

Perspectives actuarielles pour le péril incendie, en particulier pour les constructions en bois massif

Mémoire

Jérémie Barde

Sous la direction de:

Etienne Marceau, directeur de recherche
Hélène Cossette, codirectrice de recherche

Résumé

Cet essai vise à comprendre et à quantifier le risque d'incendie, en se concentrant particulièrement sur l'assurance construction pour les chantiers en bois massif. Pour ce faire, on utilise diverses bases de données de sinistres incendie, notamment celles de la ville de Toronto et la *National Fire Information Database*. À l'aide de ces bases de données, on crée plusieurs modèles probabilistes en utilisant différentes familles paramétriques. On examine l'impact des types de construction sur les pertes totales en sinistre incendie en calculant la part allouée à chaque type en se basant sur la théorie du partage des risques (*risk sharing*).

Table des matières

Résumé	ii
Table des matières	iii
Liste des tableaux	v
Liste des illustrations	viii
Remerciements	x
1 Introduction	1
1.1 Introduction	1
1.2 Revue de la littérature	2
1.3 Transfert de risque incendie via l'assurance IARD	5
1.4 Tarification en assurance IARD	7
1.5 Loi Pareto	10
1.6 Théorie des valeurs extrêmes	11
1.7 Mesures de risque de distorsion	14
2 Description de deux bases de données sinistres incendie	17
2.1 Introduction	17
2.2 Base de données de périls incendie de la ville de Toronto	17
2.3 Base de données de la <i>National Fire Information Data Base</i>	23
3 Modélisation des pertes pour le péril incendie (sévérité)	32
3.1 Introduction	32
3.2 Modélisation préliminaire des montants de perte avec une loi lognormale . .	33
3.3 Méthode <i>Peak-over-Threshold</i> (POT)	33
3.4 Nouvelles lois construites par raccordement de lois	39
3.5 Test des lois	46
3.6 Intervalles de confiance	50
3.7 Informations sur les lois sélectionnées pour la base de données DataToronto	50
3.8 Résumé de l'analyse avec la base de données DataNFID	51
3.9 Modélisation par année des pertes en sinistres incendie avec la base de don- nées DataToronto	54
3.10 Modélisation par année des pertes en sinistres incendie avec la base de don- nées DataNFID	58

4 Partage du risque lié aux pertes en sinistre incendie	63
4.1 Introduction	63
4.2 Modélisation par classe des pertes en sinistre incendie	65
4.3 Règle de partage de risques	66
4.4 Résultats	68
Conclusion	71
A Modélisation du processus d'avènements des incendies	72
B Autres bases de données	74
B.1 Données de San Francisco des pertes en sinistre incendie	74
B.2 Données suédoises des pertes en sinistre incendie	75
B.3 Données norvégiennes des pertes en sinistre incendie	75
C Les sinistres les plus élevés selon le type de construction	77
D Liste des variables explicatives	80
D.1 Données de sinistres incendie de la ville de Toronto	80
D.2 <i>National Fire Information Data Base (NFID)</i>	81
E Caractéristiques des lois	83
E.1 Caractéristiques des lois de probabilité	83
E.2 Caractéristiques des raccordements de lois de probabilité	91
F Développement des formes analytiques pour les mesures de risque	94
F.1 Loi Pareto généralisée	94
Bibliographie	97

Liste des tableaux

1.1	Principaux secteurs d'activité en assurance IARD au Canada, avec leurs pourcentages de primes nettes écrites	5
1.2	Comparaison de la probabilité d'avoir un sinistre supérieur à 1 million	10
1.3	Forme analytique des mesures de risque <i>VaR</i> , <i>TVaR</i> , <i>PHT</i> , <i>WT</i> et <i>GS</i> pour les lois exponentielle, lognormales tronquée et Pareto généralisée	16
2.1	Description de 7 variables explicatives de la base de données DataToronto	18
2.2	Statistiques descriptives des $n = 13\,571$ pertes estimées retenues de la base de données DataToronto	18
2.3	Nombre de données dans des tranches de montants de sinistres	20
2.4	Catégories des variables explicatives et exemple pour certaines	23
2.5	Statistiques descriptives des $n = 102\,635$ pertes retenues de la base de données DataNFID	24
2.6	Nombre de données par tranche de montants de sinistres	25
2.7	Pertes moyennes pour les sinistres supérieures à 1 million selon la présence d'un système d'arroseur	26
2.8	Valeur moyenne du bâtiment, pour les pertes supérieures à 1 million selon la présence d'un système d'arroseur	26
2.9	Code pour la variable GENCONST	27
2.10	Code pour la variable PROPGRP	27
2.11	Les 10 sinistres avec les montants de pertes les plus élevés en millions et leurs caractéristiques	29
2.12	Les 10 sinistres avec les montants de pertes les plus élevés, en millions, pour les constructions en bois et leurs caractéristiques	30
2.13	Statistiques descriptives des $n = 182$ pertes sur des bâtiments en construction .	30
2.14	Nombre de sinistres sur des bâtiments en construction selon le type de construction	30
3.1	Paramètres estimés pour la loi lognormale	33
3.2	Seuil trouvé avec la méthode PLFIT selon différents vecteurs de u	36
3.3	Paramètres de la Pareto généralisée	36
3.4	Fonction de répartition empirique au seuil, nombre d'observations supérieures au seuil, pourcentage des montants de sinistre supérieurs au seuil au montant total de sinistres	37
3.5	Statistique descriptive pour les montants de sinistres pour les bâtiments en construction et pour les bâtiments en construction et rénovations	37
3.6	Paramètres de la Pareto généralisée pour les bâtiments en construction et en construction et rénovation	38
3.7	Valeurs pour les mesures <i>VaR</i> et <i>TVaR</i>	38

3.8	Paramètres estimés pour la loi composite lognormale-Pareto généralisée	41
3.9	Paramètres pour la loi composite lognormale-Pareto généralisée	42
3.10	Paramètres pour la loi composite Coxienne-2-Pareto généralisée	44
3.11	Paramètres pour la loi composite bêta généralisée de type 2-Pareto généralisée	45
3.12	Résultats des tests W^2 , A^2 , AIC et BIC	49
3.13	Intervalle de confiance estimé avec 14 000 <i>bootstrap</i>	50
3.14	Espérance et variance	50
3.15	Valeurs de $VaR_\kappa(X)$ et $TVaR_\kappa(X)$	51
3.16	Résultats des tests W^2 , A^2 , AIC et BIC	52
3.17	Fonction de répartition empirique évaluée au seuil, nombre d'observations supérieures au seuil, pourcentage des montants de sinistre supérieurs au seuil au montant total de sinistres	52
3.18	Paramètres estimés pour les lois LN-PaG et GB2-PaG	52
3.19	Espérance et variance	53
3.20	Valeurs de $VaR_\kappa(X)$ et $TVaR_\kappa(X)$	53
3.21	Paramètres estimés par maximum de vraisemblance des raccordements de loi pour les pertes incendies de la base de données DataToronto	54
3.22	Test d'ajustement de modèles pour les données canadiennes de périls incendie .	56
3.23	Mesures d'information des modèles ajustés pour la base de données DataToronto (2011-2018)	57
3.24	Mesures VaR et $TVaR$ (2ièmes lignes) au niveau de confiance 95%, pour la base de données DataToronto (2012-2018)	57
3.25	Valeurs prédites et actuelles de $\Pr(B > 100\,000)$, pour la base de données DataToronto (2012-2018)	58
3.26	Paramètres estimés par maximum de vraisemblance des raccordements de loi pour les pertes incendies de la base de données DataNFID	59
3.27	Tests d'ajustements des modèles pour la base de données DataNFID	61
3.28	Critères d'informations des modèles ajustés pour la base de données DataNFID (2010-2015)	61
3.29	Mesure VaR et TCM (2ieme lignes) au niveau de confiance 90%, pour la base de données DataNFID (2009-2015)	62
3.30	Valeurs prédites et actuelles de $\Pr(B > 100\,000)$, pour la base de données DataNFID (2009-2015)	62
4.1	Classes basés sur le type de construction	63
4.2	Statistiques descriptives pour chacun des classes	64
4.3	Paramètres estimés pour chacun des classes	65
4.4	Nombre $n_{\theta,j}$ de sinistres de la classe j avec pertes supérieures au seuil θ . . .	65
4.5	Paramètres estimés du processus de Poisson homogène pour chaque classe . .	66
4.6	Contribution de chaque classe pour une valeur de s (paramètre α)	69
4.7	Probabilité que la v.a. S soit plus grande 5, 10, 30, 50 et 117.54	69
4.8	Contribution à la mesure $TVaR$ pour chacune des classes	70
B.1	Statistiques descriptives de la base de données de San Francisco de sinistres incendies	74
B.2	Paramètres de la Pareto généralisée pour les montants de réclamations de la ville de San Francisco	74

B.3	Statistiques descriptives de la base de données suédoise de $n = 215$ sinistres incendies. Les montants de sinistres sont en unité de millions SKR.	75
B.4	Paramètres de la Pareto généralisée pour les montants de réclamations suédois	75
B.5	Statistiques descriptives de la base de données norvégienne de $n = 9181$ sinistres incendies	75
B.6	Paramètres de la Pareto généralisée pour les montants de réclamations norvégien	75
C.1	Construction combustible	77
C.2	Construction combustible protégé	78
C.3	Construction en bois massif	78
C.4	Construction non combustible	78
C.5	Construction non combustible protégé	79

Liste des illustrations

1.1	Conventions de réassurance pour deux chantiers impliquant deux compagnies d'assurance et trois compagnies de réassurance	6
1.2	Processus de tarification	8
1.3	Processus d'analyse des données	8
1.4	Comparaison de la densité de deux lois Pareto selon $\alpha \in (1.5, 3)$	11
1.5	Exemple de la méthode <i>Peak-Over-Threshold</i> (POT)	14
2.1	Contribution des plus grosses pertes estimées au total des pertes estimées	18
2.2	Pertes estimées selon la date du sinistre	19
2.3	Histogramme des pertes et du log des pertes	20
2.4	Fonction de répartition empirique des montants de sinistre	21
2.5	<i>QQplots</i> exponentielle, lognormale et Pareto	22
2.6	Pertes estimées selon la date du sinistre	25
2.7	Proportion des plus grosses pertes estimées du total des pertes estimées	26
2.8	Pertes totales, pertes moyennes, valeurs moyennes et nombre d'incendies selon le type de construction	28
2.9	Pertes totales, pertes moyennes, valeurs moyennes et nombre d'incendies selon le secteur d'utilité du bâtiment	29
2.10	Pertes et valeurs moyennes pour les bâtiments en construction selon le type de construction	31
3.1	Fonction de répartition empirique et fonction de répartition de la loi lognormale	33
3.2	<i>QQplot</i> Pareto et fonction d'excès moyen	34
3.3	Comparaison fonction de répartition de la loi Pareto généralisée, de la fonction de répartition de la loi lognormale avec la fonction de répartition empirique	37
3.4	Comparaison de la fonction de répartition de la loi LN-PaG avec la fonction de répartition empirique	42
3.5	Comparaison fonction de répartition de la loi We-PaG avec la fonction de répartition empirique	43
3.6	Comparaison fonction de répartition de la loi We-PaG avec la fonction de répartition empirique	44
3.7	Comparaison fonction de répartition de la loi We-PaG avec la fonction de répartition empirique	45
3.8	<i>QQplot</i> lognormale-Pareto généralisée	46
3.9	<i>QQplot</i> pour la loi Weibull-Pareto généralisée	47
3.10	<i>QQplot</i> pour la loi Weibull-Pareto généralisée avec continuité	47
3.11	<i>QQplot</i> pour la loi Coxienne-2-Pareto généralisée	48
3.12	<i>QQplot</i> pour la loi Coxienne-2-Pareto généralisée avec continuité	48

3.13	<i>QQplot</i> pour la loi GB2-Pareto généralisée avec continuité	49
3.14	<i>QQplots</i> pour les observations sous le seuil (ligne 1) et pour les observations au-dessus du seuil (ligne 2), pour le modèle LN-PaG (2011-2018)	55
3.15	<i>QQplots</i> pour les observations sous le seuil (ligne 1) et pour les observations au-dessus du seuil (ligne 2), pour le modèle We-PaG (2011-2018)	55
3.16	<i>QQplots</i> pour les observations sous le seuil (ligne 1) et pour les observations au-dessus du seuil (ligne 2), pour le modèle GB2-PaG (2011-2018)	56
3.17	<i>QQplots</i> pour les observations sous le seuil (ligne 1) et pour les observations au-dessus du seuil (ligne 2), pour le modèle LN-PaG (2011-2018)	60
3.18	<i>QQplots</i> pour les observations sous le seuil (ligne 1) et pour les observations au-dessus du seuil (ligne 2), pour le modèle GB2-PaG (2011-2018)	60
4.1	Pertes selon le temps pour chaque classe	64
4.2	Fonction de répartition empirique des pertes comparativement à la fonction de répartition des pertes du modèles	65
4.3	Évolution de la contribution selon la valeur s pour la règle de partage de risque <i>conditional mean</i>	70
A.1	Test d'adéquation graphique pour la fréquence d'arrivée des sinistres incendies .	72
A.2	Pour les sinistres supérieurs à 1 million	73

Remerciements

Premièrement, je tiens à remercier mes directeurs de maîtrise Etienne Marceau et Hélène Cossette pour m'avoir encouragé et donné la possibilité de poursuivre mes études au deuxième cycle. Je suis extrêmement reconnaissant de la chance que vous m'avez donnée. Je vous remercie également pour le temps et le support que vous m'avez fourni pour que je puisse m'améliorer et atteindre mon plein potentiel. Je remercie également le professeur Ilie Radu Mitric d'avoir accepté de corriger l'essai.

Je suis profondément reconnaissant pour le soutien financier qui m'a permis de me consacrer entièrement à mes études. J'ai énormément de gratitude pour la Chaire Industrielle de Recherche sur la Construction Écoresponsable du Bois (CIRCERB) qui m'a ouvert ses portes pour joindre une équipe très dynamique. Je la remercie grandement pour avoir financé ma maîtrise en grande majorité. Je remercie également l'École d'actuariat et la Chaire d'actuariat pour le financement et les locaux qui ont été mis à ma disposition.

Enfin, je veux remercier mes amis et ma famille pour m'avoir supporté pendant toute la durée de mes études. Je me dois également de remercier mes collèges du laboratoire, André Orelion Chuisseu Tchuisseu, Benjamin Côté, Alexandre Dubeau et Philippe Leblanc, dont leur présence a été d'une aide précieuse. Ils ont grandement aidé à améliorer la qualité de cet essai.

Chapitre 1

Introduction

1.1 Introduction

L’assurance de dommage remonte à plusieurs 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, appelée *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 leurs portes. Il est évident que l’assurance incendie a grandement contribué au développement de l’assurance IARD. Encore aujourd’hui, l’assurance incendie reste un défi pour les actuaires. Plusieurs personnes ont été amenées à travailler sur ce risque et à proposer différentes techniques pour mieux le comprendre. Dans Ramachandran (1974), l’auteur propose d’utiliser la théorie des valeurs extrêmes pour modéliser les montants de pertes élevées dues aux incendies. Dans Brazauskas and Kleefeld (2016), les auteurs utilisent la technique de raccordement de loi pour essayer de mieux modéliser la queue de la distribution et testent cette technique sur des données de sinistres incendies. Bien qu’il soit important de modéliser la sévérité, il est aussi important de modéliser la distribution du nombre de sinistres. Les auteurs de Corradin and Verbrigghe (2001) et Corradin et al. (2002) proposent d’utiliser la loi binomiale négative pour la fréquence. Lin (2005) quant à lui propose plutôt d’utiliser le processus de Poisson pour modéliser la fréquence.

1.2 Revue de la littérature

Dans Longley-Cook (1952), l'auteur examine les pertes élevées résultant des incendies, étant l'un des premiers à aborder ce sujet. 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 de supposer 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'est 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, aux États-Unis, il y en avait 38. Il propose aussi de standardiser la tarification en assurance de propriété.

Dans Benckert and Sternberg (1957) les auteurs introduisent la loi Pareto. Ensuite, Mandelbrot (1964) pousse les travaux plus loin. Dans Charpentier and Flachaire (2021) les auteurs montrent l'utilisation de la loi Pareto dans les domaines de l'assurance, de la réassurance et 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 du risque 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 individuel.

Une excellente introduction à la théorie du risque est présentée dans Antal and Re (2007). On y présente les outils mathématiques pour la tarification en assurance de dommages et la modélisation. De plus, l'auteur propose explication du capital et une introduction à la tarification des contrats de réassurance.

Le professeur G. 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 pertes élevées dues aux incendies, mais qu'il faudrait l'utiliser avec prudence.

Plusieurs chercheurs ont tenté de modéliser les pertes provenant de sinistre incendie. 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. En outre, il utilise le fait que la fonction d'intensité du feu devrait être décroissante pour montrer que la loi lognormale est un meilleur choix puisque celle 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 la loi 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 la méthode de raccordement de lois, par exemple les lois lognormale-Pareto généralisée et Weibull-Pareto Généralisée. Ils proposent aussi d'utiliser la famille de distributions *Folded-t*. 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. De façon similaire, les auteurs de Reynkens et al. (2017) proposent d'utiliser une loi mélange d'erlang pour modéliser la portion sous le seuil tout en gardant la loi Pareto généralisée pour la portion au-dessus du seuil. Par ailleurs, les auteurs de Majid and Ibrahim (2021) proposent différents modèles par raccordement de lois, incluant la loi bêta généralisée de type 2 pour la portion sous le seuil, dans le but de modéliser les revenus par foyer en Malaisie. Enfin, la méthode par raccordement de loi a également été utilisée dans Grün and Miljkovic (2019) où les auteurs testent plus de 250 lois par raccordement pour modéliser la base *Danish fire loss date*.

Une autre approche présentée dans Laudagé et al. (2019) est de modéliser les pertes au-dessus d'un seuil par une loi Pareto généralisée et de modéliser les pertes sous le seuil par un GLM lognormale. Cette méthode présente l'avantage d'incorporer des covariables dans le modèle par régression pour les données situées en dessous du seuil, offrant ainsi la possibilité d'expliquer une portion du montant des pertes observées.

Pour Biffis and Chavez (2014), les auteurs utilisent une base de données obtenue auprès de 2 syndicats de *Lloyd's of London*. Ils ont pour objectif d'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 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'auteur montre qu'il est envisageable de modéliser la fréquence des incendies par un processus de Poisson. Il présente les résultats obtenu avec des données en provenance de Taiwan. Il teste le 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.

Dans un article récent, Chao (2022), l'auteur donne une façon de calculer une prime de réassurance en utilisant la méthode *Peak-Over-Threshold*. Il donne des formes analytiques pour le calcul de la prime d'un contrat, et ce, que le niveau de rétention soit égal ou supérieur au seuil spécifié.

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 Maxwell Boltzmann Bose Einstein Fermi Dirac (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 la réassurance pour le risque d'incendie. Cette méthode est aussi utilisée par Abramson (2022) pour la tarification dans le milieu de la construction.

On recommande Denuit et al. (2022) pour se familiariser avec la notion de partage de risque. Dans cet article, les auteurs introduisent diverses règles de partage de risque, notamment *Proportional mean* et *Conditional mean*. En plus, ils énumèrent les propriétés souhaitables que devraient avoir de bonnes règles de partage de risque. Dans Denuit and Robert (2020), les auteurs abordent le comportement de la règle de partage *Conditional mean* lorsqu'elle est utilisée avec les pertes majeures. Ils montrent qu'il y a deux cas de figure pour les pertes les plus élevées, soit un risque va prendre le dessus sur les autres ou l'évolution des contributions aura une tendance linéaire. L'article de Denuit and Robert (2022) offre des exemples concrets dans le domaine de l'assurance *peer-to-peer* (P2P), illustrant ainsi ces concepts de manière tangible et pratique.

1.3 Transfert de risque incendie via l'assurance IARD

1.3.1 Assurance IARD

Dans le glossaire du Bureau d'Assurance du Canada BAC (2023), on classifie l'industrie de l'assurance en deux branches. La première est l'assurance vie et maladie, qui n'est pas abordée dans ce rapport. La deuxième branche est l'assurance IARD (Incendies, Accidents et Risques Divers), appelée aussi assurance de dommages. Dans ce rapport on utilisera l'appellation assurance IARD. Cette branche de l'industrie couvre, en outre, l'assurance habitation, l'assurance automobile et l'assurance aux entreprises. L'ensemble de ces secteurs d'activité est présenté au Tableau 1.1.

Secteur d'activité	Primes nettes écrites
Assurance automobile	39 %
Assurance habitation ligne personnelle	22 %
Assurance aux entreprises	17 %
Assurance responsabilité	12 %
Autres	9 %

TABLEAU 1.1 – Principaux secteurs d'activité en assurance IARD au Canada, avec leurs pourcentages de primes nettes écrites

Dans la colonne 2 du Tableau 1.1, on trouve les pourcentages de primes nettes écrites pour chaque secteur d'activité pour l'année 2023. Ces pourcentages sont extraits d'*Assurances de dommages au Canada*¹, publié par le Bureau d'Assurance du Canada. On remarque que l'assurance automobile est une composante prépondérante de l'assurance IARD, représentant près de la moitié des primes nettes écrites. Dans ce rapport, on s'intéressera davantage à l'assurance aux entreprises, qui fait partie de l'assurance IARD.

1.3.2 Réassurance IARD

Une compagnie d'assurance IARD a la capacité financière pour émettre des contrats d'assurance aux particuliers et, pour la plupart d'entre elles, de rencontrer les engagements attachés à ces contrats. Par contre, il lui serait impossible d'émettre des contrats d'assurance chantier, ou plusieurs autres types de contrats d'assurance aux entreprises, et de remplir la totalité des engagements de ces contrats sans mettre en péril sa santé financière. Dans ce contexte, la compagnie d'assurance transfère une part du risque lié à ces contrats en achetant elle-même de l'assurance auprès de compagnies de réassurance.

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

1. Bureau d'Assurance du Canada, [www.ibc.ca/fr/ns/ressources/état-de-l-industrie/publication-assurances-de-dommages-au-canada](http://www.ibc.ca/fr/ns/ressources/etat-de-l-industrie/publication-assurances-de-dommages-au-canada)

proposent une excellente introduction à la réassurance et mentionnent une liste de raisons qui mènent les compagnies d'assurance à recourir à des compagnies de réassurance. Ces dernières permettent 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.

Les compagnies de réassurance permettent également de stabiliser les résultats commerciaux. En effet, une difficulté de l'assurance est de prévoir les coûts futurs qu'une police pourrait engendrer. Avec les compagnies de réassurance, ces coûts futurs inconnus sont remplacés par une prime, qui elle est connue d'avance.

En faisant affaire avec des compagnies de réassurance, les compagnies d'assurance 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 compagnies d'assurance peuvent aussi augmenter leur capacité à la souscription. En effet, en transférant une partie du risque aux compagnies de réassurance, elles peuvent accepter plus de contrats ou accepter des contrats qu'elles n'auraient pas pu prendre originellement.

Les compagnies d'assurance opèrent généralement dans une région relativement restreinte. Les compagnies de réassurance leur permettent d'accéder à des possibilités de diversification plus importantes. En effet, puisque les compagnies de réassurance opèrent généralement à l'international, les compagnies d'assurance qui font affaire avec elles ont accès à des marchés qui seraient autrement trop risqués pour elles.

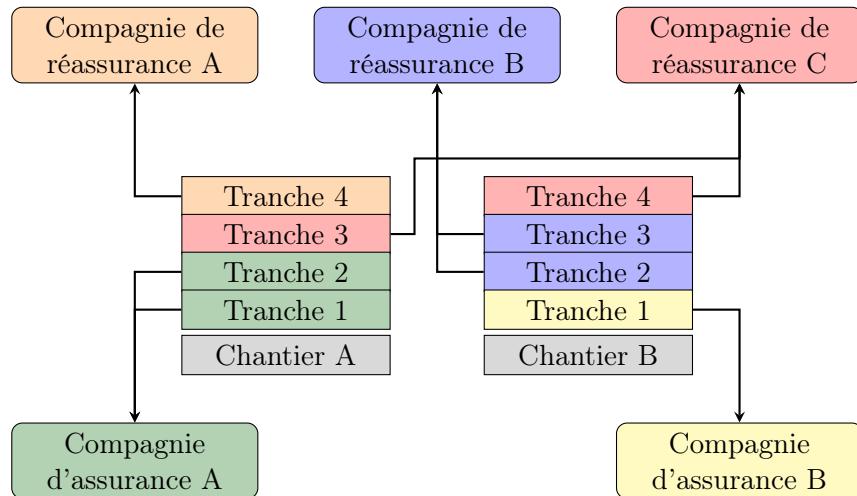


ILLUSTRATION 1.1 – Conventions de réassurance pour deux chantiers impliquant deux compagnies d'assurance et trois compagnies de réassurance

L'Illustration 1.1 présente un exemple d'échange entre des chantiers, des compagnies d'assurance et des compagnies de réassurance. Le défi pour l'actuaire de chacune des compagnies

d'assurance et de réassurance est de calculer la prime associée à chaque tranche à laquelle cette compagnie est exposée.

1.3.3 Défi de l'actuaire

Le défi des actuaires au sein d'une compagnie d'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'y paraît, car l'actuaire ne sait pas combien la compagnie d'assurance devra payer en dédommagement 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 qui est donné par

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

Par contre, en assurance, cette équation est un peu plus complexe. En effet, il est également possible de décomposer une prime d'assurance comme suit :

$$\text{Prime} = \text{Sinistre} + \left. \begin{array}{l} \text{Frais de règlement de sinistre} + \\ \text{Frais de souscription de sinistre} + \end{array} \right\} \text{Coût} + \underbrace{\text{Profit de souscription}}_{\text{Profit}}.$$

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 la compagnie d'assurance pourrait ne pas être suffisant.

1.4 Tarification en assurance IARD

Dans cette section, on présente les grandes lignes de la procédure de tarification en assurance IARD. Une attention particulière est portée au processus d'établissement des coûts, en mettant l'accent sur l'objectif de modélisation des pertes en sinistres incendie. À cet effet, on introduit le modèle de fréquence-sévérité.

1.4.1 Processus de tarification

Comme mentionné à la Section 1.3.3, le défi des actuaires est d'établir une prime juste et équitable. Pour y parvenir, les actuaires suivent un processus de tarification. L'Illustration 1.2, 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 de tarification. Plusieurs ouvrages ont traité de ce sujet, Parodi (2014), Denuit et al. (2007) et Ohlsson and Johansson (2010) pour l'assurance IARD et Michaelides et al. (1997) pour l'assurance aux entreprises.

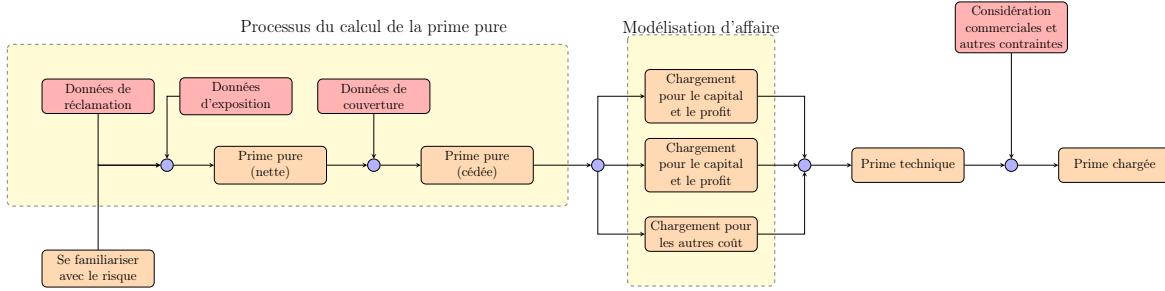


ILLUSTRATION 1.2 – Processus de tarification

Le processus de tarification se compose de plusieurs étapes. Il y a le processus de souscription, qui consiste à se familiariser avec le risque ou les activités de l'assuré, suivi du processus d'établissement des coûts, qui consiste à estimer le coût du risque à l'aide des données historiques disponibles. Dans les coûts, il faut aussi prendre en considération les dépenses générales de la compagnie, les revenus de placement ainsi que les marges de profit.

