



HAL
open science

Modèles statistiques à vraisemblance pivotale

Francis Maisonneuve

► **To cite this version:**

| Francis Maisonneuve. Modèles statistiques à vraisemblance pivotale. 2013. hal-00519962v7

HAL Id: hal-00519962

<https://hal.science/hal-00519962v7>

Preprint submitted on 25 May 2013

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.

Modèles statistiques à vraisemblance pivotale

Francis Maisonneuve

MINES ParisTech

Résumé

Quitte à se limiter à quelques modèles statistiques usuels (mais importants) qui vérifient une condition forte de “vraisemblance pivotale”, on peut étendre la méthode des pivots utilisée pour les intervalles de confiance à l’estimation de paramètre (réel ou vectoriel) par *régions les plus vraisemblables*. La procédure proposée ne fait intervenir aucun élément ou choix arbitraire (estimation *intrinsèque*) ; elle est de plus invariante par reparamétrage, à l’instar des estimations ponctuelles au maximum de vraisemblance attachées à ces estimations ensemblistes.

Bien que la démarche soit de facture purement classique (fréquentiste), on vérifiera en dernière partie de l’étude qu’il est possible de *probabiliser a posteriori* l’espace des paramètres de tels modèles à vraisemblance pivotale, au moins de manière partielle en vue de *formuler* les résultats d’estimation, et même complètement pour les exemples considérés, qualifiés de *parfaits*. Allant au-delà de la notion contestée de *probabilité fiduciaire* introduite par R.A. Fisher, on pourra établir pour eux un lien direct avec l’approche bayésienne en dégageant une notion de loi a priori non informative ; on rejoindra à cette occasion l’approche de l’école bayésienne appelée *exact matching prior*.

Abstract

Statistical Models with Pivotal Likelihood

We consider a few common (however important) statistical models satisfying a strong condition of “pivotal likelihood”. For these, we can extend the pivotal method designed for the confidence intervals to the (real or vectorial) parameter estimation based on *highest likelihood regions*. The proposed procedure involve no arbitrary element or choice (*intrinsic* estimation); furthermore it is equivariant under reparameterization, as do maximum likelihood point estimations linked with these set estimations.

Although this work relies only on classical (frequentist) statistics, we verify in the final section of the paper that it is possible to post-probabilize the parameter space of such models having pivotal likelihood, at least partially in order to *express* the estimation results, and even completely for the previous examples that we call *perfect models*. Going further than the highly contested *fiducial probability* introduced by R.A. Fisher, we will establish for them a direct link with Bayesian modelling by identifying a notion of uninformative prior distribution ; we will meet on this occasion the Bayesian approach called *exact matching prior*.

Table des matières

1	Introduction : généralités sur la vraisemblance	3
1.A	Cadre de l'étude	3
1.B	Paramètres utiles et nuisibles	3
2	Vraisemblance pivotale et régions de confiance	4
2.A	L'hypothèse de vraisemblance pivotale	4
2.B	Réflexions sur le choix de l'hypothèse	6
2.C	Statistiques exhaustives et factorisation vraie de ϕ	7
2.D	Cas des modèles exponentiels réels	8
3	Exemples	9
3.A	Une situation asymptotique assez générale	9
3.B	Modèles à paramètre de translation pour T	10
3.C	Modèles réels à paramètre de position pour T	12
3.D	Modèles réels à paramètre d'échelle pour T	13
3.E	Modèles des lois gaussiennes à un paramètre nuisible	16
4	Probabilisation possible de Θ	17
4.A	La famille des pré-probabilités sur Θ sachant x	18
4.B	Factorisation de ϕ et probabilité sur Θ	19
4.C	Lien avec la statistique bayésienne	20
4.D	Exemples de modèles parfaits	22
5	Annexes	24
5.A	Statut de la fonction de répartition de T dans le cas réel	24
5.B	Le modèle de seuillage d'une variable aléatoire	25
5.C	Information a priori en vision fréquentiste ?	27
6	Conclusion	27
	Références	28

1. Introduction : généralités sur la vraisemblance

1.A. Cadre de l'étude

On considère un modèle statistique paramétrique dominé par une mesure σ -finie μ :

Θ ouvert connexe de \mathbb{R}^d ($d \geq 1$), $(\Omega, \mathcal{A}, (P_\theta)_{\theta \in \Theta})$ et X variable aléatoire de loi $f_\theta \mu$ pour P_θ .

Pour ne pas multiplier les notations, un X -échantillon (X_1, \dots, X_n) de taille n sera encore noté X ; et sa densité par rapport à la mesure $\mu^{n \otimes}$ sera encore notée f_θ :

$$x = (x_1, \dots, x_n) \xrightarrow{f_\theta} \prod_{i=1}^n f_\theta(x_i) \text{ (et en notation croisée } f_x \text{ est la fonction de vraisemblance } \theta \mapsto f_\theta(x)).$$

Δ désigne un borélien porteur de μ ($\mu(\Delta^c) = 0$), en général inclus dans le support de μ . On ne considérera couramment que des $x = (x_1, \dots, x_n) \in \Delta^n$.

Rappelons que la (fonction de) vraisemblance sachant x est en fait une notion définie à un facteur multiplicatif (strictement positif) près pour chaque valeur fixée de x , ce qu'on écrit $\cdot \propto f_x$; de sorte qu'elle n'est pas affectée par le remplacement de la mesure dominante μ par une mesure équivalente. Elle exprime essentiellement une relation d'ordre sur Θ (variable avec x), sur laquelle est fondée l'estimation du paramètre θ au maximum de vraisemblance. Elle détermine aussi la famille des *régions les plus vraisemblables* pour θ sachant $x \in \Delta^n$

$$h \geq 0, \{f_x > h\} \subseteq \Theta,$$

qui peuvent constituer d'intéressantes estimations régionales de θ , dans la mesure où on peut leur attribuer des niveaux de confiance justifiés et précis.

Soulignons le caractère essentiellement fréquentiste de cette analyse. Il existe des estimations apparentées en inférence bayésienne, fondées sur les grandes valeurs de la *densité a posteriori du paramètre*, avec comme on le sait l'obligation pour rester cohérent de modifier convenablement la mesure de référence en cas de reparamétrage.

1.B. Paramètres utiles et nuisibles

Dans le cas d'un modèle à paramètre θ vectoriel ($d \geq 2$), il est fréquent de s'intéresser plus spécialement à l'estimation de certaines composantes de θ , les autres ayant des valeurs connues ou non. Précisons les relations entre estimation complète et de telles estimations partielles de θ par régions les plus vraisemblables : pour tout sous-ensemble strict J de $\{1, \dots, d\}$ de cardinal d_J et pour tout $\theta \in \Theta$, soit θ_J et θ_{J^c} le d_J et le d_{J^c} -uplet "projections" de θ selon J et J^c respectivement, constitués de ses composantes d'indices dans J et dans $J^c = \{1, \dots, d\} \setminus J$; de sorte qu'on peut identifier θ au couple (θ_J, θ_{J^c}) .

– Si θ_{J^c} est connu, la famille des régions les plus vraisemblables pour θ_J sachant $x \in \Delta^n$ est

$$h \geq 0, \{f_x(\cdot, \theta_{J^c}) > h\};$$

$\{f_x(\cdot, \theta_{J^c}) > h\}$ est la section à θ_{J^c} de la région la plus vraisemblable pour θ sachant x , $\{f_x > h\}$. Ceci revient, dans le cadre du modèle de référence, à prendre comme fonction de vraisemblance de θ_J la *vraisemblance conditionnelle* à θ_{J^c}

$$\theta_J \mapsto f_x(\theta_J, \theta_{J^c}).$$

Une justification consiste à se placer dans le cadre du modèle de référence "conditionné à θ_{J^c} " :

$$(\Omega, \mathcal{A}, (P_{(\theta_J, \theta_{J^c})})_{\theta_J}) \text{ où } \forall \theta_J, f_{(\theta_J, \theta_{J^c})} \mu \text{ est la loi de } X \text{ pour } P_{(\theta_J, \theta_{J^c})},$$

dont le paramètre θ_J admet la fonction précédente comme vraisemblance sachant x .

- Si θ_{J^c} est inconnu (θ_{J^c} paramètre *nuisible*), la famille des régions les plus vraisemblables pour θ_J sachant $x \in \Delta^n$ est

$$h \geq 0, \bigcup_{\theta_{J^c}} \{f_x(\cdot, \theta_{J^c}) > h\};$$

$\bigcup_{\theta_{J^c}} \{f_x(\cdot, \theta_{J^c}) > h\}$ est la projection selon J de la région la plus vraisemblable pour θ sachant x , $\{f_x > h\}$: on adopte ceci comme région la plus vraisemblable, car à hauteur h fixée, on veut y trouver tous les θ_J tels que $f_x(\theta_J, \theta_{J^c}) > h$ pour au moins une valeur de θ_{J^c} . Ceci revient, dans le cadre du modèle de référence, à prendre comme fonction de vraisemblance de θ_J sachant x la *vraisemblance marginale*, qu'on suppose à valeurs finies,

$$\theta_J \mapsto \sup_{\{\theta_{J^c} : (\theta_J, \theta_{J^c}) \in \Theta\}} f_x(\theta_J, \theta_{J^c}).$$

2. Vraisemblance pivotale et régions de confiance

2.A. L'hypothèse de vraisemblance pivotale

On considère dans la suite un modèle statistique à d paramètres réels tel que

$$\forall x = (x_1, \dots, x_n) \in \Delta^n, M(x) \stackrel{\text{déf}}{=} \|f_x\|_\infty = \sup_{\theta \in \Theta} f_x(\theta) \text{ vérifie } 0 < \|f_x\|_\infty < +\infty;$$

ce qui permet de définir la *vraisemblance normalisée* $\phi = \phi^{(n)} : \Delta^n \times \Theta \rightarrow [0, 1]$ par

$$\forall (x, \theta) \in \Delta^n \times \Theta, \phi_x(\theta) = \frac{1}{\|f_x\|_\infty} f_x(\theta) \text{ (en abrégé } \phi = \frac{f}{M}\text{)}$$

(ou en cas de paramètre nuisible θ_{J^c} , de définir la *vraisemblance marginale normalisée* $\phi = \phi_J^{(n)}$ par

$$\phi_x(\theta_J) = \frac{1}{\|f_x\|_\infty} \sup_{\{\theta_{J^c} : (\theta_J, \theta_{J^c}) \in \Theta\}} f_x(\theta_J, \theta_{J^c}).$$

Définition 2.1 *On dira que le modèle vérifie l'hypothèse de vraisemblance pivotale si ϕ est une fonction pivotale, autrement dit si ϕ vérifie la propriété pivotale qui s'écrit en notation croisée ($\phi_\theta(x) = \phi_x(\theta)$) : $\forall \theta \in \Theta$, la variable aléatoire composée $\phi_\theta(X_1, \dots, X_n)$ (qui dépend de θ) est telle que sa loi $\eta = \eta^{(n)}$ n'en dépend pas :*

$$\forall \theta \in \Theta \text{ et } \forall B \in \mathcal{B}(\mathbb{R}), P_\theta(\phi_\theta(X_1, \dots, X_n) \in B) = \eta(B);$$

ce qu'on note en abrégé $\forall \theta \in \Theta, (\phi_\theta(X))(P_\theta) = \eta$.

η est ainsi la probabilité image de P_θ par $\phi_\theta(X_1, \dots, X_n)$, ou encore celle de la loi $f_\theta \mu^{n \otimes} = M \phi_\theta \mu^{n \otimes}$ de (X_1, \dots, X_n) par ϕ_θ ; elle est portée par $\phi_\theta(\Delta^n)$ pour tout $\theta \in \Theta$, a fortiori par $\phi(\Delta^n \times \Theta) \subseteq [0, 1]$.

L'hypothèse très forte de vraisemblance pivotale ci-dessus, qui exprime que la fonction de vraisemblance normalisée est une fonction pivotale, est semble-t-il originale; les exemples détaillés dans les sections suivantes, où on pourra identifier la loi η , attesteront que cette hypothèse n'est pas totalement saugrenue!

On a d'après la propriété pivotale $\forall \theta \in \Theta$ et $\forall B \in \mathcal{B}(\mathbb{R}), P_\theta(\phi_X^{-1}(B) \ni \theta) = P_\theta(\phi_X(\theta) \in B) = \eta(B)$: $\phi_X^{-1}(B)$ est une *région aléatoire de confiance* pour θ au niveau $\gamma \stackrel{\text{déf}}{=} \eta(B)$. Conformément au point de vue fréquentiste habituel, si on renouvelait un grand nombre de fois l'expérience aléatoire pour obtenir une suite d'échantillons $x^j \in \Delta^n$, on aurait $\phi_{x^j}^{-1}(B) \ni \theta$ avec une fréquence expérimentale voisine de γ ; ce qui conduit pour un échantillon donné $x \in \Delta^n$ à prendre comme estimation ensembliste de θ au niveau

(exact) de confiance γ

$$\Theta_x^B \stackrel{\text{d\u00e9f}}{=} \phi_x^{-1}(B) = \{\phi_x \in B\} = \{\theta \in \Theta : \phi_\theta(x_1, \dots, x_n) \in B\},$$

appel\u00e9 *r\u00e9gion de confiance pour θ sachant l'\u00e9chantillon x au niveau $\gamma = \eta(B)$* , ind\u00e9pendant de x .

On peut noter que le *principe de vraisemblance* n'est pas \u00e0 proprement parler respect\u00e9 dans cette d\u00e9finition, du fait qu'on doit se r\u00e9f\u00e9rer \u00e0 la mesure η pour la d\u00e9termination des niveaux de confiance (voir toutefois le paragraphe 4.C \u00e0 ce sujet).

En fait, pour \u00e9viter tout paradoxe apparent dans l'estimation par r\u00e9gions de confiance, on peut compl\u00e9ter l'hypoth\u00e8se de vraisemblance pivotale par deux conditions sur la mesure η :

- $\forall x \in \Delta^n$, η est port\u00e9e par $\phi_x(\Theta)$, autrement dit $\eta(\phi_x(\Theta)^c) = 0$.

En effet sinon il pourrait exister $B \in \mathcal{B}(\mathbb{R})$ tel que $\eta(B) > 0$ v\u00e9rifiant $B \subseteq \phi_x(\Theta)^c$ pour certains \u00e9chantillons $x \in \Delta^n$; de sorte qu'on aurait pour ceux-ci $\Theta_x^B = \emptyset$: curieuse estimation ensembliste! (cf. \u00e0 ce sujet le paragraphe 3.C). Plus profond\u00e9ment, la valeur du niveau de confiance $\eta(B)$ de Θ_x^B d\u00e9pend ainsi de B seulement via Θ_x^B : en effet $\Theta_x^B = \Theta_x^{B'} \implies \phi_x(\Theta_x^B) = \phi_x(\Theta_x^{B'})$, c'est-\u00e0-dire $B \cap \phi_x(\Theta) = B' \cap \phi_x(\Theta)$; de sorte que

$$\Theta_x^B = \Theta_x^{B'} \implies \eta(B) = \eta(B \cap \phi_x(\Theta)) = \eta(B' \cap \phi_x(\Theta)) = \eta(B').$$

- η diffuse, ou au moins $\eta(\{0\}) = 0$.

La relation $\eta(\{0\}) = 0$ assure que pour tout $x \in \Delta^n$, la r\u00e9gion de vraisemblance nulle $\{\phi_x = 0\}$ a un niveau de confiance nul : ce qui para\u00eet \u00eatre une exigence minimum de coh\u00e9rence entre ces deux notions!

Ces deux conditions se r\u00e9sument en : $\forall x \in \Delta^n$, $\phi_x(\Theta) \cap]0, 1] = \phi_x(\{\phi_x > 0\})$ est porteur de η . Nous dirons alors que ϕ est une fonction pivotale *\u00e9tal\u00e9e*.

Remarque 1 Il est clair que la vraisemblance normalis\u00e9e ϕ , et de l\u00e0 la mesure η , ne d\u00e9pendent ni du choix de la mesure dominante μ du mod\u00e8le, ni d'un reparam\u00e9trage $\theta \in \Theta \xrightarrow{\delta} \tilde{\theta} \in \tilde{\Theta}$, o\u00f9 δ est une bijection bi-mesurable de Θ sur un autre ouvert $\tilde{\Theta}$ de \mathbb{R}^d (au sens o\u00f9 $\phi_{\tilde{\theta}} = \phi_\theta$ et $\tilde{\eta} = \eta$). Et les r\u00e9gions al\u00e9atoires de confiance pour θ au niveau γ d\u00e9termin\u00e9es par ϕ ne d\u00e9pendent pas non plus du choix de la normalisation de ϕ \u00e0 1 plut\u00f4t qu'\u00e0 une autre constante $k > 0$, ou m\u00eame du remplacement de ϕ par $\phi' = \ell \circ \phi$ o\u00f9 $\ell : \mathbb{R} \rightarrow \mathbb{R}$ est un hom\u00e9omorphisme croissant quelconque (au sens o\u00f9 $\{\phi'_x \in B'\} = \{\phi_x \in \ell^{-1}(B')\}$, avec $\eta' = \ell(\eta)$).

Remarque 2 La fonction ϕ est \u00e9tal\u00e9e si $\eta(\{0\}) = 0$ et si on est dans l'un ou l'autre des deux cas suivants :

- $\phi_x(\Theta)$ ne d\u00e9pend pas de $x \in \Delta^n$ (car il co\u00efncide alors avec $\phi(\Delta^n \times \Theta)$ qui porte η);
- $\forall x \in \Delta^n$, $\phi_x(\Theta) \supseteq]0, 1]$, ou $\eta(\{1\}) = 0$ et $\forall x \in \Delta^n$, $\phi_x(\Theta) \supseteq]0, 1[$ (car $]0, 1[$ porte η).

