

Analyse actuarielle du péril incendie

Jérémie Barde*

sous la supervision de Prof. Hélène Cossette et de Prof. Etienne Marceau
École d'actuariat, Université Laval, Québec, Canada

8 août 2024

Résumé

Le péril incendie correspond à un péril majeur en assurance dommages. Il est une des raisons qui ont poussé au développement de l'assurance de dommage. Dans ce rapport, on utilise plusieurs familles paramétriques pour modéliser la sévérité des données canadiennes de sinistre incendie. Les lois de probabilité considérées sont la lognormale-Pareto généralisée, Weibull-Pareto généralisée, Coxienne-2-Pareto généralisée et la Bêta généralisée de type 2-Pareto généralisée. On utilise le maximum de vraisemblance pour estimer les paramètres des lois et l'on utilise des tests d'adéquations graphiques en plus de tests quantitatifs pour juger de la performance des lois ajustées aux données. On calcule les mesures VaR et $TVaR$ pour avoir de l'information sur la queue de la distribution. Pour le processus d'avènement des incendies, on utilise un processus de Poisson homogène pour la modélisation. Après avoir trouvé une loi pour le processus d'avènement des sinistres et une loi pour le montant des sinistres, on veut créer un fonds fictif pour analyser les entrées et sorties de fonds. À l'aide de la méthode de simulation Monte-Carlo, on peut approximer la probabilité de ruine ce qui permettra d'avoir une idée des chances que le fonds devienne insolvable.

Mots-clés : Actuariat dommages, risque incendie, loi Pareto généralisée, processus de Poisson,

*Corresponding author, jeremie.barde.1@ulaval.ca

Table des matières

1	Introduction générale	1
2	Mise en contexte	2
2.1	Assurance	2
2.2	Réassurance	2
2.3	Défi de l'assurance	3
2.4	Tarification assurance dommages	3
2.4.1	Processus de tarification	3
2.4.2	Modélisation actuarielle	4
3	Revue de la littérature	6
4	Données de sinistre incendie de la ville de Toronto	7
4.1	Description des données	7
4.2	Analyse descriptive	8
5	Modélisation des pertes pour le péril incendie (sévérité)	12
5.1	Modélisation préliminaire de la distribution avec une loi lognormale	12
5.2	Méthode <i>Peak-over-Threshold</i> (POT)	12
5.2.1	Déterminer le seuil	13
5.2.2	Paramètres	15
5.2.3	Mesure de risque	16
5.3	Raccordement de lois	17
5.3.1	Définition	17
5.3.2	Lognormale-Pareto généralisée	17
5.3.3	Weibull-Pareto généralisée	18
5.3.4	Coxienne-2-Pareto généralisée	19
5.3.5	Bêta généralisée de type 2-Pareto généralisée	21
6	Test des lois	22
6.1	Test graphique	22
6.2	Test quantitatif	24
7	Informations sur les lois LN-PaG et GB2-PaG	25
8	Modélisation du processus d'avènements des incendies	26
9	Étude avec une mutuelle fictive	27
10	Conclusion	29
11	Remerciements	29
A	Autres données	30
A.1	Données suédoises de sinistres incendies	30
A.2	Données norvégiennes de sinistres incendies	30
B	Caractéristiques des lois de probabilité	32
B.1	Loi lognormale	32
B.2	Loi de Weibull	33
B.3	Loi bêta	34
B.4	Loi Coxienne-2	36
B.5	Loi Bêta généralisée de type 2	37
B.6	Loi Pareto généralisée	38

C	Caractéristiques des raccordements de lois de probabilité	39
C.1	Loi Lognormale-Pareto généralisée	39
C.2	Loi Weibull-Pareto généralisée	40
C.3	Loi GB2-Pareto généralisée	41
D	Liste complète des variables explicatives de la base de données de sinistres incendies de la ville de Toronto	42
E	Description du projet pour la demande au CIMMUL	43

Liste des illustrations

1	Exemple de patron de la réassurance	3
2	Processus de tarification	4
3	Processus d'analyse des données	4
4	Contribution des plus grosses pertes estimées au total des pertes estimées	8
5	Pertes estimées selon la date du sinistre	9
6	Histogramme des pertes et du log des pertes	10
7	Fonction de répartition empirique des montants de sinistre	11
8	<i>QQplots</i> exponentielle, lognormale et Pareto	12
9	Fonction de répartition empirique comparée à la fonction de répartition de la loi lognormale	12
10	<i>QQplot</i> Pareto et fonction d'excès moyen	13
11	Comparaison fonction de répartition de la loi Pareto généralisée, de la fonction de répartition de la loi lognormale avec la fonction de répartition empirique	15
12	Comparaison de la fonction de répartition de la loi LN-PaG avec la fonction de répartition empirique	18
13	Comparaison fonction de répartition de la loi We-PaG avec la fonction de répartition empirique	19
14	Comparaison fonction de répartition de la loi We-PaG avec la fonction de répartition empirique	20
15	Comparaison fonction de répartition de la loi We-PaG avec la fonction de répartition empirique	21
16	<i>QQplot</i> lognormale-Pareto généralisée	22
17	<i>QQplot</i> pour la loi Weibull-Pareto généralisée	22
18	<i>QQplot</i> pour la loi Weibull-Pareto généralisée avec continuité	23
19	<i>QQplot</i> pour la loi Coxienne-2-Pareto généralisée	23
20	<i>QQplot</i> pour la loi Coxienne-2-Pareto généralisé avec continuité	24
21	<i>QQplot</i> pour la loi GB2-Pareto généralisée avec continuité	24
22	Test d'adéquation graphique pour la fréquence d'arrivée des sinistres incendies	26
23	Pour les sinistres supérieurs à 1 million	27
24	Évolution d'un fonds d'urgence pour le risque incendie avec les sinistres de la base de données canadienne et des simulations de la loi LN-PaG, en chargeant la prime pure quotidiennement	28

Liste des tableaux

1	Principales lignes d'affaires pour un assureur de dommages au Canada en 2021, ainsi que leurs pourcentages de primes nettes écrites	2
2	Description de 7 variables explicatives des données de sinistres incendie de la ville de Toronto	7
3	Statistiques descriptives des $n = 13571$ pertes estimées retenues de la base de données de sinistres incendies de la ville de Toronto	8
4	Nombre de données dans des tranches de montants de sinistres	9
5	Paramètres pour la loi lognormale	12
6	Seuil trouvé avec la méthode PLFIT selon différents vecteurs de u	14
7	Paramètres de la Pareto généralisée	15
8	Fonction de répartition empirique évaluée au seuil, nombre d'observations supérieures au seuil, pourcentage des montants de sinistre supérieurs au seuil	16
9	Statistique descriptive pour les montants de sinistres pour les bâtiments en construction et pour les bâtiments en construction et rénovations	16
10	Paramètres de la Pareto généralisée pour les bâtiments en construction et en construction et rénovation	16
11	Valeurs pour les mesures VaR et $TVaR$	17
12	Paramètres pour la loi composite lognormale-Pareto généralisée	18
13	Paramètres pour la loi composite lognormale-Pareto généralisée	19
14	Paramètres pour la loi composite Coxienne-2-Pareto généralisée	20
15	Paramètres pour la loi composite bêta généralisée de type 2-Pareto généralisée	21
16	Résultats des tests W^2 , A^2 , AIC et BIC	25
17	Intervalle de confiance estimé avec 14 000 <i>bootstrap</i>	25
18	Espérance et variance	25
19	Valeurs de la mesure VaR et de la mesure $TVaR$	26
20	Statistiques descriptives de la base de données suédoise de $n = 215$ sinistres incendies. Les montants de sinistres sont en unité de millions SKR.	30
21	Paramètres de la Pareto généralisée pour les montants de réclamations suédois	30
22	Statistiques descriptives de la base de données norvégienne de $n = 9181$ sinistres incendies . .	30
23	Paramètres de la Pareto généralisée pour les montants de réclamations norvégien	30

1 Introduction générale

L'assurance de dommage remonte à plusieurs d'années dans le passé. En effet, en Allemagne dans les années 1600 les premières sociétés mutuelles d'assurance font leur apparition. Cependant, le grand incendie de Londres en 1666, qui a brûlé les quatre cinquièmes de la ville, est un des événements qui a vraiment propulsé le besoin pour l'assurance. Un fait intéressant est que plusieurs des premières compagnies d'assurance sont créées pour faire face au risque incendie. En effet, la première compagnie d'assurance au Canada fait son apparition en 1809 et s'appelle la *Fire Association of Halifax*. Elle se spécialise principalement dans le risque incendie. À son tour en 1819 le Québec voit sa première compagnie d'assurance ouvrir ses portes et elle se nomme la *Québec Fire Insurance Company*. Puis, à compter de 1833, des sociétés mutuelles québécoises, établies à Montréal et en Estrie et spécialisées en assurance contre les incendies, ouvrent leur porte. On constate que l'assurance incendie a joué un gros morceau dans le développement de l'assurance dommages. Aujourd'hui, encore l'assurance incendie reste un défi pour les actuaires. Plusieurs personnes se sont vu amener à travailler sur ce risque et proposer différentes techniques pour bien le comprendre. Dans [Ramachandran, 1974] l'auteur propose d'utiliser la théorie des valeurs extrêmes pour modéliser les grosses pertes dues aux incendies. Dans [Brazauskas and Kleefeld, 2016] les auteurs utilisent le raccordement de loi pour essayer de mieux modéliser la queue de la distribution et ils testent sur des données de sinistres incendies. Bien qu'il soit important de modéliser la sévérité, il est aussi important de regarder la distribution du nombre de sinistres. Les auteurs de [Corradin and Verbrigghe, 2001] et [Corradin et al., 2002] proposent d'utiliser la binomiale négative comme loi pour la fréquence. Il y a aussi dans [Lin, 2005] où l'auteur propose d'utiliser le processus de Poisson pour modéliser la fréquence.

L'objectif de ce rapport est de modéliser le risque incendie avec les données canadiennes de sinistres incendies pour ensuite être en mesure d'analyser le risque et de calculer certaines mesures de risque. Plus précisément, il faudra comprendre et analyser la base de données canadienne. Après avoir identifié quelques modèles potentiels, il faudra faire des tests pour déterminer lequel est le plus adéquat. Après avoir choisi le modèle le plus adéquat, il sera possible de développer les caractéristiques de celui-là. Il sera intéressant de faire une analyse introductive de la fréquence pour déterminer si elle peut être modélisée par un processus de Poisson homogène. Après avoir modélisé la fréquence et la sévérité, il faudra utiliser la simulation pour analyser comment un fonds commun pour les montants de sinistres incendies évolue.

Dans la section 2, on présente une introduction à l'assurance. On montre les principales lignes d'affaires en assurance dommages, on introduit et motive la réassurance et on aborde les défis liés à l'assurance. On regarde aussi en surface la tarification en assurance dommages d'un point de vue global et d'un point de vue actuariel. La section 3 est constituée d'une revue non exhaustive de la littérature sur différents aspects du risque incendie. La section 4 a pour objet de présenter la base de données canadienne. On présente les transformations faites aux données pour les rendre utilisables. On poursuit avec une analyse descriptive où l'on regarde l'histogramme de la densité, l'histogramme de la log densité, la fonction de répartition empirique et quelques *QQplots*. Dans la section 5, on utilise la loi lognormale comme première loi pour modéliser les pertes incendies. Ensuite, on présente la méthode *Peak-over-Threshold* (POT) et les démarches pour arriver à trouver le bon seuil pour, par la suite, utiliser cette méthode sur les données canadiennes et pouvoir faire un premier calcul de la mesure VaR et $TVaR$. Plus loin, on utilise le raccordement de lois pour avoir une meilleure adéquation avec nos données. Les raccordements suivants seront testés : Lognormale-Pareto généralisée (LN-PaG), Weibull-Pareto généralisée (We-PaG), Coxiennne-2-Pareto généralisée (Cox2-PaG) et la bêta généralisée de type 2-Pareto généralisée (GB2-PaG). Pour la section 6, on fait un test d'adéquation graphique et des tests statistiques pour déterminer la loi parmi celles testées qui offre la meilleure adéquation aux données canadiennes de sinistres incendies. Pour la section 7 on présente la moyenne, la variance, la mesure VaR et la mesure $TVaR$ pour le raccordement de loi LN-PaG. On passera en suite à la section 8 dans laquelle on tente de montrer que la fréquence d'arrivée des sinistres de la base de données canadienne suit possiblement un processus de Poisson homogène. On terminera avec la section 9 où l'on utilise les sinistres de la base de données canadienne et la simulation de Monte-Carlo pour analyser l'évolution journalière du fonds d'une mutuelle fictive pour les montants de sinistres incendies.

2 Mise en contexte

2.1 Assurance

En assurance, il y a principalement deux grandes lignes : l'assurance vie (appelée aussi assurance de personnes) et l'assurance dommages. L'assurance vie est principalement composée de l'assurance vie, l'assurance invalidité, etc. Pour ce présent rapport, on se concentre sur l'assurance dommages. L'assurance dommages est constituée de différentes lignes d'affaires. Par exemple, la plus importante est la ligne personnelle, constituée en grande partie par l'assurance automobile et l'assurance habitation. L'ensemble des lignes d'affaires sont présentées au tableau 1.

Ligne d'affaire	Primes nettes écrites
Assurance automobile	43 %
Assurance habitation ligne personnelle	22 %
Assurance aux entreprises	15 %
Assurance responsabilité	11 %
Autres	11 %

TABEAU 1 – Principales lignes d'affaires pour un assureur de dommages au Canada en 2021, ainsi que leurs pourcentages de primes nettes écrites

À la colonne 2 du tableau 1, on trouve les pourcentages de primes nettes écrits pour chaque ligne d'affaires. Ces pourcentages sont extraits de *Assurances de dommages au Canada*¹, publié par le Bureau d'Assurance du Canada. On constate que l'assurance automobile représente un gros morceau de l'assurance dommages puisqu'elle représente un peu moins de la moitié des primes nettes écrites. Dans ce rapport, on s'intéressera davantage à l'assurance aux entreprises.

2.2 Réassurance

Un autre aspect important dans le milieu de l'assurance est la réassurance. Les réassureurs sont en quelques sortes les assureurs des assureurs. En effet, les assureurs vont transférer la portion du risque qu'ils ne sont pas en mesure de supporter aux réassureurs. Cela peut être fait de différentes façons, par exemple par un pourcentage de la perte ou par la portion supérieure à une limite préétablie. Dans [Albrecher et al., 2017], les auteurs proposent une excellente introduction à la réassurance.

Voici quelques raisons, mentionnées dans ce livre, qui poussent les assureurs à avoir recours à la réassurance. Elle leur permet d'abord de réduire la probabilité de subir des pertes difficiles à assumer. En d'autres mots, la réassurance permet de diminuer l'exposition au risque des assureurs.

Elle permet aussi de stabiliser les résultats commerciaux. En effet, une difficulté de l'assurance est de prévoir les coûts futurs qu'une police pourrait engendrer. Cependant, avec la réassurance, ces coûts futurs inconnus se transforment en une prime connue d'avance par les assureurs.

En faisant affaire avec des réassureurs, les assureurs peuvent diminuer leur capital requis, c'est-à-dire qu'ils ont besoin de moins d'argent pour un certain contrat. Cela leur permet de prendre davantage de contrats ou d'avoir une plus grande possibilité d'investissement.

Les assureurs peuvent aussi augmenter leur capacité à la souscription. En effet, en transférant une partie du risque aux réassureurs, ils peuvent accepter plus de contrats en plus d'accepter des contrats qu'ils n'auraient pas pu prendre seuls.

Les assureurs opèrent généralement dans une région relativement restreinte. La réassurance leur permet d'accéder à des pools de diversification plus importants. En effet, puisque les réassureurs opèrent généralement

1. Bureau d'Assurance du Canada, www.ibc.ca/fr/ns/ressources/état-de-l'industrie/publication-assurances-de-dommages-au-canada

à l'international, les assureurs qui font affaire avec eux ont accès à des marchés qui seraient autrement trop risqués pour eux.

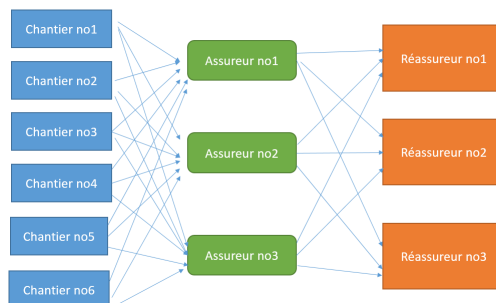


ILLUSTRATION 1 – Exemple de patron de la réassurance

L'illustration 1 présente un exemple d'échange entre des clients, des assureurs et des réassureurs. On voit que différents chantiers transfèrent leurs risques à différents assureurs, qui transfèrent ensuite les risques qu'ils ne peuvent se permettre de couvrir aux réassureurs.

2.3 Défi de l'assurance

Le défi de l'assurance est de déterminer une prime juste et équitable qui permettra à la compagnie de répondre à ses obligations futures. Ce défi est plus complexe qu'il n'en a l'air, car un assureur ne sait pas à l'avance combien il devra payer dans le futur. Pour une entreprise standard, c'est plus simple, car les coûts sont en partie connus d'avance. On a donc une équation simple pour identifier le prix d'un produit,

$$\text{Prix} = \text{Coût} + \text{Profit}.$$

Par contre, en assurance, cette équation est un peu plus complexe,

$$\begin{aligned} \text{Prime} = & \text{Sinistre} + \\ & \text{Frais de règlement de sinistre} + \\ & \text{Frais de souscription de sinistre} + \\ & \text{Profit de souscription}. \end{aligned}$$

Les sinistres constituent la portion difficile à évaluer puisqu'ils ne sont pas connus d'avance.

L'assurance aux entreprises amène aussi ces défis. En assurance de particulier, l'exposition au risque est beaucoup moins importante pour un assureur comparativement à l'assurance aux entreprises. En général, une police d'assurance pour un particulier ne dépassera pas le capital de l'assureur. Cependant, pour une police d'assurance aux entreprises, le capital de l'assureur pourrait ne pas être suffisant. Les assureurs peuvent donc aussi avoir une aversion au risque et avoir recours aux réassureurs.

2.4 Tarification assurance dommages

2.4.1 Processus de tarification

Comme mentionné plus tôt, le défi de l'assurance est d'établir la prime. Pour y parvenir, il faut suivre un processus de tarification. Plusieurs excellents ouvrages ont traité du sujet, autant pour l'assurance dommages [Parodi, 2014] que pour l'assurance commerciale [Michaelides et al., 1997] ou l'assurance dommages au particulier [Denuit et al., 2007] et [Ohlsson and Johansson, 2010].

