

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ESTIMATION BAYÉSIENNE EMPIRIQUE POUR LES
PLANS D'EXPÉRIENCE NON ÉQUILIBRÉS

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN MATHÉMATIQUES

PAR

AHMED EL-HABTI

JUIN 2007

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Tout d'abord, je tiens à remercier mes directeurs de recherche, les professeurs Glenn Shorrock et Brenda MacGibbon. Mme Brenda MacGibbon, pour son soutien académique et financier par le biais des octrois de recherche CRSNG et FQRNT, pour sa continuelle disponibilité, sa grande patience ainsi que sa gentillesse. M. Glenn Shorrock, pour son soutien scientifique et ses précieux conseils dans ma recherche. Je tiens aussi à remercier M. Michel Adès pour le soutien financier de son octroi CRSNG.

Enfin, j'exprime toute ma reconnaissance envers mes frères, ma sœur et particulièrement ma mère pour leurs encouragements, leur soutien et surtout l'amour dont ils ont toujours su m'entourer.

TABLE DES MATIÈRES

LISTE DES FIGURES	v
LISTE DES TABLEAUX	vi
RÉSUMÉ	vii
INTRODUCTION	1
CHAPITRE I	
L'ANALYSE BAYÉSIENNE ET BAYÉSIENNE EMPIRIQUE	5
1.1 Introduction	5
1.2 Analyse bayésienne	6
1.3 L'analyse bayésienne hiérarchique	14
1.4 L'analyse bayésienne empirique	16
1.5 Le lien entre les estimateurs bayésiens empiriques et les estimateurs du type James-Stein pour la loi multinomiale	19
CHAPITRE II	
LOI NORMALE	21
2.1 Estimateurs bayésiens empiriques du vecteur moyen de la loi normale	21
2.2 Estimation bayésienne empirique	23
2.3 Proposition d'un modèle robuste.	26
2.4 Application.	30
2.5 Un estimateur pour le cas non équilibré	32
2.6 Application	39
CHAPITRE III	
LOI DE POISSON	42
3.1 Aspect théorique	42
3.2 Différentes méthodes d'estimation bayésienne empirique paramétrique et non paramétrique	47
3.3 Estimation bayésienne empirique pour les échantillons de tailles inégales	53
CHAPITRE IV	

LOI EXPONENTIELLE	58
4.1 Approche paramétrique	58
4.2 Approche non paramétrique	65
4.2.1 Étude de simulation	66
APPENDICE A	
PROGRAMMATION	72
RÉFÉRENCES	99

LISTE DES FIGURES

3.1	La variation de l'estimateur $\hat{\beta}$ en fonction de n dans le cas de la loi Poisson-gamma	57
4.1	Risque intégré de l'estimateur bayésien empirique	60
4.2	L'estimateur bayésien empirique et la vraie valeur de θ	61
4.3	L'estimation de la loi <i>a priori</i> utilisant le modèle exponentiel <i>a priori</i> versus la densité exponentielle avec paramètre $\lambda = 0.2$	63
4.4	L'estimation de la densité <i>a priori</i> utilisant un modèle <i>a priori</i> gamma inapproprié versus la densité exponentielle avec paramètre $\lambda = 0.2$	64

LISTE DES TABLEAUX

2.1	Estimation bayésienne empirique pour les 7 joueurs	31
2.2	Estimation bayésienne empirique pour les 7 joueurs avec présence d'une valeur aberrante en utilisant la loi de Student t comme mesure à <i>priori</i>	32
2.3	Estimateur minimax dans le cas non équilibré	41
3.1	Les fréquences regroupées pour la méthode de khi-deux	50
3.2	Estimation pour la loi de Poisson	52
3.3	Le cas de taille inégale pour la loi Poisson-gamma	56
4.1	Estimateur en fonction de la taille de l'échantillon	62
4.2	EQM pour les trois estimateurs en fonction de n et λ	68
4.3	EQM pour les deux estimateurs en fonction de n dans le cas gamma	68
4.4	$\Pr(EQM2 \leq EQM1)$ avec le nombre de simulations égal à 10^6	69

RÉSUMÉ

Dans notre travail de mémoire, nous présentons l'approche bayésienne empirique dans l'inférence statistique. Les premiers articles dans ce domaine ont été publiés par Robbins (1951, 1955, 1980). Robbins a utilisé une méthode non paramétrique (Maritz et Lwin (1989)) pour estimer la densité marginale. Plus tard, Morris (1983) a introduit l'approche paramétrique bayésienne empirique (voir aussi Efron et Morris (1973) (a), Casella (1985)). Nous décrivons les deux approches pour l'estimation de la moyenne de la loi gaussienne, de la loi de Poisson et de la loi exponentielle. Pour le cas gaussien, nous adaptons une méthodologie proposée par Angers (1992) pour l'estimation bayésienne hiérarchique à l'estimation bayésienne empirique dans le but d'obtenir des estimations plus robustes. Nous nous intéressons à l'estimation de la moyenne gaussienne et de la moyenne de la loi de Poisson quand les tailles des groupes sont inégales. Pour le cas gaussien, nous utilisons un estimateur du type James-Stein d'après Berger et Bock (1976) pour incorporer les tailles inégales. Dans le cas de la loi de Poisson, nous utilisons une méthode proposée par Maritz et Lwin (1989). Nous étudions également les estimateurs bayésiens empiriques pour estimer une moyenne exponentielle. Pour ce cas, nous avons introduit un nouvel estimateur bayésien empirique qui semble prometteur. Dans le cas gaussien, nous illustrons les approches en utilisant des banques de données pertinentes. Dans les autres cas, nous effectuons les études de simulation.

Mots-clés : Analyse bayésienne, bayésien empirique, estimateur de stein, estimateur minimax, plans d'expérience non équilibrés.

INTRODUCTION

Une des méthodologies très importante pour faire de l'inférence statistique paramétrique est l'analyse bayésienne (Berger (1985) et Robert(1992, 2001)). Au lieu d'utiliser les données pour estimer un paramètre directement, on suppose que le paramètre suit une loi *a priori* et on utilise la loi *a posteriori* pour l'estimation. Une critique de cette méthode est que l'on ne peut pas connaître la loi *a priori* avec certitude. L'approche bayésienne pour modéliser l'incertitude consiste à mettre les mesures *a priori* sur les paramètres de la loi *a priori* originale. Ce type de modélisation bayésienne est appelé la méthode bayésienne hiérarchique (Berger (1985) et Robert (1992, 2001)). Par contre, la méthode d'analyse bayésienne empirique, qui est un mélange de l'inférence bayésienne et fréquentiste, est une approche pour rectifier l'incertitude concernant la loi *a priori*. La méthode a été introduite par Robbins (1951, 1955, 1964, 1980). Dans cette méthode, il faut estimer la loi marginale. Robbins (1955) a proposé une approche non paramétrique pour estimer cette loi. Étant donné que l'estimateur non paramétrique à noyau de la fonction de densité (Devroye et al. (1985)) requiert un échantillon assez grand pour avoir une bonne approximation, les chercheurs ont essayé de trouver d'autres approches. Morris (1983) a introduit la méthodologie bayésienne empirique paramétrique et cette méthodologie a joué un rôle très important dans beaucoup de domaines.

Nous avons choisi d'étudier la méthodologie bayésienne empirique appliquée à l'estimation des moyennes de la loi gaussienne, de la loi de Poisson et de la loi exponentielle. La plupart des modèles statistiques pour les variables continues utilisent la loi gaussienne. La loi de Poisson joue un rôle très important dans la modélisation des données de comptage et la loi exponentielle joue un rôle très important dans l'étude des données en analyse de survie. Pour la loi gaussienne, nous présentons premièrement l'approche

d'Efron et Morris (1977) et de Casella (1985) pour l'estimation bayésienne empirique de la moyenne. Deuxièmement, étant donné l'importance de la robustesse dans la statistique bayésienne d'après Berger (1985) et Angers (1992), nous adaptons l'approche robuste d'Angers(1992) à la méthodologie bayésienne empirique en utilisant la loi de Student t comme la loi *a priori*.

Il faut noter que Efron et Morris (1977) et Casella (1985) ont travaillé avec des tailles d'échantillons égales. Il n'est pas réaliste de penser qu'en général on aura exactement les même tailles pour chaque échantillon. Cela nous a mené à la recherche d'une autre approche pour les tailles inégales. Étant donné le lien entre l'estimation minimax de James-Stein et l'estimation bayésienne empirique, nous utilisons donc l'approche de Berger et Bock (1976) pour obtenir des estimateurs dans le cas de tailles d'échantillons inégales que l'on appelle le cas non équilibré.

Pour la loi de Poisson, nous suivons premièrement l'approche originale de Robbins (1955). Deuxièmement, nous développons l'approche bayésienne empirique paramétrique de Morris (1983) pour estimer la moyenne.

La loi exponentielle joue un rôle fondamental dans la modélisation des données de survie. De plus, les temps d'attente entre les arrivées distribuées selon la loi de Poisson suivent la loi exponentielle. Aussi, cette loi peut modéliser les temps d'attente entre les inclusions dans un essai clinique. Pour la loi exponentielle, nous utilisons l'approche bayésienne empirique paramétrique.

Nous nous servons de deux bases de données très intéressantes pour illustrer la méthodologie décrite dans ce mémoire. Premièrement, nous étudions l'estimation d'une moyenne gaussienne. Nous choisissons des données de baseball qui ont été analysées originalement par Efron et Morris (1977) et plus tard par Casella(1985). La provenance de ces données peut être décrite de la manière suivante. Étant donné que le baseball est un sport très populaire en Amérique du Nord, il y a beaucoup de statistiques disponibles concernant ce sport. La moyenne au « bâton » est un score d'un grand intérêt pour les amateurs de baseball. Les chercheurs s'intéressent à la prédiction du score à la fin

de la saison en utilisant les scores au milieu de la saison. Efron et Morris (1977) et Casella (1985) ont utilisé ces données pour élucider le phénomène de Stein et l'analyse bayésienne empirique.

L'autre base de données pour l'estimation des moyennes gaussiennes dans le cas non équilibré concerne la durée de vie des cellules d'aluminium dans une usine après une grève. La provenance de ces données peut être décrite de la manière suivante. À Baie Comeau, en 1967, la grève des employés de la compagnie Reynolds provoque l'interruption du processus d'électrolyse. Ce processus consiste à faire fondre l'aluminium à très haute température à l'aide d'un courant électrique : l'anode est une électrode de carbone et la cathode est un énorme creuset de carbone où des morceaux d'acier, servant de conducteurs, y sont fixés. Ces « cuves » sont appelées cellules et elles ont une durée de vie qui dépend de la déformation du creuset. La compagnie a affirmé que lorsque les grévistes ont coupé l'électricité, le refroidissement précipité des cuves a endommagé les cellules d'électrolyse et donc réduit leur durée de vie. Un procès intenté contre la Confédération des Syndicats Nationaux par les dirigeants de l'usine a duré plus de dix ans et a fait appel à plusieurs spécialistes, y incluant des statisticiens qui ont étudié la perte de vie des cellules pour savoir si la compagnie avait perdu de l'argent. L'article de Gentleman et Whitmore (1982) présente ces données et nous avons modélisé ces données en supposant que la durée de vie suit une loi log-normale.

La motivation pour l'étude de l'estimateur bayésien empirique du paramètre de la loi de Poisson est liée à un problème en analyse séquentielle des essais cliniques mentionné dans MacGibbon *et al.*(1988). Plus précisément, ça peut être très important d'avoir un « bon » estimateur du taux d'inclusion dans un essai clinique pour implanter leur méthodologie.

Ce mémoire est organisé de la façon suivante. Afin de rendre la lecture plus facile, nous incluons une introduction à l'analyse bayésienne et bayésienne empirique dans le premier chapitre. Le deuxième chapitre contient la méthode de l'estimation bayésienne empirique pour la moyenne d'une loi gaussienne y compris des exemples utilisant les

deux bases de données décrites plus haut. Dans le troisième chapitre, nous décrivons les méthodes bayésiennes empiriques pour l'estimation du paramètre d'une loi de Poisson. Dans le quatrième chapitre nous étudions les estimateurs bayésiens empiriques de la moyenne de la loi exponentielle. Pour ce cas, nous décrivons un nouvel estimateur. Nous illustrons la méthodologie dans le deuxième chapitre avec des bases de données pertinentes et nous effectuons des études de simulation dans les deux chapitres. Les programmes se trouvent en annexe.

CHAPITRE I

L'ANALYSE BAYÉSIENNE ET BAYÉSIENNE EMPIRIQUE

1.1 Introduction

Dans ce chapitre, nous présentons trois méthodes d'estimation : l'analyse bayésienne classique, l'analyse bayésienne hiérarchique et l'analyse bayésienne empirique.

La méthode bayésienne empirique est utilisée pour l'estimation des paramètres d'une loi quand les observations sont des variables aléatoires indépendantes identiquement distribuées qui suivent cette loi. On peut mentionner que cette technique n'est pas une méthode bayésienne pure car elle fait appel à des approximations fréquentistes dans le cas où l'information *a priori* n'est pas disponible ou est insuffisante. Cependant un résultat montre qu'on peut obtenir des résultats asymptotiquement équivalents à ceux du modèle bayésien hiérarchique. L'analyse bayésienne empirique peut être vue comme une bonne combinaison des méthodes fréquentiste et bayésienne.

Les définitions et exemples donnés ici se retrouvent dans plusieurs références. Nous notons Berger (1985) et Robert (1992) pour la partie bayésienne et bayésienne hiérarchique et Maritz et Lwin (1989) pour la partie bayésienne empirique. Nous indiquons également le lien entre les estimateurs bayésiens empiriques et les estimateurs de type James-Stein.

1.2 Analyse bayésienne

Définition 1.1

Un modèle statistique bayésien est composé d'une fonction de densité paramétrique $f(x|\theta)$, d'une fonction de répartition *a priori* $G(\theta)$ et d'une densité *a priori* $g(\theta)$ définie sur Θ , l'espace des paramètres.

On suppose que

- les observations X_1, \dots, X_n sont indépendantes et identiquement distribuées avec densité $f(x|\theta)$;
- $f(x|\theta)$ est la fonction de densité de la variable aléatoire X sachant la valeur de la variable aléatoire θ ;
- $\tilde{x} = (x_1, \dots, x_n)$ sont les valeurs observées de (X_1, \dots, X_n) ;
- $g(\theta)$ modélise la fonction de densité *a priori* sur θ ;
- $G(\theta)$ modélise la fonction de répartition de θ ;
- la loi *a posteriori* sur laquelle l'inférence bayésienne est basée, peut s'écrire :

$$\begin{aligned} \pi(\theta|\tilde{x}) &= \frac{\text{densité conjointe de } \theta \text{ et } \tilde{x}}{\text{densité marginale de } \tilde{x}} \\ &= \frac{g(\theta) \prod_{i=1}^n f(x_i|\theta)}{\int g(\theta) \prod_{i=1}^n f(x_i|\theta) d\theta} \end{aligned}$$

- $f(x) = \int_{\Theta} f(x|\theta)g(\theta)d\theta$ est la densité marginale de X .

La densité marginale *a posteriori* $\pi(\theta|\tilde{x})$ satisfait les deux propriétés fondamentales suivantes.

Théorème 1.1 (Lehmann(1983))

Soient x_1, \dots, x_n et y_1, \dots, y_m deux séries d'observations provenant de variables aléatoires indépendantes et identiquement distribuées. Alors la densité *a posteriori* $\pi(\theta|\tilde{x})$ respecte le principe d'exhaustivité et le principe de vraisemblance, c'est-à-dire :

- si T est une statistique exhaustive pour θ telle que $T(x_1, \dots, x_n) = T(y_1, \dots, y_m)$ alors les deux échantillons amènent à la même inférence sur θ .

– si $l(\theta|x_1, \dots, x_n) \propto l(\theta|y_1, \dots, y_m)$ où l est la fonction de vraisemblance de $f(\tilde{x}|\theta)$ alors les deux échantillons mènent à la même inférence sur θ .

En effet, on pose $\tilde{x} = (x_1, \dots, x_n)$ et $\tilde{y} = (y_1, \dots, y_m)$; soit T une statistique exhaustive pour θ . Alors, d'après le théorème de factorisation de Lehmann (voir Lehmann(1982) Théorème 5.2) il existe une fonction $h(\tilde{x})$ qui ne dépend pas de θ et une fonction de T sachant θ , $u(t|\theta)$ telle que $f(\tilde{x}|\theta) = h(\tilde{x})u(t|\theta)$; donc

$$\begin{aligned}\pi(\theta|\tilde{x}) &= \frac{f(\tilde{x}|\theta)g(\theta)}{\int_{\Theta} f(\tilde{x}|\theta)g(\theta)d\theta} \\ &= \frac{h(\tilde{x})u(t|\theta)g(\theta)}{\int h(\tilde{x})u(t|\theta)g(\theta)d\theta} \\ &= \frac{u(t|\theta)g(\theta)}{\int u(t|\theta)g(\theta)d\theta} \\ &= \pi(\theta|t).\end{aligned}$$

Et puisque

$$l(\theta|\tilde{x}) \propto l(\theta|\tilde{y}),$$

on déduit :

$$\begin{aligned}\pi(\theta|\tilde{x}) &= \frac{l(\theta|\tilde{x})g(\theta)}{\int_{\Theta} l(\theta|\tilde{x})g(\theta)d\theta} \\ &\propto \frac{l(\theta|\tilde{y})g(\theta)}{\int_{\Theta} l(\theta|\tilde{y})g(\theta)d\theta} \\ &= \pi(\theta|\tilde{y}).\end{aligned}$$

Définition 1.2

Une règle de décision, notée $\delta(\tilde{x})$, est une fonction de S vers D : $\tilde{x} \mapsto \delta(\tilde{x})$ où S est l'espace échantillonnal et D l'espace de toutes les décisions possibles.

Définition 1.3

Une fonction de perte, dénotée par $L(\delta(x), \theta)$:

$$\begin{aligned} L : D \times \Theta &\longrightarrow [0, \infty) \\ (\delta(x), \theta) &\longrightarrow L(\delta(x), \theta) \end{aligned}$$

est une fonction mesurable habituellement continue et convexe.

La fonction de perte usuelle est la fonction de perte quadratique $L(\delta(x), \theta) = (\delta(x) - \theta)^2$, où δ est une règle de décision et un estimateur de θ . De temps en temps, on utilise une fonction de perte quadratique pondérée (habituellement pondérée par la variance). L'importance intuitive de ce type de pondération a été soulignée dans les articles de Efron et Morris (1973b, 1975, 1976a, 1976b) et par Berger et Bock (1976).

La fonction de risque fréquentiste de δ associée à $L(\delta(x), \theta)$ est définie comme :

$$R_\delta(\theta) = \int L(\delta(x), \theta) f(x|\theta) dx.$$

Définition 1.4

- i. Le risque minimax associé à la fonction de perte $L(\delta(x), \theta)$ est la valeur \bar{R} telle que $\bar{R} = \inf_{\delta \in D} \sup_{\theta \in \Theta} R_\delta(\theta)$.
- ii. Un estimateur δ_0 est minimax si $R_{\delta_0}(\theta) = \inf_{\delta \in D} \sup_{\theta \in \Theta} R_\delta(\theta)$.

Théorème 1.2 (Robert(2001))

Soit D , l'ensemble de décisions, est un ensemble convexe de \mathbb{R}^k . Si $L(\delta, \theta)$ est une fonction de perte continue et convexe en $\theta \forall \theta \in \Theta$, alors il existe un estimateur minimax. La démonstration de ce théorème se trouve dans (Robert(2001)).

Définition 1.5

Un estimateur $\delta(x)$ est admissible s'il n'existe aucun autre estimateur $\delta_2(x)$ tel que $R_{\delta_2}(\theta) \leq R_\delta(\theta) \forall \theta \in \Theta$ et $R_{\delta_2}(\theta) < R_\delta(\theta)$ pour un certain $\theta \in \Theta$.

On a besoin d'un critère pour choisir un tel δ admissible, d'où la nécessité de l'approche bayésienne.

Définition 1.6

La fonction de risque intégrée de δ est définie par :

$$W(\delta) = \int \int L(\delta(x), \theta) f(x|\theta) g(\theta) dx d\theta.$$

L'approche bayésienne emploie $W(\delta)$ pour évaluer l'estimateur $\delta(x)$ en tenant compte de toutes les valeurs possibles de θ .

Définition 1.7

Si un estimateur δ^B existe tel que

$$W(\delta^B) \leq W(\delta)$$

pour tout autre estimateur δ de θ , alors δ^B est appelé l'estimateur bayésien.

Lemme 1.1 (Robert(2001))

Si un estimateur δ de θ domine la règle de Bayes δ^B , c'est-à-dire que

$$\forall \theta \in \Theta, \quad R_\delta(\theta) \leq W(\delta^B),$$

alors il est un estimateur minimax.

Lemme 1.2 (Robert(2001))

S'il existe une suite de densités *a priori* propres π_m , $m \geq 1$ telle que la règle de Bayes par rapport à la loi impropre π satisfait

$$R_{\delta^{B\pi}}(\theta) \leq \lim_{n \rightarrow \infty} W(\delta^{B\pi_m}) < \infty \quad \forall \theta \in \Theta,$$

alors δ^B est minimax.

Proposition 1.1 (Robert(2001))

S'il existe un unique estimateur δ_0 minimax, cet estimateur est admissible.

Démonstration : Supposons que δ_0 n'est pas admissible, alors il existe δ_1 telle que

$$R_{\delta_1}(\theta) \leq R_{\delta_0}(\theta),$$

$\forall \theta \in \Theta$ avec inégalité stricte pour au moins une valeur de θ d'où :

$$\sup_{\theta} R_{\delta_1}(\theta) \leq \sup_{\theta} R_{\delta_0}(\theta) = \inf_{\delta} \sup_{\theta} R_{\delta}(\theta),$$

parce que δ_0 est minimax, en contradiction avec l'hypothèse que δ_0 est l'unique estimateur minimax. Donc δ_0 est minimax.

Voici la proposition 1.2 qui donne une propriété importante des estimateurs admissibles avec un risque constant.

Proposition 1.2 (Robert(2001))

Si un estimateur est admissible avec un risque constant alors il est minimax.

Démonstration : voir Robert(2001).

Comme exemple, si on considère la fonction de perte usuelle :

$$L(\delta(x), \theta) = (\delta(x) - \theta)^2,$$

alors

$$\begin{aligned}
 W(\delta) &= \int \int L(\delta(x), \theta) f(x|\theta) g(\theta) d\theta dx \\
 &= \int \int (\delta(x) - \theta)^2 f(x|\theta) g(\theta) d\theta dx \\
 &= \int \int (\delta(x) - \delta^B(x) + \delta^B(x) - \theta)^2 f(x|\theta) g(\theta) d\theta dx \\
 &= \int \int (\delta(x) - \delta^B(x))^2 f(x|\theta) g(\theta) d\theta dx \\
 &\quad + \int \int (\delta^B(x) - \theta)^2 f(x|\theta) g(\theta) d\theta dx \\
 &\quad + 2 \int \int (\delta(x) - \delta^B(x)) (\delta^B(x) - \theta) f(x|\theta) g(\theta) d\theta dx \\
 &= W(\delta^B) + \int \int (\delta^B(x) - \theta)^2 f(x|\theta) g(\theta) d\theta dx \\
 &\quad + 2 \int \int (\delta(x) - \delta^B(x)) (\delta^B(x) - \theta) f(x|\theta) g(\theta) d\theta dx.
 \end{aligned}$$

Donc, pour minimiser $W(\delta)$ par rapport à δ , il suffit que :

$$\frac{\partial}{\partial \delta} (2 \int \int (\delta(x) - \delta^B(x)) (\delta^B(x) - \theta) f(x|\theta) g(\theta) d\theta dx) = 0,$$

ou encore, il suffit que

$$\int (\delta^B(x) - \theta) f(x|\theta) g(\theta) d\theta = 0.$$

Ceci équivaut à

$$\begin{aligned}
 \delta^B(x) &= \frac{\int \theta f(x|\theta) g(\theta) d\theta}{\int f(x|\theta) g(\theta) d\theta} \\
 &= \frac{\int \theta f(x|\theta) g(\theta) d\theta}{f(x)} \\
 &= \int \theta \pi(\theta|x) d\theta \\
 &= E_{\theta|x}(\theta).
 \end{aligned}$$

L'estimateur bayésien ou la moyenne *a posteriori* est alors noté $E_{\theta|x}(\theta)$.

