

Mémoire présenté devant
l'UFR de Mathématique et Informatique
pour l'obtention du du Diplôme Universitaire d'Actuaire de Strasbourg
et l'admission à l'Institut des Actuaire
le Jeudi 2 décembre 2021

Par : Antoine PIERRE

Titre: Modélisation de la sévérité des traités en excédent de sinistre, approche par la théorie des valeurs extrêmes.

Confidentialité : NON OUI Durée : 1 an 2 ans 3 ans 4 ans 5 ans

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Signature :

Membres du jury de l'Unistra :

Entreprise : Covéa

P. ARTZNER
J. BERARD
A. COUSIN
K.-T. EISELE
M. MAUMY-BERTRAND

Directeur de mémoire en entreprise:

Nom : GILBERG Noémie

Signature : 

Invité :

Nom :

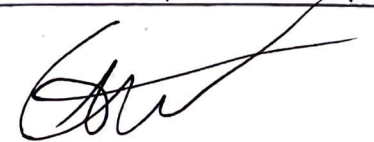
Signature :

Jury de l'Institut des
Actuaire :

Jean MODRY
Magali KELLE VIGON

**Autorisation de publication et de
mise en ligne sur un site de
diffusion de documents
actuariels (après expiration de
l'éventuel délai de confidentialité)**

Signature du responsable entreprise



Secrétariat : Mme Stéphanie Richard

Bibliothèque : Mme Christine Disdier

Signature du candidat





Diplôme Universitaire des Actuaire de Strasbourg

MÉMOIRE D'ACTUARIAT

PIERRE Antoine
Diplôme Universitaire des Actuaire de Strasbourg

Table des matières

Remerciements	4
Résumé	5
Abstract	6
Introduction	7
1 Contexte de l'étude	9
1.1 Généralités sur la réassurance	9
1.1.1 Les besoins et objectifs de la réassurance	9
1.1.2 Fonctionnement de la réassurance	10
1.1.3 La réassurance proportionnelle	11
1.1.4 La réassurance non-proportionnelle	15
1.2 Le groupe Covéa	18
1.2.1 L'histoire du groupe	18
1.2.2 La réassurance non-vie chez Covéa	19
1.3 Problématique de l'étude : la modélisation de la sévérité des traités en excédent de sinistre	20
1.3.1 Notion de branches courtes et branches longues	20
1.3.2 Notion de tranches travaillantes ou non travaillantes	21
1.3.3 Contexte du renouvellement	21
1.3.4 Cotation	21
1.3.5 Limites de la tarification à l'expérience : le <i>burning cost</i>	22
1.3.6 Prime pure, prime technique et prime commerciale	25
2 La théorie des valeurs extrêmes	26
2.1 Convergence en loi du maximum d'un échantillon et loi des valeurs extrêmes	26
2.1.1 Théorie classique et loi du maximum	26
2.1.2 Loi des valeurs extrêmes : <i>Generalized Extreme Value (GEV)</i>	27
2.2 Théorie des excédents au-delà de seuils élevés	29
2.2.1 La loi des excès : méthode <i>Peaks Over the Threshold</i>	29
2.2.2 La distribution de Pareto Généralisée	30
2.2.3 Le théorème de Pickands, Balkema, De Hann	30
2.3 Estimation paramétriques de la loi de Pareto Généralisée	31
2.3.1 La méthode du maximum de vraisemblance	31
2.3.2 La méthode des moments	32
2.3.3 La méthode des moments pondérés	32
2.4 Estimation semi-paramétrique de l'indice des valeurs extrêmes	33
2.4.1 Estimateur de Hill	33
2.4.2 Estimateur de Pickands	33
2.4.3 Estimateur de DEdH	34
2.4.4 Détermination du domaine d'attraction	34
2.5 Détermination du seuil des extrêmes	35
2.5.1 La fonction des excès moyens	35
2.5.2 Le graphe de l'indice des valeurs extrêmes de la <i>GPD</i>	37

2.5.3	Le graphique de Hill et de Pickands	38
2.5.4	Le graphique de Gerstengarbe	40
2.5.5	Méthode de minimisation de l'AMSE	41
3	Modélisation de la sinistralité	43
3.1	Modèle individuel et Modèle collectif	43
3.1.1	Le modèle individuel	43
3.1.2	Le modèle collectif : approche probabiliste fréquence et sévérité	44
3.2	Seuil de modélisation et fréquence au dessus du seuil	44
3.2.1	Sélection du seuil de modélisation	44
3.2.2	Calcul de la fréquence	45
3.3	Modélisation de la sévérité	45
3.3.1	La fonction de répartition empirique	46
3.3.2	Les lois usuelles pour la modélisation de la sévérité en réassurance	47
3.3.3	Prise en compte du seuil de modélisation : lois tronquées	50
3.3.4	Quantile des lois tronquées	51
3.4	Estimation paramétrique	51
3.4.1	La méthode du maximum de vraisemblance	52
3.4.2	La méthode des moments	53
3.5	Évaluation des ajustements des lois	54
3.5.1	Graphique Quantile-Quantile	54
3.5.2	Les tests statistiques	56
3.5.3	Critères statistiques	58
3.6	Modélisation par deux distributions	59
3.6.1	Distribution usuelle et loi de Pareto Généralisée	59
3.6.2	Approche alternative : mélange d'Erlang et loi de Pareto Généralisée	62
3.6.3	Quantile des modèles à deux lois	63
3.7	Calcul du montant à charge dans une tranche	66
3.7.1	Modèle à une loi	66
3.7.2	Modèle à deux lois usuelles	68
3.7.3	Modèle d'Erlang	70
4	Application	72
4.1	Sur données simulées	72
4.1.1	Identification des paramètres du modèle à une loi	73
4.1.2	Identification du seuil des extrêmes	74
4.2	Sur données réelles	77
4.2.1	Revalorisation des sinistres : la mise en <i>as if</i>	78
4.2.2	Analyse descriptive des données	79
4.2.3	Sélection du modèle à une loi	80
4.2.4	Détermination du domaine d'attraction	83
4.2.5	Sélection du seuil des extrêmes	84
4.2.6	Construction du modèle à deux lois usuelles	87
4.2.7	Comparaison modèle à deux lois usuelles et modèle d'Erlang	89
4.2.8	La technique du bootstrap	90
4.2.9	Comparaison du montant à charge des trois modèles	95
4.2.10	Sensibilité au seuil des extrêmes dans le modèles à deux lois usuelles	96
4.2.11	Estimation de la période de retour	98
	Conclusion	101
	Liste des figures	104
	Liste des tableaux	105
	Bibliographie	107

Annexes	108
A Calcul des coût à charge des lois usuelles	108
B Présentation de l'outil de modélisation R Shiny	115

Remerciements

Je tiens à exprimer ma reconnaissance aux personnes qui ont contribué au bon déroulement de mon alternance et à la réalisation de mon mémoire :

Je remercie, en premier lieu, ma tutrice en entreprise, **Noémie Gilberg** pour son suivi sans faille, ses conseils bienveillants, ses remarques pertinentes et ses relectures.

Je tiens également à remercier mon manager **Christophe Guilloux** pour sa pédagogie, ses critiques constructives, son partage de connaissances et son aide quant à la rédaction du mémoire.

Je saisis l'occasion de remercier **Gwenaël Pedron**, récemment arrivé dans l'équipe de tarification, pour ses idées, son implication et les discussions techniques que nous avons eu ensemble.

Je veux remercier tout l'équipe pédagogique du DUAS pour la qualité de leurs enseignements et la passion transmise pour l'actuariat. En particuliers, je remercie ma tutrice universitaire, **Myriam Maumy-Bertrand** pour sa relecture et son suivi durant mes trois années de formation.

Je n'oublie pas **mon équipe de relecture** : **Charline, Juline** et **mes parents**.

Merci à tous.

Résumé

L'objectif de ce mémoire est de compléter et d'optimiser l'outil de modélisation en étendant les possibilités d'ajustement des sinistres à des lois de probabilité, dans le cadre de la tarification des parties non travaillantes des programmes de type excédent de sinistre par risque ou par évènement.

Dans le modèle collectif de l'outil de tarification actuel, la sévérité est uniquement modélisée par la distribution de Pareto. Dans un premier temps, le nombre de lois est étendu à six. Pour sélectionner la loi la plus adaptée aux données, des comparaisons graphiques, des critères numériques et des tests statistiques sont utilisés.

Pour améliorer la tarification des tranches hautes, un modèle à deux lois est proposé. Ce modèle repose sur la théorie des valeurs extrêmes et implique la détermination d'un seuil de raccordement. Il existe différents outils permettant de guider l'actuaire dans le choix de ce seuil, comme le graphique de Hill ou le graphique des excès moyens.

L'impact du modèle à deux lois est illustré par l'application sur des données réelles d'un traité couvrant de l'incendie où une modélisation classique par une loi à queue fine, ne permet pas d'estimer convenablement le coût à charge de la réassurance dans les tranches hautes, souvent non-travaillantes. Cependant, le contexte de la réassurance rend les estimations des différentes quantités volatiles à cause du faible nombre de données disponibles.

Les développements théoriques se sont matérialisés par la création d'une nouvelle interface de modélisation des sinistres sur *R Shiny*. Cette application est intégrée à l'outil de tarification déjà en place et est opérationnelle pour la prochaine période de renouvellement.

Mots-clés : réassurance non-vie, modélisation, sévérité, excédent de sinistre, seuil, théorie des valeurs extrêmes, Pareto Généralisée, tranches hautes, mélange d'Erlang, modèle à deux lois.

Abstract

The objective of this paper is to complement and optimise the modelling tool by extending the possibilities of adjusting claims to probability distributions, in the context of pricing the non-working parts of excess of loss per risk or per event programmes.

In the collective model of the current pricing tool, severity can only be modelled by the Pareto distribution. In a first step, the number of distributions is extended to six. To select the most suitable distributions for the data, graphical comparisons, numerical criteria and statistical tests are used.

In order to improve the pricing of high layer, a spliced-distribution model is proposed. This model is based on the extreme value theory and involves the determination of a connection threshold. There are different tools to guide the actuary in the choice of this threshold, such as the Hill plot or the mean excess plot.

The impact of the splice-distribution model is illustrated by the application to real data of a fire treaty where a classical fine-tail distribution model does not allow for an adequate estimation of the reinsurance cost in the high layer, often non-working. However, the context of reinsurance makes the estimates of the various quantities volatile due to the small amount of data available.

Theoretical developments have resulted in the creation of a new interface for claims modelling on R Shiny. This application is integrated with the existing pricing tool and is operational for the next renewal period.

Keywords : non life reinsurance, modelling, severity, excess of loss, threshold, extreme value theory, Generalized Pareto, non working layer, Erlang mixture, spliced distribution.

Introduction

Le marché de la réassurance est en pleine croissance. Sur les cinq dernières années, les primes de réassurance mondiales ont augmenté en moyenne de 5,60% annuellement. Cette croissance s'explique par des hausses tarifaires et par l'achat de nouvelles couvertures de la part des assureurs. Ces derniers souhaitent renforcer leur protection contre les risques existants et anticiper leur protection contre les risques émergents. Pour suivre la croissance du marché de la réassurance et développer leur compétitivité, le pôle de réassurance non-vie du groupe Covéa souhaite améliorer la tarification des traités en excédent de sinistre, en particulier dans les tranches hautes pour lesquelles il y a peu ou pas de données historiques disponibles. Dans ces cas, un modèle collectif est implémenté avec d'une part l'estimation d'une fréquence, et d'autre part l'estimation d'un coût moyen à charge de la réassurance entre la priorité et la portée du traité.

L'objet du mémoire se concentre sur la modélisation de la sévérité. Lorsque la tranche de réassurance est touchée par plusieurs sinistres durant la période d'observation, il est possible d'estimer un coût à charge pour la réassurance à l'aide d'une seule distribution de probabilité qui modélise l'ensemble de la sinistralité de la cédante. Cependant, lorsqu'une tranche de réassurance n'est jamais touchée par des sinistres, l'actuaire doit estimer un coût à charge dans des grandeurs qui appartiennent à la queue de distribution des sinistres. Dans ce cas, le modèle à une loi dont l'estimation est focalisée sur la partie centrale de la distribution ne permet pas de refléter correctement le risque pour ces tranches non-travaillantes. L'objectif est d'améliorer la modélisation de la sévérité, et in fine la tarification des tranches non-travaillantes. Pour répondre à cette problématique, un modèle à deux lois fondé sur la théorie des valeurs extrêmes est proposé. En effet, [Pickands \(1975\)](#) suggère qu'il existe un seuil à partir duquel les données sont distribuées selon une loi de Pareto Généralisée. L'intérêt est ensuite de construire un modèle à deux lois, avec d'une part une loi classique qui permet de représenter le risque des tranches moyennes et de tarifier convenablement la partie proche de la sinistralité connue, et une loi dédiée aux valeurs extrêmes pour représenter le risque des tranches hautes et tarifier correctement la partie au-delà d'un certain seuil. Une alternative à l'utilisation d'une loi classique pour modéliser les sinistres inférieurs au seuil des extrêmes est d'utiliser une loi mélange d'Erlang comme le propose [Albrecher et al. \(2017\)](#).

La principale difficulté dans la construction des modèles à deux lois réside dans le choix du seuil de raccordement. Ce seuil correspond à un point de rupture dans la distribution des sinistres scindant les sinistres fréquents, mais de faible sévérité des sinistres rares, mais de sévérité intense. Pour déterminer le seuil de jonction, l'actuaire dispose d'outils graphiques et numériques issus de la théorie des valeurs extrêmes.

Du point de vue opérationnel, l'exposé théorique se concrétise par un développement informatique via le déploiement d'une application de modélisation sur l'interface *R Shiny*. Cette application est intégrée à l'outil de tarification déjà existant au sein du pôle de réassurance non-vie du groupe Covéa. L'application sera utilisée lors de la prochaine période de renouvellement des traités.

Le premier chapitre permet de décrire le cadre de travail et de développer la problématique de tarification des tranches des traités en excédent de sinistre faisant intervenir la sinistralité extrême. Le chapitre qui suit permet de développer la théorie des valeurs extrêmes, en particulier la justification de son utilisation et les différentes méthodes qui permettent de sélectionner un seuil des extrêmes. Les modèles et les hypothèses retenus sont décrits dans le troisième chapitre. Enfin, le mémoire se termine par l'application des modèles sur des données simulées pour valider la modélisation et sur des données

réelles pour montrer l'apport d'un modèle à deux lois par rapport à un modèle classique à une loi.

Chapitre 1

Contexte de l'étude

L'assurance est l'opération par laquelle un assureur organise en mutualité une multitude d'assurés exposés à la réalisation de certains risques et indemnise ceux d'entre eux pour qui le risque se réalise grâce à la masse commune des primes collectées. C'est le principe fondamental de mutualisation des risques d'une compagnie d'assurance. Cette mutualisation s'effectue sur une vaste gamme de produits d'assurance : habitation, automobile, vie, santé etc. Cependant, pour certains risques le principe de mutualisation peut être fragile et l'assurance peut se tourner vers une réassurance pour renforcer sa couverture. L'objectif de ce chapitre est de présenter les mécanismes de la réassurance et d'introduire la problématique du mémoire.

1.1 Généralités sur la réassurance

La réassurance est une opération technique et juridique qui permet à l'industrie des assurances de réaliser les impératifs de diversification et de dispersion des risques pour être en conformité avec les différentes normes qui encadrent l'activité des assureurs. La réassurance est donc l'opération par laquelle une entreprise d'assurance se fait assurer à son tour pour tout ou partie des risques qu'elle demeure seule à garantir à l'égard de l'assuré. Il n'y a aucune relation juridique entre l'assuré et le réassureur. En effet, l'article L111-3 du code des assurances énonce : *Dans tous les cas où l'assureur se réassure contre les risques qu'il a assurés [...], il reste seul responsable vis-à-vis de l'assuré.*

La réassurance permet une dilution des risques en laissant à chaque assureur la seule charge de ce qu'il doit conserver pour son propre compte. La réassurance est par essence une transaction financière impliquant un transfert de risque, appelé cession, d'une compagnie d'assurance définie comme la cédante, vers une compagnie de réassurance en échange du paiement d'une prime de réassurance. Le contrat qui régit les rapports entre cédant et réassureur est appelé *traité de réassurance*.

Les mécanismes de réassurance sont classés en deux catégories : la réassurance de sommes (proportionnelle) et la réassurance de dommages (non-proportionnelle) selon que la réassurance s'applique sur le risque ou le sinistre. Ces deux mécanismes sont présentés plus en détails dans la suite.

1.1.1 Les besoins et objectifs de la réassurance

Les assureurs font appel à de la réassurance pour diverses raisons :

- Se protéger contre une sinistralité rare et intense.
- Supporter l'entrée sur un nouveau marché. La proposition d'un produit sur un marché nouveau représente un risque et engendre des coûts pour l'assurance. Ces deux éléments peuvent être difficiles à maîtriser dès le début. La réassurance permet de répondre à l'incertitude qui entoure le développement sur un nouveau marché.
- Se protéger contre les cumuls de risques sur des événements CATNAT (exemple : ouragans, inondations).

- Augmenter sa capacité de souscription, en particulier pour permettre à l'assurance de couvrir de très grands risques qu'il ne serait pas judicieux de conserver pour son propre compte.
- Réduire le besoin en capital. La directive Solvabilité 2 impose aux compagnies d'assurance de détenir un certain niveau de fond propre. Une manière d'alléger le niveau de fond exigé est d'avoir recours à de la réassurance. Dans la grande majorité des cas, le risque de contrepartie généré par la réassurance est très largement compensé par une diminution du capital de solvabilité requis du module souscription.
- Stabiliser le résultat annuel en réduisant la volatilité de la sinistralité.
- Diversifier des risques individuels avec des sommes assurées élevées pour lesquels la mutualisation à l'échelle de l'assurance n'est pas possible (exemple : un satellite).
- Bénéficier d'une expertise technique pour des risques nouveaux difficiles à quantifier, par exemple le risque cyber ou le terrorisme.

Pour offrir ces services et être en mesure d'assumer la volatilité des risques, les réassureurs travaillent à l'échelle internationale à minima. La diversification est effectuée à cette échelle géographique, mais également sur toutes les branches d'assurance : dommage aux biens, automobile, responsabilité civile, transport etc. Cela permet à la réassurance de compenser les écarts de sinistralité qu'il peut y avoir sur un risque spécifique.

1.1.2 Fonctionnement de la réassurance

Les sens de la réassurance

La réassurance a deux sens selon si la compagnie cède le risque ou l'accepte.

En réassurance passive, l'assureur cède les risques qu'il n'est pas en mesure de couvrir seul et qui excède sa capacité de souscription. Les risques sont cédés à un réassureur. C'est la cession en réassurance. Inversement, en réassurance active, l'assureur (ou le réassureur en général) accepte les risques qu'une compagnie souhaite céder. C'est l'acceptation en réassurance.

Les formes de réassurance

Il existe quatre formes de réassurance qui se distinguent par le caractère obligatoire ou facultatif de la cession ou de l'acceptation.

En réassurance obligatoire (par traité), toutes les affaires sont obligatoirement cédées et acceptées par la cédante et le réassureur. La cédante est obligée de verser au traité tous les risques définis dans ce dernier. Le réassureur s'oblige à accepter ces risques. Chaque partie connaît le risque qu'elle cède et qu'elle accepte. Tous les risques du portefeuille sont couverts. La gestion administrative de la réassurance obligatoire est simple, il n'y a pas d'anti-sélection.

En réassurance facultative, la cession et l'acceptation sont facultatives, elles se font individuellement police par police. Le cédant est libre de proposer au réassureur le risque s'il ne figure pas dans les risques automatiquement couverts par le traité ou s'il dépasse la capacité proposée par le réassureur. Le réassureur prend le temps d'analyser le risque par le biais d'études techniques. Il peut accepter ou refuser la police associée. La réassurance facultative se heurte au problème d'anti-sélection si la cédante cède uniquement les mauvais risques. De plus, il y a une lourdeur dans la gestion administrative car les sinistres doivent être suivis précisément.

Enfin, il existe la réassurance FACOB où la cession est facultative et l'acceptation obligatoire. L'asymétrie d'information est importante car la cédante peut décider quels risques réassurer tandis que le réassureur doit les accepter. Inversement, la cession peut être obligatoire et l'acceptation facultative, c'est le cas de la réassurance OB-FAC.

TABLE 1.1: Les quatre formes de réassurance

	Acceptation (réassureur)	
Cession (assureur)	Facultative	Obligatoire
Facultative	Facultative	FACOB
Obligatoire	OB-FAC	Traité

Les modes de distribution

En assurance, les modes de distribution des produits sont nombreux : en direct, par internet, via des agents généraux, via des courtiers, via des bancassurance etc.

En réassurance, les modes de distribution sont plus restreints. Les deux canaux de distribution sont en direct ou via des courtiers de réassurance indépendants. C'est la cédante qui décide d'opter pour l'un ou l'autre mode de distribution. Le fait de souscrire un programme de réassurance en direct permet d'établir une relation de long terme entre l'assureur et le réassureur. Cependant, si l'assureur souhaite limiter sa charge de travail, il peut préférer les affaires qui passent par un courtier.

Le courtier est avant tout mandaté par la cédante. En ce sens, le courtier fournit une liste de services pour l'assureur comme :

- définir les besoins de couverture.
- rédiger les contractuels.
- aider à la préparation des documents permettant le renouvellement des affaires.
- chercher les meilleures conditions sur le marché de la réassurance.
- être l'interlocuteur direct entre l'assureur et le réassureur pour limiter la lourdeur administrative.
- Réaliser les études d'exposition à différents périls naturels pour les programmes CAT.

Historiquement, le métier de courtier en réassurance présentait un aspect commercial important. Au fil des années, le métier a évolué. Aujourd'hui, de nombreux cabinets de courtage en réassurance proposent des analyses de risques pour les assureurs. Ce métier tend à devenir de plus en plus technique. On peut citer AON Re, Guy Carpenter ou Willis Re comme les cabinets de courtage les plus connus.

Le fonctionnement de la réassurance étant présenté, la sous-section suivante aborde les différents types de réassurance.

1.1.3 La réassurance proportionnelle

En réassurance proportionnelle, le transfert de risque est établi à partir de base de proportionnalité et s'applique aux risques couverts par la cédante. C'est-à-dire qu'à un pourcentage de risques cédés correspond un même pourcentage de primes cédées. Un taux de cession τ est défini pour chacune des polices en portefeuille. Le réassureur reçoit une portion de la prime originale correspondant au taux de cession et il paie la même portion des sinistres éventuels qui touchent la police d'assurance.

Si le taux de cession est identique pour chacune des polices, on parle de Quote-Part (*Quota-Share*). S'il varie selon la somme assurée de chaque police, on parle d'Excédent de plein (*Surplus*). Il convient alors de définir le plein de souscription et le plein de rétention.

Définition 1 (Le plein de souscription). *C'est la somme maximale totale que l'assureur s'engage à garantir à l'égard de l'assuré.*

Définition 2 (Le plein de rétention). *C'est le capital maximum garanti conservé par l'assureur pour son propre compte, le surplus étant cédé en réassurance dans la limite du plein de souscription.*

Dans la suite, les différents mécanismes de réassurance sont illustrés sur le portefeuille fictif suivant de 8 polices assurées.

TABLE 1.2: Portefeuille fictif pour illustrer les formes de réassurance

Assuré	1	2	3	4	5	6	7	8
Montant de garantie	98	120	53	165	80	85	40	200

La Quote-Part

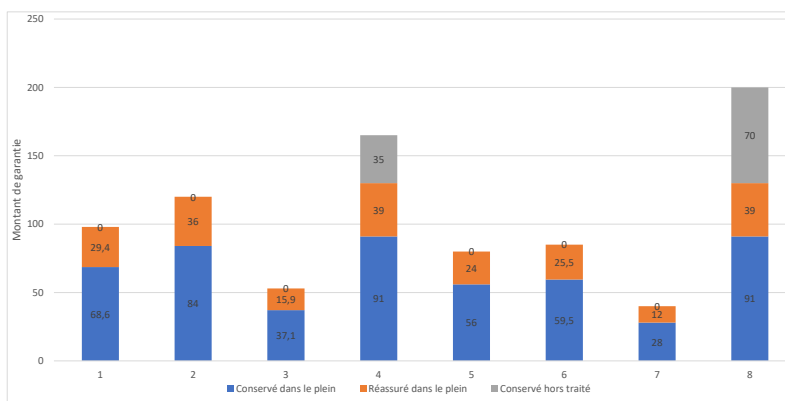
Le taux de cession τ est contractuellement le même pour tous les risques en portefeuille. La cédante et le réassureur s'accordent sur un taux de cession appelé la quote-part. Les primes et sinistres sont cédés à proportion de cette quote-part.

La réassurance en Quote-Part est illustrée sur le portefeuille fictif en considérant un taux de cession de 30% et un plein de souscription de 130.

TABLE 1.3: Quote-Part à 30% et plein de souscription à 130

Assuré	1	2	3	4	5	6	7	8
Montant de garantie	98	120	53	165	80	85	40	200
Conservé dans le plein	68,6	84	37,1	91	56	59,5	28	91
Réassuré dans le plein	29,4	36	15,9	39	24	25,5	12	39
Conservé hors traité	0	0	0	35	0	0	0	70
Total conservé	68,6	84	37,1	126	56	59,5	28	161
Total réassuré	29,4	36	15,9	39	24	25,5	12	39

FIGURE 1.1: Quote-Part à 30% et plein de souscription à 130



La réassurance en Quote-Part présente de nombreux avantages. On peut citer la simplicité de gestion du traité, la réduction du risque à hauteur de la Quote-Part, notamment pour le lancement d'un nouveau produit sur un marché où l'assureur a peu d'expériences. Le réassureur est un réel partenaire de l'assureur. Cependant, la réassurance en Quote-Part possède des limites. Le volume de primes cédées à la réassurance est important et la Quote-Part n'offre pas réellement de couverture pour les sinistres extrêmes.

L'autre forme de réassurance proportionnelle est l'excédent de plein.

L'excédent de plein

Le taux de cession varie en fonction de la somme assurée des risques en portefeuille et de la rétention maximale que la cédante souhaite conserver. Cette rétention est aussi appelée "le plein". Dans ce type de réassurance, seuls les risques présentant une somme assurée supérieure au plein sont cédés au réassureur selon un taux de cession choisi de manière à ce que le sinistre maximal n'entraîne pas une perte supérieure au plein pour la cédante. Le taux de cession se définit pour chacune des polices i comme :

$$\tau_i = \max\left(0, 1 - \frac{R}{Sa_i}\right)$$

où :

R : le plein de rétention.

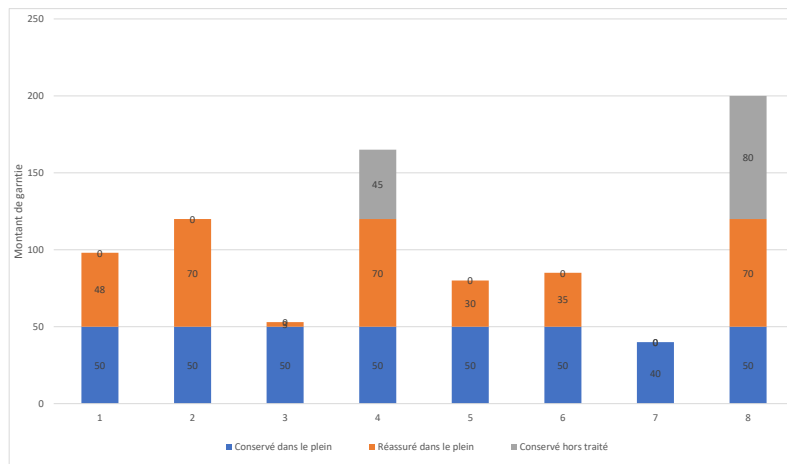
Sa_i : la somme assurée de la police i .

La réassurance en Excédent de plein est illustrée sur le portefeuille fictif en considérant un plein de rétention de 50 et un plein de souscription de 120.

TABLE 1.4: Excédent de plein avec rétention de 50 et plein de souscription à 120

Assuré	1	2	3	4	5	6	7	8
Montant de garantie	98	120	53	165	80	85	40	200
Conservé dans le plein	50	50	50	50	50	50	40	50
Réassuré dans le plein	48	70	3	70	30	35	0	70
Conservé hors traité	0	0	0	45	0	0	0	80
Total conservé	50	50	50	95	50	50	40	130
Total réassuré	48	70	3	70	30	35	0	70

FIGURE 1.2: Excédent de plein avec rétention de 50 et plein de souscription à 120



L'excédent de plein permet à l'assureur d'augmenter significativement sa capacité de souscription. L'autre avantage de cette forme de réassurance proportionnelle est qu'elle permet à l'assureur d'être libre

dans la gestion des risques inférieurs à son plein de rétention qu'il peut garantir seul. Cependant, ce type de réassurance n'est pas possible dans le cas d'une exposition illimitée, comme cela peut-être le cas en responsabilité civile automobile. De plus, l'assureur peut tout de même être affecté en cas de sinistre très important.

Il est également possible de mixer ces deux formes de couvertures.

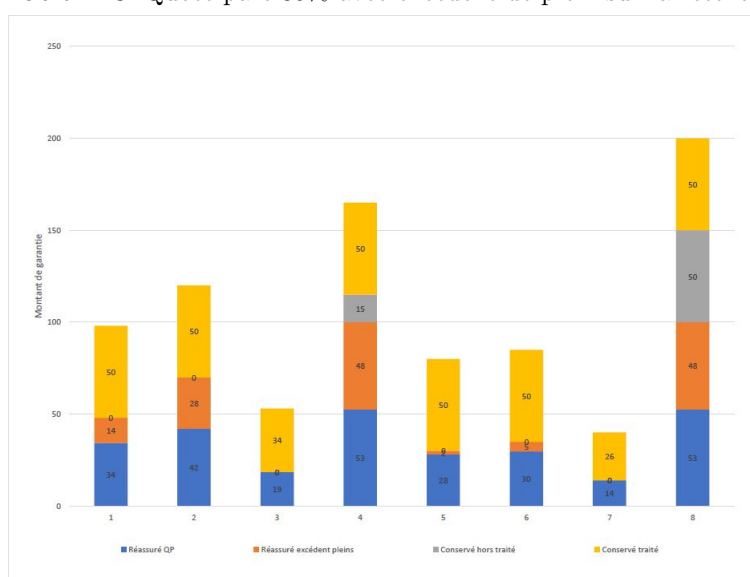
Quote-part avec excédent de plein sur la rétention

Considérons une quote-part à 35% avec un excédent de plein sur la rétention dont le plein de rétention est de 50. Le plein de souscription global est fixé à 150.

TABLE 1.5: Quote-part 35% avec excédent de plein sur la rétention

Assuré	1	2	3	4	5	6	7	8
Montant de garantie	98	120	53	165	80	85	40	200
Couvert dans le traité	98	120	53	150	80	85	40	150
Réassuré par la QP	34	42	19	53	28	30	14	53
Soumis à l'excédent de plein	64	78	34	98	52	55	26	98
Réassuré par l'excédent de plein	14	28	0	48	2	5	0	48
Conservé hors traité	0	0	0	15	0	0	0	50
Conservé dans traité	50	50	34	50	50	50	26	50
Total conservé	50	50	34	65	50	50	26	100
Total réassuré	48	70	19	100	30	35	14	100

FIGURE 1.3: Quote-part 35% avec excédent de plein sur la rétention



Les clauses affectant les traités proportionnels

Les traités proportionnels sont souvent assortis de différentes clauses. On peut citer le taux de commission fixe ou à échelle. Dans le deuxième cas, plus le ratio sinistres sur primes est élevé, moins le taux de commission servi par le réassureur est important. A l'inverse, un ratio combiné faible pour l'assureur est récompensé par une commission de réassurance plus importante.

Il existe également la participation aux sinistres (*loss corridor*) où au-delà d'un certain niveau du ratio combiné, la cédante intervient pour assumer partiellement ou en totalité la sinistralité associée.

Il peut aussi y avoir un report de pertes (*loss carry forward*). En cas de mauvais résultat une année, le résultat est reporté l'année suivante pour éviter que l'assureur ne paye une commission de réassurance élevée.

Une clause de participation aux bénéfices peut aussi figurer dans les traités proportionnels.

1.1.4 La réassurance non-proportionnelle

La réassurance non-proportionnelle ne repose pas sur les risques couverts mais sur les sinistres réalisés. Il n'existe plus aucune proportionnalité entre la prime d'assurance et la prime de réassurance d'une part et les sinistres originaux et l'engagement du réassureur dans les montants de sinistres d'autre part.

Il convient de distinguer deux formes de réassurance non-proportionnelle :

- Les traités limitant le reste à charge sur un sinistre défini : l'excédent de sinistre (*excess of loss*).
- Les traités limitant le reste à charge sur tous les sinistres confondus : l'excédent de perte (*stop loss*).

L'excédent de sinistre

Dans un traité en excédent de sinistre, noté XS , le réassureur s'engage à payer la partie du sinistre supérieur à un certain montant appelé *franchise* (D) dans la limite d'une certaine couverture qui s'exprime par la *portée* (C) et qui représente la limite maximale d'engagement pour le sinistre. La couverture associée est notée $C - XS - D$. Il est possible que la portée soit infinie, c'est le cas des branches responsabilité civile automobile.

Il existe trois formes de traité XS :

1. L' XS par risque : les conditions du traité s'appliquent à chacune des polices en portefeuille. Chaque police représente un risque.
2. L' XS par événement : les conditions du traité s'appliquent à la perte totale provoquée par la réalisation d'un même événement touchant plusieurs polices du portefeuille. La notion d'événement doit être clairement définie dans le traité. L' XS par événement est souvent utilisé pour se couvrir contre les catastrophes naturelles.
3. L' XS par risque et par événement : prend en compte les deux types de sinistralité.

Mathématiquement, notons X le montant d'un sinistre survenu au cours de l'année. Considérons un traité $C - XS - D$. Le montant à charge du réassureur S s'exprime comme :

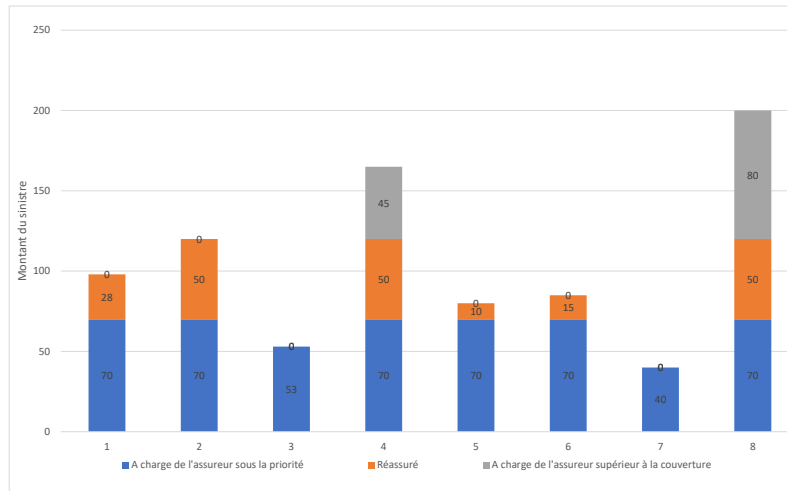
$$S = \min \left(\max \left(X - D; 0 \right); C \right)$$

La réassurance en excédent de sinistre est illustrée sur le portefeuille fictif en considérant un $50 - XS - 70$.

TABLE 1.6: Excédent de sinistre $50 - XS - 70$

Assuré	1	2	3	4	5	6	7	8
Montant du sinistre	98	120	53	165	80	85	40	200
Réassuré dans l' XS	28	50	0	50	10	15	0	50
A charge de l'assureur sous la priorité	70	70	53	70	70	70	40	70
A charge de l'assureur supérieur à la couverture	0	0	0	45	0	0	0	80

FIGURE 1.4: Excédent de sinistre 50 XS 70



Le traité en excédent de sinistre présente certains avantages. Contrairement au traité proportionnel en excédent de plein couvrant des sommes assurées, l'XS ne couvre pas les petits sinistres ce qui permet de limiter le volume de prime cédée à la réassurance. La couverture XS permet une protection contre les grands sinistres de pointe. La gestion du traité est facilitée par la prime fixée lors du renouvellement, il suffit de suivre les sinistres excédant la priorité. Le réassureur a la possibilité de limiter son engagement en introduisant un AAL (*Annual Aggregate Limit*). De même que l'assureur peut limiter le coût de l'XS en absorbant les premières pertes grâce à un AAD (*Annual Aggregate Deductible*).

Le principal inconvénient de l'XS est la volatilité du prix qui est fonction de l'expérience de l'assureur. En effet, l'estimation de la prime d'assurance en excédent de sinistre peut parfois être difficile, en particulier pour les tranches de réassurance n'ayant jamais été touchées dans le passé. Cette problématique sera développée dans la dernière section de ce chapitre.

L'excédent de perte

L'autre forme de réassurance non-proportionnelle est le traité en excédent de perte (*stop-loss*). La sinistralité rattachée au traité n'est pas calculée affaire par affaire, mais globalement sur l'ensemble du portefeuille. La réassurance prend en compte l'ensemble des sinistres d'un exercice qui dépasse un pourcentage déterminé des primes qui y sont afférentes. C'est la sinistralité annuelle qui est prise en compte et non un sinistre individuel. Le traité en excédent de perte permet à l'assurance de se couvrir contre une fréquence anormalement élevée de sinistres. En effet, un nombre important de petits sinistres peut être aussi dangereux pour l'assurance que la réalisation de quelques gros sinistres. La réassurance en excédent de perte se rapproche de l'excédent de sinistre, la différence est que la franchise et la portée du traité sont souvent exprimées en pourcentage de l'encaissement de la cédante.

Mathématiquement, en notant d la priorité du traité exprimée en pourcentage, c la portée du traité exprimée en pourcentage. Soit $\frac{S}{P}$ le rapport sinistres à primes de l'année avec S la somme des sinistres de la cédante et P la somme des primes acquises par la cédante. La couverture du réassureur s'exprime comme :

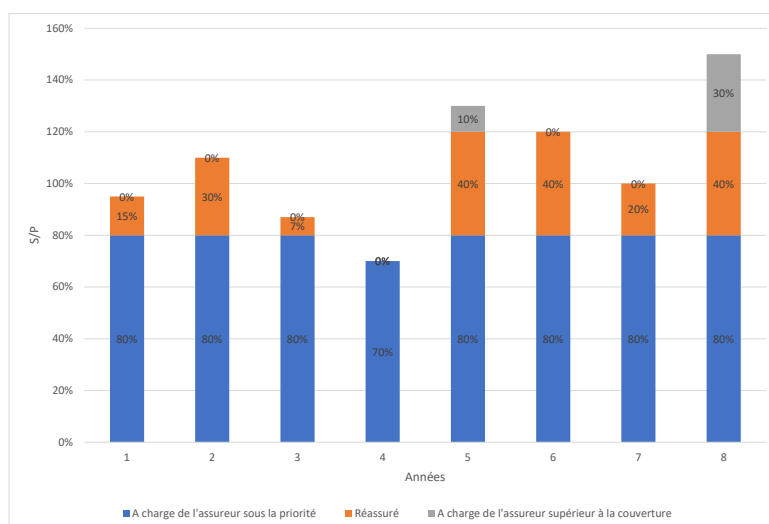
$$\min \left(\max \left(0, \frac{S}{P} - d \right), c \right) \times P$$

La réassurance en excédent de perte est illustrée sur un 40% XS 80 %.

TABLE 1.7: Excédent de perte 40% XS 80 %

Année	1	2	3	4	5	6	7	8
S/P	95%	110%	87%	70%	130%	120%	100%	150%
Conservé traité sous la priorité	80%	80%	80%	70%	80%	80%	80%	80%
Réassuré	15%	30%	7%	0%	40%	40%	20%	40%
Conservé traité au-delà de la portée	0%	0%	0%	0%	10%	0%	0%	30%

FIGURE 1.5: Excédent de perte 40% XS 80 %



Le traité en excédent de perte peut être utilisé pour remplacer l'excédent de sinistre par événement lorsque celui-ci est difficile à définir précisément. C'est le cas pour la sécheresse dans le domaine agricole, ou bien la grêle. La gestion d'un traité en excédent de perte est simple, il suffit de suivre la sinistralité globale de la cédante. Ce type de couverture est idéal pour l'assureur qui souhaite protéger son résultat annuel. Cependant, l'excédent de perte est soumis à l'aléa moral. Lorsque la priorité est atteinte, certaines cédante pourraient être laxistes dans la gestion des sinistres, se sachant couvertes par le *stop-loss* à hauteur de la portée.

Les clauses affectant les traités non-proportionnels

Certaines clauses de réassurance peuvent être ajoutées dans un traité. Ces clauses sont le fruit des négociations entre les assureurs et réassureurs. Les clauses les plus courantes sont présentées brièvement.

Clause de stabilité Cette clause s'applique essentiellement au traité en excédent de sinistre couvrant des branches à développement long où la durée entre la date de survenance et la date du paiement du sinistre peut être de plusieurs années. La clause permet de faire évoluer la franchise et la portée du traité pour partager l'inflation future entre le réassureur et l'assureur. Ce dernier pourrait avoir tendance à traîner dans le règlement des sinistres, laissant la valeur du sinistre augmenter du fait de l'inflation et ainsi être défavorable à la réassurance.

Franchise annuelle Elle est communément appelée AAD pour *Annual Aggregate Deductible*. Cette clause permet à l'assureur de diminuer la prime d'assurance qui décide de supporter la charge liée à la réalisation des premiers sinistres. Une fois l'AAD entièrement consommé, le traité de réassurance fonctionne normalement.

Limite annuelle Elle est communément appelée AAL pour *Annual Aggregate Limit*. Elle permet au réassureur de limiter son engagement maximal au montant fixé de l'AAL. Cela fait diminuer le coût de la réassurance.

Clause de reconstitution Il arrive que la limite annuelle s'exprime sous la forme d'un multiple n de la portée. On dit qu'il y a n reconstitutions, c'est-à-dire que la garantie peut être rétablie n fois. Ces reconstitutions peuvent être gratuites ou payantes. La reconstitution s'applique dès qu'une partie de l'engagement du réassureur a été utilisé. Dans le cas d'une reconstitution payante, l'assureur verse une prime de reconstitution calculée au prorata de la portion de couverture absorbée par la tranche.

La section suivante présente le groupe Covéa et plus particulièrement le pôle d'acceptation en réassurance non-vie.

1.2 Le groupe Covéa

Le groupe Covéa est un groupe d'assurance mutualiste régit par le code des assurances. Le groupe est reconnu à travers ses trois marques : MAAF, MMA et GMF.

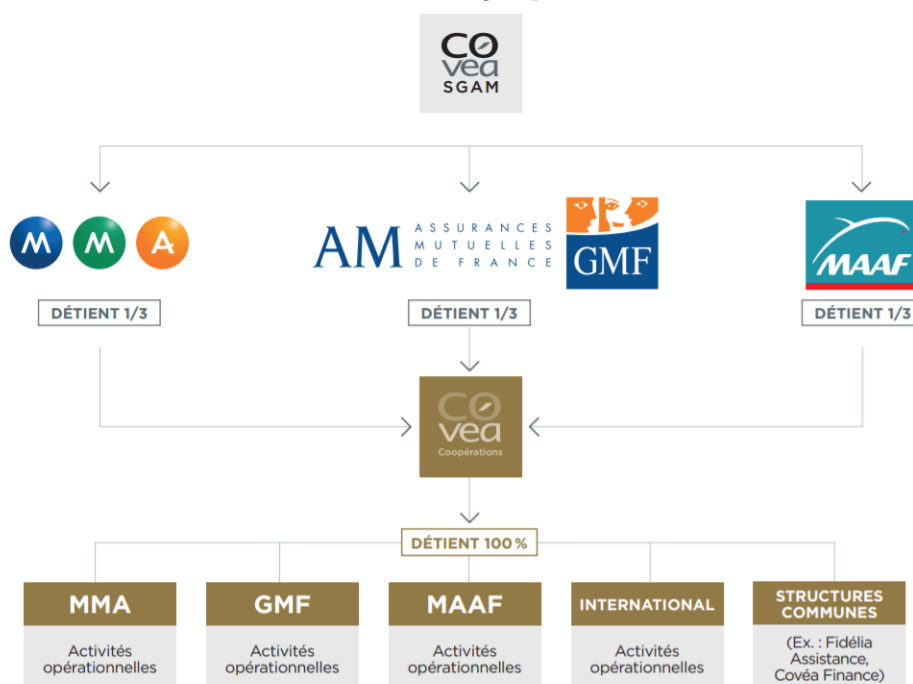
1.2.1 L'histoire du groupe

L'urbanisation et l'industrialisation d'une part, le déclin de la solidarité familiale d'autre part, favorisent l'essor remarquable des compagnies d'assurance et des mutuelles au cours du 19ème siècle et du début du 20ème siècle. De cette période, Covéa a hérité de trois grandes marques reconnues aujourd'hui.

C'est en 1828 que Les Mutuelles du Mans Assurances (MMA) sont créées. S'en suit en 1934 la création de la Garantie Mutuelle des Fonctionnaires (GMF) puis la création de la Mutuelle d'Assurance des Artisans de France (MAAF) en 1950.

C'est en 2003 que MAAF, MMA et GMF s'unissent pour créer la SGAM Covéa. Cette création permet la mise en place d'une stratégie de groupe commune et donne naissance à des liens de solidarité forts.

