



HAL
open science

Estimations vraisemblables et fonctions pivotales

Francis Maisonneuve

► **To cite this version:**

| Francis Maisonneuve. Estimations vraisemblables et fonctions pivotales. 2011. hal-00519962v5

HAL Id: hal-00519962

<https://hal.science/hal-00519962v5>

Preprint submitted on 15 Nov 2011 (v5), last revised 25 May 2013 (v7)

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.

Estimations vraisemblables et fonctions pivotales

Résumé : Quitte à se limiter à quelques modèles statistiques usuels (et 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.

Pour les modèles exponentiels à un paramètre réel qui ne s’y prêtent pas, on propose un mode d’estimation de substitution adossé à des intervalles de confiance bien choisis. Ceci demande de nouveaux développements dans le cas où la loi de l’estimateur efficace est discrète. On retrouve en les ayant justifiées certaines façons de procéder habituelles.

Ce travail statistique est de facture purement classique (fréquentiste), aucune densité de probabilité bayésienne n’étant introduite dans l’espace des paramètres à côté ou à la place de la fonction de vraisemblance. On indiquera pourtant dès le début de l’étude une façon élémentaire (mais apparemment inusitée) de prendre correctement en compte de l’*information a priori* sur le paramètre.

On vérifiera aussi en fin d’étude qu’on peut, sur un plan formel, *probabiliser a posteriori* l’espace des paramètres des modèles à vraisemblance pivotale, mais seulement de manière “superficielle” en vue de *formuler* les résultats d’estimation. On peut même probabiliser véritablement l’espace des paramètres de certains modèles exponentiels à un paramètre réel, mais en commettant à notre avis un abus : on retombe alors sur la notion de *probabilité fiduciaire* introduite par R.A. Fisher, qui a tenté sans y parvenir vraiment d’en donner une définition générale et d’en faire un outil rigoureux d’inférence statistique.

Mots-clés : statistique paramétrique ; théorie de l’estimation ; fonction pivotale ; région de confiance ; région vraisemblable ; invariance par reparamétrage ; information a priori ; probabilité fiduciaire.

Statistical models with pivotal likelihood

Abstract : We consider a few common (and 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 *likelihood regions*, that is to say sets of relatively high likelihood. The proposed procedure involve no arbitrary element or choice (*intrinsic* estimation) ; furthermore it is equivariant under reparameterization, as do maximum likelihood point estimations associated with these set estimations.

For exponential models with one real parameter which does not satisfy the aforementioned condition, we propose an alternative estimation mode, connected with convenient confidence intervals. It requires new developments when the law of the efficient estimator is discrete. We get back to some usual methods with a full justification.

This work relies only on classical (frequentist) statistics : no Bayesian probability density is introduced into the parameter space, neither alongside nor in place of the likelihood function. However, we point out at the beginning an elementary (but apparently unusual) way of taking proper account of some *a priori information* on the parameter.

At the end of this study, we also check it is possible, on a formal level, to put an *a posteriori probability* on the parameter space of the models having pivotal likelihood. We can do it only superficially, in order to *express* the estimation results. For some exponential models with one real parameter, it can even be done in a genuine way we however believe unreasonable : we in fact retrieve the *fiducial probability* introduced by R.A. Fisher, who attempted without complete success to give this notion a general definition and make it a rigorous tool for statistical inference.

Keywords : parametric statistics ; estimation theory ; pivotal function ; confidence region ; likelihood region ; equivariance under reparameterization ; a priori information ; fiducial probability.

Sommaire

1	Généralités sur la vraisemblance	4
1.A	Cadre général	4
1.B	Paramètres utiles et nuisibles	4
1.C	Information a priori	5
2	Vraisemblance pivotale et régions de confiance	6
2.A	L'hypothèse de vraisemblance pivotale	6
2.B	Réflexions sur le choix de l'hypothèse	9
2.C	Cas de statistiques exhaustives	10
2.D	Cas des modèles exponentiels réels	11
3	Premiers exemples	12
3.A	Une situation asymptotique assez générale	12
3.B	Le modèle des exponentielles translatées	14
3.C	Le modèle des lois uniformes	15
3.D	Paramètre d'échelle sur les lois Gamma—Weibull	16
4	Exemple du modèle des lois gaussiennes	18
4.A	Estimation du couple (espérance, écart-type)	18
4.B	Estimation de l'espérance à écart-type connu ou inconnu	20
4.C	Estimation de l'écart-type à espérance connue ou inconnue	21
5	Modèle exponentiel réel, cas diffus	23
5.A	Une fonction pivotale intrinsèque	23
5.B	Estimation de substitution	24
5.C	Exemple du modèle des lois Gamma	26
6	Modèle exponentiel réel, cas discret	28
6.A	Une nouvelle fonction pivotale intrinsèque	28
6.B	Estimation dans le cas discret	31
6.C	Exemple du modèle exponentiel pile ou face	37
6.D	Comparaison avec d'autres méthodes pour pile ou face	40
7	Probabilisation possible de Θ	43
7.A	La famille des probabilités a posteriori	43
7.B	Cas des modèles exponentiels réels	45
	Index alphabétique	47

1 Généralités sur la vraisemblance

1.A Cadre général

On considère un modèle statistique 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,
avec $\forall \theta \in \Theta$, $f_\theta \mu$ loi de X selon P_θ .

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

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

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

Introduisons la *notation croisée* (ou *inversée*) pour la fonction de vraisemblance

$$\theta \in \Theta, f_x(\theta) \stackrel{\text{d\u00e9f}}{=} f_\theta(x_1, \dots, x_n).$$

Rappelons que la vraisemblance en x est une notion définie à un facteur multiplicatif (strictement positif) près pour chaque valeur fixée de x . Ainsi lorsqu'on remplace la mesure dominante μ par n'importe quelle mesure $\bar{\mu}$ équivalente de densité $b > 0$ par rapport à μ , on a en notant $f_\theta \bar{\mu} = f_\theta \mu$ la loi de X selon P_θ pour tout $\theta \in \Theta$ et en utilisant la notation croisée :

$$\bar{f}_x = \frac{1}{B(x)} f_x \quad \text{o\u00f9 } B(x) = \prod_{i=1}^n b(x_i);$$

ce qu'on écrit en abrégé $\bar{f}_x \propto f_x$ (\propto pour *positivement proportionnel* à).

La fonction de vraisemblance f_x en x exprime en fait essentiellement une relation d'ordre sur Θ (variable avec x), sur laquelle est fondée l'estimation du paramètre θ au maximum de vraisemblance. Cette relation d'ordre est inchangée lorsqu'on compose f_x par une fonction strictement croissante (typiquement le logarithme). Elle détermine la famille des *r\u00e9gions vraisemblables* pour θ en $x \in \Delta^n$

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

qui sont autant d'*estimations de θ par r\u00e9gions vraisemblables* en l'\u00e9chantillon x , dont il reste \u00e0 pr\u00e9ciser le niveau de confiance.

Soulignons le caract\u00e8re essentiellement fr\u00e9quentiste de cette analyse. Les estimations apparent\u00e9es en conception bay\u00e9sienne sont elles fond\u00e9es sur les grandes valeurs de la *densit\u00e9 a posteriori du param\u00e8tre*; ce qui conduit d'ailleurs \u00e0 certains "paradoxes" bien connus, dus au manque d'invariance des densit\u00e9s par reparam\u00e9trage...

1.B Param\u00e8tres utiles et nuisibles

Dans le cas d'un mod\u00e8le \u00e0 param\u00e8tre θ vectoriel ($d \geq 2$), il est fr\u00e9quent de s'int\u00e9resser plus sp\u00e9cialement \u00e0 l'estimation de certaines composantes de θ , les autres ayant des

valeurs connues ou non. Précisons les relations entre estimation globale et de telles estimations partielles de θ par régions vraisemblables :

Pour tout sous-ensemble strict J de $\{1, \dots, d\}$ de cardinal d_J ($1 \leq d_J \leq d-1$) et pour tout $\theta \in \Theta$, soit θ_J le d_J -uplet “projection selon J ” de θ constitué de ses composantes d’indices dans J et θ_{J^c} le d_{J^c} -uplet constitué de ses composantes d’indices dans le complémentaire $J^c = \{1, \dots, d\} \setminus J$ de J ; par abus, le couple (θ_J, θ_{J^c}) désignera le d -uplet θ de départ, en identifiant Θ à un sous-ensemble de $\Theta_J \times \Theta_{J^c}$ quitte à réordonner les composantes.

- Si θ_{J^c} est connu, la famille des régions vraisemblables pour θ_J en $x \in \Delta^n$ est

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

$\{f_x(\cdot, \theta_{J^c}) > h\}$ est la **section** à θ_{J^c} de la région vraisemblable pour θ en x $\{f_x > h\}$. Ceci revient à prendre comme fonction de vraisemblance de θ_J en x dans le cadre du modèle de référence la *vraisemblance conditionnelle* à θ_{J^c} (qui est semi-continue inférieurement, en abrégé *sci*, dès que f_x l’est) :

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

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

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

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

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

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

$\bigcup_{\theta_{J^c} \in \Theta_{J^c}} \{f_x(\cdot, \theta_{J^c}) > h\}$ est la **projection** sur Θ_J de la région vraisemblable pour θ en x $\{f_x > h\}$: on adopte ceci comme région 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 à prendre comme fonction de vraisemblance de θ_J en x dans le cadre du modèle de référence la *vraisemblance marginale* (qui est *sci* dès que f_x l’est), qu’on suppose à valeurs finies,

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

1.C Information a priori

En ce qui concerne l’expression d’une éventuelle information a priori sur le paramètre θ , il nous semble que la notion la plus pertinente n’est pas une densité de probabilité (approche bayésienne) mais une vraisemblance, pour les deux raisons suivantes :

- 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 changement de variable ; tel est le cas d’une vraisemblance, mais pas d’une densité de probabilité. . .

- l'absence complète d'information a priori s'exprime parfaitement par une vraisemblance constante : on échappe à la difficulté de fond que constitue en bayésien l'introduction de *lois a priori non informatives*, qui ne peuvent pas l'être vraiment...

On peut (comme en bayésien) facilement tirer parti de plusieurs sources indépendantes d'information a priori, en faisant le **produit** des fonctions de vraisemblance correspondantes : considérons ainsi 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 $\pi = f_{x'} : \theta \mapsto f_\theta(x')$

(il peut d'ailleurs être pratique de considérer les log-vraisemblances qui se combinent plus simplement par addition).

Dans cet esprit, on peut s'il y a lieu demander de spécifier une *vraisemblance a priori* π en imposant de la choisir dans la famille $(f_{x'})_{x' \in \bigcup_{m \in \mathbb{N}^*} \Delta^m}$, et adjoindre à l'échantillon obtenu $x \in \Delta^n$ un échantillon $x' \in \Delta^m$ (réel ou fictif, pas forcément déterminé de manière unique à l'ordre près) tel que $f_{x'} = \pi$; la fonction produit πf_x , qualifiée de *vraisemblance a posteriori en x* , étant simplement la vraisemblance en l'échantillon réunion (x', x) .

2 Vraisemblance pivotale et régions de confiance

2.A L'hypothèse de vraisemblance pivotale

On considère dans la suite un modèle statistique $(\Theta$ ouvert connexe de \mathbb{R}^d) tel que

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

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

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

(ou en cas de paramètre nuisible θ_{J^c} , de définir la vraisemblance marginale normalisée

$$\phi = \phi_J^{(n)} : \Delta^n \times \Theta_J \rightarrow [0, 1] \text{ par } \phi_x(\theta_J) = \frac{1}{\|f_x\|_\infty} \sup_{\{\theta_{J^c} : (\theta_J, \theta_{J^c}) \in \Theta\}} f_x(\theta_J, \theta_{J^c}).$$

Définition : 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 :

$\forall \theta \in \Theta$, la variable aléatoire composée $\phi_\theta(X_1, \dots, X_n)$ (qui dépend de θ) est telle que sa **loi** $\eta = \eta^{(n)}$ par rapport à P_θ ne **dépend pas** de θ :

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

ce qu'on note en abrégé

$$\boxed{\forall \theta \in \Theta, (\phi_\theta(X))(P_\theta) = \eta}.$$

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

L'hypothèse très forte de vraisemblance pivotale ci-dessus, qui exprime que la fonction de vraisemblance normalisée à 1 pour tout $x \in \Delta^n$ 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 :

$$\forall \theta \in \Theta \text{ et } \forall B \in \mathcal{B}(\mathbb{R}), \underline{P_\theta(\phi_X^{-1}(B) \ni \theta) = P_\theta(\phi_X(\theta) \in B) = \eta(B)} :$$

$\phi_X^{-1}(B)$ est pour tout $B \in \mathcal{B}(\mathbb{R})$ 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 la fréquence γ ; ce qui conduit pour un échantillon donné $x \in \Delta^n$ à prendre comme estimation ensembliste de θ au niveau de risque γ

$$\underline{\Theta_B^x \stackrel{\text{déf}}{=} \phi_x^{-1}(B) = \{\phi_x \in B\} = \{\theta \in \Theta : \phi_\theta(x_1, \dots, x_n) \in B\}},$$

appelé *région de confiance pour θ en l'échantillon x au niveau $\gamma = \eta(B)$* .

En fait pour éviter tout **paradoxe** dans l'estimation par régions de confiance, on va compléter l'hypothèse de vraisemblance pivotale par des conditions sur la mesure η :

- $\forall x \in \Delta^n$, η est portée 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 $0 < \eta(B) < \max_{x \in \Delta^n} \eta(\phi_x(\Theta)^c)$ vérifiant $B \subseteq \phi_x(\Theta)^c$ pour certains échantillons $x \in \Delta^n$; de sorte qu'on aurait pour ceux-ci $\Theta_B^x = \emptyset$ (pauvre estimation ensembliste!). Et surtout grâce à cette condition supplémentaire, la valeur du niveau de confiance $\eta(B)$ de Θ_B^x pour x et B quelconques n'a rien d'artificielle puisque

$$\Theta_B^x = \Theta_{B'}^x \iff B \cap \phi_x(\Theta) = \phi_x(\Theta_B^x) = \phi_x(\Theta_{B'}^x) = B' \cap \phi_x(\Theta);$$

de sorte que $\Theta_B^x = \Theta_{B'}^x \implies \eta(B) = \eta(B \cap \phi_x(\Theta)) = \eta(B' \cap \phi_x(\Theta)) = \eta(B')$.

Il s'agit ainsi d'une condition qu'on devrait exiger de toute fonction pivotale, et qui rend naturelle la normalisation de la vraisemblance.

- η diffuse, ou au moins $\eta(\{0\}) = \eta(\{1\}) = 0$.

La condition $\eta(\{0\}) = 0$ assure que pour tout $x \in \Delta^n$, la région de vraisemblance nulle $\{\phi_x = 0\}$ a un niveau de confiance nul : ce qui paraît être une condition minimum de cohérence entre ces deux notions! La condition $\eta(\{1\}) = 0$ assure de son côté un niveau de confiance nul à l'estimation (ponctuelle ou ensembliste) $\{\phi_x = 1\}$ de θ au maximum de vraisemblance en $x \dots$

Ces deux conditions se résument en : $\forall x \in \Delta^n, \phi_x(\Theta) \cap]0, 1[$ est porteur de η .

Nous dirons alors que ϕ est une fonction pivotale *correcte*.

Remarques

- Les notions de *région aléatoire de confiance pour θ au niveau γ* et de *fonction pivotale correcte* ne dépendent pas du choix de normalisation de ϕ à 1 plutôt qu'à une autre constante $k > 0$ (car $\phi' = k\phi$ implique $\phi_X'^{-1}(kB) = \phi_X^{-1}(B)$ et donc $\eta'(kB) = \eta(B)$ pour tout borélien B , et en particulier pour $B = \phi_x(\Theta)$).
Plus généralement, pour tout homéomorphisme $\ell : \mathbb{R} \rightarrow \mathbb{R}$, $\phi' = \ell \circ \phi$ est une fonction pivotale correcte, de loi fixe $\eta' = \ell(\eta)$ portée par $\phi'_x(\Theta)$ pour tout x ; et les régions aléatoires de confiance pour θ au niveau γ selon ϕ et ϕ' sont les mêmes ($\phi_X'^{-1}(\ell(B)) = \phi_X^{-1}(B)$ et donc $\eta'(\ell(B)) = \eta(B)$ pour tout borélien B).
- Comme $\phi(\Delta^n \times \Theta)$ porte η , ϕ est *correcte* si $\eta(\{0\}) = \eta(\{1\}) = 0$ et si on est dans l'un ou l'autre des cas suivants :
 - * $\phi_x(\Theta)$ ne dépend pas de $x \in \Delta^n$ (puisqu'il coïncide alors avec $\phi(\Delta^n \times \Theta)$);
 - * $\forall x \in \Delta^n, \phi_x(\Theta) \supseteq]0, 1[$ (puisque $\phi(\Delta^n \times \Theta) \subseteq [0, 1]$ et $\eta(\{0\}) = \eta(\{1\}) = 0$).
- En excluant le cas de paramètre nuisible, si f est telle que pour tout $x \in \Delta^n$, $\inf f_x(\Theta) = 0$ et f_x est continue sur Θ , il en est de même pour ϕ ; de sorte que la condition ci-dessus $\forall x \in \Delta^n, \phi_x(\Theta) \supseteq]0, 1[$ est automatiquement vérifiée.

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

$$y \in [0, 1[, \boxed{\{\phi_x > y\} \subseteq \Theta}, \text{ au niveau } \boxed{\gamma(y) = \eta(]y, 1])} \text{ indépendant de } x.$$

La fonction γ (prolongée par la valeur 1 sur \mathbb{R}_- et 0 sur $[1, +\infty[$) est la fonction de répartition complémentaire de la probabilité η sur \mathbb{R} , continue en 0 et en 1 puisque $\eta(\{0\}) = \eta(\{1\}) = 0$; si elle est continue sur \mathbb{R} (cas où η est diffuse), elle prend en décroissant toutes les valeurs de $]0, 1[$ et on a pour tout $x \in \Delta^n$

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

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

De par sa définition (fonction de vraisemblance normalisée), la fonction ϕ ne dépend pas du choix de la mesure dominante μ du modèle ($\forall x \in \Delta^n, \bar{f}_x \propto f_x \implies \bar{\phi}_x = \phi_x$); donc la mesure η et de là pour tout $x \in \Delta^n$, $\gamma \circ \phi_x$ n'en dépendent pas non plus.

Si on remplaçait ϕ par $\phi' = k\phi$, on aurait d'après la première remarque ci-dessus $\gamma'(ky) = \gamma(y)$ pour tout $y \in [0, 1[$; donc pour tout $x \in \Delta^n$, $\gamma' \circ \phi'_x = \gamma \circ \phi_x$. Plus généralement, pour tout homéomorphisme **croissant** $\ell : \mathbb{R} \rightarrow \mathbb{R}$, si on remplaçait ϕ par $\phi' = \ell \circ \phi$ on aurait

$$\forall y \in [0, 1[, \{\phi_x > y\} = \{\phi'_x > \ell(y)\} \text{ de niveau commun } \gamma(y) = \gamma'(\ell(y)) ;$$

donc à nouveau pour tout $x \in \Delta^n$, $\gamma' \circ \phi'_x = \gamma \circ \phi_x$.

Enfin soit $\theta \in \Theta \xrightarrow{\delta} \tilde{\theta} \in \tilde{\Theta}$ un homéomorphisme de Θ sur un autre ouvert $\tilde{\Theta}$ de \mathbb{R}^d . Comme pour tout $x \in \Delta^n$ $\tilde{f}_x(\tilde{\theta}) = f_x(\theta)$ on a aussi $\tilde{\phi}_x(\tilde{\theta}) = \phi_x(\theta)$, autrement dit $f_x = \tilde{f}_x \circ \delta$ et $\phi_x = \tilde{\phi}_x \circ \delta$ (on parle d'*invariance par reparamétrage*) ; donc η n'est pas affectée par δ . En conséquence on a $\forall y \in [0, 1[$, $\delta(\{\phi_x > y\}) = \{\tilde{\phi}_x > y\}$ (invariance par reparamétrage des régions vraisemblables pour θ en x) ; et comme la fonction γ est comme η non affectée par δ , on a aussi invariance par reparamétrage pour $\gamma \circ \phi_x : \gamma \circ \phi_x = (\gamma \circ \tilde{\phi}_x) \circ \delta$.

En **résumé**, sous l'hypothèse très forte de *vraisemblance pivotale correcte*, la famille $(\{\phi_x > y\})_{y \in [0, 1]}$ des régions vraisemblables de confiance pour θ en x est *intrinsèque* au modèle statistique, au sens où la définition de ces régions et de leur niveau de confiance ne comportent aucune part d'arbitraire ; contrairement par exemple aux intervalles de confiance d'un paramètre réel qui dépendent d'un choix conventionnel de répartition du risque (en général "à risque équilibré" faute de mieux), et aussi plus fondamentalement du choix arbitraire d'une fonction pivotale parmi de nombreuses possibles ! De plus, les niveaux de confiance ne dépendent que de la "hauteur" y , la fonction $\gamma \circ \phi_x$ synthétisant l'information pour tout $x \in \Delta^n$.

L'estimation par régions vraisemblables pour θ nous semble "optimale" au sens évident où elle maximise la vraisemblance du paramètre. Comme pour l'estimation $\hat{\theta}$ au maximum de vraisemblance 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 au choix d'un paramétrage particulier ; au contraire des 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 privilégie ou non les cas où un paramétrage particulier s'impose, cette invariance apparaît comme un défaut ou au contraire comme une qualité structurelle supplémentaire.