Voici une proposition très pratique pour les calculs des risques intégrés, qui est une déduction immédiate de la démonstration de la proposition 1.2 .

Proposition 1.3 (Maritz et Lwin (1989))

$$W(\delta) = W(\delta^B) + \int (\delta(x) - \delta^B(x))^2 f(x) dx ,$$

où

$$\delta^B(x) = \frac{\int \theta f(x|\theta)g(\theta)d\theta}{\int f(x|\theta)g(\theta)d\theta} .$$

Donnons un exemple simple d'un tel estimateur bayésien. Soit X une variable aléatoire qui, sachant λ , suit une loi de Poisson avec paramètre λ et λ une variable aléatoire qui suit une loi de gamma de paramètres (α, β) . Dans ce cas on a,

$$f(x|\lambda) = e^{-\lambda} \frac{\lambda^x}{x!}$$

et

$$g(\lambda) = \Gamma_{(\alpha, \beta)}(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} ,$$

où on suppose que $\alpha, \beta (> 0)$ sont connus. Donc

$$\begin{aligned} \pi(\lambda|x) &\propto f(x|\lambda)g(\lambda) \\ &\propto e^{-\lambda} \frac{\lambda^x}{x!} \lambda^{\alpha-1} e^{-\beta\lambda} \\ &\propto \lambda^{x+\alpha-1} e^{-\lambda(\beta+1)} \\ &\propto \Gamma_{(x+\alpha, \beta+1)}(\lambda) . \end{aligned}$$

Des calculs simples donnent

$$\delta^B(x) = E_{\lambda|x}(\lambda) = \frac{x + \alpha}{\beta + 1} .$$

Le risque de cet estimateur est donné par :

$$\begin{aligned}
 W(\delta^B) &= \int_0^\infty \sum_{x=0}^\infty \left(\frac{x+\alpha}{\beta+1} - \lambda\right)^2 e^{-\lambda} \frac{\lambda^x}{x!} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} d\lambda \\
 &= \int_0^\infty \frac{\lambda^{\alpha-1} \beta^\alpha e^{-\beta\lambda}}{\Gamma(\alpha)} \sum_{x=0}^\infty \left(\frac{x+\alpha}{\alpha+1} - \lambda\right)^2 \frac{e^{-\lambda} \lambda^x}{x!} d\lambda \\
 &= \int_0^\infty \frac{\lambda^{\alpha-1} \beta^\alpha e^{-\beta\lambda}}{\Gamma(\alpha)} \frac{1}{(\beta+1)^2} E_X(x + \alpha - \lambda(\beta+1))^2 d\lambda \\
 &= \int_0^\infty \frac{\lambda^{\alpha-1} \beta^\alpha e^{-\beta\lambda}}{\Gamma(\alpha)} \frac{1}{(\beta+1)^2} (\lambda + (\alpha - \beta\lambda))^2 d\lambda \\
 &= \frac{1}{(\beta+1)^2} E_{\Gamma(\alpha, \frac{1}{\beta})} (\lambda + (\alpha - \beta\lambda))^2 \\
 &= \frac{1}{(\beta+1)^2} \left\{ \frac{\alpha}{\beta} + \beta^2 \left(\alpha \frac{1}{\beta^2} \right) \right\} \\
 &= \frac{\alpha}{\beta(\beta+1)}.
 \end{aligned}$$

La fonction marginale qui est très importante dans l'approche bayésienne empirique est donnée par :

$$\begin{aligned}
 f(x) &= \int_0^\infty f(x|\lambda) dG(\lambda) \\
 &= \int_0^\infty \frac{e^{-\lambda} \lambda^x}{x!} \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta\lambda} \lambda^{\alpha-1} d\lambda \\
 &= \frac{1}{x!} \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{-\lambda(\beta+1)} \lambda^{x+\alpha-1} d\lambda \\
 &= \frac{1}{x!} \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(x+\alpha)}{(\beta+1)^{x+\alpha}} \\
 &= \frac{\Gamma(x+\alpha)}{x! \Gamma(\alpha)} \left(\frac{\beta}{\beta+1} \right)^\alpha \left(\frac{1}{\beta+1} \right)^x.
 \end{aligned}$$

Ce qui correspond à une densité de la loi binomiale négative, $x = 0, 1, 2, \dots$ avec

$$E(X) = \frac{\alpha}{\beta} \quad \text{et} \quad \text{Var}(X) = \frac{\alpha(1+\beta)}{\beta^2}.$$

1.3 L'analyse bayésienne hiérarchique

Un des problèmes d'une approche bayésienne classique noté par les fréquentistes est l'incertitude concernant la loi *a priori*. Une approche qui essaie de rectifier ce problème est l'analyse bayésienne hiérarchique qui met des mesures *a priori* sur les paramètres de la loi *a priori* $\pi(\theta)$.

Définition 1.8

On appelle modèle bayésien hiérarchique de niveau n , un modèle statistique bayésien avec densité conditionnelle de X sachant θ , $f(x|\theta)$, la densité *a priori* $\pi(\theta)$ décomposée en n densités conditionnelles et une densité marginale, c'est-à-dire :

$$\begin{aligned} X|\theta &\sim f(x|\theta), \\ \theta|\theta_1 &\sim \pi_1(\theta|\theta_1), \\ \theta_1|\theta_2 &\sim \pi_2(\theta_1|\theta_2), \\ &\vdots \\ \theta_{(n-1)}|\theta_n &\sim \pi_n(\theta_{(n-1)}|\theta_n), \\ \theta_n &\sim \pi_{n+1}(\theta_n). \end{aligned}$$

Les $\theta_1, \theta_2, \dots, \theta_n$ sont appelés hyperparamètres. Par conséquent

$$\pi(\theta) = \int_{\Theta_1 \times \dots \times \Theta_n} \pi_1(\theta|\theta_1) \dots \pi_n(\theta_{(n-1)}|\theta_n) \pi_{n+1}(\theta_n) d\theta_1 \dots d\theta_n.$$

Habituellement, on se limite au cas où $n=1$.

Dans ce cas on a :

$$\begin{aligned}\pi(\theta|x) &= \int \pi(\theta|\theta_1, x)\pi(\theta_1|x)d\theta_1, \\ \pi(\theta|\theta_1, x) &= \frac{f(x|\theta)\pi_1(\theta|\theta_1)}{m_1(x|\theta_1)}, \\ m_1(x|\theta_1) &= \int f(x|\theta)\pi_1(\theta|\theta_1)d\theta, \\ \pi(\theta_1|x) &= \frac{m_1(x|\theta_1)\pi_2(\theta_1)}{m(x)}, \\ m(x) &= \int m_1(x|\theta_1)\pi_2(\theta_1)d\theta_1.\end{aligned}$$

Nous donnons deux méthodes générales pour construire des estimateurs bayésiens hiérarchiques.

La première est la *Méthode de Monte-Carlo* (Robert(2001)). Soient $\theta_1, \dots, \theta_n$, n variables aléatoires indépendantes et identiquement distribuées de fonction de densité $\pi(\theta|x)$. Alors, la moyenne

$$\frac{1}{n} \sum_{i=1}^n g(\theta_i)$$

converge vers

$$\int_{\Theta} g(\theta) \frac{f(x|\theta)\pi(\theta)}{m(x)} d\theta,$$

quand $n \rightarrow \infty$.

Donc

$$E(g(\theta)|x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n g(\theta_i).$$

La deuxième méthode est la *Méthode de Monte-Carlo avec fonction d'importance* (Robert(2001)). Dans le cas où l'on est incapable de générer facilement à partir de $\pi(\theta|x)$, on peut utiliser la fonction d'importance de $h(\theta)$ qui vérifie les conditions suivantes :

- le support de $\pi(\theta|x)$ est inclus dans le support de $h(\theta)$;
- il est facile de générer à partir de $h(\theta)$;
- $h(\theta)$ doit être proche de $\pi(\theta|x)$, soit $w(\theta) = \frac{\pi(\theta)f(x|\theta)}{h(\theta)}$. Alors

$$E(g(\theta)|x)$$

va être approché par :

$$\frac{\sum_{i=1}^n g(\theta_i) w(\theta_i)}{\sum_{i=1}^n w(\theta_i)}$$

– les ailes de $h(\theta)$ doivent être plus élevées que celles de $\pi(\theta|x)$, sinon $\lim_{n \rightarrow \infty} W(\theta) = +\infty$.

1.4 L'analyse bayésienne empirique

L'estimation bayésienne empirique peut être vue comme un complément du modèle hiérarchique : l'estimation bayésienne empirique consiste à estimer la densité *a posteriori* quand les hyperparamètres sont inconnus.

On considère \tilde{x} , un vecteur d'observations indépendantes et identiquement distribuées de vraisemblance $f(x|\theta)$ avec θ un vecteur de paramètres inconnus de fonction de densité $g(\theta|\lambda)$ et λ un vecteur de paramètres. Si λ est connu, alors la densité *a posteriori* de θ est

$$\pi(\theta|x) = \frac{f(x|\theta)g(\theta|\lambda)}{m(x|\lambda)},$$

où $m(x|\lambda)$ est la distribution marginale

$$m(x|\lambda) = \int f(x|\theta)g(\theta|\lambda)d\theta,$$

et on estime θ avec la méthode de Bayes ($E^{\theta|x,\lambda}(\theta)$) ou avec la méthode du maximum de vraisemblance.

Si λ est inconnu, dans l'approche bayésienne on considère la distribution *hyperpriori* de λ , $\pi(\lambda)$, et nous définissons la distribution *a posteriori* de la manière suivante :

$$\begin{aligned} \pi(\theta|x) &= \frac{\int f(x|\theta)g(\theta|\lambda)\pi(\lambda)d\lambda}{\int \int f(x|\theta)g(\theta|\lambda)\pi(\lambda)d\theta d\lambda} \\ &= \int \pi(\theta|x, \lambda)\pi(\lambda|x)d\lambda. \end{aligned}$$

L'approche bayésienne empirique paramétrique (Morris(1983)) consiste à estimer les paramètres λ de la densité marginale $m(x|\lambda)$ en se basant sur les observations (par exemple,

par la méthode de vraisemblance maximale, la méthode des moments, la méthode de minimisation de χ^2 ou par la méthode de l'estimation à noyau de la fonction de densité). Ces différentes méthodes sont développées dans les chapitres 2, 3 et 4, mais nous donnons par la suite, comme exemple, le modèle de Poisson. Nous avons alors :

$$\begin{aligned}x|\theta &\sim \text{Poisson}(\theta); \\ \theta|\lambda &\sim \exp(\lambda), \text{ avec } \lambda \text{ inconnu.}\end{aligned}$$

On veut estimer θ , la moyenne poissonnienne basée sur n observations indépendantes et identiquement distribuées selon $\text{Poisson}(\theta)$. La densité marginale est donc :

$$\begin{aligned}m(x_i|\lambda) &= \int_0^\infty e^{-\theta} \frac{\theta^{x_i}}{x_i!} \lambda \exp(-\theta\lambda) d\theta \\ &= \frac{\lambda}{(1+\lambda)^{x_i}} \\ &= \left(\frac{1}{1+\lambda}\right)^{x_i} \frac{\lambda}{1+\lambda}.\end{aligned}$$

Donc $x_i|\lambda$ est une géométrique de moyenne $\frac{1}{\lambda}$.

Le logarithme de la fonction de vraisemblance pour les n premières observations (x_1, x_2, \dots, x_n) est donné par :

$$\begin{aligned}\log\left(\prod_{i=0}^n m(x_i|\lambda)\right) &= \log\left(\left(\frac{1}{1+\lambda}\right)^{\sum x_i} \left(\frac{\lambda}{1+\lambda}\right)^n\right) \\ &= -\sum x_i \log(1+\lambda) + n\left(\log\left(\frac{\lambda}{1+\lambda}\right)\right).\end{aligned}$$

On a que :

$$\begin{aligned}\frac{\partial}{\partial \lambda} \log\left(\prod_{i=0}^n m(x_i|\lambda)\right) &= \frac{-\sum x_i}{1+\lambda} + \frac{n}{\lambda} - \frac{n}{1+\lambda} \\ &= 0;\end{aligned}$$

ce qui est équivalent à

$$\hat{\lambda}(\bar{x}) = \frac{1}{\bar{x}}.$$

Sachant que $\lambda = \widehat{\lambda}(\tilde{x}) = \frac{1}{\tilde{x}}$ et \tilde{x} , la distribution *a posteriori* est estimée par :

$$\begin{aligned}\pi(\theta|\widehat{\lambda}, x_{n+1}) &\propto (e^{-\theta} \frac{\theta^{x_{n+1}}}{x_{n+1}!}) \widehat{\lambda} \exp(-\theta \widehat{\lambda}) \\ &\propto \theta^{x_{n+1}} \exp(-\theta(\widehat{\lambda} + 1)) \\ &= \Gamma(x_{n+1} + 1, \widehat{\lambda} + 1).\end{aligned}$$

L'estimateur bayésien empirique paramétrique de θ est donné par :

$$\begin{aligned}E^{\theta|\widehat{\lambda}, x_{n+1}}(\theta) &= \frac{x_{n+1} + 1}{\widehat{\lambda} + 1} \\ &= \frac{(x_{n+1} + 1)\bar{x}}{1 + \bar{x}}.\end{aligned}$$

Une autre façon de résoudre la difficulté de l'incertitude dans la mesure *a priori* est d'utiliser une approche bayésienne empirique de Robbins (1955) qui est essentiellement non paramétrique et que l'on appelle estimation bayésienne empirique non paramétrique.

Considérons à présent le cas suivant. Les observations passées sont : x_1, x_2, \dots, x_n . L'observation présente est x_{n+1} . Intuitivement, on peut écrire l'estimateur bayésien empirique de la façon suivante : on suppose $n(x_{n+1})$ le nombre de fois où on a observé x_{n+1} parmi x_1, x_2, \dots, x_n .

Alors,

$$f(x_{n+1}) \text{ est estimé par } \frac{n(x_{n+1})}{n},$$

et si on inclut l'observation courante alors

$$f(x) \text{ est estimé par } \frac{1 + n(x_{n+1})}{n + 1},$$

et

$$f(x_{n+1} + 1) \text{ est estimé par } \frac{n(x_{n+1} + 1)}{n + 1},$$

d'où l'estimateur bayésien empirique de θ :

$$\delta(x_{n+1}) = \frac{(x_{n+1} + 1)n(x_{n+1} + 1)}{1 + n(x_{n+1})}.$$

Plus précisément, dans le cas précédent de l'estimation d'une moyenne de la loi de Poisson, la loi marginale $f(x)$ est égale à :

$$f(x) = \int_0^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} g(\lambda) d\lambda.$$

Si l'on suppose que x_{n+1} suit une loi de Poisson de paramètre λ_{n+1} , l'estimateur bayésien classique de λ_{n+1} est :

$$\begin{aligned}\delta^B(x_{n+1}) &= E(\lambda|x_{n+1}) \\ &= \frac{\int_0^\infty e^{-\lambda} \frac{\lambda^{x_{n+1}+1}}{x_{n+1}!} g(\lambda) d\lambda}{\int_0^\infty e^{-\lambda} \frac{\lambda^{x_{n+1}}}{x_{n+1}!} g(\lambda) d\lambda} \\ &= \frac{f(x_{n+1} + 1)}{f(x_{n+1})} (x_{n+1} + 1).\end{aligned}$$

Donc, l'estimateur bayésien empirique de λ_{n+1} est donné par :

$$\delta_n(x_{n+1}) = \frac{(1 + x_{n+1})n(1 + x_{n+1})}{1 + n(x_{n+1})}$$

Le théorème suivant (Robert(2001)) montre le lien asymptotique entre les méthodes bayésienne empirique et bayésienne hiérarchique.

Théorème 1.4 (Robert (2001))

Soit X une variable aléatoire de densité conditionnelle $f(x|\theta)$ alors :

$$E^{\pi_2(\lambda|x)}[E^{\pi_1(\theta|\lambda,x)}(\theta)] = E^{\pi_1(\theta|\hat{\lambda},x)}(\theta) + o(n^{-1}),$$

où $\hat{\lambda} = \operatorname{argmax} m(x|\lambda)$.

1.5 Le lien entre les estimateurs bayésiens empiriques et les estimateurs du type James-Stein pour la loi multinomiale

Une conséquence de la proposition 1.1 est que s'il existe un seul estimateur minimax, il est admissible (Robert 1992). Réciproquement, si un estimateur minimax δ_0 du risque constant n'est pas admissible, il existe d'autres estimateurs minimax δ qui le dominent uniformément, c'est-à-dire, que $R_\delta(\theta) \leq R_{\delta_0}(\theta)$ pour tout θ de Θ et l'inégalité est stricte pour au moins une valeur de θ (sous quelques conditions de régularité, voir Brown, (1976)).

Avant 1955, on pensait que si $\tilde{X} \sim N(\theta, I)$, alors l'estimateur des moindres carrés $\delta_0(\tilde{X}) = \tilde{X}$ était admissible, puisque c'est le seul estimateur minimax de risque constant.

Stein (1955) a montré que ce résultat ne peut être vrai pour un vecteur de plus de deux composantes. Plus précisément, le risque de l'estimateur du type James-Stein

$$\delta^{JS}(\tilde{X}) = \left(1 - \frac{p-2}{\|\tilde{X}\|^2}\right)\tilde{X}$$

est plus petit ou égal au risque de $\delta_0(\tilde{X}) = \tilde{X}$ pour tout θ et $R_{\delta^{JS}}(\bar{0}) < R_{\delta_0}(\bar{0})$.

Si l'on suppose que $\tilde{\theta} \sim N(\tilde{\theta}, \tau^2 I)$, la loi marginale de \tilde{X} est alors $N(\tilde{\theta}, (1 + \tau^2)I)$.

L'estimateur du maximum de vraisemblance de τ^2 est :

$$\begin{cases} \frac{\|\tilde{X}\|^2}{p} - 1 & \text{si } \|\tilde{X}\|^2 > p, \\ 0 & \text{sinon.} \end{cases}$$

donc l'estimateur de Bayes empirique de θ est :

$$\begin{aligned} \delta^E(\tilde{X}) &= \frac{\hat{\tau}^2 \tilde{X}}{1 + \hat{\tau}^2} \\ &= \left(1 - \frac{p}{\|\tilde{X}\|^2}\right)^+ \tilde{X}, \end{aligned}$$

qui est donc un estimateur du type James-Stein tronqué (Roberts 2001).

CHAPITRE II

LOI NORMALE

Dans ce chapitre, on considère le problème d'estimation de la moyenne ou du vecteur moyen d'une loi gaussienne. On utilise l'approche bayésienne empirique décrite dans Casella (1985) et Efron et Morris (1977). Aussi, étant donné que l'on s'intéresse aussi à la robustesse des estimateurs, nous introduisons un estimateur robuste, un analogue de celui d'Angers(1992) dans le cas de l'analyse bayésienne. On compare son comportement avec celui de Efron et Morris (1977). On va considérer deux modèles bayésiens empiriques : le premier selon la loi *a priori* normale et le deuxième selon la loi de Student. Finalement, nous considérons le problème des échantillons non équilibrés.

2.1 Estimateurs bayésiens empiriques du vecteur moyen de la loi normale

En suivant l'approche de Casella (1985), on suppose qu'on a p observations de variables aléatoires de moyennes différentes, mais de même variance, c'est-à-dire que :

$$X_i \sim \mathcal{N}(\theta_i, \sigma^2), \quad \text{pour } i = 1, \dots, p. \quad (2.1)$$

Les résultats suivants sont très bien connus mais nous donnons la démonstration pour faciliter la lecture du mémoire.

Lemme 2.1

Si $X_i \sim \mathcal{N}(\theta_i, \sigma^2)$ et θ_i , pour $i = 1, \dots, p$, a pour la loi *a priori* la loi $\mathcal{N}(\mu, \tau^2)$, pour les constantes μ et σ , alors l'estimateur bayésien, $\delta^B(X_i)$, est :

$$\delta^B(X_i) = [\sigma^2/(\sigma^2 + \tau^2)]\mu + [\tau^2/(\sigma^2 + \tau^2)]X_i. \quad (2.2)$$

Démonstration :

La loi *a priori* de θ_i sachant X_i , $\pi(\theta_i|X_i)$, est proportionnelle à :

$$\begin{aligned} \pi(\theta_i|X_i) &\propto \exp\left(-\frac{1}{2}\left(\frac{X_i - \theta_i}{\sigma}\right)^2 - \frac{1}{2}\left(\frac{\theta_i - \mu}{\tau}\right)^2\right) \\ &\propto \exp\left(-\frac{1}{2}\left(\frac{\theta_i^2}{\sigma^2} + \frac{\theta_i^2}{\tau^2} - 2\frac{X_i\theta_i}{\sigma^2} - 2\frac{\mu\theta_i}{\tau^2}\right)\right) \\ &\propto \exp\left(-\frac{1}{2}\left(\frac{\theta_i^2}{\alpha^2} - 2\theta_i\left(-\frac{X_i}{\sigma^2} + \frac{\mu}{\tau^2}\right)\right)\right) \\ &\propto \exp\left(-\frac{1}{2}\frac{1}{\alpha^2}\left(\theta_i^2 - 2\theta_i\alpha^2\left(\frac{X_i}{\sigma^2} + \frac{\mu}{\tau^2}\right)\right)\right) \\ &\propto \exp\left(-\frac{1}{2\alpha^2}\left(\theta_i - \alpha^2\left(\frac{X_i}{\sigma^2} + \frac{\mu}{\tau^2}\right)\right)^2\right), \end{aligned}$$

avec $\frac{1}{\alpha^2} = \frac{1}{\sigma^2} + \frac{1}{\tau^2}$, or

$$\delta^B(X_i) = E_{\theta_i|X_i}(\theta_i).$$

Soit alors

$$\begin{aligned} \delta^B(X_i) &= \alpha^2\left(\frac{X_i}{\sigma^2} + \frac{\mu}{\tau^2}\right) \\ \delta^B(X_i) &= \frac{\alpha^2}{\sigma^2}X_i + \frac{\alpha^2}{\tau^2}\mu \\ \delta^B(X_i) &= \frac{\tau^2}{\sigma^2 + \tau^2}X_i + \frac{\sigma^2}{\sigma^2 + \tau^2}\mu. \end{aligned}$$

Déduisons aussi que cette loi *a priori* de θ_i sachant X_i est

$$\pi(\theta_i|X_i) \sim N\left(\delta^B(X_i), \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}\right).$$

De même, la distribution de la fonction marginale f de X est déduite de la façon

suivante :

$$\begin{aligned}
f_X(x) &= \int f(x|\theta)\pi(\theta)d\theta \\
&\propto \int e^{-\frac{1}{2\sigma^2}(x-\theta)^2} e^{-\frac{1}{2\tau^2}(\theta-\mu)^2} d\theta \\
&\propto \int e^{-\frac{1}{2}\left(\frac{x^2}{\sigma^2} - \frac{2x\theta}{\sigma^2} + \frac{\theta^2}{\sigma^2}\right)} e^{-\frac{1}{2}\left(\frac{\theta^2}{\tau^2} - \frac{2\mu\theta}{\tau^2} + \frac{\mu^2}{\tau^2}\right)} d\theta \\
&\propto \int e^{-\frac{1}{2}\left(\frac{x^2}{\sigma^2} + \theta^2\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right) - 2\theta\left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right)\right)} d\theta \\
&\propto e^{-\frac{1}{2}\left(\frac{x^2}{\sigma^2}\right)} e^{\frac{1}{2}\alpha^2\left(\frac{1}{\alpha^2}\left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right)\right)^2} \int e^{-\frac{1}{2}\alpha^2\left(\theta - \frac{1}{\alpha^2}\left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right)\right)^2} d\theta \\
&\propto e^{-\frac{1}{2}\left(\frac{x^2}{\sigma^2}\right)} e^{\frac{1}{2}\alpha^2\left(\frac{1}{\alpha^2}\left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right)\right)^2} \\
&\propto \exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2} + \frac{1}{2}\frac{1}{\alpha^2}\left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right)^2\right) \\
&\propto \exp\left(-\frac{1}{2}\left(x^2\left(\frac{1}{\sigma^2} - \frac{1}{\alpha^2\sigma^2}\right) - \frac{2x\mu}{\alpha^2\sigma^2\tau^2}\right)\right) \\
&\propto \exp\left(-\frac{1}{2}\left(x^2a - \frac{2x\mu}{\alpha^2\sigma^2\tau^2}\right)\right) \quad \text{où } a = \frac{1}{\sigma^2 + \tau^2}, \\
&\propto \exp\left(-\frac{1}{2}a\left(x - \frac{2x\mu}{a\alpha^2\sigma^2\tau^2}\right)^2\right).
\end{aligned}$$

Nous remarquons que

$$a\alpha^2\sigma^2\tau^2 = 1.$$

On obtient alors

$$f_X(x) \propto \exp\left(-\frac{1}{2}a(x - \mu)^2\right). \quad (2.3)$$

Par conséquent, f_X est la fonction de densité d'une normale $\mathcal{N}(\mu, \sigma^2 + \tau^2)$.