Il y a différentes étapes au processus de tarification. L'illustration 2, inspirée par l'illustration 1.8 de [Parodi, 2014] et tirée de [Blier-Wong, 2018] à la page 2, illustration 1.1, présente une vision d'ensemble du processus.

Le processus suivra le modèle de fréquence-sévérité sous forme de processus agrégé défini comme suit :

$$S(t) = \begin{cases} \sum_{k=1}^{N(t)} B_k & , N(t) > 0 \\ 0 & , N(t) = 0, \end{cases}$$

où $N(t)$ est un processus aléatoire qui modélise la fréquence et où B_k est la variable aléatoire (v.a.) qui modélise la sévérité. On utilise un modèle dynamique pour décrire l'évolution dans le temps des coûts d'un portefeuille. Les hypothèses classiques de ce modèle sont :

- $\underline{B} = \{B_1, B_2, \dots\}$ forme une suite de v.a.
- B_1, B_2, \dots ont la même distribution que B .
- la v.a B et le processus de comptage $N(t)$ sont indépendants.

Ces hypothèses permettent de calculer facilement la prime pure, qui se trouve à être l'espérance du processus agrégé \underline{S} ,

$$E[S(t)] = E[N(t)] \times E[B].$$

Cependant, pour y arriver, en général on modélise le processus de comptage et la v.a. pour la sévérité séparément. Plusieurs choix sont possibles pour la loi de la sévérité.

Fréquence

Pour le processus de comptage, on pourrait par exemple choisir un processus de poisson homogène. On aurait les caractéristiques suivantes :

- $N(t) \sim Po(\lambda t)$
- $E[N(t)] = \lambda t$
- Distribution totale des sinistres : processus de Poisson homogène composé

Sévérité

Pour la v.a B , on verra souvent la loi Gamma. Par contre, si on travaille avec des pertes très élevées, on choisira plutôt des lois à queue lourde comme les lois lognormale et Pareto. Les caractéristiques de ces lois sont :

- $B \sim Gamma(\alpha, \lambda)$ avec $E[B] = \frac{\alpha}{\lambda}$
- $B \sim Pa(\alpha, \lambda)$ avec $E[B] = \frac{\lambda}{\alpha-1}$, $\alpha > 1$
- $B \sim LN(\mu, \sigma)$ avec $E[B] = e^{\mu + \frac{\sigma^2}{2}}$

Pour mieux connaître le risque, on a recours à certaines mesures de risque. Des mesures couramment utilisées en science actuarielle sont les mesures VaR et $TVaR$. La mesure VaR est définie comme suit :

$$VaR_{\kappa}(B) = \inf\{x \in \mathbb{R} : F_B(x) \geq \kappa\}, \quad \kappa \in (0, 1)$$

Il faut voir cette mesure comme étant la plus petite prime à charger pour être sûr à un niveau 100κ % qu'on est capable de payer si une réclamation survient. Le défaut de cette mesure est qu'elle ne donne pas d'information sur des valeurs supérieures à $VaR_{\kappa}(B)$. Pour remédier à cela, on va calculer la mesure $TVaR$, qui est définie comme suit :

$$TVaR_{\kappa}(B) = \frac{1}{1 - \kappa} \int_{\kappa}^1 VaR_u(B) du, \quad u \in (0, 1).$$

Elle peut être considérée comme la moyenne des $VaR_{\kappa}(B)$ pour des valeurs de u supérieur à κ . Elle donne donc de l'information pour les valeurs supérieures à $VaR_{\kappa}(B)$.

3 Revue de la littérature

Dans [Longley-Cook, 1952], l’auteur aborde le sujet des pertes élevées dues au feu. Il est mentionné qu’il est difficile de trouver une loi pour modéliser les pertes élevées, mais qu’il est possible d’utiliser des données pour créer un modèle empirique. L’auteur réussit à construire une base de données qu’il regroupe en différentes catégories selon le montant de perte. Il calcule le coût au-dessus d’une certaine franchise préalablement choisie pour chaque montant de perte et remarque qu’il est possible d’assumer que la deuxième dérivée du log des pertes décroît de manière géométrique. Cela permettait donc d’estimer l’espérance des pertes pour chaque groupe. Il mentionne cependant que ce modèle n’était pas excellent. Dans [Longley-Cook, 1951], l’auteur présente les principales difficultés dans la tarification incendie à cette époque. Il critique principalement le fait que les procédures mises en place étaient trop basées sur le jugement et qu’il était donc difficile de savoir si le taux trouvé était juste et équitable. Une autre critique est que les méthodes n’étaient pas standardisées entre les états. En effet, certains états considéraient seulement deux types de bâtiments, soit les bâtiments en brique et ceux en bois, alors que d’autres états avaient beaucoup plus de catégories. Il propose justement dans [Longley-Cook, 1959] d’avoir un seul bureau de tarification puisqu’à cette époque, au États-Unis, il y en avait 38. Il propose aussi de standardiser la tarification en assurance de propriété.

[Benckert and Sternberg, 1957] introduit la loi Pareto, [Mandelbrot, 1964] pousse les travaux plus loin. Dans [Charpentier and Flachaire, 2021] les auteurs montrent l’utilisation de la loi Pareto dans le domaine de l’assurance, de la réassurance, de la finance ainsi que son utilisation pour des mesures de risques. Ils modélisent aussi, à l’aide de la loi Pareto, des sinistres élevés à la suite d’un incendie avec la base de données *Danish dataset*.

Dans [Almer, 1963], l’auteur utilise la théorie des risques pour analyser le risque incendie. Il mentionne que cela se fait naturellement puisque les assurés pour le risque incendie sont généralement indépendants et cela s’applique bien à la théorie du risque individuelle.

L’auteur de [Antal and Re, 2007] propose une excellente introduction, plus moderne, à la théorie des risques. Il présente les outils mathématiques pour la tarification dommages et la modélisation. De plus, il propose une introduction à la théorie du risque, une excellente explication du capital et une introduction à la tarification des contrats de réassurance.

Ramachandran a énormément travaillé et écrit sur le risque incendie. Dans [Ramachandran, 1974], il évoque que la théorie des valeurs extrêmes pourrait être un bon outil pour modéliser les grosses pertes dues aux incendies, mais qu’il faudrait l’utiliser avec prudence. Dans [Rydman, 2018] l’auteur utilise la méthode *Peak-over-Threshold* pour modéliser une base de données de montants sinistres en assurance automobile. L’auteur présente deux techniques pour trouver les paramètres de la Pareto généralisée, soit la méthode du maximum de vraisemblance et la méthode *Probability Weighted Moments*.

Plusieurs chercheurs ont tenté de modéliser les pertes dues aux incendies. Par exemple, [Shpilberg, 1977] critique que la majorité des lois soient bonnes uniquement pour modéliser la portion centrale, soit autour de la médiane, et moins bonnes pour la queue de la distribution. Il propose de tester les lois Lognormale, quasi-lognormale et Pareto. Il conclut qu’il est très difficile de choisir la meilleure loi puisque, graphiquement, les résultats sont similaires. Il utilise le fait que la fonction densité du feu devrait être décroissante pour montrer que la lognormale est un meilleur choix puisque la fonction d’intensité de la Pareto est une constante.

Dans [Cummins et al., 1990], les auteurs proposent la famille de distributions GB2 pour la modélisation des pertes. Cette famille contient les lois gamma généralisée, bêta généralisée de type 1 et bêta généralisée de type 2. Cette famille de distributions a l’avantage d’être très flexible. Les auteurs mentionnent qu’avec cette famille, on peut modéliser la sévérité, mais aussi la distribution des montants agrégés.

Pour [Brazauskas and Kleefeld, 2016], les auteurs visent à modéliser les montants de pertes d’incendie de la base de données norvégienne d’incendie. Ils proposent le raccordement de lois, c’est-à-dire de modéliser différentes portions de la distribution avec différentes lois, par exemple les lois lognormale-Pareto généralisée et Weibull-Pareto Généralisée. Ils proposent aussi d’utiliser la famille de distributions *Folded-t* en appliquant une transformation aux données. Ils présentent différents tests statistiques pour sélectionner la meilleure loi. Dans leur cas, ce sont les lois avec la Pareto généralisée et le modèle avec la famille de distributions *Folded-t* qui étaient les meilleurs.

Pour [Biffis and Chavez, 2014], les auteurs utilisent une base de données obtenue auprès de 2 syndicats de *Lloyd's of London*. Ils veulent estimer l'index de la queue de la distribution pour avoir de l'information sur la lourdeur des montants de sinistre en assurance commerciale. Deux méthodes sont présentées, une méthode de régression et l'utilisation de l'estimateur de Hill. Ils trouvent l'indice de la queue de la distribution pour différents secteurs comme les secteurs résidentiel, commercial et manufacturier.

Dans [Corradin and Verbrigghe, 2001] et [Corradin et al., 2002], les auteurs proposent aussi une analyse d'une base de données de montants de sinistre incendie. Ils utilisent le modèle de fréquence-sévérité pour modéliser le montant de pertes au-dessus d'un montant préétabli. Les auteurs utilisent la loi binomiale négative pour modéliser la fréquence et la loi Pareto généralisée pour modéliser la sévérité. Ils poussent plus loin en utilisant leur modèle pour faire de la simulation et ainsi tester différents scénarios de réassurance et voir l'impact sur le capital.

Dans [Lin, 2005], l'auteure montre qu'il est envisageable de modéliser la fréquence des incendies par un processus de Poisson. Elle présente les résultats qu'elle a obtenus avec des données en provenance de Taiwan. Elle teste la processus de Poisson sur différents types de bâtiments et montre que les bâtiments industriels ont le plus gros taux d'incendie par unité de surface du sol.

Une méthode de tarification utilisée en réassurance est *exposure rating*. Elle est expliquée dans le chapitre 21 de [Parodi, 2014]. On utilise une courbe d'exposition au lieu d'une courbe de sévérité. Elle donne le pourcentage de risque qui est conservé par l'assureur si une franchise est imposée. Une courbe d'exposition couramment utilisée est la MBBEFD, introduite dans [Bernegger, 1997], car elle a l'avantage de pouvoir produire d'autres courbes selon sa paramétrisation. [Riegel, 2010] présente une application de cette technique dans le cas de réassurance pour le risque d'incendie. Cette méthode est aussi utilisée par [Abramson, 2022] pour la tarification dans le milieu de la construction.

4 Données de sinistre incendie de la ville de Toronto

4.1 Description des données

Pour modéliser le risque incendie, on a utilisé une base de données canadienne provenant de Toronto. La base de données est accessible sur le site [City of Toronto's Open Data Portal](https://open.toronto.ca/dataset/fire-incidents/)². Le jeu de données contient de l'information sur 17 536 cas d'incendies. Pour chaque incident, il y a 43 variables explicatives. Dans le tableau 2, on présente 7 de ces 43 variables explicatives et la liste complète se trouve dans l'annexe D.

Variable	Définition
Emplacement d'origine :	Brève description de l'emplacement d'origine du feu.
Perte civile :	Nombre de personnes décédées suite à l'incendie.
Pertes estimées :	Montants de sinistre estimé.
Contagion :	Description sur la contagion de l'incendie.
Nombre de répondants à l'appel :	Nombre de personnes présentes pour régler l'incendie.
État de l'incendie à l'arrivée :	Éteint, avec fumé, contagion.
Type de bâtiment :	Normal, en construction, en rénovation, etc.

TABEAU 2 – Description de 7 variables explicatives des données de sinistres incendie de la ville de Toronto

Pour la présente analyse, seule la variable *pertes estimées* sera utilisée. Il est important de souligner qu'il s'agit de pertes estimées, ce qui explique qu'il y a plusieurs occurrences d'un même montant. Afin de faciliter l'analyse, il a été nécessaire de retirer les 1 909 données, dont les pertes estimées, sont indisponibles. De plus, on a aussi retiré les 2 056 données dont les pertes estimées sont de 0. Donc, on fait notre analyse avec $n = 13571$ données.

2. Fire Services, <https://open.toronto.ca/dataset/fire-incidents/>

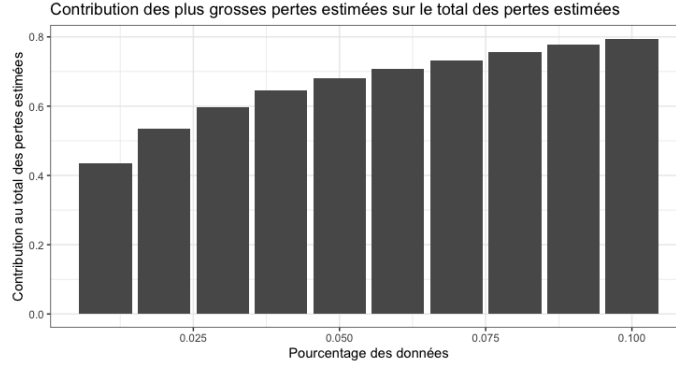


ILLUSTRATION 4 – Contribution des plus grosses pertes estimées au total des pertes estimées

Au tableau 3, on présente les statistiques descriptives de la variable *pertes estimées* après les modifications.

Min.	1 ^{er} Qu.	Médiane	Moyenne	Variance	3 ^{er} Qu.	Max.
1	1000	5000	40 296	2.38×10^{11}	20 000	5×10^7

TABEAU 3 – Statistiques descriptives des $n = 13571$ pertes estimées retenues de la base de données de sinistres incendies de la ville de Toronto

La moyenne des pertes est de 40 296 \$ CAN et elle est considérablement plus élevée que la médiane des pertes à 5 000 \$ CAN. De son côté, la variance des pertes est extrêmement élevée. Cela pourrait potentiellement indiquer qu'elle n'existe pas. Comme on voit à l'illustration 4, plus de 65 % des pertes totales sont dues à 5 % des données et environ 80 % des pertes totales sont dues à 10 % des données. De plus, l'illustration 5 permet de bien voir que certaines valeurs sont extrêmement élevées comparativement aux autres. En effet, l'illustration 5a montre les pertes totales de 2011 à 2019. La ligne horizontale rouge représente le 99.9 % percentile qui est de 2.7 millions. On remarque qu'il y a quand même 14 pertes qui sont au-dessus du 99.9 % percentile. Les illustrations 5b, 5c et 5d montrent les sinistres supérieurs à 0.5, 1, et 5 millions. La ligne horizontale rouge est la borne inférieure. On constate qu'il y a 69 pertes supérieures ou égales à 1 million. Sur l'illustration 5, on observe que la perte maximale de 50 millions n'est arrivée qu'une seule fois durant ces années. Par contre, cette observation ne peut pas être considérée comme aberrante, car elle est typique des données de sinistres incendies. On observe ce phénomène dans plusieurs bases de données de sinistres incendie. Dans la section A on présente les données suédoises de sinistres incendies et les données norvégiennes de sinistres incendies et on peut constater aux tableaux 20 et 22 que la valeur maximale pour chaque base de données est extrêmement élevée comparativement à la moyenne. De plus, on observe aussi phénomène dans la célèbre base données *Danish data on large fire insurance losses* présenté dans [McNeil, 1997a]. En effet, dans cette base de données la moyenne est de 3.39 et la valeurs maximale est de 263.25.

4.2 Analyse descriptive

Afin de mieux comprendre les données, on commence par une analyse descriptive qui donnera des pistes de futures lois à utiliser. L'histogramme de la densité et de la log densité, la fonction de répartition empirique et des *QQplots* seront produits.

Au tableau 4, on constate qu'il y a peu de montants de sinistres extrêmement élevés en comparaison avec les sinistres plus normaux, mais les sinistres élevés restent responsables d'une grande partie des pertes totales.

On observe à l'illustration 6a que l'histogramme de la densité n'est pas facile à lire en raison des valeurs extrêmes. Il est à noter que l'axe des x arrête à 300 000 pour que la lecture soit plus facile. Il semble y avoir

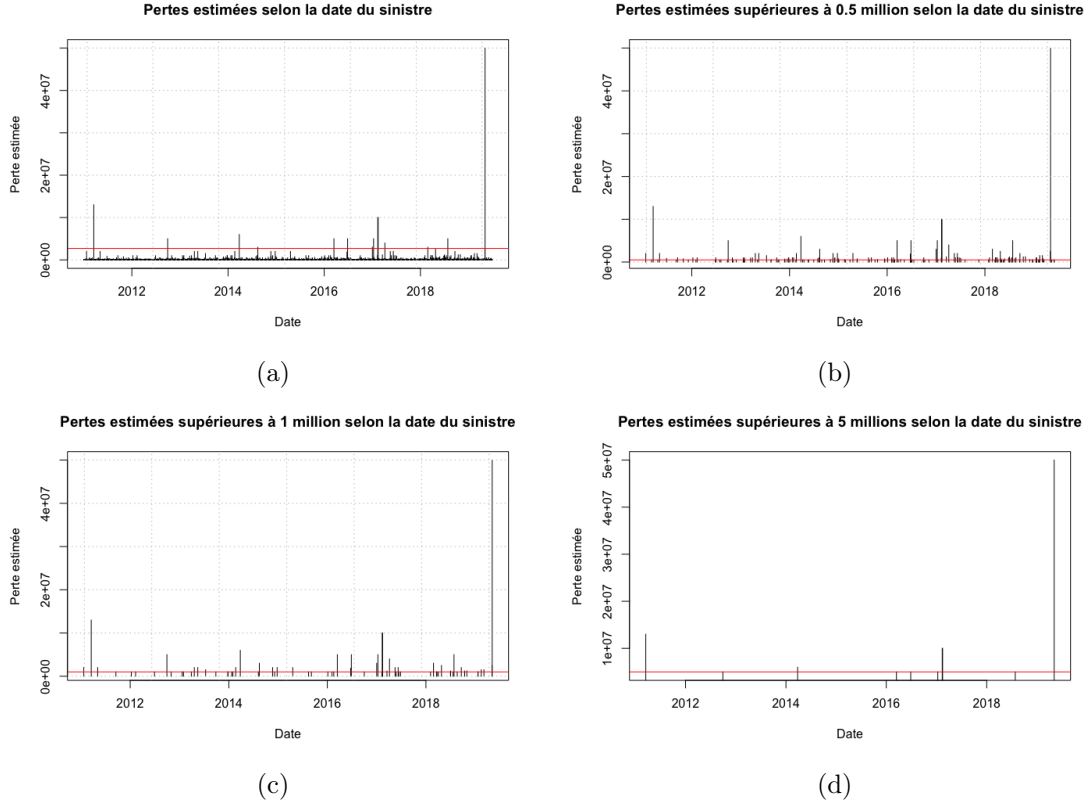


ILLUSTRATION 5 – Pertes estimées selon la date du sinistre

des modes à certaines valeurs, par exemple 50 000, 100 000 et 150 000. Cela est probablement dû au fait que le montant de sinistres est estimé et qu'on a plusieurs occurrences d'une même valeur. Cela crée donc des masses à certaines valeurs qui sont peut-être des montants de perte plus courants.