FIGURE 1.6: Le groupe Covéa



Le groupe Covéa est un leader sur le marché français avec près d'un ménage sur trois assuré auprès d'une des trois marques. Le groupe propose des assurances de biens et responsabilité avec notamment 10,7 millions de véhicules et 8,1 millions d'habitations assurés en France au 31 décembre 2020. Le groupe est présent en santé et prévoyance avec près de 3 millions de bénéficiaires en santé au 31 décembre 2020. Le groupe est également présent en assurance vie en proposant des produits d'épargne et de retraite.

En plus de ces activités classiques d'assurance, le groupe a développé l'activité de réassurance.

1.2.2 La réassurance non-vie chez Covéa

La réassurance non-vie du groupe Covéa est gérée à la Direction Réassurance Acceptations Internationales (DRAI) située sur le site de Strasbourg. Le département de réassurance non-vie est composé d'une quarantaine de collaborateurs répartis entre trois pôles :

- Le pôle commercial, où les souscripteurs ont pour mission de développer le portefeuille et échangent directement avec les courtiers et les cédantes pour négocier et recueillir les informations nécessaires au renouvellement des traités.
- Le pôle de pilotage économique et performance, où les comptables enregistrent les flux sortants et entrants des différents traités.
- Le pôle tarification et actuariat, composé de 7 personnes en charge des différentes modélisations et de la tarification de tous les traités de réassurance.

Au premier trimestre 2021, le portefeuille de réassurance se compose de 1 756 traités souscrits dans 41 pays auprès de 281 cédantes.

FIGURE 1.7: Répartition des branches de réassurance au sein de la DRAI

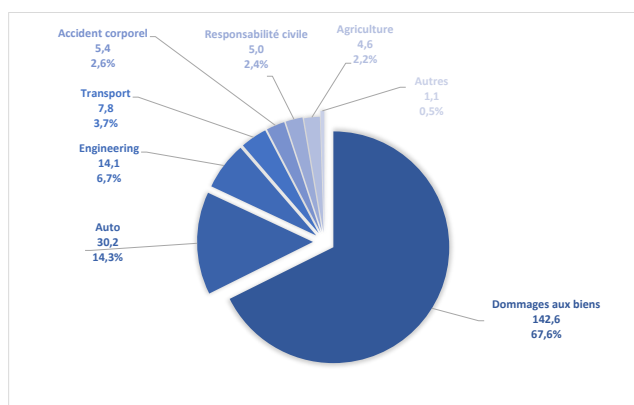
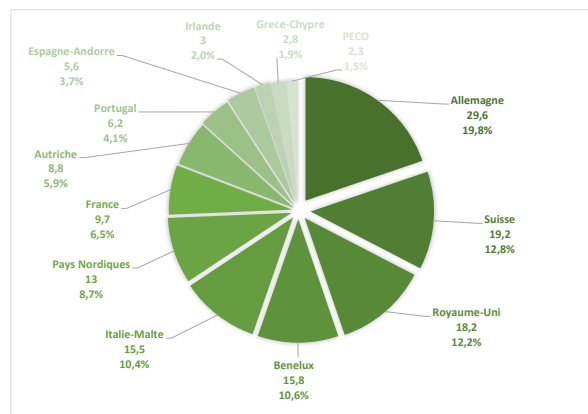


FIGURE 1.8: Répartition des pays Européens au sein de la DRAI



Les chiffres d'affaire sont exprimés en millions d'euros.

Les branches de réassurance les plus importantes sont la couverture des dommages aux biens et l'automobile. La direction de réassurance travaille également avec des pays hors de l'Europe tel que l'Afrique du Sud, le Canada, le Japon etc.

1.3 Problématique de l'étude : la modélisation de la sévérité des traités en excédent de sinistre

En réassurance non-proportionnelle de type excédent de sinistre, le réassureur doit estimer la sinistralité future d'une cédante à partir des informations passées. Pour ce faire, l'actuaire reçoit des informations de renouvellement dans lesquelles, entre autres, un historique de sinistres est disponible. Dans le cas de tranches travaillantes, il existe des méthodes simples et intuitives pour estimer un taux de prime. En revanche, pour les tranches hautes n'ayant jamais été touchées par des sinistres, il est malaisé d'estimer un prix. L'actuaire peut alors utiliser des distributions probabilistes. L'objectif de cette section est de présenter le contexte de la tarification de traités XS et de développer la problématique du mémoire.

1.3.1 Notion de branches courtes et branches longues

Il est possible de scinder les branches d'assurance en deux catégories selon la durée qui sépare la date de survenance d'un sinistre, la date de déclaration de ce dernier et la date de son règlement.

Les branches courtes

Ce sont les branches d'assurances pour lesquelles la durée entre la date de déclaration d'un sinistre, l'estimation du montant à indemniser et le règlement effectif est courte. Cette durée n'excède en général pas plus d'une année. Les branches courtes sont celles liées à la propriété, les plus connues sont les branches couvrant l'incendie ou le vol par exemple. Pour tarifier convenablement ces branches, l'actuaire a besoin entre 5 et 7 années d'historique.

Les branches longues

Ce sont les branches d'assurance pour lesquelles la durée entre la date de survenance, de déclaration, l'estimation et le règlement d'un sinistre est longue. Cette durée peut être de plusieurs années, ne serait-ce qu'entre la date de survenance et la date de déclaration. C'est notamment le cas des branches

responsabilité civile. Ces branches font l'objet d'un traitement spécifique par l'actuaire car l'estimation du coût à la date de déclaration peut être très différente de l'indemnisation effective qui sera payée quelques années plus tard. L'actuaire a recours à des triangles de liquidation pour estimer la charge ultime du règlement. Des provisions sont donc constituées et suivies tout le long du processus de gestion du sinistre. Pour tarifier convenablement ces branches, l'actuaire a besoin d'un minimum de 8 années d'historique.

1.3.2 Notion de tranches travaillantes ou non travaillantes

Très souvent, les programmes de réassurance en excédent de sinistre sont scindés en plusieurs tranches de réassurance. La raison principale qui explique ce découpage est l'appétence au risque des réassureurs. Certains préfèrent offrir des couvertures sur les tranches basses fréquemment sinistrées (tranches travaillantes) tandis que d'autres préfèrent se focaliser dans la souscription de part sur des tranches hautes (tranches non travaillantes).

Une tranche non travaillante est tout ou partie d'un programme de réassurance dont la probabilité d'occurrence d'un sinistre est faible. La notion de tranche non travaillante est subjective, plusieurs facteurs permettent de suspecter qu'une tranche est non travaillante :

- La sinistralité est faible ou inexistante sur le passé.
- Le rapport entre la prime de réassurance et la couverture est faible. C'est le *rate on line* (ROL). Lorsque que le ROL est faible, cela signifie que la tranche a une faible chance d'être complètement traversée par un ou plusieurs sinistres, autrement dit le tarif proposé est bas au regard de l'engagement associé à la tranche.
- La priorité est élevée.

En tarification, le calcul de la prime pure des tranches non travaillantes est le plus complexe car l'actuaire dispose de peu de données pour l'estimer.

1.3.3 Contexte du renouvellement

La plupart des traités de réassurance ont une période d'effet d'un an et se renouvellent chaque année au premier janvier en Europe (au premier juillet au Moyen-Orient et au premier avril au Japon). La période de renouvellement la plus importante est celle s'étalant d'octobre à fin décembre, c'est durant cette période que la majorité du chiffre d'affaire annuel est générée. De fait, cette période est marquée par un surcroît de travail intense pour les réassureurs et les courtiers de réassurance.

Dans la pratique, les réassureurs sont approchés par des cabinets de courtage en réassurance ou directement par des cédantes pour effectuer des cotations actives ou de contrôle. Le réassureur indique également la part du programme qu'il est prêt à souscrire. Les cédantes reçoivent ensuite plusieurs propositions de prix et décident de placer l'affaire chez un ou plusieurs réassureurs. Il y a la désignation d'un réassureur apériteur, souvent celui dont le taux a été retenu ou dont la part souscrite est majoritaire.

Une fois le traité de réassurance signé, c'est le souscripteur qui prend en charge sa gestion jusqu'à l'extinction des polices couvertes du portefeuille.

1.3.4 Cotation

La prime de réassurance payée par la cédante pour la couverture accordée par le réassureur fait l'objet d'une tarification appelée *cotation*. En réassurance non-proportionnelle de type excédent de sinistre, la cotation permet de fixer un prix pour une tranche donnée. Pour effectuer une cotation, l'actuaire dispose des informations de renouvellement de la cédante, en particuliers des statistiques historiques lui sont transmises. En général, parmi ces statistiques il y a l'historique de l'encaissement de la cédante, une mesure d'exposition (en nombre ou en somme assurée) du portefeuille à réassurer et surtout l'historique sur plusieurs années de la sinistralité ayant touché le traité dans le passé. Une des difficultés est la forme et

le contenu que ces informations peuvent prendre selon les cédantes et les pays. Il n'y pas d'informations de renouvellement "types" communes à toutes les cédantes. L'actuaire doit s'adapter aux spécificités de chaque assureur. Il arrive parfois que l'information transmise soit incomplète, l'actuaire doit échanger avec le courtier de l'affaire pour obtenir les données nécessaires afin de pouvoir effectuer une cotation convenable.

In fine, les informations de renouvellement permettent à l'actuaire de modéliser la sinistralité de la cédante et d'estimer un tarif de réassurance tout en prenant en compte l'incertitude autour des estimations effectuées liées aux informations qui peuvent être en faible quantité.

1.3.5 Limites de la tarification à l'expérience : le *burning cost*

Une méthode classique de tarification de traités XS est la méthode du *burning cost*. Cette méthode repose sur les sinistres passés à partir desquels un taux pur d'XS est estimé. Pour utiliser cette méthode, l'actuaire a besoin d'un historique de sinistres et de primes sur plusieurs années, selon la branche à tarifier. Les primes et les sinistres sont ensuite revalorisés à l'année à tarifier pour pouvoir les comparer indépendamment de leur année d'origine : c'est la mise en *as-if*. Cependant la méthode du *burning cost* présente certaines limites, notamment lors de l'estimation du coût à charge de la réassurance pour des tranches non travaillantes.

La méthode du *burning cost*

La méthode du *burning cost* permet d'estimer un taux pur de réassurance en calculant le rapport entre les sinistres à charge du réassureur et les assiettes de primes sur le nombre d'années d'historique disponibles.

Mathématiquement, considérons une tranche C XS D, sans AAD, sans AAL, en supposant les reconstitutions gratuites et illimitées, avec les caractéristiques suivantes :

- n : le nombre d'années d'historique disponibles pour la tarification.
- $(N_j)_{1 \leq j \leq n}$: le nombre de sinistres survenus l'année j .
- $(X_{i,j})_{1 \leq i \leq N_j \text{ et } 1 \leq j \leq n}$: le montant *as-if* du sinistre i survenu durant l'année j .
- $(S_j)_{1 \leq j \leq n}$: la charge totale supportée par la réassurance pour l'année j .
- $(EPI_j)_{1 \leq j \leq n}$: l'assiette de primes *as-if* de l'année j .

Il est possible d'exprimer le montant à charge de la réassurance pour chacune des années $j \in \llbracket 1 ; n \rrbracket$ par :

$$S_j = \sum_{i=1}^{N_j} \min(\max(X_{i,j} - D ; 0) ; C)$$

Le taux pur du *burning cost*, noté BC_{pur} , est la moyenne des montants à charge pondérée par l'assiette de primes *as-if* de l'année associée :

$$BC_{\text{pur}} = \frac{1}{n} \sum_{j=1}^n \frac{S_j}{EPI_j}$$

La prime de réassurance est obtenue en multipliant le taux pur par l'assiette de primes estimée pour l'année à tarifier. La méthode du *burning cost* représente finalement le ratio entre les pertes enregistrées par le réassureur et les encaissements de la cédante sur une période de plusieurs années. Cette méthode est simple mais présente une limite importante qui est illustrée à travers le cas pratique suivant.

Cas pratique pour exposer les limites de la méthode

Pour illustrer une des limites de la tarification par la méthode du *burning cost*, considérons l'exemple fictif suivant où une cédante dispose de 9 années d'expérience. La cédante souhaite se couvrir avec un programme XS par risque ayant une première tranche T_1 : 2 000 000 XS 1 000 000, et une tranche T_k : 30 000 000 XS 15 000 000.

Pour tarifier ce programme, l'actuaire dispose des montants de sinistres revalorisés et des encaissements de la cédante revalorisés. Dans la suite, l'encaissement de la cédante est noté EPI pour *Earn Premium Income* en anglais.

TABLE 1.8: Calcul du taux *burning cost* pour la tranche 2 000 000 XS 1 000 000

Année	Sinistres	Charge annuelle du sinistre	Charge annuelle cumulée	EPI de l'année	Taux BC annuel
2010	2 000 000 1 500 000	1 000 000 500 000	1 500 000	75 140 635	1,996%
2011	700 000 1 400 000	0 400 000	400 000	80 562 254	0,497%
2012	2 900 000	1 900 000	1 900 000	82 014 789	2,317%
2013	1 600 000 500 000 1 250 000	600 000 0 250 000	850 000	81 562 247	1,042%
2014	1 000 000 1 500 000 2 200 000 1 260 000	0 500 000 1 200 000 260 000	1 960 000	84 584 012	2,317%
2015	800 000	0	0	85 047 896	0,000%
2016	1 150 000 1 600 000	150 000 600 000	750 000	88 047 963	0,852%
2017	1 350 000 950 000 1 030 000 1 180 000 2 250 000	350 000 0 30 000 180 000 1 250 000	1 810 000	92 546 870	1,956%
2018	1 960 000 1 200 000 1 750 000 650 000	960 000 200 500 750 000 0	1 910 500	95 358 720	2,003%
2019	1 150 000 2 485 000	150 000 1 485 000	1 635 000	95 358 720	1,599%

La tranche T_1 est une tranche travaillante. Chaque année, au moins un sinistre atteint la priorité de 1 000 000€, sauf en 2015 où le sinistre le plus important vaut 800 000€. Pour chacune des années, le montant à charge de chaque sinistre est calculé. Ces montants sont ensuite additionnés annuellement pour obtenir la charge annuelle cumulée. Cette charge annuelle cumulée est divisée par l'EPI de l'année pour obtenir le taux du *burning cost* associé.

Le taux pur calculé par la méthode vaut 1,458%. En supposant que l'EPI 2020 de la cédante vaut 100 459 896€, la prime de réassurance associée à la tranche 2 000 000 XS 1 000 000 vaut 1 464 583€. Cependant, on remarque que le montant maximum d'un sinistre à charge de la réassurance vaut 1 900 000€, c'est le sinistre de 2012. Autrement dit, la tarification effectuée est celle d'une tranche 1 900 000 XS 1 000 000 car il n'y a pas de données observée au-delà de cette portée. La tranche n'est jamais traversée entièrement, la méthode par expérience ne permet pas de fournir un prix pour l'excédent de couverture 100 000 XS 2 900 000 qui n'a jamais été touché.

Ce constat est accentué lorsque l'actuaire souhaite tarifier une tranche non-travaillante.

TABLE 1.9: Calcul du taux *burning cost* sur une tranche 30 000 000 XS 15 000 000

Année	Sinistres	Charge annuelle du sinistre	Charge annuelle cumulée	EPI de l'année	Taux BC annuel																																																																																				
2010	2 000 000	0	0	75 140 635	0,000%																																																																																				
	1 500 000	0				2011	700 000	0	0	80 562 254	0,000%	1 400 000	0	2012	2 900 000	0	0	82 014 789	0,000%	2013	1 600 000	0	0	81 562 247	0,000%	500 000	0	1 250 000	0	2014	1 000 000	0	0	84 584 012	0,000%	1 500 000	0	2 200 000	0	1 260 000	0	2015	800 000	0	0	85 047 896	0,000%	2016	1 150 000	0	0	88 047 963	0,000%	1 600 000	0	2017	1 350 000	0	0	92 546 870	0,000%	950 000	0	1 030 000	0	1 180 000	0	2 250 000	0	2018	1 960 000	0	0	95 358 720	0,000%	1 200 000	0	1 750 000	0	650 000	0	2019	1 150 000	0	0	95 358 720	0,000%	2 485 000	0
2011	700 000	0	0	80 562 254	0,000%																																																																																				
	1 400 000	0				2012	2 900 000	0	0	82 014 789	0,000%	2013	1 600 000	0	0	81 562 247	0,000%	500 000	0		1 250 000	0				2014	1 000 000	0	0		84 584 012	0,000%				1 500 000	0	2 200 000	0	1 260 000	0	2015	800 000	0	0	85 047 896	0,000%	2016	1 150 000	0	0	88 047 963	0,000%	1 600 000	0		2017	1 350 000				0	0	92 546 870	0,000%	950 000	0	1 030 000	0		1 180 000	0				2 250 000	0	2018	1 960 000	0	0	95 358 720	0,000%	1 200 000	0	1 750 000	0	650 000	0
2012	2 900 000	0	0	82 014 789	0,000%																																																																																				
2013	1 600 000	0	0	81 562 247	0,000%																																																																																				
	500 000	0																																																																																							
	1 250 000	0																																																																																							
2014	1 000 000	0	0	84 584 012	0,000%																																																																																				
	1 500 000	0																																																																																							
	2 200 000	0																																																																																							
	1 260 000	0																																																																																							
2015	800 000	0	0	85 047 896	0,000%																																																																																				
2016	1 150 000	0	0	88 047 963	0,000%																																																																																				
	1 600 000	0																																																																																							
2017	1 350 000	0	0	92 546 870	0,000%																																																																																				
	950 000	0																																																																																							
	1 030 000	0																																																																																							
	1 180 000	0																																																																																							
	2 250 000	0																																																																																							
2018	1 960 000	0	0	95 358 720	0,000%																																																																																				
	1 200 000	0																																																																																							
	1 750 000	0																																																																																							
	650 000	0																																																																																							
2019	1 150 000	0	0	95 358 720	0,000%																																																																																				
	2 485 000	0																																																																																							

Cette tranche n'est jamais touchée par un sinistre. La méthode du *burning cost* fournit des taux nuls. Néanmoins, à partir du moment où il existe une exposition sur cette tranche, c'est-à-dire des risques ou cumul de risques pouvant dépasser 15 000 000 €, le coût de la couverture de ces risques est non nul. C'est la limite de la méthode, elle ne permet pas de calculer un prix de réassurance pour les tranches non-travaillantes. De plus, cette tranche est très haute par rapport au montant du sinistre maximum observé. Pour estimer la prime pure d'une tranche non-travaillante, l'actuaire utilise les modèles probabilistes en modélisant d'une part la fréquence, et d'autre part la sévérité des sinistres. Ce mémoire se concentre sur la modélisation de la sévérité des traités de réassurance en excédent de sinistre.

Enjeux pour le pôle tarification de la DRAI

Actuellement, les actuaires de la DRAI utilisent un outil de tarification XS implémenté en VBA sur Excel. Cet outil permet entre autres, d'estimer un taux pur de réassurance par l'approche probabiliste. Les montants des sinistres sont modélisés par une loi de Pareto lorsque les données ne sont pas en nombre suffisant pour utiliser la méthode du *burning cost*. L'inconvénient de l'outil est que seule cette loi est disponible pour modéliser la sévérité des tranches non travaillantes. La loi de Pareto est la loi classique en réassurance, elle permet de modéliser convenablement la majorité des traités XS. Cependant, il arrive que cette loi ne soit pas adaptée sans que d'autres ajustements ne soient disponibles. Le pôle de tarification souhaite pouvoir étendre la possibilité d'ajustement des montants de sinistres à d'autres lois de probabilité afin d'extrapoler l'information de la manière la plus juste et fournir un taux de réassurance représentatif du risque, en particulier pour les tranches hautes non travaillantes. L'objectif du mémoire est d'améliorer et d'optimiser l'outil de tarification déjà en place.

Pour répondre à cette problématique du point de vue technique, le nombre de lois de probabilité pour ajuster la sinistralité est étendu passant d'une seule loi à six. De plus, afin d'améliorer la tarification des tranches hautes non travaillantes, un modèle à deux lois fondé sur la théorie des valeurs extrêmes est proposé. La théorie des valeurs extrêmes offre un cadre mathématique solide qui justifie son utilisation pour estimer le risque sous-jacent aux tranches non travaillantes, parfois très hautes par rapport au montant maximum de sinistre observé. Cette théorie est développée dans le chapitre suivant.

Pour répondre à cette problématique du point de vue opérationnel, une application développée dans

le langage de programmation *R Shiny* est implémentée. Cette application permet de modéliser la sévérité des sinistres en proposant des méthodes de comparaison d'ajustement entre les lois. L'application permet également de modéliser la sinistralité par deux lois. Cette application se devait d'être facile d'utilisation pour les utilisateurs non initiés au langage R tout en permettant une modélisation rigoureuse. L'application terminale est déployée et est fonctionnelle pour le prochain renouvellement. Elle est présentée en annexe [B](#).

1.3.6 Prime pure, prime technique et prime commerciale

En réassurance, il est commun que la prime s'exprime sous la forme d'un taux. Ce taux est ensuite appliqué sur l'encaissement estimé de la cédante pour obtenir la prime. Le taux commercial effectif auquel la cédante est soumise repose sur un taux technique, lui-même basé sur un taux pur.

Le mémoire se concentre sur l'estimation du taux pur de réassurance, la prime pure associée reflète l'espérance des coûts de sinistres. Le taux pur est ensuite augmenté par différentes charges telles que :

- Le coefficient de sécurité, noté $c_s\%$, permet de couvrir la volatilité autour du taux pur liée aux écarts entre la sinistralité future observée et prédite. Cette volatilité est souvent la conséquence d'un faible nombre de données pour estimer les paramètres du modèle. Ajouté au taux pur, le coefficient de sécurité permet d'obtenir le taux technique.
- Les frais de gestion du traité, notés $f_g\%$, permettent de couvrir les frais engendrés par l'acceptation d'une affaire, comme par exemple l'enregistrement comptable des flux ou la gestion des sinistres.
- Les frais de courtage, notés $f_c\%$, permettent de rémunérer le courtier en réassurance qui est intervenu pour placer l'affaire.
- Les frais financiers, notés $f_r\%$, diminuent le taux et représentent les intérêts générés par le placement des provisions de primes et sinistres du portefeuille réassuré.

Mathématiquement, le taux commercial s'exprime ainsi :

$$\text{Taux commercial} = \underbrace{\text{Taux pur} \times (1 + c_s\%)}_{\text{Taux technique}} \times (1 + f_g\%) \times \frac{1}{1 - f_c\%} \times \frac{1}{1 + f_r\%}$$

Chapitre 2

La théorie des valeurs extrêmes

La tarification de traité XS consiste à calculer un coût moyen à charge entre deux valeurs qui sont la priorité et le plafond de la tranche du traité de réassurance. L'une des difficultés est que dans certains cas, la tranche de réassurance en question n'est pas travaillante, c'est-à-dire qu'il n'y a aucun sinistre de montant supérieur à la priorité dans l'historique de la cédante. L'actuaire qui s'attache à la tarification de cette tranche doit alors être en mesure d'estimer un coût moyen à charge entre des grandeurs qui appartiennent à la queue de distribution des sinistres, pour laquelle il ne dispose pas d'observations.

L'étude de la queue de distribution des montants de sinistres est cruciale pour appréhender le risque et mieux l'évaluer. La théorie des valeurs extrêmes permet de répondre à cette problématique de meilleure maîtrise des risques. En effet, alors que la théorie "classique" s'intéresse essentiellement à la partie dite centrale de la loi modélisant au mieux les sinistres, la théorie des valeurs extrêmes permet d'étudier les "grandes" valeurs, c'est-à-dire la queue de distribution de la loi. La théorie des valeurs extrêmes propose un cadre théorique solide pour l'étude de ces valeurs qualifiées d'extrêmes.

La théorie des valeurs extrêmes peut se scinder en deux parties : une première par une approche *maxima par bloc* où une loi des valeurs extrêmes est ajustée sur un échantillon de maximums. Une seconde partie par une approche *dépassement de seuil* où la loi de Pareto Généralisée est ajustée sur un échantillon d'observations supérieures à un certain seuil.

L'objectif de ce chapitre est de présenter les principaux résultats de la théorie des valeurs extrêmes qui seront ensuite appliqués à la problématique étudiée.

2.1 Convergence en loi du maximum d'un échantillon et loi des valeurs extrêmes

Dans la suite, on considère X_1, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées représentant les montants de sinistres fournis par une cédante. Elles ont la même loi que la variable aléatoire X de fonction de répartition F . L'échantillon ordonné par ordre croissant est noté $X_{(1)} \leq \dots \leq X_{(n)}$.

2.1.1 Théorie classique et loi du maximum

La théorie classique s'intéresse principalement à la variable aléatoire $S_n = \sum_{i=1}^n X_i$. En supposant que la variable X est intégrable, la loi faible des grands nombres assure la convergence en probabilité suivante :

$$\frac{S_n}{n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \mathbb{E}(X)$$

En supposant de plus que la variable X est de carré intégrable, le Théorème Centrale Limite permet d'établir la convergence en loi suivante :

$$\frac{S_n - n\mathbb{E}(X)}{\sqrt{n\mathbb{V}(X)}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

Le Théorème Centrale Limite permet d'obtenir un résultat sur le comportement de la somme normalisée de variables aléatoires et donc sur la partie centrale de la distribution.

En réassurance, il s'avère que les plus gros montants de sinistres gouvernent la somme des montants de sinistres. Dans le cadre de la modélisation de la sévérité dans les tranches non travaillantes, c'est la queue de distribution des montants de sinistres qui intéresse l'actuaire. La théorie des valeurs extrêmes permet d'étudier le maximum normalisé de l'échantillon. Le maximum de l'échantillon est la variable $X_{(n)}$, sa fonction de répartition est donnée par :

$$\mathbb{P}(X_{(n)} \leq x) = F^n(x)$$

Soit $x_F = \sup\{x \text{ tel que } F(x) < 1\}$ le point terminal de F . Lorsque n tend vers l'infini, pour $x < x_F$, $F^n(x) \rightarrow 0$ et pour $x \geq x_F$, $F^n(x) \rightarrow 1$. Autrement dit, le maximum converge en probabilité vers le point terminal de l'échantillon, mais ce résultat ne fournit pas d'informations importantes. C'est pourquoi dans la suite, la convergence en loi du maximum est étudiée.

2.1.2 Loi des valeurs extrêmes : *Generalized Extreme Value* (GEV)

Avant d'énoncer le théorème sur la loi des valeurs extrêmes, il convient de définir un domaine d'attraction.

Définition 3 (Domaine d'attraction). *Soit F une fonction de répartition et H une fonction de répartition non-dégénérée (c'est à dire une fonction de répartition qui n'est pas associée à une variable constante presque-sûrement). On dit que F appartient au domaine d'attraction de H (on note $F \in \mathcal{DA}(H)$) s'il existe deux suites de normalisation $a_n > 0$ et b_n telles qu'en tout point de continuité x de H ,*

$$\lim_{n \rightarrow +\infty} F^n(a_n x + b_n) = H(x) \Leftrightarrow \frac{X_{(n)} - b_n}{a_n} \xrightarrow{\mathcal{L}} Y$$

où la variable aléatoire Y admet H pour fonction de répartition.

Ainsi, en considérant $a_n > 0$ et b_n des constantes de normalisation, la théorie classique des valeurs extrêmes s'intéresse à la distribution limite H qui satisfait la convergence en loi suivante :

$$\mathbb{P}\left(\frac{X_{(n)} - b_n}{a_n} \leq x\right) \xrightarrow{\mathcal{L}} H(x), \quad \forall x$$

Le résultat fondamental de la théorie des valeurs extrêmes, exposé par Fisher-Tippett (1928) a permis de caractériser la famille de distributions à laquelle appartient H : *Generalized Extreme Value* (GEV). Ce théorème énoncé ci-dessous, est à rapprocher du Théorème Centrale Limite.

Théorème 1 (Théorème de Fischer-Tippett). *Soit F une fonction de répartition. Si $F \in \mathcal{DA}(G)$ où G est une fonction de répartition non-dégénérée, alors G est de même type que la fonction de répartition de la loi GEV notée H_γ définie pour tout x tel que $1 + \gamma x > 0$*

$$H_\gamma(x) = \begin{cases} \exp[-(1 + \gamma x)^{-\frac{1}{\gamma}}] & \text{si } \gamma \neq 0 \\ \exp[-e^{-x}] & \text{si } \gamma = 0 \end{cases}$$

La loi limite du maximum dépend donc d'un seul paramètre de forme γ , appelé indice des valeurs extrêmes et qui contrôle la forme de la queue de distribution :

- $\gamma > 0$: domaine d'attraction de Fréchet. Le point terminal est infini. La fonction de survie converge vers 0 à une vitesse polynomiale. Ce sont les distributions dites à queue épaisse.
- $\gamma = 0$: domaine d'attraction de Gumbel. Le point terminal est fini ou infini. La fonction de survie converge vers 0 à une vitesse exponentielle. Ce sont les distributions à queue légère.
- $\gamma < 0$: domaine d'attraction de Weibull. Le point terminal est fini. La fonction de survie converge vers 0 à une vitesses polynomiale. Ce sont les distributions à queue fine.

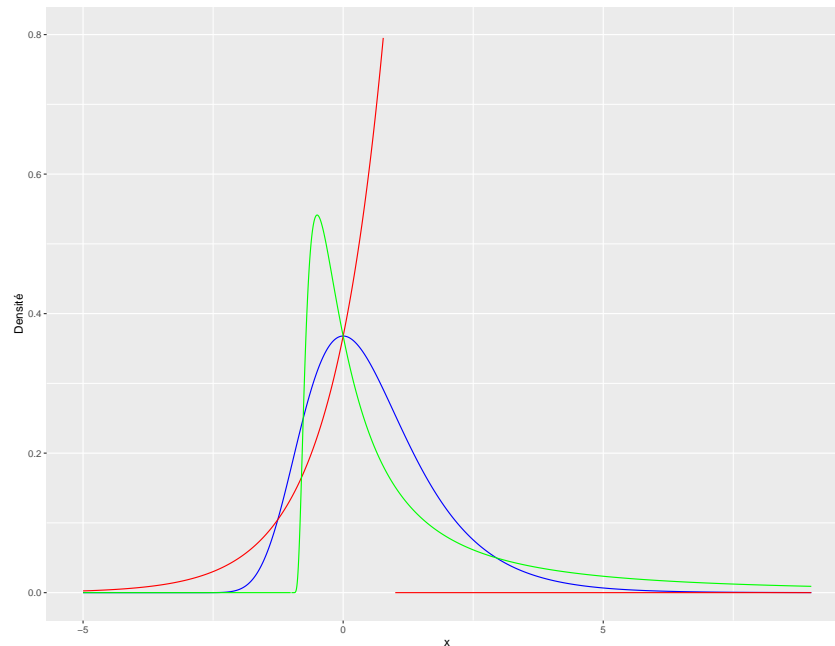
Le tableau suivant regroupe quelques lois usuelles en fonction de leur domaine d'attraction.

TABLE 2.1: Domaine d'attractions des lois usuelles

Fréchet	Gumbel	Weibull
Pareto	Normale	Uniforme
Log-gamma	Exponentielle	Beta
Student	Log-normale	Burr inverse
Burr	Gamma	
	Weibull	

Dans le contexte de réassurance, ce sont les lois à queue épaisse qui intéressent l'actuaire, c'est-à-dire les lois qui appartiennent au domaine d'attraction de Fréchet. Le graphique suivant illustre les densités d'une loi GEV pour un indice des valeurs extrêmes égal à -1 (courbe rouge), 0 (courbe bleue) et 1 (courbe verte).

FIGURE 2.1: Densité de la loi des valeurs extrêmes pour des valeurs différentes de l'indice des extrêmes



Remarque : Il existe une représentation à trois paramètres de la loi GEV, c'est la représentation de Jenkinson-Von-Mises. Elle se définit comme $H_{\gamma,\mu,\sigma}(x) = H_{\gamma}(\frac{x-\mu}{\sigma})$ où μ est un paramètre de position et σ un paramètre de dispersion.

Cette approche classique de la théorie des valeurs extrêmes qui s'appuie sur l'étude du maximum n'est pas assez générale dans le contexte de la réassurance. En effet, l'actuaire en charge de la tarification est intéressé par une information plus large que celle relative au maximum de l'échantillon. C'est pour ces problématiques qu'une seconde approche de la théorie des valeurs extrêmes a été développée. C'est

l'approche par *dépassement de seuil*. Elle n'est pas sans lien avec l'étude de la distribution du maximum et fait l'objet de la section suivante.

2.2 Théorie des excédents au-delà de seuils élevés

La section précédente présente la théorie des valeurs extrêmes par l'étude de la distribution du maximum d'un échantillon. Cette approche est trop restrictive pour l'actuaire en réassurance. En effet, l'actuaire préfère étudier l'ensemble de la queue de distribution des montants de sinistres, et ne pas se limiter à l'étude du maximum. L'approche POT (*Peaks Over the Threshold*) permet de répondre à ce besoin. Cette méthode repose sur la loi de Pareto Généralisée. Le théorème de Pickands, Balkema et De Hann permet de faire le lien avec la théorie des valeurs extrêmes classique qui étudie le comportement du maximum.

2.2.1 La loi des excès : méthode *Peaks Over the Threshold*

Pickands (1975) présente une théorie analogue à celle de la théorie classique des valeurs extrêmes. Cette théorie ne se base plus sur l'étude du maximum et s'intéresse plutôt au comportement asymptotique de la distribution des excès $X - u | X > u$ lorsque $u \rightarrow +\infty$.

Définition 4 (La loi des excès). *Soit X une variable aléatoire de fonction de répartition F . Pour tout seuil $u < x_F$, la loi des excès de X au dessus du seuil u est :*

$$F_u(x) = \mathbb{P}(X - u \leq x | X > u)$$

En développant l'expression de la fonction de répartition des excès au-delà du seuil u il vient que :

$$\begin{aligned} F_u(x) &= \frac{F(x+u) - F(u)}{1 - F(u)} \\ &= 1 - \frac{1 - F(x+u)}{1 - F(u)} \end{aligned}$$

En supposant l'approximation du maximum par une loi GEV de fonction de répartition $H_{\gamma, \mu, \sigma}$ grâce au théorème de Fisher-Tippett, on peut approcher :

$$F^n(x) \approx \exp\left[-\left(1 + \gamma\left(\frac{x - \mu}{\sigma}\right)^{-\frac{1}{\gamma}}\right)\right]$$

Puis :

$$\ln(F(x)) \approx -\frac{1}{n}\left(1 + \gamma\left(\frac{x - \mu}{\sigma}\right)^{-\frac{1}{\gamma}}\right)$$

Or on travail avec des valeurs de x élevées donc on a l'approximation $\ln(F(x)) \approx -(1 - F(x))$. Ainsi :

$$1 - F(x) \approx \frac{1}{n}\left(1 + \gamma\left(\frac{x - \mu}{\sigma}\right)^{-\frac{1}{\gamma}}\right)$$

En revenant à l'expression de $F_u(x)$ il vient que :

$$\begin{aligned} F_u(x) &\approx 1 - \left(\frac{1 + \gamma\left(\frac{x+u-\mu}{\sigma}\right)^{-\frac{1}{\gamma}}}{1 + \gamma\left(\frac{u-\mu}{\sigma}\right)^{-\frac{1}{\gamma}}}\right)^{-\frac{1}{\gamma}} \\ &\approx 1 - \left(1 + \frac{\gamma x}{\tilde{\sigma}}\right)^{-\frac{1}{\gamma}} \end{aligned}$$

Avec $\tilde{\sigma} = \sigma + \gamma(u - \mu)$. Cette expression est celle de la fonction de répartition d'une loi de Pareto Généralisée.

Autrement dit, la distribution de Pareto Généralisée fournit une bonne approximation de la loi des excès $X - u | X > u$

2.2.2 La distribution de Pareto Généralisée

Une variable aléatoire X suit une distribution Pareto Généralisée (GPD pour *Generalized Pareto Distribution*) de paramètres $\gamma \in \mathbb{R}$ et $\sigma > 0$, que l'on note $X \sim \mathcal{GPD}(\gamma, \sigma)$, si elle admet une fonction de densité de la forme :

$$\forall x > 0 \text{ et tel que } 1 + \frac{\gamma x}{\sigma} > 0, \quad g_{\gamma, \sigma}(x) = \begin{cases} \frac{1}{\sigma} (1 + \frac{\gamma x}{\sigma})^{-\frac{\gamma+1}{\sigma}} & \text{si } \gamma \neq 0 \\ \frac{1}{\sigma} e^{-\frac{x}{\sigma}} & \text{si } \gamma = 0 \end{cases}$$

La fonction de répartition associée s'exprime pour tout $x > 0$ et tel que $1 + \frac{\gamma x}{\sigma} > 0$ par :

$$G_{\gamma, \sigma}(x) = \begin{cases} 1 - (1 + \frac{\gamma x}{\sigma})^{-\frac{1}{\sigma}} & \text{si } \gamma \neq 0 \\ 1 - e^{-\frac{x}{\sigma}} & \text{si } \gamma = 0 \end{cases}$$

L'espérance d'une Pareto Généralisée est donnée par :

$$\forall \gamma < 1, \quad \mathbb{E}(X) = \frac{\sigma}{1 - \gamma}$$

La variance d'une Pareto Généralisée est donnée par :

$$\forall \gamma < \frac{1}{2}, \quad \mathbb{V}(X) = \frac{\sigma^2}{(1 - \gamma)^2(1 - 2\gamma)}$$

Remarque :

- Cette loi peut être étendue à 3 paramètres en ajoutant un paramètre de position μ . Il suffit de considérer $G_{\gamma, \sigma, \mu}(x) = G_{\gamma, \sigma}(x - \mu)$.
- Dans le cas où $\gamma = 0$, on retrouve une loi exponentielle de paramètre $\frac{1}{\sigma}$.
- La loi de Pareto Généralisée peut s'exprimer en fonction de la loi des valeurs extrêmes par : $G_{\gamma, \sigma}(x) = 1 + \ln[H_{\gamma}(\frac{x}{\sigma})]$.

2.2.3 Le théorème de Pickands, Balkema, De Hann

La loi de Pareto Généralisée peut donc être vue comme la distribution asymptotique des excédents au-delà d'un certain seuil qui tend vers l'infini. Le théorème de Pickands, Balkema et De Hann formalise ce résultat.

Théorème 2 (Théorème de Pickands, Balkema, De Hann). *La fonction de répartition F de point terminal x_F appartient au domaine d'attraction de H_{γ} si et seulement si il existe une fonction positive σ telle que :*

$$\lim_{x \rightarrow x_F} \sup_{0 < x < x_F - u} |F_u(x) - G_{\gamma, \sigma(u)}(x)| = 0$$

Ce résultat peut se traduire par le fait que la distribution GEV H_{γ} décrit le comportement du maximum de l'échantillon si et seulement si $G_{\gamma, \sigma(u)}$ décrit les excédents au-delà d'un seuil élevé. L'indice des valeurs extrêmes γ de la loi GPD est le même que pour la loi GEV, autrement dit le comportement de leur queue de distribution est identique.

Ce théorème suggère que pour un certain seuil u suffisamment grand, la loi de Pareto Généralisée est une bonne approximation de la loi des excès. La fonction de répartition $F_u(x - u)$ peut être approchée par $G_{\gamma, \sigma(u), u}$.

Finalement, en choisissant un seuil adéquat, une loi de Pareto Généralisée peut être ajustée aux données dépassant ce seuil. Dans notre contexte, ce sera aux montants des sinistres supérieurs à ce seuil.

Apparaît alors la problématique d'un choix de seuil optimal. Le problème d'un choix approprié du seuil u n'est pas aisé en pratique. En effet, soit u est trop grand et l'actuaire ne dispose pas de suffisamment de données pour l'estimation des paramètres du modèle. Soit u est trop petit et l'estimation

des paramètres est biaisée car l'actuaire utilise des données pour lesquelles le modèle n'est pas valide. Différentes méthodes pour sélectionner un seuil convenable seront présentées à partir de la page 35. La prochaine section présente des méthodes pour estimer les paramètres de la loi de Pareto Généralisée dans l'approche *Peaks Over the Threshold*.

2.3 Estimation paramétriques de la loi de Pareto Généralisée

Dans le cadre d'étude de ce mémoire, c'est l'approche *Peaks Over the Threshold* qui est retenue. Les paramètres de la loi de Pareto Généralisée doivent être estimés non pas à partir de l'échantillon initial X_1, \dots, X_n mais à partir de l'échantillon des données supérieures à un seuil u correspondant à la k_n valeurs de l'échantillon initial. L'échantillon des excès sur lequel les paramètres sont estimés est noté :

$$\forall i \in \llbracket 1; k_n \rrbracket, \quad Y_i = X_{(n-k_n+i)} - X_{(n-k_n)}$$

Cette section présente trois méthodes pour l'estimation des paramètres de la loi de Pareto Généralisée à partir de l'échantillon des excès Y_1, \dots, Y_{k_n} supposés indépendants et de même loi $\mathcal{GPD}(\gamma, \sigma)$.

2.3.1 La méthode du maximum de vraisemblance

Considérons que les variables aléatoires Y_1, \dots, Y_{k_n} soient indépendantes et identiquement distribuées. Le théorème de Pickands, Balkema et De Hann permet de supposer que ces excès suivent une loi de Pareto Généralisée. La méthode du maximum de vraisemblance permet d'estimer la valeur des paramètres γ et σ qui maximise la fonction de log-vraisemblance suivante :

$$l(\gamma, \sigma | y_1, \dots, y_{k_n}) = \begin{cases} -k_n \ln(\sigma) - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^{k_n} \ln\left(1 + \gamma \frac{y_i}{\sigma}\right) & \text{si } \gamma \neq 0 \\ -k_n \ln(\sigma) - \frac{1}{\sigma} \sum_{i=1}^{k_n} y_i & \text{si } \gamma = 0 \end{cases}$$

Les estimateurs du maximum de vraisemblance sont obtenus en déterminant la valeur des paramètres qui annule les dérivées partielles de la fonction de log-vraisemblance. Dans le cas où $\gamma = 0$, l'estimation du paramètre σ est donné par :

$$\hat{\sigma} = \frac{\sum_{i=1}^{k_n} y_i}{k_n}$$

Dans le cas où $\gamma \neq 0$, l'estimation des paramètres $\hat{\gamma}$ et $\hat{\sigma}$ s'obtient en résolvant le système d'équation suivant :

$$\begin{cases} -\frac{k_n}{\sigma} + \sum_{i=1}^{k_n} \frac{y_i(1+\gamma)}{\sigma^2 + \sigma\gamma y_i} = 0 \\ \sum_{i=1}^{k_n} \ln\left(1 + \frac{\gamma y_i}{\sigma}\right) + \gamma(1-\gamma) \sum_{i=1}^{k_n} \frac{y_i}{\sigma + \gamma y_i} = 0 \end{cases}$$

Les solutions de ce système ne sont pas explicites, des méthodes numériques itératives sont utilisées pour approcher la solution.

De plus, Smith (1984) a montré la normalité asymptotique de ces estimateurs dans le cas où $\gamma > -\frac{1}{2}$. La loi asymptotique des estimateurs est donnée par :

$$\sqrt{k_n} \begin{pmatrix} \hat{\sigma} \\ \hat{\gamma} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} \sigma \\ \gamma \end{pmatrix}, \begin{pmatrix} 2\sigma^2(1+\gamma) & -\sigma(1+\gamma) \\ -\sigma(1+\gamma) & (1+\gamma)^2 \end{pmatrix}\right)$$

Ce résultat permettra dans la suite de construire des intervalles de confiance pour l'estimation des quantiles. Dans notre étude, c'est cette méthode d'estimation qui est retenue. Cependant deux autres méthodes régulièrement utilisées sont présentées. C'est la méthode des moments et la méthode des moments pondérés.

2.3.2 La méthode des moments

Cette méthode est valable uniquement dans les cas où $\gamma < \frac{1}{2}$.

La méthode des moments consiste à égaliser les deux premiers moments théoriques de la loi de Pareto Généralisée avec les deux premiers moments empiriques. Il suffit ensuite de résoudre le système de deux équations à deux inconnues pour exprimer les paramètres de la loi en fonction des moments empiriques.

Dans le cas où une variable aléatoire X suit une loi de Pareto Généralisée de paramètres γ et σ , les deux premiers moments théoriques s'expriment pour tout $\gamma < \frac{1}{2}$ par :

$$m_1 = \mathbb{E}(X) = \frac{\sigma}{1 - \gamma}$$

$$m_2 = \mathbb{E}(X^2) = \frac{2\sigma^2}{(1 - \gamma)(1 - 2\gamma)}$$

En faisant coïncider ces moments théoriques avec les moments empiriques suivant :

$$\hat{m}_1 = \frac{1}{k_n} \sum_{i=1}^{k_n} y_i = \bar{Y}$$

$$\hat{m}_2 = \frac{1}{k_n} \sum_{i=1}^{k_n} y_i^2 = \hat{s}^2 + \bar{Y}^2$$

où \hat{s} est l'écart-type de l'échantillon des excès. Il vient comme estimateur des moments des paramètres de la distribution de Pareto Généralisée, sous réserve que $\gamma < \frac{1}{2}$:

$$\hat{\gamma} = \frac{1}{2} \left[1 - \frac{\bar{Y}^2}{\hat{s}^2} \right]$$

$$\hat{\sigma} = \bar{Y} \left[\frac{1 + \frac{\bar{Y}^2}{\hat{s}^2}}{2} \right]$$

2.3.3 La méthode des moments pondérés

Cette méthode, présentée par [Hosking et al. \(1985\)](#) est valable uniquement dans les cas où $\gamma < 1$. Dans un premier temps, la notion de moments pondérés est définie.

