

Modélisation de la couverture du portefeuille Non-Marine  
d'un réassureur : Cas pratique Société centrale de  
réassurance

Maissoro Malam Idi

23 juin 2022

# Table des matières

<b>Remerciements</b>	<b>3</b>
<b>Résumé</b>	<b>4</b>
<b>Abstrack</b>	<b>5</b>
<b>Introduction</b>	<b>6</b>
<b>Chapitre 1 Chapitre préliminaire</b>	<b>7</b>
1.1 Présentation générale de la réassurance . . . . .	7
1.1.1 Les différentes formes de réassurance . . . . .	8
1.1.2 Le marché de réassurance . . . . .	9
1.2 Société centrale de réassurance . . . . .	10
1.2.1 Aperçu général . . . . .	11
1.2.2 Portefeuille client . . . . .	12
1.2.3 Analyse historique par catégorie . . . . .	13
1.2.4 Programme de Rétrocession . . . . .	14
<b>Chapitre 2 Les contrats classiques de réassurance</b>	<b>17</b>
2.1 Réassurance Proportionnelle . . . . .	17
2.1.1 Quote-Part . . . . .	17
2.1.2 Excédent de plein (EP) . . . . .	18
2.2 Réassurance non proportionnelle . . . . .	20
2.2.1 Excédent de sinistres . . . . .	20
2.2.2 Excédent de perte . . . . .	22
2.2.3 Les méthodes de Modélisation du contrat XS . . . . .	24
<b>Chapitre 3 Modélisation de la sinistralité</b>	<b>26</b>
3.1 Statistiques As-If . . . . .	27
3.2 Choix du seuil . . . . .	28
3.2.1 La loi des valeurs extrêmes - GEV(Generalized Extreme Value) . . . . .	28
3.2.2 Théorie des excédents au-delà d'un certain seuil . . . . .	30
3.2.3 Estimation paramétrique . . . . .	32
3.2.4 Estimation semi-paramétrique de l'indice des valeurs extrême . . . . .	33
3.2.5 Détermination du domaine d'attraction . . . . .	34
3.2.6 Détermination du seuil des extrêmes . . . . .	35
3.3 Les modèles de modélisation . . . . .	37
3.3.1 Modèle collectif . . . . .	37
3.3.2 Modèle collectif séparé . . . . .	38

3.4	Modélisation de la sévérité . . . . .	39
3.4.1	Fonction de répartition empirique . . . . .	39
3.4.2	Lois usuelles en réassurance . . . . .	40
3.4.3	Analyse préliminaire des distributions candidates : . . . . .	43
3.4.4	QQ-Plot . . . . .	45
3.4.5	Maximum de vraisemblance : Estimation des paramètres . . . . .	47
3.4.6	PP-Plot . . . . .	49
3.4.7	Test de Kolmogorov-Smirnov . . . . .	49
3.4.8	Test de d'Anderson-Darling . . . . .	50
3.4.9	Spécification du modèle . . . . .	50
3.4.10	Loi de la fréquence . . . . .	53
3.5	Application sur des données réels . . . . .	56
3.5.1	Statistiques As-If de nos données . . . . .	57
3.5.2	Statistiques descriptives . . . . .	58
3.5.3	Choix du seuil . . . . .	63
3.5.4	Modélisation de la sévérité . . . . .	67
3.5.5	Modélisation de la loi de fréquence . . . . .	73
3.5.6	Établissement de la distribution à deux lois . . . . .	79
3.6	Méthode de simulation de monte Carlo . . . . .	85
<b>Chapitre 4</b>	<b>Exemples d'utilisation de la modélisation</b>	<b>89</b>
4.1	Pricing probabiliste d'un contrat XS . . . . .	89
4.1.1	Exemple d'application . . . . .	90
4.2	Optimisation de la couverture de rétrocession XS sur un critère de mesure de risk . . . . .	91
4.2.1	Choix du nombre de réassureur . . . . .	92
4.2.2	Spécification de la problématique . . . . .	94
4.2.3	Résolution . . . . .	94
<b>Conclusion</b>		<b>97</b>
<b>Références</b>		<b>98</b>
<b>Annexes</b>		<b>99</b>

# Table des figures

1.1	illustration de l'activité de la réassurance . . . . .	8
1.2	Évolution des primes - Source données : Swiss Re et Apref-Marché Mondial . . . . .	9
1.3	Source SCR - Logo . . . . .	10
1.4	Source SCR-Organigramme de la SCR . . . . .	11
1.5	Source SCR-Affaire Marocaine . . . . .	12
1.6	Source SCR-Primes Activité Vie . . . . .	13
1.7	Source SCR-Primes Activité non vie . . . . .	13
2.1	Types de réassurance . . . . .	17
2.2	Illustration Quote-Part . . . . .	18
2.3	Simulation d'un portefeuille . . . . .	19
2.4	Illustration Excédent de plein . . . . .	19
2.5	Illustration XS . . . . .	20
2.6	Illustration XL . . . . .	23
3.1	Résumé sur le domaines d'attraction des lois usuelles . . . . .	30
3.2	Simulation du quantile plot généralisé . . . . .	34
3.3	Fonction de répartition empirique . . . . .	39
3.4	Représentation de la distribution sur un échantillon simulé d'une loi log-normale . . . . .	43
3.5	Exemple graphe de Cullen and Frey . . . . .	45
3.6	Exemple d'un QQ-PLOT . . . . .	46
3.7	Comparaison des pertes après revalorisation selon les marchés . . . . .	58
3.8	Comparaison des pertes . . . . .	59
3.9	Sinistres - Marché national . . . . .	60
3.10	Sinistres-Marché international . . . . .	61
3.11	Courbes-Comparaison du nombre de sinistres par an selon les marchés . . . . .	62
3.12	Histogramme de densité . . . . .	62
3.13	Quartile plot généralisé - Marché Marocain . . . . .	63
3.14	Quartile plot généralisé des coûts - Marché International . . . . .	64
3.15	Hill Plot Marché Marocain . . . . .	65
3.16	Pickands-Plot Marché Marocain . . . . .	65
3.17	Hill Plot Marché International . . . . .	66
3.18	Pickands-Plot Marché International . . . . .	66
3.19	Histogramme de densité et graphe cumulative de distribution_MA . . . . .	68
3.20	Graphe de Cullen et Frey avec bootstrap_MA . . . . .	69
3.21	QQ-Plot selon les lois usuelles . . . . .	70
3.22	PP-Plot selon les lois . . . . .	71
3.23	Comparaison entre la fonction de répartition empirique des données extrêmes et la fonction de répartition théorique de la loi log-normale . . . . .	72

3.24	Modélisation des sinistres supérieurs au seuil des extrêmes . . . . .	73
3.25	Densité et QQ-Plot d'une loi de Poisson (bleue) superposée à celle du jeu de données (noire) . . . . .	74
3.26	Densité et QQ-Plot d'une loi binomiale négative (rouge) superposée à celle du jeu de données (noire) . . . . .	75
3.27	Densité et QQ-Plot d'une loi de Poisson (bleue) superposée à celle du jeu de données (noire) . . . . .	76
3.28	Densité et QQ-Plot d'une loi binomiale négative (rouge) superposée à celle du jeu de données (noire) . . . . .	77
3.29	Densité et QQ-Plot d'une loi de Poisson (bleue) superposée à celle du jeu de données (noire) . . . . .	78
3.30	Densité et QQ-Plot d'une loi binomiale négative (rouge) superposée à celle du jeu de données (noire) . . . . .	79
3.31	Modélisation des sinistres . . . . .	82
3.32	Adéquation de alpha la loi normale et sensibilité de alpha par rapport à la loi globale(modèle) . . . . .	84
4.1	Quartile Plot généralisé - Sinistres réels Marché Marocain . . . . .	99
4.2	Hill Plot - Sinistres réels Marché Marocain . . . . .	100

## Liste des tableaux

3.1	Table de réévaluation par inflation . . . . .	28
3.2	Échantillon des pertes du lob incendie sur le marché marocain . . . . .	56
3.3	Échantillon des pertes du lob incendie sur le marché international . . . . .	57
3.4	Table de revalorisation par inflation . . . . .	57
3.5	Revalorisation des pertes sur le marché Marocain . . . . .	58
3.6	Statistiques descriptives de la sinistralité Incendie . . . . .	59
3.7	Statistiques descriptives - Marché marocain . . . . .	60
3.8	Échantillon-Nombre de sinistres par an . . . . .	61
3.9	Seuils . . . . .	67
3.10	coefficients d'asymétrie et de pointicité_MA . . . . .	68
3.11	Paramètres de la loi log-normale . . . . .	71
3.12	Paramètres Loi GPD . . . . .	72
3.13	Paramètres de la $\mathcal{BN}$ - Nombre de sinistres attritionnels . . . . .	75
3.14	Paramètres de la $\mathcal{BN}$ - Fréquence globale . . . . .	79
4.1	Tarification d'un XS . . . . .	91
4.2	Optimisation par dichotomie-En DH . . . . .	96

# Remerciements & Dédicaces

Tout d'abord, je tiens à remercier Monsieur YOUSSEFI Anasse de la Société Centrale de Réassurance(SCR) pour m'avoir proposé ce sujet ainsi que pour excellent encadrement tout au long de cette étude. Je remercie également toute l'équipe Entreprise Risque Management pour son accueil chaleureux et son aide pendant l'étude.

Je tiens ensuite à remercier mon encadrant interne Monsieur Fouad Marri pour sa disponibilité et son aide pendant l'étude.

---

Ce travail est dédié à :

Mes chers parents qui m'ont toujours soutenus sans condition.

Mes sœurs et mon frère, à qui je souhaite toute la réussite et le bonheur.

# Résumé

---

L'objectif de ce mémoire est de compléter l'outil de modélisation de la sinistralité afin d'améliorer la mise en place d'une couverture ou l'optimisation de cette ouverture selon certains critères.

Classiquement en réassurance, la sévérité de la sinistralité est modélisée uniquement par une seule loi de probabilité, généralement par une Log-normale ou une loi de Pareto généralisée. Cependant, cette modélisation n'est pas toujours parfaite. Ainsi nous proposons une modélisation par deux lois afin de pouvoir modéliser la partie attritionnelle (faibles montants et fréquents) séparément de la partie extrême (grands montants et rares). Cette modélisation repose sur la théorie des valeurs extrêmes et implique la détermination d'un seuil de raccordement. Ayant des données présentant les caractéristiques d'existence des valeurs extrêmes, nous avons fait appel à des outils mathématique de la théorie des valeurs extrêmes dont le graphique Hill afin de déterminer le seuil des extrêmes et séparer la partie extrême de l'attritionnelle. Cela étant fait, nous avons modélisation séparément par les outils de l'inférence classique la distribution de données attritionnelles et extrêmes. Pour sélectionner la loi la plus adaptée aux données, des comparaisons graphiques, des critères numériques et des tests statistiques sont utilisés. Après quoi, nous construisons une variable aléatoire regroupant les deux types de distribution de données.

Ayant modélisé nos sinistres, il devient alors possible de tarifer n'importe quel contrat en excédent de sinistres en passant par bien d'autres manipulations. Nous présentons également une optimisation de la rétrocession contrat XS de la SCR, qui permet de contrôler son appétit au risque à travers la value at risk.

---

**Mots clés :** *SCR (Société centrale de réassurance), Réassurance, rétrocession, modélisation, fréquence, sévérité, théorie des valeurs extrêmes, seuil des extrêmes, loi usuelle, Pareto Généralisée, GPD (Loi généralisée de Pareto), simulation de Monte Carlo, XS (Excédent de sinistre), Var (Value at risk)*

**Toutes les figures présentées dans ce rapport ont été élaborées par nos soins sauf mention contraire.**

# Abstrack

---

The objective of this report is to complete the claims modeling tool in order to improve the implementation of coverage or its optimization according to certain criteria.

Classically in reinsurance, the severity of claims is modeled solely by a single probability law, generally by a Log-normal or a generalized Pareto law. However, this modeling is not always perfect. Thus we propose a modeling by two laws in order to be able to model the attritional part (low amounts and frequent) separately from the extreme part (large amounts and rare). This modeling is based on the theory of extreme values and involves the determination of a connection threshold. Having data presenting the characteristics of existence of the extreme values, we appealed to mathematical tools of the theory of the extreme values of which the graph Hill in order to determine the threshold of the extremes and to separate the extreme part from the attritional. This being done, we modeled separately by classical inference tools the distribution of attritional and extreme data. To select the most suitable law for the data, graphical comparisons, numerical criteria and statistical tests are used. After which, we construct a random variable grouping the two types of data distribution.

Having modeled our claims, it then becomes possible to price any contract in excess of claims by going through many other manipulations. We also present an optimization of the retrocession contract XS of the SCR, which makes it possible to control its risk appetite through the value at risk.

---

**Key words :** *SCR(Central Reinsurance Company), Reinsurance, Retrocession, Modeling, Frequency, Severity, Extreme Value Theory, Extreme Threshold, Generalized Pareto, GPD (Generalized Pareto Law), Monte Carlo Simulation, XS (Excess of Loss, Value at risk (Var)*

*All figures presented in this report were produced by us unless otherwise stated.*

# Introduction

Activité remontant au moyen âge, L'assurance telle que nous la connaissons aujourd'hui, est un service basé sur un contrat par lequel l'assureur s'engage envers l'assuré, pour une période donnée, à le couvrir financièrement lors de la survenance d'un événement incertain (appelé sinistre) préalablement déterminé par le dit contrat, en contrepartie d'un montant appelé "prime d'assurance". Sur le marché des assurances, il existe deux types de sociétés majeures qui constituent le socle de l'activité, à savoir les compagnies d'assurance et les sociétés de réassurance. Dans les compagnies d'assurance, les assurés sont des personnes physiques, alors que les sociétés de réassurance couvrent les sociétés d'assurance.

Un des enjeux majeurs des compagnies d'assurance est la bonne évaluation des risques liés à leurs portefeuilles afin de garantir leur solvabilité ainsi que leur gain d'affaire. Pour atteindre leur but elle se tourne vers des sociétés de réassurance. Moyennement le paiement d'une prime elle cède sous certain condition une partie de leur risque à ces sociétés.

Cependant dans la même vision de couverture et d'homogénéisation du risque, un réassureur ne peut pas concentrer plusieurs risques dans son portefeuille, lui aussi a une certaine capacité de souscription, c'est pourquoi il rétrocède une part des risques à d'autres réassureurs. Ce processus est appelé la rétrocession. Tout cela dans l'optique de couvrir les portefeuilles avec une forte dispersion des risques.

En l'occurrence, la SCR (Société Centrale de Réassurance) nous propose dans le cadre de notre travail de fournir une modélisation de la sinistralité lui permettant d'améliorer les différents paramètres de la couverture de son portefeuille Non-Marine. Le but de notre démarche est de fournir un outil permettant de modéliser parfaitement la sinistralité afin d'établir la meilleure couverture de rétrocession pour le dit portefeuille.

Tout d'abord, nous présenteront de façon générale la réassurance (son marché, les types de contrats) ainsi que les principaux points de la Société Centrale de Réassurance. Puis, nous expliciterons les différents mécanismes de couverture de réassurance. Ensuite, nous présenterons notre cheminement pour trouver la loi des sinistres en fréquence et en coût, afin de pouvoir simuler des sinistres ayant les mêmes propriétés statistiques que ceux de la base de données de notre portefeuille. Et pour finir nous donneront des exemples d'utilisation de la modélisation notamment dans le cadre de la tarification d'une couverture ou de l'optimisation cette couverture sous certaines retenue de cette dernière.

# Chapitre 1

## Chapitre préliminaire

### 1.1 Présentation générale de la réassurance

Les premiers contrats de réassurance remontent au 14<sup>ème</sup> siècle lorsque les transferts de marchandises transcontinentaux s'effectuaient essentiellement par voie maritime. Ces contrats de couverture ont été motivés à l'origine par des risques liés aux attaques de pirates et/ou par des risques de naufrage assez conséquents. Ainsi en 1370 le premier contrat fut signé par un assureur vénitien exerçant dans le domaine du transport maritime qui a décidé de céder la partie la plus risquée du voyage : autour du Détroit de Gibraltar et lors de l'escale à Cadix en Espagne, à un autre assureur. Dès lors, les contrats de réassurance sur quelques cessions, essentiellement de la branche maritime, se sont multipliés en Italie, en France et au Royaume-Uni. La réassurance telle que nous la connaissons aujourd'hui apparaît en Allemagne vers la fin du 19<sup>ème</sup> siècle, au moment de la révolution industrielle.

Pour toute définition simple, la réassurance est l'assurance des assureurs. Elle est définie par P.Blanc comme étant un "contrat intervenant pour réaliser la compensation des écarts, soit par insuffisance du nombre de risques, soit par dépassement anormal des sinistres espérés". La problématique du cycle de production inversé amène les assureurs à être méticuleux sur la couverture des leur différents portefeuilles, la réassurance est le système le plus efficace pour garder sa compétitivité sur le marché et protéger son portefeuille en cas d'écarts importants entre la sinistralité prévue et la sinistralité réelle.

En effet, certaines sociétés d'assurance doivent se protéger contre des risques qui peuvent dépasser leur propre capacité, ou risques extrêmes tels que les catastrophes naturelles, qui pourraient déstabiliser leur bilan et leur compte de résultat, et les entraîner vers l'insolvabilité. C'est là qu'intervient la réassurance, qui, à travers divers contrats, permet à toute société d'assurance de pouvoir transférer une partie des risques souscrits aux termes d'une ou plusieurs polices de réassurance. La cédante paie alors une prime au réassureur, qui s'engage en contrepartie à indemniser une partie de sa sinistralité. Par le même principe, un réassureur peut se décharger d'une partie des risques qu'il assume, auprès d'un ou plusieurs réassureurs alors appelés rétrocessionnaires.

La réassurance a des avantages systémiques sur l'économie. Elle permet entre autres :

- Réduire la probabilité d'insolvabilité des assureurs.

- Assister les assureurs ‘a se lancer dans un nouveau business
- Augmenter ses souscriptions et donc sa croissance, en augmentant sa capacité(montant qu’il peut garantir pour une catégorie de sinistres)

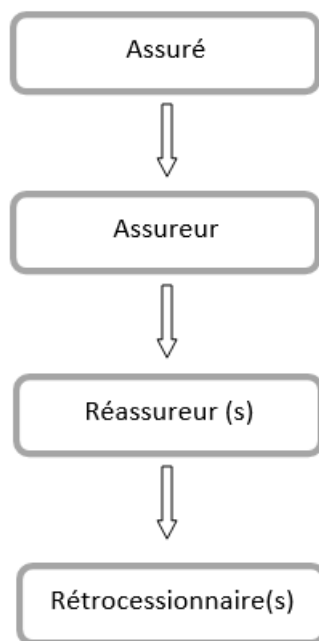


FIGURE 1.1 – illustration de l’activité de la réassurance

### 1.1.1 Les différentes formes de réassurance

On distingue principalement deux formes de réassurance. La **réassurance facultative** et la **réassurance obligatoire**. La **réassurance facultative** fait état des contrats dans lesquels la cédante peut choisir les risques qu’elle veut céder comme l’assureur peut choisir de les accepter ou pas. Cette forme de réassurance est utilisée lorsque le portefeuille est trop petit, ou lorsque la taille ou la nature du risque justifient une approche individuelle. La **réassurance obligatoire** qui est la forme la plus utilisée, fait état des contrats de réassurance qui s’effectuent sur un groupe de risques, ou sur la totalité d’un portefeuille, selon des modalités définies. Ces types de contrats sont appelés des **traités**. Ils imposent une obligation réciproque, la cédante ( Un assureur ou un réassureur) s’engage à céder tous les risques évoqués dans le traité et la réassurance se doit de les accepter.

En considérant l’aspect technique, on peut regrouper les contrats de réassurance en deux types, à savoir la **réassurance proportionnelle** et la **réassurance non proportionnelle**.

La **réassurance proportionnelle** : On parle de réassurance proportionnelle lorsque les sinistres payés par le réassureur sont proportionnels aux primes qu’il a reçues. Sous cette forme la cédante est très liée à son réassureur.

La **réassurance non proportionnelle** : Les traités non proportionnels sont définis par une priorité ou déductible (correspondant à une franchise) et une portée ou limite. Le réassureur assume tout

ou partie du sinistre qui excède la priorité et dans la limite de la portée. La cédante et le réassureur sont beaucoup moins liées puisque au cours d'une année d'exercice, la cédante peut être en perte, mais pas forcément le réassureur (ce qui aurait dû être le cas dans le cadre d'un contrat en quote-part).

### 1.1.2 Le marché de réassurance

Comme dans la quasi totalité des marchés, le marché de réassurance s'est vu impacté par la crise socio-économique que le monde a connue suite à la pandémie. Aussi la persistance des taux d'intérêt bas ainsi que la hausse de la sinistralité des catastrophes naturelles ont créé un climat d'incertitude lié à la montée de risques majeurs imprévisibles, complexes et interdépendants.

Le marché mondial de l'assurance enregistre 6 287 milliards USD de primes en 2020, réparties entre les branches vie (46%) et non vie (54%). Un peu plus de 5% des primes totales, soit environ 320 milliards USD, sont cédées en réassurance. L'activité de réassurance non vie représente 69% de ce montant alors que moins d'un tiers de l'aliment, soit 31%, concerne la branche vie.

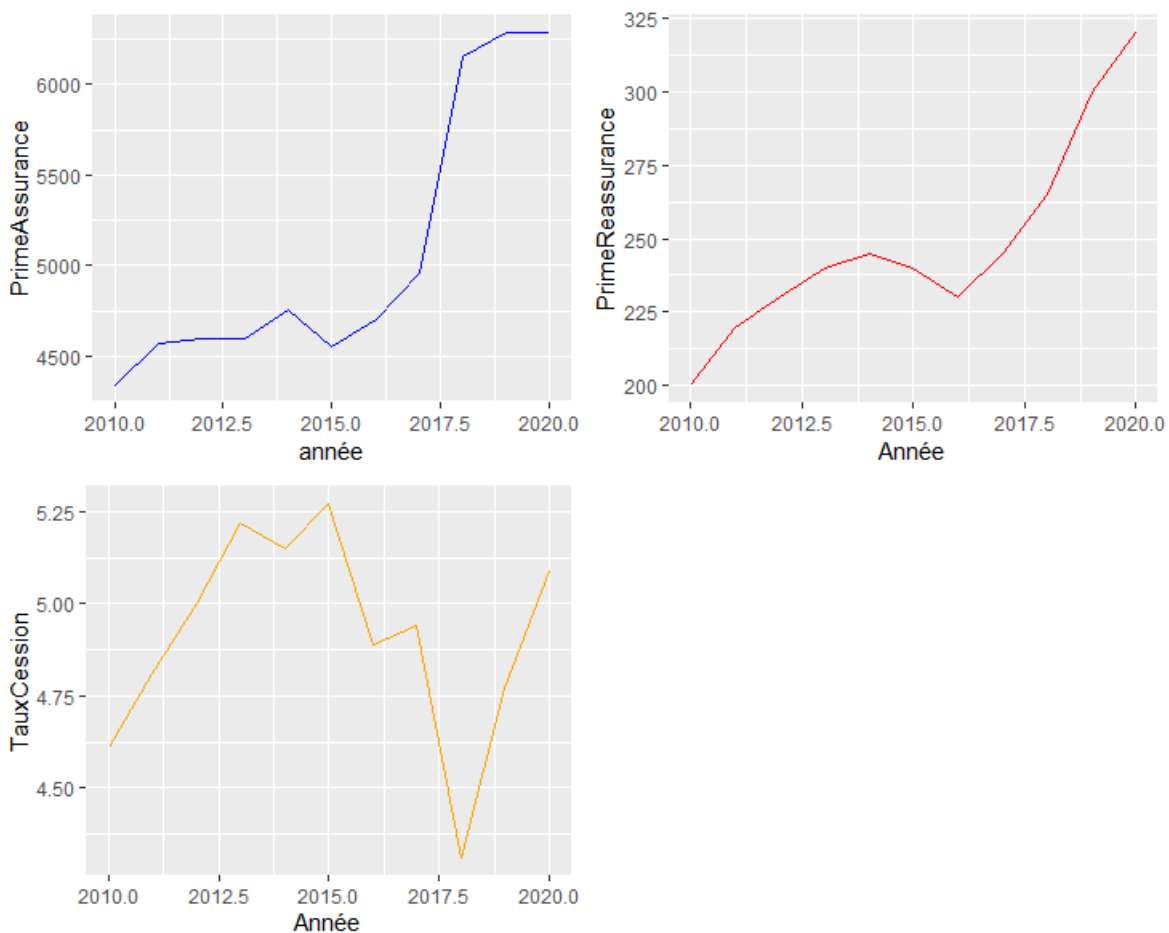


FIGURE 1.2 – Évolution des primes - Source données : Swiss Re et Apref-Marché Mondial

Que cela soit sur le marché des assurances ou de réassurances, les primes sont en hausses depuis

les années 2016. Le taux cession quant à lui n'a pas de tendance particulière mais nous observons une chute importante entre 2016-2018.

Les grands réassureurs comme unich Re, Swiss Re, Hannover Re, Scor et Lloyd's ont enregistré une hausse des primes en fin 2021 et affichent de bons résultats semestriels, après avoir provisionné la sinistralité Covid-19 en 2020.

Confronté à un regain de pertes liées aux catastrophes naturelles, à une réévaluation des périls secondaires, à une sinistralité Covid-19 et à un faible rendement des placements, les réassureurs n'ont pas d'autres choix que d'augmenter leurs tarifs. De ce fait, la discipline tarifaire imposée lors du renouvellement de janvier 2021, avec des taux en augmentation de 5 à 8%, devrait être reconduite en 2022. Lors du précédent renouvellement (janvier 2021), les principaux réassureurs ont imposé des améliorations de primes et de conditions pour l'ensemble des branches et zones géographiques. A titre d'exemple, Scor a enregistré une hausse tarifaire moyenne de 7,8%, Swiss Re 6,5% et Hannover Re 5,5%. Toujours en 2021, les traités des branches dommages et responsabilités, affectés par une forte sinistralité catastrophe naturelle ou industrielle ont vu leurs taux augmenter de deux chiffres. En revanche, la hausse s'est située dans une fourchette comprise entre 3% et 5% pour les traités n'ayant pas subi de pertes.

Beaucoup d'acteurs sur le marché se préoccupe de la montée de la sinistralité suite à l'évolution des problèmes systémiques générales comme le réchauffement climatique, trouvant ainsi le niveau de tarification bas.

## 1.2 Société centrale de réassurance



FIGURE 1.3 – Source SCR - Logo

Avant toute chose, il est important d'avoir une vision générale sur la société d'étude ainsi que ces principales activités notamment dans le cadre de la rétrocession.

## 1.2.1 Aperçu général

La SCR agit en tant que réassureur généraliste sur le marché domestique et à l'international. Elle propose divers catégories de couverture en réassurance vie et non vie soit donc des traités et des facultatives proportionnels et non proportionnels. La Société Centrale de Réassurance est l'œuvre d'une convention signée entre l'état marocain et la Caisse de Dépôt et de Gestion le 09 Mars 1960, modifié et renouvelée le 28 Novembre 2000. C'est une société anonyme détenue majoritairement par la CDG avec un chiffre d'affaires de 2,61 Milliards de MAD en 2021. Elle est considérée comme étant le premier acteur du marché de réassurance marocain avec une part de 70%, assurant ses missions de régulation et sécurisation face aux turbulences internationales en participant à la conservation des primes au niveau national et à la mobilisation de l'épargne dans l'économie du pays.

Dans l'objectif de renforcer l'assise financière de la SCR sur le marché national et international, une assemblée générale extraordinaire des actionnaires s'est réunie le 24 juin 2021 et a décidé d'augmenter le capital social actuel de la société en numéraire, d'un montant de 700 Millions de dirhams. Le capital social de la société est désormais porté à 2,2 Mds de dirhams.

A l'international, elle se démarque par la qualité de sa gestion ainsi que par sa performance globale et a reçu le prix "Meilleur Réassureur en Afrique pour l'année 2021, la plus haute distinction remise annuellement, par *l'International Business Magazine – catégorie Finance*. La SCR est notée par des agences de notations internationales tel que l'agence AM Best (B++(Good)) et Fitch Ratings (AAA(Prime)).

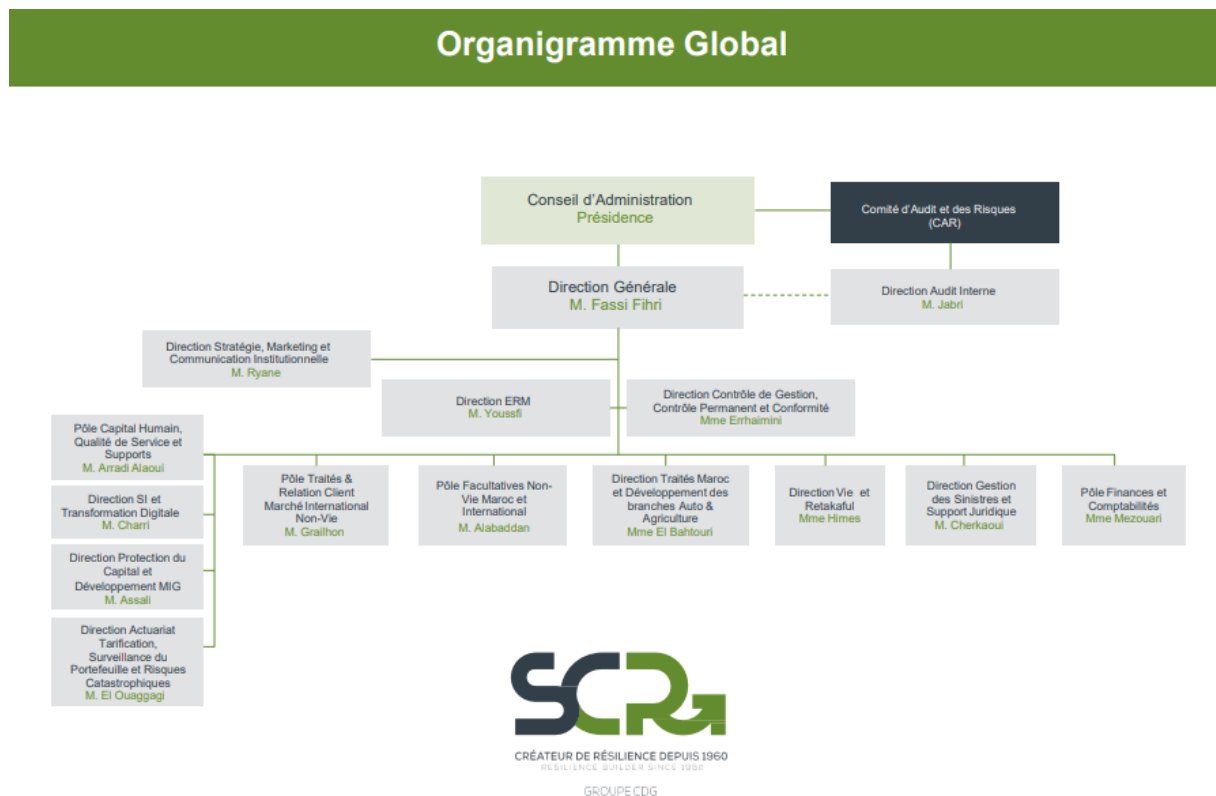


FIGURE 1.4 – Source SCR-Organigramme de la SCR

## ERM : Entreprise Risk Management :

ERM est un dispositif de la SCR dont l'objectif principal est de gérer l'ensemble des risques individuellement et conjointement via le processus d'évaluation interne des risques et de la solvabilité. Ce pôle travaille aussi dans le cadre de certaines missions avec d'autres équipes, en l'occurrence, afin d'optimiser le programme de rétrocession et d'étudier l'adéquation de la structure actuelle aux besoins de la SCR, des travaux de modélisation ont été conjointement menés avec les équipes ERM, rétrocession et les courtiers.

