple, on a  $\mathbb{V}(X_1) = \mathbb{V}(X_2) = 1$  et  $\boldsymbol{\rho} = \boldsymbol{\Sigma}$ . L'égalité  $N \times \mathbb{V}(\hat{\boldsymbol{\mu}}) = \boldsymbol{\rho}$  indique que les EMV  $\hat{\mu}_1$  et  $\hat{\mu}_2$  sont des variables aléatoires dépendantes, et ce de par l'existence d'une corrélation entre  $X_1$  et  $X_2$ . Cet exemple entérine l'idée selon laquelle les estimateurs  $\hat{\boldsymbol{\alpha}}_1^{\text{MV}}, \dots, \hat{\boldsymbol{\alpha}}_d^{\text{MV}}$  et  $\hat{\boldsymbol{\beta}}^{\text{MV}}$  ne sont pas indépendants. Les fonctions  $\hat{f}_{X_1}(\cdot | \mathcal{F}_1, \hat{\boldsymbol{\alpha}}_1^{\text{MV}}), \dots, \hat{f}_{X_d}(\cdot | \mathcal{F}_d, \hat{\boldsymbol{\alpha}}_d^{\text{MV}})$  et  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{F}_{\text{cop}}, \hat{\boldsymbol{\beta}}^{\text{MV}})$  estimées avec la méthode du maximum de vraisemblance ne peuvent donc pas être considérées comme indépendantes.

Il avait été également dit dans le Chapitre 1 que l'estimation conjointe des paramètres  $\bar{\boldsymbol{\alpha}}_1, \dots, \bar{\boldsymbol{\alpha}}_d$  et  $\bar{\boldsymbol{\beta}}$  telle que proposée par la méthode du maximum de vraisemblance devient difficile à déployer lorsque la dimension du problème augmente. Des alternatives séquentielles telles que la méthode *Inference Functions for Margins* (IFM) de JOE et XU [1996] ou la méthode semi-paramétrique (SP) de GENEST et collab. [1995] sont alors privilégiées pour simplifier la phase de descente de gradient. Pour ces deux approches, on se propose d'examiner l'évolution du lien de dépendance entre les paramètres estimés. Pour la méthode IFM, on rappelle qu'elle s'exécute en trois temps :

(IFM1) Pour chaque densité marginale, calcul de  $\hat{\boldsymbol{\alpha}}_j^{\text{MV}}$  à partir de la colonne  $\mathbf{X}_{\text{obs}}^{[\cdot, j]}$ .

(IFM2) Construction de la matrice  $\hat{\mathbf{U}}_{\text{obs}}$  des pseudo-observations de la copule :

$$\hat{\mathbf{U}}_{\text{obs}} = \left[ \hat{U}_j^{(i)} \right]_{\substack{1 \leq i \leq N \\ 1 \leq j \leq d}} \quad \text{avec} \quad \hat{U}_j^{(i)} = \hat{F}_{X_j} \left( X_j^{(i)} \mid \mathcal{F}_j, \hat{\boldsymbol{\alpha}}_j^{\text{MV}} \right). \quad (4.20)$$

(IFM3) Calcul de  $\hat{\boldsymbol{\beta}}^{\text{MV}}$  à partir de la colonne  $\hat{\mathbf{U}}_{\text{obs}}$ .

En revenant éventuellement au cas bidimensionnel gaussien présenté plus haut, on comprend facilement que la phase initiale (IFM1) va à nouveau impliquer une dépendance statistique entre  $\hat{\boldsymbol{\alpha}}_1^{\text{MV}}, \dots, \hat{\boldsymbol{\alpha}}_d^{\text{MV}}$  en dépit du fait que chaque  $\hat{\boldsymbol{\alpha}}_j^{\text{MV}}$  est calculé à partir de sa propre colonne  $\mathbf{X}_{\text{obs}}^{[\cdot, j]}$ . Par ailleurs, la phase (IFM2) de renormalisation s'appuie directement sur les fonctions de répartition marginales estimées. De par ce séquençage,  $\hat{\boldsymbol{\beta}}^{\text{MV}}$  dépend des valeurs prises par les  $\hat{\boldsymbol{\alpha}}_j^{\text{MV}}$  puisque ce sont elles qui déterminent la qualité des pseudo-observations utilisées pour calculer  $\hat{\boldsymbol{\beta}}^{\text{MV}}$ . Ces deux arguments indiquent qu'on est à nouveau en présence d'une technique d'estimation des paramètres qui introduit une dépendance entre composantes fonctionnelles.

La méthode SP ne change rien à la dépendance statistique existant entre les densités marginales estimées mais tend à éliminer la dépendance introduite par (IFM2). L'étape de renormalisation est exécutée sans faire appel au résultat de l'apprentissage des marginales :

(SP2) Construction de la matrice  $\hat{\mathbf{U}}_{\text{obs}}$  des pseudo-observations de la copule :

$$\hat{\mathbf{U}}_{\text{obs}} = \left[ \hat{U}_j^{(i)} \right]_{\substack{1 \leq i \leq N \\ 1 \leq j \leq d}} \quad \text{avec} \quad \hat{U}_j^{(i)} = \frac{1}{N+1} \sum_{k=1}^N \mathbb{1}_{\{X_j^{(k)} \leq X_j^{(i)}\}}. \quad (4.21)$$

L'idée derrière (SP2) est d'éliminer de potentielles erreurs introduites par un choix erroné de famille marginale  $\mathcal{F}_j$  ou une estimation imprécise de l'un des paramètres  $\bar{\boldsymbol{\alpha}}_j$ . Pour cela, on utilise la notion de copule empirique de DEHEUVELS [1979]. Les pseudo-observations de la copule sont fabriquées par renormalisation des colonnes de  $\mathbf{X}_{\text{obs}}$  avec les fonctions de répartition empiriques. L'estimation de la densité de copule n'est alors plus influencée par la modélisation des lois marginales. Si recourir à (SP2) semble atténuer la dépendance entre  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{F}_{\text{cop}}, \hat{\boldsymbol{\beta}}^{\text{MV}})$  et la collection des densités marginales estimées  $\hat{f}_{X_1}(\cdot | \mathcal{F}_1, \hat{\boldsymbol{\alpha}}_1^{\text{MV}}), \dots, \hat{f}_{X_d}(\cdot | \mathcal{F}_d, \hat{\boldsymbol{\alpha}}_d^{\text{MV}})$ , rien ne permet d'établir une parfaite indépendance entre ces deux entités. Il ne faut pas perdre de vue qu'elles découlent toujours d'une même matrice aléatoire  $\mathbf{X}_{\text{obs}}$ .

Ainsi, les trois principales méthodes d'estimation paramétrique vues au Chapitre 1 semblent impliquer la création d'une dépendance statistique entre les paramètres estimés prenant ensuite la forme d'une dépendance fonctionnelle entre les composantes de la densité conjointe. On peut se demander si cette remarque reste valable pour un modèle d'apprentissage  $\mathcal{M}$  intégrant des outils non-paramétriques. Il devient alors plus difficile d'appréhender la notion de dépendance fonctionnelle car on ne peut plus réduire le comportement d'une composante à celui de son vecteur de paramètres. Reprenons le cas jouet du vecteur  $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  et examinons l'apprentissage des deux densités marginales à l'aide de noyaux gaussiens. Pour  $X_1$ , on a :

$$\hat{f}_{X_1}(x_1 | \mathcal{M}_1, \mathbf{X}_{\text{obs}}) = \hat{f}_{\text{KDE}}\left(x_1 \mid \mathbf{X}_{\text{obs}}^{[\bullet 1]}, \hat{h}_1\right) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\hat{h}_1} \mathcal{K}_{\mathcal{N}}\left(\frac{x_1 - X_1^{(i)}}{\hat{h}_1}\right). \quad (4.22)$$

$\mathcal{K}_{\mathcal{N}} : z \mapsto \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$  représente le noyau gaussien tandis que la fenêtre de lissage  $\hat{h}_1$  découle d'une application de la règle de SILVERMAN [1986]. Le moment d'ordre 1 de la densité estimée vaut alors :

$$\hat{m}_1 = \int_{\mathbb{R}} x_1 \hat{f}_{X_1}(x_1 | \mathcal{M}_1, \mathbf{X}_{\text{obs}}) dx_1 = \frac{1}{\hat{h}_1 \times N} \sum_{i=1}^N \int_{\mathbb{R}} x_1 \mathcal{K}_{\mathcal{N}}\left(\frac{x_1 - X_1^{(i)}}{\hat{h}_1}\right) dx_1, \quad (4.23)$$

$$= \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}} (\hat{h}_1 z_1 + X_1^{(i)}) \mathcal{K}_{\mathcal{N}}(z_1) dz_1, \quad (4.24)$$

$$= \frac{1}{N} \sum_{i=1}^N \left( \hat{h}_1 \underbrace{\int_{\mathbb{R}} z_1 \mathcal{K}_{\mathcal{N}}(z_1) dz_1}_{=0} + X_1^{(i)} \underbrace{\int_{\mathbb{R}} \mathcal{K}_{\mathcal{N}}(z_1) dz_1}_{=1} \right), \quad (4.25)$$

$$= \frac{1}{N} \sum_{i=1}^N X_1^{(i)} = \hat{\boldsymbol{\mu}}_1^{\text{MV}}. \quad (4.26)$$

Le calcul est d'ailleurs valable quelle que soit la technique d'estimation de la fenêtre de lissage que l'on décide d'utiliser. Le vecteur  $\hat{\boldsymbol{m}}$  est finalement égal à  $\hat{\boldsymbol{\mu}}^{\text{MV}}$  étudié plus haut. Les moments d'ordre 1 des densités marginales estimées sont donc distribués suivant une loi gaussienne de matrice de covariance  $\boldsymbol{\Sigma}_N$ .

L'examen de ce petit exemple indique que la dépendance fonctionnelle entre densités marginales continue d'exister en présence de modèles d'apprentissage  $\mathcal{M}_1$  et  $\mathcal{M}_2$  non-paramétriques. Pour ce qui est de la dépendance entre la collection de densités marginales et la densité de copule, sa nature est déterminée par la méthode de renormalisation choisie. Dans l'esprit de (IFM2), on peut construire  $\hat{\mathbf{U}}_{\text{obs}}$  à partir des fonctions de répartition  $\hat{F}_{X_j}(\cdot | \mathcal{M}_j, \mathbf{X}_{\text{obs}})$  et introduire, de par ce séquençage, un fort lien de dépendance fonctionnelle. Au contraire, si on fait à nouveau appel à une renormalisation de type (SP2), le caractère de cette dépendance semble plus ténu.

#### 4.2.2.2 Énoncé d'une hypothèse sur la distribution fonctionnelle des composantes

La section précédente vient de montrer que les techniques mentionnées dans la littérature pour estimer une densité multidimensionnelle introduisent, quel que soit le modèle d'apprentissage  $\mathcal{M}$  considéré, une dépendance fonctionnelle entre les différentes composantes de la séparation copule-marginales. À la Figure 4.2, la propagation de la variabilité d'échantillonnage a été schématisée comme suit :

$$\begin{array}{ccc} \mathbf{X}_{\text{obs}} & \xrightarrow{\boxed{\mathcal{L}}} & \hat{f}_{\mathbf{X}}(\cdot | \mathcal{M}, \mathbf{X}_{\text{obs}}) \\ \sim (\boldsymbol{\mu}_{\mathbf{X}})^{\otimes N} & & \sim \boldsymbol{\mu}_{\hat{f}} \end{array} \quad (4.27)$$

Les arguments présentés dans la section précédente nous amènent à dire que la loi  $\boldsymbol{\mu}_{\hat{f}}$  évolue en fonction du modèle d'apprentissage  $\mathcal{M}$ , ce qui est normal et attendu, mais aussi de mécanismes inhérents à la technique d'estimation. Ces derniers introduisent une dépendance fonctionnelle entre les densités marginales et la densité de copule.

On pourrait essayer de simuler cette loi, c'est-à-dire de produire de nouvelles réalisations de la variable fonctionnelle  $\hat{f}_{\mathbf{X}}(\cdot | \mathbf{X}_{\text{obs}})$ . Il faudrait pour cela disposer d'un générateur de la loi de  $\mathbf{X}$  :

$$\begin{aligned} G_{\mathbf{X}} : \mathbb{N} &\longrightarrow \mathbb{R}^d \times \dots \times \mathbb{R}^d \\ n &\longmapsto \underbrace{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}}_{n\text{-échantillon de } \mathbf{X}} . \end{aligned} \quad (4.28)$$

Ce générateur serait le point de départ de la propagation de la variabilité d'échantillonnage à travers l'algorithme d'apprentissage. Pour une technique d'estimation donnée, la répétition d'apprentissages pour différents échantillons  $\mathbf{x}_{\text{obs}}$  permettrait de simuler une série de densités jointes suivant  $\mu_{\hat{f}}$ . On se reporte à l'Algorithme 1 pour le pseudo-code associé.

<b>Algorithme 1</b> : Simulation de $\hat{f}_{\mathbf{X}}(\cdot   \mathcal{M}, \mathbf{X}_{\text{obs}})$	
$G_{\mathbf{X}}$	$\Rightarrow$ Générateur de $\mathbf{X}$
<b>Entrées :</b> $\mathcal{M}_j$	$\Rightarrow$ Modèle d'apprentissage pour la densité marginale $f_{X_j}$
$\mathcal{M}_{\text{cop}}$	$\Rightarrow$ Modèle d'apprentissage pour la densité de copule $c_{\mathbf{U}}$
<ul style="list-style-type: none"> <li>▶ Simuler un N-échantillon <math>\mathbf{x}_{\text{obs}}</math> à l'aide de <math>G_{\mathbf{X}}</math></li> </ul>	
<b>pour chaque</b> $j \in \{1, \dots, d\}$ <b>faire</b>	
<ul style="list-style-type: none"> <li>▶ Estimer <math>\hat{f}_{X_j}(\cdot   \mathcal{M}_j, \mathbf{x}_{\text{obs}}^{[j]})</math></li> </ul>	
<ul style="list-style-type: none"> <li>▶ Fabriquer un N-échantillon de pseudo-observations <math>\hat{\mathbf{u}}_{\text{obs}}</math> en utilisant (au choix) : <ul style="list-style-type: none"> <li>▷ Les fonctions de répartition préalablement estimées</li> <li>▷ Les fonctions de répartition empiriques</li> </ul> </li> </ul>	
<ul style="list-style-type: none"> <li>▶ Estimer <math>\hat{c}_{\mathbf{U}}(\cdot   \mathcal{M}_{\text{cop}}, \hat{\mathbf{u}}_{\text{obs}})</math></li> </ul>	
<ul style="list-style-type: none"> <li>▶ Ré-assembler les composantes pour former la densité conjointe <math>\hat{f}_{\mathbf{X}}(\cdot   \mathcal{M}, \mathbf{x}_{\text{obs}})</math></li> </ul>	
<b>Sorties :</b> $\hat{f}_{\mathbf{X}}(\cdot   \mathcal{M}, \mathbf{x}_{\text{obs}})$	$\Rightarrow$ Une observation de l'apprentissage de $f_{\mathbf{X}}$

Évidemment, cette simulation ne peut pas être conduite car le générateur théorique  $G_{\mathbf{X}}$  est inaccessible. Quand bien même on aurait la capacité d'exécuter l'Algorithme 1, la nature de la dépendance fonctionnelle entre les objets créés resterait difficile à appréhender et à manipuler. Effectuer une AS sur l'application  $\Psi$  définie par l'équation (4.12) pour répondre au Problème 4.2 est donc une tâche ardue, d'autant plus si on ambitionne de proposer une solution pouvant convenir à toutes les procédures inférentielles. Une simplification évidente serait de supposer que les composantes fonctionnelles issues de l'apprentissage sont indépendantes :

$$[\hat{f}_{X_1} \quad \dots \quad \hat{f}_{X_d} \quad \hat{c}_{\mathbf{U}}] \sim \mu_{\hat{f}_1} \otimes \dots \otimes \mu_{\hat{f}_d} \otimes \mu_{\hat{c}} . \quad (4.29)$$

On sait d'ores et déjà que cette hypothèse n'est pas valable pour des techniques d'estimation telles que le maximum de vraisemblance ou les méthodes IFM et SP. Il est alors légitime de se demander s'il existe un contexte pour lequel l'indépendance des composantes est effectivement vérifiée. Elle peut être envisagée si les lots d'observations servant à l'estimation des composantes proviennent de sources indépendantes. C'est par exemple le cas si on choisit de partitionner  $\mathbf{X}_{\text{obs}}$  en  $d+1$  blocs d'observations. Pour cela, il faut prendre une suite d'entiers  $1 < n_1 < n_2 < \dots < n_d < N$  et poser :

$$\mathbf{X}_{\text{obs}}^{\text{B}_1} = [\mathbf{X}_j^{(i)}]_{\substack{1 \leq i \leq n_1 \\ 1 \leq j \leq d}} ; \quad \mathbf{X}_{\text{obs}}^{\text{B}_2} = [\mathbf{X}_j^{(i)}]_{\substack{n_1 < i \leq n_2 \\ 1 \leq j \leq d}} ; \quad \dots ; \quad \mathbf{X}_{\text{obs}}^{\text{B}_{\text{cop}}} = [\mathbf{X}_j^{(i)}]_{\substack{n_d < i \leq N \\ 1 \leq j \leq d}} . \quad (4.30)$$

Comme le montre la Figure 4.3, chaque bloc est alors réservé à l'apprentissage d'une composante spécifique. Les lignes  $\mathbf{X}_{\text{obs}}^{[i \cdot]}$  de  $\mathbf{X}_{\text{obs}}$  étant des observations indépendantes de  $\mu_{\mathbf{X}}$ , les différents blocs

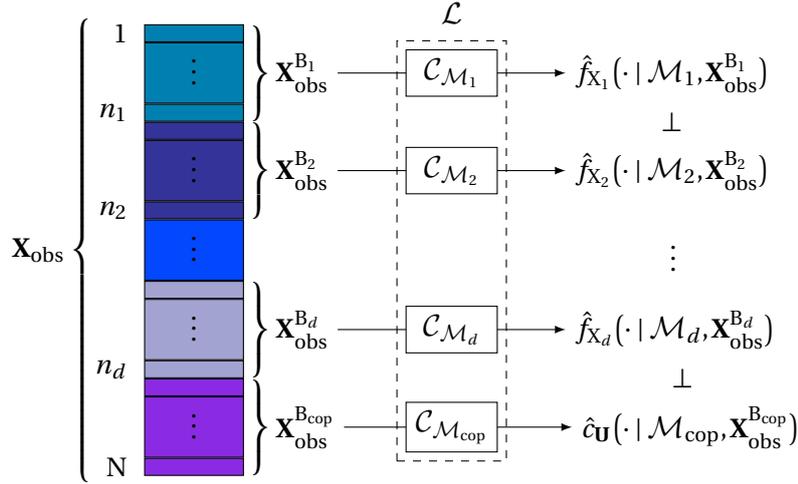


FIGURE 4.3 – Partitionnement de l'échantillon d'apprentissage en plusieurs blocs.

constitués sont indépendants les uns des autres et il en va de même pour les composantes fonctionnelles résultant de cette technique d'estimation. Évidemment, cette méthode n'est pas optimale au sens où  $\mathbf{X}_{\text{obs}}$  est sous-exploité. En effet, dans chacun des blocs de type  $\mathbf{X}_{\text{obs}}^{\text{B}_j}$ , la  $j$ -ème colonne est utilisée tandis que les autres sont abandonnées. Par ailleurs, chaque composante est estimée avec un plus petit budget d'observations. Dans un contexte "small data", implémenter cette technique d'estimation est assez peu recommandable. Toutefois, on prend conscience du fait que l'indépendance fonctionnelle n'est pas une hypothèse irréalisable. Il est d'ailleurs souhaitable de s'y ramener car :

- (a) Elle colle à l'intuition initiale du théorème de Sklar qui vise à une déconnexion totale de la modélisation de la distribution de copule et des distributions marginales.
- (b) La procédure inférentielle gagne en robustesse et en lisibilité car on cloisonne les erreurs.

On pourrait d'ailleurs chercher à mettre au point des techniques d'estimation permettant d'aboutir à l'indépendance fonctionnelle tout en exploitant de manière optimale le budget d'observations fourni par  $\mathbf{X}_{\text{obs}}$ . Dans la suite de ce travail, on va chercher à résoudre le Problème 4.2 sous l'hypothèse d'indépendance (4.29). On s'intéresse alors à la propagation de la variabilité d'échantillonnage à travers un algorithme d'apprentissage "parfait" calibrant le modèle  $\mathcal{M}$  choisi sans introduire la moindre dépendance fonctionnelle entre les composantes.

#### 4.2.2.3 Transformation de la chaîne de propagation des incertitudes

Pour synthétiser, on pourrait dire qu'il y a deux causes à la présence de dépendance fonctionnelle entre les composantes d'une densité multidimensionnelle. D'une part, si on disloque les colonnes de la matrice  $\mathbf{X}_{\text{obs}}$ , les vecteurs aléatoires résultants ne sont pas indépendants. D'autre part, la matrice  $\hat{\mathbf{U}}_{\text{obs}}$  des pseudo-observations de la copule est fabriquée à partir de  $\mathbf{X}_{\text{obs}}$ . Une stratégie pour obtenir l'indépendance serait de remplacer  $\mathbf{X}$  et  $\mathbf{U}$  par des vecteurs  $\mathbf{Z}$  et  $\mathbf{V}$  tels que :

$$\mathbf{Z} = [Z_1, \dots, Z_d] \in \mathbb{R}^d \quad \text{avec} \quad f_{\mathbf{Z}}(\mathbf{z}) := \prod_{j=1}^d f_{X_j}(z_j), \quad (4.31)$$

$$\mathbf{V} = [V_1, \dots, V_d] \in \mathcal{H}_d \quad \text{avec} \quad c_{\mathbf{V}}(\mathbf{v}) := c_{\mathbf{U}}(\mathbf{v}) \text{ et } \mathbf{V} \perp \mathbf{Z}. \quad (4.32)$$

$\mathbf{Z}$  est doté des mêmes lois marginales que  $\mathbf{X}$  mais ses variables constitutives  $Z_j$  sont mutuellement indépendantes. De son côté, le vecteur  $\mathbf{V}$  suit la même loi que  $\mathbf{U}$  tout en étant indépendant de  $\mathbf{Z}$ . Notons  $\mathbf{Z}_{\text{obs}}$  (resp.  $\mathbf{V}_{\text{obs}}$ ) un  $N$ -échantillon de  $\mathbf{Z}$  (resp.  $\mathbf{V}$ ) produit par un générateur  $G_{\mathbf{Z}}$  (resp.  $G_{\mathbf{V}}$ ) et agencé sous la forme d'une matrice aléatoire de  $\mathbb{R}^{N \times d}$ . Utiliser les colonnes de  $\mathbf{Z}_{\text{obs}}$  permet de construire une collection d'estimateurs  $\hat{f}_{X_j}(\cdot | \mathcal{M}_j, \mathbf{Z}_{\text{obs}})$  mutuellement indépendants. De même,

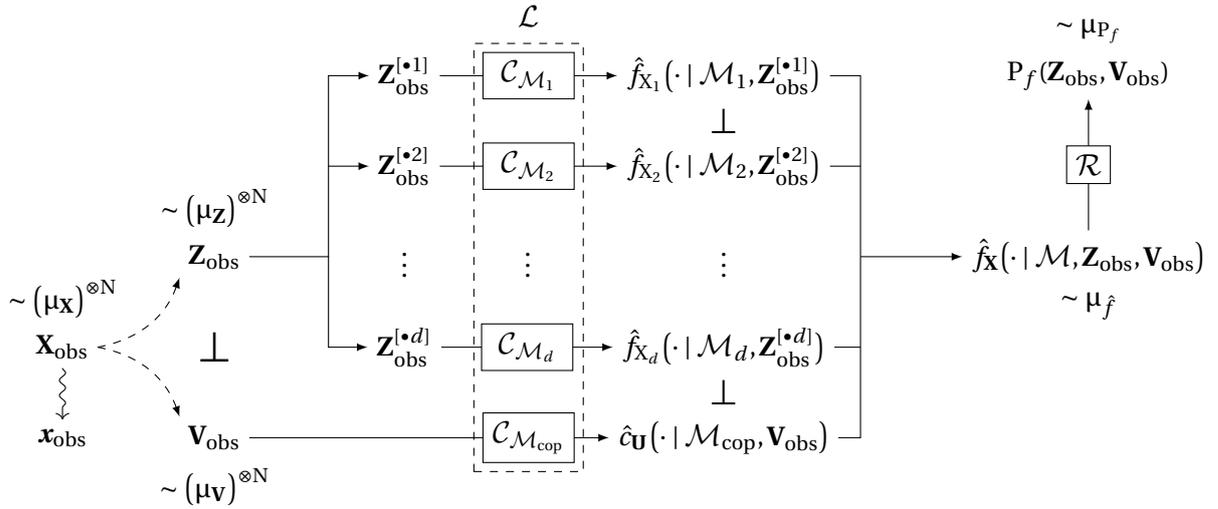


FIGURE 4.4 – Propagation de la variabilité d'échantillonnage après séparation des sources d'apprentissage.

l'indépendance entre  $\mathbf{V}$  et  $\mathbf{Z}$  garantit celle de l'estimateur  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{V}_{\text{obs}})$  vis-à-vis des estimateurs des densités marginales. On comprend donc que la clé de l'élimination de la dépendance fonctionnelle réside dans la séparation des sources d'apprentissage. Un pseudo-code pour cette procédure est fourni par l'Algorithme 2. Cela conduit à une transformation de la chaîne de propagation initiale de la variabilité d'échantillonnage (voir Figure 4.2) pour laquelle la phase d'apprentissage statistique  $\mathcal{L}$  était exécutée sur  $\mathbf{X}_{\text{obs}}$ . La nouvelle version de la chaîne de propagation, représentée à la Figure 4.4, inclut au contraire une parallélisation de toutes les tâches d'apprentissage. Le risque de défaillance est modélisé par une variable  $P_f(\mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}})$  issue de la propagation de la mesure d'échantillonnage  $(\mu_{\mathbf{z}})^{\otimes N} \otimes (\mu_{\mathbf{v}})^{\otimes N}$  à travers les algorithmes  $\mathcal{L}$  et  $\mathcal{R}$ . L'objectif est maintenant de comprendre comment simuler une série d'observations des matrices  $\mathbf{Z}_{\text{obs}}$  et  $\mathbf{V}_{\text{obs}}$  à partir d'un unique échantillon d'apprentissage  $\mathbf{x}_{\text{obs}}$ .

**Algorithme 2 :** Simulation de  $\hat{f}_{\mathbf{X}}(\cdot | \mathcal{M}, \mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}})$

**Entrées :**

- $G_{\mathbf{Z}} \Rightarrow$  Générateur de  $\mathbf{Z}$
- $G_{\mathbf{V}} \Rightarrow$  Générateur de  $\mathbf{V}$
- $\mathcal{M}_j \Rightarrow$  Modèle d'apprentissage pour la densité marginale  $f_{\mathbf{X}_j}$
- $\mathcal{M}_{\text{cop}} \Rightarrow$  Modèle d'apprentissage pour la densité de copule  $c_{\mathbf{U}}$

► Simuler un N-échantillon  $\mathbf{z}_{\text{obs}}$  à l'aide de  $G_{\mathbf{Z}}$

**pour chaque**  $j \in \{1, \dots, d\}$  **faire**

- Estimer  $\hat{f}_{\mathbf{X}_j}(\cdot | \mathcal{M}_j, \mathbf{z}_{\text{obs}}^{[j]})$
- Simuler un N-échantillon  $\mathbf{v}_{\text{obs}}$  à l'aide de  $G_{\mathbf{V}}$
- Estimer  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{v}_{\text{obs}})$
- Ré-assembler les composantes pour former la densité conjointe  $\hat{f}_{\mathbf{X}}(\cdot | \mathcal{M}, \mathbf{z}_{\text{obs}}, \mathbf{v}_{\text{obs}})$

**Sorties :**  $\hat{f}_{\mathbf{X}}(\cdot | \mathcal{M}, \mathbf{z}_{\text{obs}}, \mathbf{v}_{\text{obs}}) \Rightarrow$  Une observation de l'apprentissage de  $f_{\mathbf{X}}$

#### 4.2.2.4 Approximation de la variabilité d'échantillonnage

Lorsqu'on dispose d'un N-échantillon  $\mathbf{x}_{\text{obs}}$  d'observations d'une loi inconnue  $\mu_{\mathbf{X}}$ , et qu'on cherche à fabriquer une copie de ce N-échantillon sans faire d'hypothèse sur la loi sous-jacente, une méthode appelée *bootstrap* consiste à ré-échantillonner suivant la distribution empirique :

$$\hat{\mu}[\mathbf{x}_{\text{obs}}] = \frac{1}{N} \sum_{i=1}^N \delta_{\mathbf{x}_{\text{obs}}^{[i\bullet]}} , \quad (4.33)$$

où  $\delta_{\mathbf{x}}$  désigne la mesure de Dirac associée à  $\mathbf{x} \in \mathbb{R}^d$ . D'abord développée par EFRON [1982], cette technique est désormais très couramment utilisée en statistique. Bien que l'on sache que  $\mathbf{X}$  est absolument continu, on considère que les lignes  $\mathbf{x}_{\text{obs}}^{[i\bullet]}$  sont les seules valeurs qu'il peut prendre et qu'elles sont toutes équiprobables. Chaque observation  $\mathbf{x}_{\text{obs}}^{[i\bullet]}$  devient alors un atome de la loi discrète  $\hat{\mu}[\mathbf{x}_{\text{obs}}]$ . Pour simuler une nouvelle observation de  $\mathbf{X}$ , il suffit alors de faire un tirage uniforme (avec remise) d'une ligne de  $\mathbf{x}_{\text{obs}}$ . En répétant N fois cette opération, on fabrique un "clone"  $\tilde{\mathbf{x}}_{\text{obs}}$  de  $\mathbf{x}_{\text{obs}}$ . Une même ligne  $\mathbf{x}_{\text{obs}}^{[i\bullet]}$  peut être sélectionnée plusieurs fois et être insérée à plusieurs endroits de la version dupliquée  $\tilde{\mathbf{x}}_{\text{obs}}$ . Pour de plus amples détails sur le *bootstrap*, on pourra se référer à l'ouvrage de GOOD [2006] ou à celui de SHAO et TU [2012]. Le *bootstrap* est souvent utilisé (faute de mieux) pour évaluer la sensibilité d'une sortie à des paramètres d'entrées calculés à partir d'un échantillon  $\mathbf{x}_{\text{obs}}$ . Un exemple est fourni dans MONTGOMERY et collab. [2012] où la loi conjointe des paramètres d'une régression linéaire est simulée à l'aide d'une série de répliques *bootstrap* de la matrice d'incidence. Selon le même principe, le *bootstrap* peut servir à construire l'intervalle de confiance d'une quantité initialement estimée à l'aide d'un unique jeu de données  $\mathbf{x}_{\text{obs}}$ .

On cherche à utiliser le *bootstrap* afin d'imiter le fonctionnement des générateurs  $G_{\mathbf{X}}$ ,  $G_{\mathbf{Z}}$  et  $G_{\mathbf{V}}$  impliqués dans les Algorithmes 1 et 2. Pour imiter  $G_{\mathbf{X}}$ , il suffit d'utiliser la loi empirique  $\hat{\mu}[\mathbf{x}_{\text{obs}}]$ . De nouvelles observations de  $\mathbf{X}$  sont obtenues en tirant aléatoirement des lignes de  $\mathbf{x}_{\text{obs}}$ . Pour simuler  $\mathbf{Z}$ , dont les variables sont mutuellement indépendantes, on a besoin d'éliminer la structure de dépendance reliant les colonnes du N-échantillon  $\mathbf{x}_{\text{obs}}$  de  $\mathbf{X}$ . Pour cela, on crée autant de mesures empiriques qu'il y a de colonnes  $\mathbf{x}_{\text{obs}}^{[i\bullet j]}$  et on les laisse exister indépendamment les unes des autres. En pratique, il suffit de procéder à un ré-échantillonnage colonne par colonne de  $\mathbf{x}_{\text{obs}}$  pour imiter un N-échantillon  $\mathbf{z}_{\text{obs}}$  issu de  $\mathbf{Z}$ . Pour simuler  $\mathbf{V}$ , il suffit d'utiliser la loi empirique  $\hat{\mu}[\hat{\mathbf{u}}_{\text{obs}}]$  au sein d'une seconde boucle de ré-échantillonnage. Mathématiquement parlant, la loi conjointe  $\mu_{\mathbf{ZV}}$  est alors approchée par :

$$\hat{\mu}_{\mathbf{ZV}}[\mathbf{x}_{\text{obs}}] = \underbrace{\hat{\mu}[\mathbf{x}_{\text{obs}}^{[\bullet 1]}]}_{\text{pour } Z_1} \otimes \underbrace{\hat{\mu}[\mathbf{x}_{\text{obs}}^{[\bullet 2]}]}_{\text{pour } Z_2} \otimes \dots \otimes \underbrace{\hat{\mu}[\mathbf{x}_{\text{obs}}^{[\bullet d]}]}_{\text{pour } Z_d} \otimes \underbrace{\hat{\mu}[\hat{\mathbf{u}}_{\text{obs}}]}_{\text{pour } \mathbf{V}} . \quad (4.34)$$

Le mécanisme de ré-échantillonnage peut alors être vu comme l'application stochastique :

$$\mathcal{B}: \begin{array}{ccc} \mathbb{R}^{N \times d} & \longrightarrow & \mathbb{R}^N \times \dots \times \mathbb{R}^N \times [0, 1]^N \\ \mathbf{x}_{\text{obs}} & \longmapsto & \tilde{\mathbf{z}}_1, \dots, \tilde{\mathbf{z}}_d \quad \tilde{\mathbf{v}} \end{array} , \quad (4.35)$$

incluant les trois étapes suivantes :

(B1a) Pour  $1 \leq j \leq d$ , tirer un N-échantillon  $\tilde{\mathbf{z}}_j$  suivant  $\hat{\mu}[\mathbf{x}_{\text{obs}}^{[\bullet j]}]$ .

(B1b) Calculer la matrice des pseudo-observations  $\hat{\mathbf{u}}_{\text{obs}}$  en appliquant (SP2) à  $\mathbf{x}_{\text{obs}}$ .

(B1c) Tirer un N-échantillon  $\tilde{\mathbf{v}}$  suivant  $\hat{\mu}[\hat{\mathbf{u}}_{\text{obs}}]$ .

De façon complètement équivalente, on aurait pu remplacer les deux dernières étapes par :

(B1b) Tirer un N-échantillon  $\tilde{\mathbf{x}}_{\text{obs}}$  suivant  $\hat{\mu}[\mathbf{x}_{\text{obs}}]$ .

(B1c) Calculer  $\tilde{\mathbf{v}}$  en appliquant (SP2) à  $\tilde{\mathbf{x}}_{\text{obs}}$ .

La Figure 4.5 représente les manipulations matricielles associées à la première version de  $\mathcal{B}$ . On peut remarquer que cette procédure de ré-échantillonnage est en fait constituée de  $d + 1$  étapes

de *bootstrap* parallélisables. Les objets  $\tilde{\mathbf{Z}}_1, \dots, \tilde{\mathbf{Z}}_d$  et  $\tilde{\mathbf{V}}$  peuvent être considérés comme exactement indépendants car ils sont simulés suivant la loi empirique  $\hat{\mu}_{\mathbf{ZV}}[\mathbf{x}_{\text{obs}}]$  qui met en jeu des ré-échantillonnages déconnectés les uns des autres. La matrice  $\mathbf{x}_{\text{obs}}$  est figée et peut être vue comme un paramétrage de  $\mathcal{B}$ . Cette capacité à générer des échantillons d'apprentissage indépendants donne ensuite la possibilité de simuler les composantes fonctionnelles indépendantes. En effet, on s'est ramené à une situation conforme à l'hypothèse (4.29) dans laquelle :

$$[\hat{f}_{X_1}(\cdot | \mathcal{M}_1, \tilde{\mathbf{Z}}_1) \dots \hat{f}_{X_d}(\cdot | \mathcal{M}_d, \tilde{\mathbf{Z}}_d) \hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \tilde{\mathbf{V}})] \sim \mu_{\hat{f}_1} \otimes \dots \otimes \mu_{\hat{f}_d} \otimes \mu_{\hat{c}} \quad (4.36)$$

Le pseudo-code combinant  $\mathcal{B}$  et  $\mathcal{L}$  est donné dans l'Algorithme 3. On est ainsi en mesure, à partir du seul  $\mathbf{x}_{\text{obs}}$ , de fabriquer autant d'entrées indépendantes de l'application  $\Psi$  que souhaité.

<b>Algorithme 3</b> : Simulation des entrées de l'application $\Psi$ à partir de $\mathbf{x}_{\text{obs}}$	
$\mathbf{x}_{\text{obs}}$	$\Rightarrow$ Échantillon d'apprentissage initial
<b>Entrées :</b> $\mathcal{M}_j$	$\Rightarrow$ Modèle d'apprentissage pour la densité marginale $f_{X_j}$
$\mathcal{M}_{\text{cop}}$	$\Rightarrow$ Modèle d'apprentissage pour la densité de copule $c_{\mathbf{U}}$
<b>pour chaque</b> $j \in \{1, \dots, d\}$ <b>faire</b>	
▶	Extraire la $j$ -ème colonne $\mathbf{c}_j := \mathbf{x}_{\text{obs}}^{[\cdot j]}$
▶	Simuler un N-échantillon $\tilde{\mathbf{z}}_j$ suivant $\hat{\mu}[\mathbf{c}_j]$
▶	Estimer $\hat{f}_{X_j}(\cdot   \mathcal{M}_j, \tilde{\mathbf{z}}_j)$
▶	Fabriquer un N-échantillon de pseudo-observations $\hat{\mathbf{u}}_{\text{obs}}$ en renormalisant $\mathbf{x}_{\text{obs}}$ .
▶	Simuler un N-échantillon $\tilde{\mathbf{v}}$ à l'aide de $\hat{\mu}[\hat{\mathbf{u}}_{\text{obs}}]$ .
▶	Estimer $\hat{c}_{\mathbf{U}}(\cdot   \mathcal{M}_{\text{cop}}, \tilde{\mathbf{v}})$
▶	Ré-assembler les composantes pour former la densité conjointe $\hat{f}_{\mathbf{X}}(\cdot   \mathcal{M}, \tilde{\mathbf{z}}_1, \dots, \tilde{\mathbf{z}}_d, \tilde{\mathbf{v}})$
<b>Sorties :</b>	
$\hat{f}_{X_j}(\cdot   \mathcal{M}_j, \tilde{\mathbf{z}}_j)$	$\Rightarrow$ Une observation de l'apprentissage de $f_{X_j}$
$\hat{c}_{\mathbf{U}}(\cdot   \mathcal{M}_{\text{cop}}, \tilde{\mathbf{v}})$	$\Rightarrow$ Une observation de l'apprentissage de $c_{\mathbf{U}}$
$\hat{f}_{\mathbf{X}}(\cdot   \mathcal{M}, \tilde{\mathbf{z}}_1, \dots, \tilde{\mathbf{z}}_d, \tilde{\mathbf{v}})$	$\Rightarrow$ Une observation de l'apprentissage de $f_{\mathbf{X}}$

## 4.2.3 Implémentation de l'algorithme d'ASF

### 4.2.3.1 Justification du choix des indices de sensibilité

On rappelle que l'on cherche à analyser la sensibilité de l'application (4.12) dont les entrées fonctionnelles vérifient l'hypothèse (4.29) si elles sont simulées avec l'Algorithme 3. Dès lors que l'on est assuré de l'indépendance entre les entrées, l'utilisation des indices de Sobol semble être la meilleure alternative. Leur interprétation est simple et il existe une large gamme de techniques pour les estimer. Par ailleurs, ils sont très permissifs quant à la nature des objets dont ils peuvent quantifier l'influence. Toute application  $\Psi : E_1 \times \dots \times E_r \rightarrow \mathbb{R}$  peut convenir à condition que l'espace d'entrée soit mesurable et que la variable de sortie soit de carré intégrable. Il est d'ailleurs montré dans [GAMBOA et collab. \[2014\]](#) que la notion d'indice de Sobol se généralise facilement au cas où la sortie de  $\Psi$  existe dans un espace de Hilbert. Des travaux comme ceux conduits par [IOOSS et RIBATET \[2009\]](#) ou [FRUTH et collab. \[2015\]](#) mettent en avant l'utilisation des indices de Sobol pour résoudre des problèmes d'analyse de sensibilité impliquant des variables fonctionnelles. Dans le

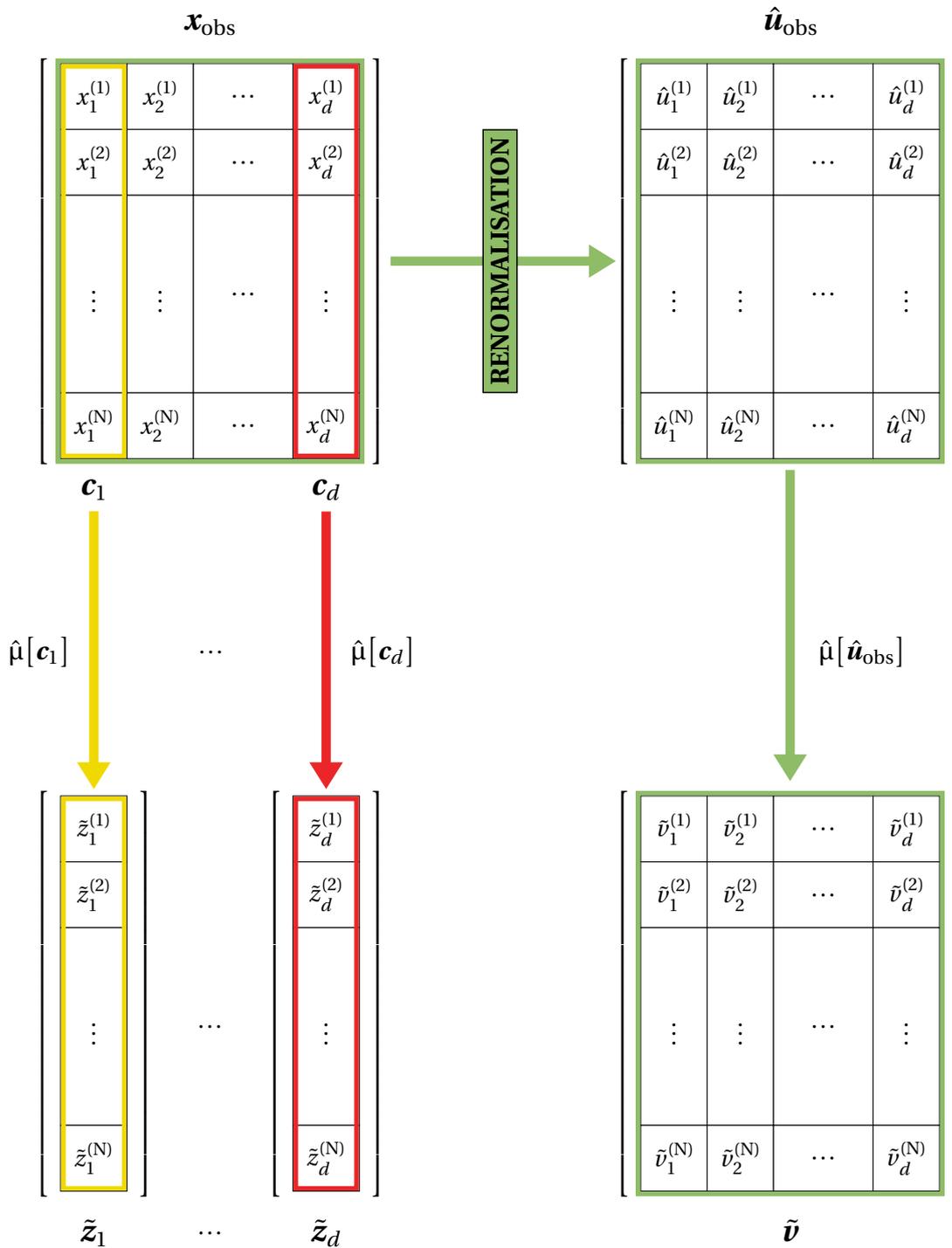


FIGURE 4.5 – Mécanisme de ré-échantillonnage  $\mathcal{B}$  permettant d'imiter les générateurs  $G_Z$  et  $G_V$ .

contexte d’ASF propre à ce travail, les indices de Sobol du premier ordre sont alors donnés par :

$$S_j := \frac{\mathbb{V} \left( \mathbb{E} \left[ P_f(\mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}}) \mid \hat{f}_{X_j}(\cdot \mid \mathcal{M}_j, \mathbf{Z}_{\text{obs}}^{[j]}) \right] \right)}{\mathbb{V} (P_f(\mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}}))}, \quad (4.37)$$

$$S_{\text{cop}} := \frac{\mathbb{V} \left( \mathbb{E} \left[ P_f(\mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}}) \mid \hat{c}_{\mathbf{U}}(\cdot \mid \mathcal{M}_{\text{cop}}, \mathbf{V}_{\text{obs}}) \right] \right)}{\mathbb{V} (P_f(\mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}}))}. \quad (4.38)$$

Comme on a pu le voir au Chapitre 3, les indices de Sobol peuvent être estimés avec une multitude de schémas numériques. Tous sont compatibles avec la procédure d’ASF développée dans cette section. Pour autant, les simulations numériques qui seront conduites dans la suite feront principalement appel au schéma *Pick-Freeze* dont les propriétés asymptotiques sont détaillées dans JANON et collab. [2014] et GAMBOA et collab. [2014]. Dans certaines situations, on adaptera le schéma *Correlation 2* imaginé par OWEN [2013] afin d’améliorer la précision de l’estimation des petits indices.

#### 4.2.3.2 Assemblage des différentes briques de calcul

À partir du jeu de données initial  $\mathbf{x}_{\text{obs}}$ , l’obtention d’un appariement entrées-sortie de l’application  $\Psi$  passe par l’exécution de trois étapes de calcul :

$$\mathbf{x}_{\text{obs}} \xrightarrow{\mathcal{B}} \{\bar{\mathbf{z}}, \bar{\mathbf{v}}\} \xrightarrow{\mathcal{L}} \{\hat{f}_{X_1}(\cdot \mid \bar{\mathbf{z}}_1), \dots, \hat{f}_{X_d}(\cdot \mid \bar{\mathbf{z}}_d), \hat{c}_{\mathbf{U}}(\cdot \mid \bar{\mathbf{v}})\} \xrightarrow{\text{Sklar}} \hat{f}_{\mathbf{X}}(\cdot \mid \bar{\mathbf{z}}, \bar{\mathbf{v}}) \xrightarrow{\mathcal{R}} \hat{P}_f(\bar{\mathbf{z}}, \bar{\mathbf{v}}), \quad (4.39)$$

$$\{\hat{f}_{X_1}(\cdot \mid \bar{\mathbf{z}}_1), \dots, \hat{f}_{X_d}(\cdot \mid \bar{\mathbf{z}}_d), \hat{c}_{\mathbf{U}}(\cdot \mid \bar{\mathbf{v}})\} \xrightarrow{\Psi} \hat{P}_f(\bar{\mathbf{z}}, \bar{\mathbf{v}})$$

qui correspondent respectivement à :

1. Un ré-échantillonnage  $\mathcal{B}$  :  $\mathbb{R}^{N \times d} \rightarrow \mathbb{R}^N \times \dots \times \mathbb{R}^N \times [0, 1]^{N \times d}$ ,
2. Un apprentissage  $\mathcal{L}$  :  $\mathbb{R}^N \times \dots \times \mathbb{R}^N \times [0, 1]^{N \times d} \rightarrow \mathbb{L}^2(\mathbb{R}) \times \dots \times \mathbb{L}^2(\mathbb{R}) \times \mathbb{L}^2(\mathcal{H}_d)$ ,
3. Une analyse de fiabilité  $\mathcal{R}$  :  $\mathbb{L}^2(\mathbb{R}) \times \dots \times \mathbb{L}^2(\mathbb{R}) \times \mathbb{L}^2(\mathcal{H}_d) \rightarrow [0, 1]$ .

Le *workflow* résultant de l’enchaînement de ces différentes briques de calcul est représenté à la Figure 4.6 où apparaît également le schéma d’estimation  $\mathcal{A}$  des indices de sensibilité. La collecte des données du retour d’expérience permet de constituer l’échantillon d’apprentissage  $\mathbf{x}_{\text{obs}}$ . C’est une étape préalable, considérée comme fixée pour toute cette étude. Ensuite, une exécution séquentielle de  $\mathcal{B}$ ,  $\mathcal{L}$  et  $\mathcal{R}$  permet d’obtenir autant d’échantillons entrées-sortie de  $\Psi$  qu’en exige le schéma d’estimation  $\mathcal{A}$ . En termes de coût de calcul, les tâches de clonage *bootstrap* de  $\mathcal{B}$  ainsi que les tâches d’inférence statistique de  $\mathcal{L}$  sont peu contraignantes. En début de section, il a été supposé que  $\mathcal{R}$  n’implique aucune charge de calcul. Cette “gratuité” vaut donc aussi pour l’application  $\Psi$  qui encapsule  $\mathcal{R}$  dans le contexte d’analyse de sensibilité qui est le nôtre. Toutefois, il faut garder à l’esprit que c’est précisément cette dernière phase de calcul qui est susceptible d’occasionner les surcoûts les plus conséquents si on s’émancipe des hypothèses initiales.

Pour la plupart des indices de sensibilité, les schémas d’estimation les plus efficaces sont ceux qui permettent le calcul des indices à partir d’un M-échantillon d’entrées et du M-échantillon des sorties associées. Seules M évaluations de  $\Psi$  (c’est-à-dire M appels à  $\mathcal{R}$ ) sont ainsi nécessaires pour obtenir une estimation de la collection désirée d’indices. Dans le cas des indices de Sobol, un schéma de ce type a été récemment développé par GAMBOA et collab. [2020] en utilisant la notion de corrélation de CHATTERJEE [2020]. Toutefois, cette méthode exige que chaque espace d’entrée  $E_k$  de l’application  $\Psi : E_1 \times \dots \times E_r \rightarrow \mathbb{R}$  puisse être muni d’une relation d’ordre. Ce n’est pas possible dans le cas de l’application (4.12) car les espaces  $E_1 = \dots = E_d = \mathbb{L}^2(\mathbb{R})$  et  $E_{d+1} = \mathbb{L}^2(\mathcal{H}_d)$  sont fonctionnels. L’estimation des indices de Sobol nécessite alors l’utilisation de schémas d’estimation plus coûteux, à l’instar du *Pick-Freeze*.

Il a été vu à la Section 3.2.2 que la méthode du *Pick-Freeze* s’appuie sur une réécriture des indices de Sobol après duplication des entrées de  $\Psi$ . À partir de deux lots indépendants  $(\mathbf{Z}_{\text{obs}}^A, \mathbf{V}_{\text{obs}}^A)$  et

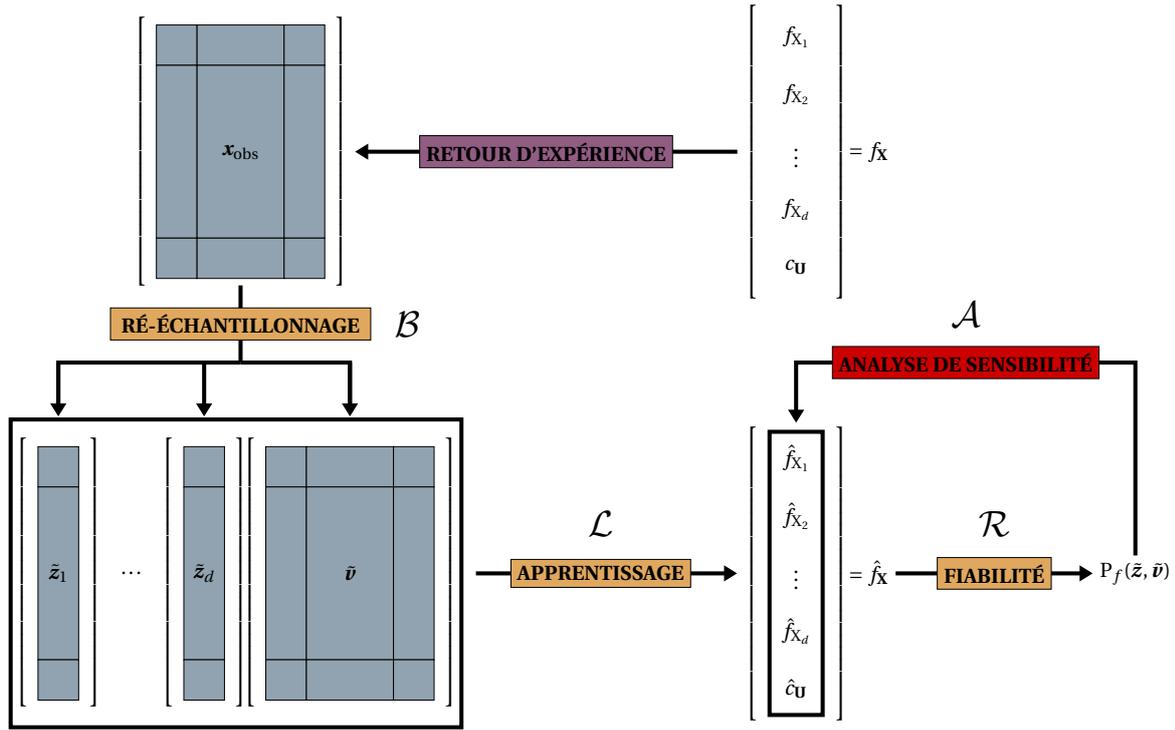


FIGURE 4.6 – Illustration des différentes briques de calcul intervenant dans l’algorithme d’ASF.

$(\mathbf{Z}_{\text{obs}}^B, \mathbf{V}_{\text{obs}}^B)$  d’échantillons d’apprentissage, on obtient deux séquences indépendantes de composantes fonctionnelles. Celles-ci sont alors désassemblées pour constituer un tableau  $T_{AB}$  composé de  $d + 2$  entrées recombinaées. Sur la Figure 4.7, on note  $\mathcal{P}$  l’opérateur permettant la construction de  $T_{AB}$  à partir des deux séquences d’entrées de  $\Psi$ . Sur chaque ligne du tableau à l’exception de la première, il consiste à “geler” une composante de la première séquence puis à “piocher” dans la seconde séquence pour compléter la ligne. En vertu du théorème de Sklar, chaque ligne donne naissance à une densité conjointe après ré-assemblage des différentes composantes fonctionnelles. Les probabilités de défaillance associées à ces densités permettent d’exprimer les indices de Sobol (4.37) et (4.38) sans utiliser les espérances conditionnelles :

$$S_1 = \frac{\text{Cov}\left(\mathbf{p}_f^{A\dots A}, \mathbf{p}_f^{AB\dots B}\right)}{\mathbb{V}\left(\mathbf{p}_f^{A\dots A}\right)} \quad \dots \quad S_d = \frac{\text{Cov}\left(\mathbf{p}_f^{A\dots A}, \mathbf{p}_f^{B\dots BAB}\right)}{\mathbb{V}\left(\mathbf{p}_f^{A\dots A}\right)} \quad \text{et} \quad S_{\text{cop}} = \frac{\text{Cov}\left(\mathbf{p}_f^{A\dots A}, \mathbf{p}_f^{B\dots BA}\right)}{\mathbb{V}\left(\mathbf{p}_f^{A\dots A}\right)}. \quad (4.40)$$

Les estimateurs se construisent alors naturellement à partir des variances et des covariances empiriques. Si on souhaite faire intervenir  $M$  termes dans le calcul de ces quantités, l’estimation de tous les indices de Sobol du premier ordre représente un coût total de  $M \times (d + 2)$  appels à  $\mathcal{R}$ . En pratique, le schéma *Pick-Freeze* présente certaines limites, notamment lorsque la quasi-totalité des indices est de petite taille, c’est-à-dire lorsqu’ils sont tous compris dans un sous-intervalle  $[0, t] \subset [0, 1]$  où  $t$  est très petit devant 1. D’une part, la formule (4.40) ne garantit pas la positivité des petits indices. D’autre part, le manque de précision sur ces petits indices empêche souvent l’utilisateur d’établir un classement robuste alors même qu’il serait fortement souhaitable de trouver des résultats sensiblement identiques d’une exécution de l’algorithme d’ASF à l’autre.

Le schéma d’estimation *Correlation 2* apporte davantage de précision dans l’estimation des petits indices mais sans toutefois éradiquer complètement le risque de l’apparition d’une valeur négative au sein de la collection d’indices. Dans le même esprit que le *Pick-Freeze*, l’idée de *Correlation 2* est de simuler trois collections d’entrées puis de construire un tableau  $T_{ABC}$  constitué de

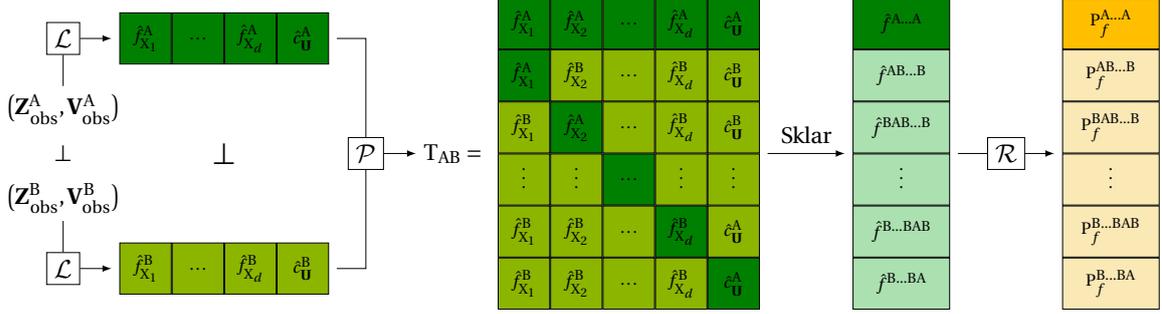


FIGURE 4.7 – Schématisation du principe de *Pick-Freeze* pour l'application  $\Psi$ .

$2d + 3$  entrées recombinaées. On a alors :

$$S_1 = \frac{\mathbb{E} \left[ \left( p_f^{A...A} - p_f^{CA...A} \right) \left( p_f^{AB...B} - p_f^{B...B} \right) \right]}{\mathbb{V} \left( p_f^{A...A} \right)} \quad \dots \quad S_d = \frac{\mathbb{E} \left[ \left( p_f^{A...A} - p_f^{A...ACA} \right) \left( p_f^{B...BAB} - p_f^{B...B} \right) \right]}{\mathbb{V} \left( p_f^{A...A} \right)} \quad (4.41)$$

$$\text{et } S_{\text{cop}} = \frac{\mathbb{E} \left[ \left( p_f^{A...A} - p_f^{A...AC} \right) \left( p_f^{B...BA} - p_f^{B...B} \right) \right]}{\mathbb{V} \left( p_f^{A...A} \right)} .$$

Le coût d'estimation de tous les indices du premier ordre est alors de  $M \times (2d + 3)$  appels à  $\mathcal{R}$ . Ce schéma prend toute son utilité lorsque  $d$  augmente et que le nombre de petits indices s'accroît.

Pour conclure sur l'assemblage des briques de cet algorithme, on peut voir qu'il est composé de quatre parties  $\mathcal{B}$ ,  $\mathcal{L}$ ,  $\mathcal{R}$  et  $\mathcal{A}$ . Le mécanisme de ré-échantillonnage  $\mathcal{B}$  permet de simuler des échantillons d'apprentissage indépendants à partir de la connaissance de  $\mathbf{x}_{\text{obs}}$ . L'algorithme d'apprentissage  $\mathcal{L}$  est choisi par l'utilisateur qui veut analyser la sensibilité de son processus d'évaluation des risques à de tels choix de modélisation. La phase  $\mathcal{R}$ , supposée déterministe, permet un calcul exact de la probabilité de défaillance à partir de la densité apprise. C'est sur cette partie qu'on réfléchira dans la Section 4.4 pour envisager l'extension de l'algorithme d'ASF à des contextes pour lesquels aucun algorithme déterministe  $\mathcal{R}$  ne peut être déployé et pour lesquels il serait vain de croire que l'on peut répéter  $k \times M$  analyses de fiabilité avec  $k \in \{d + 2, 2d + 3\}$  pour les deux schémas d'estimation vus précédemment. Enfin, le choix du schéma  $\mathcal{A}$  est laissé libre. Pour des cas simples, on peut sereinement envisager que les résultats en seront indépendants. Pour des problèmes d'analyse de sensibilité plus complexes, où la robustesse des classements n'est plus assurée, il faudra accorder plus d'importance à la mise en place d'un schéma d'estimation permettant d'atteindre un compromis entre précision et budget de calcul.

#### 4.2.4 Adaptation du mécanisme de ré-échantillonnage pour conduire une ASF après assimilation de données supplémentaires

Imaginons que l'on ait exécuté l'algorithme d'ASF pour un échantillon  $\mathbf{x}_{\text{obs}}$  donné et un modèle d'apprentissage  $\mathcal{M}$  fixé. On dispose alors d'une collection d'indices  $\{\hat{S}_1, \dots, \hat{S}_d, \hat{S}_{\text{cop}}\}$  rendant compte de l'impact de chacune des composantes fonctionnelles sur la variabilité de la prédiction du risque de défaillance. On aimerait vérifier que les indices calculés rendent effectivement compte de la propagation de la variabilité d'apprentissage à travers  $\Psi$ . Avec l'inégalité triangulaire, on peut remarquer que :

$$\left| P_f(\mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}}) - P_f \right| = \left| \int_{\mathcal{D}_f} \left( \hat{f}_{\mathbf{X}}(\mathbf{x} \mid \mathcal{M}, \mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}}) - f_{\mathbf{X}}(\mathbf{x}) \right) d\mathbf{x} \right|, \quad (4.42)$$

$$\leq \left\| \hat{f}_{\mathbf{X}}(\cdot \mid \mathcal{M}, \mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}}) - f_{\mathbf{X}} \right\|_{L^1(\mathcal{D}_f)} .$$

Si le modèle d'apprentissage  $\mathcal{M}$  est choisi raisonnablement, lorsqu'on augmente la taille  $N$  des échantillons  $\mathbf{Z}_{\text{obs}}$  et  $\mathbf{V}_{\text{obs}}$  utilisés pour l'apprentissage, il faut s'attendre à ce que la densité estimée  $\hat{f}_{\mathbf{X}}(\cdot | \mathcal{M}, \mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}})$  approche avec plus de précision et moins de variabilité la densité théorique  $f_{\mathbf{X}}$ . Cela sera vrai partout sur  $\mathbb{R}^d$  mais seule la qualité de l'estimation sur  $D_f$  importe ici. Ainsi, en rajoutant des données supplémentaires, on peut s'attendre à ce que la distribution de  $P_f(\mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}})$  soit davantage resserrée autour de la valeur théorique  $P_f$ . Comme il est souvent difficile de collecter des observations supplémentaires de la loi conjointe  $\mu_{\mathbf{ZV}}$ , on peut essayer de collecter des données pour une seule loi marginale  $\mu_{Z_j}$ . On appellera cette stratégie "acquisition ciblée". En pratique, les données récupérées émanent bien sûr de  $\mu_{X_j}$ . Sous réserve que l'algorithme d'ASF précédemment développé donne les résultats escomptés, deux tendances "extrêmes" sont alors attendues.

- (a) Si  $\hat{S}_j \gg \hat{S}_k$  pour tout  $k \neq j$  :
  - $\Rightarrow \hat{f}_{X_j}(\cdot | \mathcal{M}_j, \mathbf{Z}_{\text{obs}}^{[j]})$  véhicule une très grande part de la variabilité de  $P_f(\mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}})$ .
  - $\Rightarrow$  Trouver d'autres observations de  $Z_j$  permet de réduire la dispersion de  $P_f(\mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}})$ .
- (b) Si  $\hat{S}_j \ll \hat{S}_k$  pour tout  $k \neq j$  :
  - $\Rightarrow \hat{f}_{X_j}(\cdot | \mathcal{M}_j, \mathbf{Z}_{\text{obs}}^{[j]})$  véhicule une toute petite part de la variabilité de  $P_f(\mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}})$ .
  - $\Rightarrow$  Trouver d'autres observations de  $Z_j$  n'a aucun effet sur la dispersion de  $P_f(\mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}})$ .

Pour vérifier ces intuitions, on souhaite construire une seconde procédure d'ASF à partir de  $\mathcal{M}$ ,  $\mathbf{x}_{\text{obs}}$  et des observations supplémentaires de la variable  $X_j$  sélectionnée. L'objectif est d'obtenir une seconde collection d'indices de Sobol et de la comparer à la première. Dans le cas (a), l'ajout de nouvelles données doit conduire à une rétrogradation de l'indice  $\hat{S}_j$  car la proportion de variabilité engendrée par l'apprentissage de  $f_{X_j}$  chute suite à l'assimilation des données supplémentaires. Si les données apportées sont suffisamment nombreuses pour permettre une neutralisation efficace de la composante initialement identifiée comme la plus influente, une autre composante fonctionnelle émerge alors en tant que contributrice principale de la variabilité d'apprentissage restante. Dans le cas (b), étant donné que  $\hat{S}_j$  est le plus petit indice, on peut s'attendre à ce que l'ajout de nouvelles données abaisse encore un peu plus  $\hat{S}_j$  sans toutefois perturber les autres classements. Pour implémenter cette seconde ASF, il suffit de réutiliser les outils mis en place pour effectuer la première ASF. Cependant, le mécanisme de ré-échantillonnage  $\mathcal{B}$  doit être adapté au contexte de l'acquisition ciblée.

#### 4.2.4.1 Adaptation dans le cas de données supplémentaires pour une loi marginale

Les données supplémentaires sont modélisées par une matrice aléatoire  $\mathbf{X}_{\text{extra}}$  constituée de  $N'$  observations supplémentaires de la loi conjointe  $\mu_{\mathbf{X}}$ . Par ailleurs, notons  $\mathbf{X}_{\text{tot}}$  la matrice obtenue virtuellement en rassemblant  $\mathbf{X}_{\text{obs}}$  et  $\mathbf{X}_{\text{extra}}$ . Dans le cadre de l'acquisition ciblée de nouvelles données pour la variable  $X_j$ , tout fonctionne comme si on ne disposait que d'une réalisation  $x_{\text{extra}}^{[j]}$  de la  $j$ -ème colonne de la matrice  $\mathbf{X}_{\text{extra}}$ . Il faut alors chercher à reproduire le fonctionnement des générateurs  $G_{\mathbf{Z}}$  et  $G_{\mathbf{V}}$  en ne se basant que sur la connaissance de  $\mathbf{x}_{\text{obs}}$  et  $x_{\text{extra}}^{[j]}$ . Pour  $G_{\mathbf{Z}}$ , il suffit d'effectuer des ré-échantillonnages *bootstrap* indépendants à partir des observations marginales de chaque variable. On distingue deux cas :

- Pour  $X_j \Rightarrow$  on dispose des  $N + N'$  observations de  $x_{\text{tot}}^{[j]} = \mathbf{x}_{\text{obs}}^{[j]} \cup x_{\text{extra}}^{[j]} = \left[ a_j^{(i)} \right]_{1 \leq i \leq N+N'}$ .
- Pour les autres  $X_k \Rightarrow$  on se contente des  $N$  observations initiales de  $\mathbf{x}_{\text{obs}}^{[k]} = \left[ x_k^{(i)} \right]_{1 \leq i \leq N}$ .

Pour  $G_{\mathbf{V}}$ , l'idée est à nouveau d'approcher  $\mu_{\mathbf{V}}$  par la mesure empirique  $\hat{\mu}[\hat{\mathbf{u}}_{\text{obs}}]$  et de réaliser des tirages *bootstrap* parmi l'ensemble des pseudo-observations de la copule. Toutefois, comme on dispose de données supplémentaires  $x_{\text{extra}}^{[j]}$  pour la variable  $X_j$ , la renormalisation permettant de construire  $\hat{\mathbf{u}}_{\text{obs}}$  est légèrement différente. En effet, pour transformer la  $j$ -ème colonne de  $\mathbf{x}_{\text{obs}}$ , on applique la fonction de répartition empirique construite à partir de  $\mathbf{x}_{\text{tot}}^{[j]}$  pour impliquer l'information apportée par les observations supplémentaires. La matrice  $\hat{\mathbf{u}}_{\text{obs}}$  est alors calculée à l'aide

des deux formules suivantes :

$$\hat{u}_j^{(i)} = \frac{1}{N+N'+1} \sum_{l=1}^{N+N'} \mathbb{1}_{\{a_j^{(l)} \leq x_j^{(i)}\}} \quad \text{et} \quad \hat{u}_k^{(i)} = \frac{1}{N+1} \sum_{l=1}^N \mathbb{1}_{\{x_k^{(l)} \leq x_k^{(i)}\}}. \quad (4.43)$$

Finalement, la loi conjointe  $\mu_{ZV}$  est approchée par :

$$\hat{\mu}_{ZV} \left[ \mathbf{x}_{\text{obs}} \cup \mathbf{x}_{\text{extra}}^{[\bullet j]} \right] = \bigotimes_{k \neq j} \underbrace{\hat{\mu} \left[ \mathbf{x}_{\text{obs}}^{[\bullet k]} \right]}_{\text{pour tout } Z_k} \otimes \underbrace{\hat{\mu} \left[ \mathbf{x}_{\text{tot}}^{[\bullet j]} \right]}_{\text{pour } Z_j} \otimes \underbrace{\hat{\mu} \left[ \hat{\mathbf{u}}_{\text{obs}} \right]}_{\text{pour } \mathbf{V}}. \quad (4.44)$$

La nouvelle technique de ré-échantillonnage est notée  $\mathcal{B}_j$  et s'articule comme suit :

$$\mathcal{B}_j: \begin{array}{ccccccc} \mathbb{R}^{N \times d} & \times & \mathbb{R}^{N'} & \longrightarrow & \mathbb{R}^N \times \dots \times \mathbb{R}^N & \times & \mathbb{R}^{N+N'} & \times & \mathbb{R}^N \times \dots \times \mathbb{R}^N & \times & [0, 1]^N \\ \mathbf{x}_{\text{obs}} & & \mathbf{x}_{\text{extra}}^{[\bullet j]} & \longmapsto & \tilde{\mathbf{z}}_1, \dots, \tilde{\mathbf{z}}_{j-1} & & \tilde{\mathbf{z}}_j & & \tilde{\mathbf{z}}_{j+1}, \dots, \tilde{\mathbf{z}}_d & & \tilde{\mathbf{v}} \end{array}. \quad (4.45)$$

Un appel à cette application stochastique donne lieu à la séquence de tâches listée ci-dessous :

- (B2a) Tirer un  $(N + N')$ -échantillon  $\tilde{\mathbf{z}}_j$  suivant  $\hat{\mu} \left[ \mathbf{x}_{\text{tot}}^{[\bullet j]} \right]$ .
- (B2b) Pour  $k \neq j$ , tirer un  $N$ -échantillon  $\tilde{\mathbf{z}}_k$  suivant  $\hat{\mu} \left[ \mathbf{x}_{\text{obs}}^{[\bullet k]} \right]$ .
- (B2c) Calculer la matrice des pseudo-observations  $\hat{\mathbf{u}}_{\text{obs}}$  avec les formules (4.43).
- (B2d) Tirer un  $N$ -échantillon  $\tilde{\mathbf{v}}$  suivant  $\hat{\mu} \left[ \hat{\mathbf{u}}_{\text{obs}} \right]$ .

On pourra se référer à la Figure 4.8 pour une représentation schématique des manipulations matricielles dans le cas où  $j = 1$ . Pour mener la procédure d'ASF, il suffit de répéter un nombre  $M$  de ré-échantillonnages  $\mathcal{B}_j$  puis d'exécuter les briques de calcul  $\mathcal{L}$ ,  $\mathcal{R}$  et  $\mathcal{A}$  exactement comme indiqué dans la Section 4.2.3.2 pour la procédure initiale d'ASF.

#### 4.2.4.2 Adaptation dans le cas de données supplémentaires pour la copule

Le mécanisme de ré-échantillonnage  $\mathcal{B}_j$  permet d'incorporer des observations supplémentaires  $x_{\text{extra}}^{[\bullet j]}$  de la variable  $X_j$ . L'échantillon  $\tilde{\mathbf{Z}}_j$  fabriqué par  $\mathcal{B}_j$  comporte  $N + N'$  observations alors que les autres échantillons créés ne sont constitués que de  $N$  observations. Cela doit conduire, au cours de la seconde ASF, à l'obtention d'une collection d'indices de Sobol mettant en exergue la réduction de la variabilité d'apprentissage véhiculée par  $\hat{f}_{X_j}(\cdot | \mathcal{M}_j, \mathbf{Z}_{\text{obs}}^{[\bullet j]})$ .

Par analogie, on souhaite construire un mécanisme  $\mathcal{B}_{\text{cop}}$  permettant de prendre en compte un plus grand nombre d'observations de la copule. L'objectif est alors de vérifier que l'impact de  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{V}_{\text{obs}})$  peut lui aussi être réduit par l'ajout de données. L'acquisition de données supplémentaires pour la seule copule est une situation qui n'a pas de justification pratique. En effet, toute observation supplémentaire de la copule provient d'une observation supplémentaire de la loi conjointe et peut donc aussi servir à affiner la connaissance des lois marginales. Tout se passe comme si on disposait d'une réalisation complète  $\mathbf{x}_{\text{extra}}$  de la matrice  $\mathbf{X}_{\text{extra}}$  mais qu'on faisait le choix de restreindre l'utilisation de cette information supplémentaire à l'amélioration de la connaissance de la structure de dépendance.

La technique de ré-échantillonnage  $\mathcal{B}_{\text{cop}}$  se déduit facilement de ce qui a été fait pour  $\mathcal{B}$ . Pour les échantillons d'apprentissage  $\tilde{\mathbf{Z}}_j$  des lois marginales, on effectue des tirages *bootstrap* sur les colonnes de  $\mathbf{x}_{\text{obs}}$ . Pour l'échantillon d'apprentissage  $\tilde{\mathbf{V}}$ , des répliques *bootstrap* sont simulées à partir d'une version enrichie  $\hat{\mathbf{u}}_{\text{tot}}$  de la matrice des pseudo-observations. Pour construire  $\hat{\mathbf{u}}_{\text{tot}}$ , il suffit d'effectuer une renormalisation de  $\mathbf{x}_{\text{tot}}$  avec les fonctions de répartition empiriques. Elles incorporent  $N + N'$  atomes et permettent de ramener  $\mathbf{x}_{\text{tot}}$  dans  $\mathcal{H}_d$ . On a ainsi :

$$\mathbf{x}_{\text{tot}} = \left[ a_j^{(i)} \right]_{\substack{1 \leq i \leq N+N' \\ 1 \leq j \leq d}} \quad \text{et} \quad \hat{\mathbf{u}}_{\text{tot}} = \left[ \hat{u}_j^{(i)} \right]_{\substack{1 \leq i \leq N+N' \\ 1 \leq j \leq d}} \quad \text{avec} \quad \hat{u}_j^{(i)} = \frac{1}{N+N'+1} \sum_{l=1}^{N+N'} \mathbb{1}_{\{a_j^{(l)} \leq a_j^{(i)}\}}. \quad (4.46)$$

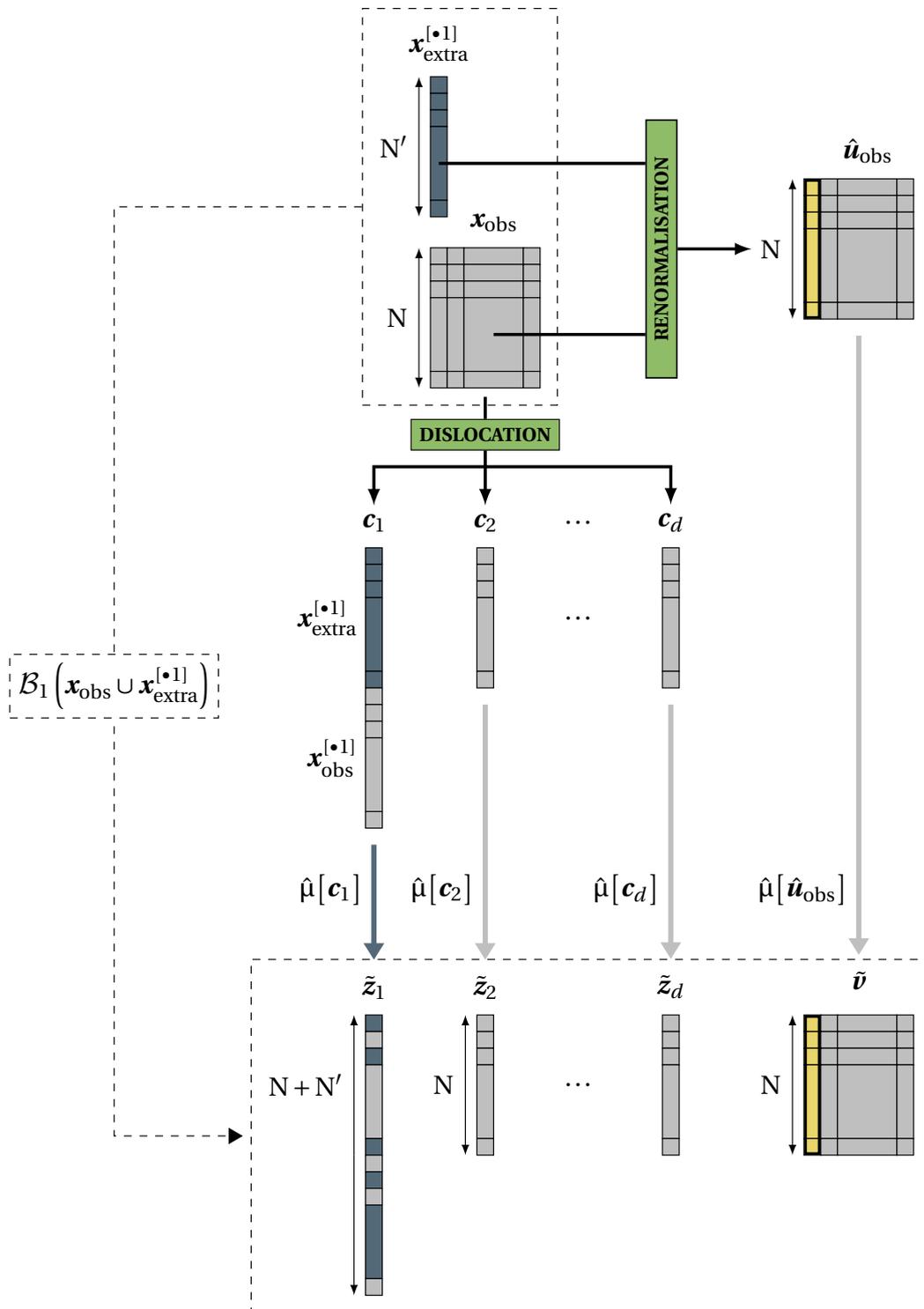


FIGURE 4.8 – Mécanisme de ré-échantillonnage  $\mathcal{B}_1$  en présence d'observations supplémentaires de  $X_1$ .

La loi conjointe  $\mu_{ZV}$  est alors approchée par :

$$\hat{\mu}_{ZV}[\mathbf{x}_{\text{obs}} \cup \mathbf{x}_{\text{extra}}] = \underbrace{\hat{\mu}[\mathbf{x}_{\text{obs}}^{[\bullet 1]}]}_{\text{pour } Z_1} \otimes \dots \otimes \underbrace{\hat{\mu}[\mathbf{x}_{\text{obs}}^{[\bullet d]}]}_{\text{pour } Z_d} \otimes \underbrace{\hat{\mu}[\hat{\mathbf{u}}_{\text{tot}}]}_{\text{pour } V}. \quad (4.47)$$

Le mécanisme de ré-échantillonnage  $\mathcal{B}_{\text{cop}}$  prend la forme suivante :

$$\mathcal{B}_{\text{cop}} : \begin{array}{ccc} \mathbb{R}^{N \times d} & \times & \mathbb{R}^{N' \times d} & \longrightarrow & \mathbb{R}^N \times \dots \times \mathbb{R}^N & \times & [0, 1]^{N+N'} \\ \mathbf{x}_{\text{obs}} & & \mathbf{x}_{\text{extra}} & \longmapsto & \tilde{\mathbf{z}}_1, \dots, \tilde{\mathbf{z}}_d & & \tilde{\mathbf{v}} \end{array}, \quad (4.48)$$

et s'exécute selon la séquence décrite ci-dessous :

- (B3a) Pour  $1 \leq j \leq d$ , tirer un N-échantillon  $\tilde{\mathbf{z}}_j$  suivant  $\hat{\mu}[\mathbf{x}_{\text{obs}}^{[\bullet j]}]$ .
- (B3b) Calculer la matrice des pseudo-observations  $\hat{\mathbf{u}}_{\text{tot}}$  avec la formule (4.46).
- (B3c) Tirer un N-échantillon  $\tilde{\mathbf{v}}$  suivant  $\hat{\mu}[\hat{\mathbf{u}}_{\text{tot}}]$ .

La Figure 4.9 propose un schéma des manipulations matricielles associées à  $\mathcal{B}_{\text{cop}}$ . Comme on avait déjà pu le faire remarquer pour  $\mathcal{B}_j$ , la procédure d'ASF est ensuite exactement calquée sur ce qui est fait dans la procédure d'ASF initiale.

Dans cette section, on a mis en place des adaptations  $\mathcal{B}_j$  et  $\mathcal{B}_{\text{cop}}$  du mécanisme initial  $\mathcal{B}$ . Ces versions adaptées permettent de conduire une seconde ASF après une acquisition ciblée de données, c'est-à-dire après la collecte d'observations supplémentaires qui vont être exclusivement réservées à la réduction de la variabilité d'apprentissage d'une des composantes fonctionnelles. Les deux collections d'indices de Sobol ont vocation à être comparées et doivent mettre en évidence l'impact de l'assimilation des nouvelles données. La lisibilité de ce phénomène sera d'autant plus marquée en présence de facteurs facilitant la réduction de la variabilité d'apprentissage tels que :

- (a) Le rang de l'indice  $\hat{S}_j$  au sein de la première ASF
- (b) Le nombre  $N'$  de données supplémentaires collectées pour constituer  $X_{\text{extra}}^{[\bullet j]}$ .
- (c) La vitesse avec laquelle  $\mathcal{M}_j$  fait converger  $\hat{f}_{X_j}$  vers  $f_{X_j}$  à budget fixé d'observations.

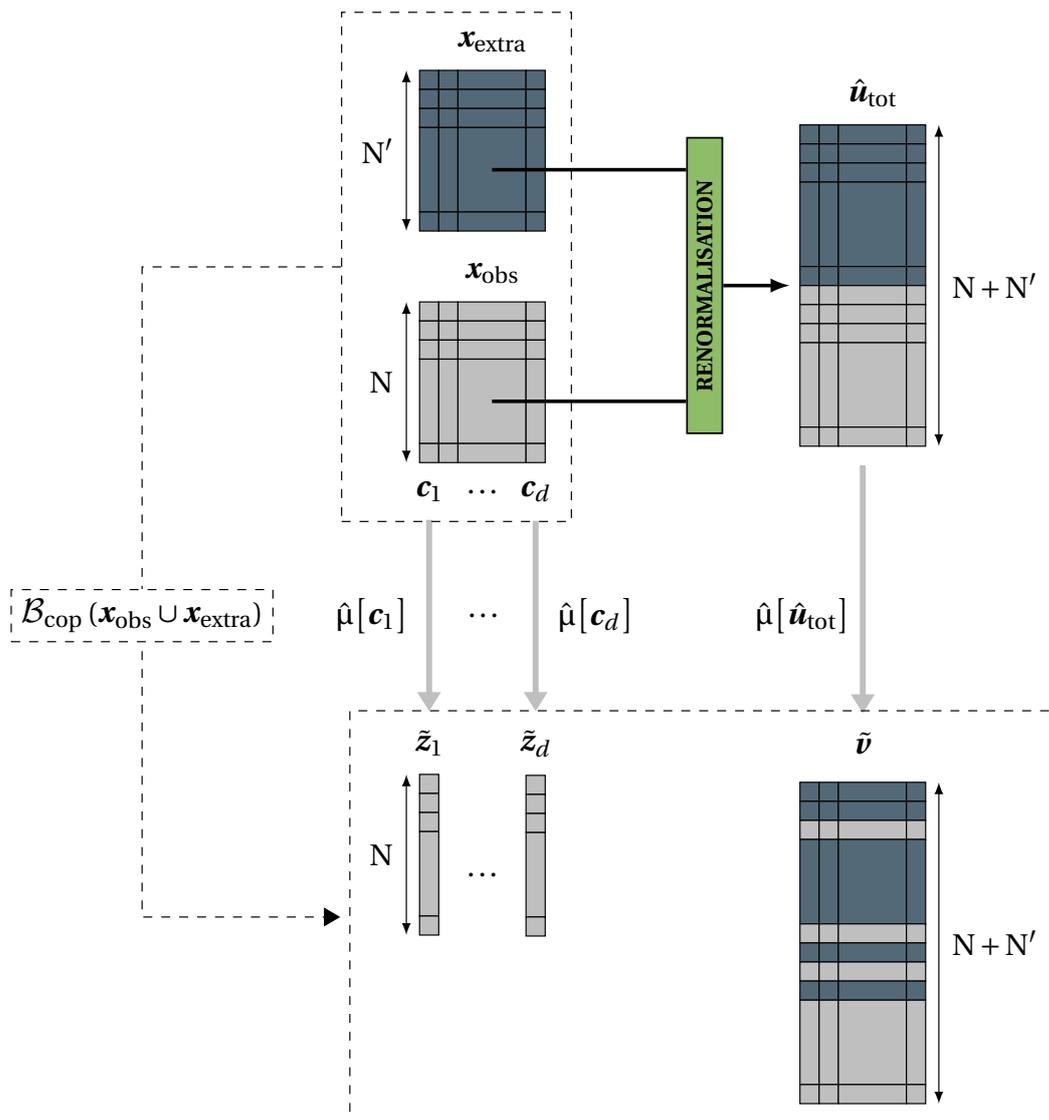


FIGURE 4.9 – Mécanisme de ré-échantillonnage  $\mathcal{B}_{\text{cop}}$  en présence d'observations supplémentaires de  $\mathbf{U}$ .

### 4.3 Validation de la procédure d'ASF

Dans cette section, on cherche à valider les résultats fournis par l'algorithme d'ASF construit à la Section 4.2 en vérifiant qu'ils apportent une réponse adéquate au Problème 4.2 de décomposition de la variabilité d'apprentissage. Cette étape de validation passe par l'étude des simulations numériques conduites pour deux cas tests analytiques. Le premier met en jeu des entrées gaussiennes et une fonction de performance linéaire. Cette simplicité autorise divers développements théoriques qui guident notre appréciation des résultats fournis par l'algorithme d'ASF. À l'aide de ce premier cas test, on compte étalonner les hyperparamètres, valider le bien-fondé de l'utilisation des techniques de ré-échantillonnage basées sur le *bootstrap* et prouver qu'une procédure d'acquisition ciblée peut raisonnablement être envisagée en pratique, c'est-à-dire sans solliciter un nombre excessivement important de nouvelles observations de la composante fonctionnelle à raffiner. Pour le second cas test, la loi d'entrée est une distribution de mélange tandis que la fonction de performance s'écrit comme une somme de fonctions unidimensionnelles. Cette situation est exposée pour mettre en avant le fait que les indices de sensibilité produits par l'algorithme d'ASF sont caractéristiques des modèles  $\mathcal{M}_1, \dots, \mathcal{M}_d$  et  $\mathcal{M}_{\text{cop}}$  retenus pour l'apprentissage.

#### 4.3.1 Étude d'un cas test linéaire

##### 4.3.1.1 Description

Dans ce premier cas test, on considère une fonction de performance  $\phi$  linéaire appliquée à un vecteur d'entrées  $\mathbf{X} \in \mathbb{R}^3$ . La sortie aléatoire  $Y$  est alors définie par :

$$Y = \phi(\mathbf{X}) = a_1X_1 + a_2X_2 + a_3X_3 = \mathbf{a}^\top \mathbf{X} \quad \text{avec} \quad \mathbf{a} = [a_1, a_2, a_3]^\top. \quad (4.49)$$

La loi conjointe  $\mu_{\mathbf{X}}$  est gaussienne :

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{avec} \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \quad \text{et} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix}. \quad (4.50)$$

Dans cette situation très simple, la valeur théorique de la probabilité de défaillance  $P_f$  peut être déduite de calculs probabilistes élémentaires :

$$P_f = \mathbb{P}(Y > T) = \mathbb{P}(\mathbf{a}^\top \mathbf{X} > T) = 1 - \Phi\left(\frac{T - \mu_{\mathbf{a}}}{\sigma_{\mathbf{a}}}\right) \quad \text{avec} \quad \begin{cases} \mu_{\mathbf{a}} = \mathbf{a}^\top \boldsymbol{\mu} \\ \sigma_{\mathbf{a}}^2 = \mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a} \end{cases}, \quad (4.51)$$

où  $\Phi$  est la fonction de répartition de la loi gaussienne unidimensionnelle centrée et réduite. Étant donné que  $P_f$  est censée être la probabilité d'un évènement rare, le seuil  $T$  est naturellement supposé plus grand que  $\mu_{\mathbf{a}}$ .

Essayons d'envisager cette configuration selon le formalisme d'une analyse de fiabilité conduite avec l'algorithme FORM présenté au Chapitre 2. Sachant que  $\mathbf{X}$  est gaussien et que l'état limite est linéaire, on est dans l'une des rares situations où FORM donne la valeur exacte de la probabilité de défaillance. Dans ce cas test, l'algorithme de fiabilité  $\mathcal{R}$  peut donc être assimilé (au choix) à une formule analytique ou à la méthode FORM. Si on s'en tient à FORM, un changement d'échelle de type linéaire permet de transformer le vecteur initial  $\mathbf{X}$  en un vecteur gaussien  $\mathbf{W}$  centré et réduit :

$$\mathbf{W} = \mathcal{T}(\mathbf{X}) = \mathbf{L}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \quad \text{avec} \quad \boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^\top. \quad (4.52)$$

$\mathcal{T}$  désigne la transformée de NATAF [1962] tandis que  $\mathbf{L}$  est la matrice triangulaire inférieure issue de la factorisation de Cholesky de  $\boldsymbol{\Sigma}$ . La variable de sortie  $Y$  peut alors s'écrire en fonction de  $\mathbf{W}$  :

$$Y = \phi(\mathbf{X}) = \mathbf{a}^\top (\mathbf{L}\mathbf{W} + \boldsymbol{\mu}), \quad (4.53)$$

et il en va de même pour la probabilité de défaillance :

$$P_f = \mathbb{P}(\phi(\mathbf{X}) > T) = \mathbb{P}(H(\mathbf{W}) > T) \quad \text{avec} \quad H(\mathbf{W}) = [\phi \circ \mathcal{T}^{-1}](\mathbf{W}) = \mathbf{a}^\top (\mathbf{L}\mathbf{W} + \boldsymbol{\mu}). \quad (4.54)$$

Dans l'espace normal standard, l'état limite est la courbe d'équation  $H(\mathbf{w}) = T$ . Le domaine de défaillance  $D_f^{\mathcal{N}}$  est quant à lui donné par :

$$D_f^{\mathcal{N}} = \left\{ \mathbf{w} \in \mathbb{R}^d : H(\mathbf{w}) > T \right\}. \quad (4.55)$$

On rappelle que la méthodologie FORM comprend une première étape permettant de déterminer le point de conception  $\mathbf{w}^*$  par résolution du problème d'optimisation :

$$\mathbf{w}^* = \underset{H(\mathbf{w})=T}{\operatorname{argmin}} \|\mathbf{w}\|_2, \quad (4.56)$$

puis une seconde consistant à calculer l'indice de fiabilité d'Hasofer-Lindt  $\beta$  et la probabilité de défaillance  $P_f$  à partir des formules suivantes :

$$\beta = \|\mathbf{w}^*\|_2 \quad \text{et} \quad P_f = 1 - \Phi(\beta). \quad (4.57)$$

La méthode FORM n'est exacte que dans le cas où  $H$  est linéaire. En effet, la linéarité de l'état limite permet alors de résoudre analytiquement le problème d'optimisation (4.56). Avec des outils rudimentaires de géométrie vectorielle et de calcul différentiel, on aboutit à :

$$\mathbf{w}^* = (T - \mu_a) \frac{\mathbf{b}}{\|\mathbf{b}\|_2^2} \quad \text{avec} \quad \mathbf{b}^\top = \mathbf{a}^\top \mathbf{L} \quad \text{et} \quad \beta = \frac{T - \mu_a}{\sigma_a}. \quad (4.58)$$

On retrouve tout naturellement le résultat de l'équation (4.51) et cela justifie l'applicabilité de FORM dans la situation étudiée. Une illustration est proposée à la Figure 4.10a pour rappeler au lecteur l'interprétation géométrique des différentes notions évoquées. Les valeurs numériques assignées aux paramètres de  $\phi$  et  $f_{\mathbf{X}}$  sont listées ci-dessous :

- Fonction de performance  $\phi$  :  $a_1 = a_2 = 2; a_3 = 1$
- Vecteur moyen  $\boldsymbol{\mu}$  :  $\mu_1 = \mu_2 = \mu_3 = 0$
- Matrice de covariance  $\boldsymbol{\Sigma}$  :  $\sigma_1 = \sigma_2 = \sigma_3 = 1; \rho_{12} = 0.5; \rho_{13} = \rho_{23} = 0.3$

Avec  $T = 11$ , la valeur théorique de la probabilité de défaillance est  $P_f = 2.53 \times 10^{-3}$ .