2.2 Estimation bayésienne empirique

Sous ces hypothèses et ces résultats, on construit alors l'estimateur bayésien pour la quantité (2.3) associée. Pour cela on a besoin du résultat suivant :

Lemme 2.1 Suivant les hypothèses et les résultats précédents, on a :

$$E(\bar{X}_i) = \mu, \quad (2.4)$$

$$\text{et} \quad E\left(\frac{(p-3)\sigma^2}{\sum_{i=1}^p (X_i - \bar{X})^2}\right) = \frac{\sigma^2}{\sigma^2 + \tau^2}. \quad (2.5)$$

Démonstration :

On a évidemment (2.4). Pour le deuxième résultat on pose

$$Y = \frac{\sum_1^p (X_i - \bar{X})^2}{\sigma^2 + \tau^2}, \quad (2.6)$$

dont la loi est χ_{p-1}^2 et la fonction de densité

$$f_Y(y) = \frac{1}{\Gamma(\frac{p-1}{2}) 2^{\frac{p-1}{2}}} y^{\frac{p-1}{2}-1} e^{-\frac{y}{2}} \quad \text{où } 0 \leq y \leq \infty, \quad (2.7)$$

d'où

$$\begin{aligned} E\left(\frac{1}{Y}\right) &= \int_0^\infty \frac{1}{y} f_Y(y) dy \\ &= \frac{1}{\Gamma(\frac{p-1}{2}) 2^{\frac{p-1}{2}}} \int_0^\infty y^{\frac{p-1}{2}-2} e^{-\frac{y}{2}} dy \\ &= \frac{\Gamma(\frac{p-3}{2}) 2^{\frac{p-3}{2}}}{\Gamma(\frac{p-1}{2}) 2^{\frac{p-1}{2}}} \\ &= \frac{1}{p-3}. \end{aligned}$$

Soit alors

$$E\left(\frac{(p-3)\sigma^2}{\sum_1^p (X_i - \bar{X})^2}\right) = \frac{\sigma^2}{\sigma^2 + \tau^2}. \quad (2.8)$$

Par conséquent, de ce lemme 2.1 et l'estimation sans biais de $\delta^B(X_i)$, (voir la formule (2.2)), on peut construire alors l'estimateur bayésien empirique :

Définition 2.1

Un estimateur bayésien empirique, $\delta_i^E(\mathbf{X})$, de θ_i est défini comme suit :

$$\delta_i^E(\mathbf{X}) = \left[\frac{(p-3)\sigma^2}{\sum_1^p (X_i - \bar{X})^2} \right] \bar{X} + \left[1 - \frac{(p-3)\sigma^2}{\sum_1^p (X_i - \bar{X})^2} \right] X_i. \quad (2.9)$$

Casella (1985) a montré qu'on obtient alors avec l'estimateur bayésien empirique un risque quadratique meilleur qu'avec l'estimateur usuel, en se basant sur des exemples de base des données. ce qui amène à énoncer que :

Proposition 2.1

Soit δ_i^E est un estimateur bayésien empirique de θ_i , pour tout $p \geq 4$, alors

$$E\left(\sum_1^p [\theta_i - \delta_i^E(\mathbf{X})]^2\right) < E\left(\sum_1^p [\theta_i - X_i]^2\right) \quad \text{pour tout } \theta_i \quad (2.10)$$

On applique le lemme suivant, démontré dans Casella et Berger (1990), pour la preuve de (2.10).

Lemme 2.2 (Lemme de Stein)

Si $X \sim \mathcal{N}(\theta, \sigma^2)$ et g dérivable telle que $E(|g'(x)|) < \infty$, alors

$$E[g(x)(x - \theta)] = \sigma^2 E g'(x) \quad (2.11)$$

Démonstration :

Démontrons alors la proposition précédente. En utilisant (2.9), l'erreur quadratique moyenne de θ_i est alors égale à :

$$\begin{aligned} E\left\{\sum_1^p [\theta_i - \delta_i^E(\mathbf{X})]^2\right\} &= E\left\{\sum_1^p \left([\theta_i - X_i] - \left[\frac{(p-3)\sigma^2}{\sum_1^p (X_i - \bar{X})^2} \right] (\bar{X} - X_i) \right)^2\right\} \\ &= E\left\{\sum_1^p [(\theta_i - X_i)^2] + (p-3)^2 \sigma^4 \times E\left[\frac{1}{\sum_1^p (X_i - \bar{X})^2}\right] \right. \\ &\quad \left. - \underbrace{2(p-3)\sigma^2 \times E\left\{\sum_1^p \frac{(X_i - \theta_i)(X_i - \bar{X})}{\sum_1^p (X_i - \bar{X})^2}\right\}}_{(\text{Terme } *)} \right\}. \end{aligned}$$

Maintenant, on utilise le lemme 2.2 pour le (Terme *)

$$(\text{Terme } *) = 2\sigma^4(p-3) \sum_1^p E\left\{\frac{\partial}{\partial X_i} \left(\frac{\bar{X} - X_i}{\sum_1^p (X_i - \bar{X})^2} \right)\right\}, \quad (2.12)$$

et du fait que

$$\frac{\partial}{\partial X_i} \sum_1^p (X_i - \bar{X})^2 = \frac{\partial}{\partial X_i} \sum_1^p X_i^2 - p\bar{X}^2 = 2(X_i - \bar{X}), \quad (2.13)$$

ce terme donne

$$(\text{Terme } *) = \sigma^4(p-3) \sum_1^p E\left\{\frac{\left(\frac{1}{p} - 1\right) \sum_1^p (X_i - \bar{X})^2 + 2(X_i - \bar{X})^2}{\left(\sum_1^p (X_i - \bar{X})^2\right)^2}\right\}. \quad (2.14)$$

Soit aussi

$$\begin{aligned} (\text{Terme } *) &= \sigma^4(p-3)E\left\{\frac{1-p+2}{\sum_1^p(X_i - \bar{X})^2}\right\} \\ &= -2\sigma^4(p-3)^2E\left\{\frac{1}{\sum_1^p(X_i - \bar{X})^2}\right\}. \end{aligned}$$

De cela on déduit :

$$E\left\{\sum_1^p[\theta_i - \delta_i^E(\mathbf{X})]^2\right\} - E\left\{\sum_1^p(\theta_i - X_i)^2\right\} = -\sigma^4(p-3)^2E\left\{\frac{1}{\sum_1^p(X_i - \bar{X})^2}\right\}, \quad (2.15)$$

qui est négatif et on déduit alors le résultat voulu. ■

2.3 Proposition d'un modèle robuste.

Dans l'approche bayésienne selon Berger (1985), il peut être intéressant de rendre les estimateurs plus robustes. Angers (1992) a proposé d'utiliser la loi de Student t comme loi *a priori* dans le problème d'estimation bayésienne d'une moyenne gaussienne. Nous proposons ici d'adapter cette idée de Angers (1992) à l'approche bayésienne empirique.

Plus particulièrement,

$$\begin{cases} X_i|\theta_i \sim N(\theta_i, \sigma^2), \\ \theta_i \sim T_\alpha(\mu, \tau^2). \end{cases} \iff \begin{cases} X_i|\theta_i \sim N(\theta_i, \sigma^2), \\ \theta_i \sim N(\mu, \tau^2/\lambda_i), \\ \lambda_i \sim \Gamma(\alpha/2, \alpha/2). \end{cases}$$

On cherche les estimateurs de vraisemblance maximale de μ et τ^2 . On a que :

$$\begin{aligned} m(x_i|\mu, \tau^2, \lambda) &= \int_{\Theta} f(x_i|\theta_i)\pi(\theta_i|\mu, \tau^2)d\theta_i \\ &\sim N(\mu, \sigma^2 + \tau^2/\lambda), \end{aligned}$$

et

$$m(x_i|\mu, \tau^2) = \int_0^\infty \frac{1}{\sqrt{2\pi a}} e^{-\frac{1}{2a}(x_i-\mu)^2} \Gamma(\lambda) d\lambda \quad \text{où} \quad a = \sigma^2 + \frac{\tau^2}{\lambda} \quad \text{et} \quad \Gamma(\lambda) = \frac{(\alpha/2)^{\alpha/2}}{\Gamma(\alpha/2)} \lambda^{\alpha/2-1} e^{-\alpha\lambda/2}.$$

La fonction de vraisemblance est à un constant près :

$$l(x_1, x_2, \dots, x_p) = \prod_{i=1}^p m(x_i|\mu, \tau^2),$$

dont

$$\begin{aligned} \frac{\partial}{\partial \mu} \log(l(x_1, x_2, \dots, x_p)) &= \sum_{i=1}^p \frac{1}{m(x_i|\mu, \tau^2)} \frac{\partial}{\partial \mu} m(x_i|\mu, \tau^2) \\ &= \sum_{i=1}^p \frac{1}{m(x_i|\mu, \tau^2)} \frac{\partial}{\partial \mu} \int_0^\infty \frac{1}{\sqrt{a}} e^{-\frac{1}{2a}(x_i-\mu)^2} \Gamma(\lambda) d\lambda. \end{aligned}$$

On a

$$\frac{\partial}{\partial \mu} \frac{1}{\sqrt{a}} e^{-\frac{1}{2a}(x_i-\mu)^2} \Gamma(\lambda) = \frac{1}{a\sqrt{a}} (x_i - \mu) e^{-\frac{1}{2a}(x_i-\mu)^2} \Gamma(\lambda).$$

On montre que :

$$\frac{\partial}{\partial \mu} \int_0^\infty \frac{1}{\sqrt{a}} e^{-\frac{1}{2a}(x_i-\mu)^2} \Gamma(\lambda) d\lambda = \int_0^\infty \frac{\partial}{\partial \mu} \frac{1}{\sqrt{a}} e^{-\frac{1}{2a}(x_i-\mu)^2} \Gamma(\lambda) d\lambda.$$

Il suffit de montrer que les conditions du théorème de convergence dominée de Lebesgue s'appliquent.

Théorème de convergence dominée de Lebesgue (Rudin(1991))

Si $f(x, \theta)$ est dérivable en θ et $\exists g(x, \theta)$, telle que :

$$\begin{cases} |f(x, \theta)| \leq g(x, \theta) \\ \int_{-\infty}^{\infty} g(x, \theta) dx < \infty, \end{cases}$$

alors

$$\frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} f(x, \theta) dx = \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} f(x, \theta) dx.$$

En effet, dans notre cas $a = \sigma^2 + \tau^2/\lambda$, ce qui donne

$$\frac{1}{a^{3/2}} \leq \frac{1}{(\sigma^2)^{3/2}} = \frac{1}{\sigma^3} \quad \text{et} \quad e^{-\frac{1}{2a}(x_i-\mu)^2} \leq 1.$$

Ainsi,

$$\left| -\frac{(x_i - \mu)}{a^{3/2}} e^{-\frac{1}{2a}(x_i-\mu)^2} \Gamma(\lambda) \right| \leq \frac{|x_i - \mu|}{\sigma^3} \Gamma(\lambda),$$

et

$$\begin{aligned} \int_0^\infty \frac{|x_i - \mu|}{\sigma^3} \Gamma(\lambda) d\lambda &= \frac{|x_i - \mu|}{\sigma^3} \int_0^\infty \Gamma(\lambda) d\lambda \\ &= \frac{|x_i - \mu|}{\sigma^3} < \infty, \end{aligned}$$

ce qu'il fallait démontrer.

Donc,

$$\begin{aligned} \frac{\partial}{\partial \mu} \log(l(x_1, x_2, \dots, x_p)) &= \sum_{i=1}^p \frac{1}{m(x_i|\mu, \tau^2)} \int_0^\infty -\frac{x_i - \mu}{a^{3/2}} e^{-\frac{1}{2a}(x_i - \mu)^2} \Gamma(\lambda) d\lambda, \\ &= 0, \end{aligned}$$

ce qui implique que

$$\mu = A/B, \quad (2.16)$$

où

$$\begin{cases} A = \sum_{i=1}^p \frac{x_i}{m(x_i|\mu, \tau^2)} \int_0^\infty -\frac{1}{a^{3/2}} e^{-\frac{1}{2a}(x_i - \mu)^2} \Gamma(\lambda) d\lambda, \\ B = \sum_{i=1}^p \frac{1}{m(x_i|\mu, \tau^2)} \int_0^\infty -\frac{1}{a^{3/2}} e^{-\frac{1}{2a}(x_i - \mu)^2} \Gamma(\lambda) d\lambda. \end{cases}$$

On montre de même que :

$$\frac{\partial}{\partial \tau^2} \int_0^\infty \frac{1}{\sqrt{a}} e^{-\frac{1}{2a}(x_i - \mu)^2} \Gamma(\lambda) d\lambda = \int_0^\infty \frac{\partial}{\partial \tau^2} \left(\frac{1}{\sqrt{a}} e^{-\frac{1}{2a}(x_i - \mu)^2} \Gamma(\lambda) \right) d\lambda.$$

On a :

$$\frac{\partial}{\partial \tau^2} a = \frac{1}{\lambda} \quad \text{et} \quad \frac{\partial}{\partial \tau^2} \frac{1}{\sqrt{a}} = -\frac{a'}{a^{3/2}} = -\frac{1}{2\lambda a^{3/2}}.$$

Alors

$$\begin{aligned} \frac{\partial}{\partial \tau^2} \left(\frac{1}{\sqrt{a}} e^{-\frac{1}{2a}(x_i - \mu)^2} \Gamma(\lambda) \right) &= -\frac{1}{2\lambda a^{3/2}} e^{-\frac{1}{2a}(x_i - \mu)^2} \Gamma(\lambda) \\ &\quad + \frac{1}{2\lambda a^{5/2}} (x_i - \mu)^2 e^{-\frac{1}{2a}(x_i - \mu)^2} \Gamma(\lambda) \\ &= \left[-\frac{1}{2a^{3/2}} + \frac{1}{2a^{5/2}} \right] \frac{1}{\lambda} e^{-\frac{1}{2a}(x_i - \mu)^2} \Gamma(\lambda). \end{aligned}$$

Ce dernier terme en valeur absolue est inférieur à

$$\left(\frac{1}{2\sigma^3} + \frac{1}{2\sigma^5} \right) \frac{1}{\lambda} \Gamma(\lambda),$$

et

$$\int_0^\infty \frac{1}{\lambda} \Gamma(\lambda) d\lambda \propto \int_0^\infty \Gamma(\alpha/2 - 1, \alpha/2) d\lambda < \infty.$$

Puisque $\alpha/2 > 1$, alors le théorème de convergence dominée de Lebesgue est applicable, d'où

$$\frac{\partial}{\partial \tau^2} \int_0^\infty \frac{1}{\sqrt{a}} e^{-\frac{1}{2a}(x_i-\mu)^2} \Gamma(\lambda) d\lambda = \int_0^\infty \frac{\partial}{\partial \tau^2} \frac{1}{\sqrt{a}} e^{-\frac{1}{2a}(x_i-\mu)^2} \Gamma(\lambda) d\lambda,$$

et

$$\begin{aligned} \frac{\partial}{\partial \tau^2} l(x_1, x_2, \dots, x_p) &= -\sum_1^p \frac{1}{m(x_i|\mu, \tau^2)} \int_0^\infty -\frac{1}{2\lambda a^{3/2}} e^{-\frac{1}{2a}(x_i-\mu)^2} \Gamma(\lambda) d\lambda \\ &\quad + \sum_1^p \frac{1}{m(x_i|\mu, \tau^2)} \int_0^\infty -\frac{1}{2\lambda a^{5/2}} (x_i - \mu)^2 e^{-\frac{1}{2a}(x_i-\mu)^2} \Gamma(\lambda) d\lambda \\ &= -C + \frac{1}{\tau^2} D \quad (\text{disons}) \\ &= 0. \end{aligned}$$

Soit donc :

$$\tau^2 = \frac{D}{C}, \quad (2.17)$$

où

$$\begin{cases} C = \sum_{i=1}^p \frac{x_i}{m(x_i|\mu, \tau^2)} \int_0^\infty -\frac{1}{2\lambda a^{3/2}} e^{-\frac{1}{2a}(x_i-\mu)^2} \Gamma(\lambda) d\lambda, \\ D = \tau^2 \sum_{i=1}^p \frac{1}{m(x_i|\mu, \tau^2)} \int_0^\infty -\frac{1}{2\lambda a^{5/2}} e^{-\frac{1}{2a}(x_i-\mu)^2} \Gamma(\lambda) d\lambda. \end{cases}$$

On évalue les intégrales avec la méthode de Monte-Carlo et on estime $\hat{\mu}$ et $\hat{\tau}^2$ de la manière suivante (voir programme (A.0.10)). On peut démontrer que, sauf pour un ensemble de mesure de Lebesgue nulle, $m(x_i|\mu, \tau^2)$ est une fonction continue qui tend vers zéro quand $\mu \rightarrow \pm\infty$ et $\tau \rightarrow 0$ ou $\tau \rightarrow \infty$. Donc, sur un ensemble compact, $m(x_i|\mu, \tau^2)$ a un maximum. On a choisi le rectangle $[0.1, 0.5] \times [0.0001, 0.3]$ et sur le rectangle on a fait une recherche fine en utilisant une méthode itérative de calcul. Si l'on varie la valeur initiale de τ dans ce rectangle on trouve les mêmes valeurs $\hat{\mu} = .2857157$, $\hat{\tau}^2 = 0.000593$ après 10 itérations. Ces valeurs donnent un maximum local ; donc on a trouvé un estimateur de vraisemblance maximale.

Pour λ , on l'estime en utilisant le résultat que

$$E\left[\frac{(p-3)\sigma^2}{\sum(x_i - \bar{X})^2}\right] = \frac{\sigma^2}{\sigma^2 + \tau^2/\lambda},$$

et en posant que :

$$\frac{(p-3)\sigma^2}{\sum(x_i - \bar{X})^2} = \frac{\sigma^2}{\sigma^2 + \hat{\tau}^2/\hat{\lambda}}.$$

On construit alors un estimateur de θ_i , noté $\delta_i^{E2}(\mathbf{X})$, qui est égal à :

$$\begin{aligned} \delta_i^{E2}(\mathbf{X}) &= [\sigma^2/(\sigma^2 + \hat{\tau}^2/\hat{\lambda})]\hat{\mu} + [(\hat{\tau}^2/\hat{\lambda})/(\sigma^2 + \hat{\tau}^2/\hat{\lambda})]X_i \\ &= [\hat{\lambda}\sigma^2/(\hat{\lambda}\sigma^2 + \hat{\tau}^2)]\hat{\mu} + [\hat{\tau}^2/(\hat{\lambda}\sigma^2 + \hat{\tau}^2)]X_i \\ &= \left[\frac{(p-3)\sigma^2}{\sum_1^p(X_i - \bar{X})^2}\right]\hat{\mu} + \left[1 - \frac{(p-3)\sigma^2}{\sum_1^p(X_i - \bar{X})^2}\right]X_i, \end{aligned} \quad (2.18)$$

2.4 Application.

Dans notre exemple on utilise les données de baseball utilisées par Casella (1985). Ce dernier considère 7 joueurs parmi 18 : le meilleur, le pire et 5 choisis aléatoirement afin d'estimer la performance de chaque joueur à la fin de l'année :

- X_i = la performance du i ième joueur au milieu de la saison ;
- θ_i = la moyenne finale de la saison au bâton ;
- δ_i^E = l'estimateur bayésien empirique du modèle de Casella (1985) ;
- δ_i^{E2} = l'estimateur bayésien empirique avec le nouveau modèle utilisant la loi de Student t avec 25 degrés de liberté, choisie parce que cette loi est assez proche à la loi gaussienne *a priori* et on veut comparer les deux résultats ;
- $EQMP$ est l'erreur quadratique moyenne pondérée (pondérée par la variance).

On a alors $\sigma^2 = 0.0659^2$ et le terme $\frac{4\sigma^2}{\sum_1^p(x_i - \bar{x})^2} = 0.495$ et $\hat{\mu} = .2857157$. Pour l'estimateur proposé, on trouve

$$\delta_i^{E2} = (.495)(.2857157) + .505X_i ,$$

et l'estimateur bayésien empirique du modèle de Casella avec la loi *a priori* gaussienne est :

$$\delta_i^E = (.495)(.286) + .505X_i .$$

Tableau 2.1 Estimation bayésienne empirique pour les 7 joueurs

Joueur	X_i	θ_i	δ_i^E	δ_i^{E2}	$(x_i - \theta_i)^2$	$(\delta_i^E - \theta_i)^2$	$(\delta_i^{E2} - \theta_i)^2$
1	0,395	0,346	0,341	0,340904	0,002401	0,000025	0,000026
2	0,355	0,279	0,321	0,320704	0,005776	0,001764	0,0017392
3	0,313	0,276	0,3	0,299494	0,001369	0,000576	0,000552
4	0,291	0,266	0,289	0,288384	0,000625	0,000529	0,0005011
5	0,247	0,271	0,26	0,266164	0,000576	0,000121	0,0000234
6	0,224	0,266	0,255	0,254549	0,001764	0,000121	0,0001311
7	0,175	0,318	0,23	0,229804	0,020449	0,007744	0,0077785
Total					0,03296	0,01088	0,010751
<i>EQMP</i>	1,08422	0	0,03579	0,0.35365			

$$\text{Avec } EQMP = \sum_{i=1}^7 \frac{(\delta_i - \theta_i)^2}{7\sigma^2} \text{ et } \sigma^2 = 0.0659^2.$$

Le tableau 2.1 illustre les différents estimateurs pour chaque joueur. L'erreur quadratique moyenne pondérée pour le nouvel estimateur est légèrement meilleure. Étant donné qu'il y a très peu de différence entre les deux estimateurs dans cet exemple, nous avons introduit une valeur aberrante comme Angers et MacGibbon (1996) ont fait dans le cas de la régression gaussienne.

On suppose l'existence d'une valeur aberrante, soit $\bar{x} + 2\sigma$ et on remplace la performance du 6^e joueur (choisi aléatoirement) par 0,4175. La moyenne devient 0,3134 et la valeur du terme de la pondération $\frac{4\sigma^2}{\sum_1^7 (x_i - \bar{x})^2} = 0,4013$. L'estimateur bayésien empirique avec le nouveau modèle est calculé selon différents degrés de liberté de la loi de Student t . On commence avec le nombre de degrés de liberté = 25 qui donne une loi de Student t très proche de la loi gaussienne. Selon les résultats d'Angers et MacGibbon (1996), on espère avoir la meilleure amélioration avec le nombre de degrés de liberté petit ; donc on a choisi le nombre=10, 4, et 2. Dans le tableau 2.2, δ_i^{E2} désigne l'estimateur bayésienne

Tableau 2.2 Estimation bayésienne empirique pour les 7 joueurs avec présence d'une valeur aberrante en utilisant la loi de Student t comme mesure *a priori*

Joueur	$\delta_i^{E_2} dl = 25$	$\delta_i^{E_2} dl = 10$	$\delta_i^{E_2} dl = 4$	$\delta_i^{E_2} dl = 2$
1	0,361695	0,360892	0,348854	0,340829
2	0,337746	0,336943	0,324905	0,316880
3	0,312599	0,311796	0,299758	0,291733
4	0,299427	0,298624	0,286586	0,278561
5	0,273082	0,272280	0,260242	0,252216
6	0,375166	0,374364	0,362326	0,354300
7	0,229974	0,229171	0,217133	0,209108
<i>EQMP</i>	0,003690	0,003650	0,003240	0,003120

empirique avec le nouveau modèle et dl représente le degré de liberté pour la loi de Student t .

D'après les résultats suivants dans le tableau 2.2, on obtient avec ce modèle une meilleure erreur quadratique moyenne (*EQMP*).