### 1.2.2 Portefeuille client

Le chiffre d'affaires de la SCR toutes affaires confondues s'élève à 2,61 Milliards de MAD. Le chiffre d'affaires sur le marché marocain ressort de 1,97 Milliards de MAD, en baisse de -2,9% par rapport à l'exercice précédent, du fait d'un volume exceptionnel de primes sur les affaires facultatives en 2020.

	<i>En Millions de MAD</i>		<i>2020/2021</i>	
	<b>2020</b>	<b>2021</b>	<b>En MDH</b>	<b>En %</b>
<b>Affaires légales</b>	50,4	76,3	25,9	-51,40%
<b>Affaires Marocaines Traités</b>	710	722,2	12,2	1,70%
<b>Affaires Marocaines Facultatives</b>	1267,2	1170,1	-97,1	-7,70%
<b>TOTAL</b>	<b>2027,6</b>	<b>1968,6</b>	<b>-59,1</b>	<b>-2,90%</b>

FIGURE 1.5 – Source SCR-Affaire Marocaine

Sur le marché international, le chiffre d'affaires s'élève à 646,3 Millions de MAD en 2021, à comparer à 605,8 Millions de MAD en 2020, soit une hausse de +6,7% essentiellement due aux affaires traitées.

### 1.2.3 Analyse historique par catégorie

#### Activité vie

Primes émises	En Millions de MAD			2020/2021	
	2019	2020	2021	En MDH	En %
<b>Cession Légale</b>	<b>28,1</b>	<b>51,2</b>	<b>76,6</b>	<b>25,4</b>	<b>49,70%</b>
<i>Capitalisation</i>	19,7	35,3	49,3	13,9	39,50%
<i>Retraite</i>	8,1	15,1	26,8	11,8	78,10%
<i>Décès</i>	0,3	0,8	0,5	-0,3	-33,70%
<b>Affaires Marocaines</b>	<b>44,8</b>	<b>32</b>	<b>60,3</b>	<b>28,3</b>	<b>88,50%</b>
<i>Traités</i>	35,1	20,4	29,7	9,2	45,20%
<i>Facultatives</i>	9,6	11,6	30,6	19,1	165%
<b>Affaires Etrangères</b>	<b>41,1</b>	<b>36,9</b>	<b>49,6</b>	<b>12,8</b>	<b>34,60%</b>
<i>Traités</i>	29,8	18,5	45,4	26,8	144,80%
<i>Facultatives</i>	11,3	18,3	4,3	-14,1	-76,70%
<b>TOTAL</b>	<b>114</b>	<b>120,1</b>	<b>186,5</b>		
<b>Evolution en MDH</b>		6,1	66,5		
<b>Evolution en %</b>		5,40%	55,40%		

FIGURE 1.6 – Source SCR-Primes Activité Vie

le chiffre d'affaires de l'activité vie de l'année 2021 a observé une hausse de +55,4% comparativement à l'année 2020. Ceci est essentiellement du au volume de primes des affaires marocaines et de la cession légale.

#### Activité non vie

Primes émises	En Millions de MAD			2020/2021	
	2019	2020	2021	En MDH	En %
<b>Cession légale non vie</b>	<b>1</b>	<b>-0,8</b>	<b>-0,3</b>	<b>0,5</b>	<b>-56,50%</b>
<b>Traités Maroc</b>	591,7	689,6	703,2	13,6	2,00%
<b>Fac.Maroc</b>	858,2	1255,6	1128,8	-126,9	-10,10%
<b>Traités Etranger</b>	368,9	455,9	465,9	10	2,20%
<b>Fac. Etranger</b>	83,5	113,1	130,8	17,7	15,70%
<b>TOTAL</b>	<b>1903,3</b>	<b>2513,4</b>	<b>2428,4</b>		
<b>Evolution en MDH</b>		610,1	-85,5		
<b>Evolution en %</b>		32,10%	-3,40%		

FIGURE 1.7 – Source SCR-Primes Activité non vie

Le chiffre d'affaires de l'activité non vie s'établit à 2,43 Milliards de MAD, une baisse de -3,4% par rapport à l'année 2020. Cette baisse s'explique par une baisse des facultatives marocaines.

Aussi le chiffres d'affaires de l'activité non vie est constitué de 100% en affaires conventionnelles, avec 75% sur le marché marocain et 25% sur le marché international.

#### **1.2.4 Programme de Rétrocession**

La rétrocession est l'opération par laquelle un réassureur se décharge sur un ou plusieurs autres réassureurs de la totalité ou d'une fraction des risques qu'il a acceptés de couvrir. Dans le modèle économique de la SCR, la rétrocession est le principal mécanisme de transfert de risque. Elle vise la mise à disposition de capacités de souscription et la protection de ses capitaux propres.

##### **La politique de rétrocession**

La politique de rétrocession a pour principaux objectifs la réduction de l'exposition aux risques de souscription, une diminution de la volatilité des résultats financiers ainsi qu'une stabilisation de la solvabilité afin de contribuer au développement de l'entreprise. Les principes fondamentaux à la SCR de cette activité sont les suivants :

- La mise à disposition de capacités de souscription.
- Une utilisation stratégique en tant que levier du développement à l'international à travers des courtiers internationaux et des es compagnies de réassurance dont la SCR détient une partie du capital.
- Une protection des capitaux propres et de la solvabilité de la SCR.
- La gestion des risques des actifs sous-jacents à la rétrocession, entre autres les risques de contrepartie, les risques juridiques et les risques opérationnels.

La SCR dispose **d'un programme de rétrocession**. Des travaux de modélisation sont conjointement réalisés par les équipe ERM, rétrocession, et les courtier international afin de garantir de l'adéquation de celle-ci avec la structure actuelle des besoins de la SCR.

##### **Processus de rétrocession**

La rétrocession est un processus pleinement intégré dans les activités de la SCR. Ce dernier met en contribution l'ensemble des parties prenantes stratégiques en l'occurrence la direction de protection du capital, la direction juridique et de gestion des sinistres, le pôle financier, le pôle de souscription ainsi que l'exécutif et administration. Les étapes du processus sont les suivants.

##### **Élaboration de la stratégie de rétrocession**

Afin de mettre en place une stratégie de rétrocession, la SCR se base entre autres sur une analyse des besoins actuels exprimés par le pôle de souscription(Capacités, garanties, chiffre d'affaires prévisionnel...) ainsi que sur une étude quantitative du portefeuille existant. Les analyses quantitatives réalisées consistent en une étude des profils de risques et leur adéquation avec la structure actuelle, en un examen de la sinistralité expérimentée et en une modélisation des risques les plus matériels du portefeuille de la SCR. Ces analyses se font sur la base des données extraites du système préalablement validées par les équipes qui en sont responsables.

Le plan de rétrocession prévisionnel(structure et placement) ainsi mis en place est par la suite soumis pour validation à l'ensemble des instances de gouvernance(comité Risques et capital,

comité d'audit et des risques, conseil d'administration). Ce dernier est communiqué à l'ACAPS conformément à l'article 80 de la circulaire numéro 01/AS/19

### **Consultation du marché de la réassurance et placement**

La mise en œuvre de la stratégie de récession adoptée est réalisée selon le processus qui suit :

- Constitution du dossier de renouvellement et initialisation de la consultation.
- On soumet soit de façon directe, soit de façon indirecte à travers les courtiers, les besoins de récession au marché international de réassurance.
- Les dossiers de renouvellement à la base de la consultation incluent notamment les cahiers des charges des couvertures (options de couverture souhaitées, caractéristiques des portefeuilles) et les informations quantitatives sous-jacentes (Profils de risques, triangle de développement, sinistres majeurs, principaux risques, résultats de la modélisation).
- Analyse des offres et sélection des leaders des couvertures à l'aide de critères objectifs matérialisés dans un outil de scoring dédié.
- Les principaux critères sur lesquels s'appuient la SCR sont : l'adéquation de la structure cotée et les paramètres de couverture associés, son coût, les garanties, territoires, clauses contractuelles, exclusions potentielles, la qualité de la sécurité, solidité financière, expertise, gestion de sinistres, cadre juridique, ...
- L'outil de scoring aboutit donc à un classement des cotations et la SCR choisie l'offre ayant obtenue la meilleure note.
- Confirmation de la sélection et établissement des contrats.
- Le contrat réalisé fait l'objet d'une revue par la direction juridique de la SCR.
- Placement des couvertures.

### **Critère de choix des réassureurs**

Dans le cadre du programme de la récession, la SCR mutualise ses cessions en affaires facultatives auprès d'un panel de récessionnaires étrangers, après avoir alimenté en amont, sa propre rétention. La SCR peut donc se retrouver en cas de défaillance d'un récessionnaire en perte au titre des sinistres dus par ce récessionnaire sur lesdites cessions.

Afin de gérer ce risque de défaut, des organismes d'évaluation applique des notations générale aux récessionnaires sur la base d'un certain nombre de critères. Les plus connus sont **AM Best, Standard & Poor's/Fitch Ratings**. Nous dénombrons certains de ces critères entre autres :

- Chiffre d'affaires
- Assiète des primes
- Performances financières
- Célérité dans le paiement des sinistres.
- Résultat annuel et excellents indicateurs financiers.
- Conformité des engagements juridiques dans la protection des clausiers
- Absence de litiges

- Acceptabilité de mise à disposition de la SCR d'une attestation de réserve couvrant les créances admises en représentation des provisions techniques.
- Politique de souscription

# Chapitre 2

## Les contrats classiques de réassurance

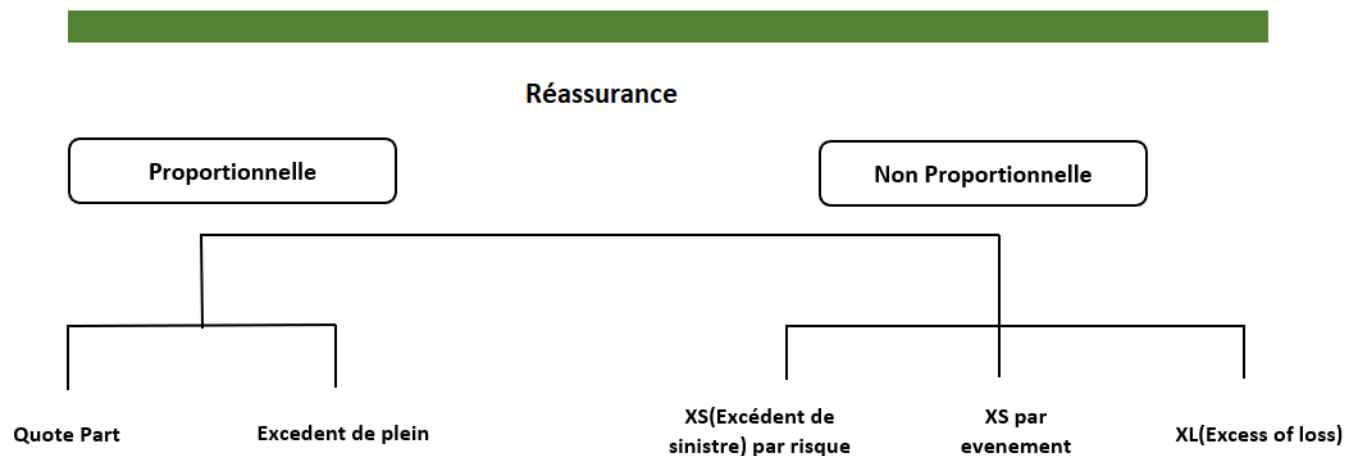


FIGURE 2.1 – Types de réassurance

### 2.1 Réassurance Proportionnelle

#### 2.1.1 Quote-Part

Le Quote-Part est la forme la plus simple de réassurance. Ce contrat consiste simplement à partager une proportion équivalente des primes et des sinistres du portefeuille de la cédante. De plus, ce mécanisme de couverture de réassurance n'est pas spécifique à une branche de l'assurance, on peut l'appliquer sur toutes les branches. Toutefois, cette méthode réduit faiblement la volatilité du portefeuille, car peu importe le montant du sinistre, l'assureur prendra une partie en charge. En effet la cession est fixe quelque soit le risque, l'assureur cède de la même façon sur les risques importants que sur les risques faibles, Le profil du portefeuille après cession de l'assureur ne change pas, mais c'est plutôt le niveau des engagements qui se modifie.

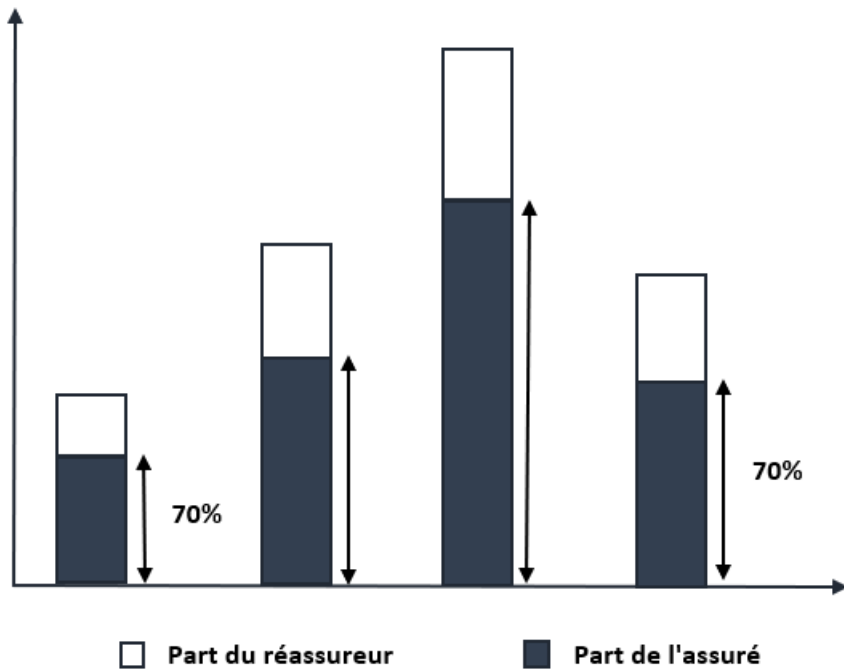


FIGURE 2.2 – Illustration Quote-Part

### 2.1.2 Excédent de plein (EP)

L'excédent de plein s'inscrit dans la même logique de cession proportionnelle. Cela dit le taux de cession est calculé police par police sur la base de la somme assurée(SI) et non des sinistres. En plus le réassureur n'intervient que si la somme à assurer dépasse un certain montant R appelé le plein de rétention. Cette méthode est en effet un peu plus bénéfique, car elle réduit les risques pris par ce dernier. En effet, l'assureur sait d'avance quel montant maximal il doit payer en cas de survenance d'un sinistre. Cependant, comme pour la quote-part, l'assureur doit toujours faire face au risque de cumulations de sinistres (nombre important de sinistres par année).

$$\alpha = \min \left\{ \max \left( 0, 1 - \frac{R}{SI} \right); \frac{nR}{SI} \right\}$$

Avec :

- $R$  : Plein de rétentions
- $SI$  : Sommes assurés
- $n$  : Nombre de pleins

Pour une simple illustration, considérons un portefeuille théorique et appliquons un excédent de plein de 6 millions de rétention. Les montants ci-après sont tous exprimés en millions de dirham.

Risques	Capital assuré	Prime d'assurance	Sinistres
1	14	0,25	8
2	5	0,12	4
3	25	0,5	17
4	8	0,2	3
5	10	0,25	7

Risques	Alpha	Sinistres cédés	Sinistres conservés
1	0,42857143	3,42857143	4,57142857
2	0	0	4
3	0,24	4,08	12,92
4	0,25	0,75	2,25
5	0,4	2,8	4,2

FIGURE 2.3 – Simulation d'un portefeuille

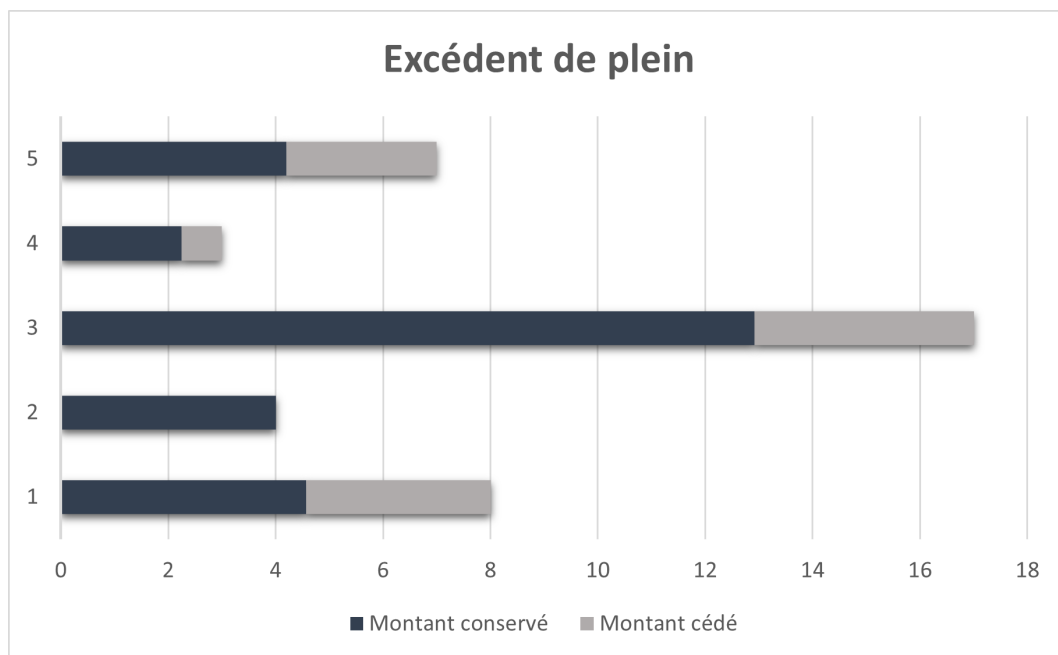


FIGURE 2.4 – Illustration Excédent de plein

L'excédent de plein permet à l'assureur d'augmenter significativement sa capacité de souscription. Aussi l'EP permet la libre gestion dans les risques inférieurs à son plein de rétention. Cependant, dans ce type de couverture, l'assureur est toujours exposé aux sinistres très importants.

## 2.2 Réassurance non proportionnelle

### 2.2.1 Excédent de sinistres

La réassurance en excédent de sinistre(XS) est un contrat de réassurance non proportionnelle dans lequel le réassureur indemnise la compagnie cédante pour les sinistres qui dépassent une limite prédéterminée. Cette limite est appelée **Priorité** notée  $P$ , mais le réassureur ne dépasse pas un certain seuil, une charge maximale, qu'on appelle portée, notée  $C$ . La somme de la priorité et de la portée correspond à une valeur appelée **plafond**. Dans le cas où la charge totale du sinistre dépasse le montant du plafond, le réassureur ne va payer que le montant de la portée  $C$ , la notation mathématique du XS est donnée comme suit :  $C \ XS \ P$

Soit  $X$  la charge de sinistre, le montant à la charge du réassureur( $X^{Reass}$ ) sera :

$$\begin{cases} 0 & \text{Si } X \leq \text{Priorité} \\ X - \text{Priorité} & \text{Si } \text{Priorité} \leq X \leq \text{Plafond} \\ \text{Portée} & \text{Si } X \geq \text{Plafond} \end{cases}$$

Il peut aussi s'écrire sous la forme suivante :  $X^{Reass} = \min[\max(X - P), C] = (X - P)^+$

Par exemple,  $15M \ XS \ 20M$  correspond à un traité en excédent de sinistre de portée  $15M$  et de priorité  $20M$ .

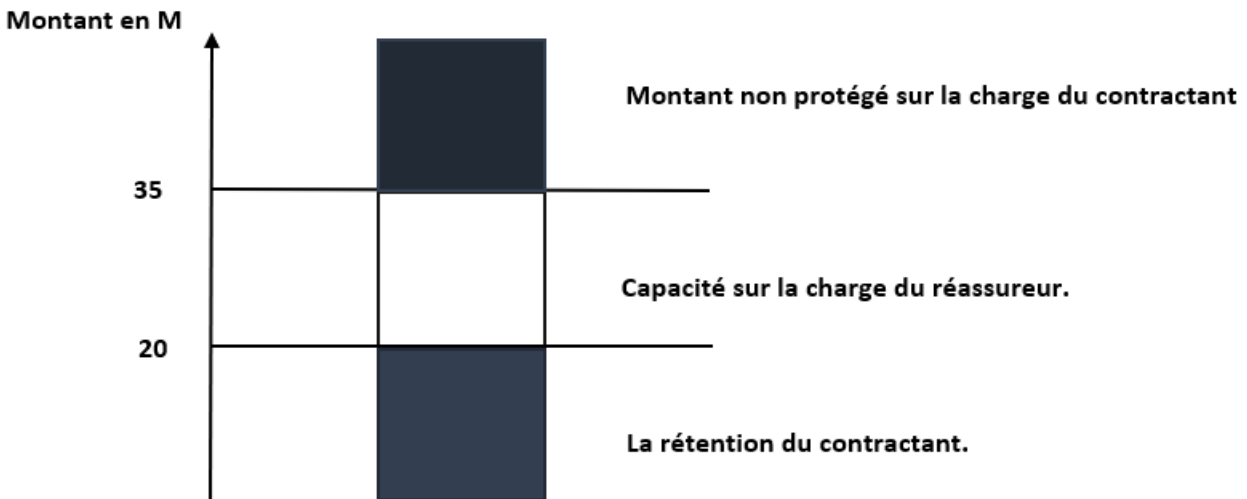


FIGURE 2.5 – Illustration XS

De façon générale, on peut ajouter sur la couverture de traité XS, des clauses supplémentaires de limitation. Ces clauses ont généralement pour but de réduire le coût de la couverture de réassurance. Les plus courantes sont les suivantes :

- Une clause de limite annuelle notée AAL (Annual Aggregate Limit) : dans l'année, le réassureur ne paiera jamais plus que ce montant.

- Des clauses de reconstitutions de capacité, gratuites ou payantes : lorsque la portée "C" est consommée, le réassureur mettra à disposition une autre portée (gratuite ou payante selon des conditions prédéfinies au traité) et ainsi de suite selon le nombre de reconstitutions définies au traités. Il peut arriver que le nombre de reconstitutions soit illimité.
- Une clause de franchise annuelle notée AAD (Annual Aggregate Deductible) : le réassureur ne commencera à prendre en charge les sinistres que lorsque la charge annuelle à la tranche dépassera ce montant.

Selon la nature de la couverture, on distingue deux types d'XS : XS par risque et XS par événement. De façon générale, on définit le risque comme étant la probabilité que survienne un dommage. Le sinistre est la matérialisation de ce risque par exemple un accident de route. Un risque peut être :

- Isolé : une cause & un risque, on parle alors de sinistre.
- Cumulatif : une cause & plusieurs risques, on parle d'événement.
- Agrégatif : plusieurs causes & plusieurs sinistres, on dit qu'on parle d'agrégat.

Ainsi,

### **XS par risque**

L'XS par risque fournit une couverture de réassurance à chaque fois qu'un risque est sinistré.

### **XS par événement**

L'XS par événement fait référence et s'applique aux dommages de plusieurs risques touchés du fait d'une même cause par exemple une catastrophe naturelle, un accident industriel. La base du traité établi, on peut augmenter certaines clauses dans le but de contrôler certains paramètres afin de limiter la charge du réassureur provenant d'un même événement. Nous pouvons par exemple citer des clauses comme :

- Clause de limitation géographique : la prise en charge sera constituée de plusieurs sinistres issus du même événement et dans un périmètre géographique limité.
- Clause de limitation dans le temps : le traité comporte une durée maximale pendant laquelle les sinistres imputables à une même cause sont pris en compte. Cette période peut généralement varier de 72 à 168 heures en fonction des événements couverts (inondations, accident industriel chimique, etc.).

L'XS par événement est généralement pris en complément d'un XS par risque lorsque l'assureur veut être sûr de ne payer qu'une fois la priorité lorsqu'un événement touche au moins deux ou trois polices de son portefeuille.

### **Primes de reconstitution**

Lorsque plusieurs sinistres surviennent dans la même année et absorbe toute la portée prévu par le contrat XS, La société cédante se retrouve à découvert. Pour éviter cela, les réassureurs ont mis en place la reconstitution de garantie, c'est-à-dire qu'après la survenance d'un sinistre, l'assureur va payer un certain montant pour bénéficier à nouveau de la protection totale du contrat XS qui lui était proposé à l'origine. Ainsi, la clause de reconstitution sert à repayer la portée après sinistre, c'est-à-dire à maintenir le même niveau de protection pour la société cédante. Une fois la portée

partiellement ou totalement absorbée par la réclamation, la clause de reconstitution de la garantie s'applique immédiatement. Lorsque le réassureur accorde N fois la reconstitution de la garantie, il s'engage à payer au maximum N+1 fois la portée durant la période de référence du traité d'indemnisation : garantie initiale + les reconstitutions. Le nombre de reconstructions peut être illimité.

Donc un contrat XS s'accompagne généralement du nombre N de reconstitutions accordées de la prime de reconstitution (en général égale à 1% de la portée), et la clause de reconstitution (généralement comprise entre 0 et 50% en prorata capital, et entre 0 et 100% en double prorata)

Supposons par un exemple C XS P.

Prime de reconstitution :  $P$

Clause de reconstitution :  $r\%$

Sinistre :  $S$

Survenance :  $i^{me}$  mois de l'année

Charge du réassureur :  $S^{Reass}$

— Prorata capital

$$S^{Reass} = \min[\max(S - P, 0), C]$$

$$\text{Coût}_{Reconstitution} = P \cdot \frac{r}{100} \cdot \frac{S^{Reass}}{C}$$

— Double prorata

$$S^{Reass} = \min[\max(S - P, 0), C]$$

$$\text{Coût}_{Reconstitution} = P \cdot \frac{r}{100} \cdot \frac{S^{Reass}}{C} \cdot \frac{12 - i}{12}$$

## 2.2.2 Excédent de perte

Aussi appelé **Stop Loss XL**, le XL, donc l'excédent de perte est un traité de couverture de réassurance basé sur quasi le même principe que le XS. Il concerne la protection du ratio de perte du contractant. On calcule donc la charge annuelle globale des sinistres en question . Avec cette charge, on calcule le ratio S/P  $\left(\frac{Sinistres}{Primes}\right)$ . Ce traité est noté mathématiquement par C XS P. Le réassureur intervient lorsque la sinistralité annuelle dépasse la priorité (P), L'intervention du réassureur est limitée à la capacité(C). La priorité P et la portée C sont en général exprimés en pourcentage des primes acquises reçues.

Pour une simple illustration, nous considérons un Stop Loss de 80% des primes acquises allant jusqu'à une perte allant à 30% que nous noterons 30% XS 80%. Soit  $(X_i)$  avec  $i$  allant de  $1, \dots, n$ ,

les sinistres de la cédante durant la période de couverture et nous notons par  $\pi$  le total des primes reçues.  $X^{Reas}$  est le montant à la charge du réassureur.

$$X^{Reass} = \min \left( \max \left( \sum_{i=1}^n X_i - P\pi; 0 \right); P\pi \right)$$

Nous avons pris pour exemple le jeu de données suivant auquel nous appliquons le traité :

Année	1	2	3	4	5	6	7	8
S/P	95%	110%	87%	70%	130%	120%	100%	150%
Conservé	80%	80%	80%	70%	80%	80%	80%	80%
Réassuré	15%	30%	7%	0%	30%	30%	20%	30%
Conservé trai	0%	0%	0%	0%	20%	10%	0%	40%

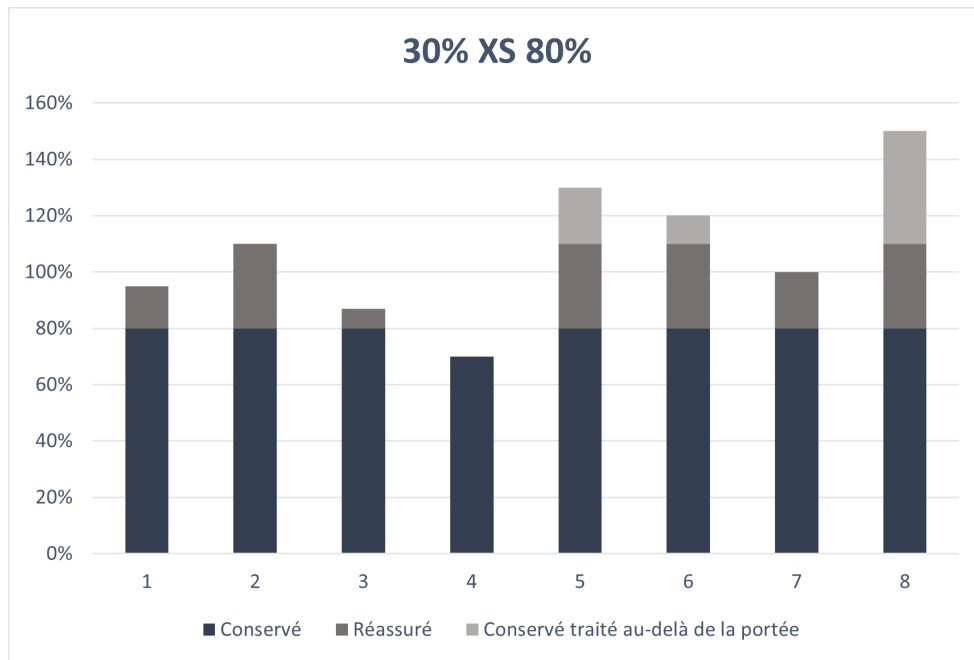


FIGURE 2.6 – Illustration XL

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.

### 2.2.3 Les méthodes de Modélisation du contrat XS

La tarification d'un contrat en assurance comme en réassurance est l'exercice qui permet de déterminer le tarif d'un contrat de couverture contre un risque. Une bonne tarification est indispensable car elle permet entre autres au contractant du risque de respecter ses engagements. On appelle prime pure le montant du sinistre moyen auquel devra faire face l'assureur ou le réassureur pour le risque. En réassurance on distingue principalement trois types de méthodes de tarification.

- La méthode du Burning Cost
- La méthode probabiliste
- La méthode par exposition

#### Burning Cost

Le Burning Cost est une méthode basée sur la sinistralité observée. On estime de façon empirique le prix du contrat en réassurance. Dans un premier temps, on actualise les données (prise en compte de l'inflation, évolution des paramètres liés au risque). On obtient alors une série de données "as if". Ensuite, on calcule la moyenne pondérée des ratios  $\frac{S}{P}$  revalorisés (le Burning Cost appelé le taux de flambage en français) et on en déduit la prime pure en multipliant le Burning Cost à l'assiette de prime estimée pour l'année considérée.

$$BC = \frac{\sum \text{Sinistres observés dans la tranche}}{\text{Total des primes perçues}}$$

Le BC est l'un des moyens les plus simples de tarification. Cela dit cette méthode présente des nombreux désavantages. Elle ne donne des résultats acceptables que si l'on se situe dans une tranche totalement travaillante, c'est à dire une tranche traversée totalement par la sinistralité. Aussi elle est perturbée lorsqu'on se retrouve avec un sinistre exceptionnel. Impossibilité de prendre en considération les clauses spéciales. Par ailleurs, il est impossible de tarifier une tranche qui n'a jamais été touchée.

