UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ESTIMATEURS À NOYAU ET THÉORIE DES VALEURS EXTRÊMES : COMPARAISON DE LEUR POUVOIR PRÉDICTIF DANS L'ANALYSE DU COÛT DES RÉCLAMATIONS EN ASSURANCE AUTOMOBILE

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN MATHÉMATIQUES

CONCENTRATION STATISTIQUE

PAR

ETIENNE DOUCET

JANVIER 2014

REMERCIEMENTS

Je tiens à remercier mon directeur Jean-Philippe Boucher sans qui ce mémoire n'aurait jamais vu le jour. Son aide fut indispensable à tous les niveaux. Merci à l'Université du Québec à Montréal avec qui il y a un peu plus de cinq ans je commençais mon aventure universitaire. Un merci spécial aux professeurs qui ont su m'inspirer tels Fabrice Larribe et Jacques Labelle. Je dois aussi remercier l'aide financière aux études qui m'a permis de me concentrer sur mes études.

Finalement, merci à toute ma famille, Alain, Renelle et François qui n'ont jamais cessé de m'encourager.

TABLE DES MATIÈRES

LISTE DES FIGURES

LISTE DES TABLEAUX

RÉSUMÉ

Ce mémoire étudie la distribution du montant des réclamations en assurance automobile à l'aide d'une base de données provenant d'un assureur de l'Ontario. Cette base de données contient différents types de réclamations : blessures corporelles, dommages aux véhicules, vol et risques divers. Les différents types de réclamations sont décrits au premier chapitre. Une attention particulière est portée à la modélisation des réclamations dites extrêmes. Une approche paramétrique est d'abord présentée et les distributions Pareto, gamma, bêta prime et Champernowne sont tour à tour étudiées. On montre que cette approche n'est pas idéale pour modéliser les coûts des réclamations pour blessures corporelles. Deux estimateurs à noyaux sont ensuite introduits. L'estimateur à noyau gamma est utilisé directement sur les données, tandis que l'estimateur à noyau bêta est utilisé pour modéliser une transformation des données. Cette transformation utilise la fonction de répartition d'une distribution Champernowne. Finalement, la théorie des valeurs extrêmes est utilisée. Une méthode de sélection automatique du seuil utilisant le produit maximum des espacements est proposée. Au dernier chapitre, on compare le pouvoir de prédiction des différents modèles à l'aide de la base de données.

Mot-clé : Estimation paramétrique, estimation non paramétrique, estimation semi-paramétrique, distribution Champernowne, distribution bêta prime, distribution Pareto généralisée, noyau gamma, noyau bêta, théorie des valeurs extrêmes, produit maximum des espacements.

CHAPITRE I

INTRODUCTION

Ce mémoire s'intéresse à la tarification en assurance de dommages et utilise le concept de décomposition de la sévérité et de la fréquence. L'étude de la fréquence vise à modéliser le nombre de réclamations d'un portefeuille de polices d'assurance alors que l'étude de la sévérité s'intéresse au montant des réclamations. Dans ce travail, la distribution du montant des réclamations en assurance automobile est étudiée. Cette étude est faite à partir d'une base de données provenant d'un assureur de l'Ontario.

Ce travail se divise comme suit. Une analyse exploratoire des données est d'abord faite au premier chapitre. Il y est montré que la base de données contient quelques réclamations particulièrement élevées par rapport à la moyenne des réclamations. Ensuite, une première modélisation paramétrique est présentée au deuxième chapitre. Les chapitres suivants s'intéressent à la modélisation des réclamations dites extrêmes. Différentes approches non paramétriques et semi-paramétriques utilisant les estimateurs à noyau sont étudiées. Entre autres, il est suggéré que les estimateurs à noyau traditionnels ne sont pas les plus appropriés pour la modélisation du montant des réclamations et que les noyaux gamma et bêta sont plus adéquats pour cette tâche. Finalement, la théorie des valeurs extrêmes sera présentée ainsi qu'une méthode de détection automatique du seuil à partir duquel il est justifié d'utiliser la loi de Pareto généralisée.

1.1 Notions de base

Une revue de quelques notions de base est faite dans cette section. Tous ces concepts sont décrits en détail et dans un contexte actuariel par Klugman, Panjer et Willmot dans (Klugman, Panjer et Willmot, 2004). Les axiomes sur les probabilités ont été pris dans (Ross, 2002).

1.1.1 Axiomes de probabilités

Pour une discussion formelle sur les probabilités, on doit nécessairement passer par la théorie de la mesure, ce qui dépasse le cadre de ce mémoire. Le lecteur intéressé peut trouver une telle discussion dans (Casella et Berger, 1990). Ici, une présentation de trois axiomes classiques de probabilités tels que décrits par Ross dans (Ross, 2002) sera faite et servira de base aux autres concepts statistiques.

Soit une expérience dont le résultat n'est pas prévisible. On appelle l'ensemble des résultats possibles, l'ensemble fondamental. Soit un évènement E, qui fait partie de l'ensemble fondamental S, c'est-à-dire $E \in S$, alors il existe un nombre $\Pr[E]$, la probabilité de l'évènement E, qui satisfait aux trois axiomes suivants :

Axiome 1. $0 \leq \Pr[E] \leq 1$

Axiome 2. $\Pr[S] = 1$

Axiome 3. *Pour chaque séquence d'événements mutuellement exclusifs* $E_1, E_2, \ldots$, *la probabilité de l'union des événements* E_i *est :*

$$\Pr\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \Pr[E_i]$$

1.1.2 Fonction de densité, fonction cumulative et espérance

Soit une variable aléatoire X, la probabilité que X prenne une valeur inférieure ou égale à x, c'est-à-dire $\Pr[X \leq x]$, est appelée fonction cumulative ou fonction de réparti-

tion. Cette fonction est, par convention, notée par une lettre majuscule. Donc pour une variable aléatoire X de fonction cumulative F, on notera $F_X(x) = \Pr[X \leq x]$.

Dans le cas d'une variable aléatoire discrète, c'est-à-dire une variable qui prend des valeurs dans un ensemble de valeurs possibles au plus dénombrable, la fonction de densité évaluée en x représente la probabilité que la variable X prenne la valeur x, donc $\Pr[X = x]$. Dans le cas d'une variable aléatoire continue, par exemple une variable prenant des valeurs dans les réels, l'interprétation de la fonction de densité n'est pas aussi simple. En effet la fonction de densité d'une variable aléatoire continue X, notée $f_X(x)$, doit être multipliée par une distance infiniment petite, dx, pour obtenir une probabilité. La fonction de densité est, par convention, notée par une lettre minuscule. Remarquez que dans le cas d'une variable discrète il est plus courant de noter $f_X(x)$ par $P_X[x]$.

Quelques relations :

$$
\begin{aligned}
f_X(x) &= \frac{dF_X(x)}{dx}, \\
F_X(x) &= \int_{-\infty}^{x} f_X(x)dx, \\
E[X] &= \int_{-\infty}^{\infty} x f_X(x)dx = \int_{-\infty}^{\infty} x dF_X(x),
\end{aligned}
$$

où $E[X]$ représente l'espérance de X.

La variance d'une variable aléatoire, notée $Var[X]$, est définie comme suit :

$$
\begin{aligned}
Var[X] &= E[(X - E[X])^2] \\
&= \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x)dx \\
&= E[X^2] - E[X]^2.
\end{aligned}
$$

L'écart-type, soit la racine carrée de la variance, est une mesure de la variabilité d'une variable aléatoire et peut être vu comme une mesure de risque. En finance, des rendements avec un plus grand écart-type sont souvent vus comme plus risqués. Voir par

exemple la théorie de Markovitz dans (Markowitz, 1970), ou le modèle d'évaluation des actifs financiers ($CAPM$) de Treynor, dans (Treynor, 1961).

Dans l'approche paramétrique, on suppose une forme pour $F_X(x)$ ou pour $f_X(x)$. Par exemple, si on suppose que X suit une loi normale on aura alors

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, x \in \mathbb{R}.$$

Cette fonction de densité est entièrement définie par deux paramètres, μ et $\sigma > 0$, qui correspondent à la moyenne et à l'écart-type respectivement.

1.1.3 Somme de variables aléatoires

Soit deux variables aléatoires X et Y. On note la densité conjointe de X et Y par $f_{X,Y}(x,y)$, i.e. la probabilité que $X = x$ et $Y = y$ simultanément. Lorsque X et Y sont indépendantes, on a $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

On peut retrouver les fonctions de densité de X et Y à partir de leur densité conjointe par :

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \tag{1.1}$$

et

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx. \tag{1.2}$$

Il arrive souvent que l'on s'intéresse à l'espérance de la somme de variables aléatoires. Soit deux variables aléatoires X et Y. L'espérance de la somme $X + Y$ est

$$\begin{aligned} E[X+Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y) f_{X,Y}(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dxdy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x,y)dxdy. \end{aligned}$$

En appliquant les équations (1.1) et (1.2) on a

$$\begin{aligned} E[X+Y] &= \int_{-\infty}^{\infty} x f_X(x)dx + \int_{-\infty}^{\infty} y f_Y(y)dy \\ &= E[X] + E[Y]. \end{aligned}$$

Donc, l'espérance de la somme de variables aléatoires est égale à la somme des espérances. Notez qu'aucune hypothèse d'indépendance entre les variables n'est nécessaire.

1.1.4 Statistiques d'ordre

Soit une variable aléatoire X et un échantillon de réalisations $\{x_i\}$ avec $i \in \{1, 2 \ldots, n\}$. On note $x_{(k)}$ la statistique d'ordre k, on a $x_{(1)} \leq x_{(1)} \leq \ldots \leq x_{(n)}$. On peut aussi parler de la variable aléatoire $X_{(k)}$, soit la distribution de la $k^{\text{ème}}$ plus grande observation d'un échantillon de taille n.

1.1.5 Quantiles et valeurs à risque

En mathématiques statistiques, on parle de quantiles alors que dans le monde de la gestion du risque (finance ou actuariat par exemple) on parle plutôt de valeur à risque. Ces deux termes représentent une seule et même chose. Pour une discussion approfondie sur la valeur à risque et d'autres mesures de risques voir (Denuit et al., 2005) et (Klugman, Panjer et Willmot, 2004).

Pour une variable aléatoire continue X, le quantile à 95%, noté $q_{0.95}$, et la valeur à risque à 95%, notée $VaR_{0.95}$ [1], représentent la valeur x pour laquelle la probabilité que X soit inférieur ou égal à x est de 95%. De façon plus générale on a :

$$q_p = VaR_p = F_X^{-1}(p).$$

Évidemment, on a donc aussi $F_X(q_p) = p$.

Comme son nom l'indique, la valeur à risque est une mesure de risque. Cette mesure a l'avantage de seulement tenir compte du risque de perte inattendue, contrairement à l'écart-type qui considère aussi le risque de gain inattendu, qui dans un contexte financier n'est pas réellement un risque. Par contre, la valeur à risque n'est pas parfaite non plus comme mesure de risque. Pour une critique de la valeur à risque voir (Dowd et Blake, 2006).

---

1. Attention, ne pas confondre la valeur à risque (VaR) et la variance (Var).

1.1.6 Estimation par maximum de vraisemblance

L'estimation par maximum de vraisemblance est une méthode paramétrique. Le statisticien doit donc décider d'une loi ou d'une famille de lois paramétriques avant d'utiliser la méthode du maximum de vraisemblance pour estimer les paramètres de la loi choisie.

Dans certains cas, le choix de la loi peut être facile. Par exemple, le nombre de "piles" obtenues en lançant une pièce de monnaie (pièce qui pourrait être biaisée) suivra une loi binomiale. Par contre, en général, il n'est pas évident de savoir *a priori* quelle loi suit une variable aléatoire. Une fois que le statisticien a choisi une forme pour la distribution de la variable aléatoire X, il doit en estimer les paramètres.

Plusieurs méthodes existent pour estimer les paramètres d'une loi paramétrique. Il y a l'approche bayésienne, où l'on suppose que les paramètres de la loi sont eux-mêmes des variables aléatoires. Il y a la méthode des moments, où l'on choisit les paramètres de façon à ce que les moments empiriques coïncident avec les moments théoriques. Il y a aussi l'approche par maximum de vraisemblance. Évidemment, cette liste est loin d'être exhaustive. Il existe beaucoup d'autres méthodes pour estimer les paramètres d'une distribution paramétrique (les termes loi et distribution sont utilisés de façon interchangeable).

Avant de décrire les estimateurs par maximum de vraisemblance, il faut définir la fonction de vraisemblance. Soit un échantillon $\boldsymbol{x} = \{x_1, \ldots, x_n\}$, où les x_i sont des réalisations indépendantes d'une variable aléatoire X ayant comme distribution la fonction $f_X(x)$ de paramètre $\boldsymbol{\theta}$ ($\boldsymbol{\theta}$ peut être un vecteur de paramètre) alors la fonction de vraisemblance pour cet échantillon est donnée par :

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = \prod_{i=1}^{n} f_X(x_i|\boldsymbol{\theta}).$$

Cette fonction peut s'interpréter comme la probabilité d'obtenir l'échantillon $\boldsymbol{x}$ si X avait comme distribution la fonction $f_X(x)$ de paramètres $\boldsymbol{\theta}$. Rappelons que cette interprétation est rigoureusement vraie seulement dans le cas d'une variable aléatoire discrète.

L'idée de l'estimation par maximum de vraisemblance est de trouver le vecteur de paramètre $\hat{\boldsymbol{\theta}}$ qui maximise la fonction de vraisemblance, c'est-à-dire $\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\text{argmax}}[L(\boldsymbol{\theta}|\boldsymbol{x})]$.

En général, on travaille plutôt avec la fonction de logvraisemblance :

$$l(\boldsymbol{\theta}|\boldsymbol{x}) = \ln(L(\boldsymbol{\theta}|\boldsymbol{x}) = \sum_{i=1}^{n} \ln(f_X(x_i|\boldsymbol{\theta})),$$

cette fonction est généralement plus facile à maximiser.

L'estimateur par maximum de vraisemblance a les propriétés suivantes, voir (Wasserman, 2003) :

1. Il est convergent, c'est-à-dire que $\hat{\boldsymbol{\theta}}$ converge vers la vraie valeur de $\boldsymbol{\theta}$ lorsque la taille de l'échantillon tend vers l'infini.
2. Il est asymptotiquement distribué selon une loi normale de moyenne $\boldsymbol{\theta}$, c'est-à-dire la vraie valeur de $\boldsymbol{\theta}$.

1.1.7 Le critère d'information d'Akaike

Le critère d'information d'Akaike, noté AIC, offre une mesure relative de la qualité de l'ajustement d'un modèle. Il peut donc être utilisé pour comparer deux modèles concurrents entre eux, mais ne permet pas d'évaluer la qualité d'un modèle de façon absolue. Autrement dit, si tous les modèles évalués sont mauvais, le AIC indiquera le moins mauvais, mais ce critère ne pourra détecter que l'ajustement du modèle choisi laisse lui-même à désirer.

L'AIC est une mesure de la perte d'information lorsqu'un modèle est utilisé pour décrire la réalité. On préfère donc le modèle offrant le plus petit AIC. L'AIC se calcule par :

$$AIC = 2k - l(\hat{\boldsymbol{\theta}}|\boldsymbol{x}),$$

où k est le nombre de paramètres dans le modèle et $l(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{x}})$ est la fonction de logvraisemblance maximisée. Donc, pour deux modèles ayant la même vraisemblance, le modèle ayant le moins de paramètres sera sélectionné. Ce qui va dans l'esprit du rasoir d'Occam, c'est-à-dire que le modèle le plus simple est aussi le plus probable.

Pour plus de détail sur l'AIC le lecteur peut se référer à (Akaike, 1974).

1.1.8 La statistique de Kolmogorov-Smirnov

La statistique de Kolmogorov-Smirnov, généralement notée par D, est utilisée pour comparer l'ajustement de différents modèles ou choix de distributions. Soit un échantillon de données $x_1, x_2, \ldots, x_n$ alors cette statistique se définit comme :

$$D = \sup_x |F_n(x) - \hat{F}(x)|,$$

où $\hat{F}(x)$ est la fonction cumulative du modèle testé et $F_n(x)$ est la fonction cumulative empirique, c'est-à-dire :

$$F_n(x) = \frac{\text{nombre d'éléments dans l'échantillon} \leq x}{n}.$$

Une statistique D plus petite est signe d'un meilleur ajustement et elle peut être utilisée pour comparer des modèles entre eux. Pour plus de renseignements sur la statistique de Kolmogorov-Smirnov et des trucs pour en accélérer le calcul, le lecteur peut se référer à (Stephens, 1970).

1.1.9 Comparaison des queues de distributions

La fonction de densité, $f_X(x)$, pour des valeurs de x élevées, est appelée la queue de la distribution. Une distribution ayant une queue plus élevée qu'une autre distribution sera considérée comme plus risquée. Une méthode couramment utilisée pour comparer la queue de deux distributions est de passer par les fonctions de survie, voir (Klugman, Panjer et Willmot, 2004).

La fonction de survie, notée $S_X(x)$, est le complément de la fonction cumulative, c'est-à-dire $F_X(x) + S_X(x) = 1$. On a aussi que $S_X(x) = \Pr[X > x]$. Soit deux variables aléatoires X_1 et X_2 de fonction de survie S_{X_1} et S_{X_2} respectivement, la distribution de X_1 aura une queue plus lourde que la distribution de X_2 si la limite

$$\lim_{x \to \infty} \frac{S_{X_1}(x)}{S_{X_2}(x)} \tag{1.3}$$

tend vers l'infini.

Il est habituel de comparer la queue d'une distribution avec la queue d'une loi normale, i.e. S_{X_2} est une loi normale. Cette comparaison est aussi souvent faite avec S_{X_2} étant la fonction de survie d'une loi exponentielle. Dans ce cas, on parle de queue surexponentielle ou sous-exponentielle, si le ratio (1.3) tend vers l'infini ou vers zéro respectivement.

Il y a beaucoup de choses à dire sur l'analyse de la queue d'une distribution, voir par exemple (Denuit et Charpentier, 2005) ou le chapitre 5 de ce mémoire sur la théorie des valeurs extrêmes.

1.2 Les données

Les données analysées proviennent d'un assureur automobile de l'Ontario. La base de données est composée de 2 346 681 transactions pour 322 174 assurés. L'analyse qui suit s'intéresse au coût des réclamations, donc seulement les transactions qui contiennent des réclamations (au minimum une réclamation) nous intéressent. Il y a 92 457 transactions ayant au moins une réclamation, tous types de réclamations confondues.

Il y a six types de réclamations possibles dans la base de données correspondant aux six couvertures auxquelles un assuré peut souscrire. Notez que certaines couvertures sont obligatoires par la loi, comme la couverture pour blessures corporelles infligées à autrui, et que d'autres sont facultatives, comme la couverture couvrant les dommages à la voiture de l'assuré. Ces couvertures sont les suivantes[2] :

1. Blessures corporelles (BC) : couvre les coûts reliés aux blessures corporelles infligées aux autres, dont l'assuré est responsable. Cette assurance couvre les frais médicaux, perte de revenu ou perte de jouissance due à la douleur. Les blessures que l'assuré s'est infligées ou les dommages causés à son véhicule ne sont pas couverts. Cette couverture est obligatoire.

2. La description de ces différentes couvertures peut être trouvée sur le site web du Bureau d'assurance du Canada, http ://www.ibc.ca/fr/

2. Responsabilité civile (RC) : couvre les coûts reliés aux dommages faits aux biens d'autrui (généralement leur voiture) dont l'assuré est responsable.
3. Collisions (Coll) : couvre les coûts reliés aux dommages subis par la voiture de l'assuré. Cette couverture n'est pas obligatoire.
4. Blessures personnelles (BP) : couvre les coûts reliés aux blessures corporelles subies par l'assuré, dont il est responsable. Les frais médicaux sont couverts, mais la réclamation maximale est généralement plus basse que celle associée à la couverture blessures corporelles.
5. Vol : couvre les coûts reliés au vol de la voiture ou de matériel dans la voiture.
6. Tout-Risque (TR) : couvre une foule d'autres coûts qui peuvent survenir et qui ne sont pas couverts par les autres couvertures. Par exemple, la chute d'un arbre sur la voiture de l'assuré.

Avant même d'aller voir les données, on peut se faire une idée de chaque risque par la description de leurs couvertures. Par exemple, il semble évident qu'il risque d'y avoir plus de réclamations extrêmes pour le chapitre BC (les termes chapitre et couverture sont utilisés de façon interchangeable) que pour tous les autres chapitres. En effet, il n'y a pratiquement pas de limites aux coûts médicaux, et à ces coûts vient potentiellement s'ajouter une prestation pour perte de salaire si la victime devient invalide suite à l'accident. En comparaison, le chapitre collision couvre les coûts de réparation de la voiture de l'assuré, ces coûts ne peuvent donc pas dépasser le coût de la voiture, coût connu par l'assureur au moment d'établir la prime. On s'attend aussi à ce que la protection pour responsabilité civile ressemble à la protection pour collisions puisqu'elles couvrent, dans la majorité des cas, le même risque (c.-à-d. dommage aux voitures).

Le tableau 1.1 présente plusieurs statistiques descriptives pour les différentes couvertures décrites plus haut. On remarque que les plus grosses réclamations sont pour les blessures corporelles suivies de près par les blessures personnelles. Du point de vue de l'assureur, ces deux couvertures peuvent avoir un impact significatif sur les résultats de la compagnie. Il est donc important d'accorder une attention particulière à la modélisation de

Tableau 1.1 Résumé de la base de données

Couverture	Nombre	Moyenne	Écart-type	$95^{ème}$ Percentiles	Maximum
Blessures corporelles	3 194	45 938	106 442	158 783	2 521 561
Responsabilité civile	41 407	3 496	3 711	9 966	102 113
Collision	27 504	4 763	4 905	13 972	56 914
Blessures personnelles	10 674	22 750	74 986	82 681	2 260 203
Vol	15 067	1 147	2 323	5 182	41 057
Tout Risque	7 098	4 778	8 182	20 262	107 034

ces deux risques. Ce mémoire se concentre donc particulièrement à la modélisation du coût des réclamations pour les blessures personnelles. Cette couverture fut choisie, car la base de données étudiée contient trois fois plus de données pour cette couverture que pour la couverture blessures corporelles.

La figure 1.1 présente les histogrammes de la distribution des coûts pour chaque couverture.

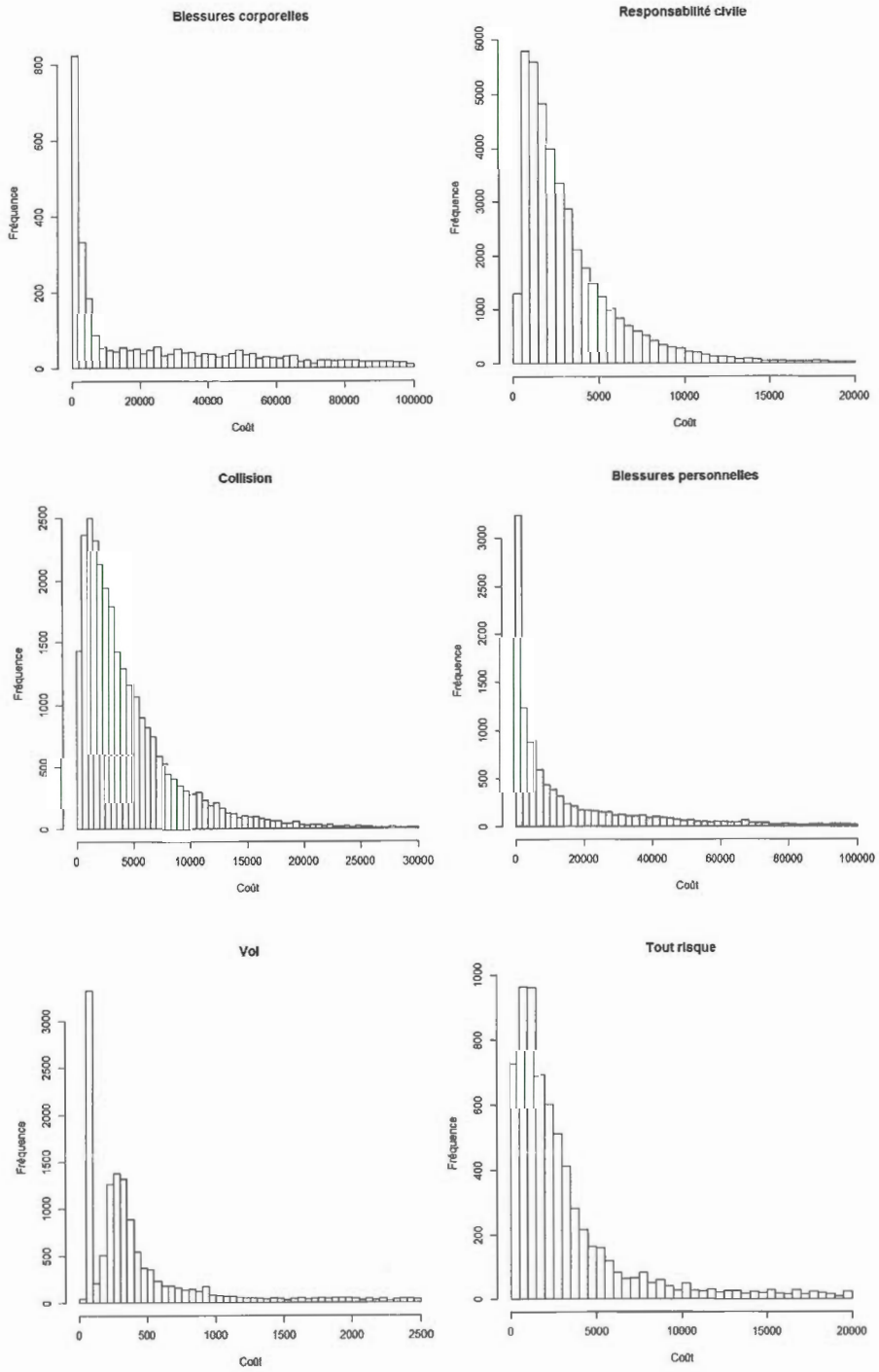

Figure 1.1 Histogrammes des données pour chaque couvertures.

CHAPITRE II

ESTIMATION DE DENSITÉ - MÉTHODES PARAMÉTRIQUES

L'approche traditionnelle de la modélisation en statistique est l'approche dite paramétrique. Cette approche est appliquée de la manière suivante :

1. Choisir une distribution ;
2. Choisir une méthode pour estimer les paramètres de la distribution choisie en 1 ;
3. (optionnel) Choisir un seuil pour séparer les données extrêmes des autres.