Nous avons remarqué qu'il n'y a pas beaucoup de différence entre l'estimateur bayésien empirique proposé par Efron et Morris (1973) et Casella (1985) et le nouvel estimateur qui utilise la loi de Student t avec 25 degrés de liberté comme la loi *a priori*. Par contre, quand le nombre de degrés de liberté est très petit, il y a une différence et l'*EQMP* est plus petite quand le degré de liberté est petit. Ceci met en évidence l'hypothèse qu'un tel estimateur bayésien empirique est utile pour la protection contre les valeurs aberrantes.

2.5 Un estimateur pour le cas non équilibré

On considère un échantillon du type suivant : soit $X_i = (X_{i1}, \dots, X_{in_i})^t$ un vecteur de variables aléatoires *i.i.d.* de loi normale $N(\theta_i, \sigma_i^2)$ pour $1 \leq i \leq p$ et les n_i ($1 \leq i \leq p$)

ne sont pas nécessairement égales. Rappelons que dans l'exemple du baseball, les tailles n_i sont égales mais habituellement ce n'est pas le cas. Pour chaque i , \bar{X}_i représente la moyenne de X_{i1}, \dots, X_{in_i} et

$$V_i = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2.$$

Soit la matrice diagonale $A = \text{diag}(A_1, \dots, A_p)$, où $A_i = \frac{\sigma_i^2}{n_i} \forall i = 1, \dots, p$.

Rappelons que $V_i/A_i \sim \chi_{n_i-1}^2$. Étant donné le lien étroit, dans le cas d'échantillons de tailles égales, entre les estimateurs minimax du type James-Stein et les estimateurs bayésiens empiriques, nous poursuivons les estimateurs minimax du type James-Stein pour le cas non équilibré défini ici.

Berger et Bock (1976) considèrent le problème où l'on estime $\tilde{\theta}$ sous la fonction de perte quadratique définie ainsi :

$$L(\tilde{\delta}, \tilde{\theta}, A) = \frac{(\tilde{\delta} - \tilde{\theta})^t Q (\tilde{\delta} - \tilde{\theta})}{\text{tr}(QA)}, \quad (2.18)$$

pour un estimateur $\tilde{\delta} = (\delta_1, \dots, \delta_p)^t$, pour $Q = \text{Diag}(q_1, \dots, q_p)$ une matrice $p \times p$ diagonale définie positive et $\text{tr}(QA)$ est la constante de pondération. Souvent Q est une matrice diagonale choisie égale à A^{-1} ou I . On définit le risque quadratique moyen comme suit :

$$\begin{aligned} R_{\tilde{\delta}}(\tilde{\theta}, A) &= E_{\tilde{\theta}, A} L(\tilde{\delta}, \tilde{\theta}, A) \\ &= E_{\tilde{\theta}, A} \left[\frac{(\tilde{\delta} - \tilde{\theta})^t Q (\tilde{\delta} - \tilde{\theta})}{\text{tr}(QA)} \right]. \end{aligned}$$

Soit $W = \text{Diag}(w_1, \dots, w_p)$ une matrice diagonale $p \times p$,

avec $w_i = V_i / (n_i - 3) =$ pour $1 \leq i \leq p$ et

$$\begin{aligned} \|\tilde{X}\|_W^2 &= \tilde{X}^t W^{-1} Q^{-1} W^{-1} \tilde{X} \\ &= \sum_{i=1}^p \frac{X_i^2}{q_i w_i^2}. \end{aligned}$$

Soit $r(\tilde{X}, W)$ une fonction mesurable et intégrable de $\mathbb{R}^p \times \mathbb{R}^p$ vers \mathbb{R} . On propose un estimateur de $\tilde{\theta}$ défini par :

$$\delta(\tilde{X}, W) = (1 - r(\tilde{X}, W)) \|\tilde{X}\|_W^{-2} Q^{-1} W^{-1} \tilde{X}. \quad (2.20)$$

Soit

$$T = \min_{1 \leq i \leq p} (\chi_{n_i-1}^2 / n_i - 1),$$

et

$$\tau(n_1, \dots, n_p) = E(T^{-1}).$$

Théorème 2.1 (Berger et Bock (1976))

Soit $\delta(\tilde{X}, W)$ défini précédemment dans (2.20) tel que :

- (i) $0 \leq r(\tilde{X}, W) \leq 2(p - 2\tau)$,
- (ii) $r(\tilde{X}, W)$ est croissante en $|\tilde{X}_i|$ pour $i = 1, \dots, p$,
- (iii) $r(\tilde{X}, W)$ est croissante en w_i pour $i = 1, \dots, p$,
- (iv) $r(\tilde{X}, W) \|\tilde{X}\|_W^{-2}$ est croissante en w_i pour $i = 1, \dots, p$.

Alors $\delta(\tilde{X}, W)$ est un estimateur minimax pour θ sous la fonction de perte (2.19).

Démonstration du théorème : (d'après Berger et Bock (1976))

Équation (2.20) implique que

$$(\tilde{\delta} - \tilde{\theta}) = (\tilde{X} - \tilde{\theta}) - r(\tilde{X}, W) \|\tilde{X}\|_W^{-2} Q^{-1} W^{-1} \tilde{X},$$

d'où :

$$\begin{aligned} R_\delta(\tilde{\theta}, A) &= E_{\theta, A} [(\tilde{X} - \tilde{\theta})^t Q (\tilde{X} - \tilde{\theta}) / \text{tr}(QA)] \\ &\quad - E_{\theta, A} [2r(\tilde{X}, W) \|\tilde{X}\|_W^{-2} (\tilde{X} - \tilde{\theta})^t W^{-1} \tilde{X} / \text{tr}(QA)] \\ &\quad + E_{\theta, A} [r^2(\tilde{X}, W) \|\tilde{X}\|_W^{-4} \tilde{X}^t W^{-1} Q^{-1} Q Q^{-1} W^{-1} \tilde{X} / \text{tr}(QA)], \end{aligned}$$

du fait que

$$\tilde{X}^t W^{-1} (\tilde{X} - \tilde{\theta}) = (\tilde{X} - \tilde{\theta})^t W^{-1} \tilde{X},$$

car ceci est un scalaire et W est une matrice diagonale, donc symétrique.

De plus on a :

$$\begin{aligned} E_{\theta,A}[(\tilde{X} - \tilde{\theta})^t Q (\tilde{X} - \tilde{\theta}) / \text{tr}(QA)] &= E_{\theta,A}[\sum q_i (X_i - \theta_i)^2 / \text{tr}(QA)] \\ &= \sum q_i A_i / \text{tr}(QA) \\ &= 1. \end{aligned}$$

et

$$E_{\theta,A} \left[\frac{r^2(\tilde{X}, W) \|\tilde{X}\|_W^{-4} \tilde{X}^t W^{-1} Q^{-1} Q Q^{-1} W^{-1} \tilde{X}}{\text{tr}(QA)} \right] = E_{\theta,A} \left[\frac{r^2(\tilde{X}, W) \|\tilde{X}\|_W^{-2}}{\text{tr}(QA)} \right],$$

d'où :

$$R_\delta(\tilde{\theta}, A) = 1 - E_{\theta,A} \left[\frac{2r(\tilde{X}, W) \|\tilde{X}\|_W^{-2} (\sum X_i (X_i - \theta_i) / w_i)}{\text{tr}(QA)} \right] + E_{\theta,A} \left[\frac{r^2(\tilde{X}, W) \|\tilde{X}\|_W^{-2}}{\text{tr}(QA)} \right].$$

Étant donné que le risque de l'estimateur usuel sous la fonction de perte (2.19) est égal à 1 et que cet estimateur est minimax, pour démontrer que δ est minimax, il suffit de montrer l'inégalité suivante :

$$E_{\theta,A} \left[\frac{2r(\tilde{X}, W) \|\tilde{X}\|_W^{-2} (\sum_{i=1}^p X_i (X_i - \theta_i) / w_i)}{\text{tr}(QA)} \right] - E_{\theta,A} \left[\frac{r^2(\tilde{X}, W) \|\tilde{X}\|_W^{-2}}{\text{tr}(QA)} \right] \geq 0 \quad (2.21)$$

Ceci est équivalent à montrer que :

$$E_{\theta,A} [2r(\tilde{X}, W) \|\tilde{X}\|_W^{-2} (\sum_{i=1}^p X_i (X_i - \theta_i) / w_i)] - E_{\theta,A} [r^2(\tilde{X}, W) \|\tilde{X}\|_W^{-2}] \geq 0. \quad (2.22)$$

En faisant l'intégration par partie pour le premier terme de l'inégalité (2.21) avec

$$u = r(X, W) \|X\|_W^{-2} X_i$$

et

$$dv = (X_i - \theta_i) / A_i f_{\theta_i}(X_i) dX_i,$$

où $f_{\theta_i}(X_i)$ représente la densité gaussienne appropriée.

On a

$$\begin{aligned} E_{\theta,A} [\{r(\tilde{X}, W) \|\tilde{X}\|_W^{-2} X_i\} \{(X_i - \theta_i) / A_i\}] &= E_{\theta,A} \left[\frac{\partial}{\partial X_i} (r(\tilde{X}, W) \|\tilde{X}\|_W^{-2} X_i) \right] \\ &= E_{\theta,A} \left[\frac{r(\tilde{X}, W)}{\|\tilde{X}\|_W^2} - \frac{2r(\tilde{X}, W) X_i^2}{\|\tilde{X}\|_W^4 q_i W_i^2} \right. \\ &\quad \left. + \frac{X_i}{\|\tilde{X}\|_W^2} \left(\frac{\partial}{\partial X_i} r(\tilde{X}, W) \right) \right], \end{aligned}$$

d'où

$$E_{\theta,A}[2r(\tilde{X}, W)\|\tilde{X}\|_W^{-2}(\sum_{i=1}^p X_i(X_i - \theta_i)/W_i)] = E_{\theta,A}[(\sum_{i=1}^p \frac{2r(\tilde{X}, W)A_i}{\|\tilde{X}\|_W^2 W_i}) - \frac{4r(\tilde{X}, W)}{\|\tilde{X}\|_W^4}(\sum_{i=1}^p \frac{X_i^2 A_i}{q_i W_i^3}) + \sum_{i=1}^p \frac{2X_i A_i}{\|\tilde{X}\|_W^2 W_i}(\frac{\partial}{\partial X_i} r(\tilde{X}, W))] . \quad (2.23)$$

D'après la condition (ii) on a $(X_i \frac{\partial}{\partial X_i} r(X, W)) \geq 0$. Donc en combinant (2.22) et (2.23) une condition suffisante pour être minimax devient

$$E_{\theta,A}[(\sum \frac{2r A_i}{\|X\|_W^2 W_i}) - \frac{4r}{\|X\|_W^4}(\sum_{i=1}^p \frac{X_i^2 A_i}{q_i W_i^3}) - \frac{r^2}{\|X\|_W^2}] \geq 0.$$

Il faut démontrer le lemme suivant.

Lemme 2.1

$$E_{\theta,A}[\frac{r A_i}{\|X\|_W^2 W_i}] = E_{\theta,A}[\frac{r}{\|X\|_W^2} - \frac{4A_i r(X, W) X_i^2}{(n_i - 3)\|X\|_W^4 q_i w_i^3} - \frac{2A_i}{(n_i - 3)\|X\|_W^2}(\frac{\partial}{\partial w_i} r(X, W))]. \quad (2.24)$$

Démonstration : (d'après Berger et Bock (1976))

On suppose que toutes les dérivées partielles de $r(\tilde{X}, W)$ existent.

Soit $U \sim \chi_{n-1}^2$, $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction absolument continue, Efron et Morris (1976a) montrent :

$$E(Ug(U)) = (n-1)E(g(U)) + 2E(Ug'(U)) , \quad (2.25)$$

si les intégrales existent et convergent. On fait le changement de variable suivant dans (2.25).

Soit $Z = cU/(n-3)$ $c > 0$ et

$$\begin{aligned} h(z) = g((n-2)z/c) &\implies h'(z) = \frac{n-3}{c} g'((n-3)z/c) \\ &\implies g'(U) = \frac{c}{n-3} h'(z) \end{aligned}$$

Donc on peut réécrire (2.25) comme

$$\begin{aligned} E((n-3)zh(z)/c) &= E(Ug(U)) \\ &= (n-3)E(h(z)) + 2E(zh'(z)) \quad (2.26) \end{aligned}$$

Notons que d'après la définition de $W_i = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 / (n_i - 3)$

$$\begin{aligned} \frac{(n_i - 3)W_i}{A_i} &= \frac{V_i}{A_i} \\ &\sim \chi_{n_i-1}^2. \end{aligned}$$

Alors en recouvrant (2.26), on a

$$E_{\theta,A}[(n_i - 3)w_i h(w_i)/A_i] = E_{\theta,A}[(n_i - 1)h(w_i)] + 2E_{\theta,A}[w_i h'(w_i)]. \quad (2.27)$$

On pose

$$h(W_i) = \frac{r(\tilde{X}, W)}{\|\tilde{X}\|_W^2 W_i}. \quad (2.28)$$

Et on a

$$\frac{\partial}{\partial w_i} \|\tilde{X}\|_W^2 = \frac{-2X_i^2}{q_i w_i^3},$$

et

$$\frac{\partial}{\partial w_i} \frac{1}{\|\tilde{X}\|_W^2 w_i} = \frac{-1}{\|\tilde{X}\|_W^2 w_i^2} + \frac{2X_i^2}{\|\tilde{X}\|_W^4 q_i w_i^4},$$

dont

$$h'(w_i) = \frac{-r(\tilde{X}, W)}{\|\tilde{X}\|_W^2 w_i^2} + \frac{2r(\tilde{X}, W)X_i^2}{\|\tilde{X}\|_W^4 q_i w_i^4} + \frac{1}{\|\tilde{X}\|_W^2 w_i} \left(\frac{\partial}{\partial w_i} r(\tilde{X}, W) \right). \quad (2.29)$$

En faisant la substitution de (2.28) et (2.29) dans (2.27) on a

$$E_{\theta,A} \left[\frac{(n_i - 3)r(\tilde{X}, W)}{A_i \|\tilde{X}\|_W^2} \right] = E_{\theta,A} \left[\frac{(n_i - 1)r(\tilde{X}, W)}{\|\tilde{X}\|_W^2 w_i} - \frac{2r(\tilde{X}, W)}{\|\tilde{X}\|_W^2 w_i} + \frac{4r(\tilde{X}, W)X_i^2}{\|\tilde{X}\|_W^4 w_i^3} + \frac{2}{\|\tilde{X}\|_W^2} \frac{\partial}{\partial w_i} r(\tilde{X}, W) \right].$$

Avec un peu d'algèbre, l'équation (2.24) est vraie, c'est à dire que :

$$E_{\theta,A} \left[\frac{r(\tilde{X}, W)A_i}{\|\tilde{X}\|_W^2 W_i} \right] = E_{\theta,A} \left[\frac{r(\tilde{X}, W)}{\|\tilde{X}\|_W^2} - \frac{4A_i r(\tilde{X}, W)X_i^2}{(n_i - 3)\|\tilde{X}\|_W^4 q_i W_i^3} - \frac{2A_i}{(n_i - 3)\|\tilde{X}\|_W^2} \left(\frac{\partial}{\partial W_i} r(\tilde{X}, W) \right) \right]. \quad (2.30)$$

En faisant la substitution de (2.30) dans (2.20), on a

$$E_{\theta,A} \left[\frac{2pr(\tilde{X}, W) - r^2(\tilde{X}, W)}{\|\tilde{X}\|_W^2} - \frac{4r(\tilde{X}, W)}{\|\tilde{X}\|_W^4} \left(\sum_{i=1}^p \left(\frac{2A_i}{n_i - 3} + A_i \right) \frac{X_i^2}{q_i W_i^3} \right) - \frac{4}{\|\tilde{X}\|_W^2} \left(\sum_{i=1}^p \frac{A_i}{n_i - 3} \frac{\partial}{\partial W_i} r(\tilde{X}, W) \right) \right]$$

De (iii) on a

$$\frac{\partial}{\partial W_i} r(\tilde{X}, W) \leq 0,$$

donc il suffit de démontrer que :

$$E_{\theta,A} \left[\frac{2pr(\tilde{X}, W) - r^2(\tilde{X}, W)}{\|\tilde{X}\|_W^2} - \frac{4r(\tilde{X}, W)}{\|\tilde{X}\|_W^4} \left(\sum_{i=1}^p \left(\frac{2A_i}{n_i - 3} + A_i \right) \frac{X_i^2}{q_i W_i^3} \right) \right] \geq 0$$

ou bien

$$E_{\theta,A} \left[\frac{r(\tilde{X}, W)}{\|\tilde{X}\|_W^2} (2p - r(\tilde{X}, W)) - \frac{4}{\|\tilde{X}\|_W^2} \left(\max_{1 \leq i \leq p} \frac{(n_i - 1)A_i}{(n_i - 3)w_i} \right) \left(\sum_{i=1}^p \left(\frac{X_i^2}{q_i w_i^2} \right) \right) \right] \geq 0.$$

Parce que $\sum_{i=1}^p \frac{X_i}{q_i w_i^2} = \|X\|_W^2$, la condition suffisante devient

$$E_{\theta,A} [r(\tilde{X}, W) \|\tilde{X}\|_W^{-2} (2p - r(\tilde{X}, W) - 4f(W))] \geq 0$$

où

$$f(W) = 4 \max_{1 \leq i \leq p} \left(\frac{(n_i - 1)A_i}{(n_i - 3)w_i} \right). \quad (2.31)$$

D'après la définition de $f(W)$ et l'hypothèse (iii), $[2p - r(\tilde{X}, W) - f(W)]$ est non décroissante en w_i . L'hypothèse (iv) implique que $\left(\frac{r(\tilde{X}, W)}{\|\tilde{X}\|_W^2} \right)$ est non décroissante en w_i pour chaque i . Donc

$$E_{A_1}^{w_1} [r(\tilde{X}, W) \|\tilde{X}\|_W^{-2} (2p - r(\tilde{X}, W) - 4f(W))] \geq (E_{A_1}^{w_1} (r(\tilde{X}, W) \|\tilde{X}\|_W^{-2})) (E_{A_1}^{w_1} (2p - r(\tilde{X}, W) - f(W)))$$

Parce que w_1 soit indépendante de w_2

$E_{A_1}^{w_1} (r(\tilde{X}, W) \|\tilde{X}\|_W^{-2})$ et $(E_{A_1}^{w_1} (2p - r(\tilde{X}, W) - f(W)))$ sont non décroissantes en w_2 .

Si l'on continue en prenant les espérances jusqu'à $i = p$, on a :

$$\begin{aligned} & E_{\theta,A}^X (E_A^W [r(\tilde{X}, W) \|\tilde{X}\|_W^{-2} (2p - r(\tilde{X}, W) - 4f(W))]) \\ & \geq E_{\theta,A}^{\tilde{X}} [E_A^W (r(\tilde{X}, W) (\|\tilde{X}\|_W^{-2}) E_A^W (2p - r(\tilde{X}, W) - 4f(W)))] \\ & = E_{\theta,A}^X [E_A^W (r(\tilde{X}, W) (\|\tilde{X}\|_W^{-2}) (2p - E_A^W (r(\tilde{X}, W)) - 4\tau))] \quad (2.32) \end{aligned}$$

Parce que

$$\frac{(n_i - 1)A_i}{(n_i - 3)w_i} = \frac{V_i}{A_i(n_i - 1)} \sim \left[\frac{\chi_{n_i-1}^2}{(n_i - 1)} \right]^{-1},$$

et donc

$$E_A^V [h(W)] = 4E[T^{-1}] = 4\tau .$$

L'hypothèse (i) implique que

$$r(\tilde{X}, W) \|\tilde{X}\|_W^{-2} \geq 0 ,$$

et que

$$(2p - E_A^W[r] - 4\tau) \geq 0$$

Maintenant (2.32) implique que (2.30) est vraie.

Donc δ est minimax. ■

Remarque 2.1 Berger et Bock (1976) ont montré que pour $p \geq 3$, il existe N tel que si $n_i \geq N$ pour chaque $i = 1, \dots, p$, alors $(p - 2\tau) \geq 0$.

Remarque 2.2 Le choix usuel de $r(\tilde{X}, W) = c$ tel que $0 \leq c \leq 2(p - 2\tau)$. On a utilisé $c = 2(p - 2\tau)$ ici.

2.6 Application

Dans cette partie, on estime les moyennes de plusieurs groupes de tailles inégales où le premier modèle des sections 2.2 et 2.3 ne s'applique pas. On fait appel à l'estimateur minimax qui tient compte des tailles inégales. Nous utilisons ces données pour démontrer la méthodologie décrite ici. Voici la description des données.

La base de données de Whitmore et Gentleman (1982) à laquelle nous appliquons la méthode décrite ici concerne la durée de vie des cellules d'aluminium dans une usine après une grève. La provenance de ces données est déjà décrite dans l'introduction. Notre but ici est d'estimer la moyenne du temps entre l'intervention et l'échec de la cellule étant donné que l'intervention précède l'échec pour toutes les cellules de chaque groupe (on a éliminé les groupes de taille moins que 3 pour respecter les hypothèses de Berger et Bock (1976)). Cela nous donne les groupes de tailles inégales. Nous supposons que les données suivent une loi log-normale et nous utilisons l'estimateur proposé par Berger et Bock (1976). (Voir programme A.0.5) les résultats sont présentés dans le tableau 2.3.

On note

– Catégorie 1 : les groupes avant la grève, c'est-à-dire, ceux qui ont vécu avant le coupage

- de l'électricité ;
- Catégorie 2 : les groupes après la grève, c'est-à-dire, ceux qui ont vécu après le coupage de l'électricité ;
 - m_i la taille du groupe ;
 - \hat{y}_i l'estimateur usuel ;
 - $\hat{m}x$ l'estimateur minimax.

On remarque que, si on prend la différence quadratique moyenne pondérée (par la variance) entre $\hat{m}x$ et \hat{y} , cette différence est plus grande pour les groupes dans la catégorie 2 que dans la catégorie 1. Cela nous permis donc de croire que le coupage de l'électricité a affecté l'âge des cellules et, en conséquence, la grève a causé la perte d'argent pour l'usine.

Tableau 2.3 Estimateur minimax dans le cas non équilibré

Catégorie 1				
i	m	$\hat{\sigma}_1$	\hat{y}_1	$mx1$
1	17	0.4190	7.1420	7.141263
2	5	0.2211	7.1492	7.147397
3	21	0.3644	7.0727	7.071710
4	19	0.3648	6.8926	6.891648
5	13	0.2777	6.8616	6.860053
6	8	0.1296	6.7820	6.775779
7	7	0.3770	6.5310	6.530326
8	6	0.1098	6.8205	6.812752
Catégorie 2				
i	m	$\hat{\sigma}_2$	\hat{y}_2	$mx2$
9	3	0.0549	7.7811	7.780210
10	16	0.0780	7.6125	7.611368
11	20	0.0915	7.4789	7.478069
12	20	0.0970	7.4427	7.441964
13	22	0.1364	7.3627	7.362328
14	23	0.1391	7.2596	7.259246
15	26	0.2200	7.3001	7.299956
16	8	0.0958	7.3000	7.299243

CHAPITRE III

LOI DE POISSON

Dans ce chapitre, on étudie l'estimation bayésienne empirique paramétrique et non paramétrique de la moyenne de la loi de Poisson. On utilise les méthodes bayésiennes empiriques exposées dans Maritz et Lwin (1989). On considère aussi les cas équilibré et non équilibré.

3.1 Aspect théorique

Soit X une variable aléatoire. Étant donné θ , X suit une loi de Poisson de moyenne θ , de fonction de densité

$$f(x|\theta) = \frac{e^{-\theta}\theta^x}{x!},$$

où θ a une distribution inconnue G . On suppose que l'on a observé X_1, X_2, \dots, X_n indépendantes et identiquement distribuées de la même loi que X et que l'on a une nouvelle observation $X_{n+1} = a$. On veut estimer θ par $E(\theta|x)$.

Soit

$$\begin{aligned} h &= E(\theta|X_{n+1} = a) \\ &= \frac{\int_0^\infty \theta f(x|\theta)g(\theta)d(\theta)}{f(a)} \\ &= \frac{(a+1)f(a+1)}{f(a)}, \end{aligned}$$

avec a un entier fixé. On pose :

$$u = u(x) = \begin{cases} 1 & \text{si } x = a. \\ 0 & \text{sinon;} \end{cases}$$

$$v = v(x) = \begin{cases} 1 & \text{si } x = a + 1. \\ 0 & \text{sinon.} \end{cases}$$

Ainsi,

$$u \sim B(1, f(a)) \quad \text{et} \quad v \sim B(1, f(a + 1)) \quad (3.1)$$

et

$$T_n = \frac{(a + 1) \sum_1^n v_i}{\sum_1^n u_i}. \quad (3.2)$$

Le lemme suivant montre que T_n est un bon estimateur de h .

