

Brice IEMMI

Modélisation du risque catastrophe et impacts en prévoyance dans le cadre de Solvabilité II

➤ ETABLISSEMENT D'ACCUEIL

optimind ::

75, boulevard Haussmann

75008 PARIS

➤ MAITRE DE STAGE

Ariane de Taillandier, Actuaire consultante, Manager

ariane.detaillandier@optimind.fr

➤ PÉRIODE DE STAGE

Du 3 Mars 2008 au 30 Juillet 2008

Sommaire

REMERCIEMENTS	5
RESUME	7
SUMMARY	9
INTRODUCTION	11
CHAPITRE 1 PRÉSENTATION DU MÉMOIRE.....	13
I CARACTÉRISTIQUES GÉNÉRALES D’UN PORTEFEUILLE PRÉVOYANCE	13
1 Nature de la prévoyance	14
1.1 Distinction entre l’assurance vie et l’assurance non vie : dispositions réglementaires	14
1.2 Diversité des garanties en prévoyance	16
2 Les contrats « classiques » de prévoyance Individuelle et Collective	17
3 Les contrats d’assurance emprunteur	17
II LES RISQUES EN PRÉVOYANCE	18
1 Panorama global des risques en assurance	18
2 Les risques spécifiques en prévoyance	18
3 Le risque catastrophe en prévoyance	19
3.1 L’assurance Incendie Accident Risques Divers (IARD)	19
3.2 Impact du risque catastrophe sur la solvabilité et la valeur intrinsèque en prévoyance	20
III LE CAPITAL ECONOMIQUE ET LE RISQUE CATASTROPHE.....	22
1 Les garanties financières de bases : les provisions techniques.....	23
2 Le capital économique	23
2.1 Mesure cohérente de risque	24
2.2 Impacter le risque catastrophe sur le capital économique	26
3 Modélisation interne des risques catastrophes.....	27
3.1 Les modèles internes en IARD.....	27
3.2 Les possibilités d’adaptation en prévoyance	28
IV LA VALEUR INTRINSEQUE ET LE RISQUE CATASTROPHE.....	30
1 Principes	30
2 Les possibilités d’études d’impacts	31
V CHOIX D’UNE PROBLÉMATIQUE PRÉCISE.....	32
VI LE PROJET SOLVABILITE II ET LA PREVOYANCE	33
1 Le projet de directive en bref.....	33
2 La prise en compte du risque catastrophe.....	34

CHAPITRE 2 SELECTION DES CATASTROPHES ET SENSIBILITE DU PORTEFEUILLE	37
I	DÉFINITION ET TYPOLOGIE 37
II	SELECTION DES COUPLES PAYS/CATASTROPHE - SENSIBILITES 39
III	CHOIX : PRIORITÉS ET EXCLUSIONS..... 41
1	Les priorités..... 41
2	Les exclusions..... 43
3	Conclusion..... 44
CHAPITRE 3 MODÉLISATION INTERNE DU RISQUE CATASTROPHE.....	45
I	APPROCHE DÉTERMINISTE ET MODÉLISATION STOCHASTIQUE 45
1	L'approche déterministe 45
2	L'approche stochastique ou « fondée sur des scénarios » 46
2.1	Approche retenue pour le risque de séisme au Japon 47
2.2	Approche retenue pour les canicules en France 48
II	BUT DES MODÉLISATIONS 49
III	GENERER DES NOMBRES ET SIMULER EN ASSURANCE 49
IV	LE RISQUE DE SEISME AU JAPON 50
1	Eléments théoriques de modélisation 50
1.1	Loi de Zipf..... 51
1.2	Théorie des valeurs extrêmes 52
2	Lois de mortalité due aux séismes 65
2.1	Choix de l'échantillon..... 65
2.2	Modélisation des mortalités extrêmes par la loi de Zipf..... 66
2.3	Calibrage d'une loi de Pareto Généralisée 69
3	Loi de fréquence de survenance de séismes meurtriers 74
3.1	La loi binomiale 74
3.2	Ajustement d'un processus de Poisson sur le processus de survenance des séismes meurtriers au Japon.. 75
3.3	Simulation de la loi de fréquence sur 30 ans 77
4	Simulations : Intérêts et possibilités 77
V	LE RISQUE DE CANICULE EN FRANCE..... 79
1	Travaux préliminaires..... 79
1.1	Problématique 79
1.2	Définition 80
1.3	Caractère meurtrier d'une vague de chaleur 82
1.4	La surmortalité engendrée par les canicules n'est-elle qu'une illusion ?..... 83
2	Proposition d'une loi de surmortalité 84
2.1	Loi de surmortalité de 2003 84
2.2	Projection de la loi de surmortalité..... 84
3	Prévision d'événements caniculaires par les séries chronologiques (loi de fréquence)..... 85

3.1	Principes et étapes de la modélisation de la loi de fréquence.....	85
3.2	Choix des données	86
3.3	Calibrage d'un modèle ARIMA sur les historiques de températures journalières	88
CHAPITRE 4 APPLICATION SUR DES GARANTIES DECES D'UN PORTEFEUILLE PREVOYANCE.....		109
I	INTRODUCTION.....	109
II	PAYS CHOISIS – SENSIBILITÉ	109
1	France.....	109
2	Japon	110
III	PORTEFEUILLES ÉTUDIÉS	110
1	France : Portefeuille Prévoyance Collective.....	110
2	Japon : Portefeuille global périmètre vie	112
IV	PRECONISATION DU QIS4 SUR LE RISQUE CATASTROPHE	112
1	Prendre en compte la formule standard du module Santé	112
2	Le choix : retenir uniquement la formule standard pour les activités Vie	113
V	MODÉLISATION INTERNE : MÉTHODES ET RÉSULTATS.....	113
1	France.....	113
1.1	Simulation d'une canicule de plus de 5 jours sur la France en 2008.....	113
1.2	Ajustement du quantile à 99,5%.....	120
1.3	Le risque industriel : pire des cas répété aujourd'hui	124
1.4	Comparaison de la charge retenue avec la préconisation du QIS4.	127
2	Japon	129
2.1	Simulation du quantile à 99,5 de la loi d'intensité – des Worst Case –Comparaison	129
2.2	Constitution d'une chronique réaliste	130
2.3	Autres catastrophes	131
CONCLUSION		132
ANNEXES.....		134
ANNEXE 1 : TESTS STATISTIQUES		135
ANNEXE 2 : HISTORIQUE CENTENAL		138
ANNEXE 3 : DESCRIPTION DES GARANTIES NON VIES.....		139
ANNEXE 4 : TYPOLOGIE DES CATASTROPHES		140
ANNEXE 5 : GENERER DES NOMBRES ET SIMULER EN ASSURANCE		142
ANNEXE 6 : L'APPROCHE ANALYTIQUE POUR LES QUANTILES EXTREMES.....		146
BIBLIOGRAPHIE		148
LISTE DES ABRÉVIATIONS		150

Remerciements

Je remercie vivement la société Optimind et son président Christophe Eberlé pour m'avoir permis de réaliser mon stage de fin d'étude. Je remercie également l'ensemble des consultants d'Optimind et de Netanswer, pour leur accueil et leur professionnalisme.

J'adresse des remerciements particuliers à ma maître de stage Ariane De Taillandier, consultante et manager, qui a su cadrer mon travail et me guider dans l'élaboration de mon mémoire d'actuariat. Je remercie également Frédérique Henge pour son soutien, son aide et ses précieuses remarques.

Je tiens à exprimer mes meilleurs sentiments aux autres stagiaires, Blanche Pépin de Bonnerive, Eva Benros et Sarah Porel, avec lesquelles les échanges sur nos sujets de stages furent très instructifs.

Je remercie également les personnes qui de plus loin ont largement contribué à la réalisation de ce mémoire dans les meilleures conditions et plus globalement à la réalisation de mes études d'actuariat.

Brice IEMMI

Modélisation du risque catastrophe et impacts en prévoyance dans le cadre de Solvabilité II

Résumé

Le risque catastrophe est un risque dont la réalisation est rare et l'intensité potentiellement grande. Trois types de catastrophes peuvent être distingués : les catastrophes naturelles, les catastrophes technologiques et les catastrophes diverses (troubles sociaux et terrorisme).

Un portefeuille prévoyance est un portefeuille au carrefour entre l'assurance vie et l'assurance non-vie. Les garanties décès sont en effet réglementairement des clauses vie tandis que les autres garanties sont des clauses non vie. Ces dernières sont diverses et engagent l'assureur à payer une prestation à l'assuré en cas d'invalidité, d'incapacité, d'hospitalisation ou encore de chômage. Les prestations peuvent être des indemnités, des rentes, des remboursements de frais de santé ou encore des remboursements de prêts.

Le mémoire tente d'analyser l'impact du risque catastrophe sur un portefeuille prévoyance. Deux études d'impacts possibles auraient pu être envisagées. La première étude analyserait la sensibilité du capital économique et donc de la probabilité de ruine au risque catastrophe. La deuxième s'intéresserait aux impacts sur la valeur intrinsèque d'un portefeuille prévoyance.

Le mémoire propose une comparaison de coûts actualisés potentiels de garanties décès à l'exigence en marge de solvabilité pour le risque catastrophe demandé dans le QIS4.

Les pays sélectionnés pour cette étude d'impacts sont la France et le Japon. Les risques catastrophes faisant l'objet d'une modélisation stochastique sont le risque de canicule en France et le risque de séisme au Japon. D'autres catastrophes sont également impactées, mais d'une manière déterministe.

Les modélisations stochastiques consistent à déterminer des lois de fréquence de survenance et d'intensité en termes de coûts humains (nombres de tués potentiels).

Les modélisations déterministes consistent à répéter le pire des cas de l'historique centennal sur le portefeuille actuel (approche « as-if »).

Les études d'impacts permettent d'estimer des coûts potentiels pour chaque catastrophe. Les mesures de risques choisies sont la Value at Risk et la Tail Value at Risk (cette dernière mesure est utilisée uniquement sur le périmètre français). L'agrégation des risques est réalisée en utilisant la Tail Value at Risk car elle possède la propriété de cohérence (les risques ne sont pas agrégés pour le portefeuille japonais, ils sont simplement mesurés séparément).

Il s'avère, pour le périmètre français, que les 0,15% de capitaux sous risques du portefeuille préconisés par le CEIOPS pour absorber le risque catastrophe semblent excessifs.

Brice IEMMI

Catastrophe modeling : Impacts on a life insurance portfolio under Solvency II

Summary

The catastrophe risk, or disaster risk, is a risk whose achievement is rare and intensity potentially great. Three types of disasters can be distinguished: natural disasters, technological disasters and diverse disasters (social confusions and terrorism).

The studied portfolio is a portfolio at the crossroads between life insurance and non-life insurance. The death guarantees are legally life clauses while other guarantees are non-life terms. The latter guarantees are varied and engage the insurer to pay benefits to an insured in case of invalidated, disability, hospitalization or unemployment. Benefits may be compensations, pensions, health's expenses or loan repayment.

The report is trying to analyze the impact of disaster risks on the studied portfolio. Two possible impacts can be considered: the sensitivity of economic capital and thus the probability of ruin at the risk disaster. The report is limited to a comparison of potential updated costs to requirement solvency margin for disaster risk requested in the 4th quantitative impacts studies.

Countries selected for this impacts study are France and Japan. Disaster risks being a stochastic modeling are the risk of heat wave in France and the risk of earthquakes in Japan. Other disasters are also impacted, but in a deterministic way.

The stochastic models determine the laws of frequency of occurrence and intensity in terms of human costs (numbers of deaths). The deterministic models repeat the worst case on the current portfolio.

Impact studies estimate costs for each potential disaster. The risk measures selected are Value at Risk and Tail Value and Risk (the latter measure is used only on the french perimeter). The aggregation of risk is performed using the Tail Value at Risk because this risk measure is coherent (the risks are not aggregated for the Japanese portfolio, they are measured separately).

It turns out, for the chosen perimeter, that 1.5 for one thousand of capital at risk of the portfolio requested by the CEIOPS to absorb the disaster risk seems exaggerated.

Introduction

Selon le groupe de réassurance Munich Re, l'année 2007 a battu des records en termes de nombre et d'intensité des catastrophes. 950 catastrophes naturelles ont en effet été répertoriées. L'année 2008 semble décidée à prendre le relais. Les bilans provisoires du cyclone en Birmanie et du séisme meurtrier en Chine sont respectivement de 15 000 et 100 000 victimes (tués, blessés et déplacés).

Alors que les avis d'expert sont nuancés sur les évolutions récentes du nombre annuel et des intensités des catastrophes, le nombre de victimes annuels est quant à lui en hausse considérable depuis 20 à 30 ans.

Les facteurs aggravant pour les décennies à venir sont clairement identifiables et rendent très probable la possibilité d'une multiplication du nombre et de l'intensité des catastrophes naturelles.

D'autres catastrophes, non naturelles, pourraient également potentiellement survenir et engendrer des dégâts matériels et humains considérables. Le risque de terrorisme est en première ligne.

De leur côté, les assureurs constatent également une augmentation relativement récente de leur frais pour répondre aux catastrophes naturelles. Certaines analyses prospectives prévoient une aggravation pour les décennies à venir et particulièrement en Indonésie, en Chine et en Inde.

La question de savoir comment les assureurs pourraient financièrement être mis à mal par les risques catastrophes dans les années à venir est donc posée. Cette interrogation vaut pour l'industrie de l'assurance toute entière (assureurs, réassureurs, courtiers etc.). Plus précisément, le problème qui se pose est de savoir si la richesse et la solvabilité d'un assureur, qui constituent deux indicateurs phares de la performance et de la santé financière, peuvent être considérablement impactés par une aggravation du risque catastrophe dans les prochaines décennies.