Ce chapitre se concentre sur les étapes 1 et 2. La dernière étape est utilisée dans le cadre de la théorie des valeurs extrêmes. Ce sujet sera traité au chapitre 5.

Dans la pratique, aucune de ces étapes n'est triviale. Beaucoup de choix s'offrent à l'actuaire, tant pour les distributions que pour les méthodes d'estimation. Pour une revue des distributions et les méthodes d'estimation de paramètres utilisées en actuariat, voir (Klugman, Panjer et Willmot, 2004).

Comme les données étudiées ici représentent des montants de réclamations, il parait logique de se limiter à des distributions ayant comme domaine $(0, \infty)$. Quatre distributions seront utilisées à tour de rôle : la loi de Pareto, la loi bêta prime, la loi gamma et la loi de Champernowne. Ces distributions ont été choisies à cause de leurs importances dans la littérature actuarielle. Le lecteur intéressé peut se référer à (Klugman, Panjer et Willmot, 2004) ou à (Kleiber et Kotz, 2003).

Ce chapitre se concentre spécialement sur la distribution du montant des réclamations pour blessures personnelles. Ces données présentent des observations très élevées. Il sera

donc intéressant de voir si une approche purement paramétrique sera suffisante pour modéliser ces données.

2.1 Distribution de Pareto

La distribution de Pareto classique, tel que définie dans (Kleiber et Kotz, 2003), est définie par sa fonction cumulative :

$$F(x) = 1 - \left(\frac{x}{x_0}\right)^{-\alpha} \text{ pour } x \geq x_0 \geq 0,$$

où $\alpha > 0$ est un paramètre de forme (plus α est petit, plus la queue de la distribution est épaisse) et x_0 est un paramètre d'échelle. La fonction de densité est donnée par :

$$f(x) = \frac{\alpha x_0^{\alpha}}{x^{\alpha+1}} \text{ pour } x \geq x_0 \geq 0.$$

2.1.1 Estimation par maximum de vraisemblance

Il y a plusieurs méthodes d'estimation des paramètres d'une distribution et la loi de Pareto ne fait pas exception. Ici, la méthode du maximum de vraisemblance sera utilisée.

La fonction de logvraisemblance à maximiser est :

$$l(x_0, \alpha|\mathbf{y}) = \sum_{i=1}^{n} \left[\log(\alpha) + \alpha \log(x_0) - (\alpha+1)\log(y_i)\right],$$

où $\mathbf{y}$ est un vecteur d'observations. Les estimateurs par maximum de vraisemblance sont donc (voir (Kleiber et Kotz, 2003)) :

$$\begin{aligned} \hat{\alpha} &= n\left[\sum_{i=1}^{n} \log\left(\frac{y_i}{\hat{x_0}}\right)\right], \\ \hat{x_0} &= y_{(1)}, \end{aligned}$$

où $y_{(1)}$ représente la plus petite donnée du vecteur $\mathbf{y}$.

2.1.2 Estimation des paramètres de la loi de Pareto pour les données d'assurance

Les estimateurs décrits plus haut ont été calculés pour les données d'assurance automobile présentées au chapitre 1. On s'intéresse plus particulièrement aux réclamations pour blessures personnelles, car elles représentent un plus grand risque pour l'assureur (voir la section 1.2).

Le tableau 2.1 présente les résultats obtenus. L'estimation fût d'abord faite sur l'ensemble des données, puis sur les réclamations supérieures à 500$, la raison en est expliquée plus bas. Notez que le premier échantillon à une taille de 10674 observations et que le deuxième à une taille de 9 492 observations.

On remarque dans le tableau 2.1 que la valeur de $\hat{\alpha}$ pour l'ensemble des données est particulièrement petite. En effet, une valeur inférieure à 1 pour le paramètre α indique une queue de distribution si lourde que l'espérance n'est pas définie. Par contre, la valeur estimée du paramètre x_0 correspondant est pour le moins surprenante. On peut se questionner à savoir s'il y a réellement eu une réclamation de 1$ ou bien s'il y a une erreur dans la base de données. Donc, une fois que toutes les réclamations inférieures à 500$ ont été supprimées de la base de données, on obtient les estimations présentées dans la partie droite du tableau 2.1.

La valeur de $\hat{\alpha}$ obtenue avec les données supérieures à 500$, bien qu'un peu plus grande, est encore très petite. Ce qui laisse croire que la distribution est dangereuse, c'est-à-dire à queue lourde. Pour une discussion sur le concept de queue de distribution et de queue lourde, veuillez vous référer au chapitre 1.

Notez que les valeurs des AIC présentées dans le tableau 2.1 ne peuvent être comparées entre elles, car le nombre de données utilisées pour l'estimation n'est pas le même. Les AIC pourront être utilisés pour comparer l'ajustement de la loi de Pareto à celui obtenu avec les autres distributions étudiées dans ce chapitre.

Tableau 2.1 Paramètres estimés pour une distribution de Pareto à l'aide des réclamations pour blessures personnelles

Paramètres	Pour l'ensemble des données	Réclamations supérieures à 500$
$\hat{\alpha}$	0.1167	0.3579
$\hat{x_0}$	1.0000	500.8000
AIC	250 100	209 544

2.1.3 Quantiles estimés

Dans la modélisation et l'estimation du risque en assurance, il est important de pouvoir estimer les quantiles de la distribution des réclamations. Ceci est d'autant plus vrai pour les risques extrêmes où une très grosse réclamation peut affecter significativement les résultats de la compagnie d'assurance.

De plus, la majorité des mesures de risque utilise les quantiles dans leurs calculs ce qui rend leurs estimations très importantes (voir l'article (Dowd et Blake, 2006)).

Dans le cadre d'une modélisation paramétrique, une fois que les paramètres de la distribution ont été estimés, l'estimation des quantiles est généralement simple et directe. Le tableau 2.2 présente quelques quantiles selon les paramètres estimés de la distribution de Pareto et les compare aux quantiles empiriques.

On voit clairement que la distribution de Pareto estimée surestime les hauts quantiles par rapport aux quantiles empiriques. Bien qu'il soit possible que le 99^e percentile empirique sous-estime le vrai 99^e percentile, il semble peu probable qu'il soit de 200$ millions. Notez par ailleurs que la réclamation maximale sur plus de 10 000 observations est de 2 260 203$.

Tableau 2.2 Comparaison des quantiles empiriques avec ceux estimés par la loi de Pareto

Quantile	Théorique	Empirique
99%	193 993 067	209 080
95%	2 161 829	88 293
90%	311 691	60 213

2.2 Distribution bêta prime

La distribution bêta prime, aussi appelée distribution bêta inverse ou distribution bêta du second type, a la fonction de densité suivante :

$$f(x) = \frac{x^{\alpha-1}(1+x)^{-\alpha-\beta}}{B(\alpha,\beta)}, x \in \mathbb{R}^+$$

où $\alpha > 0$ et $\beta > 0$ sont des paramètres de forme et $B(\alpha,\beta)$ est la fonction bêta, c'est-à-dire : $B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

2.2.1 Estimation des paramètres

Il y a très peu d'articles dans la littérature à propos de l'estimation des paramètres de la distribution bêta prime. Puisque dans la plupart des cas, une méthode numérique de maximisation de la vraisemblance a été utilisée, nous utiliserons aussi cette méthode dans le mémoire. Voir (Kleiber et Kotz, 2003) pour une revue de la littérature sur l'estimation de la distribution bêta prime

Les estimateurs obtenus sont présentés dans le tableau 2.3 (notez que l'estimation est faite à partir de la base de données où les réclamations inférieures à 500$ ont été supprimées). On remarque que le AIC est meilleur que celui obtenu avec la loi de Pareto. Ceci pourrait indiquer un meilleur ajustement des données.

Tableau 2.3 Paramètres estimés pour une distribution bêta prime

$\hat{\alpha}$	2117.27
$\hat{\beta}$	0.66
AIC	206 373

Tableau 2.4 Comparaison des quantiles empiriques avec ceux estimés par la loi bêta prime

Quantile	Théorique	Empirique
99%	2 787 950	209 080
95%	238 228	88 293
90%	81 919	60 213

2.2.2 Quantiles estimés

Les quantiles estimés avec une loi bêta prime seront maintenant présentés. Notez qu'il n'y a pas de forme fermée pour les quantiles d'une loi bêta prime et qu'une méthode d'estimation numérique a été utilisée. Le tableau 2.4 présente les quantiles estimés par la loi bêta prime. Le 99^e percentile théorique est très loin du 99^e percentile empirique.

2.3 Distribution gamma

La distribution gamma sera aussi utilisée pour modéliser les données. La fonction de densité de la distribution gamma, telle que présentée dans (Kleiber et Kotz, 2003), est la suivante :

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \tag{2.1}$$

pour $x > 0$ et $\alpha, \beta > 0$. Ici α est un paramètre de forme et $1/\beta$ un paramètre d'échelle.

2.3.1 Estimation par maximum de vraisemblance

Encore une fois, la méthode du maximum de vraisemblance sera utilisée pour estimer les paramètres de la distribution. La fonction de logvraisemblance est :

$$l(\alpha, \beta|\mathbf{y}) = n(\alpha \log(\beta) - \log(\Gamma(\alpha)) + (\alpha - 1)\sum_{i=1}^{n} \log(y_i) - \alpha \sum_{i=1}^{n} y_i$$

L'estimation du paramètre β est bien connue : $\hat{\beta} = \frac{n\hat{\alpha}}{\sum_{i=1}^{n} y_i}$, voir (Klugman, Panjer et Willmot, 2004). Par contre, l'estimation du paramètre α est un peu plus complexe et n'a pas de forme fermée. Dans (Choi et Wette, 1969), on présente une méthode itérative avec l'algorithme de Newton-Raphson pour estimer la valeur de α. La valeur de $\hat{\alpha}$ à l'étape k soit $\hat{\alpha}_k$ est donnée par :

$$\hat{\alpha}_k = \hat{\alpha}_{k-1} - \frac{\log(\hat{\alpha}_{k-1} - \psi(\hat{\alpha}_{k-1} - M)}{1/\hat{\alpha}_{k-1} - \psi'(\hat{\alpha}_{k-1})},$$

où $\psi(x)$ et $\psi'(x)$ sont les fonctions digamma et trigamma respectivement. Ces fonctions se définissent par $\psi(x) = \frac{d\Gamma(x)}{dx}$ et $\psi'(x) = \frac{\psi(x)}{dx}$. Il est montré dans (Choi et Wette, 1969) que cet algorithme converge pour toutes valeurs initiales de $0 < \hat{\alpha}_0 < \infty$.

2.3.2 Estimation des paramètres avec les données d'assurance

La méthode décrite plus haut fut appliquée aux données de réclamations pour blessures personnelles. Le tableau 2.5 résume les résultats (notez que l'estimation est faite à partir de la base de données où les réclamations inférieures à 500$ ont été supprimées).

Si l'on se fie uniquement au AIC, la distribution gamma est un moins bon choix que les distributions Pareto et bêta prime.

2.3.3 Quantiles estimés

Ici aussi, il n'y a pas de forme fermée pour les quantiles de la distribution gamma. Une méthode numérique est donc utilisée. Le tableau 2.6 présente quelques quantiles selon la distribution gamma estimée et leurs contreparties estimées empiriquement.

Tableau 2.5 Paramètres estimés pour une distribution gamma

$\hat{\alpha}$	1.1384
$\hat{\beta}$	< 0.0001
AIC	213 232

Tableau 2.6 Comparaison des quantiles empiriques avec ceux estimés par la loi gamma

Quantile	Théorique	Empirique
99%	110 345	209 080
95%	73 161	88 293
90%	56 998	60 213

Il semble que les quantiles sont mieux estimés avec la distribution gamma qu'avec la loi de Pareto. Par contre, les quantiles sont tous sous-estimés et pour le 99^{e} percentile, cette différence devient très significative. Ceci pointe encore vers une modélisation de la queue avec une distribution quelconque (Pareto par exemple) et du reste avec une autre distribution (la distribution gamma semble être un candidat intéressant). Dans le chapitre sur les estimateurs à noyau, nous explorerons des procédures qui nous permettent de bien estimer l'ensemble de la distribution sans passer par la théorie des risques extrêmes.

2.4 Distribution Weilbull

La fonction de distribution de la loi de Weibull telle que définie dans (Kleiber et Kotz, 2003) est :

$$F(x) = 1 - e^{-(x/\beta)^{\alpha}}.$$

On remarque que lorsque $\alpha = 1$, on retrouve la loi exponentielle. La distribution de Weibull fut utilisée dans un contexte actuariel entre autres dans (Cummins et al., 1990) pour modéliser le montant des pertes dues aux incendies.

Tableau 2.7 Paramètres estimés pour une distribution de Weilbull

$\hat{\alpha}$	0.5805
$\hat{\beta}$	13 114.59
AIC	177 663

2.4.1 Estimation par maximum de vraisemblance

La méthode du maximum de vraisemblance sera utilisée pour estimer les paramètres de la distribution. La fonction de logvraisemblance est :

$$l(\alpha, \beta|\mathbf{y}) = n(\log(\alpha) - \log(\beta)) + (\alpha - 1)\sum_{i=1}^{n}\left[\log\left(\frac{y_i}{\beta}\right) - \left(\frac{y_i}{\beta}\right)^{\alpha}\right],$$

où $\mathbf{y}$ est un vecteur d'observation de longueur n. Le système d'équations à résoudre est donc :

$$\frac{\partial l(\alpha, \beta|\mathbf{y})}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n}\log\left(\frac{y_i}{\beta}\right)\left(1 - \left(\frac{y_i}{\beta}\right)^{\alpha}\right) = 0$$

$$\frac{\partial l(\alpha, \beta|\mathbf{y})}{\partial \beta} = \alpha\beta^{-\alpha-1}\sum_{i=1}^{n} y_i^{\alpha} - \frac{n\alpha}{\beta} = 0.$$

Ces équations sont résolues numériquement.

2.4.2 Estimation des paramètres avec les données d'assurance

La méthode décrite plus haut fut appliquée aux données de réclamations pour blessures personnelles. Le tableau 2.7 résume les résultats (notez que l'estimation est faite à partir de la base de données où les réclamations inférieures à 500$ ont été supprimées). C'est la distribution présentant le meilleur AIC jusqu'à maintenant.

2.4.3 Quantiles estimés

Les quantiles sont donnés par :

$$q_p = (-\log(p))^{1/\alpha}\beta$$

Tableau 2.8 Comparaison des quantiles empiriques avec ceux estimés par la loi Weilbull

Quantile	Théorique	Empirique
99%	182 090	209 080
95%	86 815	88 293
90%	55 172	60 213

Au tableau 2.8 on observe que l'estimation des quantiles est du même ordre de grandeur que leurs contreparties empiriques. Par contre, ils sont tous sous-estimés. Somme toute, la distribution de Weilbull semble un choix intéressant pour la modélisation des données étudiées dans ce mémoire.

2.5 Distribution Champernowne

Finalement, la distribution Champernowne sera étudiée. Cette distribution est un peu moins bien connue par les actuaires et fut principalement utilisée pour modéliser la distribution des revenus dans les sociétés (pour plus de détails, voir (Kleiber et Kotz, 2003)). Par contre, dernièrement la distribution Champernowne est de plus en plus souvent utilisée comme estimation préliminaire dans les estimateurs à noyau semi-paramétrique (voir le chapitre 4). De plus, la queue de la distribution Champernowne converge vers une loi de Pareto ce qui fait de cette distribution un choix intéressant pour modéliser des distributions à queues lourdes.

La fonction de répartition est :

$$F(x) = \frac{(x+c)^\alpha - c^\alpha}{(x+c)^\alpha + (M+c)^\alpha - 2c^\alpha},$$

où $\alpha, M > 0$ et $c \geq 0$.

2.5.1 Estimation des paramètres

Deux méthodes d'estimation seront comparées. Pour les deux méthodes nous utiliserons le fait que $F(0.5) = M$ et nous estimerons M par la médiane empirique. La différence

entre les deux méthodes est sur l'estimation du couple de paramètres (α, c).

Dans la première méthode, on estime le couple (α, c) par maximum de vraisemblance. Cette méthode est la plus courante dans la littérature (voir (Charpentier et Oulidi, 2010) ou (Buch-Larsen et al., 2005) entre autres).

Pour la deuxième approche, on estime α de manière à ce que le $95^{\text{ème}}$ percentile de la distribution estimée coïncide avec le $95^{\text{ème}}$ percentile empirique. On estime ensuite c pour que la moyenne de la distribution empirique soit le plus près possible de la moyenne empirique. Dans certains cas, il serait peut-être plus approprié de choisir d'autres critères, mais ici, comme ces distributions sont utilisées pour la tarification en assurance, il est important de bien modéliser la queue, d'où le choix d'utiliser un quantile élevé, et que la moyenne soit bonne. Cette méthode est introduite dans (Buch-Larsen, 2005).

2.5.2 Première Méthode

La méthode utilisant le maximum de vraisemblance décrite plus haut fut appliquée aux données de réclamations pour blessures personnelles. Le tableau 2.9 résume les résultats (notez que l'estimation est faite à partir de la base de données où les réclamations inférieures à 500$ ont été supprimées).

Le AIC est le plus bas de tous les AIC calculés précédemment. Reste à voir si l'estimation des quantiles est satisfaisante.

2.5.3 Deuxième Méthode

Pour cette méthode on estime d'abord M par la médiane empirique, donc $M = 7914.32$. Ensuite, on sélectionne α de façon à ce que le $95^{\text{ème}}$ percentile estimé soit le même que le $95^{\text{ème}}$ percentile empirique. On a donc la fonction suivante :

$$F(x = 88293.14; \alpha, c) = \frac{(88293.14 + c)^{\alpha} - c^{\alpha}}{(88293.14 + c)^{\alpha} + (7914.32 + c)^{\alpha} - 2c^{\alpha}}. \tag{2.2}$$

La figure 2.1 présente le graphique de cette fonction. On peut y voir que (2.2) atteint 0.95 lorsque $\alpha = 1.214$.

Une fois les paramètres M et α choisis, il reste à déterminer c de façon à ce que la moyenne estimée soit le plus près possible de la moyenne empirique. La moyenne échantillonnale est de 25557.94. Comme la moyenne estimée augmente lorsque c augmente, et que la moyenne lorsque $c = 0$ est de 38435, la valeur de c sera donc estimée à 0.

On a alors les estimations présentées dans le tableau 2.9. Comme on peut le voir, les paramètres estimés et les AIC des deux méthodes sont très similaires. Notez que le AIC est fonction inverse de la vraisemblance, voir section 1.1.7. Il est donc normal que le AIC de la première méthode, qui vise à maximiser la vraisemblance, soit inférieur à celui de la deuxième.

La distribution Champernowne sera à nouveau rencontrée à nouveau au chapitre 4, portant sur les estimateurs à noyau semi-paramétrique. Comme les propriétés des estimateurs par maximum de vraisemblance sont mieux connues seulement cette méthode sera utilisée pour estimer les paramètres de la distribution.

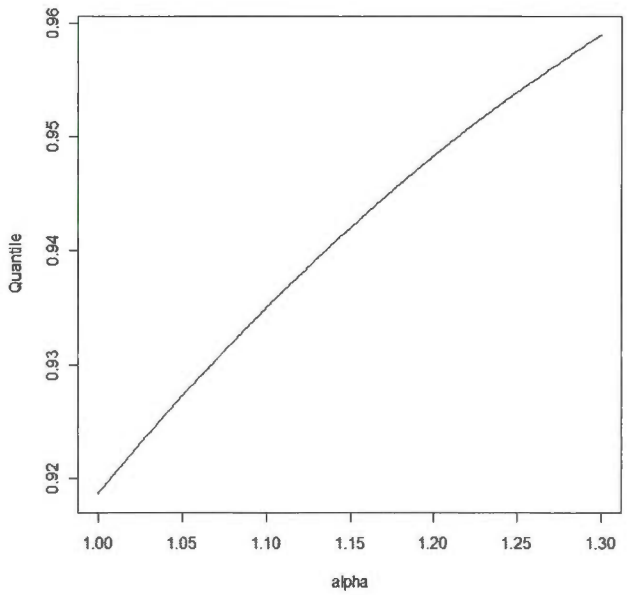

Figure 2.1 Graphique de l'équation (2.2) en fonction du paramètre α.

Tableau 2.9 Paramètres estimés pour une distribution Champernowne

Paramètres	Méthode 1	Méthode 2
$\hat{\alpha}$	1.1174	1.2140
$\hat{c}$	1.4932E-15	0.0000
$\hat{M}$	7 814.3200	7 814.3200
AIC	206 339	206 438

Tableau 2.10 Comparaison des quantiles empiriques avec ceux estimés par la loi Champernowne

Quantile	Théorique(méthode 1)	Théorique(méthode 2)	Empirique
99%	477 369.10	344 145.6	209 080.35
95%	108 966.30	88 354.82	88 293.14
90%	55 831.06	47 744.45	60 213.19

2.5.4 Quantiles estimés

Le tableau 2.10 présente quelques quantiles estimés avec la distribution Champernowne et leurs contreparties estimées empiriquement.

Ces estimations sont plus intéressantes que les précédentes. Pour les quantiles à 90% et 95%, la valeur estimée est près de la valeur observée empiriquement (pour la deuxième méthode, il est naturel que le quantile à 95% soit très près de la valeur observée).

De plus la valeur du quantile à 99% est plus élevée que la valeur observée, ce qui est souhaitable. En effet, le but de l'analyse de la queue d'une distribution est d'aller au-delà des données.

La distribution Champernowne semble donc un bon point de départ pour modéliser les réclamations pour blessures corporelles.

CHAPITRE III

ESTIMATION DE DENSITÉ - ESTIMATEUR À NOYAU

3.1 Introduction

Il y a plusieurs approches pour estimer une densité à partir de données. Il y a l'approche paramétrique, où l'objectif est d'estimer les paramètres d'une distribution connue. Cette approche fut brièvement discutée au chapitre 1. Une autre approche est l'estimation non paramétrique, la méthode la plus connue étant l'estimation à noyau. C'est cette dernière méthode qui sera appliquée dans ce chapitre [1].

Une brève description des estimateurs à noyau traditionnel sera faite à la section 3.2. Certains problèmes des noyaux traditionnels seront discutés et le noyau gamma sera alors proposé à la section 3.2.1. Les propriétés asymptotiques ainsi que le choix d'une largeur de bande optimale pour un estimateur à noyau gamma seront également traités à la section 3.2.3.

3.2 Estimateur à noyau

L'idée dans l'estimation non paramétrique est de laisser les données parler d'elles-mêmes. Donc, au lieu de supposer une loi paramétrique et d'utiliser un échantillon pour estimer les paramètres, on voudrait utiliser directement l'échantillon comme estimation de la

---

1. Bien sûr, il y a beaucoup d'autres méthodes d'estimation de densité, pensons à l'approche bayésienne par exemple.

densité.

La méthode la plus simple serait d'utiliser un histogramme comme estimation de la densité. Toutefois, un premier problème vient du fait que l'histogramme nous donne une densité qui n'est pas continue partout, alors qu'ici la variable modélisée est continue. On pourrait augmenter le nombre de barres dans l'histogramme, mais alors il y a de moins en moins d'observations dans chaque intervalle. Ce qui rend l'estimation très instable, comme observée à la figure 3.1.

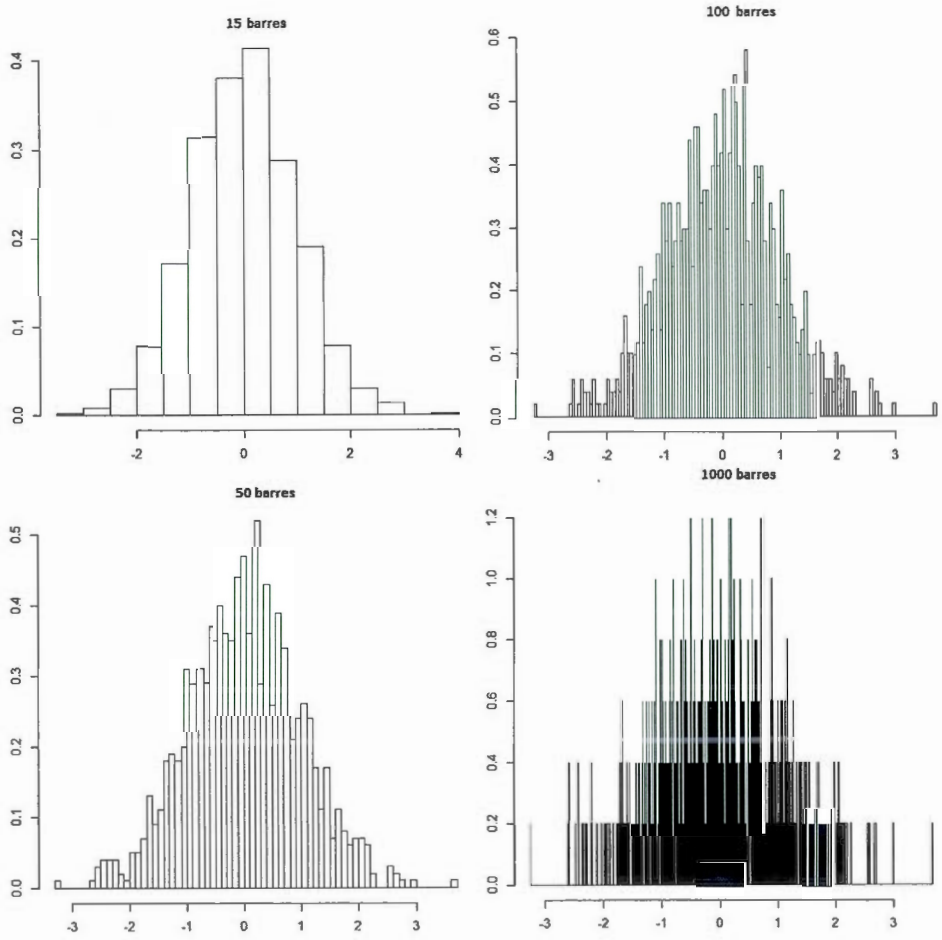

Figure 3.1 Plusieurs histogrammes avec différents nombres d'intervalles, tous calculés à partir du même échantillon de taille 1000 provenant d'une loi normale de moyenne 0 et de variance 1.

