

Mémoire présenté le :

**pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA
et l'admission à l'Institut des Actuaires**

Par : LE Thi Thanh Nga

Titre : Optimisation de l'outil de tarification des traités CAT XL par le modèle POT

Confidentialité : NON OUI (Durée : 1 an 2 ans)

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

*Membre présents du jury de l'Institut
des Actuaires*

Entreprise :

Nom : Arundo Re

Signature :

Directeur de mémoire en entreprise :

Nom : Maxime CABROL

Signature : 

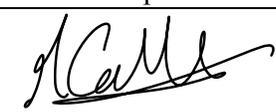
Invité :

Nom :

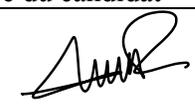
Signature :

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

Signature du responsable entreprise



Signature du candidat



Mémoire d'actuariat



S U J E T :

Optimisation de l'outil de tarification des traités CAT XL en réassurance vie par le modèle Peaks Over Threshold (POT)



Par : LE Thi Thanh Nga

Tuteur d'entreprise : Maxime CABROL

Manager Senior - Équipe Tarification Vie de la Direction Actuariat - Risque
Arundo Re

Tuteur Pédagogique : Frédéric PLANCHET

Professeur à l'Institut de Science Financière et d'Assurances

Paris, Mars 2025

Table des matières

Remerciements

Résumé

Abstract

INTRODUCTION

1	Naissance du réassureur Arundo Re	1
2	Présentation du risque catastrophe	4
2.1	Catastrophe d'origine naturelle (Cat-nat)	6
2.2	Catastrophe d'origine humaine (man-made catastrophe)	9
3	Le monde de la réassurance	11
3.1	Rôle de la réassurance	11
3.2	Marché de la réassurance	12
3.3	Types de réassurance	15
3.4	Traité excédent de sinistres par évènement catastrophe (XSC)	23
3.4.1	Terminologies générales	23
3.4.2	Clauses particulières	24
4	Explication du modèle « Peaks Over Threshold » (POT)	27
4.1	Construction des variables	29
4.1.1	Sources et structure des données	30
4.1.2	Nettoyage et prétraitement	30
4.1.3	Sélection et regroupements géographiques	31
4.1.4	Variables d'entrée	32

4.2	Modélisation de la fréquence des évènements catastrophiques	33
4.3	Modélisation de la sévérité des sinistres	36
4.3.1	Estimation du nombre de victimes	36
4.3.2	Modélisation du nombre de victimes assurées dans notre portefeuille	37
4.4	Modélisation du coût de chaque sinistre lié à l'évènement catastrophe concerné	41
4.5	Application des conditions du contrat CAT	42
5	Amélioration du modèle	44
5.1	Détermination du seuil optimal	48
5.1.1	Méthode d'analyse décisionnelle à critères multiples (MCDA)	48
5.1.2	Estimateur de Hill	54
5.1.3	Vérification de la conformité du seuil optimal	56
	Mean Excess Function (ou Fonction d'excès en moyenne)	56
	Threshold choice plot [Davison and Smith (1990)]	58
	Q-Q Plot	61
5.2	Groupement	64
5.2.1	Logique du regroupement dans notre modèle	65
5.2.2	Méthodes analytiques pour le choix des pays à regrouper	65
	Analyse globale des évènements catastrophiques	65
	Analyse des tendances mensuelles des évènements naturels	66
	K-Mean cluster	69
6	Calibrage et Sensibilité du modèle	72
6.1	Estimation de la fréquence des évènements	72
6.2	Sensibilité de l'estimation des victimes assurées	74
6.3	Ajustement par maximum de vraisemblance de la distribution de Pareto généralisée	77
7	Tarification du traité	80

CONCLUSION	84
APPENDICE	86
Notion du modèle de tarification des traités XSC en réassurance vie	86
Théorie des Valeurs Extrêmes	87
Preuve du théorème Fischet Tippet	93
Estimation de l'engagement moyen par tête	98
REFERENCE	99

Liste des figures

1	Chronologie historique d'Arundo Re	1
2	Répartition géographique des activités de réassurance non-vie et vie (2022 vs 2023)	2
3	Répartition du chiffre d'affaires au titre de l'exercice (2022 vs 2023)	3
4	Tendances de la mortalité par continent pour la période 1970 - 2020	4
5	Présentation des impacts des catastrophes de 1970 à 2020	5
6	Nombre de décès par décennie (1900 - 2020)	6
7	Nombre d'évènements dûs à des catastrophes de 1970 à 2020	7
8	Chiffres clés des catastrophes naturelles des 2 dernières décennies	8
9	Nombre de décès liés aux catastrophes naturelles par type d'évènement et par continent	8
10	Répartition des décès par origine dûs au terrorisme entre 1970 et 2020	9
11	« Man-made » évènements catastrophiques par continent	9
12	Chiffre d'affaires global sur le marché de la réassurance	12
13	Répartition du chiffre d'affaires selon les marchés de la réassurance	13
14	Evolution des primes, résultat net, fonds propres et ROE sur la période 2014 - 2023	14
15	Types de réassurance	15
16	Répartition des sinistres dans un traité de réassurance en quote-part (30%)	17
17	Illustration de traité de Excédent de Plein : 2000 XS 1000	19
18	Exemple de traité XS CAT 200 XS 50	26
19	Processus du modèle de tarification du traité XSC dans l'assurance-vie	28
20	Extrait de notre base de données de référence pour la France	33
21	Comportement de p lié sur $d(X)$ avec $p_{assume} = 0.2$	40
22	Distribution des décès des évènements catastrophiques en France	45

23	Distribution des décès en Chine	46
24	Distribution des décès en Colombie	47
25	Le résultat de test statistique global	52
26	Le résultat de tests statistiques	53
27	Plot de Stabilité de l'Estimateur de Hill	55
28	Mean Residual Life Plot	57
29	Threshold choice Plot pour la France	58
30	Threshold choice Plot pour la Chine	59
31	Threshold choi Plot pour la Chine pour les quantiles [70% - 90%]	59
32	Threshold choice Plot pour la Colombie	60
33	QQ-Plot pour la zone France	61
34	QQ-Plot de la zone Chine pour le seuil 83 et 110	62
35	QQ Plot de la zone Colombie pour des les seuils 37; 42 et 60	63
36	Nombre de évènements dans les pays voisins dus à divers événements catastrophiques	65
37	Distribution mensuelle des décès (événement naturel origine) entre la France et les pays voisins	67
38	«Heatmap» de corrélation	68
39	Map «coude» - nombre de cluster	69
40	Tableau des pays regroupés par clusters selon leur exposition aux catastrophes naturelles (résultat du K-Means)	70
41	Illustration de K-Means Cluster	70
42	Q-Q Plot comparant la loi de Poisson et la loi binomiale négative	73
43	Fonction logarithme	75
44	Changement de θ	76
45	Distribution de p	76
46	Variation des paramètres de GPD avec MLE	79

47 Visualisation des densités des distributions de Gumbel, Fréchet et Weibull 90

Remerciements

Je tiens tout d'abord à exprimer ma profonde et sincère gratitude à Monsieur Maxime CABROL, mon maître de stage et désormais mon manager, responsable de l'équipe Actuariat Vie de Arundo Re pour ses conseils, ses soutiens et sa disponibilité. Son expertise, ses corrections minutieuses ont été essentielles tout au long de ce travail. Grâce à lui, j'ai eu l'opportunité de traiter un sujet complexe, me permettant d'approfondir mes connaissances et de progresser professionnellement.

Je tiens à remercier à mes collègues, qui ont été à mes côtés tout au long de ce travail. Depuis mon arrivée, ils m'ont accueilli avec bienveillance et enthousiasme, créant un environnement de travail agréable.

Je souhaite également adresser quelques mots en anglais à mes proches.

To my family and friends — thank you for your unwavering support. Having you by my side through every up and low has been one of the greatest gifts in my life. Your support has always been a source of motivation, helping me grow and move forward every day.

Résumé

L'assurance et la réassurance jouent un rôle clé dans l'atténuation de l'impact financier des événements catastrophiques, contribuant à la stabilité économique et à la résilience des individus, des entreprises et des gouvernements. Cependant, face à l'ampleur des dommages potentiels, un seul assureur peut avoir des difficultés à absorber l'ensemble du risque. C'est dans ce contexte que **le contrat d'excédent de sinistre par évènement catastrophique (XSC)** intervient.

Cette étude vise à améliorer l'outil de tarification des événements catastrophiques en réassurance vie. Le modèle actuellement utilisé chez Arundo Re repose sur l'approche **Peaks Over Threshold (POT)**, issue de la théorie des valeurs extrêmes, développée à l'origine par Erland Ekhedén et Ola Hössjer. Cette méthode permet de modéliser la fréquence et la gravité des événements extrêmes, conditionnée à un seuil pertinent. L'objectif est d'acquérir une vision globale des schémas de risque catastrophique et de comprendre la transition entre les fondements théoriques du modèle POT et ses applications pratiques.

Dans ce mémoire, deux points majeurs ont été développés pour améliorer la robustesse du modèle. Le premier porte sur le choix optimal du seuil à partir duquel la distribution de Pareto généralisée s'applique pour estimer le nombre de victimes dans un événement catastrophique, en s'appuyant sur des tests statistiques et des outils graphiques. Le second point concerne une méthode de regroupement de pays présentant des caractéristiques similaires (géographiques, climatiques ou en termes de gravité des sinistres) afin de compenser la rareté des données au niveau national, et ainsi d'obtenir des estimations plus fiables.

Ces améliorations visent à renforcer la fiabilité des estimations de fréquence et de sévérité des sinistres, ce qui permet une tarification plus précise, même dans un contexte de rareté des données.

Mots-clés : Réassurance vie, Risque catastrophique, Modèle Peaks Over Threshold, Catastrophe naturelle, Terrorisme, Excédent de sinistre, Man-made catastrophe, Théorie des valeurs extrêmes.

Abstract

It is universally acknowledged that insurance and reinsurance play an essential role in mitigating the financial impact of catastrophic events, contributing to economic stability and resilience of individuals, businesses, and governments. However, due to the scale of potential damages, a single insurer may struggle to absorb all associated risks. This is precisely where the **Catastrophe Excess of Loss contract (XSC)** comes into play.

Within this framework, this thesis aims to improve the pricing tool used for the catastrophic events in life reinsurance. The model currently implemented at Arundo Re is based on the Peaks Over Threshold (POT) approach, derived from extreme value theory, developed originally by Erland Ekheden and Ola Hössjer. This method makes it possible to model both the frequency and the severity of extreme events above a given threshold of number of victims, using the Generalized Pareto Distribution (GPD). The objective is to gain a comprehensive understanding of catastrophic risk patterns and to bridge the gap between the theoretical foundations of the POT model and its practical applications.

Two main areas of improvement are explored throughout this work to enhance the robustness of the model. The first focuses on determining the optimal threshold above which the GPD can be reliably applied, using a combination of statistical tests and graphical tools. The second involves grouping countries with similar characteristics — whether geographic, climatic, or in terms of loss severity — in order to address data scarcity at the national level and produce more reliable estimations.

Together, these enhancements aim to improve the accuracy of frequency and severity estimations, leading to more precise pricing even in the context of limited data availability.

Keywords : Life insurance, Catastrophic risk, Model Peaks Over Threshold, Natural catastrophe, Terrorism, Excess of loss, Man-made catastrophe, Theory of extreme value.

INTRODUCTION

Les catastrophes, qu'elles soient d'origine naturelles ou humaines, représentent des défis majeurs pour le secteur de l'assurance. Leur caractère soudain, leur ampleur potentiellement dévastatrice et les incertitudes qu'elles peuvent engendrer des pertes financières considérables, non seulement pour les personnes assurées, mais également pour les assureurs. Ces risques extrêmes nécessitent la mise en place de mécanismes de réassurance adaptés. Le traité Excédent de Sinistre par Événement Catastrophe (ou XSC) constitue l'un de ces instruments de couverture, conçu pour protéger les assureurs contre des pertes exceptionnelles liées à des sinistres d'envergure. Cependant, la tarification de ces traités est complexe, en raison de la fréquence faible (qui conduit à l'imprévisibilité) et de l'extrême variabilité des événements qu'ils couvrent.

Le groupe Arundo Re (anciennement CCR-Re), acteur clé dans le domaine de la réassurance, développe des modèles avancés pour tarifer ces risques extrêmes. La tarification des événements catastrophiques est une problématique complexe, nécessitant une analyse approfondie de la fréquence et de la sévérité des sinistres. Le modèle probabiliste, Peaks Over Threshold (POT), issu de la théorie des valeurs extrêmes, offre un cadre solide pour analyser le comportement des événements rares et extrêmes. Il se focalise sur la queue de distribution, c'est-à-dire les observations qui dépassent un seuil élevé, au-delà duquel le comportement des données peut être approximé par une loi de Pareto généralisée (GPD). Autrement dit, ce modèle est utilisé en tarification pour s'appuyer sur une loi de Pareto généralisée aux sinistres les plus graves, c'est-à-dire ceux dont le nombre de victimes dépasse un seuil défini. L'efficacité de cette approche dépend cependant fortement du choix de ce seuil, qui constitue un enjeu central dans l'application pratique du modèle.

Notre modèle repose sur une approche dite fréquence – sévérité. Cela signifie que nous modélisons, d'un côté, le nombre d'événements catastrophiques ayant entraîné au moins un certain nombre de victimes sur une période donnée — généralement une année. De l'autre côté, on évalue la gravité de chacun de ces événements. Cette combinaison permet de mieux représenter les risques rares mais importants auxquels sont exposés les portefeuilles en réassurance vie

Toutefois, leur application nécessite des ajustements (par exemple : le regroupement de zones géographiques) pour améliorer la précision des estimateurs. Dans ce mémoire, nous proposerons des moyens d'amélioration pour optimiser notre modèle, notamment en déterminant un seuil optimal et en regroupant des données des pays. L'objectif est de mieux capturer la nature rare mais fortement impactante des événements extrêmes, et de les intégrer dans notre modèle de tarification.

Ce travail s'articulera autour des axes suivants :

- Présentation du rôle de la réassurance du groupe Arundo-Re ;
- Explication des caractéristiques des risques catastrophiques d'origine naturelle ou humaine ;
- Analyse détaillée du modèle POT ;
- Proposition d'améliorations pour optimiser la précision et la fiabilité du modèle ;
- Calibrage et Sensibilité du modèle.

1

Naissance du réassureur Arundo Re

L'histoire du réassureur Arundo Re commence avec l'établissement de la Caisse Centrale de Réassurance (CCR) en 1946. Détenue par l'État, CCR bénéficie de plus de 75 ans d'expertise en réassurance publique et de marché, se classant parmi les 30 plus grands réassureurs mondiaux. En tant que réassureur public, CCR propose, avec la garantie de l'État, des couvertures contre les catastrophes naturelles et les risques non assurables, au service des compagnies d'assurances en France. En 2016, le groupe CCR a créé un fidèle de réassurance de marché CCR Re. Son activité est la réassurance de portefeuilles non-vie, vie et spécialités opérant dans plus de 80 pays à travers le monde.

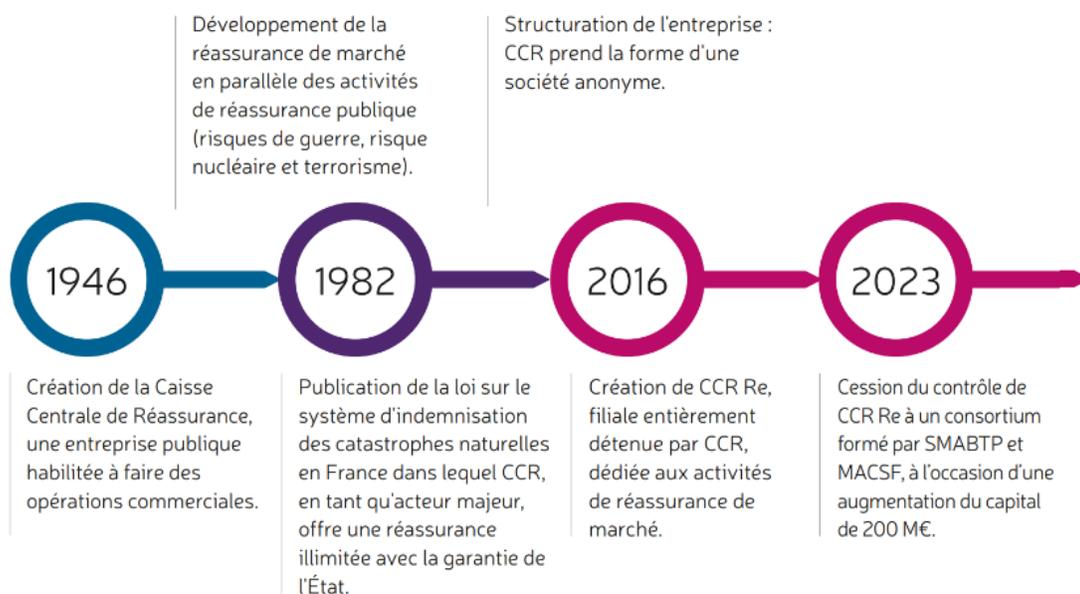


FIGURE 1 – Chronologie historique d'Arundo Re

En 2023, le groupe Caisse Centrale de Réassurance (CCR) a vendu une part majoritaire de CCR Re lors d'une levée de fonds de 200 millions d'euros, passant sous le contrôle d'un consortium composé de deux mutuelles françaises : SMABTP et MACSF. Ces dernières détiennent désormais 75% du capital, et CCR conserve une participation de 25% au capital. Cette opération financière fournit à CCR Re (qui deviendra *Arundo Re*) les ressources nécessaires pour stimuler sa croissance et pour répondre à la demande croissante en capacité des assureurs.

À partir de 2025, la société adopte le nom d'Arundo Re, marquant une nouvelle étape dans son développement international. Basée à Paris, Arundo Re est désormais un réassureur de portée mondiale, opérant dans plus de 100 pays. Spécialisée à la fois en réassurance non-vie et vie, la société souscrit des risques variés. En réassurance non-vie, elle couvre des branches traditionnelles telles que les dommages aux biens et la responsabilité civile, jusqu'aux risques les plus innovants comme le « risque cyber ». En réassurance vie, elle réassure les risques de prévoyance individuelle ou collective, tels que les contrats Credit Life ou Group Life. En santé, nous disposons d'un large portefeuille de produits comme « medical expenses » ou « critical illness », répartis selon différentes zones géographiques. Ces engagements peuvent être annuels ou par génération, avec une réassurance de cohortes allant jusqu'à l'extinction des polices d'assurance.

Arundo Re bénéficie depuis juillet 2023 d'une notation S&P A avec perspective stable et d'une notation AM Best A avec perspective stable.

Répartition par zone géographique

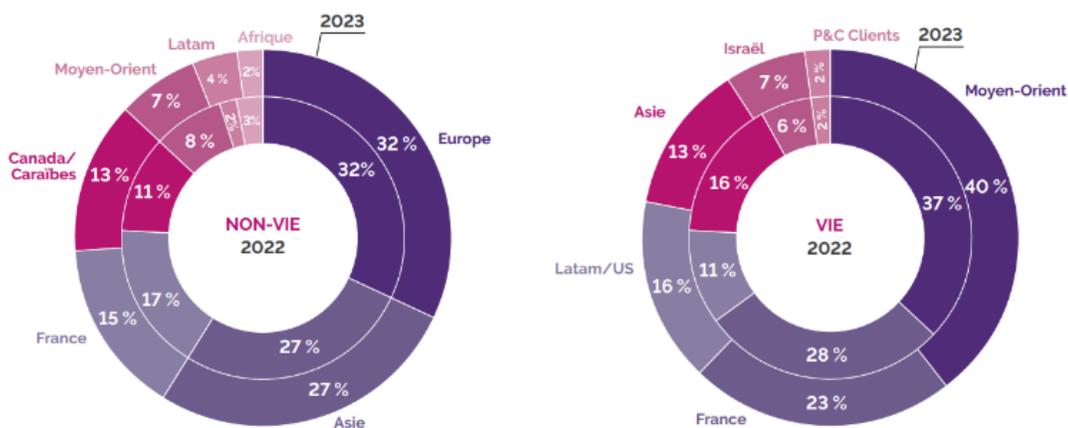


FIGURE 2 – Répartition géographique des activités de réassurance non-vie et vie (2022 vs 2023)

Entre 2022 et 2023, la répartition géographique des activités d'Arundo Re montre une relative stabilité sur ses principaux marchés, avec des ajustements ciblés. En non-vie, l'Europe et l'Asie conservent des parts dominantes (32 % et 27 %), tandis que la France connaît une légère hausse, témoignant d'un renforcement local. En vie, le Moyen-Orient progresse (de 37 % à 40 %), soulignant son rôle croissant, alors que la France recule légèrement.

Répartition par branche

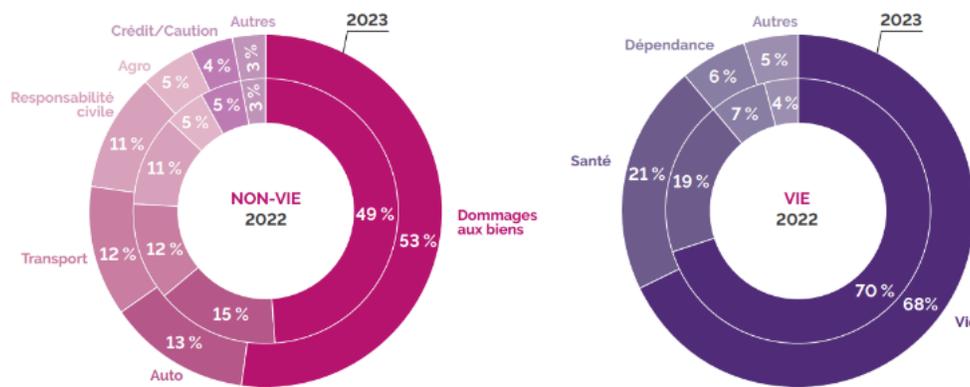


FIGURE 3 – Répartition du chiffre d'affaires au titre de l'exercice (2022 vs 2023)

Selon le rapport du chiffre d'affaires de la société, le chiffre d'affaires est de 1 186 M€, en hausse de 20% par rapport à 2022. le segment non-vie représente une part prédominante de 69% (815 M €) du chiffre d'affaires total, témoignant de sa forte présence dans ce secteur. La branche dommage est la plus importante, représentant environ 53% du chiffre d'affaires du segment non-vie en 2023. En assurance vie, la branche prévoyance est très largement majoritaire avec 70% du chiffre d'affaires. La seconde branche la plus importante est la branche santé avec 21% du chiffre d'affaires regroupant les produits "Medical expenses" et "Critical Illness". Cette répartition par branche de l'activité vie et non-vie met en lumière la diversité de notre portefeuille, adapté aux besoins des branches différents marchés.

2

Présentation du risque catastrophe

L'origine du terme « catastrophe » remonte au théâtre grec, où il était désigné par « katastrophē », un mot formé à partir de « kata- », signifiant « vers le bas », et « strophē », signifiant « tourner ». Initialement, ce terme décrivait le dénouement ou la fin dramatique d'une pièce, marquant souvent une issue tragique. Au fil du temps, la signification de « catastrophe » a dépassé du théâtre antique pour embrasser une définition plus large, référant désormais à des événements abrupts qui provoquent d'importantes perturbations ou des destructions massives.

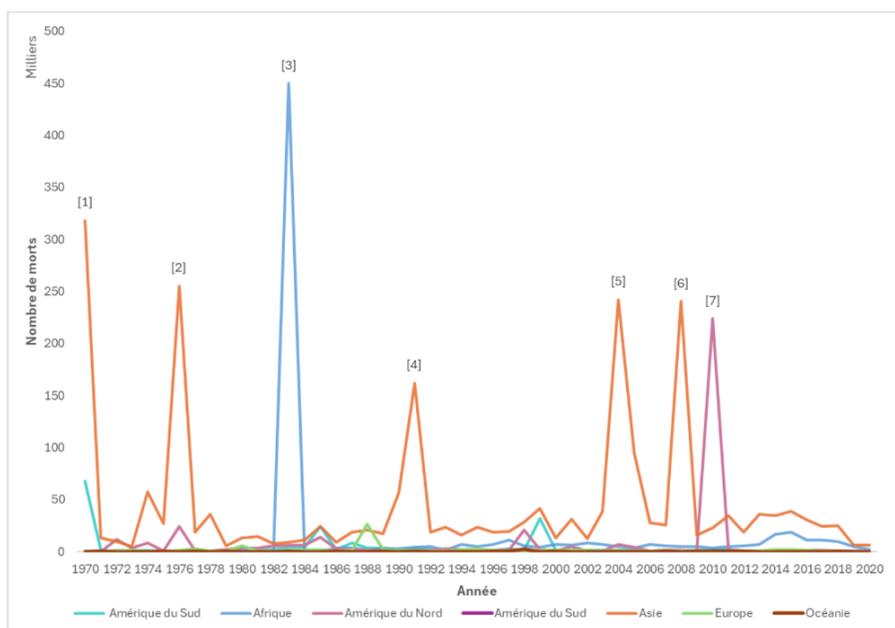


FIGURE 4 – Tendances de la mortalité par continent pour la période 1970 - 2020

La catastrophe la plus meurtrière :

- [1] Le cyclone Bhola (Bangladesh - Indes) en 1970 ;
- [2] Le séisme de Tangshan (Chine) en 1976 ;
- [3] La sécheresse en Éthiopie en 1983 ;
- [4] Le cyclone tropical qui a frappé le Bangladesh en 1991 ;
- [5] Le tsunami de l'océan Indien, survenu le 26 décembre 2004 ;
- [6] Le séisme de mai 2006 en Indonésie, et le typhon Durian, qui a frappé les Philippines fin novembre ;
- [7] Le séisme en Haïti, survenu le 12 janvier 2010.

Le graphique mentionne le nombre de décès dus aux catastrophes, en distinguant les continents. Il donne une vision globale de l'existence et de l'impact des événements catastrophiques à travers l'histoire. Par exemple, le cyclone Bhola, qui a touché le Bangladesh et l'État indien du Bengale occidental le 13 novembre 1970, illustre l'ampleur de ces catastrophes (voir le pic [1] sur le *figure 4*). Il est estimé que ce cyclone a causé la mort de 300 000 à 500 000 personnes. En 1983, l'un des événements marquants en Afrique, noté sur le graphique, est la reprise de la guerre civile au Soudan, appelée Seconde Guerre civile soudanaise.

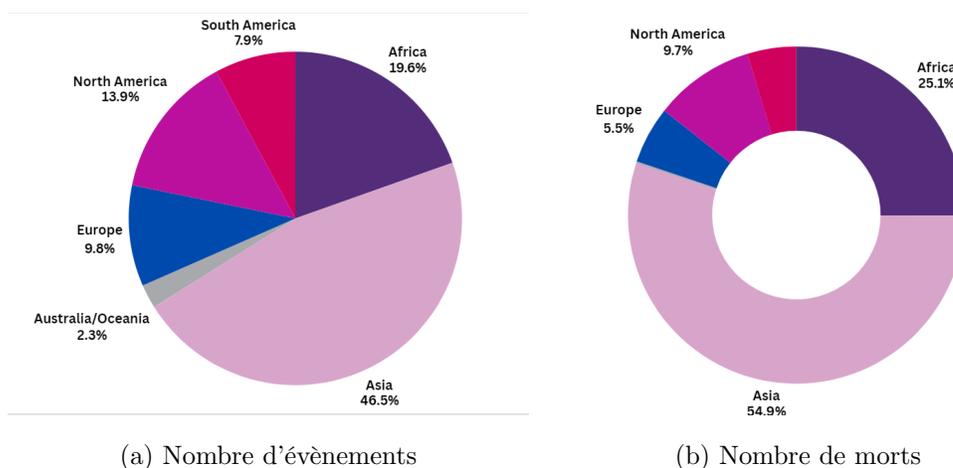


FIGURE 5 – Présentation des impacts des catastrophes de 1970 à 2020

En évoquant le nombre d'événements catastrophiques et des décès survenus durant les cinquante dernières années depuis 1970 , il apparaît que l'Asie est le continent le plus touché en nombre d'évènements et en nombre de morts.

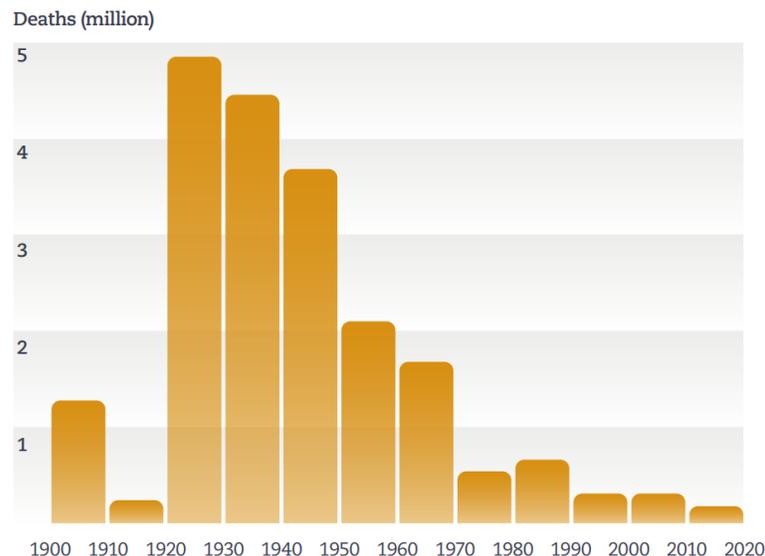
Dans « une catastrophe », nous pouvons distinguer l'impact direct de l'impact indirect. « Les impacts directs » sont les dommages physiques sur les infrastructures (transports, bâtiments, e.t.c) et sur les personnes (blessés ou décès). En revanche, « les impacts indirects » sont une série de conséquences non matérielles comme « le mass-trauma », la mise à l'arrêt des entreprises, ayant des impacts macroécono-

miques.