Lemme 3.1 (Robbins(1980))

T_n est un estimateur qui converge en probabilité vers h . De plus, $\sqrt{n}(T_n - h)$ converge en distribution vers $N(0, \sigma^2)$, où

$$\sigma^2 = n(a + 1)^2 \left[1 + \frac{n(a + 1)}{n(a)} \right] \frac{n(a + 1)}{n^2(a)} \quad (3.3)$$

et $n(j)$ est le nombre de fois qu'on a observé j parmi les x_1, x_2, \dots, x_n .

Démonstration :

On donne les indications de la démonstration de Robbins (1980). Comme $E(u) = f(a)$ et $E(v) = f(a + 1)$, on a

$$h = \frac{(a + 1)E(v)}{E(u)}, \quad (3.4)$$

et par la suite on montre que T_n est un estimateur convergent en probabilité vers h . Parce que $E(v_i) = f(a + 1)$ et $E(u_i) = f(a)$, alors $(a + 1) \sum_1^n v_i$ et $\sum_1^n u_i$ convergent en probabilité vers $(a + 1)f(a + 1)$ et $f(a)$ respectivement (Casella et Berger (1990) Théorème 5.5.2). Or la fonction $(x, y) \mapsto x/y$ est continue sur $\mathbb{R} \times \mathbb{R} \setminus \{0\}$ et par le Théorème 5.5.4 de (Casella et Berger (1990)) d'où le premier résultat du lemme en découle.

Pour le deuxième résultat, d'après le théorème de limite centrale,

$$\sqrt{n}(T_n - h) = \frac{\sum_1^n ((a+1)v_i - hu_i)}{(\sum_1^n u_i/n)\sqrt{n}}$$

converge en distribution vers $N(0, \sigma^2)$ avec

$$\begin{aligned} \sigma^2 &= \frac{E((a+1)v - hu)^2}{E^2(u)} \\ &= \frac{E((a+1)^2v^2 + h^2u^2)}{f^2(a)} \\ &= \frac{E((a+1)^2v + h^2u)}{f^2(a)} \\ &= \frac{1}{f^2(a)} \left\{ (a+1)^2 f(a+1) + \frac{(a+1)^2 f^2(a+1)}{f^2(a)} f(a) \right\} \\ &= \frac{(a+1)^2 f(a+1)}{f^2(a)} \left\{ 1 + \frac{f(a+1)}{f(a)} \right\} \\ &= n \frac{(a+1)^2 n(a+1)}{n^2(a)} \left\{ 1 + \frac{n(a+1)}{n(a)} \right\}, \end{aligned}$$

La fonction $f(a)$ est estimée par $n(a)/n$ et $f(a+1)$ par $n(a+1)/n$.

D'après ce qui précède, on obtient alors

$$\frac{T_n - h}{\sqrt{\frac{(a+1)^2 n(a+1)}{n^2(a)} \left[1 + \frac{n(a+1)}{n(a)} \right]}} \rightarrow N(0, 1).$$

On déduit aussi, pour n grand, un intervalle approximatif de confiance à 95 % pour h qui est :

$$(a+1) \frac{n(a+1)}{n(a)} \pm \frac{1.96(a+1)}{\sqrt{n(a)}} \sqrt{\frac{n(a+1)}{n(a)} \left[1 + \frac{n(a+1)}{n(a)} \right]}. \quad \blacksquare$$

C'est possible d'avoir un meilleur estimateur que T_{n_j} par exemple, si G , la mesure *a priori* de θ , est de la loi *Gamma* de paramètres (α, β) :

$$G(\theta|\alpha, \beta) = \frac{1}{\Gamma(\alpha)} \beta^\alpha \theta^{\alpha-1} e^{-\beta\theta},$$

alors la densité marginale de X est :

$$\begin{aligned} f_X(x) &= \int_0^\infty f(x|\theta) dG(\theta) \\ &= \frac{\Gamma(x+\alpha)}{x! \Gamma(\alpha)} \left(\frac{\beta}{1+\beta} \right)^\alpha \left(\frac{1}{1+\beta} \right)^x, \end{aligned}$$

et de ce qui précède, la valeur de h est alors

$$\begin{aligned} h &= \frac{(a+1)f(a+1)}{f(a)} \\ &= \frac{a+\alpha}{1+\beta}. \end{aligned} \quad (3.5)$$

Dé

$$E(X) = E(E(X|\theta)) = E(\theta) = \frac{\alpha}{\beta}$$

et

$$Var(X) = \frac{\alpha(1+\beta)}{\beta^2} = \frac{\alpha}{\beta^2},$$

on déduit

$$\beta = \frac{E(X)}{Var(X) - E(X)} \quad \text{et} \quad \alpha = \frac{E^2(X)}{Var(X) - E(X)},$$

d'où

$$h = E(X) + \left(1 - \frac{E(X)}{Var(X)}\right)(a - E(X)). \quad (3.6)$$

Si G ne suit pas une loi de *Gamma*, on pose

$$k = E(X) + \left(1 - \frac{E(X)}{Var(X)}\right)(a - E(x)), \quad (3.7)$$

et comme h n'est pas nécessairement égale à k , on cherche un estimateur qui converge en distribution vers h . En mettant les estimateurs,

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{et} \quad s_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \quad \text{pour } E(X) \text{ et } Var(X) \text{ dans (3.7)}$$

on obtient un autre estimateur W_n où

$$W_n = \bar{x}_n + \left(1 - \frac{\bar{x}_n}{s_n^2}\right)(a - \bar{x}_n).$$

Notons que T_n est asymptotiquement équivalent à W_n si $h=k$.

Robbins (1980) a montré que $\sqrt{n}(W_n - k)$ converge en distribution vers la loi $N(0, \sigma_1^2)$ où

$$\sigma_1^2 = \frac{(2E(x) - a)^2}{\mu_2} + \frac{E^2(x)(E(x) - a)^2}{\mu_2^4} (\mu_4 - \mu_2^2) - \frac{2E(x)(E(x) - a)(2E(x) - a)\mu_3}{\mu_2^3}$$

avec

$$\mu_i = E(x - E(x))^i \quad \text{pour } i \geq 2 .$$

Robbins (1980) a proposé l'estimateur Z_n suivant :

$$Z_n = \frac{W_n + n^\epsilon (T_n - W_n)^2 T_n}{1 + n^\epsilon (T_n - W_n)^2} \quad \text{pour } 1/2 < \epsilon < 1 .$$

Maintenant,

$$\begin{aligned} \sqrt{n}(Z_n - h) &= \frac{\sqrt{n}W_n + n^\epsilon (T_n - W_n)^2 T_n \sqrt{n} - h\sqrt{n} - hn^\epsilon (T_n - W_n)^2 \sqrt{n}}{1 + n^\epsilon (T_n - W_n)^2} \\ &= \frac{\sqrt{n}(W_n - h) + n^\epsilon (T_n - W_n)^2 (T_n - h) \sqrt{n}}{1 + n^\epsilon (T_n - W_n)^2} . \end{aligned}$$

o Si $h = k$, alors $\sqrt{n}(T_n - h) \rightarrow 0$ et $\sqrt{n}(W_n - h) \rightarrow 0$ en distribution, d'où

$$\sqrt{n}|T_n - W_n| \leq \sqrt{n}|T_n - h| + \sqrt{n}|W_n - h| \rightarrow 0 \quad \text{en probabilité .}$$

Ceci implique que $T_n - W_n = O(n^{-\frac{1}{2}})$ et de même pour

$$n^\epsilon (T_n - W_n)^2 = O(n^{\epsilon-1}) \rightarrow 0 \quad \text{en probabilité}$$

Finalement

$$\sqrt{n}(Z_n - h) \sim \sqrt{n}(W_n - h)$$

et Robbins(1980) a démontré que $\sqrt{n}(W_n - h) \rightarrow N(0, \sigma_1^2)$ en distribution.

o Si $h \neq k$ alors

$$|T_n - W_n| \rightarrow |h - k| \quad \text{en probabilité ,}$$

du fait que $x \mapsto |x|$ est continue et T_n et W_n convergent respectivement en probabilité vers h et k .

On peut déduire que

$$\sqrt{n}(Z_n - h) \sim \sqrt{n}(T_n - h) \rightarrow N(0, \sigma^2) . \quad \blacksquare$$

Dorénavant, on suppose que G suit une loi de *gamma* de paramètre (α, β) et on appelle ce modèle Poisson-gamma. Maritz et Lwin (1989) suggèrent plusieurs méthodes pour l'estimation de la moyenne de la loi de Poisson. Nous les citons dans la prochaine section.

3.2 Différentes méthodes d'estimation bayésienne empirique paramétrique et non paramétrique

Dans cette section on fait une étude de simulation pour illustrer ces différentes méthodes. On génère n variables aléatoires de ce modèle avec $\alpha = 10$ et $\beta = 2$ de la manière suivante : on génère une variable λ qui suit la loi de Poisson-gamma et avec cette valeur on génère 50 variables selon la loi de Poisson de moyenne λ (Prog A.0.1).

On trouve les données suivantes :

10, 5, 5, 4, 2, 4, 3, 5,3,5, 6, 4, 7, 7, 3, 3, 8, 5, 10, 6 5, 8, 9, 6, 4, 3,
4, 6, 4, 12, 14, 2, 4, 0, 10, 4, 6, 4, 5, 5, 7, 2, 3, 3, 4, 4, 1, 2, 7, 8.

On calcule $\bar{x} = 5.22$ et $\hat{\sigma}^2 = 7.807755$.

On rappelle que la densité gamma est définie par :

$$\Gamma_{(\alpha,\beta)}(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda},$$

et la densité marginale du modèle Poisson-gamma pour une observation x_i est une densité de la loi binomiale négative définie par :

$$f(x_i) = \frac{\Gamma(x_i + \alpha)}{x_i! \Gamma(\alpha)} \left(\frac{\beta}{\beta + 1}\right)^\alpha \left(\frac{1}{\beta + 1}\right)^{x_i}$$

Illustrons par la suite les différentes méthodes pour l'estimation bayésienne empirique de θ . Étant donné que les calculs sont assez longs et complexes, nous avons choisi $n = 50$ à titre d'illustration. Évidemment, il serait souhaitable dans un recherche future d'avoir un plus grand nombre de simulations.

a. Méthode des moments (Paramétrisation de G qui suit une loi gamma avec (α, β) inconnu). On utilise la méthode des moments pour l'estimation. On a

- $E(X) = \frac{\alpha}{\beta} = \bar{x}$
- $Var(X) = \frac{\alpha(1+\beta)}{\beta^2} = \hat{\sigma}^2$
- $\beta = \frac{E(X)}{Var(X) - E(X)} = 2.02$
- $\alpha = \frac{E^2(x)}{Var(x) - E(x)} = 10.53$

$$- \delta^E(X) = \frac{x+\alpha}{\beta+1} = 3.487 + 0.33X.$$

b. Méthode du maximum de vraisemblance (ML) (Paramétrisation de G qui suit une loi gamma avec (α, β) inconnu). Il faut estimer α et β par la méthode du maximum de vraisemblance. La fonction de vraisemblance associée aux 50 observations est donnée par :

$$\begin{aligned} l(x_1, \dots, x_{50}) &= \prod_{i=1}^{50} f(x_i) \\ &= \prod_{i=1}^{50} \frac{\Gamma(x_i + \alpha)}{x_i! \Gamma(\alpha)} \left(\frac{\beta}{\beta + 1}\right)^\alpha \left(\frac{1}{\beta + 1}\right)^{x_i} \end{aligned}$$

Le logarithme de la fonction de vraisemblance est donné par

$$\text{Log}(l(x_1, \dots, x_{50})) = \sum_{i=1}^{50} \log\left(\frac{\Gamma(x_i + \alpha)}{x_i! \Gamma(\alpha)} \left(\frac{\beta}{\beta + 1}\right)^\alpha \left(\frac{1}{\beta + 1}\right)^{x_i}\right).$$

Nous écrivons

$$\frac{\partial \text{Log}(l(x_1, \dots, x_{50}))}{\partial \beta} = 0.$$

On peut alors ignorer les facteurs de $l(x_1, \dots, x_{50})$ qui ne dépendent pas de β , ce qui donne :

$$\sum_{i=1}^{50} \left(\alpha \left(\frac{1}{\beta} - \frac{1}{\beta + 1}\right) - \frac{x_i}{\beta + 1}\right) = 0 \Rightarrow \hat{\alpha} = \beta \bar{x}$$

Pour résoudre, on remplace α par $\beta \times \bar{x}$ dans $\text{Log}(l(x_1, \dots, x_{50}))$, la fonction de vraisemblance devient en ignorant les facteurs qui ne dépendent pas de β :

$$\sum_{i=1}^{50} \log\left(\frac{\Gamma(x_i + 5.22\beta)}{\Gamma(5.22\beta)} \left(\frac{\beta}{\beta + 1}\right)^{5.22\beta} \left(\frac{1}{\beta + 1}\right)^{x_i}\right)$$

et on prend la dérivée à nouveau par rapport à β , on met cette dérivée = 0 et on va résoudre l'équation avec (Maple) qui va nous trouver la valeur de $\hat{\beta}$. (voir programme A.0.2).

On trouve :

$$\hat{\beta} = 2.350700930, \quad \text{d'où} \quad \hat{\alpha} = 12.27065885.$$

$$\delta^E(X) = \frac{x + \hat{\alpha}}{\hat{\beta} + 1} = 3.6621 + 0.29845X.$$

c. Méthode de vraisemblance maximale quand on suppose que la densité de X est un mélange de 3 densités de Poisson de paramètres $\lambda_1 \leq \lambda_2 \leq \lambda_3$ Maritz (1989).

Soit $f(x) = \frac{1}{3} \sum_{i=1}^3 \frac{e^{-\lambda_i} \lambda_i^x}{x!}$ et

$$\text{Log}(l) = \sum_{k=0}^{16} f_n(x_k) \log\left(\frac{1}{3} \sum_{i=1}^3 \frac{e^{-\lambda_i} \lambda_i^{x_k}}{x_k!}\right),$$

où $f_n(x_i)$ est le nombre de fois que l'on observe x_i . Étant donné que $f_n(x) < 0.001$ si $x > 16$ on l'ignore dans les calculs. Or

$$\begin{aligned} \frac{\partial \text{Log}(l)}{\partial \lambda_j} &= \sum_{k=0}^{16} f_n(x_k) \frac{\frac{1}{3} \frac{1}{x_k!} e^{-\lambda_j} \lambda_j^{x_k-1} (x_k - \lambda_j)}{\frac{1}{3} \sum_{i=1}^3 \frac{e^{-\lambda_i} \lambda_i^{x_k}}{x_k!}} \\ &= \frac{\sum_{k=0}^{16} f_n(x_k) e^{-\lambda_j} \lambda_j^{x_k-1} (x_k - \lambda_j)}{\sum_{i=1}^3 e^{-\lambda_i} \lambda_i^{x_k}} = 0 \end{aligned}$$

Ceci est vrai si et seulement si

$$\lambda_j \sum_{k=0}^{16} \frac{x_k f_n(x_k)}{\sum_{i=1}^3 e^{-\lambda_i} \lambda_i^{x_k}} = \sum_{k=0}^{16} \frac{f_n(x_k)}{\sum_{i=1}^3 e^{-\lambda_i} \lambda_i^{x_k}}.$$

Si on pose

$$D_j = \sum_{k=0}^{16} \frac{x_k f_n(x_k)}{\sum_{i=1}^3 e^{-\lambda_i} \lambda_i^{x_k}} \quad \text{et} \quad A_j = \sum_{k=0}^{16} \frac{f_n(x_k)}{\sum_{i=1}^3 e^{-\lambda_i} \lambda_i^{x_k}}$$

alors $\lambda_j = A_j/D_j$ et avec la méthode de Newton (voir programme A.0.3), on obtient :

$$\lambda_1 = 4.151728, \quad \lambda_2 = 4.151728, \quad \text{et} \quad \lambda_3 = 7.48663;$$

$$\delta^E(X) = \frac{\sum_{i=1}^3 e^{-\lambda_i} \lambda_i^{X+1}}{\sum_{i=1}^3 e^{-\lambda_i} \lambda_i^X}.$$

(La convergence de cette méthode est assurée par le même raisonnement utilisé pour les estimateurs $\hat{\mu}$ et $\hat{\tau}^2$ dans la section (2.4)).

d. Méthode de minimisation du khi-deux quand on suppose que la densité est un mélange de 3 densités poissonniennes de paramètres $\lambda_1 \leq \lambda_2 \leq \lambda_3$ comme le cas précédent. Cette méthode consiste à minimiser

$$S = n \sum_{i=1}^m \frac{[f_n(x_i)/n - \{p(x_i|\lambda_1) + p(x_i|\lambda_2) + p(x_i|\lambda_3)\}/3]^2}{p_3(x_i)}$$

Pour trouver les paramètres λ_1 , λ_2 et λ_3 on procède de la manière suivante d'après Maritz (1966) :

Soit $g_n(x_j)$ ($1 \leq j \leq m = 8$) les fréquences regroupées comme dans le tableau (3.1) ;

Tableau 3.1 Les fréquences regroupées pour la méthode de khi-deux

j	x_j	$g_n(x_j)$
1	2	6/50
2	3	7/50
3	4	11/50
4	5	8/50
5	6	5/50
6	7	4/50
7	8	3/50
8	14	6/50

x_1, \dots, x_8 les observations qui correspondent aux fréquences regroupées (voir tableau (3.1)) ; et $p_n(x_j)$ les fréquences théoriques regroupées correspondantes avec la densité $f(x) = \frac{1}{3} \sum_{i=1}^3 \frac{e^{-\lambda_i} \lambda_i^x}{x!}$; c'est à dire,

$$p_n(x_1) = f(0) + f(1) + f(2) ,$$

$$p_n(x_8) = f(9) + f(10) + f(12) + f(14) ,$$

$$p_n(x_j) = f(x_j) \text{ pour } 2 \leq j \leq 7 .$$

Ainsi dans notre cas on a

$$x_1 = 2, x_2 = 3, \dots, x_8 = 14,$$

$$g_n(x_1) = g_n(2) = (f_n(0) + f_n(1) + f_n(2))/50 = 6/50,$$

$$g_n(x_2) = g_n(3) = f_n(3)/50 = 7/50, \dots, \text{et}$$

$$g_n(x_8) = g_n(14) = (f_n(9) + f_n(10) + f_n(12) + f_n(14))/50 = (1 + 3 + 1 + 1)/50 = 6/50.$$

On pose

$$\tilde{\lambda} = (\lambda_1, \lambda_2, \lambda_3)^t \text{ et } \lambda^0 = (\lambda_1^0, \lambda_2^0, \lambda_3^0) \text{ sont les valeurs initiales.}$$

On a fait le même choix des valeurs initiales de λ^0 que Maritz (1966). Soit $\tilde{\Delta} = (\Delta_1, \Delta_2, \Delta_3)^t$.

On estime $\tilde{\lambda}$ comme suit

$$\tilde{\lambda} = \tilde{\lambda}^0 + \tilde{\Delta},$$

où $\tilde{\Delta}$ est la solution du système d'équations suivant

$$A^0 \tilde{\Delta} = \tilde{Y}^0;$$

où

$$\begin{aligned} A^0 &= (a_{ij}^0)_{1 \leq i, j \leq 3} \text{ et } \tilde{Y}^0 = (y'_1, y'_2, y'_3), \\ a_{rs}^0 &= \frac{n}{9} \sum_{j=1}^m \frac{\partial p(x_j | \lambda_r^0)}{\partial \lambda_r} \frac{\partial p(x_j | \lambda_s^0)}{\partial \lambda_s}, \quad 1 \leq r, s, \leq 3, \\ y'_r &= \frac{n}{3} \sum_{j=1}^m y_j^0 \frac{\partial p(x_j | \lambda_r^0)}{\partial \lambda_r} p_n(x_j), \quad 1 \leq r \leq 3, \\ y_j^0 &= g_n(x_j)/n - p_n(x_j), \quad 1 \leq j \leq 3. \end{aligned}$$

(Voir programme A.0.4 dans l'annexe).

Voici les résultats des simulations. Dans le tableau 3.2 suivant, on utilise les notations suivantes :

- $f_n(x)$ est l'effectif;
- $f(x)$ est la valeur de la densité marginale à x ;
- $\delta^B(x) = \frac{x+\alpha}{\beta+1}$ est l'estimateur de Bayes quand on suppose que la loi *a priori* est une loi de gamma;
- $\delta^{E_n}(x)$ est l'estimateur bayésien empirique non paramétrique défini dans le chapitre 1;

Tableau 3.2 Estimation pour la loi de Poisson

x	$f_n(x)$	f(x)	$\delta^B(x)$	$\delta^{E_n}(x)$	$\delta^{E_a}(x)$	$\delta^{E_b}(x)$	$\delta^{E_c}(x)$	$\delta^{E_d}(x)$
0	1	0.017341530	3.333333	0.500000	3.487	2.960	4.210080	2.920209
1	1	0.057805100	3.666667	1.000000	3.817	3.393	4.255494	3.095416
2	4	0.105976016	4.000000	2.400000	4.147	3.826	4.334281	3.424810
3	7	0.141301355	4.333333	3.500000	4.477	4.259	4.467053	3.974409
4	11	0.153076468	4.666667	4.583333	4.807	4.692	4.680203	4.730468
5	8	0.142871370	5.000000	5.333333	5.137	5.125	4.997096	5.549046
6	5	0.119059475	5.333333	5.833333	5.467	5.558	5.418255	6.266539
7	4	0.090711981	5.666667	6.400000	5.797	5.991	5.901742	6.826278
8	3	0.064254320	6.000000	6.750000	6.127	6.424	6.371701	7.252189
9	1	0.042836213	6.333333	5.000000	6.457	6.857	6.760155	7.579489
10	3	0.027129602	6.666667	8.250000	6.787	7.290	7.040467	7.831130
11	-	0.016442183	7.000000	0.000000	7.117	7.723	7.223530	8.021347
12	1	0.009591273	7.333333	6.500000	7.447	8.156	7.335413	8.161485
13	-	0.005410462	7.666667	0.000000	7.777	8.589	7.401044	8.262080
14	1	0.002962872	8.000000	7.500000	8.107	9.022	7.438619	8.332709
15	-	0.001580198	8.333333	0.000000	8.437	9.455	7.459834	8.381456
16	-	0.000823020	8.666667	0.000000	8.767	9.888	7.471717	8.414679
$W(\delta)$			1.67	4.168128	1.688845	1.758456	1.748632	2.244535

- $\delta^{E_a}(x)$ est l'estimateur bayésien empirique avec la méthode des moments quand on suppose que la loi *a priori* est une loi de gamma ;
- $\delta^{E_b}(x)$ est l'estimateur bayésien empirique avec la méthode du maximum de vraisemblance quand on suppose que la loi *a priori* est un mélange de trois densités poissonniennes de paramètres $\lambda_1 \leq \lambda_2 \leq \lambda_3$;
- $\delta^{E_c}(x)$ est l'estimateur bayésien empirique avec la méthode du maximum de vraisemblance quand on suppose que la densité est un mélange de 3 poissonniennes de paramètres $\lambda_1 \leq \lambda_2 \leq \lambda_3$;
- $\delta^{E_d}(x)$ est l'estimateur bayésien empirique avec la méthode de minimisation de χ_2 quand on suppose que la densité est un mélange des 3 poissonniennes de paramètres $\lambda_1 \leq \lambda_2 \leq \lambda_3$;
- On calcule aussi le risque intégré $W(\delta)$ pour chaque estimateur défini précédemment

par la façon suivante (Voir proposition 1.3) :

$$W(\delta) = W(\delta^B) + \sum (\delta(x) - \delta^B(x))^2 f(x),$$

avec

$$W(\delta^B) = \frac{\alpha}{\beta(\beta + 1)}.$$

D'après le tableau 3.2, l'estimateur de vraisemblance maximale $\delta^{E_a}(x)$ est meilleur que $\delta^{E_c}(x)$, $\delta^{E_c}(x)$ est meilleur que $\delta^{E_b}(x)$ et $\delta^{E_{cb}}(x)$ est meilleur que $\delta^{E_d}(x)$. Le pire est bien sûr $\delta^{E_n}(x)$, l'estimateur bayésien empirique non paramétrique puisqu'on suppose qu'on ne connaît pas la loi *a priori*.