1.4.2 Modélisation actuarielle

Dans le cadre de cet essai, on s'intéresse au processus d'établissement des coûts. Pour ce faire, le modèle de fréquence-sévérité est utilisé. La fréquence représente le nombre de sinistres sur un intervalle de temps choisi, généralement une année en assurance IARD. La sévérité représente les coûts engendrés par chacun des sinistres survenus. Une représentation simplifiée du coût total est de considérer le produit de la fréquence et de la sévérité. En réalité, ce calcul est beaucoup plus complexe.

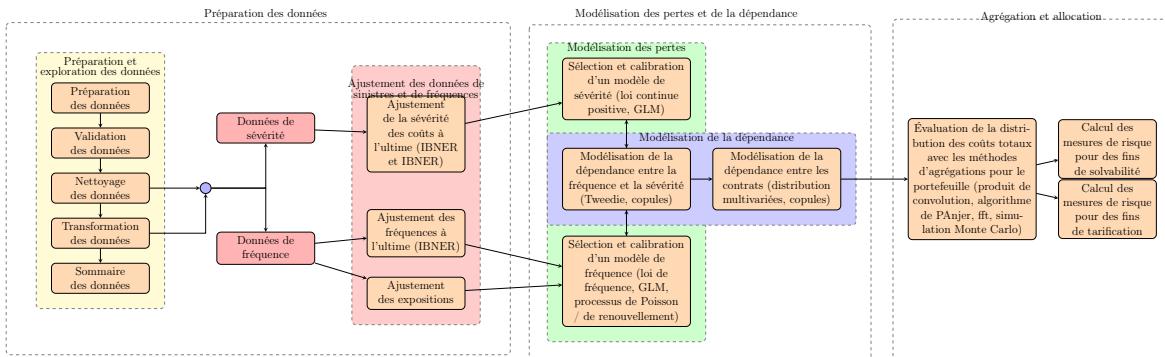


ILLUSTRATION 1.3 – Processus d'analyse des données

L'Illustration 1.3, extraite de l'Illustration 1.2 à la page 3 de Blier-Wong (2018), donne une idée globale du processus de calcul des coûts du risque.

On modélise le coût total par le processus agrégé $\underline{X} = \{X(t), t > 0\}$ où $X(0)$ et $X(t)$ sont définis selon le modèle de fréquence-sévérité

$$X(t) = \begin{cases} \sum_{k=1}^{N(t)} B_k, & N(t) > 0 \\ 0, & N(t) = 0. \end{cases} \quad (1.1)$$

On définit $\underline{N} = \{N(t), t > 0\}$ comme étant un processus de comptage où $N(0)$ et $N(t)$ sont le nombre de sinistres sur l'intervalle $(0, t)$. La variable aléatoire (v.a.) B_k est le montant du k -ième sinistres. Les hypothèses classiques de ce modèle sont :

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

Ces hypothèses permettent de calculer facilement la prime pure, qui est simplement l'espérance du processus agrégé \underline{X} ,

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

Ces hypothèses permettent également de modéliser le processus de comptage et la v.a. pour la sévérité séparément comme il n'y a aucune hypothèse de dépendance entre les deux. Pour le processus de comptage, on pourrait par exemple choisir un processus de Poisson homogène. Dans ce cas, pour la v.a. $N(t)$, on a les caractéristiques suivantes

- $N(t) \sim \text{Po}(\lambda t)$;
- $E[N(t)] = \lambda t$;
- $\text{Var}(N(t)) = \lambda t$.

Comme mentionné au Chapitre 6 de Parodi (2014), la loi gamma est fréquemment utilisée en assurance IARD pour modéliser la sévérité. Cependant, lorsqu'un actuaire travaille dans un secteur où les pertes peuvent être très élevées, par exemple en assurance aux entreprises, les lois lognormale et Pareto sont utilisées. Les caractéristiques de ces lois sont :

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

Il existe évidemment un large éventail de lois disponibles pour la modélisation de la fréquence et de la sévérité. Ici, seules les plus courantes sont présentées.

1.5 Loi Pareto

En actuariat, on a recours à la distribution Pareto pour décrire adéquatement le comportement aléatoire des montants élevés des sinistres. Plusieurs définitions existent pour la distribution Pareto. On commence en présentant la formulation dite *Pareto Lomax*. Ensuite, on décrit la représentation de cette distribution sous la forme distribution *Pareto généralisée*.

1.5.1 Distribution Pareto Lomax

La distribution Pareto Lomax est définie à l'aide de deux paramètres dont l'un est appelé l'indice de queue, que l'on désigne par la lettre grecque α et l'autre est le paramètre d'échelle, que l'on désigne par la lettre grecque λ . La fonction de répartition est donnée par

$$F_B(x) = 1 - \left(\frac{\lambda}{\lambda + x} \right)^\alpha, \quad x > 0.$$

L'espérance et la variance sont

$$E[B] = \frac{\lambda}{\alpha - 1}, \quad \alpha > 1, \quad Var(B) = \frac{\alpha \lambda^2}{(\alpha - 1)^2 (\alpha - 2)}, \quad \alpha > 2.$$

L'indice de queue α a une incidence sur la dangerosité de la distribution. Comme on l'observe à l'Illustration 1.4, la queue de la distribution de la loi Pareto lorsque l'indice de queue est de 1.5 est bien plus épaisse que celle avec un α de 3. Cela a inévitablement une incidence sur la prime, comme la probabilité d'observer des sinistres très élevés s'accroît lorsque l'indice de queue α diminue.

On présente au Tableau 1.2 la probabilité d'avoir des pertes supérieures à 1 million en faisant seulement varier l'indice de queue.

α	$\Pr(B > 1000000)$
2.5	0.003
2.0	0.010
1.5	0.032
1.2	0.063

TABLEAU 1.2 – Comparaison de la probabilité d'avoir un sinistre supérieur à 1 million

On observe bien le phénomène mentionné plutôt. Plus le α diminue, plus la probabilité d'avoir des sinistres supérieurs à 1 million augmente. En effet, lorsque le α passe de 2.5 à 1.2, cette probabilité est 21 fois plus élevée.

1.5.2 Distribution Pareto généralisée

La distribution Pareto Lomax admet aussi une représentation sous la forme de la distribution Pareto généralisée. On utilise la distribution Pareto généralisée pour modéliser le comporte-

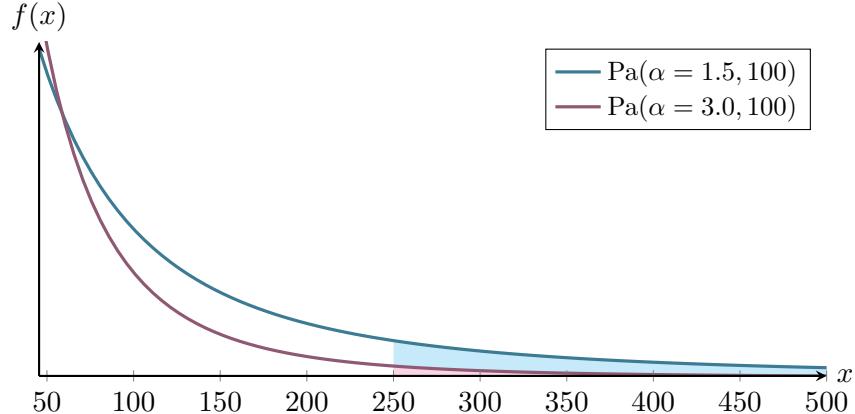


ILLUSTRATION 1.4 – Comparaison de la densité de deux lois Pareto selon $\alpha \in (1.5, 3)$

ment dans la queue de la distribution. La fonction de répartition est donnée par

$$G_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \frac{\xi}{\beta}x\right)^{-\frac{1}{\xi}}, & \xi \neq 0, \\ 1 - e^{-\frac{x}{\beta}}, & \xi = 0 \end{cases}, \quad x > 0.$$

En raison de son utilité pour la modélisation de la queue de la distribution, il est également possible d'ajouter un paramètre de translation. Alors, la fonction de répartition est donnée par

$$G_{\xi,\beta,\theta}(x) = \begin{cases} 1 - \left(1 + \frac{\xi}{\beta}(x - \theta)\right)^{-\frac{1}{\xi}}, & \xi \neq 0, \\ 1 - e^{-\frac{x-\theta}{\sigma}}, & \xi = 0 \end{cases}, \quad x > \theta.$$

L'espérance et la variance sont

$$E[B] = \frac{\beta}{1 - \xi} + \theta, \quad \xi < 1 \quad \quad Var(B) = \frac{\beta^2}{(1 - \xi)^2(1 - 2\xi)}, \quad \xi < \frac{1}{2}.$$

Afin de présenter les résultats, on utilisera les transformations suivantes :

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

Ces transformations permettent d'obtenir les paramètres plus familiers de la loi Pareto. Elles seront utilisées tout au long du rapport pour présenter les résultats relatifs à la loi Pareto généralisée.

1.6 Théorie des valeurs extrêmes

La modélisation des valeurs extrêmes est une composante essentielle du domaine de l'actuariat, qui vise à évaluer et gérer les risques extrêmes. Les valeurs extrêmes, ou événements rares ont

un impact significatif sur les résultats financiers et les réserves des compagnies d'assurance, ainsi que sur la prise de décision en matière de gestion des risques. Le risque d'incendie est un exemple où des valeurs extrêmes peuvent survenir.

La théorie des valeurs extrêmes, Pickands III (1975), est un outil mathématique essentiel pour la modélisation des événements extrêmes. En effet, on s'intéresse à la distribution généralisée de la valeur extrême (GEV), qui est justement utilisée pour décrire le comportement du maximum d'un échantillon de v.a. i.i.d. La fonction de répartition de cette distribution est donnée par

$$H_{\xi, \sigma, \theta} = \begin{cases} \exp \left\{ - \left(\frac{1}{1 + \frac{\xi}{\sigma}(x - \theta)} \right)^{\frac{1}{\xi}} \right\} & , \xi < 0, 0 < x < -\frac{\sigma}{\xi} \quad (\text{Weibull}) \\ \exp \left\{ - \exp \left\{ - \frac{x - \theta}{\sigma} \right\} \right\} & , \xi = 0, x > 0 \quad (\text{Gumbel}) \\ \exp \left\{ - \left(\frac{1}{1 + \frac{\xi}{\sigma}(x - \theta)} \right)^{\frac{1}{\xi}} \right\} & , \xi > 0, x > 0 \quad (\text{Fréchet}). \end{cases} \quad (1.2)$$

La fonction de répartition de la distribution GEV est souvent présentée sous une forme compacte,

$$H_\xi = \begin{cases} \exp \left\{ - \left(1 + \frac{\xi}{\sigma}(x - \theta) \right)^{-\frac{1}{\xi}} \right\} & , \xi \neq 0 \\ \exp \left\{ - \exp \left\{ - \frac{x - \theta}{\sigma} \right\} \right\} & , \xi = 0 \end{cases}. \quad (1.3)$$

Cette représentation est plus utile pour l'estimation comme il n'est pas nécessaire de faire une hypothèse sur la loi sous-jacente, Weibull, Gumbel ou Fréchet.

Les bases théoriques sont bien expliquées dans McNeil (1997) ou dans le livre Coles et al. (2001). Essentiellement, on considère une suite de v.a. i.i.d. X_1, X_2, \dots . On définit le maximum des n premières v.a. par

$$M_n = \max\{X_1, \dots, X_n\}.$$

On s'intéresse à la convergence en distribution du maximum. Soit $x_F = \sup\{x; F(x) < 1\}$, le point extrême droit de F . Il est possible de montrer que

$$\lim_{n \rightarrow \infty} M_n = x_F.$$

La valeur de x_F peut être finie ou infinie. Cependant, on remarque que la maximum converge vers une distribution dégénérée à x_F , cela a le désavantage de ne donner pratiquement aucune information. Pour en savoir plus sur l'ordre de grandeur du maximum, on s'intéresse à la convergence faible du maximum centré et réduit,

$$\frac{M_n - b_n}{a_n}, \quad a_n > 0.$$

On suppose que ces entiers normalisants $a_n > 0$ et b_n existent tel que

$$\Pr \left(\frac{M_n - b_n}{a_n} \leq x \right) = F(a_n x + b_n)^n \rightarrow H(x), \quad \text{quand } n \rightarrow \infty$$

où $H(x)$ est une fonction de répartition non dégénérée. Si cette condition est respectée, alors on dit que F fait partie du domaine d'attraction du maximum de la distribution H , qui est noté $F \in DAM(H)$. Le théorème de Fisher-Tippett, Fisher and Tippett (1928), indique quelle est la distribution de H .

Théorème 1 (Fisher-Tippett). *Soit X_1, X_2, \dots une suite de variables aléatoires i.i.d. S'il existe des constantes de standardisation $a_n > 0$ et $b_n \in \mathbb{R}$ et une distribution non dégénérée H tel que*

$$\frac{M_n - b_n}{a_n} \xrightarrow{\mathcal{D}} H,$$

alors la seule fonction de répartition non dégénérée H est la fonction de répartition H_ξ, σ de la distribution GEV.

Le Théorème 1 indique que si le maximum centré et réduit converge vers une distribution non dégénérée, alors il s'agit absolument de la distribution GEV. C'est là un premier résultat important de la théorie des valeurs extrêmes. Ensuite, on veut faire un lien entre la loi Pareto généralisée et la distribution GEV. On commence par regarder la distribution de l'excédent.

Définition 1 (Distribution de l'excédent). *Soit la v.a. X et la v.a. W_u représentant l'excédent d'un sinistre par rapport à un seuil θ . La v.a. W_θ est définie par*

$$W_\theta = (X - \theta | X > \theta).$$

La fonction de répartition de la v.a. W_θ est donnée par

$$F_{W_\theta}(X) = \frac{F_X(x + \theta) - F_X(\theta)}{1 - F_X(\theta)}, \quad x > 0.$$

On souhaite approximer la distribution des excédents par la loi Pareto généralisée. Pour ce faire, on utilise le théorème de Pickands–Belkema–de Haan, qui est un deuxième théorème fondamental de la théorie des valeurs extrêmes.

Théorème 2 (Pickands–Belkema–de Haan). *Pour tout $\xi \in \mathbb{R}$, $F \in MDA(H_\xi)$ si et seulement si*

$$\lim_{\theta \rightarrow \infty} \sup_{x \geq 0} |F_{W_u}(x) - G_{\xi, \sigma(\theta)}(x)| = 0$$

pour une certaine fonction σ positive et mesurable.

Le Théorème 2 donne des bases théoriques raisonnables pour approximer la distribution des excédents par la loi Pareto généralisée pour un seuil θ assez élevé. On appellera cette méthode *Peak-Over-Threshold* (POT). À l'Illustration 1.5 on présente un exemple de cette méthode. À l'Illustration 1.5a on la fonction de répartition empirique. L'objectif de la méthode POT est de garder la portion empirique pour les données sous le seuil et ensuite de modéliser la queue de la distribution par une loi Pareto généralisée, comme montré à l'Illustration 1.5b.

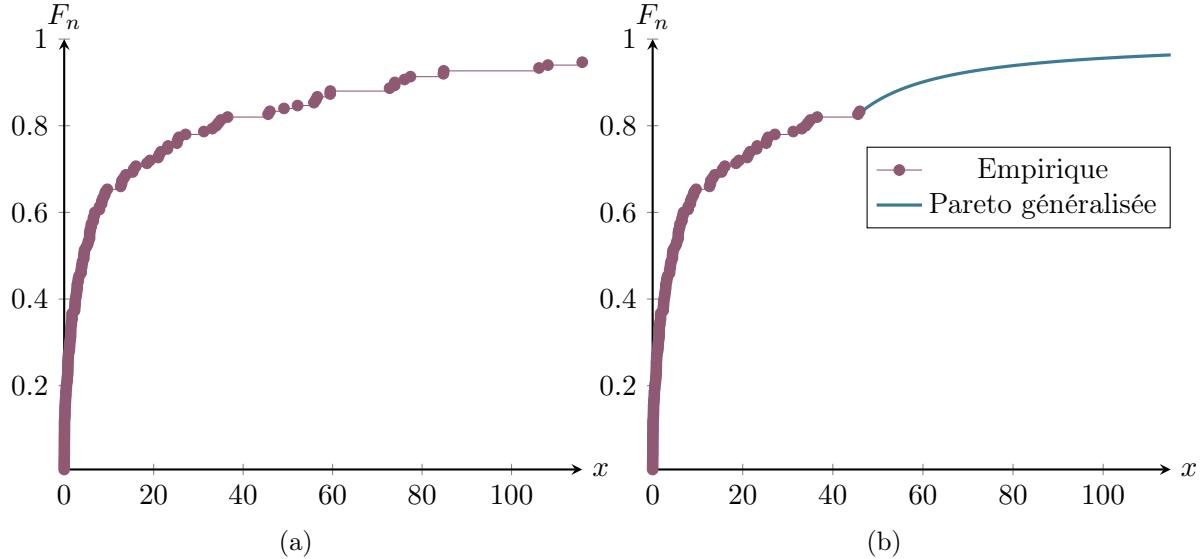


ILLUSTRATION 1.5 – Exemple de la méthode *Peak-Over-Threshold* (POT)

1.7 Mesures de risque de distorsion

Dans cette section, on se concentre sur les mesures de risque de distorsion. Ces mesures sont essentielles pour quantifier le risque une fois qu'on a sélectionné des lois appropriées pour modéliser les pertes en sinistres d'incendie.

Soit la v.a. non-négative X , qui représente les pertes pour un assureur. On définit l'espérance de X par

$$E[X] = \int_0^\infty (1 - F_X(x))dx.$$

On peut également définir une fonction de distorsion $g : [0, 1] \rightarrow [0, 1]$ non décroissante, telle que $g(0) = 0$ et $g(1) = 1$, et calculer l'espérance déformée, notée $R[X]$, défini comme

$$R[X] = \int_0^\infty g(1 - F_X(x))dx.$$

On note que $g(1 - F_X(x))$ est une fonction non décroissante de $F_X(x)$. Il existe une fonction de distorsion g qui implique que l'assureur préfère les pertes Y aux pertes X si et seulement si $R[Y] \leq R[X]$, voir Denuit et al. (2006). En d'autres termes, si la mesure de risque est plus faible pour une variable aléatoire, cela indique que le risque associé est moindre, rendant ce risque plus attrayant pour une compagnie d'assurance.

Si la fonction de distorsion g est dérivable, on peut réécrire « l'espérance déformée » comme suit :

$$R[X] = \int_0^1 F^{-1}(u)g'(1-u)du = \int_0^\infty xg'(\bar{F}_X(x))f_X(x)dx = E[Xg'(\bar{F}_X(X))].$$

Pour démontrer ce résultat, on effectue une intégration par partie avec $u = g(1 - F(x))$ et $dv = dx$:

$$\begin{aligned} R[F] &= xg(1 - F(x)) \Big|_0^\infty + \int_0^\infty xg'(1 - F(x))f(x)dx \\ &= \int_0^\infty xg'(1 - F(x))f(x)dx \quad (u = F(x)) \\ &= \int_0^1 F^{-1}(u)g'(1 - u)du \end{aligned}$$

Le facteur $g'(1 - F(x))$ agit comme un poids appliqué à la distribution de la variable aléatoire X . Comme $g(1 - F_X(x))$ est une fonction non décroissante de $F_X(x)$, on applique plus de poids aux valeurs élevées de X . Cette caractéristique s'avère particulièrement utile lors de la comparaison des risques en assurance, où les distributions ont souvent des queues lourdes.

Des mesures couramment utilisées en science actuarielle sont les mesures VaR et $TVaR$. La mesure VaR est définie comme suit :

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

Cette mesure représente la prime minimale pour garantir, avec un niveau de confiance de $100\kappa\%$, la capacité de payer une réclamation en cas de sinistre. Le défaut de cette mesure est qu'elle ne donne pas d'information sur des valeurs supérieures à $VaR_\kappa(X)$. Pour remédier à cela, on considère la mesure $TVaR$, qui est définie comme suit :

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

Elle correspond à la moyenne des mesures $VaR_\kappa(X)$ pour des valeurs de u supérieures à κ . Elle donne donc de l'information pour les valeurs supérieures à $VaR_\kappa(X)$.

D'autres mesures de risque de distorsion ont été proposées dans la littérature, notamment :

- *Proportional Hazards Transform (PHT)*, voir Wang (1995) ;
- *Wang Transformation (WT)*, voir Wang (2000) ;
- *Gini Shortfall (GS)*, voir Furman et al. (2017).

La mesure de risque *Gini Shortfall* est particulièrement intéressante, car elle permet également d'avoir un indice de la variabilité dans la queue de la distribution sans avoir besoin que le deuxième moment existe. De son côté, la mesure $TVaR$ donne seulement une idée de la moyenne dans la queue de la distribution.

La mesure *Gini Shortfall* est définie comme suit

$$GS_\kappa^\lambda(X) = ES_\kappa(X) + \lambda TGini_\kappa(X) \quad \kappa \in [0, 1], \quad \lambda \geq 0,$$

où $ES(X) = TVaR(X)$ si la v.a. X est continue et $TGini_\kappa(X)$ est la *Tail-Gini functional*, qui est définie, pour une v.a. continue, par

$$\begin{aligned} TGini_\kappa(X) &= E[|X^* - X^{**}|] \\ &= \frac{2}{(1-\kappa)^2} \int_{\kappa}^1 F_X^{-1}(u)(2u - (1+\kappa)) \, du \\ &= \frac{2}{(1-\kappa)^2} \int_{VaR_\kappa(X)}^{\infty} [(1-\kappa)\bar{F}(x) - (\bar{F}(x))^2] \, dx, \end{aligned}$$

où les variables X^* et X^{**} sont des copies indépendantes de X .

Mesure Risque	Exponentielle $Exp(x_0, \beta)$	Lognormale $LN(x_0, \mu, \sigma)$	Pareto généralisée $PaG(\xi, \theta, \beta)$
VaR	$x_0 - \beta \ln(1-\kappa)$	$x_0 + e^{\mu + \sigma \Phi^{-1}(\kappa)}$	$\frac{\beta}{\xi} \left((1-\kappa)^{-\xi} - 1 \right) + \theta$
$TVaR$	$x_0 - \beta \ln(1-\kappa) + \beta$	$x_0 + \frac{1}{1-\kappa} e^{\mu + \sigma^2/2} \Phi(\sigma - \Phi^{-1}(\kappa))$	$\frac{\beta}{\xi} \left(\frac{1}{1-\xi} (1-\kappa)^{-\xi} - 1 \right) + \theta$
PHT	$x_0 + \frac{\beta}{r}$	N.D.	$\theta + \frac{\beta}{r-\xi}, \ r > \xi.$
WT	N.D.	$x_0 + e^{\mu + \lambda \sigma + \sigma^2/2}$	N.D.
GS	$x_0 + \beta \ln(1-\kappa) + \beta(1+\delta)$	N.D.	$\frac{\beta}{\xi} \left(\frac{(1-\kappa)^{-\xi}}{1-\xi} - 1 \right) + \frac{1}{(1-\kappa)\xi} \frac{2\lambda\beta}{(1-\xi)(2-\xi)} + \theta, \ \xi < 1.$

TABLEAU 1.3 – Forme analytique des mesures de risque VaR , $TVaR$, PHT , WT et GS pour les lois exponentielle, lognormale tronquée et Pareto généralisée

Au Tableau 1.3, on présente les formes analytiques de ces mesures de risque pour les lois exponentielle tronquée, lognormale tronquée et Pareto généralisée. Pour les deux premières, on réfère le lecteur ou la lectrice à Upreeet and Brazauskas (2022). Le développement des expressions pour la loi Pareto généralisée se trouve à l'Annexe F.

Chapitre 2

Description de deux bases de données sinistres incendie

2.1 Introduction

Dans ce chapitre, on présente deux bases de données de pertes en sinistres dues au péril incendie :

- Données de périls incendie de la ville de Toronto ([DataToronto](#)) ;
- *National Fire Information Database* ([DataNFID](#))

On utilise ces bases de données afin de construire les modèles actuariels permettant de modéliser les pertes en sinistres dues au péril incendie. On utilisera la base de données DataToronto au Chapitre 3 pour faire une analyse complète de la modélisation de la sévérité. Pour la base de données DataNFID elle sera principalement utilisée au Chapitre 4 pour faire le partage de risques selon le type de construction d'un bâtiment.

2.2 Base de données de périls incendie de la ville de Toronto

2.2.1 Description des données

La base de données est accessible sur le site *City of Toronto's Open Data Portal*¹. Le jeu de données contient de l'information sur 17 536 cas d'incendies de 2011 à 2018. Pour chaque incident, il y a 43 variables explicatives, au Tableau 2.1 on présente 7 d'entre elles. La liste complète se trouve dans l'Annexe D.

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

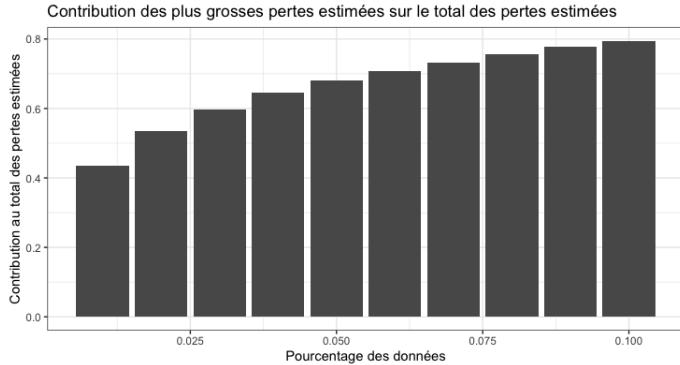


ILLUSTRATION 2.1 – Contribution des plus grosses pertes estimées au total des pertes estimées

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 estimés.
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.

TABLEAU 2.1 – Description de 7 variables explicatives de la base de données DataToronto

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 que l'on retrouve plusieurs occurrences d'un même montant. Afin de faciliter l'analyse, il a été nécessaire de retirer les 1909 données dont la perte estimée est indisponible. De plus, on a aussi retiré les 2056 données dont les pertes estimées sont de 0. Après ces retraits, il reste $n = 13\,571$ données utilisables.