On dispose ainsi pour tout $x \in \Delta^n$ de la famille des *r\u00e9gions de confiance les plus vraisemblables pour θ*

$$y \in]0, 1[, \{\phi_x > y\} \subseteq \Theta, \text{ au niveau } \gamma(y) \stackrel{\text{d\u00e9f}}{=} \eta(]y, 1]) \text{ ind\u00e9pendant de } x.$$

La fonction γ (prolong\u00e9e par les valeurs 1 sur $] - \infty, 0]$ et 0 sur $[1, +\infty[$) est la fonction de r\u00e9partition compl\u00e9mentaire de la probabilit\u00e9 η sur \mathbb{R} , continue en 0 puisque $\eta(\{0\}) = 0$; si elle est continue sur \mathbb{R} (cas o\u00f9 η est diffuse), elle prend en d\u00e9croissant toutes les valeurs de $]0, 1[$ et on a pour tout $x \in \Delta^n$

$$\forall \gamma_0 \in]0, 1[, \{\gamma \circ \phi_x < \gamma_0\} = \{\phi_x > y_0\} \text{ o\u00f9 } y_0 = \max\{y \in \mathbb{R} : \gamma(y) = \gamma_0\} :$$

la (plus petite) r\u00e9gion la plus vraisemblable pour θ sachant x au niveau de confiance γ_0 s'obtient ainsi directement par seuillage du graphe de la fonction $\gamma \circ \phi_x$ dans $\Theta \times \mathbb{R}$ sous la hauteur γ_0 .

En r\u00e9sum\u00e9, sous l'hypoth\u00e8se tr\u00e8s forte de *vraisemblance pivotale (\u00e9tal\u00e9e)*, la famille $(\{\phi_x > y\})_{y \in]0, 1[}$ des r\u00e9gions les plus vraisemblables pour θ sachant x est *intrins\u00e8que* au mod\u00e8le statistique, au sens o\u00f9 la d\u00e9finition de ces r\u00e9gions et de leur niveau de confiance ne comportent aucune part d'arbitraire; contrairement-

ment par exemple aux intervalles de confiance d'un paramètre réel qui dépendent du choix de la fonction pivotale, ainsi que du type d'intervalle retenu.

L'estimation par régions les plus vraisemblables pour θ nous semble "optimale" au sens évident où elle maximise la vraisemblance du paramètre. Comme pour l'estimation ponctuelle au maximum de vraisemblance $\hat{\theta}$ de θ que (l'adhérence de) ces régions contiennent lorsqu'elle existe¹, ce mode d'estimation n'est pas associé à un critère d'optimalité lié à un paramétrage particulier ; contrairement aux estimations classiques fondées sur la minimisation d'un risque ou d'une "taille" impliquant la mesure de Lebesgue sur Θ , ou tout autre étalon non invariant par reparamétrage. Selon qu'on se focalise sur des cas où un paramétrage particulier s'impose ou non, cette invariance apparaîtra comme un défaut ou au contraire comme une qualité structurelle supplémentaire.

2.B. Réflexions sur le choix de l'hypothèse

On part de la relation

$$\forall B \in \mathcal{B}(\mathbb{R}), P_\theta(\phi_X^{-1}(B) \ni \theta) \stackrel{\forall \theta}{=} \eta(B),$$

et pour tout échantillon $x \in \Delta^n$ on attribue à la région $\phi_x^{-1}(B)$ le niveau de confiance $\gamma = \eta(B)$. Ceci suppose impérativement que B ne dépend pas de x ; car si $B = B_x$ on devrait se référer à la quantité $P_\theta(\phi_X^{-1}(B_X) \ni \theta) = P_\theta(\phi_X(\theta) \in B_X)$ qui ne s'exprime plus au moyen de η (même si $\eta(B_x)$ ne dépend pas de x) !

C'est la raison pour laquelle on s'est résolu à poser la très stricte *hypothèse de vraisemblance pivotale* plutôt que le jeu d'hypothèses :

- $\exists \Phi : \Delta^n \times \Theta \rightarrow \mathbb{R}^{d'} \ (1 \leq d' \leq d)$ telle que $\begin{cases} \forall \theta, \Phi_\theta(X_1, \dots, X_n) \text{ est de loi fixe } \nu ; \\ \forall x, \Phi_x(\Theta) \text{ est porteur de } \nu \end{cases}$;
- $\forall x \in \Delta^n, \exists G_x : \Phi_x(\Theta) \subseteq \mathbb{R}^{d'} \rightarrow [0, 1]$ telle que $\phi_x = G_x \circ \Phi_x$, où $\phi = \frac{f}{M}$,

cette dernière hypothèse assurant simplement qu'à x fixé la valeur $\Phi_x(\theta)$ détermine la valeur $\phi_x(\theta)$ pour tout $\theta \in \Theta$; de sorte que $\forall y \in]0, 1[$, la région la plus vraisemblable pour θ , $\{\phi_x > y\}$, s'exprime au moyen de la fonction pivotale Φ par

$$\{\phi_x > y\} = \{\Phi_x \in B_x\} \quad \text{où} \quad B_x = \{G_x > y\}.$$

On peut noter que, d'un point de vue strictement formel, ce jeu d'hypothèses peu exigeantes pourrait suffire pour une partie de l'étude. Mais les niveaux de confiance qu'on attribuerait alors aux régions les plus vraisemblables pour θ sachant x seraient à la fois injustifiés (cf. ci-dessus) et non intrinsèques, du fait de la multiplicité des fonctions pivotales Φ possibles.

Par contre si G_x ne dépend pas de x , $\phi = G \circ \Phi$ est une fonction pivotale, qui est étalée si la loi ν est portée par $\{G > 0\}$ (cf. le paragraphe 2.C).

Une idée plus subtile est la particularisation suivante du jeu d'hypothèses précédent :

- $\exists \Phi : \Delta^n \times \Theta \rightarrow \mathbb{R}$ telle que $\begin{cases} \forall \theta, \Phi_\theta(X_1, \dots, X_n) \text{ est de loi fixe } \nu ; \\ \forall x, \Phi_x(\Theta) \text{ est porteur de } \nu \end{cases}$;
- $\forall x \in \Delta^n, \exists G_x : \mathbb{R} \rightarrow \mathbb{R}$ homéomorphisme croissant tel que $\phi_x = G_x \circ \Phi_x$.

1. on se réfère pour définir $\hat{\theta}$ à la version semi-continue supérieurement de f_x , $\theta \mapsto \limsup_{\theta' \rightarrow \theta} f_x(\theta')$

Il s'agit en effet d'hypothèses moins restrictives que l'hypothèse de vraisemblance pivotale (où $\Phi = \phi$ et $G_x = \text{id}_{\mathbb{R}}$) et telles cependant que la *famille* des régions les plus vraisemblables pour θ sachant $x \in \Delta^n$ est encore de la forme (cf. les paragraphes 1.A et 2.A)

$$z \in \mathbb{R}, \{\Phi_x > z\}, \text{ de niveau de confiance } z \mapsto \nu(\cdot|z, +\infty[) \text{ indépendant de } x.$$

Mais rien ne semble garantir que si (Φ', G') est un autre couple de fonctions vérifiant les hypothèses ci-dessus, donc tel que

$$\Phi'_x = G'_x{}^{-1} \circ \phi_x = G'_x{}^{-1} \circ G_x \circ \Phi_x,$$

l'homéomorphisme croissant composé $\ell_x \stackrel{\text{déf}}{=} G'_x{}^{-1} \circ G_x$ ne dépende pas de x ; or c'est seulement sous cette condition de non dépendance en x qu'on est assuré que les niveaux de confiance attribués aux régions les plus vraisemblables ne dépendent pas du choix particulier du couple (Φ, G) et donc sont bien intrinsèques au modèle.

2.C. Statistiques exhaustives et factorisation vraie de ϕ

On sait qu'une variable $T = \tau(X_1, \dots, X_n)$ à valeurs dans \mathbb{R}^p (couramment $p = d$) est une statistique exhaustive pour le modèle si on a une factorisation de la densité de (X_1, \dots, X_n) de la forme

$$\forall (x_1, \dots, x_n, \theta) \in \Delta^n \times \Theta, f_\theta(x_1, \dots, x_n) = B(x_1, \dots, x_n) f_\theta^T(t_1, \dots, t_p)$$

où $(t_1, \dots, t_p) \stackrel{\text{déf}}{=} t = \tau(x_1, \dots, x_n)$ et où $B > 0$ sur Δ^n . f_θ^T est pour tout $\theta \in \Theta$ la densité de T par rapport à la mesure dominante $\mu' = \tau(B \mu^{n \otimes})$ sur \mathbb{R}^p , image par $\tau = \tau^{(n)}$ de la mesure de densité B par rapport à $\mu^{n \otimes}$. La factorisation s'écrit $f_x = B(x) f_t^T$ avec la notation croisée $f_t^T(\theta) \stackrel{\text{déf}}{=} f_\theta^T(t_1, \dots, t_p)$; elle exprime donc simplement que $f_t^T \propto f_x$ quand $t = \tau(x)$.

En supposant toujours que $\forall x \in \Delta^n, M(x) = \|f_x\|_\infty \in]0, +\infty[$, on a $f_t^T \propto f_x \propto \phi_x$. Ainsi $\forall x, x' \in \Delta^n$ tels que $t = \tau(x) = \tau(x')$, on a $\phi_x \propto \phi_{x'}$ et donc $\phi_x = \phi_{x'}$ du fait de la normalisation. Ceci assure que la fonction ϕ ne dépend de x que via $\tau(x)$:

$$\forall x \in \Delta^n, \phi_x \stackrel{\text{déf}}{=} \phi_t^T \quad \text{où } t = \tau(x),$$

avec pour tout $t \in \tau(\Delta^n)$, $\phi_t^T \propto f_t^T$ et $\|\phi_t^T\|_\infty = 1$. Ainsi $f_x = M(x) \phi_x = M(x) \phi_t^T$: cette dernière factorisation exprime que pour tout $\theta \in \Theta$, la fonction normalisée ϕ_θ^T est exactement la densité de T dans le modèle statistique image par τ du modèle de référence lorsqu'on choisit M comme fonction B — de sorte que $\mu' = \tau(M \mu^{n \otimes})$. Enfin si on note $M^T(t) \stackrel{\text{déf}}{=} \|f_t^T\|_\infty$ pour tout $t \in \tau(\Delta^n)$, on a $M(x) = B(x) M^T(t)$.

L'hypothèse de vraisemblance pivotale étalée s'exprime donc au moyen de T par

$$\forall \theta \in \Theta, (\phi_\theta^T(T))(P_\theta) = \eta \quad \text{et} \quad \forall t \in \tau(\Delta^n), \phi_t^T(\Theta) \cap]0, 1] \text{ est porteur de } \eta;$$

et on peut si besoin remplacer ci-dessus $\tau(\Delta^n)$ par Δ' , où $\Delta' \subseteq \tau(\Delta^n)$ est n'importe quel sous-ensemble porteur de μ' . On dira encore que ϕ^T est une *fonction pivotale étalée* (du modèle image par τ). On retiendra que la relation $\phi_x = \phi_{\tau(x)}^T$ assure qu'une région la plus vraisemblable pour θ ne dépend de l'échantillon x que via la valeur $t = \tau(x)$ de n'importe quelle statistique exhaustive $T = \tau(X_1, \dots, X_n)$:

$$y \in]0, 1[, \{\phi_t^T > y\} \subseteq \Theta, \text{ au niveau de confiance } \gamma(y) = \eta(\cdot|y, 1]) \text{ indépendant de } t;$$

de sorte que cette définition respecte le *principe d'exhaustivité*.

Supposons à présent qu'il existe deux fonctions boréliennes $\Phi^T : \Delta' \times \Theta \rightarrow \mathbb{R}^d$ et $g : \Phi^T(\Delta' \times \Theta) \rightarrow \mathbb{R}_+$ bornée telles que

$$f^T = g \circ \Phi^T \quad \text{où} \quad \begin{cases} \forall \theta, \Phi_\theta^T(T) \text{ a sa loi } \nu \text{ fixe et portée par } \{g > 0\} \\ \forall t, \Phi_t^T \text{ est une bijection de } \{f_t^T > 0\} \text{ sur } \{g > 0\} \end{cases}.$$

Comme $\forall t \in \Delta'$, $f_t^T = g \circ \Phi_t^T$, on a $\|f_t^T\|_\infty = \|g\|_\infty \in]0, +\infty[$ puisque Φ_t^T est une bijection de $\{\phi_t^T > 0\}$ sur $\{g > 0\}$ et g est bornée non nulle; donc en posant $G \stackrel{\text{déf}}{=} \frac{1}{\|g\|_\infty} g : \Phi(\Delta^n \times \Theta) \subseteq \mathbb{R}^d \rightarrow [0, 1]$, on a de manière plus intrinsèque

$$\phi^T = G \circ \Phi^T \quad \text{où} \quad \begin{cases} \forall \theta, \Phi_\theta^T(T) \text{ a sa loi } \nu \text{ fixe et portée par } \{G > 0\} \\ \forall t, \Phi_t^T \text{ est une bijection de } \{\phi_t^T > 0\} \text{ sur } \{G > 0\} \end{cases}.$$

Ceci assure que ϕ^T est une fonction pivotale étalée (ainsi que $f^T = \|g\|_\infty \phi^T$), puisque $\forall \theta$, $\phi_\theta^T(T)$ a sa loi fixe $\eta = (G \circ (\Phi_\theta^T(T)))(P_\theta) = G(\nu)$ portée pour tout t par $G(\{G > 0\}) = \phi_t^T(\{\phi_t^T > 0\}) = \phi_t^T(\Theta) \cap]0, 1]$. Si U est une variable aléatoire de loi ν , alors η est la loi de $G(U)$. On désignera en abrégé ce type de factorisation de ϕ^T par (Φ^T, G) .

On a d'après ci-dessus $\forall \theta \in \Theta$ et $\forall B \in \mathcal{B}(\mathbb{R}^d)$, $P_\theta((\Phi_T^T)^{-1}(B) \ni \theta) = P_\theta(\Phi_T^T(\theta) \in B) = \nu(B)$: $(\Phi_T^T)^{-1}(B)$ est une région aléatoire de confiance pour θ au niveau $\gamma = \nu(B)$. Conformément au point de vue fréquentiste habituel, si on renouvelait un grand nombre de fois l'expérience aléatoire pour obtenir une suite d'échantillons $x^j \in \Delta^n$ déterminant la suite $t^j = \tau(x^j)$, on aurait $(\Phi_{t^j}^T)^{-1}(B) \ni \theta$ avec une fréquence expérimentale voisine de γ ; d'où pour un échantillon donné $x \in \Delta^n$ l'estimation ensembliste de θ au niveau de confiance γ , en posant $t = \tau(x)$:

$$(\Phi_t^T)^{-1}(B) = \{\Phi_t^T \in B\} = \{\theta \in \Theta : \Phi_\theta^T(t_1, \dots, t_p) \in B\}.$$

En particulier pour tout $y \in]0, 1[$, la région la plus vraisemblable pour θ sachant x est

$$\{\phi_x > y\} = \{\phi_t^T > y\} = \{\Phi_t^T \in \{G > y\}\},$$

au niveau de confiance $\gamma(y) = \eta(]y, 1]) = \nu(G > y) = P(G(U) > y)$.

On peut noter que dans le cas où la statistique exhaustive $T = \tau(X_1, \dots, X_n)$ est minimale, la factorisation est "universelle" : en effet pour toute statistique exhaustive $S = \sigma(X_1, \dots, X_n)$, il existe une application borélienne ψ telle que $T = \psi(S)$, de sorte que $\tau = \psi \circ \sigma$; alors $\phi^T = G \circ \Phi^T$ implique que $\phi^S = G \circ \Phi^S$ en posant $\Phi_\theta^S = \Phi_\theta^T \circ \psi$ pour tout θ , car $\phi_\theta^S = \phi_\theta^T \circ \psi$.

2.D. Cas des modèles exponentiels réels

On rappelle qu'il s'agit de modèles réguliers (qu'on dit réels, car $d = 1$) :

μ portée par $\Delta \subseteq \mathbb{R}$ et $\forall \theta \in \Theta$, $f_\theta(x) = \bar{b}(x) e^{\alpha(\theta) \bar{a}(x) + \beta(\theta)}$ où Θ est un intervalle ouvert de \mathbb{R} , dans lesquels α et β sont de classe C^2 sur Θ , $\alpha' \neq 0$ partout, $\bar{a}(X)$ non presque sûrement constant et $\bar{b} > 0$ sur Δ , avec $e^{-\beta(\theta)} = \int \bar{b}(x) e^{\alpha(\theta) \bar{a}(x)} d\mu(x)$.

L'estimateur efficace $T = \frac{1}{n} \sum_{i=1}^n \bar{a}(X_i)$, qui en est une statistique exhaustive minimale, est intrinsèque

au modèle exponentiel, car la fonction \bar{a} dans $f_\theta(x) = \bar{b}(x) e^{\alpha(\theta) \bar{a}(x) + \beta(\theta)}$ ne dépend pas de la mesure dominante μ du modèle est n'est évidemment pas affectée par un reparamétrage. On sait que sa densité $f_\theta^T(t) \stackrel{\text{déf}}{=} e^{n[\alpha(\theta)t + \beta(\theta)]} > 0$ (en prenant $B(x) = \prod_{1 \leq i \leq n} \bar{b}(x_i)$) est telle que pour tout $t \in \mathbb{R}$, f_t^T est strictement

monotone ou unimodale² sur Θ , comme on le voit par le changement de variable canonique $\delta \stackrel{\text{déf}}{=} \alpha$

2. au sens de strictement croissant, puis strictement décroissant

à partir de la concavité stricte de l'application $\tilde{\theta} \mapsto \frac{1}{n} \ln(f_t^T \circ \alpha^{-1})(\tilde{\theta}) = \tilde{\theta} t + \tilde{\beta}(\tilde{\theta})$ sur l'intervalle ouvert image $\tilde{\Theta} = \alpha(\Theta)$: on a en effet le résultat classique, en notant $h(\theta) \stackrel{\text{déf}}{=} -\frac{\beta'(\theta)}{\alpha'(\theta)} = E_\theta(T)$:

$$\tilde{\beta} = \beta \circ \alpha^{-1} \implies \tilde{\beta}'(\tilde{\theta}) = \frac{\beta'(\theta)}{\alpha'(\theta)} = -h(\theta) \implies -\tilde{\beta}''(\tilde{\theta}) = \frac{h'(\theta)}{\alpha'(\theta)} = n \text{var}_\theta T > 0.$$

En conséquence les régions les plus vraisemblables sont toutes des intervalles ouverts.