Le mémoire tente de cerner le risque catastrophe en trouvant une définition et une typologie qui puissent faire un certain consensus. Il propose deux modélisations stochastiques de deux catastrophes auxquels le portefeuille étudié semble particulièrement sensible. Il essaie enfin de mener une étude d'impacts pragmatique, en estimant des coûts actualisés potentiels sur un périmètre bien précis du portefeuille.

Le mémoire se décompose en 4 chapitres.

Le premier chapitre présente les caractéristiques générales du portefeuille étudié puis les risques inhérents à l'activité de prévoyance. Il donne ensuite deux possibilités d'études d'impacts. Il débouche enfin sur un choix précis de problématique et sur une présentation du projet Solvabilité II ainsi que les volets de la 4^{ème} étude quantitative d'impacts sur le risque catastrophe.

Le deuxième chapitre fixe le vocabulaire en rapport avec la notion de catastrophe. Il tente de donner une définition ainsi qu'une typologie précise. Il donne également des choix de couples pays/catastrophe qui, au regard de la présence de la compagnie d'assurance étudiée, constituent une priorité de traitement.

Le troisième chapitre propose des possibilités de modélisation du risque catastrophe en prévoyance. Deux modélisations stochastiques sont réalisées. Elles consistent à estimer des lois de fréquence (probabilités de survenance) et d'intensité (coût humain potentiel) qui permettront, si un degré de prudence est fixé à priori, de mesurer le risque modélisé.

Le quatrième et dernier chapitre donne une méthodologie pour passer du coût humain potentiel d'une catastrophe à un coût financier potentiel actualisé. Cette quantité sera ensuite comparée aux capitaux sous risques du portefeuille et à l'exigence liée au risque catastrophe telle qu'elle est établie par le CEIOPS (Committee of European Insurance and Occupational Pensions Supervisors). Cette partie tente de juger de l'utilité d'une modélisation fine (ou modélisation interne) des risques catastrophes auxquels un portefeuille est soumis, dans le cadre des exigences futures de marges de solvabilité.

Chapitre 1 Présentation du mémoire

Ce chapitre a pour but d'introduire le thème du mémoire et les possibilités d'études d'impacts. Il débouche sur le choix de problématique qui a été fait. Il présente également le cadre réglementaire, le portefeuille et les risques.

Il se concentre dans un premier temps sur les aspects généraux du portefeuille géré et dans un deuxième temps sur des risques auxquels le portefeuille est exposé. La troisième partie introduit le thème de la solvabilité et deux méthodes utilisées par une compagnie d'assurance pour gérer les pertes potentielles pouvant être engendrées par la réalisation des risques. La quatrième partie propose un autre indicateur phare d'un portefeuille qui peut être potentiellement impacté par le risque catastrophe : la valeur intrinsèque. Le cinquième paragraphe donne un choix de problématiques qui réduit le portefeuille à disposition et délimite clairement l'étude d'impacts qui sera réalisée.

Le dernier paragraphe se concentre sur le projet de directive Solvabilité II. La future réglementation va instaurer une vision novatrice du risque, dans leur évaluation comme dans leur gestion. Elle prévoit également un volet sur l'absorption des risques catastrophes.

I Caractéristiques générales d'un portefeuille prévoyance

Il convient de définir clairement la nature d'un contrat prévoyance afin d'en cerner la spécificité et de justifier la possibilité de segmentation opérée dans le chapitre 4. Il est également important de mentionner les deux périmètres présents dans le portefeuille :

- Contrats de prévoyance classique Individuels et Collectifs.
- Contrats d'assurance emprunteur (« AdE »).

Ces contrats se distinguent par les types de prestations, les aléas « déclencheurs » de prestations étant les mêmes.

1 Nature de la prévoyance

Un contrat d'assurance prévoyance est un contrat au carrefour de deux types d'assurance : l'assurance vie et l'assurance non-vie. En effet, un contrat de prévoyance regroupe des clauses ou garanties de type vie et des garanties de type non-vie. Plus précisément, certaines clauses prévoient le paiement (immédiat, différé ou étalé dans le temps etc.) d'un capital en cas de décès de l'assuré tandis que d'autres prévoient une prestation si un aléa qui n'est pas basé sur la durée de vie humaine se réalise.

1.1 Distinction entre l'assurance vie et l'assurance non vie : dispositions réglementaires

La réglementation actuelle (Solvabilité I) repose sur les directives suivantes, distinguant l'assurance vie de l'assurance non vie :

- Directive 73/239/CEE du Conseil Européen, du 24 juillet 1973, portant coordination des dispositions législatives, réglementaires et administratives concernant l'accès à l'activité de l'assurance directe autre que l'assurance sur la vie et son exercice.
- Directive 2002/83/CE du Parlement européen et du Conseil Européen du 5 novembre 2002 concernant l'assurance directe sur la vie.

1.1.1 Assurance vie

Un contrat d'assurance vie se définit comme un contrat dont l'aléa est basé sur la durée de vie humaine :

- La survie : assurance en cas de vie.
- Le décès : assurance en cas de décès

Des assurances mixtes sont souvent envisagées dans la pratique.

L'assurance vie prend également en compte les contrats d'assurance basés sur la durée de vie humaine et contenant des clauses supplémentaires mineures du type :

- L'incapacité temporaire de travail (IT).
- L'invalidité due à une maladie ou un accident.

1.1.2 Assurance non vie

L'assurance non vie correspond à tous les contrats qui ne sont ni des contrats d'assurance vie, ni des contrats de capitalisation.

L'assurance non vie est classée par branches d'activités. L'annexe 2 du mémoire donne la classification détaillée telle qu'elle figure dans la directive de 1973 (retraitée depuis). Le CEIOPS, dans la quatrième étude quantitative d'impacts (voir la section 4 de ce chapitre) donne une segmentation simplifiée :

1. Accident et santé : indemnité aux travailleurs
2. Accident et santé (assurance santé)
3. Accident et santé (autres)
4. Responsabilité automobile, tiers
5. Automobile, autres classes
6. Marine, aviation et transport
7. Incendie et autres dommages immobiliers
8. Responsabilité au tiers
9. Crédit et assurance caution
10. Dépenses juridiques
11. Assistance
12. Assurance non vie divers

1.2 Diversité des garanties en prévoyance

Les garanties non-vie d'un contrat d'assurance prévoyance sont multiples et deux options peuvent être envisagées pour effectuer une segmentation des clauses en clauses en vie et en non-vie.

Options n°1: La définition européenne de l'activité-vie permet l'approximation suivante : un contrat d'assurance prévoyance est un contrat d'assurance vie.

Options n°2: Une partie du portefeuille est un périmètre d'assurance vie (clause décès). Les autres clauses font partie des branches non vie et plus précisément des branches 1, 2 et 3 (Accidents et Santé)

L'option n°2 sous-entend une tarification distincte des garanties et un système de gestion adapté à cette distinction.

L'option n°1 sous-entend à l'inverse une tarification commune. Il se pose alors le problème de projection des prestations futures probables. Tandis que la projection des prestations dues au décès est envisageable par l'intermédiaire de tables de mortalité, la projection de flux liés au chômage de l'assuré est beaucoup difficile car les tables de maintien (en invalidité, chômage etc.) sont plus complexes à réaliser.

Il est possible de distinguer des clauses usuelles en prévoyance :

- Décès
- Perte Totale et Irréversible d'Autonomie (PTIA)
- Invalidité Permanente Totale (IPT)
- Invalidité Permanente Partielle (IPP)
- Incapacité temporaire de travail (ITT)

2 Les contrats « classiques » de prévoyance Individuelle et Collective

Les contrats classiques garantissent le versement d'une prestation en cas de réalisation d'un aléa.

En prévoyance classique, les prestations sont de natures diverses :

- Versement d'un capital décès (indemnité non étalée)
- Rente de conjoint
- Rente d'éducation
- Frais professionnel
- Frais de santé

Les contrats peuvent se conclure entre un individu et une société d'assurance. Il s'agit alors de contrats de prévoyance Individuelle. Ils peuvent se conclure à un niveau collectif et c'est une entreprise qui souscrit pour l'ensemble de ses salariés.

3 Les contrats d'assurance emprunteur

Le marché de l'assurance emprunteur est un marché qui arrive actuellement à maturité, compte tenu des établissements de crédits qui exigent aujourd'hui quasi systématiquement la souscription d'une assurance, pour autoriser l'emprunt.

Les aléas déclencheurs sont classiques. Ils sont énoncés plus haut.

Les prestations sont par contre particulières et la projection (dans le cadre de la tarification par exemple) des prestations futures probables passent par l'établissement d'un tableau d'amortissement du prêt.

De manière usuelle, les prestations sont les suivantes :

- En cas de décès et/ou PTIA et/ou IPT, paiement du capital restant dû, en fonction de la part assurée
- En cas d'IPP et/ou d'ITT et/ou perte d'emploi (chômage), prise en charge des échéances du prêt par l'assureur, en fonction de la part assurée et du taux de remboursement contractuel

II Les risques en prévoyance

1 *Panorama global des risques en assurance*

Le CEIOPS donne, à travers le projet de directive Solvabilité II, un panorama complet des risques. Ils sont clairement distinguables par classes mais sont toutefois plus ou moins corrélés entre eux.

D'autres classifications des risques sont envisageables. L'étude de 2004 réalisée par l'Association Internationale des Actuaires constitue une référence¹.

La segmentation du CEIOPS est la suivante :

- Risque de marché
- Risque de contrepartie
- Risque de souscription
- Risque opérationnel

2 *Les risques spécifiques en prévoyance*

L'activité d'assurance prévoyance, étant au carrefour de l'assurance vie et de l'assurance non vie, a un risque de souscription spécifique car il est double.

Le risque provient de la garantie décès qui est une garantie vie et des autres garanties (invalidité, incapacité, chômage etc.) qui sont des garanties non-vie. La prévoyance est donc à la fois soumise au risque de décès et de survie mais aussi principalement aux risques d'incapacité/invalidité et de chômage.

Il n'y a pas de risque spécifique propre à la prévoyance, mais plutôt une exposition multiple. Cette caractéristique a un impact direct sur le provisionnement, qui est considérablement complexifié par rapport à une activité d'assurance vie stricte (voir la partie III).

¹ A Global Framework for Insurer Solvency Assessment, International Actuarial Association (2004)

3 *Le risque catastrophe en prévoyance*

Le CEIOPS définit le risque catastrophe comme un événement extrême, rare et irrégulier. Ce terme est clairement défini dans le chapitre 2. Ce paragraphe tente d'isoler la problématique de la gestion du risque catastrophe en prévoyance afin d'en cerner des enjeux.

Il fait d'abord une parenthèse sur le cas de l'assurance IARD avant de justifier la nécessité de traiter, à court terme comme à plus long terme, le risque catastrophe en prévoyance.

3.1 L'assurance Incendie Accident Risques Divers (IARD)

Le risque catastrophe est un risque qui occupe traditionnellement le secteur de l'assurance de biens (assurance IARD). En effet, depuis le premier grand sinistre de l'ère moderne, le tremblement de terre de 1906, qui détruisit San Francisco, les assureurs et réassureurs redoutent de se retrouver dans cette situation extrême qui conduisit bon nombre d'entre eux aux limites de leur capacité.

A l'époque, environ 130 compagnies d'assurance étaient présentes aux Etats-Unis, dont 50 étrangères. Lors de l'année 1906, elles avaient encaissé 2,6 millions de dollars de primes et durent faire face à 400 millions d'euros de dégâts. L'équivalent du montant de primes des 47 dernières années avaient été engloutis.

Actuellement, si un séisme de cette ampleur survenait de nouveau, les experts estiment le coût pour l'industrie de l'assurance à 40 milliards de dollars, ce qui est équivalent à la charge du cyclone Katrina qui a balayé la Nouvelle Orléans en 2005. Selon ces mêmes experts, aucune difficulté majeure n'a été constatée chez les assureurs lorsque le cyclone survint. Cela prouve que les assureurs ont tiré les enseignements durant le 20^{ème} siècle et peuvent à court terme continuer à absorber les événements catastrophiques. En effet, les risques sont mieux modélisés (voir partie III), les exclusions de garanties plus fréquentes et des plafonds de couverture mis en place.

Cette capacité actuelle à long terme est sévèrement remise en question. Les deux facteurs suivants risquent d'aggraver considérablement les dommages matériels dans les années à venir.

- L'accumulation vertigineuse des richesses dans les zones à risques
- Le changement climatique qui pourra potentiellement décupler la puissance des catastrophes naturelles, telles que les cyclones et les inondations, dans certaines régions du monde, dans les décennies à venir.

Les progrès dans les modélisations et les plafonds de couverture risquent de devenir insuffisant pour maîtriser ces risques grandissants. De nouveaux modèles plus sophistiqués et encore plus proches de la dynamique des catastrophes devront être probablement inventés. Les risques devront être également mieux et encore plus partagés (usage de la titrisation).

3.2 Impact du risque catastrophe sur la solvabilité et la valeur intrinsèque en prévoyance

La question de savoir comment les indicateurs phares d'une compagnie d'assurance vie peuvent être touchés (capital économique, valeur intrinsèque) est moins fréquente qu'en assurance IARD. Cela vient du fait que les portefeuilles d'assurance vie sont moins sensibles aux risques catastrophes. Il est vraisemblable qu'un assureur vie, lorsqu'il estime le coût potentiel d'une catastrophe, se rend compte de l'insensibilité de ses portefeuilles si les pires catastrophes surviennent de nouveau en même temps.

Toutefois, cette conclusion risque néanmoins de ne plus se vérifier à l'avenir, principalement du fait de l'aggravation des risques dus à :

- L'accumulation vertigineuse des personnes dans les zones à risques, notamment dans les pays en voie de développement
- Le développement anarchique de ses zones, que ce soit au niveau des habitations ou des infrastructures (transports, canalisations etc.)
- Le changement climatique qui pourra potentiellement décupler la puissance des catastrophes naturelles, telles que les cyclones et les inondations, dans certaines régions du monde, dans les décennies à venir.