Ces catastrophes peuvent être regroupés en deux catégories : les catastrophes d'origine naturelles et les catastrophes d'origine humaines.

2.1 Catastrophe d'origine naturelle (Cat-nat)

EM-DAT (The Emergency Events Database) est une base de données internationale qui recense les catastrophes naturelles et technologiques majeures survenues dans le monde depuis 1900. L'EM-DAT - créé par le soutien initial du Centre de recherche sur l'épidémiologie des catastrophes (CRED), situé en Belgique - définit les catastrophes comme des événements imprévus et soudains qui causent des dommages importants, des destructions ou de lourdes souffrances humaines et nécessitent une demande d'aide au niveau national ou international. Nous pouvons citer par exemple des tempêtes de 1999 en Europe. Deux tempêtes exceptionnelles, appelées Lothar et Martin, ont eu lieu dans le Nord de l'Europe (Irlande, l'île britanniques et France) à la fin du mois de décembre 1999, avec des vents dépassant 150 km/h dans certaines régions. Elles ont entraîné des dégâts matériels importants (estimés à 19,2 milliards de dollars), des coupures d'électricité massives et la mort de 140 personnes.



Source : EM-DAT report en 2020

FIGURE 6 – Nombre de décès par décennie (1900 - 2020)

Comme son nom l'indique, les catastrophes liées à la nature sont causées par des phénomènes naturels tels que les inondations, les séismes, les éboulements, les avalanches, etc. Concernant la raison de ces événements, elles sont réellement les conséquences des processus géologiques ou météorologiques. Bien que les scientifiques aient fait des avancées dans la détection et la surveillance de certaines catastrophes naturelles, il demeure une impossibilité en ce qui concerne la prévisibilité de ces événements.

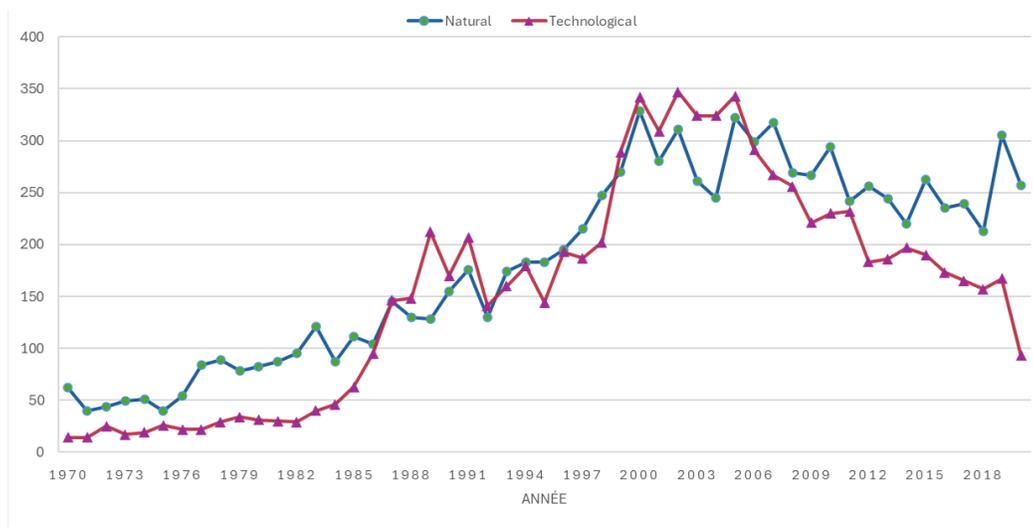
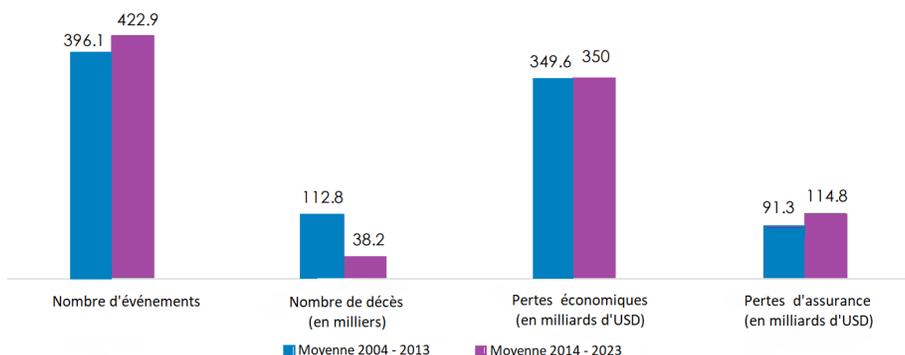


FIGURE 7 – Nombre d'événements dûs à des catastrophes de 1970 à 2020

Le figure 7 ci-dessus illustre l'évolution du nombre d'événements, qu'ils soient d'origine naturelle ou d'origine humaine. L'augmentation des événements naturels s'explique principalement par des facteurs tels que le changement climatique, la croissance démographique mondiale, ainsi que la concentration croissante dans des zones exposées aux risques.

La période 2014 - 2023 est une décennie de catastrophes naturelles :

Les catastrophes naturelles ont engendré des pertes économiques annuelles moyennes de 350 milliards USD au cours de la période 2014 - 2023. En termes de coût financier, l'année 2017 reste la plus coûteuse de cette décennie, avec 603 milliards USD de pertes. Cette année-là, trois ouragans majeurs - Harvey, Irma et Maria - ont dévasté les côtes américaines, provoquant à eux seuls des pertes économiques record de 292 milliards USD, dont 93 milliards USD pris en charge par le marché de l'assurance.



Source : Aon-Climate and Catastrophe insight

FIGURE 8 – Chiffres clés des catastrophes naturelles des 2 dernières décennies

La période 2014 - 2023 a également été marquée par le plus grand nombre de victimes humaines en 2023, avec un total de 95 000 décès. Parmi les événements les plus meurtriers figurent trois tremblements de terre majeurs au Maroc, en Turquie et en Syrie, causant 62 272 décès. De plus, plusieurs vagues de chaleur entre juillet et septembre 2023 ont particulièrement touché l'Europe de l'Ouest et du Sud, entraînant environ 15 000 décès (Selon Atlas Magazine).

	Drought	Earthquake	Extreme temperature	Flood	Landslide	Mass movement (dry)	Storm	Volcanic activity	Wildfire
Africa	695 081	7 130	372	26 710	2 841	222	6 828	2 199	335
Asia	5 477	923 307	25 861	239 184	22 086	1 498	749 533	2 310	861
Australia/Oceania	84	3 164	508	480	577	86	1 245	30	336
Europe	2	36 298	154 592	4 905	1 561	110	2 763	9	760
North America	86	269 297	4 604	12 228	4 421	100	53 832	709	389
South America	44	77 432	2 522	44 913	9 048	307	1 564	21 837	119

FIGURE 9 – Nombre de décès liés aux catastrophes naturelles par type d'événement et par continent

Les caractéristiques naturelles de chaque continent déterminent les types de catastrophes auxquelles il est le plus exposé. Ainsi, l'Afrique, caractérisée par ses vastes étendues désertiques, est principalement affectée par les sécheresses. En revanche, la vaste diversité géographique de l'Asie, incluant ses variations climatiques, sa topographie, et son activité tectonique, la rend vulnérable à de nombreux aléas naturels tels que les inondations, les tremblements de terre et les tsunamis.

2.2 Catastrophe d'origine humaine (man-made catastrophe)

Parmi les événements catastrophiques, deux catégories sont d'origine humaine : les actes terroristes et les accidents technologiques. Un acte terroriste se caractérise généralement par un ou plusieurs actes violents organisés par des individus ou des groupes, visant à instaurer un climat de peur au sein d'un groupe spécifique ou de la population en général. Par le passé, ce type de risque était exclu des garanties d'assurance en raison des difficultés d'évaluation et de gestion. Aujourd'hui, l'intégration des risques terroristes dans les contrats d'assurance résulte principalement de la demande croissante des clients, soucieux de se protéger face à ces menaces.

Au contraire, les catastrophes liées à la technologie résultent souvent d'accidents industriels, nucléaires ou biologiques. Bien que les progrès technologiques aient révolutionné notre vie quotidienne, ils augmentent également les risques d'accidents et de menaces potentielles. De même, les technologies industrielles et nucléaires, bien qu'indispensables, peuvent avoir des conséquences catastrophiques en cas de défaillance. La figure 7 montre l'augmentation progressive des événements catastrophiques d'origine humaine au fil des années, suivie d'une diminution notable à partir de 2005.

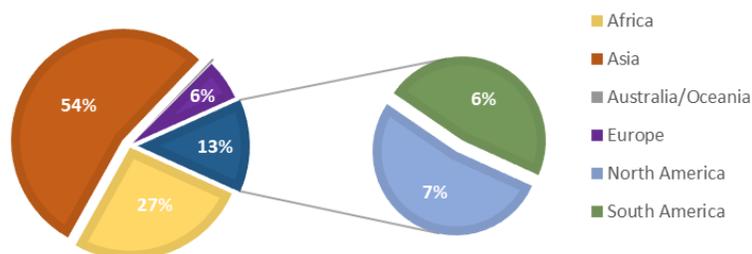


FIGURE 10 – Répartition des décès par origine dus au terrorisme entre 1970 et 2020

	Industrial accident	Miscellaneous accident	Terrorisme	Transport accident
Africa	7 066	7 065	119 475	67 446
Asia	31 436	28 913	248 381	102 648
Australia/Oceania	51	124	179	1 775
Europe	4 731	5 354	10 199	24 364
North America	2 072	4 294	32 121	15 653
South America	2 652	4 354	26 537	14 880

FIGURE 11 – « Man-made » événements catastrophiques par continent

À l'instar des catastrophes naturelles, les catastrophes provoquées par l'homme se produisent majoritairement en Asie, qui dispose d'un grand nombre d'industries. C'est «Usine du monde ». On peut citer la catastrophe de l'usine nucléaire de Fukushima (Japon, 2011), qui a entraîné l'évacuation de milliers de personnes.

Risque de Terrorisme

Le risque terroriste reste un enjeu majeur dans certaines régions du monde, notamment au Moyen-Orient, en Afrique du Nord et en Asie du Sud. Ces événements, bien que moins fréquents que les catastrophes naturelles ou industrielles, peuvent avoir un impact humain considérable. On peut citer, par exemple, les attentats de Bombay en 2008 ou ceux de Kaboul à répétition, qui ont causé plusieurs centaines de victimes. Pour étudier ce type de risque, nous nous appuyons sur la base de données reconnue **Global Terrorism Database (GTD)**, qui recense les actes terroristes dans le monde depuis 1970.

Risque de transport

Les catastrophes de transport représentent également une part importante des sinistres d'origine humaine, en particulier dans les zones à forte densité de population et de circulation. Les accidents aériens, ferroviaires ou maritimes peuvent entraîner des bilans humains lourds. Parmi les événements marquants, on peut citer le naufrage du ferry Sewol en Corée du Sud (2014), ayant causé la mort de plus de 300 personnes, en majorité des lycéens.

3

Le monde de la réassurance

La réassurance est un acteur important de l'industrie de l'assurance. Elle permet aux assureurs de transférer une partie de leurs risques à un réassureur via des structures non-proportionnelles (risque de perte par risque « XST », risque venant de catastrophe « XSC ») ou des structures proportionnelles comme le Quote-part ou l'Excédent de Plein. La réassurance permet aux assureurs de limiter leur exposition aux risques, de protéger le bilan de la compagnie et d'augmenter leur capacité à émettre des polices d'assurance.

3.1 Rôle de la réassurance

Une compagnie d'assurance est exposée à un risque d'exploitation élevé, renforcé par la nature aléatoire et souvent asymétrique de ses engagements. Elle doit non seulement garantir sa compétitivité, comme tous les types d'entreprises, mais aussi respecter ses exigences réglementaires de solvabilité, qui traduisent sa capacité à faire face à ses engagements en toutes circonstances. Bien que les primes soient perçues avant la survenance des sinistres, ce décalage temporel ne protège pas nécessairement l'assureur contre l'ampleur potentielle des pertes futures.

Dans un scénario, si une série d'événements survenait simultanément ou dans un intervalle de temps restreint, l'assureur pourrait subir des pertes importantes mettant en péril sa solidité financière. La réassurance offre une solution à ce problème, car elle permet à l'assureur de récupérer une partie des sinistres en excédent de sinistre. Elle contribue ainsi à stabiliser les sinistres importants pour la cédante, ce qui permet de mieux les évaluer et de réduire le risque de constitution de réserves de fonds propres insuffisantes face aux catastrophes.

En outre, la réassurance contribue à la protection du bilan des assureurs par faire face à des pics

exceptionnels, tels que l'explosion de l'usine AZF en 2001 à Toulouse entraînant la mort de trente et une personnes et faisant deux mille cinq cents blessés. Grâce à des traités comme le Stop-Loss, elle permet de lisser les résultats techniques sur plusieurs exercices, évitant ainsi une dégradation brutale du résultat net. Enfin, en réduisant son exposition nette, l'assureur libère du capital qu'il peut redéployer pour souscrire de nouveaux contrats sans augmenter son besoin de fonds propres.

3.2 Marché de la réassurance

Chiffres en millions USD

	2019	2020	2021	2022	2023	Evolution 2022-2023
Primes émises brutes	311 588	347 699	385 695	363 477	378 543	4,14%
Primes émises brutes non vie	216 188	235 399	264 295	256 752	279 069	8,69%
Primes émises brutes vie	95 400	112 200	121 400	106 725	99 474	-6,79%
Résultat net	40 812	18 447	46 369	13 960	77 682	456,46%
Fonds propres	869 155	942 173	1 009 632	891 936	1 060 643	18,91%
Ratio combiné en %	100,10%	104,50%	96,40%	96,20%	91,50%	5,13%
ROE (Return on Equity) en %	9,40%	2,40%	9,00%	2,50%	21,40%	856%

Source : AM Best, ratios des principaux réassureurs selon les normes IFRS 17

FIGURE 12 – Chiffre d'affaires global sur le marché de la réassurance

Croissance globale du marché de la réassurance entre 2019 et 2023

Entre 2019 et 2023, le marché de la réassurance a connu une croissance soutenue, avec une augmentation des primes brutes de 4,14%, atteignant 378 543 millions USD en 2023. On observe une hausse significative des primes émises brutes non-vie (8,69%), tandis que les primes vie accusent un recul notable (-6,79%). Malgré cette baisse, le résultat net progresse fortement en 2023 (456,46%), porté par une amélioration du ratio combiné (91,5%) et un retour sur fonds propres (ROE) en nette hausse à 21,4% (856%). Cela illustre une amélioration notable de la rentabilité globale des réassureurs.

Performance différenciée entre les branches vie et non vie

En 2023, les réassureurs non-vie, représentés par 136 compagnies, enregistrent un volume de primes de 279 milliards USD contre 256,7 milliards USD en 2022. L'activité non-vie constitue 73,7% du total du

marché (vie et non vie) en 2023, contre 70,6% en 2022, confirmant la prépondérance croissante de ce segment. Cette branche affiche une hausse des primes de 8,69% en 2023, tandis que la branche vie subit une baisse de 6,79%, passant de 106,725 milliards USD en 2022 à 99,475 milliards USD en 2023.

La réassurance vie, bien qu'elle repose souvent sur des contrats à court ou moyen terme, est fortement influencée par les évolutions économiques en raison de la nature sensible des risques couverts (longévité, mortalité, santé). Les facteurs comme l'inflation persistante, la volatilité des marchés financiers et les tensions géopolitiques influencent directement les comportements des souscripteurs, qui deviennent plus prudents dans leurs choix.

La reprise post-pandémie COVID 2019 demeure lente, aggravant la situation pour la branche vie. Celle-ci a été lourdement affectée par la hausse significative de la mortalité liée au Covid-19. Cette situation a renforcé les défis auxquels les réassureurs vie doivent faire face.

Concernant les marchés du monde :

Chiffres en millions USD

	Nombre de réassureurs (1)	2014	2018	2022	2023	Evolution 2014-2023 (2)	Parts 2023
Europe	26	104 205	123 721	155 194	164 287	5,20%	43,40%
Lloyd's	1	13 185	14 048	18 527	22 072	5,90%	5,80%
Amérique et Bermudes	39	61 464	86 150	148 790	152 193	10,60%	40,20%
Asie	22	23 792	38 623	33 496	32 395	3,49%	8,60%
Afrique	48	3 690	4 471	5 815	5 712	4,97%	1,50%
Moyen-Orient	7	958	1 047	1 655	1 884	7,80%	0,50%
Total	143	207 294	268 060	363 477	378 543	6,92%	100%

(1) Le nombre de compagnies de réassurance étudiées est le même durant toute la période

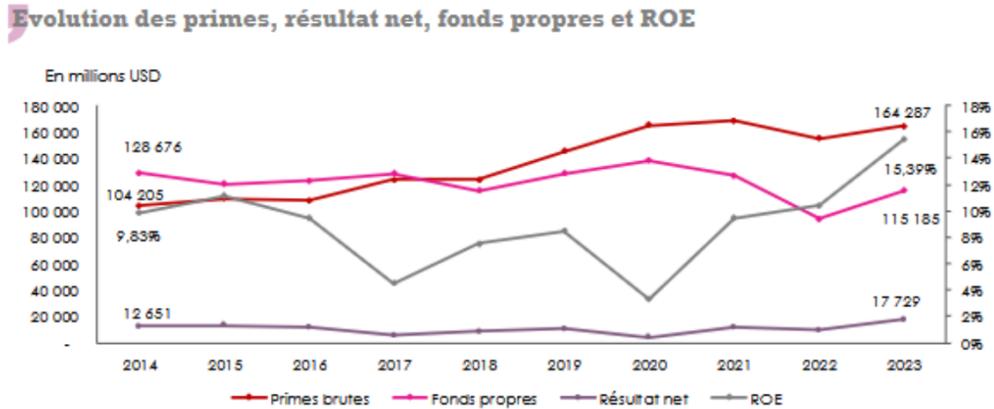
(2) Taux de croissance annuel moyen

Source : AM Best, ratios des principaux réassureurs selon les normes IFRS 17

FIGURE 13 – Répartition du chiffre d'affaires selon les marchés de la réassurance

D'après Atlas Magazine (l'Actualité de l'assurance dans le monde), l'Europe, incluant les Lloyd's, est le leader du marché mondial de la réassurance en 2023, représentant 49,2% des primes souscrites. Elle est suivie par les États-Unis et les Bermudes, qui occupent les positions suivantes. Toutefois, c'est la région

Amérique et Bermudes qui affiche la plus forte croissance au cours des dix dernières années.



Source : L'Actualité de l'assurance dans le monde

FIGURE 14 – Evolution des primes, résultat net, fonds propres et ROE sur la période 2014 - 2023

Le marché européen de la réassurance

Le marché européen de la réassurance évolue dans un environnement complexe, marqué par des défis économiques, climatiques et technologiques. Les quatre leaders du secteur, Munich Re, Swiss Re, Hannover Re et SCOR, occupent une place de premier plan tant en Europe et qu'échelle mondiale. En 2023, leur chiffre d'affaires cumulé atteint 131,3 milliards USD, soit 79,9% des souscriptions européennes et 34,7% des souscriptions mondiales. Le marché reste fortement dominé par la réassurance non-vie, qui représente 69% des primes souscrites cette même année, reflétant l'importance des branches telles que l'incendie, les dommages aux biens et la responsabilité civile.

3.3 Types de réassurance

Des modes de cession d'assurance vers la réassurance :

Les modes des cessions peuvent être proportionnels ou non-proportionnels, selon la manière dont les primes et les sinistres sont partagés entre l'assureur et le réassureur. Les contrats de réassurance peuvent également être classés en fonction du type de risque couvert, tels que la responsabilité civile.

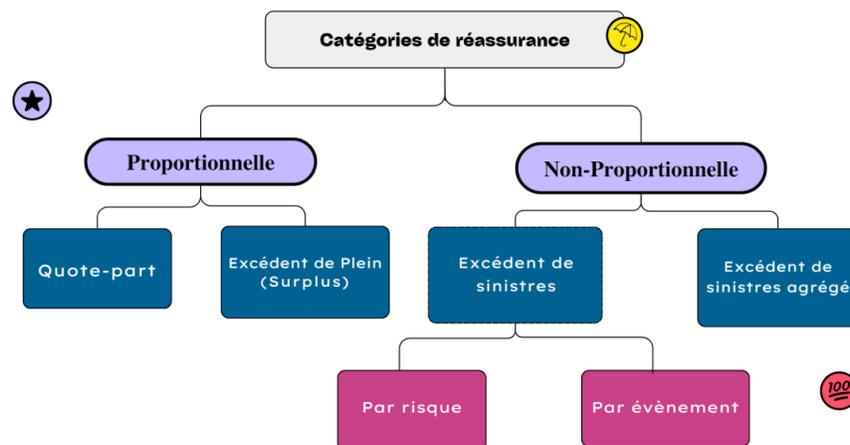


FIGURE 15 – Types de réassurance

Les acceptations : "Facultatives", "Obligatoires" et "FACOB" :

L'offre de couvertures de réassurance est aussi variée que les besoins et évolutions des assureurs. Il existe 2 types d'acceptation de réassurance : l'acceptation facultative et l'acceptation obligatoire. Pour **un contrat de réassurance obligatoire**, l'assureur doit céder tout son risque entrant dans le périmètre défini au contrat, et le réassureur doit accepter ce risque sans condition. En revanche, pour **un contrat de réassurance facultative**, la cession dépend du choix ou de la volonté de chaque partie. La cession facultative demeure plus souple : les parties peuvent négocier librement sur le risque considéré, c'est-à-dire, l'assureur étant libre de céder ce risque, le réassureur libre de l'accepter.

Parallèlement, il existe un autre mode de réassurance, la facultative - obligatoire, qui se distingue par une approche de cession où l'assureur sélectionne spécifiquement les risques à transférer au réassureur. Ce dernier est contraint d'accepter ces risques, sans possibilité de refus. Cette méthode présente un risque pour le réassureur, particulièrement si l'assureur transmet des risques de moindre qualité : Cela expose le

réassureur au risque d'anti-sélection.

La co-réassurance désigne un mécanisme par lequel plusieurs réassureurs se partagent un même risque ou portefeuille. Le réassureur apériteur sera le leader des réassureurs : c'est lui qui aura la possibilité de définir les conditions et les clauses du contrat de réassurance (entre partie d'assurance et partie de réassurance). Ce refus ou cette acceptation engagera en revanche l'ensemble des réassureurs suiveurs, ce qui n'est pas sans conséquence pour ces derniers. Ce mécanisme permet de répartir les grands risques sur plusieurs entités pour éviter qu'un seul réassureur ne porte un fardeau trop important.

Traité de réassurance proportionnelle :

D'une part, dans le cadre d'un traité de réassurance proportionnelle, la cédante partage ses polices avec le réassureur, en cédant un pourcentage $x\%$ des primes perçues. En échange, le réassureur s'engage à souscrire la même proportion $x\%$ des risques et à payer les sinistres correspondants.

Le Quote-Part

Dans un **Traité en Quote-Part**, le plus commun, un taux de cession de $x\%$ est défini entre 2 parties.

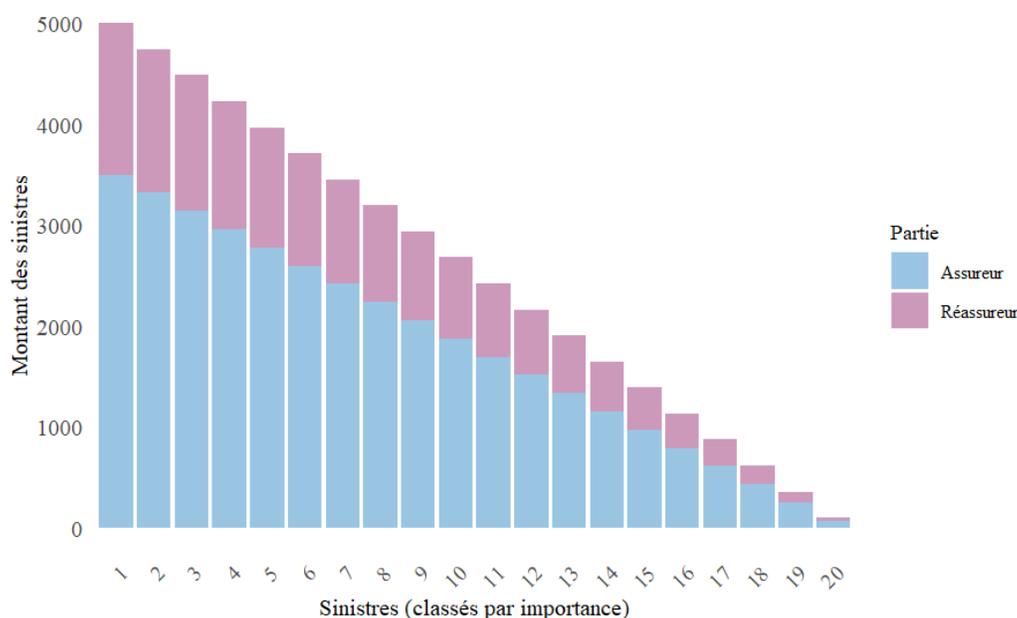


FIGURE 16 – Répartition des sinistres dans un traité de réassurance en quote-part (30%)

Dans ce type de réassurance en Quote-Part, le taux de cession est identique pour toutes les personnes dans le même portefeuille. Ce mécanisme de répartition garantit une couverture proportionnelle et constante, quel que soit le niveau du sinistre. Toutefois, certains traités de quote-part peuvent inclure une limite maximale de responsabilité, fixant un plafond au-delà duquel le réassureur n'intervient plus, laissant l'assureur prendre en charge le reste des sinistres.

On distingue ainsi plusieurs formes de traités en quote-part :

Quote-part financière :

Ce type de traité vise principalement à soulager les contraintes de solvabilité ou de capacité financière de la cédante. Le réassureur peut prendre en charge une part significative du portefeuille — pouvant aller jusqu'à 90% ou plus — ce qui permet à l'assureur direct de libérer du capital réglementaire tout en partageant les résultats techniques et la rentabilité du portefeuille.

Quote-part de développement de produit :

Ce type de traité est souvent utilisé lorsqu'une compagnie d'assurance lance un nouveau produit ou une nouvelle activité. Il permet de partager les risques d'un portefeuille encore jeune ou sans assez de recul. Dans ce cas, la compagnie d'assurance et le réassureur travaillent ensemble de manière rapprochée pour suivre les résultats et améliorer le produit si besoin.

Le Traité Vie de Génération (TVG) :

Ce mécanisme spécifique de réassurance est destiné à couvrir les risques liés à une génération particulière de contrats d'assurance vie. Contrairement aux traités annuels, le TVG offre une protection à long terme pour une groupe de personnes définie, permettant aux assureurs de mieux gérer les risques liés à la mortalité et à la morbidité. Le TVG est particulièrement adapté aux portefeuilles vie incluant des garanties décès, invalidité ou épargne, car il offre une gestion plus prévisible des engagements, tout en permettant des ajustements périodiques en fonction des évolutions actuarielles.

Dans les traités de réassurance en quote-part, une clause importante figure souvent dans le contrat : **Clause de participation aux bénéfices (PB)**. Cette clause permet à la cédante de recevoir une part des bénéfices du traité lorsque son résultat de réassurance est positif. La participation peut être exprimée en pourcentage des bénéfices réalisés et est souvent plafonnée pour garantir un équilibre entre les parties. Le résultat de la réassurance comprend le paiement des sinistres ainsi que les frais de gestion de réassureur.

L'Excédent de Plein

Quant à un **Traité en Excédent de Plein (Surplus)**, il existe le plein de rétention comme le montant maximal que l'assureur souhaite payer par sinistre.

$$x = \min \left[\max \left(\frac{\text{capital assurée} - \text{plein de rétention}}{\text{capital assuré}} ; 0 \right) ; \frac{\text{capacité de l'excédent de plein}}{\text{capital assuré}} \right]$$

Où :

- x : la proportion de la somme assurée prise en charge par le réassureur.
- Capital assurée : le montant réel de sinistre
- Plein de rétention (la franchise) : le montant activant la partie du réassureur
- Capacité de l'excédent de plein : le montant maximum que le réassureur est disposé à accepter pour un risque ou une police spécifique au-delà de la limite de rétention de l'assureur primaire.

