



HAL
open science

Etudes de modèles markoviens à variables latentes avec applications à la pharmacologie et à l'épidémiologie.

Maud Delattre

► **To cite this version:**

Maud Delattre. Etudes de modèles markoviens à variables latentes avec applications à la pharmacologie et à l'épidémiologie.. Statistiques [math.ST]. Université Paris-Saclay, 2020. tel-03202055

HAL Id: tel-03202055

<https://hal.science/tel-03202055>

Submitted on 19 Apr 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.

UNIVERSITÉ PARIS-SACLAY

École doctorale de mathématiques Hadamard (ED 574)

Unité de recherche INRAE, MaIAGE (UR 1404)

Mémoire présenté pour l'obtention du

Diplôme d'habilitation à diriger les recherches

Discipline : Mathématiques

par

Maud DELATTRE

Etudes de modèles markoviens à variables latentes avec applications
à la pharmacologie et à l'épidémiologie.

Rapporteurs :

ANTOINE CHAMBAZ
ANNE GEGOUT-PETIT
STÉPHANE ROBIN

Date de soutenance : 6 novembre 2020

Composition du jury :

CHRISTOPHE AMBROISE (Examineur)
DAVID CAUSEUR (Examineur)
ANTOINE CHAMBAZ (Rapporteur)
ANNE GEGOUT-PETIT (Rapporteuse)
NATHALIE PEYRARD (Examinatrice)
STÉPHANE ROBIN (Rapporteur)

Articles publiés dans des revues de statistique

- [A1] Delattre, M. (2010) *Inference in Mixed Hidden Markov Models and Applications to Medical Studies*. Journal de la Société Française de Statistique 151(1) p. 90-105
- [A2] Delattre, M. and Lavielle, M. (2012) *Maximum Likelihood Estimation in Discrete Mixed Hidden Markov Models using the SAEM algorithm*. Computational Statistics & Data Analysis 56(6) p. 2073-2085
- [A3] Delattre, M., Savic, R.M., Miller, R., Karlsson, M.O. and Lavielle, M. (2012) *Analysis of exposure-response of CI-945 in patients with epilepsy : application of novel mixed hidden Markov modelling methodology*. Journal of Pharmacokinetics and Pharmacodynamics 39(3) p. 263-271
- [A4] Delattre, M., Genon-Catalot, V. and Samson, A. (2013) *Maximum Likelihood Estimation for Stochastic Differential Equations with Random Effects*. Scandinavian Journal of Statistics 40(2) p. 322-343
- [A5] Delattre, M. and Lavielle, M. (2013) *Coupling the SAEM algorithm and the extended Kalman filter for maximum likelihood estimation in mixed-effects diffusion models*. Statistics and Its Interface 6(4) p. 519-532
- [A6] Delattre, M., Lavielle, M. and Poursat, M.A. (2014) *A note on BIC in mixed effects models*. Electronic Journal of Statistics 8(1) p. 456-475
- [A7] Lévy-Leduc, C., Delattre, M., Mary-Huard, T. and Robin, S. (2014) *Two-dimensional segmentation for analyzing HiC data*. Bioinformatics 30(17) p. 386-392
- [A8] Delattre, M., Genon-Catalot, V. and Samson, A. (2015) *Estimation of population parameters in stochastic differential equations with random effects in the diffusion coefficient*. ESAIM : Probability and Statistics 19 p. 671-688
- [A9] Colin, P., Micallef, S., Delattre, M., Mancini, P. and Parent, E. (2015) *Towards using a full spectrum of early clinical trial data : a retrospective analysis to compare potential longitudinal categorical models for molecular targeted therapies in oncology*. Statistics in Medicine 34(22) p. 2999-3016
- [A10] Delattre, M., Genon-Catalot, V. and Samson, A. (2016) *Mixtures of stochastic differential equations with random effects : Application to data clustering*. Journal of Statistical Planning and Inference 173 p. 109-124
- [A11] Colin, P., Delattre, M., Mancini, P. and Micallef, S. (2017) *An Escalation for Bivariate Binary Endpoints Controlling the Risk of Overtoxicity (EBE-CRO) : Managing Efficacy and Toxicity in Early Oncology Clinical Trials*. Journal of Biopharmaceutical Statistics 27(6) p. 1054-1072

- [A12] Brault, V., Delattre, M., Lebarbier, E., Mary-Huard, T. and Lévy-Leduc, C. (2017) *Estimating the number of change-points in a two-dimensional segmentation model without penalization*. Scandinavian Journal of Statistics 44(2) p. 563-580
- [A13] Delattre, M., Genon-Catalot, V. and Larédo, C. (2017) *Parametric inference for discrete observations of diffusion processes with mixed effects*. Stochastic Processes and their Applications 128(6) p. 1929-1957
- [A14] Delattre, M., Genon-Catalot, V. and Larédo, C. (2018) *Approximate maximum likelihood estimation for stochastic differential equations with random effects in the drift and the diffusion*. Metrika 81(8) p. 953-983
- [A15] Delattre, M. and Poursat, M.A. (2020) *An iterative algorithm for joint covariate and random effect selection in mixed effects models*. International Journal of Biostatistics, accepté.

Articles soumis et pré-publications

- [S1] Delattre, M. and Kuhn, E. (2019) *Estimating Fisher Information Matrix in Latent Variable models*. arXiv :1909.06094.
- [S2] Narci, R., Delattre, M., Larédo, C. and Vergu, E. (2020) *Inference for partially observed epidemic dynamics guided by Kalman filtering techniques*. hal-02475936.
- [S3] Delattre, M. (2020) *A review on asymptotic inference in stochastic differential equations with mixed-effects*. hal-02938697.

Rapports techniques

- [RT1] Delattre, M., Lavielle, M. and Poursat, M.A. (2012) *BIC selection procedures in mixed effects models*. RR-7948, INRIA.

Mémoire de thèse

- [TH] Delattre, M. *Inférence statistique dans les modèles mixtes à dynamique Markovienne* (2012)

Articles publiés ou soumis dans des revues de biologie

- [B1] Faure, M.C., Sulpice, J.C., Delattre, M., Lavielle, M., Prigent, M., Cuif, M.H., Melchior, C., Tschirhart, E., Nusse, O. and Dupré-Crochet, S. (2013) *The recruitment of p47phox and Rac2G12V at the phagosome is transient and phosphatidylserine dependent*. Biology of the Cell 105 p. 1-18
- [B2] Tharrey, M., Mariotti, F., Mashchak, A., Barbillon, P., Delattre, M. and Fraser, G. E. (2018) *Patterns of plant and animal protein intake are strongly associated with cardiovascular mortality : the Adventist Health Study-2 cohort*. International journal of epidemiology 47(5) p. 1603-1612
- [B3] Tharrey, M., Mariotti, F., Mashchak, A., Barbillon, P., Delattre, M. and Fraser, G. E. (2020) *Patterns of amino acids intake are strongly associated with cardiovascular mortality, independently of the sources of protein*. International Journal of epidemiology 49(1) p. 312-321
- [B4] Goulnik, J., Plantureux, S., Théry, M., Baude, M., Delattre, M., Van Reeth, C., Villerd, J. and Michelot-Antalik, A. (2020) *Floral trait functional diversity is related to soil characteristics and positively influences pollination function in semi-natural grasslands*. 301, 10733

Introduction	1
1 Modèles mixtes définis en temps discret	5
1.1 Introduction	5
1.1.1 Modèle général et notations	5
1.1.2 Orientations de recherche	6
1.2 Estimation paramétrique	8
1.2.1 Maximum de vraisemblance	8
1.2.2 Algorithmes SAEM et SAEM-MCMC	9
1.2.3 Modèles définis par des chaînes de Markov cachées	11
1.3 Estimation de la matrice d'information de Fisher	14
1.3.1 Introduction	14
1.3.2 Estimateur basé sur le score	15
1.3.3 Algorithme	17
1.3.4 Etude sur données simulées	18
1.4 Caractérisation de la variabilité inter-individuelle par sélection de modèle	20
1.4.1 Utilisation du BIC dans les modèles à effets mixtes	21
1.4.2 Application en pharmacocinétique	25
2 Modèles de diffusion à effets aléatoires	27
2.1 Orientations de recherche	27
2.2 Modèle général et notations	28
2.3 Diffusions observées en temps continu	29
2.3.1 Effets aléatoires de loi gaussienne dans la dérive	29
2.3.2 Effets aléatoires de loi mélange et clustering de courbes	32
2.4 Diffusions observées en des temps discrets	35
2.4.1 Contrastes dérivés du schéma d'Euler	35
2.4.2 Effet aléatoire dans la dérive	37
2.4.3 Effet aléatoire dans le coefficient de diffusion	38
2.4.4 Effets aléatoires dans la dérive et dans le coefficient de diffusion	40
2.4.5 Remarques sur le cas d'effets aléatoires multidimensionnels dans la dérive	41
2.4.6 Etudes par simulation	42

3	Description et inférence de dynamiques épidémiques	47
3.1	Contexte	47
3.2	Formulation d'un modèle linéaire gaussien partiellement observé	48
3.3	Estimation paramétrique	50
3.4	Illustration	50
3.4.1	Modèle SIR	50
3.4.2	Etude sur données simulées	51
4	Projets de recherche	55
4.1	Méthodes statistiques pour l'analyse de modèles à effets mixtes	55
4.1.1	Sélection de modèle	55
4.1.2	Estimation de la matrice d'information de Fisher	56
4.2	Modèles de diffusion à effets aléatoires	57
4.3	Modélisation et estimation de dynamiques épidémiques	57
4.4	Segmentation de trajectoires en écologie du mouvement	58

Mes activités de recherche portent sur les modèles à dynamique markovienne et à variables latentes pour l'analyse de données longitudinales. Ces modèles, complexes par leur structure, suscitent aujourd'hui un fort intérêt dans des domaines d'application aussi variés que la pharmacologie, l'épidémiologie, la biologie végétale, la génomique, l'écologie, etc. Leur attrait tient essentiellement en leur capacité à décrire finement les mécanismes sous-jacents aux données ainsi qu'en leur interprétabilité. Les variables cachées sont utiles pour tenir compte de plusieurs sources de variabilité simultanément dans la modélisation, sans qu'elles soient directement observées. Dans les études que j'ai menées, il y a deux situations. Dans la première, une seule série d'observations est disponible. Les composantes cachées du modèle sont des variables d'état, ou processus stochastiques, dont le rôle est de structurer la dépendance temporelle entre les observations. Dans la seconde, plusieurs séquences sont observées simultanément. Aux variables d'état s'ajoutent alors des effets aléatoires, eux aussi latents, dont le rôle est de prendre en compte la variabilité inter-séquences.

Du point de vue applicatif, l'objectif principal est d'obtenir de l'information sur des paramètres mécanistes d'intérêt à partir des observations, que ce soit à des fins descriptives ou prédictives. Le problème statistique qui traduit cet objectif est celui de l'estimation paramétrique. Il s'agit d'estimer le paramètre θ de la distribution conjointe de la variable observée Y et des variables latentes ϕ à partir d'une observation partielle des variables du modèle donnée par la seule observation de Y . Dans une approche fréquentiste, il est habituel d'estimer les paramètres par la méthode du maximum de vraisemblance. La fonction de vraisemblance s'obtient ici en intégrant la densité jointe $p(Y, \phi; \theta)$ sur les variables latentes, par intégrale continue ou somme discrète selon la nature de ϕ . L'estimation de θ par maximum de vraisemblance est un problème difficile, soit parce que $p(Y, \phi; \theta)$ a une expression complexe ou n'est pas explicite, soit parce que son intégrale sur ϕ n'a pas de forme analytique. Les méthodes d'estimation développées pour répondre à ce problème sont nombreuses. Elles dépendent de la nature des variables latentes et reposent, selon les cas, sur des procédures de calcul exact astucieuses et spécifiques, de type forward-backward par exemple, sur des approximations de la densité $p(Y, \phi; \theta)$ ou sur l'utilisation d'algorithmes numériques, comme l'algorithme EM et ses variantes stochastiques.

Il arrive que la caractérisation des mécanismes biologiques à l'origine des données soit un problème plus ouvert relevant de la modélisation. Un travail préalable est alors nécessaire pour proposer un modèle statistique qui soit cohérent avec la nature et la réalité des données. Il s'agit de formuler des hypothèses réalistes concernant les structures de dépendance entre observations et les mécanismes génératifs latents, mais aussi de proposer une paramétrisation du modèle statistique qui soit biologiquement informative. Dans certains cas, le modèle statistique est déjà partiellement posé, au sens où il existe une collection

de modèles statistiques candidats à la modélisation du phénomène biologique d'intérêt. Le travail relève dans ce cas de la sélection de modèle en statistique.

Mes travaux sur les modèles markoviens et à variables latentes mêlent ces deux aspects : développement méthodologique et modélisation pour l'analyse de données réelles. Mes contributions sont majoritairement inspirées de questions concrètes issues de la pharmacologie et de l'épidémiologie, et je suis ouverte à d'autres domaines applicatifs comme l'écophysiologie végétale. Elles portent sur le développement, l'étude théorique et l'implémentation de méthodes statistiques pour les modèles à effets mixtes, les modèles de Markov cachés et les modèles de diffusion. Je prête également une attention toute particulière à rendre les méthodes développées utilisables par un public non expert.

Dans ce manuscrit, mes travaux ne sont pas présentés dans l'ordre chronologique mais sont regroupés par thème.

Dans le chapitre 1, je présente mes travaux relatifs aux modèles à effets mixtes définis en temps discret. J'y aborde des questions qui trouvent leur origine dans l'analyse de relations dose-concentration-effet en pharmacologie à laquelle je me suis familiarisée pendant ma thèse. Je présente d'abord les travaux connexes à l'étude d'un anti-épileptique qui ont composé une partie de ma thèse de doctorat. Pour cette étude clinique, je propose de combiner une approche dite de population, où les effets aléatoires permettent d'expliquer la variabilité inter-individuelle, à des chaînes de Markov cachées, qui traduisent l'évolution de l'état de santé des patients à travers une succession d'états discrets ([A3]). L'estimation par maximum de vraisemblance, délicate dans ces modèles en raison des effets aléatoires et des chaînes de Markov cachées, n'avait pas trouvé de solution algorithmique satisfaisante au vu du nombre d'effets aléatoires nécessaires dans cette étude. J'ai associé l'analyse des données cliniques au développement d'une version spécifique de l'algorithme SAEM pour l'estimation des paramètres dans les modèles de Markov cachés à effets mixtes ([A2]). J'aborde ensuite la question de l'évaluation de la matrice d'information de Fisher dans des modèles à effets mixtes et autres modèles à variables latentes. Dans ce travail ([S1]), réalisé conjointement avec Estelle Kuhn (INRAE, MaIAGE), nous proposons un nouvel estimateur de la matrice d'information de Fisher, et un algorithme stochastique pour son évaluation lorsque celui-ci n'est pas explicite. L'avantage de notre approche est de n'utiliser que des dérivées au premier ordre des fonctions de vraisemblance et de vraisemblance complète. Nous étudions les propriétés théoriques de l'estimateur et de l'algorithme proposés. Je termine ce premier chapitre par une synthèse de mes travaux relatifs à la sélection de modèle dans les modèles à effets mixtes, développés en collaboration avec Marie-Anne Poursat (Université Paris-Saclay, LMO) ([A6], [A15]). La modélisation de la variabilité inter-individuelle est un aspect crucial dans la construction d'un modèle à effets mixtes, qui ne trouve pourtant de solution rigoureuse que dans des modèles à effets mixtes très spécifiques. L'objectif commun de ces travaux est de fournir des outils génériques pour une très large gamme de modèles à effets mixtes, linéaires, non linéaires ou encore à dynamique markovienne latente. Nous proposons un algorithme pour la sélection conjointe de covariables et d'effets aléatoires construit sur des critères de type BIC dont les pénalités sont spécifiques aux modèles à effets mixtes et aux différents objectifs de sélection.

Le chapitre 2 est consacré au développement méthodologique et théorique d'estimateurs pour les modèles de diffusion à effets mixtes. J'y présente le fruit de collaborations avec Valentine Genon-Catalot (Université de Paris, Paris-Descartes, MAP5) et Adeline Samson (Université Grenoble-Alpes, LJK) d'une part ([A4], [A8], [A10]) et Valentine Genon-Catalot et Catherine Larédo (INRAE, MaIAGE) d'autre part ([A13], [A14]). Ces travaux sont partis du constat de l'absence de résultats asymptotiques généraux pour l'estimateur du maximum de vraisemblance dans ces modèles. L'estimateur du maximum de vraisemblance est en effet difficile à étudier car rarement explicite. Outre la construction de contrastes explicites pour l'estimation paramétrique, la contribution essentielle de ces travaux est donc l'étude des propriétés asymptotiques des estimateurs associés. Ce chapitre s'intéresse

à une classe générale de diffusions réelles à effets mixtes, dont la dérive est linéaire en les paramètres et dont le coefficient de diffusion est multiplicatif, observées sans bruit de mesure. Les effets aléatoires sont supposés de loi gaussienne ou de loi Gamma. Selon le type d'observations, continues ou discrétisées, et selon la position des effets aléatoires dans le modèle, différents contrastes et différents cadres asymptotiques sont considérés. Un détour vers la classification non supervisée de données fonctionnelles est opéré au fil de ce chapitre en étendant certains résultats à des effets aléatoires de loi mélange et en proposant des solutions algorithmiques pour l'estimation et la classification.

Le chapitre 3 décrit un travail très récent sur la description de dynamiques épidémiques, réalisé dans le cadre de la thèse de Romain Narci (INRAE, MaIAGE) que je co-encadre avec Catherine Larédo et Elisabeta Vergu (INRAE, MaIAGE) ([S2]). Nous nous intéressons à l'estimation de paramètres épidémiques qui sont cruciaux pour la compréhension et la gestion des épidémies. Notre procédure d'estimation combine une approximation gaussienne de processus de saut normalisés, valide en grande population, aux calculs récursifs du filtre de Kalman.

Mes perspectives de recherche sont décrites dans le chapitre 4.

Certains de mes travaux, en segmentation par exemple, ne sont pas directement liés aux trois chapitres principaux de ce manuscrit mais sont référencés dans la liste de mes publications. D'autres publications sont référencées sans qu'elles aient donné lieu à des développements nouveaux en statistique. En effet, il m'arrive d'assister des collègues d'autres disciplines dans l'analyse de leurs données expérimentales. Je trouve intéressant et instructif d'être confrontée directement aux questions que se posent les utilisateurs et d'éclairer leurs analyses avec mon point de vue de statisticienne.

1.1 Introduction

1.1.1 Modèle général et notations

Une partie des recherches que j'ai entreprises depuis le début de ma thèse porte sur les modèles à effets mixtes. Ces modèles permettent d'analyser des observations collectées de façon répétée sur plusieurs individus. La variabilité intrinsèque aux données est alors attribuable à différentes sources (intra-individuelle, inter-individuelle, résiduelle) dont la prise en compte est essentielle pour caractériser sans biais les mécanismes biologiques à l'origine des observations. Ce type de données est fréquent dans de nombreux domaines d'applications, en particulier dans les études de pharmacologie qui nécessitent d'observer plusieurs patients pour étudier les relations dose-concentration-effet des médicaments. Notons N le nombre d'individus, n_i le nombre d'observations pour l'individu i , et $y_i = (y_{i1}, \dots, y_{i,n_i})^\top$ le vecteur des données y_{ij} mesurées aux temps t_{ij} pour l'individu i , $j = 1, \dots, n_i$, $i = 1, \dots, N$. Les modèles mixtes peuvent être définis comme des modèles hiérarchiques à deux niveaux (voir par exemple [Lavielle M., 2014]).

1. Le premier niveau définit le *modèle individuel*. L'évolution des observations y_i au cours du temps est décrite au moyen d'un même modèle statistique pour tous les individus :

$$Y_i \sim p(\cdot | \psi_i, x_i) \tag{1.1}$$

Ce modèle, parfois qualifié de modèle structurel, est spécifié par sa distribution $p(\cdot | \psi_i, x_i)$. Cette distribution dépend d'un paramètre $\psi_i \in \mathbb{R}^d$ propre à l'individu, et de variables x_i qui sont structurelles, intrinsèques au modèle dynamique ou à l'échantillonnage, tels que les temps de mesure t_{ij} . Pour alléger les notations, j'omettrai dans la suite la dépendance aux variables structurelles et je noterai $p(\cdot | \psi_i)$ la distribution conditionnelle de Y_i sachant ψ_i .

2. Le deuxième niveau spécifie le *modèle de population*. Les différences inter-individuelles sont décrites en définissant les paramètres individuels ψ_i comme des variables aléatoires indépendantes. En général, les paramètres individuels sont supposés gaussiens à une transformation près, tels que

$$g(\psi_i) = \mu + \beta C_i + \eta_i, \quad \eta_i \underset{i.i.d.}{\sim} \mathcal{N}(0, \Omega), \tag{1.2}$$

où $g(\psi) = (g_1(\psi_1), \dots, g_d(\psi_d))^\top : \mathbb{R}^d \rightarrow \mathbb{R}^d$ est une fonction de lien, $C_i \in \mathbb{R}^q$ est le vecteur de covariables pour l'individu i , $\mu \in \mathbb{R}^d$ est l'intercept, β est la matrice $d \times q$

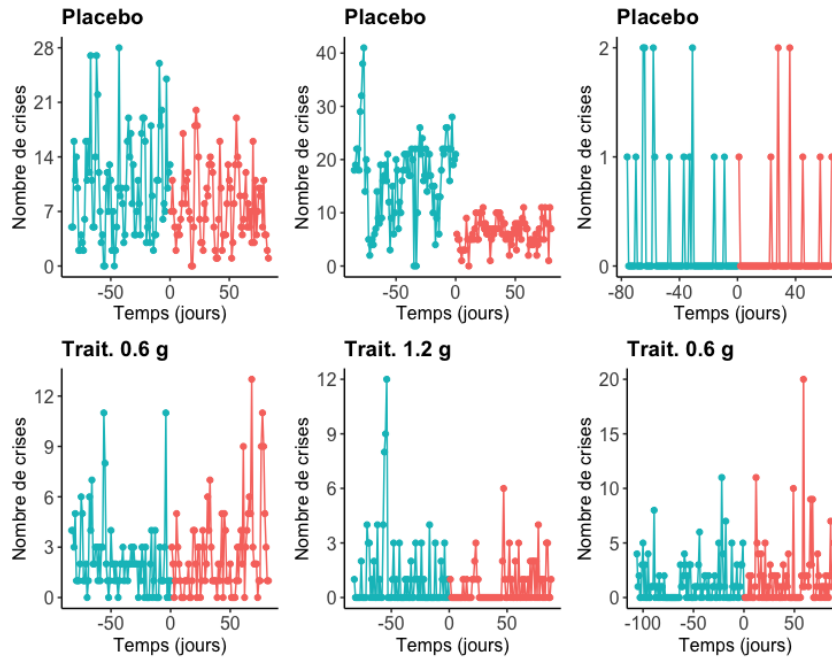


FIGURE 1.1 – Nombres de crises d’épilepsie mesurés au cours du temps dans le cadre d’un essai clinique. Chaque graphique correspond à un patient. La période d’observation sans traitement est représentée en cyan, la période de traitement est représentée en rouge. Certains patients reçoivent un placebo, d’autres reçoivent la forme active du médicament étudié. Plusieurs doses sont expérimentées.

des effets spécifiques des covariables et $\Omega = (\omega_{kl})_{1 \leq k, l \leq d}$ est une matrice de variance-covariance. Dans ce modèle, le terme βC_i décrit les différences inter-individuelles dues à des caractéristiques individuelles observées, tandis que les effets aléatoires η_i , représentent la variabilité entre les individus non attribuable aux covariables mesurées. Les paramètres μ , β et Ω caractérisent la population : les effets fixes μ et β , communs à tous les individus, définissent un profil type, tandis que la matrice Ω mesure les fluctuations aléatoires autour de ce profil type. Les paramètres $\theta = (\mu, \beta, \Omega)$ sont généralement appelés paramètres de population.

1.1.2 Orientations de recherche

Mes travaux de thèse et certains problèmes que j’étudie actuellement ont trouvé leur origine dans l’analyse de données pour la pharmacocinétique et la pharmacodynamie. Les deux exemples qui suivent offrent un panorama des questionnements biologiques qui alimentent mes développements sur les modèles à effets mixtes.

Exemple 1. Description de symptômes au cours du temps.

La Figure 1.1 représente les profils de patients d’une étude clinique visant à évaluer l’efficacité d’un nouvel anti-épileptique. Les données recueillies dénombrent les crises d’épilepsie quotidiennes. L’essai commence par une période de screening pendant laquelle les patients sont suivis sous leur traitement habituel, puis se poursuit par une période de traitement où chaque patient reçoit le nouvel anti-épileptique sous forme active ou placebo.

Ce type d’étude est destiné à prouver la pertinence clinique du médicament proposé, puis à guider le choix de sa posologie sur de nouveaux patients. Cela suppose d’utiliser un modèle statistique qui permette de *i)* décrire précisément le mécanisme de succession des crises au cours du temps, *ii)* décrire l’effet du médicament sur l’évolution des symptômes,

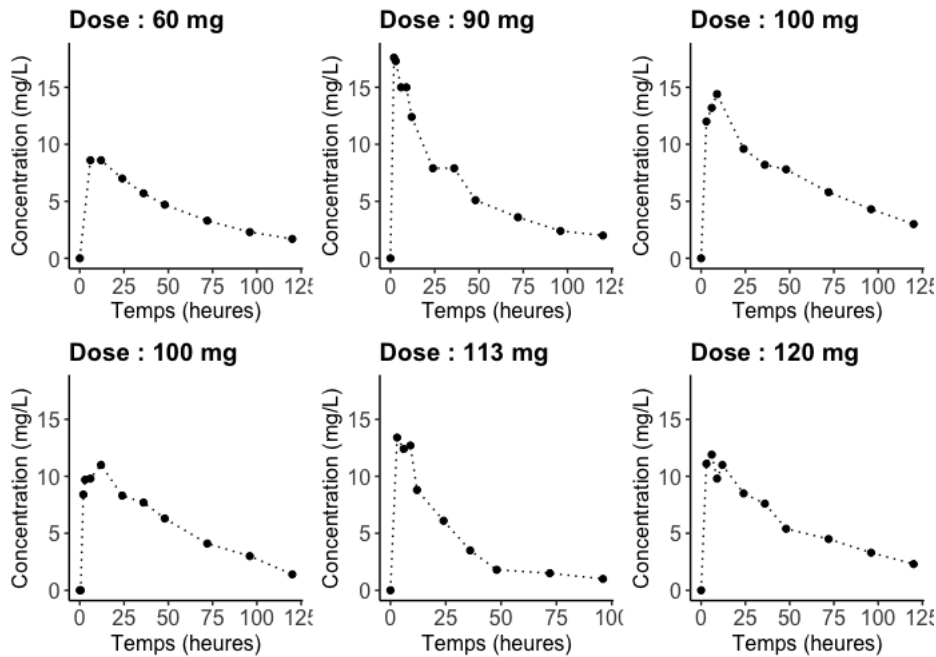


FIGURE 1.2 – Concentrations d’un anticoagulant, la warfarine, dans le plasma sanguin au cours du temps après administration orale d’une dose unique dont la valeur est indiquée au dessus de chaque courbe. Chaque graphique correspond à un patient. Ces données de pharmacocinétiques sont en accès libre dans le package R `nlmixr`.

iii) comprendre les différences de réponse entre patients.

Exemple 2. Description du devenir de médicaments dans l’organisme.

La Figure 1.2 représente des mesures de concentration d’un médicament dans le plasma sanguin obtenues sur plusieurs patients. Ces données sont typiques d’études de pharmacocinétique conduites par les laboratoires pharmaceutiques. Ces études visent à définir les modalités optimales d’administration de médicaments en fonction des profils des patients. Elles nécessitent d’évaluer des paramètres pharmacocinétiques clés, par exemple les vitesses d’absorption et d’élimination du médicament, ainsi que l’influence potentielle de caractéristiques individuelles (âge, sexe, taille, poids, ...) sur ces paramètres.