#### 4.3.1.2 Visualisation de l'impact de la variabilité d'apprentissage

Supposons que l'on utilise les générateurs théoriques  $G_Z$  et  $G_V$  pour fabriquer des échantillons d'apprentissage  $\mathbf{Z}_{\text{obs}}$  et  $\mathbf{V}_{\text{obs}}$  indépendants. On choisit d'utiliser un modèle d'apprentissage  $\mathcal{M}$  entièrement paramétrique et gaussien. Pour la  $j$ -ème loi marginale,  $\mathcal{M}_j$  représente l'estimation par maximum de vraisemblance des paramètres de la famille gaussienne  $\mathcal{F}_j$  :

$$\hat{f}_{\mathbf{X}_j}(\cdot | \mathcal{M}_j, \mathbf{Z}_{\text{obs}}^{[\bullet, j]}) = \hat{f}_{\mathbf{X}_j}(\cdot | \mathcal{F}_j, \hat{\boldsymbol{\mu}}_j^{\text{MV}}, \hat{\boldsymbol{\sigma}}_j^{\text{MV}}), \quad (4.59)$$

où les EMV sont calculés à partir de  $\mathbf{Z}_{\text{obs}}^{[\bullet, j]}$  comme suit :

$$\hat{\boldsymbol{\mu}}_j^{\text{MV}} = \frac{1}{N} \sum_{i=1}^N \mathbf{Z}_j^{(i)} \quad \text{et} \quad \hat{\boldsymbol{\sigma}}_j^{\text{MV}} = \sqrt{\frac{1}{N} \sum_{i=1}^N \left( \mathbf{Z}_j^{(i)} - \hat{\boldsymbol{\mu}}_j^{\text{MV}} \right)^2}. \quad (4.60)$$

Pour ce qui est de la copule,  $\mathcal{M}_{\text{cop}}$  désigne l'estimation par maximum de vraisemblance au sein de la famille gaussienne  $\mathcal{F}_{\text{cop}}$  :

$$\hat{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{V}_{\text{obs}}) = \hat{\mathbf{U}}(\cdot | \mathcal{F}_{\text{cop}}, \hat{\boldsymbol{\rho}}^{\text{MV}}). \quad (4.61)$$

La matrice  $\hat{\boldsymbol{\rho}}^{\text{MV}}$  s'obtient comme suit :

(R1) Envoi des données dans  $\mathbb{R}^d$  par application de  $\Phi^{-1}$  aux coefficients de  $\mathbf{V}_{\text{obs}}$  :

$$\mathbf{E}_{\text{obs}} := \left[ E_j^{(i)} \right]_{\substack{1 \leq i \leq N \\ 1 \leq j \leq d}} \quad \text{avec} \quad E_j^{(i)} = \Phi^{-1} \left( \mathbf{V}_j^{(i)} \right). \quad (4.62)$$

(R2) Utilisation de  $\mathbf{E}_{\text{obs}}$  pour l'estimation par maximum de vraisemblance de la matrice de covariance d'un vecteur gaussien centré :

$$\hat{\mathbf{\Gamma}}^{\text{MV}} := [\hat{\Gamma}_{kl}]_{1 \leq k, l \leq d} \quad \text{avec} \quad \hat{\Gamma}_{kl} = \frac{1}{N} \sum_{i=1}^N E_k^{(i)} E_l^{(i)}. \quad (4.63)$$

(R3) Extraction de la matrice de corrélation associée à  $\hat{\mathbf{\Gamma}}^{\text{MV}}$  :

$$\hat{\mathbf{\rho}}^{\text{MV}} := [\hat{\rho}_{kl}]_{1 \leq k < l \leq d} \quad \text{avec} \quad \hat{\rho}_{kl} = \frac{\hat{\Gamma}_{kl}}{\sqrt{\hat{\Gamma}_{kk}} \sqrt{\hat{\Gamma}_{ll}}}. \quad (4.64)$$

Après ré-assemblage,  $\hat{f}_{\mathbf{X}}$  n'est rien d'autre que la densité conjointe d'un vecteur gaussien :

$$\hat{f}_{\mathbf{X}}(\cdot | \mathcal{M}, \mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}}) = \hat{f}_{\mathbf{X}}(\cdot | \mathcal{F}, \hat{\boldsymbol{\mu}}^{\text{MV}}, \hat{\boldsymbol{\Sigma}}^{\text{MV}}), \quad (4.65)$$

dont les paramètres sont donnés par :

$$\hat{\boldsymbol{\mu}}^{\text{MV}} = \left[ \hat{\mu}_j^{\text{MV}} \right]_{1 \leq j \leq d} \quad \text{et} \quad \hat{\boldsymbol{\Sigma}}^{\text{MV}} = \hat{\mathbf{\Delta}}^{\text{MV}} \hat{\boldsymbol{\rho}}^{\text{MV}} \hat{\mathbf{\Delta}}^{\text{MV}} \quad \text{avec} \quad \hat{\mathbf{\Delta}}^{\text{MV}} = \begin{bmatrix} \hat{\sigma}_1^{\text{MV}} & 0 & 0 \\ 0 & \hat{\sigma}_2^{\text{MV}} & 0 \\ 0 & 0 & \hat{\sigma}_3^{\text{MV}} \end{bmatrix}. \quad (4.66)$$

Avec un tel modèle d'apprentissage, l'indépendance des échantillons d'apprentissage  $\mathbf{Z}_{\text{obs}}$  et  $\mathbf{V}_{\text{obs}}$  garantit l'indépendance mutuelle des paramètres se rapportant aux  $d + 1$  composantes :

$$\underbrace{(\hat{\mu}_1^{\text{MV}}, \hat{\sigma}_1^{\text{MV}})}_{\text{pour } \hat{f}_{X_1}} \perp \underbrace{(\hat{\mu}_2^{\text{MV}}, \hat{\sigma}_2^{\text{MV}})}_{\text{pour } \hat{f}_{X_2}} \perp \underbrace{(\hat{\mu}_3^{\text{MV}}, \hat{\sigma}_3^{\text{MV}})}_{\text{pour } \hat{f}_{X_3}} \perp \underbrace{(\hat{\rho}_{12}^{\text{MV}}, \hat{\rho}_{13}^{\text{MV}}, \hat{\rho}_{23}^{\text{MV}})}_{\text{pour } \hat{c}_{\mathbf{U}}} \quad (4.67)$$

À la fin du processus d'analyse de fiabilité décrit à la Figure 4.4, on a :

$$P_f(\hat{\boldsymbol{\mu}}^{\text{MV}}, \hat{\boldsymbol{\Sigma}}^{\text{MV}}) = 1 - \Phi \left( \frac{T - \hat{\mu}_{\mathbf{a}}}{\hat{\sigma}_{\mathbf{a}}} \right) \quad \text{avec} \quad \begin{cases} \hat{\mu}_{\mathbf{a}} = \mathbf{a}^{\top} \hat{\boldsymbol{\mu}}^{\text{MV}} \\ \hat{\sigma}_{\mathbf{a}}^2 = \mathbf{a}^{\top} \hat{\boldsymbol{\Sigma}}^{\text{MV}} \mathbf{a} \end{cases}. \quad (4.68)$$

Au vu de ce qui a déjà été fait dans la Section 4.3.1.1, une approximation du domaine de défaillance dans l'espace normal standard en découle immédiatement :

$$\hat{D}_f^{\mathcal{N}} = \{ \mathbf{w} \in \mathbb{R}^d : \hat{H}(\mathbf{w}) > T \} \quad \text{avec} : \quad (4.69)$$

$$\hat{H} := \phi \circ \hat{\mathcal{T}}^{-1} \quad ; \quad \hat{\mathcal{T}} : \mathbf{x} \mapsto \mathbf{w} := (\hat{\mathbf{L}}^{\text{MV}})^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}^{\text{MV}}) \quad \text{et} \quad \hat{\boldsymbol{\rho}}^{\text{MV}} = \hat{\mathbf{L}}^{\text{MV}} (\hat{\mathbf{L}}^{\text{MV}})^{\top}.$$

Le point de conception  $\mathbf{w}^*$  peut quant à lui être approché par :

$$\hat{\mathbf{w}}^* = (T - \hat{\mu}_{\mathbf{a}}) \frac{\hat{\mathbf{b}}}{\|\hat{\mathbf{b}}\|_2} \quad \text{avec} \quad \hat{\mathbf{b}}^{\top} = \mathbf{a}^{\top} \hat{\mathbf{L}}^{\text{MV}}. \quad (4.70)$$

La ressemblance entre l'expression (4.51) de la probabilité théorique et l'expression (4.68) de la probabilité estimée vient du fait que l'apprentissage n'est rien d'autre qu'une calibration du modèle de distribution théorique. La seule incertitude est celle introduite par le petit nombre d'observations utilisées pour ajuster les familles  $\mathcal{F}_1, \dots, \mathcal{F}_d$  et  $\mathcal{F}_{\text{cop}}$ . La variabilité de la prédiction du risque de défaillance est alors d'autant plus faible que la taille  $N$  des échantillons est suffisamment grande pour permettre une reconstruction précise des paramètres de la densité sous-jacente. On peut se reporter à la Figure 4.10b pour une visualisation de l'impact de la variabilité d'apprentissage sur les contours du domaine  $\hat{D}_f^{\mathcal{N}}$ . Pour une séquence de couples  $\{(\mathbf{z}_{\text{obs}}^{(k)}, \mathbf{v}_{\text{obs}}^{(k)})\}_{k=1}^n$  simulés à partir des générateurs théoriques  $G_{\mathbf{Z}}$  et  $G_{\mathbf{V}}$ , on applique les formules (4.59) à (4.70) pour conduire l'apprentissage puis l'analyse de fiabilité. L'état limite reste un hyperplan de  $\mathbb{R}^3$  mais sa direction et son point de conception  $\hat{\mathbf{w}}^*$  dépendent de la qualité de l'estimation paramétrique. Les dessins 2D proposés à la Figure 4.10 sont des coupes transversales judicieusement choisies.

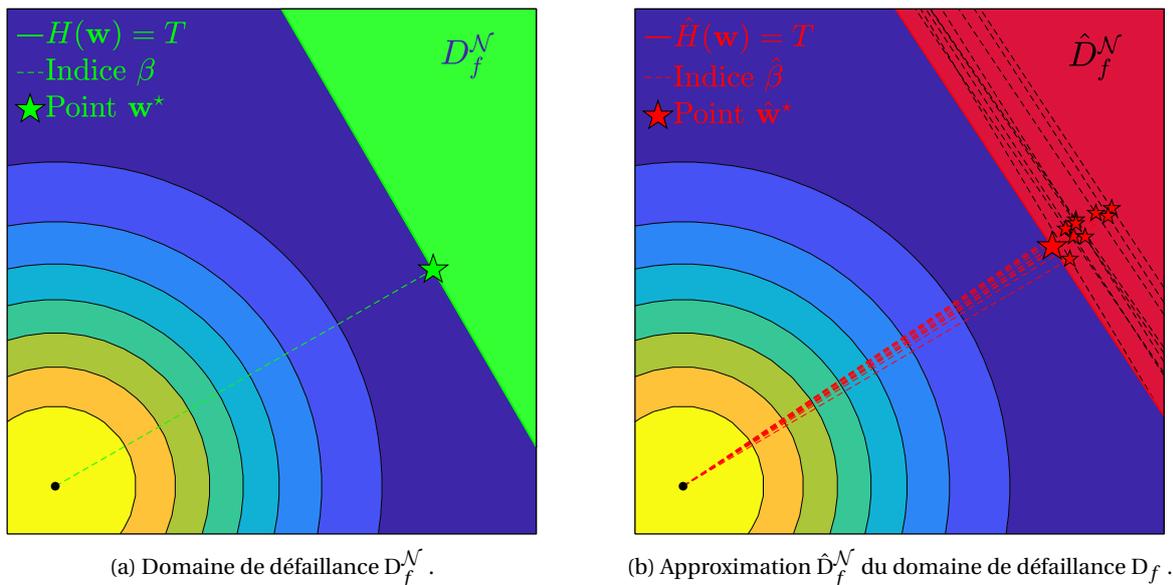


FIGURE 4.10 – Visualisation de l’impact du contexte “small data” sur l’estimation du risque de défaillance.

### 4.3.1.3 À propos des indices obtenus avec les générateurs théoriques

Dans ce paragraphe, on choisit d’alléger les notations. Pour les paramètres estimés par maximum de vraisemblance, on omet le sigle MV en exposant. Les composantes fonctionnelles estimées dans un cadre gaussien deviennent alors :

- $\hat{f}_{X_j}(\cdot | \mathcal{F}_j, \hat{\mu}_j^{MV}, \hat{\sigma}_j^{MV}) \Rightarrow \hat{f}_{\mathcal{N}}(\cdot | \hat{\mu}_j, \hat{\sigma}_j)$ ,
- $\hat{F}_{X_j}(\cdot | \mathcal{F}_j, \hat{\mu}_j^{MV}, \hat{\sigma}_j^{MV}) \Rightarrow \hat{F}_{\mathcal{N}}(\cdot | \hat{\mu}_j, \hat{\sigma}_j)$ ,
- $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{F}_{\text{cop}}, \hat{\rho}^{MV}) \Rightarrow \hat{c}_{\mathcal{N}}(\cdot | \hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23})$ .

On s’intéresse aux indices théoriques  $S_1, S_2, S_3$  et  $S_{\text{cop}}$  tels qu’ils sont exprimés aux équations (4.37) et (4.38). Avec les nouvelles notations, ils peuvent être réécrits comme suit :

$$S_j = \frac{\mathbb{V}\left(\mathbb{E}\left[P_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) \mid \hat{f}_{\mathcal{N}}(\cdot | \hat{\mu}_j, \hat{\sigma}_j)\right]\right)}{\mathbb{V}\left(P_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})\right)} = \frac{\mathbb{V}\left(\mathbb{E}\left[P_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) \mid \hat{\mu}_j, \hat{\sigma}_j\right]\right)}{\mathbb{V}\left(P_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})\right)}, \quad (4.71)$$

$$S_{\text{cop}} = \frac{\mathbb{V}\left(\mathbb{E}\left[P_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) \mid \hat{c}_{\mathcal{N}}(\cdot | \hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23})\right]\right)}{\mathbb{V}\left(P_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})\right)} = \frac{\mathbb{V}\left(\mathbb{E}\left[P_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) \mid \hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23}\right]\right)}{\mathbb{V}\left(P_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})\right)}. \quad (4.72)$$

Au regard des valeurs numériques choisies pour les paramètres de  $f_{\mathbf{X}}$  et  $\phi$ , on peut s’attendre à ce que  $S_1 = S_2$  de par les effets de symétrie qui sont à l’œuvre entre les variables  $X_1$  et  $X_2$  :

- (Sym1) **Échantillonnage :**  $\begin{cases} \mathbf{Z}_{\text{obs}}^{[1]} \sim (\mu_{X_1})^{\otimes N} \\ \mathbf{Z}_{\text{obs}}^{[2]} \sim (\mu_{X_2})^{\otimes N} \end{cases}$  avec  $\mu_{X_1} = \mu_{X_2}$
- (Sym2) **Apprentissage :**  $\begin{cases} \mathcal{F}_1 \text{ et } \mathcal{F}_2 \text{ identiques} \\ \text{Calibration par MV avec l'équation (4.60)} \end{cases}$
- (Sym3) **Motif de dépendance avec  $X_3$  :**  $\begin{cases} (\mathbf{V}_{\text{obs}}^{[1]}, \mathbf{V}_{\text{obs}}^{[3]}) \sim c_{\mathcal{N}}(\cdot | \rho_{13}) \\ (\mathbf{V}_{\text{obs}}^{[2]}, \mathbf{V}_{\text{obs}}^{[3]}) \sim c_{\mathcal{N}}(\cdot | \rho_{23}) \end{cases}$  avec  $\rho_{13} = \rho_{23}$
- (Sym4) **Fonction de performance :**  $a_1 = a_2 = 2$

Avec (Sym1) et (Sym2), il est clair que les couples  $(\hat{\mu}_1, \hat{\sigma}_1)$  et  $(\hat{\mu}_2, \hat{\sigma}_2)$  sont de même loi. De même, au vu des étapes (R1) à (R3) permettant d’obtenir  $\hat{\rho}$  à partir de  $\mathbf{V}_{\text{obs}}$ , (Sym3) justifie que les triplets  $(\hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23})$  et  $(\hat{\rho}_{12}, \hat{\rho}_{23}, \hat{\rho}_{13})$  sont identiquement distribués. Pour démontrer rigoureusement

l'égalité des deux indices, il suffit de prouver que :

$$\boxed{\mathbb{E} [P_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) \mid \hat{\boldsymbol{\mu}}_1, \hat{\sigma}_1] = \mathbb{E} [P_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) \mid \hat{\boldsymbol{\mu}}_2, \hat{\sigma}_2]} . \quad (\text{Obj})$$

L'équation (4.68) donne une formule condensée de  $P_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ . Elle est complètement équivalente à l'expression intégrale suivante :

$$P_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = \int_{D_f} \hat{f}_{\mathcal{N}}(x_1 \mid \hat{\boldsymbol{\mu}}_1, \hat{\sigma}_1) \times \hat{f}_{\mathcal{N}}(x_2 \mid \hat{\boldsymbol{\mu}}_2, \hat{\sigma}_2) \times \hat{f}_{\mathcal{N}}(x_3 \mid \hat{\boldsymbol{\mu}}_3, \hat{\sigma}_3) \dots \\ \times \hat{c}_{\mathcal{N}}\left(\hat{F}_{\mathcal{N}}(x_1 \mid \hat{\boldsymbol{\mu}}_1, \hat{\sigma}_1), \hat{F}_{\mathcal{N}}(x_2 \mid \hat{\boldsymbol{\mu}}_2, \hat{\sigma}_2), \hat{F}_{\mathcal{N}}(x_3 \mid \hat{\boldsymbol{\mu}}_3, \hat{\sigma}_3) \mid \hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23}\right) d\mathbf{x} . \quad (4.73)$$

Cette dernière nous incite à envisager  $P_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$  comme le résultat d'une application  $\xi$  conçue pour séparer les paramètres estimés en blocs indépendants :

$$P_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = \xi((\hat{\boldsymbol{\mu}}_1, \hat{\sigma}_1); (\hat{\boldsymbol{\mu}}_2, \hat{\sigma}_2); (\hat{\boldsymbol{\mu}}_3, \hat{\sigma}_3); (\hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23})) . \quad (4.74)$$

Les deux premiers arguments de  $\xi$  peuvent être permutés à condition d'opérer un réarrangement des coefficients de corrélation. On a ainsi :

$$\xi((\hat{\boldsymbol{\mu}}_2, \hat{\sigma}_2); (\hat{\boldsymbol{\mu}}_1, \hat{\sigma}_1); (\hat{\boldsymbol{\mu}}_3, \hat{\sigma}_3); (\hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23})) \\ = \int_{D_f} \hat{f}_{\mathcal{N}}(x_1 \mid \hat{\boldsymbol{\mu}}_2, \hat{\sigma}_2) \times \hat{f}_{\mathcal{N}}(x_2 \mid \hat{\boldsymbol{\mu}}_1, \hat{\sigma}_1) \times \hat{f}_{\mathcal{N}}(x_3 \mid \hat{\boldsymbol{\mu}}_3, \hat{\sigma}_3) \dots \\ \times \hat{c}_{\mathcal{N}}\left(\hat{F}_{\mathcal{N}}(x_1 \mid \hat{\boldsymbol{\mu}}_2, \hat{\sigma}_2), \hat{F}_{\mathcal{N}}(x_2 \mid \hat{\boldsymbol{\mu}}_1, \hat{\sigma}_1), \hat{F}_{\mathcal{N}}(x_3 \mid \hat{\boldsymbol{\mu}}_3, \hat{\sigma}_3) \mid \hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23}\right) d\mathbf{x} \quad (4.75)$$

$$= \int_{D_f} \hat{f}_{\mathcal{N}}(x_2 \mid \hat{\boldsymbol{\mu}}_1, \hat{\sigma}_1) \times \hat{f}_{\mathcal{N}}(x_1 \mid \hat{\boldsymbol{\mu}}_2, \hat{\sigma}_2) \times \hat{f}_{\mathcal{N}}(x_3 \mid \hat{\boldsymbol{\mu}}_3, \hat{\sigma}_3) \dots \\ \times \hat{c}_{\mathcal{N}}\left(\hat{F}_{\mathcal{N}}(x_2 \mid \hat{\boldsymbol{\mu}}_1, \hat{\sigma}_1), \hat{F}_{\mathcal{N}}(x_1 \mid \hat{\boldsymbol{\mu}}_2, \hat{\sigma}_2), \hat{F}_{\mathcal{N}}(x_3 \mid \hat{\boldsymbol{\mu}}_3, \hat{\sigma}_3) \mid \hat{\rho}_{12}, \hat{\rho}_{23}, \hat{\rho}_{13}\right) d\mathbf{x} \quad (4.76)$$

$$= \int_{D_f} \hat{f}_{\mathcal{N}}(s_1 \mid \hat{\boldsymbol{\mu}}_1, \hat{\sigma}_1) \times \hat{f}_{\mathcal{N}}(s_2 \mid \hat{\boldsymbol{\mu}}_2, \hat{\sigma}_2) \times \hat{f}_{\mathcal{N}}(s_3 \mid \hat{\boldsymbol{\mu}}_3, \hat{\sigma}_3) \dots \\ \times \hat{c}_{\mathcal{N}}\left(\hat{F}_{\mathcal{N}}(s_1 \mid \hat{\boldsymbol{\mu}}_1, \hat{\sigma}_1), \hat{F}_{\mathcal{N}}(s_2 \mid \hat{\boldsymbol{\mu}}_2, \hat{\sigma}_2), \hat{F}_{\mathcal{N}}(s_3 \mid \hat{\boldsymbol{\mu}}_3, \hat{\sigma}_3) \mid \hat{\rho}_{12}, \hat{\rho}_{23}, \hat{\rho}_{13}\right) ds \quad (4.77)$$

$$= \xi((\hat{\boldsymbol{\mu}}_1, \hat{\sigma}_1); (\hat{\boldsymbol{\mu}}_2, \hat{\sigma}_2); (\hat{\boldsymbol{\mu}}_3, \hat{\sigma}_3); (\hat{\rho}_{12}, \hat{\rho}_{23}, \hat{\rho}_{13})) . \quad (4.78)$$

Dans l'égalité (4.76), on modifie le paramétrage de la densité de copule  $\hat{c}_{\mathcal{N}}$  afin de pouvoir permuter les deux premiers arguments. Un changement de variables consistant à remplacer le triplet  $\mathbf{x} = (x_1, x_2, x_3)$  par  $\mathbf{s} = (s_2, s_1, s_3)$  conduit ensuite à l'égalité (4.77) où l'on peut remarquer que le domaine d'intégration reste inchangé par conséquence directe de (Sym4).

On se propose désormais d'expliciter le calcul de la première espérance dans le cas d'un conditionnement par l'évènement  $\{\hat{\boldsymbol{\mu}}_1 = m, \hat{\sigma}_1 = s\}$  où  $(m, s)$  est un couple pris dans  $\mathbb{R} \times [0, +\infty[$  :

$$\mathbb{E}\left[P_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) \mid \hat{\boldsymbol{\mu}}_1 = m, \hat{\sigma}_1 = s\right] \\ = \mathbb{E}\left[\xi((\hat{\boldsymbol{\mu}}_1, \hat{\sigma}_1); (\hat{\boldsymbol{\mu}}_2, \hat{\sigma}_2); (\hat{\boldsymbol{\mu}}_3, \hat{\sigma}_3); (\hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23})) \mid \hat{\boldsymbol{\mu}}_1 = m, \hat{\sigma}_1 = s\right] \text{ avec (4.74)} \\ = \mathbb{E}\left[\xi((m, s); (\hat{\boldsymbol{\mu}}_2, \hat{\sigma}_2); (\hat{\boldsymbol{\mu}}_3, \hat{\sigma}_3); (\hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23}))\right] \text{ avec (4.67)} \\ = \mathbb{E}\left[\xi((m, s); (\hat{\boldsymbol{\mu}}_1, \hat{\sigma}_1); (\hat{\boldsymbol{\mu}}_3, \hat{\sigma}_3); (\hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23}))\right] \text{ avec (Sym1) et (Sym2)} \\ = \mathbb{E}\left[\xi((m, s); (\hat{\boldsymbol{\mu}}_1, \hat{\sigma}_1); (\hat{\boldsymbol{\mu}}_3, \hat{\sigma}_3); (\hat{\rho}_{12}, \hat{\rho}_{23}, \hat{\rho}_{13}))\right] \text{ avec (Sym3)} \\ = \mathbb{E}\left[\xi((\hat{\boldsymbol{\mu}}_1, \hat{\sigma}_1); (m, s); (\hat{\boldsymbol{\mu}}_3, \hat{\sigma}_3); (\hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23}))\right] \text{ avec (4.78)} \\ = \mathbb{E}\left[\xi((\hat{\boldsymbol{\mu}}_1, \hat{\sigma}_1); (\hat{\boldsymbol{\mu}}_2, \hat{\sigma}_2); (\hat{\boldsymbol{\mu}}_3, \hat{\sigma}_3); (\hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23})) \mid \hat{\boldsymbol{\mu}}_2 = m, \hat{\sigma}_2 = s\right] \text{ avec (4.67)} \\ = \mathbb{E}\left[P_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) \mid \hat{\boldsymbol{\mu}}_2 = m, \hat{\sigma}_2 = s\right] \text{ avec (4.74)}. \quad (4.79)$$

L'égalité des espérances conditionnelles vient alors naturellement et termine la preuve de (Obj). Sachant que les blocs  $(\hat{\mu}_1, \hat{\sigma}_1)$  et  $(\hat{\mu}_2, \hat{\sigma}_2)$  sont identiquement distribués, l'égalité  $S_1 = S_2$  devient évidente. Si on applique l'algorithme d'ASF développé à la Section 4.2.3.2 pour estimer ces deux indices, on retrouvera cette relation à condition que :

- (a)  $\mathcal{B}$  imite le fonctionnement de  $G_Z$  et  $G_V$  sans introduire de biais lié à  $\mathbf{x}_{\text{obs}}$ .
- (b) Le schéma d'estimation  $\mathcal{A}$  des indices de Sobol soit suffisamment précis et robuste.

Pour le moment, on supprime la première source d'erreur. Son influence sur la précision des estimations sera étudiée dans le prochain paragraphe. On déploie donc l'algorithme d'ASF en remplaçant la phase  $\mathcal{B}$  de ré-échantillonnage *bootstrap* par des tirages aléatoires avec les générateurs théoriques  $G_Z$  et  $G_V$ . On choisit d'estimer les indices de Sobol à l'aide du schéma  $\mathcal{A}$  de type *Pick-Freeze* dont le seul hyperparamètre est le nombre  $M$  d'itérations. Son choix doit faire l'objet d'un compromis entre précision et coût de calcul.  $M$  doit être suffisamment grand pour que les estimateurs des indices de Sobol aient suffisamment convergé. Toutefois, on doit garder en tête qu'un nombre  $M$  d'itérations induit  $M \times (d + 2)$  appels à  $\mathcal{R}$  ce qui représente une charge de calcul rapidement prohibitive dès lors que la fonction de performance n'est plus analytique. On se propose de choisir empiriquement  $M$  en procédant comme suit :

1. Simulation de  $M_{\text{max}} = 5 \times 10^3$  échantillons  $\mathbf{Z}_{\text{obs}}$  et  $\mathbf{V}_{\text{obs}}$  ( $N = 300$ ) à l'aide de  $G_Z$  et  $G_V$ .
2. Apprentissage de  $M_{\text{max}}$  collections de composantes fonctionnelles.
3. Calcul des  $M_{\text{max}} \times (d + 2)$  probabilités de défaillance associées à la mise en œuvre de  $\mathcal{A}$ .
4. Estimation des indices pour une séquence de valeurs de  $M$  allant de  $M_{\text{min}} = 100$  à  $M_{\text{max}}$ .

Les résultats obtenus sont présentés à la Figure 4.11 où l'on peut voir que  $M^* := 10^3$  itérations sont suffisantes pour obtenir un classement définitif des indices et la quasi-égalité de  $\hat{S}_1$  et  $\hat{S}_2$ . On retiendra cette valeur  $M^*$  comme un premier réglage du nombre d'itérations permettant d'atteindre un niveau satisfaisant de convergence des indices. À chacune des applications futures de l'algorithme d'ASF sur un nouveau cas test, une étape préalable consistera toujours à s'assurer que  $M = 10^3$  itérations garantissent la convergence. Toutefois, on ne montrera pas les graphiques de convergence pour ne pas alourdir inutilement les explications. On se propose maintenant de conduire  $N_{\text{rep}} = 100$  estimations des indices afin d'apprécier la robustesse du schéma  $\mathcal{A}$  utilisé. On obtient alors :

ASF :	$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_{\text{cop}}$	
moyenne :	32.33 %	32.77 %	4.89 %	28.13 %	(4.80)
écart-type :	3.18 %	2.97 %	2.71 %	3.25 %	
rang :	(1)	(1)	(3)	(2)	

Au vu de ce qui a été précédemment justifié par le calcul et observé à la Figure 4.11, on ne s'étonne pas d'avoir  $\hat{S}_1 \approx \hat{S}_2$ . On peut aussi souligner le fait que  $\hat{S}_3$  est bien plus petit que  $\hat{S}_{\text{cop}}$  dont la valeur moyenne est légèrement inférieure à celle du lot  $(\hat{S}_1, \hat{S}_2)$ . Cela signifie que l'apprentissage de la densité de copule  $\hat{c}_{\mathcal{N}}(\cdot | \hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23})$  exerce une plus grande influence sur la variabilité de  $P_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$  que l'apprentissage de la densité marginale  $\hat{f}_{\mathcal{N}}(\cdot | \hat{\mu}_3, \hat{\sigma}_3)$ . Pour tous les estimateurs, on constate que les écarts-types sont relativement faibles. Le schéma d'estimation  $\mathcal{A}$  peut par conséquent être considéré comme robuste (à condition d'incorporer  $M \geq 10^3$  itérations).

#### 4.3.1.4 Étude de la dépendance à l'échantillon initial

On se replace désormais dans un cadre où  $G_Z$  et  $G_V$  sont inconnus. La simulation d'échantillons d'apprentissage  $\mathbf{Z}_{\text{obs}}$  et  $\mathbf{V}_{\text{obs}}$  émanant de ces générateurs n'est donc plus possible. La connaissance de  $\mathbf{x}_{\text{obs}}$  et l'application du mécanisme  $\mathcal{B}$  permet par contre de fabriquer des échantillons  $\tilde{\mathbf{Z}}$  et  $\tilde{\mathbf{V}}$  obéissant à la mesure empirique  $\hat{\mu}_{ZV}[\mathbf{x}_{\text{obs}}]$  définie par l'équation (4.34). Les indices de Sobol

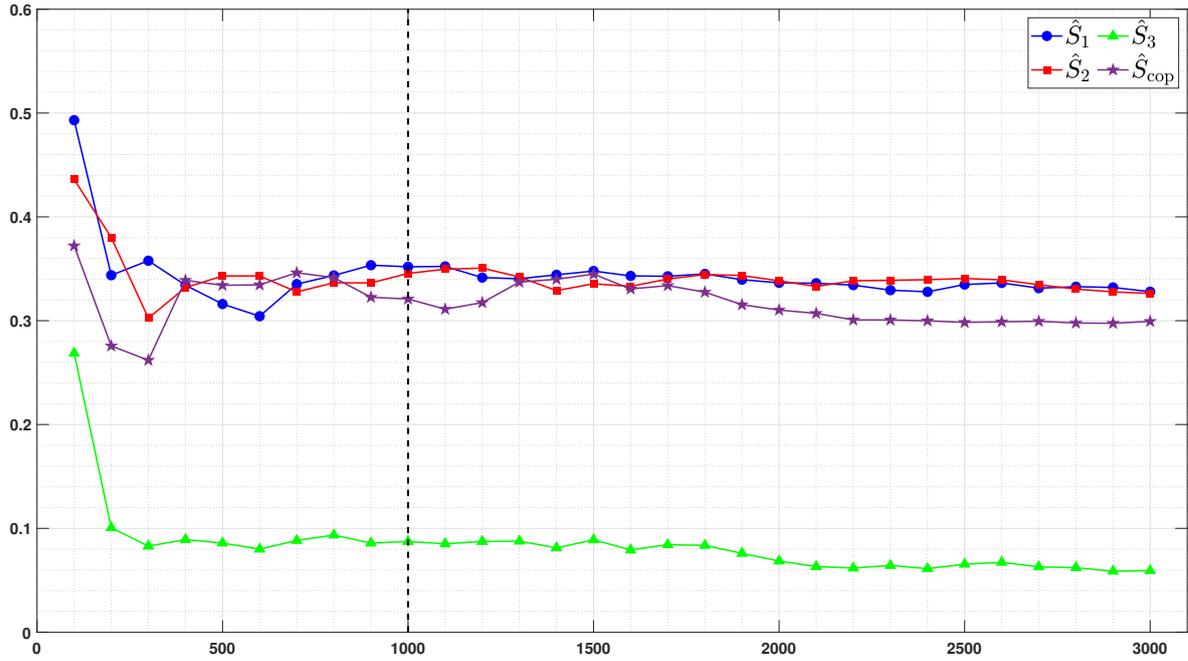


FIGURE 4.11 – Convergence des indices  $\hat{S}_j$  et  $\hat{S}_{\text{cop}}$  en fonction du nombre  $M$  d'itérations.

découlant de l'utilisation de  $\mathcal{B}$  peuvent alors s'exprimer de la manière suivante :

$$S_j^{\mathcal{B}} := \frac{\mathbb{V}\left(\mathbb{E}\left[\mathbb{P}_f(\tilde{\mathbf{Z}}, \tilde{\mathbf{V}}) \mid \hat{f}_{X_j}(\cdot \mid \mathcal{F}_j, \tilde{\mathbf{Z}}_j)\right]\right)}{\mathbb{V}\left(\mathbb{P}_f(\tilde{\mathbf{Z}}, \tilde{\mathbf{V}})\right)} \quad \text{et} \quad S_{\text{cop}}^{\mathcal{B}} := \frac{\mathbb{V}\left(\mathbb{E}\left[\mathbb{P}_f(\tilde{\mathbf{Z}}, \tilde{\mathbf{V}}) \mid \hat{c}_{\mathbf{U}}(\cdot \mid \mathcal{F}_{\text{cop}}, \tilde{\mathbf{V}})\right]\right)}{\mathbb{V}\left(\mathbb{P}_f(\tilde{\mathbf{Z}}, \tilde{\mathbf{V}})\right)}. \quad (4.81)$$

La notation  $\mathcal{B}$  en exposant des indices  $S_j^{\mathcal{B}}$  et  $S_{\text{cop}}^{\mathcal{B}}$  ne sera utilisée que dans les cas où elle permet de lever une ambiguïté. Pour valider le bien-fondé de la technique de ré-échantillonnage  $\mathcal{B}$  qui a été développée, il faut s'assurer que les indices  $\hat{S}_1^{\mathcal{B}}$ ,  $\hat{S}_2^{\mathcal{B}}$ ,  $\hat{S}_3^{\mathcal{B}}$  et  $\hat{S}_{\text{cop}}^{\mathcal{B}}$  fournis en sortie de l'algorithme d'ASF sont proches des indices théoriques  $S_1, S_2, S_3$  et  $S_{\text{cop}}$ . Pour cela, on considère trois échantillons distincts que l'on note  $\mathbf{x}_{\text{obs}}^{(k)}$  avec  $k \in \{1, 2, 3\}$  et  $N = 300$ . Pour chacun d'eux, on estime les indices de Sobol' avec l'algorithme d'ASF (réglé à  $M = 10^3$  itérations). Les résultats sont consignés dans le Tableau 4.1 après  $N_{\text{rep}} = 10$  estimations de chaque collection d'indices.

ASF exécutée avec $\mathcal{B}$ :		$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_{\text{cop}}$	
Échantillons	$\mathbf{x}_{\text{obs}}^{(1)}$	moyenne :	28.26 %	26.56 %	5.03 %	37.49 %
		écart-type :	4.18 %	4.35 %	3.85 %	4.19 %
		rang :	(2)	(3)	(4)	(1)
	$\mathbf{x}_{\text{obs}}^{(2)}$	moyenne :	28.23 %	32.52 %	3.78 %	30.75 %
		écart-type :	3.82 %	3.77 %	3.62 %	2.84 %
		rang :	(3)	(1)	(4)	(2)
	$\mathbf{x}_{\text{obs}}^{(3)}$	moyenne :	38.18 %	27.13 %	7.52 %	28.92 %
		écart-type :	2.34 %	3.88 %	3.41 %	3.91 %
		rang :	(1)	(3)	(4)	(2)

TABEAU 4.1 – Estimation des indices de Sobol' pour trois échantillons distincts.

Quel que soit l'échantillon  $\mathbf{x}_{\text{obs}}^{(k)}$  utilisé, l'ASF fait émerger un triplet formé par les indices  $\hat{S}_1^{\mathcal{B}}$ ,  $\hat{S}_2^{\mathcal{B}}$  et  $\hat{S}_{\text{cop}}^{\mathcal{B}}$ . Ainsi, l'apprentissage des composantes  $f_{X_1}$ ,  $f_{X_2}$  et  $c_{\mathbf{U}}$  engendre 95 % de la variabilité affectant  $\mathbb{P}_f(\tilde{\mathbf{Z}}, \tilde{\mathbf{V}})$  et les contributions individuelles sont globalement équivalentes. La part restante de variabilité peut alors être attribuée à l'apprentissage de  $f_{X_3}$ . Ces résultats sont tout à fait

cohérents avec ceux obtenus avec  $G_Z$  et  $G_V$ . Cela conforte l'idée selon laquelle il est tout à fait raisonnable d'utiliser le mécanisme  $\mathcal{B}$  en pratique pour pallier la méconnaissance de la distribution sous-jacente  $\mu_X$ . Toutefois, il faut aussi remarquer que la hiérarchie au sein du triplet dominant est susceptible d'évoluer d'un échantillon à l'autre :

$$\mathbf{x}_{\text{obs}}^{(1)} \Rightarrow \hat{S}_1^{\mathcal{B}} \approx \hat{S}_2^{\mathcal{B}} < \hat{S}_{\text{cop}}^{\mathcal{B}} \quad ; \quad \mathbf{x}_{\text{obs}}^{(2)} \Rightarrow \hat{S}_1^{\mathcal{B}} < \hat{S}_2^{\mathcal{B}} \approx \hat{S}_{\text{cop}}^{\mathcal{B}} \quad ; \quad \mathbf{x}_{\text{obs}}^{(3)} \Rightarrow \hat{S}_1^{\mathcal{B}} > \hat{S}_2^{\mathcal{B}} \approx \hat{S}_{\text{cop}}^{\mathcal{B}}. \quad (4.82)$$

On ne retrouve dans aucun des trois cas le classement  $\hat{S}_1 \approx \hat{S}_2 > \hat{S}_{\text{cop}}$  obtenu en utilisant les générateurs théoriques. Pour chaque échantillon  $\mathbf{x}_{\text{obs}}^{(k)}$  considéré, ré-échantillonner avec  $\mathcal{B}$  revient à simuler selon la mesure empirique  $\hat{\mu}_{ZV}[\mathbf{x}_{\text{obs}}^{(k)}]$  qui est construite exclusivement à partir des données fournies. On est alors confronté aux inéluctables limites d'une approche "*data-driven*". L'emploi de  $\hat{\mu}_{ZV}[\mathbf{x}_{\text{obs}}]$  introduit un biais qui dépend directement de  $\mathbf{x}_{\text{obs}}$ . Cette remarque est d'autant plus centrale dans un contexte "*small data*" où le budget d'observations de la loi conjointe  $\mu_X$  est restreint.  $\mathcal{B}$  doit être vu comme un générateur d'échantillons d'apprentissage conçu pour imiter le fonctionnement de  $G_Z$  et  $G_V$  à partir de la connaissance de  $\mathbf{x}_{\text{obs}}$ . La qualité de cette approximation s'améliore lorsque le nombre  $N$  d'atomes utilisés pour construire  $\hat{\mu}_{ZV}[\mathbf{x}_{\text{obs}}]$  augmente ou lorsque  $\mathbf{x}_{\text{obs}}$  est un échantillon fortement représentatif de  $\mu_X$ .

#### 4.3.1.5 Raffinement des mesures empiriques

Dans ce paragraphe, on s'attèle à montrer que ce n'est pas le *bootstrap* en lui-même qui occasionne la dépendance des résultats de l'ASF vis-à-vis de l'échantillon  $\mathbf{x}_{\text{obs}}$  fourni initialement. C'est davantage le contexte "*small data*" qui est en cause car il est responsable de la forte variabilité pouvant impacter la construction de la mesure empirique  $\hat{\mu}_{ZV}[\mathbf{X}_{\text{obs}}]$ . La maîtrise de la variabilité d'échantillonnage théorique (c'est-à-dire celle émanant de  $\mu_X$ ) peut être améliorée autant qu'on le souhaite à condition que le *bootstrap* se fasse en piochant dans une plus grande base d'observations existantes (et pas uniquement parmi celles initialement contenues dans  $\mathbf{x}_{\text{obs}}$ ). Évidemment, cette démarche est purement artificielle et sert essentiellement à montrer que le caractère "*data-driven*" des indices produits par l'ASF n'est pas un défaut de la méthodologie proposée mais plutôt une limite intrinsèque du problème dont on ne peut s'affranchir sauf, et c'est ce qui est fait dans la suite, si on s'autorise à raffiner la connaissance des mesures empiriques.

Notons  $\mathbf{X}_{\text{base}}$  une matrice aléatoire comportant  $N''$  observations de la loi conjointe  $\mu_X$  où  $N''$  peut être choisi arbitrairement grand.  $\mathbf{X}_{\text{base}}$  joue le rôle d'un réservoir d'atomes permettant d'affiner la connaissance des différentes mesures mises en jeu dans le mécanisme  $\mathcal{B}$ . La juxtaposition verticale de  $\mathbf{X}_{\text{obs}}$  et  $\mathbf{X}_{\text{base}}$  donne naissance à la matrice  $\mathbf{X}_{\text{raff}}$  qui va servir à construire une version empirique raffinée de  $\mu_{ZV}$ . Pour une matrice  $\mathbf{x}_{\text{raff}}$  fixée, on a besoin de calculer les pseudo-observations  $\hat{\mathbf{u}}_{\text{raff}}$  associées. Il suffit alors d'adapter ce qui a été fait à l'équation (4.46) en incorporant désormais  $N + N''$  atomes :

$$\mathbf{x}_{\text{raff}} = \left[ a_j^{(i)} \right]_{\substack{1 \leq i \leq N+N'' \\ 1 \leq j \leq d}} \quad \text{et} \quad \hat{\mathbf{u}}_{\text{raff}} = \left[ \hat{u}_j^{(i)} \right]_{\substack{1 \leq i \leq N+N'' \\ 1 \leq j \leq d}} \quad \text{avec} \quad \hat{u}_j^{(i)} = \frac{1}{N + N'' + 1} \sum_{l=1}^{N+N''} \mathbb{1}_{\{a_j^{(l)} \leq a_j^{(i)}\}}. \quad (4.83)$$

Les matrices  $\mathbf{x}_{\text{raff}}$  et  $\hat{\mathbf{u}}_{\text{raff}}$  peuvent alors se substituer à  $\mathbf{x}_{\text{obs}}$  et  $\hat{\mathbf{u}}_{\text{obs}}$  pour approcher la loi  $\mu_{ZV}$  :

$$\hat{\mu}_{ZV}[\mathbf{x}_{\text{raff}}] = \underbrace{\hat{\mu}[\mathbf{x}_{\text{raff}}^{\bullet 1}]}_{\text{pour tout } Z_1} \otimes \dots \otimes \underbrace{\hat{\mu}[\mathbf{x}_{\text{raff}}^{\bullet d}]}_{\text{pour tout } Z_d} \otimes \underbrace{\hat{\mu}[\hat{\mathbf{u}}_{\text{raff}}]}_{\text{pour } \mathbf{V}}. \quad (4.84)$$

L'idée est d'utiliser  $\hat{\mu}_{ZV}[\mathbf{x}_{\text{raff}}]$  pour fabriquer des échantillons d'apprentissage  $(\tilde{\mathbf{Z}}, \tilde{\mathbf{V}})$  qui soient plus à même d'imiter le comportement de  $(\mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}})$ . Chaque entité ré-échantillonnée continue à être constituée de  $N$  observations mais celles-ci sont puisées dans un stock totalisant  $N + N''$  atomes. Cette nouvelle procédure de ré-échantillonnage est notée  $\mathcal{B}_{\text{raff}}$  et respecte le schéma suivant :

$$\mathcal{B}_{\text{raff}}: \begin{array}{ccc} \mathbb{R}^{N \times d} & \times & \mathbb{R}^{N'' \times d} & \longrightarrow & \mathbb{R}^N \times \dots \times \mathbb{R}^N & \times & [0, 1]^N \\ \mathbf{x}_{\text{obs}} & & \mathbf{x}_{\text{base}} & \longmapsto & \tilde{\mathbf{z}}_1, \dots, \tilde{\mathbf{z}}_d & & \tilde{\mathbf{v}} \end{array}. \quad (4.85)$$

Elle s'exécute selon les étapes suivantes :

(B4a) Pour  $1 \leq j \leq d$ , tirer un N-échantillon  $\tilde{\mathbf{z}}_j$  suivant  $\hat{\mu}[\mathbf{x}_{\text{raff}}^{[\bullet, j]}]$ .

(B4b) Calculer la matrice des pseudo-observations  $\hat{\mathbf{u}}_{\text{raff}}$  avec la formule de (4.83).

(B4c) Tirer un N-échantillon  $\tilde{\mathbf{v}}$  suivant  $\hat{\mu}[\hat{\mathbf{u}}_{\text{raff}}]$ .

En important artificiellement un stock  $\mathbf{x}_{\text{base}}$  d'observations de la loi conjointe, l'objectif poursuivi est celui d'une convergence progressive de la mesure empirique  $\hat{\mu}_{\text{ZV}}[\mathbf{x}_{\text{raff}}]$  vers la mesure théorique  $\mu_{\text{ZV}}$ . On s'attend en effet à ce que  $\hat{\mu}_{\text{ZV}}[\mathbf{X}_{\text{raff}}]$  soit moins sensible à la variabilité d'échantillonnage que  $\hat{\mu}_{\text{ZV}}[\mathbf{X}_{\text{obs}}]$ . Le risque de surapprentissage de  $\mu_{\text{ZV}}$  évolue en effet dans le même sens que la sévérité du contexte "small data". Par ailleurs, le biais induit sur les indices de Sobol doit également s'effacer lorsque la taille  $N''$  augmente. Pour valider cette logique, on se propose de conduire l'expérimentation numérique suivante :

1. Récupérer les échantillons fabriqués précédemment :  $\mathbf{x}_{\text{obs}}^{(k)}$  avec  $k \in \{1, 2, 3\}$ .
2. Simuler un stock pour chaque échantillon :  $\mathbf{x}_{\text{base}}^{(k)}$  avec  $N'' \in \{10^3, 10^4, 10^5\}$ .
3. Implanter  $\mathcal{B}_{\text{raff}}$  dans l'algorithme d'ASF.
4. Estimer les indices de Sobol dans tous les cas :  $(k, N'') \in \{1, 2, 3\} \times \{10^3, 10^4, 10^5\}$ .

Une représentation schématique de la démarche est proposée à la Figure 4.12 pour une meilleure visualisation de la stratégie mise en œuvre. Les résultats sont présentés dans le Tableau 4.2 où l'on constate une disparition du caractère "data-driven" à mesure que le nombre d'atomes impliqués dans le raffinement de la mesure empirique  $\hat{\mu}[\mathbf{x}_{\text{obs}}]$  augmente. Lorsque  $N'' = 10^5$ , on peut voir que tous les échantillons partagent la même hiérarchie au sein du triplet d'indices dominants, à savoir  $\hat{S}_1 \approx \hat{S}_2 > \hat{S}_{\text{cop}}$ , et qu'elle correspond au classement initialement obtenu lors de l'utilisation des générateurs  $G_{\text{Z}}$  et  $G_{\text{V}}$ . Les valeurs moyennes des indices sont par ailleurs très proches des valeurs théoriques (4.80), ce qui indique que toutes les mesures empiriques  $\hat{\mu}_{\text{ZV}}[\mathbf{x}_{\text{raff}}]$  ont pratiquement convergé vers  $\mu_{\text{ZV}}$  à ce stade.

Deux idées fondamentales peuvent alors être dégagées de cette expérimentation. Tout d'abord, il est inutile d'espérer mettre en place une technique de ré-échantillonnage permettant de retrouver avec exactitude les résultats d'une ASF hypothétiquement pratiquée avec des générateurs théoriques. Le contexte "small data" propre à  $\mathbf{x}_{\text{obs}}$  nous en empêche et le volume de données supplémentaires à collecter pour tenter d'éradiquer un éventuel phénomène de surapprentissage est tout à fait déraisonnable. Ceci étant dit, il faut cependant reconnaître que le mécanisme  $\mathcal{B}$  s'avère être une solution assez efficace. Elle permet de scinder la collection des indices de Sobol en plusieurs lots, chacun d'eux regroupant les indices des composantes fonctionnelles exerçant globalement la même influence sur la variabilité de la prédiction du risque de défaillance. On peut donc estimer que, de ce point de vue, le déploiement de l'algorithme d'ASF présenté à la Section 4.2.3.2 répond au Problème 4.2 de décomposition de la variabilité d'apprentissage.

#### 4.3.1.6 Assimilation de données supplémentaires

Pour s'assurer que les résultats de l'ASF rendent bien compte de la variabilité engendrée par l'apprentissage des différentes composantes fonctionnelles, on a pu voir à la Section 4.2.4 qu'une stratégie de validation consiste à conduire une seconde ASF après avoir réduit la variabilité d'une des composantes par acquisition ciblée de données supplémentaires pour cette unique entité. Au cours de la seconde phase d'ASF, la simulation des échantillons d'apprentissage s'effectue à l'aide des mécanismes  $\mathcal{B}_j$  et  $\mathcal{B}_{\text{cop}}$  respectivement développés aux Sections 4.2.4.1 et 4.2.4.2 et permettant de tenir compte de toutes les données disponibles au moment du ré-échantillonnage. On expérimente tous les scénarios possibles. Pour cela, on va :

1. Sélectionner une composante fonctionnelle :  $\hat{f}_{X_j}$  ou  $\hat{\mathbf{u}}$
2. Simuler des observations supplémentaires :  $\mathbf{x}_{\text{extra}}^{[\bullet, j]}$  ou  $\mathbf{x}_{\text{extra}}$
3. Implanter le mécanisme de ré-échantillonnage adapté :  $\mathcal{B}_j$  ou  $\mathcal{B}_{\text{cop}}$
4. Estimer les indices et les confronter à la collection initiale.

		ASF exécutée avec $\mathcal{B}_{\text{raff}}$ :					
			$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_{\text{cop}}$	
Échantillons	$\mathbf{x}_{\text{obs}}^{(1)} \cup \mathbf{x}_{\text{base}}^{(1)}$	$N'' = 0$	moyenne :	28.26 %	26.56 %	5.03 %	37.49 %
			écart-type :	4.18 %	4.35 %	3.85 %	4.19 %
			rang :	(2)	(3)	(4)	(1)
		$N'' = 10^3$	moyenne :	30.17 %	36.00 %	5.85 %	28.25 %
			écart-type :	3.83 %	3.40 %	3.63 %	3.06 %
			rang :	(2)	(1)	(4)	(3)
		$N'' = 10^4$	moyenne :	32.61 %	33.18 %	6.76 %	29.04 %
			écart-type :	3.58 %	3.73 %	3.92 %	4.37 %
			rang :	(2)	(1)	(4)	(3)
	$N'' = 10^5$	moyenne :	33.67 %	33.67 %	6.59 %	29.20 %	
		écart-type :	1.88 %	3.16 %	2.84 %	2.18 %	
		rang :	(1)	(1)	(4)	(3)	
	$\mathbf{x}_{\text{obs}}^{(2)} \cup \mathbf{x}_{\text{base}}^{(2)}$	$N'' = 0$	moyenne :	28.23 %	32.52 %	3.78 %	30.75 %
			écart-type :	3.82 %	3.77 %	3.62 %	2.84 %
			rang :	(3)	(1)	(4)	(2)
		$N'' = 10^3$	moyenne :	30.52 %	34.16 %	5.16 %	27.77 %
			écart-type :	2.12 %	3.11 %	3.29 %	3.48 %
			rang :	(2)	(1)	(4)	(3)
		$N'' = 10^4$	moyenne :	32.14 %	34.07 %	6.91 %	29.38 %
			écart-type :	4.48 %	3.88 %	3.83 %	3.60 %
			rang :	(2)	(1)	(4)	(3)
		$N'' = 10^5$	moyenne :	32.99 %	32.60 %	4.31 %	26.43 %
			écart-type :	2.80 %	2.71 %	2.52 %	3.79 %
			rang :	(1)	(1)	(4)	(3)
	$\mathbf{x}_{\text{obs}}^{(3)} \cup \mathbf{x}_{\text{base}}^{(3)}$	$N'' = 0$	moyenne :	38.18 %	27.13 %	7.52 %	28.92 %
			écart-type :	2.34 %	3.88 %	3.41 %	3.91 %
			rang :	(1)	(3)	(4)	(2)
		$N'' = 10^3$	moyenne :	35.75 %	28.21 %	4.92 %	29.23 %
			écart-type :	2.18 %	1.89 %	1.96 %	3.02 %
			rang :	(1)	(3)	(4)	(2)
		$N'' = 10^4$	moyenne :	33.85 %	32.52 %	6.77 %	30.86 %
			écart-type :	3.70 %	1.97 %	2.70 %	2.60 %
			rang :	(1)	(2)	(4)	(3)
		$N'' = 10^5$	moyenne :	31.89 %	34.40 %	5.24 %	28.27 %
			écart-type :	2.18 %	2.77 %	2.65 %	3.43 %
			rang :	(1)	(2)	(4)	(3)

TABLEAU 4.2 – Convergence des indices de Sobol après raffinement des mesures empiriques.

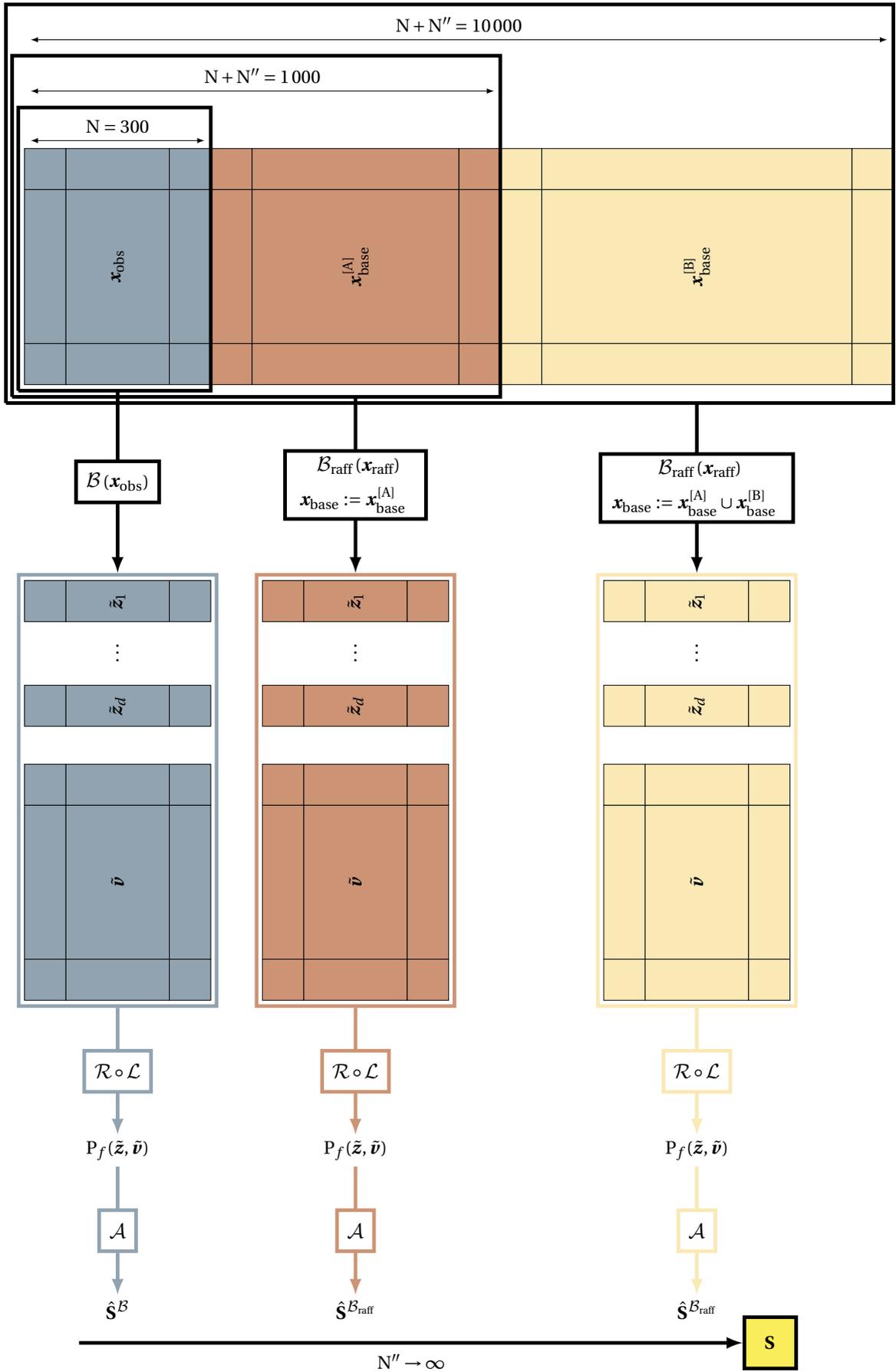


FIGURE 4.12 – Illustration du processus de raffinement des mesures empiriques.

Les résultats sont consignés dans le Tableau 4.3 après avoir mené toutes les expérimentations. Le premier étage contient les résultats de l'ASF pour l'échantillon  $\mathbf{x}_{\text{obs}}^{(1)}$  déjà utilisé dans le Tableau 4.1 et cette collection d'indices sert de référence pour étudier l'évolution des résultats de l'ASF après assimilation de nouvelles données. On rappelle que, pour la configuration traitée dans cette présente section, les incertitudes épistémiques se restreignent à l'incertitude induite par le manque de données. Il n'y a pas d'incertitude quant au modèle de distribution à utiliser pour l'apprentissage statistique. On s'attend donc à ce que l'impact d'un apport, éventuellement massif, de données supplémentaires soit clairement visible.

Le second étage du Tableau 4.3 présente les résultats obtenus pour l'assimilation d'un nombre  $N'$  croissant d'observations marginales de la variable  $X_1$ . On remarque que :

$$\hat{S}_1^{\mathcal{B}_1} < \hat{S}_1^{\mathcal{B}} \quad \text{tandis que} \quad \hat{S}_2^{\mathcal{B}_1} > \hat{S}_2^{\mathcal{B}} \quad ; \quad \hat{S}_{\text{cop}}^{\mathcal{B}_1} > \hat{S}_{\text{cop}}^{\mathcal{B}} \quad \text{et} \quad \hat{S}_3^{\mathcal{B}_1} \approx \hat{S}_3^{\mathcal{B}} . \quad (4.86)$$

D'ailleurs, cette tendance s'accroît au fur et à mesure de l'accroissement du nombre  $N'$  de données supplémentaires assimilées. L'indice  $\hat{S}_1^{\mathcal{B}_1}$  initialement membre du triplet dominant dégringole dans les classements jusqu'à atteindre la dernière position. À ce stade, la calibration de  $\hat{f}_{X_1}$  s'appuie sur un si grand nombre d'observations de  $X_1$  qu'il ne reste plus qu'une toute petite part de variabilité d'apprentissage à propager sur l'estimateur de la probabilité de défaillance. En ce qui concerne les autres indices, les positions de  $\hat{S}_2^{\mathcal{B}_1}$  et  $\hat{S}_{\text{cop}}^{\mathcal{B}_1}$  sont renforcées, avec une prédominance de  $\hat{S}_{\text{cop}}^{\mathcal{B}_1}$ , tandis que celle de  $\hat{S}_3^{\mathcal{B}_1}$  demeure inchangée. Cette évolution des classements, conforme à nos attentes, semble indiquer que les indices de sensibilité calculés rendent bien compte de l'importance des contributions respectives des composantes fonctionnelles. Cela conforte l'idée que l'algorithme d'ASF apporte une réponse pertinente au Problème 4.2 de décomposition de la variabilité affectant la prédiction du risque de défaillance. On note également que le phénomène de neutralisation de la composante enrichie s'amorce assez tôt. Il est par exemple visible lorsque l'on assimile  $N' = 500$  nouvelles observations de  $X_1$ . L'idée d'une réduction de la variabilité d'apprentissage par acquisition ciblée évoquée dans le Problème 4.3 n'est donc pas vaine et peut être mise en œuvre. Il reste alors à juger, selon le contexte, si la collecte de plusieurs centaines d'observations d'une variable est une requête coûteuse ou non.

Les étages inférieurs du Tableau 4.3 sont consacrés aux autres scénarios d'assimilation. Dans le cas de  $X_2$ , une évolution symétrique à celle décrite pour  $X_1$  est observable :

$$\hat{S}_2^{\mathcal{B}_2} < \hat{S}_2^{\mathcal{B}} \quad \text{tandis que} \quad \hat{S}_1^{\mathcal{B}_2} > \hat{S}_1^{\mathcal{B}} \quad ; \quad \hat{S}_{\text{cop}}^{\mathcal{B}_2} > \hat{S}_{\text{cop}}^{\mathcal{B}} \quad \text{et} \quad \hat{S}_3^{\mathcal{B}_2} \approx \hat{S}_3^{\mathcal{B}} . \quad (4.87)$$

Les indices  $\hat{S}_1^{\mathcal{B}_2}$  et  $\hat{S}_{\text{cop}}^{\mathcal{B}_2}$  absorbent de plus grandes parts de la variabilité restante. Ces remarques restent valables au dernier étage du tableau, lorsqu'on cherche à réduire l'influence exercée par l'apprentissage de  $\mathbf{c}_{\mathcal{U}}$ . Par contre, pour ce qui est de l'assimilation d'observations supplémentaires pour  $X_3$ , la situation est légèrement différente. Sachant que  $\hat{S}_3^{\mathcal{B}}$  est déjà le plus petit indice, sa diminution est moins visible au gré de l'apport de nouvelles données et le triplet dominant voit sa primauté renforcée encore davantage.

#### 4.3.1.7 Bilan

Dans cette section, on a étudié un cas test très simple combinant une distribution sous-jacente  $\mu_{\mathbf{X}}$  gaussienne, un modèle d'apprentissage  $\mathcal{M}$  paramétrique gaussien et une fonction de performance  $\phi$  linéaire. Ce cadre rend possible un calcul exact de toutes les probabilités de défaillance à l'aide d'un algorithme de fiabilité déterministe  $\mathcal{R}$ . Chacune des briques impliquées dans l'algorithme d'ASF a alors pu être décortiquée et validée. On retiendra les éléments suivants :

- Pour ce cas test, une estimation de type *Pick-Freeze* des indices de Sobol requiert  $M = 10^3$  itérations pour atteindre un niveau suffisant de précision et de robustesse.
- Les résultats de la procédure d'ASF souffrent d'un léger biais introduit par l'échantillon  $\mathbf{x}_{\text{obs}}$  fourni. La solution de ré-échantillonnage proposée reste cependant "optimale" au sens où elle fait au mieux dans un contexte de "*small data*".

ASF exécutée avec $\mathcal{B}$ :				$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_{\text{cop}}$
	$\mathbf{x}_{\text{obs}}^{(1)}$		moyenne :	28.26 %	26.56 %	5.03 %	37.49 %
			écart-type :	4.18 %	4.35 %	3.85 %	4.19 %
			rang :	(2)	(3)	(4)	(1)
ASF exécutée avec $\mathcal{B}_1$ :				$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_{\text{cop}}$
Observations de $X_1$	$\mathbf{x}_{\text{obs}}^{(1)} \cup \mathbf{x}_{\text{extra}}^{[\bullet 1]}$	$N' = 500$	moyenne :	14.65 %	30.94 %	4.91 %	44.99 %
			écart-type :	2.82 %	3.82 %	3.58 %	3.69 %
			rang :	(3)	(2)	(4)	(1)
		$N' = 10^3$	moyenne :	12.68 %	32.72 %	7.08 %	49.37 %
			écart-type :	2.69 %	3.58 %	3.39 %	2.95 %
			rang :	(3)	(2)	(4)	(1)
$N' = 10^4$	moyenne :	2.92 %	35.34 %	6.66 %	55.21 %		
	écart-type :	0.61 %	2.91 %	3.28 %	3.27 %		
	rang :	(4)	(2)	(3)	(1)		
ASF exécutée avec $\mathcal{B}_2$ :				$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_{\text{cop}}$
Observations de $X_2$	$\mathbf{x}_{\text{obs}}^{(1)} \cup \mathbf{x}_{\text{extra}}^{[\bullet 2]}$	$N' = 500$	moyenne :	31.51 %	14.26 %	5.31 %	44.68 %
			écart-type :	3.07 %	3.23 %	1.80 %	2.06 %
			rang :	(2)	(3)	(4)	(1)
		$N' = 10^3$	moyenne :	33.61 %	10.37 %	6.42 %	49.58 %
			écart-type :	3.52 %	2.09 %	2.43 %	3.40 %
			rang :	(2)	(3)	(4)	(1)
$N' = 10^4$	moyenne :	37.99 %	3.34 %	6.99 %	53.16 %		
	écart-type :	2.87 %	1.85 %	2.63 %	3.00 %		
	rang :	(2)	(4)	(3)	(1)		
ASF exécutée avec $\mathcal{B}_3$ :				$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_{\text{cop}}$
Observations de $X_3$	$\mathbf{x}_{\text{obs}}^{(1)} \cup \mathbf{x}_{\text{extra}}^{[\bullet 3]}$	$N' = 500$	moyenne :	28.79 %	29.50 %	4.01 %	40.94 %
			écart-type :	3.51 %	2.36 %	2.41 %	2.11 %
			rang :	(2)	(2)	(3)	(1)
		$N' = 10^3$	moyenne :	29.27 %	29.79 %	3.79 %	39.48 %
			écart-type :	3.33 %	3.29 %	2.74 %	2.32 %
			rang :	(2)	(2)	(3)	(1)
$N' = 10^4$	moyenne :	30.12 %	27.87 %	2.18 %	40.52 %		
	écart-type :	2.97 %	2.37 %	1.37 %	2.62 %		
	rang :	(2)	(3)	(4)	(1)		
ASF exécutée avec $\mathcal{B}_{\text{cop}}$ :				$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_{\text{cop}}$
Observations de $\mathbf{U}$	$\mathbf{x}_{\text{obs}}^{(1)} \cup \mathbf{x}_{\text{extra}}$	$N' = 500$	moyenne :	37.86 %	36.86 %	6.31 %	18.09 %
			écart-type :	1.97 %	3.70 %	3.26 %	3.73 %
			rang :	(1)	(1)	(3)	(2)
		$N' = 10^3$	moyenne :	41.76 %	39.31 %	8.45 %	12.24 %
			écart-type :	3.48 %	2.98 %	3.94 %	3.17 %
			rang :	(1)	(2)	(4)	(3)
$N' = 10^4$	moyenne :	46.39 %	43.86 %	8.14 %	1.59 %		
	écart-type :	2.61 %	3.05 %	1.75 %	1.47 %		
	rang :	(1)	(2)	(3)	(4)		

TABLEAU 4.3 – Impact de l'assimilation de données supplémentaires sur le classement des indices de Sobol.

- Des expérimentations numériques ont montré la possibilité d'une réduction de la variabilité d'apprentissage par acquisition ciblée puis assimilation de nouvelles observations.

### 4.3.2 Étude d'un cas test non-linéaire

#### 4.3.2.1 Description

Soit  $\mathbf{X} = (X_1, X_2)$  un couple de variables aléatoires identiquement distribuées. Leur loi commune est celle d'un mélange uniforme de deux lois gaussiennes. Par ailleurs, une copule gaussienne sous-tend la corrélation entre ces deux variables. En vertu du théorème de Sklar, la densité conjointe  $f_{\mathbf{X}}$  est construite en réunissant les composantes spécifiées ci-dessous :

$$\begin{aligned} & \bullet f_{X_1}(z) = f_{X_2}(z) = \frac{1}{2} (f_{\mathcal{N}}(z | m_1, \sigma_1) + f_{\mathcal{N}}(z | m_2, \sigma_2)) \\ & \bullet c_{\mathbf{U}}(\mathbf{u}) = c_{\mathcal{N}}(\mathbf{u} | \rho) \end{aligned} \quad \text{avec} \quad \begin{cases} m_2 = -m_1 = m = 2 \\ \sigma_1 = \sigma_2 = \sigma = 1 \\ \rho = 0.5 \end{cases} . \quad (4.88)$$

Comme on peut le voir sur la Figure 4.13a,  $f_{\mathbf{X}}$  est quadrimodale. L'existence d'une corrélation positive entre  $X_1$  et  $X_2$  accroît l'hégémonie des modes situés sur la bissectrice  $x_2 = x_1$  comparativement à ceux situés sur sa perpendiculaire. La variable de sortie  $Y$  est ici donnée par :

$$Y = \phi(\mathbf{X}) = (X_2 - \alpha) + \min_{i \in \{1,2\}} a_i (X_1 - \alpha) \quad \text{avec} \quad a_1 = \frac{1}{a_2} = \sqrt{3} . \quad (4.89)$$

La probabilité de défaillance est définie par  $P_f = \mathbb{P}(Y \geq 0)$  et le domaine de défaillance peut être décrit de la manière suivante :

$$D_f = \{\mathbf{x} \in \mathbb{R}^2 : x_2 \geq \psi_1(x_1)\} \quad \text{avec} \quad \psi_1(x_1) = \begin{cases} p_a(x_1) = -a_1(x_1 - \alpha) + \alpha & \text{si } x_1 < \alpha \\ p_b(x_1) = -a_2(x_1 - \alpha) + \alpha & \text{si } x_1 > \alpha \end{cases} . \quad (4.90)$$

Ainsi, l'état limite  $\phi(\mathbf{x}) = 0$  peut s'exprimer comme une fonction linéaire par morceaux de la variable  $x_1$ . En remarquant que  $\psi_1$  est continue et strictement décroissante, on comprend qu'il est aussi tout à fait possible d'écrire que :

$$D_f = \{\mathbf{x} \in \mathbb{R}^2 : x_1 \leq \psi_2(x_2)\} \quad \text{avec} \quad \psi_2 = \psi_1^{-1} . \quad (4.91)$$

De par les valeurs choisies pour  $a_1$  et  $a_2$ , il est facile de montrer que :

$$\phi(\mathbf{x}) \geq 0 \iff x_2 \geq \alpha - \min_{i \in \{1,2\}} a_i (x_1 - \alpha) \iff x_1 \geq \alpha - \min_{i \in \{1,2\}} a_i (x_2 - \alpha) . \quad (4.92)$$

On en déduit que  $\psi_1 = \psi_2$  et on en conclut que  $D_f$  est symétrique par rapport à la droite  $x_2 = x_1$ . Sur la Figure 4.13a, on dispose d'un aperçu visuel du positionnement et de la géométrie de  $D_f$ . Le fait de travailler en dimension  $d = 2$  et de disposer des paramétrages (4.90) et (4.91) rend possible un calcul exact de  $P_f$  par le biais d'une intégration numérique. La suite de ce paragraphe s'inspire directement de ce qui est exposé dans TANG et collab. [2013]. L'idée centrale repose sur la séquence d'égalités suivante :

$$\begin{aligned} P_f &= \int_{D_f} f_{X_1}(x_1) f_{X_2}(x_2) c_{\mathbf{U}}(F_{X_1}(x_1), F_{X_2}(x_2)) \, d\mathbf{x} \\ &= \int_{\mathbb{R}} f_{X_1}(x_1) \left( \int_{\psi_1(x_1)}^{+\infty} f_{X_2}(x_2) \frac{\partial^2 C_{\mathbf{U}}}{\partial u \partial v}(F_{X_1}(x_1), F_{X_2}(x_2)) \, dx_2 \right) dx_1 \\ &= \int_{\mathbb{R}} f_{X_1}(x_1) \left[ \frac{\partial C_{\mathbf{U}}}{\partial u}(F_{X_1}(x_1), F_{X_2}(z)) \right]_{z=\psi_1(x_1)}^{+\infty} dx_1 \\ &= \int_{\mathbb{R}} f_{X_1}(x_1) \left( 1 - \frac{\partial C_{\mathbf{U}}}{\partial u}(F_{X_1}(x_1), F_{X_2}(\psi_1(x_1))) \right) dx_1 \\ &= 1 - \int_{\mathbb{R}} f_{X_1}(x_1) \frac{\partial C_{\mathbf{U}}}{\partial u}(F_{X_1}(x_1), F_{X_2}(\psi_1(x_1))) dx_1 . \end{aligned} \quad (4.93)$$

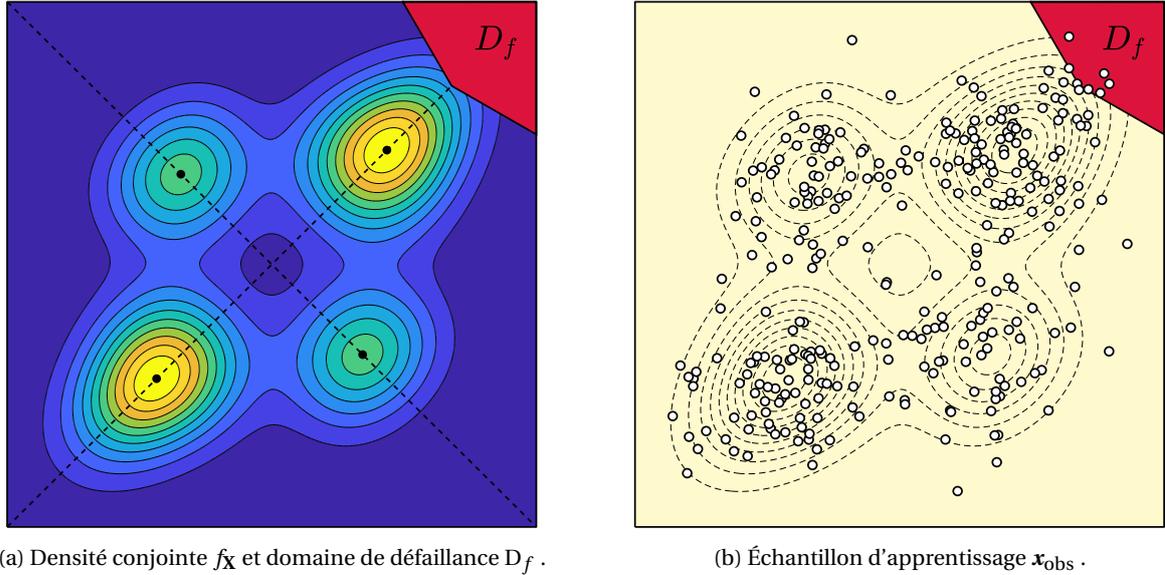


FIGURE 4.13 – Visualisation du problème d'analyse de fiabilité.

Si on connaît l'expression analytique de  $\partial C_{\mathbf{U}}/\partial u$ , ou si on a accès à une routine numérique permettant d'en évaluer les valeurs, l'intégrande est parfaitement connue. L'utilisation d'une méthode de quadrature telle que celle de **ROMBERG [1955]** donne alors une estimation de  $P_f$ . Sachant que l'erreur commise est très nettement inférieure à celle de n'importe quelle technique d'échantillonnage, on admet que le résultat fourni par l'intégration numérique est exact. En utilisant le paramétrage (4.91), l'élimination de la variable  $x_1$  aurait conduit à la formule :

$$P_f = 1 - \int_{\mathbb{R}} f_{X_2}(x_2) \frac{\partial C_{\mathbf{U}}}{\partial v} \left( F_{X_1}(\psi_2(x_2)), F_{X_2}(x_2) \right) dx_2 . \quad (4.94)$$

Ainsi, pour ce cas test, la faible dimension et la connaissance d'un paramétrage de l'état limite permettent la construction ad hoc d'un algorithme de fiabilité  $\mathcal{R}$  dont le fonctionnement est déterministe. On pose  $\alpha = k \times m$  dans l'équation (4.89) et on rappelle que  $m$  est la moyenne de  $X_2$  et l'opposé de celle de  $X_1$ . En réglant  $k$ , on peut fixer  $P_f$  au niveau de criticité souhaité pour l'étude sans toutefois perturber la symétrie de  $D_f$ . Pour  $k = 1.7$ , on obtient  $P_f = 0.016$  après avoir appliqué la démarche présentée ci-dessus.

#### 4.3.2.2 Hybridation du modèle d'apprentissage

Pour ce cas test, on souhaite comparer les résultats de l'ASF pour différentes combinaisons de modèles d'apprentissage des marginales et de la copule. Pour simplifier l'analyse des résultats, la copule est systématiquement estimée au sein de la famille gaussienne.  $\mathcal{M}_{\text{cop}}$  est donc fixé et on se contente d'intervenir sur les modèles  $\mathcal{M}_1$  et  $\mathcal{M}_2$  d'apprentissage des lois marginales. L'échantillon initial comporte  $N = 500$  observations. L'apprentissage de la densité marginale  $f_{X_1}$  à partir d'un échantillon  $\mathbf{Z}_{\text{obs}}^{[\bullet 1]}$  est ici envisagée de trois manières différentes : la première est naïve, la seconde s'appuie sur une pratique courante tandis que la troisième cherche à tirer profit d'une analyse visuelle préalable des données. En procédant de la sorte, on cherche à faire évoluer la nature et l'amplitude de la variabilité portée par la composante fonctionnelle  $\hat{f}_{X_1}(\cdot | \mathcal{M}_1, \mathbf{Z}_{\text{obs}}^{[\bullet 1]})$  en fonction du choix de  $\mathcal{M}_1$ .

La première technique n'est rien d'autre qu'une calibration  $\hat{f}_{\mathcal{N}}(\cdot | \hat{m}_1^{\text{MV}}, \hat{\sigma}_1^{\text{MV}})$  de la densité gaussienne aux observations marginales disponibles. Ce choix est infondé, extrêmement maladroit et conduit à l'apprentissage d'une loi normale unimodale dont l'espérance est située approximativement à mi-chemin entre les pics de la densité théorique  $f_{X_1}$ . C'est ce que l'on peut observer à la Figure 4.14a où le faisceau de courbes représente les résultats de plusieurs dizaines

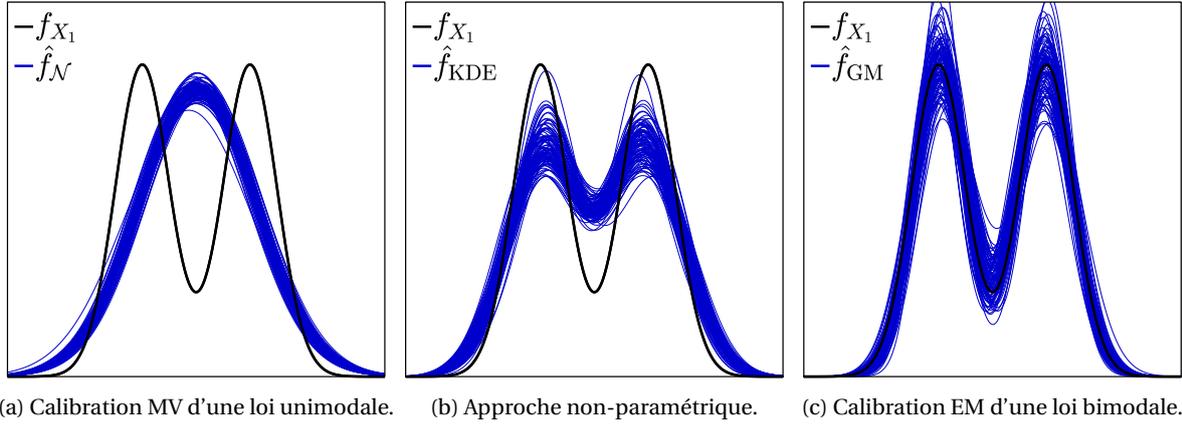


FIGURE 4.14 – Évolution de la variabilité affectant l'apprentissage de la densité marginale  $f_{X_1}$ .

d'estimations de  $f_{X_1}$  à partir d'échantillons  $\mathbf{Z}_{\text{obs}}^{[\bullet|1]}$  simulés avec  $G_{\mathbf{Z}}$ . En guise de seconde technique, pour pallier au manque de flexibilité observé précédemment, on propose de se tourner vers l'estimation par noyaux gaussiens :

$$\hat{f}_{X_1}(x_1 | \mathcal{M}_1, \mathbf{Z}_{\text{obs}}^{[\bullet|j]}) = \hat{f}_{\text{KDE}}(x_1 | \mathbf{Z}_{\text{obs}}^{[\bullet|1]}, \hat{h}_1) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\hat{h}_1} \mathcal{K}_{\mathcal{N}}\left(\frac{x_1 - Z_1^{(i)}}{\hat{h}_1}\right), \quad (4.95)$$

avec  $\mathcal{K}_{\mathcal{N}}(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$  et  $\hat{h}_1 := 1.06 \times \hat{\sigma}_1 \times N^{-\frac{1}{5}}$ .

La fenêtre de lissage  $\hat{h}_1$  est estimée de façon naïve avec la règle de SILVERMAN [1986] où  $\hat{\sigma}_1$  est l'écart-type empirique calculé à partir des observations contenues dans  $\mathbf{Z}_{\text{obs}}^{[\bullet|1]}$ . Comme on peut le constater à la Figure 4.14b, cette approche non-paramétrique "aveugle" permet de capturer la bimodalité étant donné que chaque densité  $\hat{f}_{\text{KDE}}(x_1 | \mathbf{Z}_{\text{obs}}^{[\bullet|1]}, \hat{h}_1)$  présente deux maxima locaux et un minimum local. Toutefois, on constate que le lissage produit par la méthode KDE tend à freiner l'amplitude des fluctuations de la densité estimée. À taille d'échantillon fixée, on ne peut pas espérer faire mieux. La faible vitesse de convergence (vers la densité marginale théorique) est en effet l'un des inconvénients majeurs de l'approche KDE. Pour améliorer la capacité d'apprentissage de la structure bimodale, une solution peut être l'estimation d'un mélange. Lorsque celui-ci n'implique que des lois gaussiennes, la distribution résultante est un mélange gaussien, ou *Gaussian Mixture* (GM). En examinant l'histogramme de  $\mathbf{Z}_{\text{obs}}^{[\bullet|1]}$ , on peut facilement conjecturer qu'un mélange de deux lois est suffisant même s'il serait tout à fait envisageable d'en incorporer davantage. La densité marginale calibrée au cours de l'apprentissage s'écrit alors :

$$\begin{aligned} \hat{f}_{X_1}(x_1 | \mathcal{M}_1, \mathbf{Z}_{\text{obs}}^{[\bullet|j]}) &= \hat{f}_{\text{GM}}(x_1 | \hat{\mathbf{w}}, \hat{\mathbf{m}}, \hat{\boldsymbol{\sigma}}) \\ &= \hat{w}_A \times \hat{f}_{\mathcal{N}}(x_1 | \hat{m}_A, \hat{\sigma}_A) + \hat{w}_B \times \hat{f}_{\mathcal{N}}(x_1 | \hat{m}_B, \hat{\sigma}_B), \end{aligned} \quad (4.96)$$

où les vecteurs  $\hat{\mathbf{w}}$ ,  $\hat{\mathbf{m}}$  et  $\hat{\boldsymbol{\sigma}}$  sont estimés à l'aide d'un algorithme *Expectation-Maximization* (EM) présenté en détail dans MCLACHLAN et PEEL [2004]. Le schéma itératif tire profit de l'existence d'une variable latente pour améliorer la procédure d'optimisation de la log-vraisemblance. Le faisceau des densités  $\hat{f}_{\text{GM}}(x_1 | \hat{\mathbf{w}}, \hat{\mathbf{m}}, \hat{\boldsymbol{\sigma}})$  représenté à la Figure 4.14c met en exergue un gain de précision dans la reconstruction de  $f_{X_1}$  avec l'emploi du modèle GM. Cela se traduit par une amplification des fluctuations entre extrema locaux et un raffinement de l'estimation des queues. Pour ce modèle d'apprentissage, on s'est en fait placé dans la classe de distributions incluant la distribution théorique  $\mu_{\mathbf{X}}$ . La part d'incertitude liée à la méconnaissance du modèle de distribution sous-jacent est ainsi éliminée.

### 4.3.2.3 À propos des indices obtenus avec les générateurs théoriques

En s’inspirant de ce qui a été fait à la Section 4.3.1.3, on veut être capable de comparer les valeurs théoriques des indices  $S_1$  et  $S_2$ . La définition (4.88) de  $\mu_{\mathbf{X}}$  spécifie que  $\mu_{X_1} = \mu_{X_2}$  donc (Sym1) reste vraie. La symétrie de  $D_f$  révélée par l’équation (4.92) garantit (Sym4). De plus, (Sym3) n’existe pas en dimension  $d = 2$ . L’égalité des indices  $S_1$  et  $S_2$  dépend donc uniquement de (Sym2), c’est-à-dire de l’utilisation du même modèle d’apprentissage  $\mathcal{M}_1 = \mathcal{M}_2$  pour l’estimation de  $f_{X_1}$  et  $f_{X_2}$ . Le Tableau 4.4a présente les résultats des procédures d’ASF conduites pour disposer d’estimations des indices de sensibilité théoriques associés à sept combinaisons différentes de sous-modèles  $\mathcal{M}_1$ ,  $\mathcal{M}_2$  et  $\mathcal{M}_{\text{cop}}$ . Il en résulte alors sept modèles d’apprentissage notés  $\mathcal{M}^{(A)}, \dots, \mathcal{M}^{(G)}$ . Les échantillons d’apprentissage sont simulés à partir des générateurs  $G_{\mathbf{Z}}$  et  $G_{\mathbf{V}}$ . Pour chaque nouvelle déclinaison de  $\mathcal{M}$ , les valeurs moyennes et les écarts-types sont obtenus après  $N_{\text{rep}} = 10$  exécutions de l’algorithme d’ASF. Par ailleurs, quels que soient les outils utilisés pour l’apprentissage des lois marginales, le problème du calcul des probabilités de défaillance peut toujours se ramener à celui d’une série d’intégrations numériques. Il suffit d’adapter la formule (4.93) en remplaçant chaque composante fonctionnelle par son estimation :

$$P_f(\mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}}) = 1 - \int_{\mathbb{R}} \hat{f}_{X_1}(x_1 | \mathcal{M}_1, \mathbf{Z}_{\text{obs}}^{[\bullet 1]}) \dots \times \frac{\partial \hat{\mathbf{C}}_{\mathbf{U}}}{\partial u} \left( \hat{F}_{X_1}(x_1 | \mathcal{M}_1, \mathbf{Z}_{\text{obs}}^{[\bullet 1]}), \hat{F}_{X_2}(\psi_1(x_1) | \mathcal{M}_2, \mathbf{Z}_{\text{obs}}^{[\bullet 2]}) | \mathcal{M}_{\text{cop}}, \mathbf{V}_{\text{obs}} \right) dx_1. \quad (4.97)$$

Les collections d’indices présentées au Tableau 4.4a montrent que  $\hat{S}_1 \approx \hat{S}_2$  dès lors que  $\mathcal{M}_1 = \mathcal{M}_2$ . (Sym2) est ainsi vérifiée par les modèles  $\mathcal{M}^{(A)}$ ,  $\mathcal{M}^{(B)}$  et  $\mathcal{M}^{(E)}$ . Au contraire d’importants écarts séparent  $\hat{S}_1$  et  $\hat{S}_2$  pour les modèles hybrides que sont  $\mathcal{M}^{(C)}$ ,  $\mathcal{M}^{(D)}$ ,  $\mathcal{M}^{(F)}$  et  $\mathcal{M}^{(G)}$ . Une discussion plus poussée sur l’évolution des indices en fonction des modèles d’apprentissage est proposée au prochain paragraphe.

### 4.3.2.4 Étude de l’influence du modèle d’apprentissage sur les résultats de l’ASF

On utilise un échantillon d’apprentissage  $\mathbf{x}_{\text{obs}}$  de  $N = 500$  observations à l’image de celui représenté à la Figure 4.13b en début de section. On reproduit les simulations numériques du paragraphe précédent en remplaçant les générateurs théoriques  $G_{\mathbf{Z}}$  et  $G_{\mathbf{V}}$  par le mécanisme de ré-échantillonnage  $\mathcal{B}$ . La comparaison des Tableaux 4.4a et 4.4b peut servir à mesurer l’importance du caractère “*data-driven*” des résultats de la procédure d’ASF. Les indices estimés à partir du seul  $\mathbf{x}_{\text{obs}}$  sont proches de leurs référents théoriques mais on observe un léger biais qui se caractérise par une surestimation de  $\hat{S}_1$  pour tous les modèles d’apprentissage. Cela indique que le ré-échantillonnage *bootstrap* à partir de  $\mathbf{x}_{\text{obs}}^{[\bullet 1]}$  engendre une plus grande variabilité d’apprentissage que le ré-échantillonnage *bootstrap* à partir de  $\mathbf{x}_{\text{obs}}^{[\bullet 2]}$  même si les lois marginales sous-jacentes et les modèles d’apprentissage sont identiques. En dépit de cela, les résultats produits par l’algorithme d’ASF restent satisfaisants car ils permettent à l’utilisateur, par exemple dans le cas des modèles  $\mathcal{M}^{(A)}$ ,  $\mathcal{M}^{(B)}$  et  $\mathcal{M}^{(E)}$ , d’identifier clairement que ce sont les deux lois marginales qui contribuent le plus à la variabilité d’apprentissage et que ces deux contributions sont globalement de même importance.

Pour chacun des modèles d’apprentissage proposés, on représente sur la Figure 4.15 l’histogramme des probabilités  $P_f(\tilde{\mathbf{z}}^{(k)}, \tilde{\mathbf{v}}^{(k)})$  calculées au cours de l’ASF. On peut ainsi apprécier les caractéristiques de la distribution de l’estimateur  $P_f(\tilde{\mathbf{Z}}, \tilde{\mathbf{V}})$ . La Figure 4.15a révèle que l’apprentissage de lois marginales gaussiennes contribue à surestimer lourdement la probabilité de défaillance théorique. Cela se comprend très intuitivement à partir de deux éléments :

- Figure 4.13a  $\Rightarrow$  Le domaine de défaillance  $D_f$  est situé à l’intersection des queues de distributions supérieures de  $f_{X_1}$  et  $f_{X_2}$ .
- Figure 4.14a  $\Rightarrow$  La calibration d’une densité gaussienne entraîne la surestimation systématique de la queue de distribution supérieure de  $f_{X_1}$ .

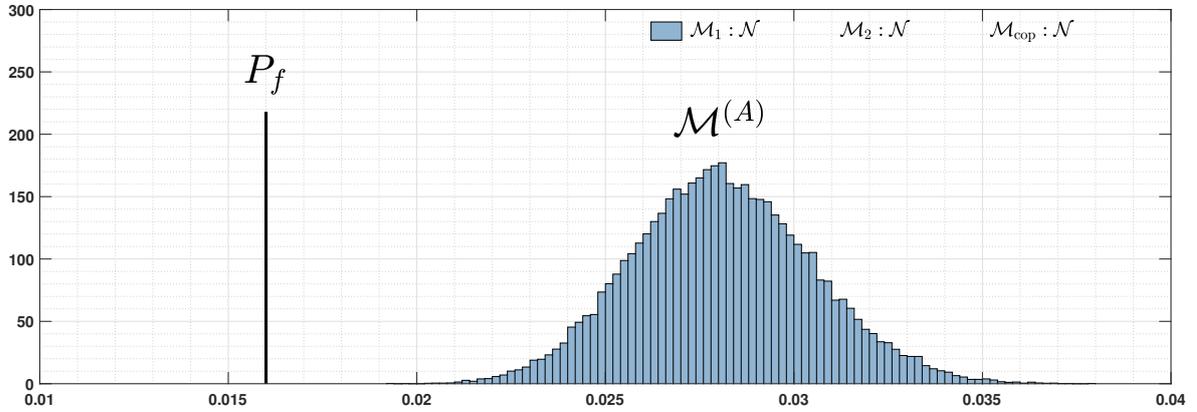
Pour les modèles  $\mathcal{M}^{(B)}$  et  $\mathcal{M}^{(E)}$  où  $\hat{f}_{X_1}$  et  $\hat{f}_{X_2}$  ne sont plus des gaussiennes, les Figures 4.15b et 4.15c

montrent que le biais  $|\mathbb{E}[P_f(\tilde{\mathbf{Z}}, \tilde{\mathbf{V}})] - P_f|$  diminue considérablement. Dans le cas de  $\mathcal{M}^{(B)}$  et d'un apprentissage non-paramétrique des lois marginales, la tendance reste à la surestimation de  $P_f$  et cela concorde avec la Figure 4.14b où l'on constate à nouveau une surestimation des queues supérieures. Le modèle  $\mathcal{M}^{(E)}$  qui reconstruit  $f_{\mathbf{X}}$  selon le modèle initial de distribution est très naturellement le moins biaisé, malgré une légère tendance à la sous-estimation imputable à la dépendance aux données disponibles dans  $\mathbf{x}_{\text{obs}}^{[\bullet, 1]}$ . Les modèles hybrides  $\mathcal{M}^{(C)}$ ,  $\mathcal{M}^{(D)}$ ,  $\mathcal{M}^{(F)}$  et  $\mathcal{M}^{(G)}$ , sont dans une situation intermédiaire où l'apprentissage unimodal de l'une des lois marginales maintient l'existence d'un biais conséquent. Tous ces éléments corroborent l'idée selon laquelle la précision de l'estimation du risque de défaillance est directement tributaire de la qualité de l'algorithme d'estimation de la densité conjointe des entrées.

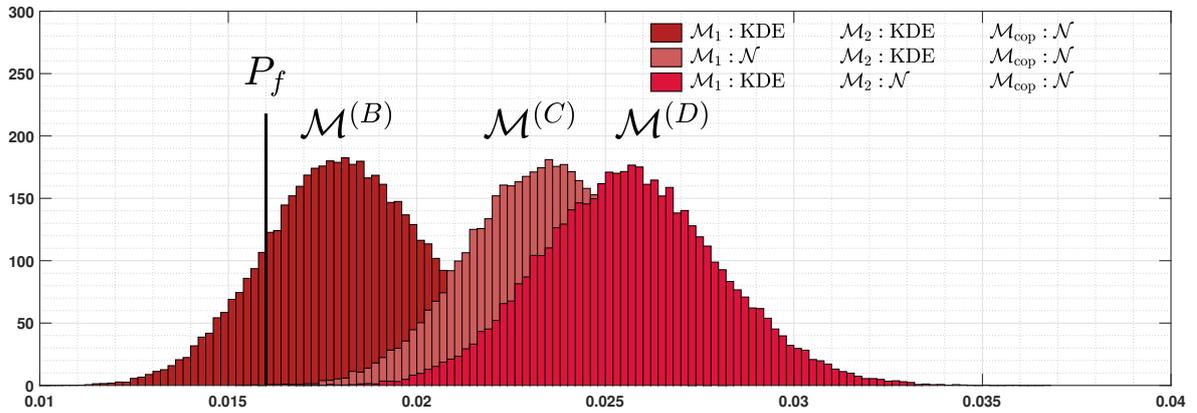
Si l'influence du modèle d'apprentissage  $\mathcal{M}$  est incontestable sur la précision de l'analyse de fiabilité, il est bon de faire remarquer que l'information apportée par la procédure d'ASF n'a pas été développée pour identifier les composantes dont l'apprentissage introduit le plus grand biais. Les indices de sensibilité de Sobol ne servent en effet qu'à quantifier des variances. Par conséquent, ils permettent d'identifier la composante fonctionnelle dont l'apprentissage engendre le plus de variabilité sur la prédiction du risque et en aucun cas la composante fonctionnelle dont l'apprentissage est le plus imprécis. Si on regarde par exemple les résultats obtenus pour le modèle hybride  $\mathcal{M}^{(D)}$  dans le Tableau 4.4b, on peut considérer que  $\hat{S}_1 \approx \hat{S}_2$  et en déduire que les composantes  $\hat{f}_{\mathcal{N}}(\cdot | \hat{m}_1, \hat{\sigma}_1)$  et  $\hat{f}_{\text{KDE}}(\cdot | \mathbf{Z}_{\text{obs}}^{[\bullet, 1]}, \hat{h}_1)$  véhiculent la même part d'incertitude. On est incapable de se rendre compte du fait que la première composante induit un biais sur l'estimation de  $P_f$  qui est nettement supérieur à celui induit par la seconde.

#### 4.3.2.5 Bilan

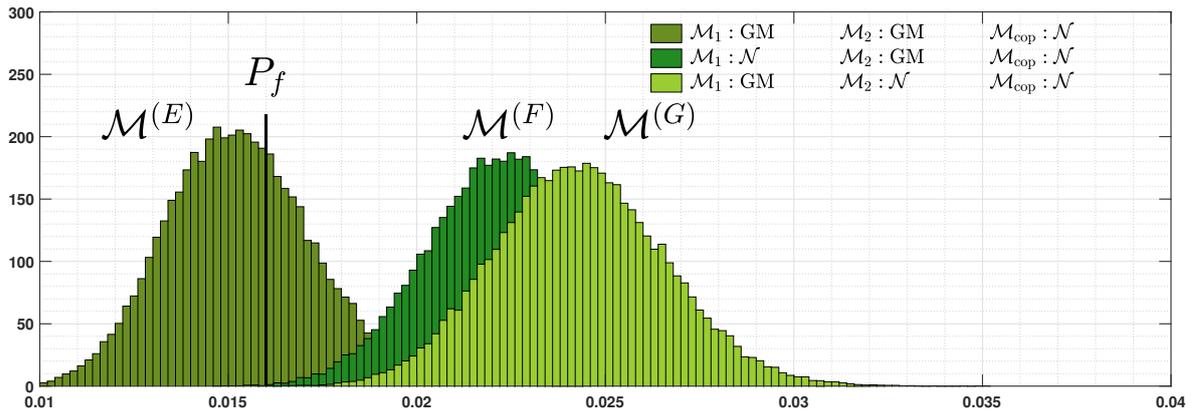
Comme ils découlent de la distribution de  $P_f(\tilde{\mathbf{Z}}, \tilde{\mathbf{V}}) = \mathcal{R}(\hat{f}_{\mathbf{X}}(\cdot | \mathcal{M}, \tilde{\mathbf{Z}}, \tilde{\mathbf{V}}))$ , les indices de sensibilité produits par l'ASF sont caractéristiques du modèle d'apprentissage  $\mathcal{M}$  utilisé. Toutefois, il ne faut pas se méprendre sur la nature de l'information qu'ils apportent. À modèle  $\mathcal{M}$  fixé, ils permettent seulement d'identifier la composante fonctionnelle pour laquelle une acquisition ciblée de données supplémentaires est susceptible d'engendrer la plus forte réduction de la variabilité du risque de défaillance. De plus, si  $\mathcal{M}$  est choisi malhablement, l'estimation du risque est alors biaisée et l'assimilation de nouvelles données ne permet de compenser cette erreur que dans une moindre mesure.