Une autre solution intuitive, pour transformer l'histogramme en une fonction continue, est de relier les points centraux des barres de l'histogramme, comme à la figure 3.2. Ici encore, il pourrait être tentant d'augmenter le nombre de barres de l'histogramme pour réduire l'espace entre les points centraux, mais le problème d'instabilité reste le même. De plus, il faudrait alors s'assurer que la fonction ainsi obtenue intègre à 1 pour que ce soit une densité.

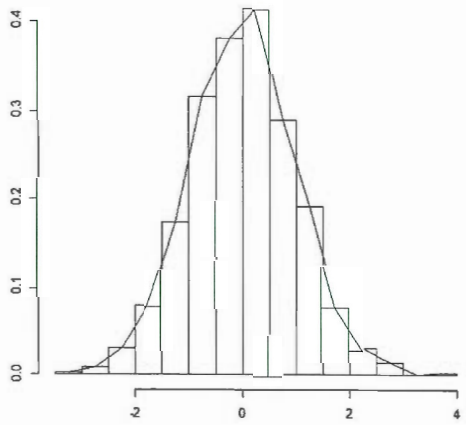

Figure 3.2 Histogramme calculé à partir d'un échantillon provenant d'une loi normale de moyenne 0 et de variance 1 dont les points centraux des barres sont reliés par une droite.

Une solution simple et élégante, pour utiliser l'idée de l'histogramme comme estimation de la fonction de densité, est ce que l'on pourrait appeler l'histogramme mobile. L'idée de l'histogramme mobile est que pour estimer la densité au point x, on construit une barre de l'histogramme ayant comme point central le point x et une largeur de $2h$. La valeur de la densité estimée évaluée au point x serait donc :

$$\hat{f}(x) = (nh)^{-1} \sum_{i=1}^{n} I_{x-h<x_i<x+h}, \tag{3.1}$$

où I est la fonction indicatrice et les x_i sont les éléments d'un échantillon et $i = 1, \ldots, n$. L'équation (3.1) correspond justement à ce qui est appelé un estimateur à noyau, dans notre cas, un noyau rectangulaire.

De façon plus générale, les estimateurs à noyau se définissent comme suit. Soit un échan-

Tableau 3.1 Exemples de noyaux symétriques

Type de noyau	Fonction $K(u)$
Epanechnikov	$(3/4)(1-u^2)I_{-1,1}(u)$
Rectangulaire	$(1/2)I_{(-1,1)}(u)$
Triangulaire	$(1-\|u\|)I_{(-1,1)}(u)$
Bi-poids	$(15/16)(1-u^2)^2 I_{(-1,1)}(u)$
Gaussien	$\exp(-u^2/2)/\sqrt{2\pi}$

tillon aléatoire $x_1, \ldots, x_n$, où les x_i sont indépendants et proviennent d'une distribution inconnue f, alors l'estimateur à noyau est donné par :

$$\hat{f}(x) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right). \tag{3.2}$$

Dans l'équation (3.2), K et h représentent la fonction noyau et la largeur de bande respectivement. K est habituellement symétrique par rapport à x (voir table 3.1 pour quelques exemples et (Parzen, 1962) pour une revue plus exhaustive). Notez que la fonction $I_{-1,1}(u)$ dans les fonctions noyau du tableau 3.1 est la fonction indicatrice et se définit comme :

$$I_{-1,1}(u) = \begin{cases} 1 & \text{si } -1 \leq u \leq 1 \\ 0 & \text{sinon} \end{cases}$$

De plus, la fonction $K(\cdot)$ doit satisfaire les conditions suivantes : $0 \leq K(t) < \infty, \forall t$ et $\int_{-\infty}^{\infty} K(t)dt = 1$. Donc $K(\cdot)$ est une fonction de densité. Il est commun d'écrire $K_h(t) = h^{-1}K(t/h)$ et alors l'estimateur de f devient :

$$\hat{f}(x) = (n)^{-1} \sum_{i=1}^{n} K_h(x - x_i). \tag{3.3}$$

3.2.1 Noyau gamma

Dans la littérature, il est généralement accepté que le choix de la largeur de bande soit plus important que le choix du noyau (voir (Bouezmarni, El Ghouch et Mesfioui, 2011)

par exemple). Par contre, lorsque le support de f est $[0, \infty)$, les noyaux traditionnels (symétriques) donnent une estimation de plus en plus biaisée lorsque x approche de 0. Ce biais est dû au fait que les noyaux symétriques donnent des poids à l'extérieur du support de f (voir figure 3.3).

À la figure 3.3, on voit que, si le domaine des données analysées est $[0, \infty)$, le noyau gaussien donne des poids pour les valeurs inférieures à zéro. Pire, il donne de plus en plus de poids aux valeurs négatives, plus on veut estimer la densité proche de zéro. Rappelons qu'un estimateur à noyau doit satisfaire $\int_{-\infty}^{\infty} K(t)dt = 1$, mais dans le cas d'une variable aléatoire $X \in [0, \infty)$, on veut avoir $\int_0^{\infty} K(t)dt = 1$, ce qui n'est pas le cas avec les noyaux symétriques, d'où le problème de biais multiplicatif aux bornes.

Comme les données analysées ici représentent des montants de réclamations, elles ont un support de $(0, \infty)$. Il est donc important de tenir compte de ce problème de biais multiplicatif aux bornes. Une solution est d'utiliser un noyau qui ne donne pas de poids aux valeurs de x négatives. (Chen, 2000) propose d'utiliser le noyau gamma. Il propose d'abord le noyau gamma suivant :

$$K_{\frac{x}{b}+1,b}(t) = \frac{t^{x/b}e^{-t/b}}{b^{\frac{x}{b}+1}\Gamma(\frac{x}{b}+1)},$$

où b est un paramètre de lissage. L'estimateur de f est donné par :

$$\hat{f}_1(x) = n^{-1}\sum_{i=1}^{n} K_{\frac{x}{b}+1,b}(x_i). \tag{3.4}$$

On voit clairement que $K_{\frac{x}{b}+1,b}(t)$ est la fonction de densité d'une loi gamma avec $\alpha = \frac{x}{b} + 1$ et $\beta = b$, où α est un paramètre de forme et β est un paramètre d'échelle. Donc, la forme du noyau utilisée change en fonction du x où l'on veut estimer la densité f, comme on peut voir sur la figure 3.4.

En observant la figure 3.4, on remarque que la forme du noyau change considérablement en fonction de la valeur de x où l'on veut estimer la densité. Pour $x = 0$, le noyau est très asymétrique et donne de grands poids pour les valeurs près de zéro, puis les poids décroissent rapidement par la suite. Par contre, pour $x = 10$, le noyau devient plus

symétrique et plus plat. Remarquez que dans tous les cas, le noyau gamma donne des poids positifs sur l'ensemble du domaine de X, soit $(0, \infty)$.

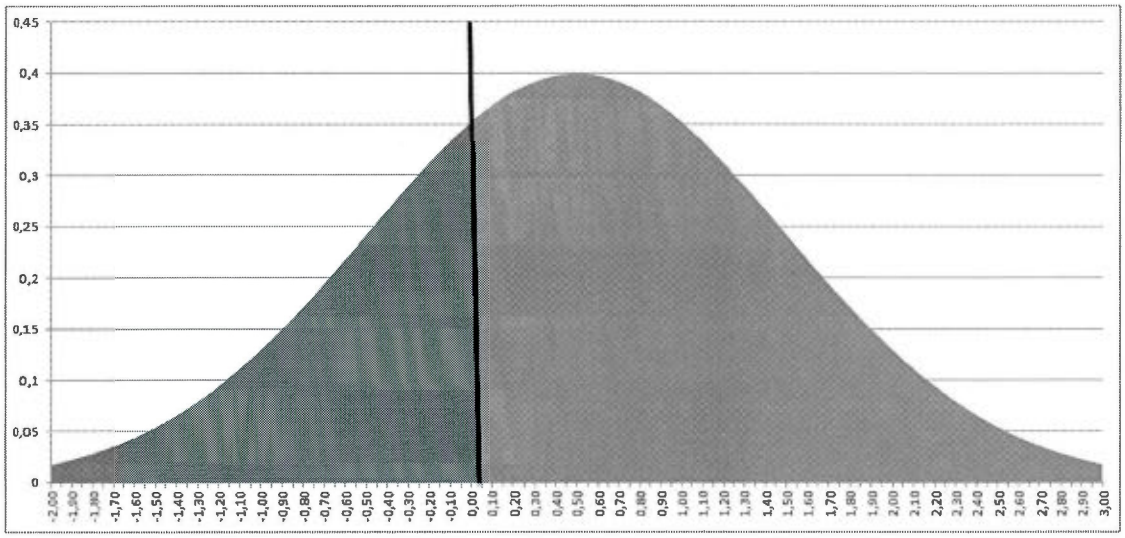

Figure 3.3 Noyau gaussien avec une largeur de bande de 1 autour du point $x = 0.5$.

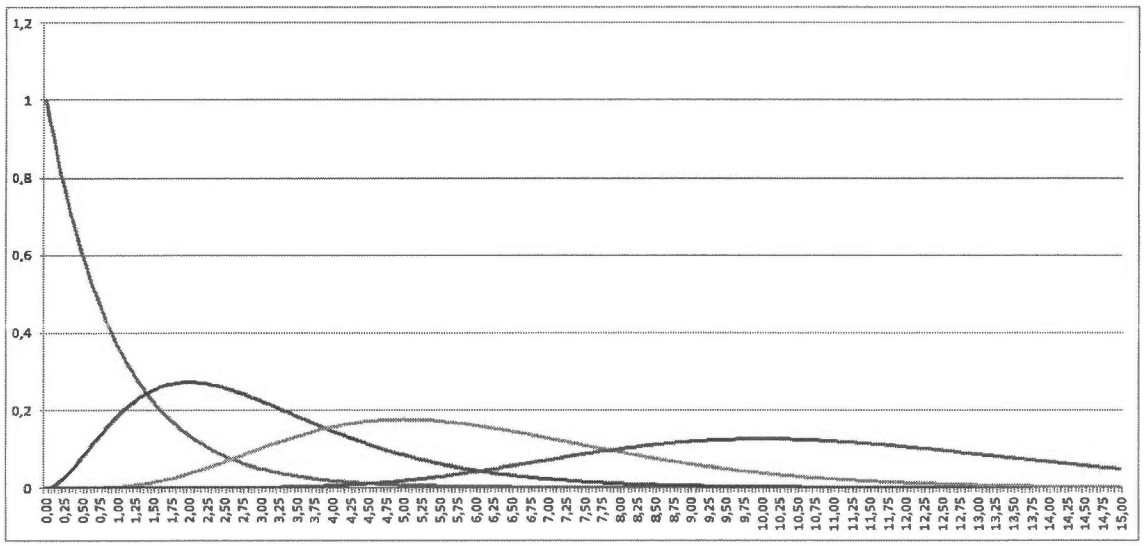

Figure 3.4 Noyaux gamma, tous avec une largeur de bande de 1, pour différentes valeurs de x, $x = 0, 2, 5$ et 10 de gauche à droite.

3.2.2 Analyse du biais de l'estimateur à noyau gamma

Afin de vérifier si l'estimateur à noyau gamma tel que défini en (3.4) n'est pas sujet au problème de biais multiplicatif aux bornes, nous avons besoin de trouver l'espérance de cet estimateur. Une fois que nous aurons l'espérance, il faudra vérifier que cet estimateur est asymptotiquement sans biais pour des valeurs de x près de 0.

En utilisant le fait que l'espérance d'une somme de variables aléatoires est égale à la somme des espérances (voir le chapitre 1), l'espérance de $\hat{f}_1(x)$ est donnée par :

$$\begin{aligned} E_X[\hat{f}_1(x)] &= E_X\left[n^{-1}\sum_{i=1}^{n} K_{x/b+1,b}(X_i)\right] \\ &= E_X\left[K_{x/b+1,b}(X)\right] \\ &= \int_0^\infty K_{x/b+1,b}(t)f(t)dt, \end{aligned}$$

où E_X dénote l'espérance prise par rapport à la distribution de X, c'est-à-dire la densité inconnue f que l'on cherche à estimer. Soit une variable aléatoire $\zeta_x \sim Gamma(x/b + 1, b)$, alors on voit que $E_X[\hat{f}_1(x)] = E_{\zeta_x}[f(\zeta_x)]$. Il est bien connu que $E[\zeta_x] = \mu = x + b$ et que $Var[\zeta_x] = xb+b^2$. Finalement, par expansion de Taylor de $f(\zeta_x)$ au tour du point μ on obtient :

$$\begin{aligned} E_X[\hat{f}_1(x)] &= E_{\zeta_x}\left[\sum_{i=0}^{\infty} \frac{f^{(i)}(\mu)}{i!}(\zeta_x - \mu)^i\right] \\ &= f(\mu) + \frac{1}{2}f^{(2)}(\mu)Var[\zeta_x] + o(b) \end{aligned} \tag{3.5}$$

où $f^{(i)}$ est la $i^{\text{ème}}$ dérivée de f, et donc :

$$E_X[\hat{f}_1(x)] = f(x+b) + \frac{1}{2}f^{(2)}(x+b)(xb+b^2) + o(b), \tag{3.6}$$

où $o(b)$ signifie petit ordre de b. Soit deux séquences a_n et b_n, $a_n = o(b_n)$ si et seulement si $\lim_{n\to\infty} a_n/b_n = 0$.

En définissant $b = h(n)$, une fonction telle que $\lim_{n\to\infty} h(n) = 0$, alors : $\lim_{n\to\infty} E_X[\hat{f}_1(x)] = f(x)$ et ce $\forall x$. L'estimateur $\hat{f}_1(x)$ n'est donc pas sujet au biais multiplicatif aux bornes et est asymptotiquement sans biais. Notez qu'il est impossible d'avoir un estimateur de

densité non biaisé, voir (Rosenblatt, 1956). Donc le mieux que l'on peut faire est d'utiliser un estimateur asymptotiquement sans biais. Comme ici on travaille avec une base de données de plusieurs milliers d'observations, il est raisonnable d'utiliser des résultats asymptotiques.

Cet estimateur ne souffre pas du problème de biais multiplicatif aux bornes. Bien qu'il soit asymptotiquement sans biais, il est possible d'améliorer son biais additif. En observant l'équation (3.6), on remarque que le biais additif vient de deux sources (si on ignore le terme $o(b)$). Il y a d'abord le b dans la fonction de densité : $f(x+b)$, puis il y a un terme qui dépend de la dérivée seconde de f : $\frac{1}{2}f^{(2)}(x+b)(xb+b^2)$. Donc, s'il était possible de se débarrasser du b dans le premier terme, c'est-à-dire avoir $f(x)$ et non $f(x+b)$, le biais de l'estimateur en serait diminué.

Ce problème est dû au fait que le point x, où l'on veut estimer la densité f, n'est pas l'espérance du noyau $K_{x/b+1,b}$ mais plutôt son mode. Ce fait devient évident lorsqu'on porte attention au passage de l'équation (3.5) à l'équation (3.6). Le noyau $K_{x/b,b}$, c'est-à-dire la fonction de densité d'une $Gamma(x/b, b)$, a bien une espérance de x, mais $K_{x/b,b} \to \infty$ lorsque $x \to 0$ ce qui contrevient aux conditions imposées aux fonctions noyaux.

Comme solution, (Chen, 2000) définit le pont suivant :

$$\rho_b(x) = \begin{cases} x/b & \text{si } x \geq 2b \\ \frac{1}{4}(x/b)^2 + 1 & \text{si } x \in [0, 2b) \end{cases}.$$

L'estimateur de f devient donc :

$$\hat{f}_2(x) = n^{-1} \sum_{1}^{n} K_{\rho_b(x),b}(x_i). \tag{3.7}$$

Cet estimateur utilise donc un noyau $K_{x/b,b}$ assez loin dans le domaine de X et $K_{\frac{1}{4}(x/b)^2+1,b}$ près de $x = 0$.

Il peut être montré, par une méthode semblable à celle utilisée en (3.6), que :

$$E_X[\hat{f}_2(x)] = \begin{cases} f(x) + b\frac{1}{2}xf^{(2)}(x) + o(b) & \text{si } x \geq 2b \\ f(x) + \xi_b(x)bf^{(1)}(x) + o(b) & \text{si } x \in [0, 2b) \end{cases}$$

où $\xi_b(x) = (1-x)(\rho_b(x) - x/b)/(1 + b\rho_b(x) - x)$.

Dans ce qui suit l'expression "noyau gamma" fait référence à $K_{\rho_b(x),b}$ et pour alléger la notation, $\hat{f}_2(x)$ est noté par $\hat{f}(x)$.

3.2.3 Choix du paramètre de lissage

Il y a essentiellement deux approches pour trouver une largeur de bande optimale. La première consiste à trouver la largeur de bande $b_f^*(x,n)$ qui minimise l'erreur quadratique moyenne (EQM) de $\hat{f}(x)$, c'est-à-dire $\text{argmin}_b E[(\hat{f}(x) - f(x))^2]$. On obtient donc une largeur de bande optimale, qui varie en fonction du x où l'on veut estimer la fonction de densité f. La seconde approche nous donne une largeur de bande optimale globale, qui ne dépend pas de x. Pour ce faire, on trouve la largeur de bande $b_f^*(n)$ qui minimise l'erreur quadratique moyenne intégrée ($EQMI$), c'est-à-dire $\text{argmin}_b \int_0^\infty E[(\hat{f}(x) - f(x))^2]dx$.

Afin de calculer la valeur de l'EQM ou de l'$EQMI$, dans (Chen, 2000), on utilise une expansion de Taylor de façon similaire à ce qui a été fait ici pour obtenir l'équation (3.6). L'EQM et l'$EQMI$ de l'estimateur à noyau gamma défini en (3.7), tels que présentés dans (Chen, 2000) sont

$$EQM[\hat{f}(x)] = \frac{1}{4}(xf^{(2)}(x))^2b^2 + \frac{1}{2\sqrt{\pi}}n^{-1}(bx)^{-1/2}f(x) + o(b^2 + (nb)^{-1/2}), \qquad (3.8)$$

$$\begin{aligned} EQMI[\hat{f}(x)] &= \frac{b^2}{4}\int_0^\infty (xf^{(2)}(x))^2dx \\ &\quad + \frac{n^{-1}b^{-1/2}}{2\sqrt{\pi}}\int_0^\infty x^{-1/2}f(x)dx + o(b^2 + n^{-1}b^{-1/2}). \end{aligned} \qquad (3.9)$$

Pour trouver la largeur de bande b qui minimise les termes principaux (on omet le $o(\cdot)$)

on dérive par rapport à b et on égale à 0 :

$$\frac{\partial EQM[\hat{f}(x)]}{\partial b} = \frac{1}{2}b(xf^{(2)}(x))^2 - \frac{1}{4\sqrt{\pi}}n^{-1}b^{-3/2}x^{-1/2}f(x) = 0.$$

Donc les fonctions qui minimisent (3.8) et (3.9) sont respectivement :

$$b_f^*(x,n) = \left(\frac{\frac{1}{2\sqrt{\pi}}x^{-1/2}f(x)}{(xf^{(2)}(x))^2 n}\right)^{2/5} \tag{3.10}$$

et

$$b_f^*(n) = \frac{\left(\frac{1}{2\sqrt{\pi}}\int_0^\infty x^{-1/2}f(x)dx\right)^{2/5}}{\left(\int_0^\infty (xf^{(2)}(x))^2 dx\right)^{2/5}} n^{-2/5}. \tag{3.11}$$

On voit clairement que les conditions, pour que $\hat{f}(x)$ soit asymptotiquement sans biais, sont respectées pour les deux largeurs de bandes optimales, c'est-à-dire $\lim_{n\to\infty} b_f^*(n) = \lim_{n\to\infty} b_f^*(x,n) = 0$ et $\lim_{n\to\infty} b_f^*(n)n = \lim_{n\to\infty} b_f^*(x,n)n = \infty$.

On remarque que f, ainsi que la dérivée seconde de f, interviennent dans (3.10) et (3.11). Toutefois le problème est que f est justement la fonction que l'on cherche à estimer.

Afin d'estimer la largeur de bande optimale, (Wand et Jones, 1995) propose de substituer les valeurs inconnues par leur estimation. Pour estimer f, il suffit d'utiliser un estimateur à noyau. Le noyau gamma semble un choix naturel dans ce cas. Pour l'estimation de la dérivée seconde de f, on ne peut utiliser le noyau gamma directement, mais il y a une manière simple d'estimer la dérivée d'une fonction de densité à l'aide des estimateurs à noyau.

Dans (Wand et Jones, 1995) on propose d'estimer $f^{(r)}$, soit la $r^{\text{ème}}$ dérivée de f, par :

$$\hat{f}^{(r)} = n^{-1}\sum_1^n \frac{\partial^r K(x_i)}{\partial x^r},$$

tant que K le permet. Cet estimateur est asymptotiquement sans biais.

La largeur de bande (3.10), qui minimise l'EQM doit être recalculée pour chaque valeur de x où l'on veut estimer la densité. Cela implique l'utilisation d'un estimateur à noyau

à deux reprises. Pour accélérer les calculs, la largeur de bande globale définie en (3.11) sera donc utilisée dans ce qui suit.

Des estimateurs à noyau seront donc utilisés pour estimer f et $f^{(2)}$ qui interviennent dans (3.2.3). Il reste à déterminer la largeur de bande utiliser dans ces estimateurs à noyau. L'idée est d'utiliser une largeur de bande optimale préliminaire en supposant que les x_i, $i = 1, \ldots, n$, proviennent d'une loi connue. On trouve ensuite la largeur de bande optimale en remplaçant f et $f^{(2)}$ dans (3.11) par la fonction de densité choisie.

Trouvons la largeur de bande optimale préliminaire en supposant que f est la fonction de densité d'une loi exponentielle de paramètre θ, c'est-à-dire $f(x) = \frac{e^{-x/\theta}}{\theta}$ alors :

$$\begin{aligned} \int_0^\infty x^{-1/2} f(x) dx &= \int_0^\infty x^{-1/2} \frac{e^{-x/\theta}}{\theta} dx \\ &= \frac{\theta^{1/2}\Gamma(1/2)}{\theta} \\ &= \frac{\sqrt{\pi}}{\theta^{1/2}}. \end{aligned} \tag{3.12}$$

De plus

$$f^{(2)}(x) = \frac{e^{-x/\theta}}{\theta^3},$$

donc de la même manière qu'en (3.12) on trouve :

$$\int_0^\infty (x f^{(2)}(x))^2 dx = \frac{1}{4\theta^3}. \tag{3.13}$$

En substituant (3.12) et (3.13) dans (3.11) on trouve :

$$b^*_{exp}(n) = \theta \left(\frac{n}{2}\right)^{-2/5}.$$

Finalement, en estimant θ par son estimateur par maximum de vraisemblance, c'est-à-dire $\hat{\theta} = \bar{x}$ où $\bar{x} = n^{-1} \sum_{i=1}^n x_i$, on a :

$$b^*_{exp}(n) = \bar{x} \left(\frac{n}{2}\right)^{-2/5}. \tag{3.14}$$

Tous les éléments sont réunis pour estimer le paramètre de lissage optimal. La méthode proposée est :

1. Poser $\tilde{b}_f^*(n) = \bar{x}\left(\frac{n}{2}\right)^{-2/5}$.
2. Calculer $\hat{b}_f^*(n) = \frac{\left(\frac{1}{2\sqrt{\pi}}\int_0^\infty x^{-1/2}\hat{f}(x)dx\right)^{2/5}}{\left(\int_0^\infty (x\hat{f}^{(2)}(x))^2 dx\right)^{2/5}} n^{-2/5}$.

À l'étape 2, $\hat{f}$ est obtenu avec (3.7) en utilisant $b = \tilde{b}_f^*(n)$ (soit une estimation préliminaire de la largeur de bande optimale), $\hat{f}^{(2)}(x) = n^{-1}\sum_1^n \frac{\partial^2 K_{x/b+1,b}(x_i)}{\partial x^2}$ et les deux intégrales sont calculées numériquement.

Le calcul de $\frac{\partial^2 K_{x/b+1,b}(x_i)}{\partial x^2}$ implique $\Gamma^{(1)}(a)$ et $\Gamma^{(2)}(a)$ qui sont obtenus à l'aide des fonctions digamma ($\psi(a)$) et trigamma ($\psi_1(a)$) respectivement. En effet :

$$\begin{aligned} \psi(a) &= \frac{d\ln\Gamma(a)}{da} = \frac{\Gamma^{(1)}(a)}{\Gamma(a)} \\ \Leftrightarrow \Gamma^{(1)}(a) &= \Gamma(a)\psi(a) \end{aligned}$$

et

$$\begin{aligned} \psi_1(a) &= \frac{d^2\ln\Gamma(a)}{da^2} = \frac{\Gamma^{(2)}(a)\Gamma(a) - \Gamma^{(1)}(a)^2}{\Gamma(a)^2} \\ \Leftrightarrow \Gamma^{(2)}(a) &= \Gamma(a)\left(\psi(a)^2 + \psi_1(a)\right). \end{aligned}$$

On remarque que le noyau $K_{x/b+1,b}$ et non $K_{\rho(x),b}$ est utilisé pour le calcul de $\hat{f}^{(2)}(x)$. Cela est dû au fait que la seconde dérivée de $K_{\rho(x),b}$ par rapport à x n'est pas définie en $x = 2b$. Donc :

$$\frac{\partial^2 K_{x/b+1,b}(t)}{\partial x^2} = \left[e^{(x/b+1)\ln t/b - t/b} \times \frac{(\ln t/b - \psi(x/b+1))^2 - \psi_1(x/b+1)}{\Gamma(x/b+1)b^2 t}\right].$$

Finalement $\tilde{b}_f^*(n)$ est une largeur de bande optimale en supposant que les données proviennent d'une loi exponentielle.

3.2.4 Résultats

La procédure décrite ci-dessus a été appliquée aux six jeux de données. Le tableau 3.2 compare le point de départ de l'algorithme et la largeur de bande obtenue pour chaque chapitre. On remarque que pour tous les chapitres, la largeur de bande optimale estimée

Tableau 3.2 Largeurs de bandes calculées

Chapitre	$b^*_{exp}(n)$	$\hat{b}^*_f(n)$
Tout risque	181.66	58.76
Vol	32.25	2.48
Collision	105.33	39.92
Responsabilité civile	65.65	12.10
Blessures corporelles (soi-même)	734.64	130.73
Blessures corporelles (autrui)	2403.58	333.89

est plus basse d'un facteur de 6, en moyenne, que la largeur de bande optimale en supposant une loi exponentielle.