Quand le capital assuré est plus petit que le Plein de rétention, le taux de cession est nul. L'excédent de plein permet donc de conserver la plus grande partie de sa prime tout en bénéficiant d'une protection proportionnelle pour ses risques les plus importants.

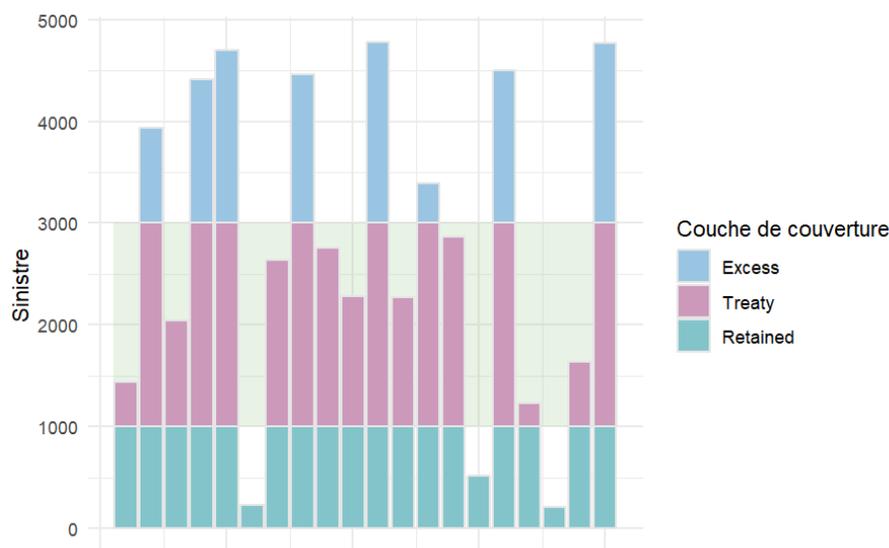


FIGURE 17 – Illustration de traité de Excédent de Plein : 2000 XS 1000

Chaque sinistre provenant du portefeuille décrit dans l'image ci-dessus sera payé si leur valeur est supérieure à 1000 (nous l'appelons 'plein' ou capacité) et le réassureur paiera au maximum 2000 pour chaque sinistre. Dans ce type de réassurance, la portée est normalement exprimée en nombre de pleins. Par

exemple, si une ligne de rétention est fixée à 1000 d'euros, le réassureur accepte de couvrir jusqu'à trois fois cette ligne, soit jusqu'à 3000 d'euros par sinistre. Ce format permet de structurer les engagements entre les assureurs et les réassureurs, ce qui rend les arrangements plus clairs et plus quantifiables. Chaque 'plein' constitue une unité de risque retenue d'abord par l'assureur, que le réassureur prend ensuite en charge jusqu'à une limite spécifiée.

Traité combiné Quote-Part et Excédent de Plein :

Dans ce traité, une part fixe de chaque risque (par exemple 30%) est cédée au réassureur via *la quote-part*. La part excédentaire, correspondant à la somme assurée dépassant un point défini (appelé «plein»), est ensuite réassurée par tranches successives dans le cadre d'un traité en excédent de plein.

Cette structure permet à une cédante de conserver une flexibilité dans la gestion de ses engagements, en mutualisant les risques selon leur taille et en optimisant la capacité de souscription sur les grands contrats.

Traité de réassurance non-proportionnelle :

Contrairement aux contrats de réassurance proportionnelle, dans les contrats de réassurance non-proportionnelle, il n'y a pas de relation de proportionnalité entre la prime cédée et la participation du réassureur aux sinistres à payer.

La réassurance permet aux assureurs de se protéger contre différents types de risques. Le risque de pointe peut être couvert par des contrats d'excédent de sinistre par tête (Excess per Head), tandis que le risque événementiel est pris en charge par des traités d'excédent de sinistre catastrophe (Excess Cat). Quant au risque de fréquence, il peut être réassuré à l'aide de structures non proportionnelles de type Stop-Loss ou excédent agrégé (Aggregate Excess).

L'Excédent de sinistre

L'**Excédent de sinistre** fonctionne avec 2 paramètres : la franchise ou priorité et la portée. Le réassureur s'engage à payer pour tous les sinistres dépassant la priorité dans la limite de la portée.

Si on note S le sinistre à charge du réassureur, **le montant x** cédé à la réassurance est définie comme :

$$x = \min [\max (X - \text{priorité} ; 0) ; \text{portée}]$$

Où : X est le montant réel pour chaque sinistre S (chaque événement dépendant de la condition du traité)

Un traité XS peut fonctionner « par risque », par « événement » ou « par risque et par événement » :

- Par risque : le traité s'applique aux sinistres liés à une seule police.
- Par événement : le traité s'applique au montant cumulé de tous les sinistres résultant du même événement. En fait, ce type de réassurance a principalement pour but de protéger la cédante de la survenance d'événements catastrophiques.
- Par risque et par événement : une combinaison des deux approches précédentes, qui permet d'affiner la couverture en tenant compte à la fois du sinistre unitaire et de l'accumulation liée à un événement

Des clauses supplémentaires :

En particulier, dans le contexte des traités excédent de sinistre, il y a 2 termes clés, ce sont *l'Aggregate Annual Limit (AAL)* et *l'Annual Aggregate Deductible (AAD)*. Alors que l'AAD est la franchise annuelle cumulative en français avant que le réassureur paye, l'AAL est la limite maximale de remboursement que le réassureur s'engage à payer pour les sinistres survenus.

Par exemple : Détail du traité XS par tête : 10.000 XS 10.000 euros

Annual Aggregate Deductible (AAD) : 10.000 euros.

Annual Aggregate Limit (AAL) : 20.000 euros.

Le processus de traitement des sinistres au cours d'une année est le suivant :

Sinistre	Montant	Retention	Couverture par traité	Excès
1	5.000	5.000	0	0
2	12.000	10.000	2.000	0
3	20.000	10.000	10.000	0
4	25.000	10.000	10.000	5.000

TABLE 1 – Illustration de traité de l'Excédent de sinistre

Explication de la fonction du traité :

- Somme de tous les montants couverts par le traité : $2.000 + 10.000 + 10.000 = 22.000$;
- Ce montant dépasse l'AAD (10 000 €), le réassureur commence donc à payer (12.000) ;
- Le montant total à la charge du réassureur est de 12 000 (inférieur à l'AAL : 20.000).

L'Excédent de Sinistre Agrégé (ou Stop Loss)

Ce type de traité permet à l'assureur de se protéger contre l'accumulation de pertes sur une année donnée, en apportant une couverture adaptée aux risques de fréquence élevée. Il est souvent utilisé dans les contextes où les sinistres répétés peuvent fragiliser la rentabilité. Elle représente une bonne solution pour les assureurs qui souhaitent une protection supplémentaire contre l'éventualité de plusieurs sinistres affectant leurs activités au cours de la même année de survenance (le risque fréquence comme la mortalité avec une pandémie).

Dans ce type de traité, les limites peuvent être définies de deux manières. Lorsqu'elles sont exprimées en montant absolu, on parle classiquement de **traité d'Excédent de sinistre agrégé**. En revanche, lorsque la couverture est déclenchée sur la base d'un ratio de sinistralité dépassant un certain seuil — formulé en pourcentage des primes acquises — le traité est qualifié de **Stop - Loss**.

Ce type de réassurance est utilisé pour non seulement protéger le bilan de l'assureur mais aussi répondre

aux exigences de solvabilité II. Dans les années exceptionnellement défavorables où une pandémie ou un événement catastrophique sévère se produit, le traité de réassurance offre une stabilisation financière en faisant face aux importants sinistres accumulés au cours d'une période donnée. Par conséquent, l'effet attendu d'un traité Stop Loss serait généralement une réduction du Capital de solvabilité requis (SCR - « Solvency Capital Requirement » en anglais).

3.4 Traité excédent de sinistres par évènement catastrophe (XSC)

Pour rappel, le traité Excédent de Sinistre par Catastrophe (XSC) constitue une protection contre des sinistres affectant simultanément plusieurs polices d'assurance dans un portefeuille, résultant d'une cause unique (« Natural perils » ou « man-made » event) et survenant dans un espace et une durée limitée. Ce type de réassurance non proportionnelle intervient uniquement après l'épuisement des couvertures excédent de sinistre par risques (XoL par risque).

Un traité de réassurance est composé de Conditions Générales et Particulières, qui précisent notamment la nature du contrat, son objectif, les risques couverts, les limites de l'engagement du réassureur et les exclusions. Ces précisions varient d'un traité à l'autre et reflètent les dynamiques du marché de l'assurance.

3.4.1 Terminologies générales

Territorialité : Les traités XSC offrent généralement une couverture mondiale, mais uniquement pour les risques liés aux polices souscrites dans la zone du pays d'origine du assuré. Toutefois, certaines conventions permettent une extension de la couverture à d'autres territoires selon les accords entre la cédante et le réassureur.

Les garanties couvertes : Les garanties offertes par ces traités sont généralement regroupées en deux principales :

- **Le risque décès ;**
- **Le risque Invalidité - Incapacité** , incluant notamment l'Invalidité Permanente Totale (IPT ou Permanent Total Disability ou PTD en anglais), définie comme une situation où l'assuré est définitivement incapable d'exercer une activité professionnelle.

Dans certaines régions, notamment en zone MENA (Moyen-Orient et Afrique du Nord), la couverture peut être étendue à des risques complémentaires tels que les frais médicaux.

L'évènement « catastrophe » est défini comme un sinistre d'origine soudaine et extérieure, entraînant des conséquences graves, principalement le décès et l'invalidité. L'évènement catastrophique est défini dans le contrat avec les conditions suivantes :

1. Le nombre minimum de victimes assurées pour activer la couverture. Il varie selon les contrats, mais il est généralement fixé à deux ou trois victimes ;
2. La durée standard maximale d'un événement catastrophe est de 48 à 72 heures consécutives. Mais pour certains sinistres, tels que les inondations et les tremblements de terre, ils peuvent être défini par une période de 168 heures consécutives.

Exclusions générales : varient selon les spécialités des régions et les traités, mais certaines restrictions sont systématiquement appliquées pour limiter l'exposition du réassureur à des risques difficiles à modéliser. Parmi les exclusions les plus courantes, nous pouvons citer :

- Les risques nucléaires, biologiques et chimiques ;
- Les risques de guerre passive et active ;
- Les épidémies et infections de grande ampleur ;
- Les risques d'origine liés aux opérations de rétrocession

Par ailleurs, certaines nouvelles exclusions apparaissent comme des exclusions aux incidents cybernétiques résultant d'attaques informatiques (Cyber Act).

3.4.2 Clauses particulières

Clause de reconstitution de garantie (REC) :

Dans le cadre d'un traité XSC, la capacité initiale fournie par le réassureur peut être utilisée intégralement rapidement en cas de survenance de plusieurs sinistres majeurs au cours de la période de couverture. Afin d'assurer que la cédante conserve une protection adéquate après l'épuisement de la garantie initiale, une clause de reconstitution peut être prévue. Cette clause permet au réassureur de rétablir automatiquement la couverture moyennant une prime additionnelle.

Mécanisme de reconstitution de garantie :

- **Activation immédiate :** Dès qu'une partie de la garantie initiale a été utilisée pour indemniser un sinistre, la reconstitution s'applique automatiquement.

- **Paiement d'une prime additionnelle** : Cette prime est généralement calculée en prorata du montant reconstitué et exprimée sous forme d'un taux appliqué à la prime de réassurance de base.

Dans certains traités, le réassureur peut accorder une (ou plusieurs) reconstitution gratuite, permettant ainsi à la cédante de bénéficier d'une couverture prolongée sans coût additionnel.

Clause de Non-Claim Bonus (NCB) : Cette clause est mentionnée dans le traité comme une réduction de la prime de réassurance ou un remboursement partiel si aucune sinistralité n'est enregistrée sur la période de couverture. Cette spécialité encourage la cédante à mieux maîtriser son risque en lui offrant un avantage financier en cas de l'absence de sinistres ou en cas de faible sinistralité.

Exemple :

Capacité du traité : 200€ par événement

Rétention de la cédante : 50€

Il faut que l'évènement affecte au minimum 3 personnes que la garantie prenne effet

Prime cédée : 10€

Nombre de reconstitution de garantie : 1 reconstitution payante à 50% de la prime initiale

NCB : 20% de la prime cédée

Scénario 1 : Aucun événement catastrophique ne dépasse la priorité pendant la durée de couverture

Selon la clause NCB, si aucun sinistre ne dépasse la rétention, la cédante bénéficie d'un remboursement de 20% de la prime de réassurance.

Résultat : La cédante reçoit un bonus de 2€ (20% de 10€), ce qui réduit son coût total de réassurance à 8€.

Scénario 2 : Une inondation majeure entraîne un coût total de 90€, dont 40€ sont pris en charge par la réassurance (hypothèse sans le frais additionnel de la clause REC)

Suite à l'inondation, l'assureur a enregistré 5 sinistres A, B, C, D, E de montants respectifs :

- A : 14 €
- B : 17 €
- C : 12 €
- D : 25 €

- E : 22 €

Le réassureur intervient quand la somme des sinistres individuels liés à un même événement catastrophique dépasse la franchise (ici 50) et ce jusqu'au plafond (ici 200).

Somme des sinistres vaut : $14 + 17 + 12 + 25 + 22 = 90$ €. Et la condition des 3 têtes est vérifiée.

Résultat de réassurance = Prime cédée – Sinistres payés = -30€

Le traité est totalement décitaire. La cédante ne va pas recevoir le NCB.

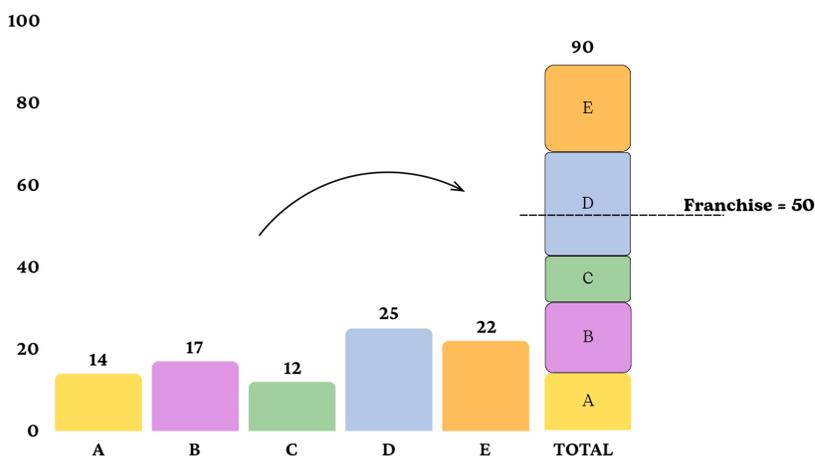


FIGURE 18 – Exemple de traité XS CAT 200 XS 50

Scénario 3 : Un tremblement de terre en mars entraîne des pertes totales de 300€, dont 250 € sont couverts par la réassurance (300€ - 50€ de rétention)

Couverture initiale : 200 € ⇒ Totalemment épuisée

Première reconstitution de 200€ activée immédiatement

→ 50€ avec la prime additionnelle 25% (= 50/200) de prime initiale.

4

Explication du modèle « Peaks Over Threshold » (POT)

Le modèle utilisé par Arundo Re s’inspire des travaux développés par Erland Ekhedén et O. Hössjier. Cette section se concentre sur le processus de tarification des traités catastrophiques (XSC) dans le secteur réassurance-vie, avec une attention particulière sur le risque de décès. L’objectif de ce chapitre est de présenter une méthodologie permettant de suivre et d’analyser l’exposition aux catastrophes naturelles et aux actes de terrorisme, en s’appuyant sur des bases de données reconnues, telles que **l’International disaster database EM-DAT** (accessible sur www.emdat.be/) et **la Global Terrorism Database** (disponible sur www.start.umd.edu/gtd/). L’analyse repose sur les données provenant des deux bases citées, sur une période allant de 1970 à 2018, soit 48 ans d’observations.

Modélisation – au cœur de cette partie, nous allons présenter un modèle structuré en plusieurs étapes clés. Nous commençons par estimer la fréquence annuelle des événements catastrophiques, à partir des données observées entre 1970 et 2018. Puis, nous évaluons la gravité de ces événements en estimant le nombre de personnes assurées décédées, ce qui permet de quantifier l’impact humain de chaque sinistre. Cette quantification est indispensable pour appréhender l’exposition globale aux risques et estimer les coûts potentiels. Au final, nous appliquons les paramètres du traité de réassurance afin d’estimer la prime pure, correspondant à la charge moyenne attendue par le réassureur :

$$\mathbb{P} = \mathbb{E}[\min\{\max(C - \text{priorité}, 0), \text{capacité}\}]$$

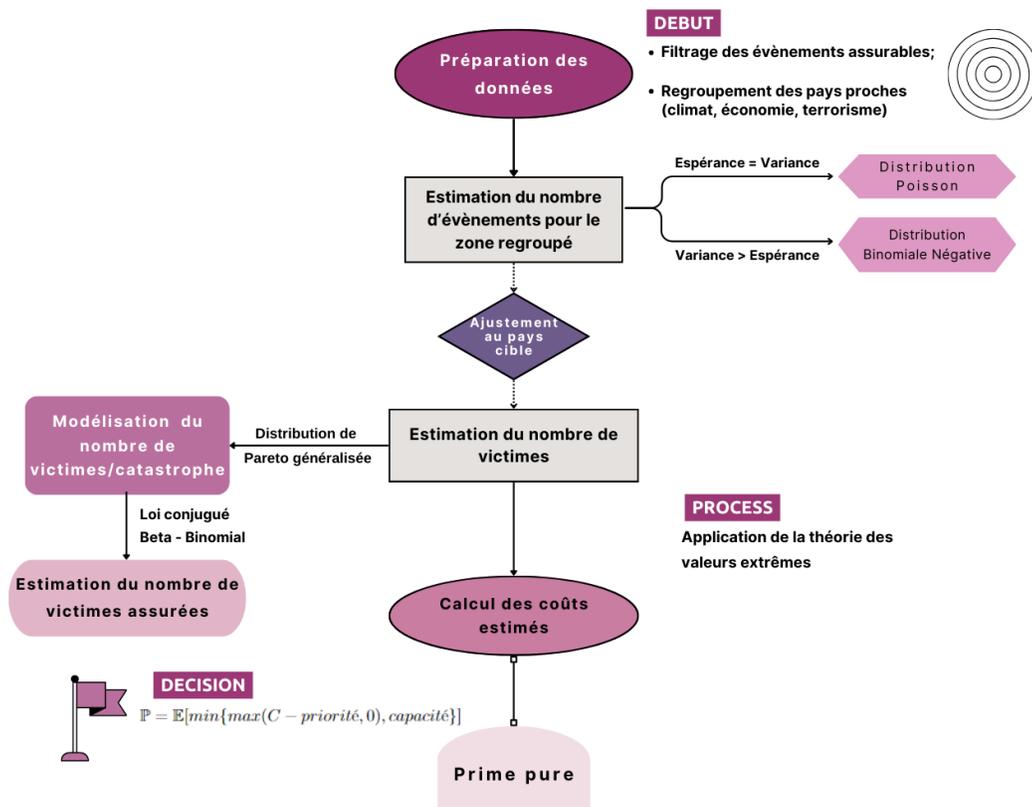


FIGURE 19 – Processus du modèle de tarification du traité XSC dans l’assurance-vie

Dans notre modèle de tarification en réassurance vie, la simulation Monte Carlo permet d’intégrer l’aléa à chaque étape – de la fréquence des événements à leur sévérité, jusqu’au nombre de victimes assurées. Elle permet ainsi de générer un grand nombre de scénarios de sinistralité, rendant possible le calcul d’indicateurs de risque tels que la Value at Risk, essentiels pour déterminer une prime de réassurance adaptée à des événements rares mais à fort impact.

Ces différentes étapes, complémentaires et interdépendantes, renforcent la cohérence de notre modélisation. Ensemble, elles permettent de mieux comprendre les dynamiques des risques catastrophiques et d’estimer de manière fiable les coûts attendus. L’objectif final est de construire un outil d’aide à la décision robuste, facilitant une gestion efficace et prospective des risques extrêmes en assurance-vie.

4.1 Construction des variables

Dans le cadre de l'étude, deux bases de données internationales ont été mobilisées : **EM-DAT**, pour les catastrophes naturelles, et **GTD**, pour les actes de terrorisme. Ces sources sont largement reconnues dans la communauté scientifique de l'assurance. Elles ont toutefois évolué au fil du temps — dans leurs méthodes de collecte, leurs critères d'inclusion ou leurs définitions d'événements — ce qui peut influencer l'interprétation des tendances observées. Comprendre ces évolutions est donc essentiel afin d'assurer une lecture correcte des phénomènes extrêmes.

EM-DAT, gérée par le Centre de recherche sur l'épidémiologie des désastres (CRED) de l'Université catholique de Louvain, recense les catastrophes naturelles et technologiques majeures. Un événement y est intégré dès lors qu'il satisfait à l'un des critères suivants :

- au moins dix décès,
- au moins cent personnes affectées,
- une déclaration d'état d'urgence,
- ou un appel à l'aide internationale.

Les données proviennent de sources multiples (rapports gouvernementaux, agences onusiennes, ONG, assureurs, médias). Au fil du temps, la structure d'EM-DAT s'est affinée. Depuis 2023, un portail documentaire précise la méthodologie de collecte et les définitions, renforçant la comparabilité des données, malgré certaines limites (précision variable des localisations, manque d'estimations économiques complètes).

La **Global Terrorism Database (GTD)**, aujourd'hui administrée par le National Consortium for the Study of Terrorism and Responses to Terrorism - START (Université du Maryland), couvre les incidents terroristes depuis 1970. Elle a connu plusieurs phases de gestion (la Pinkerton Global Intelligence Service "PGIS", le Center for Terrorism and Intelligence Studies "CTIS", le Institute for the Study of Violent Groups "ISVG", puis START depuis 2011), chaque reprise s'accompagnant d'une révision méthodologique. Certaines de ces évolutions ont influencé la fréquence enregistrée de certains types d'attaques (par ex. attaques-suicides). Une rupture notable est l'absence totale de données pour 1993, en raison d'un problème technique de migration. Malgré ces contraintes, GTD reste la base la plus exhaustive disponible à ce jour, couvrant plus de 200 000 incidents.

Ces évolutions méthodologiques ont des implications directes pour l'analyse. Ainsi, un pic ou une baisse dans les données peut traduire un changement de protocole plutôt qu'une évolution réelle du risque. Dans

le cadre du modèle POT, centré sur les événements extrêmes, il est donc crucial de veiller à la cohérence temporelle et aux définitions employées.

4.1.1 Sources et structure des données

EM-DAT fournit des informations sur des catastrophes comme les inondations, séismes, tempêtes ou sécheresses, incluant la date, le pays, le type d'événement, le nombre de victimes et, parfois, des pertes économiques.

GTD recense les actes terroristes depuis 1970, avec des variables comme le type d'attaque, la cible, l'auteur (si identifié) et le nombre de victimes.

Pour cette étude, la période retenue va de **1970 à 2018**, afin d'assurer l'homogénéité temporelle entre les deux bases. Les événements considérés doivent :

- avoir un nombre de victimes supérieur à zéro,
- être associés à un pays identifié,
- et comporter une date complète (au moins l'année).

Base de données	Type d'événement	Période utilisée	Nombre d'événements initial	Nombre d'événements après filtrage
EM-DAT	Catastrophes naturelles	1970 - 2018	21 316	17 042
GTD	Actes terroristes	1970-2018	49 711	49 520

TABLE 2 – Synthèse des bases de données utilisées

4.1.2 Nettoyage et prétraitement

Avant l'intégration dans le modèle, un traitement des données a été mené pour garantir leur cohérence et leur pertinence. Les principales opérations réalisées sont :

- **Filtrage** des événements incomplets (sans année, pays ou nombre de victimes).
- **Extraction de l'année** : seule l'année de survenance est retenue, le modèle fonctionnant sur une base annuelle.
- **Exclusion** de certains événements hors périmètre, comme les épisodes de températures extrêmes, non couverts par le traité étudié.

- **Application d'un seuil minimal de victimes (m)** pour ne retenir que les événements suffisamment graves.

4.1.3 Sélection et regroupements géographiques

Nous avons limité l'analyse aux pays ayant une fréquence d'événements suffisante pour permettre une estimation statistique robuste. Ainsi :

- aux pays comptant au moins dix événements sur la période (selon l'ancien modèle) ;
- à des regroupements de pays aux profils de risque proches (par ex. Espagne–Portugal–Italie ; Colombie–Costa Rica).

Pays / Groupe régional	Seuil (nombre de victimes)
France	35
Chine	100
...	

TABLE 3 – Exemple de seuils choisis pour la méthode POT

Zone de risque	Population 2019	Part régionale	Pays 1	Pays 2	Pays 3
FR	67 059 887	36,31 %	ES	PT	IT
ES	47 076 781	25,49 %	PT	IT	FR
PT	10 269 417	5,56 %	ES	IT	FR
IT	60 297 396	32,65 %	ES	FR	IT

TABLE 4 – Exemples de regroupements de pays utilisés dans l'analyse

Ces regroupements visent à accroître la stabilité des résultats tout en respectant une cohérence climatique, sociale et structurelle.

4.1.4 Variables d'entrée

Chaque observation correspond à un **événement unique** enregistré dans EM-DAT ou GTD. Deux variables principales le caractérisent :

- **Année de survenance** : référence temporelle unique (fréquence annuelle).
- **Nombre total de victimes** : indicateur de gravité et d'impact.

Chaque événement est rattaché à une **unité géographique** (pays isolé ou regroupement de pays), afin de garantir une meilleure représentativité statistique.

4.2 Modélisation de la fréquence des événements catastrophiques

La première étape consiste à ne retenir que les événements dont le nombre de décès est supérieur à un seuil, noté « m », qui représente le nombre minimum de victimes dans un événement catastrophique. Ce seuil, indispensable à notre modèle, nous aide à focaliser l'analyse sur les événements les plus rares et les plus graves. Une fois la sélection effectuée, nous procédons au comptage annuel des événements pour constituer une nouvelle base de données.

Claim_year	Country	Type	Morts
1973	France	Terrorism	3
1974	France	Terrorism	2
1975	France	Terrorism	2
1978	France	Terrorism	5
1978	France	Terrorism	2
1978	France	Terrorism	2
1978	France	Terrorism	9
1980	France	Terrorism	2
1980	France	Terrorism	3
1980	France	Terrorism	4
1980	France	Terrorism	2
1980	France	Terrorism	2
1981	France	Terrorism	2
1982	France	Terrorism	5
1982	France	Terrorism	6

FIGURE 20 – Extrait de notre base de données de référence pour la France

Exemple d'application du seuil $m = 2$ en France

L'image illustre l'application du seuil « m » = 2 pour filtrer les événements catastrophiques en France. La première table recense les événements individuels, indiquant l'année, le pays, la nature de l'événement et le nombre de décès. Seuls les événements ayant entraîné au moins 2 décès sont retenus. Cet exemple montre comment le seuil défini permet d'extraire une base de données épurée et adaptée à l'analyse des tendances des événements extrêmes.

Estimateurs de la fréquence des événements extrêmes

Dans cette section, nous cherchons à estimer le nombre d'événements catastrophiques sur une année.

Nous considérons deux distributions pour modéliser le nombre de victimes : *La loi de Poisson* et *la loi Binomiale Négative*. La sélection de la loi dépendra de sa capacité à modéliser des événements rares et indépendants sur un intervalle de temps, en tenant compte du niveau de dispersion des données. La loi de Poisson est adaptée si la variance est proche de la moyenne, tandis que la loi Binomiale Négative est privilégiée en cas de sur-dispersion.