Intervalles	Nombre de données									
	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total
$[1, 50\,000[$	1394	1373	1328	1387	1347	1343	1291	1458	686	11607
$[5 \times 10^4, 5 \times 10^5[$	172	181	199	207	195	219	222	269	132	1796
$[5 \times 10^5, 1 \times 10^6[$	6	7	11	14	13	16	4	17	11	99
$[1 \times 10^6, 1 \times 10^7[$	3	5	9	9	3	7	10	14	5	65
$[1 \times 10^7, 5 \times 10^7[$	1	0	0	0	0	0	2	0	0	3
$[5 \times 10^7, \infty[$	0	0	0	0	0	0	0	0	1	1

TABLEAU 4 – Nombre de données dans des tranches de montants de sinistres

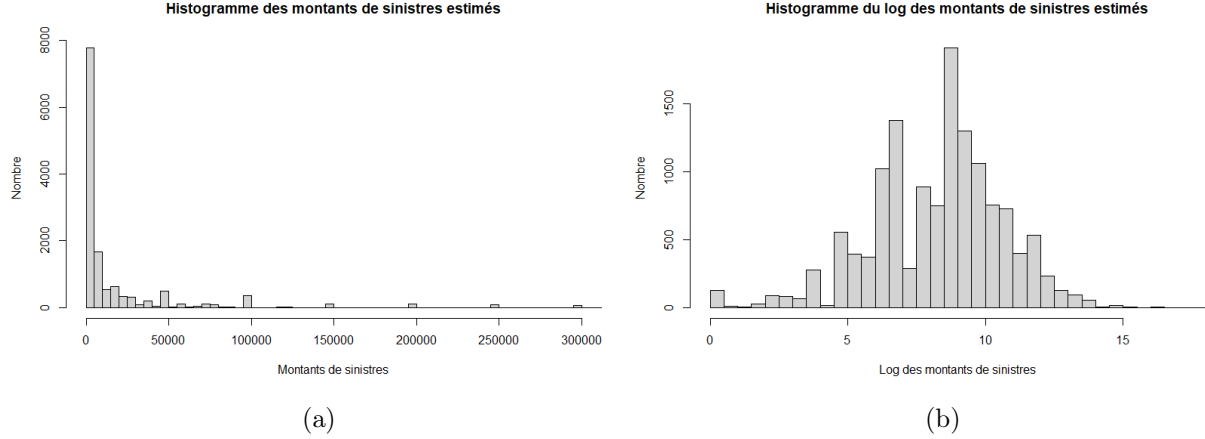


ILLUSTRATION 6 – Histogramme des pertes et du log des pertes

Afin de rendre la lecture encore plus facile, on regarde le log des montants de sinistre. De cette façon, on a des montants moins élevés, ce qui nous permet d’avoir une échelle plus facile à lire. On constate à l’illustration 6b que le montant des sinistres semble être à peu près normal quand on applique le log.

Proposition 1 *Soit une variable aléatoire X où $\ln(X) \sim N(\mu, \sigma)$. Alors X obéit à une loi lognormale de paramètres (μ, σ) , noté $X \sim LN(\mu, \sigma)$.*

Démonstration. Posons $Y = \ln(X) \sim N(\mu, \sigma)$. Alors, on a

$$F_Y(y) = \Phi\left(\frac{y - \mu}{\sigma}\right) = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right) = F_X(x), \quad x > 0$$

On obtient bien la fonction de répartition d’une loi lognormale. \square

Proposition 2 *Soit une variable aléatoire X où $\ln(X) \sim \exp(\alpha)$. Alors X obéit à une loi Pareto de paramètres $(\alpha, 1)$, noté $X \sim Pa(\alpha, 1)$ (Type 1).*

Démonstration. Posons $Y = \ln(X) \sim \exp(\alpha)$. Alors, on a

$$\bar{F}_Y(y) = e^{-\alpha y} = e^{-\alpha \ln(x)} = x^{-\alpha} = \bar{F}_X(x), \quad x > 1$$

On obtient bien la fonction de survie d’une loi Pareto de type 1. \square

La Proposition 1 donne l’intuition que le montant des sinistres espérés pourrait suivre une loi lognormale. On regarde aussi la fonction de répartition empirique à laquelle on pourra plus tard comparer les fonctions de répartition des lois testées.

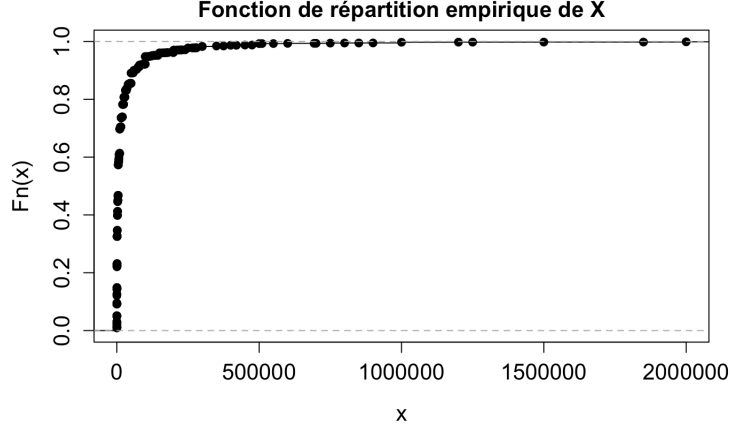


ILLUSTRATION 7 – Fonction de répartition empirique des montants de sinistre

À l'illustration 7, on constate que la fonction de répartition s'approche de 1 assez rapidement. Cependant, on voit qu'il y a encore de faibles probabilités d'avoir des sinistres très élevés. Il est déjà qu'on est une distribution à queue lourde. Cet aspect sera vérifié avec les *QQplots*. L'axe des x est coupé à 2 millions pour faciliter la lecture, mais en réalité, le domaine se rend à 50 millions.

Une façon d'avoir une idée initiale de la distribution des montants de sinistre est de regarder les *QQ-plots* de certaines lois.

Définition 1 *Le QQ-plot d'une loi exponentielle est défini comme suit :*

$$\left(-\ln \left(1 - \frac{i}{n+1} \right), X_{[i]} \right), \quad i = 1, 2, \dots, n$$

où $X_{[i]}$ est la i ième statistique d'ordre.

Définition 2 *Le QQ-plot d'une loi Pareto est défini comme suit :*

$$\left(-\ln \left(1 - \frac{i}{n+1} \right), \ln(X_{[i]}) \right), \quad i = 1, 2, \dots, n$$

où $X_{[i]}$ est la i ième statistique d'ordre. On utilise la Proposition 2

Définition 3 *Le QQ-plot d'une loi lognormale est défini comme suit :*

$$\left(\Phi \left(\frac{i}{n+1} \right), \ln(X_{[i]}) \right), \quad i = 1, 2, \dots, n$$

où $X_{[i]}$ est la i ième statistique d'ordre. On utilise la Proposition 1

En regardant l'illustration 8a, on voit bien que la distribution du montant des sinistres a une queue plus lourde qu'une loi exponentielle. En effet, la ligne bleue représente la ligne que les points devraient suivre si la distribution était exponentielle, mais les points passent bien au-dessus de cette droite. Ce résultat confirme l'intuition d'une distribution à queue lourde que la fonction de répartition empirique donnait. Il est donc intéressant d'aller regarder des *QQplots* de lois à queue lourde, soit la lognormale et la Pareto. On constate à l'illustration 8b que le *QQplots* de la loi lognormale produit une droite, ce qui est une autre bonne indication que le montant de sinistres suit une loi lognormale. Par contre, le *QQplots* de la Pareto à l'illustration 8c est aussi une droite pour certaines valeurs. Cela pourrait indiquer qu'une portion de la distribution est une loi lognormale et que l'autre est une loi Pareto. Les *QQplots* présentés à l'illustration 8 sont produits à l'aide des fonctions `ExpQQ`, `LognormalQQ` et `ParetoQQ` du package `ReIns` de R, [Reynkens and Verbelen, 2020].

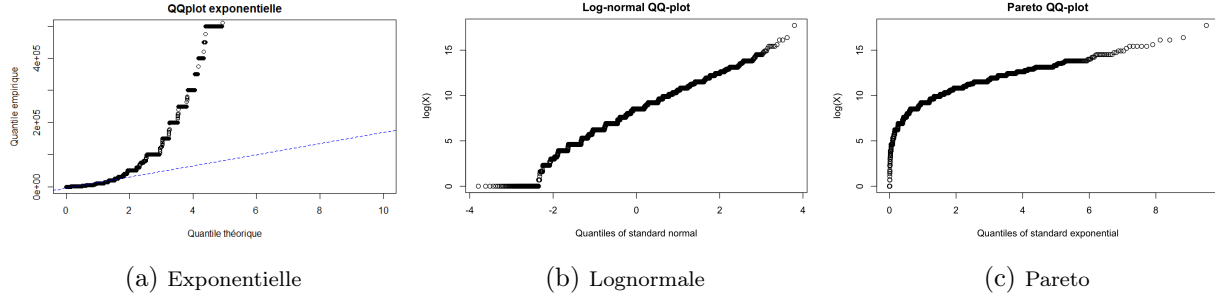


ILLUSTRATION 8 – *QQplots* exponentielle, lognormale et Pareto

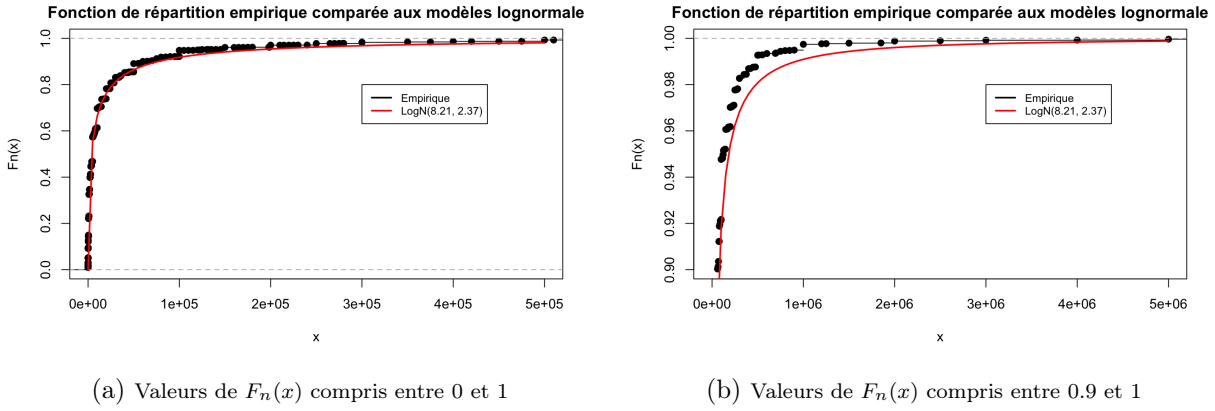


ILLUSTRATION 9 – Fonction de répartition empirique comparée à la fonction de répartition de la loi lognormale

5 Modélisation des pertes pour le péril incendie (sévérité)

5.1 Modélisation préliminaire de la distribution avec une loi lognormale

Il a été constaté dans la section 4.2 que la loi lognormale serait un choix judicieux comme première modélisation des montants de sinistre incendies. On utilise le maximum de vraisemblance (MV) pour trouver les paramètres de la loi. La méthode des moments (MOM) est utilisée pour les valeurs de départ. La fonction `constrOptim` de R est utilisée pour l’optimisation. Les paramètres obtenus sont présentés au tableau 5 :

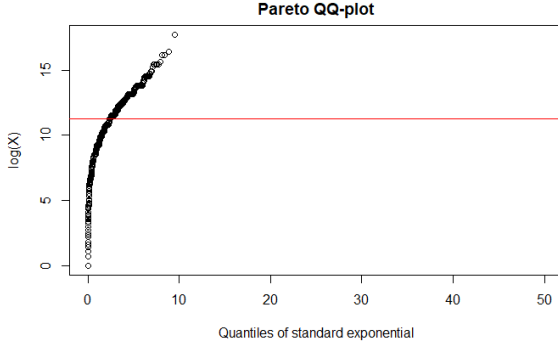
μ	σ
8.21	2.37

TABLEAU 5 – Paramètres pour la loi lognormale

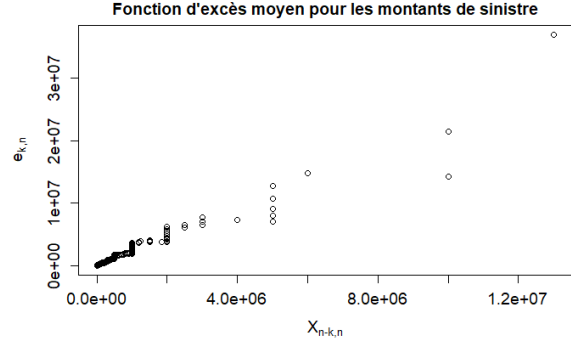
À l’illustration 9, on remarque que la loi lognormale modélise adéquatement les montants de sinistres moins élevés. Cependant, les montants de sinistres plus élevés sont surévalués. La loi lognormale ne modélise pas bien les valeurs faibles en même temps que les valeurs extrêmes. L’intuition mentionnée dans la section 4 que la queue de la distribution pourrait être Pareto mériterait d’être testée.

5.2 Méthode *Peak-over-Threshold* (POT)

Il a été mentionné dans la section 4 qu’il y avait des montants extrêmes avec de faibles probabilités de réalisation. Il a aussi été mentionné dans la section 4.2 que la distribution des montants de sinistre aurait une portion lognormale et une portion Pareto. Il est donc intéressant d’avoir recours à la théorie des valeurs



(a) *QQplot* Pareto avec ligne horizontale à $\ln(75000)$



(b) Fonction d'excès moyen

ILLUSTRATION 10 – *QQplot* Pareto et fonction d'excès moyen

extrêmes, présentée dans [McNeil, 1997b]. Elle permet de dire que, pour un seuil assez haut, la distribution des montants de sinistre suit une Pareto généralisée, dont la fonction de répartition est donnée par

$$F_X(x) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma}(x - u)\right)^{-\frac{1}{\xi}} & , \xi \neq 0 \\ 1 - \exp\left\{-\frac{x-u}{\sigma}\right\} & , \xi = 0 \end{cases}, \quad x \geq u.$$

Le défi est de trouver un seuil, u , au-dessus duquel la distribution des montants de sinistres correspond à une loi Pareto généralisée. Cependant, il faut faire attention, car si on sélectionne un seuil trop bas, les résultats risquent d'être biaisés. Par contre, si on prend un seuil trop élevé, on risque d'avoir un nombre insuffisant de données, ce qui pourrait également produire des résultats biaisés.

Le paramètre α de la Pareto généralisée est très important. On devrait, en théorie, pour le risque incendie, trouver $\alpha \approx 1.5$, comme mentionné dans [Antal and Re, 2007] à la page 11. On ramène les paramètres de la Pareto généralisée sous leur forme plus connue avec les transformations suivantes,

$$\alpha = \frac{1}{\xi}, \quad \lambda = \frac{\sigma}{\xi}.$$

Ces transformations seront utilisées tout au long de ce rapport pour indiquer les paramètres de la Pareto généralisée.

5.2.1 Déterminer le seuil

Pour déterminer le seuil, il est possible d'analyser le graphique de la fonction d'excès moyen. Il faut identifier la portion où la fonction d'excès moyen au-dessus d'un seuil est linéaire. À l'illustration 10b, on voit qu'il est difficile de repérer un tel seuil.

Il est également possible de regarder le *QQplots* Pareto et d'essayer de voir un changement de distribution. On observe, à l'illustration 8c, que la courbe se transforme tranquillement en droite. Il demeure difficile d'identifier le point précis où le changement se fait. L'illustration 10a est le même *QQplot* qu'en 8c, mais avec un changement de l'axe des x qui permet de mieux identifier où la transformation se fait. La ligne horizontale rouge se trouve à $\ln(75000)$, ce qui semble être le passage à la distribution Pareto. Le seuil u serait possiblement 75 000.

Cependant, il est aussi important de tester des méthodes de détection automatique du seuil. On testera trois méthodes soit la méthode normalité des écarts, la méthode Power-Law FIT (PLFIT) et une méthode par double *bootstrap*.

Normalité des écarts

La première méthode testée est celle présentée par [Thompson et al., 2009]. On choisit m seuils également séparés qu'on place du plus petit au plus grand. Pour chaque seuil, on calcule les estimateurs du maximum de vraisemblance $\hat{\xi}_{u_i}$, $\hat{\sigma}_{u_i}$ pour ensuite définir

$$\tau_{u_i} = \hat{\sigma}_{u_i} - \hat{\xi}_{u_i} u_i, \quad i = 1, \dots, n.$$

On peut montrer que,

$$E[\tau_{u_i} - \tau_{u_{i-1}}] \approx 0, \quad i = 2, \dots, n$$

et ainsi conclure que $\tau_{u_i} - \tau_{u_{i-1}}$ est approximativement normal. On a l'algorithme 1 suivant :

Algorithme 1 Normalité des écarts

```

1:  $u \leftarrow \text{vecteur}(\text{seuil})$ 
2: tant que  $p.\text{value} < 0.05$  faire
3:    $p.\text{value} \leftarrow \text{test.normalite}(u)$ 
4:    $u \leftarrow u[-1]$ 
5: fin tant que
6: retourne  $u[1]$ 

```

▷ Le seuil optimal est $u[1]$

Cette approche est implémentée dans la fonction TH du package `tea` de R, [Ossberger, 2020]. Cependant, cette technique s'est avérée inefficace avec la base de données canadienne. Elle produisait des résultats non constants et choisissait en général la dernière limite du vecteur.

Power-Law FIT

Une méthode présentée par [Clauset et al., 2007] a aussi été essayée. Cette approche est assez intuitive. On veut trouver la valeur de u qui minimisera l'écart entre la fonction de répartition empirique et la fonction de répartition de la loi théorique. Pour y arriver, la statistique de Kolmogorov–Smirnov (KS) a été utilisée.

Définition 4 Soit X , la loi modélisée et F_n , la fonction de répartition empirique. Alors, la statistique KS est

$$D = \max_{x \geq u} |F_n(x) - F_X(x)|.$$

On teste plusieurs valeurs de u et on trouve celle qui minimise la statistique KS. Ici, on avait un vecteur de 500 valeurs de u allant de 70 milles à 1 million. On trouve une valeur de u optimale de 116 233. À noter que, lorsque l'on change les valeurs dans le vecteur ou le nombre de valeurs, le seuil trouvé change, mais on reste dans les mêmes valeurs si on ne descend pas la borne inférieure trop basse.