Au Tableau 2.2, 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

TABLEAU 2.2 – Statistiques descriptives des $n = 13\,571$ pertes estimées retenues de la base de données DataToronto

La moyenne des pertes est de 40 296 \$ CAN, ce qui est considérablement plus élevé que la médiane des pertes, qui est de 5 000 \$ CAN. De son côté, la variance des pertes est extrêmement élevée, ce qui pourrait potentiellement indiquer qu'elle n'existe pas. On analyse également la proportion que représentent les pertes élevées par rapport au montant total des pertes. Comme le montre l'Illustration 2.1, plus de 65 % des pertes totales sont dues à 5 % des données, et

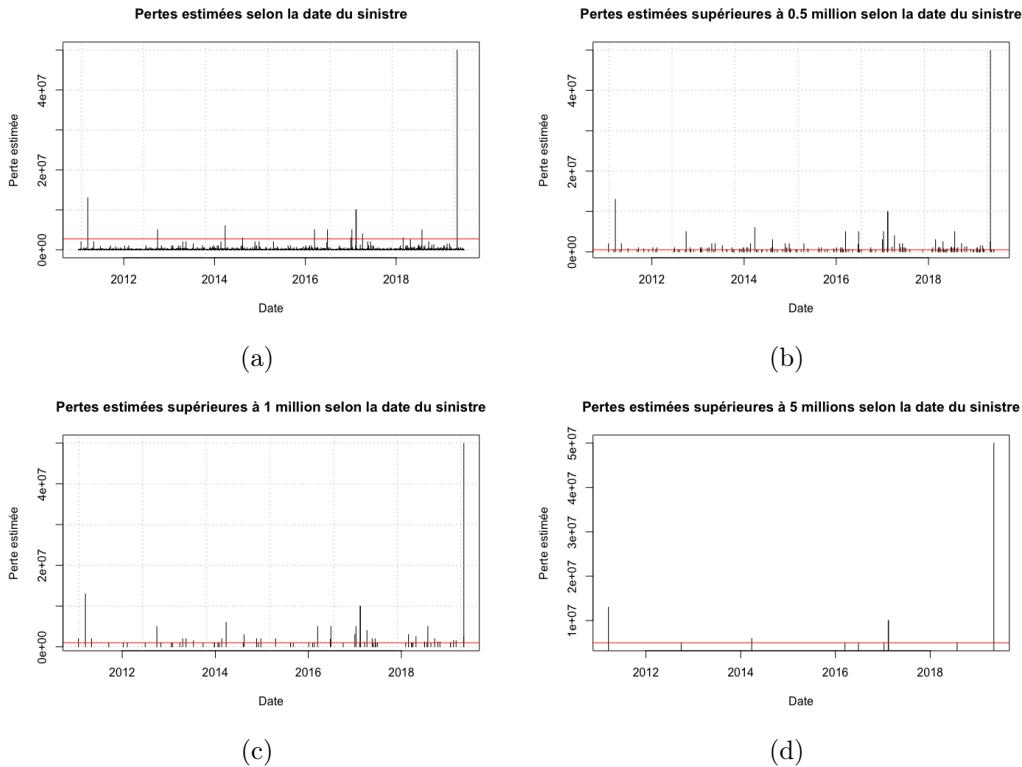


ILLUSTRATION 2.2 – Pertes estimées selon la date du sinistre

environ 80 % des pertes totales proviennent de 10 % des données. De plus, l’Illustration 2.2 permet de voir que certaines valeurs sont extrêmement élevées comparativement aux autres. En effet, l’Illustration 2.2a 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 2.2b, 2.2c et 2.2d 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 2.2, on observe que la perte maximale de 50 millions n’est arrivée qu’une seule fois durant ces années. Cette observation ne peut toutefois être considérée comme aberrante, car elle est typique des données de sinistres incendie. On observe ce phénomène dans plusieurs bases de données de sinistre incendie.

À l’Annexe B, on présente les statistiques descriptives de quatre autres bases de données et on peut constater que la valeur maximale pour chaque base de données est extrêmement élevée comparativement à la moyenne. On observe également ce phénomène dans la célèbre base de données *Danish data on large fire insurance losses* présentée dans McNeil (1997). En effet, dans cette base de données, la moyenne est de 3.39 et la valeur maximale est de 263.25.

On a aussi estimé l’intensité λ du processus de Poisson décrivant l’avènement des sinistres incendie avec la base de données DataToronto. Les détails sont fournis à l’Annexe A.

2.2.2 Analyse descriptive

Afin de mieux comprendre les données, on commence par une analyse descriptive qui fournira des indications sur les lois à utiliser ultérieurement. On produit l'histogramme de la densité et de la log densité, la fonction de répartition empirique, ainsi que les *QQplots* pour avoir une idée de la distribution des données.

Au Tableau 2.3, on observe qu'il y a peu de pertes extrêmement élevées par rapport aux sinistres plus courants, mais ces sinistres élevés représentent une grande partie des pertes totales, voir Illustration 2.1.

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 × 10 ⁴ , 5 × 10 ⁵ [172	181	199	207	195	219	222	269	132	1796
[5 × 10 ⁵ , 1 × 10 ⁶ [6	7	11	14	13	16	4	17	11	99
[1 × 10 ⁶ , 1 × 10 ⁷ [3	5	9	9	3	7	10	14	5	65
[1 × 10 ⁷ , 5 × 10 ⁷ [1	0	0	0	0	0	2	0	0	3
[5 × 10 ⁷ , ∞[0	0	0	0	0	0	0	0	1	1

TABLEAU 2.3 – Nombre de données dans des tranches de montants de sinistres

On observe à l'Illustration 2.3a que l'histogramme de la densité n'est pas facilement interprétable en raison des valeurs extrêmes. On note que l'axe des x arrête à 300 000 pour que la lecture soit plus facile. Il semble y avoir des modes à certaines valeurs, par exemple 50 000, 100 000 et 150 000. Cela est probablement dû au fait que le montant de la perte 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.

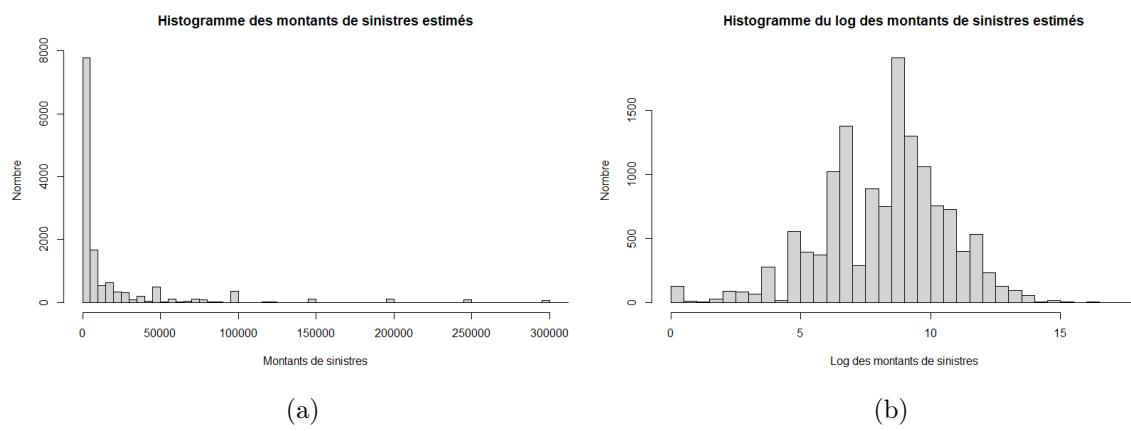


ILLUSTRATION 2.3 – Histogramme des pertes et du log des pertes

Afin de rendre la lecture plus claire, on regarde le log des montants de sinistre. De cette façon, on a des montants moins élevés, ce qui permet d'avoir une échelle plus facile à lire. On constate à l'Illustration 2.3b que le montant des sinistres semble être distribué selon une loi normale 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

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

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

À l'aide de la Proposition 1 et de l'Illustration 2.3b, on déduit que les pertes pourraient 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.

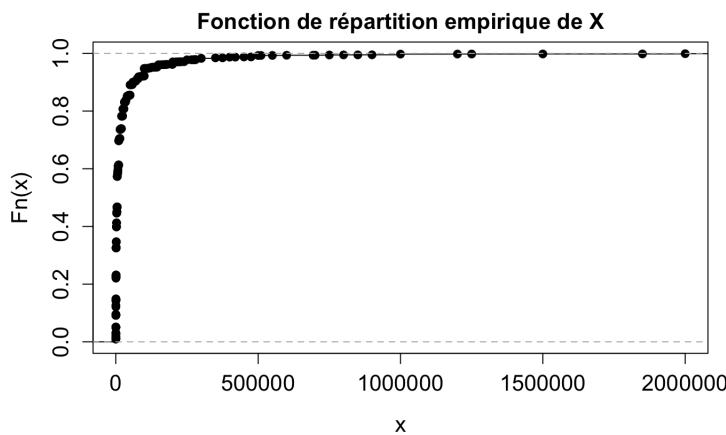


ILLUSTRATION 2.4 – Fonction de répartition empirique des montants de sinistre

À l'Illustration 2.4, on voit qu'il y a une probabilité non-négligeable d'avoir des sinistres extrêmement élevés comparativement à l'ensemble des données. Cette remarque met en évidence

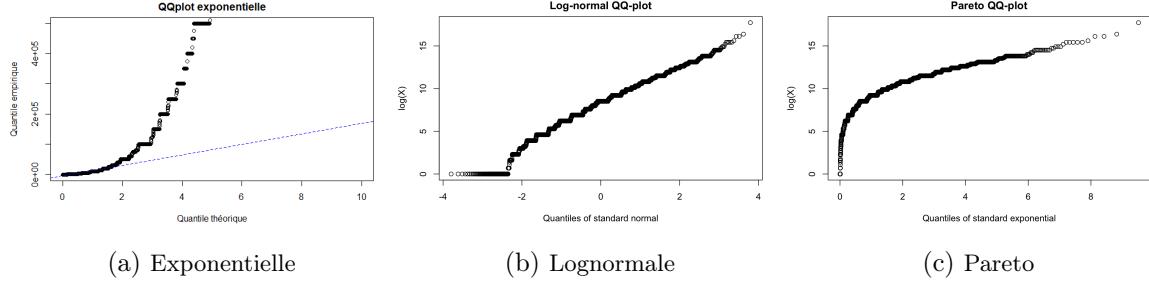


ILLUSTRATION 2.5 – *QQplots* exponentielle, lognormale et Pareto

qu'on a 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 autre technique pour obtenir une idée initiale de la distribution des pertes consiste à examiner les *QQ-plots* de certaines lois. Si les données suivent une loi particulière, le *QQ-plot* correspondant devrait montrer une droite à 45° .

Définition 2. 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 3. 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 4. 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.

À l'Illustration 2.5a, on voit que la distribution des pertes a une queue plus lourde qu'une loi exponentielle, comme les points passent bien au-dessus de la droite en bleu. Ce résultat confirme l'intuition d'une distribution à queue lourde, comme la loi exponentielle est généralement considérée comme le point milieu entre les distributions à queues légères et à queues lourdes. Il est préférable de produire des *QQ-plots* pour des lois à queue lourde, telles que la loi lognormale et la loi Pareto. On constate à l'Illustration 2.5b que le *QQ-plot* de la loi lognormale produit environ une droite, ce qui est une autre indication que les montants de sinistres suivent une loi lognormale. Par contre, sur le *QQplots* de la loi Pareto, Illustration 2.5c, on observe 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 2.5 sont produits à l'aide des fonctions `ExpQQ`, `LognormalQQ` et `ParetoQQ` du package `ReIns` de [R, Reynkens and Verbeelen \(2020\)](#).

2.3 Base de données de la *National Fire Information Data Base*

2.3.1 Description des données

Cette base de données a été mise en place par Len Garris, directeur de la recherche pour le Conseil national de sécurité incendie autochtone et professeur adjoint à l'École de culture et médias. L'objectif était de rassembler une grande quantité d'informations sur les sinistres incendie au Canada. Elle a été utilisée dans divers travaux, notamment Zheng et al. (2022).

La base de données DataNFID contient de l'information sur 467 929 cas d'incendies pour les années 2005 à 2015. Pour chaque incident, il y a 136 variables explicatives. On peut subdiviser ces variables explicatives en différentes catégories, voir Tableau 2.4. On donne des exemples de variables pour chacune des catégories pour lesquelles l'information pourra éventuellement être utile.

Catégories	Variables	Descriptions
Information sur l'incident	YEAR : MONTH : DAY : RESPONSE : CREWSIZE :	années de l'incident mois de l'incident jour de l'incident temps de réponse des secours nombre de secouristes initial
Information sur la propriété	GENCONST : PROPCLAS : PROPGRP : PROPCLAS : RISKVALA :	type de construction utilité du bâtiment secteur d'utilité du bâtiment secteur d'utilité spécifique du bâtiment valeur du bâtiment
Protection contre le feu	SPRINPRO : FIREDET :	système d'arroseur système de détection de fumé
Montant de perte du au feu	DOLLOSSA : DOLLOSSB : DOLLOSSC :	perte du au bâtiment et véhicule perte du au contenu DOLLOSSA + DOLLOSSB
Circonstance et facteur contribuant au feu		
Découverte du feu et action faite		
Information sur les victimes		
Autres données dans le domaine social		

TABLEAU 2.4 – Catégories des variables explicatives et exemple pour certaines

Au Tableau 2.4, seules les variables utiles pour ce projet sont présentées. Cependant, bien d'autres variables pourraient être utiles dans des contextes différents. Par exemple, en ingénierie, les variables suivantes pourraient être pertinentes :

- **IGNIOBJ** : objet d'allumage ;
- **FUELERGY** : combustible ou énergie associés à l'objet d'allumage ;
- **MATERIAL** : premier matériau allumé.

D'un autre côté, les services de sécurité incendie pourraient utiliser ces variables :

- **ACTION** : action faite contre l'incendie ;
- **RESPONSE** : temps de réponse des premiers pompiers ;
- **CAUSFAIL** : cause empêchant les pompiers de sortir.

2.3.2 Données manquantes

L'une des principales difficultés de cette base de données réside dans le nombre considérable de données manquantes. En effet, dans tous les cas, au minimum une variable est indisponible. Ainsi, bien que la base de données compte plus de 400 000 cas d'incendies, en réalité, on travaillera avec un nombre inférieur une fois les données manquantes retirées.

2.3.3 Variables retenues

On s'intéresse premièrement à modéliser de façon générale les pertes engendrées par les incendies. Comme on s'intéresse aux périls incendies sur des chantiers, il est plus approprié de se concentrer sur les pertes liées seulement au bâtiment et non au bâtiment et son contenu. C'est pour cela qu'on utilisera la variable **DOLLOSSA** plutôt que **DOLLOSSC**. Par contre, la variable **DOLLOSSA** inclut aussi les dommages aux véhicules. On note également que les montants de pertes sont des pertes estimées, ce ne sont pas les montants payés par un assureur.

Pour utiliser cette variable, il ne reste plus que 102 635 données. En effet, pour des fins d'analyse, on a retiré 360 181 données manquantes et 5113 données dont la perte estimée était de 0.

Au Tableau 2.5, on présente les statistiques descriptives de la variable **DOLLOSSA** après modifications.

Min.	1 ^{er} Qu.	Médiane	Moyenne	3 ^{er} Qu.	Max.
1	1000	5000	75 800	30 000	200 000 000

TABLEAU 2.5 – Statistiques descriptives des $n = 102\ 635$ pertes retenues de la base de données **DataNFID**

On observe les mêmes résultats que pour la base de données **DataToronto** : une moyenne plus élevée que la médiane. De plus, on remarque que la perte maximale observée de 200 millions

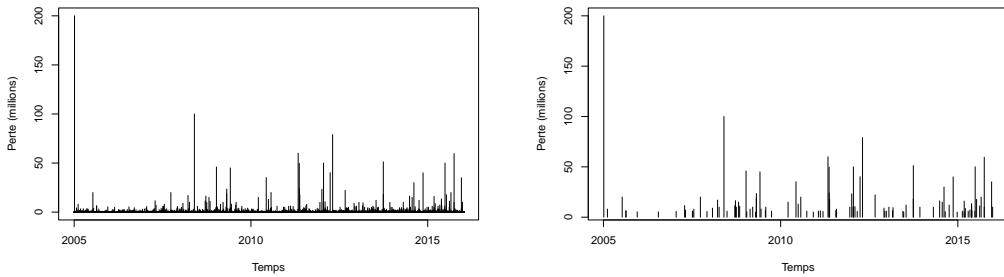


ILLUSTRATION 2.6 – Pertes estimées selon la date du sinistre

est extrêmement éloignée de la moyenne. À l'aide de l'Illustration 2.6, on constate encore que certaines pertes sont vraiment plus élevées que les autres. L'Illustration 2.6b montre l'ensemble des sinistres supérieurs à 5 millions, au nombre de 102.

Au Tableau 2.6, on présente la fréquence des sinistres dans certaines tranches pour chaque année.

Intervalles	Nombre de données						
	2009	2010	2011	2012	2013	2014	2015
[1, 50 000[7134	6869	7581	8083	7531	7234	7428
[5 × 10 ⁴ , 5 × 10 ⁵ [1594	1431	2022	1902	1778	1811	1752
[5 × 10 ⁵ , 1 × 10 ⁶ [127	124	388	159	174	145	166
[1 × 10 ⁶ , 1 × 10 ⁷ [78	58	107	98	89	83	89
[1 × 10 ⁷ , 5 × 10 ⁷ [6	4	4	6	4	6	9
[5 × 10 ⁷ , ∞[0	0	1	2	1	0	2

TABLEAU 2.6 – Nombre de données par tranche de montants de sinistres

On constate de nouveau qu'il y a peu de montants de sinistres extrêmement élevés par rapport aux sinistres plus courants, mais que ces pertes élevées représentent une grande partie des pertes totales, comme le montre l'Illustration 2.7. En effet, on remarque que 1% des plus gros sinistres représentent plus de 40% de la perte totale, alors que si on prend les 10% plus gros sinistres, ces derniers représentent plus de 80% de la perte totale. De plus, on remarque également qu'à 2.5%, on dépasse déjà les 50% du montant total de perte.

On veut également être en mesure de modéliser le processus d'avènement des sinistres. Pour cela, il sera nécessaire d'avoir l'information sur la date complète des sinistres. On utilise donc les variables **YEAR**, **MONTH**, **DAY** pour créer une nouvelle variable qu'on nommera **TEMPS**. La variable **TEMPS** sera la date complète à laquelle le sinistre est survenu. Pour cette variable, on enlève les 264 091 données manquantes et on garde seulement les observations pour lesquelles on connaît la perte estimée. Il reste 100 063 données pour la modélisation du processus d'avènement des

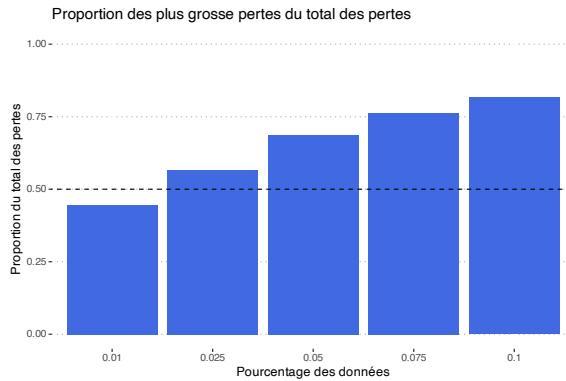


ILLUSTRATION 2.7 – Proportion des plus grosses pertes estimées du total des pertes estimées

sinistres.

Il peut également être intéressant d'examiner les variables liées à la protection contre le feu, puisqu'on a de l'information sur la présence ou non d'un système de protection. On pourra alors évaluer l'effet de ces protections sur les montants de pertes. On regarde brièvement les variables SPRINPRO et FIREDET. Au Tableau 2.7, on présente les pertes moyennes selon la présence d'un système d'arroseurs pour les pertes supérieures à 1 million. On note que pour cette analyse, on a seulement 265 cas d'incendie où toute l'information est disponible.

Arroseur	Perte moyenne	Nombre	Pourcentage
Système d'arroseurs	7 784 866	39	15%
Pas d'arroseur	2 781 761	226	85%

TABLEAU 2.7 – Pertes moyennes pour les sinistres supérieures à 1 million selon la présence d'un système d'arroseur

On constate que contrairement à ce qu'on pourrait s'attendre, la perte moyenne est plus élevée pour le bâtiment muni d'un système d'arroseur. On utilise la variable RISKVALA pour regarder la valeur des bâtiments. Comme le montre le Tableau 2.8, les bâtiments avec un système d'arroseur ont une valeur plus élevée, ce qui pourrait expliquer le résultat précédent.

Arroseur	Valeur moyenne du bâtiment
Système d'arroseurs	72 924 155
Pas d'arroseur	4 451 039

TABLEAU 2.8 – Valeur moyenne du bâtiment, pour les pertes supérieures à 1 million selon la présence d'un système d'arroseur

On fait une analyse relative de la perte moyenne selon la valeur moyenne du bâtiment. On constate que la perte moyenne avec système d'arroseur représente 10.7% de la valeur moyenne des bâtiments, alors que la perte moyenne sans système d'arroseur représente 62.5% de la

valeur moyenne des bâtiments. Donc, la présence d'un système d'arroseur a bien un effet sur la perte encourue suite à un incendie. On pourrait faire la même analyse préliminaire pour les détecteurs de feu, mais on constate que tous les bâtiments dont on a l'information possèdent un tel système.

On passe maintenant aux variables **GENCONST** et **PROPGRP**, qui donnent respectivement l'information sur le type de structure et le secteur d'activité du bâtiment. Les codes utilisés dans la base de données sont donnés au Tableau 2.9.

Classe	Description (officiel)
1 :	Combustible construction - open wood joist
2 :	Protected combustible construction - wood protected by plaster
3 :	Heavy timber construction
4 :	Non-combustible construction - exposed steel
5 :	Protected non-combustible construction - protected steel or concrete
Classe	Description (traduction)
1 :	Combustible
2 :	Combustible protégé
3 :	Bois massif
4 :	Non combustible
5 :	Non combustible protégé

TABLEAU 2.9 – Code pour la variable **GENCONST**

Classe	Description
1000 :	rassemblement
2000 :	institutionnelle
3000 :	résidentiel
4000 :	services professionnels et personnels
5000 :	bâtiment commercial
6000 :	entreprises de fabrication industrielle
7000 :	propriétés de stockage
8000 :	propriété diverses

TABLEAU 2.10 – Code pour la variable **PROPGRP**

À l'aide de ces variables, il est possible d'aller chercher de l'information sur les pertes totales, les pertes moyennes, les valeurs moyennes des bâtiments et le nombre d'incendies selon le type de construction et le secteur d'utilité du bâtiment, voir Illustration 2.8 et 2.9.

On constate à l'Illustration 2.8a que la perte totale est la plus élevée pour les constructions en bois protégé, ce qui est probablement expliqué par le fait qu'à l'Illustration 2.8b, on voit bien que 60% des incendies sont sur des bâtiments en bois protégé. Cependant, à l'Illustration 2.8c,

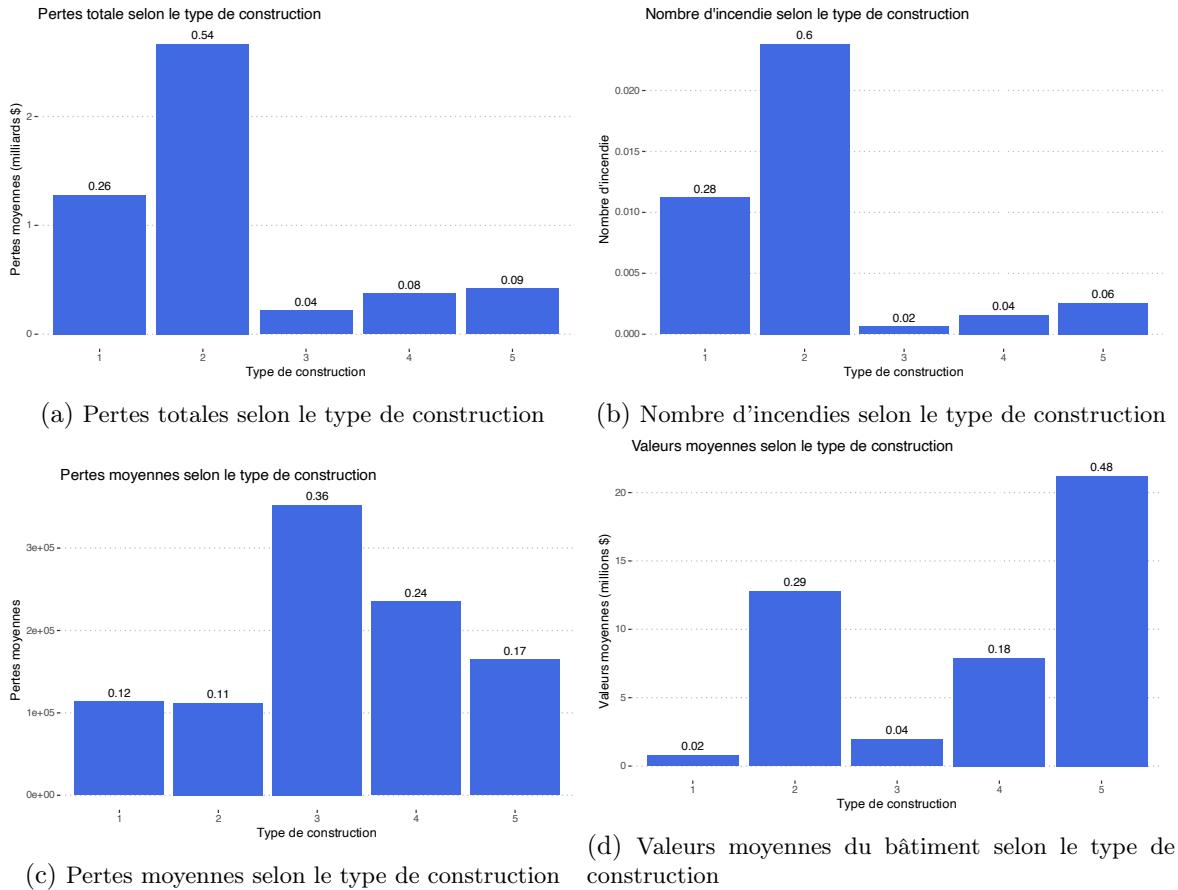


ILLUSTRATION 2.8 – Pertes totales, pertes moyennes, valeurs moyennes et nombre d’incendies selon le type de construction

on remarque que la perte moyenne est plus élevée sur les bâtiments construits en bois massif. On identifie également, à l’Illustration 2.8d, que ce sont les bâtiments construits en acier ou béton protégé qui ont la valeur moyenne la plus élevée.

On constate à l’Illustration 2.9 que la perte totale est la plus élevée pour les bâtiments résidentiels, ce qui est probablement expliqué par le fait qu’à l’Illustration 2.9b, on voit bien que 71% des incendies sont sur des bâtiments de type résidentiel. Cependant, à l’Illustration 2.9c on remarque que la perte moyenne est le plus élevée sur les bâtiments de type industriel. On identifie également, à l’Illustration 2.9d, que ce sont les bâtiments de type institutionnel qui ont la valeur moyenne la plus élevée.

Comme on s’intéressera principalement à la modélisation des pertes élevées, il est intéressant d’aller chercher les caractéristiques des 10 plus pertes les plus élevées. On présente ces sinistres au Tableau 2.11. On note qu’on garde les sinistres pour lesquels l’information sur le bâtiment était disponible, ce qui en laisse 39 775.

