

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE MED. KHIDER BISKRA
FACULTE DES SCIENCES ET SCIENCES DE L'INGENIEUR
DEPARTEMENT DE MATHEMATIQUES

MEMOIRE

Présenté Par :

BOUZIANE Nadjette

*Pour obtenir le Diplôme de Magister
En Mathématiques*

Efficacité-Robustesse des estimateurs des modèles paramétriques

Option

Analyse & Modèles Aléatoires

Devant le Jury :

Président :	B. MEZERDI	PR	Université de Biskra.
Rapporteur :	A. NECIR	MC	Université de Biskra.
Examineur :	S. BAHLALI	MACC	Université de Biskra.
Examineur :	A. BELAGOUNE	Dr MC	Université de Biskra.

Soutenu Le :.../.../2005.

Remerciements

Avec, mon grand remerciements à mon encadreur monsieur A. Necir maître de conférences à l'université de Biskra, de m'avoir suivi ce travail.

je tiens à exprimer ma profonde reconnaissance à monsieur B. Mezerdi, professeur à l'université de Biskra, d'avoir accepté de présider le jury de ce mémoire .

Je tiens à remercier honorablement monsieur S. Bahlali maître de conférences à l'université de Biskra et monsieur A. Belagoune maître de conférences, d'avoir d'être examinateurs et participés au jury .

Mes remerciements vont également à tous mes enseignants en graduation et en post graduation.

Honneur et gratitude à tous ceux qui de près ou de loin ont participe à la réalisation de mon mémoire.

Abstract

The present report is a synthesis on the research tasks concerning the effective or asymptotic effective estimate which is not always robust against outliers or robust but is not effective. Therefore, it is very interesting of research of the estimators at the same time effective and robust. For this goal Park and G.Lindsay (1999) proposed the method of the quadratic inference function. Serfling and Brazauskas introduced favorable estimators which are robust and effective.

Key words: Effectiveness, Method of maximum likelihood, Robustness, Quadratic function of inference, Method of median.

Résumé

Le présent mémoire est une synthèse sur les travaux de recherche concernant l'estimation efficace ou asymptotiquement efficace, mais pas toujours robuste ou bien robuste mais n'est pas efficaces. Donc, il est très intéressant de rechercher des estimateurs à la fois efficace et robuste. Pour ce but Park et G.Lindsay (1999) proposaient la méthode de la fonction d'inférence quadratique. D'autre part, Serfling et Brazauskas (2004) ont donnée des estimateurs favorables robustes et efficaces.

Mots clés : Efficacité, Robustesse, Fonction d'inférence quadratique, Méthode de médiane.

في هذه المذكرة ندرس المقدر الفعال و المقدر الفعال تقريبا و الذي لا يكون في غالب الأحيان صلبا ضد النقاط الشاذة، أو يكون صلب لكن ليس فعالا، إذن من المهم جدا البحث عن مقدر يكون صلب و فعال في نفس الوقت لأجل هذا الهدف قام كل من بارك و ليندساي (1999) بطرح طريقة الدالة الاستدلالية المربعة. كما قدم لنا كل من سيرفلينغ و برازوسكاس بعض المقدرات المختارة الصلبة و الفعالة.

الكلمات الجوهرية: الفعالية، الصلابة، الدالة الاستدلالية المربعة، الوسيط، طريقة الوسيط.

TABLE DES MATIÈRES

0.1. Notations	3
0.2. Introduction	4
CHAPITRE 1. Théorie de l'estimation ponctuelle, efficacité et efficacité relative des estimateurs	
1.1. Exhaustivité et information	6
1.1.1. Exhaustivité	6
1.1.2. Eléments du théorie de l'information	7
1.1.3. Relation entre l'exhaustivité et l'information	9
1.2. Estimation ponctuelle	10
1.2.1. Définitions et théorèmes sur l'estimation ponctuelle . .	10
1.2.2. Méthodes d'estimations	12
1.2.3. Efficacité et efficacité relative des estimateurs	17
CHAPITRE 2. Robustesse et estimation robuste	
2.1. Quelques définitions utiles	24
2.1.1. Distances usuelles	24
2.1.2. Différentiabilité de fonctionnelle	25
2.1.3. Dérivée au sens de Gâteaux	26
2.1.4. Différentiabilité au sens de Fréchet	26
2.2. Robustesse et estimation robuste	29
2.2.1. Définition de la robustesse	29
2.2.2. Définition qualitative de la robustesse	29
2.2.3. Estimateur robuste	31
2.2.4. Estimateur asymptotiquement efficace	37
CHAPITRE 3. Fonction d'inférence quadratique et estimations favorables	
3.1. Fonction d'inférence quadratique	44
3.1.1. Définition du fonction d'inférence quadratique	45
3.1.2. Efficacité asymptotique	49
3.1.3. Point de rupture de l'estimateur de la <i>FIQ</i>	53
3.2. Simulation	55
3.3. Estimations favorables pour le modèle de Pareto	60
3.3.1. Qualité des mesures d'ajustement	60
3.3.2. Critère de robustesse : Point de rupture	61
3.3.3. Critère d'efficacité : Variance	62

3.3.4. Estimateurs	62
CHAPITRE 4. Applications de la méthode de médiane pour les modèles de Weibull et de Paréto	65
4.1. Modèle de Weibull	65
4.1.1. Méthode de médiane	65
4.1.2. Comparaison par simulation	70
4.2. Modèle de Pareto	71
4.2.1. Méthode de médiane	71
4.2.2. Simulation	78

0.1 Notations

\xrightarrow{loi}	Convergence en loi
\xrightarrow{P}	Convergence en probabilité
$\xrightarrow{\mathcal{D}}$	Convergence en distribution
COV	Covariance
V	Variance
V_{as}	Variance asymptotique
E	Espérance
E_{θ}	Espérance de l'estimateur θ
f.r	Fonction de répartition
f	Fonction de densité
F	Fonction de répartition
CNS	Condition nécessaire et suffisante
i.i.d	Indépendante identiquement distribuées
v.a	Variable aléatoire
EMV	Estimateur du maximum de vraisemblance
Med	Médiane
EMMed	Estimateur de méthode de médiane
Σ	Matrice de covariance
$\langle \cdot, \cdot \rangle$	Produit scalaire
(\cdot, \cdot)	Intervalle ouvert
$\ \cdot\ $	Norme
O_p, o_p	Ordre d'une fonction
p.s	Presque sûrement
f.d	Fonction de distribution
$S(X, \theta), S(\theta, X), S(\theta)$	Fonction score
$d(x, y)$	Distance entre x et y

0.2 Introduction

L'un des problèmes de la statistique mathématique est la recherche des estimateurs efficaces ou asymptotiquement efficaces, c'est à dire des estimateurs à variance minimale. L'estimateur du maximum de vraisemblance possède cette propriété d'efficacité, mais malheureusement cet estimateur est très sensible contre les valeurs extrêmes, en d'autre terme un estimateur non robuste. Il est donc très intéressant de chercher des estimateurs qui sont à la fois efficaces et robustes. Dans la littérature il existe des méthodes, comme celle de la médiane, qui offrent des estimateurs robustes mais qui n'assurent pas l'efficacité. Une nouvelle méthode a été récemment développée assurant à la fois la robustesse et l'efficacité, c'est celle de la fonction d'inférence quadratique.

Pour aborder ce type de problème, nous donnons une définition introductive d'un estimateur, ces différents types et les méthodes les plus utilisées pour sa construction avec quelques théorèmes sur l'efficacité. En deuxième étape, nous exposons les différentes définitions de la robustesse, avec une analyse des estimateurs robustes, en prenant ici comme exemple le M-estimateur. Par la suite, nous présentons la méthode *d'inférence quadratique* offrant à la fois l'efficacité et la robustesse d'un estimateur. Cette méthode a été introduite récemment par Chanseok Park et Bruce G. Lindsay (1999).

Nous abordons dans le premier chapitre, deux outils de base de la statistique mathématique. Après avoir défini précisément ce que nous entendons par le mot "statistique", nous introduisons par le concept d'information. Il est naturel d'essayer de représenter l'information contenue dans un échantillon par une fonctionnelle $T(x_1, \dots, x_n)$ que l'on appellera une statistique. Cela nous conduira à donner un sens au mot "information". Comment savoir si la réduction des données opérés par la statistique T n'a pas fait perdre d'information. C'est ce type de problème que cherche à résoudre la notion d'exhaustivité.

A partir des valeurs observées x_1, \dots, x_n , nous calculons ensuite une certaine valeur numérique que nous les considérons comme une valeur approchée et que nous appellerons un estimateur. La règle qui permettra d'effectuer ce calcul est un estimateur auquel on demande de posséder certaines propriétés. Il est sans biais si son espérance est égale à la valeur de paramètre, il converge s'il converge en probabilité vers la valeur vraie, quand la taille de l'échantillon tend vers l'infini. Un estimateur optimal dans la classe des estimateurs sans biais est dit efficace. Les méthodes les plus classiques pour construire un estimateur à partir d'un échantillon sont : la méthode du maximum de vraisemblance, méthode de moment...etc.

Nous présentons dans le deuxième chapitre la définition de la robustesse selon F.Hampel (1968) et P.Huber (1964); suivie par quelques théorèmes et définitions

à savoir : la fonction d'influence, différentiabilité du fonctionnelle et l'estimation robuste...etc.

Les estimateurs classiques du paramètre θ sont peu sensibles aux déviations de la vraie distribution. Nous allons alors s'intéresser à la robustesse des méthodes d'optimisation, ou bien à celle des estimateurs ayant une forme spécifiée à priori. Pour cela, nous nous intéressons ici à la classe des M-estimateurs.

Le troisième chapitre contient deux parties :

Dans la première partie, nous exposons une nouvelle méthode, récemment développée par Chanseok Park et Bruce G. Lindsay (1999) assurant à fois la robustesse et l'efficacité d'un estimateur. Nous donnons aussi quelques définitions sur l'efficacité asymptotique ainsi que le principe du critère de la robustesse.

Nous avons réservé la deuxième partie à l'étude des estimateurs robustes et efficaces des paramètres des queues de distributions pour les modèles de Pareto et exponentielle tronquée et ce, dans le cas des grands et petits échantillons. Parmi les estimateurs robustes, nous citons entre autres : la médiane généralisée et la moyenne tronquée.

Dans le dernier chapitre nous donnons quelques applications de la méthode de la médiane pour deux types de modèles à savoir : le modèle de Weibull et le modèle de Pareto.

Chapitre 1

Théorie de l'estimation ponctuelle, efficacité et efficacité relative des estimateurs

1.1 Exhaustivité et information

1.1.1 Exhaustivité

Dans un problème statistique où figure un paramètre θ inconnu, un échantillon nous apporte une certaine information sur ce paramètre (information qui serait différente pour un autre paramètre avec le même échantillon). Lorsque l'on résume cet échantillon par une statistique, il s'agit de ne pas perdre cette information ; une statistique qui conserve l'information sera qualifiée d'exhaustive.

Soit le structure statistique¹ le couple $(\mathcal{L}, P_\theta, \theta \in \Theta)$, où \mathcal{L} l'espace des valeurs de X le v.a. du modèle, et P_θ la loi de probabilité de X . On écrit aussi (X, P_θ) . On notera $f(x, \theta)$ la densité de P_θ .

1.1.1.1-Statistique exhaustive :

Définition 1. La statistique T sera dite exhaustive si la loi conditionnelle de X sachant $T(x) = t$ et indépendant du paramètre θ . C'est-à-dire :

$$P(X/T(x) = t) \text{ ne dépend pas de } \theta.$$

1.1.1.2-Critère de factorisation : La démonstration de l'exhaustivité d'une statistique est rendue pénible par les longs calculs que demandent les probabilités conditionnelles. Il existe un théorème d'utilisation très facile, c'est le théorème de factorisation.

Théorème de factorisation (Fisher-Neyman) :

Soit le modèle (X, P_θ) , T une statistique exhaustive pour θ si et seulement si pour tout X , il existe deux fonctions mesurables positives h, g telle que g dépend de θ et t , h dépend de X , telle que :

$$f(x, \theta) = g(t, \theta)h(x).$$

Remarque : Ceci veut dire qu'une fois T connu, aucune valeur de l'échantillon ni aucune autre statistique ne nous apportera des renseignements supplémentaires sur θ .

¹modèle statistique

1.1.1.3-Famille exponentielle :

Définition 2. On appelle famille exponentielle de distribution $\{P_\theta, \theta \in \Theta\}$ si la densité de probabilité correspondante est de la forme suivante :

$$f(x, \theta) = \exp\{a(x)\alpha(\theta) + b(x) + \beta(\theta)\},$$

où α , a , b , et β des fonctions.

Théorème de Darmais : Soit X une v.a. dont le domaine de définition ne dépend pas de θ . Une CNS pour que l'échantillon (X_1, \dots, X_n) admette une statistique exhaustive T est que la forme de la densité soit :

$$f(x, \theta) = \exp\{a(x)\alpha(\theta) + b(x) + \beta(\theta)\}, \text{ (famille exponentielle).}$$

Si la densité est de cette forme et si de plus l'application $x_i \rightarrow \sum_{i=1}^n a(x_i)$ est bijective et continûment dérivable pour tout i , alors $T = \sum_{i=1}^n a(X_i)$ est une statistique exhaustive particulière.

1.1.1.4-Modèle complet, statistique complète :

Définition 3. Une statistique T est complète pour une famille de lois de probabilités $f(x, \theta)$ si :

$$\forall \theta \in \Theta \ E[h(T)] = 0 \implies h = 0 \text{ p.s.}$$

Proposition : La statistique naturelle d'une famille exponentielle (exhaustive) est complète, à condition que le domaine de $\alpha(\theta)$, pour $\theta \in \Theta$ soit d'intérieure non vide.

1.1.2 Éléments du théorie de l'information

Par analogie avec la signification du mot exhaustivité, dans la langue commun on peut donner la définition suivante : Si T est une statistique exhaustive pour θ , l'échantillon (X_1, \dots, X_n) , fournit sur θ pour une quantité d'information égale à celle fournit par T .

1.1.2.1-Information au sens de Fisher : Soit le modèle (\mathcal{L}, P_θ) , $\theta \in \Theta \subset \mathbb{R}$. On fait les hypothèses suivantes :

$$\mathbf{H}_1 : \forall x \ \forall \theta, f(x, \theta) > 0$$

$$\mathbf{H}_2 : \forall \theta \in \Theta, f(x, \theta) \text{ est dérivable par rapport à } \theta, P_\theta - p.s.$$

Définition 4. On appelle information de Fischer au point θ la quantité

$$I(\theta) = E \left[\frac{f'(x, \theta)}{f(x, \theta)} \right]^2.$$

Remarque : $f'(x, \theta)$ la dérivée par rapport à θ . L'espérance est prise par rapport à P_θ , à θ fixé, on peut écrire également :

$$I(\theta) = E \left[\frac{\partial}{\partial \theta} \log f(x, \theta) \right]^2.$$

Fonction score :

Définition 5. On appelle score la quantité :

$$S(x, \theta) = \frac{\partial}{\partial \theta} \log f(x, \theta).$$

Donc :

$$I(\theta) = E(S(x, \theta))^2.$$

Lemme : Soit S une fonction score alors :

$$\begin{aligned} E \left[\frac{f'(x, \theta)}{f(x, \theta)} \right] &= E[S] = 0, \\ E \left[\frac{f''(x, \theta)}{f(x, \theta)} \right] &= 0. \end{aligned}$$

Théorème 6. Si le domaine de définition de X ne dépend pas de θ alors :

$$I(\theta) = -E \left[\frac{\partial^2}{\partial \theta^2} \log f(x, \theta) \right] = -E \left(\frac{\partial}{\partial \theta} S(x, \theta) \right).$$

1.1.2.3-Propriétés de $I(\theta)$:

1- Positivité : $\forall \theta \in \Theta$, $I(\theta) = V \left(\frac{f'}{f} \right)$ alors $I(\theta) \geq 0$.

2- Additivité : Si on désigne par $I(\theta)$, l'information de Fisher dans une seule observation alors dans le cas de n observations où le domaine de définition ne dépend pas de θ on a :

$$I_n(\theta) = nI(\theta),$$

en effet les opérations espérance et la dérivée seconde sont linéaires.

Remarque : $I_n(\theta) = nI(\theta)$ veut dire que chaque observation a la même importance.

1.1.2.4-Propriétés de la fonction score :

1. Soit $S_n(\theta)$ le score sur θ apporté par un échantillon i.i.d. de X , les hypothèses H_1 et H_2 étant vérifiées. On a :

$$\frac{1}{n}S_n(\theta) = \frac{1}{n} \sum_{i=1}^n S(X_i, \theta).$$

D'après H_2 , l'application $u \xrightarrow{S} S(u, \theta) = \frac{\partial}{\partial \theta} \log f(u, \theta)$ est continue, (puisque $\frac{\partial^2}{\partial \theta^2} \log f(x, \theta)$ existe), et donc les $S(X_i, \theta)$, $i = 1$ à n , sont indépendants, de même loi P_θ^S image de P_θ par S .

2. D'après la Loi Forte des Grands Nombres :

$$\frac{1}{n}S_n(\theta) \xrightarrow{p.s} E(S) = 0, \text{ quand } n \rightarrow \infty,$$

alors :

$$\frac{1}{n}S_n(\theta) \xrightarrow{p.s} 0, \text{ quand } n \rightarrow \infty.$$

3. D'après le Théorème Central Limite :

$$\frac{\sqrt{n} \left(\frac{1}{n}S_n - E(S) \right)}{\sqrt{V(S)}} \xrightarrow{loi} N(0, 1), \text{ quand } n \rightarrow \infty,$$

or : $E(S) = 0$ et $V(S) = I(\theta)$, il en résulte :

$$\frac{1}{\sqrt{n}}S_n \xrightarrow{loi} N\left(0, \sqrt{I(\theta)}\right), \text{ quand } n \rightarrow \infty.$$

1.1.3 Relation entre l'exhaustivité et l'information

Le théorème suivant établissant la relation très étroite existant entre les notions de statistique exhaustive et d'information. Soit $I(\theta)$ l'information sur θ , T une statistique, P_θ^T l'image de P_θ par T , $I^T(\theta)$ l'information de la statistique T .

Théorème 7. *L'information ne peut que diminuer lorsque l'on passe dans le modèle image par une statistique :*

$$I^T(\theta) \leq I(\theta).$$

Remarque :

$$I^T(\theta) = I(\theta) \iff T \text{ exhaustive.}$$

1.2 Estimation ponctuelle

Le problème général de l'estimation est le suivant :

Soit X une variable dont la densité $f(x, \theta)$ dépend d'un paramètre inconnu θ ; comment trouver la "meilleure" valeur possible de θ à partir d'un échantillon observés x_1, \dots, x_n ?

Autrement dit, il s'agit de trouver une application (ou statistique) T fonction des variables aléatoires X_1, \dots, X_n à valeurs dans Θ espace des valeurs possibles de θ satisfaisant à des critères aussi naturels que possible. T sera appelé un estimateur de θ et la valeur prise par T pour n -uple de réalisations particulières x_1, \dots, x_n sera noté $\hat{\theta}$ et appelés estimation de θ .

1.2.1 Définitions et théorèmes sur l'estimation ponctuelle

1.2.1.1-Définition d'un estimateur :

Définition 8. On appelle estimateur de θ l'application :

$$T : \mathcal{L}^n \rightarrow \mathcal{D} \\ (X_1, \dots, X_n) \xrightarrow{T} T(X_1, \dots, X_n),$$

la valeur de T en l'échantillon observé (x_1, \dots, x_n) est l'estimation du paramètre.

1.2.1.2-Fonction de vraisemblance : Soit (X_1, \dots, X_n) i.i.d. de (X, P_θ) .

Définition 9. On appelle vraisemblance du paramètre θ , l'application $L : \Theta \rightarrow \mathbb{R}^+$ définie par :

$$\theta \in \Theta, L(x_1, \dots, x_n) = f(x_1, \dots, x_n, \theta).$$

Remarques :

1- L'échantillon étant i.i.d., on a :

$$L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta).$$

2- S désignant la fonction score :

$$\frac{\partial}{\partial \theta} \log L(x_1, \dots, x_n, \theta) = S_n(\theta) = \sum_{i=1}^n S(x_i, \theta).$$

1.2.2.4-Propriétés générales d'un estimateur : T désigne l'estimateur de θ , quand n la taille de l'échantillon intervient dans les propriétés (comportement asymptotique), nous le noterons T_n .

1- Estimateur sans biais :

Tout estimateur, on peut l'écrire comme suit :

$$E(T_n) = \theta + B(n, \theta),$$

où $B(n, \theta)$ est le biais de T_n .

Définition 10. On dit que l'estimateur T de θ est sans biais, s'il est centré en θ , c'est-à-dire :

$$E(T) = \theta.$$

Remarque : La propriété d'être sans biais exprime que dans une suite d'estimation donnée, la moyenne des estimations sera confondue avec la vraie valeur du paramètre, si cette propriété fait défaut ; on dit que l'estimation est à biais ou biaisée.

2- Estimateur asymptotiquement sans biais :

Définition 11. T_n est un estimateur asymptotiquement sans biais pour θ , c'est-à-dire :

$$\lim_{n \rightarrow +\infty} E(T_n) = \theta \iff \lim_{n \rightarrow +\infty} B(n, \theta) = 0.$$

3- Estimateur convergent :

T_n est un estimateur de θ convergent ; s'il converge en probabilité vers θ , c'est-à-dire :

$$\forall \alpha \in \mathbb{R}^+ \exists \varepsilon > 0, P[|T_n - \theta| \leq \alpha] \geq 1 - \varepsilon,$$

$$\text{où } \lim P[|T_n - \theta| \leq \alpha] = 1.$$

Théorème 12. Un estimateur T_n , dont l'espérance mathématique tend vers θ , et la variance tend vers zéro, est convergent pour θ c'est-à-dire :

$$\left\{ \begin{array}{l} E(T_n) \xrightarrow[n \rightarrow \div \infty]{} \theta \\ \text{et } V(T_n) \xrightarrow[n \rightarrow \div \infty]{} 0 \end{array} \right\} \implies T_n \xrightarrow[n \rightarrow \div \infty]{} \theta.$$

1.2.2 Méthodes d'estimations

Un paramètre possède une foule d'estimateurs pour cela, On arrêtons sur quelques méthodes générales de leurs construction, ces méthodes regroupent les approches les plus logiques du problème d'estimation et nous permettons d'acquérir les meilleurs estimateurs selon tel sur tel critère.

1.2.2.1-Méthode des moments

1. **Principe :** Cette méthode repose sur la propriété de convergence presque sûre des moments empiriques d'un échantillon i.i.d. (X_1, \dots, X_n) , extrait de X vers les moments théoriques correspondants de X .

Soit $\Theta \subset \mathbb{R}^p$, $\theta = (\theta_1, \dots, \theta_p)$, nous noterons par $m_k(\theta)$, le moment théorique d'ordre k de X ; qu'il soit centré ou non, et par $m_k(e, n)$ le moment empirique d'ordre k .

2. **Définition :** On appelle estimateur de θ obtenu par la méthode des moments (EMM), la solution θ_n^* du système :

$$\begin{cases} m_1(\theta) & = & m_1(e, n) \\ \vdots & & \vdots \\ m_p(\theta) & = & m_p(e, n). \end{cases}$$

Remarques :

- 1- Le choix des moments est guidé par la facilité de résolution du système.
- 2- On peut prendre des moments tous centrés en $E(X)$, ou tous non centrés, ou un mélange des moments centrés, ou non centrés.
- 3- Il n'y a aucune raison de retenir les p premiers moments sinon la simplicité de calcul .