(a)  $\mathcal{M}$  équivaut à l'estimation (désassemblée) d'une densité gaussienne multidimensionnelle.



(b)  $\mathcal{M}$  incorpore un apprentissage non-paramétrique de certaines lois marginales.



(c)  $\mathcal{M}$  s'appuie sur la connaissance du modèle de distribution de certaines lois marginales.

FIGURE 4.15 – Influence du modèle d'apprentissage  $\mathcal{M}$  sur la distribution de l'estimateur  $P_f(\tilde{\mathbf{Z}}, \tilde{\mathbf{V}})$ .

$\mathcal{M}$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_{\text{cop}}$	ASF :	$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_{\text{cop}}$
$\mathcal{M}^{(A)}$	$\mathcal{N}$	$\mathcal{N}$	$\mathcal{N}$	moyenne :	35.77 %	36.50 %	25.88 %
				écart-type :	2.74 %	4.69 %	3.26 %
				rang :	(2)	(1)	(3)
$\mathcal{M}^{(B)}$	KDE	KDE	$\mathcal{N}$	moyenne :	41.25 %	39.53 %	18.14 %
				écart-type :	2.86 %	2.52 %	3.52 %
				rang :	(1)	(2)	(3)
$\mathcal{M}^{(C)}$	$\mathcal{N}$	KDE	$\mathcal{N}$	moyenne :	45.72 %	33.08 %	22.67 %
				écart-type :	2.64 %	4.33 %	1.23 %
				rang :	(1)	(2)	(3)
$\mathcal{M}^{(D)}$	KDE	$\mathcal{N}$	$\mathcal{N}$	moyenne :	32.76 %	44.37 %	22.19 %
				écart-type :	1.70 %	2.08 %	3.11 %
				rang :	(2)	(1)	(3)
$\mathcal{M}^{(E)}$	GM	GM	$\mathcal{N}$	moyenne :	41.03 %	42.86 %	17.55 %
				écart-type :	2.34 %	3.10 %	3.32 %
				rang :	(2)	(1)	(3)
$\mathcal{M}^{(F)}$	$\mathcal{N}$	GM	$\mathcal{N}$	moyenne :	43.78 %	33.63 %	20.62 %
				écart-type :	3.50 %	3.82 %	3.91 %
				rang :	(1)	(2)	(3)
$\mathcal{M}^{(G)}$	GM	$\mathcal{N}$	$\mathcal{N}$	moyenne :	33.31 %	44.43 %	22.31 %
				écart-type :	2.35 %	2.32 %	2.40 %
				rang :	(2)	(1)	(3)

(a) Estimation des indices de Sobol à partir des générateurs théoriques.

$\mathcal{M}$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_{\text{cop}}$	ASF avec $\mathcal{B}$ :	$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_{\text{cop}}$
$\mathcal{M}^{(A)}$	$\mathcal{N}$	$\mathcal{N}$	$\mathcal{N}$	moyenne :	41.02 %	35.36 %	25.96 %
				écart-type :	1.94 %	2.09 %	4.44 %
				rang :	(1)	(2)	(3)
$\mathcal{M}^{(B)}$	KDE	KDE	$\mathcal{N}$	moyenne :	42.70 %	37.35 %	16.45 %
				écart-type :	2.46 %	3.05 %	2.99 %
				rang :	(1)	(2)	(3)
$\mathcal{M}^{(C)}$	$\mathcal{N}$	KDE	$\mathcal{N}$	moyenne :	45.58 %	32.54 %	20.97 %
				écart-type :	4.16 %	2.45 %	3.86 %
				rang :	(1)	(2)	(3)
$\mathcal{M}^{(D)}$	KDE	$\mathcal{N}$	$\mathcal{N}$	moyenne :	37.49 %	40.68 %	23.55 %
				écart-type :	2.19 %	3.25 %	2.37 %
				rang :	(2)	(1)	(3)
$\mathcal{M}^{(E)}$	GM	GM	$\mathcal{N}$	moyenne :	42.53 %	37.38 %	15.63 %
				écart-type :	2.50 %	3.80 %	3.01 %
				rang :	(1)	(2)	(3)
$\mathcal{M}^{(F)}$	$\mathcal{N}$	GM	$\mathcal{N}$	moyenne :	48.79 %	30.88 %	20.55 %
				écart-type :	2.00 %	3.31 %	3.40 %
				rang :	(1)	(2)	(3)
$\mathcal{M}^{(G)}$	GM	$\mathcal{N}$	$\mathcal{N}$	moyenne :	35.65 %	42.40 %	23.27 %
				écart-type :	3.61 %	1.98 %	2.83 %
				rang :	(2)	(1)	(3)

(b) Estimation des indices de Sobol par ré-échantillonnage de  $\mathbf{x}_{\text{obs}}$ .TABLEAU 4.4 – Résultats de l'ASF pour différentes versions du modèle d'apprentissage  $\mathcal{M}$ .

## 4.4 Renforcement de la procédure d'ASF pour le cadre général

Dans la Section 4.2, on a développé un algorithme d'ASF permettant de quantifier l'influence exercée par l'apprentissage de chaque composante fonctionnelle sur la variabilité d'estimation du risque de défaillance. Par ailleurs, les expérimentations numériques conduites dans la Section 4.3 ont démontré que les indices de sensibilité rendus par cet algorithme identifient correctement les composantes véhiculant le plus de variabilité d'apprentissage. La connaissance de ces indices indique alors la nature des données à collecter pour tenter de réduire la variabilité entachant le processus de prédiction. La méthode proposée n'est cependant valable que dans les rares cas où il existe un algorithme de fiabilité  $\mathcal{R}$  déterministe, exact et peu coûteux. Tout dépend de la fonction de performance  $\phi$  du problème physique traité et du modèle d'apprentissage  $\mathcal{M}$  choisi. De par sa simplicité, l'exemple étudié à la Section 4.3.1 donnait accès à une formule analytique de  $P_f(\mathbf{x}_{\text{obs}})$ . De même, à la Section 4.3.2, le fait de rester en dimension  $d = 2$  laissait la possibilité d'un recours à l'intégration numérique. Toutefois, d'une manière générale, l'éventail des situations permettant de calculer l'intégrale sur  $D_f := \phi^{-1}(\{y > T\})$  de la densité  $\hat{f}_{\mathbf{X}}(\cdot | \mathcal{M}, \mathbf{x}_{\text{obs}})$  est très réduit.

Dans cette section, on souhaite étendre l'algorithme d'ASF pour qu'il puisse s'adapter au cadre plus général et plus réaliste où il n'existe pas d'algorithme  $\mathcal{R}$  déterministe. Dès lors que  $\phi$  est un vrai code de simulation, son expression est inconnue et chaque évaluation implique un fort coût de calcul. Dans ce contexte, il faut utiliser l'un des algorithmes de simulation stochastique présentés au Chapitre 2 pour construire un estimateur non-biaisé de la quantité  $P_f(\mathbf{x}_{\text{obs}})$ . Dans la suite, on désignera par  $\mathcal{R}_w$  tout algorithme d'estimation de la probabilité d'un évènement rare qui s'appuie sur une technique d'échantillonnage. La notation  $w$  insiste sur le caractère stochastique d'un tel algorithme. En effet, à échantillon d'apprentissage  $\mathbf{x}_{\text{obs}}$  fixé, toute estimation produite par  $\mathcal{R}_w$  doit être vue comme la réalisation d'une variable aléatoire  $\hat{P}_f(\mathbf{x}_{\text{obs}})$  qui dépend des points de  $\mathbb{R}^d$  choisis au moment de l'échantillonnage. Le choix de la méthode d'échantillonnage, la taille du budget de simulation et le réglage des hyperparamètres sont trois leviers permettant d'agir sur cette variabilité intrinsèque au fonctionnement de  $\mathcal{R}_w$ . Le lien entre  $\mathbf{x}_{\text{obs}}$  et  $P_f(\mathbf{x}_{\text{obs}})$  est alors stochastique et modélisé par  $\mathcal{D}_w$  :

$$\hat{P}_f(\mathbf{x}_{\text{obs}}) = \mathcal{D}_w(\mathbf{x}_{\text{obs}}) = [\mathcal{R}_w \circ \mathcal{L}](\mathbf{x}_{\text{obs}}). \quad (4.98)$$

Le problème d'ASF consiste alors à étudier la sensibilité de l'application  $\Psi_w$  :

$$\Psi_w : \quad \mathbb{L}^2(\mathbb{R}) \quad \times \cdots \times \quad \mathbb{L}^2(\mathbb{R}) \quad \times \quad \mathbb{L}^2(\mathcal{H}_d) \quad \longrightarrow \quad [0, 1] \\ \hat{f}_{X_1}(\cdot | \mathcal{M}_1, \mathbf{X}_{\text{obs}}) \quad \dots \quad \hat{f}_{X_d}(\cdot | \mathcal{M}_d, \mathbf{X}_{\text{obs}}) \quad \hat{\mathcal{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{X}_{\text{obs}}) \quad \longmapsto \quad \hat{P}_f(\mathbf{X}_{\text{obs}}) \quad (4.99)$$

On cherche à nouveau à propager sur le risque de défaillance les incertitudes provenant de la collecte de  $\mathbf{X}_{\text{obs}}$ . Toutefois, à la différence de ce qui a été étudié précédemment, l'enchaînement des briques de calcul inclut désormais  $\mathcal{R}_w$  et n'est donc plus purement déterministe. La variabilité affectant la variable de sortie  $\hat{P}_f(\mathbf{X}_{\text{obs}})$  est alors le mélange de la variabilité engendrée par  $\mathbf{X}_{\text{obs}}$  et de la variabilité propre à  $\mathcal{R}_w$  qui peut être vue comme du bruit.

### 4.4.1 Étude des difficultés introduites par ce nouveau cadre

Effectuer l'ASF décrite par l'équation (4.99) implique de se confronter à deux principaux obstacles. Le premier est d'ordre calculatoire tandis que le second est d'ordre méthodologique.

Dans le cas de l'utilisation du schéma de type *Pick-Freeze* à  $M$  itérations, on a pu voir à la Section 4.2.3.2 que l'ASF implique  $M_t := (d + 2) \times M$  appels à  $\mathcal{R}$  pour calculer toutes les probabilités de défaillances. En effet, chaque itération commence par deux exécutions en parallèle de  $\mathcal{L} \circ \mathcal{B}$  à partir de  $\mathbf{x}_{\text{obs}}$  puis l'application d'un schéma de permutation  $\mathcal{P}$  conduit à la construction d'un tableau  $T_{AB}$  de  $d + 2$  densités conjointes recombinaées. Ce sont autant de probabilités de défaillance qu'il faut estimer à chaque itération. Lorsqu'on a recours à  $\mathcal{R}_w$  pour accomplir la tâche

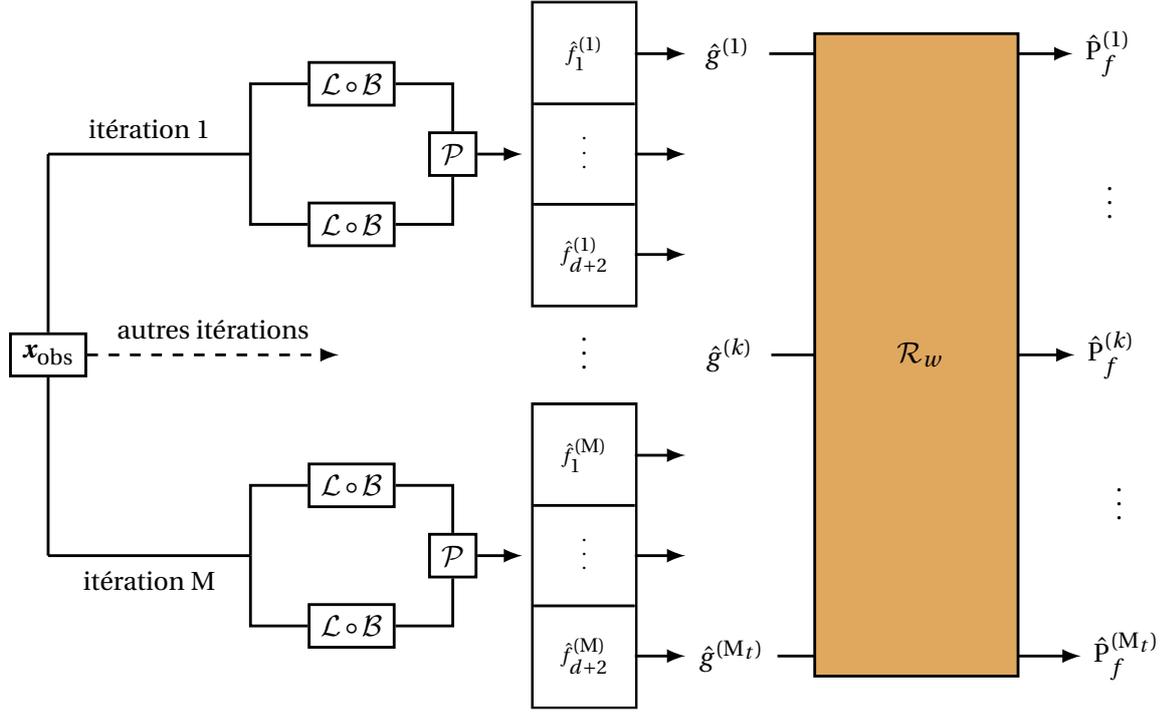


FIGURE 4.16 – Mise en évidence de la charge de calcul induite par  $\mathcal{R}_w$ .

$\hat{f}_{\mathbf{X}}(\cdot | \mathcal{M}, \mathbf{X}_{\text{obs}}) \mapsto \hat{P}_f(\mathbf{X}_{\text{obs}})$ , quel que soit l'algorithme de simulation utilisé et quel que soit le niveau de criticité de l'évènement rare, il faut évaluer  $\phi$  au moins plusieurs milliers de fois. Le cas très simple étudié à la Section 4.3.1 a montré qu'il peut être nécessaire de prendre  $M \geq 10^3$ . Ici, pour une telle valeur de  $M$ , il ne sera pas possible de lancer l'algorithme  $\mathcal{R}_w$  à  $M_t$  reprises. C'est d'autant plus vrai lorsque le nombre  $d$  de variables d'entrée augmente. Dans le cas du schéma d'estimation *Correlation 2*, pour lequel le coefficient multiplicatif  $(d+2)$  est remplacé par  $(2d+3)$ , le problème de gestion de la charge de calcul est encore plus prégnant. Sur la Figure 4.16, on propose une schématisation de la charge de calcul accompagnant l'utilisation du schéma *Pick-Freeze*. On adopte alors la notation  $\hat{g}^{(k)}$  pour désigner une densité d'entrée et cette convention est conservée dans la suite. L'enjeu est alors la construction d'un estimateur  $\hat{P}_f^{(k)}$  de la probabilité  $P_f^{(k)}$  associée à  $\hat{g}^{(k)}$  qui peut être utilisé à  $M_t$  reprises au sein de la procédure d'ASE. Une contribution de ce travail, présentée dans les Sections 4.4.2 et 4.4.3, a consisté à mettre en place une stratégie d'estimation des probabilités  $P_f^{(k)}$  s'appuyant sur certains développements de la théorie de l'échantillonnage préférentiel afin d'amoindrir le nombre d'appels à  $\phi$ .

Le second problème réside dans le fait que l'application  $\Psi_w$  dont on cherche à analyser la sensibilité est stochastique. À supposer qu'une stratégie d'estimation de toutes les probabilités  $P_f^{(k)}$  puisse être implémentée, il faut s'assurer que les indices de sensibilité calculés à partir de la variance de  $\hat{P}_f(\mathbf{X}_{\text{obs}})$  sont proches de ceux qu'on pourrait calculer à partir de la variance de  $P_f(\mathbf{X}_{\text{obs}})$  si un algorithme  $\mathcal{R}$  existait. Trois situations peuvent être distinguées :

- (a) Si  $\mathbb{V}(P_f(\mathbf{X}_{\text{obs}})) \approx \mathbb{V}(\hat{P}_f(\mathbf{X}_{\text{obs}}))$  :
  - $\Rightarrow$  La variabilité intrinsèque à  $\mathcal{R}_w$  est négligeable.
  - $\Rightarrow$  Les distributions de  $P_f(\mathbf{X}_{\text{obs}})$  et  $\hat{P}_f(\mathbf{X}_{\text{obs}})$  sont quasiment identiques.
  - $\Rightarrow$  Substituer  $\mathcal{R}_w$  à  $\mathcal{R}$  ne pose aucun problème.
- (b) Si  $\mathbb{V}(P_f(\mathbf{X}_{\text{obs}})) < \mathbb{V}(\hat{P}_f(\mathbf{X}_{\text{obs}})) < 2 \times \mathbb{V}(P_f(\mathbf{X}_{\text{obs}}))$  :
  - $\Rightarrow$  La variabilité intrinsèque à  $\mathcal{R}_w$  n'est plus négligeable.
  - $\Rightarrow$  Les distributions de  $P_f(\mathbf{X}_{\text{obs}})$  et  $\hat{P}_f(\mathbf{X}_{\text{obs}})$  diffèrent.

- ⇒ L'algorithme d'ASF est inopérant et donne des résultats erronés.
- (c) Si  $\mathbb{V}(\mathbf{P}_f(\mathbf{X}_{\text{obs}})) \ll \mathbb{V}(\hat{\mathbf{P}}_f(\mathbf{X}_{\text{obs}}))$  :
  - ⇒ La variabilité intrinsèque à  $\mathcal{R}_w$  écrase celle induite par l'apprentissage.
  - ⇒ Appliquer l'algorithme d'ASF ne sert à rien.
  - ⇒ Pour réduire  $\mathbb{V}(\hat{\mathbf{P}}_f(\mathbf{X}_{\text{obs}}))$ , il faut intervenir sur  $\mathcal{R}_w$  et non sur  $\mathbf{X}_{\text{obs}}$ .

L'algorithme d'ASF ne pourra être exécuté sereinement qu'à condition d'être dans la situation (a). Si la configuration (c) est rencontrée, les motivations initiales de ce travail n'ont plus lieu d'être. Par contre, si on est confronté à la situation (b), il faut envisager d'affiner la construction des estimateurs  $\hat{\mathbf{P}}_f^{(k)}$ . Cela passe souvent par l'utilisation de techniques d'échantillonnage plus coûteuses, et donc par un accroissement du nombre d'appels à  $\phi$ .

Adapter l'algorithme initial d'ASF pour y implanter  $\mathcal{R}_w$  relève d'un compromis entre deux tendances antagonistes. D'un côté, le besoin d'estimer  $M_t$  probabilités de défaillance encourage à favoriser des techniques d'échantillonnage peu coûteuses en termes de nombre d'appels à  $\phi$ . De l'autre, pour limiter l'influence indésirable de la stochasticité de  $\mathcal{R}_w$  sur les résultats de l'ASF, on est tenté d'utiliser les méthodes de simulation d'évènements rares les plus robustes, qui sont également les plus onéreuses d'un point de vue numérique. Les expérimentations conduites à la Section 4.5 viendront montrer que la stratégie de calcul développée dans ce travail permet d'atteindre un juste équilibre.

#### 4.4.2 Utilisation de l'échantillonnage préférentiel dans un contexte d'ASF

Comme on l'a pressenti, l'estimation de toutes les probabilités  $\mathbf{P}_f^{(k)}$  ne peut évidemment pas être envisagée avec une séquence d'appels à l'algorithme *Crude Monte Carlo* (CMC) :

$$\hat{\mathbf{P}}_f^{(k)} := \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbb{1}_{\{Y^{(i)} > \Gamma\}} \quad \text{avec} \quad \begin{cases} \mathbf{X}^{(i)} \text{ i.i.d. } \hat{\mathbf{g}}^{(k)} \\ Y^{(i)} = \phi(\mathbf{X}^{(i)}) \end{cases} . \quad (4.100)$$

La Figure 4.17a propose une schématisation de cette stratégie naïve dont la charge de calcul est exorbitante. Imaginons qu'un problème  $\mathcal{P}_0$  implique  $d = 8$  entrées aléatoires et que le risque de défaillance théorique soit  $\mathbf{P}_f = 10^{-4}$ . Pour estimer une telle probabilité avec un écart relatif de l'ordre de 10%, l'estimateur CMC de  $\mathbf{P}_f$  réclame  $N_s = 10^6$  appels à  $\phi$ . Il faudrait donc budgétiser  $N_{\text{tot}} := N_s \times (d + 2) \times M = 10^{10}$  exécutions de  $\phi$ . Réciproquement, pour un budget de simulation fixé par exemple à  $N_{\text{tot}} = 10^5$ , il ne reste que  $N_s = 10$  tirages aléatoires pour construire l'estimateur  $\hat{\mathbf{P}}_f^{(k)}$  et la variance de l'estimateur associé devient alors trop grande. Une technique très classique de réduction de la variance consiste alors à utiliser l'échantillonnage préférentiel ou *Importance Sampling* (IS). L'apparition de cette approche dans un contexte fiabiliste remonte aux travaux de SHINOZUKA [1983] puis de HARBITZ [1986]. L'idée est de remarquer que :

$$\mathbf{P}_f^{(k)} = \mathbb{E}_{\hat{\mathbf{g}}^{(k)}} [\mathbb{1}_{D_f}(\mathbf{X})] = \mathbb{E}_{h^{(k)}} \left[ \mathbb{1}_{D_f}(\mathbf{X}) \frac{\hat{\mathbf{g}}^{(k)}(\mathbf{X})}{h^{(k)}(\mathbf{X})} \right], \quad (4.101)$$

et d'utiliser la densité auxiliaire  $h^{(k)}$  pour construire l'estimateur :

$$\hat{\mathbf{P}}_f^{(k)} = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbb{1}_{\{Y^{(i)} > \Gamma\}} \frac{\hat{\mathbf{g}}^{(k)}(\mathbf{X}^{(i)})}{h^{(k)}(\mathbf{X}^{(i)})} \quad \text{avec} \quad \begin{cases} \mathbf{X}^{(i)} \text{ i.i.d. } h^{(k)} \\ Y^{(i)} = \phi(\mathbf{X}^{(i)}) \end{cases}, \quad (4.102)$$

dont la variance est nulle dans le cas où  $h^{(k)}$  est la densité optimale  $h_{\text{opt}}^{(k)}$  donnée par :

$$h_{\text{opt}}^{(k)}(\mathbf{x}) = \frac{\mathbb{1}_{D_f}(\mathbf{x}) \times \hat{\mathbf{g}}^{(k)}(\mathbf{x})}{\mathbf{P}_f^{(k)}} . \quad (4.103)$$

Le dénominateur  $\hat{P}_f^{(k)}$  étant inconnu, on a vu au Chapitre 2 qu'il fallait construire une version sous-optimale  $\hat{h}_{\text{opt}}^{(k)}$  de  $h_{\text{opt}}^{(k)}$ . Diverses techniques ont été mentionnées, les plus courantes étant celle de RUBINSTEIN et KROESE [2013] basée sur l'optimisation par entropie croisée des paramètres au sein d'une famille  $\mathcal{F}_\lambda$  de distributions ou alors celle de GIVENS et RAFTERY [1996] suggérant un apprentissage non-paramétrique de  $h^{(k)}$  à l'aide de noyaux gaussiens. Quelle que soit la méthode retenue, la construction de chaque estimateur  $\hat{P}_f^{(k)}$  passe par deux phases :

- (a) Une phase d'exploration qui aboutit à l'obtention de  $\hat{h}_{\text{opt}}^{(k)}$  et à la simulation d'échantillons  $\mathbf{X}^{(i)}$  indépendants et identiquement distribués suivant cette densité sous-optimale.
- (b) Une phase de pondération qui sert à calculer les ratios intervenant dans la formule (4.102) :

$$w^{(k)}(\mathbf{X}^{(i)}) := \frac{\hat{g}^{(k)}(\mathbf{X}^{(i)})}{\hat{h}_{\text{opt}}^{(k)}(\mathbf{X}^{(i)})}. \quad (4.104)$$

La stratégie d'ASF reposant sur une application séquentielle de l'IS à toutes les densités  $\hat{g}^{(k)}$  est décrite à la Figure 4.17b avec une disjonction des deux phases mentionnées ci-dessus. Si on reprend le problème  $\mathcal{P}_0$  évoqué plus haut ( $d = 8$  et  $P_f = 10^{-4}$ ), toute technique d'échantillonnage préférentiel réclame au moins  $N_s = 10^3$  échantillons, ce qui porte finalement la charge de calcul à  $N_{\text{tot}} = 10^7$  appels à  $\phi$ . Cela reste là encore inenvisageable en pratique. De même, pour un budget de calcul préalablement limité à  $N_{\text{tot}} = 10^5$  appels à  $\phi$ , les  $N_s = 10$  tirages alloués à  $\hat{P}_f^{(k)}$  ne seront pas suffisants pour que l'échantillonnage préférentiel joue pleinement son rôle dans la réduction de la variance. Dans le prochain paragraphe, on présente une stratégie de calcul permettant de ne pas se lancer dans  $M_f$  explorations de la zone de défaillance.

### 4.4.3 Mise en place d'une stratégie de diminution de la charge de calcul

#### 4.4.3.1 Réduction du nombre d'explorations du domaine de défaillance

Le paragraphe précédent a mis en avant les particularités d'une procédure d'ASF utilisant l'échantillonnage préférentiel pour estimer toutes les probabilités de défaillance. Tout d'abord, comme le montre l'expression (4.103), toutes les densités auxiliaires optimales partagent un support commun car le domaine de défaillance  $D_f$  est une donnée invariante du problème. Il existe donc de fait une proximité analytique entre toutes les densités auxiliaires optimales. On a aussi pu constater que le budget de simulation est exclusivement consommé par les phases d'exploration destinées à la construction de densités auxiliaires sous-optimales supposées être des approximations satisfaisantes des versions optimales. On peut alors penser qu'il serait très avantageux dans un objectif de réduction du coût de calcul d'estimer toutes les probabilités à partir d'une seule densité auxiliaire sous-optimale. Dans l'esprit des équations (4.101) et (4.102), on peut écrire que :

$$P_f^{(k)} = \mathbb{E}_{\hat{g}^{(k)}} [\mathbb{1}_{D_f}(\mathbf{X})] = \mathbb{E}_{\hat{h}_{\text{opt}}^{(1)}} \left[ \mathbb{1}_{D_f}(\mathbf{X}) \frac{\hat{g}^{(k)}(\mathbf{X})}{\hat{h}_{\text{opt}}^{(1)}(\mathbf{X})} \right], \quad (4.105)$$

où  $\hat{h}_{\text{opt}}^{(1)}$  est la densité auxiliaire associée à  $\hat{g}^{(1)}$ . Pour toute autre densité  $\hat{g}^{(k)}$ , l'estimateur du risque de défaillance est alors donné par :

$$\hat{P}_f^{(k)} = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbb{1}_{\{Y^{(i)} > T\}} \frac{\hat{g}^{(k)}(\mathbf{X}^{(i)})}{\hat{h}_{\text{opt}}^{(1)}(\mathbf{X}^{(i)})} \quad \text{avec} \quad \begin{cases} \mathbf{X}^{(i)} \text{ i.i.d. } \hat{h}_{\text{opt}}^{(1)} \\ Y^{(i)} = \phi(\mathbf{X}^{(i)}) \end{cases}. \quad (4.106)$$

Cette stratégie d'unique exploration de la zone de défaillance est représentée à la Figure 4.17c et elle est connue sous le nom d'astuce "*inverse-reverse*". Les échantillons d'entrée  $\mathbf{X}^{(i)}$  sont simulés lorsque la phase d'entraînement de  $\hat{h}_{\text{opt}}^{(1)}$  s'achève et ils sont ensuite réutilisés par tous les autres estimateurs. Il en va évidemment de même pour les sorties  $Y^{(i)}$  qui ne sont calculées qu'une seule fois à partir de  $\phi$ . D'un estimateur  $\hat{P}_f^{(k)}$  à l'autre, la seule tâche à accomplir est la mise à jour des

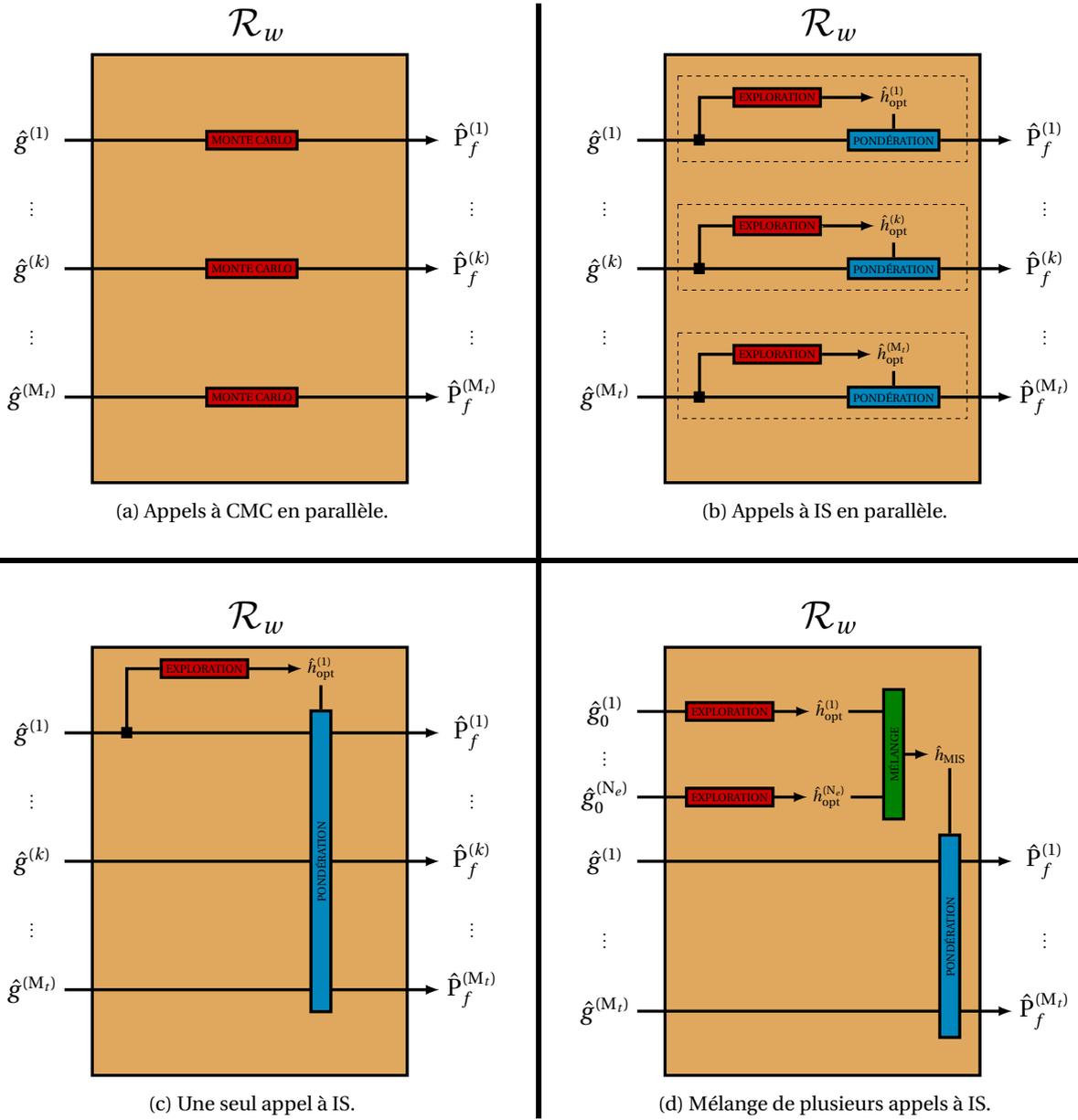


FIGURE 4.17 – Différentes stratégies pour conduire l'ASF avec  $\mathcal{R}_w$ .

ratios, notamment des numérateurs  $\hat{g}^{(k)}(\mathbf{X}^{(i)})$ . Avec cette stratégie, le budget se restreint alors à  $N_{\text{tot}} = N_s$ . Pour en revenir au problème  $\mathcal{P}_0$ , cela signifie que l'on peut dépenser l'intégralité des  $N_s = 10^5$  appels à  $\phi$  dans la phase d'exploration initiale et l'estimateur  $\hat{P}_f^{(1)}$  acquiert une très grande robustesse comparativement aux deux stratégies séquentielles présentées plus haut.

S'il est indéniable que l'astuce "inverse-reverse" permet de construire un estimateur  $\hat{P}_f^{(1)}$  de faible variance, on peut s'interroger sur la robustesse des estimateurs qui se contentent de recycler le matériel fourni par la première exploration. Pour une densité  $\hat{g}^{(k)}$  fixée, lorsqu'on remplace la densité auxiliaire sous-optimale  $\hat{h}_{\text{opt}}^{(k)}$  par  $\hat{h}_{\text{opt}}^{(1)}$ , l'estimateur reste non-biaisé mais sa variance n'est plus réduite à un niveau minimal. Si ces deux densités sont très différentes, la variance de  $\hat{P}_f^{(k)}$  peut même excéder celle d'un estimateur de type CMC. Toutefois, plusieurs éléments peuvent nous laisser penser que, dans le contexte d'ASF qui est le nôtre, il n'est pas aberrant d'utiliser  $\hat{h}_{\text{opt}}^{(1)}$  pour toute la série d'estimations. Tout d'abord, les densités initiales  $\hat{g}^{(k)}$  résultent de l'exécution d'un même code de simulation  $[\mathcal{L} \circ \mathcal{B}](\mathbf{x}_{\text{obs}})$ . Bien entendu, le mécanisme de ré-échantillonnage génère des perturbations mais on ne s'attend pas à ce que les densités apprises soient radicalement différentes. C'est dans cet esprit que MORIO [2011] a utilisé cette astuce pour implémenter un algorithme d'ASF visant à identifier les paramètres de la distribution d'entrée les plus influents sur la prédiction du risque de défaillance. De même, LEMAÎTRE et collab. [2015] ont eu recours à cette astuce pour estimer leurs indices PLI (*Perturbed-Law Indices*).

Lorsqu'on décide de mettre en œuvre cette stratégie, il faut être conscient qu'un scénario critique existe. Il survient si la densité  $\hat{g}^{(1)}$  est assez peu représentative du comportement de la fonction aléatoire  $\hat{f}_{\mathbf{X}}(\cdot \mid \mathcal{M}, \tilde{\mathbf{Z}}, \tilde{\mathbf{V}})$ . Par effet cascade, la densité auxiliaire  $\hat{h}_{\text{opt}}^{(1)}$  ne permet plus à tous les estimateurs  $\hat{P}_f^{(k)}$  qu'elle sert à construire d'afficher une faible variance. Pour une exécution de l'algorithme d'ASF, la variabilité observée sur la série des probabilités de défaillance n'est plus seulement le résultat de la propagation des incertitudes d'échantillonnage. Elle intègre aussi une part de variabilité causée par le manque de robustesse de certains estimateurs  $\hat{P}_f^{(k)}$ . Cette variabilité est indésirable étant donné qu'elle est susceptible de biaiser les estimations des indices de Sobol. Pour éviter un tel scénario, une alternative consiste à réaliser un petit nombre d'explorations et à les unifier.

#### 4.4.3.2 Renforcement de la robustesse par échantillonnage préférentiel multiple

Selon ce qui a été dit précédemment, imaginons que l'on réalise désormais une exploration pour chacune des  $N_e$  premières densités impliquées dans l'ASF. En se basant sur l'expression de l'estimateur (4.102), on définit  $N_e$  estimateurs utilisant leur propre densité auxiliaire :

$$\forall 1 \leq l \leq N_e, \quad \hat{P}_f^{(l)} = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbb{1}_{\{Y_l^{(i)} > T\}} \frac{\hat{g}^{(l)}(\mathbf{X}_l^{(i)})}{\hat{h}_{\text{opt}}^{(l)}(\mathbf{X}_l^{(i)})} \quad \text{avec} \quad \begin{cases} \mathbf{X}_l^{(i)} \text{ i.i.d. } \hat{h}_{\text{opt}}^{(l)} \\ Y_l^{(i)} = \phi(\mathbf{X}_l^{(i)}) \end{cases}, \quad (4.107)$$

et l'indexation par  $l$  de  $\mathbf{X}_l^{(i)}$  permet de mentionner explicitement l'exploration ayant fourni l'échantillon considéré. La même remarque est valable pour les sorties  $Y_l^{(i)}$ . La question est maintenant de savoir comment combiner au sein d'un même estimateur les  $N_e$  densités auxiliaires  $\hat{h}_{\text{opt}}^{(l)}$  ainsi fabriquées. Pour une densité d'entrée  $\hat{g}^{(k)}$  donnée, on pourrait estimer  $P_f^{(k)}$  en faisant la moyenne des estimations obtenues par échantillonnage préférentiel avec chacune des densités auxiliaires  $\hat{h}_{\text{opt}}^{(l)}$  mais on a vu au Chapitre 2 que cette solution n'est pas la plus attrayante pour recouper les résultats des différentes explorations. La technique dite de *Multiple Importance Sampling* (MIS) qui est présentée dans VEACH et GUIBAS [1995a,b] propose une alternative plus efficace en rattachant à chaque lot d'échantillons  $\mathbf{X}_l^{(i)}$  une fonction  $p_l$  permettant d'affiner le mécanisme de pondération. L'idée est ensuite de construire un estimateur  $\hat{P}_f^{(k)}$  qui combine les densités auxiliaires et leurs

fonctions de pondération respectives :

$$\hat{P}_f^{(k)} = \frac{1}{N_s} \sum_{l=1}^{N_e} \left( \sum_{i=1}^{N_s} p_l(\mathbf{X}_l^{(i)}) \mathbb{1}_{\{Y_l^{(i)} > T\}} \frac{\hat{g}^{(k)}(\mathbf{X}_l^{(i)})}{\hat{h}_{\text{opt}}^{(l)}(\mathbf{X}_l^{(i)})} \right) \text{ avec } \begin{cases} \mathbf{X}_l^{(i)} \text{ i.i.d. } \hat{h}_{\text{opt}}^{(l)} \\ Y_l^{(i)} = \phi(\mathbf{X}_l^{(i)}) \end{cases} . \quad (4.108)$$

La fonction  $p_l$  ne s'applique qu'aux échantillons  $\mathbf{X}_l^{(i)}$  produits par la  $l$ -ème phase d'exploration mais son système de pondération peut mettre en jeu des informations collectées au cours de chacune des  $N_e - 1$  autres phases. Parmi les options proposées dans [VEACH \[1997\]](#), la règle connue sous le nom de *Balance Heuristic* (BH) suggère que chaque fonction  $p_l$  soit proportionnelle à la densité auxiliaire  $\hat{h}_{\text{opt}}^{(l)}$ . En regroupant les différents lots d'échantillons  $\{\mathbf{X}_l^{(i)}\}_{l=1}^{N_e}$  fabriqués au cours des explorations préliminaires, l'estimateur (4.108) peut être réécrit comme suit :

$$\hat{P}_f^{(k)} = \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbb{1}_{\{Y^{(j)} > T\}} \frac{\hat{g}^{(k)}(\mathbf{X}^{(j)})}{\hat{h}_{\text{MIS}}(\mathbf{X}^{(j)})} \text{ avec } \begin{cases} N_t = N_e \times N_s \\ \mathbf{X}^{(j)} \text{ i.i.d. } \hat{h}_{\text{MIS}} = \frac{1}{N_e} \sum_{l=1}^{N_e} \hat{h}_{\text{opt}}^{(l)} \\ Y^{(j)} = \phi(\mathbf{X}^{(j)}) \end{cases} . \quad (4.109)$$

L'estimateur ainsi obtenu correspond simplement à un échantillonnage préférentiel avec une densité auxiliaire  $\hat{h}_{\text{MIS}}$  qui est la moyenne de toutes les densités  $\hat{h}_{\text{opt}}^{(l)}$ . Faire le choix de BH signifie donc opter pour un mélange uniforme des  $N_e$  densités candidates. Si on souhaite utiliser une telle approche pour conduire l'ASF, il faut suivre les instructions suivantes :

1. Exécuter  $N_e$  fois le code de simulation  $[\mathcal{L} \circ \mathcal{B}](\mathbf{x}_{\text{obs}})$  :
  - ⇒ Fabrication d'une série de  $N_e$  densités d'entrée  $\hat{g}_0^{(1)}, \dots, \hat{g}_0^{(N_e)}$  .
2. Lancer une phase d'exploration pour chaque densité  $\hat{g}_0^{(l)}$  :
  - ⇒ Construction d'une densité auxiliaire sous-optimale  $\hat{h}_{\text{opt}}^{(l)}$  .
  - ⇒ Simulation de  $N_s$  échantillons  $\mathbf{X}_l^{(i)} \text{ i.i.d. } \hat{h}_{\text{opt}}^{(l)}$  et calcul de leurs sorties  $Y_l^{(i)} = \phi(\mathbf{X}_l^{(i)})$  .
3. Calculer les valeurs prises par la densité  $\hat{h}_{\text{MIS}}$  :

$$H_l^{(i)} := \hat{h}_{\text{MIS}}(\mathbf{X}_l^{(i)}) = \frac{1}{N_e} \sum_{s=1}^{N_e} \hat{h}_{\text{opt}}^{(s)}(\mathbf{X}_l^{(i)}) . \quad (4.110)$$

4. Fusionner les différents lots d'échantillons :

$$\begin{aligned} [\mathbf{X}^{(j)}]_{1 \leq j \leq N_t} &\leftarrow [\mathbf{X}_l^{(i)}]_{\substack{1 \leq i \leq N \\ 1 \leq l \leq N_e}} ; [Y^{(j)}]_{1 \leq j \leq N_t} \leftarrow [Y_l^{(i)}]_{\substack{1 \leq i \leq N \\ 1 \leq l \leq N_e}} ; \\ [H^{(j)}]_{1 \leq j \leq N_t} &\leftarrow [H_l^{(i)}]_{\substack{1 \leq i \leq N \\ 1 \leq l \leq N_e}} . \end{aligned} \quad (4.111)$$

5. Estimer toutes les probabilités  $P_f^{(k)}$  par mise à jour des numérateurs  $\hat{g}^{(k)}(\mathbf{X}^{(j)})$  :

$$\hat{P}_f^{(k)} = \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbb{1}_{\{Y^{(j)} > T\}} \frac{\hat{g}^{(k)}(\mathbf{X}^{(j)})}{H^{(j)}} . \quad (4.112)$$

La Figure 4.17d propose une représentation simplifiée de la stratégie d'ASF décrite ci-dessus. Seules les  $N_e$  explorations préalables de la zone de défaillance représentent une charge de calcul importante. En effet, dès lors que la densité  $\hat{h}_{\text{MIS}}$  est construite, plus aucun appel à  $\phi$  n'est nécessaire. Le budget de simulation vaut donc  $N_{\text{tot}} = N_t = N_e \times N_s$ . Dans le cas du problème  $\mathcal{P}_0$  où le nombre maximal d'appels à  $\phi$  est fixé à  $N_{\text{tot}} = 10^5$ , on peut par exemple choisir de conduire  $N_e$  explorations tout en se limitant à  $N_s = 2 \times 10^4$  appels à  $\phi$  par exploration. Si on compare l'estimateur (4.106) issu de l'astuce "inverse-reverse" et l'estimateur (4.112) issu du mélange de  $N_e$  explorations préalables,

le nombre total  $N_{\text{tot}}$  d'échantillons impliqués est le même. Toutefois, dans le premier cas, il y a un fort risque de biaiser les estimations des indices de Sobol si  $\hat{h}_{\text{opt}}^{(1)}$  est une densité auxiliaire qui ne convient pas à la plupart des densités  $\hat{g}^{(k)}$  produites par l'ASF. Au contraire, dans le second cas, un tel incident a moins d'impact sur le résultat final car seule une petite part des échantillons provient de  $\hat{h}_{\text{opt}}^{(1)}$  et on s'attend à ce que les  $N_e - 1$  autres densités auxiliaires possèdent une meilleure capacité à estimer les probabilités de défaillance  $P_f^{(k)}$ . La Figure 4.18 récapitule toutes les briques de calcul intervenant alors au cours de la procédure d'ASF globale qui prend en entrée l'échantillon initial  $\mathbf{x}_{\text{obs}}$  fourni par le retour d'expérience et qui rend en sortie la collection des  $d + 1$  indices de Sobol du premier ordre.

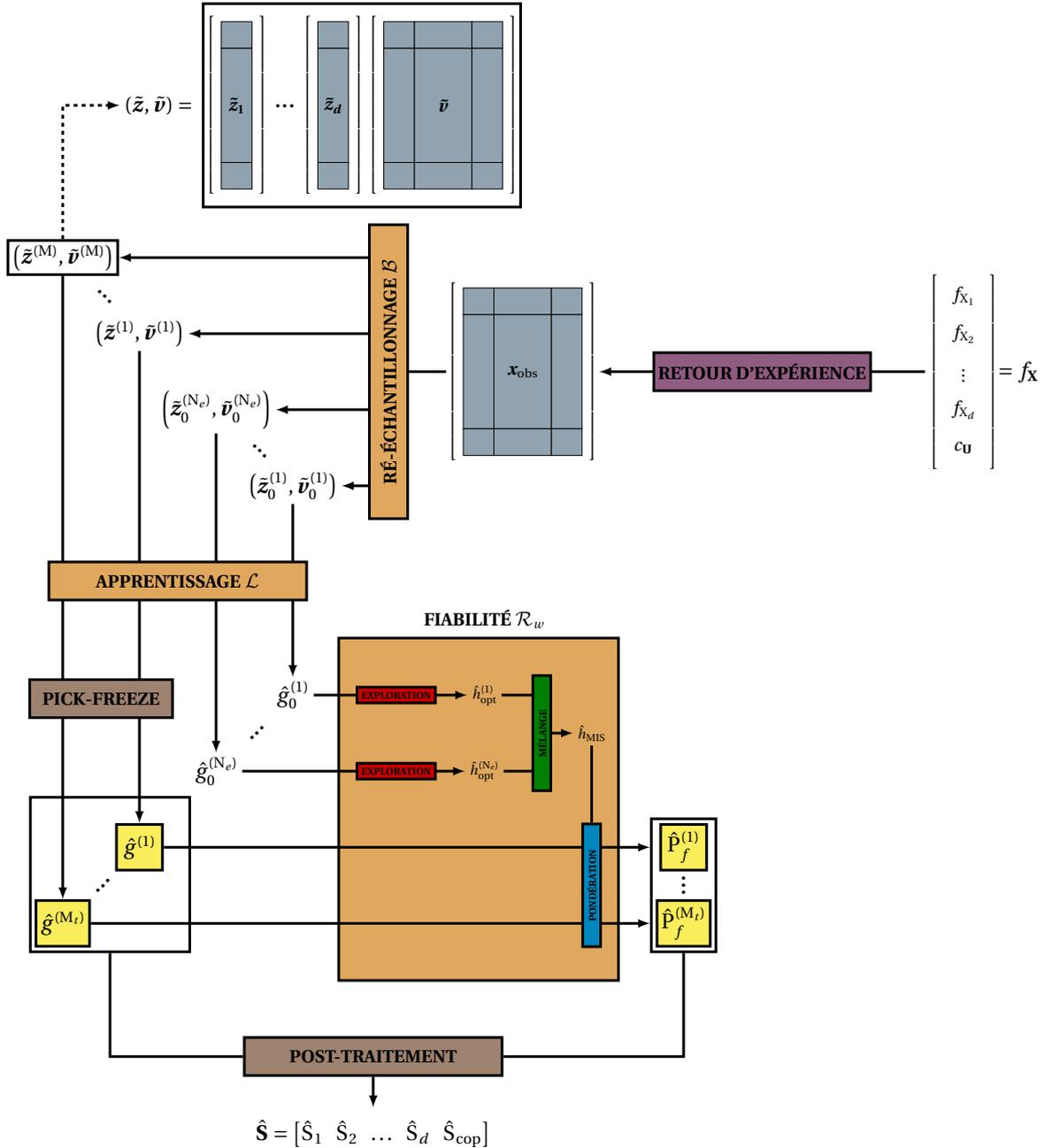


FIGURE 4.18 – Illustration des différentes briques de calcul lorsque l'ASF est exécutée avec  $\mathcal{R}_w$ .

Dans cette section, on a pu voir qu'effectuer une ASF dans un cadre général impose l'utilisation d'un algorithme de simulation  $\mathcal{R}_w$  pour accomplir la tâche  $\hat{g}^{(k)} \rightarrow \hat{P}_f^{(k)}$ . Idéalement, il faudrait être en mesure de répéter  $M_t = k \times M$  fois cette opération avec  $k \in \{d + 2, 2d + 3\}$  et  $M \geq 10^3$ ,

ce qui n'est évidemment pas envisageable en pratique. À budget de simulation contraint, on a pu voir que les approches séquentielles naïves conduisent à la construction d'une série d'estimateurs dont la variance est excessive. L'astuce "inverse-reverse", déjà appliquée par MORIO [2011] ou par LEMAITRE et collab. [2015] dans des contextes d'ASF, est une première solution pour réduire la variance des estimateurs construits. Une idée nouvelle présentée dans ce travail consiste à utiliser l'échantillonnage préférentiel multiple pour renforcer la robustesse de la précédente astuce. En construisant plus finement la densité auxiliaire servant à calculer toutes les probabilités de défaillance, on espère réduire à son plus bas niveau le biais introduit par l'utilisation de méthodes stochastiques.

## 4.5 Validation de la procédure d'ASF renforcée

Dans cette section, on cherche à s'assurer que l'algorithme d'ASF développé pour s'adapter aux contraintes liées à la charge de calcul permet d'estimer avec précision les indices de sensibilité. La démarche de validation se divise en deux parties. Dans un premier temps, on reprend le cas test linéaire de la Section 4.3.1 pour lequel l'ASF peut être conduite avec un algorithme de fiabilité  $\mathcal{R}$  déterministe et exact. L'objectif est de s'assurer que les différentes stratégies de calcul proposées pour conduire l'ASF avec  $\mathcal{R}_w$  rendent des résultats proches de ceux obtenus en conduisant l'ASF avec  $\mathcal{R}$ . Dans un second temps, on s'intéresse à un cas test non-linéaire pour lequel il n'existe pas d'algorithme  $\mathcal{R}$  offrant l'accès aux valeurs exactes des probabilités de défaillance. Les résultats obtenus pour plusieurs algorithmes  $\mathcal{R}_w$  sont alors comparés pour s'assurer que :

- (a) La méthodologie proposée est indépendante de l'algorithme  $\mathcal{R}_w$  choisi.
- (b) La variabilité intrinsèque d'un algorithme  $\mathcal{R}_w$  ne perturbe pas les résultats de l'ASF.

### 4.5.1 Retour sur le cas test linéaire

Reprenons donc le cas test étudié en détail à la Section 4.3.1 pour lequel on a :

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad ; \quad Y = \phi(\mathbf{X}) = \mathbf{a}^\top \mathbf{X} \quad \text{et} \quad \hat{f}_{\mathbf{X}}(\cdot | \mathcal{M}, \mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}}) = \hat{f}_{\mathcal{N}}(\cdot | \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}). \quad (4.113)$$

On rappelle que l'expression analytique (4.68) permet un calcul exact de la probabilité  $P_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$  associée à la densité d'entrée  $\hat{f}_{\mathcal{N}}(\cdot | \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ . On a vu qu'un algorithme  $\mathcal{R}$  existe et qu'il s'agit de FORM. Les résultats d'ASF disponibles dans le Tableau 4.1 servent désormais de référence. On s'intéresse en particulier à ceux concernant l'échantillon  $\mathbf{x}_{\text{obs}}^{(1)}$ .

Dans cette configuration très simple, de nombreux algorithmes de simulation peuvent permettre de conduire l'analyse de fiabilité. Par souci de simplicité,  $\mathcal{R}_w$  prend ici la forme d'une technique d'échantillonnage préférentiel assez rudimentaire qui consiste à utiliser une densité auxiliaire gaussienne. En effet, un estimateur assez naturel de  $P_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$  peut être construit en tirant des échantillons  $\mathbf{X}^{(i)}$  avec une loi gaussienne centrée sur le point de conception  $\mathbf{x}^*$  de l'espace physique :

$$\hat{P}_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbb{1}_{D_f}(\mathbf{X}^{(i)}) \frac{\hat{f}_{\mathcal{N}}(\mathbf{X}^{(i)} | \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})}{\varphi_3(\mathbf{X}^{(i)} - \hat{\mathbf{x}}^*)} \quad \text{avec} \quad \begin{cases} \hat{\mathbf{x}}^* := \hat{\mathcal{T}}^{-1}(\hat{\mathbf{w}}^*) \\ \mathbf{X}^{(i)} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\hat{\mathbf{x}}^*, \mathbf{I}_3) \end{cases}, \quad (4.114)$$

avec  $\varphi_3$  la densité conjointe d'un vecteur gaussien centré réduit de dimension  $d = 3$ . Pour ce qui est du choix de la matrice de covariance de la densité auxiliaire, on admet que  $\mathbf{I}_3$  est un bon réglage dans ce cas particulier. Pour une illustration de cette technique d'échantillonnage, on peut se reporter à la Figure 4.19 où elle est comparée à CMC. Les particules  $\mathbf{X}^{(i)}$  sont transférées dans l'espace normal standard où elles sont notées  $\mathbf{W}^{(i)}$  après application de la transformation de Nataf  $\hat{\mathcal{T}}$  définie à l'équation (4.69). Dans la suite, on parle de *Gaussian Importance Sampling* (GIS) pour désigner la méthode  $\mathcal{R}_w$  qui se résume à :

1. Calculer le point de conception  $\hat{\mathbf{x}}^*$  en utilisant FORM.

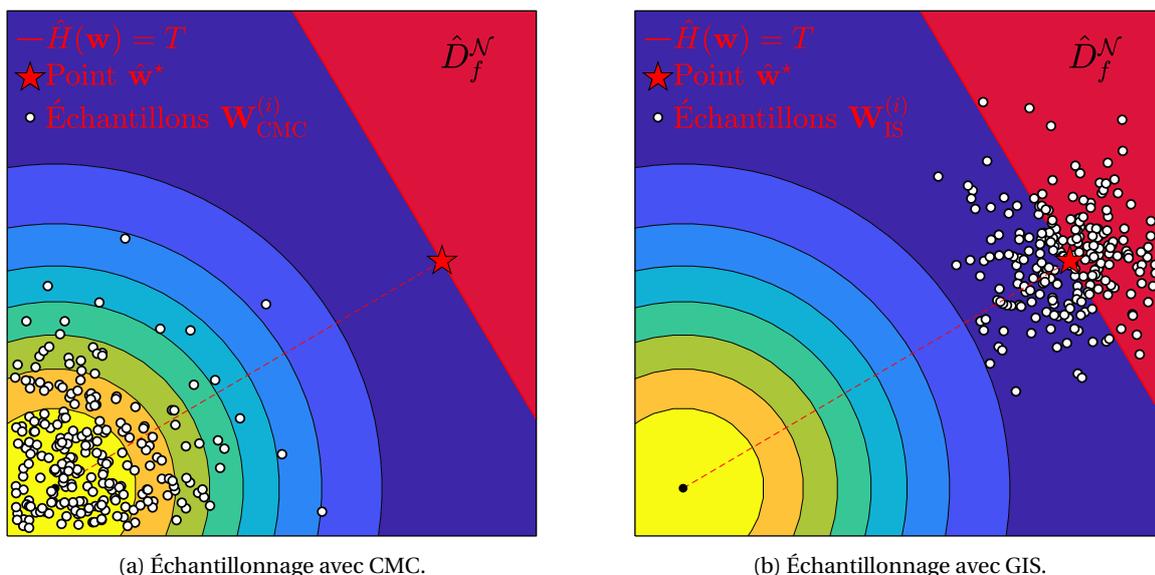


FIGURE 4.19 – Visualisation des techniques d'échantillonnage dans l'espace normal standard.

## 2. Utiliser une densité auxiliaire gaussienne centrée sur $\hat{\mathbf{x}}^*$ .

Pour ce cas test linéaire,  $\hat{\mathbf{w}}^*$  peut être calculé analytiquement avec la formule (4.70). Dans une situation où  $\hat{\mathbf{w}}^*$  serait le résultat de la résolution numérique du problème d'optimisation (4.56), il faudrait prendre en compte dans le budget de simulation les  $N_{\text{opt}}$  évaluations de  $\phi$  qu'impose chaque nouvel appel FORM. Ainsi, pour la technique GIS, la phase d'exploration désigne l'ensemble des tâches consistant à déterminer  $\hat{\mathbf{x}}^*$ , simuler les échantillons  $\mathbf{X}^{(i)}$  et calculer les sorties  $Y^{(i)}$  associées. Quelle que soit la valeur de la probabilité de défaillance, on se retrouve souvent dans une situation où  $N_{\text{opt}} \ll N_s$  quitte à éventuellement relâcher la sévérité du critère d'arrêt de l'algorithme d'optimisation couplé avec FORM. Pour cette raison, on considèrera par la suite que le budget dépensé par GIS se limite aux calculs des  $N_s$  sorties.

On lance l'algorithme d'ASF en implémentant trois stratégies différentes pour l'estimation de la série des probabilités de défaillance  $P_f^{(k)}$ , pouvant d'ailleurs s'écrire  $P_f(\hat{\boldsymbol{\mu}}^{(k)}, \hat{\boldsymbol{\Sigma}}^{(k)})$ :

- |   |               |  |   |
|---|---------------|--|---|
| (St1) Procédure séquentielle avec $M_t$ appels à GIS    | $\Rightarrow$ | $N_{\text{tot}} = M \times (d + 2) \times N_s$ | ; |
| (St2) Astuce "inverse-reverse" avec un seul appel à GIS | $\Rightarrow$ | $N_{\text{tot}} = N_s$                         | ; |
| (St3) Approche MIS à partir de $N_e = 5$ appels à GIS   | $\Rightarrow$ | $N_{\text{tot}} = N_e \times N_s$              | . |

Ici, comme une évaluation de  $\phi$  représente un coût négligeable, on ne travaille pas à budget fixé. Quelle que soit la stratégie mise en œuvre, pour chaque appel à GIS, on s'autorise la simulation de  $N_s = 3 \times 10^3$  échantillons  $\mathbf{X}^{(i)}$  dans l'espace normal standard. Pour les stratégies (St1) à (St3), les résultats obtenus à partir de  $\mathbf{x}_{\text{obs}}^{(1)}$  et après  $N_{\text{rep}} = 10$  estimations sont présentés dans le Tableau 4.5 où ils sont comparés aux résultats de référence issus de l'ASF conduite avec FORM.

On constate que les quatre algorithmes d'ASF rendent des résultats équivalents tant au niveau des valeurs moyennes que des indicateurs de dispersion. La proximité des résultats entre les deux procédures séquentielles, à savoir FORM et (St1), indique que l'on peut remplacer l'algorithme déterministe  $\mathcal{R}$  (ici FORM) par un algorithme de simulation  $\mathcal{R}_w$  (ici GIS) sans craindre de modifier les résultats de l'ASF. La stochasticité provenant du tirage aléatoire des échantillons  $\mathbf{X}^{(i)}$  n'introduit pas un bruit suffisamment important pour que les variables  $P_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$  et  $\hat{P}_f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$  ne puissent plus être considérées comme étant globalement de même loi. Par ailleurs, le fait que les stratégies (St2) et (St3) donnent les mêmes résultats que la stratégie (St1) donne du crédit à l'idée selon laquelle l'ASF peut être effectuée par échantillonnage préférentiel sans pour autant devoir réaliser  $M_t$  explorations. Si (St1) impose  $N_{\text{tot}} = 15 \times 10^6$  appels à  $\phi$ , l'astuce "inverse-reverse" utilisée par (St2)

					ASF exécutée avec $\mathcal{B}$ :							
					$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_{\text{cop}}$				
Échantillon	$\mathbf{x}_{\text{obs}}^{(1)}$	Algorithme de fiabilité	$\mathcal{R}$	FORM		moyenne :	28.26 %	26.56 %	5.03 %	37.49 %		
						écart-type :	4.18 %	4.35 %	3.85 %	4.19 %		
								rang :	(2)	(3)	(4)	(1)
			$\mathcal{R}_w$	GIS	Explorations	$N_e = M_t$		moyenne :	25.14 %	25.38 %	4.63 %	35.32 %
								écart-type :	4.64 %	4.99 %	2.82 %	6.54 %
									rang :	(2)	(2)	(3)
$N_e = 1$		moyenne :				28.48 %	26.68 %	5.01 %	37.53 %			
		écart-type :	3.51 %	5.17 %	3.40 %	2.92 %						
					rang :	(2)	(3)	(4)	(1)			
$N_e = 5$		moyenne :	27.76 %	26.71 %	5.58 %	38.01 %						
		écart-type :	3.36 %	3.23 %	2.79 %	2.50 %						
					rang :	(2)	(2)	(3)	(1)			

TABLEAU 4.5 – Résultats de l'ASF avec  $\mathcal{R}$  et avec différentes versions de  $\mathcal{R}_w$ .

permet d'obtenir les mêmes résultats avec seulement  $N_{\text{tot}} = 3 \times 10^3$  appels à  $\phi$ . Évidemment, la grande simplicité de ce cas test peut inciter à la prudence quant aux conclusions qu'on se permet d'en tirer. La question est alors de savoir si ces conclusions tiennent toujours lorsque  $\mathcal{M}$  met en jeu des outils de modélisation probabiliste plus sophistiqués ou lorsque  $\phi$  est non-linéaire.

## 4.5.2 Étude d'un cas test non-linéaire

### 4.5.2.1 Description

On considère à nouveau le vecteur d'entrée  $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  défini à l'équation (4.50) mais la fonction de performance est modifiée par ajout d'un terme d'interaction entre  $X_1$  et  $X_2$  :

$$Y = \phi(\mathbf{X}) = a_1 X_1 + a_2 X_2 + a_3 X_3 + a_4 X_1 X_2. \quad (4.115)$$

Les différents paramètres sont réglés comme suit :

- Fonction de performance  $\phi$  :  $a_1 = a_2 = 2$ ;  $a_3 = a_4 = 1$
- Vecteur moyen  $\boldsymbol{\mu}$  :  $\mu_1 = \mu_2 = \mu_3 = 3$
- Matrice de covariance  $\boldsymbol{\Sigma}$  :  $\sigma_1 = \sigma_2 = \sigma_3 = 1$ ;  $\sigma_{12} = 0.5$ ;  $\sigma_{13} = \sigma_{23} = 0.3$

Après avoir supposé que les constantes  $a_1, \dots, a_4$  sont toutes strictement positives, on adopte les notations suivantes :

$$\gamma_{\mathbf{a}} = -\frac{a_2}{a_4} \quad ; \quad \delta_{\mathbf{a}} = -\frac{a_1}{a_4} \quad ; \quad \xi_{\mathbf{a}} = \frac{T + a_2 a_4}{a_3 a_4} \quad ; \quad \kappa_{\mathbf{a}}(x_1, x_3) = \frac{T - a_1 x_1 - a_3 x_3}{a_2 + a_4 x_1}. \quad (4.116)$$

Après de minces efforts de réécriture, on peut facilement montrer que le domaine de défaillance  $D_f$  est divisible en 4 sous-domaines :

$$D_f = (D_f \cap \{x_3 < \xi_{\mathbf{a}}\}) \cup (D_f \cap \{x_3 > \xi_{\mathbf{a}}\}), \quad (4.117)$$

$$= (D_f^{A_1} \cup D_f^{A_2}) \cup (D_f^{B_1} \cup D_f^{B_2}), \quad (4.118)$$

dont les expressions sont respectivement données par :

$$\begin{aligned} D_f^{A_1} &:= \{x_1 > \gamma_{\mathbf{a}} \text{ avec } x_3 < \xi_{\mathbf{a}} \text{ et } x_2 > \kappa_{\mathbf{a}}(x_1, x_3)\}; \\ D_f^{A_2} &:= \{x_1 < \gamma_{\mathbf{a}} \text{ avec } x_3 < \xi_{\mathbf{a}} \text{ et } x_2 < \kappa_{\mathbf{a}}(x_1, x_3)\}; \\ D_f^{B_1} &:= \{x_1 > \gamma_{\mathbf{a}} \text{ avec } x_3 > \xi_{\mathbf{a}} \text{ et } x_2 > \kappa_{\mathbf{a}}(x_1, x_3)\}; \\ D_f^{B_2} &:= \{x_1 < \gamma_{\mathbf{a}} \text{ avec } x_3 > \xi_{\mathbf{a}} \text{ et } x_2 < \kappa_{\mathbf{a}}(x_1, x_3)\}. \end{aligned} \quad (4.119)$$

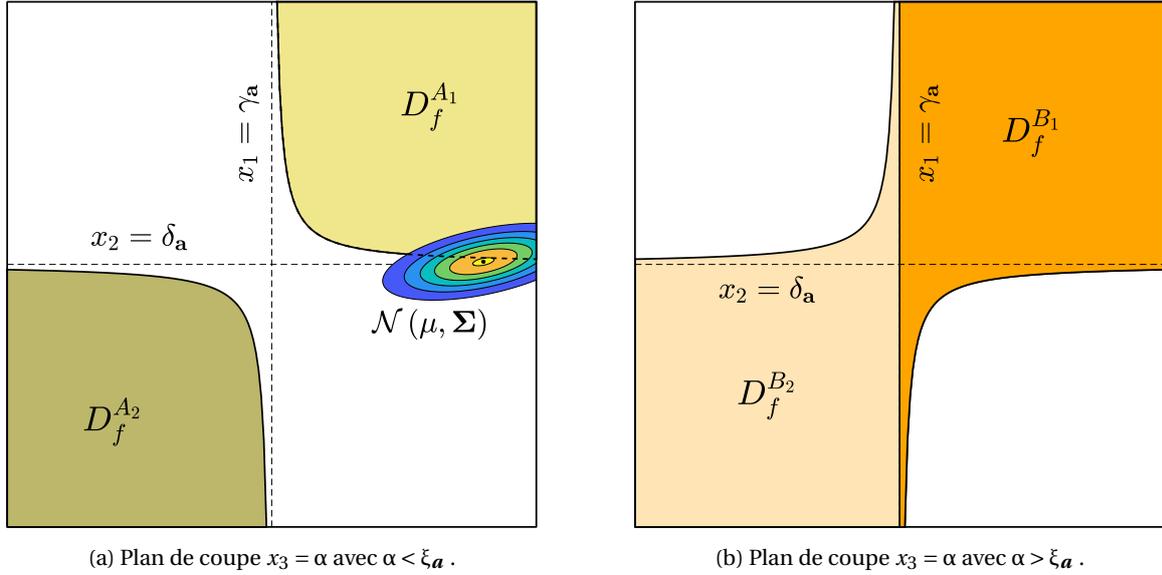


FIGURE 4.20 – Visualisation du domaine de défaillance  $D_f$ .

Le plan d'équation  $x_3 = \xi_a$  sépare la paire de sous-domaines  $D_f^{A_1}$  et  $D_f^{A_2}$  de la paire de sous-domaines  $D_f^{B_1}$  et  $D_f^{B_2}$ . On peut se reporter à la Figure 4.20 pour un aperçu des différents modes de défaillance envisageables. Les Figures 4.20a et 4.20b montrent l'évolution du profil de l'état limite de par et d'autre du plan  $x_3 = \xi_a$ . Si  $D_f^{A_1}$  et  $D_f^{A_2}$  sont disjoints,  $D_f^{B_1}$  et  $D_f^{B_2}$  sont au contraire tangents au niveau du plan  $x_1 = \gamma_a$ . Pour les valeurs numériques choisies plus haut, on a  $\mu_3 < \xi_a$  donc le vecteur moyen  $\mu$  est situé dans le demi-espace des sous-domaines  $D_f^{A_1}$  et  $D_f^{A_2}$ . Par ailleurs, les valeurs prises par  $\mu_1$  et  $\mu_2$  positionnent  $\mu$  dans le très proche voisinage de  $D_f^{A_1}$ . Ainsi, il devient tout à fait raisonnable d'écrire que :

$$P_f = \int_{D_f} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \approx \int_{D_f^{A_1}} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}. \quad (4.120)$$

Si le problème pouvait à première vue sembler faire intervenir plusieurs sous-domaines de défaillance, la nature de  $f_{\mathbf{X}}$  ne rend en fait actif que le domaine  $D_f^{A_1}$ . Pour  $T = 50$ , l'application par force brute d'une approche CMC permet d'estimer  $P_f$  avec précision. Avec  $N_s = 10^7$ , on obtient  $\hat{P}_f^{\text{CMC}} = \mathcal{R}_w(f_{\mathbf{X}}) \approx 7.8 \times 10^{-3}$ . Au vu de ce qui a été examiné dans l'espace physique, il ne fait aucun doute que le domaine de défaillance  $\hat{D}_f^{\mathcal{N}}$  dans l'espace normal standard n'est pas linéaire. Toutefois, comme on le voit sur la Figure 4.21a, il n'est pas non plus aberrant d'approcher l'état limite  $H(\mathbf{w}) = T$  par l'hyperplan  $\mathbf{w}^* \cdot (\mathbf{w} - \mathbf{w}^*) = 0$  correspondant à l'approximation faite en utilisant l'algorithme FORM. Numériquement, on obtient  $\hat{P}_f^{\text{FORM}} = \mathcal{R}(f_{\mathbf{X}}) = 8.2 \times 10^{-3}$  et cela confirme que la méthode FORM surestime légèrement  $P_f$  par convexité de  $\hat{D}_f^{\mathcal{N}}$  pour ce cas test.

#### 4.5.2.2 Visualisation de l'impact de la variabilité d'apprentissage

Comme dans la Section 4.3.1, on suppose dans un premier temps que les générateurs théoriques  $G_Z$  et  $G_V$  sont accessibles. Toutefois, on fait légèrement évoluer le modèle d'apprentissage utilisé pour estimer  $f_{\mathbf{X}}$ . L'idée est de choisir  $\mathcal{M}$  dans l'esprit de ce que pourrait faire un utilisateur lambda n'ayant aucune connaissance du modèle de distribution  $\mu_{\mathbf{X}}$  sous-jacent. Comme dans l'équation (4.61),  $\mathcal{M}_{\text{cop}}$  continue de représenter l'apprentissage par maximum de vraisemblance d'une copule gaussienne à partir de  $\mathbf{V}_{\text{obs}}$  en suivant les étapes (R1) à (R3). Ce choix, certes discutable pour les spécialistes de la théorie des copules, est cependant très courant dans de nombreuses applications. Pour l'apprentissage de la  $j$ -ème densité marginale, on choisit de se tourner

vers une méthode KDE calquée sur ce qui est fait dans l'équation (4.95) :

$$\hat{f}_{X_j}(x_j | \mathcal{M}_j, \mathbf{Z}_{\text{obs}}^{[\bullet j]}) = \hat{f}_{\text{KDE}}(x_j | \mathbf{Z}_{\text{obs}}^{[\bullet j]}, \hat{h}_j). \quad (4.121)$$

Si on peut encore considérer que la modélisation de la structure de dépendance relève d'une certaine expertise, il n'en est pas de même pour l'identification paramétrique des marginales. Les dangers d'une utilisation systématique de la famille gaussienne ont été clairement mis en lumière et les approches non-paramétriques se sont désormais largement répandues en pratique. Après ré-assemblage des différentes composantes fonctionnelles, la densité conjointe issue de l'apprentissage statistique peut être appréhendée de la manière suivante :

$$\hat{f}_{\mathbf{X}}(\cdot | \mathcal{M}, \mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}}) = \hat{f}_{\mathbf{X}}(\cdot | \mathbf{Z}_{\text{obs}}, \hat{\mathbf{h}}, \hat{\boldsymbol{\rho}}^{\text{MV}}) \quad \text{avec} \quad \hat{\mathbf{h}} := [\hat{h}_j]_{1 \leq j \leq d}. \quad (4.122)$$

L'indépendance à l'œuvre entre les colonnes de  $\mathbf{Z}_{\text{obs}}$  ainsi qu'entre  $\mathbf{Z}_{\text{obs}}$  et  $\mathbf{V}_{\text{obs}}$  garantit à nouveau l'indépendance des blocs se rapportant aux différentes composantes :

$$\underbrace{(\mathbf{Z}_{\text{obs}}^{[\bullet 1]}, \hat{h}_1)}_{\text{pour } \hat{f}_{X_1}} \perp \underbrace{(\mathbf{Z}_{\text{obs}}^{[\bullet 2]}, \hat{h}_2)}_{\text{pour } \hat{f}_{X_2}} \perp \underbrace{(\mathbf{Z}_{\text{obs}}^{[\bullet 3]}, \hat{h}_3)}_{\text{pour } \hat{f}_{X_3}} \perp \underbrace{(\hat{\boldsymbol{\rho}}_{12}^{\text{MV}}, \hat{\boldsymbol{\rho}}_{13}^{\text{MV}}, \hat{\boldsymbol{\rho}}_{23}^{\text{MV}})}_{\text{pour } \hat{\boldsymbol{\rho}}}. \quad (4.123)$$

Pour un modèle d'apprentissage  $\mathcal{M}$  combinant ainsi une estimation non-paramétrique des lois marginales et la calibration d'une copule gaussienne, il est bon de remarquer que l'incertitude liée au manque de données n'est plus la seule source d'incertitudes épistémiques. En effet, estimer par KDE la densité  $f_{X_j}$  revient à chercher le meilleur représentant  $\hat{f}_{X_j}$  au sein d'un espace fonctionnel présentant une intersection avec la famille gaussienne sans pour autant y être inclus. Une incertitude liée à la méconnaissance du modèle de distribution sous-jacent est ainsi introduite par le choix d'un apprentissage non-paramétrique des lois marginales. Dans cette situation, les indices de Sobol calculés par l'algorithme d'ASF offrent une grille de décomposition de la variabilité provoquée par l'effet combiné du "small data" et du choix de modélisation fait par le statisticien pour se prémunir contre les risques inhérents au cadre paramétrique.

Comme dans le cas test linéaire, on peut construire une approximation de  $D_f^{\mathcal{N}}$  après finalisation de la phase d'inférence. On a alors :

$$\hat{D}_f^{\mathcal{N}} = \{\mathbf{w} \in \mathbb{R}^d : \hat{H}(\mathbf{w}) > T\} \quad \text{avec} \quad \hat{H} = \phi \circ \hat{\mathcal{T}}^{-1}, \quad (4.124)$$

où  $\hat{\mathcal{T}}$  est une transformation isoprobabiliste permettant de fabriquer une variable gaussienne centrée réduite à partir d'un vecteur de densité conjointe (4.122). Une telle transformation reste dans l'éventail des cas pouvant être traités avec la méthode de Nataf qui s'exécute alors en deux étapes.  $\mathcal{T}_1$  permet de transformer les marginales issues de l'apprentissage KDE en des lois normales centrées réduites.  $\mathcal{T}_2$  permet ensuite d'éliminer la structure de dépendance que  $\mathcal{T}_1$  a laissé inchangée. Par souci de simplicité, on allège les notations des marginales estimées par KDE :

$$\begin{aligned} \hat{g}_j(x_j) &:= \hat{f}_{\text{KDE}}(x_j | \mathbf{Z}_{\text{obs}}^{[\bullet j]}, \hat{h}_j) \quad ; \quad \hat{G}_j(x_j) := \hat{F}_{\text{KDE}}(x_j | \mathbf{Z}_{\text{obs}}^{[\bullet j]}, \hat{h}_j) \\ \text{et} \quad \hat{g}(\mathbf{x}) &:= \hat{f}_{\mathbf{X}}(\mathbf{x} | \mathbf{Z}_{\text{obs}}, \hat{\mathbf{h}}, \hat{\boldsymbol{\rho}}^{\text{MV}}). \end{aligned} \quad (4.125)$$

Les opérations à exécuter pour réussir la transformation de Nataf sont détaillées ci-dessous :

$$\begin{aligned} \hat{\mathcal{T}} : \quad \mathbf{x} &\longrightarrow \mathbf{s} = [\Phi^{-1}(\hat{G}_j(x_j))]_j \longrightarrow \mathbf{w} = (\hat{\mathbf{L}}^{\text{MV}})^{-1} \mathbf{s} \\ \mathbf{X} \sim \hat{g}(\cdot) &\xrightarrow{\hat{\mathcal{T}}_1} \mathbf{S} \sim \hat{f}_{\mathcal{N}}(\cdot | \mathbf{0}, \hat{\boldsymbol{\rho}}^{\text{MV}}) \xrightarrow{\hat{\mathcal{T}}_2} \mathbf{W} \sim \varphi_3(\cdot) \end{aligned} \quad (4.126)$$

S'il était prévisible que l'état limite  $\hat{H}(\mathbf{w}) = [\phi \circ \hat{\mathcal{T}}_1^{-1} \circ \hat{\mathcal{T}}_2^{-1}](\mathbf{w}) = T$  hérite de la non-linéarité portée par  $\phi$ , il est intéressant de remarquer que la transformation  $\hat{\mathcal{T}}_1^{-1}$  des lois marginales rajoute

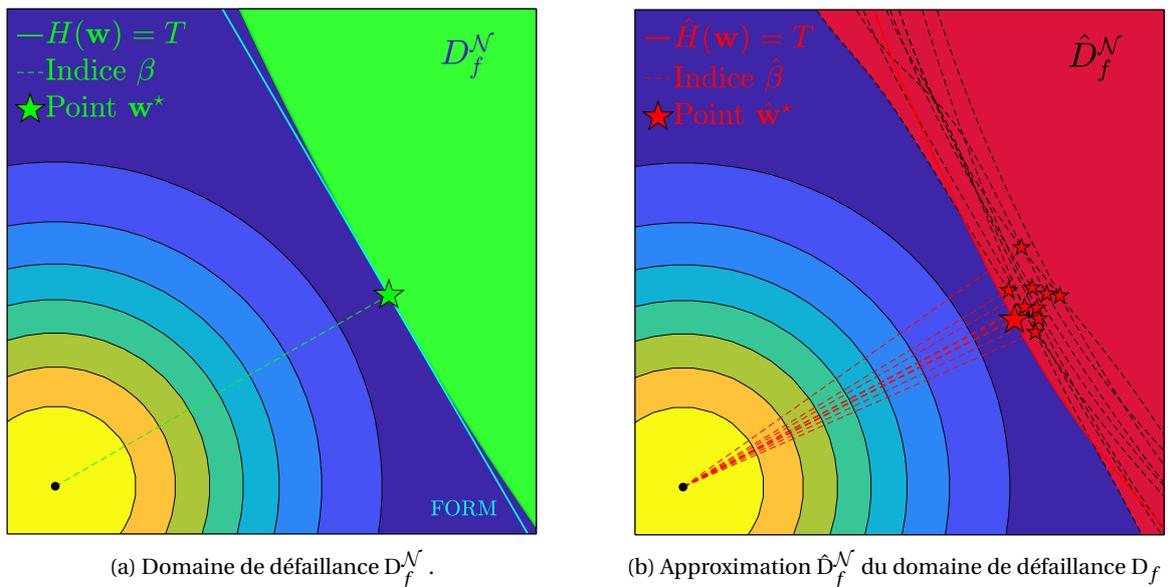


FIGURE 4.21 – Visualisation de l’impact du contexte “small data” sur l’estimation du risque de défaillance.

une forme supplémentaire de non-linéarité. L’approximation d’une loi gaussienne par un mélange de noyaux gaussiens se traduit ainsi par une déformation supplémentaire de  $D_f^N$ .

Pour le cas test linéaire, dans lequel le modèle de distribution sous-jacent est connu, on a pu observer à la Figure 4.10b que les différentes approximations  $\hat{H}(w) = T$  de l’état limite théorique  $H(w) = T$  restent linéaires. Seule la position du point de conception  $\hat{w}^*$  évolue en fonction de l’échantillon fourni. Dans la présente configuration, la proximité morphologique entre  $D_f$  et  $\hat{D}_f$  n’est plus vraie. Ce phénomène est par exemple observable à la Figure 4.21b où l’on superpose les courbes d’état limite obtenues pour différents échantillons d’apprentissage. À la Figure 4.21a, on peut constater que l’état limite théorique  $H(w) = T$  a un profil de type hyperbolique. Un examen au cas par cas des courbes de la Figure 4.21b révèle au contraire que la plupart des approximations  $\hat{H}(w) = T$  ne sont pas hyperboliques et mettent par exemple en jeu plusieurs points d’inflexion. Cela peut s’interpréter comme le contrecoup d’un apprentissage non-paramétrique des densités marginales. Certes, on ne se hasarde pas à faire une hypothèse paramétrique sur  $\mu_X$  mais il faut accepter que la géométrie de  $D_f^N$  soit déformée à coup sûr, ce qui peut devenir problématique dans un contexte “small data” où le niveau de convergence permis par KDE n’est pas toujours suffisant. Travailler dans l’espace normal standard a permis de mieux cerner le rôle joué par les différents types d’incertitudes épistémiques. Celle qui est liée à la méconnaissance du modèle sous-jacent déforme le domaine de défaillance théorique tandis que celle qui est liée au manque de données rend instable le paramétrage de la configuration géométrique induite par les choix faits au moment de l’élaboration du modèle d’apprentissage.

#### 4.5.2.3 À propos des indices obtenus avec les générateurs théoriques

Pour ce cas test, les indices de Sobol théoriques  $S_1, S_2, S_3$  et  $S_{\text{cop}}$  s’expriment comme suit :

$$S_j := \frac{\mathbb{V}\left(\mathbb{E}\left[\mathbb{P}_f(\mathbf{Z}_{\text{obs}}, \hat{\mathbf{h}}, \hat{\boldsymbol{\rho}}) \mid \mathbf{Z}_{\text{obs}}^{*j}, \hat{h}_j\right]\right)}{\mathbb{V}\left(\mathbb{P}_f(\mathbf{Z}_{\text{obs}}, \hat{\mathbf{h}}, \hat{\boldsymbol{\rho}})\right)} \quad \text{et} \quad S_{\text{cop}} := \frac{\mathbb{V}\left(\mathbb{E}\left[\mathbb{P}_f(\mathbf{Z}_{\text{obs}}, \hat{\mathbf{h}}, \hat{\boldsymbol{\rho}}) \mid \hat{\boldsymbol{\rho}}\right]\right)}{\mathbb{V}\left(\mathbb{P}_f(\mathbf{Z}_{\text{obs}}, \hat{\mathbf{h}}, \hat{\boldsymbol{\rho}})\right)}. \quad (4.127)$$

En s’inspirant de ce qui a été fait dans le paragraphe 4.3.1.3, on peut à nouveau justifier rigoureusement que  $S_1 = S_2$ . En effet, l’indépendance mutuelle des composantes fonctionnelles a été mise en évidence par l’équation (4.123) et on retrouve des effets de symétrie entre  $X_1$  et  $X_2$  :

- (Sym1) **Échantillonnage :**  $\begin{cases} \mathbf{Z}_{\text{obs}}^{[\bullet 1]} \sim (\mu_{X_1})^{\otimes N} \\ \mathbf{Z}_{\text{obs}}^{[\bullet 2]} \sim (\mu_{X_2})^{\otimes N} \end{cases}$  avec  $\mu_{X_1} = \mu_{X_2}$
- (Sym2) **Apprentissage :**  $\mathcal{M}_1$  et  $\mathcal{M}_2$  identiques : KDE calibré avec (4.95)
- (Sym3) **Motif de dépendance avec  $X_3$  :**  $\begin{cases} (\mathbf{V}_{\text{obs}}^{[\bullet 1]}, \mathbf{V}_{\text{obs}}^{[\bullet 3]}) \sim \mathcal{CN}(\cdot | \rho_{13}) \\ (\mathbf{V}_{\text{obs}}^{[\bullet 2]}, \mathbf{V}_{\text{obs}}^{[\bullet 3]}) \sim \mathcal{CN}(\cdot | \rho_{23}) \end{cases}$  avec  $\rho_{13} = \rho_{23}$
- (Sym4) **Fonction de performance :**  $a_1 = a_2 = 2$

On peut alors reprendre les différents calculs du paragraphe 4.3.1.3 pour montrer que :

$$\mathbb{E} \left[ P_f(\mathbf{Z}_{\text{obs}}, \hat{\mathbf{h}}, \hat{\boldsymbol{\rho}}) \mid \mathbf{Z}_{\text{obs}}^{[\bullet 1]}, \hat{h}_1 \right] = \mathbb{E} \left[ P_f(\mathbf{Z}_{\text{obs}}, \hat{\mathbf{h}}, \hat{\boldsymbol{\rho}}) \mid \mathbf{Z}_{\text{obs}}^{[\bullet 2]}, \hat{h}_2 \right], \quad (\text{Obj})$$

et conclure très naturellement que  $S_1 = S_2$ . On souhaite vérifier ce résultat numériquement. Pour espérer retrouver cette égalité avec suffisamment de précision, il faut évidemment avoir recours aux générateurs théoriques  $G_Z$  et  $G_V$ . L'évaluation des  $M_t$  probabilités de défaillance  $P_f(\mathbf{Z}_{\text{obs}}, \hat{\mathbf{h}}, \hat{\boldsymbol{\rho}})$  requises par l'ASF est conduite à l'aide d'une stratégie "inverse-reverse" utilisant l'algorithme GIS (avec  $N_s = 3 \times 10^3$ ) pour réaliser l'unique exploration. Après  $N_{\text{rep}} = 50$  exécutions de cette version de l'algorithme d'ASF, on obtient les résultats suivants :

ASF :	$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_{\text{cop}}$
moyenne :	42.97 %	43.60 %	3.45 %	12.10 %
écart-type :	3.40 %	3.09 %	1.94 %	3.86 %
rang :	(1)	(1)	(3)	(2)

(4.128)

On obtient comme prévu  $\hat{S}_1 \approx \hat{S}_2$ . Les résultats sont très semblables à ceux obtenus pour le cas test linéaire. Toutefois, au sein de la collection des indices de Sobol, la prédominance de  $\hat{S}_1$  et  $\hat{S}_2$  a été renforcée. Bien qu'étant toujours plus grand que  $\hat{S}_3$ ,  $\hat{S}_{\text{cop}}$  ne peut plus être considéré comme appartenant au lot des plus grands indices. Ainsi, pour ce cas test, c'est majoritairement l'apprentissage des densités marginales  $\hat{f}_{X_1}$  et  $\hat{f}_{X_2}$  qui contribue à la variabilité observée sur  $P_f(\mathbf{Z}_{\text{obs}}, \hat{\mathbf{h}}, \hat{\boldsymbol{\rho}})$ . Si on compare ces résultats aux indices théoriques (4.80) du cas test linéaire, on remarque que les écarts-type sont du même ordre de grandeur alors qu'on utilise ici  $\mathcal{R}_w$  et non plus  $\mathcal{R}$ . Preuve est donc à nouveau faite qu'un algorithme de simulation  $\mathcal{R}_w$  est tout à fait compatible avec la procédure d'ASF à condition, comme ici, que  $\mathcal{R}_w$  convienne à la configuration traitée.

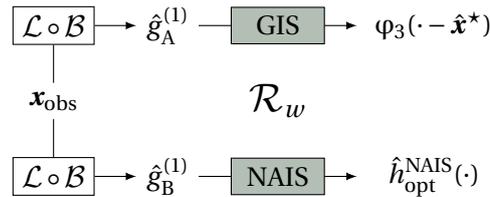
#### 4.5.2.4 Comparaison des résultats obtenus pour divers algorithmes de fiabilité

À l'instar de ce qui a été fait au paragraphe 4.5.1 avec le cas test linéaire, on cherche ici à comparer les résultats rendus par la procédure d'ASF lorsque différentes stratégies sont mises en œuvre pour estimer la série des probabilités  $P_f^{(k)}$  que l'on peut ici noter  $P_f(\bar{\mathbf{Z}}^{(k)}, \hat{\mathbf{h}}^{(k)}, \hat{\boldsymbol{\rho}}^{(k)})$ . Pour un algorithme de simulation  $\mathcal{R}_w$  donné, GIS en l'occurrence, on a pu observer que le nombre d'explorations réalisées avait assez peu d'influence sur les résultats de l'ASF. On souhaite vérifier qu'il en va de même lorsqu'on remplace GIS par une autre méthode d'échantillonnage préférentiel. Pour construire une densité auxiliaire d'une tout autre manière, on propose donc d'utiliser l'algorithme *Nonparametric Adaptive Importance Sampling* (NAIS) développé par MORIO [2012]. De nature adaptative, cet algorithme s'appuie sur plusieurs étapes de ré-échantillonnage pour faire migrer un nuage initial de  $N_s$  particules en direction du domaine de défaillance. Les échantillons arrivant à destination sont alors autant de noyaux gaussiens que l'on doit pondérer pour construire une densité auxiliaire sous-optimale  $\hat{h}_{\text{opt}}$ . Le budget de simulation consommé par NAIS pour accomplir une exploration dépend du nombre  $N_i$  d'itérations nécessaires pour atteindre  $D_f$ . Pour les réglages choisis dans ce cas test,  $N_i = 2$  itérations suffisent. On se donne un échantillon  $\mathbf{x}_{\text{obs}}$  et on lance  $N_{\text{rep}} = 10$  fois l'algorithme d'ASF pour chacune des trois stratégies suivantes :

(St1) Procédure séquentielle avec  $M_t$  appels à FORM  $\Rightarrow N_{\text{tot}} = M \times (d + 2) \times N_{\text{opt}}$  ;

- (St2) Astuce “*inverse-reverse*” avec un seul appel à GIS  $\Rightarrow N_{\text{tot}} = N_s$  ;
- (St3) Astuce “*inverse-reverse*” avec un seul appel à NAIS  $\Rightarrow N_{\text{tot}} = N_i \times N_s$  .

Une synthèse des résultats est disponible au Tableau 4.6 où l’on constate que les trois méthodes estiment de façon identique tous les indices de Sobol. La similarité des résultats produits par (St2) et (St3) est un constat particulièrement rassurant. Alors que les ressorts de la phase d’exploration sont radicalement différents entre GIS et NAIS, les deux densités auxiliaires  $\mathbf{x} \rightarrow \varphi_3(\mathbf{x} - \hat{\mathbf{x}}^*)$  et  $\mathbf{x} \rightarrow \hat{h}_{\text{opt}}^{\text{NAIS}}(\mathbf{x})$  semblent interchangeables. En effet, pour toute densité d’entrée  $\hat{g}^{(k)}$  intervenant dans l’ASF, on construit pratiquement le même estimateur  $\hat{P}_f^{(k)}$  selon qu’on applique l’astuce “*inverse-reverse*” avec l’une ou avec l’autre. C’est d’autant plus remarquable que ces deux densités ont été entraînées selon le schéma suivant :



c’est-à-dire à partir de densités initiales  $\hat{g}_A^{(1)}$  et  $\hat{g}_B^{(1)}$  différentes. L’astuce “*inverse-reverse*” donne donc des résultats satisfaisants quelle que soit la densité  $\hat{g}$  ayant servi à initier la seule exploration mais aussi quel que soit l’algorithme de simulation  $\mathcal{R}_w$  sélectionné. On peut en revanche s’étonner du fait que les résultats de (St1) soient identiques à ceux de (St2) et (St3) étant donné qu’on a vérifié numériquement et graphiquement que FORM surestime légèrement le risque de défaillance. Cela semble indiquer que l’erreur introduite par une approximation linéaire de l’état limite dans l’espace normal standard n’est pas d’assez grande amplitude pour que la distribution de  $\hat{P}_f(\bar{\mathbf{Z}}, \hat{\mathbf{h}}, \hat{\rho})$  soit réellement impactée par le fait d’utiliser FORM ou GIS. Il est aussi bon de noter que les valeurs moyennes des indices estimés à partir du seul  $\mathbf{x}_{\text{obs}}$  sont pratiquement égales aux valeurs des indices théoriques (4.128) obtenues avec  $G_Z$  et  $G_V$ . On est donc amené à penser que l’échantillon  $\mathbf{x}_{\text{obs}}$  utilisé pour l’ASF est particulièrement représentatif de la loi  $\mu_X$  au sens où réaliser des tirages aléatoires selon la mesure empirique  $\hat{\mu}_{ZV}[\mathbf{x}_{\text{obs}}]$  suffit à correctement imiter la variabilité d’échantillonnage du problème traité.

		ASF exécutée avec $\mathcal{B}$ :				$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_{\text{cop}}$		
Échantillon	$\mathbf{x}_{\text{obs}}$	Algorithme de fiabilité	$\mathcal{R}$	FORM		moyenne :	42.19 %	43.68 %	3.18 %	10.64 %	
						écart-type :	3.99 %	3.70 %	1.53 %	3.92 %	
			rang :	(1)	(1)	(3)	(2)				
			$\mathcal{R}_w$	GIS	Explorations	$N_e = 1$	moyenne :	41.75 %	42.55 %	3.67 %	17.29 %
							écart-type :	2.63 %	3.53 %	2.82 %	3.32 %
				rang :	(1)	(1)	(3)	(2)			
NAIS	Explorations	$N_e = 1$		moyenne :	40.75 %	43.93 %	3.23 %	17.95 %			
			écart-type :	2.99 %	3.83 %	3.08 %	3.15 %				
rang :	(2)	(1)	(4)	(3)							

TABLEAU 4.6 – Résultats de l’ASF avec  $\mathcal{R}$  et avec différentes versions de  $\mathcal{R}_w$  .

#### 4.5.2.5 Assimilation de données supplémentaires

On vient de valider que l’algorithme d’ASF est capable de répondre au Problème 4.2 de décomposition de la variabilité d’apprentissage dans un cadre général pouvant incorporer des distributions non-gaussiennes ainsi que des fonctions de performances non-linéaires. Il a également été montré qu’on est en mesure de gérer efficacement la charge de calcul inhérente au déploiement des outils d’analyse de fiabilité adéquats. On souhaite s’assurer que le Problème 4.3 de réduction

de la variabilité d'apprentissage peut toujours être traité en collectant puis en assimilant des observations supplémentaires.

On met en place les mêmes expérimentations numériques que celles présentées au paragraphe 4.5.2.5 pour le cas test linéaire. On reprend l'échantillon  $\mathbf{x}_{\text{obs}}$  qui a servi à obtenir les résultats du Tableau 4.6 et toutes les procédures d'ASF sont conduites en implémentant l'astuce "inverse-reverse" à partir d'une densité auxiliaire construite avec l'algorithme GIS. Chacun des  $d+1$  scénarios d'assimilation est expérimenté  $N_{\text{rep}} = 10$  fois pour un nombre  $N' \in \{500, 10^3\}$  d'observations supplémentaires. Les résultats sont présentés au Tableau 4.7 où le premier étage rappelle les résultats de l'ASF initiale. Dans le cas de l'assimilation d'observations supplémentaires pour la variable  $X_1$ , les indices évoluent suite à la stabilisation de l'apprentissage de  $f_{X_1}$  :

$$\hat{S}_1^{\mathcal{B}_1} < \hat{S}_1^{\mathcal{B}} \quad \text{tandis que} \quad \hat{S}_2^{\mathcal{B}_1} > \hat{S}_2^{\mathcal{B}} \quad ; \quad \hat{S}_{\text{cop}}^{\mathcal{B}_1} > \hat{S}_{\text{cop}}^{\mathcal{B}} \quad \text{et} \quad \hat{S}_3^{\mathcal{B}_1} \approx \hat{S}_3^{\mathcal{B}}. \quad (4.129)$$

Ce constat est similaire à celui qui avait été fait pour le cas test linéaire. Après réduction voire élimination de la variabilité induite par  $\hat{f}_{X_1}$ , la redistribution des parts de la variabilité restante profite principalement à  $\hat{f}_{X_2}$  et  $\hat{c}_{\mathbf{U}}$  qui deviennent les contributeurs principaux. Rappelons que  $\hat{f}_{X_1}$  est issue d'une estimation KDE de  $f_{X_1}$  à partir d'un échantillon  $\mathbf{x}_{\text{obs}}^{[\bullet,1]}$ . En intégrant un lot  $\mathbf{x}_{\text{extra}}^{[\bullet,1]}$  de nouvelles observations à cet apprentissage, on agit simultanément sur les deux formes d'incertitudes épistémiques. L'incertitude liée à la méconnaissance du modèle sous-jacent et l'incertitude liée au manque de données diminuent en effet conjointement au gré de l'incorporation d'un nombre croissant de noyaux et de la convergence progressive de l'estimation faite par la méthode KDE. Si on examine les étages visant à neutraliser l'influence de  $\hat{f}_{X_2}$  et de  $\hat{c}_{\mathbf{U}}$ , des tendances analogues sont observables. Par contre, en ce qui concerne l'assimilation de données supplémentaires de  $X_3$ , aucun changement n'affecte le classement des indices car la part de variabilité véhiculée par  $\hat{f}_{X_3}$  est trop faible au regard de celles véhiculées par les autres contributeurs.