2.B Réflexions sur le choix de l'hypothèse

On part de $\forall \theta \in \Theta$ et $\forall B \in \mathcal{B}(\mathbb{R})$, $P_\theta(\phi_X^{-1}(B) \ni \theta) = P_\theta(\phi_X(\theta) \in B) = \eta(B)$,

et pour tout échantillon $x \in \Delta^n$ on attribue à la région de confiance $\phi_x^{-1}(B)$ le niveau $\gamma = \eta(B)$. Ceci suppose en fait 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 pour cette raison qu'on s'est résolu à poser la très stricte *hypothèse de vraisemblance pivotale* plutôt que :

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

cette dernière hypothèse exprimant 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 vraisemblable pour θ (en x) $\{\phi_x > y\}$ s'exprime au moyen de Φ_x par

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

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

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, en plus du caractère injustifié des niveaux de confiance qu'on attribuerait alors aux régions vraisemblables pour θ en x , on perdrait aussi le caractère intrinsèque de ces niveaux du fait de la multiplicité des fonctions pivotales Φ possibles.

Par contre si G_x ne dépend pas de x , $\phi = G \circ \Phi$ vérifie la propriété pivotale, correcte si par exemple la loi image $\eta = G(v)$ est diffuse.

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

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

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 vraisemblables en $x \in \Delta^n$ est encore de la forme (cf. les paragraphes 1.A et 2.A)

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

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 $\ell_x \stackrel{\text{déf}}{=} G'_x{}^{-1} \circ G_x$ soit indépendant de x ; or c'est seulement sous cette condition d'indépendance qu'on est assuré que les niveaux de confiance attribués ne dépendent pas du choix arbitraire du couple (Φ, G) et donc sont bien intrinsèques au modèle (cf. aussi le paragraphe 2.A).

2.C Cas de statistiques exhaustives

On sait que disposer d'une **statistique exhaustive** $T = \tau(X_1, \dots, X_n)$ à valeurs dans \mathbb{R}^p pour le modèle considéré équivaut à avoir 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 . Couramment $p = d$.

f_θ^T est pour tout $\theta \in \Theta$ la densité de T par rapport à la mesure $\underline{\mu} = \tau(B \mu^{n \otimes})$ sur \mathbb{R}^p , image par τ de la mesure de densité B par rapport à $\mu^{n \otimes}$: on a en effet

$$T(P_\theta) = \tau((X_1, \dots, X_n)(P_\theta)) = \tau(f_\theta \mu^{n \otimes}) = \tau(B (f_\theta^T \circ \tau) \mu^{n \otimes}) = f_\theta^T \tau(B \mu^{n \otimes}).$$

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 $\underline{f}_t^T \propto f_x$.

En supposant toujours que $\forall x \in \Delta^n, M(x) \stackrel{\text{déf}}{=} \|f_x\|_\infty \in]0, +\infty[$, on a $\underline{f}_t^T \propto f_x \propto \phi_x$. 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 $\tau(x)$:

$$\forall (x, \theta) \in \Delta^n \times \Theta, \phi(x, \theta) = \phi^T(\tau(x), \theta) = \phi^T(t, \theta),$$

et $\phi^T : \tau(\Delta^n) \times \Theta \rightarrow \mathbb{R}_+$ est l'application borélienne vérifiant pour tout $t \in \tau(\Delta^n)$, $f_t^T \propto \phi_t^T$ et $\|\phi_t^T\|_\infty = 1$. Si on note $M^T(t) \stackrel{\text{déf}}{=} \|f_t^T\|_\infty$ pour tout $t \in \tau(\Delta^n)$, on a $f_t^T = M^T(t) \phi_t^T$ et $M(x) = B(x) M^T(t)$, donc $f_x = B(x) f_t^T = M(x) \phi_t^T$; ainsi 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 $\dot{\mu} = \tau(M \mu^{n\otimes})$.

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

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

et on peut si besoin remplacer ci-dessus $\tau(\Delta^n)$ par I , où I est n'importe quel sous-ensemble porteur de μ . On dira encore que ϕ^T est une *fonction pivotale correcte* (du modèle image).

On retiendra que la relation $\phi_x = \phi_{\tau(x)}^T$ assure qu'une région vraisemblable pour θ à un niveau de confiance $\gamma \in]0, 1[$ fixé (déterminé selon ϕ) ne **dépend de l'échantillon x que par la valeur $t = \tau(x)$ de n'importe quelle statistique exhaustive $T = \tau(X_1, \dots, X_n)$** (ce qui incite à choisir T minimale et même complète quand c'est possible).

En ce qui concerne l'information a priori (cf. le paragraphe 1.C), la vraisemblance a priori est choisie dans la famille $(f_{x'})_{x' \in \bigcup_{m \in \mathbb{N}^*} \Delta^m}$, donc aussi bien dans la famille normalisée $(\phi_{t'}^{T_m})_{t' \in \bigcup_{m \in \mathbb{N}^*} \tau_m(\Delta^m)}$ où pour tout $m \in \mathbb{N}^*$, $T_m = \tau_m(X_1, \dots, X_m)$ est une statistique exhaustive du modèle; ce qui revient à préciser simplement la valeur de $m \in \mathbb{N}^*$ et une valeur $t' \in \tau_m(\Delta^m)$.

En effet soit $x' \in \Delta^m$ tel que $\tau_m(x') = t'$; on a, en notant $t = \tau_n(x)$ où $x \in \Delta^n$ et (x', x) l'échantillon réunion,

$$\phi_{(x', x)} \propto f_{(x', x)} = f_{x'} f_x \propto \phi_{x'} \phi_x = \phi_{t'}^{T_m} \phi_t^{T_n},$$

c'est-à-dire $\phi_{(x', x)} = k' \phi_{t'}^{T_m} \phi_t^{T_n}$ où la constante $k' > 0$ est caractérisée par

$$k' = \frac{1}{\|\phi_{t'}^{T_m} \phi_t^{T_n}\|_\infty} \geq \frac{1}{\|\phi_{t'}^{T_m}\|_\infty \|\phi_t^{T_n}\|_\infty} = 1.$$

2.D Cas des modèles exponentiels réels

On rappelle qu'il s'agit du modèle régulier (qu'on dit réel, car de dimension $d = 1$)

$$\Delta \subseteq \mathbb{R} \quad \text{et} \quad \forall \theta \in \Theta, \underline{f_\theta(x) = \bar{b}(x) e^{\alpha(\theta) \bar{a}(x) + \beta(\theta)}} \quad \text{où } \Theta \text{ intervalle ouvert de } \mathbb{R},$$

dans lequel α et β sont de classe C^2 sur Θ , $\alpha' \neq 0$ partout, $\bar{a}(X)$ n'est pas presque sûrement constant et $\bar{b} > 0$ sur Δ , avec $e^{-\beta(\theta)} = \int \bar{b}(x) e^{\alpha(\theta) \bar{a}(x)} d\mu(x)$.

Soit $T = \frac{1}{n} \sum_{i=1}^n \bar{a}(X_i)$ l'estimateur efficace usuel du modèle qui en est une statistique

exhaustive, de densité $\underline{f_t^T(\theta) \stackrel{\text{déf}}{=} e^{n[\alpha(\theta)t + \beta(\theta)]} > 0$ (en prenant $B(x) = \prod_{1 \leq i \leq n} \bar{b}(x_i)$).

$\forall t \in \mathbb{R}$, f_t^T est de classe C^2 et monotone (de dérivée ne s'annulant pas) ou d'opposée unimodale sur Θ , comme on le voit par le changement de variable $\delta \stackrel{\text{déf}}{=} \alpha$ à partir de la concavité stricte de l'application

$$\tilde{\theta} \mapsto \frac{1}{n} \ln(f_t^T \circ \alpha^{-1})(\tilde{\theta}) = \tilde{\theta} t + \tilde{\beta}(\tilde{\theta}) \text{ sur l'intervalle ouvert image } \tilde{\Theta} = \alpha(\Theta) :$$

on a en effet le résultat classique, en notant $h(\theta) \stackrel{\text{déf}}{=} E_\theta(T)$,

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

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

D'un autre côté h dont la dérivée est de signe constant est un C^1 -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 d'opposée unimodale, de sorte que $h^{-1}(t)$ est l'estimation $\hat{\theta}$ de θ au maximum de vraisemblance.

Supposons à présent que $I \stackrel{\text{déf}}{=} h(\Theta)$ porte $\dot{\mu} = \tau(B \mu^{n \otimes})$; alors ϕ^T est bien définie sur $I \times \Theta$ puisque $\|f_t^T\|_\infty = e^{n[\alpha(h^{-1}(t))t + \beta(h^{-1}(t))]} \in]0, +\infty[$; et $\forall t \in I$, $\phi_t^T > 0$ est de classe C^2 , d'opposée unimodale et atteint son maximum 1 en $\hat{\theta} = h^{-1}(t)$: d'où

$$\forall (t, \theta) \in I \times \Theta, \quad \underline{-\frac{1}{n} \ln \phi_\theta^T(t) = [\alpha(h^{-1}(t)) - \alpha(\theta)] t + [\beta(h^{-1}(t)) - \beta(\theta)]}.$$

$h^{-1}(T)$ est l'estimateur de θ au maximum de vraisemblance et on a

$$\forall \theta \in \Theta, \quad -\frac{1}{n} \ln \phi_\theta^T(T) = [\alpha(h^{-1}(T)) - \alpha(\theta)] T + [\beta(h^{-1}(T)) - \beta(\theta)].$$

Ainsi lorsque $h(\Theta)$ porte $\dot{\mu}$, l'hypothèse de vraisemblance pivotale est vérifiée par le modèle 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 .

Il faut de plus que cette loi n'ait pas d'atome en 0 pour que la loi η de $\phi_\theta^T(T) = \phi_\theta(X_1, \dots, X_n)$ vérifie $\eta(\{1\}) = 0$ (la condition $\eta(\{0\}) = 0$ est automatiquement vérifiée puisque $\phi_\theta^T(T) > 0$; η est diffuse si et seulement si cette loi est diffuse). Et enfin par exemple la condition couramment vérifiée

$$\forall t \in I, \quad \underline{\inf f_t^T(\Theta) = 0}$$

suffit à assurer que $\forall t \in I$, $\phi_t^T(\Theta) \in]0, 1]$, de sorte que η est bien portée par $\phi_t^T(\Theta)$ (cf. les paragraphes 2.A, remarques et 2.C).

Si $h(\Theta)$ ne porte pas $\dot{\mu}$, f_t^T est strictement monotone en tout point t d'un sous-ensemble non $\dot{\mu}$ -négligeable A ; et ϕ_t^T n'est définie en $t \in A$ que si $\|f_t^T\|_\infty < +\infty$.

3 Premiers exemples

3.A Une situation asymptotique assez générale

On suppose ici que pour tout $x = (x_1, \dots, x_n) \in \Delta^n$ 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

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

c'est-à-dire une convergence en loi vers une variable du chi-deux qui ne **dépend pas de θ ni même du modèle**, mais seulement du nombre d de paramètres réels composant θ . $\phi = \frac{f}{M}$ vérifie donc la propriété pivotale de manière asymptotique, puisque la loi η_θ de $\phi_\theta(X_1, \dots, X_n)$ est assimilable à la loi (diffuse) η de $e^{-\frac{\chi_d^2}{2}}$ pour n assez grand.

Sous réserve par exemple que la régularité de f assure $\phi_x(\Theta) \supseteq]0, 1[$, on a d'après le paragraphe 2.A que pour tout $x \in \Delta^n$ et pour tout $\gamma \in]0, 1[$ fixé,

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

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

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

Cas particulier des modèles exponentiels réels

Comme vu au paragraphe 2.D, sous l'hypothèse que $h(\Theta)$ porte $\dot{\mu}$ et que $\forall t \in h(\Theta)$, $\inf f_t^T(\Theta) = 0$, on a donc que pour $\dot{\mu}$ -presque tout t la région vraisemblable pour θ au niveau de confiance asymptotique $\gamma = P(\chi_1^2 < z) \in]0, 1[$ est l'intervalle ouvert

$$] \theta_t^1, \theta_t^2 [= \left\{ \phi_t^T > e^{-\frac{z}{2}} \right\} = \left\{ \theta \in \Theta : \alpha(\theta)t + \beta(\theta) > \alpha(h^{-1}(t))t + \beta(h^{-1}(t)) - \frac{z}{2n} \right\}.$$

Exemple du modèle exponentiel des lois Gamma réduites

On considère le modèle exponentiel de paramètre de forme θ

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

(λ désigne la mesure de Lebesgue sur \mathbb{R}), dont l'estimateur efficace $T = \frac{1}{n} \sum_{i=1}^n \ln X_i$

a pour densité par rapport à $\dot{\mu}$

$$f_\theta^T(t) = \frac{1}{\Gamma(\theta)^n} e^{n\theta t} = e^{n[\theta t - \ln \Gamma(\theta)]}.$$

On a donc $h = -\frac{\beta'}{\alpha'} = \frac{\Gamma'}{\Gamma} \stackrel{\text{déf}}{=} \psi$, fonction *digamma*, qui est strictement croissante de \mathbb{R}_+^* sur \mathbb{R} ; ainsi $h(\Theta) = \mathbb{R}$, donc $h(\Theta)$ porte $\dot{\mu}$; et $\forall t \in \mathbb{R}$, $\inf f_t^T(\Theta) = 0$. De sorte que, bien que l'hypothèse de vraisemblance pivotale ne soit pas vérifiée, la région vraisemblable pour θ au niveau de confiance asymptotique $\gamma = P(\chi_1^2 < z)$ est

$$] \theta_t^1, \theta_t^2 [= \left\{ \theta \in \Theta : t\theta - \ln \Gamma(\theta) > C_t - \frac{z}{2n} \right\} \text{ où } C_t \stackrel{\text{déf}}{=} t\psi^{-1}(t) - \ln \Gamma(\psi^{-1}(t)).$$

Illustration pour $n = 100$, $\gamma = 0.95$ ($z \approx 3.841$) et $t = -1$: $] \theta_t^1, \theta_t^2 [\approx] 0.66, 0.92 [$.

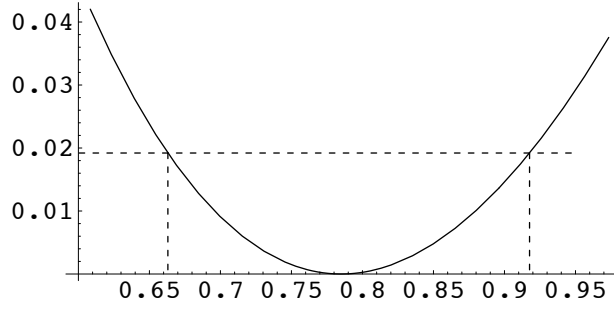
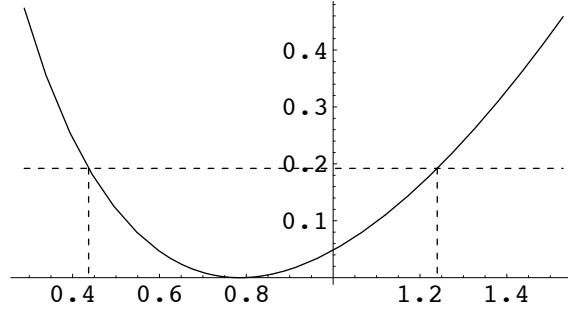


Illustration pour $n = 10$, $\gamma = 0.95$ et $t = -1$: $]\theta_t^1, \theta_t^2[\approx]0.44, 1.24[$.



3.B Le modèle des exponentielles translatées

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

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

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 puisque $\forall x \in \Delta^n, f_x \propto f_t^T$ avec $f_t^T(\theta) \stackrel{\text{déf}}{=} e^{n\theta} \mathbb{1}_{]-\infty, t[}(\theta)$; on a alors $\|f_t^T\|_\infty = e^{nt} \in]0, +\infty[$ et $\phi_t^T(\theta) = e^{-n(t-\theta)} \mathbb{1}_{]-\infty, t[}(\theta)$ (ϕ_t^T a pour image $]0, 1[$ qui ne dépend pas de t). Comme $\forall i \in \{1, \dots, n\}, X_i - \theta \stackrel{\mathcal{L}}{=} S_1$, où S_1 est la variable exponentielle de paramètre 1, et comme

$$\forall \theta \in \Theta, T - \theta = \min(X_1 - \theta, \dots, X_n - \theta),$$

on a $T - \theta \stackrel{\mathcal{L}}{=} \frac{S_1}{n}$. Puisque $\phi_\theta^T(T) = e^{-n(T-\theta)}$ du fait que $T > \theta$, on en déduit que ϕ^T vérifie la propriété pivotale avec η loi de e^{-S_1} , qui est la probabilité uniforme sur $]0, 1[$ (donc est diffuse).

On a donc $\forall y \in]0, 1[, \gamma(y) = \eta(]y, +\infty[) = 1 - y$; de sorte que pour tout $t \in \mathbb{R}$ et pour tout $\gamma \in]0, 1[$ fixé,

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

est la région vraisemblable pour θ au niveau de confiance γ .

On notera que cet intervalle de confiance diffère de l'intervalle équilibré habituel de niveau γ

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

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 : La **densité** $\Phi_x(\theta) \stackrel{\text{déf}}{=} f_\theta(x_1, \dots, x_n) = \prod_{1 \leq i \leq n} (e^{-(x_i - \theta)} \mathbb{1}_{]0, +\infty[}(x_i - \theta))$

est elle-même une fonction pivotale puisque pour tout $\theta \in \Theta$, du fait que $X > \theta$,

$$\Phi_X(\theta) = f_\theta(X_1, \dots, X_n) = \prod_{1 \leq i \leq n} e^{-(X_i - \theta)} \stackrel{\mathcal{L}}{=} \prod_{1 \leq i \leq n} e^{-S_1^{(i)}} \stackrel{\mathcal{L}}{=} e^{-S_n}$$

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 . Mais Φ ne vérifie pas la condition $\forall x \in \Delta^n$, $\Phi_x(\Theta)$ porteur de la loi (diffuse) de e^{-S_n} , puisque le support de e^{-S_n} est $[0, 1]$ alors que $\Phi_x(\Theta) = \left[0, e^{n \min(x_1, \dots, x_n) - \sum_{i=1}^n x_i} \right[$ est un intervalle de longueur strictement inférieur à 1 dès que les x_i ne sont pas tous égaux entre eux. Ceci suffit à rendre condamnable l'emploi de Φ non normalisée pour toute estimation par régions de confiance (cf. le paragraphe 2.A).

Passons outre pour s'en convaincre : un niveau $\gamma \in]0, 1[$ 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 vraisemblable pour θ au niveau de confiance γ ainsi déterminé selon Φ serait

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

Cet intervalle dépend non seulement de la valeur $t = \min(x_1, \dots, x_n)$ de la statistique exhaustive T , mais aussi de $\frac{1}{n} \sum_{i=1}^n x_i$ par sa borne inférieure, ce qui paraît saugrenu (cf. à ce sujet le paragraphe 2.C). L'intervalle peut même être vide si par aventure l'un des x_i de l'échantillon est très inférieur aux autres !

Cette analyse montre au passage que la situation n'est pas simple même dans le cas de modèles à un seul paramètre de position.

3.C Le modèle des lois uniformes

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

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

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 clairement une statistique exhaustive : on peut prendre $f_t^T \stackrel{\text{déf}}{=} f_x$, donc $\|f_t^T\|_\infty = \frac{1}{t^n}$ et $\phi_t^T(\theta) = \left(\frac{t}{\theta}\right)^n \mathbb{1}_{]t, +\infty[}(\theta)$ (ϕ_t^T a pour image $[0, 1[$ qui ne dépend pas de t). Comme

$$\forall \theta \in \Theta, \phi_\theta^T(T) = \left(\frac{T}{\theta}\right)^n = \left[\max\left(\frac{X_1}{\theta}, \dots, \frac{X_n}{\theta}\right)\right]^n$$

où les $\frac{X_i}{\theta}$ sont indépendants et de loi uniforme sur $]0, 1[$, $\left(\frac{T}{\theta}\right)^n$ a cette même loi : ϕ^T vérifie la **propriété pivotale** avec η probabilité uniforme sur $]0, 1[$ (donc diffuse).

On a donc $\forall y \in]0, 1[$, $\gamma(y) = \eta(]y, +\infty[) = 1 - y$; de sorte que pour tout $t > 0$ et pour tout $\gamma \in]0, 1[$ fixé,

$$\underline{\{\phi_t^T > 1 - \gamma\}} = \left]t, t \left(\frac{1}{1 - \gamma}\right)^{\frac{1}{n}}\right[$$

est la région vraisemblable pour θ au niveau de confiance γ .

On notera que cet intervalle de confiance diffère de l'intervalle équilibré habituel de niveau γ

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

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 : La fonction $\Phi_x(\theta) \stackrel{\text{déf}}{=} x_1 \cdots x_n f_\theta(x_1, \dots, x_n) = \prod_{1 \leq i \leq n} \left[\frac{x_i}{\theta} \mathbb{1}_{]0, 1[}\left(\frac{x_i}{\theta}\right)\right]$

est aussi une fonction pivotale puisque pour tout $\theta \in \Theta$, du fait que $X < \theta$,

$$\Phi_X(\theta) = \left(\prod_{1 \leq i \leq n} X_i\right) f_\theta(X_1, \dots, X_n) = \prod_{1 \leq i \leq n} \frac{X_i}{\theta} \stackrel{\mathcal{L}}{=} \prod_{1 \leq i \leq n} U^{(i)} \stackrel{\mathcal{L}}{=} \prod_{1 \leq i \leq n} e^{-S_1^{(i)}} \stackrel{\mathcal{L}}{=} e^{-S_n}$$

où les $U^{(i)}$ sont des variables uniformes sur $]0, 1[$ indépendantes. Mais comme dans l'exemple précédent Φ ne vérifie pas la condition $\forall x \in \Delta^n$, $\Phi_x(\Theta)$ porteur de la loi de e^{-S_n} , puisque la longueur de l'intervalle $\Phi_x(\Theta) = \left[0, \frac{x_1 x_2 \cdots x_n}{\max(x_1, \dots, x_n)^n}\right]$ est strictement inférieur à 1 dès que les x_i ne sont pas tous égaux entre eux ; ce qui suffit à rendre condamnable l'emploi de Φ pour toute estimation par régions de confiance (cf. les paragraphes 2.A et 3.B).

Cette analyse montre au passage que la situation n'est pas simple même dans le cas de modèles à un seul paramètre d'échelle...