Comme le suggèrent ces deux exemples, l’étude de données réelles soulève des questions de différents ordres allant de la modélisation à la sélection de modèle en passant par l’estimation. Si l’aspect majoritaire de mon travail est le développement de méthodes statistiques pour l’inférence au sens large de modèles à effets mixtes, c’est souvent en essayant de proposer des modèles adaptés à la structure de données réelles, ou à une problématique biologique précise, que ces questions méthodologiques ont émergé. L’essentiel des méthodes existantes pour les modèles à effets mixtes utilisent le formalisme de Pinheiro J. C. et Bates D. M. [2000] où les observations sont modélisées sous la forme :

$$Y_{ij} = f_{ij}(\psi_i) + g_{ij}(\psi_i)\varepsilon_{ij}, \varepsilon_{ij} \underset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2), i = 1, \dots, N, j = 1, \dots, n_i,$$

et ψ_i est défini selon (1.2). Dans cette équation, $f_{ij}(\cdot)$ et $g_{ij}(\cdot)$ sont des fonctions connues, potentiellement non linéaires en ψ_i , et les variables aléatoires ε_{ij} sont des résidus indépendants des paramètres individuels ψ_i . Ce formalisme très répandu suppose l’indépendance

des observations de chaque individu conditionnellement à ψ_i :

$$p(\cdot|\psi_i) = \prod_{j=1}^{n_i} p_{ij}(\cdot|\psi_i), \quad p_{ij}(\cdot|\psi_i) = \mathcal{N}(f_{ij}(\psi_i), g_{ij}(\psi_i)^2 \sigma^2).$$

Cette hypothèse n'est pas toujours cohérente. J'ai démarré ma thèse par l'analyse de données de crises d'épilepsie (cf *Exemple 1*) pour lesquelles les données consécutives d'un même sujet sont typiquement liées les unes aux autres par une structure de dépendance temporelle. La section 1.2 décrit les travaux que j'ai menés en lien avec cette étude et les modèles de chaînes de Markov cachées à effets mixtes ([A2], [A3]). Après un état de l'art succinct des techniques d'estimation existantes dans les modèles à effets mixtes, puis une description générale des algorithmes SAEM [Delyon B. et al., 1999] et SAEM-MCMC [Kuhn E. et Lavielle M., 2004], je présente une version spécifique de l'algorithme SAEM-MCMC pour les modèles de Markov cachés à effets mixtes ([A2]). L'originalité et l'intérêt de cet algorithme sont de ne nécessiter la simulation que d'une partie des variables cachées du modèle, les effets aléatoires, menant à un coût computationnel limité par rapport à ce que ferait une version plus classique de SAEM-MCMC pour ces modèles. Je décris ensuite les résultats de l'analyse des données cliniques de crises d'épilepsie obtenus avec cette méthode ([A3]). La section 1.3 résume un travail récent, conjoint avec Estelle Kuhn, sur l'estimation de la matrice d'information de Fisher dans les modèles à variables latentes [S1]. Nous y proposons un nouvel estimateur de la matrice d'information de Fisher, ne nécessitant pas de dériver deux fois la log-vraisemblance contrairement à l'estimateur usuellement utilisé en pratique. Nous proposons ensuite une procédure d'évaluation de cet estimateur imbriquée dans un algorithme SAEM lorsque son calcul exact est impossible et nous en étudions les propriétés théoriques de convergence. Enfin, j'aborde dans la section 1.4 des questions de sélection de modèle qui se posent naturellement dans la modélisation de la variabilité inter-individuelle dans une classe large de modèles à effets mixtes, à dynamique markovienne ou non. Le modèle structurel étant choisi, je me suis attachée à fournir aux modélisateurs des outils qui permettent d'identifier les covariables influentes et une structure d'effets aléatoires pertinente au vu des observations. Je propose plus précisément des critères de type BIC et un algorithme pour la sélection conjointe de covariables et d'effets aléatoires ([A6],[A15]) dont l'avantage est d'être facilement compréhensibles et utilisables par un public non statisticien.

1.2 Estimation paramétrique

Dans cette partie, l'objectif est d'estimer les paramètres de population θ à partir d'observations $y = (y_i, i = 1, \dots, N)$ où $y_i = (y_{i1}, \dots, y_{i,n_i})^\top$ dans le modèle (1.1)-(1.2). Ma contribution principale concerne la modification d'algorithmes existants pour la prise en compte de dynamiques markoviennes latentes (section 1.2.3).

1.2.1 Maximum de vraisemblance

L'estimateur du maximum de vraisemblance (EMV) est un estimateur naturel de θ :

$$\hat{\theta}_N = \arg \max_{\theta \in \Theta} L_N(\theta), \quad (1.3)$$

où l'indépendance entre les individus permet d'écrire la vraisemblance des observations, $L_N(\theta)$, comme le produit de N contributions individuelles :

$$L_N(\theta) = \prod_{i=1}^N p(y_i; \theta) = \prod_{i=1}^N \int p(y_i, \psi_i; \theta) d\psi_i. \quad (1.4)$$

Les modèles à effets mixtes sont des modèles à données incomplètes, les paramètres individuels ψ_i n'étant pas observés. Chaque vraisemblance individuelle s'obtient alors en intégrant la vraisemblance des données individuelles complètes $p(y_i, \psi_i; \theta)$ par rapport à ψ_i . Cette intégrale n'a en général pas d'expression analytique.

Pour réaliser l'estimation malgré l'absence d'expression explicite pour la vraisemblance, il est possible de maximiser une vraisemblance approchée en θ plutôt que la vraisemblance exacte. Les techniques les plus courantes reposent sur la linéarisation de $p(y_i|\psi_i, x_i)$, donnant lieu à une vraisemblance approchée dont le calcul est explicite [Sheiner L. B. et al., 1972, Beal S. L. et Sheiner L. B., 1982, Lindstrom M. J. et Bates D. M., 1990], ou sur l'évaluation numérique de l'intégrale $\int p(y_i, \psi_i; \theta)d\psi_i$ par des méthodes de quadrature [Davidian M. et Gallant A. R., 1993]. Les conditions de validité et/ou le coût algorithmique de ces méthodes en limitent néanmoins l'utilisation. L'algorithme EM [Dempster A. P. et al., 1977] et ses variantes, construits pour les modèles à variables latentes, sont des alternatives à ces méthodes d'approximation de la vraisemblance. Le principe de l'algorithme EM est de maximiser itérativement en θ l'espérance conditionnelle de la log-vraisemblance complète pour augmenter la vraisemblance des observations et ainsi approcher numériquement l'EMV. Dans beaucoup de modèles non linéaires à effets mixtes, le calcul de cette espérance conditionnelle n'est pas analytique. Plusieurs variantes stochastiques de l'algorithme EM, telles que l'algorithme MCEM [Wei G. C. G. et Tanner M. A., 1990] ou l'algorithme SAEM [Delyon B. et al., 1999], ont été développées pour contourner cette difficulté. Elles reposent sur l'évaluation de cette espérance conditionnelle par simulation des données cachées sous leur loi conditionnelle aux observations.

1.2.2 Algorithmes SAEM et SAEM-MCMC

1.2.2.1 Algorithme SAEM

L'algorithme SAEM est une variante stochastique de l'algorithme EM associant la simulation des données cachées à un schéma d'approximation stochastique de type Robbins-Monro [Delyon B. et al., 1999].

Considérons un modèle à variables latentes général, de variable observée Y et de variable latente ϕ , dont la densité jointe $p(y, \phi; \theta)$ dépend d'un paramètre θ . Après initialisation, et pour la valeur courante $\hat{\theta}_m$, l'itération $m \geq 1$ de l'algorithme SAEM consiste en trois étapes :

1. *Etape S* : simulation de $\phi^{(m)}$ selon la loi conditionnelle $p(\phi|y, \hat{\theta}_m)$,
2. *Etape SA* : approximation stochastique de l'espérance conditionnelle $\mathbb{E}[\log p(y, \phi; \theta)|y, \hat{\theta}_m]$ par :

$$Q_m(\theta) = Q_{m-1}(\theta) + \gamma_m \left(\log p(y, \phi^{(m)}; \theta) - Q_{m-1}(\theta) \right),$$

3. *Etape M* : actualisation de l'estimation du paramètre

$$\hat{\theta}_{m+1} = \arg \max_{\theta \in \Theta} Q_m(\theta),$$

où $(\gamma_m)_{m \geq 1}$ est une suite de pas décroissante, telle que $\gamma_1 = 1$ et $\lim_{m \rightarrow +\infty} \gamma_m = 0$.

Contrairement à d'autres variantes stochastiques de l'algorithme EM, l'algorithme SAEM ne nécessite de simuler qu'une seule réalisation des données cachées à chaque itération, conduisant à des temps d'exécution bien plus attractifs en pratique. Du point de vue théorique, un résultat de [Delyon B. et al., 1999] garantit la convergence de l'algorithme vers un maximum local de la log-vraisemblance des observations, lorsque $\sum_{m=1}^{+\infty} \gamma_m = +\infty$ et $\sum_{m=1}^{+\infty} \gamma_m^2 < +\infty$, et lorsque la vraisemblance complète appartient à la famille des modèles exponentiels tel que formulé par l'hypothèse :

$$(\mathbf{M1}) \quad p(y, \phi; \theta) = \exp(-\kappa(\theta) + \langle S(y, \phi), \nu(\theta) \rangle),$$

où $\kappa(\cdot)$ et $\nu(\cdot)$ sont des fonctions de θ , $S(y, \phi)$ sont les statistiques exhaustives du modèle et $\langle \cdot, \cdot \rangle$ désigne le produit scalaire. Dans les modèles vérifiant **(M1)**, la mise en œuvre de l'algorithme SAEM utilise une version simplifiée des étapes SA et M portant sur les statistiques exhaustives du modèle :

(b) *Etape SA*

$$s_m = s_{m-1} + \gamma_m \left(S(y, \phi^{(m)}) - s_{m-1} \right),$$

(c) *Etape M*

$$\hat{\theta}_{m+1} = \arg \max_{\theta \in \Theta} [-\kappa(\theta) + \langle s_m, \nu(\theta) \rangle].$$

1.2.2.2 Combinaison à des méthodes MCMC

Dans beaucoup de modèles, la simulation exacte des données cachées sous leur loi conditionnelle aux observations n'est pas possible. Kuhn E. et Lavielle M. [2004, 2005] proposent dans ce cas de réaliser l'étape S de l'algorithme SAEM au moyen de méthodes MCMC. Ces méthodes reposent sur l'utilisation de lois instrumentales $q(\cdot, \cdot)$ plus simples à échantillonner que la loi cible. L'une des plus populaires est l'algorithme de Metropolis-Hastings [Robert C. et Casella G., 2004] utilisé lorsque la distribution cible est connue à une constante près. L'itération k de cet algorithme consiste à

1. générer un candidat $\phi^{(c)}$ selon une loi de proposition $q(\phi_{k-1}, \cdot)$ pouvant dépendre de la valeur ϕ_{k-1} générée à l'itération précédente,
2. générer $U \sim \mathcal{U}([0, 1])$,
3. définir

$$\phi_k = \begin{cases} \phi^{(c)} & \text{si } U < \gamma(\phi_{k-1}, \phi^{(c)}), \\ \phi_{k-1} & \text{sinon,} \end{cases}$$

où la probabilité d'acceptation est définie selon :

$$\begin{aligned} \gamma(\phi_{k-1}, \phi^{(c)}) &= \min \left(1, \frac{q(\phi^{(c)}, \phi_{k-1}) p(\phi^{(c)} | y; \theta)}{q(\phi_{k-1}, \phi^{(c)}) p(\phi_{k-1} | y; \theta)} \right), \\ &= \min \left(1, \frac{p(y, \phi^{(c)}; \theta) q(\phi^{(c)}, \phi_{k-1})}{p(y, \phi_{k-1}; \theta) q(\phi_{k-1}, \phi^{(c)})} \right). \end{aligned} \quad (1.5)$$

La convergence de l'algorithme SAEM-MCMC est démontrée sous des conditions similaires à celles de [Delyon B. et al., 1999] et sous des hypothèses supplémentaires garantissant l'ergodicité de la chaîne de Markov générée par l'algorithme MCMC [Kuhn E. et Lavielle M., 2004, Allasonnière S. et Kuhn E., 2010]. En pratique, il arrive que la loi instrumentale $q(\cdot, \cdot)$ choisie ait une incidence importante sur la capacité de l'algorithme à explorer l'espace des paramètres. Lavielle M. [2014] recommande par exemple d'utiliser successivement plusieurs lois instrumentales afin d'affiner l'exploration de l'espace des paramètres au fil des itérations de l'algorithme SAEM-MCMC.

Dans les modèles mixtes, la mise en place de l'algorithme SAEM nécessite d'échantillonner selon les lois

$$p(\psi_i | y_i; \theta) = \frac{p(y_i, \psi_i; \theta)}{p(y_i; \theta)}, \quad i = 1, \dots, N,$$

qui ne sont pas explicites lorsque le modèle n'est pas linéaire gaussien. L'implémentation d'un algorithme de Metropolis-Hastings à l'étape S dépend du modèle considéré, et en particulier de la capacité à calculer les densités complètes $p(y_i, \psi_i; \theta)$. Lorsque $p(y_i | \psi_i; \theta)$

dépend de variables latentes, un échantillonneur de Gibbs ou un algorithme Metropolis-within-Gibbs sont généralement utilisés à la place de l'algorithme de Metropolis-Hastings (voir par exemple [Donnet S. et Samson A., 2008]).

1.2.3 Modèles définis par des chaînes de Markov cachées

Les modèles à états discrets sont utiles dans les applications médicales qui nécessitent de tenir compte des stades évolutifs d'une pathologie [Anisimov V. V. et al., 2007, Albert P. S., 1991]. Ils permettent de décrire l'évolution du patient i à partir des états $(Z_{ij})_{j=1,\dots,n_i} \in \{1, \dots, S\}^{n_i}$ successivement occupés aux différents temps d'observation. Ces modèles se déclinent sous différentes formes correspondant chacune à des hypothèses plus ou moins complexes sur les lois conditionnelles $p(Z_{ij}|Z_{i1}, \dots, Z_{i,j-1}, \psi_i)$ qui précisent la manière dont les états sont visités au cours du temps. En supposant que seul l'état courant est utile pour prédire l'état futur, *i.e.* $p(Z_{ij}|Z_{i1}, \dots, Z_{i,j-1}, \psi_i) = p(Z_{ij}|Z_{i,j-1}, \psi_i)$, les chaînes de Markov d'ordre 1 formalisent la plus simple des dépendances temporelles, mais aussi la plus populaire. Dans beaucoup d'applications, la chaîne de Markov est cachée : les observations disponibles $y_i = (y_{i1}, \dots, y_{i,n_i})^\top$ ne sont pas une mesure directe des états mais des quantités qui y sont liées. Conditionnellement aux paramètres individuels ψ_i (cf équation (1.2)), la dynamique du sujet i est décrite par

1. la loi initiale de la chaîne de Markov Z_i , $\pi_i = (\pi_i(1), \dots, \pi_i(S))$ où $\pi_i(s) = \mathbb{P}(Z_{i1} = s|\psi_i)$, $s = 1, \dots, S$,
2. l'ensemble des probabilités de transition $p_{ij}(k, \ell) = \mathbb{P}(Z_{i,j+1} = k|Z_{i,j} = \ell, \psi_i)$ telles que $\sum_{\ell=1}^S p_{ij}(k, \ell) = 1$ pour tous i, j, k ,
3. les distributions d'émission des observations dans chaque état $q_{ij}(y_{ij}|Z_{ij} = s)$ pour $s = 1, \dots, S$ et $j = 1, \dots, n_i$.

1.2.3.1 Adaptation de l'algorithme SAEM-MCMC

Dans [A2], nous proposons une version spécifique de l'algorithme SAEM pour les modèles de Markov cachés (HMM, *Hidden Markov Model*) à effets mixtes. Les algorithmes concurrents sont peu nombreux. On peut essentiellement citer [Altman R. M., 2007] qui propose des méthodes de quadrature et Monte-Carlo. Ces algorithmes sont coûteux et leur utilisation est limitée en pratique par le nombre d'effets aléatoires du modèle. L'originalité de notre approche par rapport aux versions standards de l'algorithme SAEM est de considérer les états des chaînes de Markov cachées Z_i comme des paramètres de nuisance. Ceci évite de simuler des candidats pour les états cachés à chaque itération de l'algorithme. Pour simuler les paramètres aléatoires ψ_i , nous utilisons un algorithme de Metropolis-Hastings. La difficulté repose dans le calcul des probabilités d'acceptation qui nécessite d'évaluer les vraisemblances complètes $p(y_i, \psi_i; \theta)$ (cf équation (1.5)). Rappelons que

$$p(y_i, \psi_i; \theta) = p(y_i|\psi_i)p(\psi_i; \theta).$$

Le calcul de $p(\psi_i; \theta)$ ne pose aucune difficulté sous l'hypothèse gaussienne (1.2). Quant à $p(y_i|\psi_i)$, son expression est explicite :

$$p(y_i|\psi_i) = \sum_{z_{i1}=1}^S \dots \sum_{z_{i,n_i}=1}^S \left(\pi_i(z_{i1}) \prod_{j=2}^{n_i} p_{ij}(z_{ij}, z_{i,j-1}) \prod_{j=1}^{n_i} q_{ij}(y_{ij}|z_{ij}) \right), \quad (1.6)$$

mais son calcul direct à partir de (1.6) est prohibitif (complexité $\mathcal{O}(n_i S^{n_i})$). Dans les HMM classiques, on lui préfère l'utilisation d'algorithmes de type *forward* ou *backward* [Rabiner L. R., 1989], récursifs et moins coûteux (complexité $\mathcal{O}(n_i S^2)$).

Nous proposons de coupler l'algorithme SAEM-MCMC à des méthodes *forward*. A chaque itération, les quantités $p(y_i|\psi_i)$, nécessaires à l'implémentation des étapes S et SA , sont calculées selon

$$p(y_i|\psi_i) = \sum_{s=1}^S \alpha_{i,n_i}(s),$$

où les densités $\alpha_{ij}(s) = p(y_{i1}, \dots, y_{ij}, z_{ij} = s|\psi_i)$ sont calculées récursivement grâce à la relation

$$\alpha_{ij}(s) = p(y_{ij}|z_{ij} = s, \psi_i) \sum_{s'=1}^S p(z_{ij} = s|z_{i,j-1} = s', \psi_i) \alpha_{i,j-1}(s').$$

Cette opération de marginalisation permet par ailleurs de garantir la convergence de l'algorithme SAEM-MCMC sous les mêmes hypothèses que [Kuhn E. et Lavielle M., 2004].

Dans [A2], une étude de simulation illustre les bonnes performances de l'algorithme SAEM-MCMC proposé pour les modèles de Markov cachés à effets mixtes, ainsi que son intérêt en termes de temps de calcul. L'analyse de l'essai clinique portant sur l'efficacité d'un nouvel anti-épileptique menée dans [A3], détaillée ci-dessous, confirme également la capacité de l'algorithme à estimer les paramètres de modèles complexes incluant un grand nombre d'effets aléatoires.

1.2.3.2 Application à l'étude de l'efficacité d'un anti-épileptique

En collaboration avec Marc Lavielle, Radojka Savic (UMR 738, INSERM - Université de Paris, anciennement Paris-Diderot), Raymond Miller (R. & D., Pfizer) et Mats Karlsson (Université d'Uppsala, Suède), nous avons analysé dans [A3] les données d'une étude multicentrique, menée par le laboratoire Pfizer, et visant à évaluer les apports d'un anti-épileptique par rapport aux traitements standards. Cet essai, d'une durée de 24 semaines, inclut 788 patients épileptiques. Pendant les 12 premières semaines, ils reçoivent leur traitement habituel. Ils sont ensuite répartis au hasard dans des groupes de traitement parallèles, chaque groupe correspondant à une dose différente du nouvel anti-épileptique ou à sa forme placebo. Le nouveau traitement est suivi durant 12 semaines. Pendant toute la durée de l'étude, chaque patient reporte quotidiennement le nombre de crises d'épilepsie, toutes catégories de crises confondues. Un extrait des données est représenté sur la Figure 1.1. Les données de la période de screening ont été considérées dans [Trokoniz I. F. et al., 2009]. Les auteurs incorporent une composante markovienne à un modèle binomial négatif en autorisant les paramètres des distributions individuelles à varier selon le nombre de crises d'épilepsie reportés la veille.

Modélisation. Nous proposons un modèle de Markov caché à effets mixtes. Cette modélisation est cliniquement intéressante, car elle permet de relier l'évolution des crises d'épilepsie à une succession d'états qui pourraient s'apparenter à différents stades évolutifs de la maladie. Elle permet en outre de caractériser l'effet de l'anti-épileptique aussi bien sur l'intensité des symptômes que sur la dynamique de la maladie. Comme dans [Albert P. S., 1991], nous considérons $S = 2$ états. Ce choix est cliniquement cohérent, les patients épileptiques ayant naturellement tendance à alterner des périodes de faible activité épileptique (état 1) et de forte activité épileptique (état 2), en raison d'écart potentiels à leurs traitements ou encore d'un état de forme variable. Nous considérons des émissions Poisson, de paramètres λ_{1ij} et λ_{2ij} pour le patient $i = 1, \dots, N$ à l'instant $j = 1, \dots, n_i$ dans les états 1 et 2 respectivement. La probabilité de passer de l'état s à l'état s' est notée $p_{ij}(s, s')$ pour le patient i au jour j de l'étude. Dans la suite, l'exposant B se réfère aux paramètres du HMM de référence, *i.e.* décrivant la dynamique des crises d'épilepsie en phase de screening, et l'exposant Tr aux paramètres du HMM après administration du

nouvel anti-épileptique. t_i^d est le moment auquel le sujet i reçoit pour la première fois le nouvel anti-épileptique.

1. Le *modèle individuel* est défini comme suit :

$$\text{logit}(p_{ij}(1, 1)) = \mathbb{1}_{\{t_{ij} < t_i^d\}} \text{logit}(p_i^B(1, 1)) + \mathbb{1}_{\{t_{ij} \geq t_i^d\}} \text{logit}(p_i^{Tr}(1, 1)), \quad (1.7a)$$

$$\text{logit}(p_{ij}(2, 1)) = \mathbb{1}_{\{t_{ij} < t_i^d\}} \text{logit}(p_i^B(2, 1)) + \mathbb{1}_{\{t_{ij} \geq t_i^d\}} \text{logit}(p_i^{Tr}(2, 1)), \quad (1.7b)$$

$$\log(\lambda_{1ij}) = \mathbb{1}_{\{t_{ij} < t_i^d\}} \log(\lambda_{1i}^B) + \mathbb{1}_{\{t_{ij} \geq t_i^d\}} \log(\lambda_{1ij}^{Tr}), \quad (1.7c)$$

$$\lambda_{2ij} - \lambda_{1ij} = \mathbb{1}_{\{t_{ij} < t_i^d\}} \alpha_i^B + \mathbb{1}_{\{t_{ij} \geq t_i^d\}} \alpha_{ij}^{Tr}, \quad (1.7d)$$

où :

$$\text{logit}(p_i^B(1, 1)) = \psi_{i1}, \quad \text{logit}(p_i^{Tr}(1, 1)) = \text{logit}(p_i^B(1, 1)) + \psi_{i5}, \quad (1.8a)$$

$$\text{logit}(p_i^B(2, 1)) = \psi_{i2}, \quad \text{logit}(p_i^{Tr}(2, 1)) = \text{logit}(p_i^B(2, 1)) + \psi_{i6}, \quad (1.8b)$$

$$\log(\lambda_{1i}^B) = \psi_{i3}, \quad \log(\lambda_{1ij}^{Tr}) = \log(\lambda_{1i}^B) + \psi_{i7}(1 - 2^{-(t_{ij}-t_i^d)/\psi_{i9}}), \quad (1.8c)$$

$$\log(\alpha_i^B) = \psi_{i4}, \quad \log(\alpha_{ij}^{Tr}) = \log(\alpha_i^B) + \psi_{i8}(1 - 2^{-(t_{ij}-t_i^d)/\psi_{i10}}). \quad (1.8d)$$

Précisons que la chaîne de Markov cachée évoluant dans un espace à deux états, les probabilités de transition $p_{ij}(1, 2)$ et $p_{ij}(2, 2)$ se déduisent de (1.7) grâce aux relations $p_{ij}(1, 2) = 1 - p_{ij}(1, 1)$ et $p_{ij}(2, 2) = 1 - p_{ij}(2, 1)$.

2. Le *modèle de population* est ensuite donné par :

$$\begin{aligned} \psi_{i1} &= \mu_1 + \eta_{i1} & , & & \psi_{i6} &= \delta_2 + \eta_{i6} & , \\ \psi_{i2} &= \mu_2 + \eta_{i2} & , & & \psi_{i7} &= \delta_3 + \gamma_3 D_i + \eta_{i7} & , \\ \psi_{i3} &= \log(\lambda_1) + \eta_{i3} & , & & \psi_{i8} &= \delta_4 + \gamma_4 D_i + \eta_{i8} & , \\ \psi_{i4} &= \log(\alpha) + \eta_{i4} & , & & \psi_{i9} &= \tau_1 & , \\ \psi_{i5} &= \delta_1 + \eta_{i5} & , & & \psi_{i10} &= \tau_2 & , \\ (\eta_{1i}, \dots, \eta_{8i})^\top &\underset{i.i.d.}{\sim} \mathcal{N}(0, \Omega) & , & & \Omega &= \text{diag}(\omega_1^2, \dots, \omega_8^2), \end{aligned} \quad (1.9)$$

où D_i est la dose reçue par le patient i .

Ce modèle suppose que la dynamique des crises est homogène en période de screening et devient hétérogène en période de traitement pour tenir compte de l'effet progressif de l'anti-épileptique étudié sur le mécanisme d'apparition des crises d'épilepsie. Le traitement affecte les transitions entre états de la même manière que le placebo quelle que soit la dose, via les paramètres δ_1 et δ_2 . Le traitement modifie les nombres moyens de crises d'épilepsie dans chaque état, par la combinaison d'un effet placebo mesuré par les paramètres δ_3 et δ_4 et d'un effet de la dose mesuré par les paramètres γ_3 et γ_4 . Cet effet dose est progressif, τ_1 et τ_2 mesurent la durée nécessaire à l'obtention de la moitié de l'effet maximum dans les états de faible et de forte activité épileptique respectivement. Nous montrons dans [A3] sur des simulations que ce modèle à deux états reproduit de façon satisfaisante les dynamiques de crises d'épilepsie observées durant l'essai clinique.

Résultats. L'algorithme SAEM-MCMC combiné à l'algorithme *forward* est utilisé pour en estimer les paramètres. Les résultats sont réunis dans la Table 1.1. On déduit par Monte-Carlo les probabilités de transition moyennes ainsi que les nombres moyens de crises d'épilepsie dans chaque état, avec le traitement standard ou le nouvel anti-épileptique. Les résultats montrent en particulier que le nouvel anti-épileptique n'a qu'un très faible impact sur la dynamique des états mais occasionne une diminution du nombre moyen de crises dans les états de faible et de forte activité épileptique. Par rapport au traitement standard, le nouvel anti-épileptique pris à une dose 1.8 g, la plus haute dose utilisée dans cet essai, entraîne une diminution du nombre de crises de 15% en moyenne dans l'état de faible activité épileptique et de 27% dans l'état de forte activité épileptique. Les estimations des paramètres τ_1 et τ_2 montrent que la moitié de l'effet maximal est atteint en 2 jours dans l'état de faible intensité et 16 jours dans l'état de forte intensité.

Paramètre	Estimation	RSE (%)	Paramètre	Estimation	RSE (%)
μ_1	1.96	3	ω_1^2	0.72	10
μ_2	1.56	5	ω_2^2	0.78	11
$\log(\lambda_1)$	-2.01	3	ω_3^2	1.95	7
$\log(\alpha)$	-0.08	65	ω_4^2	1.5	7
δ_1	0.17	20	ω_5^2	0.09	29
δ_2	0.27	16	ω_6^2	0.07	26
δ_3	-0.05	78	ω_7^2	0.41	9
δ_4	0.07	57	ω_8^2	0.37	12
γ_3	-1.6×10^{-4}	18			
γ_4	-2.0×10^{-4}	11			
τ_1 (jour ⁻¹)	1.70	10			
τ_2 (jour ⁻¹)	15.65	13			

TABLE 1.1 – Paramètres estimés pour le modèle (1.7)-(1.8)-(1.9) sur les données cliniques de fréquences de crises d'épilepsie. Les valeurs des colonnes 3 et 6 correspondent aux écart-types relatifs (RSE), donnés par estimation de la matrice d'information de Fisher.

1.3 Estimation de la matrice d'information de Fisher

En collaboration avec Estelle Kuhn, nous nous intéressons dans [S1] à l'estimation de la matrice d'information de Fisher dans les modèles à variables latentes. Par souci de cohérence avec le reste du chapitre, je présente ce travail avec les notations des modèles à effets mixtes (équations (1.1)-(1.2)) mais la méthodologie développée et les résultats obtenus sont également valables dans des modèles à variables latentes plus généraux.

1.3.1 Introduction

La matrice d'information de Fisher (FIM, *Fisher Information Matrix*) est indispensable aussi bien pour l'analyse statistique que pour le recueil de données. Elle intervient par exemple dans l'évaluation de la variance asymptotique de l'EMV, nécessaire à la construction d'intervalles de confiance ou encore au calcul de certaines statistiques de test. Elle permet également d'évaluer des critères d'optimalité utiles en planification d'expériences, bien en amont de l'estimation. Il est donc important de pouvoir en faire le calcul ou à défaut, de l'estimer.

Dans un modèle où les variables observées Y_1, \dots, Y_N sont *i.i.d.* (indépendantes et identiquement distribuées), sous réserve que la fonction de log-vraisemblance observée soit

différentiable sur Θ et que $\|\partial_\theta \log p(y; \theta)(\partial_\theta \log p(y; \theta))^\top\|$ soit intégrable, la matrice d'information de Fisher est définie pour tout $\theta \in \Theta$ par :

$$I(\theta) = E_\theta \left[\partial_\theta \log p(Y_1; \theta)(\partial_\theta \log p(Y_1; \theta))^\top \right]. \quad (1.10)$$

Si de plus la fonction de log-vraisemblance est deux fois différentiable sur Θ , on peut également exprimer $I(\theta)$ sous la forme :

$$I(\theta) = -E_\theta \left[\partial_\theta^2 \log p(Y_1; \theta) \right]. \quad (1.11)$$

Pour des variables Y_1, \dots, Y_N indépendantes mais non identiquement distribuées - lorsque leur loi dépend de covariables par exemple - on peut donner une définition plus générale de la FIM (voir par exemple [Silvapulle M. J. et Sen P. K., 2001]), le passage à la définition générale n'ayant aucune incidence sur la suite.