Il est fort probable qu'un assureur qui se développe dans ces zones et qui souhaitent y développer des activités de prévoyance s'expose fortement à ces risques grandissants. Les exemples suivants permettent d'illustrer de manière plus concrète les expositions futures probables.

- Les séismes en Asie pourraient devenir de plus en plus meurtriers. Le récent séisme en Chine en est une preuve.
- Les pandémies pourraient se développer, du fait de la résistance croissante de certaines bactéries, de la déforestation et du développement anarchique de certaines métropoles, notamment en Inde. L'OCDE, dans « Les risques émergents du 21^{ème} siècle » soutient cette thèse.
- Les canicules pourraient faire un nombre de victimes considérables dans les années à venir du fait de l'augmentation de l'espérance de vie, dans les pays développés.
- Le risque de terrorisme, risque par ailleurs émergent, pourrait engendrer des coûts considérables pour des activités prévoyance, si par exemple un attentat du type World Trade Center survenait à nouveau dans le monde.

Au regard des analyses pessimistes sur l'émergence de ces nouveaux risques, les compagnies possédant des portefeuilles prévoyance risquent très probablement de faire face à l'avenir à des coûts exceptionnels de plus en plus fréquents et de plus en plus importants, que ce soit pour les pays en voie de développement ou dans les pays développés.

Les impacts se situeraient entre autre à deux niveaux :

- A niveau de la richesse disponible aux actionnaires : des scénarios catastrophes réalistes permettraient d'évaluer la valeur actuelle probable de cette richesse, diminuée du coût actuel probable engendré par les catastrophes.
- Au niveau de la solvabilité : le capital économique, s'il intègre des scénarios catastrophes, peut s'accroître potentiellement. Le capital cible pour le risque catastrophe mériterait donc d'être évalué.

La prospection et la modélisation de ces risques semblent indispensables. Il est également fort possible que ces efforts deviennent rapidement urgents.

La partie qui suit introduit les techniques actuarielles usuelles pour mesurer et absorber les risques, d'abord d'un point de vue global en intégrant les risques à faible sévérité, puis états actuels d'avancement des modèles internes de gestion du risque catastrophe en assurance IARD et termine sur une introduction des possibilités de modélisation envisagées en prévoyance.

III Le capital économique et le risque catastrophe

Le premier paragraphe présente brièvement les principales provisions techniques constituées en prévoyance. Ce sont des garanties financière des bases qui reconnaissent une dette future moyenne envers les assurés.

Les provisions techniques ne sont pas suffisantes car elles couvrent contre une perte moyenne et ne peuvent donc pas absorber des aggravations de risque. Dans le cas d'une assurance décès, une Provision Mathématique (PM) sera par exemple insuffisante si la mortalité s'aggrave, pour une raison ou pour une autre.

Il est donc nécessaire de reconnaître des pertes potentielles à un niveau de prudence élevée. C'est le but du capital économique dont les principes de calcul sont donnés dans un second paragraphe.

Si le niveau de prudence requis est élevé, ce qui sera le cas dans le futur cadre réglementaire, les pertes potentielles très exceptionnelles devront être évaluées. Ces pertes sont des pertes rares et sont donc généralement due à des catastrophes. Le capital cible qui mesure la valeur actuelle de cette perte potentielle peut par ailleurs alourdir considérablement le capital économique. Le troisième paragraphe introduit les possibilités de modélisation de ce risque particulier. Ces modélisations sont l'objet du chapitre 3. Elles seront utilisées pour évaluer des capitaux cibles dans le chapitre 4.

1 Les garanties financières de bases : les provisions techniques

Les principales provisions techniques en prévoyance sont la provision mathématique et la provision pour sinistres à payer.

Tandis que la première est usuellement utilisée en vie, l'autre est généralement constituée pour les activités d'assurance IARD.

- La provision pour risques croissants (PRC) est la différence entre la valeur actuelle probable de l'engagement de l'assureur et la valeur actuelle probable de l'engagement de l'assuré. Elle correspond à la provision mathématique en assurance vie. Elle prend le nom de provision pour risques croissants en prévoyance car les risques assurés (invalidité, morbidité) sont en général croissants dans le temps.
- La provision pour sinistres à payer (PSAP) est la « *valeur estimative des dépenses en principal et en frais, tant internes qu'externes, nécessaires au règlement de tous les sinistres survenus et non payés, y compris les capitaux constitutifs des rentes non encore mises à la charge de l'entreprise* » d'après le Code des Assurances.

2 Le capital économique

La compagnie doit détenir suffisamment de fonds propres pour assurer ses engagements quand les provisions établies de manière prudentes deviennent insuffisantes à la suite d'une sinistralité exceptionnelle.

Selon la SOA (Society of Actuaries), le capital économique se définit comme « *le montant de capital suffisant pour couvrir les pertes potentielles à un niveau de tolérance au risque donné et pour un horizon de temps spécifié* ».

Toutefois, plus le capital est élevé, plus la rentabilité sur fonds propres est faible, ce qui nuit aux intérêts des actionnaires. Ces derniers ont en effet pour objectif de rentabiliser l'investissement qu'ils ont effectué en achetant des parts de la société.

Le capital économique tient compte à la fois :

- du point de vue des assurés et de la réglementation : le capital économique désigne un niveau de fonds propres minimal garantissant la continuité de l'activité dans des conditions saines
- du point de vue des actionnaires : le capital économique indique le besoin en capital suffisant par rapport aux risques que les investisseurs sont prêts à prendre pour atteindre une certaine rentabilité.

En pratique, le capital économique peut s'analyser, si l'on dispose d'une chronique de valeurs actuelles des profits futurs issue de simulations stochastiques, comme un quantile à un ordre spécifié à priori (degré de prudence) de cette distribution empirique du résultat. La chronique à disposition est issue de simulations stochastiques où les paramètres clefs d'une activité d'assurance sont variables (taux d'intérêts, rentabilité des actifs risqués, mortalité, rachats etc.).

En toute rigueur le quantile qui représente le capital économique (une perte potentielle à un niveau de prudence) est un quantile d'ordre en général élevé de la variable aléatoire $X = -$ valeur actuelle des profits futurs.

Ce capital économique permet de se prémunir contre des dérives de sinistralité, des baisses des actifs etc.

Si le degré de prudence fixé à priori est élevé, il va de soit que les scénarios envisagés par les modélisations stochastiques devront obligatoirement prendre en compte les événements rares et potentiellement coûteux.

Une possibilité pour impacter le risque catastrophe sur le capital économique est proposé dans le paragraphe 2.2 qui suit. Mais avant cela, la notion de mesure cohérente de risque est présentée.

2.1 Mesure cohérente de risque

Ce paragraphe a pour but de mettre en évidence l'opposition entre la TVaR et la VaR. Tandis que la première est toujours cohérente, la deuxième ne l'est pas sauf dans des cas particuliers.

Soit un risque modélisé par une fonction X qui à tout état de la nature ω associe le réel $X(\omega)$. Nous appelons **mesure de risque** toute application ρ associant au risque X un réel positif $\rho(X)$.

D'après Artzner, Delbaen, Eber et Heath (1998), une mesure de risque est dite cohérente si elle respecte les 4 axiomes suivants :

- **Invariance par translation** : $\rho(X + c) = \rho(X) + c$, pour toute constante c ;
- **Sous-additivité** : $\rho(X + Y) \leq \rho(X) + \rho(Y)$, quels que soient les risques X et Y ;
- **Homogénéité positive** : $\rho(cX) = c\rho(X)$, pour toute constante positive c ;
- **Monotonie** : $X \leq Y \Rightarrow \rho(X) \leq \rho(Y)$, quels que soient les risques X et Y .

Trois mesures de risques sont répandues :

- La Value at Risk
- La Tail Value at Risk
- La Conditionnal Tail Expectation

Les définitions et propriétés sont les suivantes :

- $CTE(X, \alpha) = E[X | X \geq VaR(X, \alpha)]$

Si X désigne une perte, alors la CTE au seuil α désigne la perte moyenne sachant que cette perte dépasse la VaR d'ordre α .

- $VaR(X, \alpha) = \inf \{x \in \mathbb{R}, P(X \leq x) > \alpha\}$

Il existe également une caractérisation : $VaR(X, \alpha) = F_X^{-1}(\alpha)$ où F_X^{-1} désigne la fonction quantile. Cette approximation est une caractérisation dans le cas où F est continue. L'approximation est faite dans le cas des distributions empiriques en escalier.

- La TVaR est assimilée à la CTE, dans le cas d'une distribution continue. L'approximation est également faite pour les distributions empiriques en escalier

Le caractère cohérent de la TVaR (ou CTE) permet à cette mesure de risque d'être sous additive. Une hypothèse supplémentaire d'indépendance entre X et Y ne suffit pas pour assurer

$$\text{VaR}(X+Y, \alpha) \leq \text{VaR}(X, \alpha) + \text{VaR}(Y, \alpha)$$

2.2 Impacter le risque catastrophe sur le capital économique

La méthode pour impacter le capital économique pourrait être de **traiter les risque catastrophes à part pour estimer une distribution du coût actualisé du risque catastrophe**. Le problème de cohérence de la mesure de risque se pose à deux niveaux :

1. Au niveau de l'agrégation des risques catastrophes. Si X désigne le coût actualisé de la catastrophe n°1 et Y le coût actualisé de la catastrophe n°2, alors la VaR (en pratique la fonction quantile), n'étant pas une mesure de risque cohérente, aucune proposition ne peut être faite immédiatement sur $\text{VaR}(X+Y, \alpha)$, par rapport à $\text{VaR}(X, \alpha) + \text{VaR}(Y, \alpha)$. A l'inverse, la sous-additivité de la TVaR permet pour n risques catastrophiques de poser :

$$\text{TVaR}(X_1 + \dots + X_n, \alpha) \leq \sum_{i=1}^n \text{TVaR}(X_i, \alpha)$$

2. Au niveau de l'agrégation des risques catastrophes avec la valeur actuelle des profits futurs (sans scénarios catastrophes), notée VAPF. La sous-additivité de la TVaR permet d'écrire directement

$$\text{TVaR}(X_1 + \dots + X_n - \text{VAPF}, \alpha) \leq \sum_{i=1}^n \text{TVaR}(X_i, \alpha) + \text{TVaR}(-\text{VAPF}, \alpha)$$

Alors que la fonction quantile de la variable $X_1 + \dots + X_n - \text{VAPF}$ ne peut pas être aisément comparée à la somme des fonctions quantiles (à un ordre α spécifié), la TVaR de cette même variable est directement comparable à la somme des TVaR. Cette mesure de risque conviendrait mieux pour éviter des approximations, dans les applications pratiques, lors d'agrégation de risques.

3 *Modélisation interne des risques catastrophes*

Le paragraphe précédent met clairement en évidence des possibilités pratiques d'études d'impacts du risque catastrophe sur le capital économique.

Le défi consiste à estimer une distribution empirique du coût actualisé des risques catastrophes. En assurance IARD, les modèles le permettant sont très répandus et parfois d'une complexité remarquable, nécessitant des moyens techniques de simulations considérables.

Le paragraphe suivant donne une idée des possibilités de modélisation. Le deuxième paragraphe donne des possibilités d'adaptation pour un portefeuille prévoyance.

Comme le but de ce premier chapitre consiste avant tout à définir une problématique précise sur le risque catastrophe (modélisation, étude d'impacts), il semble indispensable, avant de proposer une problématique, d'avoir une idée générale des possibilités de modélisation.

Des compléments sur les modèles de risques catastrophes sont donnés au début du chapitre 3.

3.1 Les modèles internes en IARD

Des agences sont aujourd'hui spécialisées dans ce genre de modélisation.

- AIR
- EQECAT
- RMS

Les catastrophes modélisées sont généralement des catastrophes naturelles : tempêtes tropicales et extratropicales, séismes, Inondations etc. C'est ce type de catastrophes qui est le plus coûteux et qui devrait le rester dans les prochaines décennies (voir chapitre 2 pour une typologie complète).

Les besoins de modélisation se sont accrus à la fin des années 80, à la suite de tempêtes particulièrement coûteuses pour l'industrie de l'assurance. Ces besoins ont pu être progressivement satisfaits grâce au développement des moyens informatiques.

Un modèle de catastrophe est un système expert permettant de quantifier les dommages matériels et les pertes financières dues à un péril spécifique. Ce système reproduit des sinistres réels dans un environnement informatique.

Ces systèmes croisent deux types d'informations pour en déduire des courbes de vulnérabilité.

Les informations nécessaires sont :

1. **Une connaissance des valeurs assurées.** Elles doivent être précisément situées par l'intermédiaire d'un géocodage. Elles doivent être caractérisées au mieux : type de construction, solidité, hauteur, existence d'une franchise etc.
2. **Une connaissance du péril.** Les possibilités de réalisation de l'aléa naturel doivent être connues. Chaque paramètre péril doit être caractérisé au mieux.

Dans le cas d'un cyclone il convient idéalement de connaître la fréquence, l'intensité potentielle des vents suivant la localisation par rapport au centre de la catastrophe et la topographie du terrain, les trajectoires possibles etc.

Le croisement des données propres au péril et des données propres aux valeurs assurées permettent de prévoir des scénarios probables d'endommagement et donc de coût financier. Ces croisements sont résumés par des courbes de vulnérabilité qui placent un endommagement en abscisse et une caractéristique du péril en abscisse. Ces courbes constituent une base sérieuse de connaissance du risque catastrophe étudié.

Ces modèles présentent l'inconvénient d'être très variables d'une agence à l'autre. Seule une coopération optimale entre les agences et les assureurs, en terme de communication des valeurs assurées permet d'estimer au plus juste les dommages potentiels, et notamment les pires dommages possibles.