3.D Paramètre d'échelle sur les lois Gamma — Weibull

Soient deux réels $a, b > 0$ fixés ; on considère le modèle de paramètre d'échelle $\frac{1}{\theta}$

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

Γ désignant la fonction d'Euler. Il s'agit d'un modèle exponentiel de forme standard

$$f_\theta(x) = \bar{b}(x) e^{\alpha(\theta) \bar{a}(x) + \beta(\theta)} \text{ avec}$$

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

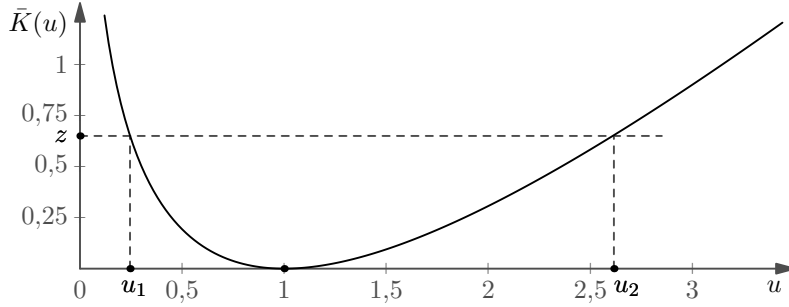
Pour $b = 1$ il s'agit des lois Gamma de paramètre de forme a ; pour $a = 1$ il s'agit des lois de Weibull de paramètre de forme b ; et pour $a = b = 1$, on a affaire au modèle des durées de vie exponentielles.

On a ici $T = \frac{1}{n} \sum_{i=1}^n X_i^b$ et $h(\theta) = \frac{a b}{\theta} \frac{1}{b \theta^{b-1}} = \frac{a}{\theta^b}$, donc $h(\Theta) = \mathbb{R}_+^* = \tau(\Delta^n)$

avec $h^{-1}(t) = \left(\frac{a}{t}\right)^{\frac{1}{b}}$; de sorte que (cf. le paragraphe 2.D)

$$\forall \theta \in \Theta, -\frac{1}{n} \ln \phi_\theta^T(T) = \left(-\frac{a}{T} + \theta^b\right) T + \left(a \ln \frac{a}{T} - a b \ln \theta\right) = a \bar{K}\left(\frac{\theta^b}{a} T\right)$$

en ayant posé $\forall u > 0, \boxed{\bar{K}(u) \stackrel{\text{déf}}{=} u - \ln u - 1}$.



\bar{K} est une fonction positive et strictement convexe sur \mathbb{R}_+^* , de minimum 0 en $u = 1$ et de limite $+\infty$ en 0 et à l'infini.

Comme $\forall i \in \{1, \dots, n\}$, θX_i a pour densité sur \mathbb{R}_+ $\frac{b}{\Gamma(a)} y^{a b-1} e^{-y^b}$, $(\theta X_i)^b$ a pour densité sur \mathbb{R}_+ $\frac{1}{\Gamma(a)} z^{a-1} e^{-z}$, c'est-à-dire $(\theta X_i)^b \stackrel{\mathcal{L}}{=} S_a$. En conséquence la fonction

$\Phi^T(t, \theta) \stackrel{\text{déf}}{=} \frac{\theta^b}{a} t$ vérifie $\Phi_\theta^T(T) = \frac{\theta^b}{a} T = \frac{1}{na} \sum_{i=1}^n (\theta X_i)^b \stackrel{\mathcal{L}}{=} \frac{S_{na}}{na} \stackrel{\text{déf}}{=} U$ où S_{na} est la variable gamma réduite de paramètre de forme na (de loi diffuse). On en déduit que $a \bar{K}\left(\frac{\theta^b}{a} T\right) = a (\bar{K} \circ \Phi_\theta^T)(T)$ a aussi sa loi (diffuse) qui ne dépend pas de θ , de sorte que **la propriété pivotale est vérifiée par ϕ^T** ; et comme $\phi_t^T(\theta) = e^{-na \bar{K}\left(\frac{\theta^b}{a} t\right)}$, ϕ_t^T a pour image $]0, 1]$ qui ne dépend pas de t .

Comme détaillé ci-dessous, pour tout $t > 0$ et pour tout $\gamma \in]0, 1[$ fixé,

$$\left\{ \phi_t^T > H(z) \stackrel{\text{déf}}{=} e^{-na z} \right\} = \left\{ \theta \in \Theta : \bar{K}\left(\frac{\theta^b}{a} t\right) < z \right\} = \left] \left(\frac{a u_1}{t}\right)^{\frac{1}{b}}, \left(\frac{a u_2}{t}\right)^{\frac{1}{b}} \right[$$

est alors la région vraisemblable pour θ au niveau de confiance γ fixé pour $u_1, u_2 > 0$ définis par

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

On notera 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 de niveau γ , à savoir

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

4 Exemple du modèle des lois gaussiennes

Le modèle des lois gaussiennes va permettre d'illustrer la démarche pratique consistant à choisir une statistique exhaustive $T = \tau(X_1, \dots, X_n)$ pour le modèle (en général un estimateur au maximum de vraisemblance), puis à chercher une factorisation de ϕ^T de la forme

$$\forall t, \phi_t^T = H \circ K \circ \Phi_t^T \quad (\text{soit } \phi^T = H \circ K \circ \Phi^T)$$

telle que

- $\Phi^T : \tau(\Delta^n) \times \Theta \rightarrow \mathbb{R}^d$ vérifie $\begin{cases} \forall \theta, \Phi_\theta^T(T) \stackrel{\mathcal{L}}{=} U \text{ de loi } v \text{ fixe diffuse et connue} ; \\ \forall t, \Phi_t^T(\Theta) = \Phi^T(\tau(\Delta^n) \times \Theta) \end{cases}$;
- $K : \Phi^T(\tau(\Delta^n) \times \Theta) \subseteq \mathbb{R}^d \rightarrow \mathbb{R}$ a une forme analytique simple, si possible ne dépendant pas de n , telle que $\forall z \in \mathbb{R}, v(\{K = z\}) = 0$;
- $H : (K \circ \Phi^T)(\tau(\Delta^n) \times \Theta) \subseteq \mathbb{R} \rightarrow \mathbb{R}_+$ est décroissante au sens strict.

On a ainsi ϕ fonction pivotale correcte de loi $\eta = (H \circ K)(v)$, et pour tout $t \in \tau(\Delta^n)$

$$\forall z, \boxed{\{\phi_t^T > H(z)\} = \{K \circ \Phi_t^T < z\}} \text{ de niveau } \boxed{\gamma = P(K(U) < z) = v(\{K < z\})}.$$

À noter que l'unicité de ϕ^T n'implique absolument pas celle de Φ^T , K ou H .

4.A Estimation du couple (espérance, écart-type)

On note m l'espérance et σ l'écart-type, à estimer ensemble ($\theta = (m, \sigma)$ dans $\Theta = \mathbb{R} \times \mathbb{R}_+^*$), et on considère le modèle (exponentiel) usuel

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

$$\text{soit } f_{(m,\sigma)}(x_1, \dots, x_n) = \left(\frac{1}{\sigma \sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2} = (2\pi)^{-\frac{n}{2}} e^{-\frac{n}{2} \left(\ln \sigma^2 + \frac{(\bar{x} - m)^2 + s^2}{\sigma^2}\right)}$$

en notant $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ et $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$; on voit que le couple $T = (\bar{X}, S)$

est une statistique exhaustive du modèle (*) et on sait que \bar{X} et S sont des variables indépendantes telles que $\forall \theta = (m, \sigma) \in \Theta$,

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

où Z est une gaussienne centrée réduite indépendante de la variable du chi-deux χ_{n-1}^2 (la loi du membre de gauche étant prise par rapport à la probabilité $P_{(m,\sigma)}$).

$\forall (\bar{x}, s) \in \mathbb{R} \times \mathbb{R}_+^*$ fixé, $f_{(\bar{x},s)}^T(m, \sigma) \stackrel{\text{déf}}{=} e^{-\frac{n}{2} \left(\ln \sigma^2 + \frac{(\bar{x} - m)^2 + s^2}{\sigma^2}\right)}$ atteint son maximum pour $\ln \sigma^2 + \frac{(\bar{x} - m)^2 + s^2}{\sigma^2}$ minimum, soit pour $m = \bar{x}$ et $\frac{2}{\sigma} - \frac{2s^2}{\sigma^3} = 0 \iff \sigma = s$; de sorte que $T = (\bar{X}, S)$ est l'estimateur de $\theta = (m, \sigma)$ au maximum de vraisemblance.

On a donc $\|f_{(\bar{x},s)}^T\|_\infty = e^{-\frac{n}{2}(1 + \ln s^2)}$ et comme $\phi_{(\bar{x},s)}^T = \frac{f_{(\bar{x},s)}^T}{\|f_{(\bar{x},s)}^T\|_\infty}$, on a $\forall (m, \sigma) \in \Theta$,

$$-\frac{2}{n} \ln \phi_{(m,\sigma)}^T(\bar{X}, S) = \ln \sigma^2 + \frac{(\bar{X} - m)^2 + S^2}{\sigma^2} - \ln S^2 - 1 = K\left(\frac{\bar{X} - m}{\sigma}, \frac{S^2}{\sigma^2}\right)$$

pour $K(u_1, u_2) \stackrel{\text{d\u00e9f}}{=} u_1^2 + u_2 - \ln u_2 - 1 = u_1^2 + \bar{K}(u_2)$ (cf. le paragraphe 3.D pour \bar{K}).

On a ainsi $\phi^T = H \circ K \circ \Phi^T$ avec $H : z \in \mathbb{R}_+ \mapsto e^{-\frac{n}{2}z}$, de sorte que ϕ^T **v\u00e9rifie la propri\u00e9t\u00e9 pivotale** puisque $\forall (m, \sigma) \in \Theta$, $\phi_{(m,\sigma)}^T(\bar{X}, S) \stackrel{\mathcal{L}}{=} e^{-\frac{n}{2}K(U)}$; et comme $\phi_{(\bar{x},s)}^T(m, \sigma) = e^{-\frac{n}{2}K\left(\frac{\bar{x}-m}{\sigma}, \frac{s^2}{\sigma^2}\right)}$, $\phi_{(\bar{x},s)}^T$ a pour image $]0, 1]$ qui ne d\u00e9pend pas de (\bar{x}, s) .

D'apr\u00e8s le d\u00e9but du paragraphe 4, pour tout $(\bar{x}, s) \in \mathbb{R} \times \mathbb{R}_+^*$ et pour tout $\gamma \in]0, 1[$ fix\u00e9,

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

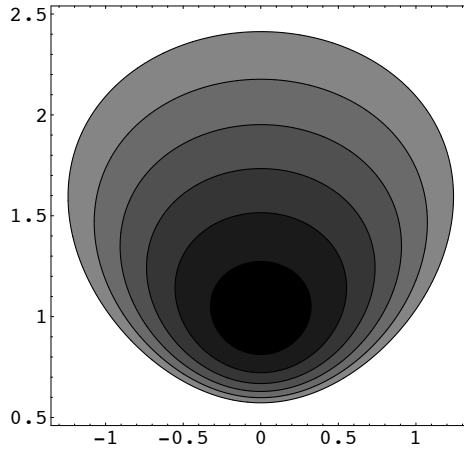
est la r\u00e9gion vraisemblable pour θ au niveau de confiance γ pour $z > 0$ tel que

$$\gamma = P(K(U) < z) = v(\{(u_1, u_2) \in \mathbb{R} \times \mathbb{R}_+^* : u_1^2 + \bar{K}(u_2) < z\});$$

sachant que la loi diffuse v de $U = \left(\frac{Z}{\sqrt{n}}, \frac{\chi_{n-1}^2}{n}\right)$ a pour densit\u00e9

$$(u_1, u_2) \in \mathbb{R}^2 \mapsto \frac{n^{\frac{n}{2}}}{2\sqrt{2\pi}\Gamma(\frac{n-1}{2})} e^{-\frac{n u_1^2}{2}} e^{-\frac{n u_2}{2}} \left(\frac{u_2}{2}\right)^{\frac{n-3}{2}} \mathbb{1}_{\mathbb{R}_+^*}(u_2).$$

$\bar{x} = 0, s = 1$; $\{\phi_{(0,1)}^T > e^{-\frac{n}{2}z}\}$ pour diverses valeurs positives de z :

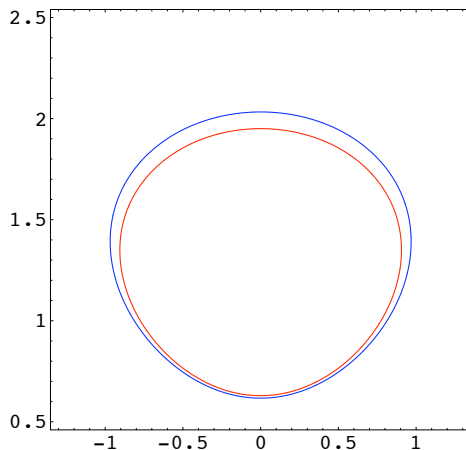


Application num\u00e9rique

Pour le niveau de confiance $\gamma = 0.95$, $n = 10$, $\bar{x} = 0$ et $s = 1$ par exemple, on trouve $z \approx 0.661$.

Si on reprend cette \u00e9tude en consid\u00e9rant (hardiment) qu'on est d\u00e9j\u00e0 aussi en situation asymptotique (cf. le paragraphe 3.A), la r\u00e9gion vraisemblable $\{\phi_{(0,1)}^T > e^{-\frac{1}{2}z_{as}}\}$ a le niveau de confiance asymptotique γ pour $P(\chi_d^2 < z_{as}) = \gamma$; ce qui donne num\u00e9riquement ici (o\u00f9 $d = 2$) $z_{as} \approx 5.99$.

Dans le sch\u00e9ma ci-dessous figurent ces deux r\u00e9gions vraisemblables d\u00e9termin\u00e9es par m\u00e9thodes directe (la plus grande) et asymptotique (la plus petite). L'\u00e9cart est assez faible d\u00e9j\u00e0 pour $n = 10$:



4.B Estimation de l'espérance à écart-type connu ou inconnu

- On veut d'abord estimer l'espérance $m \in \Theta_1 = \mathbb{R}$ à écart-type σ de valeur **connue**.

Le modèle (*) conditionné à σ (cf. le paragraphe 1.B) est exponentiel réel, de statistique exhaustive l'estimateur efficace $T = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

On a $\Phi_m^{\bar{X}}(\bar{X}) \stackrel{\text{déf}}{=} \frac{\bar{X} - m}{\sigma} \sqrt{n} \stackrel{\text{L}}{=} Z$ et \bar{X} a pour densité $f_m^{\bar{X}}(\bar{x}) \stackrel{\text{déf}}{=} e^{-\frac{n}{2\sigma^2}(\bar{x}-m)^2}$; la **propriété pivotale est vérifiée par** $\phi^{\bar{X}} = f^{\bar{X}} = H \circ K \circ \Phi^{\bar{X}}$ où $H : z \in \mathbb{R}_+ \mapsto e^{-\frac{z^2}{2}}$ et $K(u) \stackrel{\text{déf}}{=} |u|$, et l'image de $\phi^{\bar{X}}$ est $]0, 1[$, qui ne dépend pas de \bar{x} .

D'après le début du paragraphe 4, pour tout $\bar{x} > 0$ et pour tout $\gamma \in]0, 1[$ fixé,

$$\left\{ \phi_{\bar{x}}^{\bar{X}} > H(z) \right\} = \left\{ m \in \Theta_1 : \left| \frac{\bar{x} - m}{\sigma} \sqrt{n} \right| < z \right\} = \left] \bar{x} - z \frac{\sigma}{\sqrt{n}}, \bar{x} + z \frac{\sigma}{\sqrt{n}} \right[,$$

qui est la section à σ d'une région vraisemblable pour (m, σ) , est la région vraisemblable pour m au niveau de confiance γ pour $z > 0$ tel que

$$P(|Z| < z) = \gamma.$$

On retrouve l'intervalle de confiance équilibré habituel au niveau γ , du fait que Z est de loi paire.

- On veut à présent estimer l'espérance $m \in \Theta_1 = \mathbb{R}$ à écart-type σ de valeur **inconnue**.

Comme expliqué au paragraphe 1.B, on doit considérer la fonction de vraisemblance marginale pour m en x , $m \mapsto \sup_{\sigma > 0} f_x(m, \sigma)$. On a en l'occurrence pour tout $m \in \Theta_1$,

du fait que la dérivée de $\sigma \mapsto \ln \sigma^2 + \frac{(\bar{x} - m)^2 + s^2}{\sigma^2}$ s'annule pour $\sigma^2 = (\bar{x} - m)^2 + s^2$:

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

Ainsi pour $T = (\bar{X}, S)$ statistique exhaustive du modèle (*), $\phi_{(\bar{x}, s)}^T \propto \sup_{\sigma > 0} f_x(\cdot, \sigma)$ de norme sup égale à 1 est $\phi_{(\bar{x}, s)}^T(m) \stackrel{\text{déf}}{=} \left(1 + \left(\frac{\bar{x} - m}{s}\right)^2\right)^{-\frac{n}{2}}$; $\phi_{(\bar{x}, s)}^T$ a pour image $]0, 1]$, qui ne dépend pas de (\bar{x}, s) .

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

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

où T_{n-1} désigne la variable de Student à $n - 1$ degrés de liberté ($n \geq 2$). Ainsi la fonction $\phi^T = H \circ K \circ \Phi^T$ où $K(u) = |u|$ et $H : z \in \mathbb{R}_+ \mapsto \left(1 + \frac{z^2}{n-1}\right)^{-\frac{n}{2}}$ **vérifie la propriété pivotale**.

D'après le début du paragraphe 4, pour tout $x \in \Delta^n$ et pour tout $\gamma \in]0, 1[$ fixé,

$$\underline{\left\{ \phi_{(\bar{x}, s)}^T > H(z) \right\} = \left\{ m \in \Theta_1 : \left| \frac{\bar{x} - m}{s} \sqrt{n-1} \right| < z \right\} = \left] \bar{x} - z \frac{s}{\sqrt{n-1}}, \bar{x} + z \frac{s}{\sqrt{n-1}} \right[,}$$

qui est la projection d'une région vraisemblable pour (m, σ) , est la région 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 de niveau γ , du fait que T_{n-1} est de loi paire.

4.C Estimation de l'écart-type à espérance connue ou inconnue

• On veut d'abord estimer l'écart-type $\sigma \in \Theta_2 = \mathbb{R}_+^*$ à espérance m de valeur connue.

Le modèle (*) conditionné à m (cf. le paragraphe 1.B) est exponentiel réel, de statistique exhaustive l'estimateur efficace $V = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2 = S^2 + (\bar{X} - m)^2$.

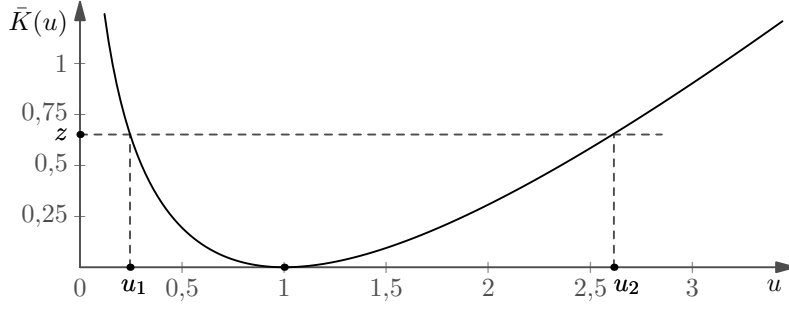
On a, par définition même de la loi du χ^2 , $\Phi_\sigma^V(V) \stackrel{\text{déf}}{=} \frac{V}{\sigma^2} \stackrel{\mathcal{L}}{=} \frac{\chi_n^2}{n}$; et V a pour densité $f_\sigma^V(v) \stackrel{\text{déf}}{=} e^{-\frac{n}{2}(\ln \sigma^2 + \frac{v}{\sigma^2})} = e^{-\frac{n}{2} \bar{K}(\frac{v}{\sigma^2})} e^{-\frac{n}{2}(1 - \ln v)}$, de sorte que ϕ_v^V de norme sup égale à 1 est $\phi_v^V(\sigma) = e^{-\frac{n}{2} \bar{K}(\frac{v}{\sigma^2})}$. **La propriété pivotale est vérifiée par** $\phi^V = H \circ \bar{K} \circ \Phi^V$ où $H : z \in \mathbb{R}_+ \mapsto e^{-\frac{n}{2}z}$, et l'image de ϕ_v^V est $]0, 1]$, qui ne dépend pas de v .

D'après le début du paragraphe 4, pour tout $v > 0$ et pour tout $\gamma \in]0, 1[$ fixé,

$$\underline{\left\{ \phi_v^V > H(z) \right\} = \left\{ \sigma \in \Theta_2 : \bar{K}\left(\frac{v}{\sigma^2}\right) < z \right\} = \left] \sqrt{\frac{v}{u_2}}, \sqrt{\frac{v}{u_1}} \right[,}$$

qui est la section à σ d'une région vraisemblable pour (m, σ) , est la région vraisemblable pour σ au niveau de confiance γ fixé pour $u_1, u_2 > 0$ définis par

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



On ne retrouve **pas** ici l'intervalle de confiance équilibré habituel de niveau γ :

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

• On veut à présent estimer l'écart-type $\sigma \in \Theta_2 = \mathbb{R}_+^*$ à espérance m de valeur **inconnue**.

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

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

Ainsi, en ne retenant que S dans la statistique exhaustive $T = (\bar{X}, S)$ du modèle (*), $\phi_s^S \propto \sup_{m \in \mathbb{R}} f_x(m, \cdot)$ de norme sup égale à 1 est $\phi_s^S(\sigma) \stackrel{\text{déf}}{=} e^{-\frac{n}{2} \bar{K}\left(\frac{s^2}{\sigma^2}\right)}$; ϕ_s^S a pour image $]0, 1]$, qui ne dépend pas de x .

Comme $\Phi_\sigma^S(S) \stackrel{\text{déf}}{=} \frac{S^2}{\sigma^2} \stackrel{\mathcal{L}}{\equiv} \frac{\chi_{n-1}^2}{n}$, la fonction $\phi^S = H \circ \bar{K} \circ \Phi^S$ où $H : z \in \mathbb{R}_+ \mapsto e^{-\frac{n}{2} z}$ vérifie de plus la propriété pivotale.