Lorsqu'aucune des expressions (1.10) et (1.11) n'est calculable analytiquement, $I(\theta)$ est estimée à partir des observations de l'échantillon. L'estimateur usuel de la FIM est un M-estimateur construit à partir de (1.11) également connu sous la dénomination *matrice d'information de Fisher observée* :

$$I_{N,obs}(\theta) = -\frac{1}{N} \sum_{i=1}^N \partial_\theta^2 \log p(y_i; \theta). \quad (1.12)$$

La formule de Louis T. A. [1982] fournit une autre expression de $I_{N,obs}(\theta)$ pour les modèles à variables latentes, reposant sur les moments conditionnels des dérivées d'ordres 1 et 2 de la log-vraisemblance complète plutôt que sur les dérivées secondes des log-vraisemblances observées :

$$\begin{aligned} \partial_\theta^2 \log p(y_i; \theta) &= \mathbb{E}_{\psi_i|y_i;\theta} \left(\partial_\theta^2 \log(p(y_i, \psi_i; \theta)) \right) \\ &+ \mathbb{E}_{\psi_i|y_i;\theta} \left(\partial_\theta \log(p(y_i, \psi_i; \theta)) (\partial_\theta \log(p(y_i, \psi_i; \theta)))^\top \right) \\ &- \mathbb{E}_{\psi_i|y_i;\theta} \left(\partial_\theta \log p(y_i, \psi_i; \theta) \right) \left(\mathbb{E}_{\psi_i|y_i;\theta} \left(\partial_\theta \log p(y_i, \psi_i; \theta) \right) \right)^\top. \end{aligned}$$

Dans les modèles non linéaires à effets mixtes, l'absence d'expression explicite pour la vraisemblance (cf équation (1.4)) empêche en général le calcul direct de la FIM et de son estimateur $I_{N,obs}$. Les solutions les plus courantes consistent à évaluer la formule de Louis par Monte-Carlo ou par approximation stochastique [Delyon B. et al., 1999], [Kuhn E. et Lavielle M., 2004]. La revue de la littérature existante montre que l'ensemble des approches développées pour évaluer la FIM sont construites sur l'estimateur $I_{N,obs}(\theta)$ et nécessitent de calculer les dérivées partielles secondes de la log-vraisemblance observée ou complète. Cet aspect peut s'avérer limitant si dériver deux fois est difficile ou coûteux, comme par exemple dans certains modèles de culture complexes ou "boîte noire", ou plus rarement si la fonction de log-vraisemblance n'est pas suffisamment régulière.

1.3.2 Estimateur basé sur le score

Nous proposons de considérer l'estimateur :

$$I_{N,sco}(\theta) = \frac{1}{N} \sum_{i=1}^N \partial_\theta \log p(y_i; \theta)(\partial_\theta \log p(y_i; \theta))^\top. \quad (1.13)$$

Construit sur le score, $I_{N,sco}(\theta)$ est un M-estimateur naturel de la FIM. Il a l'avantage d'être défini dès lors que la FIM l'est alors que $I_{N,obs}(\theta)$ requiert des conditions de régularité supplémentaires. A notre connaissance, $I_{N,sco}(\theta)$ n'a jamais été utilisé dans la

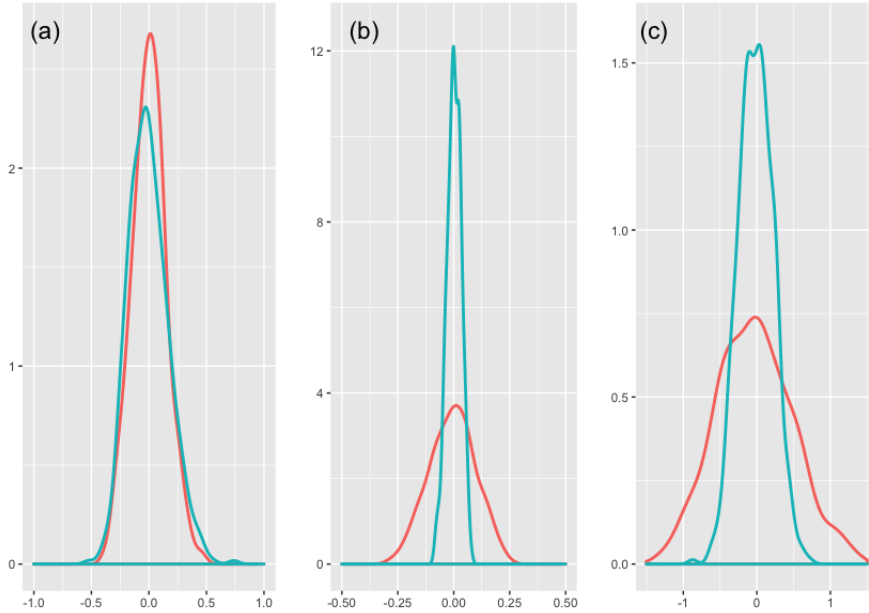


FIGURE 1.3 – Comparaison des distributions de $\tilde{I}^{sco} := \sqrt{N}(I_{N,sco}(\theta) - I(\theta))$ et $\tilde{I}^{obs} := \sqrt{N}(I_{N,obs}(\theta) - I(\theta))$.

Le modèle utilisé est un modèle de mélange de $K = 3$ lois de Poisson : $Y_i | \phi_i = k \sim \mathcal{P}(\lambda_k)$, $i = 1, \dots, N$, où Y_i et ϕ_i sont respectivement la variable observée et la variable d'état, $P(\phi_i = k) = \alpha_k$ et $\sum_{k=1}^K \alpha_k = 1$, avec pour valeurs de paramètres $\lambda_1 = 2$, $\lambda_2 = 5$, $\lambda_3 = 9$, $\alpha_1 = 0.3$ et $\alpha_2 = 0.5$. Ici le paramètre est $\theta = (\lambda_1, \lambda_2, \lambda_3, \alpha_1, \alpha_2)$.

Les graphiques représentent : (a) les distributions de \tilde{I}_{11}^{sco} (cyan) et \tilde{I}_{11}^{obs} (rouge), (b) les distributions de \tilde{I}_{22}^{sco} (cyan) et \tilde{I}_{22}^{obs} (rouge), (c) les distributions de \tilde{I}_{25}^{sco} (cyan) et \tilde{I}_{25}^{obs} (rouge), obtenues à partir de 500 échantillons simulés, de taille $N = 500$ chacun.

littérature. Nous montrons pourtant que $I_{N,sco}(\theta)$ et $I_{N,obs}(\theta)$ sont tous deux consistants et asymptotiquement gaussiens avec la même vitesse de convergence \sqrt{N} lorsque $N \rightarrow +\infty$. Des simulations menées dans différents modèles montrent qu'aucun des deux estimateurs n'est systématiquement meilleur que l'autre du point de vue du biais ou de sa covariance asymptotique (cf Figure 1.3). $I_{N,sco}$ est donc un estimateur tout à fait pertinent en théorie et en pratique.

Nous fournissons par ailleurs une formule analogue à la formule de Louis, permettant de relier $I_{N,sco}(\theta)$ aux espérances conditionnelles des dérivées de la log-vraisemblance des données complètes :

$$I_{N,sco}(\theta) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\psi_i | y_i; \theta} (\partial_\theta \log p(y_i, \psi_i; \theta)) \mathbb{E}_{\psi_i | y_i; \theta} (\partial_\theta \log p(y_i, \psi_i; \theta))^\top, \quad (1.14)$$

où $\mathbb{E}_{\psi | y; \theta}$ désigne l'espérance de ψ conditionnellement aux observations y pour une valeur de paramètre θ . Comme pour (1.13), le calcul de (1.14) a l'avantage de ne demander que des dérivations au premier ordre. La formule (1.14) est utile dans les modèles où les vraisemblances marginales $p(y_i; \theta)$ n'ont pas d'expression analytique alors que les espérances conditionnelles $\mathbb{E}_{\psi_i | y_i; \theta} (\partial_\theta \log p(y_i, \psi_i; \theta))$ sont explicites. C'est le cas par exemple de nombreux modèles de mélange. On reconnaît dans (1.14) les mêmes espérances conditionnelles que celles sur lesquelles reposent le calcul de la quantité Q de l'algorithme EM. Or, lorsque le modèle (1.1) n'est pas linéaire en ψ_i , celles-ci n'ont pas de forme analytique. Par conséquent, seul un calcul approché peut être envisagé.

1.3.3 Algorithmme

Dans le même esprit que [Delyon B. et al., 1999], nous proposons d'évaluer (1.14) par approximation stochastique. Plus précisément, l'estimation de la FIM selon (1.14) est obtenue comme sous-produit de l'algorithme SAEM mis en œuvre pour l'estimation des paramètres. Après initialisation des quantités Δ_i par Δ_i^0 pour $i = 1, \dots, N$, de Q par Q_0 et des paramètres du modèle par θ_0 , l'itération $m \geq 1$ se décompose en trois étapes :

1. *Etape S* : pour $i = 1, \dots, N$ simuler une réalisation de ψ_i^m sous la loi conditionnelle $p(\psi_i|y_i; \theta_{m-1})$ en utilisant la valeur courante du paramètre θ_{m-1} .
2. *Etape SA* : calculer les quantités

$$\begin{aligned} Q_m(\theta) &= (1 - \gamma_m)Q_{m-1}(\theta) + \gamma_m \sum_{i=1}^N \log p(y_i, \psi_i^m; \theta) \\ \Delta_i^m &= (1 - \gamma_m)\Delta_i^{m-1} + \gamma_m \partial_\theta \log p(y_i, \psi_i^m; \theta_{m-1}), \quad i = 1, \dots, N. \end{aligned}$$

3. *Etape M* : mettre à jour l'estimation du paramètre selon

$$\theta_m = \arg \max_{\theta \in \Theta} Q_m(\theta),$$

où (γ_m) est une suite de pas décroissante, telle que $\gamma_1 = 1$, $\lim_{m \rightarrow +\infty} \gamma_m = 0$.

Lorsque la convergence est atteinte, à l'itération M , on évalue alors l'estimateur de la FIM selon

$$I_{N,sco}^M = \frac{1}{N} \sum_{i=1}^N \Delta_i^M (\Delta_i^M)^\top.$$

Comme pour l'algorithme d'estimation (cf section 1.2.2), des méthodes MCMC peuvent être utilisées à l'étape S lorsque la simulation exacte sous $p(\psi_i|y_i; \theta)$ est impossible.

Nous montrons la convergence presque sûre des séquences (θ_m) et $(I_{N,sco}^m)$ produites par l'algorithme proposé en supposant

(M1') pour tout $i = 1, \dots, N$ le modèle conjoint $p(y_i, \psi_i; \theta)$ appartient à la famille exponentielle :

$$p(y_i, \psi_i; \theta) = \exp(-\kappa_i(\theta) + \langle S_i(\psi_i), \nu_i(\theta) \rangle),$$

où $\kappa_i(\cdot)$ et $\nu_i(\cdot)$ sont des fonctions de θ , $S_i(\cdot)$ sont les statistiques exhaustives du modèle prenant leur valeur dans un sous-ensemble \mathcal{S}_i de \mathbb{R}^{q_i} et $\langle \cdot, \cdot \rangle$ est le produit scalaire sur \mathbb{R}^{q_i} .

Notons que dans ce cas, les différentes étapes de l'algorithme s'expriment à partir des statistiques exhaustives des modèles individuels. Après initialisation des statistiques exhaustives par s_i^0 , $i = 1, \dots$, et des paramètres du modèle par θ_0 , l'itération $m \geq 1$ se décompose comme suit :

1. *Etape S* : pour $i = 1, \dots, N$ simuler une réalisation de ψ_i^m sous la loi conditionnelle $p(\psi_i|y_i; \theta_{m-1})$ en utilisant la valeur courante du paramètre θ_{m-1} .
2. *Etape SA* : pour $i = 1, \dots, N$, calculer les quantités

$$s_i^m = (1 - \gamma_m)s_i^{m-1} + \gamma_m S_i(\psi_i^m).$$

3. *Etape M* : mettre à jour l'estimation du paramètre selon

$$\theta_m = \arg \max_{\theta \in \Theta} \sum_{i=1}^N (-\kappa_i(\theta) + \langle s_i^m, \nu_i(\theta) \rangle) = \hat{\theta}(s^m).$$

L'estimateur de la FIM est évalué lorsque la convergence est atteinte, à l'itération M , selon la formule :

$$I_{N,sco}^M = \frac{1}{N} \sum_{i=1}^N \hat{\Delta}_i \left(\hat{\theta}(s^M) \right) \hat{\Delta}_i \left(\hat{\theta}(s^M) \right)^\top$$

où $\hat{\Delta}_i(\hat{\theta}(s)) = -\partial \kappa_i(\hat{\theta}(s)) + \left\langle s_i, \partial \nu_i(\hat{\theta}(s)) \right\rangle$ pour tout s .

Sous **(M1')** et sous des hypothèses de régularité similaires à celles qui garantissent le résultat général de convergence de SAEM dans [Delyon B. et al., 1999] et en supposant comme dans [Delyon B. et al., 1999] que pour tout $m \in \mathbb{N}$, $\gamma_m \in [0, 1]$, $\sum_{m=1}^{\infty} \gamma_m = \infty$ et $\sum_{m=1}^{\infty} \gamma_m^2 < \infty$, nous obtenons le résultat suivant :

Théorème 1 Soit $\mathcal{L}_N = \{\theta \in \Theta, \partial_\theta L_N(\theta) = 0\}$ l'ensemble des points stationnaires de la log-vraisemblance des observations. Alors, pour tout $\theta_0 \in \Theta$, et tout $N \in \mathbb{N}^*$, $\lim_m d(\theta_m, \mathcal{L}_N) = 0$ et $\lim_m d(I_{N,sco}^m, \mathcal{I}_N) = 0$ où $\mathcal{I}_N = \{I(\theta), \theta \in \mathcal{L}_N\}$.

Nous montrons également la normalité asymptotique des séquences (θ_m) et $(\bar{I}_{N,sco}^m)$ où $\bar{I}_{N,sco}^m$ est un estimateur moyen défini par $\bar{I}_{N,sco}^m = \frac{1}{N} \sum_{i=1}^N \hat{\Delta}_i(\hat{\theta}(\bar{s}^m)) \hat{\Delta}_i(\hat{\theta}(\bar{s}^m))^t$ et $\bar{s}^m = \sum_{l=1}^m s^l / m$. Par rapport au théorème 1 et au résultat de normalité asymptotique énoncé dans [Delyon B. et al., 1999], nous imposons de plus qu'il existe γ^* tel que $\lim m^\alpha / \gamma_m = \gamma^*$, et $\gamma_m / \gamma_{m+1} = 1 + O(m^{-1})$. Sous l'ensemble de ces conditions, le résultat s'énonce :

Théorème 2 Il existe un point fixe stable et régulier $\theta^* \in \Theta$ tel que $\lim_m \theta_m = \theta^*$ p.s. De plus, la suite $(\sqrt{m}(\text{Vect}(\bar{I}_{N,sco}^m) - \text{Vect}(\bar{I}_{N,sco}(\theta^*)))) \mathbb{1}_{\lim \|\theta_m - \theta^*\| = 0}$ converge en loi vers une distribution gaussienne multivariée centrée lorsque m tend vers l'infini.

1.3.4 Etude sur données simulées

Nous utilisons le modèle de pharmacocinétique suivant :

$$Y_{ij} = \frac{D_i k a_i}{V_i k a_i - Cl_i} \left[e^{-\frac{Cl_i}{V_i} t_{ij}} - e^{-k a_i t_{ij}} \right] + \varepsilon_{ij}, \quad \varepsilon_{ij} \underset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2). \quad (1.15)$$

$\psi_i = (k a_i, Cl_i, V_i)$ est le vecteur des paramètres individuels tels que $\log(\psi_i) = \mu + \eta_i$ où $\mu = (\log(k a), \log(Cl), \log(V))$, $\eta_i \underset{i.i.d.}{\sim} \mathcal{N}(0, \Omega)$ et $\Omega = \text{diag}(\omega_{ka}^2, \omega_{Cl}^2, \omega_V^2)$.

Y_{ij} représente la mesure de concentration du médicament pour l'individu i au temps t_{ij} , D_i est la dose reçue par l'individu i , et V_i , $k a_i$ et Cl_i représentent respectivement le volume du compartiment central, la constante d'absorption du médicament et la clairance du médicament pour le sujet i . Les paramètres de population sont $\theta = (k a, V, Cl, \omega_{ka}^2, \omega_V^2, \omega_{Cl}^2, \sigma^2)$.

Dans ce modèle, $I(\theta)$, $I_{N,sco}(\theta)$ et $I_{N,obs}(\theta)$ n'ont pas d'expression explicite. Néanmoins, comme les log-vraisemblances des données complètes $\log p(y_i, \psi_i; \theta)$ sont explicites, $I_{N,sco}(\theta)$ et $I_{N,obs}(\theta)$ peuvent être évalués par approximation stochastique, selon l'algorithme décrit dans la section 1.3.3 pour $I_{N,sco}(\theta)$ et la procédure de [Delyon B. et al., 1999] pour $I_{N,obs}(\theta)$.

Famille exponentielle. Lorsque tous les paramètres individuels sont aléatoires, le modèle vérifie les hypothèses **(M1')** et **(M1)** qui garantissent les propriétés théoriques des algorithmes d'approximation stochastique pour le calcul de ces deux estimateurs.

Un jeu de données de $N = 100$ sujets est simulé selon ce modèle en utilisant les valeurs suivantes pour les paramètres : $V = 31$, $k a = 1.6$, $Cl = 1.8$, $\omega_V^2 = 0.40$, $\omega_{ka}^2 = 0.40$, $\omega_{Cl}^2 = 0.40$ et $\sigma^2 = 0.75$, la même dose $d_i = 320$ et les mêmes temps d'observations (en

heures) : 0.25, 0.5, 1, 2, 3.5, 5, 7, 9, 12, 24 pour tous les sujets. On répète 500 fois les procédures d'estimation de $I_{N,sco}(\hat{\theta})$ et $I_{N,obs}(\hat{\theta})$. $\hat{\theta}$ est estimé conjointement.

Les Figures 1.4 et 1.5 représentent respectivement les biais relatifs empiriques et les écart-types relatifs empiriques calculés à chaque itération pour les composantes de $I_{N,sco}$ (courbe rouge) et $I_{N,obs}$ (courbe cyan). On observe que le biais et les écarts types des estimations des composantes des deux matrices diminuent au fil des itérations. Pour $I_{N,sco}$ comme pour $I_{N,obs}$, le biais est presque nul à convergence de l'algorithme. Ces résultats montrent qu'aucune des deux méthodes n'est systématiquement meilleure que l'autre en termes de biais ou d'écart-type.

FIGURE 1.4 – Représentation du biais relatif moyen des composantes diagonales de la matrice d'information de Fisher estimée au fil des itérations. La ligne rouge correspond à $I_{N,sco}(\theta)$ et la ligne cyan correspond à $I_{N,obs}(\theta)$. L'axe des ordonnées représente le biais relatif moyen et l'axe des abscisses représente les itérations de l'algorithme. Les itérations de chauffe des algorithmes stochastiques ne sont pas représentées.

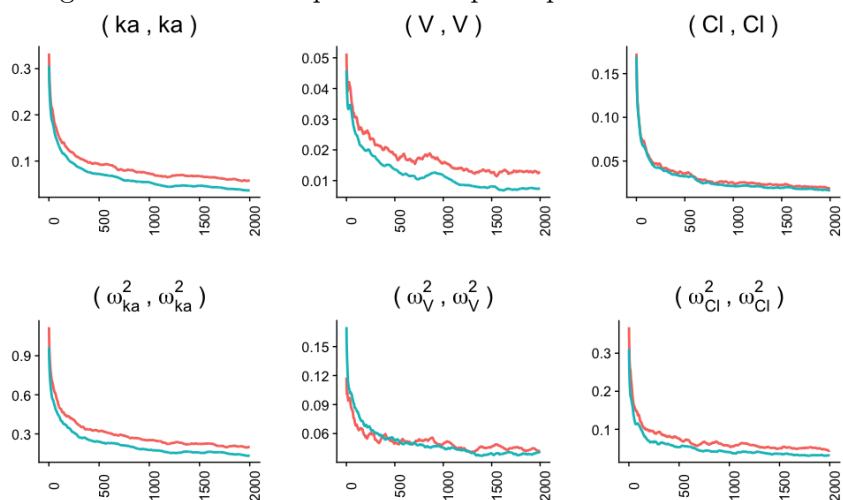
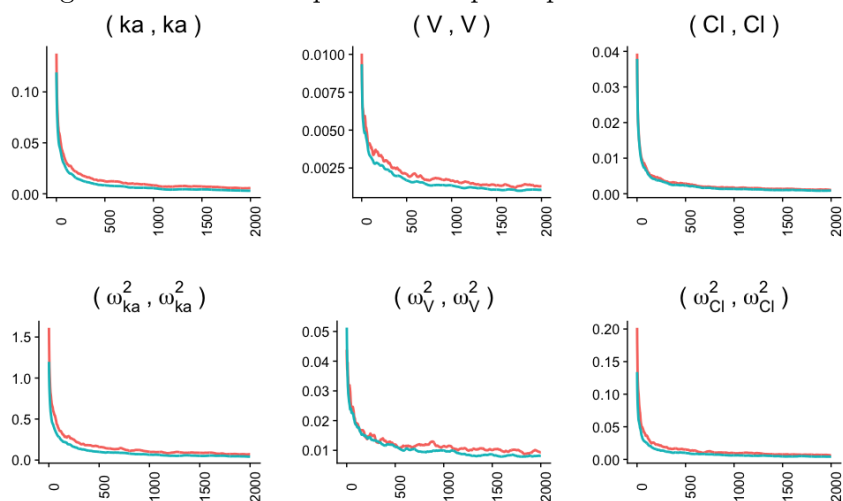


FIGURE 1.5 – Représentation de l'écart-type relatif moyen des composantes diagonales de la matrice d'information de Fisher estimée au fil des itérations. La ligne rouge correspond à $I_{N,sco}(\theta)$ et la ligne cyan correspond à $I_{N,obs}(\theta)$. L'axe des ordonnées représente l'écart-type relatif moyen et l'axe des abscisses représente les itérations de l'algorithme. Les itérations de chauffe des algorithmes stochastiques ne sont pas représentées.



Cas général. Si l'un des paramètres individuels n'est pas aléatoire, le modèle n'est plus exponentiel au sens de (M1) et (M1'). Les algorithmes stochastiques pour l'estimation de $I_{N,sco}$ et $I_{N,obs}$ peuvent être implémentés, mais leurs propriétés ne sont plus garanties en théorie. Nous répétons la même étude de simulation que précédemment en fixant $\omega_V^2 = 0$. L'algorithme converge bien en pratique malgré l'absence de garantie théorique. Par ailleurs, les évaluations de $I_{N,sco}$ permettent de construire des intervalles de confiance asymptotiques pour les paramètres. Les taux de recouvrement empiriques obtenus sont cohérents avec les niveaux théoriques des intervalles de confiance construits.

1.4 Caractérisation de la variabilité inter-individuelle par sélection de modèle

Depuis la fin de ma thèse, je m'intéresse également à des questions de choix de modèle dans les modèles à effets mixtes. Je considère plus spécifiquement la situation où le *modèle individuel* de l'équation (1.1) est fixé, et où l'objectif est de modéliser la variabilité inter-individuelle. L'équation (1.2) :

$$g(\psi_i) = \mu + \beta C_i + \eta_i, \quad \eta_i \underset{i.i.d.}{\sim} \mathcal{N}(0, \Omega),$$

décompose la variabilité inter-individuelle en deux termes principaux. Dans le premier, les effets fixes β décrivent les différences entre individus dues aux covariables C_i . Dans le second, l'effet aléatoire η_i décrit la part de variabilité non liée aux covariables mesurées, qui correspond également à la variabilité entre sujets qui partageraient les mêmes valeurs de covariables. Sélectionner les covariables les plus pertinentes et la structure des effets aléatoires la mieux adaptée, *i.e.* la position des effets aléatoires et les corrélations entre effets aléatoires, sont des questions cruciales dans la conception d'un modèle mixte. Dans (1.2), cela revient à identifier les coefficients non nuls de β et les composantes non nulles de Ω .

Lorsque j'ai commencé à m'intéresser à cette question, quelques méthodes de sélection de variables et d'effets aléatoires avaient été proposées, mais elles ne concernaient que les modèles linéaires généralisés à effets mixtes et peu d'entre elles étaient implémentées dans les logiciels (voir par exemple la revue de Muller S. et al. [2013] et les références qui y sont citées). Ces questions de sélection de modèle ont donné lieu à de nombreuses contributions par la suite, mais beaucoup restent spécifiques aux modèles linéaires généralisés à effets mixtes (*e.g.* [Hui F. et Muller S., 2017]), ou ont été développées pour répondre à des questions très particulières telles que la comparaison de groupes d'individus dans les modèles non linéaires à effets mixtes (*e.g.* [Ollier E. et al., 2016]). Par ailleurs, beaucoup des méthodes proposées sont des méthodes régularisées initialement conçues pour les contextes de grande dimension, où le nombre de variables est important devant la taille de l'échantillon ; leur utilisation est peu commune en pharmacométrie. D'une manière générale, aucune méthodologie ne fait consensus sur la façon de construire le *modèle de population* dans les modèles non linéaires à effets mixtes. L'objectif général de mes travaux sur ce sujet est de proposer une méthodologie de sélection de modèle qui soit opérationnelle dans une large gamme de modèles à effets mixtes, y compris non linéaires, justifiée par des arguments théoriques et simple d'utilisation pour un public non statisticien.

J'ai proposé deux contributions sur la sélection de covariables et d'effets aléatoires dans les modèles non linéaires à effets mixtes, toutes deux articulées autour du critère d'information bayésien (BIC) et issues de collaborations avec Marie-Anne Poursat ([A6],[A15]). Le BIC, introduit par [Schwarz G., 1978], est un outil de comparaison de modèles très populaire pour la simplicité de son utilisation. Soit un modèle m candidat au problème de modélisation considéré. La définition standard du BIC fait intervenir la log-vraisemblance observée en l'estimateur du maximum de vraisemblance $\ell(\hat{\theta}_m)$ et une pénalité dépendant

de la complexité du modèle $\dim(m)$ et de la taille de l'échantillon n_s :

$$BIC_m = -2\ell(\hat{\theta}_m) + \dim(m) \log(n_s).$$

Le meilleur modèle au sens de BIC est celui qui présente la plus petite valeur de ce critère. La pénalisation de la vraisemblance est primordiale pour ne pas systématiquement choisir le modèle le plus complexe, mais la définition de la pénalité du BIC est équivoque dans les modèles à effets mixtes. En particulier, la taille d'échantillon est difficile à définir. On trouve ainsi différentes implémentations du critère dans les logiciels, n_s étant tantôt associé au nombre de sujets N , tantôt au nombre total d'observations $n_{tot} = \sum_{i=1}^N n_i$. Par exemple, la pénalité $\log(n_{tot})$ est utilisée dans le package `nlme` de R et la procédure MIXED de SPSS, alors que la pénalité $\log(N)$ est utilisée dans MONOLIX, le package `saemix` de R ou la proc NLMIXED de SAS.

Dans [A6] et [A15], nous avons repris les fondements théoriques du BIC ([Lebarbier E. et Mary-Huard T., 2006], [Claeskens G. et Hjort N. L., 2008]) et adapté la pénalité du critère pour la sélection de covariables et la sélection des effets aléatoires dans les modèles non linéaires à effets mixtes. Nous avons ensuite proposé un algorithme pour la sélection conjointe de covariables et d'effets aléatoires.

1.4.1 Utilisation du BIC dans les modèles à effets mixtes

Dans la suite, j'appellerai *structure de covariables* d'un modèle l'ensemble

$$S_\beta = \{(k, k') \in \{1, \dots, d\} \times \{1, \dots, q\} : \beta_{k,k'} \neq 0\}$$

des positions des effets fixes non nuls de ce modèle. De même, je parlerai de *structure de covariance* en référence à l'ensemble

$$S_\Omega = \{(k, k') \in \{1, \dots, d\}^2 : \omega_{k,k'} \neq 0, k \leq k'\}$$

des positions des composantes non nulles de la matrice de covariance Ω des effets aléatoires de ce modèle. L'objectif est d'identifier parmi une collection de modèles candidats pour (1.2), notés m par la suite, celui qui fournit la meilleure description de la variabilité inter-individuelle.