3.2 Les possibilités d'adaptation en prévoyance

Les modèles de catastrophes existants sont extrêmement exigeants en termes de localisation des biens assurés et en termes de connaissance de la catastrophe étudiée.

Imaginer une adaptation conduit aux deux complications suivantes :

- Il semble dans un premier temps qu'un tel modèle expert de catastrophe soit plus complexe encore à développer pour des portefeuilles d'assurance vie, ou d'assurance prévoyance. Il est en effet très peu réaliste de fixer les personnes assurées à une adresse donnée, d'autant plus que la mobilité des personnes est globalement de plus en plus grande, d'un point de vue mondial, et de surcroît dans les pays développés.
- La deuxième complication réside dans l'établissement de fonctions de vulnérabilité. Alors que la relation entre l'intensité de la catastrophe et les dommages matériels peut être assez bien formalisée, grâce à des données d'expérience dont les réassureurs peuvent disposer assez facilement par des enquêtes de terrain après les catastrophes, les impacts d'une catastrophe sur les vies humaines sont plus complexes.

D'abord parce que les impacts sont multiples. Une catastrophe peut avoir pour conséquences :

- Des décès
- Des blessés
- Des déplacés

Le rapprochement entre ces conséquences directes d'une catastrophe et les aléas définis contractuellement qui déclenchent des prestations n'est pas évident. Comment prévoir le nombre potentiel de personnes qui seront hospitalisées ou qui connaîtront une phase de chômage à la suite d'une catastrophe ? Ces nombres peuvent-ils être, sans trop d'approximation, approchés par des nombres potentiels de blessés et de déplacés ?

La seconde raison se trouve dans la difficulté de récolter des données d'expérience sur les coûts observés (nombres de victimes, coûts financiers constatés) lors de précédentes catastrophes. Cette remarque vaut particulièrement pour le portefeuille qui sera utilisé dans les applications (chapitre 4). Le problème vaut moins pour l'assurance IARD, domaine dans lequel les compagnies commencent à connaître les courbes de vulnérabilité. Ces données sont absolument nécessaires pour extrapoler des coûts potentiels futurs. Elles permettent de faire un lien précis entre l'intensité d'une catastrophe et le nombre de victimes constatées (tués, blessés, déplacés).

IV La valeur intrinsèque et le risque catastrophe

Même si les standards de ce concept sont actuellement en mutation, et notamment tous récemment (4 juin 2008) avec la publication par le CFO Forum de nouveaux principes de calcul, la valeur intrinsèque d'un portefeuille d'assurance peut toutefois être défini et les principes de bases énoncés. La connaissance des principes généraux suffira pour justifier une étude d'impacts possible.

1 Principes

La valeur intrinsèque d'une compagnie d'assurance est **la valeur actuelle des montants futurs probables distribuables aux actionnaires**. C'est une quantité dont le calcul est minutieux et qui nécessite de multiples précautions pour obtenir une quantité qui soit qualifiée de valeur actuelle « juste ». Les montants potentiellement distribuables peuvent provenir à la fois de l'activité d'assurance mais aussi de la liquidation des capitaux propres. Le premier montant est traditionnellement clairement isolé du second montant.

La valeur intrinsèque d'un portefeuille d'assurance, parfois appelé Valeur de l'In Force (VIF) ou valeur du stock en français est une composante de la valeur intrinsèque de la compagnie qui possède ce portefeuille. Son calcul consiste, hors ajustements liés au coût de la marge de solvabilité, à déterminer la valeur actuelle des profits probables projetés distribuables aux actionnaires, actualisés à un taux représentant le coût réel du capital. D'une manière plus condensée, il s'agit d'évaluer **la valeur actuelle probable des profits futurs générés par les contrats en cours**.

- Le calcul peut être fait de manière déterministe en projetant des scénarios uniques sur un horizon donné, pour chaque flux financiers (entrées et sorties comptables). La valeur intrinsèque est obtenue en sommant les résultats actualisés.
- Il peut se faire de manière stochastique, en identifiant un très grand nombre de possibilités pour chaque paramètre de l'activité d'assurance (revenus financiers, prestations versées etc.). La valeur intrinsèque est alors l'espérance de la valeur actuelle des résultats nets futurs (générés par l'activité d'assurance, c'est-à-dire les contrats en cours). Cette espérance est en pratique déterminée à partir d'une distribution empirique de la valeur

actuelle des résultats nets futurs, obtenues par des simulations de Monte Carlo (voir les éléments théoriques de ces simulations dans le chapitre 3).

2 *Les possibilités d'études d'impacts*

Les principes de la valeur intrinsèque étant désormais établis, il apparaît clairement la possibilité d'étudier l'impact que le risque catastrophe peut avoir sur la valeur actuelle de la richesse disponible aux actionnaires. Dans le cadre d'une valeur intrinsèque stochastique, la proposition suivante peut être faite.

La valeur intrinsèque prenant en compte des scénarios catastrophes est la valeur intrinsèque sans scénarios catastrophes diminuée de la valeur actuelle probable des pertes futures dues à la réalisation du risque catastrophe.

Le mode de calcul envisagé présente des exigences :

- **Les flux qui peuvent être impactés par la catastrophe doivent être identifiés.** L'hypothèse simplificatrice selon laquelle seules les prestations sont choquées par la survenance d'une catastrophe peut être formulée. Cette idée semble farfelue dans le cas d'une attaque terroriste du type World Trade Center car les rendements financiers peuvent être considérablement atteints.
- **Des modèles réalistes de fréquence et d'intensité des risques catastrophes auxquels le portefeuille est soumis doivent être élaborés.** La loi de fréquence permettra de connaître des années touchées par une catastrophe et la loi d'intensité de simuler un nombre de victimes, et par conséquent, par l'intermédiaire d'une courbe de vulnérabilité relativement simplifiée, un coût financier probable.

V Choix d'une problématique précise

Les deux parties précédentes ont identifié deux possibilités d'études d'impacts du risque catastrophe. Elles sont toutes les deux envisageables et méritaient d'être précisées de manière plus détaillée.

Mais avant d'annoncer le choix qui a été fait, des questions préliminaires se posent :

- Un choix de périmètres de portefeuille
- Un choix de zones géographique
- Un choix de catastrophes auxquels les zones choisies sont particulièrement soumises
- Un choix de modélisation des catastrophes retenues

Ces quatre questions constituent la première étape de la problématique du mémoire. Le chapitre 3 donne le choix de modélisations. Le chapitre 2 donne les choix des zones et des catastrophes. Le chapitre 4 présente les périmètres de portefeuilles retenus.

Concernant les études d'impacts, le chapitre 4 donne de manière beaucoup plus détaillée les orientations choisies. Les choix sont basés sur des contraintes pratiques.

La décision de rejeter l'étude d'impacts sur la valeur intrinsèque de portefeuille prévoyance à disposition a été prise car un travail d'implémentation de modèles de catastrophes était nécessaire pour pouvoir extraire des conclusions chiffrées et intéressantes.

La décision de rejeter des études d'impacts complètes sur le capital économique a également été prise car le calcul du capital économique nécessite les mêmes travaux de projection et d'implémentation que le calcul de la valeur intrinsèque. L'obtention de résultats comparatifs de capital économique sans et avec scénarios catastrophes aurait nécessité une implémentation rigoureuse et complète.

La décision a été de modéliser correctement des risques catastrophes potentiellement meurtriers puis d'effectuer l'étude d'impacts suivante. Cette décision a été prise avec la compagnie cliente qui détient le portefeuille.

L'application sur les données de portefeuille consiste à chiffrer un capital cible pour les risques catastrophes auxquels les pays du monde, dans lequel la compagnie cliente est considérablement implantée, sont particulièrement sensibles. Ce capital cible est soit un coût actualisé probable (risques catastrophes en France), par l'intermédiaire de simulations de Monte Carlo, soit un coût estimé de manière déterministe, en envisageant des scénarios probables (risques catastrophes au Japon). Ces capitaux cibles estiment uniquement le coût potentiel des garanties décès du portefeuille prévoyance à disposition. Ce coût est estimé à un degré de prudence élevé.

Ces capitaux cibles sont ensuite comparés aux exigences du CEIOPS (Committee of European Insurance and Occupational Pensions Supervisors) dans la quatrième étude quantitative d'impacts du projet de directive Solvabilité II. Cette comparaison se place dans le cadre de la modélisation interne des risques. La question consiste à se demander si un modèle élaboré par la compagnie pour mesurer ses risques peut déboucher à des capitaux de solvabilité moindres que les capitaux statutaires exigibles par le régulateur dans les formules standards des études d'impacts.

VI Le projet Solvabilité II et la prévoyance

Le projet Solvabilité II, dont l'échéance est prévue pour 2012, crée **un cadre prudentiel plus adapté aux risques réels pesant sur les compagnies d'assurance.**

1 Le projet de directive en bref

La réforme « Solvabilité II » est organisée sur trois piliers :

- pilier 1 : exigences quantitatives – mesures des besoins de marge de solvabilité
- pilier 2 : contrôles internes, gestion des risques et principes de gouvernance
- pilier 3 : informations et rapports au marché (public et contrôle)

Le pilier 1 va donner une vision plus réaliste des risques et exigera que le bilan les reflète à un niveau de prudence élevée. L'idée phare de Solvabilité II dans la recherche des besoins de couverture de la solvabilité des assureurs est la mise en oeuvre d'une probabilité de ruine à moins de 0,5 %.

Cette exigence passe par une connaissance précise des risques et des lois qui régissent leurs réalisations. Cette exigence étant forte, elle nécessite de répertorier, comprendre et évaluer les risques dont la fréquence de réalisation est très faible (par exemple des périodes centennales ou bicentennales) et dont le coût peut être potentiellement très élevé.

Les études quantitatives d'impacts sont un ensemble de préconisations faites par le CEIOPS pour tester, sur les compagnies et mutuelles qui acceptent de participer à cette étude, les impacts de la nouvelle réglementation, principalement au niveau des provisions techniques, dont le mode de calcul est considérablement remanié, et au niveau de la marge réglementaire de solvabilité (statutaire, brutale et ne reflétant pas les risques dans la réglementation actuelle Solvabilité I).

La dernière en date, la quatrième (QIS4), disponible depuis avril 2008, consacre plusieurs paragraphes à la couverture des risques catastrophes. Elle propose des formules standards qui pourraient, selon le CEIOPS, suffire à absorber les risques catastrophes. Plusieurs formules sont proposées suivant le type d'activité de la compagnie (Vie, Santé et Non Vie).

Le paragraphe suivant donne les conclusions d'une analyse du QIS4.

2 La prise en compte du risque catastrophe

Une analyse de la version définitive du QIS4 met en évidence quatre formules standards pour les risques catastrophiques. Dans le cadre d'une activité prévoyance, une d'entre elles doit être clairement prise en compte. Deux sont exclues. La prise en compte de la quatrième est très incertaine.

1/ La formule qui doit être à coup sûr prise en compte se trouve dans le Module Risque de souscription vie SCRlife.

(TS.XI.H. Risque de catastrophe Lifecat P 155 des spécifications techniques en français)

2/ Les deux formules clairement exclues sont la formule du module Risque de souscription non-vie et la formule du sous-module Risque de souscription santé à long terme du module Risque de souscription santé. En effet le premier module exclut les trois premières lignes de business non vie

(LOB) et le sous-module « Risque de souscription santé à long terme » du module Risque de souscription santé ne concerne que l'Autriche et l'Allemagne.

3/ La formule dont la prise en compte est incertaine est la formule située dans le module Risque de souscription accident et santé court terme (TS.XII.C.13 Risque de catastrophe en accident et santé à court terme) et dans le module Risque de souscription en accidents du travail (TS.XII.D.32 Risque de catastrophe en accidents du travail WCompCAT).

Chapitre 2 Sélection des catastrophes et sensibilité du portefeuille

Ce chapitre propose d'abord de mettre en confrontation diverses sources afin d'établir une définition d'une catastrophe, saisir leurs propriétés essentielles et dégager une typologie relativement précise. Il donne dans un deuxième temps les résultats d'une analyse de sensibilité du portefeuille de prévoyance présenté dans le premier chapitre.

I Définition et typologie

Les sources nationales et internationales retenues sont les suivantes :

- Base de données EMDAT (The International Emergency Disasters Database) du CRED (Centre of Research and Epidemiology of Disaster). Cette source est citée par l'OCDE dans « Les risques émergents aux XXIème siècle » en 2003.
- Le Ministère de l'Environnement et du Développement Durable (MEDD), relayé par l'INSEE dans « Les risques majeurs naturels et technologiques » en 2007 (Courrier des statistiques n° 120).
- La compagnie de réassurance Swiss RE. Notamment une publication de la série SIGMA parue en 2006 : « catastrophes naturelles et techniques en 2005 ».
- L'Institut des Risques Majeurs de Grenoble (IRMa).

Ces sources présentent l'avantage de proposer des définitions précises et une classification riche. La première source a la particularité de fournir des historiques centennaux.

Pour fixer le vocabulaire, nous pouvons établir un consensus global sur le terme de risque majeur (ou désastre) et sur sa définition :

- Faible fréquence et énorme gravité : dans un seul accident, de très nombreuses victimes et/ou des dommages importants pour les biens.
- Présence d'un événement, qui est la manifestation d'un phénomène naturel ou anthropique (dû à l'Homme)
- Existence d'enjeux, qui représentent l'ensemble des personnes et des biens affectés par un phénomène.