D'après le début du paragraphe 4, pour tout $x \in \Delta^n$ et pour tout $\gamma \in]0, 1[$ fixé,

$$\left\{ \phi_s^S > H(z) \right\} = \left\{ \sigma \in \Theta_2 : \bar{K}\left(\frac{s^2}{\sigma^2}\right) < z \right\} = \left] \frac{s}{\sqrt{u_2}}, \frac{s}{\sqrt{u_1}} \left[,$$

qui est la projection d'une région vraisemblable pour (m, σ) , est la région vraisemblable pour σ au niveau de confiance γ pour $u_1, u_2 > 0$ 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 ne retrouve **pas** ici l'intervalle de confiance équilibré habituel de niveau γ :

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

Remarque : la fonction \bar{K} semble intimement liée au modèles exponentiels vérifiant l'hypothèse de vraisemblance pivotale (cf. les paragraphes 3.D, 4.A et 4.C ci-dessus) !

5 Modèle exponentiel réel, cas diffus

5.A Une fonction pivotale intrinsèque

Soit un modèle exponentiel réel ($d = 1$) quelconque et soit $\varepsilon = -1$ ou $+1$ selon que le signe constant de α' est négatif ou positif.

L'estimateur efficace $T = \frac{1}{n} \sum_{i=1}^n \bar{a}(X_i)$ (unique à une transformation affine près) 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.

Limitons-nous ici au cas où la loi $f_\theta^T \mu$ de T est **diffuse**, autrement dit, puisque la fonction $f_\theta^T(t) = e^{n[\alpha(\theta)t+\beta(\theta)]}$ est strictement positive, au cas où la mesure μ est diffuse (ce qui impose que la mesure μ le soit) ; et considérons l'application F définie sur $\mathbb{R} \times \Theta$ par

$$F(t, \theta) \stackrel{\text{d\u00e9f}}{=} \begin{cases} P_\theta(T \leq t) = P_\theta(T < t) & \text{si } \varepsilon = -1 \\ P_\theta(T > t) = P_\theta(T \geq t) & \text{si } \varepsilon = +1 \end{cases}.$$

Comme la loi de T est diffuse, F_θ est continue en tant que fonction de répartition (complémentaire si $\varepsilon = +1$) de T ; 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 par construction.

On sait par ailleurs que pour tout $t \in \mathbb{R}$, l'application positive F_t est continue et monotone croissante. La monotonie s'établit classiquement en montrant que pour tous $\theta_1 < \theta_2$ dans Θ , on a (pour $\varepsilon = \pm 1$)

$$\forall t \in \mathbb{R}, \begin{cases} F_{\theta_2}(t) \geq \lambda_t F_{\theta_1}(t) \\ 1 - F_{\theta_2}(t) \leq \lambda_t (1 - F_{\theta_1}(t)) \end{cases} \quad \text{o\u00f9 } \lambda_t = \frac{e^{n[\alpha(\theta_2)t+\beta(\theta_2)]}}{e^{n[\alpha(\theta_1)t+\beta(\theta_1)]}} = \frac{f_{\theta_2}^T(t)}{f_{\theta_1}^T(t)} ;$$

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 I le plus petit intervalle ouvert de \mathbb{R} dont l'adhérence contient le support de μ ; comme $f_\theta^T > 0$ sur \mathbb{R} , on a $I = \{0 < F_\theta < 1\}$ pour tout $\theta \in \Theta$; et comme μ est diffuse, I lui-même est un sous-ensemble porteur de μ .

Montrons que la croissance de F_t est **stricte** pour tout $t \in I$: supposons par l'absurde F_t 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 $t \in I$ exclut qu'on ait $F_{\theta_1}(t) = F_{\theta_2}(t) = 0$ ou $F_{\theta_1}(t) = F_{\theta_2}(t) = 1$. f_t^T serait donc aussi constante sur l'intervalle en question, ce qui est incompatible avec f_t^T strictement monotone ou d'opposée unimodale sur Θ (cf. le paragraphe 2.D).

Il suffit alors de rajouter l'hypothèse

$$\forall t \in I, \underline{\inf F_t(\Theta)} = 0 \text{ et } \underline{\sup F_t(\Theta)} = 1,$$

autrement dit d'après ce qu'on vient de voir l'hypothèse $\forall t \in I, F_t(\Theta) =]0, 1[$, pour que F soit une fonction pivotale correcte.

Si $\phi^T = \frac{f^T}{M^T}$ est bien définie sur $I \times \Theta$ (c'est le cas par exemple si l'intervalle ouvert $h(\Theta)$ porte $\dot{\mu}$ d'après le paragraphe 2.D) et s'il existe une application $G :]0, 1[\rightarrow]0, 1[$ telle que $\phi^T = G \circ F$ sur $I \times \Theta$, le modèle exponentiel vérifie donc l'hypothèse de vraisemblance pivotale correcte.

Inversement supposons que ϕ^T admette une factorisation telle qu'indiquée au début du paragraphe 4.A, avec ici $d = 1$

$$\forall t \in I, \phi_t^T = H \circ K \circ \Phi_t^T,$$

où $\Phi_\theta^T(T) \stackrel{\mathcal{L}}{=} U$ est de loi ν fixe diffuse et connue et où la loi fixe $\eta = (H \circ K)(\nu)$ de $\phi_\theta^T(T)$ est elle-même diffuse ; de sorte que le modèle exponentiel est à vraisemblance pivotale correcte. Si (quitte à remplacer Φ^T par son opposée), l'application partielle Φ_θ^T est croissante au sens large pour tout $\theta \in \Theta$ comme aux paragraphes 3.D, 4.B et 4.C, on a

$$\forall t \in I, \{\Phi_\theta^T(T) < \Phi_\theta^T(t)\} \subseteq \{T < t\} \subseteq \{T \leq t\} \subseteq \{\Phi_\theta^T(T) \leq \Phi_\theta^T(t)\};$$

de sorte que, la loi de U étant diffuse,

$$P_\theta(T \leq t) = P_\theta(\Phi_\theta^T(T) \leq \Phi_\theta^T(t)) = P(U \leq \Phi_\theta^T(t)).$$

Autrement dit $F = \ell \circ \Phi^T$ où ℓ est la restriction à l'image de Φ^T de la fonction de répartition de U si $\varepsilon = -1$, de sa complémentaire si $\varepsilon = +1$. Comme F_t est strictement croissante et donc injective pour tout $t \in I$, Φ_t^T et ℓ sont aussi injectives ; on a donc $\phi^T = G \circ F$ sur $I \times \Theta$ avec $G \stackrel{\text{déf}}{=} H \circ K \circ \ell^{-1} :]0, 1[\rightarrow]0, 1[$.

Nous qualifierons cette situation (modèle exponentiel réel à loi $\dot{\mu}$ diffuse tel que $\forall t \in I, F_t(\Theta) =]0, 1[$ et $\phi_t^T \circ F_t^{-1}$ ne dépend pas de t , donc à vraisemblance pivotale correcte) de *modèle exponentiel réel parfait*.

Comme on l'a constaté dans les exemples précédents, il ne sert à rien de déterminer F pour les modèles exponentiels réels lorsque ceux-ci sont à vraisemblance pivotale.

5.B Estimation de substitution

Dans le cas d'un modèle exponentiel réel qui n'est pas à vraisemblance pivotale, mais qui vérifie toujours T de loi diffuse, $\forall t \in I, F_t(\Theta) =]0, 1[$ et $\phi^T = \frac{f^T}{M^T}$ bien définie sur $I \times \Theta$, on considère pour tout $t \in I$ la fonction

$$G_t \stackrel{\text{déf}}{=} \phi_t^T \circ F_t^{-1} :]0, 1[\rightarrow]0, 1[\text{ telle que } \phi_t^T = G_t \circ F_t;$$

G_t dépend de t , et elle est continue ainsi que strictement monotone ou d'opposée unimodale comme l'est ϕ_t^T , du fait de la croissance stricte et de la continuité de F_t .

Comme remarqué au paragraphe 2.B, les régions vraisemblables pour θ en t s'expriment toujours au moyen de la fonction pivotale (ici F), mais les niveaux de confiance qu'on pourrait continuer formellement à leur attribuer ne semblent plus justifiés. Cependant F permet de faire correctement des estimations par intervalles

de confiance intrinsèques, et on peut chercher à les *harmoniser au mieux* avec la vraisemblance.

On a d'une part $\forall t \in I$ et $\forall y \in [0, 1[$, $\{\phi_t^T > y\} = \{F_t \in B_t\}$ où $B_t = \{G_t > y\}$ est un intervalle ouvert de $]0, 1[$ vu les propriétés de G_t notées ci-dessus ; d'autre part lorsque le borélien $B \in \mathcal{B}(]0, 1[)$ ne dépend pas de t ,

$$\forall \theta \in \Theta, P_\theta(F_T^{-1}(B) \ni \theta) = P_\theta(F_T(\theta) \in B) = \lambda(B),$$

puisque la loi fixe de $F_\theta(T)$ est la probabilité uniforme sur $]0, 1[$.

Fixons un niveau de confiance de référence $\tilde{\gamma} \in]0.5, 1[$, et pour tout $t \in I$ soit $y_t > 0$ tel que la longueur de l'intervalle ouvert $B_t = \{G_t > y_t\} \subseteq]0, 1[$ soit égale à $\tilde{\gamma}$; alors l'intérieur $\overset{\circ}{B}$ de l'intervalle

$$B = \bigcap_{t \in I} B_t$$

est un intervalle ouvert de $]0, 1[$ de longueur $\gamma \in [2\tilde{\gamma} - 1, \tilde{\gamma}]$: $\gamma = 2\tilde{\gamma} - 1 > 0$ dans le cas de chevauchement minimal des B_t où $\overset{\circ}{B} =]1 - \tilde{\gamma}, \tilde{\gamma}[$, et $\gamma = \tilde{\gamma}$ par exemple dans le cas *parfait* où les B_t ne dépendent pas de t . Supposons que $\bar{B} \subset]0, 1[$ pour simplifier un peu l'analyse. Une valeur particulière $t \in I$ étant fournie par l'échantillon, l'ensemble

$$\{F_t \in \overset{\circ}{B}\} \stackrel{\text{déf}}{=}]\theta_t^1, \theta_t^2[, \text{ où } \theta_t^1, \theta_t^2 \in \Theta,$$

est un intervalle de confiance pour θ au niveau γ qui a la propriété intéressante suivante : comme $B_t \supseteq \overset{\circ}{B}$, on a $\{\phi_t^T > y_t\} = \{F_t \in B_t\} \supseteq]\theta_t^1, \theta_t^2[$, donc $\phi_t^T(\theta_t^1) \geq y_t$ et $\phi_t^T(\theta_t^2) \geq y_t$ par continuité de ϕ_t^T ; en posant $y_t' \stackrel{\text{déf}}{=} \min(\phi_t^T(\theta_t^1), \phi_t^T(\theta_t^2)) \geq y_t$, on a donc, du fait que ϕ_t^T est strictement monotone ou d'opposée unimodale,

$$]\theta_t^1, \theta_t^2[\subseteq]\theta_t'^1, \theta_t'^2[\stackrel{\text{déf}}{=} \{\phi_t^T > y_t'\} \subseteq \{\phi_t^T > y_t\}$$

avec $\theta_t'^1 = \theta_t^1$ ou $\theta_t'^2 = \theta_t^2$, $\theta_t'^1 = \inf \Theta$ si ϕ_t^T est décroissante sur Θ et $\theta_t'^2 = \sup \Theta$ si ϕ_t^T est croissante sur Θ .

Ceci permet d'attribuer à la région vraisemblable pour θ en t $]\theta_t'^1, \theta_t'^2[$ un niveau de confiance au moins égal à γ (et non plus égal à $\tilde{\gamma}$, comme dans le cas parfait) ; et ceci bien que $]\theta_t'^1, \theta_t'^2[$ ne soit pas en général un intervalle de confiance pour θ . Tel est le prix (minimum) à payer pour continuer à faire de l'estimation par régions vraisemblables...

En pratique, on pourra se passer de déterminer $B = \bigcap_{t \in I} B_t$ lorsqu'on sait que B_t dépend effectivement de t , en le remplaçant par défaut par l'intervalle $]1 - \tilde{\gamma}, \tilde{\gamma}[$ ($[1 - \tilde{\gamma}, \tilde{\gamma}] \subset]0, 1[$) qui est a fortiori inclus dans les B_t . Les bornes de la région vraisemblable $]\theta_t'^1, \theta_t'^2[$ de niveau de confiance au moins égal à $\gamma \stackrel{\text{déf}}{=} 2\tilde{\gamma} - 1$ sont alors caractérisées de la manière suivante :

- On commence par déterminer θ_t^1 et θ_t^2 comme solutions dans Θ des équations

$$\boxed{F_t(\theta^1) = 1 - \tilde{\gamma} = \frac{1 - \gamma}{2}} \quad \text{et} \quad \boxed{F_t(\theta^2) = \tilde{\gamma} = \frac{1 + \gamma}{2}} ;$$

$] \theta_t^1, \theta_t^2 [$ est un intervalle de confiance pour θ au niveau γ .

- - Si $\phi_t^T(\theta_t^1) > \phi_t^T(\theta_t^2)$, on remplace θ^1 par $\theta_t'^1 = \inf\{\theta < \theta_t^1 : \phi_t^T(\theta) > \phi_t^T(\theta_t^2)\}$, et on pose $\theta_t'^2 = \theta_t^2$; de sorte que $\phi_t^T(\theta_t'^1) = \phi_t^T(\theta_t'^2)$ si $\theta_t'^1 > \inf \Theta$.
- Si $\phi_t^T(\theta_t^2) > \phi_t^T(\theta_t^1)$, on remplace θ_t^2 par $\theta_t'^2 = \sup\{\theta > \theta_t^2 : \phi_t^T(\theta) > \phi_t^T(\theta_t^1)\}$, et on pose $\theta_t'^1 = \theta_t^1$; de sorte que $\phi_t^T(\theta_t'^1) = \phi_t^T(\theta_t'^2)$ si $\theta_t'^2 < \sup \Theta$.
- Si $\phi_t^T(\theta_t^1) = \phi_t^T(\theta_t^2)$, on pose $\theta_t'^1 = \theta_t^1$ et $\theta_t'^2 = \theta_t^2$.

Remarques

- Pour $\tilde{\gamma} = 0.5$, on a $\theta_t^1 = \theta_t^2$, déterminés par l'équation $F_t(\theta) = 0.5$; cette "estimation ponctuelle médiane" sous-jacente de θ diffère de son estimation au maximum de vraisemblance et est intéressante en soi.
- On peut voir dans l'analyse ci-dessus surtout une forme de justification de l'emploi d'un intervalle "à risque équilibré" dans l'estimation par intervalle $] \theta_t^1, \theta_t^2 [$ pour θ en t obtenue au moyen de la fonction pivotale correcte F , et se contenter d'ailleurs de celle-ci plutôt que de la dégrader en la région vraisemblable $] \theta_t'^1, \theta_t'^2 [$. En résumé, continuer à faire de l'estimation de θ par régions vraisemblables dans ce contexte est finalement assez artificiel...

5.C Exemple du modèle des lois Gamma

On considère à nouveau le modèle exponentiel des lois gamma de paramètre de forme θ (cf. le paragraphe 3.A)

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

dont l'estimateur efficace $T = \frac{1}{n} \sum_{i=1}^n \ln X_i$ a pour densité par rapport à μ

$$f_\theta^T(t) = \frac{1}{\Gamma(\theta)^n} e^{n\theta t} = e^{n[\theta t - \ln \Gamma(\theta)]}.$$

La simulation de $m \times n$ variables gamma indépendantes de paramètre θ permet de simuler m réalisations indépendantes de T ; on peut alors construire sa fonction de répartition empirique complémentaire \tilde{F}_θ ($\varepsilon = +1$), qui est une bonne approximation de F_θ si m est grand (quelques dizaines de milliers). m grand est d'autant plus nécessaire qu'on se focalise sur les queues des distributions : une valeur t de T étant fournie par l'échantillon, on détermine θ_t^1 et θ_t^2 en les ajustant pour que

$$\tilde{F}_{\theta_t^1}(t) = \frac{1-\gamma}{2} \text{ et } \tilde{F}_{\theta_t^2}(t) = \frac{1+\gamma}{2}.$$

Illustration pour $n = 100$, $\gamma = 0.95$ et $t = -1$:

$$] \theta_t'^1, \theta_t'^2 [\approx] \theta_t^1, \theta_t^2 [\approx] 0.663, 0.918 [.$$

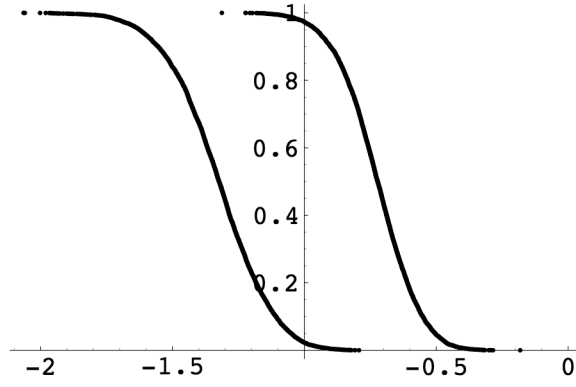
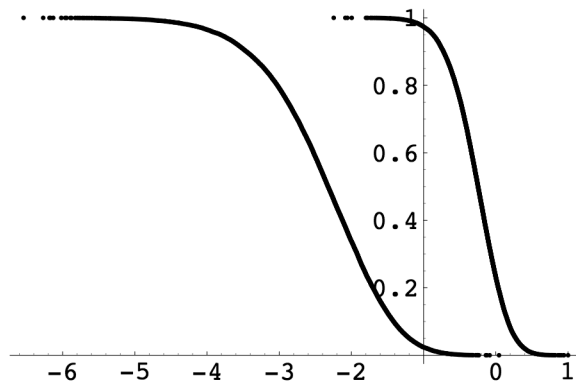


Illustration pour $n = 10$, $\gamma = 0.95$ et $t = -1$:

$$] \theta_t^1, \theta_t^2 [\approx] 0.425, 1.22 [\text{ et }] \theta_t'^1, \theta_t'^2 [\approx] 0.425, 1.26 [.$$



L'accord avec les valeurs trouvées au paragraphe 3.A est très bon, déjà pour $n = 10$.

À titre de comparaison, déterminons à présent des intervalles de confiance de niveau asymptotique γ de manière plus classique, en se référant au théorème central limite :

on sait (cf. le paragraphe 2.D) que l'estimateur efficace $T = \frac{1}{n} \sum_{i=1}^n \ln X_i$ a pour tout $\theta \in \Theta$ l'espérance $-\frac{\beta'(\theta)}{\alpha'(\theta)} = \frac{\Gamma'(\theta)}{\Gamma(\theta)} \stackrel{\text{déf}}{=} \psi(\theta)$ et de là la variance $\frac{1}{n} \frac{\psi'(\theta)}{\alpha'(\theta)} = \frac{1}{n} \psi'(\theta)$; on a donc

$$\sqrt{\frac{n}{\psi'(\theta)}} (T - \psi(\theta)) \stackrel{\mathcal{L}}{\approx} Z,$$

où Z est une variable gaussienne centrée réduite. On définit ainsi l'intervalle ouvert

$$] \theta_t^1, \theta_t^2 [= \left\{ \theta \in \Theta : |\psi(\theta) - t| < \frac{z}{\sqrt{n}} \sqrt{\psi'(\theta)} \right\} \text{ où } P(|Z| < z) = \gamma.$$

Pour $\gamma = 0.95$ ($z \approx 1.96$) et $t = -1$, on obtient $\begin{cases}] \theta_t^1, \theta_t^2 [\approx] 0.66, 0.91 [\text{ pour } n = 100 \\] \theta_t^1, \theta_t^2 [\approx] 0.37, 1.19 [\text{ pour } n = 10 \end{cases}$

l'accord est nettement moins bon pour $n = 10$.

6 Modèle exponentiel réel, cas discret

On considère un modèle exponentiel réel pour lequel μ et donc aussi $\dot{\mu} = \tau(B\mu^{n\otimes})$ (où $B(x) = \prod_{i=1}^n \bar{b}(x_i)$) sont des mesures discrètes ; de sorte que la loi de la statistique exhaustive $T = \frac{1}{n} \sum_{i=1}^n \bar{a}(X_i)$ est **discrète** pour tout $\theta \in \Theta$.

Supposons pour fixer les idées que $\varepsilon = +1$, autrement dit que le signe constant de α' est positif sur Θ ; et notons pour tout $\theta \in \Theta$

$$t \in \mathbb{R}, \underline{F_\theta^d(t) \stackrel{\text{d\u00e9f}}{=} P_\theta(T > t), \text{ inf\u00e9rieur ou \u00e9gal \u00e0 } F_\theta^g(t) \stackrel{\text{d\u00e9f}}{=} P_\theta(T \geq t),}$$

les versions continues \u00e0 droite et \u00e0 gauche de la fonction de r\u00e9partition compl\u00e9mentaire de T , qui diff\u00e8rent de $f_\theta^T(t) \dot{\mu}(\{t\})$ en tout point de $\dot{\Delta} \stackrel{\text{d\u00e9f}}{=} \{t \in \mathbb{R} : \dot{\mu}(\{t\}) > 0\}$. On a

$$\dot{\Delta} = \tau(\Delta^n) \text{ o\u00f9 } \Delta \stackrel{\text{d\u00e9f}}{=} \{x \in \mathbb{R} : \mu(\{x\}) > 0\};$$

Δ et $\dot{\Delta}$ sont les ensembles porteurs naturels de μ et $\dot{\mu}$ discr\u00e8tes, de cardinaux au plus d\u00e9nombrable.