#### Méthode probabiliste

La méthode probabiliste est une méthode basée sur les lois de probabilités. On détermine le tarif par l'approche fréquence  $\times$  sévérité. Elle consiste à modéliser séparément le coût des sinistres ( $X_i$ ) et leur fréquence  $N$ . Sous l'hypothèse que les  $X_i$  sont indépendants et identiquement distribués de même loi qu'une variable aléatoire  $X$ , et que les  $X_i$  et  $N$  sont des variables aléatoires indépendantes, on obtient le résultat fondamental de ce modèle qui est la formule de calcul de prime pure :

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

Il s'agit d'une approche paramétrique qui utilise tous les sinistres, elle résout le problème de tarification pour les tranches où l'on ne dispose pas d'assez de données. Elle permet aussi de prendre en considération les clauses particulières. Les principaux risques de cette méthode sont des risques d'erreur mathématique notamment les erreurs d'estimations ou le risque de modèle lié au choix des distributions.

En méthode probabiliste, Il est en général assez difficile de calculer analytiquement la prime pure. Aujourd'hui vue les capacités informatiques, on a principalement recours à des simulations. Une fois qu'on a modélisé la fréquence ainsi que les coûts, on procède de la façon suivante

- Nous générons un nombre de sinistres pour l'année.
- Nous simulons les montants de sinistres correspondants.
- Nous calculons la charge XS selon les caractéristiques du contrat.
- Et enfin on calcule la moyenne de la charge XS simulée, ce qui nous donne la prime pure.

Bien que présentant l'avantage de tarifer n'importe quel programme, la qualité de cette méthode en l'occurrence la qualité de l'estimation des paramètres est fortement liée à la quantité de données disponibles. Dans certains cas, on se retrouve dans l'incapacité d'obtenir des résultats pertinents notamment par exemple dans le cas des cédantes qui débutent sans statistiques sur leur portefeuille ou des cédantes qui ont subi un changement radical dans la structure de leur portefeuille. Dans ce cas, on fait généralement appel à la méthode de tarification par exposition.

### **Tarification par exposition**

Pour résoudre le problème des nouvelles cédantes ou des cédantes ayant changé radicalement de profil de risque, on utilise la tarification par exposition. Le principe est d'utiliser un portefeuille de composition similaire et pour lequel on dispose de données suffisantes. Elle ne se réfère pas à l'expérience passée de la cédante mais à son profil de risque actuel, Il s'agit d'une forme de cotation de marché, le but est de trouver le portefeuille le plus semblable possible par rapport au champ d'activité de la cédante en question. On va essayer d'estimer le tarif XS sur la base du portefeuille qui les génère (détail des sommes assurées, nombre de polices par tranches des capitaux...). La fonction d'exposition représente la probabilité pour qu'un sinistre coûte moins de  $x\%$  de la somme assurée. La prime pure s'exprime ainsi :

$$P_{\text{pure}} = F_{\text{exposition}} \left( \frac{\text{priorité de l'XS}}{\text{capitaux exposés}} \right) \times \text{Prime directe nette de chargement}$$

Pour la suite du mémoire, nous nous concentrerons sur la méthode probabiliste.

# Chapitre 3

## Modélisation de la sinistralité

Pour trouver la meilleure couverture de rétrocession ou pour tarifier un contrat de couverture ou même d'établir des mesures de risque comme la value at risk sur la charge de sinistres conservées, il faut d'abord déterminer la distribution en fréquence et en cout de notre jeu de donnée. Bien évidemment, l'objectif est de simuler plusieurs échantillons suivant les mêmes lois de probabilité que les données fournies.

Il est en effet crucial de bien mener la modélisation, encore plus dans le cadre de la rétrocession. Une mauvaise estimation des paramètres caractérisant les distributions des lois conduit à une sous-estimation ou une surestimation du risque associé au portefeuille. Cela est préjudiciable tant pour le réassureur que pour le rétrocessionnaire menant à des contrats de couvertures non adéquat. Ainsi l'actuaire doit être extrêmement prudent sur les caractéristiques des données en sa position ainsi que sur les méthodes mathématiques utilisées.

Par ailleurs, dans le cadre de la réassurance, il serait assez imprudent de fournir une modélisation à ligne complète(Globale) de la sinistralité. En effet, très souvent, dans les portefeuilles en excédent de sinistre, on observe en général deux paries, une partie du portefeuille fréquemment sinistrée avec des montants faibles, et une autre partie caractérisée par des très gros sinistres avec une faible probabilité d'occurrence. Ces types de portefeuilles ont généralement besoins de couverture. Dans ces portefeuilles, les sinistres extrêmes, bien qu'ayant une fréquence faible, influencent particulièrement la charge globale. Ces types de distribution de données sont appelés des distributions à queue lourde. Cela nous conduit à opter pour une méthode de modélisation consistant à séparer les sinistres attritionnels des sinistres extrêmes et d'étudier leurs lois séparément. Les sinistres attritionnels étant caractérisés par des montants relativement faibles et une forte probabilité d'occurrence. nonobstant, il existe des lois usuelles comme la loi de Pareto permettant d'approcher à line globale des distributions de données à queue lourde. Cependant dans la pratique, dans beaucoup des cas, les données ne sont pas bien ajustées par ces dernières, ayant soit une queue moins lourde ou plus lourde. Cela dit nous auront à tester ces lois.

La méthode de modélisation dite à deux lois est la plus poussée et semble être cohérente vis-à-vis du cadre de l'étude. Nous détailleront après le cadre théorique du choix de seuil de séparation ou d'écrêtement.

## 3.1 Statistiques As-If

Nous disposons d'une base de données des montants de sinistres s'étalant sur plusieurs années. La valeur de l'argent évoluant dans le temps, il est tout à fait normal et indispensable de prendre cela en considération pour évaluer correctement les risques. Il convient donc de revaloriser les montants des sinistres à la date de modélisation pour avoir un échantillon comparable sur une même base, on parle aussi de **"mise as-if des sinistres"**. Les sinistres passés ont un coût, l'objectif est donc de déterminer le coût de ces mêmes sinistres s'ils étaient survenus aujourd'hui.

Cette revalorisation est réalisée à partir des indices pouvant refléter l'effet d'un facteur modifiant le coût des sinistres. Les principaux facteurs as-if sont les suivants :

- L'inflation : L'inflation est l'augmentation du niveau général des prix accompagnée par une baisse des taux. Elle résulte à une perte de pouvoir d'achat de la monnaie. Dans beaucoup de pays, en l'occurrence le Maroc, l'indice la modélisant est généralement basé sur le prix à la consommation. 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.
- Des changements réglementaires : En effet un changement réglementaire intrinsèque ou systématique peut avoir des conséquences sur l'évaluation du coût d'un sinistre comme par exemple un changement du taux d'actualisation.
- Changement dans la politique de souscription : des changements dans 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. Cela dit, ces facteurs sont difficiles à quantifier et à transformer en indice.

### Méthodologie

Il existe plusieurs méthodes envisageables pour revaloriser les sinistres : des méthodes prospectives, des méthodes rétrospectives ou encore un mélange entre les deux. Nous avons privilégié une approche prospective reposant sur l'estimation des valeurs futures des indices d'inflation qui vont servir à la revalorisation. Nos 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 basé sur un indice d'inflation.

On considère  $k$  sinistres survenus durant les années  $a_1 < \dots < a_k$  avec  $a_k$  l'année de couverture à évaluer. Notons par  $i_1, \dots, i_k$  les taux d'inflations considérés, et par  $I_1, \dots, I_k$  les indices d'inflation associés. Notons également par  $r_1, \dots, r_k$  les facteurs de redressements (FR) associés aux années  $a_1, \dots, a_k$  et par  $S_1, \dots, S_k$  les montants de sinistres historiques et enfin  $\tilde{S}_1, \dots, \tilde{S}_k$  les sinistres revalorisés.

Année de survenance	Montant historique	Taux d'inflation	Indice d'inflation	FR	Montant As-If
$a_1$	$S_1$		$I_1$	$r_1 = \frac{I_k}{I_1}$	$\tilde{S}_1 = S_1 \times r_1$
$a_2$	$S_2$	$i_2$	$I_2 = (1 + i_2) \times I_1$	$r_2 = \frac{I_k}{I_2}$	$\tilde{S}_2 = S_2 \times r_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$a_k$	$S_k$	$i_k$	$I_k = (1 + i_k) \times I_{k-1}$	$r_k = 1$	$\tilde{S}_k = S_k$

TABLE 3.1 – Table de réévaluation par inflation

## 3.2 Choix du seuil

L'objectif est de déterminer un seuil permettant d'établir correctement les sinistres extrêmes. En effet l'étude de la queue de distribution des montants de sinistres est cruciale pour appréhender le risque et mieux le gérer. La théorie des valeurs extrêmes permet de répondre à cette problématique de meilleure maîtrise des risques. Justement, 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.

Nous considérons dans la suite  $X_1, \dots, X_n$  des variables aléatoires indépendantes et identiquement distribuées représentant les montants de sinistres d'un portefeuille. Nous notons  $F$  leur fonction de répartition. L'échantillon ordonné par ordre croissant est noté  $X_{(1)} \leq \dots \leq X_{(n)}$

En effet dans le théorie classique, le théorème Centrale limite permet d'obtenir un résultat sur le comportement de la somme normalisée des 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.

### 3.2.1 La loi des valeurs extrêmes - GEV(Generalized Extreme Value)

### Définition 1 (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  $H$  nous avons,

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

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  $H_\gamma$  définie pour tout  $x$  tel que  $1 + \gamma x > 0$

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

La loi limite du maximum dépend donc d'un seul paramètre  $\gamma$ , appelé l'indice des valeurs extrêmes. Selon le signe de  $\gamma$ , on définit trois domaines d'attraction :

- $\gamma > 0$  : on dit que  $F(\cdot)$  appartient au domaine d'attraction de Fréchet. Il contient les lois dont la fonction de survie décroît comme une fonction puissance. On parle aussi de lois à queue lourde dans ce domaine d'attraction, on trouve les lois de Pareto, de Student, de Cauchy, etc...
- $\gamma = 0$  :  $F(\cdot)$  est dans le domaine d'attraction de Gumbel qui regroupe les lois ayant une fonction de survie à décroissance exponentielle. C'est le cas des lois normale, gamma, exponentielle, etc...
- $\gamma < 0$  :  $F(\cdot)$  appartient au domaine d'attraction de Weibull. Ce domaine contient les lois dont le point terminal  $x_F = \inf\{x, F(x) \geq 1\}$  est fini. C'est le cas par exemple des lois uniformes, lois beta, etc...

Domaine d'attraction	Gumbel $\gamma = 0$	Fréchet $\gamma > 0$	Weibull $\gamma < 0$
Loi	Normale Exponentielle Lognormale Gamma Weibull	Cauchy Pareto Student Burr	Uniforme Beta

FIGURE 3.1 – Résumé sur le domaines d'attraction des lois usuelles

### 3.2.2 Théorie des excédents au-delà d'un certain seuil

La théorie des valeurs extrêmes ainsi présentée précédemment s'intéresse à l'étude de la distribution du maximum d'un échantillon. cela dit dans le cadre de notre étude cette approche est trop restrictive. En effet il est plus intéressant pour nous d'étudier l'ensemble de la distribution de la queue lourde. L'approche POT (*Peaks Over the Threshold*) qui se repose sur la loi de pareto généralisé permet de répondre à ce besoin. 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.

Ce théorème ne se basant plus sur l'étude du maximum, s'intéresse à l'étude du comportement asymptotique de la distribution des excès  $X - u \mid X > u$  lorsque  $u \longrightarrow +\infty$

#### Loi des excess - méthode Peaks Over the Threshold :

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 :

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

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

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

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  $X - u \mid X > u$

### Loi de Pareto généralisée

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

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

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

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

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  $\mu$ . 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 :

$$G_{\gamma,\sigma}(x) = 1 + \ln \left[ H_\gamma \left( \frac{x}{\sigma} \right) \right]$$

## 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.

*La fonction de répartition  $F$  de point terminal  $x_F$  appartient au domaine d'attraction de  $H_\gamma$  si et seulement si il existe une fonction positive  $\sigma$  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  $H_\gamma$  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  $\gamma$  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.

Il existe plusieurs méthodes pour estimer les paramètres de la loi de Pareto généralisée dans l'approche Peaks Over the Threshold. On distingue les méthodes paramétriques et non paramétriques.

### 3.2.3 Estimation paramétrique

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 estimation se base sous l'hypothèse :  $Y_1, \dots, Y_{k_n}$  supposées indépendantes et de même loi  $\mathcal{GPD}(\gamma, \sigma)$ .

On distingue donc **la méthode du maximum de vraisemblance**, **la méthode des moments** ou encore **la méthode des moments pondérés** présentée par **Hosking et kal en 1985**.

### 3.2.4 Estimation semi-paramétrique de l'indice des valeurs extrême

Les estimateurs semi-paramétriques s'appliquent à des échantillons dont la distribution n'est plus  $H_\theta$ , 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.

#### Estimateur de Hill

L'estimateur de Hill est l'estimateur le plus populaire pour estimer la valeur de l'indice des valeurs extrêmes  $\gamma$ . 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  $\lim_{n \rightarrow +\infty} \frac{k_n}{n} = 0$ , nous pouvons montrer que l'estimateur de Hill est asymptotiquement gaussien :

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

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  $\alpha$  de sorte que la relation  $\alpha = \frac{1}{\gamma}$  est vérifiée.

#### Estimateur de Pickands

Cet estimateur 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  $\gamma$ . 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 :

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

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.

### 3.2.5 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 :

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

Avec  $\hat{\gamma}_{j,n}^{(UH)} = X_{(n-j)} \times \hat{\gamma}_n^{(H)}(k_n)$  La courbe trouvée ressemble à l'une de ces trois :

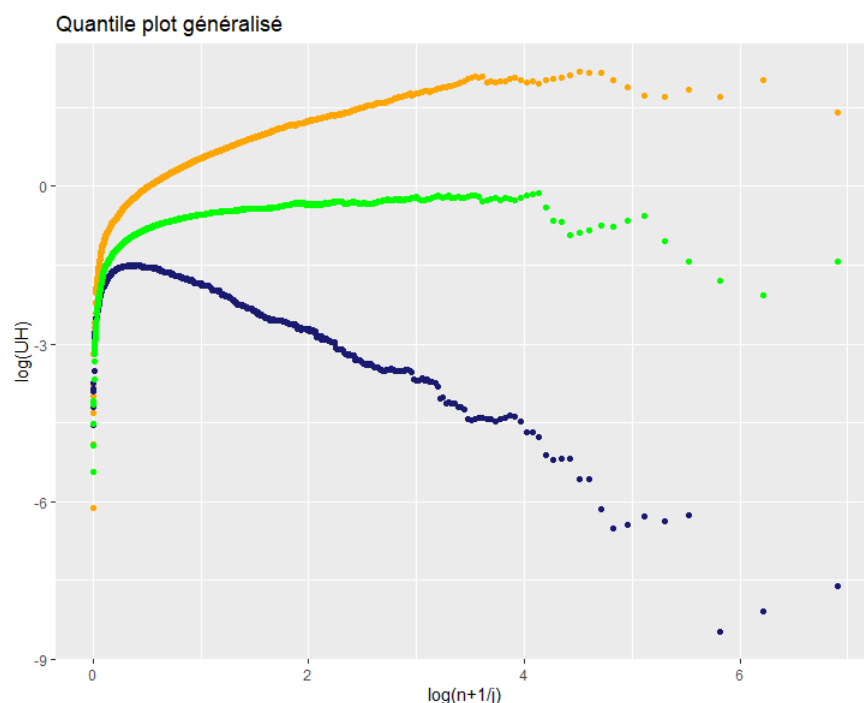


FIGURE 3.2 – Simulation du quantile plot généralisé

Si la courbe tracée avec les données ressemble à la première (**Orange**), elle appartient au domaine d'attraction de Fréchet ( $\gamma > 0$ ), si elle est comme celle en **vert**, elle appartient au domaine de

Gumbel ( $\gamma = 0$ ), et en fin si elle est comme celle en **bleu** foncé, elle appartient au domaine de weibull ( $\gamma < 0$ ).

### 3.2.6 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. Nous baserons principalement sur **le graphique de Hill et de Pickands** en raison de l'efficacité recommandée de ces méthodes.

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  $X_{(1)}, \dots, X_{(n)}$ .

#### 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.

## 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.

## Méthode de minimisation de l'AMSE

Cette méthode a été introduite par Caeiro et Gomes en 2016. Elle se base sur le principe suivant :

étant donné, l'estimateur de Hill  $\hat{\gamma}_n^{(H)}(k_n) = \frac{1}{k_n} \sum_{i=1}^{k_n} \ln(X_{(n-i+1)}) - \ln(X_{(n-k_n)})$ , on détermine le nombre d'excès  $k_0$  minimisant l'erreur quadratique moyenne de ce dernier. Il 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 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 :

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

Avec  $\lceil \cdot \rceil$  la partie entière et  $\rho$  et  $\lambda$  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 :

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

On obtient donc 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. L'avantage principale de cette méthode sur les autres est qu'elle fournit un estimateur analytique précis .

## 3.3 Les modèles de modélisation

L'objectif de notre modélisation est d'étudier la distribution de probabilité de la charge totale  $S$  de sinistres sur un portefeuille donné au cours d'une période donnée.

Comme dit dans la section précédente, nous rappelons que la considération d'un modèle simple à une seule loi s'adapte souvent mal aux données réelles issues de la réassurance, qui rappellent le comportement souvent deux types de sinistralité, la sinistralité attritionnelle qui correspond à des sinistres faibles de grande fréquence et la sinistralité majeure ou extrême qui correspond aux sinistres de forte sévérité mais de faible fréquence. Le seuil des extrêmes étant déterminé à travers les méthodes quantitatives précédemment présentées, nous présenterons un modèle prenant en compte ces deux types de sinistralité. Le coût attritionnel sera modélisé par une loi classique. Le coût extrême, quant à lui, fera l'objet d'une modélisation par une loi de Pareto généralisée, ce qui est cohérent avec les résultats de la théorie des valeurs extrêmes.

### 3.3.1 Modèle collectif

#### Définition

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.

La variable  $X$  sera estimée par une seule loi usuelle ou par une nouvelle distribution qui sera construite à partir d'une loi usuelle et d'une loi GDP en les alliant de façon à avoir une unique densité ainsi qu'une fonction de répartition unique.

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 :

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

### 3.3.2 Modèle collectif séparé

Nous avons appelé Modèle séparé, un modèle dans lequel la charge totale  $S$  est définie de la façon suivante,

$$S = S_1 + S_2 = \sum_{i=1}^{N_1} X_{1i} + \sum_{i=1}^{N_2} X_{2i}$$

Avec :

$X_{j1}, \dots, X_{jN}$  des variables aléatoires représentant les montants de chaque sinistre du portefeuille. Avec  $X_{j=1}$  la variable aléatoire représentant les coût attritionnels (soit donc  $X$  tronquée à droite au seuil  $u$ ) et  $X_{j=2}$  la variable aléatoire représentant les coûts extrêmes. Ces variables sont supposées indépendantes et identiquement distribuées de même loi qu'une variable aléatoire  $X$ .

$N_1$  et  $N_2$  sont des variables aléatoires à valeurs dans  $\mathbb{N}$  indépendantes de la suite  $(X_{ji})_{i \geq 1, j \in \{1,2\}}$ . Ces variables représentent le nombre total de sinistres survenus au sein du portefeuille respectivement pour les sinistres attritionnels et pour les sinistres extrêmes.

Comme dans le modèle collectif, 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.

$$\mathbb{E}(S) = \mathbb{E}(N_1) \times \mathbb{E}(X_1) + \mathbb{E}(N_2) \times \mathbb{E}(X_2)$$

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

$$\sigma(S) = \sqrt{\mathbb{E}(N_1) \times \sigma^2(X_1) + \mathbb{E}^2(X_1) \times \sigma^2(N_1) + \mathbb{E}(N_2) \times \sigma^2(X_2) + \mathbb{E}^2(X_2) \times \sigma^2(N_2)}$$

Dans la suite, ces deux modèles feront l'objet de discussions en fonction du résultat obtenu.

#### Remarque

En général, dans les modélisations qui ont pour objectif la tarification d'une couverture en réassurance, les données sont censurées à gauche par un seuil un seuil d'avis de sinistres. Ce seuil s'appelle le seuil de modélisation, et est par définition la valeur en dessous de laquelle les montants de sinistres inférieurs ne sont pas considérés pour la modélisation. Le choix de ce seuil de modélisation est important en l'occurrence car il permet de ne conserver que les données jugées pertinentes. L'établissement de celui-ci peut faire l'objet d'une étude. Cela dit le but de notre travail est l'établissement d'une modélisation pour différente utilisation dans l'établissement de la couverture notamment l'optimisation du modèle de rétrocession pour le portefeuille non marin pour le cas incendie. Ainsi toute sensibilité sera bonne à voir. Les données sur lesquelles nous travailleront sont des données complètes.

## 3.4 Modélisation de la sévérité

Cette section fait état de la notion d'ajustement paramétrique des coûts par des distributions de probabilité. Nous considérons un échantillon d'observations noté  $(x_i)_{1 \leq i \leq n}$ . Ces observations représentent les montants de sinistres as-if sur le portefeuille étudié. Nous considérons que ces observations sont la réalisation de la variables aléatoires  $(X_i)_{1 \leq i \leq n}$  indépendantes et identiquement distribuées suivant la même loi. Nous notons  $F$  leur fonction de répartition. L'objectif de l'actuaire est d'approcher de la meilleure manière la distribution de probabilité de  $X$ .

### 3.4.1 Fonction de répartition empirique

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  $\bar{F}$  car elle est simple à construire et surtout car le théorème de Glivenko-Cantelli assure la convergence suivante :

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

Bien qu'elle converge presque sûrement vers la fonction de répartition théorique, elle présente un inconvénient majeur. En effet 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.

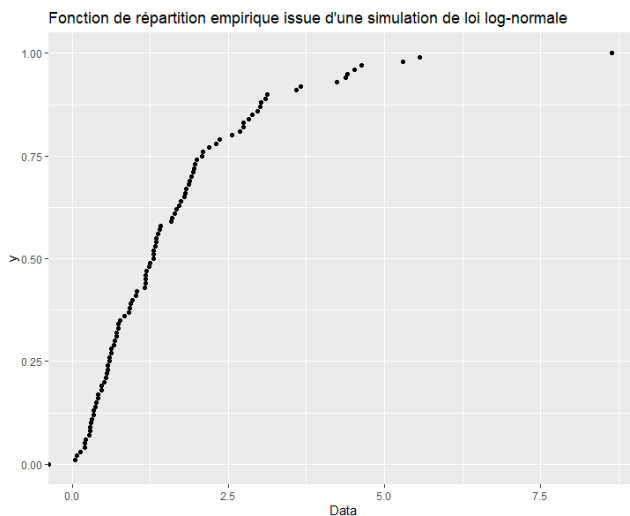


FIGURE 3.3 – Fonction de répartition empirique

### 3.4.2 Lois usuelles en réassurance

La modélisation des distributions en loi de probabilité est indispensable pour modéliser des couvertures de réassurance pour lesquelles les données d'expérience sont en faible quantité voire inexistantes. Dans cette partie, nous présentons rapidement les lois fréquemment utilisées en réassurance pour modéliser la loi du montant des sinistres.

#### Loi log-normale

La variable  $X$  suit une loi log-normale de paramètres  $\mu$  et  $\sigma$ , ( $X \sim \ln N(\mu, \sigma)$ ) si le logarithme de  $X$  suit une loi normale ( $\ln(X) \sim N(\mu, \sigma)$ ). Sa densité est donnée par l'expression suivante :

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

où  $x > 0$ ,  $\mu \in \mathbb{R}$  et  $\sigma > 0$ .

Les moments soit donc l'espérance et la variance de la loi log-normale sont donnés par :

$$\begin{aligned} E(X) &= \exp(\mu + \sigma^2/2) \\ \text{Var}(X) &= [\exp(2\mu + \sigma^2)] [\exp(\sigma^2) - 1] \end{aligned}$$

#### Loi Gamma

La variable  $X$  suit une loi Gamma de paramètres  $\alpha$  et  $\beta$ , ( $X \sim Ga(\alpha, \beta)$ ) si sa densité est de la forme :

$$f(x) = \frac{\beta^{-\alpha} x^{\alpha-1} \exp(-x/\beta)}{\Gamma(\alpha)} \mathbb{I}_{\mathbb{R}^+}(x)$$

où  $\alpha > 0$ ,  $\beta > 0$  et  $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$ .

Les moments de la loi Gamma sont donnés par :

$$\begin{aligned} E(X) &= \alpha\beta \\ \text{Var}(X) &= \alpha\beta^2 \end{aligned}$$

#### Loi de Weibull

La variable  $X$  suit une loi de Weibull de paramètres  $\alpha$  et  $\beta$ , ( $X \sim W(\alpha, \beta)$ ) si sa densité est de la forme :

$$f(x) = \alpha\beta^{-\alpha} x^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^\alpha} \mathbb{I}_{\mathbb{R}^+}(x)$$

et sa fonction de répartition :

$$F(x) = 1 - \exp\left[-\left(\frac{x}{\beta}\right)^\alpha\right]$$

où  $\alpha > 0$  et  $\beta > 0$ .

Les moments de la loi de Weibull sont donnés par :

$$E(X) = \frac{\beta}{\alpha} \Gamma\left(\frac{1}{\alpha}\right)$$
$$\text{Var}(X) = \frac{\beta^2}{\alpha} \left\{ 2\Gamma\left(\frac{2}{\alpha}\right) - \frac{1}{\alpha} \left[ \Gamma\left(\frac{1}{\alpha}\right) \right]^2 \right\}$$

### Loi de Pareto

La variable  $X$  suit une loi de Pareto de paramètres  $\theta$  et  $a$ , ( $X \sim Pa(\theta, a)$ ) si sa densité est de la forme :

$$f(x) = \frac{\theta a^\theta}{x^{\theta+1}} \mathbb{1}_{[a, +\infty[}(x)$$

et sa fonction de répartition :

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

où  $\theta > 0$  et  $a > 0$ . Les moments de la loi de Pareto sont donnés par :

$$E(X) = \frac{a\theta}{\theta - 1}$$
$$\text{Var}(X) = \frac{\theta a^2}{(\theta - 2)(\theta - 1)^2}$$

### 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 :

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

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 :

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

## Loi Pareto Généralisée

Une variable aléatoire  $Y$  suit une distribution de 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 :

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

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

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

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

pour  $x \geq 0$  si  $\gamma \geq 0$  et  $0 \leq x \leq -\sigma/\gamma$  si  $\gamma < 0$ . L'espérance de la loi de Pareto généralisée est finie si et seulement si  $\gamma < 1$ . De manière générale, si  $\gamma < 1/r$ , avec  $r \in \mathbb{N}$ , alors on a

$$E(X^r) = \frac{\sigma^r}{\gamma^{r+1}} \frac{\Gamma(\gamma^{-1} - r)}{\Gamma(1 + \gamma^{-1})} r!$$

## Lois tronquées

Soit  $X$  une variable aléatoire réelle de fonction de répartition  $F_X$  dont on cherche la loi de distribution. Soit  $s$  et  $u$  deux seuils. Le premier modélise la troncature à gauche, 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 :

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

### 3.4.3 Analyse préliminaire des distributions candidates :

Avant de passer à l'estimation analytique par les méthodes quantitatives comme la méthode de maximum de vraisemblance et où des moments, on fait généralement recours à des indicateurs ou à des représentations graphiques nous permettant d'avoir une idée plus ou moins précise de la loi théorique modélisant notre plage de données.

— **Histogramme de densité et graphe cumulative de distribution :**

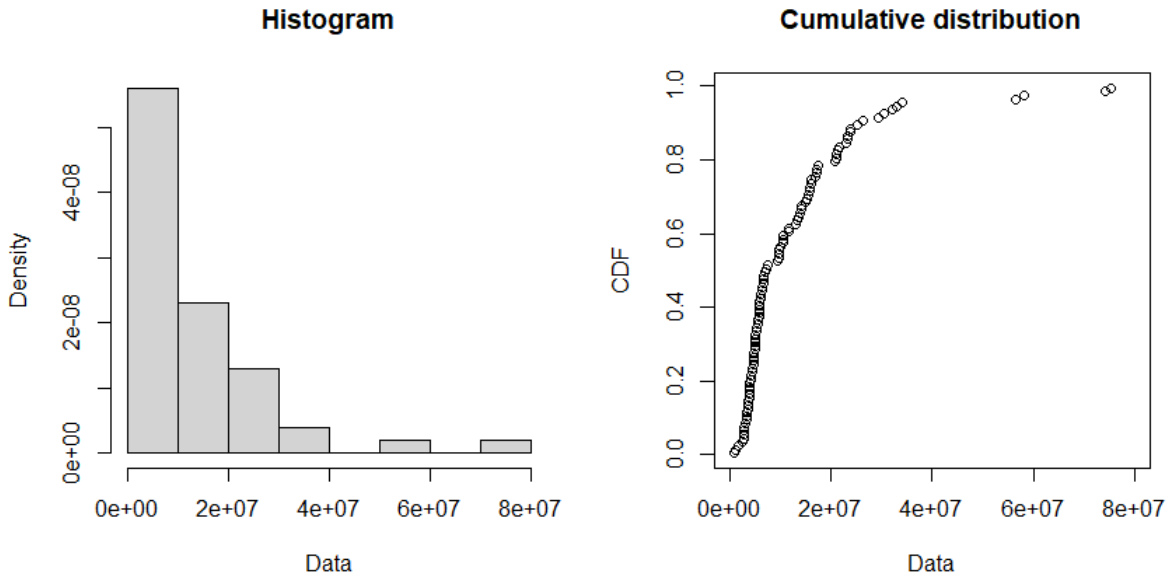


FIGURE 3.4 – Représentation de la distribution sur un échantillon simulé d'une loi log-normale

Ces deux graphiques représentent respectivement les couples  $(x(data), f(x))$  densité de probabilité et  $(x, F(x) = P(X \geq x))$  Fonction cumulative de distribution ou fonction de répartition). Ils permettent de faire une interprétation plus ou moins vague de la loi théorique possible.

### — Calcul des coefficients d'asymétrie et de pointicité

**Coefficient d'asymétrie :** En anglais skewness, le coefficient d'asymétrie nous permet de savoir comment nos données sont réparties par rapport à la valeur centrale. Une distribution est dite symétrique si les valeurs observées se répartissent de façon uniforme autour des trois valeurs centrales : la moyenne, le mode et la médiane. Sinon on dit qu'elle est asymétrique.

$$\text{skewness} = \frac{\frac{1}{n} \times \sum (x_i - \mu)^3}{\sigma^3}$$

- skewness = 0 : distribution symétrique
- skewness > 0 : asymétrie positive c'est à dire que la queue à droite est plus grosse.
- skewness < 0 : asymétrie négative soit donc une queue plus grosse à gauche.