Le CRED donne une version quantifiée et raisonnable de cette définition : c'est celle que nous pouvons retenir :

- Déclaration d'état d'urgence ou d'un appel à l'aide internationale
- Au moins 10 morts
- Au moins 100 personnes touchées

Nous constatons enfin des exclusions qui font l'unanimité. La notion même de désastre n'intègre pas différents risques dont les effets sont diffus dans l'espace et dans le temps, tels que ceux liés à l'utilisation de l'amiante ou les risques de la vie quotidienne (accidents domestiques, accidents de la route...), les risques liés aux conflits/guerres ou guerres civiles.

Une classification des risques majeurs passe d'abord par une confrontation de sources. Nous pouvons décliner les réponses en deux possibilités : oui ou non.

Le tableau suivant résume les points de vue :

groupe de catastrophes	type de catastrophes	Sources			
		EMDAT	MEDD/INSEE	SWISS RE	INRA
catastrophes naturelles	épidémie/pandémie	oui	non	non	non
	afflux d'insectes ou de parasites	oui	non	non	non
	raz de marée et tsunamis	oui	non	oui	non
	feu de forêts	oui	oui	oui	oui
	tempête de vent	oui	oui	oui	oui
	sécheresse	oui	non	oui	non
	canicule ou vague de froid	oui	non	oui	non
	montée des eaux, crue de rivière ou lac	oui	oui	oui	oui
	mouvement de terrain - avalanche	oui	oui	non	oui
	tremblement de terre	oui	oui	oui	oui
	éruption volcanique	oui	oui	non	non
	accident industriel	oui	oui	oui	oui
catastrophes technologiques	accident non industriel	oui	oui	oui	oui
	catastrophe liée aux transports	oui	oui	oui	oui
catastrophes diverses	trouble social	non	non	oui	non
	terrorisme	non	non	oui	non
	conflit/guerre/guerre civile	non	non	non	non

Les types de catastrophes sont définis comme suit :

En résumé :

- Les conflits, guerres et guerres civiles ne sont pas des risques majeurs.
- Les troubles sociaux et le risque de terrorisme apparaissent parfois formellement dans les exclusions des conditions générales de contrats de prévoyance, notamment d'assurance emprunteur. Il convient donc de consulter ces conditions et de disposer d'un avis d'expert. Sur le portefeuille étudié, les catastrophes diverses (et notamment le terrorisme) font partie des risques exclus, ceux-ci ne seront donc pas retenus par la suite
- Tous les autres types de catastrophes seront considérés comme des risques majeurs, même si pour certains, les sources françaises ne les donnent pas de manière explicite dans leurs classifications

II Sélection des couples pays/catastrophe - Sensibilités

Le but de cette section est d'établir l'ensemble des couples pays/catastrophe qui seront retenus et les couples qui seront écartés, pour chaque pays analysé.

Le choix de segmenter le portefeuille par pays est un choix fait conjointement avec la compagnie cliente, car leurs bases de données sont segmentées de cette manière.

Dans un pays donné, il est nécessaire de choisir des critères pour mesurer l'exposition réelle du portefeuille à un risque de catastrophe spécifié. Pour cela, il est important de considérer les enjeux qui peuvent être traduits par une part de marché. Par exemple, la mesure d'exposition du portefeuille turque au risque de tremblement de terre, qui en théorie est très élevée, dépendra du nombre d'assurés présents dans le pays par rapport à la population totale, c'est-à-dire de la présence de la compagnie.

La classification, résumée par le tableau qui suit plus bas, est basée sur l'estimation du nombre potentiel d'assurés touchés (décédés ou blessés) dans le pays à cause de la catastrophe. Cette estimation est faite à partir du ratio : nombre d'assurés dans le pays/population totale du pays

- Si ce nombre est inférieur à 10, l'exposition est nulle
- Si ce nombre est entre 10 et 100, l'exposition est faible
- Si ce nombre se situe entre 100 et 1000, l'exposition est moyenne
- Si ce nombre dépasse 1000, l'exposition est forte.

Cette vulnérabilité potentielle est estimée à partir d'historiques de catastrophes, fournis par le CRED qui répertorie deux types d'historiques, pour tous les pays du monde et pour toutes les catastrophes :

- Des historiques centennaux de blessés et de déplacés
- Des historiques centennaux de tués

Le tableau de sensibilité est le suivant, pour un périmètre de pays où la compagnie cliente émet une priorité. Le risque de terrorisme étant un risque relativement récent, il nécessite un traitement particulier. Soit il doit être totalement exclu si les conditions générale le prévoient, soit il doit être pris en compte dans la mesure du possible. Des remarques supplémentaires sont faites sur ce risque dans la section suivante.

		catastrophes naturelles											catastrophes technologiques			catastrophes diverses		
		épidémie/ pandémie	afflux d'insectes/ parasites	raz de marée et tsunamis	feu de forêts	tempête de vent	sécheresse	canicule vague de froid	montée des eaux, crue de rivière ou lac	mouvement de terrain avalanche	tremblement de terre	éruption volcanique	accident industriel	accident non industriel	catastrophe liée aux transports	trouble social	terrorisme	
Pays	Austria							+									Worst Case	
	Belgium							+					+	+			Worst Case	
	Chile			+					++	+	+++	+	+				Exclu	
	France							+++					++		+		Worst Case	
	Germany							+++					+				Worst Case	
	Italy							+++	+		+++			+	+		Worst Case	
	Japan			+		+		+	+		+++		+	+	+		Worst Case	
	Netherlands							+	+					+			Worst Case	
	Poland							+						+	+		Worst Case	
	Portugal							++	+									Worst Case
	Spain							++	+				++	+	+	+		Worst Case
	Taiwan					+			+		++		+					Exclu
Thailand			++					+									Exclu	
UK					+										+		Worst Case	

La section suivante donne les décisions prises avec la compagnie cliente. Les choix sont faits à partir de ce tableau d'exposition dont l'intérêt est d'éclairer sur des priorités de traitement.

III Choix : priorités et exclusions

1 Les priorités

Il apparaît clairement que les risques de canicule (température extrême) et de tremblement de terre sont les risques les plus importants d'un point de vue mondial.

Il est nécessaire de récolter des informations qualitatives concernant les évolutions probables des risques catastrophiques pour confirmer ou rejeter la conclusion qui vient d'être faite.

Trois sources ont été utilisées :

- ✓ Un rapport de Swiss Re paru dans la série SIGMA. N°2/2006
- ✓ Le rapport annuel du CRED de 2006
- ✓ Le rapport de l'OCDE « les risques émergents du XXIème siècle » publié en 2003

Les informations convergent d'une manière manifeste. La synthèse suivante des avis d'expert tente de reprendre les grands changements sociaux, démographiques et environnementaux du XXIème siècle et dont les conséquences seront, dans la mesure du possible, impactées sur l'exposition au risque du portefeuille.

- Tremblements de terre, tsunamis : même si les installations parasismiques permettent de réduire les risques d'effondrements d'immeubles, entraînant avec eux des vies humaines, l'ampleur de ces types de catastrophes est imprévisible et pourrait potentiellement surpasser les capacités d'absorption des installations parasismiques.
De plus, selon l'OCDE, plus du tiers de la population mondiale vivra dans une zone sismique à risque d'ici 40 ans
- Tempêtes, vagues de froid, vagues de chaleur : Multiplication de ces événements jusqu'en 2050 à cause du réchauffement de la planète. En effet, malgré les bonnes volontés actuelles des pays développés pour réduire les émissions de gaz à effet de serre, le réchauffement de la planète s'est enclenché et continuera sur un horizon de 40 ans au minimum. En Europe plus particulièrement, Swiss Re prévoit une tempête du type « 1999 » tous les 8 à 10 ans.

Concernant le golfe du Mexique, les Caraïbes et la Floride, un autre paramètre est à prendre en compte : la phase chaude de l'AMO (Oscillation Atlantique Multidécennale) semble lancée depuis 1995 et va encore durer de 10 à 30 ans. Une conséquence directe est une répétition de cyclones dévastateurs comme Andrew ou Katrina.

- Les accidents de transport ne vont pas décroître dans les années à venir à cause notamment des catastrophes maritimes potentielles, dues à des feries ou à des embarcations trop chargées. Les pays touchés seront les pays en voie de développement. A titre d'exemple, les Philippines ont récemment été marquées par un accident maritime ayant causé plusieurs milliers de morts.
- Sécheresse, inondations, tempêtes de vent : les pays les plus touchés sont l'Inde et la Chine, qui depuis 1987, connaissent un désastre tous les 3 ans, d'après des recoupements de données effectués par le CRED.
- Epidémies :
 - Le développement agricole, les destructions de forêts, et les constructions d'habitations constituent un facteur de développement de maladies connues (du type malaria), ou de maladies inconnues. Le Brésil et la Colombie sont des pays fortement exposés.
 - La vitesse d'extension des épidémies sera plus rapide du fait de la mobilité croissante des êtres humains et du développement du tourisme
 - En Europe et aux USA, certains experts craignent l'apparition de la malaria. Dans ces mêmes régions le risque de maladies infectieuses véhiculées par les aliments est présent et grandissant.
 - Les espoirs placés dans les progrès de la science permettent toutefois de limiter ce pessimisme.

Ces analyses permettent d'effectuer un éclairage supplémentaire sur l'exposition du portefeuille au risque catastrophe. Toutefois ces points de vue alarmistes doivent être croisés avec les préoccupations pratiques de la compagnie cliente : connaître la sensibilité des vies humaines assurés sur un périmètre géographique bien précis.

Cette analyse croisée fait apparaître clairement la nécessité de traiter :

- **Le risque de canicule en Europe.** Toutefois étant donné que la très grande majorité des personnes hospitalisées ou décédées à cause de la canicule ont plus de 65 ans, le nombre de personnes du portefeuille prévoyance potentiellement touchées par la prochaine canicule semble faible. Malgré tout, à cause du vieillissement de la population européenne le risque de canicule est une menace grandissante pour les assureurs. Cette menace est pour l'instant limitée mais pourrait potentiellement devenir extrêmement coûteuse pour les assureurs. Ce risque constitue donc une priorité de traitement.
- **Le risque de séisme** doit être pris en compte. Malgré les progrès réalisés dans les installations parasismiques, cette catastrophe géologique est potentiellement illimitée en intensité. Elle constitue la deuxième priorité de traitement.
- **Le risque d'épidémie/pandémie** doit être analysé en distinguant chaque type de pandémie. La compagnie cliente a écarté ce risque du thème de ce mémoire car elle le traite en interne.
- Les autres risques ayant une exposition ++ pourront être traités de deux manières différentes suivant si des méthodes de modélisation sont identifiables, applicables et adaptables.
- Le risque de terroriste ne sera pas traité. Il peut toutefois être modélisé et géré. L'OCDE indique :

« Différentes idées novatrices ont été mises en avant dans le but de modéliser les risques terroristes (Major, 2002 ; Woo, 2002) ; récemment, plusieurs modèles ont ainsi été élaborés. Reposant sur des opinions d'experts (recueillies par exemple en appliquant les méthodes Delphi) et/ou des modèles de la théorie des jeux appliqués au comportement, ils évaluent la probabilité pour un lieu donné de devenir la cible d'une attaque terroriste, et la probabilité de la réussite de l'attaque. Un module de simulation estime ensuite les dommages encourus.

2 Les exclusions

Certaines exclusions de risques apparaissent clairement : les conditions générales excluent textuellement les **catastrophes diverses** (guerres civiles ou étrangères, participations volontaires à des crimes, délits, mouvements populaires, attentats ou meurtres) qui prennent notamment en compte le risque de terrorisme, du champ d'assurabilité. Les pays de l'Union européenne, la Suisse, l'Islande, la Norvège, les Etats-Unis, le Japon et le Canada constituent des cas particuliers où

l'exclusion ne s'applique pas. Dans ces pays là, le coût éventuel d'une catastrophe pourra être extrapolé à partir d'un historique en retenant par exemple le pire des cas sur les vingt dernières années et en le répétant sur le portefeuille actuel (voir chapitre sur les modélisations de catastrophes).

Les **catastrophes nucléaires** font également partie des exclusions de garantie. De ce fait, comme ce type de catastrophe est classé dans les catastrophes industrielles, il faudra s'assurer avant toute chose, si un couple pays/catastrophe industrielle est retenu, que l'historique ne contient pas de catastrophe nucléaire.

3 Conclusion

Le choix final a été de traiter le risque de séisme au Japon et le risque de canicule en France par des modélisations probabilistes (voir chapitre sur les modélisations de catastrophes)

- Modélisation de la mortalité due aux séismes au Japon et de la fréquence de survenance de séismes meurtriers
- Prévision de la fréquence des canicules en France et analyse de la surmortalité lors de la canicule de 2003.

Les autres risques catastrophes majeurs en France et au Japon seront traités par une approche « Worst case », c'est-à-dire de manière déterministe.

Chapitre 3 Modélisation interne du risque catastrophe

Une modélisation interne du risque catastrophe consiste à répertorier les risques, puis à les comprendre. La compréhension d'un risque catastrophique consiste soit à envisager la pire des réalisations et à la répéter sur le portefeuille (approche Worst Case) afin d'estimer le coût actuel probable, soit à modéliser le risque de manière probabiliste afin de le mesurer par un intervalle de confiance.

- Le premier paragraphe présente, analyse et confronte les deux approches.
- Le deuxième paragraphe donne les éléments théoriques sur la génération de nombres et la simulation de lois et processus en assurance. Ces éléments sont transversaux car ils valent pour les modélisations déterministes (dans le cadre de simulations du pire des cas sur le portefeuille) et les modélisations stochastiques (simulation des lois estimées, impacts sur le portefeuille)
- Les deux derniers paragraphes proposent une modélisation de la mortalité due aux séismes au Japon et de la fréquence des canicules en France.

I Approche déterministe et modélisation stochastique

Il est possible de distinguer deux types d'approches pour évaluer les risques catastrophiques, à savoir l'approche déterministe et l'approche stochastique. Cette classification est inspirée de celle présentée par le réassureur Swiss RE.