Définition 5 (Moments pondérés). *Soit Z une variable aléatoire de fonction de répartition F . Si Z est intégrable, le moment pondéré d'ordre $r \in \mathbb{N}$ et $s \in \mathbb{N}$ de Z est :*

$$WM_Z(r, s) = \mathbb{E}[ZF^r(Z)(1 - F(Z))^s]$$

Dans la suite, on considère $r = 0$. Le résultat suivant fournit la valeur du moment pondéré d'une variable aléatoire suivant une loi de Pareto Généralisée de paramètres γ et σ .

Théorème 3 (Moments pondérés d'une Pareto Généralisée). *Soit Z une variable aléatoire de loi de Pareto Généralisée de paramètres γ et σ . Dans le cas où $\gamma < 1$, on a pour tout $s \in \mathbb{N}$:*

$$WM_Z(0, s) = v_s = \frac{\sigma}{(s + 1)(s + 1 - \gamma)}$$

Les paramètres de la loi de Pareto Généralisée s'expriment en fonction des moments pondérés d'ordre 0 et 1 par :

$$\sigma = \frac{2v_0v_1}{v_0 - 2v_1} \quad \text{et} \quad \gamma = \frac{4v_1 - v_0}{2v_1 - v_0}$$

De la même façon que pour la méthode des moments classique, il est possible d'estimer v_0 et v_1 . De manière générale, le moment pondéré d'ordre s , noté v_s peut être estimé par la quantité suivante :

$$\hat{v}_s = \frac{1}{k_n} \sum_{i=1}^{k_n} Z_{i,k_n} \frac{(k_n - i + s)! k_n!}{(k_n - i)! (k_n + s)!}$$

Une estimation des paramètres γ et σ par la méthode des moments pondérés s'obtient en remplaçant v_0 et v_1 par la valeur de leur estimateur \hat{v}_0 et \hat{v}_1 .

2.4 Estimation semi-paramétrique de l'indice des valeurs extrêmes

Les trois méthodes décrites précédemment supposent que les variables aléatoires $(Y_i)_{1 \leq i \leq k_n}$ soient distribuées selon H_θ . La famille de lois H_θ est complètement paramétrique.

Les estimateurs présentés dans cette section s'appliquent à des échantillons dont la distribution n'est plus H_θ , mais qui appartient à un certain domaine d'attraction de H . C'est pourquoi cette approche est qualifiée de semi-paramétrique. Les trois estimateurs présentés sont basés sur la statistique d'ordre $X_{(1)} \leq \dots \leq X_{(n)}$, obtenue à partir de l'échantillon de sinistres initiaux en considérant les k_n plus grandes valeurs.

2.4.1 Estimateur de Hill

L'estimateur de Hill est l'estimateur le plus populaire pour estimer la valeur de l'indice des valeurs extrêmes γ . Cependant, il est valable uniquement pour les distributions appartenant au domaine d'attraction de Fréchet, c'est-à-dire dans le cas où $\gamma > 0$. Il se définit à partir d'un nombre d'excès considéré k_n par :

$$\hat{\gamma}_n^{(H)}(k_n) = \frac{1}{k_n} \sum_{i=1}^{k_n} \ln(X_{(n-i+1)}) - \ln(X_{(n-k_n)})$$

Si k_n est choisi de sorte que $\lim_{n \rightarrow +\infty} k_n = +\infty$ et $\lim_{n \rightarrow +\infty} \frac{k_n}{n} = 0$, on peut montrer que l'estimateur de Hill est asymptotiquement gaussien :

$$\sqrt{k_n} \left(\frac{\hat{\gamma}_n^{(H)}(k_n) - \gamma}{\gamma} \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

De plus, on peut montrer que l'estimateur de Hill est construit à partir de l'estimateur du maximum de vraisemblance d'une loi de Pareto de paramètre α de sorte que la relation $\alpha = \frac{1}{\gamma}$ est vérifiée.

2.4.2 Estimateur de Pickands

L'estimateur de Pickands présente l'avantage d'être valable pour les trois domaines d'attraction, il n'y a pas de restriction sur la valeur de l'indice des valeurs extrêmes γ . Il se définit à partir d'un certain nombre d'excès considéré k_n par :

$$\hat{\gamma}_n^{(P)}(k_n) = \frac{1}{\ln(2)} \ln \left(\frac{X_{(n-k_n)} - X_{(n-2k_n)}}{X_{(n-2k_n)} - X_{(n-4k_n)}} \right)$$

On peut également montrer qu'il est asymptotiquement gaussien :

$$\sqrt{k_n} \left(\frac{\hat{\gamma}_n^{(P)}(k_n) - \gamma}{\sigma(\gamma)} \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

$$\text{où } \sigma(\gamma) = \frac{\gamma \sqrt{2^{2\gamma+1} + 1}}{2(2\gamma - 1) \ln(2)}.$$

L'estimateur de Pickands se calcule à partir de distances entre deux statistiques d'ordre sans utiliser le maximum de l'échantillon, ceci implique une perte d'information sur la queue de distribution. De plus, cela le rend plus volatile que l'estimateur de Hill qui est calculé à partir de la moyenne des logarithmes des observations.

2.4.3 Estimateur de DEdH

Le troisième estimateur de l'indice de queue est celui proposé par Dekkers, Einmahl et De Hann. Cet estimateur est une généralisation de l'estimateur de Hill, valable pour tous les domaines d'attraction. Il est défini par :

$$\hat{\gamma}_n^{(DEdH)}(k_n) = \mathcal{M}_{k_n}^{(1)} + 1 - \frac{1}{2} \left(1 - \frac{(\mathcal{M}_{k_n}^{(1)})^2}{\mathcal{M}_{k_n}^{(2)}} \right)^{-1}$$

où $\mathcal{M}_{k_n}^{(r)} = \frac{1}{k_n} \sum_{i=1}^{k_n} (\ln(X_{(n-i+1)}) - \ln(X_{(n-k_n)}))^r$. La valeur de $\mathcal{M}_{k_n}^{(1)}$ correspond à l'estimateur de Hill.

Il est possible de montrer la convergence en loi suivante :

$$\sqrt{k_n} \left(\frac{\hat{\gamma}_n^{(DEdH)}(k_n) - \gamma}{\sigma_M} \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

où :

$$\sigma_M^2 = \begin{cases} 1 + \gamma^2 & \text{si } \gamma \geq 0 \\ (1 - \gamma^2)(1 - 2\gamma) \left(4 - 8 \frac{1-2\gamma}{1-3\gamma} + \frac{(5-11\gamma)(1-2\gamma)}{(1-3\gamma)(1-4\gamma)} \right) & \text{si } \gamma < 0 \end{cases}$$

En pratique, il n'est pas facile de classer ces estimateurs. Cependant, l'estimateur de Hill présente une variance asymptotique plus faible. C'est pourquoi c'est ce dernier qui est utilisé dans la suite. L'estimateur de Hill étant valable uniquement pour les indices de queue $\gamma > 0$, il convient de s'assurer de cette hypothèse.

2.4.4 Détermination du domaine d'attraction

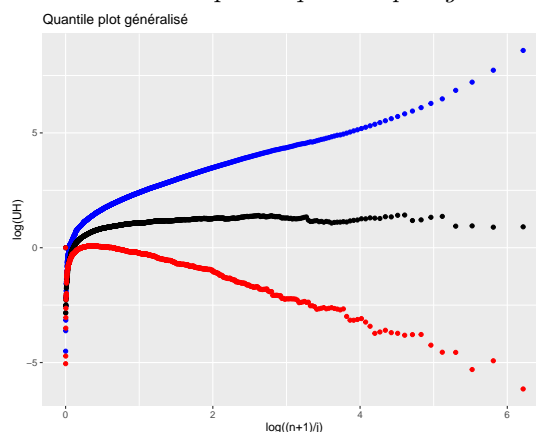
Une méthode graphique qui permet de déterminer à quel domaine d'attraction appartiennent les données consiste à tracer le *quantile plot généralisé*. Cette méthode est reprise de [Borchani \(2010\)](#). Le *quantile plot généralisé* est défini comme le graphe :

$$\left\{ \left(\ln\left(\frac{n+1}{j}\right), \ln(\hat{\gamma}_{j,n}^{(UH)}) \right), \text{ pour tout } j \in \llbracket 1; k_n \rrbracket \right\}$$

$$\text{avec } \hat{\gamma}_{j,n}^{(UH)} = X_{(n-j)} \times \hat{\gamma}_n^{(H)}(k_n).$$

La courbe obtenue ressemble à l'une des trois ci-dessous :

FIGURE 2.2: Exemple de *quantile plot généralisé*



Si la courbe est semblable à la courbe **bleue** (resp. **noire**, **rouge**), l'indice des valeurs extrêmes peut être considéré **positif** (resp. **nul**, **négatif**).

La difficulté pratique dans le calcul de ces estimateurs est le choix du nombre d'excès k_n à considérer. Si les estimateurs sont calculés avec un nombre trop important d'observations, leur biais sera élevé. Au contraire, s'ils sont calculés sur un nombre trop faible d'observations, c'est leur variance qui sera importante. De plus, le choix d'un seuil convenable est d'autant plus crucial que le théorème de Pickands, Balkema, De Hann suggère qu'au-delà d'un certain seuil suffisamment grand, les excès se comportent comme une distribution de Pareto Généralisée.

Le choix du seuil, ou par équivalence du nombre d'excès est une étape importante pour construire le modèle à deux lois. Différentes méthodes de sélection de seuils sont présentées dans la section suivante.

2.5 Détermination du seuil des extrêmes

Le choix du seuil u (ou de manière équivalente le nombre d'excès k_n à considérer avec n le nombre d'observations disponibles) est une étape primordiale dans la modélisation des montants de sinistres. Si le seuil est trop petit, alors l'approximation par une loi de Pareto Généralisée sera mauvaise. A l'inverse, si le seuil est trop grand, il y aura très peu d'observations pour le calcul des estimateurs qui auront alors une variance importante. Le bon choix du nombre d'excès k_n est celui qui représente le meilleur compromis entre le biais et la variance.

En pratique, de nombreuses méthodes existent pour sélectionner ce seuil optimal. Des méthodes graphiques et numériques sont présentées. Les méthodes graphiques nécessitent un contrôle de la part de l'utilisateur et une part de subjectivité contrairement aux méthodes numériques qui fournissent une valeur pour l'estimation du seuil.

2.5.1 La fonction des excès moyens

Appelé en anglais, *mean excess function*, cette fonction donne l'espérance de l'excédent d'une variable aléatoire X par rapport à un seuil conditionnellement au fait que ce seuil est dépassé. La fonction des excès moyens est définie pour tout seuil u inférieur au point terminal x_F par :

$$e(u) = \mathbb{E}[X - u | X > u] = \frac{\int_u^{+\infty} (1 - F(x)) dx}{1 - F(u)}$$

Soit X_1, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées, et x_1, \dots, x_n l'échantillon d'observations. Une estimation empirique de la fonction des excès moyens, noté $e_n(u)$ est définie par :

$$e_n(u) = \frac{\sum_{i=1}^n \max(0, x_i - u)}{\sum_{i=1}^n \mathbb{1}_{x_i > u}}$$

La fonction des excès moyens $e(u)$ est simplement estimée par la somme des excès au delà du seuil u divisée par le nombre d'observations supérieures au seuil u . Pour tracer le graphique des excès moyens, il faut choisir des valeurs pour le seuil u . En pratique, ce sont les différentes valeurs des observations $(x_i)_{1 \leq i \leq n}$ qui servent de seuil. Ainsi, en considérant l'échantillon d'observations ordonnées par ordre croissant $x_{(1)}, \dots, x_{(n)}$, une estimation empirique de la fonction des excès moyens vaut :

$$e_n(x_{(k)}) = \frac{1}{n-k} \sum_{j=1}^{n-k} (x_{(k+j)} - x_{(k)})$$

Le graphe de la fonction des excès moyens (*mean excess plot* en anglais) consiste à représenter les points $\{x_{(k)}, e_n(x_{(k)})\}$ pour tout $k \in \llbracket 1; n \rrbracket$.

Le *mean excess plot* est un outil graphique qui permet de déterminer le seuil des valeurs extrêmes u à partir duquel les observations se comportent comme une distribution de Pareto Généralisée. La proposition suivante permet de faire le lien entre l'allure du *mean excess plot* et le comportement d'une loi de Pareto Généralisée. En effet, le graphe des excès moyens d'une distribution de Pareto Généralisée est linéaire en u .

Proposition 1. *Soit X une variable aléatoire suivant une $\mathcal{GPD}(\gamma, \sigma)$. La fonction des excès moyens de X s'écrit pour tout $\gamma < 1$:*

$$e(u) = \frac{\gamma}{1-\gamma} \times u + \frac{\sigma}{1-\gamma}$$

Démonstration. Soit $\gamma < 1$.

$$e(u) = \frac{1}{\mathbb{P}(X > u)} \times \int_u^{+\infty} \mathbb{P}(X > x) dx$$

Avec d'une part :

$$\mathbb{P}(X > u) = \left(1 + \frac{\gamma u}{\sigma}\right)^{-\frac{1}{\gamma}}$$

Et d'autre part :

$$\int_u^{+\infty} \mathbb{P}(X > x) dx = \int_u^{+\infty} \left(1 + \frac{\gamma x}{\sigma}\right)^{-\frac{1}{\gamma}} dx$$

avec le changement de variable $y = 1 + \frac{\gamma x}{\sigma}$, il vient :

$$\int_u^{+\infty} \mathbb{P}(X > x) dx = \frac{\sigma}{1-\gamma} \left[1 + \frac{\gamma u}{\sigma}\right]^{1-\frac{1}{\gamma}}$$

Ainsi :

$$\begin{aligned} e(u) &= \frac{\sigma}{1-\gamma} \left[1 + \frac{\gamma u}{\sigma}\right]^{1-\frac{1}{\gamma}} \times \frac{1}{\left[1 + \frac{\gamma u}{\sigma}\right]^{-\frac{1}{\gamma}}} \\ &= \frac{\sigma}{1-\gamma} \left[1 + \frac{\gamma u}{\sigma}\right] \\ &= \frac{\gamma}{1-\gamma} \times u + \frac{\sigma}{1-\gamma} \end{aligned}$$

□

Si la fonction empirique des excès moyens se comporte linéairement à partir d'un certain seuil positif, cela signifie que les données supérieures à ce seuil suivent une loi de Pareto Généralisée. Il convient alors de déterminer le seuil à partir duquel le *mean excess plot* est approximativement linéaire.

Exemples

Pour illustrer cette propriété de linéarité, deux échantillons sont simulés. Le premier comporte 500 valeurs issues d'une loi de Pareto Généralisée de paramètres $\gamma = 0,8$ et $\sigma = 10000$ et le deuxième de 50 valeurs issues de cette même loi. Le *mean excess plot* est tracé pour ces deux échantillons :

FIGURE 2.3: *Mean excess plot* d'une GPD avec 500 observations.

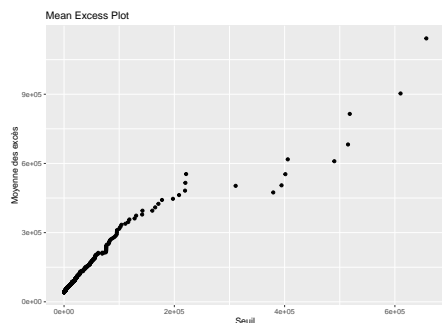
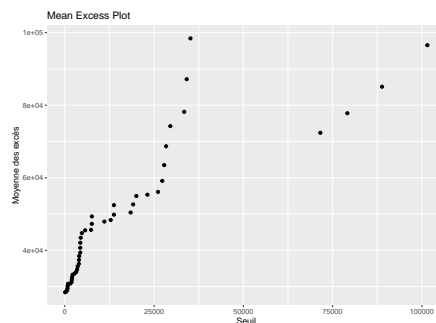


FIGURE 2.4: *Mean excess plot* d'une GPD avec 50 observations.



Pour ces deux échantillons, la courbe du *mean excess plot* est croissante et la partie droite du graphe est instable car les excédents moyens sont calculés sur un faible nombre de données. Pour l'échantillon de 500 observations, le comportement linéaire est sans équivoque. En revanche pour l'échantillon de 50 observations, il n'est pas évident d'observer un seuil au delà duquel la courbe est linéaire malgré que les données soient simulées selon une loi de Pareto Généralisée.

Finalement, le graphe des excédents moyens permet de déterminer un seuil des extrêmes grâce à la propriété de linéarité de la loi de Pareto Généralisée. Cependant, sur données réelles, qui plus est en faible quantité, la détermination de ce seuil de linéarité peut s'avérer délicat. Il convient alors de s'appuyer également sur d'autres outils.

2.5.2 Le graphe de l'indice des valeurs extrêmes de la *GPD*

Le graphe de l'indice des valeurs extrêmes d'une *GPD* (*Gamma plot*) est un autre outil graphique qui permet de déterminer un seuil au-delà duquel les données se comportent selon une distribution de Pareto Généralisée. Ce graphe repose sur la proposition de stabilité de l'indice des valeurs extrêmes d'une *GPD* par rapport au seuil.

Proposition 2. Si X , une variable aléatoire, suit une $\mathcal{GPD}(\gamma, \sigma)$ alors la variable $X - u | X > u$ suit une $\mathcal{GPD}(\gamma, \sigma + \gamma u)$.

Démonstration. Soit $x \in [0, x_F - u]$ où x_F est le point terminal de X .

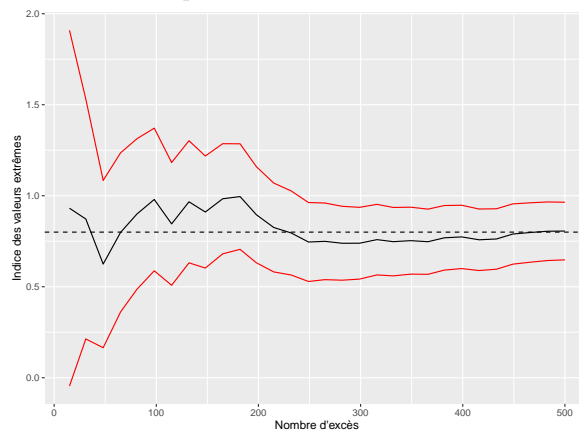
$$\begin{aligned}
\mathbb{P}(X - u < x | X > u) &= 1 - \frac{\mathbb{P}(X \geq x + u)}{\mathbb{P}(X > u)} \\
&= 1 - \frac{\left(1 + \frac{\gamma}{\sigma}(x + u)\right)^{-\frac{1}{\gamma}}}{\left(1 + \frac{\gamma}{\sigma}u\right)^{-\frac{1}{\gamma}}} \\
&= 1 - \left(1 + \frac{\frac{\gamma}{\sigma}x}{1 + \frac{\gamma}{\sigma}u}\right)^{-\frac{1}{\gamma}} \\
&= 1 - \left(1 + x\left(\frac{\gamma}{\sigma + \gamma u}\right)\right)^{-\frac{1}{\gamma}}
\end{aligned}$$

□

Le *Gamma plot* consiste à tracer l'estimation de l'indice des valeurs extrêmes obtenue par maximum de vraisemblance en fonction du nombre d'excès considérés. Si un échantillon provient d'une loi de Pareto Généralisée, il faut observer une stabilité de l'estimateur de l'indice des valeurs extrêmes γ .

Cette propriété de stabilité est illustrée sur la simulation de 5 000 observations issues d'une $\mathcal{GPD}(\gamma = 0,8; \sigma = 10000)$. Le graphique de l'indice des valeurs extrêmes associé est représenté :

FIGURE 2.5: *Gamma plot* sur données simulées selon une \mathcal{GPD} .



L'indice des valeurs extrêmes γ estimé se stabilise à la valeur réelle de 0,8 à partir de 250 excès.

Cet outil permet donc de déterminer un seuil d'extrêmes u en cherchant le nombre d'excès à partir duquel l'estimateur de l'indice des extrêmes se stabilise.

2.5.3 Le graphique de Hill et de Pickands

Ces deux méthodes graphiques de détermination du seuil reposent sur la propriété de stabilité de la loi de Pareto Généralisée. Ces graphiques consistent à tracer la valeur de l'estimateur de Hill $\hat{\gamma}_n^{(H)}$ et de Pickands $\hat{\gamma}_n^{(P)}$ en fonction du nombre d'excès k_n considérés. Le seuil à sélectionner correspond, par équivalence, au plus petit nombre d'excès pour lequel les estimateurs se stabilisent.

Soit X_1, \dots, X_n un échantillon de variables aléatoires indépendantes et identiquement distribuées. L'échantillon ordonné par ordre croissant est noté $X_{(1)}, \dots, X_{(n)}$. Le graphique de Hill et de Pickands sont présentés ci-dessous.

Le Hill Plot

L'inconvénient du Hill Plot est qu'il est valable uniquement pour les distributions de probabilités appartenant au domaine d'attraction de Fréchet, c'est-à-dire celles ayant un indice des valeurs extrêmes strictement positif. Le graphe du Hill Plot consiste à représenter les points :

$$\left\{ k_n, \hat{\gamma}_n^{(H)}(k_n) = \frac{1}{k_n} \sum_{i=1}^{k_n} \ln(X_{(n-i+1)}) - \ln(X_{(n-k_n)}) \right\}$$

Il faut sélectionner un nombre d'excès k_n pas trop petit, au risque de disposer de trop peu d'observations pour estimer convenablement les paramètres de la \mathcal{GPD} . Le nombre d'excès k_n ne doit pas non plus être trop grand au risque que l'approximation par une loi \mathcal{GPD} ne soit pas vérifiée. En théorie, le nombre d'excès optimal k_n à choisir devrait vérifier $\lim_{n \rightarrow +\infty} k_n = +\infty$ (pas trop petit) et $\lim_{n \rightarrow +\infty} \frac{k_n}{n} = 0$ (pas trop grand).

Graphiquement sur le Hill Plot, il faut sélectionner la plus petite valeur de k_n pour laquelle l'estimateur de Hill est stable.

Le Pickands Plot

Le grand avantage du Pickands Plot par rapport au Hill Plot est qu'il est valable pour les trois domaines d'attraction (Gumble, Weibull ou Fréchet). Il consiste à représenter les points :

$$\left\{ k_n, \hat{\gamma}_n^{(P)}(k_n) = \frac{1}{\ln(2)} \ln \left(\frac{X_{(n-k_n)} - X_{(n-2k_n)}}{X_{(n-2k_n)} - X_{(n-4k_n)}} \right) \right\}$$

Comme pour le Hill Plot, il convient de sélectionner le plus petit k_n pour lequel l'estimateur de Pickands se stabilise.

Exemples sur données simulées

Pour illustrer ces deux méthodes de détermination de seuil, un premier échantillon de 400 observations issues d'une distribution de Pareto de paramètres $\alpha = 1,5$ et $A = 10\,000$ est simulé. L'indice des valeurs extrêmes associé vaut $\gamma = \frac{1}{\alpha} = \frac{2}{3}$.

FIGURE 2.6: Hill Plot sur 400 observations issues d'une Pareto(1,5 ; 10 000)

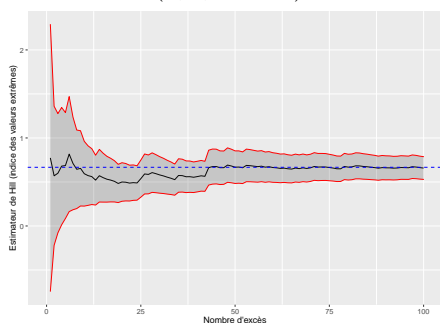
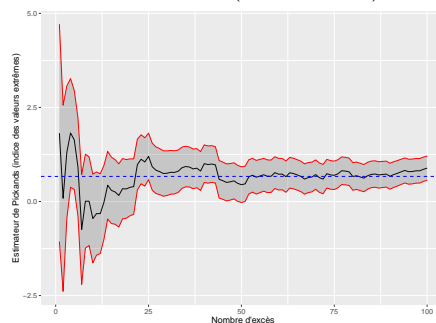


FIGURE 2.7: Pickands Plot sur 400 observations issues d'une Pareto(1,5 ; 10 000)



L'estimateur de Pickands est beaucoup plus volatile que l'estimateur de Hill, en particulier pour les plus petites valeurs du nombre d'excès k_n . Cette instabilité plus importante est accentuée en présence d'un échantillon de petite taille comme le montre ce second exemple, avec la simulation de 100 observations issues de la même loi que le premier échantillon.

FIGURE 2.8: Hill Plot sur 100 observations issues d'une Pareto(1,5 ; 10 000)

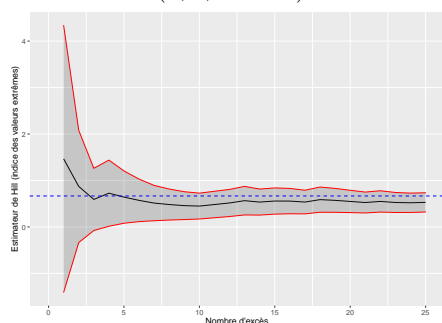
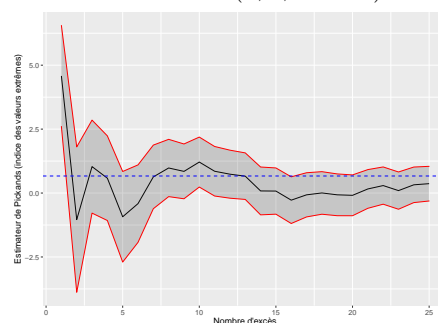


FIGURE 2.9: Pickands Plot sur 100 observations issues d'une Pareto(1,5 ; 10 000)



En réassurance, la quantité de données disponibles est souvent faible. La forte volatilité de l'estimateur de Pickands dans ce cas peut rendre difficile la détermination de la zone de stabilité. C'est pourquoi dans la suite, le Hill Plot sera préféré au Pickands Plot pour la détermination du seuil.

Cependant, même si l'estimateur de Hill est moins volatile que celui de Pickands, la phase de stabilisation de l'estimateur peut parfois être difficile à identifier, c'est pourquoi il convient de s'appuyer également sur d'autres outils de sélection du seuil.

2.5.4 Le graphique de Gerstengarbe

Le graphique de Gerstengarbe (*Gerstengarbe plot*) est issu des travaux de [Gerstengarbe and Werner \(1989\)](#). Cette méthode repose sur le test non-paramétrique de Mann-Kendall qui sert à déterminer si une tendance est identifiable dans une série temporelle. En considérant les différences de coût entre deux sinistres successifs, l'idée est qu'il semble raisonnable de s'attendre à un changement de comportement de ces écarts entre ceux issus des données extrêmes et ceux issus des données non-extrêmes. Ainsi, il existe un point de changement dans le comportement des écarts, et ce point de changement est considéré comme le point de départ des données extrêmes. Le seuil des extrêmes est alors estimé par ce point de changement.

Mathématiquement, soit x_1, \dots, x_n l'échantillon des montants sinistres et $x_{(1)} \leq \dots \leq x_{(n)}$ ce même échantillon ordonné par ordre croissant. La série des différences entre deux montants de sinistres est notée $\Delta_i = x_{(i)} - x_{(i-1)}$ pour tout $i \in \llbracket 2; n \rrbracket$. On s'attend à un changement de comportement de cette série de différences, ce point de changement est le point d'entrée dans la zone des données extrêmes.

Pour identifier ce point d'entrée, la version séquentielle du test de Mann-Kendall est utilisée. Pour $i = 1, \dots, n - 1$, la série suivante est calculée :

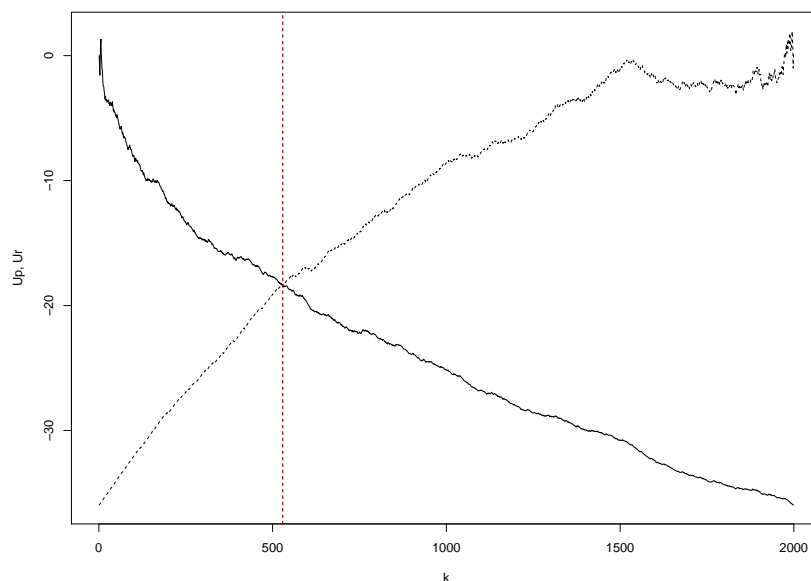
$$U_i = \frac{\sum_{k=2}^i n_k - \frac{i(i-1)}{4}}{\sqrt{\frac{i(i-1)(i+5)}{72}}}$$

où $n_k = \sum_{j=2}^k \mathbb{1}_{\Delta_j \leq \Delta_k}$, c'est-à-dire le nombre de valeurs dans $\Delta_2, \dots, \Delta_k$ inférieures à Δ_k . Une seconde série \tilde{U}_i est calculée de la même manière mais en utilisant la série décroissante des différences $\Delta_n, \dots, \Delta_2$.

Le *Gerstengarbe plot* consiste à tracer les deux séries $(U_i)_{2 \leq i \leq n}$ et $(\tilde{U}_i)_{2 \leq i \leq n}$. Le point de départ des données extrêmes correspond alors au point d'intersection des deux séries.

Pour illustrer cette méthode, un échantillon de 2 000 observations de loi $\mathcal{GPD}(\gamma = 0, 8; \sigma = 10000)$ est simulé et le *Gerstengarbe plot* associé est tracé :

FIGURE 2.10: Exemple de *Gerstengarbe plot* sur données simulées.



Le point d'intersection des deux séries indique un nombre d'excès $k = 529$ observations ce qui correspond à un seuil de 25 135.

Finalement, l'avantage du *Gerstengarbe plot* est de fournir une valeur fixe pour le seuil d'entrée dans la zone extrême, ce que ne permet pas le *mean excess plot* ou le *Hill plot*. Cependant, cette méthode n'est pas issue des travaux de la théorie des valeurs extrêmes, le seuil fourni par cette méthode est difficile à vérifier et valider. En revanche, utiliser en première approche la valeur du seuil fournie par le *Gerstengarbe plot* pour ensuite analyser les seuils proposés par les autres outils peut être une méthodologie intéressante.

2.5.5 Méthode de minimisation de l'AMSE

Cette méthode numérique de sélection d'un seuil optimal est issue des travaux de [Caeiro and Gomes \(2016\)](#). L'idée générale de cette méthode est de sélectionner un nombre d'excès k_0 qui minimise l'erreur quadratique moyenne de l'estimateur de Hill. On rappelle que pour un échantillon X_1, \dots, X_n de variables aléatoires indépendantes et de même loi appartenant au domaine d'attraction de Fréchet, l'estimateur de Hill de l'indice des valeurs extrêmes γ est donné par :

$$\hat{\gamma}_n^{(H)}(k_n) = \frac{1}{k_n} \sum_{i=1}^{k_n} \ln(X_{(n-i+1)}) - \ln(X_{(n-k_n)})$$

où $X_{(n-k_n)}$ est une estimation du seuil des extrêmes et k_n un entier compris entre 1 et n . Cette méthode consiste alors à sélectionner un seuil $\hat{u} = X_{(n-\hat{k}_0)}$ où \hat{k}_0 est une estimation de la valeur k_0 de k_n qui minimise l'erreur quadratique moyenne asymptotique (AMSE) de l'estimateur de Hill, c'est-à-dire la somme du biais au carré et de la variance de la distribution asymptotique de $\hat{\gamma}_n^{(H)}(k_n)$. L'objectif est de sélectionner un seuil qui représente un bon compromis entre le biais et la variance de l'estimateur. Sous certaines conditions de régularité qui ne sont pas présentées ici, il est possible d'exprimer l'AMSE par :

$$\text{AMSE}(\hat{\gamma}_n^{(H)}(k_n)) = \gamma^2 \left(\frac{1}{k_n} + \frac{\lambda^2}{(1-\rho)^2} \left(\frac{n}{k_n} \right)^{2\rho} \right)$$

On peut montrer que cette quantité est minimisée pour la valeur :

$$k_0 = \left\lfloor \left(\frac{(1-\rho)^2 n^{-2\rho}}{-2\rho\lambda^2} \right)^{\frac{1}{(1-2\rho)}} \right\rfloor$$

Avec $\lfloor \cdot \rfloor$ la partie entière et ρ et λ des paramètres du second ordre qui contrôle la vitesse de convergence de l'estimateur de Hill. Ces paramètres peuvent être estimés comme décrit par [Caeiro and Gomes \(2016\)](#), et ainsi obtenir une estimation du nombre d'excès à considérer pour estimer le seuil des extrêmes :

$$\hat{k}_0 = \left\lfloor \left(\frac{(1-\hat{\rho})^2 n^{-2\hat{\rho}}}{-2\hat{\rho}\hat{\lambda}^2} \right)^{\frac{1}{(1-2\hat{\rho})}} \right\rfloor$$

Finalement, cette méthode fournit le nombre d'excès optimal qui minimise l'AMSE de l'estimateur de Hill. Il existe une correspondance naturelle entre le nombre d'excès de l'échantillon et la valeur du seuil correspondant. Cette méthode de détermination de seuil présente l'avantage de fournir une valeur pour l'estimation du seuil contrairement aux méthodes graphiques qui peuvent être difficiles à interpréter.

Chapitre 3

Modélisation de la sinistralité

L'objectif de l'étude est de compléter et optimiser l'outil de tarification des traités de type excédent de sinistre en étendant les possibilités d'ajustement des sinistres à des lois de probabilité. En particulier, un modèle à deux lois, basé sur la théorie des valeurs extrêmes, est proposé pour améliorer la tarification des parties non-travaillantes des programmes de réassurance.

Ce chapitre présente les différents éléments mathématiques utilisés pour modéliser la sinistralité issue d'une cédante.

Concernant l'aspect métier, ces modèles théoriques font l'objet d'un développement informatique avec la création d'un outil de modélisation sur l'application *R Shiny*, présenté en annexe [B](#).

3.1 Modèle individuel et Modèle collectif

La tarification d'une tranche de réassurance repose sur deux composants, le nombre de sinistres qui se traduit par l'estimation d'une fréquence, et le coût d'un sinistre qui se traduit par l'estimation d'un montant moyen à charge. Ces deux éléments permettent d'estimer la charge globale du portefeuille. Il existe deux approches pour modéliser cette charge : le modèle individuel et le modèle collectif.

3.1.1 Le modèle individuel

L'objectif d'un modèle individuel est de pouvoir étudier la distribution de probabilité de la charge totale S de sinistres sur un portefeuille de réassurance au cours d'une période de référence. La charge globale des sinistres s'exprime comme la somme sur le nombre de polices n du montant total engendré par chacune des polices $(X_i)_{1 \leq i \leq n}$. Formellement, ce modèle est défini ainsi :

Définition 6 (Modèle individuel pour la charge de sinistres). *La charge totale S pour un portefeuille comportant n polices, est donnée par :*

$$S = \sum_{i=1}^n X_i$$

avec :

X_i : la somme des montants de sinistres qui ont touché la police i . La valeur de X_i peut être nulle si aucun sinistre n'a impacté la police i durant la période d'observation.

Les variables aléatoires X_1, \dots, X_n sont supposées mutuellement indépendantes mais ne sont pas supposées identiquement distribuées.

Le modèle individuel repose sur un raisonnement police par police. Ces dernières n'étant pas supposées identiquement distribuées, il est souvent malaisé d'estimer la distribution de la charge globale. Le modèle collectif permet de résoudre ce problème par une approche simplifiée.

3.1.2 Le modèle collectif : approche probabiliste fréquence et sévérité

L'objectif du modèle collectif est d'obtenir un modèle plus facile à manipuler que le modèle individuel, et fournissant une bonne approximation de celui-ci, avec la même finalité : étudier la distribution de probabilité de la charge totale S de sinistres sur un portefeuille de réassurance au cours d'une période donnée. Dans ce modèle, la charge globale S s'exprime comme la somme sur le nombre de sinistres N des montants de chaque sinistre. Formellement, le modèle collectif se définit comme :

Définition 7 (Modèle collectif pour la charge de sinistres). *La charge totale S est donnée par :*

$$S = \sum_{i=1}^N X_i$$

avec :

X_1, \dots, X_N des variables aléatoires représentant les montants de chaque sinistre du portefeuille. Ces variables sont supposées indépendantes et identiquement distribuées de même loi qu'une variable aléatoire X .

N est une variable aléatoire à valeurs dans \mathbb{N} indépendante de la suite $(X_i)_{i \geq 1}$. Cette variable représente le nombre total de sinistres survenus au sein du portefeuille.

Ce modèle repose sur un raisonnement sinistre par sinistre. Il n'y a plus de distinction entre les différentes polices. Autrement dit, une police peut donner lieu à plusieurs sinistres. De ce modèle découle la formule de la prime pure reliant cette dernière à la fréquence et à la sévérité des sinistres :

$$\mathbb{E}(S) = \mathbb{E}(N) \times \mathbb{E}(X)$$

La formule pour calculer l'écart-type de la charge s'obtient par :

$$\sigma(S) = \sqrt{\mathbb{E}(N) \times \sigma^2(X) + \mathbb{E}^2(X) \times \sigma^2(N)}$$

Dans la suite, c'est le modèle collectif qui est retenu pour modéliser la charge globale du portefeuille d'une cédante. La section suivante s'attache au choix du seuil de modélisation et à l'estimation de la fréquence des sinistres supérieurs au seuil sélectionné.

3.2 Seuil de modélisation et fréquence au dessus du seuil

Le seuil de modélisation est par définition la valeur en dessous de laquelle les montants de sinistres inférieurs ne sont pas considérés pour la modélisation. Le choix du seuil de modélisation est important en réassurance, il permet de ne conserver que les données jugées pertinentes pour le calcul de la prime pure des tranches de réassurance à tarifier. De ce seuil de modélisation découle un nombre de sinistres et donc une fréquence.

3.2.1 Sélection du seuil de modélisation

En réassurance, les données disponibles pour estimer la prime pure sont fournies par la cédante. De manière générale, la cédante ne fournit pas l'ensemble de son portefeuille de sinistres et se limite à un sous-échantillon de sinistres supérieurs à un seuil d'avis de sinistres.

Si le seuil d'avis de sinistres est jugé suffisamment élevé au regard de la priorité de la tranche de réassurance à tarifier, l'actuaire peut considérer le seuil d'avis de sinistres comme le seuil de modélisation. En revanche, si le seuil d'avis de sinistres semble petit au regard de la priorité de la tranche de réassurance à tarifier, l'actuaire peut être amené à sélectionner un seuil de modélisation différent du seuil d'avis de sinistres fourni par la cédante.

En pratique, il n'existe pas de règles spécifiques qui permettent de sélectionner un seuil de modélisation optimal par rapport à la portée de la tranche de réassurance à tarifier. La seule contrainte à respecter

dans le choix du seuil de modélisation est que ce dernier doit être au moins supérieur au seuil d'avis de sinistres et au plus égal à la priorité de la tranche de réassurance à coter. Dans le cas particulier où le seuil d'avis de sinistres est supérieur à la priorité de la tranche de réassurance à tarifier, l'actuaire ne peut correctement modéliser la sinistralité et doit demander à ce que le seuil d'avis de sinistres de la cédante soit diminué.

Le choix d'un seuil de modélisation convenable est important, car les données inférieures à ce seuil ne sont pas considérées pour la modélisation. Ainsi, sélectionner un seuil trop faible peut conduire à focaliser la modélisation sur des sinistres de trop faibles montants par rapport à la tranche de réassurance à coter, et in fine sous-estimer le risque. Inversement, sélectionner un seuil de modélisation trop élevé peut poser le problème de la perte d'informations.

Dans la suite, le seuil de modélisation retenu est noté s .

3.2.2 Calcul de la fréquence

Le seuil de modélisation s étant déterminé, il est possible d'estimer simplement le nombre de sinistres N_s supérieurs à ce seuil pour l'année à tarifier. On suppose que la variable aléatoire N_s suit une loi de Poisson de paramètre λ_s . Ainsi, l'égalité suivante est vérifiée :

$$\mathbb{E}(N_s) = \lambda_s$$

Pour estimer la valeur de λ_s , l'actuaire dispose des informations suivantes :

- n : le nombre d'années d'historique pour tarifier le traité.
- Y : l'année à tarifier.
- $(N_j)_{1 \leq j \leq n}$: le nombre de sinistres survenus l'année j .
- $(X_{i,j})_{1 \leq i \leq N_j \text{ et } 1 \leq j \leq n}$: le montant *as-if* du sinistre i survenu durant l'année j .
- $(EPI_j)_{1 \leq j \leq n}$: l'assiette de primes *as-if* des années j .
- EPI_Y : l'assiette de primes estimée pour l'année à tarifier.

L'estimation du paramètre λ_s retenue est la suivante :

$$\hat{\lambda}_s = \left[\sum_{j=1}^n \sum_{i=1}^{N_j} \mathbb{1}_{\mathbb{R}_+}(X_{i,j} - s) \right] \times \underbrace{\frac{EPI_Y}{\sum_{j=1}^n EPI_j}}_{\text{terme correctif}}$$

L'estimation du nombre de sinistres supérieurs au seuil s pour l'année à tarifier est le nombre de sinistres ayant dépassés le seuil dans les années d'expérience disponibles multiplié par un terme correctif. Ce dernier n'est pas simplement $\frac{1}{\text{nombre d'années d'historique}}$ qui ne rendrait pas compte de l'évolution du portefeuille. L'utilisation du ratio des assiettes de primes permet de prendre en considération l'évolution du portefeuille de la cédante durant les années d'expérience. Si l'assiette de primes augmente, on s'attend à observer plus de sinistres de montant supérieur au seuil de modélisation.

La fréquence ne fait pas l'objet d'une étude approfondie dans ce mémoire, cette approche est conservée de l'outil de tarification déjà implémenté. La suite s'attache à présenter les différents outils de modélisation de la sévérité, avec notamment la construction d'un modèle à deux lois.

3.3 Modélisation de la sévérité

La section précédente a permis d'introduire la sélection du seuil de modélisation et la fréquence des sinistres associée. L'étape qui suit est la modélisation des montants de sinistres supérieurs au seuil de modélisation. C'est l'axe principal de l'étude : l'amélioration de la modélisation des sinistres. Cette section introduit la notion d'ajustement paramétrique, en particulier avec la prise en compte du seuil de modélisation par des distributions de probabilité tronquées à gauche.

3.3.1 La fonction de répartition empirique

L'actuaire en charge de la tarification des traités de réassurance en excédent de sinistre dispose d'un échantillon d'observations noté $(x_i)_{1 \leq i \leq n}$. Ces observations représentent des montants de sinistres *as-if*, survenus pour une cédante donnée. L'hypothèse principale est que les observations sont la réalisation de variables aléatoires $(X_i)_{1 \leq i \leq n}$ indépendantes et identiquement distribuées de même loi qu'une variable aléatoire X de fonction de répartition F . L'objectif de l'actuaire est d'approcher de la meilleure manière la distribution de probabilité de X .

La première étape consiste à représenter la fonction de répartition empirique F_n à l'aide des observations disponibles. Elle se définit pour tout $x \in \mathbb{R}$ par :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{x_i \leq x}$$

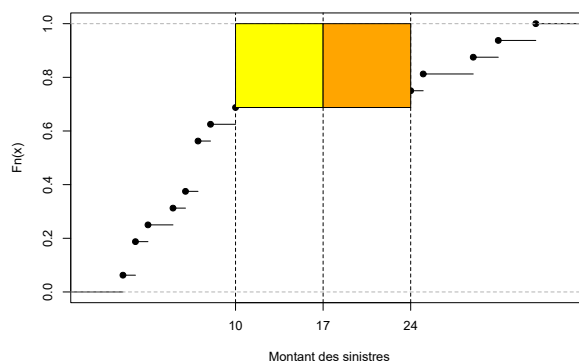
La valeur de $F_n(x)$ peut se traduire par la proportion d'observations inférieures ou égales à x . La fonction de répartition empirique est très largement utilisée en première approche de F car elle est simple à construire et surtout car le théorème de Glivenko-Cantelli assure la convergence suivante :

$$\sup_{x \in \mathbb{R}} \| F_n(x) - F(x) \| \xrightarrow[n \rightarrow +\infty]{\text{presque-sûrement}} 0$$

Autrement dit, la fonction de répartition empirique converge presque-sûrement vers la fonction de répartition théorique. Cependant, la fonction de répartition empirique présente deux inconvénients majeurs en réassurance :

1. La fonction de répartition empirique est construite à partir des observations disponibles, donc au-delà du maximum de l'échantillon, la fonction de répartition est nulle. C'est-à-dire que la probabilité d'observer un sinistre de montant supérieur au maximum de l'échantillon disponible vaut zéro. Par conséquent, si l'actuaire doit estimer le coût à charge dans une tranche de réassurance ayant une priorité supérieure au maximum des observations disponibles, le coût sera nul. Les queues de distributions sont mal modélisées par la fonction de répartition empirique.
2. La fonction de répartition empirique est une fonction en escalier, il existe alors des plateaux. Dans le cas où le plateau est long, c'est notamment le cas dans les queues de distributions où le nombre d'observations disponibles est faible, le coût à charge pour deux tranches de réassurance adjacentes de même amplitude situées sur le même plateau est identique. Ce constat n'est pas cohérent, la tranche la plus haute a une probabilité plus faible d'être touchée par un sinistre, et par conséquent l'estimation de son coût à charge doit être inférieure.