3- Propriétés :

Nous nous contenterons de citer les deux propriétés essentielles d'EMM, θ_0 désigne la vraie valeur du paramètre.

$$\theta_n^* \xrightarrow[n \rightarrow +\infty]{p.s.} \theta_0.$$

$$\sqrt{n}(\theta_n^* - \theta_0) \xrightarrow{loi} N(0, \Sigma), \text{ quand } n \rightarrow +\infty.$$

La matrice Σ dépend des moments qui étaient choisis pour la détermination de θ_n^* .

On sait que les moments empiriques convergents en loi vers loi normale, θ_n^* qui est une fonction de ces moments empiriques, convergera donc vers une loi normale, par contre Σ n'est pas $I^{-1}(\theta)$.

4. Application :

Loi de gamma :

Soit $X \rightarrow \gamma(p; \theta)$, On appliquons la méthode des moments :

$$\begin{aligned} E(X) &= \frac{p}{\theta} = \overline{X}_n, \\ V(X) &= \frac{p}{\theta^2} = S_n^2, \end{aligned}$$

on en déduit :

$$\theta_n^* = \frac{\overline{X}_n}{S_n^2} \text{ et } p_n^* = \frac{\overline{X}_n^2}{S_n^2}.$$

On va calculer la loi de (θ_n^*, P_n^*) , on sait que :

$$\sqrt{n} \left(\begin{pmatrix} m_1(e, n) \\ m_2(e, n) \end{pmatrix} - \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} \right) \xrightarrow{\text{loi}} N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma \right), \quad (*)$$

avec :

$$\Sigma = \begin{pmatrix} m_2 - m_1^2 & m_3 - m_1 m_2 \\ m_3 - m_1 m_2 & m_4 - m_2^2 \end{pmatrix}.$$

Exprime en fonction de (p, θ) , on a :

$$\Sigma = \begin{pmatrix} \frac{p}{\theta^2} & \frac{2p(p+1)}{\theta^3} \\ \frac{2p(p+1)}{\theta^3} & \frac{2p(p+1)(2p+3)}{\theta^4} \end{pmatrix}.$$

Soit alors : $f : \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R} \times \mathbb{R}^+$ définie par :

$$f(u, v) = \begin{pmatrix} \frac{u}{v - u^2} \\ \frac{u^2}{v - u^2} \end{pmatrix},$$

le choix de f tel que :

$$\begin{cases} f(m_1(e, m), m_2(e, m)) = (\theta_n^*, p_n^*), \\ f(m_1, m_2) = (\theta, p). \end{cases}$$

La matrice des dérivées premières de f calculée en m_1 et m_2 est :

$$F = \begin{pmatrix} \frac{\theta^2 (2p+1)}{p} & -\frac{\theta^3}{p} \\ 2\theta(p+1) & -\theta^2 \end{pmatrix}.$$

En utilisant le résultat (*), on obtient :

$$\sqrt{n} \begin{pmatrix} \theta_n^* - \theta \\ p_n^* - p \end{pmatrix} \xrightarrow{\text{loi}} N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{\theta^2 (2p+3)}{p} & 2\theta(p+1) \\ 2\theta(p+1) & 2p(p+1) \end{pmatrix} \right).$$

1.2.2.2-Méthode du maximum de vraisemblance :

1. **Principe :** Cette méthode consiste, étant donné un échantillon X_1, \dots, X_n on pris comme estimation de θ , la valeur de θ qui rend maximale la vraisemblance $L(x_1, \dots, x_n, \theta)$. Intuitivement, puisque L représente une densité de probabilité, cela revient à supposer que l'évènement qui s'est produit était le plus "probable".

Les justifications mathématiques sont les suivantes : En effet, on prend comme estimation de θ la solution de l'équation de la vraisemblance :

$$\frac{\partial}{\partial \theta} \log L(X, \theta) = 0.$$

2. **Définition :** Soit $L(x, \theta)$ la vraisemblance au point θ , $\theta \in \Theta$, Θ quelconque. On appelle estimateur du maximum de vraisemblance (*EMV*) pour θ , la statistique $\hat{\theta}$

$$\begin{aligned} \hat{\theta} : \mathcal{L}^n &\rightarrow \Theta \\ (x_1, \dots, x_n) &\rightarrow \hat{\theta}(x_1, \dots, x_n), \end{aligned}$$

telle que :

$$\forall \theta \in \Theta, L(x, \hat{\theta}) \geq L(x, \theta).$$

3. Propriétés :

- 1- S'il existe une statistique exhaustive U alors $L(X, \theta) = g(u, \theta) h(x)$, la résolution de $\frac{\partial}{\partial \theta} \log L = 0$ revient à résoudre $\frac{\partial}{\partial \theta} \log g = 0$ donc :

$$\hat{\theta} = f(u).$$

2- Il n'y a aucune raison pour que l'estimateur du maximum de vraisemblance soit sans biais.

- 3- L'estimateur du maximum de vraisemblance n'a aucune raison d'être unique .

4. **Exemple** : Estimation de l'espérance d'une loi normale $N(m, \sigma)$, σ connu.

$$f(x, m) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right),$$

$$L(x_1, \dots, x_n, m) = \frac{1}{\sigma^n} (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{\sum (x_i - m)^2}{2\sigma^2}\right),$$

$$\log L = -n \log \sigma - \frac{\sum (x_i - m)^2}{2\sigma^2} + \text{constante},$$

$$\frac{\partial}{\partial m} \log L = \frac{\sum (x_i - m)}{\sigma^2} = 0 \implies \hat{m} = \bar{X}.$$

5. **Propriétés asymptotiques des EMVs** :

Soit (X_1, \dots, X_n) i.i.d. et $\hat{\theta}_n$ l'EMV de θ , l'indice n rappelant que cet estimateur dépend de la taille de l'échantillon.

Théorème 13. *Supposons les hypothèses suivantes :*

- H₁** : Θ ouvert de \mathbb{R} ,
- H₂** : $\theta \neq \theta' \implies P_\theta \neq P_{\theta'}$,
- H₃** : $\forall x, \forall \theta, f(x, \theta) > 0$,
- H₄** : $\forall x, \forall \theta, \frac{\partial f(x, \theta)}{\partial \theta}$ existe.

Alors, si θ_0 est la vraie valeur du paramètre, il existe une suite $\hat{\theta}_n$ des solutions de l'équation de vraisemblance qui converge p.s vers θ_0 :

$$\hat{\theta}_n \xrightarrow{p,s} \theta_0. \text{ quand } n \rightarrow +\infty$$

Pour la démonstration voir Philippe Tassi (1987), p. 187.

Théorème 14. *Sous les hypothèses H_1, H_2, H_3, H_4 , les ci-dessus, et :*

- H₅** : $\forall x, \forall \theta, f''(x, \theta)$ existe.
- H₆** : $\forall x, \forall \theta, \frac{\partial^2}{\partial \theta^2} \log f(x, \theta)$ est continue en θ , uniformément continue en x .
- H₇** : $\forall \theta, 0 < I(\theta) < +\infty$.
- H₈** : On peut dériver, $\forall \theta$, deux fois sous le signe \int ,

$$E\left(\frac{f'}{f}\right) = 0, \quad E\left(\frac{f''}{f}\right) = 0.$$

Alors, pour toute suite de les racines de l'équation de vraisemblance, on a :

$$\sqrt{n} (\hat{\theta}_n - \theta_0) \xrightarrow{\text{loi}} N \left(\theta_0, \frac{1}{\sqrt{I(\theta_0)}} \right), \quad n \rightarrow +\infty.$$

Remarque sur la résolution de l'équation de vraisemblance :

Sous les hypothèses de la régularité de la densité $f(x, \theta)$ de P_θ , la recherche des EMVs revenait à résoudre l'équation de vraisemblance :

$$S_n(\theta) = 0,$$

où $S_n(\theta) = \frac{\partial}{\partial \theta} L$. Malheureusement, il n'est pas toujours possible de fournir une formule explicite de l'EMV $\hat{\theta}$, pour cela, il y a plusieurs méthodes, comme la méthode de Newton-Raphson, ...etc. (voir, par exemple, Philippe Tassi, 1987)

3-Estimation par intervalle : Jusqu'ici nous cherchons à remplacer le paramètre inconnu θ par une valeur numérique unique à partir d'une réalisation d'un échantillon des variable aléatoire étudiée. Maintenant, au lieu de chercher une v.a T dont on espère qu'elle prendra une valeur assez proche de θ nous cherchons un intervalle aléatoire $[T_1, T_2]$ recouvrant la valeur θ avec une probabilité fixée à l'avance. Il s'agit donc de trouver deux variables aléatoires T_1 et T_2 telle que si α est la probabilité fixée (le plus souvent choisie égale à 95% ou 99%) on ait :

$$P(T_1 \leq \theta \leq T_2) = \alpha.$$

Définition 15. T_1 et T_2 étant deux statistiques construites à partir d'un n -échantillon de la v.a dont la loi de probabilité dépend d'un paramètre inconnu θ , on dira que l'intervalle aléatoire $[T_1, T_2]$ est un intervalle de confiance de coefficient α pour θ si les deux conditions suivantes sont réalisées :

$$T_1 \leq T_2 \text{ et } P(T_1 \leq \theta \leq T_2),$$

α s'appelle le niveau ou le coefficient de confiance de l'intervalle. La probabilité de commettre une erreur est :

$$1 - \alpha = P[(\theta < T_1) \text{ ou } (T_2 < \theta)] = P(\theta < T_1) + P(T_2 < \theta),$$

si les probabilités :

$$\begin{aligned} P_1 &= P(\theta < T_1), \\ P_2 &= P(T_2 < \theta), \end{aligned}$$

sont égales, et on dira que l'intervalle $[T_1, T_2]$ est symétrique en probabilité.

1.2.3 Efficacité et efficacité relative des estimateurs

1.2.3.1-Inégalité de Frechet-Darmonis-Cramer-Rao (FDCR) :

Hypothèses de Cramer-Rao :

\mathbf{H}_1 : Θ est un ouvert de \mathbb{R} , et $\forall x, \forall \theta, f(x, \theta) > 0$.

\mathbf{H}_2 : $\forall \theta, \frac{\partial f(x, \theta)}{\partial \theta}$ et $\frac{\partial^2 f(x, \theta)}{\partial \theta^2}$ existent et sont finies, *p.s.*

\mathbf{H}_3 : $\forall A \in \mathcal{A}, \int_A f(x, \theta) dx$ est dérivable au moins deux fois par rapport à θ , sous le signe d'intégration :

$$\frac{\partial}{\partial \theta} \int_A f(x, \theta) dx = \int_A \frac{\partial}{\partial \theta} f(x, \theta) dx.$$

(Ce qui impose en particulier que $\frac{\partial f(x, \theta)}{\partial \theta}$ et $\frac{\partial^2 f(x, \theta)}{\partial \theta^2}$ sont intégrables)

\mathbf{H}_4 : $\forall \theta, 0 < I(\theta) < +\infty$.

Théorème 16. *On suppose que les hypothèses de Cramer-Rao, sont vérifiées. Soit T un estimateur sans biais de $g(\theta)$, d'une variance finie, vérifiant :*

H_5 : $\forall A \in \mathcal{A}, \int_A T(x) f(x, \theta) dx$ est dérivable par rapport à θ , sous le symbole d'intégration et :

$$\forall \theta, \int \left| T(x) \frac{\partial}{\partial \theta} f(x, \theta) \right| dx < +\infty,$$

alors :

- (i) $g(\theta)$, est dérivable,
- (ii) $\forall \theta, V(T) \geq \frac{g'^2(\theta)}{I_n(\theta)}$.

Remarques :

1- Lorsque l'on veut estimer le paramètre θ , lui même ; $g(\theta) = \theta$, et

$$V(T) \geq \frac{1}{I_n(\theta)}.$$

2- Nous appellerons borne de Fréchet, notée B_F , la quantité :

$$B_F = \frac{g'^2(\theta)}{I_n(\theta)}.$$

3- Si T est un estimateur biaisé de $g(\theta)$, on a le risque $R(T, \theta)^2$

$$R(T, \theta) = E(T - g(\theta))^2 = V(T) + B^2(\theta).$$

T est sans biais pour $u(\theta) = g(\theta) + B(\theta)$ ³, et l'inégalité de FDCR, entraîne (si $B(\theta)$ est dérivable en θ) :

$$\forall T \in \Delta, \forall \theta E(T - g(\theta))^2 \geq B^2(\theta) + \frac{\left(\frac{du(\theta)}{d\theta}\right)^2}{I_n(\theta)}.$$

(Δ : l'ensemble des estimateurs possible.)

1.2.3.2-*Estimateur efficace et asymptotiquement efficace :*

Définition 17. *Un estimateur T sans biais pour $g(\theta)$, est dit efficace si sa variance est égale à la borne de Fréchet :*

$$V(T) = \frac{g'^2(\theta)}{I_n(\theta)}.$$

Remarque : Le rapport $\frac{B_F}{V(T)}$ mesure (en pourcentage) l'efficacité (absolue) de T .

Définition 18. *Un estimateur T_n sans biais pour $g(\theta)$, est dit asymptotiquement efficace si :*

$$\lim_{n \rightarrow +\infty} \frac{B_F}{V(T_n)} = 1.$$

Théorème 19. *La borne de Cramer-Rao ne peut être atteinte que si la loi de X est de la forme exponentielle :*

$$\log f(x, \theta) = a(x)\alpha(\theta) + \beta(x) + b(\theta),$$

car T est nécessairement exhaustif pour θ .

²erreur quadratique moyenne

³ $B(\theta) = B(\theta, n)$ est le biais de T

Sur la loi de X est bien de la forme précédent, il n'existe (à une transformation linéaire près) qu'une seule fonction $h(x)$ du paramètre qui puisse être estimée efficacement : c'est $h(\theta) = \frac{-b'(\theta)}{a'(\theta)}$.

L'estimateur de $h(\theta)$ est alors :

$$T = \frac{1}{n} \sum_{i=1}^n a(X_i).$$

La variance minimale est :

$$V(T) = -\frac{1}{na'(\theta)} \frac{d}{d\theta} \left(\frac{b'(\theta)}{a'(\theta)} \right) = \frac{h'(\theta)}{na'(\theta)}.$$

Pour la démonstration, voir G. Saporta (1990), p. 299.

Théorème de Koopman :

Soit $g(\theta)$ une fonction à estimer, P_θ vérifiant les hypothèses H_1 à H_4 de Cramer-Rao :

H_6 : $g(\theta)$ non constante,

H_7 : $f'(x, \theta)$ continue en θ .

Une CNS pour qu'il existe un estimateur sans biais T efficace de $g(\theta)$, est qu'il existe trois fonctions ; $\alpha(\theta)$, $\beta(\theta)$, et $\gamma(x)$, telles que :

$$\log f(x, \theta) = \alpha(\theta) T(x) + \beta(\theta) + \gamma(x).$$

En outre :

· $\alpha(\theta)$ et $\beta(\theta)$ sont dérivables,

· $\alpha'(\theta) \neq 0$,

· $g(\theta) = -\frac{\beta'(\theta)}{\alpha'(\theta)}$.

Application à la famille exponentielle :

Soit X dont la loi P_θ appartient à la famille exponentielle a un paramètre.

$$\log f(x, \theta) = \alpha(\theta) T(x) + \beta(\theta) + \gamma(x)$$

Alors en appliquant le théorème de Koopman, on établit immédiatement que la statistique $T(X)$ (ou $\sum_{i=1}^n T(X_i)$ pour un échantillon i.i.d. (X_1, \dots, X_n) de X) est un estimateur efficace du paramètre $g(\theta) = -\frac{\beta'(\theta)}{\alpha'(\theta)}$.

Statistique complète et estimation efficace :

Théorème 20. *Si une statistique exhaustive T est complète et $\theta^* \in \mathcal{B}_b^4$ l'estimateur $\theta_T^* = E_\theta(\theta^*/T)$ est le seul estimateur efficace de \mathcal{B}_b .*

Corollaire : Si T est une statistique exhaustive complète, et $\hat{\theta}$ un estimateur sans biais, alors $\hat{\theta}$ est un estimateur efficace qui de plus est unique.

1.2.3.3-Estimateur asymptotiquement normale :

Définition 21. *Soit (T_n) suite des estimateurs d'un paramètre θ , on dit que la suite $(T_n)_{n \geq 1}$ est convergente asymptotiquement normale (CAN) si :*

$$\forall x, \forall \theta, \lim_{n \rightarrow +\infty} P(\sqrt{n}(T_n - \theta) \leq x\sqrt{V(\theta)}) = \Phi(x),$$

où $\Phi(x)$ désigne la fonction de répartition de la loi $N(0, 1)$.

$$\sqrt{n}(T_n - \theta) \rightarrow N\left(0, \sqrt{V(\theta)}\right), \text{ quand } n \rightarrow +\infty.$$

$V(\theta)$ est la variance asymptotique de la suite (T_n) .

Définition 22. $(T_n)_{n \geq 1}$ une suite est dite normale asymptotiquement efficace (NAE) si :

- (i) T_n est une CAN,
- (ii) $\forall \theta, V(\theta) = \frac{1}{I(\theta)}$.

1.2.3.4-Estimateur optimal :

Définition 23. *On appelle estimateur optimal de $g(\theta)$, dans \mathcal{B}_0 , l'estimateur T préférable à tout autre au sens de la variance :*

$$\forall T_1 \in \mathcal{B}_0, V(T) \leq V(T_1).$$

Il n'y a aucune raison, pour que l'estimateur optimal soit efficace.

Théorème 24. *Soit S une statistique, $S : \mathcal{L}^n \rightarrow \mathbb{R}$, vérifiant $E(S) = 0$, CNS, pour qu'un estimateur T soit optimal est :*

$$COV(S, T) = 0.$$

⁴classe des estimateurs biais

Théorème de Rao-Blackwell : Soit T un estimateur sans biais de $g(\theta)$; on suppose que la loi P_θ admet une statistique exhaustive U . L'estimateur $h(U)$ défini par :

$$h(U) = E(T/U = u),$$

est sans biais pour $g(\theta)$, et préférable à T :

$$\forall \theta, V(T) \geq V(h(U)).$$

Remarque : Le théorème de Rao-Blackwell ne conduit pas à l'estimateur optimal; il permet seulement d'obtenir un estimateur préférable à l'estimateur initial. Le théorème de Lehmann-Scheffe prouve que, si la statistique U est complète, $h(U)$ est optimal pour $g(\theta)$, c'est-à-dire de variance minimale.

Théorème de Lehmann-Scheffe : Dans les conditions du théorème de Rao-Blackwell, si U est complète, l'estimateur amélioré $h(U)$ est optimal pour $g(\theta)$.

Démonstration : Voir Philippe Tassi (1987), p. 142.

1.2.3.5-Efficacité relative et asymptotiquement relative : Cette notion a été introduite par E. J. Pitman (1948), et permet de comparer deux procédures statistiques.

Soit deux procédures statistiques considérées comme équivalentes, selon un certain critère, pour des tailles d'échantillons n_1 et n_2 . Le rapport $\frac{n_1}{n_2}$ sera appelé efficacité relative de la procédure 2 vis-à-vis de la procédure 1. Si cette équivalence des performances n'est atteinte que pour n_1 et n_2 infinis et si $\frac{n_1}{n_2}$ admet une limite, cette dernière sera appelé efficacité relative asymptotique (ERA) de la procédure 2 vis-à-vis de la procédure 1.

Définition 25. On appelle efficacité relative de S vis-à-vis de T , pour une taille d'échantillon finie et une alternative θ^* fixé, le rapport :

$$ERA(S, T) = \frac{N'}{N},$$

tel que :

$$\begin{aligned} N &= \min \{n / P_{\theta^*} (S \in W_n) \geq r\}, \\ N' &= \min \{n / P_{\theta^*} (T \in W'_n) \geq r\}, \end{aligned}$$

* W'_n, W_n , les régions critiques de niveau $\alpha \in]0, 1[$, du test :

$$\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta - \Theta_0 \end{cases}$$

* $r \in]\alpha, 1[$, un nombre arbitraire fixé et θ^* alternative fixée de $\Theta - \Theta_0$.

Remarque : le rapport $\frac{N'}{N} \geq 1$, on dit qu'un test basé sur S est plus efficace que celui basé sur T .

Définition 26. On appelle *efficacité relative asymptotique (ERA)*, de la suite de tests (S_{n_j}) relativement à $(T_{n'_j})$ la limite :

$$ERA(S, T) = \lim_{j \rightarrow \infty} \frac{n'_j}{n_j},$$

* les hypothèses qui sont considérés :

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta_j \in \Theta_j, \quad j \in \mathbb{N}^*, \end{cases}$$

* (θ_j) suite alternative, telle que :

$$\theta_j \xrightarrow{j \rightarrow +\infty} \theta_0,$$

* pour un niveau $\alpha \in]0, 1[$, fixé. On noterons $S_{n_j}, T_{n'_j}$, deux statistiques permettant de construire ce test, et $\eta_{n_j}(\theta_j), \eta_{n'_j}(\theta_j)$, leurs puissances respectives.

* les suites $(n_j), (n'_j)$ sont choisies de telle sorte que ces deux tests aient même niveau limite α , et :

$$\alpha < \lim_{j \rightarrow +\infty} \eta_{n_j}(\theta_j) = \lim_{j \rightarrow +\infty} \eta_{n'_j}(\theta_j) < 1,$$

* le rapport $\frac{n'_j}{n_j}$ admette une limite.

Conditions de Pitman :

Soit T et S deux statistiques admettent une même loi limite, nous énoncerons un résultat permettant de calculer aisément cette quantité $ERA(S, T)$. Nous appellerons "Condition de Pitman" les conditions suivantes :

1. $\frac{S_{n_j} - \mu_j(\theta_j)}{\sigma_j(\theta_j)}, \frac{S_{n_j} - \mu_j(\theta_0)}{\sigma_j(\theta_0)}, \frac{T_{n'_j} - \mu'_j(\theta_j)}{\sigma'_j(\theta_j)}, \frac{T_{n'_j} - \mu'_j(\theta_0)}{\sigma'_j(\theta_0)}$, ont même loi limite quand $j \rightarrow +\infty$, où on a posé $\mu_j(\theta) = E_\theta(S_{n_j}), \mu'_j(\theta) = E_\theta(T_{n'_j}), \sigma_j^2(\theta) = V_\theta(S_{n_j}), \sigma_j'^2(\theta) = V_\theta(T_{n'_j}),$

2. $\lim_{n \rightarrow +\infty} \frac{\sigma_j(\theta_j)}{\sigma_j(\theta_0)} = \lim_{n \rightarrow +\infty} \frac{\sigma'_j(\theta_j)}{\sigma'_j(\theta_0)} = 1,$
3. $\Phi_n(\theta) = \frac{d\mu_n(\theta)}{d\theta}$ et $\Phi'_n(\theta) = \frac{d\mu'_n(\theta)}{d\theta}$ existent, continues dans un voisinage fermé de θ_0 avec $\Phi_n(\theta_0) \Phi'_n(\theta_0) \neq 0.$
4. $\lim_{n \rightarrow +\infty} \frac{\Phi_{n_j}(\theta_j)}{\Phi_{n_j}(\theta_0)} = \lim_{n \rightarrow +\infty} \frac{\Phi'_{n_j}(\theta_j)}{\Phi'_{n_j}(\theta_0)} = 1,$
5. $\lim_{n \rightarrow +\infty} \frac{\Phi_n(\theta_0)}{\sqrt{n \sigma_n^2(\theta_0)}} = K_S > 0$ et $\lim_{n \rightarrow +\infty} \frac{\Phi'_n(\theta_0)}{\sqrt{n \sigma_n'^2(\theta_0)}} = K_T > 0.$

Théorème de Noether (1955) :

Si les conditions de Pitman sont réalisées, alors :

$$ERA = \frac{K_S^2}{K_T^2}.$$

Chapitre 2

Robustesse et estimation robuste

2.1 Quelques définitions utiles

2.1.1 Distances usuelles

Il y a plusieurs distances usuelles, mais dans ce paragraphe on va définir quelques distances, telle que on va supposer que Ω est un espace polonais¹.