1 *L'approche déterministe*

Cette approche est la plus simple pour estimer le potentiel d'un sinistre. Il faut recourir à des événements majeurs antérieurs qui sont rapportés aux valeurs assurées aujourd'hui. Cette analyse est qualifiée d'analyse « as-if ». Il s'agit par exemple de répéter le pire des cas (« Worst Case ») sur le portefeuille actuel.

L'inconvénient évident est que tous les événements possibles ne sont pas pris en compte.

2 L'approche stochastique ou « fondée sur des scénarios »

Les méthodes probabilistes d'évaluation des risques ont peu à peu émergé comme des compléments indispensables aux études déterministes classiques, auxquelles elles apportent une vision plus équilibrée et réaliste des situations de risque.

Les modèles doivent impérativement se baser sur les sinistres antérieurs, car c'est la seule source d'information pour estimer :

- La fréquence d'occurrence de la catastrophe
- L'intensité des catastrophes passées
- Des extrapolations

Par nature, une modélisation stochastique passe par deux étapes :

- Estimation d'une loi de probabilité ou d'un processus stochastique sur la fréquence et l'intensité du risque étudié
- Simulation de cette loi ou de ce processus.

Les estimations effectuées permettent de générer, par l'intermédiaire de simulations de Monte Carlo, des réalisations du risque qui n'ont jamais eu lieu mais qui pourraient se produire à l'avenir avec une forte probabilité. En effet, une modélisation stochastique est considérée comme une vision prospective dans le sens où elle envisage des sinistres possibles pour les années à venir.

Les réassureurs ont adopté cette vision prospective de l'aléa en simulant toutes ses caractéristiques (trajectoires, intensité etc.) au moyen d'outils informatiques perfectionnés. Swiss Re explique dans le cas des cyclones :

*« Pour générer la série d'événements, les trajectoires des cyclones **antérieurs** sont modifiées selon un mouvement pseudo-aléatoire (directed random walk), à savoir une simulation mathématique basée sur des nombres aléatoires (méthode de Monte Carlo). La validation des séries d'événements*

générées artificiellement se fait en comparant les données climatiques simulées et celles enregistrées par le passé. »

Cet esprit se trouve au cœur de la motivation des modélisations et simulations stochastiques. Dans le cadre de ce mémoire, des simulations simplifiées seront effectuées. L'idée est de générer artificiellement, à partir de l'historique, des événements qui ont de grandes chances de se produire à l'avenir. Par l'intermédiaire de simulations, une vision prospective est intégrée tout en respectant l'expérience passée.

Les modélisations déterministes peuvent par ailleurs être considérées comme une échappatoire en cas d'échec de la modélisation stochastique. En effet il est parfois difficile, impossible ou critiquable de calibrer des lois ou processus sur les caractéristiques d'un risque catastrophe. Les modélisations déterministes permettent d'éviter la modélisation stochastique en se basant sur le pire des cas de l'historique.

2.1 Approche retenue pour le risque de séisme au Japon

L'historique du nombre de décès dus aux séismes est une série chronologique marquée par des excès très ponctuels et une majorité d'années peu touchées voire pas du tout. L'objectif est de trouver une ou des lois qui reflète(nt) bien ce genre d'observations, au niveau de l'intensité comme au niveau de la fréquence. L'intérêt est de pouvoir ensuite simuler les lois calibrées pour générer des scénarios possibles pour les années à venir. Ces scénarios étant générés à partir d'observations passées, les graphiques obtenus seront très proches du graphique de l'historique

2.1.1 Modélisation de l'intensité

L'approche retenue est de calibrer une loi de type Pareto sur les mortalités au-delà d'un seuil. D'après la théorie des extrêmes, la loi dite de Pareto Généralisée (GPD en anglais) est une loi naturelle des excès au-delà d'un seuil.

2.1.2 Modélisation de la fréquence

La fréquence de survenance de séismes meurtriers (1 victime au moins) est considérée comme un processus stochastique. Il a été choisi de calibrer un processus de Poisson sur ce processus de survenance car imaginer des temps d'attente distribués par une loi exponentielle semble assez vraisemblable.

En résumé, le modèle calibré est un modèle « Peak Over the Threshold » (POT). Stricto sensu, ce modèle ajuste une loi GPD sur les excès au-delà d'un seuil et le processus de survenance de ces excès par un processus de Poisson. Le modèle calibré est donc en réalité un modèle POT adapté.

2.2 Approche retenue pour les canicules en France

Le risque de canicule est un risque considérablement difficile à traiter puisque la France n'a connu que deux épisodes de ce type depuis 100 ans : celui de 1976 et celui de 2003. De ce fait, il est impossible de calibrer une quelconque loi d'intensité. Il est par contre envisageable de développer un modèle de prévision réaliste. L'intensité d'une future canicule, si elle survient dans les prochaines années, sera supposée identique à celle de 2003, dans le sens où les impacts démographiques, c'est-à-dire les surmortalités observées, seront les mêmes qu'en 2003. Il est par contre envisageable de développer un modèle de prévision réaliste. La fréquence sera modélisée en calibrant un modèle de type ARIMA sur l'historique des températures journalières. La simulation de ce modèle permettra de mettre en évidence une fréquence potentielle.

2.2.1 Approche retenue pour les autres catastrophes

Les risques d'accidents industriels (risque catastrophe mineur en France), de tempêtes et d'inondations (risques catastrophes mineurs au Japon) seront gérés en adoptant une approche « as-if », c'est-à-dire en répétant le pire des cas ou un des pires cas sur le portefeuille actuel. Répéter le pire des cas consiste à simuler la catastrophe sur le portefeuille de la compagnie exposée ou à le choquer de manière brutale, par un simple prorata. Ces approches seront abordées dans le chapitre 4.

II But des modélisations

Les travaux de mesure de risques catastrophiques entrepris doivent répondre au problème spatio-temporel suivant : estimer la loi de fréquence et la loi d'intensité des catastrophes retenues dans les pays sélectionnés.

Dans le cadre de problématiques d'assurance, et notamment de gestion actif/passif, ces modélisations permettront de réaliser les travaux suivants :

- La loi de fréquence permet de tirer une année au hasard sur 30 ans et la loi d'intensité de tirer au hasard une mortalité pour l'année générée. Dans le cadre d'une Embedded Value stochastique, ces calibrages pourront être intéressants pour analyser l'impact des catastrophes sur la valeur intrinsèque.
- La loi de fréquence permet également de connaître la probabilité de survenance de la catastrophe pour l'année à venir. La loi d'intensité permet ensuite de déduire des Value at Risk.

III Générer des nombres et simuler en assurance

Cette section permet de fixer les éléments théoriques de base nécessaires pour toute simulation stochastique.

La méthode de simulation de Monte Carlo, basée sur la loi des grands nombres permet de simuler une loi de probabilité ou de manière très fréquente en pratique et notamment en finance, de générer des trajectoires d'un processus stochastique, au préalable discrétisé. Dans notre cas, le but est de disposer d'un très grand nombre de scénarios afin d'en déduire empiriquement, par le biais de la loi des grands nombres, les quantiles désirés de la variable aléatoire simulée. La génération de trajectoires passe obligatoirement par deux étapes : d'abord générer des réalisations d'une variable aléatoire de loi uniforme, puis en déduire une réalisation de la loi de probabilité, en inversant la fonction de répartition

La génération des trajectoires nécessite la génération de nombres aléatoires. De manière pratique, il s'agit de générer des réalisations d'une variable aléatoire uniforme sur le segment $[0,1]$. En effet, si

u est une telle réalisation alors $F^{-1}(u)$ peut s'apparenter à une réalisation d'une variable aléatoire de fonction de répartition F . La technique d'inversion de la fonction de répartition permet ainsi, à partir de réalisations de variables uniformes, d'obtenir des réalisations d'autres variables aléatoires.

Lorsqu'il n'y a pas de formule explicite pour F^{-1} , des algorithmes d'approximation de cette fonction ou des algorithmes spécifiques à la loi traitée. C'est précisément le cas pour l'étude des séismes, où aucune loi à priori n'est connue

IV Le risque de séisme au Japon

Ce pays, à la jonction de quatre plaques tectoniques, est extrêmement soumis au risque de séismes.

L'histoire du Japon est marquée par deux séismes extrêmement meurtriers :

- Le séisme de 1995 à Kobe qui fit 5297 victimes
- Le tremblement de terre de 1923 où les secousses causèrent la mort de 143 000 personnes

Cette partie tente de modéliser la fréquence et l'intensité meurtrière des séismes au Japon. Au préalable, les éléments théoriques utilisés sont présentés dans une première section.

1 *Éléments théoriques de modélisation*

Les éléments théoriques utilisés sont au nombre de 2 :

- La loi de Zipf est une loi caractéristique de quantités. Le hasard fait que le nombre de victimes dû aux séismes, quelque soit par ailleurs la zone géographique considérée, suit assez bien cette loi. Cette loi est une loi de type puissance.
- La théorie des valeurs extrêmes et notamment plus particulièrement la méthode « Peak Over the Threshold » est présentée. Elle est particulièrement adaptée pour caractériser les processus de survenance d'événements rares (du point de vue de la fréquence comme du point de vue de l'intensité).

1.1 Loi de Zipf

Supposons que l'on dispose de n quantités ordonnées de la plus grande à la plus petite :

$$x_1 \geq x_2 \geq \dots \geq x_r \geq \dots \geq x_n$$

Ces quantités vérifient la loi de Zipf si l'on peut trouver des coefficients α et C tels que, pour i de 1 à r , avec r de préférence pas trop grand (d'après Gordon Woo).

$$x_i = \frac{C}{i^\alpha}$$

Ce qui s'écrit de manière équivalente :

$$\text{Log}(x_i) = -\alpha \log(i) + c + \varepsilon$$

Avec $c = \log(C)$

Cette loi se retrouve dans les domaines suivants :

- Nombre de fois x_r que le mot numéro r est utilisé dans les langues vivantes
- Populations x_r de la r ème ville la plus peuplée

α est souvent proche de 1 pour les quantités qui vérifient la loi de Zipf. ε est un terme d'erreur.

La loi de Zipf a une expression probabiliste. En effet, la densité de la variable aléatoire X , dont les réalisations ordonnées sont $x_1 \geq x_2 \geq \dots \geq x_r \geq \dots \geq x_n$, peut s'écrire sous la forme suivante :

$$f_X(x) = \frac{C^{1/\alpha}}{\alpha n} * \frac{1}{x^{(1/\alpha)+1}} = \frac{A}{x^\beta}$$

Avec $A = \frac{C^{1/\alpha}}{\alpha n}$ et $\beta = (1/\alpha) + 1$

X est donc distribuée selon une loi de type puissance. On dira donc que des quantités ordonnées suivent la loi de Zipf si la loi de la variable X , dont les quantités observées sont des réalisations, est une loi de type puissance.

1.2 Théorie des valeurs extrêmes

1.2.1 Caractérisation des extrêmes : trois théorèmes principaux

Les théorèmes qui suivent sont donnés sans démonstration. Le lecteur intéressé peut consulter Paul Embrechts et al. « Modelling Extremal Events for insurance and finance ».

L'application sur les séismes se base sur ces théories.

i Théorème fondamental de convergence en loi des extrêmes

Ce théorème marque le début de la théorie des valeurs extrêmes en donnant la loi limite des maximas. Il a été énoncé dans :

Fisher, R.A and Tippett, L.H.C (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample.

La première preuve rigoureuse a été donnée par Gnedenko en 1943.

Soit :

$(X_1, \dots, X_n)$ un échantillon de variables aléatoires indépendantes et identiquement distribuées.

M_n désigne $\max(X_1, \dots, X_n)$

Le théorème de Fisher-Tippett se base sur ces hypothèses simples et énonce une convergence en loi fondamentale :

S'il existe deux suites normalisantes de réels $c_n > 0$ et $d_n \in \mathfrak{R}$ et une fonction de distribution H tel que

$$\frac{M_n - d_n}{c_n} \xrightarrow{\text{loi}} H,$$

alors H est forcément de l'une de trois formes suivantes :

Fréchet	Weibull	Gumbel
$\Phi_{\alpha}(x) = \begin{cases} \exp(-x^{-\alpha}) & \text{si } x > 0, \\ 0 & \text{sinon} \end{cases}$ $\alpha > 0$	$\psi_{\alpha}(x) = \exp(-(-x)^{\alpha})$ si $x \leq 0$, 1 sinon $\alpha > 0$	$\Delta(x) = \exp(-e^{-x})$ $x \in \mathbb{R}$

Si F désigne la fonction de répartition des X_i , alors F est dite « dans le domaine d'attraction de H ».

Ces trois lois caractérisent chacune un comportement spécifique des valeurs extrêmes (au préalable correctement normalisées).

La question de généraliser les trois domaines d'attraction se pose immédiatement. C'est-à-dire trouver une formule analytique à la fonction de distribution suivante :

$$H_{\xi} = \begin{cases} \Phi_{1/\xi} & \text{si } \xi > 0 \\ 1 & \text{si } \xi = 0 \\ \Psi_{-1/\xi} & \text{si } \xi < 0 \end{cases}$$

Autrement dit, il s'agit de trouver une généralisation qui englobe les trois familles énoncées dans le tableau :

Fréchet	Weibull	Gumbel
$\Phi_{\alpha}(x) = \exp(-x^{-1/\xi})$ si $x > 0$, 0 sinon $\xi = \alpha^{-1} > 0$	$\psi_{\alpha}(x) = \exp(-(-x)^{1/\xi})$ si $x \leq 0$, 1 sinon $\xi = -\alpha^{-1} < 0$	$\Delta(x) = \exp(-e^{-x})$ $\xi = 0$

ii Généralisation de Jenckison Von Mises : distribution généralisée des valeurs extrêmes

La représentation de Jenkinson-Von Mises fournit une caractérisation synthétique des lois extrêmes: la distribution généralisée des valeurs extrêmes (Generalized Extreme Value, ou GEV). La fonction de répartition GEV a la forme suivante :

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

- μ est un paramètre de localisation, il est directement lié à la valeur la plus probable de la loi, il indique donc approximativement où se trouve le cœur de la distribution.
- σ est un paramètre de dispersion, il indique l'étalement des extrêmes.
- ξ est l'indice de queue ou indice des valeurs extrêmes.