FIGURE 3.1: Illustration du coût à charge de deux tranches de même amplitude situées sur un plateau de la fonction de répartition empirique



Ce graphique illustre ce problème, l'estimation du coût à charge d'une tranche de réassurance est représentée par l'aire au-dessus de la fonction de répartition. Pour tarifer une tranche 7 XS 10 ou 7 XS 17, l'aire sous la courbe est similaire, le coût est identique.

Pour résoudre ces deux problèmes, l'actuaire utilise des lois de probabilité. Cette méthode de *fitting* consiste à ajuster une distribution de probabilité continue sur la distribution empirique. De cette manière, l'actuaire est en mesure d'estimer un coût à charge entre la priorité et la portée de n'importe quels traités de réassurance. Les lois couramment utilisées en réassurance pour réaliser cet ajustement sont présentées.

3.3.2 Les lois usuelles pour la modélisation de la sévérité en réassurance

Les distributions de probabilité sont indispensables pour l'actuaire qui doit coter des couvertures de réassurance pour lesquelles les données d'expérience sont en faible quantité voire inexistantes. Ces modèles permettent de déterminer un coût à charge pour une tranche de réassurance donnée à l'aide de formules fermées. Les distributions statistiques les plus utilisées en réassurance pour modéliser les sinistres sont décrites.

La loi de Pareto

La loi de Pareto a été mise en évidence pour ses applications en réassurance par [Hagstøem \(1925\)](#). Cette loi est très régulièrement utilisée par les actuaires en réassurance pour modéliser la sinistralité afin d'obtenir une estimation du montant moyen à charge dans les tranches de réassurance en excédent de sinistre.

En effet, la distribution de Pareto a la réputation de bien ajuster les grands sinistres, c'est-à-dire ceux qui intéressent l'actuaire lors de la tarification des traités en excédent de sinistre. De plus, la distribution de Pareto présente le grand avantage d'être facile à manipuler pour obtenir des formules fermées. Ces éléments font donc le succès de la distribution de Pareto auprès des actuaires en réassurance.

Expression mathématique Une variable aléatoire X suit une distribution de Pareto de paramètres $\alpha > 0$ et A , que l'on note $X \sim \mathcal{P}(\alpha, A)$, si elle admet une fonction de densité de la forme :

$$f(x) = \alpha A^\alpha x^{-\alpha-1} \mathbb{1}_{x \geq A}$$

La fonction de répartition associée s'exprime pour tout $x \geq A$ par :

$$F(x) = 1 - A^\alpha x^{-\alpha}$$

L'espérance d'une Pareto est donnée par :

$$\forall \alpha > 1, \quad \mathbb{E}(X) = \frac{A\alpha}{\alpha - 1}$$

Autrement dit, l'espérance d'une Pareto n'existe pas si $\alpha \leq 1$. Ceci implique que la valeur moyenne ne peut pas être utilisée pour estimer le coût moyen dans une tranche de réassurance à portée illimitée si le paramètre de forme est inférieur ou égal à 1.

La variance d'une Pareto est donnée par :

$$\forall \alpha > 2, \quad \mathbb{V}(X) = A^2 \frac{\alpha}{(\alpha - 1)^2(\alpha - 2)}$$

De manière plus générale, les moments d'une Pareto sont donnés par :

$$\forall \alpha > k, \quad \mathbb{E}(X) = A^k \frac{\alpha}{\alpha - k}$$

Valeur de référence pour le paramètre α Le marché de la réassurance s'accorde sur des valeurs de référence pour le paramètre α selon la branche de l'affaire.

TABLE 3.1: Valeur de marché du paramètre α

Branche	Valeur de référence
Incendie	$\alpha \sim 2$
Incendie risques industriels	$\alpha \sim 1,5$
CAT	$\alpha \sim 1,1$
RC automobile	$\alpha \sim 2,5$
RC générale	$\alpha \sim 1,8$
Transport	$\alpha \sim 1,5$

Le paramètre α est le paramètre de forme de la distribution de Pareto. Une petite valeur de α implique une plus grande probabilité de grands sinistres. Cette caractéristique est illustrée avec l'exemple suivant : Considérons le paramètre $A = 1\,000$ et calculons la probabilité de survie $\mathbb{P}(X > x)$ pour différentes valeurs de x et du paramètre de forme α :

TABLE 3.2: Probabilité de survie de la Pareto en fonction de α

α	x				
	2 000	4 000	6 000	8 000	10 000
1,1	0,4665	0,2176	0,1393	0,1015	0,0794
1,5	0,3535	0,1250	0,0680	0,0442	0,0316
1,8	0,2872	0,0825	0,0397	0,0237	0,0158
2,2	0,2500	0,0625	0,0278	0,0156	0,0100
2,5	0,1768	0,0312	0,0113	0,0055	0,0032

La loi de Weibull

Une variable aléatoire X suit une distribution de Weibull de paramètres $a > 0$ et $b > 0$, que l'on note $X \sim \mathcal{W}(a, b)$, si elle admet une fonction de densité de la forme :

$$f(x) = ab^{-a} x^{a-1} e^{-\left(\frac{x}{b}\right)^a} \mathbf{1}_{x \geq 0}$$

La fonction de répartition associée s'exprime pour tout $x \geq 0$ par :

$$F(x) = 1 - e^{-\left(\frac{x}{b}\right)^a}$$

L'espérance d'une Weibull est donnée par :

$$\mathbb{E}(X) = \frac{b}{a} \Gamma\left(\frac{1}{a}\right)$$

La variance d'une Weibull est donnée par :

$$\mathbb{V}(X) = \frac{b^2}{a} \left[2\Gamma\left(\frac{2}{a}\right) - \frac{1}{a} \Gamma^2\left(\frac{1}{a}\right) \right]$$

La loi Log-Normale

Une variable aléatoire X suit une distribution de Log-Normale de paramètres $\mu \in \mathbb{R}$ et $\sigma > 0$, que l'on note $X \sim \mathcal{LN}(\mu, \sigma)$, si elle admet une fonction de densité de la forme :

$$\forall x > 0, \quad f(x) = \frac{1}{\sqrt{2\pi\sigma x}} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2}$$

Il existe la propriété suivante, $X \sim \mathcal{LN}(\mu, \sigma)$ si la variable aléatoire $Y = \ln(X)$ suit une distribution normale $\mathcal{N}(\mu, \sigma)$. Ainsi la variable aléatoire $\frac{\ln(X) - \mu}{\sigma} \sim \mathcal{N}(0, 1)$. On peut donc exprimer la fonction de répartition d'une loi Log-Normale à partir de la fonction de répartition d'une loi Normale centrée réduite :

$$F(x) = \varphi\left(\frac{\ln(x) - \mu}{\sigma}\right)$$

avec φ : la fonction de répartition d'une loi Normale centrée réduite.

L'espérance d'une Log-Normale est donnée par :

$$\mathbb{E}(X) = e^{\mu + \frac{\sigma^2}{2}}$$

La variance d'une Log-Normale est donnée par :

$$\mathbb{V}(X) = e^{2\mu + 2\sigma^2} \left[e^{\sigma^2} - 1 \right]$$

La loi Exponentielle

Une variable aléatoire X suit une distribution Exponentielle de paramètres $\lambda > 0$, que l'on note $X \sim \mathcal{E}(\lambda)$, si elle admet une fonction de densité de la forme :

$$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{x \geq 0}$$

La fonction de répartition associée s'exprime pour tout $x \geq 0$ par :

$$F(x) = 1 - e^{-\lambda x}$$

L'espérance d'une Exponentielle est donnée par :

$$\mathbb{E}(X) = \frac{1}{\lambda}$$

La variance d'une Exponentielle est donnée par :

$$\mathbb{V}(X) = \frac{1}{\lambda^2}$$

La loi Gamma

Une variable aléatoire X suit une distribution Gamma de paramètres $k > 0$ et $\theta > 0$, que l'on note $X \sim \mathcal{G}(k, \theta)$, si elle admet une fonction de densité de la forme :

$$f(x) = \frac{x^{k-1}}{\theta^k \Gamma(k)} e^{-\frac{x}{\theta}} \mathbf{1}_{x \geq 0} \quad \text{où } \Gamma(k) = \int_0^{\infty} t^{k-1} e^{-t} dt$$

En définissant pour tout $x \geq 0$ et $a > 0$ la fonction Gamma incomplète supérieure par :

$$\Gamma(x, k) = \int_x^{\infty} e^{-t} t^{k-1} dt$$

Il est possible d'exprimer la fonction de répartition d'une loi Gamma pour tout $x \geq 0$ par :

$$F(x) = 1 - \frac{1}{\Gamma(k)} \Gamma\left(\left(\frac{x}{\theta}\right), k\right)$$

L'espérance d'une Gamma est donnée par :

$$\mathbb{E}(X) = k\theta$$

La variance d'une Gamma est donnée par :

$$\mathbb{V}(X) = k\theta^2$$

3.3.3 Prise en compte du seuil de modélisation : lois tronquées

Pour modéliser la sévérité des sinistres, l'actuaire dispose de données tronquées à gauche au seuil de modélisation. Durant tout le processus de modélisation, l'actuaire doit tenir compte de ce seuil de modélisation, c'est pourquoi les distributions de probabilités sont normalisées pour prendre en considération la troncature. Dans un premier temps, le cadre général d'une troncature à gauche et à droite est présenté, puis le cas particulier d'une seule troncature à gauche est exposé. C'est ce dernier cas qui est utilisé pour modéliser la sévérité des sinistres.

La troncature à gauche et à droite

Soit s et u deux seuils. Le premier, sera en pratique le seuil de modélisation, le second permet d'illustrer la troncature à droite. Notons X la variable aléatoire représentant le montant des sinistres modélisés et F sa fonction de répartition. La fonction de répartition associée aux données tronquées à gauche et à droite aux seuils s et u n'est plus égale à F mais vaut ici pour tout $x \in \mathbb{R}$:

$$F_{s,u}(x) = \mathbb{P}(X \leq x \mid s \leq X \leq u)$$

La fonction de répartition tronquée $F_{s,u}$ peut s'exprimer à l'aide de la fonction de répartition non tronquée F par la relation suivante :

$$F_{s,u}(x) = \begin{cases} 0 & \text{si } x \leq s \\ \frac{F(x)-F(s)}{F(u)-F(s)} & \text{si } s \leq x \leq u \\ 1 & \text{sinon} \end{cases}$$

Démonstration.

$$F_{s,u}(x) = \mathbb{P}(X \leq x \mid s \leq X \leq u)$$

Par la formule de Bayes :

$$= \frac{\mathbb{P}(s \leq X \leq \min(x, u))}{\mathbb{P}(s \leq X \leq u)}$$

En revenant à la fonction de répartition non tronquée F :

$$= \frac{F(\min(x, u)) - F(s)}{F(u) - F(s)} \\ = \begin{cases} 0 & \text{si } x \leq s \\ \frac{F(x)-F(s)}{F(u)-F(s)} & \text{si } s \leq x \leq u \\ 1 & \text{sinon} \end{cases}$$

□

En notant f la fonction de densité des données non tronquées et $f_{s,u}$ celle des données tronquées, il vient par dérivation :

$$f_{s,u}(x) = \frac{f(x)}{F(u) - F(s)} \times \mathbb{1}_{s \leq x \leq u}$$

Cas particulier de la troncature à gauche

Dans la suite, l'estimation des paramètres des lois se fait également sur des données uniquement tronquées à gauche au seuil de modélisation s . Cela revient à faire tendre u vers $+\infty$, et donc à considérer $\lim_{u \rightarrow +\infty} F(u) = 1$. Dans ce cas particulier, et en reprenant le raisonnement précédent, il vient pour la fonction de répartition associée :

$$F_s(x) = \begin{cases} 0 & \text{si } x \leq s \\ \frac{F(x) - F(s)}{1 - F(s)} & \text{si } s \leq x \end{cases}$$

Pour la fonction de densité :

$$f_s(x) = \frac{f(x)}{1 - F(s)} \times \mathbb{1}_{x \geq s}$$

Remarque : La quantité $1 - F(s)$ peut aussi être notée $\bar{F}(s)$.

Dans la suite, la notion de lois tronquées réfère à la troncature à gauche au seuil de modélisation s .

3.3.4 Quantile des lois tronquées

Les distributions statistiques étant tronquées, les quantiles associés sont modifiés pour prendre en compte le seuil de modélisation. Avant d'exprimer le quantile d'une loi tronquée, il convient de définir l'inverse généralisée d'une fonction croissante et continue à droite.

Définition 8. Soit Φ une fonction croissante et continue à droite sur \mathbb{R} . L'inverse généralisée de Φ est définie par :

$$\Phi^{\leftarrow}(y) := \inf\{x | \Phi(x) \geq y\}$$

Dans le cas où Φ est une fonction continue, l'inverse généralisée Φ^{\leftarrow} coïncide avec l'inverse classique Φ^{-1} . Nous pouvons désormais donner la définition d'un quantile d'une fonction de répartition F continue.

Définition 9. Soit F une fonction de répartition. Le quantile d'ordre α de la fonction de répartition F est défini pour $\alpha \in [0, 1]$ par :

$$q(\alpha) = F^{-1}(1 - \alpha) = \inf\{x | F(x) \geq 1 - \alpha\}$$

Appliquons cette définition à une fonction de répartition continue et tronquée à gauche au seuil s . Considérons $y \in [0, 1]$ et $x \in \mathbb{R}$.

Le cas où $x < s$ mène trivialement à un quantile nul.

Supposons alors $x \geq s$ et on résout l'équation suivante en x :

$$\begin{aligned} F_s(x) &= y \\ F(x) - F(s) &= y \times \bar{F}(s) \\ F(x) &= F(s) + y \times \bar{F}(s) \\ x &= F^{-1}\left[F(s) + y \times \bar{F}(s)\right] \end{aligned}$$

Il vient alors : $\forall \alpha \in [0, 1]$

$$q(\alpha) = F^{-1}\left[F(s) + (1 - \alpha) \times \bar{F}(s)\right]$$

Les modèles de probabilités étant présentés, les méthodes utilisées pour estimer leurs paramètres sont décrites dans la section suivante.

3.4 Estimation paramétrique

L'approche paramétrique consiste à sélectionner un modèle connu pour étudier un phénomène aléatoire. Dans notre cas, ce sont les montants de sinistres fournis par une cédante qui sont modélisés. Il convient alors d'estimer les paramètres du modèle retenu en se basant sur les observations disponibles. En termes mathématiques, l'actuaire cherche à estimer un vecteur de paramètres $\theta = (\theta_1, \dots, \theta_p)$ d'une fonction de densité de probabilité, que l'on note f_θ , à partir d'un échantillon de variables aléatoires $X = (X_1, \dots, X_n)$. Les observations à disposition de l'actuaire pour estimer les paramètres du modèle

sont notées $\underline{x} = (x_1, \dots, x_n)$.

Il existe plusieurs méthodes statistiques pour estimer les paramètres du modèle :

- la méthode des moments.
- la méthode du maximum de vraisemblance.
- la méthode des quantiles.
- la méthode bayésienne.

Le choix d'une méthode repose avant tout sur les propriétés recherchées pour l'estimateur des paramètres du modèle. Selon les propriétés voulues, l'actuaire peut avoir à sa disposition plusieurs estimateurs d'un même vecteur de paramètres.

Dans notre cadre d'étude d'estimation des paramètres d'une loi de probabilité, c'est la méthode du maximum de vraisemblance qui est privilégiée, notamment pour ses propriétés statistiques qui sont rappelées dans la suite. Cependant, pour certaines lois de probabilité, il n'existe pas de formules fermées pour obtenir l'estimateur du maximum de vraisemblance. Nous avons alors recours à des algorithmes d'optimisation qui sont initialisés à partir des estimateurs de la méthode des moments.

3.4.1 La méthode du maximum de vraisemblance

Définition

Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire tel que les variables aléatoires $(X_i)_{1 \leq i \leq n}$ sont supposées indépendantes et identiquement distribuées de fonction de densité f_θ . L'échantillon d'observations de la réalisation des variables $(X_i)_{1 \leq i \leq n}$ dont dispose l'actuaire pour estimer les paramètres du modèle est noté (x_1, \dots, x_n) . La vraisemblance associée à l'échantillon s'exprime comme :

$$\mathcal{L}(x_1, \dots, x_n | \theta) = \prod_{i=1}^n f_\theta(x_i)$$

L'estimateur du maximum de vraisemblance de θ est la valeur $\hat{\theta}_n$ qui maximise la fonction de vraisemblance $\mathcal{L}(x_1, \dots, x_n | \theta)$ par rapport à θ :

$$\hat{\theta}_n = \operatorname{argmax}_\theta (\mathcal{L}(x_1, \dots, x_n | \theta))$$

Pour des raisons de commodités de calcul, on introduit généralement la fonction log-vraisemblance définie par :

$$l(\theta) = \sum_{i=1}^n \ln(f_\theta(x_i))$$

En pratique, cette fonction est souvent plus simple à manipuler et donc plus simple à maximiser que la fonction de vraisemblance. L'estimateur du maximum de vraisemblance est défini comme :

$$\hat{\theta}_n = \operatorname{argmax}_\theta (l(\theta))$$

On définit alors les fonction de score comme :

$$\forall j \in \llbracket 1 ; p \rrbracket, \quad S_j(\theta) = \frac{\partial}{\partial \theta_j} \ln(l(\theta))$$

De sorte que maximiser la fonction log-vraisemblance revient à résoudre les p équations suivantes :

$$\forall j \in \llbracket 1 ; p \rrbracket, \quad S_j(\theta) = 0$$

Pour certaines lois, il n'existe pas d'expression analytique pour ces équations, il faut utiliser des méthodes numériques avec des algorithmes d'optimisation. L'algorithme utilisé est celui de Nelder-Mead, il est initialisé à partir de la valeur de l'estimateur obtenu grâce à la méthode des moments.

Propriétés de l'estimateur du maximum de vraisemblance

Sous des hypothèses générales qui ne sont pas détaillées ici mais dont le lecteur intéressé peut consulter Chesneau (2017), l'estimateur du maximum de vraisemblance dispose de nombreuses propriétés. Soit $\hat{\theta}_n$ l'estimateur du maximum de vraisemblance de θ sur un échantillon de taille n .

Propriété 1 (Invariance). *Pour toute fonction bijective g , $g(\hat{\theta}_n)$ est l'estimateur du maximum de vraisemblance de $g(\theta)$.*

Propriété 2 (Asymptotiquement sans biais et efficace). *Pour tout $1 \leq i \leq p$:*

$$\lim_{n \rightarrow +\infty} \mathbb{E}[\hat{\theta}_{n,i}] = \theta_i \quad \text{et} \quad \lim_{n \rightarrow +\infty} \mathbb{V}[\hat{\theta}_{n,i}] = \frac{1}{I_n(\theta)_{i,i}}$$

Avec $I_n(\theta)$ la matrice d'information de Fisher de taille $p \times p$ définie par :

$$I_n(\theta) = \mathbb{E} \left[\left(\frac{\partial l(\theta)}{\partial \theta} \right)^2 \right]$$

Propriété 3 (Consistance). *L'estimateur $\hat{\theta}_n$ converge presque sûrement vers θ :*

$$\mathbb{P} \left(\lim_{n \rightarrow +\infty} \hat{\theta}_n = \theta \right) = 1$$

Propriété 4 (Normalité asymptotique). *La distribution asymptotique de $\hat{\theta}_n$ est une loi normale multivariée de moyenne θ et de matrice de variance-covariance égale à l'inverse de la matrice d'information de Fisher $I_n(\theta)^{-1}$*

$$\hat{\theta}_n \xrightarrow{L} \mathcal{N}(\theta, I_n(\theta)^{-1})$$

Cette propriété de normalité asymptotique est un résultat central dans la construction d'intervalles de confiance pour l'estimation de θ

C'est pour toutes ces propriétés que l'estimateur du maximum de vraisemblance est préféré à l'estimateur des moments dans la suite.

Cas d'une densité tronquée à gauche

Dans la pratique en réassurance, l'actuaire dispose d'un échantillon d'observations tronquées à gauche au seuil de modélisation s fourni par la cédante. La densité tronquée $f_{\theta,s}$ s'exprime alors comme :

$$f_{\theta,s}(x) = \frac{f_{\theta}(x)}{1 - F_{\theta}(s)} \times \mathbf{1}_{x \geq s}$$

La fonction log-vraisemblance associée est définie pour tout $1 \leq i \leq n$ tel que $x_i \geq s$ par :

$$l(\theta) = \sum_{i=1}^n \ln(f_{\theta}(x_i)) - n \ln \left[1 - F_{\theta}(s) \right]$$

3.4.2 La méthode des moments

La méthode des moments est une méthode utilisée afin d'obtenir des estimateurs ponctuels pour les paramètres d'une loi de probabilité. C'est une méthode qui présente certains avantages, elle est simple, intuitive et rapide à implémenter et ne nécessite pas beaucoup de temps de calcul. Cependant, les estimateurs qui découlent de cette méthode ne disposent pas de propriétés aussi intéressantes que ceux issus de la méthode du maximum de vraisemblance.

C'est pourquoi la méthode des moments est souvent utilisée en complément de la méthode du maximum de vraisemblance, en particulier pour initialiser les valeurs des paramètres lorsque la vraisemblance est maximisée par un algorithme itératif.

Définition

La méthode des moments repose sur l'égalité entre les p premiers moments théoriques et les p premiers moments empiriques pour obtenir l'estimateur des moments du vecteur de paramètres $\theta = (\theta_1, \dots, \theta_p)$.

Mathématiquement, en notant X une variable aléatoire et en considérant $k \in \mathbb{N}^*$, le k -ème moment de la distribution de X est défini comme :

$$m_k = \mathbb{E}[X^k]$$

Si X suit une loi de probabilité connue, la valeur des moments m_k s'exprime en fonction des paramètres de la loi. Les estimateurs empiriques des moments m_k se définissent à partir des observations $\underline{x} = (x_1, \dots, x_n)$ comme :

$$\hat{m}_k = \frac{1}{n} \sum_{i=1}^n x_i^k$$

Pour obtenir les estimateurs des paramètres $\theta_1, \dots, \theta_p$, il ne reste qu'à égaliser les p premiers moments théoriques $(m_k)_{1 \leq k \leq p}$ avec les p premiers moments empiriques $(\hat{m}_k)_{1 \leq k \leq p}$

Exemple : Loi Exponentielle

Soit (x_1, \dots, x_n) un échantillon d'observations issues d'une variable aléatoire $X \sim \mathcal{E}(\lambda)$. L'espérance d'une distribution exponentielle vaut $\frac{1}{\lambda}$. L'estimateur des moments $\hat{\lambda}$ est solution de :

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

Finalement, les estimateurs des moments sont populaires pour leur facilité d'utilisation plus que pour leurs propriétés intrinsèques.

3.5 Évaluation des ajustements des lois

L'actuaire dispose de plusieurs distributions de probabilité pour modéliser les sinistres. Son objectif est de sélectionner le modèle qui s'adapte le mieux aux observations. Il existe différents outils pour juger de la qualité d'une distribution de probabilité. Les graphiques Quantile-Quantile permettent d'évaluer graphiquement l'ajustement, les tests et les critères statistiques permettent de quantifier la qualité de l'ajustement. Ces méthodes statistiques sont des outils d'aide à la décision pour l'actuaire qui modélise les sinistres.

3.5.1 Graphique Quantile-Quantile

Le graphique Quantile-Quantile (QQ-plot) est un outil qui permet d'évaluer la pertinence de l'ajustement d'une distribution donnée à un modèle théorique. Pour ce faire, le graphique Quantile-Quantile consiste à représenter les quantiles empiriques en fonction des quantiles théoriques afin d'étudier visuellement la linéarité entre ces deux quantités.

Mathématiquement, considérons (X_1, \dots, X_n) un échantillon de variables aléatoires supposées indépendantes et identiquement distribuées de même loi qu'une variable aléatoire X . Notons (x_1, \dots, x_n) l'échantillon d'observations et $(x_{(1)}, \dots, x_{(n)})$ le même vecteur ordonné par ordre croissant, c'est-à-dire tel que $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$. Notons F_n la fonction de répartition empirique de l'échantillon et F la fonction de répartition théorique que l'on souhaite comparer.

Le graphique Quantile-Quantile s'attache à analyser graphiquement la linéarité entre les points $(F_n^{-1}(p), F^{-1}(p))$ pour des valeurs de $p \in [0, 1]$.

En pratique, le graphique Quantile-Quantile de (X_1, \dots, X_n) est le graphe des points suivants :

$$\left\{ \left(x_{(i)}, F^{-1}\left(\frac{i}{n+1}\right) \right), \text{ pour tout } i \in \llbracket 1; n \rrbracket \right\}$$

Si l'échantillon est identiquement distribué selon la même loi que X , alors les points doivent être alignés sur une droite. Dans ce cas les quantiles empiriques sont égaux aux quantiles théorique de la loi analysée. C'est une condition nécessaire pour estimer que la distribution des données disponibles peut être ajustée par la loi théorique en question. En revanche, si le graphe du QQ-plot est concave, cela indique que l'échantillon est issu d'une distribution à queue plus épaisse, et inversement, un graphe convexe indique que l'échantillon est issu d'une loi à queue plus fine.

Le graphique Quantile-Quantile peut être adapté à de nombreuses distributions de probabilité afin d'effectuer un premier test graphique d'adéquation de loi.

- Pour la loi exponentielle, le graphique consiste à tracer les points :

$$\left\{ \left(x_{(i)}, -\ln\left(1 - \frac{i}{n+1}\right) \right), \text{ pour tout } i \in \llbracket 1; n \rrbracket \right\}$$

- Pour la loi de Pareto, le graphique consiste à tracer les points :

$$\left\{ \left(\ln(x_{(i)}), -\ln\left(1 - \frac{i}{n+1}\right) \right), \text{ pour tout } i \in \llbracket 1; n \rrbracket \right\}$$

- Pour la loi Log-Normale, en utilisant le lien avec la loi Normale centrée réduite et sa fonction de répartition φ le graphique consiste à tracer les points :

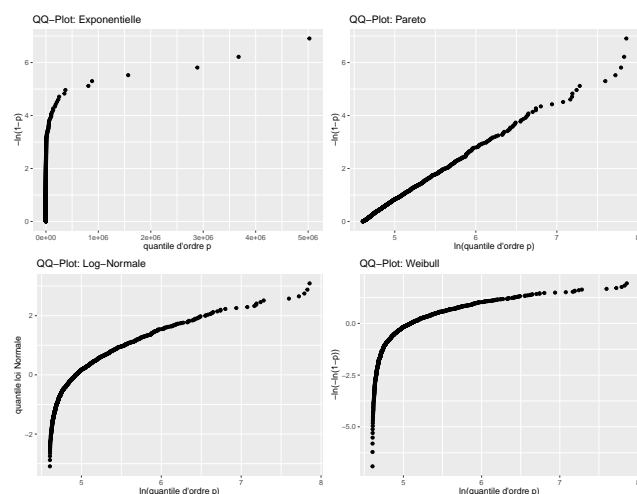
$$\left\{ \left(\ln(x_{(i)}), \varphi^{-1}\left(\frac{i}{n+1}\right) \right), \text{ pour tout } i \in \llbracket 1; n \rrbracket \right\}$$

- Pour la loi de Weibull, le graphique consiste à tracer les points :

$$\left\{ \left(\ln(x_{(i)}), \ln\left(-\ln\left(1 - \frac{i}{n+1}\right)\right) \right), \text{ pour tout } i \in \llbracket 1; n \rrbracket \right\}$$

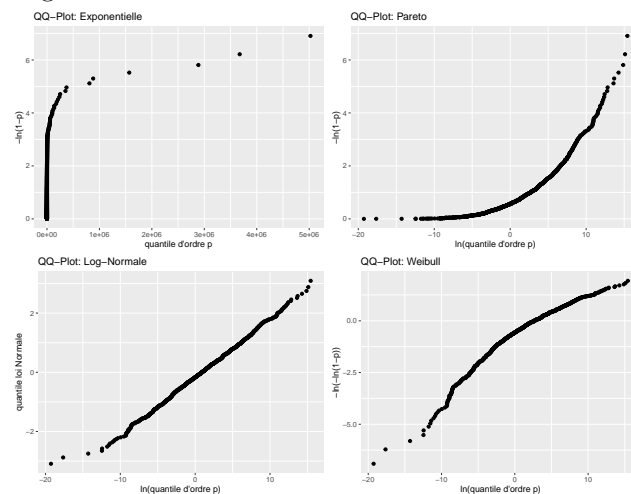
Pour illustrer ces différents graphiques Quantile-Quantile, deux échantillons de 1 000 observations sont simulés. Le premier est issu d'une distribution de Pareto de paramètre $\alpha = 2$ et $A = 100$ et le second d'une distribution Log-Normale de paramètre $\mu = 1$ et $\sigma = 2$.

FIGURE 3.2: QQ-Plot sur simulation d'une loi de Pareto.



Les QQ-Plot exponentielle, log-normale et weibull sont concaves car la loi simulée de l'échantillon a une queue plus épaisse. Le QQ-Plot de pareto est linéaire ce qui fournit une première preuve graphique de l'adéquation de l'échantillon à cette loi.

FIGURE 3.3: QQ-Plot sur simulation d'une loi Log-Normale.



Les QQ-Plot exponentielle et weibull sont concaves car la loi simulée de l'échantillon a une queue plus épaisse. En revanche, le QQ-Plot de pareto est convexe car la loi log-normale simulée a une queue plus fine. Le QQ-Plot log-normale est logiquement linéaire.

3.5.2 Les tests statistiques

Pour modéliser les montants de sinistres de la cédante, nous disposons d'un échantillon (X_1, \dots, X_n) de variables aléatoires supposées indépendantes et identiquement distribuées (i.i.d). Nous supposons que cet échantillon est issu d'une distribution de probabilité inconnue \mathcal{P} . Nous voulons alors tester l'hypothèse selon laquelle \mathcal{P} est égale à une distribution particulière. Pour ce faire, nous avons recours à deux tests d'adéquation, celui de Kolmogorov-Smirnov et celui d'Anderson-Darling.

Pour accepter ou rejeter les tests, nous nous baserons sur la valeur de la p-valeur. La p-valeur est la probabilité que la statistique du test dépasse, sous l'hypothèse nulle, la valeur du seuil de la région critique. Plus cette p-valeur est proche de 0, plus l'hypothèse nulle tend à être rejetée. La p-valeur est utilisée au seuil de 5%, de sorte qu'au delà de ce seuil, l'hypothèse nulle ne peut être rejetée.

Le test de Kolmogorov-Smirnov

Soit (X_1, \dots, X_n) un échantillon i.i.d de fonction de répartition F . Considérons une fonction de répartition F_0 supposée continue. On veut tester l'hypothèse suivante :

$$\begin{cases} H_0 : F = F_0 \\ H_1 : F \neq F_0 \end{cases}$$

Ce test repose sur les propriétés asymptotiques de la fonction de répartition empirique F_n définie comme :

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \leq t}$$

La statistique du test de Kolmogorov-Smirnov étudie la plus grande différence en valeur absolue entre la fonction de répartition empirique F_n et la fonction de répartition F_0 de la loi testée. On note D_n cette statistique, elle s'exprime ainsi :

$$D_n = \sup_t |F_n(t) - F_0(t)|$$

Considérons l'échantillon ordonné $X_{(1)} \leq \dots \leq X_{(n)}$, pour se calculer plus facilement, la statistique D_n peut se réécrire comme :

$$D_n = \max_i (D_n^+ ; D_n^-)$$

où $D_n^+ = |\frac{i}{n} - F_0(X_{(i)})|$ et $D_n^- = |\frac{i-1}{n} - F_0(X_{(i)})|$.

La région critique du test au seuil α est définie par :

$$\{D_n \geq d_\alpha\}$$

où d_α est le quantile $(1 - \alpha)$ de la table de Kolmogorov -Smirnov.

Sous l'hypothèse nulle H_0 ($F = F_0$), la loi de D_n est indépendante de F_0 . Cela permet de tabuler la loi de D_n sous H_0 et donc ne pas avoir à construire une table pour chaque loi de probabilité testée. Plus la valeur de D_n est proche de 0, plus la fonction empirique est proche de la fonction de répartition théorique F_0 postulée, ainsi la p-value sera proche de 1.

Le test d'Anderson Darling

Le test d'Anderson Darling est une alternative du test de Kolmogorov-Smirnov, il repose également sur le calcul d'une distance entre la fonction de répartition empirique de l'échantillon et la fonction de répartition théorique testée. La statistique du test d'Anderson Darling s'exprime comme :

$$A_n^2 = \int_{-\infty}^{+\infty} \frac{(F_n(x) - F_0(x))^2}{F_0(x)(1 - F_0(x))} dF_0(x)$$

Cette statistique s'exprime aussi à l'aide de l'échantillon ordonné $X_{(1)} \leq \dots \leq X_{(n)}$, par la formule suivante :

$$A_n^2 = -n + \sum_{i=1}^n \frac{(2i-1)}{n} \left(\ln(F_0(X_{(i)})) + \ln(1 - F_0(X_{(n+1-i)})) \right)$$

La région critique du test au seuil α est définie par :

$$\{A_n^2 \geq c_\alpha\}$$

où c_α est le quantile $(1 - \alpha)$ de la table d'Anderson-Darling.

Comme pour la statistique D_n du test de Kolmogorov-Smirnov, la loi de A_n^2 est indépendante de F_0 sous l'hypothèse nulle H_0 , ce qui permet de la tabuler.

La statistique d'Anderson Darling prend en compte l'étendue de la distribution et notamment la queue de distribution relative aux montants des sinistres les plus importants. Le test d'Anderson-Darling accorde une importance plus grande aux queues de distribution, il est naturellement plus approprié pour notre problématique de modélisation de la sévérité. En cas de conclusion contradictoire entre le test de Kolmogorov-Smirnov et Anderson Darling, c'est la conclusion de ce dernier qui est préférée.

Remarque

Les données utilisées pour la modélisation sont tronquées à gauche. Les tests sont adaptés pour tenir compte du seuil de modélisation en modifiant la fonction de répartition théorique. Les fonctions de répartitions théoriques testées sont de la forme $F_{0,s}(x) = \frac{F_0(x) - F_0(s)}{1 - F_0(s)}$ avec s le seuil de troncature qui représente le seuil de modélisation.

3.5.3 Critères statistiques

Le critère d'information d'Akaike

Le critère de l'AIC (pour *Akaike Information Criterion* en anglais) est un critère statistique qui s'applique aux modèles estimés par une méthode du maximum de vraisemblance. Il se définit comme :

$$AIC = -2 \ln(\tilde{L}(\theta)) + 2 \times p$$

où \tilde{L} est la vraisemblance maximisée et p le nombre de paramètres à estimer dans le modèle.

L'AIC peut être simplement perçu comme un critère de vraisemblance incluant une pénalité en fonction du nombre de paramètres. Le meilleur modèle étant celui avec la vraisemblance la plus élevée, il faut choisir le modèle ayant l'AIC le plus faible. Un AIC plus faible s'interprète comme le signe d'un modèle plus proche de la véritable loi engendrant les données.

L'AIC représente un compromis entre le biais qui diminue avec le nombre de paramètres, et la parcimonie qui est la volonté de décrire les données avec le moins de paramètres possibles.

Le critère d'information Bayésien

Le critère du BIC (pour *Bayesian Information Criterion* en anglais) fonctionne de façon similaire à l'AIC. Il se définit comme :

$$BIC = -2 \ln(\tilde{L}(\theta)) + \ln(n) \times p$$

où n représente le nombre d'observations.

La pénalité du BIC multiplie le nombre de paramètres par $\ln(n)$ au lieu de le multiplier par 2 comme l'AIC. La pénalité croît plus rapidement en fonction du nombre de paramètres. Dans le cas où $\ln(n) > 2$, la pénalité du BIC est plus importante que celle de l'AIC. Ce critère est en général plus parcimonieux que le critère de l'AIC.

Il faut sélectionner le modèle ayant la valeur du BIC la plus faible.

L'erreur quadratique moyenne

L'erreur quadratique moyenne (pour *Mean Square Error* (MSE) en anglais) est un autre indicateur qui permet d'évaluer la qualité d'un ajustement. C'est une mesure simple de l'écart entre les valeurs observées et les valeurs prédites par un modèle.

Soit $(x_i)_{1 \leq i \leq n}$ les montants des sinistres disponibles, F_n la fonction de répartition empirique associée à ces données et \hat{F}_s la fonction de répartition théorique tronquée à gauche estimée que l'on souhaite évaluer. Le MSE se définit comme la moyenne des écarts au carré entre les valeurs observées $(F_n(x_i))_{1 \leq i \leq n}$ et les valeurs théoriques estimées $(\hat{F}_s(x_i))_{1 \leq i \leq n}$ par :

$$MSE = \frac{1}{n} \sum_{i=1}^n \left(F_n(x_i) - \hat{F}_s(x_i) \right)^2$$

Il faut sélectionner un modèle pour lequel l'erreur quadratique moyenne est la plus faible.

Plusieurs critères d'évaluation d'ajustement des lois ont été présentés, en pratique les graphiques Quantiles-Quantiles peuvent être analysés en premier. Les tests permettent d'orienter son choix en éliminant les lois pour lesquelles les tests sont significatifs au seuil $\alpha = 5\%$. Si les tests conduisent à accepter plusieurs distributions de probabilité, les critères du MSE, de l'AIC et du BIC permettent de faciliter la prise de décision en sélectionnant le modèle pour lequel ces quantités sont minimisées.

L'un des objectifs du mémoire est de compléter l'outil de tarification, en particulier la modélisation de la sévérité des sinistres. Initialement, seule la loi de Pareto était disponible. Le nouvel outil de modélisation permet de modéliser les sinistres par d'autres distributions. L'outil intègre différents critères qui permettent à l'actuaire de sélectionner la distribution la plus adéquate aux données. Cependant, les distributions ajoutées à l'outil de tarification appartiennent au domaine d'attraction de Gumbel, caractérisé par des lois de probabilité à queue fine. Ces lois représentent mal le risque dans les tranches hautes non-travaillantes. Un modèle à deux lois est proposé pour palier à ce problème.

3.6 Modélisation par deux distributions

En réassurance, la modélisation des sinistres peut être malaisée car la plage de valeurs des montants de sinistres peut-être large, et les quelques sinistres les plus extrêmes peuvent impacter la modélisation. Une modélisation classique à une loi peut alors s'avérer mauvaise. Une attention particulière doit être apportée aux sinistres extrêmes sans pour autant focaliser la modélisation sur ces sinistres. C'est pourquoi une modélisation des montants de sinistres par deux lois est proposée. Le modèle à deux lois implémenté dans le nouvel outil mixte une loi usuelle avec une loi de Pareto Généralisée. Une approche alternative est exposée avec la construction d'un modèle à deux lois qui mixte une loi mélange d'Erlang et une loi de Pareto Généralisée.

L'approche d'une modélisation de la sinistralité par deux distributions consiste à construire une loi de probabilité basée sur deux distributions raccordées à certain un seuil.

La difficulté principale réside dans le choix du seuil, que l'on qualifiera par la suite de "seuil des extrêmes". Le choix est basé sur la théorie des valeurs extrêmes, en particulier sur le théorème de Pickands, Balkema, De Hann. On rappelle que ce théorème suggère qu'il existe un seuil au-delà duquel les observations se comportent selon une distribution de Pareto Généralisée. Le seuil des extrêmes sera déterminé grâce aux différents outils qui ont été présentés (*Hill Plot, Mean Excess Plot* etc..). Dans toute la suite de cette section, sauf mention contraire, le seuil des extrêmes est noté u .

3.6.1 Distribution usuelle et loi de Pareto Généralisée

L'idée générale de la démarche consiste à modéliser la sinistralité à l'aide de deux fonctions de densité f_1 et f_2 puis d'en former une troisième f_3 en raccordant les deux premières au niveau du seuil des extrêmes. Une façon de faire est de considérer la densité mixte f_3 comme une combinaison linéaire convexe de f_1 et de f_2 de la forme suivante :

$$f_3 = \alpha f_1 + (1 - \alpha) f_2, \quad \text{avec } \alpha \in [0, 1] \quad (3.1)$$

Une fonction de densité est une fonction positive, ce qui est le cas pour f_3 comme combinaison linéaire de fonctions positives. De plus, l'intégrale d'une fonction de densité doit valoir 1 sur son domaine de définition, ici :

$$\begin{aligned} \int_{\mathbb{R}} f_3 &= \int_{\mathbb{R}} \left(\alpha f_1 + (1 - \alpha) f_2 \right) \\ &= \alpha \int_{\mathbb{R}} f_1 + (1 - \alpha) \int_{\mathbb{R}} f_2 \end{aligned}$$

Les fonctions f_1 et f_2 étant des densités, il vient que :

$$\begin{aligned} \int_{\mathbb{R}} f_3 &= \alpha + (1 - \alpha) \\ &= 1 \end{aligned}$$

L'objectif à ce stade est de choisir une valeur de α . Cette valeur n'est pas unique et dépend des attentes pour le modèle. La première approche est celle retenue dans l'outil de modélisation final, la seconde approche permet d'illustrer le fait que le choix du coefficient α peut être différent selon ce que l'actuaire recherche pour son modèle.

Modèle retenu

Dans le cadre de ce mémoire, deux modèles sont comparés. Le premier où la sinistralité est modélisée par un modèle classique à une loi tronquée au seuil de modélisation s et le second où la sinistralité est modélisée par un modèle plus technique à deux lois. Dans cette version retenue du modèle à deux lois, la modélisation effectuée par le modèle à une loi est conservée pour ajuster la sinistralité sous le seuil.

Dans un premier temps, l'actuaire modélise l'ensemble de la sinistralité à l'aide d'une loi usuelle classique et valide l'ajustement grâce aux différents outils présentés. Dans ce modèle à une loi, le seuil de modélisation est pris en compte, la densité associée est de la forme :

$$f_s(x) = \frac{f(x)}{\bar{F}(s)} \times \mathbf{1}_{x \geq s}$$

Dans un second temps, l'actuaire sélectionne le seuil des extrêmes noté u , et modélise les montants de sinistres supérieurs à ce seuil par une distribution de Pareto Généralisée.

Ensuite, l'actuaire tronque à droite au seuil u la densité sélectionnée durant la première étape. Mathématiquement, en reprenant les notations du modèle à deux lois introduit en (3.1) :

$$f_1(x) = \beta \times \frac{f(x)}{\bar{F}(s)} \times \mathbf{1}_{[s,u]}. \text{ La densité tronquée à gauche du modèle à une loi est tronquée à droite au seuil des extrêmes. Pour conserver une densité, un coefficient de proportionnalité } \beta \text{ est introduit.}$$

$$f_2(x) = f^{\mathcal{GPD}}(x) \text{ pour tout } x \geq u.$$

Il est possible de déterminer la valeur de β qui permet à la fonction f_1 d'être une densité sur \mathbb{R} . En effet, en intégrant f_1 sur \mathbb{R} il vient que :

$$\begin{aligned} \int_{\mathbb{R}} f_1(t) dt &= \frac{\beta}{\bar{F}(s)} \times \int_s^u f(t) dt \\ 1 &= \frac{\beta}{\bar{F}(s)} \times [F(u) - F(s)] \\ \Rightarrow \beta &= \frac{\bar{F}(s)}{F(u) - F(s)} \end{aligned}$$

Ainsi, la densité du modèle à deux lois s'écrit comme la combinaison suivante :

$$f_3(x) = \alpha \times \frac{\bar{F}(s)}{F(u) - F(s)} \times \frac{f(x)}{\bar{F}(s)} \times \mathbf{1}_{[s,u]} + (1 - \alpha) f^{\mathcal{GPD}}(x) \times \mathbf{1}_{x \geq u}$$

En choisissant $\alpha = \frac{1}{\beta} = \frac{F(u) - F(s)}{\bar{F}(s)}$ il vient que :

$$f_3(x) = \underbrace{\frac{f(x)}{\bar{F}(s)}}_{f_s(x)} \times \mathbf{1}_{[s,u]} + \left(1 - \frac{F(u) - F(s)}{\bar{F}(s)}\right) f^{\mathcal{GPD}}(x) \times \mathbf{1}_{x \geq u}$$

De cette façon, la modélisation effectuée par le modèle à une loi est conservée dans le modèle à deux lois pour la sinistralité inférieure au seuil des extrêmes u . La fonction de répartition de ce modèle à deux lois s'exprime ainsi :

$$F_3(x) = \begin{cases} \frac{F(x) - F(s)}{\bar{F}(s)} & \text{si } x \in [s, u] \\ \frac{F(u) - F(s)}{\bar{F}(s)} + \left(1 - \frac{F(u) - F(s)}{\bar{F}(s)}\right) \times F^{\mathcal{GPD}}(x) & \text{si } x \geq u \end{cases}$$

Il reste à vérifier que la fonction F_3 est bien une fonction de répartition :

1. $\lim_{x \rightarrow -\infty} F_3(x) = \lim_{x \rightarrow s} \frac{F(x) - F(s)}{\bar{F}(s)} = 0.$
2. $\lim_{x \rightarrow +\infty} F_3(x) = \frac{F(u) - F(s)}{\bar{F}(s)} + \left(1 - \frac{F(u) - F(s)}{\bar{F}(s)}\right) \times \lim_{x \rightarrow +\infty} F^{\mathcal{GPD}}(x) = 1$
3. De plus la fonction F_3 est croissante et continue sur son domaine de définition.