\mathcal{B} : désigne la tribu borélienne de Ω .

\mathcal{P} : ensemble convexe de toutes les probabilités sur (Ω, \mathcal{B}) , ou sur \mathcal{F} l'ensemble des f.r.

2.1.1.1-Distance de Prohorov : Introduit en 1956, elle est définie sur \mathcal{P} , A un sous ensemble non vide de Ω , et ε un réel strictement positif. On appelle ε -voisinage fermé de A ; le sous ensemble A^ε de Ω tel que :

$$A^\varepsilon = \{x \in \Omega / \inf_{y \in A} d(x, y) \leq \varepsilon\}.$$

Définition 27. Soit A un évènement quelconque de \mathcal{B} . On définit l'application, $p : \mathcal{B} \rightarrow \mathbb{R}^+$, par :

$$p(A) = \inf\{\varepsilon \geq 0, P(A) \leq Q(A^\varepsilon) + \varepsilon\},$$

alors :

$$d_p(P, Q) = \sup_{A \in \mathcal{B}} p(A).$$

2.1.1.2-Distance de Kolmogorov :

Définition 28. La distance de Kolmogorov ce n'est pas que la restriction de la distance en variation² à la classe \mathcal{F} des intervalles ouverts de la forme $] -\infty, x[$ alors la distance de Kolmogorov égale à :

$$d_K(F, G) = \sup_x (P(]-\infty, x]) - Q(]-\infty, x]) = \sup_x |F(x) - G(x)|,$$

où F et G sont des fonctions de répartition de P et Q respectivement.

¹espace polonais :est un espace topologique muni de la distance d complet et séparable (qui contient un sous ensemble dénombrable dense)

² $d_v(P, Q) = \sup_{A \in \mathcal{B}} |P(A) - Q(A)|$

2.1.1.3-Distance de Hellinger :

Définition 29. La distance de Hellinger $H(P, Q)$ entre deux probabilités P et Q de \mathcal{P} est définie par :

$$H^2(P, Q) = \int \left(\sqrt{dP} - \sqrt{dQ} \right)^2.$$

Corollaire : Si P et Q admettent des densités f et g (par rapport à une certaine mesure μ , σ -finie), on a :

$$H^2(P, Q) = \int \left(\sqrt{f} - \sqrt{g} \right)^2 d\mu = 2 \left(1 - \int \sqrt{fg} d\mu \right).$$

2.1.2 Différentiabilité de fonctionnelle

2.1.2.1-fonctionnelle statistique : En 1947, von Mises a introduit une classe de statistique, c'est les fonctionnelles statistique, alors on dit une fonctionnelle statistique tout statistique T_n pouvant s'écrire comme fonction de P_n , telle que :

$$T_n = T(P_n).$$

2.1.2.2-Contamination de deux probabilités :

Définition 30. On appelle contaminée de la probabilité P par la probabilité Q au taux t , $t \in [0, 1]$, la probabilité; notée $P_t(Q)$ définie par :

$$P_t(Q) = (1 - t)P + tQ.$$

Remarques :

- 1- $(P, Q) \in \mathcal{P}^2$, \mathcal{P} sera supposé muni d'une structure d'espace vectoriel et d'une distance δ .
- 2- Q est appelée la probabilité contaminante.
- 3- \mathcal{P} ensemble convexe et P_t est une combinaison linéaire de deux probabilités P et Q alors P_t est une probabilité.
- 4- La contamination est parfaite si $P_t(Q) = Q$.
- 5- On dit qu'il existe une absence de contamination si $P_t(Q) = P$.

2.1.3 Dérivée au sens de Gâteaux

La fonctionnelle T est dite dérivable au sens de Gâteaux si la quantité $T'_Q(P)$ définie par :

$$T'_Q(P) = \lim_{t \rightarrow 0} \frac{T(P + t(Q - P)) - T(P)}{t},$$

existe; $T'_Q(P)$ s'appelle la dérivée de Gâteaux de T en P dans la direction Q .

Remarques :

1- $T'_Q(P)$ n'est autre que la dérivée usuelle de :

$$\varphi(t) = T(P + t(Q - P)),$$

prise au point $t = 0$.

2- Il est possible de généraliser cette définition à l'ordre k :

$$T_Q^{(k)}(P) = \left[\frac{d^k}{dt^k} T(P + t(Q - P)) \right]_{t=0} = \varphi^{(k)}(0).$$

3- Par construction même, la dérivée de T en P dans la direction P est nulle.

2.1.4 Différentiabilité au sens de Fréchet

Définition 31. La fonctionnelle T est dite différentiable au sens de Fréchet en P s'il existe une fonctionnelle linéaire $L_P(Q - P)$, telle que :

$$T(Q) - T(P) - L_P(Q - P) = o(\delta(P, Q)).$$

Nous admettrons les résultats suivants :

1. L_P étant linéaire, il existe une fonction g intégrable vérifiant :

$$L_P(Q - P) = \int g d(Q - P).$$

2. Si T est différentiable en P au sens de Fréchet, elle est différentiable pour tout Q au sens de Gâteaux, (i.e. $T'_Q(P)$ existe pour tout Q de \mathcal{P}) et :

$$T'_Q(P) = L_P(Q - P).$$

Pour la démonstration voir R-von Mises (1947) ou R-Serfling (1980).

Théorème 32. Soit T fonctionnelle différentiable au sens de Fréchet en P ; P_n distribution empirique associée à un échantillon (X_1, \dots, X_n) . On suppose qu'elle vérifie :

$$\sqrt{n}\delta(P_n, P) \xrightarrow{P} 0, \text{ quand } n \rightarrow +\infty.$$

Alors : si $V_P(T_x(P)) > 0$:

$$\sqrt{n}(T(P_n) - T(P)) \xrightarrow{loi} N\left(0, \sqrt{V_P(T_x(P))}\right); \text{ quand } n \rightarrow +\infty.$$

2.1.4.1-Fonction d'influence : Soit T fonctionnelle quelconque, supposons qu'il existe une fonction h telle que :

$$T(P_n) - T(P) = \frac{1}{n} \sum_{i=1}^n h(P, X_i) + R_n,$$

$h(P, X_n)$ et R vérifient les hypothèses suivantes :

$$\begin{cases} V_P(h(P, X_i)) > 0, \\ \text{et } \sqrt{n}R_n \xrightarrow{P} 0. \end{cases}$$

Définition de la fonction d'influence :

Supposons que l'on possède une approximation de la différence $T(P_n) - T(P)$ de la forme :

$$T(P_n) - T(P) = \frac{1}{n} \sum_{j=1}^n h(P, X_j) + R_n;$$

avec : $R_n = O_P\left(\frac{1}{\sqrt{n}}\right)$.

L'erreur commise en estimant $T(P)$ par $T(P_n)$ est approximativement,

$$\frac{1}{n} \sum_{j=1}^n h(P, X_j),$$

ainsi $h(P, X_j)$ peut être considérée comme la contribution de X_j à l'erreur approchée $T(P_n) - T(P)$ ou l'influence de X_j sur cette erreur.

Or on vient d'établir que si la fonctionnelle T est différentiable au sens de Fréchet, on a :

$$T(P_n) - T(P) = \frac{1}{n} \sum_{j=1}^n T_{x_j}^{\wedge}(P) + R_n.$$

Remarque : R_n représente le reste du développement de Taylor de $T(P_n)$ en P .

Définition 33. On appelle fonction d'influence de la fonctionnelle T en P la dérivée au sens de Gâteaux de T en P dans la direction δ_x , soit en notant :

$$P_t(\delta_x) = (1-t)P + t\delta_x,$$

$$IF(x; P, T) = T_x^\wedge(P) = \lim_{t \rightarrow 0} \frac{T(P_t(\delta_x)) - T(P)}{t}.$$

Remarques :

- 1 $E_P(IF) = \int IF(x; P, T) dP = 0$.
- 2 L'influence "moyenne" est nulle pour tout point supplémentaire x , réalisation de mécanisme aléatoire P .
- 3 Si $V_P(IF) = E_P(IF^2)$ est strictement positive :

$$\sqrt{n}(T(P_n) - T(P)) \xrightarrow{\text{loi}} N\left(0, \sqrt{V_P(IF)}\right), \text{ quand } n \rightarrow +\infty.$$

La variance asymptotique $\sqrt{n}(T(P_n) - T(P))$ n'est autre que la variance de la fonction d'influence; de façon heuristique, on pourra approximer $V(T(P_n))$ par $\frac{1}{n}V_P(IF)$.

Application : Soit $T_n = T(P_n) = T(X_1, \dots, X_n)$. L'impact d'une observation supplémentaire Y sur T est défini par :

$$I(T_n, Y) = T_{n+1}(X_1, \dots, X_n, Y) - T(X_1, \dots, X_n).$$

En notant

$$P_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i} \text{ et } P_{n+1} = \frac{1}{n+1} \left(\delta_y + \sum_{i=1}^n \delta_{x_i} \right),$$

on a :

$$P_{n+1} = \frac{n}{n+1}P_n + \frac{1}{n+1}\delta_y = \left(1 - \frac{1}{n+1}\right)P_n + \frac{1}{n+1}\delta_y,$$

P_{n+1} est donc la contaminée de P_n par δ_y , au taux $\frac{1}{n+1}$.

Ainsi :

$$I(T_n, Y) = T(P_{n+1}) - T(P_n) = T\left(\left(1 - \frac{1}{n+1}\right)P_n + \frac{1}{n+1}\delta_y\right) - T(P_n).$$

L'impact $I(T_n, Y)$ est donc égale, au coefficient $(n + 1)$ près à la version discrétisée de la fonction d'influence.

Cette dernière est :

$$\frac{T\left(\left(1 - \frac{1}{n+1}\right)P_n + \frac{1}{n+1}\delta_y\right) - T(P_n)}{1/(n+1)},$$

dont la limite en probabilité, quand $n \rightarrow +\infty$ est $IF(y, P, T)$.

2.1.4.2-Sensibilité de T_n :

Définition 34. On appelle sensibilité de T_n , l'échantillon (X_1, \dots, X_n) étant donné, la quantité :

$$s(T_n) = \sup_Y |I(T_n, Y)|.$$

2.2 Robustesse et estimation robuste

2.2.1 Définition de la robustesse

Il existe plusieurs approches diverses de la robustesse mais nous retiendrons la définition admise depuis les travaux de P.Huber (1964) et F.Hampel (1968), qui repose sur la continuité de la fonctionnelle T au voisinage de la loi P .

Définition 35. Une procédure statistique sera robuste si ses performances sont peu modifiées par des faibles modifications de la loi P , modèle des observations. Cette approche est celle de F.Hampel et P.Huber.

Définition 36. P.Huber, en 1972, écrit : "La robustesse est une sorte d'assurance : Je suis prêt à payer une perte d'efficacité de 5 à 10% par rapport au modèle idéal pour me protéger de mauvais effets de petites déviations de celui-ci : je serai bien sûr heureux que ma procédure statistique fonctionne bien sous de gros écarts, mais je n'y prête pas réellement attention car faire de l'inférence à partir d'un modèle aussi faux n'a que peu de signification concrète".

2.2.2 Définition qualitative de la robustesse

Soit X une v.a. à valeurs sur $(\mathcal{L}, \mathcal{B}, P)$, (X_1, \dots, X_n) les v.a. associées au modèle d'échantillonnage $(\mathcal{L}, \mathcal{B}, P)^{\otimes n}$, T_n une statistique définie sur $(\mathcal{L}, \mathcal{B}, P)^{\otimes n}$ et à valeurs dans $(\mathbb{R}^k, \mathcal{R}^k)$, \mathcal{P} étant muni de la distance de Prohorov d_p .

Définition 37. La suite $T_n = T(X_1, \dots, X_n)$ est dite robuste en P si :

$$(\forall \varepsilon > 0); (\exists \alpha > 0), (\forall Q \in \mathcal{P}), (\forall n \in \mathbb{N}^*) \\ d_p(P, Q) \leq \alpha \implies d_p(P^{T_n}, Q^{T_n}) \leq \varepsilon.$$

Remarque : Cette définition exprime que si deux lois sont proches, les lois images par la statistique robuste T_n sont aussi proches.

2.2.2.1-Indicateurs de mesure de la robustesse : La robustesse qualitative d'une statistique T_n , associée à la notion de continuité, est la qualité minimale requise. Dans le cas où la fonction d'influence existe, c'est-à-dire lorsque la fonctionnelle T est non seulement continue mais dérivable, on peut quantifier cette notion de robustesse au moyen de certains indicateurs. Une statistique robuste correspondra en général au fait que la notion d'influence est bornée et continue, c'est-à-dire que l'influence de valeurs extrêmes est limitée et varie continûment.

2.2.2.2-Utilisation de la fonction d'influence : Soit $IF(x, P, T)$ la fonction d'influence. On appelle sensibilité globale en P de la fonctionnelle T la quantité :

$$\gamma^* = \sup_x |IF(x; P, T)|.$$

On appelle sensibilité locale de T la quantité :

$$\lambda^* = \sup_{\substack{x, y \\ x \neq y}} \left| \frac{IF(y; P, T) - IF(x; P, T)}{y - x} \right|.$$

Remarque : Ce deuxième indicateur révèle l'effet d'un déplacement local d'une observation de x en y , ce qui est le cas lorsqu'on tronque ou arrondit les données.

2.2.2.3-Robustesse au sens de la variance asymptotique : Soit T_n une suite converge en loi, avec $V_{as}(P, T)$ la variance asymptotique. On posera :

$$\sigma^2(T) = \sup_{\mathcal{P}_0} V_{as}(P, T),$$

où \mathcal{P}_0 est un ensemble de lois de probabilité; usuellement, \mathcal{P}_0 est un voisinage de la loi vraie du modèle.

Définition 38. Tout suite T_n soit robuste au sens de la variance asymptotique minimale pour toute suite d'estimateurs T_n^* de θ si :

$$\sigma^2(T_n) \leq \sigma^2(T_n^*).$$

2.2.2.4-Point de rupture : Revenons à la définition qualitative de la robustesse. De façon heuristique, on pourrait avoir α petit tant que ε reste inférieur à un seuil ε^* ; ce seuil s'appelle le point de rupture et est défini par :

$$\varepsilon^* = \sup \{ \varepsilon / d_p(P, Q) < \varepsilon \Rightarrow d_p(P^{T_n}, Q^{T_n}) < \alpha \}.$$

Remarque : Le point de rupture définit le plus grand voisinage possible de P dans lequel toute déviation n'entraîne pas d'effets "désastreux" (définis par ε) pour la statistique T_n .

2.2.3 Estimateur robuste

2.2.3.1-M-estimateurs : La classe des M-estimateurs a été introduite par P.Huber (1964) et ses principales propriétés sont exposées dans P.Huber (1981).

Définition 39. Un M-estimateur T_n est la solution du problème :

$$\min_{\theta} \sum_{i=1}^n \rho(X_i, \theta). \quad (1)$$

Si ρ est une fonction dérivable par rapport à θ , T_n est solution de l'équation du premier ordre :

$$\sum_{i=1}^n \Psi(X_i, \theta) = 0. \quad (2)$$

Où $\Psi(x, \theta) = \frac{\partial}{\partial \theta} \rho(x, \theta)$; ainsi, en divisant (2) par $\frac{1}{n}$, l'estimateur T_n vérifie :

$$\frac{1}{n} \sum_{i=1}^n \Psi(X_i, T_n) = 0.$$

Ce qui n'est autre que :

$$E_{p_n} [\Psi(X, T_n)] = 0.$$

On associe à la fonctionnelle statistique T_n , la fonctionnelle T M-fonctionnelle définie sur \mathcal{P} comme solution de l'équation :

$$\int \Psi(X, T(P)) dP = E_P [\Psi(X, T(P))] = 0.$$

2.2.3.2-Calcul de la fonction d'influence : Soit $P_t = (1 - t)P + t\delta_x$, $T(P_t)$ est solution de :

$$\int \Psi(X, T(P_t)) d((1 - t)P + t\delta_x) = 0.$$

On dérivons par rapport à t , avant de dériver, on trouve par linéarité que la formule précédente égale à :

$$(1 - t) \int \Psi(x, T(P_t)) dP + t\Psi(x, T(P_t)) = 0,$$

car :

$$t \int \Psi(x, T(P_t)) d\delta_x = t\Psi(x, T(P_t)),$$

alors :

$$\begin{aligned} - \int \Psi(x, T(P_t)) dP + \Psi(x, T(P_t)) + (1 - t) \frac{dT(P_t)}{dt} \int \left[\frac{\partial}{\partial \theta} \Psi(x, \theta) \right]_{\theta=T(P_t)} dP \\ + t \frac{dT(P_t)}{dt} \left[\frac{\partial}{\partial \theta} \Psi(x, \theta) \right]_{\theta=T(P_t)} = 0. \end{aligned}$$

Puis, en faisant $t = 0$, nous obtenons :

$$IF(x, P, T) = \frac{\Psi(x, T(P))}{- \int \left[\frac{\partial}{\partial \theta} \Psi(x, \theta) \right]_{\theta=T(P)} dP}. \quad (3)$$

Remarque : Si θ paramètre de translation, $\Psi(x, \theta) = \Psi(x - \theta)$, on a :

$$IF(x, P, T) = \frac{\Psi(x - T(P))}{\int \Psi'(x - T(P)) dP}.$$

2.2.3.3-Point de rupture : Si Ψ n'est pas bornée, le M-estimateur associé n'est pas robuste et $\varepsilon^* = 0$, si Ψ est bornée et impaire, ε^* prend la valeur maximale $\frac{1}{2}$.

Exemple : La médiane

Notons par $Med(P) = T(P) = \theta$, la médiane de la loi P . On fait le choix de la fonction critère $\rho(t) = |t|$.

La médiane empirique $T_n = Med_n$, est la solution de :

$$\min \sum_{i=1}^n |X_i - \theta|.$$

La fonction ρ n'est pas dérivable en $t = 0$. On peut cependant définir $\rho' = \Psi$ comme la fonction *sign* définie par :

$$\rho'(t) = \Psi(t) = \text{sign}(t) = \begin{cases} 1 & \text{si } t > 0 \\ 0 & \text{si } t = 0 \\ -1 & \text{si } t < 0. \end{cases}$$

Nous allons calculer la fonction d'influence, en admettant l'interversion des opérations de dérivation et d'intégration.

Alors :

$$\int \Psi(x, \theta) dP = \int \text{sign}(x - \theta) dP = - \int_{-\infty}^{\theta} dP + \int_{\theta}^{+\infty} dP = 1 - 2F(\theta),$$

si F admet une densité f , on obtient alors, en dérivant l'équation précédente :

$$\frac{\partial}{\partial \theta} \int \Psi(x, \theta) dP = -2f(\theta),$$

d'où :

$$IF(x, P, Med) = \frac{\text{sign}(x - Med)}{2f(Med)}. \quad (4)$$

Nous voyons bien que Med_n , la médiane empirique, est un estimateur robuste car $IF(x, P, Med)$ est bornée, ce qui n'était pas le cas de la moyenne empirique. Le point de rupture est $\varepsilon^* = \frac{1}{2}$ et la sensibilité globale $\gamma^* = \frac{1}{2}f(Med)$, $\gamma^* = 1,253$ pour loi $N(0, 1)$. Remarquons que $IF^2(x, P, Med) = \frac{1}{4f^2(Med)}$ est une constante, le résultat de convergence est :

$$\sqrt{n}(Med_n - Med) \xrightarrow{loi} N\left(0, \frac{1}{2f(Med)}\right), \text{ quand } n \rightarrow +\infty.$$

Si par exemple : $P = N(m, \sigma)$, alors $Med = m$ et :

$$\sqrt{n}(Med_n - m) \xrightarrow{loi} N\left(0, \sigma\sqrt{\frac{\pi}{2}}\right), \text{ quand } n \rightarrow +\infty.$$

Remarque : Dans le cas du médiane, pour une loi quelconque P , le calcul direct de la fonction d'influence peut être mené à partir de la définition.

Posons $P_t = (1 - t)P + t\delta_x$; en notant F_t la f.r. de P_t , la médiane $T(P_t)$ est solution de l'équation $F_t(u) = \frac{1}{2}$.

Soit :

$$\frac{1}{2} = (1 - t)F(u) + t1_{]-\infty, x[}(u).$$

On obtient :

$$T(P_t) = \begin{cases} F^{-1}\left[\frac{1}{2(1-t)}\right] & \text{si } x \geq T(P_t) \\ F^{-1}\left[\frac{1-2t}{2(1-t)}\right] & \text{si } x < T(P_t). \end{cases}$$

On voit que :

$$T(P_t) \xrightarrow[t \rightarrow 0]{} F^{-1}\left(\frac{1}{2}\right) = Med,$$

alors $T(P_t)$ est une fonction de t continue en $t = 0$.

Nous pouvons écrire :

$$\begin{aligned} \frac{1}{2(t-1)} &= \frac{1}{2} + \frac{t}{2} + O(t^2) \\ \text{et } \frac{1-2t}{2(1-t)} &= \frac{1}{2} - \frac{t}{2} + O(t^2). \end{aligned}$$

Ainsi :

$$\frac{1}{t} [T(P_t) - T(P)] \xrightarrow[t \rightarrow 0]{} \frac{1}{2} \text{sign}[x - Med] \left[\frac{dF^{-1}(x)}{dx} \right]_{x=Med}.$$

On retrouve :

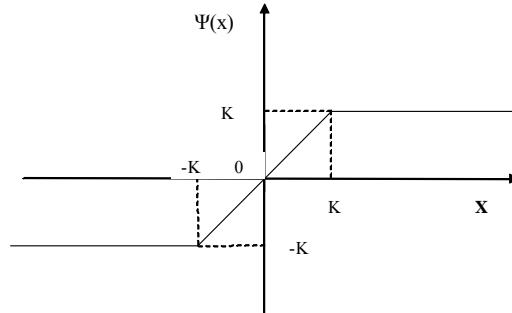
$$IF(x, P, Med) = \frac{\text{sign}(x - Med)}{2f(Med)}.$$

2.2.3.4-Choix de Ψ : On supposons ici, que la loi P symétrique, telle que, dans le cas où θ est un paramètre de translation, on prend comme fonction Ψ , une fonction impaire et croissante au sens large.

En 1964, P.Huber a suggéré de prendre comme fonction Ψ :

$$\Psi_H(x) = \begin{cases} x & \text{si } |x| \leq k \\ k \text{sign}(x) & \text{si } |x| > k. \end{cases}$$

Où k est nombre réel positif fixé.