La loi initiale F influe sur la distribution limite H au travers de la forme de ses queues de distribution, qui détermine les paramètres μ et σ .

iii Distribution au-delà d'un seuil : la distribution de Pareto Généralisée

La distribution de Pareto généralisée est définie par la fonction de densité suivante :

$$G_{\xi, \beta}(x) = 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi}$$

Avec $\beta > 0$ et $x \in D(\xi, \beta) = \begin{cases} [0, -\beta/\xi] \\ [0, +\infty) \end{cases}$ si $\xi \geq 0$

Pickand a énoncé une propriété de convergence en loi de la variable aléatoire avec u un seuil choisi de manière judicieuse. Ce théorème donne l'équivalence suivante :

Pour tout $\xi \in \mathbb{R}$, F appartient au domaine d'attraction de H_{ξ} si et seulement si :

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

Avec :

- $x_F = \sup\{x / F(x) < 1\}$
- $1 - F_u(x) = P(X - u > x | X > u)$

Cette convergence peut être écrite de manière plus simplifiée :

Lorsque u est grand, c'est-à-dire proche du point terminal x_F , alors on peut trouver un $\beta(u) > 0$ et un ξ tel que :

$$F_u(x) \approx G_{\xi, \beta(u)}(x)$$

Signalons une propriété de linéarité importante :

$$e(u) = E(X - u | X > u) = \frac{\beta + \xi u}{1 - \xi}$$

Avec $u \in D(\xi, \beta)$, $\xi < 1$

La moyenne empirique de excès au-delà du seuil est un estimateur sans biais de $e(u)$:

$$e_n(u) = \frac{1}{N_u} \sum_{i \in \Delta_n(u)} (X_i - u)$$

Avec $N_u = \text{card}\{i : i = 1, \dots, n, X_i > u\} = \text{card}\Delta_n(u)$

Cette propriété sera utilisée par la suite pour justifier le choix d'un seuil u .

iv Première conclusion

Deux caractérisations ont été mises en évidence :

- Les maximums correctement normalisés suivent une loi limite $GEV_{H_{\xi}}$ quand la taille de l'échantillon n est grande. Le paramètre réel ξ est appelé indice de queue ou indice des extrêmes
- Les excès au-delà d'un seuil sont distribués selon une loi de Pareto Généralisée pour u grand. Les paramètres de cette loi sont ξ et $\beta(u) > 0$. Cette loi change avec le seuil choisi et le paramètre ξ est le même que le paramètre de la loi GEV : c'est l'indice des extrêmes.

Naturellement, l'étape suivante consiste à se demander comment estimer les paramètres ξ et β

1.2.2 Une méthode d'estimation des paramètres d'une loi de Pareto Généralisée.

i Identification préalable du type de valeur extrême

L'indice ξ des valeurs extrêmes, qui constitue le premier paramètre à estimer, est soit nul, soit positif, soit négatif. Les trois cas correspondent respectivement aux domaines d'attraction de Gumbel, Fréchet et Weibull.

Avant de proposer une estimation, il convient de se demander quelle est la loi des extrêmes du phénomène étudié. Le recours à un diagramme quantile-quantile (QQ-plot en anglais) est judicieux car ce graphique présente l'avantage d'être simple à construire et donne une première information sur le type de valeurs extrêmes.

Un diagramme quantile-quantile, par construction, porte les quantiles x_i de la distribution observée en ordonnée tandis que l'axe des abscisses porte les quantiles x_i^* correspondants de la loi théorique. Lorsque la distribution théorique proposée est une bonne représentation des observations alors le nuage de points (x_i^*, x_i) s'aligne sur une droite.

Par la suite, un diagramme quantile-quantile de la loi de Pareto et un diagramme quantile-quantile de la loi exponentielle seront construits.

Le premier diagramme permettra d'identifier le type de valeurs extrêmes tandis que le deuxième trouvera sa place dans le calibrage d'un modèle Peak Over the Threshold (voir les parties 4 et 5).

ii L'estimateur de Hill

- *Définition*

L'estimateur de Hill est fondé sur la statistique d'ordre $X_{k,n} \leq \dots \leq X_{1,n}$ obtenue à partir de la série initiale en considérant les k valeurs les plus grandes.

L'estimateur de Hill n'est utilisable que pour $\xi > 0$. Il est défini par la statistique suivante :

$$\hat{\xi}_{k,n}^H = \frac{1}{k-1} \sum_{j=1}^{k-1} \ln \left(\frac{X_{j,n}}{X_{k,n}} \right)$$

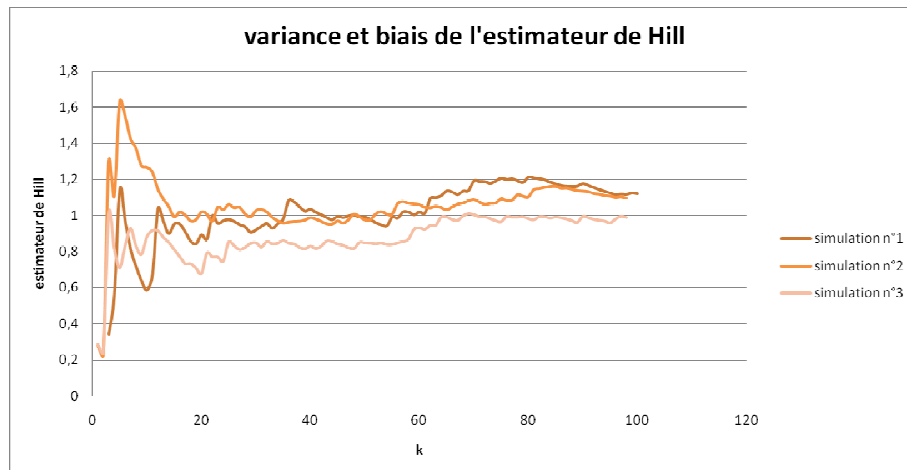
En choisissant k et n tendant vers l'infini et k/n tendant vers 0, il est possible de démontrer les deux propriétés suivantes :

$\hat{\xi}_{k,n}^H$ converge presque sûrement vers ξ

La variable $\sqrt{k} \frac{\hat{\xi}_{k,n}^H - \xi}{\xi}$ converge en loi vers une loi normale centrée réduite $N(0,1)$

iii Choix du nombre k

Le choix du nombre k est crucial car l'estimateur possède la spécificité, d'après un théorème de convergence en loi, d'être volatil lorsque k est plutôt petit et biaisé lorsque k est plutôt grand. Pour illustrer ce défaut, voici un graphique qui représente l'estimateur de Hill en fonction de k , pour 3 simulations d'une loi GEV de paramètre $\xi = 1$



Pour k inférieur à 20, l'estimateur est volatil tandis qu'il devient biaisé pour des valeurs de k excédant 80.

La question du choix optimal de k se pose donc. Il est par ailleurs équivalent au choix du seuil u à partir duquel les observations sont considérées comme extrêmes.

Le critère le plus standard consiste à choisir le k qui minimise la moyenne du carré des erreurs :

$$E[(\xi_{k,n}^H - \xi)^2]$$

Le paramètre ξ est un paramètre de référence à partir duquel les erreurs vont être minimisées. Planchet donne des indications dans « Utilisation de la théorie des valeurs extrêmes dans le contexte de Solvabilité II ».

De notre côté, nous ne calculerons pas explicitement ce coefficient, pour ne pas surcharger les modélisations. Le choix de k se fera par une double appréciation graphique :

- Recours au théorème de linéarité des excès au-delà d'un seuil
- Visualisation du graphique de l'estimateur de Hill en fonction de k

Donnons le théorème de linéarité de l'espérance :

$$e(u) = E(X - u | X - u) = \frac{\beta + \xi u}{1 - \xi}$$

Avec $u \in D(\xi, \beta)$, $\xi < 1$

En construisant le graphique portant en ordonnée les

$$e_n(u) = \frac{1}{N_u} \sum_{i \in \Delta_n(u)} (X_i - u)$$

et en abscisse le seuil u , il est possible de juger visuellement à partir de quel seuil u la courbe devient une droite, c'est-à-dire $e_n(u)$ est linéaire par rapport à u .

Cette première impression est dans un deuxième temps confrontée à une visualisation du graphique de l'estimateur de Hill en fonction de u (ou k de manière équivalente). Il conviendrait de se trouver dans la zone de la courbe qui se situe après la zone de grande volatilité et avant celle de stabilisation où l'estimateur est biaisé. Cette zone de la courbe est approximativement la zone du k optimal, là où un compromis est fait entre biais et variance.

Le paramètre ξ est supposé estimé par l'estimateur de Hill. Il reste à estimer le paramètre $\beta(u) > 0$, pour tout u . Pour cela, une méthode envisageable consiste à maximiser la vraisemblance des excès observés au-delà du seuil u .

Ce paragraphe donne les généralités sur la méthode d'estimation par maximisation de la vraisemblance, puis propose une application dans le cadre de l'estimation de β .

iv Une méthode d'estimation de β par maximisation de la vraisemblance des observations

Il existe trois méthodes usuelles d'estimation de paramètre de lois. Les estimateurs qui en découlent peuvent être comparés selon des critères rigoureux. Les méthodes sont les suivantes :

1. La méthode des moments
2. La méthode des moindres carrés

3. La méthode du maximum de vraisemblance.

La méthode des moindres carrés est adaptée à des systèmes linéaires. Elle ne concurrence donc pas les deux autres.

Nous retiendrons l'estimateur du maximum de vraisemblance pour estimer β car cet estimateur surpasse l'estimateur des moments. En effet, il dispose de plus de qualités. Les critères de comparaison sous entendus sont les propriétés usuelles dont disposent les estimateurs.

Soit $\{P_\theta\}$ une famille de lois de probabilité continues sur $\mathfrak{R}$ et n un entier. Notons f_θ la densité de probabilité de la loi P_θ . On appelle vraisemblance associée à la famille $\{P_\theta\}$, la fonction qui à un n -uplet $(x_1, \dots, x_n)$ d'éléments de $\mathbb{C}$ et à une valeur θ du paramètre associe la quantité :

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

L'interprétation est la suivante : Considérons un échantillon théorique $(X_1, \dots, X_n)$ de la loi continue P_θ . Soit ε un réel strictement positif (petit). La probabilité que l'échantillon théorique $(X_1, \dots, X_n)$ ait une réalisation proche « à ε -près » de l'échantillon observé $(x_1, \dots, x_n)$ peut s'écrire :

$$P[X_1 \in [x_1 - \varepsilon/2, x_1 + \varepsilon/2] \text{ et } \dots \text{ et } X_n \in [x_n - \varepsilon/2, x_n + \varepsilon/2]] = \prod_{i=1}^n \int_{x_i - \varepsilon/2}^{x_i + \varepsilon/2} f_\theta(x) dx = \prod_{i=1}^n f_\theta(x_i) \varepsilon = \varepsilon^n L(x_1, \dots, x_n, \theta)$$

Estimer un paramètre par la méthode du maximum de vraisemblance, c'est proposer comme valeur de ce paramètre celle qui rend maximale la vraisemblance, à savoir la probabilité d'observer les données comme réalisation d'un échantillon de la loi P_θ . C'est un problème d'optimisation.

Si les conditions suivantes sont regroupées, elles permettent de conclure à $\hat{\theta}$ estimateur du maximum de vraisemblance de θ .

1. L est dérivable (ce qui n'est pas toujours le cas)
2. La dérivée première s'annule en un $\theta = \hat{\theta}$
3. En ce point, la dérivée seconde est négative

En pratique, la vraisemblance étant positive, on peut réduire la recherche et se contenter de maximiser le logarithme népérien de cette fonction.

On pose $Y_1, \dots, Y_{N_u}$ les excès au delà du seuil u , avec N_u (variable aléatoire) le nombre d'excès au-delà du seuil u indépendant des Y_i .

La vraisemblance de β au vue des observations $y_1, \dots, y_N$, avec N la réalisation de N_u est la quantité :

$$L(x_1, \dots, x_n; \beta) = \prod_{i=1}^N f_Y(y_i; \beta) = \prod_{i=1}^n \frac{1}{\beta} \left(1 + \hat{\xi}_{k,n}^H \frac{y_i}{\beta} \right)^{-1/\hat{\xi}_{k,n}^H - 1}$$

La log-vraisemblance s'en déduit immédiatement :

$$l(x_1, \dots, x_n; \beta) = -n \ln \beta - \left(\frac{1}{\hat{\xi}_{k,n}^H} + 1 \right) \sum_{i=1}^n \ln \left(1 + \frac{\hat{\xi}_{k,n}^H}{\beta} y_i \right)$$

Une résolution simple à l'aide d'un solveur donne :

$$\hat{\beta}_{MV} = \max_{\beta > 0} (l(x_1, \dots, x_n; \beta))$$

La loi de Pareto Généralisé étant estimée, il reste à connaître les quantiles de la loi.

1.2.3 Détermination des quantiles extrêmes

La formule analytique donne une expression exacte du quantile extrême (voir Annexe 8). L'alternative est de recourir à une simulation de la loi estimée, par inversion de la fonction de répartition. Le but est de disposer d'un très grand nombre de scénarios afin d'en déduire empiriquement, par le biais de la loi des grands nombres, les quantiles désirés de la variable aléatoire simulée. La génération de réalisations passe obligatoirement par deux étapes : d'abord générer des réalisations d'une variable

aléatoire de loi uniforme, puis en déduire une réalisation de la loi de probabilité, en inversant la fonction de répartition.