Mettons \mathbf{K} , le nombre d'évènements catastrophes dans un an.

$$\begin{cases} \mathbb{E}[K] = \text{Var}[K] & \Rightarrow \text{Distribution Poisson} \\ \text{Var}[X] \gg \mathbb{E}[X] & \Rightarrow \text{Distribution Binomiale négative} \end{cases}$$

La loi de Poisson suppose un taux d'occurrence constant des événements. Si les événements catastrophiques se produisent à un rythme relativement constant sur une longue période, la loi de Poisson est un modèle approprié.

$$\begin{aligned} K &\stackrel{\mathcal{L}}{\sim} \text{Poi}(\lambda) \\ \Pr[K = k] &= \frac{e^{-\lambda} \lambda^k}{k!} \\ \mathbb{E}[K] &= \text{Var}[K] = \lambda \end{aligned}$$

En revanche, **la loi Binomiale Négative** est souvent utilisée lorsque les données présentent une sur-dispersion, c'est-à-dire lorsque la variance des données est plus grande que l'espérance. Dans le cas des catastrophes, il est possible que la fréquence des événements varie considérablement d'une année à l'autre, rendant la loi Binomiale Négative plus appropriée que la loi de Poisson.

$$\begin{aligned} \mathbf{K} &\stackrel{\mathcal{L}}{\sim} \text{NB}(\mathbf{n}, \mathbf{p}) \\ \Pr[K = k] &= \binom{k+r-1}{k} p^r (1-p)^k \\ \mathbb{E}[K] &= \frac{n(1-p)}{p} \\ \text{Var}[K] &= \frac{n(1-p)}{p^2} \end{aligned}$$

Regroupement

Lorsqu'un pays couvert par notre traité de réassurance ne dispose pas d'un historique de sinistres suffisamment significatif, il est judicieux de l'associer à d'autres pays présentant des caractéristiques comparables. Cette approche s'appelle « *sous - échantillonnage* ». Elle permet d'obtenir des paramètres de fréquence et de sévérité plus robustes en mutualisant les données, garantissant ainsi une meilleure estimation du risque pour l'ensemble du groupement. Pour déterminer le nombre d'événements CAT entraînant au moins « *m* » décès dans la région et pouvant affecter le pays couvert, on introduit le paramètre pr_{region} , défini comme suit :

$$pr_{\text{region}} = \frac{\text{Population totale du pays}}{\text{Population totale de la region}}$$

Avec ce paramètre, le nombre K' d'événements CAT causant au moins « *m* » morts dans un pays spécifique suit une loi binomiale de paramètres K et pr_{region} , soit :

$$K' \stackrel{\mathcal{L}}{\sim} \text{Bin}(K; pr_{\text{region}})$$

En résumé, nous utilisons la loi de probabilité sélectionnée pour calculer la fréquence annuelle des événements sur la zone regroupée avec le pays cible. Une fois ce chiffre estimé, nous appliquons la distribution binomiale pour estimer le nombre d'événements dans le pays concerné.

4.3 Modélisation de la sévérité des sinistres

L'approche Peaks Over Threshold (POT), issue de la théorie des valeurs extrêmes, la distribution des excès au-dessus d'un seuil élevé u peut, sous certaines conditions, être approximée par une loi limite. Plus précisément, si X est une variable aléatoire représentant la sévérité d'un sinistre, alors la distribution conditionnelle des excès $\mathbf{Y} = X - u \mid X > u$ converge vers une distribution de Pareto généralisée (GPD), lorsque $u \rightarrow \infty$. Cette distribution est définie par :

$$f_Y(y) = \mathbb{P}[\mathbf{Y} = X - u \mid X > u] = \frac{1}{\beta} \left(1 + \frac{\xi y}{\beta}\right)^{-\left(1 + \frac{1}{\xi}\right)}, \quad y > 0, \beta > 0, \text{ et } 1 + \frac{\xi y}{\beta} > 0$$

4.3.1 Estimation du nombre de victimes

Une fois le nombre d'événements catastrophiques annuels K' estimé, nous simulerons le nombre de victimes par événement à l'aide de la distribution GPD (Generalized Pareto Distribution). Issue de la théorie des valeurs extrêmes, cette distribution est particulièrement adaptée pour modéliser la probabilité d'occurrence des valeurs rares et extrêmes, dépassant un seuil « m » prédéfini.

Mettons X_k , le nombre de décès dans le $k^{\text{ième}}$ scénario d'évènement catastrophe avec $k = 1; 2; \dots; K'$. On suppose que tous les évènements sont indépendantes et identiquement distribuées (i.i.d).

Estimation du nombre des victimes en chaque évènement

Nous nous concentrons ici sur une phase essentielle de notre analyse des risques : l'évaluation du nombre de victimes pour chaque catastrophe. Dans chaque scénario, la loi de distribution de Pareto généralisée (GPD) est appliquée pour estimer, de manière probabiliste, le nombre de victimes par événement extrême qui est au-delà d'un certain seuil « m ». Donc, le seuil « m » est un élément clé du modèle POT, car il définit la frontière entre les valeurs courantes et les événements extrêmes. Cette considération sera examinée et fera l'objet de tests analytiques dans le chapitre suivant **Amélioration du modèle**.

Nous prenons la variable \mathbf{X} comme le nombre de victimes d'une catastrophe. Si X est plus grand que la localisation « m », X suit la distribution GDP comme suit :

$$X \stackrel{\mathcal{L}}{\sim} \text{GPD}(\xi, \sigma, \mu)$$

ξ : la forme/le shape

σ : l'échelle/le scale

μ : la localisation/le seuil

La fonction de répartition :

$$f(x) = \begin{cases} (1 + \xi z)^{-\frac{\xi+1}{\xi}} & \text{si } \xi \neq 0 \\ e^{-z} & \text{si } \xi = 0 \end{cases}$$

La fonction de densité :

$$F(x) = \begin{cases} 1 - (1 + \xi z)^{-\frac{1}{\xi}} & \text{si } \xi \neq 0 \\ e^{-z} & \text{si } \xi = 0 \end{cases}$$

La GPD, reconnue pour sa capacité à modéliser les valeurs extrêmes et les queues lourdes des distributions, permet d'estimer le nombre de victimes par sinistre. Dans certains scénarios, le nombre estimé de décès peut être inférieur au seuil défini par notre contrat de réassurance. Il ne s'agit pas du seuil utilisé dans notre modèle, mais de celui spécifié dans le cadre contractuel. Dans ces cas, nous ajustons les résultats en fixant le nombre de décès à 0 pour ces scénarios, afin de respecter les exigences de notre cadre d'analyse.

4.3.2 Modélisation du nombre de victimes assurées dans notre portefeuille

L'étape suivante de notre modèle consiste à déterminer le nombre de victimes d'un événement catastrophe de notre portefeuille d'assurance. Cette démarche est essentielle pour compléter notre compréhension des impacts spécifiques aux assurés. Il s'agit de la nouvelle variable :

$$q = \frac{\text{Nombre d'assurés dans notre portefeuille}}{\text{Nombre de personnes dans la population}}$$

L'approche « sous - échantillonnage » est assez similaire à celle qui consiste à estimer le nombre d'événements dans un pays spécifique dans l'ensemble de la région (premier étape). Le paramètre « q » représente la part de marché.

$$Y|X \stackrel{\mathcal{L}}{\sim} \text{Bin}(X, q)$$

$$\Pr[Y = y] = \binom{X}{y} q^y (1 - q)^{(X-y)} \text{ avec } y \in 0, 1, \dots, X$$

Pour les grandes catastrophes, quand $X \rightarrow \infty$, nous avons la relation suivante :

$$\frac{Y}{X} \simeq q \text{ lorsque } X \rightarrow \infty$$

Elle découle de la loi des grands nombres. A mesure que la taille de l'échantillon augmente, la moyenne des résultats observés tend à se rapprocher de l'espérance théorique q .

Toutefois, la distribution Binomiale n'est qu'appliquée avec des essais indépendants, où les couvertures d'assurance sont uniformément réparties dans la population. Plus exactement, on dit que :

- Chaque victime représente un essai indépendant,
- La probabilité d'être assurée (notée q) est identique pour toutes les victimes.

En fait, ces hypothèses sont souvent éloignées de la réalité du marché de l'assurance, en particulier dans le cas des contrats collectifs. En effet, la souscription d'assurances peut être fortement concentrée au sein de certains segments de population ou dans des zones géographiques spécifiques.

Lorsqu'un événement catastrophique touche une région où un assureur détient une forte part de marché via des contrats de groupe (par exemple, une entreprise, une administration, etc.), la proportion de victimes assurées peut alors s'écarter significativement de la moyenne théorique. Avec cette situation, nous abordons ici « le risque de concentration ».

L'inférence bayésienne : quand la loi Bêta agit comme une distribution a priori
conjuguée pour une variable aléatoire suivant une distribution binomiale

L'une des méthodes pour résoudre ce problème est l'approche Bayésienne. Grâce à cette approche, les compagnies d'assurance peuvent intégrer leurs connaissances ou leurs croyances préalables sur la concentration du portefeuille d'assurance. On a donc :

$$Y|X \stackrel{\mathcal{L}}{\sim} \text{Bin}(n = X, p)$$

$$p|X \sim \text{Beta}[d(X)q, d(X)(1 - q)]$$

$$E[p|X] = \frac{d(X)q}{d(X)q + d(X)(1 - q)} = q$$

$$\text{Var}(p|X) = \frac{q(1 - q)}{d(X) + 1}$$

Avec cette forme, $\alpha = d(X)q$ représente la probabilité qu'une victime soit assurée. Donc si α est plus grand, la distribution est penchée vers 1 indiquant une plus grande probabilité que les victimes soient assurées. La fonction $d(X)$ permet de moduler la probabilité conditionnelle qu'une victime soit assurée, en fonction de l'ampleur de la catastrophe, capturant le risque de concentration.

On va construire la fonction $d(X)$ sous l'hypothèse :

- Lorsque $d(X) \rightarrow +\infty$, la variance de la loi bêta tend vers zéro. Dans ce cas, p est quasiment constant et se rapproche de q . On retrouve alors une loi binomiale classique :

$$Y | X \stackrel{\mathcal{L}}{\sim} \text{Bin}(X, q)$$

Cela correspond à une situation d'indépendance totale entre le statut d'assurance des victimes, et reflète le cas où l'événement est de grande ampleur, rendant l'estimation de p plus stable (loi des grands nombres).

- Lorsque $d(X) \rightarrow 0$, la variance de la loi bêta devient maximale. Le paramètre p tend alors vers les valeurs extrêmes 0 ou 1. Concrètement, cela signifie que :

$$\mathbb{P}(Y = 0 | X) = 1 - q \quad \text{et} \quad \mathbb{P}(Y = X | X) = q$$

Autrement dit, soit aucune victime n'est assurée, soit toutes le sont. Cette configuration illustre un cas de dépendance extrême, et traduit un risque de concentration élevé, typique des événements de faible ampleur ou très ciblés.

Pour visualiser le comportement de p :

Lorsque la valeur de $d(X)$ est suffisamment grande, la distribution de p tend à se rapprocher de celle de q . A l'inverse, la valeur de $d(X)$ étant petite, le comportement de p est varié.

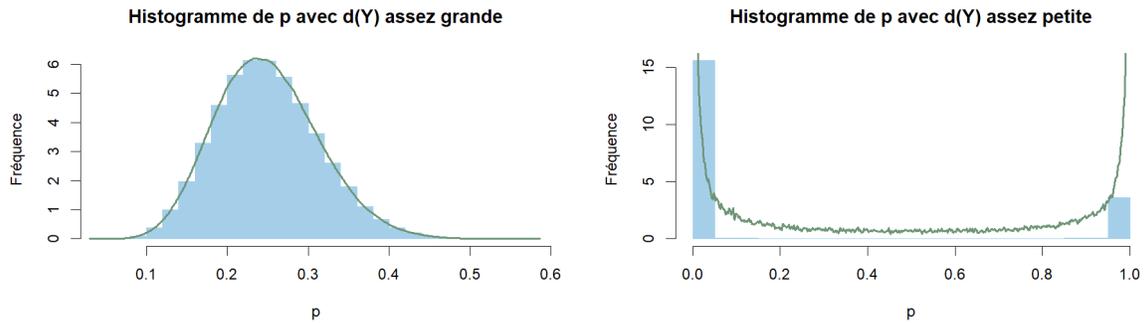


FIGURE 21 – Comportement de p lié sur $d(X)$ avec $p_{assume} = 0.2$

D'après Ekheden, on définit la forme de $d(X)$ comme :

$$d(X) = \theta \log(X)$$

De plus, le paramètre θ est primordial : pour X grand, si θ est grand, il y a indépendance totale alors que si θ est petit, on note une grande dépendance. Nous prendrons en compte le changement de paramètre dans le chapitre 6 « Calibration et sensibilité du modèle ».

4.4 Modélisation du coût de chaque sinistre lié à l'évènement catastrophe concerné

Il ne nous reste plus qu'à estimer l'engagement moyen associé à chaque portefeuille de risque. Nous allons utiliser une distribution exponentielle puis une distribution gamma pour estimer la somme des engagements. Dans un premier temps, une distribution exponentielle est utilisée pour représenter un engagement par assuré constant et indépendant. Toutefois, afin de mieux capturer l'hétérogénéité des montants assurés selon les individus ou les contrats, nous privilégions ensuite l'utilisation d'une distribution gamma, plus flexible, permettant de modéliser à la fois la moyenne et la variance des engagements.

La distribution exponentielle permet de représenter un engagement individuel supposé constant et indépendant. Elle est souvent utilisée en assurance pour modéliser des variables continues et positives, comme un montant moyen par assuré, avec un paramètre $\lambda = 1$, ce qui correspond à un coût moyen unitaire. Cependant, si un sinistre touche un groupe de personnes, il devient plus pertinent de modéliser la somme des engagements individuels. Dans ce cas, la distribution gamma est mieux adaptée, car elle permet de représenter la somme de plusieurs variables exponentielles indépendantes, reflétant ainsi l'engagement total d'un groupe d'assurés.

Mathématiquement, on considère la somme sous risque d'un assuré noté z , c'est :

$$p^{(1)}(z) = f(z) = \theta e^{-\theta z} = e^{-z} \text{ (la loi exponentielle avec paramètre 1)}$$

Quand on suppose l'indépendance des décès, la densité de la somme sous risque pour y décès va suivre la loi Gamma de paramètre y et $\theta = 1$. La densité s'écrit suivante :

$$p^{(y)}(z) = \frac{z^{y-1} e^{-z}}{\Gamma(y)} = \frac{z^{y-1} e^{-z}}{\Gamma(y)} \text{ avec } z \geq 0 \text{ et } y \in \mathbb{N}^*$$

Où la fonction Gamma $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ et avec pour tout $y \in \mathbb{N}^*$:

$$f(y) = \Gamma(y) = \begin{cases} \Gamma(1) = 1! \\ \Gamma(y+1) = y\Gamma(y) = y(y-1)! = y! \end{cases}$$

Dans ce cas, en utilisant une distribution Gamma pour estimer l'engagement moyen par tête et que on multiplie ensuite cette estimation par le capital sous risque moyen par chaque tête, nous obtenons une estimation de l'engagement potentiel total en cas de survenance de l'évènement.

Concernant le capital sous risque moyen, en tant que réassurance, on est informé des effectifs par tranche de capitaux sous forme de classe. Finalement :

$$\mathbf{Z} = \Gamma(\mathbf{z}, \mathbf{y}, \theta = 1) * \mathbf{SAR}_{\text{moyen}}$$

4.5 Application des conditions du contrat CAT

Introduction d'un contrat d'hypothèse :

- P : portée du traité définie par contrat
- M : nombre minimal de têtes requis pour activer le traité
- R : priorité/ rétention du traité
- SAR : capitaux sous risque
- AAL : (fois) le nombre de reconstitution
- L : limité = $P*(1 + AAL)$

$$Z' = \begin{cases} 0 & \text{si } Z \in [0; R] \\ Z - R & \text{si } Z \in [R, L] \\ L - R & \text{si } Z \in [L; +\infty) \end{cases}$$

- C : montant à payer par le réassureur lorsqu'il y a un sinistre (z étant le montant du sinistre)

Pour chaque scénario,

$$C = \sum_{k=1}^{K_M} Z_k$$

En basant sur le nombre de scénarios simulés, on peut obtenir les différentes situations possibles. La dernière étape, on souhaite calculer l'espérance et la variance de notre coût de chaque scénario.

$$\mathbb{E}[C] = \mathbb{E}[Y] * \mathbb{E}[Z]$$

$$\text{Var}[C] = \mathbb{E}[Y] * \text{Var}[Z] + \mathbb{E}[Z^2] * \text{Var}[Y]$$

La prime pure est juste que le coût moyen pour couvrir nos pertes attendues, c'est-à-dire :

$$\pi_{\text{pure}} = \mathbb{E}[C]$$

Cependant, la prime pure ne tient pas compte de la variabilité ou de l'incertitude des sinistres. En ajoutant l'écart-type à un niveau de confiance élevé (comme 99.5%), on prend en compte la volatilité des sinistres. Par exemple : Les événements catastrophiques, bien que rares, peuvent entraîner des pertes massives. Cela signifie que la prime est calculée de manière à couvrir non seulement les pertes moyennes, mais aussi la majorité des scénarios de pertes plus élevées que la moyenne. Ainsi, nous définissons **la prime de risque** avec la formule :

$$\pi_{\text{risque}} = \mathbb{E}[C] + \text{chargement de sécurité} * \sigma_C$$

Le chargement de sécurité correspond un niveau de sécurité que le réassureur choisit comme nécessaire. Dans cette partie, on peut expliquer plus de terme « marge technique ». La marge technique est nécessaire pour prendre en compte le risque lié à la qualité des données, le risque de volatilité ainsi que le risque de modèle. Au Arundo Re, selon les experts, nous prenons un chargement de sécurité égal à 10%.

La prime commerciale :

$$\pi_{\text{commerciale}} = \frac{\pi_{\text{risque}}}{1 - \text{Courtage} - \text{Taxe} - \text{Frais de gestion}} \times (1 + \text{Chargement de Rentabilité})$$

Dans un traité catastrophe XSC, le *Rate on Line* (ROL) est un indicateur qui permet de mesurer le prix pour 1€ de capacité. Ce prix s'exprime en pourcentage. Le ROL permet ainsi d'évaluer le coût unitaire de la protection réassurance, facilitant la comparaison entre plusieurs structures ou niveaux de couverture. Il permet aux assureurs et réassureurs de trouver un équilibre entre le prix tarifé (calculé par les modèles) et le prix souscrit sur le marché.

$$\text{ROL} = \frac{\text{Prime Commerciale}}{\text{Portée}}$$

5

Amélioration du modèle

Nous modélisons le modèle pour tarifier à la traité XSC en réassurance vie en utilisant une approche structurée en quatre étapes :

- [1] Estimer le nombre total de catastrophes K pouvant survenir durant la période du contrat T (généralement T est annuel) ;
- [2] Evaluer le nombre de décès X_k , traduisant la gravité de l'événement ;
- [3] Déterminer le nombre de sinistres Y_k découlant de ces décès, en prenant en compte la part des victimes couvertes par l'assureur ;
- [4] Calculer le coût total C associé aux réclamations issues de chaque catastrophe.

Notre modèle utilise la théorie des valeurs extrêmes et a été développé par les mathématiciens Erland Ekhedén et Ola Hössjer. Le modèle original, est connu également sous le nom de POT (Peaks Over Threshold). Ce cadre probabiliste est particulièrement adapté à la modélisation des événements rares et extrêmes, notamment en réassurance. Il permet de décrire avec précision la fréquence et la sévérité des sinistres catastrophiques à partir de données empiriques couvrant la période de 1970 à 2018. Grâce à cette structure théorique robuste et à l'intégration de données réelles, le modèle offre une représentation fidèle du risque, garantissant ainsi des résultats plus fiables et utiles pour la prise de décision en assurance-vie.

Cependant, malgré sa pertinence, le modèle présente certaines limitations qu'il convient d'améliorer :

- Le choix du seuil pour une application de la GPD repose aujourd'hui largement sur l'expertise. Nous proposerons des méthodes statistiques et graphiques pour déterminer ce seuil de manière plus objective.
- La rareté des données par pays nous amène à envisager des regroupements de pays partageant des

caractéristiques similaires, afin d'enrichir les échantillons et d'améliorer la stabilité des estimations.

Ces deux problématiques sont l'objet de ce chapitre. Dans cette section, nous utiliserons les données de la France, de la Chine et de la Colombie comme exemples. Nous testerons des solutions statistiques et illustrerons comment nous avons amélioré la recherche du seuil ou le problème du regroupement.

Pour bien comprendre l'application du modèle à chaque pays, nous commencerons par analyser les caractéristiques descriptives des trois pays étudiés, à savoir la France, la Chine et la Colombie.

o La France

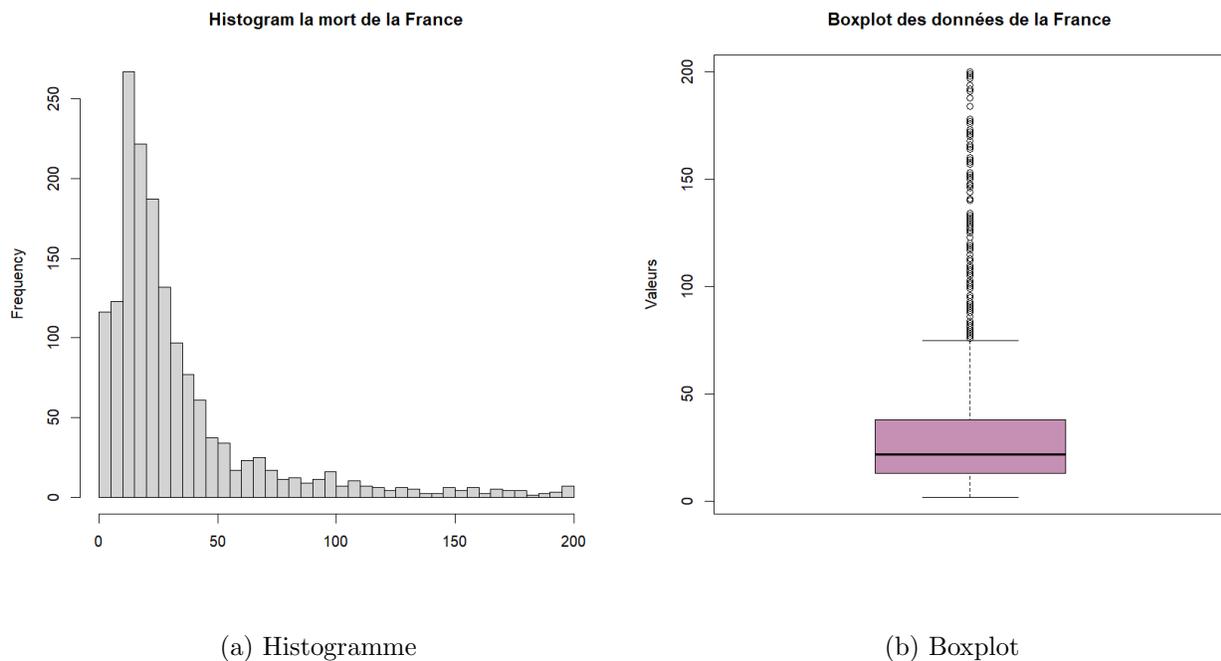


FIGURE 22 – Distribution des décès des événements catastrophiques en France

La **Figure 22** présente la distribution des décès en France à travers deux visualisations : un histogramme (a) et un boxplot (b).

- L'histogramme révèle une asymétrie marquée à droite, signifiant que la majorité des événements ont un nombre de décès inférieur à 50. Toutefois, la présence d'une queue longue dans la distribution indique qu'on a un faible nombre d'événements extrêmes engendrant jusqu'à près de 200 décès.
- Le boxplot confirme cette forte dispersion, avec une médiane autour de 8 décès, un premier quartile

avoisinant 3 décès et un troisième quartile autour de 30 décès. De nombreux outliers, dépassant les 100 décès, confirment l'existence d'événements catastrophiques rares mais significatifs.

o La Chine

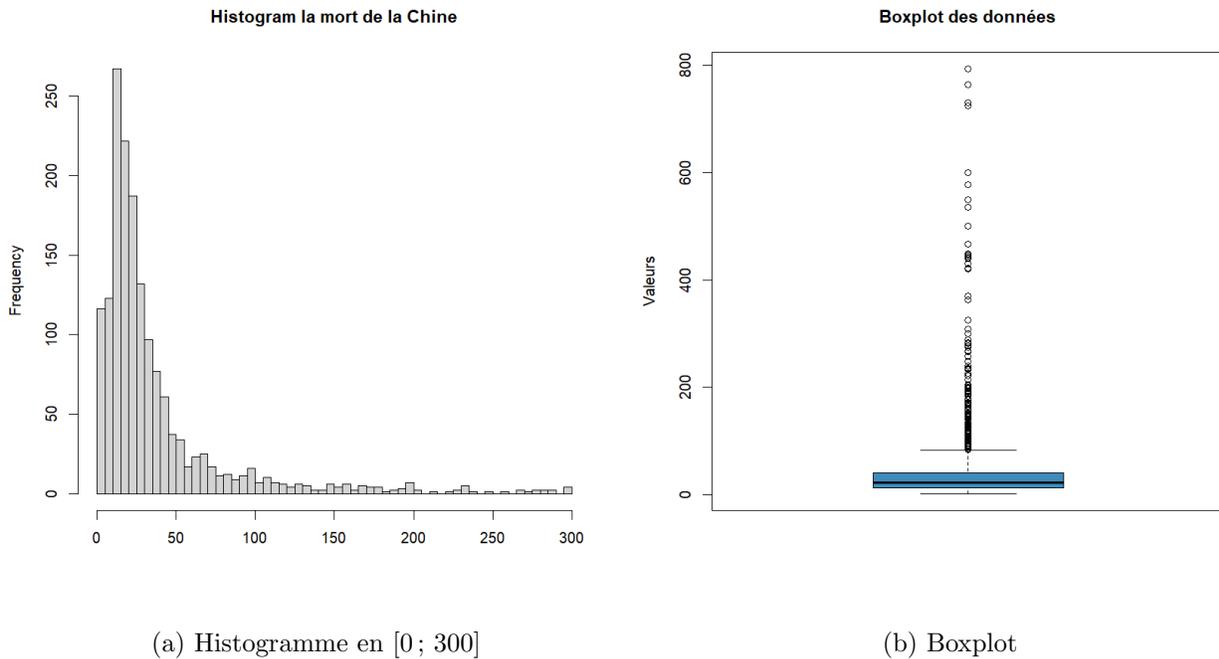


FIGURE 23 – Distribution des décès en Chine

La **figure 23** illustre la distribution du nombre de décès en Chine. Le **boxplot** met en évidence un grand nombre de valeurs extrêmes, avec certains sinistres dépassant 800 décès (maximum 3656 morts dans un évènement), ce qui reflète la présence d'événements catastrophiques de très grande ampleur.

Par rapport à la France, la dispersion des décès est plus marquée en Chine, avec un maximum bien plus élevé et une proportion plus importante de sinistres extrêmement meurtriers. Cela est dû à la densité de population élevée dans les zones à risque accru, à la croissance industrielle et à la gravité des catastrophes naturelles affectant le pays.

○ La Colombie

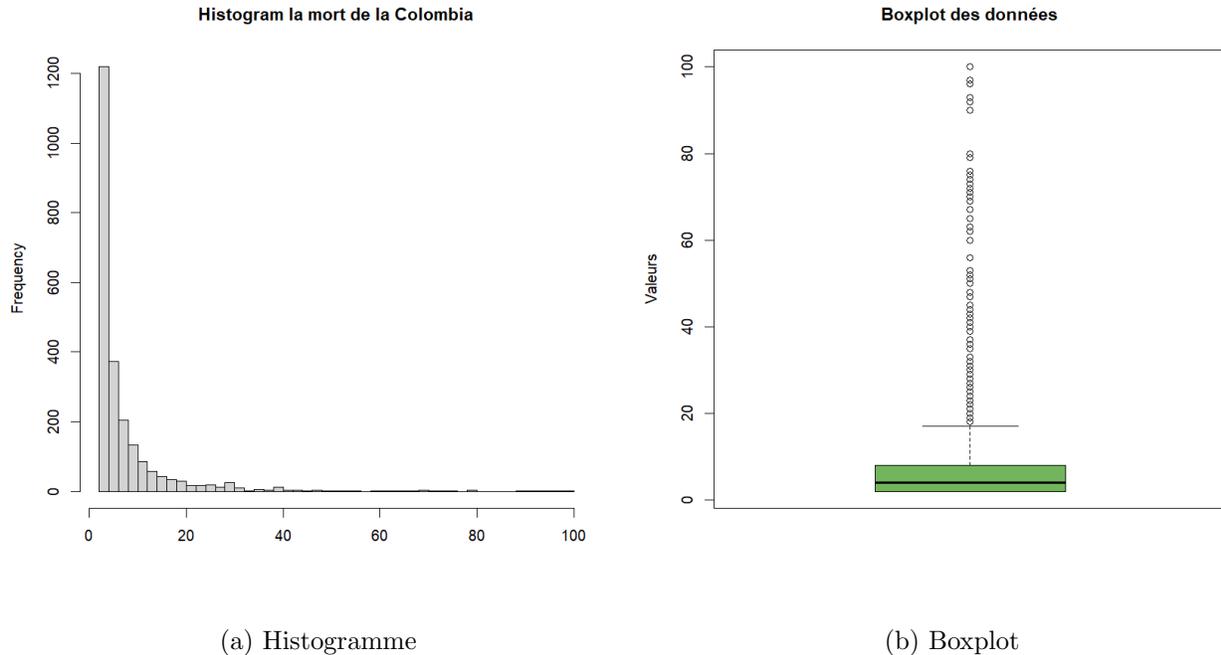


FIGURE 24 – Distribution des décès en Colombie

Par rapport à la France, la Colombie présente des décès par événement plus faibles, avec une majorité des sinistres n'excédant pas 20 victimes, alors que la France enregistre des sinistres dépassant fréquemment 50 décès. Et par rapport à la Chine, où les événements les plus extrêmes dépassent 800 décès, la Colombie affiche un plafond de mortalité nettement plus bas.

Caractéristiques des événements : en Colombie, les catastrophes sont souvent liées aux conflits armés, aux glissements de terrain et aux inondations.