Les figures 3.5 à 3.9 présentent les densités obtenues pour chaque chapitre.

3.2.5 Observations

On remarque, en regardant les figures 3.5 à 3.9, que les chapitres dont les couvertures se ressemblent ont des distributions similaires. Par contre, on voit que le chapitre tous risques est le seul à être bimodal. De plus, on peut voir que l'estimation de la queue des distributions est particulièrement instable, et ce pour toutes les distributions à l'exception de la distribution de la couverture collision. Ce problème sera adressé au chapitre 4.

Pour une description de ce qui est couvert par chaque couverture, veuillez vous référer au chapitre 1.

1. Couverture collisions (figure 3.5)

   Le mode de la distribution estimée est environ 1200$, la moyenne estimée est de 4800$ et l'écart type est de 4924$. Ces valeurs sont en ligne avec l'idée que la distribution du coût des réclamations pour la couverture collisions ne présente pas

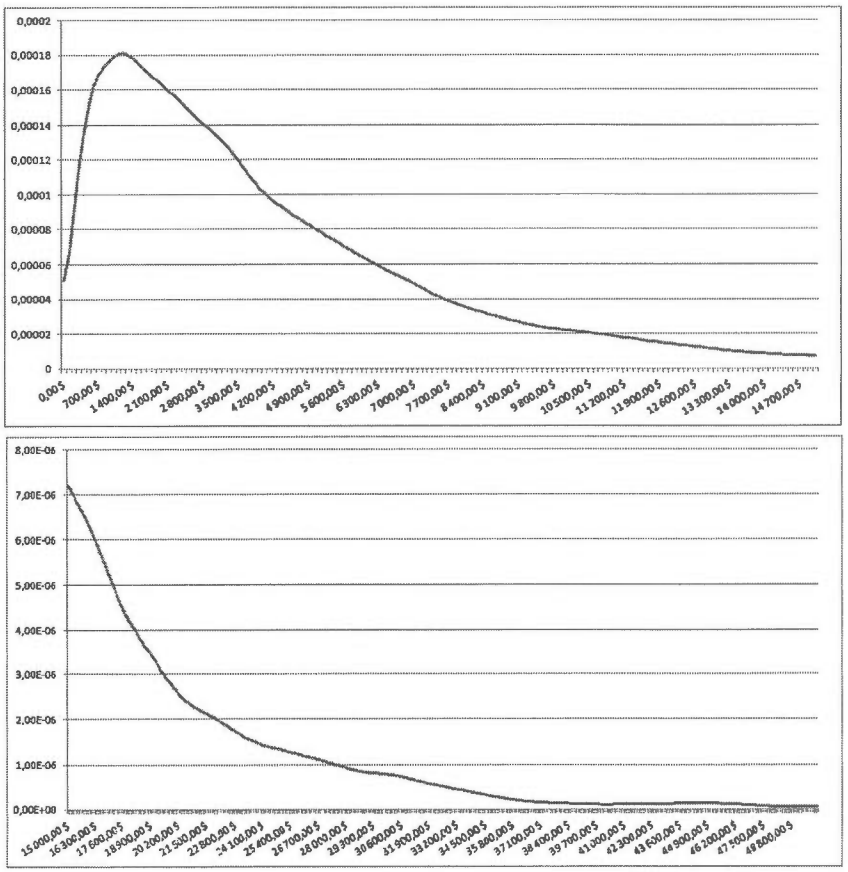

Figure 3.5 Densité estimée pour la couverture collision avec un zoom sur la queue dans le graphique du bas.

de risques extrêmes.

2. Couverture responsabilité civile (figure 3.6)

   Le mode de la distribution estimée est d'environ 720$ alors que la moyenne est de 3508$ et l'écart type est de 3717$. Ces valeurs sont toutes plus basses que pour la couverture collision, par contre elles sont proches.

3. Couverture tous risques (figure 3.7)

   Ce qui saute aux yeux lorsqu'on regarde la distribution estimée, c'est qu'elle est bimodale. Un des modes de la distribution estimée est d'environ 100$ et le deuxième est de 260$. La moyenne est de 1150$ et l'écart type est de 2324$.

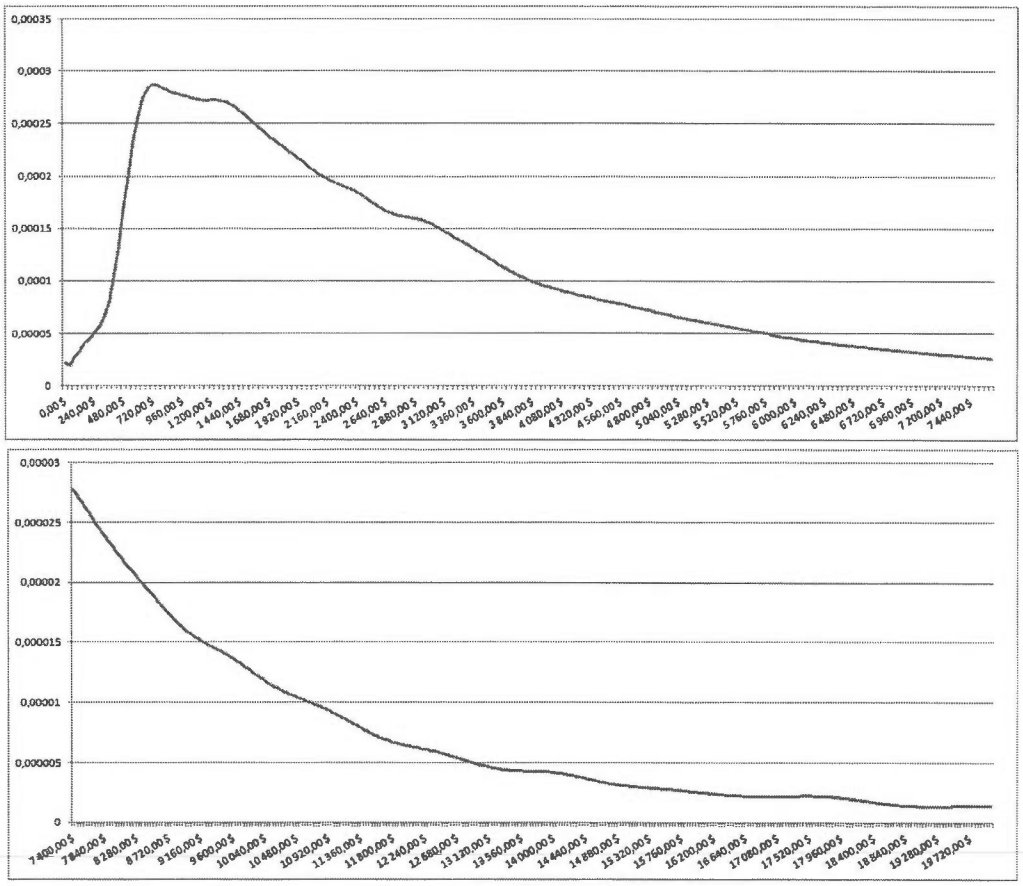

Figure 3.6 Densité estimée pour la couverture responsabilité civile avec un zoom sur la queue dans le graphique du bas.

Notez que la base de données contient plus de 7000 réclamations pour le chapitre tous risques. Donc la distribution sous-jacente est probablement réellement bimodale.

4. Couverture blessures corporelles (figure 3.8)

   Le mode de la distribution estimée est à 0$ alors que la moyenne est de 46268$ et l'écart-type est de 106500$. La moyenne est beaucoup plus élevée que pour les autres couvertures, et l'écart-type est plus de deux fois la moyenne.

5. Couverture blessures personnelles (figure 3.9)

   Le mode de la distribution estimée est à 0$ alors que la moyenne est de 22882$

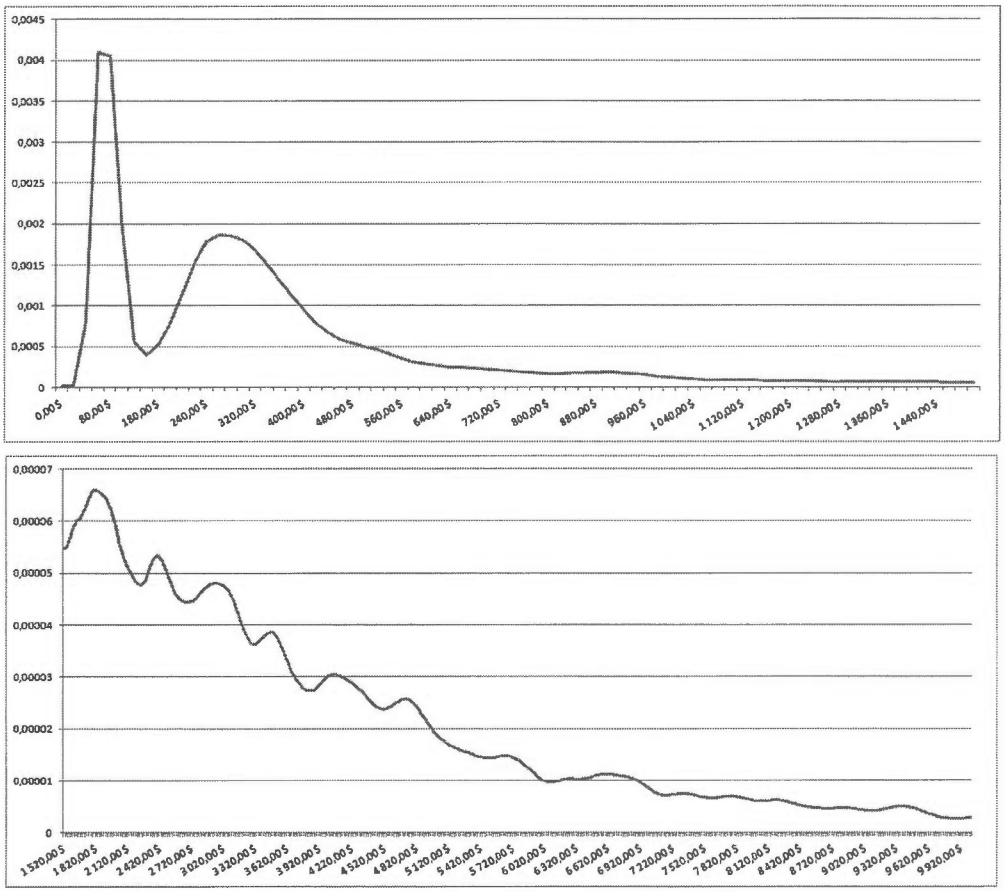

Figure 3.7 Densité estimée pour la couverture tous risques avec un zoom sur la queue dans le graphique du bas.

et l'écart type est de 75002$. Ici aussi, tout comme avec la couverture blessures corporelles, la moyenne et l'écart type sont élevés.

Il est clair, en regardant les figures 3.8 et 3.9, que les distributions estimées des montants de réclamations pour les couvertures blessures corporelles et blessures personnelles se ressemblent.

L'estimateur à noyau gamma a permis de comparer entre elles différentes distributions sans avoir à supposer une forme *a priori* pour la distribution. Ceci a permis, entre autres, de constater que la distribution des réclamations pour la couverture tous risques est probablement bimodale, ce qui aurait été pratiquement impossible avec une approche

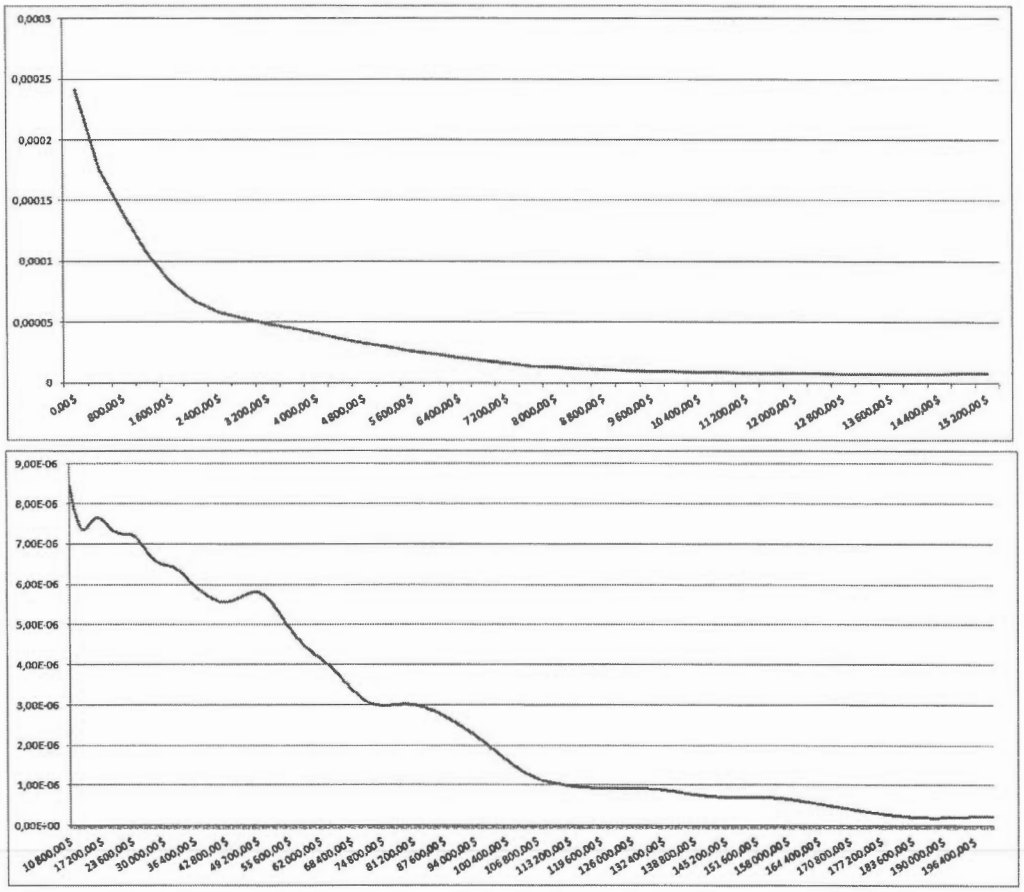

Figure 3.8 Densité estimée pour la couverture blessures corporelles avec un zoom sur la queue dans le graphique du bas.

purement paramétrique.

De plus, cette approche permet de voir les ressemblances et différences entre les distributions étudiées. Le tableau 3.3 présente quelques caractéristiques des distributions estimées à l'aide de l'estimateur à noyau gamma.

Comme les couvertures pour blessures corporelles et blessures personnelles présentent toutes deux un risque de réclamations extrêmes, elles méritent une attention particulière. Dans les chapitres qui suivent, une analyse plus approfondie de la distribution des montants de réclamations pour blessures personnelles sera faite. Cette couverture fut

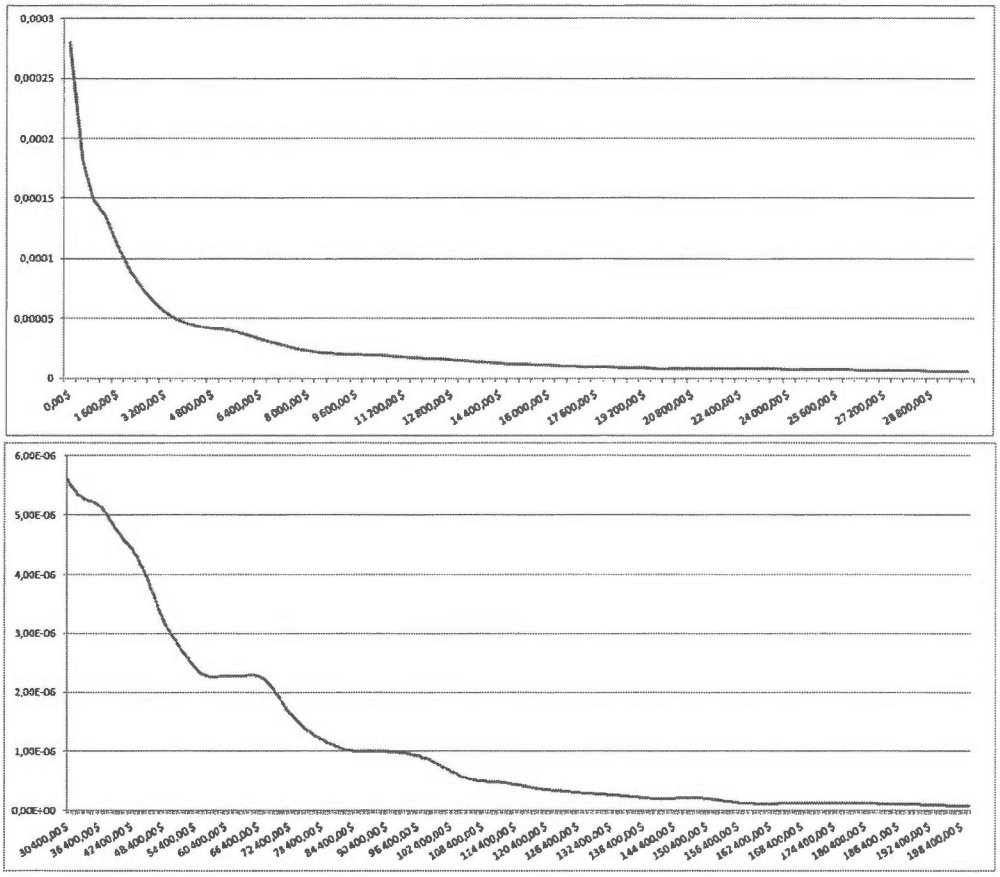

Figure 3.9 Densité estimée pour la couverture blessures personnelles avec un zoom sur la queue dans le graphique du bas.

choisie, car la base de données contient près de quatre fois plus de données pour cette couverture que pour la couverture blessures corporelles.

Tableau 3.3 Comparaison des différentes distributions estimées à l'aide du noyau gamma

Couverture	mode(s)	moyenne	écart type
Collisions	1200$	4800$	4924$
Responsabilité civile	720$	3508$	3717$
Tous risques	100$ et 260$	1150$	2324$
Blessures corporelles	0$	46268$	106500$
Blessures personnelles	0$	22882$	75002$

CHAPITRE IV

TRANSFORMATION ET ESTIMATEUR À NOYAU

Au chapitre précédent, les estimateurs à noyau traditionnel symétriques furent présentés. Comme ces estimateurs souffrent du problème de biais aux bornes lorsque le domaine de la distribution étudiée n'est pas $(-\infty, \infty)$, nous nous sommes tournés vers des estimateurs à noyau asymétriques pour régler ce problème.

Par contre, une autre solution aurait été d'estimer la densité, toujours à l'aide d'un estimateur à noyau, sur une transformation des données. En effet, il est possible d'estimer une densité à l'aide de données transformées, voir (Wand et Jones, 1995). Pour éviter le problème de biais aux bornes, il suffit alors d'utiliser une transformation qui va des réels positifs aux réels. Cette solution sera explorée dans la section 4.1.

Enfin, à la section 4.2, nous verrons comment utiliser une fonction cumulative comme transformation. Cette transformation est en quelque sorte une estimation préliminaire de la densité inconnue.

4.1 Transformation avec la famille de puissance décalée

L'estimateur à noyau gamma décrit au chapitre 3 n'a pas le problème de biais aux bornes des estimateurs à noyau classiques (noyau gaussien par exemple). Par contre, comme la largeur de bande utilisée reste constante sur tout le domaine de f, l'estimation de la queue de la distribution est très instable. En effet, plus la valeur de x est extrême, moins il y a de données disponibles à proximité de x pour effectuer l'estimation de la

densité. Dans cette section, il est montré que l'utilisation du principe de transformation semi-paramétrique de (Wand et Jones, 1995) améliore l'estimation dans la queue de la distribution.

Rappelons que l'estimateur à noyau traditionnel est :

$$\hat{f}(x) = (n)^{-1} \sum_{i=1}^{n} K_h (x - x_i),$$

où $K_h(t) = h^{-1}K(t/h)$ est la fonction noyau et h la largeur de bande.

Soit $\mathbf{F} = \{F_\theta | \theta \in \Theta\}$ un ensemble de fonctions F_θ de paramètre θ (généralement des fonctions cumulatives de probabilité, mais pas obligatoirement). Alors, par la méthode de transformation, tel que décrit dans (Wand et Jones, 1995), l'estimateur à noyau devient :

$$\hat{f}(x) = F_\theta'(x)(n)^{-1} \sum_{i=1}^{n} K_h (F_\theta(x) - F_\theta(x_i)). \tag{4.1}$$

Cet estimateur est approximativement un estimateur à noyau traditionnel appliqué aux données transformées, mais avec une largeur de bande variable égale à $h(F_\theta'(x))^{-1}$. Notez que la largeur de bande utilisée à l'intérieur de la fonction noyau reste constante afin que $\hat{f}(x)$ intègre à 1, voir (Bolancé, Guillén et Perch Nielsen, 2000). Donc, si la fonction de transformation F_θ est telle que $F_\theta'(x)$ tend vers 0 lorsque x tend vers l'infini, on aura une largeur de bande qui tend vers l'infini lorsque x tend vers l'infini, ce qui garantit une plus grande stabilité dans l'estimation de la queue de la distribution.

Dans (Bolancé, Guillén et Nielsen, 2003), on propose d'utiliser un estimateur à noyau avec la famille de puissance décalée (*shifted power transformation*) comme fonction de transformation dans le contexte de l'assurance automobile.

La transformation proposée est la suivante :

$$y_i = g_\lambda(x_i) = \begin{cases} (x_i + \lambda_1)^{\lambda_2} & \text{si } \lambda_2 \neq 0 \\ \ln(x_i + \lambda_1) & \text{si } \lambda_2 = 0 \end{cases}, \tag{4.2}$$

où $\lambda = (\lambda_1, \lambda_2)$ avec $\lambda_1 > -\min(x_1, \ldots, x_n)$ et $\lambda_2 < 1$. Les paramètres λ_1 et λ_2 sont appelés paramètres de transformation et doivent être spécifiés par l'utilisateur. L'esti-

mateur de la densité f, c'est-à-dire la fonction de distribution des X_i, est alors :

$$\hat{f}(x) = g'_\lambda(x)(n)^{-1} \sum_{i=1}^{n} K_h \left(g_\lambda(x) - g_\lambda(x_i) \right). \tag{4.3}$$

Afin d'estimer f avec (4.3), dans (Bolancé, Guillén et Perch Nielsen, 2000), on propose une méthode pour sélectionner les paramètres de transformation. Premièrement, afin de faciliter le choix de la largeur de bande plus tard, on se limite aux valeurs de λ_1 et λ_2 qui génèrent des données transformées de coefficient de dissymétrie nul, c'est-à-dire :

$$\hat{\gamma}_y = \frac{n^{-1} \sum_{i=1}^{n} (y_i - \bar{y})^3}{(n^{-1} \sum_{i=1}^{n} (y_i - \bar{y})^2)^{3/2}} = 0$$

où les y_i sont tels que définis en (4.2) et $\bar{y}$ est la moyenne échantillonnale des y_i.

Par la sélection du vecteur de paramètre λ, on vise à minimiser l'erreur quadratique moyenne intégrée (EQMI) de $\hat{f}(x)$. L'EQMI peut s'estimer par (voir (Wand et Jones, 1995) sur l'EQMI asymptotique) :

$$\frac{5}{4} \left(\mu_2(K) R(K)^2 \right)^{2/5} R(f''_y)^{1/5} n^{-4/5}, \tag{4.4}$$

où $\mu_2(K) = \int_{-\infty}^{\infty} z^2 K(z) dz$ et $R(t) = \int_{-\infty}^{\infty} t^2 dt$. Comme seulement $R(f''_y)$ dépend du vecteur de paramètre λ, il est suffisant de minimiser $R(f''_y)$. Par contre, comme la fonction f''_y est inconnue, l'estimateur proposé par (Hall, Marron et Sheather, 1987) sera utilisé pour estimer $R(f''_y)$. L'estimateur est défini par :

$$\hat{R}(f''_y) = n^{-1}(n-1)^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} c^{-5} K \left(c^{-1}(y_i - y_j) \right)^2, \tag{4.5}$$

où c est la largeur de bande utilisée pour cette estimation. En supposant que f_y est une distribution normale, alors, la largeur de bande c, qui minimise l'erreur quadratique moyenne de $\hat{R}(f''_y)$, est : $\hat{c} = \hat{\sigma}_y \left(\frac{21}{40\sqrt{2}n^2} \right)^{1/13}$, où $\hat{\sigma}_y = \sqrt{n^{-1} \sum_{i=1}^{n} (y_i - \bar{y})^2}$ (voir (Wand et Jones, 1995)). Rappelons que nous nous limitons aux valeurs de λ qui donnent un coefficient de dissymétrie de zéro. Il parait donc acceptable de supposer une distribution normale pour les y_i lors du choix de la largeur de bande.

Une fois que les paramètres de transformation ont été sélectionnés, il reste à déterminer la largeur de bande h à utiliser dans l'équation (4.3). Ici encore, dans (Bolancé, Guillén

et Perch Nielsen, 2000), on propose de supposer une distribution normale pour les y_i et d'utiliser $\hat{h} = 1.059\hat{\sigma}_y n^{-1/5}$.

4.1.1 Résultats

La procédure décrite précédemment fut appliquée aux montants de réclamations pour blessures corporelles. Le graphique 4.1 présente la valeur du coefficient de dissymétrie des données transformées pour différentes valeurs de λ_1 et λ_2. Ensuite, pour chaque couple (λ_1, λ_2) qui donne un coefficient de dissymétrie de zéro la fonction $\hat{R}(f_y'')$ fut évaluée et le couple de paramètres minimisant $\hat{R}(f_y'')$ fut sélectionné. Les valeurs obtenues sont $\lambda_1 = 2716$ et $\lambda_2 = -0.48$.

La figure 4.2 présente la densité estimée à l'aide des données transformées. Remarquez que le but de la transformation des données est d'améliorer (stabiliser) l'estimation de la queue de la distribution. La figure 4.3 compare la queue de la distribution estimée avec des données transformées (à gauche), avec la queue de la distribution estimée par le noyau gamma (à droite). Il est évident que l'estimation de la queue est beaucoup plus stable lorsque les données transformées sont utilisées.

Comme ici on s'intéresse particulièrement à quantifier le risque de réclamations extrêmes, l'estimation de la queue de la distribution est particulièrement importante. L'utilisation de la transformation de famille de puissance décalée semble donc préférable au noyau gamma de la section précédente.

4.2 Transformation avec la distribution Champernowne

Dans cette section, un estimateur à noyau appliqué à une transformation des données est aussi utilisé. Par contre, l'idée derrière la transformation est différente.