6.A Une nouvelle fonction pivotale intrins\u00e8que

\u00c0 pr\u00e9sent les lois de $F_\theta^d(T)$ et de $F_\theta^g(T)$ ne sont plus que des approximations — elles-m\u00eames non diffuses — de la loi uniforme sur $]0, 1[$, qui en plus d\u00e9pendent de θ ! Les consid\u00e9rations des paragraphes 5.A et 5.B ne sont donc plus valides en l'\u00e9tat, et on cherche \u00e0 les \u00e9tendre.

Commen\u00e7ons par \u00e9tablir :

Lemme : Soit T une variable r\u00e9elle de fonctions de r\u00e9partition compl\u00e9mentaires continue \u00e0 droite F^d et continue \u00e0 gauche $F^g \geq F^d$; et soit U une variable uniforme sur $]0, 1[$ ind\u00e9pendante de T .

La variable compos\u00e9e $\underline{\psi(T, U) \stackrel{\text{d\u00e9f}}{=} F^d(T) + U (F^g(T) - F^d(T))}$ v\u00e9rifie $\underline{\psi(T, U) \stackrel{\mathcal{L}}{=} U}$.

Preuve : La fonction ψ d\u00e9finie sur $\mathbb{R} \times [0, 1]$ par $\psi(t, u) = F^d(t) + u (F^g(t) - F^d(t))$ v\u00e9rifie

- d'une part $\forall t \in \mathbb{R}$ et $\forall u \in]0, 1[$, $F^g(t) = \psi(t, 1) \geq \psi(t, u) \geq F^d(t) = \psi(t, 0)$, avec $F^g(t) > \psi(t, u) > F^d(t)$ si $F^g(t) > F^d(t)$;
- d'autre part $\forall t' < t$ dans \mathbb{R} , $\psi(t', 0) \geq \psi(t, 1)$ car $F^d(t') \geq F^g(t)$.

Montrons que la fonction de r\u00e9partition continue \u00e0 gauche de $\psi(T, U)$ vaut y en tout point $y \in]0, 1[$; ce qui suffira pour conclure que $\psi(T, U) \stackrel{\mathcal{L}}{=} U$:

- Supposons d'abord que $y \in F^d(\mathbb{R})$, autrement dit $y = F^d(t) \in]0, 1[$ pour un $t \in \mathbb{R}$; notons $\bar{t} = \sup\{s \geq t : F^d(s) = F^d(t)\}$ et distinguons trois cas :

- Soit $\bar{t} = t$; auquel cas $\{\psi(T, U) < F^d(t)\} = \{T > t\}$,
donc $P(\psi(T, U) < y) = F^d(t) = y$.
- Soit $\bar{t} > t$ et $F^d(t) = F^g(\bar{t}) = F^d(\bar{t})$ (pas de saut en \bar{t}); auquel cas
 $\{\psi(T, U) < F^d(t)\} = \{T > \bar{t}\}$,
donc $P(\psi(T, U) < y) = F^d(\bar{t}) = F^d(t) = y$.
- Soit $\bar{t} > t$ et $F^d(t) = F^g(\bar{t}) > F^d(\bar{t})$ (saut en \bar{t}); auquel cas, vu que $U < 1$,
 $\{\psi(T, U) < F^d(t)\} = \{F^d(T) \leq F^d(\bar{t})\} = \{T \geq \bar{t}\}$
donc $P(\psi(T, U) < y) = F^g(\bar{t}) = F^d(t) = y$.

• Supposons à présent que $y \notin F^d(\mathbb{R})$, de sorte que $y \in]F^d(t), F^g(t)]$ pour un $t \in \mathbb{R}$ où il y a saut (car $y \in]0, 1[$); ainsi $y = F^d(t) + u(F^g(t) - F^d(t))$ avec $0 < u \leq 1$.

On a $\{\psi(T, U) < y\} = \{T > t\} \uplus (\{T = t\} \cap \{U < u\})$,
donc $P(\psi(T, U) < y) = F^d(t) + u(F^g(t) - F^d(t)) = y$. \square

Considérons l'application ψ définie pour $(t, u) \in \mathbb{R} \times [0, 1]$ et $\theta \in \Theta$ par

$$\psi_\theta(t, u) = \underline{F_\theta^d(t) + u P_\theta(T = t)} \in [0, 1],$$

de sorte que $\psi_\theta(T, U) \stackrel{\mathcal{L}}{=} U$ d'après le lemme : cette fonction ψ à une variable supplémentaire est une **fonction pivotale**, intrinsèque par construction.

$\forall t \in \dot{\Delta}$, $F_t^g = F_t^d + \dot{\mu}(\{t\}) f_t^T$, donc f_t^T est bornée (par $\frac{1}{\dot{\mu}(\{t\})}$); on peut donc définir $\phi_t^T \propto f_t^T$ pour $t \in \dot{\Delta}$, et comme $\|\phi_t^T\|_\infty = 1$ on a $\forall t \in \dot{\Delta}$, $\phi_t^T = \frac{F_t^g - F_t^d}{\|F_t^g - F_t^d\|_\infty}$.

Pour tout $t \in \mathbb{R}$, les applications F_t^d et F_t^g sont continues et croissantes au sens large sur Θ , et elles sont même **strictement croissantes** pour tout $t \in I$, où I est le plus petit intervalle ouvert de \mathbb{R} dont l'adhérence contient $\dot{\Delta}$, qui vérifie pour tout $\theta \in \Theta$ $I = \{0 < F_\theta^d \leq F_\theta^g < 1\}$ (même preuve que dans le cas exponentiel diffus, cf. le paragraphe 5.A). En conséquence l'application

$$\theta \in \Theta \xrightarrow{\psi_t^u} \psi_\theta(t, u) = \underline{(1 - u) F_\theta^d(t) + u F_\theta^g(t)}$$

est elle aussi **continue et croissante** au sens large sur Θ pour tout $(t, u) \in \mathbb{R} \times [0, 1]$ fixé, avec croissance **au sens strict** lorsque $t \in I$.

À présent I n'est plus forcément porteur de $\dot{\mu}$, car $I \not\supseteq \dot{\Delta}$ dès que $\dot{\Delta}$ a soit un minimum $t_{\min} \in \mathbb{R}$, soit un maximum $t_{\max} \in \mathbb{R}$, soit les deux.

Dans le premier cas, on a $P_\theta(T \geq t_{\min}) = 1$ pour tout $\theta \in \Theta$, c'est-à-dire $F_{t_{\min}}^g = 1$ sur Θ . Comme $\forall t \in \dot{\Delta}$, $F_t^g = F_t^d + \dot{\mu}(\{t\}) f_t^T$, on a donc $F_{t_{\min}}^d = 1 - \dot{\mu}(\{t_{\min}\}) f_{t_{\min}}^T$ avec $\dot{\mu}(\{t_{\min}\}) > 0$. Et comme $F_{t_{\min}}^d$ est croissante au sens large, $f_{t_{\min}}^T$ n'est pas d'opposée unimodale sur Θ ni strictement croissante. Par élimination $f_{t_{\min}}^T$ est donc **strictement décroissante** (cf. le paragraphe 2.D) ainsi que $\phi_{t_{\min}}^T$; ce qui permet

de conclure que $F_{t_{\min}}^d$ et donc aussi $\psi_{t_{\min}}^u$ pour tout $u \in [0, 1[$ sont **strictement croissantes** sur Θ .

De manière analogue dans le second cas, on a $P_\theta(T > t_{\max}) = 0$ pour tout $\theta \in \Theta$, c'est-à-dire $F_{t_{\max}}^d = 0$ sur Θ . On a donc $F_{t_{\max}}^g = \dot{\mu}(\{t_{\max}\}) f_{t_{\max}}^T$, et on en déduit que $f_{t_{\max}}^T$, $\phi_{t_{\max}}^T$, $F_{t_{\max}}^g$ et donc aussi $\psi_{t_{\max}}^u$ pour tout $u \in]0, 1]$ sont **strictement croissantes** sur Θ .

On fait les hypothèses complémentaires suivantes pour que ψ soit une fonction pivotale **correcte** :

- $\forall t \in I, \underline{\inf F_t^g(\Theta) = 0}$ et $\underline{\sup F_t^d(\Theta) = 1}$,

d'où $\inf F_t^d(\Theta) = 0$ et $\sup F_t^g(\Theta) = 1$; ainsi $\forall t \in I, F_t^d(\Theta) = F_t^g(\Theta) =]0, 1[$.

Comme $\forall t \in \dot{\Delta}, f_t^T = \frac{1}{\dot{\mu}(\{t\})} (F_t^g - F_t^d)$, f_t^T a donc des limites nulles en les extrémités gauche et droite de Θ dans $\bar{\mathbb{R}}$ lorsque de plus $t \in I$; ce qu'on écrit

$$\forall t \in I \cap \dot{\Delta}, \underline{\lim}_g f_t^T = \underline{\lim}_d f_t^T = 0.$$

On en déduit que f_t^T et donc aussi ϕ_t^T sont nécessairement d'opposée **unimodale** pour $t \in I \cap \dot{\Delta}$ (de sorte que $I \cap \dot{\Delta} \subseteq h(\Theta)$).

Et comme $\psi_t^u = (1 - u) F_t^d + u F_t^g$, on a $\forall (t, u) \in I \times [0, 1], \underline{\psi_t^u(\Theta) =]0, 1[}$.

- Si $\dot{\Delta}$ a un minimum $t_{\min} \in \mathbb{R}$, on sait que $F_{t_{\min}}^g = 1$ sur Θ ; on suppose

$$\underline{\inf F_{t_{\min}}^d(\Theta) = 0}$$
 et $\underline{\sup F_{t_{\min}}^d(\Theta) = 1}$, autrement dit $F_{t_{\min}}^d(\Theta) =]0, 1[$.

Comme $f_{t_{\min}}^T = \frac{1}{\dot{\mu}(\{t_{\min}\})} (1 - F_{t_{\min}}^d)$, on a $\underline{\lim}_d f_{t_{\min}}^T = 0$ et $\lim_d \phi_{t_{\min}}^T = 0$.

Et comme $\psi_{t_{\min}}^u = (1 - u) F_{t_{\min}}^d + u$, on a $\forall u \in [0, 1[, \psi_{t_{\min}}^u(\Theta) =]u, 1[$.

- Si $\dot{\Delta}$ a un maximum $t_{\max} \in \mathbb{R}$, on sait que $F_{t_{\max}}^d = 0$ sur Θ ; on suppose

$$\underline{\inf F_{t_{\max}}^g(\Theta) = 0}$$
 et $\underline{\sup F_{t_{\max}}^g(\Theta) = 1}$, autrement dit $F_{t_{\max}}^g(\Theta) =]0, 1[$.

Comme $f_{t_{\max}}^T = \frac{1}{\dot{\mu}(\{t_{\max}\})} F_{t_{\max}}^g$, on a $\underline{\lim}_g f_{t_{\max}}^T = 0$ et $\lim_g \phi_{t_{\max}}^T = 0$.

Et comme $\psi_{t_{\max}}^u = u F_{t_{\max}}^g$, on a $\forall u \in]0, 1], \psi_{t_{\max}}^u(\Theta) =]0, u[$.

Exemple du modèle exponentiel des lois de Poisson de paramètre $\theta \in]0, +\infty[$:

$$\Delta = \mathbb{N}, \mu = \sum_{k=0}^{\infty} \delta_k, f_\theta(x) = e^{-\theta} \frac{\theta^x}{x!} = \frac{1}{x!} e^{x \ln \theta - \theta}$$

où $\alpha(\theta) = \ln \theta$ a sa dérivée $\alpha'(\theta) = \frac{1}{\theta}$ de signe positif, donc $\varepsilon = +1$.

La statistique exhaustive $T = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$ est l'estimateur efficace de $-\frac{\beta'(\theta)}{\alpha'(\theta)} = \theta$,

et on a

$$\forall t = \bar{x} \geq 0 \text{ et } \forall \theta > 0, f_t^T(\theta) = \theta^{nt} e^{-n\theta};$$

pour $t > 0$, $\hat{\theta} = t$ est l'estimation de θ au maximum de vraisemblance.

On a $\hat{\Delta} = \tau(\Delta^n) = \frac{1}{n} \mathbb{N}$; f_t^T est bien strictement décroissante de limite nulle en $+\infty$ pour $t_{\min} = 0$ et d'opposée unimodale de limite nulle en 0 et $+\infty$ pour $t > 0$.

Comme nT est une variable de Poisson de paramètre $n\theta$, on a

$$\forall j \in \mathbb{N}, P_\theta\left(T = \frac{j}{n}\right) = e^{-n\theta} \frac{(n\theta)^j}{j!};$$

$$\text{de sorte que } F_t^d(\theta) = 1 - P_\theta(T \leq t) = 1 - \sum_{j=0}^{\lfloor nt \rfloor} P_\theta\left(T = \frac{j}{n}\right) = 1 - e^{-n\theta} \sum_{j=0}^{\lfloor nt \rfloor} \frac{(n\theta)^j}{j!}$$

$$\text{et pour } t > 0, F_t^g(\theta) = 1 - P_\theta(T < t) = 1 - \sum_{j=0}^{\lceil nt \rceil - 1} P_\theta\left(T = \frac{j}{n}\right) = 1 - e^{-n\theta} \sum_{j=0}^{\lceil nt \rceil - 1} \frac{(n\theta)^j}{j!}$$

F_t^d et F_t^g vérifient bien les hypothèses complémentaires :

- $\forall t \in I =]0, +\infty[, \quad \inf F_t^g(\Theta) = 0 \text{ et } \sup F_t^d(\Theta) = 1;$
- $F_0^d(\theta) = 1 - e^{-n\theta}$, donc $\inf F_0^d(\Theta) = 0 \text{ et } \sup F_0^d(\Theta) = 1.$

On étudiera en détail au paragraphe 6.C le **modèle pile ou face** plus intéressant, qui vérifie aussi toutes les hypothèses.

6.B Estimation dans le cas discret

Comme indiqué dans la dernière remarque du paragraphe 5.B, il est artificiel de greffer de l'estimation de θ par régions vraisemblables sur de l'estimation par intervalle de confiance puisqu'on procède par simple dégradation de cette dernière. Nous n'approfondirons donc pas cette possibilité dans le cas discret, et nous concentrerons plutôt sur l'utilisation de la nouvelle fonction pivotale intrinsèque ψ pour l'estimation de θ par intervalle de confiance.

On a pour tout $B \in \mathcal{B}(]0, 1[)$ ne dépendant pas de (t, u) ,

$$\forall \theta \in \Theta, P_\theta((\psi_T^U)^{-1}(B) \ni \theta) = P_\theta(\psi_\theta(T, U) \in B) = \lambda(B),$$

puisque la loi fixe de $\psi_\theta(T, U)$ est la probabilité uniforme sur $]0, 1[$ (paragraphe 6.A).

Fixons un niveau de confiance $\gamma \in]0, 1[$ et considérons l'intervalle "à risque équilibré" associé $B = \left] \frac{1-\gamma}{2}, \frac{1+\gamma}{2} \right[$. Pour tout $(t, u) \in \hat{\Delta} \times]0, 1[$, $(\psi_t^u)^{-1}(B) = \{\psi_t^u \in B\}$ est un intervalle ouvert $\left] \theta_{(t,u)}^1, \theta_{(t,u)}^2 \right[$, éventuellement vide si $t = t_{\min}$ ou $t = t_{\max}$, avec pour tout $u \in]0, 1[$

$$\left\{ \begin{array}{l} \forall t \in I \cap \hat{\Delta}, \psi_t^u(\theta_{(t,u)}^1) = \frac{1-\gamma}{2} \text{ et } \psi_t^u(\theta_{(t,u)}^2) = \frac{1+\gamma}{2} \\ \theta_{(t_{\min}, u)}^1 = \inf \Theta \text{ si } u > \frac{1-\gamma}{2} \text{ et } \theta_{(t_{\min}, u)}^2 = \inf \Theta \text{ si } u > \frac{1+\gamma}{2} \\ \theta_{(t_{\max}, u)}^1 = \sup \Theta \text{ si } u < \frac{1-\gamma}{2} \text{ et } \theta_{(t_{\max}, u)}^2 = \sup \Theta \text{ si } u < \frac{1+\gamma}{2} \end{array} \right.$$

Comme $(\psi_T^U)^{-1}(B) =]\theta_{(T,U)}^1, \theta_{(T,U)}^2[$, on a

$$\forall \theta \in \Theta, \gamma = P_\theta(] \theta_{(T,U)}^1, \theta_{(T,U)}^2[\ni \theta) = \int_{\mathbb{R} \times]0,1[} \mathbb{1}_{] \theta_{(t,u)}^1, \theta_{(t,u)}^2[}(\theta) f_\theta^T(t) d\mu(t) du ;$$

$$\text{soit } \sum_{t \in \dot{\Delta}} P_\theta(T = t) \tilde{\Theta}^t(\theta) = \gamma \quad \text{où } \tilde{\Theta}^t(\theta) \stackrel{\text{d\u00e9f}}{=} \int_0^1 \mathbb{1}_{] \theta_{(t,u)}^1, \theta_{(t,u)}^2[}(\theta) du.$$

$\forall t \in \dot{\Delta}$, l'application $\tilde{\Theta}^t$, \u00e0 valeurs dans $[0, 1]$ est un *intervalle flou* de *support* et *noyau*

$$\left\{ \begin{array}{l} \{\tilde{\Theta}^t > 0\} = \bigcup_{0 < u < 1}] \theta_{(t,u)}^1, \theta_{(t,u)}^2[=] \theta_{(t,1)}^1, \theta_{(t,0)}^2[\\ \{\tilde{\Theta}^t = 1\} = \bigcap_{0 < u < 1}] \theta_{(t,u)}^1, \theta_{(t,u)}^2[= [\theta_{(t,0)}^1, \theta_{(t,1)}^2] \cap \Theta \quad (= \emptyset \text{ si } \theta_{(t,0)}^1 > \theta_{(t,1)}^2) \end{array} \right\}$$

les notations $\theta_{(t,1)}^1 \stackrel{\text{d\u00e9f}}{=} \inf_{0 < u < 1} \theta_{(t,u)}^1$, $\theta_{(t,0)}^2 \stackrel{\text{d\u00e9f}}{=} \sup_{0 < u < 1} \theta_{(t,u)}^2$, $\theta_{(t,0)}^1 \stackrel{\text{d\u00e9f}}{=} \sup_{0 < u < 1} \theta_{(t,u)}^1$ et $\theta_{(t,1)}^2 \stackrel{\text{d\u00e9f}}{=} \inf_{0 < u < 1} \theta_{(t,u)}^2$ \u00e9tant ainsi choisies du fait de la croissance (stricte) de l'application partielle $u \mapsto \psi_\theta(t, u)$ pour tous $(t, \theta) \in \dot{\Delta} \times \Theta$ qui entraine la d\u00e9croissance des applications partielles $u \mapsto \theta_{(t,u)}^1$ et $u \mapsto \theta_{(t,u)}^2$ (d\u00e9croissance stricte pour $t \in I \cap \dot{\Delta}$). De sorte que $\forall t \in \dot{\Delta}$, $\theta_{(t,1)}^1$ et $\theta_{(t,0)}^2$, qu'on notera plus simplement dans la suite $\dot{\theta}_t^1$ et $\dot{\theta}_t^2$, sont les solutions des \u00e9quations

$$\left\{ \begin{array}{l} F_t^g(\theta^1) = \frac{1-\gamma}{2} \quad (t \neq t_{\min}) \quad \text{et} \quad F_t^d(\theta^2) = \frac{1+\gamma}{2} \quad (t \neq t_{\max}) \\ \text{avec } \dot{\theta}_{t_{\min}}^1 = \inf \Theta \quad \quad \quad \text{et} \quad \dot{\theta}_{t_{\max}}^2 = \sup \Theta \end{array} \right\}.$$

Et $\forall t \in \dot{\Delta}$, $\theta_{(t,0)}^1$ et $\theta_{(t,1)}^2$, qu'on notera plus simplement dans la suite $\ddot{\theta}_t^1$ et $\ddot{\theta}_t^2$ et qui v\u00e9rifient $\ddot{\theta}_t^1, \ddot{\theta}_t^2 \in] \dot{\theta}_t^1, \dot{\theta}_t^2[$, sont les solutions des \u00e9quations

$$\left\{ \begin{array}{l} F_t^d(\theta^1) = \frac{1-\gamma}{2} \quad (t \neq t_{\max}) \quad \text{et} \quad F_t^g(\theta^2) = \frac{1+\gamma}{2} \quad (t \neq t_{\min}) \\ \text{avec } \ddot{\theta}_{t_{\min}}^1 > \ddot{\theta}_{t_{\min}}^2 = \inf \Theta \quad \text{et} \quad \ddot{\theta}_{t_{\max}}^2 < \ddot{\theta}_{t_{\max}}^1 = \sup \Theta \end{array} \right\} ;$$

la condition $\{\tilde{\Theta}^t = 1\} \neq \emptyset$ n'\u00e9tant v\u00e9rifi\u00e9e que pour $t \in I \cap \dot{\Delta}$ sous r\u00e9serve que $F_t^g(\dot{\theta}_t^1) - F_t^d(\dot{\theta}_t^1) \equiv P_{\dot{\theta}_t^1}(T = t) \leq \gamma$ ou $F_t^g(\ddot{\theta}_t^2) - F_t^d(\ddot{\theta}_t^2) \equiv P_{\ddot{\theta}_t^2}(T = t) \leq \gamma$, qui assure $\dot{\theta}_t^1 \leq \ddot{\theta}_t^2$; auquel cas $\{\tilde{\Theta}^t = 1\} = [\dot{\theta}_t^1, \ddot{\theta}_t^2]$.