D'un autre côté la fonction h , dont la dérivée est de signe constant, est un difféomorphisme de Θ sur l'intervalle ouvert image $h(\Theta)$; et $\forall t \in h(\Theta)$, la dérivée de f_t^T s'annule en θ tel que $\alpha'(\theta)t + \beta'(\theta) = 0 \iff \theta = h^{-1}(t)$. On en déduit par élimination (cf. ci-dessus) que $\forall t \in h(\Theta)$, f_t^T est unimodale, de sorte que $h^{-1}(t)$ est l'estimation $\hat{\theta}$ de θ au maximum de vraisemblance.

Supposons que $\Delta' \stackrel{\text{déf}}{=} h(\Theta)$ porte $\mu' = \tau(B \mu^{n\infty})$; alors ϕ^T est bien définie sur $\Delta' \times \Theta$ puisque $\|f_t^T\|_\infty = e^{n[\alpha(h^{-1}(t))t + \beta(h^{-1}(t))]} \in]0, +\infty[$; et $\forall t \in \Delta'$, $\phi_t^T > 0$ est unimodale et atteint son maximum 1 en $\hat{\theta} = h^{-1}(t)$. D'où $\forall (t, \theta) \in h(\Theta) \times \Theta$, $-\frac{1}{n} \ln \phi_\theta^T(t) = [\alpha(h^{-1}(t)) - \alpha(\theta)]t + [\beta(h^{-1}(t)) - \beta(\theta)]$.

Ainsi lorsque $h(\Theta)$ porte μ' , $h^{-1}(T)$ est l'estimateur de θ au maximum de vraisemblance et l'hypothèse de vraisemblance pivotale est vérifiée si et seulement si la loi de $[\alpha(h^{-1}(T)) - \alpha(\theta)]T + [\beta(h^{-1}(T)) - \beta(\theta)]$ ne dépend pas de θ . Alors la loi η de $\phi_\theta^T(T) = \phi_\theta(X_1, \dots, X_n)$ vérifie $\eta(\{0\}) = 0$ puisque $\phi_\theta^T(T) > 0$.

Enfin par exemple la condition

$$\forall t \in h(\Theta), \inf f_t^T(\Theta) = 0$$

suffit à assurer que $\forall t \in h(\Theta)$, $\phi_t^T(\Theta) =]0, 1]$, de sorte que ϕ^T est étalée (cf. la remarque 2 et le paragraphe 2.C). Si $h(\Theta)$ ne portait pas μ' , la fonction f_t^T serait strictement monotone en tout point t d'un sous-ensemble non μ' -négligeable A ; et ϕ_t^T ne serait définie en $t \in A$ que si $\|f_t^T\|_\infty < +\infty$.

3. Exemples

3.A. Une situation asymptotique assez générale

On suppose ici que la fonction f , telle que $\forall x \in \Delta^n$, $M(x) = \|f_x\|_\infty \in]0, +\infty[$, est suffisamment régulière pour que s'applique le *théorème du rapport des vraisemblances maximales*, qui énonce

$$\forall \theta \in \Theta, -2 \ln \frac{f_\theta}{M}(X_1, \dots, X_n) \xrightarrow{\text{CVL}} \chi_d^2 \text{ lorsque } n \text{ tend vers l'infini ;}$$

c'est-à-dire une convergence en loi vers une variable du chi-deux qui ne dépend pas des valeurs de θ ni même du modèle, mais seulement du nombre d de paramètres réels (indépendants) constituant θ qui est égal au nombre de degrés de liberté du chi-deux. $\phi = \frac{f}{M}$ vérifie donc la propriété pivotale de manière asymptotique, puisque la loi η_θ de $\phi_\theta(X_1, \dots, X_n)$ est assimilable pour n assez grand à la loi fixe η de $e^{-\frac{\chi_d^2}{2}}$. Sous réserve par exemple que la régularité de f assure que $\phi_x(\Theta) \supseteq]0, 1[$ pour tout $x \in \Delta^n$ (cf. la remarque 2, η étant diffuse), on a donc que pour tout $\gamma \in]0, 1[$ fixé,

$$\{\phi_x > \gamma \stackrel{\text{déf}}{=} e^{-\frac{z}{2}}\} = \{f_x > h \stackrel{\text{déf}}{=} e^{-\frac{z}{2}} \|f_x\|_\infty\}$$

est la région la plus vraisemblable pour θ sachant x au niveau de confiance asymptotique γ , pour $z > 0$ défini par

$$P(\chi_d^2 < z) = \gamma.$$

3.B. Modèles à paramètre de translation pour T

Soit un modèle tel que l'ouvert $\Theta \subseteq \mathbb{R}^d$ est un groupe localement compact, de mesure de Haar à gauche notée Λ , et soit T une statistique exhaustive (minimale) telle que $\Delta' = \Theta$ et

$$\forall \theta \in \Theta, \theta^{-1} T \stackrel{\mathcal{L}}{=} U \text{ où } U \text{ a pour loi } \nu = g \Lambda \text{ avec la densité } g \text{ bornée.}$$

Comme $T \stackrel{\mathcal{L}}{=} \theta U = \tau_\theta(U)$ où τ_θ désigne l'opérateur de translation de θ à gauche, on a les égalités entre lois $T(P_\theta) = \tau_\theta(\nu) = (g \circ \tau_{\theta^{-1}}) \tau_\theta(\Lambda) = (g \circ \tau_{\theta^{-1}}) \Lambda$; de sorte qu'en choisissant Λ comme mesure dominante μ' (la relation précédente assure en pratique que Λ est dans la bonne classe d'équivalence), on a $f_\theta^T = g \circ \tau_{\theta^{-1}}$, soit $f_\theta^T(t) = g(\theta^{-1} t)$ pour tout (t, θ) . On a donc

$$f^T = g \circ \Phi^T \text{ où } \begin{cases} \forall \theta, \Phi_\theta^T(T) \text{ a sa loi } \nu \text{ fixe et portée par } \{g > 0\} \\ \forall t, \Phi_t^T \text{ est une bijection de } \{f_t^T > 0\} \text{ sur } \{g > 0\} \end{cases} \text{ pour } \Phi^T(t, \theta) \stackrel{\text{déf}}{=} \theta^{-1} t;$$

de sorte qu'en posant $G \stackrel{\text{déf}}{=} \frac{1}{\|g\|_\infty} g$, (Φ^T, G) est une *factorisation* de ϕ^T (cf. le paragraphe 2.C).

Ainsi pour tout $t \in \Delta'$ et pour tout $y \in]0, 1[$, en notant i l'involution $\theta \mapsto \theta^{-1}$ sur Θ ,

$$\{\phi_t^T > y\} = \{\theta \in \mathbb{R} : G(\theta^{-1} t) > y\} = (\tau_t \circ i) (\{G > y\})$$

est la région la plus vraisemblable pour θ au niveau de confiance $\gamma(y) = P(G(U) > y)$.

Remarque 3 Tout reparamétrage $\theta \in \Theta \xrightarrow{\delta} \tilde{\theta} \in \tilde{\Theta}$, où δ est un homéomorphisme de Θ sur un autre ouvert $\tilde{\Theta}$ de \mathbb{R}^d , permet de transporter la structure de groupe de Θ sur $\tilde{\Theta}$; de sorte que $\tilde{\Theta}$ est encore un groupe localement compact de mesure de Haar à gauche $\tilde{\Lambda} = \delta(\Lambda)$ car $\tau_{\tilde{\theta}} \circ \delta = \delta \circ \tau_\theta$. De plus $\tilde{T} \stackrel{\text{déf}}{=} \delta(T)$ est comme T une statistique exhaustive (minimale) telle que $\tilde{\Delta}' = \tilde{\Theta}$ et on a

$$\forall \tilde{\theta} \in \tilde{\Theta}, (\tilde{\theta})^{-1} \tilde{T} = \delta(\theta^{-1} T) \stackrel{\mathcal{L}}{=} \delta(U) \text{ où } \delta(U) \text{ est de loi fixe } \delta(\nu) = \tilde{g} \tilde{\Lambda} \text{ avec } \tilde{g} = g \circ \delta^{-1} \text{ bornée.}$$

L'hypothèse considérée n'est donc pas liée à un paramétrage particulier.

Exemple du modèle à deux paramètres réels des lois gaussiennes

On considère le modèle (exponentiel) usuel des lois gaussiennes

$$\Delta = \mathbb{R}, \mu = \lambda, f_{(\sigma, m)}(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} > 0; \quad (1)$$

soit $f_{(\sigma, m)}(x_1, \dots, x_n) = \left(\frac{1}{\sigma \sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2} = (2\pi)^{-\frac{n}{2}} e^{-\frac{n}{2} (\ln \sigma^2 + \frac{(\bar{x} - m)^2 + s^2}{\sigma^2})}$, en ayant noté $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ et $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$. On veut estimer ensemble m et σ ($\theta = (\sigma, m)$ dans $\Theta = \mathbb{R}_+^* \times \mathbb{R}$).

On voit (et on sait !) que le couple $T = (S, \bar{X})$ est une statistique exhaustive (minimale) du modèle (1) telle que $\Delta' = \Theta$. On va considérer cet ensemble comme un sous-groupe du *groupe affine* $ax + b$, qui est l'ensemble $\mathbb{R}^* \times \mathbb{R}$ muni de la loi de composition interne (non commutative)

$$(a, b)(a', b') \stackrel{\text{déf}}{=} (aa', ab' + b), \text{ dont } (1, 0) \text{ est l'élément neutre et } \left(\frac{1}{a}, -\frac{b}{a}\right) \text{ est l'élément inverse de } (a, b).$$

$$\text{On a } \forall \theta = (\sigma, m) \in \Theta, \theta^{-1} T = \left(\frac{1}{\sigma}, -\frac{m}{\sigma}\right)(S, \bar{X}) = \left(\frac{S}{\sigma}, \frac{\bar{X} - m}{\sigma}\right) \stackrel{\mathcal{L}}{=} \left(\frac{\sqrt{\chi_{n-1}^2}}{\sqrt{n}}, \frac{Z}{\sqrt{n}}\right) \stackrel{\text{déf}}{=} U$$

où Z est une gaussienne centrée réduite indépendante de la variable du chi-deux χ_{n-1}^2 , d'après ce qu'on sait sur les lois des variables indépendantes \bar{X} et S . La variable U admet donc pour densité par rapport à

la mesure de Lebesgue sur Δ' la fonction $u = (u_1, u_2) \mapsto k u_1^{n-2} e^{-\frac{n}{2}u_1^2} e^{-\frac{n}{2}u_2^2}$ où $k = \frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}-1}\sqrt{\pi}\Gamma(\frac{n-1}{2})}$,
 et comme “la” mesure de Haar à gauche Λ sur Δ' est $\frac{1}{u_1^2} du_1 du_2$, on a $g(u_1, u_2) = k u_1^n e^{-\frac{n}{2}(u_1^2+u_2^2)}$; d'où

$$G(u_1, u_2) = \frac{1}{\|g\|_\infty} g(u_1, u_2) = u_1^n e^{-\frac{n}{2}(u_1^2-1+u_2^2)} = e^{-\frac{n}{2}K(u_1, u_2)},$$

en posant $K(u_1, u_2) \stackrel{\text{d\u00e9f}}{=} u_2^2 + \bar{K}(u_1^2)$ où $\forall u > 0, \bar{K}(u) \stackrel{\text{d\u00e9f}}{=} \ln \frac{1}{u} + u - 1$.

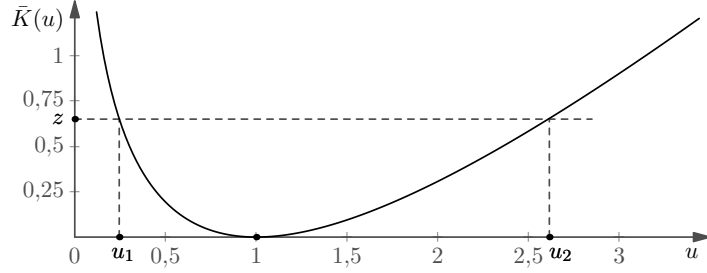


FIGURE 1. Graphe de la fonction \bar{K}

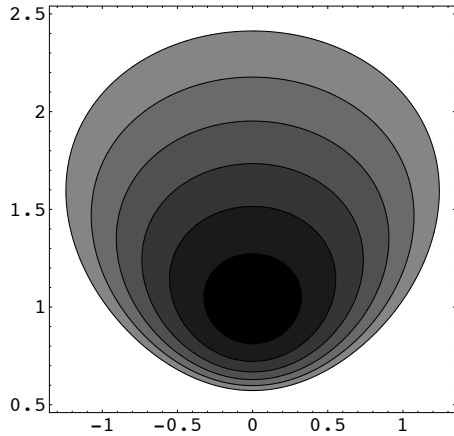
La fonction \bar{K} est positive et strictement convexe sur \mathbb{R}_+^* , de minimum 0 en $u = 1$ et de limite $+\infty$ quand $u \rightarrow 0$ et quand $u \rightarrow +\infty$.

Ainsi pour tout $(s, \bar{x}) \in \mathbb{R}_+^* \times \mathbb{R}$ et pour tout $\gamma \in]0, 1[$ fixé,

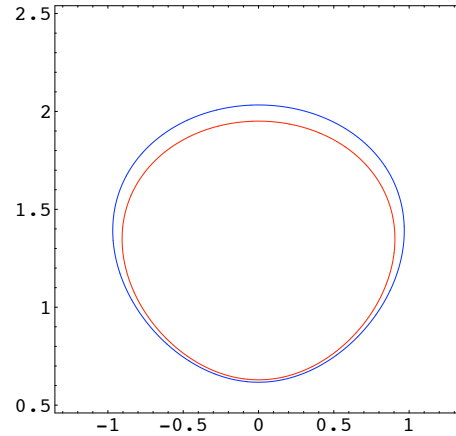
$$\{\phi_{(s, \bar{x})}^T > e^{-\frac{n}{2}z}\} = \left\{ (\sigma, m) \in \Theta : K\left(\frac{s}{\sigma}, \frac{\bar{x} - m}{\sigma}\right) \equiv \frac{(\bar{x} - m)^2 + s^2}{\sigma^2} + \ln \sigma^2 - \ln s^2 - 1 < z \right\}$$

est la région la plus vraisemblable pour (σ, m) au niveau de confiance γ , pour $z > 0$ tel que

$$\gamma = P(G(U) > e^{-\frac{n}{2}z}) = P(K(U) < z) = \nu(\{(u_1, u_2) \in \mathbb{R} \times \mathbb{R}_+^* : u_2^2 + \bar{K}(u_1^2) < z\}).$$



(a) $\{\phi_{(1,0)}^T > e^{-\frac{n}{2}z}\}$ à n fixé pour divers niveaux $\gamma = P(K(U) < z)$



(b) $\{\phi_{(1,0)}^T > e^{-\frac{n}{2}z}\}$ et $\{\phi_{(1,0)}^T > e^{-\frac{1}{2}z_{as}}\}$ à $n = 10$ pour le niveau $\gamma = 0.95$

FIGURE 2. Régions vraisemblables pour (σ, m) sachant $(s, \bar{x}) = (1, 0)$

Application numérique pour $(s, \bar{x}) = (1, 0)$ et $\gamma = 0.95$, dans le cas $n = 10$: on trouve $z \approx 0.661$.

Si on reprend cette étude en considérant (hardiment) qu'on est déjà en situation asymptotique (cf. le paragraphe 3.A), la région la plus vraisemblable $\{\phi_{(1,0)}^T > e^{-\frac{1}{2} z_{as}}\}$ a le niveau de confiance asymptotique γ pour $P(X_d^2 < z_{as}) = \gamma$; ce qui donne numériquement ici (où $d = 2$) $z_{as} \approx 5.99$.

La figure 2 (b) montre ces deux régions les plus vraisemblables, la plus **grande** étant au niveau de confiance γ **exact** et la plus **petite** au même niveau γ **asymptotique**. L'écart est déjà assez faible pour $n = 10$.

3.C. Modèles réels à paramètre de position pour T

C'est un cas particulier de modèle à paramètre de translation pour T , où $\Delta' = \Theta$ est le groupe $(\mathbb{R}, +)$:

$$T - \theta \stackrel{\mathcal{L}}{\equiv} U \text{ où } U \text{ est de loi fixe } \nu = g \lambda \text{ avec la densité } g \text{ bornée,}$$

λ désignant la mesure de Haar – Lebesgue sur \mathbb{R} . Alors pour tout $t \in \mathbb{R}$ et pour tout $y \in]0, 1[$,

$$\{\phi_t^T > y\} = \{\theta \in \mathbb{R} : G(t - \theta) > y\} = t - \{G > y\}$$

est la région la plus vraisemblable pour θ au niveau de confiance $\gamma(y) = P(G(U) > y)$.

Exemple du modèle des exponentielles translatées

On considère le modèle (non régulier) de paramètre de position θ pour X

$$\Delta = \mathbb{R}, \mu = \lambda, f_\theta(x) = e^{-(x-\theta)} \mathbb{1}_{]0, +\infty[}(x - \theta) \text{ où } \Theta = \mathbb{R},$$

tel que $\forall x = (x_1, \dots, x_n), f_\theta(x_1, \dots, x_n) = e^{n\theta - \sum_{i=1}^n x_i} \mathbb{1}_{]-\infty, \min(x_1, \dots, x_n)[}(\theta)$.

Soit $T = \min(X_1, \dots, X_n)$ l'estimateur de θ au maximum de vraisemblance, qui est une statistique exhaustive (minimale) puisque $\forall x \in \Delta^n, f_x \propto (\theta \mapsto e^{n\theta} \mathbb{1}_{]-\infty, t[}(\theta))$.