Force est de constater que les Tableaux 4.3 et 4.7 livrent les mêmes enseignements quant à l'efficacité de l'implémentation d'une stratégie de réduction de la variabilité d'apprentissage par assimilation ciblée de nouvelles données. Dans le cas de l'utilisation d'un algorithme de simulation  $\mathcal{R}_w$ , il n'était pourtant pas certain que les résultats de la seconde ASF soient cohérents avec ceux de l'ASF initiale. En effet, rien ne garantit a priori que la variabilité d'apprentissage, une fois réduite, demeure plus importante que la variabilité induite par la stochasticité de  $\mathcal{R}_w$ . Si  $X_1$  fait l'objet d'une acquisition ciblée de données, on peut en effet se retrouver dans une situation où :

$$\underbrace{\mathbb{V}\left(\mathbb{P}_f(\mathbf{X}_{\text{obs}} \cup \mathbf{X}_{\text{extra}}^{[\bullet,1]})\right) = \mathbb{V}\left(\mathcal{R}\left(\hat{f}_{\mathbf{X}}(\cdot | \mathbf{X}_{\text{obs}} \cup \mathbf{X}_{\text{extra}}^{[\bullet,1]})\right)\right)}_{\text{variabilité d'apprentissage}} \approx \underbrace{\mathbb{V}\left(\mathcal{R}_w(f_{\mathbf{X}})\right) = \mathbb{V}(\hat{\mathbb{P}}_f)}_{\text{stochasticité de } \mathcal{R}_w}, \quad (4.130)$$

et où les distributions des variables  $\mathbb{P}_f(\mathbf{X}_{\text{obs}} \cup \mathbf{X}_{\text{extra}}^{[\bullet,1]})$  et  $\hat{\mathbb{P}}_f(\mathbf{X}_{\text{obs}} \cup \mathbf{X}_{\text{extra}}^{[\bullet,1]})$  ne peuvent plus être considérées comme identiques. Les indices de Sobol rendus par l'algorithme d'ASF ne constituent alors plus une grille de décomposition pertinente de la variabilité d'apprentissage. Dans le Tableau 4.7, la cohérence entre les résultats de l'ASF initiale et ceux obtenus après l'exécution des différents scénarios d'assimilation indique que l'algorithme GIS n'introduit pas assez de stochasticité pour perturber la compréhension que l'on peut avoir de la variabilité d'apprentissage.

#### 4.5.2.6 Bilan

L'étude détaillée d'un cas test non-linéaire a montré que l'algorithme d'ASF initial peut être adapté à la réalité des problèmes rencontrés en fiabilité industrielle. Dès lors qu'on dispose d'une méthode d'échantillonnage préférentiel suffisamment performante pour construire un estimateur non-biaisé et robuste de la probabilité de défaillance, l'ASF peut être conduite dans d'aussi bonnes conditions que s'il existait un algorithme  $\mathcal{R}$  permettant d'accomplir cette tâche de manière exacte et déterministe. Sous réserve d'un réglage optimal des hyperparamètres de  $\mathcal{R}_w$ , la stochasticité introduite par l'échantillonnage ne transforme que dans une très faible mesure la

ASF exécutée avec $\mathcal{B}$ :				$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_{\text{cop}}$
	$\mathbf{x}_{\text{obs}}$		moyenne :	41.75 %	42.55 %	3.67 %	17.29 %
			écart-type :	2.63 %	3.53 %	2.82 %	3.32 %
			rang :	(1)	(1)	(3)	(2)
ASF exécutée avec $\mathcal{B}_1$ :				$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_{\text{cop}}$
Pour $X_1$	$\mathbf{x}_{\text{obs}} \cup \mathbf{x}_{\text{extra}}^{[*1]}$	$N' = 500$	moyenne :	30.89 %	47.55 %	1.73 %	19.49 %
			écart-type :	4.69 %	4.79 %	1.36 %	2.77 %
			rang :	(2)	(1)	(4)	(3)
		$N' = 10^3$	moyenne :	19.36 %	55.57 %	3.44 %	23.92 %
			écart-type :	3.94 %	3.29 %	1.03 %	2.81 %
			rang :	(3)	(1)	(4)	(2)
ASF exécutée avec $\mathcal{B}_2$ :				$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_{\text{cop}}$
Pour $X_2$	$\mathbf{x}_{\text{obs}} \cup \mathbf{x}_{\text{extra}}^{[*2]}$	$N' = 500$	moyenne :	52.03 %	25.54 %	3.57 %	22.37 %
			écart-type :	3.12 %	4.77 %	2.21 %	4.14 %
			rang :	(1)	(2)	(4)	(3)
		$N' = 10^3$	moyenne :	57.43 %	17.76 %	2.98 %	24.02 %
			écart-type :	2.65 %	4.37 %	2.05 %	3.46 %
			rang :	(1)	(3)	(4)	(2)
ASF exécutée avec $\mathcal{B}_3$ :				$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_{\text{cop}}$
Pour $X_3$	$\mathbf{x}_{\text{obs}} \cup \mathbf{x}_{\text{extra}}^{[*3]}$	$N' = 500$	moyenne :	39.52 %	43.10 %	3.02 %	15.57 %
			écart-type :	4.23 %	3.55 %	1.21 %	4.09 %
			rang :	(2)	(1)	(4)	(3)
		$N' = 10^3$	moyenne :	40.28 %	40.86 %	1.79 %	15.98 %
			écart-type :	2.07 %	3.21 %	2.02 %	3.52 %
			rang :	(2)	(1)	(4)	(3)
ASF exécutée avec $\mathcal{B}_{\text{cop}}$ :				$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_{\text{cop}}$
Pour $\mathbf{U}$	$\mathbf{x}_{\text{obs}} \cup \mathbf{x}_{\text{extra}}$	$N' = 500$	moyenne :	46.11 %	45.43 %	2.33 %	7.16 %
			écart-type :	4.53 %	4.48 %	1.33 %	2.85 %
			rang :	(1)	(1)	(3)	(2)
		$N' = 10^3$	moyenne :	46.79 %	48.43 %	4.83 %	6.43 %
			écart-type :	4.52 %	4.05 %	2.19 %	4.16 %
			rang :	(1)	(1)	(3)	(2)

TABEAU 4.7 – Impact de l’assimilation de données supplémentaires sur le classement des indices de Sobol.

distribution de l'estimateur du risque de défaillance. On a aussi pu constater que des méthodes d'échantillonnage préférentiel concurrentes donnent des résultats identiques. Toutefois, on peut souligner la simplicité du cas test non-linéaire décortiqué dans cette section. L'évaluation de  $\phi$  est gratuite tandis que les Figures 4.20a et 4.21a mettent en exergue l'intervention d'un très faible degré de non-linéarité. La prochaine section est donc consacrée à l'application de l'algorithme d'ASF à une situation réelle où  $\phi$  découle d'une théorie scientifique.

## 4.6 Application au flambage d'une plaque composite

Après avoir étudié deux cas tests purement analytiques, on s'intéresse désormais à un problème de fiabilité très courant dans les applications industrielles : le flambage d'une plaque rectangulaire. Les plaques sont des éléments de structure essentiels dans la plupart des ouvrages d'art de l'ingénierie aéronautique, civile ou offshore et leur étude remonte aux prémices de l'ère industrielle avec les travaux de CAUCHY [1822], NAVIER [1823] ou POISSON [1838] pour ne citer que quelques-uns. Si les plaques sont réputées très résistantes aux efforts en traction, elles sont par contre beaucoup plus sensibles aux charges en compression. En réaction à de fortes sollicitations en compression, leur comportement élastique se caractérise par une soudaine perte de stabilité, appelée le flambage. Les conséquences sur la sécurité de la structure inspectée peuvent évidemment être désastreuses. On conseille de se reporter à ESLAMI et collab. [2018] pour disposer d'un ouvrage récent sur le flambage des poutres, des plaques et des coques.

### 4.6.1 Description du cas test

#### 4.6.1.1 Plaque composite

On considère une plaque rectangulaire posée en appui simple sur ses quatre bords. Ainsi, la rotation suivant l'axe  $\vec{z}$  est bloquée mais les cinq autres degrés de liberté restent autorisés. Comme on peut le voir sur la partie gauche de la Figure 4.22, la compression prend la forme d'une charge linéique  $L_x$  qui est appliquée de part et d'autre de la plaque dans la direction de l'axe  $\vec{x}$ . C'est une configuration assez classique pour modéliser le flambage pouvant survenir au niveau des baies entre les raidisseurs d'un panneau d'aile d'avion. Par souci de simplicité, on supposera que la plaque est carrée ( $a = b$ ) dans la suite de cette section.

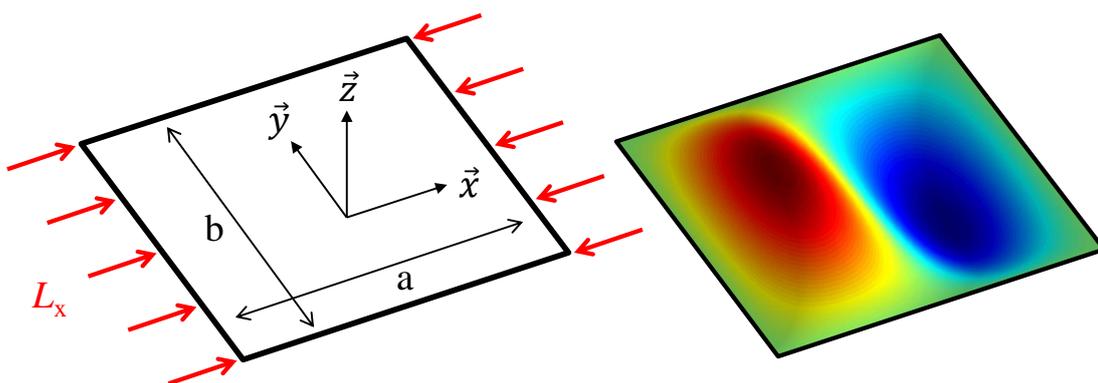


FIGURE 4.22 – Flambage d'une plaque en appui simple sous un effort  $L_x$  de compression uniaxiale.

La plaque étudiée est un stratifié composé de huit couches, aussi appelées des plis. Chaque pli est préalablement découpé dans un rouleau de préimprégné en carbone-époxy. Il s'agit d'un matériau composite dont les fibres de carbone sont agglutinées les unes aux autres avec une résine époxy. Un stratifié est caractérisé par son nombre de plis mais surtout par l'orientation des fibres dans chacun des plis. Notons  $\mathcal{B} := (\vec{x}, \vec{y}, \vec{z})$  la base orthonormée qui est attachée au repère de travail initial. Des axes dits "principaux"  $(\vec{x}_i, \vec{y}_i)$  sont affectés à chaque couche :  $\vec{x}_i$  indique la

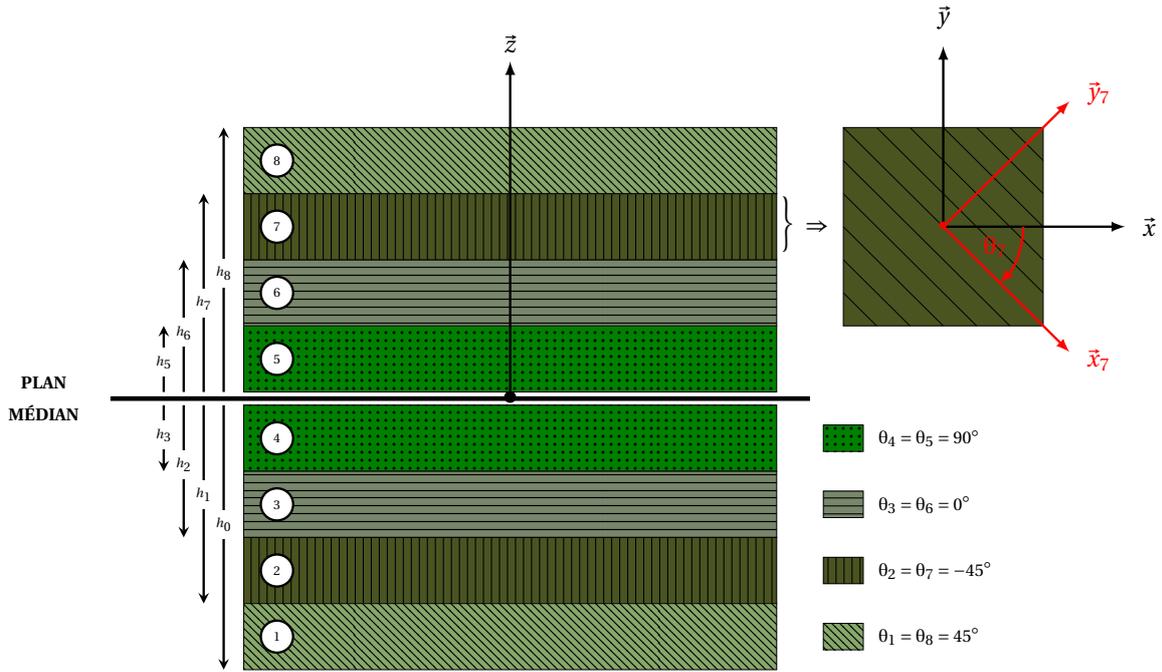


FIGURE 4.23 – Plan de coupe d'un stratifié à huit couches respectant la symétrie "miroir".

direction des fibres au sein de la  $i$ -ème couche tandis que sa perpendiculaire  $\vec{y}_i$  est choisie de manière à ce que  $\mathcal{B}_i := (\vec{x}_i, \vec{y}_i, \vec{z})$  soit orthonormée directe. On note  $\theta_i$  l'angle permettant d'obtenir  $\mathcal{B}_i$  à partir de  $\mathcal{B}$  après une rotation autour de l'axe  $\vec{z}$ . Une illustration est proposée à la Figure 4.23 dans le cas du 7-ème pli. Sachant que le carbone-époxy offre une meilleure tenue dans la direction des fibres, on fabrique un stratifié en superposant des couches orientées dans plusieurs directions préférentielles. Généralement, l'assemblage des couches se fait ensuite par polymérisation ou par collage. Au final, le stratifié offre d'excellentes performances mécaniques tout en gardant un rapport rigidité-densité très avantageux. La plaque étudiée dans cette section est schématisée à la Figure 4.23 où l'on peut observer que l'empilement respecte une symétrie "miroir" qui fait de la séparation entre le 4-ème pli et le 5-ème pli un plan de symétrie du stratifié. Deux couches équidistantes du plan médian sont ainsi orientées dans la même direction.

#### 4.6.1.2 Flambage

Le comportement élastique d'une plaque stratifiée peut être résumé par son équation constitutive. Celle-ci stipule qu'on est capable de calculer les résultantes  $\mathbf{N}$  et les moments  $\mathbf{M}$  en tout point dès lors qu'on connaît les déformations en membrane  $\boldsymbol{\epsilon}$  et les courbures  $\boldsymbol{\kappa}$  en ce point précis et qu'on a accès à la matrice de rigidité  $\mathbf{H}$  du stratifié :

$$\begin{bmatrix} \mathbf{N} \\ \mathbf{M} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{D} \end{bmatrix}}_{\mathbf{H}} \begin{bmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\kappa} \end{bmatrix}. \quad (4.131)$$

On construit  $\mathbf{H} \in \mathbb{R}^{6 \times 6}$  à l'aide de trois blocs :  $\mathbf{A} \in \mathbb{R}^{3 \times 3}$  la matrice de rigidité en membrane,  $\mathbf{B} \in \mathbb{R}^{3 \times 3}$  la matrice de couplage membrane-flexion-torsion et  $\mathbf{D} \in \mathbb{R}^{3 \times 3}$  la matrice de rigidité en flexion. L'obtention de l'équation constitutive (4.131) est détaillée dans BERTHELOT [2010]. Dans la suite, on fait l'hypothèse qu'il n'y a pas de couplage entre la rigidité en membrane et la rigidité en flexion :

$$\mathbf{B} = \mathbf{0}. \quad (4.132)$$

Par ailleurs, de par la symétrie “miroir” régissant l’empilement des couches, le stratifié présente trois plans de symétrie orthogonaux entre eux. On dit alors qu’il est orthotrope et sa matrice  $\mathbf{D}$  de rigidité en flexion ne dépend plus que de quatre coefficients (parmi neuf possibles) :

$$\mathbf{D} = \begin{bmatrix} D_{11} & D_{12} & 0 \\ D_{12} & D_{22} & 0 \\ 0 & 0 & D_{33} \end{bmatrix}. \quad (4.133)$$

Le flambage est un phénomène d’instabilité qui survient lorsqu’une plaque se déforme en flexion pour échapper à une charge en compression. Elle exploite alors un certain mode de déformation, caractérisé par  $m$  demi-ondulations dans la direction de l’axe  $\vec{x}$  et  $n$  demi-ondulations dans la direction de l’axe  $\vec{y}$ . Le mode de flambage  $(m, n) = (2, 1)$  est par exemple représenté sur la partie droite de la Figure 4.22 où la palette des couleurs montre le déplacement suivant l’axe vertical  $\vec{z}$ . Le flambage est un phénomène intrinsèquement structural. Il se produit quel que soit le matériau utilisé, à condition que l’effort appliqué soit d’intensité suffisante. Réciproquement, lorsqu’on choisit un matériau et qu’on fixe la charge appliquée, ce sont les dimensions de la structure qui conditionnent sa tenue au flambage.

Dans TIMOSHENKO et GERE [2009], il est établi que la charge minimale engendrant le mode de flambage  $(m, n)$  peut être calculée en fonction des coefficients de la matrice  $\mathbf{D}$  :

$$N_0(m, n) = \frac{\pi^2}{m^2 a^2} [m^4 D_{11} + 2 m^2 n^2 (D_{12} + 2 D_{33}) + n^4 D_{22}]. \quad (4.134)$$

Cette formule analytique reste valable tant que les hypothèses (4.132) et (4.133) sont vérifiées. En pratique, si les coefficients  $D_{13}$  et  $D_{23}$  ne peuvent plus être considérés comme étant égaux à zéro mais que leurs valeurs restent faibles, la formule (4.134) fournit tout de même une approximation assez précise de la charge critique. Si on souhaite s’affranchir de toute hypothèse, il faut privilégier une résolution semi-analytique par la méthode de Rayleigh-Ritz ou une résolution numérique par la méthode des éléments finis. Des approches de ce type sont adoptées dans les études menées par IKHENAZEN et collab. [2010] ou MAJEED et ABED [2019]. Pour un mode de déformation  $(m, n)$  donné, le facteur de flambage  $\Lambda_{m,n}$  désigne le rapport entre la charge critique et l’effort appliqué :

$$\Lambda_{m,n} = \frac{N_0(m, n)}{L_x}. \quad (4.135)$$

Pour prévenir tous les modes de flambage, il faut s’intéresser au plus petit facteur de flambage car c’est celui qui survient en premier lorsqu’on augmente l’intensité  $L_x$  de l’effort appliqué :

$$\Lambda = \Lambda_{m^*, n^*} = \min_{m, n} \Lambda_{m, n}. \quad (4.136)$$

La paire  $(m^*, n^*)$  définit le mode critique. Expérimentalement, on constate que l’on a toujours  $m^* = 1$  ou  $n^* = 1$  et que l’ondulation se fait le long de la plus grande dimension. Ici, comme  $a = b$ , on fait l’hypothèse que la déformation ondulatoire se produit dans la direction de l’axe  $\vec{x}$ . On distingue alors les deux situations suivantes :

$$\begin{cases} \Lambda > 1 & \Rightarrow \text{Le flambage n'est pas prévu.} \\ \Lambda \leq 1 & \Rightarrow \text{Le flambage se produit.} \end{cases} \quad (4.137)$$

Dans la suite, on notera  $\phi_1$  l’application  $\mathbf{D} \rightarrow \Lambda$  qui compile les étapes (4.134) à (4.136).

#### 4.6.1.3 Calcul analytique de la matrice de rigidité en flexion

La matrice de rigidité en flexion  $\mathbf{D}$  intervenant dans l’équation (4.134) pour le calcul du facteur de flambage  $\Lambda$  n’est pas directement fournie par le fabricant. Toutefois, elle peut être calculée facilement à partir des propriétés du stratifié. Il suffit de connaître les grandeurs caractérisant le

comportement élastique des plis, l'épaisseur des plis et l'orientation des fibres dans les différents plis. Pour le  $i$ -ème pli, la matrice de rigidité réduite  $\mathbf{Q}_i$  exprimée dans la base  $\mathcal{B}_i$  vaut :

$$\mathbf{Q}_i = \mathbf{S}^{-1} \quad \text{avec} \quad \mathbf{S} := \begin{bmatrix} \frac{1}{E_{11}} & -\frac{\nu_{12}}{E_{11}} & 0 \\ -\frac{\nu_{12}}{E_{11}} & \frac{1}{E_{22}} & 0 \\ 0 & 0 & \frac{1}{G_{12}} \end{bmatrix}. \quad (4.138)$$

Les coefficients de la matrice de souplesse  $\mathbf{S}$  dépendent des modules de l'ingénieur que sont :

- $E_{11}$  le module d'Young dans le sens des fibres,
- $E_{22}$  le module d'Young perpendiculairement au sens des fibres,
- $G_{12}$  le module de cisaillement,
- $\nu_{12}$  le coefficient de Poisson.

L'expression  $\mathbf{Q}'_i$  de  $\mathbf{Q}_i$  dans la base initiale  $\mathcal{B}$  s'obtient avec la formule :

$$\mathbf{Q}'_i = [\mathbf{T}_\sigma^{-1}(\theta_i)] \mathbf{Q}_i [\mathbf{T}_\epsilon(\theta_i)], \quad (4.139)$$

qui fait intervenir  $\mathbf{T}_\epsilon(\theta_i)$  la matrice de changement de base dans le champ des déformations et  $\mathbf{T}_\sigma(\theta_i)$  son équivalent dans le champ des contraintes :

$$\mathbf{T}_\epsilon(\theta_i) = \begin{bmatrix} \cos^2(\theta_i) & \sin^2(\theta_i) & \sin(\theta_i) \cos(\theta_i) \\ \sin^2(\theta_i) & \cos^2(\theta_i) & -\sin(\theta_i) \cos(\theta_i) \\ -2 \sin(\theta_i) \cos(\theta_i) & 2 \sin(\theta_i) \cos(\theta_i) & \cos^2(\theta_i) - \sin^2(\theta_i) \end{bmatrix}, \quad (4.140)$$

$$\mathbf{T}_\sigma(\theta_i) = \begin{bmatrix} \cos^2(\theta_i) & \sin^2(\theta_i) & 2 \sin(\theta_i) \cos(\theta_i) \\ \sin^2(\theta_i) & \cos^2(\theta_i) & -2 \sin(\theta_i) \cos(\theta_i) \\ -\sin(\theta_i) \cos(\theta_i) & \sin(\theta_i) \cos(\theta_i) & \cos^2(\theta_i) - \sin^2(\theta_i) \end{bmatrix}.$$

La matrice de rigidité  $\mathbf{D}$  du stratifié dépend des matrices de rigidité réduites  $\mathbf{Q}'_i$  mais aussi de l'épaisseur des plis. On a besoin d'introduire un paramétrage de l'empilement des couches au sein du stratifié. Les côtes des frontières inférieure et supérieure du  $i$ -ème pli sont respectivement notées  $h_{i-1}$  et  $h_i$  sur l'axe vertical  $\bar{z}$ . Comme on peut le voir sur la Figure 4.23, la côte  $h_4$  du plan médian sert d'origine à l'axe et sa valeur est par conséquent fixée à 0. Si on note  $e$  l'épaisseur du préimprégné, la formule  $h_i = e \times (i - 4)$  donne accès à toutes les côtes. Ces valeurs interviennent ensuite dans le calcul de  $\mathbf{D}$  :

$$\mathbf{D} = \frac{1}{3} \sum_{i=1}^8 (h_i^3 - h_{i-1}^3) \mathbf{Q}'_i. \quad (4.141)$$

Finalement, le calcul de  $\mathbf{D}$  à partir des propriétés du stratifié est résumé par la fonction :

$$\phi_0 : \underbrace{\mathbb{R}^4}_{\mathbf{M}} \times \underbrace{\mathbb{R}^8}_{\boldsymbol{\theta}} \times \mathbb{R} \longrightarrow \underbrace{\mathbb{R}^4}_{\mathbf{D}} \\ \left[ E_{11}, E_{22}, G_{12}, \nu_{12} \right] \times \left[ \theta_1, \dots, \theta_8 \right] \times e \longmapsto \left[ D_{11}, D_{22}, D_{33}, D_{12} \right]. \quad (4.142)$$

qui encapsule le paramétrage vertical de la plaque, la construction de la matrice de souplesse  $\mathbf{S}$  à partir des modules  $\mathbf{M}$ , l'expression des matrices de rigidité réduites  $\mathbf{Q}'_i$  des différents plis dans une base commune  $\mathcal{B}$  et le calcul final de la matrice de rigidité  $\mathbf{D}$  du stratifié.

## 4.6.2 Description du problème d'analyse de fiabilité

### 4.6.2.1 Identification et modélisation des incertitudes

On rappelle que la fabrication du stratifié comprend trois étapes : le découpage des plis dans un rouleau de préimprégné en carbone-époxy, la superposition des plis selon les orientations imposées et l'assemblage des plis par polymérisation. En fin de processus, on peut calculer la matrice

de rigidité  $\mathbf{D}$  du stratifié en utilisant le code de calcul  $\phi_0$ . Il prend en entrée un total de 13 paramètres issus de la modélisation macroscopique de la plaque.

Si on dispose de valeurs nominales pour les entrées de  $\phi_0$ , il faut toutefois tenir compte des incertitudes qui les accompagnent. Les valeurs des modules  $E_{11}$ ,  $E_{22}$ ,  $G_{12}$  et  $\nu_{12}$  du matériau composite sont procurées par le fournisseur qui précise des intervalles de confiance. Il est assez habituel de représenter l'erreur commise au cours de la spécification de ces grandeurs par une loi gaussienne. On peut supposer que le vecteur  $\mathbf{M}$  rassemblant les quatre modules prend la même valeur dans toutes les couches. En effet, avant d'être empilés, les plis sont tous découpés dans un même rouleau de préimprégné. On pourrait discuter de la variabilité spatiale de  $\mathbf{M}$  au sein d'un rouleau ou de la variabilité de  $\mathbf{M}$  d'un rouleau à l'autre mais cela dépasse le cadre de ce travail. La phase d'assemblage des plis induit quant à elle une légère déviation de l'orientation des fibres. Il est assez habituel d'utiliser une dispersion uniforme centrée sur la valeur  $\theta_i$  pour modéliser cette source d'erreur. Contrairement aux autres paramètres, l'épaisseur d'un pli n'est pas fournie avec le rouleau de préimprégné. Elle est calculée après polymérisation et à partir de l'épaisseur finale du stratifié. L'incertitude sur l'épaisseur d'un pli est donc considérée comme le résultat des erreurs introduites au cours de la fabrication. Une distribution gaussienne est conventionnellement choisie pour représenter la dérive de ce paramètre. Le recensement des incertitudes conduit au modèle probabiliste suivant :

$$\underbrace{(E_{11} \perp E_{22} \perp G_{12} \perp \nu_{12})}_{\mathbf{M}=[M_i]_{1 \leq i \leq 4}} \perp \underbrace{(\Theta_1 \perp \dots \perp \Theta_8)}_{\Theta=[\theta_i]_{1 \leq i \leq 8}} \perp e \quad \text{avec} \quad \begin{cases} M_i \sim \mathcal{N}(m_i, \sigma_i^2) \\ \Theta_i \sim \mathcal{U}([\theta_i \pm \eta]) \\ e \sim \mathcal{N}(m_e, \sigma_e^2) \end{cases} \quad (4.143)$$

Les incertitudes répertoriées en entrée de  $\phi_0$  se propagent sur la matrice  $\mathbf{D}$ . Le comportement conjoint de ses quatre coefficients  $D_{11}$ ,  $D_{22}$ ,  $D_{33}$  et  $D_{12}$  est alors modélisé par une densité multidimensionnelle  $f_{\mathbf{D}} \in \mathbb{L}^2(\mathbb{R}^4)$  qui est a priori inconnue. La dispersion constatée sur les variables  $\Theta_i$  peut susciter des interrogations quant à la validité du calcul  $\Lambda = \phi_1(\mathbf{D})$  tel qu'on l'a détaillé plus haut. L'erreur commise au moment de positionner le  $i$ -ème pli n'est pas la même que celle commise au moment de positionner son pli symétrique par rapport au plan médian. Les variables  $\Theta_1$  et  $\Theta_8$  peuvent par exemple être modélisées par une même loi uniforme mais elles ne sont jamais égales. La symétrie "miroir" et l'orthotropie du stratifié sont donc remises en question par les incertitudes issues du processus de fabrication. L'hypothèse (4.133) ne tient plus car les coefficients  $D_{13}$  et  $D_{23}$  ne sont plus égaux à zéro. Toutefois, on suppose que leurs valeurs restent faibles et l'expression (4.134) utilisée pour le calcul de la charge critique  $N_0(m, n)$  demeure une approximation raisonnable. Si on poursuit la chaîne de propagation des incertitudes aléatoires qui est représentée à la Figure 4.24, il faut aussi définir la densité  $f_{\Lambda} \in \mathbb{L}^2(\mathbb{R})$  du facteur de flambage  $\Lambda$ . Sachant qu'on se base sur le critère (4.137), la probabilité de défaillance s'écrit :

$$\begin{aligned} P_f &= \mathbb{P}(N_0(m^*, n^*) \leq L_x) = \mathbb{P}(\Lambda \leq 1) = \int_{\mathbb{R}} \mathbb{1}_{\{y \leq 1\}} f_{\Lambda}(y) dy \\ &= \mathbb{P}(\phi(\mathbf{D}) \leq 1) = \int_{\mathbb{R}^4} \mathbb{1}_{\{\phi_1(\mathbf{x}) \leq 1\}} f_{\mathbf{D}}(\mathbf{x}) d\mathbf{x}. \end{aligned} \quad (4.144)$$

On pourrait tout à fait choisir d'examiner la probabilité  $\mathbb{P}(\Lambda \leq T)$  avec  $T \neq 1$ . En effet, ce n'est pas une aberration d'un point de vue physique. Il s'agit simplement de la probabilité que le flambage se produise lorsque l'effort linéique  $L_x$  est multiplié par un nombre (sans dimension) égal à  $T$  :

$$\mathbb{P}(\Lambda \leq T) = \mathbb{P}\left(\frac{N_0(m^*, n^*)}{L_x} \leq T\right) = \mathbb{P}\left(\frac{N_0(m^*, n^*)}{T \times L_x} \leq 1\right). \quad (4.145)$$

#### 4.6.2.2 Choix d'une fonction de performance

Dans la suite, l'application  $\phi_1 : \mathbf{D} \mapsto \Lambda$  devient la fonction de performance  $\phi$ . Ce choix est motivé par le contexte dans lequel on souhaite appliquer l'algorithme d'ASF développé. On veut en

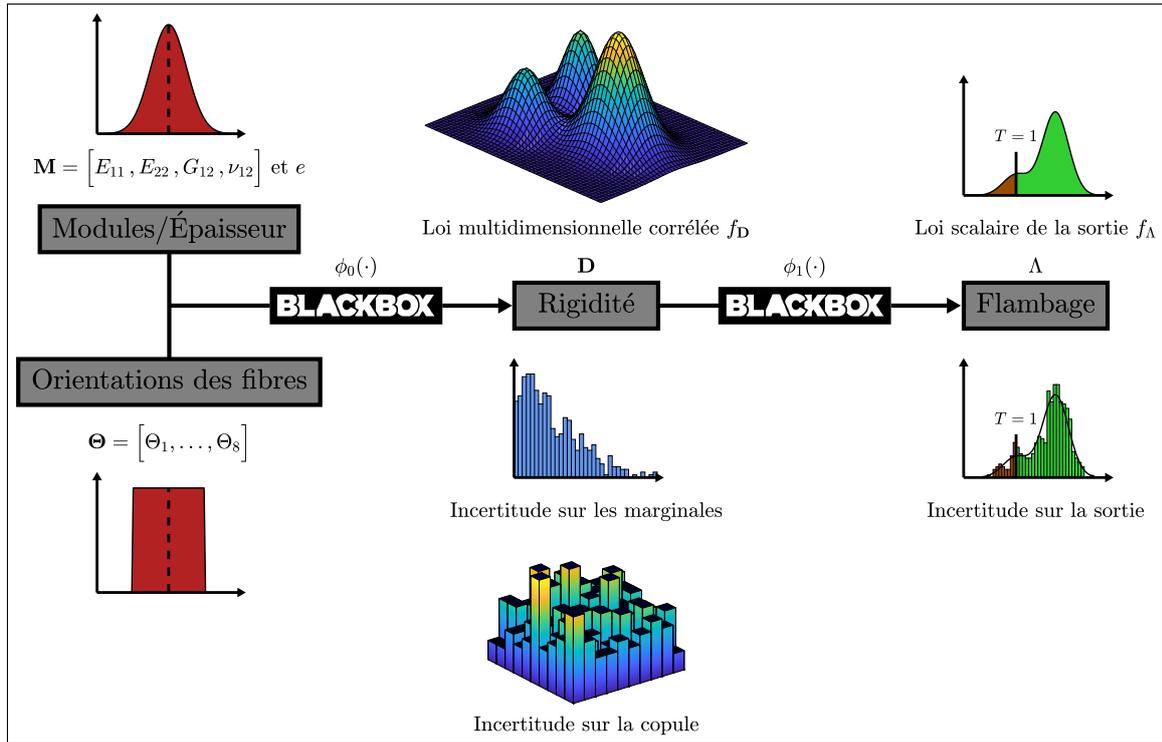


FIGURE 4.24 – Contexte de propagation des incertitudes pour le flambage d’une plaque composite.

effet travailler dans un cadre où la distribution des variables d’entrée n’appartient à aucun modèle de distribution connu. On sait seulement que la mesure  $\mu_D$  est l’image de la mesure  $\mu_M \otimes \mu_\Theta \otimes \mu_e$  par l’application  $\phi_0$ . Le fonctionnement interne de  $\phi_0$  laisse d’ailleurs penser qu’une structure de dépendance non triviale sous-tend le comportement conjoint des coefficients de la matrice **D**. Contrairement à ce qui est fait dans les Sections 4.3.1 et 4.5.2, la construction du modèle d’apprentissage  $\mathcal{M}$  n’est ici pas feinte et elle constitue une étape clé du processus de prédiction des risques. On peut aussi remarquer que l’on dispose d’une routine pour simuler la matrice **D** en initiant des tirages aléatoires à partir des variables **M**, **Θ** et  $e$ . Ce mécanisme offre la possibilité de constituer un échantillon d’apprentissage mais permet aussi d’estimer la probabilité de défaillance théorique  $P_f$  avec une approche de type CMC :

$$\hat{P}_f = \frac{1}{N_s} \sum_{k=1}^{N_s} \mathbb{1}_{\{\Lambda^{(k)} \leq 1\}} \quad \text{avec} \quad \begin{cases} (\mathbf{M}^{(k)}, \Theta^{(k)}, e^{(k)}) \stackrel{\text{i.i.d.}}{\sim} \mu_M \otimes \mu_\Theta \otimes \mu_e \\ \Lambda^{(k)} = [\phi_1 \circ \phi_0](\mathbf{M}^{(k)}, \Theta^{(k)}, e^{(k)}) \end{cases} \quad (4.146)$$

#### 4.6.2.3 Calibration de la distribution conjointe des variables d’entrée

On souhaite fixer les paramètres du modèle probabiliste qui a été esquissé à l’équation (4.143) sans être complètement spécifié. On peut se reporter aux réglages préconisés dans **ROLLET [2007]** pour traiter un exemple similaire. L’incertitude entourant chaque variable  $y$  est décrite avec des bornes symétriquement disposées de part et d’autre d’une valeur nominale. Dans notre situation, étant donné que certaines variables sont supposées gaussiennes, une légère adaptation s’impose. Elle consiste à supposer que les bornes extrêmes sont positionnées à une distance de  $3\sigma$  de la moyenne  $\mu$ . Le Tableau 4.8 présente les paramètres obtenus en procédant de la sorte. Pour cette calibration, on souhaite identifier les entrées qui exercent la plus grande influence sur  $\Lambda$ . Il suffit d’effectuer une AS globale sur l’application :

$$\begin{aligned} \phi_1 \circ \phi_0 : \quad \mathbb{R}^4 \times \mathbb{R}^8 \times \mathbb{R} &\longrightarrow \mathbb{R} \\ (\mathbf{M}, \Theta, e) &\longmapsto \Lambda \end{aligned} \quad (4.147)$$

Entrée	Description (Unité)	Type	Paramètres
$E_{11} = M_1$	Module d'Young des fibres (MPa)	$\mathcal{N}(m_1, \sigma_1^2)$	$m_1 = 1.81 \times 10^5$ $\frac{\sigma_1}{m_1} = 0.01$
$E_{22} = M_2$	Module d'Young de la matrice (MPa)	$\mathcal{N}(m_2, \sigma_2^2)$	$m_2 = 1.03 \times 10^4$ $\frac{\sigma_2}{m_2} = 0.02$
$G_{12} = M_3$	Module de cisaillement (MPa)	$\mathcal{N}(m_3, \sigma_3^2)$	$m_3 = 7.17 \times 10^3$ $\frac{\sigma_3}{m_3} = 0.04$
$\nu_{12} = M_4$	Coefficient de Poisson (1)	$\mathcal{N}(m_4, \sigma_4^2)$	$m_4 = 0.28$ $\frac{\sigma_4}{m_4} = 0.01$
$\Theta_i$	Orientation des fibres dans la $i$ -ème couche (°)	$\mathcal{U}([\theta_i \pm \eta])$	$\eta = 2^\circ$
$e$	Épaisseur (mm)	$\mathcal{N}(m_e, \sigma_e^2)$	$m_e = 1.52$ $\frac{\sigma_e}{m_e} = 0.02$

TABLEAU 4.8 – Modèle probabiliste affecté aux variables macroscopiques décrivant la plaque composite.

Comme les entrées sont mutuellement indépendantes, le calcul d'indices de Sobol est suffisant. Les résultats obtenus après  $N_{\text{rep}} = 10$  estimations des indices du premier ordre sont disponibles au premier étage du Tableau 4.9 où l'on peut voir que l'incertitude sur l'épaisseur  $e$  permet d'expliquer la quasi-totalité de la variabilité en sortie des codes de calcul  $\phi_0$  et  $\phi_1$  étant donné que  $\mathbb{V}(\mathbb{E}[\Lambda | e]) = 0.97 \times \mathbb{V}(\Lambda)$ . On remarque que la part de variabilité engendrée par la dispersion de l'orientation des fibres est négligeable. Pour cette raison, on ne se hasarde pas à essayer de quantifier l'écart-type des estimateurs  $\hat{S}_{\Theta_i}$  associés. Le rôle insignifiant des angles  $\Theta_i$  nous conforte dans l'idée qu'il est raisonnable de considérer que le stratifié est orthotrope et ce, même si le processus de fabrication engendre une légère dissymétrie dans l'empilement des couches.

On souhaite aussi avoir une idée de la loi  $\mu_{\mathbf{D}}$  induite par la première étape de propagation des incertitudes. Pour cela, on simule  $N = 500$  échantillons de la matrice  $\mathbf{D}$  :

$$\left[ \mathbf{D}^{(i)} \right]_{1 \leq i \leq N} = \left[ D_{11}^{(i)}, D_{22}^{(i)}, D_{33}^{(i)}, D_{12}^{(i)} \right]_{1 \leq i \leq N} =: \left[ x_j^{(i)} \right]_{\substack{1 \leq i \leq N \\ 1 \leq j \leq d}} = \mathbf{x}_{\text{obs}} \in \mathbb{R}^{N \times 4}. \quad (4.148)$$

Un graphique de type “*pair plot*” est proposé à la Figure 4.25a et rassemble plusieurs outils de visualisation des lois unidimensionnelles et bidimensionnelles. Les cellules disposées sur la diagonale sont réservées aux histogrammes et aux estimations KDE des densités marginales. Les cellules de la partie supérieure accueillent le tracé des nuages de paires de points tandis que les cellules de la partie inférieure représentent les lignes d'isovaleurs des densités construites à partir de ces mêmes points après utilisation d'une méthode bidimensionnelle d'estimation par noyaux. Visuellement, on constate une très forte corrélation des coefficients de la matrice  $\mathbf{D}$ . Cette tendance est confirmée par les résultats des calculs empiriques :

$$\hat{\rho} = \begin{bmatrix} 1 & 0.95 & 0.74 & 0.97 \\ 0.95 & 1 & 0.87 & 0.98 \\ 0.74 & 0.87 & 1 & 0.88 \\ 0.97 & 0.98 & 0.88 & 1 \end{bmatrix}. \quad (4.149)$$

Chercher à apprendre une telle distribution  $\mu_{\mathbf{D}}$  à partir d'un échantillon  $\mathbf{x}_{\text{obs}}$  peut être périlleux car la copule sous-jacente est proche du cas limite de comonotonicité. Pour contourner ce problème, une solution peut consister à neutraliser, ou du moins à atténuer l'influence exercée par l'épaisseur  $e$  dont la valeur semble piloter celle de tous les coefficients de  $\mathbf{D}$ . Désormais, on choisit de fixer à 0.2% le coefficient de variation de la loi gaussienne  $\mathcal{N}(m_e, \sigma_e^2)$ . Disposer d'un tel niveau de connaissance sur l'épaisseur n'est pas inatteignable en pratique. En effet, si l'assemblage des plis est réalisé avec une presse plutôt qu'avec un autoclave, il est vrai que l'on peut disposer de mesures plus précises de l'épaisseur totale.

$$\hat{\rho} = \begin{bmatrix} 1 & 0.45 & -0.59 & 0.58 \\ 0.45 & 1 & 0.20 & 0.73 \\ -0.59 & 0.20 & 1 & 0.31 \\ 0.58 & 0.73 & 0.31 & 1 \end{bmatrix}. \quad (4.150)$$

		Indices de Sobol								
Loi de l'épaisseur $e$	$\frac{\sigma_e}{m_e} = 2\%$	$\hat{S}_{E_{11}}$	$\hat{S}_{E_{22}}$	$\hat{S}_{G_{12}}$	$\hat{S}_{V_{12}}$	$\hat{S}_e$				
		moyenne :	2.24 %	0.04 %	0.03 %	0.00 %	97.07 %			
		écart-type :	0.06 %	×	×	×	1.57 %			
		rang :	(2)	(3)	(4)	(7)	(1)			
	$\hat{S}_{\Theta_1}$	$\hat{S}_{\Theta_2}$	$\hat{S}_{\Theta_3}$	$\hat{S}_{\Theta_4}$	$\hat{S}_{\Theta_5}$	$\hat{S}_{\Theta_6}$	$\hat{S}_{\Theta_7}$	$\hat{S}_{\Theta_8}$		
	moyenne :	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %
	écart-type :	×	×	×	×	×	×	×	×	×
	rang :	(5)	(9)	(11)	(13)	(12)	(10)	(8)	(6)	
$\frac{\sigma_e}{m_e} = 0.2\%$	$\hat{S}_{E_{11}}$	$\hat{S}_{E_{22}}$	$\hat{S}_{G_{12}}$	$\hat{S}_{V_{12}}$	$\hat{S}_e$					
		moyenne :	68.13 %	1.26 %	0.96 %	0.02 %	29.62 %			
		écart-type :	1.33 %	0.02 %	0.02 %	×	0.75 %			
		rang :	(1)	(3)	(4)	(7)	(2)			
	$\hat{S}_{\Theta_1}$	$\hat{S}_{\Theta_2}$	$\hat{S}_{\Theta_3}$	$\hat{S}_{\Theta_4}$	$\hat{S}_{\Theta_5}$	$\hat{S}_{\Theta_6}$	$\hat{S}_{\Theta_7}$	$\hat{S}_{\Theta_8}$		
	moyenne :	0.03 %	0.01 %	0.00 %	0.00 %	0.00 %	0.00 %	0.01 %	0.03 %	
	écart-type :	×	×	×	×	×	×	×	×	
	rang :	(5)	(9)	(11)	(13)	(12)	(10)	(8)	(6)	

TABLEAU 4.9 – Analyse de la sensibilité du facteur de flambage  $\Lambda$  aux entrées  $\mathbf{M}$ ,  $\Theta$  et  $e$ .

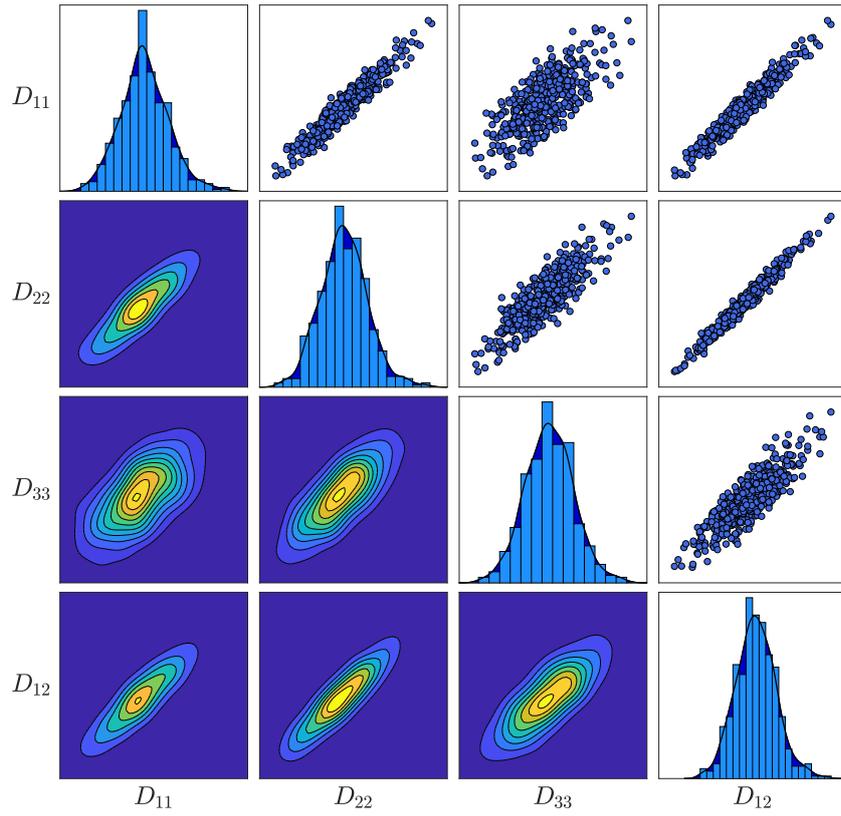
Par ailleurs, l'évolution du “*pair plot*” entre les Figures 4.25a et 4.25b est conforme aux attentes. Une plus grande diversité de motifs de dépendance transparait lorsqu'on étudie les échantillons des différentes paires de variables. Lorsqu'on réitère l'AS globale sur l'application (4.147), le second étage du Tableau 4.9 montre que l'influence de l'épaisseur  $e$  a été divisée par trois. C'est désormais le module d'Young  $E_{11}$  qui devient prépondérant dans l'explication de  $\mathbb{V}(\Lambda)$ . On peut toutefois regretter que l'action menée sur la variable initialement dominante n'ait pas permis d'activer un lot plus fourni de variables d'entrée. Cet ajustement du rapport  $\sigma_e/m_e$  est entériné pour la suite de cette section.

#### 4.6.2.4 Réglage du niveau de criticité du risque de défaillance

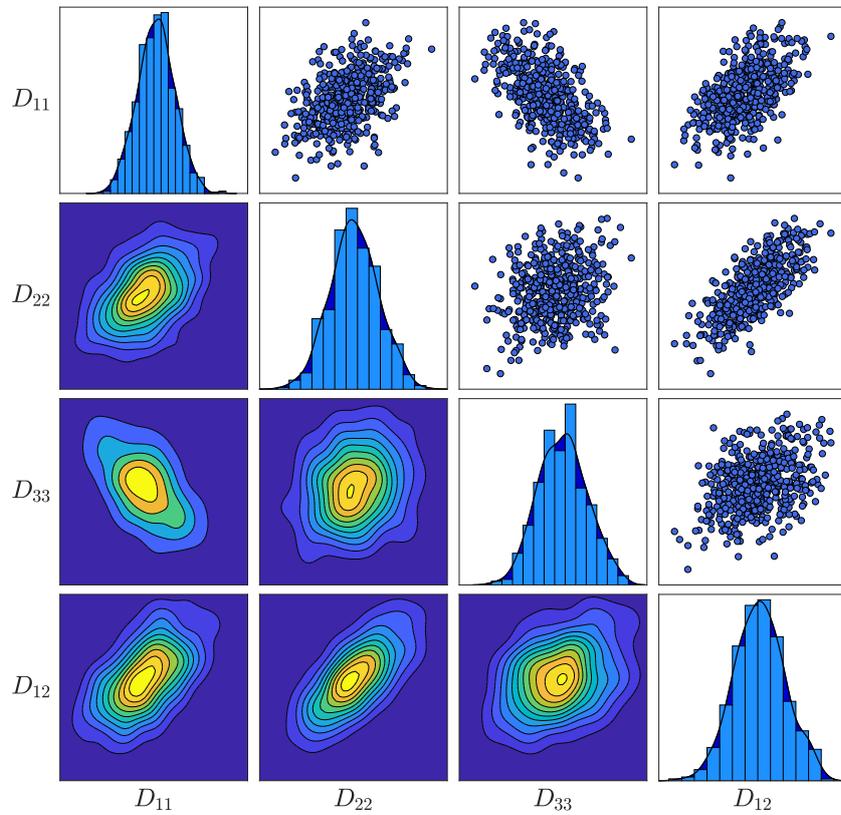
On utilise une plaque de côté  $a = 1$  m. On peut commencer par supposer que la charge linéique vaut  $L_0 = 1$  N/mm. À partir de  $N_{CMC} = 10^6$  échantillons simulés  $\Lambda^{(k)}$ , on obtient  $\mathbb{E}[\Lambda] \approx 499$ . Il n'y a alors aucun risque de flambage. Pour ne pas tâtonner dans le réglage de  $L_x$ , il suffit de choisir un seuil  $T$  pour lequel la quantité  $\mathbb{P}(\Lambda \leq T)$  atteint le niveau de criticité requis. Avec  $T = 485$ , on obtient par exemple  $\mathbb{P}(\Lambda \leq T) \approx 4.6 \times 10^{-3}$ . En vertu de l'équation (4.145), cela signifie qu'appliquer un effort de compression  $L_x = T \times L_0 = 485$  N/mm engendre un risque théorique de flambage de  $\mathbb{P}(\Lambda \leq 1) \approx 4.6 \times 10^{-3}$ .

#### 4.6.2.5 Sélection d'un modèle d'apprentissage

On veut désormais sélectionner un modèle d'apprentissage  $\mathcal{M}$  susceptible de nous permettre d'estimer avec précision la loi  $\mu_{\mathbf{D}}$  à partir d'un échantillon  $\mathbf{x}_{\text{obs}}$ . Ici, et contrairement à ce qui est fait dans les cas tests analytiques étudiés précédemment, la méconnaissance du modèle sous-jacent ne se cantonne pas aux lois marginales et s'étend également à la copule. Les histogrammes de la Figure 4.25b suggèrent que les densités marginales sont unimodales mais aucun d'eux ne laisse penser que l'utilisation de la famille gaussienne soit suffisante pour tenir compte des asymétries et de la diversité des vitesses de décroissance dans les différentes queues de distribution. On décide donc de recourir à une estimation KDE des densités marginales. De même, on préfère ne pas faire



(a) Lois marginales et bidimensionnelles lorsque  $\frac{\sigma_e}{m_e} = 2\%$ .



(b) Lois marginales et bidimensionnelles lorsque  $\frac{\sigma_e}{m_e} = 0.2\%$ .

FIGURE 4.25 – Visualisation de la sensibilité de la loi  $\mu_{\mathbf{D}}$  à la modélisation de l'épaisseur  $e$ .

d'hypothèse quant à une éventuelle gaussianité de la copule sous-jacente. Pour disposer d'une meilleure capacité de description de la structure de dépendance, une bonne pratique consiste à utiliser un modèle de copule adapté à la plus grande dimension. Les copules R-vignes présentées à la Section 1.5.4.1 en sont un exemple. Elles permettent de prendre en compte la polymorphie des liens de dépendance pouvant exister au sein d'un lot de plusieurs variables. En ce sens, elles apportent une réponse aux restrictions constatées chez les copules elliptiques et archimédiennes. La manipulation de ces modèles rend bien plus complexe le processus de transformation de la variabilité d'échantillonnage en variabilité d'apprentissage. La compréhension de la mesure  $\mu_{\hat{c}}$  induite par  $(\mu_V)^{\otimes N}$  ne se limite plus à la maîtrise de l'incertitude propagée par l'estimateur  $\hat{\theta}$  des paramètres d'une famille paramétrique  $\mathcal{F}$  fixée. À chaque nouvel échantillon  $\mathbf{v}_{\text{obs}}$  d'observations de la copule, la procédure d'inférence statistique de [DISSMANN et collab. \[2013\]](#) associe une nouvelle séquence d'arbres  $\hat{\mathbf{T}}_{\text{vg}}$ , un nouvel assortiment de familles de copules bidimensionnelles  $\hat{\mathbf{F}}_{\text{vg}}$  et un nouvelle collection de paramètres  $\hat{\theta}_{\text{vg}}$ . La densité conjointe résultant de la mise en œuvre du modèle d'apprentissage  $\mathcal{M}$  est ainsi constituée des composantes fonctionnelles suivantes :

$$\hat{f}_{X_j}(\cdot | \mathcal{M}_j, \mathbf{Z}_{\text{obs}}^{[\cdot, j]}) = \hat{f}_{\text{KDE}}(x_j | \mathbf{Z}_{\text{obs}}^{[\cdot, j]}, \hat{h}_j) \quad \text{et} \quad \hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{V}_{\text{obs}}) = \hat{c}_{\text{vg}}(\cdot | \hat{\mathbf{T}}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}}, \hat{\theta}_{\text{vg}}). \quad (4.151)$$

### 4.6.3 Implémentation de la méthodologie développée

#### 4.6.3.1 Examen des résultats fournis par l'algorithme d'ASF

L'algorithme d'ASF développé à la Section 4.4 est utilisé pour identifier la composante fonctionnelle de la densité conjointe  $\hat{f}_{\mathbf{D}}(\cdot | \mathcal{M}, \mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}})$  dont l'apprentissage engendre le plus de variabilité sur l'estimation du risque de flambage du stratifié orthotrope. On effectue  $N_{\text{rep}} = 10$  estimations des indices de Sobol du premier ordre à l'aide du schéma *Correlation 2* limité à  $M = 10^3$  itérations. La stratégie retenue pour supporter la charge de calcul consiste à entreprendre  $N_e = 5$  explorations préliminaires du domaine de défaillance avec l'algorithme NAIS puis à utiliser la technique d'échantillonnage préférentiel multiple décrite à la Section 4.4.3.2 pour construire une densité auxiliaire  $\hat{h}_{\text{MIS}}$ . À chaque exploration, NAIS est lancé à partir de  $N_s = 3 \times 10^3$  particules et converge en  $N_i = 3$  itérations, ce qui porte le budget de simulation à  $N_{\text{tot}} = N_e \times (N_i \times N_s) = 45 \times 10^3$  appels à la fonction de performance  $\phi = \phi_1$ .

Les résultats sont présentés au premier étage du Tableau 4.10 où l'on préserve la notation habituelle des indices de Sobol, le vecteur  $\mathbf{X}$  désignant ainsi  $\mathbf{D} = [D_{11}, D_{22}, D_{33}, D_{12}]$ . On peut voir que l'indice  $\hat{S}_{\text{cop}}$  occupe la première place du classement. La variabilité de  $\hat{P}_f(\mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}})$  s'explique donc avant tout à travers celle de  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{V}_{\text{obs}})$ . Viennent ensuite, dans l'ordre, les indices  $\hat{S}_4$ ,  $\hat{S}_1$ ,  $\hat{S}_2$  et  $\hat{S}_3$ . Pour ce cas test, on constate que l'estimation d'une loi marginale  $f_{X_j}$  représente une part allant du quinzième au quart de la variabilité d'apprentissage totale. Par ailleurs, les faibles valeurs des écarts-types montrent à nouveau que la construction de la densité auxiliaire  $\hat{h}_{\text{MIS}}$ , reconduite à chaque nouvelle estimation des indices de Sobol, introduit une incertitude dont l'amplitude est insignifiante et ne détériore pas la robustesse des estimateurs  $\hat{S}_j$  et  $\hat{S}_{\text{cop}}$ .

Comme pour les cas tests analytiques, on souhaite vérifier que les indices de Sobol produits par l'ASF rendent correctement compte de la propagation des incertitudes épistémiques. À nouveau, la démarche de validation consiste à s'assurer qu'il est possible d'amoindrir l'influence de l'une des composantes fonctionnelles par acquisition puis assimilation ciblée de données supplémentaires. Les deuxième et troisième étages du Tableau 4.10 détaillent les résultats produits par l'algorithme d'ASF après avoir procédé à l'assimilation de nouvelles observations pour  $X_4$  et  $\mathbf{U}$ . Dans chacun des cas, la stratégie de calcul reste la même et s'appuie sur la construction d'une densité auxiliaire  $\hat{h}_{\text{MIS}}$  par échantillonnage préférentiel multiple en combinant l'information de  $N_e = 5$  explorations préalables réalisées avec NAIS. Au second étage, on constate que l'ajout de  $N' = 4 \times 10^3$  données supplémentaires pour la variable  $X_4$  conduit  $\hat{S}_4$  de la seconde à la dernière place du classement des indices de Sobol. Un phénomène analogue est observable au troisième étage où l'indice  $\hat{S}_{\text{cop}}$  chute de plusieurs positions. Dans les deux cas, l'évolution du classement

des indices est cohérente avec l'évolution des incertitudes que les indices sont supposés quantifier. On peut toutefois noter que quatre fois moins de données ont été utilisées au moment de réduire l'influence de  $\hat{c}_U$ . Ce n'est pas étonnant car c'est précisément l'estimation de la copule qui a été identifiée comme la tâche d'apprentissage rendue la plus instable par le contexte "small data". Si on compare les résultats de scénarios d'assimilation ciblant différentes composantes, il faut en effet s'attendre à ce que la composante générant le plus de variabilité d'apprentissage soit celle qui tire le plus grand bénéfice de l'information supplémentaire ainsi apportée.

ASF exécutée avec $\mathcal{B}$ :		$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_4$	$\hat{S}_{\text{cop}}$
$\mathbf{x}_{\text{obs}}$	moyenne :	20.95 %	15.08 %	7.84 %	23.65 %	38.51 %
	écart-type :	1.54 %	1.57 %	1.02 %	3.14 %	3.10 %
	rang :	(3)	(4)	(5)	(2)	(1)
ASF exécutée avec $\mathcal{B}_4$ :		$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_4$	$\hat{S}_{\text{cop}}$
$\mathbf{x}_{\text{obs}} \cup \mathbf{x}_{\text{extra}}^{[4]}$ $N' = 4 \times 10^3$	moyenne :	23.37 %	17.45 %	9.12 %	11.35 %	48.23 %
	écart-type :	2.53 %	1.68 %	0.64 %	1.83 %	3.66 %
	rang :	(2)	(3)	(5)	(4)	(1)
ASF exécutée avec $\mathcal{B}_{\text{cop}}$ :		$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_4$	$\hat{S}_{\text{cop}}$
$\mathbf{x}_{\text{obs}} \cup \mathbf{x}_{\text{extra}}$ $N' = 10^3$	moyenne :	29.04 %	20.32 %	11.10 %	35.96 %	14.08 %
	écart-type :	2.84 %	4.36 %	2.68 %	5.89 %	1.64 %
	rang :	(2)	(3)	(5)	(1)	(4)

TABLEAU 4.10 – Résultats de l'ASF initiale et des scénarios d'assimilation de données.

#### 4.6.3.2 Bilan

La méthodologie d'ASF mise en place dans les Sections 4.4 et 4.5 a été appliquée dans un cadre assujéti à la réalité du terrain. La variabilité affectant la prédiction du risque de flambage d'un stratifié orthotrope a été examinée à l'aune de l'incertitude entourant la spécification de la distribution des coefficients de la matrice de rigidité. On s'est placé dans un contexte où l'état de connaissance de la loi d'entrée se limite véritablement à la possession d'un échantillon d'apprentissage. Il a été vérifié que l'algorithme d'ASF permet une estimation robuste des indices de sensibilité et que ces derniers proposent une décomposition rigoureuse de la variabilité d'apprentissage entre les différentes entités fonctionnelles.

Si la pertinence des résultats produits par l'algorithme d'ASF n'est pas à remettre en question, on peut toutefois souligner que le cas test proposé reste très simple de par la faible dimension du problème et la quasi-gratuité d'un appel à  $\phi$ . Les expressions (4.134) à (4.136) suggèrent d'ailleurs que le comportement de  $\phi$  s'éloigne finalement assez peu de celui d'une fonction linéaire. On a également remarqué que la loi des entrées n'est en fait que le résultat du mélange des incertitudes gaussiennes portées par le module d'Young  $E_{11}$  et l'épaisseur  $e$  lorsqu'on les propage à travers un code de calcul préliminaire. Il est alors assez légitime de se demander s'il ne suffirait pas de modéliser la distribution des coefficients de la matrice de rigidité par une loi gaussienne multi-dimensionnelle. Expérimentalement, le niveau de précision atteint sur la prédiction du risque de flambage s'avère tout aussi bon. Ceci souligne la trivialité du problème d'inférence statistique.

La façon dont le problème d'analyse de fiabilité est posé peut aussi susciter certaines critiques. L'étude de la propagation des incertitudes selon un schéma  $(\mathbf{M}, \boldsymbol{\Theta}, e) \mapsto \Lambda$  correspond à un contexte classique en mécanique des structures composites. En revanche, s'intéresser à un schéma de propagation de type  $\mathbf{D} \mapsto \Lambda$  n'a pas de réalité pratique. La matrice de rigidité ne s'inscrit pas dans un processus de mise en valeur de la donnée. En effet, ses coefficients ne peuvent pas être mesurés directement. Ils se déduisent des caractéristiques du stratifié et jouent souvent le rôle de quantités intermédiaires dans le calcul d'autres grandeurs.

## 4.7 Conclusion

La méthodologie proposée dans ce chapitre doit être vue comme un outil d'aide à la décision applicable à tout contexte d'analyse de fiabilité où l'état de connaissance de la distribution des entrées se limite à la possession d'un jeu de données de petite taille. La propagation des incertitudes épistémiques se fait à travers un algorithme d'estimation de la densité conjointe des entrées puis à travers un algorithme d'estimation de la probabilité d'un évènement rare. Une procédure d'ASF a été développée pour identifier la composante de la description copule-marginales qui engendre le plus de variabilité sur la prédiction du risque de défaillance du fait de l'incertitude entourant sa modélisation et son estimation. L'algorithme proposé s'appuie sur trois rouages principaux :

1. Grâce à un mécanisme de ré-échantillonnage, on parvient à simuler des réalisations indépendantes de chacune des composantes de la densité estimée.
2. Des stratégies telles que l'astuce "*inverse-reverse*" ou l'échantillonnage préférentiel multiple permettent de s'accommoder de la charge de calcul inhérente à toute ambition d'ASF.
3. Les indices de Sobol sont appliqués dans un cadre fonctionnel qui dépasse leur champ d'utilisation le plus traditionnel.

La méthodologie proposée a pu être validée sur plusieurs cas tests analytiques. Les indices calculés s'avèrent être caractéristiques du modèle d'apprentissage  $\mathcal{M}$  mais indépendants de l'algorithme de fiabilité  $\mathcal{R}_w$  utilisé à condition que celui-ci soit robuste et non-biaisé. Par ailleurs, le caractère "*data-driven*" de cette procédure reste limité au sens où les résultats d'une ASF conduite à partir d'un échantillon fixé peuvent être extrapolés à tout échantillon de la même loi sous-jacente.

Les différents scénarios d'assimilation de données supplémentaires ont donné des résultats cohérents qui mettent en lumière la possible réduction de la part d'apprentissage véhiculée par la composante fonctionnelle dominante. Sur des cas tests analytiques, on a pu constater qu'il n'est pas inhabituel qu'une loi marginale soit l'entité la plus importante à cibler dans le but de réduire la variabilité d'apprentissage. Toutefois, avec la montée en dimension, il est prévisible que la copule devienne petit à petit l'entité dominante. La seule manière de neutraliser son influence consiste alors à collecter des données supplémentaires pour toutes les variables. Une telle action aurait pu être menée sans même attendre la réponse de l'ASF et ne s'inscrit plus dans l'idée d'une acquisition ciblée, et donc moins coûteuse, de nouvelles observations des variables du problème. Dans le Chapitre 5, on cherche donc à subdiviser la copule en plusieurs sous-entités fonctionnelles et on étudie la possibilité d'un raffinement de l'état de connaissance de ces sous-entités.

## Références

- AKAIKE, H. 1974, «A new look at the statistical model identification», *IEEE transactions on automatic control*, vol. 19, n° 6, p. 716–723.
- BERTHELOT, J.-M. 2010, «Mécanique des matériaux et structures composites», *Institut Supérieur des Matériaux et Mécaniques Avancés*.
- BOGACHEV, V. I. 2007, *Measure theory*, vol. 1, Springer Science & Business Media.
- CAUCHY, A. L. B. 1822, *Recherches sur l'équilibre et le mouvement intérieur des corps solides ou fluides, élastiques ou non élastiques*.
- CHATTERJEE, S. 2020, «A new coefficient of correlation», *Journal of the American Statistical Association*, p. 1–21.
- DEHEUVELS, P. 1979, «La fonction de dépendance empirique et ses propriétés. un test non paramétrique d'indépendance», *Bulletins de l'Académie Royale de Belgique*, vol. 65, n° 1, p. 274–292.

- DISSMANN, J., E. C. BRECHMANN, C. CZADO et D. KUROWICKA. 2013, «Selecting and estimating regular vine copulae and application to financial returns», *Computational Statistics & Data Analysis*, vol. 59, p. 52–69.
- EFRON, B. 1982, *The jackknife, the bootstrap and other resampling plans*, SIAM.
- ESLAMI, M. R., J. ESLAMI et JACOBS. 2018, *Buckling and postbuckling of beams, plates, and shells*, Springer.
- FRUTH, J., O. ROUSTANT et S. KUHN. 2015, «Sequential designs for sensitivity analysis of functional inputs in computer experiments», *Reliability Engineering & System Safety*, vol. 134, p. 260–267.
- GAMBOA, F., P. GREMAUD, T. KLEIN et A. LAGNOUX. 2020, «Global sensitivity analysis : a new generation of mighty estimators based on rank statistics», *arXiv preprint arXiv :2003.01772*.
- GAMBOA, F., A. JANON, T. KLEIN, A. LAGNOUX et collab.. 2014, «Sensitivity analysis for multidimensional and functional outputs», *Electronic Journal of Statistics*, vol. 8, n° 1, p. 575–603.
- GENEST, C., K. GHOUDI et L.-P. RIVEST. 1995, «A semiparametric estimation procedure of dependence parameters in multivariate families of distributions», *Biometrika*, vol. 82, n° 3, p. 543–552.
- GIVENS, G. H. et A. E. RAFTERY. 1996, «Local adaptive importance sampling for multivariate densities with strong nonlinear relationships», *Journal of the American Statistical Association*, vol. 91, n° 433, p. 132–141.
- GOOD, P. I. 2006, *Resampling methods*, Springer.
- GRUBER, L., C. CZADO et collab.. 2015, «Sequential Bayesian model selection of regular vine copulas», *Bayesian Analysis*, vol. 10, n° 4, p. 937–963.
- HARBITZ, A. 1986, «An efficient sampling method for probability of failure calculation», *Structural safety*, vol. 3, n° 2, p. 109–115.
- IKHENAZEN, G., M. SAIDANI et A. CHELGHOU. 2010, «Finite element analysis of linear plates buckling under in-plane patch loading», *Journal of Constructional Steel Research*, vol. 66, n° 8-9, p. 1112–1117.
- IOOSS, B. et M. RIBATET. 2009, «Global sensitivity analysis of computer models with functional inputs», *Reliability Engineering & System Safety*, vol. 94, n° 7, p. 1194–1204.
- JANON, A., T. KLEIN, A. LAGNOUX, M. NODET et C. PRIEUR. 2014, «Asymptotic normality and efficiency of two Sobol index estimators», *ESAIM : Probability and Statistics*, vol. 18, p. 342–364.
- JOE, H. et J. J. XU. 1996, «The estimation method of inference functions for margins for multivariate models», cahier de recherche, Department of Statistics, University of British Columbia.
- KANG, Y.-J., Y. NOH et O.-K. LIM. 2019, «Integrated statistical modeling method : part I-statistical simulations for symmetric distributions», *Structural and Multidisciplinary Optimization*, vol. 60, n° 5, p. 1719–1740.
- LEMAÎTRE, P., E. SERGIENKO, A. ARNAUD, N. BOUSQUET, F. GAMBOA et B. IOOSS. 2015, «Density modification-based reliability sensitivity analysis», *Journal of Statistical Computation and Simulation*, vol. 85, n° 6, p. 1200–1223.
- MAJEED, W. I. et Z. A. K. ABED. 2019, «Buckling analysis of laminated composite plate with different boundary conditions using modified Fourier series», *Journal of Engineering*, vol. 25, n° 8, p. 1–18.

- MCLACHLAN, G. J. et D. PEEL. 2004, *Finite mixture models*, John Wiley & Sons.
- MIN, A. et C. CZADO. 2010, «Bayesian inference for multivariate copulas using pair-copula constructions», *Journal of Financial Econometrics*, vol. 8, n° 4, p. 511–546.
- MIN, A. et C. CZADO. 2011, «Bayesian model selection for D-vine pair-copula constructions», *Canadian Journal of Statistics*, vol. 39, n° 2, p. 239–258.
- MONTGOMERY, D. C., E. A. PECK et G. G. VINING. 2012, *Introduction to linear regression analysis*, vol. 821, John Wiley & Sons.
- MORIO, J. 2011, «Influence of input PDF parameters of a model on a failure probability estimation», *Simulation Modelling Practice and Theory*, vol. 19, n° 10, p. 2244–2255.
- MORIO, J. 2012, «Extreme quantile estimation with nonparametric adaptive importance sampling», *Simulation Modelling Practice and Theory*, vol. 27, p. 76–89.
- NATAF, A. 1962, «Détermination des distributions dont les marges sont données», *Comptes Rendus de l'Académie des Sciences*, vol. 225, p. 42–43.
- NAVIER, C. 1823, «Mémoire sur les lois du mouvement des fluides», *Mémoires de l'Académie Royale des Sciences de l'Institut de France*, vol. 6, n° 1823, p. 389–440.
- OWEN, A. B. 2013, «Better estimation of small Sobol'sensitivity indices», *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, vol. 23, n° 2, p. 1–17.
- POISSON, S. D. 1838, *Traité de mécanique*, Société belge de librairie.
- ROLLET, Y. 2007, *Vers une maîtrise des incertitudes en calculs des structures composites*, thèse de doctorat, École Polytechnique.
- ROMBERG, W. 1955, «Vereinfachte numerische integration», *Norske Vid. Selsk. Forh.*, vol. 28, p. 30–36.
- RUBINSTEIN, R. Y. et D. P. KROESE. 2013, *The cross-entropy method : a unified approach to combinatorial optimization, Monte Carlo simulation and machine learning*, Springer Science & Business Media.
- SANKARARAMAN, S. et S. MAHADEVAN. 2013, «Distribution type uncertainty due to sparse and imprecise data», *Mechanical Systems and Signal Processing*, vol. 37, n° 1-2, p. 182–198.
- SHAO, J. et D. TU. 2012, *The jackknife and bootstrap*, Springer Science & Business Media.
- SHINOZUKA, M. 1983, «Basic analysis of structural safety», *Journal of Structural Engineering*, vol. 109, n° 3, p. 721–740.
- SILVERMAN, B. W. 1986, *Density estimation for statistics and data analysis*, vol. 26, CRC press.
- TANG, X.-S., D.-Q. LI, C.-B. ZHOU et L.-M. ZHANG. 2013, «Bivariate distribution models using copulas for reliability analysis», *Proceedings of the Institution of Mechanical Engineers, Part O : Journal of Risk and Reliability*, vol. 227, n° 5, p. 499–512.
- TIMOSHENKO, S. P. et J. M. GERE. 2009, *Theory of elastic stability*, Courier Corporation.
- VEACH, E. 1997, *Robust Monte Carlo methods for light transport simulation*, thèse de doctorat, Stanford University.
- VEACH, E. et L. GUIBAS. 1995a, «Bidirectional estimators for light transport», dans *Photorealistic Rendering Techniques*, Springer, p. 145–167.

VEACH, E. et L. J. GUIBAS. 1995b, «Optimally combining sampling techniques for Monte Carlo rendering», dans *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, p. 419–428.

WOODRUFF, B. W. et A. H. MOORE. 1988, «Application of goodness-of-fit tests in reliability», *Handbook of Statistics*, vol. 7, p. 113–120.



## Chapitre 5

# Analyse de la variabilité d'un risque de défaillance suite à l'apprentissage de la structure de corrélation des entrées

### Sommaire

---

<b>5.1 Introduction</b> . . . . .	<b>218</b>
<b>5.2 Formulation d'un problème d'ASF</b> . . . . .	<b>220</b>
5.2.1 Étude des difficultés inhérentes à l'usage des copules bidimensionnelles . . . . .	220
5.2.2 Transformation du problème d'ASF initial . . . . .	225
<b>5.3 Développement d'une procédure d'ASF pour le sous-problème</b> . . . . .	<b>233</b>
5.3.1 Prise en compte de la dépendance entre les variables d'entrée de l'ASF . . . . .	233
5.3.2 Construction et estimation d'un indicateur de stochasticité . . . . .	240
<b>5.4 Développement d'une stratégie d'assimilation de nouvelles données</b> . . . . .	<b>249</b>
5.4.1 Adaptation du ré-échantillonnage aux données supplémentaires . . . . .	250
5.4.2 Adaptation de l'apprentissage aux données supplémentaires . . . . .	251
<b>5.5 Validation des deux procédures d'ASF</b> . . . . .	<b>257</b>
5.5.1 Retour sur le cas test linéaire . . . . .	257
5.5.2 Retour sur le cas test non-linéaire . . . . .	264
<b>5.6 Application au flambage d'une plaque composite</b> . . . . .	<b>275</b>
5.6.1 Transformation du problème initial . . . . .	275
5.6.2 Implémentation de la méthodologie développée . . . . .	276
<b>5.7 Conclusion</b> . . . . .	<b>278</b>

---

## 5.1 Introduction

Au Chapitre 4, on a cherché à étudier l'impact de l'incertitude d'échantillonnage sur l'estimateur  $\hat{f}_{\mathbf{X}}(\cdot | \mathbf{X}_{\text{obs}})$  de la densité conjointe puis sur l'estimateur  $\hat{P}_f(\mathbf{X}_{\text{obs}})$  de la probabilité de défaillance. On a développé un algorithme d'ASF capable d'identifier la composante fonctionnelle (la densité de copule *vs* l'une des densités marginales) de  $\hat{f}_{\mathbf{X}}(\cdot | \mathbf{X}_{\text{obs}})$  qui engendre le plus de variabilité sur  $\hat{P}_f(\mathbf{X}_{\text{obs}})$ . Si le principal contributeur de la variabilité d'apprentissage est la copule  $\hat{c}_{\mathbf{U}}$ , le processus d'acquisition de nouvelles données ne peut plus être qualifié de "ciblé" car il consiste à collecter des observations de toute la loi conjointe  $\mu_{\mathbf{X}}$  et celles-ci peuvent alors être utilisées pour raffiner la connaissance de la copule ainsi que la connaissance de toutes les lois marginales. On souhaite désormais se focaliser uniquement sur les incertitudes épistémiques véhiculées par l'estimation de la copule. Les lois marginales sont alors supposées connues. C'est un point de vue adopté dans des ouvrages tels que RÜSCHENDORF [1985] ou BENES et STEPÁN [2012]. La chaîne de propagation des incertitudes épistémiques devient alors :

$$\mathbf{U}_{\text{obs}} \sim (\mu_{\mathbf{U}})^{\otimes N} \longrightarrow \hat{c}_{\mathbf{U}}(\cdot | \mathbf{U}_{\text{obs}}) \longrightarrow P_f(\mathbf{U}_{\text{obs}}). \quad (5.1)$$

La matrice aléatoire  $\mathbf{U}_{\text{obs}}$  n'est pas constituée de pseudo-observations de la copule. En effet, ses lignes  $\mathbf{U}_{\text{obs}}^{[i \cdot]}$  sont indépendantes et identiquement distribuées suivant la loi  $\mu_{\mathbf{U}}$  de support  $\mathcal{H}_d$ . La phase d'apprentissage statistique  $\mathcal{L}$  ne désigne désormais que l'estimation de la copule. Quant à l'analyse de fiabilité  $\mathcal{R}$ , elle consiste à intégrer sur  $D_f$  la densité conjointe  $\hat{f}_{\mathbf{X}}(\cdot | \mathbf{U}_{\text{obs}})$  issue de la connaissance des lois marginales et de l'estimation de la copule. On est dans un contexte où la copule sous-jacente est inconnue et où l'on ne dispose que d'un unique échantillon d'apprentissage  $\mathbf{u}_{\text{obs}} \in [0, 1]^{N \times d}$  fourni par un retour d'expérience industriel. La propagation de l'incertitude d'échantillonnage à partir de ce seul matériel  $\mathbf{u}_{\text{obs}}$  n'est plus un véritable problème au regard des solutions proposées au Chapitre 4 où l'utilisation du *bootstrap* est fortement mise en avant.

La comparaison avec le chapitre précédent s'arrête dès que l'on prend conscience qu'il n'est pas possible de s'appuyer sur un résultat universel de la puissance du théorème de SKLAR [1959]. En effet, la densité  $\hat{c}_{\mathbf{U}}(\cdot | \mathbf{U}_{\text{obs}})$  et la fonction de répartition  $\hat{C}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  ne peuvent pas être réécrites à partir de sous-composantes fonctionnelles constituant une partition rigoureuse de l'information sur la structure de dépendance. Pour tout  $\mathbf{a} = \{a_1, \dots, a_r\} \subseteq \{1, \dots, d\}$  où  $2 \leq r < d$ , la sous-structure de dépendance entre les variables de  $\mathbf{U}_{\mathbf{a}}$  est décrite par la  $r$ -copule  $c_{\mathbf{U}_{\mathbf{a}}}$  :

$$\forall \mathbf{u}_{\mathbf{a}} \in \mathcal{H}_r, \quad c_{\mathbf{U}_{\mathbf{a}}}(\mathbf{u}_{\mathbf{a}}) = \int_{\mathcal{H}_{d-r}} c_{\mathbf{U}}(\mathbf{u}) d\mathbf{u}_{-\mathbf{a}} \quad \text{avec} \quad \mathbf{u}_{-\mathbf{a}} := [u_j]_{j \notin \mathbf{a}}. \quad (5.2)$$

et peut mériter d'être isolée et étudiée. Au total,  $2^d - (d + 2)$  sous-motifs de dépendance peuvent être extraits de la copule principale  $c_{\mathbf{U}}$ . Il serait vain et redondant de vouloir tous les estimer. Une idée peut consister à s'intéresser à toutes les  $r$ -copules de  $c_{\mathbf{U}}$  avec  $r \geq 2$  fixé :

$$\mathcal{C}_r(c_{\mathbf{U}}) := \{c_{\mathbf{U}_{\mathbf{a}}} \text{ avec } \mathbf{a} \subseteq \{1, \dots, d\} \text{ et } \text{Card}(\mathbf{a}) = r\}. \quad (5.3)$$

Quel que soit  $r$ , cette grille de description de  $c_{\mathbf{U}}$  n'est pas satisfaisante pour trois raisons :

1. **Dénaturation.**  $\mathcal{C}_r(c_{\mathbf{U}})$  ne représente qu'une partie de l'information permettant de caractériser la structure de dépendance. En général, la connaissance de toutes les 2-copules ne suffit pas à reconstruire les 3-copules, etc.
2. **Redondance.** Dès lors que  $d \geq 4$  et  $r \geq 3$ , l'information sur les  $s$ -copules (avec  $2 \leq s < r$ ) est portée par plusieurs éléments de  $\mathcal{C}_r(c_{\mathbf{U}})$ . Pour  $d = 4$  et  $r = 3$ , on peut ainsi voir que  $c_{\mathbf{U}_{12}}$  est englobée dans  $c_{\mathbf{U}_{123}}$  et dans  $c_{\mathbf{U}_{124}}$ .
3. **Indivisibilité.** Il n'existe pas de formule permettant d'écrire  $c_{\mathbf{U}}$  en fonction des seuls éléments de la classe  $\mathcal{C}_r(c_{\mathbf{U}})$ .

Par ailleurs, si on essaie de rassembler plusieurs classes, on limite la dénaturation mais on renforce le phénomène de redondance. On voit donc qu'il y a une difficulté à identifier des sous-entités fonctionnelles de la copule pouvant être déconnectées les unes des autres.

#### Problème 5.1: Compréhension de la variabilité d'apprentissage

Comment partitionner l'information portée par  $\hat{c}_{\mathbf{U}}(\cdot | \mathbf{U}_{\text{obs}})$  ?

Si on se contente de considérer  $\hat{c}_{\mathbf{U}}(\cdot | \mathbf{U}_{\text{obs}})$  comme une copule, au sens mathématique du terme, aucune des classes  $\mathcal{C}_r(c_{\mathbf{U}})$  ne semble devoir être privilégiée. Toutefois, il faut se rappeler que  $\hat{c}_{\mathbf{U}}(\cdot | \mathbf{U}_{\text{obs}})$  n'existe qu'à travers un certain modèle  $\mathcal{M}_{\text{cop}}$  de distribution de copule. Or, la plupart des modèles utilisés en pratique et présentés dans la Section 1.5.4 s'appuient principalement sur l'assemblage ou l'imbrication de copules bidimensionnelles. Certaines de ces copules bidimensionnelles représentent directement la dépendance entre des paires de variables  $(U_i, U_j)$  tandis que d'autres sont affectées à la modélisation de motifs de dépendance dont l'interprétation est moins immédiate. Extraire de  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  toutes les copules bidimensionnelles  $\hat{c}_{\mathbf{U}_{ij}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  peut ainsi constituer un bon descriptif de l'information stockée au sein de cette  $d$ -copule après l'exécution de l'algorithme  $\mathcal{L}$  sur  $\mathbf{U}_{\text{obs}}$ . La  $d$ -copule initiale donne ainsi naissance à une collection rassemblant  $p := C_d^2 = \frac{1}{2}d(d-1)$  copules bidimensionnelles. Dans le prolongement du travail réalisé au Chapitre 4, on va chercher à identifier la copule  $\hat{c}_{\mathbf{U}_{ij}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  qui exerce la plus grande influence sur la distribution de l'estimateur  $P_f(\mathbf{U}_{\text{obs}})$ .

#### Problème 5.2: Décomposition de la variabilité d'apprentissage

Comment quantifier l'influence de chaque copule  $\hat{c}_{\mathbf{U}_{ij}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  sur  $P_f(\mathbf{U}_{\text{obs}})$  ?

Ce problème s'inscrit à nouveau dans un contexte d'ASE. La difficulté vient du fait que les objets  $\hat{c}_{\mathbf{U}_{ij}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  en entrée sont fonctionnels et dépendants. Par ailleurs, contrairement au Chapitre 4 où la reformulation du problème d'ASF dans un cadre plus favorable a permis de se ramener à une situation d'indépendance entre les objets fonctionnels, rien ne permet ici d'adopter une stratégie analogue. En effet, le mécanisme d'inférence qui est à l'œuvre lors de l'apprentissage de la  $d$ -copule  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  introduit systématiquement une dépendance fonctionnelle entre les copules bidimensionnelles constituantes. Sauf dans des cas très particuliers, les copules bidimensionnelles extraites de  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  ne sont pas des entités mutuellement indépendantes. La résolution du problème d'ASF doit donc tenir compte de cette contrainte.

Supposons qu'on parvienne à identifier la copule  $\hat{c}_{\mathbf{U}_{ij}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  dont l'existence au sein de  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  a le plus fort impact sur la distribution de  $P_f(\mathbf{U}_{\text{obs}})$ . Se pose alors la question de savoir si l'amélioration de l'état de connaissance de cette copule bidimensionnelle peut permettre une réduction importante de la variabilité de prédiction du risque de défaillance.

#### Problème 5.3: Réduction de la variabilité d'apprentissage

L'acquisition ciblée de nouvelles données permet-elle de réduire la variabilité sur  $P_f(\mathbf{U}_{\text{obs}})$  ?

Après avoir collecté des observations supplémentaires pour la paire  $(U_i, U_j)$  qui a été ciblée par l'ASE, il faut comprendre comment ces nouvelles données peuvent être incorporées au sein d'un algorithme d'apprentissage statistique  $\mathcal{L}$  initialement conçu pour recevoir un échantillon  $\mathbf{U}_{\text{obs}}$  d'observations conjointes du vecteur  $\mathbf{U}$ . Dans certains cas, la copule  $\hat{c}_{\mathbf{U}_{ij}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  a une expression explicite et invariante, par exemple sous la forme d'une famille paramétrique  $\mathcal{F}_{\text{cop}}$ . Il est alors assez simple d'utiliser les données supplémentaires pour affiner la connaissance des paramètres de  $\mathcal{F}_{\text{cop}}$ . Au contraire, dans de nombreux autres cas,  $\hat{c}_{\mathbf{U}_{ij}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  présente une expression implicite, bien plus complexe, qui fait intervenir des paramètres ne pouvant pas être

estimés à partir d'observations de la paire  $(U_i, U_j)$ . Il est alors impossible de tirer profit des nouvelles données pour recalibrer  $\hat{c}_{U_{ij}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ . Par ailleurs, pour de nombreux modèles, la nature de l'expression de  $\hat{c}_{U_{ij}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  est susceptible d'évoluer en fonction de  $\mathbf{U}_{\text{obs}}$ . Face à ces contraintes, des stratégies d'assimilation doivent donc être mises au point de manière à ce que l'apport de nouvelles données puisse être valorisé.

Au sein de ce chapitre, la Section 5.2 vise à étudier les difficultés inhérentes à l'identification de la copule bidimensionnelle  $\hat{c}_{U_{ij}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  la plus influente sur  $P_f(\mathbf{U}_{\text{obs}})$ . Une version simplifiée du problème initial est notamment proposée. Elle consiste à examiner la propagation des incertitudes épistémiques véhiculées par les coefficients de corrélation empiriques sur  $P_f(\mathbf{U}_{\text{obs}})$ . Dans la Section 5.3, un algorithme d'ASF est développé pour répondre à ce sous-problème. De nombreux éléments déjà présentés au Chapitre 4 (ré-échantillonnage *bootstrap*, interfaçage des briques de calcul, gestion optimale du budget de simulation) sont réutilisés sans être détaillés à nouveau. En revanche, une attention plus particulière est portée aux spécificités de ce nouveau contexte d'ASF (utilisation des  $\delta$ -indices de Borgonovo et estimation d'un indicateur de stochasticité). Des stratégies d'assimilation de données supplémentaires sont exposées à la Section 5.4 et conviennent notamment aux copules gaussiennes et aux copules R-vignes. Les différentes méthodologies développées sont validées à la Section 5.5 par des cas tests analytiques avant d'être mises en application à la Section 5.6 sur le cas test physique précédemment étudié au Chapitre 4.

## 5.2 Formulation d'un problème d'ASF

Désormais, la copule bidimensionnelle  $\hat{c}_{U_{ij}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  extraite de la  $d$ -copule conjointe  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  est notée  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$  pour alléger les notations. Par ailleurs, on suppose pour le moment que le calcul de  $P_f(\mathbf{U}_{\text{obs}})$  à partir de  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  est une tâche triviale qui peut être exécutée à partir d'un algorithme de fiabilité  $\mathcal{R}$  déterministe et sans coût de calcul.

### 5.2.1 Étude des difficultés inhérentes à l'usage des copules bidimensionnelles

Conformément à ce qui a été dit précédemment, on cherche à expliquer la variabilité observée sur  $P_f(\mathbf{U}_{\text{obs}})$  à partir de celles portées par toutes les copules bidimensionnelles  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$ . La chaîne de propagation des incertitudes épistémiques est représentée à la Figure 5.1 et on retrouve le phénomène de transformation de la variabilité d'échantillonnage, d'abord en variabilité d'apprentissage, puis en variabilité de prédiction du risque de défaillance. Pour décrire le comportement aléatoire des différents objets mathématiques mis en jeu, on utilise les notations suivantes :

- $(\mu_{\mathbf{U}})^{\otimes N} \implies$  La mesure d'échantillonnage,
- $\mu_{\hat{c}} \implies$  L'image de  $(\mu_{\mathbf{U}})^{\otimes N}$  après estimation de  $c_{\mathbf{U}}$  selon  $\mathcal{M}_{\text{cop}}$ ,
- $\mu_{\hat{c}_{ij}} \implies$  L'image de  $\mu_{\hat{c}}$  après extraction de la copule associée à la paire  $(U_i, U_j)$ ,
- $\mu_{P_f} \implies$  L'image de  $\mu_{\hat{c}}$  après calcul de la probabilité de défaillance.

Dans le Chapitre 4, l'algorithme d'apprentissage statistique  $\mathcal{L}$  fournit une estimation des différentes composantes fonctionnelles de la séparation copule-marginales et celles-ci sont ensuite ré-assemblées pour donner naissance à une estimation de la densité conjointe  $\hat{f}_{\mathbf{X}}(\cdot | \mathcal{M}, \mathbf{X}_{\text{obs}})$ . La mesure  $\mu_{P_f}$  est alors une transformation directe des mesures  $\mu_{\hat{f}_1}, \dots, \mu_{\hat{f}_d}$  et  $\mu_{\hat{c}}$  après franchissement d'un opérateur déterministe. Ici, c'est une mécanique différente qui est étudiée. L'algorithme  $\mathcal{L}$  produit une estimation de la densité conjointe  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  et on en déduit des estimations  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$  pour toutes les copules bidimensionnelles. On peut alors se demander s'il est possible d'expliquer la mesure  $\mu_{P_f}$  à partir de la collection des mesures  $\mu_{\hat{c}_{ij}}$ . Comme indiqué à la Figure 5.1, on s'intéresse à l'application :

$$\Psi: \quad \mathbb{L}^2(\mathcal{H}_2) \quad \times \quad \mathbb{L}^2(\mathcal{H}_2) \quad \times \dots \times \quad \mathbb{L}^2(\mathcal{H}_2) \quad \longrightarrow \quad [0, 1] \\ \hat{c}_{12}(\cdot | \mathbf{U}_{\text{obs}}) \quad \hat{c}_{13}(\cdot | \mathbf{U}_{\text{obs}}) \quad \dots \quad \hat{c}_{(d-1)d}(\cdot | \mathbf{U}_{\text{obs}}) \quad \longmapsto \quad P_f(\mathbf{U}_{\text{obs}}) \quad . \quad (5.4)$$

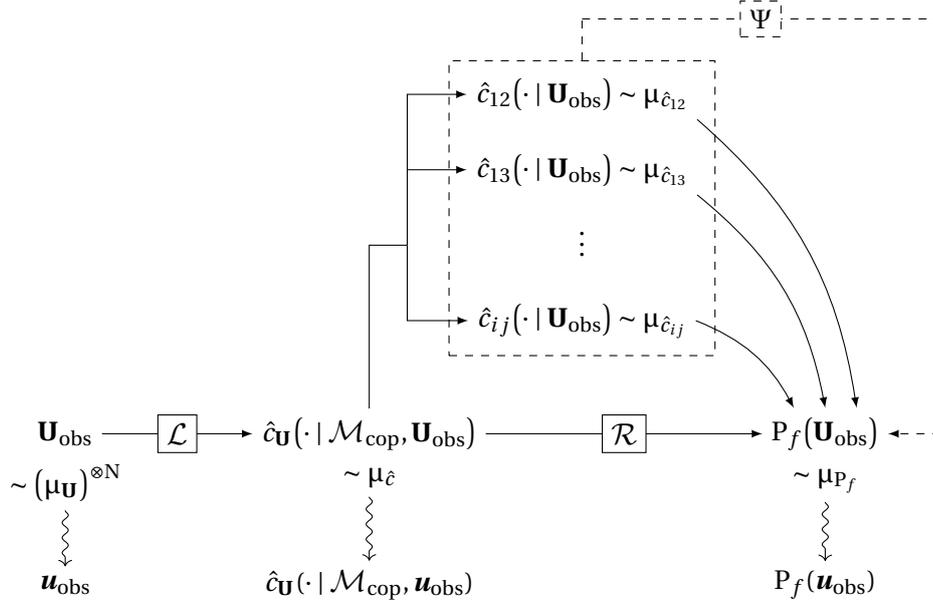


FIGURE 5.1 – Propagation de la variabilité d'échantillonnage.

dont on veut analyser la sensibilité lorsque  $\mathbf{U}_{\text{obs}}$  propage l'incertitude d'échantillonnage à la fois sur ses entrées et sa sortie. Une première difficulté réside dans le fait que cette propagation est en général indirecte. En effet, l'incertitude portée par  $P_f(\mathbf{U}_{\text{obs}})$  ne peut pas être appréhendée comme le résultat de la propagation des incertitudes portées par les copules bidimensionnelles  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$ . Les incertitudes en entrée et en sortie de  $\Psi$  ne peuvent pas être mises en relation sans remonter à leur source commune qui est la  $d$ -copule  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ . L'application  $\Psi$  est donc potentiellement stochastique. Par ailleurs, une seconde difficulté vient du fait qu'il existe une dépendance fonctionnelle que l'on ne sait pas caractériser entre les copules bidimensionnelles  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$ . Dans la suite, on se propose d'examiner plusieurs cas jouets en dimension  $d = 3$  pour mieux comprendre la situation. L'objectif est de montrer que les deux difficultés venant d'être décrites dépendent du modèle d'apprentissage  $\mathcal{M}_{\text{cop}}$  utilisé.

### Situation 1. Apprentissage d'une copule archimédienne

Dans l'esprit de ce qui est par exemple fait dans ZHANG et SINGH [2007], on cherche à estimer  $c_{\mathbf{U}}$  à l'aide d'une copule de Gumbel. La distribution calibrée à partir de  $\mathbf{U}_{\text{obs}}$  dépend d'un unique paramètre  $\hat{\theta} \geq 1$  et une expression synthétique en est fournie par :

$$\forall \mathbf{u} \in \mathcal{H}_3, \hat{C}_{\mathbf{U}}(\mathbf{u} | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}}) = \hat{C}_G(\mathbf{u} | \hat{\theta}) = \exp\left(-\left[\sum_{k=1}^3 (-\log(u_k))^{\hat{\theta}}\right]^{1/\hat{\theta}}\right), \quad (5.5)$$

où  $\hat{C}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  est la fonction de répartition associée à la densité  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ . On déduit de façon triviale l'expression des fonctions de répartition des copules bidimensionnelles :

$$\begin{aligned} \forall i \neq j, \hat{C}_{ij}(u_i, u_j | \mathbf{U}_{\text{obs}}) &= \hat{C}_{\mathbf{U}}(\mathbf{v} | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}}) \text{ avec } \mathbf{v} = [v_k]_{1 \leq k \leq 3} \text{ et } v_k = \begin{cases} u_i & \text{si } k = i \\ u_j & \text{si } k = j \\ 1 & \text{sinon} \end{cases} \\ &= \exp\left(-\left[(-\log(u_i))^{\hat{\theta}} + (-\log(u_j))^{\hat{\theta}}\right]^{1/\hat{\theta}}\right) = \hat{C}_G(u_{ij} | \hat{\theta}). \end{aligned} \quad (5.6)$$

Toute copule bidimensionnelle est une copule de Gumbel de paramètre  $\hat{\theta}$ . On est donc dans une situation d'extrême dépendance fonctionnelle qui se traduit par une égalité parfaite entre toutes

les copules  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$ . On peut remarquer que l'application  $\Psi$  est alors complètement déterministe. La connaissance d'une des copules bidimensionnelles donne accès au paramètre  $\hat{\theta}$ , ce qui permet de reconstruire la  $d$ -copule  $\hat{c}_G(\cdot | \hat{\theta})$  puis de calculer la probabilité de défaillance  $P_f(\mathbf{U}_{\text{obs}})$ . Dans ce contexte, analyser la sensibilité de  $\Psi$  n'est pas d'un grand intérêt. Quelle que soit la copule bidimensionnelle ciblée par l'ASF, la réduction de sa variabilité d'apprentissage passe par une réduction de la dispersion du paramètre  $\hat{\theta}$  qui va également profiter à toutes les autres copules bidimensionnelles. Cette première situation n'est en fait jamais rencontrée en pratique car les copules archimédiennes ne sont plus utilisées dès lors qu'on monte en dimension, leur unique paramètre n'offrant plus suffisamment de flexibilité dans l'apprentissage des motifs de dépendance.