Partition des paramètres. Les travaux [A6] et [A15] reposent sur une partition spécifique des paramètres de population θ , suggérée par Nie L. [2007] dans son étude des propriétés asymptotiques de l'EMV dans les modèles non linéaires à effets mixtes. L'idée est d'utiliser une formulation dégénérée du *modèle de population* (1.2), obtenue en réorganisant les paramètres individuels selon la structure de covariance S_Ω du modèle. Plus précisément ψ_i est divisé en $\psi_{F,i}$, ses composantes non aléatoires, et $\psi_{R,i}$, ses composantes aléatoires. $\psi_{R,i}$ correspond aux termes diagonaux (k, k) , $k = 1, \dots, d$ de Ω dans S_Ω , alors que $\psi_{F,i}$ correspond aux termes de variance nuls de Ω . Si $\psi_{F,i} \neq \emptyset$, la matrice Ω est singulière et peut s'écrire $\Omega = \text{diag}(0, \Omega_R)$. Cette réorganisation des paramètres individuels implique la décomposition suivante pour les paramètres de population :

$$\mu = \begin{pmatrix} \mu_F \\ \mu_R \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_F \\ \beta_R \end{pmatrix}, \quad \Omega = \text{diag}(0, \Omega_R), \quad (1.16)$$

où la notation $\beta = \begin{pmatrix} \beta_F \\ \beta_R \end{pmatrix}$ (resp. $\mu = \begin{pmatrix} \mu_F \\ \mu_R \end{pmatrix}$) signifie que la matrice β (resp. μ) s'écrit comme la concaténation verticale des deux matrices β_F et β_R (resp. μ_F and μ_R), d'où :

$$g \begin{pmatrix} \psi_{F,i} \\ \psi_{R,i} \end{pmatrix} = \begin{pmatrix} \mu_F \\ \mu_R \end{pmatrix} + \begin{pmatrix} \beta_F \\ \beta_R \end{pmatrix} C_i + \begin{pmatrix} 0 \\ \eta_{R,i} \end{pmatrix}, \quad \eta_{R,i} \underset{i.i.d.}{\sim} \mathcal{N}(0, \Omega_R).$$

Critères. Dans [A6], nous proposons un critère de type BIC pour la sélection de covariables :

$$BIC_\beta = -2\ell(\hat{\mu}, \hat{\beta}, \hat{\Omega}_R) + \dim(\{\mu_R, \beta_R\}) \log N + \dim(\{\mu_F, \beta_F\}) \log n_{\text{tot}}, \quad (1.17)$$

où $\ell(\hat{\mu}, \hat{\beta}, \hat{\Omega}_R)$ est la log-vraisemblance des observations en l'estimateur du maximum de vraisemblance des paramètres et $\dim(A)$ dénombre les éléments non nuls de l'ensemble A . La pénalité de BIC_β est une combinaison linéaire des logarithmes du nombre de sujets et du nombre total d'observations dont les coefficients dépendent de la paramétrisation et des niveaux de variabilité exprimés dans le modèle via la structure de covariance des effets aléatoires S_Ω . Il s'agit en réalité d'un BIC conditionnel, puisque la pénalité suppose S_Ω connue et fixée. Il ne permet donc de comparer que des *modèles de population* de même structure d'effets aléatoires. Ce résultat est original dans le sens où il offre un compromis entre les définitions empiriques du BIC implémentées dans les logiciels pour les modèles à effets mixtes. Dans [A15], nous montrons qu'un BIC classique de pénalité $\log(N)$ permet la sélection d'effets aléatoires dans (1.2) :

$$BIC_\Omega = -2\ell(\hat{\mu}, \hat{\beta}, \hat{\Omega}_R) + \text{card}(S_\Omega) \log(N), \quad (1.18)$$

où $\text{card}(S_\Omega)$ est le cardinal de S_Ω .

Idées de preuve. Les preuves de (1.17) et (1.18) reposent sur l'évaluation asymptotique des vraisemblances conditionnelles $p(m|y)$ obtenues par approximation de Laplace des vraisemblances observées $p(y|m)$:

$$\log p(y|m) \approx \log p(y|\hat{\theta}_m, m) - \frac{1}{2} \log \det H_{\hat{\theta}_m},$$

où $\hat{\theta}_m$ et $H_{\hat{\theta}_m}$ désignent respectivement l'estimateur du maximum de vraisemblance de θ et la matrice d'information observée. Dans les modèles classiques à effets fixes, sous des hypothèses de régularité usuelles sur la log-vraisemblance, la Hessienne $H_{\hat{\theta}_m}$ se comporte comme N fois la matrice d'information du modèle $\mathcal{I}_{\hat{\theta}_m}$, d'où il vient :

$$\frac{1}{2} \log \det H_{\hat{\theta}_m} \approx \frac{\dim(\theta)}{2} \log(N) + \frac{1}{2} \log \det \mathcal{I}_{\hat{\theta}_m}.$$

Le second terme se comporte comme une constante lorsque $N \rightarrow +\infty$. On en déduit l'approximation asymptotique du BIC classique

$$\log p(y|m) \approx \log p(y|\hat{\theta}_m, m) - \frac{\dim(\theta)}{2} \log(N).$$

Cette évaluation n'est pas toujours exacte dans les modèles à effets mixtes, selon que $\min n_i \rightarrow +\infty$ ou non. Sans perte de généralité, nous supposons que $n_i = n$ pour tout $i = 1, \dots, N$. Nous utilisons les résultats de Nie L. [2007] pour montrer que l'ordre de grandeur des composantes de $H_{\hat{\theta}_m}$ dépend à la fois du nombre de sujets N et des nombres d'observations par sujet n_i . Lorsque la structure des effets aléatoires est fixée, non sujette à sélection, on peut montrer que lorsque $N, n \rightarrow +\infty$ et $n/N \rightarrow +\infty$,

$$\log \det H_{\hat{\theta}_m} \approx \log \det(N\mathcal{I}_1) + \log \det(Nn\mathcal{I}_2) + \mathcal{O}(N/n),$$

où \mathcal{I}_1 et \mathcal{I}_2 sont des matrices de plein rang d'ordre $\mathcal{O}(1)$ de dimensions $\dim(\{\mu_R, \beta_R\}) \times \dim(\{\mu_R, \beta_R\})$ et $\dim(\{\mu_F, \beta_F\}) \times \dim(\{\mu_F, \beta_F\})$ respectivement, ce dont on déduit :

$$\log p(y|m) \approx \log p(y|\hat{\theta}_m, m) - \dim(\{\mu_R, \beta_R\}) \log(N) - \dim(\{\mu_F, \beta_F\}) \log(Nn),$$

et l'expression de BIC_β pour la sélection de covariables. Ce résultat s'obtient sous des conditions de régularité sur les log-vraisemblances $\log p(y_i; \theta)$ et log-vraisemblances complètes $\log p(y_i, \psi_i; \theta)$.

Lorsque la sélection porte sur les effets aléatoires du modèle, on évalue $\log \det H_{\hat{\theta}_m}$ de la même façon. Cette fois, on obtient que

$$\log p(y|m) \approx \log p(y|\hat{\theta}_m, m) - \dim(\Omega_R) \log(N),$$

d'où l'expression de BIC_Ω .

Evaluation numérique des critères. Dans [A6], nous proposons une étude de simulation dans un modèle linéaire à effets mixtes avec trois effets aléatoires, dont l'objectif est de comparer les performances de BIC_β avec les deux versions usuelles du BIC

$$BIC_N = -2\ell(\hat{\theta}) + \dim(\theta) \log(N), \quad (1.19)$$

$$BIC_{n_{tot}} = -2\ell(\hat{\theta}) + \dim(\theta) \log(n_{tot}), \quad (1.20)$$

dans différentes situations combinant différents nombres et positions de covariables et d'effets aléatoires. Les résultats montrent que BIC_β se comporte dans les diverses situations simulées au moins aussi bien que le meilleur des deux critères standards BIC_N et $BIC_{n_{tot}}$. Les performances de BIC_N et $BIC_{n_{tot}}$ dépendent fortement de la structure de covariance des effets aléatoires. Ainsi, en adaptant automatiquement sa pénalité en fonction de la structure de covariance du modèle, BIC_β est un compromis utile entre BIC_N et $BIC_{n_{tot}}$ pour la sélection des covariables dans une approche de population. Le critère proposé est globalement très intéressant pour le problème de sélection de covariables considéré. A l'occasion du travail [A15], nous vérifions également sur des simulations que BIC_Ω est un critère intéressant en pratique pour l'identification de la structure de covariance des effets aléatoires. Par souci de concision, cette étude n'est pas insérée dans la publication [A15].

Algorithme de sélection conjointe. Les critères BIC_β et BIC_Ω permettent de traiter séparément la sélection de covariables et la sélection des effets aléatoires. En général, les deux questions se posent simultanément. D'un point de vue théorique, BIC_N (1.19) présente la pénalité la plus cohérente pour comparer des modèles qui diffèrent à la fois par leurs structures de covariables et par leurs structures d'effets aléatoires. Toutefois, lorsque le modèle contient d paramètres individuels et lorsque q covariables sont mesurées pour chaque individu, la combinatoire entre covariables et effets aléatoires peut conduire à un nombre considérable de modèles candidats, y compris pour des petites valeurs de q et d . Par exemple, si $q = 6$ et $d = 3$, il y a un peu plus de deux millions de modèles possibles pour (1.2). Une recherche exhaustive n'est donc pas envisageable. En pratique, le choix du *modèle de population* dans les modèles non linéaires à effets mixtes est traité en deux temps par les modélisateurs. Dans les approches les plus courantes, une première étape consiste à sélectionner les covariables, puis une seconde à sélectionner les effets aléatoires et leur structure de variance-covariance. Ces deux étapes sont, selon les communautés, exécutées dans des ordres différents et avec des outils différents (voir par exemple [Lavielle M. et Mentré F., 2007] pour des modèles non linéaires à effets mixtes ou encore la fonction `step` du package R `lmerTest` [Kuznetsova A. et al., 2017] pour les modèles linéaires à effets mixtes). Une fois la structure choisie à la deuxième étape, la sélection obtenue à l'étape précédente n'est généralement pas remise en question. Toutefois, compte-tenu de la relation (1.2), la modification de la structure des effets aléatoires peut conduire à des sélections différentes des covariables, et inversement. Dans [A15], nous proposons une procédure itérative, alternant :

1. une étape de *sélection de covariables* à structure d'effets aléatoires fixée, en envisageant l'inclusion ou l'élimination d'une covariable à la fois parmi un ensemble de covariables proposées par l'utilisateur, dans l'esprit des procédures *stepwise* usuelles,

2. une étape de *sélection des effets aléatoires*, par sélection d'une structure de covariance parmi une collection de structures de covariances proposées par l'utilisateur.

Ces étapes sont répétées jusqu'à convergence, c'est-à-dire jusqu'à ce que les structures S_β et S_Ω retenues ne changent plus d'une itération à la suivante. Les deux types de sélection y sont traités avec les BIC_β et BIC_Ω établis plus haut. L'Algorithme 1 décrit en détail la procédure proposée.

Algorithme 1 : Algorithme de sélection de type stepwise pour la sélection conjointe de covariables et d'effets aléatoires [A15].

Initialisation. Choisir un modèle initial M_0 pour (1.2), de structure de covariables $S_\beta^{(0)}$ et de structure d'effets aléatoires $S_\Omega^{(0)}$. Les effets fixes et la matrice de covariance des effets aléatoires de M_0 sont $\beta^{(0)}$ et $\Omega_R^{(0)}$ respectivement.

Itération t . Soient $S_\beta^{(t)}$ et $S_\Omega^{(t)}$ les structures de covariables et d'effets aléatoires retenues à l'itération $t - 1$.

1. **Sélection de la structure des effets aléatoires.** On définit une collection de modèles par combinaison de la structure d'effets fixes $S_\beta^{(t)}$ retenue à l'itération précédente avec toutes les structures de covariance candidates, $(S_\Omega^k)_{k=1,\dots,K}$, définies en amont par l'utilisateur. Chaque S_Ω^k correspond à une matrice de covariance $\Omega_{R,k}$.
 - Estimer les K modèles et évaluer leurs log-vraisemblances respectives $\ell(\widehat{\mu}^{(t)}, \widehat{\beta}^{(t)}, \widehat{\Omega}_{R,k})$, $k = 1, \dots, K$.
 - Choisir la meilleure des K structures au sens de BIC_Ω , notée $S_\Omega^{(t+1)}$.
 2. **Sélection de la structure de covariables.** On définit une collection de modèles par combinaison de la structure d'effets aléatoires $S_\Omega^{(t+1)}$ retenue à l'itération précédente avec les L_t structures de covariables obtenues par l'ajout ou la suppression d'une covariable dans $S_\beta^{(t)}$. Chacune d'elle correspond à un effet fixe β_l , $l = 1, \dots, L_t$.
 - Estimer les L_t modèles et évaluer leurs log-vraisemblances $\ell(\widehat{\mu}_l, \widehat{\beta}_l, \widehat{\Omega}_R^{(t+1)})$, $l = 1, \dots, L_t$.
 - Choisir la meilleure des L_t structures au sens de BIC_β , notée $S_\beta^{(t+1)}$.
 3. **Répéter 1. and 2.** jusqu'à convergence.
-

Evaluation numérique de l'algorithme. Dans [A15], nous menons une étude de simulations approfondie pour évaluer les performances de cet algorithme. Nous considérons un modèle linéaire à effets mixtes, un modèle de Poisson à effets mixtes et un modèle non linéaire à effets mixtes avec différents nombres de sujets N et différents nombres d'observations par sujet n . Dans le modèle linéaire à effets mixtes, nous étudions la sensibilité de l'algorithme au nombre de covariables considérées et aux niveaux de variances des effets aléatoires. Selon les modèles considérés et la complexité du problème de sélection, nous comparons l'Algorithme 1 à une recherche exhaustive par le critère BIC_N (1.19) et à la méthode régularisée, rPQL, de Hui F. et Muller S. [2017]. Les résultats de notre étude montrent que notre algorithme est tout aussi performant qu'une recherche exhaustive par le critère BIC_N et présente une alternative efficace sur le plan du temps de calcul dans les situations où le nombre de covariables et le nombre d'effets aléatoires définissent un nombre élevé de modèles candidats. Dans les modèles linéaires généralisés à effets mixtes, notre algorithme présente des performances comparables à celles de la méthode rPQL. Les procédures stepwise sont réputées sensibles à l'initialisation, ici le choix de M_0 . Nous ne retrouvons pas ce comportement sur les modèles simulés dans cette expérience. En

conclusion, l'algorithme que nous proposons est une très bonne alternative aux méthodes existantes pour la sélection conjointe des covariables et des effets aléatoires. C'est à notre connaissance la première méthode opérationnelle dans les modèles non linéaires à effets mixtes reposant sur des critères issus d'une étude théorique.

La consistance des critères et de l'algorithme proposés restent des questions non résolues qui alimentent à l'heure actuelle mes projets de recherche à court terme (voir chapitre 4).

1.4.2 Application en pharmacocinétique

Pour illustrer la méthodologie proposée, nous reprenons dans [A15] l'étude publiée dans [Burdet C. et al., 2015]. Dans cette étude, 60 patients atteints de pneumopathie sous ventilation assistée (PVA) reçoivent un antibiotique en perfusion avec une dose unique de 20 mg par kg de poids corporel. La concentration d'antibiotique dans le sang est ensuite mesurée à plusieurs reprises. Douze covariables sont également mesurées pour chaque patient pendant l'expérience, dont l'âge (*age*), le sexe (*sex*), le poids corporel total (*w*), un indicateur de détresse respiratoire (*P/F*), la clairance de la créatinine sur 4 heures (*ClCr*). L'objectif principal de cette étude est de caractériser la pharmacocinétique de l'antibiotique ainsi que sa variabilité entre les patients souffrant de PVA grâce à un modèle statistique qui pourrait ensuite être utilisé pour choisir les doses à administrer à de nouveaux patients. Le modèle structurel (*modèle individuel*) retenu pour les mesures de concentration Y_{ij} est un modèle à deux compartiments défini selon :

$$Y_{ij} = f_{ij} + (u + v f_{ij}) \epsilon_{ij}, \quad f_{ij} = f(D_i, t_{ij}, t_{D,i}, T_{inf,i}, \psi_i), \quad \epsilon_{ij} \underset{i.i.d.}{\sim} \mathcal{N}(0, 1),$$

où

$$f(D, t, t_D, T_{inf}, \psi) =$$

$$\frac{D}{T_{inf}} \left[\frac{A}{a} (1 - e^{-\alpha(t-t_D)}) + \frac{B}{b} (1 - e^{-\beta(t-t_D)}) \right] \text{ si } t - t_D \leq T_{inf},$$

$$\frac{D}{T_{inf}} \left[\frac{A}{a} (1 - e^{-aT_{inf}}) e^{-a(t-t_D-T_{inf})} + \frac{B}{b} (1 - e^{-bT_{inf}}) e^{-b(t-t_D-T_{inf})} \right] \text{ sinon,}$$

$$A = \frac{1}{V_1} \frac{a - Q/V_2}{a - b}; \quad B = \frac{1}{V_1} \frac{b - Q/V_2}{b - a}; \quad a = \frac{QCl}{V_1 V_2 b};$$

$$b = \frac{1}{2} \left[\frac{Q}{V_1} + \frac{Q}{V_2} + \frac{Cl}{V_1} - \sqrt{\left(\frac{Q}{V_1} + \frac{Q}{V_2} + \frac{Cl}{V_1} \right)^2 - 4 \frac{Q}{V_2} \frac{Cl}{V_1}} \right].$$

Dans ce modèle, les paramètres individuels ψ sont Q, Cl, V_1, V_2 , où Cl est la clairance de l'antibiotique, V_1 est le volume de distribution du compartiment central, V_2 est le volume de distribution du compartiment périphérique, et Q est la clairance inter-compartimentale. Les variables structurelles, notées x_i dans (1.1), sont la dose D_i , les temps de mesure t_{ij} , l'instant d'administration de la dose $t_{D,i}$ et la durée de la perfusion $T_{inf,i}$. Pour respecter les contraintes de positivité sur les paramètres pharmacocinétiques, le *modèle de population* est donné par $\log \psi_i = \mu + \beta C_i + \eta_i$, où $\eta_i \underset{i.i.d.}{\sim} \mathcal{N}(0, \Omega)$ et C_i est le vecteur des covariables pour l'individu i . Dans [Burdet C. et al., 2015], la construction du *modèle de population* s'appuie sur la procédure décrite dans [Lavielle M. et Mentré F., 2007]. Cette procédure traite séparément la sélection des covariables et la sélection des effets aléatoires, en associant des tests de Wald et des tests de rapport de vraisemblance pour la sélection des covariables, et en utilisant un critère BIC pour la sélection de la structure de covariance des effets aléatoires. Nous utilisons l'Algorithme 1 pour identifier les covariables influentes et la structure d'effets aléatoires la plus appropriée pour caractériser la variabilité inter-individuelle, puis comparons les résultats obtenus avec ceux publiés dans [Burdet C. et al.,

TABLE 1.2 – Comparaison des *modèles de population* retenus avec notre algorithme de sélection (★) et l'étude de [Burdet C. et al., 2015] (●).

	<i>ClCr</i>	<i>age</i>	<i>sex</i>	<i>w</i>	<i>P/F</i>	η
<i>CL</i>	★ ●	★				★ ●
<i>V</i> ₁				★ ●	★ ●	★ ●
<i>Q</i>			★			●
<i>V</i> ₂	★					★ ●

TABLE 1.3 – Comparaison de modèles sur les données de [Burdet C. et al., 2015].

Model	BIC
\mathcal{M}_1	1492.624
\mathcal{M}_2	1643.676
\mathcal{M}_3	1513.123

2015]. Pour éviter les difficultés numériques dues à la complexité du modèle et au nombre élevé de covariables disponibles, nous avons choisi de démarrer l'algorithme avec le modèle sans covariables. Comme le nombre de structures d'effets aléatoires pour ψ_i est élevé, nous avons choisi de limiter l'algorithme aux matrices de covariance diagonales et avons ensuite essayé, *a posteriori*, d'ajouter des corrélations entre les effets aléatoires du modèle retenu. En excluant les données censurées, $N = 53$ et $n_{tot} = 247$. Notre procédure de sélection a conduit à un modèle à trois effets aléatoires indépendants :

$$\begin{aligned}
 \log CL_i &= \mu_{CL} + \beta_{CL,c} ClCr_i + \beta_{CL,a} age_i + \eta_{CL,i}, \\
 \log V_{1,i} &= \mu_{V1} + \beta_{V1,w} w_i + \beta_{V1,P} P/F_i + \eta_{V1,i}, \\
 \log Q_i &= \mu_Q + \beta_{Q,s} sex_i, \\
 \log V_{2,i} &= \mu_{V2} + \beta_{V2,c} ClCr_i + \eta_{V2,i}.
 \end{aligned} \tag{1.21}$$

Le tableau 1.2 résume les covariables (colonnes *ClCr* à *P/F*) et effets aléatoires (colonne η) retenus avec notre approche (symbole ★) et l'étude de [Burdet C. et al., 2015] (symbole ●). Avec notre procédure, nous retenons donc un modèle avec plus de covariables et moins d'effets aléatoires que dans [Burdet C. et al., 2015]. En particulier, la sélection de la covariable *sex* sur le paramètre *Q* permet d'expliquer la variabilité inter-sujets sur *Q* sans effet aléatoire. Ce résultat est intéressant dans la mesure où les effets aléatoires sont supposés décrire la partie de la variabilité inter-individuelle qui n'est pas reliée aux covariables observées. Nous comparons ensuite le modèle retenu avec notre approche (\mathcal{M}_1), et le modèle de [Burdet C. et al., 2015] (\mathcal{M}_2) avec un modèle intermédiaire (\mathcal{M}_3) qui combine la structure des effets aléatoires de \mathcal{M}_1 et les trois covariables de \mathcal{M}_2 . Comme ces trois modèles diffèrent à la fois par leur structure de covariables et d'effets aléatoires, la comparaison utilise un BIC classique dont la pénalité est donnée par $D \times \log(N)$ où $D = \dim(\{\mu, \beta\}) + \text{card}(S_\Omega)$. Les valeurs du BIC pour les trois modèles sont données dans le tableau 1.3. Notre modèle (\mathcal{M}_1) a la plus petite valeur de BIC. Sans prétendre avoir identifié le meilleur *modèle de population*, cet exemple illustre la capacité de notre procédure à trouver un compromis intéressant entre covariables et effets aléatoires dans la description de la variabilité inter-individuelle.

2.1 Orientations de recherche

Dans ce chapitre, je décris mes travaux relatifs aux modèles à effets mixtes définis par des processus de diffusion. Historiquement, l'étude de ces modèles a été motivée par des applications en pharmacocinétique et le besoin d'améliorer la modélisation de la variabilité intrinsèque à la distribution du médicament dans l'organisme (voir par exemple Overgaard R. V. et al. [2005], Tornøe C. W. et al. [2005]). Les modèles de diffusion à effets aléatoires suscitent un intérêt croissant en sciences du vivant et trouvent maintenant des applications en neurobiologie [Dion C., 2016] ou encore en modélisation de données de croissance [Strathe A. B. et al., 2011].

Le fil conducteur de mes travaux sur les modèles à effets mixtes définis par des processus de diffusion est celle de l'estimation des paramètres. Je me suis d'abord intéressée à cette question pendant ma thèse, dans le cadre d'applications en pharmacocinétique. Après avoir proposé une approche de modélisation alternative à celles de la littérature, j'ai développé une version spécifique de l'algorithme SAEM qui repose sur des opérations de marginalisation analogues à celles présentées dans le chapitre précédent pour les modèles de Markov cachés à effets mixtes [A5]. Alors que de nombreux autres algorithmes sont proposés dans la littérature (voir par exemple la revue proposée par Donnet S. et Samson A. [2013]), la question de l'estimation des paramètres dans ce type de modèle n'est quasiment jamais abordée sous un angle théorique, excepté dans [Ditlevsen S. et De Gaetano A., 2005] où les auteurs fournissent un résultat spécifique au mouvement brownien avec dérive aléatoire et non généralisable à d'autres modèles. Je me suis donc spécifiquement intéressée au comportement asymptotique des estimateurs des paramètres dans des classes plus générales de modèles de diffusions à effets mixtes. Je ne présente dans ce chapitre que ces études asymptotiques, à savoir [A4], [A8], [A10], [A13] et [A14], qui composent actuellement l'essentiel de mes contributions sur ces modèles. Soucieuse de proposer des outils utilisables par la communauté, j'ai également implémenté les contrastes proposés dans le package `R MsdeParEst`.

Dans ce chapitre, les diffusions sont observées sans bruit de mesure. Je commence par présenter dans la section 2.2 le modèle général considéré dans ces études asymptotiques et les difficultés liées à l'inférence. La section 2.3 réunit ensuite les résultats publiés dans [A4] et [A10]. Dans cette section, les diffusions individuelles sont observées en temps continu. Nous étudions d'abord dans [A4] la consistance et la normalité asymptotique de l'estimateur du maximum de vraisemblance puis nous proposons dans [A10] une extension originale de cette étude à la classification de données fonctionnelles. Cette approche ne

permet néanmoins de considérer que des effets aléatoires dans la dérive. La section 2.4 propose une synthèse de [A8], [A13] et [A14] dont le point commun est de considérer des estimateurs obtenus par optimisation de contrastes construits à partir du schéma d'Euler des diffusions individuelles. Les observations sont discrétisées et différents cadres asymptotiques sont considérés pour étudier les estimateurs selon que la dérive ou le coefficient de diffusion ou les deux simultanément dépendent d'effets aléatoires.

2.2 Modèle général et notations

Soient N processus stochastiques à valeurs réelles $(X_i(t), t \geq 0)$, $i = 1, \dots, N$, dont la dynamique est décrite par les équations différentielles stochastiques :

$$dX_i(t) = b(X_i(t), \phi_i)dt + \sigma(X_i(t), \Gamma_i) dW_i(t), \quad X_i(0) = x, \quad i = 1, \dots, N, \quad (2.1)$$

où (W_1, \dots, W_N) sont N mouvements Browniens indépendants, $\psi_i = (\phi_i, \Gamma_i)$, $i = 1, \dots, N$, sont N effets aléatoires *i.i.d.* dans $\mathbb{R}^d \times (0, +\infty)$, $(\psi_i)_{i=1, \dots, N}$ et $\{W_i\}_{i=1, \dots, N}$ sont mutuellement indépendants. La condition initiale x est une valeur réelle connue, supposée commune aux N processus. Le coefficient de diffusion $\sigma(\cdot, \gamma) : \mathbb{R} \rightarrow \mathbb{R}$ et la fonction de dérive $b(\cdot, \varphi) : \mathbb{R} \rightarrow \mathbb{R}$ sont des fonctions connues aux paramètres γ et φ près. Chaque processus $X_i(t)$ décrit l'évolution d'un individu au cours du temps et ψ_i est l'ensemble des paramètres régissant l'équation différentielle stochastique de l'individu i . Les variables aléatoires ψ_i sont de même loi, de densité $g(\psi, \theta)$ sur $\mathbb{R}^d \times (0, +\infty)$ où θ est un paramètre inconnu dans un ensemble Θ .

J'aborde dans ce chapitre la question de l'estimation de θ . Les observations sont des observations non bruitées des trajectoires $X_i(t)$, réalisations de (2.1), obtenues sur un intervalle de temps fixe $[0, T]$, $T > 0$, alors que les paramètres individuels ψ_i sont des données latentes. La vraisemblance des observations s'obtient comme dans le chapitre 1, en utilisant l'indépendance des N individus et en intégrant les densités conditionnelles des observations individuelles par rapport à la distribution des effets aléatoires. Notons $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_N)$ l'ensemble des observations disponibles sur les N trajectoires. L'estimateur du maximum de vraisemblance de θ , noté $\hat{\theta}_N$, est ainsi défini selon

$$\hat{\theta}_N = \arg \max_{\theta \in \Theta} L(\mathbf{X}; \theta),$$

où

$$L(\mathbf{X}; \theta) := L_N(\mathbf{X}; \theta) = \prod_{i=1}^N L(\mathbf{X}_i; \theta) = \prod_{i=1}^N \int p(\mathbf{X}_i | \psi) g(\psi, \theta) d\psi. \quad (2.2)$$

L'expression de $p(\mathbf{X}_i | \psi)$ dépend de la nature des observations.

Comme dans les modèles considérés dans le chapitre précédent, le calcul de $\hat{\theta}_N$, et par conséquent l'étude théorique de ses propriétés asymptotiques, sont des problèmes délicats car ni la solution de (2.1) ni l'expression de $L(\mathbf{X}; \theta)$ ne sont explicites en général. A ma connaissance, seuls les travaux de Nie L. et Yang M. [2005] et Nie L. [2006, 2007] abordent l'estimation des paramètres par maximum de vraisemblance sous un angle théorique dans les modèles non linéaires à effets mixtes. Cependant, les résultats de ces études reposent sur des hypothèses difficiles à vérifier en pratique, en particulier dans les modèles de diffusion à effets mixtes, et leur généralisation au modèle (2.1) est équivoque.