Comme $\forall i \in \{1, \dots, n\}, X_i - \theta \stackrel{\mathcal{L}}{\equiv} S_1$, où S_1 est la variable exponentielle de paramètre 1, et comme les $X_i - \theta$ sont indépendantes, on a

$$\forall \theta \in \Theta, T - \theta = \min(X_1 - \theta, \dots, X_n - \theta) \stackrel{\mathcal{L}}{\equiv} U \text{ où } U \stackrel{\text{déf}}{=} \frac{S_1}{n} \text{ a pour densité } g(u) = n e^{-nu} \mathbb{1}_{]0, +\infty[}(u).$$

Ainsi $G(u) = e^{-nu} \mathbb{1}_{]0, +\infty[}(u)$, donc $G(U) = e^{-S_1}$ et η est la probabilité uniforme sur $]0, 1[$. On en déduit que $\forall y \in]0, 1[, \gamma(y) = \eta(]y, 1]) = 1 - y$; de sorte que pour tout $t \in \mathbb{R}$ et pour tout $\gamma \in]0, 1[$ fixé,

$$\{\phi_t^T > 1 - \gamma\} = t - \{G > y\} = \left] t - \frac{1}{n} \ln \frac{1}{1 - \gamma}, t \right[$$

est la région la plus vraisemblable pour θ au niveau de confiance γ . On note que cet intervalle de confiance diffère de l'intervalle équilibré habituel

$$\left] t - \frac{1}{n} \ln \frac{2}{1 - \gamma}, t - \frac{1}{n} \ln \frac{2}{1 + \gamma} \right[,$$

qui est de plus grande taille et dont l'adhérence ne contient pas l'estimation au maximum de vraisemblance $\hat{\theta} = t$ de θ .

Remarque 4 La densité $\psi_x(\theta) \stackrel{\text{déf}}{=} f_\theta(x_1, \dots, x_n) = \prod_{1 \leq i \leq n} (e^{-(x_i - \theta)} \mathbb{1}_{]0, +\infty[}(x_i - \theta))$ est elle-même une fonction pivotale puisque pour tout $\theta \in \Theta$, du fait que $\forall i, X_i \stackrel{\text{ps}}{>} \theta$,

$$\psi_X(\theta) \stackrel{\text{ps}}{=} \prod_{1 \leq i \leq n} e^{-(X_i - \theta)} \stackrel{\mathcal{L}}{\equiv} \prod_{1 \leq i \leq n} e^{-S_1^{(i)}} \stackrel{\mathcal{L}}{\equiv} e^{-S_n},$$

où les $S_1^{(i)}$ sont des variables exponentielles de paramètre 1 indépendantes et où S_n est la variable gamma

réduite de paramètre de forme n . Le support de la loi (diffuse) de e^{-S_n} est $[0, 1]$, mais la fonction pivotale ψ n'est pas normalisée et a fortiori n'est pas étalée, puisque le point 1 est à une distance strictement positive de $\psi_x(\Theta) = [0, e^{n \min(x_1, \dots, x_n) - \sum_{i=1}^n x_i}]$ dès que les x_i ne sont pas tous égaux entre eux. Comme analysé au paragraphe 2.A, ceci suffit à rendre contestable l'emploi de cette fonction pivotale pour toute estimation par régions de confiance.

Passons outre pour s'en convaincre : un niveau de confiance $\gamma \in]0, 1[$ étant fixé, soit $y \in]0, 1[$ tel que $P(e^{-S_n} > y) = \gamma$. Pour tout $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, la région la plus vraisemblable pour θ sachant x au niveau de confiance γ ainsi déterminée par ψ serait

$$\{\psi_x > y\} = \left\{ \theta < \min(x_1, \dots, x_n) : e^{n\theta - \sum_{i=1}^n x_i} > y \right\} = \left] \frac{1}{n} \sum_{i=1}^n x_i + \frac{1}{n} \ln y, t \right[.$$

Cet intervalle dépend non seulement de la valeur $t = \min(x_1, \dots, x_n)$ de la statistique exhaustive T , mais aussi de $\frac{1}{n} \sum_{i=1}^n x_i$ via sa borne inférieure ; ce qui paraît saugrenu (cf. à ce sujet le paragraphe 2.C). Et il peut même être vide si d'aventure moyenne et borne inférieure de l'échantillon sont assez différentes !

Exemple du modèle des lois gaussiennes à écart-type connu

On considère le modèle (exponentiel) des lois gaussiennes à écart-type connu σ , de paramètre de position $\theta = m$ pour X

$$\Delta = \mathbb{R}, \mu = \lambda, f_m(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} > 0 \text{ où } \Theta = \mathbb{R}.$$

On sait que l'estimateur efficace $T = \bar{X} \stackrel{\text{déf}}{=} \frac{1}{n} \sum_{i=1}^n X_i$ est une statistique exhaustive minimale, et que $\bar{X} - m \stackrel{\mathcal{L}}{=} U$ où $U = \frac{\sigma}{\sqrt{n}} Z$, Z étant une gaussienne centrée réduite.

On a donc $g(u) = \frac{\sqrt{n}}{\sigma \sqrt{2\pi}} e^{-\frac{nu^2}{2\sigma^2}}$, d'où $G(u) = e^{-\frac{nu^2}{2\sigma^2}}$ et $G(U) = e^{-\frac{z^2}{2}}$. On en déduit

$$\forall z > 0, \{G > e^{-\frac{z^2}{2}}\} = \{u \in \mathbb{R} : nu^2 < \sigma^2 z^2\} = \left] -z \frac{\sigma}{\sqrt{n}}, z \frac{\sigma}{\sqrt{n}} \right[\text{ et } P(G(U) > e^{-\frac{z^2}{2}}) = P(|Z| < z);$$

de sorte que pour tout $\bar{x} \in \mathbb{R}$ et pour tout $\gamma \in]0, 1[$ fixé,

$$\{\phi_{\bar{x}}^{\bar{X}} > e^{-\frac{z^2}{2}}\} = \bar{x} - \{G > e^{-\frac{z^2}{2}}\} = \left] \bar{x} - z \frac{\sigma}{\sqrt{n}}, \bar{x} + z \frac{\sigma}{\sqrt{n}} \right[\text{ pour } z > 0 \text{ tel que } P(|Z| < z) = \gamma$$

(qui est la section à σ d'une région la plus vraisemblable pour (σ, m) , cf. le paragraphe 1.B), est la région la plus vraisemblable pour m au niveau de confiance γ . On retrouve l'intervalle de confiance équilibré habituel, du fait que Z est de loi paire.

3.D. Modèles réels à paramètre d'échelle pour T

C'est un cas particulier de modèle à paramètre de translation pour T , où $\Delta' = \Theta$ est le groupe (\mathbb{R}_+^*, \times) :

$$\frac{T}{\theta} \stackrel{\mathcal{L}}{=} U \text{ où } U > 0 \text{ est de loi fixe } \nu = g \Lambda \text{ avec la densité } g \text{ bornée sur } \mathbb{R}_+^*,$$

Λ désignant la mesure de Haar $\frac{1}{u} du$ sur \mathbb{R}_+^* . Alors pour tout $t \in \mathbb{R}_+^*$ et pour tout $y \in]0, 1[$,

$$\{\phi_t^T > y\} = \left\{ \theta \in \mathbb{R}_+^* : G\left(\frac{t}{\theta}\right) > y \right\} = \left\{ \frac{t}{u} : u \in \mathbb{R}_+^*, G(u) > y \right\}$$

est la région la plus vraisemblable pour θ au niveau de confiance $\gamma(y) = P(G(U) > y)$.

Exemple du modèle des lois uniformes

On considère le modèle (non régulier) de paramètre d'échelle θ pour X

$$\Delta = \mathbb{R}_+^*, \mu = \mathbb{1}_\Delta \lambda, f_\theta(x) = \frac{1}{\theta} \mathbb{1}_{]0,1[}\left(\frac{x}{\theta}\right) \quad \text{où } \Theta = \mathbb{R}_+^*,$$

tel que $\forall x = (x_1, \dots, x_n), f_\theta(x_1, \dots, x_n) = \frac{1}{\theta^n} \mathbb{1}_{]\max(x_1, \dots, x_n), +\infty[}(\theta)$.

Soit $T = \max(X_1, \dots, X_n)$ l'estimateur de θ au maximum de vraisemblance, qui est une statistique exhaustive (minimale) puisque $\forall x \in \Delta^n, f_x = \left(\theta \mapsto \frac{1}{\theta^n} \mathbb{1}_{]t, +\infty[}(\theta)\right)$.

Comme $\forall i \in \{1, \dots, n\}, \frac{X_i}{\theta}$ est une variable uniforme sur $]0, 1[$ et comme les $\frac{X_i}{\theta}$ sont indépendantes, $\forall \theta \in \Theta, \frac{T}{\theta} = \max\left(\frac{X_1}{\theta}, \dots, \frac{X_n}{\theta}\right) \stackrel{\mathcal{L}}{=} U$ où U a pour densité $g(u) = u^n \mathbb{1}_{]0,1[}(u)$ par rapport à Λ .

Ainsi $G(u) = u^n \mathbb{1}_{]0,1[}(u)$, donc la loi η de $G(U)$ est la probabilité uniforme sur $]0, 1[$. On en déduit que $\forall y \in]0, 1[, \gamma(y) = P(G(U) > y) = 1 - y$; de sorte que pour tout $t \in \mathbb{R}_+^*$ et pour tout $\gamma \in]0, 1[$ fixé,

$$\{\phi_t^T > 1 - \gamma\} = \left\{ \frac{t}{u} : u \in \mathbb{R}_+^*, G(u) > y \right\} = \left] t, \frac{t}{(1-\gamma)^{\frac{1}{n}}} \right[$$

est la région la plus vraisemblable pour θ au niveau de confiance γ . On note que cet intervalle de confiance diffère de l'intervalle équilibré habituel

$$\left] t \left(\frac{2}{1+\gamma} \right)^{\frac{1}{n}}, t \left(\frac{2}{1-\gamma} \right)^{\frac{1}{n}} \right[,$$

qui est de plus grande taille et dont l'adhérence ne contient pas l'estimation au maximum de vraisemblance $\hat{\theta} = t$ de θ .

Remarque 5 La fonction $\psi_x(\theta) \stackrel{\text{déf}}{=} x_1 \cdots x_n f_\theta(x_1, \dots, x_n) = \prod_{1 \leq i \leq n} \left[\frac{x_i}{\theta} \mathbb{1}_{]0,1[}\left(\frac{x_i}{\theta}\right) \right]$ est elle-même une fonction pivotale puisque pour tout $\theta \in \Theta$, du fait que $\forall i, X_i \stackrel{\text{ps}}{<} \theta$,

$$\psi_X(\theta) \stackrel{\text{ps}}{=} \prod_{1 \leq i \leq n} \frac{X_i}{\theta} \stackrel{\mathcal{L}}{=} \prod_{1 \leq i \leq n} U^{(i)} \stackrel{\mathcal{L}}{=} \prod_{1 \leq i \leq n} e^{-S_1^{(i)}} \stackrel{\mathcal{L}}{=} e^{-S_n}$$

où les $U^{(i)}$ sont des variables uniformes sur $]0, 1[$ indépendantes. Le support de la loi (diffuse) de e^{-S_n} est $[0, 1]$, mais comme dans l'exemple du modèle des exponentielles translatées (cf. le paragraphe 3.C), la fonction pivotale ψ n'est pas normalisée et a fortiori n'est pas étalée puisque le point 1 est à une distance strictement positive de $\psi_x(\Theta) = \left[0, \frac{x_1 x_2 \cdots x_n}{\max(x_1, \dots, x_n)^n} \right]$ dès que les x_i ne sont pas tous égaux entre eux. Ceci suffit à rendre contestable l'emploi de cette fonction pivotale pour toute estimation par région de confiance.

Exemple du modèle des lois gamma – Weibull

Soient deux réels $a, b > 0$ fixés; on considère le modèle de paramètre d'échelle θ pour X

$$\Delta = \mathbb{R}_+^*, \mu = \mathbb{1}_\Delta \lambda, f_\theta(x) = \frac{b}{\Gamma(a)} \frac{1}{\theta} \left(\frac{x}{\theta}\right)^{ab-1} e^{-\left(\frac{x}{\theta}\right)^b} \quad \text{où } \Theta = \mathbb{R}_+^*,$$

Γ étant la fonction d'Euler. Il s'agit d'un modèle exponentiel standard $f_\theta(x) = \bar{b}(x) e^{\alpha(\theta) \bar{a}(x) + \beta(\theta)}$ avec

$$\bar{a}(x) = x^b, \bar{b}(x) = \frac{b}{\Gamma(a)} x^{ab-1}, \alpha(\theta) = -\frac{1}{\theta^b} \text{ et } \beta(\theta) = -ab \ln \theta.$$

Pour $b = 1$ il s'agit des lois gamma (non réduites) de paramètre de forme a ; pour $a = 1$ il s'agit des lois de Weibull de paramètre de forme b ; et pour $a = b = 1$, on a affaire au modèle des durées de vie exponentielles.

Comme $\forall i \in \{1, \dots, n\}$, $\frac{X_i}{\theta}$ a pour densité $y \mapsto \frac{b}{\Gamma(a)} y^{a-b-1} e^{-y^b} \mathbb{1}_{]0, +\infty[}(y)$, $\left(\frac{X_i}{\theta}\right)^b$ a pour densité $z \mapsto \frac{1}{\Gamma(a)} z^{a-1} e^{-z} \mathbb{1}_{]0, +\infty[}(z)$, c'est-à-dire $\left(\frac{X_i}{\theta}\right)^b \stackrel{\mathcal{L}}{\equiv} S_a$ où S_a est la variable gamma réduite de paramètre de forme a . Comme les $\frac{X_i}{\theta}$ sont indépendantes, la statistique exhaustive minimale $T = \frac{1}{n} \sum_{i=1}^n X_i^b$ vérifie donc $\frac{T}{\theta^b} = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i}{\theta}\right)^b \stackrel{\mathcal{L}}{\equiv} U$ où $U = \frac{S_{na}}{n}$; de sorte que θ^b (et non θ) est un paramètre d'échelle pour T .

Varions la méthode pour la détermination de G : comme $h(\theta) = -\frac{\beta'(\theta)}{\alpha'(\theta)} = \frac{ab}{\theta} \frac{\theta^{b+1}}{b} = a\theta^b$ (cf. le paragraphe 2.D), on a $h(\Theta) = \mathbb{R}_+^*$ avec $h^{-1}(t) = \left(\frac{t}{a}\right)^{\frac{1}{b}}$; de sorte que $\forall (t, \theta) \in h(\Theta) \times \Theta$, on a $-\frac{1}{n} \ln \phi_{\theta}^T(t) = \left(-\frac{a}{t} + \frac{1}{\theta^b}\right)t + \left(a \ln \frac{a}{t} + ab \ln \theta\right) = a\bar{K}\left(\frac{1}{a\theta^b}t\right)$ et donc $\phi_{\theta}^T(t) = G\left(\frac{t}{\theta^b}\right)$ en ayant posé $G(u) \stackrel{\text{déf}}{=} e^{-na\bar{K}\left(\frac{u}{a}\right)}$ (la fonction \bar{K} a été définie au paragraphe 3.B).

Ainsi $G(U) = e^{-na\bar{K}\left(\frac{S_{na}}{na}\right)}$ et on obtient pour $u_1 < u_2$ définis pour tout $z > 0$ par $\bar{K}(u_1) = \bar{K}(u_2) = z$,

$$\{G > e^{-naz}\} = \left\{u \in \mathbb{R}_+^* : \bar{K}\left(\frac{u}{a}\right) < z\right\} =]au_1, au_2[$$

$$\text{et } P(G(U) > e^{-naz}) = P\left(\bar{K}\left(\frac{S_{na}}{na}\right) < z\right) = P\left(u_1 < \frac{S_{na}}{na} < u_2\right);$$

de sorte que pour tout $t \in \mathbb{R}_+^*$ et pour tout $\gamma \in]0, 1[$ fixé,

$$\{\phi_t^T > e^{-naz}\} = \left\{\left(\frac{t}{u}\right)^{\frac{1}{b}} : u \in \mathbb{R}_+^*, G(u) > e^{-naz}\right\} = \left]\left(\frac{t}{au_2}\right)^{\frac{1}{b}}, \left(\frac{t}{au_1}\right)^{\frac{1}{b}}\right[$$

est la région la plus vraisemblable pour θ au niveau de confiance γ , pour $u_1 < u_2$ définis par

$$\bar{K}(u_1) = \bar{K}(u_2) \text{ et } P\left(u_1 < \frac{S_{na}}{na} < u_2\right) = \gamma.$$

On note à nouveau que cet intervalle de confiance au niveau γ diffère, déjà dans le cas $a = b = 1$ des durées de vie exponentielles, de l'intervalle de confiance équilibré habituel, à savoir

$$\left]\frac{\bar{x}}{u_2}, \frac{\bar{x}}{u_1}\right[\text{ où } P(S_n \leq nu_1) = P(S_n \geq nu_2) = \frac{1-\gamma}{2}.$$

Exemple du modèle des lois gaussiennes à moyenne connue

On considère le modèle (exponentiel) des lois gaussiennes à moyenne connue m , de paramètre d'échelle $\theta = \sigma$ pour $X - m$

$$\Delta = \mathbb{R}, \mu = \lambda, f_{\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} > 0 \text{ où } \Theta = \mathbb{R}_+^*.$$

Par définition de la loi du χ^2 , la statistique exhaustive minimale $V \stackrel{\text{déf}}{=} \frac{1}{n} \sum_{i=1}^n (X_i - m)^2$ vérifie $\frac{V}{\sigma^2} \stackrel{\mathcal{L}}{\equiv} U$

où $U = \frac{\chi_n^2}{n}$, χ_n^2 étant la variable du chi-deux à n degrés de liberté; de sorte que σ^2 (et non $\theta = \sigma$) est un paramètre d'échelle pour V .