Finalement, cette construction présente l'avantage de conserver le travail effectué par la modélisation classique à une loi et accorde une place particulière à la sinistralité extrême. Cette double modélisation, qui repose en partie sur la théorie des valeurs extrêmes permet ensuite d'améliorer la tarification des tranches hautes peu ou pas travaillantes.

Autre modèle possible

Le choix du modèle à deux lois peut également se construire de cette façon. La première étape consiste à définir le seuil des extrêmes noté u à partir des outils de la théorie des valeurs extrêmes. Si en théorie le choix du seuil peut sembler facile, en pratique, il peut être délicat de réussir à le sélectionner. Une attention particulière doit être apportée à cette étape de manière à ce que le seuil des extrêmes retenu tend à être cohérent avec les différents résultats fournis par les outils disponibles.

Ensuite, les montants de sinistres inférieurs au seuil des extrêmes sont modélisés par une loi classique tronquée à gauche au seuil de modélisation s et à droite au seuil des extrêmes u . La sinistralité supérieure au seuil des extrêmes u est modélisée via une loi \mathcal{GPD} .

En résumé, l'actuaire modélise les montants de sinistres sur l'intervalle $[s, u]$ et sur l'intervalle $[u, +\infty[$ puis il définit une nouvelle densité à partir des deux précédentes. En reprenant les notations du modèle à deux lois introduit en [\(3.1\)](#) :

$$f_1(x) = \frac{f(x)}{F(u) - F(s)} \times \mathbb{1}_{s \leq x \leq u}$$

$$f_2(x) = f^{\mathcal{GPD}}(x) \text{ pour tout } x \geq u.$$

Il vient alors pour densité du modèle à deux lois :

$$f_3 = \alpha \times \frac{f(x)}{F(u) - F(s)} \times \mathbb{1}_{s \leq x \leq u} + (1 - \alpha) \times f^{\mathcal{GPD}}(x) \times \mathbb{1}_{x \geq u}$$

En choisissant $\alpha = F(u) - F(s)$ de sorte que :

$$f_3 = f(x) \times \mathbb{1}_{s \leq x \leq u} + \left(1 + F(s) - F(u)\right) \times f^{\mathcal{GPD}}(x) \times \mathbb{1}_{x \geq u}$$

L'inconvénient de ce modèle est que la modélisation à une loi classique n'est pas prise en compte, cela demande un nouvel ajustement de la sinistralité sur l'intervalle $[s, u]$. En période de renouvellement, le temps accordé pour la modélisation des sinistres d'une cédante est restreint, c'est pourquoi cette nouvelle modélisation n'est pas adaptée au besoin métier. L'avantage du modèle retenu est que la sinistralité sous le seuil des extrêmes est modélisée de la même manière dans le modèle à une et deux lois. Seule la queue de distribution est ajustée différemment, cela permet d'étudier l'impact du modèle à deux lois sur la tarification des tranches hautes. La sous-section suivante expose une approche alternative pour construire un modèle à deux distributions.

3.6.2 Approche alternative : mélange d'Erlang et loi de Pareto Généralisée

L'objectif de l'actuaire est d'estimer une fonction de répartition qui permet de modéliser convenablement l'ensemble de la sinistralité, aussi bien dans la partie centrale que dans la queue de distribution des sinistres. Le modèle à deux lois présenté dans la sous-section précédente permet d'utiliser une distribution usuelle (Log-Normale, Weibull etc) pour modéliser la sinistralité sous le seuil des extrêmes, puis de compléter la distribution par une loi de Pareto Généralisée plus adaptée pour modéliser la queue de distribution.

L'alternative proposée par [Albrecher et al. \(2017\)](#) consiste à construire un modèle à deux lois où la sinistralité sous le seuil des extrêmes est modélisée par une loi mélange d'Erlang. La justification de l'utilisation de cette loi pour modéliser les sinistres sous le seuil repose sur le théorème de [Tijms \(1994\)](#) qui énonce qu'il est possible d'approximer toute distribution avec support positif par une loi mélange d'Erlang. Les sinistres supérieurs au seuil sont toujours modélisés par une loi de Pareto Généralisée.

Loi d'Erlang et loi mélange d'Erlang

La distribution d'Erlang est une distribution Gamma où le paramètre de forme est un entier naturel strictement positif. Sa densité s'exprime comme :

$$f_E(x) = \frac{\lambda^r x^{r-1}}{(r-1)!} e^{-\lambda x} \mathbb{1}_{x>0}, \quad \lambda > 0 \text{ et } r \in \mathbb{N}^*$$

La densité de la loi mélange d'Erlang s'exprime comme la somme de densités d'Erlang par :

$$f_{ME}(x) = \sum_{j=1}^M \beta_j \frac{\lambda^{r_j}}{(r_j-1)!} x^{r_j-1} e^{-\lambda x}, \quad x > 0$$

où les entiers (r_1, \dots, r_M) sont les paramètres de forme tels que $0 < r_1 < r_2 < \dots < r_M$. Le vecteur $\beta = (\beta_1, \dots, \beta_M)$ est le vecteur des poids du mélange qui satisfont l'égalité suivante :

$$\sum_{j=1}^M \beta_j = 1$$

La fonction de répartition de la loi mélange d'Erlang s'exprime par :

$$F_{ME}(x) = \sum_{j=1}^M \beta_j \left(1 - \sum_{n=0}^{r_j-1} e^{-\lambda x} \frac{(\lambda x)^n}{n!} \right)$$

La densité et la fonction de répartition de la loi mélange d'Erlang étant présentées, le paragraphe suivant s'attache à exposer la construction de la fonction de répartition mixte.

Construction de la distribution mixte : mélange Erlang et loi de Pareto Généralisée

Pour construire ce modèle, la valeur du seuil des extrêmes qui permet de raccorder les deux distributions doit être sélectionnée. Le choix de ce seuil repose sur les différents outils issus de la théorie des valeurs extrêmes. Ce seuil de raccordement est noté u . Le seuil de modélisation, qui permet de prendre en compte la troncature à gauche des données, est noté s .

Le modèle consiste à modéliser les montants de sinistres entre s et u par une loi de mélange d'Erlang tronquée à gauche et à droite. Les sinistres supérieurs au seuil des extrêmes sont modélisés par une loi de Pareto Généralisée. Les deux densités sont ensuite pondérées par un coefficient α , comme exposé en [3.1](#) pour former la densité du modèle. Mathématiquement, la densité du modèle ainsi construit est définie par :

$$f_3(x) = \alpha \left(\frac{f_{ME}(x)}{F_{ME}(u) - F_{ME}(s)} \right) \mathbb{1}_{x \in [s, u]} + (1 - \alpha) f^{\mathcal{GPD}} \mathbb{1}_{x > u}$$

La fonction de répartition du modèle s'exprime par :

$$F_3(x) = \begin{cases} \alpha \left(\frac{F_{ME}(x) - F_{ME}(s)}{F_{ME}(u) - F_{ME}(s)} \right) & \text{si } x \in [s, u] \\ \alpha + (1 - \alpha) F^{\mathcal{GPD}}(x) & \text{si } x \geq u \end{cases}$$

Albrecher et al. (2017) propose de prendre comme coefficient de pondération α la proportion du nombre de sinistres inférieurs au seuil des extrêmes. Autrement dit, si le seuil des extrêmes correspond à un nombre d'excès k_n , le coefficient de pondération se détermine par $\alpha = 1 - \frac{k_n}{n}$.

Il est malaisé d'estimer les paramètres de la loi de mélange d'Erlang en maximisant directement la vraisemblance associée. Pour estimer les paramètres des lois de mélange, un algorithme d'Espérance-Maximisation, présenté par Dempster et al. (1977), est souvent utilisé. La première étape de cet algorithme consiste à initialiser la valeur des paramètres. Pour la deuxième étape, il faut exprimer l'espérance conditionnelle de la log-vraisemblance par rapport aux observations et aux paramètres estimés lors de l'itération précédente. La troisième étape consiste à déterminer les paramètres qui maximisent l'espérance conditionnelle. La deuxième et troisième étape sont répétées jusqu'à convergence des estimations. Un critère d'arrêt permet d'éviter les problèmes de sur-apprentissage, ici c'est le BIC. Pour de plus amples informations sur le fonctionnement de l'algorithme, le lecteur intéressé peut consulter Goffard (2019).

Les paramètres de la loi de Pareto Généralisée sont estimés par la méthode classique du maximum de vraisemblance dans l'approche *Peaks Over the Threshold* présentée en 2.3.1.

Finalement, deux méthodes pour construire le modèle à deux lois ont été présentées. La première utilise une loi usuelle pour modéliser la sinistralité sous le seuil des extrêmes. Ce modèle présente l'avantage de conserver la modélisation du modèle à une loi. La deuxième méthode permet d'introduire la loi de mélange d'Erlang. Cette loi présente l'avantage d'être plus flexible qu'une loi usuelle, mais nécessite suffisamment de données pour pouvoir estimer convenablement les paramètres associés. Dans le contexte de réassurance, il arrive régulièrement que les cédantes fournissent très peu de données pour tarifier les traités en excédent de sinistre. En ce sens, dans l'outil de modélisation déployé, seul le modèle à deux lois avec une distribution usuelle est implémenté pour l'instant. Dans la suite, ce modèle est nommé *modèle à deux lois usuelles*. Le modèle faisant intervenir la distribution de mélange d'Erlang fait l'objet d'une comparaison dans le dernier chapitre d'application sur des données réelles. Ce modèle est nommé dans la suite *modèle d'Erlang*.

La prochaine section s'attache à étudier la construction des quantiles dans les modèles à deux lois.

3.6.3 Quantile des modèles à deux lois

Expression du quantile du modèle à deux lois

Dans les modèles à deux lois, la fonction de répartition s'exprime sous la forme générale suivante :

$$F_3(x) = \begin{cases} \alpha F_1(x) & \text{si } x \in [s, u] \\ \alpha + (1 - \alpha) \times F^{\mathcal{GPD}}(x) & \text{si } x \geq u \end{cases}$$

De cette façon, pour tout $x \geq u$ la quantité $\alpha + (1 - \alpha) \times F^{\mathcal{GPD}}(x)$ peut-être utilisée comme approximation de la queue de distribution de la variable aléatoire X représentant la sinistralité.

Pour déterminer la fonction quantile associée à cette modélisation à deux lois, il faut être en mesure d'inverser la fonction de répartition F_3 . Considérons alors $x \geq u$ et $y \in [0, 1]$ tel que :

$$\begin{aligned}
F_3(x) &= y \\
\alpha + (1 - \alpha) \times F^{\mathcal{GPD}}(x) &= y \\
\Rightarrow x &= F_{\mathcal{GPD}}^{-1}\left(\frac{y - \alpha}{1 - \alpha}\right)
\end{aligned}$$

or si $X \sim \mathcal{GPD}(\mu, \gamma, \sigma)$ alors :

$$F_{\mathcal{GPD}}^{-1}(y) = \begin{cases} \mu + \frac{\sigma}{\gamma} \left[(1 - y)^{-\gamma} - 1 \right] & \text{si } \gamma \neq 0 \\ \mu - \sigma \ln(1 - y) & \text{si } \gamma = 0 \end{cases}$$

Ainsi, en reprenant le calcul du quantile, il vient $\forall \gamma \neq 0$:

$$\begin{aligned}
x &= u + \frac{\sigma}{\gamma} \left[\left(1 - \frac{y - \alpha}{1 - \alpha} \right)^{-\gamma} - 1 \right] \\
&= u + \frac{\sigma}{\gamma} \left[\left(\frac{1 - y}{1 - \alpha} \right)^{-\gamma} - 1 \right]
\end{aligned}$$

Finalement, on obtient $F_3^{-1}(y) = u + \frac{\sigma}{\gamma} \left[\left(\frac{1 - y}{1 - \alpha} \right)^{-\gamma} - 1 \right]$ donc un quantile d'ordre y :

$$q(y) = F_3^{-1}(1 - y) = u + \frac{\sigma}{\gamma} \left[\left(\frac{y}{1 - \alpha} \right)^{-\gamma} - 1 \right] \quad (3.2)$$

Intervalle de confiance pour le quantile du modèle à deux lois

L'expression précédente du quantile dans le modèle à deux lois, permet en remplaçant la valeur des paramètres par leurs estimations, d'obtenir une estimation du quantile. Cependant, l'actuaire peut souhaiter obtenir une information plus large et vouloir une mesure de l'incertitude sur le quantile estimé. C'est pourquoi l'actuaire peut être amené à construire des intervalles de confiance. Dans ce cadre d'étude, la construction d'intervalles de confiance repose sur la méthode Delta et les propriétés asymptotiques des paramètres de la loi de Pareto Généralisée estimés par la méthode du maximum de vraisemblance.

La méthode Delta La méthode Delta est une méthode permettant d'obtenir une approximation de la distribution asymptotique de la transformée d'une variable aléatoire, le plus souvent elle même asymptotiquement normale. On peut considérer cette méthode comme une extension du théorème central limite. Cette méthode est présentée dans le cadre général :

Théorème 4 (Méthode Delta). *Soit (X_n) une suite de variables aléatoires et (v_n) une suite de réels tendant vers $+\infty$. Supposons qu'il existe un réel a et une variable X tels que :*

$$v_n(X_n - a) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$$

Si g est une fonction dérivable au point a , alors :

$$v_n(g(X_n) - g(a)) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} g'(a)X$$

Dans la suite, la méthode Delta est utilisée dans le cas d'un vecteur aléatoire dans un cadre asymptotiquement normale.

Corolaire 1 (Méthode Delta, cas multivarié). *Soit X_1, \dots, X_n une suite de vecteurs aléatoires de \mathbb{R}^d , $g : \mathbb{R}^d \rightarrow \mathbb{R}^s$ une fonction différentiable en θ . Supposons que $\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}_d(0, \Sigma)$ où $\mathcal{N}_d(0, \Sigma)$ désigne une loi normale d -dimensionnelle centrée de matrice de variance covariance Σ . Dans ce cas, la méthode Delta s'écrit :*

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{\mathcal{L}} \mathcal{N}_s\left(0, \mathcal{J}_{g(\theta)} \Sigma \mathcal{J}_{g(\theta)}^T\right)$$

où $\mathcal{J}_{g(\theta)}$ est la matrice jacobienne de g en θ

Intervalle de confiance du quantile L'objectif est de déterminer un intervalle de confiance d'un quantile du modèle à deux lois. Les paramètres (γ, σ) sont estimés par la méthode du maximum de vraisemblance, et d'après (Smith, 1984), on rappelle que pour tout $\gamma > -\frac{1}{2}$, ces estimateurs du maximum de vraisemblance sont asymptotiquement normaux de variance asymptotique :

$$\mathbb{V}_{as} \begin{pmatrix} \hat{\sigma} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} 2\sigma^2(1+\gamma) & -\sigma(1+\gamma) \\ -\sigma(1+\gamma) & (1+\gamma)^2 \end{pmatrix}$$

où \mathbb{V}_{as} est l'inverse de l'information de Fisher. Il y a donc la convergence en loi suivante :

$$\sqrt{n} \begin{pmatrix} \hat{\sigma} - \sigma \\ \hat{\gamma} - \gamma \end{pmatrix} \xrightarrow{L} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbb{V}_{as} \begin{pmatrix} \hat{\sigma} \\ \hat{\gamma} \end{pmatrix} \right)$$

D'après l'équation (3.2), le quantile du modèle à deux lois peut être estimé par :

$$\hat{q}(y) = u + \frac{\hat{\sigma}}{\hat{\gamma}} \left[\left(\frac{y}{1-\alpha} \right)^{-\hat{\gamma}} - 1 \right]$$

Pour des raisons de commodité, posons $a = \frac{y}{1-\alpha}$ de sorte que $\hat{q}(y) = u + \frac{\hat{\sigma}}{\hat{\gamma}} [a^{-\hat{\gamma}} - 1]$. La méthode Delta est appliquée en dérivant l'expression du quantile estimé par rapport à $\hat{\gamma}$ et $\hat{\sigma}$:

$$\begin{cases} \frac{\partial \hat{q}}{\partial \hat{\sigma}} = \frac{a^{-\hat{\gamma}} - 1}{\hat{\gamma}} \\ \frac{\partial \hat{q}}{\partial \hat{\gamma}} = -\frac{\hat{\sigma}}{\hat{\gamma}} \left[\frac{a^{-\hat{\gamma}} - 1}{\hat{\gamma}} + \ln(a) \times a^{-\hat{\gamma}} \right] \end{cases}$$

Ainsi, la matrice Jacobienne de \hat{q} est la suivante :

$$\mathcal{J}(\hat{q}) = \begin{pmatrix} \frac{a^{-\hat{\gamma}} - 1}{\hat{\gamma}} & -\frac{\hat{\sigma}}{\hat{\gamma}} \left[\frac{a^{-\hat{\gamma}} - 1}{\hat{\gamma}} + \ln(a) \times a^{-\hat{\gamma}} \right] \end{pmatrix}$$

Il ne reste plus que deux produits matriciels à calculer, le premier qui vaut :

$$\mathcal{J}(\hat{q}) \times \mathbb{V}_{as} \begin{pmatrix} \hat{\sigma} \\ \hat{\gamma} \end{pmatrix} = \left(2 \times \left(\frac{\partial \hat{q}}{\partial \hat{\sigma}} \right) \hat{\sigma}^2 (1 + \hat{\gamma}) - \left(\frac{\partial \hat{q}}{\partial \hat{\gamma}} \right) \hat{\sigma} (1 + \hat{\gamma}) + \left(\frac{\partial \hat{q}}{\partial \hat{\gamma}} \right) (1 + \hat{\gamma})^2 \right)$$

Et le deuxième :

$$\mathcal{J}(\hat{q}) \times \mathbb{V}_{as} \begin{pmatrix} \hat{\sigma} \\ \hat{\gamma} \end{pmatrix} \times \mathcal{J}^T(\hat{q}) = 2 \times \left(\frac{\partial \hat{q}}{\partial \hat{\sigma}} \right)^2 \hat{\sigma}^2 (1 + \hat{\gamma}) - 2 \times \left(\frac{\partial \hat{q}}{\partial \hat{\gamma}} \right) \left(\frac{\partial \hat{q}}{\partial \hat{\sigma}} \right) \hat{\sigma} (1 + \hat{\gamma}) + \left(\frac{\partial \hat{q}}{\partial \hat{\gamma}} \right)^2 (1 + \hat{\gamma})^2$$

Finalement, la formule de la variance asymptotique de l'estimateur du quantile est la suivante :

$$\mathbb{V}_{as}(\hat{q}) = 2 \times \left(\frac{\partial \hat{q}}{\partial \hat{\sigma}} \right)^2 \hat{\sigma}^2 (1 + \hat{\gamma}) - 2 \times \left(\frac{\partial \hat{q}}{\partial \hat{\gamma}} \right) \left(\frac{\partial \hat{q}}{\partial \hat{\sigma}} \right) \hat{\sigma} (1 + \hat{\gamma}) + \left(\frac{\partial \hat{q}}{\partial \hat{\gamma}} \right)^2 (1 + \hat{\gamma})^2$$

Un intervalle de confiance asymptotique de niveau δ pour le quantile d'ordre y du modèle à deux lois $q(y)$ est donc :

$$\left[\hat{q}(y) \pm \frac{\sqrt{\mathbb{V}_{as}(\hat{q}(y))}}{\sqrt{n}} \times q^{\mathcal{N}} \left(1 - \frac{\delta}{2} \right) \right]$$

avec $q^{\mathcal{N}} \left(1 - \frac{\delta}{2} \right)$ le quantile d'ordre $1 - \frac{\delta}{2}$ d'une loi normale centrée réduite et n le nombre d'excès considérés pour estimer les paramètres de la loi de Pareto Généralisée.

Cette section a permis d'exposer la modélisation à deux lois. Le nouvel outil de tarification intègre la construction du modèle à deux lois avec l'utilisation d'une loi usuelle pour modéliser la sinistralité sous le seuil des extrêmes. L'actuaire peut désormais alourdir la queue de distribution des lois à queue fine et ainsi représenter plus convenablement le risque porté par les tranches hautes non-travaillantes.

Lors de la tarification des traités en excédent de sinistre, l'actuaire doit être en mesure d'estimer le montant à charge de la réassurance pour une tranche de réassurance donnée. La section suivante présente la méthode pour calculer cette estimation.

3.7 Calcul du montant à charge dans une tranche

Soit X la variable aléatoire représentant le montant *from ground up* d'un sinistre, c'est-à-dire sans l'application des franchises du traité. Supposons de plus qu'il n'y a ni clause de limitation annuelle ni franchise annuelle. Considérons alors un traité en excédent de sinistre de franchise D et de portée C . Ce traité est noté $C \text{ XS } D$. Soit Y la variable aléatoire représentant le coût à la charge du réassureur après application des conditions du traité. Si le coût du sinistre est inférieur au montant de la franchise D , le réassureur ne supporte aucune charge. Si le montant de sinistre est compris entre D et $D + C$, le réassureur a à sa charge le montant $X - D$. Enfin, si le montant du sinistre est supérieur à $D + C$ alors le montant à charge du réassureur est plafonné à C . Mathématiquement, l'expression du montant à charge du réassureur s'exprime comme :

$$Y = \begin{cases} 0 & \text{si } X \leq D \\ X - D & \text{si } D \leq X \leq D + C \\ C & \text{si } X \geq D + C \end{cases}$$

La portée du traité peut être infinie, c'est notamment le cas des traités couvrant de la responsabilité automobile. La définition précédente s'adapte naturellement au cas où $C = +\infty$

On se place dans un modèle collectif comme présenté à la sous-section [3.1.2](#). En notant S le montant total à charge du réassureur et N la variable représentant le nombre de sinistres, l'actuaire souhaite estimer la quantité suivante :

$$\mathbb{E}(S) = \mathbb{E}(Y) \times \mathbb{E}(N)$$

Le cadre de l'étude porte sur l'amélioration de la modélisation des montants des sinistres d'une cédante. La fréquence de sinistres ne fait pas l'objet d'une nouvelle modélisation. C'est pourquoi dans toute la suite, la fréquence est supposée unitaire, de sorte que l'estimation de la charge totale S revient à estimer le coût d'un sinistre pour le réassureur :

$$\mathbb{E}(S) = \mathbb{E}(Y) \times 1$$

Dans la suite, nous allons exposer le fait qu'il est possible de tarifier tout type de traité en excédent de sinistre à l'aide des modélisations à une et deux lois effectuées.

3.7.1 Modèle à une loi

Dans ce modèle, la sinistralité est entièrement modélisée à l'aide d'une seule distribution de probabilité. Pour tarifier un traité $C \text{ XS } D$, il convient de distinguer deux cas : d'une part le cas où la portée du traité est illimitée, d'autre part le cas où la portée du traité est finie.

Cas d'un traité à portée infinie

En reprenant les différentes notations et en notant $\pi(D, \infty)$ le coût moyen à charge d'un traité $\infty \text{ XS } D$, il vient :

$$\pi(D, \infty) = \mathbb{E}((X - D)^+) = \int_D^\infty (x - D) f_s(x) dx = \int_D^\infty \bar{F}_s(x) dx$$

avec les fonctions f_s et F_s respectivement la fonction de densité et la fonction de répartition de la variable aléatoire X avec prise en compte de la troncature à gauche au seuil de modélisation s .

Cas d'un traité à portée finie

De la même façon, lorsque la portée du traité est finie, c'est à dire $C \in \mathbb{R}$, la formule précédente s'adapte en remarquant que le montant à charge du réassureur peut s'écrire $Y = (X - D)\mathbb{1}_{D \leq X \leq D+C} + C\mathbb{1}_{X \geq D+C}$

Dans ce cas, en notant $\pi(D, C)$ le coût moyen à charge d'un traité C XS D, il vient que :

$$\begin{aligned}\pi(D, C) &= \mathbb{E}(Y) \\ &= \mathbb{E}((X - D)\mathbf{1}_{D \leq X \leq D+C}) + C\mathbb{E}(\mathbf{1}_{X \geq D+C}) \\ &= \mathbb{E}((X - D)\mathbf{1}_{D \leq X \leq D+C}) + \underbrace{C\mathbb{P}(X \geq D + C)}_{=\bar{F}_s(D+C)}\end{aligned}$$

Avec :

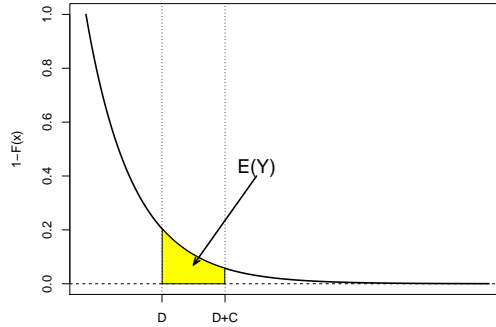
$$\begin{aligned}\mathbb{E}((X - D)\mathbf{1}_{D \leq X \leq D+C}) &= \int_D^{D+C} (x - D)f_s(x)dx \\ &= \int_D^{D+C} xf_s(x)dx - D \int_D^{D+C} f_s(x)dx \\ &= \left[x(F_s(x) - 1) \right]_D^{D+C} - \int_D^{D+C} (F_s(x) - 1)dx - D \int_D^{D+C} f_s(x)dx \\ &= D\bar{F}_s(D) - (D + C)\bar{F}_s(D + C) + \int_D^{D+C} \bar{F}_s(x)dx - D \left[\bar{F}_s(D) - \bar{F}_s(D + C) \right] \\ &= \int_D^{D+C} \bar{F}_s(x)dx - C\bar{F}_s(D + C)\end{aligned}$$

Finalement, on obtient la formule suivante :

$$\pi(D, C) = \int_D^{D+C} \bar{F}_s(x)dx$$

Graphiquement, le coût moyen dans une tranche de réassurance correspond à l'aire sous la courbe de la fonction de survie de la variable aléatoire représentant la sinistralité.

FIGURE 3.4: Illustration de la prime pure d'un C XS D



Les démonstrations des formules fermées des coûts dans la tranche dans le modèle à une loi sont disponibles dans l'Annexe [A](#).

Avec les deux formules précédemment définies, il est possible d'estimer le coût à la charge du réassureur pour tous les traités en excédent de sinistre avec le modèle à une loi. Dans la suite, le coût moyen à charge du réassureur dans le modèle à deux lois est défini.

Remarque Il est bon de noter qu'il existe l'égalité suivante $\pi(D, C) = \pi(D, \infty) - \pi(D + C, \infty)$

3.7.2 Modèle à deux lois usuelles

Dans ce modèle, un seuil de sinistres extrêmes, noté u est déterminé, puis en dessous de ce seuil, la sinistralité est modélisée via la distribution usuelle tronquée à gauche du modèle à une loi. Au-delà du seuil, les sinistres sont modélisés par à une loi de Pareto Généralisée. Une nouvelle densité qui est le mélange des deux précédentes est alors définie. La densité de ce modèle à deux lois s'exprime comme :

$$f_3(x) = \underbrace{\frac{f(x)}{\bar{F}(s)}}_{f_s(x)} \times \mathbb{1}_{[s,u]} + \left(1 - \frac{F(u) - F(s)}{\bar{F}(s)}\right) f^{\mathcal{GPD}}(x)$$

En notant $\alpha = \frac{F(u) - F(s)}{\bar{F}(s)}$, il vient que :

$$f_3(x) = f_s(x) \times \mathbb{1}_{[s,u]} + (1 - \alpha) f^{\mathcal{GPD}}(x)$$

L'objectif est de calculer une estimation du coût moyen à charge du réassureur pour un traité en excédent de sinistre C XS D. Pour ce modèle, on note π le coût moyen à charge du réassureur estimé selon le modèle à deux lois, π_1 le coût moyen à charge du réassureur estimé selon le modèle à une loi tronquée à gauche et π_2 le coût moyen à charge du réassureur estimé pour une variable aléatoire X suivant une loi $\mathcal{GPD}(u, \gamma, \sigma)$.

Il est possible d'exprimer π en fonction de π_1 et π_2 selon la valeur de la franchise D et de la portée C du traité.

Cas où $D + C \leq u$

La modélisation au delà du seuil u n'est pas utilisée. Le coût moyen à charge du réassureur dans le modèle à deux lois est identique au coût moyen à charge du réassureur dans le modèle à une loi :

$$\boxed{\forall D, C \in \mathbb{R}, \text{ tel que } D + C \leq u, \quad \pi(D, C) = \pi_1(D, C)}$$

Cas où $D \geq u$

La modélisation en dessous du seuil u n'est pas utilisée. Le coût moyen à charge du réassureur est le coût moyen estimé à l'aide d'une loi de Pareto Généralisée :

$$\boxed{\forall D, C \in \mathbb{R}, \text{ tel que } D \geq u, \quad \pi(D, C) = (1 - \alpha) \times \pi_2(D, C)}$$

Il reste alors à définir le coût moyen à charge du réassureur quand la sinistralité est modélisée à l'aide d'une loi de Pareto Généralisée $\mathcal{GPD}(u, \gamma, \sigma)$.

On rappelle que si $X \sim \mathcal{GPD}(u, \gamma, \sigma)$ alors :

$$F^{\mathcal{GPD}}(x) = \begin{cases} 1 + \left[1 + \frac{\gamma}{\sigma}(x - u)\right]^{-\frac{1}{\gamma}} & \text{si } \gamma \neq 0 \\ 1 - \exp\left(-\frac{x-u}{\sigma}\right) & \text{si } \gamma = 0 \end{cases}$$

Cas où la portée est infinie Dans ce cas, on a la formule suivante :

$$\pi_2(D, \infty) = \int_D^\infty \bar{F}_s^{\mathcal{GPD}}(x) dx$$

$\forall \gamma > 0$ et sous réserve que $\gamma < 1$

$$\begin{aligned}
\pi_2(D, \infty) &= \int_D^\infty \left(1 + \frac{\gamma}{\sigma}(x - u)\right)^{-\frac{1}{\gamma}} dx \\
&= \frac{\sigma}{\gamma} \left[\frac{\left(1 + \frac{\gamma}{\sigma}(x - u)\right)^{\frac{\gamma-1}{\gamma}}}{\frac{\gamma-1}{\gamma}} \right]_D^\infty \\
&= \frac{\sigma}{1-\gamma} \left(1 + \frac{\gamma(D-u)}{\sigma}\right)^{\frac{\gamma-1}{\gamma}}
\end{aligned} \tag{3.3}$$

Cas où la portée est finie Dans ce cas, on a la formule suivante :

$$\pi_2(D, \infty) = \int_D^{D+C} \bar{F}_s^{\mathcal{GPD}}(x) dx$$

$\forall \gamma > 0$

$$\begin{aligned}
\pi_2(D, \infty) &= \int_D^{D+C} \left(1 + \frac{\gamma}{\sigma}(x - u)\right)^{-\frac{1}{\gamma}} dx \\
&= \frac{\sigma}{1-\gamma} \left[\left(1 + \frac{\gamma(D-u)}{\sigma}\right)^{\frac{\gamma-1}{\gamma}} - \left(1 + \frac{\gamma(D+C-u)}{\sigma}\right)^{\frac{\gamma-1}{\gamma}} \right]
\end{aligned}$$

Ces résultats se résument ainsi :

$\forall D, C \in \mathbb{R}$, tel que $D \geq u$, $\pi(D, C) = (1 - \alpha) \times \pi_2(D, C)$ où :

- Si $C = \infty$, $\pi_2(D, \infty) = \frac{\sigma}{1-\gamma} \left(1 + \frac{\gamma(D-u)}{\sigma}\right)^{\frac{\gamma-1}{\gamma}}$.
- Si $C \in \mathbb{R}$, $\pi_2(D, C) = \frac{\sigma}{1-\gamma} \left[\left(1 + \frac{\gamma(D-u)}{\sigma}\right)^{\frac{\gamma-1}{\gamma}} - \left(1 + \frac{\gamma(D+C-u)}{\sigma}\right)^{\frac{\gamma-1}{\gamma}} \right]$

Remarque Dans le cas où une variable aléatoire suit une loi $\mathcal{GPD}(u, \gamma, \sigma)$, sa moyenne vaut $u + \frac{\sigma}{1-\gamma}$. Il est intéressant de remarquer que lorsque $C = \infty$ et si l'on considère que $D = u$ alors on retrouve la moyenne de la loi GPD à laquelle on soustrait le montant de la franchise D (ici égale à u). Autrement dit, du point de vue actuarielle, si l'on modélise la sinistralité avec une $\mathcal{GPD}(u, \gamma, \sigma)$, et que l'on veut tarifier un traité en excédent de sinistre ∞ XS u , le coût moyen dans la tranche est la moyenne de la variable aléatoire à laquelle on soustrait la franchise. Ce résultat est logique et nous permet de vérifier que la formule est cohérente.

Cas où $D \leq u$ et $D + C \geq u$

Ce dernier cas est le plus technique car il fait intervenir les deux lois du modèle. Le coût moyen à charge du réassureur s'exprime comme la combinaison suivante :

$$\forall D, C \in \mathbb{R}, \text{ tel que } D \leq u \text{ et } D + C \geq u, \quad \pi(D, C) = \underbrace{\pi_1(D, u - D)}_{u-D \text{ XS } D \text{ selon modèle 1}} + (1 - \alpha) \times \pi_2^*(u, D + C - u)$$

Il faut alors justifier la valeur de $\pi_2^*(u, D + C - u)$ dans la configuration où $D \leq u$ et $D + C \geq u$. Pour cela, supposons que la sinistralité est modélisée dans son ensemble par une variable aléatoire X suivant une loi de Pareto Généralisée de paramètre (u, γ, σ) . Dans ce cas, il vient que :

$$\begin{aligned}\pi(D, C) &= \int_D^{D+C} \bar{F}^{\mathcal{GPD}}(x) dx \\ &= \underbrace{\int_u^{D+C} \bar{F}^{\mathcal{GPD}}(x) dx}_{\pi_2^*(u, D+C-u)} + \underbrace{\int_D^u \bar{F}^{\mathcal{GPD}}(x) dx}_{u-D}\end{aligned}$$

Or, revenons à la modélisation à deux lois. La partie sous le seuil u est modélisée par une loi que l'on peut qualifier de classique. Autrement dit, la valeur de $\int_D^u \bar{F}^{\mathcal{GPD}}(x) dx$ est déjà pris en compte dans le modèle par $\pi_1(D, u - D)$. La valeur de $\pi_2^*(u, D + C - u)$ se calcule facilement à l'aide des formules définies en page [68](#).

Ces résultats se résument, sous réserve que $0 < \gamma < 1$, par :

$\forall D, C \in \mathbb{R}$, tel que $D \leq u$ et $D+C \geq u$, $\pi(D, C) = \pi_1(D, u-D) + (1-\alpha) \times \pi_2^*(u, D+C-u)$
où :

- Si $C = \infty$, $\pi_2^*(u, \infty) = \frac{\sigma}{1-\gamma}$.
- Si $C \in \mathbb{R}$, $\pi_2^*(u, D+C-u) = \frac{\sigma}{1-\gamma} \left[1 - \left(1 + \frac{\gamma(D+C-u)}{\sigma} \right)^{\frac{\gamma-1}{\gamma}} \right]$

3.7.3 Modèle d'Erlang

Dans ce modèle, un seuil de sinistres extrêmes, noté u est déterminé, puis en dessous de ce seuil, la sinistralité est modélisée via une distribution de mélange d'Erlang tronquée à gauche et à droite. Au-delà du seuil, les sinistres sont modélisés par à une loi de Pareto Généralisée. Une nouvelle densité qui est le mélange des deux précédentes est alors définie. La fonction de répartition de ce modèle à deux lois s'exprime comme :

$$F_3(x) = \begin{cases} \alpha \left(\frac{F_{ME}(x) - F_{ME}(s)}{F_{ME}(u) - F_{ME}(s)} \right) & \text{si } x \in [s, u] \\ \alpha + (1-\alpha) F^{\mathcal{GPD}}(x) & \text{si } x \geq u \end{cases}$$

L'objectif est de calculer une estimation du coût moyen à charge du réassureur pour un traité en excédent de sinistre C XS D . Pour ce modèle, on note π le coût moyen à charge du réassureur estimé selon le modèle à deux lois, π_1 le coût moyen à charge du réassureur estimé selon le modèle de mélange d'Erlang tronquée à droite et à gauche, π_1^* le coût moyen à charge selon un modèle de mélange d'Erlang non tronquée et π_2 le coût moyen à charge du réassureur estimé pour une variable aléatoire X suivant une loi $\mathcal{GPD}(u, \gamma, \sigma)$. La propriété suivante est utilisée :

$$\pi(D, C) = \pi(D, \infty) - \pi(D + C, \infty)$$

De sorte qu'il suffit de déterminer la valeur pour toute franchise D de $\int_D^\infty (1 - F_3(x)) dx$ pour être en mesure d'estimer le coût à charge de toutes les tranches de réassurance selon ce modèle. Il est possible d'exprimer π en fonction de π_1^* et π_2 selon la valeur de la franchise D du traité.

Cas où $D \geq u$

Dans ce cas, comme dans le précédent modèle à deux lois, seule la modélisation par la distribution de Pareto Généralisée est utilisée.

$$\pi(D, \infty) = (1-\alpha) \pi_2(D, \infty)$$

La valeur de $\pi_2(D, \infty)$ est définie par l'équation [3.3](#)

Cas où $D < u$

Dans ce cas, les deux modélisations sont importantes. Le coût s'exprime par :

$$\begin{aligned}
 \pi(D, \infty) &= \int_D^u (1 - \alpha F_1(x)) dx + \int_u^\infty (1 - (\alpha + (1 - \alpha) F^{\mathcal{GPD}}(x))) dx \\
 &= (u - D) - \alpha \int_D^u F_1(x) dx + (1 - \alpha) \int_u^\infty (1 - F^{\mathcal{GPD}}(x)) dx \\
 &= (u - D) - (u - D)\alpha + \alpha \int_D^u (1 - F_1(x)) dx + (1 - \alpha) \int_u^\infty (1 - F^{\mathcal{GPD}}(x)) dx \\
 &= (1 - \alpha)(u - D) + \alpha \pi_1(D, u - D) + (1 - \alpha) \pi_2(u, \infty)
 \end{aligned}$$

La valeur de $\pi_1(D, u - D)$ est le coût à charge de la réassurance pour une tranche $u - D$ XS D lorsque les données sont modélisées par une loi de mélange d'Erlang tronquée à gauche et à droite. Ce coût s'exprime par :

$$\pi_1(D, u - D) = \int_D^u \left(1 - \frac{F_{ME}(x) - F_{ME}(s)}{F_{ME}(u) - F_{ME}(s)} \right) dx$$

Cette expression peut s'écrire en fonction de π_1^* qui représente le coût à charge de la réassurance lorsque les données sont modélisées selon une loi de mélange d'Erlang non tronquée. Ainsi :

$$\pi_1(D, u - D) = \frac{(F_{ME}(u) - 1)(u - D) + (\pi_1^*(D, \infty) - \pi_1^*(u, \infty))}{F_{ME}(u) - F_{ME}(s)}$$

Avec pour expression de la fonction de répartition de la loi mélange d'Erlang :

$$F_{ME}(x) = \sum_{j=1}^M \beta_j \left(1 - \sum_{n=0}^{r_j-1} e^{-\lambda x} \frac{(\lambda x)^n}{n!} \right)$$

Et [Albrecher et al. \(2017\)](#) fournit la valeur pour $\pi_1^*(D, \infty)$:

$$\pi_1^*(D, \infty) = \frac{1}{\lambda} e^{\lambda D} \sum_{n=0}^{M-1} \sum_{k=n}^{M-1} \left(\sum_{j=k+1}^M \beta_j \right) \frac{(\lambda D)^n}{n!}$$

Les différents outils utilisés et les résultats théoriques obtenus pour la modélisation des montants de sinistres étant présentés, le chapitre suivant s'attache à l'application des modèles sur données simulées et réelles.

Chapitre 4

Application

Ce chapitre illustre la mise en place des modélisations à une et à deux lois. Dans un premier temps, le modèle à une loi et le modèle à deux lois usuelles sont implémentés sur des données simulées afin de vérifier que leurs paramètres sont identifiables. Dans un second temps, une modélisation sur des données réelles issues d'une cédante est présentée. L'application sur données réelles permet de mettre en pratique la seconde modélisation à deux lois via le modèle d'Erlang. Cette modélisation sur données réelles aboutit à une analyse de l'impact de la modélisation à deux lois sur le niveau des quantiles et du coût à charge de la réassurance. Une étude de sensibilité des quantiles et du coût à charge en fonction du seuil des extrêmes sélectionné dans le modèle à deux lois est également présentée.

4.1 Sur données simulées

Les données simulées permettent de vérifier que les paramètres du modèle à deux lois usuelles sont identifiables grâce aux différents outils mathématiques présentés. Dans ce modèle à deux distributions, l'actuaire doit être en mesure d'estimer le seuil des extrêmes u , les paramètres de la loi usuelle modélisant la sinistralité sous le seuil, et les paramètres de forme γ et d'échelle σ de la loi de Pareto Généralisée modélisant la sinistralité au-delà du seuil.

La simulation de données qui suivent une distribution particulière repose sur le lemme suivant.

Lemme 1 (Lemme de transformation des quantiles). *Soit $U : (\Omega, \mathcal{F}, \mathbb{P}) \mapsto ([0, 1], \mathcal{B}[0, 1])$ une variable aléatoire de loi uniforme sur $[0, 1]$. Soit $F : \mathbb{R} \mapsto [0, 1]$ une fonction de répartition. Alors la variable aléatoire $X = F^{-1} \circ U : (\Omega, \mathcal{F}, \mathbb{P}) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est une variable aléatoire de fonction de répartition F .*

Démonstration. Soit $t \in \mathbb{R}$. On a :

$$\mathbb{P}(X \leq t) = \mathbb{P}(F^{-1}(U) \leq t)$$

Or F est une fonction croissante et continue à droite, son inverse est une fonction croissante et continue à gauche donc :

$$\mathbb{P}(X \leq t) = \mathbb{P}(U \leq F(t))$$

De plus, $\forall t \quad F(t) \in [0, 1]$ et comme la variable aléatoire $U \sim U[0, 1]$:

$$\mathbb{P}(X \leq t) = F(t)$$

□

De cette manière, un échantillon de 500 données distribuées selon le modèle à deux lois suivant est simulé :

- Un seuil des extrêmes $u = 20$.
- Une loi de Pareto de paramètres $\alpha = 2$ et $A = 10$ pour modéliser la sinistralité sous le seuil.
- Une loi de Pareto Généralisée de paramètres $\gamma = 0,5$ et $\sigma = 10$ pour modéliser la sinistralité au-delà du seuil.

La suite permet de vérifier que les paramètres du modèles peuvent être estimés.

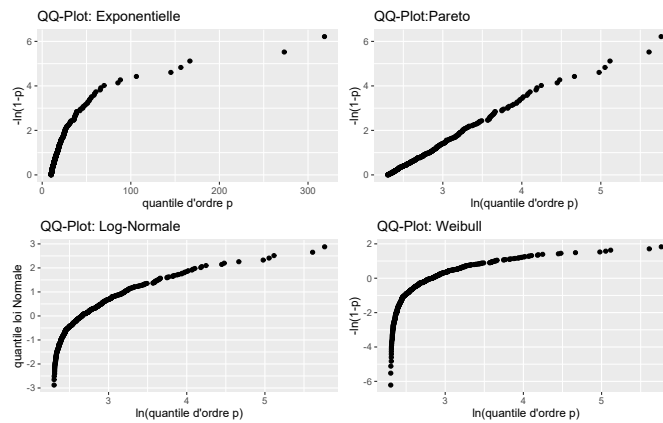
4.1.1 Identification des paramètres du modèle à une loi

Dans un premier temps, il faut réussir à identifier la loi de Pareto et les paramètres associés. Les outils statistiques présentés dans le chapitre précédent sont utilisés.

Graphique Quantile-Quantile

Les QQ-plot permettent une première analyse graphique de l'ajustement à une loi donnée.

FIGURE 4.1: QQ-plot sur données simulées



Le QQ-plot associé à la distribution de Pareto est linéaire. Les trois autres QQ-plot sont concaves car la loi simulée est à queue plus épaisse. La conclusion graphique de ces QQ-plot est que les données semblent bien issues d'une distribution de Pareto. Pour consolider cette hypothèse, les résultats des tests statistiques sont utilisés.

Test d'adéquation

Le tableau ci-dessous récapitule la conclusion du test de Kolmogorov-Smirnov et d'Anderson Darling sur l'échantillon de données simulées.

TABLE 4.1: Résultats des tests statistiques sur les données simulées

Distribution	Kolmogorov-Smirnov	Anderson-Darling
Pareto	Accepté	Accepté
Weibull	Rejeté	Rejeté
Exponentielle	Rejeté	Rejeté
Log-Normale	Rejeté	Rejeté
Gamma	Rejeté	Rejeté
Normale	Rejeté	Rejeté

Les tests ne sont pas significatif au seuil de 5% uniquement pour la distribution de Pareto. Les résultats vont dans le sens de l'analyse des QQ-Plot. Enfin, les critères de l'erreur quadratique moyenne, de l'AIC et du BIC sont calculés.

Critère du MSE, AIC et BIC

Le tableau suivant fournit la valeur de l'erreur quadratique moyenne, de l'AIC et du BIC calculés sur les données simulées.