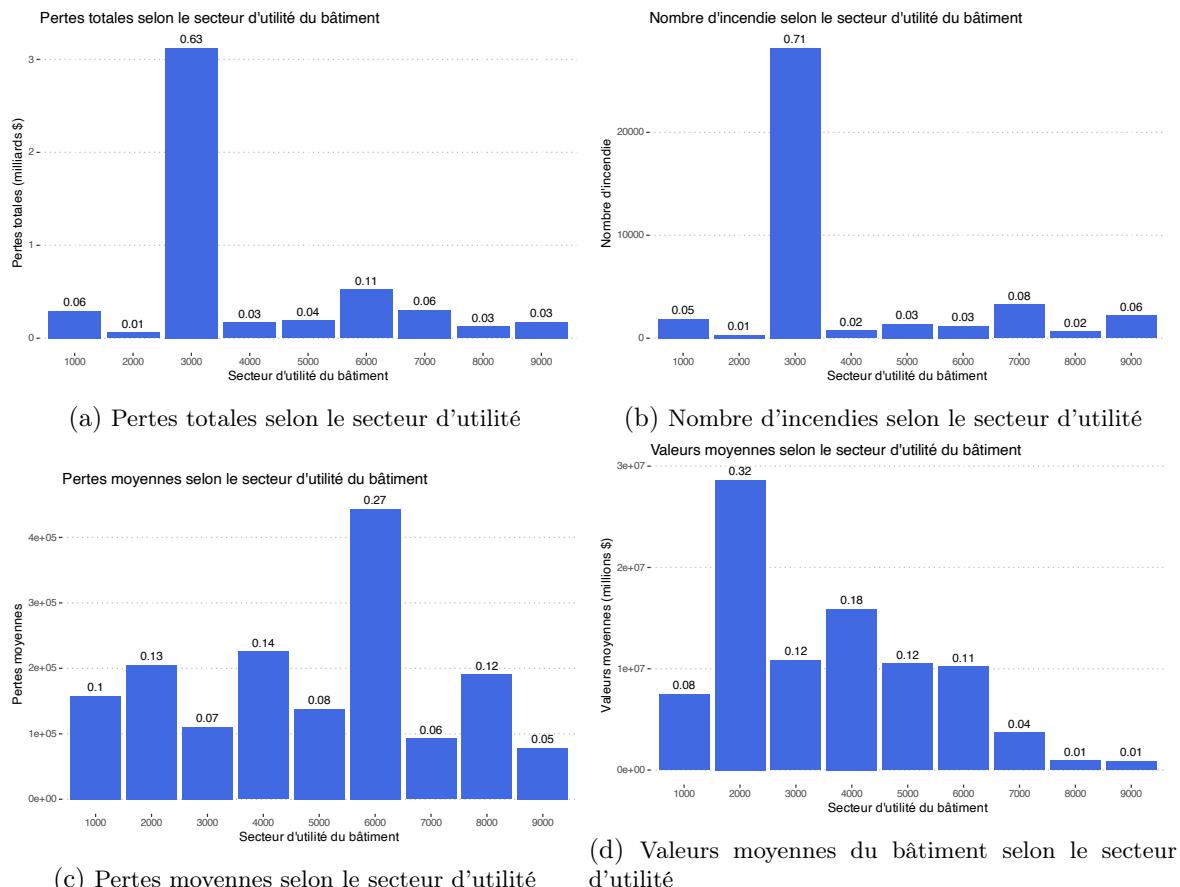


ILLUSTRATION 2.9 – Pertes totales, pertes moyennes, valeurs moyennes et nombre d'incendies selon le secteur d'utilité du bâtiment

Perte	Valeur	Construction	Secteur	Utilité
100	100	Bois massif	Industrielle	Production matériel en bois
60	60	Combustible	Diverse	Bâtiments en construction
59.6	59.6	Combustible protégé	Services	Bureau médical, de santé, de recherche ou scientifique
50	100	Combustible protégé	Industrielle	Production matériel en Bois
50		Bois massif	Industrielle	Production matériel en Bois
46	250	Non combustible	Industrielle	Raffinerie de pétrole
45	150	Non combustible protégé	Industrielle	Raffinerie de pétrole
40	400	Non combustible protégé	Résidentiel	Hôtel à l'année longue
35.2	38.7	Combustible protégé	Résidentiel	Logement utilisable toute l'année
35	35	Combustible protégé	Institutionnelle	Maison de retraite

TABLEAU 2.11 – Les 10 sinistres avec les montants de pertes les plus élevés en millions et leurs caractéristiques

On regarde également les plus grosses pertes pour les structures en bois, voir Tableau 2.12.

Les pertes en bleu dans le Tableau 2.12 représentent les plus gros sinistres pour les constructions en bois parmi les 10 plus importants sinistres toutes constructions confondues. On constate que 7 sinistres se trouvent dans les deux tableaux. Dans l'Annexe C, on présente

Perte	Valeur	Construction	Secteur	Utilité
100	100	Bois massif	Industrielle	Production matériel en bois
60	60	Combustible	Diverses	Bâtiments en construction
59.6	59.6	Combustible protégé	Services	Bureau médical, de santé, de recherche ou scientifique
50	100	Combustible protégé	Industrielle	Production matériel en bois
50		Bois massif	Industrielle	Production matériel en Bois
35.2	38.7	Combustible protégé	Résidentiel	Logement utilisable toute l'année
35	35	Combustible protégé	Institutionnelle	Maison de retraite
23.5	40	Combustible protégé	Résidentiel	Appartement, 5-10 unités
19	25	Combustible	Résidentiel	Appartement, 20 unité et plus
16.3	16	Combustible	Résidentiel	Appartement, 20 unité et plus

TABLEAU 2.12 – Les 10 sinistres avec les montants de pertes les plus élevés, en millions, pour les constructions en bois et leurs caractéristiques

les dix plus gros sinistres pour chacun des types de construction.

Dans le cadre de ce projet, on voudrait éventuellement travailler principalement sur le péril incendie pour les chantiers de construction. Dans la base de données DataNFID, il est possible d'aller chercher les incendies qui ont eu lieu sur un bâtiment en construction à l'aide de la variable PROPCLAS. Malheureusement, la base de données possède seulement 182 données pour des bâtiments en construction.

On regarde également les statistiques descriptives pour les incendies sur les bâtiments en construction. On présente les statistiques au Tableau 2.13.

Min.	1 ^{er} Qu.	Médiane	Moyenne	Variance	3 ^{er} Qu.	Max.
2	1625	22 625	528 473	2.02×10^{13}	118 418	60 000 000

TABLEAU 2.13 – Statistiques descriptives des $n = 182$ pertes sur des bâtiments en construction

Au Tableau 2.14, on présente le nombre d'incendies sur des bâtiments en construction selon le type de construction.

	1	2	3	4	5
	109	51	2	9	11

TABLEAU 2.14 – Nombre de sinistres sur des bâtiments en construction selon le type de construction

Ensuite, à l'Illustration 2.10, on s'intéresse à la perte moyenne.

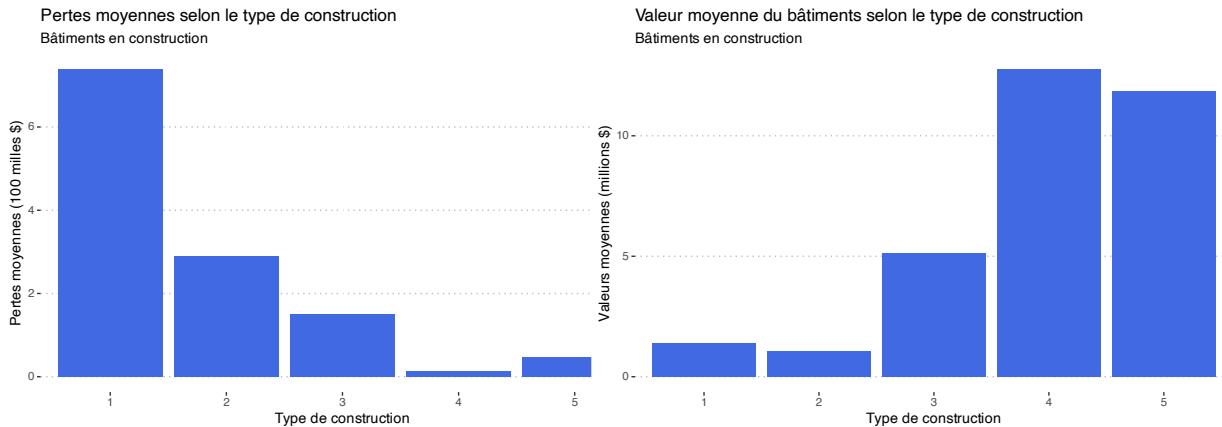


ILLUSTRATION 2.10 – Pertes et valeurs moyennes pour les bâtiments en construction selon le type de construction

On observe que la perte moyenne est plus élevée pour les catégories qui ont eu le plus de sinistres, soit les constructions en bois non protégé et celles en bois protégé. On constate quand même que la perte moyenne est plus élevée pour les constructions en bois massif que celles en acier, malgré un nombre de sinistres moins élevé. Il est également intéressant de constater que la valeur moyenne des bâtiments en acier est plus élevée que celle en bois, mais leurs pertes moyennes sont moins élevées que celles des bâtiments en bois.

Chapitre 3

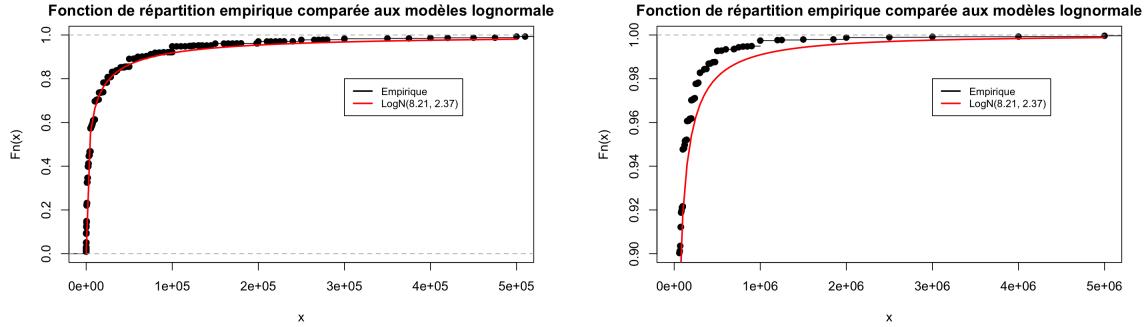
Modélisation des pertes pour le péril incendie (sévérité)

3.1 Introduction

Dans cette section, l'objectif est de modéliser les montants de pertes en sinistres d'incendie. Pour ce faire, on utilise les deux bases de données présenté au chapitre 2 : **DataToronto** et **DataNFID**. On commence par utiliser les résultats de la Section 2.2.2 et l'on teste la loi lognormale. Ensuite, on emploie la méthode *Peak-over-Threshold* avec la méthode de raccordement de lois pour tester des lois plus complexes et potentiellement meilleures pour la queue de la distribution. Les lois par raccordement testées sont :

- Lognormal-Pareto généralisée (LN-PaG) ;
- Weibull-Pareto généralisée (We-PaG) ;
- Coxienne-2-Pareto généralisée (Cox2-PaG) ;
- Bêta généralisée de type 2-Pareto généralisée (GB2-PaG).

Ensuite, on utilise un test graphique et des tests statistiques pour déterminer quelles lois offrent la meilleure adéquation aux données de la base de données **DataToronto**. Une fois les lois les plus appropriées sélectionnées, on présente leurs caractéristiques, telles que l'espérance et les mesures de risque *VaR* et *TVaR*. Sans refaire toute l'analyse en détail, car elle est similaire, on présente également les caractéristiques des modèles choisis pour la base de données **DataNFID**. On terminera avec la modélisation par années des montants de pertes pour les bases **DataToronto** et **DataNFID**.



(a) Valeurs de $F_n(x)$ compris entre 0 et 1

(b) Valeurs de $F_n(x)$ compris entre 0.9 et 1

ILLUSTRATION 3.1 – Fonction de répartition empirique et fonction de répartition de la loi lognormale

3.2 Modélisation préliminaire des montants de perte avec une loi lognormale

Il a été constaté dans la Section 2.2.2 que la loi lognormale serait un choix judicieux pour modéliser les montants de pertes de la base de données DataToronto. On utilise la méthode du maximum de vraisemblance (MV) pour estimer les paramètres de la loi. La méthode des moments (MOM) est utilisée obtenir les valeurs de départ des paramètres. La fonction `constrOptim` de **R** est utilisée pour l'optimisation. Les paramètres obtenus sont présentés au Tableau 3.1.

$\hat{\mu}$	$\hat{\sigma}$
8.21	2.37

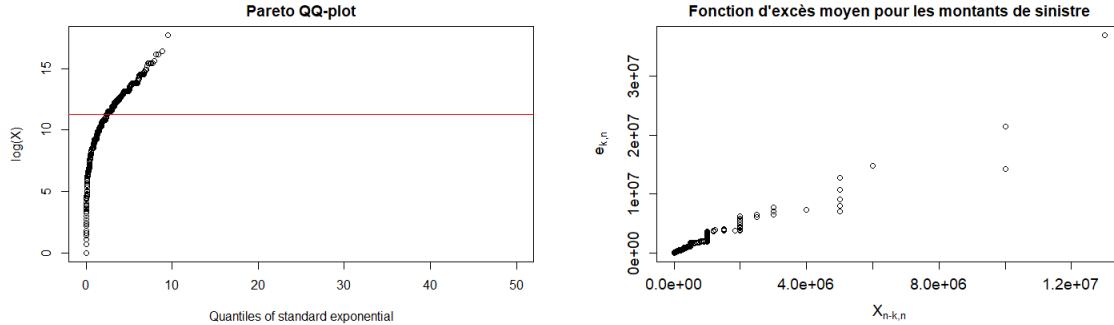
TABLEAU 3.1 – Paramètres estimés pour la loi lognormale

À l'Illustration 3.1, on remarque que la loi lognormale modélise adéquatement les montants de sinistres moins élevés. Cependant, elle surévalue les montants de sinistres plus élevés, indiquant qu'elle ne modélise pas bien simultanément les valeurs faibles et les valeurs extrêmes. L'intuition mentionnée dans la Section 2.2.2, selon laquelle la queue de la distribution pourrait être de loi Pareto, mérite d'être investiguée.

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

Il a été mentionné dans la Section 2.2 qu'il y avait des pertes extrêmes avec une probabilités non-négligeable, ce qui suggère que la distribution des montants de sinistres pourrait comporter une portion lognormale et une portion Pareto. Pour cette raison, pour modéliser la queue de la distribution, on utilise la méthode POT, présentée dans la Section 1.6.

Le défi de cette méthode est de d'identifier un seuil, θ , au-dessus duquel la distribution des



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

(b) Fonction d'excès moyen

ILLUSTRATION 3.2 – *QQplot* Pareto et fonction d'excès moyen

montants de sinistres correspond à une loi Pareto généralisée. Cependant, il faut faire attention, car si l'on sélectionne un seuil trop bas, les résultats risquent d'être biaisés puisque certaines données ne seront pas distribuées selon une loi Pareto généralisée. Par contre, si l'on prend un seuil trop élevé, on risque d'avoir un nombre insuffisant de données, ce qui pourrait également produire des résultats biaisés.

Comme mentionné à la Section 1.5, le paramètre α de la loi Pareto généralisée est très important. Dans la littérature actuarielle, on observe fréquemment que $\alpha \approx 1.5$ pour les pertes dues aux incendies en assurance commerciale, voir Antal and Re (2007) à la page 11.

3.3.1 Déterminer le seuil

Pour déterminer le seuil, une méthode est 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 3.2b, 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 2.5c, que la courbe se transforme tranquillement en droite. Il demeure difficile d'identifier le point précis où le changement se fait. L'Illustration 3.2a est le même *QQplot* qu'en 2.5c, 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 θ 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 $\hat{\xi}_{\theta_i}$, $\hat{\sigma}_{\theta_i}$ pour ensuite définir

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

On peut montrer que,

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

et ainsi conclure que $\tau_{\theta_i} - \tau_{\theta_{i-1}}$ est approximativement normal. On a l'Algorithme 1 suivant :