**Situation 2. Apprentissage d'une copule elliptique**

On imagine désormais que  $c_{\mathbf{U}}$  est estimée à l'aide d'une copule gaussienne comme c'est le cas dans [CLEMEN et REILLY \[1999\]](#). La fonction de répartition issue de l'apprentissage s'exprime alors :

$$\forall \mathbf{u} \in \mathcal{H}_3, \hat{c}_{\mathbf{U}}(\mathbf{u} | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}}) = \hat{c}_{\mathcal{N}}(\mathbf{u} | \hat{\boldsymbol{\rho}}) = \Phi_3(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \Phi^{-1}(u_3) | \hat{\boldsymbol{\rho}}) ,$$

$$\text{avec } \hat{\boldsymbol{\rho}} = \begin{bmatrix} 1 & \hat{\rho}_{12} & \hat{\rho}_{13} \\ \hat{\rho}_{12} & 1 & \hat{\rho}_{23} \\ \hat{\rho}_{13} & \hat{\rho}_{23} & 1 \end{bmatrix} \text{ et } \begin{cases} \Phi(\cdot) & \text{la fonction de répartition de } \mathcal{N}(0, 1) , \\ \Phi_3(\cdot | \hat{\boldsymbol{\rho}}) & \text{la fonction de répartition de } \mathcal{N}(\mathbf{0}, \hat{\boldsymbol{\rho}}) . \end{cases} \quad (5.7)$$

Les fonctions de répartition des copules bidimensionnelles s'obtiennent comme précédemment :

$$\forall i \neq j, \hat{c}_{ij}(u_i, u_j | \mathbf{U}_{\text{obs}}) = \hat{c}_{\mathbf{U}}(\mathbf{v} | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$$

$$= \Phi_2(\Phi^{-1}(u_i), \Phi^{-1}(u_j) | \hat{\rho}_{ij}) = \hat{c}_{\mathcal{N}}(u_i, u_j | \hat{\rho}_{ij}) . \quad (5.8)$$

Ainsi, toute copule  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$  est gaussienne et paramétrée par un coefficient  $\hat{\rho}_{ij}$  de corrélation linéaire. Pour la copule gaussienne, il y a autant de paramètres que de paires de variables. C'est une situation idéale dans laquelle chaque copule bidimensionnelle est décrite par un unique paramètre qui lui est exclusivement réservé. Analyser la sensibilité de  $P_f(\mathbf{U}_{\text{obs}})$  à l'apprentissage des copules bidimensionnelles  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$  revient à analyser la sensibilité de  $P_f(\mathbf{U}_{\text{obs}})$  à l'estimation des paramètres  $\hat{\rho}_{ij}$ . On se ramène alors à un problème d'ASF où les entrées sont scalaires mais dépendantes. Comme on a pu le voir à la Section 4.3.1.2, l'estimation par maximum de vraisemblance (MV) du coefficient de corrélation  $\rho_{ij}$  d'une copule gaussienne repose sur la formule :

$$\hat{\rho}_{ij} = \frac{\hat{\Gamma}_{ij}}{\sqrt{\hat{\Gamma}_{ii}}\sqrt{\hat{\Gamma}_{jj}}} \text{ avec } \hat{\Gamma}_{ij} = \frac{1}{N} \sum_{n=1}^N E_i^{(n)} E_j^{(n)} \text{ et } \forall s \in \{i, j\}, E_s^{(n)} = \Phi^{-1}(U_s^{(n)}) . \quad (5.9)$$

Il est alors évident que les coefficients  $(\hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23})$  sont dépendants car ils sont calculés à partir des colonnes  $(\mathbf{U}_{\text{obs}}^{[\bullet 1]}, \mathbf{U}_{\text{obs}}^{[\bullet 2]}, \mathbf{U}_{\text{obs}}^{[\bullet 3]})$  qui sont dépendantes. Le fait de réduire chaque copule bidimensionnelle  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$  à son paramètre  $\hat{\rho}_{ij}$  transforme les entrées fonctionnelles en des entrées scalaires mais n'élimine pas la dépendance statistique entre les entrées. Pour résoudre le problème d'ASF, les indices de Sobol ne sont plus adaptés et il faut avoir recours à l'une des alternatives présentées à la Section 3.2.3. Par ailleurs, on est à nouveau dans une situation où  $\Psi$  est complètement déterministe. La connaissance des copules bidimensionnelles  $\hat{c}_{\mathcal{N}}(\cdot | \hat{\rho}_{ij})$  permet de reconstruire la matrice de corrélation  $\hat{\boldsymbol{\rho}}$  puis la  $d$ -copule  $\hat{c}_{\mathcal{N}}(\cdot | \hat{\boldsymbol{\rho}})$  et enfin de calculer  $P_f(\mathbf{U}_{\text{obs}})$ . Un raisonnement analogue aurait pu être mené à partir de l'estimation d'une  $d$ -copule de Student :

$$\forall \mathbf{u} \in \mathcal{H}_3, \hat{c}_{\mathbf{U}}(\mathbf{u} | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}}) = \hat{c}_t(\mathbf{u} | \hat{\boldsymbol{\rho}}, \hat{\nu}) \text{ avec } \begin{cases} \hat{\boldsymbol{\rho}} & \text{la matrice de corrélation,} \\ \hat{\nu} & \text{le nombre de degrés de liberté.} \end{cases} \quad (5.10)$$

Le résultat énoncé dans [FANG \[2018\]](#) sur la nature des lois marginales d'une distribution elliptique permet d'affirmer que toute copule bidimensionnelle  $\hat{c}_{ij}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  est une copule de Student de paramètres  $\hat{\rho}_{ij}$  et  $\hat{\nu}$ . La bijection entre les copules bidimensionnelles et les paramètres

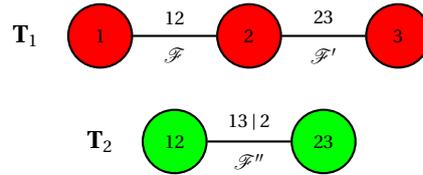
ne tient plus étant donné que  $\hat{v}$  est maintenant impliqué dans la description de toutes les paires de variables. On est alors dans une situation intermédiaire où le problème d'ASF est à mi-chemin entre celui découlant de l'apprentissage d'une copule de Gumbel et celui découlant de l'apprentissage d'une copule gaussienne. On peut choisir d'analyser la sensibilité de  $(\hat{\rho}, \hat{v}) \mapsto P_f(\mathbf{U}_{\text{obs}})$  mais l'éventuelle identification de  $\hat{v}$  en tant que paramètre le plus influent ne donne aucune indication sur la paire de variables à cibler. S'il est en revanche décidé d'analyser la sensibilité de  $\hat{\rho} \mapsto P_f(\mathbf{U}_{\text{obs}})$ , le lien entre les paramètres et la probabilité de défaillance devient stochastique étant donné que l'incertitude liée à la dispersion de  $\hat{v}$  n'est plus prise en compte. Si l'on dispose d'une méthodologie d'ASF appropriée à ce contexte, on peut cibler un certain coefficient de corrélation  $\hat{\rho}_{ij}$  et remonter à sa copule bidimensionnelle  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$ .

**Situation 3. Apprentissage d'une copule D-vigne**

On se place désormais dans le cas de l'apprentissage d'une copule D-vigne dont la séquence d'arbres  $\mathbf{T}_{\text{vg}}$  et les familles paramétriques associées  $\mathbf{F}_{\text{vg}}$  sont spécifiées en amont :

$$\mathbf{T}_{\text{vg}} := \{\mathbf{T}_1, \mathbf{T}_2\} \quad \text{avec} \quad \begin{cases} \mathbf{E}_1 = \{\{\{1\}, \{2\}\}, \{\{2\}, \{3\}\}\} \\ \mathbf{E}_2 = \{\{\{1, 2\}, \{2, 3\}\}\} \end{cases} \quad \text{et} \quad \mathbf{F}_{\text{vg}} := \{\mathcal{F}, \mathcal{F}', \mathcal{F}''\}. \quad (5.11)$$

Les familles paramétriques  $\mathcal{F}$  et  $\mathcal{F}'$  sont respectivement affectées à la modélisation des copules bidimensionnelles  $c_{12}$  et  $c_{23}$  tandis que la famille paramétrique  $\mathcal{F}''$  sous-tend l'apprentissage de la copule conditionnelle  $c_{13|2}$  :



On suppose que chaque famille de  $\mathbf{F}_{\text{vg}}$  admet un paramétrage scalaire. Les paramètres sont respectivement notés  $\alpha$ ,  $\beta$  et  $\gamma$ . La copule construite à l'aide de ce modèle  $\mathcal{M}_{\text{cop}}$  est caractérisée par la densité suivante :

$$\begin{aligned} \forall \mathbf{u} \in \mathcal{H}_3, \quad c_{\mathbf{U}}(\mathbf{u} | \mathcal{M}_{\text{cop}}) &= c_{\text{vg}}(\mathbf{u} | \mathbf{T}_{\text{vg}}, \mathbf{F}_{\text{vg}}, \boldsymbol{\theta}_{\text{vg}}) \\ &= c_{12}(u_1, u_2 | \mathcal{F}, \alpha) \times c_{23}(u_2, u_3 | \mathcal{F}', \beta) \dots \\ &\quad \times c_{13|2}(h_{12}^{(D)}(u_1, u_2), h_{23}^{(G)}(u_2, u_3) | \mathcal{F}'', \gamma) \end{aligned} \quad (5.12)$$

$$\text{avec} \quad \begin{cases} h_{12}^{(D)}(u_1, u_2) = \frac{\partial C_{12}}{\partial u_2}(u_1, u_2 | \mathcal{F}, \alpha), \\ h_{23}^{(G)}(u_2, u_3) = \frac{\partial C_{23}}{\partial u_2}(u_2, u_3 | \mathcal{F}', \beta). \end{cases}$$

Compte tenu du fait que  $\mathbf{T}_{\text{vg}}$  et  $\mathbf{F}_{\text{vg}}$  sont déjà sélectionnés, l'apprentissage statistique  $\mathcal{L}$  consiste à estimer les paramètres  $\boldsymbol{\theta}_{\text{vg}} = (\alpha, \beta, \gamma)$  en maximisant la log-vraisemblance donnée par :

$$\ell_N(\boldsymbol{\theta}_{\text{vg}}) = \sum_{i=1}^N \log \left( c_{\text{vg}}(\mathbf{U}^{(i)} | \mathbf{T}_{\text{vg}}, \mathbf{F}_{\text{vg}}, \boldsymbol{\theta}_{\text{vg}}) \right). \quad (5.13)$$

Après calibration du modèle  $\mathcal{M}_{\text{cop}}$  aux observations de l'échantillon  $\mathbf{U}_{\text{obs}}$ , les copules bidimensionnelles sont extraites de la  $d$ -copule  $\hat{c}_{\text{vg}}(\cdot | \mathbf{T}_{\text{vg}}, \mathbf{F}_{\text{vg}}, \hat{\boldsymbol{\theta}}_{\text{vg}})$ . Les paires reliées au sein de l'arbre  $\mathbf{T}_1$  indiquent les copules dont on peut obtenir une description explicite :

$$\hat{c}_{12}(\cdot | \mathbf{U}_{\text{obs}}) = \hat{c}(\cdot | \mathcal{F}, \hat{\alpha}) \quad \text{et} \quad \hat{c}_{23}(\cdot | \mathbf{U}_{\text{obs}}) = \hat{c}(\cdot | \mathcal{F}', \hat{\beta}). \quad (5.14)$$

Au contraire, la paire non représentée dans  $\mathbf{T}_1$  n'a qu'une expression implicite :

$$\hat{c}_{13}(u_1, u_3 | \mathbf{U}_{\text{obs}}) = \int_0^1 \hat{c}_{\text{vg}}(\mathbf{u} | \mathbf{T}_{\text{vg}}, \mathbf{F}_{\text{vg}}, \hat{\boldsymbol{\theta}}_{\text{vg}}) du_2 = \lambda(u_1, u_3 | \hat{\alpha}, \hat{\beta}, \hat{\gamma}). \quad (5.15)$$

La fonction  $\lambda$  dépend des trois paramètres de  $\hat{\boldsymbol{\theta}}_{\text{vg}}$  et constitue une estimation de la densité de la copule bidimensionnelle  $c_{13}$ . Le calcul intégral étant impraticable,  $\lambda$  n'a pas d'expression simple. Pour un point  $(\bar{u}_1, \bar{u}_3)$  de l'hypercube  $\mathcal{H}_2$ , la valeur  $\lambda(\bar{u}_1, \bar{u}_3 \mid \hat{\alpha}, \hat{\beta}, \hat{\gamma})$  peut toutefois être approchée par intégration numérique sur l'intervalle  $[0, 1]$  ou par Monte-Carlo. Contrairement à la copule gaussienne où le paramétrage  $\boldsymbol{\rho}$  est construit sur la base d'un cloisonnement entre les différentes paires de variables, la situation est beaucoup plus complexe au sein de la D-vigne :

- Le paramètre  $\alpha$  (resp.  $\beta$ ) est initialement prévu pour modéliser la copule  $c_{12}$  (resp.  $c_{23}$ ) mais il intervient également dans la modélisation de la copule  $c_{13}$ .
- Le paramètre  $\gamma$  est destiné au pilotage de la copule conditionnelle  $c_{13|2}$  et touche donc une information beaucoup plus subtile que celle portée par  $c_{13}$ . On rappelle que :

$$\forall \mathbf{u} \in \mathcal{H}_3, \quad c_{13|2} \left( \frac{\partial C_{12}}{\partial u_2}(u_1, u_2), \frac{\partial C_{23}}{\partial u_2}(u_2, u_3) \right) = \frac{c_{\mathbf{U}}(\mathbf{u})}{c_{12}(u_1, u_2) \times c_{23}(u_2, u_3)}. \quad (5.16)$$

$c_{13|2}$  peut être interprétée comme la structure de dépendance résiduelle après élimination des motifs de dépendance déjà décrits par  $c_{12}$  et  $c_{23}$ . C'est une caractéristique du vecteur  $\mathbf{U}$  et non une simple caractéristique de la paire  $(U_1, U_3)$ . D'un point de vue purement analytique, on pourrait uniquement rattacher  $\gamma$  à la copule  $c_{13}$  mais il faut bien comprendre que ce paramètre a été incorporé à  $\boldsymbol{\theta}_{\text{vg}}$  dans un autre dessein.

La construction d'un modèle  $\mathcal{M}_{\text{cop}}$  plus flexible a nécessité la mise en place d'un paramétrage  $\boldsymbol{\theta}_{\text{vg}}$  au sein duquel le rôle des paramètres n'est plus uniquement limité à la description des liens de dépendance entre paires de variables. Analyser la sensibilité de  $\hat{\boldsymbol{\theta}}_{\text{vg}} \mapsto P_f(\mathbf{U}_{\text{obs}})$  peut conduire à identifier un paramètre intervenant au sein de plusieurs copules bidimensionnelles. Il est alors difficile de décider quelle est la paire à cibler en priorité. Par ailleurs, si on choisit néanmoins d'étudier la réponse de  $P_f(\mathbf{U}_{\text{obs}})$  aux variations de  $\hat{\boldsymbol{\theta}}_{\text{vg}}$ , on doit tenir compte du fait que les paramètres  $\hat{\alpha}$ ,  $\hat{\beta}$  et  $\hat{\gamma}$  sont rendus dépendants par le processus d'estimation par MV. Si les difficultés sont accrues dans le cas d'une D-vigne, il est bon de remarquer que  $\Psi$  reste toutefois une application déterministe. Considérons une réalisation des trois copules bidimensionnelles. D'une part,  $\hat{c}_{12}(\cdot \mid \mathbf{u}_{\text{obs}})$  et  $\hat{c}_{23}(\cdot \mid \mathbf{u}_{\text{obs}})$  donnent accès aux valeurs de  $\hat{\alpha}$  et  $\hat{\beta}$ . D'autre part, connaissant  $\hat{\alpha}$  et  $\hat{\beta}$ , il n'y a qu'une seule valeur de  $\hat{\gamma}$  qui peut conduire à  $\hat{c}_{13}(\cdot \mid \mathbf{u}_{\text{obs}}) = \lambda(\cdot \mid \hat{\alpha}, \hat{\beta}, \hat{\gamma})$ . On en déduit qu'il n'existe qu'une seule D-vigne de type  $\hat{c}_{\text{vg}}(\cdot \mid \mathbf{T}_{\text{vg}}, \mathbf{F}_{\text{vg}}, \hat{\boldsymbol{\theta}}_{\text{vg}})$  pouvant engendrer ces trois copules bidimensionnelles. La probabilité de défaillance  $P_f(\mathbf{u}_{\text{obs}})$  qui en découle est alors la seule sortie possible de  $\Psi$ .

Imaginons maintenant que l'apprentissage de la D-vigne se fasse différemment. La séquence d'arbres  $\mathbf{T}_{\text{vg}}$  reste fixée mais le choix des familles de copules bidimensionnelles est libre. Cela signifie que la  $d$ -copule estimée s'écrit  $\hat{c}_{\mathbf{U}}(\cdot \mid \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}}) = \hat{c}_{\text{vg}}(\cdot \mid \mathbf{T}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}}, \hat{\boldsymbol{\theta}}_{\text{vg}})$  où  $\hat{\mathbf{F}}_{\text{vg}} = \{ \hat{\mathcal{F}}, \hat{\mathcal{F}}', \hat{\mathcal{F}}'' \}$  rassemble les différentes familles retenues au cours de l'apprentissage. Pour chaque copule bidimensionnelle impliquée dans la factorisation en D-vigne, une famille paramétrique est ainsi sélectionnée en maximisant un critère d'information au sein d'un catalogue de familles usuelles. La procédure séquentielle de descente des arbres  $\mathbf{T}_1$  et  $\mathbf{T}_2$  décrite dans [AAS et collab. \[2009\]](#) permet de choisir  $\hat{\mathbf{F}}_{\text{vg}}$  à partir de  $\mathbf{U}_{\text{obs}}$ . Une fois que la structure  $(\mathbf{T}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}})$  de la D-vigne est figée, les paramètres  $\boldsymbol{\theta}_{\text{vg}}$  sont recalés par MV. Dans ce contexte, il est beaucoup plus complexe de mettre en place une méthodologie d'ASF. L'expression analytique de chaque copule  $\hat{c}_{ij}(\cdot \mid \mathbf{U}_{\text{obs}})$  est conditionnée à la valeur prise par une variable discrète latente  $Z_{ij}$  indiquant la famille sélectionnée pour la paire  $(U_i, U_j)$ . L'identification des copules bidimensionnelles au moyen des paramètres du modèle  $\mathcal{M}_{\text{cop}}$  est une méthode qui est insuffisante lorsque l'ensemble  $\mathbf{F}_{\text{vg}}$  est imposé, et qui devient impossible lorsque l'ensemble  $\hat{\mathbf{F}}_{\text{vg}}$  est sélectionné automatiquement à partir de  $\mathbf{U}_{\text{obs}}$ . Une situation encore plus complexe peut être envisagée si la séquence d'arbres  $\mathbf{T}_{\text{vg}}$  est inconnue. Il faut alors utiliser l'algorithme de [DISSMANN et collab. \[2013\]](#) pour estimer la  $d$ -copule  $\hat{c}_{\text{vg}}(\cdot \mid \hat{\mathbf{T}}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}}, \hat{\boldsymbol{\theta}}_{\text{vg}})$ . Une même paire de variables  $(U_i, U_j)$  peut alors être modélisée explicitement puis implicitement selon le squelette de l'arbre  $\mathbf{T}_1$ .

En dimension  $d = 3$ , l'étude de trois modèles d'apprentissage de la copule  $c_{\mathbf{U}}$  a permis de mieux cerner la nature des difficultés rencontrées lors de la mise en œuvre d'une ASF pour  $\Psi$  :

- Il existe une dépendance fonctionnelle entre les copules bidimensionnelles  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$  quel que soit le modèle  $\mathcal{M}_{\text{cop}}$  choisi et la technique de calibration retenue.
- La situation la plus simple à traiter est celle de la copule gaussienne où chaque paire  $(U_i, U_j)$  est pilotée par un paramètre scalaire qui lui est exclusivement réservé. Le problème d'ASF à entrées fonctionnelles se résume alors à un problème d'ASF à entrées scalaires.
- Les modèles de copules adaptés à la grande dimension impliquent une description bien plus sophistiquée de la structure de dépendance. Le système de paramétrage est généralement rendu évolutif par un mécanisme d'arbres permettant d'adapter le squelette de la structure de dépendance aux données rencontrées. De même, les familles des différentes copules bidimensionnelles impliquées ne sont pas spécifiées et doivent être sélectionnées.
- Un résumé en dimension finie de la  $d$ -copule  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  impliquerait donc des variables discrètes (pour la modélisation des arbres), des variables latentes (pour la modélisation des familles) et des variables continues (pour la modélisation des paramètres). Analyser la sensibilité de  $P_f(\mathbf{U}_{\text{obs}})$  à toutes ces variables est un problème excessivement complexe et actuellement laissé sans solution.

Dans la prochaine section, on propose une démarche permettant de simplifier le problème d'ASF et d'envisager de le résoudre quel que soit le modèle  $\mathcal{M}_{\text{cop}}$  utilisé pour l'apprentissage.

## 5.2.2 Transformation du problème d'ASF initial

### 5.2.2.1 Étude des méthodes d'AS pour entrées fonctionnelles dépendantes

Dans la section précédente, on a compris qu'il est rarement possible d'utiliser le paramétrage de la  $d$ -copule  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  pour se ramener à un problème en dimension finie et construire une procédure ad hoc. La question est alors de savoir s'il est possible d'implémenter l'ASF sur  $\Psi$  à partir des copules bidimensionnelles  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$ . Examinons ce qui est proposé dans la littérature pour analyser la sensibilité d'une sortie scalaire à des entrées fonctionnelles dépendantes. Dans les travaux de [NANTY \[2015\]](#), les techniques d'AS pour entrées dépendantes sont étudiées dans un cadre fonctionnel. Tout repose sur la décomposition fonctionnelle des entrées du problème. Imaginons un code de calcul  $\phi_f$  qui relie  $p$  entrées fonctionnelles et une sortie scalaire :

$$[f_1(\cdot), \dots, f_p(\cdot)] \xrightarrow{\phi_f} Y . \quad (5.17)$$

Chaque entrée fonctionnelle  $f_i(\cdot)$  est décomposée sur une base de fonctions  $\{g_k(\cdot)\}_{k \geq 1}$ . Se posent alors les questions du choix de la base et de la dépendance des résultats de l'AS à la base retenue mais elles dépassent le cadre de ce travail. Une fois obtenue, la décomposition de  $f_i(\cdot)$  est tronquée, c'est-à-dire qu'elle est limitée à ses  $q$  premières fonctions de base. On a ainsi :

$$\forall i \in \{1, \dots, p\}, \quad f_i(\cdot) = \sum_{k=1}^{\infty} \alpha_{ik} g_k(\cdot) \approx \sum_{k=1}^q \alpha_{ik} g_k(\cdot) , \quad (5.18)$$

et on note  $\alpha_i := [\alpha_{ik}]_{1 \leq k \leq q}$  le vecteur des coordonnées de  $f_i(\cdot)$  dans la base sélectionnée. Dans un contexte de quantification des incertitudes, [ANSTETT-COLLIN et collab. \[2015\]](#) utilisent la décomposition de Karhunen-Loève tandis que [NANTY et collab. \[2017\]](#) préfèrent l'emploi de la décomposition par moindres carrés partiels. Après avoir décomposé toutes les entrées fonctionnelles, on peut rassembler tous les coefficients dans une même matrice aléatoire :

$$\alpha := [\alpha_i]_{1 \leq i \leq p} = [\alpha_{ik}]_{\substack{1 \leq i \leq p \\ 1 \leq k \leq q}} . \quad (5.19)$$

Celle-ci constitue un résumé dans  $\mathbb{R}^{p \times q}$  du comportement simultané des  $p$  entrées fonctionnelles. Comme les coefficients  $\alpha_{ik}$  sont des variables aléatoires dépendantes et potentiellement nombreuses, il a été proposé dans [JACQUES et PREDA \[2014\]](#) d'apprendre leur distribution conjointe  $\mu_{\alpha}$

à l'aide d'un mélange de lois gaussiennes multidimensionnelles. Si on dispose d'une série d'observations des entrées fonctionnelles, on peut fabriquer par projection une série d'observations du vecteur  $\alpha$  puis obtenir une estimation  $\hat{f}_\alpha$  de la densité conjointe des coefficients. On est alors en mesure de simuler de nouvelles observations des coefficients puis des entrées fonctionnelles. Si la question de la propagation des incertitudes semble réglée, il reste à savoir comment quantifier l'influence de chaque entrée fonctionnelle  $f_i(\cdot)$  sur la sortie  $Y$ . **NANTY et collab. [2016]** proposent de calculer un indice de Sobol pour chaque bloc de coefficients :

$$\forall i \in \{1, \dots, p\}, \quad S_i := \frac{\mathbb{V}(\mathbb{E}[Y | \alpha_i])}{\mathbb{V}(Y)}. \quad (5.20)$$

Étant donné que les variables fonctionnelles  $f_i(\cdot)$  sont dépendantes, les blocs de coefficients  $\alpha_i$  qui permettent de les représenter le sont aussi. Pour pallier ce problème, une procédure d'estimation a été développée par **JACQUES et collab. [2006]** puis améliorée par **GILQUIN et collab. [2015]**.

Dans le contexte d'ASF propre à l'application  $\Psi$ , la démarche de **NANTY [2015]** semble difficile à mettre en œuvre. Tout d'abord, il faut bien noter que les  $p$  entrées fonctionnelles  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$  sont des copules bidimensionnelles, c'est-à-dire des fonctions définies sur l'hypercube  $\mathcal{H}_2 = [0, 1]^2$  et vérifiant des conditions aux bords bien spécifiques :

$$\int_0^1 \hat{c}_{ij}(u_i, u_j | \mathbf{U}_{\text{obs}}) du_i = \mathbb{1}_{[0,1]}(u_j) \quad \text{et} \quad \int_0^1 \hat{c}_{ij}(u_i, u_j | \mathbf{U}_{\text{obs}}) du_j = \mathbb{1}_{[0,1]}(u_i) \quad . \quad (5.21)$$

Dans l'idéal, on aimerait disposer d'une base de copules  $\{g_k(\cdot)\}_{k \geq 1}$  mais il n'en existe pas dans la littérature à notre connaissance. En revanche, certaines techniques d'estimation de la densité d'une copule bidimensionnelle ont été développées en s'appuyant sur les possibilités offertes par les bases fonctionnelles traditionnelles. À partir d'un échantillon de pseudo-observations fabriquées avec la technique de renormalisation empirique présentée à la Section 1.5.1 et notée (SP2) dans le Chapitre 4, **GENEST et collab. [2009]** proposent de lisser la copule empirique par applications successives d'un filtrage par ondelettes. Une idée de **MARX et EILERS [2005]** ensuite reprise par **KAUERMANN et collab. [2013]** consiste à construire une base de fonctions  $\{g_k(\cdot)\}_{k \geq 1}$  où chaque fonction  $g_k : \mathcal{H}_2 \rightarrow \mathbb{R}$  est obtenue par produit tensoriel de deux B-splines définies sur l'intervalle  $[0, 1]$ . Dans ces travaux, il n'est toutefois jamais question de résumer une copule bidimensionnelle par ses  $q$  premiers coefficients sur une base donnée. Imaginons qu'il soit quand même décidé de mettre en place une décomposition de toutes les copules bidimensionnelles :

$$\forall 1 \leq i < j \leq d, \quad \hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}}) \approx \sum_{k=1}^q \alpha_{ijk}(\mathbf{U}_{\text{obs}}) g_k(\cdot) \quad . \quad (5.22)$$

Si une copule  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$  est connue de façon explicite, notamment sous une forme paramétrique  $\hat{c}_{ij}(\cdot | \mathcal{F}, \theta)$ , la décomposition peut être vue comme une complexification inutile du problème. En effet, l'incertitude portée par  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$  qui était initialement décrite par  $\hat{\theta} \in \mathbb{R}^r$  ( $r \leq 3$ ) est désormais décrite par un vecteur aléatoire  $q$ -dimensionnel. Par ailleurs, pour les copules bidimensionnelles connues de façon implicite, la décomposition peut s'avérer très coûteuse d'un point de vue calculatoire. Quelle que soit la stratégie employée, le calcul des coefficients sur la base nécessite d'être capable d'évaluer  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$  en n'importe quel point de  $\mathcal{H}_2$ . Si le coefficient  $\alpha_{ijk}(\mathbf{U}_{\text{obs}})$  résulte d'un produit scalaire sur  $\mathcal{H}_2$  entre  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$  et la fonction de base  $g_k(\cdot)$ , il faut être capable d'intégrer le produit de ces deux fonctions. Que l'intégration soit numérique ou stochastique, des centaines voire des milliers d'évaluations de  $\hat{c}_{ij}(\cdot | \mathcal{M}_{\text{cop}})$  sont alors nécessaires. Il en va de même pour les autres techniques de calcul des coefficients. Or, lorsqu'une copule bidimensionnelle  $\hat{c}_{ij}(\cdot | \mathcal{M}_{\text{cop}})$  est définie implicitement, on a vu qu'il est impossible de calculer  $\hat{c}_{ij}(\bar{u}_i, \bar{u}_j | \mathcal{M}_{\text{cop}})$  avec  $(\bar{u}_i, \bar{u}_j) \in \mathcal{H}_2$  sans le concours d'une approche de type Monte-Carlo. Le calcul du coefficient  $\alpha_{ijk}(\mathbf{U}_{\text{obs}})$  nécessite donc l'estimation par Monte-Carlo des valeurs prises par  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$  en plusieurs milliers de points de  $\mathcal{H}_2$ . Si on cherchait à tirer profit de la méthodologie

mise en place par [NANTY \[2015\]](#), il faudrait analyser la sensibilité de l'application suivante :

$$\Psi_{\alpha}: \mathbb{R}^q \times \dots \times \mathbb{R}^q \longrightarrow [0, 1] \\ \left\{ \alpha_{12k}(\mathbf{U}_{\text{obs}}) \right\}_{1 \leq k \leq q} \dots \left\{ \alpha_{(d-1)dk}(\mathbf{U}_{\text{obs}}) \right\}_{1 \leq k \leq q} \longrightarrow P_f(\mathbf{U}_{\text{obs}}) . \quad (5.23)$$

Pour une copule  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$  définie implicitement, le calcul de  $\alpha_{ij1}(\mathbf{U}_{\text{obs}})$  est très coûteux mais tous les autres coefficients  $\alpha_{ijk}(\mathbf{U}_{\text{obs}})$  avec  $2 \leq k \leq q$  peuvent en être déduits sans aucun surcoût en réutilisant intelligemment les points d'intégration. Par contre, ce travail doit être effectué pour toutes les copules bidimensionnelles définies implicitement et leur proportion augmente avec le nombre  $d$  de variables d'entrée du problème. Dans le cas d'une copule R-vigne, le rapport entre le nombre de copules bidimensionnelles définies implicitement et le nombre total de copules bidimensionnelles vaut :

$$\frac{\text{nb de paires de variables} - \text{nb d'arêtes dans l'arbre } \mathbf{T}_1}{\text{nb de paires de variables}} = \frac{C_d^2 - (d-1)}{C_d^2} = \frac{d-2}{d} \xrightarrow{d \rightarrow \infty} 1 . \quad (5.24)$$

En grande dimension, le nombre de copules bidimensionnelles dont la décomposition est coûteuse numériquement devient pratiquement égal au nombre total de copules bidimensionnelles. Par ailleurs, comme l'objectif est d'analyser la sensibilité de  $\Psi_{\alpha}$ , des centaines voire des milliers de décompositions fonctionnelles de toutes les entrées doivent pouvoir être exécutées. On comprend alors que le coût de calcul total de cette ASF est exorbitant, avant même d'avoir pris en compte le coût de calcul lié à l'utilisation de la boîte noire  $\phi$ .

Comme on ne peut pas envisager de répondre au problème initial à partir des résultats d'une ASF pratiquée sur  $\Psi_{\alpha}$ , une autre stratégie va être explorée dans la section suivante. On rappelle que l'ASF est particulièrement facile à mettre en œuvre dans le cas de l'estimation d'une  $d$ -copule gaussienne. Il est alors possible de remplacer chaque copule bidimensionnelle par son coefficient de corrélation. On aimerait se ramener à cette situation quel que soit le modèle  $\mathcal{M}_{\text{cop}}$  utilisé. Pour cela, il faut être capable d'extraire de la  $d$ -copule estimée  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  un résumé matriciel de sa structure de dépendance.

### 5.2.2.2 Choix d'un résumé matriciel de la structure de dépendance

Soit  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  la  $d$ -copule estimée à partir de l'échantillon  $\mathbf{U}_{\text{obs}}$ . Un résumé matriciel de cet objet fonctionnel est donné par :

$$\hat{\Delta}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}}) := \left[ \delta(V_i, V_j) \right]_{1 \leq i, j \leq d} \quad \text{avec} \quad \mathbf{V} \sim \hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}}) , \quad (5.25)$$

où  $\delta$  est une mesure de dépendance (au sens de la Définition 1.2.3) qui reste à choisir. De par (M2) et (M4), cette matrice est symétrique et ses coefficients diagonaux sont tous égaux à 1. Elle est donc assimilable à un vecteur aléatoire de dimension  $p = C_d^2$ . À la Section 1.2.7.3, on a pu constater l'existence de plusieurs mesures de dépendance. On choisit d'utiliser le tau de Kendall  $\tau_K$  parce qu'il joue un rôle central dans les algorithmes proposés par [BRECHMANN et collab. \[2012\]](#) puis par [DISSMANN et collab. \[2013\]](#) pour estimer des copules R-vignes en l'absence de connaissances préalables sur les arbres  $\mathbf{T}_{\text{vg}}$  et les familles  $\mathbf{F}_{\text{vg}}$ . L'équation (5.25) devient alors :

$$\hat{\tau}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}}) := \left[ \tau_K(V_i, V_j) \right]_{1 \leq i, j \leq d} \quad \text{avec} \quad \mathbf{V} \sim \hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}}) . \quad (5.26)$$

L'expression intégrale des coefficients non diagonaux fait intervenir les copules bidimensionnelles :

$$\hat{\tau}_{ij} := \tau_K(V_i, V_j) = 4 \int_{\mathcal{H}_2} \hat{C}_{ij}(u_i, u_j | \mathbf{U}_{\text{obs}}) \hat{c}_{ij}(u_i, u_j | \mathbf{U}_{\text{obs}}) du_i du_j - 1 . \quad (5.27)$$

Ici, il faut bien comprendre que  $\hat{\tau}_{ij}$  est un estimateur du tau de Kendall théorique  $\tau_{ij} := \tau_K(U_i, U_j)$  mais dans un sens assez inhabituel. En effet, l'intégration sur  $\mathcal{H}_2$  est supposée exacte mais elle

Famille $\mathcal{F}_{\text{cop}}$	Domaine $D_{\hat{\theta}}$	Coefficient $\hat{\tau}_{ij}$
Gumbel	$\hat{\theta} \in [1, +\infty[$	$\hat{\tau}_{ij} = \frac{\hat{\theta} - 1}{\hat{\theta}}$
Clayton	$\hat{\theta} \in ]0, +\infty[$	$\hat{\tau}_{ij} = \frac{\hat{\theta}}{\hat{\theta} + 2}$
Frank	$\hat{\theta} \in ]0, +\infty[$	$\hat{\tau}_{ij} = 1 - \frac{4}{\hat{\theta}} \left[ 1 - D_1(\hat{\theta}) \right]$
AMH	$\hat{\theta} \in [-1, 1[$	$\hat{\tau}_{ij} = \frac{3\hat{\theta} - 2}{3\hat{\theta}} - \frac{2(1 - \hat{\theta})^2}{3\hat{\theta}} \log(1 - \hat{\theta})$
Joe	$\hat{\theta} \in [1, +\infty[$	$\hat{\tau}_{ij} = 1 + \frac{4}{\hat{\theta}} J(\hat{\theta})$

$$\text{avec } D_n(\hat{\theta}) = \frac{n}{\hat{\theta}^n} \int_0^{\hat{\theta}} \frac{z^n}{e^z - 1} dz \quad \text{et} \quad J(\hat{\theta}) = \frac{1}{\hat{\theta}} \int_0^1 z \log(z) (1 - z)^{2(1-\hat{\theta})/\hat{\theta}} dz.$$

TABLEAU 5.1 – Lien entre les coefficients  $\hat{\tau}_{ij}$  et le paramètre  $\hat{\theta}$  pour différentes copules archimédiennes.

s'opère sur la fonction  $\mathbf{u} \mapsto \hat{C}_{ij}(u_i, u_j \mid \mathbf{U}_{\text{obs}}) \hat{c}_{ij}(u_i, u_j \mid \mathbf{U}_{\text{obs}})$  au lieu de s'opérer sur la fonction  $\mathbf{u} \mapsto C_{ij}(u_i, u_j) c_{ij}(u_i, u_j)$ . Le coefficient  $\hat{\tau}_{ij}$  est donc une caractéristique intrinsèque de la copule bidimensionnelle  $\hat{c}_{ij}(\cdot \mid \mathbf{U}_{\text{obs}})$  issue de l'apprentissage. Selon la nature de  $\mathcal{M}_{\text{cop}}$ , il peut être plus ou moins facile d'accéder aux coefficients  $\hat{\tau}_{ij}$  composant la matrice  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ . Dans le cas où  $\mathcal{M}_{\text{cop}}$  est assimilable à une famille paramétrique et où la densité de copule s'écrit sous la forme  $\hat{c}_{\mathbf{U}}(\cdot \mid \mathcal{F}_{\text{cop}}, \hat{\boldsymbol{\beta}})$ , il existe souvent une expression analytique des coefficients  $\hat{\tau}_{ij}$  en fonction de  $\hat{\boldsymbol{\beta}}$ . Dans le cas d'une  $d$ -copule gaussienne  $\hat{c}_{\mathcal{N}}(\cdot \mid \hat{\boldsymbol{\rho}})$ , on sait depuis [DANIELS et KENDALL \[1947\]](#) que :

$$\hat{\tau}_{ij} = \frac{2}{\pi} \arcsin(\hat{\rho}_{ij}). \quad (5.28)$$

Il est précisé dans [LINDSKOG et collab. \[2003\]](#) que ce résultat est valable pour toutes les copules elliptiques. Ainsi, la copule gaussienne  $\hat{c}_{\mathcal{N}}(\cdot \mid \hat{\boldsymbol{\rho}})$  et la copule de Student  $\hat{c}_t(\cdot \mid \hat{\boldsymbol{\rho}}, \hat{\nu})$  conduisent à la même matrice  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ . Pour les différentes familles de copules archimédiennes, les coefficients  $\hat{\tau}_{ij}$  peuvent être calculés à partir du paramètre scalaire  $\hat{\theta}$ . Plusieurs formules sont recensées dans le [Tableau 5.1](#) et les preuves sont disponibles dans [NELSEN \[2007\]](#). Dans le cas d'une copule R-vigne  $\hat{c}_{\text{vg}}(\cdot \mid \mathbf{T}_{\text{vg}}, \mathbf{F}_{\text{vg}}, \hat{\boldsymbol{\theta}}_{\text{vg}})$ , il est parfois possible d'accéder aux valeurs exactes des coefficients  $\hat{\tau}_{ij}$ , en particulier lorsque les deux conditions suivantes sont vérifiées :

- (a) la séquence d'arbres  $\mathbf{T}_{\text{vg}}$  est fixée,
- (b) les familles paramétriques  $\mathbf{F}_{\text{vg}}$  sont fixées et ne sont que des copules elliptiques.

La matrice  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  peut alors être calculée à partir des différents coefficients de corrélation linéaire impliqués dans le paramétrage des copules elliptiques. Ce résultat n'est pas immédiat et on peut se reporter aux ouvrages de [KUROWICKA et COOKE \[2006\]](#) et de [COOKE et collab. \[2010\]](#) pour de plus amples explications à ce sujet. Dans la suite de ce travail, on notera  $\mathcal{E}$  l'opérateur permettant d'obtenir la matrice  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  à partir de la  $d$ -copule  $\hat{c}_{\mathbf{U}}(\cdot \mid \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ .

Dès lors qu'on est dans une configuration d'apprentissage où l'opérateur  $\mathcal{E}$  est bien défini, l'accès à un résumé matriciel de  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  permet de simplifier le problème d'ASF initial. On cherche désormais à étudier le lien entre la variabilité des mesures de dépendance  $\hat{\tau}_{ij}$  et la variabilité de la probabilité de défaillance  $P_f(\mathbf{U}_{\text{obs}})$ . Le nouveau cadre d'ASF peut être formalisé comme suit :

$$\Psi_{\boldsymbol{\tau}} : \begin{matrix} [-1, 1] & \times & [-1, 1] & \times \dots \times & [-1, 1] & \longrightarrow & [0, 1] \\ \hat{\tau}_{12} & & \hat{\tau}_{13} & \dots & \hat{\tau}_{(d-1)d} & \longmapsto & P_f(\mathbf{U}_{\text{obs}}) \end{matrix}. \quad (5.29)$$

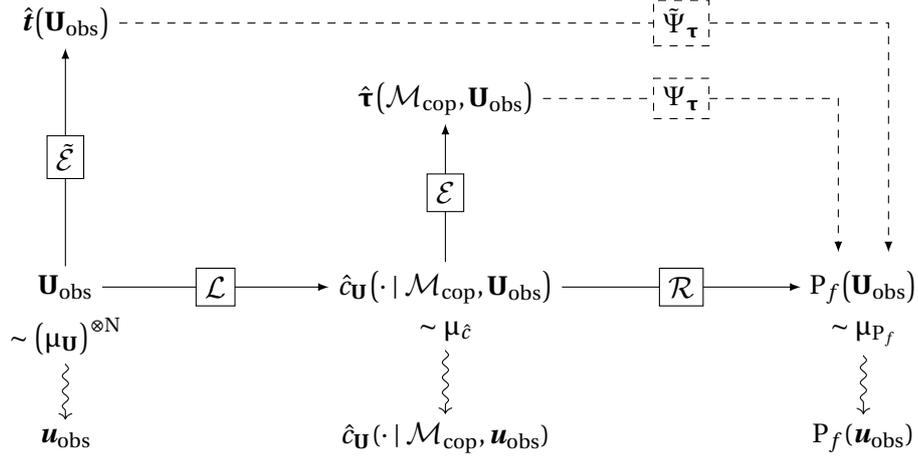


FIGURE 5.2 – Propagation de la variabilité d'échantillonnage via les coefficients de corrélation.

Comme on avait déjà pu le remarquer pour l'application  $\Psi$  définie à l'équation (5.4), la propagation des incertitudes est indirecte au sens où la sortie ne découle pas directement des entrées. À partir de la  $d$ -copule  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ , les opérateurs déterministes  $\mathcal{E}$  et  $\mathcal{R}$  permettent d'obtenir séparément la matrice  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  et la probabilité de défaillance  $P_f(\mathbf{U}_{\text{obs}})$ . Cette situation est représentée à la Figure 5.2 où l'appariement entrée-sortie engendré par  $\Psi_{\boldsymbol{\tau}}$  est représenté en pointillé. Il est facile de voir que  $\Psi_{\boldsymbol{\tau}}$  a un caractère stochastique car un même résumé matriciel peut être extrait de deux copules différentes. Avec le résultat de [LINDSKOG et collab. \[2003\]](#) sur les copules elliptiques, on sait par exemple que la  $d$ -copule gaussienne  $\hat{c}_{\mathcal{N}}(\cdot | \hat{\boldsymbol{\rho}})$  et la  $d$ -copule de Student  $\hat{c}_t(\cdot | \hat{\boldsymbol{\rho}}, \hat{\nu})$  engendrent la même matrice  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ . En effet, le nombre de degrés de liberté  $\hat{\nu}$  détermine la nature des queues de distribution multidimensionnelles tout en préservant les valeurs attribuées aux coefficients de corrélation de Kendall. Le paramètre  $\hat{\nu}$  est donc invisible du point de vue de l'opérateur  $\mathcal{E}$  mais il joue un rôle central lorsqu'on fait appel à l'algorithme de fiabilité  $\mathcal{R}$ . Cette stochasticité de  $\Psi_{\boldsymbol{\tau}}$  est entérinée par [TANG et collab. \[2013\]](#) qui montrent qu'il ne suffit pas de figer  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  pour éliminer l'incertitude accompagnant  $P_f(\mathbf{U}_{\text{obs}})$ . À corrélations fixées, l'incertitude liée au choix du modèle  $\mathcal{M}_{\text{cop}}$  peut avoir des conséquences importantes sur l'estimation du risque de défaillance et cette tendance s'accroît avec la rareté de l'évènement redouté. Un autre élément important dans la description de  $\Psi_{\boldsymbol{\tau}}$  est la dépendance existant entre les coefficients  $\hat{\tau}_{ij}$ . C'est une conséquence immédiate de la définition donnée à l'équation (5.26) à partir du vecteur  $\mathbf{V} \sim \hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ . On peut par exemple voir que :

- Les coefficients  $\hat{\tau}_{ij} = \tau_K(V_i, V_j)$  et  $\hat{\tau}_{jk} = \tau_K(V_j, V_k)$  sont dépendants car ils sont calculés à partir d'une même variable aléatoire uniforme  $V_j$ .
- Les coefficients  $\hat{\tau}_{ij} = \tau_K(V_i, V_j)$  et  $\hat{\tau}_{kl} = \tau_K(V_k, V_l)$  sont dépendants car  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  induit une dépendance entre les deux paires de variables.

Pratiquer une ASF sur  $\Psi_{\boldsymbol{\tau}}$  implique donc d'être capable d'appréhender les difficultés engendrées par la stochasticité et la dépendance.

### 5.2.2.3 Construction empirique d'un résumé matriciel de la structure de dépendance

Dans le cas général d'un modèle  $\mathcal{M}_{\text{cop}}$  conçu pour la grande dimension, il n'existe pas d'opérateur  $\mathcal{E}$  car il n'y a pas d'expression explicite de  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  en fonction des paramètres estimés. Les coefficients  $\hat{\tau}_{ij}$  peuvent être estimés par Monte-Carlo à condition d'être en mesure de :

- Simuler un grand nombre  $N_s$  d'échantillons  $\mathbf{V}^{(k)}$  selon la  $d$ -copule  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ .
- Évaluer les fonctions de répartition  $\hat{C}_{ij}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  des copules bidimensionnelles.

En exploitant l'expression intégrale (5.27), on a alors :

$$\forall 1 \leq i < j \leq d, \quad \hat{\tau}_{ij}^{\text{CMC}} = \frac{4}{N_s} \sum_{k=1}^{N_i} \hat{C}_{ij}(V_i^{(k)}, V_j^{(k)} | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}}) - 1 . \quad (5.30)$$

En général, les deux conditions ne sont pas réunies. Dans le cas de l'apprentissage d'une R-vigne, on est capable d'échantillonner suivant  $\hat{c}_{\mathbf{U}}(\cdot | \mathbf{U}_{\text{obs}})$  mais les fonctions  $\hat{C}_{ij}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  sont inconnues. Face à ce problème, il faut trouver une autre collection de variables scalaires permettant d'offrir un résumé de la structure de dépendance. Toutefois, on aimerait que cette collection puisse être calculée indépendamment du degré de sophistication de  $\mathcal{M}_{\text{cop}}$ . Cette condition est par exemple vérifiée par les estimateurs empiriques des coefficients de corrélation de Kendall. Ils ont l'avantage de pouvoir être calculés à partir de la seule connaissance de l'échantillon  $\mathbf{U}_{\text{obs}}$ . En vertu de la Définition 1.2.4, le coefficient  $\tau_{ij}$  peut s'écrire comme suit :

$$\tau_{ij} = \mathbb{P}((U_i - \tilde{U}_i)(U_j - \tilde{U}_j) > 0) - \mathbb{P}((U_i - \tilde{U}_i)(U_j - \tilde{U}_j) < 0) , \quad (5.31)$$

avec  $(U_i, U_j)$  et  $(\tilde{U}_i, \tilde{U}_j)$  deux copies indépendantes d'une paire de variables qui est distribuée suivant la copule bidimensionnelle  $c_{ij}$ . Un estimateur empirique de  $\tau_{ij}$  est alors donné par :

$$\hat{\tau}_{ij} = \frac{\text{nb de paires concordantes} - \text{nb de paires discordantes}}{\text{nb total de paires}} \quad (5.32)$$

$$= \frac{\text{Card}(\mathcal{C}_{\text{obs}}) - \text{Card}(\mathcal{D}_{\text{obs}})}{C_N^2} = \frac{4 \text{Card}(\mathcal{C}_{\text{obs}})}{N(N-1)} - 1 = 1 - \frac{4 \text{Card}(\mathcal{D}_{\text{obs}})}{N(N-1)} , \quad (5.33)$$

où les ensembles  $\mathcal{C}_{\text{obs}}$  et  $\mathcal{D}_{\text{obs}}$  sont construits à partir des observations des colonnes  $\mathbf{U}_{\text{obs}}^{[\bullet i]}$  et  $\mathbf{U}_{\text{obs}}^{[\bullet j]}$  :

$$\mathcal{C}_{\text{obs}} := \left\{ 1 \leq k < l \leq d : (U_i^{(k)} - U_i^{(l)})(U_j^{(k)} - U_j^{(l)}) > 0 \right\} , \quad (5.34)$$

$$\mathcal{D}_{\text{obs}} := \left\{ 1 \leq k < l \leq d : (U_i^{(k)} - U_i^{(l)})(U_j^{(k)} - U_j^{(l)}) < 0 \right\} . \quad (5.35)$$

En pratique, le calcul de  $\hat{\tau}_{ij}$  implique de prendre chaque paire  $(U_i^{(k)}, U_j^{(k)})$  avec  $1 \leq k \leq N-1$  et de la comparer à toutes les autres paires  $(U_i^{(l)}, U_j^{(l)})$  avec  $k < l \leq N$ . Cette approche naïve engendre un temps de calcul d'ordre  $\mathcal{O}(N^2)$ . Il est montré dans **KENDALL [1948]** qu'une autre façon de procéder peut consister à :

1. Construire une matrice  $\mathbf{U}_{\text{obs}}^{[\bullet(ij)]}$  en accolant les colonnes  $\mathbf{U}_{\text{obs}}^{[\bullet i]}$  et  $\mathbf{U}_{\text{obs}}^{[\bullet j]}$ .
2. Réorganiser les lignes de  $\mathbf{U}_{\text{obs}}^{[\bullet(ij)]}$  pour que les observations de  $U_i$  soient ordonnées :

$$U_i^{(1)} < U_i^{(2)} < \dots < U_i^{(N)} . \quad (5.36)$$

3. Compter le nombre  $s$  de permutations permettant de trier les observations de  $U_j$ .

On peut facilement voir que  $s = \text{Card}(\mathcal{D}_{\text{obs}})$  et la troisième version de l'équation (5.33) donne :

$$\hat{\tau}_{ij} = 1 - \frac{4s}{N(N-1)} . \quad (5.37)$$

La construction de  $\hat{\tau}_{ij}$  se résume donc au tri de deux listes. Afin de réduire le temps de calcul, **KNIGHT [1966]** propose d'utiliser des algorithmes de tri de complexité  $\mathcal{O}(N \log N)$ . Un état de l'art des routines de calcul permettant d'obtenir  $\hat{\tau}_{ij}$  est disponible dans **CHRISTENSEN [2005]**. Dans la suite, on note  $\tilde{\mathcal{E}}$  tout algorithme permettant de calculer les  $p := C_d^2 = \frac{1}{2}d(d-1)$  estimateurs empiriques  $\hat{\tau}_{ij}$ . On peut ainsi construire empiriquement un nouveau résumé matriciel :

$$\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}}) := [\hat{\tau}_{ij}]_{1 \leq i, j \leq d} . \quad (5.38)$$

Le remplacement de la matrice  $\hat{\tau}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  par la matrice  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$  modifie le problème d'ASF :

$$\begin{array}{ccccccc} \tilde{\Psi}_{\tau} : & [-1, 1] & \times & [-1, 1] & \times \dots \times & [-1, 1] & \longrightarrow & [0, 1] \\ & \hat{t}_{12} & & \hat{t}_{13} & \dots & \hat{t}_{(d-1)d} & \longmapsto & P_f(\mathbf{U}_{\text{obs}}) \end{array} \quad (5.39)$$

On peut se reporter à la Figure 5.2 pour une meilleure compréhension du mécanisme de propagation des incertitudes. À l'instar de  $\Psi_{\tau}$ , l'application  $\tilde{\Psi}_{\tau}$  est stochastique car  $P_f(\mathbf{U}_{\text{obs}})$  ne peut pas être calculée à partir de la matrice  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$ . On peut d'ailleurs remarquer que  $\tilde{\Psi}_{\tau}$  a un caractère stochastique plus marqué. Pour le justifier, on s'appuie sur la Figure 5.3 qui est obtenue en réorganisant la Figure 5.2 de manière à ce que  $\Psi_{\tau}$  soit vue comme une étape de  $\tilde{\Psi}_{\tau}$ . On a pu voir précédemment que la stochasticité de  $\Psi_{\tau}$  provient de la perte d'information liée à l'utilisation d'un résumé matriciel de  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ . Dans le cas de  $\tilde{\Psi}_{\tau}$ , une stochasticité supplémentaire est introduite par l'estimation empirique des coefficients de corrélation de Kendall. En effet, le modèle  $\mathcal{M}_{\text{cop}}$  n'est jamais conçu de manière ce que la matrice empirique  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$  devienne la future matrice de corrélation de Kendall de la  $d$ -copule  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ . On a donc :

$$\underbrace{\mathcal{E}\left(\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})\right)}_{\text{extraction analytique}} = \hat{\tau}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}}) \neq \underbrace{\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})}_{\text{extraction empirique}} = \tilde{\mathcal{E}}(\mathbf{U}_{\text{obs}}) \quad (5.40)$$

Si le choix du modèle d'apprentissage est pertinent au regard de la nature de  $c_{\mathbf{U}}$ , et si  $N$  n'est pas excessivement petit, les deux matrices sont proches mais elles ne sont pas pour autant égales. Le lien  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}}) \longmapsto \hat{\tau}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  est donc stochastique. On en veut pour preuve la situation se produisant en présence d'un échantillon  $\mathbf{u}_{\text{obs}}$  constitué de  $N$  observations d'une copule gaussienne inconnue que l'on cherche à reconstruire par MV. Avec l'algorithme de **IMAN et CONOVER [1982]**, on peut générer une séquence d'échantillons  $\{\tilde{\mathbf{u}}^{(k)}\}_{1 \leq k \leq K}$  partageant les mêmes coefficients empiriques de corrélation de Kendall :

$$\hat{\mathbf{t}}(\mathbf{u}_{\text{obs}}) = \hat{\mathbf{t}}(\tilde{\mathbf{u}}^{(1)}) = \hat{\mathbf{t}}(\tilde{\mathbf{u}}^{(2)}) = \dots = \hat{\mathbf{t}}(\tilde{\mathbf{u}}^{(K)}) \quad (5.41)$$

En revanche, lors de l'estimation par MV, chaque échantillon  $\tilde{\mathbf{u}}^{(k)}$  conduit à sa propre  $d$ -copule gaussienne  $\hat{c}_{\mathcal{N}}(\cdot | \hat{\rho}^{(k)})$ . Comme  $\mathcal{E}$  consiste à calculer  $\hat{\tau}(\mathcal{F}_{\mathcal{N}}, \tilde{\mathbf{u}}^{(k)})$  à partir de  $\hat{\rho}^{(k)}$  en appliquant la formule (5.28) à tous les coefficients  $\hat{\rho}_{ij}^{(k)}$ , on en déduit que les matrices  $\hat{\tau}(\mathcal{F}_{\mathcal{N}}, \tilde{\mathbf{u}}^{(k)})$  sont toutes différentes. On est donc dans une situation où une même matrice  $\hat{\mathbf{t}}(\mathbf{u}_{\text{obs}})$  correspond à une infinité de matrices  $\hat{\tau}(\mathcal{F}_{\mathcal{N}}, \tilde{\mathbf{u}}^{(k)})$ . Cela justifie le second niveau de stochasticité introduit par le remplacement des coefficients  $\hat{\tau}_{ij}$  par les coefficients  $\hat{t}_{ij}$ . Le passage de  $\Psi_{\tau}$  à  $\tilde{\Psi}_{\tau}$  amplifie donc la stochasticité du problème d'ASF mais les variables d'entrée restent dépendantes.

**CLIFF et CHARLIN [1991]** ont entrepris le calcul analytique de la matrice de covariance des coefficients de  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$ . Pour énoncer ces résultats, il faut adopter quelques nouvelles notations. Après avoir introduit la fonction "signe" :

$$\forall x \in \mathbb{R}, \quad \text{sgn}(x) = \begin{cases} 1 & \text{si } x > 0 \\ 0 & \text{si } x = 0 \\ -1 & \text{si } x < 0 \end{cases} \quad (5.42)$$

la définition (5.31) peut être réécrite sous la forme d'une espérance :

$$\begin{aligned} \tau_{ij} &= \mathbb{P}((U_i - \tilde{U}_i)(U_j - \tilde{U}_j) > 0) - \mathbb{P}((U_i - \tilde{U}_i)(U_j - \tilde{U}_j) < 0) \\ &= 1 \times \mathbb{P}(\text{sgn}(U_i - \tilde{U}_i) \times \text{sgn}(U_j - \tilde{U}_j) = 1) + (-1) \times \mathbb{P}(\text{sgn}(U_i - \tilde{U}_i) \times \text{sgn}(U_j - \tilde{U}_j) = -1) \end{aligned} \quad (5.43)$$

$$= \mathbb{E}[D_i D_j] = \mathbb{E}[T_{ij}], \quad (5.44)$$

avec  $D_k := \text{sgn}(U_k - \tilde{U}_k)$  pour  $k \in \{i, j\}$  et  $T_{ij} = D_i D_j$ . En utilisant ce formalisme, l'estimateur  $\hat{t}_{ij}$  défini à l'équation (5.33) peut s'écrire :

$$\hat{t}_{ij} = \frac{2}{N(N-1)} \sum_{k < l} T_{ij}^{(kl)} \quad \text{avec} \quad T_{ij}^{(kl)} = \text{sgn}(U_i^{(k)} - U_i^{(l)}) \times \text{sgn}(U_j^{(k)} - U_j^{(l)}) \quad (5.45)$$

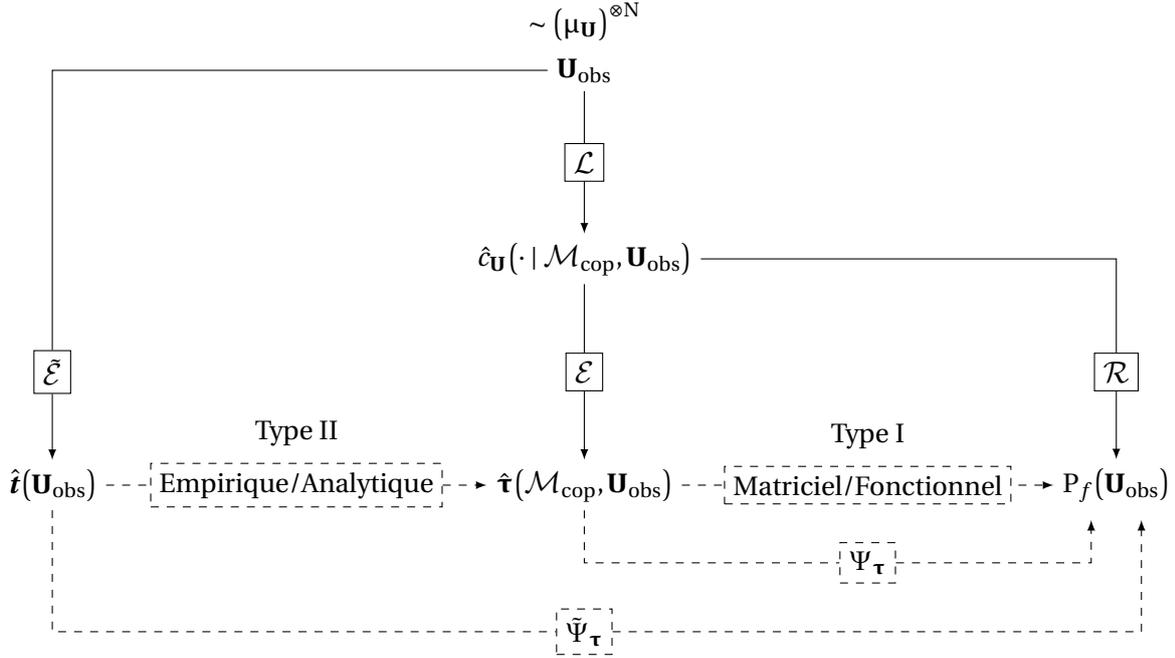


FIGURE 5.3 – Décomposition de la stochasticité intervenant dans l'application  $\tilde{\Psi}_\tau$ .

On a aussi besoin de définir la variable  $T_{ij}^{(r\bullet)}$  qui rend compte de la cohérence moyenne d'une paire d'observations  $(U_i^{(r)}, U_j^{(r)})$  par rapport à toutes les autres paires d'observations  $(U_i^{(s)}, U_j^{(s)})$  :

$$T_{ij}^{(r\bullet)} = \frac{1}{N-1} \sum_{s \neq r} T_{ij}^{(rs)}. \quad (5.46)$$

Pour  $1 \leq r \leq N$ , les variables  $T_{ij}^{(r\bullet)}$  ne sont pas égales mais elles sont toutes de même loi. Elles ont notamment une variance commune notée  $\bar{V}_{ij}$  :

$$\bar{V}_{ij} := \mathbb{V}(T_{ij}^{(1\bullet)}) = \mathbb{V}(T_{ij}^{(2\bullet)}) = \dots = \mathbb{V}(T_{ij}^{(N\bullet)}). \quad (5.47)$$

Le résultat de **CLIFF et CHARLIN [1991]** porte sur la matrice :

$$\Gamma(\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})) := \left[ \text{Cov}(\hat{t}_{ij}, \hat{t}_{kl}) \right]_{\substack{1 \leq i < j \leq N \\ 1 \leq k < l \leq N}} \in \mathbb{R}^{p \times p} \quad \text{avec} \quad p = C_d^2. \quad (5.48)$$

Les termes diagonaux sont les variances des coefficients  $\hat{t}_{ij}$  :

$$\mathbb{V}(\hat{t}_{ij}) = \frac{4(N-2)}{N(N-1)} \times \bar{V}_{ij} + \frac{2}{N(N-1)} \times (1 - \tau_{ij}^2). \quad (5.49)$$

Pour les termes non diagonaux, correspondant aux covariances entre les coefficients  $\hat{t}_{ij}$ , on distingue deux cas. Lorsque  $\{i, j\} \cap \{k, l\} = \emptyset$ , on a :

$$\text{Cov}(\hat{t}_{ij}, \hat{t}_{kl}) = \frac{4(N-2)}{N(N-1)} \times \text{Cov}(T_{ij}^{(r\bullet)}, T_{kl}^{(r\bullet)}) + \frac{2}{N(N-1)} \times (\mathbb{E}[T_{ij}T_{kl}] - \tau_{ij}\tau_{kl}). \quad (5.50)$$

Si l'un des indices est partagé par les deux paires d'indices, la formule précédente se simplifie :

$$\text{Cov}(\hat{t}_{ij}, \hat{t}_{jk}) = \frac{4(N-2)}{N(N-1)} \times \text{Cov}(T_{ij}^{(r\bullet)}, T_{jk}^{(r\bullet)}) + \frac{2}{N(N-1)} \times (\tau_{ik} - \tau_{ij}\tau_{kl}). \quad (5.51)$$

Ces formules prouvent l'existence d'une dépendance statistique entre les différents coefficients de la matrice  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$ . Elles sont utilisées dans **LONG et CLIFF [1997]** pour construire des intervalles

de confiance des corrélations de Kendall théoriques  $\tau_{ij}$ . On souhaite faire remarquer que les formules (5.49) à (5.51) ne sont valables que si  $\mathbf{U}_{\text{obs}}$  est un N-échantillon d'une loi  $\mu_{\mathbf{U}}$  absolument continue. En effet, on a alors :

$$\forall 1 \leq j \leq d, \quad \forall r \neq s, \quad \mathbb{P}\left(\text{sgn}(U_j^{(r)} - U_j^{(s)}) = 0\right) = \mathbb{P}(U_j^{(r)} = U_j^{(s)}) = 0, \quad (5.52)$$

ce qui implique que deux paires d'observations ne peuvent jamais être égales. Elles sont soit concordantes, soit discordantes. Cette situation conduit par exemple au résultat suivant :

$$\mathbb{V}(T_{ij}) = \mathbb{E}[T_{ij}^2] - (\mathbb{E}[T_{ij}])^2 = 1 - \tau_{ij}^2. \quad (5.53)$$

Dans notre contexte, cette hypothèse est vérifiée par toute matrice  $\mathbf{u}_{\text{obs}}$  simulée à l'aide du générateur théorique  $G_{\mathbf{U}}$ . Au contraire, si on manipule une matrice  $\tilde{\mathbf{u}}$  issue de la réplication *bootstrap* d'un échantillon initial  $\mathbf{u}_{\text{obs}}$ , l'utilisation de la mesure empirique  $\hat{\mu}[\mathbf{u}_{\text{obs}}]$  peut engendrer la création de doublons, c'est-à-dire de plusieurs occurrences d'une même ligne au sein de  $\tilde{\mathbf{u}}$ . Les coefficients  $\hat{t}_{ij}$  restent alors dépendants mais l'expression analytique de  $\Gamma(\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}}))$  doit être modifiée. Les formules adaptées à cette configuration sont disponibles dans l'ouvrage de CLIFF [2014].

Les deux applications stochastiques  $\Psi_{\tau}$  et  $\tilde{\Psi}_{\tau}$  sont construites pour simplifier le problème d'ASF initial où  $\Psi$  met en relation l'ensemble des copules bidimensionnelles  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$  et la probabilité de défaillance  $P_f(\mathbf{U}_{\text{obs}})$ . Dans les deux cas, la simplification s'appuie sur l'utilisation d'un résumé en dimension finie de la  $d$ -copule  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ . Celui-ci est assimilable à un vecteur aléatoire dont les variables sont dépendantes.  $\Psi_{\tau}$  est une première sous-version du problème d'ASF qui nécessite d'être capable d'extraire la matrice  $\hat{\mathbf{t}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  de la  $d$ -copule estimée. Cela limite son emploi à certains modèles d'apprentissage pour lesquels des résultats analytiques sont disponibles. Au contraire,  $\tilde{\Psi}_{\tau}$  s'articule autour de la matrice  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$  qui peut être obtenue indépendamment de  $\mathcal{M}_{\text{cop}}$ . Toutefois, cette seconde sous-version du problème d'ASF induit un surcroît de stochasticité. Dans la prochaine section, on met en place une procédure permettant d'analyser la sensibilité de  $\Psi_{\tau}$  et de  $\tilde{\Psi}_{\tau}$ . Pour ce qui touche à la propagation des incertitudes épistémiques et à la gestion du budget de simulation, de nombreux éléments sont empruntés au Chapitre 4 et sont alors mentionnés succinctement. L'objectif est davantage de se focaliser sur les difficultés propres au nouveau contexte d'ASF, à savoir le traitement de variables d'entrée dépendantes et la manipulation d'applications stochastiques.

### 5.3 Développement d'une procédure d'ASF pour le sous-problème

Afin de répondre clairement aux nouveaux enjeux de l'ASF, cette section est divisée en deux parties. La première est consacrée à la sélection d'indices de sensibilité adaptés aux entrées dépendantes. Cela implique aussi de choisir un schéma d'estimation pour les indices retenus et d'examiner les modalités de son intégration au sein de la procédure d'ASF. Dans l'esprit de ce qui a été fait au Chapitre 4, l'interfaçage des différentes briques de calcul permet alors l'implémentation d'un algorithme d'ASF dont la seule entrée est l'échantillon  $\mathbf{u}_{\text{obs}}$ . La seconde partie s'attèle à la question délicate de la mesure du niveau de stochasticité des applications  $\Psi_{\tau}$  et  $\tilde{\Psi}_{\tau}$ . La solution proposée repose sur la construction d'un indicateur de stochasticité pour lequel on propose un schéma d'estimation. Dans les deux parties, les raisonnements sont conduits pour  $\tilde{\Psi}_{\tau}$  car cette configuration ne souffre d'aucune restriction. Sauf contre-indication à ce sujet, tout ce qui est mis en place pour  $\tilde{\Psi}_{\tau}$  reste valable pour  $\Psi_{\tau}$ .

#### 5.3.1 Prise en compte de la dépendance entre les variables d'entrée de l'ASF

Afin de traiter les difficultés les unes après les autres, on procède comme dans le Chapitre 4 et on se restreint dans un premier temps à des contextes fiabilistes pour lesquels la probabilité de défaillance  $P_f$  peut être calculée à moindre coût par un algorithme  $\mathcal{R}$  déterministe. Entre la

Section 5.3.1.1 et la Section 5.3.1.2, une ébauche de l’algorithme d’ASF est ainsi développée sous cette hypothèse. Le contexte plus réaliste de l’utilisation d’un algorithme de simulation  $\mathcal{R}_w$  est abordé dans un second temps, en particulier à la Section 5.3.1.4 où la question de la gestion de la charge de calcul est remise à l’honneur.

### 5.3.1.1 Justification du choix des indices de sensibilité

On choisit de quantifier l’influence que chaque coefficient  $\hat{t}_{ij}$  exerce sur  $P_f(\mathbf{U}_{\text{obs}})$  à l’aide de la mesure d’importance de BORGONOVO [2007] qui a été présentée à la Section 3.2.3.2. Plusieurs raisons motivent cette décision :

- La construction théorique des  $\delta$ -indices de Borgonovo n’est pas assujettie à une hypothèse d’indépendance sur les variables d’entrée. Quel que soit le motif de dépendance existant entre les variables d’entrée, les  $\delta$ -indices sont définis rigoureusement. On peut alors se fier aux résultats des estimations qui en sont faites.
- Contrairement à la mesure d’importance de SOBOL [2001] qui quantifie l’influence des entrées à partir de la variance de la sortie, celle de BORGONOVO [2007] prend en compte l’intégralité de la distribution de la sortie.
- La littérature à ce sujet propose des schémas d’estimation aux propriétés complémentaires.
- Certains schémas d’estimation peuvent être mis en œuvre à partir de la seule connaissance d’un M-échantillon d’entrées et de sorties.

Évidemment, chacune de ces propriétés n’est pas spécifique aux  $\delta$ -indices. Toutefois, on apprécie le fait qu’elles soient toutes vérifiées par la même mesure d’importance.

La Section 3.2.3 a révélé l’existence de nombreuses méthodes d’AS conçues pour les variables dépendantes. Les procédures opérant un partage de la variance (décomposition hiérarchique dans CHASTAING et collab. [2012], analyse de la covariance dans SUDRET et CANIOU [2013], techniques d’orthogonalisation dans MARA et collab. [2015]) imposent que des conditions spécifiques soient vérifiées et mettent en jeu des indices qui peuvent être difficiles à interpréter ou qui ne quantifient pas directement l’influence des entrées. Lorsqu’on souhaite une compréhension plus fine de l’impact exercé sur la distribution de sortie, BORGONOVO et collab. [2016] évoquent plusieurs alternatives. On peut mentionner les  $\beta$ -indices de BAUCCELLS et BORGONOVO [2013] au même titre que les indices construits dans BORGONOVO et collab. [2014] à l’aide des divergences de Csiszár. Ces indices ne sont en fait que des variantes ou des généralisations des  $\delta$ -indices. Leur plus-value est alors discutable, d’autant que tous ne peuvent pas être estimés aussi facilement que les  $\delta$ -indices.

Un inconvénient majeur des  $\delta$ -indices réside dans le fait que seul leur schéma d’estimation en “double boucle” présente des garanties de convergence. Par ailleurs, la somme des  $\delta$ -indices ne vaut pas 1. Pour s’y ramener, on peut suivre la méthode de SARAZIN et collab. [2020] et calculer des valeurs de Shapley à partir des  $\delta$ -indices à tous les ordres. Toutes les propriétés souhaitées sont alors vérifiées mais des schémas d’estimation spécifiques sont nécessaires pour accéder aux  $\delta$ -indices d’ordre supérieur. Par simplicité, et parce que la propriété de sommation n’est en rien indispensable à la résolution d’un problème d’AS, cette méthode n’est pas implémentée.

Les différents arguments avancés jusqu’ici laissent penser que les  $\delta$ -indices semblent être la solution la plus attrayante pour analyser la sensibilité de  $P_f(\mathbf{U}_{\text{obs}})$  aux variations affectant les coefficients  $\hat{t}_{ij}$ . Dans un contexte d’ASF, l’idée d’utiliser ou de s’inspirer de la mesure de Borgonovo n’est pas incongrue. Les indices de sensibilité fiabilistes de CUI et collab. [2010] ont été conçus dans cet esprit. Plus récemment, DERENNES et collab. [2019] choisissent de calculer des  $\delta$ -indices conditionnellement à la réalisation d’une défaillance. Dans la suite, pour faciliter la lisibilité et la compréhension des formules, on adopte les notations suivantes :

$$\tilde{\mathbf{X}} := \left[ \tilde{X}_{ij} \right]_{1 \leq i < j \leq d} \quad \text{avec} \quad \tilde{X}_{ij} := \hat{t}_{ij} \quad \text{tandis que} \quad \tilde{\mathbf{Y}} := P_f(\mathbf{U}_{\text{obs}}) \quad . \quad (5.54)$$

Les indices de Borgonovo ont été présentés au Chapitre 3 dans le cadre de la propagation d'incertitudes aléatoires suivant un schéma  $\phi : \mathbf{X} \longrightarrow Y$ . Dans notre situation où des incertitudes épistémiques sont propagées suivant le schéma  $\tilde{\Psi}_\tau : \tilde{\mathbf{X}} \longrightarrow \tilde{Y}$ , rien ne change et on impose la similarité des notations pour une meilleure analogie. L'impact exercé par chaque coefficient  $\hat{t}_{ij}$  sur la probabilité  $P_f(\mathbf{U}_{\text{obs}})$  est alors quantifié par un indice  $\delta_{ij}^F$  que l'on définit comme :

$$\delta_{ij}^F := \delta\left(P_f(\mathbf{U}_{\text{obs}}) \mid \hat{t}_{ij}\right) = \delta\left(\tilde{Y} \mid \tilde{X}_{ij}\right) = \frac{1}{2} \mathbb{E}\left[\text{shift}(\tilde{X}_{ij})\right]. \quad (5.55)$$

Il s'agit de l'indice de Borgonovo du premier ordre de  $\tilde{X}_{ij}$  par rapport à  $\tilde{Y}$ . En tenant compte du support des différentes variables, la fonction  $\text{shift}(\cdot)$  vaut :

$$\forall x \in [-1, 1], \quad \text{shift}(x) = \left\| f_{\tilde{Y}}(\cdot) - f_{\tilde{Y}|\tilde{X}_{ij}}(\cdot | x) \right\|_{L^1} = \int_0^1 \left| f_{\tilde{Y}}(y) - f_{\tilde{Y}|\tilde{X}_{ij}}(y | x) \right| dy, \quad (5.56)$$

où  $\|\cdot\|_{L^1}$  représente la restriction de la norme  $L^1$  à l'intervalle  $[0, 1]$ . La définition (5.55) peut alors s'exprimer sous la forme d'une intégrale double :

$$\delta_{ij}^F := \frac{1}{2} \int_{-1}^1 \int_0^1 \left| f_{\tilde{Y}}(y) - f_{\tilde{Y}|\tilde{X}_{ij}}(y | x) \right| f_{\tilde{X}_{ij}}(x) dy dx, \quad (5.57)$$

qui dépend des densités marginales  $f_{\tilde{X}_{ij}}$  et  $f_{\tilde{Y}}$  ainsi que de la densité conditionnelle  $f_{\tilde{Y}|\tilde{X}_{ij}}$ . Pour que les indices  $\delta_{ij}^F$  puissent s'exprimer selon l'équation (5.57), toutes les densités mentionnées précédemment doivent exister. On sait que :

- (a)  $\mathbf{U}_{\text{obs}}$  est formé de  $N$  observations indépendantes d'une loi  $\mu_{\mathbf{U}}$  absolument continue.
- (b) Les entrées  $\tilde{\mathbf{X}}$  et la sortie  $\tilde{Y}$  sont calculées à partir de  $\mathbf{U}_{\text{obs}}$  :

$$\tilde{\mathbf{X}} = \hat{\mathbf{t}}(\mathbf{U}_{\text{obs}}) = \tilde{\mathcal{E}}(\mathbf{U}_{\text{obs}}) \quad \text{et} \quad \tilde{Y} = P_f(\mathbf{U}_{\text{obs}}) = [\mathcal{R} \circ \mathcal{L}](\mathbf{U}_{\text{obs}}). \quad (5.58)$$

Les lois  $\mu_{\tilde{\mathbf{X}}}$ ,  $\mu_{\tilde{Y}}$  et  $\mu_{\tilde{\mathbf{X}}\tilde{Y}}$  sont donc des images de la mesure d'échantillonnage  $(\mu_{\mathbf{U}})^{\otimes N}$ . Leur absolue continuité dépend de la régularité des opérateurs mis en jeu. On examine ci-après chacun des cas.

- Toutes les lois marginales  $\mu_{\tilde{X}_{ij}}$  sont discrètes. En effet,  $\tilde{X}_{ij}$  a été défini à l'équation (5.33) à partir du seul cardinal de l'ensemble  $\mathcal{C}_{\text{obs}}$ . Dans le formalisme de l'estimateur (5.45), c'est la présence de la fonction "signe" qui explique la transformation de la mesure d'échantillonnage  $(\mu_{\mathbf{U}})^{\otimes N}$  en une mesure discrète. La loi conjointe des entrées  $\mu_{\tilde{\mathbf{X}}}$  ainsi que les lois conditionnelles de type  $\mu_{\tilde{Y}|\tilde{X}_{ij}}$  ne sont donc pas absolument continues. Toutefois, il faut bien noter que la variable  $\tilde{X}_{ij}$  peut prendre  $C_N^2$  valeurs différentes. Même pour une petite taille d'échantillon ( $N \approx 50$ ), le nombre d'atomes de la loi discrète  $\mu_{\tilde{X}_{ij}}$  demeure conséquent ( $C_N^2 > 10^3$ ). Ainsi, l'histogramme d'une séquence de réalisations de l'entrée  $\tilde{X}_{ij}$  peut être raisonnablement assimilé à l'histogramme d'une loi à densité.
- En ce qui concerne la loi  $\mu_{\tilde{Y}}$ , la situation est plus compliquée. L'opérateur  $\mathcal{R}$  représentant l'intégration sur le domaine  $D_f$  est continu. Toutefois, la continuité de  $\mathcal{L}$  n'est pas toujours assurée et dépend de la nature du modèle  $\mathcal{M}_{\text{cop}}$ . Pour le cas test linéaire gaussien utilisé à la Section 4.3.1, l'estimation par MV d'une copule gaussienne  $\hat{\mathcal{C}}_{\mathcal{N}}(\cdot | \hat{\boldsymbol{\rho}})$  garantit l'absolue continuité de la mesure  $\mu_{\tilde{Y}}$ .

Pour les entrées  $\tilde{X}_{ij}$  comme pour la sortie  $\tilde{Y}$ , on ne peut pas affirmer catégoriquement que les lois sous-jacentes sont absolument continues. Dans la suite, les échantillons simulés à partir de ces lois sont toutefois traités comme si c'était effectivement le cas.

### 5.3.1.2 Sélection d'un schéma d'estimation des indices de sensibilité

Après avoir défini des indices de sensibilité, il faut sélectionner un schéma d'estimation pouvant être adapté à la structure de l'application stochastique  $\tilde{\Psi}_\tau$ . En effet, il s'agit d'une configuration inhabituelle pour laquelle la propagation des incertitudes épistémiques est indirecte. La

connaissance des entrées  $\tilde{X}_{ij}$  n'est alors pas suffisante pour calculer la sortie  $\tilde{Y}$ . L'objectif est donc de déployer un schéma d'estimation s'accordant avec cette contrainte. Pour l'instant, on suppose que l'on dispose d'un générateur  $G_{\mathbf{U}}$  de la copule. On est ainsi capable de simuler de nouvelles réalisations de la matrice  $\mathbf{U}_{\text{obs}}$ . Le cas d'un échantillon unique  $\mathbf{u}_{\text{obs}}$  est traité à la Section 5.3.1.3 où l'on discute notamment de l'utilisation du *bootstrap* au sein du schéma d'estimation des indices. Comme indiqué dans BORGONOV0 [2007], l'estimateur le plus naturel de  $\delta_{ij}^F$  est obtenu à partir de la définition (5.55) en suivant l'approche dite de la "double boucle" :

$$\hat{\delta}_{ij}^F = \frac{1}{2N_s} \sum_{k=1}^{N_s} \left\| \hat{f}_{\tilde{Y}}(\cdot) - \hat{f}_{\tilde{Y}|\tilde{X}_{ij}}(\cdot | Z^{(k)}) \right\|_{L^1} \quad \text{avec } Z^{(k)} \stackrel{\text{i.i.d.}}{\sim} f_{\tilde{X}_{ij}}. \quad (5.59)$$

Il est bon de remarquer que cette méthode impose l'estimation de la densité de sortie  $f_{\tilde{Y}}$  ainsi que celle des différentes densités conditionnelles  $f_{\tilde{Y}|\tilde{X}_{ij}}(\cdot | Z^{(k)})$ . Pour la première, il suffit de :

1. Simuler une séquence d'échantillons  $\{\mathbf{U}_{\text{obs}}^{(k)}\}_{k=1}^M$  à l'aide du générateur théorique  $G_{\mathbf{U}}$ .
2. Lancer successivement l'algorithme d'apprentissage  $\mathcal{L}$  puis l'algorithme de fiabilité  $\mathcal{R}$  pour calculer les différentes sorties  $\tilde{Y}^{(k)} = P_f(\mathbf{U}_{\text{obs}}^{(k)})$ .
3. Estimer  $f_{\tilde{Y}}$  à partir des échantillons  $\{\tilde{Y}^{(k)}\}_{k=1}^M$ .

Pour les densités conditionnelles  $f_{\tilde{Y}|\tilde{X}_{ij}}(\cdot | Z^{(k)})$ , la situation n'est pas aussi simple. Comme on ne dispose d'aucune information sur la distribution de  $P_f(\mathbf{U}_{\text{obs}})$  conditionnellement au fait que le coefficient  $\hat{t}_{ij}$  prenne une valeur donnée  $z \in [-1, 1]$ , on ne sait pas simuler des échantillons de la densité  $f_{\tilde{Y}|\tilde{X}_{ij}}(\cdot | z)$  et l'estimation d'une telle densité est alors impossible. On ne peut donc pas utiliser l'approche en "double boucle" pour l'estimation des indices  $\delta_{ij}^F$ . La méthode de la "simple boucle" est initiée dans WEI et collab. [2013] et s'inspire de la réécriture de l'équation (5.57) à l'aide du théorème de Bayes :

$$\delta_{ij}^F = \frac{1}{2} \int_{-1}^1 \int_0^1 \left| f_{\tilde{X}_{ij}}(x) f_{\tilde{Y}}(y) - f_{\tilde{X}_{ij}\tilde{Y}}(x, y) \right| dy dx. \quad (5.60)$$

Les indices  $\delta_{ij}^F$  peuvent alors être estimés en exécutant la procédure suivante :

- (Bg1) Simuler une séquence d'échantillons  $\{\mathbf{U}_{\text{obs}}^{(k)}\}_{k=1}^M$  à l'aide du générateur théorique  $G_{\mathbf{U}}$ .
- (Bg2) Calculer les matrices  $\hat{\mathbf{t}}^{(k)} := \hat{\mathbf{t}}(\mathbf{U}_{\text{obs}}^{(k)})$  avec l'algorithme  $\tilde{\mathcal{E}}$ .
- (Bg3) Calculer les probabilités  $P_f^{(k)} := P_f(\mathbf{U}_{\text{obs}}^{(k)})$  en appliquant successivement l'algorithme d'apprentissage  $\mathcal{L}$  puis l'algorithme de fiabilité  $\mathcal{R}$ .
- (Bg4) Estimer  $f_{\tilde{Y}}$  à partir des échantillons  $\{\tilde{Y}^{(k)}\}_{k=1}^M$ .
- (Bg5) Pour chaque paire d'entiers vérifiant  $1 \leq i < j \leq d$  :
  - (Bg5a) Estimer  $f_{\tilde{X}_{ij}}$  à partir des échantillons  $\{\tilde{X}_{ij}^{(k)}\}_{k=1}^M$ .
  - (Bg5b) Estimer  $f_{\tilde{X}_{ij}\tilde{Y}}$  à partir des échantillons  $\{(\tilde{X}_{ij}^{(k)}, \tilde{Y}^{(k)})\}_{k=1}^M$ .
  - (Bg5c) Estimer  $\delta_{ij}^F$  par échantillonnage préférentiel avec  $N_s$  points :

$$\hat{\delta}_{ij}^F = \frac{1}{2N_s} \sum_{k=1}^{N_s} \left| \frac{\hat{f}_{\tilde{X}_{ij}}(Z_1^{(k)}) \hat{f}_{\tilde{Y}}(Z_2^{(k)})}{\hat{f}_{\tilde{X}_{ij}\tilde{Y}}(Z_1^{(k)}, Z_2^{(k)})} - 1 \right| \quad \text{avec } (Z_1^{(k)}, Z_2^{(k)}) \stackrel{\text{i.i.d.}}{\sim} \hat{f}_{\tilde{X}_{ij}\tilde{Y}}. \quad (5.61)$$

Rien n'empêche l'application directe de ce schéma d'estimation. Toutefois, on peut remarquer que le support de la densité conjointe  $f_{\tilde{X}_{ij}\tilde{Y}}$  se restreint au pavé  $[-1, 1] \times [0, 1]$ . Les densités marginales  $f_{\tilde{X}_{ij}}$  et  $f_{\tilde{Y}}$  ont également un support compact. Par conséquent, les étapes (Bg4), (Bg5a) et (Bg5b) doivent être conduites avec des techniques spécifiques afin d'éviter les effets de bord. Les solutions proposées dans SCHUSTER [1985] ou SILVERMAN [1986] peuvent être implémentées sans

difficulté particulière si on souhaite à tout prix utiliser le schéma en “simple boucle”. Toutefois, pour estimer les indices  $\delta_{ij}^F$ , il semble beaucoup plus simple d'utiliser l'approche développée par [WEI et collab. \[2014\]](#) et reposant sur l'estimation de la copule reliant les variables  $\tilde{X}_{ij}$  et  $\tilde{Y}$ . Pour cela, on introduit les notations suivantes :

$$\tilde{\mathbf{R}} := \left[ \tilde{R}_{ij} \right]_{1 \leq i < j \leq d} \quad \text{avec} \quad \tilde{R}_{ij} := F_{\tilde{X}_{ij}}(\tilde{X}_{ij}) \quad \text{tandis que} \quad \tilde{S} := F_{\tilde{Y}}(\tilde{Y}) . \quad (5.62)$$

Un simple changement de variables dans l'équation (5.60) permet d'exprimer  $\delta_{ij}^F$  à partir de la seule copule bidimensionnelle  $c_{\tilde{\mathbf{R}}_{ij}\tilde{S}}$  :

$$\delta_{ij}^F := \frac{1}{2} \int_{\mathcal{H}_2} \left| c_{\tilde{\mathbf{R}}_{ij}\tilde{S}}(u, v) - 1 \right| du dv . \quad (5.63)$$

Un schéma d'estimation “orienté copule” est obtenu en remplaçant les étapes (Bg4) et (Bg5) par :

(Bg4) Fabriquer des pseudo-observations de la copule  $c_{\tilde{\mathbf{R}}\tilde{S}} \in \mathcal{C}_{p+1}$  :

$$\forall i < j, \quad \tilde{R}_{ij}^{(k)} = \frac{1}{M+1} \sum_{l=1}^M \mathbb{1}_{\{\hat{r}_{ij}^{(l)} \leq \hat{r}_{ij}^{(k)}\}} \quad \text{et} \quad \tilde{S}^{(k)} = \frac{1}{M+1} \sum_{l=1}^M \mathbb{1}_{\{P_f^{(l)} \leq P_f^{(k)}\}} . \quad (5.64)$$

(Bg5) Pour chaque paire d'indices telle que  $i < j$  :

(Bg5a) Estimer  $c_{\tilde{\mathbf{R}}_{ij}\tilde{S}}$  à partir des données  $\left\{ (\tilde{R}_{ij}^{(k)}, \tilde{S}^{(k)}) \right\}_{k=1}^M$  .

(Bg5b) Estimer  $\delta_{ij}^F$  par Monte-Carlo :

$$\hat{\delta}_{ij}^F = \frac{1}{2N_s} \sum_{k=1}^{N_s} \left| \hat{c}_{\tilde{\mathbf{R}}_{ij}\tilde{S}}(W_1^{(k)}, W_2^{(k)}) - 1 \right| \quad \text{avec} \quad (W_1^{(k)}, W_2^{(k)}) \stackrel{\text{i.i.d.}}{\sim} \mathcal{U}([0, 1])^{\otimes 2} . \quad (5.65)$$

D'un point de vue inférentiel, l'estimation des  $p$  indices de sensibilité  $\delta_{ij}^F$  est équivalente à l'estimation de  $p$  copules bidimensionnelles. À titre de comparaison, la méthode en “simple boucle” impose l'estimation de  $2p + 1$  densités à support compact. En outre, on peut considérer que l'estimation des  $p$  densités conjointes présente un degré de difficulté tout à fait comparable à celui de l'estimation de  $p$  copules bidimensionnelles. Le schéma basé sur l'estimation des copules est donc plus efficace en termes de mise en œuvre numérique. Dans [WEI et collab. \[2014\]](#), l'étape (Bg5a) est effectuée à l'aide d'une méthode KDE pour les copules bidimensionnelles. Toute autre technique d'estimation non-paramétrique d'une densité de copule est également envisageable. Le lecteur peut se reporter à [NAGLER et collab. \[2017\]](#) pour une présentation des alternatives possibles. Pour l'algorithme d'ASE, on choisit d'utiliser la méthode des noyaux transformés, et plus particulièrement une version améliorée développée par [GEENENS et collab. \[2017\]](#). Le schéma d'estimation complet est noté  $\mathcal{A}$  dans la suite du chapitre. Au niveau de la charge de calcul, l'instruction (Bg3) mérite une attention particulière étant donné qu'elle implique le calcul de  $M$  probabilités de défaillance. Au contraire, les instructions (Bg4) et (Bg5) peuvent être vues comme un simple post-traitement consécutif à l'obtention d'un  $M$ -échantillon d'entrées et de sorties de  $\tilde{\Psi}_{\mathbf{r}}$ . Par ailleurs, sachant que le générateur  $G_{\mathbf{U}}$  est inaccessible en pratique, la prochaine section s'attèle à montrer que le *bootstrap* permet d'exécuter (Bg1) à partir de la seule connaissance de  $\mathbf{u}_{\text{obs}}$ .

### 5.3.1.3 Assemblage des différentes briques de calcul

Quand on ne dispose d'aucune information sur  $\mu_{\mathbf{U}}$ , il faut construire une approximation  $\hat{\mu}_{\mathbf{U}}$  à partir des données disponibles. À l'instar de ce qui est entrepris au Chapitre 4, on choisit d'utiliser la mesure empirique définie par :

$$\hat{\mu}_{\mathbf{U}} := \hat{\mu}[\mathbf{u}_{\text{obs}}] = \frac{1}{N} \sum_{i=1}^N \delta_{\mathbf{u}_{\text{obs}}^{[i*]}} , \quad (5.66)$$

où  $\delta_{\mathbf{u}}$  est la mesure de Dirac d'un point  $\mathbf{u} \in \mathcal{H}_d$ . Le mécanisme de ré-échantillonnage est trivial car il s'agit simplement de fabriquer des répliques *bootstrap* de  $\mathbf{u}_{\text{obs}}$ . Dans le Chapitre 4, on cherchait à ce que les densités marginales  $\hat{f}_{X_j}(\cdot | \mathcal{M}_j, \tilde{\mathbf{Z}}_j)$  et la densité de copule  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \tilde{\mathbf{V}})$  issues d'un ré-échantillonnage soient des objets mutuellement indépendants. Ici, l'objectif visé consiste uniquement à fabriquer de nouvelles réalisations de  $\mathbf{U}_{\text{obs}}$ . Il n'est pas question de modifier la structure de dépendance existant entre les coefficients empiriques  $\hat{t}_{ij}$  de corrélation de Kendall. Par ailleurs, il faut remarquer que le remplacement de la mesure théorique  $\mu_{\mathbf{U}}$  par la mesure empirique  $\hat{\mu}_{\mathbf{U}}$  perturbe le cadre mathématique établi par l'équation (5.58) car les entrées  $\tilde{\mathbf{X}}$  et la sortie  $\tilde{\mathbf{Y}}$  sont désormais calculées comme suit :

$$\tilde{\mathbf{X}} = \tilde{\mathcal{E}}(\tilde{\mathbf{U}}) \quad \text{et} \quad \tilde{\mathbf{Y}} = [\mathcal{R} \circ \mathcal{L}](\tilde{\mathbf{U}}) \quad \text{avec} \quad \tilde{\mathbf{U}} \sim (\hat{\mu}_{\mathbf{U}})^{\otimes N} . \quad (5.67)$$

On a déjà discuté des difficultés émanant du manque de régularité des opérateurs  $\tilde{\mathcal{E}}$  et  $\mathcal{R} \circ \mathcal{L}$ . Une difficulté supplémentaire vient du fait que  $(\hat{\mu}_{\mathbf{U}})^{\otimes N}$  est une mesure discrète. Toute matrice  $\tilde{\mathbf{u}}$  créée par *bootstrap* de l'échantillon  $\mathbf{u}_{\text{obs}}$  appartient à l'ensemble :

$$\mathbf{M}_{\mathcal{B}}(\mathbf{u}_{\text{obs}}) := \left\{ \tilde{\mathbf{u}} \in [0, 1]^{N \times d} : \forall 1 \leq i \leq N, \exists 1 \leq k_i \leq N, \tilde{\mathbf{u}}^{[i \bullet]} = \mathbf{u}_{\text{obs}}^{[k_i \bullet]} \right\} . \quad (5.68)$$

Chaque matrice  $\tilde{\mathbf{u}}$  est alors caractérisée par :

1. Le nombre  $r$  de lignes distinctes qui la constituent. On les retrouve dans  $\mathbf{u}_{\text{obs}}$  en s'aidant d'une liste d'indices  $\{l_1, \dots, l_r\}$ . À  $r$  fixé, on peut constituer  $C_N^r$  listes différentes.
2. Le nombre  $n_i$  d'occurrences de chaque ligne  $\mathbf{u}_{\text{obs}}^{[l_i \bullet]}$  au sein de  $\tilde{\mathbf{u}}$ . À  $r$  fixé, compter le nombre de listes  $\{n_1, \dots, n_r\}$  revient à calculer le cardinal de l'ensemble :

$$\mathbf{A}_r := \left\{ 1 \leq n_1, \dots, n_r \leq N : \sum_{j=1}^r n_j = N \right\} . \quad (5.69)$$

Par conséquent, le nombre de matrices  $\tilde{\mathbf{u}}$  pouvant être simulées par  $(\hat{\mu}_{\mathbf{U}})^{\otimes N}$  est donné par :

$$\mathbf{N}_{\mathcal{B}} := \text{Card}(\mathbf{M}_{\mathcal{B}}(\mathbf{u}_{\text{obs}})) = \sum_{r=1}^N C_N^r \times \text{Card}(\mathbf{A}_r) \gg \sum_{r=1}^N C_N^r = 2^N - 1 . \quad (5.70)$$

Même pour une petite taille d'échantillon ( $N \approx 50$ ), le nombre de matrices pouvant être formées par *bootstrap* à partir de  $\mathbf{u}_{\text{obs}}$  est exorbitant ( $\mathbf{N}_{\mathcal{B}} > 10^{15}$ ) et il dépasse très nettement le nombre de ré-échantillonnages qu'on prévoit d'effectuer au sein d'un schéma numérique destiné à l'estimation d'indices de sensibilité. Si l'utilisation du *bootstrap* induit une discrétisation des lois impliquées dans l'expression (5.57) des indices  $\hat{\delta}_{ij}^F$ , les atomes sont bien trop nombreux pour que des outils adaptés au traitement des lois discrètes puissent être déployés. Étant donné que la probabilité d'occurrence de doublons est faible, le parti pris est de travailler comme si toutes les lois étaient absolument continues.