Intervalle de u testés ³	Seuil trouvé
$[2 \times 10^4, 1 \times 10^6]$	21964
$[4 \times 10^4, 1 \times 10^6]$	116954
$[7 \times 10^4, 1 \times 10^6]$	116233
$[7 \times 10^4, 5 \times 10^6]$	119399

TABEAU 6 – Seuil trouvé avec la méthode PLFIT selon différents vecteurs de u

Quelques résultats sont présentés au tableau 6. On voit que, si on prend une borne trop basse, la méthode sélectionne un seuil trop petit. On voit aussi qu'elle donne des résultats qui varient légèrement. La valeur

3. Les intervalles sont constitués de 500 seuils différents.

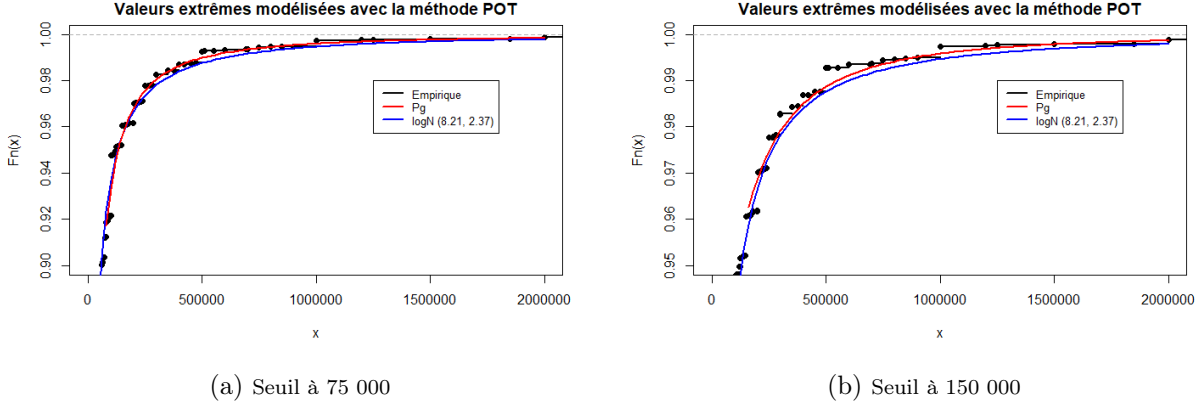


ILLUSTRATION 11 – Comparaison fonction de répartition de la loi Pareto généralisée, de la fonction de répartition de la loi lognormale avec la fonction de répartition empirique

sélectionnée le plus fréquemment est celle de 116 000. On opte donc pour un seuil de 116 000. Cependant, le test de Kolmogorov-Smirnov n'est pas l'idéal lorsque l'on a des données identiques, ce qui représente un désavantage de cette méthode avec la base de données canadienne. À noter qu'une autre statistique aurait pu être utilisée, par exemple celle de Anderson-Darling (A^2), mais comme mentionné par [Clauset et al., 2009], cette statistique est trop conservatrice pour la détection du seuil.

Double *bootstrap*

Il est aussi possible d'utiliser une approche par double *bootstrap*. Alors que la méthode avec la statistique KS est basée sur une approche plus statistique, la méthode par double *bootstrap* tente de produire une MSE optimale. La fonction `gomes` du package `tea` de R [Ossberger, 2020] est utilisée pour mettre en oeuvre la méthode de double *bootstrap* proposée par [Caeiro and Gomes, 2015]. Avec 1000 *bootstrap*, on trouve un seuil u de 150 000. Cette méthode est beaucoup plus stable, car même si on change le nombre de *bootstrap*, on trouve toujours 150 000. Pour cette raison, cette approche est privilégiée aux autres méthodes de détection automatique.

5.2.2 Paramètres

Pour trouver les paramètres pour la Pareto généralisée, on utilise la fonction `gpdFit` du package `tea` de R, [Ossberger, 2020], qui maximise la fonction de vraisemblance. Comme il était difficile de trancher entre un seuil à 75 000 ou un seuil à 150 000, on trouve les paramètres pour les deux seuils. Les paramètres sont présentés au tableau 7.

α	λ	u
1.41	117 038	75 000

(a) Seuil à 75 000

α	λ	u
1.94	387 881	150 000

(b) Seuil à 150 000

TABEAU 7 – Paramètres de la Pareto généralisée

Il est maintenant possible de comparer les fonctions de répartition des deux Pareto généralisée avec la fonction de répartition empirique. On observe sur l'illustration 11 que la Pareto généralisée avec le seuil à 75 000 va mieux chercher les observations élevées que celle avec le seuil à 150 000. Sans surprise, elle est aussi meilleure que la lognormale. De plus, avec le seuil à 75 000, on obtient $\alpha = 1.41$ ce qui est plus près de 1.5 que $\alpha = 1.94$ du seuil à 150 000. Pour cette raison, il est préférable de continuer avec le seuil à 75 000. On a les informations suivantes :

$F_n(u)$	Nb. obs. $> u$	% Sinistre $> u$
0.9122	1192	77 %

TABEAU 8 – Fonction de répartition empirique évaluée au seuil, nombre d’observations supérieures au seuil, pourcentage des montants de sinistre supérieurs au seuil

On peut donc voir à l’illustration 8 que plus de 90 % de nos données sont sous le seuil. Il nous reste 1192 données supérieures au seuil, donc amplement pour faire une analyse. On constate aussi que les 1192 montants de sinistre supérieurs à u représentent 77 % des sinistres totaux.

Cette approche a aussi été utilisée pour les bâtiments en construction et ceux en construction et rénovation. On a 127 bâtiments en construction et 579 bâtiments en construction et rénovation. Au tableau 9, on présente les autres statistiques descriptives.

Type	Min.	$1^{er}Qu.$	Médiane	Moyenne	Variance	$3^{er}Qu.$	Max.	n
Const.	1	1000	10 000	156 461	2.13×10^{11}	50 000	3 000 000	127
Const. et rénov.	1	1100	10 000	92 902	9.33×10^{10}	50 000	3 000 000	579

TABEAU 9 – Statistique descriptive pour les montants de sinistres pour les bâtiments en construction et pour les bâtiments en construction et rénovations

On applique la méthode POT et on trouve les paramètres présentés au tableau 10.

Type	α	λ
Const.	1.4853	387 183
Const. et rénov.	1.4461	226 104

TABEAU 10 – Paramètres de la Pareto généralisée pour les bâtiments en construction et en construction et rénovation

À des fins de validation, cette approche a été testée sur d’autres bases de données de sinistres d’incendies pour vérifier si l’on trouve aussi $\alpha \approx 1.5$. Les résultats sont présentés dans l’annexe A.

5.2.3 Mesure de risque

On utilise les données empiriques pour la portion sous le seuil et on utilise la Pareto généralisée pour la portion au-dessus du seuil. On a la fonction de répartition suivante :

$$F_X(x) = F_n(u) + (1 - F_n(u)) \left(1 - \left(1 + \frac{\xi}{\sigma}(x - u) \right)^{-\frac{1}{\xi}} \right).$$

De cette fonction de répartition, on trouve les expressions des mesures VaR et $TVaR$:

$$VaR_{\kappa}(X) = \frac{\sigma}{\xi} \left(\left(\frac{1 - \kappa}{F_n(u)} \right)^{-\xi} - 1 \right) + u, \quad \kappa \geq F_n(u)$$

$$TVaR_{\kappa}(X) = \frac{\sigma}{\xi} \left(\frac{1}{1 - \xi} \left(\frac{1 - \kappa}{F_n(u)} \right)^{-\xi} - 1 \right) + u, \quad \kappa \geq F_n(u).$$

Au tableau 11, les valeurs des mesures VaR et $TVaR$ sont présentées pour certains quantiles.

κ	0.95	0.99	0.995	0.999
VaR	132 641	506 275	855 356	2 774 855
$TVaR$	561 849	1 853 546	3 06 0361	9 69 6210

TABLEAU 11 – Valeurs pour les mesures VaR et $TVaR$

5.3 Raccordement de lois

Comme vue à la section 5.2 la Pareto généralisée offre une bonne adéquation pour les valeurs au-dessus de 75 000. Cependant, on ne veut pas garder la portion empirique pour les valeurs sous le seuil. Il est donc utile de raccorder des lois ensemble pour modéliser les montants de sinistre.

5.3.1 Définition

Définition 5 Soit X une variable qui suit une loi composite et soient $\{X_j, j = 1, \dots, n\}$ des variables non identiquement distribuées. Alors,

$$f_X(x) = \begin{cases} w_1 \frac{f_{X_1}(x)}{F_{X_1}(u_1) - F_{X_1}(u_0)} & , u_0 < x \leq u_1 \\ \dots & \\ w_n \frac{f_{X_n}(x)}{F_{X_n}(u_n) - F_{X_n}(u_{n-1})} & , u_{n-1} < x \leq u_n \end{cases}$$

où $w_i > 0$ et $\sum_{i=1}^n w_i = 1$.

Comme montré par [Brazauskas and Kleefeld, 2016] il est possible d'assurer la continuité et la dérivabilité aux points u_i . Cela a aussi comme effet de réduire le nombre de paramètres à optimiser. Par exemple pour une combinaison de deux lois on obtient.

$$\begin{cases} w \frac{f_{X_1}(u)}{F_{X_1}(u)} = (1 - w) \frac{f_{X_2}(u)}{F_{X_2}(u)} & \text{(continuité)} \\ w \frac{f'_{X_1}(u)}{F_{X_1}(u)} = (1 - w) \frac{f'_{X_2}(u)}{F_{X_2}(u)} & \text{(dérivabilité)} \end{cases} \quad (1)$$

5.3.2 Lognormale-Pareto généralisée

On teste la loi LN-PaG, on a X_1 qui est lognormale et X_2 qui est Pareto généralisée.

Définition 6 La fonction de densité de la loi composite lognormale-Pareto généralisée est

$$f_X(x) = \begin{cases} w \frac{f_{X_1}(x)}{F_{X_1}(u)} & , 0 < x \leq u \\ (1 - w) f_{X_2}(x) & , x > u \end{cases}$$

où

$$f_{X_1}(x) = \frac{1}{xr\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{r}\right)^2}$$

et

$$f_{X_2}(x) = \frac{1}{\sigma} \left(1 + \frac{\xi}{\sigma}(x - u)\right)^{\left(-\frac{1}{\xi} - 1\right)}.$$

Si on tient compte de la continuité et de la dérivabilité, on résout le système d'équations 1 et on trouve :

$$\begin{aligned} \mu &= \ln(u) - r^2 u \frac{1 + \xi}{\sigma}, \\ w &= \left(\sigma \frac{f_{X_1}(u)}{F_{X_1}(u)} + 1 \right)^{-1}. \end{aligned}$$

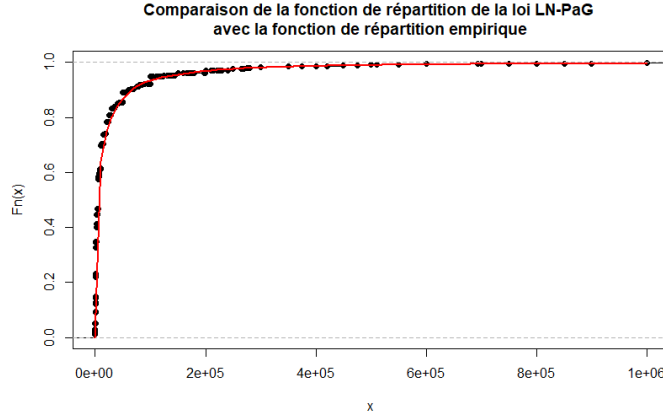


ILLUSTRATION 12 – Comparaison de la fonction de répartition de la loi LN-PaG avec la fonction de répartition empirique

Il a été observé que si l'on tient compte de la continuité l'optimisation est difficile même s'il y a moins de paramètres. En effet, si l'on change les valeurs de départ les paramètres obtenus sont significativement différents. Si l'on tient compte de la continuité et de la dérivabilité, les résultats obtenus ne sont pas bons, car le seuil trouvé par optimisation était toujours trop élevé. Pour ces raisons, seulement les paramètres pour la loi LN-PaG sans continuité et dérivabilité sont présentés.

Avec la méthode du maximum de vraisemblance, on trouve les paramètres suivants :

μ	r	α	λ	w
8.65	2.69	1.41	117 236	0.91

TABEAU 12 – Paramètres pour la loi composite lognormale-Pareto généralisée

On peut voir à l'illustration 12 que la loi LN-PaG à une bonne adéquation avec la fonction de répartition empirique.

5.3.3 Weibull-Pareto généralisée

Même si on a une bonne intuition que les montants de sinistres suivent une loi lognormale il peut être intéressant de tester d'autres lois pour la portion sous le seuil, par exemple la Weibull.

Définition 7 La fonction de densité de la loi composite Weibull-Pareto généralisée est

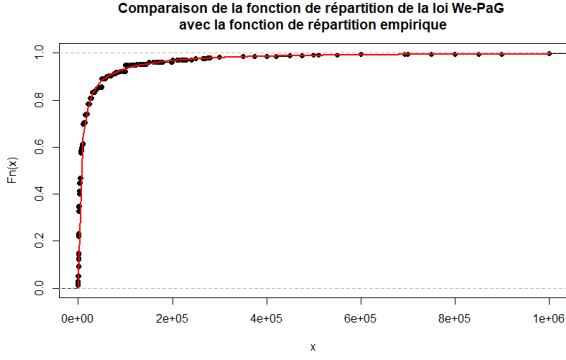
$$f_X(x) = \begin{cases} w \frac{f_{X_1}(x)}{F_{X_1}(u)} & , 0 < x \leq u \\ (1-w)f_{X_2}(x) & , x \geq u \end{cases}$$

où

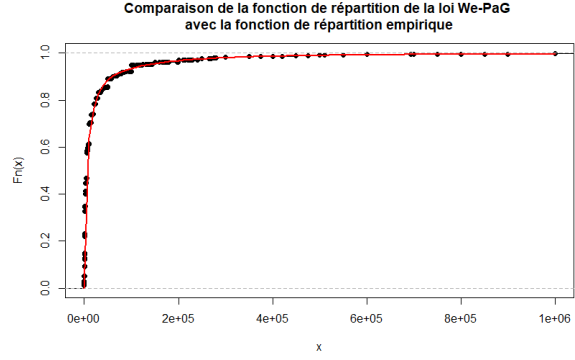
$$f_{X_1}(x) = \frac{\tau}{\beta^\tau} x^{\tau-1} e^{-\left(\frac{x}{\beta}\right)^\tau}$$

et

$$f_{X_2}(x) = \frac{1}{\sigma} \left(1 + \frac{\xi}{\sigma} (x-u) \right)^{\left(-\frac{1}{\xi}-1\right)}$$



(a) Sans continuité et dérivabilité



(b) Avec continuité

ILLUSTRATION 13 – Comparaison fonction de répartition de la loi We-PaG avec la fonction de répartition empirique

Il est toujours possible d'imposer la continuité et la dérivabilité au point u .

$$\beta = \left[\frac{1}{\tau u^\tau} \left(u \frac{\xi + 1}{\sigma} - \tau + 1 \right) \right]^{-\frac{1}{\tau}}$$

$$w = \left(\sigma \frac{f_{X_1}(u)}{F_{X_1}(u)} + 1 \right)^{-1}$$

Avec la loi We-PaG le cas avec continuité était assez stable, donc il vaut la peine d'être regardé. Encore fois, si l'on tient compte de la continuité et de la dérivabilité le seuil trouvé par optimisation est trop élevé et les résultats sont mauvais.

Avec la méthode du maximum de vraisemblance, on trouve les paramètres suivants :

τ	β	α	λ	w
0.55	7831	1.42	118 801	0.91

(a) Sans continuité et dérivabilité

τ	β	α	λ	w
0.54	8258	1.61	160 729	0.92

(b) Avec continuité

TABEAU 13 – Paramètres pour la loi composite lognormale-Pareto généralisée

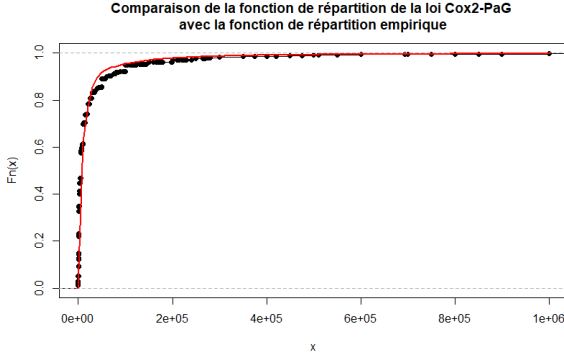
À l'illustration 13 on constate que la loi We-PaG suit quand même bien la fonction de répartition empirique tant dans les valeurs faibles que dans les valeurs élevées. Il est difficile de dire laquelle des deux courbes est la meilleure juste en se basant sur ces graphiques. Pour l'instant, la loi We-PaG et We-PaG avec continuité semble équivalente.

5.3.4 Coxienne-2-Pareto généralisée

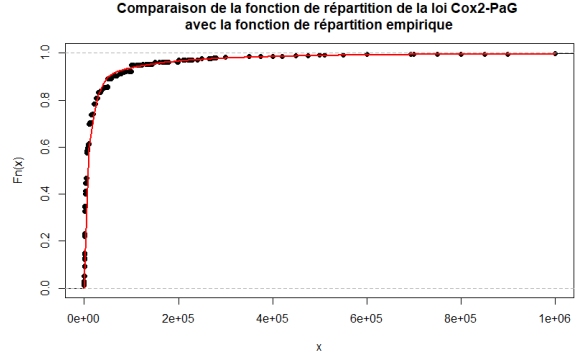
Le mélange d'Erlang a été testé, mais les résultats obtenus étaient visiblement moins bons que pour les autres lois testées. Cependant, il était intéressant de regarder un cas particulier de cette classe, soit la coxienne d'ordre deux. Présenté par [Klugman et al., 2013] cette distribution à l'avantage d'avoir des formes analytiques simples comparativement au mélange d'Erlang classique, voir annexe B.4. On utilise la coxienne d'ordre deux pour limiter le nombre de paramètres.