Varions encore la méthode pour la détermination de G : comme V a pour densité (par rapport à μ') $f_{\sigma}^V(v) = e^{-\frac{n}{2}\left(\ln \sigma^2 + \frac{v}{\sigma^2}\right)} = e^{-\frac{n}{2}\bar{K}\left(\frac{v}{\sigma^2}\right)} e^{-\frac{n}{2}(1+\ln v)}$, on a $\phi_{\sigma}^V(\sigma) = e^{-\frac{n}{2}\bar{K}\left(\frac{v}{\sigma^2}\right)} = G\left(\frac{v}{\sigma^2}\right)$ pour $G \stackrel{\text{déf}}{=} e^{-\frac{n}{2}\bar{K}}$.

Ainsi $G(U) = e^{-\frac{n}{2}\bar{K}\left(\frac{\chi_n^2}{n}\right)}$ et on obtient pour $u_1 < u_2$ définis pour tout $z > 0$ par $\bar{K}(u_1) = \bar{K}(u_2) = z$,

$$\{G > e^{-\frac{n}{2}z}\} = \{\bar{K} < z\} =]u_1, u_2[\text{ et } P(G(U) > e^{-\frac{n}{2}z}) = P\left(\bar{K}\left(\frac{\chi_n^2}{n}\right) < z\right) = P\left(u_1 < \frac{\chi_n^2}{n} < u_2\right);$$

de sorte que pour tout $t \in \mathbb{R}_+^*$ et pour tout $\gamma \in]0, 1[$ fixé,

$$\{\phi_v^V > e^{-\frac{n}{2}z}\} = \left\{ \left(\frac{v}{u}\right)^{\frac{1}{2}} : u \in \mathbb{R}_+^*, G(u) > e^{-\frac{n}{2}z} \right\} = \left] \sqrt{\frac{v}{u_2}}, \sqrt{\frac{v}{u_1}} \right[$$

(qui est la section à m d'une région la plus vraisemblable pour (σ, m) , cf. le paragraphe 1.B), est la région la plus vraisemblable pour σ au niveau de confiance γ , pour $u_1 < u_2$ définis par

$$\bar{K}(u_1) = \bar{K}(u_2) \text{ et } P\left(u_1 < \frac{\chi_n^2}{n} < u_2\right) = \gamma.$$

On note à nouveau que cet intervalle de confiance au niveau γ diffère de l'intervalle de confiance équilibré habituel

$$\left] \sqrt{\frac{v}{u_2}}, \sqrt{\frac{v}{u_1}} \right[\text{ où } P(\chi_n^2 \leq n u_1) = P(\chi_n^2 \geq n u_2) = \frac{1-\gamma}{2}.$$

3.E. Modèles des lois gaussiennes à un paramètre nuisible

On va être confronté ici à des factorisations de ϕ^T moins triviales.

Cas de l'estimation de la moyenne à écart-type inconnu

Comme expliqué au paragraphe 1.B, on doit considérer la fonction de vraisemblance marginale pour m sachant x , $m \mapsto \sup_{\sigma > 0} f_x(\sigma, m)$. On a en l'occurrence pour tout $m \in \Theta_2 = \mathbb{R}$, du fait que la dérivée de

$\sigma \mapsto \ln \sigma^2 + \frac{(\bar{x} - m)^2 + s^2}{\sigma^2}$ s'annule pour $\sigma^2 = (\bar{x} - m)^2 + s^2$ (cf. le modèle (1) du paragraphe 3.B) :

$$\begin{aligned} \sup_{\sigma > 0} f_x(\sigma, m) &= (2\pi)^{-\frac{n}{2}} e^{-\frac{n}{2} \inf_{\sigma > 0} \left(\ln \sigma^2 + \frac{(\bar{x} - m)^2 + s^2}{\sigma^2} \right)} = (2\pi)^{-\frac{n}{2}} e^{-\frac{n}{2} (\ln[(\bar{x} - m)^2 + s^2] + 1)} \\ &= (2\pi e)^{-\frac{n}{2}} \left((\bar{x} - m)^2 + s^2 \right)^{-\frac{n}{2}} = (2\pi e s^2)^{-\frac{n}{2}} \left(1 + \frac{(\bar{x} - m)^2}{s^2} \right)^{-\frac{n}{2}}. \end{aligned}$$

Ainsi pour $T = (S, \bar{X})$ statistique exhaustive minimale du modèle (1), la vraisemblance marginale normalisée $\phi_{(s, \bar{x})}^T \propto \sup_{\sigma > 0} f_x(\cdot, \sigma)$ vaut $\phi_{(s, \bar{x})}^T(m) = \left(1 + \frac{(\bar{x} - m)^2}{s^2} \right)^{-\frac{n}{2}}$; donc pour $G(u) \stackrel{\text{déf}}{=} \left(1 + \frac{u^2}{n-1} \right)^{-\frac{n}{2}}$

et $\Phi_m^T(s, \bar{x}) \stackrel{\text{déf}}{=} \frac{\bar{x} - m}{s} \sqrt{n-1}$, on a

$$\phi^T = G \circ \Phi^T \text{ où } \begin{cases} \forall m, \Phi_m^T(S, \bar{X}) \text{ a sa loi } \nu \text{ fixe et portée par } \{G > 0\} = \mathbb{R} \\ \forall (s, \bar{x}), \Phi_{(s, \bar{x})}^T \text{ est une bijection de } \{\phi_{(s, \bar{x})}^T > 0\} = \mathbb{R} \text{ sur } \{G > 0\} \end{cases};$$

en effet comme $\forall (\sigma, m) \in \Theta$, $\left(\frac{S}{\sigma}, \frac{\bar{X} - m}{\sigma} \right) \stackrel{\mathcal{L}}{=} \left(\frac{\sqrt{\chi_{n-1}^2}}{\sqrt{n}}, \frac{Z}{\sqrt{n}} \right)$ où Z et χ_{n-1}^2 sont indépendantes, on a

$$\Phi_m^T(S, \bar{X}) = \frac{\bar{X} - m}{S} \sqrt{n-1} \stackrel{\mathcal{L}}{=} \frac{Z}{\sqrt{\frac{\chi_{n-1}^2}{n-1}}} \stackrel{\mathcal{L}}{=} T_{n-1}$$

où T_{n-1} désigne la variable de Student à $n-1$ degrés de liberté ($n \geq 2$).

Ainsi pour tout $\bar{x} \in \mathbb{R}$ et pour tout $\gamma \in]0, 1[$ fixé,

$\{\phi_{(s,\bar{x})}^T > G(z)\} = \{|\Phi_{(s,\bar{x})}^T| < z\} = \left\{m \in \Theta_2 : \left| \frac{\bar{x} - m}{s} \sqrt{n-1} \right| < z \right\} = \left] \bar{x} - z \frac{s}{\sqrt{n-1}}, \bar{x} + z \frac{s}{\sqrt{n-1}} \right[$
 (qui est la projection sur Θ_2 d'une région la plus vraisemblable pour (σ, m) , cf. le paragraphe 1.B), est la région la plus vraisemblable pour m au niveau de confiance γ pour $z > 0$ tel que $P(|T_{n-1}| < z) = \gamma$. On retrouve l'intervalle de confiance équilibré habituel au niveau γ , du fait que T_{n-1} est de loi paire.

Cas de l'estimation de l'écart-type à moyenne inconnue

On considère cette fois la fonction de vraisemblance marginale pour σ sachant x , $\sigma \mapsto \sup_{m \in \mathbb{R}} f_x(\sigma, m)$.

On a en l'occurrence pour tout $\sigma \in \Theta_1 = \mathbb{R}_+^*$,

$$\sup_{m \in \mathbb{R}} f_x(\sigma, m) = (2\pi)^{-\frac{n}{2}} e^{-\frac{n}{2}} \inf_{m \in \mathbb{R}} \left(\ln \sigma^2 + \frac{s^2 + (\bar{x} - m)^2}{\sigma^2} \right) = (2\pi)^{-\frac{n}{2}} e^{-\frac{n}{2}} \left(\ln \sigma^2 + \frac{s^2}{\sigma^2} \right) = (2\pi)^{-\frac{n}{2}} e^{-\frac{n}{2}} \left[\bar{K} \left(\frac{s^2}{\sigma^2} \right) + 1 + \ln s^2 \right].$$

Ainsi, en ne retenant que S dans la statistique $T = (S, \bar{X})$, la vraisemblance marginale normalisée $\phi_s^S \propto \sup_{m \in \mathbb{R}} f_x(m, \cdot)$ vaut $\phi_s^S(\sigma) \stackrel{\text{d\u00e9f}}{=} e^{-\frac{n}{2}} \bar{K} \left(\frac{s^2}{\sigma^2} \right)$; on a donc pour $G \stackrel{\text{d\u00e9f}}{=} e^{-\frac{n}{2}} \bar{K}$ et $\Phi_\sigma^S(s) \stackrel{\text{d\u00e9f}}{=} \frac{s^2}{\sigma^2}$,

$$\phi^S = G \circ \Phi^S \quad \text{o\u00f9} \quad \begin{cases} \forall \sigma, \Phi_\sigma^S(S) \text{ a sa loi } \nu \text{ fixe et port\u00e9e par } \{G > 0\} = \mathbb{R}_+^* \\ \forall s, \Phi_s^S \text{ est une bijection de } \{\phi_s^S > 0\} = \mathbb{R}_+^* \text{ sur } \{G > 0\} \end{cases},$$

puisque $\forall \sigma \in \Theta_1, \Phi_\sigma^S(S) = \frac{S^2}{\sigma^2} \stackrel{\mathcal{L}}{=} \frac{\chi_{n-1}^2}{n}$. Ainsi pour tout $x \in \mathbb{R}^n$ et pour tout $\gamma \in]0, 1[$ fixé,

$$\{\phi_s^S > e^{-\frac{n}{2}} z\} = \left\{ \sigma \in \Theta_1 : \bar{K} \left(\frac{s^2}{\sigma^2} \right) < z \right\} = \left] \frac{s}{\sqrt{u_2}}, \frac{s}{\sqrt{u_1}} \right[$$

(qui est la projection sur Θ_1 d'une région la plus vraisemblable pour (σ, m) , cf. le paragraphe 1.B), est la région la plus vraisemblable pour σ au niveau de confiance γ , pour $u_1 < u_2$ définis par

$$\bar{K}(u_1) = \bar{K}(u_2) = z \quad \text{et} \quad P \left(u_1 < \frac{\chi_{n-1}^2}{n} < u_2 \right) = \gamma.$$

On note à nouveau que cet intervalle de confiance au niveau γ diffère de l'intervalle de confiance équilibré habituel

$$\left] \frac{s}{\sqrt{u_2}}, \frac{s}{\sqrt{u_1}} \right[\quad \text{o\u00f9} \quad P(\chi_{n-1}^2 \leq n u_1) = P(\chi_{n-1}^2 \geq n u_2) = \frac{1 - \gamma}{2}.$$

4. Probabilisation possible de Θ

Un des reproches faits par les tenants du paradigme bayésien à l'estimation par intervalles en statistique classique est de ne pas pouvoir exprimer les résultats obtenus sous la forme naturelle à laquelle on pense, à savoir "le paramètre se trouve dans tel intervalle de confiance (déterministe) avec telle probabilité", du fait du refus de principe de considérer le paramètre comme aléatoire.

On va montrer qu'il est possible, dans le cadre des modèles à vraisemblance pivotale étalée, de déterminer pour chaque échantillon $x \in \Delta^n$ une probabilité intrinsèque sur Θ qui résout formellement cette difficulté; toutefois cette probabilité dépendant de x ne sera définie que sur la sous-tribu de la tribu de Borel $\mathcal{B}(\Theta)$ engendrée par les régions les plus vraisemblables pour θ sachant x .

Prolonger cette *pré-probabilité sachant x* à toute la tribu de Borel $\mathcal{B}(\Theta)$ est facile lorsque la vraisemblance pivotale se factorise, mais se pose alors un problème d'interprétation; de plus on risque d'avoir une multiplicité de prolongements possibles quand il y a plus d'un paramètre réel ($d \geq 2$), ce qui constitue une difficulté de fond. On peut voir ces résultats comme une généralisation de la très discutée *probabi-*

lité fiduciaire introduite par R.A. Fisher ; l'introduction d'une condition supplémentaire permettra, nous semble-t-il, d'échapper à certaines difficultés inhérentes à cette notion.

4.A. La famille des pré-probabilités sur Θ sachant x

On considère ici un modèle de dimension $d \geq 1$ quelconque, vérifiant l'hypothèse de vraisemblance pivotale étalée.

Proposition 4.1 : *Pour tout échantillon $x \in \Delta^n$, il existe sur la sous-tribu $\phi_x^{-1}(\mathcal{B}(\mathbb{R}))$ de $\mathcal{B}(\Theta)$ engendrée par les régions les plus vraisemblables sachant x une (unique) probabilité $\check{\Pi}_x$ portée par $\{\phi_x > 0\}$ et intrinsèque au modèle, qu'on appellera pré-probabilité sur Θ sachant x , caractérisée par*

$$\forall B \in \mathcal{B}(\mathbb{R}), \check{\Pi}_x(\phi_x^{-1}(B)) \stackrel{\text{déf}}{=} \eta(B), \text{ c'est-à-dire } \phi_x(\check{\Pi}_x) = \eta ;$$

de sorte que $\forall y \in]0, 1[$, le niveau de confiance $\gamma(y) = \eta(]y, 1])$ de la région la plus vraisemblable $\phi_x^{-1}(]y, 1]) = \{\phi_x > y\}$ coïncide avec sa pré-probabilité $\check{\Pi}_x(\{\phi_x > y\})$.

Preuve On a en toute généralité $B \cap \phi_x(\Theta) = \phi_x(\phi_x^{-1}(B))$; donc comme η est portée par $\phi_x(\Theta)$, on a

$$\forall B \in \mathcal{B}(\mathbb{R}), \eta(B) = \eta(B \cap \phi_x(\Theta)) = \eta[\phi_x(\phi_x^{-1}(B))],$$

ce qui permet bien de définir $\check{\Pi}_x(\phi_x^{-1}(B))$ comme indiqué. De plus, si $(\phi_x^{-1}(B_n))_{n \in \mathbb{N}}$ est une suite de boréliens de Θ deux à deux disjoints, les $B_n \cap \phi_x(\Theta)$ sont aussi deux à deux disjoints puisque pour $n \neq m$,

$$B_n \cap B_m \cap \phi_x(\Theta) = \phi_x(\phi_x^{-1}(B_n \cap B_m)) = \phi_x(\phi_x^{-1}(B_n) \cap \phi_x^{-1}(B_m)) = \phi_x(\emptyset) = \emptyset.$$

Comme en toute généralité $\{\phi_x \in B\} = \phi_x^{-1}(B) = \phi_x^{-1}(B \cap \phi_x(\Theta))$, on a

$$\biguplus_{n \in \mathbb{N}} \phi_x^{-1}(B_n) = \phi_x^{-1}\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \phi_x^{-1}\left(\left(\bigcup_{n \in \mathbb{N}} B_n\right) \cap \phi_x(\Theta)\right) = \phi_x^{-1}\left(\biguplus_{n \in \mathbb{N}} (B_n \cap \phi_x(\Theta))\right) ;$$

donc $\check{\Pi}_x\left(\biguplus_{n \in \mathbb{N}} \phi_x^{-1}(B_n)\right) = \eta\left(\biguplus_{n \in \mathbb{N}} (B_n \cap \phi_x(\Theta))\right) = \sum_{n=0}^{\infty} \eta(B_n \cap \phi_x(\Theta)) = \sum_{n=0}^{\infty} \check{\Pi}_x(\phi_x^{-1}(B_n))$ par définition de $\check{\Pi}_x$, et on a montré que $\check{\Pi}_x$ est σ -additive. Enfin $\check{\Pi}_x$ est positive, $\check{\Pi}_x(\{\phi_x = 0\}) = \eta(\{0\}) = 0$ et $\check{\Pi}_x(\Theta) = \eta([0, 1]) = 1$, donc $\check{\Pi}_x$ est bien une probabilité portée par $\{\phi_x > 0\}$; et on a par définition $\eta = \phi_x(\check{\Pi}_x)$, probabilité image de $\check{\Pi}_x$ par ϕ_x .

Comme la fonction ϕ et la mesure η ne dépendent pas du choix de la mesure dominante μ du modèle, les $\check{\Pi}_x$ n'en dépendent pas non plus. On a la même chose en cas de remplacement de ϕ par $\phi' = \ell \circ \phi$ où $\ell : \mathbb{R} \rightarrow \mathbb{R}$ est un homéomorphisme croissant ; et on contrôle aussi la propriété d'invariance par reparamétrage pour les $\check{\Pi}_x$. \square

Remarque 6 On note la relation de dualité entre les variables x et θ déterminée par la vraisemblance pivotale ϕ

$$\phi_x(\check{\Pi}_x) \stackrel{\forall x}{=} \eta \stackrel{\forall \theta}{=} \phi_\theta(X(P_\theta)).$$

Remarque 7 Lorsque η est diffuse, on a $\phi_x(\check{\Pi}_x) = \eta \implies (\gamma \circ \phi_x)(\check{\Pi}_x) = \gamma(\eta) = \bar{\lambda}$ où $\bar{\lambda}$ désigne ici la probabilité uniforme sur $]0, 1[$ (propriété d'anamorphose dans la méthode d'inversion), puisque γ est la fonction de répartition complémentaire de η (cf. le paragraphe 2.A). Cette relation exprime que pour tout $\gamma_0 \in]0, 1[$, le niveau de confiance γ_0 de la région la plus vraisemblable $\{\gamma \circ \phi_x < \gamma_0\}$ vaut $\check{\Pi}_x(\{\gamma \circ \phi_x < \gamma_0\})$.