Pour coller aux notations du Chapitre 4, la phase de ré-échantillonnage est notée  $\mathcal{B}$ . Elle est implantée à la source d'un processus de propagation des incertitudes épistémiques :

$$\begin{array}{ccccccc} \mathbf{u}_{\text{obs}} & \xrightarrow{\mathcal{B}} & \tilde{\mathbf{u}} & \xrightarrow{\mathcal{L}} & \hat{c}_{\mathbf{U}}(\cdot | \tilde{\mathbf{u}}) & \xrightarrow{\text{Sklar}} & \hat{f}_{\mathbf{X}}(\cdot | \tilde{\mathbf{u}}) & \xrightarrow{\mathcal{R}} & \mathbf{P}_f(\tilde{\mathbf{u}}) \\ & & \tilde{\mathbf{u}} & \xrightarrow{\tilde{\mathcal{E}}} & \hat{\mathbf{t}}(\tilde{\mathbf{u}}) & & & & \end{array} . \quad (5.71)$$

Une schématisation du *workflow* résultant de l'interfaçage des différentes briques de calcul est représentée à la Figure 5.4 où apparaît également l'algorithme  $\mathcal{A}$  d'estimation des  $\delta$ -indices. En répétant la séquence algorithmique décrite par l'équation (5.71), on parvient à construire un échantillon associant des entrées  $\tilde{\mathbf{X}}^{(k)}$  et des sorties  $\tilde{\mathbf{Y}}^{(k)}$  de l'application  $\tilde{\Psi}_{\mathbf{T}}$ . Pour alimenter correctement  $\mathcal{A}$ , il faut fournir un nombre suffisant d'associations  $(\tilde{\mathbf{X}}^{(k)}, \tilde{\mathbf{Y}}^{(k)})$ . Pour cela, chacune des tâches  $\hat{f}_{\mathbf{X}}(\cdot | \tilde{\mathbf{u}}) \rightarrow \mathbf{P}_f(\tilde{\mathbf{u}})$  doit pouvoir être exécutée avec un algorithme de fiabilité déterministe  $\mathcal{R}$  dont le coût de calcul est faible. Lorsque le modèle d'apprentissage  $\mathcal{M}_{\text{cop}}$  ou la fonction de performance  $\phi$  ne le permettent pas, on doit se résoudre à utiliser un algorithme de simulation  $\mathcal{R}_w$  capable d'appréhender la configuration examinée. La maîtrise du budget de simulation devient alors une contrainte majeure.

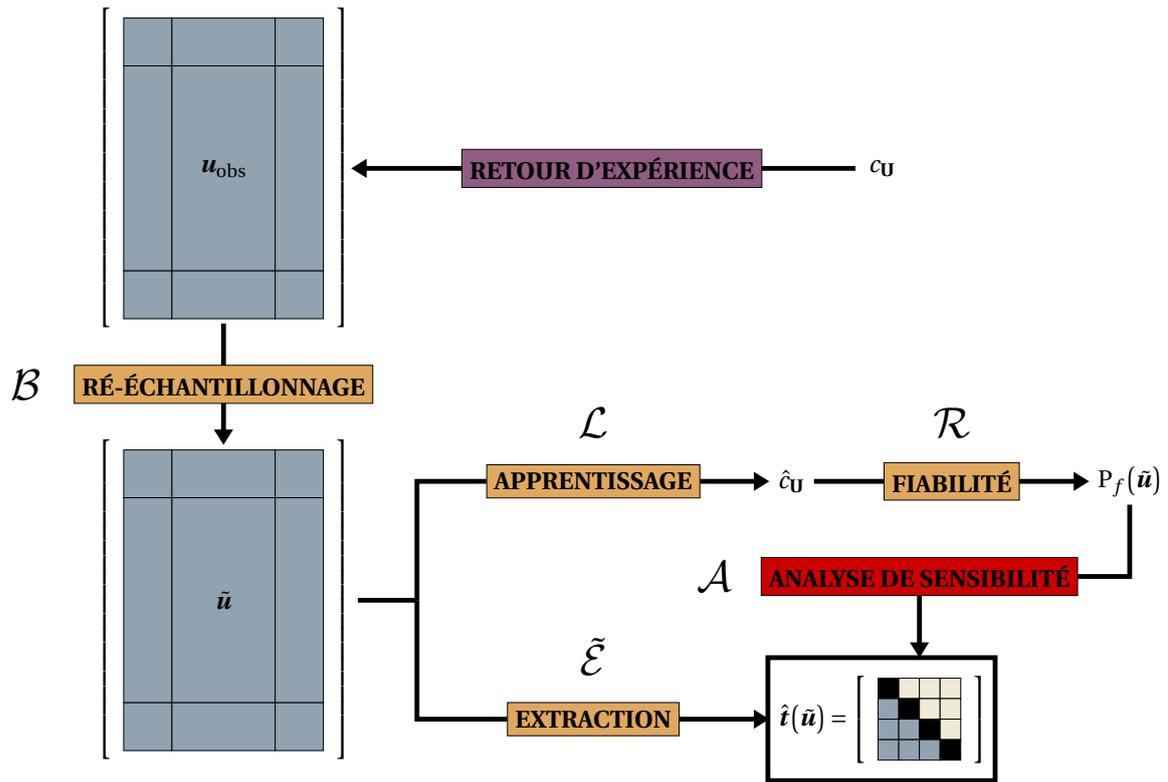
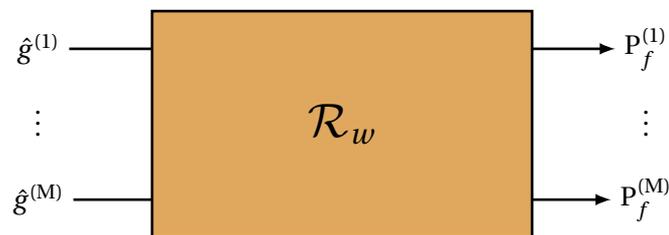


FIGURE 5.4 – Illustration des différentes briques de calcul intervenant dans l’algorithme d’ASF.

### 5.3.1.4 Mise en place d’une stratégie de gestion du budget de simulation

De par l’utilisation d’un algorithme de simulation  $\mathcal{R}_w$ , la valeur exacte de chaque probabilité  $P_f^{(k)}$  impliquée dans l’ASF est désormais remplacée par une estimation  $\hat{P}_f^{(k)}$ . En se basant sur les conclusions de la Section 4.5.1, on suppose que  $\mathcal{R}_w$  est réglé de manière à ce que sa stochasticité intrinsèque n’impacte pas l’estimation des indices  $\delta_{ij}^F$ . Cela revient à supposer que les lois des variables  $P_f(\mathbf{U}_{\text{obs}})$  et  $\hat{P}_f(\mathbf{U}_{\text{obs}})$  sont pratiquement identiques.

La procédure d’ASF développée pour  $\tilde{\Psi}_\tau$  ressemble sous de nombreux aspects à celle développée dans le Chapitre 4. La charge de calcul est sensiblement de la même nature. À l’instar de ce qui est proposé à la Figure 4.16, elle peut être résumée par le bloc de tâches  $\hat{g}^{(k)} \rightarrow P_f^{(k)}$ .



Toutefois, on peut noter deux différences importantes :

- Pour les indices de Sobol utilisés au Chapitre 4, si on décide d’effectuer  $M$  itérations du schéma *Pick-Freeze*, il faut estimer  $M_t = (d + 2) \times M$  probabilités de défaillance. Ici, au regard du schéma  $\mathcal{A}$  retenu à la Section 5.3.1.2, le nombre de probabilités de défaillance à estimer est égal au nombre  $M$  d’itérations.
- Les densités conjointes  $\hat{g}^{(k)}$  sont issues du ré-assemblage de la copule estimée  $\hat{c}_U(\cdot | \tilde{\mathbf{u}})$  et des lois marginales définies à l’avance comme des constantes du problème :

$$\forall 1 \leq k \leq M, \quad \hat{g}^{(k)}(\mathbf{x}) = \left[ \prod_{j=1}^d f_{X_j}(x_j) \right] \times \hat{c}_U(F_1(x_1), \dots, F_d(x_d) | \tilde{\mathbf{u}}^{(k)}) . \quad (5.72)$$

Dans ce contexte, les solutions à mettre en œuvre sont exactement les mêmes que celles présentées à la Section 4.4 et on renvoie notamment à la Figure 4.17 où plusieurs stratégies basées sur l'échantillonnage préférentiel sont proposées. Il faut commencer par sélectionner une méthode d'échantillonnage préférentiel compatible avec le problème traité. On est alors capable de construire une densité auxiliaire sous-optimale  $\hat{h}_{\text{opt}}^{(k)}$  pour toute densité d'entrée  $\hat{g}^{(k)}$  impliquée dans l'ASF. Une approche séquentielle étant trop coûteuse, des stratégies telle que l'astuce "inverse-reverse" ou l'échantillonnage préférentiel multiple (MIS) doivent être mises en place pour optimiser la dépense du budget de simulation. On rappelle que l'approche MIS consiste à :

1. Exécuter  $N_e$  fois le code de simulation  $[\mathcal{L} \circ \mathcal{B}](\mathbf{u}_{\text{obs}})$  :
  - ⇒ Fabrication d'une série de  $N_e$  densités de copule  $\hat{c}_{\mathbf{U}}(\cdot | \tilde{\mathbf{u}}_0^{(1)}), \dots, \hat{c}_{\mathbf{U}}(\cdot | \tilde{\mathbf{u}}_0^{(N_e)})$ .
  - ⇒ Transformation en une série de  $N_e$  densités d'entrée  $\hat{g}_0^{(1)}, \dots, \hat{g}_0^{(N_e)}$ .
2. Lancer une phase d'exploration pour chaque densité  $\hat{g}_0^{(l)}$  :
  - ⇒ Fabrication d'une densité sous-optimale  $\hat{h}_{\text{opt}}^{(l)}$ .
3. Construire l'estimateur MIS :
  - ⇒ Fabrication d'une densité auxiliaire  $\hat{h}_{\text{MIS}}$  par mélange des densités sous-optimales  $\hat{h}_{\text{opt}}^{(l)}$ .
  - ⇒ Recyclage du résultat des différentes explorations pour constituer une séquence :

$$\left\{(\mathbf{X}^{(j)}, Y^{(j)}, H^{(j)})\right\}_{j=1}^{N_t} \text{ avec } \mathbf{X}^{(j)} \stackrel{\text{i.i.d.}}{\sim} \hat{h}_{\text{MIS}} \text{ et } \begin{cases} Y^{(j)} = \phi(\mathbf{X}^{(j)}) \\ H^{(j)} = \hat{h}_{\text{MIS}}(\mathbf{X}^{(j)}) \end{cases} . \quad (5.73)$$

4. Estimer toutes les probabilités  $P_f^{(k)}$  par mise à jour des numérateurs  $\hat{g}^{(k)}(\mathbf{X}^{(j)})$  :

$$\forall 1 \leq k \leq M, \quad \hat{P}_f^{(k)} = \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbb{1}_{\{Y^{(j)} > T\}} \frac{\hat{g}^{(k)}(\mathbf{X}^{(j)})}{H^{(j)}} . \quad (5.74)$$

Pour un algorithme d'ASF intégrant une stratégie MIS pour l'estimation des probabilités  $P_f^{(k)}$ , le budget de simulation vaut  $N_{\text{tot}} = N_e \times (N_i \times N_s)$  avec  $N_e$  le nombre d'explorations,  $N_i$  le nombre d'itérations de  $\mathcal{R}_w$  avant d'atteindre  $D_f$  et  $N_s$  le nombre de particules impliquées dans chaque itération. Une synthèse de cette procédure d'ASF est représentée à la Figure 5.5 dans le cas le plus général où les coefficients de corrélation de Kendall sont calculés empiriquement avec  $\tilde{\mathcal{E}}$ .

### 5.3.2 Construction et estimation d'un indicateur de stochasticité

Dans la première partie de cette section, on a mis en place un algorithme d'ASF capable de s'accommoder à la dépendance statistique existant entre les entrées des applications  $\Psi_{\tau}$  et  $\tilde{\Psi}_{\tau}$ . Il est aussi souligné à la Section 5.2.2 que ces deux applications ont un caractère stochastique. Cela signifie que des sorties différentes peuvent être obtenues à partir d'un même lot d'entrées. L'AS en présence d'une boîte noire stochastique est une thématique de recherche en plein essor dans le domaine de la quantification des incertitudes. D'une part, la modélisation de certains phénomènes naturels peut passer par ce type d'approches. C'est par exemple le cas pour l'étude des réactions chimiques en chaîne dans GILLESPIE [1976] ou pour la description des systèmes biologiques dans WILKINSON [2009]. D'autre part, l'avènement des stratégies de métamodélisation des codes de simulation déterministes par des processus gaussiens a donné un écho bien plus grand à cette problématique. Les techniques d'AS doivent alors être adaptées à l'aléa supplémentaire introduit par la stochasticité du code de simulation étudié. Des méthodologies appropriées pour l'estimation des indices de Sobol sont par exemple proposées dans MARREL et collab. [2009].

Pour la question du traitement de la dépendance entre les entrées, on a vu que rien n'empêche l'application naïve des algorithmes d'AS développés sous une hypothèse d'indépendance. Il faut simplement être conscient des risques encourus lorsqu'on s'affranchit du cadre établi. De

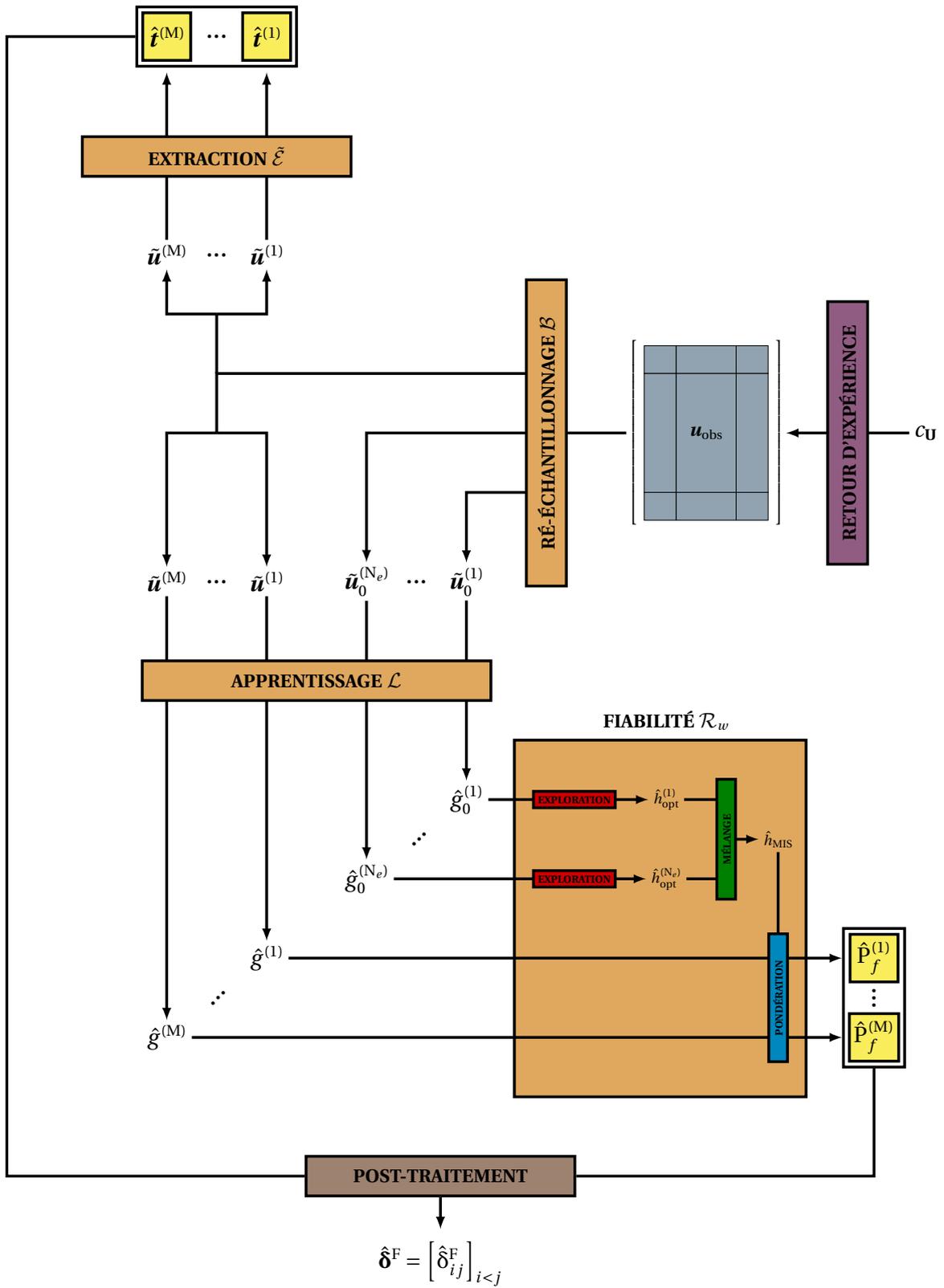


FIGURE 5.5 – Illustration des différentes briques de calcul lorsque l’ASF est exécutée avec  $\mathcal{R}_w$ .

même, pour la question du traitement de la stochasticité de la boîte noire, l'utilisateur peut décider d'appliquer sans précaution les mêmes techniques d'AS que celles dont il dispose dans le cas déterministe. D'un point de vue purement numérique, l'estimation des  $\delta$ -indices est tout à fait envisageable. Reste ensuite à savoir si on peut se fier aux résultats et en tirer des conclusions.

### 5.3.2.1 Étude d'un exemple analytique

Examinons un cas jouet pour mieux comprendre les enjeux de la maîtrise de la stochasticité. On garde les notations  $\tilde{\mathbf{X}}$  et  $\tilde{Y}$  même si elles sont ici déconnectées du contexte fiabiliste et on pose :

$$\tilde{Y} = a_1 \tilde{X}_1 + a_2 \tilde{X}_2 + a_3 \tilde{X}_3 = h_w(\tilde{X}_1, \tilde{X}_2) \quad \text{avec} \quad (\tilde{X}_1, \tilde{X}_2, \tilde{X}_3) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}) \quad \text{et} \quad \mathbf{\Gamma} := \begin{bmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (5.75)$$

La variable  $\tilde{Y}$  est traitée comme la sortie d'une application stochastique  $h$  dont les entrées  $\tilde{X}_1$  et  $\tilde{X}_2$  sont les seules variables observables.  $\tilde{X}_3$  est intégrée au fonctionnement interne de  $h$  et peut être vue comme une variable bruitant la sortie indépendamment des valeurs prises par les entrées. À partir de résultats élémentaires sur les vecteurs gaussiens, les différents  $\delta$ -indices du premier ordre s'obtiennent comme suit :

$$\delta_i := \delta_{\tilde{X}_i}^{\tilde{Y}} = \frac{1}{2} \int_{\mathcal{H}_2} |c_{\mathcal{N}}(\mathbf{u} | \rho_{\tilde{X}_i \tilde{Y}}) - 1| \, d\mathbf{u} \quad \text{avec} \quad \rho_{\tilde{X}_i \tilde{Y}} = \frac{\text{Cov}(\tilde{X}_i, \tilde{Y})}{\sqrt{\mathbb{V}(Y)}} \quad (5.76)$$

$$\text{où} \quad \text{Cov}(\tilde{X}_i, \tilde{Y}) = \begin{cases} a_1 + a_2 \rho & \text{si } i = 1 \\ a_2 + a_1 \rho & \text{si } i = 2 \\ a_3 & \text{si } i = 3 \end{cases} \quad \text{et} \quad \mathbb{V}(Y) = a_1^2 + a_2^2 + a_3^2 + 2 a_1 a_2 \rho .$$

La loi de  $\mathbf{X}_{12} := (X_1, X_2)$  est fixée avec  $\rho = 0.8$  tandis que  $h_w$  est réglée avec  $a_1 = a_2 = 1$  et  $a_3 = 2$ . Par Monte-Carlo, on montre facilement que  $\delta_1 = \delta_2 \approx 0.19$  tandis que  $\delta_3 \approx 0.43$  et cela conduit à faire plusieurs remarques quant à l'application des  $\delta$ -indices.

- Le passage d'une boîte noire déterministe à une boîte noire stochastique est élémentaire. La définition des  $\delta$ -indices reste la même. La seule différence réside dans le fait qu'il n'est pas possible d'estimer l'indice associé aux variables responsables du caractère stochastique.
- En pratique, les séquences  $\{\tilde{X}_1^{(k)}\}_{k=1}^M$  et  $\{\tilde{X}_2^{(k)}\}_{k=1}^M$  permettent d'estimer  $\delta_1$  et  $\delta_2$  à partir d'un schéma numérique "simple boucle" ou "orienté copule".
- Les  $\delta$ -indices des variables  $\tilde{X}_1$  et  $\tilde{X}_2$  offrent une grille de description incomplète bien que non erronée. La connaissance de  $\hat{\delta}_1$  et  $\hat{\delta}_2$  ne permet pas de comprendre que  $\tilde{X}_3$  est la principale source d'incertitude sur la sortie  $\tilde{Y}$ .

Si les variables  $\tilde{X}_j$  étaient mutuellement indépendantes, une décomposition ANOVA pourrait être mise en place et elle se simplifierait de la manière suivante :

$$\sum_{\mathbf{u} \subseteq \{1,2,3\}} S_{\mathbf{u}} = S_1 + S_2 + S_{12} + S_3^T = 1. \quad (5.77)$$

L'indice de Sobol d'ordre total  $S_3^T$  permettrait de quantifier la part de la variance  $\mathbb{V}(Y)$  expliquée par la variable cachée  $\tilde{X}_3$ . L'idée serait d'abord d'écrire que  $S_3^T = 1 - (S_1 + S_2 + S_{12})$  puis de chercher à estimer les termes du membre de droite. Dans le contexte particulier de l'application stochastique  $h_w : \tilde{\mathbf{x}}_{12} \mapsto \tilde{y}$ , on ne peut fabriquer qu'un M-échantillon  $\{(\tilde{X}_1^{(k)}, \tilde{X}_2^{(k)}, \tilde{Y}^{(k)})\}_{k=1}^M$  associant des entrées observables et des sorties. Il est alors difficile d'estimer les indices  $S_1$ ,  $S_2$  et  $S_{12}$  car les schémas d'estimation traditionnels (comme le *Pick-Freeze*) réclament l'accès à un matériel initial différent. La méthode proposée par [GAMBOA et collab. \[2020\]](#) permet d'estimer les indices du premier ordre  $S_1$  et  $S_2$  mais ne peut pas être appliquée à l'estimation de l'indice croisé  $S_{12}$ .

Dans le cas général de variables  $\tilde{X}_j$  dépendantes, les propriétés des  $\delta$ -indices énoncées dans BORGONOVO [2007] et rappelées au Théorème 3.2.1 offrent la possibilité d'évaluer le niveau de stochasticité de  $h_w$ . Pour cela, il faut examiner le  $\delta$ -indice de la paire  $(\tilde{X}_1, \tilde{X}_2)$  :

$$\delta_{12} := \delta_{\tilde{X}_{12}}^{\tilde{Y}} = \frac{1}{2} \int_{\mathcal{H}_3} |c_{\mathcal{N}}(\mathbf{u}, v | \Lambda_{12}) - c_{\mathcal{N}}(\mathbf{u} | \rho)| \, d\mathbf{u} dv \quad \text{avec} \quad \Lambda_{12} := \begin{bmatrix} 1 & \rho & \rho_{\tilde{X}_1 \tilde{Y}} \\ \rho & 1 & \rho_{\tilde{X}_2 \tilde{Y}} \\ \rho_{\tilde{X}_1 \tilde{Y}} & \rho_{\tilde{X}_2 \tilde{Y}} & 1 \end{bmatrix}. \quad (5.78)$$

De par la propriété de croissance des  $\delta$ -indices, on sait que  $\delta_i \leq \delta_{12}$  pour  $i \in \{1, 2\}$ . Par ailleurs,  $\delta_{12}$  est majoré par l'indice d'ordre total  $\delta_{123}$ . Une estimation par Monte-Carlo permet alors de vérifier que  $0.20 \approx \delta_{12} < \delta_{123} = 1$ . La faible valeur prise par  $\delta_{12}$  indique que les variables observables  $\tilde{X}_1$  et  $\tilde{X}_2$  ne suffisent pas à expliquer correctement la distribution de  $\tilde{Y}$ . L'écart entre  $\delta_{12}$  et 1 donne une idée de l'influence sur  $\tilde{Y}$  de la variable non observable  $\tilde{X}_3$ . Pour des valeurs croissantes données au paramètre  $a_3$  réglant l'amplitude de la stochasticité interne, on obtient :

$a_3$	$\delta_1$	$\delta_2$	$\delta_3$	$\delta_{12}$	$\delta_{123}$
0.1	0.62	0.62	0.02	0.93	1
0.5	0.54	0.54	0.08	0.68	1
1.0	0.43	0.43	0.17	0.49	1
3.0	0.18	0.18	0.43	0.20	1

(5.79)

L'augmentation du caractère stochastique de  $h_w$  se traduit simultanément par une augmentation de  $\delta_3$  et une diminution de  $\delta_{12}$ . En outre, si on supprime le troisième terme dans l'équation (5.75),  $h_w$  devient une fonction déterministe  $h$  des variables  $\tilde{X}_1$  et  $\tilde{X}_2$ . On a alors  $\delta_{12} = 1$  car  $\delta_{12}$  devient l'indice d'ordre total après suppression de  $\tilde{X}_3$ . À partir d'une séquence d'observations conjointes  $\{(\tilde{X}_1, \tilde{X}_2)\}_{k=1}^M$ , l'estimation de  $\delta_{12}$  permet ainsi de mesurer le degré de stochasticité par comparaison à une valeur de référence, en l'occurrence 1. Dans le prochain paragraphe, cette idée est généralisée et appliquée pour mesurer la stochasticité des applications  $\Psi_{\tau}$  et  $\tilde{\Psi}_{\tau}$  décrivant les sous-problèmes d'ASF. On peut ensuite conforter ou remettre en cause la stratégie visant à enrichir l'état de connaissance des entrées du problème. En effet, si la stochasticité interne écrase les autres sources d'incertitude, il vaut mieux chercher à intervenir directement sur le code de calcul stochastique, par exemple en choisissant un meilleur résumé matriciel de  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ .

### 5.3.2.2 Définition d'un indicateur de stochasticité

Pour mesurer la stochasticité de l'application  $\tilde{\Psi}_{\tau}$ , on s'intéresse au  $\delta$ -indice d'ordre total des variables observables, c'est-à-dire au  $\delta$ -indice associé à l'ensemble des coefficients  $\hat{t}_{ij}$  constituant la matrice  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$ . Un indicateur de stochasticité de  $\tilde{\Psi}_{\tau}$  est ainsi défini par :

$$\xi^{\text{F}} := \delta\left(\mathbb{P}_f(\mathbf{U}_{\text{obs}}) \mid \hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})\right) = \delta(\tilde{\mathbf{Y}} \mid \tilde{\mathbf{X}}) = \frac{1}{2} \mathbb{E}[\text{shift}(\tilde{\mathbf{X}})], \quad (5.80)$$

en réutilisant les notations  $\tilde{\mathbf{X}}$  et  $\tilde{\mathbf{Y}}$  introduites à l'équation (5.58). La fonction  $\text{shift}(\cdot)$  s'écrit alors :

$$\forall \mathbf{x} \in [-1, 1]^p, \quad \text{shift}(\mathbf{x}) = \|f_{\tilde{\mathbf{Y}}}(\cdot) - f_{\tilde{\mathbf{Y}}|\tilde{\mathbf{X}}}(\cdot | \mathbf{x})\|_{L^1} = \int_0^1 |f_{\tilde{\mathbf{Y}}}(y) - f_{\tilde{\mathbf{Y}}|\tilde{\mathbf{X}}}(y | \mathbf{x})| \, dy. \quad (5.81)$$

Dans l'esprit des équations (5.57), (5.60) et (5.63), on obtient les égalités suivantes :

$$\xi^{\text{F}} = \frac{1}{2} \int_{[-1, 1]^p} \left( \int_0^1 |f_{\tilde{\mathbf{Y}}}(y) - f_{\tilde{\mathbf{Y}}|\tilde{\mathbf{X}}}(y | \mathbf{x})| \, dy \right) f_{\tilde{\mathbf{X}}}(\mathbf{x}) \, d\mathbf{x} \quad (5.82)$$

$$= \frac{1}{2} \int_{[-1, 1]^p \times [0, 1]} |f_{\tilde{\mathbf{X}}\tilde{\mathbf{Y}}}(\mathbf{x}, y) - f_{\tilde{\mathbf{X}}}(\mathbf{x}) f_{\tilde{\mathbf{Y}}}(y)| \, d\mathbf{x} dy \quad (5.83)$$

$$= \frac{1}{2} \int_{\mathcal{H}_{p+1}} |c_{\tilde{\mathbf{R}}\tilde{\mathbf{S}}}(\mathbf{u}, v) - c_{\tilde{\mathbf{R}}}(\mathbf{u})| \, d\mathbf{u} dv. \quad (5.84)$$

Dans l'exemple analytique exhibé à la Section 5.3.2.1, la stochasticité de  $h_w$  provient de la non prise en compte de la variable  $\tilde{X}_3$  qui est supposée inobservable. La quantité  $1 - \delta_{12}$  permet de mesurer l'amplitude du bruit causé par l'action cachée de  $\tilde{X}_3$ . Sa définition fait intervenir implicitement  $\tilde{X}_3$  mais son estimation ne requiert que la connaissance d'une série de réalisations des entrées observables et des sorties associées. Réciproquement, toute application stochastique peut être réécrite sous la forme d'une application déterministe dont certaines variables sont inobservables. Ce formalisme est notamment utilisé dans [HART et collab. \[2017\]](#) où l'incertitude affectant la sortie est décomposée entre l'incertitude provenant des entrées observables et l'incertitude provenant de la stochasticité interne. Le cas de l'application  $\tilde{\Psi}_\tau$  est différent de celui de  $h_w$  car le caractère stochastique provient de la propagation indirecte des incertitudes épistémiques à travers les opérateurs  $\tilde{\mathcal{E}}$  et  $\mathcal{R}$ . Afin de réinterpréter  $\tilde{\Psi}_\tau$ , on introduit un vecteur aléatoire  $\mathring{\mathbf{X}}$  issu du stockage de l'information portée par la  $d$ -copule estimée  $\hat{c}_\mathbf{U}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ . La dimension  $q$  de  $\mathring{\mathbf{X}}$  dépend évidemment de la nature de  $\mathcal{M}_{\text{cop}}$ . Plus précisément :

- Si  $\hat{c}_\mathbf{U}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}}) = \hat{c}_\mathcal{N}(\cdot | \hat{\boldsymbol{\rho}})$  alors  $q = \frac{1}{2}d(d-1)$  car  $\mathring{\mathbf{X}} = [\hat{\rho}_{ij}]_{i < j}$ .
- Si  $\hat{c}_\mathbf{U}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  est estimée par KDE,  $q = Nd + \frac{1}{2}d(d+1)$  car  $\mathring{\mathbf{X}}$  rassemble :
  - ⇒ Les  $Nd$  observations  $U_j^{(i)}$  de l'échantillon  $\mathbf{U}_{\text{obs}}$ .
  - ⇒ Les  $\frac{1}{2}d(d+1)$  coefficients  $H_{ij}$  de la fenêtre de lissage multidimensionnelle  $\mathbf{H}$ .
- Si  $\hat{c}_\mathbf{U}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  est une R-vigne et que l'on choisit de suivre le système de spécification proposé dans [DISSMANN et collab. \[2013\]](#) ou dans [JOE \[2014\]](#) et implémenté dans le package R `VineCopula`<sup>1</sup>,  $q = d^2 + d(d-1)$  car  $\mathring{\mathbf{X}}$  met bout à bout :
  - ⇒ Les  $d^2$  variables discrètes impliquées dans le codage de  $\hat{\mathbf{T}}_{\text{vg}}$  et de  $\hat{\mathbf{F}}_{\text{vg}}$ .
  - ⇒ Les  $d(d-1)$  variables continues formant le vecteur des paramètres  $\hat{\boldsymbol{\theta}}_{\text{vg}}$ .

On note  $S$  et  $S^{-1}$  les opérateurs représentant respectivement le stockage et le déstockage de l'information de  $\hat{c}_\mathbf{U}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  dans le vecteur  $\mathring{\mathbf{X}}$ . On peut alors voir que :

$$\tilde{Y} = P_f(\mathbf{U}_{\text{obs}}) = \mathcal{R}[\hat{c}_\mathbf{U}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})] = \mathcal{R}[S^{-1}(\mathring{\mathbf{X}})] = [\mathcal{R} \circ S^{-1}](\mathring{\mathbf{X}}). \quad (5.85)$$

La dernière égalité indique que  $\tilde{Y}$  est une fonction déterministe de  $\mathring{\mathbf{X}}$ . Par extension,  $\tilde{Y}$  est aussi une fonction déterministe de  $(\tilde{\mathbf{X}}, \mathring{\mathbf{X}})$  dans laquelle  $\tilde{\mathbf{X}}$  n'intervient pas. On note  $\mathring{\Psi}_\tau$  cette application :

$$\begin{aligned} \mathring{\Psi}_\tau : [-1, 1]^p \times \mathbb{R}^q &\longrightarrow [0, 1] \\ (\tilde{\mathbf{X}}, \mathring{\mathbf{X}}) &\longmapsto \tilde{Y} = [\mathcal{R} \circ S^{-1}](\mathring{\mathbf{X}}) \end{aligned} \quad (5.86)$$

$\mathring{\Psi}_\tau$  est déterministe tant que la spécification  $\mathring{\mathbf{X}}$  est fournie. Si ce n'est plus le cas,  $\mathring{\Psi}_\tau$  devient stochastique et se transforme en  $\tilde{\Psi}_\tau$ . Il faut bien noter que les indices  $\delta_{ij}^F$  et  $\xi^F$  sont définis de façon équivalente lorsqu'on adopte le point de vue de  $\tilde{\Psi}_\tau$  ou celui de  $\mathring{\Psi}_\tau$  :

$$\begin{aligned} \delta_{ij}^F &= \delta(\tilde{\Psi}_\tau(\tilde{\mathbf{X}}) | \tilde{\mathbf{X}}_{ij}) = \delta(\mathring{\Psi}_\tau(\tilde{\mathbf{X}}, \mathring{\mathbf{X}}) | \tilde{\mathbf{X}}_{ij}) \\ \xi^F &= \delta(\tilde{\Psi}_\tau(\tilde{\mathbf{X}}) | \tilde{\mathbf{X}}) = \delta(\mathring{\Psi}_\tau(\tilde{\mathbf{X}}, \mathring{\mathbf{X}}) | \tilde{\mathbf{X}}) \end{aligned} \quad (5.87)$$

car la distribution des entrées observables  $\tilde{\mathbf{X}}$  et celle de la sortie  $\tilde{Y}$  restent inchangées. On peut se reporter à la Figure 5.6 pour une confrontation des applications  $\tilde{\Psi}_\tau$  et  $\mathring{\Psi}_\tau$ . Bien que le vecteur  $\tilde{\mathbf{X}}$  ne soit pas impliqué dans le calcul de la sortie  $P_f(\mathbf{U}_{\text{obs}})$ , les  $\delta$ -indices associés aux différents sous-vecteurs de  $\tilde{\mathbf{X}}$  ne sont pas nécessairement nuls. Cette situation se produirait si  $\tilde{\mathbf{X}}$  et  $\mathring{\mathbf{X}}$  étaient indépendants. Ce n'est évidemment pas le cas étant donné que :

$$\tilde{\mathbf{X}} = \tilde{\mathcal{E}}(\mathbf{U}_{\text{obs}}) \quad \text{et} \quad \mathring{\mathbf{X}} = [S \circ \mathcal{L}](\mathbf{U}_{\text{obs}}) \quad (5.88)$$

La redondance d'information entre  $\tilde{\mathbf{X}}$  et  $\mathring{\mathbf{X}}$  n'affecte en rien les valeurs prises par les  $\delta$ -indices. Il faut cependant remarquer que l'on a les propriétés suivantes :

1. <https://cran.r-project.org/package=VineCopula>

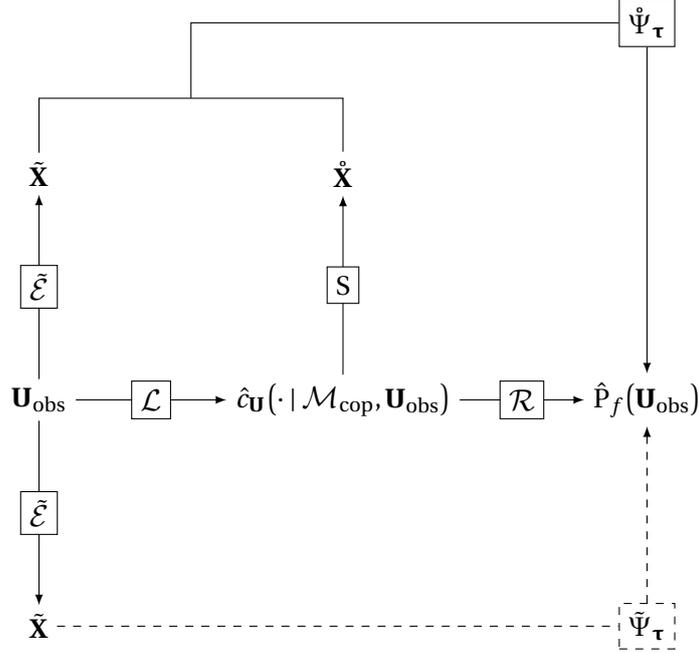


FIGURE 5.6 – Transformation de  $\tilde{\Psi}_{\tau}$  en une application déterministe  $\hat{\Psi}_{\tau}$  incomplète.

$$\boxed{\delta_A} \quad \delta(\tilde{Y} \mid \mathring{\mathbf{X}}) = 1 \text{ car } \mathring{\mathbf{X}} \text{ suffit à calculer la sortie } \tilde{Y}.$$

$$\boxed{\delta_B} \quad \delta(\tilde{Y} \mid \tilde{\mathbf{X}}, \mathring{\mathbf{X}}) = \dots = \delta(\tilde{Y} \mid \tilde{X}_{ij}, \mathring{\mathbf{X}}) = \delta(\tilde{Y} \mid \mathring{\mathbf{X}}) = 1.$$

$$\boxed{\delta_C} \quad \delta(\tilde{Y} \mid \tilde{\mathbf{X}}) \leq 1 \text{ avec égalité lorsqu'il existe une fonction } g : \mathbb{R}^p \longrightarrow \mathbb{R}^q \text{ telle que } \mathring{\mathbf{X}} = g(\tilde{\mathbf{X}}).$$

Le cas d'égalité dans  $\boxed{\delta_C}$  correspond à une configuration idéale où le résumé  $\tilde{\mathbf{X}}$  contient autant d'information que la spécification complète  $\mathring{\mathbf{X}}$  de la  $d$ -copule  $\hat{c}_{\mathbf{U}}(\cdot \mid \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ . Pour  $\tilde{\Psi}_{\tau}$ , cette situation n'existe pas car aucun modèle  $\mathcal{M}_{\text{cop}}$  n'est construit à partir de l'information portée par la matrice empirique  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$  des corrélations de Kendall. Pour  $\Psi_{\tau}$ , l'utilisation de la matrice  $\hat{\mathbf{t}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  rend ce recoupement d'information envisageable dans certains cas particuliers. On renvoie par exemple au cas de la copule gaussienne ou des copules archimédiennes. L'indicateur de stochasticité  $\xi^{\text{F}}$  peut donc être vu comme un  $\delta$ -indice d'ordre  $p$  calculé à partir d'une application déterministe  $\hat{\Psi}_{\tau}$  impliquant  $p + q$  entrées scalaires potentiellement dépendantes. Dans le prochain paragraphe, on discute de la mise en place d'un schéma d'estimation. Comme  $\xi^{\text{F}}$  est un  $\delta$ -indice d'ordre supérieur, de nouvelles difficultés apparaissent avec la montée en dimension et les schémas numériques doivent alors incorporer des outils capables d'y résister.

### 5.3.2.3 Mise en place d'un schéma d'estimation de l'indicateur de stochasticité

Au vu des expressions intégrales (5.82), (5.83) et (5.84), on comprend que les différentes méthodes d'estimation présentées à la Section 3.2.3.2 sont applicables à première vue. À l'instar de ce qui est fait à l'équation (5.59), l'estimateur "double boucle" est donné par :

$$\hat{\xi}^{\text{F}} = \frac{1}{2N_s} \sum_{i=1}^{N_s} \left\| \hat{f}_{\tilde{Y}}(\cdot) - \hat{f}_{\tilde{Y} \mid \tilde{\mathbf{X}}}(\cdot \mid \mathbf{Z}^{(k)}) \right\|_{L^1} \quad \text{avec } \mathbf{Z}^{(k)} \stackrel{\text{i.i.d.}}{\sim} f_{\tilde{\mathbf{X}}}, \quad (5.89)$$

et passe par l'estimation de  $N_s$  densités conditionnelles de type  $f_{\tilde{Y} \mid \tilde{\mathbf{X}}}(\cdot \mid \mathbf{z})$  pour  $\mathbf{z} \in [-1, 1]^p$ . Cela implique d'être capable de simuler des échantillons de  $P_f(\mathbf{U}_{\text{obs}})$  sous la contrainte d'une matrice  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$  fixée. Comme  $P_f(\mathbf{U}_{\text{obs}}) = [\mathcal{R} \circ \mathcal{L}](\mathbf{U}_{\text{obs}})$  avec  $\mathcal{R}$  et  $\mathcal{L}$  des opérateurs déterministes, le cœur du problème est en fait de savoir simuler  $\mathbf{U}_{\text{obs}}$  conditionnellement au fait que  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$  soit constante. On a déjà dit que  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$  ne peut prendre qu'un nombre fini de valeurs matricielles.

Par ailleurs, la valeur prise par  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$  ne dépend en fait que des rangs au sein des colonnes de  $\mathbf{U}_{\text{obs}}$ . On introduit la matrice des rangs de  $\mathbf{U}_{\text{obs}}$  :

$$\text{rg}(\mathbf{U}_{\text{obs}}) := \mathbf{R}_{\text{obs}} = [\mathbf{R}_j^{(i)}]_{\substack{1 \leq i \leq N \\ 1 \leq j \leq d}} \quad \text{avec} \quad \mathbf{R}_j^{(i)} = \sum_{k=1}^N \mathbb{1}_{\{U_j^{(k)} \leq U_j^{(i)}\}}. \quad (5.90)$$

Il est facile de voir que  $\tilde{\mathcal{E}}(\mathbf{U}_{\text{obs}}) = \tilde{\mathcal{E}}(\mathbf{R}_{\text{obs}})$ . On comprend ainsi que la seule connaissance de  $\mathbf{R}_{\text{obs}}$  permet de calculer la matrice empirique des corrélations de Kendall. Contrairement à  $\mathbf{U}_{\text{obs}}$  qui est sous-tendue par la mesure d'échantillonnage  $(\mu_{\mathbf{U}})^{\otimes N}$ ,  $\mathbf{R}_{\text{obs}}$  suit une loi discrète à  $(N!)^d$  atomes. Toutefois, comme une permutation des lignes de  $\mathbf{R}_{\text{obs}}$  laisse  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$  inchangée, il n'y a en réalité que  $(N!)^{d-1}$  matrices à prendre en compte. Elles sont regroupées au sein d'un ensemble noté  $\mathcal{R}$ . Imaginons que l'on cherche à simuler  $\mathbf{U}_{\text{obs}}$  conditionnellement au fait que la matrice aléatoire  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$  soit égale à  $\mathbf{t}_0$ . Pour y parvenir, il faut pouvoir accomplir les trois tâches énoncées ci-après.

(T1) Retrouver toutes les matrices de  $\mathcal{R}$  conduisant à  $\mathbf{t}_0$  :

$$\tilde{\mathcal{E}}^{-1}(\{\mathbf{t}_0\}) = \{\mathbf{R}_0 \in \mathcal{R} : \tilde{\mathcal{E}}(\mathbf{R}_0) = \mathbf{t}_0\} = \{\mathbf{R}_0^{(1)}, \dots, \mathbf{R}_0^{(s)}\}. \quad (5.91)$$

(T2) Calculer les probabilités d'apparition des matrices de  $\tilde{\mathcal{E}}^{-1}(\{\mathbf{t}_0\})$  :

$$\forall 1 \leq i \leq s, \quad w_i := \mathbb{P}(\mathbf{R}_{\text{obs}} = \mathbf{R}_0^{(i)}). \quad (5.92)$$

(T3) Simuler  $\mathbf{U}_{\text{obs}}$  conditionnellement à un évènement de type  $\{\mathbf{R}_{\text{obs}} = \mathbf{R}_0^{(i)}\}$ .

À supposer que l'on soit capable d'exécuter facilement toutes ces tâches, la simulation de  $\mathbf{U}_{\text{obs}}$  conditionnellement à  $\{\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}}) = \mathbf{t}_0\}$  requiert alors deux étapes :

(a) Exécuter (T1) et (T2) pour simuler  $\mathbf{R}_{\text{obs}}$  conditionnellement à  $\{\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}}) = \mathbf{t}_0\}$ . On utilise une loi discrète sur  $\tilde{\mathcal{E}}^{-1}(\{\mathbf{t}_0\})$ . Les probabilités élémentaires sont alors données par :

$$p_i = \mathbb{P}(\mathbf{R}_{\text{obs}} = \mathbf{R}_0^{(i)} \mid \hat{\mathbf{t}}(\mathbf{U}_{\text{obs}}) = \mathbf{t}_0) \quad (5.93)$$

$$= \frac{\mathbb{P}(\mathbf{R}_{\text{obs}} = \mathbf{R}_0^{(i)})}{\mathbb{P}(\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}}) = \mathbf{t}_0)} = \frac{\mathbb{P}(\mathbf{R}_{\text{obs}} = \mathbf{R}_0^{(i)})}{\mathbb{P}(\mathbf{R}_{\text{obs}} \in \tilde{\mathcal{E}}^{-1}(\{\mathbf{t}_0\}))} = \frac{w_i}{\sum_{j=1}^s w_j}. \quad (5.94)$$

(b) Exécuter (T3) pour simuler  $\mathbf{U}_{\text{obs}}$  conditionnellement à  $\mathbf{R}_{\text{obs}}$ .

Ce processus de simulation est illustré à la Figure 5.7 où l'on explicite les deux étages du conditionnement. Examinons désormais la faisabilité des tâches (T1) à (T3).

- **Tâche (T1)** : il n'existe pas d'algorithme permettant de retrouver toutes les matrices  $\mathbf{R}_0^{(i)}$  constituant  $\tilde{\mathcal{E}}^{-1}(\{\mathbf{t}_0\})$ . Seule une approche par force brute peut être envisagée. Elle consiste à balayer les  $(N!)^{d-1}$  matrices de  $\mathcal{R}$  et à retenir celles qui vérifient  $\tilde{\mathcal{E}}(\mathbf{R}) = \mathbf{t}_0$ . C'est évidemment impossible en pratique, même lorsque  $N \leq 50$ .
- **Tâche (T2)** : chaque probabilité  $w_i$  s'exprime en fonction de la  $d$ -copule sous-jacente  $c_{\mathbf{U}}$ . En effet, en se rappelant que  $\mathbf{U}_{\text{obs}} \sim (\mu_{\mathbf{U}})^{\otimes N}$ , on peut écrire que :

$$w_i = \int_{A(\mathbf{R}_0^{(i)})} c_{\mathbf{U}}(u_1^{(1)}, \dots, u_d^{(1)}) \dots c_{\mathbf{U}}(u_1^{(N)}, \dots, u_d^{(N)}) d\mathbf{u}^{(1)} \dots d\mathbf{u}^{(N)} \quad (5.95)$$

où  $A(\mathbf{R}_0^{(i)})$  est un sous-domaine de  $[0, 1]^{N \times d}$  qui est défini par :

$$A(\mathbf{R}_0^{(i)}) := \bigcap_{j=1}^d \left\{ 0 < u_j^{(k_j^1)} < u_j^{(k_j^2)} < \dots < u_j^{(k_j^N)} < 1 \right\}, \quad (5.96)$$

avec  $\{k_j^1, k_j^2, \dots, k_j^N\}$  la permutation réciproque de la  $j$ -ème colonne de  $\mathbf{R}_0^{(i)}$ . Sachant que  $c_{\mathbf{U}}$  est une inconnue du problème, on ne peut pas calculer les probabilités élémentaires. À supposer que (T1) puisse être exécutée via une recherche exhaustive, la méconnaissance de  $c_{\mathbf{U}}$  bloque l'étape (a) du processus de simulation.

- **Tâche (T3) :** IMAN et CONOVER [1982] proposent une méthode très simple afin de produire un échantillon dont la matrice des rangs est imposée.

(a) Simuler une matrice  $\mathbf{V}_\perp$  dont tous les coefficients sont uniformes et indépendants :

$$\mathbf{V}_\perp = [V_j^{(i)}]_{\substack{1 \leq i \leq N \\ 1 \leq j \leq d}} \sim \left( \mathcal{U}(\mathcal{H}_d) \right)^{\otimes N}. \quad (5.97)$$

(b) Réordonner les colonnes de  $\mathbf{V}_\perp$  au sein d'une matrice  $\mathbf{W}$  :

$$\mathbf{W} = [W_j^{(i)}]_{\substack{1 \leq i \leq N \\ 1 \leq j \leq d}} \quad \text{avec} \quad W_j^{(1)} < W_j^{(2)} < \dots < W_j^{(N)} \quad \forall j \in \{1, \dots, d\}. \quad (5.98)$$

La variable  $W_j^{(i)}$  occupe alors le  $i$ -ème rang au sein de la colonne  $\mathbf{W}^{[ \bullet j ]}$ .

(c) Définir  $\mathbf{U}_{\text{obs}}$  en réordonnant  $\mathbf{W}$  en fonction des rangs imposés par  $\mathbf{R}_{\text{obs}}$  :

$$\mathbf{U}_{\text{obs}} = [U_j^{(i)}]_{\substack{1 \leq i \leq N \\ 1 \leq j \leq d}} \quad \text{avec} \quad U_j^{(i)} = W_j^{(\mathbf{R}^{(i)})}. \quad (5.99)$$

La matrice  $\mathbf{U}_{\text{obs}}$  ainsi fabriquée vérifie bien  $\text{rg}(\mathbf{U}_{\text{obs}}) = \mathbf{R}_{\text{obs}}$ .

Finalement, seule la tâche (T3) semble facile à mettre en œuvre. Il n'est donc pas possible d'estimer  $\xi^F$  à l'aide d'une approche en "double boucle". Un estimateur en "simple boucle" est obtenu en adaptant l'équation (5.61) :

$$\hat{\xi}^F = \frac{1}{2N_s} \sum_{k=1}^{N_s} \left| \frac{\hat{f}_{\tilde{\mathbf{X}}}(Z_A^{(k)}) \hat{f}_{\tilde{\mathbf{Y}}}(Z_B^{(k)})}{\hat{f}_{\tilde{\mathbf{X}\tilde{\mathbf{Y}}}}(Z_A^{(k)}, Z_B^{(k)})} - 1 \right| \quad \text{avec} \quad (Z_A^{(k)}, Z_B^{(k)}) \stackrel{\text{i.i.d.}}{\sim} \hat{f}_{\tilde{\mathbf{X}\tilde{\mathbf{Y}}}}. \quad (5.100)$$

Comme pour les indices  $\delta_{ij}^F$  du premier ordre, rien n'empêche l'implémentation de ce schéma d'estimation. La possession d'un échantillon associant des entrées  $\tilde{\mathbf{X}}^{(k)}$  à des sorties  $\tilde{\mathbf{Y}}^{(k)}$  rend possible l'estimation des densités  $f_{\tilde{\mathbf{Y}}}$ ,  $f_{\tilde{\mathbf{X}}}$  et  $f_{\tilde{\mathbf{X}\tilde{\mathbf{Y}}}}$ . Toutefois, on est de nouveau confronté au problème de l'apprentissage de distributions dont les supports sont des sous-domaines compacts de  $\mathbb{R}$ ,  $\mathbb{R}^p$  et  $\mathbb{R}^{p+1}$ . Il est alors beaucoup plus pratique de suivre une approche "orientée copule". En vertu de l'équation (5.84), un estimateur de  $\xi^F$  est donné par :

$$\hat{\xi}^F = \frac{1}{2N_s} \sum_{k=1}^{N_s} \left| \hat{c}_{\tilde{\mathbf{R}}\tilde{\mathbf{S}}}(\mathbf{W}_A^{(k)}, \mathbf{W}_B^{(k)}) - \hat{c}_{\tilde{\mathbf{R}}}(\mathbf{W}_A^{(k)}) \right| \quad \text{avec} \quad (\mathbf{W}_A^{(k)}, \mathbf{W}_B^{(k)}) \stackrel{\text{i.i.d.}}{\sim} \mathcal{U}([0, 1])^{\otimes (d+1)}. \quad (5.101)$$

L'estimation de  $\xi^F$  peut être greffée au schéma d'estimation des indices  $\delta_{ij}^F$ . À l'issue des étapes (Bg1) à (Bg5), il suffit d'exécuter une ultime étape :

(Bg6) Estimer  $\xi^F$  à partir du matériel pré-existant :

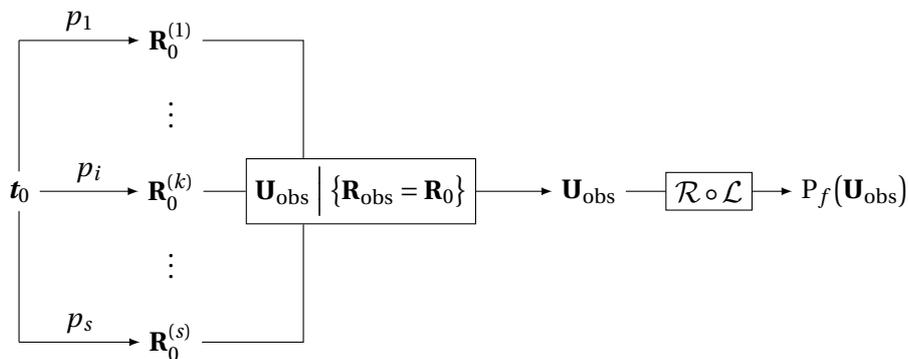


FIGURE 5.7 – Idée d'une méthode pour la simulation de  $P_f(\mathbf{U}_{\text{obs}})$  conditionnellement à  $\{\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}}) = \mathbf{t}_0\}$ .

(Bg6a) Estimer  $c_{\tilde{\mathbf{R}}}$  et  $c_{\tilde{\mathbf{R}}\tilde{\mathbf{S}}}$  à partir des échantillons  $\{(\tilde{\mathbf{R}}^{(k)}, \tilde{\mathbf{S}}^{(k)})\}_{k=1}^M$  de l'étape (Bg4).

(Bg6b) Appliquer la formule (5.101) pour estimer  $\xi^F$ .

Ainsi, un simple post-traitement d'échantillons déjà disponibles permet l'estimation de  $\xi^F$ . Il est notamment important de noter que l'étape (Bg3) qui est la plus coûteuse de par ses appels à  $\mathcal{R}$  et donc à  $\phi$  n'a pas besoin d'être relancée. Le principal écueil rencontré lors de la mise en place du schéma numérique "orienté copule" est l'étape (Bg6a) d'estimation des densités de copule.

Une première difficulté réside dans le fait que la copule  $c_{\tilde{\mathbf{R}}\tilde{\mathbf{S}}}$  est dégénérée si  $\tilde{\Psi}_\tau$  n'est pas stochastique. En effet, si  $\tilde{\Psi}_\tau$  est déterministe, il existe une fonction  $\lambda : \mathbb{R}^p \rightarrow \mathbb{R}$  telle que  $\tilde{\mathbf{S}} = \lambda(\tilde{\mathbf{R}})$ . Le vecteur  $(\tilde{\mathbf{R}}, \tilde{\mathbf{S}})$  s'écrit alors  $(\tilde{\mathbf{R}}, \lambda(\tilde{\mathbf{R}}))$  et son support effectif est un sous-domaine de  $\mathcal{H}_{p+1}$  négligeable au sens de la mesure de Lebesgue. Des exemples de copules dégénérées ont été exhibés au Théorème 1.2.7 avec la copule monotone  $M_2$  ou la copule anti-monotone  $W_2$ . Les supports de ces deux copules bidimensionnelles sont respectivement les segments  $u_2 = u_1$  et  $u_2 = -u_1$  de  $\mathcal{H}_2$ . Le fait que  $c_{\tilde{\mathbf{R}}\tilde{\mathbf{S}}}$  puisse être dégénérée pose un problème car la plupart des modèles d'apprentissage  $\mathcal{M}_{\text{cop}}$  n'englobent pas les copules de cette nature. Il faut donc rester conscient du fait que l'estimation de  $\xi^F$  nécessite qu'il existe une vraie stochastocité interne au sein de l'application  $\tilde{\Psi}_\tau$ . Si elle est pseudo-déterministe au sens où le bruit induit par  $\tilde{\mathbf{X}}$  est négligeable,  $\xi^F$  risque d'être très proche de 1 et son estimation peut s'avérer compliquée car il faut estimer une copule  $c_{\tilde{\mathbf{R}}\tilde{\mathbf{S}}}$  qui tend vers la dégénérescence.

La seconde et principale difficulté vient de la rapide montée en dimension induite par le fait que  $\tilde{\mathbf{X}}$  est un vecteur aléatoire de  $\mathbb{R}^p$  avec  $p = C_d^2$ . Pour  $d = 5$ , l'estimation de  $\xi^F$  implique l'estimation des copules  $c_{\tilde{\mathbf{R}}}$  et  $c_{\tilde{\mathbf{R}}\tilde{\mathbf{S}}}$  appartenant respectivement à  $\mathcal{C}_{10}$  et  $\mathcal{C}_{11}$ . L'utilisation de familles paramétriques (elliptiques ou archimédiennes) est à proscrire si on souhaite disposer d'une flexibilité suffisante. De même, la généralisation naïve des méthodes non-paramétriques ne résiste pas à la montée en dimension. Pour exécuter l'étape (Bg6a), il faut se tourner vers des modèles  $\mathcal{M}_{\text{cop}}$  capables de s'accommoder à la polymorphie accrue des motifs de dépendance. Dans un contexte légèrement différent où la sortie  $\tilde{Y}$  est multidimensionnelle, ZHOU et collab. [2018] ont proposé de renforcer le schéma numérique "orienté copule" en choisissant d'utiliser des R-vignes paramétriques à l'étape (Bg6a). Cette idée est ensuite reprise par SARAZIN et collab. [2020] pour estimer les  $\delta$ -indices d'ordre supérieur d'un code de calcul  $\phi$ . La précision des estimations fournies est validée expérimentalement pour plusieurs cas tests de référence. Par ailleurs, au vu des résultats énoncés par NAGLER et CZADO [2016], il est conseillé d'utiliser des R-vignes non-paramétriques pour deux raisons principales :

- **Robustesse** : le couplage entre l'algorithme traditionnel d'inférence des R-vignes et les techniques d'estimation non-paramétrique des copules bidimensionnelles aboutit à la création d'une méthode d'estimation qui n'est plus entravée par le fléau de la dimension.
- **Convergence** : Sous réserve que la copule sous-jacente  $c_{\tilde{\mathbf{R}}\tilde{\mathbf{S}}}$  vérifie l'hypothèse "simplificatrice", l'erreur d'estimation peut être bornée et un TCL peut être établi.

Notons  $\hat{c}_{\text{vg}}(\cdot | \hat{\mathbf{T}}_{\text{vg}}, \mathbf{U}_{\text{obs}})$  la densité estimée lorsque  $\mathcal{M}_{\text{cop}}$  représente l'apprentissage d'une R-vigne non-paramétrique. Au sujet du choix de la technique d'estimation non-paramétrique à déployer sur chaque branche de  $\hat{\mathbf{T}}_{\text{vg}}$ , NAGLER et collab. [2017] indiquent qu'il n'en existe pas une surpassant les autres en toutes circonstances. En conséquence, en accord avec ce qui a été proposé pour les indices  $\delta_{ij}^F$ , on utilise la méthode des noyaux transformés de GEENENS et collab. [2017]. Une routine de calcul pour l'algorithme final est disponible dans le package R `kdevine`<sup>2</sup>.

Il s'avère néanmoins que le modèle des R-vignes non-paramétriques n'est pas une solution infaillible pour l'estimation des copules en grande dimension. Dans le cas test aérospatial présenté en conclusion de SARAZIN et collab. [2020], les  $\delta$ -indices d'ordre supérieur sont très nettement sous-estimés malgré l'emploi de R-vignes non-paramétriques au sein d'un schéma numérique "orienté copule". Cette sous-estimation s'explique par le fait que les copules mises en jeu ne sont

2. <https://cran.r-project.org/package=kdevine>

pas compatibles avec l’hypothèse “simplificatrice”. La factorisation en R-vigne constitue alors une représentation erronée de la structure de dépendance. Contrairement au schéma numérique en “double boucle” où la convergence est garantie pour une taille d’échantillon  $M$  suffisante, l’approche “orientée copule” est tributaire des capacités d’apprentissage de  $\mathcal{M}_{\text{cop}}$  et du niveau de complexité de la structure de dépendance du vecteur  $(\hat{\mathbf{R}}, \hat{\mathbf{S}})$ . Face à l’écueil de la non-vérification de l’hypothèse “simplificatrice”, plusieurs alternatives sont proposées dans la littérature.

- Des tests ont été développés pour savoir si une copule conditionnelle de type :

$$\mathbf{u} \mapsto c_{\mathbf{U}_{12}|\mathbf{U}_3}(u_{1|3}, u_{2|3} \mid u_3), \quad (5.102)$$

dépend ou non de la valeur  $u_3$  prise par la covariable  $U_3$ . Si la dépendance est avérée, l’hypothèse “simplificatrice” ne peut pas être retenue. Un état de l’art des tests paramétriques et non-paramétriques est proposé dans [DERUMIGNY et FERMANIAN \[2017\]](#). Les différents travaux qui y sont mentionnés se cantonnent au test de l’influence d’une unique covariable. Au contraire, [KURZ et SPANHEL \[2017\]](#) montrent qu’on peut aussi tester l’influence d’un lot de plusieurs covariables. Ces mêmes auteurs en déduisent un algorithme permettant de valider l’hypothèse “simplificatrice” pour toutes les copules conditionnelles impliquées au sein d’une copule R-vigne. [KRAUS et CZADO \[2017\]](#) y ont vu une opportunité d’améliorer la procédure d’inférence de [DISSMANN et collab. \[2013\]](#). Leur idée consiste à choisir  $\hat{\mathbf{T}}_{\text{vg}}$  de manière à minimiser le nombre de fois où l’hypothèse “simplificatrice” est acceptée à tort.

- Il y a une volonté croissante de construire des R-vignes sans recourir à l’hypothèse “simplificatrice”. En effet, la prise en compte de covariables est naturelle du point de vue mathématique de la factorisation (1.175). L’hypothèse “simplificatrice” n’a en fait été introduite que pour faciliter l’implémentation de procédures inférentielles. La mise en place d’un système de description de l’influence des covariables reste un problème ouvert même si plusieurs variantes des R-vignes ont déjà vu le jour. Un apprentissage des copules conditionnelles à partir de B-splines est exploré dans [SCHELLHASE et SPANHEL \[2018\]](#) en reprenant une méthode développée au départ par [KAUERMANN et collab. \[2013\]](#). De plus, [HAN et collab. \[2017\]](#) ainsi que [VATTER et NAGLER \[2018\]](#) se sont intéressés à des R-vignes pour lesquelles les paramètres des copules bidimensionnelles sont exprimés en fonction des covariables suite à la calibration de modèles linéaires généralisés. Ces différentes initiatives ont été valorisées au sein des packages R `gamCopula`<sup>3</sup> et `pencopulaCond`<sup>4</sup>.

L’estimation de  $\xi^{\text{F}}$  à l’aide d’une approche “orientée copule” a l’avantage d’être facile à mettre en place dès lors qu’on dispose d’un modèle d’apprentissage  $\mathcal{M}_{\text{cop}}$  suffisamment robuste pour estimer avec précision la densité des copules  $c_{\hat{\mathbf{R}}}$  et  $c_{\hat{\mathbf{R}}\hat{\mathbf{S}}}$ . Toutefois, le choix de  $\mathcal{M}_{\text{cop}}$  mérite une attention particulière qui reste à la charge de l’utilisateur.

## 5.4 Développement d’une stratégie d’assimilation de nouvelles données

L’exécution de la procédure d’ASF développée à la Section 5.3 produit une collection d’indices  $\hat{\delta}_{ij}^{\text{F}}$  qui est assortie d’un indicateur de stochasticité  $\hat{\xi}^{\text{F}}$ . Une faible valeur prise par cet indicateur peut indiquer que la matrice  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$  constitue un résumé peu pertinent de la  $d$ -copule estimée  $\hat{c}_{\mathbf{U}}(\cdot \mid \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  et conduire à remettre en cause l’idée de prendre une décision à partir d’indices conçus pour expliquer la distribution de  $P_f(\mathbf{U}_{\text{obs}})$  à partir de celle des coefficients  $\hat{t}_{ij}$ . Toutefois, indépendamment des discussions suscitées par  $\hat{\xi}^{\text{F}}$ , on peut s’inspirer de la Section 4.2.4 pour vérifier que les indices  $\hat{\delta}_{ij}$  identifient correctement la paire  $(U_i, U_j)$  dont l’apprentissage selon  $\mathcal{M}_{\text{cop}}$  engendre le plus de variabilité sur  $P_f(\mathbf{U}_{\text{obs}})$ . À nouveau, l’idée consiste à réaliser une seconde ASF après avoir collecté puis assimilé des observations supplémentaires de la paire ciblée

3. <https://cran.r-project.org/package=gamCopula>

4. <https://cran.r-project.org/package=pencopulaCond>

par l'ASF initiale. On s'attend à ce que la seconde collection d'indices mette en exergue une diminution de l'indice  $\hat{\delta}_{ij}^F$ , étant donné que l'estimation de la copule bidimensionnelle  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$  peut éventuellement s'appuyer sur un plus grand budget d'observations. Toutefois, on rappelle que les  $\delta$ -indices ne vérifient pas la propriété de sommation à 1 des indices de Sobol. Par conséquent, l'évolution des classements est potentiellement plus subtile. Dans le Chapitre 4, la neutralisation d'une composante fonctionnelle déclenche systématiquement un ré-équilibrage des parts de la variance restante en faveur de l'une des autres composantes fonctionnelles. Ici, au contraire, on peut tout à fait envisager que la diminution de l'indice  $\hat{\delta}_{ij}^F$  ne s'accompagne pas de l'augmentation compensatoire d'un autre indice. Ce scénario peut survenir si l'apport des nouvelles observations entraîne une diminution de la capacité de  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$  à résumer la  $d$ -copule  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  et un renforcement du rôle joué par les variables inaccessibles rassemblées dans  $\hat{\mathbf{X}}$ . Cette situation est facilement détectable car elle coïncide avec un infléchissement de  $\hat{\xi}^F$  lors de la seconde ASF. Avant de pouvoir exécuter une seconde ASF, trois rouages doivent être mis en place.

1. Des données supplémentaires doivent être récupérées pour la paire  $(U_i, U_j)$  dont on souhaite améliorer l'état de connaissance. Si on note  $\mathbf{U}_{\text{extra}}$  une matrice aléatoire rassemblant  $N'$  observations supplémentaires de la  $d$ -copule  $c_{\mathbf{U}}$ , le résultat du processus de collecte est une réalisation  $\mathbf{u}_{\text{extra}}^{[\bullet(ij)]} \in [0, 1]^{N' \times 2}$  de la restriction de  $\mathbf{U}_{\text{extra}}$  à ses  $i$ -ème et  $j$ -ème colonnes.
2. Le mécanisme de ré-échantillonnage  $\mathcal{B}$  ne se limite plus au tirage d'une réplique  $\tilde{\mathbf{u}}$  de  $\mathbf{u}_{\text{obs}}$ . Il doit aussi inclure la création d'un clone  $\tilde{\mathbf{v}}_{ij}$  de  $\mathbf{u}_{\text{extra}}^{[\bullet(ij)]}$ . Seule une légère adaptation de  $\mathcal{B}$  est nécessaire. Elle est présentée à la Section 5.4.1 et tout repose de nouveau sur le *bootstrap*.
3. La phase d'apprentissage statistique  $\mathcal{L}$  doit être transformée car le modèle initial  $\mathcal{M}_{\text{cop}}$  prévoit d'estimer  $c_{\mathbf{U}}$  à partir d'un  $N$ -échantillon de  $\mathbf{U}$ . Le nombre d'observations de chaque variable  $U_j$  est donc supposé être le même. Cette difficulté n'est pas présente au Chapitre 4 car les différents modèles  $\mathcal{M}$  examinés sont basés sur la décomposition copule-marginales et peuvent donc assimiler tout matériel de type :

$$(\tilde{\mathbf{z}}_1, \tilde{\mathbf{z}}_2, \dots, \tilde{\mathbf{z}}_d, \tilde{\mathbf{v}}) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \dots \times \mathbb{R}^{n_d} \times [0, 1]^{n_{d+1} \times d} . \quad (5.103)$$

Tous les scénarios d'ASF du Chapitre 4 entrent dans ce cadre, avec en particulier :

- l'ASF initiale :  $n_1 = \dots = n_{d+1} = N$ ,
- l'ASF ciblant une marginale  $\hat{f}_{X_j}$  :  $n_j = N + N'$  et  $n_k = N$  sinon,
- l'ASF ciblant la copule  $\hat{c}_{\mathbf{U}}$  :  $n_{d+1} = N + N'$  et  $n_k = N$  sinon.

Grâce au théorème de Sklar, les sous-modèles  $\mathcal{M}_1, \dots, \mathcal{M}_d$  et  $\mathcal{M}_{\text{cop}}$  qui constituent  $\mathcal{M}$  peuvent être déployés simultanément, y compris lorsque  $n_1 \neq n_2 \neq \dots \neq n_{d+1}$ . Ainsi, la phase d'apprentissage  $\mathcal{L}$  peut être appliquée identiquement dans les trois situations d'ASF. Ici, comme suggéré à l'équation (5.103), l'échantillon d'apprentissage requis par  $\mathcal{M}_{\text{cop}}$  est une matrice de taille  $n_{d+1} \times d$  et rien ne permet a priori la prise en compte du matériel supplémentaire apporté par  $\mathbf{u}_{\text{extra}}^{[\bullet(ij)]}$ . Comme on a déjà pu le signaler, l'inexistence d'une décomposition universelle de la  $d$ -copule  $c_{\mathbf{U}}$  à l'aide des copules bidimensionnelles  $c_{ij}$  empêche de déconnecter l'estimation des différentes paires de variables. Il faut réfléchir à une transformation du modèle  $\mathcal{M}_{\text{cop}}$  de manière à ce qu'il permette l'incorporation des observations contenues dans  $\mathbf{u}_{\text{extra}}^{[\bullet(ij)]}$  et donc le raffinement de l'état de connaissance de l'une des copules bidimensionnelles. À notre connaissance, il n'existe pas de réponse passe-partout à cette question, les modifications à apporter étant dépendantes de  $\mathcal{M}_{\text{cop}}$ . La Section 5.4.2.1 s'intéresse au cas particulier de la copule gaussienne tandis que la Section 5.4.2.2 est consacrée à celui des R-vignes.

### 5.4.1 Adaptation du ré-échantillonnage aux données supplémentaires

Le nouveau mécanisme de ré-échantillonnage consiste simplement à simuler par *bootstrap* une copie de  $\mathbf{u}_{\text{extra}}^{[\bullet(ij)]}$  après avoir fait de même pour  $\mathbf{u}_{\text{obs}}$ . Cela peut être formalisé à l'aide d'un

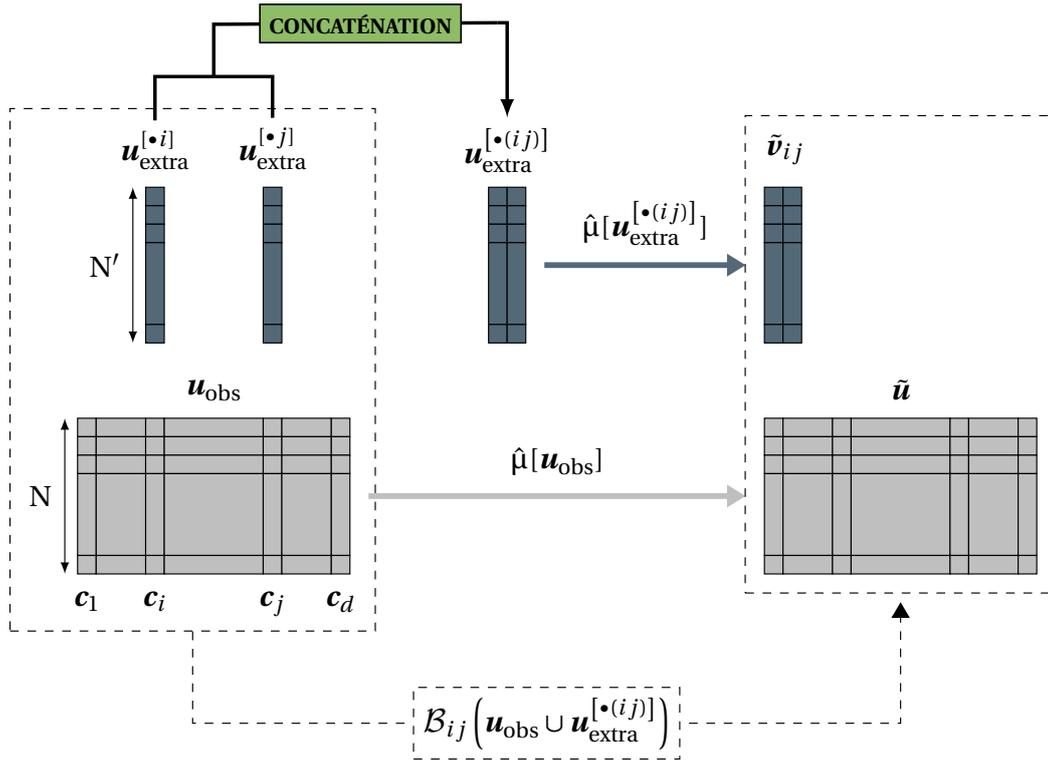


FIGURE 5.8 – Mécanisme de ré-échantillonnage  $\mathcal{B}_{ij}$  en présence d'observations supplémentaires de  $\mathbf{U}_{ij}$ .

opérateur  $\mathcal{B}_{ij}$  respectant le schéma suivant :

$$\mathcal{B}_{ij} : \begin{array}{ccc} [0, 1]^{N \times d} & \times & [0, 1]^{N' \times 2} \\ \mathbf{u}_{\text{obs}} & & \mathbf{u}_{\text{extra}}^{[\bullet(ij)]} \end{array} \longrightarrow \begin{array}{ccc} [0, 1]^{N \times d} & \times & [0, 1]^{N' \times 2} \\ \tilde{\mathbf{u}} & & \tilde{\mathbf{v}}_{ij} \end{array} . \quad (5.104)$$

Les objets ré-échantillonnés obéissent à une loi empirique construite à partir de l'ensemble des observations fournies :

$$(\tilde{\mathbf{u}}, \tilde{\mathbf{v}}_{ij}) \sim \left( \hat{\mu}[\mathbf{u}_{\text{obs}}] \right)^{\otimes N} \otimes \left( \hat{\mu}[\mathbf{u}_{\text{extra}}^{[\bullet(ij)]}] \right)^{\otimes N'} . \quad (5.105)$$

La Figure 5.8 donne un aperçu des manipulations matricielles accompagnant l'utilisation de  $\mathcal{B}_{ij}$ .

## 5.4.2 Adaptation de l'apprentissage aux données supplémentaires

### 5.4.2.1 Cas particulier de la copule gaussienne

On a déjà vu que la copule gaussienne est une famille paramétrique dont l'estimation est triviale car elle obéit à la formule (5.9) qui agglomère les étapes (R1) à (R3) détaillées au Chapitre 4. L'interprétation des paramètres estimés est également très simple car le comportement d'une paire est entièrement résumé par son coefficient de corrélation. À ce sujet, on peut remarquer que l'estimation de  $\rho_{ij}$  s'appuie simplement sur des observations conjointes de la paire  $(U_i, U_j)$  et non des observations conjointes du vecteur  $\mathbf{U}$ . On peut alors penser qu'il est possible de déconnecter l'estimation des différents coefficients et d'utiliser des budgets d'observations de tailles inégales. Supposons qu'on dispose d'un N-échantillon  $\mathbf{U}_{\text{obs}}$  et d'observations supplémentaires  $\mathbf{U}_{\text{extra}}^{[\bullet(ij)]}$  pour une paire  $(U_i, U_j)$  sur laquelle il est décidé d'intervenir. Un estimateur  $\hat{\rho}$  de la matrice

de corrélation peut être construit de la manière suivante :

$$\hat{\rho}_{kl} = \frac{\hat{\Gamma}_{kl}}{\sqrt{\hat{\Gamma}_{kk}}\sqrt{\hat{\Gamma}_{ll}}} \quad \text{avec} \quad \hat{\Gamma}_{kk} = \frac{1}{N_1} \sum_{n=1}^{N_1} \left( E_k^{(n)} \right)^2 \quad \text{où} \quad N_1 = \begin{cases} N + N' & \text{si } k \in \{i, j\} \\ N & \text{sinon} \end{cases}, \quad (5.106)$$

$$\text{et} \quad \hat{\Gamma}_{kl} = \frac{1}{N_2} \sum_{n=1}^{N_2} E_k^{(n)} E_l^{(n)} \quad \text{et} \quad N_2 = \begin{cases} N + N' & \text{si } (k, l) = (i, j) \\ N & \text{sinon} \end{cases}.$$

Contrairement à l'estimateur initial donné par la formule (5.9), la nouvelle matrice  $\hat{\rho}$  n'est plus définie positive. Au départ,  $\hat{\rho}$  est calculée à partir du seul N-échantillon  $\mathbf{U}_{\text{obs}}$  et  $N_1 = N_2 = N$ . En adoptant un formalisme matriciel, on peut écrire que :

$$\hat{\Gamma} = \frac{1}{N} \sum_{n=1}^N \left( \mathbf{E}_{\text{obs}}^{[n\bullet]} \right)^\top \mathbf{E}_{\text{obs}}^{[n\bullet]} \quad \text{avec} \quad \mathbf{E}_{\text{obs}} = \begin{bmatrix} E_j^{(i)} \\ \vdots \\ E_j^{(j)} \\ \vdots \\ E_j^{(d)} \end{bmatrix}_{\substack{1 \leq i \leq N \\ 1 \leq j \leq d}}. \quad (5.107)$$

Pour tout vecteur  $\mathbf{a} \in \mathbb{R}^d$ , on a alors :

$$\mathbf{a}^\top \hat{\Gamma} \mathbf{a} = \mathbf{a}^\top \left( \frac{1}{N} \sum_{n=1}^N \left( \mathbf{E}_{\text{obs}}^{[n\bullet]} \right)^\top \mathbf{E}_{\text{obs}}^{[n\bullet]} \right) \mathbf{a} = \frac{1}{N} \sum_{n=1}^N \mathbf{a}^\top \left( \mathbf{E}_{\text{obs}}^{[n\bullet]} \right)^\top \mathbf{E}_{\text{obs}}^{[n\bullet]} \mathbf{a} = \frac{1}{N} \sum_{n=1}^N \left( \mathbf{E}_{\text{obs}}^{[n\bullet]} \mathbf{a} \right)^2 \geq 0. \quad (5.108)$$

Par ailleurs, on peut remarquer que :

$$\mathbf{a}^\top \hat{\Gamma} \mathbf{a} = 0 \iff \sum_{n=1}^N \left( \mathbf{E}_{\text{obs}}^{[n\bullet]} \mathbf{a} \right)^2 = 0 \iff \mathbf{E}_{\text{obs}}^{[n\bullet]} \mathbf{a} = 0 \text{ pour } n \in \{1, \dots, N\} \iff \mathbf{E}_{\text{obs}} \mathbf{a} = \mathbf{0}. \quad (5.109)$$

La dernière égalité équivaut à  $\mathbf{a} = \mathbf{0}$  à condition que le rang de  $\mathbf{E}_{\text{obs}}$  soit égal à  $d$ . Cet évènement se produit presque sûrement à condition que la copule  $c_{\mathbf{U}}$  sous-jacente ne soit pas dégénérée. Dans le cas le plus général, on a donc :

$$\mathbb{P} \left( \det(\hat{\Gamma}) > 0 \right) = 1. \quad (5.110)$$

Par ailleurs, sachant que  $\hat{\Gamma} = \hat{\Delta} \hat{\rho} \hat{\Delta}$  où  $\hat{\Delta}$  est la matrice diagonale des écarts-types, il vient :

$$\det(\hat{\Gamma}) = \left( \det(\hat{\Delta}) \right)^2 \times \det(\hat{\rho}) \quad \text{et donc} \quad \mathbb{P} \left( \det(\hat{\rho}) > 0 \right) = 1. \quad (5.111)$$

Les équations (5.107) à (5.111) ne sont plus valables lorsque la matrice  $\hat{\rho}$  inclut un coefficient  $\hat{\rho}_{ij}$  qui est estimé avec  $N + N'$  observations. Dans cette situation, rien ne garantit que  $\hat{\rho}$  soit définie positive, ce qui est pourtant nécessaire pour que la copule gaussienne  $c_{\mathcal{N}}(\cdot | \hat{\rho})$  ait un sens du point de vue mathématique. Pour reprendre la terminologie de KUROWICKA et COOKE [2003],  $\hat{\rho}$  est une matrice de proto-corrélation. Elle est symétrique, ses coefficients diagonaux sont tous égaux à 1, ses coefficients non diagonaux sont tous compris dans l'intervalle  $] -1, 1[$  mais les axiomes de définition et de positivité peuvent ne pas être vérifiés. L'analyse numérique propose certains outils pour transformer une matrice de proto-corrélation en matrice de corrélation. HIGHAM [2002] cherche par exemple à construire :

$$\gamma(\hat{\rho}) := \arg \min_{\mathbf{R} \in \mathcal{R}_c} \|\hat{\rho} - \mathbf{R}\|_F, \quad (5.112)$$

où  $\mathcal{R}_c$  est l'ensemble des matrices de corrélation (de taille  $d \times d$ ) tandis que  $\|\cdot\|_F$  est la norme matricielle de Frobenius. L'algorithme associé est implémenté au sein du package R `Matrix`<sup>5</sup>. On fait le choix de ne pas utiliser cette méthode car on veut rester capable d'exprimer les coefficients  $\hat{\rho}_{ij}$  à l'aide d'une formule analytique.

Pour tirer profit de l'information supplémentaire apportée par  $\mathbf{U}_{\text{extra}}^{[\bullet(ij)]}$ , on a choisi d'explorer une autre piste. L'idée est d'exploiter le fait que le paramétrage d'une copule gaussienne est complètement équivalent à celui d'une copule R-vigne dont toutes les familles de  $F_{\text{vg}}$  sont gaussiennes.

5. <https://cran.r-project.org/package=Matrix>

Prenons un vecteur gaussien  $\mathbf{X}$  de moyenne  $\boldsymbol{\mu}$  et de matrice de covariance  $\boldsymbol{\Sigma}$ . On rappelle que la densité conjointe est donnée par :

$$f_{\mathbf{X}}(\mathbf{x}) = \left[ \prod_{j=1}^d f_{\mathcal{N}}(x_j | \mu_j, \sigma_j) \right] \times c_{\mathcal{N}}(\mathbf{u} | \boldsymbol{\rho}) \quad \text{avec} \quad \mathbf{u} := \left[ F_{\mathcal{N}}(x_j | \mu_j, \sigma_j) \right]_{1 \leq j \leq d} . \quad (5.113)$$

On cherche à connaître la nature des copules bidimensionnelles impliquées dans la factorisation en R-vigne de  $c_{\mathcal{N}}$ , et en particulier la nature des copules bidimensionnelles conditionnelles. Cela revient à connaître toute loi de la forme :

$$X_i, X_j \mid X_{k_1}, X_{k_2}, \dots, X_{k_c} \quad \text{avec} \quad 1 \leq k_c \leq d-2 . \quad (5.114)$$

On introduit les notations suivantes :

$$\mathbf{Z}_1 = [X_i \ X_j]^\top, \quad \mathbf{Z}_2 = [X_{k_1} \ X_{k_2} \ \dots \ X_{k_c}]^\top \quad \text{et} \quad \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{bmatrix} . \quad (5.115)$$

Les variables conditionnées et conditionnantes sont respectivement rangées dans  $\mathbf{Z}_1$  et  $\mathbf{Z}_2$ . Comme  $\mathbf{Z}$  est un sous-vecteur de  $\mathbf{X}$ , il s'agit également d'un vecteur gaussien. En séparant les paramètres décrivant  $\mathbf{Z}_1$  et ceux décrivant  $\mathbf{Z}_2$ , la loi de  $\mathbf{Z}$  s'écrit :

$$\mathbf{Z} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{Z}}, \boldsymbol{\Sigma}_{\mathbf{Z}}) \quad \text{avec} \quad \boldsymbol{\mu}_{\mathbf{Z}} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \quad \text{et} \quad \boldsymbol{\Sigma}_{\mathbf{Z}} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} . \quad (5.116)$$

Par ailleurs, la loi de  $\mathbf{Z}_1 | \mathbf{Z}_2$  est un résultat élémentaire, appelé théorème du conditionnement gaussien dans RASMUSSEN [2003] ou STEIN [2012]. Pour toute valeur  $\mathbf{z}_2 \in \mathbb{R}^{k_c}$  des covariables de  $\mathbf{Z}_2$ , la loi de  $\mathbf{Z}_1$  conditionnellement à l'évènement  $\{\mathbf{Z}_2 = \mathbf{z}_2\}$  reste gaussienne :

$$\mathbf{Z}_1 | \mathbf{Z}_2 = \mathbf{z}_2 \sim \mathcal{N}(\boldsymbol{\eta}(\mathbf{z}_2), \boldsymbol{\Gamma}(\mathbf{z}_2)) \quad \text{et} \quad \begin{cases} \boldsymbol{\eta}(\mathbf{z}_2) = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{z}_2 - \boldsymbol{\mu}_2) \\ \boldsymbol{\Gamma}(\mathbf{z}_2) = \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \end{cases} . \quad (5.117)$$

On note  $\pi(\mathbf{z}_2)$  le coefficient de corrélation pouvant être extrait de  $\boldsymbol{\Gamma}(\mathbf{z}_2)$ . Suite au conditionnement de  $\mathbf{Z}_1$  par  $\{\mathbf{Z}_2 = \mathbf{z}_2\}$ , le théorème de Sklar devient :

$$f_{\mathbf{Z}_1 | \mathbf{Z}_2}(\mathbf{z}_1 | \mathbf{z}_2) = \left[ \prod_{k \in \{i, j\}} f_{X_k | \mathbf{Z}_2}(x_k | \mathbf{z}_2) \right] \times c_{X_i X_j | \mathbf{Z}_2}(u(\mathbf{z}_2), v(\mathbf{z}_2) | \mathbf{z}_2) , \quad (5.118)$$

avec  $u(\mathbf{z}_2) = F_{X_i | \mathbf{Z}_2}(x_i | \mathbf{z}_2)$  et  $v(\mathbf{z}_2) = F_{X_j | \mathbf{Z}_2}(x_j | \mathbf{z}_2)$ . Le résultat (5.117) met en exergue deux points essentiels. D'une part, la gaussianité de  $\mathbf{Z}_1 | \mathbf{Z}_2 = \mathbf{z}_2$  induit celle de la copule. D'autre part, comme  $\boldsymbol{\Gamma}(\mathbf{z}_2)$  est constamment égale à  $\boldsymbol{\Gamma}_0 = \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}$ , le coefficient  $\pi(\mathbf{z}_2)$  ne dépend pas de  $\mathbf{z}_2$  et le paramétrage de la copule est donc invariant :

$$\forall \mathbf{z}_2 \in \mathbb{R}^{k_c}, \quad \forall \mathbf{v} \in \mathcal{H}_2, \quad c_{X_i X_j | \mathbf{Z}_2}(\mathbf{v} | \mathbf{z}_2) = c_{\mathcal{N}}(\mathbf{v} | \pi(\mathbf{z}_2)) = c_{\mathcal{N}}(\mathbf{v} | \pi_0) . \quad (5.119)$$

On est dans l'une des rares situations où l'hypothèse "simplificatrice" est vérifiée. Sachant que ce résultat est valable pour toutes les lois de type (5.114), toute factorisation de la densité d'une copule gaussienne en une densité de copule R-vigne est exacte. Prenons l'exemple en dimension  $d = 4$  de la copule D-vigne qui conduit à la factorisation suivante :

$$c_{\mathcal{N}}(\mathbf{u} | \hat{\boldsymbol{\rho}}) = c_{\mathcal{N}}(u_1, u_2 | \rho_{12}) \times c_{\mathcal{N}}(u_2, u_3 | \rho_{23}) \times c_{\mathcal{N}}(u_3, u_4 | \rho_{34}) \dots \\ \times c_{\mathcal{N}}(u_{1|2}, u_{3|2} | \rho_{13|2}) \times c_{\mathcal{N}}(u_{2|3}, u_{4|3} | \rho_{24|3}) \times c_{\mathcal{N}}(u_{1|23}, u_{4|23} | \rho_{14|23}) . \quad (5.120)$$

La séquence d'arbres  $\mathbf{T}_{\text{vg}}$  associée à cette factorisation est représentée à la Figure 5.9 où le paramétrage par la matrice de corrélation  $\boldsymbol{\rho}$  est remplacé par un système de paramétrage impliquant des coefficients de corrélation conditionnelle :

$$\boldsymbol{\theta}_{\text{vg}} = [\rho_{12} \ \rho_{23} \ \rho_{34} \ ; \ \rho_{13|2} \ \rho_{24|3} \ ; \ \rho_{14|23}] . \quad (5.121)$$

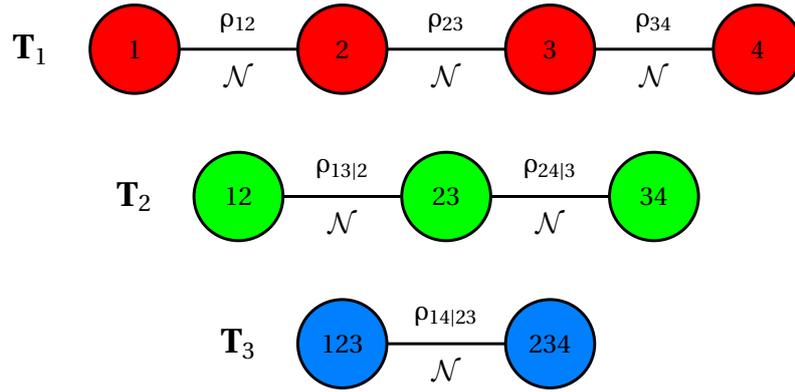


FIGURE 5.9 – Transformation d'une copule gaussienne en une copule D-vigne.

Il y a une parfaite équivalence entre  $\rho$  et  $\theta_{vg}$  de par les formules suivantes :

$$\rho_{13|2} = \frac{\rho_{13} - \rho_{23}\rho_{12}}{\sqrt{(1-\rho_{23}^2)(1-\rho_{12}^2)}} ; \rho_{24|3} = \frac{\rho_{24} - \rho_{23}\rho_{34}}{\sqrt{(1-\rho_{23}^2)(1-\rho_{34}^2)}} ; \rho_{14|23} = \frac{\rho_{14} - \omega_{12}}{\sqrt{(1-\omega_{11})(1-\omega_{22})}} \quad (5.122)$$

avec  $\begin{bmatrix} \omega_{11} \\ \omega_{22} \\ \omega_{12} \end{bmatrix} = \frac{1}{1-\rho_{23}^2} \begin{bmatrix} a_{11} \\ a_{22} \\ a_{12} \end{bmatrix}$  et  $\begin{cases} a_{11} = \rho_{12}^2 + \rho_{13}^2 - 2\rho_{12}\rho_{13}\rho_{23} \\ a_{22} = \rho_{24}^2 + \rho_{34}^2 - 2\rho_{23}\rho_{24}\rho_{34} \\ a_{12} = \rho_{12}\rho_{24} + \rho_{13}\rho_{34} + \rho_{23}(\rho_{12}\rho_{34} - \rho_{13}\rho_{24}) \end{cases}$ ,

que l'on peut retrouver dans l'ouvrage de JOE [2014]. En dimension quelconque, quelle que soit la factorisation en R-vigne qui est retenue, une généralisation de (5.122) permet de calculer les  $C_{d-1}^2$  coefficients de corrélation conditionnelle.

Le modèle d'apprentissage qui est construit dans la suite cherche à exploiter le fait qu'une copule gaussienne est une copule R-vigne dont les familles sont gaussiennes. Comme on a pu le voir à la Section 1.5.4.1, lors de l'apprentissage d'une copule R-vigne, la majorité des tâches sont exécutées dans un cadre bidimensionnel. C'est notamment le cas pour les quatre tâches élémentaires mentionnées ci-après :

- le calcul d'un coefficient empirique de corrélation de Kendall,
- la sélection d'une famille paramétrique par maximisation du critère AIC,
- l'estimation par MV des paramètres d'une famille donnée,
- la fabrication d'observations conditionnelles à l'aide d'une h-fonction.

Elles sont d'ailleurs suffisantes pour que l'algorithme de descente inventé par AAS et collab. [2009] puisse fonctionner. Ce contexte semble favorable à l'incorporation des données supplémentaires apportées par  $\mathbf{U}_{\text{extra}}^{[i,j]}$ . Le nouveau modèle est noté  $\mathcal{M}_{\text{cop}}^{(i,j)}$  pour indiquer qu'il s'agit d'une adaptation de  $\mathcal{M}_{\text{cop}}$  favorisant l'amélioration de l'estimation de la copule  $c_{ij}$ .