Algorithme 1 Normalité des écarts

```

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

Donc, l'algorithme 1 vérifie la normalité des écarts. Si le test est rejeté, on retire le seuil le plus bas et on refait le test jusqu'à ce qu'il ne soit plus rejeté. 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 θ 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 5. Soit X , la v.a. modélisant les pertes et F_n , la fonction de répartition empirique. Alors, la statistique KS est

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

On utilise un vecteur de 500 valeurs de θ allant de 70 milles à 1 million. On trouve une valeur de θ optimale de 116 233. Il est important de noter que lorsque les valeurs dans le vecteur ou le nombre de valeurs sont modifiés, le seuil trouvé change également, mais reste généralement dans des valeurs similaires, tant que la borne inférieure n'est pas trop basse.

Intervalle de θ testés ¹	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

TABLEAU 3.2 – Seuil trouvé avec la méthode PLFIT selon différents vecteurs de u

Quelques résultats sont présentés au Tableau 3.2. 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 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 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 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 θ 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.

3.3.2 Estimation des paramètres

Pour estimer les paramètres pour la Pareto généralisée, on utilise la fonction `gpdFit` du package `tea` de 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 3.3.

$\hat{\alpha}$	$\hat{\lambda}$	θ
1.41	117 038	75 000

(a) Seuil à 75 000

$\hat{\alpha}$	$\hat{\lambda}$	θ
1.94	387 881	150 000

(b) Seuil à 150 000

TABLEAU 3.3 – Paramètres de la Pareto généralisée

1. Les intervalles sont constitués de 500 seuils différents.

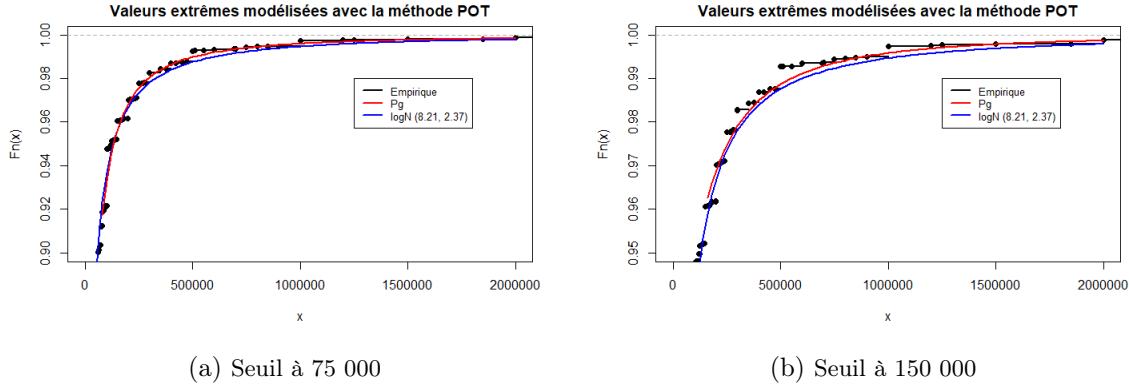


ILLUSTRATION 3.3 – Comparaison fonction de répartition de la loi Pareto généralisée, de la fonction de répartition de la loi lognormale avec la fonction de répartition empirique

On compare les fonctions de répartition des deux lois Pareto généralisée avec la fonction de répartition empirique. On observe à l'Illustration 3.3 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 $\alpha = 1.94$ pour le seuil à 150 000. Pour ces raisons, il est préférable de continuer avec le seuil à 75 000. Au Tableau 3.4, on présente des informations relatives au seuil choisi.

$F_n(u)$	Nb. obs. $> \theta$	$\sum_{i=1}^n x_i \times 1_{\{X > \theta\}} / \sum_{i=1}^n x_i$
0.9122	1192	77 %

TABLEAU 3.4 – Fonction de répartition empirique au seuil, nombre d'observations supérieures au seuil, pourcentage des montants de sinistre supérieurs au seuil au montant total de sinistres

On voit au Tableau 3.4 que plus de 90 % des données sont inférieures au seuil. Il est essentiel de vérifier combien de données restent au-dessus du seuil, et dans ce cas, il reste 1192 données, ce qui est raisonnable pour mener une analyse. On constate également que le montant total des sinistres pour les 1192 données supérieures à θ représente 77 % du montant total des sinistres.

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 3.5, on présente les statistiques descriptives pour ces deux cas.

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

TABLEAU 3.5 – 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 obtient les paramètres estimés présentés au Tableau 3.6.

Type	$\hat{\alpha}$	$\hat{\lambda}$
Const.	1.49	387 183
Const. et rénov.	1.45	226 104

TABLEAU 3.6 – Paramètres de la Pareto généralisée pour les bâtiments en construction et en construction et rénovation

On constate que les valeurs de α obtenues correspondent une fois de plus étroitement à ce que l'on s'attendrait selon la littérature actuarielle. Cependant, le nombre de données pour faire cette analyse n'est pas élevé.

À des fins de validation, cette approche a été testée sur d'autres bases de données de sinistres d'incendie afin de vérifier si l'on trouve également des valeurs de α près de 1.5. Les résultats sont présentés dans l'Annexe B.

3.3.3 Applications avec les mesures VaR et TVaR

On veut calculer les mesures de risque VaR et $TVaR$ en utilisant les données pour la portion sous le seuil et la loi Pareto généralisée pour la portion au-dessus du seuil. On a la fonction de répartition suivante :

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

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

$$\begin{aligned} VaR_\kappa(X) &= \frac{\sigma}{\xi} \left(\left(\frac{1 - \kappa}{\bar{F}_n(\theta)} \right)^{-\xi} - 1 \right) + \theta, \quad \kappa \geq F_n(\theta); \\ TVaR_\kappa(X) &= \frac{\sigma}{\xi} \left(\frac{1}{1 - \kappa} \left(\frac{1 - \kappa}{\bar{F}_n(\theta)} \right)^{-\xi} - 1 \right) + \theta, \quad \kappa \geq F_n(\theta). \end{aligned}$$

Au Tableau 3.7, 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 3.7 – Valeurs pour les mesures VaR et $TVaR$

Comme mentionné à la Section 1.7, il est important de calculer les deux mesures de risque pour avoir le plus d'information possible. On voit bien ici qu'il y a une différence considérable entre les deux mesures.

3.4 Nouvelles lois construites par raccordement de lois

Comme constaté dans la Section 3.3, la loi de Pareto généralisée offre une bonne adéquation pour les valeurs supérieures à 75 000. Cependant, il n'est pas souhaitable de conserver la portion empirique pour les pertes inférieures à ce seuil, comme on veut avoir un modèle complet pour les pertes en sinistre incendie. Par conséquent, on utilise la méthode par raccordement de lois afin d'obtenir une loi pour la portion sous le seuil et une autre pour la portion au-dessus du seuil. Cette méthode a été utilisée dans divers travaux, notamment par Brazauskas and Kleefeld (2016), qui visaient à modéliser les pertes de la base de données norvégienne d'incendie. De plus, Reynkens et al. (2017) propose d'utiliser un mélange d'Erlang pour la portion sous le seuil et la loi de Pareto généralisée pour la portion au-dessus du seuil. Par ailleurs, Majid and Ibrahim (2021) suggèrent d'employer la loi bêta généralisée de type 2 pour la portion sous le seuil et la loi de Pareto de type I pour la portion au-dessus du seuil. Enfin, Grün and Miljkovic (2019) testent 256 modèles par raccordement sur les données de pertes d'incendie danoises (*Danish fire loss data*).

3.4.1 Définitions

On commence par définir les lois par raccordement.

Définition 6. Soit X une v.a. qui suit une loi composite et soient $\{X_j, j = 1, \dots, n\}$ des variables non identiquement distribuées. Alors, la fonction de densité de X est

$$f_X(x) = \begin{cases} w_1 \frac{f_{X_1}(x)}{F_{X_1}(\theta_1) - F_{X_1}(\theta_0)} & , \theta_0 < x \leq \theta_1 \\ \dots \\ w_n \frac{f_{X_n}(x)}{F_{X_n}(\theta_n) - F_{X_n}(\theta_{n-1})} & , \theta_{n-1} < x \leq \theta_n \end{cases}$$

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 θ_i . Cela a aussi comme effet de réduire le nombre de paramètres à optimiser. Par exemple, pour une combinaison de deux lois, on obtient

$$\begin{cases} w \frac{f_{X_1}(\theta)}{F_{X_1}(\theta)} = (1-w) \frac{f_{X_2}(\theta)}{F_{X_2}(\theta)} & (\text{continuité}) \\ w \frac{f'_{X_1}(\theta)}{F_{X_1}(\theta)} = (1-w) \frac{f'_{X_2}(\theta)}{F_{X_2}(\theta)} & (\text{dérivabilité}). \end{cases} \quad (3.1)$$

Pour maximiser la fonction de vraisemblance avec un outil numérique, il faut obtenir la fonction de log-vraisemblance, qui est dans le cas général tronquée à gauche :

$$\begin{aligned}\ln \mathcal{L}(\theta, \dots | x_1, \dots, x_n) &= \sum_{i=1}^n \ln f_1(x_i) \times \mathbb{1}\{x_i \leq \theta\} + \sum_{i=1}^n \ln f_2(x_i) \times \mathbb{1}\{x_i > \theta\} \\ &\quad + \ln\left(\frac{w}{F_1(\theta)}\right) \sum_{i=1}^n \mathbb{1}\{x_i \leq \theta\} + \ln(1-w) \sum_{i=1}^n \mathbb{1}\{x_i > \theta\} \\ &\quad - n \ln\left(1 - w \frac{F_1(x_i)}{F_1(\theta)}\right) \times \mathbb{1}\{d \leq \theta\} - n \ln((1-w)(1-F_2(d))) \times \mathbb{1}\{d > \theta\}.\end{aligned}$$

Démonstration.

$$\begin{aligned}\ln \mathcal{L}(\theta, \dots | x_1, \dots, x_n) &= \sum_{i=1}^n \ln\left(\frac{f(x_i)}{1-F(d)}\right) \\ &= \sum_{i=1}^n \ln\left(w \frac{f_1(x_i)}{F_1(\theta)} \times \mathbb{1}\{x_i \leq \theta\} + (1-w)f_2(x_i) \times \mathbb{1}\{x_i > \theta\}\right) - n \ln(1-F(x_i)) \\ &= \sum_{i=1}^n \ln\left(w \frac{f_1(x_i)}{F_1(\theta)}\right) \times \mathbb{1}\{x_i \leq \theta\} + \sum_{i=1}^n \ln((1-w)f_2(x_i)) \times \mathbb{1}\{x_i > \theta\} - n \ln(1-F(x_i)) \\ &= \sum_{i=1}^n \ln f_1(x_i) \times \mathbb{1}\{x_i \leq \theta\} + \sum_{i=1}^n \ln f_2(x_i) \times \mathbb{1}\{x_i > \theta\} \\ &\quad + \ln\left(\frac{w}{F_1(\theta)}\right) \sum_{i=1}^n \mathbb{1}\{x_i \leq \theta\} + \ln(1-w) \sum_{i=1}^n \mathbb{1}\{x_i > \theta\} \\ &\quad - n \ln\left(1 - w \frac{F_1(x_i)}{F_1(\theta)} \times \mathbb{1}\{d \leq \theta\} - (w + (1-w)F_2(d)) \times \mathbb{1}\{d > \theta\}\right) \\ &= \sum_{i=1}^n \ln f_1(x_i) \times \mathbb{1}\{x_i \leq \theta\} + \sum_{i=1}^n \ln f_2(x_i) \times \mathbb{1}\{x_i > \theta\} \\ &\quad + \ln\left(\frac{w}{F_1(\theta)}\right) \sum_{i=1}^n \mathbb{1}\{x_i \leq \theta\} + \ln(1-w) \sum_{i=1}^n \mathbb{1}\{x_i > \theta\} \\ &\quad - n \ln\left(1 - w \frac{F_1(x_i)}{F_1(\theta)}\right) \times \mathbb{1}\{d \leq \theta\} - n \ln((1-w)(1-F_2(d))) \times \mathbb{1}\{d > \theta\}.\end{aligned}$$

□

3.4.2 Lognormale-Pareto généralisée

On suppose X_1 de loi lognormale, X_2 de loi Pareto généralisée et l'on teste la loi LN-PaG.

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

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

où

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

et

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

Si on tient compte de la continuité et de la dérivabilité, on résout le système d'équations 3.1 et on trouve :

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

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 est 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 obtient les paramètres estimés suivants :

$\hat{\mu}$	\hat{r}	$\hat{\alpha}$	$\hat{\lambda}$	\hat{w}
8.65	2.69	1.41	117 236	0.91

TABLEAU 3.8 – Paramètres estimés pour la loi composite lognormale-Pareto généralisée

On peut voir à l'Illustration 3.4 que la loi LN-PaG a une bonne adéquation avec la fonction de répartition empirique.

3.4.3 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 loi de Weibull.

Définition 8. La fonction de densité de la loi composite Weibull-Pareto généralisée est

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

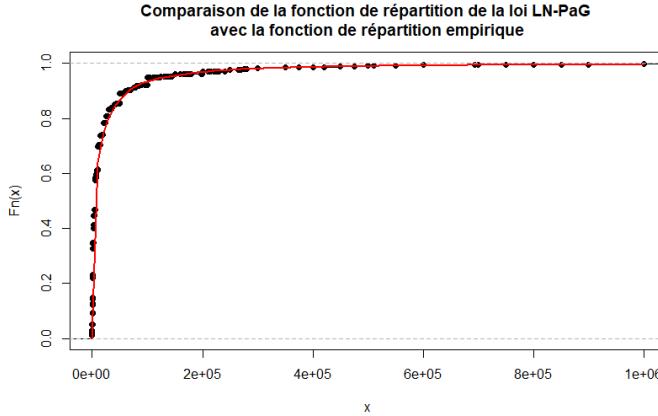


ILLUSTRATION 3.4 – Comparaison de la fonction de répartition de la loi LN-PaG avec la fonction de répartition empirique

où

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

et

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

Il est toujours possible d'imposer la continuité et la dérivabilité au point θ . Dans ce cas, on obtient :

$$\begin{aligned} \beta &= \left[\frac{1}{\tau\theta^\tau} \left(\theta \frac{\xi+1}{\sigma} - \tau + 1 \right) \right]^{-\frac{1}{\tau}} \\ w &= \left(\sigma \frac{f_{X_1}(\theta)}{F_{X_1}(\theta)} + 1 \right)^{-1}. \end{aligned}$$

Avec la loi We-PaG, le cas avec continuité était assez stable, donc il vaut la peine d'être regardé. Encore une 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 :

$\hat{\tau}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\lambda}$	\hat{w}
0.55	7831	1.42	118 801	0.91

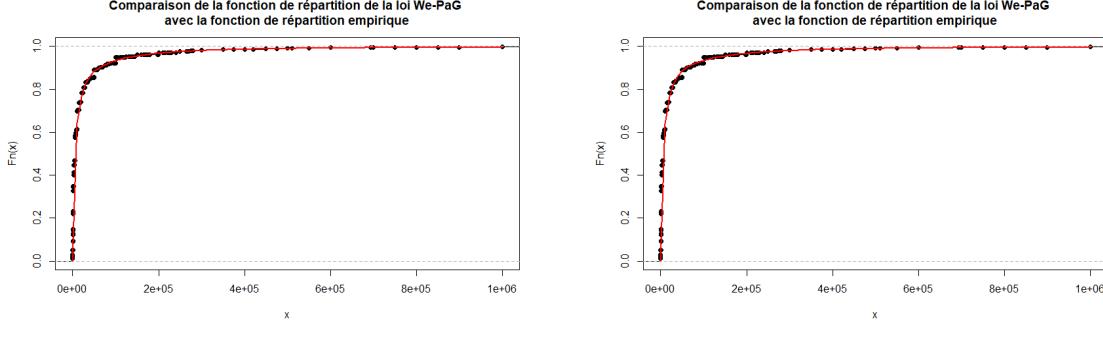
(a) Sans continuité et dérivabilité

$\hat{\tau}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\lambda}$	\hat{w}
0.54	8258	1.61	160 729	0.92

(b) Avec continuité

TABLEAU 3.9 – Paramètres pour la loi composite lognormale-Pareto généralisée

À l'Illustration 3.5, on constate que la loi We-PaG suit bien la fonction de répartition empirique, tant pour les valeurs faibles que pour les valeurs élevées. Il est difficile de dire laquelle des deux



(a) Sans continuité et dérivalibilité

(b) Avec continuité

ILLUSTRATION 3.5 – Comparaison fonction de répartition de la loi We-PaG avec la fonction de répartition empirique

courbes est la meilleure juste en se basant sur ces graphiques. Pour l'instant, les lois We-PaG et We-PaG avec continuité semblent équivalentes.

3.4.4 Coxienne-2-Pareto généralisée

Le mélange d'Erlang a été testé, mais les résultats obtenus étaient nettement moins bons que ceux des autres lois testées. Cependant, il était intéressant de regarder un cas particulier de cette classe, soit la loi coxienne d'ordre deux. Présenté par Klugman et al. (2013), cette distribution a l'avantage d'avoir des formes analytiques simples comparativement au mélange d'Erlang classique, voir Annexe E.1.4. On utilise la coxienne d'ordre deux pour limiter le nombre de paramètres.

Définition 9. La fonction de densité de la loi composite Coxienne-2-Pareto généralisée est

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

où

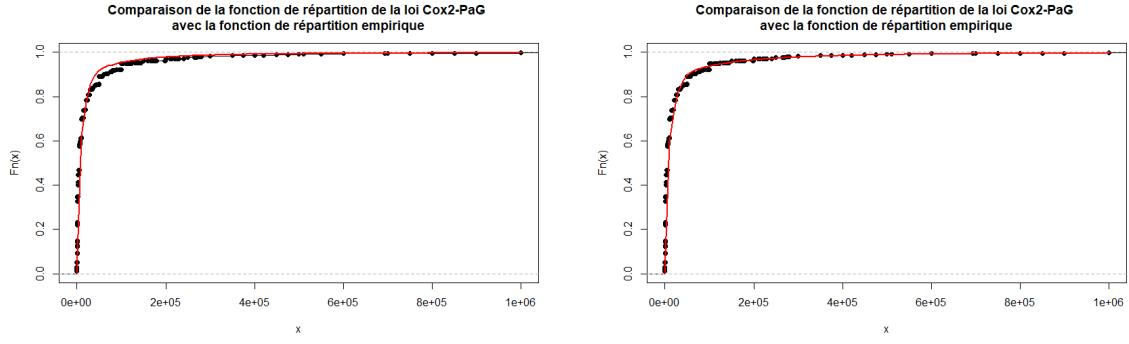
$$f_{X_1}(x) = p\beta_1 e^{-\beta_1 x} + (1-p) \left(\frac{\beta_2}{\beta_2 - \beta_1} \beta_1 e^{-\beta_1 x} + \frac{\beta_1}{\beta_1 - \beta_2} \beta_2 e^{-\beta_2 x} \right)$$

et

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

Il est toujours possible de tenir compte de la continuité. Cependant, pour la dérivalibilité, l'expression est très complexe et difficilement travaillable. L'expression pour assurer la continuité est donné par

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



(a) Sans continuité et dérivabilité

(b) Avec continuité

ILLUSTRATION 3.6 – Comparaison fonction de répartition de la loi We-PaG avec la fonction de répartition empirique

On utilise la méthode du maximum de vraisemblance pour estimer les paramètres présentés au Tableau 3.10.

\hat{p}	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\alpha}$	$\hat{\lambda}$	\hat{w}
0.40	1.32×10^{-3}	6.17×10^{-5}	1.41	118 442	0.94

(a) Sans continuité et dérivabilité

\hat{p}	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\alpha}$	$\hat{\lambda}$	\hat{w}
0.41	1.18×10^{-3}	5.34×10^{-5}	1.98	261 116	0.93

(b) Avec continuité

TABLEAU 3.10 – Paramètres pour la loi composite Coxienne-2-Pareto généralisée

À l’Illustration 3.6, 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 alpha est de 1.98 et qu’on s’attend davantage à avoir un alpha près de 1.5.

3.4.5 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, étant donné ses 4 paramètres.

Définition 10. *La fonction de densité de la loi composite bêta généralisée de type 2-Pareto généralisée est*

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

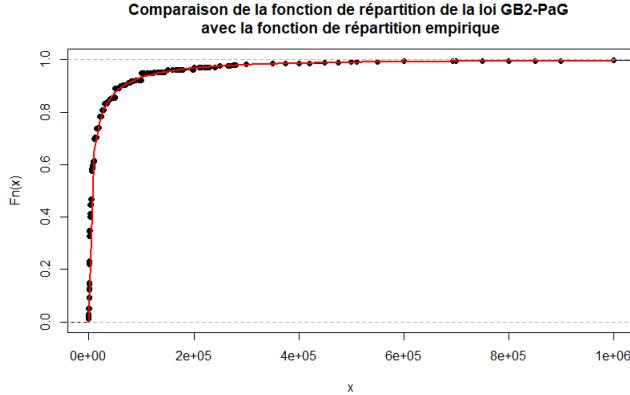


ILLUSTRATION 3.7 – Comparaison fonction de répartition de la loi We-PaG avec la fonction de répartition empirique

où

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

et

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

Il est toujours possible de tenir compte de la continuité. Cependant, pour la dérivation, l'expression est très complexe et difficilement travaillable. L'expression pour assurer la continuité est donné par

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

On utilise la méthode du maximum de vraisemblance pour estimer les paramètres présentés au Tableau 3.11. Pour trouver les paramètres initiaux, on utilise la fonction `m1.gb2` du package GB2 de [R](#), Graf and Nedyalkova. (2022).

\hat{a}	\hat{b}	\hat{p}	\hat{q}	$\hat{\alpha}$	$\hat{\lambda}$	\hat{w}
0.47	19 990	1.57	2.89	1.42	120 149	0.91

TABLEAU 3.11 – Paramètres pour la loi composite bêta généralisée de type 2-Pareto généralisée

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

À l'Illustration 3.7, on constate que l'adéquation est très bonne sur tout le domaine. Il faudra donc utiliser des tests graphiques et statistiques pour définir le modèle le plus adapté à ces données.

3.5 Test des lois

Afin de sélectionner la loi offrant l'ajustement le plus adéquat aux données canadiennes, on utilise un test d'adéquation graphique ainsi que des tests statistiques, tels que Cramer-Von Mises (W^2), Anderson-Darling (A^2), AIC et BIC.

3.5.1 Test graphique

Comme à la Section 2.2, on analyse les *QQplots* des raccordements de lois. Si l'ajustement est bon, les données suivront la ligne rouge.

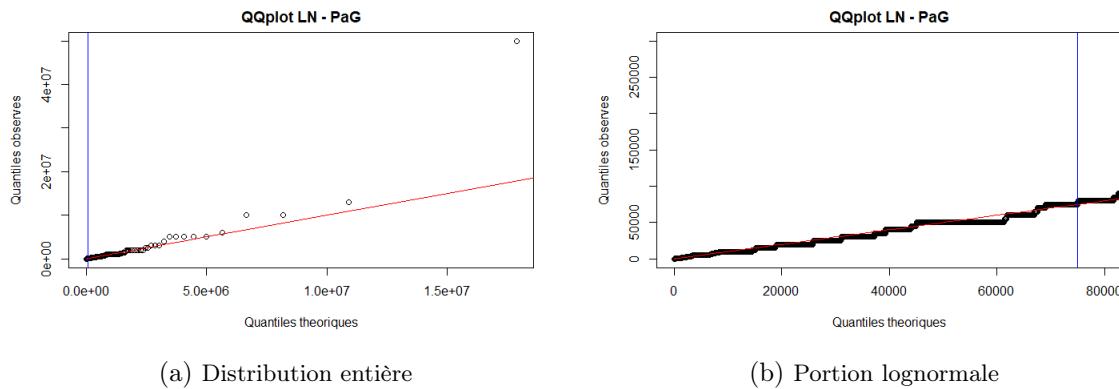


ILLUSTRATION 3.8 – *QQplot* lognormale-Pareto généralisée

On observe à l'Illustration 3.8b que pour les valeurs inférieures à 75 000, ce qui correspond à la ligne verticale bleue, l'adéquation de la loi est bonne. Pour les valeurs supérieures 75 000, on trouve aussi une assez bonne adéquation, même pour les valeurs extrêmes qui restent près de la droite. Il semble y avoir peu de données sous 75 000, mais on rappelle qu'en réalité, plus de 90 % de nos données sont inférieures à 75 000.

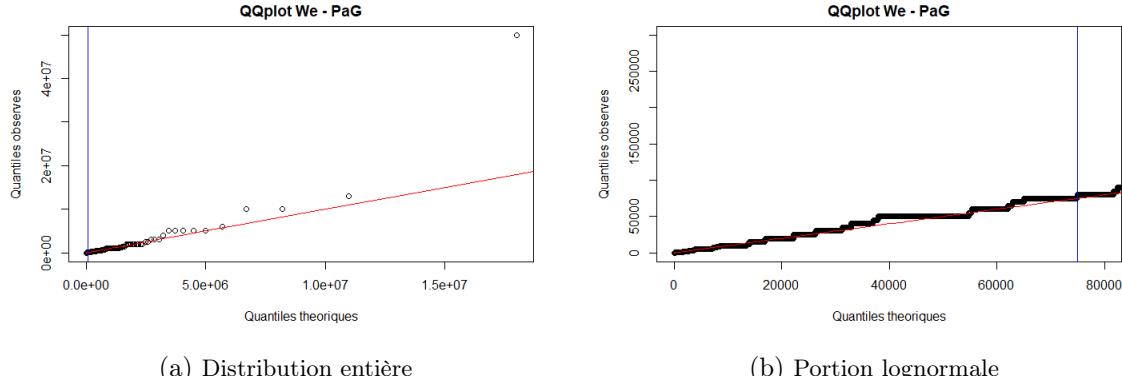


ILLUSTRATION 3.9 – *QQplot* pour la loi Weibull-Pareto généralisée

La loi **We-PaG** présentée à l’Illustration 3.9 offre une bonne adéquation. Il est difficile de dire quelle loi est la plus adéquate entre la loi LN-PaG et la loi **We-PaG**, car leur adéquation semble être très similaire.

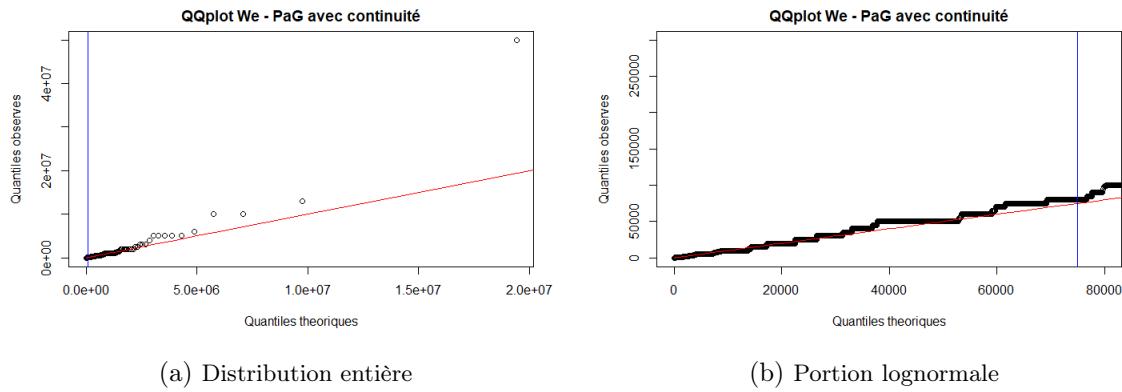


ILLUSTRATION 3.10 – *QQplot* pour la loi Weibull-Pareto généralisée avec continuité

Lorsqu’on tient compte de la continuité avec la loi **We-PaG**, on remarque à l’illustration 3.10 que l’adéquation est légèrement moins bonne que pour les deux autres lois, car on sous-estime plus les valeurs, tant dans la portion modélisée par la loi Weibull que dans la portion modélisée par la loi Pareto généralisée.

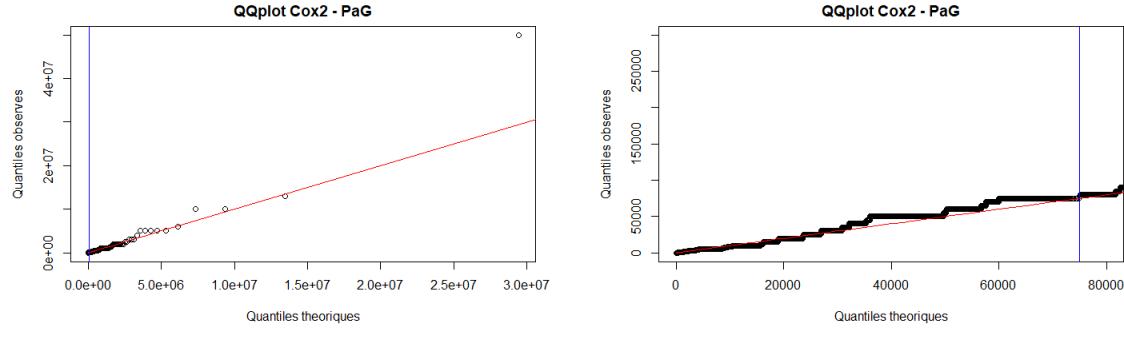


ILLUSTRATION 3.11 – *OOfplot* pour la loi Coxienne-2-Pareto généralisée

On remarque à l'Illustration 3.11 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 près des lois LN-PaG et We-PaG comme les paramètres de la loi Pareto généralisée sont presque identiques.

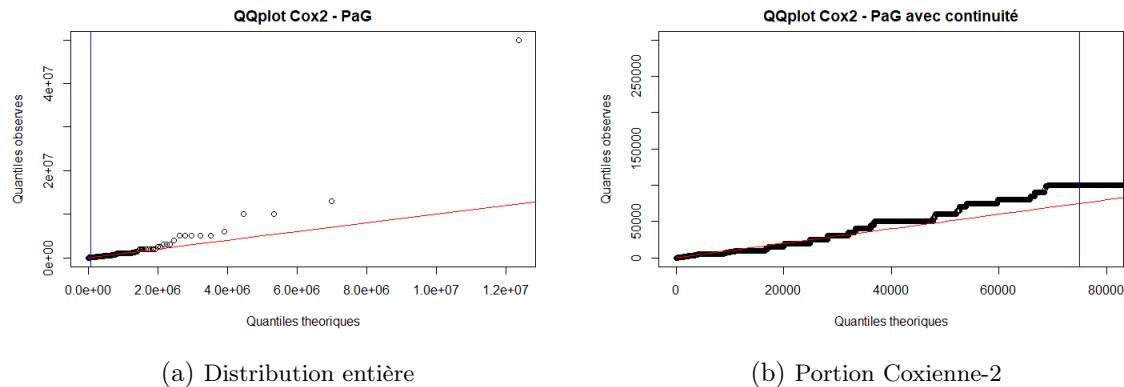


ILLUSTRATION 3.12 – *QQplot* pour la loi Coxienne-2-Pareto généralisée avec continuité

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 alpha de 2.

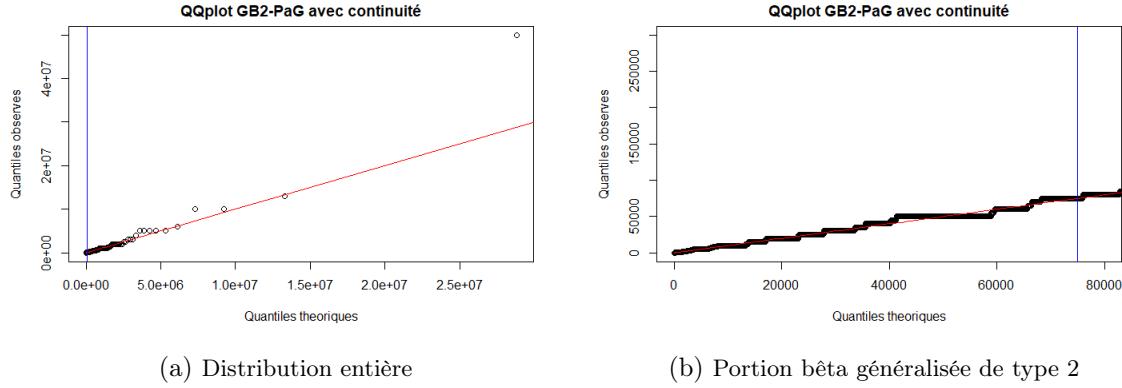


ILLUSTRATION 3.13 – *QQplot* pour la loi GB2-Pareto généralisée avec continuité

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 laquelle est la meilleure. Cependant, on peut déjà observer que les lois **LN-PaG**, **We-PaG** et **GB2-PaG** avec continuité semblent être les meilleures.

3.5.2 Tests statistiques

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 utilisés.

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 3.12 – 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 3.12 que la meilleure loi selon ces tests est la loi **GB2-PaG**. Pour les tests AIC et BIC, on veut aussi le résultat le plus petit. La loi **GB2-PaG** est encore la meilleure selon ces tests.

La loi **LN-PaG** serait également un bon choix, car elle se classe deuxième selon les tests effectués. De plus, le Graphique 2.3b de la log-densité et le *QQplot* 2.5b lognormale donnent une bonne

intuition que la distribution puisse être de loi lognormale. Finalement, comme les deux modèles sont assez équivalents, quelqu'un pourrait préférer le modèle ayant le moins de paramètres.

3.6 Intervalles de confiance

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]
$\hat{\sigma}$	83 219	[74 635, 92 024]

TABLEAU 3.13 – Intervalle de confiance estimé avec 14 000 *bootstrap*

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 dûs au fait que, si les données élevées ne se retrouvent pas dans le sous-échantillon, les résultats changent considérablement.

3.7 Informations sur les lois sélectionnées pour la base de données DataToronto

On utilise les lois sélectionnées dans la Section 3.5, à savoir LN-PaG et GB2-PaG, pour comparer l'espérance ainsi que les mesures *VaR* et *TVaR* aux valeurs empiriques.

Distribution	$E[X]$	Var(X)
Empirique	40 296	238 004 336 993
LN-PaG	41 562	∞
GB2-PaG	41 077	∞

TABLEAU 3.14 – Espérance et variance

Dans le Tableau 3.14, on constate que l'espérance obtenue avec les deux lois est très près de la moyenne empirique. En revanche, la variance n'existe pas pour les deux lois, comme vu aux Sections 3.4.2 et 3.4.5 l'indice de queue α est inférieur à deux. On calcule également les mesures *VaR* et *TVaR*, qu'on compare aux valeurs empiriques.

Distribution	κ				
	0.90	0.95	0.99	0.995	0.999
Empirique	60 000	125 000	500 000	1 000 000	2 715 000
LN-PaG	66 841	132 752	506 608	855 718	2 774 167
GB2-PaG	65 445	134 862	513 663	865 066	2 780 455

(a) Mesure de risque VaR

Distribution	κ				
	0.90	0.95	0.99	0.995	0.999
Empirique	321 120	562 888	2 227 303	4 720 000	9 071 428
LN-PaG	327 028	561 683	1 851 936	3 056 788	9 677 751
GB2-PaG	328 020	562 678	1 841 745	3 028 301	9 495 836

(b) Mesure de risque $TVaR$

TABLEAU 3.15 – Valeurs de $VaR_\kappa(X)$ et $TVaR_\kappa(X)$

Pour la mesure VaR , voir Tableau 3.15a, les valeurs obtenues avec les deux lois sont similaires aux valeurs obtenues avec l'échantillon sauf pour le percentile 0.995. Comme le montre le Tableau 3.15b, les deux lois sous-estiment considérablement la mesure $TVaR$ pour les percentiles 0.99 et 0.995, tandis qu'elles surestiment le percentile 0.999. On constate également qu'avec les lois on obtient des valeurs qui sont similaires. Les formes analytiques des caractéristiques sont présentées à l'Annexe E.2.1 et E.2.3.

3.8 Résumé de l'analyse avec la base de données DataNFID