Remarque 8 Dans le cas où on dispose d'une statistique exhaustive pour le modèle, on sait (cf. le paragraphe 2.C) que $\forall x \in \Delta^n$, $\phi_x = \phi_t^T$ où $t = \tau(x)$. Comme par définition $\check{\Pi}_x$ est pour tout $x \in \Delta^n$ l'unique probabilité sur la tribu $\phi_x^{-1}(\mathcal{B}(\mathbb{R}))$ de Θ telle que

$$\phi_x(\check{\Pi}_x) = \eta, \text{ autrement dit ici } \phi_t^T(\check{\Pi}_x) = \eta,$$

la pré-probabilité $\check{\Pi}_x$ sachant x ne dépend en fait que de $t = \tau(x)$; on la note en conséquence aussi $\check{\Pi}_t^T$, telle que $\phi_t^T(\check{\Pi}_t^T) = \eta$.

4.B. Factorisation de ϕ et probabilité sur Θ

On suppose de plus comme au paragraphe 2.C qu'il existe une statistique exhaustive minimale T et deux fonctions boréliennes $\Phi^T : \Delta' \times \Theta \rightarrow \mathbb{R}^d$ et $G : \Phi^T(\Delta' \times \Theta) \subseteq \mathbb{R}^d \rightarrow [0, 1]$ telles que

$$\phi^T = G \circ \Phi^T \quad \text{où} \quad \begin{cases} \forall \theta, \Phi_\theta^T(T) \stackrel{\mathcal{L}}{=} U \text{ a sa loi } \nu \text{ fixe et portée par } \{G > 0\} \\ \forall t, \Phi_t^T \text{ est une bijection de } \{\phi_t^T > 0\} \text{ sur } \{G > 0\} \end{cases},$$

de sorte que ϕ^T est une fonction pivotale étalée, admettant la factorisation (Φ^T, G) .

Alors $\forall t \in \Delta'$, Φ_t^T permet de définir une probabilité $\bar{\Pi}_t$ sur la tribu de Borel $\mathcal{B}(\Theta)$, qu'on appellera *probabilité sur Θ sachant t* , en transportant la mesure ν sur $\{G > 0\}$ par la bijection réciproque de Φ_t^T :

$$\forall t \in \Delta', \bar{\Pi}_t \stackrel{\text{déf}}{=} (\Phi_t^T)^{-1}(\nu).$$

On a donc $\forall t \in \Delta'$, $\Phi_t^T(\bar{\Pi}_t) = \nu$ et comme $\phi^T = G \circ \Phi^T$, on en déduit que $\phi_t^T(\bar{\Pi}_t) = G(\nu) = \eta$: la restriction de $\bar{\Pi}_t$ à la sous-tribu $(\phi_t^T)^{-1}(\mathcal{B}(\mathbb{R}))$ de $\mathcal{B}(\Theta)$ coïncide avec la pré-probabilité $\check{\Pi}_t$ sur Θ sachant t , définie au paragraphe 4.A. En particulier, $\bar{\Pi}_t$ est portée par $\{\phi_t^T > 0\}$. On contrôle aussi la propriété d'invariance par reparamétrage pour les $\bar{\Pi}_t$.

La relation $\forall t \in \Delta'$, $\Phi_t^T(\bar{\Pi}_t) = \nu$ exprime que $\forall B \in \mathcal{B}(\mathbb{R}^d)$, le niveau $\gamma = \nu(B)$ de la région de confiance $(\Phi_t^T)^{-1}(B)$ (cf. le paragraphe 2.C) coïncide avec sa probabilité $\bar{\Pi}_t((\Phi_t^T)^{-1}(B))$. Comme tout sous-ensemble borélien $\Theta' \subseteq \Theta$ fixé vérifie pour tout $t \in \Delta'$, $\Theta' \cap \{\phi_t^T > 0\} = (\Phi_t^T)^{-1}(B_t)$ pour $B_t \stackrel{\text{déf}}{=} \Phi_t^T(\Theta' \cap \{\phi_t^T > 0\})$, il est tentant d'interpréter la probabilité $\bar{\Pi}_t(\Theta') = \nu(B_t)$ comme *la probabilité que θ se trouve dans Θ' sachant t* ; mais ceci ne découle pas du modèle fréquentiste (cf. sur le même sujet le début de paragraphe 2.B), et il s'agit nous semble-t-il du point de rupture avec les modèles de type bayésien (cf. le paragraphe suivant).

On note la nouvelle relation de *dualité* entre les variables t et θ déterminée par la fonction pivotale Φ^T ,

$$\Phi_t^T(\bar{\Pi}_t) \stackrel{\forall t}{=} \nu \stackrel{\forall \theta}{=} \Phi_\theta^T(T(P_\theta)).$$

Si S est n'importe quelle autre statistique exhaustive, il existe une application borélienne ψ telle que $T = \psi(S)$ puisque T est minimale, et on a la même factorisation avec $\Phi_s^S \stackrel{\text{déf}}{=} \Phi_{\psi(s)}^T$ (cf. le paragraphe 2.C); de sorte que Φ_s^S détermine la même probabilité sachant s que Φ_t^T sachant $t = \psi(s)$, à savoir $(\Phi_s^S)^{-1}(\nu) = \bar{\Pi}_{\psi(s)}$.

Supposons à présent qu'on dispose d'une autre factorisation de la fonction pivotale de même type (qu'on peut associer à la même statistique exhaustive minimale T , comme vu juste ci-dessus)

$$(\Phi'^T, G'), \text{ soit } \phi'^T = G' \circ \Phi'^T \quad \text{où} \quad \begin{cases} \forall \theta, \Phi'_\theta(T) \text{ a sa loi } \nu' \text{ fixe et portée par } \{G' > 0\} \\ \forall t, \Phi'_t(T) \text{ est une bijection de } \{\phi'_t > 0\} \text{ sur } \{G' > 0\} \end{cases}.$$

On a ainsi $\forall t \in \Delta'$, $\Phi_t'^T = \ell_t \circ \Phi_t^T$ sur $\{\phi_t^T > 0\}$, où $\ell_t \stackrel{\text{déf}}{=} \Phi_t'^T \circ (\Phi_t^T)^{-1} : \{G > 0\} \rightarrow \{G' > 0\}$ est une bijection; de sorte que $\forall t, G = \phi_t^T \circ (\Phi_t^T)^{-1} = G' \circ \ell_t$ sur $\{G > 0\}$.

Si $\ell_t = \ell$ ne dépend pas de t , on a $\forall t \in \Delta'$, $\Phi_t'^T = \ell \circ \Phi_t^T$ sur $\{\phi_t^T > 0\}$; donc

$$\nu' \stackrel{\forall \theta}{=} (\Phi'_\theta(T))(P_\theta) = \ell[(\Phi_\theta^T(T))(P_\theta)] \stackrel{\forall \theta}{=} \ell(\nu).$$

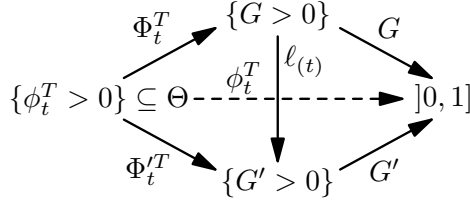


FIGURE 3. Factorisations de la fonction ϕ^T

On en déduit que $\forall t \in \Delta', \bar{\Pi}'_t \stackrel{\text{déf}}{=} (\Phi_t'^T)^{-1}(\nu') = (\Phi_t^T)^{-1}[\ell^{-1}(\ell(\nu))] = (\Phi_t^T)^{-1}(\nu) = \bar{\Pi}_t$. Inversement pour toute bijection ℓ de $\{G > 0\} \subseteq \mathbb{R}^d$ sur un sous-ensemble $\text{im } \ell$ de \mathbb{R}^d (strict si $\{\phi_t^T = 0\} \neq \emptyset$), on définit une nouvelle factorisation de ϕ^T en posant

$$\forall t \in \Delta', \Phi_t'^T \stackrel{\text{déf}}{=} \begin{cases} \ell \circ \Phi_t^T & \text{sur } \{\phi_t^T > 0\} \\ \text{des } e \in \mathbb{R}^d \setminus \text{im } \ell & \text{sur } \{\phi_t^T = 0\} \end{cases} \quad \text{et } G' \stackrel{\text{déf}}{=} \begin{cases} G \circ \ell^{-1} & \text{sur } \text{im } \ell \\ 0 & \text{sur } \mathbb{R}^d \setminus \text{im } \ell \end{cases};$$

on désignera en abrégé cette nouvelle factorisation par $(\ell \circ \Phi^T, G \circ \ell^{-1})$.

En dimension $d = 1$ la relation $G = G' \circ \ell_t$ sur $\{G > 0\}$ permet de conclure que $\ell_t = \ell$ ne dépend pas de t —ce qui assure le caractère intrinsèque des probabilités $\bar{\Pi}_t$ sur la tribu de Borel $\mathcal{B}(\Theta)$ entière—, du moins en renforçant légèrement les hypothèses de factorisation : on peut par exemple imposer aux $\{\phi_t^T > 0\}$ et à $\{G > 0\}$ d'être des intervalles ouverts de \mathbb{R} , à G d'être (strictement) monotone par morceaux sur $\{G > 0\}$ et aux bijections Φ_t^T d'être monotones, toutes croissantes ou toutes décroissantes.

Par contre dès que $d \geq 2$ aucune condition simple ne semble pouvoir assurer que ℓ_t ne dépend pas de t , même si $G' = G$, et donc que les familles de probabilités sachant t coïncident nécessairement. Il faut alors pouvoir justifier ou au moins motiver le choix fait dans chaque cas particulier, comme cela sera illustré au paragraphe 4.D.

Remarque 9 En fait, on pourrait construire bien d'autres familles de probabilités $(\bar{\Pi}'_t)_{t \in \Delta'}$ sur Θ telles que $\forall t \in \Delta' \bar{\Pi}'_t$ restreint à la sous-tribu $(\phi_t^T)^{-1}(\mathcal{B}(\mathbb{R}))$ coïncide avec $\check{\Pi}_t$, autrement dit telles que $\phi_t^T(\bar{\Pi}'_t) = \eta$. En effet, étant donné une seule factorisation (Φ^T, G) , on a pour toute probabilité ν' portée par $\{G > 0\}$ telle que $G(\nu') = G(\nu)$:

$$\forall t \in \Delta', \bar{\Pi}'_t \stackrel{\text{déf}}{=} (\Phi_t^T)^{-1}(\nu') \text{ vérifie } \phi_t^T(\bar{\Pi}'_t) = G(\Phi_t^T(\bar{\Pi}'_t)) = G(\nu') = \eta;$$

se restreindre à la loi fixe ν de $\Phi_\theta^T(T)$ évite donc une démultiplication inutile des choix et préserve le lien avec les niveaux de confiance ; mais on mesure combien la donnée des $\check{\Pi}_t^T$ ne suffit pas à déterminer les $\bar{\Pi}_t$.

4.C. Lien avec la statistique bayésienne

Dans le cadre précédent, si U admet la densité f_U par rapport à la mesure de Lebesgue λ sur \mathbb{R}^d (soit $\nu = f_U \lambda$) et si les Φ_t^T sont des difféomorphismes entre les ouverts $\{\phi_t^T > 0\}$ et $\{G > 0\}$, de Jacobien $J_{\Phi_t^T}$, la formule de changement de variables dans les intégrales multiples permet d'écrire

$$\forall t, \bar{\Pi}_t = (\Phi_t^T)^{-1}(f_U \lambda) = f_U \circ \Phi_t^T (\Phi_t^T)^{-1}(\lambda) = f_U \circ \Phi_t^T |J_{\Phi_t^T}| \mathbb{1}_\Theta \lambda;$$

c'est-à-dire que $\bar{\Pi}_t$ admet la densité $\pi_t^\lambda = |J_{\Phi_t^T}| f_U \circ \Phi_t^T$ par rapport à la mesure de Lebesgue $\mathbb{1}_\Theta \lambda$ sur $\Theta \subseteq \mathbb{R}^d$.

L'idée fort séduisante proposée par R.A. Fisher est que dans une situation où on dispose sur Θ de probabilités "naturelles" $\bar{\Pi}_t$, $t \in \Delta'$, comme au paragraphe 4.B (et Fisher utilise pour cela dans le cas

réel la fonction de répartition de T comme fonction Φ , cf. le paragraphe 5.A), il deviendrait légitime de considérer pleinement le paramètre θ comme un point aléatoire ; et que si ces probabilités admettent des densités — qu'il appelle *densités fiduciaires* — π_t , elles pourraient servir de densités bayésiennes a priori à l'occasion d'un nouvel échantillonnage : ce qui permettrait de faire la jonction entre statistique fréquentiste et statistique bayésienne.

Toutefois on doit noter que si la densité fiduciaire π_t est fournie par un premier échantillonnage connu $x = (x_1, \dots, x_n)$ et si on fait un second tirage $x' = (x'_1, \dots, x'_m)$, la densité a posteriori $\pi_{(x,x')}$ sur Θ qui en découlerait en bayésien, censée pouvoir servir de densité a priori pour un troisième tirage, serait

$$\pi_{(x,x')} \propto \pi_t f_{t'}^{T(m)} \quad \text{où } t = \tau^{(n)}(x) \text{ et } t' = \tau^{(m)}(x').$$

Le problème est que $\pi_{(x,x')}$ et $\pi_{(x',x)}$ n'ont aucune raison de coïncider dans le cas général ; or il paraît aberrant que l'ordre dans lequel les deux premiers tirages x et x' ont eu lieu puisse affecter cette densité ! De fait ce n'est pas le cas dans le paradigme bayésien authentique, où la densité a posteriori est proportionnelle à

$$(\pi f_t^{T(n)}) f_{t'}^{T(m)} = (\pi f_{t'}^{T(m)}) f_t^{T(n)},$$

π désignant la densité a priori, qui fait partie intégrante du modèle bayésien considéré.

Je propose en conséquence de se restreindre, pour cette interprétation fiduciaire, au cas où on fait l'hypothèse supplémentaire suivante sur les $\bar{\Pi}_t$, compatible avec le fait que $\bar{\Pi}_t$ est portée par $\{\phi_t^T > 0\}$:

$$\text{les mesures } \frac{1}{\phi_t^T} \bar{\Pi}_t, t \in \Delta' \text{ sont toutes proportionnelles,}$$

autrement dit il existe une mesure σ -finie $\bar{\Pi}$ sur Θ (définie à un facteur multiplicatif près) telle que

$$\forall t \in \Delta', \bar{\Pi}_t = \bar{\pi}_t \bar{\Pi}, \text{ où } \bar{\pi}_t \propto \phi_t^T.$$

En effet ceci permet de “recoller” au paradigme bayésien en considérant la densité a priori $\bar{\pi} = 1$ par rapport à $\bar{\Pi}$, puisqu'alors $\bar{\pi}_t \propto \bar{\pi} f_t^T$ (du moins en cas d'absence de paramètre nuisible, où $\phi_t^T \propto f_t^T$). Cette propriété est conservée par tout reparamétrage $\theta \in \Theta \xrightarrow{\delta} \tilde{\theta} \in \tilde{\Theta}$ (δ bijection bi-mesurable), à condition de remplacer la mesure de référence $\bar{\Pi}$ par son image $\tilde{\bar{\Pi}}$ par δ : en effet

$$\forall t \in \Delta', \tilde{\bar{\Pi}}_t = \delta(\bar{\Pi}_t) = \delta(\bar{\pi}_t \bar{\Pi}) = (\bar{\pi}_t \circ \delta^{-1}) \delta(\bar{\Pi}),$$

soit la densité $\tilde{\bar{\pi}}_t = \bar{\pi}_t \circ \delta^{-1}$ relativement à $\tilde{\bar{\Pi}} = \delta(\bar{\Pi})$, avec $\tilde{\bar{\pi}}_t \propto \tilde{\phi}_t^T$, puisque $\bar{\pi}_t \propto \phi_t^T$ et $\tilde{\phi}_t^T = \phi_t^T \circ \delta^{-1}$.

En prenant $\bar{\Pi}$ comme loi a priori (au sens généralisé si ce n'est pas une mesure finie), les estimations bayésiennes au maximum de probabilité a posteriori et par régions de plus forte densité a posteriori coïncident respectivement avec leurs homologues fréquentistes au maximum de vraisemblance et par régions les plus vraisemblables, avec les mêmes niveaux de confiance pour les régions : choisir $\bar{\Pi}$ comme mesure de référence sur Θ conduit ainsi à exprimer l'absence d'information a priori sur le modèle par la densité a priori $\bar{\pi} = 1$. Et par reparamétrage $\theta \in \Theta \xrightarrow{\delta} \tilde{\theta} \in \tilde{\Theta}$, toutes les densités (a priori et sachant t) par rapport à $\bar{\Pi}$ et $\tilde{\bar{\Pi}}$ conservent les mêmes expressions $\tilde{\pi}_{(t)}(\tilde{\theta}) = \pi_{(t)}(\theta)$, puisqu'on a comme ci-dessus

$$\forall t \in \Delta', \tilde{\pi}_{(t)} \tilde{\bar{\Pi}} = \tilde{\bar{\Pi}}_{(t)} = \delta(\bar{\Pi}_{(t)}) = \delta(\pi_{(t)} \bar{\Pi}) = (\pi_{(t)} \circ \delta^{-1}) \tilde{\bar{\Pi}}.$$

Je dirai en conséquence que $\bar{\Pi}$ est *basiquement non informative* pour le modèle et j'appellerai *parfait* un tel modèle. La mesure $\bar{\Pi}$ détermine en retour les probabilités $\bar{\Pi}_t \propto \phi_t^T \bar{\Pi}$, et on a pour tout $t \in \Delta'$

$$\Phi_t^T(\bar{\Pi}) \propto \Phi_t^T\left(\frac{1}{\phi_t^T} \bar{\Pi}_t\right) = \frac{1}{G} \Phi_t^T(\bar{\Pi}_t) = \frac{1}{G} \nu, \text{ d'où } \phi_t^T(\bar{\Pi}) \propto G\left(\frac{1}{G} \nu\right) = \frac{1}{\text{id}_{\mathbb{R}}} \eta, \text{ mesures indépendantes de } t;$$

et comme $\bar{\Pi}_t \propto \phi_t^T \bar{\Pi}$, on a $\int_{\Theta} \phi_t^T d\bar{\Pi} < +\infty$.