Définition 8 La fonction de densité de la loi composite Coxienne-2-Pareto généralisée est

$$f_X(x) = \begin{cases} w \frac{f_{X_1}(x)}{F_{X_1}(u)} & , 0 < x \leq u \\ (1 - w) f_{X_2}(x) & , x > u \end{cases}$$



(a) Sans continuité et dérivabilité



(b) Avec continuité

ILLUSTRATION 14 – Comparaison fonction de répartition de la loi We-PaG avec la fonction de répartition empirique

où

$$f_{X_1}(x) = p\beta_1 e^{-\beta_1 x} + (1-p) \left(\frac{\beta_2}{\beta_2 - \beta_1} \beta_1 e^{-\beta_1 x} + \frac{\beta_1}{\beta_1 - \beta_2} \beta_2 e^{-\beta_2 x} \right)$$

et

$$f_{X_2}(x) = \frac{1}{\sigma} \left(1 + \frac{\xi}{\sigma}(x-u) \right)^{(-\frac{1}{\xi}-1)}$$

Il est toujours possible de tenir compte de la continuité. Cependant pour la dérivabilité l'expression est très complexe et difficilement travaillable.

$$w = \left(\sigma \frac{f_{X_1}(u)}{F_{X_1}(u)} + 1 \right)^{-1}$$

On utilise le maximum de vraisemblance pour trouver les paramètres présentés au tableau 14.

p	β_1	β_2	α	λ	w
0.40	1.32×10^{-3}	6.17×10^{-5}	1.41	118 442	0.94

(a) Sans continuité et dérivabilité

p	β_1	β_2	α	λ	w
0.41	1.18×10^{-3}	5.34×10^{-5}	1.98	261 116	0.93

(b) Avec continuité

TABEAU 14 – Paramètres pour la loi composite Coxienne-2-Pareto généralisée

À l'illustration 14 on constate que la loi Cox2-PaG sans continuité semble être un peu moins bonne, que la loi Cox2-PaG avec continuité, dans les valeurs sous le seuil. L'adéquation semble généralement meilleure en tenant compte de la continuité ce qui est surprenant comme le α est de 1.98 et qu'on s'attend davantage à avoir un α proche de 1.5.

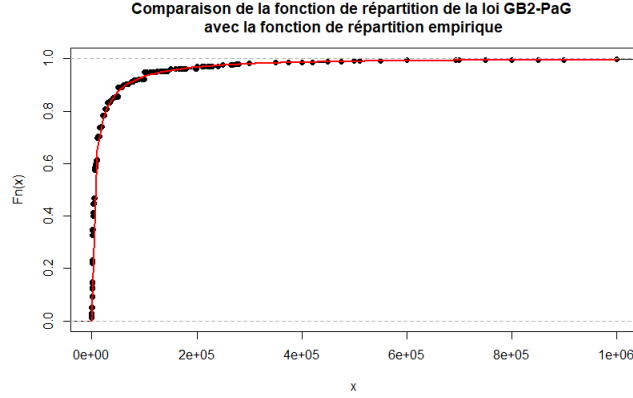


ILLUSTRATION 15 – Comparaison fonction de répartition de la loi We-PaG avec la fonction de répartition empirique

5.3.5 Bêta généralisée de type 2-Pareto généralisée

La loi bêta généralisée de type 2 (GB2) est aussi testée pour la portion sous le seuil. Présentée dans [Cummins et al., 1990] la motivation derrière l'utilisation de la GB2 est qu'elle est très flexible, comme elle a 4 paramètres.

Définition 9 La fonction de densité de la loi composite bêta généralisée de type 2-Pareto généralisée est

$$f_X(x) = \begin{cases} w \frac{f_{X_1}(x)}{F_{X_1}(u)} & , 0 < x \leq u \\ (1-w)f_{X_2}(x) & , x > u \end{cases}$$

où

$$f_{X_1}(x) = \frac{ax^{ap-1}}{b^{ap}I(p, q)(1 + (\frac{x}{b})^a)^{p+q}}$$

et

$$f_{X_2}(x) = \frac{1}{\sigma} \left(1 + \frac{\xi}{\sigma}(x-u) \right)^{\left(-\frac{1}{\xi}-1\right)}$$

Il est toujours possible de tenir compte de la continuité. Cependant pour la dérivabilité l'expression est très complexe et difficilement travaillable.

$$w = \left(\sigma \frac{f_{X_1}(u)}{F_{X_1}(u)} + 1 \right)^{-1}$$

On utilise le maximum de vraisemblance pour trouver les paramètres présentés au tableau 15. Pour trouver les paramètres initiaux pour l'optimisation on utilise la fonction `m1.gb2` du package GB2 de R, [Graf and Nedyalkova., 2022].

a	b	p	q	α	λ	w
0.47	19 990	1.57	2.89	1.42	120 149	0.91

TABEAU 15 – Paramètres pour la loi composite bêta généralisée de type 2-Pareto généralisée

On choisit de garder seulement la loi GB2-PaG avec continuité comme les paramètres sont pratiquement identiques sans ou avec la continuité. De plus, en gardant celle-là on a seulement besoin d'optimiser 6 paramètres au lieu de 7.

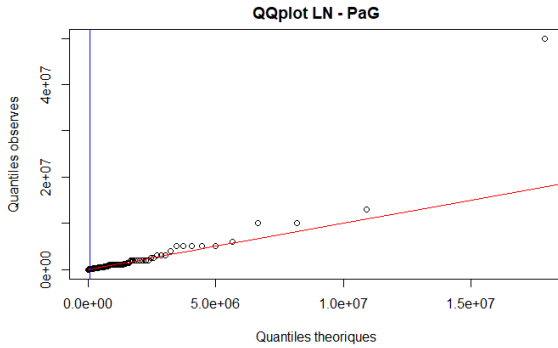
À l'illustration 15 on constate que l'adéquation est très bonne sur tout le domaine. Il faudra donc utiliser des tests

6 Test des lois

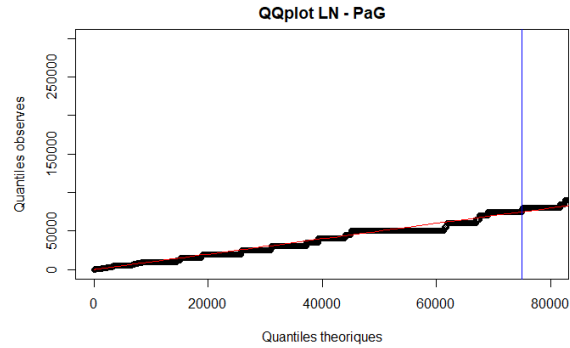
Afin de sélectionner la loi offrant la description la plus adéquate aux données canadiennes, on a recours à un test d'adéquation graphique et aux tests quantitatifs, Cramer-Von Mises (W^2), Anderson-Darling (A^2), AIC et BIC.

6.1 Test graphique

Comme à la section 4, on analyse les *QQplots* des raccordements de lois trouvées. Une bonne loi aura des points très près de la droite rouge.



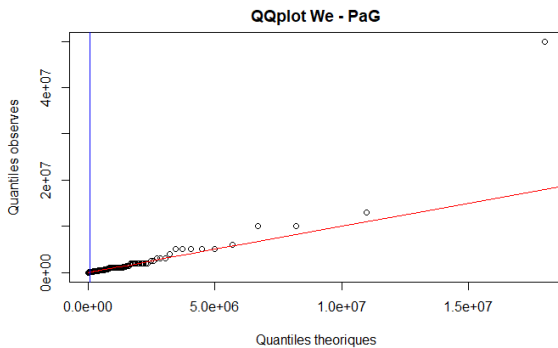
(a) Distribution entière



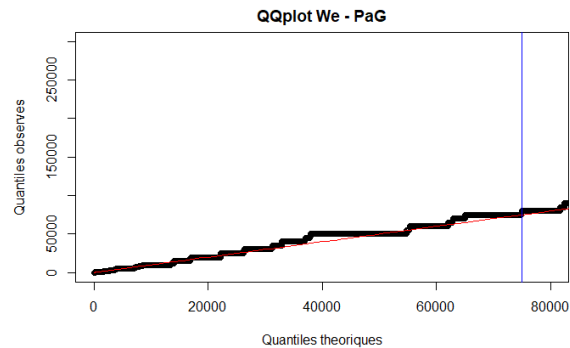
(b) Portion lognormale

ILLUSTRATION 16 – *QQplot* lognormale-Pareto généralisée

On observe à l'illustration 16b que pour les valeurs sous 75 000, ligne verticale bleue, l'adéquation de la loi est bonne. Pour les valeurs passées 75 000, on trouve aussi une assez bonne adéquation, même pour les valeurs extrêmes qui restent proches de la droite. Il semble y avoir peu de données sous les 75 000, mais on rappelle qu'en réalité plus de 90 % de nos données sont sous 75 000.



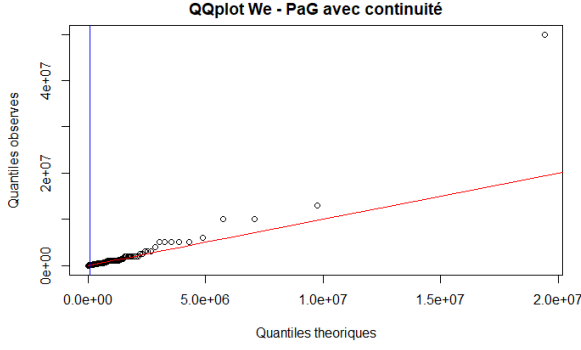
(a) Distribution entière



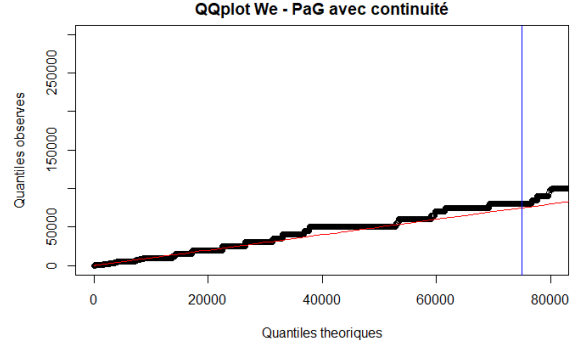
(b) Portion lognormale

ILLUSTRATION 17 – *QQplot* pour la loi Weibull-Pareto généralisée

La loi We-PaG présentée à l'illustration 17 offre une bonne adéquation. Il est difficile de dire quelles lois est la plus adéquate entre la loi LN-PaG et la loi We-PaG, car leur adéquation semble être très similaire.



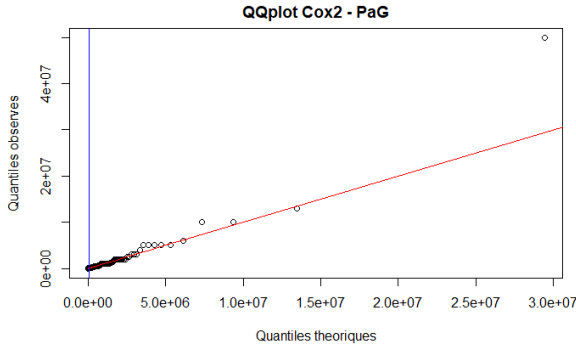
(a) Distribution entière



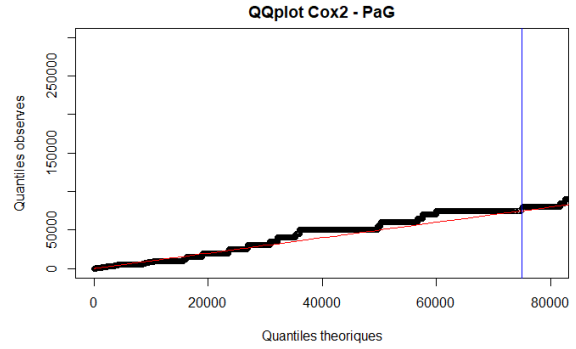
(b) Portion lognormale

ILLUSTRATION 18 – *QQplot* pour la loi Weibull-Pareto généralisée avec continuité

Lorsqu'on tient compte de la continuité, illustration 18, avec la loi We-PaG l'adéquation est légèrement moins bonne que pour les deux autres lois. On sous-estime plus les valeurs tant dans la portion modélisée par la loi Weibull que dans la portion modélisées par la loi Pareto généralisée.



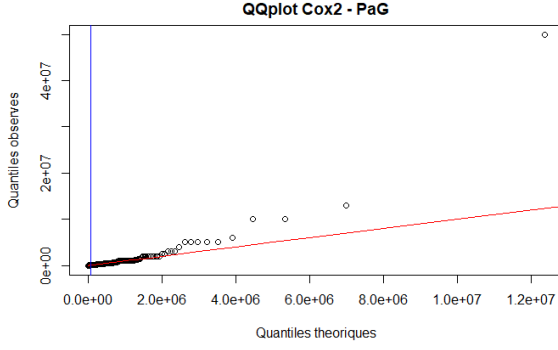
(a) Distribution entière



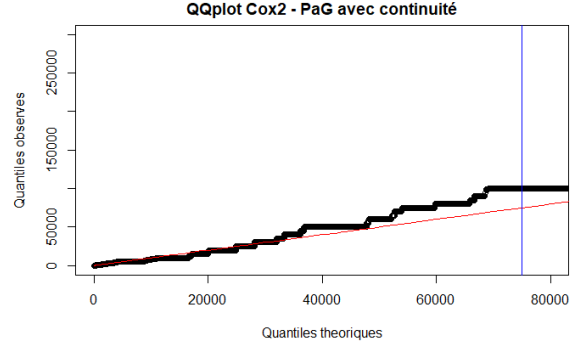
(b) Portion Coxienne-2

ILLUSTRATION 19 – *QQplot* pour la loi Coxienne-2-Pareto généralisée

On remarque à l'illustration 19 que l'adéquation pour la portion sous le seuil est moins bonne avec la loi Coxienne-2 que les autres lois, on sous-estime beaucoup plus les valeurs. Pour la portion au-dessus du seuil les résultats sont assez proches de la loi LN-PaG et We-PaG comme les paramètres de la loi Pareto généralisée sont presque identiques.



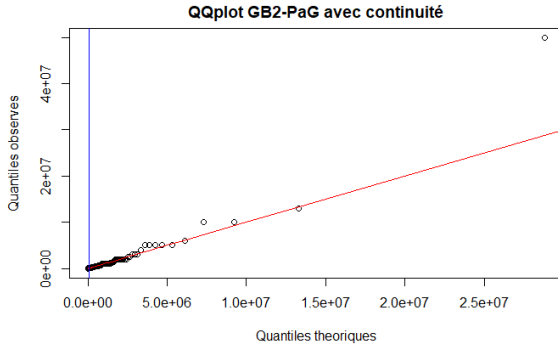
(a) Distribution entière



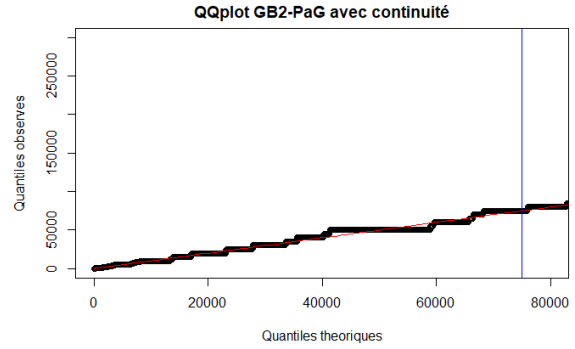
(b) Portion Coxienne-2

ILLUSTRATION 20 – *QQplot* pour la loi Coxienne-2-Pareto généralisé avec continuité

Cependant lorsqu'on tient compte de la continuité avec la loi coxienne-2 l'adéquation est aussi moins bonne pour la portion Pareto généralisée. Il était prévisible qu'on allait sous-estimer cette portion comme on obtient un α de 2.



(a) Distribution entière



(b) Portion bêta généralisée de type 2

ILLUSTRATION 21 – *QQplot* pour la loi GB2-Pareto généralisée avec continuité

Pour la loi GB2-PaG, on observe que l'adéquation est très bonne tant pour la portion modélisée par la loi bêta généralisée de type 2 que dans la portion modélisée par la loi Pareto généralisée.

Comme certaines lois ont des adéquations assez proches, il faudra utiliser des tests quantitatifs pour savoir lequel est le meilleur. Cependant, on peut déjà observer que les lois LN-PaG, We-PaG et GB2-PaG avec continuité semblent être les meilleurs.

6.2 Test quantitatif

Pour être en mesure de départager les lois qui ont une adéquation similaire les statistiques W^2 et A^2 sont utilisées. La statistique W^2 accorde un poids égal le long de la distribution tandis que la statistique A^2 accorde plus de poids aux observations dans la queue, ce qui est important dans le cas présent comme on cherche à modéliser les valeurs extrêmes. Les critères classiques AIC et BIC sont aussi performés.

Loi	W^2	A^2	AIC	BIC
LN-PaG	6.02	33.24	284 503	284 465
We-PaG	6.18	34.22	284 502	284 464
We-PaG avec cont.	6.56	36.88	284 529	284 499
Cox2-PaG	10.70	86.64	285 494	285 449
Cox2-PaG avec cont.	13.75	85.41	285 516	285 478
GB2-PaG avec cont.	4.74	24.61	284 319	284 274

TABLEAU 16 – Résultats des tests W^2 , A^2 , AIC et BIC

Pour les tests W^2 et A^2 on veut la statistique la plus petite possible. On voit dans le tableau 16 que la meilleure loi selon ces tests est la loi GB2-PaG. Pour le test AIC et BIC on veut aussi le résultat le plus petit. La loi GB2-PaG est encore la meilleure selon ces tests.

La loi à choisir serait donc la loi GB2-PaG. La loi LN-PaG ne serait pas un mauvais choix non plus, comme elle est la seconde meilleure loi selon les tests effectués. De plus, le graphique de log-densité 6b et le *QQplot* 8b lognormale donnait une bonne intuition que la distribution pouvait être lognormale.

7 Informations sur les lois LN-PaG et GB2-PaG

On s'intéresse plus particulièrement aux valeurs extrêmes de la distribution. Pour cette raison il peut être pertinent de trouver un intervalle de confiance pour les paramètres de la portion modélisée par la loi Pareto généralisée. On estime l'intervalle de confiance avec la méthode *Bias-Corrected and Accelerated Bootstrap* (BC_a).

	Valeur	Intervalle
$\hat{\xi}$	0.71	(0.61, 0.81)
$\hat{\sigma}$	83219	(74635, 92024)

TABLEAU 17 – Intervalle de confiance estimé avec 14 000 *bootstrap*

On observe au tableau 17 que les intervalles sont assez grands. Les fonctions `boot` et `boot.ci` du package `boot` de R, [Canty and Ripley, 2021], ont été utilisés pour faire le *bootstrap*. Les grands intervalles sont dus au fait que si les grosses données ne se retrouvent pas dans le sous-échantillon les résultats changent considérablement.

Maintenant que la loi est sélectionnée, il est intéressant d'aller calculer certaines quantités et de les comparer aux quantités empiriques.