On obtient ainsi un estimateur robuste, a un point de rupture $\varepsilon^* = \frac{1}{2}$ et une sensibilité globale :

$$\gamma^* = \frac{k}{[2F(k) - 1]},$$

où F est la f.r. d'une loi P .

Un tel choix permet d'atténuer l'effet de les valeurs extrêmes aberrantes, ou provenant d'une distribution à queue importante, pour l'estimation de paramètre θ

Un choix plus radical peut être :

$$\Psi_H^*(x) = \begin{cases} 0 & \text{si } |x| > k \\ x & \text{si } |x| \leq k. \end{cases}$$

Ainsi ; si n_1 (resp. n_2) est le nombre des valeurs inférieures à $-k$ (resp. supérieures à k), en utilisant (2) pour Ψ_H :

$$\hat{\theta}_H = \frac{1}{n} \sum_{i=1}^n X_i 1_{[-k, k]}(X_i) + \frac{(n_2 - n_1)k}{n},$$

et pour Ψ_H^* :

$$\hat{\theta}_H^* = \frac{1}{n - n_1 - n_2} \sum_{i=1}^n X_i 1_{[-k, k]}(X_i).$$

De nombreuses autres fonction Ψ ont été proposées. Citons à titre d'exemple, celle de Tukey :

$$\Psi_T(x) \begin{cases} x(1 - (x/k)^2)^2 & \text{si } |x| \leq k \\ 0 & \text{si } |x| > k. \end{cases}$$

Ou celle de Andrews :

$$\Psi_A(x) \begin{cases} \sin\left(\frac{x}{k}\right) & \text{si } |x| \leq k\pi \\ 0 & \text{si } |x| > k\pi. \end{cases}$$

J.Barthoulot (1983) considère deux formulations générales pour Ψ .

a) La classe de fonctions Ψ impaires :

$$\Psi(u) \begin{cases} \text{sign}(u) \gamma(\alpha + \beta) k & \text{si } |u| \geq k \\ \alpha u + \beta k \text{sign}(u) & \text{si } |u| \leq k. \end{cases}$$

b) La classe de fonctions Ψ asymétriques :

$$\Psi(u) = \begin{cases} u & \text{si } u \leq 0 \\ \alpha u + \beta k & \text{si } 0 < u \leq k \\ \gamma k(\alpha + \beta) & \text{si } u > k. \end{cases}$$

Remarque : Pour $\alpha = 1$, $\beta = 0$, $\gamma = 1$, on trouve la fonction Ψ_H de Huber pour la classe de fonction Ψ impaire.

2.2.3.5-Choix d'un estimateur robuste de dispersion : Il y a deux principes d'estimation possible pour une loi de probabilités dépend d'un paramètre de position θ et d'un paramètre d'échelle ou de dispersion, noté σ :

a) Soit estimer simultanément θ et σ , c'est-à-dire résoudre en T_n et S_n :

$$\begin{aligned} \sum_{i=1}^n \Psi\left(\frac{X_i - T_n}{S_n}\right) &= 0, \\ \sum_{i=1}^n \mathcal{X}\left(\frac{X_i - T_n}{S_n}\right) &= 0. \end{aligned}$$

b) Soit estimer θ est en résolvant en T_n l'équation :

$$\sum_{i=1}^n \Psi\left(\frac{X_i - T_n}{S_n}\right) = 0,$$

ayant déterminé l'estimation S_n de σ par ailleurs

Dans ce dernier cas S_n doit posséder des propriétés de robustesse, soit par exemple, S_n est l'écart-type empirique :

$$S_n = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n - 1}}.$$

Si S_n est grand ³ $(X_i - T_n)/S_n$ peut être réduit pour ces observation.

Par exemple, si on retient la fonction Ψ_H de Huber, il se peut des variables X_i dont la contribution devrait être ramenée à $\pm k$ soient prises en compte dans le calcul du M-estimateur.

J. Barthoulot (1983) à étudié les propriétés de trois estimateurs usuels de la dispersion :

$$\left\{ \begin{array}{l} S_n \text{ écart-type empirique,} \\ D_n \text{ écart absolu moyen, } D_n = \frac{1}{n} \sum_{i=1}^n |X_i - \bar{X}|, \\ E_n \text{ distance semi-interquartile, } E_n = \frac{1}{2} [X_{0.75} - X_{0.25}] \end{array} \right.$$

Les propriétés en contamination sont résumées dans le tableau ci-après, où la première ligne est le biais asymptotique, la seconde est l'écart quadratique moyen asymptotique.

loi	S_n	D_n	E_n
$N(0, 1)$	0 0.5	0 0.571	0.0007 1.369
$0.95N(0, 1) + 0.05N(0, 4)$	0.323 5.560	0.038 1.672	0.075 1.580
$0.95N(0, 1) + 0.05N(2, 1)$	0.091 0.717	-0.040 0.949	0.038 1.550
$0.90N(0, 1) + 0.1N(4, 1)$	0.562 2.622	-0.129 3.091	0.149 2.318

2.2.4 Estimateur asymptotiquement efficace

La construction d'un estimateur robuste du paramètre θ dépend de la famille \mathcal{F} l'ensemble de les f.r. F . On ne peut pas espérer que l'estimateur robuste pour tout élément F de \mathcal{F} ait des propriétés d'optimalité globale; le coût de la robustesse est une efficacité parfois faible.

Par contre, si on se place au point de vue local, \mathcal{F} étant un voisinage d'une distribution donnée, on obtiendra fréquemment un estimateur du minimax pour un voisinage \mathcal{F} bien choisi. Afin d'introduire simplement quelques résultats, sur ce point on se propose d'étudier le cas d'un paramètre de translation.

Théorème 40. *Une CNS pour que $T(P_n)$ soit un estimateur asymptotiquement efficace de $\theta = T(P)$ est que :*

$$IF(x, P_\theta, T) = \frac{1}{I(P_\theta)} \frac{\partial \log f_\theta}{\partial \theta}.$$

³c'est-à-dire s'il existe un certain nombre d'observation X très dispersées,

2.2.4.1-Cas particulier du paramètre de translation : Pour $F_\theta(x) = F(x - \theta)$, où F est une f.r. donnée, on a $f_\theta(x) = f_0(x - \theta)$.

Pour le M-estimateur, on peut obtenir un estimateur asymptotiquement efficace du paramètre $\theta = T(P_\theta)$ au moyen des fonctionnelles T_0 définies ci-après :

M-estimateur :

T_0 est obtenue en prenant :

$$\Psi(x) = -c \frac{f'_0(x)}{f_0(x)}, \quad c \neq 0.$$

En effet

$$\begin{aligned} E_\theta(\Psi') &= \int \Psi'(x - \theta) dP_\theta = - \int \Psi(x - \theta) f'_0(x - \theta) dx \\ &= c \int \left(\frac{f'_0}{f_0} \right)^2 f_0 dx = \frac{1}{c} E_\theta(\Psi^2), \end{aligned}$$

et

$$V_{as}(P_\theta, T) = \frac{E_\theta(\Psi^2)}{[E_\theta(\Psi'^2)]^2} = \frac{c^2}{E_\theta(\Psi^2)} = \frac{1}{\int \left(\frac{f'_\theta}{f_\theta} \right)^2 f_0 dx} = \frac{1}{I(P_\theta)}.$$

2.2.4.2-Estimateur minimax-robuste : La plupart des estimateurs $T_n = T(P_n)$ sont asymptotiquement normaux. On peut donc utiliser leur variance asymptotique comme mesure de leur performance :

$$V_{as}(P, T) = \int IF^2(x, P, T) dP.$$

Remarque : La variance asymptotique peut être maximale lorsque P appartient à un certain ensemble \mathcal{P}_0 de lois de probabilité :

$$\sigma^2(T) = \sup_{P \in \mathcal{P}_0} V_{as}(P, T).$$

Définition 41. On dit qu'un estimateur associé à une fonctionnelle T_0 est minimax-robuste ⁴ si T_0 réalise :

$$\sigma^2(T_0) = \inf_{T \in \mathcal{F}} \sup_{P \in \mathcal{P}} V_{as}(P, T),$$

tel que :

$$\forall T \in \mathcal{F}, \quad \sigma^2(T_0) \leq \sigma^2(T).$$

⁴robuste au sens de la variance asymptotique minimale

2.2.4.3-Cas particulier du paramètre de translation : Soit (X_1, \dots, X_n) un échantillon d'une loi $F(x - \theta)$ où θ est un paramètre inconnu à estimer et F un élément inconnu d'un ensemble de f.r.

Soit $F_0 \in \mathcal{F}$ qui minimise l'information de Fisher $I(F)$ pour un paramètre de translation. On considérera dans tout ce qui suit les deux spécifications de \mathcal{F} ci-dessous :

Modèle de contamination de G :

$$\mathcal{F}'_G = \{F/F = (1 - t)G + tH, H \in \mathcal{P}\},$$

où G est une f.r. fixée, H étant une f.r. décrivant un ensemble donné \mathcal{P} .

Modèle de Kolmogorov :

$$\mathcal{F}_G = \{F/ d_k(F, G) = \sup |F(x) - G(x)| \leq t\},$$

t nombre positif fixé.

M-estimateur :

On sait que l'estimateur T_n est solution de $\int \Psi(x - T_n) dF_n = 0$ et associe à la fonctionnelle T solution de l'équation implicite :

$$\int \Psi(x - T(F)) dF(x) = 0.$$

Nous considérons le modèle de contamination \mathcal{F}'_G précédemment défini.

On suppose que G admet une densité g deux fois dérivable, $(-\log g)$ est convexe si le support de g est compact. Si G est symétrique, g est dite fortement unimodale.

Nous avons, bien entendu :

$$V_{as}(F, T) = \frac{E_F[\Psi^2(X - T(F))]}{[E_F[\Psi'(X - T(F))]]^2} = \frac{\int \Psi^2(x - T(F)) dF(x)}{[\int \Psi'(x - T(F)) dF(x)]^2} \geq \frac{1}{I(F)}.$$

Dans le cas d'une distribution symétrique, $T(F)$ est égal à 0. Nous allons donc remplacer \mathcal{F}'_G par le sous-ensemble :

$$\mathcal{F}_I = \{f \in \mathcal{F} / I(F) < \infty, T(F) = 0\}.$$

Nous pouvons alors énoncer :

a) Théorème d'existence : Il existe une distribution $F_0 \in \mathcal{F}_I$ telle que :

$$\sigma^2(T_0) = \sup_{F \in \mathcal{F}_I} V_{as}(F, T_0) = \frac{1}{I(F_0)} = \inf_{T \in \mathcal{F}} V_{as}(F_0, T) = V_{as}(F_0, T_0).$$

T_0 étant le M-estimateur du minimax défini à partir de :

$$\Psi_0 = -\frac{f'_0}{f_0},$$

où f_0 est la densité de F_0 .

b) Recherche de la solution minimax :

Nous allons déterminer la solution F_0 qui doit donc vérifier la condition :

$$\int (2\Psi'_0 - \Psi_0^2) (dF_1 - dF_2) \geq 0, \forall F_2 \in \mathcal{F}_I$$

où

$$\Psi_0 = -\frac{f'_0}{f_0}.$$

Notons que

$$2\Psi'_0 - \Psi_0^2 = \frac{-4}{\sqrt{f_0}} \left(\sqrt{f_0} \right)''.$$

La solution f_0 va être telle que f atteigne la frontière $(1 - \varepsilon)g$ dans la partie centrale et que :

$$\frac{(\sqrt{f_0})''}{\sqrt{f_0}},$$

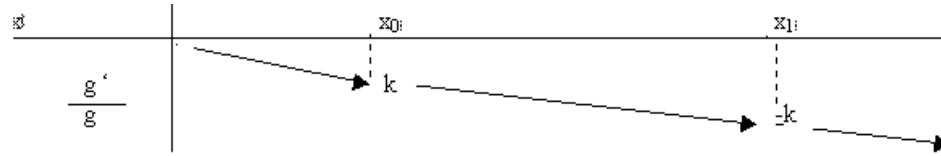
soit constante dans les queues de distribution, ce qui correspond à une solution exponentielle de la forme

$$f_0(x) = C e^{-\lambda|x|}.$$

La solution continue f_0 est :

$$f_0(x) = \begin{cases} (1 - \varepsilon) g(x_0) e^{k(x-x_0)} & \text{si } x \leq x_0 \\ (1 - \varepsilon) g(x) & \text{si } x_0 \leq x \leq x_1 \\ (1 - \varepsilon) g(x_1) e^{-k(x-x_1)} & \text{si } x \geq x_1, \end{cases}$$

où $x_0 = \inf\{x / \left| \frac{g'(x)}{g(x)} \right| \leq k\}$, $x_1 = \sup\{x / \left| \frac{g'(x)}{g(x)} \right| \leq k\}$.



Pour que $\int f_0(x) dx = 1$, il faut que :

$$(1 - \varepsilon) \frac{g(x_0)}{k} + (1 - \varepsilon) \int_{x_0}^{x_1} g(x) dx + (1 - \varepsilon) \frac{g(x_1)}{k} = 1,$$

donc k est défini comme fonction implicite de ε par :

$$\frac{1}{k} [g(x_0) + g(x_1)] + \int_{x_0}^{x_1} g(x) dx = \frac{1}{(1 - \varepsilon)}.$$

On peut vérifier (voir Lecoutre-Tassi-Trognon, (1986)) que F_0 appartient à \mathcal{F}_G et que pour toute densité f d'une f.r. F de \mathcal{F}_I , on a :

$$\int (2\Psi'_0 - \Psi_0^2) (f - f_0) dx \geq 0.$$

Le M-estimateur associé à cette distribution à information minimale est défini par Ψ_0 telle que :

$$\Psi_0(x) = -[\log f_0(x)]' = -\frac{f'_0(x)}{f_0(x)} = \begin{cases} -k & \text{si } x \leq x_0 \\ -\frac{g'(x)}{g(x)} & \text{si } x_0 < x < x_1 \\ k & \text{si } x \geq x_1. \end{cases}$$

c) Cas particulier

si $G = \Phi$, la loi normale centrée réduite, la densité f_0 qui minimise l'information de Fisher est définie par :

$$f_0(x) = \begin{cases} \frac{1 - \varepsilon}{\sqrt{2\pi}} e^{-x^2/2} & \text{si } |x| \leq k \\ \frac{1 - \varepsilon}{\sqrt{2\pi}} e^{(k^2/2) - k|x|} & \text{si } |x| > k, \end{cases}$$

soit :

$$\rho(t) = \begin{cases} \frac{t^2}{2} & \text{si } |t| < k \\ k|t| - \frac{1}{2}k^2 & \text{si } |t| \geq k, \end{cases}$$

avec

$$\int_{-k}^k \varphi(x) dx + \frac{1}{k} [\varphi(k) - \varphi(-k)] = \frac{1}{(1-\varepsilon)},$$

$k = k(\varepsilon)$ est défini par l'équation implicite :

$$1 - 2\Phi(-k) + 2\frac{\varphi(k)}{k} = \frac{1}{(1-\varepsilon)} \text{ ou } 2\left|\frac{\varphi(k)}{k} - \Phi(-k)\right| = \frac{\varepsilon}{(1-\varepsilon)}.$$

Le M-estimateur correspondant est défini par $\Psi_0 = \Psi_H$

La quantité d'information associée est :

$$I(F_0) = E_{p_0} \left(\Psi_0'(x) f_0(x) \right) dx = \int 1_{[-k,k]}(x) f_0(x) dx.$$

D'où :

$$I(F_0) = \int_{-k}^k f_0(x) dx = F_0(k) - F_0(-k) = 1 - 2F_0(-k),$$

avec

$$F_0(-k) = \frac{1-\varepsilon}{\sqrt{2\pi}} e^{(k^2/2)} \int_{-\infty}^{-k} e^{kx} dx = \frac{1-\varepsilon}{k\sqrt{2\pi}} e^{(-k^2/2)} = (1-\varepsilon) \frac{\Phi(k)}{k}.$$

D'après l'équation implicite définissant $k = k(\varepsilon)$:

$$I(F_0) = (1-\varepsilon) [1 - 2\Phi(-k)] = (1-\varepsilon) [2\Phi(k) - 1].$$

ε	0	0,01	0,10	0,5	1
$V_{as}(F_0, T_0)$	1	1,065	1,490	5,928	∞

Remarque : Lorsque ε tend vers 1, k tend vers 0 et F_0 est dégénérée puisque f_0 converge vers 0. Cependant, le M-estimateur associé admet une limite, puisque :

$$\frac{\Psi(x)}{k} \rightarrow \begin{cases} -1 & \text{si } x < x^* \\ +1 & \text{si } x > x^*, \end{cases}$$

où x^* est tel que $\frac{g'(x^*)}{g(x^*)} = 0$; donc $\Psi_0(x) = \text{sign}(x - x^*)$ et T_n est solution de :

$$\sum_{i=1}^n \text{sign}(X_i - T_n - x^*) = 0,$$

soit $T_n = Md_n(X_i) - x^*$ ($\rho(t) = |t|$).

De même, lorsque $\varepsilon \rightarrow 0$, alors k tend vers l'infini, F_0 converge vers Φ et le M-estimateur associé est la moyenne empirique :

$$T_n = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (\rho(t) = t^2).$$

Chapitre 3

Fonction d'inférence quadratique et estimations favorables

3.1 Fonction d'inférence quadratique

Dans cette section, nous présentons la fonction d'inférence quadratique (FIQ) et ses propriétés. Hansen (1982) a étudié les grandes propriétés des estimateurs de la méthode des moments généralisée et Qu (1998) a proposé la FIQ basée sur ces estimateurs.

Supposons $\theta = (\theta_1, \dots, \theta_p)$ et $S(\theta, X) = (S_1(\theta, X), \dots, S_k(\theta, X))^T$ un vecteur des fonctions scores prolongées, satisfait :

$$E(S(\theta, X)) = 0 \quad \text{avec } k \geq p.$$

Alors l'estimateur de la méthode des moments généralisée est défini comme classe d'estimateur indiqué par W :

$$\frac{1}{n} \sum_{i=1}^n S(\theta, X_i)^T W^{-1} \frac{1}{n} \sum_{i=1}^n S(\theta, X_i). \quad (1)$$

Hansen (1982) a prouvé que la variance d'estimateur de la méthode des moments généralisée est réduit au minimum quand :

$$W = V \left(\frac{1}{\sqrt{n}} S(\theta, X_i) \right).$$

Lee (1996) interprète ceci en normalisant les poids : pour que W ci-dessus soit optimal, en réduisant au minimum (1), alors il vaut mieux de normaliser $\sum_{i=1}^n S(\theta, X_i)$.

Chamberlain (1987) prouve que l'estimateur de la méthode des moments généralisée avec W optimal est efficace dans l'état donné :

$$E(S(\theta, X)) = 0$$

avec des observations i.i.d. La forme quadratique dans (1) avec W optimal s'appelle la fonction d'inférence quadratique. Nous pouvons récrire (1) en ignorant la constante

$1/n$. et alors ceci donne une interprétation géométrique que nous présenterons dans le théorème 42. Pour tous vecteurs indépendants $v_1, \dots, v_k \in \mathbb{R}^n$ avec $v_j = (v_{j1}, \dots, v_{jn})^T$, la FIQ est définie comme suit :

$$FIQ(v_1, \dots, v_k) = \begin{pmatrix} \sum v_{1i} \\ \sum v_{2i} \\ \vdots \\ \sum v_{ki} \end{pmatrix}^T \begin{pmatrix} \sum v_{1i}^2 & \cdots & \sum v_{1i}v_{ki} \\ \vdots & & \vdots \\ \sum v_{ki}v_{1i} & \cdots & \sum v_{ki}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum v_{1i} \\ \sum v_{2i} \\ \vdots \\ \sum v_{ki} \end{pmatrix}. \quad (2)$$

L'interprétation est que la FIQ est la longueur carrée de la projection de 1 sur le sous espace engendré par $\{v_1, v_2, \dots, v_k\}$, où :

$$v_i = (S_i(\theta, X_1) \dots, S_i(\theta, X_n))^T, \quad i = 1, \dots, k.$$

3.1.1 Définition du fonction d'inférence quadratique

Théorème 42. Soit v_1, v_2, \dots, v_k des vecteurs indépendants dans \mathbb{R}^n et P_v est une matrice de projection orthogonale sur V , le sous espace engendré par $\{v_1, \dots, v_k\}$. Alors la fonction d'inférence quadratique FIQ est donnée par :

$$FIQ(v_1, \dots, v_k) = \|P_v 1\|^2,$$

où $1 = (1, \dots, 1)^T$.

Propriétés générales :

1- Bornétude : Soit U l'ensemble des vecteurs indépendants $v_1, \dots, v_k \in \mathbb{R}^n$. Alors :

$$\forall U_0 \subset U, \quad FIQ(U_0) \leq FIQ(U) \leq n.$$

2- Invariance : Pour tout scalaire $a_i \in \mathbb{R}$, $a_i \neq 0$, $\forall i = 1, \dots, k$.

$$FIQ(a_1 v_1, \dots, a_k v_k) = FIQ(v_1, \dots, v_k).$$

3- Décomposition : Supposons V_1, \dots, V_q sont des sous espaces orthogonaux de \mathbb{R}^n , engendrés respectivement par :

$$U_1 = \{v_{11}, \dots, v_{1k_1}\}, \dots, U_q = \{v_{q1}, \dots, v_{qk_q}\}.$$

Alors la FIQ est décomposable, c'est-à-dire :

$$FIQ(U_1, \dots, U_q) = \sum_{i=1}^q FIQ(U_i).$$

Exemple : Si on utilise le seul vecteur v_i , nous trouvons que :

$$FIQ(v_i) = \frac{\langle v_i, 1 \rangle^2}{\|v_i\|^2}.$$

Ici, la FIQ satisfait l'inégalité suivante :

$$\max \left(\frac{\langle v_1, 1 \rangle^2}{\|v_1\|^2}, \dots, \frac{\langle v_k, 1 \rangle^2}{\|v_k\|^2} \right) \leq FIQ(v_1, \dots, v_k) \leq n.$$

De la condition $FIQ = n$ (i.e. $1 = \sum_{i=1}^k a_i v_i$), on peut voir que la $FIQ = n$ est toujours atteint si $n = k$. Remarquons que si $n \leq k$, la FIQ n'existe pas.

Soit θ un paramètre à p -dimensions et :

$$S(\theta, X) = (S_1(\theta, X), \dots, S_k(\theta, X))^T$$

vecteur des fonctions scores prolongées, vérifiant :

$$E(S(\theta, X)) = 0.$$

On remplaçons $v_i = (S_i(\theta, X_1), \dots, S_i(\theta, X_n))^T$ dans (2), pour $i = 1, \dots, k$. On obtenons la FIQ basé sur les fonctions scores prolongées.

$$FIQ(S(\theta, \cdot), X) = n \bar{S}_n(\theta, X)^T C_n^{-1}(\theta, X) \bar{S}_n(\theta, X), \quad (3)$$

où :

1. $\bar{S}_n(\theta, X) = \frac{1}{n} \sum_{i=1}^n S(\theta, X_i),$
2. $C_n(\theta, X) = \frac{1}{n} \sum_{i=1}^n S(\theta, X_i) S(\theta, X_i)^T.$

La matrice inverse n'existe pas puisque le rang de C_n est moins de k , la FIQ n'existe pas même si la projection sur le sous espace engendré par le vecteur colonne de $S(\theta, X) = (S(\theta, X_1), \dots, S(\theta, X_n))^T$ existe.