**Coefficient de pointicité :** En théorie de probabilité, le coefficient de pointicité (kurtosis) est une mesure directe de l'acuité et une mesure indirecte de l'aplatissement de la distribution d'une variable aléatoire réelle. Il est donné par :

$$\text{Kurtosis} = \frac{\frac{1}{n} \times \sum (x_i - \mu)^4}{\sigma^4}$$

- kurtosis = 3 : dist. mésokurtique (distribution normale)
- kurtosis > 3 : dist. leptokurtique (pointue et à queues épaisses)
- kurtosis < 3 : dist. platikurtique (plate)

### — Graphe de Cullen et Frey avec bootstrap

Le graphique de Cullen et Frey est un graphique représentant le couple de valeurs (skewness, Kurtosis) pour différentes distributions connues afin d'aider au choix des distributions en fonction des données. Pour certaines distributions (normale, uniforme, logistique, exponentielle), il n'y a qu'une seule valeur possible pour le skewness et le kurtosis. Ainsi, la distribution est représentée par un point unique sur le graphique. Pour d'autres distributions, les zones des valeurs possibles sont représentées avec lignes (comme pour les distributions gamma et log-normale), ou des zones plus grandes (comme pour la distribution bêta).

Afin d'avoir un meilleur résultat d'interprétation, on applique au graphe la technique de bootstrap. Les techniques de bootstrap sont des méthodes d'inférence statistique basées sur la réplification multiple des données à partir du jeu de données étudié, selon les techniques de rééchantillonnage. Le principe est le suivant, lorsqu'on veut étudier ou apprécier les variations d'estimation d'une statistique donnée, on génère un échantillon de données en effectuant sur

notre échantillon initial un tirage aléatoire avec remise. On parle d'échantillon bootstrap. Ensuite, on ré-estime à partir ce dernier notre statistique d'étude, on se retrouve ainsi avec plusieurs valeurs de l'estimation qu'on pourra utiliser pour diverse interprétation. **Les valeurs bootstrap dans le graphe de cullen et frey sont colorées en jaune foncé.**

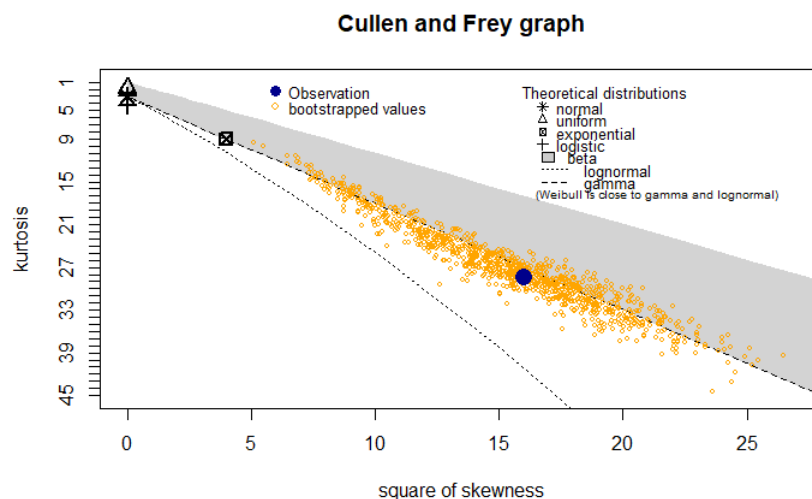


FIGURE 3.5 – Exemple graphe de Cullen and Frey

Selon le résultat de ce graphique, les distributions probables ajustant nos données sont les lois log normale, gamma, weibull.

### 3.4.4 QQ-Plot

Le quantile quantile Plot est un outil mathématique permettant de comparer et d'évaluer l'ajustement d'un modèle de distribution théorique à un modèle empirique en traçant leurs quantiles l'un par rapport à l'autre afin d'étudier visuellement la linéarité entre ces deux quantités. Si on considère  $(X_1, \dots, X_n)$  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, mathématiquement et pour les lois qui nous intéressent, on a :

Le QQ Plot le plus utilisé est vraisemblablement celui de la loi exponentielle, où  $F(x) = 1 - \exp(-x/\lambda)$  pour  $x \geq 0$ . La fonction quantile est donnée par  $F^{-1}(p) = -\lambda \log(1 - p)$ ,  $p \in ]0, 1[$ . Le QQ Plot exponentiel consiste alors à tracer les points :

$$\left( -\ln \left( 1 - \frac{i}{n+1} \right), x_{(i)} \right) \quad i = 1, \dots, n$$

Le QQ Plot de la loi de Pareto consiste à tracer les points :

$$\left( -\ln \left( 1 - \frac{i}{n+1} \right), \ln(x_{(i)}) \right) \quad i = 1, \dots, n$$

Concernant la loi lognormale, on utilise le fait que si l'échantillon suit une loi log-normale, alors son logarithme suit une loi normale. Le QQ Plot Log-normal consiste donc à tracer les points :

$$\left( \phi^{-1} \left( \frac{i}{n+1} \right), \ln(x_{(i)}) \right) \quad i = 1, \dots, n$$

où  $\phi$  est l'inverse de la fonction de répartition de la loi normale.

Enfin, si on considère la loi de Weibull, le QQ Plot consiste à tracer les points

$$\left( \ln \left( -\ln \left( 1 - \frac{i}{n+1} \right) \right), \ln(x_{(i)}) \right) \quad i = 1, \dots, n$$

Les figures suivantes illustrent l'usage des différents QQ-Plots que nous venons de présenter. Nous les avons appliqués à des simulations de loi log-normale.

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.

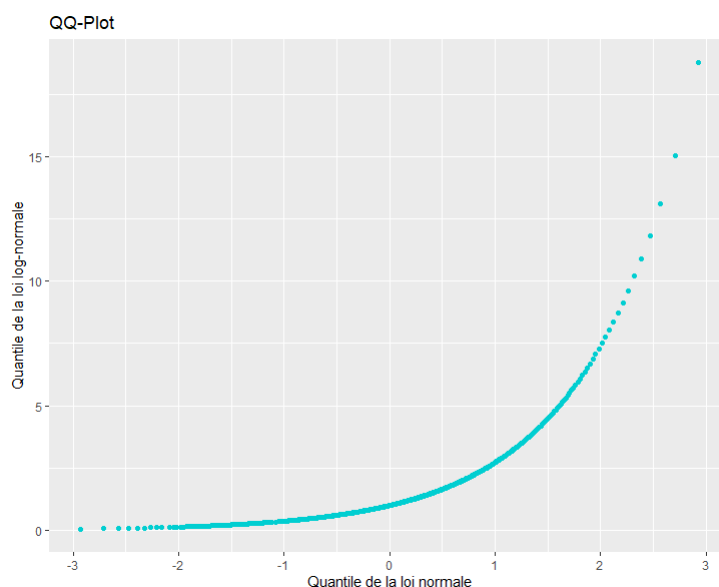


FIGURE 3.6 – Exemple d'un QQ-PLOT

Le QQ-Plot représenté ci-dessus est celui entre une loi normale en abscisses et une loi log-normale en ordonnées. Cette allure convexe nous traduit non seulement que ces lois ne sont pas les mêmes mais aussi que la loi log-normale a une queue plus épaisse que la loi normale, résultat classiquement connu. Attention cette interprétation n'est pas systématique car si on inverse les axes alors la courbe devient prendre une allure concave. Cela étant nous travaillerons de la manière précédemment décrite.

### 3.4.5 Maximum de vraisemblance : Estimation des paramètres

Afin d'estimer les paramètres de la loi modélisant notre jeu de données, plusieurs méthodes sont possibles comme la méthode des moments ou celle de vraisemblance. Il est admis dans la littérature que la méthode maximum de vraisemblance fournit des résultats de meilleures qualités. Nous utiliserons cette méthode.

On considère un échantillon  $X_1, \dots, X_n$  de densité  $f_X$  et dont la loi a pour paramètre  $\theta$ , le maximum de vraisemblance consiste à maximiser la fonction suivante :

$$\mathcal{L}(\theta) = \prod_{i=1}^n f_X(x_i, \theta)$$

En pratique, on utilise la fonction log-vraisemblance  $l$  qui est le logarithme népérien de  $L$  et qui est plus facile à mettre en œuvre car elle fait intervenir des sommes au lieu de produits. On écrit :

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

Dans le cas où, on ne disposons pas des  $X_1, \dots, X_n$ , mais des

$Y_1, \dots, Y_n$  tronqués, le coefficient correcteur intervient dans la fonction du maximum de vraisemblance : et la log vraisemblance s'écrit :

$$\mathcal{L} = \prod_{i=1}^n f_Y(x_i) = \prod_{i=1}^n \frac{f_X(x_i)}{F_X(u) - F_X(s)}$$

$$l = \sum_{i=1}^n \ln[f_X(x_i)] - n \ln[F_X(u) - F_X(s)]$$

L'estimateur du maximum de vraisemblance de  $\theta$  est la valeur  $\hat{\theta}_n$  qui maximise la fonction de vraisemblance  $\mathcal{L}$  par rapport à  $\theta$  :

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

Il est donné par :

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

On définit alors les fonction de score comme :

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

Maximiser  $\mathcal{L}$  revient à résoudre les  $p$  équations suivantes :

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

En général, pour la plus part des lois, on est capable de déterminer de façon explicite l'estimateur de maximum de vraisemblance, cependant il existe des cas où il est impossible de fournir celui-ci de façon explicite, on a recours à des méthodes numériques utilisant des algorithmes d'optimisations. L'algorithme le plus utilisé est celui de Nelder-Mead, il est initialisé à partir de la valeur de l'estimateur obtenu grâce à la méthode des moments.

On note par  $\hat{\theta}_n$  l'estimateur du maximum de vraisemblance de  $\theta$  sur un échantillon de taille  $n$ , il satisfait les propriétés suivantes :

-- **Asymptotiquement sans biais et efficace** : Pour tout  $1 \leq i \leq p$  :

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

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

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

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

-- **Consistance** : L'estimateur  $\hat{\theta}_n$  converge presque sûrement vers  $\theta$  :

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

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

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

Cette propriété de normalité asymptotique est un résultat central dans la construction d'intervalles de confiance pour l'estimation de  $\theta$ . C'est pour toutes ces propriétés que l'estimateur du maximum de vraisemblance est préféré à l'estimateur des moments dans la suite.

Une fois les paramètres estimés, on peut effectuer des représentations graphiques comme le QQ-Plot (quantile-quantile plot) ou le PP-Plot (probability-probability plot) ainsi que des tests statistiques afin de savoir quel ajustement choisir. Pour tester l'adéquation à des lois continues, on a souvent recours aux tests de Kolmogorov-Smirnov et d'Anderson-Darling. Ces tests sont construits à partir de la fonction de répartition empirique. Le test d'Anderson-Darling est particulièrement intéressant car il accorde plus de poids aux queues de distribution que Kolmogorov-Smirnov.

### 3.4.6 PP-Plot

Le diagramme Probability-Probability Plot quant à lui est un graphique qui agence l'une sur l'autre les fonctions de distributions cumulatives. La fonction de répartition théorique  $F_n$  et la fonction de répartition  $F$ . La ligne de comparaison est la droite  $x = y$ . Les distributions sont égales si et seulement si le graphique tombe sur cette ligne – tout écart indique une différence entre les distributions.

### 3.4.7 Test de Kolmogorov-Smirnov

Le test d'ajustement de Kolmogorov – Smirnov est un test non paramétrique basé sur la distance maximale entre une fonction de répartition théorique (entièrement déterminée par les valeurs de ses paramètres) et la fonction de répartition empirique de l'échantillon. Il se traduit mathématiquement de la façon suivante, on définit par  $(X_1, \dots, X_n)$  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 :

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

On note par  $D_n$  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  $D_n^+ = \left| \frac{i}{n} - F_0(X_{(i)}) \right|$  et  $D_n^- = \left| \frac{i-1}{n} - F_0(X_{(i)}) \right|$ . La région critique du test au seuil  $\alpha$  est définie par :

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

où  $d_\alpha$  et 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.

Dans le cas où on aura à travailler avec des lois tronquées, il faut modifier la fonction de répartition théorique :  $F'_0(x) = \frac{F_0(x) - F_0(d)}{F_0(s) - F_0(d)}$  et le test devient alors

$$\begin{cases} H_0 : F_X = F'_0 \\ H_1 : F_X \neq F'_0 \end{cases}$$

### 3.4.8 Test de 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 suit :

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

Cette statistique peut également s'écrire :

$$A_n^2 = -n - \frac{1}{n} \sum_{i=1}^n [(2i - 1) \ln(Z_i) + (2n + 1 - 2i) \ln(1 - Z_i)]$$

où  $Z_i = F_0(X_i)$ . Le test au seuil  $\alpha$  associé à cette statistique est défini par la région critique de la forme :  $\{A_n^2 \geq c_\alpha\}$  où  $c_\alpha$  est le quantile  $(1 - \alpha)$  de la table d'Anderson-Darling.

Si on travaille avec des lois tronquées, il faut modifier la fonction de répartition théorique pour en tenir compte de la même manière que pour le test de Kolmogorov-Smirnov :  $F_0'(x) = \frac{F_0(x) - F_0(d)}{F_0(s) - F_0(d)}$ .

Bien que le test d'Anderson Darling et le test de Kolmogorov ont la même application, la différence entre eux réside dans le fait que pour le deuxième test seul l'écart maximum entre la distribution empirique et la distribution d'ajustement compte, alors que l'indicateur d'écart du premier test capte mieux l'ensemble des données dans la mesure où toute la somme des écarts intervient. On peut donc déduire que le test d'Anderson-Darling donne une importance plus grande aux queues de distribution.

### 3.4.9 Spécification du modèle

En générale, dans les modélisation visant à l'étude des couvertures, les quelques sinistres les plus extrêmes peuvent avoir un impacte important. Ainsi notre but est de donner une attention particulière à ces sinistres sans pour autant mettre un lourd accent sur ces dernières. Nous cherchons à construire mathématiquement un modèle basé soit sur une seule loi de probabilité usuelle, soit sur une loi de probabilité construite à partir d'une loi usuelle et d'une loi de Pareto Généralisée, soit sur deux variables aléatoires représentant les charges  $S_1$  et  $S_2$  chacune modélisant un type particulier de sinistres(attributionnels ou extrêmes).

La première difficulté est la détermination de ce seuil. Le choix de notre seuil à été basé sur la théorie des valeurs extrêmes particulièrement sur les outils quantitatifs comme le Hill-Plot le Pickands-Plot et la méthode "Minimizing the AMSE of the Hill estimator". Sauf mention contraire, ce seuil sera noté  $u$ .

Il faut maintenant estimer les paramètres de deux lois. L'estimation des paramètres de la première loi, par exemple une loi log-normale, relève des méthodes classiques telles que le maximum de vraisemblance ou la méthode des moments. La loi de Pareto généralisée comporte trois paramètres. Le premier paramètre définit le début du support de la distribution. Nous prenons donc le seuil  $u$  précédemment déterminé par les méthodes graphiques. Il reste donc deux paramètres à

estimer : le paramètre de queue  $\gamma$  et le paramètre d'échelle  $\sigma$ . Il est possible de déterminer conjointement ces deux paramètres par la méthode du maximum de vraisemblance ou par la méthode des moments. Pour vérifier que ces deux lois conviennent bien, il est nécessaire d'effectuer des tests d'adéquation. Pour la sinistralité attritionnelle, nous pouvons utiliser le test de Kolmogorov-Smirnov ou d'Anderson-Darling. Pour la sinistralité extrême, il faudrait également effectuer un test d'adéquation, cependant la littérature ne nous offre pas de réponse satisfaisante. Le théorème de Picands-Balkema-De Hann est le garant du bien fondé de l'utilisation de la distribution de Pareto généralisée pour la modélisation des excès après le seuil, de la même manière que d'après le théorème central limite, nous pouvons modéliser par une loi normale la moyenne de  $n$  variables aléatoires indépendantes et identiquement distribuées, sous des hypothèses convenables et pour  $n$  assez grand.

### Modèle à deux lois intégrées

$$S = \sum_i^N X_i$$

Mathématiquement, nous voulons construire une distribution globale s'ajustant parfaitement à la sévérité  $X$ , composée d'une distribution de Pareto généralisée (comme le suggère le théorème de Picands-Balkema-De Hann.) qui modélise les coûts au delà d'un seuil soit donc sur  $[u, +\infty]$  et d'une distribution usuelle tronquée à ce point sur  $[0, u]$ . Nous devons prendre des précautions et effectuer un raisonnement clair sur la construction de la loi globale, en effet, La juxtaposition des fonctions de répartition sans précaution conduit à une aberration dans la pratique, puisque en  $u$ , nous passerions d'une valeur strictement positive à 0 et nous n'aurions plus une fonction croissante, autrement dit nous perdons les propriétés de base d'une fonction de répartition. il faut éviter que le saut de valeurs dans l'espérance de la sinistralité lorsque la priorité ou la portée varie très légèrement, donc nous cherchons également une fonction continue. L'idée générale est la suivante,

Soit  $F$  la fonction de répartition de la variable aléatoire  $X$ . Nous démontrons :

$$\begin{aligned} F(x) &= \mathbb{P}(X \leq x) = \mathbb{P}(X \leq x \cap X < u) + \mathbb{P}(X \leq x \cap X \geq u) \\ &= \mathbb{P}(X \leq u) \mathbb{P}(X \leq x \mid X \leq u) + (1 - \mathbb{P}(X \leq u)) \mathbb{P}(X \leq x \mid X > u) \\ &= \mathbb{P}(X \leq u) F^{UsuelleTronquee}(x) + (1 - \mathbb{P}(X \leq u)) F_u(x - u) \end{aligned}$$

Avec,  $F_u(x) = \mathbb{P}(X - u \leq x \mid X > u)$  explicitée dans la partie (3.2.2), et  $F^{UsuelleTronquee}(x) = \mathbb{P}(X \leq x \mid X < u)$ . Nous savons que d'après le théorème de Pickands-Balkema-de Haan, pour  $u$  assez grand,  $F_u(x - u)$  tend vers  $G_{\gamma, \sigma}(x - u) = G_{\gamma, u, \sigma}(x)$  à support, nous noterons  $\mathbb{P}(X \leq u) = \alpha$ , alors notre équation peut se réécrire :

$$F^{Globale}(x) = \alpha F^{UsuelleTronquee}(x) + (1 - \alpha) G_{\gamma, u, \sigma}(x)$$

Afin de garder une clarté dans les écritures, nous notons par  $f^{UsuelleTronquee}$  la densité tronquée associée à la variable aléatoire  $X$  dont le support est  $[0, u]$ ,  $f^{GPD}$  la densité de la loi de Pareto généralisée de seuil  $u$ , donc dont le support est  $]u, +\infty[$ . Nous avons par dérivation :

$$f^{Globale} = \alpha f^{UsuelleTronque} + (1 - \alpha) f^{GPD}$$

$$\int_0^\infty f^{Globale} = \alpha \int_0^u f^{UsuelleTronque} + (1 - \alpha) \int_u^\infty f^{GPD} = \alpha + (1 - \alpha) = 1$$

Il est également possible d'exprimer la densité tronquée en fonction de la densité globale. En effet comme nous l'avons notifié dans le paragraphe dédié aux lois tronquées, sur  $[0, u]$ , les deux densités sont proportionnelles :

$$f^{UsuelleTronque} = \frac{1}{F^{Usuelle}(u) - F^{Usuelle}(s)} f^{Usuelle} \mathbb{1}_{[s, u]}$$

Comme nous sommes dans un cas où nos données sont uniquement tronquées à droite ,alors  $F^{Usuelle}(s) = 0$ , nous aurons alors :

$$f^{UsuelleTronque} = \frac{1}{F^{Usuelle}(u)} f^{Usuelle} \mathbb{1}_{[0, u]}$$

La densité globale dévient :

$$f^{Globale} = \alpha \frac{f^{Usuelle}}{F^{Usuelle}(u)} \mathbb{1}_{[0, u]} + (1 - \alpha) f^{GPD} \mathbb{1}_{]u, \infty[}$$

Nous rappelons que  $\alpha = \mathbb{P}(X \leq u) = F^{Usuelle}(u)$ ,  $\alpha$  peut s'interpréter comme un coefficient de pondération qui modélise la probabilité qu'un sinistre soit plus petit que le seuil  $u$  en utilisant la loi de probabilité modélisant la sinistralité attritionnel. Nous aboutissons donc à :

$$f^{Globale} = f^{Usuelle} \mathbb{1}_{[0, u]} + (1 - F^{Usuelle}(u)) f^{GPD} \mathbb{1}_{]u, \infty[}$$

Nous avons bien une densité car :

$$\int_0^\infty f^{Globale} dx = \int_0^u f^{Usuelle} dx + (1 - F^{Usuelle}(u)) \int_u^\infty f^{GPD} dx = F^{Usuelle}(u) + (1 - F^{Usuelle}(u)) = 1$$

Par intégration, nous obtenons la fonction de répartition  $F^{Globale}$  associée à la densité  $f^{Globale}$  :

$$\begin{aligned} F^{Globale}(x) &= \int_{-\infty}^x f^{Globale}(x) dx = \int_{-\infty}^x f^{Usuelle} \mathbb{1}_{[0, u]} dx + (1 - F^{Usuelle}(u)) \int_{-\infty}^x f^{GPD} \mathbb{1}_{]u, +\infty[} dx \\ &= F^{Usuelle}(\min(x, u)) + (1 - F^{Usuelle}(u)) F^{GPD}(x) \mathbb{1}_{]u, +\infty[} \end{aligned}$$

Soit donc :

$$F^{Globale}(x) = \begin{cases} F^{Usuelle}(x) & \text{si } x \leq u \\ F^{Usuelle}(u) + (1 - F^{Usuelle}(u)) F^{GPD}(x) & \text{si } x > u \end{cases}$$

Vérifions les propriétés de base d'une fonction de répartition.

$$— \lim_{x \rightarrow -\infty} F^{Globale} = \lim_{x \rightarrow -\infty} F^{Usuelle}(x) = 0.$$

$$— \lim_{x \rightarrow +\infty} F^{Globale} = F^{Usuelle}(u) + (1 - F^{Usuelle}(u)) \times \lim_{x \rightarrow +\infty} F^{GPD}(x) = 1$$

— La fonction  $F_{Globale}$  est croissante et continue sur son domaine de définition.

Ceci démontré, le modèle est clairement. Aussi il existe d'autres approches proposées par des mathématiciens pour construire ce type modèle à deux loi dans le cas où on est assujetti à des données extrêmes. Nous pensons particulièrement à l'approche qui permet d'utiliser un mélange d'Erlang et une loi de Pareto généralisée.

### Modèle à deux lois séparées

$$S = S_1 + S_2 = \sum_{i=1}^{N_1} X_i + \sum_{i=1}^{N_2} Y_i$$

La principale différence entre celui-ci et le modèle à deux lois intégrées est qu'il n'y a pas de construction d'une nouvelle densité par raccordement. En effet, nous modélisons en fréquence et en coût les deux types des sinistralités. Mathématiquement, la variable aléatoire  $X$  représente le coût des sinistres attritionnels. Elle sera modélisée par une loi de probabilité usuelle tronquée à droite de densité  $f_X$  à support  $[0, u]$ . nous notons  $F_X$  sa fonction de répartition. Nous rappelons que nous pouvons exprimer  $f_X$  en fonction de la densité usuelle globale :

$$f_X = \frac{1}{F_X^{Usuelle}(u)} f_X^{Usuelle} \mathbb{1}_{[0, u]}$$

Quant aux coûts excédent le seuil des extrêmes  $u$ , ils seront modélisés par la loi de probabilité de Pareto généralisée. Notons sa densité  $f_Y$  et  $F_Y$  sa fonction de répartition . Elle est définie sur le support  $]u, +\infty[$ .

Cette approche est particulièrement efficace lorsque nous disposons assez des données notamment dans la modélisation des deux fréquences.

### 3.4.10 Loi de la fréquence

Dans un portefeuille construit à partir des traités et des facultatives, l'horizon d'observation des données est généralement réduit à quelques années. Par conséquent, l'échantillon observé du nombre annuel des sinistres est petit. Aussi ces événements sont rares, les lois les plus utilisées pour modéliser les événements rares sont la loi de poisson et la loi binomiale négative. Que l'on cherche à modéliser le modèle à deux lois intégrées ou le modèle à deux lois séparées, la procédure d'ajustement de la loi de fréquence est la même. Le but est d'ajuster par une loi usuelle la variable aléatoire  $N$  représentant le nombre des sinistres. On aura recours au test de sur-dispersion liée à l'indice de dispersion de Fisher ainsi qu'à des représentations graphiques pour confirmer le choix de la distribution.

## Loi de poisson

La loi de Poisson est une loi de probabilité discrète qui décrit la probabilité qu'un événement se réalise durant un intervalle de temps donné, lorsque la probabilité de réalisation d'un événement est très faible et que le nombre d'essais est très grand.

Soit  $X$  une variable aléatoire suivant la loi de poisson. Le nombre moyen d'événements dans un intervalle de temps fixé est noté  $\lambda$ . On écrit  $X \sim \mathcal{P}(\lambda)$  avec  $\lambda > 0$ .

Alors la probabilité qu'il existe exactement  $k$  événements ( $k$  entier =  $1, 2, \dots$ ,) est :

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Les moments d'ordre 1 et 2 sont :

$$\mathbb{E}(X) = \lambda \quad \text{et} \quad \mathbb{V}(X) = \lambda$$

## Loi binomiale négative

Soit une expérience décrivant une situation de Bernoulli, une série de tirages indépendants, donnant un « succès » avec probabilité  $p$  (constante durant toute l'expérience) et un « échec » avec une probabilité complémentaire  $(1 - p)$ . Et on poursuit cette expérience jusqu'à l'obtention d'un nombre  $n$  de succès. Soit  $X$  une variable aléatoire représentant le nombre d'échecs avant l'obtention du nombre donné  $n$  de succès. On dit que  $X$  suit alors une loi binomiale négative. Ses paramètres sont  $n$ , le nombre de succès attendus, et  $p$ , la probabilité d'un succès. On écrit  $X \sim \mathcal{BN}(n, p)$  avec  $n \in \mathbb{N}^*$

$p \in [0, 1]$  et on note  $q = 1 - p$ , soit  $k$  une valeur sur le support  $\mathbb{N}$ , la fonction de masse de probabilité est donnée par :

$$P(X = k) = \binom{k+n-1}{k} p^n q^k \quad \text{Où} \quad \binom{k+n-1}{k} \text{ est le coefficient binomial.}$$

Les moments de premier ordre et de deuxième ordre sont :

$$\mathbb{E}(X) = \frac{nq}{p} \quad \text{et} \quad \mathbb{V}(X) = \frac{nq}{p^2}$$

La loi binomiale négative est connue comme une alternative intéressante à la loi de poisson. Elle est particulièrement utile pour des données discrètes, à valeurs dans un ensemble positif non-borné, dont la variance empirique excède la moyenne empirique. Si une Poisson est utilisée pour modéliser de telles données, la moyenne et la variance doivent être égales. Dans ce cas, les observations sont « sur-dispersées » par rapport au modèle Poisson. Puisque la loi binomiale négative possède un paramètre supplémentaire, il peut être utilisé pour ajuster la variance indépendamment de la moyenne.

**On utilisera la méthode de maximum de vraisemblance pour estimer les paramètres**

## Test de sur-dispersion liée à l'indice de dispersion de Fisher

Le test de sur dispersion comme son nom l'indique est un test qui permet de mettre en évidence une sur-dispersion ou non des données par rapport à la loi les modélisant. Pour une variable aléatoire discrète  $N$ , la loi de la fréquence. Les hypothèses du test sont :

$$\begin{cases} H_0 : \mathbb{E}(N) = \mathbb{V}(N) & \text{(Absence de sur dispersion)} \\ H_1 : \mathbb{E}(N) > \mathbb{V}(N) & \text{(Sur dispersion)} \end{cases}$$

La statistique  $T$  pour estimer l'indice de dispersion  $\frac{\mathbb{V}(N)}{\mathbb{E}(N)}$  est donnée par :

$$T = \frac{(n-1)\bar{S}^2}{\bar{X}}$$

Avec et respectivement  $\bar{X}$  et  $\bar{S}$  la moyenne et la variance empirique de la loi de fréquence.

Sous l'hypothèse, Hoel (1943) a montré que la statistique  $IF = nT$ , connue sous le nom d'indice de dispersion de Fisher, est asymptotiquement distribuée comme une loi de khi2 à  $(n - 1)$  degrés de liberté avec  $n$  le nombre d'observations.

Comme nous l'avons énoncé précédemment si la loi de poisson est utilisée pour modéliser une place de données à valeurs dans un ensemble positif non-borné, dont la variance empirique excède la moyenne empirique, la moyenne et la variance doivent être égales. Nous avons expliqué que la loi binomiale serait plus adaptée dans ce cas -ci. Ainsi ce test nous permettra de choisir la plus adaptée de ces deux lois.

## Test du rapport de vraisemblance

Le test de rapport de vraisemblance permet de tester si l'échantillon de  $N$  suit une distribution donnée contre une distribution alternative. Dans cette étude, on s'intéressera à tester l'hypothèse suivante :

$$\begin{cases} H_0 : N \text{ suit une loi de poisson} \\ H_1 : N \text{ suit une loi binomiale} \end{cases}$$

La statistique du rapport de vraisemblance est donnée par :

$$T = 2 (\ln(L_1) - \ln(L_0))$$

avec :

- $L_0$  : La vraisemblance de la loi de poisson
- $L_1$  : La vraisemblance de la binomiale négative.

La statistique  $T$  tend asymptotiquement en loi vers une loi du Khi-deux à un degré de liberté.

### Test d'adéquation de Khi-deux

On observe un échantillon de données  $y_1, \dots, y_N$  d'une variable aléatoire  $Y$  qui prend un nombre fini  $J$  de valeurs. On veut tester  $H_0$  : les probabilités que  $Y$  prenne les valeurs 1 à  $J$  sont respectivement  $p_1, \dots, p_J$  avec  $\sum_{j=1}^J p_j = 1$ . On appelle  $\hat{p}_j$  la probabilité empirique que  $Y$  prenne la valeur  $j$ , c'est-à-dire le nombre d'observations qui prennent la valeur  $j$  dans l'échantillon divisé par le nombre total d'observations  $N$  :

$$\hat{p}_j = \frac{1}{N} \sum_{i=1}^N \mathbb{1}(y_i = j)$$

On peut alors définir la statistique du

$$\chi^2 : T = \sum_{j=1}^J \frac{(N \hat{p}_j - N p_j)^2}{N p_j}$$

Sous  $H_0$ , cette statistique suit asymptotiquement une loi du  $\chi^2$  à  $J - 1$  degrés de liberté. On construit alors un test de niveau  $\alpha$  en rejetant  $H_0$  lorsque  $T \geq F_{\chi^2(J-1)}^{-1}(1 - \alpha)$  (avec  $F_{\chi^2(J-1)}^{-1}(1 - \alpha)$  le quantile d'ordre  $1 - \alpha$  de la loi du  $\chi^2$  à  $J - 1$  degrés de liberté). En terme de p-value, on rejette  $H_0$  lorsque p-value  $\leq 5\%$ .

### 3.5 Application sur des données réels

Nous allons mettre en application les méthodes décrites précédemment sur des données réelles. Nous travaillerons sur les données de sinistralité du portefeuille Non-Marin de la SCR. Spécifiquement sur la sinistralité se rapportant aux traités ainsi qu'aux facultatives de la branche incendie de la SCR sur le marché marocain(national) et sur le marché international. Ci dessous, un échantillon présentant les données selon les marchés marocains.