TABLE 4.2: MSE, AIC et BIC sur les données simulées

Distribution	MSE	AIC	BIC
Pareto	0,0414	3 116	3 124
Weibull	9,850	3 925	3 933
Exponentielle	14,559	3 998	4 002
Log-Normale	3,798	3 533	3 541
Gamma	7,339	3 780	3 789
Normale	37,210	4 484	4 493

C'est pour la loi de Pareto que les critères statistiques sont minimisés. Finalement, la loi utilisée pour modéliser la sinistralité sous le seuil est identifiable. Il reste à être en mesure d'estimer les paramètres α et A de la loi.

Estimation des paramètres

Les paramètres de la loi de Pareto sont estimés par la méthode du maximum de vraisemblance. Cette méthode permet de construire un intervalle de confiance à 95% pour le paramètre de forme α en utilisant les propriétés de normalité asymptotique de l'estimateur du maximum de vraisemblance.

TABLE 4.3: Estimation des paramètres de la loi de Pareto sur les données simulées

Paramètre	Valeur réelle	Valeur estimée	Intervalle de confiance à 95%
A	10	10,00196	\times
α	2	1,996864	[1, 83594 ; 2, 18871]

Finalement, la loi qui modélise la sinistralité sous le seuil est identifiable. De plus, l'estimation des paramètres est proche des valeurs réelles. La prochaine étape est de réussir à identifier le seuil des extrêmes, fixé à $u = 20$.

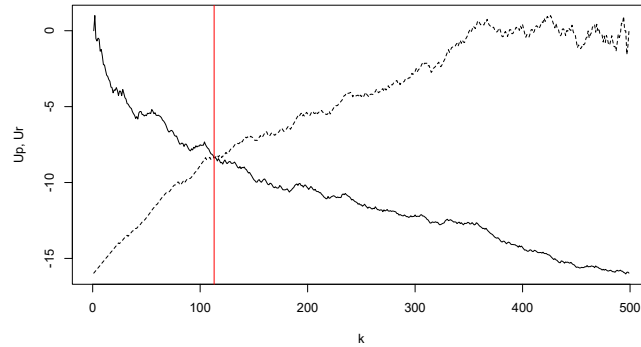
4.1.2 Identification du seuil des extrêmes

Une fois que la loi qui permet de modéliser la sinistralité sous le seuil est estimée, il faut être capable d'identifier le seuil des extrêmes. Pour cela, le graphique du Gerstengarbe est utilisé en premier pour fournir une valeur. Ce graphique n'étant pas construit à partir de la théorie des valeurs extrêmes, la valeur obtenue est ensuite comparée avec le seuil estimé visuellement grâce au *Hill plot*, au *Mean Excess plot*, et au *Gamma plot*.

Gerstengarbe plot

La valeur du seuil d'entrée dans la zone des données extrêmes correspond à l'intersection des deux courbes du *Gerstengarbe plot*.

FIGURE 4.2: *Gerstengarbe plot* sur les données simulées

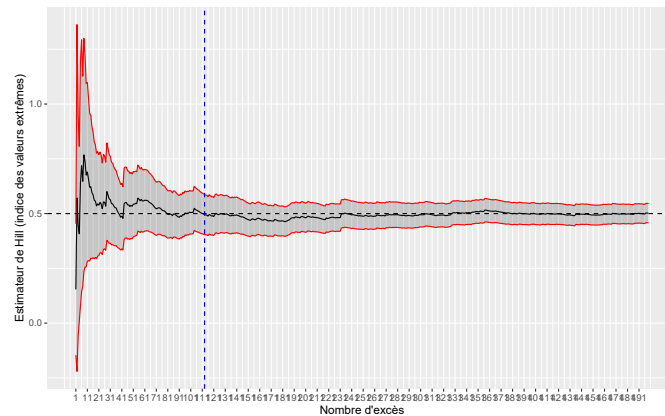


Le graphique de Gerstengarbe indique un nombre d'excès à considérer égal à 113 ce qui correspond à un seuil de 20,83612. Cette valeur est reportée sur les graphiques suivants pour pouvoir le valider ou non.

Hill plot

La valeur du seuil d'entrée dans la zone des données extrêmes correspond au plus petit nombre d'excès à partir duquel le graphique de Hill se stabilise.

FIGURE 4.3: *Hill plot* sur les données simulées

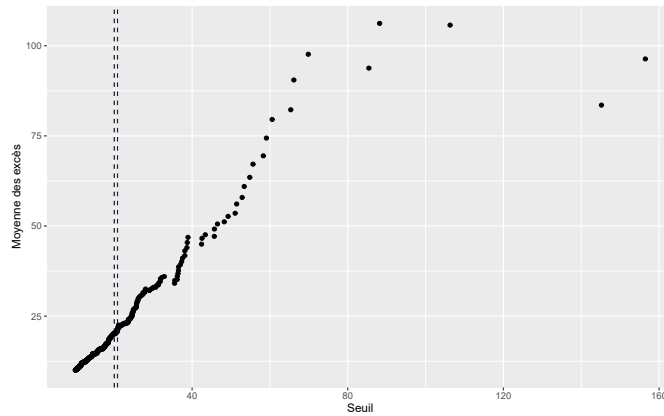


A partir d'un nombre d'excès de 113, l'estimateur de Hill se stabilise. La valeur du seuil fournie par le *Gerstengarbe plot* est cohérente avec le graphique de Hill.

Mean Excess plot

La valeur du seuil d'entrée dans la zone des données extrêmes correspond à la valeur à partir de laquelle le *Mean Excess plot* est approximativement linéaire.

FIGURE 4.4: *Mean Excess plot* sur les données simulées

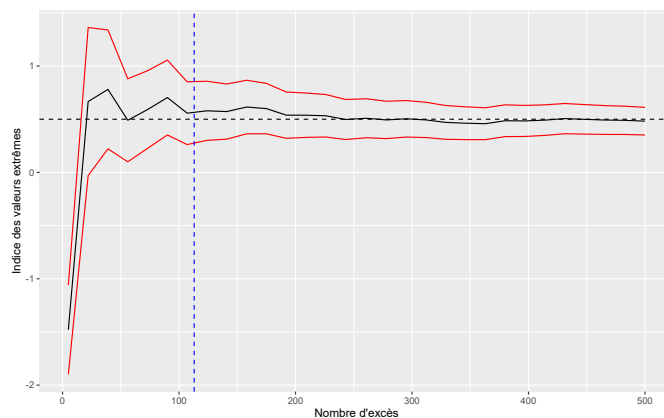


La droite verticale **bleue** correspond au seuil fourni par le *Gerstengarbe plot* et la droite verticale **noire** indique la valeur réelle du seuil. Sur le *Mean Excess plot*, une légère cassure est observée au-delà de laquelle le graphique est approximativement linéaire.

Gamma plot

Un dernier outil graphique est utilisé pour confirmer le seuil fourni par le graphique de Gerstengarbe, c'est le *Gamma plot*. Le seuil d'entrée dans la zone des données extrêmes correspond à la plus petite valeur à partir de laquelle le *Gamma plot* se stabilise.

FIGURE 4.5: *Gamma plot* sur les données simulées



A partir des 113 excès, le *Gamma plot* se stabilise, la valeur du seuil fourni par le *Gerstengarbe plot* est cohérente avec les différents outils graphiques issus de la théorie des valeurs extrêmes. Le seuil estimé est de 20, 83612 par rapport à la valeur réelle de 20.

Il reste à être capable d'estimer les paramètres γ et σ de la loi de Pareto Généralisée utilisée pour modéliser la sinistralité supérieure au seuil des extrêmes.

Estimation des paramètres de la loi de Pareto Généralisée

Les paramètres de la loi de Pareto Généralisée sont estimés par la méthode du maximum de vraisemblance. Cette méthode permet de construire un intervalle de confiance à 95% pour les paramètres de forme γ et d'échelle σ en utilisant les propriétés de normalité asymptotique de l'estimateur du maximum de vraisemblance.

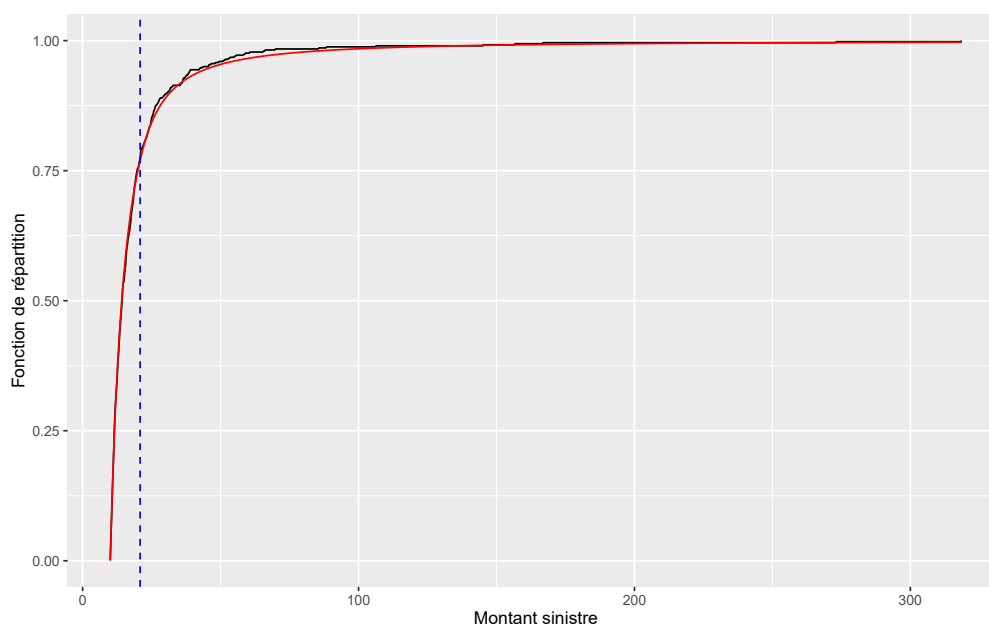
TABLE 4.4: Estimation des paramètres de la loi de Pareto Généralisée sur les données simulées

Paramètre	Valeur réelle	Valeur estimée	Intervalle de confiance à 95%
σ	10	9,38	[6, 44 ; 12, 32]
γ	0,5	0,59	[0, 30 ; 0, 89]

Finalement, le seuil des extrêmes est identifiable. De plus, l'estimation des paramètres est proche des valeurs réelles. Sur les données simulées, il est possible d'estimer convenablement tous les paramètres du modèle à deux lois grâce aux outils issus de la théorie des valeurs extrêmes.

Pour terminer la modélisation à deux lois sur données simulées, la fonction de répartition empirique et la fonction de répartition à deux lois usuelles estimée sont représentées sur le même graphique.

FIGURE 4.6: Modélisation à deux lois sur des données simulées



L'adéquation graphique est excellente. L'étape de validation du modèle sur des données simulées est concluante. La prochaine étape est de modéliser des sinistres réels issus d'une cédante.

4.2 Sur données réelles

La section précédente a permis de valider la construction du modèle à deux lois usuelles sur des données simulées. En effet, la distribution engendrant les données était connue. Nous avons simplement vérifié que la distribution et ses paramètres étaient identifiables. Cette section s'attache à présenter le processus de modélisation de la sévérité sur des données réelles issues d'une cédante européenne. Dans le cadre d'une cotation sur des données réelles, l'actuaire n'a pas connaissance de la loi engendrant les données. Il cherche à extrapoler une loi lui permettant de modéliser le phénomène aléatoire représentant la sinistralité et ainsi estimer un coût à charge pour une tranche de réassurance. L'ensemble de cette section s'attache à présenter la méthodologie utilisée par l'actuaire voulant modéliser les sinistres par un modèle à deux lois. Cette modélisation sur données réelles permet d'appliquer le second modèle à deux lois basé sur la distribution mélange d'Erlang.

Le traité utilisé pour illustrer la modélisation de la sévérité réelle est un traité Dommages aux Biens. Ce traité a une structure par Risque et par Évènement couvrant tous les périls des polices Incendie, Bris de Machine et Tous Risques Chantier. L'analyse présentée se restreint au portefeuille Incendie. La

couverture de réassurance à évaluer est un traité en excédent de sinistre.

La première étape dans la modélisation consiste à revaloriser les sinistres passés de l'inflation pour connaître leurs montants s'ils avaient lieu durant l'année de cotation.

4.2.1 Revalorisation des sinistres : la mise en *as if*

Pour modéliser les montants des sinistres d'une cédante, l'actuaire dispose d'un historique de montants de sinistres pouvant s'étaler sur plusieurs dizaines d'années. La valeur de l'argent évolue au cours du temps, l'inflation doit être prise en compte pour évaluer le risque. Il convient de revaloriser les montants des sinistres à la date de modélisation pour avoir un échantillon comparable sur une même base. Les sinistres passés ont un coût, l'objectif est donc de répondre à la question suivante : Quels seraient les montants de ces sinistres s'ils avaient lieu aujourd'hui ?

Ce retraitement des données historiques est la mise en *as-if*.

Les facteurs de revalorisation

Les montants des sinistres sont sujets à trois facteurs principaux :

- L'inflation économique : Elle est souvent décrite par un indice de prix à la consommation. Dans beaucoup de pays, cet indice est le mieux documenté pour suivre l'évolution monétaire dans le temps. D'autres indices peuvent également être utilisés comme un indice d'évolution des salaires, un indice du coût de la construction ou encore un indice d'évolution des coûts d'hospitalisation. Le choix de l'indice dépend du pays et de la branche à laquelle appartient les sinistres à modéliser.
- La politique de souscription : Les changements dans la politique de souscription doivent être considérés lors de la mise en *as-if* des montants des sinistres. Les garanties, les limites et couvertures sont des éléments typiques qui ont un impact direct sur le portefeuille et qu'il convient de prendre en compte. Cependant, ces facteurs sont difficiles à quantifier et à transformer en indice.
- L'environnement juridique et réglementaire : Comme pour la politique de souscription, un changement juridique ou une évolution de la jurisprudence peut avoir un impact important sur les montants des sinistres. Une évolution réglementaire peut être la modification du taux d'actualisation dans le cadre de règlement de sinistres corporels graves en automobile.

La mise en *as-if*

La revalorisation des sinistres est réalisée en utilisant un ou plusieurs indices qui permettent de quantifier les facteurs mentionnés précédemment. Il existe différentes méthodes pour revaloriser les montants des sinistres passés : des méthodes prospectives ou des méthodes rétrospectives. C'est cette dernière méthode qui est retenue dans la suite. La méthode rétrospective suppose que l'inflation passée reflète l'inflation future.

Méthodologie L'objectif de la revalorisation des montants des sinistres passés est de déterminer le coût de ces mêmes sinistres s'ils étaient survenus l'année étudiée. Pour cela, les montants des sinistres historiques sont redressés de l'inflation en fonction de leur année de survenance. La méthode repose sur le calcul de facteurs de redressement (OLF) basé sur un indice d'inflation.

Mathématiquement, considérons k sinistres survenus durant les années $y_1 < \dots < y_k$ avec y_k l'année de couverture à évaluer. Notons i_1, \dots, i_k les taux d'inflations considérés, et I_1, \dots, I_k les indices d'inflation associés. Notons également O_1, \dots, O_k les facteurs de redressements associés aux années y_1, \dots, y_k et enfin S_1, \dots, S_k les montants de sinistres historiques et $\tilde{S}_1, \dots, \tilde{S}_k$ les sinistres revalorisés.

TABLE 4.5: Méthodologie de la mise en *as-if*

Année de survenance	Montant historique	Taux d'inflation	Indice d'inflation	OLF	Montant <i>As-If</i>
y_1	S_1		I_1	$O_1 = \frac{I_k}{I_1}$	$\tilde{S}_1 = S_1 \times O_1$
y_2	S_2	i_2	$I_2 = (1 + i_2) \times I_1$	$O_2 = \frac{I_k}{I_2}$	$\tilde{S}_2 = S_2 \times O_2$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
y_k	S_k	i_k	$I_k = (1 + i_k) \times I_{k-1}$	$O_k = 1$	$\tilde{S}_k = S_k$

Application Pour terminer la présentation de la mise en *as-if*, la méthode est appliquée sur les données réelles de la cédante européenne. La valeur revalorisée des sinistres survenus entre 2012 et 2020 est évaluée. Le taux d'inflation utilisé est l'indice d'évolution du prix à la consommation. Il est récupéré à partir du site internet du Fond Monétaire International.

TABLE 4.6: Evolution de l'indice du prix à la consommation

Année	Evolution	Indice	OLF
2012		104,41	1,1083
2013	1,80%	106,30	1,0886
2014	2,11%	108,54	1,0662
2015	-0,04%	108,50	1,0665
2016	0,74%	109,30	1,0587
2017	0,76%	110,13	1,0508
2018	1,60%	111,89	1,0342
2019	2,05%	114,18	1,0135
2020	1,35%	115,72	1

Les sinistres de la cédante sont revalorisés grâce à cet indice. Afin d'illustrer ce travail, pour chacune des années entre 2012 et 2020, la valeur historique et la valeur *as-if* d'un sinistre sont indiquées.

TABLE 4.7: Mise en *as-if* sur données réelles

Année de survenance	Montant historique	Montant <i>As-If</i>
2012	1 265 380€	1 402 396€
2013	Aucun sinistre	
2014	1 029 441€	1 097 539€
2015	Aucun sinistre	
2016	2 132 797€	2 258 072€
2017	1 000 000€	1 050 758€
2018	3 966 599€	4 102 376€
2019	32 100 000€	32 532 948€

Pour cette cédante européenne, le sinistre le plus ancien est augmenté de près de 10% pour témoigner de l'inflation des dernières années. Les données étant désormais comparables sur une même base, il est possible de commencer la modélisation par une première analyse descriptive.

4.2.2 Analyse descriptive des données

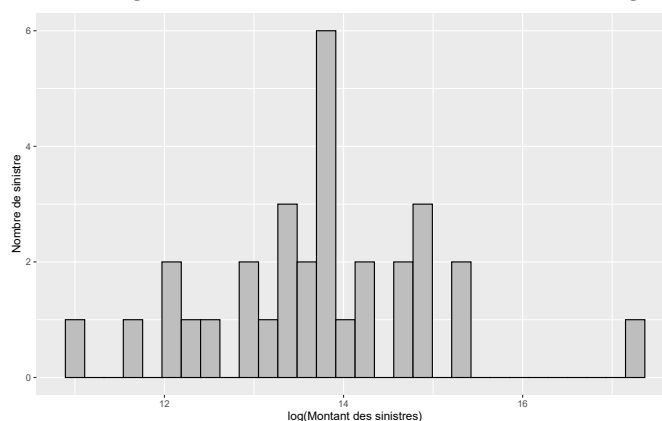
Avant de débiter le processus de modélisation des sinistres, il est indispensable d'analyser quelques statistiques descriptives des données Incendie disponibles pour la cotation. Le seuil d'avis de sinistres de la cédante est de 50 000 €. Autrement dit, la cédante fournit uniquement les sinistres supérieurs à 50 000 €.

TABLE 4.8: Statistiques descriptives de la sinistralité Incendie

Statistique	Valeur
Nombre de sinistres	31
Minimum	62 451€
Premier quartile	483 537€
Médiane	925 189€
Moyenne	2 249 891€
Troisième quartile	1 905 665€
Maximum	32 532 948€
Écart-type	5 729 455€

Les données disponibles sont typiques de la réassurance. Le nombre d'observations est faible. Les montants des sinistres sont très étalés avec un minimum de 62 451€ et un maximum à 32 532 948€. L'écart-type est par conséquent important. La médiane est relativement basse contrairement à la moyenne qui est tirée vers le haut par le plus gros sinistre.

FIGURE 4.7: Histogramme des sinistres Incendie en échelle logarithmique



Les montants des sinistres sont répartis de manière homogène, il ne semble pas avoir de tendance particulière. Seul le plus gros sinistre à 32 532 948 € se détache des autres. C'est un incendie qui s'est déroulé en 2019 dans une usine industrielle.

De manière plus générale, l'hypothèse d'indépendance des sinistres n'est pas remise en cause. La suite s'attache à déterminer la loi qui modélise au mieux la sinistralité sous le seuil des extrêmes. Ce seuil sera sélectionné à la suite.

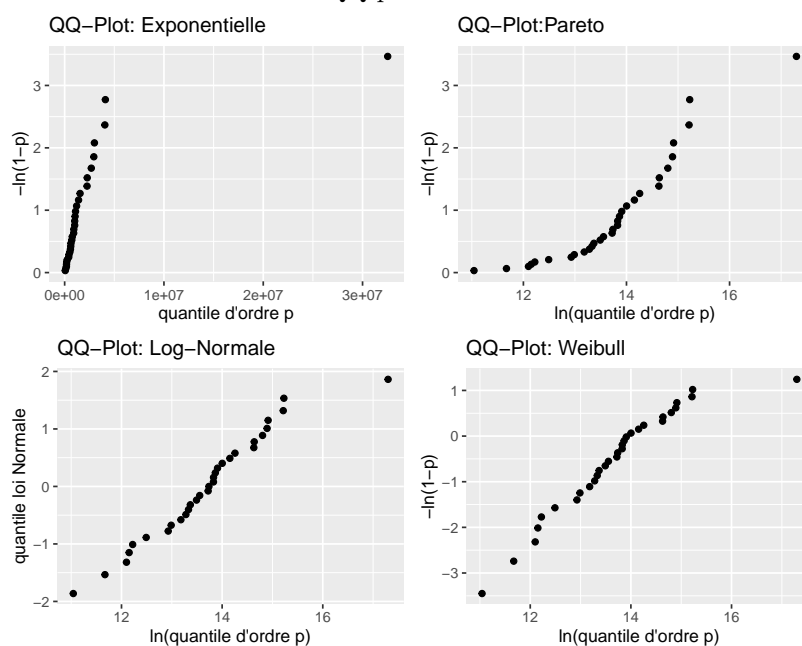
4.2.3 Sélection du modèle à une loi

La première étape de la modélisation consiste à déterminer la loi qui modélisera la sinistralité sous le seuil des extrêmes, qui sera défini dans la suite. Pour ce faire, différents outils statistiques sont utilisés. En premier lieu, une analyse graphique de l'adéquation est réalisée grâce au QQ-Plot.

Graphique Quantile-Quantile

En première approche, une distribution de probabilité peut être jugée pertinente pour modéliser un phénomène aléatoire si l'allure de son QQ-Plot est linéaire.

FIGURE 4.8: QQ-plot sur données réelles



Le QQ-plot de la loi Exponentielle à une allure concave ce qui suggère que les données sont issues d'une distribution à queue plus épaisse. A l'inverse le QQ-plot de la loi de Pareto est convexe et suggère que les données sont issues d'une distribution à queue plus fine. Les QQ-plot de la loi Log-Normale et de la loi de Weibull sont approximativement linéaires. L'analyse des QQ-plot permet de mettre en avant deux distributions :

- La distribution Log-Normale.
- La distribution de Weibull.

A ce stade, il est difficile de sélectionner la loi la plus adéquate aux données. Les résultats des tests statistiques sont utilisés, cependant, il convient d'estimer les paramètres de chacune des distributions pour évaluer la qualité de l'ajustement.

Estimation des paramètres

La méthode du maximum de vraisemblance est utilisée pour estimer les paramètres des différentes distributions tronquées à gauche au seuil de modélisation $s = 50\,000$. Pour certaines lois (Log-Normale, Gamma, Weibull, Normale), il est malaisé d'obtenir une formule fermée pour le couple de paramètres qui maximise la fonction de vraisemblance. La méthode numérique de Nelder-Mead est alors utilisée, elle nécessite des valeurs initiales pour la première itération. Ces valeurs initiales sont calculées grâce à la méthode des moments.

TABLE 4.9: Estimation des paramètres des distributions sur données réelles

Distribution	Paramètres	
Pareto	$\hat{\alpha} = 0,38$	$\hat{A} = 62\,451$
Weibull	$\hat{a} = 0,64$	$\hat{b} = 1\,433\,276$
Exponentielle	$\hat{\lambda} = 4,54e-7$	×
Log-Normale	$\hat{\mu} = 13,72$	$\hat{\sigma} = 1,19$
Gamma	$\hat{k} = 0,77$	$\hat{\theta} = 3\,042\,980$
Normale	$\hat{\mu} = 2\,249\,890$	$\hat{\sigma} = 5\,729\,455$

Les paramètres optimaux de chacune des lois étant estimés, l'actuaire peut évaluer la qualité de l'ajustement grâce aux tests statistiques d'abord.

Test d'adéquation

Pour conforter l'actuaire dans le choix de la distribution sous le seuil des extrêmes, les tests statistiques de Kolmogorov-Smirnov et d'Anderson Darling sont utilisés. Le tableau ci-dessous récapitule la conclusion de ces deux tests sur les données réelles.

TABLE 4.10: Résultats des tests statistiques sur les données réelles

Distribution	Kolmogorov-Smirnov	Anderson-Darling
Pareto	Refusé	Refusé
Weibull	Accepté	Accepté
Exponentielle	Rejeté	Rejeté
Log-Normale	Accepté	Accepté
Gamma	Rejeté	Rejeté
Normale	Rejeté	Rejeté

Les tests conduisent à rejeter la distributions de Pareto, Exponentielle, Gamma et Normale. Les tests ne permettent pas de conclure au seuil de 5% que les données ne sont pas issues d'une distribution Log-Normale ou d'une distribution de Weibull.

Les résultats des tests rejoignent la première analyse graphique des QQ-plot. Les lois de Weibull et Log-Normale semblent les mieux adaptées pour modéliser les sinistres de la cédante européenne. Afin de sélectionner la loi la plus appropriée, l'actuaire se base sur d'autres critères statistiques.

Critère du MSE, AIC et BIC

Ces critères sont des aides à la décision pour l'actuaire lorsque les QQ-plot et les tests statistiques ne permettent pas de sélectionner une loi. L'actuaire en charge de la modélisation peut s'appuyer sur ces trois critères en sélectionnant la loi qui les minimise.

TABLE 4.11: MSE, AIC et BIC sur les données réelles

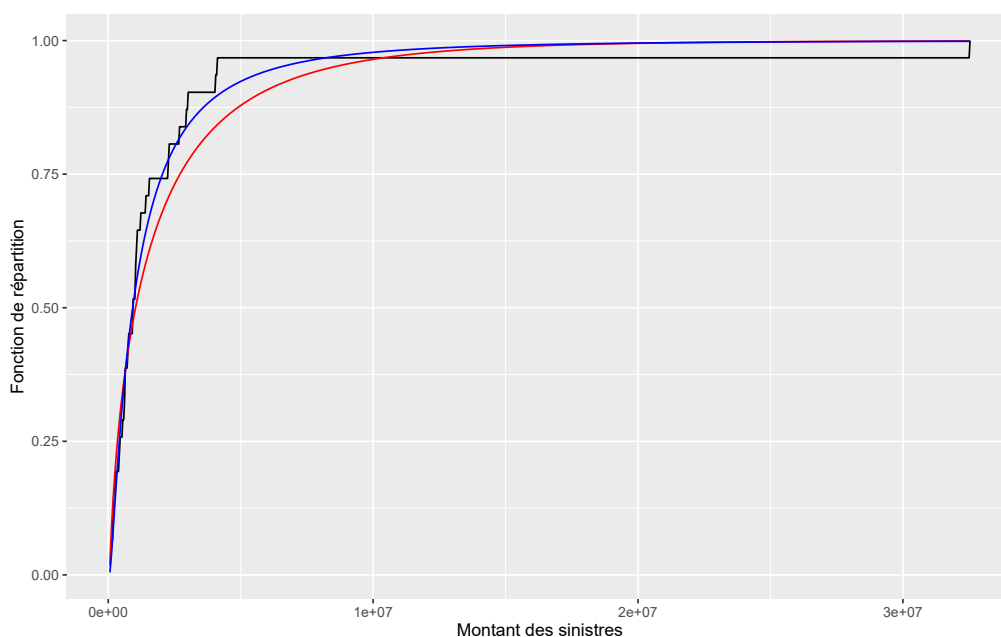
Distribution	MSE	AIC	BIC
Pareto	0,794	976	979
Weibull	0,181	957	960
Exponentielle	0,774	969	971
Log-Normale	0,059	953	955
Gamma	0,710	966	969
Normale	5,230	1 029	1 031

L'erreur quadratique, l'information d'Akaike et l'information Bayésienne sont minimisées par la distribution Log-Normale. La distribution Log-Normale semble être la plus adaptée pour modéliser la sinistralité.

Evaluation graphique

Pour terminer le choix de la loi, la fonction de répartition empirique des données et les deux fonctions de répartition théoriques **Log-Normale** (courbe **bleue**) et **Weibull** (courbe **rouge**) sont tracées sur le même graphique :

FIGURE 4.9: Ajustement graphique Log-Normale et Weibull sur données réelles



La distribution Log-Normale apparaît plus adéquate pour modéliser les sinistres. La loi de Weibull tend à être trop prudente et surestime la probabilité de survenance des sinistres de montants compris entre 1 000 000 € et 10 000 000 €.

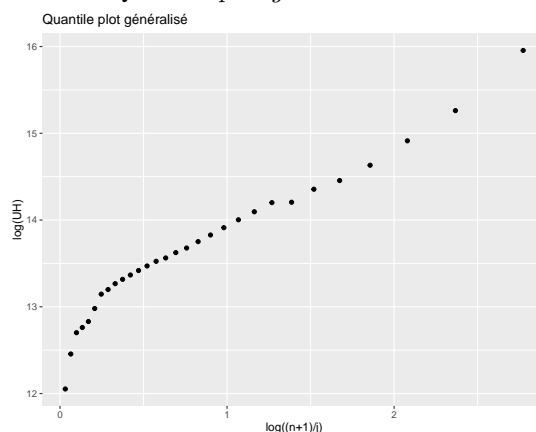
Finalement c'est une loi Log-Normale de paramètres $\mu = 13,72$ et $\sigma = 1,19$ qui sera utilisée pour modéliser la sinistralité sous le seuil des extrêmes. Cependant, à partir d'environ 20 000 000 €, la probabilité d'avoir un sinistre de montant supérieur est presque nulle ($\approx 0,002$). Or la cédante dispose d'un sinistre à près de 32 000 000 € dans son portefeuille de 31 sinistres. Si l'actuaire conserve cette modélisation, le coût à charge pour les tranches hautes sera quasiment nul et cela n'est pas approprié. Cette conséquence résulte du fait que la loi Log-Normale est une loi à queue fine. Pour palier à ce problème de tarification dans les tranches élevées, un modèle à deux lois est mis en place. La première étape est de déterminer à quel domaine d'attraction appartiennent les données. Ensuite, un seuil des extrêmes sera sélectionné pour construire une fonction de répartition à deux lois.

4.2.4 Détermination du domaine d'attraction

La détermination du domaine d'attraction des données est importante pour deux raisons. La première est que l'actuaire qui modélise la sinistralité souhaite savoir si la queue de distribution des données est épaisse ou fine. Dans le premier cas, l'indice des valeurs extrêmes est strictement positif, dans le second cas, l'indice est négatif. L'intérêt de la construction d'un modèle à deux lois se trouve dans le cas où les données appartiennent au domaine d'attraction de Fréchet. La seconde raison est que certains outils permettant de sélectionner le seuil des extrêmes sont valides uniquement pour certaines valeurs de l'indice des valeurs extrêmes. C'est notamment le cas du *Hill plot* valable pour un indice strictement positif.

La détermination du domaine d'attraction des données repose sur l'allure du *quantile plot généralisé* comme définit à la sous-section [2.4.4](#)

FIGURE 4.10: *Quantile plot généralisé* sur données réelles



L'allure du *quantile plot généralisé* est croissante. Les données peuvent être considérées comme appartenant au domaine d'attraction de Fréchet avec un indice des valeurs extrêmes strictement positif. La prochaine étape s'attache à sélectionner le seuil des extrêmes à l'aide des différents outils graphiques et numériques issus de la théorie des valeurs extrêmes.

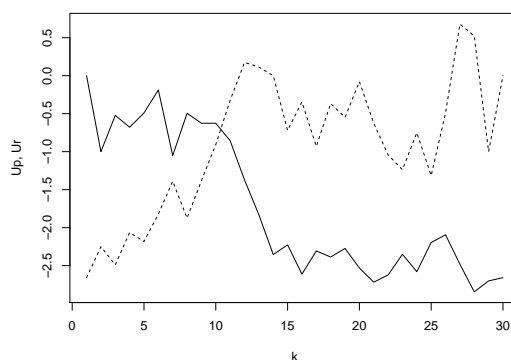
4.2.5 Sélection du seuil des extrêmes

Le théorème de Pickands, Balkema et De Hann suggère qu'il existe un certain seuil au-delà duquel les données se comportent comme une distribution de Pareto Généralisé. Une fois déterminé, ce seuil des extrêmes permettra de construire une fonction de répartition mêlant deux lois. Cependant, le choix du seuil est parfois malaisé. Une méthodologie est présentée pour guider l'actuaire dans son choix.

Gerstengarbe plot

Sur données réelles, il peut être difficile de trouver un seuil des extrêmes convenable, c'est-à-dire ni trop petit, ni trop grand. Pour faciliter l'analyse de l'actuaire dans la sélection de ce seuil, la première étape consiste à récupérer la valeur du seuil d'entrée dans la zone extrême fournie par le *Gerstengarbe plot*.

FIGURE 4.11: *Gerstengarbe plot* sur les données réelles

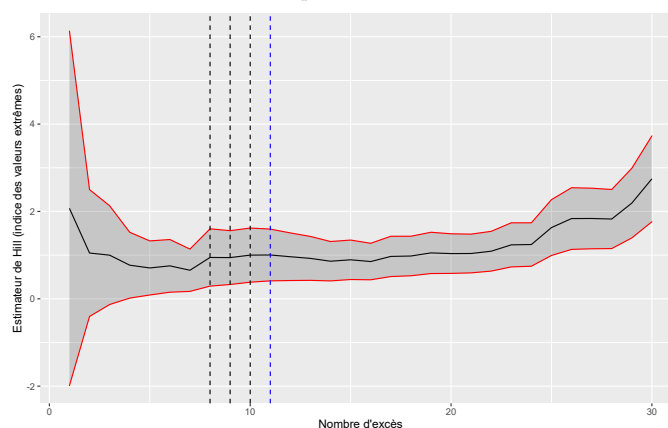


Le *Gerstengarbe plot* suggère un nombre d'excès à considérer pour l'entrée dans la zone extrêmes de 11, ce qui correspond à un seuil de 1 205 809 €. Cette première valeur permet d'orienter l'analyse de l'actuaire sur ce seuil lors de l'étude des graphiques suivants.

Hill plot

Le domaine d'attraction auquel appartiennent les données est supposé être celui de Fréchet. Ce domaine d'attraction est caractérisé par un indice des valeurs extrêmes γ strictement positif. Le graphique de Hill est valable.

FIGURE 4.12: Hill plot sur les données réelles



La droite verticale bleue représente le nombre d'excès suggéré par le *Gerstengarbe plot*. Les droites verticales noires sont des nombres d'excès à partir desquels l'actuaire peut considérer que le *Hill plot* se stabilise.

TABLE 4.12: Seuils des extrêmes suggérés par le *Hill plot*

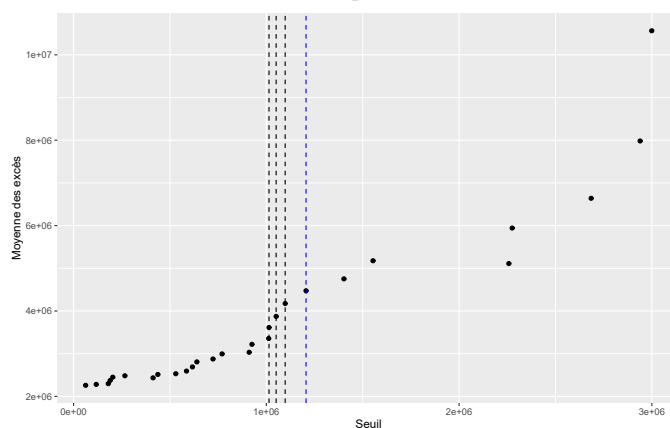
Nombre d'excès	Seuil
8	2 258 072 €
9	1 553 258 €
10	1 402 396 €
11	1 205 809 €

Le *Hill plot* ne permet pas à lui seul de sélectionner un seuil des extrêmes. La valeur fournie par le *Gerstengarbe plot* est cohérente avec la stabilité du graphique de Hill mais trois autres seuils pourraient convenir. D'autres graphiques sont analysés.

Mean Excess plot

Le graphique de la moyenne des excès est un outil graphique utilisé pour aider l'actuaire à sélectionner le seuil des extrêmes. Un seuil convenable est un seuil à partir duquel le *Mean Excess plot* se comporte approximativement de façon linéaire.

FIGURE 4.13: *Mean Excess plot* sur les données réelles



La droite verticale bleue représente la valeur du seuil suggéré par le *Gerstengarbe plot*. Les droites verticales noires sont des seuils à partir duquel l'actuaire peut considérer que le *Mean Excess plot* devient linéaire.

TABLE 4.13: Seuils des extrêmes suggérés par le *Mean Excess plot*

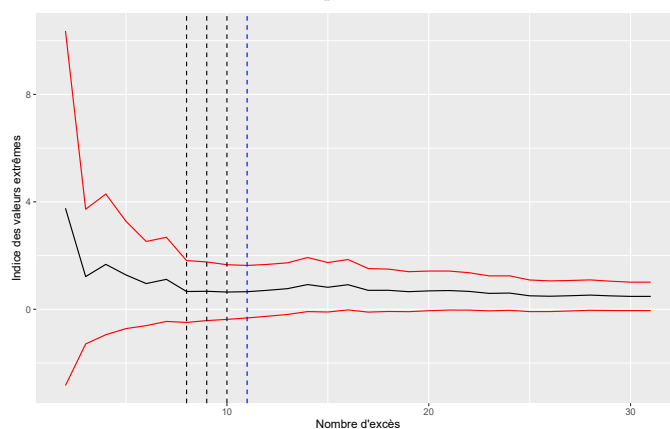
Seuil
1 205 809 €
1 097 539 €
1 050 758 €
1 013 741 €

Les seuils suggérés par la fonction de la moyenne des excès sont plus faibles que ceux suggérés par le graphique de Hill. Jusqu'ici le seuil indiqué par le *Gerstengarbe plot* est un bon compromis.

Gamma plot

Le dernier outil graphique utilisé pour aider l'actuaire dans le choix du seuil des extrêmes est le *Gamma plot*. Une valeur convenable de seuil correspond à un nombre d'excès à partir duquel le *Gamma plot* se stabilise.

FIGURE 4.14: *Gamma plot* sur les données réelles



La droite verticale bleue représente le nombre d'excès suggéré par le *Gerstengarbe plot*. Les droites verticales noires sont des nombres d'excès à partir desquels l'actuaire peut considérer que le *Gamma plot*

se stabilise.

Les nombres d'excès suggérés par le *Gamma plot* sont identiques à ceux suggérés par le *Hill plot*. Pour terminer l'analyse des outils d'aide au choix du seuil des extrêmes, la valeur numérique de minimisation de l'AMSE est calculée.

Minimisation de l'AMSE

Cette méthode numérique fournit la valeur du seuil qui minimise l'erreur quadratique moyenne asymptotique de l'estimateur de Hill.

FIGURE 4.15: Minimisation de l'AMSE sur données réelles

```
> damse(data_r_exemple_2$montant_sinistre)
$second.order.par
[1] 0.9893705 -0.9270152

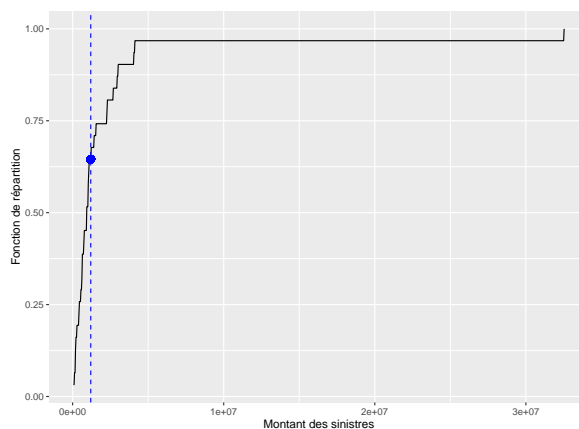
$k0
[1] 11

$threshold
[1] 1205809
```

Le seuil numérique fournit par cette méthode est de 1 205 809 €.

Finalement, l'étude des différents outils de sélection de seuils mène à retenir plusieurs seuils candidats. Cependant, un seul seuil est commun aux différents outils, il vaut 1 205 809€. La valeur retenue pour le seuil des extrêmes utilisé pour construire le modèle à deux distributions est de 1 200 000€.

FIGURE 4.16: Fonction de répartition empirique avec le seuil des extrêmes



Le seuil des extrêmes sélectionné correspond à un point de cassure de la fonction de répartition empirique. Il y a 35% des données qui sont supérieures au seuil des extrêmes. Ces 35% des données représentent 83% de la charge totale de la cédante. La suite s'attache à modéliser les sinistres supérieurs au seuil et à construire la fonction de répartition à deux lois.

4.2.6 Construction du modèle à deux lois usuelles

Le modèle à deux lois usuelles repose sur le choix d'un seuil des extrêmes. En dessous de ce seuil, les données sont modélisées par une loi classique, ici une distribution Log-Normale. Les données supérieures au seuil sont modélisées par une loi de Pareto Généralisée. Une fonction de répartition est ensuite construite à partir des deux distributions.

Modélisation des sinistres supérieurs au seuil des extrêmes

Une fois le choix du seuil effectué, l'actuaire modélise les sinistres supérieurs au seuil à l'aide d'une distribution de Pareto Généralisée, comme le suggère le théorème de Pickands, Balkema et De Hann.

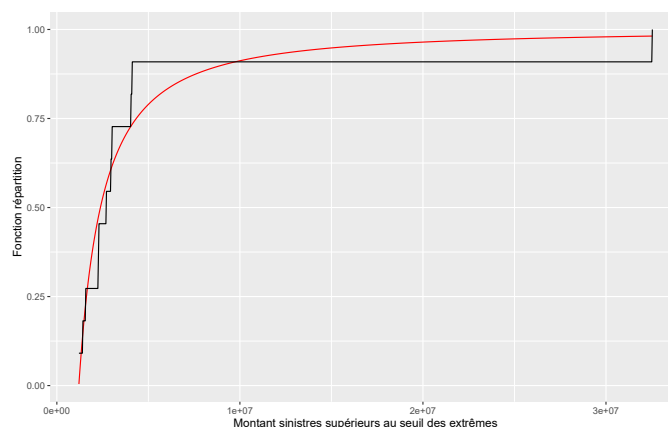
Il y a deux paramètres à estimer. Un paramètres d'échelle σ et un paramètre de forme γ . Le paramètre de forme est l'indice des valeurs extrêmes, celui-ci doit être positif car nous avons supposé que le domaine d'attraction auquel appartiennent les données est celui de Fréchet. Les deux paramètres sont estimés par la méthode du maximum de vraisemblance et celle des moments pondérés.

TABLE 4.14: Estimation des paramètres de la loi de Pareto Généralisée sur les données réelles

Paramètres	Méthode d'estimation	
	Maximum de vraisemblance	Moments pondérés
γ	0,7329	0,6592
σ	1 304 933	1 387 967

Dans la suite, ce sont les paramètres estimés par la méthode du maximum de vraisemblance qui sont utilisés. Le graphique ci-dessous illustre la modélisation des données supérieures au seuil des extrêmes par la loi de Pareto Généralisée.

FIGURE 4.17: Modélisation des sinistres supérieurs au seuil des extrêmes



Visuellement, l'adéquation est bonne. Il reste à construire la fonction de répartition à deux lois usuelles.

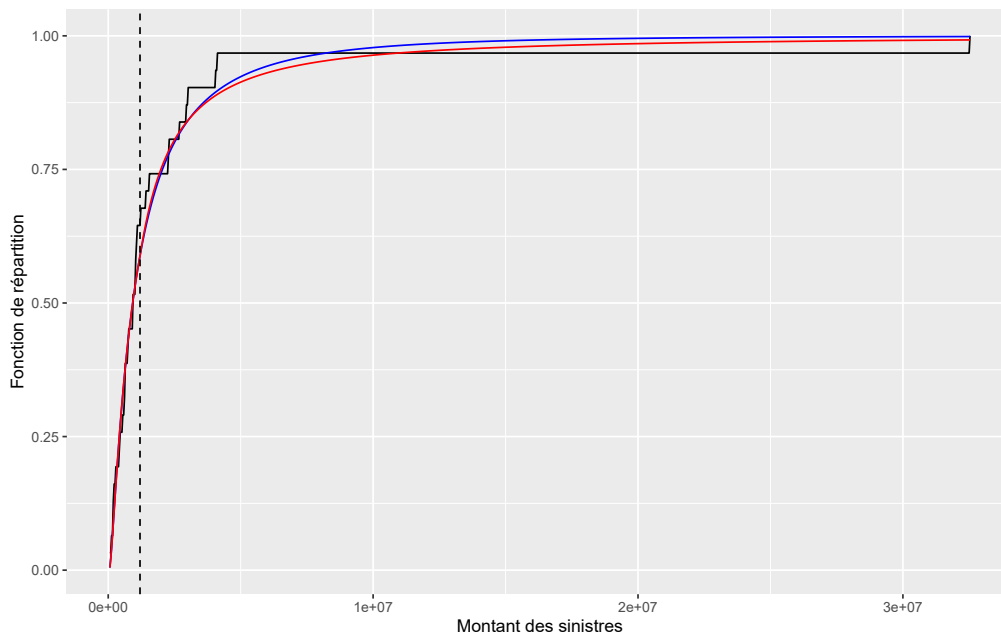
Construction du modèle à deux distributions usuelles

La modélisation finale à deux lois est la combinaison des deux lois suivantes : Une distribution Log-Normale tronquée à gauche au seuil de 50 000, de paramètres $\mu = 13,72$ et $\sigma = 1,19$, de fonction de répartition notée F^{LN} . Une distribution de Pareto Généralisée au seuil de 1 200 000, de paramètres $\gamma = 0,7329$ et $\sigma = 1 304 933$, de fonction de répartition notée $F^{\mathcal{GPD}}$. La fonction de répartition à deux lois associée s'exprime comme :

$$F(x) = \begin{cases} \frac{F^{LN}(x) - F^{LN}(50\,000)}{F^{LN}(50\,000)} & \text{si } x \in [50\,000, 1\,200\,000] \\ \frac{F^{LN}(1\,200\,000) - F^{LN}(50\,000)}{F^{LN}(50\,000)} + \left(1 - \frac{F^{LN}(1\,200\,000) - F^{LN}(50\,000)}{F^{LN}(50\,000)}\right) \times F^{\mathcal{GPD}}(x) & \text{si } x \geq 1\,200\,000 \end{cases}$$

Le graphique ci-dessous illustre la fonction de répartition empirique avec la fonction de répartition du **modèle à une loi** (en **bleue**) et la fonction de répartition du **modèle à deux lois usuelles** (en **rouge**). La droite verticale noire identifie le seuil des extrêmes.