L'analyse des décès liés à des événements extrêmes **en France, en Chine et en Colombie** met en évidence l'importance du **choix du seuil** et du **regroupement des pays** pour appliquer la Théorie des Valeurs Extrêmes (TVE). Un seuil trop bas inclut des événements non extrêmes, tandis qu'un seuil trop élevé réduit excessivement le nombre d'observations, compromettant la fiabilité des estimations. De plus, la rareté naturelle des événements catastrophiques impose de regrouper les données de plusieurs pays aux caractéristiques homogènes (géographie, climat ou géopolitique) afin de garantir la stabilité statistique des ajustements. Cette double approche - optimisation du seuil et regroupement raisonné des données — s'avère donc essentielle pour une modélisation fiable et robuste des sinistres extrêmes en réassurance vie.

5.1 Détermination du seuil optimal

Selon notre approche, le seuil représente le nombre minimum de sinistre dans un évènement catastrophe en global. Dans notre modèle, ce seuil sert de point de référence pour identifier les valeurs extrêmes dans nos données des évènements.

Pour rappel, l'évènement catastrophe est l'un des événements représentatifs de la théorie des valeurs extrêmes. La théorie des valeurs extrême est utilisée pour quantifier le comportement stochastique d'un processus à des niveaux inhabituellement grands ou petits. Sur la base de la méthode POT, l'idée principale est d'extraire les valeurs dépassant un seuil, et puis, les données extraites (le nombre de décès lors d'un évènement catastrophique) peuvent être ajustées par la distribution de Pareto généralisée (GPD). Donc, le problème principe de cette méthode est de déterminer le seuil.

Un seuil optimal doit équilibrer deux aspects : il doit être suffisamment élevé pour garantir que seules les valeurs réellement extrêmes sont prises en compte, mais pas trop élevé pour éviter de perdre des informations importantes et de réduire excessivement la taille de l'échantillon.

Pour résoudre ce problème, les parties principes qu'on va suivre :

Étape 1 : Utilisation des méthodes pour identifier l'intervalle du seuil

- Méthode d'analyse décisionnelle à critères multiples (MCDA)
- Méthode de Hill

Étape 2 : Utilisation des plots afin de vérifier la conformité du seuil

- Mean Residual Plots
- Threshold selection Plots
- QQ Plot

5.1.1 Méthode d'analyse décisionnelle à critères multiples (MCDA)

Les approches statistiques apportent une objectivité dans le choix du seuil, basé sur des tests d'adéquation pour une gamme de seuils différents. Ces tests visent à mesurer la distance entre la fonction de distribution empirique de l'échantillon et une distribution théorique de référence, telle que la loi de Pareto généralisée. Parmi les outils les plus couramment utilisés figurent le test de Kolmogorov-Smirnov (K-S), sensible à la différence maximale entre les deux courbes de répartition ; le test du chi-carré de Pearson (χ^2), basé sur les écarts entre fréquences observées et attendues ; et le RMSE (Root Mean Squared Error),

qui quantifie l'erreur moyenne entre les densités.

Pour sélectionner un seuil optimal dans la méthode POT, il est essentiel de combiner plusieurs critères d'évaluation. L'analyse décisionnelle à critères multiples (MCDA) est une méthode adaptée à ce type de problème car les différents tests peuvent présenter des résultats contradictoires i.e ces statistiques peuvent ne pas avoir la même monotonie en fonction du seuil. En pratique, ces trois tests seront calculés, et une fonction des trois tests (MCDA) permettra d'identifier le seuil optimal.

Test de Kolmogorov-Smirnov (K-S) :

Le test de Kolmogorov-Smirnov (K-S) permet d'évaluer la qualité de l'ajustement d'un modèle statistique à un échantillon de données. Dans notre cas, il est utilisé pour vérifier si la distribution des excès suit bien une distribution de Pareto généralisée (GPD).

La statistique du test est définie comme :

$$K_{max} = \max_{1 \leq j \leq n-1} |F(x_j) - G(x_j)|, |F(x_{j+1}) - G(x_j)|$$

où :

- $F(x)$ est la fonction de répartition empirique (issue des données) des excès au-dessus d'un seuil,
- $G(x)$ est la fonction de répartition théorique (de la GPD estimée),
- x_i sont les données triées par ordre croissant.

Dans notre cas, ce test permet de valider la pertinence de l'ajustement GPD sur les excès au-dessus du seuil choisi. Plus cette valeur est faible, meilleur est l'ajustement.

Le principal avantage du test de K-S est qu'il est **non paramétrique**, c'est-à-dire qu'il ne suppose aucune forme particulière pour la distribution des données. Cela le rend très polyvalent et robuste, en particulier dans le cas de données extrêmes ou atypiques.

Test du Chi-carré de Pearson :

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - Q_i)^2}{Q_i}$$

où :

- O_i est nombre d'observations empiriques dans la classe i (fréquence observée),
- Q_i est nombre attendue dans la classe i selon la GPD estimée,
- k est le nombre total de classes.

Le test du Chi-carré de Pearson est basé sur la comparaison entre les fréquences observées et attendues dans plusieurs intervalles de valeurs. Le but est d'identifier le seuil optimal à partir duquel la GPD fournit un bon ajustement aux données, en particulier dans la queue de distribution, où se concentrent les événements les plus graves.

Par exemple : Supposons que nous avons les excès $X - u$ pour les événements avec **plus de 50 victimes** (seuil $u = 50$) dans les données GTD.

Nous avons un extrait de **100 observations** où $X > 50$, et tu veux voir si la GPD est bien ajustée.

Nous divisons ces données excédentaires en **4 intervalles** de gravité croissante :

CLasse ($X - u$)	Nombre observé O_i	Nombre attendu Q_i (par la GPD)
[0 – 20]	40	38
[20 – 40]	30	32
[40 – 70]	20	22
≥ 70	10	8

Ensuite, nous calculons la statistique du test du Chi-carré :

$$\chi^2 = \sum_{i=1}^4 \frac{(O_i - Q_i)^2}{Q_i} = \frac{(40 - 38)^2}{38} + \frac{(30 - 32)^2}{32} + \frac{(20 - 22)^2}{22} + \frac{(10 - 8)^2}{8}$$

Cela nous donne une valeur $\chi^2 = 0,74$

Erreur Quadratique Moyenne (Root Mean Squared Error - RMSE)

$$\delta_{pdf} = \frac{1}{n^2} \sum_{i=1}^n (f(x_i) - g(x_i))^2$$

Le *RMSE (Root Mean Square Error)* est une mesure largement utilisée pour évaluer la différence entre les valeurs prédites par un modèle et les valeurs réellement observées. Il représente l'écart quadratique moyen entre les deux ensembles de données, exprimé dans la même unité que les observations, ce qui facilite son interprétation.

Pour chaque seuil candidat, une valeur de test sera calculée en fonction du critère de test correspondant. Compte tenu des différents principes des critères d'essai, plus la valeur de la pondération est élevée, plus le critère d'essai est important.

Et puis, on va avoir la matrice de décision R :

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix}$$

Nous considérons un ensemble de m candidats, correspondant aux différentes valeurs possibles de seuils. Pour chaque seuil, nous calculons les performances selon $n = 3$ critères statistiques (K-S, χ^2 , RMSE). Enfin, r_{ij} désigne la valeur de performance associée au j -ème candidat pour le i -ème critère.

Les étapes suivantes consistent à normaliser les valeurs de la matrice de décision R . Le vecteur (w_1, w_2, \dots, w_n) représente l'importance relative et les poids associés aux critères choisis, qui peuvent être dérivés à partir de la matrice de décision. Afin de mieux illustrer les poids attribués à chaque critère, la matrice de décision R est transformée selon les pas suivants :

- Normalisation des scores :

$$p_{ij} = \frac{r_{ij}}{\sum_{i=1}^m r_{ij}}$$

- Entropie du critère j :

Dans la prise de décision, on va utiliser la valeur d'entropie pour calculer le poids des indicateurs. L'entropie de Shannon, largement utilisée en analyse multicritère, nous permet de déterminer objectivement l'importance relative de chaque test statistique (critère), en se fondant sur la variabilité de leurs résultats selon les seuils considérés. Cette approche permet de mesurer le degré de dispersion des performances associées à chaque critère :

- Si les résultats d'un critère varient fortement selon les seuils testés, cela signifie que ce critère est informatif et utile pour discriminer entre de bons et de mauvais seuils.
- Si tous les seuils ont des performances similaires sur un critère, ce dernier n'aide pas à la décision → il aura une entropie élevée et donc un poids faible.

$$E_j = -\frac{1}{\ln(m)} \sum_{i=1}^m p_{ij} \ln(p_{ij})$$

o Poids du critère j :

$$w_j = \frac{(1 - E_j)}{\sum_{j=1}^m (1 - E_j)}$$

La vecteur du poids $W = [w_1, w_2, \dots, w_n]^T$. Finalement, on estime la valeur du test pour chaque candidat du seuil par l'indicateur d'évaluation globale :

$$S = R * W$$

Pour opter quel candidat est mieux, plus l'indicateur d'évaluation globale est petit, meilleur est l'effet d'ajustement de la qualité de l'ajustement obtenu.

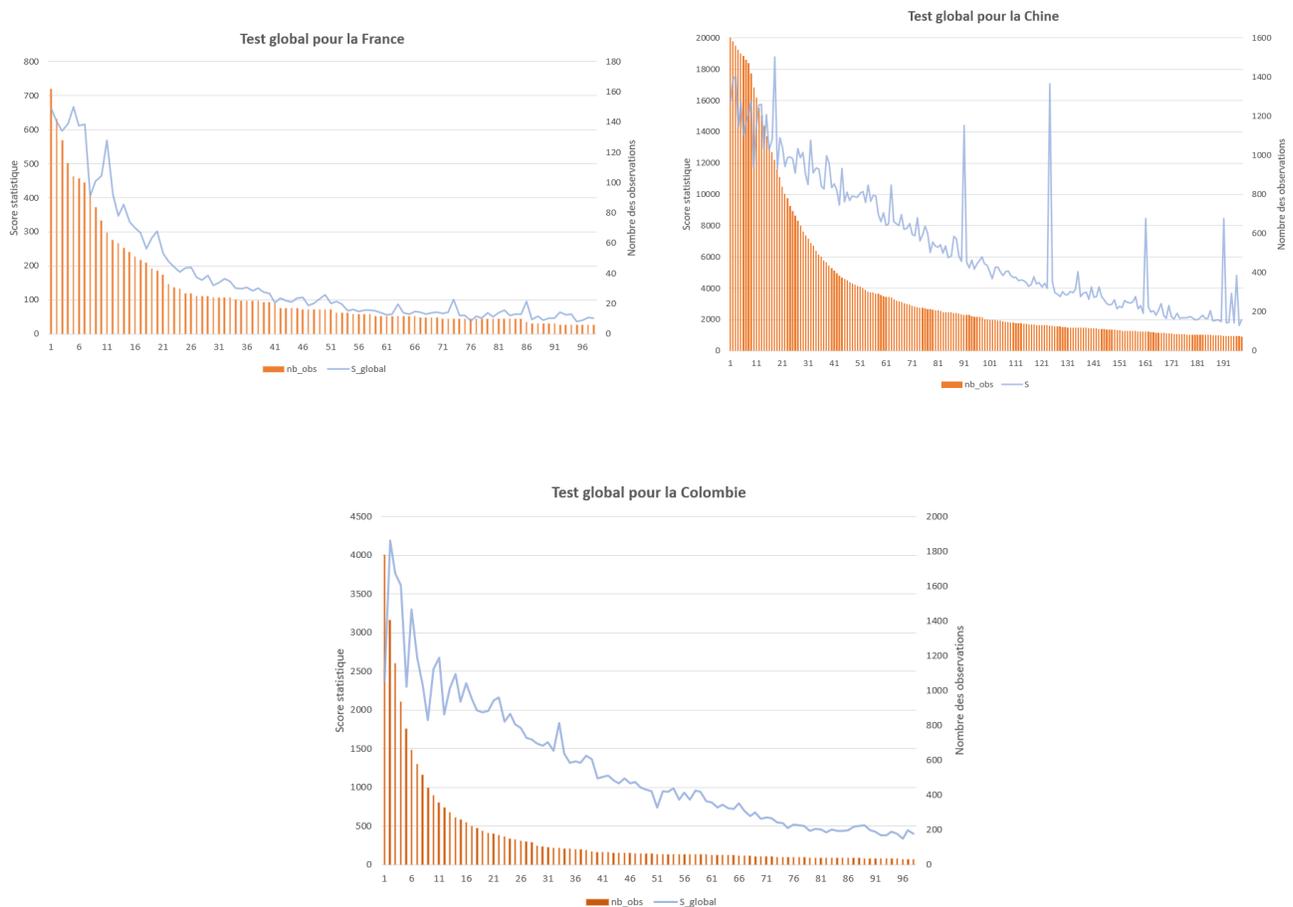


FIGURE 25 – Le résultat de test statistique global

D'après le diagramme ci-dessus, la valeur S représente une mesure statistique globale. Nous observons

que, plus la valeur du seuil augmente, plus la valeur du test diminue. Cela s'explique par le fait que le nombre d'observations disponibles diminue lorsque le seuil augmente.

Une autre manière d'interpréter les résultats de cette méthode est d'examiner le graphique comparatif des tests statistiques. L'objectif principal de ce graphique est de représenter les courbes des valeurs obtenues p_{ij} , situées dans l'intervalle $[0,1]$. Nous cherchons à identifier le point d'intersection des trois courbes, qui pourrait indiquer un seuil optimal.

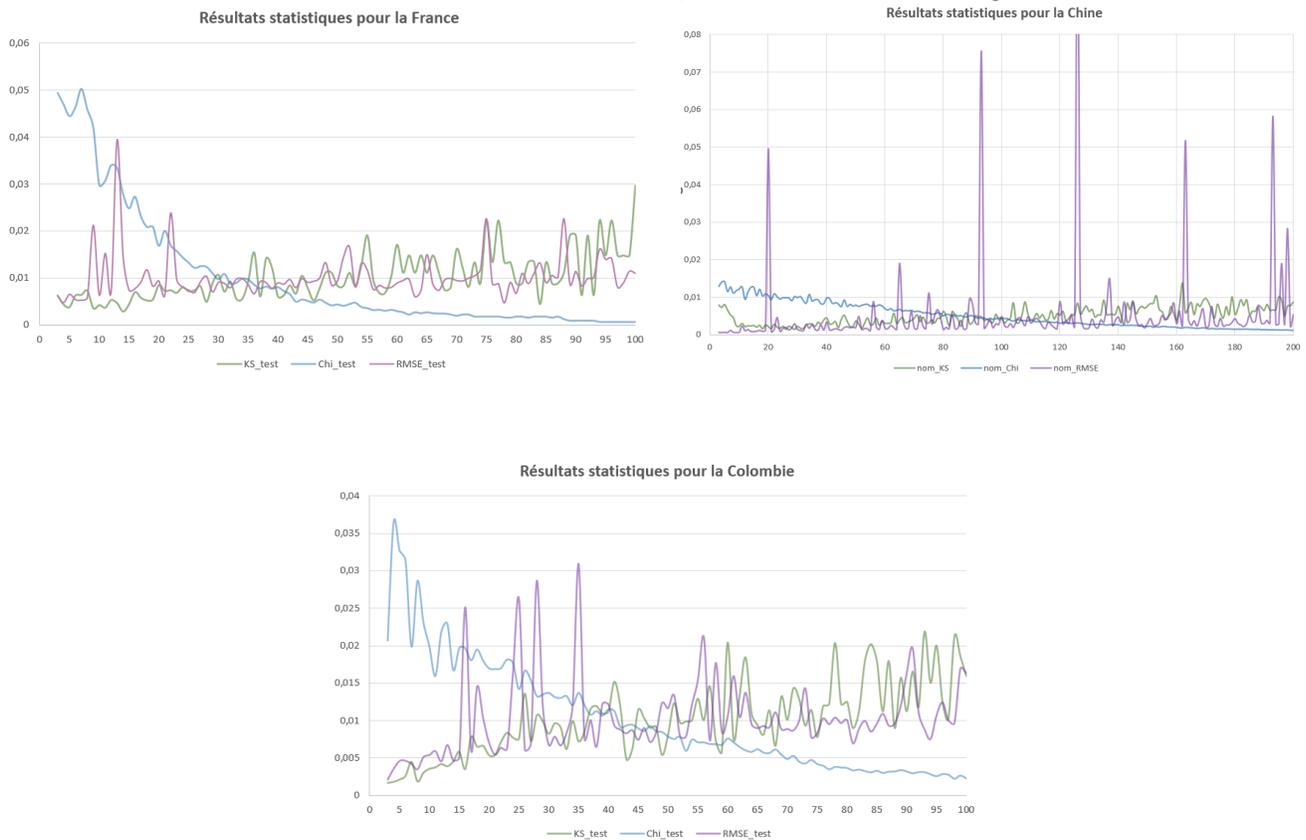


FIGURE 26 – Le résultat de tests statistiques

L'analyse des tests statistiques permet d'identifier des zones de seuil fiables pour l'application du modèle POT. Pour la France, la zone de stabilité se situe entre **28 et 43**, avec une attention particulière sur le seuil **33**, où les trois tests convergent de manière optimale. Concernant la Chine, la plage recommandée est plus large, entre **83 et 110**, avec un intérêt particulier pour l'intervalle **[83 ; 100]**, où les variations des tests sont plus modérées. Enfin, pour la Colombie, la zone fiable est comprise entre **37 et 50**, avec un seuil optimal identifié autour de **42**. Ces choix garantissent un équilibre entre la représentativité des valeurs extrêmes et la stabilité des estimations statistiques.

5.1.2 Estimateur de Hill

L'estimateur de Hill se révèle particulièrement efficace pour les distributions qui relèvent du domaine d'attraction de Fréchet, autrement dit, dans les cas où l'indice de queue ξ est positif.

La formule pour calculer l'estimateur de Hill :

$$\xi_{k,n}^{\hat{H}} = \frac{1}{k} \sum_{i=1}^k \ln \frac{X_{n-i+1}}{X_{n-k}} = \frac{1}{k} \sum_{i=1}^k \ln(X_{n-i+1}) - \ln(X_{n-k})$$

où : X_{n-i+1} les i plus grandes valeurs de l'échantillon

L'utilisation de l'estimateur de Hill se déroule de la manière suivante :

1. **Tri des données** : Les données sont triées par ordre décroissant pour se concentrer sur les valeurs les plus extrêmes ;
2. **Calcul des indices de queue** : Les indices de queue de Hill, notés $\xi_{k,n}^{\hat{H}}$ sont calculés pour chaque k , où k représente le nombre d'observations extrêmes utilisées. Ce paramètre varie entre 2 et n , afin d'explorer le comportement des queues selon différents seuils. ;
3. **Visualisation graphique** : Un graphique est tracé en représentant les paires ordonnées $(k, \xi_{k,n}^{\hat{H}})$. Ce graphique permet d'identifier les zones de stabilité de l'estimateur, essentielles pour déterminer un seuil optimal.

Pour déterminer le seuil des valeurs extrêmes à l'aide de cette méthode, appliquée à une distribution appartenant au domaine d'attraction de Fréchet, l'objectif est d'identifier les intervalles où l'estimateur de Hill $\xi_{k,n}^{\hat{H}}$ varie peu, formant une zone de stabilité relative. Cette stabilité se traduit graphiquement par une portion de la courbe approximativement horizontale, ou avec une pente proche de zéro. Elle indique que les données extrêmes respectent les hypothèses asymptotiques du modèle, rendant l'estimation plus fiable pour le choix du seuil optimal.

Par ailleurs, les zones où l'estimateur fluctue fortement doivent être évitées, car elles signalent soit un seuil trop bas, soit des données qui ne respectent pas les hypothèses asymptotiques nécessaires pour un ajustement fiable par une loi Pareto Généralisée.

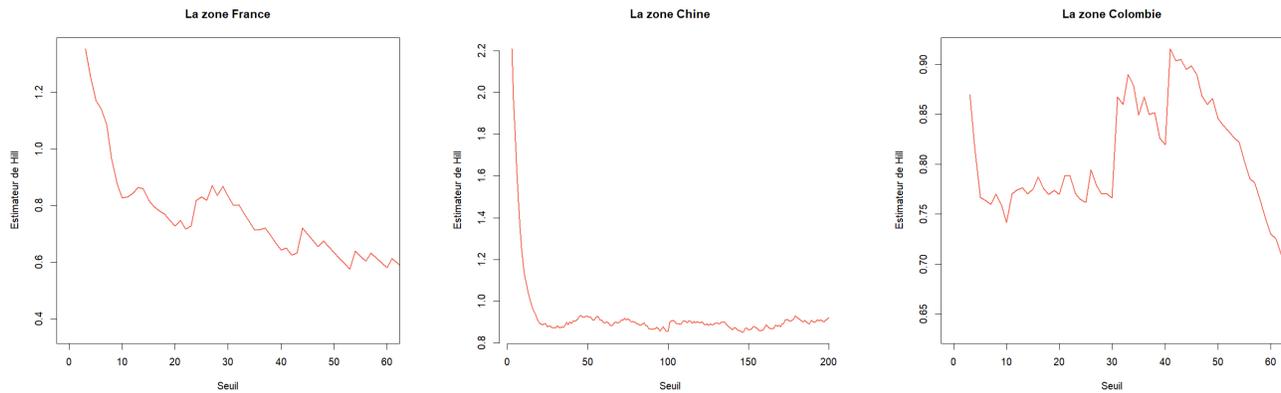


FIGURE 27 – Plot de Stabilité de l'Estimateur de Hill

Les trois graphiques montrent une phase initiale de forte variation, suivie d'une zone plus stable où l'estimateur fluctue légèrement.

- **La France**, la courbe commence avec une valeur élevée de l'estimateur, supérieure à 1,2, et décroît rapidement. Entre 30 et 50, l'estimateur oscille autour de 0,6 à 0,7, avant une légère tendance descendante. Après 50, la courbe semble maintenir une certaine stabilisation avec une pente moins marquée.
- **La Chine** : La courbe démarre à 2,2 et décroît rapidement jusqu'à le threshold environ 50, où elle atteint un palier plus stable. Entre 50 et 100, l'estimateur de Hill fluctue autour de 0,8 - 1, avec de légères variations. Après 100, la courbe conserve une tendance stable sans trop de pics.
- **La Colombie** : La courbe présente une pente descendante dans les premiers seuils (inférieurs à 20), atteignant une phase plus stable. Entre 30 et 50, la courbe montre une stabilisation relative, mais avec quelques pics. Après 50, une hausse soudaine est visible, suivie de fortes variations.

5.1.3 Vérification de la conformité du seuil optimal

Mean Excess Function (ou Fonction d'excès en moyenne)

La Mean Excess Function (MEF) est une méthode d'analyse des valeurs extrêmes qui, lorsque la variable X suit une distribution Pareto Généralisée (GPD), révèle des informations cruciales sur la nature des queues de distribution. La fonction d'excès en moyenne permet de décrire la prédiction du dépassement du seuil u lorsqu'un excès se produit. Elle est définie par :

$$e(u) = \mathbb{E}[X - u | X > u] \text{ avec } 0 \leq u \leq x_F$$

Où x_F est le point terminal de X .

Empiriquement, elle est estimée par le rapport entre la somme des excès dépassant un seuil élevé u et le nombre d'observations qui dépassent ce seuil :

$$\hat{e}_u = \frac{\sum_{i=1}^n (X_i - u)^+}{\sum_{i=1}^n \mathbf{1}_{[X_i > u]}}$$

Avec $(X_i - u)^+ = \max(X_i - u, 0)$

La MEF, fonction linéaire de u , peut présenter trois tendances distinctes à partir d'un certain seuil : décroissante, stable, ou croissante.

- **Décroissante** : Lorsque la MEF est en décroissance, cela indique une distribution à queue légère, souvent associée aux lois du domaine d'attraction de Weibull. Cela signifie que les valeurs extrêmes sont moins prononcées.
- **Stable** : Une MEF horizontale suggère que les données suivent une distribution exponentielle, typique des lois relevant du domaine d'attraction de Gumbel. Cela indique une régularité dans la distribution des valeurs extrêmes.
- **Croissante** : Si la MEF augmente au-delà d'un seuil spécifié, les données sont probablement issues d'une distribution à queue lourde, caractéristique des lois du domaine d'attraction de Fréchet. Cette tendance indique une probabilité accrue de survenance d'événements extrêmes.

Pour identifier le seuil optimal dans l'analyse des valeurs extrêmes, particulièrement pour les distributions relevant du domaine de Fréchet, l'objectif est de repérer les zones où la MEF se manifeste comme

une droite ascendante avec une pente positive. Cette approche aide à déterminer le seuil à partir duquel les observations commencent à montrer un comportement extrême, fournissant ainsi une base solide pour l'analyse des risques et la prise de décisions.

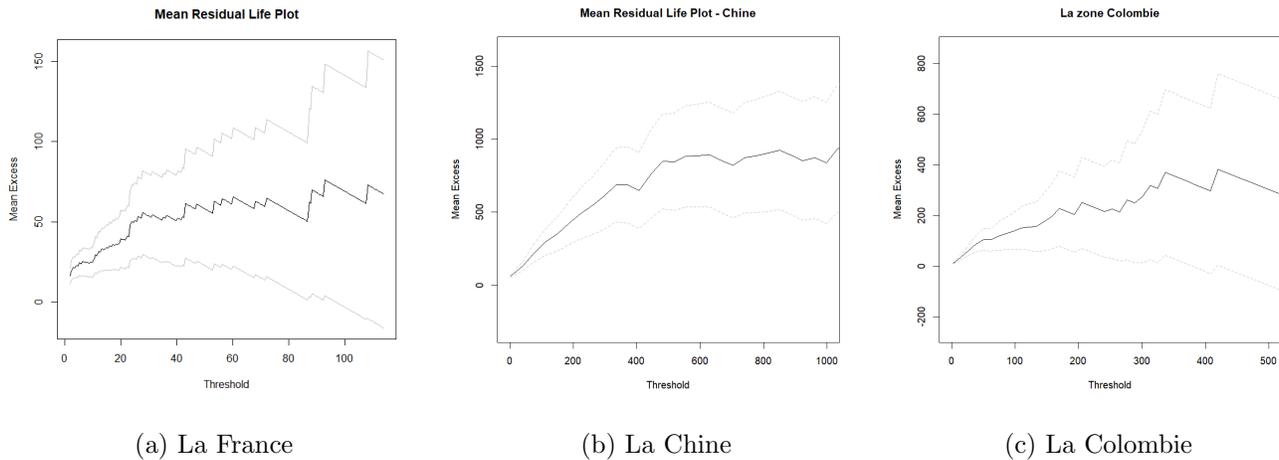


FIGURE 28 – Mean Residual Life Plot

En France, l'intervalle $[28;42]$ présente une progression relativement linéaire, ce qui constitue un indicateur clé de stabilité dans l'estimation. Cette stabilité suggère que cet intervalle correspond à une plage de seuils fiable. En dehors de cette plage, notamment au-delà de 80, la courbe devient moins cohérente et montre des variations importantes, indiquant que ces seuils sont probablement inadaptés en raison d'une perte de robustesse dans les estimations.

Comme en France, **la Chine et la Colombie** présentent une augmentation continue et marquée de l'excès moyen, indiquant un comportement compatible avec une queue lourde.

Pour identifier le seuil optimal, nous utilisons le Mean Residual Life Plot, qui repose sur la détection d'une zone où la pente devient linéaire et stable, indiquant que les valeurs extrêmes suivent une loi de Pareto généralisée (GPD). Un seuil trop bas inclurait des observations non extrêmes, biaisant l'estimation, tandis qu'un seuil trop élevé réduirait le nombre de données disponibles, rendant l'estimation instable.

Par ailleurs, les barres de confiance restent relativement étroites dans cet intervalle, ce qui confirme que les moyennes résiduelles sont estimées avec une précision satisfaisante et une incertitude réduite.

Threshold choice plot [Davison and Smith (1990)]

Cette approche vise à aider au choix d'un seuil approprié pour l'analyse des valeurs extrêmes. Les estimations pour chaque seuil, de la forme ξ et de l'échelle modifiée σ devraient rester constantes pour tout seuil $x_j > x_i$, en supposant que la distribution Pareto généralisée (GPD) au-delà de x_i est valide. Cette méthode permet de visualiser comment différentes statistiques ou paramètres estimés varient en fonction des choix de seuils.

Par ailleurs, ces graphiques servent à vérifier que le seuil choisi appartient à un intervalle de seuils appropriés. En examinant les courbes, on cherche une zone où les paramètres commencent à afficher un comportement stable. Ce comportement stable indique que les données au-delà de ce seuil respectent les hypothèses asymptotiques de la distribution GPD. Une zone est donc dite « stable » si les estimations restent cohérentes et que la bande de confiance demeure raisonnablement étroite.