Marché Marocain		
CONTRACT NUMBER	YEAR	Incurred Loss
0001182363	2012	261 979
0000450268	2012	933 540
0200006010	2012	2 706 347
0200014510	2013	29 686
0000937846	2013	377 222
0200014510	2014	140 733
0000450359	2014	225 305
0000450268	2015	1 895 523

TABLE 3.2 – Échantillon des pertes du lob incendie sur le marché marocain

<b>Marché International</b>		
CONTRACT NUMBER	YEAR	Incurred Loss
0300161190	2012	1 231 245
0000775684	2012	3 050
0001426243	2012	6 704
0001335516	2012	748 665
0001488032	2013	4 750,10
0001583218	2014	2 553,00
0001540175	2014	9 585,94
0001075529	2013	53 512

TABLE 3.3 – Échantillon des pertes du lob incendie sur le marché international

### 3.5.1 Statistiques As-If de nos données

Comme nous l'avons notifié dans le paragraphe théorique, les statistiques As-If sont la première étape pour évaluer des données qui représentent des quantités monétaires. En effet des nombreux facteurs dont l'inflation, les changements réglementaires ainsi que la politique de souscription, font que la valeur de l'argent change dans le temps.

**Dans le cas de la SCR, la réglementation ainsi que la politique de souscription sont stables sur la plage d'années d'études.**

la revalorisation as-if est appliquée à des sinistres survenus entre 2012 et 2022. Le taux d'inflation utilisé est l'indice d'évolution du prix à la consommation. les évolutions ont été récupérées à partir du site internet [donneesmondiales.com](http://donneesmondiales.com)

Année	Evolution	Indice	Facteur de redressement
2012		100	1,123967819
2013	0,0129	101,29	1,109653291
2014	0,0188	103,194252	1,089176768
2015	0,0044	103,6483067	1,084405385
2016	0,0156	105,2652203	1,067748508
2017	0,0164	106,9915699	1,05051998
2018	0,0075	107,7940067	1,042699732
2019	0,018	109,7342988	1,024262998
2020	0,003	110,0635017	1,0211994
2021	0,0071	110,8449526	1,014
2022	0,014	112,3967819	1

TABLE 3.4 – Table de revalorisation par inflation

Les sinistres sont revalorisés à partir du facteur de redressement ci-dessus calculé. Les sinistres les plus anciens ont subis une augmentation de près de 11%. Ci-dessous la mise as-if de l'échantillon

précédemment présenté.

Marché Marocain			
CONTRACT NUMBER	YEAR	Incurred Loss	Revalue Incurred Loss
0001182363	2012	261 979	294 456
0000450268	2012	933 540	1 049 269
0200013020	2012	339 473	381 557
0200006010	2012	2 706 347	3 041 847
0200014510	2013	29 686	32 941
0000937846	2013	377 222	418 586
0200006010	2013	5 595 903	6 209 513
0200014510	2014	140 733	153 283
0001468649	2014	70 758	77 068
0000450359	2014	225 305	245 397
0000450268	2015	1 895 523	2 055 515

TABLE 3.5 – Revalorisation des pertes sur le marché Marocain

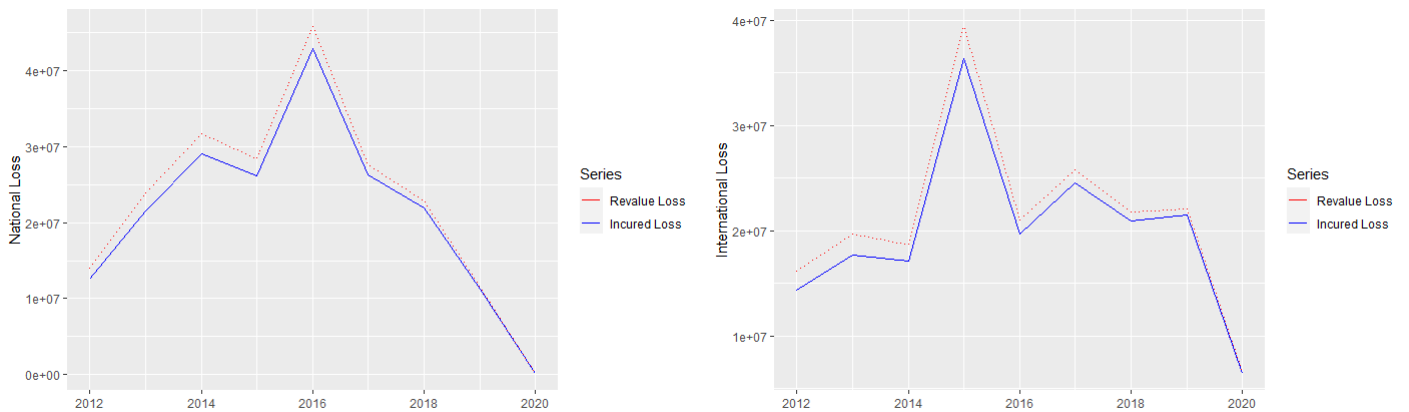


FIGURE 3.7 – Comparaison des pertes après revalorisation selon les marchés

Ces deux graphiques nous permettent de mettre en évidence que le Maroc n'a pas subi de déflation et à subi une inflation de taux de croissance plus ou moins stable depuis 2012. Notamment depuis 2019, l'inflation est assez faible avec un taux d'évolution petit n'influençant pas énormément la sinistralité de ces dernières années.

Maintenant que la base de nos sinistres est ramenée à la même référence, nous pouvons procéder à l'étude quantitative.

### 3.5.2 Statistiques descriptives

Les statistiques descriptives ont pour but de décrire la structure des données. Il est en effet nécessaire pour un meilleur traitement des données de se familiariser avec les données. Comme

notifié précédemment, nos données d'étude sont disposé selon deux marchés, le marché domestique(marocain) et le marché international.

Données brutes - Marché Marocain	
Statistique	Valeur
Minimum	0
Premier quartile	4 588
Médiane	38 058
Moyenne	872 007
Troisième quartile	315 537
Maximum	42 816 574

Données Marché-International	
Statistique	Valeur
Minimum	0
Premier quartile	1 245
Médiane	7 640
Moyenne	115 141
Troisième quartile	54 375
Maximum	7 092 818

TABLE 3.6 – Statistiques descriptives de la sinistralité Incendie

La première chose qui frappe sur les deux marchés est la dispersion des montants de pertes. En effet, le sinistre minimum est de l'ordre de 0,1... arrondie à zero, très petit, alors que nous observons une perte toute chose étant égale par ailleurs de l'ordre de 42.816.57 DH sur la marché marocain et 7.092.818 DH sur le marché international. Aussi sur les deux marchés, la médiane est nettement inférieur à la moyenne. Ceci décrit parfaitement la caractéristique d'un portefeuille de réassurance, beaucoup de petits sinistres et quelques sinistres très importants.

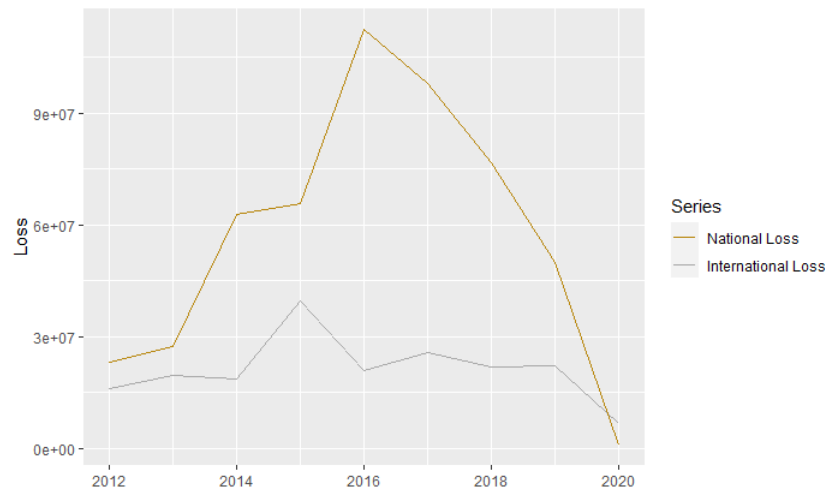


FIGURE 3.8 – Comparaison des pertes

Nous avons tracé ce graphique afin de comparer les pertes sur nos deux marchés. La différence est très illustrée. Les pertes sur le marché national sont largement supérieures et nettement plus conséquentes que sur le marché international. Ceci est principalement dû à la grande majorité des contrats facultatives sur le marché marocain. Le portefeuille est clairement non homogène, il est nécessaire de les étudiés séparément. Dans la suite de notre travail, notre méthode de modélisation sera appliquée aux deux marchés.

Une caractéristique importante de ce portefeuille est que La Société Centrale de Réassurance applique sur celui-ci un taux cession  $\alpha$  confidentiel pour les affaires facultatives marocaines. Ainsi, nous avons multiplié par  $(1 - \alpha_i)$  tous les sinistres venant des contrats facultatives marocaines. La suite de notre travail se fera sur la base de ces sinistres nettes pour en effet se mettre dans la situation réelle du portefeuille. Ci-dessous, une statistique descriptive après application du taux de cession retenu des affaires facultatives marocaines.

<b>Données nettes - Marché Marocain</b>	
Statistique	Valeur
Minimum	0
Premier quartile	1 126
Médiane	10 980
Moyenne	347 174
Troisième quartile	129 967
Maximum	9 455 775

TABLE 3.7 – Statistiques descriptives - Marché marocain

La grande étendue des valeurs est toujours présente, cela dit le portefeuille est devenu plus homogène selon les deux marchés. Dans la suite et pour des raisons d’efficacité et d’optimisation discutées avec l’encadrant, nous avons décidé de garder la segmentation. En effet dans la pratique, la couverture de rétrocession qui couvre ce portefeuille est achetée selon que le marché est national ou international.

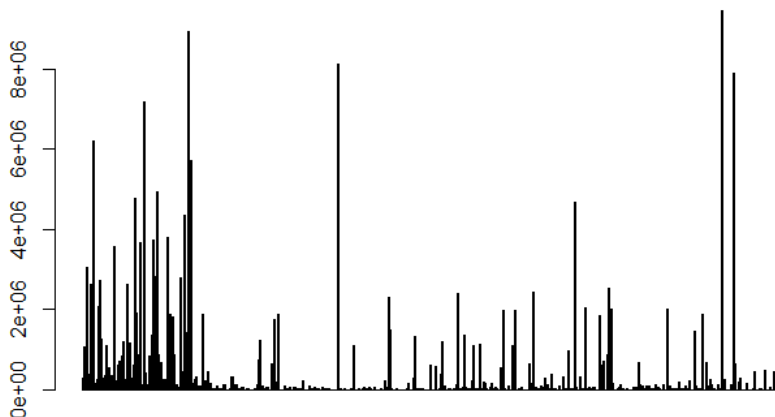


FIGURE 3.9 – Sinistres - Marché national

Ce graphique en bar ci-dessus représentant les sinistres nous permet de mettre en évidence les sinistres importants et leur nombre d’occurrence. Il nous permet également de voir s’il y a un regroupement des sinistres importants ou une tendance particulière dans leur fréquence d’apparition, ce qui remettrait en cause notre hypothèse d’indépendance. Nous remarquons sur ce marché

que les sinistres étaient plus sévères et plus fréquents dans les années 2012-2014, après, ils ont une tendance plus ou moins aléatoire avec une fréquence plus faible. Il est clair qu'aucune tendance de regroupement ne mise en évidence sur ce marché.

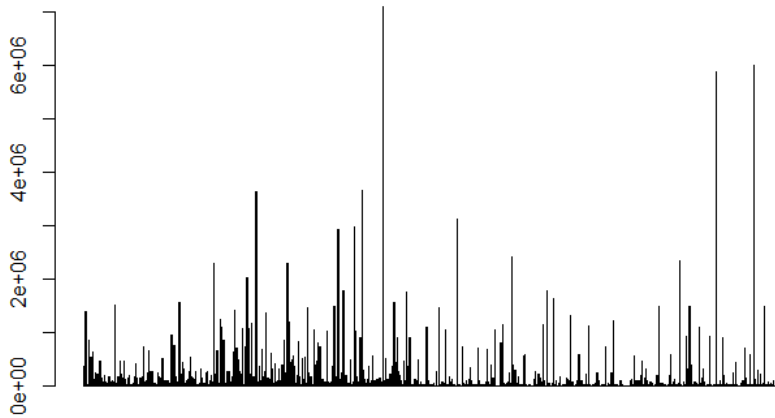


FIGURE 3.10 – Sinistres-Marché international

Sur le marché international, les montants extrêmes de ce portefeuille semble moins fréquents que sur le marché domestique, mais les sinistres faibles sont plus denses. De même, nous ne remarquons aucune tendance particulière dans l'apparition des sinistres qu'ils soient attritionnels ou extrêmes.

Année	Nombre de sinistres - Marché Marocain	Nombre de sinistres - Marché international
2000	63	83
2001	65	73
2002	56	69
2003	58	56
2004	58	97
2005	58	135
2006	58	205
2007	58	286
2008	56	261
2009	56	260
2010	58	249
2011	58	271
2012	62	249
2013	64	86
2014	54	105

TABLE 3.8 – Échantillon-Nombre de sinistres par an

Nous disposons d'une plage de nombre de sinistres par année de 1990 à 2021. Les nombres sont

aussi disposées selon les marchés.

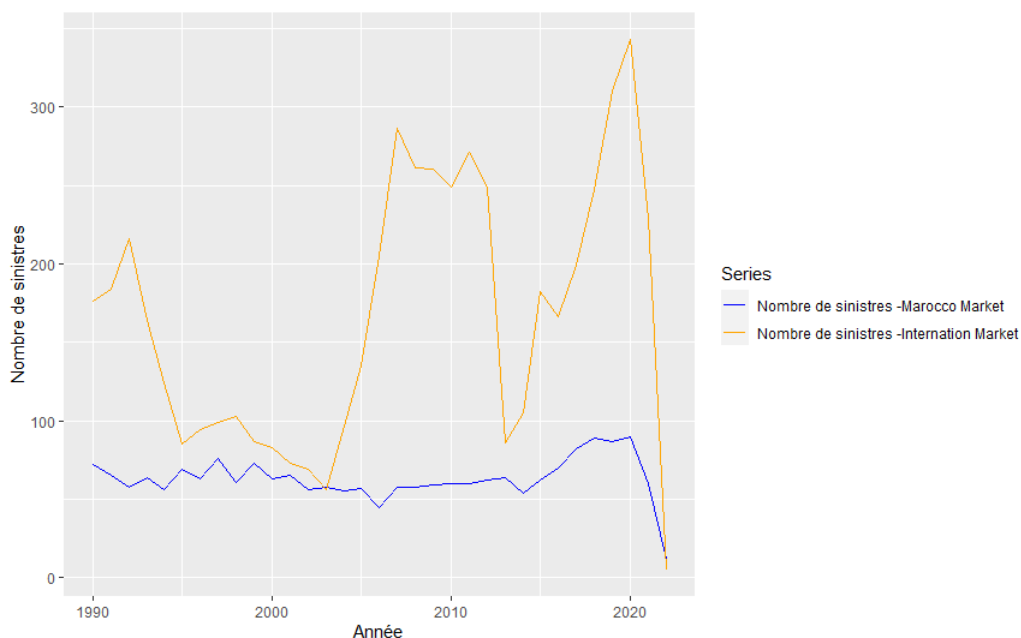


FIGURE 3.11 – Courbes-Comparaison du nombre de sinistres par an selon les marchés

Sur la courbe tracée ci-dessus, nous observons une nette différence à partir des années 2005 entre le nombre de sinistres international des sinistres et le nombre de sinistres national. Sur le marché international, les nombres sont largement supérieurs.

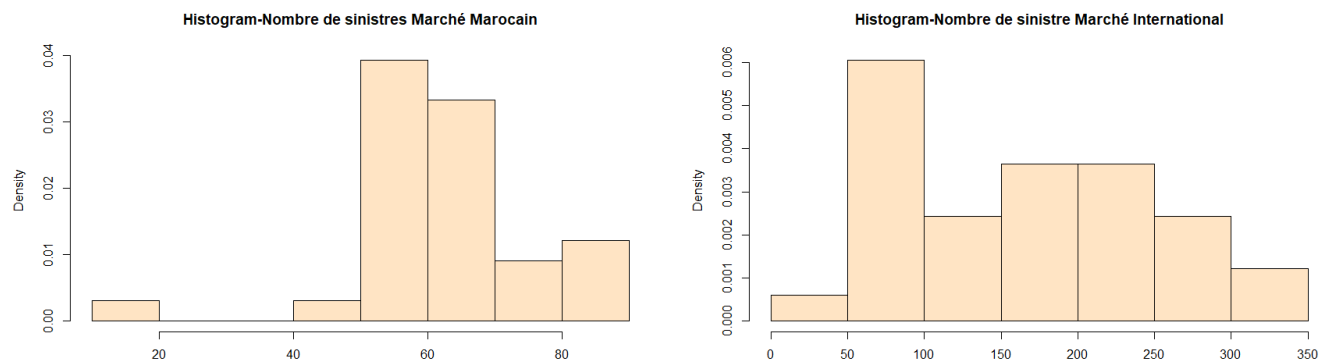


FIGURE 3.12 – Histogramme de densité

Pour rappel, on définit la densité comme étant la proportion d’agents statistiques par unité de mesure. Sur le marché marocain, la densité selon les intervalles du nombre de sinistres est plus ou moins bien répartie. Sur l’intervalle  $[0, 40]$  sa densité est quasi nulle, elle est existante à partir de 50 avec une forte densité entre  $[50, 70]$ . Cependant sur la marché international, la densité est plus ou moins bien répartie, nous observons spécialement une forte densité entre  $[50, 100]$

Dans la suite nous cherchons à ajuster nos données de fréquence comme de coût par les méthodes décrites au 2.7 et particulièrement au 2.7.3. Nous commencerons par choisir le seuil à partir duquel nous pouvons considérer nos montants de sinistres comme extrêmes..

*Hypothèse de modélisation : Nous supposons que le portefeuille est constant sur l'horizon de l'analyse.*

### 3.5.3 Choix du seuil

#### A. Détermination du domaine d'attraction

Tout d'abord le domaine d'attraction. la détermination du domaine d'attraction est une étape importante car l'intérêt de construire un modèle à deux lois est justifié lorsque la queue de distribution des données est épaisse, auquel cas l'indice des valeurs extrêmes est strictement positif(Domaine de Fréchet), et négatif ou nul dans les cas contraires. Aussi pour certain outil, la sélection du seuil des extrêmes est valide uniquement pour certaines valeurs de l'indice des valeurs extrêmes comme par exemple le Hill-plot, l'outil le plus utilisé. Afin de statuer, nous traçons le quantile plot généralisé(*présenté la partie théorique choix du seuil*)

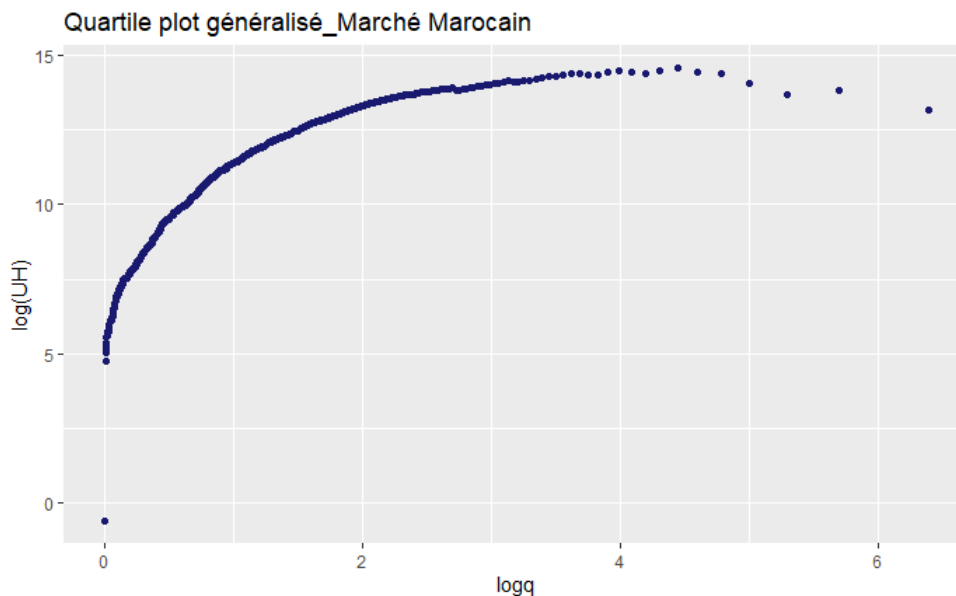


FIGURE 3.13 – Quartile plot généralisé - Marché Marocain

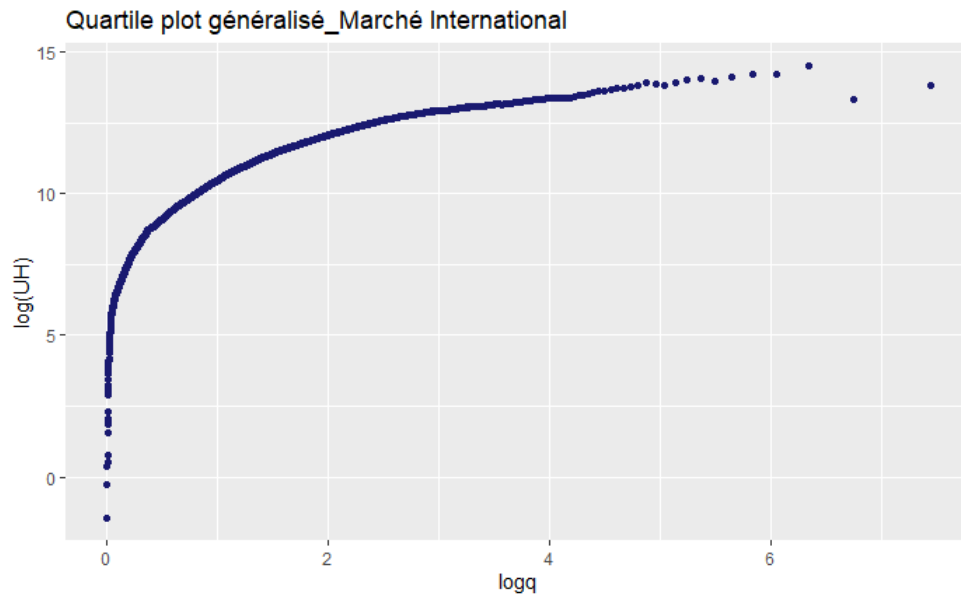


FIGURE 3.14 – Quartile plot généralisé des coûts - Marché International

La courbe 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.

Nous utiliserons différents outils graphiques et numériques décrits issus de la théorie des valeurs extrêmes valables dans ce domaine afin de sélectionner le 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 mélangeant deux lois.

## B. Hill-Plot et Pickands-Plot

Le Hill-plot ainsi que le pickands-Plot sont valables dans le domaine d'attraction de Fréchet. Nous avons tracé le Hill plot en fonction du seuil  $u$  et du nombre d'excès  $k$  explicités en 2.5. l'estimateur de l'indice des valeurs extrêmes est sur un intervalle de confiance à 95%. Ci-dessous les graphes réalisés sur R. 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 et de même pour le Pickand's Plot.

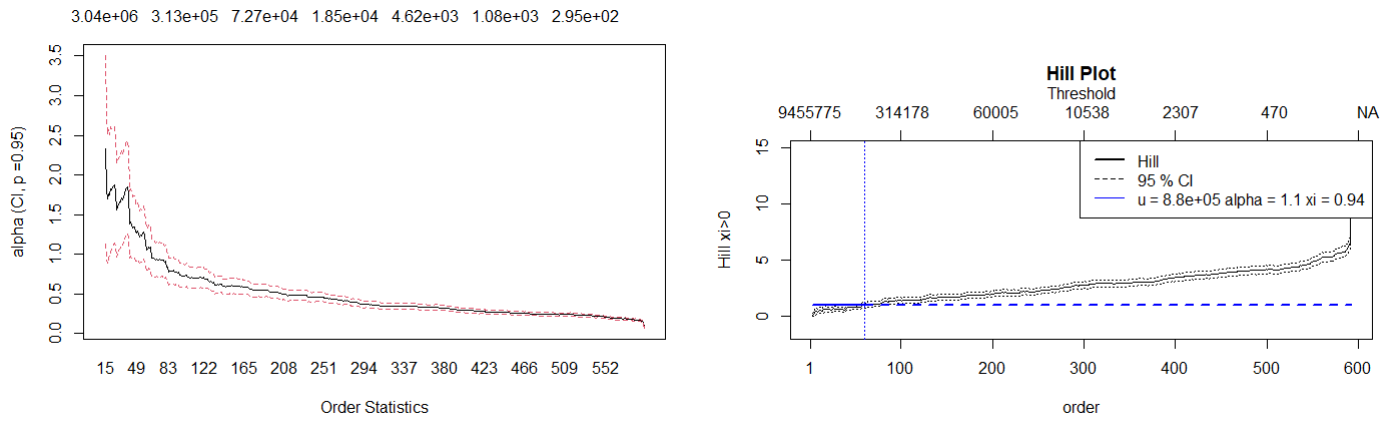


FIGURE 3.15 – Hill Plot Marché Marocain

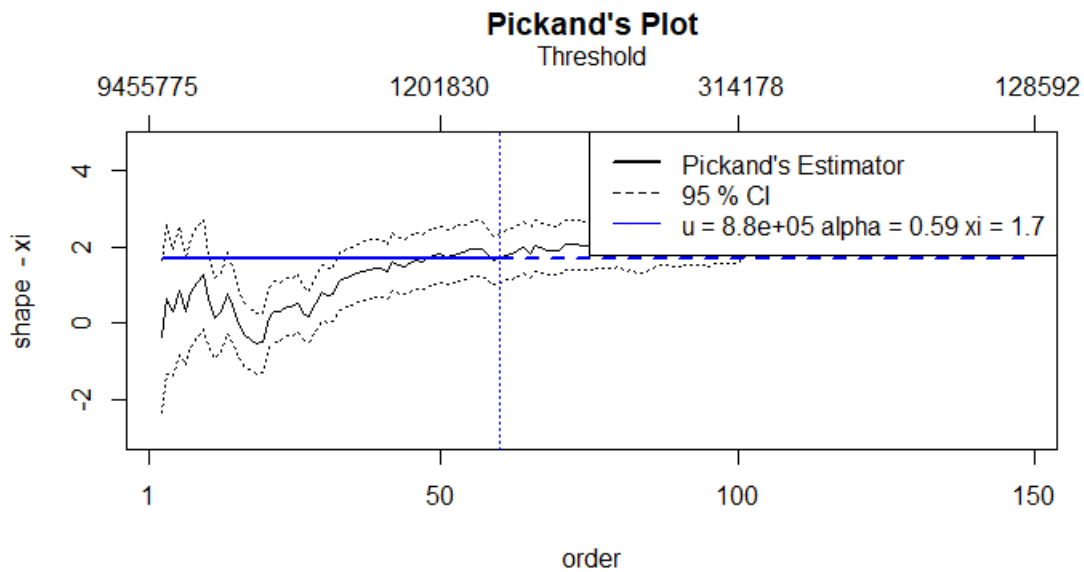


FIGURE 3.16 – Pickands-Plot Marché Marocain

Sur le **Marché Marocain**, qu'il s'agisse du Hill-plot ou du Pickand's-Plot nous voyons clairement le moment où l'estimateur de l'indice des valeurs extrêmes  $\gamma$  commence à être stable. Autrement dit le point à partir duquel nos données peuvent être ajustées par la loi de Pareto généralisée. Les deux outils graphiques nous suggère le même **seuil de 880 000 DH**. L'actuaire peut ré-choisir un autre seuil sur le graphique pour des raisons d'efficacité de la modélisation remarquées.

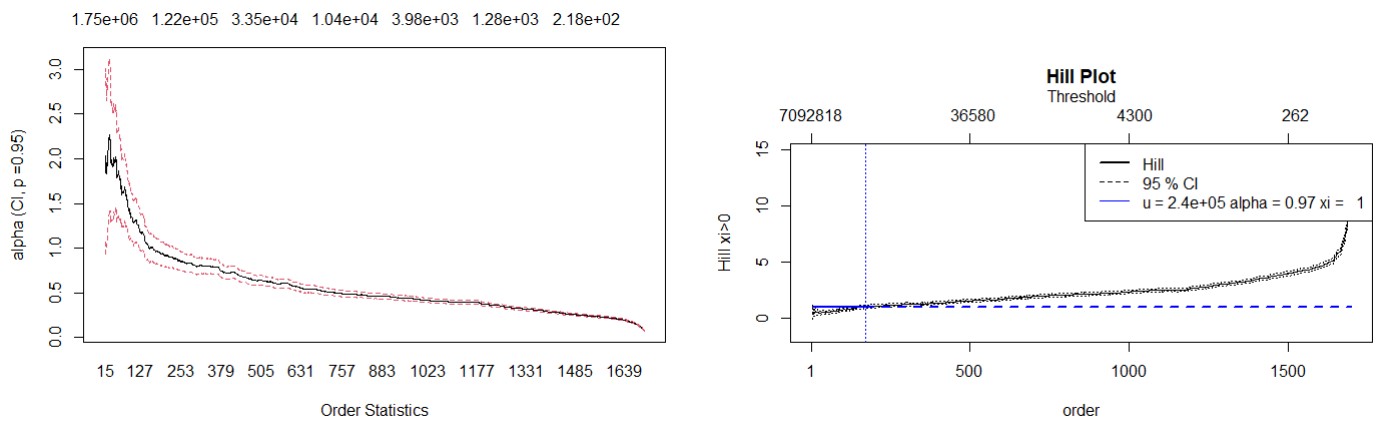


FIGURE 3.17 – Hill Plot Marché International

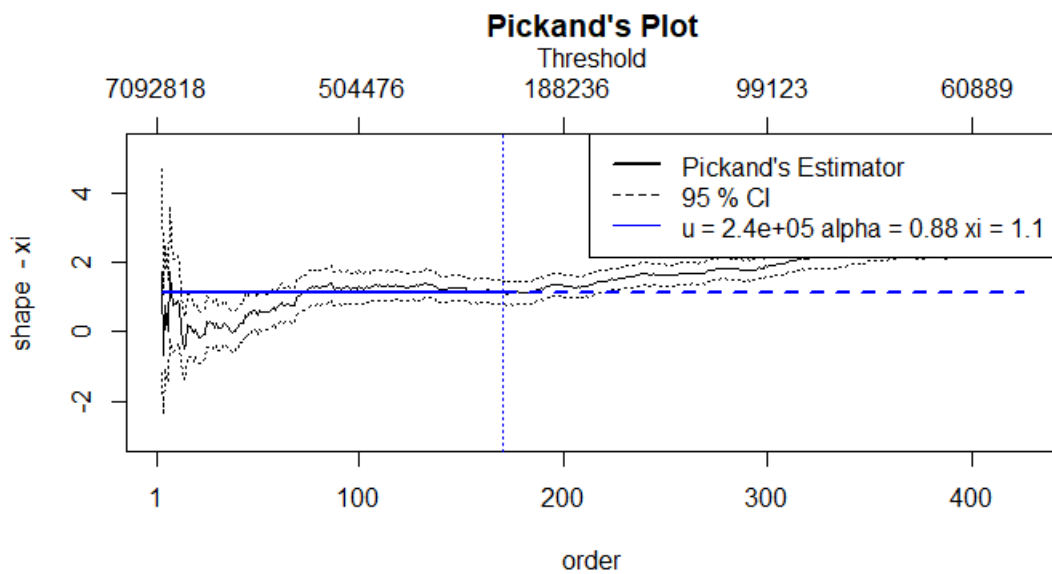


FIGURE 3.18 – Pickands-Plot Marché International

En dupliquant la même logique sur le **Marché International**, le seuil retenu est de **240.000 DH**

### C. Minimizing the AMSE of the Hill estimator (dAMSE)

Cet outil repose sur la minimisation de l'erreur quadratique de l'estimateur de Hill. L'avantage principal de celui-ci est le fait qu'il soit analytique.