On note que la d\u00e9croissance stricte en t et la croissance en θ des applications

$$(t, \theta) \in \dot{\Delta} \times \Theta \mapsto F_t^g(\theta) \quad \text{et} \quad (t, \theta) \in \dot{\Delta} \times \Theta \mapsto F_t^d(\theta)$$

assure que les 4 applications $t \mapsto \dot{\theta}_t^1$, $t \mapsto \dot{\theta}_t^2$, $t \mapsto \ddot{\theta}_t^1$ et $t \mapsto \ddot{\theta}_t^2$ sont strictement croissantes sur $\dot{\Delta}$. Notons aussi que pour $t \in I \cap \dot{\Delta}$

- si $\dot{\theta}_t^1 < \ddot{\theta}_t^2$, on a

$$(F_t^d + F_t^g)(\dot{\theta}_t^1) < \frac{1+\gamma}{2} + \frac{1-\gamma}{2} = 1 \quad \text{et} \quad (F_t^d + F_t^g)(\ddot{\theta}_t^2) > \frac{1-\gamma}{2} + \frac{1+\gamma}{2} = 1 ;$$

- si $\ddot{\theta}_t^1 > \ddot{\theta}_t^2$, on a

$$(F_t^d + F_t^g)(\ddot{\theta}_t^1) > \frac{1+\gamma}{2} + \frac{1-\gamma}{2} = 1 \quad \text{et} \quad (F_t^d + F_t^g)(\ddot{\theta}_t^2) < \frac{1-\gamma}{2} + \frac{1+\gamma}{2} = 1.$$

L'application $F_t^d + F_t^g$ étant strictement croissante sur Θ , on en déduit qu'il existe une unique valeur θ_t^{med} de θ dite *estimation médiane de θ* telle que

$$(F_t^d + F_t^g)(\theta) = 1 \iff \underline{P_\theta(T < t)} = \overline{P_\theta(T > t)};$$

et que dans tous les cas $\theta_t^{\text{med}} \in [\ddot{\theta}_t^1, \ddot{\theta}_t^2] \subseteq]\dot{\theta}_t^1, \dot{\theta}_t^2[$ (avec $\theta_t^{\text{med}} = \ddot{\theta}_t^1$ si $\ddot{\theta}_t^1 = \ddot{\theta}_t^2$).

Ainsi le caractère flou de la situation résulte de la fonction ψ qui constitue une forme de *brouillage* de la fonction de répartition de T . Or on ne peut pas se contenter d'intervalles de confiance flous pour l'estimation de θ , on veut disposer de vrais intervalles.

Une première solution simpliste est de ne pas entrer dans ces considérations et de se référer directement à l'intervalle aléatoire de confiance $] \theta_{(T,U)}^1, \theta_{(T,U)}^2 [$ de niveau γ pour retenir l'intervalle $] \theta_{(t,u)}^1, \theta_{(t,u)}^2 [$. Mais il paraît inacceptable que pour une même réalisation $t \in \dot{\Delta}$ de T celui-ci varie en fonction d'une valeur aléatoire u indépendante du modèle étudié et qui donc ne porte aucune information le concernant ! On cherche à définir des intervalles d'estimation déterministes...

Une seconde solution plus intéressante est de remarquer que

$$(\forall (t, u) \in \dot{\Delta} \times]0, 1[,] \dot{\theta}_t^1, \dot{\theta}_t^2 [\supseteq] \theta_{(t,u)}^1, \theta_{(t,u)}^2 [) \implies] \dot{\theta}_T^1, \dot{\theta}_T^2 [\supseteq] \theta_{(T,U)}^1, \theta_{(T,U)}^2 [;$$

de sorte que $\forall \theta \in \Theta$, $P_\theta(] \dot{\theta}_T^1, \dot{\theta}_T^2 [\ni \theta) \geq \gamma$, ce qui exprime qu'on peut attribuer à l'intervalle aléatoire $] \dot{\theta}_T^1, \dot{\theta}_T^2 [$ un niveau de confiance **au moins égal** à γ ; on peut donc prendre comme intervalle d'estimation pour θ en tout $t \in \dot{\Delta}$ de niveau au moins égal à γ l'intervalle $] \dot{\theta}_t^1, \dot{\theta}_t^2 [$. L'inconvénient de cette approche est un manque certain de précision, les intervalles retenus étant inutilement surdimensionnés.

On cherche en conséquence à améliorer cette solution en réduisant autant qu'il est possible les intervalles $] \dot{\theta}_t^1, \dot{\theta}_t^2 [$ à niveau de confiance au moins égal à γ . Pour cela on se réfère à la formule élémentaire de calcul

$$\forall \theta \in \Theta, P_\theta(] \dot{\theta}_T^1, \dot{\theta}_T^2 [\ni \theta) = \sum_{t \in \dot{\Delta} : \dot{\theta}_t^1 < \theta < \dot{\theta}_t^2} P_\theta(T = t),$$

et on va s'employer à définir au mieux $\theta_t^1 < \theta_t^2$ dans $[\dot{\theta}_t^1, \dot{\theta}_t^2]$ pour tout $t \in \dot{\Delta}$ vérifiant encore

$$P_\theta(] \theta_T^1, \theta_T^2 [\ni \theta) = \sum_{t \in \dot{\Delta} : \theta_t^1 < \theta < \theta_t^2} P_\theta(T = t) \geq \gamma$$

pour tout $\theta \in \Theta$ (sauf éventuellement un nombre dénombrable d'entre eux).

Avant de commencer, on peut noter qu'on a aussi

$$\forall \theta \in \Theta, P_\theta(\dot{\theta}_T^1 \geq \theta) = \sum_{t \in \dot{\Delta} : \dot{\theta}_t^1 \geq \theta} P_\theta(T = t) \quad \text{et} \quad P_\theta(\dot{\theta}_T^2 \leq \theta) = \sum_{t \in \dot{\Delta} : \dot{\theta}_t^2 \leq \theta} P_\theta(T = t);$$

donc

$$\left\{ \begin{array}{l} \forall t \neq t_{\min}, P_{\dot{\theta}_t^1}(\dot{\theta}_T^1 \geq \dot{\theta}_t^1) = \sum_{s \geq t} P_{\dot{\theta}_t^1}(T = s) = P_{\dot{\theta}_t^1}(T \geq t) = F_t^g(\dot{\theta}_t^1) = \frac{1-\gamma}{2} \\ \forall t \neq t_{\max}, P_{\dot{\theta}_t^2}(\dot{\theta}_T^2 \leq \dot{\theta}_t^2) = \sum_{s \leq t} P_{\dot{\theta}_t^2}(T = s) = P_{\dot{\theta}_t^2}(T \leq t) = 1 - F_t^d(\dot{\theta}_t^2) = \frac{1-\gamma}{2} \end{array} \right. ,$$

et la relation $P_{\dot{\theta}_t^1}(\dot{\theta}_T^1 \geq \dot{\theta}_t^1) = P_{\dot{\theta}_t^2}(\dot{\theta}_T^2 \leq \dot{\theta}_t^2) \iff P_{\dot{\theta}_t^1}(T < t) = P_{\dot{\theta}_t^2}(T > t)$ peut être pensée comme une *condition d'équilibre* portant sur les bornes de l'intervalle $]\dot{\theta}_t^1, \dot{\theta}_t^2[$.

D'autre part la formule élémentaire précédente se précise alors en

$$\forall \theta \in \Theta, P_\theta(]\dot{\theta}_T^1, \dot{\theta}_T^2[\ni \theta) = \frac{\sum_{t \in \dot{\Delta}: t' \leq t \leq t''} P_\theta(T = t) = F_{t'}^g(\theta) - F_{t''}^d(\theta)}{\quad}$$

où $t' \leq t''$ sont caractérisés dans $\dot{\Delta}$ par les conditions $\ddot{\theta}_{t'}^2 \leq \theta < \dot{\theta}_{t'}^2$ et $\dot{\theta}_{t''}^1 < \theta \leq \ddot{\theta}_{t''}^1$:

en effet $\forall t < t'$ dans $\dot{\Delta}$, on a l'inclusion $\{T \geq t'\} \subseteq \{T > t\}$, donc

$$F_t^d(\dot{\theta}_t^2) = \frac{1+\gamma}{2} = F_{t'}^g(\ddot{\theta}_{t'}^2) \leq F_t^d(\ddot{\theta}_{t'}^2) \implies \dot{\theta}_t^2 \leq \ddot{\theta}_{t'}^2 ;$$

de même $\forall t > t''$ dans $\dot{\Delta}$, on a l'inclusion $\{T \leq t''\} \subseteq \{T < t\}$, donc

$$F_t^g(\dot{\theta}_t^1) = \frac{1-\gamma}{2} = F_{t''}^d(\ddot{\theta}_{t''}^1) \geq F_t^d(\ddot{\theta}_{t''}^1) \implies \dot{\theta}_t^1 \geq \ddot{\theta}_{t''}^1.$$

On suppose dorénavant que $\dot{\Delta}$ est fini ou au moins localement fini, de sorte que chaque $t \in \dot{\Delta}$ a (pour la relation d'ordre sur \mathbb{R}) un prédécesseur qu'on notera t_- (sauf si $t = t_{\min}$) et un successeur qu'on notera t_+ (sauf si $t = t_{\max}$). Comme $\{T \geq t\} = \{T > t_-\}$ et $\{T \geq t_+\} = \{T > t\}$, on a

$$F_t^d(\ddot{\theta}_t^1) = \frac{1-\gamma}{2} = F_{t_+}^g(\dot{\theta}_{t_+}^1) = F_t^d(\dot{\theta}_{t_+}^1) \text{ et } F_t^g(\dot{\theta}_t^2) = \frac{1+\gamma}{2} = F_{t_-}^d(\dot{\theta}_{t_-}^2) = F_t^g(\dot{\theta}_{t_-}^2) ;$$

de sorte que (si $t \neq t_{\max}$) $\dot{\theta}_t^1 = \dot{\theta}_{t_+}^1$ et (si $t \neq t_{\min}$) $\ddot{\theta}_t^2 = \ddot{\theta}_{t_-}^2$,

puisque $t \neq t_{\max}$, respectivement $t \neq t_{\min}$, assure que l'application F_t^d , respectivement F_t^g , est strictement croissante sur Θ .

* Considérons d'abord les couples $(t', t'') \in \dot{\Delta}^2$ tels que

$$t' < t'' \text{ et } \dot{\theta}_{t'_-}^2 = \ddot{\theta}_{t'}^2 \leq \dot{\theta}_{t''}^1 < \dot{\theta}_{t''}^2 \leq \ddot{\theta}_{t''}^1 = \dot{\theta}_{t''_+}^1$$

(les égalités extrêmes ci-dessus ne figurant pas si $t' = t_{\min}$ ou/et $t'' = t_{\max}$).

D'après la fomule précédente, on a pour tout $\theta \in]\dot{\theta}_{t''}^1, \dot{\theta}_{t'}^2[$, du fait que $\theta > \ddot{\theta}_{t'}^2$ et $\theta < \ddot{\theta}_{t''}^1$,

$$\forall \theta \in]\dot{\theta}_{t''}^1, \dot{\theta}_{t'}^2[, F_{t'}^g(\theta) - F_{t''}^d(\theta) = P_\theta(]\dot{\theta}_T^1, \dot{\theta}_T^2[\ni \theta) \geq \gamma.$$

D'autre part, la fonction $F_{t'}^d + F_{t''}^g$ est strictement croissante sur $[\dot{\theta}_{t''}^1, \dot{\theta}_{t'}^2]$, avec

$$(F_{t'}^d + F_{t''}^g)(\dot{\theta}_{t''}^1) < F_{t'}^d(\dot{\theta}_{t'}^2) + F_{t''}^g(\dot{\theta}_{t''}^1) = \frac{1+\gamma}{2} + \frac{1-\gamma}{2} = 1 < (F_{t'}^d + F_{t''}^g)(\dot{\theta}_{t'}^2).$$

Il existe donc une unique valeur intermédiaire $\theta^{\text{int}} \in]\dot{\theta}_{t''}^1, \dot{\theta}_{t'}^2[$ de θ telle qu'on ait $(F_{t'}^d + F_{t''}^g)(\theta^{\text{int}}) = 1$; θ^{int} ne dépend pas de la valeur de γ , et quand $\gamma > 0$ décroît les intervalles $]\dot{\theta}_{t''}^1, \dot{\theta}_{t'}^2[$ aussi jusqu'à ce que $\dot{\theta}_{t''}^1 = \dot{\theta}_{t'}^2 = \theta^{\text{int}}$. On a alors

$$\left\{ \begin{array}{l} P_{\theta^{\text{int}}}(\dot{\theta}_T^1, \dot{\theta}_T^2 \ni \theta^{\text{int}}) = \sum_{t' < t < t''} P_{\theta^{\text{int}}}(T = t) = F_{t'}^d(\theta^{\text{int}}) - F_{t''}^g(\theta^{\text{int}}) \stackrel{\text{d\u00e9f}}{=} \gamma_{\min} \\ P_{\theta^{\text{int}}}(\dot{\theta}_T^1 \geq \theta^{\text{int}}) = \sum_{t \geq t''} P_{\theta^{\text{int}}}(T = t) = F_{t''}^g(\theta^{\text{int}}) = \frac{1 - \gamma_{\min}}{2} \\ P_{\theta^{\text{int}}}(\dot{\theta}_T^2 \leq \theta^{\text{int}}) = \sum_{t \leq t'} P_{\theta^{\text{int}}}(T = t) = 1 - F_{t'}^d(\theta^{\text{int}}) = 1 - \frac{1 + \gamma_{\min}}{2} = \frac{1 - \gamma_{\min}}{2} \end{array} \right. .$$

On a ainsi la *condition d'équilibre* $P_{\theta^{\text{int}}}(\dot{\theta}_T^1 \geq \theta^{\text{int}}) = P_{\theta^{\text{int}}}(\dot{\theta}_T^2 \leq \theta^{\text{int}})$, et on souhaite remplacer dès que possible (c'est-à-dire pour des $\gamma > \gamma_{\min}$) $\dot{\theta}_{t''}^1$ et $\dot{\theta}_{t'}^2$ par cette valeur intermédiaire θ^{int} . Pour cela, soit

$$\left\{ \begin{array}{l} \theta_{\text{prov}}^1 = \text{la plus grande valeur de } \theta \geq \dot{\theta}_{t''}^1, \text{ telle que } F_{t'}^g - F_{t''}^g \geq \gamma \text{ sur }]\dot{\theta}_{t''}^1, \theta[\\ \theta_{\text{prov}}^2 = \text{la plus petite valeur de } \theta \leq \dot{\theta}_{t'}^2, \text{ telle que } F_{t'}^d - F_{t''}^d \geq \gamma \text{ sur }]\theta, \dot{\theta}_{t'}^2[\end{array} \right. ,$$

qui vérifient $\theta_{\text{prov}}^1 > \dot{\theta}_{t''}^1$ dès que $\dot{\theta}_{t'}^2 < \dot{\theta}_{t''}^1$ et $\theta_{\text{prov}}^2 < \dot{\theta}_{t'}^2$ dès que $\dot{\theta}_{t''}^1 > \dot{\theta}_{t'}^2$:

en effet $\dot{\theta}_{t'}^2 < \dot{\theta}_{t''}^1$ assure $F_{t'}^g(\dot{\theta}_{t''}^1) - F_{t''}^g(\dot{\theta}_{t''}^1) > F_{t'}^g(\dot{\theta}_{t'}^2) - F_{t''}^g(\dot{\theta}_{t''}^1) = \frac{1 + \gamma}{2} - \frac{1 - \gamma}{2} = \gamma$ si $t' \neq t_{\min}$ et $F_{t_{\min}}^g(\dot{\theta}_{t''}^1) - F_{t''}^g(\dot{\theta}_{t''}^1) = 1 - \frac{1 - \gamma}{2} > \gamma$; donc dans les deux cas $F_{t'}^g - F_{t''}^g > \gamma$ aussi au voisinage (droit) de $\dot{\theta}_{t''}^1$, par continuité de $F_{t'}^g - F_{t''}^g$. De même $\dot{\theta}_{t''}^1 < \dot{\theta}_{t'}^2$ assure $F_{t'}^d(\dot{\theta}_{t''}^2) - F_{t''}^d(\dot{\theta}_{t''}^2) > F_{t'}^d(\dot{\theta}_{t'}^2) - F_{t''}^d(\dot{\theta}_{t''}^2) = \frac{1 + \gamma}{2} - \frac{1 - \gamma}{2} = \gamma$ si $t'' \neq t_{\max}$ et $F_{t''}^d(\dot{\theta}_{t'}^2) - F_{t_{\max}}^d(\dot{\theta}_{t'}^2) = \frac{1 + \gamma}{2} > \gamma$; donc dans les deux cas $F_{t'}^d - F_{t''}^d > \gamma$ aussi au voisinage (gauche) de $\dot{\theta}_{t'}^2$ par continuité de $F_{t'}^d - F_{t''}^d$.

On remplace finalement $\dot{\theta}_{t''}^1$ et $\dot{\theta}_{t'}^2$ par

$$\theta_{t''}^1 \stackrel{\text{d\u00e9f}}{=} \min(\theta^{\text{int}}, \theta_{\text{prov}}^1) \text{ et } \theta_{t'}^2 \stackrel{\text{d\u00e9f}}{=} \max(\theta^{\text{int}}, \theta_{\text{prov}}^2).$$

On réduit ainsi au maximum les intervalles d'estimation de θ sous les contraintes de maintien d'une part des enchevêtrements entre intervalles et d'autre part d'un niveau de confiance au moins égal à γ pour tous les θ (hormis éventuellement $\theta_{t''}^1$ et $\theta_{t'}^2$); en effet puisque $\theta_{t''}^1 \leq \theta_{\text{prov}}^1$ et $\theta_{t'}^2 \geq \theta_{\text{prov}}^2$ on a pour les θ affectés par les remplacements :

$$\left\{ \begin{array}{l} \forall \theta \in]\dot{\theta}_{t''}^1, \theta_{t''}^1[, \sum_{t \in \dot{\Delta}, t' \leq t < t''} P_{\theta}(T = t) = F_{t'}^g(\theta) - F_{t''}^g(\theta) \geq \gamma \\ \forall \theta \in]\theta_{t'}^2, \dot{\theta}_{t'}^2[, \sum_{t \in \dot{\Delta}, t' < t \leq t''} P_{\theta}(T = t) = F_{t'}^d(\theta) - F_{t''}^d(\theta) \geq \gamma \end{array} \right. .$$

La condition $\theta_{t''}^1 \leq \theta^{\text{int}} \leq \theta_{t'}^2$ de maintien des enchevêtrements entre intervalles est indispensable, car sinon on aurait :

- soit $]\theta_{t'}^2, \theta_{t''}^1[\subset]\dot{\theta}_{t''}^1, \dot{\theta}_{t'}^2[$, et le niveau de confiance des $\theta \in]\theta_{t'}^2, \theta_{t''}^1[$ serait

$$\begin{aligned} \sum_{t \in \dot{\Delta}, t' < t < t''} P_{\theta}(T = t) &= F_{t'}^d(\theta) - F_{t''}^g(\theta) \\ &< F_{t'}^d(\dot{\theta}_{t'}^2) - F_{t''}^g(\dot{\theta}_{t''}^1) = \frac{1 + \gamma}{2} - \frac{1 - \gamma}{2} = \gamma. \end{aligned}$$

- soit $\theta_{t''}^1 > \theta_{t'}^2$ et le niveau de confiance des $\theta \in]\dot{\theta}_{t'}^2, \theta_{t''}^1[$ voisins de $\dot{\theta}_{t'}^2$ serait

$$\begin{aligned} \sum_{t \in \dot{\Delta}, t' < t < t''} P_{\theta}(T = t) &= F_{t'}^d(\theta) - F_{t''}^g(\theta) \approx F_{t'}^d(\dot{\theta}_{t'}^2) - F_{t''}^g(\dot{\theta}_{t'}^2) \\ &< F_{t'}^d(\dot{\theta}_{t'}^2) - F_{t''}^g(\dot{\theta}_{t''}^1) = \gamma. \end{aligned}$$

par continuité de $F_{t'}^d$ et $F_{t''}^g$.

- soit $\theta_{t'}^2 < \dot{\theta}_{t''}^1$ et le niveau de confiance des $\theta \in]\theta_{t'}^2, \dot{\theta}_{t''}^1[$ voisins de $\dot{\theta}_{t''}^1$ serait

$$\begin{aligned} \sum_{t \in \dot{\Delta}, t' < t < t''} P_{\theta}(T = t) &= F_{t'}^d(\theta) - F_{t''}^g(\theta) \approx F_{t'}^d(\dot{\theta}_{t''}^1) - F_{t''}^g(\dot{\theta}_{t''}^1) \\ &< F_{t'}^d(\dot{\theta}_{t'}^2) - F_{t''}^g(\dot{\theta}_{t''}^1) = \gamma. \end{aligned}$$

D'autre part la chaîne d'inégalités $\ddot{\theta}_{t'}^2 \leq \dot{\theta}_{t''}^1 \leq \theta_{t''}^1 \leq \theta_{t'}^2 \leq \dot{\theta}_{t'}^2 \leq \ddot{\theta}_{t''}^1$, montre que les intervalles modifiés contiennent toujours les noyaux des intervalles flous et sont toujours inclus dans leurs supports.

Enfin la définition de $\theta_{t''}^1$ et $\theta_{t'}^2$ assure que les intervalles $]\theta_{t'}^2, \theta_{t''}^1[$ décroissent avec γ avant de disparaître.

* Examinons ensuite toutes les autres extrémités des intervalles d'estimation de θ en considérant

- d'une part les $t'' \in \dot{\Delta} \setminus \{t_{\max}\}$ pour lesquels il existe $t' < t''$ (t' unique dans $\dot{\Delta}$) vérifiant

$$\dot{\theta}_{t'_-}^2 = \ddot{\theta}_{t'}^2 \leq \dot{\theta}_{t''}^1 < \ddot{\theta}_{t''}^1 = \dot{\theta}_{t''_+}^1 < \dot{\theta}_{t'}^2$$

(la première égalité ne figurant pas si $t' = t_{\min}$); pour eux $\theta_{t''}^2 \in [\dot{\theta}_{t''_+}^1, \dot{\theta}_{t'}^2]$ a été défini à l'étape précédente. On pose de même

- $\theta^{\text{int}} \in]\dot{\theta}_{t''}^1, \dot{\theta}_{t'}^2[$ unique solution θ de l'équation $(F_{t'}^d + F_{t''}^g)(\theta) = 1$; θ^{int} ne dépend pas de γ , et il est strictement inférieur au θ^{int} de l'étape précédente, du fait que la fonction $F_{t''}^g$ est strictement supérieure à celle de l'étape précédente.
- $\theta_{\text{prov}}^1 =$ la plus grande valeur de $\theta \in [\dot{\theta}_{t''}^1, \dot{\theta}_{t''_+}^1]$ t.q. $F_{t''}^g - F_{t''_+}^g \geq \gamma$ sur $]\dot{\theta}_{t''}^1, \theta[$, qui vérifie comme précédemment $\theta_{\text{prov}}^1 > \dot{\theta}_{t''_+}^1$ dès que $\ddot{\theta}_{t''}^2 < \dot{\theta}_{t''_+}^1$.
- $\theta_{t''}^1 \stackrel{\text{déf}}{=} \min(\theta^{\text{int}}, \theta_{\text{prov}}^1)$.