Comme mentionné à l'introduction de cette section, on présentera seulement les résultats et les conclusions de l'analyse de la base de données DataNFID. Pour cette base de données, on a testé seulement les lois de raccordement suivantes :

- Lognormal-Pareto généralisée (LN-PaG) ;
- Weibull-Pareto généralisée (We-PaG) ;
- Bêta généralisée de type 2-Pareto généralisée (GB2-PaG).

Pour sélectionner les modèles les plus appropriés, on sait également baser sur les tests statistiques de Cramer-Von-Mises (W^2) et Anderson-Darling (A^2), ainsi que les critères d'information AIC et BIC. Dans le Tableau 3.16 on présente les résultats de chacun des tests.

Distribution	W^2	A^2	AIC	BIC
LN-GPD	13	83	-2 237 688	-2 237 640
We-GPD	68	428	-2 242 921	-2 242 874
GB2-GPD	12	74	-2 237 440	-2 237 374

TABLEAU 3.16 – Résultats des tests W^2 , A^2 , AIC et BIC

Sur la base de ces tests, il apparaît que la loi GB2-PaG est la meilleure pour la base DataBFID. Toutefois, il convient de noter que la loi LN-PaG présente une performance relativement similaire selon les résultats de ces tests. De plus, cette loi n'a que 5 paramètres contrairement à la loi GB2-PaG qui en a 7. Pour ces raisons les lois LN-PaG et GB2-PaG sont sélectionnées.

Pour le choix du seuil, les mêmes approchent qu'à la Section 3.3.1 on était utilisés. On choisit pour cette base de données un seuil à 350 000. Au Tableau 3.17, on présente des informations relatives au seuil choisi.

$F_n(u)$	Nb. obs. $> \theta$	$\sum_{i=1}^n x_i \times 1_{\{X>\theta\}} / \sum_{i=1}^n x_i$
0.9621	3889	64 %

TABLEAU 3.17 – Fonction de répartition empirique évaluée au seuil, nombre d'observations supérieures au seuil, pourcentage des montants de sinistre supérieurs au seuil au montant total de sinistres

On voit au Tableau 3.17 que plus de 96 % des données sont inférieures au seuil. Il est essentiel de vérifier combien de données restent au-dessus du seuil, et dans ce cas, il reste 3889 données, ce qui est raisonnable pour mener une analyse. On constate également que le montant total des sinistres pour les 3889 données supérieures à θ représente 64 % du montant total des sinistres.

Les paramètres estimés par la méthode du maximum de vraisemblance sont présentés au Tableau 3.18.

$\hat{\mu}$	\hat{r}	$\hat{\alpha}$	$\hat{\lambda}$	\hat{w}	\hat{a}	\hat{b}	\hat{p}	\hat{q}	$\hat{\alpha}$	$\hat{\lambda}$	\hat{w}
8.74	2.66	1.36	332 209	0.96	0.24	25 244	4.76	6.57	1.40	343 089	0.96

(a) Loi LN-GPD

(b) Loi GB2-GPD

TABLEAU 3.18 – Paramètres estimés pour les lois LN-PaG et GB2-PaG

On s'intéresse principalement à l'indice de queue α de la portion Pareto généralisée, on remarque que celui pour la loi LN-PaG est légèrement inférieur à celui pour la loi GB2-PaG. Par contre, dans les deux cas, l'indice de queue est inférieur à deux ; donc la variance n'existe pas pour les deux lois.

Distribution	$E[X]$
Empirique	75 800
LN-GPD	76 274
GB2-GPD	73 187

TABLEAU 3.19 – Espérance et variance

Dans le Tableau 3.19, on constate que l'espérance obtenue avec les deux lois est relativement près de la moyenne empirique. On calcule également les mesures VaR et $TVaR$, qu'on compare aux valeurs empiriques.

Distribution	κ				
	0.90	0.95	0.99	0.995	0.999
Empirique	130 000	290 000	855 895	1 500 000	5 000 000
LN-GPD	132 343	277 560	903 834	1 494 471	4 852 422
GB2-GPD	125 968	273 232	901 700	1 476 447	4 657 006

(a) Mesure de risque VaR

Distribution	κ				
	0.90	0.95	0.99	0.995	0.999
Empirique	619 893	1 042 588	3 378 455	5 881 809	18 180 293
LN-GPD	618 499	1 045 204	3 385 605	5 630 593	18 394 037
GB2-GPD	592 403	999 682	3 154 707	5 176 620	16 365 563

(b) Mesure de risque $TVaR$

TABLEAU 3.20 – Valeurs de $VaR_\kappa(X)$ et $TVaR_\kappa(X)$

Pour la mesure VaR , voir Tableau 3.20a, les valeurs obtenues avec les deux lois sont similaires à celles obtenues avec l'échantillon. Cependant, on remarque que les valeurs obtenues avec la loi GB2-PaG sont légèrement moins proches des valeurs empiriques que celles obtenues avec la loi LN-PaG.

Comme le montre le Tableau 3.20b, la loi GB2-PaG sous-estime considérablement les percentiles 0.995 et 0.999, tandis que la loi LN-PaG donne des valeurs assez proches des valeurs empiriques. Cela est probablement dû au fait que l'indice de queue pour la loi LN-PaG est inférieur à celle de la loi GB2-PaG. Les formes analytiques des caractéristiques sont présentées à l'Annexe E.2.1 et E.2.3.

3.9 Modélisation par année des pertes en sinistres incendie avec la base de données DataToronto

Dans cette section, on souhaite faire une analyse similaire, mais en modélisant les pertes pour les années 2011 à 2018. Cette analyse permettra de déterminer si le modèle est généralement applicable ou si, pour certaines années, il semble inapproprié.

3.9.1 Modélisation de la sévérité

Pour modéliser les pertes par années de la base de données DataToronto, on applique la même méthode qu'à la Section 3.3. Les lois qui seront testées sont les suivantes : LN-PaG, We-PaG et GB2-PaG. Encore une fois, pour estimer les paramètres, on utilisera une méthode numérique pour maximiser la fonction de log-vraisemblance. Pour obtenir le seuil pour chacune des années, on ajoutera le seuil comme paramètre dans la fonction de vraisemblance, comme proposé dans [Brazauskas and Kleefeld \(2016\)](#). Comme valeur de départ pour le seuil, on utilise le seuil obtenu à la Section 3.3.1, qui est de 75 000.

	2011	2012	2013	2014	2015	2016	2017	2018
LN-PaG								
μ	8.26	8.44	8.24	8.56	8.68	8.61	8.85	9.04
r	2.63	2.76	2.48	2.69	2.72	2.57	2.69	2.80
α	1.61	2.14	1.58	1.35	2.31	1.61	1.10	1.31
λ	133 909	178 184	143 275	122 003	230 095	161 367	99 800	116 401
w	0.93	0.92	0.91	0.92	0.93	0.91	0.92	0.90
θ	75 000	75 000	79 999	80 068	80 000	80 000	80 004	83 319
We-PaG								
τ	0.54	0.54	0.57	0.55	0.57	0.58	0.57	0.58
β	6122	6559	6167	7091	7218	7387	8146	8990
α	1.57	1.73	1.53	1.29	1.95	1.42	1.10	1.31
λ	133 116	144 386	129 907	109 428	163 092	119 967	92 171	108 796
w	0.93	0.93	0.90	0.91	0.91	0.90	0.91	0.88
θ	80 000	100 000	75 000	80 000	75 000	75 000	75 000	75 000
GB2-PaG								
a	0.50	0.49	0.49	0.47	0.58	0.54	0.45	0.77
b	12 947	23 551	11 936	19 350	21 116	16 705	16 436	22103
p	1.46	1.40	1.65	1.61	1.15	1.37	1.77	0.77
q	2.64	3.17	2.78	3.08	2.52	2.61	2.70	1.82
α	1.55	1.74	1.37	1.28	1.86	1.20	1.08	1.04
λ	128 536	144 342	123 289	106 466	155 004	103 518	94051	86866
w	0.93	0.93	0.92	0.91	0.92	0.91	0.92	0.90
θ	79 987	99 996	99 000	79 988	79 987	99 000	80 842	99 999

TABLEAU 3.21 – Paramètres estimés par maximum de vraisemblance des raccordements de loi pour les pertes incendies de la base de données DataToronto

On présente au Tableau 3.21 les paramètres pour les raccordements de lois construits. On observe que peu importe le raccordement de lois, le paramètre θ reste entre 75 000 et 100 000, même s'il est majoritairement 75 000. Pour le paramètre α , qui influence la queue de la distribution, on remarque qu'il est pratiquement toujours inférieur à 2. Il y a seulement pour le modèle LN-PaG pour les années 2011 et 2015 que le paramètre α est supérieur à 2. Donc, dans la grande majorité des modèles, la variance est infinie.

3.9.2 Validation des modèles

Afin de sélectionner la loi offrant l'ajustement le plus adéquat au données de la base de données **DataToronto**, on utilise un test d'adéquation graphique ainsi que des tests statistiques, tels que Kolmogorov-Smirnov (D), Anderson-Darling (A^2), AIC et BIC.

Graphiques quantile-quantile

On commence par valider graphiquement l'ajustement des modèles à l'aide des graphiques quantile-quantile présentés aux Illustrations 3.14, 3.15 et 3.16.

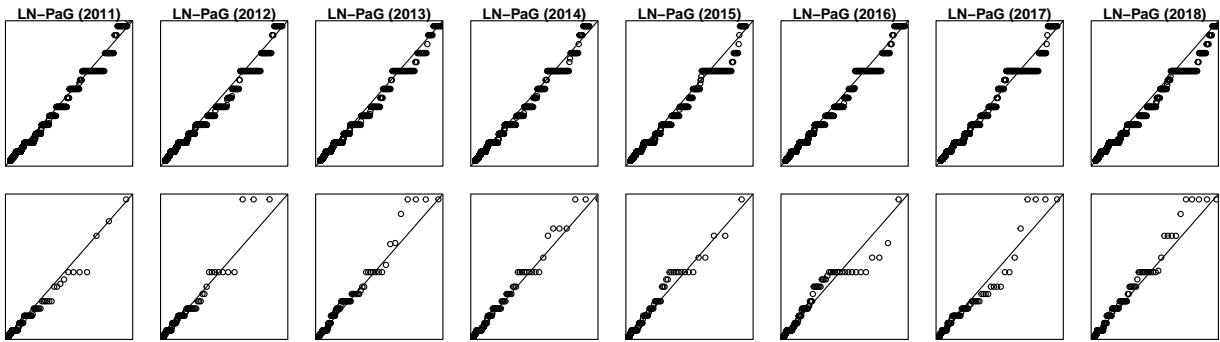


ILLUSTRATION 3.14 – *QQplots* pour les observations sous le seuil (ligne 1) et pour les observations au-dessus du seuil (ligne 2), pour le modèle LN-PaG (2011-2018)

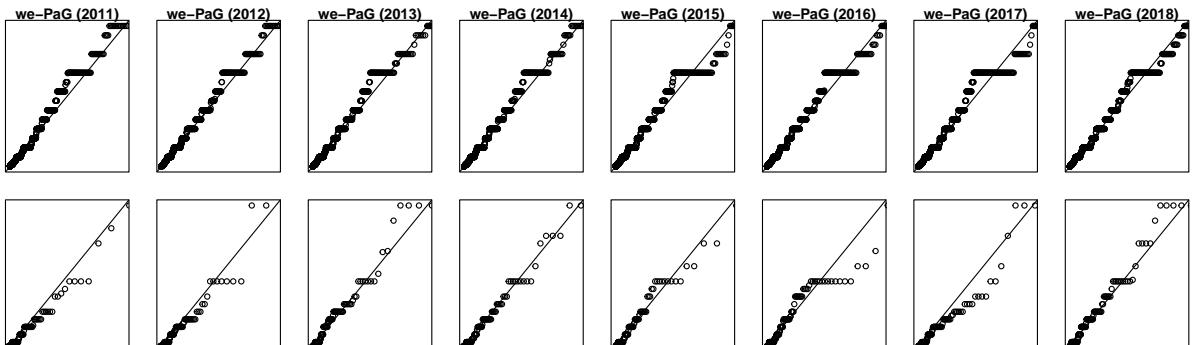


ILLUSTRATION 3.15 – *QQplots* pour les observations sous le seuil (ligne 1) et pour les observations au-dessus du seuil (ligne 2), pour le modèle We-PaG (2011-2018)

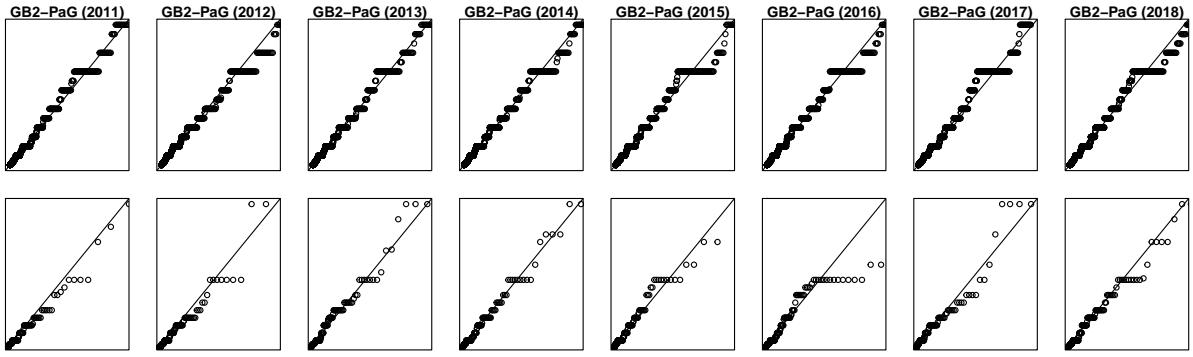


ILLUSTRATION 3.16 – *QQplots* pour les observations sous le seuil (ligne 1) et pour les observations au-dessus du seuil (ligne 2), pour le modèle **GB2-PaG** (2011-2018)

On remarque que l’ajustement est globalement assez bon. Par contre, pour les trois modèles, l’ajustement semble moins bon pour l’année 2017. En effet, on remarque que les points passent davantage sous la droite comparativement aux autres années. Il est donc difficile avec ce test de dire quel est le meilleur modèle parmi ceux-là pour cette base de données.

Tests statistiques

Une autre façon de vérifier l’ajustement des modèles et de regarder des statistiques d’ajustement. On utilise encore les statistiques D et A^2 .

	2011	2012	2013	2014	2015	2016	2017	2018
Statistique D								
LN-PaG	0.06	0.06	0.06	0.06	0.07	0.06	0.06	0.07
We-PaG	0.07	0.07	0.08	0.07	0.06	0.07	0.07	0.06
GB2-PaG	0.06	0.07	0.07	0.06	0.06	0.07	0.06	0.06
Statistique A^2								
LN-PaG	4.55	3.53	3.84	3.53	4.81	4.07	3.90	5.73
We-PaG	5.04	3.74	4.48	4.15	3.79	4.18	4.43	3.52
GB2-PaG	3.13	3.36	3.21	2.94	3.02	3.49	3.16	3.67

TABLEAU 3.22 – Test d’ajustement de modèles pour les données canadiennes de périls incendie

Le Tableau 3.22 présente les valeurs obtenues pour les deux statistiques concernant les trois modèles. On observe que pour la statistique D , le modèle **We-PaG** est légèrement moins bon que les deux autres, qui semblent assez équivalents. Pour la statistique A^2 , le modèle **GB2-PaG** est meilleur que les deux autres. Selon ces tests, le modèle **GB2-PaG** serait celui à utiliser.

Critère d’information

Une autre approche est de comparer la log-vraisemblance et les critères d’information AIC et BIC pour chacun des trois modèles.

	2011	2012	2013	2014	2015	2016	2017	2018
Log-vraisemblance								
LN-PaG	16 069	16 095	16 103	16 905	16 226	16 723	16 140	18 881
We-PaG	16 072	16 087	16 110	16 898	16 215	16 728	16 140	18 860
GB2-PaG	16 050	16 083	16 089	16 894	16 208	16 707	16 130	18 851
Akaike Information Criterion, AIC								
LN-PaG	32 150	32 202	32 217	33 822	32 465	33 457	32 292	37 774
We-PaG	32 155	32 187	32 232	33 807	32 442	33 467	32 291	37 732
GB2-PaG	32 116	32 182	32 194	33 805	32 432	33 429	32 277	37 717
Bayesian Information Criterion, BIC								
LN-PaG	32 183	32 234	32 249	33 855	32 497	33 489	32 324	37 806
We-PaG	32 187	32 219	32 264	33 840	32 475	33 499	32 323	37 765
GB2-PaG	32 159	32 224	32 237	33 848	32 475	33 472	32 320	37 761

TABLEAU 3.23 – Mesures d’information des modèles ajustés pour la base de données DataToronto (2011-2018)

Au Tableau 3.23, on présente les valeurs obtenues pour ces différents critères. On remarque que le modèle **GB2-PaG** est meilleur que les deux autres selon le critère AIC et BIC, alors que les modèles **LN-PaG** et **We-PaG** sont relativement équivalents. En effet, pour certaines années, le modèle **LN-PaG** est meilleur que celui **We-PaG** alors que pour d’autres c’est l’inverse. Selon ce test, le modèle **GB2-PaG** serait le modèle à sélectionner.

Conclusion

Selon les différents tests effectués, autant graphiques que quantitatifs, le modèle **GB2-PaG** serait le meilleur modèle parmi les trois testés. Pour les deux autres, il n'est pas évident de déterminer lequel est le meilleur.

3.9.3 Mesure de la queue de la distribution

Risque au fil du temps

On calcule la mesure *VaR* et la mesure *TVaR*, dont les définitions sont données à la Section 1.7, pour les 3 modèles.

	2011	2012	2013	2014	2015	2016	2017	2018
LN-PaG	104 969	114 281	148 013	135 497	117 150	148 591	124 776	167 301
	373 714	304 698	511 814	639 655	321 718	528 539	1 543 882	815 485
We-PaG	109 768	129 393	150 519	146 140	126 429	154 148	145 589	172 558
	397 209	366 292	538 085	741 309	351 805	624 356	1 775 095	842 771
GB2-PaG	109 057	129 444	152 992	145 159	122 951	158 853	124 164	180 194
	398 128	365 802	633 490	758 282	352 247	977 715	1 880 610	3 927 449

TABLEAU 3.24 – Mesures *VaR* et *TVaR* (2ièmes lignes) au niveau de confiance 95%, pour la base de données DataToronto (2012-2018)

Les résultats pour ces mesures sont présentés au Tableau 3.24. On peut constater que pour cette base de données, les années 2013, 2014, 2016 et 2018 ont les mesures les plus élevées. On pourrait peut-être croire qu'il y a une faible tendance à la hausse au fil des années comme les mesures en 2011 et 2012 sont moins élevées que celle des années suivantes, alors que celles de 2017 et 2018 sont plus élevées que les autres.

Prédiction

Afin de comparer le pouvoir prédictif de nos 3 modèles, on utilise les données de l'année précédente afin d'estimer $\text{Pr}(B > 100\,000)$ pour l'année actuel.

Modèle	2012	2013	2014	2015	2016	2017	2018
Actuel	0.05	0.06	0.05	0.04	0.06	0.05	0.06
LN-PaG	0.05	0.06	0.08	0.07	0.06	0.07	0.06
We-PaG	0.06	0.07	0.08	0.07	0.06	0.08	0.07
GB2-PaG	0.05	0.07	0.08	0.07	0.06	0.09	0.06

TABLEAU 3.25 – Valeurs prédites et actuelles de $\text{Pr}(B > 100\,000)$, pour la base de données DataToronto (2012-2018)

Au Tableau 3.25 on présente le résultat de cette probabilité pour chacun des modèles. On remarque que chaque modèle prédit de façon assez similaire. De plus, tous les modèles ont tendance à prédire une probabilité plus élevée que ce qui est vraiment survenu, principalement pour les années 2014, 2015 et 2017. Pour les autres années, on obtient quelque chose de très près de ce qui est survenu.

3.10 Modélisation par année des pertes en sinistres incendie avec la base de données DataNFID

3.10.1 Modélisation de la sévérité

Pour modéliser les montants de pertes de cette base de données, on utilise la même approche que pour la Section 3.9.1. On notera aussi que le seuil avait déjà été analysé et évalué à 350 000, voir Section 3.8. On utilisera cette valeur comme seuil de départ pour l'optimisation.

	2009	2010	2011	2012	2013	2014	2015
LN-PaG							
μ	8.74	8.72	9.06	8.91	8.93	9.07	9.24
σ	2.57	2.60	2.55	2.51	2.51	2.58	3.43
α	0.87	0.63	0.50	0.89	0.80	0.87	0.92
λ	275 433	244 808	244 807	244 807	245 082	244 808	253 558
w	0.97	0.96	0.93	0.96	0.96	0.97	0.96
θ	355 206	350 021	399 957	399 841	399 975	399 974	377 058
GB2-PaG							
a	0.19	0.21	0.17	0.28	0.26	0.27	0.68
b	18493	18417	23656	24074	25147	48452	38102
p	8.34	6.35	9.93	4.19	4.56	4.06	0.66
q	10.21	7.99	11.79	5.67	6.14	6.30	1.46
α	0.96	0.63	0.50	0.89	0.80	0.87	0.91
λ	273 219	244 850	244 577	244 827	244 835	244 784	253 695
w	0.97	0.96	0.93	0.96	0.96	0.97	0.96
θ	399 978	349 992	400 000	399 848	396 500	399 996	370 000

TABLEAU 3.26 – Paramètres estimés par maximum de vraisemblance des raccordements de loi pour les pertes incendies de la base de données DataNFID

On constate au Tableau 3.26 que tous les α pour les deux modèles sont en dessous de 1, ce qui implique que l'espérance et la variance n'existent pas. De plus, on remarque aussi que les paramètres α et λ sont similaires pour le modèle LN-PaG et le modèle GB2-PaG. On observe aussi que le seuil, pour les années 2010-2014, désormais à 400 000.

3.10.2 Validation des modèles

Afin de sélectionner la loi offrant l'ajustement le plus adéquat au données de la base de données DataNFID, on utilise un test d'adéquation graphique ainsi que des tests statistiques, tels que Kolmogorov-Smirnov (D), Anderson-Darling (A^2), AIC et BIC.

Graphiques quantile-quantile

On commence par valider graphiquement l'ajustement des modèles à l'aide des graphiques quantile-quantile présentés aux Illustrations 3.17 et 3.18.

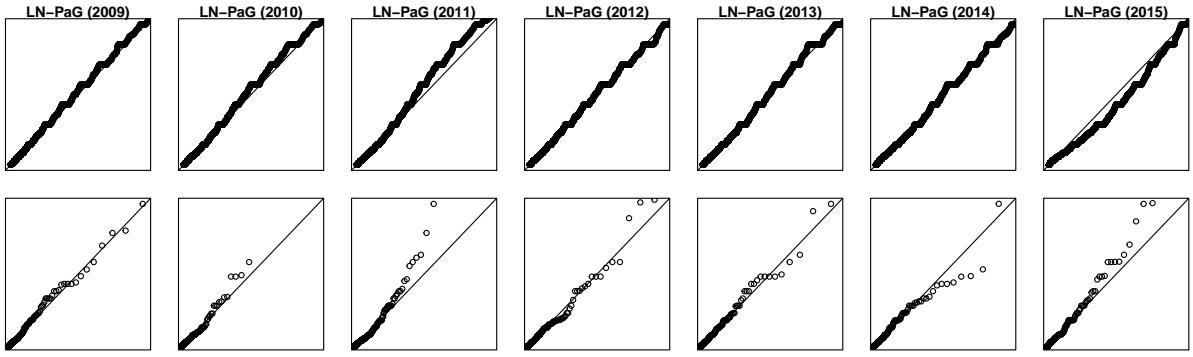


ILLUSTRATION 3.17 – *QQplots* pour les observations sous le seuil (ligne 1) et pour les observations au-dessus du seuil (ligne 2), pour le modèle LN-PaG (2011-2018)

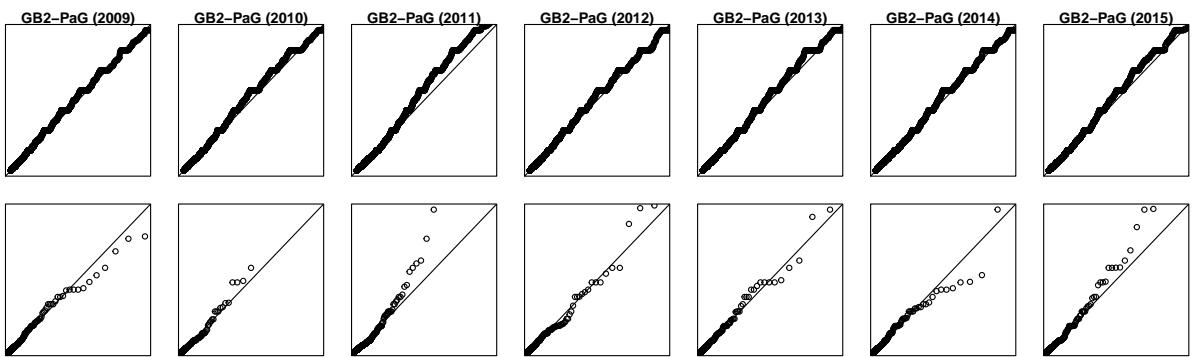


ILLUSTRATION 3.18 – *QQplots* pour les observations sous le seuil (ligne 1) et pour les observations au-dessus du seuil (ligne 2), pour le modèle GB2-PaG (2011-2018)

L'ajustement est globalement bon pour les deux modèles, bien que moins performant pour l'année 2011. On constate aussi que les deux modèles s'ajustent moins bien au-dessus du seuil, car pour l'année 2014, les points passent visiblement sous la droite. Il est encore une fois difficile de dire quel est le meilleur modèle parmi les deux testés sur les données de la base de données DataNFID.

Tests statistiques

Une autre façon de vérifier l'ajustement des modèles et de regarder des statistiques d'ajustement. On utilise encore les statistiques D et A^2 .

Le Tableau 3.27 présente les valeurs obtenues pour les deux statistiques concernant les deux modèles. On constate que pour la statistique D , les valeurs sont extrêmement similaires, sauf pour les années 2013 et 2015 où le modèle GB2-PaG est meilleur. Pour la statistique A^2 , le modèle GB2-PaG est meilleur sur toutes les années sauf 2009. Il est aussi intéressant de constater que les deux modèles semblent moins bons pour l'année 2015, où les statistiques sont bien plus élevées que pour les autres années. Selon ces tests, le modèle GB2-PaG serait celui à utiliser.

	2009	2010	2011	2012	2013	2014	2015
Statistique D							
LN-PaG	0.04	0.04	0.03	0.04	0.04	0.04	0.07
GB2-PaG	0.04	0.04	0.03	0.04	0.03	0.04	0.04
Statistique A^2							
LN-PaG	7.20	7.30	7.88	8.40	7.24	7.94	57.78
GB2-PaG	7.58	7.17	7.80	7.69	6.04	6.09	12.07

TABLEAU 3.27 – Tests d’ajustements des modèles pour la base de données DataNFID

Critère d’information

On cherche à comparer la log-vraisemblance et les critères d’information AIC et BIC pour chacun des 3 modèles.

	2009	2010	2011	2012	2013	2014	2015
Log-vraisemblance							
LN-PaG	97 512	92 308	114 854	113 518	106 300	103 512	103 642
GB2-PaG	97 511	92 304	114 854	113 506	106 290	103 495	103 440
Akaike Information Criterion, AIC							
LN-PaG	195 035	184 629	229 719	227 048	212 612	207 036	207 296
GB2-PaG	195 037	184 624	229 724	227 028	212 595	207 006	206 896
Bayesian Information Criterion, BIC							
LN-PaG	195 036	184 630	229 720	227 049	212 613	207 037	207 297
GB2-PaG	195 038	184 625	229 725	227 029	212 596	207 007	206 897

TABLEAU 3.28 – Critères d’informations des modèles ajustés pour la base de données DataNFID (2010-2015)

Pour les critères d’informations, on constate, au Tableau 3.28, que les valeurs sont très proches pour les deux modèles. Cependant, les valeurs du modèle GB2-PaG sont plus basses pour la majorité des années, sauf pour l’année 2011 où le modèle LN-PaG est meilleur. Avec ce test on choisit encore une fois le modèle GB2-PaG.

Conclusion

Comme pour la base de données canadienne de périls incendies, le modèle à sélectionner serait le modèle GB2-PaG. Il faut toutefois noter que les modèles étaient assez équivalents et que quelqu’un pourrait choisir le modèle Ln-PaG sous prétexte qu’il est moins complexe.

3.10.3 Mesure de risque

Risque au fil du temps

On calcule la mesure VaR et la mesure TCM, cette dernière étant définie au même niveau de confiance comme :

$$TCM_{0.90}[B] = \text{med}[B|B > VaR_{0.90}(B)] = VaR_{0.95}(B),$$

voir So and Wong (2012). Il a été mentionné précédemment que tous les α étaient inférieurs à 1, rendant impossible le calcul de la mesure $TVaR$. C'est pour cette raison qu'on utilise la mesure TCM.

	2009	2010	2011	2012	2013	2014	2015
LN-PaG	124 679	124 803	255 829	143 172	148 252	154 937	182 285
	260 561	261 245	508 444	300 120	313 850	306 233	327 310
GB2-PaG	122 645	121 814	252 683	135 606	141 059	145 117	137 745
	265 530	264 512	508 524	291 984	307 738	296 373	301 814

TABLEAU 3.29 – Mesure VaR et TCM (2ieme lignes) au niveau de confiance 90%, pour la base de données DataNFID (2009-2015)

Les résultats pour ces mesures sont présentés au Tableau 3.29. Il n'y a pas de tendance apparente, soit à la hausse ou à la baisse. On constate aussi que les deux modèles donnent des mesures assez similaires. Cependant, on constate que l'année 2011 est celle où les mesures de risque sont les plus élevées, ce qui correspond à l'année où la perte moyenne est la plus élevée.

Prédiction

Afin de comparer le pouvoir prédictif des deux modèles, on utilise les données de l'année précédente afin estimer $\Pr(B > 100 000)$ pour l'année actuelle.

	2010	2011	2012	2013	2014	2015
Actuel	0.12	0.17	0.12	0.12	0.13	0.12
LN-PaG	0.12	0.12	0.18	0.13	0.13	0.14
GB2-PaG	0.12	0.12	0.17	0.13	0.13	0.13

TABLEAU 3.30 – Valeurs prédites et actuelles de $\Pr(B > 100 000)$, pour la base de données DataNFID (2009-2015)

On présente le résultat de cette probabilité au Tableau 3.30. On remarque qu'on prédit moins bien les valeurs pour les années 2011 et 2012. Cela peut être expliqué par le fait que l'année 2011 était un peu hors du commun. Ainsi, les résultats de l'année 2010 prédisent moins bien ceux de 2011, tandis que les résultats de 2011 prédisent à la hausse ceux de 2012.

Chapitre 4

Partage du risque lié aux pertes en sinistre incendie

4.1 Introduction

Dans ce chapitre, on vise à étudier le partage du risque lié aux pertes en sinistre incendie. Pour ce faire, on utilise les modèles construits au Chapitre 3 à partir de la base de données DataNFID.

On débute par créer des classes basées sur les types de construction des bâtiments, tels que fournis dans la base de données DataNFID. L'objectif sera de calculer la part de chacune des classes, voir Tableau 4.1, au montant total des pertes.

Classe j	Construction	n_j
1	Combustible	12 242
2	Combustible protégé	26 039
3	Bois massif	738
4	Non combustible	1797
5	Non combustible protégé	3041

TABLEAU 4.1 – Classes basés sur le type de construction

Dans chacune des classes j , il y a n_j cas d'incendie. Au Tableau 4.2, on présente quelques caractéristiques pour chacune des classes. Il faut noter que pour l'entièreté de cette section, les montants de pertes sont sur une base de 100 000, pour faciliter les calculs numériques.