### Marché Marocain

```
## $second.order.par : [1] 0.9830785 -0.6715868
##
## $k0 : [1] 53
##
## $threshold : [1] 1110264
##
## $tail.index : [1] 1.210993
```

**Seuil sur le marché marocain = 1 110 264**

### Marché International

```
## $second.order.par : [1] 1.0055600 -0.7720594
##
## $k0 : [1] 120
##
## $threshold : [1] 413152.7
##
## $tail.index : [1] 1.180968
```

**Seuil sur le marché international = 413 153**

	Hill & Pickand's Plot	dAMSE
Marché Marocain	880 000	1 110 265
Marché International	240 000	413 153

TABLE 3.9 – Seuils

En pratique le choix d'un seuil est toujours subjectif, les outils ne donnent pas en général des seuils quasi proches. Ces outils sont des supports d'aide à la décision. Aussi un paramètre important est la quantité de données qui restent après l'application de celui-ci. Ce dernier paramètre nous a poussé à sélectionner pour la suite de notre étude le seuil donné par le graphique de Hill. En appliquant ce seuil à nos données, il s'illustre que 90% de sinistres sont attritionnels contre 10% de sinistres extrêmes.

### 3.5.4 Modélisation de la sévérité

Maintenant que notre seuil des extrêmes est déterminé, nous pouvons modéliser la sinistralité attritionnelle à l'aide d'une loi usuelle tronquée, ainsi que la sinistralité extrême, à l'aide d'une loi de Pareto généralisée. Le but de ce paragraphe est de fournir un ajustement adéquat du montant de la sinistralité. Notre portefeuille est en effet segmenté pour des raisons énoncées précédemment en deux marchés, cela dit nous présenterons l'exemple de la modélisation des lois utilisées sur le marché marocain. Pour des raisons d'efficacité afin d'avoir un rapport concis, la même méthodologie est suivie pour le marché international.

## Marché Marocain

## A. Analyse préliminaire des distributions candidates

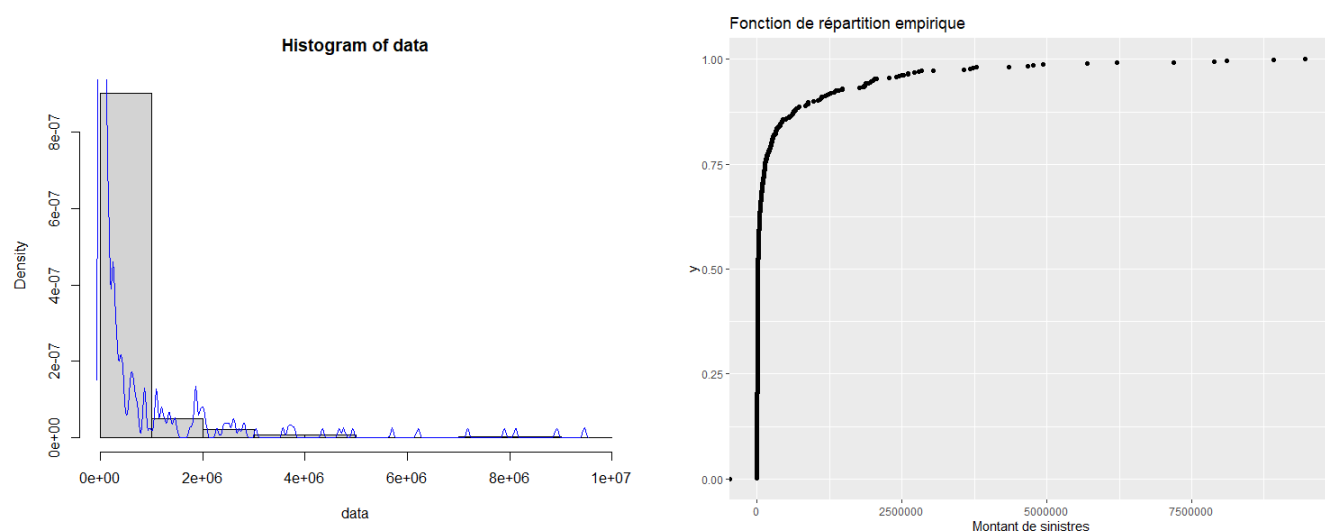


FIGURE 3.19 – Histogramme de densité et graphe cumulative de distribution\_MA

Ces graphiques nous suggèrent d'éliminer certains lois de probabilité comme la loi normale, uniforme, logistique dont les allures sont classiquement connues de tous.

## B. Coefficients d'asymétries et Graphe de Cullen et Frey

Coefficient d'asymétrie(Skewness)	Coefficient de pointicité(Kurtosis)
5,165	34,69

TABLE 3.10 – coefficients d'asymétrie et de pointicité\_MA

Le coefficient d'asymétrie Skewness est supérieur à zéro et le kurtosis est largement supérieur à 3. Cela traduit une distribution asymétrique positive leptokurtique, soit, une distribution de probabilité avec une lourde queue de distribution à droite.

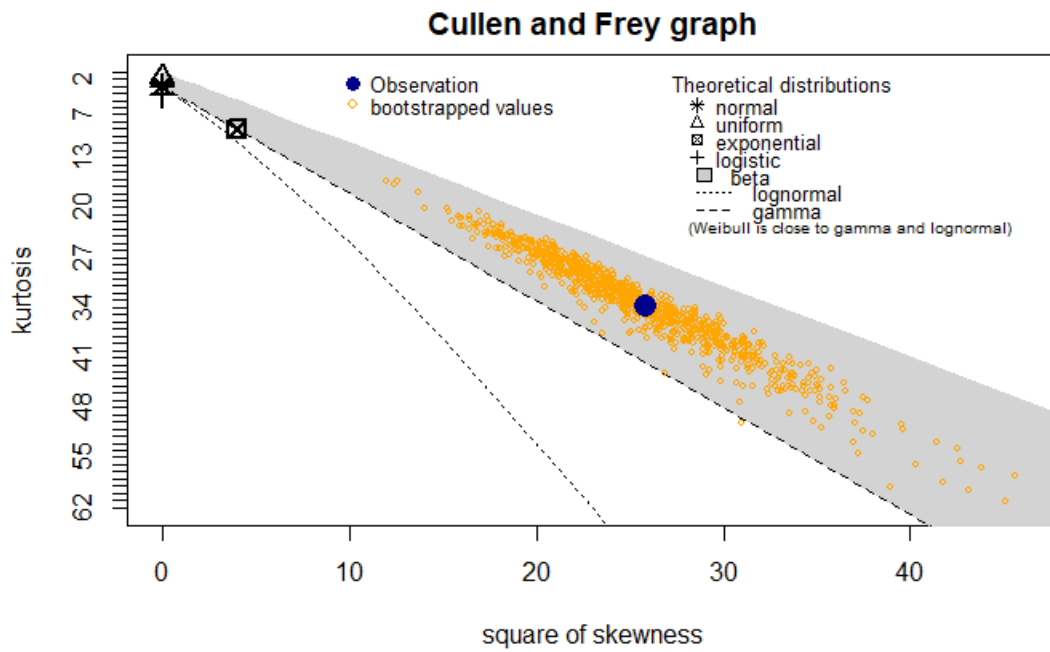


FIGURE 3.20 – Graphe de Cullen et Frey avec bootstrap<sub>MA</sub>

Le graphe Cullen and Frey ci-dessus ne donne pas d'informations précises. L'analyse de cet indicateur nous suggère principalement la loi gamma, de weibull et la loi log-normale et éventuellement la loi exponentielle. Ceci dit, ce n'est pas un outil très puissant. Ainsi pour plus de pertinence, traçons le QQ-Plot ces lois afin d'avoir des informations plus précises.

## C. QQ-Plot

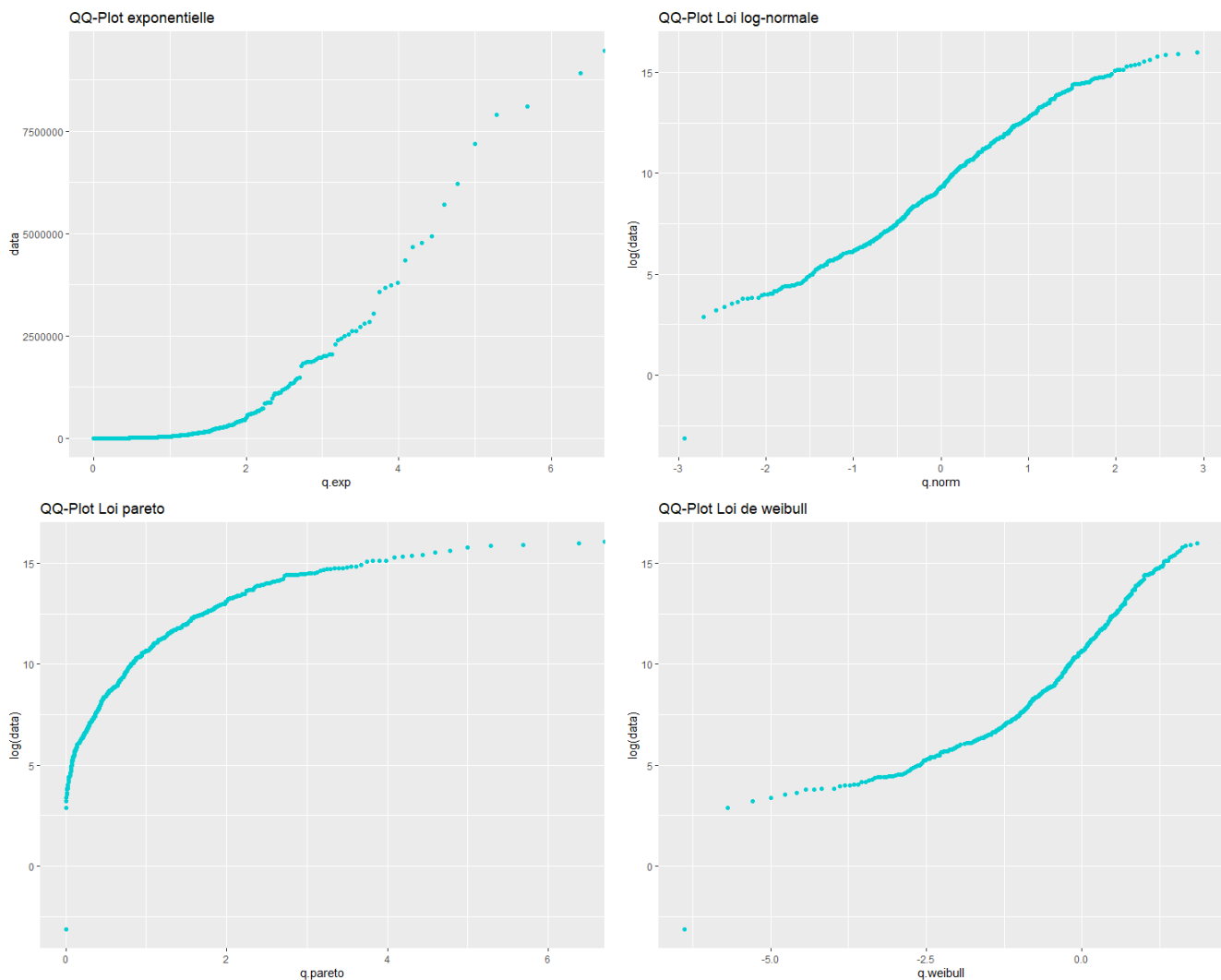


FIGURE 3.21 – QQ-Plot selon les lois usuelles

Alors, nous notons que la loi exponentielle et la loi de Pareto ne sont pas adaptées pour nos données. Leurs allures de courbe nous montrent que nos données ont une queue de distribution plus lourde que la loi exponentielle et moins lourde que la loi de Pareto. Les seules qui y ressemblent à une droite sont la loi log-normale et loi de Weibull. La loi log-normale est clairement la plus adaptée. Ceci dit, elle n'est pas parfaitement adaptée, cela nous confirme l'intérêt de faire appel à la modélisation à deux lois.

## D. Estimation

Loi usuelle	Test de Kolmogorov-Smirnov	Test d'Anderson-Darling
Log-normale	Acceptation ( $D = 0.045$ , <b>P-value</b> = 0.20)	Acceptation ( $A = 1.9574$ , <b>P-value</b> = 0.0969)
Gamma	Rejet	Rejet
Weibull	Rejet	Rejet
Pareto	Rejet	Rejet

Après estimation par la méthode de maximum de vraisemblance, la seule loi qui a été acceptée à la fois par le test de Kolmogorov-Smirnov et d'Anderson-Darling pour nos données est la loi log-normale. Nos données peuvent donc être modélisées par cette dernière. Ces paramètres sont :

MeanLog	SdLog
9,526	3,073

TABLE 3.11 – Paramètres de la loi log-normale

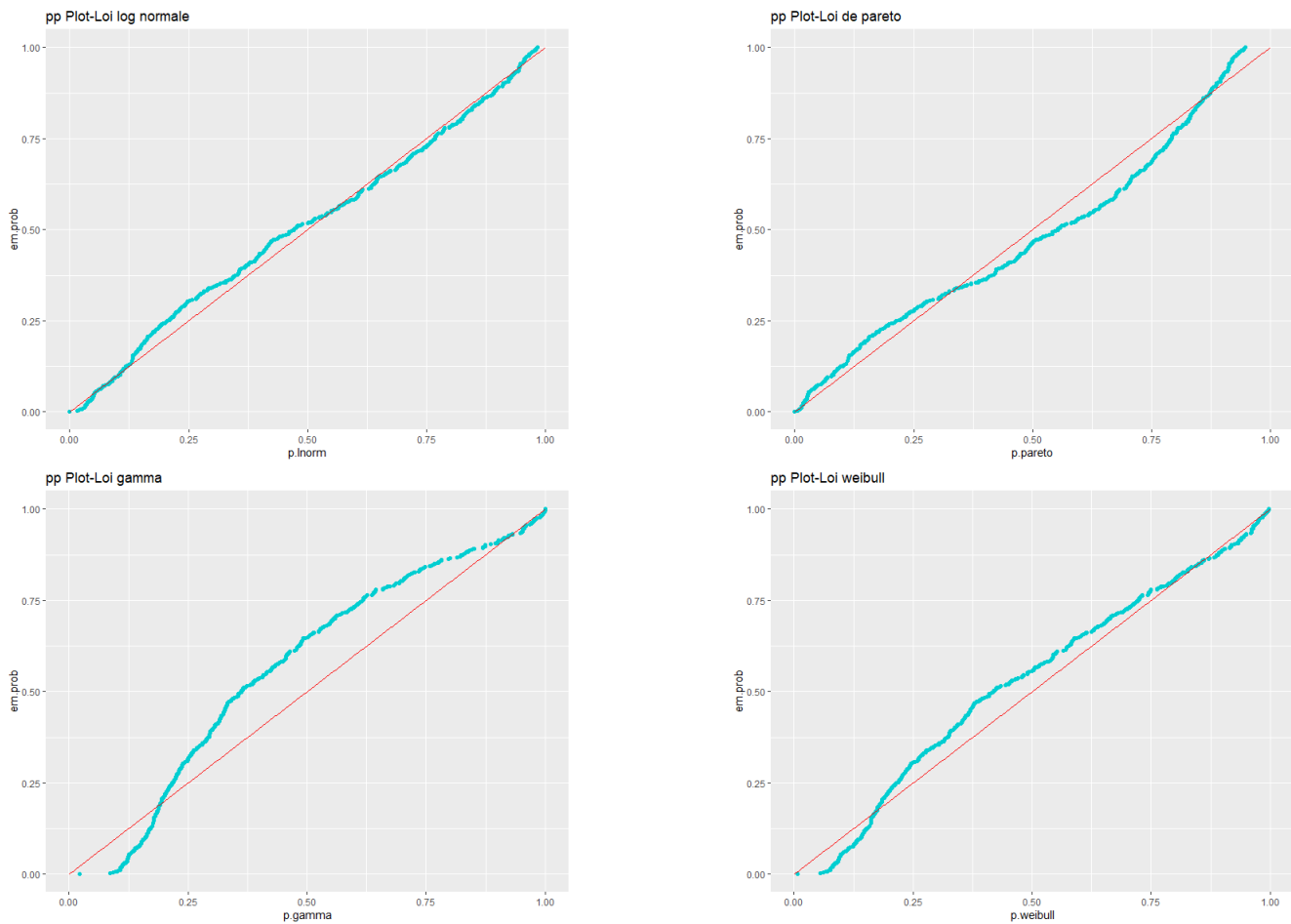


FIGURE 3.22 – PP-Plot selon les lois

Le PP-plot qui est une courbe tracée entre la fonction de répartition empirique et la fonction de répartition théorique avec les paramètres estimés, montre clairement que la loi log-normale estimée est adaptée, elle suit bien la première bissectrice.

Dans la suite, nous traçons une courbe entre la fonction de répartition de notre loi théorique et la fonction de répartition empirique sur les données au dessus du seuil afin de juger son adéquation dans les valeurs extrêmes.

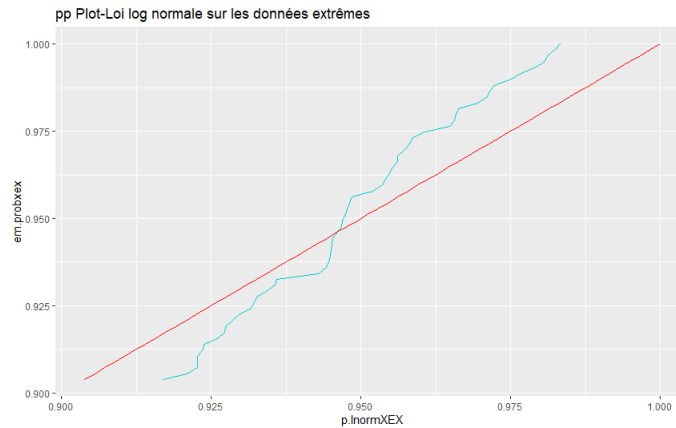


FIGURE 3.23 – Comparaison entre la fonction de répartition empirique des données extrêmes et la fonction de répartition théorique de la loi log-normale

Ce graphique confirme que la loi log-normale de paramètres décrits plus haut n’est pas très bien adaptée. En effet comme nous le voyons sur la figure, les grandes valeurs ne sont pas bien ajustés par cette loi, Si l’actuaire conserve cette modélisation, le coût à charge pour les tranches hautes sera quasiment nul et cela n’est pas approprié. L’ajustement par le modèle à deux lois pallie ce problème. Dans notre modèle, on tronque cette dernière au seuil de telle sorte qu’elle ajuste la partie attritionnelle (Usuelle) des nos montants. Connaissant notre seuil  $u$ , estimons alors les paramètres de la loi de Pareto généralisée(GPD).

Paramètres de la loi de Pareto généralisée		
$u(\text{seuil})$	$\gamma$	$\sigma$
880.000	0,0827	1.795.547

TABLE 3.12 – Paramètres Loi GPD

Dans toute la suite nous considérons :

- La fonction de répartition **théorique en rouge**
- La fonction de répartition **empirique en noire**

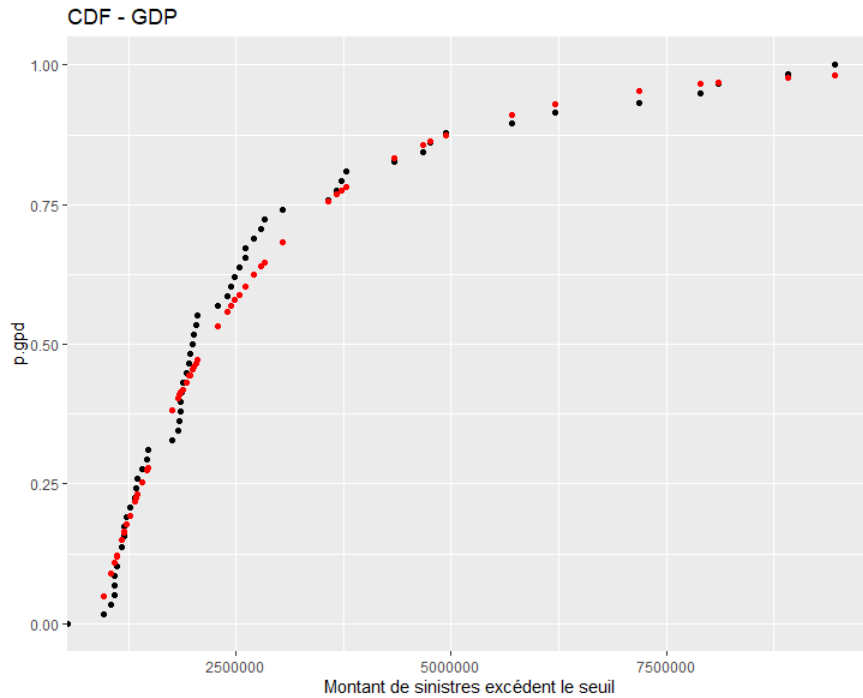


FIGURE 3.24 – Modélisation des sinistres supérieurs au seuil des extrêmes

Le graphique ci-dessus est la représentation de la fonction de répartition empirique des montants de sinistres excédant le seuil  $u$  et de la fonction de répartition théorique de la GDP précédemment estimée. L'adéquation est retenue.

### 3.5.5 Modélisation de la loi de fréquence

*Hypothèse de modélisation : Nous supposons que le portefeuille est constant sur l'horizon de l'analyse.*

Nous rappelons que les données utilisées sont ceux du marché Marocain. Les résultats qui suivent, décrits par la méthodologie théorique précédemment présentée peuvent être facilement retrouvés pour le marché international.

Nous rappelons que nous de testerons les deux lois statistiques classiques utilisées, pour ajuster le nombre de sinistres :

- Une loi de Poisson
- Une loi Binomiale négative

Nous estimons les paramètres de nos lois par le maximum de vraisemblance et nous réalisons des graphiques ainsi des tests d'adéquation pour juger l'ajustement.

Lorsqu'on modélise le nombre de sinistres, il est important de prendre en considération l'exposition du portefeuille. Car au cours des années, cette dernière peut changer et ainsi influencer la distribution de portefeuille. Alors il est important de travailler avec des données les plus récentes possibles. Le portefeuille non-Marin du lob incendie sur lequel nous travaillons n'as pas connu de changement d'exposition depuis 2012. Cela étant, nous modéliserons le nombre de sinistres à partir de l'année 2012.

## Modélisation du nombre de sinistres attritionnels

### ❖ Loi de Poisson

Les hypothèses sont les suivantes :

$$H = \begin{cases} H_0 : \text{La loi des sinistres attritionnels en fréquence est une loi de Poisson} \\ H_1 : \text{Non, elle ne suit pas une loi de poisson} \end{cases}$$

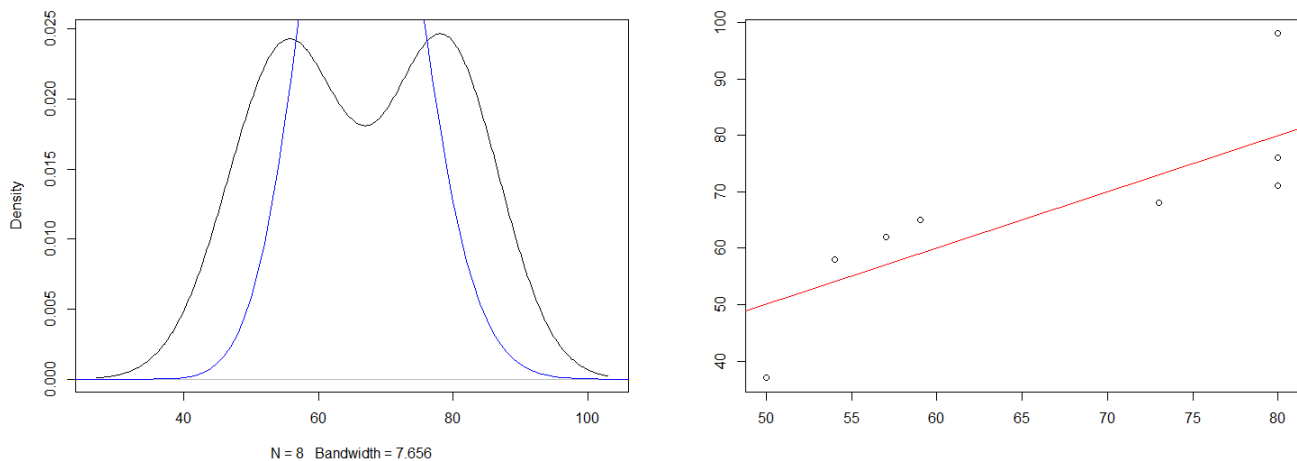


FIGURE 3.25 – Densité et QQ-Plot d'une loi de Poisson (bleue) superposée à celle du jeu de données (noire)

Nous observons clairement que la loi de Poisson n'est pas adaptée, sa densité ne coïncide pas du tout avec celle des sinistres extrêmes. Tout comme la densité, le QQ-plot n'est pas correct : les points ne sont pas situés sur la droite  $y = x$ .

### ❖ Loi Binomiale Négative

Les hypothèses sont les suivantes :

$$H = \begin{cases} H_0 : \text{La loi des sinistres attritionnels en fréquence est une loi binomiale négative} \\ H_1 : \text{Non, elle ne suit pas cette loi} \end{cases}$$

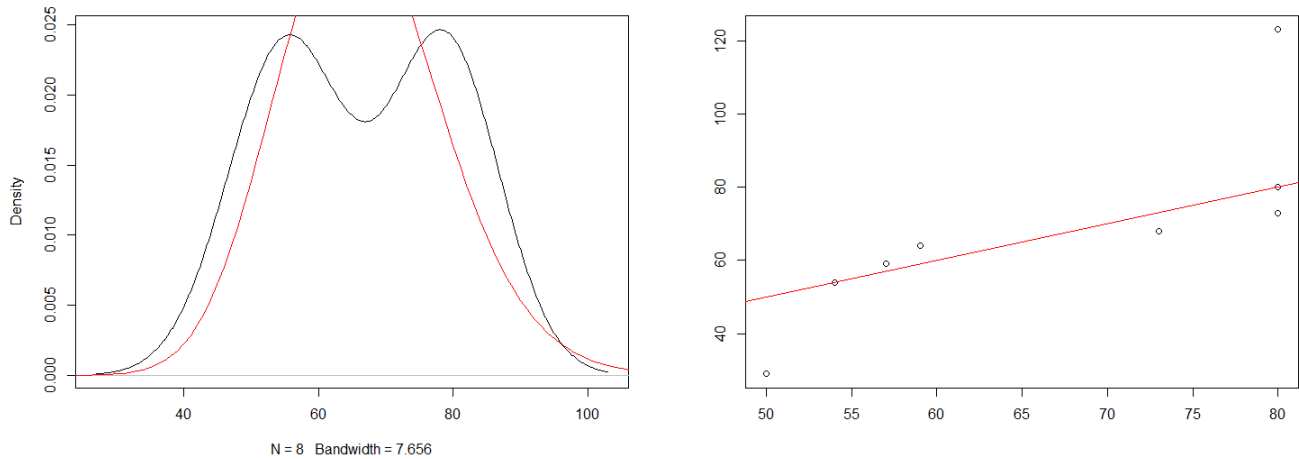


FIGURE 3.26 – Densité et QQ-Plot d’une loi binomiale négative (rouge) superposée à celle du jeu de données (noire)

Nous voyons tout de suite que la loi Binomiale négative est mieux ajustée que la loi de Poisson pour le nombre des sinistres attritionnels. Toutefois elle n’est pas parfaite, mais comme le jeu de données ne contient que 8 années, nous avons considéré que la loi Binomiale négative est la plus adaptée pour modéliser la fréquence des sinistres attritionnels.