L'idée est d'utiliser une estimation paramétrique préliminaire de la distribution f inconnue comme fonction de transformation, puis d'utiliser un estimateur à noyau pour corriger cette estimation préliminaire.

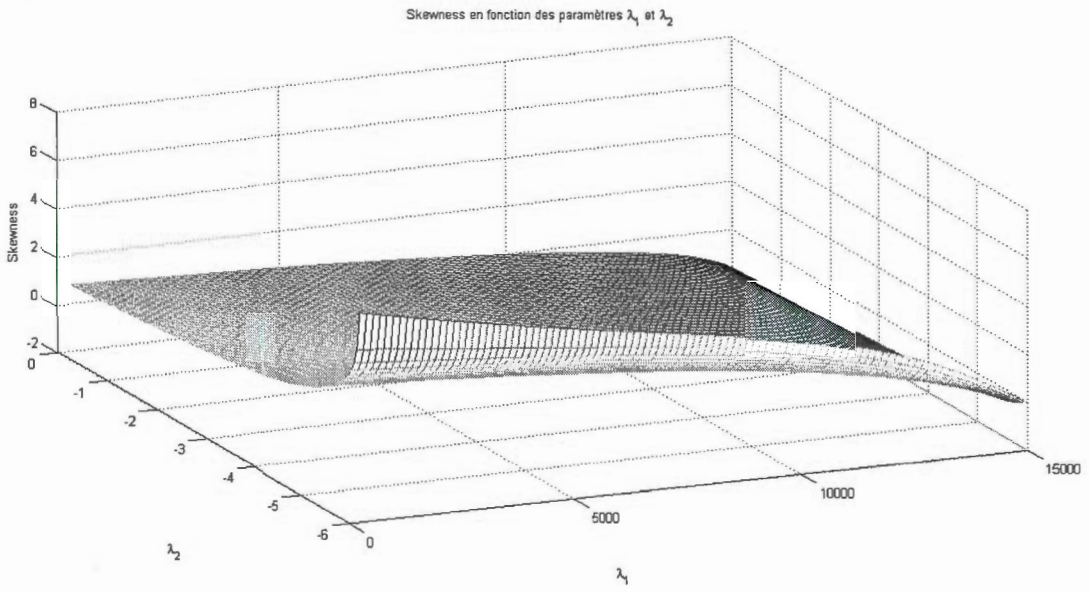

Figure 4.1 Coefficient de dissymétrie en fonction de λ_1 et λ_2 pour les réclamations pour blessures personnelles.

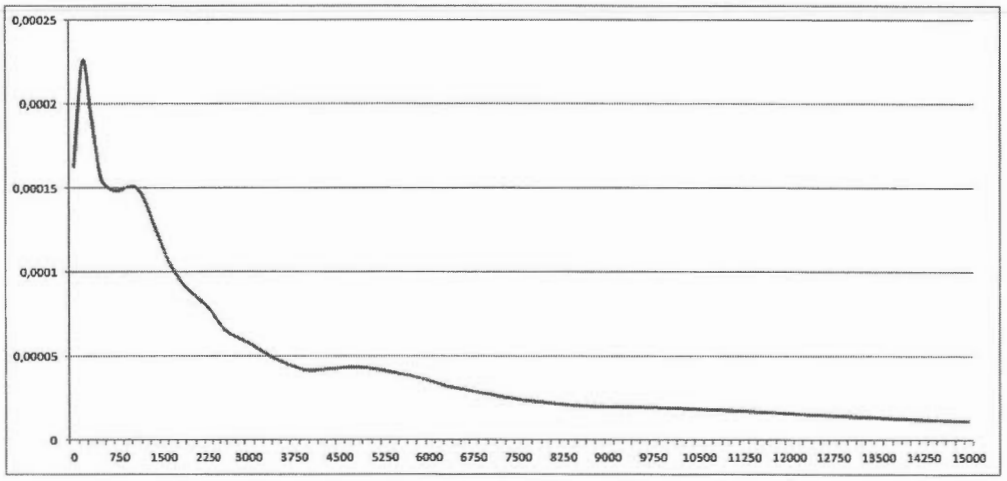

Figure 4.2 Densité estimée avec un noyau Epanechnikov et une transformation de famille de puissance décalée pour les données sur les blessures personnelles.

Dans (Buch-Larsen et al., 2005) on propose d'utiliser la distribution Champernowne

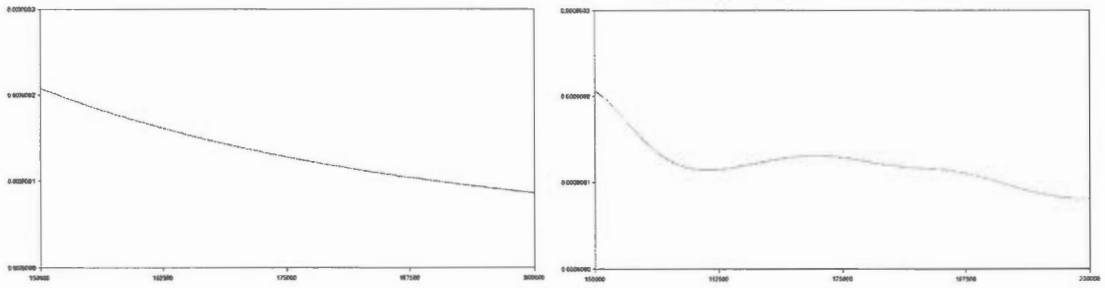

Figure 4.3 Queue de la distribution estimée avec les données transformées (à gauche) et avec noyau gamma (à droite), pour les données blessures personnelles.

comme estimation préliminaire. La fonction de répartition de la Champernowne est :

$$F_{Champ}(x) = \frac{(x+c)^\alpha - c^\alpha}{(x+c)^\alpha + (M+c)^\alpha - 2c^\alpha}.$$

Comme il en a été discuté dans le chapitre 2 sur l'estimation paramétrique, il y a plusieurs méthodes pour estimer les paramètres de la distribution Champernowne. Dans (Buch-Larsen et al., 2005), M est estimé par la médiane empirique. Ensuite, on estime les paramètres α et c par maximum de vraisemblance. Pour plus d'informations sur l'estimation des paramètres de la distribution Champernowne, le lecteur peut se référer au chapitre 2 ou à (Buch-Larsen et al., 2005).

Une fois les paramètres de la distribution estimés, on transforme les données originales (les x_i) en posant : $y_i = \hat{F}_{Champ}(x_i)$. Un estimateur à noyau est ensuite utilisé pour estimer la distribution des y_i.

Nous pouvons faire deux observations sur les données ainsi transformées. Premièrement, les y_i seront nécessairement entre zéro et un (c.-à-d. $y_i \in [0,1], \forall i$). Deuxièmement, si les x_i de départ proviennent effectivement d'une distribution Champernowne, c'est-à-dire que la distribution f inconnue est en fait une distribution Champernowne, alors les y_i auront une distribution uniforme (si on ignore les erreurs d'estimation dues au fait que l'on a seulement un échantillon).

Comme les estimateurs à noyau traditionnels supposent que le domaine de la distribution

estimée est $(-\infty, \infty)$ et qu'ici le domaine est $[0, 1]$, on doit tenir compte du problème de biais aux bornes tel que décrit précédemment à la section 3.2.1.

Dans (Buch-Larsen, 2005) et (Buch-Larsen et al., 2005), on propose d'utiliser un noyau Epanechnikov avec correction pour biais aux bornes, c'est-à-dire :

$$\hat{f}(y) = \frac{1}{nk_y} \sum_{i=1}^{n} K_b(y - y_i),$$

où n est le nombre d'observations, $K_b(z) = K(z/b)/b$ avec b comme largeur de bande, $K(x)$ est la fonction noyau Epanechnikov et k_y est défini par :

$$k_y = \int_{\min(-1, \frac{-y}{b})}^{\min(1, \frac{1-y}{b})} K(u)du.$$

Par contre, dans (Charpentier et Oulidi, 2010), on propose plutôt d'utiliser un noyau bêta modifié pour estimer la densité des y_i. Le noyau bêta modifié est asymétrique et ne souffre pas de biais aux bornes. Cet estimateur se définit comme :

$$\hat{f}_{BM,b}(x) = \frac{1}{n} \sum_{i=1}^{n} K_{B(\rho_{b,0}(x), \rho_{b,1}(x))}(X_i), \tag{4.6}$$

où $K_{B(\alpha,\beta)}(u) = u^{\alpha-1}(1-u)^{\beta-1}/B(\alpha, \beta)$ est la densité d'une distribution bêta de paramètre $\alpha, \beta > 0$, $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$,

$$\rho_{b,0}(x) = \begin{cases} \rho_b(x) \text{ pour } x \in [0, 2b) \\ x/b \text{ pour } x \in [2b, 1] \end{cases},$$

$$\rho_{b,1}(x) = \begin{cases} (1-x)/b \text{ pour } x \in [0, 1-2b] \\ \rho_b(1-x) \text{ pour } x \in (1-2b, 1] \end{cases},$$

$$\rho_b(x) = 2b^2 + \frac{5}{2} - \sqrt{4b^4 + 6b^2 + \frac{9}{4} - x^2 - \frac{x}{b}},$$

$b > 0$ est un paramètre de lissage.

Une fois que la densité des données transformées est estimée, on retrouve la densité estimée pour les données d'origine par :

$$\hat{f}(x) = \frac{\hat{f}_{trans}(\hat{F}_{Champ}(x))}{\hat{F}'^{-1}_{Champ}(\hat{F}_{Champ}(x))}.$$

En résumé, la méthode proposée est la suivante :

1. Calculer les paramètres de la distribution Champernowne à l'aide des données originales (les x_i).
2. Poser $y_i = \hat{F}_{Champ}(x_i)$, où $\hat{F}_{Champ}$ est la fonction de répartition d'une distribution Champernowne, avec les paramètres estimés à l'étape précédente.
3. Estimer la densité des données transformées à l'aide de l'estimateur à noyau bêta modifié, c'est-à-dire $\hat{f}_{trans}(y) = \frac{1}{n}\sum_{i=1}^{n} K_{B(\rho_{b,0}(x),\rho_{b,1}(x))}(x_i)$.
4. Finalement, on retrouve la densité pour les données d'origine en posant :

$$f(x) = \frac{\hat{f}_{trans}(\hat{F}_{Champ}(x))}{\hat{F}^{-1'}_{Champ}(\hat{F}_{Champ}(x))}.$$

4.2.1 Choix du paramètre de lissage

Tout comme pour le choix du paramètre de lissage du noyau gamma, le choix de la largeur de bande b sera fait en minimisant l'erreur quadratique moyenne intégrée. Le calcul de ce paramètre optimal est fait dans (Chen, 1999). Il y est montré que l'erreur quadratique moyenne intégrée pour le noyau bêta modifié est :

$$\begin{aligned} EQMI(\hat{f}_{BM,b}) &= \frac{1}{4}b^2 \int_0^1 \left(x(1-x)f''(x)\right)^2 dx \\ &\quad + \frac{1}{2\sqrt{\pi}} n^{-1} b^{-1/2} \int_0^1 \left(x(1-x)\right)^{-1/2} f(x)dx + o(n^{-1}b^{-1/2} + b^2). \end{aligned} \tag{4.7}$$

La valeur de b qui minimise (4.7) est :

$$b^* = \left[\frac{\frac{1}{2\sqrt{\pi}} \int_0^1 \left(x(1-x)\right)^{-1/2} f(x)dx}{\int_0^1 \left(x(1-x)f''(x)\right)^2 dx} \right]^{2/5} n^{-2/5}. \tag{4.8}$$

Dans (4.8), la densité inconnue f (que l'on tente justement d'estimer) intervient. Donc, afin d'estimer le paramètre de lissage b optimal, on doit avoir une estimation préliminaire de f et de f'' (la dérivée seconde de f prise par rapport à x). Comme il a été vu dans la section 3.2.3, il est possible d'estimer ces quantités à l'aide des estimateurs à noyau. $f''(x)$ sera donc estimé par :

$$\hat{f}''(x) = n^{-1} \sum_{i=1}^{n} K''(x_i).$$

Par contre, comme le noyau bêta modifié n'a pas de dérivée seconde pour certains points entre 0 et 1, le noyau bêta sera utilisé. Le noyau bêta tel que défini dans (Chen, 1999) est :

$$\hat{f}_{B,b}(x) = n^{-1}\sum_{i=1}^{n} K_{x/b+1,(1-x)/b+1}(x_i),$$

où $K_{\alpha,\beta}$ est la fonction de densité d'une loi bêta de paramètre α et β, c'est-à-dire : $K_{\alpha,\beta}(x) = x^{\alpha-1}(1-x)^{\beta-1}\Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$.

On trouve donc comme dérivée seconde :

$$K''_{x/b+1,(1-x)/b+1}(x_i) = \frac{h_1''(x)h_2(x) + 2h_1'(x)h_2'(x) + h_1(x)h_2''(x)}{\Gamma(1/b+1)},$$

où :

$$\begin{aligned}
h_1(x) &= x_i^{x/b}(1-x_i)^{(1-x)/b} \\
h_2(x) &= \Gamma(x/b+1)\Gamma((1-x)/b+1) \\
h_1'(x) &= -\frac{x_i^{x/b}\ln\frac{1-x_i}{x_i}}{b(1-x_i)^{(x-1)/b}} \\
h_2'(x) &= \frac{h_2(x)[\psi(x/b+1) - \psi((1-x)/b+1)]}{b} \\
h_1''(x) &= \frac{x_i^{x/b}(\ln\frac{1-x_i}{x_i})^2}{b^2(1-x_i)^{(x-1)/b}} \\
h_2''(x) &= \frac{h_2'(x)[\psi(x/b+1) - \psi((1-x)/b+1)]}{b} + \frac{h_2(x)[\psi'(x/b+1) + \psi'((1-x)/b+1)]}{b^2}.
\end{aligned}$$

On peut maintenant estimer le paramètre de lissage optimal. Notez que les intégrales dans (4.8) sont effectuées de façon numérique.

4.2.2 Résultats

La procédure décrite ci-haut fut appliquée aux montants de réclamation pour la couverture blessures corporelles (autrui).

La largeur de bande optimale calculée est : $b^* = 0.0517014$. La figure 4.4 présente la densité estimée des données transformées. On remarque que cette densité ne semble pas

uniforme. Par contre, l'estimateur à noyau est justement là pour corriger cette estimation préliminaire.

Finalement, la figure 4.5 présente la densité estimée pour les données d'origines. On remarque que la densité estimée ressemble à celle estimée avec le noyau gamma ou avec la transformation de famille puissance décalée. Cependant cette estimation semble plus lisse. Notez que ceci est peut-être dû à la différence entre les largeurs de bande utilisées.

Comme il a été montré à la section précédente, la transformation de famille puissance décalée stabilise l'estimation de la queue par rapport à l'estimation obtenue avec un noyau gamma. Il est donc naturel de vérifier si cette propriété est conservée par la transformation Champernowne. La figure 4.6 montre que cette propriété est effectivement conservée.

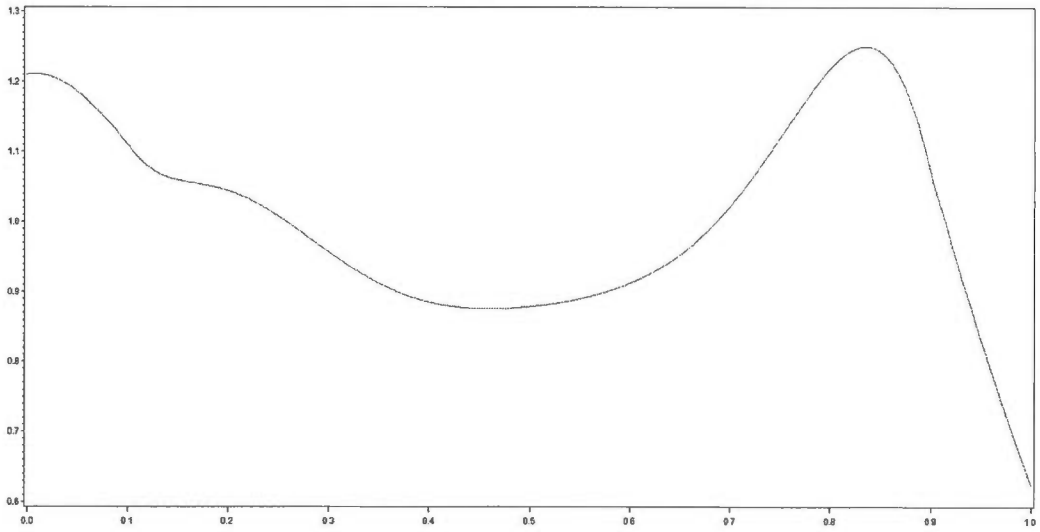

Figure 4.4 Densité estimée des données transformées.

4.2.3 Estimation de quantiles à l'aide des estimateurs à noyau

L'importance de pouvoir estimer les quantiles d'une distribution, pour l'analyse du risque, fut soulignée au chapitre portant sur l'estimation paramétrique. Il est possible d'obtenir une estimation des quantiles à l'aide des estimateurs à noyau.

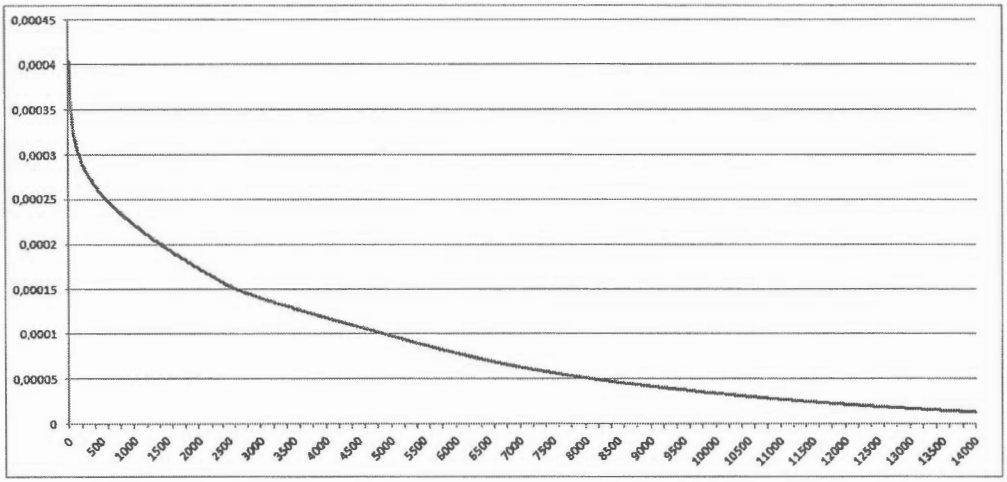

Figure 4.5 Densité estimée des données originales.

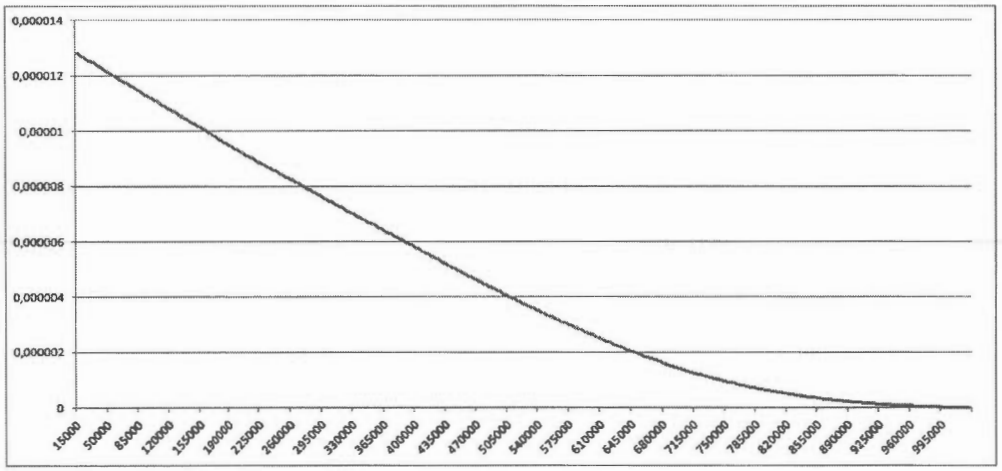

Figure 4.6 Queue de la densité estimée des données originales.

Tout d'abord, il faut définir un estimateur à noyau de la fonction de distribution F_X associé à la densité inconnue f.

$$\hat{F_X}(x) = \int_0^x \hat{f_X}(t)dt \tag{4.9}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\int_0^x K_h(x_i;t)dt, \tag{4.10}$$

où $K_h(x_i;t)$ est une fonction noyau avec une largeur de bande h évaluée au point x_i autour du point t. Cette définition est générale et le noyau bêta modifié défini en (4.6),

voir (Charpentier et Oulidi, 2010), est utilisé. On doit donc calculer l'intégrale

$$\int_0^x K_{B(\rho_{h,0}(t),\rho_{h,1}(t))}(X_i)dt,$$

où la fonction à intégrer est telle que définie en (4.6). Une méthode d'intégration numérique est utilisée pour le calcul de cette intégrale.

Une fois qu'on a la fonction de distribution d'une variable aléatoire X, il suffit de l'inverser pour obtenir la fonction quantile, c'est-à-dire $Q_X(p) = F_X^{-1}(p)$. Par contre, il n'y a pas de formule fermée pour l'estimation des quantiles avec les estimateurs à noyau. On doit donc procéder par inversion numérique, en utilisant le fait que $\hat{F}_n(\hat{Q}_n(p)) = p$, où $\hat{Q}_n(p)$ est la fonction quantile qu'on cherche à estimer. De plus, $\hat{F}_n$ est un estimateur de la fonction de distribution et $0 < p < 1$. Dans ce qui suit $\hat{F}_n$ est obtenu à l'aide d'un estimateur à noyau.

De plus, rappelons que la densité estimée à l'aide de l'estimateur à noyau bêta modifié est en fait la densité sur l'intervalle $(0, 1)$ des données transformées et non la densité des données originales sur l'intervalle $(0, \infty)$.

Soit une transformation H et une variable aléatoire Y, si la fonction H est strictement croissante, alors le $p^{\text{ème}}$ quantile de $H(Y)$ est égal à $H(Q_Y(p))$. Donc l'idée, voir (Charpentier et Oulidi, 2010), est de transformer l'échantillon initial $\{X_1, \ldots, X_n\}$ en un échantillon $\{Y_1, \ldots, Y_n\} = \{H(X_1), \ldots, H(X_n)\}$ prenant des valeurs dans l'intervalle $(0,1)$. Ensuite, on estime la densité des Y_i à l'aide de l'estimateur bêta modifié. Finalement, on obtient les quantiles pour les données originales par $Q_X(p) = H^{-1}(Q_Y(p))$.

La transformation H utilisée est la fonction de distribution d'une loi Champernowne telle que proposée par (Buch-Larsen et al., 2005). Les paramètres de la distribution Champernowne sont estimés par maximum de vraisemblance à l'aide de l'échantillon original tel qu'expliqué plus haut. Comme on souhaite que la distribution des données transformées soit aussi près d'une distribution uniforme que possible, on désire que le graphique de la fonction de distribution estimée soit une droite, c'est-à-dire $F(x) = x$ où $x \in [0, 1]$. La figure 4.7 présente la distribution estimée par la procédure décrite

Tableau 4.1 Comparaison des quantiles empiriques avec ceux estimés

Quantile	Théorique	Empirique
99%	366 650	209 080.35
95%	98 550	88 293.14
90%	53 503	60 213.19

précédemment.

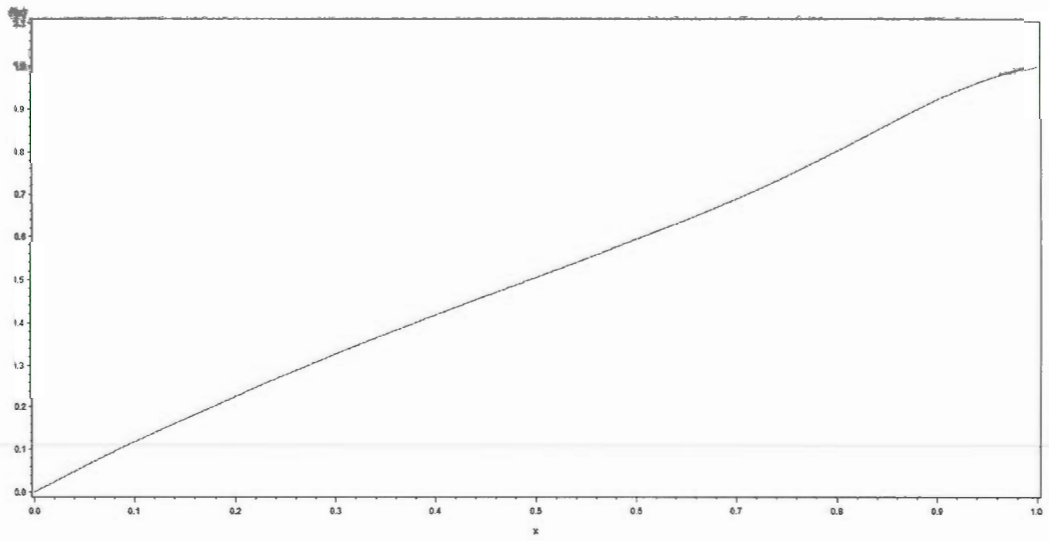

Figure 4.7 Fonction de distribution estimée à l'aide des données transformées.

Il est finalement possible d'estimer les quantiles de la distribution originale, c'est-à-dire les montants de réclamation pour blessures personnelles. Le tableau 4.1 présente les quantiles estimés et leurs contreparties empiriques.

CHAPITRE V

ESTIMATION DE DENSITÉ - THÉORIE DES VALEURS EXTRÊMES

En assurance non-vie, quelques réclamations extrêmes peuvent grandement influencer les résultats d'une compagnie d'assurance. C'est le cas, entre autres, des réclamations pour blessures corporelles infligées à autrui qui sont analysées dans ce mémoire. C'est pourquoi il est important de bien modéliser ces risques extrêmes. Cela est généralement fait à l'aide de la théorie des valeurs extrêmes.

Dans la théorie des valeurs extrêmes, on s'intéresse particulièrement à la distribution au-dessus d'un seuil, c'est-à-dire la distribution de $[X|X > u]$ où X est une variable aléatoire et u un seuil. On s'intéresse aussi à l'estimation de quantiles élevés et à l'estimation de la perte maximale probable.