Remarque 10 On rejoint à ce stade l'approche bayésienne des lois a priori non informatives dites lois a priori coïncidentes (exact matching priors).

Remarque 11 Comme vu au paragraphe 4.B, rien n'assure en dimension $d \geq 2$ l'unicité de la famille $(\bar{\Pi}_t)_{t \in \Delta'}$, et pas plus celle de $\bar{\Pi}$ (à un facteur multiplicatif près). Par contre, à l'unicité de la famille de leurs restrictions $(\check{\Pi}_t)_{t \in \Delta'}$, à la sous-tribu $(\phi_t^T)^{-1}(\mathcal{B}(\mathbb{R}))$ de $\mathcal{B}(\Theta)$ répond l'unicité, à un facteur multiplicatif près, de la restriction de $\bar{\Pi}$ à $(\phi_t^T)^{-1}(\mathcal{B}(\mathbb{R}))$ pour tout $t \in \Delta'$ (puisque $\phi_t^T(\bar{\Pi}) \propto \frac{1}{\text{id}_{\mathbb{R}}} \eta$). Celle-ci apparaît comme authentiquement non informative, d'où le choix du terme *basiquement non informative* pour qualifier $\bar{\Pi}$.

4.D. Exemples de modèles parfaits

Nous réexaminons les modèles vus précédemment, pour lesquels les propriétés ci-dessus sont vérifiées :

- Les **modèles à paramètre de translation** pour T (cf. le paragraphe 3.B)

$\Delta' = \Theta \subseteq \mathbb{R}^d$ groupe localement compact, de mesure de Haar à gauche Λ , et

$$\forall \theta \in \Theta, \theta^{-1} T \stackrel{\mathcal{L}}{=} U \text{ où } U \text{ est de loi fixe } \nu = g \Lambda \text{ avec } g \text{ bornée.}$$

Comme $u = \Phi_t^T(\theta) = \theta^{-1} t \iff u^{-1} = t^{-1} \theta \iff (\Phi_t^T)^{-1}(u) = \theta = t u^{-1}$, on a $(\Phi_t^T)^{-1} = \tau_t \circ i$ (où $i : u \mapsto u^{-1}$ et $\tau_t : v \mapsto t v$); on en déduit que $\forall t \in \Delta', \bar{\Pi}_t = (\Phi_t^T)^{-1}(\nu) = \tau_t(i(\nu))$, translatée à gauche par t de la mesure fixe $i(\nu)$. Or $T \stackrel{\mathcal{L}}{=} \theta U \iff T(P_\theta) = \tau_\theta(\nu)$, de sorte que

$$\forall c \in \Theta, (cT)(P_\theta) = \tau_c(T(P_\theta)) = \tau_{c\theta}(\nu) = T(P_{c\theta}) = T(P_{\tau_c(\theta)}) :$$

le modèle statistique est donc invariant sous l'action du groupe des translations à gauche; et comme

$$\forall t \in \Delta', \bar{\Pi}_{\tau_c(t)} = \tau_{ct}(i(\nu)) = \tau_c(\bar{\Pi}_t),$$

le *principe d'invariance* est vérifié.

On a de plus

$$\forall t \in \Delta', \bar{\Pi}_t = (\Phi_t^T)^{-1}(g \Lambda) = (g \circ \Phi_t^T)(\Phi_t^T)^{-1}(\Lambda) = (g \circ \Phi_t^T) \tau_t(i(\Lambda)) = (g \circ \Phi_t^T) \Delta(t) i(\Lambda),$$

où $\Delta > 0$ est la *fonction modulaire* du groupe Θ et où $i(\Lambda)$ est "la" mesure de Haar à droite sur Θ : ainsi $\forall t \in \Delta'$, la probabilité $\bar{\Pi}_t$ sur Θ sachant t admet par rapport à la mesure $\bar{\Pi} \stackrel{\text{déf}}{=} i(\Lambda)$ la densité $\bar{\pi}_t = \Delta(t) g \circ \Phi_t^T = \|g\|_\infty \Delta(t) \phi_t^T$ proportionnelle à ϕ_t^T ; de sorte que le modèle à paramètre de position pour T est *parfait*. Et on note que $\bar{\Pi}$ dépend seulement de la structure de groupe sur Θ et pas du tout de la loi particulière ν de U , ce qui constitue une *propriété intrinsèque de non information*.

Remarque 12 De $(\Phi_t^T)^{-1}(\Lambda) = \Delta(t) i(\Lambda)$ on déduit $\Lambda = \Delta(t) \Phi_t^T(i(\Lambda))$. Ainsi la relation $\bar{\Pi} = i(\Lambda)$ suffit à assurer $\Phi_t^T(\bar{\Pi}) = \frac{1}{\Delta(t)} \Lambda \propto \Lambda$, de sorte que (cf. le paragraphe 4.C) $\frac{1}{G} \nu \propto \Lambda$, soit $\nu = g \Lambda$ où $g = \|g\|_\infty G$.

Exemple du modèle des lois gaussiennes

La mesure basiquement non informative sur le sous-groupe $\Theta = \mathbb{R}_+^* \times \mathbb{R}$ du groupe $ax + b$ est donc la mesure de Haar à droite $\bar{\Pi}(d\sigma, dm) = \frac{1}{\sigma} d\sigma dm$, qui fait de σ et m des paramètres a priori "indépendants".

On note que $\bar{\Pi}$ n'est pas (à un facteur près) la *loi de Jeffreys* $\frac{1}{\sigma^2} d\sigma dm$, qui est elle la mesure de Haar à gauche sur Θ . On calcule les densités $\bar{\pi}_{(s,\bar{x})}$ sachant $g(u_1, u_2) = k u_1^n e^{-\frac{n}{2}(u_1^2 + u_2^2)}$ et $k = \frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}-1} \sqrt{\pi} \Gamma(\frac{n-1}{2})}$

(cf. le paragraphe 3.B) en vérifiant que $\Delta(s, \bar{x}) = \frac{1}{s}$, ou en se référant au début du paragraphe 4.C :

$$\forall (s, \bar{x}) \in \mathbb{R}_+^* \times \mathbb{R}, \bar{\pi}_{(s, \bar{x})}(\sigma, m) = k \frac{s^{n-1}}{\sigma^n} e^{-\frac{n}{2} \frac{(\bar{x}-m)^2 + s^2}{\sigma^2}}.$$

- Le cas particulier où il s'agit d'une **translation de position** (cf. paragraphe 3.C). La mesure basiquement non informative sur le groupe commutatif $(\Theta = \mathbb{R}, +)$ est la mesure de Haar–Lebesgue $\bar{\Pi} = \lambda$. Pour ces modèles, on a $\forall t \in \mathbb{R}, \bar{\pi}_t(\theta) = g(t - \theta)$, puisque la fonction modulaire Δ vaut 1.

Exemples du **modèle des exponentielles translatées** et du **modèle des lois gaussiennes à écart-type connu**.

Pour ce second modèle (qui est exponentiel), on a $E_\theta(T) - m = E(U)$, soit $h(m) = E(U) + m$; d'autre part $\alpha(m) = -\frac{m}{\sigma^2}$. Comme l'*information de Fisher* vaut $I^T = n \alpha' h' = \frac{n}{\sigma^2}$ qui est une constante, $\bar{\Pi}$ est aussi (à un facteur près) la *loi de Jeffreys* $\sqrt{I^T} \lambda$ sur Θ .

- Le cas particulier où il s'agit d'une **translation d'échelle** (cf. paragraphe 3.D). La mesure basiquement non informative sur le groupe commutatif $(\Theta = \mathbb{R}_+^*, \times)$ est la mesure de Haar $\bar{\Pi}(d\theta) = \frac{1}{\theta} d\theta$. Pour ces modèles, on a $\forall t \in \mathbb{R}_+^*, \bar{\pi}_t(\theta) = g\left(\frac{t}{\theta}\right)$, puisque la fonction modulaire Δ vaut 1.

Exemples du **modèle des lois uniformes**; du **modèle gamma-Weibull**; et du **modèle des lois gaussiennes à moyenne connue**.

Pour les deux derniers modèles (exponentiels), on a respectivement $h(\theta) = E_\theta(T) = \theta^b E(U) = a \theta^b$ et $\alpha(\theta) = -\frac{1}{\theta^b}$; $h(\theta) = E_\theta(V) = \theta^2 E(U) = \theta^2$ et $\alpha(\theta) = -\frac{1}{2\theta^2}$. Comme $I^T = n \alpha' h'$, on a dans les deux cas $\sqrt{I^T} \propto \left(\theta \mapsto \frac{1}{\theta}\right)$: la mesure $\bar{\Pi}$ est aussi (à un facteur près) la *loi de Jeffreys* $\sqrt{I^T} \lambda$ sur Θ .

- Le **modèle des lois gaussiennes à écart-type inconnu**, avec $\Phi_m^T(s, \bar{x}) \stackrel{\text{déf}}{=} \frac{\bar{x} - m}{s} \sqrt{n-1}$ (cf. paragraphe 3.E).

$\forall (s, \bar{x}) \in \Delta' = \mathbb{R}, \Phi_{(s, \bar{x})}^T$ est un difféomorphisme de $\{\phi_{(s, \bar{x})}^T > 0\} = \Theta_2 = \mathbb{R}$ sur $\{G > 0\} = \mathbb{R}$, de Jacobien $J_{\Phi_{(s, \bar{x})}^T}(m) = -\frac{\sqrt{n-1}}{s}$. Comme $\Phi_m^T(S, \bar{X}) \stackrel{\mathcal{L}}{=} T_{n-1}$, on a $f_U(u) = \frac{1}{\sqrt{(n-1)\pi}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \left(1 + \frac{u^2}{n-1}\right)^{-\frac{n}{2}}$ (où $\nu = f_U \lambda$); d'où l'expression de la densité sur Θ_2 sachant (s, \bar{x}) par rapport à la mesure de Lebesgue (cf. le début du paragraphe 4.C)

$$\bar{\pi}_{(s, \bar{x})}(m) = \frac{1}{s \sqrt{\pi}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \left(1 + \left(\frac{\bar{x} - m}{s}\right)^2\right)^{-\frac{n}{2}}.$$

Ainsi $\bar{\pi}_{(s, \bar{x})} \propto \phi_{(s, \bar{x})}^T$ puisque $\phi_m^T(s, \bar{x}) = \left(1 + \left(\frac{\bar{x} - m}{s}\right)^2\right)^{-\frac{n}{2}}$: on a affaire à un modèle parfait, dont la mesure basiquement non informative sur $(\Theta_2 = \mathbb{R}, +)$ est la mesure de Haar–Lebesgue $\bar{\Pi} = \lambda$.

- Le **modèle des lois gaussiennes à moyenne inconnue**, avec $\Phi_\sigma^S(s) \stackrel{\text{déf}}{=} \frac{s^2}{\sigma^2}$ (cf. paragraphe 3.E).

$\forall s \in \Delta' = \mathbb{R}_+^*, \Phi_s^S$ est un difféomorphisme de $\{\phi_s^S > 0\} = \Theta_1 = \mathbb{R}_+^*$ sur $\{G > 0\} = \mathbb{R}_+^*$, de Jacobien $J_{\Phi_s^S}(m) = -\frac{2s^2}{\sigma^3}$. Comme $\Phi_\sigma^S(S) \stackrel{\mathcal{L}}{=} \frac{\chi_{n-1}^2}{n}$, on a $f_U(u) = \frac{n}{2^{\frac{n-1}{2}} \Gamma(\frac{n-1}{2})} e^{-\frac{nu}{2}} (nu)^{\frac{n-1}{2}-1} \mathbb{1}_{\mathbb{R}_+^*}(u)$; d'où l'expression de la densité sur Θ_1 sachant s par rapport à la mesure de Lebesgue (cf. le début du paragraphe 4.C)

$$\bar{\pi}_s(\sigma) = \frac{2s^2}{\sigma^3} \frac{n}{2^{\frac{n-1}{2}} \Gamma(\frac{n-1}{2})} e^{-\frac{ns^2}{2\sigma^2}} \left(\frac{ns^2}{2\sigma^2}\right)^{\frac{n-1}{2}-1} = \frac{n^{\frac{n-1}{2}}}{2^{n-3} \Gamma(\frac{n-1}{2})} \frac{s^{n-1}}{\sigma^n} e^{-\frac{ns^2}{2\sigma^2}}.$$

Ainsi $\bar{\pi}_s \propto \phi_s^S$ puisque $\phi_\sigma^S(s) = e^{-\frac{n}{2} K(\frac{s^2}{\sigma^2})} = e^{\frac{n}{2} \frac{s^n}{\sigma^n}} e^{-\frac{ns^2}{2\sigma^2}}$: on a affaire à un modèle parfait, dont la mesure basiquement non informative $\bar{\Pi}$ sur $(\Theta_1 = \mathbb{R}_+^*, \times)$ est inhabituelle, puisque c'est la mesure de Lebesgue sur \mathbb{R}_+^* ! Ceci est lié au fait que dans ce modèle à paramètre nuisible la vraisemblance marginale ϕ_s^S n'est pas proportionnelle à la fonction f_s^S (où f_σ^S désigne la densité de S à σ fixé) ; en effet la variable

S est indépendante de \bar{X} pour toute valeur de m et $S \stackrel{\mathcal{L}}{\equiv} \sigma \frac{\sqrt{\chi_{n-1}^2}}{\sqrt{n}}$ (cf. le paragraphe 3.B), donc

$$f_\sigma^S(s) = \frac{n^{\frac{n-1}{2}}}{2^{\frac{n-3}{2}} \Gamma(\frac{n-1}{2})} \frac{s^{n-2}}{\sigma^{n-1}} e^{-\frac{ns^2}{2\sigma^2}} \mathbb{1}_{\mathbb{R}_+^*}(s).$$

5. Annexes

5.A. Statut de la fonction de répartition de T dans le cas réel

On se place dans le cadre d'une factorisation (Φ^T, G) de la vraisemblance pivotale en dimension $d = 1$, en supposant de plus que $\nu = f_U \lambda$ et que les Φ_t^T sont des difféomorphismes tous croissants ou tous décroissants ; de sorte que les densités π_t^λ par rapport à la mesure de Lebesgue sur Θ sont bien définies et vérifient d'après les relations de dualité et de changement de variable (cf. paragraphes 4.B et 4.C)

$$\Phi_t^T(\pi_t^\lambda \mathbb{1}_\Theta \lambda) \stackrel{\forall t}{\equiv} f_U \lambda \stackrel{\forall \theta}{\equiv} \Phi_\theta^T(f_\theta^T \mu') \quad \text{avec} \quad \pi_t^\lambda = \left| \frac{d\Phi_t^T}{d\theta} \right| f_U \circ \Phi_t^T.$$

On suppose en outre que $\{f_U > 0\} = \{G > 0\}$ est un intervalle ouvert sur lequel f_U est continue ; de sorte que la restriction ℓ de la fonction de répartition de la mesure ν à cet intervalle est strictement croissante et de classe C^1 . La fonction ℓ est donc un difféomorphisme croissant de $\{G > 0\}$ sur $]0, 1[$ et la mesure image $\ell(\nu)$ coïncide avec la probabilité $\bar{\lambda}$ uniforme sur $]0, 1[$ (propriété d'anamorphose).

Considérons la nouvelle factorisation $(F \stackrel{\text{déf}}{=} \ell \circ \Phi^T, G' \stackrel{\text{déf}}{=} G \circ \ell^{-1})$ (avec F_t prolongée par continuité et monotonie de part et d'autre de $\{G > 0\}$ par les valeurs constantes 0 et 1 si nécessaire) ; on a

$$\{G' > 0\} =]0, 1[\quad \text{et} \quad F_t(\pi_t^\lambda \mathbb{1}_\Theta \lambda) \stackrel{\forall t}{\equiv} \bar{\lambda} \stackrel{\forall \theta}{\equiv} F_\theta(f_\theta^T \mu') \quad \text{avec} \quad \pi_t^\lambda = \left| \frac{dF_t}{d\theta} \right|.$$

Ainsi F_t est pour tout $t \in \Delta'$ la fonction de répartition de $\bar{\Pi}_t = \pi_t^\lambda \mathbb{1}_\Theta \lambda$ (ou pour tout $t \in \Delta'$ sa complémentaire). Si de plus les fonctions Φ_θ^T sont toutes croissantes ou toutes décroissantes, il en est de même des $F_\theta = \ell \circ \Phi_\theta^T$; et on déduit de $F_\theta(f_\theta^T \mu') \stackrel{\forall \theta}{\equiv} \bar{\lambda}$ que F_θ est pour tout $\theta \in \Theta$ la fonction de répartition de T (ou pour tout $\theta \in \Theta$ sa complémentaire), et aussi que la mesure $f_\theta^T \mu'$ est diffuse.

Inversement limitons-nous au cas d'un modèle exponentiel réel ($d = 1$) dont la mesure μ' est diffuse (donc μ aussi) et portée par l'intervalle ouvert $h(\Theta)$ (cf. le paragraphe 2.D). Posons $\varepsilon = \text{sgn}(\alpha')$, c'est-à-dire $\varepsilon = -1$ ou $+1$ selon que le signe constant de α' est négatif ou positif, et considérons l'application F définie sur $\mathbb{R} \times \Theta$ par

$$F(t, \theta) \stackrel{\text{déf}}{=} \begin{cases} P_\theta(T \leq t) = P_\theta(T < t) & \text{si } \varepsilon = -1 \\ P_\theta(T > t) = P_\theta(T \geq t) & \text{si } \varepsilon = +1 \end{cases},$$

de sorte que les fonctions F_θ^T sont toutes croissantes si $\varepsilon = -1$ ou toutes décroissantes si $\varepsilon = +1$.

Comme la loi de T est diffuse, F_θ est continue ; on a donc (propriété d'anamorphose)

$$\forall \theta \in \Theta, F_\theta(T) \text{ est de loi uniforme sur }]0, 1[,$$

de sorte que F est une fonction pivotale, intrinsèque au modèle exponentiel par construction.