FIGURE 4.18: Distribution empirique, modèle à une loi et à deux lois usuelles sur données réelles



Par construction, la modélisation sous le seuil des extrêmes est identique dans les deux modèles. Entre le seuil et 5 000 000, les deux modélisations sont proches. Au-delà de 5 000 000, le modèle à deux lois usuelles se distingue par une plus grande prudence des montants élevés. Cette plus grande prudence est en adéquation avec la possibilité que des sinistres importants surviennent, comme ce fût le cas avec le sinistre à 32 000 000. Graphiquement l'impact du modèle à deux lois est perceptible mais son intérêt est accru dans son application à l'estimation des quantiles et surtout du coût à charge de la réassurance pour des tranches hautes. La sous-section suivante compare le modèle à deux lois usuelles et le modèle d'Erlang.

4.2.7 Comparaison modèle à deux lois usuelles et modèle d'Erlang

Les modèles basés sur deux distributions sont différents par rapport à la loi utilisée pour modéliser la sinistralité sous le seuil des extrêmes. Dans le modèle à deux lois usuelles, la loi utilisée est une loi classique Log-Normale. Le modèle d'Erlang utilise la distribution mélange d'Erlang pour représenter les sinistres sous le seuil des extrêmes. La première comparaison est graphique, les suivantes sont numériques via les résultats des tests statistiques et la valeur de l'erreur quadratique moyenne dans les deux modèles.

Les graphiques suivants permettent de comparer visuellement la fonction de répartition empirique et les fonctions de répartition estimées des modèles à deux lois. La comparaison est effectuée sur la modélisation sous le seuil des extrêmes et sur la modélisation au-delà du seuil.

La courbe **bleue** est la fonction de répartition du **modèle à une loi**, la fonction de répartition du **modèle à deux lois usuelles** est représentée **rouge** et la fonction de répartition du **modèle d'Erlang** est représentée en **verte**. La droite verticale noire identifie le seuil des extrêmes.

FIGURE 4.19: Comparaison graphique des modèles à deux lois sous le seuil des extrêmes

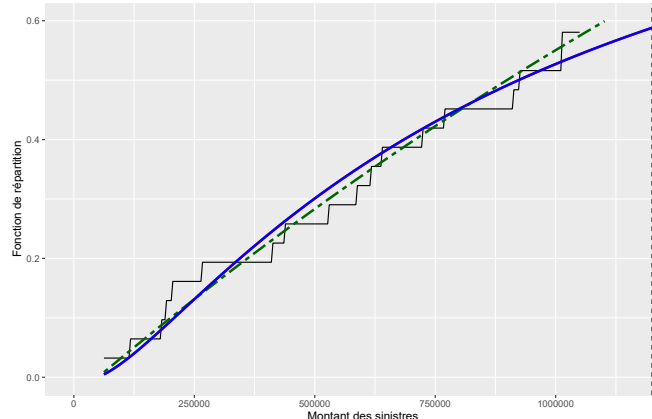
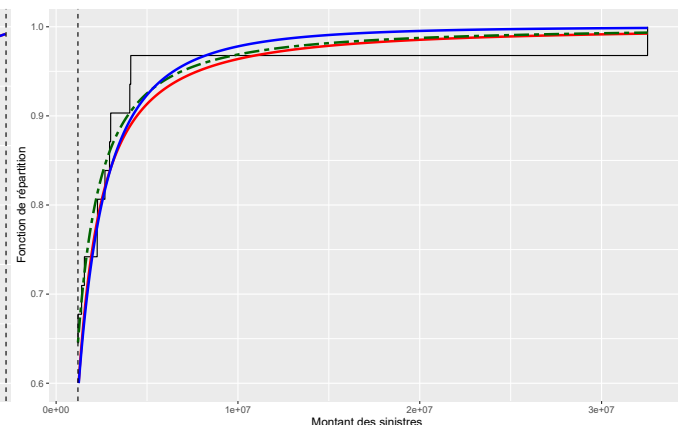


FIGURE 4.20: Comparaison graphique des modèles à deux lois au-delà du seuil des extrêmes



Sous le seuil des extrêmes, la fonction de répartition du modèle à une loi et du modèle à deux lois usuelles sont identiques par construction. Le modèle d'Erlang semble légèrement plus prudent tout en étant plus proche de la sinistralité observée. L'écart entre les deux modèles s'accroît à partir de 825 000 quand la sinistralité s'approche du seuil des extrêmes. Cela s'explique par le fait que dans le modèle d'Erlang, les sinistres sous le seuil sont modélisés par une densité mélange d'Erlang tronquée à gauche et à droite, seuls les sinistres entre le seuil de modélisation et le seuil des extrêmes sont pris en compte pour estimer les paramètres de la loi. Dans le modèle à deux lois usuelles, la sinistralité sous le seuil est modélisée par une loi Log-Normale dont les paramètres sont estimés sur l'ensemble des sinistres, la fonction de répartition de ce modèle est tirée vers le bas par les plus gros sinistres, cet effet est accentué en s'approchant du seuil des extrêmes. Au-delà du seuil des extrêmes, jusqu'à 5 000 000, le modèle d'Erlang est moins prudent mais semble plus en adéquation avec la sinistralité observée. Pour les sinistres supérieurs à 5 000 000, les modèles à deux lois sont plus prudents. Le modèle d'Erlang semble être un bon compromis entre la prudence et l'adéquation aux données. Les tests statistiques et l'erreur quadratique moyenne permettent de terminer la comparaison entre le modèle à deux lois usuelles et le modèle d'Erlang.

TABLE 4.15: Critères numériques de comparaison entre les modèles à deux lois

Modèle à deux lois	Critères	Test de Kolmogorov	Test d'Anderson	MSE
Usuelles	Distance	0,0875	0,298	0,001885
	p-valeur	0,955	0,940	
Erlang	Distance	0,0702	0,170	0,000799
	p-valeur	0,995	0,9965	

La p-valeur conduit à ne pas rejeter l'hypothèse nulle des tests. Les distances de Kolmogorov-Smirnov et d'Anderson-Darling sont inférieures dans le modèle d'Erlang. L'erreur quadratique moyenne du modèle d'Erlang est également inférieure à celle du modèle à deux lois usuelles.

Finalement, après l'étude des différents critères de comparaison entre les deux modèles, celui d'Erlang semble plus adéquat pour modéliser les sinistres de la cédante européenne.

La sous-section suivante permet de comparer les différents modèles par rapport à l'estimation des quantiles et des coûts à charge de la réassurance, en particuliers pour des tranches hautes.

4.2.8 La technique du bootstrap

Le problème de l'actuaire en réassurance est le peu de données dont il dispose pour modéliser la sinistralité. Dans le cadre de la tarification de traité en excédent de sinistre par l'approche probabi-

liste, l'actuaire détermine la loi la plus adéquate aux données et estime les paramètres associés. Le faible nombre de données impacte négativement la précision des estimations. Cette imprécision se répercute sur l'estimation du coût à charge de la réassurance pour une tranche de réassurance donnée. Dès lors, il est essentiel de mesurer le degré d'incertitude autour des estimations afin d'apprécier le risque convenablement.

Le bootstrap repose sur le ré-échantillonnage des observations disponibles. Soit (X_1, \dots, X_n) un échantillon de variables aléatoires indépendantes et identiquement distribuées, (x_1, \dots, x_n) sa réalisation, θ le paramètre d'intérêt à estimer et $\hat{\theta}(x_1, \dots, x_n)$ un estimateur de θ . La technique de ré-échantillonnage consiste à construire un nombre $B \leq n^n$ d'échantillons de taille n à partir des données de l'échantillon initial. Chacun des B échantillons est obtenu par un tirage aléatoire avec remise dans l'échantillon disponible initialement où toutes les observations ont la même probabilité $\frac{1}{n}$ d'être tirées. On obtient alors B nouveaux échantillons d'observations $(x_1^b, \dots, x_n^b)_{1 \leq b \leq B}$. Puis pour chaque échantillon, une estimation $\hat{\theta}^b$ de θ est calculée. Finalement, on dispose alors de B estimations $(\hat{\theta}^b)_{1 \leq b \leq B}$ du paramètre d'intérêt.

En posant ensuite :

$$\hat{\mu}(\hat{\theta}^1, \dots, \hat{\theta}^B) = \frac{1}{B} \sum_{i=1}^B \hat{\theta}^i$$

Et :

$$\hat{\sigma}^2(\hat{\theta}^1, \dots, \hat{\theta}^B) = \frac{1}{B-1} \sum_{i=1}^B (\hat{\theta}^i - \hat{\mu})^2$$

Le théorème central limite permet de construire un intervalle de confiance asymptotique de niveau $1 - \alpha$ par :

$$IC_\alpha = \left[\hat{\mu} - \frac{\hat{\sigma}}{\sqrt{B}} \varphi^{-1}\left(1 - \frac{\alpha}{2}\right) ; \hat{\mu} + \frac{\hat{\sigma}}{\sqrt{B}} \varphi^{-1}\left(1 - \frac{\alpha}{2}\right) \right]$$

où φ^{-1} est le quantile d'ordre α de la loi normale centrée réduite.

Dans la suite, la technique du bootstrap est appliquée pour étudier l'incertitude sur le coût à charge de la réassurance dans le modèle à une loi mais également sur l'estimation des quantiles dans le modèle.

Application sur le coût à charge de la réassurance dans le modèle à une loi

Afin de mesurer l'incertitude sur l'estimation du coût à charge de la réassurance, dans le cas où les données sont modélisées par une loi Log-Normale, la technique du bootstrap est appliquée sur l'échantillon de 31 sinistres. Pour ce faire, $B = 10\,000$ valeurs du couple de paramètres (μ, σ) sont estimées.

FIGURE 4.21: Histogramme des estimations bootstrap du paramètre μ de la Log-Normale

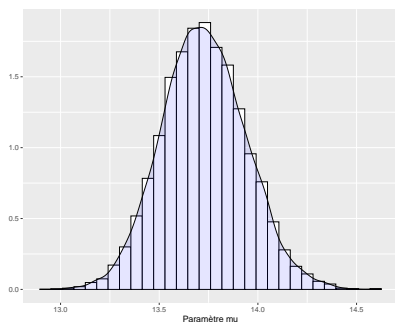


FIGURE 4.22: QQ-plot Normale des estimations bootstrap du paramètre μ de la Log-Normale

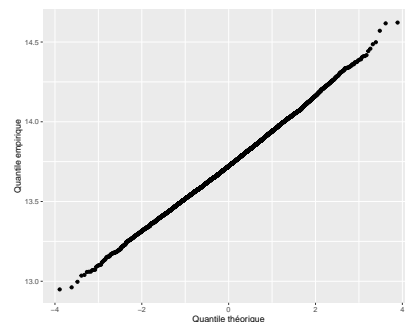


FIGURE 4.23: Histogramme des estimations bootstrap du paramètre σ de la Log-Normale

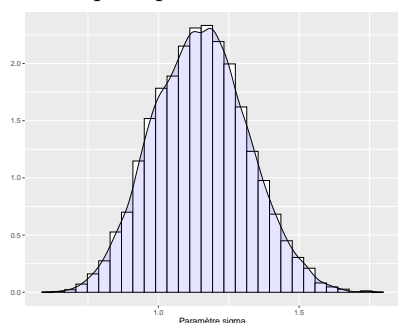
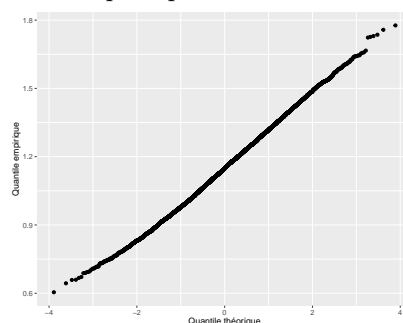


FIGURE 4.24: QQ-plot Normale des estimations bootstrap du paramètre σ de la Log-Normale



Les distributions empiriques des deux paramètres d'intérêt laisse supposer une distribution gaussienne. Les QQ-plot sont linéaires et le test de Kolmogorov-Smirnov ne permet pas de rejeter l'hypothèse nulle selon laquelle les données suivent une loi Normale.

La valeur de l'espérance bootstrap, du coefficient de variation, des paramètres μ et σ estimés sur les données initiales ainsi qu'un intervalle de confiance basé sur la normalité asymptotique de l'échantillon bootstrap sont fournis dans le tableau suivant :

TABLE 4.16: Statistiques sur les paramètres bootstrap de la Log-Normale

Statistiques	Paramètre μ	Paramètre σ
Espérance	13,7254	1,1519
Coefficient de variation	1,58%	14,80%
Valeur estimée sur données initiales	13,7245	1,1867
Intervalle de confiance à 95%	[13, 30 ; 14, 15]	[0, 887 ; 1, 556]

Lors de la modélisation de la sévérité, l'incertitude autour des paramètres de la loi Log-Normale impacte directement l'estimation du coût à charge de la réassurance. L'objectif est de fournir des éléments supplémentaires à la simple estimation ponctuelle d'un coût à charge pour une tranche donnée. Lorsque les sinistres sont modélisés par une loi Log-Normale, le coût à charge de la réassurance pour une tranche de priorité D supérieure au seuil de modélisation s et de portée C, notée $\pi(D, C)$ s'exprime par une formule fermée :

$$\pi(D, C) = \frac{1}{\varphi\left(\frac{\mu - \ln(s)}{\sigma}\right)} \left(e^{\mu + \frac{\sigma^2}{2}} \left[\varphi\left(\frac{\sigma^2 - \ln(D) + \mu}{\sigma}\right) - \varphi\left(\frac{\sigma^2 - \ln(D+C) + \mu}{\sigma}\right) \right] + (D+C)\varphi\left(\frac{\mu - \ln(D+C)}{\sigma}\right) - D\varphi\left(\frac{\mu - \ln(D)}{\sigma}\right) \right)$$

Autrement dit, en remplaçant les 10 000 valeurs des couples (μ, σ) , le coût à charge de la réassurance peut être vu comme une variable aléatoire dont il est possible d'exprimer certaines caractéristiques.

TABLE 4.17: Analyse du coût à charge dans le modèle à une loi Log-Normale pour différentes tranches

	Tranche			
	T1 : 500k XS 100k	T2 : 1M XS 1M	T3 : 5M XS 5M	T4 : 10M XS 10M
Espérance	403 259€	341 964€	213 231€	124 753€
Écart-type	22 515€	75 373€	151 357€	129 637€
Coefficient de variation	6%	22%	71%	104%
Coût estimé sur l'échantillon initial	400 790€	348 242€	208 713€	105 161€

Ce tableau illustre l'impact de l'incertitude liée à l'estimation des paramètres sur l'estimation du coût à charge de la réassurance. Plus la tranche est haute, c'est-à-dire moins il y a de données d'expérience disponibles, plus l'incertitude sur l'estimation du coût à charge est grande. Cette incertitude est reflétée par le coefficient de variation. Le coefficient de variation est relativement faible pour la tranche basse T_1 mais augmente rapidement sur les tranches plus hautes pour atteindre jusqu'à 104% sur la tranche T_4 . Il est également possible de fournir la distribution cumulée pour ces quatre tranches afin de mieux évaluer l'incertitude.

TABLE 4.18: Analyse des quantiles du coût à charge dans le modèle à une loi Log-Normale pour différentes tranches

Quantile	Tranche			
	T1 : 500k XS 100k	T2 : 1M XS 1M	T3 : 5M XS 5M	T4 : 10M XS 10M
0%	304 085€	94 501€	943€	24€
5%	366 228€	217 619€	34 643€	6 759€
10%	374 662€	243 621€	50 314€	11 894€
15%	380 279€	261 234€	65 267€	17 496€
20%	384 785€	276 965€	80 164€	24 373€
25%	388 660€	289 818€	95 946€	32 131€
30%	392 045€	301 426€	113 837€	40 533€
35%	395 051€	312 202€	129 767€	50 702€
40%	397 989€	322 389€	147 422€	61 464€
45%	400 817€	331 455€	164 288€	72 228€
50%	403 510€	341 687€	181 781€	84 094€
55%	406 328€	351 419€	199 338€	96 496€
60%	409 087€	362 164€	218 600€	111 077€
65%	411 946€	372 033€	239 010€	127 371€
70%	414 797€	382 971€	264 087€	147 352€
75%	418 378€	393 968€	292 571€	171 331€
80%	422 244€	406 397€	326 517€	202 578€
85%	426 143€	421 529€	367 261€	241 102€
90%	431 777€	439 833€	424 243€	295 748€
95%	439 461€	464 185€	508 961€	385 567€
100%	480 668€	625 772€	1 078 135€	1 186 288€
Écart maximum	58%	562%	114 174%	4 893 820%

Sur la tranche la plus basse T_1 , l'estimation du coût à charge de la réassurance varie de 304 085€ à 480 668€, soit un écart de 58%. Cet écart est d'autant plus important que la tranche est haute. L'incertitude liée à l'estimation des paramètres a donc un impact important sur l'incertitude qui entoure l'estimation du coût à charge de la réassurance. Pour prendre en compte cette incertitude dans le modèle, un coefficient de sécurité est ajouté au taux pur de réassurance. L'expression retenue pour ce coefficient de sécurité est :

$$CS = \min \left(5\% \times \frac{\hat{\sigma}_S}{\hat{\mu}_S} ; 20\% \right)$$

avec $\hat{\sigma}_S$ et $\hat{\mu}_S$ l'écart-type et la moyenne de la charge S pour la réassurance. Ces deux quantités sont estimées par un bootstrap lors de la modélisation.

Application aux quantiles

Dans les modèle à deux lois, les intervalles de confiance sont construits à partir de la normalité asymptotique des estimateurs du maximum de vraisemblance des paramètres de la loi de Pareto Généralisée.

Pour le modèle à une loi, dans le cas d'une distribution Log-Normale, il est malaisé d'obtenir une formule fermée pour l'intervalle de confiance des quantiles. La technique du bootstrap est utilisée grâce

au $B = 10\,000$ couples de paramètres (μ, σ) estimés précédemment. Pour chaque couple de paramètres, le quantile est calculé, de fait, un échantillon de 10 000 quantiles est construit. Cet échantillon est noté $(\hat{q}_i(\alpha))_{1 \leq i \leq B}$. L'hypothèse de normalité de l'échantillon bootstrap des quantiles n'est pas vérifiée. La construction d'intervalle de confiance basé sur l'hypothèse de normalité ne peut être utilisée. Pour s'affranchir de cette hypothèse, les intervalles de confiance sont construits grâce à la méthode des percentiles.

La méthode des percentiles La méthode des percentiles est simple à mettre en place. Les 10 000 estimations ponctuelles obtenues sont classées par ordre croissant et les bornes de l'intervalle de confiance sont constituées des estimations correspondantes. L'intervalle de confiance ainsi construit est un intervalle de confiance empirique du B -échantillon bootstrap. En notant $\hat{q}_{1,B}(\alpha) \leq \dots \leq \hat{q}_{B,B}(\alpha)$ les estimations ponctuelles des quantiles bootstrap, l'intervalle de confiance de niveau α s'exprime comme :

$$IC_\alpha = \left[\hat{q}_{\lfloor \frac{B\alpha}{2} \rfloor, B} ; \hat{q}_{\lfloor (1-\frac{\alpha}{2})B \rfloor, B} \right]$$

où $\lfloor \cdot \rfloor$ est la fonction partie entière.

TABLE 4.19: Estimation des quantiles sur données réelles

Ordre du quantile	Modèle une loi : Log-Normale	Modèle deux lois usuelles	Modèle d'Erlang	Empirique
90%	4 198 251€	4 444 216€	3 924 486€	2 999 267€
95%	6 456 361€	7 770 507€	6 902 159€	4 076 896€
99%	14 481 118€	26 585 156€	23 707 779€	24 003 776€
99,5%	19 465 604€	44 568 479€	39 742 377€	28 268 362€

Les quantiles estimés du modèle à deux lois usuelles sont supérieurs aux quantiles estimés du modèle à une loi. Le modèle à une loi sous-estime le risque, en particulier dans les tranches hautes. Cela s'explique par le fait que la distribution Log-Normale est une loi à queue fine contrairement aux modèles à deux lois où la queue de distribution est modélisée par une loi de Pareto Généralisée à queue épaisse. Les trois premiers quantiles du modèle d'Erlang sont relativement proches des quantiles empiriques tout en étant supérieurs. Le dernier quantile est grandement surestimé par les modèles à deux lois tandis que le modèle à une loi le sous-estime nettement. De manière générale, le modèle à deux lois usuelles semble surestimer le risque par rapport au modèle d'Erlang. Ce dernier apparaît être un bon compris entre les trois modèles.

Le tableau suivant fournit des intervalles de confiance pour ces quantiles.

TABLE 4.20: Intervalle de confiance des quantiles sur données réelles

Ordre du quantile	Modèle une loi : Log-Normale		Modèle deux lois usuelles		Modèle d'Erlang	
	Borne inférieure	Borne supérieure	Borne inférieure	Borne supérieure	Borne inférieure	Borne supérieure
90%	2 166 976€	8 283 636€	3 768 857€	5 119 576€	3 467 355€	4 381 618€
95%	3 029 279€	14 054 541€	5 596 981€	9 944 032€	5 283 013€	8 521 305€
99%	5 526 653€	37 739 461€	9 897 996€	43 272 317€	10 299 583€	37 115 976€
99,5%	6 898 025€	54 210 120€	9 370 626€	79 766 332€	11 073 890€	68 410 864€

Les intervalles de confiance des quantiles sont très larges, que ce soit par la construction bootstrap du modèle à une loi ou par la construction basée sur la normalité asymptotique des estimateurs du maximum de vraisemblance dans le modèle à deux lois. L'amplitude importante des intervalles de confiance reflète l'incertitude des estimations liées à la faible quantité d'informations dont l'actuaire en réassurance dispose pour effectuer les modélisations. Ces intervalles de confiance présentent l'avantage d'indiquer à l'actuaire que les estimations obtenues sont relativement imprécises.

Finalement, l'étude des quantiles dans les trois modèles révèle que le modèle à une loi sous-estime le risque. L'intérêt des modèles à deux lois est de construire une fonction de répartition à queue épaisse afin d'évaluer convenablement le risque des tranches hautes non travaillantes. La sous-section suivante permet de comparer les trois modèles au niveau des coût à charge de la réassurance pour différentes tranches.

4.2.9 Comparaison du montant à charge des trois modèles

La modélisation à une loi avec la distribution Log-Normale est comparée aux modèles à deux distributions avec introduction de la loi de Pareto Généralisée au seuil de 1 200 000€. Cette comparaison s'effectue sur le coût moyen à charge du réassureur pour différentes tranches de réassurance.

TABLE 4.21: Comparaison du coût à charge de la réassurance entre les trois modèles

Tranche	Modèle une loi	Modèle deux lois usuelles	Modèle d'Erlang
T_1 : 500 000 XS 100 000	400 790€	400 790€	404 255€
T_2 : 1 000 000 XS 500 000	488 208€	486 905€	465 606€
T_3 : 3 000 000 XS 3 000 000	287 490€	310 320€	267 340€
T_4 : 7 000 000 XS 5 000 000	245 224€	340 243€	292 842€
T_5 : 10 000 000 XS 10 000 000	105 161€	226 312€	194 522
T_6 : 20 000 000 XS 20 000 000	39 284€	180 788€	155 105€
T_7 : 30 000 000 XS 40 000 000	9 758€	117 634€	100 735€
T_8 : 40 000 000 XS 60 000 000	3 829€	93 881€	80 301€

La première tranche concerne la sinistralité sous le seuil des extrêmes fixé à 1 200 000€, elle fait uniquement intervenir la modélisation sous le seuil dans les modèles à deux lois. Les primes sont identiques par construction dans le modèle à une loi et le modèle à deux lois usuelles. Le modèle d'Erlang estime un coût très proche. La deuxième tranche fait intervenir la sinistralité de part et d'autre du seuil des extrêmes, l'estimation du coût à charge est proche dans les trois modèles. Les tranches moyennes T_3 et T_4 et T_5 , concernent la sinistralité supérieure au seuil des extrêmes. L'écart entre le modèle à une loi et les modèles à deux lois croît très rapidement. Le modèle à deux lois usuelles semble surestimer le coût. Les trois dernières tranches sont des tranches qui appartiennent à la queue de distribution des sinistres, les deux dernières sont non-travaillantes, il n'y pas de données d'expérience disponibles. L'écart est très important entre le modèle à une loi et les deux autres. Pour les tranches très hautes, le modèle d'Erlang estime des coûts proches du modèle à deux lois usuelles.

Pour les tranches affectant la sinistralité supérieure au seuil des extrêmes, le modèle à une loi a tendance à sous-estimer le risque. C'est tout l'intérêt de la construction d'un modèle à deux lois pour améliorer la tarification des tranches peu voir par travaillantes, où une simple loi ne peut refléter le risque convenablement. Dans les tranches moyennes, le modèle à deux lois usuelles semble surestimer le risque, le modèle d'Erlang apparaît être un bon compromis pour ces tranches. Dans les tranches hautes, les modèles à deux lois convergent vers des coûts similaires. Une autre manière de comparer les modèles est de considérer le coût à charge de la réassurance pour des tranches de portées égales.

TABLE 4.22: Comparaison du coût à charge de la réassurance entre les trois modèles pour des tranches de portées constantes

Tranche	Modèle une loi	Modèle deux lois usuelles	Modèle d'Erlang
T_1 : 1 000 000 XS 1 000 000	348 242€	342 803€	299 663€
T_2 : 1 000 000 XS 2 000 000	202 159€	197 768€	170 443€
T_3 : 1 000 000 XS 3 000 000	130 868€	133 928€	115 404€
T_4 : 1 000 000 XS 4 000 000	90 704€	99 022€	85 303€
T_5 : 1 000 000 XS 5 000 000	65 919€	77 370€	66 631€
T_6 : 1 000 000 XS 10 000 000	19 973€	34 030€	29 270€
T_7 : 1 000 000 XS 15 000 000	8 580€	20 461€	17 582€
T_8 : 1 000 000 XS 20 000 000	4 406€	14 139€	12 140€
T_9 : 1 000 000 XS 40 000 000	703€	5 688€	4 874€
T_{10} : 1 000 000 XS 50 000 000	363€	4 225€	3 618€
T_{11} : 1 000 000 XS 75 000 000	100€	2 453€	2 098€
T_{12} : 1 000 000 XS 100 000 000	37€	1 665€	1 422€

L'estimation des coûts à charge pour la réassurance diminue avec la hauteur de la tranche car la probabilité que la tranche soit touchée par un sinistre diminue également. Pour les quatre premières tranches, le modèle d'Erlang est le moins prudent, en particulier pour T_1 et T_2 . Pour ces quatre premières tranches, les résultats du modèle à une loi et du modèle à deux lois usuelles sont très proches. Pour les deux tranches suivantes, les estimations des modèles à deux lois sont supérieures et l'écart se creuse avec la hauteur de la tranche. A partir de la tranche T_7 , la différence entre le modèle à une loi et les modèles à deux lois est notable. Pour les dernières tranches non travaillantes, le modèle à une loi estime un coût à charge très faible par rapport aux modèles à deux lois. Le modèle d'Erlang est plus prudent que le modèle à deux lois usuelles, mais dans les tranches hautes, les estimations de ces deux modèles sont proches.

Finalement, le modèle à une loi est basé sur une distribution Log-Normale à queue fine. Cette distribution apparaît inadaptée pour la tarification des tranches hautes. Dans ces tranches, le modèle à une loi sous-estime le risque porté par le réassureur. Les modèles à deux lois permettent de capter le risque dans ces tranches hautes, peu ou pas travaillantes. L'introduction de la loi de Pareto Généralisée pour modéliser la queue de distribution trouve alors tout son intérêt lors de la tarification de tranches hautes. Le modèle d'Erlang est moins prudent que le modèle à deux lois usuelles pour les tranches basses et moyennes. En revanche, pour les tranches hautes les modèles à deux lois estiment des coûts relativement proches.

La sous-section suivante est une étude de sensibilité au choix du seuil des extrêmes. Cette analyse permet de rendre-compte de l'impact du choix de la valeur du seuil sur l'estimation des quantiles et notamment sur l'estimation des coûts à charge de la réassurance.

4.2.10 Sensibilité au seuil des extrêmes dans le modèles à deux lois usuelles

La modélisation sur les données réelles a montré que le choix du seuil n'est pas unanime et est sujet à une part de subjectivité. Deux actuaires pourraient sélectionner des seuils différents pour les mêmes observations. Cette sous-section permet d'étudier la sensibilité à la valeur du seuil de l'estimation des quantiles et du coût à charge de la réassurance. Pour ce faire, cinq seuils sont sélectionnés, ils valent 900 000€, 1 000 000€, 1 200 000€, 1 400 000€ et 1 500 000€. Pour chacun des seuils, les paramètres de la loi de Pareto Généralisée associée sont estimés par la méthode du maximum de vraisemblance. Le modèle à deux lois usuelles correspondant est ensuite construit.

TABLE 4.23: Estimation des paramètres γ et σ de la loi de Pareto Généralisée en fonction du seuil des extrêmes

Seuil des extrêmes	$\hat{\gamma}$	$\hat{\sigma}$	Nombre d'excès
900 000€	1,030	586 415	17
1 000 000€	1,097	609 560	15
1 200 000€	0,733	1 304 934	11
1 400 000€	0,771	1 287 812	10
1 500 000 €	0,721	1 523 519	9

Pour la modélisation des sinistres réels de la cédante, le seuil des extrêmes a été sélectionné à 1 200 000€. On remarque qu'à partir de cette valeur, l'estimation du paramètres d'indice de queue $\hat{\gamma}$ se stabilise. Ce résultat illustre la propriété de stabilité de l'indice des valeurs extrêmes.

Impact sur le niveau des quantiles

Les cinq modèles à deux lois sont différents par leur seuil des extrêmes. Pour ces modèles, les quantiles sont estimés et analysés.

TABLE 4.24: Sensibilité de l'estimation des quantiles à la valeur du seuil des extrêmes

Seuil des extrêmes	Quantile 90%	Quantile 95%	Quantile 99%	Quantile 99,5%
900 000€	3 371 771€	6 542 717€	32 954 814€	66 974 902€
1 000 000€	3 498 401€	6 976 252€	38 606 970€	82 063 431€
1 200 000€	4 444 216€	7 770 507€	26 585 156€	44 568 479€
1 400 000€	4 233 323€	7 415 948€	26 323 812€	45 119 342€
1 500 000€	4 496 637€	7 809 875€	26 271 064€	43 704 322€
Écart relatif maximum sur les 5 seuils	33,4%	19,37%	46,96%	87,77%
Écart relatif maximum sur les 3 derniers seuils	6,22%	5,31%	1,20%	3,24%

Avec un seuil des extrêmes à 900 000€ et 1 000 000€, l'estimation des quantiles les plus hauts est supérieure à l'estimation de ces mêmes quantiles avec un seuil des extrêmes supérieur à 1 200 000€. Pour les deux premiers seuils, l'indice des valeurs extrêmes estimé, qui contrôle la forme de la queue de distribution, est supérieur à celui estimé pour les trois derniers seuils. De fait, la queue de distribution des modèles aux seuils des extrêmes à 900 000€ et 1 000 000€ est plus épaisse, ce qui a pour conséquence une hausse du niveau des quantiles élevés. Pour les trois derniers seuils, l'écart relatif sur les quantiles ne dépasse pas les 7%, les estimations sont proches. Finalement, avec un seuil des extrêmes à 900 000€ et 1 000 000€, il y a un risque de surestimation de la sinistralité. Bien qu'en tarification, une attention particulière est accordée à la prudence dans l'estimation des quantités, il convient tout de même de refléter au mieux le risque pour ne pas le sur-tarifier et ne pas réussir à souscrire une part du traité. De plus, pour ces deux seuils, le nombre d'excès utilisé est important, en sachant que seulement 31 sinistres sont disponibles pour la modélisation. La notion de seuil ni trop petit ni trop grand apparaît.

Impact sur le coût à charge de la réassurance

L'impact de la valeur du seuil des extrêmes sélectionné est également illustré au travers l'estimation du coût à charge de la réassurance.

TABLE 4.25: Sensibilité de l'estimation du coût à charge à la valeur du seuil des extrêmes

Seuil des extrêmes	Tranches (priorité et portée en million d'euros)					
	1 XS 1	2 XS 2	5 XS 5	20 XS 10	30 XS 30	40 XS 60
900 000€	269 414€	238 066€	226 631€	358 749€	229 675€	171 699€
1 000 000€	283 172€	246 761€	242 188€	401 001€	269 668€	208 625€
1 200 000€	342 803€	331 697€	275 972€	337 156€	157 440€	93 881€
1 400 000€	342 823€	317 928€	260 176€	326 117€	159 413€	98 578€
1 500 000€	347 187€	337 927€	278 189€	334 744€	153 480€	90 345€
Écart relatif maximum sur les 5 seuils	28,87%	41,95%	22,75%	22,96%	75,70%	130,92%
Écart relatif maximum sur les 3 derniers seuils	1,28%	6,29%	6,92%	3,39%	3,87%	9,11%

Le constat est similaire à l'analyse de la sensibilité de l'estimation des quantiles au seuil des extrêmes. Pour les deux seuils les plus faibles, le modèle surestime le coût à charge dans les tranches hautes. En revanche, pour les tranches basses, le modèle à deux distributions avec les seuils à 900 000€ et 1 000 000€ sous-estime le coût à charge de la réassurance. En effet, par exemple le modèle à une loi Log-Normale estime un coût à charge pour une tranche 1 000 000 XS 1 000 000 de 348 242€. Pour ce modèle à une loi, l'adéquation a été validée en amont pour modéliser la sinistralité sous le seuil. Ainsi, avec les deux premiers seuils, le risque est sous-évalué pour les tranches basses. A partir du seuil des extrêmes à 1 200 000€, les estimations des coûts à charge sont stables. L'écart maximum est de 9,11% pour la tranche la plus haute.

Finalement, à travers cette étude de sensibilité, l'arbitrage dans le choix du seuil est parfaitement illustré. Le choix du seuil a un impact important sur les quantités estimées. Si le seuil sélectionné est trop faible, le modèle tend à sous-estimer le risque dans les montants bas mais à surestimer le risque dans les montants hauts. En revanche, au-delà d'une certaine valeur de seuil, l'estimation des quantités se stabilise. Le constat est similaire pour la sensibilité des résultats à la valeur du seuil dans le modèle d'Erlang. Ainsi, dans le cas où les outils de sélection de seuils conduisent à retenir plusieurs candidats, il est préférable de ne pas conserver les seuils les plus faibles.

La sous-section suivante présente la notion de période de retour qui introduit une mesure de récurrence dans la survenance des sinistres.

4.2.11 Estimation de la période de retour

L'estimation des quantiles permet de décrire le comportement de la queue de distribution du phénomène étudié. En gestion des risques, en particulier dans le domaine de la réassurance, la notion de quantile est complétée par la notion de période de retour.

La période de retour T_m d'un sinistre de niveau z_m s'interprète comme la durée moyenne qui sépare deux sinistres supérieurs ou égaux au montant z_m . Un niveau de sinistre z_m de période de retour T_m est donc un montant de sinistre qui est en moyenne dépassé toutes les T_m unités de temps. L'unité de temps généralement utilisée est l'année. La période de retour peut être liée à une probabilité : une période de retour T_m a une probabilité $p_m = \frac{1}{T_m}$ d'être dépassée chaque année.

Appliquons cette définition dans le cas où la variable $X|X > u$ est modélisée par une loi de Pareto Généralisée avec un indice des valeurs extrême positif. Le niveau de retour z_m associé à une période de retour de m années (c'est-à-dire le niveau de sinistre qui sera en moyenne dépassé toutes les m années) est solution de l'équation suivante :

$$\begin{aligned}\mathbb{P}(X > z_m) &= \frac{1}{m} \\ \mathbb{P}(X > u) \times \left[1 + \frac{\gamma(z_m - u)}{\sigma}\right]^{-\frac{1}{\gamma}} &= \frac{1}{m}\end{aligned}$$

En notant $\epsilon_u = \mathbb{P}(X > u)$ et en isolant z_m , il vient que :

$$z_m = u + \frac{\sigma}{\gamma} \left[(m \times \epsilon_u)^\gamma - 1 \right]$$

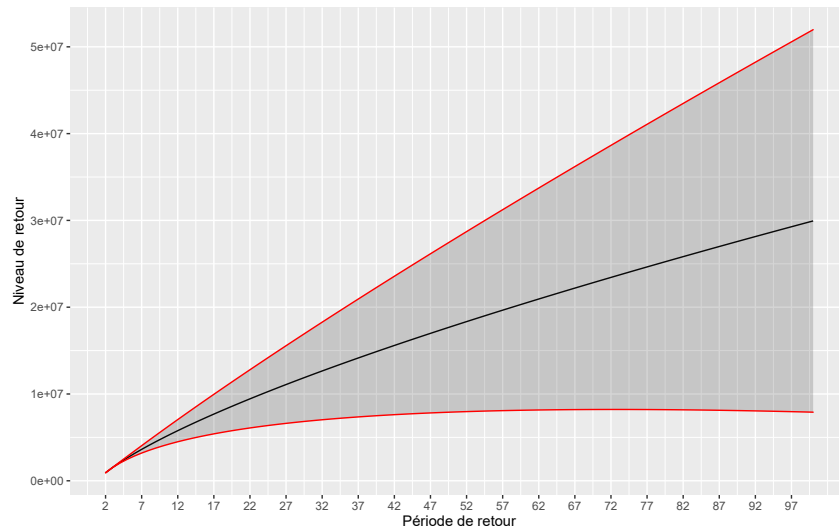
En remplaçant la valeur de γ et σ par les estimations obtenues par la méthode du maximum de vraisemblance, on obtient une estimation du niveau de retour \hat{z}_m associé à une période de retour de m années. La méthode Delta présentée en [3.6.3](#) permet de construire un intervalle de confiance pour l'estimation de la période de retour. Le tableau ci-dessous fournit les résultats obtenus sur les données réelles de la cédante modélisées :

TABLE 4.26: Estimation de la période de retour

Période de retour	Borne inférieure	Niveau de retour	Borne supérieure
2	916 530	924 348	932 165
3	1 537 006	1 548 257	1 559 508
4	2 044 176	2 115 611	2 187 046
5	2 477 348	2 644 307	2 811 265
6	2 857 165	3 144 106	3 431 047
7	3 196 102	3 621 088	4 046 074
8	3 502 395	4 079 370	4 656 346
9	3 781 826	4 521 906	5 261 986
10	4 038 644	4 950 908	5 863 173
11	4 276 079	5 368 090	6 460 101
12	4 496 657	5 774 812	7 052 966
13	4 702 401	6 172 178	7 641 956
14	4 894 955	6 561 102	8 227 249
15	5 075 683	6 942 348	8 809 012
16	5 245 729	7 316 565	9 387 401
17	5 406 064	7 684 312	9 962 560
18	5 557 519	8 046 071	10 534 623
19	5 700 816	8 402 266	11 103 715
20	5 836 583	8 753 267	11 669 952
21	5 965 372	9 099 407	12 233 441
22	6 087 671	9 440 978	12 794 285
23	6 203 915	9 778 245	13 352 575
24	6 314 491	10 111 446	13 908 401
25	6 419 748	10 440 796	14 461 845
26	6 520 001	10 766 492	15 012 983
27	6 615 534	11 088 712	15 561 889
28	6 706 609	11 407 620	16 108 632
29	6 793 460	11 723 367	16 653 274
30	6 876 305	12 036 092	17 195 879
31	6 955 342	12 345 922	17 736 502

Plus la période de retour est élevée, plus l'intervalle de confiance est large. Cela s'explique par le fait qu'une période de retour importante implique un niveau de retour associé élevé qui appartient à la queue de distribution. La faible quantité de données disponibles rend l'estimation de ces quantités plus volatile. Bien que l'intervalle de confiance soit large pour certaines périodes de retour, cela permet tout de même à l'actuaire d'avoir un aperçu du risque et d'avoir conscience de la volatilité liée au faible nombre d'observations fournies par la cédante. Le graphique suivant permet de mieux rendre compte de la difficulté de juger de la pertinence de l'estimation du niveau de retour pour les hautes périodes de retour tant l'intervalle est large :

FIGURE 4.25: Estimation de la période de retour



Au-delà d’une période de retour de 20 années, l’estimation du niveau de retour est très volatile. Pour les périodes de retour inférieures, il est possible de faire certaines analyses.

En moyenne, tous les trois ans, un sinistre de montant supérieur ou égal à 1 548 257€ devrait se produire. Le niveau de retour associé à une période de retour de 10 années vaut 4 985 908€ avec un intervalle de confiance entre 4 038 644€ et 5 863 173€. Autrement dit, en moyenne tous les dix ans, il devrait y avoir un sinistre de montant supérieur à ce niveau.

Les données fournies par la cédante font état d’un montant de sinistre de 32 532 948€. Ce montant est associé à une période de retour estimée à 112 années, soit une probabilité de ne pas être dépassé en moyenne pour une année donnée de $1 - \frac{1}{112} = 0,99107$. Ainsi, en huit ans, en supposant la survenance des sinistres indépendante, la probabilité de ne pas observer de sinistre supérieur à 32 532 948€ est de $0,99107^8 = 0,9308$ soit une probabilité de $1 - 0,9308 = 0,0692$ d’observer au moins un sinistre supérieur à 32 532 948€ en 8 ans. Le fait d’avoir observé ce montant est relativement inattendu.

En reprenant ce raisonnement pour une période de retour de 10 ans associée au niveau de retour de 4 950 908€, la probabilité d’observer au moins un sinistre supérieur à 4 950 908€ en 8 ans est de 0,569. Il y a donc plus d’une chance sur deux d’observer au moins un sinistre supérieur à ce montant. La cédante en a observé trois.

Finalement, l’estimation des périodes de retour fournit une information complémentaire à l’actuaire en ajoutant la dimension temporelle aux montants des sinistres. Cependant il convient de souligner que le faible nombre de données rend les estimations volatiles à partir d’une certaine période de retour. Les intervalles de confiance reflètent l’incertitude liée à la faible quantité d’observations. De plus, un niveau de retour de z_m associé à une période de retour de T_m signifie qu’en moyenne, sur une période de T_m unités de temps, un sinistre de montant supérieur ou égal à z_m devrait être observé. Cette interprétation est en moyenne, il est tout à fait possible d’observer aucun sinistre supérieur à z_m durant plusieurs périodes de T_m unités de temps et subitement observer plusieurs sinistres supérieurs à z_m durant T_m unité de temps.

Conclusion

Le pôle tarification de la direction de réassurance non-vie du groupe Covéa souhaitait compléter et améliorer son outil de tarification des traités en excédent de sinistre, en particulier afin d'estimer plus justement le coût à charge de la réassurance pour les tranches qualifiées de non-travaillantes.

L'outil de tarification déjà en place permettait d'estimer la prime pure en utilisant le modèle collectif où la fréquence était modélisée par une loi de Poisson, et la sévérité par une loi de Pareto. Cependant, dans certains cas la distribution de Pareto n'était pas adaptée pour modéliser le montant des sinistres sans qu'aucune autre distribution ne soit disponible. Dans ces cas, la loi de Pareto était utilisée par défaut. Pour compléter l'outil de tarification, de nouvelles lois ont été implémentées pour modéliser le montant des sinistres. Le nouvel outil de modélisation propose différentes méthodes pour sélectionner la loi la plus adéquate aux données et les paramètres associés. L'inconvénient des lois ajoutées est qu'elles ont une queue de distribution légère. Autrement dit, ces lois sont adaptées pour estimer la sévérité dans une tranche de réassurance située dans la partie centrale de la distribution. Pour les tranches non-travaillantes faisant intervenir la queue de distribution, il est difficile d'estimer convenablement la charge pour la réassurance.

Pour palier à ce problème et améliorer l'outil de tarification, un modèle à deux distributions a été proposé. Ce modèle est basé sur la théorie des valeurs extrêmes, en particulier sur l'approche des excédents au-delà d'un seuil exposée par [Pickands \(1975\)](#). Cette théorie propose de modéliser les sinistres supérieurs à un certain seuil par une loi de Pareto Généralisée. Le point d'attention de la méthode réside dans le choix du seuil, qui est le point où les deux distributions sont raccordées. Pour définir ce seuil, différents outils issus de la théorie des valeurs extrêmes ont été mis en place. La méthode de Gerstengarbe notamment, qui alliée aux outils graphiques permet de guider l'actuaire dans son choix. Néanmoins, le choix du seuil peut être malaisé et plusieurs valeurs peuvent sembler convenir. L'étude de la valeur du seuil sur la sensibilité des quantités estimées propose alors de conserver un seuil intermédiaire.