Dans [A4], [A8], [A10], [A13] et [A14], nous construisons des contrastes dérivés de la vraisemblance et étudions le comportement asymptotique des estimateurs associés à ces

contrastes avec des techniques classiques de statistique des processus. Nous faisons des hypothèses spécifiques sur la fonction de dérive et la distribution des effets aléatoires :

- 1) La fonction de dérive est linéaire en les effets aléatoires $b(x, \varphi) = \varphi^\top b(x)$, $b(\cdot) = (b_1(\cdot), \dots, b_d(\cdot))^\top : \mathbb{R} \rightarrow \mathbb{R}^d$, et le coefficient de diffusion est multiplicatif de la forme $\sigma(x, \gamma) = (\gamma)^{-1/2} \sigma(x)$, $\sigma(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$. Le modèle (2.1) devient alors :

$$dX_i(t) = \phi_i^\top b(X_i(t)) dt + (\Gamma_i)^{-1/2} \sigma(X_i(t)) dW_i(t), \quad X_i(0) = x, \quad i = 1, \dots, N. \quad (2.3)$$

- 2) Les effets aléatoires de la dérive (ϕ_1, \dots, ϕ_N) , sont gaussiens ou conditionnellement gaussiens, et les effets aléatoires du coefficient de diffusion $(\Gamma_1, \dots, \Gamma_N)$, sont de loi Gamma.

Bien que simplificatrices, ces hypothèses définissent une classe relativement générale de diffusions et auront l'avantage de permettre des calculs explicites. Dans la suite, la notation θ_0 désignera la vraie valeur du paramètre θ . Par souci de simplicité, je me contente dans ce chapitre de présenter les résultats lorsque φ ou ϕ_i sont scalaires ($d = 1$). Je renvoie aux articles où ils sont également énoncés et étudiés dans le cas où ϕ_i est multidimensionnel ($d > 1$); les résultats sont analogues.

2.3 Diffusions observées en temps continu

Cette partie regroupe les résultats de [A4] et [A10] issus d'une collaboration avec Valentine Genon-Catalot et Adeline Samson. Nous y considérons des observations continues des N diffusions individuelles sur un intervalle de temps fixe. Bien que peu réaliste en pratique, ce type d'observation est habituel en statistique des diffusions et permet l'étude théorique de l'estimateur du maximum de vraisemblance exact. Les densités conditionnelles $p(\mathbf{X}_i | \psi_i)$ s'expriment à partir de la formule de Girsanov, qui nécessite de supposer les paramètres du coefficient de diffusion connus. Dans cette partie, nous considérons donc le modèle (2.3) dans le cas particulier où $\Gamma_i = \gamma$ pour tout $i = 1, \dots, N$ avec $\gamma \in \mathbb{R}$ connu :

$$dX_i(t) = \phi_i b(X_i(t)) dt + \gamma^{-1/2} \sigma(X_i(t)) dW_i(t), \quad X_i(0) = x, \quad i = 1, \dots, N.$$

Pour simplifier les notations, nous désignerons par $\sigma_\gamma(\cdot)$ le coefficient de diffusion $\sigma_\gamma(\cdot) = \gamma^{-1/2} \sigma(\cdot)$, et par $p_\gamma(\cdot | \varphi)$ la vraisemblance conditionnelle aux effets dans la dérive $p_\gamma(\cdot | \varphi) = p(\cdot | \psi = (\varphi, \gamma))$.

2.3.1 Effets aléatoires de loi gaussienne dans la dérive

2.3.1.1 Vraisemblance exacte

Existence. Pour aborder la question de l'estimation, il faut montrer au préalable l'existence de solutions au système d'équations différentielles stochastiques (2.3). Les hypothèses standards ne suffisent pas ici en raison des effets aléatoires ϕ_i . Nous considérons la filtration $\mathcal{F}_t = \sigma(\phi_i, W_i(s), s \leq t, i = 1, \dots, N), t \geq 0$ et les hypothèses :

(H1) (i) La fonction $(x, \varphi) \rightarrow b(x, \varphi)$ est C^1 sur $\mathbb{R} \times \mathbb{R}$, et telle que :

$$\exists K > 0, \forall (x, \varphi) \in \mathbb{R} \times \mathbb{R}, \quad b^2(x, \varphi) \leq K(1 + x^2 + |\varphi|^2),$$

(ii) La fonction $\sigma(\cdot)$ est C^1 sur \mathbb{R} et

$$\forall x \in \mathbb{R}, \quad \sigma^2(x) \leq K(1 + x^2).$$

Sous **(H1)**, les équations différentielles stochastiques considérées admettent pour tout $i = 1, \dots, N$ une unique solution forte $(X_i(t), t \geq 0)$ adaptée à la filtration $(\mathcal{F}_t, t \geq 0)$ et les processus $(X_i(t), t \geq 0), i = 1, \dots, N$ sont indépendants. De plus, pour tout $k \geq 1$,

$\mathbb{E}|\phi_i|^{2k} < \infty$, et pour tout $T > 0$, $\sup_{t \in [0, T]} \mathbb{E}[X_i(t)]^{2k} < \infty$. Notons que si $b(x, \varphi) = \varphi b(x)$, l'hypothèse **(H1)** n'est pas vraie, mais l'existence et l'unicité des solutions s'obtiennent par étude directe.

Vraisemblance. La vraisemblance des observations continues $X_i = (X_i(t), t \in [0, T])$, $i = 1, \dots, N$ conditionnellement aux effets aléatoires est donnée par la formule de Girsanov en supposant que pour tout $i = 1, \dots, N$, $\int_0^T b^2(X_i(t), \varphi) / \sigma_\gamma^2(X(t)) dt < +\infty$:

$$p_\gamma(X_i | \varphi) = \exp \left(\int_0^{T_i} \frac{b(X_i(s), \varphi)}{\sigma_\gamma^2(X_i(s))} dX_i(s) - \frac{1}{2} \int_0^{T_i} \frac{b^2(X_i(s), \varphi)}{\sigma_\gamma^2(X_i(s))} ds \right).$$

On en déduit, d'après (2.2), la vraisemblance de X_i :

$$L(X_i, \theta) = \int \exp \left(\int_0^{T_i} \frac{b(X_i(s), \varphi)}{\sigma_\gamma^2(X_i(s))} dX_i(s) - \frac{1}{2} \int_0^{T_i} \frac{b^2(X_i(s), \varphi)}{\sigma_\gamma^2(X_i(s))} ds \right) g(\varphi; \theta) d\varphi. \quad (2.4)$$

(2.4) n'a pas toujours de forme explicite. L'hypothèse de linéarité de la dérive en les paramètres, *i.e.* $b(x, \varphi) = \varphi b(x)$, permet d'exprimer $p_\gamma(X_i | \varphi)$ en fonction des statistiques

$$U_i := U_i(T) = \int_0^T \frac{b(X_i(s))}{\sigma_\gamma^2(X_i(s))} dX_i(s), \quad V_i := V_i(T) = \int_0^T \frac{b^2(X_i(s))}{\sigma_\gamma^2(X_i(s))} ds. \quad (2.5)$$

On obtient

$$L(X_i, \theta) = \int \exp \left(\varphi U_i - \frac{1}{2} \varphi^2 V_i \right) g(\varphi; \theta) d\varphi. \quad (2.6)$$

Si (ϕ_1, \dots, ϕ_N) sont *i.i.d.* de densité $g(\varphi, \theta)$ gaussienne sur \mathbb{R} :

$$g(\varphi, \theta) d\varphi = \mathcal{N}(\mu, \omega^2),$$

les paramètres à estimer sont $\theta = (\mu, \omega^2)$ et $L(X_i, \theta)$ est explicite :

$$L(X_i, \theta) = \frac{1}{\sqrt{1 + \omega^2 V_i}} \exp \left[-\frac{1}{2} \left(\mu - \frac{U_i}{V_i} \right)^2 \frac{V_i}{1 + \omega^2 V_i} \right] \exp \left(\frac{1}{2} \frac{U_i^2}{V_i} \right).$$

Par indépendance entre les N processus individuels, il vient

$$L_N(\theta) = \prod_{i=1}^N L(X_i, \theta).$$

Score. La fonction de score est donnée par

$$\frac{\partial}{\partial \mu} \log L_N(\theta) = \sum_{i=1}^N \delta_i(\theta), \quad \frac{\partial}{\partial \omega^2} \log L_N(\theta) = \frac{1}{2} \sum_{i=1}^N (\delta_i(\theta)^2 - I_i(\omega^2)), \quad (2.7)$$

où

$$\delta_i(\theta) = \frac{U_i - V_i \mu}{1 + \omega^2 V_i} \quad \text{et} \quad I_i(\omega^2) = \frac{V_i}{1 + \omega^2 V_i}.$$

2.3.1.2 Résultats de consistance et normalité asymptotique

Pour étudier le comportement asymptotique de l'estimateur du maximum de vraisemblance $\hat{\theta}_N$, les hypothèses suivantes sont nécessaires :

- (H2) La fonction $b(\cdot)/\sigma_\gamma(\cdot)$ n'est pas constante. Sous la distribution Q du processus canonique, la variable aléatoire (U_1, V_1) admet une densité $f(u, v)$ par rapport à la mesure de Lebesgues sur $\mathbb{R} \times (0, +\infty)$ continue et positive sur un ouvert de $\mathbb{R} \times (0, +\infty)$.
- (H3) L'espace des paramètres Θ est un compact de $\mathbb{R} \times (0, +\infty)$.
- (H4) La vraie valeur θ_0 appartient à $\overset{\circ}{\Theta}$.
- (H5) La matrice d'information de Fisher $\mathcal{I}(\theta_0)$ est inversible, où pour tout θ de Θ , $\mathcal{I}(\theta)$ est la matrice de covariance du vecteur

$$\begin{pmatrix} \delta_1(\theta) \\ \frac{1}{2}(\delta_1^2(\theta) - I_1(\theta)) \end{pmatrix}.$$

Vérifier (H2) fait appel à des outils de calcul de Malliavin et nécessite certaines hypothèses de régularité sur $b(\cdot)$ et $\sigma_\gamma(\cdot)$. Cette hypothèse est nécessaire pour démontrer l'identifiabilité des paramètres. Les hypothèses (H3) à (H5) sont des hypothèses classiques dans l'étude asymptotique de l'EMV.

Théorème 3 (Identifiabilité) Notons $K(Q_{\theta_0}^1, Q_\theta^1)$ la divergence de Kullback de $Q_{\theta_0}^1$ par rapport à Q_θ^1 .

- (i) Sous (H1)-(H2), $Q_\theta^1 = Q_{\theta_0}^1$ implique que $\theta = \theta_0$. Donc, $\theta \rightarrow K(Q_{\theta_0}^1, Q_\theta^1)$ admet un unique minimum en $\theta = \theta_0$.
- (ii) Sous (H1), la fonction $\theta \rightarrow K(Q_{\theta_0}^1, Q_\theta^1)$ est continue sur $\mathbb{R} \times (0, +\infty)$.

Théorème 4 (Consistance et normalité asymptotique)

1. Supposons les hypothèses (H1) à (H3) vraies. Sous Q_{θ_0} , $\hat{\theta}_N$ converge en probabilité vers θ_0 .
2. Supposons les hypothèses (H1) à (H5) vraies. L'estimateur du maximum de vraisemblance est asymptotiquement gaussien lorsque $N \rightarrow +\infty$:

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \rightarrow_{\mathcal{D}} \mathcal{N}(0, \mathcal{I}^{-1}(\theta_0)).$$

Les schémas de preuve sont classiques. La consistance de $\hat{\theta}_N$ s'obtient en étudiant le module de continuité de la log-vraisemblance (cf van der Vaart A.W. [2000]). La preuve de la normalité asymptotique de $\hat{\theta}_N$ utilise le développement de Taylor du score autour de $\hat{\theta}_N$ et découle de la convergence en loi du score normalisé, de la convergence en probabilité de la matrice d'information de Fisher observée et de la gestion des termes de reste. Un point important est l'étude des statistiques $\delta_i(\theta)$ et $I_i(\omega^2)$ et de leurs moments.

On peut remarquer que contrairement au cadre plus usuel des diffusions à effets fixes où l'on n'observerait qu'une seule trajectoire, il n'est pas nécessaire de se placer en temps long et de faire tendre T vers $+\infty$ pour que les paramètres de la dérive soient estimés de façon consistante.

2.3.1.3 Discrétisation haute fréquence des observations

Comme la fonction de score est explicite (cf équation (2.7)), il est possible de calculer $\hat{\theta}_N$ à partir des observations, au moins numériquement. Néanmoins, le cas d'observations continues est peu réaliste en pratique. On dispose plutôt d'observations discrètes des processus $X_{i,n} = X_i = (X_i(t_j), j = 1, \dots, n)$ faites en des temps déterministes

$0 < t_1 < t_2 < \dots < t_n = T$. Pour calculer effectivement l'estimateur du maximum de vraisemblance, il faut donc discrétiser les intégrales $U_i(T)$ et $V_i(T)$ qui permettent de définir les équations de vraisemblance :

$$U_i^n = \sum_{k=0}^{n-1} \frac{b(X_i(t_k))}{\sigma_\gamma^2(X_i(t_k))} (X_i(t_{k+1}) - X_i(t_k)), \quad (2.8)$$

$$V_i^n = \sum_{k=0}^{n-1} \frac{b^2(X_i(t_k))}{\sigma_\gamma^2(X_i(t_k))} (t_{k+1} - t_k). \quad (2.9)$$

Notons $\hat{\theta}_N^{(n)}$ la version discrétisée de $\hat{\theta}_N$ ainsi obtenue. Nous vérifions les garanties de cette approche lorsque le nombre d'observations n tend vers l'infini. Pour simplifier les notations, nous supposons que les N processus sont observés de manière synchrone à pas de temps réguliers $t_k^n = t_k = k\frac{T}{n}$, $k = 0, 1, \dots, n$. Pour étudier $\hat{\theta}_N^{(n)}$, il suffit d'étudier les différences $U_i(T) - U_i^n$, $V_i(T) - V_i^n$.

Lemme 1 *Supposons que $b(\cdot)/\sigma(\cdot)$ est borné et Lipschitz, que $\sigma(\cdot) \geq \epsilon > 0$, que $b(\cdot)$ et $\sigma(\cdot)$ sont Lipschitz, alors pour tout $p \geq 1$ et pour tout $i = 1, \dots, N$, il existe une constante C telle que*

$$\mathbb{E}_{\theta_0} (|V_i(T) - V_i^n|^p + |U_i(T) - U_i^n|^p) \leq \frac{C}{n^{p/2}}.$$

On peut alors montrer le résultat suivant :

Proposition 1 *Si $n \rightarrow +\infty$, alors $\hat{\theta}_N - \hat{\theta}_N^{(n)} = o_{P_{\theta_0}}(1)$. Si $n = n(N) \rightarrow +\infty$ tel que $\frac{n}{N} \rightarrow +\infty$, alors $\sqrt{N}(\hat{\theta}_N - \hat{\theta}_N^{(n)}) = o_{P_{\theta_0}}(1)$.*

2.3.1.4 Expériences numériques

Nous avons mené une étude de simulation dans [A4] qui illustre les bonnes propriétés de $\hat{\theta}_N$ à partir de son calcul discrétisé obtenu avec un pas de discrétisation suffisamment petit. Nous avons considéré différentes fonctions de dérive $b(\cdot)$ et différents coefficients de diffusion $\sigma_\gamma(\cdot)$. Le biais et la variance empirique de l'estimateur diminuent lorsque N augmente. Les simulations montrent également que la valeur de T qui définit la longueur de l'intervalle d'observation ne semble pas jouer sur la qualité des estimations.

2.3.2 Effets aléatoires de loi mélange et clustering de courbes

2.3.2.1 Contexte

[A10] fait suite à [A4]. Nous y proposons une extension naturelle et originale des résultats précédents à la classification non supervisée de courbes. Proposer une typologie d'individus à partir de mesures d'une ou plusieurs variable(s) est une question statistique régulièrement considérée dans de nombreux cadres applicatifs. Depuis quelques années, avec l'accroissement des capacités de stockage, il est devenu important de définir des classes d'individus à partir de données longitudinales. De nombreuses méthodes ont été considérées pour ce problème dont les mélanges de modèles à effets mixtes. La variabilité inter-individuelle y est structurée en groupes homogènes d'individus en définissant les effets aléatoires du modèle comme des variables aléatoires de distribution mélange. Des mélanges de modèles linéaires à effets mixtes ont par exemple été considérés dans [Celeux G. et al., 2005], [Luts J. et al., 2012] et [Proust-Lima C. et Jacqmin-Gadda H., 2005] pour faire du clustering de gènes, de données de surveillance épidémiologique ou encore de données de croissance. Lavielle M. et Mbogning C. [2014] ont étendu cette approche à des mélanges de modèles non linéaires à effets mixtes pour des applications en pharmacologie,

les composantes des mélanges permettant de définir différentes catégories de répondants à un traitement à partir de données pharmacocinétiques.

En suivant la même idée, nous considérons des données fonctionnelles observées de façon continue sur un intervalle de temps fixe $[0, T]$ et modélisées par une équation différentielle stochastique à effets aléatoires définie par (2.3)¹. L'originalité par rapport aux travaux précédemment cités est de baser la classification sur un modèle défini en temps continu. Nous considérons un mélange de distributions gaussiennes pour les effets aléatoires. Le modèle étudié est donc le suivant :

$$\begin{aligned} dX_i(t) &= \phi_i b(X_i(t))dt + \sigma_\gamma(X_i(t)) dW_i(t), \quad X_i(0) = x, \quad i = 1, \dots, N, \\ \phi_i &\sim g(\varphi, \theta)d\varphi, \end{aligned}$$

où

$$g(\varphi, \theta) = \sum_{k=1}^K \pi_k n(\varphi, \tau_k), \quad n(\varphi, \tau_k)d\varphi = \mathcal{N}(\mu_k, \omega_k^2), \quad \tau_k = (\mu_k, \omega_k^2).$$

K est le nombre de composantes du mélange, les proportions du mélange $(\pi_k)_{k=1, \dots, K}$ vérifient $\sum_{k=1}^K \pi_k = 1$, μ_k et ω_k^2 sont la moyenne et la variance de la k -ième gaussienne du mélange. Les composantes du mélange représentent ici des classes d'individus dont les trajectoires sont *i.i.d.*, c'est-à-dire des classes d'individus dont les comportements présentent des similitudes au cours du temps. Le nombre de composantes du mélange K est inconnu et doit être estimé, au même titre que les paramètres $\tau_k = (\mu_k, \omega_k^2)$ de chaque composante du mélange ainsi que leurs proportions π_k . Comme dans la section 2.3.1, le paramètre du coefficient de diffusion γ est connu. Ainsi, lorsque K est connu, $\theta = \{(\pi_k, k = 1, \dots, K-1), (\tau_k, k = 1, \dots, K)\}$.

2.3.2.2 Propriétés asymptotiques

On garantit l'existence de solutions grâce à l'hypothèse suivante qui est classique :

(H1b) La fonction de dérive $b(\cdot)$ est lipschitzienne sur \mathbb{R} et la fonction $\sigma(\cdot)$ est α -höldérienne sur \mathbb{R} de coefficient $\alpha \in [1/2, 1]$.

Le choix de considérer un mélange gaussien pour les effets aléatoires permet, comme dans la section 2.3.1, d'obtenir une forme explicite pour la vraisemblance. Pour exprimer la vraisemblance de X_i , on procède de la même façon. On obtient d'abord la vraisemblance de X_i conditionnellement à $\phi_i = \varphi$ grâce au théorème de Girsanov, puis on intègre cette densité conditionnelle par rapport à la distribution mélange de ϕ_i . Ces étapes font apparaître les mêmes statistiques U_i et V_i (équation (2.5)). Par analogie avec le cas purement gaussien de la section 2.3.1, il vient :

$$L(X_i, \theta) = \sum_{k=1}^K \pi_k \lambda(X_i, \tau_k),$$

où

$$\lambda(X_i, \tau_k) = \frac{1}{\sqrt{\det(1 + V_i \omega_k^2)}} \exp \left[-\frac{1}{2} \left(\mu_k - \frac{U_i}{V_i} \right)^2 \frac{V_i}{1 + \omega_k^2 V_i} \right] \exp \left(\frac{1}{2} \frac{U_i^2}{V_i} \right).$$

Comme dans la section 2.3.1, la fonction de score est explicite. Sous réserve que l'hypothèse **(H2)** introduite dans la section 2.3.1 soit vérifiée, nous montrons que le modèle est identifiable. Nous montrons également la consistance faible de $\hat{\theta}_N$.

1. La fonction de dérive considérée dans [A10] est en réalité plus générale, de la forme $b(x, \varphi) = \varphi^\top b(x) + a(x)$ où $a(\cdot)$ est une fonction connue. Je restreins cette présentation au cas $a(\cdot) \equiv 0$. Les résultats sont analogues dans le cas plus général.

2.3.2.3 Algorithme d'estimation

L'estimateur du maximum de vraisemblance est bien défini et possède des propriétés théoriques qui en justifient l'utilisation sur des données. En pratique, son calcul numérique nécessite comme précédemment de discrétiser les intégrales U_i et V_i mais n'est pas direct pour autant. En effet, K est inconnu en pratique, de même que les compositions des K groupes. Nous proposons une méthodologie classique pour estimer les paramètres du modèle, estimer le nombre de composantes du mélange et classer les individus.

Estimation des paramètres. Lorsque le nombre de classes K est connu, nous proposons d'utiliser l'algorithme EM pour calculer l'estimateur du maximum de vraisemblance. Cet algorithme est couramment utilisé pour l'estimation des paramètres dans les modèles de mélange. Dans le modèle que nous considérons, la quantité Q sur laquelle repose l'algorithme est explicite. A l'itération m , pour l'estimation courante $\hat{\theta}^{(m)}$ du paramètre, l'algorithme comporte deux étapes :

(a) *Etape E* : calculer

$$Q(\theta, \hat{\theta}^{(m)}) = \sum_{i=1}^N \sum_{k=1}^K \tilde{\pi}_k(X_i, \hat{\theta}^{(m)}) \log(\pi_k \lambda(X_i, \tau_k)),$$

où $\tilde{\pi}_k(X_i, \theta')$ est la probabilité *a posteriori* :

$$\tilde{\pi}_k(X_i, \theta') := \mathbb{P}(Z_i = k | X_i, \theta') = \frac{\pi'_k \lambda(X_i, \tau'_k)}{L(X_i, \theta')}, \quad (2.10)$$

$Z_i \in \{1, \dots, K\}$ étant la composante du mélange régissant l'équation du sujet i .

(b) *Etape M* : maximiser

$$\hat{\theta}^{(m+1)} = \arg \max_{\theta \in \Theta} Q(\theta, \hat{\theta}^{(m)}).$$

En suivant la preuve proposée dans McLachlan G. et Krishnan T. [1996], nous montrons que les estimations $(\hat{\theta}^{(m)})$ obtenues au fil des itérations de l'algorithme EM convergent vers un point stationnaire de la vraisemblance.

Estimation du nombre de composantes du mélange. Lorsque K n'est pas connu, nous sélectionnons le nombre de composantes du mélange par minimisation d'un critère BIC

$$BIC(K, \theta(K)) = -2\ell(\theta(K)) + \alpha(K) \log(K),$$

où $\alpha(K)$ est le nombre total de paramètres à estimer pour K composantes, $\theta(K)$ et $\ell(\theta(K))$ sont le paramètre et la log-vraisemblance associés. Les propriétés théoriques du BIC pour la sélection du nombre de composantes d'un mélange fini sont connues. Keribin C. [2000] a montré en particulier que le nombre de composantes sélectionné par BIC converge presque sûrement vers le vrai nombre de composantes lorsque $N \rightarrow +\infty$.

Règle de classification. Notons \hat{K} le nombre de composantes estimé et $\hat{\theta}$ l'estimation des paramètres obtenue à convergence de l'algorithme EM pour \hat{K} composantes. Pour prédire la classe d'appartenance de l'individu i , nous estimons les probabilités *a posteriori* $\tilde{\pi}_k(X_i, \theta)$ selon (2.10) et classons i dans le cluster k si

$$\tilde{\pi}_k(X_i, \hat{\theta}) = \arg \max_{1 \leq k' \leq \hat{K}} \tilde{\pi}_{k'}(X_i, \hat{\theta}). \quad (2.11)$$

2.3.2.4 Application

Cette procédure est utilisée sur des données simulées dans plusieurs modèles où les effets aléatoires sont distribués selon un mélange de 3 gaussiennes. Plusieurs situations sont considérées où les composantes du mélange sont bien séparées, moins bien séparées et très mal séparées. Lorsque les composantes du mélange sont bien séparées les résultats illustrent la consistance de l'estimateur du maximum de vraisemblance et la consistance du BIC pour le choix du nombre de composantes. Les résultats se détériorent (estimations biaisées, sous-estimation du nombre de composantes) lorsque les composantes du mélange sont peu séparables. Notre étude montre qu'appliquer la même méthodologie sur les réalisations des effets aléatoires gaussiens ϕ_i plutôt que sur les observations des trajectoires donne des résultats très proches. Nous utilisons cette approche pour décrire des données de croissance de poulets. Le BIC nous conduit à choisir 4 groupes. La Table 2.1 montre que la classification obtenue est fortement liée aux lignées génétiques représentées dans les données.

Classes estimées	Lignées génétiques				Total
	LH	LL	HL	HH	
1	17	13	0	0	30
2	14	40	5	1	60
3	0	2	26	9	37
4	0	0	4	22	26
Total	31	55	35	32	153

TABLE 2.1 – Données de croissance de poulets : comparaison des classes estimées et des lignées génétiques.

2.4 Diffusions observées en des temps discrets

2.4.1 Contrastes dérivés du schéma d'Euler

La vraisemblance exacte dérivée de la formule de Girsanov ne permet pas d'estimer un paramètre inconnu dans le coefficient de diffusion. Pour étudier les situations plus générales où le paramètre de diffusion est inconnu, fixe ou aléatoire, il est plus naturel de considérer des observations discrétisées des processus. Ce point de vue est par ailleurs plus réaliste que celui d'observations continues. Nous supposons donc que les trajectoires $X_i(t)$ sont observées en des instants déterministes sur un intervalle de temps fixe $[0, T]$: $0 = t_{i1} < t_{i2} < \dots < t_{in} = T$. Ainsi, les observations sont les n -uples $\mathbf{X}_i = (X_i(t_{ij}))_{j=1, \dots, n}$, $i = 1, \dots, N$. Pour simplifier les notations mais sans perte de généralité, je présenterai les résultats pour des observations synchrones et régulières avec un pas de discrétisation $\Delta = T/n$, c'est-à-dire $t_{ij} = t_j = j\Delta$. Dans ce cas, par la propriété de Markov du processus $(X_i(t))$, la vraisemblance conditionnelle des observations a pour expression :

$$p(\mathbf{X}_i | \psi) := p(\mathbf{X}_i | \psi_i = \psi, \theta) = p(X_i(t_1); \psi) \prod_{j=1}^{n-1} p(X_i(t_{j+1}) | X_i(t_j); \psi). \quad (2.12)$$

Les densités de transition $p(X_i(t_{j+1}) | X_i(t_j); \psi)$ ne sont pas explicites en général, rendant impossible l'estimation des paramètres de (2.3) par maximum de vraisemblance exacte. Une solution classique dans les diffusions à effets fixes consiste à définir une pseudo-vraisemblance en approchant les équations différentielles par leur schéma d'Euler, ce qui

revient à approcher les densités de transition $p(X_i(t_{j+1})|X_i(t_j); \psi)$ par des densités gaussiennes :

$$\tilde{p}(\mathbf{X}_i|\psi) \propto \gamma^{n/2} \exp\left[-\frac{\gamma}{2}(S_{i,n} + \varphi^2 V_{i,n} - 2\varphi U_{i,n})\right], \quad (2.13)$$

où

$$S_{i,n} = \frac{1}{\Delta} \sum_{j=1}^n \frac{(X_i(t_j) - X_i(t_{j-1}))^2}{\sigma^2(X_i(t_{j-1}))}, \quad (2.14)$$

$$U_{i,n} = \sum_{j=1}^n \frac{b(X_i(t_{j-1}))(X_i(t_j) - X_i(t_{j-1}))}{\sigma^2(X_i(t_{j-1}))}, \quad (2.15)$$

$$V_{i,n} = \Delta \sum_{j=1}^n \frac{b^2(X_i(t_{j-1}))}{\sigma^2(X_i(t_{j-1}))}. \quad (2.16)$$

Cette approximation n'est valable que si le pas de discrétisation Δ est suffisamment petit, ce qui nous conduira à nous placer dans des cadres asymptotiques où $n \rightarrow +\infty$ et $N \rightarrow +\infty$ conjointement. Il est intéressant de noter que ces vraisemblances conditionnelles approchées s'expriment en fonction des discrétisées des statistiques U_i et V_i définies dans le cas continu (équation (2.5)). Le choix d'effets aléatoires de loi gaussienne dans la dérive et de loi Gamma dans le coefficient de diffusion permet de définir des contrastes explicites en intégrant les $\tilde{p}(\mathbf{X}_i|\psi)$ par rapport aux distributions des effets aléatoires :

$$\tilde{\ell}_{N,n}(\theta) = \sum_{i=1}^N \log \int \tilde{p}(\mathbf{X}_i|\psi) g(\psi, \theta) d\psi. \quad (2.17)$$

En collaboration avec Valentine Genon-Catalot, Catherine Larédo et Adeline Samson nous étudions dans [A8] et [A13] l'estimateur du maximum de pseudo-vraisemblance défini par :

$$G_{N,n}(\tilde{\theta}_{N,n}) = 0, \quad (2.18)$$

où $G_{N,n}(\theta)$ est le pseudo-score associé à $\tilde{\ell}_{N,n}(\theta)$:

$$G_{N,n}(\theta) = \nabla_{\theta} \tilde{\ell}_{N,n}(\theta).$$

Selon que la dérive ou le coefficient de diffusion dépendent d'effets aléatoires, les asymptotiques considérées et les vitesses de convergence obtenues sont différentes.