5.1 Loi du maximum

Théorème 1. *Soit une suite $X_1, \ldots, X_n$ de variables aléatoires de même fonction de répartition F. S'il existe des constantes $a_n \in \mathbb{R}$ et $b_n > 0$ telles que $\frac{X_{(n)}-a_n}{b_n} \rightarrow_{loi} G$, alors G aura une des trois formes suivantes :*

$$G_+(x) = \exp(-(1+\xi x)^{-1/\xi}) \text{ pour } x \geq -1/\xi \text{ et } \xi > 0$$
$$\Lambda(x) = \exp(-\exp(-x)) \text{ pour } x \in \mathbb{R}$$
$$G_-(x) = \exp(-(1+\xi x)^{-1/\xi}) \text{ pour } x \leq -1/\xi \text{ et } \xi < 0$$

Notez que les fonctions de répartition G_+, Λ et G_- correspondent aux lois de Fréchet,

de Gumbel et de Weibull respectivement. De plus, ces trois lois sont des cas particuliers de la loi $GEV_{\mu,\sigma,\xi}$. La loi $GEV_{\mu,\sigma,\xi}$ se définit comme

$$GEV_{\mu,\sigma,\xi}(x) = \begin{cases} \exp(-(1+\xi\frac{x-\mu}{\sigma})^{-1/\xi}), \text{ si } \xi \neq 0 \\ \exp(-\exp(-\frac{x-\mu}{\sigma})), \text{ si } \xi = 0 \end{cases}.$$

5.1.1 Estimation des paramètres de la loi $GEV_{\mu,\sigma,\xi}$

Le problème avec l'estimation de la loi du maximum, c'est que peu importe la taille de l'échantillon, il n'y aura toujours qu'un seul maximum, ce qui rend délicate l'estimation des paramètres.

La méthode proposée dans (Gumbel, 1958) est de subdiviser l'échantillon original de n données en m sous-échantillons de n/m données chacun. Alors, en notant y_i le maximum du $i^{\text{ème}}$ sous-échantillon, on peut écrire la fonction de logvraisemblance de $GEV_{\mu,\sigma,\xi}$ comme :

$$\begin{aligned} l = &- m\ln\sigma - (1+\xi^{-1})\sum_{i=1}^{m}\ln\left(1+\xi\frac{y_i-\mu}{\sigma}\right) \\ &- \sum_{i=1}^{m}\ln\left(1+\xi\frac{y_i-\mu}{\sigma}\right)^{-1/\xi}. \end{aligned}$$

Les montants de réclamation pour blessures corporelles infligées à autrui ont été divisés aléatoirement en 100 blocs de données et 100 maximums furent obtenus. À l'aide de ces 100 maximum, les paramètres de la loi $GEV_{\mu,\sigma,\xi}$ furent estimés par maximum de vraisemblance. Le tableau 5.1 présente les résultats de l'estimation par maximum de vraisemblance.

5.2 Distribution de Pareto généralisée

La loi de Pareto généralisée à deux paramètres telle que définie dans (Cebrian, Denuit et Lambert, 2003) :

$$G_{\xi,\beta}(x) = G_{\xi}\left(\frac{x}{\beta}\right), \tag{5.1}$$

Tableau 5.1 Paramètres estimés de la loi $GEV_{\mu,\sigma,\xi}$

Paramètre	Valeur estimée	Écart-type
μ	195 985.60	15 476.4
σ	137 567.30	20 090.02
ξ	0.9466	0.1216

où $\beta > 0$ et :

$$G_\xi(x) = \begin{cases} 1 - (1+\xi x)^{-1/\xi} & \text{si } \xi \neq 0 \\ 1 - e^{-x} & \text{si } \xi = 0 \end{cases}.$$

Le paramètre ξ est appelé l'index de la distribution.

Soit l'excès au-dessus d'un seuil défini comme $[X - u|X > u]$. On peut montrer que, pour des distributions à queue lourde, l'excès au-dessus d'un seuil peut être traité comme une variable aléatoire de loi Pareto généralisée lorsque le seuil u est suffisamment grand. Dans ce cas, on dit que la queue de la loi appartient au max-domaine d'attraction de G_ξ. Plus spécifiquement, citons le théorème de Pickands-Balkema-de Haanm, tel que donné dans (Denuit et Charpentier, 2005) :

Théorème 2. *La fonction de répartition F appartient au max-domaine d'attraction de G_ξ si, et seulement si,*

$$\lim_{u \to x_F} \sup_{0<x<x_F} \{|\Pr[X - u \leq x|X > u] - G_{\xi,\beta(u)}(x)|\} = 0,$$

où $\beta(u)$ est une fonction positive.

5.2.1 Estimation de ξ

L'estimateur proposé pour ξ est l'estimateur de Hill. Cet estimateur utilise les statistiques d'ordre. L'idée est que si la fonction de survie $S(x) = \Pr[X > x]$ peut s'écrire comme $S(x) = x^{1/\xi}L(x)$, où $L(x)$ est une fonction à variation lente, alors la fonction

quantile s'exprime comme $F^{-1}(1-p) = p^{-\xi}L^*(1/p)$, où :

$$\ln(F^{-1}(1-p)) = -\xi \ln(p) + \ln(L^*(1/p)).$$

Notons qu'un estimateur de $F^{-1}(1-k/(n+1))$ est la statistique d'ordre $X_{(n-k+1)}$ (voir la section 1.1.4). Alors les points

$$(\ln(X_{(n-k+1)}), \ln(k/(n+1))), k = 1, \ldots, m,$$

devraient être sur une droite de pente ξ. L'estimateur de la pente est donc

$$\hat{\xi} = \frac{\frac{1}{m}\sum_{i=1}^{m} \ln(X_{(n-i+1)}) - \ln(X_{(n-m)})}{\frac{1}{m}\sum_{i=1}^{m} \ln(\frac{i}{n+1}) - \ln(\frac{m}{n+1})}.$$

Lorsque m est suffisamment grand, le dénominateur est approximativement égal à 1 et on a donc l'estimateur de Hill :

$$\hat{\xi}_{m,n}^{Hill} = \frac{1}{m}\sum_{i=1}^{m} \ln(X_{(n-i+1)}) - \ln(X_{(n-m)}).$$

En plus de l'estimateur de Hill, deux autres estimateurs sont souvent utilisés dans la littérature. Il y a l'estimateur de Pickands

$$\hat{\xi}_{m,n}^{Pickands} = \frac{1}{\ln 2} \ln\left(\frac{X_{(n-m)} - X_{(n-2m)}}{X_{(n-2m)} - X_{(n-4m)}}\right),$$

et l'estimateur de Dekkers-Einmalh-de Haan

$$\hat{\xi}_{m,n}^{DEdH} = \xi_{m,n}^{H(1)} + 1 - \frac{1}{2}\left(1 - \frac{(\xi_{m,n}^{H(1)})^2}{\xi_{m,n}^{H(2)}}\right)^{-1},$$

où

$$\xi_{m,n}^{H(r)} = \frac{1}{m}\sum_{i=1}^{m-1} (\ln X_{(n-i)} - \ln X_{(n-m)})^r.$$

Pour chacun de ces estimateurs, il faut déterminer une valeur de m appropriée correspondant au nombre de données utilisées dans l'estimation. Des méthodes pour déterminer la valeur de m seront discutées plus bas.

5.3 Détection des queues lourdes

Comme les méthodes statistiques présentées dans ce chapitre supposent que la distribution des données analysées a une queue lourde, il est important de pouvoir vérifier cette hypothèse. Deux méthodes graphiques seront proposées. Ces méthodes sont tirées de (Cebrian, Denuit et Lambert, 2003).

La première méthode utilise l'espérance de l'excès au-dessus d'un seuil, c'est-à-dire $E[X-u|X>u]$. On peut facilement montrer que si la variable aléatoire X est de loi exponentielle de moyenne μ alors $E[X-u|X>u] = E[X] = \mu$. Le graphique de cette fonction sera une droite. Si la distribution de X à une queue légère (sous-exponentielle) le graphique de $E[X-u|X>u]$ aura une pente descendante alors que si au contraire, la queue de la distribution est lourde (surexponentielle), la pente sera ascendante. Comme ici la fonction $E[X-u|X>u]$ n'est pas connue explicitement, la moyenne empirique des $[x_i - u|x_i > u]$ sera utilisée comme estimation de $E[X-u|X>u]$.

La deuxième méthode est le graphique quantile-quantile (*QQ-plot*) exponentiel. Ce graphique compare les quantiles empiriques aux quantiles théoriques d'une loi exponentielle. Son interprétation est simple. Si les données proviennent d'une loi exponentielle, le graphique sera sur la droite, par contre, si la queue est plus lourde (plus légère) la courbe sera au-dessus (en dessous) de la droite.

Les figures 5.1 et 5.2 montrent les deux méthodes graphiques pour les montants de réclamation pour blessures corporelles infligées à autrui. Dans les deux cas, on voit clairement que la distribution des x_i a une queue lourde.

5.4 Choix du seuil

Afin d'estimer les paramètres de la loi de $[X-u|X>u]$, on doit déterminer un seuil u. Ce seuil doit être suffisamment élevé pour que la distribution de $[X-u|X>u]$ soit une Pareto généralisée, mais pas trop élevé pour pouvoir avoir suffisamment de données pour que l'estimation des paramètres soit stable. Ces deux buts sont en conflit et il faut

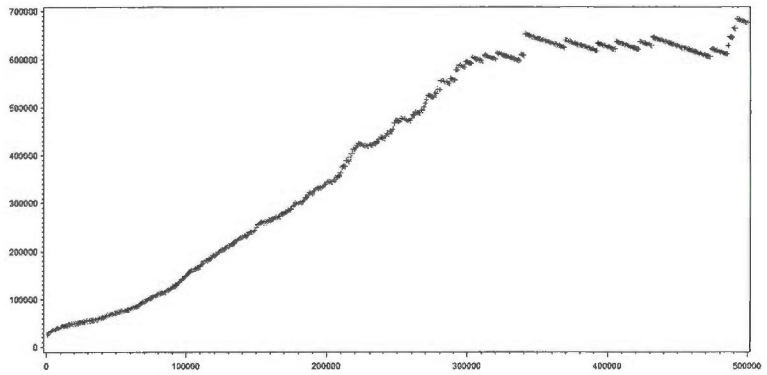

Figure 5.1 Espérance de l'excès sur l'ordonné et le seuil sur l'abscisse.

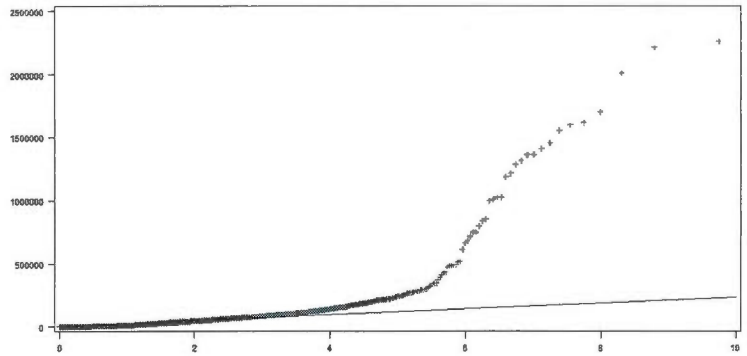

Figure 5.2 Quantile-Quantile exponentiel.

trouver le meilleur compromis.

Ici encore, les auteurs de (Cebrian, Denuit et Lambert, 2003) proposent d'utiliser des méthodes graphiques pour l'estimation du seuil. Trois méthodes sont proposées :

1. **Moyenne empirique de l'excès.** On peut montrer que lorsque X suit une loi de Pareto généralisée de fonction $G_{\xi,\beta}$, l'espérance de l'excès est une fonction linéaire du seuil u, c'est-à-dire : $E[X - u|X > u] = \frac{\beta}{1-\xi} + \frac{\xi}{1-\xi}u$. Donc, l'idée est de déterminer visuellement à l'aide d'un graphique le seuil u à partir duquel la moyenne empirique de l'excès devient approximativement linéaire. La figure 5.1 présente cette méthode.

2. **Graphique de l'index de la distribution de Pareto généralisée.** Cette mé-

thode utilise le fait que, si une variable aléatoire X a la fonction de répartition $G_{\xi,\beta}$, la variable aléatoire $[X - u|X > u]$ aura $G_{\xi,\beta+\xi u}$ comme fonction de répartition. L'idée est de calculer l'estimateur $\hat{\xi}^{Hill}_{m,n}$, ou un autre estimateur de ξ, pour des seuils u de plus en plus petits et de vérifier visuellement à partir de quel seuil $\hat{\xi}$ semble se stabiliser. La figure 5.3 présente cette méthode appliquée aux données de réclamations pour blessures corporelles. Notez que l'abscisse représente le seuil du plus grand au plus petit. Par exemple, l'ordonnée associée à la valeur 500 correspond aux estimateurs pour un seuil u égal à la 500^{e} plus grande réclamation.

En observant la figure 5.3, on remarque la difficulté d'appliquer cette méthode et son côté un peu arbitraire. Un peu avant la donnée numérotée 500 et ce jusqu'à la fin du graphique, on sent qu'il y a une légère tendance à la baisse pour l'estimateur de Hill et une légère tendance à la hausse pour l'estimateur Deckers-Einmalh-de Haan. Pour les données entre 300 et 600, on peut considérer les estimations comme stables, ou du moins sans tendance claire. Pour les estimations en deçà de 200, l'estimation devient particulièrement instable, car le nombre de données utilisées pour l'estimation devient (trop) petit. Dans ce mémoire, l'auteur considère que l'estimation se stabilise à la donnée 350.

3. **Graphique de Gertensgarbe.**

   Soit une série de différences $\Delta_i = x_{(i)} - x_{(i-1)}$. Le début de la région des points extrêmes sera détecté par un changement dans la série des Δ_i.

   Afin d'identifier le point de changement dans la série des Δ_i, on définit une série de U_i normalisée comme : $U_i = \frac{U_i^* - \frac{i(i-1)}{4}}{\sqrt{\frac{i(i-1)(i+5)}{72}}}$ où $U_i^* = \sum_{k=1}^{i} n_k$ et n_k est le nombre de valeurs dans $\Delta_1, \ldots, \Delta_k$ inférieur à Δ_k. Ensuite, la même procédure est appliquée à la série des différences dans l'ordre inverse, c'est-à-dire à la série $\Delta_n, \ldots, \Delta_1$ plutôt qu'à la série $\Delta_1, \ldots, \Delta_n$. Le point d'intersection des deux séries de U_i est considéré comme le début des observations extrêmes. La figure 5.4 présente le résultat obtenu lorsqu'on applique cette méthode aux données d'assurance pour blessures personnelles.

   Notez que cette procédure revient à effectuer des tests de Mann-Kendall succes-

sivement sur les n premières données de la serie des différences, soit $\Delta_1, \ldots, \Delta_n$, pour $n = 2, \ldots, N$ où N représente le nombre total de différences Δ_i.

En général, le seuil estimé par la méthode du graphique de Gertensgarbe est plus petit que le seuil estimé par les autres méthodes présentées ici. C'est effectivement le cas pour les données utilisées dans ce travail. En effet, le seuil estimé par cette méthode est plus petit que les autres seuils estimés (voir le tableau 5.2).

Comme chacune de ces méthodes donne une valeur approximative du seuil u, il est recommandé d'utiliser les trois méthodes simultanément (voir (Cebrian, Denuit et Lambert, 2003)). Le tableau 5.2 résume le résultat obtenu pour chacune des méthodes discutées précédemment. Comme on peut le constater, deux méthodes offrent une estimation similaire, mais la troisième donne une estimation plus petite que les deux autres. On peut donc être persuadé que les données au-dessus de 100000 se comportent comme une Pareto généralisée.

Il est utile de rappeler que le choix du seuil est un compromis entre une valeur qui nous assure que les données supérieures au seuil se comportent comme une Pareto généralisée et une valeur qui laisse suffisamment d'observations supérieures au seuil pour avoir une bonne estimation des paramètres. De plus, si le vrai seuil est u^*, alors pour tous les seuils $u \geq u^*$ les données supérieures à u auront aussi un comportement Pareto généralisé avec le même index ξ. Donc, de choisir un seuil plus grand que u^* est une erreur moins grande que de choisir un seuil $u < u^*$. Dans le cas des données d'assurance étudiées dans ce travail, il y a plus de 350 réclamations supérieures à 100000$ ce qui est suffisant pour avoir confiance aux estimations obtenues en utilisant un seuil $u = 100000$.

5.5 Estimation de quantile avec la loi de Pareto généralisée

Dans les sections précédentes, un seuil fut choisi afin de séparer les réclamations extrêmes, celles au-dessus du seuil, des autres réclamations. Il est maintenant possible d'estimer les paramètres de la distribution de Pareto généralisée associés aux données d'assurance, pour blessures personnelles, en utilisant un seuil de 100000$. Les estimateurs

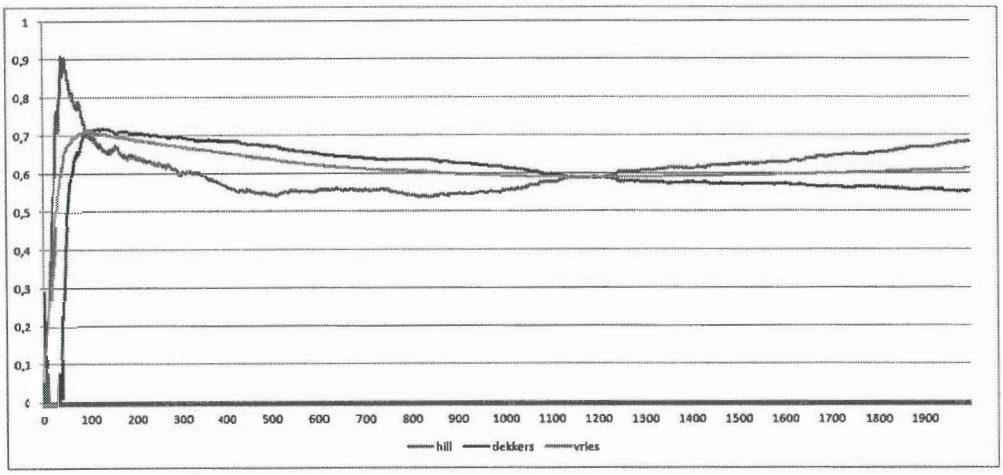

Figure 5.3 Graphique de l'index de la distribution GPD. L'abscisse représente le seuil, et l'ordonnée, la valeur estimée de l'index ξ.

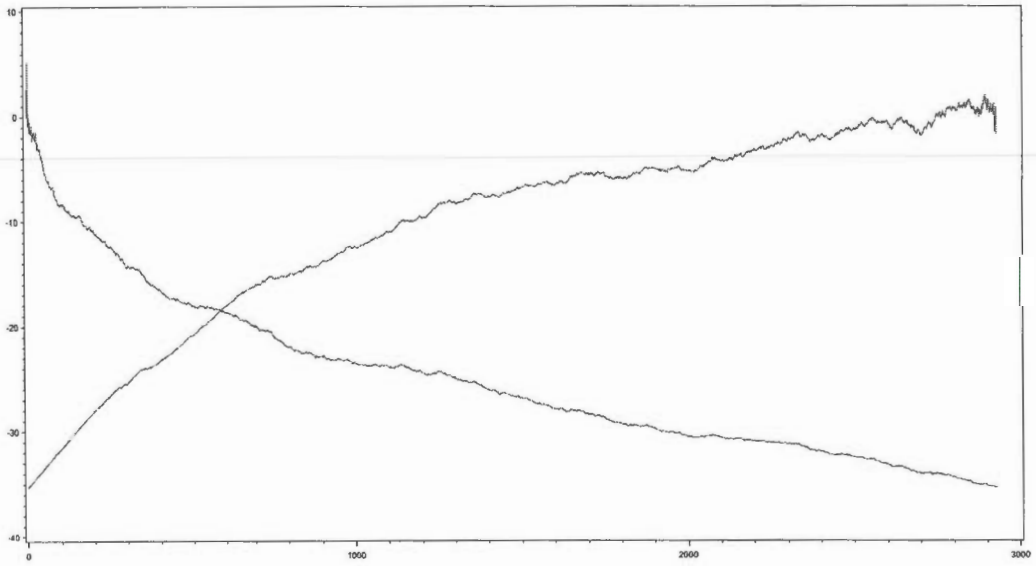

Figure 5.4 Graphique de Gertensgarbe. L'indice des données en absisse et la valeur des U_i correspondants en ordonnée.

obtenus par maximum de vraisemblance sont présentés au tableau 5.3.

Une fois les paramètres estimés, il est aisé de trouver les quantiles élevés. Pour une loi

Tableau 5.2 Estimation du seuil pour la distribution Pareto généralisée

Méthode	Valeur estimée du seuil
Moyenne empirique de l'excès	≈ 110000
Graphique de l'index de la GPD	≈ 100000
Graphique de Gertensgarbe	≈ 78635

de Pareto généralisée, les quantiles sont donnés par $q_p = G^{-1}_{\xi,\sigma}(p)$, c'est-à-dire

$$G_{\xi,\sigma}(q_p) = 1 - (1 + \xi\frac{q_p}{\sigma}) = p$$
$$\Leftrightarrow q_p = \left((1-p)^{-\xi} - 1\right)\frac{\sigma}{\xi},$$

où q_p représente le p^{e} quantile. Notez que la distribution de Pareto généralisée estimée s'applique seulement conditionnellement à ce que la réclamation soit plus élevée que le seuil u. Donc, les quantiles estimés par cette distribution sont des quantiles conditionnels, c'est-à-dire sachant que la réclamation est supérieure à u. Par contre, l'objectif est d'estimer les quantiles élevés non conditionnels.

Soit $F(x)$ la distribution non conditionnelle d'une variable aléatoire X et $S(x) = 1-F(x)$ la fonction de survie, alors

$$P[X - u \geq x | X > u] = \frac{S(u+x)}{S(u)},$$

en approximant $P[X - u \geq x|X > u]$ avec une Pareto généralisée on obtient :

$$S(u+x) = S(u)(1 - G_{\xi,\sigma}(x)).$$

Maintenant, le terme $S(u)$ peut être estimé par n_u/n où n_u et n représentent le nombre d'observations supérieures à u et le nombre d'observations totales respectivement. On obtient :

$$S(u+x) = \frac{n_u}{n}(1 - G_{\xi,\sigma}(x)),$$

ou pour une valeur $y > u$:

$$S(y) = \frac{n_u}{n}(1 - G_{\xi,\sigma}(y-u)).$$

Tableau 5.3 Paramètres estimés d'une loi Pareto généralisée pour les données d'assurance avec un seuil de 100000$

Paramètre	Valeur estimée
$\hat{\xi}$	0.8082
$\hat{\sigma}$	46591

On peut alors trouver la fonction des quantiles de la distribution non conditionnels qui s'exprime comme :

$$q_p = u + \left(\left(\frac{n}{n_u}(1-p) \right)^{-\xi} - 1 \right) \frac{\sigma}{\xi}. \tag{5.2}$$

La figure 5.5 compare les quantiles théoriques estimés obtenus avec l'équation (5.2), en utilisant les paramètres estimés avec un seuil $u = 100000$ tels que présentés dans le tableau 5.3, aux quantiles empiriques. On remarque que les quantiles estimés sont près des quantiles empiriques pour les quantiles inférieurs à 1500000. Puis les quantiles estimés deviennent beaucoup plus élevés que les quantiles empiriques par la suite.

5.6 Détection automatique du seuil

La procédure décrite précédemment permet d'utiliser le théorème de Pickands-Balkema-de Haan. Par contre, les méthodes graphiques sont subjectives. La méthode du graphique de Gertensgarbe repose sur des heuristiques.

L'article de (Tung et Li, 2009) traite du choix automatique du seuil dans l'utilisation du théorème de Pickands-Balkema-de Haan pour la modélisation de risques extrêmes. Une approche utilisant le produit maximum des espacements (traduction libre de *maximum product of spacings*) est proposée. Cette approche offre plusieurs avantages : elle n'est pas subjective, elle modélise l'ensemble du domaine de la distribution et non seulement les valeurs extrêmes et par conséquent, utilise l'ensemble des données dans la modélisation et non seulement les données supérieures à un seuil. De plus, cette méthode converge vers le vrai seuil lorsque le nombre d'observations tend vers l'infini.

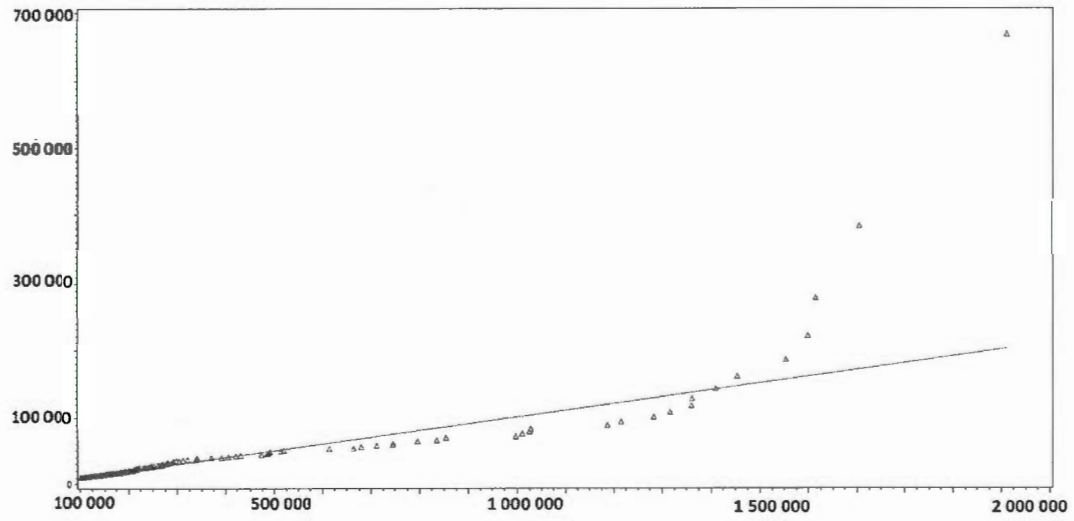

Figure 5.5 Graphique quantiles-quantiles, quantiles théoriques en ordonnée et quantiles empiriques en abscisse.

5.6.1 Description de la méthode

Comme il a été mentionné, le but est d'utiliser la théorie des valeurs extrêmes, donc le théorème de Pickands-Balkema-de Haan, mais tout en modélisant l'ensemble de la distribution et non pas seulement les valeurs supérieures à un seuil u.