- d'autre part les $t' \in \dot{\Delta} \setminus \{t_{\min}\}$ pour lesquels il existe $t'' > t'$ (t'' unique dans $\dot{\Delta}$) vérifiant

$$\dot{\theta}_{t''}^1 < \dot{\theta}_{t'_-}^2 = \ddot{\theta}_{t'}^2 < \dot{\theta}_{t'}^2 \leq \ddot{\theta}_{t''}^1 = \theta_{t''_+}^1$$

(la dernière inégalité ne figurant pas si $t'' = t_{\max}$); pour eux $\theta_{t'}^1 \in [\dot{\theta}_{t''}^1, \dot{\theta}_{t'_-}^2]$ a été défini à l'étape précédente. On pose de même

- $\theta^{\text{int}} \in]\dot{\theta}_{t''}^1, \dot{\theta}_{t'_-}^2[$ unique solution θ de l'équation $(F_{t''}^d + F_{t'}^g)(\theta) = 1$; θ^{int} ne dépend pas de γ , et il est strictement supérieur au θ^{int} de l'étape précédente, du fait que la fonction $F_{t'}^d$ est strictement inférieure à celle de l'étape précédente.
- $\theta_{t'}^2 =$ la plus petite valeur de $\theta \in [\theta_{t'_-}^2, \dot{\theta}_{t''}^1]$ t.q. $F_{t'}^d - F_{t'_-}^d \geq \gamma$ sur $]\theta, \dot{\theta}_{t''}^1[$, qui vérifie comme précédemment $\theta_{t'}^2 < \dot{\theta}_{t''}^1$ dès que $\ddot{\theta}_{t''}^1 > \dot{\theta}_{t''}^1$.
- $\theta_{t'}^2 \stackrel{\text{déf}}{=} \max(\theta^{\text{int}}, \theta_{\text{prov}}^2)$.

On conserve ainsi le même enchevêtrement des intervalles d'estimation, et ceux-ci contiennent toujours les noyaux des intervalles flous et sont toujours inclus dans leurs supports : on a ainsi procédé à une forme efficace de *défuzzification* des intervalles flous $\tilde{\Theta}^t$, $t \in \hat{\Delta}$ en des intervalles déterministes $\Theta^t =]\theta_t^1, \theta_t^2[$ tous décroissants avec le niveau de confiance γ .

Contrairement au cas diffus, faire tendre $\gamma > 0$ vers 0 ne conduit pas à des estimations ponctuelles de θ : en effet par principe aucune valeur possible de θ n'est exclue des estimations par intervalles, autrement dit la réunion de ceux-ci coïncide toujours avec Θ . Toutefois ces intervalles $]\theta_t^1, \theta_t^2[$, $t \in \hat{\Delta}$ tendent à être 2 à 2 disjoints et contigus (et les intervalles $\Theta^t =]\theta_t^1, \theta_t^2[$ le sont effectivement pour γ plus petit qu'une certaine valeur γ_0 et sont alors stabilisés), puisqu'on a pour tout $t \in \hat{\Delta}$ distinct de t_{\max}

$$F_t^d(\dot{\theta}_t^2) = \frac{1+\gamma}{2} \approx \frac{1-\gamma}{2} = F_t^d(\ddot{\theta}_t^1) \implies \dot{\theta}_t^2 \approx \ddot{\theta}_t^1 = \dot{\theta}_{t_+}^1.$$

Les points intermédiaires $\theta_t^{\text{int}} \stackrel{\text{déf}}{=} \theta_t^2 = \theta_{t_+}^1$ séparant ces intervalles d'estimation contigus — intervalles qui apparaissent comme les “zones d'influence” de chaque point $t \in \hat{\Delta}$ — correspondent à $\gamma_{\min} = 0$ et sont donc caractérisés par

$$\forall t \in \hat{\Delta} \setminus \{t_{\max}\}, F_{t_+}^g(\theta_t^{\text{int}}) = \frac{1}{2} = 1 - F_t^d(\theta_t^{\text{int}});$$

ce qu'on retrouve en écrivant la relation d'équilibre

$$\forall t \in \hat{\Delta} \setminus \{t_{\min}, t_{\max}\}, P_{\dot{\theta}_t^1}(T < t) = P_{\dot{\theta}_t^2}(T > t) = \frac{1+\gamma}{2} \text{ avec ici } \gamma = 0.$$

Enfin le niveau de confiance associé est au moins égal à

$$\gamma_0 = \min_{\theta \in \Theta \setminus \{\theta_t^{\text{int}}\}} P_{\theta}(\cdot] \theta_T^1, \theta_T^2 [\ni \theta) = \min_{\theta \in \Theta \setminus \{\theta_t^{\text{int}}\}} P_{\theta}(T = t_{\theta}) > 0,$$

où t_{θ} est l'unique élément $t \in \hat{\Delta}$ tel que $\theta_{t_-}^{\text{int}} < \theta < \theta_t^{\text{int}}$, avec $\begin{cases} \theta_{t_{\min}-}^{\text{int}} \stackrel{\text{déf}}{=} \inf \Theta \\ \theta_{t_{\max}}^{\text{int}} \stackrel{\text{déf}}{=} \sup \Theta \end{cases}$.

Ceci sera illustré par le **modèle pile ou face** à la fin du paragraphe suivant.

6.C Exemple du modèle exponentiel pile ou face

Il s'agit d'estimer le paramètre $\theta \in \Theta =]0, 1[$ d'une variable de Bernoulli X , autrement dit la probabilité θ d'obtenir pile ($X = 1$) plutôt que face ($X = 0$) en lançant une pièce :

$$\Delta = \{0, 1\}, \mu = \delta_0 + \delta_1, f_{\theta}(x) = \theta^x (1 - \theta)^{1-x} = e^{x \ln \frac{\theta}{1-\theta} + \ln(1-\theta)},$$

où $\alpha(\theta) = \ln \frac{\theta}{1-\theta}$ a sa dérivée $\alpha'(\theta) = \frac{1}{\theta} + \frac{1}{1-\theta}$ de signe positif, donc $\varepsilon = +1$.

La statistique exhaustive $T = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$, fréquence empirique du nombre de

piles pour n tirages, est l'estimateur efficace de $-\frac{\beta'(\theta)}{\alpha'(\theta)} = \theta$ et on a

$$\forall t = \bar{x} \in [0, 1], f_t^T(\theta) = \theta^{nt} (1 - \theta)^{n(1-t)};$$

pour $t \in]0, 1[$, $\hat{\theta} = t$ est l'estimation de θ au maximum de vraisemblance.

On a $\hat{\Delta} = \tau(\Delta^n) = \{0, \frac{1}{n}, \frac{2}{n}, \dots, 1\}$; f_t^T est bien strictement décroissante de limite nulle en 1 pour $t_{\min} = 0$, d'opposée unimodale de limite nulle en 0 et 1 pour $0 < t < 1$ et strictement croissante de limite nulle en 0 pour $t_{\max} = 1$.

Comme nT est une variable binomiale de paramètre θ , on a

$$\forall j \in \{0, \dots, n\}, P_\theta\left(T = \frac{j}{n}\right) = \binom{n}{j} \theta^j (1 - \theta)^{n-j};$$

$$\text{de sorte que } F_t^g(\theta) = P_\theta(T \geq t) = \sum_{j=\lceil nt \rceil}^n P_\theta\left(T = \frac{j}{n}\right) = \sum_{j=\lceil nt \rceil}^n \binom{n}{j} \theta^j (1 - \theta)^{n-j}$$

$$\text{et } F_t^d(\theta) = 1 - P_\theta(T \leq t) = 1 - \sum_{j=0}^{\lfloor nt \rfloor} P_\theta\left(T = \frac{j}{n}\right) = 1 - \sum_{j=0}^{\lfloor nt \rfloor} \binom{n}{j} \theta^j (1 - \theta)^{n-j}.$$

F_t^d et F_t^g vérifient bien les hypothèses complémentaires :

- $\forall t \in I =]0, 1[$, $\inf F_t^g(\Theta) = 0$ et $\sup F_t^d(\Theta) = 1$

car $\theta \mapsto \frac{F_t^g(\theta)}{\theta}$ et $\theta \mapsto \frac{1 - F_t^d(\theta)}{1 - \theta}$ sont des polynômes ;

- $F_0^d(\theta) = 1 - (1 - \theta)^n$, donc $\inf F_0^d(\Theta) = 0$ et $\sup F_0^d(\Theta) = 1$;

- $F_1^g(\theta) = \theta^n$, donc $\inf F_1^g(\Theta) = 0$ et $\sup F_1^g(\Theta) = 1$.

La démarche présentée au paragraphe 6.B s'applique donc parfaitement à cet exemple, et on peut illustrer les résultats en examinant les graphes des applications

$$\theta \in]0, 1[\mapsto \varepsilon(\theta) = P_\theta(\cdot] \hat{\theta}_T^1, \hat{\theta}_T^2 [\neq \theta) = 1 - \sum_{j=0}^n \binom{n}{j} \theta^j (1 - \theta)^{n-j} \mathbb{1}_{\hat{\Theta}^{t_j}}(\theta)$$

et

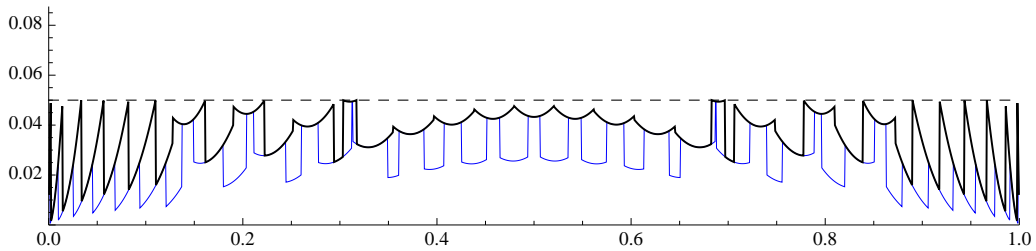
$$\theta \in]0, 1[\mapsto \varepsilon(\theta) = P_\theta(\cdot] \theta_T^1, \theta_T^2 [\neq \theta) = 1 - \sum_{j=0}^n \binom{n}{j} \theta^j (1 - \theta)^{n-j} \mathbb{1}_{\Theta^{t_j}}(\theta)$$

où $\forall j \in \{0, \dots, n\}$, $t_j = \frac{j}{n}$, $\hat{\Theta}^{t_j} = \cdot] \hat{\theta}_{t_j}^1, \hat{\theta}_{t_j}^2 [$ et $\Theta^{t_j} = \cdot] \theta_{t_j}^1, \theta_{t_j}^2 [$.

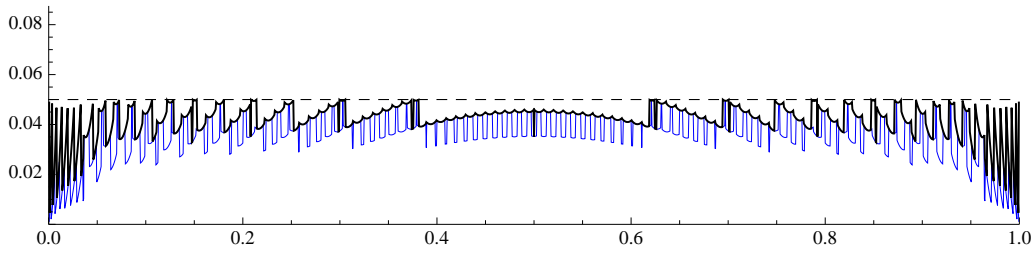
Considérons d'abord les cas de $n = 25$ et $n = 100$ tirages avec le niveau $\gamma = 0.95$:

On obtient les graphes symétriques suivants de $\theta \mapsto \varepsilon(\hat{\theta})$ et $\theta \mapsto \varepsilon(\theta)$, qui illustrent respectivement les méthodes non améliorée (en trait maigre) et améliorée (en gras) :

$n = 25$:



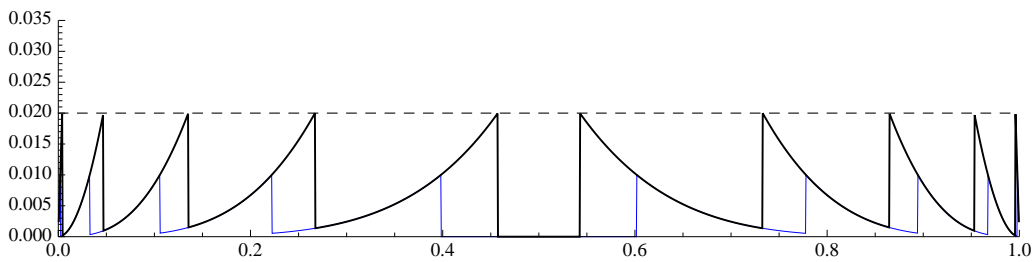
$n = 100$:



Voici d'autres situations avec n très petit !

Cas de $n = 5$ tirages avec les niveaux $\gamma = 0.98, 0.96, 0.8, 0.4, 0.3$:

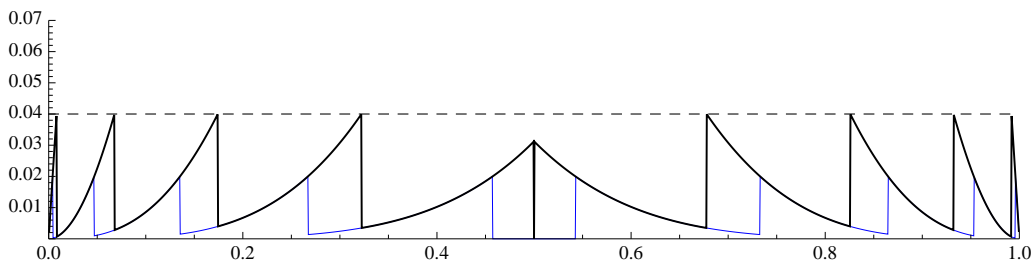
$\gamma = 0.98$:



... qui correspond aux intervalles d'estimation améliorés $\Theta^{t_j}, 0 \leq j \leq 5$:

$]0, 0.542695[$, $]0.00403239, 0.732939[$, $]0.0468944, 0.864733[$, $]0.135267, 0.953106[$,
 $]0.267061, 0.995968[$, $]0.457305, 1[$.

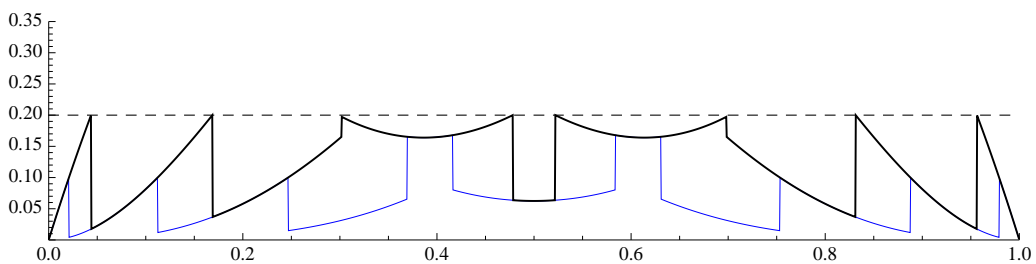
$\gamma = 0.96$:



... qui correspond aux intervalles d'estimation améliorés $\Theta^{t_j}, 0 \leq j \leq 5$:

$]0, 0.5[$, $]0.00813116, 0.677791[$, $]0.0677578, 0.825819[$, $]0.174181, 0.932242[$,
 $]0.322209, 0.991869[$, $]0.5, 1[$.

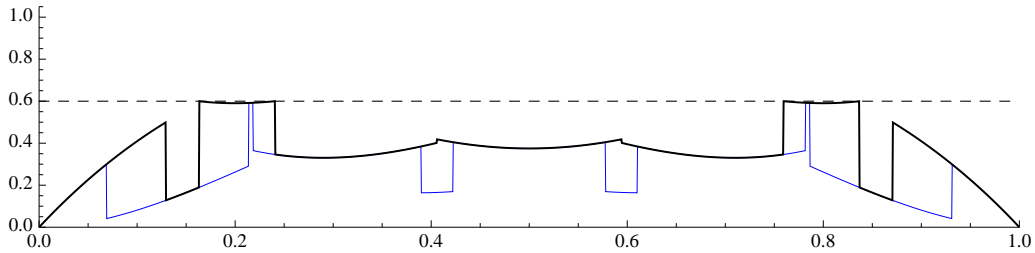
$\gamma = 0.8$:



... qui correspond aux intervalles d'estimation améliorés Θ^{t_j} , $0 \leq j \leq 5$:

$$]0, 0.298614[,]0.0436475, 0.521956[,]0.168609, 0.701386[,]0.298614, 0.831391[, \\]0.478044, 0.956352[,]0.701386, 1[.$$

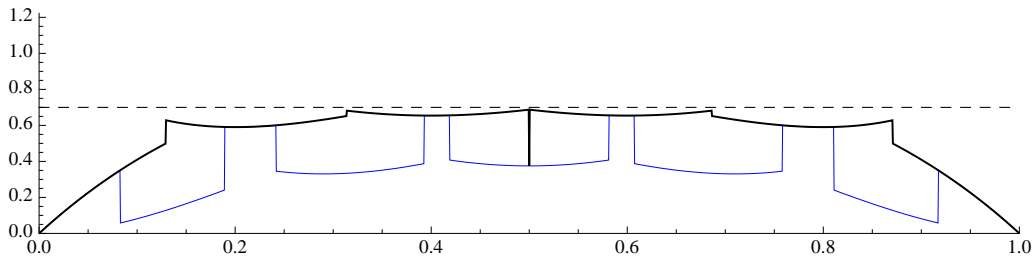
$\gamma = 0.4$:



... qui correspond aux intervalles d'estimation améliorés Θ^{t_j} , $0 \leq j \leq 5$:

$$]0, 0.163001[,]0.163001, 0.389818[,]0.240782, 0.610182[,]0.389818, 0.759218[, \\]0.610182, 0.836999[,]0.836999, 1[.$$

$0.3 \approx \gamma_0$:



... qui correspond aux intervalles d'estimation améliorés contigus Θ^{t_j} , $0 \leq j \leq 5$:

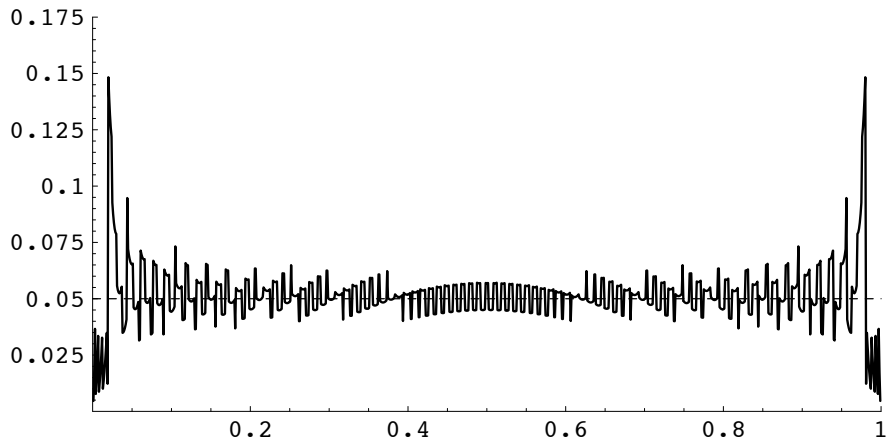
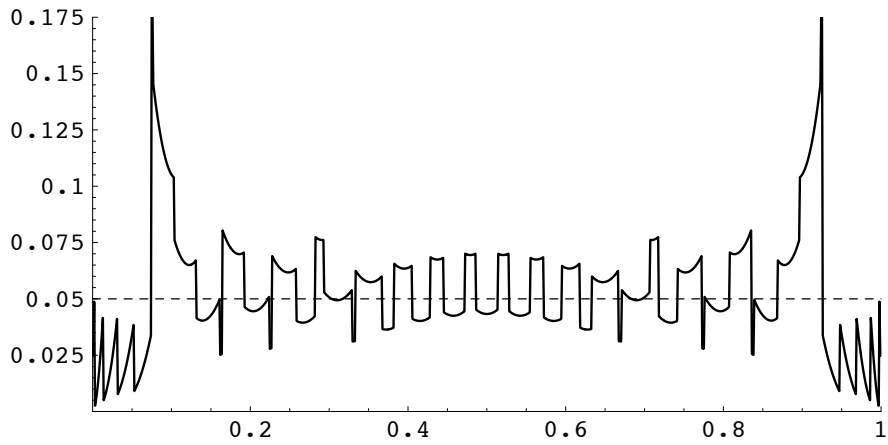
$$]0, 0.129449[,]0.129449, 0.31381[,]0.31381, 0.5[,]0.5, 0.68619[, \\]0.68619, 0.870551[,]0.870551, 1[.$$

6.D Comparaison avec d'autres méthodes pour pile ou face

- Méthode asymptotique fondée sur le théorème du rapport des vraisemblances maximales (paragraphe 3.A).

Comme vu alors, la région vraisemblable pour $\theta \{f_t^T > e^{-\frac{z}{2}} \|f_t^T\|_\infty\}$ a pour tout $t \in \dot{\Delta}$ le niveau de confiance asymptotique γ pour $z > 0$ tel que $P(\chi_d^2 < z) = \gamma$; on choisit ici $\gamma = 0.95$, d'où ($d = 1$) $z \approx 3.841$.