3.3 Estimation bayésienne empirique pour les échantillons de tailles inégales

Ici, nous suivons la méthodologie décrite par Maritz et Lwin (1989) pour le cas général de l'estimateur bayésien empirique. On suppose X_1, \dots, X_m , m valeurs indépendantes de l'observation courante X de moyenne λ , et n échantillons de valeurs indépendantes de tailles m_i , $1 \leq i \leq n$: X_{i1}, \dots, X_{im_i} distribuées selon la loi de Poisson de moyenne $\lambda_1, \dots, \lambda_n$ respectivement. Notons que $\lambda_1, \dots, \lambda_n$ et λ sont de même loi, disons Λ . Pour une bonne application de la théorie bayésienne empirique, on doit tenir compte de la taille de l'échantillon de l'observation courante car dans le cas non équilibré, les tailles différentes (m_i) agissent comme des paramètres de nuisance et la complexité du problème augmente.

- Cas $m \leq m_i (\forall i)$: $\hat{\lambda}$ est l'estimateur de λ obtenu par la méthode de vraisemblance maximale. On sélectionne m observations pour chaque échantillon d'observations précédentes et on estime avec la méthode de vraisemblance maximale $\hat{\lambda}_i$ de λ_i . On obtient donc $\hat{\lambda}_1, \dots, \hat{\lambda}_n$ et on estime λ avec la même méthode bayésienne empirique que dans le cas simple ; c'est-à-dire, en utilisant tous les cas possibles. Il y a $\prod \binom{m_i}{m}$ cas possibles d'estimateurs et on déduit un estimateur, par exemple, la moyenne de tous les estimateurs.

Supposons que la loi *a priori* G est une loi paramétrique et que $T(x_1, \dots, x_m)$ est une statistique exhaustive étant donné l'observation courante et que $T_i(x_{i1}, \dots, x_{im})$, est une statistique exhaustive correspondante pour l'échantillon précédent pour chaque $i = 1, \dots, n$. Pour les observations passées pour chaque i , ($1 \leq i \leq n$), la distribution de T_i dépend de λ et m_i . Sa densité est $f(t_i|\lambda, m_i)$ et $G(\lambda, \tilde{\theta})$ est sa loi *a priori* qui dépend de $\tilde{\theta}$ de dimension k ; c'est-à-dire, le nombre de paramètres à estimer est égal à k . Par exemple, dans le cas de la loi *gamma*, $k = 2$ et $\tilde{\theta} = (\alpha, \beta)$. La fonction de vraisemblance de t_1, \dots, t_n est :

$$L(t, \theta) = \prod_{i=1}^n \int f(t_i|\lambda, m_i)g(\lambda_i, \theta)d\lambda_i .$$

On estime θ en maximisant $l(t, \theta)$. On pourra aussi calculer θ par la méthode des moments. Par la suite, on va donner un exemple d'estimation de G par simulation, pour le cas d'une loi de Poisson. Soit $X \sim \text{Poisson}(\lambda)$ et

$$T_i = \sum_{j=1}^n \frac{X_{ij}}{m_i},$$

une statistique exhaustive pour λ_i . Alors

$$E(T_i|\lambda_i) = \lambda_i$$

et

$$\begin{aligned} E(T_i^2|\lambda_i) &= E^2(T_i|\lambda_i) + \text{Var}(T_i|\lambda_i) \\ &= \lambda_i^2 + \frac{1}{m_i^2}m_i\lambda_i \\ &= \lambda_i^2 + \lambda_i/m_i . \end{aligned}$$

Si de plus $\lambda|(\alpha, \beta) \sim \Gamma_{(\alpha, \beta)}(\lambda)$, alors on estime α et β empiriquement des observations précédentes avec la méthode des moments de la manière suivante :

$$\begin{aligned} E(\sum_{j=1}^n T_i) &= E(E(\sum_{j=1}^n T_i|\lambda_i)) \\ &= \sum_{j=1}^n E(\lambda_i) \\ &= n \frac{\alpha}{\beta} , \end{aligned}$$

aussi .

$$\begin{aligned}
 E(\Sigma_{j=1}^n T_i^2) &= E(E(\Sigma_{j=1}^n T_i^2 | \lambda_i)) \\
 &= \Sigma_{j=1}^n E(E(T_i^2 | \lambda_i)) \\
 &= \Sigma_{j=1}^n E(\lambda_i^2 + \frac{\lambda_i}{m_i}) \\
 &= n(\frac{\alpha}{\beta^2} + \frac{\alpha^2}{\beta^2}) + (\frac{\alpha}{\beta}) \Sigma_{j=1}^n \frac{1}{m_i}.
 \end{aligned}$$

Des deux équations précédentes on déduit que :

$$\begin{aligned}
 \frac{\hat{\alpha}}{\hat{\beta}} &= \bar{T}, \\
 n(\frac{\hat{\alpha}}{\hat{\beta}^2} + \frac{\hat{\alpha}^2}{\hat{\beta}^2}) + (\frac{\hat{\alpha}}{\hat{\beta}}) \Sigma_{j=1}^n \frac{1}{m_i} &= \Sigma_{j=1}^n T_i^2 ;
 \end{aligned}$$

D'où

$$n(\bar{T} \frac{1}{\hat{\beta}} + \bar{T}^2) + \bar{T} \Sigma_{i=1}^n \frac{1}{m_i} = \Sigma_{i=1}^n n T_i^2 ;$$

donc

$$\hat{\beta} = n\bar{T} / [\Sigma_i^n T_i^2 - \bar{T} \Sigma_i^n \frac{1}{m_i} - n\bar{T}] \quad (3.8)$$

et

$$\hat{\alpha} = \hat{\beta} \bar{T}. \quad (3.9)$$

Dans le tableau 3.3, on donne un exemple de l'estimation de α et β . On prend $n=10$ et on génère 10 valeurs aléatoires de taille m_i entre 50 et 1000, et 10 valeurs de $\lambda_i \sim \Gamma_{20,0.5}$, ($\alpha = 20$, $\beta = 0.5$) et pour chaque λ_i on génère un échantillon P_{i1}, \dots, P_{im_i} de loi de *Poisson*(λ_i) respectivement. On calcule T_i et T_i^2 .

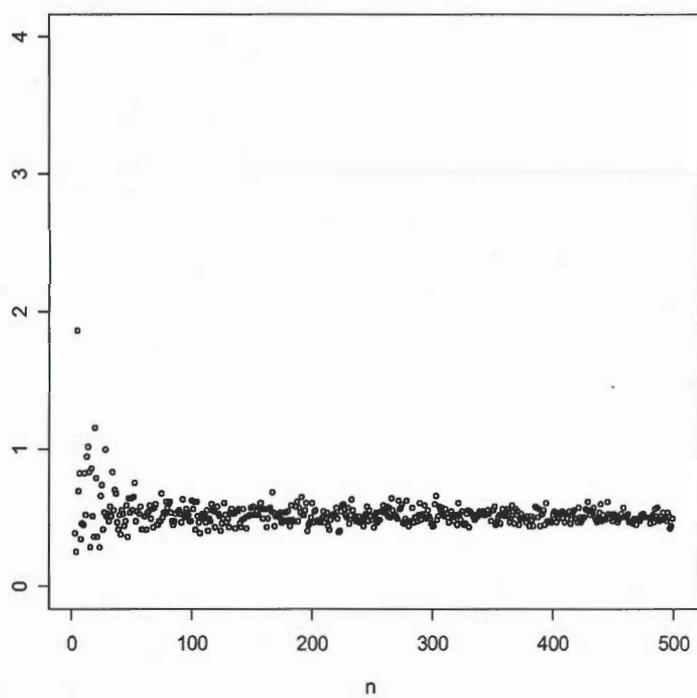
Le tableau 3.3 démontre que T_i est un bon estimateur de λ_i pour chaque i . En utilisant les équations (3.8) et (3.9) on obtient comme estimateurs de α et β les valeurs $\hat{\alpha} = 28.48802$ et $\hat{\beta} = 0.678019$.

Tableau 3.3 Le cas de taille inégale pour la loi Poisson-gamma

i	m_i	λ_i	T_i	T_i^2
1	67	48.15998	48.26866	2329.863
2	987	42.61722	42.57548	1812.672
3	124	45.87999	45.03226	2027.904
4	460	49.14388	48.95870	2396.954
5	550	52.26041	52.34545	2740.047
6	593	31.85459	31.67960	1003.597
7	553	50.17172	49.33273	2433.718
8	920	37.59565	37.89674	1436.163
9	491	67.12693	67.26477	4524.549
10	783	63.17349	63.36143	4014.671

Pour chaque valeur de n entre 10 et 500, on génère 10 valeurs m_i entre 50 et 1000 et 10 valeurs $\lambda_i \sim \Gamma_{20,0.5}$ et pour chaque λ_i on génère des échantillons P_{i1}, \dots, P_{im_i} de la loi de Poisson(λ_i) respectivement. La figure 3.1 donne le graphique de l'estimateur $\hat{\beta}$ comme une fonction de n . L'estimateur $\hat{\beta}$ varie avec β , mais quand n est grand la variation diminue comme la figure 3.1 le démontre.

Figure 3.1 La variation de l'estimateur $\hat{\beta}$ en fonction de n dans le cas de la loi Poisson-gamma



CHAPITRE IV

LOI EXPONENTIELLE

On considère le modèle exponentiel comme suit :

$$(4.1) \begin{cases} x|\theta \sim \wp(\theta), \\ \theta|\lambda \sim \text{exp}(\lambda), \\ \lambda \text{ inconnu.} \end{cases}$$

L'objet est d'estimer θ . Premièrement, on suppose que le modèle est connu ; deuxièmement, dans le but de vérifier la nécessité de l'approche non paramétrique quand on ignore la paramétrisation de θ , on estime le paramètre avec deux approches. Dans les deux cas, on suppose la même hypothèse qu'au chapitre 1, c'est-à-dire qu'on a n observations, $(x_1, \theta_1), \dots, (x_n, \theta_n)$ du même modèle et une observation récente (x, θ_{n+1}) pour laquelle on veut estimer son paramètre θ_{n+1} . On compare les deux méthodes, puis on tire une conclusion.

4.1 Approche paramétrique

Dans la section 1.4 du chapitre 1, on sait que l'estimateur du maximum de vraisemblance de λ est

$$\hat{\lambda} = \frac{1}{\bar{X}};$$

et que la distribution marginale de $X_{n+1}|\hat{\lambda}$ est la loi géométrique du paramètre

$$p = \frac{\hat{\lambda}}{1 + \hat{\lambda}};$$

et que l'estimateur bayésien empirique de θ_{n+1} est :

$$\begin{aligned}\delta^E(X_{n+1}) &= E_{\theta|\hat{\lambda}, X_{n+1}}(\theta_{n+1}) \\ &= \frac{(1 + X_{n+1})\bar{X}}{1 + \bar{X}}.\end{aligned}$$

D'après la section 1.2 du chapitre 1,

$$\delta^B(X_{n+1}) = \frac{1 + X_{n+1}}{1 + \lambda},$$

et le risque intégré de l'estimateur de Bayes est

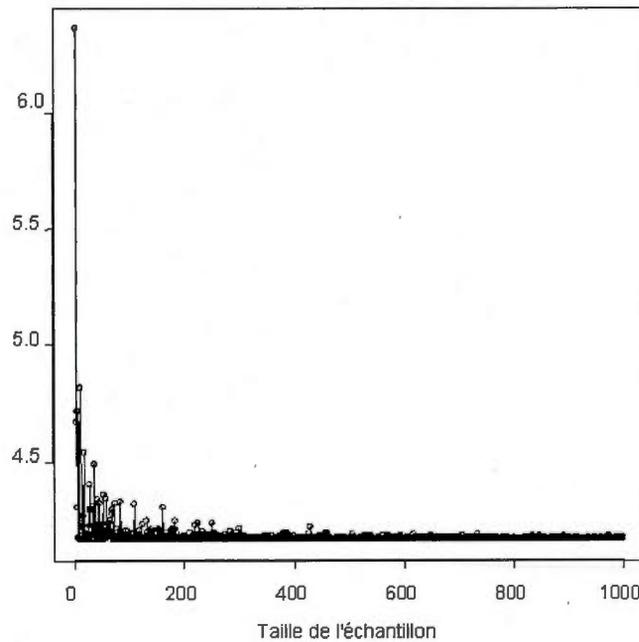
$$\begin{aligned}W(\delta^B) &= \int_0^\infty \sum_{x=0}^\infty \left(\frac{x+1}{\lambda+1} - \theta\right)^2 e^{-\theta} \frac{\theta^x}{x!} \lambda e^{-\lambda\theta} d\theta \\ &= \frac{1}{\lambda(1+\lambda)}.\end{aligned}$$

Ainsi

$$\begin{aligned}W(\delta^E) &= W(\delta^B) + E^{X_{n+1}|\hat{\lambda}}((\delta^E - \delta^B)^2) \\ &= W(\delta^B) + E\left(\left(\frac{(1 + X_{n+1})\bar{X}}{1 + \bar{X}} - \frac{(1 + X_{n+1})}{1 + \lambda}\right)^2\right) \\ &= \left(\frac{\bar{X}}{1 + \bar{X}} - \frac{1}{1 + \lambda}\right)^2 E((1 + X_{n+1})^2) \\ &= W(\delta^B) + \left(\frac{\bar{X}}{1 + \bar{X}} - \frac{1}{1 + \lambda}\right)^2 (Var((1 + X_{n+1})^2) + E^2(1 + X_{n+1})) \\ &= W(\delta^B) + \left(\frac{\bar{X}}{1 + \bar{X}} - \frac{1}{1 + \lambda}\right)^2 \left(\frac{1}{\lambda^2} + \frac{1}{\lambda} + \frac{1}{\lambda^2} + \frac{4}{\lambda} + 4\right) \\ &= \frac{1}{\lambda(\lambda+1)} + \left(\frac{\bar{X}}{1 + \bar{X}} - \frac{1}{1 + \lambda}\right)^2 \left(\frac{2}{\lambda^2} + \frac{5}{\lambda} + 4\right).\end{aligned}$$

Avant d'introduire la 2^e méthode d'estimation, on fait une étude de simulation en donnant une valeur à λ et en l'estimant par la méthode du maximum de vraisemblance décrite ci-haut.

Soit $\lambda = 0.2$. Pour cette valeur de $\lambda = 0.2$, on génère 1000 échantillons aléatoires de taille n pour tout n entre 2 et 1000 selon le modèle (4.1). Par la suite, on calcule δ^E , l'estimateur bayésien empirique de θ_{n+1} ainsi que le risque intégré correspondant, sachant



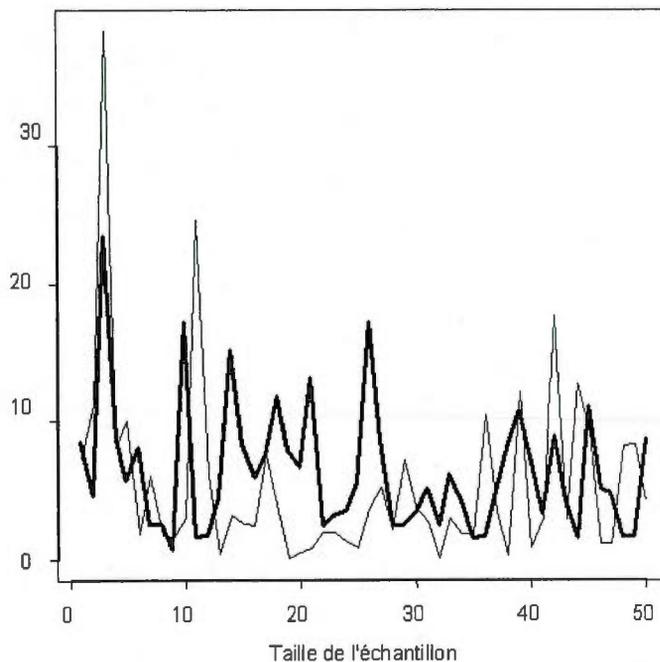
l'observation X_{n+1} afin de voir la performance de l'estimateur bayésien empirique de θ_{n+1} , c'est-à-dire, la variation du risque intégré en fonction de la taille de l'échantillon n . On obtient un graphique de $W(\delta^E)$ en fonction de n à la figure 4.1. Elle nous donne le comportement asymptotique de l'estimateur. Ainsi on remarque que la convergence du risque intégré asymptotique est très forte vers le risque de l'estimateur bayésien (qui est égal à $\frac{1}{\lambda(1+\lambda)} = 4.16$).

Le deuxième point d'intérêt est la comparaison de l'estimateur bayésien empirique avec la vraie valeur de θ_{n+1} en fonction de la taille de l'échantillon. On présente cette comparaison dans la figure 4.2. L'estimateur bayésien empirique est un trait foncé et celui de la vraie valeur de θ_{n+1} une ligne plus fine. Cette figure nous montre que les deux graphiques sont assez proches, un résultat souhaitable qui montre l'efficacité de l'estimateur.

Le programme A.0.7 en annexe A donne les différents calculs originaux pour le nouvel estimateur introduit dans la section 4.2, avec le logiciel *R*. Pour ces deux figures, nous avons utilisé l'approche paramétrique, c'est à dire qu'on suppose connue la famille de la loi *a priori*. On a supposé que cette loi est :

$$\theta|\lambda \sim \exp(\lambda) .$$

Figure 4.2 L'estimateur bayésien empirique et la vraie valeur de θ



Le tableau 4.1 donne une autre façon d'illustrer le comportement de l'estimateur bayésien empirique en fonction de la taille l'échantillon. Pour plusieurs valeurs de n et $\lambda = 2$ on génère $(X_1, \theta_1), \dots, (X_n, \theta_n)$ et $((X_{n+1}, \theta_{n+1}))$. Le tableau 4.1 donne pour les 4 différentes valeurs de n , θ_{n+1} et X_{n+1} ainsi que l'estimateur bayésien classique, l'estimateur bayésien empirique et son risque. Clairement, à partir de $n = 50$, le risque de l'estimateur bayésien empirique approche le risque de la règle de Bayes ($\simeq 4.16$).

Une autre approche qui démontre l'efficacité de l'estimateur bayésien empirique est de comparer les graphiques de la densité *a priori* exponentielle connue et l'estimateur de cette densité basé sur un échantillon des estimateurs bayésiens empiriques de θ ; c'est-à-dire, pour chaque $n = 1, \dots, 10000$, on génère n valeurs de θ par la méthode bayésienne empirique décrite ici. Donc, on aura $\hat{\theta}_1, \dots, \hat{\theta}_{10000}$, estimateurs de θ et on utilise une méthode non paramétrique pour estimer la fonction de densité de θ . On utilise la méthode d'estimation de noyau.

Tableau 4.1 Estimateur en fonction de la taille de l'échantillon

n	θ_{n+1}	X_{n+1}	δ^B	δ^E	$W(\delta^E)$
10	15.20730	9	3.3	7.6744	4.5096
50	3.802563	3	5.2	3.3548	4.1689
100	9.042732	8	5.19	7.5460	4.1687
1000	3.996712	3	4.814	4.1666	4.1689

Voici une description de cette méthode. Il faut noter que les $\hat{\theta}_i$ jouent le rôle des z_i dans cette méthode et que $k = n = 10000$. Soient z_1, \dots, z_n n observations indépendantes d'une variable aléatoire X continue de densité $f(z)$. La méthode d'estimation de noyau pour l'estimation de $f(x)$ proposée par Parzen(1962) et détaillée dans Maritz et Lwin (1989 chapitres 2 et 3), donne un estimateur de densité de noyau $f_n(z)$ donné par :

$$f_k(z) = \frac{1}{nh(n)} \sum_{i=1}^k K\left(\frac{z - z_i}{h(k)}\right).$$

Il y a plusieurs choix usuels pour le noyau K , par exemple, la loi gaussienne $\mathcal{N}(0, 1)$. La fenêtre $h(k)$ doit être choisie afin d'avoir le lissage désiré. Cette fenêtre doit vérifier les deux conditions suivantes afin d'assurer la convergence de la fonction de densité :

$$h(n) \rightarrow 0 \text{ et } nh(n) \rightarrow \infty .$$

Cette méthode est déjà programmée dans le logiciel R avec la commande " density(V, bw =0.18) " pour un vecteur de données V .

La figure 4.3 fait une comparaison entre la densité estimée de l'estimateur bayésien empirique (voir programme A.0.8 dans l'annexe) et la densité exponentielle avec le même paramètre. La figure montre qu'il y a une très grande efficacité de l'estimateur bayésien empirique quand on utilise la bonne loi *a priori*.

Par contre, on a répété une expérience analogue dont la vraie loi *a priori* était exponentielle mais on a supposé qu'elle est une loi gamma de même moyenne. La figure 4.4

Figure 4.3 L'estimation de la loi *a priori* utilisant le modèle exponentiel *a priori* versus la densité exponentielle avec paramètre $\lambda = 0.2$

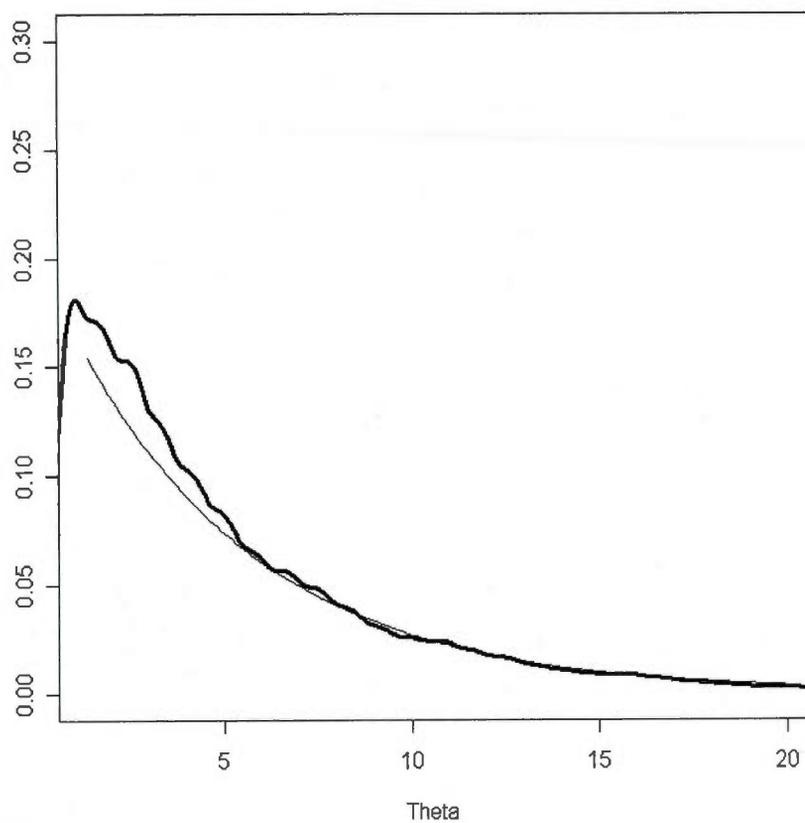
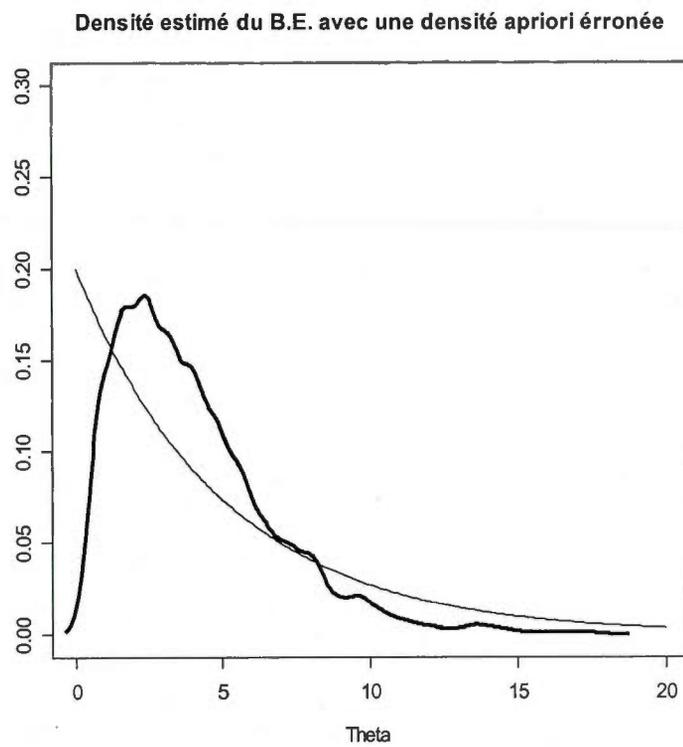


Figure 4.4 L'estimation de la densité *a priori* utilisant un modèle *a priori* gamma inapproprié versus la densité exponentielle avec paramètre $\lambda = 0.2$



démontre que l'estimateur de la densité *a priori* basé sur l'estimateur bayésien empirique est très loin de la vraie densité. Donc on peut conclure que dans le cas où on ne connaît pas la densité *a priori*, une méthode non paramétrique est préférable.