Par conséquent, pour définir la FIQ en utilisant l'inverse généralisé de Moore-Penrose et la mesure de probabilité.

Remarque : La matrice inverse généralisée de Moore-Penrose A^+ existe et unique pour chaque matrice A .

Définition 43. La fonction d'inférence quadratique basé sur la fonction score $S(\cdot)$ est :

$$\begin{aligned} Q(\theta, S, F) &= \int S(\theta, \cdot)^T dF \left[\int S(\theta, \cdot) S(\theta, \cdot)^T dF \right]^+ \int S(\theta, \cdot) dF \\ &= E \left(S(\theta, X)^T \right) \left[E \left(S(\theta, X) S(\theta, X)^T \right) \right]^+ E(S(\theta, X)), \end{aligned} \quad (4)$$

où, θ est un paramètre de p -dimension et $S(\theta, X) = (S_1(\theta, X), \dots, S_k(\theta, X))^T$ est un vecteur des fonctions scores prolongées satisfait : $E(S(\theta, X)) = 0$ au modèle du paramètre vraie.

Exemple : Si on utilise la mesure empirique :

$$F_n = n^{-1} \sum \delta_x,$$

où δ_x est la masse de Dirac au point x . On a :

$$Q(\theta, S, F_n) = \bar{S}_n(\theta, X)^T C_n^+(\theta, X) \bar{S}_n(\theta, X). \quad (5)$$

Remarques :

- 1- Si C_n est inversible, $nQ(\theta, S, F_n)$ est égale à (3)
- 2- Si X_1, \dots, X_n sont i.i.d. avec distribution $F(\cdot)$, alors :

$$Q_n(\theta, S, X) = Q(\theta, S, F_n) \xrightarrow{n \rightarrow \infty} Q(\theta, S, F).$$

Lemme 1 : Soit U sous espace fermé de \mathcal{H} ; l'espace de Hilbert.

- 1- Chaque $f \in \mathcal{H}$ a une décomposition unique :

$$f = P(f) + P(f)^\perp, \quad \forall P(f) \in U \text{ et } P^\perp(f) \in U^\perp.$$

- 2- $P(f)$ et $P(f)^\perp$ sont les points plus proches de f dans U et U^\perp , respectivement.
- 3- Les applications ; $P : H \rightarrow U$ et $P^\perp : H \rightarrow U^\perp$, sont linéaires.
- 4- $\|f\|^2 = \|P(f)\|^2 + \|P^\perp(f)\|^2$.

Pour la démonstration : voir le théorème 4-11 dans Rubin (1987).

Lemme 2 : Soit U un sous espace engendré par $\{S_1, \dots, S_k\}$, de l'espace de Hilbert \mathcal{H} avec le produit scalaire $\langle f, S \rangle = \int f S d\mu$, où μ est une mesure positive. Alors la projection de f sur U est donnée par :

$$P(f) = \int f S^T d\mu \left[\int S S^T d\mu \right]^+ S.$$

Remarque : Pour la démonstration voir Chanseok Park et Bruce G.Lindsay (1999).

Théorème 44. Soit $Q(\cdot)$ est la FIQ défini dans la définition 43.

- 1) $Q(\theta, S, F) = \|P(1)\|^2$, telle que $P(1)$ est la projection orthogonale de 1 sur le sous-espace engendré par $\{S_1, \dots, S_k\}$.
- 2) $Q(\theta, S_0, F) \leq Q(\theta, S, F) \leq 1$, où $S_0(\cdot)$ est la composante du vecteur score prolongé $S(\cdot)$.
- 3) $Q(\theta, S, F) = \sum_{i=1}^q Q(\theta, S_i, F)$, où $S^T = (S_1^T, \dots, S_q^T)$ et tout deux sous espaces engendrés par S_i et S_j sont orthogonaux pour $i \neq j$.

Lemme 3 : Si les fonctions scores prolongées S_1, \dots, S_r sont indépendantes linéairement¹, alors :

$$\int S(\theta, X) S(\theta, X)^T d\mu \text{ n'est pas singulier pour tout mesure positive } \mu.$$

Théorème 45. Supposons qu'il existe r fonctions scores prolongées indépendantes linéairement dans $\{S_1, \dots, S_k\}$ et nous les notons généralement par $S^* = (S_1, \dots, S_r)^T$. Alors :

$$Q(\theta, S, F) = Q(\theta, S^*, F).$$

Démonstration : On écrit $S = AS^*$, où A est une matrice $(k \times r)$ de rang r . Alors nous avons :

$$Q(\theta, S, F) = E(AS^*)^T E \left[(AS^*)(AS^*)^T \right]^+ E(AS^*).$$

¹ $a_1 S_1, \dots, a_r S_r = 0 \iff a_1 = \dots = a_r = 0$.

Soit A est une matrice ($k \times r$) de rang r et B aussi une matrice ($r \times k$) de rang r , alors les résultats suivants retiennent de Graybill, (1983), (Théorèmes 6.2.16 et 6.2.17) :

$$\begin{aligned} (AB)^+ &= A^+B^+, \\ A^+A &= I, \\ BB^+ &= I. \end{aligned}$$

Employons ces résultats et le lemme 3, nous obtenons :

$$Q(\theta, S, F) = E(S^{*T}) E(S^* S^{*T}) E(S^*) = Q(\theta, S^*, F).$$

3.1.2 Efficacité asymptotique

Notre but dans ce paragraphe est de construire un estimateur d'une FIQ , asymptotiquement efficace avec les propriétés de la robustesse.

Soit $X = (X_1, \dots, X_n)$ un échantillon des variables aléatoires et $F_n(\cdot)$ la fonction de distribution empirique donnée par :

$$F_n(X) = \frac{1}{n} \sum_{i=1}^n I(x_i \leq X).$$

L'estimateur de la FIQ basé sur la fonction score prolongée $S(\cdot)$ est défini comme suit :

$$\hat{\theta}_n = \arg \inf_{\theta} Q(\theta, S, F_n) = \arg \inf_{\theta} Q_n(\theta, S, X). \quad (6)$$

Théorème 46. *Soit $S(\cdot)$ un vecteur des fonctions scores prolongées, à k -dimensions et θ paramètre de p -dimensions avec $k \geq p$. Soit X_1, \dots, X_n un échantillon de v.a. avec distribution dépend d'un paramètre $\theta \in \Theta$. Supposons que $S(\theta, x)$ inclue les fonctions scores du maximum de vraisemblance et que l'espace d'échantillon \mathcal{X} ne dépend pas du paramètre θ . Alors $\hat{\theta}_n$ l'estimateur de la FIQ de θ est asymptotiquement efficace.*

Démonstration : Hensen (1982, Théorème.3.1) et Qu (1998, Théorème.3.5.1) montraient que :

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} N_p\left(0, (d_0^T C^{-1} d_0)^{-1}\right),$$

où

$$d_0 = E[S(\theta, X)] \text{ et } C = E[S(\theta, X) S(\theta, X)^T].$$

Par conséquent il suffit de montrer que :

$$d_0^T C^{-1} d_0 = I(\theta),$$

tel que $I(\theta)$ est l'information de Fisher.

Soit $F(x, \theta)$ est une fonction de distribution avec densité $f(x, \theta)$. Sans oublier de noter généralement :

$$S_j(\theta, x) = \frac{\partial}{\partial \theta_j} \log f(x, \theta), \quad j = 1, \dots, p.$$

A partir de : $E[S(\theta, x)] = 0$, $\int S_i(\theta, x) f(x, \theta) dx = 0$, pour $i = 1, \dots, k$ et la dérivation par rapport à $\theta_j, \forall j = 1, \dots, p$.

On trouve :

$$\int S_i \frac{\partial f}{\partial \theta_j} dx = \int S_i S_j f dx = - \int \frac{\partial S_i}{\partial \theta_j} f dx.$$

On a :

$$E(S_i S_j) = -E\left(\frac{\partial S_i}{\partial \theta_j}\right), \quad i = 1, \dots, k, \quad j = 1, \dots, p.$$

Alors il est facile de montrer que $d_0 = CA$ telle que A une matrice $(k \times p)$ satisfait :

$$A = \begin{pmatrix} -1 \\ 0 \end{pmatrix}.$$

Alors on a $d_0^T C^{-1} d_0 = I(\theta)$ car :

$$d_0^T C^{-1} d_0 = d_0^T A = \left(-E\left(\frac{\partial S_i}{\partial \theta_j}\right) \right) = \left(-E\left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(x, \theta)\right) \right),$$

avec $i, j = 1, \dots, p$.

Si C est singulier, on choisit r fonctions scores indépendantes, $S^* = (S_1^*, \dots, S_r^*)$ vérifiait :

$$Q(\theta, S, F) = Q(\theta, S^*, F)$$

On utilise le théorème 45, Alors $C^* = E[S^*(\theta, x) S^{*T}(\theta, x)]$ est non-singulier, d'où le résultat.

Remarque : Ce théorème indique que nous pouvons construire un estimateur asymptotiquement efficace et robuste en choisissant les fonctions scores robustes et les fonctions scores du maximum de vraisemblance. Par exemple, nous pouvons choisir la médiane et les scores moyennes pour le modèle normale.

Dans le prochain paragraphe, nous étudions le point de rupture d'un estimateur comme le M-estimateur qui satisfait l'équation implicite :

$$\sum_{i=1}^n \Psi(X_i, t_n) = 0.$$

Soit les statistiques à valeurs réelles $t_n(x_1, \dots, x_n) = t_n(F_n)$, de la différentiabilité de $Q(\cdot)$ par rapport à θ , on a :

$$\nabla Q(\theta, S, F_n)|_{\theta=t_n} = 2\nabla \overline{S}_n^T C_n^+ \overline{S}_n + \overline{S}_n^T \nabla C_n^+ \overline{S}_n = 0,$$

qui implique que :

$$\sum_{i=1}^n \left(2\nabla \overline{S}_n^T C_n^+ S_n(t_n, x_i) + \overline{S}_n^T \nabla C_n^+ S_n(t_n, x_i) \right) = 0.$$

Ainsi, notre fonction $\Psi(\cdot)$ est :

$$\Psi(x_i, t_n(F_n)) = 2\nabla \overline{S}_n^T C_n^+ S_n(t_n, x_i) + \overline{S}_n^T \nabla C_n^+ S_n(t_n, x_i).$$

Soit $t_n(\cdot)$ un estimateur qui est fonctionnelle², ou on peut remplacer asymptotiquement par une fonctionnelle. Ceci signifie qu'il existe une fonction $t(\cdot)$ tels que :

$$t_n(F_n) \rightarrow t(F)$$

quand les observations sont i.i.d. de la vraie distribution F dans le domaine de (t) . Nous supposons toujours que les fonctionnelles étudiées sont consistes au sens de Fisher³. En considérons $E[S(\theta, X)] = 0$ et ignorons la constante 2 dans $\Psi(x_i, t_n(F_n))$, nous obtenons :

$$\Psi(x_i, t(F)) = E \left[\nabla S(\theta, X)^T \right] \left(E \left[S(\theta, X) S(\theta, X)^T \right] \right)^+ S(\theta, x). \quad (7)$$

Nous trouvons la même conclusion qui a été trouvée dans le théorème 46, en considérant la fonction d'influence quand les fonctions scores prolongées incluent les fonctions scores du maximum de vraisemblance.

²ie, $t_n(F_n) = t(F)$ pour tous n et F_n

³ie, $t_n(F_\theta) = \theta, \forall \theta \in \Theta$

Théorème 47. *Supposons qu'il existe des fonctions scores du maximum de vraisemblance dans les fonctions scores prolongées $S(\theta, x)$ et que l'espace de l'échantillon ne dépend pas du paramètre θ . Alors la fonction d'influence de l'estimateur de la FIQ est la même pour l'estimateur du maximum de vraisemblance. Ce signifie que l'estimateur de la FIQ est asymptotiquement efficace.*

Démonstration :

Soit F_θ une distribution, sa densité est définie par f_θ . Supposons que $S(\theta, x)$ a une fonction score du maximum de vraisemblance, indiqué par :

$$S_1(\theta, x) = \nabla \log f_\theta(x).$$

A partir de $E[S(\theta, X)] = 0$, $\int S_j(\theta, x) f_\theta(x) dx = 0$, $\forall j = 1, \dots, k$, et la différentiabilité par rapport à θ , nous obtenons :

$$\int S_j \nabla f_\theta dx = - \int \nabla S_j f_\theta dx. \quad (8)$$

En utilisant (8), nous trouvons que :

$$E[S_1(\theta, X) S_j(\theta, X)] = -E[\nabla S_j(\theta, X)].$$

Soit $C = E[S(\theta, X) S(\theta, X)^T]$, et C est supposé non-singulier et soit :

$$a^T = E[\nabla S(\theta, X)^T] C^{-1},$$

ou d'une manière équivalente $Ca = E|\nabla S|$, où $a = (a_1, \dots, a_k)^T$, on suit que :

$$\begin{pmatrix} -E\nabla S_1 & -E\nabla S_2 & \dots & -E\nabla S_k \\ -E\nabla S_2 & E[S_2 S_2] & \dots & E[S_2 S_k] \\ \vdots & \vdots & & \vdots \\ -E\nabla S_k & E[S_k S_2] & \dots & E[S_k S_k] \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix} = \begin{pmatrix} E\nabla S_1 \\ E\nabla S_2 \\ \vdots \\ E\nabla S_k \end{pmatrix}.$$

En résolvant pour a , nous trouvons $a_1 = -1$, $a_2 = a_3 = \dots = a_k = 0$ et la solution est unique puisque C est non-singulier. Puisque $\Psi(x_i, \theta) = a^T S(\theta, x)$, nous trouvons :

$$\Psi(x_i, \theta) = -S_1(\theta, x) = -\nabla \log f_\theta(x).$$

Alors nous avons :

$$IF(x, \Psi, F_\theta) = \frac{\Psi(x_i, \theta)}{-\int \frac{\partial}{\partial \theta} \Psi(x_i, \theta) dF_\theta(x)} = \frac{\nabla \log f_\theta(x)}{\int (\nabla \log f_\theta(x))^2 dF_\theta(x)}.$$

Si C singulier, nous choisissons r fonctions scores indépendantes $S^* = (S_1^*, \dots, S_r^*)$ satisfait :

$$Q(\theta, S, F) = Q(\theta, S^*, F).$$

En utilisant le théorème 45. Alors $C^* = E \left[S^*(\theta, X) S^*(\theta, X)^T \right]$ est non singulier.

3.1.3 Point de rupture de l'estimateur de la FIQ

Le point de rupture d'une fonction statistique est la plus petite fraction de la contamination dans les données qui peuvent causer une valeur arbitrairement extrême dans l'estimation. Il y a quelques variations sur la définition. Un de la plupart est appelé le point de rupture de ε -remplacement qui était présenté par Donoho et Huber (1983). Soit $t_n(x)$ un estimateur de l'échantillon x . Nous remplaçons les m points des données dans l'échantillon $x = (x_1, \dots, x_n)$ par des valeurs arbitraires x_1^*, \dots, x_m^* . Nous notons $x^{(m)} = (x_1^*, \dots, x_m^*, x_{m+1}, \dots, x_n)$, qui est appelé l'échantillon de ε -corrompu.

Avant définir le point de rupture, nous définissons le biais maximum qui peut être causer par ε -corruption :

$$Biais(m, t_n, x) = \sup_{x^{(m)}} |t_n(x^{(m)}) - t_n(x)|, \quad (8)$$

où le suprémum est pris sur tout échantillon de ε -corrompu possible.

Définition 48. ε_n^* le point de rupture fini de l'estimateur $t_n(\cdot)$ de l'échantillon $x = (x_1, \dots, x_n)$ est donné par :

$$\varepsilon_n^*(t, x) = \frac{1}{n} \min_m \{m : Biais(m, t, x) = \infty\}. \quad (9)$$

Remarque : En d'habitude, le point de rupture ne dépend pas de l'échantillon $x = (x_1, \dots, x_n)$, et dépend seulement de n la taille de l'échantillon .

Définition 49. Le point de rupture asymptotique est donné par :

$$\varepsilon^* = \lim_{n \rightarrow \infty} \varepsilon_n^*(t, x).$$

Théorème 50. Soit la FIQ basé sur telle fonction score comme $S_1(\theta, x) = x - \theta$ et $S_2(\theta, x) = \Psi(x - \theta)$, où Ψ est monotone, bornée avec :

$$\begin{cases} \Psi(\infty) = -\Psi(-\infty) \\ \Psi(0) = 0. \end{cases}$$

Alors l'estimateur de FIQ pour les modèles de position $\{F_\theta(x) = F_0(x - \theta), \theta \in \mathbb{R}\}$ a un point de rupture asymptotique $\varepsilon^* = \frac{1}{4}$.

Démonstration : le biais infini dans l'équation (9) suggère que nous changions les valeurs des m observations en grand $M > 0$.

Soit $x^{(m)} = (x_1^*, \dots, x_m^*, x_1, \dots, x_{n-m})$, où $x_i^* = M$ pour $i = 1, \dots, m$ et x_i sont l'échantillon de $F_0(x)$.

Soit $\varepsilon = \frac{m}{n}$, $\theta = M\eta$, et $A = \Psi(\infty) = -\Psi(-\infty)$. Alors nous avons les résultats suivantes si $\frac{n}{M} \rightarrow \infty$.

$$\begin{aligned} \frac{1}{n} \sum_{i=0}^n S_1(\theta, x_i) &= (\varepsilon_n - \eta) M + o(M), \\ \frac{1}{n} \sum_{i=0}^n S_2(\theta, x_i) &= \begin{cases} A\varepsilon_n + a_n + o(1) & : \eta = 0 \\ A(2\varepsilon_n - 1) + o(1) & : \eta \in (0, 1) \\ A(\varepsilon_n - 1) + o(1) & : \eta = 1, \end{cases} \\ \frac{1}{n} \sum_{i=0}^n S_1(\theta, x_i)^2 &= (\eta^2 - 2\varepsilon_n\eta + \varepsilon_n) M^2 + o(M^2), \\ \frac{1}{n} \sum_{i=0}^n S_1(\theta, x_i) S_2(\theta, x_i) &= A(\eta - 2\varepsilon_n\eta + \varepsilon_n) M + o(M), \\ \frac{1}{n} \sum_{i=0}^n S_2(\theta, x_i)^2 &= \begin{cases} A^2\varepsilon_n + b_n + o(1) & : \eta = 0 \\ A^2 + o(1) & : \eta \in (0, 1) \\ A^2(1 - \varepsilon_n) + o(1) & : \eta = 1, \end{cases} \end{aligned}$$

où

$$\begin{aligned} a_n &= (1 - \varepsilon_n) \frac{1}{n - m} \sum_{i=1}^{n-m} \Psi(x_i) \\ \text{et } b_n &= (1 - \varepsilon_n) \frac{1}{n - m} \sum_{i=1}^{n-m} \Psi(x_i)^2. \end{aligned}$$

En remplaçant ce qui précède dans (5), nous trouvons le résultat :

$$Q_n(\theta, S_1, S_2, x^{(m)}) = \begin{cases} \varepsilon_n + a_n^2/b_n + o(1) & : \eta = 0 \\ (2\varepsilon_n - 1)^2 + o(1) & : \eta = \frac{1}{2} \\ 1 - \varepsilon_n + o(1) & : \eta = 1 \\ 1 + o(1) & : \text{si non,} \end{cases} \quad (10)$$

Pendant que $M \rightarrow \infty$. Nous voyons que nous avons trois minimum locaux et l'estimateur décompose quand $(2\varepsilon_n - 1)^2 \leq \varepsilon_n + a_n^2/b_n$ ou $1 - \varepsilon_n \leq \varepsilon_n + a_n^2/b_n$. Ce se produit quand $\varepsilon_n \geq \left(5 - \sqrt{9 + 16a_n^2/b_n}\right)/8$. Puis le point de rupture fini est donné par :

$$\varepsilon_n^* = \frac{1}{n} \left[\frac{n}{8} \left(5 - \sqrt{9 + 16a_n^2/b_n} \right) \right],$$

où $[x]$ est le plus petit entier ne pas moins de x . En considérons $a_n \rightarrow \int \Psi(x) dx = 0$, $n \rightarrow \infty$, nous avons le point de rupture asymptotique $\varepsilon^* = \frac{1}{4}$.

3.2 Simulation

Dans cette section nous présentons une vaste étude numérique à évaluer la performance de l'estimateur proposé. Nous comparons la performance d'une variété des estimateurs pour les données continues avec et sans contamination.

ESTIMATION DU PARAMETRE DE POSITION :

Nous comparons l'estimateur de la *FIQ* qui nous avons défini dans les sections passé avec des autres estimateurs de position, ce sont :

l'EMV, la moyenne, la médiane, M-estimateur de Huber et R-estimateur de Hodge-Lehmann.

Nous révisons brièvement ces estimateurs :

L'estimateur de Hodge-Lehmann est défini comme suit :

$$Med_{i \leq j} \left(\frac{X_i + X_j}{2} \right)$$

est un compromis entre la moyenne et la médiane. Il est asymptotiquement efficace sous la distribution logistique et son efficacité relative asymptotique (ERA) par rapport à la moyenne est au moins de 0.864. Il est également robuste avec un point de rupture asymptotique $\varepsilon^* = 1 - \frac{1}{\sqrt{2}} \approx 0.293$.

Le M-estimateur de position de Huber est défini comme solution de l'équation suivante :

$$\sum_{i=1}^n \Psi_k \left(\frac{X_i - \theta}{S_n} \right) = 0,$$

où $\Psi_k(t) = \min\{k, \max(t, -k)\}$ et S_n est un estimateur d'échelle robuste. Ici nous utilisons :

$$\begin{aligned} S_n &= D_n AM = \frac{1}{\Phi^{-1}\left(\frac{3}{4}\right)} \text{Med} \left\{ \left| X_i - \text{Med}(X_j) \right| \right\} \\ &\approx 1.4826 \text{Med} \left\{ \left| X_i - \text{Med}(X_j) \right| \right\}. \end{aligned}$$

où DAM est la déviation absolue médiane. Cet estimateur possède la propriété de $\varepsilon^* = \frac{1}{2}$.

Chan et He (1994) ont proposé un estimateur de position basé sur la combinaison linéaire convexe de la moyenne \bar{X}_n et de la médiane \tilde{X}_n ,

donné par :

$$\hat{\theta} = \pi \bar{X}_n + (1 - \pi) \tilde{X}_n$$

avec $0 \leq \pi \leq 1$. Cet estimateur fait une combinaison adaptative dans le sens de réduire au minimum la variance asymptotique dans la classe des toutes les combinaisons linéaires de la moyenne et de la médiane. Nous l'appellerons l'estimateur de Chan-He.

En considérant la distribution asymptotique commune de les quantiles et de la moyenne de l'échantillon X_1, \dots, X_n (Lin et al. (1980)), donné par :

$$\sqrt{n} \begin{pmatrix} \bar{X}_n - \mu \\ \tilde{X}_n - \xi \end{pmatrix} \xrightarrow{\mathcal{D}} N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \frac{\eta}{2f(\xi)} \\ \frac{\eta}{2f(\xi)} & \frac{1}{4f(\xi)^2} \end{pmatrix} \right), \quad (11)$$

nous avons :

$$\sqrt{n} \hat{\theta} \xrightarrow{\mathcal{D}} N \left(0, \pi^2 \sigma^2 + \frac{(1 - \pi)^2}{4f(\xi)^2} + \frac{\pi(1 - \pi)\eta}{f(\xi)} \right), \quad (12)$$

où $\mu = E(X)$, $\xi = \text{Med}(X)$, $\sigma^2 = E(X - \mu)^2$ et $\eta = E|X - \mu|$.