### Test d’adéquation de khi-deux de Pearson

```
#Pearson’s Chi-squared test
#
#data:  tableaucontingence
# X-squared = 34.667
# p-value = 0.255
```

La p-value du test de khi-deux est supérieur à deux alors nous pouvons accepter l’hypothèse  $H_0$  et donc la loi binomiale négative, les paramètres sont :

Size( $N_1$ )	$\mathbb{E}(N_1)$
55	66,625

TABLE 3.13 – Paramètres de la  $\mathcal{BN}$ - Nombre de sinistres attritionnels

$$N_1 \sim \mathcal{BN}(55; 0, 45)$$

## Modélisation du nombre extrême

### ❖ Loi de Poisson

Les hypothèses sont les suivantes :

$$H = \begin{cases} H_0 : \text{La loi des sinistres attritionnels en fréquence est une loi de Poisson} \\ H_1 : \text{Non, elle ne suit pas une loi de poisson} \end{cases}$$

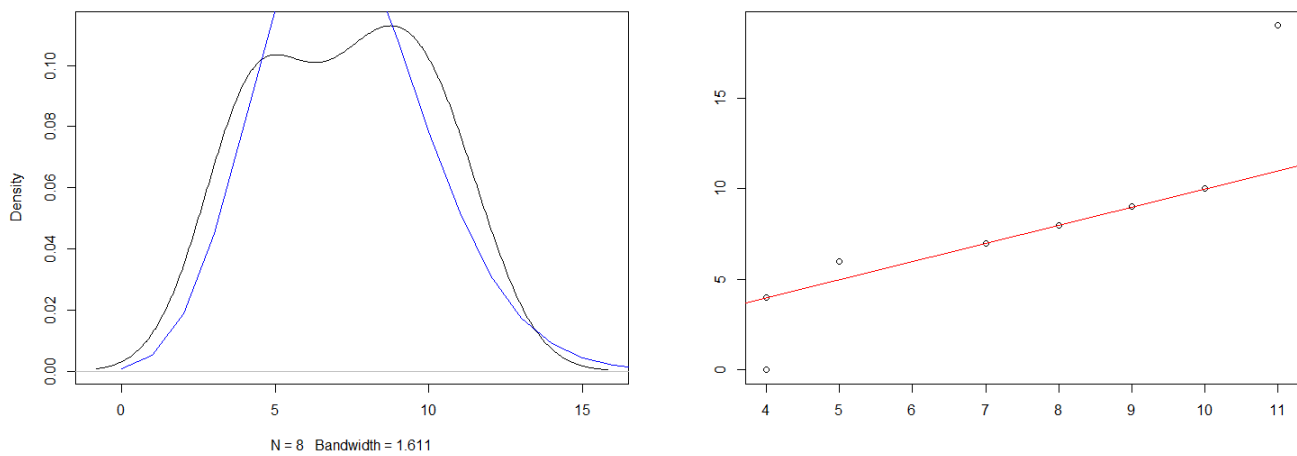


FIGURE 3.27 – Densité et QQ-Plot d’une loi de Poisson (bleue) superposée à celle du jeu de données (noire)

### ❖ Loi Binomiale Négative

Les hypothèses sont les suivantes :

$$H = \begin{cases} H_0 : \text{La loi des sinistres attritionnels en fréquence est une loi binomiale négative} \\ H_1 : \text{Non, elle ne suit pas une loi binomiale négative} \end{cases}$$

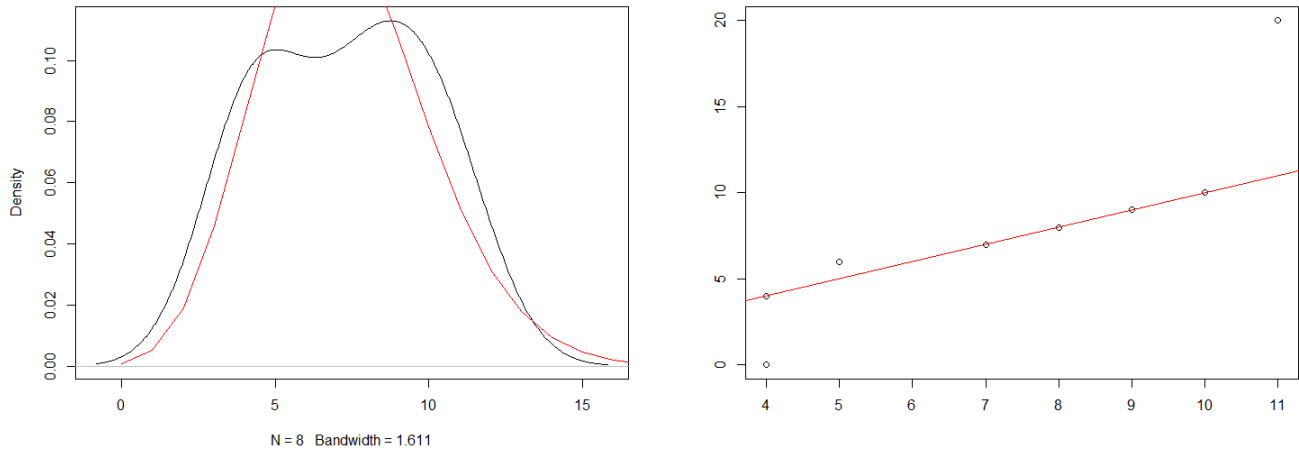


FIGURE 3.28 – Densité et QQ-Plot d’une loi binomiale négative (rouge) superposée à celle du jeu de données (noire)

Visuellement dans le cas de la fréquence de la sinistralité extrême, il nous est pas possible de distinguer graphiquement laquelle des deux lois testées précédemment est la meilleure. En effet les deux ajustent assez correctement le données. Faisons un test de sur-dispersion :

On rappelle que les hypothèses sont les suivantes :

$$\begin{cases} H_0 : \mathbb{E}(N) = \mathbb{V}(N) & (\text{Absence de sur dispersion}) \\ H_1 : \mathbb{E}(N) > \mathbb{V}(N) & (\text{Sur dispersion}) \end{cases}$$

La statistique du test est donnée par :

$$IF = n * \frac{(n-1)\bar{S}^2}{\bar{N}} \sim \chi_{(n-1),0.95}^2$$

Nous trouvons :

$$IF = 7,10 < \chi_{(n-1),0.95}^2 = 14,06714$$

Ainsi, nous acceptons  $H_0$ , donc la loi adaptée pour ce jeu de données est la loi de poissons.

$$N_2 \sim \mathcal{P}(7, 25)$$

### Modélisation du nombre de sinistres global

Nous référons au nombre de sinistres total sur le portefeuille. Nous modélisons cette fréquence dans le cadre du modèle global  $\left( S = \sum_i^N X \right)$

### ❖ Loi de Poisson

Les hypothèses sont les suivantes :

$$H = \begin{cases} H_0 : \text{La loi des sinistres attritionnels en fréquence est une loi de Poisson} \\ H_1 : \text{Non, elle ne suit pas une loi de poisson} \end{cases}$$

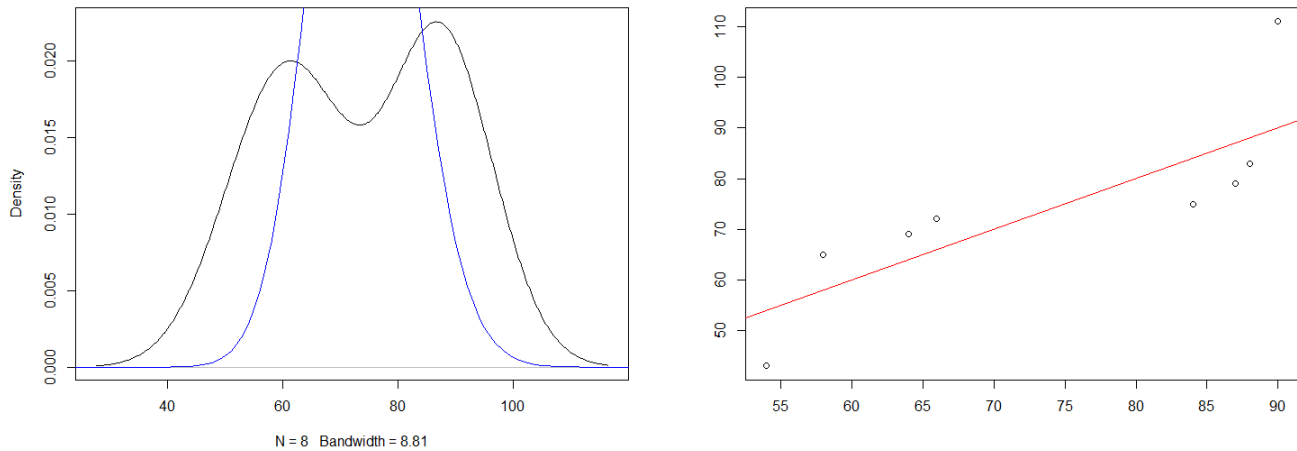


FIGURE 3.29 – Densité et QQ-Plot d’une loi de Poisson (bleue) superposée à celle du jeu de données (noire)

Il semble que la loi de poisson n’est pas adaptée. En effet ni la densité, ni le QQ-Plot ne montre une adéquation particulière.

### ❖ Loi Binomiale Négative

Les hypothèses sont les suivantes :

$$H = \begin{cases} H_0 : \text{La loi des sinistres attritionnels en fréquence est une loi binomiale négative} \\ H_1 : \text{Non, elle ne suit pas une loi binomiale négative} \end{cases}$$

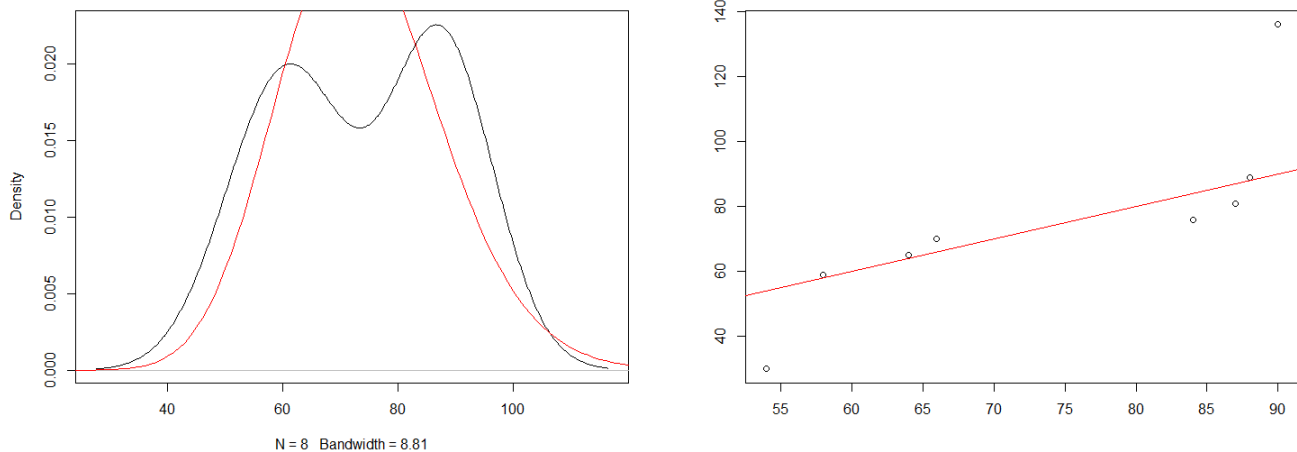


FIGURE 3.30 – Densité et QQ-Plot d’une loi binomiale négative (rouge) superposée à celle du jeu de données (noire)

La loi binomiale négative semble être plus adaptée. Ainsi nous utiliserons cette loi pour la fréquence globale. Faisons le test de Khi-deux de Pearson.

```
#Pearson's Chi-squared test
#
#data:  tableaucontingence
# X-squared = 40
# p-value = 0.2578
```

Nous acceptons ainsi  $H_0$  qui, nous rappelons est : ” Le loi s’ajuste bien ”.

Size( $N$ )	$\mathbb{E}(N)$
44,25	73,87

TABLE 3.14 – Paramètres de la  $\mathcal{BN}$ - Fréquence globale

$$N \sim \mathcal{BN}(44, 25; 0, 37)$$

### 3.5.6 Établissement de la distribution à deux lois

Maintenant que nous avons effectué la modélisation des nos lois en fréquence et en coût. Nous pouvons construire notre modèle clairement.

## Modèle 1

Nous appellerons modèle 1, le modèle de distribution à deux lois présenté précédemment. Nous rappelons qu'il s'écrit de la façon suivante :

$$S = S_1 + S_2 = \sum_{i=1}^{N_1} X_i + \sum_{i=1}^{N_2} Y_i$$

où  $S_1$  est la sinistralité attritionnelle et  $S_2$ , la sinistralité extrême.

La variable aléatoire  $X$  est à support sur  $[0, u]$  et est modélisée par une loi log-normale tronquée de densité  $f^{lnorm}$  et de fonction de répartition  $F^{lnorm}$ . Nous écrivons alors :

$$f_X = \frac{1}{F^{lnorm}(u)} f^{lnorm}(x) \mathbb{1}_{[0,u]} \quad \text{et} \quad F_X = \frac{1}{F^{lnorm}(u)} F^{lnorm}(x) \mathbb{1}_{[0,u]}$$

Les paramètres de la loi log-normale en question sont :

MeanLog	SdLog
9,526	3,073

$$\mathbb{E}(X) = \frac{1}{F^{lnorm}(u)} \int_0^u x f^{lnorm}(x) dx$$

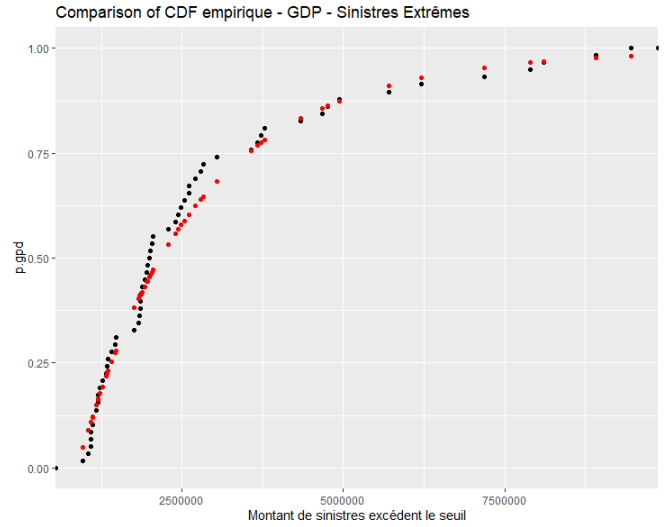
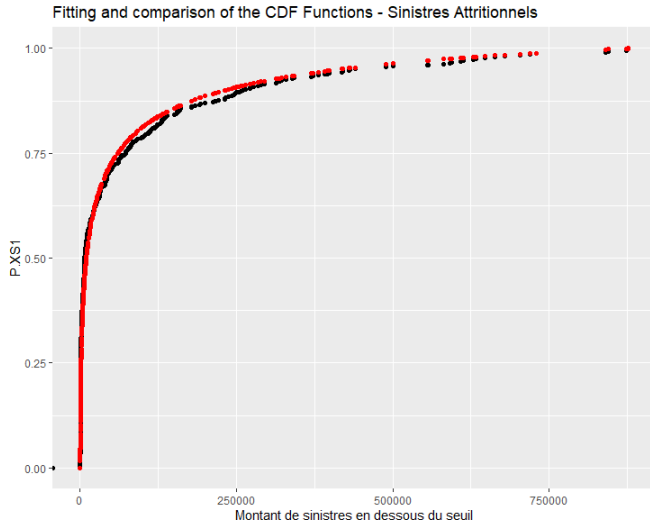
La variable aléatoire  $Y$  est à support sur  $]u, \infty[$ . Elle suit une loi de pareto généralisée de densité  $f^{GPD}$  et de fonction de répartition  $F^{GPD}$ .

$$Y \sim \mathcal{GPD}(\gamma = 0.08275, u = 880\,000, \sigma = 1\,795\,547)$$

Étant donné :  $\gamma < 1$

$$E(X^r) = \frac{\sigma}{\gamma^2} \frac{\Gamma(\gamma^{-1} - 1)}{\Gamma(1 + \gamma^{-1})}$$

Nous calculerons ces espérances par la méthode de simulation de Monte Carlo ([voir partie 4](#))



L'adéquation entre la fonction de répartition empirique(**en noire**) et la fonction de répartition théorique(**en rouge**) est très bonne que cela soit pour les sinistres attritionnels(à gauche)ou extrêmes(à droite).

La variable aléatoire  $N_1$  est modélisée par une loi binomiale négative.

$$N_1 \sim \mathcal{BN}(55; 0, 45)$$

La variable aléatoire  $N_2$  est modélisée par une loi de poisson.

$$N_2 \sim \mathcal{P}(7, 25)$$

Nous avons :

$$S_1 = \sum_{i=1}^{N_1} X_i$$

$$\mathbb{E}(S_1) = \mathbb{E}(N_1) \times \mathbb{E}(X) = 4\,819\,302 \quad \text{et} \quad \mathbb{E}(S_2) = \mathbb{E}(N_2) \times \mathbb{E}(Y) = 14\,192\,229$$

$$E(S) = \mathbb{E}(S_1) + \mathbb{E}(S_2) = 25\,369\,767 \text{ DH}$$

## Modèle 2

Nous appellerons modèle 2 le modèle à deux lois intégrées comme présenté au point (3.4.9). Nous rappelons qu'il s'écrit de la façon suivante :

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

avec :

$$f_X = f^{\text{lnorm}} \mathbf{1}_{[0,u]} + (1 - F^{\text{lnorm}}(u)) f^{\text{GPD}} \mathbf{1}_{]u, \infty[}$$

$$F(x) = \begin{cases} F^{lnorm}(x) & \text{si } x \leq u \\ F^{lnorm}(u) + (1 - F^{lnorm}(u)) F^{GPD}(x) & \text{si } x > u \end{cases}$$

Avec respectivement lnorm(Log-normale) et GPD(Loi de Pareto de généralisée). Les paramètres sont :

Paramètres de la loi log-normale	
MeanLog	SdLog
9,526	3,073

Paramètres de la loi de Pareto généralisée		
$u(\text{seuil})$	$\gamma$	$\sigma$
880 000	0,0827	1 795 547

$$\mathbb{E}(X) = \int_0^\infty x f(x) dx = \int_0^u x f^{lnorm}(x) dx + (1 - F^{lnorm}(u)) \int_u^\infty x f^{GPD}(x) dx$$

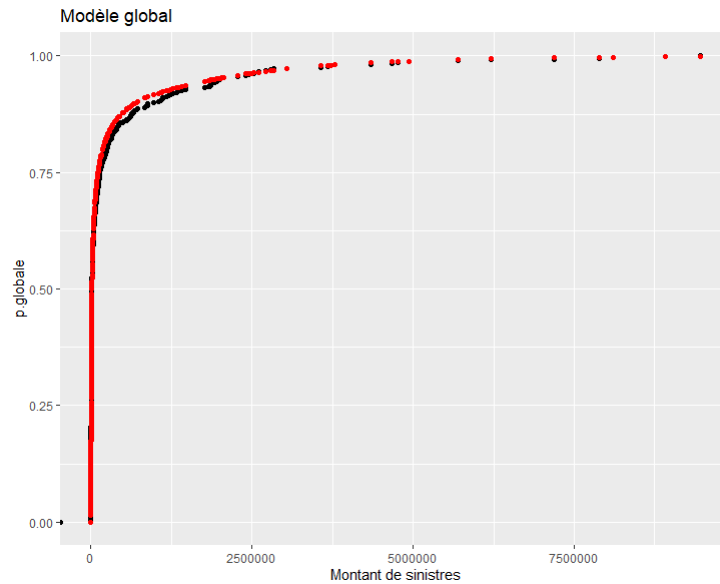


FIGURE 3.31 – Modélisation des sinistres

Graphiquement l'impact du modèle à deux lois est perceptible. La fonction de répartition théorique s'ajuste très bien.

La variable aléatoire  $N$ , la fréquence, est modélisée par une loi binomiale négative.

$$N_1 \sim \mathcal{BN}(44, 25; 0, 37)$$

Nous avons donc :

$$\mathbb{E}(S) = 24\,597\,617 \text{ DH}$$

L'avantage principal de ce modèle(modèle 2) est sa forme simple. Cela augmente l'intérêt de son application dans l'estimation des quantiles principalement dans des traitements avec la valeur at risk et tout ce qui est autour. Cependant, nous avons développé mathématiquement la fonction de répartition afin d'obtenir une écriture avec un coefficient de raccordement  $\alpha$ , ce paramètre introduit à première vue une sensibilité, en effet, nous rappelons que nous avons écrit :

$$f^{Globale} = \alpha f^{lnorm} + (1 - \alpha) f^{GPD}$$

$\alpha$  est un paramètre donnant le juste poids à chaque type de densité, il égale à la probabilité d'être en dessous ou au dessus du seuil des extrêmes, probabilité qui est donnée par  $F^{lnorm}(u)$ . A titre d'indication, nous trouvons  $\alpha = 0,91$ . Bien que le modèle construit est excellent, ce raccordement et par extension l'hypothèse selon laquelle la probabilité qu'un sinistre soit attritionnel est de 0,91 est éventuellement assez forte. Ceci étant, afin d'évaluer la sensibilité cette écriture mathématique par rapport à  $\alpha$ , étudions la sensibilité de  $\alpha$  en utilisant la méthode de bootstrap. Nous rappelons que le principe de Bootstrap est le suivant :

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,  $\theta$  le paramètre d'intérêt à estimer et  $\hat{\theta}(x_1, \dots, x_n)$  un estimateur de  $\theta$ . 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  $\theta$  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  $\varphi^{-1}$  est le quantile d'ordre  $\alpha$  de la loi normale centrée réduite.

Donc pour mesurer l'incertitude sur l'estimation du paramètre  $\alpha$ , nous avons généré 1 000 échantillons bootstrap à partir de nos données de base(données sur les montants de sinistres) et pour chacun de celui-ci, nous estimons la valeur de  $\alpha$ . Nous construisons à partir de ces estimations un échantillon. La moyenne empirique, l'écart type empirique, ainsi l'intervalle de confiance de niveau  $1 - \beta = 0.95$  sont :

$\hat{\mu}(\alpha)$	$\hat{\sigma}(\alpha)$	$\hat{IC}(\alpha)$
0,9167	0,007674	[0, 9162692; 0, 9171757]

Le paramètre qui nous intéresse, soit donc l'écart-type est très faible, ce qui nous indique que la sensibilité de notre  $\alpha$  est excellente, ceci se reflète très bien dans l'intervalle de confiance. Nous pouvons donc et déjà affirmer que le modèle 2 est très bien construit.

Aussi, la distribution empirique du paramètre  $\alpha$  laisse supposer une distribution gaussienne. Le test de Kolmogorov-Smirnov nous confirme cela avec une p-value = 0,9467 nous permettant d'accepter l'hypothèse selon laquelle les données suivent une loi normale de paramètre  $\hat{\mu}(\alpha)$  et  $\hat{\sigma}(\alpha)$

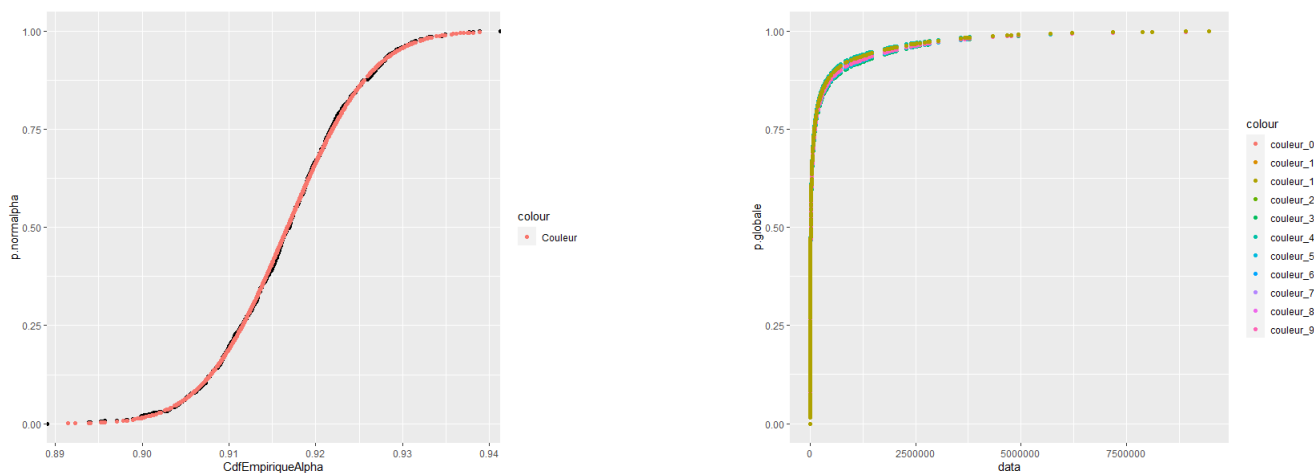


FIGURE 3.32 – Adéquation de alpha la loi normale et sensibilité de alpha par rapport à la loi globale(modèle)

Ce graphique nous montre et nous conforte dans l'idée de l'excellence de l'estimation du paramètre  $\alpha$ .

## Modèle 1 ou Modèle 2

Clairement les deux modèles sont proches et peuvent être utilisés dans des calculs mathématiques faisant intervenir la sinistralité d'un portefeuille. Le principal avantage du modèle 1 est qu'il sépare en somme la sinistralité attritionnels et la sinistralité extrêmes, ayant modélisé indépendamment chaque partie, nous pouvons faire des traitements sur une partie en particulier et ce indépendamment de l'autre. Cela dit, il devient difficile d'utilisation lorsque nous cherchons à modéliser certaine couverture introduisant par exemple des clauses avec des conditions logiques et/ou probabiliste nécessitant des calculs analytiques, dans ces conditions, le modèle 2 a largement l'avantage. Ceci dit, nous avons atteint notre but en fournissant à l'entreprise une modélisation sous deux formes pouvant être utilisée selon l'actuaire et la tâche sur des projets futurs nécessitant celle-ci.

## 3.6 Méthode de simulation de monte Carlo

La définition, la simulation est le fait d'utiliser un modèle, c'est à dire une représentation abstraite d'un système ou d'un problème, et d'étudier l'évolution de ce modèle sans faire fonctionner le système réel. Nommée par Stanisław Ulam et John von Neumann, la simulation de monte carlo est une méthode simple d'utilisation qui permet d'estimer des quantités numériques qui utilisent des nombres aléatoires. Elle est utilisée lorsque la complexité du modèle envisagé rend extrêmement difficile, voire impossible, toute approche analytique. Par exemple, pour  $X$  une variable aléatoire, approcher analytiquement la valeur de la moyenne de  $\min(\max(0, X - f), p)$  est extrêmement difficile.

Le principe de monte Carlo se base principalement sur trois propriétés.

### Propriété 1

*On note par  $F$  une fonction de répartition d'une variable continue et définie sur  $\mathbb{R}$ , Si  $X$  est une variable aléatoire de fonction de répartition  $F$ , alors  $F(X)$  suit une loi uniforme sur  $[0, 1]$ .*

$$F(X) \sim \mathcal{U}(0, 1)$$

Ainsi pour simuler la variable aléatoire  $X$ , il suffit de simuler un nombre entre 0 et 1, puis d'inverser la fonction de répartition à fin d'obtenir la réalisation associée.

Une fois qu'on a simulé les valeurs prises par une variable aléatoire suivant une loi donnée, la méthode de Monte Carlo consiste à approcher  $\mathbb{E}(X)$

### Propriété 2(La loi forte des grands nombres)

*Par définition, la loi des grands suppose qu'à mesure que la taille d'un échantillon décrivant une caractéristique  $X$  d'une population  $P$  augmente, la probabilité que la moyenne de l'échantillon(généralement notée  $\bar{X}$ ) soit différente de la moyenne de la population converge vers zéro. On écrit mathématiquement : soit  $(X_n)$  une suite de variables intégrables, indépendantes deux à deux, et identiquement distribuées. Soit  $\mathbb{E}(X)$  leur espérance commune, on note :*

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

*Alors la suite  $(\bar{X})$  tend presque sûrement(i.e probabilité du contraire quasi nulle) vers  $\mathbb{E}(X)$*

Ainsi la méthode de Monte Carlo consiste à approcher  $E(X)$  par la moyenne empirique des simulations de la variable aléatoire  $X$ .

Puisqu'on effectue des simulations, à priori en effectuant deux expériences similaires indépendantes l'une de l'autre, nous devrions obtenir deux résultats différents. Ainsi, à posteriori, la question de l'erreur d'estimation se pose, on estime cette erreur à l'aide du théorème central limite.

### Propriété 3 (Théorème centrale limite)

Le théorème central des limites stipule que sur un échantillon assez grande  $n$ , la distribution  $\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}$  converge dans la distribution vers la distribution normale standard, quelle que soit la distribution de la population.

Ainsi la loi gaussienne limite est caractérisée par deux paramètres, l'espérance de  $X$ , qui est justement ce que l'on cherche à estimer, et sa variance, qui détermine la dispersion par rapport à la moyenne. La variance est plus précisément l'espérance d'une nouvelle variable aléatoire  $Y$ . La variance est souvent inconnue, mais peut cependant être facilement estimée en même temps que la moyenne dans l'algorithme que nous avons décrit. On peut aussi construire des intervalles de confiance nous donnant l'information nécessaire sur la dispersion en estimant l'appartenance de notre estimation dans un intervalle donné avec une certaine probabilité (généralement de 0,95).

### Algorithme

1. Demander le nombre  $n$  de points à générer ;
2. Initialiser un compteur sum à 0 ;
3. Pour  $i$  de 1 à  $n$ 
  - générer une copie  $X_i$  de  $X$
  - ajouter la valeur  $X_i$  à sum
4. L'estimateur est  $\text{sum}/n$ , la moyenne des valeurs générées.

### Simulation Modèle 1

$$S = \sum_{i=1}^{N_1} X_i + \sum_{i=1}^{N_2} Y_i$$

Cet algorithme est suivi pour les parties de  $S$

- On fixe le nombre de simulations  $J$ , par exemple  $J = 100000$
- On crée un vecteur  $N$  contenant  $J$  simulations de même loi que la fréquence  $N = (n_1 \dots n_j \dots n_J)$  avec  $n_j$  la jème réalisation de la simulation par la loi  $N$
- Pour  $j = 1$  à  $J$
- Simuler  $n_j$  variables  $x_1^{n_j}, x_2^{n_j}, \dots, x_{n_j}^{n_j}$  issues de la variable aléatoire modélisant le coût.
- Pour  $i = 1$  à  $n_j$
- $S_j = \sum_{i=1}^{n_j} x_i^{n_j}$
- Appliquer une moyenne empirique au vecteur  $\{S_j\}_{1 \leq j \leq J}$  pour estimer l'espérance de la charge :  $\bar{S} = \frac{1}{J} \sum_{j=1}^J S_j$

Afin de simuler  $X$  et  $Y$ , on simule un échantillon aléatoire entre  $[0, 1]$ , échantillon auquel on appliquera l'inverse de  $F_X$  et  $F_Y$  pour trouver les montants. La fonction de répartition de  $X$  est donnée par :  $F_X(x) = \frac{1}{F^{lnorm}(u)} \times F^{lnorm}(x) \mathbf{1}_{[0,u]}$ , calculons son inverse :

$$\begin{aligned} z = \frac{1}{F^{lnorm}(u)} * F^{lnorm}(x) &\implies z \times F^{lnorm}(u) = F^{lnorm}(x) \\ &\implies F^{lnorm^{-1}}[F^{lnorm}(u) \times z] = x \end{aligned}$$

avec  $F^{lnorm^{-1}}$  l'inverse d'une fonction de répartition classique issue d'une loi log-normal,  $z$  une probabilité issue de la loi uniforme  $[0, 1]$  et  $x \in [0, u]$ . Pour la variable aléatoire  $Y$ , elle a une fonction de répartition dont l'inverse est classiquement connue, les montants seront  $F^{GPD^{-1}}[\mathcal{U}(0, 1)]$ .

## Simulation Modèle 2

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

On utilise en effet le même algorithme décrit précédemment. La fonction de répartition de  $X$  est donnée par :

$$F(x) = \begin{cases} F^{lnorm}(x) & \text{si } x \leq u \\ F^{lnorm}(u) + (1 - F^{lnorm}(u)) F^{GPD}(x) & \text{si } x > u \end{cases}$$

On cherche à simuler les valeurs de  $X$ , pour ce faire, on sait que d'après la propriété 1 que  $F_X \sim \mathcal{U}(0, 1)$ .  $F_X$  est une fonction par intervalle, :

### Sur $[0, u]$

$$\begin{aligned} 0 \leq x \leq u &\iff 0 \leq F(x) \leq F(u) \leq 1 \\ &\iff 0 \leq F(x) \leq F^{lnorm}(u) \end{aligned}$$

Pour tout  $y \in [0, F^{lnorm}(u)]$

$$y = F^{lnorm}(x) \iff x = F^{lnorm^{-1}}(y)$$

avec  $F^{lnorm^{-1}}$  l'inverse d'une fonction de répartition classique de la loi de log-normale.

### Sur $]u, \infty]$

$$\begin{aligned} u < x \leq \infty &\iff F(u) < F(x) \leq 1 \\ &\iff F^{lnorm}(u) + (1 - F^{lnorm}(u)) \times \lim_{x \rightarrow u} F^{GPD}(x) < F(x) \leq 1 \\ &\iff F^{lnorm}(u) \leq F(x) \leq 1 \end{aligned}$$

Ainsi pour tout  $y \in ]F^{lnorm}(u), 1]$

$$\begin{aligned} y = F^{lnorm}(u) + (1 - F^{lnorm}(u)) F^{GPD}(x) &\implies \frac{y - F^{lnorm}(u)}{1 - F^{lnorm}(u)} = F^{GPD}(x) \\ &\implies F^{GPD^{-1}}\left(\frac{y - F^{lnorm}(u)}{1 - F^{lnorm}(u)}\right) = x \end{aligned}$$

Dans les deux on simule  $y$  et on applique de la fonction en question, et on object les montants simulés de  $X$ .