L'application des modèles sur les données réelles de cédantes a permis d'illustrer le fait que dans certains cas, le modèle à une loi tend à sous estimer le risque dans les tranches hautes non sinistrées. Le modèle à deux lois permet de résoudre ce problème en évaluant plus justement la sinistralité prédite au regard de la sinistralité observée. Cette comparaison entre les deux modèles est sans équivoque lorsque la distribution du modèle à une loi est à queue fine. Dans le cas où elle est à queue épaisse, l'intérêt d'une modélisation à deux lois n'est pas aussi marqué. Le modèle à deux distributions peut également permettre de faciliter le choix du seuil de modélisation. L'actuaire peut utiliser tous les sinistres et appliquer une distribution à deux lois. Cela permet d'éviter le choix du seuil de modélisation et de prendre en compte toute l'information disponible pour refléter convenablement le risque.

Les modèles implémentés présentent certaines limites. Ils sont, en particulier, sensibles au nombre d'observations disponibles pour effectuer la modélisation. Les incertitudes sont mesurées par du bootstrap ou en utilisant les propriétés asymptotiques des estimateurs. Le niveau d'incertitude augmente avec la hauteur de la tranche. Les incertitudes permettent à l'actuaire de prendre conscience de la volatilité des résultats obtenus et d'en tenir compte dans la cotation. Certaines méthodes permettent de réduire l'incertitude des estimations dans le modèle collectif, c'est notamment le cas des modèles de Crédibilité présentés par exemple dans le mémoire de [Poulin \(2012\)](#). Dans le cas où l'actuaire juge que la quantité de données à sa disposition est trop faible, il peut décider de se tourner vers d'autres méthodes de tarification comme celle par exposition basée sur des profils de portefeuille.

Pour améliorer la tarification des tranches peu ou pas travaillantes, une fonction de répartition à deux lois a été implémentée pour modéliser la sévérité. La distribution mélange d'Erlang a été proposée pour modéliser les sinistres inférieurs au seuil. Une autre façon aurait pu être de conserver le modèle à une loi puis d'utiliser des fonctions de distorsions pour alourdir la queue de distribution et ainsi charger plus fortement la prime pure dans les tranches hautes. Ce sont les mesures de risque de Wang, présentées par exemple par Therond (2005).

L'étude réalisée dans le cadre de ce mémoire a donné lieu au développement d'un outil de modélisation complet reposant sur l'interface de programmation de *R Shiny*. L'application dispose de tous les outils mathématiques présentés pour permettre à l'actuaire de sélectionner la loi la plus adéquate aux données et construire un modèle à deux lois. Le développement de l'application va donner lieu au déploiement d'autres applications similaires dans les années à venir et ainsi s'inscrire dans un projet plus global de revue des outils de tarification de la Direction Réassurance Acceptations Internationales.

Table des figures

1.1	Quote-Part à 30% et plein de souscription à 130	12
1.2	Excédent de plein avec rétention de 50 et plein de souscription à 120	13
1.3	Quote-part 35% avec excédent de plein sur la rétention	14
1.4	Excédent de sinistre 50 XS 70	16
1.5	Excédent de perte 40% XS 80 %	17
1.6	Le groupe Covéa	19
1.7	Répartition des branches de réassurance au sein de la DRAI	20
1.8	Répartition des pays Européens au sein de la DRAI	20
2.1	Densité de la loi des valeurs extrêmes pour des valeurs différentes de l'indice des extrêmes	28
2.2	Exemple de <i>quantile plot généralisé</i>	35
2.3	<i>Mean excess plot</i> d'une GPD avec 500 observations.	37
2.4	<i>Mean excess plot</i> d'une GPD avec 50 observations.	37
2.5	<i>Gamma plot</i> sur données simulées selon une <i>GPD</i> .	38
2.6	Hill Plot sur 400 observations issues d'une Pareto(1,5; 10 000)	39
2.7	Pickands Plot sur 400 observations issues d'une Pareto(1,5; 10 000)	39
2.8	Hill Plot sur 100 observations issues d'une Pareto(1,5; 10 000)	40
2.9	Pickands Plot sur 100 observations issues d'une Pareto(1,5; 10 000)	40
2.10	Exemple de <i>Gerstengarbe plot</i> sur données simulées.	41
3.1	Illustration du coût à charge de deux tranches de même amplitude situées sur un plateau de la fonction de répartition empirique	46
3.2	QQ-Plot sur simulation d'une loi de Pareto.	56
3.3	QQ-Plot sur simulation d'une loi Log-Normale.	56
3.4	Illustration de la prime pure d'un C XS D	67
4.1	QQ-plot sur données simulées	73
4.2	<i>Gerstengarbe plot</i> sur les données simulées	75
4.3	<i>Hill plot</i> sur les données simulées	75
4.4	<i>Mean Excess plot</i> sur les données simulées	76
4.5	<i>Gamma plot</i> sur les données simulées	76
4.6	Modélisation à deux lois sur des données simulées	77
4.7	Histogramme des sinistres Incendie en échelle logarithmique	80
4.8	QQ-plot sur données réelles	81
4.9	Ajustement graphique Log-Normale et Weibull sur données réelles	83
4.10	<i>Quantile plot généralisé</i> sur données réelles	84
4.11	<i>Gerstengarbe plot</i> sur les données réelles	84
4.12	<i>Hill plot</i> sur les données réelles	85
4.13	<i>Mean Excess plot</i> sur les données réelles	86
4.14	<i>Gamma plot</i> sur les données réelles	86
4.15	Minimisation de l'AMSE sur données réelles	87
4.16	Fonction de répartition empirique avec le seuil des extrêmes	87
4.17	Modélisation des sinistres supérieurs au seuil des extrêmes	88
4.18	Distribution empirique, modèle à une loi et à deux lois usuelles sur données réelles	89

4.19	Comparaison graphique des modèles à deux lois sous le seuil des extrêmes	90
4.20	Comparaison graphique des modèles à deux lois au-delà du seuil des extrêmes	90
4.21	Histogramme des estimations bootstrap du paramètre μ de la Log-Normale	91
4.22	QQ-plot Normale des estimations bootstrap du paramètre μ de la Log-Normale	91
4.23	Histogramme des estimations bootstrap du paramètre σ de la Log-Normale	92
4.24	QQ-plot Normale des estimations bootstrap du paramètre σ de la Log-Normale	92
4.25	Estimation de la période de retour	100
26	Importation des sinistres	115
27	Statistiques descriptives	115
28	Choix de la modélisation	116
29	Adéquation graphique de la loi sélectionnée	116
30	Choix du seuil des extrêmes	116
31	Modélisation au-delà du seuil des extrêmes	117
32	Comparaison de la modélisation à une et deux lois	117
33	Estimation du taux pur pour une fréquence unitaire	118
34	Exportation des résultats	118

Liste des tableaux

1.1 Les quatre formes de réassurance	11
1.2 Portefeuille fictif pour illustrer les formes de réassurance	12
1.3 Quote-Part à 30% et plein de souscription à 130	12
1.4 Excédent de plein avec rétention de 50 et plein de souscription à 120	13
1.5 Quote-part 35% avec excédent de plein sur la rétention	14
1.6 Excédent de sinistre 50 XS 70	15
1.7 Excédent de perte 40% XS 80 %	17
1.8 Calcul du taux <i>burning cost</i> pour la tranche 2 000 000 XS 1 000 000	23
1.9 Calcul du taux <i>burning cost</i> sur une tranche 30 000 000 XS 15 000 000	24
2.1 Domaine d'attractions des lois usuelles	28
3.1 Valeur de marché du paramètre α	48
3.2 Probabilité de survie de la Pareto en fonction de α	48
4.1 Résultats des tests statistiques sur les données simulées	73
4.2 MSE, AIC et BIC sur les données simulées	74
4.3 Estimation des paramètres de la loi de Pareto sur les données simulées	74
4.4 Estimation des paramètres de la loi de Pareto Généralisée sur les données simulées	77
4.5 Méthodologie de la mise en <i>as-if</i>	79
4.6 Evolution de l'indice du prix à la consommation	79
4.7 Mise en <i>as-if</i> sur données réelles	79
4.8 Statistiques descriptives de la sinistralité Incendie	80
4.9 Estimation des paramètres des distributions sur données réelles	81
4.10 Résultats des tests statistiques sur les données réelles	82
4.11 MSE, AIC et BIC sur les données réelles	82
4.12 Seuils des extrêmes suggérés par le <i>Hill plot</i>	85
4.13 Seuils des extrêmes suggérés par le <i>Mean Excess plot</i>	86
4.14 Estimation des paramètres de la loi de Pareto Généralisée sur les données réelles	88
4.15 Critères numériques de comparaison entre les modèles à deux lois	90
4.16 Statistiques sur les paramètres bootstrap de la Log-Normale	92
4.17 Analyse du coût à charge dans le modèle à une loi Log-Normale pour différentes tranches	92
4.18 Analyse des quantiles du coût à charge dans le modèle à une loi Log-Normale pour différentes tranches	93
4.19 Estimation des quantiles sur données réelles	94
4.20 Intervalle de confiance des quantiles sur données réelles	94
4.21 Comparaison du coût à charge de la réassurance entre les trois modèles	95
4.22 Comparaison du coût à charge de la réassurance entre les trois modèles pour des tranches de portées constantes	95
4.23 Estimation des paramètres γ et σ de la loi de Pareto Généralisée en fonction du seuil des extrêmes	96
4.24 Sensibilité de l'estimation des quantiles à la valeur du seuil des extrêmes	97
4.25 Sensibilité de l'estimation du coût à charge à la valeur du seuil des extrêmes	97
4.26 Estimation de la période de retour	99

Bibliographie

- Albrecher, H., Beirlant, J., and Teugels, J. (2017). *Reinsurance : Actuarial and Statistical Aspects (Wiley Series in Probability and Statistics)*. Wiley, 1 edition.
- Borchani, A. (2010). Statistiques des valeurs extrêmes dans le cas de lois discrètes. Centre de recherche de l'ESSEC ISSN : 1291-9616 10009.
- Caeiro, F. and Gomes, M. (2016). *Threshold Selection in Extreme Value Analysis*, pages 69–86. Chapman and Hall/CRC 2007.
- Charpentier, A. (2013). Actuariat iard - partie 8 : réassurance. *Présentation de cours, Université du Québec à Montréal*.
- Charpentier, A. and Dutang, C. (2012). L'actuariat avec r. *Support de cours*.
- Chesneau, C. (2017). Sur l'Estimateur du Maximum de Vraisemblance (emv). Lecture.
- Deme, E. H. (2013). *Quelques contributions à la Théorie univariée des Valeurs Extrêmes et Estimation des mesures de risque actuariel pour des pertes à queues lourdes*. Theses, Université Gaston Berger, Saint Louis, Sénégal.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society : Series B (Methodological)*, 39(1) :1–22.
- Deprey, E. and Godzinski, A. (2007). Problématique de seuil dans la modélisation de la sinistralité en réassurance non vie. *Mémoire d'actuaire, ENSAE*.
- Finas, C. (2016). Les tarifs des réassureurs sont-ils crédibles ? *Mémoire d'actuaire, ISFA*.
- Gerstengarbe, F. and Werner, P. (1989). A method for the statistical definition of extreme-value regions and their application to meteorological time series. *Zeitschrift fuer Meteorologie ;(German Democratic Republic)*, 39(4).
- Goffard, P. (2019). Modélisation charge sinistre m2 actuariat - chapitre 3 : Modèle collectif. *Présentation de cours, ISFA*.
- Gourmelon, V. (2014). Modélisation des sinistres graves en assurance automobile responsabilité civile. *Mémoire d'actuaire, EURIA*.
- Guillou, A. and You, A. (2011). Introduction à la théorie des valeurs extrêmes : Applications en actuariat. *Ecole d'été, Université de Strasbourg et Société Générale Assurance*.
- Hagstoem, K.-G. (1925). La loi de pareto et la réassurance. *Scandinavian Actuarial Journal*, 1925(2) :65–88.
- Hosking, J. R. M., Wallis, J. R., and Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3) :251–261.
- Ittah, M. (2016). Calibrage de la sinistralité non vie - cotation et optimisation de la réassurance non-proportionnelle. *Mémoire d'actuaire, DUAS*.

- Jal, P. (2001). Obtention d'intervalles de confiance en réassurance par la méthode du bootstrap. *Consulting Division Gecalux Group*.
- Laureille, X., Quennelle, P., and Ittah, M. (2019). Détermination des seuils de sinistres graves et optimisation de la réassurance. *Présentation 100% actuair 100% data science*, Paris.
- Maumy-Bertrand, M. and Bertrand, F. (2018). *Initiation à la statistique avec R - 3e éd.* DUNOD.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of statistics*, 3 :119–131.
- Piera-Martinez, M. (2008). *Modélisation des comportements extrêmes en ingénierie*. Theses, Université Paris Sud - Paris XI.
- Planchet, F. (2018). Modèles financiers en assurance et analyses dynamiques. *Support de cours, ISFA*.
- Poulin, M. (2012). Analyse des solutions actuarielles en tarification des traités de réassurance non-proportionnels non-vie. *Mémoire d'actuaire, CEA*.
- Smith, R. L. (1984). *Threshold Methods for Sample Extremes*, pages 621–638. Springer Netherlands, Dordrecht.
- Therond, P. (2005). Mesures et comparaison de risques. *Support de cours, ISFA*.
- Thuillier, M. (2021). Calcul de la valeur contrat sur la branche multirisque immeuble comme aide opérationnelle à la relation client. *Mémoire d'actuaire, ENSAE*.
- Tijms, H. (1994). *Stochastic Models : An Algorithmic Approach*. Wiley, Hoboken, NJ, États-Unis.
- Walhin, J. and Leflot, J. (2012). *La réassurance*. Éditions Larcier.
- Wang, Y., Hobæk Haff, I., and Huseby, A. (2020). Modelling extreme claims via composite models and threshold selection methods. *Insurance : Mathematics and Economics*, 91 :257–268.
- Zorba, R. (2017). Tarification des tranches illimitées des traités en excédent de sinistres de la branche responsabilité civile automobile. *Mémoire d'actuaire, DUAS*.

Annexes

A Calcul des coût à charge des lois usuelles

Dans cette annexe, les démonstrations des formules fermées pour l'obtention du coût à charge de la réassurance pour un traité C XS D sont présentées. En notant s le seuil de modélisation, l'actuaire ne peut tarifier que des tranches où $D \geq s$. Si ce n'est pas le cas, l'actuaire ne peut pas tarifier convenablement la tranche, il doit demander à ce que la cédante fournisse ses sinistres à un seuil d'avis plus faible.

La loi Log-Normale

Si $X \sim \mathcal{LN}(\mu, \sigma)$ avec $\mu \in \mathbb{R}$ et $\sigma > 0$ alors sa densité est de la forme :

$$\forall x > 0, \quad f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2}$$

Si $X \sim \mathcal{LN}(\mu, \sigma)$ alors la variable aléatoire $Y = \ln(X)$ suit une loi $\mathcal{N}(\mu, \sigma)$. Ainsi, $\frac{\ln(X)-\mu}{\sigma}$ suit une loi $\mathcal{N}(0, 1)$

La densité d'une variable aléatoire suivant une loi Log-Normale tronquée à gauche au seuil s devient alors :

$$f_s(x) = \frac{\frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2}}{\mathbb{P}(X > s)} \times \mathbb{1}_{x \geq s}$$

où $\mathbb{P}(X > s) = \mathbb{P}\left(\frac{\ln(X)-\mu}{\sigma} > \frac{\ln(s)-\mu}{\sigma}\right) = \varphi\left(\frac{\mu-\ln(s)}{\sigma}\right)$ avec φ la fonction de répartition d'une variable aléatoire qui suit une loi Normale centrée réduite.

Considérons alors un traité ∞ XS D . Supposons que $D \geq s$. Le coût moyen dans la tranche de ce traité vaut :

$$\pi(D, \infty) = \frac{\int_D^\infty \frac{(x-D)}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2} dx}{\varphi\left(\frac{\mu-\ln(s)}{\sigma}\right)}$$

Il faut alors calculer la quantité suivante :

$$\int_D^\infty \frac{(x-D)}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2} dx = \underbrace{\int_D^\infty \frac{x}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2} dx}_{(1)} - D \underbrace{\int_D^\infty \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2} dx}_{(2)}$$

D'une part :

$$(2) = \mathbb{P}(X \geq D) = \varphi\left(\frac{\mu - \ln(D)}{\sigma}\right)$$

D'autre part :

$$(1) = \int_D^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2} dx$$

Avec le changement de variable $t = \frac{\ln(x) - \mu}{\sigma}$

$$\begin{aligned} &= \int_{\frac{\ln(D) - \mu}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} e^{\sigma t + \mu} dt \\ &= e^{\mu} \int_{\frac{\ln(D) - \mu}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2} + \sigma t} dt \end{aligned}$$

En remarquant que $-\frac{t^2}{2} + \sigma t = -\frac{(t - \sigma)^2 - \sigma^2}{2}$ et en faisant le changement de variable $y = t - \sigma$

$$\begin{aligned} &= e^{\mu} \int_{\frac{\ln(D) - \mu}{\sigma} - \sigma}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[y^2 - \sigma^2]} dy \\ &= e^{\mu + \frac{\sigma^2}{2}} \int_{\frac{\ln(D) - \mu}{\sigma} - \sigma}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\ &= e^{\mu + \frac{\sigma^2}{2}} \times \mathbb{P}(Y \geq \frac{\ln(D) - \mu}{\sigma} - \sigma) \quad \text{avec } Y \sim \mathcal{N}(0, 1) \\ &= e^{\mu + \frac{\sigma^2}{2}} \times \mathbb{P}(Y \leq \sigma - \frac{\ln(D) - \mu}{\sigma}) \\ &= e^{\mu + \frac{\sigma^2}{2}} \varphi\left(\frac{\sigma^2 - \ln(D) + \mu}{\sigma}\right) \end{aligned}$$

Finalement, il vient que :

$$\pi(D, \infty) = \frac{1}{\varphi\left(\frac{\mu - \ln(s)}{\sigma}\right)} \left[e^{\mu + \frac{\sigma^2}{2}} \varphi\left(\frac{\sigma^2 - \ln(D) + \mu}{\sigma}\right) - D \varphi\left(\frac{\mu - \ln(D)}{\sigma}\right) \right]$$

Et pour un traité C XS D , avec la relation suivante $\pi(D, C) = \pi(D, \infty) - \pi(D + C, \infty)$ il vient que :

$$\pi(D, C) = \frac{1}{\varphi\left(\frac{\mu - \ln(s)}{\sigma}\right)} \left(e^{\mu + \frac{\sigma^2}{2}} \left[\varphi\left(\frac{\sigma^2 - \ln(D) + \mu}{\sigma}\right) - \varphi\left(\frac{\sigma^2 - \ln(D + C) + \mu}{\sigma}\right) \right] + \right. \\ \left. (D + C) \varphi\left(\frac{\mu - \ln(D + C)}{\sigma}\right) - D \varphi\left(\frac{\mu - \ln(D)}{\sigma}\right) \right)$$

La loi de Pareto

Si $X \sim \mathcal{P}(\alpha, A)$, avec $\alpha > 0$ et $A > 0$ alors sa densité est de la forme :

$$f(x) = \alpha A^\alpha x^{-\alpha-1} \mathbb{1}_{x \geq A}$$

Et sa fonction de répartition :

$$F(x) = (1 - A^\alpha x^{-\alpha}) \mathbb{1}_{x \geq A}$$

La fonction de répartition d'une variable aléatoire suivant une loi de Pareto tronquée à gauche au seuil s devient alors :

$$F_s(x) = \frac{F(x) - F(s)}{\bar{F}(s)} \Rightarrow \bar{F}_s(x) = \frac{\bar{F}(x)}{\bar{F}(s)} = A^\alpha x^{-\alpha} \mathbb{1}_{x \geq A} + \mathbb{1}_{x < A}$$

En effet, les paramètres sont estimés par maximum de vraisemblance. Pour une loi de Pareto, $\hat{A} = x_{1:n}$ avec $x_{1:n}$ la plus petite valeur observée au dessus du seuil s . Ainsi, on a $\hat{A} \geq s$ et donc $\bar{F}(s) = 1$.

Considérons alors un traité ∞ XS D et supposons $\alpha > 1$. Le coût moyen dans la tranche de ce traité vaut :

$$\pi(D, \infty) = \int_D^\infty \bar{F}_s(x) dx$$

Dans le cas où $D \geq A$:

$$\begin{aligned} \pi(D, \infty) &= A^\alpha \int_D^\infty x^{-\alpha} dx \\ &= A^\alpha \left[\frac{x^{1-\alpha}}{1-\alpha} \right]_D^\infty \\ &= \frac{A^\alpha D^{1-\alpha}}{\alpha-1} \end{aligned}$$

Pour résumer ces résultats, et en rappelant la relation $\pi(D, C) = \pi(D, \infty) - \pi(D + C, \infty)$ il vient que :

Cas	$\pi(D, C)$	$\pi(D, \infty)$
$D \geq A$	$\frac{A^\alpha D^{1-\alpha}}{1-\alpha} \left[\left(1 + \frac{C}{D}\right)^{1-\alpha} - 1 \right]$	$\frac{A^\alpha D^{1-\alpha}}{\alpha-1}$

La loi Exponentielle

Si $X \sim \mathcal{E}(\lambda)$, avec $\lambda > 0$ alors sa densité est de la forme :

$$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{x \geq 0}$$

Et sa fonction de répartition :

$$F(x) = (1 - e^{-\lambda x}) \mathbf{1}_{x \geq 0}$$

La fonction de répartition d'une variable aléatoire suivant une loi Exponentielle tronquée à gauche au seuil s devient alors :

$$F_s(x) = \frac{F(x) - F(s)}{\bar{F}(s)} \Rightarrow \bar{F}_s(x) = \frac{\bar{F}(x)}{\bar{F}(s)} = e^{-\lambda(x-s)} \mathbf{1}_{x \geq s} + 1 \mathbf{1}_{x < s}$$

Considérons alors un traité ∞ XS D . Le coût moyen dans la tranche de ce traité vaut :

$$\pi(D, \infty) = \int_D^\infty \bar{F}_s(x) dx$$

Dans le cas où $D \geq s$:

$$\pi(D, \infty) = \int_D^\infty e^{-\lambda(x-s)} dx$$

Avec le changement de variable $t = \lambda(x - s)$

$$\begin{aligned} &= \frac{1}{\lambda} \int_{\lambda(D-s)}^\infty e^{-t} dt \\ &= \frac{1}{\lambda} \left[-e^{-t} \right]_{\lambda(D-s)}^\infty \\ &= \frac{e^{-\lambda(D-s)}}{\lambda} \end{aligned}$$

Pour résumer ces résultats, et en rappelant la relation $\pi(D, C) = \pi(D, \infty) - \pi(D + C, \infty)$ il vient que :

Cas	$\pi(D, C)$	$\pi(D, \infty)$
$D \geq s$	$\frac{e^{-\lambda(D-s)}}{\lambda} [1 - e^{-\lambda C}]$	$\frac{e^{-\lambda(D-s)}}{\lambda}$

La loi de Weibull

Si $X \sim \mathcal{W}(a, b)$, avec $a > 0$ et $b > 0$ alors sa densité est de la forme :

$$f(x) = ab^{-a}x^{a-1}e^{-\left(\frac{x}{b}\right)^a} \mathbf{1}_{x \geq 0}$$

Et sa fonction de répartition :

$$F(x) = \left(1 - e^{-\left(\frac{x}{b}\right)^a}\right) \mathbf{1}_{x \geq 0}$$

La fonction de répartition d'une variable aléatoire suivant une loi de Weibull tronquée à gauche au seuil s devient alors :

$$F_s(x) = \frac{F(x) - F(s)}{\bar{F}(s)} \Rightarrow \bar{F}_s(x) = \frac{\bar{F}(x)}{\bar{F}(s)} = e^{-\frac{1}{b^a}[x^a - s^a]} \mathbf{1}_{x \geq s} + \mathbf{1}_{x < s}$$

Considérons alors un traité ∞ XS D . Le coût moyen dans la tranche de ce traité vaut :

$$\pi(D, \infty) = \int_D^\infty \bar{F}_s(x) dx$$

Dans le cas où $D \geq s$:

$$\begin{aligned} \pi(D, \infty) &= \int_D^\infty e^{-\frac{1}{b^a}[x^a - s^a]} dx \\ &= e^{\left(\frac{s}{b}\right)^a} \times \int_D^\infty e^{-\left(\frac{x}{b}\right)^a} dx \end{aligned}$$

Avec le changement de variable $t = \left(\frac{x}{b}\right)^a$

$$\begin{aligned} &= \frac{a}{b} e^{\left(\frac{s}{b}\right)^a} \int_{\left(\frac{D}{b}\right)^a}^\infty t^{\frac{1}{a}-1} e^{-t} dt \\ &= \frac{a}{b} e^{\left(\frac{s}{b}\right)^a} \Gamma\left(\left(\frac{D}{b}\right)^a; \frac{1}{a}\right) \end{aligned}$$

Avec $\forall x \geq 0, a > 0, \Gamma(x, a)$ la fonction Gamma incomplète supérieure définie comme :

$$\Gamma(x, a) = \int_x^\infty e^{-t} t^{a-1} dt$$

Sur le logiciel R, cette fonction est définie dans le package *pracma* avec la fonction *gammainc*.

Pour résumer ces résultats, et en rappelant la relation $\pi(D, C) = \pi(D, \infty) - \pi(D + C, \infty)$ il vient que :

Cas	$\pi(D, C)$	$\pi(D, \infty)$
$D \geq s$	$\left(\frac{b}{a}\right) e^{\left(\frac{s}{b}\right)^a} \left[\Gamma\left(\left(\frac{D}{b}\right)^a; \frac{1}{a}\right) - \Gamma\left(\left(\frac{D+C}{b}\right)^a; \frac{1}{a}\right) \right]$	$\frac{a}{b} e^{\left(\frac{s}{b}\right)^a} \Gamma\left(\left(\frac{D}{b}\right)^a; \frac{1}{a}\right)$

La loi Gamma

Si $X \sim \mathcal{G}(k, \theta)$, avec $k > 0$ et $\theta > 0$ alors sa densité est de la forme :

$$f(x) = \frac{x^{k-1}}{\theta^k \Gamma(k)} e^{-\frac{x}{\theta}} \mathbf{1}_{x \geq 0}$$

avec $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$.

La fonction de densité d'une variable aléatoire suivant une loi Gamma tronquée à gauche au seuil s devient alors :

$$f_s(x) = \frac{f(x)}{\bar{F}(s)} \mathbf{1}_{x \geq s}$$

avec $\bar{F}(s) = \frac{1}{\Gamma(k)} \Gamma\left(\left(\frac{s}{\theta}\right); k\right)$.

Considérons alors un traité ∞ XS D . Le coût moyen dans la tranche de ce traité vaut :

$$\pi(D, \infty) = \int_D^\infty (x - D) \frac{x^{k-1}}{\bar{F}(s) \theta^k \Gamma(k)} e^{-\frac{x}{\theta}} \mathbf{1}_{x \geq s} dx$$

Dans le cas où $D \geq s$:

$$\pi(D, \infty) = \frac{1}{\bar{F}(s) \theta^k \Gamma(k)} \left[\int_D^\infty x^k e^{-\frac{x}{\theta}} dx - D \int_D^\infty x^{k-1} e^{-\frac{x}{\theta}} dx \right]$$

Avec le changement de variable $t = \frac{x}{\theta}$

$$\begin{aligned} &= \frac{1}{\bar{F}(s) \theta^k \Gamma(k)} \left[\int_{\frac{D}{\theta}}^\infty \theta^{k+1} t^k e^{-t} dt - D \int_{\frac{D}{\theta}}^\infty \theta^k t^{k-1} e^{-t} dt \right] \\ &= \frac{1}{\bar{F}(s) \theta^k \Gamma(k)} \left[\theta^{k+1} \Gamma\left(\left(\frac{D}{\theta}\right); k+1\right) - D \theta^k \Gamma\left(\left(\frac{D}{\theta}\right); k\right) \right] \\ &= \frac{1}{\bar{F}(s) \Gamma(k)} \left[\theta \Gamma\left(\left(\frac{D}{\theta}\right); k+1\right) - D \Gamma\left(\left(\frac{D}{\theta}\right); k\right) \right] \end{aligned}$$

Pour résumer ces résultats, et en rappelant la relation $\pi(D, C) = \pi(D, \infty) - \pi(D + C, \infty)$ il vient que, pour tout traité de portée finie $C \in \mathbb{R}$:

Cas	$\pi(D, C)$
$D \geq s$	$\frac{1}{\bar{F}(s) \Gamma(k)} \left\{ \theta \left[\Gamma\left(\left(\frac{D}{\theta}\right); k+1\right) - \Gamma\left(\left(\frac{D+C}{\theta}\right); k+1\right) \right] + (D+C) \left[\Gamma\left(\left(\frac{D+C}{\theta}\right); k\right) - \Gamma\left(\left(\frac{D}{\theta}\right); k\right) \right] \right\}$

La loi Normale

Si $X \sim \mathcal{N}(\mu, \sigma)$ avec $\mu \in \mathbb{R}$ et $\sigma > 0$ alors sa densité est de la forme :

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

Dans le cas particulier d'une loi normale centrée et réduite, on notera φ sa fonction de répartition. Ainsi si $X \sim \mathcal{N}(\mu, \sigma)$ alors on a la relation suivante $\mathbb{P}(X \leq t) = \varphi\left(\frac{t-\mu}{\sigma}\right)$.

La densité d'une variable aléatoire suivant une loi Normale tronquée à gauche au seuil s devient alors :

$$f_s(x) = \frac{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\mathbb{P}(X > s)} \mathbf{1}_{x \geq s} = \frac{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\varphi\left(\frac{\mu-s}{\sigma}\right)} \mathbf{1}_{x \geq s}$$

Considérons alors un traité ∞ XS D . Supposons que $D \geq s$. Le coût moyen dans la tranche de ce traité vaut :

$$\pi(D, \infty) = \frac{\int_D^\infty \frac{(x-D)}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx}{\varphi\left(\frac{\mu-s}{\sigma}\right)}$$

Il faut alors calculer la quantité suivante :

$$\int_D^\infty \frac{(x-D)}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \underbrace{\int_D^\infty \frac{x}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx}_{(1)} - D \underbrace{\int_D^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx}_{\varphi\left(\frac{\mu-D}{\sigma}\right)}$$

Pour l'intégrale de gauche, en posant $y = \frac{x-\mu}{\sigma}$ il vient que :

$$\begin{aligned} (1) &= \int_{\frac{D-\mu}{\sigma}}^\infty \frac{\sigma y}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy + \mu \int_{\frac{D-\mu}{\sigma}}^\infty \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy \\ &= \frac{\sigma}{\sqrt{2\pi}} \left[-e^{-\frac{1}{2}y^2} \right]_{\frac{D-\mu}{\sigma}}^\infty + \mu \varphi\left(\frac{\mu-D}{\sigma}\right) \\ &= \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{D-\mu}{\sigma}\right)^2} + \mu \varphi\left(\frac{\mu-D}{\sigma}\right) \end{aligned}$$

Finalement, il vient que :

$$\pi(D, \infty) = \frac{1}{\varphi\left(\frac{\mu-s}{\sigma}\right)} \times \left[\frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{D-\mu}{\sigma}\right)^2} + (\mu-D)\varphi\left(\frac{\mu-D}{\sigma}\right) \right]$$

Et pour un traité C XS D avec la relation suivante $\pi(D, C) = \pi(D, \infty) - \pi(D+C, \infty)$ il vient que :

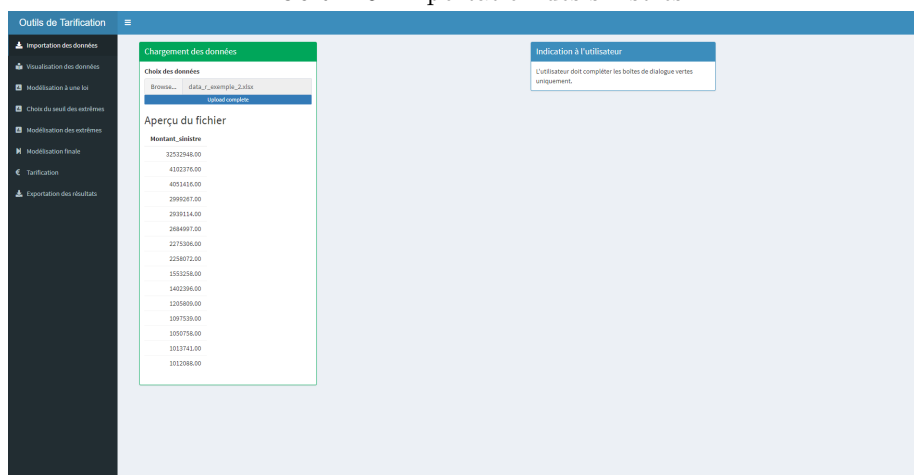
$$\pi(D, C) = \frac{1}{\varphi\left(\frac{\mu-s}{\sigma}\right)} \times \left[\frac{\sigma}{\sqrt{2\pi}} \left(e^{-\frac{1}{2}\left(\frac{D-\mu}{\sigma}\right)^2} - e^{-\frac{1}{2}\left(\frac{D+C-\mu}{\sigma}\right)^2} \right) + (\mu-D)\varphi\left(\frac{\mu-D}{\sigma}\right) - (\mu-D-C)\varphi\left(\frac{\mu-D-C}{\sigma}\right) \right]$$

B Présentation de l'outil de modélisation R Shiny

Le sujet présente un aspect opérationnel important car une application R Shiny a été développée. Cette application permet la modélisation de la sévérité des sinistres. L'application est opérationnelle pour la prochaine période de renouvellement et est intégrée à l'outil de tarification Excel déjà en place. La contrainte à respecter était de créer un outil de modélisation facile d'utilisation, même pour les personnes non initiées au langage R. Cette annexe présente l'interface de l'outil.

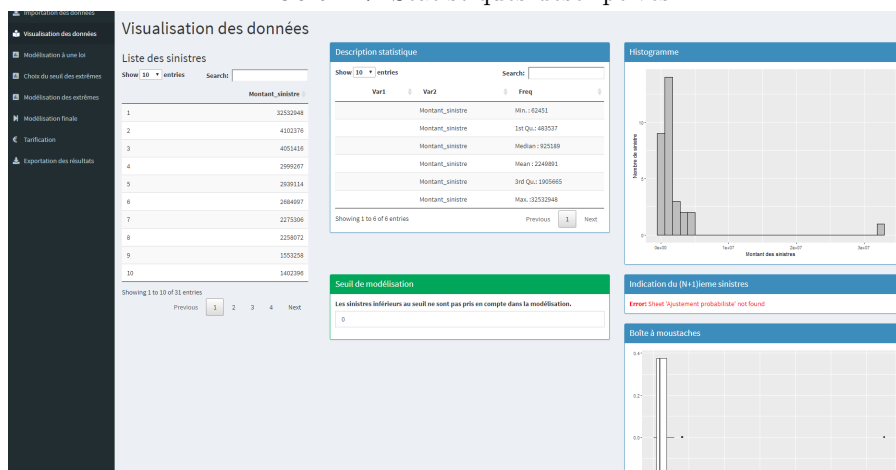
La première feuille de l'application permet à l'utilisateur de sélectionner le fichier de cotation Excel où se trouve l'ensemble des sinistres. L'outil récupère l'ensemble des montants et les stocke en mémoire pour la suite de la modélisation.

FIGURE 26: Importation des sinistres



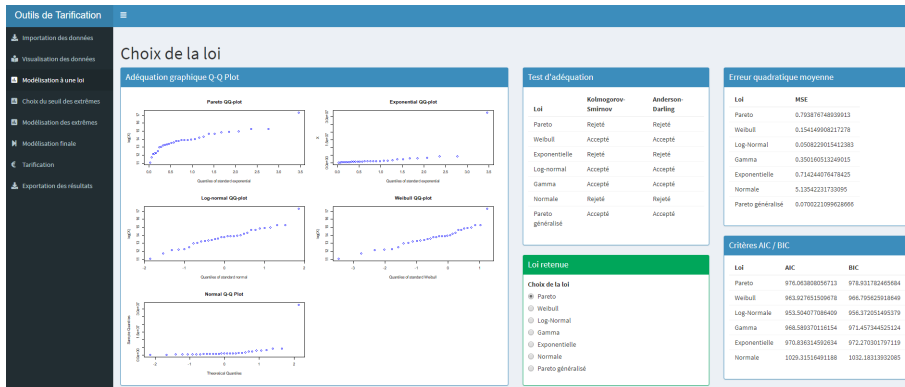
La seconde feuille de l'application fournit des statistiques descriptives sur les sinistres renseignés. L'utilisateur doit sélectionner un seuil de modélisation.

FIGURE 27: Statistiques descriptives



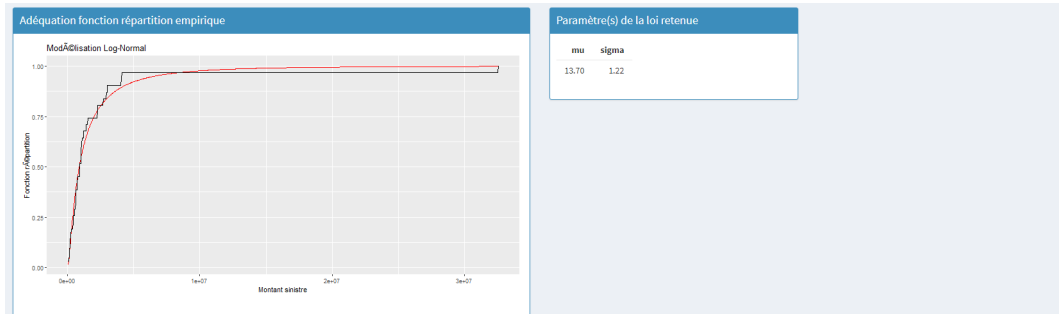
La troisième feuille de l'application permet à l'utilisateur de sélectionner la loi du modèle à une loi (qui est aussi la loi sous le seuil des extrêmes du modèle à deux lois). L'utilisateur est guidé dans son choix par différents outils statistiques. L'utilisateur doit ensuite cocher la loi qu'il souhaite conserver.

FIGURE 28: Choix de la modélisation



Les paramètres de la loi sélectionnée sont indiqués à l'utilisateur et le graphique de la fonction de répartition empirique et de la loi choisie est tracé.

FIGURE 29: Adéquation graphique de la loi sélectionnée



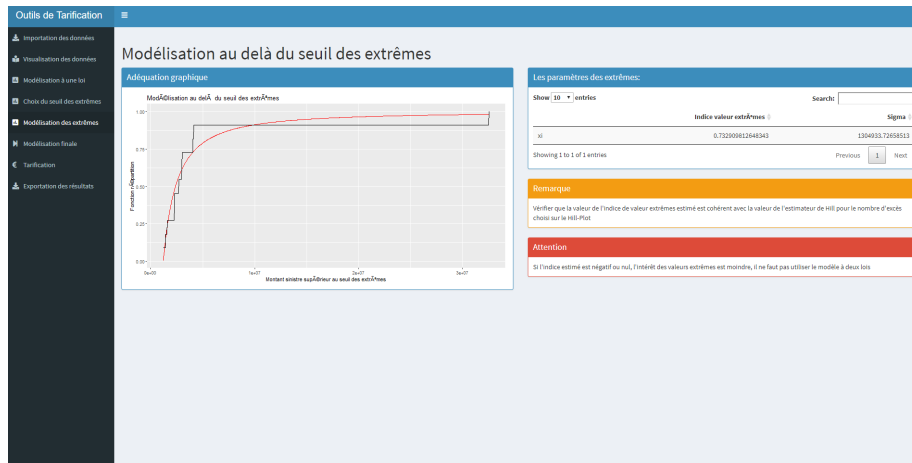
La quatrième feuille de l'application permet à l'utilisateur de sélectionner le seuil des extrêmes. La valeur du *Gerstengarbe plot* est indiquée, le *Hill plot* et le *Mean Excess Plot* permettent d'aider l'utilisateur dans le choix du seuil. L'utilisateur peut également renseigner un nombre d'excès pour obtenir la valeur du seuil correspondant. Une fois le seuil des extrêmes sélectionné, les deux graphiques circulaires fournissent une information sur la proportion du nombre de sinistres supérieurs au seuil et la charge totale associée.

FIGURE 30: Choix du seuil des extrêmes



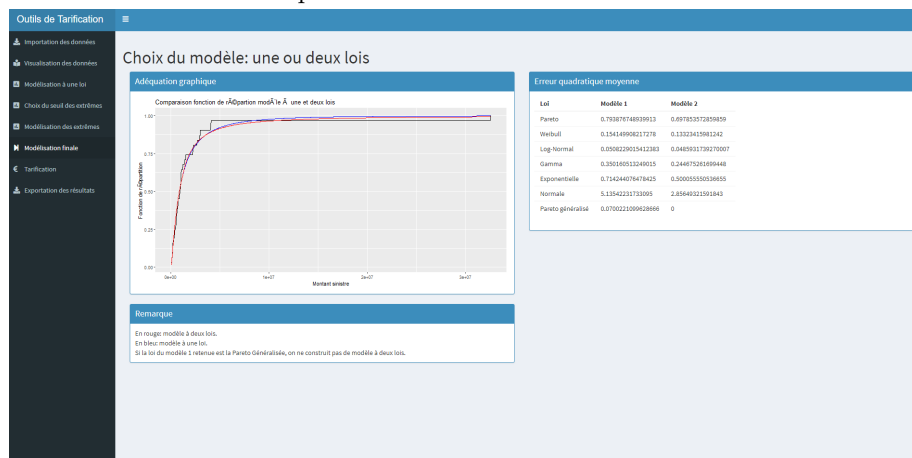
La feuille suivante de l'application permet à l'utilisateur de vérifier graphiquement l'adéquation de la loi de Pareto Généralisée sur l'échantillon des excès. L'utilisateur peut notamment vérifier que la valeur de l'indice des extrêmes estimée est cohérente avec la valeur pour laquelle le *Hill plot* se stabilise.

FIGURE 31: Modélisation au-delà du seuil des extrêmes



Cette feuille permet à l'utilisateur de comparer graphiquement les fonctions de répartition du modèle à une loi et du modèle à deux lois. Ces deux modèles sont également comparés grâce à l'erreur quadratique moyenne.

FIGURE 32: Comparaison de la modélisation à une et deux lois



Cette feuille permet à l'utilisateur d'estimer un taux pur de réassurance pour des tranches données en renseignant l'encaissement estimée de la cédante, les priorités et les portées. La fréquence est supposée unitaire.

FIGURE 33: Estimation du taux pur pour une fréquence unitaire

Tarification de traité en excédent de sinistre

EPI
Epi à tarifier
5000000

Taux

Tranche	Modèle une loi	Modèle deux lois
T1	0.6642%	0.6498%
T2	0.4328%	0.5406%
T3	0.0000%	0.0000%
T4	0.0000%	0.0000%
T5	0.0000%	0.0000%
T6	0.0000%	0.0000%
T7	0.0000%	0.0000%
T8	0.0000%	0.0000%

T1
Portée de la T1
2000000
Franchise de la T1
3000000

T2
Portée de la T2
5000000
Franchise de la T2
5000000

T3
Portée de la T3
1
Franchise de la T3
1

T4
Portée de la T4
1
Franchise de la T4
1

T5
Portée de la T5
1
Franchise de la T5
1

T6
Portée de la T6
1
Franchise de la T6
1

Enfin, la dernière feuille permet à l'utilisateur d'exporter l'ensemble des résultats pour les renseigner dans l'outil de tarification Excel. L'outil de tarification Excel déjà en place a été adapté pour pouvoir prendre en compte les différentes lois ajoutées et l'introduction du modèle à deux lois.

FIGURE 34: Exportation des résultats

Outils de Tarification

Copy Ctrl Excel PDF Paste

Search:

Tranche	Modèle une loi	Modèle deux lois	Pareto	Weibull	Log-Normal	Gamma	Exponentielle	Normale	GPD	GPD_2	conf_pareto	conf_weibull	conf	
beta	T1	0.6642%	0.6498%	0.375348934171975	0.70746960234802	13.700672250254	0.65829280819734	4.44460504503454	22.4989564316425	0.450554783714098	0.729209812648343	0.67102263759346	0.58000409508519	0.58
beta.1	T2	0.4328%	0.5406%	62451	1433075.67133271	1.1252752582827	3417937.65273165	0	5720454.65660011	1074926.74367609	1304933.72658513	0	0	
X	T3	0.0000%	0.0000%	0	0	0	0	0	0	0	0	0	0	
X.1	T4	0.0000%	0.0000%	0	0	0	0	0	0	0	0	0	0	
X.2	T5	0.0000%	0.0000%	0	0	0	0	0	0	0	0	0	0	
X.3	T6	0.0000%	0.0000%	0	0	0	0	0	0	0	0	0	0	
X.4	T7	0.0000%	0.0000%	0	0	0	0	0	0	0	0	0	0	
X.5	T8	0.0000%	0.0000%	0	0	0	0	0	0	0	0	0	0	

Showing 1 to 8 of 8 entries

Previous 1 Next