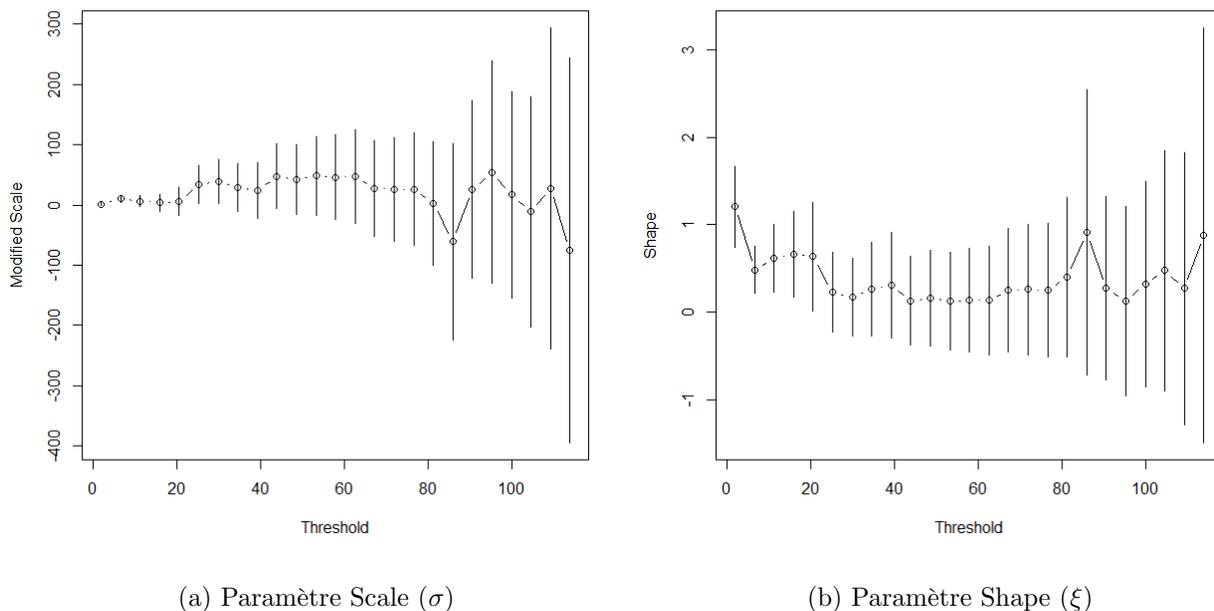


FIGURE 29 – Threshold choice Plot pour la France

Commentaire : Intervalle que l'on tend à choisir [28 ; 42] est liée à l'intervalle stable basée sur la méthode TCP.

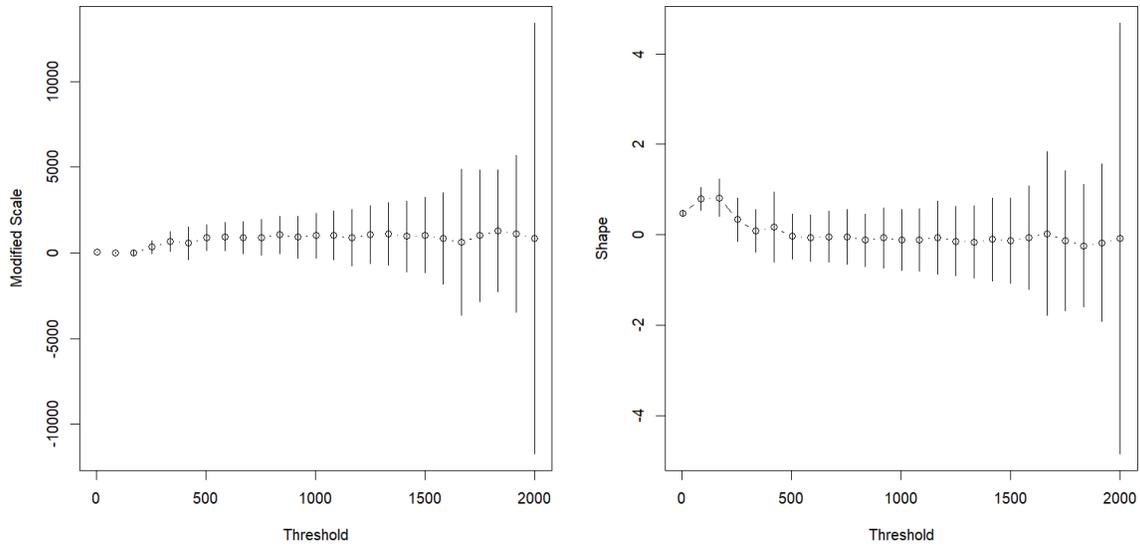


FIGURE 30 – Threshold choice Plot pour la Chine

Commentaire : On observe des intervalles de confiance très larges, en particulier pour les valeurs les plus élevées, ce qui indique une forte incertitude dans l'estimation des paramètres. De plus, la présence de warnings d'optimisation suggère des problèmes de convergence pour certains seuils. On va voir l'intervalle sur les quantiles 70% - 90%

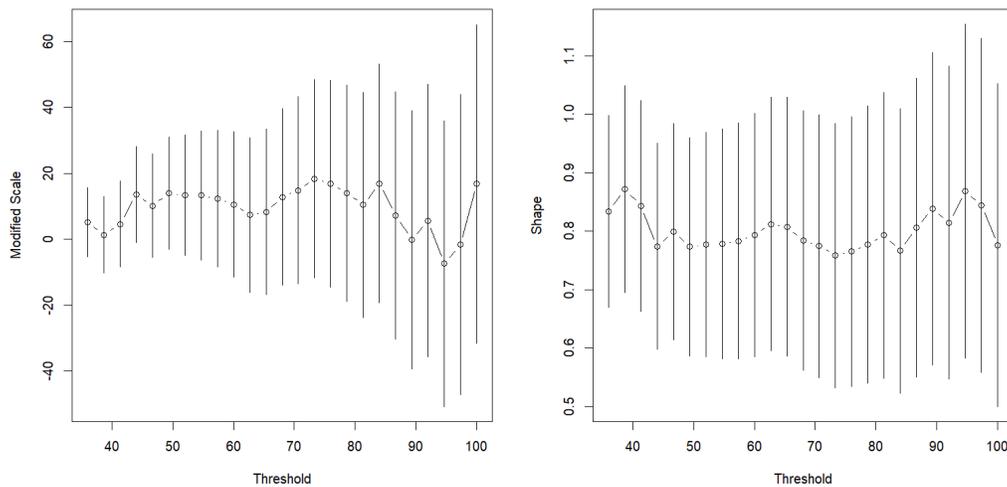


FIGURE 31 – Threshold choice Plot pour la Chine pour les quantiles [70% - 90%]

Dans la seconde image, nous avons recentré l'analyse sur la plage de seuils correspondant aux quantiles 70% - 90%, où les paramètres apparaissent plus stables. L'absence de warnings d'optimisation confirme que cette zone est plus appropriée pour la sélection du seuil. Cependant, bien que les intervalles de confiance soient réduits, ils restent encore relativement larges, en particulier pour le paramètre scale, ce qui implique une incertitude non négligeable. Donc, pour garantir un choix optimal, une validation croisée avec d'autres méthodes est essentielle afin de confirmer la stabilité du modèle et d'assurer une estimation fiable des valeurs extrêmes.

Pour la Colombie :

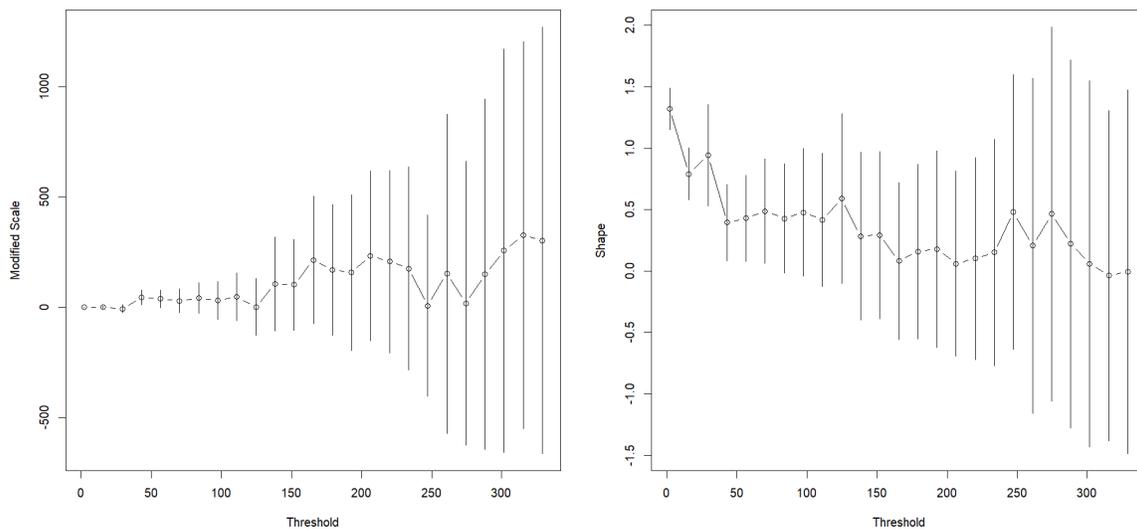


FIGURE 32 – Threshold choice Plot pour la Colombie

Commentaire : On observe une relative stabilité des paramètres estimés dans l'intervalle [40 ;100], bien que la variance des intervalles de confiance s'élargisse progressivement pour les valeurs de seuil plus élevées. Cependant, des problèmes d'optimisation apparaissent pour les quantiles supérieurs 98% (le point 71). La zone comprise entre 40 et 60 est un choix approprié pour la modélisation du POT, en équilibrant la robustesse et la fiabilité des estimations.

Q-Q Plot

Dans cette partie, nous utilisons les QQ-plots pour visualiser le niveau d'adéquation entre les observations et les simulations basées sur la distribution GPD pour l'intervalle appropriée de chaque zone. Ces graphiques comparent les quantiles empiriques avec les quantiles théoriques, permettant d'évaluer si les données suivent correctement le comportement attendu d'une loi de Pareto généralisée.

Les étapes principales de cette méthode sont le test et le calcul du x -ième quantile. Avec la distribution GPD, pour une probabilité $\alpha \in [0; 1]$, le x -ième quantile q_α est donné par :

$$q_\alpha = \mu + \frac{\sigma}{\xi} [(1 - \alpha)^{-\xi} - 1]$$

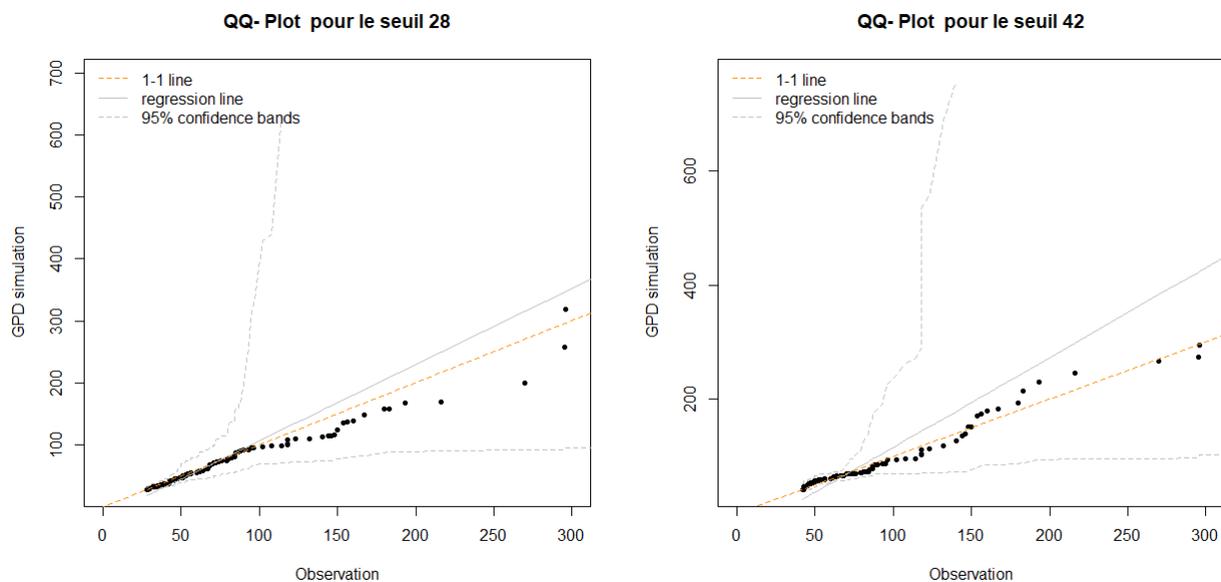


FIGURE 33 – QQ-Plot pour la zone France

D'après les graphiques, la distribution avec un seuil de 42 correspond mieux aux observations que celle avec 28. Cela est visible par la proximité accrue des points avec la ligne 1 :1 pour 42, indiquant une meilleure adéquation du modèle. Cependant, ce seuil plus élevé peut être affecté par un problème de surajustement, capturant davantage de bruit que de signal pertinent. Pour éviter cela, nous considérons que le seuil optimal se situe entre ces deux valeurs.

En examinant davantage la stabilité des paramètres et les résultats globaux, nous choisissons 35 comme seuil optimal. Ce choix est basé sur une meilleure stabilité des estimations, une réduction du risque

de surajustement, et l'accord avec les critères définis pour détecter les événements catastrophiques en France. Ce seuil garantit que les données au-delà capturent correctement les extrêmes tout en évitant une inadéquation due à un seuil trop bas ou trop élevé.

Pour la Chine :

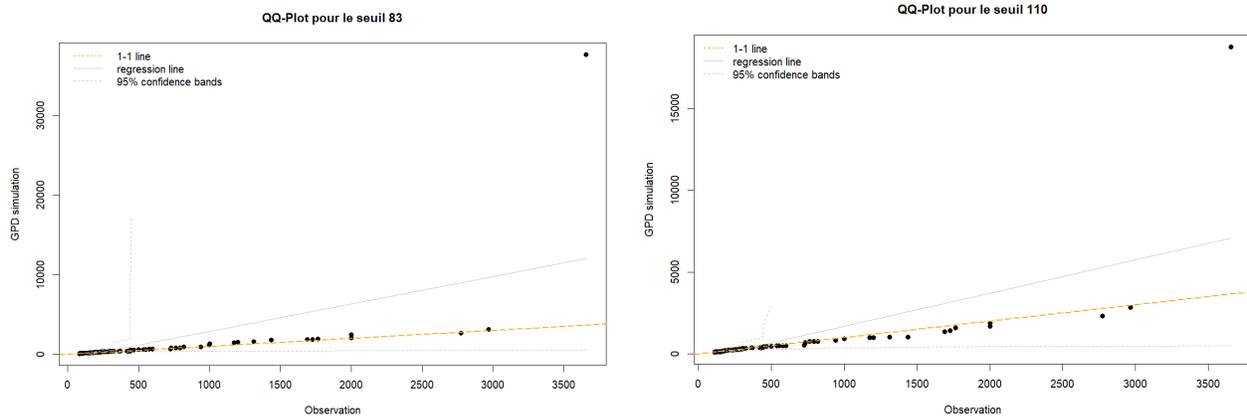


FIGURE 34 – QQ-Plot de la zone Chine pour le seuil 83 et 110

Pour le seuil 83 (graphique de gauche) et le seuil 110 (graphique de droite), la majorité des points suit la ligne de référence, indiquant une bonne adéquation globale du modèle. En comparant les deux graphiques, on observe qu'un point aberrant extrême est présent dans les deux cas, mais il est moins éloigné de la droite pour le seuil 110, suggérant un ajustement plus stable pour les valeurs extrêmes. Nous pouvons supposer que dans l'intervalle [83 ; 110], le modèle POT correspond assez bien à notre observations.

Pour la Colombie :

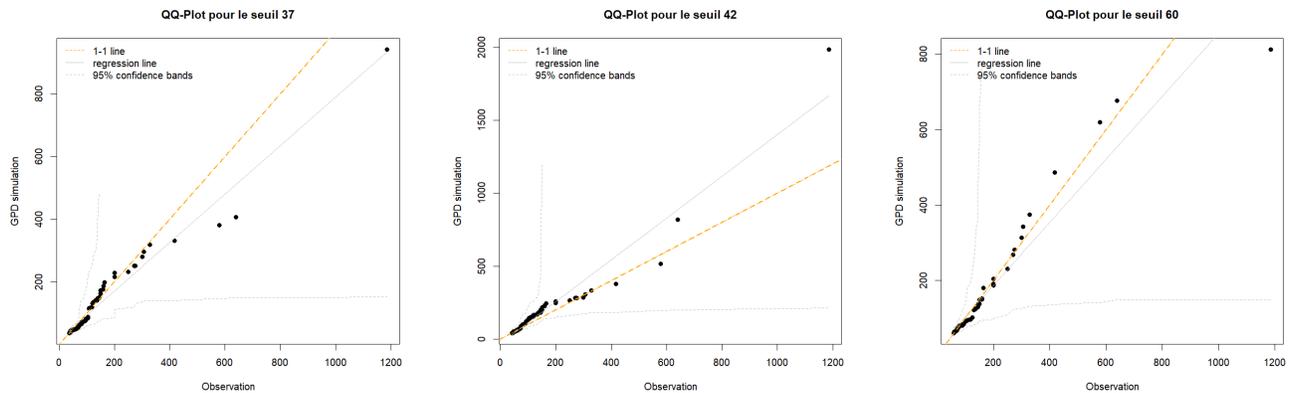


FIGURE 35 – QQ Plot de la zone Colombie pour des les seuils 37 ; 42 et 60

Commentaire : Parmi les trois seuils analysés pour la zone Colombie, le seuil 42 apparaît comme le choix optimal. En effet, les QQ-Plots montrent un sous-prix notable pour les seuils 37 et 60, où les observations associées au nombre de victimes tendent à se situer sous la ligne 1 :1. Cela indique une sous-estimation du modèle GPD dans ces cas. En revanche, pour le seuil 42, l’alignement des points avec la ligne de régression et la bande de confiance est plus stable, suggérant une meilleure adéquation entre les données observées et les valeurs simulées.

5.2 Groupement

Cette section a pour objectif d'améliorer la robustesse de notre modèle statistique, actuellement limitée par le faible volume de données disponibles. En effet, la rareté des événements catastrophiques rend difficile l'estimation fiable des paramètres des distributions utilisées, telles que la loi de Poisson, la Binomiale Négative ou la distribution de Pareto généralisée (GPD). Cette contrainte statistique affecte directement la précision des simulations et des projections.

Pour surmonter cette contrainte, nous proposons une méthodologie consistant à regrouper les données provenant de plusieurs pays proches. Ce regroupement vise à augmenter la taille de l'échantillon tout en garantissant la cohérence des données pour estimer les paramètres du modèle.

Dans la première partie, nous portons une attention particulière aux événements naturels extrêmes, car les conditions climatiques tendent à présenter des similitudes parmi les pays d'un même continent, ce qui justifie leur regroupement continental dans notre analyse. Concernant les événements terroristes, nous supposons que les pays géographiquement proches sont affectés de manière similaire en raison de menaces et de défis sécuritaires communs liés à leur proximité.

Dans la section «Détermination du seuil optimal», le choix du seuil a été étudié sur trois zones distinctes : la France, la Chine et la Colombie, des pays ayant un volume d'observations suffisant pour des résultats fiables. Cependant, pour la Chine et la Colombie, les pays voisins sont généralement de plus petite taille et ne permettent pas une comparaison précise, notamment en raison de la différence dans la nature des événements extrêmes (par exemple : catastrophes naturelles et terrorisme). Ainsi, la suite de l'analyse se concentre sur la zone France et ses pays voisins, où le regroupement des données est plus pertinent pour améliorer l'estimation des événements extrêmes.

La France, situé en Europe de l'Ouest, est bordé par la mer Méditerranée au sud, l'océan Atlantique à l'ouest, et la Manche au nord, et partage des frontières terrestres avec plusieurs pays, notamment la Belgique, l'Allemagne, la Suisse, l'Italie et l'Espagne. En regroupant des pays partageant des conditions climatiques similaires à celles de la France, nous garantissons que les observations combinées restent pertinentes pour l'analyse des événements extrêmes.

5.2.1 Logique du regroupement dans notre modèle

Cette approche consiste à regrouper les pays en fonction de leur proximité géographique par rapport à un pays de référence. Caractéristiques communes : géographiques, climatiques, économiques et les niveaux d'industrialisation.

Ce regroupement des pays renforce la robustesse de l'analyse en élargissant la base de données. Il facilite l'étude des tendances et des modèles à une échelle régionale, offrant une compréhension plus globale que l'analyse isolée d'un seul pays. Cette méthode permet également de mieux anticiper des scénarios inédits, en intégrant des informations provenant de plusieurs contextes similaires.

Parallèlement, nous intégrons le paramètre pr_{region} :

$$pr_{region} = \frac{\text{Nombre des événements survenus dans le pays cible}}{\text{Nombre des événements survenus dans toute la region regroupée}}$$

L'objectif principal est de simuler une variété de scénarios pour chaque année, en choisissant soit *la loi de Poisson*, soit *la loi Binomiale Négative*, avant d'appliquer la loi Binomiale en fonction de la probabilité p . Cette technique est dénommée « sampling échantillon ».

$$K' \sim Bin(K, pr_{region})$$

5.2.2 Méthodes analytiques pour le choix des pays à regrouper

Analyse globale des événements catastrophiques

	Drought	Earthquake	Extreme temperature	Flood	Landslide	Storm	Terrorism	Wildfire
Austria				33	63	7	29	
Belgium		2	1 873	5		27	63	
France			5 263	136	162	338	420	20
Germany		1	33	27	5	122	130	
Ireland						26	55	
Italy		332	20	466	164	92	277	7
Netherlands			400			28	28	
Portugal			462	71		39	17	41
Spain			88	102		82	728	11
Switzerland				9	76	17	67	
United Kingdom			3 331	16		233	1 475	

FIGURE 36 – Nombre de évènements dans les pays voisins dus à divers événements catastrophiques

La France présente une vulnérabilité marquée aux risques climatiques et naturels, tels que les tempêtes, les inondations, les vagues de chaleur ou encore les glissements de terrain. Cette exposition est comparable à celle de plusieurs de ses voisins : l'Italie, par exemple, est également sujette aux inondations et aux glissements de terrain, tandis que le Royaume-Uni connaît une fréquence plus élevée de tempêtes et a été affecté par plusieurs actes de terrorisme.

Ces similitudes soulignent l'intérêt d'une analyse comparative entre pays, afin de mieux comprendre les dynamiques des événements extrêmes et d'en améliorer l'anticipation dans une perspective actuarielle et assurantielle.

Analyse des tendances mensuelles des événements naturels

Notre démarche méthodologique commence par le traitement et la normalisation des données sur le nombre de décès et le nombre d'événements extrêmes, sur une base mensuelle pour chaque pays. Cette normalisation, réalisée dans l'intervalle $[0,1]$, permet de standardiser les données afin de garantir leur comparabilité. Elle neutralise les biais potentiels liés aux différences de population ou à la fréquence absolue des événements, rendant ainsi les analyses plus robustes et pertinentes.

$$P_m = \frac{D_m}{D_{\text{total}}} \times 100$$

avec :

- P_m : la proportion (%) des décès survenus au mois (par exemple : Janvier, Février,...)
- D_m : le nombre total de morts au mois,
- D_{total} : le nombre total de morts tous mois et toutes années confondus.

Nous traçons le graphique suivant :

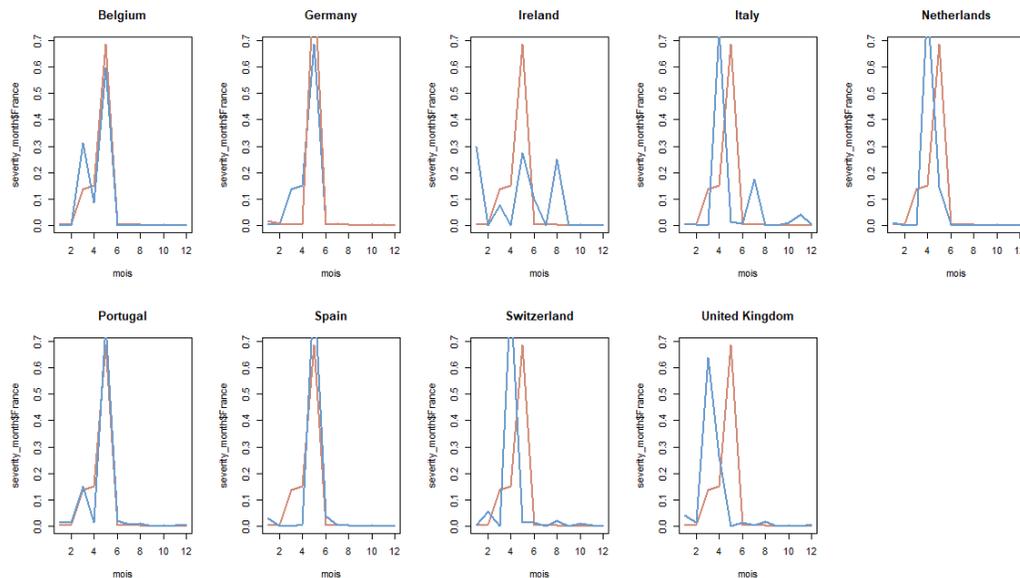


FIGURE 37 – Distribution mensuelle des décès (événement naturel origine) entre la France et les pays voisins

Des graphiques sont produits pour chaque pays, offrant une visualisation des tendances, des saisonnalités aux décès mensuels causés par des événements naturels extrêmes. Ces visualisations montrent une saisonnalité marquée, avec des pics observés principalement entre les mois de mai et juillet, le mois de juin étant souvent le mois de pointe pour la majorité des pays. La France présente un profil similaire à des pays tels que l'Allemagne, la Suisse et les Pays-Bas, avec des distributions mensuelles très proches, indiquant des contextes climatiques et saisonniers comparables. En revanche, des pays comme le Portugal et l'Espagne affichent des pics légèrement décalés vers mai, probablement en raison de spécificités climatiques propres à l'Europe du Sud. L'Irlande et le Royaume-Uni, quant à eux, présentent des distributions plus étalées ou légèrement décalées, reflétant des différences géographiques et climatiques.

Ces observations visuelles servent de base à des analyses statistiques complémentaires. Une combinaison d'approches, incluant l'analyse de corrélation et le clustering hiérarchique, est utilisée pour regrouper les pays selon leurs similarités climatiques et saisonnières. Le clustering hiérarchique permet d'identifier des regroupements naturels entre pays, tandis que l'analyse de corrélation quantifie la force des relations entre leurs distributions mensuelles. Pour enrichir l'interprétation, une "heatmap" est utilisée pour visualiser les corrélations, offrant un aperçu clair des tendances communes et des écarts observés entre les pays.

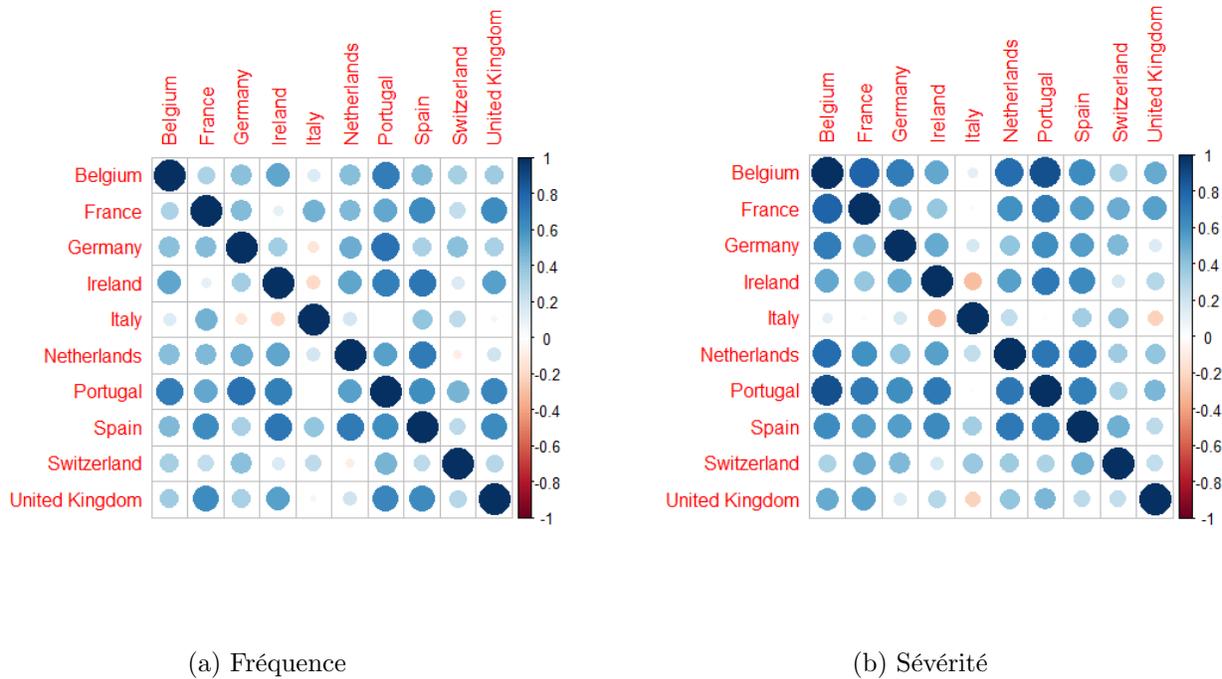


FIGURE 38 – «Heatmap» de corrélation

La matrice de corrélation révèle des similitudes variées entre les distributions mensuelles des pays européens. Des corrélations élevées, comme entre la France, l'Allemagne et les Pays-Bas, reflètent des cycles climatiques similaires. En revanche, des pays comme l'Irlande et le Royaume-Uni affichent des corrélations plus faibles, probablement en raison de leurs particularités géographiques et climatiques, comme un climat océanique modéré et des variations saisonnières moins prononcées.