On obtient comme graphes de $\theta \mapsto \varepsilon(\theta)$ pour $n = 25$ et $n = 100$ tirages les deux courbes symétriques suivantes, où le niveau du test n'est cette fois pas majoré par $1 - \gamma$, et ceci de manière inacceptable pour les petites valeurs de θ et $1 - \theta$:



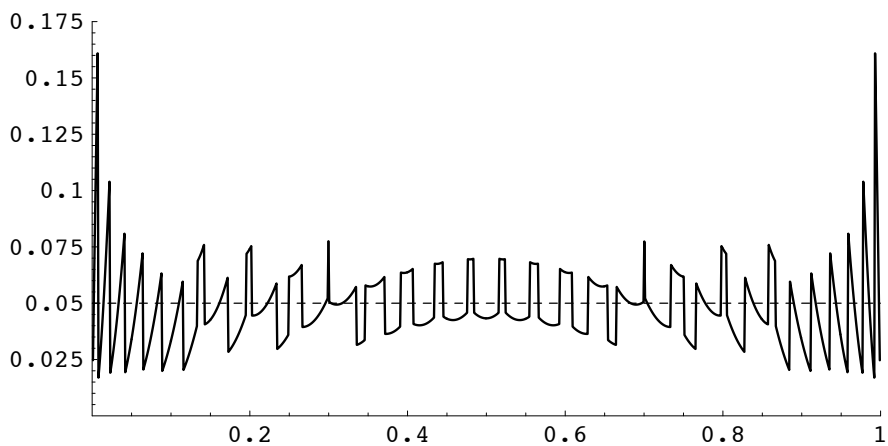
- Méthode asymptotique fondée sur le théorème central limite : on sait que $T = \bar{X}$ a pour tout $\theta \in \Theta$ l'espérance θ et la variance $\frac{\theta(1-\theta)}{n}$; on a donc

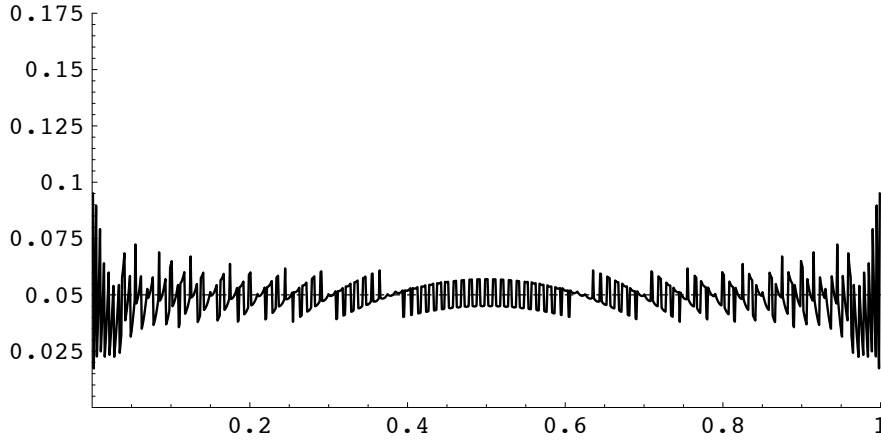
$$\sqrt{\frac{n}{\theta(1-\theta)}} (T - \theta) \stackrel{\mathcal{L}}{\approx} Z,$$

où Z est une variable gaussienne centrée réduite. On définit ainsi l'intervalle ouvert

$$\Theta^t = \left\{ \theta \in \Theta : |\theta - t| < \frac{z}{\sqrt{n}} \sqrt{\theta(1-\theta)} \right\} \quad \text{où } P(|Z| < z) = \gamma.$$

On obtient comme graphes de $\theta \mapsto \varepsilon(\theta)$ pour $n = 25$ et $n = 100$ tirages les deux courbes symétriques suivantes, moins fautives que les précédentes :



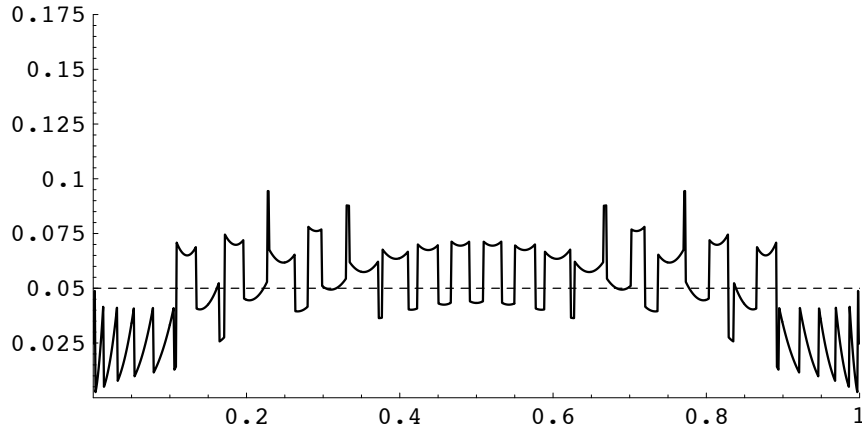


- Approche bayésienne en l'absence d'information préalable, modélisée ici par le choix de la mesure de Lebesgue $\mathbb{1}_\Theta \lambda$ (à un facteur multiplicatif près) comme mesure de référence et d'une constante comme densité a priori du paramètre θ .

Alors $\forall t \in \dot{\Delta}$, la densité a posteriori π_t sur Θ est proportionnelle à l'application f_t^T ; de sorte que l'estimation bayésienne de θ par région de plus grande densité a posteriori de niveau γ est $\Theta^t = \{\theta \in \Theta : f_t^T(\theta) > k_t\}$, où k_t est déterminé par la condition

$$(\pi_t \lambda)(\Theta^t) = \gamma \iff \int_{\Theta^t} f_t^T(\theta) d\theta = \gamma \int_{\Theta} f_t^T(\theta) d\theta.$$

On obtient comme graphe de $\theta \mapsto \varepsilon(\theta)$ pour $n = 25$ tirages la courbe symétrique :



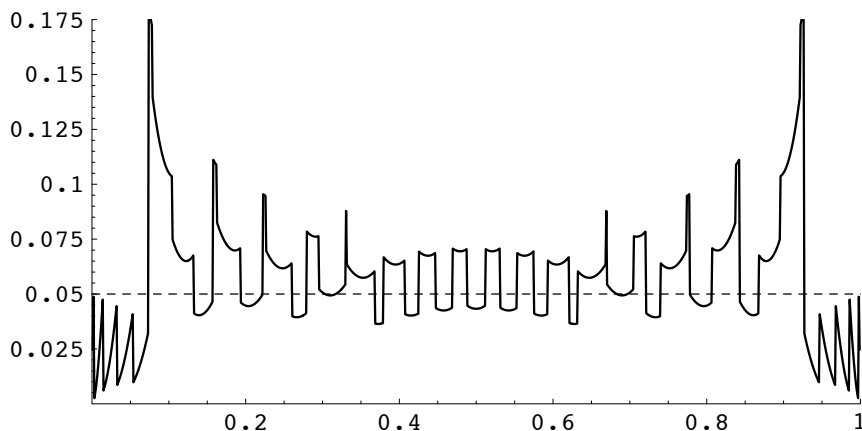
- Approche bayésienne en l'absence d'information préalable, modélisée à présent par le choix comme mesure de référence sur Θ de la mesure de Jeffreys $\sqrt{I_1} \lambda$, où $I_1(\theta) = \frac{1}{\theta(1-\theta)}$ est l'information de Fisher pour $n = 1$, et d'une constante comme densité a priori du paramètre θ .

Rappelons que la mesure de Jeffreys (définie à un facteur multiplicatif près) est réputée non informative pour le modèle et qu'on conserve avec elle la propriété d'invariance par reparamétrage; ce qui devrait imposer de la choisir comme mesure dominante sur Θ .

Alors $\forall t \in \dot{\Delta}$, $\Theta^t = \{\theta \in \Theta : f_t^T(\theta) > k_t\}$ où k_t est déterminé à présent par la condition

$$(\pi_t \lambda)(\Theta^t) = \gamma \iff \int_{\Theta^t} \frac{f_t^T(\theta)}{\sqrt{\theta(1-\theta)}} d\theta = \gamma \int_{\Theta} \frac{f_t^T(\theta)}{\sqrt{\theta(1-\theta)}} d\theta.$$

On obtient comme graphe de $\theta \mapsto \varepsilon(\theta)$ pour $n = 25$ tirages la courbe symétrique suivante, voisine de celle obtenue par la première méthode asymptotique et donc décevante :



7 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 avec telle probabilité", du fait du refus de principe de considérer le paramètre comme aléatoire.

On va montrer dans le cadre des modèles à vraisemblance pivotale qu'il est possible de déterminer pour chaque échantillon $x \in \Delta^n$ une probabilité intrinsèque sur Θ qui résoud cette difficulté ; toutefois cette *probabilité a posteriori* ne sera définie que sur la sous-tribu de la tribu de Borel $\mathcal{B}(\Theta)$ engendrée par les régions vraisemblables pour θ en x ; tout prolongement à une tribu plus vaste (en particulier à $\mathcal{B}(\Theta)$ entière, en associant à cette probabilité une densité ou une variable aléatoire) comporte en général une part d'arbitraire irréductible (perte du caractère intrinsèque), malgré l'élégance éventuelle d'une telle construction.

La situation est différente dans le cas particulier des modèles exponentiels réels de loi μ diffuse étudiés au paragraphe 5.A, car la fonction pivotale F permet de définir une probabilité intrinsèque directement sur toute la tribu $\mathcal{B}(\Theta)$, qui prolonge la probabilité précédente dans le cas *parfait* de vraisemblance pivotale ; on retombe en fait dans ce cadre sur la très discutée *probabilité fiduciaire* introduite par R.A. Fisher.

7.A La famille des probabilités a posteriori

On considère ici un modèle de dimension $d \geq 1$ quelconque, vérifiant l'hypothèse de vraisemblance pivotale, en particulier η portée par $\phi_x(\Theta)$ pour tout $x \in \Delta^n$.

Proposition : 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 vraisemblables en x une **probabilité Π_x intrinsèque** au modèle, qu'on appellera *probabilité a posteriori* sur Θ , caractérisée par

$$\forall B \in \mathcal{B}(\mathbb{R}), \boxed{\Pi_x(\phi_x \in B) \stackrel{\text{d\u00e9f}}{=} \eta(B)} \iff \underline{\phi_x(\Pi_x) = \eta};$$

de sorte que $\forall y \in]0, 1[$, le niveau de confiance $\gamma(y)$ de la r\u00e9gion vraisemblable $\{\phi_x > y\}$ co\u00efncide avec sa probabilit\u00e9 a posteriori $\underline{\Pi_x(\{\phi_x > y\})}$.

Preuve On a en toute g\u00e9n\u00e9ralit\u00e9 $B \cap \phi_x(\Theta) = \phi_x(\phi_x^{-1}(B))$; donc comme η est port\u00e9e par $\phi_x(\Theta)$, on a

$$\forall B \in \mathcal{B}(\mathbb{R}), \eta(B) = \eta(B \cap \phi_x(\Theta)) = \underline{\eta(\phi_x(\{\phi_x \in B\}))},$$

ce qui permet bien de d\u00e9finir $\Pi_x(\phi_x \in B)$ comme indiqu\u00e9. De plus, si $(\phi_x^{-1}(B_n))_{n \in \mathbb{N}}$ est une suite de bor\u00e9liens de Θ deux \u00e0 deux disjoints, les $B_n \cap \phi_x(\Theta)$ sont aussi deux \u00e0 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 g\u00e9n\u00e9ralit\u00e9 $\{\phi_x \in B\} = \phi_x^{-1}(B) = \phi_x^{-1}(B \cap \phi_x(\Theta))$, on a

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

donc

$$\Pi_x\left(\bigsqcup_{n \in \mathbb{N}} \phi_x^{-1}(B_n)\right) = \eta\left(\bigsqcup_{n \in \mathbb{N}} (B_n \cap \phi_x(\Theta))\right) = \sum_{n=0}^{\infty} \eta(B_n \cap \phi_x(\Theta)) = \sum_{n=0}^{\infty} \Pi_x(\phi_x^{-1}(B_n))$$

et Π_x est σ -additive. Enfin Π_x est positive et $\Pi_x(\Theta) = \Pi_x(\phi_x \in \mathbb{R}) = \eta(\mathbb{R}) = 1$, donc Π_x est bien une probabilit\u00e9; et par construction $\eta = \phi_x(\Pi_x)$, probabilit\u00e9 image de Π_x par ϕ_x .

Comme la fonction ϕ et la mesure η ne d\u00e9pendent pas du choix de la mesure dominante μ du mod\u00e8le, les Π_x n'en d\u00e9pendent pas non plus. Si on rempla\u00e7ait ϕ par $\phi' = \ell \circ \phi$ o\u00f9 $\ell : \mathbb{R} \rightarrow \mathbb{R}$ est un hom\u00e9omorphisme croissant, on aurait $\eta' = \ell(\eta)$ comme vu au paragraphe 2.A; donc $\forall x \in \Delta^n$ et $\forall B \in \mathcal{B}(\mathbb{R})$,

$$\Pi_x(\phi'_x \in B) = \Pi_x(\phi_x \in \ell^{-1}(B)) = \eta(\ell^{-1}(B)) = \eta'(B), \text{ de sorte que } \Pi'_x = \Pi_x.$$

Enfin soit $\theta \in \Theta \xrightarrow{\delta} \tilde{\theta} \in \tilde{\Theta}$ un hom\u00e9omorphisme de Θ sur un autre ouvert $\tilde{\Theta}$ de \mathbb{R}^d . On a $\phi_x = \tilde{\phi}_x \circ \delta$ et η n'est pas affect\u00e9e par δ (cf. le paragraphe 2.A); d'o\u00f9

$$\eta = \phi_x(\Pi_x) = \tilde{\phi}_x(\delta(\Pi_x)), \text{ et donc } \tilde{\Pi}_x = \delta(\Pi_x) \text{ sur } (\tilde{\phi}_x)^{-1}(\mathcal{B}(\mathbb{R}^d));$$

ce qui est bien la formule de transport d'une mesure par une application mesurable. \square

Remarques

- Lorsque η est diffuse, on a $\phi_x(\Pi_x) = \eta \implies \underline{(\gamma \circ \phi_x)(\Pi_x) = \gamma(\eta) = \lambda}$ o\u00f9 λ d\u00e9signe ici la probabilit\u00e9 uniforme sur $]0, 1[$ (propri\u00e9t\u00e9 d'anamorphose), puisque γ est la fonction de r\u00e9partition compl\u00e9mentaire 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 r\u00e9gion vraisemblable $\{\gamma \circ \phi_x < \gamma_0\}$ co\u00efncide avec sa probabilit\u00e9 a posteriori $\underline{\Pi_x(\{\gamma \circ \phi_x < \gamma_0\})}$.

- Conformément à l'analyse développée au paragraphe 2.B, interpréter la relation $\Pi_x(\phi_x \in B) \stackrel{\text{déf}}{=} \eta(B)$ comme le fait que la probabilité a posteriori de $\{\phi_x \in B\}$ vaut $\eta(B)$ n'apparaît *légitime* que si le borélien B est lui-même défini *a priori*, au sens qu'il ne dépend pas de la valeur de l'échantillon x . Si ceci est bien exact, il s'agit d'une limitation profonde à l'emploi de cette probabilité. . .
- Dans le cas où on dispose d'une statistique exhaustive pour le modèle, on sait (cf. le paragraphe 2.C) que $\forall (x, \theta) \in \Delta^n \times \Theta$, $\phi(x, \theta) = \phi^T(t, \theta)$ où $t = \tau(x)$. Comme par définition Π_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

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

la probabilité a posteriori Π_x ne dépend en fait que de t ; on la notera en conséquence aussi bien $\underline{\Pi}_t$.

7.B Cas des modèles exponentiels réels

On se place dans le cadre des modèles exponentiels réels présentés au paragraphe 5.A, où T est de loi diffuse et où pour tout $t \in I$, $F_t(\Theta) =]0, 1[$.

On sait que pour tout $t \in I$, F_t est un homéomorphisme croissant de Θ sur $]0, 1[$, ce qui permet de définir simplement une probabilité intrinsèque Π_t sur la tribu de Borel $\mathcal{B}(\Theta)$ en transportant la loi de $F_\theta(T)$ uniforme sur $]0, 1[$ par F_t^{-1} définie sur $]0, 1[$:

$$\Pi_t \stackrel{\text{déf}}{=} F_t^{-1}(\bar{\lambda}),$$

$\bar{\lambda}$ désignant la probabilité uniforme sur $]0, 1[$.

Comme F est une fonction pivotale correcte de loi fixe λ vu les hypothèses, on peut pour tout $B \in \mathcal{B}(]0, 1[)$ attribuer à la région $\{F_t \in B\}$ le niveau de confiance $\bar{\lambda}(B)$, et celui-ci coïncide alors avec sa probabilité $\Pi_t(\{F_t \in B\})$ puisque $F_t(\Pi_t) = \bar{\lambda}$.

On a pour tous $\theta_1 < \theta_2$ dans Θ ,

$$\Pi_t(] \theta_1, \theta_2]) = \bar{\lambda}(F_t(] \theta_1, \theta_2])) = \bar{\lambda}(] F_t(\theta_1), F_t(\theta_2)]) = F_t(\theta_2) - F_t(\theta_1);$$

de sorte que Π_t a F_t pour **fonction de répartition** et donc $\pi_t \stackrel{\text{déf}}{=} \frac{\partial F}{\partial \theta}(t, \cdot)$ pour **densité** par rapport à la mesure de Lebesgue $\mathbb{1}_\Theta \lambda$ (sous réserve d'existence de cette dérivée partielle).

De plus, dans le cas d'un modèle exponentiel *parfait* où $G = \phi_t^T \circ F_t^{-1}$ ne dépend pas de t , on a

$$\forall \theta \in \Theta, \eta = \phi_\theta^T(T)(P_\theta) = G(F_\theta(T)(P_\theta)) = G(\bar{\lambda});$$

donc $\forall t \in I$, $\phi_t^T(\Pi_t) = G(F_t(\Pi_t)) = G(\bar{\lambda}) = \eta$: la restriction de Π_t à la sous-tribu $(\phi_t^T)^{-1}(\mathcal{B}(\mathbb{R}))$ de $\mathcal{B}(\Theta)$ engendrée par les régions vraisemblables en t **coïncide** avec la probabilité a posteriori définie en dernière remarque du paragraphe 7.A et désignée par la même notation.

L'idée fort séduisante proposée par R.A. Fisher, est que dans une situation de ce type — dont on ne verrait ici qu'un cas particulier — il deviendrait légitime de considérer pleinement le paramètre θ comme un **point aléatoire** de Θ ayant π_t pour densité, $\mathbb{I}_\Theta \lambda$ étant prise comme mesure de référence sur Θ ; et cette *densité fiduciaire* π_t pourrait servir de densité bayésienne a priori à l'occasion d'un nouvel échantillonnage : ce qui permettrait enfin de faire la jonction entre statistique fréquentiste et statistique bayésienne.

Nous contestons cette idée et considérons définitivement le modèle bayésien comme structurellement plus riche d'informations pour les deux raisons suivantes :

- Si la densité fiduciaire $\pi_{t'}$ est fournie par un premier échantillonnage connu donnant $t' = \tau(x'_1, \dots, x'_m)$ et si on fait un second tirage donnant $t = \tau(x_1, \dots, x_n)$, la densité a posteriori $\pi_{(t',t)}$ sur Θ qui en découlerait en bayésien, censée pouvoir servir de densité a priori pour un tirage ultérieur, serait

$$\pi_{(t',t)} \propto \frac{\partial F'}{\partial \theta}(t', \cdot) f_t^T.$$

Or il paraît aberrant que l'ordre dans lequel les deux premiers tirages t' et t ont eu lieu puisse affecter cette densité ($\pi_{(t',t)} \neq \pi_{(t,t')}$). Tel n'est pas le cas dans la proposition exposée aux paragraphes 1.C et 2.C, où la vraisemblance a posteriori est $f_{(t',t)}^{T''} = f_{t'}^{T'} f_t^T$; ni dans le paradigme bayésien authentique où la densité a posteriori est

$$\pi_{(t',t)} \propto \pi f_{t'}^{T'} f_t^T,$$

π désignant la densité a priori partie intégrante du modèle bayésien de départ.

- Il nous semble de plus que subsiste la difficulté de fond déjà signalée (seconde remarque du paragraphe 7.A), à savoir ici que la relation $\Pi_t(\{F_t \in B\}) = \lambda(B)$ n'a le sens de "la probabilité a posteriori de $\{F_t \in B\}$ vaut $\eta(B)$ " que si le borélien B ne dépend pas de t .

Index alphabétique

D

densité fiduciaire 46

E

exhaustif 10

F

fonction \bar{K} 17, 21, 22

fonction pivotale 6, 23, 29

fonction pivotale correcte . 8, 11, 24,
30

H

hypothèse de vraisemblance pivotale
6, 9, 12

I

intrinsèque au modèle . 9, 10, 23, 29

invariance par reparamétrage 4, 5, 9,
42

M

mesure de Jeffreys 42

modèle des exponentielles translatées
14

modèle des lois de Poisson 30

modèle des lois Gamma 13, 26

modèle des lois gaussiennes . 18, 20,
21

modèle des lois uniformes 15

modèle exponentiel réel parfait .. 24

modèle Gamma-Weibull 16

modèle pile ou face 37

N

niveau de confiance 7

notation croisée 4

P

probabilité a posteriori 43

probabilité fiduciaire 43

probabilités a posteriori 43

propriété pivotale 6

R

région de confiance pour θ en x ... 7

région vraisemblable pour θ en x 4, 8

rapport des vraisemblances maximales
12

région aléatoire de confiance 7

S

situation asymptotique .. 12, 19, 27,
40, 41

statistique exhaustive 10

V

vraisemblance a posteriori en x ... 6

vraisemblance a priori 6

vraisemblance conditionnelle 5

vraisemblance marginale .. 5, 20, 22

vraisemblance pivotale 6, 9, 12