Lorsque les deux termes dépendent d'effets aléatoires, la pseudo-vraisemblance n'est pas toujours définie. Dans [A14], nous proposons deux contrastes alternatifs, tous deux dérivés de $\tilde{\ell}_{N,n}$, pour estimer les paramètres θ de la loi jointe de (ϕ_i, Γ_i) , et démontrons la consistance et la normalité asymptotique des estimateurs ainsi définis.

Dans la suite, les preuves ne sont pas détaillées. Pour étudier le comportement asymptotique de l'estimateur proposé, on utilise un schéma de preuve classique (voir par exemple Kessler M. et al. [2012]) reposant sur un développement de Taylor du pseudo-score. Les propriétés asymptotiques de l'estimateur découlent de la convergence en loi du pseudo-score convenablement normalisé, de la convergence en probabilité de la pseudo-matrice d'information de Fisher et de la gestion des termes de reste. L'étude des termes de reste est le point le plus technique. Il consiste en l'étude de moments de statistiques pivotales, celles-ci étant différentes selon les composantes aléatoires de ψ_i .

Dans la suite, nous faisons l'hypothèse suivante :

(A1) Les fonctions $b(\cdot)$ et $\sigma(\cdot)$ à valeurs réelles sont C^2 sur \mathbb{R} et leurs dérivées premières et secondes sont majorées par L . La fonction $\sigma(\cdot)$ est de plus minorée : $\exists \sigma_0 > 0, \forall x \in \mathbb{R}, \sigma(x) \geq \sigma_0$.

(A2) Il existe une constante K telle que, $\forall x \in \mathbb{R}, |b(x)| + |\sigma(x)| \leq K$.

Les hypothèses **(A1)** et **(A2)** garantissent l'existence pour tout $i = 1, \dots, N$ d'une unique solution forte $(X_i(t), t \geq 0)$ adaptée à la filtration $(\mathcal{F}_t, t \geq 0)$, où $\mathcal{F}_t = \sigma(\psi_i, W_i(s), s \leq t, i = 1, \dots, N), t \geq 0$, et que les processus $(X_i(t), t \geq 0), i = 1, \dots, N$ sont indépendants et identiquement distribués. L'hypothèse **(A2)** n'est pas classique. Usuellement, les fonctions $b(\cdot)$ et $\sigma(\cdot)$ sont supposées localement lipschitziennes et à croissance sous-linéaire. Ce n'est pas suffisant ici du fait que l'on a N trajectoires.

2.4.2 Effet aléatoire dans la dérive

Je présente les résultats obtenus dans [A13] lorsque $\Gamma_i = \gamma$ pour tout $i = 1, \dots, N$, où $\gamma > 0$, et $\phi_i \underset{i.i.d.}{\sim} \mathcal{N}(\mu, \gamma^{-1}\omega^2)$ dans (2.3). Le modèle considéré est donc

$$\begin{aligned} dX_i(t) &= \phi_i b(X_i(t))dt + \gamma^{-1/2} \sigma(X_i(t)) dW_i(t), \quad X_i(0) = x, \quad i = 1, \dots, N, \\ \phi_i &\underset{i.i.d.}{\sim} \mathcal{N}(\mu, \gamma^{-1}\omega^2). \end{aligned}$$

Les paramètres à estimer sont $\theta = (\gamma, \mu, \omega^2)$. Sous ces hypothèses, la pseudo-log-vraisemblance (2.17) a pour expression :

$$\begin{aligned} \tilde{\ell}_{N,n}(\theta) &= \frac{Nn}{2} \log \gamma - \frac{1}{2} \sum_{i=1}^N \log(1 + \omega^2 V_{i,n}) \\ &\quad - \frac{\gamma}{2} \sum_{i=1}^N \left[S_{i,n} + \left(\mu - \frac{U_{i,n}^2}{V_{i,n}} \right)^2 \left(\frac{V_{i,n}}{1 + \omega^2 V_{i,n}} \right) - \frac{U_{i,n}^2}{V_{i,n}} \right]. \end{aligned}$$

Ici, les composantes du pseudo-score $G_{N,n}(\theta) = \left(\frac{\partial}{\partial \gamma} \tilde{\ell}_{N,n}(\theta), \frac{\partial}{\partial \mu} \tilde{\ell}_{N,n}(\theta), \frac{\partial}{\partial \omega^2} \tilde{\ell}_{N,n}(\theta) \right)$ ont pour expression

$$\begin{aligned} \frac{\partial}{\partial \gamma} \tilde{\ell}_{N,n}(\theta) &= \frac{Nn}{2\gamma} - \frac{1}{2} \sum_{i=1}^N \left[S_{i,n} + \left(\mu - \frac{U_{i,n}^2}{V_{i,n}} \right)^2 \frac{V_{i,n}}{1 + \omega^2 V_{i,n}} - \frac{U_{i,n}^2}{V_{i,n}} \right], \\ \frac{\partial}{\partial \mu} \tilde{\ell}_{N,n}(\theta) &= \gamma \sum_{i=1}^N A_i, \\ \frac{\partial}{\partial \omega^2} \tilde{\ell}_{N,n}(\theta) &= -\frac{1}{2} \sum_{i=1}^N B_i + \frac{\gamma}{2} \sum_{i=1}^N A_i^2, \end{aligned}$$

où

$$\begin{aligned} B_i &= B_{i,n} = (V_{i,n}^{-1} + \omega^2)^{-1}, \\ A_i &= A_{i,n} = (1 + \omega^2 V_{i,n})^{-1} (U_{i,n} - V_{i,n} \mu) = B_i (V_{i,n}^{-1} U_{i,n} - \mu). \end{aligned}$$

L'une des clés est de montrer que les statistiques $B_{i,n}$ (resp. $A_{i,n}$) ont une limite, celles-ci étant données respectivement par

$$\begin{aligned} B_i(T; \omega^2) &= (V_i(T)^{-1} + \omega^2)^{-1}, \\ A_i(T; \mu, \omega^2) &= B_i(T; \omega^2) (V_i(T)^{-1} U_i(T) - \mu), \end{aligned}$$

où $U_i(T)$ et $V_i(T)$ sont les intégrales stochastiques définies en temps continu dans l'équation (2.5). On démontre ainsi la consistance et la normalité asymptotique de $\tilde{\theta}_{N,n}$. L'estimateur de γ d'une part et les estimateurs de μ, ω^2 ont des vitesses de convergence différentes.

Théorème 5 (Consistance et normalité asymptotique) *Supposons les hypothèses (A1)-(A2) vraies et $I(\theta_0)$ inversible, où pour tout $\theta \in \Theta$, $I(\theta)$ est la matrice de covariance de*

$$\begin{pmatrix} \gamma A_1(T; \mu, \omega^2) \\ \frac{1}{2}(\gamma A_1^2(T; \mu, \omega^2) - B_1(T; \omega^2)) \end{pmatrix}.$$

Alors, si N, n tendent conjointement vers l'infini tels que $N/n \rightarrow 0$, il existe une solution $\tilde{\theta}_{N,n}$ à (2.18) telle que

1. $\tilde{\theta}_{N,n}$ converge en probabilité vers θ_0 ,
2. $D_{N,n}^{-1}(\tilde{\theta}_{N,n} - \theta_0)$ converge en loi sous \mathbb{P}_{θ_0} vers $\mathcal{N}(0, \mathcal{I}^{-1}(\theta_0))$ où

$$D_{N,n} = \begin{pmatrix} \frac{1}{\sqrt{Nn}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{N}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{N}} \end{pmatrix}, \quad \mathcal{I}(\theta) = \begin{pmatrix} \frac{1}{2\gamma^2} & \boldsymbol{\theta} \\ \boldsymbol{\theta} & I(\theta) \end{pmatrix}.$$

Le cas particulier où γ est connu est traité dans la section 2.3 par maximum de vraisemblance exact lorsque les observations sont continues. Il est intéressant de noter que si γ est connu, la discrétisation de la vraisemblance (2.6) coïncide avec $\tilde{\ell}_{N,n}$. μ et ω^2 sont estimés avec les mêmes vitesses avec les deux approches.

2.4.3 Effet aléatoire dans le coefficient de diffusion

Je présente les résultats asymptotiques obtenus dans le cas où il y a des effets aléatoires dans le coefficient de diffusion mais pas dans la dérive. Ce cadre d'étude a fait l'objet de deux publications [A8] et [A13]. Dans la première ([A8]), nous avons considéré que la dérive était connue et avons proposé deux approches pour estimer les paramètres de la loi des effets aléatoires du coefficient de diffusion. Dans la seconde ([A13]), nous avons considéré un cas plus général où la dérive dépend d'effets fixes inconnus.

Cas général. Le modèle général s'écrit

$$\begin{aligned} dX_i(t) &= \varphi b(X_i(t))dt + \Gamma_i^{-1/2} \sigma(X_i(t)) dW_i(t), \quad X_i(0) = x, \quad i = 1, \dots, N, \\ \Gamma_i &\underset{i.i.d.}{\sim} G(a, \lambda). \end{aligned}$$

Les paramètres de la dérive sont déterministes et tous égaux à une même valeur inconnue φ ($\phi_i = \varphi$ pour tout $i = 1, \dots, N$) tandis que les paramètres du coefficient de diffusion sont aléatoires et distribués selon une loi Gamma. Les paramètres à estimer sont $\theta = (\lambda, a, \varphi)$. La pseudo-vraisemblance (2.17) a pour expression :

$$\begin{aligned} \tilde{\ell}_{N,n}(\theta) &= Na \log(\lambda) + N \log(\Gamma(a + n/2)) - N \log(\Gamma(a)) \\ &\quad - (a + n/2) \sum_{i=1}^N \log \left(\lambda + \frac{1}{2}(S_{i,n} - 2\varphi U_{i,n} + \varphi^2 V_{i,n}) \right). \end{aligned} \quad (2.19)$$

On vérifie que le terme $S_{i,n} - 2\varphi U_{i,n} + \varphi^2 V_{i,n}$ est toujours strictement positif :

$$S_{i,n} + \varphi^2 V_{i,n} - 2\varphi U_{i,n} = \sum_{j=1}^n \frac{(X_i(t_j) - X_i(t_{j-1}) - \Delta\varphi b(X_i(t_{j-1})))^2}{\Delta\sigma^2(X_i(t_{j-1}))},$$

ce qui garantit que la fonction $\tilde{\ell}_{N,n}(\theta)$ est bien définie. Le pseudo-score $G_{N,n}(\theta)$ est défini par :

$$\begin{aligned}\frac{\partial}{\partial \lambda} \tilde{\ell}_{N,n}(\theta) &= \sum_{i=1}^N \left(\frac{a}{\lambda} - \zeta_i^{-1} \right), \\ \frac{\partial}{\partial a} \tilde{\ell}_{N,n}(\theta) &= \sum_{i=1}^N (\log \lambda - \xi(a) - \log \zeta_i) + N (\xi(a + (n/2)) - \log(a + (n/2))), \\ \frac{\partial}{\partial \varphi} \tilde{\ell}_{N,n}(\theta) &= \sum_{i=1}^N \zeta_i^{-1} (U_{i,n} - \varphi V_{i,n}),\end{aligned}$$

où $\xi(z) = \Gamma'(z)/\Gamma(z)$ désigne la fonction Digamma et

$$\zeta_i(\tau) = \zeta_i = \frac{\lambda + \frac{1}{2} (S_{i,n} - 2\varphi U_{i,n} + \varphi^2 V_{i,n})}{a + (n/2)}. \quad (2.20)$$

Théorème 6 (Consistance et normalité asymptotique) *Supposons les hypothèses (A1)-(A2) vraies, $a > 5$ et N, n tendant vers l'infini tels que $N/n \rightarrow 0$. Alors il existe une solution $\tilde{\theta}_{N,n}$ à (2.18) telle que*

1. $\tilde{\theta}_{N,n}$ converge en probabilité vers θ_0 ,
2. $\sqrt{N}(\tilde{\theta}_{N,n} - \theta_0)$ converge en loi sous \mathbb{P}_{θ_0} vers $\mathcal{N}(0, \mathcal{V}^{-1}(\theta_0))$ où

$$\mathcal{V}(\theta) = \left(\begin{array}{c|c} I_0(\lambda, a) & \mathbf{0} \\ \hline \mathbf{0} & \mathbb{E}_{\theta_0}(\Gamma V(T)) \end{array} \right),$$

Γ désigne une variable aléatoire de loi Gamma $G(a, \lambda)$, $V(T)$ est définie par (2.5) et $I_0(\lambda, a)$ correspond à la matrice d'information de Fisher d'un modèle i.i.d. de distribution Gamma $G(a, \lambda)$

$$I_0(\lambda, a) = \left(\begin{array}{cc} \frac{a}{\lambda^2} & -\frac{1}{\lambda} \\ -\frac{1}{\lambda} & \xi'(a) \end{array} \right).$$

Pour les composantes $\tilde{a}_{N,n}$ et $\tilde{\lambda}_{N,n}$ de $\tilde{\theta}_{N,n}$, on peut relâcher les contraintes sur N et n et obtenir les mêmes résultats en supposant uniquement $N/n^2 \rightarrow 0$. Soulignons que pour les paramètres a et λ , sous réserve que $N/n^2 \rightarrow +\infty$, un échantillonnage haute fréquence des trajectoires X_i permet d'obtenir un estimateur asymptotiquement équivalent à l'EMV exact basé sur l'observation directe des Γ_i . Contrairement au cas d'un paramètre de diffusion fixe, tous les paramètres sont estimés ici à la même vitesse \sqrt{N} .

Cas particulier où la dérive est connue ($\varphi = 0$). Dans [A8], qui est antérieur, nous avons considéré le cas particulier où $\varphi = 0$. Nous avons défini le même contraste pour estimer les paramètres a et λ et obtenu le même résultat que dans le théorème 6 sous la contrainte plus forte $N/n \rightarrow 0$ et sans considérer l'estimation de paramètres additionnels dans la dérive. Dans [A8], nous avons étudié complètement le cas où, de plus, $\sigma(\cdot) \equiv 1$. La vraisemblance du schéma d'Euler coïncide alors avec la vraisemblance exacte. Nous avons établi que les estimateurs des paramètres du coefficient de diffusion a et λ sont consistants lorsque $N \rightarrow +\infty$ même si le nombre d'observations par trajectoire, n , est fixe. Ils sont également asymptotiquement gaussiens avec une vitesse de convergence \sqrt{N} comme dans le théorème 6.

2.4.4 Effets aléatoires dans la dérive et dans le coefficient de diffusion

Le cas d'effets aléatoires conjointement dans la dérive et dans le coefficient de diffusion est traité dans [A14] dans le modèle

$$\begin{aligned} dX_i(t) &= \phi_i b(X_i(t))dt + \Gamma_i^{-1/2} \sigma(X_i(t)) dW_i(t), \quad X_i(0) = x, \quad i = 1, \dots, N, \\ \Gamma_i &\underset{i.i.d.}{\sim} G(a, \lambda), \\ \phi_i | \Gamma_i = \gamma &\underset{ind.}{\sim} \mathcal{N}(\mu, \gamma^{-1} \omega^2). \end{aligned}$$

Les paramètres à estimer ici sont $\theta = (\lambda, a, \mu, \omega^2)$.

Le calcul de la pseudo-log-vraisemblance $\tilde{\ell}_{N,n}$ soulève une difficulté technique, $\tilde{p}(X_i|\psi)$ n'étant défini que si $S_{i,n} + T_{i,n}(\mu, \omega^2) \geq 0$ pour tout $i = 1, \dots, N$ où $S_{i,n}$ est défini par (2.14) et

$$T_{i,n}(\mu, \omega^2) = \left(\mu - \frac{U_{i,n}}{V_{i,n}} \right)^2 \frac{V_{i,n}}{1 + \omega^2 V_{i,n}} - \frac{U_{i,n}^2}{V_{i,n}}.$$

Nous proposons deux contrastes.

Premier contraste. L'idée est de restreindre le calcul intégral de (2.17) à un sous-ensemble où il est défini. Cela nécessite une hypothèse supplémentaire de bornitude sur l'espace des paramètres :

(A3) Il existe des constantes $m, c_0, c_1, \ell_0, \ell_1, \alpha_0, \alpha_1$ telles que $|\mu| \leq m, \omega^2 \leq c_1, 0 < \ell_0 \leq \lambda \leq \ell_1, 0 < \alpha_0 \leq a \leq \alpha_1$.

On introduit les statistiques

$$Z_{i,n} = Z_{i,n}(\theta) = \frac{\lambda + (S_{i,n}/2) + (T_{i,n}(\mu, \omega^2)/2)}{a + n/2}, \quad M_{i,n} = \max\{c_1 + 2, 2m^2\}(1 + |U_{i,n}|^2), \quad (2.21)$$

et les ensembles $F_{i,n} = \{S_{i,n} - M_{i,n} \geq \alpha\sqrt{n}\}$ où $\alpha > 0$ est une constante à définir. On peut montrer que sous **(A4)**, $Z_{i,n}$ est strictement positif sur $F_{i,n}$. On en déduit le contraste

$$\mathbf{U}_{N,n}(\theta) = \sum_{i=1}^N \mathbf{U}_n(X_i, \theta),$$

où

$$\begin{aligned} \mathbf{U}_n(X_i, \theta) &= a \log \lambda - \log(\Gamma(a)) + \log(\Gamma(a + (n/2))) - (a + (n/2)) \log(a + (n/2)) \\ &\quad - \frac{1}{2} \log(1 + \omega^2 V_{i,n}) - (a + (n/2)) \mathbb{1}_{F_{i,n}} \log(Z_{i,n}). \end{aligned}$$

L'estimateur $\tilde{\theta}_{N,n}$ associé à ce contraste est défini par

$$\mathcal{G}_{N,n}(\tilde{\theta}_{N,n}) = 0, \quad (2.22)$$

où

$$\mathcal{G}_{N,n}(\theta) = \nabla_{\theta} \mathbf{U}_{N,n}(\theta).$$

Théorème 7 *Supposons vérifiées les hypothèses (A1) à (A3), $a_0 > 6$, et $n, N = N(n)$ tendant conjointement vers l'infini tels que $N/n \rightarrow 0$. Supposons de plus $J(\theta_0)$ inversible, où $J(\theta)$ est la matrice de covariance du vecteur*

$$\left(\begin{array}{c} \Gamma_1 A_1(T; \mu, \omega^2) \\ \frac{1}{2} (\Gamma_1 A_1^2(T, \mu, \omega^2) - B_1(T; \omega^2)) \end{array} \right). \quad (2.23)$$

Alors, il existe une solution $\tilde{\theta}_{N,n}$ à (2.22) faiblement consistante et telle que $\sqrt{N}(\tilde{\theta}_{N,n} - \theta_0)$ converge en loi vers $\mathcal{N}(0, \mathcal{J}^{-1}(\theta_0))$ sous \mathbb{P}_{θ_0} , où

$$\mathcal{J}(\theta) = \left(\begin{array}{c|c} I_0(\lambda, a) & \mathbf{0} \\ \hline \mathbf{0} & J(\theta) \end{array} \right).$$

Pour les deux premières composantes de θ , la contrainte $N/n^2 \rightarrow 0$ est suffisante.

Un point important est de montrer que dans l'asymptotique considérée, $\mathbb{P}_{\theta_0}(\cap_{i=1}^N F_{i,n}) \rightarrow 1$ pour tout θ_0 .

Deuxième contraste. Le deuxième contraste que nous proposons est plus heuristique. Il vient de l'approximation de la pseudo-log-vraisemblance par

$$\mathbf{V}_{N,n}(\mathbf{X}, \vartheta) = \mathbf{V}_{N,n}^{(1)}(\mathbf{X}, \lambda, a) + \mathbf{V}_{N,n}^{(2)}(\mathbf{X}, \mu, \Omega), \quad (2.24)$$

où

$$\begin{aligned} \mathbf{V}_{N,n}^{(1)}(\mathbf{X}, \lambda, a) &= -(a + \frac{n}{2}) \sum_{i=1}^N \log \left(\lambda + \frac{S_{i,n}}{2} \right) + N \left\{ a \log \lambda + \log \Gamma(a + \frac{n}{2}) - \log \Gamma(a) \right\}, \\ \mathbf{V}_{N,n}^{(2)}(\mathbf{X}, \mu, \omega^2) &= - \sum_{i=1}^N \left\{ \frac{n}{2 S_{i,n}} \mathbb{1}_{S_{i,n} \geq k\sqrt{n}} \left(\mu - \frac{U_{i,n}}{V_{i,n}} \right)^2 \frac{V_{i,n}}{1 + \omega^2 V_{i,n}} \right. \\ &\quad \left. + \frac{1}{2} \log (\det(1 + \omega^2 V_{i,n})) \right\}. \end{aligned}$$

Cette décomposition permet d'estimer les paramètres de dérive (μ, ω^2) et les paramètres de diffusion (λ, a) séparément :

$$\nabla_{\lambda, a} \mathbf{V}_{N,n}^{(1)}(\mathbf{X}, \tilde{\lambda}_{N,n}, \tilde{a}_{N,n}) = 0 \quad , \quad \nabla_{\mu, \omega^2} \mathbf{V}_{N,n}^{(2)}(\mathbf{X}, \tilde{\mu}_{N,n}, \tilde{\omega}_{N,n}^2) = 0.$$

Nous montrons que ces estimateurs sont asymptotiquement équivalents à ceux obtenus par le premier contraste. L'avantage de cette seconde méthode réside dans la possibilité de traiter séparément l'estimation des paramètres de dérive et l'estimation des paramètres de diffusion et dans la meilleure stabilité numérique qui en résulte.

2.4.5 Remarques sur le cas d'effets aléatoires multidimensionnels dans la dérive

Les contrastes proposés sont étendus aux cas où l'effet aléatoire ϕ_i de la dérive est multidimensionnel ($d > 1$). Nous obtenons des résultats analogues à ceux énoncés dans les théorèmes 5 et 7 sous réserve que les matrices $V_i(T)$:

$$V_i(T) = \left(\int_0^T \frac{b_k(X_i(s)) b_\ell(X_i(s))}{\sigma^2(X_i(s))} ds \right)_{1 \leq k, \ell \leq d}, \quad (2.25)$$

soient définies positives presque sûrement. Je renvoie aux publications [A13] et [A14] pour plus de détails.

Un point remarquable est que les approximations de la vraisemblance restent bien définies même si la matrice de covariance des variables ϕ_i n'est pas inversible, ce qui permet de considérer conjointement l'estimation d'effets fixes inconnus et d'effets aléatoires dans la dérive. Les effets fixes de la dérive sont estimés avec une vitesse \sqrt{N} comme les moyennes et variances des composantes non dégénérées de ϕ_i .

2.4.6 Etudes par simulation

Des études de simulation sont menées séparément dans [A8], [A13] et [A14] pour étudier numériquement chaque méthode d'estimation proposée ainsi que les propriétés des estimateurs associés. Plusieurs modèles sont systématiquement considérés. Je donne un aperçu synthétique des résultats obtenus dans deux modèles de diffusion standards. Je renvoie aux différentes publications pour des résultats plus complets sur des modèles caractérisés par d'autres formes de dérivées et de coefficients de diffusion.

1. Mouvement Brownien avec dérive :

$$dX_i(t) = \phi_i dt + \Gamma_i^{-1/2} dW_i(t), \quad X_i(0) = 0,$$

où

a) $\Gamma_i \underset{i.i.d.}{\sim} G(a, \lambda), \phi_i = \mu$ (Table 2.2)

b) $\Gamma_i \underset{i.i.d.}{\sim} G(a, \lambda), \phi_i | \Gamma_i = \gamma \underset{i.i.d.}{\sim} \mathcal{N}(\mu, \omega^2/\gamma)$ (Table 2.3)

2. Ornstein-Uhlenbeck :

$$dX_i(t) = (\phi_{i1} X_i(t) + \phi_{i2}) dt + \Gamma_i^{-1/2} dW_i(t), \quad X_i(0) = 0,$$

où

a') $\Gamma_i = \gamma$ et $\phi_{i1} = \mu_1, \phi_{i2} \underset{i.i.d.}{\sim} \mathcal{N}(\mu, \omega^2/\gamma)$ (Table 2.4)

b') $\Gamma_i = \gamma, \phi_{i1}$ et ϕ_{i2} indépendants tels que $\phi_{i1} \underset{i.i.d.}{\sim} \mathcal{N}(\mu_1, \omega_1^2/\gamma), \phi_{i2} \underset{i.i.d.}{\sim} \mathcal{N}(\mu_2, \omega_2^2/\gamma)$ (Table 2.4)

c') $\Gamma_i \underset{i.i.d.}{\sim} G(a, \lambda), \phi_{i1} = \mu_1, \phi_{i2} | \Gamma_i = \gamma \underset{ind.}{\sim} \mathcal{N}(\mu_2, \omega_2^2/\gamma)$ (Table 2.5)

Plusieurs valeurs de paramètres et plusieurs valeurs de N (nombre de trajectoires) et n (nombre d'observations par trajectoire) sont envisagées. Dans chaque situation, 100 jeux de données sont simulés. On compare les estimations obtenues avec les contrastes issus du schéma d'Euler à celles qui auraient été obtenues à partir de l'observation directe des effets aléatoires. Le cas a) relève du contraste présenté dans la section 2.4.3, les cas a'), b') du contraste présenté dans la section 2.4.2 et les cas b) et c') des contrastes proposés dans la section 2.4.4. Les résultats sont reportés dans les Tables 2.2 à 2.5.

Des expériences préliminaires montrent que les estimateurs proposés pour les paramètres a et λ sont instables alors que les paramètres de la dérive sont correctement estimés. Les paramètres a et λ sont également mal estimés à partir d'observations directes des effets aléatoires Γ_i , ce phénomène étant peu documenté dans la littérature. Nous reparamétrisons donc la distribution Gamma au moyen de ses paramètres naturels $m = a/\lambda$ et $t = \xi(a) - \log(\lambda)$, où $\xi(\cdot)$ est la fonction Digamma, et procédons à l'estimation de m et t plutôt que de a et λ . Les paramètres m et t sont quant à eux très bien estimés.

Résultats. Les résultats illustrent les bonnes propriétés des estimateurs, avec notamment une amélioration de la précision des estimations à mesure que N augmente. Le rôle de n est peu perceptible, ce qui est cohérent avec les résultats théoriques. Les estimateurs sont \sqrt{N} -consistants sauf celui du paramètre de diffusion, lorsqu'il est fixe, qui est \sqrt{Nn} -consistant.

Les estimations sont presque aussi bonnes que les estimations que l'on aurait obtenues à partir d'observations directes des effets aléatoires. Pour les effets aléatoires du coefficient de diffusion, les résultats obtenus à partir des trajectoires et de l'observation directe des effets aléatoires coïncident, illustrant l'équivalence asymptotique démontrée dans les théorèmes 6 et 7.

Dans les situations où des effets aléatoires interviennent conjointement dans la dérive et dans le coefficient de diffusion, les performances des deux contrastes proposés sont très proches (Tables 2.3 et 2.5). Il est néanmoins possible que le deuxième critère soit numériquement plus stable sur des modèles plus complexes.

Il est intéressant de noter que les résultats sont également très satisfaisants dans les exemples a') et b') (Table 2.4), où il y a à la fois des effets aléatoires et des effets fixes dans la dérive, ce qui illustre la validité du contraste proposé dans la section 2.4.2 y compris dans des situations où la matrice de covariance des effets aléatoires de la dérive n'est pas inversible.

Remarques. Lorsque des effets aléatoires interviennent conjointement dans la dérive et dans le coefficient de diffusion, les deux procédures d'estimation proposées à partir de \mathbf{U}_n et \mathbf{V}_n nécessitent une troncature. La troncature définie au moyen des ensembles $F_{i,n}$ utilisée dans \mathbf{U}_n , bien qu'elle donne de bons résultats en théorie, est difficilement utilisable en pratique. On constate en effet qu'un grand nombre de trajectoires n'appartiennent pas aux ensembles $F_{i,n}$ y compris pour de très petites valeurs de α , ce qui dégrade les estimations. Nous avons donc remplacé $\mathbb{1}_{F_{i,n}}$ par $\mathbb{1}_{(Z_{i,n} \geq c/\sqrt{n})}$ (voir équation (2.21)) dans la définition de \mathbf{U}_n et avons optimisé numériquement le contraste correspondant à cette modification. On utilise la valeur 0.1 pour les constantes k et c qui interviennent dans les troncatures requises pour les définitions des deux approximations de la vraisemblance, mais nous constatons que les résultats des estimations sont peu sensibles aux valeurs choisies.

Finissons par remarquer que les hypothèses **(A1)** et **(A2)** ne sont pas satisfaites dans l'Ornstein-Uhlenbeck. Néanmoins, les résultats numériques montrent que les contrastes que nous avons développés peuvent être utilisés avec succès, même si ces hypothèses ne sont pas vérifiées.