4.2 Approche non paramétrique

Nous décrivons ici un nouvel estimateur en utilisant une approche non paramétrique de l'estimateur bayésien empirique. L'estimateur de Bayes de la moyenne θ d'une loi de Poisson,

$$\delta_n(X_{n+1}) = \frac{f(X_{n+1} + 1)}{f(X_{n+1})} (X_{n+1} + 1),$$

est obtenu en estimant $f(x)$ à partir des observations. L'estimateur est donné par :

$$\delta_n(X_{n+1}) = \frac{(1 + X_{n+1})n(1 + X_{n+1})}{1 + n(X_{n+1})},$$

où f est la densité marginale de X_{n+1} et $n(X_i)$ dénote le nombre de fois qu'on a observé X_i pour $i \geq 1$.

Notre nouvel estimateur empirique bayésien, $\delta_n(X_{n+1})$ donne une autre façon d'estimer. On estime le rapport $\frac{f(1+X_{n+1})}{f(X_{n+1})}$ en se basant sur les anciennes observations X_1, \dots, X_n , sachant la $(n + 1)$ ième observation X_{n+1} .

Soit $y = \text{Max}(X_i)$, $1 \leq i \leq n$. La densité marginale f de X_{n+1} est inconnue. On estime f au point $(X_{n+1} + 1)$ et X_{n+1} de la façon suivante.

Le théorème d'accroissement fini en analyse dit que, pour une fonction réelle de classe C^1 dans un intervalle contenant x et y , il existe c entre x et y tel que la tangente en ce point soit parallèle à la droite qui joint $(x, f(x))$ et $(y, f(y))$, et on a $f(x) = f(y) + (x - y)f'(c)$. Donc il existe c_1 et c_2 telles que :

$$\begin{cases} f(X_{n+1}) = f(y) - (y - X_{n+1})f'(c_1) \\ f(1 + X_{n+1}) = f(y) - (y - X_{n+1})f'(c_2). \end{cases}$$

Puisque $y = \text{Max}(X_i)$ la valeur extrême que peut prendre X_{n+1} , alors les pentes des droites qui joignent les points $(X_{n+1}, f(X_{n+1}))$ et $(y, f(y))$ d'une part et

$(1 + X_{n+1}, f(1 + X_{n+1}))$ et $(y, f(y))$ d'autre part, sont presque parallèles, parce que $f(y) \simeq 0$ par rapport à $f(x)$. Donc :

$$\frac{f(1 + X_{n+1}) - (y - X_{n+1} - 1)f'(c_1)}{f(X_{n+1}) - (y - X_{n+1})f'(c_2)} \simeq \frac{(y - X_{n+1} - 1)}{(y - X_{n+1})}$$

Ainsi, on estime $\frac{f(1+X_{n+1})}{f(X_{n+1})}$ par $\frac{y-X_{n+1}-1}{y-X_{n+1}} = \frac{1}{1+1/K}$ où $y = \text{Max}(X_i)$, $1 \leq i \leq n$ et $K = \text{Max}(y - X_{n+1} - 1, 1)$.

Dans l'étude de simulation suivante, on obtient des résultats impressionnants du point de vue de l'erreur quadratique moyenne, EQM, en comparant le nouvel estimateur avec l'estimateur bayésien empirique non paramétrique ordinaire. Dans la prochaine section nous décrivons cette étude.

4.2.1 Étude de simulation

On considère d'abord le cas exponentiel avec les notations suivantes.

Soient :

- *BEP* : estimateur bayésien empirique paramétrique ;
- *BEN1* : estimateur bayésien empirique non paramétrique avec la 1^{ière} méthode ;
- *BEN2* : estimateur bayésien empirique non paramétrique avec la 2^{ième} méthode ;
- $EQM = \frac{1}{n} \sum_x (\delta(x) - \lambda)^2$, l'erreur quadratique moyenne d'un estimateur $\delta(x)$ de λ ;
- *EQM* : l'erreur quadratique moyenne de *BEP* ;
- *EQM1* : l'erreur quadratique moyenne de *BEN1* ;
- *EQM2* : l'erreur quadratique moyenne de *BEN2*.

Pour faire une comparaison entre les trois estimateurs: le *BEP* (bayésien empirique paramétrique), le *BEN1* (bayésien empirique non paramétrique avec la méthode classique) et le *BEN2* (notre estimateur), on génère chaque fois n échantillons avec $n = 10, 50$ et 100 et on calcule l'erreur quadratique moyenne (*EQM*) de chaque estimateur, selon le modèle exponentiel ou gamma. On suppose toujours que la variance est plus que 1 puisque les x suivent la loi de Poisson (si ce n'est pas le cas, on peut faire une transformation exponentielle par exemple pour obtenir la variance plus grande que 1). Ainsi, on

remarque que $BEN2$ est toujours meilleur que $BEN1$ et la différence devient de plus en plus grande quand la variance de θ croit. De plus, $BEN2$ a à peu près le même risque que BEP (parfois on trouve $EQM2 \leq EQM$, parfois $EQM \leq EQM2$), ce qui représente un résultat impressionnant car on a un estimateur pour lequel il n'est pas nécessaire d'estimer les paramètres de θ . Les résultats sont présentés dans le tableau 4.2.

Dans le deuxième tableau, on résume une expérience analogue mais pour la loi *a priori* gamma. On génère n ($n=10,50,100$) échantillons aléatoires selon la loi gamma, on calcule les deux estimateurs non paramétriques $BEN1$ et $BEN2$ et on fait le calcul des EQM . Aussi, on suppose toujours que la variance est plus que 1. On fixe alors $\alpha = 10$ et on fait varier β . On remarque que $BEN2$ est toujours meilleur que $BEN1$ et l'écart entre eux devient de plus en plus grand quand la variance devient de plus en plus grande. Les résultats sont présentés dans le tableau 4.3.

Nous effectuons une étude de simulation de grande taille, en simulant chaque fois 10^6 échantillons de taille n . On suppose ici que $\theta \sim \Gamma_{(\alpha,\beta)}$, on fixe $\beta = 2$ et on varie α aléatoirement selon la loi uniforme de sorte que la variance soit ≥ 1 . On calcule $EQM1$ et $EQM2$ et on estime la probabilité que $EQM2 \leq EQM1$, le nombre de fois que $EQM2 \leq EQM1$ divisé par 10^6 en variant les paramètres dans tous les cas. On trouve que cette probabilité est supérieure à 0.99. Le tableau 4.4 donne les résultats d'une simulation faite avec R qui indiquent que $BEN2$ est beaucoup meilleur que $BEN1$ dans ce cas.

Tableau 4.2 EQM pour les trois estimateurs en fonction de n et λ

λ	BEP	BEN1	BEN2
n=10			
2	3.1006673	9.8114100	1.2421020
5	6.574349	27.722092	4.252959
10	3.975605	179.625580	21.345856
20	5.999196	480.706648	22.343553
n=50			
2	1.158001	8.990074	1.844042
5	4.332596	37.534897	5.569131
10	4.809585	149.110019	15.01061
20	14.75369	546.77411	14.24742
n=100			
2	0.8266057	5.2200837	1.7495451
5	3.723064	22.944294	5.484415
10	8.058279	212.465898	9.213698
20	27.18515	801.48687	25.84790

Tableau 4.3 EQM pour les deux estimateurs en fonction de n dans le cas gamma

(α, β)	BEN1	BEN2	BEN1	BEN2	BEN1	BEN2
	n=10		n=50		n=100	
(10,3)	15.481697	2.106750	4.128589	1.475843	3.345140	2.066361
(10,2)	30.421124	1.927994	12.478132	2.335402	7.741661	2.901637
(10,1)	133.519198	7.673893	86.261734	4.508419	58.131561	7.411624

Tableau 4.4 $\Pr(EQM2 \leq EQM1)$ avec le nombre de simulations égal à 10^6

α	n=10	n=50	n=100
8	0.9983	0.9998	0.9998
unif[4,10]	0.9942	0.9952	0.9952
unif[4,50]	0.9992	0.9994	0.9994
unif[10,50]	0.9999	1	1

CONCLUSION

Nous avons présenté l'approche bayésienne empirique à l'estimation de la moyenne de la loi gaussienne, de la loi de poisson et de la loi exponentielle. Nous avons discuté l'approche bayésienne empirique non paramétrique de Robbins (1955, 1980) et l'approche bayésienne empirique paramétrique plus récente de Morris (1983) (aussi Efron et Morris (1973) et Casella (1985)).

Dans le cas de l'estimation gaussienne, nous avons développé une approche bayésienne empirique robuste inspirée de la méthodologie d'Angers (1992) dans le cas de l'estimation bayésienne hiérarchique. Nous avons illustré l'effet de la présence d'une valeur aberrante introduite dans les données de baseball, analysées par Casella (1985), sur l'estimateur robuste. Les résultats sont prometteurs. Nous avons considéré le cas où les observations sont arrangées dans des groupes de tailles inégales. Dans ce cas non équilibré, nous avons exploité le lien entre les estimateurs du type James-Stein et les estimateurs bayésiens empiriques pour obtenir un estimateur minimax approprié selon Berger et Bock (1976). Nous avons illustré la méthodologie avec la base de données publiée par Whitmore et Gentleman (1982) concernant l'effet d'une grève sur la survie de cellules d'aluminium.

Pour l'estimation d'une moyenne de la loi de Poisson, nous avons utilisé plusieurs méthodes bayésiennes empiriques et nous avons illustré leurs propriétés dans une étude de simulation. Nous avons utilisé l'approche proposée par Maritz et Lwin (1989) pour l'estimation dans le cas non équilibré.

Pour l'estimation de la moyenne de la loi exponentielle nous avons considéré les approches paramétriques et non paramétriques. Nous avons également introduit une nouvelle méthode pour l'estimation dans ce cas en utilisant les techniques d'analyse numérique.

Le nouvel estimateur a un comportement supérieur à ceux de quelques autres estimateurs déjà proposés dans la littérature.

En conclusion, nous trouvons que l'estimation bayésienne empirique est une méthode très utile pour l'estimation des moyennes de la loi gaussienne, de la loi de Poisson et de la loi exponentielle dans les cas équilibré et non équilibré.

APPENDICE A

PROGRAMMATION

Programme A.0.1 :

On génère 50 variables aléatoires selon le modèle hiérarchique de Poisson-Gamma. On note que le logiciel "R" a la même notation pour la densité de la loi Gamma que nous. Dans la simulation on prend $\alpha = 10$ et $\beta = 2$, pour évaluer la performance des différentes méthodes étudiées.

```
# On génère 50 variables aléatoires selon le modèle hiérarchique  
de Poisson-Gamma.
```

```
g<-1  
p<-1  
for (i in 1 :50)  
{  
g[i]<-rgamma(1,10,2)  
p[i]<-rpois(1,g[i])  
}
```

```
#calcul des fréquences  
gn<-40
```

```
#fn sert à calculer les fréquences f<- 0.01734153
```

```

#f(x)la densité marginale de X

m<-max(p)
x<-c(0, 1:16)
for(i in 1 :m)
{
gn[i]<-sum(1-(abs(p-i+1))/((1+abs(p-i+1)+abs(1-abs(p-i+1)) )/2))
}
gn[m+1]<-50-sum(gn)
  for(i in 1:17)
  {
f[i]<-((gamma(i+9))/(factorial(i-1)*(gamma(10)))) *((2/3)^10)*
((1/3)^(i-1))
  }
f[0]<- 0.01734153
fn<-c(1 :17)
fn<-0*fn

for(i in 1:17)
  {
  if (i <= m+1)
  {
fn[i]<-gn[i]
  }
else {fn[i]=0
}
}

# estimateurs empiriques

```

```

B<-function(i)
  {b<-(i+10)/3
  b
  }
BEn<-function(i)
  {be<-(i+1)*fn[i+1]/(1+fn[i+1])
  be
  }
BEa<-function(i)
  {be<- 3.487+0.33*i
  be
  }
BEb<-function(i)
  {be<- 2.96+0.433*i
  be
  }
# Fonction de masse de loi de Poisson (P(u,x))
P<-function(u,x)
  {p<- (exp(-u)*(u^x))/factorial(x)
  p
  }
# fonction moyenne de 3 Poisson
G<- function(a,b,c,x)
  {g<-(P (a,x)+ P(b,x)+P(c,x))/3
  g
  }
#L'estimateur de Bayes empirique
Bayes<- function(a,b,c,x)
  {g<-(a*P (a,x)+ b*P(b,x)+c*P(c,x))/(3*G(a,b,c,x))
  g
  }

```

```

    }

    BEc<- function(i)
  { be<- Bayes(4.151728, 4.151728, 7.48663,i)
    be
  }

    BEd<- function(i)
  { be<- Bayes(5.620835, 2.749312, 8.482241,i)
    be
  }

  # Calcul des risques intégrés des estimateurs de Bayes empiriques

  Wa<-1.67+sum(f*(BEa(x)-B(x))^2)
  Wa
  Wb<-1.67+sum(f*(BEb(x)-B(x))^2)
  Wb
  Wc<-1.67+sum(f*(BEc(x)-B(x))^2)
  Wc
  Wd<-1.67+sum(f*(BED(x)-B(x))^2)
  Wd
  Wn<-1.67+sum(f*(BEn(x)-B(x))^2)
  Wn

  #Rassemblement de tous les calculs dans un seul tableau.
  Table2<-data.frame(x,fn,f,B(x),BEn(x),BEa(x),BEb(x),BEc(x),BED(x))
  Table2

```

Programme A.0.2 : *Maximum de vraisemblance avec Maple pour estimer $\hat{\beta}$.*

#le vecteur d'observations qu'on a simulées dans le programme précédent:

```
X:=( [10.0, 5.0, 5.0, 4.0, 2.0, 4.0, 3.0, 5.0, 3.0, 5.0, 6.0,
4.0, 7.0, 7.0, 3.0, 3.0, 8.0, 5.0, 10.0, 6.0, 5.0, 8.0, 9.0,
6.0, 4.0, 3.0, 4.0, 6.0, 4.0, 12.0, 14.0, 2.0, 4.0, 0.0,
10.0, 4.0, 6.0, 4.0, 5.0, 5.0, 7.0, 2.0, 3.0, 3.0, 4.0, 4.0,
1.0, 2.0, 7.0, 8.0] );
```

```
# Log de la vraisemblance (On ignore les facteurs qui ne dépendent
pas de beta):
```

```
l:= proc (beta) options operator, arrow;
sum('5.22* beta*log(beta)-log(GAMMA(5.22* beta))
+log(GAMMA(X[i]+5.22* beta))-(X[i]+5.22* beta)*log(beta+1)',
'i' = 1 .. 50)
end proc ;
```

```
#Dérivation de la vraisemblance:
```

```
ld1 := diff(l(beta), beta);
```

```
#Résolution de la valeur de beta:
```

```
evalf(solve({ld1 = 0}, beta));
```

Programme A.0.3 : *Maximum de vraisemblance selon la méthode de Parzen, Maritz et Lwin (1989). Ce programme estime les 3 paramètres λ_1 , λ_2 , et λ_3 de la densité marginale $f(x)$ définie dans le paragraphe 3.2 (c).*

```

p<-c(10, 5, 5, 4, 2, 4, 3, 5, 3, 5, 6, 4, 7, 7, 3, 3, 8, 5, 10,
6, 5, 8, 9, 6, 4, 3, 4, 6, 4, 12, 14, 2, 4, 0, 10, 4, 6, 4, 5, 5, 7,
2, 3, 3, 4, 4, 1, 2, 7, 8)
#les observations
x<-c(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16)
# les effectifs
fn<-c(1, 1, 4, 7, 11, 8, 5, 4, 3, 1, 3, 0, 1, 0, 1, 0, 0)
#Initiation des valeurs des lambdas
u10<-1
u20<-2
u30<-3
u1<-u10
u2<-u20
u3<-u30
#Itération avec la méthode de Newton
for(j in 1:1000)
{
u1[j+1]<-(sum(x*fn*u1[j]^(x-1)/(exp(-u1[j])*u1[j]^x+
exp(-u2[j])*u2[j]^x+ exp(-u3[j])*u3[j]^x)))/sum(fn*
u1[j]^(x-1)/(exp(-u1[j])*u1[j]^x+ exp(-u2[j])*u2[j]^x+
exp(-u3[j])*u3[j]^x))
u2[j+1]<-(sum(x*fn*u2[j]^(x-1)/(exp(-u1[j])*u1[j]^x+ exp(-u2[j])*u2[j]^x+
exp(-u3[j])*u3[j]^x)))/sum(fn* u2[j]^(x-1)/(exp(-u1[j])*u1[j]^x+
exp(-u2[j])*u2[j]^x+ exp(-u3[j])*u3[j]^x))
u3[j+1]<-(sum(x*fn*u3[j]^(x-1)/(exp(-u1[j])*u1[j]^x+ exp(-u2[j])*u2[j]^x+
exp(-u3[j])*u3[j]^x)))/sum(fn* u3[j]^(x-1)/(exp(-u1[j])*u1[j]^x+
exp(-u2[j])*u2[j]^x+ exp(-u3[j])*u3[j]^x))
}

```

Programme A.0.4 : *Méthode de khi-deux (Maritz (1966))*. Ce programme estime les 3 paramètres λ_1 , λ_2 , et λ_3 de la densité marginale $f(x)$ définie dans le paragraphe 3.2 (d).

```
# On définit plusieurs fonctions dont on a besoin dans notre
```

```
#calculs
```

```
#Fonction de masse d'une loi de Poisson
```

```
P<-function(u,x)
{
p<- (exp(-u)*(u^x))/factorial(x)
P
}
```

```
#4) fonction moyenne de 3 exponentiels
```

```
G<- function(a,b,c,x)
{
g<-(P (a,x)+ P(b,x)+P(c,x))/3
g
}
```

```
#L'estimateur de Bayes
```

```
Bayes<- function(a,b,c,x)
{
g<-(a*P (a,x)+b*P(b,x)+c*P(c,x))/(3*G(a,b,c,x))
g
}
```

```
#Regroupement des fréquences
```

```
y<-c(2,3,4,5,6,7,8,14)
```

```

yn<-c(6,7,11,8,5,4,3,6)
#fonction G3 (regroupement des fréquences
#théoriques correspondantes)

g3<-function(a,b,c,x)
{
  if (x==2)
g<-G(a,b,c,0) + G(a,b,c,1)+ G(a,b,c,2) if (x==14) g<-
G(a,b,c,9)+G(a,b,c,10)+ G(a,b,c,11)+ G(a,b,c,12)+ G(a,b,c,13)+
G(a,b,c,14)
  if (3<=x &x<=8)
g<-G(a,b,c,x)
  g
}

#G3
G3<-function(a,b,c,y)
{
  k<-length(y)
  h<-rep(0,k)
  for (i in 1:k)
  {
    h[i]<-g3(a,b,c,y[i])
  }
  h
}

#Fonctions intermédiaires (H1 et H2)
#h1 h1<-function(u,x)
{
  if (x==14)

```

```

h<-(P(u,9)*9+P(u,10)*10+P(u,11)*11+ P(u,12)*12+ P(u,13)*13+
P(u,14)*14)/u
if (2<=x & x<=8)
h<-(P(u,x)*x)/u
h
}

#H1
H1<-function(u,y)
{
k<-length(y)
h<-rep(0,k)
for (i in 1:k)
{
h[i]<-h1(u,y[i])
}
h
}

#h2
h2<-function(u,x)
{
if (x==14)

h<-( P(u,9)+P(u,10)+ P(u,11)+ P(u,12)+ P(u,13)+ P(u,14))/u
if (1<=x & x<=8)
h<-P(u,x)/u
h
}

#H2
H2<-function(u,y)
{

```

```

k<-length(y)
h<-rep(0,k)
for
(i in 1:k)
{
h[i]<-h2(u,y[i])
}
h
}

#Prendre la dérivée de la loi de Poisson dP(u,x)

dP<-function(u,x)
{
p<-(exp(-u)*(u^(x-1))/factorial(x))*(x-u)
p
}

#Dérivée regroupée (DP3)

dP3 <-function(u,x)
{
if (x==2)

h<- dP(u,0)+dP(u,1)+ dP(u,2)
if (x==14)

h<- dP(u,9)+ dP(u,10)+ dP(u,11)+dP(u,12)+ dP(u,13) +dP(u,14)
if (2<=x & x<=8)
h<-dP(u,x)
h
}

```

```

}

DP3<-function(u,y)
{
k<-length(y)
h<-rep(0,k)
for (i in 1:k)
{
h[i]<-dP3(u,y[i])
}
h
}

#Valeurs initiales

A<-matrix(0,nrow=3,ncol=3)
u<-c(3.0,5.0,7.0)
z<-c(1 :8)
b<-c(1 :3)

#Matrice des résultats

chi2<- data.frame(matrix(sample(64),nrow=8))

dimnames(chi2) <-list(rownames=c('g1','g2','g3','g4','g5', 'g6',
'g7', 'g8'),colnames=c("y","yn"," DP3(u[1],y)"," DP3(u[2],y)","
DP3(u[3],y)","G3(u[1],u[2],u[3],y)", "v","w"))

chi2[,1]<-y chi2[,2]<-yn chi2[,3]<-DP3(u[1],y)
chi2[,4]<-DP3(u[2],y) chi2[,5]<-DP3(u[3],y)
chi2[,6]<-G3(u[1],u[2],u[3],y)

```

```

# Calcul du vecteur u

for(i in 1:3)
  {for(j in 1:3)
    {A[i,j]<- (50/9)*sum(DP3(u[i],y)*DP3(u[j],y)/ G3(u[1],u[2],u[3],y))
      }
    }
  for (i in 1:8) {
    z[i]<-yn[i]/50- G3(u[1],u[2],u[3],y[i])
  }
  for (i in 1:3){
    b[i]<-(50/3)*sum(z*DP3(u[i],y)/G3(u[1],u[2],u[3],y))
  }

print(solve(A)%*%b)
u<-u+ solve(A)%*%b
print(u)

```

Programme A.0.5 *L'estimateur minimax du théorème 2.1 pour la base de données de la section 2.6 pour la durée de vie des cellules d'aluminium*

```

#Calcul de tau pour la Catégorie 1

#On génère des Gamma pour chaque groupe pour évaluer
#les'intégrales

A1<-rgamma(1000000,8.5,0.5)
A2<-rgamma(1000000,2.5,0.5)
A3<-rgamma(1000000,10.5,0.5)
A4<-rgamma(1000000,9.5,0.5)