La validation des groupes de pays voisins est une étape essentielle, effectuée à travers une série de vérifications visant à assurer la cohérence et la distinction entre les pays.

K-Mean cluster

Pour regrouper les pays en fonction de leur exposition aux événements naturels, nous utiliserons l'algorithme K-Means, qui permet de former des clusters homogènes en se basant sur l'impact des événements naturels dans chaque pays. Dans un premier temps, les données, incluant le nombre d'événements et de décès par pays, seront collectées pour obtenir une vue d'ensemble claire. Ensuite, un ajustement sera effectué pour standardiser l'échelle de chaque variable. Cet ajustement, basé sur la méthode de standardisation Z-score, permet de rendre les variables comparables en prenant en compte leur moyenne et écart-type, ce qui est comparable pour éviter qu'une variable domine le regroupement.

$$x_{normalise} = \frac{x - \mu}{\sigma}$$

Où μ est la moyenne des valeurs de la caractéristique de x et σ est l'écart-type de x.

Pour décider combien de groupes (ou « clusters ») créer, nous allons utiliser la méthode du Coude. Cela consiste à regarder comment les distances entre les points et le centre de leur groupe changent quand on augmente le nombre de groupes. On cherche le moment où ajouter plus de groupes ne réduit plus beaucoup ces distances.

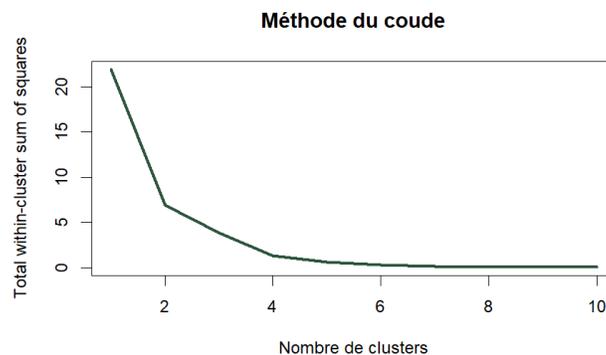


FIGURE 39 – Map «coude» - nombre de cluster

Nous utilisons l'algorithme K-Means sur nos données après les avoir normalisées, en choisissant le nombre de groupes, que l'on voudrait créer. K-Means va regrouper les pays les plus proches.

L'analyse de clustering K-Means distingue clairement deux groupes de pays en fonction du nombre total d'événements et de décès liés aux catastrophes. Le Cluster 1 regroupe la majorité des pays, tels que

	Pays	Nb_Event	Nb_Total_Deaths	Cluster
1	Austria	28	427	1
2	Belgium	26	2085	1
3	Cyprus	7	32	1
4	France	153	7270	2
5	Germany	69	613	1
6	Ireland	18	408	1
7	Italy	107	3279	1
8	Netherlands	19	627	1
9	Portugal	28	960	1
10	Spain	234	2332	2
11	Switzerland	19	324	1
12	United Kingdom	409	5453	2

FIGURE 40 – Tableau des pays regroupés par clusters selon leur exposition aux catastrophes naturelles (résultat du K-Means)

l’Autriche, la Belgique et l’Italie, caractérisés par un impact modéré avec un nombre limité d’événements et de décès. En revanche, le Cluster 2 comprend des pays comme la France, le Royaume-Uni et l’Espagne, qui présentent des valeurs extrêmes, avec un nombre élevé d’événements et de décès, reflétant une exposition et une vulnérabilité accrues. La France se distingue par le plus grand nombre de décès (7270), tandis que le Royaume-Uni enregistre le plus grand nombre d’événements (409).

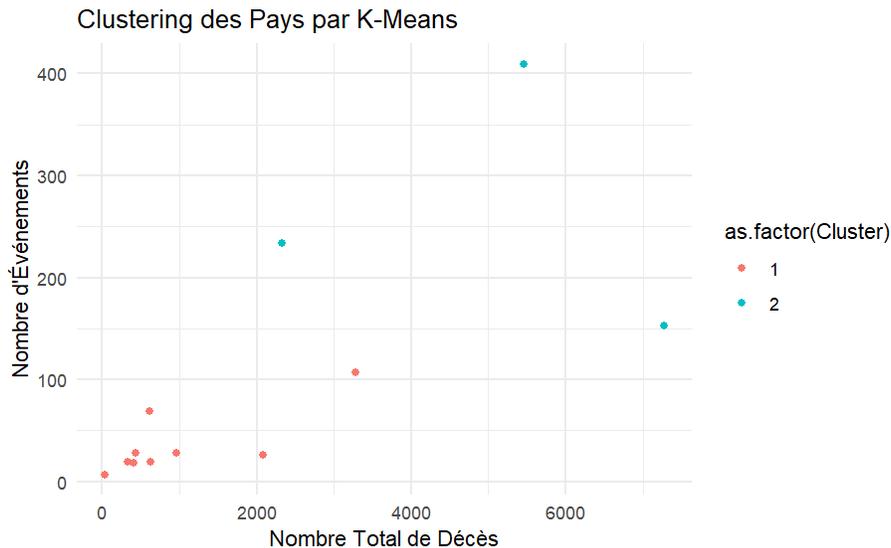


FIGURE 41 – Illustration de K-Means Cluster

Pour résumer : Le regroupement des pays en fonction de la gravité des catastrophes naturelles repose sur plusieurs analyses complémentaires. Tout d’abord, une analyse des événements climatiques montre

que la France et l'Angleterre ont subi les plus grandes pertes lors d'événements extrêmes tels que les vagues de chaleur et les tempêtes. Ensuite, en observant l'évolution du nombre d'événements et de décès sur une période de 48 ans, des tendances similaires apparaissent entre la France et l'Allemagne, la France et la Belgique, la France et l'Espagne, ainsi que la France et le Portugal. Cette proximité est confirmée par des graphiques de corrélation qui mettent en évidence une relation positive entre ces pays sur les critères de fréquence et de sévérité des catastrophes. Enfin, l'application de la méthode de classification K-Means permet d'identifier des groupes de pays partageant un niveau de gravité similaire, confirmant que l'Espagne, l'Italie et la France peuvent être regroupées en raison de leurs profils de risques comparables.

	Probabilité « p »
France	25,50%
Spain	17,90%
Portugal	3,90%
Italy	22,92%
Belgium	4,37%
United Kingdom	25,41%

6

Calibrage et Sensibilité du modèle

Ce chapitre a pour objectif d'évaluer la robustesse et la fiabilité du modèle en analysant l'influence des choix méthodologiques sur les résultats obtenus. Il s'agit d'identifier les sensibilités aux hypothèses formulées, afin de garantir la stabilité des estimations et la cohérence globale du modèle.

Concrètement, nous comparons deux modélisations de la fréquence des événements catastrophiques (distribution de Poisson vs. Binomiale négative), nous analysons l'impact du paramètre de structure θ sur le nombre de victimes assurées, et nous validons le seuil retenu à l'aide du critère de maximum de vraisemblance appliqué à la loi de Pareto généralisée (GPD).

Pour valider notre approche, nous utilisons les données recensant les catastrophes naturelles et humaines survenues entre 1970 et 2018, en France et dans les pays voisins. Bien que le modèle ait été calibré à l'échelle de l'ensemble du portefeuille international, l'analyse approfondie est ici centrée sur la France.

6.1 Estimation de la fréquence des événements

Les estimateurs de fréquence et de sévérité sont calculés à partir de cette base de données. Reste à savoir si les lois de probabilités choisies dans notre modèle reflètent bien les tendances de notre jeu de données empirique.

Pour rappel, la variable aléatoire K_M (nombre de catastrophes causant au moins M décès) suit :

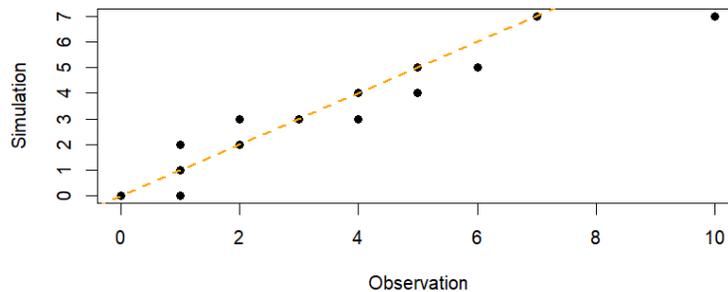
- Une loi de Poisson de paramètre K_M si $E[K_M] \approx \mathbf{Var}[K_M] = \lambda_M$
- Une loi binomiale de paramètres α_M et μ_M si $\mathbf{E}[K_M] < \mathbf{Var}[K_M]$

Le choix entre une loi de Poisson et une loi Binomiale Négative peut être déterminé en comparant l'espérance $E[K_M]$ et la variance $\mathbf{Var}[K_M]$.

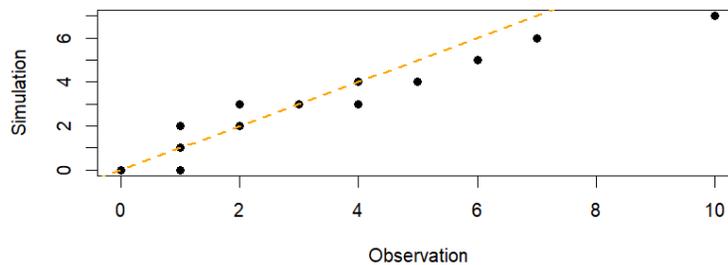
Dans le cas de la zone France (avec le seuil 35) on trouve ainsi, $E[K_M] = 2,4$ et $Var[K_M] = 3,8$.

Ce qui donne un ratio $\frac{Var[K_M]}{E[K_M]} = 1,6$. L'espérance et la variance ne sont pas loins, il ne semblerait pas aberrant de choisir une loi de Poisson.

Nous utilisons la méthode graphique du QQ-plot qui trace les quantiles observés et ajustés. Cela permet de valider la conformité entre la distribution empirique et théorique en observant l'alignement des points sur la droite 1 : 1 qui a la forme : $y = x$. Plus les points sont alignés, meilleure est l'approximation. Voici le QQ-plot pour la loi de Poisson et la binomiale négative :



(a) Loi de Poisson



(b) Loi Binomiale Négative

FIGURE 42 – Q-Q Plot comparant la loi de Poisson et la loi binomiale négative

A travers *le graphique 42*, il est raisonnable de choisir la distribution de Poisson pour notre modèle avec les données de la zone France.

Dans certaines régions d'Asie et d'Amérique du Sud, la variance peut être jusqu'à 10 fois supérieure à l'espérance. Cette forte surdispersion s'explique par une variabilité extrême des catastrophes naturelles

(CAT) : certaines sont de faible intensité, tandis que d'autres entraînent des dommages considérables. Ces régions enregistrent un grand nombre d'événements CAT de tailles très variables, rendant la modélisation du risque particulièrement complexe.

6.2 Sensibilité de l'estimation des victimes assurées

Pour rappeler, Afin de modéliser la proportion de victimes assurées parmi les victimes totales d'un événement donné, nous notons q la part de marché du portefeuille assuré — c'est-à-dire la probabilité théorique qu'une victime soit assurée dans le portefeuille étudié.

Cependant, lors d'un événement donné ayant causé X victimes, la proportion réelle de victimes assurées — notée $p | X$ — peut varier autour de q , notamment en fonction de l'intensité de la catastrophe et de la répartition géographique des assurés.

Pour modéliser cette incertitude, nous supposons que la variable aléatoire $p | X$ suit une loi bêta centrée en q :

$$p | X \stackrel{\mathcal{L}}{\sim} \text{Beta}(d(X)q, d(X)(1 - q))$$

Et puis, le nombre de victimes assurées lors d'une catastrophe est estimé en appliquant un facteur d'ajustement $d(X)$, qui reflète la proportion des assurés parmi les victimes. Cette proportion suit une relation logarithmique avec le nombre total de décès, définie par la fonction $d(X) = \theta * \log(X)$, où θ est un paramètre à calibrer.

La fonction logarithmique $\log(X)$: L'utilisation de $\log(X)$ signifie que l'impact comme une croissance décroissante de pour victime supplémentaire :

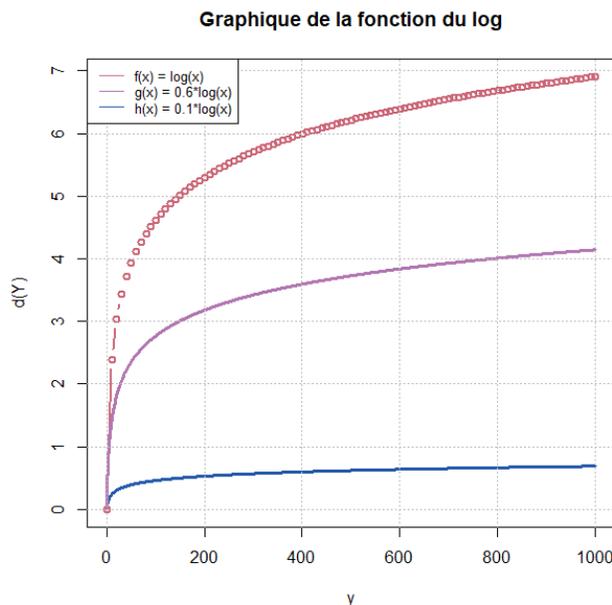


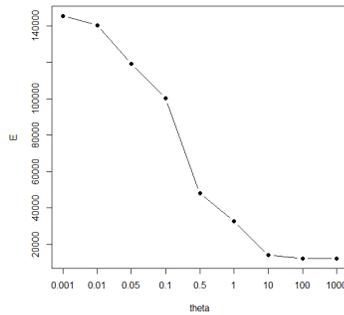
FIGURE 43 – Fonction logarithme

A côté de la fonction logarithme, le coefficient θ : Le coefficient θ est un facteur d'échelle qui mesure l'impact du nombre de victimes sur la fiabilité de l'estimation de p .

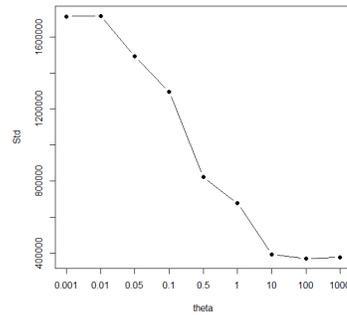
- Portefeuille à 100% Collectif : Dans un portefeuille collectif, les risques sont mutualisés entre un grand nombre d'individus appartenant à une même région géographique. La faible valeur de θ traduit cette stabilité : même si le nombre de victimes augmente, la dispersion de l'estimation de $d(X)$ reste limitée. Le risque est lissé, et l'estimation de $p | X$ reste proche de sa moyenne.
- Portefeuille à 100% Individuel : La valeur élevée de θ reflète le fait que le supplément ou la diminution d'individus au portefeuille a un impact plus significatif sur la confiance dans l'estimation de p , en raison de la grande variabilité des risques individuels.

Ce paramètre est construit par le fait que plus le traité possède des assurés indépendants élevé (les risques couverts sont largement dispersés), à propos de la (ré)assureur, plus sa probabilité d'être lourdement impacté par un évènement CAT est faible. Du graphique de la fonction logarithme, on peut voir la sensibilité quand on change la valeur de θ . Lorsque θ est grand, la valeur de $d(Y)$ est plus grande et également, la sensibilité est plus significative.

Concernant du changement du coût pure en valeur de θ :



(a) Changement du $E(C)$ en lien de θ



(b) Changement du $Std(C)$ en lien de θ

FIGURE 44 – Changement de θ

On peut facilement voir la décroissance du coût de $\theta = 0.01$ à $\theta = 10$. Si tous les éléments restent les mêmes, on peut voir 2 cas principales : totalement dépendante (100% contrat collective) et totalement indépendante (100% contrat individuel).

Si on montre par graphique la relation entre θ et q :

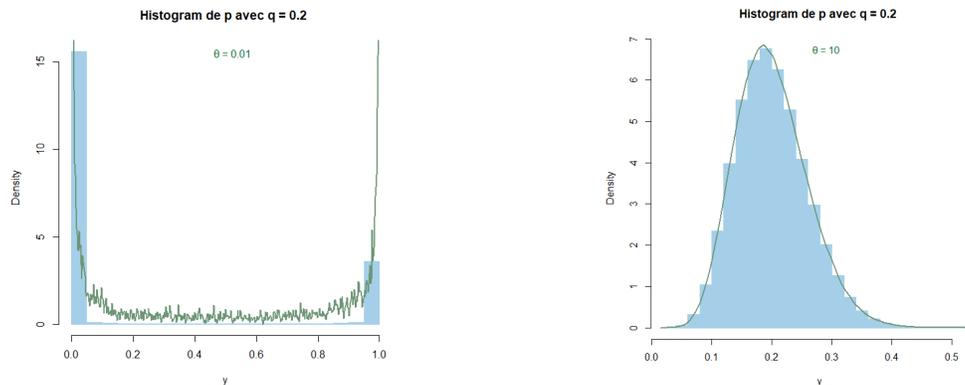


FIGURE 45 – Distribution de p

Pour un portefeuille combinant l'élément individuel et l'élément collectif, le paramètre θ_{mixte} est défini comme une pondération de θ , ajustée en fonction du niveau moyen de la composante collective dans le traité. La formulation :

$$\theta_{mixte} = \theta_{groupe} * Pourcentage_{group} + \theta_{individuel} * (1 - Pourcentage_{group})$$

assure une transition harmonieuse entre les évaluations des risques au niveau macro et au niveau micro.

6.3 Ajustement par maximum de vraisemblance de la distribution de Pareto généralisée

A travers la présentation du modèle, l'un des problèmes provient de l'optimisation des paramètres. L'estimation des paramètres dans notre modèle rencontre plusieurs défis qui peuvent affecter la robustesse des résultats obtenus :

- **Non-optimisation des paramètres** : Dans certains cas, l'algorithme d'optimisation ne parvient pas à converger vers une solution optimale, ce qui entraîne une instabilité des estimations et un ajustement inadéquat aux données. Ce problème survient principalement lorsque le seuil sélectionné est trop élevé, réduisant ainsi le nombre de points utilisés pour l'estimation et rendant l'optimisation impossible.
- **Problème lié au paramètre de forme (scale) ξ** : Lorsqu'on obtient une estimation de supérieure à 1 ($\xi > 1$), la variance de la distribution ne sont plus définies.

$$E(X) = \frac{\sigma}{1 - \xi}, \quad \text{pour } \xi < 1$$

$$\text{Var}(X) = \frac{\sigma^2}{(1 - \xi)^2(1 - 2\xi)}, \quad \text{pour } \xi < \frac{1}{2}$$

Cette absence de moments finis entraîne une incapacité à contrôler la queue de la distribution, ce qui peut conduire à une sur-simulation des valeurs extrêmes et biaiser l'estimation des quantiles de risque. On peut accepter ξ dans l'intervalle $[0,5;1]$, car cela correspond à une distribution à queue lourde, impliquant une forte probabilité d'observer des valeurs extrêmes.

- **Problème de convergence locale et globale** : L'optimisation par maximum de vraisemblance peut conduire à un minimum local plutôt qu'au minimum global. Cela signifie que l'algorithme peut s'arrêter prématurément sur une solution sous-optimale, limitant ainsi la qualité de l'ajustement.

On va commencer par la fonction de vraisemblance pour un échantillon indépendant x_1, x_2, \dots, x_n suivant une distribution de Pareto généralisée (GPD) est donnée par :

$$L(\xi, \sigma) = \prod_{i=1}^n f(x_i; \xi, \sigma)$$

En remplaçant par la densité de probabilité de la GPD :

$$L(\xi, \sigma) = \prod_{i=1}^n \frac{1}{\sigma} \left(1 + \xi \frac{x_i - u}{\sigma} \right)^{-\frac{1}{\xi} - 1}$$

En prenant le logarithme pour obtenir la fonction de log-vraisemblance :

$$\ell(\xi, \sigma) = \sum_{i=1}^n \log f(x_i; \xi, \sigma)$$

En substituant $f(x; \xi, \sigma)$:

$$\ell(\xi, \sigma) = \sum_{i=1}^n \left[-\log \sigma - \left(\frac{1}{\xi} + 1 \right) \log \left(1 + \xi \frac{x_i - u}{\sigma} \right) \right]$$

En développant la somme :

$$\ell(\xi, \sigma) = -n \log \sigma - \left(\frac{1}{\xi} + 1 \right) \sum_{i=1}^n \log \left(1 + \xi \frac{x_i - u}{\sigma} \right)$$

Nous allons vérifier nos estimateurs de vraisemblance en analysant la sensibilité des paramètres optimisés aux conditions initiales. Pour cela, nous suivons les étapes suivantes :

- **Variation des points de départ** ($\xi_0 ; \beta_0$) : nous testons plusieurs valeurs initiales pour ξ et β afin d'évaluer la robustesse de l'optimisation.
- **Traçage de la log-vraisemblance en fonction de ξ** : cette visualisation permet d'identifier si l'algorithme atteint un maximum global ou s'il est piégé dans un maximum local.
- **Analyse de la stabilité des paramètres estimés** : si l'estimation de ξ varie fortement en fonction du point de départ, cela indique une sensibilité élevée et donc une instabilité potentielle du modèle.

Le graphique ci-dessous illustre la stabilité de l'estimation du paramètre ξ en fonction de sa valeur de départ dans l'optimisation par maximum de vraisemblance (MLE). Sur le premier graphique, nous observons que l'estimateur de ξ reste constant, indépendamment de la valeur initiale choisie. Cela indique que l'optimisation converge systématiquement vers un même point, ce qui suggère une absence de sensibilité au choix du point de départ.

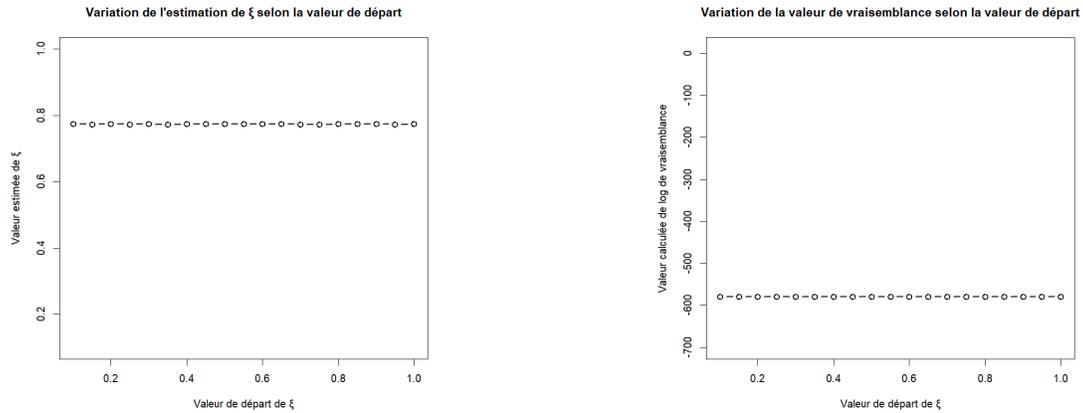


FIGURE 46 – Variation des paramètres de GPD avec MLE

Le second graphique montre que la valeur du log-vraisemblance obtenue reste également constante pour toutes les valeurs initiales de ξ . Ce comportement révèle une convergence robuste et confirme l'existence d'un maximum global stable pour la fonction de vraisemblance. Toutefois, cette invariabilité peut également traduire un possible manque de flexibilité du modèle ou une optimisation contrainte par la nature des données utilisées.

7

Tarification du traité

Nous reprenons les terminologies du traité introduit précédemment. Rappelons en les caractéristiques :

- Risque couvert
- Pays de souscription
- Priorité
- Portée
- Nombre de têtes déclenchant de traité
- Nombre d'assurés, donc q
- Somme sous risque moyenne
- Nombre de reconstitutions
- Composition du portefeuille (% collectif, % individuel)

À partir des informations du traité, nous initierons notre outil de tarification en nous basant sur les estimations présentées ci-dessus. Dans le cadre de notre étude, nous nous concentrerons sur trois principales zones : la France, la Chine et la Colombie. En nous basant sur le graphique QQ Plot, nous modéliserons la fréquence des événements CAT par an à l'aide de la distribution de Poisson.

	Estimation des paramètres				
Zone	Seuil	P_{region}	Poi.λ	GPD.β	GPD.ξ
France	35	25.50%	2.39	29.09	0.74
Chine	100	100%	3.45	76.84	0.88
Colombie	42	100%	0.67	183.00	0.05

Dans notre périmètre de recherche, je sélectionnerai 20 exemples représentatifs des traités de ces trois

pays afin d'obtenir une vision globale des améliorations possibles. Les informations sur les traités sont présentées dans le tableau ci-dessous :

Traité	Retention	Limit	Rein Tête.min Pays q		P_{grp}	SAR_{gr}	SAR_{in}	
1	1 000 000	24 000 000	2	2	FR 0,04	0,13	40 681,94	40 681,94
2	5 000 000	45 000 000	0	2	FR 0,16	0,25	8 638,20	91 732,52
3	2 215 736	38 000 000	1	3	FR 0,00	0,18	60 396,33	60 397,00
4	2 278 434	64 000 000	1	2	FR 0,05	1,00	47 331,98	47 331,98
5	3 200 000	60 400 000	1	2	FR 0,02	0,95	78 500,72	1 491 513,76
6	63 600 000	20 000 000	1	2	FR 0,02	0,95	78 500,72	1 491 513,76
7	3 000 000	60 600 000	1	2	FR 0,02	0,95	78 500,72	1 491 513,76
8	2 800 000	60 800 000	1	2	FR 0,02	0,95	78 500,72	1 491 513,76
9	2 600 000	61 000 000	1	2	FR 0,02	0,95	78 500,72	1 491 513,76
10	6 000 000	9 000 000	4	2	CN 0,27	0,16	320 152,43	518 804,08
11	15 000 000	15 000 000	3	2	CN 0,27	0,16	320 152,43	518 804,08
12	30 000 000	20 000 000	3	2	CN 0,27	0,16	320 152,43	518 804,08
13	50 000 000	50 000 000	2	2	CN 0,27	0,16	320 152,43	518 804,08
14	100 000 000	50 000 000	2	2	CN 0,27	0,16	320 152,43	518 804,08
15	150 000 000	100 000 000	2	2	CN 0,27	0,16	320 152,43	518 804,08
16	250 000 000	50 000 000	2	2	CN 0,27	0,16	320 152,43	518 804,08
17	35 000 000	2 500 000	4	1	CO 0,01	0,61	39 641 851,54	17 915 885,17
18	200 000 000	4 800 000 000	1	3	CO 0,01	0,61	39 641 851,54	17 915 885,17
19	5 000 000 000	7 000 000 000	1	3	CO 0,01	0,61	39 641 851,54	17 915 885,17
20	12 000 000 000	3 000 000 000	1	3	CO 0,01	0,61	39 641 851,54	17 915 885,17

En sortie du modèle on trouve :

Traité	Prime pure [Nouvel]	Prime pure [Ancien]	Différence
1	29 480	19 711	9 769
2	227 806	152 294	75 512
3	1 016	547	469
4	100 111	59 355	40 756
5	57 587	37 117	20 470
6	2 818	1 558	1 260
7	61 713	38 424	23 289
8	62 629	39 860	22 769
9	60 149	41 432	18 717
10	26 756 537	40 265 063	-13 508 466
11	28 325 500	33 091 736	-4 766 236
12	19 458 413	22 140 267	-2 681 854
13	23 512 745	-	23 512 745
14	12 621 429	12 809 995	-188 566
15	14 816 954	14 140 994	675 960
16	5 019 167	4 552 158	467 009
17	7 903 332	12 481 009	-4 577 677
18	15 432 955	43 200 626	-27 767 671
19	4 839 868	8 152 616	-3 313 533
20	1 103 096	1 931 858	-828 762

TABLE 5 – Comparaison des primes pures entre le nouveau et l’ancien modèle

Pour la France : En augmentant le seuil, on élimine les petits sinistres et on se focalise uniquement sur les pertes extrêmes. Cela signifie que la distribution est recalibrée uniquement à partir des sinistres les plus coûteux (l’impact sur le paramètre β), ce qui a pour conséquence une augmentation des valeurs moyennes et extrêmes.

La Chine est un cas particulier dans notre modèle car, contrairement à la France, elle ne bénéficie pas d’un regroupement avec d’autres pays. Cela signifie que ses paramètres sont estimés indépendamment, ce

qui peut entraîner des variations importantes par rapport au modèle initial.