Loi	$E[X]$	$\text{Var}(X)$
empirique	40296	238004336993
LN-PaG	41562	∞
GB2-PaG	41077	∞

TABLEAU 18 – Espérance et variance

Au tableau 18 on constate que l'espérance n'est pas loin de celle empirique. De son côté la variance n'existe pas. Ce n'est pas surprenant comme la variance empirique était extrêmement élevée.

Loi \ κ	0.90	0.95	0.99	0.995	0.999
empirique	60000	125000	500000	1000000	2715000
LN-PaG	66841	132752	506608	855718	2774167
GB2-PaG	65445	134862	513663	865066	2780455

(a) Valeurs de la mesure VaR

Loi \ κ	0.90	0.95	0.99	0.995	0.999
empirique	321120	562888	2227303	4720000	9071428
LN-PaG	327028	561683	1851936	3056788	9677751
GB2-PaG	328020	562678	1841745	3028301	9495836

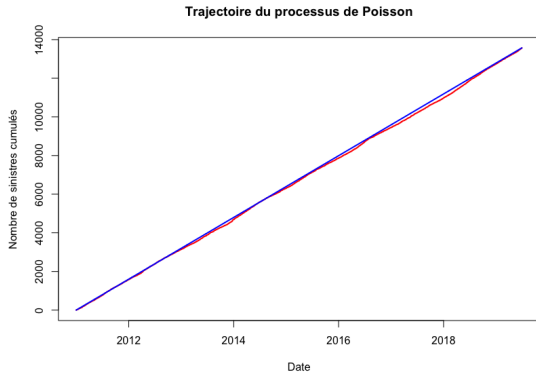
(b) Mesure de la mesure $TVaR$

TABLEAU 19 – Valeurs de la mesure VaR et de la mesure $TVaR$

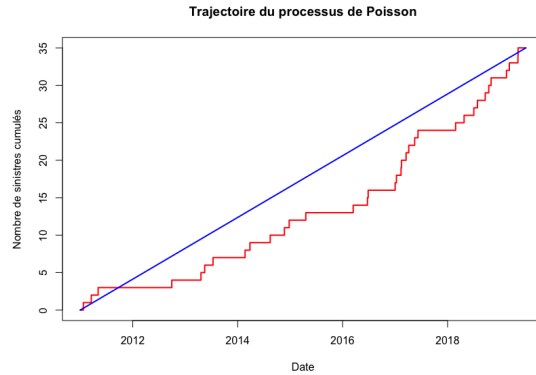
On remarque au tableau 19b qu'on sous-estime beaucoup les valeurs de la mesure $TVaR$ pour les quantiles 0.99 et 0.995 alors qu'on surestime le quantile 0.999 et ce pour les deux lois. Les formes analytiques des expressions sont présentées dans les section C.1 et C.3 de l'annexe.

8 Modélisation du processus d'avènements des incendies

Avec la variable *TFS Alarm Time*, qui donne l'heure à laquelle le système d'alarme s'est déclenché, il est possible de faire une analyse graphique pour la fréquence d'arrivée des sinistres incendies.



(a) Pour tous les sinistres



(b) Pour les sinistres au-dessus de 1 million

ILLUSTRATION 22 – Test d'adéquation graphique pour la fréquence d'arrivée des sinistres incendies

On utilise la méthode du maximum de vraisemblance pour estimer le paramètre lambda du processus de Poisson.

$$\hat{\lambda} = \frac{\sum_{i=1}^n x_i}{nt}$$

On choisit de trouver un lambda sur une base quotidienne. On trouve pour tous les sinistres incendies $\lambda = 4.37$ et pour les sinistres incendies supérieurs à 1 million on trouve $\lambda = 0.011$.

À l'illustration 22a la ligne bleue représente la moyenne du nombre de sinistres cumulés en théorie et la ligne rouge représente le nombre de sinistres cumulés observés avec la base de données de sinistres incendies.

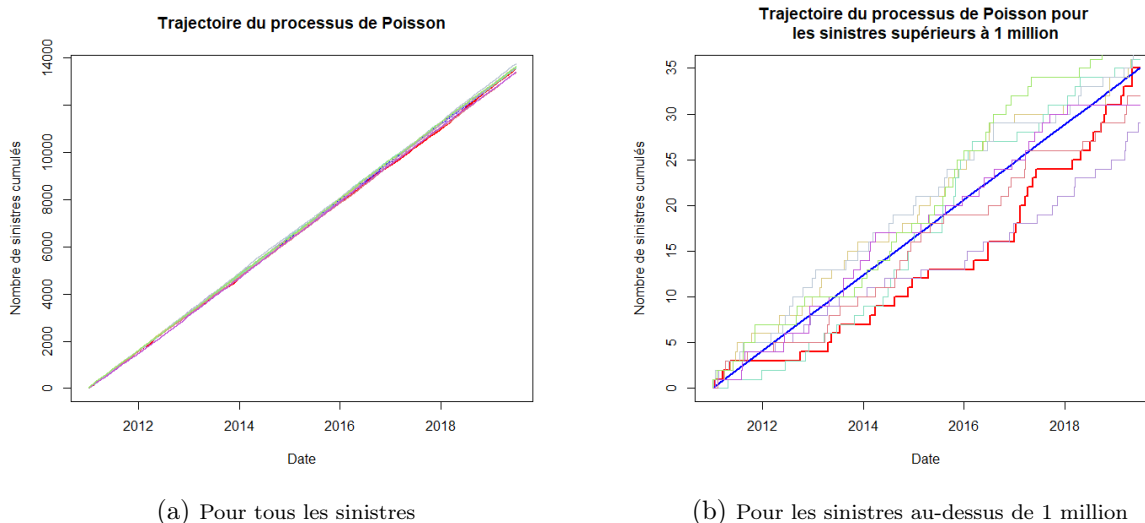


ILLUSTRATION 23 – Pour les sinistres supérieurs à 1 million

canadienne. On constate qu'on est très proche de la valeur moyenne théorique, ce qui est un bon indicatif que la fréquence d'arrivée de sinistres incendies de la base de données canadienne est un processus de Poisson. De plus, si on simule d'autres parcours en utilisant le même λ , on constate effectivement à l'illustration 23a que les autres parcours aussi semblent suivre un processus de Poisson homogène. Il est aussi intéressant de regarder la fréquence d'arrivée des sinistres au-dessus d'un certain montant. À l'illustration 22b on peut voir que les fréquences d'arrivée des sinistres ne semblent pas suivre la valeur théorique. Cependant, il faut noter qu'on a seulement 35 sinistres au-dessus de 1 million, ce qui n'est pas beaucoup pour avoir une analyse juste. Pour vérifier s'il est plausible de dire que les sinistres supérieurs à 1 million suivent un processus de Poisson homogène on utilise encore la simulation de d'autres parcours de processus de Poisson avec le λ trouvé préalablement. À l'illustration 23 on remarque que la situation qu'on a avec les sinistres de la base de données canadienne semble être un chemin plausible comme il n'est pas complètement à l'écart des autres. On ne peut donc pas écarter l'hypothèse que les sinistres au-dessus de 1 million suivent un processus de Poisson.

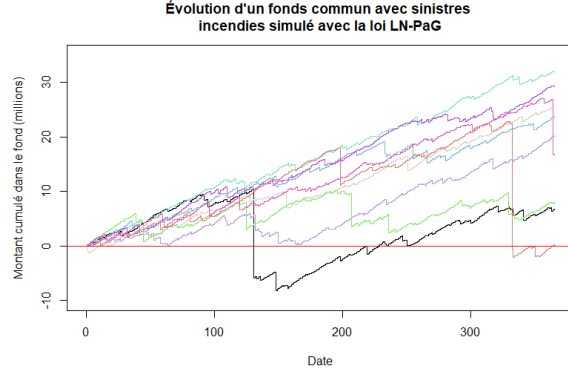
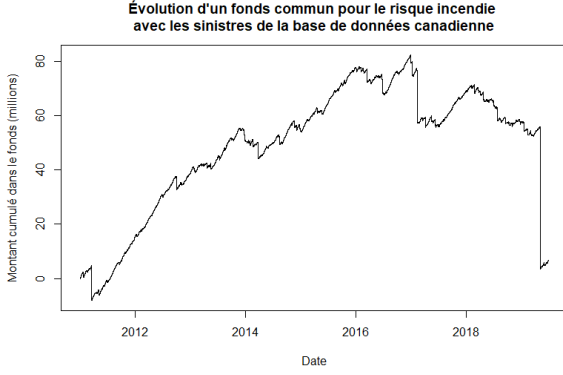
9 Étude avec une mutuelle fictive

Maintenant qu'on a modélisé la sévérité et qu'on a une bonne idée de distribution de la fréquence, on peut regarder ce qui se passe si l'on crée un fonds commun pour les sinistres incendies. L'objectif étant de voir si l'on crée une mutuelle d'assurance comment pourrait évoluer la valeur du fonds.

On commence par regarder le cas où l'on demande seulement la prime pure quotidiennement et sans mettre de fonds initiaux. On prend $\lambda = 4.37$ et l'espérance de la loi LN-PaG, 41 561, et on obtient une prime pure de 181 770 \$ CAD par jour.

L'illustration 24a présente l'évolution de ce fonds. On constate que dès le début on passe sous la barre de zéro. Ensuite, le fonds reprend de la valeur jusqu'à un peu plus de 80 millions, mais le sinistre de 50 millions le fait considérablement redescendre et il se termine à 6 625 971 \$ CAD. Il est important de constater qu'il ne faut pas sortir de l'argent du fonds, car il est toujours possible qu'un sinistre très élevé se produise. Il serait possible de pousser plus loin en regardant qu'est qui se passe si l'on charge la prime pure plus une ne surcharge ou s'il l'on met un montant initial dans le fonds.

On a regardé l'évolution du fonds avec les données canadiennes de sinistres incendies. Maintenant, on peut utiliser la simulation de Monte-Carlo avec la loi LN-PaG pour regarder 10 évolutions d'un fonds commun. On utilise la même prime quotidienne et l'on simule sur l'horizon de 1 an. L'illustration 24b l'évolution de 10 fonds simulés à partir de la loi LN-PaG. On constate que le moment d'arriver des sinistres à une



(a) Évolution d'un fonds avec les données canadiennes (b) Évolution d'un fonds avec des simulation de la loi LN-PaG

ILLUSTRATION 24 – Évolution d'un fonds d'urgence pour le risque incendie avec les sinistres de la base de données canadienne et des simulations de la loi LN-PaG, en chargeant la prime pure quotidiennement

incidence sur l'évolution du fond. En effet, si un sinistre très élevé survient au début de l'année le fonds n'est pas encore assez rempli pour y faire face et il tombe sous la barre de 0. Il n'est pas réaliste de seulement charger la prime pure, car le fond tombe en-dessous de zéro rapidement si un sinistre très élevé survient. On calcule la probabilité de ruine qui nous donnera la probabilité que la compagnie ne puisse plus faire face à ses engagements au cours de la prochaine période, ici 1 an.

Définition 10 Soit une variable aléatoire S qui représentant le coût d'un portefeuille. Alors la probabilité de ruine est

$$\zeta(x) = \Pr\left(\inf_{t \geq 0} U_t < 0 | U_0 = x\right)$$

où U_t est le revenu de prime et x est la mise de fonds initial.

En utilisant 100 milles simulations, on approxime la probabilité de ruine et on obtient 0.27. Il y a donc une probabilité de 27 % de ne pas être en mesure de tenir ces engagements en demandant seulement la prime pure sans mise de fonds initiale.

10 Conclusion

On a été en mesure de modéliser les données canadiennes de sinistre incendie à l'aide de différent raccordement de lois. Il a fallu trouver le seuil le plus approprié pour les données. Finalement, c'est en regardant le *QQplot* Pareto qu'on a trouvé le seuil à 75 000. On conclut que les lois lognormale-Pareto généralisée et Bêta généralisée de type 2-Pareto généralisée sont les lois les plus adaptées, dans celles testées, pour les données canadiennes de sinistres incendies. On privilégie dans ce rapport l'utilisation de la loi LN-PaG pour sa simplicité. On a ensuite utilisé cette loi pour calculer l'espérance, les mesures *VaR* et *TVaR*.

Pour la fréquence, on a fait un test d'adéquation pour venir à la conclusion que l'hypothèse selon laquelle la fréquence suit un processus de Poisson homogène était plausible. On a ensuite utilisé le processus de Poisson homogène et la loi LN-PaG, pour simuler des sinistres incendie et regarder l'évolution d'un fond commun sur une période de 1 an. On a constaté que charger seulement la prime pure n'était pas envisageable. Il aurait été intéressant de faire la même analyse, mais en choisissant la loi GB2-PaG qui était tout de même mieux, selon les tests, pour modéliser les données canadiennes de risque incendies.

Dans une optique d'aller plus loin, il faudrait faire davantage de tests pour s'assurer que la distribution de la fréquence est bien un processus de Poisson homogène. De plus, il faudrait tester l'évolution d'un fonds commun avec différents paramètres et horizon temporel. Par exemple, on pourrait tester différents scénarios en changeant une prime et mettant une mise de fonds initiales différente pour chaque scénarios. Il serait aussi possible de tester sur différents horizons temporels, par exemple 1 semaine, 1 mois ou 5 ans. Il y aurait aussi possibilité de rajouter de la réassurance dans ces scénarios. On pourrait tester ce qui se passe si l'on décide de réassurer les sinistres supérieurs à 1 million.

Dans ce présent rapport, on s'intéresse aux sinistres incendies en général. On ne souhaitait pas faire de distinction entre une maison en pierre ou une maison en bois par exemple. Dans le but ultime de modéliser les montants de sinistres de chantier en bois, il serait intéressant d'aller plus loin et de sélectionner seulement les sinistres qui proviennent d'une construction en bois.

11 Remerciements

Je remercie sincèrement le Centre Interdisciplinaire de Modélisation Mathématique de l'Université Laval (CIMMUL) pour une **Bourse de premier cycle du CIMMUL** qui a financé une partie des travaux de ce rapport et qui m'a permis de travailler à temps plein durant l'été.

Je tiens aussi à remercier la Prof. Hélène Cossette et le Prof. Etienne Marceau pour m'avoir donné cette merveilleuse opportunité. Vous m'avez permis de découvrir de nouveaux aspects de la science actuarielle tout en me permettant de m'améliorer.

Je suis aussi reconnaissant envers Christopher Blier-Wong, étudiant au doctorat en actuariat, pour les échanges fructueux et pour ses conseils qui m'ont permis de produire un rapport de meilleure qualité. Je suis également reconnaissant envers André Orelieu Chuisseu Tchuisseu, étudiant au doctorat en actuariat, pour avoir pris le temps de lire le rapport, d'avoir aidé avec la correction et la vérification d'une bonne portion du code informatique.

Je remercie aussi le CIRCERB de m'avoir ouvert leurs portes et de m'avoir permis de participer à différentes activités, qui m'ont permis d'en apprendre beaucoup sur les enjeux de la construction en bois.

Merci à l'École d'actuariat de m'avoir fourni des locaux à l'Université Laval, ce qui a permis de faciliter des échanges avec mes collègues.

Laboratoire ACT&RISK pour le cadre de travail.

Annexe

A Autres données

A.1 Données suédoises de sinistres incendies

Pour valider les résultats obtenus avec la base de données de Toronto, la base de données suédoises des montants de sinistres a été utilisée. Elle a été initialement utilisée par [Embrechts and Schmidli, 1994], pour faire une analyse de la fonction d'excès moyen dans le contexte de la modélisation d'événement extrême. La base de données contient de l'information sur 218 montants de réclamations qui se sont produits en 1982. Pour faciliter l'analyse, les trois montants de 0 ont été retirés. Donc, $n = 215$ données ont réellement servi pour l'analyse. Le tableau 20 présente les statistiques descriptives de la base de données.

Min.	1 ^{er} Qu.	Médiane	Moyenne	Variance	3 ^{er} Qu.	Max.
0.088	0.627	1	2.314	17.485	2	34

TABEAU 20 – Statistiques descriptives de la base de données suédoise de $n = 215$ sinistres incendies. Les montants de sinistres sont en unité de millions SKR.

On applique la méthode POT aux montants de sinistres de la base de données suédoises et on obtient les paramètres présentés au tableau 21.

α	λ
1.4308	1.925

TABEAU 21 – Paramètres de la Pareto généralisée pour les montants de réclamations suédois

On obtient bien un α proche de 1.5 et proche du α trouvé pour la base de données de canadienne, qui était de 1.41.

A.2 Données norvégiennes de sinistres incendies

Il y avait aussi la possibilité de se valider avec la base de données norvégienne de réclamation de sinistre. Cette base de données a été utilisée par Beirlant, Teugels and Vynckier (1996). Elle contient $n = 9181$ montants de sinistres incendies. Il faut noter qu'une franchise de 500 \$ a été appliquée, donc on a les montants de sinistres supérieurs à celle-là. Le tableau 22 présente les statistiques descriptives de la base de données.

Min.	1 ^{er} Qu.	Médiane	Moyenne	Variance	3 ^{er} Qu.	Max.
500	700	1020	2217	60 217 194	1800	465 365

TABEAU 22 – Statistiques descriptives de la base de données norvégienne de $n = 9181$ sinistres incendies

On applique la méthode POT aux montants de sinistres incendies de la base de données suédoises et on obtient les paramètres présentés au tableau 23.

α	λ
1.3766	1846

TABEAU 23 – Paramètres de la Pareto généralisée pour les montants de réclamations norvégien

On obtient bien un α proche de 1.5 et proche du α trouvé pour la base de données canadienne, qui était de 1.41.