```

```
A5<-rgamma(1000000,6.5,0.5)
A6<-rgamma(1000000,4,0.5)
A7<-rgamma(1000000,3.5,0.5)
A8<-rgamma(1000000,3,0.5)

#Estimation des intégrales avec la méthode de Monte-Carlo pour le
#calcul de tau

mc1<-(17/A1)*(1-pgamma(A1,2.5,0.5))*(1-pgamma(A1,10.5,0.5))*
(1-pgamma(A1,9.5,0.5))*(1-pgamma(A1,6.5,0.5))*(1-pgamma(A1,4,0.5))
*(1-pgamma(A1,3.5,0.5))*(1-pgamma(A1,3,0.5))

mc2<-(5/A2)*(1-pgamma(A2,8.5,0.5))*(1-pgamma(A2,10.5,0.5))*
(1-pgamma(A2,9.5,0.5))*(1-pgamma(A2,6.5,0.5))*(1-pgamma(A2,4,0.5))
*(1-pgamma(A2,3.5,0.5))*(1-pgamma(A2,3,0.5))

mc3<-(21/A3)*(1-pgamma(A3,8.5,0.5))*(1-pgamma(A3,2.5,0.5))*
(1-pgamma(A3,9.5,0.5))*(1-pgamma(A3,6.5,0.5))*(1-pgamma(A3,4,0.5))
*(1-pgamma(A3,3.5,0.5))*(1-pgamma(A3,3,0.5))

mc4<-(19/A4)*(1-pgamma(A4,8.5,0.5))*(1-pgamma(A4,2.5,0.5))*
(1-pgamma(A4,10.5,0.5))*(1-pgamma(A4,6.5,0.5))*(1-pgamma(A4,4,0.5))
*(1-pgamma(A4,3.5,0.5))*(1-pgamma(A4,3,0.5))

mc5<-(13/A5)*(1-pgamma(A5,8.5,0.5))*(1-pgamma(A5,2.5,0.5))*
(1-pgamma(A5,10.5,0.5))*(1-pgamma(A5,9.5,0.5))*(1-pgamma(A5,4,0.5))
*(1-pgamma(A5,3.5,0.5))*(1-pgamma(A5,3,0.5))

mc6<-(8/A6)*(1-pgamma(A6,8.5,0.5))*(1-pgamma(A6,2.5,0.5))*
(1-pgamma(A6,10.5,0.5))*(1-pgamma(A6,9.5,0.5))*(1-pgamma(A6,6.5,0.5))
```

```

*(1-pgamma(A6,3.5,0.5))*(1-pgamma(A6,3,0.5))

mc6<-(8/A6)*(1-pgamma(A6,8.5,0.5))*(1-pgamma(A6,2.5,0.5))*
(1-pgamma(A6,10.5,0.5))*(1-pgamma(A6,9.5,0.5))*(1-pgamma(A6,6.5,0.5))
*(1-pgamma(A6,3.5,0.5))*(1-pgamma(A6,3,0.5))

mc7<-(7/A7)*(1-pgamma(A7,8.5,0.5))*(1-pgamma(A7,2.5,0.5))*
(1-pgamma(A7,10.5,0.5))*(1-pgamma(A7,9.5,0.5))*(1-pgamma(A7,6.5,0.5))
*(1-pgamma(A7,4,0.5))*(1-pgamma(A7,3,0.5))

mc8<-(6/A8)*(1-pgamma(A8,8.5,0.5))*(1-pgamma(A8,2.5,0.5))*
(1-pgamma(A8,10.5,0.5))*(1-pgamma(A8,9.5,0.5))*(1-pgamma(A8,6.5,0.5))
*(1-pgamma(A8,4,0.5))*(1-pgamma(A8,3.5,0.5))
mean(mc1+mc2+mc3+mc4+mc5+mc6+mc7+mc8)

#l'estimateur minimax On a 8 groupes, n représente la taille des
#groupes, v l'écart-type de chaque groupe et y les moyennes.

n1<-c(17,5,21,19,13,8,7,6)

v1<-c(0.4190,0.2211,0.3644,0.3648,0.2777,0.1296, 0.3770, 0.1098)

y1<-c(7.1420, 7.1492, 7.0727, 6.8926, 6.8616,6.7820,6.5310,6.8205)

p<-8
t1<- 2.694656
mx1<-(1-2*(p-2*t1)/((n1*(v1^2)/(n1-2))*sum(((y1^2)*(n1-2)^2)/((n1^2)*v1^4))))*y1
mx1

```

Calcul de tau Pour la Catégorie 2

#On génère des Gamma pour chaque groupe pour évaluer
#les'intégrales

B1<-rgamma(1000000,1.5,0.5)

B2<-rgamma(1000000,8,0.5)

B3<-rgamma(1000000,10,0.5)

B4<-rgamma(1000000,10,0.5)

B5<-rgamma(1000000,11,0.5)

B6<-rgamma(1000000,11.5,0.5)

B7<-rgamma(1000000,13,0.5)

B8<-rgamma(1000000,12.5,0.5)

#Estimation des intégrales avec la méthode de Monte-Carlo pour le
#calcul de tau

Mc21<-(3/B1)*(1-pgamma(B1,8,0.5))*(1-pgamma(B1,10,0.5))*
(1-pgamma(B1,10,0.5))*(1-pgamma(B1,11,0.5))*(1-pgamma(B1,11.5,0.5))
(1-pgamma(B1,13,0.5))(1-pgamma(B1,12.5,0.5))

mc22<-(16/B2)*(1-pgamma(B2,1.5,0.5))*(1-pgamma(B2,10,0.5))
(1-pgamma(B2,10,0.5))(1-pgamma(B2,11,0.5))*(1-pgamma(B2,11.5,0.5))
(1-pgamma(B2,13,0.5))(1-pgamma(B2,12.5,0.5))

mc23<-(20/B3)*(1-pgamma(B3,1.5,0.5))*(1-pgamma(B3,8,0.5))*
(1-pgamma(B3,10,0.5))*(1-pgamma(B3,11,0.5))*(1-pgamma(B3,11.5,0.5))
(1-pgamma(B3,13,0.5))(1-pgamma(B3,12.5,0.5))

mc24<-(20/B4)*(1-pgamma(B4,1.5,0.5))*(1-pgamma(B4,8,0.5))*

```
(1-pgamma(B4,10,0.5))*(1-pgamma(B4,11,0.5))*(1-pgamma(B4,11.5,0.5))
*(1-pgamma(B4,13,0.5))*(1-pgamma(B4,12.5,0.5))
```

```
mc25<-(22/B5)*(1-pgamma(B5,1.5,0.5))*(1-pgamma(B5,8,0.5))*
(1-pgamma(B5,10,0.5))*(1-pgamma(B5,10,0.5))*(1-pgamma(B5,11.5,0.5))
*(1-pgamma(B5,13,0.5))*(1-pgamma(B5,12.5,0.5))
```

```
mc26<-(23/B6)*(1-pgamma(B6,1.5,0.5))*(1-pgamma(B6,8,0.5))*
(1-pgamma(B6,10,0.5))*(1-pgamma(B6,10,0.5))*(1-pgamma(B6,11,0.5))
*(1-pgamma(B6,13,0.5))*(1-pgamma(B6,12.5,0.5))
```

```
mc27<-(26/B7)*(1-pgamma(B7,1.5,0.5))*(1-pgamma(B7,8,0.5))*
(1-pgamma(B7,10,0.5))*(1-pgamma(B7,10,0.5))*(1-pgamma(B7,11,0.5))
*(1-pgamma(B7,11.5,0.5))*(1-pgamma(B7,12.5,0.5))
```

```
mc28<-(25/B8)*(1-pgamma(B8,1.5,0.5))*(1-pgamma(B8,8,0.5))*
(1-pgamma(B8,10,0.5))*(1-pgamma(B8,10,0.5))*(1-pgamma(B8,11,0.5))
*(1-pgamma(B8,11.5,0.5))*(1-pgamma(B8,13,0.5))
```

```
t2<-mean(mc21+mc22+mc23+mc24+mc25+mc26+mc27+mc28)
```

#L'estimateur minimax, On a 8 groupes, n représente la taille des
#groupes, v l'écart-type de chaque groupe et y les moyennes.

```
mx<-(1-2*(p-2*t)/((n*(v^2)/(n-2))*sum(((y^2)*(n-2)^2)/((n^2)*v^4))))*y
```

```
mx
```

```
n2<-c(3,16,20,20,22,23,26,8)
```

```
v2<-c(0.0549,0.0780,0.0915,0.0970,0.1364,0.1391,0.2200,0.0958)
```

```

y2<-c(7.7811,7.6125,7.4789,7.4427,7.3627,7.2596,7.3001,7.3000)

p<-8
t2<- 2.993452
mx2<-(1-2*(p-2*t2)/((n2*(v2^2)/(n2-2))*sum(((y2^2)*(n2-2)^2)/((n2^2)*v2^4))))*y2
mx2

#Sortie de tous les calculs dans un seul tableau

Table2<-data.frame(n1,v1,y1,mx1,n2,v2,y2,mx2)
Table2

```

Programme A.0.6 *Simulation de taille inégale dans le cas de Poisson paragraphe 3.3, pour l'estimation de la loi a priori G.*

On note n le nombre d'échantillons, m_i les tailles des groupes, $al=\alpha$, $be=\beta$, $m=(m_1, \dots, m_n)$.

On génère n tailles d'échantillons entre 50 et 1000, pour chaque i en génère m_i variables qui suivent la loi de Poisson avec ($\alpha = 20, \beta = 0.5$) et $1 \leq i \leq n$. T_i est la somme des P_{ij} .

```

n<-10
al<-20
be<-0.5
a<-c(1:n)
t<-list(a)
T<-a
m<-sample(50:1000,n)
p<-rgamma(n,al,be)
for(i in 1:n)

```

```

{ t[[i]]<-rpois(m[i],p[i])

  T[i]<-mean(t[[i]])
}

#calcul de l'estimateur beta

b<-n*mean(T)/( sum(T^2)-mean(T)*(sum (1/m)) -n*(mean(T))^2)

# calcul de l'estimateur alpha
a<-b*mean(T)
a
b

#Mettre les données dans un tableau

Poisson<-data.frame(n,p,T,p^2+p/n,T^2)
dimnames(Poisson)<-list(c(1 :10)c("n", "p", "T", "( p^2+p/n) ", "
(p^2+p/n )" ))
Poisson

```

Programme A.0.7 *Cas exponentiel du chapitre 4*

```

#On définit quelques fonctions dont on a besoin pour notre
#estimateur

# Indicateur J
J1<-function(x,y){a<-0

```

```

if (-0.5<=(x-y) & (x-y) < 0.5) a<-1
a
}
J<-function(u,y)
{k<-length(y)
h<-rep(0,k)
for (i in 1:k) {h[i]<-J1(u,y[i])}
h}
# Indicateur K
K1<-function(x,y,z){a<-0
if (x<= y & y<= z) a<-1
if (z<= y & y<= x) a<-1
a
}
K<-function(u,y,z)
{k<-length(y)
h<-rep(0,k)
for (i in 1:k) {h[i]<-K1(u,y[i],z)}
h}

```

```

#Calculs des différents estimateurs Bayésiens empiriques et
#l'estimateur de Bayes détaillés dans les paragraphes 4.1 et 4.2

```

```

BE10<-function(n,a)
{ b1<-rep(0,n)
be2<-rep(0,n)
m3<-rep(0,n)
wb4<-rep(0,n)
wbe5<-rep(0,n)
bp6<-rep(0,n)

```

```

b7<-rep(0,n)
b8<-rep(0,n)
ben9<-rep(0,n)
bn10<-rep(0,n)
eqmp11<-0
eqmn12<-0
benh<-rep(0,n)
eqmah<- rep(0,n)
eqmh<-0
eqmax<- rep(0,n)
eqmx<-0
for(i in 1:n)
{
theta<-rexp(i,a)
x<-rpois(length(theta),theta)
t<-rexp(1,a)
  xi<-rpois(1,t)
  m<-max(x,1)
  xb<-mean(x)
pn<-sum(I(xi,x))
pm<-sum(I(xi+1,x))
  hn<-sum(K(xi,x,xb))
  hm<-sum(K(xi+1,x,xb))
m<-max(x,1)
xb<-mean(x)
b1[i]<-(xi+1)/(1+a)
be2[i]<-(1+xi)*xb/(1+xb)
m3[i]<-xb
wb4[i]<- 1/(a*(1+a))
wbe5[i]<- 1/(a*(1+a))+(2/(a^2)+5/a+4)*(xb/(1+xb)-1/(1+a))^2

```

```

bp6[i]<-(t-be2[i])^2
b7[i]<-t
b8[i]<-xi
ben9[i]<-(1+xi)*pm/(1+pn)
bn10[i]<- (t-ben9[i])^2

# Calculs des erreurs quadratiques moyennes pour chaque estimateur

eqmax[i]<-(t-xi)^2 benh[i]<-(1+xi)*hm/(1+hn)
eqmah[i]<-(t-benh[i])^2
}
w<-list(b1=b1,be2=be2,m3=m3,wb4=wb4,wbe5=wbe5,bp6=bp6,b7=b7,b8=b8,
ben9=ben9, bn10=bn10,eqmp11=mean(bp6),
eqmn12=mean(bn10), eqmx=mean(eqmax), benh= benh, eqmh= mean(eqmah))
w
}

#Graphique de l'estimateur en fonction de la taille de
#l'échantillon

x<-seq(2,1000) plot(x,BE10(1000,0.2)$wbe5,type="o",xlab ="Taille
de l'échantillon ", ylab=" ")

title("Risque intégré de l'estimateur bayésien empirique ")

#Graphique de l'estimateur bayésien empirique et theta

x<-seq(1,50) plot(x,BE6(50,0.2)$b7,type="l",xlab ="Taille de
l'échantillon ", ylab=" ")

```

```

points(x,BE6(50,0.2)$be2, type="l",lwd=3)
title("L'estimateur bayésien empirique et la vraie valeur de
theta")

# col="blue" pour la couleur bleue

# pour que le tracé soit plus foncé lwd=2

```

Programme A.0.8 *Graphique de la densité estimée et la densité exponentielle.*

```

h<-function(x)
{
  b<-dgamma(x,shape=1,scale=5)
  b
}
H<-function(x)
{
  k<-length(x)
  b<-rep(0,k)
  for (i in 1:k)
  {
    b[i]<-h(x[i])
  }
  b
}

my.tn <- function()
{
  a<-0.2
  theta <- rgamma(10000,1,a)

```

```

x <- rpois(length(theta),theta)
t <- rgamma(1,1,a)
xn <- rpois(1,t)
tn <- (1+xn)*mean(x)/(1+mean(x))
return(tn)
}

#replicate(10, my.tn())
th1Ex<-replicate(10000, my.tn())
d1<-density(th1Ex, type="l",bw =0.4)
x<-seq(0,20,0.1)
plot(x,H(x),type="l", xlab ="Theta ", ylab=" ", xlim=c(0,20),
ylim=c(0,0.3))
points (d1,type="l", lwd=3)

title("Densité estimée du B.E. et la loi exponentielle")

```

Programme A.0.9 *Graphique avec loi a priori erronée*

```

my.tn <- function()
{
theta <- rgamma(10000,2,0.5)
x <- rpois(length(theta),theta)
t <- rgamma(10000,2,0.5)
xn <- rpois(1,t)
tn <- (1+xn)*mean(x)/(1+mean(x))
return(tn)
}

```

```

}
replicate(100, my.tn())
th<-replicate(1000, my.tn())
th
dthEP<- density(th, bw = 0.4)
x<-seq(0,20,0.1)
plot(x,H(x),type="l", xlab ="Theta ", ylab=" ",
xlim=c(0,20), ylim=c(0,0.3)) points(dthEP, type="l", lwd=3)
title("Densité estimé du B.E. avec une densité apriori érronée")

```

Programme A.0.10 *Calcul des estimateurs $\hat{\mu}$ et $\hat{\tau}^2$ pour le modèle robuste (paragraphe 2.3).*

Le but est d'estimer les paramètres τ^2 et μ du modèle robuste avec la méthode du maximum de vraisemblance basée sur 7 observations, les dérivées partielles (par rapport à τ^2 et μ) égalant zéro. On déduit (τ^2 et μ) en fonction de (τ^2 et μ), puis on procède avec la méthode de point fixe en faisant des itérations successives jusqu'à la convergence. $t = \tau^2$, $s = \sigma^2$ $u =$ la moyenne $\alpha =$ degré de liberté

```
#Programme lorsque le degré de liberté est 6.
```

```
# les valeurs initiales des paramètres
```

```
s<- 0.00434281
```

```
u0<-0.1
```

```
t0<- 0.001
```

```
t<-rbind(t0)
```

```
u<-rbind(u0)
```

```
m<-numeric(7)
```

```
Ap<-numeric(7)
```

```
Bp<-numeric(7)
Cp<-numeric(7)
Dp<-numeric(7)
A<-numeric(7)
B<-numeric(7)
C<-numeric(7)
D<-numeric(7)

# les données de baseball

x<-c(0.395, 0.355, 0.313, 0.291, 0.247,0.224, 0.175)

# On génère 100000 Gamma(alpha/2, alpha/2) pour évaluer les
#intégrales avec Monte-Carlo

g<-rgamma(100000,3,3)

a<-s+t/g
# Approximation des intégrales avec la méthode de
#Monte-Carlo
for(i in 1:7)
{

m[i]<-mean((1/sqrt(a))*exp((-1/(2*a))*(x[i]-u)^2))

Bp[i]<- mean((1/(sqrt(a^3)))*exp((-1/(2*a))*(x[i]-u)^2))

Ap[i]<-x[i]*Bp[i]
```

```
Cp[i]<-mean((1/(2*g*sqrt(a^3)))*exp((-1/(2*a))*(x[i]-u)^2))
```

```
Dp[i]<-mean( ( ((x[i]-u)^2)/(2*g*sqrt(a^5))
)*exp((-1/(2*a))*(x[i]-u)^2) )
```

```
A[i]<-Ap[i]/m[i]
```

```
B[i]<-Bp[i]/m[i]
```

```
C[i] <-Cp[i]/m[i]
```

```
D[i] <-Dp[i]/m[i]
```

```
}
```

```
# Pour terminer, on calcule des estimateurs t et u
```

```
u<-sum(A)/sum(B)
```

```
t<-t*sum(D)/sum(C)
```

```
u
```

```
t
```

```
# On refait les itérations successives jusqu'à la convergence
```

```
g<-rgamma(100000,3,3) a<-s+t/g for(i in 1:7) {
```

```
m[i]<-mean((1/sqrt(a))*exp((-1/(2*a))*(x[i]-u)^2))
```

```
Bp[i]<- mean((1/(sqrt(a^3)))*exp((-1/(2*a))*(x[i]-u)^2))

Ap[i]<-x[i]*Bp[i]

Cp[i]<-mean((1/(2*g*sqrt(a^3)))*exp((-1/(2*a))*(x[i]-u)^2))

Dp[i]<-mean( ( ((x[i]-u)^2)/(2*g*sqrt(a^5))
)*exp((-1/(2*a))*(x[i]-u)^2) )

A[i]<-Ap[i]/m[i]

B[i]<-Bp[i]/m[i]

C[i] <-Cp[i]/m[i]

D[i] <-Dp[i]/m[i]

}

u<-sum(A)/sum(B)
t<-t*sum(D)/sum(C)

u
t
```

RÉFÉRENCES

- Abramowitz, M. et I.A. K. Stegun 2002. « *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* ». U.S. : Department of Commerce.
- Angers, J. -F. 1992. « Use of Student-t prior for the estimation of normal means : A computational approach. ». *Bayesian Statistics* , Vol. 4, Oxford, pp. 567-575.
- Angers, J. -F. et B. MacGibbon 1996. « Protection against Outliers in Bayesian Linear Models with Econometric Applications. ». *Advances in Econometrics* , Vol. II, A, pp. 211-229.
- Berger, J. O. et M. E. Bock 1976. « Combining Independent Normal Mean Estimation Problems with Unknown Variances ». *The Annals of Statistics* , Vol. 4, N. 3, p. 642-648.
- Berger, J. O. 1985 « *Statistical Decision Theory and Bayesian Analysis* ». New York : Springer.
- Brown, L. D. 1976 « *Notes on Statistical Decision Theory* ». New York : Unpublished Lecture Note.
- Casella, G. 1985. « An Introduction to Empirical Bayes Data Analysis ». *American Statistics*, Vol. 39, No. 2, p. 83-87.
- Casella, G. et R. Berger. 1990. « *Statistical Inference* ». Pacific Grove, CA : Duxbury Advanced Series.
- ramer, H. 1946. « *Mathematical Methodes of Statistics* ». Princeton University Press, CA : Princeton, NJ.
- D'Agostino, R. B. et M. A. Stephens. 1986. « *Goodness-of-Fit Techniques* ». New York : Marcel Dekker.
- Devroye, L. et L. Gyorfi. 1986. « *Nonparametric Density Estimation* ». New York : Wiley Series in Probability and Mathematical Statistics.
- Douglas G. C. et Robbins, H. 1951. « Minimum Variance Estimation Without Regularity Assumptions. ». *The Annals of Mathematical Statistics* , Vol. 22, N. 4, p. 581-586.
- Efron, B. et C. Morris. 1973 (a). « Empirical Bayes on Vector Observations : An Extension of Stein's Method ». *Biometrika*, Vol. 59, p. 335-347.

- Efron, B. et C. Morris. 1973 (b). « Stein's Estimation Rule and its Competitors—An Empirical Bayes Approach ». *Journal of American Statistics*, Vol. 68, p. 117-130.
- Efron, B. et C. Morris. 1975. « Data Analysis Using Stein's Estimator and its Generalizations ». *Journal of the American Statistical Association*, Vol. 70, p. 311-319.
- Efron, B. et C. Morris. 1976 (a). « Families of Minimax Estimators of the Mean of a Multivariate Normal Distribution ». *Annals of Statistics*, Vol. 4, p. 11-21.
- Efron, B. et C. Morris. 1976 (b). « Multivariate Empirical Bayes and Estimation of Covariance Matrices ». *Annals of Statistics*, Vol. 4, p. 22-32.
- Efron, B. et C. Morris. 1977. « Stein's Paradox in Statistics ». *Scientific American*, Vol. 236, p. 119-127.
- Estrada R. et Ram P. K. 1994. « *Asymptotic Analysis : a Distributional Approach* ». Boston : Birkhäuser.
- Lehmann E. L. 1983. « *Theory of Point Estimation* ». New York ; Toronto : Wiley Series in Probability and Mathematical Statistic.
- Gentleman, J. F. et G.A. Whitmore. 1982. « Études de cas en analyse des données. ». *La Revue Canadienne de Statistique*, Vol. 10, No 4, p. 237-259.
- MacGibbon, B. et S. Groshen, et J.-G. Levreault, 1988. « Numerical Algorithms for Exact Calculations of Early Stopping Probabilities in One-Sample Clinical Trials With Censored Exponential Responses ». *Computing Science and Statistics, Proceedings of the 20th Symposium on the Interface, Interface '88*, p. 404-409.
- Maritz, J. S. 1966. « Smooth Empirical Bayes Estimation for One-Parameter Discrete Distributions ». *Biometrika*, Vol. 53, p. 417-429.
- Maritz, J. S. 1967. « Smooth Empirical Bayes Estimation for Continuous Distributions ». *Biometrika*, Vol. 54, p. 435-450.
- Maritz, J. S. 1968. « On the Smooth Empirical Bayes Approach to Testing of Hypotheses and the Compound Decision Problem ». *Biometrika*, Vol. 55, p. 83-100.
- Maritz, J. S. et T. Lwin. 1989. « *Empirical Bayes Methods* ». London : Chapman and Hall.
- Marsden, J. E. et A. Tromba 1981. « *Vector Calculus* ». New York : W. H. Freeman.
- Morris, C. N. 1983. « Parametric Empirical Bayes Inference : Theory and Applications ». *Journal of the American Statistical Association*, Vol. 78, N. 381, p. 47-65.
- Parzen, E. 1962. « On Estimation of a Probability Density Function and Mode ». *Annals of Mathematical statistics*, Vol. 33, p. 1065-1076.

- Robbins, H. 1951. « Asymptotically Subminimax Solutions of Compound Statistical Decision Problems. ». *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, p. 131-148.
- Robbins, H. 1955. « An empirical Bayes Approach to Statistics. ». *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, N. 1, p. 157-163.
- Robbins, H. 1964. « The Empirical Bayes Approach to Statistical Decision Problems. ». *The Annals of Mathematical Statistics*, Vol. 35, N. 1, p. 1-20.
- Robbins, H. 1980. « An Empirical Bayes Estimation Problem. ». *Proceedings of the National Academy of Sciences*, Vol. 77, N. 12, p. 6988-6989.
- Robert C. P. 1992. « *L'analyse Statistique Bayésienne* ». Paris : Économica.
- Robert C. P. et G. Casella. 1999. « *Monte Carlo Statistical Methods* ». New York : Springer.
- Robert C. P. 2001. « *The Bayesian Choice* ». New York : Springer.
- Rudin, W. 1975. « *Analyse réelle et complexe* ». Paris : Masson.
- Rudin, W. 1991. « *Functional Analysis* ». New York ; Montréal : McGraw-Hill.
- Sibony, M. 1988. « *Itérations et approximations* ». Paris : Hermann.
- Sibony, M. et J.-C. Mardon. 1982. « *Systèmes linéaires et non linéaires* ». Paris : Hermann.
- Stein, C. 1955. « Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. ». *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, p. 197-206.