La variance asymptotique de $\sqrt{n}\hat{\theta}$ atteint son minimum à :

$$\pi_0 = \frac{(1 - 2\eta f)}{(1 - 4\eta f + 4\sigma^2 f^2)}.$$

$\pi \in [0, 1]$, $\pi = \min\{1, \max(\pi_0, 0)\}$ est utilisée comme valeur optimale de π . La variance empirique S^2 et la déviation absolue moyenne :

$$\hat{\eta} = \frac{1}{n} \sum |X_i - \bar{X}_n|$$

sont utilisés comme estimateurs consistents pour σ^2 et η . Quand à l'estimateur de densité $f(\xi)$, on utilise l'estimateur de kernel à la médiane. est donné comme suit :

$$\hat{f}(\tilde{X}_n) = \frac{1}{nh_n} \sum_{i=1}^n \Phi\left(\frac{X_i - \tilde{X}_n}{h_n}\right),$$

où $\Phi(\cdot)$ est la densité de la distribution normale standard, $h_n = 1.06n^{-1/5}\hat{\sigma}$. Pour $\hat{\sigma}$, Silverman (1986, p.47) a recommandé un estimateur d'échelle adaptatif,

$$\hat{\sigma} = \min(\text{déviat}ion\ standard, \text{distance d'interquartile}/1.35). \quad (13)$$

Alors, son point de rupture est 0.25 et peut obtenir de près de 0.5 si on choisit un haut rupture pour l'estimateur d'échelle tel que la déviation absolue médiane (DAM) de la médiane d'échantillon au lieu de $\hat{\sigma}$ dans (13).

Nous incluons les estimateurs basés sur une norme généralisées de rang signée qui est définie comme suit :

$$\|x\|_{\varphi^+} = \sum_{i=1}^n a^+(R(|x_i|)) |x_i|, \quad (14)$$

où $R(|x_i|)$ est un rang de $|x_i|$ dans l'échantillon $x = (x_1, \dots, x_n)$ et

$$a^+(i) = \varphi^+(i/(n+1))$$

où $\varphi^+(u)$ est une fonction positive, non décroissante et carrée intégrable défini sur l'intervalle $[0, 1]$. Alors l'estimateur basé sur cette norme est donné par :

$$\hat{\theta}_{\varphi^+} = \arg \inf_{\theta} \|X - \theta 1\|_{\varphi^+}.$$

L'estimateur $\hat{\theta}_{\varphi^+}$ est asymptotiquement efficace sous le modèle de position symétrique (Hettmansperger et McKean, 1998, p. 45).

En choisissant la fonction score optimale donnée par :

$$\varphi_f^+(u) = -\frac{f' \left(F^{-1} \left(\frac{u+1}{2} \right) \right)}{f \left(F^{-1} \left(\frac{u+1}{2} \right) \right)}.$$

Ceci mène à la médiane pour la distribution double exponentielle, à l'estimateur de Hodges-Lehmann pour la distribution logistique, et à l'estimateur score normale pour la distribution normale. Pour l'estimateur score normal, nous obtenons un point de rupture asymptotique :

$$\varepsilon^* = 2\Phi \left(-\sqrt{\ln 4} \right) \approx 0.239$$

et pour le cas de Cauchy, nous avons $\varepsilon^* = 0.5$.

Nous appellerons cet estimateur R-estimateur optimale. La relation lineaire entre les estimateurs sont comme suit :

Distribution	les estimateurs	
	EMV	R-estimateur optimal
Normale	moyenne	score normale
Exponentielle Double	médiane	médiane
Logistique	$\Psi(u) = \tanh(u/2)$	Hodges-Lehmann
Cauchy	$\Psi(u) = 2u/(1+u^2)$	$\varphi^+(u) = -\sin(2\pi u)$

Pour l'estimateur de *FIQ*, nous besoin de choisir une fonction score. Pour cela nous utilisons le score moyenne et la fonction score basé sur probabilité d'intégrale transforme d'une distribution normale. Ceux-ci sont donnés par

$$\begin{aligned} S_1(\theta, x) &= x - \theta, \\ S_2(\theta, x) &= 2\Phi \left(\frac{x - \theta}{\tau} \right) - 1, \end{aligned}$$

où $\Phi(\cdot)$ est la distribution normale standard.

$$S_2(\theta, x) \xrightarrow{\tau \rightarrow 0} \text{sign}(x - \theta)$$

$\text{sign}(\cdot)$ est la fonction score de la médiane. Cette fonction score peut être pensée comme approximation douce de la médiane. Il a l'avantage que la *FIQ* basé sur ce

score est différentiable tandis que la *FIQ* basé sur le score médiane n'est pas. Il a également comme conséquence d'inférence stable. Nous utilisons cette *FIQ* pour tout les modèle données, nous pouvons obtenir les meilleurs estimateurs si on choisissons le score du maximum de vraisemblance selon le modèle donné au lieu d'employer simplement le $S(x, \theta) = x - \theta$. Nous comparons les deux estimateurs adaptatives : un basé sur la *FIQ* et l'autre basé sur une combinaison lineaire convexe de la moyenne et la médiane. Les résultats numériques montraient que l'estimateur adaptatives de la moyenne et la médiane basé sur la *FIQ* exécutent mieux, particulièrement dans le cas contaminé, que cela basé sur la combinaison convexe linéaire proposée par Chan et He (1994).

Dans le tableau 1, nous présentons les biaise estimés et les erreurs carrées moyennes (ErCM) pour des divers estimateurs basés sur les 10.000 échantillons aléatoires de la distribution normale sans contamination et avec plusieurs dimensions de l'échantillon. Les paramètres de position et d'échelle sont 0 et 1 pour cette distribution.

Dans le tableaux 2, nous répétons les expériences de cette distributions avec contamination $t = 0.1$.

La distribution du contamination est donnée par

$$F(x) = (1 - t)F(x) + t\Delta_{\varsigma},$$

où Δ_{ς} , est une distribution dégénérée avec ς donner par 99.99% quantile de la distribution correspondant, c'est-à-dire :

$$\varsigma = F^{-1}(0.9999) = 3.719,$$

pour la distribution normale. Quand les données viennent du modèle vraie, l'EMV exécute mieux que n'importe quel autre estimateur. Mais quand les données sont contaminées, particulièrement dans le modèle normale, l'EMV exécute tout à fait mal.

Dans le cas normale, notre estimateur perd l'efficacité pour les petites dimensions de l'échantillon, mais il est presque aussi bon que l'EMV quand le dimension de l'échantillon est grand. Dans le cas contaminé, l'estimateur proposé bat ses concurrents. Bien que les estimateurs de Hodges-Lehmann, de Chan-He et de score normale soient robustes, ceux-ci sont très pauvre quand la distribution est asymétrique. Ils sont également des pauvres quand la contamination est asymétrique.

3.3 Estimations favorables pour le modèle de Pareto

La distribution de Pareto d'un paramètre simple joue un rôle très significatif dans les modèles actuels dûe a sa facilité d'application dans la pratique. La fonction de distribution du modèle de Pareto $P(\sigma, \alpha)$ est donnée par :

$$F(x) = 1 - (\sigma/x)^\alpha, \quad x > \sigma \quad (1)$$

où $\alpha > 0$ est un paramètre de forme qui caractérise la queue de la distribution et $\sigma > 0$ est le paramètre d'échelle.

Si on suppose dans $P(\sigma, \alpha)$, σ connu alors le modèle de Pareto à un paramètre simple. L'acceptation de σ connu est tout à fait typique dans la littérature actuelle, par exemple l'états de Philbrick (1985), "bien qu'il puisse y avoir des situations où cette valeur doit être estimée, pratiquement, dans toutes les applications d'assurance, σ sera choisie à l'avance".

Brazauskas et Serfling (2000a, b ; 2001) ont traité l'estimation robuste et efficace de l'index α de le queue, dans le cas où l'échantillon est grand et le modèle a un paramètre ou deux et dans le cas lorsque le modèle a un ou deux paramètres et l'échantillon est petit. Les développements présentés ici, ont utilisé une relation d'équivalence bien connue entre le modele (1) et la distribution d'exponentielle tronquée $E(\mu, \theta)$:

$$G(z) = \left[1 - \exp\left(-\frac{(z - \mu)}{\theta}\right) \right] \times 1\{z > \mu, \theta > 0, -\infty < \mu < +\infty\}, \quad (2)$$

Si la variable aléatoire X a f.d. F donné par (1) alors la variable $Z = \log X$ a f.d. G donné par (2), avec $\mu = \log \sigma$ et $\theta = \alpha^{-1}$.

3.3.1 Qualité des mesures d'ajustement

Soit (X_1, \dots, X_n) un échantillon des variables aléatoires supposés ordonnés d'une façon croissante $X_{(1)} \leq \dots \leq X_{(n)}$ et la fonction de distribution empirique est :

$$F_n(x) = \frac{1}{n} \sum 1\{X_i \leq x\} \quad -\infty \leq x \leq +\infty.$$

Pour l'estimateur $\hat{\alpha}$, soit $\hat{F}(X_{(j)})$ la probabilité de $X_{(j)}$ par le modèle $P(\sigma, \hat{\alpha})$, $\forall j = 1, \dots, n$. Notons que $F_n(X_{(j)}) = \frac{j}{n}$, $\forall j = 1, \dots, n$.

La qualité des statistiques d'ajustement sont alors définies comme suit :

D_n est la statistique de Kolmogorov-Smirnov (KS) :

$$\begin{aligned}
D_n^+ &= \max_{1 \leq j \leq n} \left(\frac{j}{n} - \hat{F}(X_{(j)}) \right), \\
D_n^- &= \max_{1 \leq j \leq n} \left(\hat{F}(X_{(j)}) - \frac{j-1}{n} \right), \\
D_n &= \max\{D_n^+, D_n^-\}.
\end{aligned}$$

W_n^2 est la statistique de Cramer-von Mises (CvM) :

$$W_n^2 = \sum_{j=1}^n \left(\hat{F}(X_{(j)}) - \frac{2j-1}{n} \right)^2 + \frac{1}{12n}.$$

A_n^2 est la statistique d'Anderson-Darling :

$$A_n^2 = -n - \frac{1}{n} \sum_{j=1}^n \left\{ (2j-1) \log \hat{F}(X_{(j)}) + (2n-2j+1) \log \left(1 - \hat{F}(X_{(j)}) \right) \right\}.$$

où $\hat{\alpha}_{MV}$ est l'estimateur du maximum vraisemblance de α .

Remarque : Les résultats développés sont pour le modèle $E(\mu, \theta)$, en raison de la relation d'équivalence entre (1) et (2), ils peuvent également directement appliqués au modèle $P(\sigma, \alpha)$.

3.3.2 Critère de robustesse : Point de rupture

Un critère populaire et efficace pour la robustesse d'un estimateur est son point de rupture (PR). Alors le PR. d'un estimateur mesure le degré de résistance de l'influence des observations périphériques qui représentent probablement la contamination des données plutôt l'exception des observations extrêmes produites par le modèle paramétrique.

Types des points de ruptures :

1. **PRI (inférieure) :** La plus grande proportion d'inférieure observation de l'échantillon qui peut être prise à une limite inférieure sans prendre l'estimateur à une limite selon le paramètre étant estimé.
2. **PRS (supérieure) :** La plus grande proportion de supérieure observation de l'échantillon qui peut être prise à une limite supérieure sans prendre l'estimateur à une limite selon le paramètre étant estimé.

Remarque : Dans le traitement nous favorisons les estimateurs qui ont les PRS non nul.

3.3.3 Critère d'efficacité : Variance

Si les observations d'échantillon suivent le modèle paramétrique postulé, alors il est bien connu que pour des grandes données et parmi une grande classe des estimateurs, l'EMV atteint le possible minimum de la variance . Par conséquent, il peut être considéré comme un repère quantitatif pour la considération d'efficacité.

En particulier, pour le modèle de Pareto, l'EMV de α est donné en Arnold (1983) par :

$$\hat{\alpha}_{MV} = \frac{n}{\sum_{i=1}^n \log \left(\frac{X_i}{\sigma} \right)}.$$

$\frac{2n\alpha}{\hat{\alpha}_{MV}}$ suit la loi de χ_{2n}^2 . Où χ_{2n}^2 est la distribution de chi-deux avec $2n$ le degré de liberté.

Nous trouvons $\hat{\alpha}_{MVS}$ l'estimateur sans biais de α , si nous multiplions $\hat{\alpha}_{MV}$ par le facteur $\frac{(n-1)}{n}$.

Pour d'autre détails sur les théories du distribution exacte de $\hat{\alpha}_{MV}$, voir par exemple Rytgaard (1990)].

Remarque : $\hat{\alpha}_{MVS}$ est asymptotiquement normale avec moyenne α et variance α^2/n , pour n grand. Même résultat pour des estimateurs ont même moyenne que $\hat{\alpha}_{MVS}$ et variance $c\alpha^2/n$, $c > 1$. $\hat{\alpha}_{MVS}$ est une estimateur efficace mais non robuste.

3.3.4 Estimateurs

3.3.4.1-Estimateur des quantiles : Nous décrivons cette approche pour le cas où σ est traité comme un paramètre inconnu. Les estimateurs de quantiles basé sur $k \geq 2$ niveaux des quantiles, $0 < p_1 < \dots < p_k < 1$ sont définies comme suit :

$$\hat{\alpha}_Q = \left(\sum_{i=1}^k b_i \log X_{[np_i]} \right)^{-1}, \quad \hat{\sigma} = \exp\{\log X_{[np_1]} - u_1/\hat{\alpha}_Q\},$$

avec :

$$\begin{cases} b_1 = \frac{1}{L} \frac{u_2 - u_1}{e^{u_2} - e^{u_1}}, \\ b_i = \frac{1}{L} \left[\frac{u_i - u_{i-1}}{e^{u_i} - e^{u_{i-1}}} - \frac{u_{i+1} - u_i}{e^{u_{i+1}} - e^{u_i}} \right], \quad 2 \leq i \leq k-1, \\ b_k = \frac{1}{L} \frac{u_k - u_{k-1}}{e^{u_k} - e^{u_{k-1}}}, \end{cases}$$

et :

$$L = \sum_{i=2}^k \frac{(u_i - u_{i-1})^2}{e^{u_i} - e^{u_{i-1}}},$$

où $u_i = -\log(1 - p_i)$, $1 \leq i \leq k$, et $[x]$ est le moins entier $\geq x$. Tels estimateurs étaient présentés et étudiés pour le modèle $P(\sigma, \alpha)$ par Quandt (1966) pour $k = 2$ et par Koutrouvelis (1981) pour le cas général $k \geq 2$.

En choisissant le minimum de le déterminant de la matrice covariance asymptotique pour les estimateurs de σ et α comme un critère d'optimalité, Koutrouvelis (1981) trouvait que le choix optimal de p_1 est toujours :

$$p_1^\circ = \frac{1}{n + 0.5}.$$

Les restants des niveaux de quantiles optimales sont :

Pour $k = 2$, $p_2^\circ = 1 - (1 - p_1^\circ) \exp(-1.5936) \approx 80$.

$$\text{Pour } k = 5, \begin{cases} p_2^\circ = 1 - (1 - p_1^\circ) \exp(-0.6003) \approx 45, \\ p_3^\circ = 1 - (1 - p_1^\circ) \exp(-1.3544) \approx 74, \\ p_4^\circ = 1 - (1 - p_1^\circ) \exp(-2.3721) \approx 91, \\ p_5^\circ = 1 - (1 - p_1^\circ) \exp(-3.9657) \approx 98. \end{cases}$$

Pour $k = 5$, $p_1 = 13$, $p_2 = 315$, $p_3 = 50$, $p_4 = 685$, $p_5 = 87$.

On note par $\alpha_Q^{opt,k}$ l'estimateur optimale de α , et on considère aussi le cas d'un estimateur non optimale noté par α_Q^* .

Remarque : Quand le nombre de quantiles k est choisi pour égaliser le nombre de paramètres inconnu du modèle, cette méthode correspond à ce qu'on appelle le percentile d'assortiment (Keugman, Panjer, et Willmot (1998)).

3.3.4.2-Estimateur de la moyenne tronquée : Soit β_1 et β_2 avec $\beta_1 \geq 0$, $\beta_2 < \frac{1}{2}$ une moyenne tronquée est constitué en jetant les observations les plus basses de la proportion β_1 et les observations les plus élevées de la proportion β_2 et en faisant la moyenne des restants. pour α , on présente l'estimateur de la moyenne tronquée comme suit :

$$\hat{\alpha}_T = \left(\sum_{i=1}^n c_{ni} \log \left(\frac{X_{(i)}}{\sigma} \right) \right)^{-1},$$

avec :

$$c_{ni} = 0, \text{ pour } 1 \leq i \leq [n\beta_1], c_{ni} = 0 \text{ pour } n - [n\beta_2] + 1 \leq i \leq n,$$

$$\text{et } c_{ni} = \frac{1}{d(\beta_1, \beta_2, n)} \text{ pour } [n\beta_1] + 1 \leq i \leq n - [n\beta_2].$$

Où :

$$d(\beta_1, \beta_2, n) = \sum_{j=[n\beta_1]+1}^{n-[n\beta_2]} \sum_{i=0}^{j-1} (n-i)^{-1}.$$

Remarque : Ces estimateurs de la moyenne tronquée étaient étudiés par Kimber (1983a,b) pour le problème d'estimation équivalent à $\theta = \alpha^{-1}$ dans le modèle $E(\mu, \theta)$ avec μ connu. Les c_{ni} sont bien choisis pour construire un estimateur sans biais $\hat{\theta}_T = \hat{\alpha}_T^{-1}$ pour $\theta = \alpha^{-1}$.

3.3.4.3-Estimateur de la médiane généralisée : Les statistiques de la médiane généralisée sont définies de prendre la médiane de les $\binom{n}{k}$ évaluations d'un noyau donné $h(x_1, \dots, x_k)$ sur tout k ensemble des données (voir Serfling (1984, 2000) pour une discussion général).

Dans Brazauskas et Serfling (2000a), l'estimateur était considéré pour α où σ connu est :

$$\hat{\alpha}_{MG} = Med\{h(X_{i_1}, \dots, X_{i_k})\},$$

avec un noyau particulier $h(x_1, \dots, x_k)$:

$$h(x_1, \dots, x_k, \sigma) = \frac{1}{C_k} \frac{k}{\sum_{j=1}^k \log(x_j/\sigma)},$$

où C_k est un facteur multiplicatif de médiane non biais, c'est-à-dire nous le choisissons pour que la distribution de $h(X_{i_1}, \dots, X_{i_k}, \sigma)$ a un médiane α , les valeurs de C_k , pour $k = 2, \dots, 9$, sont fournis dans le tableau suivant :

k	2	3	4	5	6	7	8	9
C_k	1.1916	1.1219	1.0893	1.0705	1.0582	1.0495	1.0431	1.0382

Chapitre 4

Applications de la méthode de médiane pour les modèles de Weibull et de Paréto

4.1 Modèle de Weibull

La distribution de Weibull est fréquemment employée dans les modèles de temps d'échec en sciences médicales et biologiques aussi dans la technologie. Les applications de la distribution de temps d'échec peuvent être trouvées dans la recherche qui implique les maladies humaines comme le cancer et la maladie de Hodgkin (Bartolucci et Dickey, 1977)...etc.

La fonction de densité de la distribution de Weibull de deux paramètre est :

$$f(x, \lambda, \beta) = \lambda\beta (\lambda x)^{\beta-1} \exp \left[-(\lambda x)^\beta \right]. \quad (1)$$

La fonction de risque : de distribution de Weibull a une forme plus simple :

$$\lambda(x) = \lambda\beta (\lambda x)^{\beta-1},$$

cette fonction est décroissante pour $\beta < 1$, croissante pour $\beta > 1$ et constante pour $\beta = 1$. Où β est le paramètre de forme et λ le paramètre d'échelle.

Une méthode d'estimation pour les paramètres de Weibull est la méthode du maximum de vraisemblance, mais elle manque les points aberrants supérieures et inférieures de la robustesse, spécialement si $\beta > 0$.

4.1.1 Méthode de médiane

Le paramètre $\theta = (\lambda, \beta)$ doit estimer à partir de l'échantillon aléatoire X_1, \dots, X_n . la fonction score de la distribution (1) est :

$$S_\theta(x) = \frac{\partial \log f(x, \theta)}{\partial \theta}.$$

Alors :

$$S_\theta(x) = \left(\begin{array}{c} (\beta/\lambda) \left(1 - (\lambda x)^\beta\right) \\ \left(1 + \left(1 - (\lambda x)^\beta\right) \log (\lambda x)^\beta\right) / \beta \end{array} \right).$$

La méthode de médiane égalise la médiane d'échantillon de $S_\theta(X_i)$ avec la médiane de population. Puisque $Y = (\lambda X)^\beta$ a une distribution exponentielle avec moyenne égale à 1, l'estimation de $(\hat{\lambda}, \hat{\beta})$ satisfait :

$$Med_i \left((\lambda X_i)^\beta \right) = \log 2, \quad (2)$$

$$Med_i \left[\left(1 - (\lambda X_i)^\beta \right) \log \left((\lambda X_i)^\beta \right) \right] = c, \quad (3)$$

où :

$$c = Med \left((1 - Y) \log Y \right) \approx -0.51,$$

à partir de (2), il trouve que $\hat{\lambda} = (\log 2)^{1/\hat{\beta}} / Med_i\{X_i\}$. Par conséquent, le calcul est réduit à résoudre l'équation (3) avec le paramètre de forme β . Dans les applications caractéristiques, le côté à gauche de (3) est une fonction de β décroissante, qui la fait facile pour trouver la solution. Dans le cas où la moitié d'observations ou plus, dans l'échantillon prennent la même valeur, aucune solution peut être trouvée.

L'estimateur de la méthode de médiane tombe dans la classe des M-estimateurs (Hampel et al. 1986) sous forme de :

$$\sum_i h(S_\theta(X_i) - m_\theta) = 0, \quad (4)$$

où m_θ est déterminé par :

$$Eh(S_\theta(X) - m_\theta) = 0.$$

Le choix du $h(x) = x$ correspond à l'EMV, mais $h(x) = \text{sgn}(x)$ donne l'EMMed. D'autres choix de h sont également possibles, mais ils exigent plutôt le calcul numérique d'une fonction compliquée m_θ .