B Caractéristiques des lois de probabilité

B.1 Loi lognormale

- Notation : $X \sim LN(\mu, \sigma^2)$
- Paramètres : $-\infty < \mu < \infty, \sigma^2 > 0$
- Support : $x \in \mathbb{R}^+$
- Fonction de densité : $f(x) = \frac{1}{x\sqrt{2\pi}\sigma} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$
- Fonction de répartition : $F(x) = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right)$
- Espérance : $E[X] = e^{\mu + \frac{\sigma^2}{2}}$
- Variance : $\text{Var}(X) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$
- Fonction génératrice des moments : forme non analytique
- Moments d'ordre k : $E[X^k] = e^{k\mu + k^2 \frac{\sigma^2}{2}}$
- Espérance tronquée : $E[X \times 1_{\{X \leq d\}}] = \exp(\mu + \sigma^2/2) \Phi\left(\frac{\ln d - \mu - \sigma^2}{\sigma}\right)$
- Mesure VaR : $VaR_\kappa(X) = \exp(\mu + \sigma VaR_\kappa(Z))$
- Mesure $TVaR$:

$$TVaR_\kappa(X) = \frac{1}{1 - \kappa} e^{\mu + \sigma^2/2} (1 - \Phi(VaR_\kappa(Z) - \sigma))$$

- Fonction *stop-loss* :

$$\pi_d(X) = e^{\mu + \sigma^2/2} (1 - \Phi\left(\frac{\ln d - \mu - \sigma^2}{\sigma}\right)) - d[1 - \Phi\left(\frac{\ln d - \mu}{\sigma}\right)]$$

- Fonction d'excès-moyen :

$$e_d(X) = \frac{1}{[1 - \Phi\left(\frac{\ln d - \mu}{\sigma}\right)]} e^{\mu + \sigma^2/2} (1 - \Phi\left(\frac{\ln d - \mu - \sigma^2}{\sigma}\right)) - d$$

- Espérance limitée :

$$E[\min(X; d)] = e^{\mu + \sigma^2/2} \Phi\left(\frac{\ln d - \mu - \sigma^2}{\sigma}\right) + d[1 - \Phi\left(\frac{\ln d - \mu}{\sigma}\right)]$$

- Loi associée : $X = e^Y$, où $Y \sim N(\mu, \sigma^2)$, impliquant $E[X^k] = M_Y(k)$

B.2 Loi de Weibull

- Notation : $X \sim We(\tau, \beta)$
- Paramètres : $\tau > 0, \beta > 0$
- Support : $x \in \mathbb{R}^+$
- Fonction de densité : $f(x) = \beta\tau(\beta x)^{\tau-1} e^{-(\beta x)^\tau}$
- Fonction de répartition : $F(x) = 1 - e^{-(\beta x)^\tau}$
- Fonction de survie : $\bar{F}(x) = e^{-(\beta x)^\tau}$
- Espérance : $E[X] = \frac{1}{\beta} \Gamma\left(1 + \frac{1}{\tau}\right)$
- Variance : $\text{Var}(X) = \frac{1}{\beta^2} \Gamma\left(1 + \frac{2}{\tau}\right) - \left(\frac{1}{\beta} \Gamma\left(1 + \frac{1}{\tau}\right)\right)^2$
- Fonction génératrice des moments (pour $\alpha > 1$) :

$$\mathcal{M}_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{\beta^k k!} \Gamma\left(1 + \frac{k}{\tau}\right)$$

- Moments d'ordre k : $E[X^k] = \frac{1}{\beta^k} \Gamma\left(1 + \frac{k}{\tau}\right)$
- Espérance tronquée : $E[X \times 1_{\{X \leq d\}}] = \frac{1}{\beta} \Gamma\left(1 + \frac{1}{\tau}\right) H(d^\tau; 1 + \frac{1}{\tau}, \beta^\tau)$
- Mesure VaR : $VaR_\kappa(X) = \frac{1}{\beta} (-\ln(1 - \kappa))^{\frac{1}{\tau}}$
- Mesure $TVaR$: $TVaR_\kappa(X) = \frac{1}{\beta(1-\kappa)} \Gamma\left(1 + \frac{1}{\tau}\right) \bar{H}(-\ln(1 - \kappa); 1 + \frac{1}{\tau}, 1)$
- Fonction *stop-loss* : $\pi_d(X) = \frac{1}{\beta} \Gamma\left(1 + \frac{1}{\tau}\right) \bar{H}(d^\tau; 1 + \frac{1}{\tau}, \beta^\tau) - d e^{-(\beta d)^\tau}$
- Fonction d'excès-moyen : $e_d(X) = \frac{e^{(\beta d)^\tau}}{\beta} \Gamma\left(1 + \frac{1}{\tau}\right) \bar{H}(d^\tau; 1 + \frac{1}{\tau}, \beta^\tau) - d$
- Espérance limitée : $E[\min(X; d)] = \frac{1}{\beta} \Gamma\left(1 + \frac{1}{\tau}\right) H(d^\tau; 1 + \frac{1}{\tau}, \beta^\tau) + d e^{-(\beta d)^\tau}$
- Cas particuliers :
 - la loi exponentielle est un cas particulier de la loi Weibull avec $\tau = 1$;
 - la loi Raleigh est un cas particulier de la loi Weibull avec $\tau = 2$.

B.3 Loi b eta

- Notation : $X \sim \text{B eta}(\alpha, \beta)$
- Param etres : $\alpha > 0, \beta > 0$
- Support : $x \in [0, 1]$
- Fonction b eta incompl te : $I(x; \alpha, \beta) = \int_0^x u^{\alpha-1} (1-u)^{\beta-1} du, x \in [0, 1]$
- Fonction b eta compl te : $I(\alpha, \beta) = I(1; \alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$
- Fonction de densit  : $f_X(x) = \frac{1}{I(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \times 1_{\{x \in [0, 1]\}}$
- Fonction de r partition : $F_X(x) = \frac{I(x; \alpha, \beta)}{I(\alpha, \beta)}$, not e $B(x; \alpha, \beta)$

$$\text{- Si } \beta = 1, F(x) = \begin{cases} 0, & x < 0 \\ x^\alpha, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$$

$$\text{- Si } \alpha = 1, F(x) = \begin{cases} 0, & x < 0 \\ 1 - (1-x)^\beta, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$$

$$\text{- Si } \alpha, \beta \in \mathbb{N}^+,$$

$$F(x) = \begin{cases} 0, & x < 0 \\ \sum_{j=\alpha}^{\alpha+\beta-1} \frac{(\alpha+\beta-1)!}{j!(\alpha+\beta-1-j)!} x^j (1-x)^{\alpha+\beta-1-j}, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$$

- Esp rance : $E[X] = \frac{\alpha}{\alpha+\beta}$
- Variance : $\text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
- Fonction g n ratrice des moments :

$$\mathcal{M}_X(t) = 1 + \sum_{k=1}^{\infty} \left(\prod_{j=0}^{k-1} \frac{\alpha+j}{\alpha+\beta+j} \right) \frac{t^k}{k!}$$

- Moments d'ordre k : $E[X^k] = \frac{\Gamma(\alpha+k)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+k)}$
- Esp rance tronqu e : $E[X \times 1_{\{X \leq d\}}] = \frac{\alpha}{\alpha+\beta} B(d; \alpha+1, \beta), \alpha \leq d \leq \beta$

$$\bullet \text{ Si } \beta = 1, E[X \times 1_{\{X \leq d\}}] = \frac{\alpha d^{\alpha+1}}{\alpha+1}$$

$$\bullet \text{ Si } \alpha = 1, E[X \times 1_{\{X \leq d\}}] = -d(1-d)^\beta + \frac{1-(1-d)^{\beta+1}}{\beta+1}$$

- Mesure VaR : outil d'optimisation

$$\bullet \text{ Si } \beta = 1, VaR_\kappa(X) = \kappa^{\frac{1}{\alpha}}$$

$$\bullet \text{ Si } \alpha = 1, VaR_\kappa(X) = 1 - (1-\kappa)^{\frac{1}{\beta}}$$

- Mesure $TVaR$: $TVaR_\kappa(X) = \frac{1}{(1-\kappa)} \frac{\alpha}{\alpha+\beta} (1 - B(VaR_\kappa(X); \alpha+1, \beta))$

$$\bullet \text{ Si } \beta = 1, TVaR_\kappa(X) = \frac{1}{(1-\kappa)} \frac{\alpha}{\alpha+1} (1 - \kappa^{(\alpha+1)/\alpha})$$

$$\bullet \text{ Si } \alpha = 1, TVaR_\kappa(X) = 1 - \frac{\beta}{\beta+1} (1-\kappa)^{\frac{1}{\beta}}$$

- Fonction *stop-loss* : $\pi_d(X) = \frac{\alpha}{\alpha+\beta} (1 - B(d; \alpha+1, \beta)) - d(1 - B(d; \alpha, \beta)), d \in [0, 1]$

$$\bullet \text{ Si } \beta = 1, \pi_d(X) = \frac{\alpha}{\alpha+1} (1 - d^{\alpha+1}) - d(1 - d^\alpha)$$

$$\bullet \text{ Si } \alpha = 1, \pi_d(X) = \frac{(1-d)^{\beta+1}}{1+\beta}$$

- Fonction d'excès-moyen : $e_d(X) = \frac{\alpha}{\alpha+\beta} \frac{1-B(d;\alpha+1,\beta)}{1-B(d;\alpha,\beta)} - d, d \in [0, 1]$
 - Si $\beta = 1, e_d(X) = \frac{\alpha}{\alpha+1} \frac{1-d^{\alpha+1}}{1-d^\alpha} - d$
 - Si $\alpha = 1, e_d(X) = \frac{(1-d)}{1+\beta}$
- Espérance limitée : $E[\min(X; d)] = \frac{\alpha}{\alpha+\beta} B(d; \alpha + 1, \beta) + \beta(1 - B(d; \alpha, \beta)), d \in [0, 1]$
 - Si $\beta = 1, E[\min(X; d)] = \frac{\alpha}{\alpha+1} d^{\alpha+1} + d(1 - d^\alpha)$
 - Si $\alpha = 1, E[\min(X; d)] = \frac{1-(1-d)^{\beta+1}}{\beta+1}$
- Loi associée : la loi uniforme avec $a = 0$ et $b = 1$ est un cas particulier de la loi bêta avec $\alpha = 1$ et $\beta = 1$.

B.4 Loi Coxienne-2

- Notation : $Cox2(p, \beta_1, \beta_2)$
- Paramètres : $\beta_1 > 0, \beta_2 > 0, p \in (0, 1), \beta_1 \neq \beta_2$
- Support : $x \in \mathbb{R}^+$
- Fonction de densité : $f_X(x) = p\beta_1 e^{-\beta_1 x} + (1-p) \left(\frac{\beta_2}{\beta_2 - \beta_1} \beta_1 e^{-\beta_1 x} + \frac{\beta_1}{\beta_1 - \beta_2} \beta_2 e^{-\beta_2 x} \right)$
- Fonction de répartition : $F_X(x) = p(1 - e^{-\beta_1 x}) + (1-p) \left(\frac{\beta_2}{\beta_2 - \beta_1} (1 - e^{-\beta_1 x}) + \frac{\beta_1}{\beta_1 - \beta_2} (1 - e^{-\beta_2 x}) \right)$
- Fonction de survie : $\bar{F}_X(x) = p e^{-\beta_1 x} + (1-p) \left(\frac{\beta_2}{\beta_2 - \beta_1} e^{-\beta_1 x} + \frac{\beta_1}{\beta_1 - \beta_2} e^{-\beta_2 x} \right)$
- Espérance : $E[X] = p \frac{1}{\beta_1} + (1-p) \left(\frac{1}{\beta_1} + \frac{1}{\beta_2} \right)$
- Variance : $Var(X) = 2 \left[p \frac{1}{\beta_1^2} + (1-p) \left(\frac{1}{\beta_1^2} + \frac{1}{\beta_2^2} + \frac{1}{\beta_1 \beta_2} \right) \right] - \left[p \frac{1}{\beta_1} + (1-p) \left(\frac{1}{\beta_1} + \frac{1}{\beta_2} \right) \right]^2$
- Fonction génératrice des moments : $\mathcal{M}_X(t) = p \frac{\beta_1}{\beta_1 - t} + (1-p) \left(\frac{\beta_1}{\beta_1 - t} \right) \left(\frac{\beta_2}{\beta_2 - t} \right)$
- Moments d'ordre k : $p \left(\frac{1}{\beta} \right)^k k! + (1-p) \prod_{i=1}^2 \frac{\Gamma(k+1)}{\beta^k}$
- Espérance tronquée :

$$E[X \times 1_{\{X \leq d\}}] = p \left(\frac{1}{\beta_1} (1 - e^{-\beta_1 d}) - d e^{-\beta_1 d} \right) + (1-p) \sum_{i=1}^2 \left(\prod_{j=1, j \neq i}^2 \frac{\beta_j}{\beta_j - \beta_i} \right) \left(-d e^{-\beta_i d} + \frac{1 - e^{-\beta_i d}}{\beta_i} \right)$$

- Mesure VaR : outil d'optimisation
- Mesure $TVaR$:

$$TVaR_\kappa(X) = \frac{p}{1-\kappa} e^{-\beta c} \left(\frac{1}{\beta} + c \right) + \frac{1-p}{1-\kappa} \sum_{i=1}^2 \left(\prod_{j=1, j \neq i}^2 \frac{\beta_j}{\beta_j - \beta_i} \right) \left(c e^{-\beta_i c} + \frac{e^{-\beta_i c}}{\beta_i} \right)$$

où $c = VaR_\kappa(X)$.

B.5 Loi Bêta généralisée de type 2

- Notation : $GB2(a, b, p, q)$
- Paramètres : $a > 0, b > 0, p > 0, q > 0$
- Support : $x \in \mathbb{R}^+$
- Fonction de densité : $f_X(x) = \frac{ax^{ap-1}}{b^ap I(p, q)(1+(\frac{x}{b})^a)^{p+q}}$, où $I(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$
- Fonction de répartition : $F_X(x) = B\left(\frac{(\frac{x}{b})^a}{1+(\frac{x}{b})^a}; p, q\right)$, où $B(x; \alpha, \beta)$ est la fonction de répartition d'une loi bêta
- Espérance : $E[X] = \frac{bI(p+\frac{1}{a}, q-\frac{1}{a})}{I(p, q)} = \frac{b\Gamma(p+\frac{1}{a})\Gamma(q-\frac{1}{a})}{\Gamma(p)\Gamma(q)}$
- Variance : $Var(X) = \frac{b^2}{I(p, q)} \left[I(p+\frac{2}{a}, q-\frac{2}{a}) - \frac{(I(p+\frac{1}{a}, q-\frac{1}{a}))^2}{I(p, q)} \right]$
- Moments d'ordre k : $E[X^k] = \frac{b^k I(p+\frac{k}{a}, q-\frac{k}{a})}{I(p, q)}$
- Espérance tronquée : $E[X \times 1_{\{X \leq d\}}] = \frac{bI(p+\frac{1}{a}, q-\frac{1}{a})}{I(p, q)} B\left(\frac{(\frac{d}{b})^a}{1+(\frac{d}{b})^a}; p+\frac{1}{a}, q-\frac{1}{a}\right)$
- Mesure Var : outil d'optimisation
- Mesure $TVaR$: $TVaR_\kappa(X) = \frac{1}{1-\kappa} \frac{bI(p+\frac{1}{a}, q-\frac{1}{a})}{I(p, q)} \bar{B}\left(\frac{(\frac{VaR_\kappa(X)}{b})^a}{1+(\frac{VaR_\kappa(X)}{b})^a}; p+\frac{1}{a}, q-\frac{1}{a}\right)$

B.6 Loi Pareto généralisée

- Notation : $PaG(u, \xi, \sigma)$
- Paramètres : $u \leq 0, \xi > 0, \sigma > 0$
- Support : $x \geq u$
- Fonction de densité : $f_X(x) = \frac{1}{\sigma} \left(1 + \frac{\xi}{\sigma}(x - u)\right)^{\left(-\frac{1}{\xi}-1\right)}$
- Fonction de répartition : $F_X(x) = 1 - \left(1 + \frac{\xi}{\sigma}(x - u)\right)^{-\frac{1}{\xi}}$
- Fonction de survie : $\bar{F}_X(x) = \left(1 + \frac{\xi}{\sigma}(x - u)\right)^{-\frac{1}{\xi}}$
- Espérance : $E[X] = \frac{\sigma}{1-\xi} + u, \xi < 1$
- Variance : $Var(X) = \frac{\sigma^2}{(1-\xi)^2(1-2\xi)}, \xi < \frac{1}{2}$
- Fonction génératrice des moments : $\mathcal{M}_X(t) = e^{tu} \sum_{i=0}^{\infty} \left(\frac{(t\sigma)^j}{\prod_{j=0}^i} (1 - j\xi) \right), \xi < \frac{1}{j}$
- Espérance tronquée : $E[X \times 1_{\{X \leq d\}}] = u - d(1 + \frac{\xi}{\sigma}(d - u))^{-\frac{1}{\xi}} - \frac{\sigma}{1-\xi} \left((1 + \frac{\xi}{\sigma}(d - u))^{1-\frac{1}{\xi}} - 1 \right)$
- Mesure VaR : $VaR_{\kappa}(X) = \frac{\sigma}{\xi} \left((1 - \kappa)^{-\xi} - 1 \right) + u$
- Mesure $TVaR$: $TVaR_{\kappa}(X) = \frac{\sigma}{\xi} \left(\frac{1}{1-\xi} (1 - \kappa)^{-\xi} - 1 \right) + u$

C Caractéristiques des raccordements de lois de probabilité

C.1 Loi Lognormale-Pareto généralisée

- Notation : $LNPaG(\mu, r, \xi, \sigma, w)$, avec $X_1 \sim LN(\mu, r)$ et $X_2 \sim PaG(u, \xi, \sigma)$
- Paramètres : $\mu > 0, r > 0, \xi > 0, \sigma > 0, w \in (0, 1)$
- Support : $x \in \mathbb{R}^+$
- Fonction de densité :

$$f_X(x) = \begin{cases} \frac{w}{xr\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{r}\right)^2} \frac{1}{\Phi\left(\frac{\ln(u)-\mu}{r}\right)} & , x \leq u \\ (1-w)\frac{1}{\sigma} \left(1 + \frac{\xi}{\sigma}(x-u)\right)^{-\left(\frac{1}{\xi}+1\right)} & , x > u \end{cases}$$

- Fonction de répartition :

$$F_X(x) = \begin{cases} w \frac{\Phi\left(\frac{\ln(x)-\mu}{r}\right)}{\Phi\left(\frac{\ln(u)-\mu}{r}\right)} & , x \leq u \\ w + (1-w) \left(1 - \left(1 + \frac{\xi}{\sigma}(x-u)\right)^{-\frac{1}{\xi}}\right) & , x > u \end{cases}$$

- Fonction de survie :

$$\bar{F}_X(x) = \begin{cases} 1 - w \frac{\Phi\left(\frac{\ln(x)-\mu}{r}\right)}{\Phi\left(\frac{\ln(u)-\mu}{r}\right)} & , x \leq u \\ (1-w) \left(1 + \frac{\xi}{\sigma}(x-u)\right)^{-\frac{1}{\xi}} & , x > u \end{cases}$$

- Espérance :

$$E[X] = we^{\mu + \frac{r^2}{2}} \frac{\Phi\left(\frac{\ln(u)-\mu-r^2}{r}\right)}{\Phi\left(\frac{\ln(u)-\mu}{r}\right)} + (1-w) \left(\frac{\sigma}{1-\xi} + u\right)$$