Dans la littérature sur les copules R-vignes, il est souvent conseillé d'intégrer à l'arbre  $T_1$  les paires de variables qu'il faut modéliser avec le plus de soin. Ici, au vu de l'objectif poursuivi, on impose l'existence d'une arête entre les sommets  $i$  et  $j$  de l'arbre  $T_1$ . En effet, en procédant de la sorte, l'estimation de la copule  $c_{\mathcal{N}}(\cdot | \rho_{ij})$  est directe et le calcul de  $\hat{\rho}_{ij}$  peut s'appuyer sur les  $N+N'$  observations de la paire  $(U_i, U_j)$  qui sont disponibles. En découle la densité  $\hat{c}_{\mathcal{N}}(\cdot | \hat{\rho}_{ij})$  ainsi que les h-fonctions  $\hat{h}_{\mathcal{N}}^{(G)}(\cdot | \hat{\rho}_{ij})$  et  $\hat{h}_{\mathcal{N}}^{(D)}(\cdot | \hat{\rho}_{ij})$ . On peut alors fabriquer  $N+N'$  observations conditionnelles de  $U_{j|i}$  et  $\hat{U}_{i|j}$ . Toutefois, lors de l'estimation des coefficients de corrélation conditionnelle de l'arbre  $T_2$ , elles sont croisées avec des  $N$ -échantillons d'observations conditionnelles issues des autres copules bidimensionnelles de l'arbre  $T_1$ . Les  $N'$  observations conditionnelles supplémentaires qui ont été fabriquées avec  $\mathbf{U}_{\text{extra}}^{[i,j]}$  deviennent inutiles et sont abandonnées. Les coefficients

de corrélation conditionnelle des copules gaussiennes présentes dans les arbres  $\mathbf{T}_2$  à  $\mathbf{T}_{d-1}$  sont ensuite calculés en adoptant l'approche séquentielle de [AAS et collab. \[2009\]](#). Ci-dessous, on récapitule les étapes permettant d'aboutir à  $\hat{c}_{\mathbf{U}}(\cdot \mid \mathcal{M}_{\text{cop}}^{(ij)}, \mathbf{U}_{\text{obs}}, \mathbf{U}_{\text{extra}}^{[\bullet(ij)]})$  lorsque  $\mathcal{M}_{\text{cop}}$  désigne l'estimation d'une copule gaussienne.

(G1) Établir le système de pondération suivant :

$$p_{kl} = \begin{cases} +\infty & \text{si } (k, l) = (i, j) \\ \hat{t}_{kl} & \text{sinon} \end{cases}, \quad (5.123)$$

puis résoudre le problème d'arbre couvrant de poids maximal. Le choix de  $p_{ij} = +\infty$  impose la construction d'une arête entre les sommets  $i$  et  $j$  de l'arbre  $\mathbf{T}_1$ .

(G2) Estimer les coefficients de corrélation  $\rho_{kl}$  des copules gaussiennes présentes dans  $\mathbf{T}_1$ .

- Utiliser les  $N + N'$  observations de  $\mathbf{U}_{\text{tot}}^{[\bullet(ij)]} := \mathbf{U}_{\text{obs}}^{[\bullet(ij)]} \cup \mathbf{U}_{\text{extra}}^{[\bullet(ij)]}$  pour calculer  $\hat{\rho}_{ij}$ .
- Utiliser les  $N$  observations de  $\mathbf{U}_{\text{obs}}^{[\bullet(kl)]}$  pour calculer tout autre coefficient  $\hat{\rho}_{kl}$ .

(G3) Pour chaque paire  $(U_k, U_l)$  impliquée dans  $\mathbf{T}_1$ , fabriquer  $N$  observations conditionnelles de  $U_{k|l}$  et  $U_{l|k}$  en appliquant les h-fonctions adéquates :

$$\forall 1 \leq n \leq N, \quad U_{k|l}^{(n)} = \hat{h}_{\mathcal{N}}^{(D)}(U_k^{(n)}, U_l^{(n)} \mid \hat{\rho}_{kl}) \quad \text{et} \quad U_{l|k}^{(n)} = \hat{h}_{\mathcal{N}}^{(G)}(U_k^{(n)}, U_l^{(n)} \mid \hat{\rho}_{kl}) . \quad (5.124)$$

(G4) Appliquer l'approche séquentielle de [AAS et collab. \[2009\]](#) pour estimer les coefficients de corrélation conditionnelle des arbres  $\mathbf{T}_2$  à  $\mathbf{T}_{d-1}$ .

Une illustration est proposée à la Figure 5.10 dans le cas particulier où la paire ciblée est  $(U_1, U_2)$  et où la factorisation s'opère sous la forme d'une D-vigne. Il faut bien noter que la copule R-vigne  $\hat{c}_{\mathcal{N}}(\cdot \mid \hat{\mathbf{T}}_{\text{vg}}, \hat{\boldsymbol{\theta}}_{\text{vg}})$  qui résulte de l'exécution des étapes (G1) à (G4) n'est plus équivalente à une copule gaussienne étant donné que  $\hat{\rho}_{ij}$  est calculé à partir d'un plus grand budget d'observations que les autres paramètres de  $\hat{\boldsymbol{\theta}}_{\text{vg}}$ . Par conséquent, au cours de la descente dans les arbres  $\mathbf{T}_2$  à  $\mathbf{T}_{d-1}$ , l'estimation des coefficients de corrélation conditionnelle doit s'effectuer par MV et non en appliquant la généralisation des formules (5.122) car celles-ci ne sont valables que pour une copule gaussienne rigoureusement paramétrée. De même, il n'est pas possible d'envisager un recalage final de  $\boldsymbol{\theta}_{\text{vg}}$  par maximisation de la log-vraisemblance  $\ell_N(\boldsymbol{\theta}_{\text{vg}})$  issue du choix de  $\hat{\mathbf{T}}_{\text{vg}}$  :

$$\hat{\boldsymbol{\theta}}_{\text{vg}} = \arg \max_{\boldsymbol{\theta}_{\text{vg}}} \ell_N(\boldsymbol{\theta}_{\text{vg}}) = \arg \max_{\boldsymbol{\theta}_{\text{vg}}} \sum_{n=1}^N \log \left( \hat{c}_{\mathcal{N}}(\mathbf{U}_{\text{obs}}^{[n\bullet]} \mid \hat{\mathbf{T}}_{\text{vg}}, \boldsymbol{\theta}_{\text{vg}}) \right). \quad (5.125)$$

En effet, l'écriture de  $\ell_N(\boldsymbol{\theta}_{\text{vg}})$  n'implique pas les observations supplémentaires apportées par  $\mathbf{U}_{\text{extra}}^{[\bullet(ij)]}$  et la résolution numérique de (5.125) conduit à une estimation moins précise de  $\rho_{ij}$ . Finalement,  $\mathcal{M}_{\text{cop}}^{(ij)}$  est une adaptation structurelle du modèle d'apprentissage initial  $\mathcal{M}_{\text{cop}}$ . Suite à l'incorporation des données fournies par l'acquisition ciblée, la  $d$ -copule finale  $\hat{c}_{\mathcal{N}}(\cdot \mid \hat{\mathbf{T}}_{\text{vg}}, \hat{\boldsymbol{\theta}}_{\text{vg}})$  ressemble à une copule gaussienne mais elle vit dans un espace dont la délimitation mathématique est plus permissive. Dans le prochain paragraphe, on reprend cette démarche dans un contexte plus général où il est question de l'apprentissage d'une copule R-vigne.

#### 5.4.2.2 Cas particulier des copules R-vignes

Dans ce paragraphe,  $\mathcal{M}_{\text{cop}}$  représente l'apprentissage d'une copule R-vigne pour laquelle rien n'est spécifié au préalable. L'algorithme  $\mathcal{L}$  englobe alors le choix d'une séquence d'arbres  $\hat{\mathbf{T}}_{\text{vg}}$ , la sélection de familles  $\hat{\mathbf{F}}_{\text{vg}}$  et l'estimation des paramètres  $\boldsymbol{\theta}_{\text{vg}}$  induits par  $(\hat{\mathbf{T}}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}})$ . L'idée consiste à nouveau à utiliser une factorisation de  $c_{\mathbf{U}}$  qui place  $c_{ij}$  dans l'arbre  $\mathbf{T}_1$ . Grâce à cette stratégie, on favorise l'assimilation des nouvelles données apportées par  $\mathbf{U}_{\text{extra}}^{[\bullet(ij)]}$ . En effet,  $N + N'$  observations conjointes de la paire  $(U_i, U_j)$  peuvent être utilisées pour sélectionner une famille paramétrique  $\mathcal{F}_{ij}$  et construire l'estimateur  $\hat{\boldsymbol{\theta}}_{ij}$  de ses paramètres. D'un point de vue algorithmique, la procédure induite par  $\mathcal{M}_{\text{cop}}^{(ij)}$  est très proche de celle présentée pour la copule gaussienne.

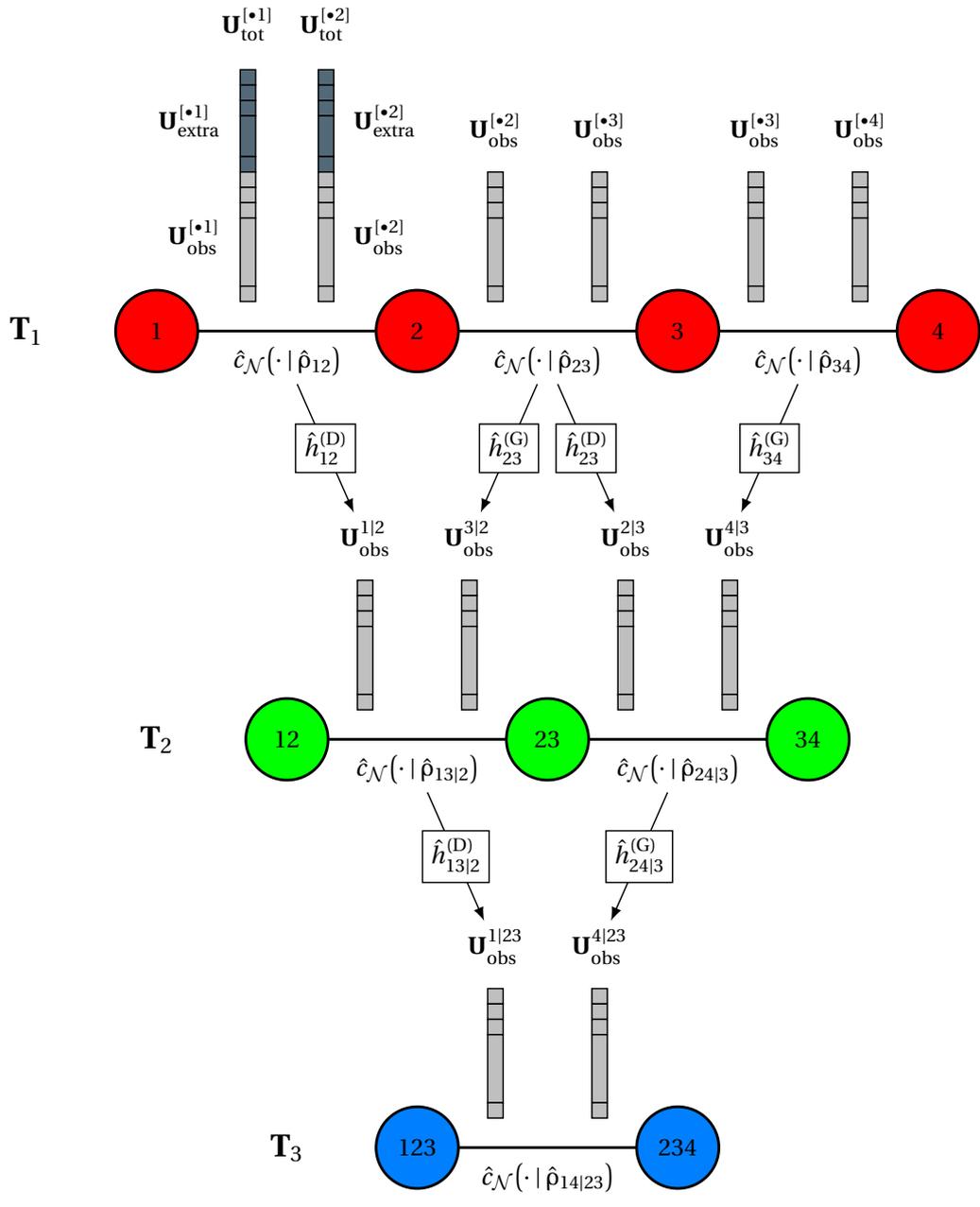


FIGURE 5.10 – Incorporation des données supplémentaires lors de l'estimation d'une copule gaussienne.

- (V1) Sélectionner l'arbre  $\mathbf{T}_1$  exactement comme à l'étape (G2).
- (V2) Sélectionner les familles  $\hat{\mathcal{F}}_{kl}$  des copules  $c_{kl}$  de  $\mathbf{T}_1$  par maximisation du critère AIC.
- Utiliser les  $N + N'$  observations de  $\mathbf{U}_{\text{tot}}^{[\bullet(ij)]} := \mathbf{U}_{\text{obs}}^{[\bullet(ij)]} \cup \mathbf{U}_{\text{extra}}^{[\bullet(ij)]}$  pour la copule  $c_{ij}$ .
  - Utiliser les  $N$  observations de  $\mathbf{U}_{\text{obs}}^{[\bullet(kl)]}$  pour tout autre copule  $c_{kl}$ .
- (V3) Calculer les paramètres  $\hat{\boldsymbol{\theta}}_{kl}$  associés aux familles  $\hat{\mathcal{F}}_{kl}$ .
- Utiliser les  $N + N'$  observations de  $\mathbf{U}_{\text{tot}}^{[\bullet(ij)]} := \mathbf{U}_{\text{obs}}^{[\bullet(ij)]} \cup \mathbf{U}_{\text{extra}}^{[\bullet(ij)]}$  pour la copule  $c_{ij}$ .
  - Utiliser les  $N$  observations de  $\mathbf{U}_{\text{obs}}^{[\bullet(kl)]}$  pour tout autre copule  $c_{kl}$ .
- (V4) Pour chaque paire  $(U_k, U_l)$  impliquée dans  $\mathbf{T}_1$ , fabriquer  $N$  observations conditionnelles de  $U_{k|l}$  et  $U_{l|k}$  en appliquant les h-fonctions adéquates :

$$\forall 1 \leq n \leq N, \quad \begin{cases} U_{k|l}^{(n)} = \hat{h}_{kl}^{(D)}(U_k^{(n)}, U_l^{(n)} \mid \hat{\mathcal{F}}_{kl}, \hat{\boldsymbol{\theta}}_{kl}) \\ U_{l|k}^{(n)} = \hat{h}_{kl}^{(G)}(U_k^{(n)}, U_l^{(n)} \mid \hat{\mathcal{F}}_{kl}, \hat{\boldsymbol{\theta}}_{kl}) \end{cases} . \quad (5.126)$$

- (V5) Appliquer l'approche séquentielle de **AAS et collab. [2009]** pour sélectionner les familles et estimer les paramètres au sein des arbres  $\mathbf{T}_2$  à  $\mathbf{T}_{d-1}$ .

La seule nouveauté réside dans le fait que la sélection de  $\hat{\mathbf{F}}_{\text{vg}}$  vient s'intercaler dans les rouages de l'inférence. Pour chacune des copules bidimensionnelles manipulées, le simple calcul de  $\hat{\rho}_{kl}$  est remplacé par la sélection de  $\hat{\mathcal{F}}_{kl}$  puis l'estimation de  $\hat{\boldsymbol{\theta}}_{kl}$ . Pour les copules R-vignes, le passage de  $\mathcal{M}_{\text{cop}}$  à  $\mathcal{M}_{\text{cop}}^{(ij)}$  n'engendre pas de changement de nature du modèle de distribution. La  $d$ -copule  $\hat{c}_{\text{vg}}(\cdot \mid \hat{\mathbf{T}}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}}, \hat{\boldsymbol{\theta}}_{\text{vg}})$  qui est obtenue après avoir conduit les étapes (V1) à (V5) demeure en effet une R-vigne. Toutefois, plusieurs différences sont notables :

- (a) La séquence d'arbres est soumise à une contrainte supplémentaire.
- (b) Certaines entités du modèle sont déduites à l'aide d'un plus grand budget d'observations.
- (c) La phase finale de recalage est supprimée.

Dans la Section 5.5, les différentes procédures d'ASF développées dans ce chapitre sont appliquées à des cas tests analytiques. Les objectifs poursuivis sont multiples :

- Contrôler la robustesse de l'estimation des indices  $\hat{\delta}_{ij}^{\text{F}}$  et de l'indicateur de stochasticité  $\hat{\xi}^{\text{F}}$ .
- Comparer les résultats d'ASF obtenus à partir des applications  $\Psi_{\mathbf{T}}$  et  $\tilde{\Psi}_{\mathbf{T}}$ .
- Étudier les modalités d'acquisition puis d'assimilation de données supplémentaires.

## 5.5 Validation des deux procédures d'ASF

### 5.5.1 Retour sur le cas test linéaire

#### 5.5.1.1 Rappels

Dans la Section 4.3.1, on s'est intéressé à une probabilité de défaillance  $\mathbb{P}(Y > T)$  émanant de l'application d'une fonction linéaire  $\phi$  à un vecteur gaussien  $\mathbf{X}$  :

$$Y = \phi(\mathbf{X}) = \mathbf{a}^\top \mathbf{X} \quad \text{et} \quad \mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) . \quad (5.127)$$

Les lois marginales  $\mathcal{N}(\mu_i, \sigma_i^2)$  sont désormais supposées connues ce qui donne accès à :

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \quad \text{et} \quad \boldsymbol{\Delta} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix} . \quad (5.128)$$

On dispose d'un échantillon  $\mathbf{U}_{\text{obs}} \sim (\boldsymbol{\mu}_{\mathbf{U}})^{\otimes N}$  d'observations conjointes du vecteur  $\mathbf{U}$ . La sortie  $Y$  peut d'ailleurs être directement exprimée en fonction de  $\mathbf{U}$  :

$$Y = \sum_{j=1}^3 a_j \left( \sigma_j \Phi^{-1}(U_j) + \mu_j \right). \quad (5.129)$$

La phase d'apprentissage statistique  $\mathcal{L}$  se limite à l'estimation par MV de la matrice de corrélation  $\boldsymbol{\rho}$  d'une copule gaussienne. Par ailleurs, l'analyse de fiabilité  $\mathcal{R}$  est triviale car la probabilité de défaillance  $P_f(\hat{\boldsymbol{\rho}})$  est une fonction déterministe de  $\hat{\boldsymbol{\rho}}$  donnée par :

$$P_f(\hat{\boldsymbol{\rho}}) = 1 - \Phi\left(\frac{T - \boldsymbol{\mu}_a}{\hat{\sigma}_a}\right) \quad \text{avec} \quad \begin{cases} \boldsymbol{\mu}_a = \mathbf{a}^\top \boldsymbol{\mu} \\ \hat{\sigma}_a^2 = \mathbf{a}^\top \boldsymbol{\Delta} \hat{\boldsymbol{\rho}} \boldsymbol{\Delta} \mathbf{a} \end{cases}. \quad (5.130)$$

On conserve les valeurs numériques choisies dans le Chapitre 4, à savoir :

- Fonction de performance  $\phi$  :  $a_1 = a_2 = 2; a_3 = 1$
- Vecteur moyen  $\boldsymbol{\mu}$  :  $\mu_1 = \mu_2 = \mu_3 = 0$
- Matrice de covariance  $\boldsymbol{\Sigma}$  :  $\sigma_1 = \sigma_2 = \sigma_3 = 1; \rho_{12} = 0.5; \rho_{13} = \rho_{23} = 0.3$

Pour ces réglages, un seuil  $T = 11$  conduit à la probabilité de défaillance  $P_f = 2.53 \times 10^{-3}$ .

### 5.5.1.2 À propos des indices obtenus avec le générateur théorique

Pour le moment, on ne s'intéresse qu'à l'application  $\Psi_{\boldsymbol{\tau}}$  qui est définie à l'équation (5.29) et qui associe le résumé exact  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  de la  $d$ -copule  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  à la probabilité de défaillance  $P_f(\mathbf{U}_{\text{obs}})$ . Pour la configuration présentée,  $\Psi_{\boldsymbol{\tau}}$  est très simple à analyser. Au vu du modèle d'apprentissage  $\mathcal{M}_{\text{cop}}$ , la définition (5.26) se transforme en :

$$\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}}) = \left[ \tau_{\mathbf{K}}(V_i, V_j) \right]_{1 \leq i, j \leq d} \quad \text{avec} \quad \mathbf{V} \sim \hat{c}_{\mathcal{N}}(\cdot | \hat{\boldsymbol{\rho}}). \quad (5.131)$$

En vertu de la formule (5.28), on a finalement :

$$\hat{\tau}_{ij} = \tau_{\mathbf{K}}(V_i, V_j) = \frac{2}{\pi} \arcsin(\hat{\rho}_{ij}) \quad \text{ou de façon équivalente} \quad \hat{\rho}_{ij} = \sin\left(\frac{\pi}{2} \hat{\tau}_{ij}\right). \quad (5.132)$$

Il existe donc une formule analytique pour l'opérateur  $\mathcal{E}$  et la notation  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  peut être simplifiée en  $\hat{\boldsymbol{\tau}}(\hat{\boldsymbol{\rho}})$ . Par ailleurs, comme la matrice  $\hat{\boldsymbol{\rho}}$  peut être reconstruite à partir de  $\hat{\boldsymbol{\tau}}(\hat{\boldsymbol{\rho}})$ , on peut affirmer que  $\mathcal{E}$  est inversible. Par conséquent,  $\Psi_{\boldsymbol{\tau}}$  peut être décomposée de la manière suivante :

$$\Psi_{\boldsymbol{\tau}} : \begin{array}{ccc} [-1, 1]^3 & \longrightarrow & [-1, 1]^3 & \longrightarrow & [0, 1] \\ (\hat{\tau}_{12}, \hat{\tau}_{13}, \hat{\tau}_{23}) & \xrightarrow{\mathcal{E}^{-1}} & (\hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23}) & \xrightarrow{\mathcal{R}} & P_f(\hat{\boldsymbol{\rho}}) \end{array}. \quad (5.133)$$

On a donc  $\Psi_{\boldsymbol{\tau}} = \mathcal{E}^{-1} \circ \mathcal{R}$  et cela indique que  $\Psi_{\boldsymbol{\tau}}$  est complètement déterministe. Aucune des deux formes de stochasticité illustrées à la Figure 5.3 n'est présente ici. C'est l'une des rares situations où l'on pourrait écrire  $P_f[\hat{\boldsymbol{\tau}}(\hat{\boldsymbol{\rho}})]$  au lieu de  $P_f(\hat{\boldsymbol{\rho}})$ . En reprenant les notations de la Section 5.3, la collection des  $\delta$ -indices du premier ordre s'écrit :

$$\forall 1 \leq i < j \leq 3, \quad \delta_{ij}^F := \frac{1}{2} \int_{\mathcal{H}_2} |c_{\hat{\mathbf{R}}_{ij}}(u, v) - 1| \, du \, dv. \quad (5.134)$$

Les valeurs numériques listées à la fin de la Section 5.5.1.1 induisent des symétries entre  $\hat{\tau}_{13}$  et  $\hat{\tau}_{23}$ . On s'attend à ce que  $\delta_{13}^F = \delta_{23}^F$  et on se propose de le démontrer. Deux symétries suffisent à obtenir le résultat souhaité :

$$\text{(Sym1) Motif de dépendance avec } \mathbf{U}_3 : \begin{cases} \left( \mathbf{U}_{\text{obs}}^{[\bullet 1]}, \mathbf{U}_{\text{obs}}^{[\bullet 3]} \right) \sim c_{\mathcal{N}}(\cdot | \rho_{13}) \\ \left( \mathbf{U}_{\text{obs}}^{[\bullet 2]}, \mathbf{U}_{\text{obs}}^{[\bullet 3]} \right) \sim c_{\mathcal{N}}(\cdot | \rho_{23}) \end{cases} \quad \text{avec } \rho_{13} = \rho_{23}$$

(Sym2) **Fonction de performance :**  $a_1 = a_2 = 2$

On a vu à la Section 4.3.1.3 que (Sym1) donne la même loi aux triplets  $(\hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23})$  et  $(\hat{\rho}_{12}, \hat{\rho}_{23}, \hat{\rho}_{13})$ . De plus, (Sym2) implique la commutativité de  $P_f(\hat{\boldsymbol{\rho}})$  par rapport aux deux dernières corrélations :

$$\forall (\hat{\rho}_a, \hat{\rho}_b, \hat{\rho}_c) \in [-1, 1]^3, \quad P_f(\hat{\rho}_a, \hat{\rho}_b, \hat{\rho}_c) = P_f(\hat{\rho}_a, \hat{\rho}_c, \hat{\rho}_b) . \quad (5.135)$$

En effet, à partir de la formule (5.130), il est facile de voir que :

$$P_f(\hat{\rho}_a, \hat{\rho}_b, \hat{\rho}_c) = P_f(\hat{\rho}_a, \hat{\rho}_c, \hat{\rho}_b) \iff (a_1 - a_2)(\sigma_1 \rho_{13} - \sigma_2 \rho_{23}) = 0 . \quad (5.136)$$

Par conséquent, (Sym2) permet à l'égalité fonctionnelle (5.135) d'être vérifiée sur l'ensemble du pavé  $[-1, 1]^3$ . Dans la suite, on cherche à montrer que les paires  $(\tilde{X}_{13}, \tilde{Y})$  et  $(\tilde{X}_{23}, \tilde{Y})$  sont de même loi. Pour cela, on commence par s'intéresser aux variables  $\hat{\rho}_{13}$ ,  $\hat{\rho}_{23}$  et  $P_f(\hat{\boldsymbol{\rho}})$ . Pour toute fonction borélienne  $g$ , on peut écrire :

$$\begin{aligned} \mathbb{E} \left[ g(\hat{\rho}_{13}, P_f(\hat{\boldsymbol{\rho}})) \right] &= \mathbb{E} \left[ g(\hat{\rho}_{13}, P_f(\hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23})) \right] \\ &= \mathbb{E} \left[ g(\hat{\rho}_{13}, P_f(\hat{\rho}_{12}, \hat{\rho}_{23}, \hat{\rho}_{13})) \right] \text{ car (Sym2) } \implies (5.135) \\ &= \mathbb{E} \left[ g(\hat{\rho}_{23}, P_f(\hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23})) \right] \text{ car (Sym1) } \implies (\hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23}) \stackrel{\text{loi}}{=} (\hat{\rho}_{12}, \hat{\rho}_{23}, \hat{\rho}_{13}) \\ &= \mathbb{E} \left[ g(\hat{\rho}_{23}, P_f(\hat{\boldsymbol{\rho}})) \right] . \end{aligned} \quad (5.137)$$

On obtient ainsi l'égalité des deux paires de variables. Pour finir la preuve, il suffit de voir que :

$$\begin{aligned} (\hat{\rho}_{13}, P_f(\hat{\boldsymbol{\rho}})) \stackrel{\text{loi}}{=} (\hat{\rho}_{23}, P_f(\hat{\boldsymbol{\rho}})) &\iff (\hat{\tau}_{13}, P_f(\hat{\boldsymbol{\rho}})) \stackrel{\text{loi}}{=} (\hat{\tau}_{23}, P_f(\hat{\boldsymbol{\rho}})) \iff (\tilde{X}_{13}, \tilde{Y}) \stackrel{\text{loi}}{=} (\tilde{X}_{23}, \tilde{Y}) \\ &\implies c_{\tilde{X}_{13}\tilde{Y}}(\cdot) = c_{\tilde{X}_{23}\tilde{Y}}(\cdot) \\ &\implies \delta_{13}^F = \delta_{23}^F . \end{aligned} \quad (5.138)$$

Cette égalité n'est valable qu'à condition que la matrice  $\mathbf{U}_{\text{obs}}$  obéisse à la mesure d'échantillonnage théorique  $(\mu_{\mathbf{U}})^{\otimes N}$ . C'est en effet ce qui sous-tend (Sym1). En pratique, les indices  $\delta_{ij}^F$  rendus par la procédure d'ASF découlent d'une matrice  $\tilde{\mathbf{U}}$  simulée avec la mesure d'échantillonnage empirique et celle-ci ne garantit pas (Sym1).

À l'instar de ce qui est fait à la Section 4.3.1.3, on va utiliser l'égalité  $\delta_{12} = \delta_{23}$  pour régler les hyperparamètres des procédures d'ASF, notamment la taille  $M$  de la séquence d'échantillons qui doit être constituée à l'étape (Bg1). On a recours à une approche empirique pour déterminer la valeur  $M^*$  qui engendre un coût de calcul minimal tout en garantissant un niveau de précision suffisant. On fabrique une série de  $M_{\text{max}} = 5 \times 10^3$  échantillons  $\mathbf{u}_{\text{obs}}^{(k)}$  avec le générateur théorique  $G_{\mathbf{U}}$ . Appliquer  $\mathcal{L}$  puis  $\mathcal{E}$  et  $\mathcal{R}$  donne ensuite naissance à  $M_{\text{max}}$  appariements  $(\tilde{\mathbf{X}}^{(k)}, \tilde{\mathbf{Y}}^{(k)})$ . À partir de ce matériel, on estime les indices  $\delta_{ij}^F$  pour une séquence de valeurs de  $M$  allant de  $M_{\text{min}} = 100$  à  $M_{\text{max}}$ . On renvoie à la Figure 5.11 pour une visualisation des résultats obtenus. On observe que  $\hat{\delta}_{13}^F \approx \hat{\delta}_{23}^F$  dès lors que  $M \geq M^* := 10^3$ . De plus, les différentes courbes d'évolution sont stabilisées à ce stade. La valeur  $M^*$  est en vigueur dans toutes les autres expérimentations numériques de ce chapitre. Par souci de concision, la justification n'est pas reproduite à chaque fois. Néanmoins, lorsqu'un changement intervient dans la fonction de performance  $\phi$  ou dans le modèle d'apprentissage  $\mathcal{M}_{\text{cop}}$ , il est important de s'assurer que brider le schéma numérique  $\mathcal{A}$  à  $M^*$  itérations n'interrompt pas la convergence des estimations. Pour évaluer la robustesse des estimateurs  $\hat{\delta}_{ij}^F$ , on exécute  $N_{\text{rep}} = 50$  fois la procédure d'ASF et les résultats sont synthétisés ci-dessous :

ASF :	$\hat{\delta}_{12}^F$	$\hat{\delta}_{13}^F$	$\hat{\delta}_{23}^F$	.	(5.139)
moyenne :	0.37	0.29	0.29		
écart-type :	0.01	0.01	0.01		
rang :	(1)	(2)	(2)		

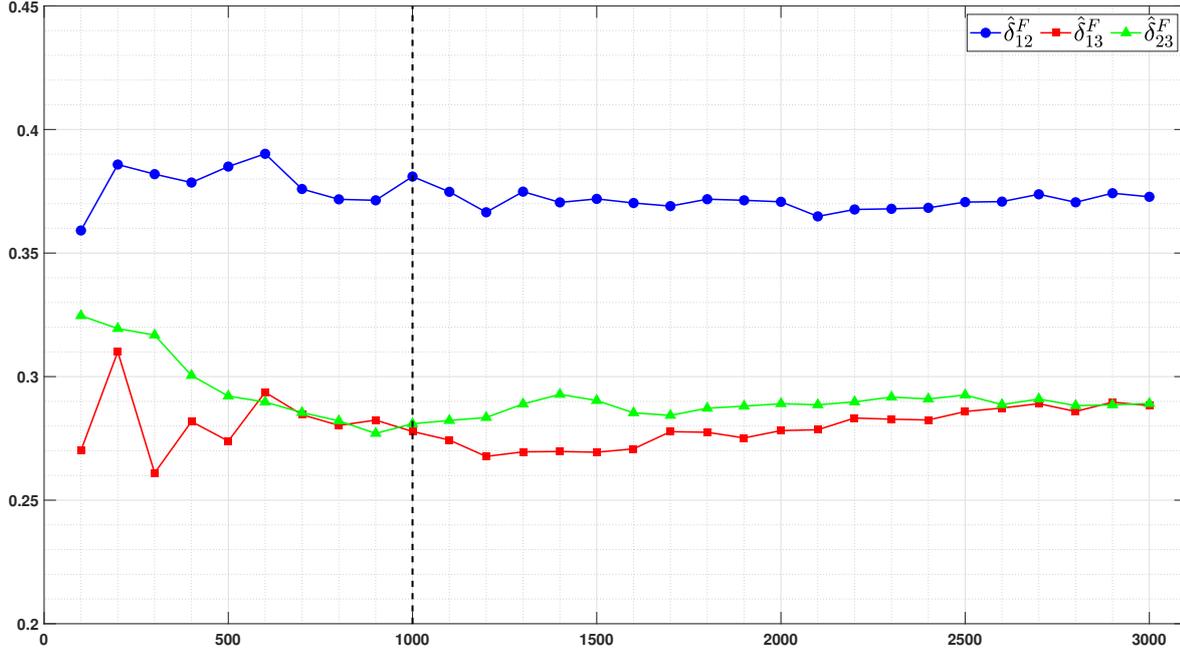


FIGURE 5.11 – Convergence des indices  $\hat{\delta}_{ij}^F$  en fonction du nombre  $M$  d'itérations.

On constate que les écarts-types sont très faibles. Cela conforte l'idée selon laquelle il faut fixer le nombre d'itérations à  $M^*$  et utiliser la méthode non-paramétrique de [GEENENS et collab. \[2017\]](#) pour estimer les copules  $c_{\tilde{R}_{ij}\tilde{S}}$ . Au niveau des classements, on retrouve sans surprise le fait que  $\hat{\delta}_{13} = \hat{\delta}_{23}$ . Par ailleurs, on voit que l'indice dominant est  $\hat{\delta}_{12}^F$ . L'identité du vainqueur de l'ASF doit être mise en relation avec l'expression (5.129) de la sortie  $Y$ . Sachant que l'on a  $\boldsymbol{\mu} = \mathbf{0}$ ,  $\boldsymbol{\Delta} = \mathbf{I}_3$  et  $\mathbf{a}^\top = [2 \ 2 \ 1]$ , le coefficient  $\tau_{12}$  pilote la dépendance entre les deux termes de la somme les plus susceptibles de prendre des grandes valeurs. Comme  $P_f = \mathbb{P}(Y > T)$ , il semble cohérent que  $\hat{\tau}_{12}$  soit identifié par l'ASF comme la variable la plus influente sur  $P_f(\hat{\boldsymbol{\rho}})$  au sein de  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ . Par conséquent, si une acquisition ciblée de données supplémentaires est décidée, elle doit se concentrer sur la paire  $(U_1, U_2)$ . Ce type de scénario est approfondi à la Section 5.5.1.4 où plusieurs stratégies d'assimilation de nouvelles observations sont comparées.

### 5.5.1.3 Étude de la dépendance à l'échantillon initial

On revient désormais à une situation plus réaliste où le générateur théorique  $G_{\mathbf{U}}$  n'est pas disponible. Les indices  $\hat{\delta}_{ij}^F$  sont calculés à partir d'une série de  $M$  répliques *bootstrap*  $\tilde{\mathbf{u}}^{(k)}$  de l'échantillon initial  $\mathbf{u}_{\text{obs}}$ . On veut savoir dans quelle mesure le remplacement de  $\boldsymbol{\mu}_{\mathbf{U}}$  par  $\hat{\boldsymbol{\mu}}[\mathbf{u}_{\text{obs}}]$  peut biaiser les estimations des indices  $\delta_{ij}^F$ . Pour cela, on considère trois échantillons de taille  $N = 300$  qui sont notés  $\mathbf{u}_{\text{obs}}^{(k)}$  avec  $k \in \{1, 2, 3\}$ . Dans chacun des cas, les propriétés des estimateurs  $\hat{\delta}_{ij}^F$  sont évaluées après  $N_{\text{rep}} = 10$  répétitions de l'algorithme d'ASF. Chaque répétition sollicite  $M = 10^3$  appels à la procédure de ré-échantillonnage *bootstrap*  $\mathcal{B}$ . Le Tableau 5.2 confronte les résultats d'ASF obtenus à partir de ces trois échantillons.

Les écarts-types sont aussi faibles que ceux rencontrés lors de l'estimation (5.139) des indices théoriques. Cela confirme que  $M^*$  itérations suffisent à faire converger le schéma numérique  $\mathcal{A}$ , y compris lorsque la série d'échantillons d'apprentissage  $\tilde{\mathbf{u}}^{(k)}$  est une série de répliques *bootstrap* d'un échantillon fixé. On observe que les classements évoluent en fonction de l'échantillon  $\mathbf{u}_{\text{obs}}^{(k)}$  fourni en entrée de l'algorithme d'ASF :

$$\mathbf{u}_{\text{obs}}^{(1)} \Rightarrow \hat{\delta}_{12}^F > \hat{\delta}_{13}^F \approx \hat{\delta}_{23}^F \quad ; \quad \mathbf{u}_{\text{obs}}^{(2)} \Rightarrow \hat{\delta}_{12}^F \approx \hat{\delta}_{13}^F > \hat{\delta}_{23}^F \quad ; \quad \mathbf{u}_{\text{obs}}^{(3)} \Rightarrow \hat{\delta}_{12}^F \approx \hat{\delta}_{23}^F > \hat{\delta}_{13}^F \quad . \quad (5.140)$$

		ASF exécutée avec $\mathcal{B}$ :	$\hat{\delta}_{12}^F$	$\hat{\delta}_{13}^F$	$\hat{\delta}_{23}^F$
Échantillons	$\mathbf{u}_{\text{obs}}^{(1)}$	moyenne :	0.38	0.29	0.29
		écart-type :	0.01	0.02	0.01
		rang :	(1)	(2)	(2)
	$\mathbf{u}_{\text{obs}}^{(2)}$	moyenne :	0.35	0.36	0.28
		écart-type :	0.02	0.01	0.02
		rang :	(2)	(1)	(3)
	$\mathbf{u}_{\text{obs}}^{(3)}$	moyenne :	0.35	0.31	0.37
		écart-type :	0.01	0.01	0.01
		rang :	(2)	(3)	(1)

TABLEAU 5.2 – Estimation des  $\delta$ -indices de Borgonovo pour trois échantillons distincts.

De par les valeurs moyennes et les classements qu’il engendre, l’échantillon  $\mathbf{u}_{\text{obs}}^{(1)}$  semble très représentatif du comportement de la matrice aléatoire  $\mathbf{U}_{\text{obs}} \sim (\mu_{\mathbf{U}})^{\otimes N}$ . Au contraire, l’utilisation de la mesure empirique construite à partir de  $\mathbf{u}_{\text{obs}}^{(2)}$  (resp.  $\mathbf{u}_{\text{obs}}^{(3)}$ ) introduit un biais qui n’impacte pas  $\hat{\delta}_{12}^F$  mais qui fait émerger  $\hat{\delta}_{13}^F$  (resp.  $\hat{\delta}_{23}^F$ ) au rang d’indice dominant. On retrouve la remarque faite à la Section 4.3.1.4 sur les problèmes causés par la mise en œuvre d’une approche “*data-driven*”. Par ailleurs, cette difficulté augmente lorsque le nombre  $N$  d’observations diminue. Néanmoins, il ne faut pas dénigrer la procédure d’ASF qui est développée dans ce chapitre car on peut considérer qu’elle remplit les objectifs formulés. En effet, il faut remarquer que les valeurs théoriques (5.139) des différents indices  $\delta_{ij}^F$  sont globalement très proches. Cela indique que les fonctions aléatoires que sont les copules bidimensionnelles  $\hat{c}_{\mathcal{N}}(\cdot | \hat{\rho}_{ij})$  ont pratiquement toutes la même influence sur  $P_f(\hat{\rho})$ . Aucune de ces copules n’exerce une influence qui occulte les autres et, réciproquement, aucune de ces copules ne peut être négligée. Les résultats consignés dans le Tableau 5.2 restituent bien cette information car tous les indices  $\hat{\delta}_{ij}^F$  sont compris dans l’intervalle  $]0.25, 0.40[$  qui est finalement assez restreint au regard de la plage  $[0, 1]$  délimitée par les bornes extrêmes. En guise de conclusion, on peut dire que la méthodologie “*data-driven*”, bien qu’elle présente des limites indéniables, préserve les informations les plus centrales de l’ASF. Il faut toutefois rester prudent et ne pas tirer de conclusion définitive à partir du classement lorsque les indices sont proches. Seule l’acquisition de nouvelles observations de la  $d$ -copule permettrait d’affiner le classement et de limiter le caractère “*data-driven*”.

#### 5.5.1.4 Assimilation de données supplémentaires

Pour vérifier que chaque indice  $\hat{\delta}_{ij}$  quantifie correctement l’influence de  $\hat{\tau}_{ij}$  sur  $P_f(\hat{\rho})$ , on va raffiner l’estimation de  $\hat{c}_{\mathcal{N}}(\cdot | \hat{\rho}_{ij})$  avec des observations supplémentaires de la paire  $(U_i, U_j)$  puis exécuter une seconde ASF. La réduction de la variabilité d’apprentissage véhiculée par  $\hat{c}_{\mathcal{N}}(\cdot | \hat{\rho}_{ij})$  doit s’accompagner d’une diminution de l’indice  $\hat{\delta}_{ij}^F$  associé. Dans le contexte de ce cas test, l’analyse de fiabilité  $\mathcal{R}$  est déterministe et le modèle d’apprentissage  $\mathcal{M}_{\text{cop}}$  coïncide avec le modèle de copule sous-jacent. Par conséquent, la seule incertitude entachant  $P_f(\hat{\rho})$  est celle liée au manque de données. On s’attend donc à ce que l’assimilation de données supplémentaires ait un impact manifeste sur la distribution de  $P_f(\hat{\rho})$  et que cela se traduise par des changements ostensibles au niveau du classement des  $\delta$ -indices. Par ailleurs, on a vu à la Section 5.4 qu’il fallait opérer deux changements au sein de l’algorithme d’ASF initial :

- (a) Remplacer le mécanisme de ré-échantillonnage initial  $\mathcal{B}$  par son extension  $\mathcal{B}_{ij}$ .
- (b) Transformer le modèle d’apprentissage  $\mathcal{M}_{\text{cop}}$  en une version  $\mathcal{M}_{\text{cop}}^{(ij)}$  capable de digérer les données fournies par un processus d’acquisition ciblée.

Ici, il ne sera pas question d’implémenter le second point car il existe une stratégie d’assimilation beaucoup plus simple. Elle consiste à exploiter l’expression analytique de  $P_f(\hat{\rho})$ . Après avoir réuni

des observations supplémentaires de la paire ciblée, les coefficients de la matrice  $\hat{\rho}$  sont calculés à partir de (5.106) puis directement réinjectés dans la formule (5.130) pour obtenir une mise à jour de la prédiction du risque de défaillance. On fait abstraction de la notion de copule gaussienne et on envisage la probabilité de défaillance comme une fonction de paramètres estimés  $\hat{\rho}_{ij}$  sur lesquels on peut directement intervenir. Pour cette raison, on choisit de ne pas mettre en application le modèle d'apprentissage  $\mathcal{M}_{\text{cop}}^{(ij)}$  présenté à la Section 5.4.2.1 et basé sur l'estimation d'une copule R-vigne dont toutes les familles sont gaussiennes.

Le Tableau 5.3 présente les résultats obtenus après l'expérimentation de plusieurs scénarios d'assimilation de données supplémentaires. Toute l'étude est menée à partir du même échantillon  $\mathbf{u}_{\text{obs}}^{(1)}$  emprunté au paragraphe précédent. Les résultats de l'ASF initiale pour cet échantillon sont indiqués au premier étage du Tableau 5.3 et on rappelle qu'il était question d'un échantillon particulièrement représentatif du comportement moyen de  $\mathbf{U}_{\text{obs}}$ . Le deuxième étage indique les résultats de la seconde ASF pour un nombre  $N'$  croissant d'observations de la paire  $(U_1, U_2)$ . C'est le scénario d'assimilation naturel sachant que  $\hat{\delta}_{12}$  est initialement l'indice dominant. La diminution attendue de l'indice  $\hat{\delta}_{12}$  n'est pas vraiment visible lorsqu'on se contente de collecter  $N' = 50$  nouvelles observations de la paire  $(U_1, U_2)$ . Il faut attendre que  $\mathbf{u}_{\text{extra}}^{[* (ij)]}$  soit de taille  $N' \geq 200$  pour que l'indice  $\hat{\delta}_{12}^F$  soit dépassé par  $\hat{\delta}_{13}^F$  et  $\hat{\delta}_{23}^F$ . On constate à l'occasion que l'augmentation de ces deux indices s'opère exactement au même rythme. Aux étages inférieurs, des phénomènes analogues sont observables après collection puis assimilation de données supplémentaires pour les paires  $(U_1, U_3)$  et  $(U_2, U_3)$ . Les résultats obtenus pour ce cas test apportent des éléments de réponse au Problème 5.3 de réduction de la variabilité affectant  $P_f(\mathbf{U}_{\text{obs}})$ . Il semble en effet possible d'agir efficacement sur la dispersion de  $P_f(\mathbf{U}_{\text{obs}})$  sans qu'il ne soit nécessaire de solliciter un volume massif d'observations conjointes de  $\mathbf{U}$ . Toutefois, on doit garder à l'esprit la simplicité de la configuration examinée. Dans la Section 5.5.2, un problème plus épineux est abordé où  $\hat{\mathbf{t}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  n'est qu'un résumé partiel de  $\hat{\mathbf{c}}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  et où l'analyse de sensibilité requiert l'utilisation d'un algorithme de simulation  $\mathcal{R}_w$ .

### 5.5.1.5 Comparaison des résultats obtenus avec $\Psi_{\tau}$ et $\tilde{\Psi}_{\tau}$

Après avoir étudié le problème d'ASF posé par  $\Psi_{\tau}$ , on considère celui posé par  $\tilde{\Psi}_{\tau}$ . Comme indiqué à l'équation (5.39), il s'agit d'analyser la sensibilité de  $P_f(\hat{\rho})$  aux variations des coefficients de la matrice empirique  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$ . L'extraction analytique  $\mathcal{E}$  de  $\hat{\mathbf{t}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  est remplacée par l'extraction empirique  $\tilde{\mathcal{E}}$  de  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$ . Le caractère aléatoire de la proximité entre les matrices  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$  et  $\hat{\mathbf{t}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  rend l'application  $\tilde{\Psi}_{\tau}$  stochastique. La Figure 5.12 permet de comparer la nature des applications  $\Psi_{\tau}$  et  $\tilde{\Psi}_{\tau}$  dans la configuration triviale de ce cas test.

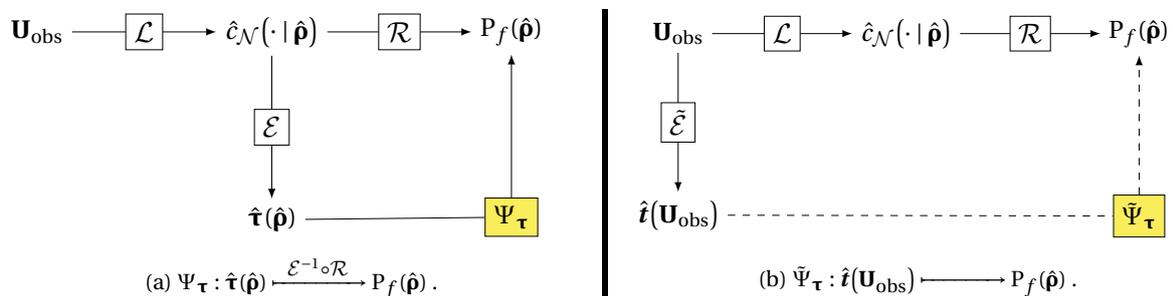


FIGURE 5.12 – Décomposition schématique des applications  $\Psi_{\tau}$  et  $\tilde{\Psi}_{\tau}$ .

Pour quantifier la stochasticité de  $\tilde{\Psi}_{\tau}$ , on utilise l'indicateur  $\xi^F$  qui est défini à la Section 5.3.2.2 comme le  $\delta$ -indice d'ordre total des variables observables. On exécute à  $N_{\text{rep}} = 10$  reprises l'algorithme d'ASF complet incluant les étapes (Bg1) à (Bg6) afin de disposer d'estimations de tous les indices  $\delta_{ij}^F$  et de l'indicateur  $\xi^F$ . Les résultats sont présentés dans le Tableau 5.4 où l'on rappelle à titre de comparaison les résultats de l'ASF conduite pour  $\Psi_{\tau}$ . On note une grande proximité entre les deux collections d'indices  $\hat{\delta}_{ij}^F$ , ce qui incite à penser que les matrices  $\hat{\mathbf{t}}(\hat{\rho})$  et  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$  offrent

ASF exécutée avec $\mathcal{B}$ :				$\hat{\delta}_{12}^F$	$\hat{\delta}_{13}^F$	$\hat{\delta}_{23}^F$
	$\mathbf{u}_{\text{obs}}^{(1)}$		moyenne :	0.38	0.29	0.29
			écart-type :	0.01	0.02	0.01
			rang :	(1)	(2)	(2)
ASF exécutée avec $\mathcal{B}_{12}$ :				$\hat{\delta}_{12}^F$	$\hat{\delta}_{13}^F$	$\hat{\delta}_{23}^F$
Paire $\mathbf{U}_{12}$	$\mathbf{u}_{\text{obs}}^{(1)} \cup \mathbf{u}_{\text{extra}}^{[\bullet(12)]}$	$N' = 50$	moyenne :	0.36	0.30	0.29
			écart-type :	0.01	0.02	0.01
			rang :	(1)	(2)	(2)
		$N' = 200$	moyenne :	0.30	0.32	0.32
			écart-type :	0.02	0.01	0.01
			rang :	(2)	(1)	(1)
		$N' = 10^3$	moyenne :	0.19	0.38	0.38
			écart-type :	0.01	0.01	0.01
			rang :	(2)	(1)	(1)
ASF exécutée avec $\mathcal{B}_{13}$ :				$\hat{\delta}_{12}^F$	$\hat{\delta}_{13}^F$	$\hat{\delta}_{23}^F$
Paire $\mathbf{U}_{13}$	$\mathbf{u}_{\text{obs}}^{(1)} \cup \mathbf{u}_{\text{extra}}^{[\bullet(13)]}$	$N' = 50$	moyenne :	0.39	0.26	0.28
			écart-type :	0.01	0.02	0.01
			rang :	(1)	(3)	(2)
		$N' = 200$	moyenne :	0.40	0.22	0.27
			écart-type :	0.01	0.01	0.01
			rang :	(1)	(3)	(2)
		$N' = 10^3$	moyenne :	0.45	0.14	0.26
			écart-type :	0.01	0.01	0.02
			rang :	(1)	(3)	(2)
ASF exécutée avec $\mathcal{B}_{23}$ :				$\hat{\delta}_{12}^F$	$\hat{\delta}_{13}^F$	$\hat{\delta}_{23}^F$
Paire $\mathbf{U}_{23}$	$\mathbf{u}_{\text{obs}}^{(1)} \cup \mathbf{u}_{\text{extra}}^{[\bullet(23)]}$	$N' = 50$	moyenne :	0.39	0.29	0.26
			écart-type :	0.01	0.01	0.01
			rang :	(1)	(2)	(3)
		$N' = 200$	moyenne :	0.41	0.27	0.22
			écart-type :	0.01	0.01	0.01
			rang :	(1)	(2)	(3)
		$N' = 10^3$	moyenne :	0.46	0.27	0.13
			écart-type :	0.01	0.02	0.01
			rang :	(1)	(2)	(3)

TABLEAU 5.3 – Impact de l’assimilation de données supplémentaires sur le classement des  $\delta$ -indices.

approximativement la même description du transfert d'incertitude entre  $\hat{c}_{\mathcal{N}}(\cdot | \hat{\boldsymbol{\rho}})$  et  $P_f(\hat{\boldsymbol{\rho}})$ . C'est d'autant plus intéressant que l'indicateur de stochasticité vaut en moyenne  $\hat{\xi}^F \approx 0.6 < 1$  et signale ainsi qu'une part de la variabilité de  $P_f(\hat{\boldsymbol{\rho}})$  ne peut pas être décrite par les coefficients empiriques  $\hat{t}_{ij}$ . Par ailleurs, l'écart-type de l'estimateur de  $\hat{\xi}^F$  est aussi faible que celui des estimateurs  $\hat{\delta}_{ij}^F$ , alors que l'obtention de  $\hat{\xi}^F$  implique l'apprentissage non-paramétrique des copules  $c_{\mathbf{R}} \in \mathcal{C}_3$  et  $c_{\mathbf{R}\tilde{\mathbf{S}}} \in \mathcal{C}_4$ . Cette stabilité entérine le choix des méthodes numériques implantées dans  $\mathcal{A}$ .

### 5.5.1.6 Bilan

Dans cette section, on a étudié une configuration triviale qui rassemble une copule théorique  $c_{\mathbf{U}}$  gaussienne, un modèle d'apprentissage  $\mathcal{M}_{\text{cop}}$  paramétrique gaussien et une fonction de performance  $\phi$  linéaire. Dans cette situation extrêmement favorable, on a vérifié le bien-fondé des idées impulsées depuis le début du chapitre. On retient les points suivants :

- L'algorithme d'ASF doit inclure au moins  $M = 10^3$  itérations pour garantir une convergence numérique suffisante des estimateurs  $\hat{\delta}_{ij}^F$  et  $\hat{\xi}^F$ .
- L'utilisation d'une approche "orientée copule" au sein du schéma  $\mathcal{A}$  doit s'accompagner d'un déploiement d'outils non-paramétriques :
  - ⇒ La méthode des noyaux transformés de [GEENENS et collab. \[2017\]](#) pour les indices  $\delta_{ij}^F$ .
  - ⇒ Les copules R-vignes non-paramétriques de [NAGLER et CZADO \[2016\]](#) pour  $\xi^F$ .
- L'assimilation de données supplémentaires peut être envisagée et mise en œuvre.
- Lorsque la matrice  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  est inaccessible, elle peut être remplacée par  $\hat{\boldsymbol{\tau}}(\mathbf{U}_{\text{obs}})$ .

## 5.5.2 Retour sur le cas test non-linéaire

### 5.5.2.1 Rappels

Dans la Section 4.5.2, la probabilité de défaillance  $\mathbb{P}(Y > T)$  découlait des éléments suivants :

$$Y = \phi(\mathbf{X}) = a_1X_1 + a_2X_2 + a_3X_3 + a_4X_1X_2 \quad \text{et} \quad \mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad . \quad (5.141)$$

Pour l'apprentissage de la distribution d'entrée, le modèle global  $\mathcal{M}$  consistait à estimer la copule  $c_{\mathbf{U}}$  avec un modèle  $\mathcal{M}_{\text{cop}}$  paramétrique gaussien et chacune des marginales  $f_{X_j}$  avec un modèle non-paramétrique  $\mathcal{M}_j$  de type KDE. Toutes les composantes fonctionnelles véhiculaient une incertitude épistémique liée au manque de données. De plus, l'utilisation d'une approche KDE pour les marginales donnait naissance à une incertitude supplémentaire liée à la méconnaissance de la famille paramétrique sous-jacente. Ici, on reprend sous un autre angle le même cas test.

- Les lois marginales sont supposées connues et fixées :

$$\forall 1 \leq j \leq 3, \quad X_j \sim \mathcal{N}(\mu_j, \sigma_j^2) \quad \text{avec} \quad \mu_j = 3 \quad \text{et} \quad \sigma_j = 1. \quad (5.142)$$

- La connaissance de  $c_{\mathbf{U}}$  se limite à celle d'un échantillon  $\mathbf{u}_{\text{obs}} \in [0, 1]^{N \times 3}$ .

Tous les autres paramètres du problème conservent leurs valeurs initiales :

$$a_1 = a_2 = 2 \quad ; \quad a_3 = a_4 = 1 \quad ; \quad T = 50 \quad . \quad (5.143)$$

		ASF exécutée avec $\mathcal{B}$ :						
		$\hat{\delta}_{12}^F$	$\hat{\delta}_{13}^F$	$\hat{\delta}_{23}^F$	$\hat{\xi}^F$			
Échantillon	$\mathbf{u}_{\text{obs}}^{(1)}$	Résumé de $c_{\mathbf{U}}$	$\mathcal{E}$	moyenne :	0.38	0.29	0.29	1
				écart-type :	0.01	0.02	0.01	×
				rang :	(1)	(2)	(3)	×
		$\tilde{\mathcal{E}}$	moyenne :	0.33	0.26	0.25	0.60	
			écart-type :	0.01	0.01	0.02	0.01	
			rang :	(1)	(2)	(3)	×	

TABEAU 5.4 – Comparaison des résultats obtenus pour les deux sous-problèmes d'ASF.

### 5.5.2.2 Présentation de l'échantillon initial

Dans ce cas test, on souhaite n'avoir aucune connaissance préalable sur la distribution  $\mu_{\mathbf{U}}$ . La démarche idéale consiste à trouver un phénomène physique dont la variabilité spatiale ou temporelle engendre une réelle dépendance statistique entre les variables permettant de le décrire. Pour cela, on doit récupérer un échantillon de référence dans un article où l'estimation d'une distribution de copule est une problématique de première importance. On choisit de se procurer le jeu de données "uranium" évoqué dans [KRAUS et CZADO \[2017\]](#) et disponible dans le package R `copula`<sup>6</sup>. Il a été constitué à partir de  $N_0 = 655$  prélèvements d'eau à proximité de la ville de Grand Junction dans l'État du Colorado. Chaque prélèvement est l'occasion de mesurer la concentration de 7 composés chimiques dans le fleuve Colorado :

- Uranium (U)
- Lithium (Li)
- Cobalt (Co)
- Potassium (K)
- Césium (Cs)
- Scandium (Sc)
- Titanium (Ti)

Le jeu de données "uranium" fournit le logarithme de chacune de ces concentrations. Pour les besoins de notre étude, seuls trois composés chimiques sont nécessaires. Très naturellement, on cherche à ce que les différentes paires de variables présentent des corrélations importantes. À ce titre, les log-concentrations en Cobalt, Scandium et Titanium sont isolées des autres variables afin de former un sous-échantillon  $\mathbf{z}_{\text{tot}} \in \mathbb{R}^{N_0 \times 3}$  du vecteur :

$$\mathbf{Z} = \left[ \log([\text{Co}]), \log([\text{Sc}]), \log([\text{Ti}]) \right]^T. \quad (5.144)$$

La visualisation de  $\mathbf{z}_{\text{tot}}$  à la Figure 5.13 donne un aperçu du motif de dépendance régissant les log-concentrations des trois composés retenus. Dans la suite, on pose  $\mathbf{c}_{\mathbf{U}} = \mathbf{c}_{\mathbf{Z}}$  afin de se placer dans une configuration où la copule est parfaitement inconnue. Comme les lois marginales  $\mu_{Z_j}$  sont indésirables, on utilise la formule de renormalisation empirique (SP2) pour construire un échantillon  $\mathbf{u}_{\text{tot}}$  à partir de  $\mathbf{z}_{\text{tot}}$ . L'échantillon final est alors constitué de  $N_0$  pseudo-observations de  $\mathbf{c}_{\mathbf{U}}$ . Par ailleurs, en prévision des scénarios d'assimilation de données, on pose  $N_0 = N + N'$  et on partitionne  $\mathbf{u}_{\text{tot}}$  en deux sous-échantillons (avec  $N = 200$  et  $N' = 455$ ) :

$$\mathbf{u}_{\text{obs}} \in [0, 1]^{200 \times 3} \quad \text{et} \quad \mathbf{u}_{\text{extra}} \in [0, 1]^{455 \times 3}. \quad (5.145)$$

### 5.5.2.3 Choix d'un modèle d'apprentissage

Comme la nature de la copule sous-jacente est parfaitement inconnue, il faut sélectionner un modèle d'apprentissage  $\mathcal{M}_{\text{cop}}$  pertinent, tout en étant conscient qu'il n'est pas possible d'empêcher la création d'une erreur liée à la méconnaissance de  $\mathbf{c}_{\mathbf{U}}$ . Comme les motifs de dépendance bidimensionnels représentés à la Figure 5.13 sont assez classiques, on décide d'apprendre une copule R-vigne. On rappelle qu'en dimension  $d = 3$ , l'ensemble des R-vignes est tout autant équivalent à celui des C-vignes que des D-vignes. Comme la matrice  $\hat{\mathbf{t}}(\mathbf{u}_{\text{obs}})$  implique les coefficients  $\hat{t}_{12} = 0.54$ ,  $\hat{t}_{13} = 0.36$  et  $\hat{t}_{23} = 0.44$ , la résolution du problème d'arbre couvrant de poids maximal conduit à prendre  $\mathbf{E}_1 = \{\{\{1\}, \{2\}\}; \{\{2\}, \{3\}\}\}$  et cela entraîne  $\mathbf{E}_2 = \{\{\{1, 2\}, \{2, 3\}\}\}$ . De plus, une descente de la séquence d'arbres  $\mathbf{T}_{\text{vg}} := \{\mathbf{T}_1, \mathbf{T}_2\}$  montre que la copule de Student est la famille paramétrique la plus adaptée à la modélisation de chacune des trois copules bidimensionnelles impliquées dans la factorisation en R-vigne. Il semble donc raisonnable de figer  $\mathbf{F}_{\text{vg}} := \{t, t, t\}$ . On note  $\mathcal{M}_{\text{cop}}^{(A)}$  le modèle d'apprentissage consistant à estimer par MV les paramètres  $\boldsymbol{\theta}_{\text{vg}}$  d'une copule R-vigne dont le squelette  $(\mathbf{T}_{\text{vg}}, \mathbf{F}_{\text{vg}})$  est imposé. La copule  $\hat{c}_{\mathbf{U}}(\cdot \mid \mathcal{M}_{\text{cop}}^{(A)}, \mathbf{U}_{\text{obs}})$  produite par la phase d'apprentissage peut alors se noter  $\hat{c}_{\text{vg}}(\cdot \mid \mathbf{T}_{\text{vg}}, \mathbf{F}_{\text{vg}}, \hat{\boldsymbol{\theta}}_{\text{vg}})$  avec  $\hat{\boldsymbol{\theta}}_{\text{vg}}$  qui agglomère 6 paramètres :

$$\hat{\boldsymbol{\theta}}_{\text{vg}} = \left[ \hat{\boldsymbol{\rho}} \mid \hat{\boldsymbol{\nu}} \right] = \left[ \hat{\rho}_a \quad \hat{\rho}_b \quad \hat{\rho}_c \mid \hat{\nu}_a \quad \hat{\nu}_b \quad \hat{\nu}_c \right]. \quad (5.146)$$

6. <https://cran.r-project.org/package=copula>

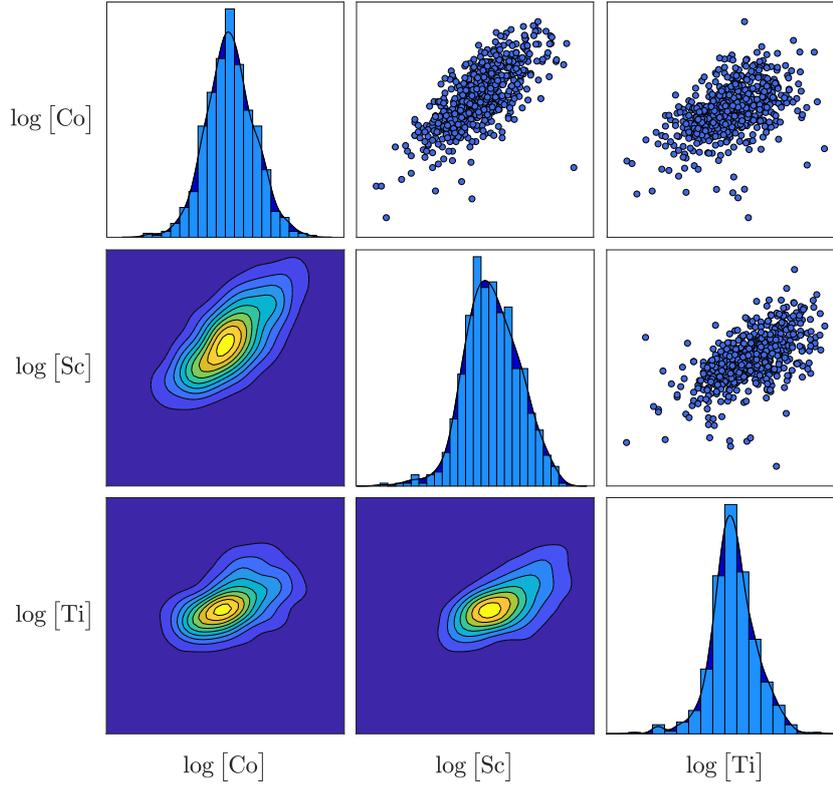
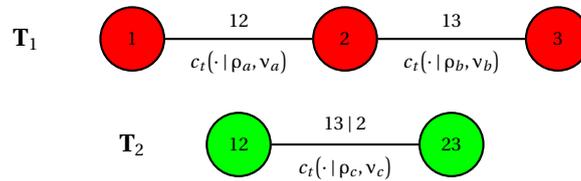


FIGURE 5.13 – Visualisation des lois marginales et bidimensionnelles de  $\mathbf{Z}$  à partir de  $\mathbf{z}_{\text{tot}}$ .

Utiliser  $\mathcal{M}_{\text{cop}}^{(A)}$  revient en fait à calibrer à  $\mathbf{u}_{\text{obs}}$  la densité  $c_{\text{vg}}(\cdot \mid \boldsymbol{\rho}, \mathbf{v})$  de la copule R-vigne ci-dessous :



Toutefois, il faut bien noter que  $c_{\text{vg}}(\cdot \mid \boldsymbol{\rho}, \mathbf{v})$  n'est pas équivalente à la densité d'une copule de Student de dimension  $d = 3$ . En effet, le paramétrage de cette famille implique trois coefficients de corrélation mais un unique nombre  $\nu_0$  de degrés de liberté. L'idée du modèle  $\mathcal{M}_{\text{cop}}^{(A)}$  est de coller autant que possible aux informations qui peuvent être tirées de l'échantillon brut  $\mathbf{u}_{\text{tot}}$ . Toutefois, de par le contexte "small data", il est tout à fait légitime de penser que la connaissance de  $\mathbf{u}_{\text{tot}}$  n'est pas suffisante pour figer définitivement  $(\mathbf{T}_{\text{vg}}, \mathbf{F}_{\text{vg}})$ . À chaque nouvel échantillon, on est libre d'exécuter l'algorithme de [DISSMANN et collab. \[2013\]](#). Le modèle d'apprentissage est alors noté  $\mathcal{M}_{\text{cop}}^{(B)}$  et la densité associée  $\hat{c}_{\mathbf{U}}(\cdot \mid \mathcal{M}_{\text{cop}}^{(B)}, \mathbf{U}_{\text{obs}})$  peut s'écrire  $\hat{c}_{\text{vg}}(\cdot \mid \hat{\mathbf{T}}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}}, \hat{\boldsymbol{\theta}}_{\text{vg}})$ .

#### 5.5.2.4 Choix d'un algorithme de fiabilité

Si on cherche à estimer  $P_f(\mathbf{u}_{\text{obs}})$ , une approche de type CMC convient parfaitement :

$$\hat{P}_f(\mathbf{u}_{\text{obs}}) = \frac{1}{N_s} \sum_{k=1}^{N_s} \mathbb{1}_{\{Y^{(i)} > T\}} \quad \text{avec} \quad \begin{cases} Y^{(i)} = \phi(\mathbf{X}^{(i)}) \\ X_j^{(i)} = \sigma_j \Phi^{-1}(U_j^{(i)}) + \mu_j \\ \mathbf{U}^{(i)} \sim \hat{c}_{\mathbf{U}}(\cdot \mid \mathcal{M}_{\text{cop}}, \mathbf{u}_{\text{obs}}) \end{cases} \quad (5.147)$$

Pour  $\mathcal{M}_{\text{cop}}^{(A)}$ , on obtient ainsi  $\hat{P}_f(\mathbf{u}_{\text{obs}}) \approx 1.2 \times 10^{-2}$  en prenant  $N_s = 10^5$ . Le nombre de sorties  $Y^{(k)}$  simulées n'est pas un problème étant donné que  $\phi$  est analytique. Toutefois, même dans cette

situation très simple où l'évaluation de  $\phi$  est peu coûteuse, répéter l'analyse de fiabilité (5.147) à  $M = 10^3$  reprises représente une perte de temps importante. Le recours à d'autres algorithmes de simulation permet de réduire drastiquement la charge de calcul induite par l'estimation d'une probabilité de défaillance. Toutefois, lorsqu'on emploie cette gamme de techniques, il faut surveiller avec vigilance le niveau de précision des estimations ainsi que l'amplitude de la stochasticité supplémentaire qui est introduite. Pour ce cas test, afin de s'inscrire dans la continuité de ce qui est fait à la Section 4.5.2.2,  $\mathcal{R}_w$  désigne la méthode d'échantillonnage préférentiel GIS. On rappelle qu'il s'agit de calculer le point de conception  $\mathbf{x}^* \in \mathbb{R}^3$  avec FORM puis de construire une densité auxiliaire gaussienne centrée sur ce point. Ici, la difficulté réside dans l'application de FORM pour une densité conjointe  $\hat{f}_{\mathbf{X}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{u}_{\text{obs}})$  impliquant une copule R-vigne :

$$\hat{f}_{\mathbf{X}}(\mathbf{x} | \mathcal{M}_{\text{cop}}, \mathbf{u}_{\text{obs}}) = \left[ \prod_{j=1}^3 f_{\mathcal{N}}(\mathbf{x} | \mu_j, \sigma_j) \right] \times \hat{c}_{\mathbf{U}}(\mathbf{u} | \mathcal{M}_{\text{cop}}, \mathbf{u}_{\text{obs}}) \quad (5.148)$$

avec  $\mathbf{u} = \left[ \Phi\left(\frac{x_j - \mu_j}{\sigma_j}\right) \right]_j$  et  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{u}_{\text{obs}}) = \begin{cases} \hat{c}_{\text{vg}}(\cdot | \hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\nu}}) & \text{pour } \mathcal{M}_{\text{cop}}^{(\text{A})} \\ \hat{c}_{\text{vg}}(\cdot | \hat{\mathbf{T}}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}}, \hat{\boldsymbol{\theta}}_{\text{vg}}) & \text{pour } \mathcal{M}_{\text{cop}}^{(\text{B})} \end{cases}.$

Il faut trouver une transformation isoprobabiliste  $\mathcal{T}$  qui aboutit au vecteur gaussien standard  $\mathbf{W}$  après une intervention sur le vecteur  $\mathbf{X}$  de densité conjointe  $\mathbf{x} \mapsto \hat{f}_{\mathbf{X}}(\mathbf{x} | \mathcal{M}_{\text{cop}}, \mathbf{u}_{\text{obs}})$ . Dans la suite, on adopte des notations plus synthétiques pour alléger les équations :

$$\begin{aligned} \bullet \quad \forall \mathbf{x} \in \mathbb{R}^3, \quad \hat{g}_1(\mathbf{x}) &= \hat{f}_{\mathbf{X}}(\mathbf{x} | \mathcal{M}_{\text{cop}}, \mathbf{u}_{\text{obs}}) & ; \\ \bullet \quad \forall \mathbf{u} \in \mathcal{H}_3, \quad \hat{g}_2(\mathbf{u}) &= \hat{c}_{\mathbf{U}}(\mathbf{u} | \mathcal{M}_{\text{cop}}, \mathbf{u}_{\text{obs}}) & ; \\ \bullet \quad \forall \mathbf{v} \in \mathcal{H}_3, \quad g_3(\mathbf{v}) &= \mathbb{1}_{\mathcal{H}_3}(\mathbf{v}) = \prod_{j=1}^3 \mathbb{1}_{[0,1]}(v_j) & ; \\ \bullet \quad \forall \mathbf{w} \in \mathbb{R}^3, \quad g_4(\mathbf{w}) &= \varphi_3(\mathbf{w}) = \prod_{j=1}^3 \varphi_1(w_j) & . \end{aligned} \quad (5.149)$$

La transformation isoprobabiliste  $\mathcal{T}$  recherchée peut être décomposée de la manière suivante :

$$\begin{array}{ccccccc} \hat{\mathcal{T}} : & \mathbf{x} & \longrightarrow & \mathbf{u} & \longrightarrow & \mathbf{v} & \longrightarrow & \mathbf{w} \\ & \mathbf{X} \sim \hat{g}_1(\cdot) & \xrightarrow{\mathcal{T}_a} & \mathbf{U} \sim \hat{g}_2(\cdot) & \xrightarrow{\hat{\mathcal{T}}_b} & \mathbf{V} \sim g_3(\cdot) & \xrightarrow{\mathcal{T}_c} & \mathbf{W} \sim g_4(\cdot) \end{array}, \quad (5.150)$$

et chaque transformation intermédiaire exécute une tâche bien spécifique :

- $\mathcal{T}_a$  : Transformation des lois marginales  $\mathcal{N}(\mu_j, \sigma_j)$  en des lois marginales  $\mathcal{U}([0, 1])$ .
- $\hat{\mathcal{T}}_b$  : Élimination de la structure de dépendance portée par  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{u}_{\text{obs}})$ .
- $\mathcal{T}_c$  : Transformation des lois marginales  $\mathcal{U}([0, 1])$  en des lois marginales  $\mathcal{N}(0, 1)$ .

Si  $\mathcal{T}_a$  et  $\mathcal{T}_c$  découlent d'une application triviale de la méthode de la transformée inverse, la mise en place de  $\hat{\mathcal{T}}_b$  fait appel à des outils bien plus complexes. Sachant que  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{u}_{\text{obs}})$  est une copule R-vigne, la transformation de NATAF [1962] ou la généralisation qu'en proposent LEBRUN et DUTFOY [2009] sont dépassées. Seule la transformation de ROSENBLATT [1952] qui a été rencontrée à la Section 2.2.1.4 peut être appliquée. Le lien entre les vecteurs  $\mathbf{U}$  et  $\mathbf{V}$  est établi par :

$$\mathbf{V} := \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = \hat{\mathcal{T}}_b(\mathbf{U}) \quad \text{avec} \quad \begin{cases} V_{k_1} = U_{k_1} \\ V_{k_2} = F_{k_2|k_1}(U_{k_2} | U_{k_1}) \\ V_{k_3} = F_{k_3|k_1 k_2}(U_{k_3} | U_{k_1}, U_{k_2}) \end{cases}, \quad (5.151)$$

où  $\{k_1, k_2, k_3\}$  est une permutation de  $\{1, 2, 3\}$  qui est choisie de manière à ce que les fonctions de répartition conditionnelles  $F_{k_2|k_1}$  et  $F_{k_3|k_1, k_2}$  puissent s'exprimer avec le plus de simplicité possible. Il est montré dans AAS et collab. [2009] que la définition (5.151) du vecteur  $\mathbf{V}$  se marie parfaitement

avec l'écriture factorisée d'une copule R-vigne. En effet,  $F_{k_2|k_1}$  et  $F_{k_3|k_1 k_2}$  peuvent s'exprimer uniquement à partir des h-fonctions des copules bidimensionnelles de la factorisation. Néanmoins, il faut que les conditionnements successifs soient effectués en accord avec la séquence d'arbres. Le travail de [TORRE et collab. \[2019\]](#) emprunte cette approche pour rendre indépendantes les observations d'un échantillon destiné à la construction d'un métamodèle par chaos polynomial. Prenons l'exemple de la spécification (5.11) qui est rappelée ci-dessous :

$$\mathbf{T}_{\text{vg}} := \{\mathbf{T}_1, \mathbf{T}_2\} \quad \text{avec} \quad \begin{cases} \mathbf{E}_1 = \{\{1\}, \{2\}\}; \{2\}, \{3\}\} \\ \mathbf{E}_2 = \{\{1, 2\}, \{2, 3\}\} \end{cases} \quad \text{et} \quad \mathbf{F}_{\text{vg}} := \{\mathcal{F}, \mathcal{F}', \mathcal{F}''\}. \quad (5.152)$$

En posant  $\{k_1, k_2, k_3\} = \{2, 1, 3\}$  et en appliquant (5.151),  $\mathbf{V}$  est donné par :

$$\begin{cases} V_2 = U_2 \\ V_1 = F_{1|2}(U_1 | U_2) \\ V_3 = F_{3|12}(U_3 | U_1, U_2) \end{cases}. \quad (5.153)$$

Avant d'aller plus loin, on souhaite introduire une notation alternative pour les h-fonctions :

$$\begin{cases} h_{ij}^{(G)}(u_i, u_j) = \frac{\partial C_{ij}}{\partial u_i}(u_i, u_j) = \mathbb{P}(U_j \leq u_j | U_i = u_i) = h_{j|i}(u_j | u_i) \\ h_{ij}^{(D)}(u_i, u_j) = \frac{\partial C_{ij}}{\partial u_j}(u_i, u_j) = \mathbb{P}(U_i \leq u_i | U_j = u_j) = h_{i|j}(u_i | u_j) \end{cases}. \quad (5.154)$$

Pour toute valeur  $u_0 \in [0, 1]$ , les fonctions  $h_{j|i}(\cdot | u_0)$  et  $h_{i|j}(\cdot | u_0)$  sont continues et strictement croissantes. Elles sont donc inversibles et on est en droit d'utiliser leurs réciproques. En utilisant les nouvelles notations,  $V_1 = \hat{h}_{1|2}(U_1 | U_2)$  et il vient facilement que :

$$\begin{aligned} V_3 &= \hat{F}_{3|12}(U_3 | U_1, U_2) = \frac{\partial \hat{C}_{13|2}}{\partial u_{1|2}}(\hat{F}_{1|2}(U_1 | U_2), \hat{F}_{3|2}(U_3 | U_2)) \\ &= \hat{h}_{3|12}(\hat{F}_{3|2}(U_3 | U_2) | \hat{F}_{1|2}(U_1 | U_2)) = \hat{h}_{3|12}(\hat{h}_{3|2}(U_3 | U_2) | \hat{h}_{1|2}(U_1 | U_2)). \end{aligned} \quad (5.155)$$

Pour la copule R-vigne étudiée, la transformation isoprobabiliste  $\hat{\mathcal{T}}_b$  se construit comme suit :

$$\mathbf{V} = \hat{\mathcal{T}}_b(\mathbf{U}) \quad \text{avec} \quad \begin{cases} V_2 = U_2 \\ V_1 = \hat{h}_{1|2}(U_1 | U_2) \\ V_3 = \hat{h}_{3|12}(\hat{h}_{3|2}(U_3 | U_2) | \hat{h}_{1|2}(U_1 | U_2)) \end{cases}. \quad (5.156)$$

De plus, le système d'égalités (5.156) s'inverse facilement et la transformation inverse  $\hat{\mathcal{T}}_b^{-1}$  s'écrit :

$$\mathbf{U} = \hat{\mathcal{T}}_b^{-1}(\mathbf{V}) \quad \text{avec} \quad \begin{cases} U_2 = V_2 \\ U_1 = \hat{h}_{1|2}^{-1}(V_1 | V_2) \\ U_3 = \hat{h}_{3|2}^{-1}(\hat{h}_{3|12}^{-1}(V_3 | V_1) | V_2) \end{cases}. \quad (5.157)$$

Les équations (5.156) et (5.157) finalisent l'implémentation de la transformation isoprobabiliste directe  $\hat{\mathcal{T}} = \mathcal{T}_c \circ \hat{\mathcal{T}}_b \circ \mathcal{T}_a$  et de la transformation isoprobabiliste inverse  $\hat{\mathcal{T}}^{-1} = \mathcal{T}_a^{-1} \circ \hat{\mathcal{T}}_b^{-1} \circ \mathcal{T}_c^{-1}$ . Une fois parvenu dans l'espace normal standard, les formules de la Section 4.5.2.2 restent valables :

$$\hat{\mathbf{D}}_f^{\mathcal{N}} = \{\mathbf{w} \in \mathbb{R}^d : \hat{\mathbf{H}}(\mathbf{w}) > \mathbf{T}\} \quad \text{avec} \quad \hat{\mathbf{H}} = \phi \circ \hat{\mathcal{T}}^{-1} \quad \text{et} \quad \hat{\mathbf{w}}^* = \underset{\hat{\mathbf{H}}(\mathbf{w})=\mathbf{T}}{\text{argmin}} \|\mathbf{w}\|_2. \quad (5.158)$$

De même, l'estimateur GIS de  $P_f(\mathbf{u}_{\text{obs}})$  est conçu exactement dans le même esprit :

$$\hat{P}_f(\mathbf{u}_{\text{obs}}) = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbb{1}_{D_f}(\mathbf{X}^{(i)}) \frac{\hat{f}_{\mathbf{X}}(\mathbf{X}^{(i)} | \mathcal{M}_{\text{cop}}, \mathbf{u}_{\text{obs}})}{\varphi_3(\mathbf{X}^{(i)} - \hat{\mathbf{x}}^*)} \quad \text{avec} \quad \begin{cases} \hat{\mathbf{x}}^* := \hat{\mathcal{T}}^{-1}(\hat{\mathbf{w}}^*) \\ \mathbf{X}^{(i)} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\hat{\mathbf{x}}^*, \mathbf{I}_3) \end{cases}. \quad (5.159)$$

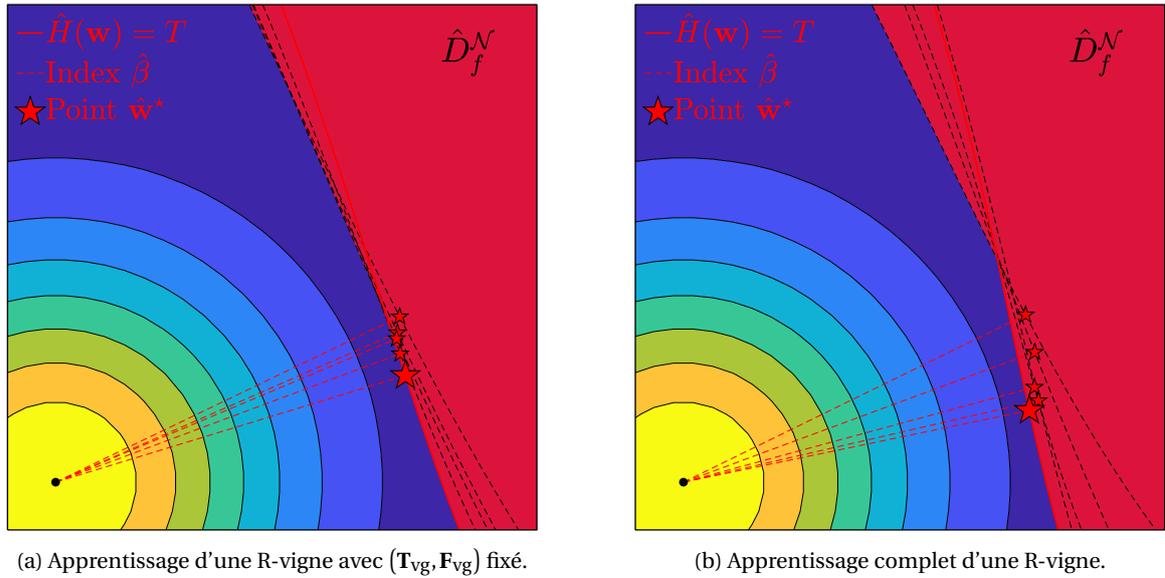


FIGURE 5.14 – Visualisation de l'influence du modèle  $\mathcal{M}_{\text{cop}}$  et de l'échantillon  $\tilde{\mathbf{U}}$  sur l'approximation  $\hat{D}_f^{\mathcal{N}}$ .

À nouveau, il est fait l'hypothèse que la matrice  $\mathbf{I}_3$  est une covariance tout à fait satisfaisante pour la densité auxiliaire. Dans la suite, pour tout échantillon  $\tilde{\mathbf{u}} \sim \hat{\mu}[\mathbf{u}_{\text{obs}}]$  survenant au cours d'une procédure d'ASE, l'estimation de  $P_f(\tilde{\mathbf{u}})$  passe par l'exécution de GIS.

Comme on connaît  $\hat{\mathcal{T}}^{-1}$ , on peut chercher à évaluer visuellement l'impact de la variabilité d'apprentissage sur l'estimation du risque de défaillance. On constitue une petite série de copies *bootstrap* de  $\mathbf{u}_{\text{obs}}$ . Pour chaque  $\tilde{\mathbf{u}}$  ainsi créé, la copule  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \tilde{\mathbf{u}})$  est apprise, le point de conception  $\tilde{\mathbf{w}}^*$  est calculé et la transformation  $\hat{\mathcal{T}}^{-1}$  est ré-assemblée. On peut alors tracer dans l'espace normal standard les courbes  $\hat{H}(\mathbf{w}) = T$  et les superposer. Elles constituent un faisceau d'approximations de l'état limite théorique  $H(\mathbf{w}) = T$ . On renvoie à la Figure 5.14 où ce travail est réalisé pour  $\mathcal{M}_{\text{cop}}^{(A)}$  puis  $\mathcal{M}_{\text{cop}}^{(B)}$ . Contrairement à ce qui peut être fait dans la Section 4.5.2.2, on ne peut pas visualiser la frontière théorique du domaine de défaillance  $D_f^{\mathcal{N}}$  car la copule sous-jacente  $c_{\mathbf{U}}$  est inconnue. Les courbes  $\hat{H}(\mathbf{w}) = T$  ne peuvent donc pas être confrontées à une courbe de référence mais on peut tout de même apprécier l'amplitude de leurs déviations. Il semble ainsi que l'apprentissage complet d'une R-vigne selon  $\mathcal{M}_{\text{cop}}^{(B)}$  soit plus enclin à déformer l'état limite et donc à perturber les résultats de l'analyse de fiabilité.

### 5.5.2.5 Construction d'un résumé matriciel

Comme on a pu le voir à la Section 5.6.1, le problème d'ASF initial décrit par  $\Psi$  est insoluble et il faut traiter l'un des deux sous-problèmes  $\Psi_{\boldsymbol{\tau}}$  ou  $\tilde{\Psi}_{\boldsymbol{\tau}}$  pour identifier la copule  $c_{ij}$  dont l'apprentissage impacte le plus la distribution de  $P_f(\mathbf{U}_{\text{obs}})$ . On privilégie  $\Psi_{\boldsymbol{\tau}}$  lorsqu'il existe un opérateur  $\mathcal{E}$  capable de calculer  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  à partir de  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ . Dans les autres situations, on est contraint de se limiter à l'étude de l'application  $\tilde{\Psi}_{\boldsymbol{\tau}}$  dont la stochasticité est plus prononcée. Pour ce cas test, on montre qu'un opérateur  $\mathcal{E}$  existe pour  $\mathcal{M}_{\text{cop}}^{(A)}$  mais pas pour  $\mathcal{M}_{\text{cop}}^{(B)}$ . Dans ce second cas, une solution alternative est néanmoins proposée pour construire un résumé de  $\hat{c}_{\text{vg}}(\cdot | \hat{\mathbf{T}}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}}, \hat{\boldsymbol{\theta}}_{\text{vg}})$  qui soit plus pertinent que celui apporté par la matrice empirique  $\hat{\boldsymbol{\tau}}(\mathbf{U}_{\text{obs}})$ . À l'issue de ce paragraphe, l'objectif est d'être capable de mettre en œuvre les deux procédures d'ASF pour l'un et l'autre des modèles d'apprentissage.

**Situation 1.** Cas du modèle  $\mathcal{M}_{\text{cop}}^{(A)}$

On renvoie à la Section 5.5.2.3 pour une description de la spécification  $(\mathbf{T}_{\text{vg}}, \mathbf{F}_{\text{vg}})$  correspondant

à  $\mathcal{M}_{\text{cop}}^{(A)}$ . Le détail de la factorisation associée à  $\hat{c}_t(\cdot | \hat{\boldsymbol{\rho}}, \hat{\mathbf{v}})$  est rappelé ci-dessous :

$$\begin{aligned} \hat{c}_{\text{vg}}(\mathbf{u} | \hat{\boldsymbol{\rho}}, \hat{\mathbf{v}}) &= \hat{c}_t(u_1, u_2 | \hat{\rho}_a, \hat{v}_a) \times \hat{c}_t(u_2, u_3 | \hat{\rho}_b, \hat{v}_b) \times \hat{c}_t(u_{1|2}, u_{3|2} | \hat{\rho}_c, \hat{v}_c) \\ \text{avec } u_{1|2} &= \frac{\partial \hat{C}_t}{\partial u_2}(u_1, u_2 | \hat{\rho}_a, \hat{v}_a) = \hat{h}_t^{(D)}(u_1, u_2 | \hat{\rho}_a, \hat{v}_a) \\ \text{et } u_{3|2} &= \frac{\partial \hat{C}_t}{\partial u_2}(u_2, u_3 | \hat{\rho}_b, \hat{v}_b) = \hat{h}_t^{(G)}(u_2, u_3 | \hat{\rho}_b, \hat{v}_b). \end{aligned} \quad (5.160)$$

Pour construire  $\mathcal{E}$ , on utilise un résultat d'abord énoncé dans **KUROWICKA et COOKE [2001]** puis complété dans **KUROWICKA et COOKE [2007]** et **MORALES et collab. [2008]**.

### Théorème 5.5.1 (Égalité des corrélations partielle et conditionnelle)

Soit  $\mathbf{U} = [U_j]_{1 \leq j \leq 3}$  un vecteur aléatoire à valeurs dans  $\mathcal{H}_3$  tel que :

- (i) La copule  $c_{12}$  de la paire  $(U_1, U_2)$  est elliptique.
- (ii) La copule  $c_{23}$  de la paire  $(U_2, U_3)$  est elliptique.
- (iii) La copule  $c_{13|2}$  vérifie l'hypothèse "simplificatrice" :

$$\forall \mathbf{u} \in \mathcal{H}_3, \frac{c_{\mathbf{U}}(u_1, u_2, u_3)}{c_{12}(u_1, u_2)c_{23}(u_2, u_3)} = c_{13|2}(u_{1|2}, u_{3|2} | u_2) = c_{13|2}(u_{1|2}, u_{3|2}). \quad (5.161)$$

On a alors  $\rho_{13;2} = \rho_{13|2}$  avec  $\begin{cases} \rho_{13;2} & \text{le coefficient de corrélation partielle,} \\ \rho_{13|2} & \text{le coefficient de corrélation conditionnelle.} \end{cases}$

Supposons qu'un vecteur  $\mathbf{V}$  soit distribué suivant la copule  $\hat{c}_{\text{vg}}(\cdot | \hat{\boldsymbol{\rho}}, \hat{\mathbf{v}})$  dont l'expression factorisée est présentée à l'équation (5.160). Les assertions (i) à (iii) sont alors toutes vérifiées :

- $c_{12} = \hat{c}_t(\cdot | \hat{\rho}_a, \hat{v}_a)$  est une copule elliptique.
- $c_{23} = \hat{c}_t(\cdot | \hat{\rho}_b, \hat{v}_b)$  est une copule elliptique.
- $c_{13|2} = \hat{c}_t(\cdot | \hat{\rho}_c, \hat{v}_c)$  vérifie l'hypothèse "simplificatrice" car  $\hat{c}_{\text{vg}}(\cdot | \hat{\boldsymbol{\rho}}, \hat{\mathbf{v}})$  est une R-vigne.

On en déduit que :

$$\hat{\rho}_a = \hat{\rho}_{12} \quad ; \quad \hat{\rho}_b = \hat{\rho}_{23} \quad ; \quad \hat{\rho}_c = \hat{\rho}_{13|2} = \hat{\rho}_{13;2}. \quad (5.162)$$

Le coefficient de corrélation  $\hat{\rho}_{13}$  n'est pas une information fournie par le paramétrage de la copule  $\hat{c}_{\text{vg}}(\cdot | \hat{\boldsymbol{\rho}}, \hat{\mathbf{v}})$  mais on peut y accéder grâce à la formule récursive de **YULE et KENDALL [1965]** qui est applicable quelle que soit la loi sous-jacente :

$$\hat{\rho}_{13} = \hat{\rho}_{12}\hat{\rho}_{23} + \hat{\rho}_{13;2}\sqrt{(1 - \hat{\rho}_{12}^2)(1 - \hat{\rho}_{23}^2)}. \quad (5.163)$$

On introduit la fonction  $\zeta$  pour établir le lien entre le paramétrage  $\hat{\boldsymbol{\theta}}_{\text{vg}}$  et la matrice  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}^{(A)}, \mathbf{U}_{\text{obs}})$  :

$$\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}^{(A)}, \mathbf{U}_{\text{obs}}) = \zeta(\hat{\boldsymbol{\theta}}_{\text{vg}}) = \zeta(\hat{\boldsymbol{\rho}}; \hat{\mathbf{v}}) = \zeta(\hat{\rho}_a, \hat{\rho}_b, \hat{\rho}_c; \hat{v}_a, \hat{v}_b, \hat{v}_c). \quad (5.164)$$

Numériquement, on conjecture que  $\mathbf{v} \mapsto \zeta(\boldsymbol{\rho}; \mathbf{v})$  est une application constante. Par conséquent, on est en droit de passer à la limite dans l'égalité précédente :

$$\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}^{(A)}, \mathbf{U}_{\text{obs}}) = \lim_{\substack{\hat{v}_a \rightarrow \infty \\ \hat{v}_b \rightarrow \infty \\ \hat{v}_c \rightarrow \infty}} \zeta(\hat{\boldsymbol{\rho}}; \hat{\mathbf{v}}). \quad (5.165)$$

La situation où  $\hat{v}_a = \hat{v}_b = \hat{v}_c = \infty$  est un cas limite de  $\mathcal{M}_{\text{cop}}^{(A)}$  pour lequel la copule estimée est une R-vigne dont toutes les familles sont gaussiennes. Par conséquent, sous réserve que  $\hat{\rho}_{12}$ ,  $\hat{\rho}_{13}$  et  $\hat{\rho}_{23}$  constituent une matrice de corrélation qui soit définie positive, la copule finale  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}^{(A)}, \mathbf{U}_{\text{obs}})$  est une copule gaussienne et on a donc :

$$\forall 1 \leq i < j \leq 3, \hat{\tau}_{ij} = \frac{2}{\pi} \arcsin(\hat{\rho}_{ij}) \quad \text{avec} \quad \begin{cases} \hat{\rho}_{12} = \hat{\rho}_a \\ \hat{\rho}_{23} = \hat{\rho}_b \\ \hat{\rho}_{13} = \hat{\rho}_a \hat{\rho}_b + \hat{\rho}_c \sqrt{(1 - \hat{\rho}_a^2)(1 - \hat{\rho}_b^2)} \end{cases}. \quad (5.166)$$

Cette formule justifie l'existence d'un opérateur  $\mathcal{E}$  pour  $\mathcal{M}_{\text{cop}}^{(A)}$ . On remarque que l'information apportée par les degrés de liberté  $\hat{\mathbf{v}}$  est perdue lors de l'application de  $\mathcal{E}$ . Le vecteur  $\hat{\boldsymbol{\theta}}_{\text{vg}}$  ne peut donc pas être reconstruit à partir de la seule matrice  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ . Contrairement au cas test de la Section 5.5.1,  $\mathcal{E}$  n'est pas inversible et  $\Psi_{\boldsymbol{\tau}}$  n'est pas déterministe. C'était assez prévisible étant donné que le nombre de paramètres constituant  $\hat{\boldsymbol{\theta}}_{\text{vg}} = (\hat{\boldsymbol{\rho}}, \hat{\mathbf{v}}) \in \mathbb{R}^6$  est plus grand que le nombre de coefficients utilisés dans  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}}) \in \mathbb{R}^3$  pour résumer la  $d$ -copule estimée.