- Cela peut être dû à l'augmentation forte du seuil, qui élimine de nombreux sinistres plus petits et ne conserve que les plus importants.
- La baisse du paramètre de Poisson λ indique une diminution significative de la fréquence des sinistres extrêmes pris en compte dans le modèle.
- L'augmentation de ξ montre que les sinistres extrêmes en Chine sont devenus encore plus significatifs, ce qui impacte fortement l'évaluation du risque.
- Et comme la France, on va prendre en compte les sinistres sont désormais bien plus coûteux qu'avant.

Globalement, on observe que lorsque la rétention augmente, le prix du traité diminue et l'écart entre le nouveau et l'ancien modèle se réduit, sauf pour le traité 13.

Pour la Colombie, le seuil augmentant légèrement par rapport à la Chine, on peut se rendre compte que globalement le prix diminue en raison de la baisse du niveau de couverture des traités.

Conclusion

Le modèle Peaks Over Threshold (POT) a été redéveloppé et mis en œuvre chez Arundo Re depuis 2018, en exploitant des données provenant de deux principales sources publiques : EMDAT et la Global Terrorism Database (GTD). Ce mémoire se concentre principalement sur l'adaptation du modèle POT afin de mieux appréhender les risques dans les pays partenaires d'Arundo Re, garantissant ainsi son applicabilité à diverses zones géographiques et profils de risque.

Pour donner un aperçu global, notre modèle POT fonctionne en estimant la fréquence et la gravité des événements catastrophiques. Ces deux caractères du modèle sont déterminés en appliquant sur des données aux excès dépassant un seuil prédéfini, un élément fondamental garantissant la cohérence asymptotique dans l'estimation des queues lourdes. Cependant, le choix d'un seuil approprié demeure un défi crucial, car il influence directement la fiabilité et la stabilité du modèle. Un seuil trop bas risque d'introduire du bruit excessif en capturant des événements non extrêmes, diluant ainsi la pertinence statistique du comportement des queues. À l'inverse, un seuil trop élevé réduit le nombre d'observations disponibles, entraînant une variance accrue dans l'estimation et une instabilité potentielle du modèle.

Notre travail vise à explorer des approches pour optimiser la détermination du seuil, en conciliant robustesse statistique et applicabilité pratique à la modélisation des catastrophes réelles. Au-delà de la sélection du seuil, un deuxième objectif majeur de notre étude est d'identifier des groupes de pays partageant des conditions similaires dans des domaines clés : les régimes climatiques, le niveau de développement scientifique et technologique, ainsi que les facteurs de risque géopolitiques, y compris leur exposition au terrorisme.

Ce mémoire met en évidence les limites de notre outil, soulignant la nécessité de mises à jour et d'améliorations continues dans la tarification. D'une part, nous avons proposé des tests analytiques pour choisir le seuil optimal de chaque pays, garantissant ainsi la stabilité des paramètres. D'autre part, nous avons étudié la possibilité de regrouper des pays selon leurs similarités climatiques, économiques ou géopolitiques, afin d'augmenter la robustesse de nos estimations. Cette mutualisation des données permet d'améliorer la représentativité statistique sans compromettre la cohérence du modèle.

Les résultats obtenus dans notre étude nous ont permis de mieux comprendre le marché de l'assurance,

en particulier les tendances de risque des marchés couverts. Ils nous offrent une vision globale du risque, facilitant ainsi l'ajustement et l'amélioration de l'outil de tarification.

Cependant, certaines limites persistent. Tout d'abord, en raison de la rareté des événements catastrophiques extrêmes, il est difficile d'en caractériser pleinement les propriétés statistiques. Ensuite, notre modèle repose sur l'hypothèse que tous les événements catastrophiques sont indépendants et identiquement distribués. Cette hypothèse constitue une simplification, car dans la réalité, les catastrophes présentent souvent des dépendances temporelles et spatiales. Cette hypothèse peut conduire à une sous-estimation du risque lorsqu'un événement augmente la probabilité d'un autre dans un délai proche. Par exemple, des répliques sismiques après un séisme majeur augmentent la probabilité de nouveaux tremblements de terre dans la même région. Une autre limite importante tient à l'exclusion des risques d'incapacité ou d'invalidité, pourtant pertinents en assurance vie. Cette omission s'explique par l'indisponibilité de données fiables dans les bases utilisées (EM-DAT et GTD), alors même que ces événements peuvent générer des charges financières considérables à long terme.

En conclusion, nous pouvons affirmer que le modèle POT constitue l'outil le plus optimal pour comprendre et gérer les risques auxquels le secteur de la réassurance doit faire face.

APPENDICE

Notion du modèle de tarification des traités XSC

Le modèle des dépassements au-dessus d'un seuil (Peak Over Threshold, POT) permet de modéliser la queue d'une distribution en s'intéressant uniquement aux valeurs extrêmes, c'est-à-dire les observations qui dépassent un certain seuil.

Voici les principales notations et paramètres utilisés dans notre cadre d'analyse :

- K : nombre d'évènements catastrophes pour une région incluant le pays spécifique dans un an
- K' : nombre d'évènements CAT causant au moins « m » morts dans un pays spécifique
- pr_{region} : proportion d'appartenance de la région d'intérêt dans un groupe de pays,
- m : seuil à partir duquel on considère qu'un événement est extrême,
- X : le nombre de victimes dans le k^{ime} scénario d'évènement catastrophe avec $k = 1; 2; \dots; K'$,
- Y : le nombre de victimes assurées,
- q : la part de marché,
- θ : l'indicateur représente de risque de concentration,
- Z : Le montant à charge de réassurance,
- P : la portée du traité définie par contrat
- M : le nombre minimal de têtes requis pour activer le traité
- R : la priorité/ la rétention du traité
- SAR : les capitaux sous risque
- AAL : (fois) le nombre de reconstitution
- L : la limité = $P*(1 + AAL)$
- C : l'expression du coût total

Appendice 1 : Théorie des Valeurs Extrêmes

La théorie des valeurs extrêmes est une branche de la statistique qui se concentre sur l'analyse des événements rares ou anormaux dans les distributions statistiques. En étudiant les extrémités des distributions, elle vise à comprendre et à anticiper les événements peu fréquents mais à fort impact. Dans le cadre de notre recherche, ces événements correspondent aux sinistres majeurs dans le secteur des assurances et de la réassurance. Son utilité est particulièrement notable dans ces domaines, où elle contribue à évaluer et à gérer les risques liés aux catastrophes.

Deux approches fondamentales de cette théorie méritent une attention particulière. La première repose sur *le théorème de Fisher-Tippett*, qui décrit la distribution asymptotique du maximum d'un ensemble de variables aléatoires. Contrairement au théorème central limite, qui conduit généralement à une distribution normale, cette approche aboutit à **la distribution de valeurs extrêmes généralisée (GEV)**. La seconde approche est liée à *la fonction des excès (ou "excess function")*, qui analyse les valeurs excédant un certain seuil et révèle la distribution de Pareto généralisée. Cette méthode est essentielle pour estimer la probabilité et l'impact des événements extrêmes, fournissant ainsi des données essentielles pour une gestion efficace des risques dans des contextes critiques.

Tout d'abord, on va développer le modèle qui se base sur la pierre angulaire de la théorie des valeurs extrêmes. Le modèle se concentre sur le comportement statistique de

$$\mathbf{M}_n = \max\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n\},$$

Où $X_1, X_2 \dots, X_n$, est une séquence de variables aléatoires indépendantes ayant une fonction de distribution commune F . Dans l'application, on considère ces variables aléatoires comme le nombre de victimes pour chaque événement catastrophique et M_n pourrait représenter le sinistre le plus important (en termes de coût) parmi n sinistres observés.

$$\begin{aligned}\Pr(M_n \leq z) &= \Pr(X_1 \leq z, \dots, X_n \leq z) \\ &= \Pr(X_1 \leq z) \dots \Pr(X_n \leq z) \\ &= [F(z)]^n\end{aligned}$$

Étant donné que la fonction de distribution F est inconnue, une méthode pour dériver M_n consiste à rechercher des modèles approximatifs pour F_n . Cette approche est analogue à l'approximation classique

de la distribution des moyennes d'échantillon par une distribution normale, comme l'explique le théorème central limite, un résultat fondamental en statistique.

Théorème de Fischer-Tippett (Le théorème de la limite des maxima) :

Supposons qu'il existe des constantes de normalisation $a_n \in R$ et $b_n > 0$, et une loi non dégénérée de fonction de répartition H telles que :

$$\frac{M_n - a_n}{b_n} \xrightarrow{(d)} H$$

Alors H est du même type qu'une des trois lois suivantes (données par leur fonction de répartition) :

- Loi de Fréchet : $\Phi_\alpha(x) = \exp(-x - \alpha)1_{\{x>0\}}$ $\alpha > 0$
- Loi de Weibull : $\Psi_\alpha(x) = \begin{cases} \exp(-x - \alpha) & \text{Si } x \leq 0 \\ 0 & \text{Sinon} \end{cases}$
- Loi de Gumbel : $\Lambda(x) = \exp(-\exp(-x))$

Ces trois lois sont en fait les trois cas particuliers de la distribution GEV (Generalized Extreme Value) qui s'écrit :

$$H_{\xi,\mu,\sigma}(x) = \begin{cases} \exp\left(-\left[1 - \frac{\xi(x-\mu)}{\sigma}\right]^{1/\xi}\right) & \text{si } \xi \neq 0 \\ \exp\left(-\exp\left[-\frac{x-\mu}{\sigma}\right]\right) & \text{si } \xi = 0 \end{cases}$$

Définition 1 : Max-domain d'attraction (MDA) S'il existe des constantes de normalisation $a_n \in \mathbb{R}$ et $b_n > 0$, et une loi non dégénérée GEV (ξ) telles que :

$$\frac{M_n - a_n}{b_n} \xrightarrow{(d)} H(\xi) \text{ sous loi (convergence en loi)}$$

Où GEV est la distribution des extrêmes généralisée. Alors, on dit que F_X appartient au max-domain d'attraction de GEV(ξ) (on écrira par la suite : $F_X \in MDA(H_\xi)$)

Remarque 1 : *Le Paramètre de Queue dans la Loi GEV* : Le paramètre ξ dans la loi des Valeurs Extrêmes Généralisées (GEV) est désigné comme le paramètre de queue. Il joue un rôle crucial dans la caractérisation de la distribution des extrémités de la distribution de $F(x)$:

- Si $F(x)$ appartient au MDA de $H(\xi)$ avec $\xi = 0$, cela indique que $F(x)$ est dans le domaine d'attraction de Gumbel. Dans ce cas, la distribution de X a une queue exponentielle ou modérément décroissante. Des exemples de telles distributions incluent la loi exponentielle, la loi normale et la loi log-normale.
- Si $F(x)$ appartient au MDA de $H(\xi)$ avec $\xi < 0$, $F(x)$ se trouve dans le domaine d'attraction de Weibull. Cela signifie que la distribution de X est bornée supérieurement, c'est-à-dire limitée sur le côté droit. La loi uniforme et la loi bêta sont des exemples de telles distributions.
- Si $F(x)$ appartient au MDA de $H(\xi)$ avec $\xi > 0$, cela signifie que $F(x)$ se trouve dans le domaine d'attraction de Fréchet. Dans ce cas, la distribution de X possède une queue lourde, caractérisée par des événements extrêmes avec des probabilités non négligeables même pour de très grandes valeurs. Les distributions typiques dans ce domaine incluent la loi de Pareto et la loi de Cauchy.

Le tableau ci-dessous représente quelques exemples de lois usuelles selon le domaine d'attraction auquel elles appartiennent :

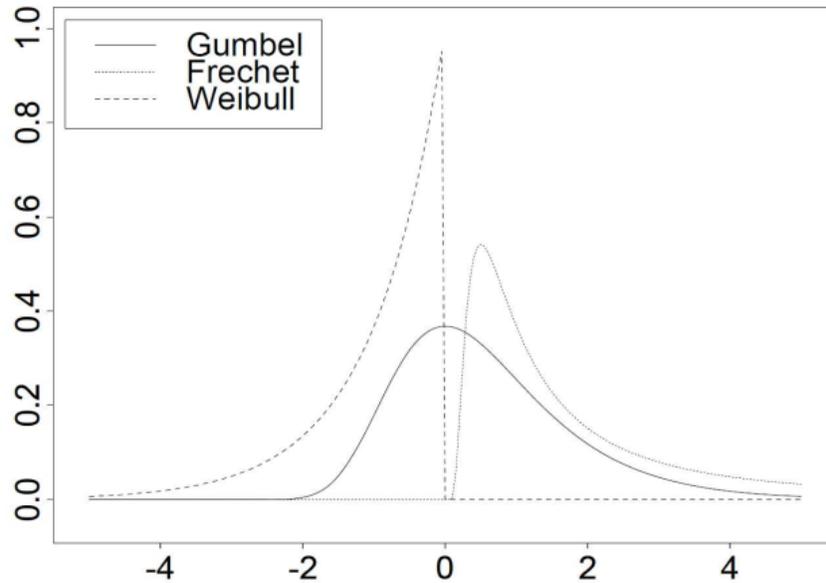


FIGURE 47 – Visualisation des densités des distributions de Gumbel, Fréchet et Weibull

Domaine d'attraction de Weibull	Domaine d'attraction de Gumbel	Domaine d'attraction de Fréchet
Uniforme	Exponentielle	Cauchy
Beta	Normale	Généralisée Pareto
Pareto Type II	Gamma	Student
	Log - Normale	

Distribution de Pareto généralisée

Théorème (Pickands, Balkema, De Hann) : Si F appartient à l'un des trois domaines d'attraction de la loi des valeurs extrêmes (DA(Fréchet), DA(Gumbel) ou DA(Weibull)), alors il existe une fonction $\sigma(u)$ positive, dénie à une équivalence près quand $u \nearrow x_F$, et un réel ξ tels que :

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

Où $G_{\xi, \sigma(x)}$ est la fonction de répartition de la loi de Pareto généralisée.

Corollaire : Von Mises condition : Soit F une fonction de densité absolument continue dont la densité f est satisfaisante :

$$\lim_{x \rightarrow \infty} \frac{x \cdot f(x)}{1 - F(x)} = \alpha > 0$$

Alors $F \in MDA(\Phi_\alpha)$

Preuve mathématique du corollaire : La distribution Pareto Généralisée est une MDA de la loi Fréchet

$Y \sim \text{GPD}(\xi, \sigma, \mu)$

ξ : la forme/le shape
 σ : l'échelle/le scale
 μ : la localisation/le seuil

La fonction de répartition :

$$f(x) = \begin{cases} (1 + \xi z)^{-\frac{\xi+1}{\xi}} & \text{si } \xi \neq 0 \\ e^{-z} & \text{si } \xi = 0 \end{cases}$$

La fonction de densité :

$$F(x) = \begin{cases} 1 - (1 + \xi z)^{-\frac{1}{\xi}} & \text{si } \xi \neq 0 \\ e^{-z} & \text{si } \xi = 0 \end{cases}$$

$$\lim_{x \rightarrow \infty} \frac{x \cdot f(x)}{1 - F(x)} = \lim_{x \rightarrow \infty} \frac{x \cdot (1 + \xi x)^{-\frac{\xi+1}{\xi}}}{(1 + \xi x)^{-\frac{1}{\xi}}} = \lim_{x \rightarrow \infty} \frac{x}{1 + \xi x} \stackrel{(L)}{\rightarrow} \lim_{x \rightarrow \infty} \frac{1}{\xi} = \frac{1}{\xi} > 0 \quad \text{avec } \xi > 0$$

La fonction de répartition des excès au-delà du seuil u est définie par :

$$\Pr[X > x] = \varphi_u[1 - \Pr(X < x | X > u)]$$

$x \geq 0, u \leq x_F$, où x_F est le point terminal droit de la distribution de X , $x_F = \sup\{x \in R : F(x) < 1\}$

Le modèle Poisson-GPD est un modèle statistique clé pour l'analyse des événements rares et extrêmes, tels que les catastrophes naturelles ou les pertes financières importantes, qui constituent le cœur de

notre travail. Il repose sur le processus ponctuel de Poisson (PPP), qui permet de décrire les événements extrêmes à l'aide de trois paramètres : la localisation μ , l'échelle σ , et la forme ξ .

Une caractéristique majeure de ce modèle est son indépendance par rapport au choix du seuil u . En effet, les propriétés fondamentales (telles que la forme de la distribution des excès) restent inchangées même lorsque le seuil u est ajusté. Cette flexibilité garantit la robustesse et la cohérence des résultats dans divers contextes analytiques. Cela simplifie considérablement les adaptations à des modèles présentant des effets non stationnaires ou des variations aléatoires.

Le taux d'excès ϕ_u est dérivé directement de ces paramètres. Une des propriétés importantes du GPD est sa stabilité de seuil : pour tout seuil $v > u$, les excès suivent une distribution GPD de même forme mais avec une échelle ajustée, donnée par $\phi_v = \phi_u + \xi(v - u)$. Cette propriété, souvent vérifiée à l'aide de la méthode *Mean Excess Loss*, permet d'évaluer et de choisir un seuil optimal.

Traditionnellement, une approche de seuil fixe était utilisée pour équilibrer le biais et la variance. Un seuil élevé améliore la précision de l'approximation du GPD en réduisant le biais. Cependant, il augmente la variance en raison d'une réduction de la taille de l'échantillon analysé. Le choix du seuil étant un élément crucial de la modélisation, nous y reviendrons plus en détail dans le chapitre suivant.

En effet, supposons que l'approximation de la distribution du maximum par une distribution GEV soit satisfaisante :

$$\Pr[M_n \leq x] \approx H(x)$$

Du théorème Pickands, Balkema, De Hann :

$$\begin{aligned} F^n(x) &\approx \exp \left[- \left(1 + \xi \frac{x - \mu}{\sigma} \right)^{-1/\xi} \right] \\ \Rightarrow n \ln[F(x)] &\approx - \left(1 + \xi \frac{x - \mu}{\sigma} \right)^{-1/\xi} \end{aligned}$$

Mais on a $\lim_{x \rightarrow 0} (1 + x) = \ln(x)$. Quand x grand, $F(x)$ est petite.

Donc, on obtient $1 - F(x) \approx -\ln[F(x)]$. Cela aboutit à une approximation simplifiée de $F(x)$:

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

Appendice 2 : Théorème Fischet Tippet :

Supposons qu'il existe des constantes de normalisation $a_n \in \mathbb{R}$ et $b_n > 0$, et une loi non dégénérée de fonction de répartition H telles que :

$$\frac{M_n - a_n}{b_n} \xrightarrow{(d)} H$$

Alors H est du même type qu'une des trois lois suivantes (données par leur fonction de répartition) :

- Loi de Fréchet : $\Phi_\alpha(x) = \exp(-x - \alpha)1_{\{x>0\}}$ $\alpha > 0$
- Loi de Weibull : $\Psi_\alpha(x) = \begin{cases} \exp(-x - \alpha) & \text{Si } x \leq 0 \\ 0 & \text{Sinon} \end{cases}$
- Loi de Gumbel : $\Lambda(x) = \exp(-\exp(-x))$

De la théorie de deux distributions de même type, on a :

1. $F^{[nt]}(a_{[nt]}x + b_{[nt]}) \rightarrow G(x)$

Supposons :

2. $F^{[nt]}(a_n x + b_n) = [F^n(a_n x + b_n)]^{\frac{[nt]}{n}} \rightarrow G^{[t]}(x)$

La convergence vers les théorèmes types s'applique alors nous sommes assurés de l'existence $\alpha(t) > 0, \beta(t) \in \mathbb{R}, t > 0$

1. $\lim_{x \rightarrow \infty} \frac{a_n}{a_{[nt]}} = \alpha(t)$
2. $\lim_{x \rightarrow \infty} \frac{b_n}{b_{[nt]}} = \beta(t)$
3. $G^{[t]}(x) = G[\alpha(t)x + \beta(t)]$

On est sûr que les fonctions $\alpha(t)$ et $\beta(t)$ sont mesurables.

Comme la fonction 3., si on a $t > 0$ et $s > 0$

$$G^{ts}(x) = G[\alpha(ts)x + \beta(ts)]$$

D'autre part,

$$\begin{aligned}G^{ts}(x) &= [G(\alpha(t)x + \beta(t))]^s \\ \Leftrightarrow G^{ts}(x) &= G(\alpha(s).[\alpha(t)x + \beta(t)] + \beta(s)) \\ \Leftrightarrow G^{ts}(x) &= G(\alpha(s).\alpha(t).x + \alpha(s).\beta(t) + \beta(s))\end{aligned}$$

Et puis, similaire, on a :

$$\begin{aligned}[1] \alpha(ts) &= \alpha(s).\alpha(t) \\ [2] \beta(ts) &= \alpha(s).\beta(t) + \beta(s) = \alpha(t).\beta(s) + \beta(t)\end{aligned}$$

où l'équation [1] est une forme d'équation fonctionnelle de Hamel. La forme simple [1] est $\alpha(t) = t^{-\theta}$.
Donc, [1] devient :

$$[1] \Rightarrow \alpha(ts) = (ts)^{-\theta} = t^{-\theta}.s^{-\theta}$$

On considère 3 cases : $\theta = 0$, $\theta > 0$ et $\theta < 0$

Le cas $\theta = 0$: $\alpha(t) = t^{-0} = 1$

L'équation [2] : $\beta(ts) = \beta(s) + \beta(t)$

L'équation fonctionnelle de Hamel avec la forme simple : $\beta(t) = -c * \log(t)$ avec $t > 0$

Donc :

$$\begin{aligned} G^t(x) &= G[\alpha(t)x + \beta(t)] \\ &= G[x - c.\log(t)] \end{aligned}$$

Si $c = 0$, l'équation $G^t(x) = G(x)$ implique que l'application répétée de la fonction ne modifie pas ses valeurs, quelle que soit la valeur de t . Cela peut se produire dans deux cas principaux :

1. $G(x) = 0$ ou $G(x) = 1$ partout,
2. $G(x)$ est une fonction dégénérée, où toute la probabilité est concentrée en un seul point.

On sait que $G^t(x)$ est une fonction non croissante en t (parce que $0 \leq G(x) \leq 1$). Pour conclure, $\Rightarrow c > 0$

- Substituer $x = 0$:

$$G^t(0) = G[-c * \log(t)]$$

- Mettre $\exp(-e^{-p}) = G(0) \in (0; 1)$
 - $u = -c * \log(t)$
- $\Rightarrow t = e^{-u.c^{-1}}$ avec $t \in (0; \infty)$ et $u \in (-\infty; +\infty)$

$$\begin{aligned} \Rightarrow G(u) &= G^t(0) = [\exp(-e^{-p})]^t = \exp(-e^{-p}t) \\ &= \exp(-e^{-c^{-1}u+p}) \\ &= \Lambda(c^{-1}u + p) \quad (\text{La distribution Gumbel}) \end{aligned}$$

Le cas $\theta > 0$:

$$\alpha(t) \cdot \beta(s) + \beta(t) = \alpha(s) \cdot \beta(t) + \beta(s)$$

$$\Rightarrow (\alpha(t) - 1)\beta(s) = (\alpha(s) - 1)\beta(t)$$

$$\Rightarrow \frac{\beta(s)}{1 - \alpha(s)} = \frac{\beta(t)}{1 - \alpha(t)} = c$$

$$\Rightarrow \beta(t) = \frac{\beta(s)}{1 - \alpha(s)} \cdot [1 - \alpha(t)]$$

$$\Rightarrow \beta(t) = c \cdot (1 - t^{-\theta})$$

$$\Rightarrow G^t(x) = G[t^{-\theta} \cdot x + c \cdot (1 - t^{-\theta})]$$

$$\Rightarrow G^t(x) = G[t^{-\theta}(x - c) + c]$$

$$\Rightarrow G^t(x + c) = G(t^{-\theta}x + c)$$

On établit : $H(x) = G(x + c)$ (les fonctions H et G sont du même type)

$$\Rightarrow H^t(x) = H[\alpha(t)x] = H(t^{-\theta}x)$$

• Si $x = 0$:

$$\rightarrow H(0) = H^t(0)$$

$$\Rightarrow \log H(0) = t \cdot \log(H[0])$$

$$\Rightarrow \begin{cases} \log H(0) = -\infty \Rightarrow H(0) = 0, \\ \log H(0) = 0 \Rightarrow H(0) = 1(\text{impossible}). \end{cases}$$

$$\Rightarrow H(0) = 0$$

• Considérer $H^t(x) = H[t^{-\theta}x]$ avec $x = 1$:

$$\Rightarrow H^t(1) = H(t^{-\theta})$$

$$\begin{cases} H(1) = 1 \Rightarrow H \equiv 1, \\ H(1) = 0 \Rightarrow H(0) = H(1) \end{cases}$$

Puisque H est une fonction non-dégénérée, donc, on a $H(1) \in (0;1)$

- Mettre $u = t^{-\theta} \Rightarrow t = u^{-\frac{1}{\theta}}$,

$$H(1) = c \exp(-p^{-\gamma}),$$

$$\gamma = \theta^{-1}.$$

$\Rightarrow H(u) = \exp(-p^{-\gamma}t) = \exp(-(pu)^{-\gamma})$ (la distribution Fréchet avec $x = 1$)

- Cas général pour $H(x)$

On suppose que la relation $H^t(x) = [H(x)]^t$ doit être satisfaite. Cela signifie que la fonction $H(x)$ doit vérifier l'équation suivante :

$$H(t^{-\theta} \cdot x) = [H(x)]^t.$$

Pour résoudre cette équation, appliquons le logarithme des deux côtés (valide car $H(x) > 0$) :

$$\log H(t^{-\theta} \cdot x) = t \cdot \log H(x).$$

Cela implique que $H(x)$ doit avoir une forme qui respecte la propriété d'échelle. Pour montrer que la solution satisfait l'équation $H(t^{-\theta} \cdot x) = [H(x)]^t$, on met $H(x) = \exp(-x^{-\alpha})$ dans cette équation :

- Calcul du côté gauche :

$$H(t^{-\theta} \cdot x) = \exp\left(-\left(t^{-\theta} \cdot x\right)^{-\alpha}\right) = \exp\left(-t^{\theta\alpha} \cdot x^{-\alpha}\right).$$

- Calcul du côté droit :

$$[H(x)]^t = [\exp(-x^{-\alpha})]^t = \exp(-t \cdot x^{-\alpha}).$$

- Égalité des deux côtés : Pour que $H(t^{-\theta} \cdot x) = [H(x)]^t$, il faut que :

$$t^{\theta\alpha} = t,$$

ce qui donne $\theta \cdot \alpha = 1$.

En remplaçant $\alpha = \frac{1}{\theta}$, la forme finale de $H(x)$ est :

$$H(x) = \begin{cases} 0, & \text{si } x < 0, \\ \exp(-x^{-\alpha}), & \text{si } x \geq 0. \end{cases}$$

Appendice 3 : Estimation de l'engagement moyen par tête

Nous prouvons la généralisation de la densité à x décès. On fait l'hypothèse qu'il existe une indépendance dans les décès.

Soient $X = (X_1, X_2, \dots, X_x)$ des variables aléatoires i.i.d. suivant une loi exponentielle de paramètre 1. On cherche à prouver que $\sum_{k=1}^x X_k$ suit une loi Gamma de paramètre $(x, 1)$. Nous allons utiliser la fonction génératrice des moments.

$$M_X(t) = \mathbb{E} [e^{tX}] = \mathbb{E} [e^{t(X_1+X_2+\dots+X_x)}] = \mathbb{E} \left[\prod_{k=1}^x e^{tX_k} \right]$$

Or les X_k , $k \in [1; x]$ sont i.i.d., donc :

$$= \prod_{k=1}^x \mathbb{E} [e^{tX_k}] = \prod_{k=1}^x M_{X_k}(t)$$

Or X_k suit une loi exponentielle de paramètre 1, donc :

$$M_{X_k}(t) = \frac{1}{1-t}$$

$$\begin{aligned} M_X(t) &= \prod_{k=1}^x \left(\frac{1}{1-t} \right) \\ &= \left(\frac{1}{1-t} \right)^x \end{aligned}$$

qui n'est autre que la fonction génératrice des moments d'une loi Gamma de paramètres $(x, 1)$.

REFERENCE

- [1] - Erland Ekheden et Ola Hössjer
Pricing catastrophe risk in life (re)insurance
- [2] - Guangrun Wu et Wenliang Qiu
Threshold Selection for POT Framework in the Extreme Vehicle Loads Analysis Based on Multiple Criteria.
- [3] - Stuart Coles
An Introduction to Statistical Modeling of Extreme Values.
- [4] - Emilie Deprey et Alexandre Godzinski
[Mémoire d'Actuariat] – Problématique de seuil dans la modélisation de la sinistralité en Réassurance Non-Vie
- [5] - Sidney I. Resnick
“Extreme Values, Regular Variation and Point Processus”.
- [6] - Assurance Audit Actuariat sous la direction de Jacques BLONDEAU et Christian PARTRAT
LA REASSURANCE - Approche technique
- [7] - Anne Pellerin
[Mémoire d'Actuariat] –Une modélisation du risqué sous-jacent aux traités CAT XL par évènement en reassurance vie
- [8] - Actualité de l'Assurance dans le monde
“Atlas Magazine N°217”.