2.1. Fonction d'influence : Pour mesurer la sensibilité d'un estimateur, Hampel et al. (1986) avaient donné le concept de la fonction d'influence. En général, la fonction d'influence $IF(x; T)$ pour un estimateur T indique combien l'estimateur peut changer en réponse pour avoir un nombre petit des observations placées au point x . Le changement est mesuré relativement à la proportion de les points contaminés. Pour une définition exacte, (voir Hampel et al., 1986, p. 84). Comme était montré en Hampel et al. (1986), la fonction d'influence d'un tel M-estimateur est :

$$IF(x, h) = -V^{-1}h(S_\theta(x) - m_\theta), \quad (5)$$

où :

$$V = \left. \frac{dE_\theta h(\Psi_\theta(x) - m_\theta)}{dt} \right|_{t=\theta}$$

est une matrice (2×2) . Dans le cas du maximum de vraisemblance, $-V$ est la matrice de l'information de Fisher :

$$I_{\theta} = \begin{pmatrix} (\beta/\lambda)^2 & i_{12}/\lambda \\ i_{12}/\lambda & i_{22}/\beta \end{pmatrix},$$

où :

$$i_{12} = E\{(1 - Y)(1 + \log Y - Y \log Y)\} \approx 0.4228,$$

$$i_{22} = E(1 + \log Y - Y \log Y)^2 \approx 1.8237,$$

et la fonction d'influence est $IF(x, EMV) = (I_{\theta})^{-1}S_{\theta}(x)$. De plus, l'influence n'est pas bornée car x se déplace vers 0 ou l'infini, indiquant que l'EMV pourrait être sévèrement déformé par un petit nombre des points aberrants (outliers).

Pour obtenir la fonction d'influence de l'estimateur de la méthode de médiane, nous devons seulement calculer la matrice V

$$V_{Med} = \begin{pmatrix} v_{11}\beta/\lambda & v_{12}/\beta \\ v_{21}\beta/\lambda & v_{22}/\beta \end{pmatrix},$$

où :

$$\begin{cases} v_{11} = -E\{sign(\log 2 - Y)(1 - Y)\} \approx -0.6931, \\ v_{12} = -E\{sign(\log 2 - Y)(1 + \log Y - Y \log Y)\} \approx 0.2541, \\ v_{21} = -E\{sign(1 - Y)(\log Y - Y \log Y - c)\} \approx -0.0354, \\ v_{22} = -E\{sign(1 + \log Y - Y \log Y)(\log Y - Y \log Y - c)\} \approx -0.8368. \end{cases}$$

La fonction d'influence de médiane est donc donnée par :

$$IF(x, Med) = |v_{11}v_{22} - v_{12}v_{21}|^{-1} \begin{pmatrix} v_{22}\lambda/\beta & -v_{11}\lambda/\beta \\ -v_{12}\beta & v_{11}\beta \end{pmatrix} \times$$

$$\begin{pmatrix} sign(\log 2 - (\lambda x)^{\beta}) \\ sign\left(\left(1 - (\lambda x)^{\beta}\right) \log(\lambda x)^{\beta}\right) \end{pmatrix},$$

est bornée pour tout $x \in (0, \infty)$.

Remarque : Pour la matrice V_{Med} , le premier élément diagonal est donnée par la dérivative de :

$$g(\tilde{\lambda}) = \int_0^{\infty} sign(\log 2 - (\tilde{\lambda}x)^{\beta}) \lambda \beta (\lambda x)^{\beta-1} \exp[-(\lambda x)^{\beta}] dx$$

$$= \int_0^{\infty} sign(\log 2 - y) \left(\lambda/\tilde{\lambda}\right)^{\beta} \exp\left[-\left(\lambda/\tilde{\lambda}\right)^{\beta} y\right] dy$$

comme une fonction de $\tilde{\lambda}$, $\tilde{\lambda} = \lambda$, qui est égale à $-(\beta/\lambda)E\{sign(\log 2 - Y)(1 - Y)\}$. Les autres éléments de V est calculé de la même méthode.

2.2-Robustesse : Puisque l'estimateur de la méthode de médiane est obtenu par des opérations médianes, il a une propriété de robustesse intéressante quand la fraction des observations extrêmes est bien censurée. Pour énoncer le résultat plus précisément, soit c_1 et c_2 les solutions de l'équation $(1 - y) \log(y) = c$, où c est la même constante utilisée dans (3). Avec des calculs simples, on montre que $c_1 \approx 0.417$ et $c_2 \approx 1.8379$.

Soient l'échantillon X_1, \dots, X_n , $(\hat{\lambda}, \hat{\beta})$ l'estimateur de la méthode de médiane, et $y_i = (\hat{\lambda}x_i)^\beta$ les indices estimés. Alors, l'estimation restera sans changement si n'importe quelle observation est déplacée à une valeur plus petite, avec $y_j < c_1$ ou si quelle observation est relevée, avec $y_j > c_2$.

Sous le modèle de weibull :

1. $P((\lambda X)^\beta < c_1) \approx 0.34 \implies$ si moins de 34% les plus petites observations sont censurées à gauche.
2. $P((\lambda X)^\beta > c_2) \approx 0.16 \implies$ si moins de 16% les plus grandes observations sont censurées à droite

Remarque : Nous pouvons calculer l'EMMed comme si aucune censure ne s'était présentée.

Quelque soit l'ensemble des données spécifiques, nous pouvons confirmer que l'estimation n'est pas affectée par la censure gauche ou droite, en calculant $y_j = (\hat{\lambda}x_j)^\beta$ pour les observations censurées. Si elles ont tous moins de 0.417 ou grandes que 1.8379, la censure n'a aucun effet sur l'estimation.

Dans la littérature de robustesse, le point de rupture est souvent employé comme une mesure globale dure de la robustesse contre les points aberrants multiples. Dans ce cas, il est défini pour être la plus petite proportion des observations qui doivent être remplacées par des valeurs arbitraires pour l'estimateur de λ ou β , arbitrairement près de zéro ou l'infini.

Il est facile de voir que le point de rupture pour l'EMV pour les paramètres de weibull est $1/n$, signifiant qu'un point aberrant simple pourrait détruire les estimations, l'EMMed a un point de rupture $\varepsilon^* = 1/2$, c'est-à-dire, l'estimation de l'un ou l'autre paramètre restera dans un intervalle fini bornée loin de 0 ou l'infini à condition que moins de la moitié des observations soient contaminées.

2.3-Efficacité : Quelque soit la fonction score discontinue utilisée pour la méthode de médiane, l'analyse asymptotique de l'estimateur est encore possible sous le cadre de travail de Huber (1967). Une application très facile du corollaire 2.2 de He et Shao (1996) et quelques calculs courants montrent que :

$$\sqrt{n}(\hat{\theta}_{n,Med} - \theta) \xrightarrow[n \rightarrow \infty]{} N(0, \Sigma),$$

où :

$$\Sigma = (V_{Med}^{-1}) \begin{pmatrix} 1 & b \\ b & 1 \end{pmatrix} (V_{Med}^{-1})^T \approx \begin{pmatrix} 1.759(\lambda/\beta)^2 & -0.176\lambda \\ -0.176\lambda & 1.440\beta^2 \end{pmatrix},$$

avec :

$$b = E\{sgn(\log 2 - Y)(\log Y - Y \log Y - c)\} \approx -0.3636.$$

Pour la comparaison, la matrice de covariance asymptotique de l'EMV est :

$$I_{\theta}^{-1} = \begin{pmatrix} 1.109(\lambda/\beta)^2 & -0.257\lambda \\ -0.257\lambda & 0.608\beta^2 \end{pmatrix}.$$

L'efficacité relative asymptotique de la méthode de médiane contre l'EMV soit égale à 63% pour le paramètre d'échelle et 42% pour le paramètre de forme.

La faiblesse de l'efficacité asymptotique de $\hat{\theta}_{n,Med}$, soit conformée à la grande variabilité de l'estimation dans les petits échantillons.

2.4-Inférence robuste : L'estimation du paramètre n'est pas habituellement l'objectif final dans l'analyse des données. Par exemple, nous pouvons examiner si la fonction de risque augmente, ou détecter les valeurs relatives périphériques à la majorité des données. Dans ce paragraphe, nous considérons comment la méthode de médianes peut être utile dans ces applications. En particulier, nous discutons sur l'utilisation des intervalles de confiance.

Intervalle de confiance : L'intervalle de confiance approximatif était obtenu basé sur les distributions asymptotiques de les estimateurs de paramètres. Dans la plupart de temps, on s'intéresse sur le paramètre de forme β . Dans cet égard, nous avons une transformation stabilisante d'une variance simple :

$$\sqrt{n} \left(\log \hat{\beta}_{n,Med} - \log \beta \right) \rightarrow N(0, 1.440).$$

Le même résultat pour $\log(\hat{\beta}_{n,EMV})$ sauf que la distribution limitée est $N(0, 0.608)$. Par conséquent, nous obtenons un intervalle de confiance pour β , basé sur la Med de niveau $(1 - \alpha) 100\%$, comme suit :

$$\left(\hat{\beta}_{n,Med} \exp\{-1.2z_{\alpha/2}/\sqrt{n}\}, \hat{\beta}_{n,Med} \exp\{1.2z_{\alpha/2}/\sqrt{n}\} \right).$$

Par contre, l'intervalle de confiance basé sur l'EMV est :

$$\left(\hat{\beta}_{n,EMV} \exp\{-0.78z_{\alpha/2}/\sqrt{n}\}, \hat{\beta}_{n,EMV} \exp\{0.78z_{\alpha/2}/\sqrt{n}\} \right).$$

Les intervalles de confiance basé sur *Med* sont généralement plus larges à moins que $\hat{\beta}_{n,EMV}$ soit beaucoup gonflé vers le haut par les points aberrants. Cependant, la perte dans les niveaux de confiance pour les intervalles basés sur l'EMV sont souvent substantielle. En fait, la probabilité d'assurance peut être petite avec les intervalles basés sur l'EMV quand les points aberrants présents.

4.1.2 Comparaison par simulation

Cette section rapporte une partie de notre expérience avec les procédures proposées dans les sections passées. Elles sont comparées avec la méthode du maximum de vraisemblance. Pour l'estimation des paramètres, nous rapportons :

1. l'erreur carrée moyenne (ErCM) pour l'estimation normalisée $\hat{\lambda}(\sqrt{n}\beta_0/\lambda_0)$,
2. l'erreur carrée moyenne (ErCM) pour l'estimation normalisée $\hat{\beta}(\sqrt{n}/\beta_0)$,

où λ_0 et β_0 sont les valeurs des paramètres pour la distribution fondamentale de Weibull. Les estimations sont normalisées de sorte que l'ErCM soit sur la même échelle pour différentes valeurs de λ_0 et β_0 . Noter que sous le modèle de Weibull, les variances asymptotiques de ces estimations normalisées sont indépendant.

Nous étudions la situation que la distribution fondamentale de Weibull est contaminée par une autre distribution de Weibull ou une distribution uniforme. La première peut exister quand un nombre d'observations dans l'échantillon viennent d'une population différente, la dernière peut modeler l'occurrence de quelques erreurs négligentes.

Soit λ_1 et β_1 les paramètres de la distribution contaminante si c'est de Weibull. Sans perte de généralité, nous plaçons toujours $\lambda_0 = 1$. Nous rapportons les résultats avec $n = 20$ et $n = 100$ avec pourcentage de contamination étant 10%. Cinq mille répétitions sont employées pour chaque cas. Le générateur pseudo-aléatoire de nombre de Wichmann et de Hill (1982) est utilisé.

Le tableau 3 donne les erreurs carrées moyennes des estimations normalisées et leurs erreurs standards pour plusieurs cas différent quand $n = 20$. Le cas *a* n'a aucune contamination, les cas *b* et *e* contiennent les points aberrants modestes supérieures de $\lambda_1 = 0.2$. Les cas *f* et *h* contiennent des points aberrants supérieures et inférieures avec $\beta = 0.2$ ou 0.5 , et les cas restants contiennent les points aberrants de la distribution uniforme $U(0, 20)$. Les détails de ces cas sont mentionnés sous le tableau 3. Par conclusion nous avons :

- (i) Quand il y a de contamination (cas *a*), l'EMV exécute mieux comme prévu. L'EMMed est moins efficace.

(ii) Quand $\beta_0 \leq 1$, l'EMV n'est pas beaucoup affecté par les points aberrants modestes dans les échantillons contaminées.

(iii) Quand le paramètre de forme β_0 est croissant, la méthode de médiane est fortement robuste. C'est évident pour les cas d, c, g et h avec $\beta_0 = 3$ ou 10 que l'ErCM de l'EMV est beaucoup de fois plus haut que ceux du méthode basé sur la médiane. De plus grandes valeurs de β , sont tout à fait commune dans la pratique. rendraient l'EMV plus sensible aux points aberrants.

(iv) Quand le contaminant est une distribution uniforme $U(0,20)$, on observe le manque de la robustesse pour l'EMV même lorsque $\beta_0 = 1$.

Tableau 4 donne les résultats de la simulation quand l'échantillon de taille $n = 100$. Il est clair que la méthode de médiane est nettement meilleur dans les cas avec ou sans contamination dans les modèles avec le paramètre β très petit.

4.2 Modèle de Pareto

Dans les applications du théorie des valeurs extrêmes en estimant les quantiles élevés, les probabilités et les points limites des queues extrêmes, une approche généralement utilisée dans la littérature pour modeler des excès au-dessus d'un seuil élevé est la distribution de Pareto généralisée. C'est-à-dire, nous supposons que les variables excessives X_1, \dots, X_n sont i.i.d et la fonction de distribution est :

$$F_{\gamma,\sigma}(x) = 1 - \left(1 + \frac{\gamma x}{\sigma}\right)^{-1/\gamma}, \quad (1)$$

où $\sigma > 0$, $\gamma \in \mathbb{R}$, $0 \leq x < \infty$, pour $\gamma \geq 0$ et $0 \leq x \leq -\sigma/\gamma$ pour $\gamma < 0$ (voir Embrechts, Klüppelberg et Mikosch, (1997)).

4.2.1 Méthode de médiane

He et Fung (1999) proposaient, pour les paramètres de Weibull, une procédure d'estimation robuste. Ils l'ont appelée la méthode de médiane. L'idée de base de cette méthode est d'égaliser la médiane de l'échantillon avec chaque composante de la fonction score à la médiane correspondante de population.

Nous allons appliquer cette méthode pour les paramètres de la distribution de Pareto.

2.1-Dérivation des estimateurs : Il est facile de vérifier que la densité de $F_{\gamma,\sigma}$ est :

$$f_{\gamma,\sigma}(x) = \frac{1}{\sigma} \left(1 + \frac{\gamma x}{\sigma}\right)^{-\frac{1}{\gamma}-1} \{1(x > 0, \gamma \geq 0) + 1(0 < x < -\sigma/\gamma, \gamma < 0)\}.$$

Les composantes de la fonction score du modèle (1) sont :

$$\begin{aligned}\frac{\partial}{\partial \sigma} \log f_{\gamma, \sigma}(x) &= \left\{ -\frac{1}{\sigma} + \frac{1+\gamma}{\sigma^2} \frac{x}{1+\gamma x/\sigma} \right\} \times \{1(x > 0, \gamma \geq 0) + \\ &\quad 1(0 < x < -\sigma/\gamma, \gamma < 0)\}, \\ \frac{\partial}{\partial \gamma} \log f_{\gamma, \sigma}(x) &= \left\{ \frac{\log(1+\gamma x/\sigma)}{\gamma^2} - \frac{(1+\gamma)x}{\sigma\gamma + \gamma^2 x} \right\} \times \{1(x > 0, \gamma \geq 0) + \\ &\quad 1(0 < x < -\sigma/\gamma, \gamma < 0)\}.\end{aligned}$$

Nous remarquons que, pour $\gamma \neq -1$

$$P \left\{ -\frac{1}{\sigma} + \frac{1+\gamma}{\sigma^2} \frac{X_i}{1+\gamma X_i/\sigma} \leq -\frac{1}{\sigma} + \frac{1+\gamma}{\sigma\gamma} (1-2^{-\gamma}) \right\} = \frac{1}{2},$$

et $-\frac{1}{\sigma} + \frac{1+\gamma}{\sigma^2} \frac{x}{1+\gamma x/\sigma}$ est une fonction monotone d'une variable x .

En appliquant la méthode de médiane sur la première composante de la fonction score, nous obtenons :

$$-\frac{1}{\sigma} + \frac{1+\gamma}{\sigma^2} \frac{\text{Med}(X_i)}{1+\frac{\gamma}{\sigma}\text{Med}(X_i)} = -\frac{1}{\sigma} + \frac{1+\gamma}{\sigma\gamma} (1-2^{-\gamma})$$

équivalente à :

$$\sigma = \frac{\gamma}{2^\gamma - 1} \text{Med}(X_i). \quad (2)$$

La dérivation de (2) suivie par $F_{\gamma, \sigma}^{-1}(1/2) = \sigma\gamma^{-1}(2^\gamma - 1)$. Pour la deuxième composante de la fonction score, on pose :

$$U = \left(1 + \frac{\gamma X_i}{\sigma} \right)^{-1/\gamma},$$

U est uniformément distribué sur $(0, 1)$. Puisque :

$$\frac{\log(1+\gamma X_i/\sigma)}{\gamma^2} - \frac{(1+\gamma)X_i}{\sigma\gamma + \gamma^2 X_i} \stackrel{D}{=} -\frac{\log U}{\gamma} - \frac{1+\gamma}{\gamma^2} (1-U^\gamma),$$

on obtien :

$$\text{Med} \left(\frac{\log(1+\gamma X_i/\sigma)}{\gamma^2} - \frac{(1+\gamma)X_i}{\sigma\gamma + \gamma^2 X_i} \right) = z(\gamma), \quad (3)$$

où $z(\gamma)$ satisfait :

$$\int_{\{0 < y < 1, -\frac{\log y}{\gamma} - \frac{1+\gamma}{\gamma^2}(1-y^\gamma) > z(\gamma)\}} dy = \frac{1}{2}. \quad (4)$$

Donc, l'estimateur de la méthode de médiane $(\hat{\gamma}, \hat{\sigma})$ est une solution de les équations (2), (3) et (4).

$z(\gamma)$ est définie implicitement et de structure compliquée qui à rendu (4) à usage difficile. Nous pouvons obtenir une autre simplification de l'équation (4) utilisable. On notons la fonction

$$h(y) = -\frac{\log y}{\gamma} - \frac{(1+\gamma)(1+y^\gamma)}{\gamma^2},$$

où h est une fonction monotone si $\gamma \leq 1$, mais si $\gamma > -1$ nous pouvons montrer que :

$$\begin{cases} h'(y) < 0 \text{ si } y \in \left(0, \left(\frac{1}{1+\gamma}\right)^{1/\gamma}\right), \\ h'(y) > 0 \text{ si } y \in \left(\left(\frac{1}{1+\gamma}\right)^{1/\gamma}, 1\right), \\ h(0) \rightarrow \infty, h(1) = 0 \text{ et } h\left(\left(\frac{1}{1+\gamma}\right)^{1/\gamma}\right) < 0. \end{cases}$$

Par conséquence (4) est équivalente à l'existence de $y_1 = y_1(\gamma) \in \left(0, \frac{1}{2}\right]$ tel que :

$$z(\gamma) = -\frac{\log y_1}{\gamma} - \frac{1+\gamma}{\gamma^2}(1-y_1^\gamma), \quad (5)$$

où y_1 satisfait : $y_1 = 1/2$ si $\gamma \leq -1$ ou si $\gamma > -1$ et $h(1/2) \geq 0$ et

$$-\frac{\log y_1}{\gamma} - \frac{1+\gamma}{\gamma^2}(1-y_1^\gamma) = -\frac{\log(y_1 + 1/2)}{\gamma} - \frac{1+\gamma}{\gamma^2}(1-(y_1 + 1/2)^\gamma) \text{ si non.} \quad (6)$$

Après, nous montrons que $h(1/2) = \gamma^{-2} \{\gamma \log 2 - (1+\gamma)(1-2^{-\gamma})\} < 0$ pour tout $\gamma > -1$, ainsi nous pouvons éliminer la condition alternative pour laquelle $y_1 = 1/2$ dans (6).

Posons :

$$s(\gamma) = \gamma \log 2 - (1+\gamma)(1-2^{-\gamma}).$$

La fonction $s(\gamma)$ est présentée dans la figure 1. Alors :

$$s'(\gamma) = \log 2 - (1 - 2^{-\gamma}) - (1 + \gamma) 2^{-\gamma} \log 2,$$

et :

$$s''(\gamma) = 2^{-\gamma} (\log 2)^2 \left\{ \gamma + 1 - \frac{2}{\log 2} \right\}.$$

Lorsque :

$$\left\{ \begin{array}{l} s' \left(\frac{2}{\log 2} - 1 \right) = \log 2 - 1 - 2^{1-2/\log 2} < 0, \\ s'(-1) = 1 + \log 2 > 0, \\ s'(\infty) = \log 2 - 1 < 0, \\ s''(\gamma) < 0 \text{ quand } -1 < \gamma < \frac{2}{\log 2} - 1, \\ \text{et } s''(\gamma) > 0 \text{ quand } \gamma > \frac{2}{\log 2} - 1, \end{array} \right.$$

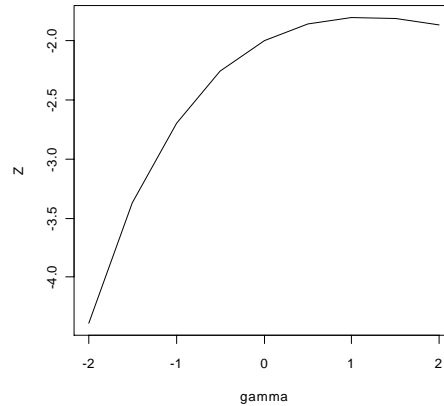
alors il existe γ_0 unique où $\gamma_0 \in \left(-1, \frac{2}{\log 2} - 1 \right)$ telle que $s'(\gamma_0) = 0$, $s'(\gamma) > 0$ si $\gamma \in (-1, \gamma_0)$ et $s'(\gamma) < 0$ si $\gamma \in (\gamma_0, \infty)$.

Puisque $s'(0) = 0$, nous avons $\gamma_0 = 0$. $s(\gamma)$ est une fonction croissante sur $(-1, 0)$ et décroissante sur $(0, \infty)$, remarquons que $s(-1) < 0$, $s(0) = 0$ et $s(\infty) < 0$, nous avons $s(\gamma) < 0$ pour tout $\gamma \in (-1, 0) \cup (0, \infty)$.

Par conséquent nous concluons que $h(1/2) < 0$ pour tout $\gamma > 1$ par notation on a $h(0) = -\log 2/2 < 0$. (6) est équivalente à :

$$\begin{aligned} y_1 &= 1/2 \text{ si } \gamma \leq -1 \text{ et} \\ -\frac{\log y_1}{\gamma} - \frac{1+\gamma}{\gamma^2} (1 - y_1^\gamma) &= -\frac{\log(y_1 + 1/2)}{\gamma} - \frac{1+\gamma}{\gamma^2} (1 - (y_1 + 1/2)^\gamma) \text{ si non(7)} \end{aligned}$$

Figure 1



Donc, l'EMMed $(\hat{\gamma}, \hat{\sigma})$ est la solution de les équations (2), (3), (5) et (7) où $y_1 = y_1(\gamma)$ dans (5) et (7) satisfaite $y_1 \in (0, 1/2]$.

2.2-Propriétés théoriques : Posons $\theta = (\gamma, \sigma)$ et

$$\Psi_{\theta}(x) = \left(\frac{\partial}{\partial \gamma} f(x), \frac{\partial}{\partial \sigma} f(x) \right)^T.$$

Il est facile de montrer que notre estimateur peut être écrit sur la forme d'un M-estimateur (voir Hampel et al. (1986)) comme suit :

$$\sum_{i=1}^n \text{sgn}(\Psi_{\theta}(X_i) - m_{\theta}) = 0, \quad (8)$$

où $\text{sgn}(x)$ est la fonction signe de x qui est appliqué à la composante sage et m_{θ} déterminé par :

$$E_{\theta} \text{sgn}(\Psi_{\theta}(X_1) - m_{\theta}) = 0,$$

pour assurer la consistance de Fisher. Ceci montre que l'EMMed est de même forme que le M-estimateur instandardisé, B-optimal avec la fonction de Huber Ψ remplacé par la fonction *sign* limitée.