Brownien avec dérive (a)						
$dX_i(t) = \mu_1 dt + \Gamma_i^{-1/2} dW_i(t), \Gamma_i \sim G(a, \lambda)$						
		$N = 100$		$N = 500$		
θ	θ_0		$(n = 500)$	$(n = 1000)$	$(n = 500)$	$(n = 1000)$
m	4	X	4.02 (0.14)	4.01 (0.15)	3.99 (0.07)	3.99 (0.07)
		Γ	4.01 (0.15)	4.01 (0.15)	3.99 (0.07)	3.99 (0.07)
t	1.32	X	1.30 (0.04)	1.31 (0.04)	1.30 (0.02)	1.31 (0.02)
		Γ	1.32 (0.04)	1.32 (0.04)	1.32 (0.02)	1.32 (0.02)
μ_1	-1	X	-1.00 (0.02)	-1.00 (0.02)	-1.00 (0.01)	-1.00 (0.01)
		Γ	-	-	-	-

TABLE 2.2 – Moyennes empiriques et écart-types (entre parenthèses) des estimations des paramètres obtenues pour différentes valeurs de N et n en cas d'effet fixe dans la dérive et d'effet aléatoire dans le coefficient de diffusion. Pour chaque paramètre et chaque couple (N, n) , la ligne X donne les résultats estimés à partir des trajectoires avec le contraste de la section 2.4.3, et la ligne Γ présente les résultats qui auraient été obtenus à partir des observations directes des effets aléatoires simulés pour générer chaque trajectoire.

Brownien avec dérive (b)						
$dX_i(t) = \phi_i dt + \Gamma_i^{-1/2} dW_i(t), X_i(0) = 0$						
		$N = 50$			$N = 100$	
θ	θ_0		$(n = 200)$	$(n = 1000)$	$(n = 200)$	$(n = 1000)$
μ	-0.50	X1	-0.51 (0.06)	-0.51 (0.06)	-0.50 (0.05)	-0.50 (0.05)
		X2	-0.49 (0.06)	-0.51 (0.06)	-0.48 (0.05)	-0.50 (0.05)
		ψ	-0.51 (0.06)	-0.51 (0.06)	-0.50 (0.04)	-0.50 (0.04)
ω^2	0.50	X1	0.49 (0.12)	0.50 (0.13)	0.48 (0.10)	0.48 (0.10)
		X2	0.45 (0.11)	0.49 (0.12)	0.44 (0.09)	0.48 (0.10)
		ψ	0.50 (0.09)	0.50 (0.09)	0.49 (0.08)	0.49 (0.08)
m	4.00	X1	4.00 (0.24)	3.99 (0.23)	4.02 (0.15)	4.00 (0.15)
		X2	3.85 (0.22)	3.96 (0.23)	3.87 (0.14)	3.97 (0.14)
		ψ	3.99 (0.23)	3.99 (0.23)	3.99 (0.14)	3.99 (0.14)
t	1.32	X1	1.28 (0.06)	1.31 (0.06)	1.28 (0.04)	1.31 (0.04)
		X2	1.24 (0.06)	1.30 (0.06)	1.24 (0.04)	1.30 (0.04)
		ψ	1.32 (0.06)	1.32 (0.06)	1.32 (0.04)	1.32 (0.04)

TABLE 2.3 – Moyennes empiriques et écart-types (entre parenthèses) des estimations des paramètres obtenues pour différentes valeurs de N et n . Pour chaque paramètre et chaque couple (N, n) , les lignes X1 et X2 donnent les résultats estimés à partir des trajectoires avec le premier contraste ($\mathbf{U}_{N,n}$ section 2.4.4) et le deuxième contraste ($\mathbf{V}_{N,n}$ section 2.4.4) respectivement, et la ligne ψ présente les résultats qui auraient été obtenus à partir des observations directes des effets aléatoires simulés pour générer chaque trajectoire.

Ornstein-Uhlenbeck									
$dX_i(t) = (\phi_{i1}X_i(t) + \phi_{i2})dt + \gamma^{-1/2}dW_i(t)$									
θ	θ_0	$N = 50$				$N = 100$			
		$(n = 500)$		$(n = 1000)$		$(n = 500)$		$(n = 1000)$	
<i>Un effet aléatoire et un effet fixe dans la dérive (a')</i>									
		X	ϕ	X	ϕ	X	ϕ	X	ϕ
μ_1	-0.1	-0.11 (0.02)	-	-0.10 (0.01)	-	-0.10 (0.02)	-	-0.10 (0.01)	-
μ_2	1	1.01 (0.08)	1.00 (0.05)	1.02 (0.06)	1.00 (0.04)	1.01 (0.06)	1.00 (0.03)	1.00 (0.05)	1.00 (0.03)
ω_2^2	0.4	0.40 (0.14)	0.39 (0.06)	0.40 (0.12)	0.41 (0.09)	0.41 (0.10)	0.41 (0.07)	0.40 (0.09)	0.40 (0.06)
γ	4	4.01 (0.04)	-	4.01 (0.02)	-	4.01 (0.03)	-	4.01 (0.02)	-
<i>Deux effets aléatoires dans la dérive (b')</i>									
		X	ϕ	X	ϕ	X	ϕ	X	ϕ
μ_1	-0.1	-0.10 (0.04)	-0.10 (0.02)	-0.11 (0.02)	-0.10 (0.01)	-0.10 (0.02)	-0.10 (0.01)	-0.10 (0.01)	-0.10 (0.01)
ω_1^2	0.04	0.04 (0.02)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)
μ_2	1	0.98 (0.08)	1.00 (0.04)	1.01 (0.06)	1.00 (0.04)	1.00 (0.05)	1.00 (0.03)	1.00 (0.05)	1.00 (0.03)
ω_2^2	0.4	0.37 (0.15)	0.40 (0.07)	0.37 (0.13)	0.41 (0.07)	0.39 (0.11)	0.40 (0.07)	0.41 (0.10)	0.42 (0.06)
γ	4	4.02 (0.03)	-	4.00 (0.03)	-	4.01 (0.02)	-	4.00 (0.08)	-

TABLE 2.4 – Moyennes empiriques et écart-types (entre parenthèses) des estimations des paramètres obtenues pour différentes valeurs de N et n en cas d'effets aléatoires dans la dérive et d'effet fixe dans le coefficient de diffusion. Pour chaque couple (N, n) , la colonne X donne les résultats estimés à partir des trajectoires avec le contraste de la section 2.4.2, et la ligne ϕ présente les résultats qui auraient été obtenus à partir des observations directes des effets aléatoires simulés pour générer chaque trajectoire.

Ornstein-Uhlenbeck (c')						
$dX_i(t) = (\mu_1 - \phi_{i2}X_i(t))dt + \Gamma_i^{-1/2}dW_i(t)$						
		$N = 50$			$N = 100$	
θ	θ_0		$(n = 200)$	$(n = 1000)$	$(n = 200)$	$(n = 1000)$
μ_1	1.00	X1	1.00 (0.08)	1.00 (0.08)	0.99 (0.05)	1.00 (0.05)
		X2	0.99 (0.07)	1.00 (0.08)	0.98 (0.05)	0.99 (0.05)
		ψ	-	-	-	-
μ_2	0.50	X1	0.50 (0.06)	0.50 (0.06)	0.49 (0.04)	0.50 (0.05)
		X2	0.49 (0.06)	0.50 (0.06)	0.48 (0.04)	0.49 (0.04)
		ψ	0.50 (0.02)	0.50 (0.02)	0.50 (0.01)	0.50 (0.01)
ω_2^2	0.10	X1	0.10 (0.05)	0.09 (0.04)	0.10 (0.04)	0.10 (0.04)
		X2	0.08 (0.04)	0.09 (0.04)	0.09 (0.03)	0.09 (0.04)
		ψ	0.10 (0.02)	0.10 (0.02)	0.10 (0.01)	0.10 (0.01)
m	4.00	X1	4.10 (0.22)	4.01 (0.21)	4.12 (0.16)	4.03 (0.16)
		X2	3.95 (0.21)	3.99 (0.21)	3.97 (0.14)	4.01 (0.14)
		ψ	4.01 (0.21)	4.01 (0.21)	4.01 (0.12)	4.01 (0.12)
t	1.32	X1	1.30 (0.05)	1.31 (0.05)	1.31 (0.04)	1.32 (0.04)
		X2	1.26 (0.05)	1.30 (0.05)	1.27 (0.04)	1.31 (0.03)
		ψ	1.33 (0.05)	1.33 (0.05)	1.33 (0.03)	1.33 (0.03)

TABLE 2.5 – Moyennes empiriques et écart-types (entre parenthèses) des estimations des paramètres obtenues pour différentes valeurs de N et n . Pour chaque paramètre et chaque couple (N, n) , les lignes X1 et X2 donnent les résultats estimés à partir des trajectoires avec le premier contraste ($\mathbf{U}_{N,n}$ section 2.4.4) et le deuxième contraste ($\mathbf{V}_{N,n}$ section 2.4.4) respectivement, et la colonne ψ présente les résultats qui auraient été obtenus à partir des observations directes des effets aléatoires simulés pour générer chaque trajectoire.

Description et inférence de dynamiques épidémiques

Cette partie synthétise mes travaux récents sur la description de dynamiques épidémiques, réalisés dans le cadre de la thèse de Romain Narci entamée en octobre 2018, que je co-encadre avec Catherine Larédo et Elisabeta Vergu.

3.1 Contexte

Les données de surveillance des maladies infectieuses constituent une source riche d'informations pour comprendre, prévoir et contrôler les épidémies actuelles ou futures. Néanmoins, ces données sont souvent incomplètes et entachées d'erreurs d'observation, comme conséquence de mécanismes variés : le processus infectieux n'est que partiellement et/ou imparfaitement observé (on n'observe que les nouveaux individus infectieux symptomatiques et pas les guéris, par exemple) ; le taux de déclaration des cas est relativement faible (*e.g.* les personnes infectées ne consultent pas nécessairement leur médecin) ; les tests diagnostics sont imparfaits. Dans ce contexte, l'intérêt et l'impact de la modélisation mathématique et des approches statistiques en épidémiologie n'ont cessé de croître, comme l'atteste la littérature abondante et récente sur ce sujet [Britton T. et al., 2019].

Les systèmes compartimentaux offrent un cadre structurel naturel pour l'étude de la propagation des épidémies. Dans cette représentation, les différents compartiments représentent des catégories d'individus dans différents états par rapport à la maladie (voir Figure 3.1 pour un exemple). La description de leurs effectifs au cours du temps fait communément appel à des modèles dynamiques mécanistes. Il en existe de nombreuses variantes, développées dans un formalisme déterministe (équations différentielles ordinaires, équations aux dérivées partielles) ou stochastique (processus markoviens de saut, processus de branchement). Les paramètres de ces modèles dynamiques sont directement reliés aux mécanismes de transmission et d'infectiosité qui caractérisent l'épidémie. Leur estimation est donc capitale. Elle repose sur la construction d'un modèle statistique approprié pour les observations, mais est rendue difficile par la nature incomplète et bruitée des données disponibles. Diverses approches ont été développées ces dernières années pour les dynamiques épidémiques partiellement observées et donnent lieu au développement et/ou à l'utilisation de techniques variées pour l'inférence [Britton T. et Giardina F., 2016]. L'augmentation des données, les méthodes sans vraisemblance (*e.g.* les méthodes ABC) et les méthodes basées sur du filtrage particulière ont ouvert des voies prometteuses pour l'amélioration de l'inférence [Anderson H. et Britton T., 2000], [Breto C. et al., 2009], [McKinley T. et al., 2009]. Néanmoins, ces méthodes ne permettent pas de contourner définitivement les problèmes liés aux données incomplètes, puisque la taille des données manquantes et le

nombre important de paramètres algorithmiques à ajuster en limitent souvent la mise en œuvre.

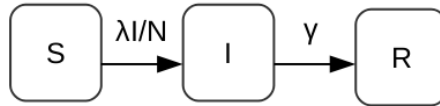


FIGURE 3.1 – Modèle épidémiologique à 3 compartiments : S (individus susceptibles de contracter la maladie), I (individus infectieux), R (individus guéris et/ou immunisés contre la réinfection). Un individu peut passer du compartiment S au compartiment I en contractant le pathogène, ou du compartiment I au compartiment R par guérison. Les équations différentielles ordinaires correspondant à ce type de dynamique figurent dans la section 3.4.

Dans ce contexte, notre contribution est essentiellement méthodologique. Nous proposons une nouvelle méthode d’estimation des paramètres épidémiques clés, reposant sur une approximation gaussienne des modèles d’observation et de dynamique épidémique en grande population. L’originalité consiste en l’utilisation du cadre des diffusions à petite variance et du filtre de Kalman, ce dernier étant peu exploité pour l’inférence de dynamiques épidémiques partiellement observées et bruitées. L’intérêt computationnel de notre approche est de ne nécessiter ni l’augmentation des données, ni l’ajustement de paramètres de réglage algorithmique.

3.2 Formulation d’un modèle linéaire gaussien partiellement observé

Nous considérons un modèle à $p+1$ compartiments dans une population fermée de taille N . $q < p$ compartiments sont observés aux instants discrets $0 = t_0 < t_1 < \dots < t_n = T$, avec erreur de report et autres erreurs de mesure. Pour simplifier, mais sans perte de généralité, on suppose l’échantillonnage régulier, avec un pas de temps Δ , *i.e.* $t_k = k\Delta$ où $T = n\Delta$. Les observations sont ici des observations de prévalence, au sens où elles apportent une information sur les effectifs aux instants d’observations. Les observations d’incidence, relatives aux nouveaux cas, n’entrent pas dans le cadre de modélisation proposé.

Modèle de dynamique épidémique. Le processus markovien de saut à temps continu est le modèle stochastique le plus naturel pour décrire des dynamiques épidémiques. On suppose qu’à chaque instant $t \geq 0$ les effectifs dans chaque compartiment $\mathcal{Z}(t) = (\mathcal{Z}_1(t), \dots, \mathcal{Z}_p(t))$ sont régis par un processus de Markov à valeurs dans $E = \{0, \dots, N\}^p$ et de générateur $\mathcal{Q} = (Q_{ij})$, $i, j \in E$ paramétré par η . Notons qu’en population fermée, pour une dynamique épidémique décrite par $p+1$ stades d’infection (ou compartiments), $\mathcal{Z}_1(t) + \dots + \mathcal{Z}_{p+1}(t) = N$ pour tout $t \geq 0$, ce qui permet de ne considérer que les p premiers stades.

Si l’estimation de η à partir de l’observation des instants de saut et de la chaîne de Markov sous-jacente est directe, l’inférence de processus de saut partiellement observés est en revanche une question difficile. Nous nous inspirons des travaux de [Guy R. et al., 2014, 2015, 2016] qui utilisent des processus de diffusion multidimensionnels à petite variance. Ces processus fournissent une approximation pertinente des processus de saut à temps continu et permettent d’estimer plus facilement les paramètres épidémiques. Nous nous plaçons dans ce cadre, en adoptant ici une approximation gaussienne du processus de saut à temps continu décrivant la dynamique épidémique.

Introduisons pour cela quelques notations et hypothèses. Soit $E^- = \{-N, \dots, 0, \dots, N\}^p$ l’ensemble des tailles de saut. Pour le processus $\mathcal{Z}(t)$, chaque taille de saut $\ell \in E^-$ est as-

sociée une fonction de saut $\alpha_\ell(\cdot)$ définie par

$$\alpha_\ell(k) = Q_{k,k+\ell} \quad \text{pour } k, k+\ell \in E, \ell \neq (0, \dots, 0)^\top. \quad (3.1)$$

On suppose que le processus de saut $\mathcal{Z}(t)$ est densité-dépendant, c'est-à-dire qu'il existe $\beta_\ell, \ell \in \{-N, \dots, 0, \dots, N\}^p$ tels que :

1. $\forall \ell \in E^-, \beta_\ell \in C^2([0, 1]^p)$,
2. $\forall \ell \in E^-, \forall x \in [0, 1]^p, \frac{1}{N}\alpha_\ell([Nx]) \xrightarrow{N \rightarrow +\infty} \beta_\ell(x)$.

On peut alors définir le processus normalisé $\mathcal{Z}_N(t) = \frac{\mathcal{Z}(t)}{N}$. On note, pour $x \in [0, 1]^p$,

$$b(x) = \sum_{\ell \in E^-} \ell \beta_\ell(x); \quad \Sigma(x) = \sum_{\ell \in E^-} \beta_\ell(x) \ell \ell^\top; \quad \Sigma(x) = \sigma(x)^\top \sigma(x).$$

Un théorème central limite [Guy R. et al., 2016] énonce la convergence en loi de $\sqrt{N}(\mathcal{Z}_N(t) - x(\eta, t))$ vers un processus gaussien $g(\eta, t)$, où $x(\eta, t)$ est la solution du système d'équations différentielles ordinaires correspondant au processus de saut :

$$\frac{dx(\eta, t)}{dt} = b(x(\eta, t))dt, \quad (3.2)$$

lorsque la taille de population N tend vers l'infini. Ce théorème et le développement stochastique du processus normalisé permettent d'obtenir l'approximation :

$$\mathcal{Z}_N(t) \underset{N \rightarrow +\infty}{\simeq} x(\eta, t) + \frac{1}{\sqrt{N}} g(\eta, t), \quad (3.3)$$

où $g(\eta, t)$ est un processus gaussien que l'on sait caractériser à partir des fonctions $b(\cdot)$, $\sigma(\cdot)$ et de la matrice résolvante $\Phi(\cdot, \cdot)$ du système différentiel (3.2).

Notons $X_k = \mathcal{Z}_N(t_k)$. Par discrétisation de (3.3), il vient :

$$X_k = F(\eta, \Delta) + A_{k-1}(\eta, \Delta)X_{k-1} + U_k, \quad U_k \sim \mathcal{N}_2(0, T_k(\eta, \Delta)), \quad (3.4)$$

où

$$F(\eta, \Delta, t_k) = x(\eta, t_k) - \Phi(\eta, t_k, t_{k-1})x(t_{k-1}), \quad A_{k-1}(\eta, \Delta) = \Phi(\eta, t_k, t_{k-1}),$$

$$T_k(\eta, \Delta) = \frac{1}{N} \int_{t_{k-1}}^{t_k} \Phi(\eta, t_k, s) \Sigma(\eta, x(\eta, s)) \Phi^\top(\eta, t_k, s) ds,$$

et $\Phi(\eta, \cdot, \cdot)$ est la résolvante de l'équation différentielle ordinaire (3.2). La discrétisation (3.4) fait apparaître un terme de petite variance, de l'ordre de $1/N$.

Modèle d'observation. Nous modélisons l'observation partielle des compartiments épidémiques \mathcal{Z} au moyen d'un opérateur de projection : $B\mathcal{Z}$ où B est une matrice $q \times p$ dont les composantes sont des 0 et 1. Cette modélisation fait intervenir des paramètres d'erreur $\delta = (p_i, \tau_i^2)_{i=1, \dots, q}$. Les paramètres $(p_i, i = 1, \dots, q)$ sont des probabilités qui représentent les taux de report pour chaque composante observée. Nous notons $P(\delta) = \text{diag}(p_i)_{1 \leq i \leq q}$ la matrice des taux de report. Nous proposons l'approximation gaussienne suivante des observations :

$$Y_k = B(\delta)\mathcal{Z}_N(t_k) + V_k, \quad V_k \sim \mathcal{N}_q(0, \Omega_k(\eta, \delta)), \quad (3.5)$$

où (V_k) sont des variables aléatoires indépendantes telles que, pour tout k , V_k est indépendant de $\mathcal{Z}_N(t_k)$, et

$$B(\delta) = P(\delta)B \text{ et } \Omega_k(\eta, \delta) = \frac{1}{N} \text{diag}((p_i(1-p_i) + \tau_i^2)(Bx(\eta, t_k))_i).$$

Cette approximation s'obtient en considérant dans chaque compartiment épidémique observé la combinaison d'une approximation gaussienne d'un tirage binomial $\mathcal{B}(\mathcal{Z}_i, p_i)$ normalisé et d'une erreur de mesure gaussienne de variance hétéroscédastique $\tau_i^2 \mathcal{Z}_{N,i}$.

3.3 Estimation paramétrique

L'objectif est d'estimer les paramètres de la dynamique épidémique η et les paramètres d'erreur et de report δ à partir d'observations y_0, \dots, y_n obtenues aux instants t_k , $k = 0, \dots, n$. Notons $\theta = (\eta, \delta)$ l'ensemble des paramètres à estimer. Nous utilisons le contraste associé à la fonction de pseudo-log-vraisemblance du modèle gaussien approché défini par les équations (3.4)-(3.5). Cette fonction de pseudo-log-vraisemblance approchée s'exprime comme suit :

$$\tilde{\ell}(y_0, \dots, y_n; \theta) = \log \tilde{p}(y_0; \theta) + \sum_{k=1}^n \log \tilde{p}(y_k | y_0, \dots, y_{k-1}; \theta). \quad (3.6)$$

D'après les équations (3.4)-(3.5), les densités conditionnelles $\tilde{p}(y_k | y_0, \dots, y_{k-1}; \theta)$, $k = 1, \dots, n$ sont gaussiennes. Le calcul de (3.6) nécessite uniquement la connaissance des espérances \hat{X}_k et variances $\hat{\Xi}_k$ de ces densités. Celles-ci s'obtiennent récursivement en utilisant les prédictions d'un filtre de Kalman décentré dont l'itération k , $k = 1, \dots, n$ s'écrit :

$$\begin{aligned} \epsilon_k &= Y_k - B\hat{X}_k, \\ \Gamma_k &= B\hat{\Xi}_k B^t + Q_k, \\ H_k &= A_k \hat{\Xi}_k B^t \Gamma_k^{-1}, \\ \hat{X}_{k+1} &= F_{k+1} + A_k \hat{X}_k + H_k \epsilon_k, \\ \hat{\Xi}_{k+1} &= (A_k - H_k B) \hat{\Xi}_k A_k^t + T_{k+1}, \end{aligned}$$

sous réserve d'inversibilité de Γ_k .

3.4 Illustration

3.4.1 Modèle SIR

Dans [S2], nous illustrons notre approche sur l'exemple simple d'un système SIR. Il comprend trois compartiments correspondant respectivement aux individus susceptibles d'être infectés (S), infectieux (I), et immunisés après guérison (R) (voir Figure 3.1). Il constitue la brique de base des modèles compartimentaux de dynamiques épidémiques.

Dans le formalisme des équations différentielles ordinaires (équation (3.2)), il est défini par

$$\begin{aligned} \frac{ds}{dt} &= -\lambda \frac{si}{N}, \\ \frac{di}{dt} &= -\lambda \frac{si}{N} - \gamma i, \\ \frac{dr}{dt} &= \gamma i, \end{aligned}$$

avec $s(t) + i(t) + r(t) = N$ pour tout $t \geq 0$. Ici, $x(t) = (s(t), i(t))$. Dans ces équations, les différents termes traduisent le caractère mécaniste du modèle, λ est le taux de transmission, γ est le taux de guérison (ou l'inverse de la durée d'infectiosité) et N est la taille de la population.

Dans le formalisme stochastique des processus de saut, $\mathcal{Z}(t) = (S(t), I(t))$ est le vecteur des effectifs dans les compartiments S et I à l'instant t . Les taux de transition correspondant aux deux événements possibles *infection* et *guérison* sont respectivement :

$$Q_{(S,I),(S-1,I+1)} = \lambda S \frac{I}{N} = \alpha_{(-1,1)}(S, I) \text{ et } Q_{(S,I),(S,I-1)} = \gamma I = \alpha_{(0,-1)}(S, I).$$

Il s'agit d'un modèle densité dépendant où $\beta_{(-1,1)} = \lambda$ et $\beta_{(0,-1)} = \gamma$.

3.4.2 Etude sur données simulées

Nous étudions les performances de notre méthode d'estimation sur des jeux de données simulés. Dans chaque situation, 500 épidémies de type SIR sont simulées selon le modèle de processus de saut associé. Pour chaque épidémie, n observations régulièrement espacées sont simulées selon un tirage binomial de probabilité p sur le compartiment des infectés. Aucune erreur de mesure supplémentaire n'est considérée ($\tau = 0$). On suppose par ailleurs l'absence de personnes immunisées au début des épidémies, c'est-à-dire $s_0 + i_0 = 1$. Les paramètres inconnus $\theta = (\lambda, \gamma, p, i_0)$ sont estimés sur chaque trajectoire par optimisation numérique de la pseudo-vraisemblance gaussienne (section 3.3). Les estimations sont comparées avec celles obtenues par Maximum Iterated Filtering [Ionides E. L. et al., 2006], un algorithme de Monte-Carlo séquentiel. Différents scénarios correspondant à différentes valeurs des paramètres λ et γ sont considérés dans notre étude, ainsi que différentes valeurs de N (taille de population), de n (nombre d'observations par épidémie) et de p (taux de report des infectés). Je montre dans les Tables 3.1 et 3.2 les résultats obtenus pour $\lambda = 1$, $\gamma = 1/3$, avec pour conditions initiales $s_0 = 0.95$ et $i_0 = 0.05$. La Table 3.1 correspond à un taux de report $p = 0.8$ et la Table 3.2 correspond à un taux de report $p = 0.3$.

Résultats. Les résultats des Tables 3.1 et 3.2 illustrent les bonnes propriétés des estimateurs obtenus par maximisation de la vraisemblance du modèle gaussien approché. On constate une amélioration de la précision des estimations à mesure que N augmente ; celle-ci est liée à la validité en grande population de l'approximation par processus gaussien des dynamiques épidémiques. Le nombre d'observations par épidémie joue aussi un rôle important dans la qualité des estimations de certains paramètres, avec diminution du biais et des écart-types lorsque n augmente. Le paramètre de report p détermine en quelques sortes l'intensité du signal : plus sa valeur est faible, plus l'information disponible sur les dynamiques épidémiques est faible. Sans surprise, le biais et la précision des estimations sont meilleures pour $p = 0.8$ (Table 3.1) que pour $p = 0.3$ (Table 3.2). La modification de p affecte de la même façon la qualité des résultats fournis par Maximum Iterated Filtering.

Les estimations obtenues par notre approche et celles obtenues par Maximum Iterated Filtering sont par ailleurs très proches dans l'ensemble. Précisons que contrairement à notre approche, cette seconde méthode estime ici les paramètres à partir du modèle génératif associant processus de saut pour la dynamique épidémique et tirage binomial sur le compartiment des infectés pour les observations. La mise en œuvre du Maximum Iterated Filtering nécessite de nombreux réglages algorithmiques, en particulier celui du nombre de particules. Ces réglages sont difficiles pour des utilisateurs non spécialistes et s'avèrent déterminants pour estimer correctement les paramètres. Nous avons également remarqué une forte sensibilité du Maximum Iterated Filtering aux initialisations des paramètres. Notre approche ne nécessite pas ces réglages délicats.

Nous avons envisagé d'autres valeurs de paramètres, en particulier pour λ et γ ; les résultats ne sont pas reportés dans ce manuscrit mais j'en donne les idées principales. L'objectif était d'examiner l'impact sur les estimations de la valeur du ratio $R_0 = \lambda/\gamma$ qui définit le taux de reproduction de base, une quantité centrale en épidémiologie correspondant au nombre de cas secondaires générés par chaque personne infectieuse. Si $R_0 < 1$, le pathogène ne peut pas déclencher d'épidémie. Pour $\lambda = 1$ et $\gamma = 1/3$ (Tables 3.1 et 3.2), ce ratio vaut 3 et correspond à des épidémies à propagation relativement rapide. Nous avons également considéré un taux de reproduction de base plus faible, en utilisant $\lambda = 0.6$ et $\gamma = 0.4$, soit $R_0 = 1.2$. Les résultats sont légèrement dégradés par rapport à ceux obtenus lorsque $R_0 = 3$ mais révèlent les mêmes tendances en termes de biais et de précision des

estimations. L'estimation des paramètres épidémiques est plus difficile pour des valeurs de R_0 plus faibles. En effet, plus le R_0 est petit, plus l'intensité du signal disponible à travers les observations des personnes infectées est faible dans l'ensemble, les épidémies se propageant plus lentement, avec un pic épidémique moins marqué et avec une probabilité d'extinction précoce plus élevée.

D'autre part, observer le compartiment des susceptibles (S) plutôt que le compartiment des infectés (I) donne lieu à des résultats similaires en ce qui concerne les influences respectives de N , n , p et R_0 sur les résultats.

Cette étude de simulations approfondie révèle tout de même des difficultés à estimer les conditions initiales s_0 et i_0 des dynamiques épidémiques. Cette difficulté est rarement mentionnée dans la littérature, et est généralement contournée par le recours à des *a priori* bayésiens. Ceci soulève la question de l'identifiabilité des conditions initiales.

TABLE 3.1 – Résultats obtenus pour $\lambda = 1$, $\gamma = 1/3$, $s_0 = 0.95$, $i_0 = 0.05$ et $p = 0.8$. Pour chaque combinaison (N, n) , les résultats consistent en la moyenne empirique des estimations et leur écart-type empirique (entre parenthèses) obtenus par notre approche (K) et par Maximum Iterated Filtering (M).