**Situation 2.** Cas du modèle  $\mathcal{M}_{\text{cop}}^{(B)}$

Étant donné que  $\hat{\mathbf{F}}_{\text{vg}}$  peut inclure des copules non-elliptiques, les points (i) et (ii) ne sont pas systématiquement vérifiés. De plus, le passage à la limite (5.165) n'est pas vrai dans le cas général. Pour  $\mathcal{M}_{\text{cop}}^{(B)}$ , il n'est pas possible d'aboutir à une formule analytique fournissant  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}^{(B)}, \mathbf{U}_{\text{obs}})$  à partir du triplet  $(\hat{\mathbf{T}}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}}, \hat{\boldsymbol{\theta}}_{\text{vg}})$ . Toutefois, il est très facile de simuler un grand nombre de nouvelles observations du vecteur  $\mathbf{V} \sim \hat{c}_{\text{vg}}(\cdot | \hat{\mathbf{T}}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}}, \hat{\boldsymbol{\theta}}_{\text{vg}})$  surtout pour  $d = 3$ . Pour accomplir une tâche de ré-échantillonnage à partir d'une R-vigne préalablement déterminée, il faut utiliser l'algorithme de AAS et collab. [2009] ou celui de DISSMANN et collab. [2013]. En revenant à la définition générale donnée à l'équation (5.26), un estimateur  $\hat{\boldsymbol{\kappa}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  de  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  vient naturellement :

$$\hat{\boldsymbol{\kappa}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}}) = [\hat{\kappa}_{ij}]_{1 \leq i < j \leq d} \quad \text{avec} \quad \hat{\kappa}_{ij} = \frac{2}{N_s(N_s - 1)} \sum_{k < l} T_{ij}^{(kl)}$$

$$\text{où} \quad \begin{cases} T_{ij}^{(kl)} = \text{sgn}(V_i^{(k)} - V_i^{(l)}) \times \text{sgn}(V_j^{(k)} - V_j^{(l)}) \\ \mathbf{V}^{(n)} \stackrel{\text{i.i.d.}}{\sim} \hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}}) \quad \forall 1 \leq n \leq N_s \end{cases} \quad (5.167)$$

Il faut bien comprendre que les matrices  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  et  $\hat{\boldsymbol{\kappa}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  sont très différentes bien qu'elles utilisent le même mécanisme empirique de calcul des corrélations de Kendall. D'un côté,  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  est basée sur  $\mathbf{U}_{\text{obs}}$ , c'est-à-dire sur un petit nombre  $N$  d'observations réelles de la copule théorique  $c_{\mathbf{U}}$ . De l'autre, on trouve  $\hat{\boldsymbol{\kappa}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  qui s'appuie sur un grand nombre  $N_s$  d'observations simulées de la  $d$ -copule estimée  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  et qui incorpore ainsi un supplément d'information propre à l'apprentissage statistique  $\mathcal{L}$  via  $\mathcal{M}_{\text{cop}}$ . En utilisant  $\hat{\boldsymbol{\kappa}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ , on ne cherche pas à imiter l'opérateur empirique  $\hat{\mathcal{E}}$  mais plutôt à construire une version légèrement dégradée de l'opérateur analytique  $\mathcal{E}$ . Par ailleurs, on veut attirer l'attention sur le fait qu'un estimateur  $\hat{\boldsymbol{\kappa}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  de la matrice  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  peut être fabriqué quel que soit  $\mathcal{M}_{\text{cop}}$  sous réserve qu'il existe une routine permettant de simuler un très grand nombre d'observations de  $\mathbf{V} \sim \hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ . Dans la suite, la production d'une instance de  $\hat{\boldsymbol{\kappa}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  à partir de  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  est notée  $\hat{\mathcal{E}}$  pour insister sur le fait que l'enjeu est d'imiter au mieux le fonctionnement de l'opérateur  $\mathcal{E}$ .

### 5.5.2.6 Comparaison des résultats obtenus avec $\Psi_{\boldsymbol{\tau}}$ et $\tilde{\Psi}_{\boldsymbol{\tau}}$

On étudie les résultats des problèmes d'ASF décrits  $\Psi_{\boldsymbol{\tau}}$  et  $\tilde{\Psi}_{\boldsymbol{\tau}}$  lorsque la phase d'apprentissage statistique  $\mathcal{L}$  s'appuie sur les modèles  $\mathcal{M}_{\text{cop}}^{(A)}$  et  $\mathcal{M}_{\text{cop}}^{(B)}$ . L'échantillon  $\mathbf{u}_{\text{obs}}$  qui est constitué à l'équation (5.145) sert de matière première à toutes les expérimentations numériques. En ce qui concerne le schéma d'estimation  $\mathcal{A}$  des indices  $\delta_{ij}^F$  et de l'indicateur  $\xi^F$ , on impose le recours à  $M = 10^3$  itérations. Pour chaque densité  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \tilde{\boldsymbol{\mu}})$  émanant de l'exécution de la chaîne  $\mathcal{L} \circ \mathcal{B}$ , la probabilité de défaillance  $P_f(\tilde{\boldsymbol{\mu}})$  est estimée avec l'algorithme GIS. Bien que ce ne soit pas indispensable en présence d'une fonction de performance  $\phi$  analytique, l'astuce "inverse-reverse" est implémentée pour limiter la charge de calcul. Les résultats sont consignés dans le Tableau 5.5 où chaque étage correspond au couplage entre un modèle d'apprentissage et un choix de résumé matriciel de la  $d$ -copule estimée  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ . Les propriétés des estimateurs sont évaluées après  $N_{\text{rep}} = 10$  répétitions de chaque *workflow* d'ASF.

Une première information réside dans le fait que  $\hat{\delta}_{12}^F$  est l'indice dominant dans chacune des quatre configurations étudiées. En s'inspirant de l'équation (5.129), la sortie  $Y$  peut être réécrite

				ASF exécutée avec $\mathcal{B}$ :						
				$\hat{\delta}_{12}^F$	$\hat{\delta}_{13}^F$	$\hat{\delta}_{23}^F$	$\hat{\xi}^F$			
Échantillon	$\mathbf{u}_{\text{obs}}$	Modèle de R-vigne	$\mathcal{M}_{\text{cop}}^{(A)}$	Résumé de $\mathcal{U}$	$\mathcal{E}$	moyenne :	0.53	0.29	0.08	0.66
						écart-type :	0.01	0.01	0.01	0.01
						rang :	(1)	(2)	(3)	×
					$\tilde{\mathcal{E}}$	moyenne :	0.23	0.13	0.04	0.25
				écart-type :	0.01	0.01	0.01	0.01		
				rang :	(1)	(2)	(3)	×		
			$\hat{\mathcal{E}}$	moyenne :	0.39	0.20	0.03	0.44		
				écart-type :	0.08	0.02	0.01	0.09		
	rang :	(1)	(2)	(3)	×					
	$\tilde{\mathcal{E}}$	moyenne :	0.29	0.12	0.03	0.30				
	écart-type :	0.04	0.01	0.01	0.04					
	rang :	(1)	(2)	(3)	×					

TABLEAU 5.5 – Comparaison des résultats obtenus pour les deux sous-problèmes d'ASF

en fonction du  $\mathbf{U}$  :

$$Y = \sum_{j=1}^3 a_j \left( \sigma_j \Phi^{-1}(U_j) + \mu_j \right) + a_4 \prod_{j=1}^2 \left( \sigma_j \Phi^{-1}(U_j) + \mu_j \right). \quad (5.168)$$

Comme les lois marginales gaussiennes ont été paramétrées identiquement à l'équation (5.142), on peut anticiper plus facilement l'impact des copules bidimensionnelles  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$  sur  $\hat{P}_f(\mathbf{U}_{\text{obs}})$ . Si la copule  $\hat{c}_{12}(\cdot | \mathbf{U}_{\text{obs}})$  pilote la dépendance entre les deux termes de la somme pour lesquels les coefficients  $a_j$  sont les plus grands, elle exerce aussi un rôle de stimulation du terme non-linéaire. Celui-ci a davantage de chances de prendre de grandes valeurs lorsque la paire  $(U_1, U_2)$  est fortement corrélée. Au regard de ces deux arguments, il n'est pas étonnant que  $\hat{\delta}_{12}^F$  soit un bien plus large vainqueur de l'ASF que dans le Tableau 5.4 où  $\phi$  est seulement linéaire.

Les classements sont identiques dans les quatre situations étudiées mais les valeurs moyennes des indices  $\hat{\delta}_{ij}^F$  évoluent considérablement au gré de la stochasticité reliant le résumé matriciel choisi et la probabilité de défaillance. L'indicateur  $\hat{\xi}^F$  fait office de borne supérieure pour la collection d'indices  $\hat{\delta}_{ij}^F$ , ce qui est normal étant donné qu'il s'agit du  $\delta$ -indice d'ordre total. La proximité entre  $\hat{\delta}_{12}^F$  et  $\hat{\xi}^F$  est une preuve supplémentaire du rôle prédominant de la copule  $\hat{c}_{12}(\cdot | \mathbf{U}_{\text{obs}})$ . Toutefois, comme les incertitudes épistémiques sont quantifiées avec la mesure d'importance de Borgonovo, cette proximité ne signifie pas que la copule  $\hat{c}_{12}(\cdot | \mathbf{U}_{\text{obs}})$  véhicule à elle seule la quasi-totalité de la variabilité d'apprentissage. Pour tirer une telle conclusion, il faudrait que les indices  $\hat{\delta}_{13}^F$  et  $\hat{\delta}_{23}^F$  soient pratiquement nuls et ce n'est pas ce qui est observé.

On se propose désormais de se pencher sur la quantification du niveau de stochasticité intervenant dans les différentes procédures d'ASF. En effet,  $\hat{\xi}^F$  semble pousser à la hausse ou à la baisse la collection des indices  $\hat{\delta}_{ij}^F$ . En particulier, le Tableau 5.5 délivre les trois inégalités suivantes :

$$\hat{\xi}^F(\mathcal{M}_{\text{cop}}^{(A)}; \mathcal{E}) > \hat{\xi}^F(\mathcal{M}_{\text{cop}}^{(B)}; \hat{\mathcal{E}}), \quad (A)$$

$$\hat{\xi}^F(\mathcal{M}_{\text{cop}}^{(B)}; \tilde{\mathcal{E}}) > \hat{\xi}^F(\mathcal{M}_{\text{cop}}^{(A)}; \tilde{\mathcal{E}}), \quad (B)$$

$$\hat{\xi}^F(\mathcal{M}_{\text{cop}}^{(A)}; \mathcal{E}) > \hat{\xi}^F(\mathcal{M}_{\text{cop}}^{(A)}; \tilde{\mathcal{E}}). \quad (C)$$

L'inégalité (A) indique que l'application  $\Psi_{\tau}$  comporte une plus grande part de stochasticité lorsque l'apprentissage statistique  $\mathcal{L}$  obéit au modèle  $\mathcal{M}_{\text{cop}}^{(B)}$ . À la Section 5.5.2.5, on a vu que résumer la densité  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}^{(A)}, \mathbf{U}_{\text{obs}}) = \hat{c}_{\text{vg}}(\cdot | \hat{\boldsymbol{\rho}}, \hat{\mathbf{v}})$  par la matrice  $\hat{\boldsymbol{\tau}}(\mathcal{M}_{\text{cop}}^{(A)}, \mathbf{U}_{\text{obs}}) = \hat{\boldsymbol{\tau}}(\hat{\boldsymbol{\rho}})$  entraîne la perte de l'information liée à  $\hat{\mathbf{v}}$ . Pour rendre  $\Psi_{\tau}$  déterministe, il suffirait d'ajouter le vecteur  $\hat{\mathbf{v}}$  à ses entrées. La perte d'information est beaucoup plus conséquente lorsque l'on décide de résumer la copule

Contexte de l'ASF			Incertitudes			
Apprentissage	Résumé	Application	Type I	Type II	Type III	Indicateur
$\mathcal{M}_{\text{cop}}^{(A)}$	$\mathcal{E}$	$\Psi_{\tau}$	+	-	+	0.66
$\mathcal{M}_{\text{cop}}^{(A)}$	$\tilde{\mathcal{E}}$	$\tilde{\Psi}_{\tau}$	+	++	+	0.25
$\mathcal{M}_{\text{cop}}^{(B)}$	$\hat{\mathcal{E}}$	$\Psi_{\tau}$	+	+	+	0.44
$\mathcal{M}_{\text{cop}}^{(B)}$	$\tilde{\mathcal{E}}$	$\tilde{\Psi}_{\tau}$	+	++	+	0.30

TABLEAU 5.6 – Bilan des incertitudes indésirables introduites au cours de l'ASF.

$\hat{\mathbf{c}}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}^{(B)}, \mathbf{U}_{\text{obs}}) = \hat{\mathbf{c}}_{\text{vg}}(\cdot | \hat{\mathbf{T}}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}}, \hat{\boldsymbol{\theta}}_{\text{vg}})$  par la matrice  $\hat{\mathbf{k}}(\mathcal{M}_{\text{cop}}^{(B)}, \mathbf{U}_{\text{obs}})$ . En effet, il y a trois sources de stochasticité qui s'accumulent :

- Pour un échantillon  $\mathbf{u}_{\text{obs}}$  fixé,  $\hat{\mathbf{k}}(\mathcal{M}_{\text{cop}}^{(B)}, \mathbf{U}_{\text{obs}})$  est seulement un estimateur de  $\hat{\mathbf{t}}(\mathcal{M}_{\text{cop}}^{(B)}, \mathbf{u}_{\text{obs}})$ .
- À supposer qu'on puisse accéder à  $\hat{\mathbf{t}}(\mathcal{M}_{\text{cop}}^{(B)}, \mathbf{u}_{\text{obs}})$ , on ne pourrait pas reconstruire  $(\hat{\mathbf{T}}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}})$ .
- Si  $\hat{\boldsymbol{\theta}}_{\text{vg}}$  implique plus que  $p = 3$  paramètres, les coefficients  $\hat{t}_{ij}$  ne sont pas assez nombreux pour retranscrire une information équivalente.

Le point (c) explique totalement la stochasticité de  $\Psi_{\tau}$  pour  $\mathcal{M}_{\text{cop}}^{(A)}$  mais pas pour  $\mathcal{M}_{\text{cop}}^{(B)}$ . Au vu de ses éléments, l'inégalité (A) apparaît donc comme tout à fait naturelle. L'inégalité (B) montre étonnamment que la tendance s'inverse pour  $\tilde{\Psi}_{\tau}$ . La part de stochasticité semble en effet être plus grande lorsqu'on utilise  $\mathcal{M}_{\text{cop}}^{(A)}$  pour l'apprentissage. Une explication peut résider dans le fait que la matrice  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$  n'apporte aucune information utile à la reconstruction de  $\hat{\mathbf{c}}_{\text{vg}}(\cdot | \hat{\boldsymbol{\rho}}, \hat{\mathbf{v}})$  alors qu'elle permet au moins de retrouver  $\hat{\mathbf{T}}_{\text{vg}}$  dans le cadre d'une reconstruction de  $\hat{\mathbf{c}}_{\text{vg}}(\cdot | \hat{\mathbf{T}}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}}, \hat{\boldsymbol{\theta}}_{\text{vg}})$ . Quant à l'inégalité (C), elle illustre à nouveau le fait que le degré de stochasticité affectant  $\tilde{\Psi}_{\tau}$  est supérieur à celui  $\Psi_{\tau}$ . Cette inégalité est écrite pour  $\mathcal{M}_{\text{cop}}^{(A)}$  mais elle est aussi vérifiée par  $\mathcal{M}_{\text{cop}}^{(B)}$ .

Comme les probabilités de défaillance sont estimées avec GIS, les applications  $\Psi_{\tau}$  et  $\tilde{\Psi}_{\tau}$  sont aussi rendues stochastiques par le bruit intrinsèque de l'algorithme  $\mathcal{R}_w$ . Même si tout est fait pour que cette stochasticité supplémentaire soit réduite à son niveau minimal, son impact sur  $\hat{P}_f(\mathbf{U}_{\text{obs}})$  contribue à abaisser la valeur de l'indicateur  $\hat{\xi}^F$ . Dans le Tableau 5.6, on propose une synthèse des incertitudes indésirables créant de la stochasticité au sein des applications analysées. Pour cela, on utilise la terminologie suivante :

- Type I :** incertitude introduite en résumant  $\hat{\mathbf{c}}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  par  $\hat{\mathbf{t}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ .
- Type II :** incertitude introduite en remplaçant  $\hat{\mathbf{t}}(\mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$  par  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$ .
- Type III :** incertitude introduite en utilisant un algorithme de simulation  $\mathcal{R}_w$ .

### 5.5.2.7 Assimilation de données supplémentaires

À partir de  $\mathbf{u}_{\text{obs}}$ , on exécute deux scénarios d'assimilation d'observations supplémentaires à partir du stock  $\mathbf{u}_{\text{extra}}$  constitué à l'équation (5.145). Celui-ci ne peut comporter qu'un maximum de  $N' = 455$  nouvelles observations de  $\mathbf{U}$ . Contrairement aux situations étudiées jusqu' alors où les données de  $\mathbf{u}_{\text{extra}}$  ont été simulées en nombre suffisant pour que l'évolution des classements soit clairement visible, on est ici dans une situation assez proche de la pratique avec  $N'$  qui est fixé à l'avance et relativement petit. On choisit d'étudier l'impact du matériel apporté par  $\mathbf{u}_{\text{extra}}$  sur l'application  $\tilde{\Psi}_{\tau}$  car il s'agit en quelque sorte de la procédure d'ASF par défaut. Pour ce qui est de l'apprentissage  $\mathcal{L}$ , on utilise  $\mathcal{M}_{\text{cop}}^{(B)}$  qui a vocation à être appliqué à d'autres problèmes contrairement à  $\mathcal{M}_{\text{cop}}^{(A)}$  qui est spécifiquement conçu pour le jeu de données "uranium". Au regard des différents classements exposés dans le Tableau 5.5, on décide de cibler les paires  $(U_1, U_2)$  et  $(U_1, U_3)$ . On ne peut pas procéder comme à la Section 5.5.1.4 où les données supplémentaires  $\mathbf{u}_{\text{extra}}^{[i,j]}$  d'une paire

$(U_i, U_j)$  permettent d'affiner la connaissance du paramètre  $\hat{\rho}_{ij}$  avant que celui-ci ne soit réinjecté dans la formule analytique établissant la valeur prise par  $P_f(\hat{\rho})$ . Pour incorporer un matériel de type  $\mathbf{u}_{\text{extra}}^{[\bullet(ij)]}$ , il faut transformer  $\mathcal{M}_{\text{cop}}^{(B)}$  selon ce qui est décrit à la Section 5.4.2.2 pour les R-vignes. On renvoie au Tableau 5.7 qui présente les résultats obtenus après avoir mené l'assimilation de  $N' = 200$  puis  $N' = 455$  nouvelles observations pour chacune des deux paires sélectionnées. Le premier étage provient du Tableau 5.5 et représente l'étape d'ASF initiale. Le second et le troisième étages montrent l'impact sur les résultats d'ASF de l'incorporation progressive de données supplémentaires pour les paires  $(U_1, U_2)$  et  $(U_1, U_3)$ .

ASF exécutée avec $\mathcal{B}$ :				$\hat{\delta}_{12}^F$	$\hat{\delta}_{13}^F$	$\hat{\delta}_{23}^F$	$\hat{\xi}^F$
	$\mathbf{u}_{\text{obs}}$		moyenne :	0.29	0.12	0.03	0.30
			écart-type :	0.04	0.01	0.01	0.04
			rang :	(1)	(2)	(3)	×
ASF exécutée avec $\mathcal{B}_{12}$ :				$\hat{\delta}_{12}^F$	$\hat{\delta}_{13}^F$	$\hat{\delta}_{23}^F$	$\hat{\xi}^F$
Paire $\mathbf{U}_{12}$	$\mathbf{u}_{\text{obs}} \cup \mathbf{u}_{\text{extra}}^{[\bullet(12)]}$	$N' = 200$	moyenne :	0.12	0.09	0.04	0.15
			écart-type :	0.01	0.01	0.02	0.01
			rang :	(1)	(2)	(3)	×
		$N' = 455$	moyenne :	0.12	0.09	0.06	0.16
			écart-type :	0.02	0.01	0.01	0.01
			rang :	(1)	(2)	(3)	×
ASF exécutée avec $\mathcal{B}_{13}$ :				$\hat{\delta}_{12}^F$	$\hat{\delta}_{13}^F$	$\hat{\delta}_{23}^F$	$\hat{\xi}^F$
Paire $\mathbf{U}_{13}$	$\mathbf{u}_{\text{obs}} \cup \mathbf{u}_{\text{extra}}^{[\bullet(13)]}$	$N' = 200$	moyenne :	0.24	0.06	0.04	0.26
			écart-type :	0.01	0.01	0.01	0.01
			rang :	(1)	(2)	(3)	×
		$N' = 455$	moyenne :	0.26	0.05	0.03	0.28
			écart-type :	0.01	0.01	0.01	0.01
			rang :	(1)	(2)	(3)	×

TABEAU 5.7 – Impact de l'assimilation de données supplémentaires sur le classement des  $\delta$ -indices.

On constate que la part de stochasticité de l'application  $\tilde{\Psi}_{\tau}$  est susceptible d'évoluer entre la première et la seconde ASF. Il s'agit d'une augmentation de stochasticité et cette tendance est marquée par une diminution de l'indicateur  $\hat{\xi}^F$ . L'affaiblissement de  $\hat{\xi}^F$  peut s'expliquer en retraçant l'évolution des différentes incertitudes qui se combinent pour engendrer la distribution de  $\hat{P}_f(\mathbf{U}_{\text{obs}})$ . Si l'assimilation de données a eu l'effet escompté, les incertitudes épistémiques portées par la copule raffinée doivent être plus faibles que celles portées par la copule initiale. En revanche, le bruit introduit par la mise en service de l'algorithme GIS garde la même amplitude au cours des deux procédures d'ASF. Par conséquent, la part de la distribution de  $\hat{P}_f(\mathbf{U}_{\text{obs}})$  qui peut être expliquée par la matrice  $\hat{\mathbf{i}}(\mathbf{U}_{\text{obs}})$  s'atrophie lorsqu'on élimine une partie de la variabilité d'apprentissage. Il faut remarquer que la sévérité de la diminution de  $\hat{\xi}^F$  dépend de la paire ciblée par l'action d'enrichissement. Cette diminution est presque imperceptible pour la paire  $(U_1, U_3)$  alors qu'elle conduit à diviser  $\hat{\xi}^F$  de moitié pour la paire  $(U_1, U_2)$ . Une forte diminution de  $\hat{\xi}^F$  indique une forte réduction des incertitudes créées par l'estimation de la copule et se veut synonyme d'une action réussie sur la paire concernée. À ce titre, les résultats du Tableau 5.7 montrent qu'il vaut mieux intervenir sur  $(U_1, U_2)$  et cela vient corroborer les classements déjà établis lors de la première ASF.

De par la propriété de croissance des  $\delta$ -indices, la diminution de  $\hat{\xi}^F$  entraîne celle de toute la collection des indices  $\hat{\delta}_{ij}$ . De ce fait, il est alors plus compliqué de comprendre l'évolution des parts d'influence entre l'ASF initiale et une seconde ASF exécutée après un apport de données supplémentaires. Par exemple, lorsque la paire  $(U_1, U_2)$  est ciblée, les indices  $\hat{\delta}_{12}^F$  et  $\hat{\delta}_{13}^F$  diminuent.

L'incorporation des données  $\mathbf{u}_{\text{extra}}^{(12)}$  justifie la première diminution mais pas la seconde. Toutefois, il faut remarquer que  $\hat{\delta}_{13}^F$  chute beaucoup moins vite que  $\hat{\delta}_{12}^F$  et  $\hat{\xi}^F$ . À l'issue de la seconde ASF, l'indice  $\hat{\delta}_{13}^F$  conserve un classement intermédiaire mais l'écart avec l'indice dominant  $\hat{\delta}_{12}^F$  ainsi que l'écart avec l'indice d'ordre total  $\hat{\xi}^F$  se sont amoindris. C'est un signe du renforcement de l'influence de la copule  $\hat{c}_{13}(\cdot | \mathbf{U}_{\text{obs}})$  au sein du lot des copules bidimensionnelles  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$ . Par ailleurs, l'augmentation de l'indice  $\hat{\delta}_{23}^F$  indique une progression du rôle joué par  $\hat{c}_{23}(\cdot | \mathbf{U}_{\text{obs}})$  bien que ce rôle reste mineur. Une évolution analogue peut-être est observée lorsque l'enrichissement est mené sur la paire  $(U_1, U_3)$ .

Il est intéressant de noter que le Tableau 5.7 met en avant les bienfaits et les limites de la mise en œuvre des stratégies d'acquisition ciblée de nouvelles données. Assimiler  $N' = 200$  observations supplémentaires d'une paire influente permet une réduction de la variabilité d'apprentissage qui se traduit par une diminution de l'indicateur de stochasticité  $\hat{\xi}^F$  ainsi que par certains changements au sein de la collection des indices  $\hat{\delta}_{ij}^F$ . Néanmoins, si on augmente le nombre d'observations supplémentaires de cette paire pour atteindre  $N' = 455$ , les résultats d'ASF restent globalement identiques. Pour continuer à réduire l'influence d'un coefficient  $\hat{\tau}_{ij}$  sur  $P_f(\mathbf{U}_{\text{obs}})$ , il faudrait disposer d'un très grand volume d'observations supplémentaires pour la paire  $(U_i, U_j)$  et le jeu de données "uranium" ne le permet pas.

### 5.5.2.8 Bilan

Dans cette section, un certain nombre d'écueils rappellent les difficultés rencontrées en pratique. De par la non-linéarité de la fonction  $\phi$ , l'estimation des probabilités de défaillance requiert l'utilisation d'un algorithme de simulation  $\mathcal{R}_w$ . De plus, la loi de la copule sous-jacente  $c_{\mathbf{U}}$  est parfaitement inconnue et on se propose de l'estimer à l'aide de deux modèles d'apprentissage reposant sur les R-vignes. Le premier implique une spécification préalable de la séquence d'arbres  $\mathbf{T}_{\text{vg}}$  et de l'ensemble  $\mathbf{F}_{\text{vg}}$  des familles paramétriques tandis que le second s'appuie uniquement sur l'échantillon  $\mathbf{U}_{\text{obs}}$ . On a tiré différents enseignements des expérimentations numériques réalisées :

- Pour les copules R-vignes, il est rare que l'on puisse accéder à l'expression analytique de  $\mathcal{E}$ .
- Si on dispose d'une routine pour simuler  $\mathbf{V} \sim \hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}})$ ,  $\hat{\mathcal{E}}$  peut se substituer à  $\mathcal{E}$ .
- Les indices  $\hat{\delta}_{ij}^F$  obtenus pour  $\tilde{\Psi}_{\tau}$  sont plus petits que ceux obtenus pour  $\Psi_{\tau}$  mais cela n'a pas d'impact sur le classement final.
- L'utilisation d'un algorithme de simulation  $\mathcal{R}_w$  introduit une stochasticité supplémentaire qui doit être maîtrisée au risque d'écraser les indices  $\hat{\delta}_{ij}$ .
- Des procédures d'assimilation de données supplémentaires peuvent être mises en place mais la validation numérique de leur efficacité est assez contre-intuitive.

Dans la prochaine section, on confronte la méthodologie développée à un problème d'ingénierie. La copule  $c_{\mathbf{U}}$  et la fonction de performance  $\phi$  découlent alors de la modélisation d'un phénomène physique.

## 5.6 Application au flambage d'une plaque composite

### 5.6.1 Transformation du problème initial

On reprend la plaque composite orthotrope dont une illustration est fournie à la Figure 4.23 et on renvoie à la Section 4.6 pour plus de détails. On rappelle que la chaîne de propagation des incertitudes inclut deux étapes :

$$\begin{array}{ccccccc}
 \mathbb{R}^4 \times \mathbb{R}^8 \times \mathbb{R} & \longrightarrow & \mathbb{R}^4 & \longrightarrow & \mathbb{R} & & \\
 (\mathbf{M}, \boldsymbol{\Theta}, e) & \xrightarrow{\phi_0} & \mathbf{D} & \xrightarrow{\phi_1} & \Lambda & \cdot & \\
 & & \mathbf{X} & \xrightarrow{\phi} & \mathbf{Y} & & 
 \end{array} \quad (5.169)$$

Les incertitudes aléatoires sont propagées par les quatre modules de l'ingénieur, l'orientation des fibres dans les huit couches du stratifié et l'épaisseur commune des plis. La fonction  $\phi_0$  permet le calcul de la matrice de rigidité  $\mathbf{D}$  du stratifié tandis que la fonction  $\phi_1$  donne ensuite accès au facteur de flambage  $\Lambda$  qui est la variable d'intérêt du problème. Dans le Chapitre 4, on a posé  $\phi = \phi_0$  et on a cherché à quantifier le risque de flambage  $\mathbb{P}(\Lambda \leq 1)$  à partir de la connaissance d'un échantillon  $\mathbf{x}_{\text{obs}}$  des coefficients de  $\mathbf{D}$ . Lorsque l'apprentissage  $\mathcal{L}$  de la densité conjointe  $f_{\mathbf{D}}$  consiste à estimer les marginales avec la méthode KDE puis à estimer la copule avec une R-vigne, il a été montré que  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{V}_{\text{obs}})$  est la composante fonctionnelle qui véhicule le plus de variabilité d'apprentissage sur  $\hat{P}_f(\mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}})$ . Face à ce constat, on souhaite désormais identifier la copule bidimensionnelle  $c_{ij}$  dont l'apprentissage a le plus d'impact sur la distribution de l'estimateur du risque de défaillance. Toutefois, la configuration (5.169) doit être légèrement transformée pour qu'on puisse tester la méthodologie du présent chapitre. En effet, les lois marginales de  $\mathbf{D}$  sont inconnues alors qu'il est nécessaire qu'on en ait une connaissance exacte afin de circonscrire l'incertitude d'apprentissage à la seule copule. Sachant que la loi des entrées  $(\mathbf{M}, \Theta, e)$  est spécifiée au Tableau 4.8, la matrice de rigidité  $\mathbf{D}$  peut être simulée. On effectue alors  $N'' = 10^5$  tirages et on note  $\mathbf{x}_{\text{base}}$  une réalisation de la matrice aléatoire ci-dessous :

$$\mathbf{X}_{\text{base}} := [X_j^{(i)}]_{\substack{1 \leq i \leq N'' \\ 1 \leq j \leq 4}} \quad \text{avec} \quad \begin{cases} (\mathbf{M}^{(i)}, \Theta^{(i)}, e^{(i)}) \stackrel{\text{i.i.d.}}{\sim} \mu_{\mathbf{M}} \otimes \mu_{\Theta} \otimes \mu_e \\ \mathbf{X}^{(i)} = \mathbf{D}^{(i)} = \phi_0(\mathbf{M}^{(i)}, \Theta^{(i)}, e^{(i)}) \end{cases}, \quad (5.170)$$

et on renvoie à l'équation (4.143) ainsi qu'au Tableau 4.8 pour plus de détails au sujet du modèle probabiliste correspondant à la loi du vecteur  $(\mathbf{M}, \Theta, e)$ . La matrice  $\mathbf{x}_{\text{base}}$  stocke ainsi un nombre  $N''$  exorbitant d'observations conjointes de  $\mathbf{D}$ . Celles-ci sont utilisées pour construire par KDE des estimations très précises des densités marginales  $f_{X_j}$  :

$$\forall 1 \leq j \leq 4, \quad \hat{k}_j(x_j) := \hat{f}_{\text{KDE}}(x_j | \mathbf{x}_{\text{base}}^{[\cdot j]}, \hat{h}_j). \quad (5.171)$$

Les fonctions  $\hat{k}_j$  sont ensuite figées et considérées comme les densités marginales théoriques. Par conséquent, la copule sous-jacente est désormais la distribution du vecteur :

$$\mathbf{U} = [U_j]_{1 \leq j \leq 4} \quad \text{avec} \quad U_j = \hat{K}_j(X_j) = \hat{F}_{\text{KDE}}(X_j | \mathbf{x}_{\text{base}}^{[\cdot j]}, \hat{h}_j). \quad (5.172)$$

Partant de ce principe, l'échantillon  $\mathbf{u}_{\text{obs}}$  s'obtient directement à partir de l'échantillon  $\mathbf{x}_{\text{obs}}$  en appliquant les fonctions de répartition  $\hat{K}_j$ . Ces manipulations permettent d'aboutir à une chaîne de propagation des incertitudes de type  $\mathbf{U}_{\text{obs}} \mapsto P_f(\mathbf{U}_{\text{obs}})$ . Cette nouvelle configuration est représentée à la Figure 5.15 et s'avère propice à la mise en œuvre de la seconde étape de l'ASF.

## 5.6.2 Implémentation de la méthodologie développée

Pour rester dans la continuité de ce qui est fait à la Section 4.6.2.5,  $\mathcal{M}_{\text{cop}}$  désigne l'apprentissage complet d'une copule R-vigne. La  $d$ -copule estimée s'écrit alors  $\hat{c}_{\text{vg}}(\cdot | \hat{\mathbf{T}}_{\text{vg}}, \hat{\mathbf{F}}_{\text{vg}}, \hat{\Theta}_{\text{vg}})$  et on utilise l'opérateur empirique  $\hat{\mathcal{E}}$  pour en construire un résumé matriciel  $\hat{\mathbf{t}}(\mathbf{U}_{\text{obs}})$ . Pour estimer les différentes probabilités de défaillance, on conserve la technique de fiabilité implémentée pour la première procédure d'ASF. Elle consiste à exécuter  $N_e = 5$  explorations préliminaires du domaine de défaillance  $D_f$  à l'aide de l'algorithme NAIS qui est initialisé avec  $N_s = 3 \times 10^3$  particules. L'échantillonnage préférentiel multiple permet ensuite de mettre en commun le matériel produit par les différentes explorations et de fabriquer une densité auxiliaire  $\hat{h}_{\text{MIS}}$  pour laquelle on dispose de  $N_e \times N_s$  échantillons  $\mathbf{X}^{(k)}$  ainsi que des sorties  $Y^{(k)}$  associées. Après cette étape, tout calcul de fiabilité peut être effectué sans avoir à solliciter  $\phi$ . Sachant que les hyperparamètres de NAIS ne sont pas modifiés entre les deux chapitres, la charge de calcul totale s'élève à  $N_{\text{tot}} = N_e \times (N_j \times N_s) = 45 \times 10^3$  appels à  $\phi$  ce qui représente une charge égale à celle du premier algorithme d'ASF. Par ailleurs, comme pour les précédents cas tests, on lance  $M = 10^3$  répliques *bootstrap* de  $\mathbf{u}_{\text{obs}}$  pour garantir la convergence du schéma numérique  $\mathcal{A}$ . Le premier étage du Tableau 5.8 présente les résultats obtenus après  $N_{\text{rep}} = 10$  exécutions de ce *workflow*. On constate

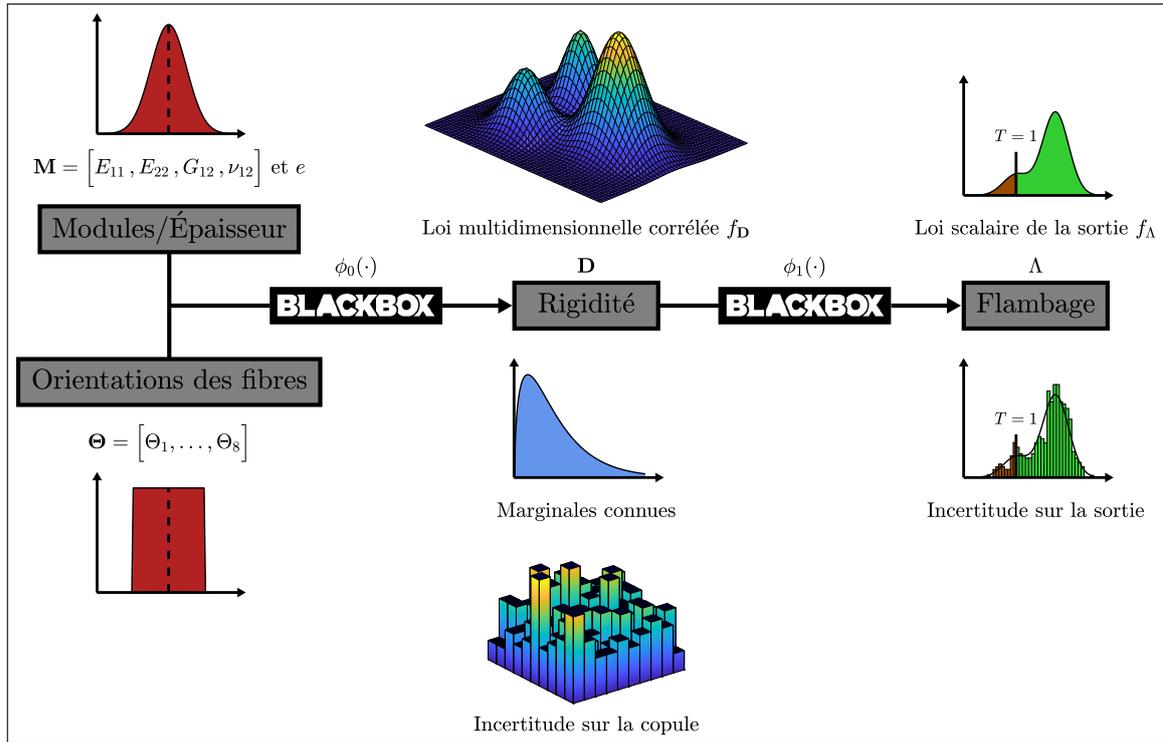


FIGURE 5.15 – Adaptation de la chaîne de propagation des incertitudes.

que les écarts-types des indices  $\hat{\delta}_{ij}^F$  et de l'indicateur de stochasticité  $\hat{\xi}^F$  sont très petits et cela vient appuyer certaines affirmations déjà formulées.

- La fabrication à chaque exécution de l'algorithme d'ASF d'une nouvelle densité auxiliaire  $\hat{h}_{\text{MIS}}$  ne détériore pas la robustesse des estimateurs  $\hat{\delta}_{ij}^F$ .
- L'emploi d'une méthode "orientée copule" combinée avec des R-vignes non-paramétriques garantit une estimation précise du  $\delta$ -indice d'ordre total  $\hat{\xi}^F$ .

On remarque que l'indice  $\hat{\delta}_{13}$  est désigné comme un vainqueur très net de l'ASF. Il semble ainsi que ce soit la copule  $\hat{c}_{13}(\cdot | \mathbf{U}_{\text{obs}})$  qui exerce le plus d'influence sur la distribution de  $\hat{P}_f(\mathbf{U}_{\text{obs}})$ . Si un processus d'acquisition ciblée est mis en place, il doit obligatoirement viser la paire  $(U_1, U_3)$ . Lorsqu'on examine le Tableau 4.10 où la copule  $\hat{c}_{\mathbf{U}}$  est identifiée comme la composante fonctionnelle la plus influente, ce ne sont pas les densités  $\hat{f}_{X_1}$  et  $\hat{f}_{X_3}$  qui occupent les deux premières places au sein du classement interne à l'apprentissage des lois marginales. Pour trouver la copule bidimensionnelle  $c_{ij}$  qui véhicule le plus d'incertitude au moment de son apprentissage, il ne suffit pas de regarder quels sont les deux plus grands indices de Sobol  $\hat{S}_i$  et  $\hat{S}_j$ . Il y a donc une véritable plus-value pour l'utilisateur dans la réponse délivrée par la seconde procédure d'ASF.

Un scénario d'assimilation de données supplémentaires est expérimenté pour  $(U_1, U_3)$  et les résultats sont consignés au second étage du Tableau 5.8 pour deux valeurs différentes de  $N'$ . Un échantillon  $\mathbf{x}_{\text{extra}}$  est construit en suivant la méthode décrite dans l'équation (5.170) puis une re-normalisation avec les fonctions de répartitions  $\hat{K}_j$  donne  $\mathbf{u}_{\text{extra}}$ . En termes d'ASF, l'incorporation de  $\mathbf{u}_{\text{extra}}^{[\cdot(ij)]}$  se traduit par une diminution de  $\hat{\delta}_{13}^F$  accompagnée d'une stagnation de tous les autres indices. Contrairement à ce qui a été observé et commenté à la Section 5.5.2.7,  $\hat{\xi}^F$  prend une valeur relativement élevée et celle-ci ne chute pas après l'assimilation de nouvelles données. On peut en tirer plusieurs enseignements :

- L'application  $\tilde{\Psi}_{\tau}$  a un caractère stochastique limité malgré l'utilisation de  $\tilde{\mathcal{E}}$  et de  $\mathcal{R}_w$ .
- La matrice  $\hat{\mathbf{i}}(\mathbf{U}_{\text{obs}})$  permet d'expliquer une large part de l'incertitude affectant  $\hat{P}_f(\mathbf{U}_{\text{obs}})$ .

			ASF exécutée avec $\mathcal{B}$ :						$\hat{\xi}^F$	
			$\hat{\delta}_{12}^F$	$\hat{\delta}_{13}^F$	$\hat{\delta}_{14}^F$	$\hat{\delta}_{23}^F$	$\hat{\delta}_{24}^F$	$\hat{\delta}_{34}^F$		
$\mathbf{u}_{\text{obs}}$	N = 300	moyenne :	0.12	0.41	0.09	0.13	0.22	0.13	0.66	
		écart-type :	0.02	0.01	0.02	0.02	0.01	0.01	0.06	
		rang :	(5)	(1)	(6)	(4)	(2)	(3)	×	
			ASF exécutée avec $\mathcal{B}_{13}$ :						$\hat{\xi}^F$	
Paire $\mathbf{U}_{13}$	$\mathbf{u}_{\text{obs}} \cup \mathbf{u}_{\text{extra}}^{[*](13)}$	N' = $2 \times 10^3$	moyenne :	0.07	0.36	0.03	0.12	0.22	0.12	0.60
			écart-type :	0.02	0.02	0.02	0.02	0.02	0.02	0.02
			rang :	(5)	(1)	(6)	(3)	(2)	(4)	×
		N' = $10^4$	moyenne :	0.08	0.21	0.04	0.12	0.33	0.10	0.58
			écart-type :	0.03	0.02	0.01	0.03	0.03	0.04	0.02
			rang :	(5)	(2)	(6)	(3)	(1)	(4)	×

TABLEAU 5.8 – Résultats de l’ASF initiale et du scénario d’assimilation de données.

- (c) La stochasticité interne à  $\mathcal{R}_w$  est réduite à un niveau où elle reste minoritaire même après une première action visant à réduire la variabilité d’apprentissage.

Les trois points mentionnés vont dans le sens d’une validation globale de la seconde procédure d’ASF proposée.

## 5.7 Conclusion

Le point de départ de ce chapitre est le constat selon lequel l’estimation de la copule peut être la tâche la plus importante au moment de la réalisation d’une prédiction fiable. On a supposé que les lois marginales étaient connues, quitte à les figer après un apprentissage statistique préliminaire. Comme il n’existe pas une sous-décomposition universelle et pleinement satisfaisante de la copule, le parti pris de ce travail a résidé dans le choix de s’intéresser aux incertitudes véhiculées par les copules bidimensionnelles. Contrairement au problème de la quantification de l’influence des composantes de la séparation copule-marginales qui a été traité au Chapitre 4 en manipulant des entités fonctionnelles, le problème de la quantification de l’influence des copules bidimensionnelles n’a pas été résolu avec une approche fonctionnelle. On a préféré construire un résumé matriciel de la copule estimée et transformer le problème initial pour se ramener à des entrées scalaires.

Une procédure d’ASF a été développée pour être capable d’identifier le coefficient de corrélation de Kendall dont l’estimation a le plus d’impact sur la distribution de l’estimateur du risque de défaillance. Si certaines briques de calcul comme le ré-échantillonnage *bootstrap* ou l’astuce “*inverse-reverse*” ont été directement empruntées au Chapitre 4, ce second algorithme d’ASF inclut aussi certaines spécificités telles que l’emploi des indices de Borgonovo et la mise en œuvre des schémas d’estimation associés. La collection des  $\delta$ -indices du premier ordre permet de comparer l’influence des différents coefficients de corrélation et, à travers eux, de comparer l’influence des différentes copules bidimensionnelles. Par ailleurs, le  $\delta$ -indice d’ordre total fournit une mesure de la part d’incertitude qui ne peut pas être expliquée par la seule connaissance du résumé matriciel de la copule.

Les différents cas tests du Chapitre 4 ont été repris et prolongés pour valider la nouvelle méthodologie. Dans un premier temps, on a pu vérifier que le couplage des différentes briques de calcul conduit à des estimateurs robustes des  $\delta$ -indices. Dans un second temps, les scénarios d’assimilation de données supplémentaires ont montré une évolution cohérente des résultats d’ASF qui s’est caractérisée par une réduction globale de la variabilité d’apprentissage et une redistribution des parts d’influence.

## Références

- AAS, K., C. CZADO, A. FRIGESSI et H. BAKKEN. 2009, «Pair-copula constructions of multiple dependence», *Insurance : Mathematics and economics*, vol. 44, n° 2, p. 182–198.
- ANSTETT-COLLIN, F., J. GOFFART, T. MARA et L. DENIS-VIDAL. 2015, «Sensitivity analysis of complex models : coping with dynamic and static inputs», *Reliability Engineering & System Safety*, vol. 134, p. 268–275.
- BAUCELLS, M. et E. BORGONOVO. 2013, «Invariant probabilistic sensitivity analysis», *Management Science*, vol. 59, n° 11, p. 2536–2549.
- BENES, V. et J. STEPÁN. 2012, *Distributions with given marginals and moment problems*, Springer Science & Business Media.
- BORGONOVO, E. 2007, «A new uncertainty importance measure», *Reliability Engineering & System Safety*, vol. 92, n° 6, p. 771–784.
- BORGONOVO, E., G. B. HAZEN et E. PLISCHKE. 2016, «A common rationale for global sensitivity measures and their estimation», *Risk Analysis*, vol. 36, n° 10, p. 1871–1895.
- BORGONOVO, E., S. TARANTOLA, E. PLISCHKE et M. MORRIS. 2014, «Transformations and invariance in the sensitivity analysis of computer experiments», *Journal of the Royal Statistical Society : Series B : Statistical Methodology*, p. 925–947.
- BRECHMANN, E. C., C. CZADO et K. AAS. 2012, «Truncated regular vines in high dimensions with application to financial data», *Canadian Journal of Statistics*, vol. 40, n° 1, p. 68–85.
- CHASTAING, G., F. GAMBOA, C. PRIEUR et collab.. 2012, «Generalized Hoeffding-Sobol decomposition for dependent variables-application to sensitivity analysis», *Electronic Journal of Statistics*, vol. 6, p. 2420–2448.
- CHRISTENSEN, D. 2005, «Fast algorithms for the calculation of Kendall's  $\tau$ », *Computational Statistics*, vol. 20, n° 1, p. 51–62.
- CLEMEN, R. T. et T. REILLY. 1999, «Correlations and copulas for decision and risk analysis», *Management Science*, vol. 45, n° 2, p. 208–224.
- CLIFF, N. 2014, *Ordinal methods for behavioral data analysis*, Psychology Press.
- CLIFF, N. et V. CHARLIN. 1991, «Variances and covariances of Kendall's tau and their estimation», *Multivariate Behavioral Research*, vol. 26, n° 4, p. 693–707.
- COOKE, R. M., H. JOE et K. AAS. 2010, «Vines arise», dans *Dependence Modeling : Vine Copula Handbook*, World Scientific, p. 37–71.
- CUI, L., Z. LÜ et X. ZHAO. 2010, «Moment-independent importance measure of basic random variable and its probability density evolution solution», *Science China Technological Sciences*, vol. 53, n° 4, p. 1138–1145.
- DANIELS, H. et M. G. KENDALL. 1947, «The significance of rank correlations where parental correlation exists», *Biometrika*, vol. 34, n° 3/4, p. 197–208.
- DERENNES, P., J. MORIO et F. SIMATOS. 2019, «Simultaneous estimation of complementary moment independent sensitivity measures for reliability analysis», *arXiv preprint arXiv :1911.02488*.
- DERUMIGNY, A. et J.-D. FERMANIAN. 2017, «About tests of the “simplifying” assumption for conditional copulas», *Dependence Modeling*, vol. 5, n° 1, p. 154–197.

- DISSMANN, J., E. C. BRECHMANN, C. CZADO et D. KUROWICKA. 2013, «Selecting and estimating regular vine copulae and application to financial returns», *Computational Statistics & Data Analysis*, vol. 59, p. 52–69.
- FANG, K. W. 2018, *Symmetric multivariate and related distributions*, CRC Press, 42–48 p.
- GAMBOA, F., P. GREMAUD, T. KLEIN et A. LAGNOUX. 2020, «Global sensitivity analysis : a new generation of mighty estimators based on rank statistics», *arXiv preprint arXiv :2003.01772*.
- GEENENS, G., A. CHARPENTIER, D. PAINDAVEINE et collab.. 2017, «Probit transformation for non-parametric kernel estimation of the copula density», *Bernoulli*, vol. 23, n° 3, p. 1848–1873.
- GENEST, C., E. MASIELLO et K. TRIBOULEY. 2009, «Estimating copula densities through wavelets», *Insurance : Mathematics and Economics*, vol. 44, n° 2, p. 170–181.
- GILLESPIE, D. T. 1976, «A general method for numerically simulating the stochastic time evolution of coupled chemical reactions», *Journal of computational physics*, vol. 22, n° 4, p. 403–434.
- GILQUIN, L., C. PRIEUR et E. ARNAUD. 2015, «Replication procedure for grouped Sobol' indices estimation in dependent uncertainty spaces», *Information and Inference : A Journal of the IMA*, vol. 4, n° 4, p. 354–379.
- HAN, D., K. S. TAN et C. WENG. 2017, «Vine copula models with GLM and sparsity», *Communications in Statistics-Theory and Methods*, vol. 46, n° 13, p. 6358–6381.
- HART, J. L., A. ALEXANDERIAN et P. A. GREMAUD. 2017, «Efficient computation of Sobol' indices for stochastic models», *SIAM Journal on Scientific Computing*, vol. 39, n° 4, p. A1514–A1530.
- HIGHAM, N. J. 2002, «Computing the nearest correlation matrix : a problem from finance», *IMA journal of Numerical Analysis*, vol. 22, n° 3, p. 329–343.
- IMAN, R. L. et W.-J. CONOVER. 1982, «A distribution-free approach to inducing rank correlation among input variables», *Communications in Statistics-Simulation and Computation*, vol. 11, n° 3, p. 311–334.
- JACQUES, J., C. LAVERGNE et N. DEVICTOR. 2006, «Sensitivity analysis in presence of model uncertainty and correlated inputs», *Reliability Engineering & System Safety*, vol. 91, n° 10–11, p. 1126–1134.
- JACQUES, J. et C. PREDÀ. 2014, «Functional data clustering : a survey», *Advances in Data Analysis and Classification*, vol. 8, n° 3, p. 231–255.
- JOE, H. 2014, *Dependence modeling with copulas*, CRC press.
- KAUERMANN, G., C. SCHELLHASE et D. RUPPERT. 2013, «Flexible copula density estimation with penalized hierarchical B-splines», *Scandinavian Journal of Statistics*, vol. 40, n° 4, p. 685–705.
- KENDALL, M. G. 1948, «Rank correlation methods.», *Griffin*.
- KNIGHT, W. R. 1966, «A computer method for calculating Kendall's tau with ungrouped data», *Journal of the American Statistical Association*, vol. 61, n° 314, p. 436–439.
- KRAUS, D. et C. CZADO. 2017, «Growing simplified vine copula trees : improving Dissmann's algorithm», *arXiv preprint arXiv :1703.05203*.
- KUROWICKA, D. et R. COOKE. 2001, «Conditional, partial and rank correlation for elliptical copula», *Dependence Modeling in Uncertainty Analysis, Proc. of ESREL 2001*.
- KUROWICKA, D. et R. COOKE. 2003, «A parameterization of positive definite matrices in terms of partial correlation vines», *Linear Algebra and its Applications*, vol. 372, p. 225–251.

- KUROWICKA, D. et R. M. COOKE. 2006, *Uncertainty analysis with high dimensional dependence modelling*, John Wiley & Sons.
- KUROWICKA, D. et R. M. COOKE. 2007, «Sampling algorithms for generating joint uniform distributions using the vine-copula method», *Computational statistics & data analysis*, vol. 51, n° 6, p. 2889–2906.
- KURZ, M. S. et F. SPANHEL. 2017, «Testing the simplifying assumption in high-dimensional vine copulas», *arXiv preprint arXiv :1706.02338*.
- LEBRUN, R. et A. DUTFOY. 2009, «A generalization of the Nataf transformation to distributions with elliptical copula», *Probabilistic Engineering Mechanics*, vol. 24, n° 2, p. 172–178.
- LINDSKOG, F., A. MCNEIL et U. SCHMOCK. 2003, «Kendall's tau for elliptical distributions», dans *Credit Risk*, Springer, p. 149–156.
- LONG, J. D. et N. CLIFF. 1997, «Confidence intervals for Kendall's tau», *British Journal of Mathematical and Statistical Psychology*, vol. 50, n° 1, p. 31–41.
- MARA, T. A., S. TARANTOLA et P. ANNONI. 2015, «Non-parametric methods for global sensitivity analysis of model output with dependent inputs», *Environmental modelling & software*, vol. 72, p. 173–183.
- MARREL, A., B. IOOSS, B. LAURENT et O. ROUSTANT. 2009, «Calculations of Sobol indices for the Gaussian process metamodel», *Reliability Engineering & System Safety*, vol. 94, n° 3, p. 742–751.
- MARX, B. D. et P. H. EILERS. 2005, «Multidimensional penalized signal regression», *Technometrics*, vol. 47, n° 1, p. 13–22.
- MORALES, O., D. KUROWICKA et A. ROELEN. 2008, «Eliciting conditional and unconditional rank correlations from conditional probabilities», *Reliability Engineering & System Safety*, vol. 93, n° 5, p. 699–710.
- NAGLER, T. et C. CZADO. 2016, «Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas», *Journal of Multivariate Analysis*, vol. 151, p. 69–89.
- NAGLER, T., C. SCHELLHASE et C. CZADO. 2017, «Nonparametric estimation of simplified vine copula models : comparison of methods», *Dependence Modeling*, vol. 5, n° 1, p. 99–120.
- NANTY, S. 2015, *Quantification des incertitudes et analyse de sensibilité pour codes de calcul à entrées fonctionnelles et dépendantes*, thèse de doctorat, Université de Grenoble.
- NANTY, S., C. HELBERT, A. MARREL, N. PÉROT et C. PRIEUR. 2016, «Sampling, metamodeling, and sensitivity analysis of numerical simulators with functional stochastic inputs», *SIAM/ASA Journal on Uncertainty Quantification*, vol. 4, n° 1, p. 636–659.
- NANTY, S., C. HELBERT, A. MARREL, N. PÉROT et C. PRIEUR. 2017, «Uncertainty quantification for functional dependent random variables», *Computational Statistics*, vol. 32, n° 2, p. 559–583.
- NATAF, A. 1962, «Détermination des distributions dont les marges sont données», *Comptes Rendus de l'Académie des Sciences*, vol. 225, p. 42–43.
- NELSEN, R. B. 2007, *An introduction to copulas*, Springer Science & Business Media.
- RASMUSSEN, C. E. 2003, «Gaussian processes in machine learning», dans *Summer school on machine learning*, Springer, p. 63–71.
- ROSENBLATT, M. 1952, «Remarks on a multivariate transformation», *The annals of mathematical statistics*, vol. 23, n° 3, p. 470–472.

- RÜSCHENDORF, L. 1985, «Construction of multivariate distributions with given marginals», *Annals of the Institute of Statistical Mathematics*, vol. 37, n° 2, p. 225–233.
- SARAZIN, G., P. DERENNES et J. MORIO. 2020, «Estimation of high-order moment-independent importance measures for Shapley value analysis», *Applied Mathematical Modelling*, vol. 88, p. 396–417.
- SCHELLHASE, C. et F. SPANHEL. 2018, «Estimating non-simplified vine copulas using penalized splines», *Statistics and Computing*, vol. 28, n° 2, p. 387–409.
- SCHUSTER, E. F. 1985, «Incorporating support constraints into nonparametric estimators of densities», *Communications in Statistics-Theory and methods*, vol. 14, n° 5, p. 1123–1136.
- SILVERMAN, B. W. 1986, *Density estimation for statistics and data analysis*, vol. 26, CRC press.
- SKLAR, A. 1959, «Fonctions de répartition à  $n$  dimensions et leurs marges», *Publ. inst. statist. univ. Paris*, vol. 8, p. 229–231.
- SOBOL, I. M. 2001, «Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates», *Mathematics and computers in simulation*, vol. 55, n° 1-3, p. 271–280.
- STEIN, M. L. 2012, *Interpolation of spatial data : some theory for kriging*, Springer Science & Business Media.
- SUDRET, B. et Y. CANIOU. 2013, *Analysis of covariance (ANCOVA) using polynomial chaos expansions*, CRC Press/Balkema.
- TANG, X.-S., D.-Q. LI, G. RONG, K.-K. PHOON et C.-B. ZHOU. 2013, «Impact of copula selection on geotechnical reliability under incomplete probability information», *Computers and Geotechnics*, vol. 49, p. 264–278.
- TORRE, E., S. MARELLI, P. EMBRECHTS et B. SUDRET. 2019, «A general framework for data-driven uncertainty quantification under complex input dependencies using vine copulas», *Probabilistic Engineering Mechanics*, vol. 55, p. 1–16.
- VATTER, T. et T. NAGLER. 2018, «Generalized additive models for pair-copula constructions», *Journal of Computational and Graphical Statistics*, vol. 27, n° 4, p. 715–727.
- WEI, P., Z. LU et J. SONG. 2014, «Moment-independent sensitivity analysis using copula», *Risk Analysis*, vol. 34, n° 2, p. 210–222.
- WEI, P., Z. LU et X. YUAN. 2013, «Monte Carlo simulation for moment-independent sensitivity analysis», *Reliability Engineering & System Safety*, vol. 110, p. 60–67.
- WILKINSON, D. J. 2009, «Stochastic modelling for quantitative description of heterogeneous biological systems», *Nature Reviews Genetics*, vol. 10, n° 2, p. 122–133.
- YULE, G. et M. KENDALL. 1965, *An introduction to the Theory of Statistics*, Charles Griffin & Co, London.
- ZHANG, L. et V. P. SINGH. 2007, «Gumbel-Hougaard copula for trivariate rainfall frequency analysis», *Journal of Hydrologic Engineering*, vol. 12, n° 4, p. 409–419.
- ZHOU, Y., Z. LU, Y. SHI et K. CHENG. 2018, «A vine copula-based method for analyzing the moment-independent importance measure of the multivariate output», *Proceedings of the Institution of Mechanical Engineers, Part O : Journal of Risk and Reliability*, p. 1748 006X18781 121.

# Conclusion

## Résumé des principales contributions

Dans un contexte où l'état de connaissance de la loi des entrées d'un système complexe était limité à un unique échantillon  $\mathbf{x}_{\text{obs}} \in \mathbb{R}^{N \times d}$  formé par un petit nombre d'observations conjointes, la motivation principale de cette thèse était la mise en place d'une démarche statistico-fiabiliste d'estimation robuste de la probabilité de défaillance. Les objectifs établis dans l'introduction de ce travail étaient doubles. D'une part, on voulait anticiper et évaluer l'impact de la variabilité d'échantillonnage sur les résultats de l'analyse de fiabilité. D'autre part, sachant qu'il était fort probable que le niveau de dispersion des estimations soit important, on souhaitait développer, implémenter et expérimenter des techniques de réduction de la variabilité des estimations. Dans ce but, un intérêt particulier était porté aux stratégies basées sur l'acquisition ciblée de données supplémentaires pour certaines variables d'entrée (idéalement une unique variable ou une paire de variables). À cet égard, on a compris qu'il fallait procéder en plusieurs étapes :

- Diviser la distribution d'entrée en plusieurs composantes (paramétriques ou fonctionnelles) pouvant être raffinées à partir d'observations supplémentaires de certaines variables.
- Concevoir un algorithme d'ASF dans le but d'identifier la ou les composante(s) qui exerce(nt) le plus d'influence sur l'estimation de la probabilité de défaillance.
- Acquérir des données supplémentaires pour la composante ciblée, les assimiler et s'assurer qu'elles conduisent à une réduction de la variabilité des estimations.

## Analyse de la sensibilité à l'apprentissage des lois marginales et de la copule

Dans la première partie de ce travail (Chapitre 4), on suggère de diviser la distribution d'entrée à l'aide de la décomposition copule-marginales énoncée par le théorème de SKLAR [1959]. Il faut alors quantifier la variabilité engendrée par chaque composante fonctionnelle lors d'une phase d'apprentissage (déterministe) impliquant un échantillon aléatoire  $\mathbf{X}_{\text{obs}}$ . Le problème d'ASF peut ainsi être résumé par :

$$\Psi_1 : \quad \mathbb{L}^2(\mathbb{R}) \quad \times \dots \times \quad \mathbb{L}^2(\mathbb{R}) \quad \times \quad \mathbb{L}^2(\mathcal{H}_d) \quad \longrightarrow \quad [0, 1] \\ \hat{f}_{X_1}(\cdot | \mathcal{M}_1, \mathbf{X}_{\text{obs}}) \quad \dots \quad \hat{f}_{X_d}(\cdot | \mathcal{M}_d, \mathbf{X}_{\text{obs}}) \quad \hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{X}_{\text{obs}}) \longmapsto \hat{P}_f(\mathbf{X}_{\text{obs}}) \quad , \quad (\text{ASF 1})$$

avec  $\hat{f}_{X_j}(\cdot | \mathcal{M}_j, \mathbf{X}_{\text{obs}})$  la  $j$ -ème densité marginale (estimée selon le modèle d'apprentissage  $\mathcal{M}_j$ ) et  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{X}_{\text{obs}})$  la densité de la copule (estimée selon le modèle d'apprentissage  $\mathcal{M}_{\text{cop}}$ ). Les entrées de  $\Psi_1$  sont des fonctions aléatoires potentiellement dépendantes que l'on est incapable de simuler étant donné que la loi  $\mu_{\mathbf{X}}$  est inconnue. Une version simplifiée du problème (ASF 1) a été examinée en supposant qu'il était possible de fabriquer des échantillons d'apprentissage indépendants  $\mathbf{Z}_{\text{obs}}^{[1]} \perp \dots \perp \mathbf{Z}_{\text{obs}}^{[d]} \perp \mathbf{V}_{\text{obs}}$  pour les différentes composantes fonctionnelles :

$$\Psi_2 : \quad \mathbb{L}^2(\mathbb{R}) \quad \times \dots \times \quad \mathbb{L}^2(\mathbb{R}) \quad \times \quad \mathbb{L}^2(\mathcal{H}_d) \quad \longrightarrow \quad [0, 1] \\ \hat{f}_{X_1}(\cdot | \mathcal{M}_1, \mathbf{Z}_{\text{obs}}^{[1]}) \quad \dots \quad \hat{f}_{X_d}(\cdot | \mathcal{M}_d, \mathbf{Z}_{\text{obs}}^{[d]}) \quad \hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{V}_{\text{obs}}) \longmapsto \hat{P}_f(\mathbf{Z}_{\text{obs}}, \mathbf{V}_{\text{obs}}) \quad . \quad (\text{ASF 2})$$

Le problème (ASF 2) a l'avantage de mettre en jeu des composantes fonctionnelles indépendantes les unes des autres. Un algorithme d'ASF a été mis au point (voir Figure 4.18 pour le *workflow*) en rassemblant plusieurs briques de calcul :

1. Un mécanisme empirique de ré-échantillonnage (voir Figure 4.5) qui cherche à imiter la variabilité d'échantillonnage théorique et qui fabrique des échantillons rendant possible un apprentissage indépendant des marginales et de la copule.
2. Le calcul d'indices de Sobol pour des entrées fonctionnelles indépendantes grâce au schéma d'estimation *Pick-Freeze* (voir Figure 4.7).
3. Plusieurs stratégies d'économie du budget de simulation (voir Figure 4.17) gravitant autour de la construction d'une densité auxiliaire d'échantillonnage préférentiel.

L'application de la procédure d'ASF à différents cas tests analytiques a permis de vérifier que les indices de sensibilité estimés quantifient de manière appropriée la part de variabilité induite par chaque composante fonctionnelle suite à son apprentissage statistique. Cette procédure peut convenir à tout type de modèle d'apprentissage  $\mathcal{M}$  (paramétrique, non-paramétrique, hybride) dès lors que celui-ci s'articule autour d'une séparation explicite des marginales et de la copule. De même, tout algorithme d'échantillonnage préférentiel peut être implanté au sein de la procédure à condition que son utilisation soit pertinente dans la configuration fiabiliste examinée.

Les indices de sensibilité obtenus en fin de procédure sont caractéristiques du modèle  $\mathcal{M}$  ayant guidé l'apprentissage mais ils sont indépendants de l'algorithme de fiabilité ayant permis l'estimation (séquentielle ou non) des probabilités de défaillance. Par ailleurs, il faut souligner que ces indices présentent une dépendance envers l'échantillon initial  $\mathbf{x}_{\text{obs}}$ , ce qui était inéluctable étant donné que l'incertitude d'échantillonnage est propagée à l'aide de mesures empiriques. Cet aspect "*data-driven*" des résultats s'accroît avec la sévérité du contexte "*small data*". Toutefois, pour les tailles d'échantillons considérées (de l'ordre de la centaine d'observations), il s'avère que le biais introduit par  $\mathbf{x}_{\text{obs}}$  est assez réduit et qu'il n'empêche pas d'aboutir aux mêmes conclusions que celles qui auraient pu être tirées si toute une séquence d'échantillons avait été fournie.

À chaque fois que l'ASF a désigné une loi marginale comme étant la composante fonctionnelle de plus grande influence, des scénarios d'acquisition ciblés ont été mis en œuvre. L'exécution d'une seconde procédure (de validation) a systématiquement mis en évidence une réduction des incertitudes épistémiques véhiculées par la composante enrichie. Ces résultats ont conforté l'idée qu'il est envisageable de réduire la variabilité sur la prédiction du risque de défaillance en menant une politique d'enrichissement sur une seule loi marginale. Toutefois, on a pu déplorer qu'il soit parfois nécessaire de solliciter plusieurs centaines d'observations supplémentaires de la variable identifiée, c'est-à-dire un budget d'observations comparable à celui de l'échantillon initial.

Cet algorithme d'ASF a donné lieu à un article accepté dans la revue *Reliability Engineering & System Safety* à l'occasion du numéro "*Special Issue on Sensitivity Analysis of Model Output*" :

◇ G. Sarazin, J. Morio, A. Lagnoux, M. Balesdent, L. Brevault, 2020.  
 "Reliability-oriented sensitivity analysis in presence of data-driven epistemic uncertainty".  
*Reliability Engineering and System Safety*.

## Analyse de la sensibilité à l'apprentissage des copules bidimensionnelles

Comme il a été choisi de décomposer la variabilité d'apprentissage à partir de la séparation copule-marginales, il est tout à fait possible que ce soit la densité de copule  $\hat{c}_{\mathbf{U}}(\cdot \mid \mathcal{M}_{\text{cop}}, \mathbf{X}_{\text{obs}})$  qui soit désignée comme la composante fonctionnelle de plus grande influence. Cette situation n'est pas compatible avec une acquisition ciblée de données supplémentaires car l'amélioration de l'état de connaissance de la copule passe par la collection d'observations conjointes. Ce constat est venu motiver la seconde partie de ce travail (Chapitre 5) qui est consacrée au développement d'une procédure d'ASF complémentaire. Afin de se concentrer sur la variabilité d'apprentissage de la copule, il est supposé que les lois marginales sont connues (ou du moins figées). Dans ce contexte, le matériel fourni se résume à un unique échantillon  $\mathbf{u}_{\text{obs}} \in [0, 1]^{N \times d}$  de la copule.

Pour essayer de diviser la copule multidimensionnelle en sous-composantes fonctionnelles, on s'est intéressé à l'information portée par les différentes copules bidimensionnelles. Ce choix est apparu comme assez naturel pour deux raisons.

- D'une part, l'état de connaissance d'une copule bidimensionnelle peut être raffiné assez facilement à partir d'observations supplémentaires de la paire de variables correspondante. Cette action apparaît alors comme une simple généralisation de la politique d'acquisition ciblée de données supplémentaires entreprise dans la première partie.
- D'autre part, les modèles d'apprentissage de la copule multidimensionnelle reposent le plus souvent sur l'assemblage de plusieurs copules bidimensionnelles et accordent donc une place importante à leur estimation.

En bref, les copules bidimensionnelles sont impliquées structurellement dans l'expression de la densité de copule multidimensionnelle et il est possible d'intervenir pour réduire leur variabilité d'apprentissage. Il a donc semblé judicieux de quantifier l'influence de l'apprentissage des copules bidimensionnelles sur l'estimation de la probabilité de défaillance. Ce nouveau problème d'ASF peut être formalisé de la manière suivante :

$$\Psi_3: \quad \mathbb{L}^2(\mathcal{H}_2) \quad \times \quad \mathbb{L}^2(\mathcal{H}_2) \quad \times \dots \times \quad \mathbb{L}^2(\mathcal{H}_2) \quad \longrightarrow \quad [0, 1] \\ \hat{c}_{12}(\cdot | \mathbf{U}_{\text{obs}}) \quad \hat{c}_{13}(\cdot | \mathbf{U}_{\text{obs}}) \quad \dots \quad \hat{c}_{(d-1)d}(\cdot | \mathbf{U}_{\text{obs}}) \quad \longmapsto \quad \hat{P}_f(\mathbf{U}_{\text{obs}}) \quad . \quad (\text{ASF 3})$$

On retrouve les difficultés du problème (ASF 1), à savoir la manipulation de fonctions aléatoires potentiellement dépendantes que l'on ne sait pas simuler (par méconnaissance de la loi  $\mu_{\mathbf{U}}$ ). À cela, s'ajoute une difficulté supplémentaire qui vient du fait qu'il n'y a pas de réécriture universelle de la copule multidimensionnelle à partir de ses copules bidimensionnelles. De plus, comme les copules bidimensionnelles ne permettent pas de décrire toute l'information portée par la copule multidimensionnelle, l'application  $\Psi_3$  est stochastique. La solution proposée pour répondre au problème (ASF 3) consiste à transformer les entrées fonctionnelles en des entrées scalaires. Pour cela, on a décidé de synthétiser chaque copule  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$  par son coefficient empirique  $\hat{t}_{ij}$  de corrélation de Kendall. Il en découle une version dégradée du problème (ASF 3) :

$$\Psi_4: \quad [-1, 1] \quad \times \quad [-1, 1] \quad \times \dots \times \quad [-1, 1] \quad \longrightarrow \quad [0, 1] \\ \hat{t}_{12} \quad \hat{t}_{13} \quad \dots \quad \hat{t}_{(d-1)d} \quad \longmapsto \quad \hat{P}_f(\mathbf{U}_{\text{obs}}) \quad . \quad (\text{ASF 4})$$

L'idée est donc d'identifier le coefficient de corrélation de Kendall de plus grande influence sur l'estimation de la probabilité de défaillance et d'en déduire la copule bidimensionnelle de plus grande influence. La perte d'information causée par l'usage d'un résumé matriciel de la copule multidimensionnelle entraîne une augmentation du niveau de stochasticité entre les applications  $\Psi_3$  et  $\Psi_4$ . Dans cette situation, un algorithme d'ASF a été imaginé (voir Figure 5.5 pour le *workflow*) en réunissant les éléments suivants :

1. Le *framework* du premier algorithme d'ASF pour ce qui est du ré-échantillonnage *bootstrap* et de la construction astucieuse d'une densité auxiliaire d'échantillonnage préférentiel.
2. Le calcul d'indices de Boronovo pour une application stochastique dont les entrées sont scalaires mais dépendantes. Un schéma d'estimation "orienté copule" a été mis en œuvre.
3. La définition et le calcul d'un indicateur de stochasticité permettant de mesurer la perte d'information provoquée par le recours à un résumé matriciel de la copule. L'estimation de cet indicateur a nécessité le développement d'un schéma numérique spécifique.

La réponse fournie par ce second algorithme d'ASF n'est en fait pas suffisante pour mener à bien toute la démarche de réduction de la variabilité d'apprentissage. En effet, contrairement à ce qui est rendu possible par la séparation copule-marginales dans la première partie, l'absence d'une décomposition rigoureuse de la copule multidimensionnelle est un frein à l'exploitation d'éventuelles données supplémentaires collectées pour une certaine paire de variables. Pour y parvenir, il faut réussir à adapter le modèle d'apprentissage pour qu'il soit en mesure de traiter un échantillon non rectangulaire. Des solutions allant dans ce sens ont été proposées dans deux cas particuliers (la copule gaussienne et les copules R-vignes). Dans chacun des cas, il a été vérifié sur des cas tests analytiques que l'assimilation de données supplémentaires engendre effectivement une réduction du volume d'incertitudes épistémiques.

La méthodologie consistant à dégrader le problème (ASF 3) pour étudier le problème (ASF 4) et en extrapoler les résultats a fait l'objet d'une communication ainsi que d'une publication dans les actes de la conférence internationale *European Safety and Reliability* :

- ◊ G. Sarazin, J. Morio, A. Lagnoux, M. Balesdent, L. Brevault.  
"Sensitivity analysis of risk assessment with data-driven dependence modelling".  
*29th European Safety and Reliability Conference*, Hannovre, Septembre 2019.

Par ailleurs, l'estimation de l'indicateur de stochasticité s'inscrit dans le contexte bien plus large de l'estimation des indices de Borgonovo d'ordre supérieur. Le schéma numérique développé est donc en soi un apport scientifique notable. Ce travail a été réalisé en collaboration avec **Pierre Derennes**<sup>7</sup>. Certaines idées peuvent donc être retrouvées dans le dernier chapitre de sa thèse<sup>8</sup> de doctorat. La démarche proposée et les résultats obtenus ont mérité une publication dans le journal *Applied Mathematical Modelling* :

- ◊ G. Sarazin, P. Derennes et J. Morio, 2020.  
"Estimation of high-order moment-independent importance measures for Shapley value analysis".  
*Applied Mathematical Modelling*, vol. 88, p. 396-417.

Les deux algorithmes développés dans cette thèse doivent être vus comme des outils d'aide à la décision. Ils permettent à l'ingénieur d'identifier la composante fonctionnelle (loi marginale ou copule bidimensionnelle) dont la connaissance doit être améliorée afin de réduire la variabilité sur la prédiction du risque de défaillance. Un schéma de synthèse de la procédure complète est représenté à la Figure 4. Les deux phases de cette démarche d'ASF ont été appliquées sur un cas test physique proposé par le Département MATériaux et Structures (DMAS) du centre ONERA de Châtillon. Cette étude concernait une plaque composite (stratifié orthotrope dont les huit couches sont empilées symétriquement) soumise à une compression uniaxiale et exposée à un risque de flambage. On a alors pu mettre en lumière la très bonne applicabilité des méthodes proposées.

## Perspectives

### Améliorations de la première étape de la procédure d'ASF

À court terme, en supposant que l'on souhaite prolonger l'étude du problème (ASF 2), des progrès pourraient être envisagés dans la manière dont est choisie la densité auxiliaire  $\hat{h}_{\text{opt}}$  utilisée pour l'estimation par échantillonnage préférentiel des différentes probabilités de défaillance (qui sont des sorties de l'application  $\Psi_2$ ). On rappelle que l'objectif est l'estimation des probabilités de défaillance  $P_f^{(1)}, \dots, P_f^{(M)}$  associées à une séquence  $\hat{g}^{(1)}, \dots, \hat{g}^{(M)}$  de densités d'entrée obtenues en répétant des tâches de ré-échantillonnage et d'apprentissage à partir de l'échantillon  $\mathbf{x}_{\text{obs}}$ . Il faudrait tendre vers une mise à jour dynamique de la densité auxiliaire  $\hat{h}_{\text{opt}}$  à chaque fois que l'on rencontre une densité  $\hat{g}^{(i^*)}$  que l'on juge très différente des précédentes. L'astuce "inverse-reverse" préconise de prendre  $\hat{h}_{\text{opt}} = \hat{h}^{(1)}$  où  $\hat{h}^{(1)}$  est la densité sous-optimale fournie par l'algorithme de fiabilité pour la première densité d'entrée  $\hat{g}^{(1)}$ . Chaque nouvelle probabilité de défaillance  $P_f^{(i)}$  est estimée à partir de  $\hat{h}_{\text{opt}}$  en supposant que seule une très légère perturbation est intervenue entre les densités  $\hat{g}^{(1)}$  et  $\hat{g}^{(i)}$ . Dans cette thèse, en présence d'une forte variabilité d'apprentissage liée au contexte "small data", il a été décidé de lancer  $N_e$  exécutions préliminaires de l'algorithme de fiabilité. Un mélange uniforme des différentes densités auxiliaires fabriquées permet ensuite de construire  $\hat{h}_{\text{opt}}$ . En procédant de la sorte, on limite le risque d'utiliser une densité  $\hat{h}_{\text{opt}}$  uniquement adaptée à l'estimation de  $P_f^{(1)}$ . De plus, en ayant recours à un échantillonnage préférentiel multiple à base de  $N_e$  densités auxiliaires, notre intuition a été de se dire que la densité auxiliaire finale  $\hat{h}_{\text{opt}}$  allait conférer une forte robustesse à la majorité des estimateurs  $\hat{P}_f^{(i)}$ . Cela a été vérifié en pratique sur le cas test physique où il a été choisi de prendre  $N_e = 5$ . Toutefois, le risque de rencontrer

7. Ex-doctorant (2016-2019) rattaché à l'IMT et encadré par Jérôme Morio (ONERA) et Florian Simatos (ISAE).

8. <http://thesesups.ups-tlse.fr/4334/1/2019T0U30039.pdf>

une densité d'entrée inhabituelle  $\hat{g}^{(i^*)}$  (pour laquelle  $\hat{h}_{\text{opt}}$  ne convient pas) n'a pas complètement disparu. Face à ce constat, une façon différente de travailler pourrait être de :

- Exécuter l'algorithme de fiabilité avec  $\hat{g}^{(1)}$  pour obtenir la densité auxiliaire  $\hat{h}^{(1)}$  et utiliser  $\hat{h}_{\text{opt}} = \hat{h}^{(1)}$  jusqu'à nouvel ordre.
- À chaque nouvelle densité d'entrée  $\hat{g}^{(i)}$ , exécuter une procédure qui détermine si  $\hat{h}_{\text{opt}}$  est adaptée à l'estimation de  $P_f^{(i)}$ . Si ce n'est pas le cas, l'algorithme de fiabilité est relancé pour estimer  $\hat{P}_f^{(i)}$  et la densité auxiliaire  $\hat{h}^{(i)}$  est mise en stock. Elle peut par exemple être mélangée avec  $\hat{h}^{(1)}$  pour mettre à jour  $\hat{h}_{\text{opt}}$ .

Pour implémenter cette démarche, il faut être capable de détecter les densités  $\hat{g}^{(i)}$  pour lesquelles l'utilisation de  $\hat{h}_{\text{opt}}$  va conduire à un estimateur  $\hat{P}_f^{(i)}$  de trop grande variance. OWEN [2013] propose un diagnostic d'échantillonnage préférentiel qui consiste à examiner la variance empirique des poids (c'est-à-dire des rapports de vraisemblance entre les densités  $\hat{g}^{(i)}$  et  $\hat{h}_{\text{opt}}$ ). Cette technique de diagnostic repose sur le calcul d'une taille d'échantillon "effective"  $N_{\text{eff}}$  qui indique le nombre de tirages aléatoires suivant  $\hat{g}^{(i)}$  qui permettrait de construire un estimateur CMC de variance égale à l'estimateur obtenu en utilisant la densité auxiliaire  $\hat{h}_{\text{opt}}$ . Un tel diagnostic vient discréditer toute densité auxiliaire conduisant à une petite valeur de  $N_{\text{eff}}$ . Si l'idée semble attrayante, sa mise en application se heurte à de nombreuses embûches.

- Il est difficile de fixer la valeur limite de  $N_{\text{eff}}$  en dessous de laquelle l'algorithme de fiabilité doit être relancé. Il faut trouver un compromis pour que le mécanisme de mise à jour de  $\hat{h}_{\text{opt}}$  soit opérationnel (détection fréquente) sans être intempestif (détection systématique).
- Contrairement aux stratégies utilisées jusqu'à présent, le coût de calcul induit par une mise à jour dynamique de  $\hat{h}_{\text{opt}}$  n'est connu qu'à la fin de la procédure. On pourrait fixer un nombre maximal d'appels à l'algorithme de fiabilité mais cela ouvre un débat sur la manière dont il faut estimer les probabilités de défaillance une fois que le budget de simulation est épuisé.

À plus long terme, une perspective ambitieuse serait de répondre au problème (ASF 1) sans passer par le problème (ASF 2). Même s'il a été justifié au Chapitre 4 que la situation d'indépendance statistique des composantes fonctionnelles issues de la séparation copule-marginales n'est pas irréalisable (et qu'elle est en plus souhaitable), les pratiques standards du statisticien (comme la calibration d'une famille paramétrique multidimensionnelle par maximum de vraisemblance ou la fabrication des pseudo-observations de la copule avec des fonctions de répartitions calibrées ou empiriques) introduisent une dépendance statistique entre les composantes fonctionnelles et rendent inappropriée l'hypothèse d'indépendance. Au Chapitre 4, on a utilisé un mécanisme de ré-échantillonnage spécifique pour se ramener à une situation d'indépendance entre les objets simulés. Idéalement, il faudrait être capable d'isoler et de quantifier la variabilité d'apprentissage affectant chaque composante fonctionnelle dans le cadre d'une procédure inférentielle au cours de laquelle aucune précaution n'est prise pour déconnecter l'estimation des lois marginales et de la copule. Afin de construire un algorithme d'ASF susceptible d'apporter une réponse rigoureuse au problème (ASF 1), on pourrait :

- Simuler de nouveaux échantillons d'apprentissage par *bootstrap* "classique" de  $\mathbf{x}_{\text{obs}}$ .
- Appliquer une méthode d'analyse de sensibilité pour entrées fonctionnelles dépendantes.

Sur ce second point, la littérature est assez peu fournie. On peut mentionner les travaux de thèse de NANTY [2015] où il est question de projeter les différentes fonctions aléatoires dans des bases de fonctions puis de manipuler des blocs de coefficients aléatoires. Cette méthode semble complexe à mettre en œuvre dans le cadre de notre problème, notamment car cela impose de projeter la densité de copule dans une base appropriée. Une autre idée pourrait être le recours aux valeurs de SHAPLEY [1953] mais cela supposerait de disposer d'une fonction caractéristique pouvant être calculée facilement pour des entrées fonctionnelles dépendantes. Dans le cas des indices de Sobol, l'estimation de la variance des espérances conditionnelles est à première vue un problème sans solution immédiate pour des entrées  $\hat{f}_{X_1}(\cdot | \mathcal{M}_1, \mathbf{X}_{\text{obs}}), \dots, \hat{f}_{X_d}(\cdot | \mathcal{M}_d, \mathbf{X}_{\text{obs}})$  et  $\hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{X}_{\text{obs}})$ .

## Améliorations de la seconde étape de la procédure d'ASF

La pierre d'achoppement de la méthode proposée au Chapitre 5 est l'idée de synthétiser toute l'information portée par une copule bidimensionnelle  $\hat{c}_{ij}(\cdot | \mathbf{U}_{\text{obs}})$  à l'aide d'une seule variable scalaire, en l'occurrence le coefficient de corrélation de Kendall (analytique ou empirique). On aboutit alors à un résumé de dimension  $p := C_d^2 = \frac{1}{2}d(d-1)$  de la copule multidimensionnelle. Il pourrait être intéressant d'examiner les résultats de l'ASF en fabriquant un résumé matriciel avec une autre mesure de dépendance que le tau de Kendall. De même, afin de rendre moins brutale la transition entre les problèmes (ASF 3) et (ASF 4), on pourrait utiliser une plus large gamme de statistiques pour résumer chaque copule bidimensionnelle. Par exemple, en incluant le rho de Spearman et les deux coefficients de dépendance de queue, on utiliserait  $r = 4$  statistiques pour décrire chaque copule bidimensionnelle :

$$\forall 1 \leq i < j \leq d, \quad \mathbf{D}_{ij} := \begin{bmatrix} \tau_K(V_i, V_j) \\ \rho_S(V_i, V_j) \\ \lambda_L(V_i, V_j) \\ \lambda_U(V_i, V_j) \end{bmatrix} \quad \text{avec} \quad \mathbf{V} \sim \hat{c}_{\mathbf{U}}(\cdot | \mathcal{M}_{\text{cop}}, \mathbf{U}_{\text{obs}}) .$$

La copule multidimensionnelle serait alors résumée à l'aide d'une matrice  $\mathbf{D}$  de taille  $r \times p$ . Le *workflow* global resterait le même (voir Figure 5.5) sachant que les  $r$  mesures de dépendance peuvent être estimées empiriquement et que cela n'exige aucun appel au code de simulation. La seule différence résiderait dans la nature et dans l'ampleur des tâches à exécuter au moment de la phase de post-traitement. Pour identifier la copule bidimensionnelle de plus grande influence, plusieurs démarches pourraient alors être envisagées.

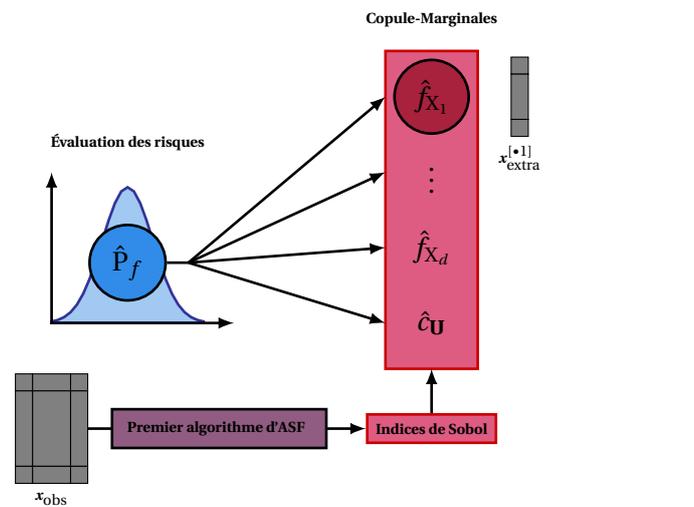
- (a) Estimer l'indicateur de stochasticité pour chaque mesure de dépendance afin d'identifier celle qui correspond à la plus petite perte d'information. Elle est ensuite retenue pour la prise de décision finale consistant à chercher le plus grand des  $\delta$ -indices du premier ordre.
- (b) Estimer le  $\delta$ -indice du premier ordre pour tous les coefficients de la matrice  $\mathbf{D}$ . On dispose alors d'une grille de lecture plus fine. En effet, pour une même copule bidimensionnelle, on peut tout à fait imaginer que le tau de Kendall et un coefficient de dépendance de queue aient des degrés d'influence complètement différents. Si toutes les mesures de dépendance ne désignent pas le même vainqueur, l'identification de la copule bidimensionnelle de plus grande influence peut faire débat.
- (c) Estimer le  $\delta$ -indice pour toutes les colonnes  $\mathbf{D}_{ij}$  de la matrice  $\mathbf{D}$ . En procédant de la sorte, on espère mieux quantifier l'influence globale de chaque copule bidimensionnelle.

Quelle que soit l'approche privilégiée, l'ASF reposerait systématiquement sur l'estimation d'un nombre conséquent de  $\delta$ -indices avec un schéma "orienté copule". En particulier, la mise en œuvre des méthodes (a) et (c) impliquerait l'estimation de  $\delta$ -indices d'ordre  $p$  et  $r$ . Il faut donc rester conscient que la qualité des résultats de ces procédures d'ASF dépendrait principalement de notre capacité à estimer des copules en dimension  $p+1$  et  $r+1$ .

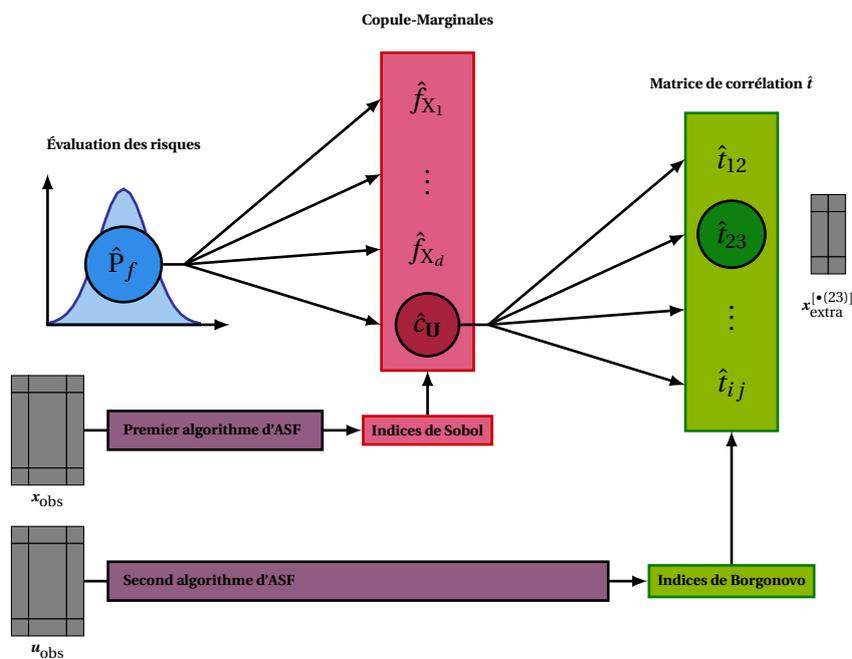
L'une des spécificités du problème (ASF 3) réside dans le fait que la phase d'identification de la copule bidimensionnelle de plus grande influence doit être suivie par une phase d'aménagement du modèle d'apprentissage  $\mathcal{M}_{\text{cop}}$  afin de permettre l'incorporation de nouvelles données pour la paire de variables qui a été ciblée. Au Chapitre 5, on a examiné le cas particulier de la copule gaussienne (voir Schéma 5.10) ainsi que celui des copules R-vignes. On pourrait essayer de faire de même pour d'autres modèles de copules adaptés à la grande dimension.

- Pour les modèles CFM (voir Section 1.5.4.2) ou les modèles PBC (voir Section 1.5.4.4), la calibration des paramètres passe par la maximisation d'une log-vraisemblance conjointe. Il n'est donc pas possible de segmenter l'algorithme d'apprentissage en un ensemble de tâches bidimensionnelles.

- Pour les modèles HAC (voir Section 1.5.4.3), la méthode de OKHRIN et collab. [2013] semble laisser la possibilité d'assimiler des données supplémentaires. Pour cela, il faudrait imposer que les deux variables de la paire ciblée soient des feuilles de l'arbre binaire. Le premier niveau pourrait être calibré en utilisant toutes les données disponibles. On abandonnerait ensuite le surplus de données pour permettre la poursuite de la procédure inférentielle au contact des autres variables. D'une certaine manière, on retrouve le mécanisme que l'on a mis en place pour les R-vignes avec la construction obligatoire d'une arête entre les sommets correspondant aux variables de la paire ciblée. Reste ensuite à savoir comment se comporte le modèle HAC ainsi calibré au regard de la condition d'emboîtement (voir Théorème 1.5.3).
- Pour les modèles PCBN (voir Section 1.5.4.5), en raison de la proximité qu'ils entretiennent avec les copules R-vignes, on peut penser qu'ils sont également compatibles avec ce type d'approche mais cela reste à vérifier.



(a) Cas de l'identification de la densité marginale  $\hat{f}_{X_1}$ .



(b) Cas de l'identification de la copule  $\hat{c}_U$  puis de la copule bidimensionnelle  $\hat{c}_{23}$ .

FIGURE 4 – Synthèse de l'imbrication des deux procédures d'ASF

## Références

- NANTY, S. 2015, *Quantification des incertitudes et analyse de sensibilité pour codes de calcul à entrées fonctionnelles et dépendantes*, thèse de doctorat, Université de Grenoble.
- OKHRIN, O., Y. OKHRIN et W. SCHMID. 2013, «On the structure and estimation of hierarchical Archimedean copulas», *Journal of Econometrics*, vol. 173, n° 2, p. 189–204.
- OWEN, A. B. 2013, «Monte Carlo theory : methods and examples», .
- SHAPLEY, L. S. 1953, «A value for  $n$ -person games», *Contributions to the Theory of Games*, vol. 2, n° 28, p. 307–317.
- SKLAR, A. 1959, «Fonctions de répartition à  $n$  dimensions et leurs marges», *Publ. inst. statist. univ. Paris*, vol. 8, p. 229–231.

## Résumé

Dans de nombreux domaines industriels, l'évaluation de la fiabilité d'un système complexe nécessite l'utilisation d'un code de simulation très coûteux en temps de calcul que l'on peut appréhender comme une fonction de performance déterministe. À partir d'un ensemble de paramètres d'entrée, la sortie du code indique si le système est défaillant. De par leur variabilité naturelle, les entrées sont entachées d'incertitudes aléatoires et leur comportement simultané est décrit par une distribution de probabilité multidimensionnelle. La probabilité de défaillance quantifie alors le risque de dépassement d'un seuil extrême en sortie et elle peut être estimée grâce à un algorithme de fiabilité. Les travaux conduits dans cette thèse s'inscrivent dans un cadre plus large où la distribution d'entrée est inconnue. Le seul matériel disponible est un petit échantillon d'observations conjointes à partir duquel la loi d'entrée peut être estimée via un algorithme d'apprentissage statistique. Afin de garantir la précision de l'estimation de la probabilité de défaillance, il est désormais communément admis qu'il faut capturer la structure de dépendance entre les variables d'entrée en s'appuyant sur la séparation copule-marginales. Dans cette situation, la prédiction fiabiliste découle directement des données fournies et elle est donc impactée par la variabilité d'échantillonnage. Dans la littérature, ce second niveau d'incertitude est qualifié d'épistémique et il est principalement étudié dans le cadre de distributions paramétriques en adoptant le point de vue bayésien. Le principal objectif de ce travail est la construction d'une démarche d'analyse de sensibilité globale permettant d'identifier la composante fonctionnelle de la distribution d'entrée qui engendre le plus de variabilité au moment de son estimation. La réponse à ce problème est très utile si on cherche à réduire la variabilité de la prédiction fiabiliste car elle permet de se focaliser uniquement sur la recherche de données supplémentaires pour la composante ciblée. La première contribution est une procédure d'analyse de sensibilité visant à séparer les incertitudes issues de l'estimation de la densité de copule et celles issues de l'estimation des densités marginales. Pour cela, il faut être capable de caractériser la variabilité d'échantillonnage à partir d'un unique jeu de données. À la suite de cette première analyse, si la copule est identifiée comme la composante la plus influente, il est assez difficile d'envisager une amélioration de l'état de connaissance de la loi d'entrée. Comme la plupart des modèles de copules conçus pour la montée en dimension s'articulent autour d'un assemblage de copules bidimensionnelles, il est naturel de chercher la copule bidimensionnelle dont l'estimation est la plus influente sur la prédiction fiabiliste. La principale difficulté réside dans le fait qu'il n'existe aucune décomposition systématique permettant de réécrire une copule en fonction de toutes ses copules bidimensionnelles. La seconde contribution est une procédure d'analyse de sensibilité dans laquelle l'incertitude de chaque copule bidimensionnelle est résumée par celle de son coefficient de corrélation de Kendall. Enfin, les deux procédures d'analyse de sensibilité fiabiliste sont appliquées au flambage d'une plaque composite.

**Mots clés :** incertitudes aléatoires et épistémiques • analyse de sensibilité • estimation de probabilités d'évènements rares • méthodes de simulation de Monte-Carlo • statistique inférentielle • théorie des copules.

## Abstract

In many industrial fields, when a highly technical system is under study, reliability assessment requires resorting to a computationally expensive simulation code which is used as a deterministic performance function. From a given set of input physical parameters, the associated output indicates whether failure has occurred or not. Inputs are affected by aleatory uncertainty due to natural variability and their random simultaneous behavior is modeled by a multivariate probability distribution. The failure probability is then defined as the risk of exceeding an extreme threshold of the output distribution and it may be estimated with a wide range of reliability algorithms. The research work conducted in this thesis comes within a broader framework because the input distribution is unknown. It can only be appreciated through a small amount of data and it must be estimated with inferential statistics. In order to secure accurate reliability assessment, it is now generally agreed that the dependence structure among input variables must be dealt with through the copula-marginal separation. In this situation, the estimated failure probability directly derives from the provided dataset and is therefore impacted by sampling variability. In the literature, this second level of uncertainty is referred to as epistemic and it has mainly been studied in the context of parametric models supplemented with Bayesian prior knowledge. The main objective of this thesis is to develop a global sensitivity analysis approach in order to identify which entity of the input distribution conveys the largest share of epistemic uncertainty during reliability analysis. Such an answer may be a valuable piece of information because collecting additional data only for the targeted entity may offer a convenient way of mitigating variability in reliability assessment. The first contribution is a sensitivity analysis procedure designed to apportion epistemic uncertainty between the estimated copula distribution and the estimated marginal distributions. Addressing this problem is not an easy task because uncertainty originates from the split of a correlated dataset. Following the first sensitivity analysis, if the copula is appointed to be the most influential component of the input distribution, the state of knowledge about the theoretical input distribution can barely be improved. Since most multivariate copula models incorporate pair-copulas as efficient building blocks, the idea is to look for the most influential estimated pair-copula. The main difficulty lies in the fact that no systematic decomposition allows to rewrite the copula distribution in terms of all pair-copulas. As a second contribution, a sensitivity analysis procedure is put forward where each pair-copula is summarized by its Kendall's rank correlation coefficient. Finally, the two reliability-oriented sensitivity analysis procedures are applied to the buckling of a composite plate.

**Keywords:** aleatory and epistemic uncertainties • sensitivity analysis • rare-event probability estimation • Monte Carlo simulation methods • inferential statistics • copula theory.