On sait par ailleurs qu'à t fixé, l'application F_t est continue et croissante (au sens large) sur Θ . La croissance s'établit classiquement en montrant que pour tous $\theta_1 < \theta_2$ dans Θ , on a (pour $\varepsilon = \pm 1$)

$$\forall t \in \mathbb{R}, \begin{cases} F_{\theta_2}(t) \geq \lambda_t F_{\theta_1}(t) \\ 1 - F_{\theta_2}(t) \leq \lambda_t (1 - F_{\theta_1}(t)) \end{cases} \quad \text{where } \lambda_t \stackrel{\text{déf}}{=} \frac{e^{n[\alpha(\theta_2)t + \beta(\theta_2)]}}{e^{n[\alpha(\theta_1)t + \beta(\theta_1)]}} = \frac{f_{\theta_2}^T(t)}{f_{\theta_1}^T(t)},$$

du fait de la croissance de l'application $s \in \mathbb{R} \mapsto \varepsilon \lambda_s$; de sorte que $F_{\theta_2}(t) \geq F_{\theta_1}(t)$, d'après la première inégalité si $\lambda_t \geq 1$ et d'après la seconde inégalité si $\lambda_t < 1$. Soit $\Delta' \subseteq h(\Theta)$ le plus petit intervalle ouvert de \mathbb{R} dont l'adhérence contient le support de μ' ; comme $f_\theta^T > 0$ partout, on a $\Delta' = \{0 < F_\theta < 1\}$ pour tout $\theta \in \Theta$; et comme μ' est diffuse, Δ' est un sous-ensemble porteur de μ' . Montrons que la croissance de F_t sur Θ est stricte pour tout $t \in \Delta'$: supposons par l'absurde l'existence d'un $t \in \Delta'$ tel que F_t est constante sur un intervalle non trivial de Θ ; pour tous $\theta_1 < \theta_2$ dans cet intervalle, l'égalité $F_{\theta_1}(t) = F_{\theta_2}(t)$ impliquerait dans les inégalités ci-dessus $\lambda_t = 1$ et donc $f_{\theta_1}^T(t) = f_{\theta_2}^T(t)$, puisque $t \in \Delta'$ exclut qu'on ait $F_{\theta_1}(t) = F_{\theta_2}(t) \in \{0, 1\}$. La fonction f_t^T serait donc aussi constante sur l'intervalle, ce qui est incompatible avec f_t^T unimodale sur Θ .

Il suffit alors, pour que F soit une fonction pivotale étalée, de rajouter l'hypothèse

$$\forall t \in \Delta', \inf F_t(\Theta) = 0 \text{ et } \sup F_t(\Theta) = 1,$$

qui équivaut vu la continuité de F_t à la condition $\forall t \in \Delta', F_t(\Theta) =]0, 1[$.

La fonction F est une parfaite fonction candidate à la factorisation de la vraisemblance pivotale, au moyen d'une fonction $G :]0, 1[\rightarrow]0, 1[$ unimodale. On aurait pu se référer à F dans les modèles réels parfaits présentés aux paragraphes 3.C et 3.D, mais F n'a pas en général d'expression explicite simple.

5.B. Le modèle de seuillage d'une variable aléatoire

Soit f_Y la densité d'une variable aléatoire réelle Y telle que l'ensemble $\{f_Y > 0\}$ est un intervalle ouvert sur lequel f_Y est continue. On considère les modèles (non réguliers) de *paramètre de seuil* θ

– modèle de seuillage inférieur au seuil haut $\theta : X \stackrel{\text{déf}}{=} (Y | Y < \theta)$, tel que

$$\Delta = \{f_Y > 0\}, \mu = \lambda, f_\theta(x) = \frac{f_Y(x)}{F_Y(\theta)} \mathbb{1}_{]0, +\infty[}(\theta - x) \quad \text{où } \Theta = \Delta,$$

F_Y désignant la fonction de répartition de Y ;

– modèle de seuillage supérieur au seuil bas $\theta : X \stackrel{\text{déf}}{=} (Y | Y > \theta)$, qui se distingue du cas précédent par

$$f_\theta(x) = \frac{f_Y(x)}{F_Y(\theta)} \mathbb{1}_{]0, +\infty[(x - \theta) \quad \text{où } F_Y \text{ est la fonction de répartition complémentaire de } Y.$$

Le reparamétrage $\theta \in \Theta \xrightarrow{\delta} \tilde{\theta} = F_Y(\theta) \in \tilde{\Theta} =]0, 1[$ (δ difféomorphisme) ramène les deux cas à $X \stackrel{\text{déf}}{=} (Y | F_Y(Y) < \tilde{\theta})$, dont l'image par F_Y est le modèle de seuillage inférieur au seuil haut $\tilde{\theta}$:

$$\bar{X} \stackrel{\text{déf}}{=} (\bar{Y} | \bar{Y} < \tilde{\theta}) \quad \text{où } \bar{Y} \text{ est la variable uniforme sur }]0, 1[, \text{ d'où } f_{\bar{Y}} = \mathbb{1}_{]0, 1[} \text{ et } F_{\bar{Y}}(\bar{y}) = \bar{y} \text{ sur }]0, 1[;$$

de sorte que $\bar{\Delta} =]0, 1[, \bar{\mu} = \bar{\lambda}, \bar{f}_{\tilde{\theta}}(\bar{x}) = \frac{1}{\tilde{\theta}} \mathbb{1}_{\{0 < \bar{x} < \tilde{\theta} < 1\}}$.

On retombe sur le modèle des lois uniformes (cf. paragraphe 3.D), à ceci près qu'on a ici $\bar{\Delta} = \tilde{\Theta} =]0, 1[$ au lieu de \mathbb{R}_+^* ; on en déduit que la vraisemblance pivotale

$$\bar{\phi}^T = G \circ \bar{\Phi}^T \quad \text{où } \bar{T} = \max(\bar{X}_1, \dots, \bar{X}_n), \quad \bar{\Phi}^T(\bar{t}, \bar{\theta}) = \frac{\bar{t}}{\bar{\theta}} \text{ sur } \bar{\Delta} \times \bar{\Theta} \text{ et } G(u) = u^n \mathbb{1}_{]0,1[}(u)$$

n'est *pas* étalée : en effet $\forall \bar{t} \in \bar{\Delta}' = \bar{\Delta}$, $\bar{\Phi}^T(\bar{\Theta}) =]\bar{t}, +\infty[$, donc $\bar{\phi}^T(\bar{\Theta}) = \{0\} \cup](\bar{t})^n, 1[$ ne porte pas la loi η de $G(U) = U^n$ qui est la probabilité uniforme $\bar{\lambda}$ sur $]0, 1[$.

En conséquence, $\bar{\Pi}_{\bar{t}} \stackrel{\text{déf}}{=} (\bar{\Phi}^T)^{-1}(\nu)$ est la sous-probabilité de masse $1 - (\bar{t})^n$ portée par $]\bar{t}, 1[$, trace sur cet intervalle de la probabilité correspondante $\bar{\Pi}_{\bar{t}}$ du modèle des lois uniformes. Sa densité $\bar{\pi}_{\bar{t}}$ par rapport à la restriction à $]0, 1[$ de la mesure de Haar sur \mathbb{R}_+^* , $\bar{\Pi}(d\bar{\theta}) = \frac{1}{\bar{\theta}} \mathbb{1}_{]0,1[}(\bar{\theta}) d\bar{\theta}$, est encore proportionnelle à la vraisemblance $\bar{\phi}^T$ sur $]\bar{t}, 1[$ et est nulle hors de $]\bar{t}, 1[$. On pourrait parler de *modèle sous-parfait*, de mesure de référence $\bar{\Pi}$ sur $\bar{\Theta}$.

De manière concrète, pour tout $\bar{t} = \max(\bar{x}_1, \dots, \bar{x}_n) \in \bar{\Delta}$ et pour tout $\gamma \in]0, 1[$ fixé, la région la plus vraisemblable pour $\bar{\theta}$ au niveau de confiance γ est

$$\{\bar{\phi}^T > 1 - \gamma\} = \left\{ \bar{\theta} \in]0, 1[: \left(\frac{\bar{t}}{\bar{\theta}} \right)^n > 1 - \gamma \text{ et } \bar{\theta} > \bar{t} \right\} =]\bar{t}, \frac{\bar{t}}{(1-\gamma)^{\frac{1}{n}}} \wedge 1[;$$

l'imprécision maximale de cette région la plus vraisemblable dès que $\gamma \geq 1 - (\bar{t})^n$, à savoir $]\bar{t}, 1[$, est en relation directe avec le non étalement de $\bar{\phi}^T$. Mais comme $\bar{\phi}^T$ est normalisée et que $\bar{\phi}^T(\bar{\Theta})$ contient même un intervalle non vide de la forme $]y_{\bar{t}}, 1[$ pour tout $\bar{t} \in \bar{\Delta}$, on échappe aux difficultés de fond rencontrées avec les fonctions ψ aux paragraphes 3.C et 3.D.

En remontant aux deux modèles initiaux, on en déduit

– seuil haut θ : pour tout $t = \max(x_1, \dots, x_n) \in \Delta$ et pour tout $\gamma \in]0, 1[$ fixé, la région la plus vraisemblable pour θ au niveau de confiance γ est

$$\{\phi_t^T > 1 - \gamma\} = \left\{ \theta \in \Theta : \frac{F_Y(t)^n}{F_Y(\theta)^n} > 1 - \gamma \text{ et } \theta > t \right\} =]t, \theta_{t,\gamma}[,$$

$$\text{où } \theta_{t,\gamma} = \sup \left\{ \theta \in \Theta : F_Y(\theta) < \frac{F_Y(t)}{(1-\gamma)^{\frac{1}{n}}} \right\} = \begin{cases} F_Y^{-1} \left(\frac{F_Y(t)}{(1-\gamma)^{\frac{1}{n}}} \right) & \text{pour } \gamma < 1 - F_Y(t)^n \\ \sup \Delta & \text{dès que } \gamma \geq 1 - F_Y(t)^n \end{cases}.$$

– seuil bas θ : pour tout $t = \min(x_1, \dots, x_n) \in \Delta$ et pour tout $\gamma \in]0, 1[$ fixé, la région la plus vraisemblable pour θ au niveau de confiance γ est

$$\{\phi_t^T > 1 - \gamma\} = \left\{ \theta \in \Theta : \frac{F_Y(t)^n}{F_Y(\theta)^n} > 1 - \gamma \text{ et } \theta < t \right\} =]\theta_{t,\gamma}, t[,$$

$$\text{où } \theta_{t,\gamma} = \inf \left\{ \theta \in \Theta : F_Y(\theta) < \frac{F_Y(t)}{(1-\gamma)^{\frac{1}{n}}} \right\} = \begin{cases} F_Y^{-1} \left(\frac{F_Y(t)}{(1-\gamma)^{\frac{1}{n}}} \right) & \text{pour } \gamma < 1 - F_Y(t)^n \\ \inf \Delta & \text{dès que } \gamma \geq 1 - F_Y(t)^n \end{cases}.$$

Dans les deux cas, la mesure de référence sur Θ est $\bar{\Pi} = F_Y^{-1}(\bar{\Pi}) = \frac{1}{F_Y} F_Y^{-1}(\bar{\lambda}) = \frac{f_Y}{F_Y} \mathbb{1}_{\Theta} \lambda$, qui cette fois dépend de la loi particulière de Y . Et on vérifie facilement que pour tout $t \in \Delta$, l'application $\theta \in \Theta \mapsto \frac{F_Y(t)^n}{F_Y(\theta)^n} \wedge 1$ est la fonction de répartition de la sous-probabilité $\bar{\Pi}_t$ de masse $1 - F_Y(t)^n$ dans le cas de seuil bas θ , et sa complémentaire dans le cas de seuil haut θ .

5.C. Information a priori en vision fréquentiste ?

En ce qui concerne l'expression d'une éventuelle information a priori sur le paramètre θ , il pourrait sembler, en adoptant un point de vue fréquentiste intransigeant, que la notion la plus pertinente n'est pas une densité de probabilité (approche bayésienne) mais une vraisemblance, pour les deux raisons suivantes :

- l'absence complète d'information a priori s'exprime parfaitement par une vraisemblance constante : on échappe ainsi à la difficulté de fond des lois a priori non informatives en bayésien ;
- un niveau de croyance en la possibilité qu'une valeur de θ soit la vraie valeur du paramètre est par nature invariant par reparamétrage, au sens où ce niveau est le même pour la valeur correspondante de θ dans le nouveau paramétrage ; tel est le cas d'une vraisemblance, mais pas d'une densité de probabilité (à moins de changer de mesure de référence avec le paramétrage).

On peut (comme en bayésien) tirer parti de plusieurs sources indépendantes d'information, en faisant le produit des fonctions de vraisemblance correspondantes : considérons les réalisations successives $x' = (x'_1, \dots, x'_m)$ et $x = (x_1, \dots, x_n)$ de deux échantillons indépendants l'un de l'autre issus d'un même modèle, ainsi que l'échantillon formé par leur réunion. On a la relation entre densités de ces échantillons (par rapport aux mesures respectives $\mu^{(m+n)\otimes}$, $\mu^{m\otimes}$ et $\mu^{n\otimes}$) :

$$\forall \theta \in \Theta, f_\theta(x', x) = f_\theta(x') f_\theta(x) ;$$

de sorte qu'une façon de tenir compte de la connaissance acquise sur le paramètre après le premier échantillonnage consiste à remplacer pour tout x la fonction de vraisemblance $f_x : \theta \mapsto f_\theta(x)$ par le produit πf_x où $\pi = f_{x'} : \theta \mapsto f_\theta(x')$.

On peut donc spécifier une *vraisemblance a priori* π choisie dans la famille $(f_{x'})_{x' \in \bigcup_{m \in \mathbb{N}^*} \Delta_m}$, d'autant plus directive que l'entier m est grand, et adjoindre à l'échantillon obtenu $x \in \Delta^n$ un échantillon (fictif) $x' \in \Delta^m$ tel que $f_{x'} = \pi$; la fonction produit πf_x , qu'on pourrait qualifier de *vraisemblance a posteriori sachant x* , étant simplement la vraisemblance en l'échantillon réunion (x', x) dans le cadre du modèle.

En cas de statistiques exhaustives $T^{(m)} = \tau^{(m)}(X_1, \dots, X_m)$ pour tout $m \in \mathbb{N}^*$, on peut aussi bien choisir π dans la famille des vraisemblances normalisées $(\phi_{t'}^{T^{(m)}})_{t' \in \bigcup_{m \in \mathbb{N}^*} \tau^{(m)}(\Delta^m)}$, ce qui revient à préciser simplement la valeur de $m \in \mathbb{N}^*$ et une valeur (fictive) $t' \in \tau^{(m)}(\Delta^m)$: en effet soit $x' \in \Delta^m$ arbitraire tel que $\tau^{(m)}(x') = t'$; on a, en notant $t = \tau^{(n)}(x)$ où $x \in \Delta^n$ et $t'' = \tau^{(m+n)}(x', x)$ où (x', x) est l'échantillon réunion,

$$\phi_{t''}^{T^{(m+n)}} = \phi_{(x', x)} \propto f_{(x', x)} = f_{x'} f_x \propto \phi_{x'} \phi_x = \phi_{t'}^{T^{(m)}} \phi_t^{T^{(n)}}.$$

Il faut toutefois reconnaître que le fait de devoir choisir π dans ces familles prédéfinies limite la capacité d'expression de l'information a priori.

6. Conclusion

Cette étude de statistique paramétrique, qui concerne quelques modèles "parfaits" d'usage courant — et en premier lieu le modèle gaussien —, a pour ambition d'une part de montrer qu'on peut faire de l'estimation fréquentiste des paramètres par régions de confiance optimales selon le critère intrinsèque de vraisemblance ; et d'autre part d'analyser en détail le passage d'une vision fréquentiste à une vision bayésienne, via la construction d'une loi a priori aussi peu informative que possible pour ces modèles. Dans le cas où il y a plusieurs paramètres, nous avons dû nous appuyer sur une structure supplémentaire de groupe pour assurer le caractère intrinsèque de la construction.

La démarche ne semble pas pouvoir se généraliser à beaucoup d'autres modèles, et il est illusoire d'espérer combler vraiment le fossé entre les deux grands courants de la démarche statistique. Outre la difficulté posée par la notion de loi a priori non informative, la conception aléatoire des paramètres, propre au paradigme bayésien, paraît irréductible à la démarche fréquentiste : elle constitue un choix de modélisation plus riche, mais (et) dont la validation (la réfutation) expérimentale paraît impossible.

Références

- [1] J.O. Berger, J.M. Bernardo, D. Sun (2009), The formal definition of reference priors, *Annals of Statistics* 37 (2), 905–938.
- [2] J.M. Bernardo (1979), Reference posterior distributions for Bayesian inference, *Journal of the Royal Statistical Society, Series B* 41 (2), 113–147.
- [3] G S. Datta, M. Ghosh, R. Mukerjee (2000), Some new results on probability matching priors, *Calcutta Statist. Assoc. Bull.* 50, 179–92.
- [4] R.A. Fisher (1956), *Statistical Methods and Scientific Inference*, Oliver and Boyd, Edinburgh.
- [5] D.A.S. Fraser (1961), On fiducial inference, *Ann. Math. Statist.* 32 (3), 661-676.
- [6] E.T. Jaynes (1968), Prior probabilities, *IEEE Transactions on Systems Science and Cybernetics* 4 (3), 227–251.
- [7] H. Jeffreys (1946), An invariant form for the prior probability in estimation problems, *Proceedings of the Royal Society of London (Ser. A)* 186, 453–461.
- [8] R.E. Kass, L. Wasserman (1996), Formal rules of selecting prior distributions : a review and annotated bibliography, *J. Amer. Statist. Assoc.* 91, 1343–1370.
- [9] C. P. Robert (2007), *The Bayesian choice, from decision-theoretic foundations to computational implementation*, Springer Science+Business Media, LLC.
- [10] B.L. Welch, H.W. Peers (1963), On formulae for confidence points based on integrals of weighted likelihoods, *J. Roy. Statist. Soc., Ser. B* 25, 318–329.