θ		$N = 1000$		$N = 2000$		$N = 10000$	
		$n = 30$	$n = 100$	$n = 30$	$n = 100$	$n = 30$	$n = 100$
$\lambda = 1$	K	1.02 (0.08)	1.00 (0.07)	1.03 (0.07)	0.98 (0.06)	1.01 (0.04)	1.00 (0.03)
	M	1.00 (0.07)	1.00 (0.06)	1.02 (0.05)	1.01 (0.05)	1.01 (0.02)	1.01 (0.02)
$\gamma = 1/3$	K	0.32 (0.04)	0.33 (0.03)	0.31 (0.03)	0.34 (0.02)	0.32 (0.02)	0.34 (0.02)
	M	0.32 (0.03)	0.34 (0.02)	0.31 (0.03)	0.34 (0.02)	0.32 (0.02)	0.34 (0.01)
$p = 0.8$	K	0.76 (0.11)	0.80 (0.07)	0.75 (0.09)	0.81 (0.06)	0.77 (0.06)	0.81 (0.06)
	M	0.76 (0.09)	0.81 (0.04)	0.73 (0.07)	0.81 (0.03)	0.75 (0.05)	0.80 (0.04)
$i_0 = 0.05$	K	0.056 (0.011)	0.054 (0.011)	0.052 (0.008)	0.045 (0.009)	0.052 (0.004)	0.050 (0.005)
	M	0.055 (0.009)	0.052 (0.005)	0.054 (0.007)	0.052 (0.004)	0.053 (0.004)	0.050 (0.003)

TABLE 3.2 – Résultats obtenus pour $\lambda = 1$, $\gamma = 1/3$, $s_0 = 0.95$, $i_0 = 0.05$ et $p = 0.3$. Pour chaque combinaison (N, n) , les résultats consistent en la moyenne empirique des estimations et leur écart-type empirique (entre parenthèses) obtenus par notre approche (K) et par Maximum Iterated Filtering (M).

θ		$N = 1000$		2000		10000	
		$n = 30$	$n = 100$	$n = 30$	$n = 100$	$n = 30$	$n = 100$
$\lambda = 1$	K	1.00 (0.10)	1.03 (0.08)	1.02 (0.07)	1.02 (0.06)	1.00 (0.04)	1.02 (0.03)
	M	1.02 (0.06)	1.03 (0.08)	1.01 (0.05)	1.01 (0.05)	1.00 (0.02)	1.01 (0.02)
$\gamma = 1/3$	K	0.33 (0.04)	0.32 (0.05)	0.32 (0.04)	0.32 (0.04)	0.34 (0.03)	0.32 (0.02)
	M	0.32 (0.06)	0.32 (0.05)	0.33 (0.03)	0.33 (0.03)	0.34 (0.02)	0.31 (0.02)
$p = 0.3$	K	0.28 (0.05)	0.27 (0.05)	0.29 (0.04)	0.29 (0.04)	0.31 (0.03)	0.29 (0.03)
	M	0.28 (0.04)	0.26 (0.04)	0.29 (0.03)	0.27 (0.03)	0.30 (0.02)	0.28 (0.02)
$i_0 = 0.05$	K	0.047 (0.012)	0.045 (0.011)	0.051 (0.007)	0.049 (0.007)	0.048 (0.004)	0.052 (0.004)
	M	0.048 (0.010)	0.052 (0.009)	0.051 (0.007)	0.053 (0.006)	0.050 (0.004)	0.053 (0.003)

Je présente maintenant quelques perspectives de recherche. Comme le montrent les chapitres principaux de ce document de synthèse, mes activités combinent développement méthodologique, recherche théorique, et applications. Je souhaite maintenir cet équilibre stimulant et ainsi continuer à travailler sur des problèmes statistiques inspirés de questions appliquées. Si jusqu'à présent, l'essentiel de mes recherches a été alimenté de questions issues de la pharmacologie et de l'épidémiologie, mes activités d'enseignant-chercheur à AgroParisTech et ma délégation dans l'unité INRAE MaIAGE de Jouy-en-Josas m'ont également sensibilisée aux questions statistiques soulevées par d'autres applications. Ainsi certaines perspectives évoquées dans mon projet de recherche concernent les sciences du végétal ou encore l'écologie du mouvement. Mes perspectives de recherche s'articulent autour de plusieurs axes.

4.1 Méthodes statistiques pour l'analyse de modèles à effets mixtes

4.1.1 Sélection de modèle

La sélection de variables et d'effets aléatoires dans les modèles à effets mixtes est toujours un thème de recherche très fécond. Dans la continuité des travaux présentés dans le chapitre 1, je m'intéresse à trois questions.

Consistance du BIC. Les critères de type BIC que j'ai proposés pour la sélection de covariables et d'effets aléatoires dans les modèles à effets mixtes présentent de bonnes performances en pratique. Si leurs pénalités respectives proviennent d'une étude rigoureuse, il reste à montrer que ces critères sont consistants pour les différentes questions de sélection pour lesquelles ils ont été conçus. Les schémas de preuve classiques sont construits sur des ensembles de modèles emboîtés et utilisent les propriétés de la distance de Kullback-Leibler sous cette condition [Burnham K. P. et Anderson D. R., 2002]. Or, le formalisme des modèles non linéaires à effets mixtes et les questions de sélection examinées ne permettent pas de faire cette hypothèse. Il est peut-être possible de s'inspirer de l'astuce de reparamétrisation utilisée par [Keribin C., 2000] pour contourner des problèmes d'identifiabilité qui rendent difficile l'étude du BIC dans le cadre de modèles de mélanges finis.

Sélection en grande dimension. Dans certaines applications, la description de la variabilité inter-individuelle implique des données génomiques de grande dimension. Com-

prendre les rôles du génotype, des conditions environnementales et de leurs interactions dans les processus de développement des plantes est un enjeu majeur en amélioration des plantes. Or, sélectionner les marqueurs influents nécessite d’aborder l’inférence statistique des modèles à effets mixtes dans un cadre où le nombre de covariables est plus grand que le nombre d’individus. Cette question a suscité de nombreuses contributions ces dernières années mais celles-ci concernent spécifiquement les modèles linéaires généralisés à effets mixtes (voir par exemple [Schell dorfer J. et al., 2011]). A ma connaissance, la question n’a jamais été abordée lorsque le modèle est non linéaire. Je commence à travailler avec Laure Sansonnet (AgroParisTech) sur la sélection de covariables dans les modèles non linéaires à effets mixtes dans ce contexte de grande dimension. Il s’agit d’abord d’établir un critère de régularisation pour la sélection parcimonieuse des prédicteurs parmi les marqueurs moléculaires et d’en étudier les propriétés théoriques. Il est sans doute possible de tirer parti de la linéarité des paramètres génétiques en les effets aléatoires, classique dans ces modèles (voir équation (1.2) dans le chapitre 1), pour adapter les méthodes développées dans le cadre des modèles linéaires à effets mixtes et la théorie des estimateurs correspondants. Ce contexte de grande dimension pose également des questions d’ordre algorithmique pour l’optimisation numérique du critère retenu. Les méthodes classiques, algorithmes ABC (Approximate Bayesian Computation) ou méthodes MCMC, nécessitent à chaque itération des simulations intensives qui occasionneraient des temps de calcul beaucoup trop longs en contexte de grande dimension. Le développement d’algorithmes spécifiques permettant de passer à l’échelle, *i.e.* de traiter des jeux de données importants en un temps raisonnable, représente une partie capitale du travail à réaliser. Une piste consiste à s’inspirer de [Allasonnière S. et Kuhn E., 2015] et à coupler des algorithmes d’approximation stochastique avec des méthodes d’échantillonnage adaptées à la grande dimension. Des travaux théoriques nouveaux pour en délimiter rigoureusement les garanties de convergence seront sans doute nécessaires.

Critère non asymptotique. A ma connaissance, l’ensemble des méthodes proposées pour la sélection de covariables et d’effets aléatoires dans les modèles à effets mixtes reposent sur des critères asymptotiques dont les propriétés sont établies lorsque la taille de l’échantillon tend vers l’infini. Ces méthodes sont largement utilisées en pratique sur des échantillons de taille réduite où ces garanties théoriques ne sont plus valides. Je souhaiterais me placer dans un cadre non asymptotique, plus cohérent avec de nombreux contextes d’études, pour aborder la question de la sélection de covariables et d’effets aléatoires dans les modèles à effets mixtes. Pour construire un critère approprié et en étudier les propriétés, l’une des difficultés sera sans doute l’absence de forme explicite pour la vraisemblance.

4.1.2 Estimation de la matrice d’information de Fisher

Les travaux que j’ai réalisés avec Estelle Kuhn sur l’estimation de l’information de Fisher dans les modèles à variables latentes (section 1.3) ouvrent également des perspectives. Nous allons appliquer notre approche à des modèles de culture en amélioration des plantes. La précision d’estimation des paramètres génétiques est déterminante en sélection génomique mais les modèles de culture sont généralement des modèles complexes dans lesquels les dérivées des log-vraisemblances complètes ne sont pas explicites. On pourrait imaginer utiliser des schémas numériques tels que la méthode des différences finies à la place d’un calcul exact, mais il faudrait étudier rigoureusement l’impact d’une approximation numérique des dérivées de la log-vraisemblance sur les performances de l’algorithme d’approximation stochastique décrit dans la section 1.3. Il s’agirait en particulier d’étudier les distances entre la solution discrète donnée par approximation numérique et la dérivée exacte lorsque les pas du maillage tendent vers 0.

4.2 Modèles de diffusion à effets aléatoires

Distributions non spécifiques pour les effets aléatoires. Une partie importante des recherches que j’ai entreprises depuis ma thèse concerne l’estimation paramétrique dans les modèles de diffusion à effets mixtes, et en particulier l’étude des propriétés asymptotiques de M-estimateurs (voir chapitre 2). Aucun résultat théorique général n’avait, à ma connaissance, été proposé avant l’ensemble de ces travaux. Les hypothèses de distribution gaussienne et Gamma pour les effets aléatoires y sont néanmoins cruciales. Je démarre une collaboration avec Hiroki Masuda (Université de Kyushu, Japon) pour étendre ces résultats à des distributions plus générales pour les effets aléatoires. La pseudo-vraisemblance issue du schéma d’Euler n’est plus explicite, mais il est possible de définir des contrastes génériques et explicites à partir de fonctions d’estimation martingales.

Prise en compte d’un bruit de mesure. Une autre extension intéressante des travaux présentés dans le chapitre 2 consisterait à considérer des observations bruitées des N trajectoires, mais le bruit de mesure fait intervenir une étape de filtrage supplémentaire dont la prise en compte dans l’étude théorique des estimateurs des paramètres est peu aisée.

Sélection de modèle. Enfin, nous n’avons pas considéré de covariables dans les modèles de diffusion que nous avons étudiés. Comme évoqué dans la section 1.4 de ce manuscrit, l’inclusion de covariables peut néanmoins s’avérer essentielle pour la description de la variabilité inter-sujets. Le travail récent publié dans [Ruse M. G. et al., 2020] considère l’estimation des paramètres en présence de covariables dans la dérive pour des applications neuronales et cliniques. Dans des modèles de diffusion généraux incluant covariables et effets aléatoires, se posent les mêmes questions de sélection de modèle que dans la section 1.4. Les critères de type BIC que j’ai développés sont cependant mal adaptés aux modèles de diffusion. Avec Hiroki Masuda, nous souhaiterions proposer des critères de sélection de modèle plus appropriés pour la sélection de covariables et d’effets aléatoires dans les modèles de diffusion à effets mixtes. Nous devrions en particulier pouvoir nous inspirer des travaux de [Eguchi S. et Masuda H., 2018].

4.3 Modélisation et estimation de dynamiques épidémiques

J’ai commencé à m’intéresser à la description de dynamiques épidémiques pendant mes deux années de délégation dans l’unité INRAE MaIAGE. J’ai ensuite démarré le co-encadrement d’une thèse sur cette thématique, dont les premiers résultats font l’objet du chapitre 3. De nombreuses questions émergent de ce travail préliminaire, et devraient pour certaines d’entre elles être considérées dans la suite de la thèse de Romain Narci.

Observation des nouveaux infectés. Dans [S2], nous proposons un cadre statistique spécifique pour l’estimation de dynamiques épidémiques à partir d’observations d’un ou plusieurs compartiments en des temps discrets. Le recueil des données par les réseaux de surveillance donne plus souvent accès à une mesure d’incidence, à travers par exemple un nombre de nouveaux infectés sur des intervalles de temps successifs. Notre travail ne permet pas de considérer ce type d’observation. Les diffusions hypoelliptiques offriraient en revanche un cadre statistique plus satisfaisant. Il doit être possible de s’inspirer des travaux de Ditlevsen S. et Samson A. [2019]. L’outil principal de leurs méthodes est une pseudo-vraisemblance basée sur un schéma de discrétisation spécifique garantissant la convergence vers la vraisemblance exacte. Des développements théoriques supplémentaires sont néanmoins nécessaires en raison de la petite variance des diffusions pour montrer que ce schéma de discrétisation présente les mêmes propriétés dans les modèles que nous considérerons, ou dans le cas contraire identifier un schéma de discrétisation plus approprié. De plus, une

extension supplémentaire serait nécessaire pour la prise en compte d'un bruit de mesure, non considéré dans Ditlevsen S. et Samson A. [2019].

Propriétés théoriques du filtre de Kalman et petite variance. Dans [S2], nous utilisons le filtre de Kalman dans le cadre peu usuel de processus à petites variances. A ma connaissance, aucun résultat n'existe sur le comportement du filtre de Kalman lorsque la variance des termes de bruit est petite. L'étude théorique du filtre de Kalman dans ce cadre serait une extension tout à fait naturelle à notre travail.

Prise en compte de la variabilité géographique. La deuxième partie de la thèse de Romain Narci est consacrée à l'estimation de paramètres épidémiques pour des épidémies multi-sites. Pour tenir compte de la variabilité géographique dans la propagation des épidémies, une approche courante consiste à diviser les compartiments des modèles standards en plusieurs compartiments correspondant chacun à une zone géographique différente. Une autre façon de prendre en compte la dynamique spatiale dans le modèle consisterait à définir certains paramètres du modèle comme des paramètres aléatoires avec une structure de dépendance adaptée. A ma connaissance, les modèles à effets mixtes sont peu utilisés en épidémiologie, et cette approche est tout à fait originale dans ce type d'application. Un des objectifs est ici de développer des algorithmes adaptés à ces modèles, et peu gourmands en temps de calcul, pour estimer les paramètres épidémiques clé et leur variabilité. Cela demande en parallèle de mener une étude théorique précise des propriétés des estimateurs des paramètres dans les différentes modélisations possibles, ce qui rejoint les travaux que je décris dans le chapitre 2 et les perspectives de recherche évoquées dans la section 4.2. En ce qui concerne le volet méthodologique, les processus gaussiens fournissant une approximation pertinente des dynamiques épidémiques, il sera possible d'étendre l'algorithme SAEM couplé au filtre de Kalman proposé dans [A5]. Une autre approche possible pourrait être de considérer un algorithme SAEM couplé à des méthodes particulières [Donnet S. et Samson A., 2014]. Dans les deux cas, des adaptations pourraient être nécessaires pour tenir compte de la petite variance des diffusions.

4.4 Segmentation de trajectoires en écologie du mouvement

Je termine la description de mon projet de recherche en évoquant des travaux que j'ai entrepris avec Marie-Pierre Etienne (Agrocampus Ouest, IRMAR) en lien avec la segmentation de trajectoires pour la description des activités de bateaux de pêche. Les positions des bateaux de pêche durant leurs opérations en mer sont connues et mesurées grâce à un système de positionnement GPS disponible de manière réglementaire sur la plupart des navires. La modélisation de la dynamique de ces navires doit permettre d'évaluer l'effort de pêche, qui est ensuite utilisé pour normaliser les captures commerciales. Il est important de pouvoir faire la différence entre un navire qui fait route sur une zone de pêche, un navire qui recherche de façon active du poisson et un navire en train de conduire une opération de pêche. La question a déjà été abordée dans la littérature (Gloaguen P. et al. [2015]) avec des modèles de Markov cachés en temps discrets. Cependant, ces modèles font deux hypothèses fortes qui limitent leur utilisation en pratique : les positions des bateaux doivent être recueillies de façon régulière et les changements de régime sont supposés synchrones aux observations. Nous proposons ici de modéliser les vitesses de déplacement des bateaux en temps continu au moyen de diffusions à changements de régime markoviens pour une meilleure prise en compte de l'asynchronie entre changements de régime et observations. Plus précisément, les paramètres des équations différentielles stochastiques sont régis par un processus de saut markovien, dont les états représentent les différents comportements en mer. En utilisant cette modélisation, les objectifs sont *i*) estimer les paramètres du modèle, *ii*) reconstruire les états cachés, *iii*) estimer les instants de changement d'activité, à partir

des observations discrètes et haute-fréquence d'une trajectoire GPS. Les algorithmes EM et dérivés sont classiques pour estimer les paramètres en présence de variables latentes. Dans les modèles que nous considérons, y compris les plus simples, l'étape E de l'algorithme EM n'est cependant pas explicite. Simuler les données cachées pour approcher l'étape E , comme dans les algorithmes MCEM et SAEM, est une alternative bien connue. Cette phase de simulation des données cachées est très délicate dans les diffusions à changement de régime car en plus de simuler les états de la chaîne de Markov cachée, il est nécessaire de simuler les temps de séjour dans les états consécutifs. Nous considérons d'abord un modèle très simple défini par un mouvement brownien avec dérive et une chaîne de Markov cachée à deux états. Nous proposons une version particulière de l'algorithme EM reposant sur l'algorithme de PaRIS (Olsson J. et Westerborn J. [2015]) pour la simulation des données cachées et une approximation de la log-vraisemblance des données complètes donnant accès à des statistiques exhaustives explicites et à une approximation de la loi d'importance optimale. Une fois les paramètres estimés, il est assez direct de reconstruire les séquences d'états cachés et les instants de saut par FFBSi (forward filtering backward simulation) en réutilisant les données simulés par l'algorithme de PaRIS. Les premiers résultats obtenus sur quelques jeux de données simulées sont convaincants, mais la méthode nécessite d'être généralisée pour d'autres processus comme les Ornstein-Uhlenbeck. Un point d'achoppement est l'approximation de la loi d'importance optimale qui est très spécifique au modèle de diffusion considéré.

- Albert P. S. A Two-State Markov Mixture Model for a Time Series of Epileptic Seizure Counts. *Biometrics*, 47(4) :1371–1381, 1991.
- Allasonnière S. et Kuhn E. Construction of Bayesian deformable models via a stochastic approximation algorithm : a convergence study. *Bernoulli*, 16(3) :641–678, 2010.
- Allasonnière S. et Kuhn E. Convergent Stochastic Expectation Maximization algorithm with efficient sampling in high dimension. Application to deformable template model estimation. *Computational Statistics and Data Analysis*, 91 :4–19, 2015.
- Altman R. M. Mixed Hidden Markov Models : An Extension of the Hidden Markov Model to the Longitudinal Data Setting. *Journal of the American Statistical Association*, 102 (477) :201–210, 2007.
- Anderson H. et Britton T. *Stochastic Epidemic Models and Their Statistical Analysis*. Springer, New York, NY, 2000.
- Anisimov V. V., Maas H. J., Danhof M., et Della Pasqua O. Analysis of responses in migraine modelling using hidden Markov models. *Statistics in Medicine*, 26 :4163–4178, 2007.
- Beal S. L. et Sheiner L. B. Estimating population kinetics. *Crit. Rev. Biomed. Eng.*, 8(3) : 195–222, 1982.
- Breto C., He D., Ionides E. L., et King A. Time series analysis via mechanistic models. *The Annals of Applied Statistics*, 3(1) :319–348, 2009.
- Britton T. et Giardina F. Introduction to statistical inference for infectious diseases. *Journal de la Société Française de Statistique*, 157(1) :53–70, 2016.
- Britton T., Pardoux E., Ball F., Larédo C., Sirl D., et Tran V. C. *Stochastic Epidemic Models With Inference*. Lecture Notes in Mathematics. Springer Nature Switzerland AG, first edition edition, 2019.
- Burdet C., Pajot O., Couffignal C., Armand-Lefevre L., Foucrier A., Laouénan C., Wolff M., Massias M., et Mentré F. Population pharmacokinetics of single-dose amikacin in critically ill patients with suspected ventilator-associated pneumonia. *European Journal of Clinical Pharmacology*, 71 :75–83, 2015.
- Burnham K. P. et Anderson D. R. *Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach*. Springer-Verlag New York, second edition edition, 2002.

- Celeux G., Martin O., et Lavergne C. Mixture of linear mixed models for clustering gene expression profiles from repeated microarray experiments. *Statistical Modelling*, 5 :1–25, 2005.
- Claeskens G. et Hjort N. L. *Model Selection and Model Averaging*. Cambridge University Press, 2008.
- Davidian M. et Gallant A. R. The nonlinear mixed effects model with a smooth random effects density. *Biometrika*, 80(3) :475–488, 1993.
- Delyon B., Lavielle M., et Moulines E. Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics*, 27(1) :94–128, 1999.
- Dempster A. P., Laird N. M., et Rubin D. B. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of Royal Statistical Society. Series B*, 39(1) :1–38, 1977.
- Dion C. Nonparametric estimation in a mixed-effect ornstein-uhlenbeck model. *Metrika*, 79(8) :919–951, 2016.
- Ditlevsen S. et De Gaetano A. Mixed effects in stochastic differential equation models. *REVSTAT*, 3(2) :137–153, 2005.
- Ditlevsen S. et Samson A. Hypoelliptic diffusions : discretization, filtering and inference from complete and partial observations. *Journal of Royal Statistical Society : Series B*, 81 :361–384, 2019.
- Donnet S. et Samson A. Parametric inference for mixed models defined by stochastic differential equations. *ESAIM :PS*, 12 :196–218, 2008.
- Donnet S. et Samson A. A review on estimation of stochastic differential equations for pharmacokinetic/pharmacodynamic models. *Advanced Drug Delivery Reviews*, 65(7) :929–939, 2013.
- Donnet S. et Samson A. Using PMCMC in EM algorithm for stochastic mixed models : theoretical and practical issues. *Journal de la Société Française de Statistique*, 155 :49–72, 2014.
- Eguchi S. et Masuda H. Schwarz type model comparison for LAQ models. *Bernoulli*, 24 (3) :2278–2327, 2018.
- Gloaguen P., Mahévas S., Rivot E., Woillez M., Guitton J., Vermard Y., et Etienne M. P. An autoregressive model to describe fishing vessel movement and activity. *Environmetrics*, 26(1) :17–28, 2015.
- Guy R., Larédo C., et Vergu E. Parametric inference for discretely observed multidimensional diffusions with small diffusion coefficient. *Stochastic Processes and their Application*, 124 :51–80, 2014.
- Guy R., Larédo C., et Vergu E. Approximation of epidemic models by diffusion processes and their statistical inference. *Journal of Mathematical Biology*, 70 :621–646, 2015.
- Guy R., Larédo C., et Vergu E. Approximation and inference of epidemic dynamics by diffusion processes. *Journal de la Société Française de Statistique*, 157(1) :71–100, 2016.
- Hui F. et Muller S. Joint Selection in Mixed Models using Regularized PQL. *Journal of the American Statistical Association*, 112(519) :1323–1333, 2017.
- Ionides E. L., Breto C., et King A. Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences of the United States of America*, 103(49), 2006.

- Keribin C. Consistent estimation of the order of mixture models. *Sankhya : The Indian Journal of Statistics*, 62 :49–66, 2000.
- Kessler M., Lindner A., et Sorensen M. *Statistical Methods for Stochastic Differential Equations*. Chapman & Hall/CRC Monographs on Statistics and Applied Probability. Chapman and Hall/CRC, 2012.
- Kuhn E. et Lavielle M. Coupling a stochastic approximation version of EM with an MCMC procedure. *ESAIM :PS*, 8 :115–131, 2004.
- Kuhn E. et Lavielle M. Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics and Data Analysis*, 49 :1020–1038, 2005.
- Kuznetsova A., Brockhoff P. B., et Christensen R. H. B. lmerTest Package : Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 2017.
- Lavielle M. *Mixed Effects Models for the Population Approach. Models, Tasks, Methods and Tools*. Chapman and Hall/CRC Biostatistics Series. Chapman and Hall/CRC, 2014.
- Lavielle M. et Mbogning C. An improved SAEM algorithm for maximum likelihood estimation in mixtures of non linear mixed effects models. *Statistics and Computing*, 24(5) : 693–707, 2014.
- Lavielle M. et Mentré F. Estimation of population pharmacokinetics parameters of saquinavir in hiv patients with the Monolix software. *Journal of Pharmacokinetics and Pharmacodynamics*, 34(2) :229–249, 2007.
- Lebarbier E. et Mary-Huard T. Une introduction au critère BIC : fondements théoriques et interprétation. *Journal de la société française de statistique*, 147(1) :39–57, 2006.
- Lindstrom M. J. et Bates D. M. Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46(3) :673–687, 1990.
- Louis T. A. Finding the Observed Information Matrix when Using the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(2) :226–233, 1982.
- Luts J., Molenberghs G., Verbeke G., Van Huffel S., et Suykens J. A. K. A mixed effects least squares support vector machine model for classification of longitudinal data. *Computational Statistics and Data Analysis*, 56(3) :611–628, 2012.
- McKinley T., Cook A. R., et Deardon R. Inference in Epidemic Models without Likelihoods. *The International Journal of Biostatistics*, 5(1), 2009.
- McLachlan G. et Krishnan T. *The EM Algorithm and Extensions*, volume 249 of *Wiley Series in Probability and Statistics*. Wiley, 1996.
- Muller S., Scealy J. L., et Welsh A. H. Model Selection in Linear Mixed Models. *Statistical Science*, 28(2) :135–167, 2013.
- Nie L. Strong consistency of the maximum likelihood estimator in generalized linear and nonlinear mixed-effects models. *Metrika*, 63(2) :123–143, 2006.
- Nie L. Convergence rate of the MLE in generalized linear and nonlinear mixed-effects models : Theory and applications. *Journal of Statistical Planning and Inference*, 137 (6) :1787–1804, 2007.
- Nie L. et Yang M. Strong consistency of the MLE in nonlinear mixed-effects models with large cluster size. *Sankhya : The Indian Journal of Statistics*, 67(4) :736–763, 2005.

- Ollier E., Samson A., Delavenne X., et Viallon V. A SAEM Algorithm for Fused Lasso Penalized Non Linear Mixed Effect Models : Application to Group Comparison in Pharmacokinetic. *Computational Statistics and Data Analysis*, 96 :207–221, 2016.
- Olsson J. et Westerborn J. An efficient particle-based online EM algorithm for general state-space models. *IFAC-PapersOnLine*, 48(28) :963–968, 2015.
- Overgaard R. V., Jonsson N., Tornoe C. W., et Madsen H. Non-linear mixed-effects models with stochastic differential equations : Implementation of an estimation algorithm. *Journal of Pharmacokinetics and Pharmacodynamics*, 32(1) :85–107, 2005.
- Pinheiro J. C. et Bates D. M. *Mixed-Effects Models in S and S-PLUS*. Statistics and Computing. Springer-Verlag New York Inc., 1st ed. edition, 2000.
- Proust-Lima C. et Jacqmin-Gadda H. Estimation of linear mixed models with a mixture of distribution for the random-effects. *Computer Methods and Programs in Biomedicine*, 78(2) :165–73, 2005.
- Rabiner L. R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2) :257–286, 1989.
- Robert C. et Casella G. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer, 2004.
- Ruse M. G., Samson A., et Ditlevsen S. Inference for biomedical data by using diffusion models with covariates and mixed effects. *Journal of Royal Statistical Society. Series C*, 2020.
- Schelldorfer J., Buhlmann P., et Van De Geer S. Estimation for High-Dimensional Linear Mixed-Effects Models Using L1-Penalization. *Scandinavian Journal of Statistics*, 38 :197–214, 2011.
- Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*, 6(2) :461–464, 1978.
- Sheiner L. B., Rosenberg B., et Melmon K. L. Modelling of individual pharmacokinetics for computer-aided drug dosage. *Comput. Biomed. Res.*, 5(5) :411–459, 1972.
- Silvapulle M. J. et Sen P. K. *Constrained Statistical Inference. Inequality, Order, and Shape Restrictions*. Wiley Series in Probability and Statistics. Wiley, 2001.
- Strathe A. B., Danfoer A., Nielsen B., Klim S., et Sorensen H. Population based growth curve analysis : a comparison between models based on ordinary or stochastic differential equations implemented in a nonlinear mixed effect framework. In *Modelling nutrient digestion and utilisation in farm animals.*, pages 22–30. Wageningen Academic Publishers, Wageningen, 2011.
- Tornoe C. W., Overgaard R. V., Agerso H., Nielsen H. A., Madsen H., et Jonsson N. Stochastic differential equations in NONMEM : Implementation, application, and comparison with ordinary differential equations. *Pharmaceutical Research*, 22(8) :1247–1258, 2005.
- Trokoniz I. F., Plan E., Miller R., et Karlsson M. O. Modelling overdispersion and Markovian features in count data. *Journal of Pharmacokinetics and Pharmacodynamics*, 36 (5) :461–477, 2009.
- van der Vaart A.W. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2000.

Wei G. C. G. et Tanner M. A. A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms. *Journal of the American Statistical Association*, 85(411) :699–704, 1990.