- Variance : L'expression de la variance est complexe, mais elle n'existe pas pour $\xi \geq \frac{1}{2}$.
- Espérance tronquée :

$$E[X \times 1_{\{X \leq d\}}] = \begin{cases} \frac{w}{F_{X_1}(u)} E[X_1 \times 1_{\{X_1 \leq u\}}] & , d \leq u \\ \frac{w}{F_{X_1}(x)} E[X_1 \times 1_{\{X_1 \leq u\}}] + (1-w) E[X_2 \times 1_{\{X_2 \leq d\}}] & , d > u \end{cases}$$

- Mesure VaR :

$$VaR_\kappa(X) = \begin{cases} \exp\{\mu + \sigma \Phi^{-1}\left(\frac{\kappa F_{X_1}(u)}{w}\right)\} & , \kappa \leq w \\ \frac{\sigma}{\xi} \left(\left(\frac{1-\kappa}{1-w}\right)^{-\xi} - 1\right) + u & , \kappa > w \end{cases}$$

- Mesure $TVaR$:

$$TVaR_\kappa(X) = \begin{cases} \frac{w}{F_{X_1}(u)(1-\kappa)} (E[X_1 \times 1_{\{X_1 \leq u\}}] - E[X_1 \times 1_{\{X_1 \leq c\}}]) + \frac{1-w}{(1-\kappa)} E[X_2 \times 1_{\{X_2 > u\}}] & , c \leq u \\ \frac{1-w}{(1-\kappa)} E[X_2 \times 1_{\{X_2 > c\}}] = (1-w) TVaR_\kappa(X_2) & , c > u \end{cases}$$

C.2 Loi Weibull-Pareto généralisée

- Notation : $WePaG(\tau, \beta, \xi, \sigma, w)$, avec $X_1 \sim We(\tau, \beta)$ et $X_2 \sim PaG(u, \xi, \sigma)$
- Paramètres : $\tau > 0, \beta > 0, \xi > 0, \sigma > 0, w \in (0, 1)$
- Support : $x \in \mathbb{R}^+$
- Fonction de densité :

$$f_X(x) = \begin{cases} \frac{\tau}{\beta} \left(\frac{x}{\beta}\right)^{\tau-1} \frac{e^{-\left(\frac{x}{\beta}\right)^\tau}}{1 - e^{-\left(\frac{u}{\beta}\right)^\tau}} & , x \leq u \\ (1-w) \frac{1}{\sigma} \left(1 + \frac{\xi}{\sigma}(x-u)\right)^{-\left(\frac{1}{\xi}+1\right)} & , x > u \end{cases}$$

- Fonction de répartition :

$$F_X(x) = \begin{cases} w \frac{1 - e^{-\left(\frac{x}{\beta}\right)^\tau}}{1 - e^{-\left(\frac{u}{\beta}\right)^\tau}} & , x \leq u \\ w + (1-w) \left(1 - \left(1 + \frac{\xi}{\sigma}(x-u)\right)^{-\frac{1}{\xi}}\right) & , x > u \end{cases}$$

- Fonction de survie :

$$\bar{F}_X(x) = \begin{cases} 1 - w \frac{1 - e^{-\left(\frac{x}{\beta}\right)^\tau}}{1 - e^{-\left(\frac{u}{\beta}\right)^\tau}} & , x \leq u \\ (1-w) \left(1 + \frac{\xi}{\sigma}(x-u)\right)^{-\frac{1}{\xi}} & , x > u \end{cases}$$

- Espérance :

$$E[X] = \frac{w\beta}{1 - e^{-\left(\frac{u}{\beta}\right)^\tau}} \Gamma\left(1 + \frac{1}{\tau}\right) H\left(u^\tau, 1 + \frac{1}{\tau}, \frac{1}{\beta^\tau}\right) + (1-w) \left(\frac{\sigma}{1-\xi} + u\right)$$

- Variance : L'expression de la variance est complexe, mais elle n'existe pas pour $\xi > \frac{1}{2}$.
- Espérance tronquée :

$$E[X \times 1_{\{X \leq d\}}] = \begin{cases} \frac{w}{F_{X_1}(u)} E[X_1 \times 1_{\{X_1 \leq u\}}] & , d \leq u \\ \frac{w}{F_{X_1}(x)} E[X_1 \times 1_{\{X_1 \leq u\}}] + (1-w) E[X_2 \times 1_{\{X_2 \leq d\}}] & , d > u \end{cases}$$

- Mesure Var :

$$Var_\kappa(X) = \begin{cases} \sigma \left(-\ln\left(1 - \frac{kF_x(u)}{w}\right)\right)^{\frac{1}{\tau}} & , \kappa \leq w \\ \frac{\sigma}{\xi} \left(\left(\frac{1-\kappa}{1-w}\right)^{-\xi} - 1\right) + u & , \kappa > w \end{cases}$$

- Mesure $TVaR$:

$$TVaR_\kappa(X) = \begin{cases} \frac{w}{F_{X_1}(u)(1-\kappa)} (E[X_1 \times 1_{\{X_1 \leq u\}}] - E[X_1 \times 1_{\{X_1 \leq c\}}]) + \frac{1-w}{(1-\kappa)} E[X_2 \times 1_{\{X_{Pg} > u\}}] & , c \leq u \\ \frac{1-w}{(1-\kappa)} E[X_2 \times 1_{\{X_2 > c\}}] = (1-w) TVaR_\kappa(X_2) & , c > u \end{cases}$$

C.3 Loi GB2-Pareto généralisée

- Notation : $GB2PaG(a, b, p, q, \xi, \sigma, w)$, avec $X_1 \sim GB2(a, b, p, q)$ et $X_2 \sim PaG(u, \xi, \sigma)$
- Paramètres : $a > 0, b > 0, p > 0, q > 0, \xi > 0, \sigma > 0, w \in (0, 1)$
- Support : $x \in \mathbb{R}^+$
- Fonction de densité :

$$f_X(x) = \begin{cases} w \frac{ax^{ap-1}}{b^{ap}I(p, q)(1+(\frac{x}{b})^a)^{p+q}} \frac{1}{B\left(\frac{(u/b)^a}{1+(u/b)^a}; p, q\right)} & , x \leq u \\ (1-w) \frac{1}{\sigma} \left(1 + \frac{\xi}{\sigma}(x-u)\right)^{\left(-\frac{1}{\xi}-1\right)} & , x > u \end{cases}$$

- Fonction de répartition :

$$F_X(x) = \begin{cases} wB\left(\frac{(x/b)^a}{1+(x/b)^a}; p, q\right) / B\left(\frac{(u/b)^a}{1+(u/b)^a}; p, q\right) & , x \leq u \\ w + (1-w) \left(1 - \left(1 + \frac{\xi}{\sigma}(x-u)\right)^{-\frac{1}{\xi}}\right) & , x > u \end{cases}$$

- Fonction de survie :

$$\bar{F}_X(x) = \begin{cases} 1 - wB\left(\frac{(x/b)^a}{1+(x/b)^a}; p, q\right) / B\left(\frac{(u/b)^a}{1+(u/b)^a}; p, q\right) & , x \leq u \\ (1-w) \left(1 - \left(1 + \frac{\xi}{\sigma}(x-u)\right)^{-\frac{1}{\xi}}\right) & , x > u \end{cases}$$

- Espérance :

$$E[X] = \frac{w}{B\left(\frac{(u/b)^a}{1+(u/b)^a}; p, q\right)} \frac{bI\left(p + \frac{1}{a}, q - \frac{1}{a}\right)}{I(p, q)} B\left(\frac{(u/b)^a}{1+(u/b)^a}; p + \frac{1}{a}, q - \frac{1}{a}\right) + (1-w) \left(\frac{\sigma}{1-\xi} + u\right)$$

- Espérance tronquée :

$$E[X \times 1_{\{X \leq d\}}] = \begin{cases} \frac{w}{F_{X_1}(u)} E[X_1 \times 1_{\{X_1 \leq u\}}] & , d \leq u \\ \frac{w}{F_{X_1}(u)} E[X_1 \times 1_{\{X_1 \leq u\}}] + (1-w) E[X_2 \times 1_{\{X_2 \leq d\}}] & , d > u \end{cases}$$

- Mesure VaR : outil d'optimisation
- Mesure $TVaR$:

$$TVaR_\kappa(X) = \begin{cases} \frac{w}{F_{X_1}(u)(1-\kappa)} (E[X_1 \times 1_{\{X_1 \leq u\}}] - E[X_1 \times 1_{\{X_1 \leq c\}}]) + \frac{1-w}{(1-\kappa)} E[X_2 \times 1_{\{X_{Pg} > u\}}] & , c \leq u \\ \frac{1-w}{(1-\kappa)} E[X_2 \times 1_{\{X_2 > c\}}] = (1-w) TVaR_\kappa(X_2) & , c > u \end{cases}$$

D Liste complète des variables explicatives de la base de données de sinistres incendies de la ville de Toronto

Pour ce rapport, on s'est concentré sur la modélisation de la variable *Estimated Dollar Loss*. On aussi utilisé la variable *Building Status* pour faire une analyse sur les bâtiments en construction et ceux en construction et rénovation. On a aussi utilisé la variable *TFS Alarm Time* pour faire l'analyse de la fréquence. On peut constater qu'on a énormément d'informations sur les sinistres incendies avec cette base de données.

- | | |
|---|--|
| 1. <i>Area of Origin</i> | 23. <i>Latitude</i> |
| 2. <i>Building Status</i> | 24. <i>Level Of Origin</i> |
| 3. <i>Business Impact</i> | 25. <i>Longitude</i> |
| 4. <i>Civilian Casualties</i> | 26. <i>Material First Ignited</i> |
| 5. <i>Count of Persons Rescued</i> | 27. <i>Method Of Fire Control</i> |
| 6. <i>Estimated Dollar Loss</i> | 28. <i>Number of responding apparatus</i> |
| 7. <i>Estimated Number Of Persons Displaced</i> | 29. <i>Number of responding personnel</i> |
| 8. <i>Exposures</i> | 30. <i>Possible Cause</i> |
| 9. <i>Ext agent app or defer time</i> | 31. <i>Property Use</i> |
| 10. <i>Extent Of Fire</i> | 32. <i>Smoke Alarm Impact on Persons Evacuating Impact on Evacuation</i> |
| 11. <i>Final Incident Type</i> | 33. <i>Smoke Alarm at Fire Origin</i> |
| 12. <i>Fire Alarm System Impact on Evacuation</i> | 34. <i>Smoke Alarm at Fire Origin Alarm Failure</i> |
| 13. <i>Fire Alarm System Operation</i> | 35. <i>Smoke Alarm at Fire Origin Alarm Type</i> |
| 14. <i>Fire Alarm System Presence</i> | 36. <i>Smoke Spread</i> |
| 15. <i>Fire Under Control Time</i> | 37. <i>Sprinkler System Operation</i> |
| 16. <i>Ignition Source</i> | 38. <i>Sprinkler System Presence</i> |
| 17. <i>Incident Number</i> | 39. <i>Status of Fire On Arrival</i> |
| 18. <i>Incident Station Area</i> | 40. <i>TFS Alarm Time</i> |
| 19. <i>Incident Ward</i> | 41. <i>TFS Arrival Time</i> |
| 20. <i>Initial CAD Event Type</i> | 42. <i>TFS Firefighter Casualties</i> |
| 21. <i>Intersection</i> | 43. <i>X id</i> |
| 22. <i>Last TFS Unit Clear Time</i> | |

E Description du projet pour la demande au CIMMUL

Titre : Analyse actuarielle du risque incendie

Description : Pour les actuaires, la modélisation mathématique des coûts suite à un incendie est un défi important. Dans ce projet, on examine une base de données de sinistre incendie de source canadienne. On fait une revue de certains modèles pour décrire adéquatement le risque incendie. On teste aussi certains modèles sur les données canadiennes de sinistre incendie.

Programme : Bourse au premier cycle

Références

- [Abramson, 2022] Abramson, D. (2022). A nonproportional premium rating method for construction risks. *North American Actuarial Journal*, pages 1–20.
- [Albrecher et al., 2017] Albrecher, H., Beirlant, J., and Teugels, J. L. (2017). *Reinsurance : actuarial and statistical aspects*. John Wiley & Sons.
- [Almer, 1963] Almer, B. (1963). Individual risk theory and risk statistics as applied to fire insurance. *ASTIN Bulletin : The Journal of the IAA*, 2(3) :365–379.
- [Antal and Re, 2007] Antal, D. P. and Re, S. (2007). Mathematical methods in reinsurance. *Lecture Notes, ETH Zurich*.
- [Benckert and Sternberg, 1957] Benckert, L.-G. and Sternberg, I. (1957). An attempt to find an expression for the distribution of fire damage amount. In *XVth International Congress of Actuaries, New York*.
- [Bernegger, 1997] Bernegger, S. (1997). The swiss re exposure curves and the mbbfd distribution class1. *ASTIN Bulletin : The Journal of the IAA*, 27(1) :99–111.
- [Biffis and Chavez, 2014] Biffis, E. and Chavez, E. (2014). Tail risk in commercial property insurance. *Risks*, 2(4) :393–410.
- [Blier-Wong, 2018] Blier-Wong, C. (2018). Correction des primes d’assurance habitation en présence de dépendance spatiale.
- [Brazauskas and Kleefeld, 2016] Brazauskas, V. and Kleefeld, A. (2016). Modeling severity and measuring tail risk of norwegian fire claims. *North American Actuarial Journal*, 20(1) :1–16.
- [Caeiro and Gomes, 2015] Caeiro, F. and Gomes, M. I. (2015). Threshold selection in extreme value analysis. *Extreme value modeling and risk analysis : Methods and applications*, pages 69–82.
- [Canty and Ripley, 2021] Canty, A. and Ripley, B. D. (2021). *boot : Bootstrap R (S-Plus) Functions*. R package version 1.3-28.
- [Charpentier and Flachaire, 2021] Charpentier, A. and Flachaire, E. (2021). Pareto models for risk management. In *Recent Econometric Techniques for Macroeconomic and Financial Data*, pages 355–387. Springer.
- [Clauset et al., 2009] Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, 51(4) :661–703.
- [Clauset et al., 2007] Clauset, A., Young, M., and Gleditsch, K. S. (2007). On the frequency of severe terrorist events. *Journal of Conflict Resolution*, 51(1) :58–87.
- [Corradin et al., 2002] Corradin, S., Spa, R., and di Gruppo, P. R. (2002). Economic risk capital and reinsurance : an extreme value theory’s application to fire claims of an insurance company. *University Catholique de Louvain : Institute of statistics*.
- [Corradin and Verbrigghe, 2001] Corradin, S. and Verbrigghe, B. (2001). Economic risk capital and reinsurance : an application to fire claims of an insurance company. *RAS, Pianificazione Reddittività di Gruppo, December*.
- [Cummins et al., 1990] Cummins, J. D., Dionne, G., McDonald, J. B., and Pritchett, B. M. (1990). Applications of the gb2 family of distributions in modeling insurance loss processes. *Insurance : Mathematics and Economics*, 9(4) :257–272.

- [Denuit et al., 2007] Denuit, M., Maréchal, X., Pitrebois, S., and Walhin, J.-F. (2007). *Actuarial modelling of claim counts : Risk classification, credibility and bonus-malus systems*. John Wiley & Sons.
- [Embrechts and Schmidli, 1994] Embrechts, P. and Schmidli, H. (1994). Modelling of extremal events in insurance and finance. *Zeitschrift für Operations Research*, 39(1) :1–34.
- [Graf and Nedyalkova., 2022] Graf, M. and Nedyalkova., D. (2022). *GB2 : Generalized Beta Distribution of the Second Kind : Properties, Likelihood, Estimation*. R package version 2.1.1.
- [Klugman et al., 2013] Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2013). *Loss models : Further topics*. John Wiley & Sons.
- [Lin, 2005] Lin, Y.-S. (2005). Estimations of the probability of fire occurrences in buildings. *Fire safety journal*, 40(8) :728–735.
- [Longley-Cook, 1951] Longley-Cook, L. H. (1951). Problems of fire insurance rate making. In *Proceedings of the Casualty Actuarial Society*, volume 38, pages 94–102.
- [Longley-Cook, 1952] Longley-Cook, L. H. (1952). A statistical study of large fire losses with application to a problem in catastrophe insurance. In *Proceedings of the Casualty Actuarial Society*, volume 39, pages 77–83.
- [Longley-Cook, 1959] Longley-Cook, L. H. (1959). Notes on some actuarial problems of property insurance. In *Proceedings of the Casualty Actuarial Society*, volume 46, pages 66–81.
- [Mandelbrot, 1964] Mandelbrot, B. (1964). Random walks, fire damage amount and other paretian risk phenomena. *Operations Research*, 12(4) :582–585.
- [McNeil, 1997a] McNeil, A. J. (1997a). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin : The Journal of the IAA*, 27(1) :117–137.
- [McNeil, 1997b] McNeil, A. J. (1997b). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin*, 27(1) :117–137.
- [Michaelides et al., 1997] Michaelides, N., Brown, P., Chacko, F., Graham, M., Haynes, J., Hindley, D., Howard, S., Johnson, H., Morgan, K., Pettengell, C., et al. (1997). The premium rating of commercial risks. In *General Insurance Convention*, pages 397–491.
- [Ohlsson and Johansson, 2010] Ohlsson, E. and Johansson, B. (2010). *Non-life insurance pricing with generalized linear models*, volume 2. Springer.
- [Ossberger, 2020] Ossberger, J. (2020). *tea : Threshold Estimation Approaches*. R package version 1.1.
- [Parodi, 2014] Parodi, P. (2014). *Pricing in general insurance*. CRC press.
- [Ramachandran, 1974] Ramachandran, G. (1974). Extreme value theory and large fire losses. *ASTIN Bulletin : The Journal of the IAA*, 7(3) :293–310.
- [Reynkens and Verbelen, 2020] Reynkens, T. and Verbelen, R. (2020). *ReIns : Functions from "Reinsurance : Actuarial and Statistical Aspects"*. R package version 1.0.10.
- [Riegel, 2010] Riegel, U. (2010). On fire exposure rating and the impact of the risk profile type. *ASTIN Bulletin : The Journal of the IAA*, 40(2) :727–777.
- [Rydman, 2018] Rydman, M. (2018). Application of the peaks-over-threshold method on insurance data.
- [Shpilberg, 1977] Shpilberg, D. C. (1977). The probability distribution of fire loss amount. *Journal of Risk and Insurance*, pages 103–115.
- [Thompson et al., 2009] Thompson, P., Cai, Y., Reeve, D., and Stander, J. (2009). Automated threshold selection methods for extreme wave analysis. *Coastal Engineering*, 56(10) :1013–1021.