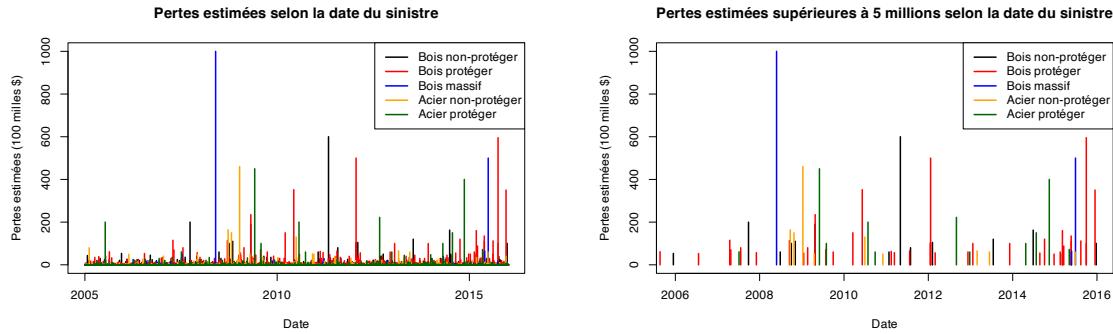


ILLUSTRATION 4.1 – Pertes selon le temps pour chaque classe

	1	2	3	4	5
Médiane :	0.17	0.11	0.30	0.10	0.04
Moyenne :	1.10	1.09	3.80	2.18	1.54
Percentiles :					
0.95	4.43	4.32	8.00	8.98	5.00
0.99	11.89	10.00	23.15	35.92	20.60
0.999	58.41	60.00	631.50	152.84	200.00
Pertes maximales :					
	190	352	65	150	222
	200	500	500	164	400
	600	596	1000	460	450
Proportion :	0.25	0.53	0.05	0.07	0.09

TABLEAU 4.2 – Statistiques descriptives pour chacun des classes

On constate que la médiane, la moyenne et les percentiles sont nettement plus élevés pour la classe 3. En examinant les pertes maximales, on observe que la plus grande perte, s'élevant à 1000, se trouve dans la classe 3. Cependant, bien qu'il y ait un potentiel de pertes très élevées, les montants diminuent rapidement, comme illustrés dans l'Annexe C. En effet, après les deux premières pertes maximales, les autres pertes sont considérablement plus petites comparées à celles des autres classes.

On présente également à l'Illustration 4.1, les pertes au fil du temps pour toutes les classes. On voit sur cette illustration qu'il y a plusieurs sinistres qui sont considérablement plus élevés que les autres. On rappelle l'importance de ne pas considérer ces sinistres comme aberrants, puisqu'ils sont typiques des données de péril incendie.

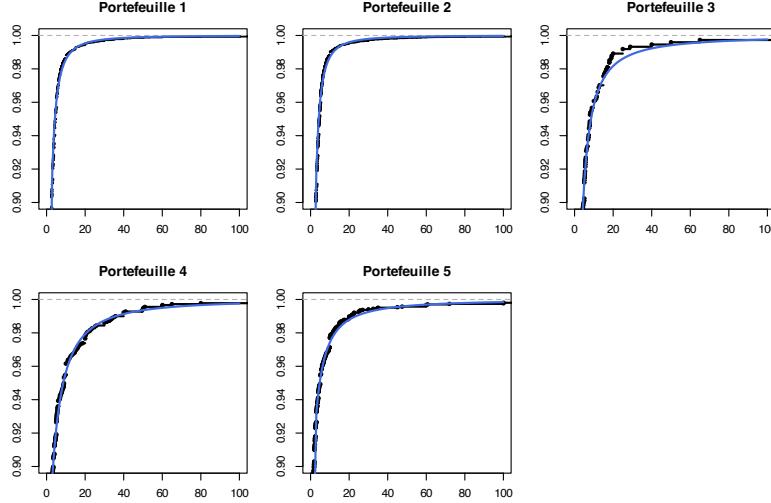


ILLUSTRATION 4.2 – Fonction de répartition empirique des pertes comparativement à la fonction de répartition des pertes du modèles

4.2 Modélisation par classe des pertes en sinistre incendie

Pour calculer la contribution de chaque classe au montant total des pertes, il est nécessaire de modéliser chacune des classes indépendamment. On utilise le modèle LN-PaG présenté à la Section 3.4, dont le paramètre α estimé pour chacune des classes est présenté au Tableau 4.3.

Classe j	1	2	3	4	5
$\hat{\alpha}_j$	1.75	1.75	1.30	1.52	1.40

TABLEAU 4.3 – Paramètres estimés pour chacun des classes

Pour obtenir ces paramètres, on fait le choix d'un seuil à 3 pour toutes les classes. Pour simplifier la modélisation, qui n'est pas le point central de cette section, il semblait raisonnable de faire ce choix. On peut voir à l'Illustration 4.2 que la loi LN-PaG offre une bonne adéquation pour l'ensemble des classes, sauf pour la classe 3. Cela peut être expliqué par le fait que le nombre de données supérieur au seuil est de 106 pour ce classe, ce qui n'est pas beaucoup comparativement aux autres classes, voir Tableau 4.4. Idéalement, il aurait fallu choisir un seuil plus haut pour la classe 3, mais le problème du manque de données aurait été empêtré.

Classe j	1	2	3	4	5
$n_{\theta,j}$	975	2111	106	190	205

TABLEAU 4.4 – Nombre $n_{\theta,j}$ de sinistres de la classe j avec pertes supérieures au seuil θ

Pour créer un modèle de fréquence-sévérité, comme présenté à l'équation (1.1), il est nécessaire de modéliser le processus d'avènement des sinistres. On suppose que le processus de Poisson

homogène décrit adéquatement l'avènement des sinistres incendie dans le temps pour chacune des classes. On estime l'intensité λ_j , $j \in \{1, \dots, 5\}$, de ces processus par la méthode du maximum de vraisemblance à partir de la base de données DataToronto. La procédure générale d'estimation est décrite à l'Annexe A. Les paramètres estimés pour chaque classe sont présentés dans le Tableau 4.5.

Classe j	1	2	3	4	5
$\hat{\lambda}_j$	3.05	6.48	0.18	0.45	0.76

TABLEAU 4.5 – Paramètres estimés du processus de Poisson homogène pour chaque classe

Cependant, il faut mentionner que dans la base de données DataNFID, l'information sur l'exposition n'est pas donnée. On ne connaît pas le nombre d'assurées dans chacune des j classes. On fait l'hypothèse que la composition des classes reste semblable. Pour le moment, c'est le mieux qu'on peut faire avec les données à disposition.

4.3 Règle de partage de risques

Pour calculer la contribution de chacun des j classes, il est nécessaire de faire un choix sur la règle de partage à utiliser. Plusieurs choix sont disponibles, en voici quelques exemples

- *Uniform risk-sharing* ;
- *Proportional mean* ;
- *Order statistics risk-sharing* ;
- *Conditional mean*.

Pour d'autres mesures, voir Denuit et al. (2022). Les règles de partage de risques ont des propriétés souhaitables qu'elles devraient respecter, telles que :

1. *Reshuffling* : l'ordre dans laquelle sont placés les risques ne change pas leur contribution.
2. Normalisation : le fait de ne pas avoir de pertes devrait entraîner une contribution nulle.
3. Transitivité : Si la perte d'un participant augmente par un montant fixe, uniquement sa contribution devrait augmenter.
4. Homogénéité positive : Si le risque considéré augmente par la multiplication d'une constante, alors la règle de partage de risque devrait aussi être multipliée par cette constante.

Pour plus d'information sur les propriétés souhaitables des règles de partage de risque, également voir Denuit et al. (2022). Pour cette analyse, on se concentre sur les règles *Proportional mean* et *Conditional mean*. La règle de partage de risques *Proportional mean* respecte les propriétés 1, 2, et 4, alors que la règle *Conditional mean* les respecte toutes.

4.3.1 Proportional mean

On commence par regarder une règle de partage de risques pour laquelle tous les participants acceptent de payer un pourcentage de la perte totale basé sur l'espérance des pertes qu'ils amènent. On appelle cette règle *Proportional mean*.

Définition 11. Soit la v.a $S_d = X_1 + \dots + X_d$ représentant le coût total, alors la contribution de chaque classe X_j , selon la règle de partage Proportional mean, est donnée par la v.a.

$$C_{j,d} = \frac{E[X_j]}{E[S_d]} S_d, \quad j = 1, 2, \dots, d.$$

On peut montrer que l'espérance de la variable C_j est égale à l'espérance de la v.a de X_j ,

$$E[C_{j,d}] = E\left[\frac{E[X_j]}{E[S_d]} S_d\right] = \frac{E[X_j]}{E[S_d]} E[S_d] = E[X_j].$$

Par conséquent, l'espérance de la contribution de la classe j est égale à l'espérance des coûts éventuels couverts par le contrat de la classe j .

4.3.2 Conditional mean

On regarde maintenant une règle de partage où les participants paient l'espérance des coûts qu'ils amènent en sachant la perte totale.

Définition 12. Soit la v.a S représentant le coût total, alors la contribution de chacune des classes X_j , selon la règle de partage Conditional mean, est donnée par la v.a.

$$C_{j,d} = E[X_j|S_d], \quad j = 1, 2, \dots, d.$$

Pour calculer les valeurs de $C_{j,d}$, on doit être capable de calculer

$$E[X_j|S_d = s] = \frac{E[X_j \times 1_{\{S_d=s\}}]}{f_{S_d}(s)},$$

ce qui peut être une tâche assez volumineuse à accomplir numériquement. Pour y parvenir, on utilise l'approche présentée dans Blier-Wong et al. (2022). Essentiellement, on utilise la fonction génératrice ordinaire des allocations espérées, qu'on définit comme

$$\mathcal{P}_S^{[j]}(t) := \sum_{j=1}^{\infty} E[X_j \times 1_{\{S_d=k\}}] t^k, \quad j = 1, \dots, d.$$

À l'exemple 3.6 de Blier-Wong et al. (2022), les auteurs donnent la forme de $\mathcal{P}_S^{[j]}(t)$ pour une loi Poisson composée :

$$\begin{aligned} \mathcal{P}_{S_d}^{[j]}(t) &= \lambda_j t \mathcal{P}'_{B_j}(t) \mathcal{P}_S(t) \\ &= \lambda_j t \left(\sum_{k=0}^{\infty} (k+1) f_{B_j}(k+1) t^k \right) \left(\sum_{k=0}^{\infty} f_{S_d}(k) t^k \right). \end{aligned} \tag{4.1}$$

Avec l'équation 4.1 et la transformée de Fourier rapide (FFT), il est très simple de calculer les allocations espérées. De plus, cette approche est intéressante, car elle fonctionne aussi dans le cas où les v.a. X_j seraient dépendantes.

4.3.3 Méthode de discrétilisations

Pour être en mesure de calculer les valeurs de ces deux règles de partage de risques, il est nécessaire de connaître la loi de la v.a. S_d . En pratique, il est généralement très difficile de calculer la loi de S_d . En effet, pour obtenir la loi de S_d , il faudrait faire la convolution des variables X_j $j = 1, 2, \dots, d$. Cependant, comme on utilise une loi de Poisson composée avec sévérité lognormale-Pareto généralisée, il n'est pas possible de trouver des formes analytiques pour les caractéristiques de la loi de S_d . De plus, si on veut avoir les valeurs exactes, il faudrait que la loi de la sévérité soit discrète pour utiliser la FFT, mais on utilise une loi continue. Il est donc nécessaire d'utiliser une méthode de discrétilisation sur la v.a. B_j . On travaillera avec les v.a. discrétilisées \tilde{B}_j et \tilde{S} avec un pas de discrétilisation $h = 0.01$, ce qui implique que les valeurs calculées ne sont pas les valeurs exactes.

4.4 Résultats

On définit la v.a. S_5 , qui représente le montant total des pertes pour une journée. On regarde pour différentes valeurs de s quelle sera la contribution de chacune des classes selon les deux principes présentés.

Dans le Tableau 4.6, on présente les contributions de chaque classe selon le montant s et dans le Tableau 4.7, on présente les probabilités que la v.a S_5 soit plus grande que ces valeurs. On constate que les deux méthodes ne produisent pas exactement les mêmes résultats, et cette différence devient plus marquée à mesure que le montant s augmente.

En effet, la contribution selon la règle de partage de risque *proportional mean* ne change pas en fonction du montant s et reste constante. En revanche, la contribution selon la règle de partage de risque *conditional mean* dépend de la valeur de s , ce qui signifie que la contribution varie avec la valeur de s . Comme montré à l'Illustration 4.3, le pourcentage de contribution change considérablement en fonction de la valeur de s . Dans ce cas, il est préférable d'utiliser la règle de partage de risque *conditional mean*. Cette variation s'explique bien à l'aide de la classe 3. Plus tôt, il a été mentionné que la classe 3 avait produit deux sinistres très élevés, mais que généralement les autres sinistres provenant de cette classe sont plutôt bas. Donc, plus le montant de perte est élevé, plus il y a de chance que la classe 3 ait produit des sinistres élevés, alors il est logique qu'elle contribue davantage. Il est également nécessaire d'examiner des valeurs élevées de s , même si la probabilité de leur réalisation est faible. En effet, bien qu'il soit peu probable qu'un sinistre de 50 se produise, seulement 2 % de chance, il est important

de considérer qu'au quotidien, il y a tout de même 2 jours sur 100 où l'on pourrait anticiper un sinistre de plus de 50. À l'échelle d'une année, ces sinistres ne sont donc pas négligeables.

Classes ($s = 5$)						Classes ($s = 10$)					
	1	2	3	4	5		1	2	3	4	5
	(1.75)	(1.75)	(1.30)	(1.52)	(1.40)		(1.75)	(1.75)	(1.30)	(1.52)	(1.40)
Cond. Mean	1.47	2.98	0.09	0.18	0.29		2.88	5.98	0.21	0.38	0.55
Prop. Mean	1.30	2.61	0.19	0.41	0.49		2.60	5.22	0.38	0.82	0.98
Classes ($s = 30$)						Classes ($s = 50$)					
	1	2	3	4	5		1	2	3	4	5
	(1.75)	(1.75)	(1.30)	(1.52)	(1.40)		(1.75)	(1.75)	(1.30)	(1.52)	(1.40)
Cond. Mean	7.99	15.87	1.04	2.48	2.63		12.22	23.39	2.24	6.07	6.08
Prop. Mean	7.81	15.67	1.14	2.46	2.93		13.01	26.11	1.91	4.09	4.88
Classes ($s = 117.54$)											
	1	2	3	4	5		1	2	3	4	5
	(1.75)	(1.75)	(1.30)	(1.52)	(1.40)		(1.75)	(1.75)	(1.30)	(1.52)	(1.40)
Cond. Mean	24.03	44.14	7.61	20.62	21.13						
Prop. Mean	30.58	61.39	4.48	9.62	11.46						

TABLEAU 4.6 – Contribution de chaque classe pour une valeur de s (paramètre α)

k	5	10	30	50	117.54
$\Pr(S_5 > k)$	0.77	0.45	0.07	0.02	0.005

TABLEAU 4.7 – Probabilité que la v.a. S soit plus grande 5, 10, 30, 50 et 117.54

On observe le même phénomène que pour la contribution au montant total : plus on calcule la contribution à la mesure $TVaR$ à un quantile élevé, plus la contribution de la classe 3 augmente et celle de classe 1 et 2 diminue.

Un avantage de la règle de partage de risque *conditional mean* est que toutes les allocations espérées sont déjà calculées, ce qui rend très simple le calcul de la contribution à la mesure $TVaR$ de la classe j . La contribution à la mesure $TVaR$ est donnée par

$$TVaR(X_j, S_d) = \frac{1}{1 - \kappa} (E[X_j \times 1_{\{S \geq VaR_\kappa(S)\}}] + VaR_\kappa(X_j, S_d) [F_{S_d}(VaR_\kappa(S_d)) - \kappa]),$$

où $VaR_\kappa(X_j, S_d) = E[X_j | S = VaR_\kappa(S_d)]$ et $E[X_j \times 1_{\{S > VaR_\kappa(S)\}}] = \sum_{k \in (VaR_\kappa(S_d), \infty)} E[X_j \times 1_{\{S=k\}}]$. On présente les résultats au Tableau 4.8.

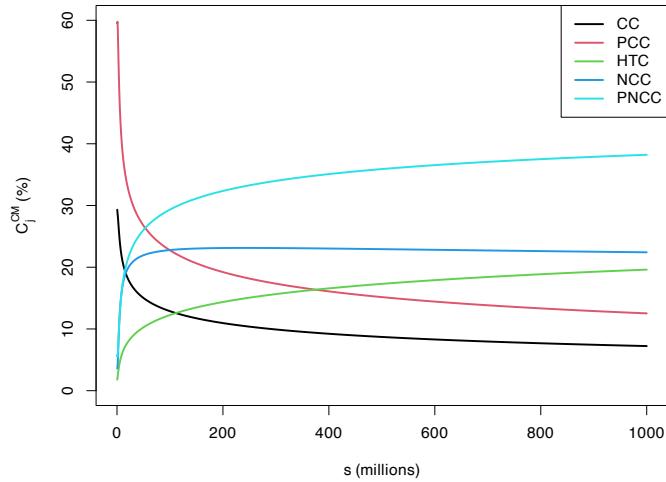


ILLUSTRATION 4.3 – Évolution de la contribution selon la valeur s pour la règle de partage de risque *conditional mean*

κ	$TVaR_\kappa(\tilde{X}_i; \tilde{S})$					$TVaR_\kappa(\tilde{S})$
	1 (1.75)	2 (1.75)	3 (1.30)	4 (1.52)	5 (1.40)	
0.900	11.92 (0.22)	22.80 (0.42)	3.32 (0.06)	7.48 (0.14)	8.37 (0.16)	53.90
0.950	16.09 (0.20)	30.18 (0.38)	5.65 (0.07)	12.61 (0.16)	14.25 (0.18)	78.80
0.990	33.22 (0.17)	60.21 (0.30)	19.60 (0.10)	39.79 (0.20)	47.83 (0.24)	200.65
0.995	46.11 (0.15)	82.79 (0.27)	33.52 (0.11)	63.92 (0.21)	79.82 (0.26)	306.16
0.999	100.96 (0.12)	178.53 (0.21)	115.33 (0.14)	186.39 (0.22)	257.08 (0.31)	838.36

TABLEAU 4.8 – Contribution à la mesure $TVaR$ pour chacune des classes

Conclusion

Au courant de cet essai, on a analysé en profondeur la base de données des sinistres incendies de la ville de Toronto ainsi que la *National Fire Information Database*, ce qui a permis d'avoir une meilleure compréhension du risque incendie. À l'aide des bases de données, on a pu tester différents modèles et déterminé lesquelles étaient les meilleures. Pour les deux bases de données, il s'est avéré que les modèles LN-PaG et GB2-PaG étaient les meilleurs. Ensuite, avec ces modèles on a calculé les mesures de risque VaR et $TVaR$, ce qui a permis d'avoir une idée du risque dans la queue de la distribution pour le péril incendie. Il s'avère que le risque est extrêmement élevé, puisqu'on obtient des paramètres α autour de 1.4 pour l'ensemble des bases de données testées.

On a également fait une application de partage de risque. On a créé cinq classes selon le type de matériaux utiliser pour la construction d'un bâtiment. À l'aide de technique récente, il a été possible de calculer rapidement la règle de partage de risque *Conditional Mean*. Cette application a permis de constater que le type de matériaux a bien un effet sur le montant total de perte.

Il y a évidemment d'autres pistes de recherche qui auraient pu être envisagées. Pour faire le raccordement de loi il y serait possible d'utiliser un GLM ou autre type de modèles prédictif pour modéliser la portion sous le seuil. Cette approche permettrait d'utiliser les variables explicatives. Dans le même esprit, il serait intéressant d'examiner la possibilité d'ajouter des covariables à la portion modélisée par loi Pareto généralisée. Cela permettrait d'identifier les types de protection susceptibles de réduire le risque dans la queue de la distribution. D'un côté plus pratique, rechercher ou avoir accès à des données d'assurance où l'exposition serait fournie. Avec les données d'exposition, les calculs de partage de risque seraient beaucoup plus appropriés et utiles pour la création d'une assurance *peer-to-peer* destinée aux acteurs du secteur de la construction en bois massif.

Annexe A

Modélisation du processus d'avènements des incendies

Pour obtenir un modèle complet de fréquence-sévérité, il est essentiel de modéliser le processus d'avènements des sinistres. La variable `TFS Alarm Time` de la base `DataToronto` indique l'heure à laquelle le système d'alarme s'est déclenché, permettant ainsi une analyse graphique de la fréquence d'arrivée des sinistres.

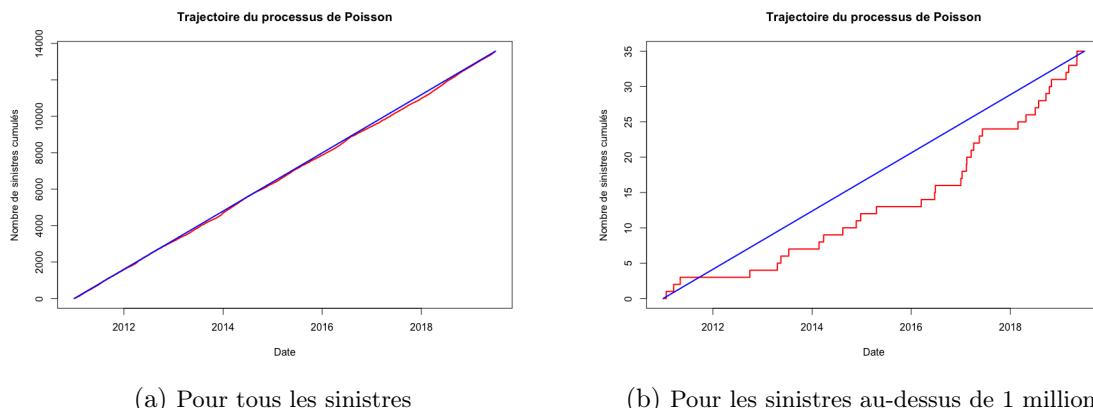


ILLUSTRATION A.1 – Test d'adéquation graphique pour la fréquence d'arrivée des sinistres incendies

On utilise la méthode du maximum de vraisemblance pour estimer le paramètre λ du processus de Poisson :

$$\hat{\lambda} = \frac{\sum_{i=1}^m x_i}{m},$$

où x_i est le nombre de sinistres par jour et m est le nombre total de sinistres pour la durée complète de la base de données. On choisit d'estimer le paramètre lambda sur une base journalière. Pour l'ensemble des sinistres incendie, on obtient $\hat{\lambda} = 4.37$ (par jour), tandis que pour les sinistres incendie supérieurs à 1 million, on obtient $\hat{\lambda} = 0.011$ (par jour).

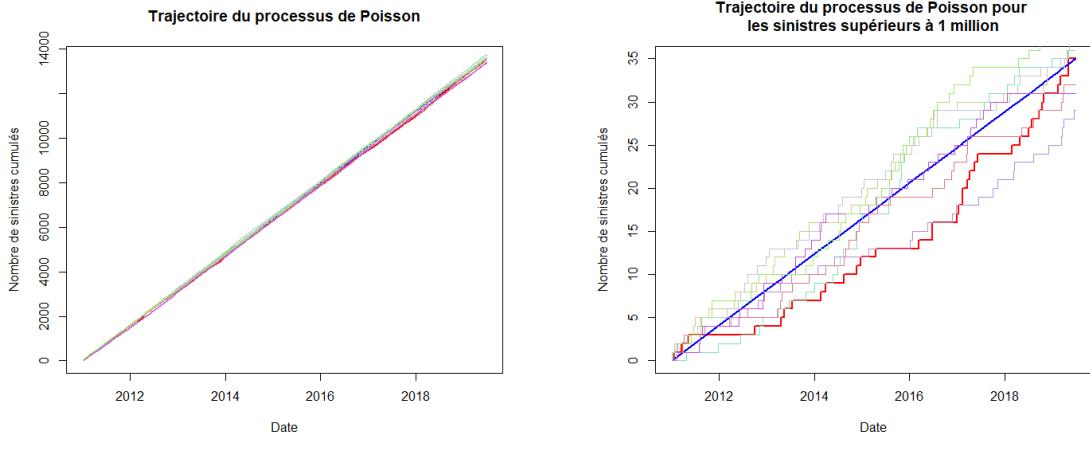


ILLUSTRATION A.2 – Pour les sinistres supérieurs à 1 million

À l’Illustration A.1a, la ligne bleue représente la moyenne théorique du nombre de sinistres cumulés, tandis que la ligne rouge représente le nombre de sinistres cumulés observés de la base de données DataToronto. On constate qu’on est très près de la valeur moyenne théorique. En simulant d’autres parcours de ce processus, on constate effectivement, comme le montre l’Illustration A.2a, que ces autres parcours semblent également suivre un processus de Poisson homogène.

Il est également intéressant d’examiner la fréquence d’arrivée des sinistres dépassant un certain seuil, ici 1 million de dollars. À l’Illustration A.1b on voit que la fréquence d’arrivée des sinistres diffère de la valeur théorique. Cependant, il est important de noter qu’on a seulement 35 sinistres dépassant le seuil de 1 million de dollars, ce qui est insuffisant pour une analyse précise. Pour déterminer si l’hypothèse selon laquelle les sinistres supérieurs à 1 million suivent un processus de Poisson homogène est plausible, on utilise à nouveau la simulation de différents parcours de processus de Poisson. À l’Illustration A.2, on remarque que le parcours obtenu, en rouge, semble être un chemin plausible comme il n’est pas complètement à l’écart des autres. Une autre façon de vérifier cette hypothèse est d’analyser le temps entre les sinistres. Si l’on modélise l’avènement des sinistres par un processus de Poisson, alors le temps entre deux sinistres successifs obéit à une loi exponentielle de paramètre λ . La variable TFS Alarm Time indique les temps observés entre deux sinistres. L’estimateur selon la méthode du maximum de vraisemblance de λ est $\hat{\lambda} = \frac{\sum_{i=1}^m y_i}{m} = 0.011$, où y_i est le temps écoulé entre le sinistre $i - 1$ et le sinistre i , pour $i \in \{1, \dots, m\}$, et avec $y_0 = 0$. Comme la valeur observée de la statistique de Kolmogorov-Smirnov est 0.374, on ne rejette pas l’hypothèse que le temps inter-sinistre obéit à une loi exponentielle confirmant ainsi le choix du processus de Poisson homogène.

On a adapté la procédure décrite ci-dessus pour déterminer les valeurs estimées des paramètres λ_j , $j \in \{1, \dots, 5\}$, que l’on fournit dans le Tableau 4.5.

Annexe B

Autres bases de données

Pour valider les résultats obtenus avec les bases de données DataToronto et DataNFID, on utilise d'autres bases de données de sinistres incendie. Ainsi, il sera possible de comparer les résultats. Pour ce faire, les données en provenance de San Francisco, les données suédoise ainsi que les données norvégienne.

B.1 Données de San Francisco des pertes en sinistre incendie

La base de données des pertes en sinistre incendie en provenance de San Francisco est disponible sur le site San Francisco Open Data¹. Le jeu de données contient de l'information sur 628 147 cas d'incendies. Pour chaque incident, il y a 64 variables explicatives. Toutefois, 495 705 données de la variable `Estimated.Property.Loss` sont manquantes et 11 770 données indiquent une perte nulle. Ces données ont été retirées de l'analyse. Il reste 20 659 données pour faire l'analyse. Dans le Tableau B.1, on présente les statistiques descriptives de la base de données.

Min.	1 ^{er} Qu.	Médiane	Moyenne	3 ^{er} Qu.	Max.
1	50	1000	47 077	6000	10×10^7

TABLEAU B.1 – Statistiques descriptives de la base de données de San Francisco de sinistres incendies

On applique la méthode POT aux montants de sinistres de la base de données et on obtient les paramètres présentés au Tableau B.2.

$\hat{\alpha}$	$\hat{\lambda}$
1.45	394 274

TABLEAU B.2 – Paramètres de la Pareto généralisée pour les montants de réclamations de la ville de San Francisco

1. DataSF, https://data.sfgov.org/Public-Safety/Fire-Incidents/wr8u-xric/about_data

B.2 Données suédoises des pertes en sinistre incendie

La base de données suédoises des montants de sinistres 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 B.3 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

TABLEAU B.3 – 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 B.4.

$\hat{\alpha}$	$\hat{\lambda}$
1.4308	1.925

TABLEAU B.4 – Paramètres de la Pareto généralisée pour les montants de réclamations suédois

B.3 Données norvégiennes des pertes en sinistre incendie

La base de données norvégienne a été utilisée dans Beirlant et al. (1996). Elle contient $n = 9181$ montants de pertes en sinistre incendie. Il faut noter qu'une franchise de 500 \$ a été appliquée, donc on a les montants de sinistres supérieurs à celle-là. Le Tableau B.5 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

TABLEAU B.5 – 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 B.6.

α	λ
1.3766	1846

TABLEAU B.6 – Paramètres de la Pareto généralisée pour les montants de réclamations norvégien

Pour toutes les base de données testée 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.

Annexe C

Les sinistres les plus élevés selon le type de construction

Les sinistres en bleu font partie des dix montants de perte les plus élevés toutes catégories confondues.

Perte	Valeur	Secteur	Utilité
60	60.0	Diverses	Bâtiments en construction
20.0		Résidentiel	Appartement, 20 unité et plus