Nous pouvons récrire (8) sur la forme suit :

$$\sum_{i=1}^n W_{\theta}(X_i) (\Psi_{\theta}(X_i) - m_{\theta}) = 0,$$

où $W_{\theta}(X_i) = \text{diag}(\text{sgn}(\Psi_{\theta}(X_i) - m_{\theta}) / (\Psi_{\theta}(X_i) - m_{\theta}))$ avec la convention que :

$$0/0 = 1.$$

Il est facile de calculer $W_{\hat{\theta}}(X_i)$ pour l'EMMed $\hat{\theta}$ et d'explorer quelles observations dans l'échantillon sont étés des poids plus basses. Nous pouvons explorer les poids de les composantes d'estimation séparée et/ou les mesures combinées de ces poids tels que la norme euclidienne de $W_{\hat{\theta}}(X_i)$ avec le déterminant $|W_{\hat{\theta}}(X_i)|$.

D'après Depuis (1999), nous choisissons le seuil pour que tout les poids de la robustesse soient proches de 1. Dans ce cas, l'estimateur serait faisable avec les données.

On remarquons que le choix de seuil est pour qu'il n'aura pas des points aberrants présents. En général, quand nous avons des points aberrants présents, le besoin d'étant faisable avec les données et la robustesse peut-être incompatible. Spécifiquement, avec les données contaminées, si un point aberrant peut-être hors de la gamme faisable pour la distribution de noyau, nous rendrions une estimation robuste faisable avec cette distribution plutôt avec les données observées.

Suivant Hampel et al. (1998), nous pouvons mesurer la sensibilité d'un estimateur en examinant sa fonction d'influence, la fonction d'influence de notre estimateur de la méthode de médiane est :

$$IF(x) = -V^{-1} \text{sgn}(\Psi_{\theta}(x) - m_{\theta}),$$

où :

$$V = \frac{\partial E_{\theta} \text{sgn}(\Psi_t(X_1) - m_{\theta})}{\partial t} \Big|_{t=\theta} := (\nu_{ij})$$

est une matrice (2×2) et $t = (\gamma_1, \sigma_1)$. On prouve que V est non-singulier, la fonction d'influence de l'EMMed est bornée.

Après on donnera une expression explicite pour ν_{ij} . Avec $y_1(\gamma)$ était déterminé par l'équation (7), alors on trouve que :

$$\left\{ \begin{array}{l} \nu_{11} = -\frac{2y_1(\gamma) \log y_1(\gamma)}{\gamma^2} + \frac{2(y_1(\gamma) + 1/2) \log(y_1(\gamma) + 1/2)}{\gamma^2} + \frac{\gamma}{\gamma^2}, \\ \nu_{12} = \frac{2}{\gamma\sigma} \left\{ (y_1(\gamma) + 1/2)^{\gamma+1} - y_1(\gamma)^{\gamma+1} \right\} - \frac{1}{\gamma\sigma}, \\ \nu_{21} = \frac{2^{-\gamma} + \log 2 - 1}{\gamma^2} \operatorname{sgn}(\gamma + 1), \\ \nu_{22} = \frac{2^{-\gamma} - 1}{\gamma\sigma} \operatorname{sgn}(\gamma + 1), \end{array} \right.$$

à condition que $\gamma \neq -1$, pour une dérivation détaillé de ν_{11} , voir Peng et Walsh (2001).

En fin, sous les conditions de M-estimateur normale (voir, par exemple, Huber (1981), p.130-133), nous avons que :

$$\sqrt{n} \left\{ (\hat{\gamma}, \hat{\sigma})^T - (\gamma(F), \sigma(F))^T \right\} \xrightarrow{d} N(0, \Sigma(F)),$$

où :

$$\theta(F) = (\gamma(F), \sigma(F))^T,$$

satisfait :

$$E_F \left\{ \operatorname{sgn}(\Psi_{\theta(F)}(X_1) - m_{\theta(F)}) \right\} = 0,$$

et :

$$\Sigma(F) = V(F)^{-1} E_F \left\{ \operatorname{sgn}(\Psi_{\theta(F)}(X_1) - m_{\theta(F)}) \operatorname{sgn}(\Psi_{\theta(F)}(X_1) - m_{\theta(F)})^T \right\} V(F)^{-1}$$

avec :

$$V(F) = \left. \frac{\partial E_F \operatorname{sgn}(\Psi_t(X_1) - m_t)}{\partial t} \right|_{t=\theta(F)}.$$

En général pour estimer $V(F)$, on utilise la formule suivante :

$$\frac{1}{2hn} \sum_{i=1}^n \left\{ \operatorname{sgn}(\Psi_{\hat{\theta}+h}(X_i) - m_{\hat{\theta}+h}) - \operatorname{sgn}(\Psi_{\hat{\theta}-h}(X_i) - m_{\hat{\theta}-h}) \right\},$$

où $h > 0$ est un paramètre lissant, qui doit être choisi approximativement. Une approche alternative est d'exploiter la robustesse de l'estimateur et d'utiliser la variance asymptotique sous le modèle de Pareto généralisé encore quand la distribution est actuelle est seulement de voisinage de se modèle. Lorsque le modèle de Pareto généralisé est tenu, nous avons que :

$$\sqrt{n} \left\{ (\hat{\gamma}, \hat{\sigma})^T - (\gamma, \sigma)^T \right\} \xrightarrow{d} N(0, \Sigma),$$

où $\Sigma = V^{-1}BV^{-T}$, avec $B = (b_{ij})$, $b_{11} = b_{22} = 1$ et

$$\begin{aligned} b_{12} &= b_{21} \\ &= Esgn \left[\left\{ \frac{(1+\gamma)X_1}{1+\gamma X_1/\sigma} - \frac{\sigma(1+\gamma)}{\gamma} (1-2^{-\gamma}) \right\} \times \right. \\ &\quad \left. \left\{ \frac{\log(1+\gamma X_1/\sigma)}{\gamma^2} - \frac{(1+\gamma)X_1}{\sigma\gamma + \gamma^2 X_1} - z(\gamma) \right\} \right]. \end{aligned}$$

Il est facile d'estimer Σ et alors de l'utiliser pour construire d'inférence asymptotique.

4.2.2 Simulation

Dans cette section on va étudier l'exécution de l'EMV, l'ERBO et de l'EMMed quand la distribution de Pareto généralisée fondamentale est contaminée avec une distribution de Pareto généralisée différente. Pour l'ERBO, on utilise le procédé itératif décrit par Depuis (1999) avec $c = 4$, $\eta = 0.001$ et l'EMV comme un estimateur initial. On a produit cinq cents échantillons de taille $n = 100$ où la distribution contaminée est $0.90F + 0.1F$.

On calcul les trois estimateurs pour chaque échantillon, et on rapporte les biais et les erreurs standards pour l'un de ces estimateurs dans tout les cas.

Dans le tableau 5, on donne les résultats de simulation pour l'EMV, l'ERBO et l'EMMed. Ces résultats présentés sont la racine des erreurs carrées moyennes de plus de 500 répliques. Où les paramètres $(\gamma_0, \sigma_0, \gamma_1, \sigma_1)$ viennent de la distribution contaminée, (γ_0, σ_0) de la distribution de Pareto généralisée fondamentale qui est contaminée par une autre distribution de Pareto généralisée avec des paramètres (γ_1, σ_1) .

Ces distributions sont choisies pour représenter trois distributions de noyau, du cas régulier ($\gamma_0 = 1$), un cas près de la limite de régularité ($\gamma = -1/3$) et un cas non régulier ($\gamma_0 = -2$). Chacune de ces trois distributions est présentée sous la forme propre, est contaminée par une distribution avec un paramètre de forme différent, $\gamma_1 \neq \gamma_0$ et contaminée par une distribution avec un paramètre d'échelle différent, $\sigma_1 \neq \sigma_0$. Dans tous les cas, la contamination est choisie pour être assez douce.

Nous présentons les racines des erreurs carrées moyennes des estimateurs dans le tableau 5. Considérons d'abord le cas que la distribution de noyau est régulier ($\gamma_0 = 1$), alors s'il n'y a aucune contamination, comme prévu, pour les deux paramètres, l'EMV est le meilleur estimateur suivi par l'ERBO et puis l'EMMed.

La contamination par une distribution avec un paramètre de forme différent, a peu d'effet sur les estimateurs. L'effet de la contamination par une distribution d'un seul paramètre d'échelle différent, sur les estimateurs de γ est semblable. Mais pour les estimateurs de σ , on a l'ERBO exécute mieux que l'EMV et l'EMMed.

Quand $\gamma_0 = -1/3$, il y'a deux cas : sans contamination, l'EMV est le plus efficace pour estimer l'un ou l'autre paramètre, mais la méthode de médiane pour l'estimation de γ , exécute mieux que l'ERBO. Avec contamination, l'EMV exécute toujours mieux pour tous les deux paramètres, mais la méthode de médiane est plus efficace que l'ERBO pour γ . cependant l'ERBO est plus efficace que l'EMMed de σ . Quand $\gamma_0 = -2$, l'EMMed est le plus efficace et l'EMV est le moins efficace, soit avec ou sans contamination.

LES TABLEAUX

Tableau 1 : Biais et ErCMs des estimateurs pour 10.000 échantillons aléatoires ont été tirés de la loi normale $N(0, 1)$ avec la taille de l'échantillon $n = 50, 100$.

Estimateur	n=50		n=100	
	Biais	ErCM	Biais	ErCM
EMV (Moyenne)	-0.0001	0.0196	-0.0002	0.0099
Médiane	-0.0004	0.0300	-0.0004	0.0155
Chan-He	-0.0004	0.0198	-0.0001	0.0100
Hodges-Lehmann	0.0001	0.0206	0.0000	0.0105
Score Normale	-0.0001	0.0198	-0.0003	0.0100
Huber : $k = 1.2$	-0.0002	0.0212	0.0000	0.0107
Huber : $k = 2.0$	0.0000	0.0199	-0.0001	0.0101
FIQ : $r = 1.0$	-0.0003	0.0219	-0.0005	0.0105
FIQ : $r = 0.5$	-0.0007	0.0222	-0.0002	0.0106

Tableau 2 : Biais et ErCMs des estimateurs pour 10.000 échantillons aléatoires ont été tirés de la loi normale $0.9N(0, 1) + 0.1\Delta_c$ avec la taille de l'échantillon $n = 50, 100$.

Estimateur	n=50		n=100	
	Biais	ErCM	Biais	ErCM
EMV (Moyenne)	0.3712	0.1557	0.3720	0.1473
Médiane	0.1381	0.0530	0.1400	0.0370
Chan-He	0.2731	0.1023	0.2721	0.0886
Hodges-Lehmann	0.2070	0.0665	0.2096	0.0558
Score Normale	0.2738	0.0981	0.2798	0.0898
Huber : $k = 1.2$	0.1789	0.0558	0.1823	0.0452
Huber : $k = 2.0$	0.2507	0.0866	0.2562	0.0776
FIQ : $r = 1.0$	0.0565	0.0345	0.0544	0.0185
FIQ : $r = 0.5$	0.1173	0.0535	0.1132	0.0325

Tableau 3 : l'ErCM pour $n = 20$. erreurs standards dans les parenthèses

Cases	Med		EMV	
	λ_0	β_0	λ_0	β_0
a	2.242 (0.078)	2.695 (0.110)	1.396 (0.046)	0.960 (0.027)
b	2.136 (0.091)	2.296 (0.096)	1.345 (0.043)	0.812 (0.022)
c	1.969 (0.053)	2.166 (0.082)	1.703 (0.029)	0.938 (0.021)
d	3.142 (0.066)	2.143 (0.066)	14.210 (0.175)	5.257 (0.035)
e	4.073 (0.223)	2.093 (0.053)	220.507 (2.129)	13.020 (0.063)
f	3.309 (0.165)	1.957 (0.069)	3.070 (0.145)	1.779 (0.030)
g	2.253 (0.056)	1.817 (0.065)	5.149 (0.117)	3.512 (0.049)
h	2.193 (0.056)	1.877 (0.066)	53.936 (1.492)	8.901 (0.091)
i	2.210 (0.050)	1.992 (0.068)	2.697 (0.036)	1.430 (0.017)
j	2.965 (0.063)	2.129 (0.075)	27.996 (0.306)	8.067 (0.051)
k	3.561 (0.089)	2.048 (0.051)	360.684 (3.542)	13.946 (0.076)

les valeurs de paramètres $(\lambda_0, \beta_0, \lambda_1, \beta_1)$ pour les cas a-h sont : a : (1, 1, 1, 1), b : (1, 0.7, 0.2, 0.7), c : (1, 1, 0.2, 1), d : (1, 3, 0.2, 3), e : (1, 10, 0.2, 10), f : (1, 0.7, 1, 0.2), g : (1, 3, 1, 0.5), h : (1, 10, 1, 0.5).

Les cas i-k sont contaminés par $U(0, 20)$ avec les paramètres fondamentaux : i : (1, 1), j : (1, 3), k : (1, 10).

Tableau 4 : l'ErCM pour estimateurs $n = 100$. erreurs standards dans les parenthèses

Cases	Med		EMV	
	λ_0	β_0	λ_0	β_0
a	1.859 (0.043)	1.611 (0.036)	1.152 (0.026)	0.666 (0.014)
b	2.164 (0.036)	1.542 (0.032)	2.321 (0.032)	1.146 (0.018)
c	3.006 (0.046)	1.602 (0.032)	4.695 (0.055)	2.877 (0.032)
d	6.235 (0.097)	2.814 (0.045)	66.910 (0.424)	31.561 (0.062)
e	6.781 (0.110)	2.931 (0.046)	1078.642 (4.905)	73.184 (0.023)
f	2.252 (0.055)	1.770 (0.030)	2.601 (0.051)	10.664 (0.110)
g	2.135 (0.043)	1.979 (0.033)	9.307 (0.180)	27.073 (0.192)
h	2.241 (0.045)	2.307 (0.037)	201.318 (3.164)	67.288 (0.198)
i	4.335 (0.062)	2.206 (0.036)	10.909 (0.088)	7.251 (0.040)
j	5.643 (0.088)	2.677 (0.042)	141.203 (0.766)	48.809 (0.057)
k	6.294 (0.102)	2.864 (0.045)	1882.879 (8.481)	81.693 (0.029)

Tableau 5 : Résultats du simulation des trois estimateurs

(γ_0, σ_0)	(γ_1, σ_1)	EMV	ERBO	Méthode des médianes
(1, 1)	(1, 1)	$\hat{\gamma} = 0.201$	$\hat{\gamma} = 0.216$	$\hat{\gamma} = 0.421$
		$\hat{\sigma} = 0.212$	$\hat{\sigma} = 0.213$	$\hat{\sigma} = 0.275$
	(2, 1)	$\hat{\gamma} = 0.250$	$\hat{\gamma} = 0.260$	$\hat{\gamma} = 0.425$
		$\hat{\sigma} = 0.215$	$\hat{\sigma} = 0.218$	$\hat{\sigma} = 0.274$
	(1, 2)	$\hat{\gamma} = 0.203$	$\hat{\gamma} = 0.218$	$\hat{\gamma} = 0.426$
		$\hat{\sigma} = 0.248$	$\hat{\sigma} = 0.240$	$\hat{\sigma} = 0.298$
$\left(-\frac{1}{3}, 1\right)$	$\left(-\frac{1}{3}, 1\right)$	$\hat{\gamma} = 0.102$	$\hat{\gamma} = 0.218$	$\hat{\gamma} = 0.185$
		$\hat{\sigma} = 0.142$	$\hat{\sigma} = 0.160$	$\hat{\sigma} = 0.174$
	$\left(-\frac{2}{3}, 1\right)$	$\hat{\gamma} = 0.098$	$\hat{\gamma} = 0.208$	$\hat{\gamma} = 0.170$
	$\left(-\frac{1}{3}, 2\right)$	$\hat{\sigma} = 0.132$	$\hat{\sigma} = 0.154$	$\hat{\sigma} = 0.169$
		$\hat{\gamma} = 0.177$	$\hat{\gamma} = 0.243$	$\hat{\gamma} = 0.203$
	$\left(-\frac{1}{3}, 2\right)$	$\hat{\sigma} = 0.132$	$\hat{\sigma} = 0.134$	$\hat{\sigma} = 0.170$
(-2, 1)	(-2, 1)	$\hat{\gamma} = 1.153$	$\hat{\gamma} = 0.664$	$\hat{\gamma} = 0.296$
		$\hat{\sigma} = 0.576$	$\hat{\sigma} = 0.333$	$\hat{\sigma} = 0.148$
	(-4, 1)	$\hat{\gamma} = 1.061$	$\hat{\gamma} = 0.702$	$\hat{\gamma} = 0.433$
		$\hat{\sigma} = 0.531$	$\hat{\sigma} = 0.351$	$\hat{\sigma} = 0.217$
	(-2, 2)	$\hat{\gamma} = 1.485$	$\hat{\gamma} = 1.535$	$\hat{\gamma} = 1.278$
		$\hat{\sigma} = 0.469$	$\hat{\sigma} = 0.467$	$\hat{\sigma} = 0.295$

BIBLIOGRAPHIE

- [1] Adatia, A. and Chan L.K. : Robust procedures for estimating the scale parameter and predicting future order statistics of the Weibull distribution. *IEEE Transactions on Reliability*, **31**, 491-498, (1982).
- [2] Arnold, B. C. : Pareto distributions. International Cooperative Publishing House. Fairland, Maryland, (1983).
- [3] Brazauskas, V. and Serfting, R. : Robust and efficient estimation of the tail index of a single-parameter Pareto distribution. *North American Actuarial Journal*, **4**, 12-27, (2000a).
- [4] Brazauskas, V. and Serfling, R. : Robust estimation of tail parameters for two-parameter Pareto and exponential models via generalized quantile statistics. *Extremes*, **3**, 231-249, (2000b).
- [5] Brazauskas, V. and Serfling, R. : Small sample performance of robust estimators of tail parameters for Pareto and exponential models. *Journal of Statistical Computation and Simulation*, **70**, 1-19, (2001).
- [6] Brazauskas, V. and Serfling, R. : Favorable estimators for fitting Pareto model : A study using goodness-of-fit measures with actual data. *ASTIN Bulletin*, (2004).
- [7] Bunday, B., and Al-Mutwali, I. : Direct optimization for calculating maximum likelihood estimates of parameters of the Weibull distribution. *IEEE Transactions on Reliability*, **30**, 367-369, (1981).
- [8] Chamberlain, G. : Asymptotic efficiency in estimation with conditional moment restrictions, *Journal of Econometrics*, **34**, 305-334 (1987).
- [9] Chanseok P. and Bruce G. Lindsay. : Robust Estimation and Tests Based on Quadratic Inference Function. Center for Likelihood Studies, University Park, (1999).
- [10] D'Agostino, R. B. and Stephens, M. A. : Goodness-of-Fit Techniques. Marcel Dekker, New York, (1986).
- [11] Dupuis, D. J. : Estimating the probability of obtaining nonfeasible parameter estimates of the generalized Pareto distribution. *Journal of Statistical Computation and Simulation*, **54**, 197-209, (1996).
- [12] Dupuis, D. j. and Field, C.A. : Robust estimation of extremes. *The Canadian Journal of Statistics*, **26**, 199-215, (1998).

- [13] Durbin, J. : Kolmogorov-Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests on spacings. *Biometrika*, **62**, 5-22, (1975).
- [14] Embrechts, P., Klüppelberg, C., and Mikosch, T. : *Modelling extremal events for insurance and Finance*, Springer : Berlin, (1997).
- [15] Ferguson, T. S. : A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities. *Annals of Mathematical Statistics*, **29**, 1046-1062, (1958).
- [16] Hampel, F. R. R., E., Rousseeuw, P.J., and Stahel, W. : *Robust Statistics : The approach based on influence functions*, John Wiley & Sons : New York, (1986).
- [17] He, X. and Fung, W.K. : Method of medians for life time data with Weibull models. *Statistics in Medicine*, **18**, 1993-2009, (1999).
- [18] He, X. and Wang, G. : Cross-checking using the minimum volume ellipsoid estimator. *Statistica Sinica*, **6**, 367-374, (1996).
- [19] Hogg, R. V. and Klugman, S. A. : *Loss Distributions*. Wiley, New York, (1984).
- [20] Huber, P. : The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and probability*, **1**, 221-233, (1967).
- [21] Kimber, A. C. : Trimming in gamma samples. *Applied Statistics*, **32**, 7-14, (1983a).
- [22] Kimber, A. C. : Comparison of some robust estimators of scale in gamma samples with known shape. *Journal of Statistical Computation and Simulation*, **18**, 273-286, (1983b).
- [23] Klugman, S.A., Panjer, H.H. and Willmot, G.E. : *Loss Models : From Data to Decisions*. Wiley, New York, (1998).
- [24] Koutrouvelis, I. A. : Large-sample quantile estimation in Pareto laws. *Communications in Statistics, Part A—Theory and Methods*, **10**, 189-201, (1981).
- [25] Lee, M. J. : *Methods of Moments and Semiparametric Econometrics for Limited Dependent Variable Models*. Springer-Verlag, New York, (1996).
- [26] Peng, L. and Welch, A.H. : *Robust Estimation of The generalized Pareto distribution*. Centre for Mathematics, and its Applications, Australian National University. Australia, (2001).

- [27] Lin, P. E., Wu, K.T. and Ahmad, I. A. : Asymptotic joint distribution of sample quantiles and sample mean with application. *Communication in Statistics : Theory and Methods*, 9, 51-60, (1980).
- [28] Neyman, J. : Contribution to the Theory of the χ^2 test. In *Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, (1949).
- [29] Partrik, G. : Estimating casualty insurance loss amount distribution. *Proceedings of the Casualty Actuarial Society*, **LXVII**, 57-109, (1980).
- [30] Philbrick, S. W. : A practical guide to the single parameter Pareto distribution. *Proceedings of the Casualty Actuarial Society*, **LXXII**, 44-84, (1985).
- [31] Philbrick, S.W. and Jurschak, J. : Discussion of "Estimating casualty insurance loss amount distributions." *Proceedings of the Casualty Actuarial Society*, **LXVIII**, 101-106, (1981).
- [32] TASSI, Ph. : *Méthode statistique*, Ecole nationale de la statistique et de l'administration économique, (2^e édition), (1989).
- [33] TASSI, Ph. et LECOUTRE, J.P. : *Statistique Non Paramétrique et Robustesse*, Ecole nationale de la statistique et de l'administrational économique, (1987).
- [34] Qu, P. : *Adaptive Generalized Estimating Equations*. Ph.D. thesis, Department of Statistics, The Pennsylvania State Univesity, University Park, PA 16802, (1998).
- [35] Saporta, G. : *Théories et Méthode de La statistique*, Bublication de l'institut Français de pétrole, (1978).
- [36] Serfing , R.J. : *Aproximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York, (1980).
- [37] Shier, D.R., and Lawrence, K.D. : A comparison of robust regression techniques for the estimation of Weibull parameters. *Communications in Statistics, B-Simulation and Computation*, 13, 743-750, (1984).