Soit une variable aléatoire X, on peut écrire $\Pr[X \leq x]$ comme

$$\Pr[X \leq x] = \Pr[X \leq x \cap X \leq u] + \Pr[X \leq x \cap X > u]$$
$$\Leftrightarrow \Pr[X \leq x] = \Pr[X \leq x \cap X \leq u] + \Pr[X > u] \Pr[X \leq x | X > u].$$

Autrement dit, la probabilité que X soit inférieure à un certain x se décompose en deux parties : soit la probabilité que X soit inférieure à x et à u et la probabilité que X soit inférieure à x, mais supérieure à u.

Si la distribution de X est dans le domaine d'attraction de la Pareto, alors, pour une valeur de u suffisamment élevée, on peut estimer le terme de droite avec une Pareto généralisée. Le terme de gauche est modélisé avec une distribution paramétrique L ayant le vecteur de paramètres θ. Le choix de la distribution L est laissé à l'utilisateur. Pour

des exemples de distributions, voir le chapitre 2 sur l'estimation paramétrique.

On a donc le modèle à seuil suivant :

$$F_X(x;\theta,\xi,\sigma) = \begin{cases} L(x;\theta), & x \leq u \\ L(u;\theta) + (1 - L(u;\theta))G_{\xi,\sigma,u}(x), & x > u \end{cases}. \tag{5.3}$$

Pour utiliser ce modèle, il faut d'abord choisir une distribution L, puis on doit estimer tous les paramètres. Ces paramètres sont : le vecteur θ dont la longueur dépend de la distribution L choisie, les deux paramètres de la Pareto généralisée, soient ξ et σ et finalement on doit estimer le seuil u.

La méthode d'estimation des paramètres θ, ξ et σ proposée dans (Tung et Li, 2009) est décrite à la section 5.6.2. Pour ce qui est du choix de la distribution L, plusieurs distributions sont testées et celle qui semble offrir le meilleur ajustement est sélectionnée.

5.6.2 Produit maximum des espacements

La méthode proposée dans (Tung et Li, 2009) utilise l'ensemble des données disponibles pour estimer tous les paramètres du modèle à seuil, c'est-à-dire $\hat{u}$, le vecteur $\hat{\theta}, \hat{\xi}$ et $\hat{\sigma}$.

Avant de discuter du choix de $\hat{u}$, une définition du produit maximum des espacements et de la statistique de Moran sera faite. Cette méthode suppose que le seuil u est connu, ou une estimation $\hat{u}$ de ce seuil.

Soit une estimation $\hat{u}$ du seuil u, alors les paramètres $\hat{\theta}, \hat{\xi}$ et $\hat{\sigma}$ sont sélectionnés en maximisant la fonction :

$$M(\theta,\xi,\sigma) = \sum_{i=1}^{n+1} \ln(F(x_{(i)}) - F(x_{(i-1)})), \tag{5.4}$$

où $F(x_{(0)}) = 0$, $F(x_{(n+1)}) = 1$ et $x_{(1)} \leq \ldots \leq x_{(n)}$ sont les statistiques d'ordre. Notez que si $x_{(j)} = x_{(j-1)}$ alors on remplace $F(x_{(j)}) - F(x_{(j-1)})$ par $f(x_j)$. La fonction définie en (5.4) est appelée statistique de Moran. Elle est maximisée par des méthodes numériques (voir (Wong et Li, 2006)) et dans cette section la méthode de Nelder-Mead sera utilisée pour trouver les paramètres maximisant la statistique de Moran.

L'algorithme de Nelder-Mead est une procédure d'optimisation de fonction non linéaire. Pour une description de l'algorithme, voir (Olsson et Nelson, 1975). Cet algorithme est disponible dans le logiciel statistique R.

5.6.3 Choix du seuil u

L'estimation du seuil $\hat{u}$ est obtenue en choisissant tour à tour chacun des $x_{(i)}$ comme candidat du seuil $\hat{u}$, c'est-à-dire que l'on maximise la fonction (5.4) pour chaque seuil $\hat{u}_i = x_{(i)}$. Finalement, on sélectionne le seuil $\hat{u}_i$ qui maximise la fonction (5.4) maximisée comme estimation de $\hat{u}$.

Cette estimation de u est en quelque sorte un maximum de maximum. En effet, on se retrouve avec une valeur de la statistique de Moran maximisée avec $\hat{u}_i = x_{(i)}$ pour chaque x_i, puis on choisit le seuil qui donne la plus grande statistique de Moran.

Les auteurs de (Tung et Li, 2009) présentent ensuite deux théorèmes.

Théorème 3. *Sous certaines conditions de régularité, $\hat{u}$ est super cohérent (traduction libre de super consistent), avec $\hat{u} - u = O(n^{-1})$.*

Théorème 4. *Si $\xi > -1/2$ l'estimation $\hat{\xi}$ et $\hat{\sigma}$ de ξ et σ obtenue en maximisant* (5.4) *est asymptotiquement normale avec*

$$\left[\sqrt{k}(\hat{\xi} - \xi), \sqrt{k}(\hat{\sigma} - \sigma)\right] \rightarrow^{D} N(0, V),$$
$$V = (1 + \xi) \begin{pmatrix} 1 + \xi & -\sigma \\ -\sigma & 2\sigma^2 \end{pmatrix}.$$

Le lecteur intéressé trouvera les preuves de ces théorèmes dans (Tung et Li, 2009).

Le premier théorème confirme que le seuil obtenu par cette méthode converge bel et bien vers le vrai seuil, ce qui rend cette méthode particulièrement attrayante. Le deuxième théorème montre que non seulement le seuil converge vers le vrai seuil, mais les paramètres de la Pareto généralisée, $\hat{\xi}$ et $\hat{\sigma}$, convergent eux aussi vers leur vraie valeur.

5.6.4 Application aux données d'assurance

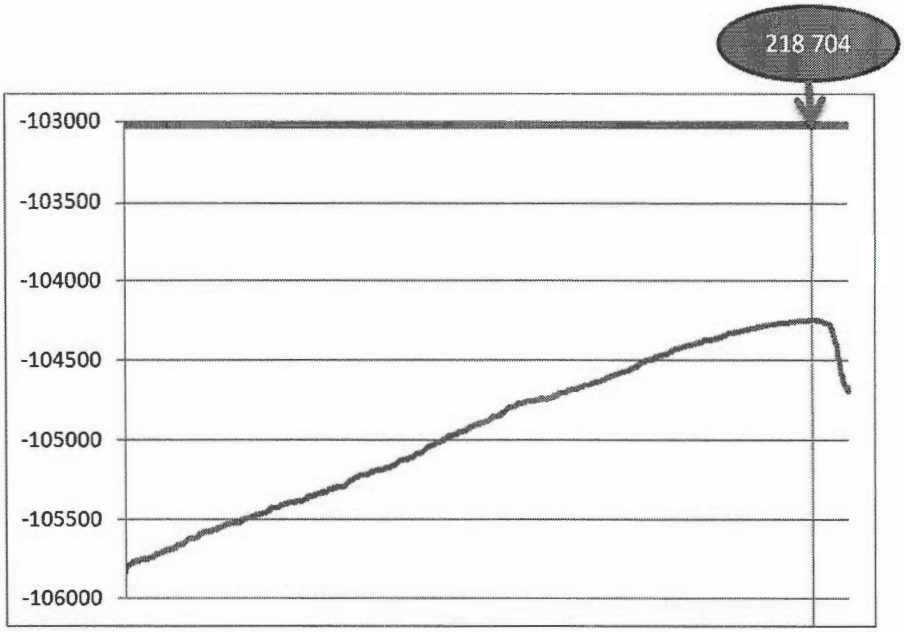

Figure 5.6 Valeur de la statistique de Moran maximisée pour des seuils de plus en plus élevés avec L étant une distribution gamma.

La méthode décrite précédemment fut appliquée aux montants de réclamation pour blessures personnelles. Bien que cette méthode permette un choix automatique du seuil u, il faut tout de même choisir une distribution L pour les données inférieures au seuil u. Tel qu'indiqué plus tôt, l'utilisateur essaie différentes distributions puis sélectionne celle qui offre le meilleur ajustement.

Trois distributions seront utilisées tour à tour pour la fonction cumulative L de l'équation (5.3), soit la distribution gamma, la distribution de Weibull et la distribution bêta prime. Toutes ces distributions sont présentées au chapitre 2.

Les figures 5.6 à 5.8 présentent la valeur maximisée de la statistique de Moran définie en (5.4) pour des seuils $\hat{u}$ de plus en plus élevés et pour différentes distributions L. Pour chacune de ces figures, l'abscisse représente le seuil $\hat{u}_i$ et l'ordonnée représente la valeur de la statistique de Moran maximisée en utilisant $u = \hat{u}_i$.

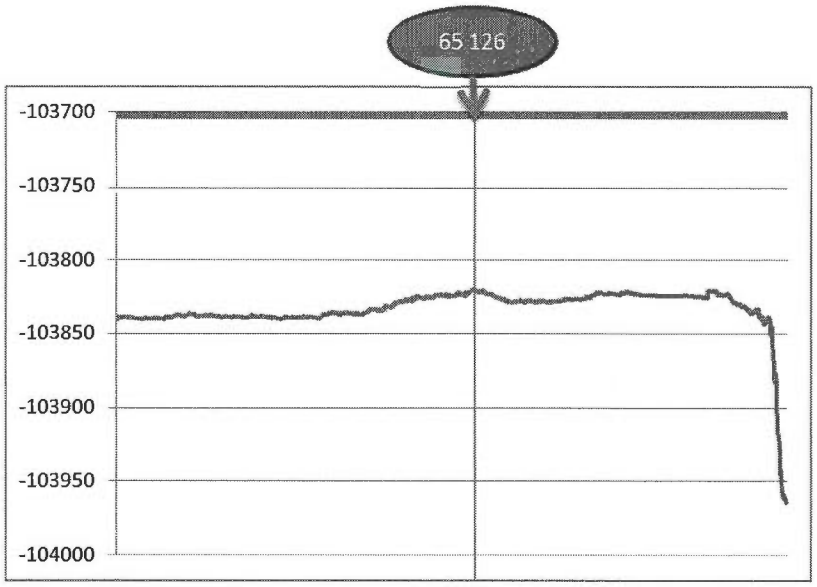

Figure 5.7 Valeur de la statistique de Moran maximisée pour des seuils de plus en plus élevés avec L étant une distribution Weibull.

Pour les distributions gamma et bêta prime, on voit que la valeur de la statistique de Moran maximisée augmente de façon assez linéaire jusqu'à son maximum. Tandis que pour la distribution Weilbull, la statistique de Moran maximisée aux différents seuils est assez stable.

Le tableau 5.4 résume les résultats observés sur les figure 5.6 à 5.8 et présente aussi les valeurs de $\hat{\xi}$ et $\hat{\sigma}$. En se fiant uniquement à la statistique de Moran, la distribution Weibull serait le meilleur choix parmi ces distributions.

Les figures 5.9 à 5.11 présentent les graphiques quantiles-quantiles pour les différentes distributions estimées, soit la distribution gamma, Weibull et bêta prime respectivement. Les quantiles théoriques sont en ordonnée et les quantiles empiriques sont en abscisse. La ligne droite représente le cas où l'ajustement est parfait, c'est-à-dire que les quantiles théoriques et empiriques sont les mêmes. Lorsque les points sont sous la droite c'est que la distribution estimée sous-estime les quantiles empiriques et vice-versa lorsque les points sont au-dessus de la droite. On remarque que la distribution gamma sous-estime

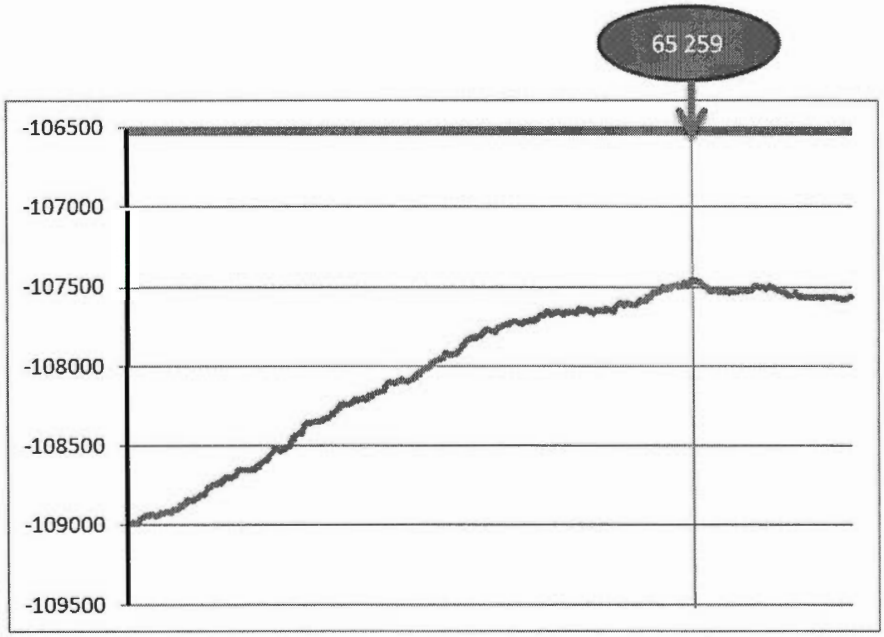

Figure 5.8 Valeur de la statistique de Moran maximisée pour des seuils de plus en plus élevés avec L étant une distribution bêta prime.

les quantiles empiriques sur l'ensemble du domaine. La distribution bêta prime, pour sa part, surestime systématiquement les quantiles empiriques, tandis que la distribution Weilbull se colle particulièrement aux quantiles inférieurs à 500 000$.

Pour conclure la comparaison entre les différents choix pour la distribution de L de l'équation (5.3), les statistiques de Kolmogorov-Smirnov furent calculées et sont présentées au tableau 5.5. Ici encore, la distribution de Weilbull offre le meilleur ajustement parmi les trois distributions étudiées.

En plus d'avoir la statistique de Moran la plus élevée et la statistique de Kolmogorov-Smirnov la plus basse, la distribution de Weibull présente également l'estimation du seuil μ la plus basse. Cela est avantageux, car la distribution de Pareto généralisée est alors estimée à l'aide d'un nombre plus élevé de données.

Tableau 5.4 Résultats

Distribution L	seuil $\hat{u}$	Statistique de Moran maximum	$\hat{\xi}$	$\hat{\sigma}$
Gamma	218704.24	-104246	0.8487	167353
Weibull	65126.48	-103820	0.6964	29966
Bêta prime	65259.99	-107454	0.6803	30423

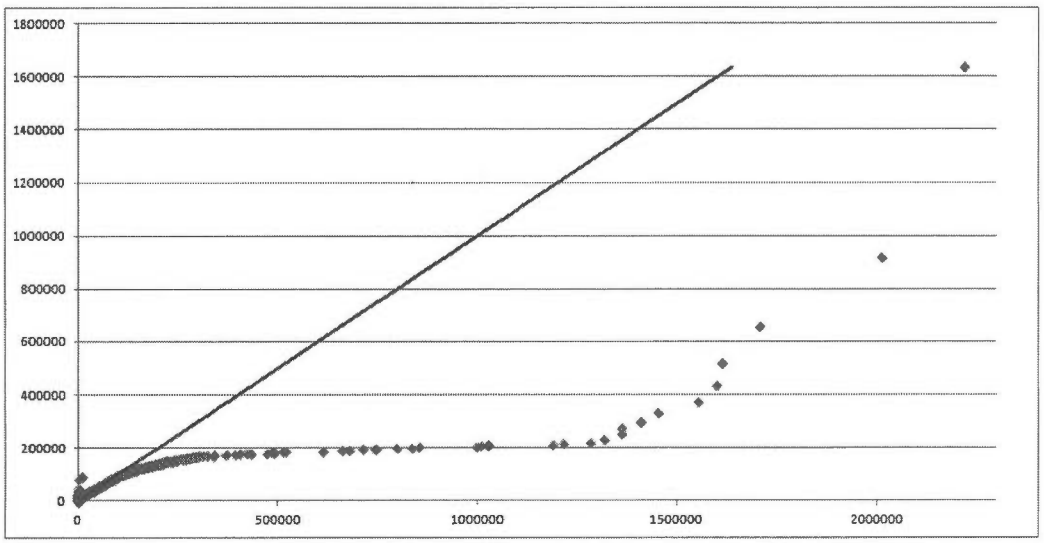

Figure 5.9 Graphique quantile-quantile pour la loi gamma. Les quantiles empiriques sont en abscisse et les quantiles théoriques sont en ordonnée.

5.6.5 Les avantages de cette approche

L'approche décrite précédemment a plusieurs avantages par rapport à celle décrite à la section 5.4. Premièrement, l'ensemble du domaine de la distribution est modélisé, ce qui permet d'utiliser l'ensemble de l'échantillon de données disponibles. Deuxièmement, dans les cas des données étudiées ici, c'est-à-dire le montant des réclamations pour blessures personnelles, le seuil estimé est plus bas ce qui permet un ajustement plus fiable de la queue de la distribution. Finalement, il y a le théorème 3 qui nous assure que le seuil estimé converge vers le vrai seuil. Ce dernier point est ce qui distingue le plus cette approche de celle de la section 5.4 qui utilise des heuristiques pour sélectionner le seuil.

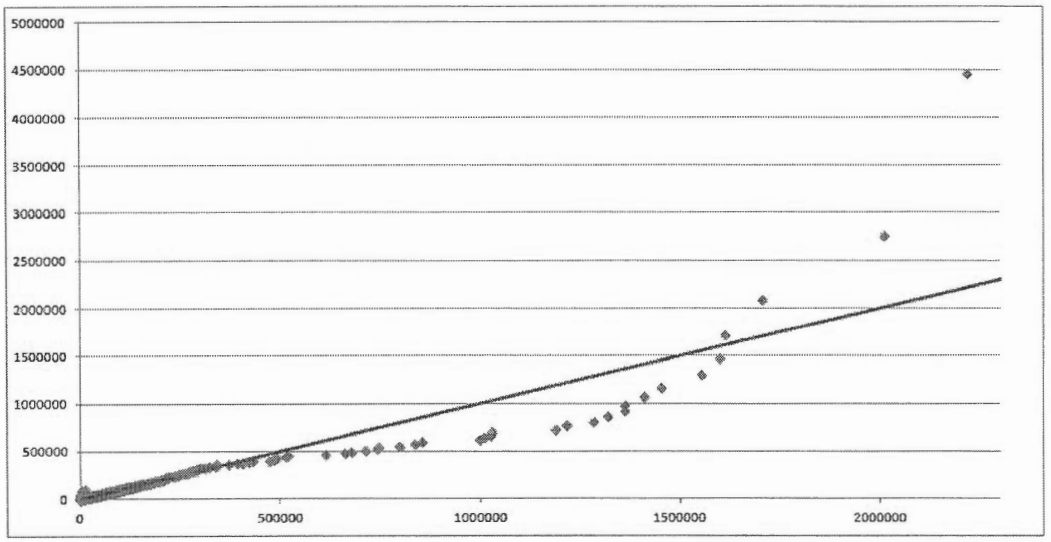

Figure 5.10 Graphique quantile-quantile pour la loi Weibull. Les quantiles empiriques sont en abscisse et les quantiles théoriques sont en ordonnée.

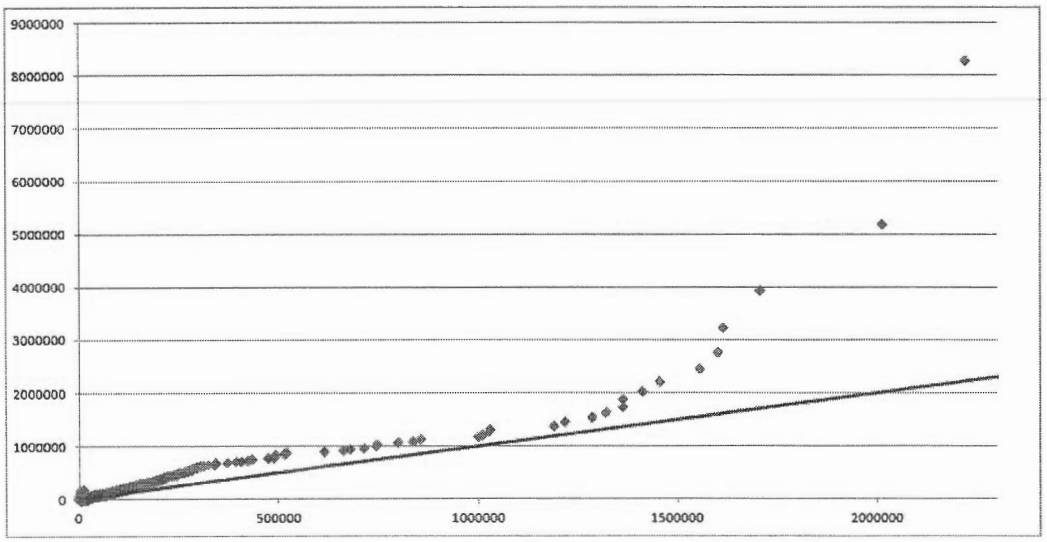

Figure 5.11 Graphique quantile-quantile pour la loi bêta prime. Les quantiles empiriques sont en abscisse et les quantiles théoriques sont en ordonnée.

Le tableau 5.6 présente les quantiles estimés par la première méthode, c'est-à-dire la méthode décrite à la section 5.4 et dans (Cebrian, Denuit et Lambert, 2003), ainsi que

Tableau 5.5 Kolmogorov-Smirnov

Distribution L	Statistique de Kolmogorov-Smirnov
Gamma	0.0819
Weibull	0.0375
Bêta prime	0.1741

Tableau 5.6 Comparaison des quantiles

	Première approche	Deuxième approche	Empirique
99.9%	1 023 583	874 901	1 331 419
99%	194 957	193 694	209 080.35
97.5%	115 122	112 795	115 649

les quantiles estimés par la deuxième méthode, soit celle décrite à la section 5.6 et dans (Wong et Li, 2006). En observant le tableau 5.6, on remarque qu'il y a peu de différence entre les deux méthodes. La différence ne se fait sentir que pour les quantiles très élevés.

CHAPITRE VI

POUVOIR DE PRÉDICTION DES DIFFÉRENTS MODÈLES

Dans les chapitres précédents, plusieurs méthodes d'estimations de fonction de densité furent présentées. Dans ce chapitre, une comparaison du pouvoir de prédiction de ces modèles sera effectuée. L'idée est d'ajuster les modèles à l'aide d'une partie de la base de données, puis de tester les modèles sur le reste de la base de données, c'est-à-dire avec des données que les modèles n'ont jamais vues avant.

La base de données utilisée dans ce travail contient des montants de réclamations pour des accidents qui ont eu lieu dans les années 2003 à 2007 inclusivement. Dans ce chapitre, les données des années 2003 à 2005 inclusivement seront utilisées pour ajuster les modèles. Les données des années 2006 et 2007 seront utilisées pour évaluer les modèles. Bien que les modèles développés dans ce mémoire ne sont pas des modèles temporels, il est intéressant de voir si une compagnie à la fin de 2005 aurait pu utiliser cette information pour estimer les quantiles des réclamations de 2006 et 2007.

Notez que dans ce chapitre seulement les montants de réclamations pour blessures personnelles seront étudiés.

6.1 Estimateur à noyau

Aux chapitres 3 et 4 concernant les estimateurs à noyau, plusieurs estimateurs furent développés. Dans cette section, l'estimateur à noyau bêta, utilisé sur une transformation Champernowne des données, sera utilisé. Pour une description de cet estimateur, voir la

Tableau 6.1 Paramètres estimés

Paramètre	Valeur estimée
α	0.9395
M	4 670.93
c	1.7292

Tableau 6.2 Résultats

Distribution L	seuil $\hat{u}$	Statistique de Moran maximum	$\hat{\xi}$	$\hat{\sigma}$
Weibull	94 378	-103 817	0.8554	39 552

section 4.2.

Afin de tester le pouvoir de prédiction de cet estimateur, les paramètres de la transformation Champernowne furent estimés à partir des données des années 2003 à 2005 inclusivement. Les paramètres sont présentés au tableau 6.1.

6.2 Théorie des valeurs extrêmes

Dans le chapitre 5 sur les valeurs extrêmes, deux méthodes furent étudiées. Dans cette section la deuxième méthode utilisant la technique du produit maximum des espacements, voir section 5.6, sera utilisée avec la distribution de Weibull comme distribution L.

La figure 6.1 et le tableau 6.2 résument les résultats de l'estimation effectuée à l'aide des données des années 2003 à 2005 inclusivement. Afin de savoir comment interpréter la figure 6.1 voir la section 5.6.4.

6.3 Estimation des quantiles

Une comparaison du pouvoir de prédiction des quantiles des deux modèles décrits plus haut sera faite dans cette section. Le lecteur est référé à la section 1.1.5 pour une dis-

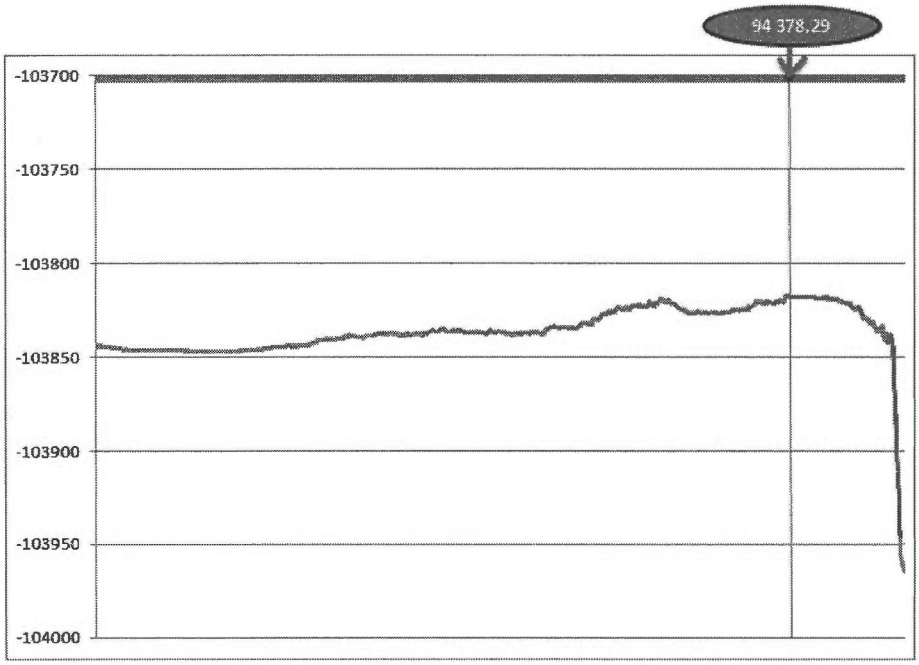

Figure 6.1 Statistique de Moran maximisée pour des seuils de plus en plus élevés.

cussion de l'importance des quantiles dans un contexte de gestion de risques.