19.0	25.0	Résidentiel	Appartement, 20 unité et plus
16.2	16.0	Résidentiel	Appartement, 20 unité et plus
12.0	22.0	Résidentiel	Appartement, 20 unité et plus
11.0	6.0	Industrielle	Propriétés de fabrication industrielle
10.5	24.0	Résidentiel	Appartement, 20 unité et plus
10.1	16.8	Résidentiel	Hôtel à l'année longue
10.0	25.0	Résidentiel	Appartement, 20 unité et plus
8.0	11.5	Résidentiel	Appartement, 20 unité et plus

TABLEAU C.1 – Construction combustible

Perte	Valeur	Secteur	Utilité
59.6	59.6	Services	Bureau médical, de santé, de recherche ou scientifique
50.0	100.0	Industrielle	Production matériel en bois
35.2	38.7	Résidentiel	Logement utilisable toute l'année
35	35.0	Institutionnelle	Maison de retraite
23.5	40.0	Résidentiel	Appartement, 5-20 unités
16.0	65.8	Résidentiel	Logement utilisable toute l'année
15.0	28.0	Résidentiel	Appartement, 20 unité et plus
13.5	44.9	Résidentiel	Appartement, 20 unité et plus
12.0		Résidentiel	Résidentiel
11.5	5.0	Résidentiel	Appartement, 20 unité et plus

TABLEAU C.2 – Construction combustible protégé

Perte	Valeur	Secteur	Utilité
100.0	100.0	Industrielle	Production matériel en bois
50.0		Industrielle	Production matériel en bois
6.5	15.0	Résidentiel	Appartement, 20 unité et plus
5.0	5.0	Stockage	Stockage des produits agricoles
4.0	4.0	Stockage	Élévateurs à grains commerciaux
2.9	2.9	Commercial	Buanderie en libre-service
2.5	2.0	Rassemblement	Aréna
2.5	2.5	Rassemblement	Boîte de nuit
2.0	2.0	Stockage	Garage
1.9	1.9	Rassemblement	Aréna

TABLEAU C.3 – Construction en bois massif

Perte	Valeur	Secteur	Utilité
46	250.0	Industrielle	Raffinerie de pétrole
16.4	16.4	Industrielle	Raffinerie de pétrole
15.0	15.0	Industrielle	Fabrication de matériaux plastiques
13.0	48.0	Industrielle	Fabrication de papier
8.0	17.6	Stockage	Stockage réfrigéré
6.5	8.5	Industrielle	Production de nourriture
6.0	3.0	Commercial	Magasin avec département de meubles
6.0	23.0	Industrielle	Propriétés de fabrication industrielle
5.1	100.0	Industrielle	Raffinerie de pétrole
5.0	2.5	Industrielle	Fabrication de matériaux plastiques

TABLEAU C.4 – Construction non combustible

Perte	Valeur	Secteur	Utilité
45.0	150.0	Industrielle	Raffinerie de pétrole
40.0	400.0	Résidentiel	Hôtel à l'année longue
22.2	22.2	Résidentiel	Appartement, 20 unité et plus
20.0	28.0	Rassemblement	École secondaire
20.0		Stockage	Stockage réfrigéré
15.0	15.0	Institutionnelle	Hôpital générale
10.0	10.0	Services	Bureau d'ingénierie, d'architecture et de technologie
10.0	9.0	Stockage	Stockage du bétail
7.2	7.2	Rassemblement	École secondaire
6.1	8.0	Résidentiel	Appartement, 20 unité et plus

TABLEAU C.5 – Construction non combustible protégé

Annexe D

Liste des variables explicatives

D.1 Données de sinistres incendie de la ville de Toronto

- | | |
|---|--|
| 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> | |

D.2 National Fire Information Data Base (NFID)

1. <i>LINK_ID</i>	34. <i>YEARCONS</i>	67. <i>DOLLOSSB</i>
2. <i>INCDNTID</i>	35. <i>HEIGHT</i>	68. <i>DOLLOSSC</i>
3. <i>JURIS</i>	36. <i>FLRAREA</i>	69. <i>DETECT</i>
4. <i>YEAR</i>	37. <i>NUMBOCC</i>	70. <i>TRANSALM</i>
5. <i>MONTH</i>	38. <i>RISKVALA</i>	71. <i>ACTION</i>
6. <i>DATE</i>	39. <i>RISKVALB</i>	72. <i>PERFORM</i>
7. <i>DAY</i>	40. <i>RISKVALC</i>	73. <i>CONTMETH</i>
8. <i>TIME</i>	41. <i>MANPROT</i>	74. <i>METHDGRP</i>
9. <i>INCIDLOC</i>	42. <i>SPRINPRO</i>	75. <i>SADPERF</i>
10. <i>RESPONSE</i>	43. <i>FIXEDSYS</i>	76. <i>IMPACT</i>
11. <i>MUTLAID</i>	44. <i>AUTODET</i>	77. <i>OCCPNTSA</i>
12. <i>CREWSIZE</i>	45. <i>FIREDET</i>	78. <i>OCCPNTSB</i>
13. <i>RESPSUB</i>	46. <i>OUTPROT</i>	79. <i>OCCPNTSC</i>
14. <i>SUBSSIZE</i>	47. <i>WATER</i>	80. <i>OCCPNTSD</i>
15. <i>NUMBENG</i>	48. <i>SERVICE</i>	81. <i>CSD</i>
16. <i>NUMBAER</i>	49. <i>IGNIOBJ</i>	82. <i>CSD_name</i>
17. <i>NUMBTANK</i>	50. <i>IGNOBGRP</i>	83. <i>csdtype</i>
18. <i>NUMBCFR</i>	51. <i>FUELERGY</i>	84. <i>cmaca</i>
19. <i>DISTANCE</i>	52. <i>ENERGY</i>	85. <i>CMA_name</i>
20. <i>STATARR</i>	53. <i>MATERIAL</i>	86. <i>sactyp</i>
21. <i>NUMBRESC</i>	54. <i>MATERGRP</i>	87. <i>ct</i>
22. <i>NUMDEATH</i>	55. <i>ACTOROM</i>	88. <i>fedres</i>
23. <i>NUMBINJ</i>	56. <i>ACTOMGRP</i>	89. <i>da</i>
24. <i>NUMBCAS</i>	57. <i>ORIGIN</i>	90. <i>postcd</i>
25. <i>MAJOCC</i>	58. <i>ORIGGRP</i>	91. <i>PSDDWELL_CMA6</i>
26. <i>MAJOCGRP</i>	59. <i>LEVELOR</i>	92. <i>PSDDWELL_CMA11</i>
27. <i>PROPCLAS</i>	60. <i>FLAMEINT</i>	93. <i>PSDDWELL_CSD6</i>
28. <i>PROPGRP</i>	61. <i>FLAMEEVER</i>	94. <i>PSDDWELL_CSD11</i>
29. <i>PROPSUBG</i>	62. <i>FLAMEHOR</i>	95. <i>ELDERPOP_CMA</i>
30. <i>PROPOWN</i>	63. <i>SMOKSPRD</i>	96. <i>ELDERPOP_CSD</i>
31. <i>ILLUSE</i>	64. <i>FIREEXT</i>	97. <i>YOUNGPOP_CMA</i>
32. <i>GENCONST</i>	65. <i>DAMEXT</i>	98. <i>YOUNGPOP_CSD</i>
33. <i>METHCONS</i>	66. <i>DOLLOSSA</i>	99. <i>LONEPAR_CMA6</i>

- | | | |
|----------------------------|----------------------------|---------------------------|
| 100. <i>LONEPAR_CMA11</i> | 113. <i>MEDINC_CSD6</i> | 126. <i>LOWEDUC_CSD11</i> |
| 101. <i>LONEPAR_CSD6</i> | 114. <i>MEDINC_CSD11</i> | 127. <i>CRIMRATE_CMA</i> |
| 102. <i>LONEPAR_CSD11</i> | 115. <i>EMPLOY_CMA6</i> | 128. <i>CRIMRATE_CSD</i> |
| 103. <i>AVNUMPER_CMA6</i> | 116. <i>EMPLOY_CMA11</i> | 129. <i>PROPCRIM_CMA</i> |
| 104. <i>AVNUMPER_CMA11</i> | 117. <i>EMPLOY_CSD6</i> | 130. <i>PROPCRIM_CSD</i> |
| 105. <i>AVNUMPER_CSD6</i> | 118. <i>EMPLOY_CSD11</i> | 131. <i>VIOLCRIM_CMA</i> |
| 106. <i>AVNUMPER_CSD11</i> | 119. <i>UNEMPLOY_CMA6</i> | 132. <i>VIOLCRIM_CSD</i> |
| 107. <i>MOVERS1_CMA6</i> | 120. <i>UNEMPLOY_CMA11</i> | 133. <i>UCRCSI_CMA</i> |
| 108. <i>MOVERS1_CMA11</i> | 121. <i>UNEMPLOY_CSD6</i> | 134. <i>UCRCSI_CSD</i> |
| 109. <i>MOVERS5_CMA6</i> | 122. <i>UNEMPLOY_CSD11</i> | 135. <i>respcode</i> |
| 110. <i>MOVERS5_CMA11</i> | 123. <i>LOWEDUC_CMA6</i> | |
| 111. <i>MEDINC_CMA6</i> | 124. <i>LOWEDUC_CMA11</i> | |
| 112. <i>MEDINC_CMA11</i> | 125. <i>LOWEDUC_CSD6</i> | |

Annexe E

Caractéristiques des lois

E.1 Caractéristiques des lois de probabilité

E.1.1 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 $E[X^k] = M_Y(k)$

E.1.2 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 $\tau = 2$.

E.1.3 Loi bêta

- Notation : $X \sim Beta(\alpha, \beta)$
- Paramètres : $\alpha > 0, \beta > 0$
- Support : $x \in [0, 1]$
- Fonction bêta incomplète : $I(x; \alpha, \beta) = \int_0^x u^{\alpha-1} (1-u)^{\beta-1} du, x \in [0, 1]$
- Fonction bêta complète : $I(\alpha, \beta) = I(1; \alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$
- Fonction de densité : $f_X(x) = \frac{1}{I(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \times 1_{\{x \in [0, 1]\}}$
- Fonction de répartition : $F_X(x) = \frac{I(x; \alpha, \beta)}{I(\alpha, \beta)}$, notée $B(x; \alpha, \beta)$

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

- Espérance : $E[X] = \frac{\alpha}{\alpha+\beta}$
- Variance : $\text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
- Fonction génératrice 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 : $E[X^k] = \frac{\Gamma(\alpha+k)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+k)}$
- Espérance tronquée : $E[X \times 1_{\{X \leq d\}}] = \frac{\alpha}{\alpha+\beta} B(d; \alpha+1, \beta), \alpha \leq d \leq \beta$
 - Si $\beta = 1$, $E[X \times 1_{\{X \leq d\}}] = \frac{\alpha d^{\alpha+1}}{\alpha+1}$
 - 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

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

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

Source : Cossette and Marceau (2022).

E.1.4 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 : $p(\frac{1}{\beta})^k k! + (1-p) \prod_{i=1}^2 \frac{\Gamma(k+1)}{\beta^k}$
- Espérance tronquée :

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

- 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 $c = VaR_\kappa(X)$.

Source : Lepage and Zito (2019).

E.1.5 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 $TVaR$: $TVaR_\kappa(X) = \frac{1}{1-\kappa} \frac{bI(p+\frac{1}{a}, q-\frac{1}{a})}{I(p, q)} \bar{B}\left(\frac{\left(\frac{VaR_\kappa(X)}{b}\right)^a}{1+\left(\frac{VaR_\kappa(X)}{b}\right)^a}; p + \frac{1}{a}, q - \frac{1}{a}\right)$

Source : Chan et al. (2018).

E.1.6 Loi Pareto généralisée

- Notation : $PaG(\theta, \xi, \sigma)$
- Paramètres : $\theta \leq 0, \xi > 0, \sigma > 0$
- Support : $x \geq \theta$
- Fonction de densité : $f_X(x) = \frac{1}{\sigma} \left(1 + \frac{\xi}{\sigma}(x - \theta)\right)^{\left(-\frac{1}{\xi}-1\right)}$
- Fonction de répartition : $F_X(x) = 1 - \left(1 + \frac{\xi}{\sigma}(x - \theta)\right)^{-\frac{1}{\xi}}$
- Fonction de survie : $\bar{F}_X(x) = \left(1 + \frac{\xi}{\sigma}(x - \theta)\right)^{-\frac{1}{\xi}}$
- Espérance : $E[X] = \frac{\sigma}{1-\xi} + \theta, \xi < 1$
- 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 : $E[X \times 1_{\{X \leq d\}}] = \theta - d(1 + \frac{\xi}{\sigma}(d - \theta))^{-\frac{1}{\xi}} - \frac{\sigma}{1-\xi} \left((1 + \frac{\xi}{\sigma}(d - \theta))^{1-\frac{1}{\xi}} - 1 \right)$
- Mesure VaR : $VaR_{\kappa}(X) = \frac{\sigma}{\xi} \left((1 - \kappa)^{-\xi} - 1 \right) + \theta$
- Mesure $TVaR$: $TVaR_{\kappa}(X) = \frac{\sigma}{\xi} \left(\frac{1}{1-\xi} (1 - \kappa)^{-\xi} - 1 \right) + \theta$

Source : Chinhamu et al. (2015).

E.2 Caractéristiques des raccordements de lois de probabilité

E.2.1 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 $TVaR$:

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

Source : Grün and Miljkovic (2019).

E.2.2 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 :

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

- Mesure $TVaR$:

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

Source : Grün and Miljkovic (2019).

E.2.3 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 :

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

- Fonction de survie :

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

- Espérance :

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

- 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 $TVaR$:

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

Source : Grün and Miljkovic (2019).

Annexe F

Développement des formes analytiques pour les mesures de risque

F.1 Loi Pareto généralisée

F.1.1 Proportional Hazard Transform (PHT)

Démonstration.

$$\begin{aligned}
PHT[F, r] &= \int_0^\infty \left(1 + \frac{\xi}{\beta}(x - \theta)\right)^{-\frac{r}{\xi}} dx \\
&= \theta + \int_\theta^\infty \left(1 + \frac{\xi}{\beta}(x - \theta)\right)^{-\frac{r}{\xi}} dx \\
&= \theta + \int_\theta^\infty \left(1 + \frac{\xi}{\beta}(x - \theta)\right)^{-\frac{r}{\xi}} dx \\
&= \theta + \int_0^\infty \left(1 + \frac{\xi}{\beta}x\right)^{-\frac{r}{\xi}} dx \\
&= \theta + \frac{\beta}{\xi} \int_1^\infty u^{-\frac{r}{\xi}} du \\
&= \theta + \frac{\beta}{\xi} \frac{u^{-\frac{r}{\xi}+1}}{-\frac{r}{\xi}+1} \Big|_1^\infty \\
&= \theta - \frac{\beta}{\xi} \frac{1}{-\frac{r}{\xi}+1}, \quad \xi < r \\
&= \theta - \frac{\beta}{\xi} \frac{\xi}{\xi - r}, \quad \xi < r \\
&= \theta - \frac{\beta}{\xi - r}, \quad \xi < r \\
&= \theta + \frac{\beta}{r - \xi}, \quad \xi < r
\end{aligned}$$

□

F.1.2 Gini Shortfall (GS)

On réécrit la mesure de risque *Gini Shortfall* sous la forme

$$\begin{aligned} GS[F, (1 - \kappa), \delta] &= TVaR_{1-(1-\kappa)}(X) + \frac{2\delta}{(1-\kappa)^2} \int_c^\infty [1 - F(x)][(1 - \kappa) - (1 - F(x))] dx \\ &= TVaR_{1-(1-\kappa)}(X) + \underbrace{\frac{2\delta}{(1-\kappa)^2} \int_c^\infty [(1 - \kappa)(1 - F(x)) - (1 - F(x))^2] dx}_{\mathbf{A}}, \end{aligned}$$

où $(1 - \kappa) = 1 - \kappa$ et $c = VaR_\kappa(X)$. Comme la mesure *TVaR* à déjà été calculée, il reste à trouver le deuxième terme.

Démonstration. On a

$$\begin{aligned} \mathbf{A} &= \frac{2\delta}{(1-\kappa)^2} \int_c^\infty \left((1 - \kappa) \left(1 + \frac{\xi}{\beta}(x - \theta) \right)^{-\frac{1}{\xi}} - \left(1 + \frac{\xi}{\beta}(x - \theta) \right)^{-\frac{2}{\xi}} \right) dx \\ &= \frac{2\delta}{(1-\kappa)^2} \left[\underbrace{(1 - \kappa) \int_c^\infty \left(1 + \frac{\xi}{\beta}(x - \theta) \right)^{-\frac{1}{\xi}} dx}_{B_1} - \underbrace{\int_c^\infty \left(1 + \frac{\xi}{\beta}(x - \theta) \right)^{-\frac{2}{\xi}} dx}_{B_2} \right]. \end{aligned}$$

On commence par résoudre l'intégrale générale B_r :

$$\begin{aligned} B_r &= \int_c^\infty \left(1 + \frac{\xi}{\beta}(x - \theta) \right)^{-\frac{r}{\xi}} dx \\ &= \int_{c-\theta}^\infty \left(1 + \frac{\xi}{\beta}x \right)^{-\frac{r}{\xi}} dx \\ &= \frac{\beta}{\xi} \int_{1+\frac{\xi}{\beta}(c-\theta)}^\infty u^{-\frac{r}{\xi}} du \\ &= \frac{\beta}{r-\xi} \left(1 + \frac{\xi}{\beta}(c-\theta) \right)^{1-\frac{r}{\xi}} \\ &= \frac{\beta}{r-\xi} \left(1 + \frac{\xi}{\beta} \left(\frac{\beta}{\xi} ((1-\kappa)^{-\xi} - 1) + \theta - \theta \right) \right)^{1-\frac{r}{\xi}} \\ &= \frac{\beta}{r-\xi} (1-\kappa)^{r-\xi}. \end{aligned}$$

Ainsi, l'expression de \mathbf{A} devient

$$\begin{aligned} \mathbf{A} &= \frac{2\delta}{(1-\kappa)^2} \left[(1 - \kappa) \frac{\beta}{1 - \xi} (1 - \kappa)^{1 - \xi} - \frac{\beta}{2 - \xi} (1 - \kappa)^{2 - \xi} \right] \\ &= 2\delta \left[\frac{\beta}{1 - \xi} (1 - \kappa)^{-\xi} - \frac{\beta}{2 - \xi} (1 - \kappa)^{-\xi} \right] \\ &= \frac{2\delta\beta}{(1-\kappa)^\xi} \left[\frac{1}{1-\xi} - \frac{1}{2-\xi} \right] \\ &= \frac{1}{(1-\kappa)^\xi} \frac{2\delta\beta}{(1-\xi)(2-\xi)}, \quad \xi < 1. \end{aligned}$$

On obtient

$$GS[F, (1 - \kappa), \delta] = \frac{\beta}{\xi} \left(\frac{(1 - \kappa)^{-\xi}}{1 - \xi} - 1 \right) + \frac{1}{(1 - \kappa)^\xi} \frac{2\delta\beta}{(1 - \xi)(2 - \xi)} + \theta, \quad \xi < 1.$$

□

Bibliographie

- Abramson, D. (2022). A nonproportional premium rating method for construction risks. *North American Actuarial Journal*, 26(4) :626–645.
- Albrecher, H., Beirlant, J., and Teugels, J. L. (2017). *Reinsurance : Actuarial and Statistical Aspects*. John Wiley & Sons.
- 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, D. P. and Re, S. (2007). Mathematical methods in reinsurance. *Lecture Notes, ETH Zurich*.
- Beirlant, J., Teugels, J. L., and Vynckier, P. (1996). *Practical Analysis of Extreme Values*. Leuven University Press.
- 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, S. (1997). The Swiss Re exposure curves and the MBBEFD distribution class1. *ASTIN Bulletin : The Journal of the IAA*, 27(1) :99–111.
- Biffis, E. and Chavez, E. (2014). Tail risk in commercial property insurance. *Risks*, 2(4) :393–410.
- Blier-Wong, C. (2018). *Correction des primes d'assurance habitation en présence de dépendance spatiale*. mathesis, Université Laval.
- Blier-Wong, C., Cossette, H., and Marceau, E. (2022). Generating function method for the efficient computation of expected allocations. *arXiv preprint arXiv :2207.02654*.
- 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, F. and Gomes, M. I. (2015). Threshold selection in extreme value analysis. *Extreme Value Modeling and Risk Analysis : Methods and Applications*, 69–82.

Canty, A. and Ripley, B. D. (2021). *boot : Bootstrap R (S-Plus) Functions*. R package version 1.3-28.

Chan, J., Choy, S., Makov, U., and Landsman, Z. (2018). Modelling insurance losses using contaminated generalised beta type-ii distribution. *ASTIN Bulletin : The Journal of the IAA*, 48(2) :871–904.

Chao, W. (2022). Pricing catastrophe reinsurance under the standard deviation premium principle. *AIMS Math*, 7 :4472–4484.

Charpentier, A. and Flachaire, E. (2021). Pareto models for risk management. *Recent Econometric Techniques for Macroeconomic and Financial Data*, 355–387.

Chinhamu, K., Huang, C.-K., Huang, C.-S., and Chikobvu, D. (2015). Extreme risk, value-at-risk and expected shortfall in the gold market. *The International Business & Economics Research Journal*, 14(1) :107.

Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM Review*, 51(4) :661–703.

Clauset, A., Young, M., and Gleditsch, K. S. (2007). On the frequency of severe terrorist events. *Journal of Conflict Resolution*, 51(1) :58–87.

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.

Corradin, S., Spa, R., and di Grutto, P. R. (2002). Economic risk capital and reinsurance : an extreme value theory's application to fire claims of an insurance company. *University Catholic de Louvain : Institute of Statistics*.

Corradin, S. and Verbrugge, B. (2001). Economic risk capital and reinsurance : an application to fire claims of an insurance company. *RAS, Pianificazione Redditività di Gruppo, December*.

Cossette, H. and Marceau, E. (2022). *Mathématiques actuarielles du risque : modèles, mesures de risque et méthodes quantitatives*. Monographie.

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, M., Dhaene, J., Goovaerts, M., and Kaas, R. (2006). *Actuarial Theory for Dependent Risks : Measures, Orders and Models*. John Wiley & Sons.

Denuit, M., Dhaene, J., and Robert, C. Y. (2022). Risk-sharing rules and their properties, with applications to peer-to-peer insurance. *Journal of Risk and Insurance*, 89(3) :615–667.

- 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.
- Denuit, M. and Robert, C. Y. (2020). Large-loss behavior of conditional mean risk sharing. *ASTIN Bulletin : The Journal of the IAA*, 50(3) :1093–1122.
- Denuit, M. and Robert, C. Y. (2022). Collaborative insurance with stop-loss protection and team partitioning. *North American Actuarial Journal*, 26(1) :143–160.
- Embrechts, P. and Schmidli, H. (1994). Modelling of extremal events in insurance and finance. *Zeitschrift für Operations Research*, 39(1) :1–34.
- Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2) :180–190.
- Furman, E., Wang, R., and Zitikis, R. (2017). Gini-type measures of risk and variability : Gini shortfall, capital allocations, and heavy-tailed risks. *Journal of Banking & Finance*, 83 :70–84.
- Graf, M. and Nedyalkova., D. (2022). *GB2 : Generalized Beta Distribution of the Second Kind : Properties, Likelihood, Estimation*. R package version 2.1.1.
- Grün, B. and Miljkovic, T. (2019). Extending composite loss models using a general framework of advanced computational tools. *Scandinavian Actuarial Journal*, 2019(8) :642–660.
- Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2013). *Loss Models : Further Topics*. John Wiley & Sons.
- Laudagé, C., Desmettre, S., and Wenzel, J. (2019). Severity modeling of extreme insurance claims for tariffication. *Insurance : Mathematics and Economics*, 88 :77–92.
- Lepage, A. and Zito, A. (2019). ACT-2101 : Projet de recherche. Technical report, Université Laval.
- Lin, Y.-S. (2005). Estimations of the probability of fire occurrences in buildings. *Fire Safety Journal*, 40(8) :728–735.
- Longley-Cook, L. H. (1951). Problems of fire insurance rate making. *Proceedings of the Casualty Actuarial Society*, 38 :94–102.
- Longley-Cook, L. H. (1952). A statistical study of large fire losses with application to a problem in catastrophe insurance. *Proceedings of the Casualty Actuarial Society*, 39 :77–83.

- Longley-Cook, L. H. (1959). Notes on some actuarial problems of property insurance. *Proceedings of the Casualty Actuarial Society*, 46 :66–81.
- Majid, M. H. A. and Ibrahim, K. (2021). Composite pareto distributions for modelling household income distribution in malaysia. *Sains Malaysiana*, 50(7) :2047–2058.
- Mandelbrot, B. (1964). Random walks, fire damage amount and other paretian risk phenomena. *Operations Research*, 12(4) :582–585.
- McNeil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin : The Journal of the IAA*, 27(1) :117–137.
- 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. *General Insurance Convention*, 397–491.
- Ohlsson, E. and Johansson, B. (2010). *Non-Life Insurance Pricing With Generalized Linear Models*. Springer.
- Ossberger, J. (2020). *tea : Threshold Estimation Approaches*. R package version 1.1.
- Parodi, P. (2014). *Pricing in General Insurance*. CRC press.
- Pickands III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 119–131.
- Ramachandran, G. (1974). Extreme value theory and large fire losses. *ASTIN Bulletin : The Journal of the IAA*, 7(3) :293–310.
- Reynkens, T. and Verbeelen, R. (2020). *ReIns : Functions from « Reinsurance : Actuarial and Statistical Aspects »*. R package version 1.0.10.
- Reynkens, T., Verbeelen, R., Beirlant, J., and Antonio, K. (2017). Modelling censored losses using splicing : A global fit strategy with mixed erlang and extreme value distributions. *Insurance : Mathematics and Economics*, 77 :65–77.
- 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.
- Shpilberg, D. C. (1977). The probability distribution of fire loss amount. *Journal of Risk and Insurance*, 103–115.
- So, M. K. and Wong, C.-M. (2012). Estimation of multiple period expected shortfall and median shortfall for risk management. *Quantitative Finance*, 12(5) :739–754.
- Thompson, P., Cai, Y., Reeve, D., and Stander, J. (2009). Automated threshold selection methods for extreme wave analysis. *Coastal Engineering*, 56(10) :1013–1021.

Upretee, S. and Brazauskas, V. (2022). Computing and estimating distortion risk measures : How to handle analytically intractable cases ? *North American Actuarial Journal*, pages 1–21.

Wang, S. (1995). Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance : Mathematics and Economics*, 17(1) :43–54.

Wang, S. S. (2000). A class of distortion operators for pricing financial and insurance risks. *Journal of Risk and Insurance*, 15–36.

Zheng, A., Garis, L., and Pike, I. (2022). Fire severity outcome comparison of apartment buildings constructed from combustible and non-combustible construction materials. *Fire technology*, 58(4) :1815–1825.