# Chapitre 4

## Exemples d'utilisation de la modélisation

### 4.1 Pricing probabiliste d'un contrat XS

Le pricing est le fait de donner un prix à un contrat. Ayant modélisé intégralement la sinistralité, alors il devient possible de tarifer n'importe quel contrat en excédent de sinistre(noté XS). Pour rappel, le contrat XS est un contrat de couverture utilisé par les sociétés d'assurance ou de réassurance pour céder les risques qu'ils auront du mal à assuré et ce afin d'améliorer leur résultat global(financier et réputation d'engagement). Pour plus de détail consulter la partie([Chapitre 2 Point 2.2.1](#)). Nous considérons un contrat XS de portée  $p$  et de franchise  $f$ , nous notons  $pXSf$ . Le montant supporté par le contractant(le réassureur ou le rétrocessionnaire) est le suivant :

$$X^R = \begin{cases} 0 & , X < f \\ X - f & , f < X < f + p \\ p & , f + p < X \end{cases}$$

Nous notons par  $N$  le nombre de sinistres,  $X_i^R$  les différents sinistres et  $S^R$  la sinistralité, qui a donc pour expression  $S^R = \sum_{i=1}^N X_i^R$ . Pour simplifier nos calculs, pour le moment, faisons l'hypothèse que l'espérance du nombre de sinistre  $\mathbb{E}(N)$  est égale à 1 . Dans ce cadre, la prime pure est simplement,

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

Notons  $f_X$  la densité de  $X$ ,  $F_X$  sa fonction de répartition et  $\bar{F}_X = 1 - F_X$  sa fonction de survie, nous avons :

$$\begin{aligned}
\mathbb{E}(X^R) &= \int_f^{f+p} (x-f)f_X(x)dx + \int_{f+p}^{+\infty} p \cdot f_X(x)dx \\
&= \int_f^{f+p} x \cdot f_X(x)dx - f \int_f^{f+p} f_X(x)dx + p \int_{f+p}^{+\infty} f_X(x)dx \\
&= \int_f^{f+p} x \cdot f_X(x)dx - f \cdot (F_X(f+p) - F_X(f)) + p \cdot (1 - F_X(f+p)) \\
&= [x \cdot (F_X(x) - 1)]_f^{f+p} - \int_f^{f+p} (F_X(x) - 1) dx - f \cdot (F_X(f+p) - F_X(f)) + p \cdot (1 - F_X(f+p)) \\
&= \int_f^{f+p} \bar{F}_X(x)dx
\end{aligned}$$

Nous pouvons donc géométriquement interpréter la prime pure d'un contrat XS. elle est égale à l'aire située au dessus de la courbe de la fonction de répartition, plus précisément l'aire délimitée par la courbe de la fonction de répartition, les droites verticales d'équation  $y = f$  et  $y = f + p$ , et la droite horizontale d'équation  $x = 1$ .

#### 4.1.1 Exemple d'application

Nous considérons le modèle 1 décrit précédemment, nous avons :

$$S = \sum_{i=1}^{N_1} X_i + \sum_{i=1}^{N_2} Y_i$$

Nous notons par  $S^R$ , la charge du récessionnaire,

$$S^R = S_1^R + S_2^R = \sum_{i=1}^{N_1} \min[\max(X_i - f, 0), p] + \sum_{i=1}^{N_2} \min[\max(Y_i - f, 0), p]$$

Le prix d'un contrat XS de franchise  $f$  et de portée  $p$  est :  $\mathbb{E}(S^R)$

$$\mathbb{E}(S^R) = \mathbb{E}(S_1^R) + \mathbb{E}(S_2^R) = \mathbb{E}(N_1) * \mathbb{E}[\min(\max(X - f, 0), p)] + \mathbb{E}(N_2) * \mathbb{E}[\min(\max(Y - f, 0), p)]$$

Afin de calculer l'espérance de la variable aléatoire  $\min(\max(X - f, 0), p)$  une approche analytique est quasi impossible. Ainsi nous utiliserons la méthode de Monté Carlo pour l'estimer. Nous suivrons l'algorithme suivant pour calculer l'espérance de  $S_1^R$  et de  $S_2^R$ .

- Nous fixons le nombre de simulations  $J$ , par exemple  $J = 100000$
- Nous fixons une portée et priorité
- Nous créons un vecteur  $N$  contenant  $J$  simulations de même loi que la fréquence  $N = (n_1 \dots n_j \dots n_J)$  avec  $n_j$  la jème réalisation de la simulation par la loi  $N$
- Pour  $j = 1$  à  $J$ 
  - Simuler  $n_j$  variables  $x_1^{n_j}, x_2^{n_j} \dots x_{n_j}^{n_j}$  issues de la variable aléatoire modélisant le coût.
  - Pour  $i = 1$  à  $n_j$

- $\min[\max(x_i^{n_j} - f, 0), p]$
- $S_j = \sum_{i=1}^{n_j} \min[\max(x_i^{n_j} - f, 0), p]$
- Appliquer une moyenne empirique au vecteur  $\{S_j\}_{1 \leq j \leq J}$  pour estimer l'espérance de la charge nette du réassureur :  $\bar{S}^R = \frac{1}{J} \sum_{j=1}^J S_j$

Contrat	Tarif Pur DH
1 200 000 XS 1 000 000	5 982 658
1 200 000 XS 2 000 000	3 523 061
1 200 000 XS 5 000 000	815 060,8
1 200 000 XS 10 000 000	101 245,8
1 200 000 XS 15 000 000	17 723,47
1 200 000 XS 20 000 000	3 112,8
1 200 000 XS 30 000 000	230,33
1 200 000 XS 40 000 000	48

TABLE 4.1 – Tarification d'un XS

## 4.2 Optimisation de la couverture de rétrocession XS sur un critère de mesure de risk

Un réassureur, un assureur, une banque ou une institution financière, au cours d'une année, admet un montant maximum de pertes qu'il peut subir avec une probabilité de  $p$ . On appelle ce montant la value-at Risk. En effet, nous avons choisi d'utiliser cet indicateur car bien que l'écart-type soit une mesure incontournable du risque qui permet de calculer la dispersion autour de la moyenne, celui-ci présente le désavantage de ne pas capter l'asymétrie des distributions et ne tient pas compte de la disposition des masses de probabilité sachant qu'en rétrocession, les queues des distributions des sinistres sont assez lourdes. De ce fait, il est largement plus intéressant utilisé la value-at-risk.

La problématique est la suivante, pour une probabilité  $p$  donnée (dans notre cas bicentennale,  $p = 0.995$ ), pour une portée ou pour une priorité donnée, nous devons trouver la priorité ou la portée optimale avec laquelle la prise en charge de notre société, est inférieure à  $\text{Var}(\text{Value-at-risk})$  avec une probabilité  $p$ . Autrement dit optimiser le contrat XS de tel sort que le quantile à  $p\%$  obtenue après couverture soit inférieur à l'appétit de perte qu'on s'est fixé.

On définit mathématiquement la value at risk de la façon suivante :

$$P(\text{Loss} \leq \text{Var}) = p \iff F(\text{Var}) = 0,995$$

Donc

$$\text{Var}_p(\text{Loss}) = F^{-1}(p)$$

Où  $Loss$  est la variable aléatoire modélisant les pertes après couverture et  $F$  sa fonction de répartition.

La value at risk est calculable par simulation une fois la loi distribution des pertes est connue. Cela dit dans certains cas, son utilisation peut présenter un désavantage notamment lorsqu'on cherche à faire des manipulations sur la value at risk de deux portefeuilles fusionnés. En effet la Var n'est pas sous additive

$$\text{Var}(X_1 + X_2) \geq \text{Var}(X_1) + \text{Var}(X_2)$$

#### 4.2.1 Choix du nombre de réassureur

Un rétrocessionnaire ou un réassureur peut ne pas pouvoir supporter l'ensemble de la couverture voulue par la cédante. Il est alors normal de choisir plusieurs réassureurs. Cela dit, il est nécessaire de nous poser la question quant à l'efficacité de ce découpage de la couverture. En effet, on pourrait penser que le coût de reconstitution pourrait alors être supérieur avec plusieurs réassureurs plutôt qu'avec un unique réassureur. Nous allons montrer que le fait d'avoir 1 ou  $n$  réassureurs n'influe pas sur le coût de réassurance ni sur le risque subi à terme.

Alors pour ainsi faire, on suppose un "rétrocessionnaire global" qui a accepté de prendre en charge toute la couverture et  $n$  rétrocessionnaires qui vont supporter la même couverture voulue par le réassureur découpée sur plusieurs intervalles qui ne se chevauchent pas. On les notera respectivement  $R_0$  et  $R_i$  pour  $i \in \llbracket 1, n \rrbracket$ , on précise que  $n$  est un nombre entier non nul. On notera  $f$  la franchise (priorité) et  $P$  la portée de la couverture en XS. On suppose que la clause de reconstitution notée  $cl$  est identique pour le monde et que la prime de reconstitution est égale à 1% de la portée de chaque contrat.

**On écrit les différents contrat de réassurance :**

$$R_0 : P \mathbf{XS} f$$

$$R_i : \frac{P}{n} \mathbf{XS} f + (i - 1) \frac{P}{n}$$

Soit  $S$  la sinistralité

— Si  $S > f + p$

— Charge =  $CR_0 = P$

—  $\forall i \in \llbracket 1, n \rrbracket$  Charge de  $R_i = CR_i = \frac{P}{n}$

— Cout de reconstitution rétrocessionnaire 0 =  $RR_0 = 0,01 \times P \times Cl \times \frac{P}{P} = 0,01 \times P \times Cl$

—  $\forall i \in \llbracket 1, n \rrbracket$ , cout de reconstitution rétrocessionnaire  $i = RR_i = 0,01 \times \frac{P}{n} \times Cl \times \frac{P}{\frac{P}{n}} = 0,01 \times \frac{P}{n} \times Cl$

On remarque alors :

$$\sum_i^n RR_i = 0,01 \times P \times Cl = RR_0$$

—  $\forall k \in \llbracket 1, n \rrbracket$ , si  $f + \frac{P \times (k-1)}{n} < S \leq f + \frac{P \times k}{n}$

—  $CR_0 = S - f$ ,  $RR_0 = 0,01 \times (S - f) \times Cl$

—  $\forall j \in \llbracket 1, k-1 \rrbracket$ ,  $CR_j = \frac{P}{n}$ ,  $CR_k = S - f - (k-1) \times \frac{P}{n}$

—  $\forall j \in \llbracket k+1, n \rrbracket$ ,  $CR_j = 0$

—  $\forall j \in \llbracket 1, k-1 \rrbracket$ ,  $RR_j = 0,01 \times \frac{P}{n} \times Cl \times \frac{\frac{P}{n}}{\frac{P}{n}} = 0,01 \times \frac{P}{n} \times Cl$

—  $RR_k = 0,01 \times \frac{P}{n} \times Cl \times \frac{S - f - (k-1) \times \frac{P}{n}}{\frac{P}{n}} = 0,01 \times \left( S - f - (k-1) \times \frac{P}{n} \right) \times Cl$

—  $\forall j \in \llbracket k+1, n \rrbracket$ ,  $RR_j = 0$

On aura :

$$\begin{aligned} \sum_i^n RR_i &= \sum_j^{k-1} RR_j + RR_k \\ &= \left( \sum_j^{k-1} 0,01 \times \frac{P}{n} \times Cl \right) + 0,01 \times \left( S - f - (k-1) \times \frac{P}{n} \right) \times Cl \\ &= \left( \sum_j^{k-1} 0,01 \times \frac{P}{n} \times Cl \right) + 0,01 \times \frac{P}{n} \times Cl + 0,01 \times \left( S - f - k \times \frac{P}{n} \right) \times Cl \\ &= k \times 0,01 \times \frac{P}{n} \times Cl + 0,01 \times \left( S - f - k \times \frac{P}{n} \right) \times Cl \\ &= 0,01 \times (S - f) \times Cl = RR_0 \end{aligned}$$

D'où

$$\sum_i^n RR_i = RR_0$$

— Si  $S < f$

—  $CR_0 = 0$ ,  $RR_0 = 0$

—  $\forall i \in \llbracket 1, n \rrbracket$ ,  $CR_i = 0$ ,  $RR_i = 0$

$$\sum_i^n RR_i = RR_0$$

Finalement, pour  $S$  et pour tout  $n$ ,

$$RR_0 = \sum_i^n RR_i = RR_0$$

On vient de montrer que toute chose étant égale par ailleurs, le découpage en plusieurs tranches n'influe pas sur le cout. Être assuré par un seul rétrocessionnaire ou diviser la couverture sur plusieurs tranches génère les mêmes conséquences sur le portefeuille tant et aussi longtemps que tous les rétrocessionnaires peuvent assurer leur engagement.

## 4.2.2 Spécification de la problématique

On rappelle qu'étant donné une Value at risk, un portée ou un priorité, le but est de savoir quel priorité ou portée optimale par rapport à la valeur maximale de pertes qu'on se fixe avec une probabilité de 0,995. Mathématiquement, on a :

$$\mathbb{P}(Loss < Var) = 0,995 \iff F_L(Var) = 0,995$$

Avec  $F_L$  la fonction de répartition de la variable aléatoire  $Loss$

$$Loss = S - S^R = \sum_i^N X - \sum_i^N \min(\max(X - f, 0), p)$$

On a :

$$Var = F_L^{-1}(0,995) \iff Var - F^{-1}(0,995) = 0$$

Soit  $g$  une fonction définie par :  $g(f, p) = Var - F^{-1}(0,995)$ , étant donné une priorité  $f^*$ , la portée qui satisfait nos conditions est celle qui annule la fonction  $g(f^*, p)$ , soit donc, La portée optimale est telle que  $g(f^*, p) = 0$ . De même pour la priorité  $g(f, p^*) = 0$

## 4.2.3 Résolution

Afin de résoudre cette problématique, nous ferons appel à la méthode de dichotomie. La dichotomie est une méthode permettant d'approcher la solution d'une équation de type  $f(x) = 0$ . Précisément, nous supposons que la fonction  $f$  est continue sur un intervalle  $[a, b]$ , avec  $f(a) < 0$  et  $f(b) > 0$ . On sait donc qu'il existe au moins un réel  $c \in [a, b]$  tel que  $f(c) = 0$ . Le principe est le suivant :

- Si  $f\left(\frac{a+b}{2}\right) < 0$  alors  $c \in \left[\frac{a+b}{2}, b\right]$
- Si  $f\left(\frac{a+b}{2}\right) > 0$  alors  $c \in \left[a, \frac{a+b}{2}\right]$  Concrètement l'algorithme consiste à construire les suites  $(a_n)$  et  $(b_n)$  comme suit :
- $a_0 = a$  et  $b_0 = b$
- Si  $f\left(\frac{a_n + b_n}{2}\right) < 0$ , alors  $a_{n+1} = \frac{a_n + b_n}{2}$  et  $b_{n+1} = b_n$
- Sinon,  $a_{n+1} = a_n$  et  $b_{n+1} = \frac{a_n + b_n}{2}$  Les suites  $(a_n)$  et  $(b_n)$  converge donc vers  $c$  la solution de  $f(x) = 0$ . L'intervalle  $[a_n, b_n]$  est de longueur  $\frac{b_n - a_n}{2^n}$ .

Il est clair que les équations  $g(f^*, p) = 0$  ou  $g(f, p^*) = 0$  admettent une solution sur le domaine de définition de la fonction  $g : [0, X_m] \times [0, \infty[$ , avec en effet  $X_m = \mathbb{E}(\max(X))$  calculable par la méthode de monté Carlo.

## Calcul du quantile de la charge $L$ à $p\%$

Le calcul de  $F^{-1}(0,995)$  se fait par simulation. Connaissant la distribution de la sinistralité en fréquence et en coût, on peut ainsi appliquer la simulation de la façon suivante.

- On fixe le nombre de simulations  $J$ , par exemple  $J = 100000$
- on fixe une portée et priorité
- On crée un vecteur  $N$  contenant  $J$  simulations de même loi que la fréquence  $N = (n_1 \dots n_j \dots n_J)$  avec  $n_j$  la jème réalisation de la simulation par la loi  $N$
- Pour  $j = 1$  à  $J$ 
  - Simuler  $n_j$  variables  $x_1^{n_j}, x_2^{n_j} \dots x_{n_j}^{n_j}$  issues de la variable aléatoire modélisant le coût.
  - Pour  $i = 1$  à  $n_j$ 
    - $x_i^{n_j} - \min[\max(x_i^{n_j} - f, 0), p]$
    - $L_j = \sum_{i=1}^{n_j} x_i^{n_j} - \sum_{i=1}^{n_j} \min[\max(x_i^{n_j} - f, 0), p]$
- On trie alors le vecteur obtenu  $\{L_j\}_{1 \leq j \leq J}$  par ordre croissant, et retenir la valeur qui représente 99,5% de la taille de l'échantillon.

### En fixant la priorité :

Tout d'abord, il est important de noter que si la priorité choisie est trop élevée, alors même si le contrat comprend une portée infinie, les conditions ne pourront pas être respectées. Il existe donc une priorité maximale à partir de laquelle même avec une portée infinie, les conditions ne pourront pas être respectées. Ainsi nous avons écrit une fonction de recherche en utilisant l'algorithme de dichotomie avec une portée infinie, qui renvoie de ce fait la priorité maximale qu'on puisse choisir.

La cédante peut alors choisir une priorité, qui est inférieure à la priorité maximale. On crée par la suite l'algorithme qui permet de résoudre par la dichotomie notre équation, algorithme qui va renvoyer la portée optimale pour la priorité choisie. Son temps d'exécution est de 2 min en moyenne. Pour compléter la méthodologie de monte Carlo, on crée une boucle sur ce dernier algorithme d'au moins 35 occurrences et on calcul la moyenne des portées optimales obtenues.

### En fixant la portée

, la même méthodologie est suivie que dans le cadre de la fixation de la priorité. En effet son choix n'est pas complètement libre. Pour une portée relativement faible, le montant à charge de la cédante(SCR) a un grand risque d'être supérieur à la perte maximale Var désirée. Afin de s'assurer qu'il est possible que la cédante puisse avoir une perte au maximum égale à Var avec une probabilité 0,995, il faut lui imposer une portée minimale. En considérant le cas extrême où la priorité est nulle, c'est-à-dire que tous les sinistres extrêmes vont être pris en charge par la cédante(dans notre cas la SCR), on détermine la portée qui va permettre de remplir les conditions fixées, une fois qu'on a déterminé la portée minimale, on peut alors choisir une qui lui est supérieure et appliquer la méthodologie expliquée précédemment pour trouver la priorité optimale.

En fixant la priorité - Var = 50 millions	
Priorité(francise) maximale	1 755 560
Priorité choisie	800 000
Portée optimale	1 351 374

En fixant la portée - Var = 50 millions	
Portée minimale	375 478,3
Portée choisie	1 500 000
Priorité optimale	952 705,5

En fixant la priorité - Var = 100 millions	
Priorité(francise) maximale	31 368 170
Priorité choisie	20 000 000
Portée optimale	0,985 492 5

En fixant la portée - Var = 100 millions	
Portée minimale	0,862 245 3
Portée choisie	100
Priorité optimale	47 333 395

TABLE 4.2 – Optimisation par dichotomie-En DH

Lorsque la Value at risk est égale à 100 millions, avec une priorité de 20 millions, nous trouvons une portée optimale très faible. De même lorsque nous fixons une portée faible, nous trouvons une priorité optimale très grande. Nous pouvons affirmer qu'à ce niveau d'appétit de risque, notre portefeuille n'a pas besoin de couverture.

# Conclusion

La sinistralité issue de l'activité de rétrocession est très complexe. Elle nécessite de faire appel à des modèles et à des méthodes mathématiques sophistiquées. Les outils qui ont été utilisés sont ceux de l'inférence classique et de la théorie des valeurs extrêmes. En effet la problématique réside de la modélisation de la sinistralité en prenant en considération la queue de distribution à sa juste valeur. Cela nous a conduit à la séparation des sinistres en catégories par un seuil pour en faire une partie attritionnelle et une autre extrême. L'outil de séparation 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. 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. Le plus efficace dans la littérature étant le Hill-Plot, nous avons retenu son utilisation pour la détermination du seuil, en suivant le protocole : en partant des observations les plus extrêmes, rechercher une zone de stabilité de l'estimateur de Hill, puis prendre le seuil à la fin de cette zone de stabilité, juste avant un changement de comportement. Et Chacune de ces deux types de sinistralité à été modélisées par une loi de probabilité adéquate.

Cette modélisation de la sinistralité extrême permet d'effectuer des manipulations plus ou moins complexes sur les caractéristiques de la couverture en rétrocession du portefeuille. Nous avons pris pour exemple la tarification d'un contrat XS en rétrocession. En effet, connaissant les lois de notre distribution de données en fréquence et en coût, on peut ainsi calculer en appliquant la simulation de monte Carlo l'espérance de n'importe quelle manipulation mathématique de la sinistralité. Une autre application qui a été présentée est l'optimisation d'une rétrocession XS sous condition de l'appétit de risque de la SCR. La problématique a été résolue par la création d'un algorithme basé sur la méthode de dichotomie. L'étude réalisée dans le cadre de ce mémoire permet d'avoir une modélisation beaucoup plus cohérente que la modélisation classique, elle nous permettra ainsi d'améliorer nos résultats que cela soit dans le cadre de la mise en place d'une couverture ou dans le cadre de l'étude des indicateurs de rentabilité après cette dernière.

# Références

1. ALEXANDER J. MCNEIL, Departement Mathematlk, ETH Zentrum, CH-8092 Zurich, March 7, 1997 - Estimating the tails of loss severity distributioins using extreme value theory
2. Borchani-Statistiques des valeurs extrêmes dans le cas de lois
3. Chesneau, C. (2017). Sur l'Estimateur du Maximum de Vraisemblance (emv). Lecture.
4. Claude Blisle-Table de khi-deux
5. Cours sur les techniques de réassurance-<https://fr.scribd.com/doc/32142159/Reassurance>
6. Goffard-Modelisation charge sinistre m2 actuariat - chapitre 3 : Modèle collectif. Présentation de cours, ISFA. 2019
7. Gourmelon, V. (2014). Modélisation des sinistres graves en assurance automobile responsabilité civile. Mémoire d'actuaire, EURIA.
8. Introduction à la théorie des valeurs extrêmes : Applications en actuariat. Guillou, Ecole d'été, Université de Strasbourg et Société Générale Assurance.
9. J. Beirlant-J. Teugels-P. Vynckier, Practical analysis of extreme values
10. Mickael ITTAH - Calibrage de la sinistralité non vie – Cotation et optimisation de la réassurance non proportionnelle-Mémoire Actuariat Université de Strasbourg
11. Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of statistics*, 3 :119–131
12. Ressource R pour les valeurs extrêmes
13. Université Laval - Modélisation des distributions de sinistres avec R
14. Yapi Arnaud Carmel-Analyse de la sinistralité d'une base de données à l'aide de la théorie des valeurs extrêmes et d'un outil de simulation.

# Annexes

## Annexe 1

### Quelques remarques sur le marché marocain avant l'application du taux de cession réglementaire des affaires facultatives marocaines.

Il serait intéressant pour les lecteurs de connaître quelques résultats de notre modélisation sur le marché marocain avant l'application du taux de cession réglementaire de la SCR. En effet dans notre étude, les résultats sont spécifiques à la SCR.

### Seuil des extrêmes

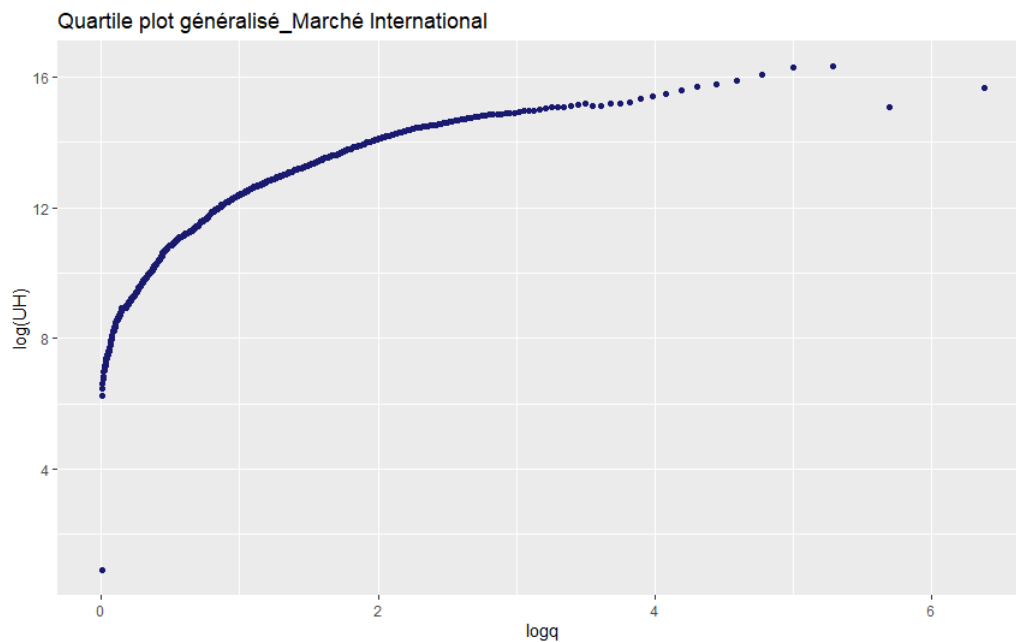


FIGURE 4.1 – Quartile Plot généralisé - Sinistres réels Marché Marocain

Les données appartiennent bien au domaine d'attraction de Fréchet.

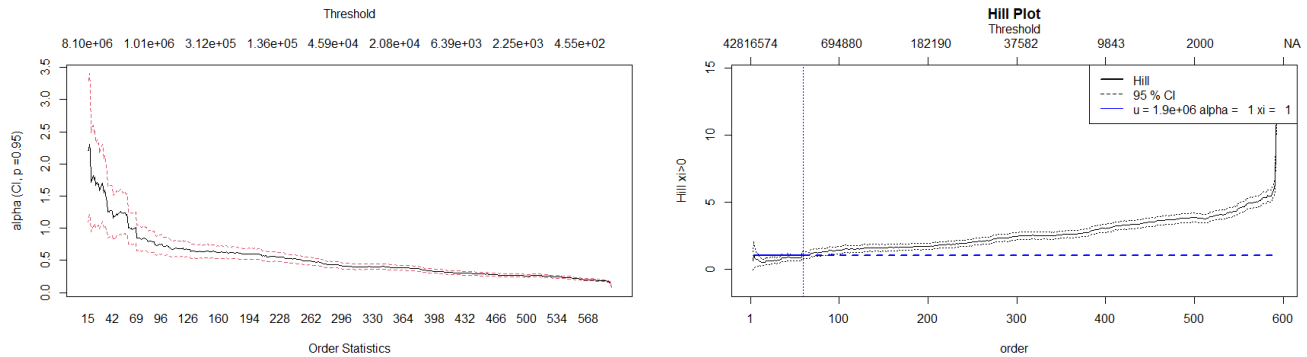


FIGURE 4.2 – Hill Plot - Sinistres réels Marché Marocain

Le seuil des extrêmes trouvé est :  $u = 1\,900\,000$ . Ce seuil deux fois supérieur au seuil trouvé lors de notre étude c'est à dire après application du taux de cession des affaires facultatives de la SCR. Cela dit la proportion des sinistres extrêmes n'a pas changé, ces sinistres représente 10% des nos données.

### Loi de modélisation

En appliquant la modélisation à deux lois, on trouve dans ce cas ci que les sinistres attritionnels sont modélisés par la loi log-normale tronquée( de paramètres  $\mu = 10,84$  et  $\sigma = 2,783$  ) à support sur  $[0, u]$ . Et les sinistres extrêmes sont modélisés par une loi généralisée de Pareto de paramètre  $\lambda = 0,346$  et  $\sigma = 3370709$  à support  $]0, \infty]$ . En appliquant la méthode de monté Carlo, on trouve un sinistre annuel moyen :

$$\mathbb{E}(S) = 56\,466\,801$$

Le taux de cession appliqué par la SCR aux affaires facultatives marocains change la quantité de perte et de risque à gérer. On rappelle qu'on travail sur le lob incendie, la scr travail sur un sinistralité annuelle moyenne de l'ordre de 25 millions de dirhams contre 56 millions sur le marché réel.

## Annexe 2

### Démonstration du coûts moyen des lois usuelles

#### 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  $\varphi$  la fonction de répartition d'une variable aléatoire qui suit une loi Normale centrée réduite.

Considérons alors un traité  $\infty$  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  $y = t - \sigma$

$$\begin{aligned} &= e^\mu \int_{\frac{\ln(D)-\mu}{\sigma} - \sigma}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[y^2 - \sigma^2]} dy \\ &= e^{\mu + \frac{\sigma^2}{2}} \int_{\frac{\ln(D)-\mu}{\sigma} - \sigma}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\ &= e^{\mu + \frac{\sigma^2}{2}} \times \mathbb{P}\left(Y \geq \frac{\ln(D) - \mu}{\sigma} - \sigma\right) \quad \text{avec } Y \sim \mathcal{N}(0, 1) \\ &= e^{\mu + \frac{\sigma^2}{2}} \times \mathbb{P}\left(Y \leq \sigma - \frac{\ln(D) - \mu}{\sigma}\right) \\ &= e^{\mu + \frac{\sigma^2}{2}} \varphi\left(\frac{\sigma^2 - \ln(D) + \mu}{\sigma}\right) \end{aligned}$$

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 :

$$\begin{aligned} \pi(D, C) = \frac{1}{\varphi\left(\frac{\mu - \ln(s)}{\sigma}\right)} & \left( e^{\mu + \frac{\sigma^2}{2}} \left[ \varphi\left(\frac{\sigma^2 - \ln(D) + \mu}{\sigma}\right) - \varphi\left(\frac{\sigma^2 - \ln(D + C) + \mu}{\sigma}\right) \right] + \right. \\ & \left. (D + C) \varphi\left(\frac{\mu - \ln(D + C)}{\sigma}\right) - D \varphi\left(\frac{\mu - \ln(D)}{\sigma}\right) \right) \end{aligned}$$

## 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  $\infty$  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	$\pi(D, C)$	$\pi(D, \infty)$
$D \geq A$	$\frac{A^\alpha D^{1-\alpha}}{1-\alpha} \left[ \left(1 + \frac{C}{D}\right)^{1-\alpha} - 1 \right]$	$\frac{A^\alpha D^{1-\alpha}}{\alpha - 1}$

## Loi de Weibull

Si  $X \sim \mathcal{W}(a, b)$ , avec  $a > 0$  et  $b > 0$  alors sa densité est de la forme :

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

Et sa fonction de répartition :

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

La fonction de répartition d'une variable aléatoire suivant une loi de Weibull tronquée à gauche au seuil  $s$  devient alors : Considérons alors un traité  $\infty$  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	$\pi(D, C)$	$\pi(D, \infty)$
$D \geq s$	$\left(\frac{b}{a}\right) e^{\left(\frac{s}{b}\right)^a} \left[ \Gamma\left(\left(\frac{D}{b}\right)^a; \frac{1}{a}\right) - \Gamma\left(\left(\frac{D+C}{b}\right)^a; \frac{1}{a}\right) \right]$	$\frac{a}{b} e^{\left(\frac{s}{b}\right)^a} \Gamma\left(\left(\frac{D}{b}\right)^a; \frac{1}{a}\right)$