Les figures 6.2 et 6.3 présentent des graphiques quantiles-quantiles. Les quantiles empiriques pour les années 2006 et 2007 sont en abscisse, et les quantiles théoriques estimés avec les données de 2003 à 2005 sont en ordonnée.

On remarque que l'estimateur à noyau surestime les quantiles inférieurs à 100 000, mais qu'il colle mieux les quantiles empiriques par la suite. Par contre, le modèle utilisant la théorie des valeurs extrêmes semble offrir une meilleure prédiction des quantiles sur l'ensemble du domaine. Autre point à remarquer, le quantile le plus élevé estimé par la théorie des valeurs extrêmes est de 2 820 000, celui estimé par l'estimateur à noyau est de 1 788 000 tandis que le quantile empirique correspondant est de 1 614 000. Ce quantile correspond à la deuxième plus grande réclamation pour blessures personnelles pour les années 2006 et 2007, la plus grande étant de 2 011 000. Donc, pour les quantiles très élevés la théorie des valeurs extrêmes est plus conservatrice que l'estimateur à noyau.

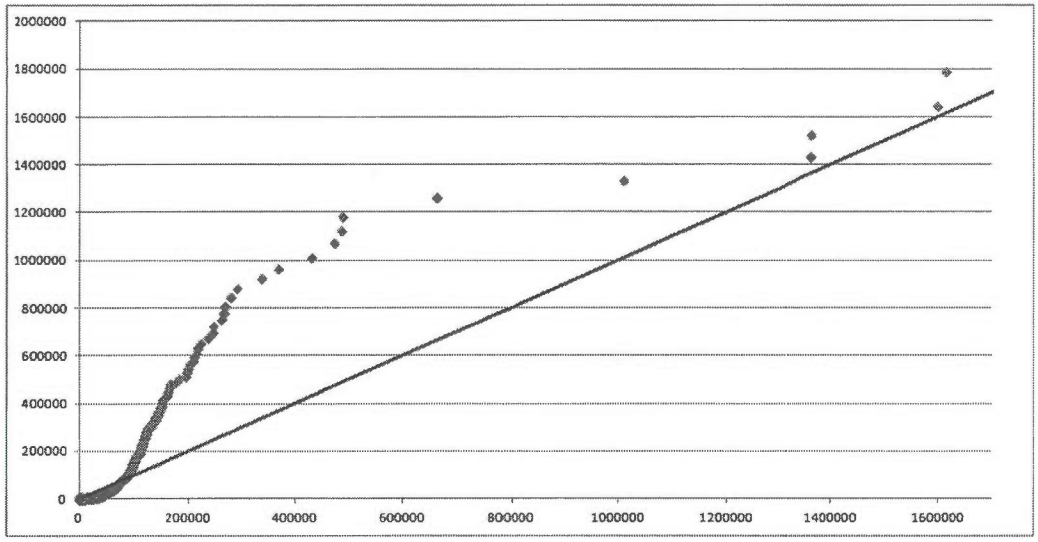

Figure 6.2 Graphique quantile-quantile. Les quantiles empiriques sont en abscisse et les quantiles théoriques sont en ordonnée.

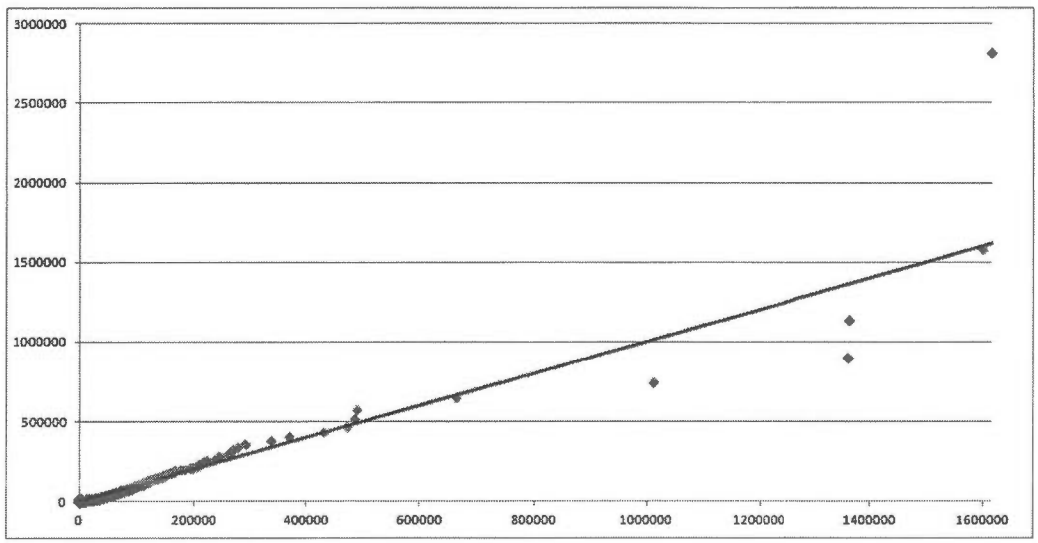

Figure 6.3 Graphique quantile-quantile. Les quantiles empiriques sont en abscisse et les quantiles théoriques sont en ordonnée.

6.4 Espérance de l'excès

Une autre mesure de risque importante en gestion des risques est l'espérance de l'excès au-dessus d'un seuil ω, soit $E[X - \omega | X > \omega]$. Pour une discussion sur l'utilisation de

l'espérance de l'excès en actuariat voir (Dowd et Blake, 2006) ou (Denuit et al., 2005). Dans cette section le pouvoir de prédiction de l'espérance de l'excès de chaque modèle sera comparé pour différents seuils.

L'espérance de l'excès au-dessus de ω empirique se calcule comme suit :

$$E[X-\omega|X>\omega] = \frac{\sum_{x_i>\omega}(x_i-\omega)}{N_\omega},$$

où N_ω est le nombre d'observations supérieures à ω. L'espérance de l'excès empirique est calculée à l'aide des données des années 2006 et 2007.

Pour le modèle utilisant la théorie des valeurs extrêmes, on peut utiliser les propriétés de la loi de Pareto généralisée, tant que $\omega > u$. En effet, si $X-u>u$ suit une Pareto généralisée de paramètres ξ et σ alors la variable $(X-\omega)$ suivra aussi une Pareto généralisée de paramètre ξ et $\sigma^* = \sigma + \xi(\omega-u)$ si $\omega > u$, voir (Denuit et Charpentier, 2005). L'espérance de l'excès est donnée par :

$$E[X-\omega|X>\omega] = \frac{\sigma+\xi(\omega-\mu)}{1-\xi}. \tag{6.1}$$

Il suffit ensuite de remplacer les paramètres de (6.1) par les paramètres estimés à la section 6.2.

Pour le modèle utilisant l'estimateur semi-paramétrique, il n'y a pas de formule fermée pour l'espérance de l'excès. En utilisant le fait que, pour une variable aléatoire continue X de densité f, l'on ait

$$f(X=x|X>u) = \begin{cases} \frac{f(x)}{\int_u^\infty f(x)dx} & \text{si } x>u \\ 0 & \text{sinon} \end{cases},$$

on peut écrire

$$E[X-\omega|X>\omega] = \int_\omega^\infty \frac{f(x)}{\int_\omega^\infty f(t)dt}dx. \tag{6.2}$$

En remplaçant f dans (6.2) par l'estimateur à noyau décrit à la section 6.1 et en utilisant une méthode numérique pour résoudre les intégrales on trouve une estimation de l'espérance de l'excès.

Tableau 6.3 Comparaison des espérance de l'excès pour différents seuils ω

ω	Estimateur à noyau	Théorie des valeurs extrêmes	Empirique
100 000	274 343	306 785	134 865
250 000	665 061	1 194 129	504 375
500 000	1 080 769	2 673 036	874 425

Tableau 6.4 Statistiques de Kolmogorov-Smirnov

Modèle	Statistique
Estimateur à noyau	0.1898
Théorie des valeurs extrêmes	0.1217

Le tableau 6.3 présente les résultats obtenus. Comme pour les quantiles élevés, les deux modèles surestiment l'espérance de l'excès.

6.5 Statistique de Kolmogorov-Smirnov

Afin de comparer le pouvoir de prédiction de l'estimateur à noyau à celui de la théorie des valeurs extrêmes, la statistique de Kolmogorov-Smirnov fut calculée pour chacun des modèles. Notez que les modèles furent calibrés avec les données des années 2003 à 2005 et que les statistiques furent calculées avec les données des années 2006 et 2007. Pour une description de la statistique de Kolmogorov-Smirnov voir la section 1.1.8.

Le tableau 6.4 présente les statistiques obtenues. La théorie des valeurs extrêmes offre une statistique plus basse que l'estimateur à noyau bêta avec une transformation Champernowne, ce qui suppose un meilleur ajustement.

6.6 Choix du modèle

À la lumière de ces tests, il n'y a pas de conclusion sans équivoque pour le choix du modèle. Tout d'abord les graphiques quantiles-quantiles montrent que le modèle utilisant

la théorie des valeurs extrêmes de la section 6.2 estime mieux l'ensemble du domaine que l'estimateur à noyau de la section 6.1. Ceci est confirmé par le test de Kolmogorov-Smirnov.

Par contre, les deux modèles offrent une bonne prédiction des quantiles élevés. Pour l'espérance de l'excès, l'estimateur à noyau est le modèle qui surestime le moins les valeurs empiriques. La théorie des valeurs extrêmes peut être vue comme plus conservatrice que l'estimateur à noyau.

CONCLUSION

Au cours des différents chapitres de ce mémoire, plusieurs méthodes furent proposées pour modéliser la distribution des montants de réclamations en assurance automobile. Ces montants de réclamations sont associés à différentes couvertures. Dans la base de données étudiée dans ce mémoire, six couvertures étaient présentes, soit : collisions, responsabilité civile, vol, tous risques, blessures corporelles et blessures personnelles.

Il fut rapidement constaté que certaines couvertures se ressemblent entre elles, et que d'autres sont très différentes. Il fut également remarqué que les couvertures dont les descriptions se ressemblent ont aussi une distribution similaire. Par exemple, les couvertures collisions et responsabilité civile couvrent toutes deux les dommages causés à un véhicule suite à un accident. Et, sans grande surprise, les distributions des coûts des réclamations pour ces deux couvertures se ressemblent aussi.

Estimation paramétrique

Il était incontournable, dans une étude sur l'estimation de la densité des montants de réclamations en assurance automobile, de commencer par une approche purement paramétrique. Premièrement, dans l'esprit du rasoir d'Occam, les modèles les plus simples devraient être préférés. Deuxièmement, ces méthodes plus simples peuvent servir de comparaison avec d'autres modèles, souvent plus complexes.

Cinq distributions ont été utilisées à tour de rôle : la loi de Pareto, la loi bêta prime, la loi gamma, la loi de Weilbull et la loi de Champernowne. Ces distributions ont été choisies à cause de leurs importances dans la littérature actuarielle. La distribution gamma sous-estime clairement le risque associé aux données d'assurance étudiées, tandis que la distribution Pareto surestime ce risque. Pour les distributions Champernowne et bêta prime, l'estimation des quantiles élevés semble plus plausible bien qu'encore loin

des quantiles empiriques.

Estimateurs à noyau

Dans ces chapitres, différentes méthodes d'estimation de densité à l'aide d'estimateur à noyau furent discutées et analysées.

Tout d'abord, les estimateurs à noyau traditionnel furent introduits (voir le tableau 3.1). Cette classe d'estimateurs est particulièrement intéressante lorsque nous n'avons pas d'indices *a priori* sur la densité à estimer. Comme ces estimateurs ne font pas d'hypothèses sur la forme de la densité à estimer, ils sont dits non paramétriques. Il est généralement accepté que le choix du noyau n'a pas de grands impacts sur l'estimation obtenue. Donc l'utilisateur n'a qu'à spécifier une largeur de bande. Par contre, ce choix de largeur de bande ne va pas de soi, (Silverman, 1986) discute par ailleurs de différentes méthodes pour le choix de la largeur de bande.

Bien que le choix de la fonction noyau à utiliser soit considéré comme moins important lorsqu'on estime une densité ayant comme domaine $(-\infty, \infty)$, ce choix devient primordial lorsque le domaine de la distribution a au moins une borne finie, telle que $(0, \infty)$ ou $(0, 1)$ par exemple. Dans ces cas, les noyaux symétriques traditionnels souffrent tous du problème de biais aux bornes. C'est pourquoi le noyau gamma fut introduit par (Chen, 2000). Ce noyau asymétrique n'a pas la même forme si l'on estime la densité près de 0 ou si l'on estime la densité loin de 0. Cet estimateur est spécifiquement conçu pour estimer une densité ayant un domaine $(0, \infty)$, ce qui le rend particulièrement intéressant pour la modélisation de la distribution des pertes en assurance. Dans ce chapitre, les propriétés asymptotiques de l'estimateur à noyau gamma furent développées ainsi que le choix d'une largeur de bande optimale. Ensuite, la procédure fut appliquée aux montants de réclamations pour blessures personnelles en assurance automobile (les données utilisées sont décrites dans le chapitre sur les données).

Avec le noyau gamma, on peut estimer la distribution des réclamations sans le problème de biais aux bornes des noyaux symétriques traditionnels. Par contre, un problème de-

meure lorsque la distribution estimée présente des queues lourdes, c'est-à-dire dans le cas de risques extrêmes. En effet, le noyau gamma décrit dans ce chapitre devient instable pour l'estimation de la queue de la distribution. Ce problème est majeur puisqu'en général on s'intéresse justement à l'estimation de la queue de la distribution. C'est pourquoi l'idée de transformer les données et d'utiliser un estimateur à noyau sur les données transformées fut introduite. Deux transformations furent étudiées dans ce chapitre.

Le problème d'instabilité dans l'estimation de la queue vient principalement du fait qu'on utilise une largeur de bande constante sur l'ensemble du domaine de la distribution. Par contre, comme il y a moins de données dans la queue de la distribution, il serait logique d'augmenter la largeur de bande utilisée, afin d'englober plus de données, pour l'estimation de la queue. C'est ce que visent (Bolancé, Guillén et Nielsen, 2003) en proposant d'utiliser la famille de puissance décalée comme transformation.

La transformation avec la famille de puissance décalée a plusieurs buts simultanés. Premièrement, bien que la largeur de bande utilisée pour estimer la densité des données transformées reste constante, pour les données originales cela revient à utiliser une largeur de bande qui tend vers l'infini, lorsque le point estimé tend vers l'infini. Deuxièmement, cette transformation, disons H, est définie comme $H : (0, \infty) \to \mathbb{R}$. On peut donc utiliser un estimateur à noyau symétrique traditionnel sans problème de biais aux bornes. Finalement, les paramètres de la transformation sont choisis de manière à ce que la distribution des données transformées soit le plus près possible d'une distribution normale. Ce qui simplifie le choix de la largeur de bande.

Finalement, une méthode semi-paramétrique fut proposée. L'idée est d'utiliser une transformation $H : (0, \infty) \to (0, 1)$, où H est une estimation paramétrique de la fonction de distribution des données originales, disons les X_i où $i = 1, \ldots, n$. Le but est d'obtenir un échantillon $\{Y_1, \ldots, Y_n\} = \{H(X_1), \ldots, H(X_n)\}$ le plus uniforme possible, puis d'estimer la densité des données transformées, les Y_i, à l'aide d'un estimateur à noyau. Comme le domaine des Y_i est $(0, 1)$ un noyau bêta modifié, tel que défini dans (Charpentier et Oulidi, 2010), fut utilisé. Selon (Buch-Larsen et al., 2005), la distribution

Champernowne est une bonne distribution à utiliser comme estimation préliminaire. La distribution Champernowne avec ses trois paramètres offre suffisamment de flexibilité et est dans le domaine d'attraction de la Pareto généralisée, ce qui en fait donc un bon choix lorsque l'on soupçonne la distribution à estimer d'avoir une queue lourde. Dans ce chapitre, une largeur de bande optimale pour le noyau bêta modifié fut développée et une méthodologie complète pour obtenir la fonction de densité, la fonction de distribution et les quantiles d'une distribution inconnue fut présentée.

En conclusion, ce chapitre montre que les estimateurs à noyau offrent une solution de rechange à l'utilisation de la théorie des risques extrêmes, et ce, en évitant d'avoir à estimer un seuil pour distinguer les données extrêmes des autres.

Théorie des valeurs extrêmes

Dans le chapitre 5, la distribution des réclamations pour blessures personnelles fut de nouveau étudiée (pour une description des données, voir le chapitre 1). Cette fois-ci, la théorie des valeurs extrêmes fut utilisée.

À première vue, cette méthode peut ressembler à l'approche paramétrique. En effet, on suppose une distribution puis on estime ses paramètres par maximum de vraisemblance, mais la ressemblance s'arrête là. Premièrement, la théorie des valeurs extrêmes modélise seulement la queue de la distribution, contrairement à l'approche paramétrique qui suppose une loi pour l'ensemble du domaine de la distribution. Cependant, la force de la théorie des valeurs extrêmes est le théorème de Pickands-Balkema-de Haan. Ce théorème nous assure que si la distribution étudiée, disons la distribution d'une variable aléatoire X, remplit certaines conditions, la distribution de $[X-u|x>u]$ pour un u suffisamment élevé sera une Pareto généralisée. C'est une situation complètement différente à l'approche paramétrique où l'on cherche, souvent par essais et erreurs, une loi paramétrique qui convient aux données étudiées.

La théorie des valeurs extrêmes s'applique en trois grandes étapes qui furent toutes présentées à tour de rôle dans ce chapitre. Les deux premières étapes découlent directement

du théorème de Pickands-Balkema-de Haan. La troisième est simplement l'utilisation de la distribution estimée aux étapes précédentes.

La première étape consiste à s'assurer que les données étudiées ont bel et bien une distribution à queue lourde et qu'il est donc approprié d'utiliser la théorie des valeurs extrêmes. Pour ce faire, deux méthodes graphiques furent présentées. Dans les deux cas, la distribution des réclamations pour blessures personnelles avait clairement les caractéristiques d'une distribution à queue lourde.

Une fois suffisamment convaincu que l'on travaille avec des données présentant des valeurs extrêmes, pour appliquer le théorème de Pickands-Balkema-de Haan, il ne reste qu'à trouver un seuil u au-delà duquel les données se comportent comme une Pareto généralisée. Encore une fois, des méthodes graphiques furent proposées pour estimer le seuil u. Dans le cas des données étudiées dans ce chapitre, ces méthodes montrent qu'un seuil $\hat{u} = 100000$ est adéquat.

Une fois le seuil sélectionné, il ne reste qu'à estimer les paramètres de la distribution de Pareto généralisée. Dans ce cas, la méthode du maximum de vraisemblance fut utilisée.

Finalement, lorsque le seuil et les paramètres sont estimés, il est alors possible d'estimer les quantiles élevés de la distribution étudiée. Par contre, seulement les quantiles élevés peuvent être estimés, car la théorie des valeurs extrêmes modélise seulement les données au-dessus d'un seuil et non l'ensemble du domaine de la distribution.

Par la suite, une méthode de détection automatique du seuil fut introduite. Cette méthode offre le double avantage de converger vers le vrai seuil lorsque le nombre de données tend vers l'infini et de modéliser l'ensemble du domaine. En effet, ce modèle utilise une distribution pour les données inférieures à un seuil, cette distribution est choisie par l'utilisateur, et la distribution de Pareto généralisée pour les données au-dessus de ce seuil.

Comparaison des modèles

Dans le dernier chapitre, une comparaison entre les principaux modèles développés dans les chapitres précédents est faite. Le but étant de déterminer le modèle ayant le meilleur pouvoir de prédiction.

Le premier modèle testé est l'estimateur à noyau semi-paramétrique, avec un noyau bêta et une transformation Champernowne, telle que définie à la section 4.2. Le second modèle est la théorie des valeurs extrêmes avec détection automatique du seuil tel que défini à la section 5.6.

La comparaison des modèles donne un avantage à la théorie des valeurs extrêmes, qui estime mieux l'ensemble du domaine.

Recherche future

Dans ce mémoire seulement les données sur les montants de réclamations furent utilisées. Dans la pratique, l'assureur possède une foule d'autres informations sur une réclamation. Par exemple, il connait le sexe et l'âge de l'assuré, la marque et la couleur de la voiture impliquée, etc. Toutes ces informations sont appelées variables exogènes.

La suite naturelle de ce mémoire serait donc de chercher à incorporer ces variables exogènes dans les modèles développés dans ce mémoire. Il est déjà connu que l'on peut utiliser les variables exogènes dans l'application du théorème de Pickands-Balkema-de Haan, voir par exemple (Cebrian, Denuit et Lambert, 2003). Il serait intéressant d'essayer de combiner l'utilisation des variables exogènes avec la méthode de détection automatique du seuil présenté à la section 5.6.

BIBLIOGRAPHIE

Akaike, H. 1974. « A new look at the statistical model identification », *IEEE Transactions on Automatic Control*, vol. 19, p. 716–723.

Bolancé, C., M. Guillén, et J. P. Nielsen. 2003. « Kernel density estimation of actuarial loss functions », *Insurance : Mathematics and Economics*, vol. 32, no. 1, p. 19–36.

Bolancé, C., M. Guillén, et J. Perch Nielsen. 2000. « Kernel density estimation of actuarial loss functions », *University of Aarhus, Aarhus School of Business, Department of Business Studies*, no. 00-4.

Bouezmarni, T., A. El Ghouch, et M. Mesfioui. 2011. « Gamma kernel estimators for density and hazard rate of right-censored data », *Journal of Probability and Statistics*, vol. 2011.

Buch-Larsen, T. 2005. « Claims cost estimation of large insurance losses », *SSRN eLibrary.*

Buch-Larsen, T., J. P. Nielsen, M. Guillén, et C. Bolancé. 2005. « Kernel density estimation for heavy-tailed distributions using the champernowne transformation », *Statistics*, vol. 39, no. 6, p. 503–516.

Casella, G. et R. L. Berger. 1990. *Statistical Inference.* Duxbury Press.

Cebrian, A. C., M. Denuit, et P. Lambert. 2003. « Generalized pareto fit to the Society of Actuaries large claims database », *North American Actuarial Journal*, vol. 7, p. 18–36.

Charpentier, A. et A. Oulidi. 2010. « Beta kernel quantile estimators of heavy-tailed loss distributions », *Statistics and Computing*, vol. 20, p. 35–55.

Chen, S. 2000. « Probability density function estimation using gamma kernels », *Annals of the Institute of Statistical Mathematics*, vol. 52, no. 3, p. 471–480.

Chen, S. X. 1999. « Beta kernel estimators for density functions », *Computational Statistics & Data Analysis*, vol. 31, no. 2, p. 131 – 145.

Choi, S. C. et R. Wette. 1969. « Maximum likelihood estimation of the parameters of the gamma distribution and their bias », *Technometrics*, vol. 11, no. 4, p. pp. 683–690.

Cummins, J., G. Dionne, J. B. McDonald, et B. Pritchett. 1990. « Applications of the gb2 family of distributions in modeling insurance loss processes », *Insurance : Mathematics and Economics*, vol. 9, no. 4, p. 257 – 272.

Denuit, M. et A. Charpentier. 2005. *Mathématiques de l'assurance non-vie : Tarification et provisionnement.* Coll. « Economie et statistiques avançées ». Economica.

Denuit, M., J. Dhaene, M. Goovaerts, et R. Kaas. 2005. *Actuarial Theory for Dependent Risks : Measures, Orders and Models.* Wiley.

Dowd, K. et D. Blake. 2006. « After VaR : The theory, estimation, and insurance applications of quantile-based risk measures », *Journal of Risk & Insurance*, vol. 73, no. 2, p. 193–229.

Gumbel, E. 1958. *Statistics of Extremes.* Coll. « Dover books on mathematics ». Columbia University Press.

Hall, P., J. S. Marron, et S. Sheather. 1987. « Estimation of integrated squared density derivatives », *Stat. Prob. Lett*, p. 109–115.

Kleiber, C. et S. Kotz. 2003. *Statistical Size Distributions in Economics and Actuarial Sciences.* Coll. « Wiley Series in Probability and Statistics ». Wiley.

Klugman, S., H. Panjer, et G. Willmot. 2004. *Loss models.* Coll. « Wiley series in probability and statistics ». Hoboken, NJ : Wiley Interscience, 2. ed édition.

Markowitz, H. 1970. *Portfolio selection.* Coll. « Monograph / Cowles Foundation for Research in Economics at Yale University », no 16. New Haven, Conn. [u.a.] : Yale Univ. Press, 2. printing édition.

Olsson, D. M. et L. S. Nelson. 1975. « The Nelder-Mead simplex procedure for function minimization », *Technometrics*, vol. 17, no. 1, p. pp. 45–51.

Parzen, E. 1962. « On estimation of a probability density function and mode », *The Annals of Mathematical Statistics*, vol. 33, no. 3, p. pp. 1065–1076.

Rosenblatt, M. 1956. « Remarks on some nonparametric estimates of a density function », *The Annals of Mathematical Statistics*, vol. 27, no. 3, p. pp. 832–837.

Ross, S. 2002. *A First course in probability.* Upper Saddle River, NJ : Prentice Hall, 6. ed édition.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis (Chapman & Hall/CRC Monographs on Statistics & Applied Probability).* Chapman and Hall/CRC.

Stephens, M. A. 1970. « Use of the Kolmogorov-Smirnov, cramer-von mises and related statistics without extensive tables », *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 32, no. 1, p. pp. 115–122.

Treynor, J. 1961. *Toward a Theory of Market Value of Risky Assets.*

Tung, T. S. et W. K. Li. 2009. « A threshold approach for peaks-over-threshold modeling using maximum product of spacings », *Statistica Sinica*, vol. 20, p. 1257–1272.

Wand, M. et C. Jones. 1995. *Kernel smoothing.* Coll. « Monographs on Statistics and Applied Probability ». Chapman and Hall.

Wasserman, L. 2003. *All of Statistics : A Concise Course in Statistical Inference (Springer Texts in Statistics).* Springer.

Wong, T. S. T. et W. K. Li. 2006. « A note on the estimation of extreme value distributions using maximum product of spacings », *Lecture Notes-Monograph Series*, vol. 52, p. pp. 272–283.