1.2.4 La méthode Peak Over the Threshold

i Protocole

La méthode POT (Peak Over the Threshold) est une méthode utilisée par les hydrologistes depuis plus de 30 ans, afin de modéliser les temps d'apparition de crues et leurs niveaux extrêmes.

Le modèle POT est usuellement formulé de la manière suivante :

Le processus d'apparition des excès au-delà d'un seuil est un processus de Poisson. En d'autres termes, c'est une suite de variables aléatoires indépendantes et identiquement distribuées, dont les temps d'attente suivent une loi exponentielle (voir loi de fréquence).

Les excès au-delà du seuil sont indépendants et sont distribués selon une loi de Pareto Généralisée.

Les temps d'occurrences des excès et les excès sont indépendants entre eux.

Signalons par ailleurs que ce modèle implique nécessairement que la loi des maxima suit une loi GEV.

Le calibrage d'un modèle POT sur une série d'observation nécessite donc les estimations suivantes :

1. Estimation de ξ et β de la loi GPD
2. Estimation du paramètre λ
3. Un choix de seuil u

Les propositions suivantes peuvent être faites, pour les trois problèmes qui viennent d'être posés

$\hat{\xi}_{k,n}^H$ est un estimateur de ξ et $\hat{\beta}_{MV}(\hat{\xi}_{k,n}^H)$ un estimateur de β

Le paramètre λ est estimé par maximisation de la vraisemblance des observations de temps d'attente (voir loi de fréquence)

Dans l'application qui suit, une extension du modèle POT sera proposée dans le sens où le processus de Poisson sera calibré sur les séismes faisant une victime ou plus et non uniquement sur les séismes les plus meurtriers.

Le processus de Poisson estimé doit par ailleurs ensuite être testé. Un diagramme quantile-quantile de la loi exponentielle, un test de Kolmogorov et de Mann-Whitney permettront de tester la bonne adéquation du processus à l'historique des survenances.

ii Loi de fréquence

Ce paragraphe s'intéresse à la modélisation des temps d'apparition d'excès. Il donne donc la définition d'un processus de Poisson et les propriétés qui lui sont propres.

- Loi exponentielle

La loi exponentielle de paramètre $\lambda > 0$ est une loi continue avec densité donnée par :

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{si } t \geq 0 \\ 0 & \text{si } t \leq 0 \end{cases}$$

Il s'agit bel et bien d'une densité de probabilité puisqu'on a :

$$f(t) \geq 0 \quad \forall t \in \mathbb{R} \quad \text{et} \quad \int_{-\infty}^{+\infty} f(t) dt = 1$$

- Loi de Poisson

La loi de Poisson de paramètre $\nu \geq 0$ est la loi discrète avec fonction de masse suivante donnée par :

$$p(k) = \begin{cases} e^{-\nu} \frac{\nu^k}{k!} & \text{si } k \in \{0, 1, 2, \dots\} \\ 0 & \text{si } k \notin \{0, 1, 2, \dots\} \end{cases}$$

Il s'agit bien d'une probabilité sur l'ensemble des entiers non négatifs puisque d'une part on a $p(k) > 0$ pour tout k entier positif et d'autre part $\sum_{k=0}^{+\infty} e^{-\nu} \nu^k / k! = 1$

- Processus de Poisson

Définition d'un processus de dénombrement

Considérons $T_1, T_2, \dots$ des variables aléatoires satisfaisant $P(T_j > 0) = 1$ pour tout $j \geq 1$. Posons $S_0 = 0$ et pour tout n supérieur ou égal à 1, posons $S_n = \sum_{j=1}^n T_j$. Enfin, pour tout $t \geq 0$, posons $N(t) = \max(n \geq 0 : S_n \leq t)$. Le processus aléatoire $(N(t); t \geq 0)$ s'appelle un processus de dénombrement

Remarque

Les variables aléatoires $T_1, T_2, \dots$ sont appelées durée de vie ou temps d'attente suivant le phénomène étudié.

Définition d'un processus de renouvellement

Un processus de renouvellement est un processus de dénombrement pour lequel les durées de vie sont des variables aléatoires indépendantes et identiquement distribuées.

Définition d'un processus de Poisson

Un processus de Poisson avec intensité λ est un processus de renouvellement dont la distribution des durées de vie est la loi exponentielle(λ).

Les éléments théoriques utilisés dans l'application ont tous été présentés. Le travail réalisé par la suite consiste à mettre en application le modèle POT, c'est-à-dire à construire des maquettes Excel

VBA pour estimer les paramètres nécessaires à une complète connaissance de l'intensité meurtrière et de la fréquence d'apparition des séismes au Japon.

2 *Lois de mortalité due aux séismes*

Cette partie tente de proposer des pistes d'estimation de la loi de mortalité annuelle due aux séismes. Un séisme est meurtrier lorsqu'il cause la mort d'au moins une personne. Les lois estimées dans cette partie ne prennent pas en compte les années sans victimes. Ce sont des lois conditionnelles à l'événement : « l'année est marquée par un séisme meurtrier ».

2.1 Choix de l'échantillon

Avant de procéder à des calibrages de lois, il est d'abord nécessaire de regrouper un historique à partir duquel des ajustements statistiques a un sens.

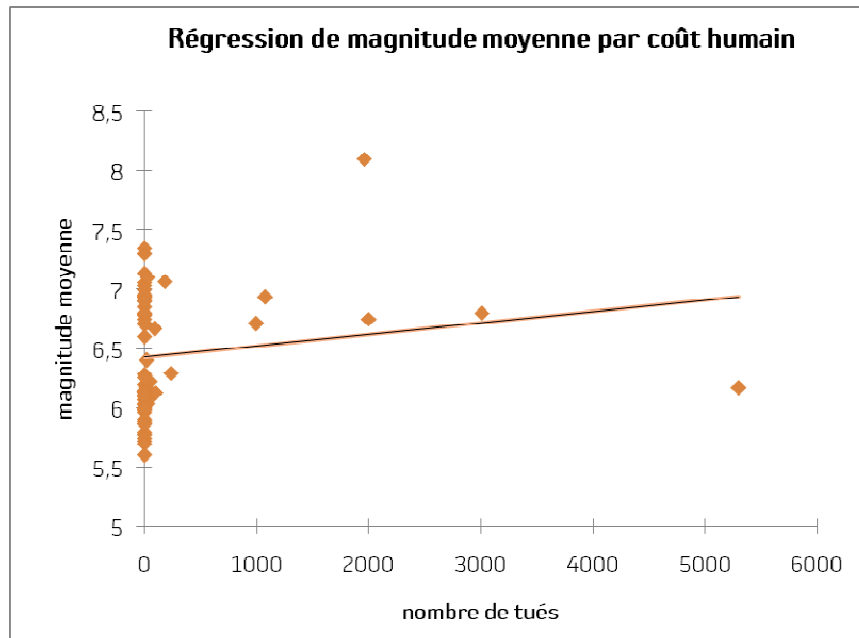
Deux critères doivent être pris en compte :

- L'étendue des observations : les séismes très meurtriers étant des événements très rares dont la probabilité d'occurrence est de l'ordre d'une fois tous les deux siècles, il est nécessaire de remonter le plus loin possible dans le passé.
- La nature des observations : soit l'historique centennal des nombres annuels de tués de l'EMDAT, soit sur des données d'intensité (sur l'échelle de Richter) des séismes passés.

Concernant le premier critère, l'étendue retenue est de 100 ans environ car 1900 est en général la date à partir de laquelle les historiques d'intensité et de mortalité commencent. Cette durée excède largement la période de prévision. Toutefois, le processus sous-jacent aux séismes, à savoir la tectonique des plaques, est un processus très lent. De ce fait, un historique de 100 ans peut paraître bien dérisoire.

Concernant le deuxième critère, les données retenues sont des données de mortalité. En effet, les données d'intensité sont éliminées car elles présentent un inconvénient majeur : même si des historiques très complets peuvent être obtenus, notamment sur le site de l'USGS (United States Geological Survey), il est très difficile de trouver une corrélation simple entre les intensités et les

nombre de victimes donnés par le CRED. En effet, la régression de la variable « magnitude moyenne » par la variable « nombre de victimes » donne un R^2 égal à 0,021.



2.2 Modélisation des mortalités extrêmes par la loi de Zipf

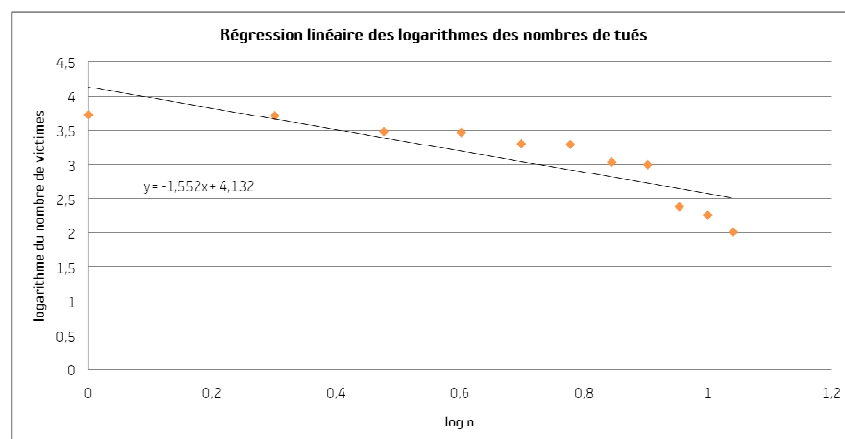
Selon Gordon Woo, célèbre mathématicien spécialiste des événements extrêmes, la fiabilité des données de mortalité est très douteuse car les chiffres sont souvent exagérés ou à l'inverse parfois atténués (dans «The mathematics of natural catastrophes»). Selon lui, l'obtention de données fiables passent par une compilation optimale de sources variées (humanitaires, gouvernementales...).

En conséquence, les mortalités affichées dans l'historique de l'EMDAT sont scindés en deux :

- Les années marquées par pas ou peu de victimes sont conservées telles quelles. Les observations de mortalité relatives à ces années constituent la loi des mortalités faibles.
- Les années marquées par un nombre considérable de victimes sont considérées comme douteuses : la loi de Zipf est estimée à partir des ces observations

En conséquence :

- Les années pour lesquelles le nombre de décès est compris entre 1 et 103 sont conservées.
- Par contre, pour les mortalités qui excèdent 103 victimes, les paramètres de la loi de Zipf sont estimés à partir de ces données de mortalité, en prenant de plus la précaution d'éliminer l'année 1923 pendant laquelle 143000 personnes ont perdu la vie dans des séismes. En effet, compte tenu des progrès réalisés dans le domaine des installations parasismiques, la répétition de ce genre d'événements est considérée comme impossible. Le graphique suivant est obtenu à l'aide du complément Excel XLSTAT : le logarithme des décès est placé en ordonnée et le logarithme du rang n est placé en abscisse.



La régression linéaire précédente nous donne une estimation des paramètres c et α de la loi de Zipf. Les paramètres $C = \exp(c)$ et β en découlent de manière immédiate. De ce fait, la loi des mortalités extrêmes est entièrement estimée.

Le tableau suivant en découle :

probabilité	mortalité
1,00E-02	[103;182]
3,74E-03	[183;241]
1,13E-02	[242;998]
3,87E-04	[999;1083]
2,27E-03	[1084;1961]
6,15E-05	[1962;2000]
1,05E-03	[2001;2925]
6,74E-05	[2926;3008]
1,08E-03	[3009;5131]
5,33E-05	[5132;5297]

Les probabilités inscrites dans le tableau ont été obtenues en multipliant la probabilité donnée par la loi de Zipf par le coefficient 0,03. Ce coefficient est la probabilité d'avoir un séisme causant la mort de plus de 103 personnes et a été directement déduite de l'historique en se plaçant toutefois dans une vision prospective optimiste.

Le choix du seuil de 103 victimes est pour l'instant un seuil arbitraire, qui est par ailleurs équivalent au choix du nombre d'observations considérées comme extrêmes par rapport au nombre d'observations total, est un problème auquel la théorie des extrêmes apporte des réponses. Il sera abordé dans la partie suivante qui tentera de modéliser les séismes extrêmes par une loi de Pareto Généralisée, très adaptée pour les événements qui surviennent tous les 30 ou 50 ans, voire de manière encore moins fréquente.

La loi de la mortalité est enfin complétée pour prendre en compte tous les événements. Les probabilités d'occurrence de séismes causant la mort de moins de 103 personnes sont estimées directement à partir de l'historique de l'EMDAT, en donnant toutefois volontairement plus de poids pour les petits séismes. Cette loi est conditionnelle à l'événement : « un séisme meurtrier survient.

probabilité	mortalité
8,40E-01	entre 1 et 50 victimes
3,60E-02	[50;93]
4,00E-03	[94;102]
4,00E-02	[103;182]
1,49E-02	[183;241]
4,52E-02	[242;998]
1,55E-03	[999;1083]
9,09E-03	[1084;1961]
2,46E-04	[1962;2000]
4,19E-03	[2001;2925]
2,70E-04	[2926;3008]
4,31E-03	[3009;5131]
2,13E-04	[5132;5297]

Cette loi est ensuite simulée pour obtenir une distribution empirique

Les simulations de Monte Carlo donnent la fonction de répartition du coût humain (mortalité) cumulé sur 30 ans. Le nombre de simulations effectuées est de 50 000 et le nombre d'années de prévision de 30 ans.

Les quantiles principaux sont les suivants :

quantile	ordre du quantile
479	75%
561	80%
836	85%
1337	90%
3232	99%

2.3 Calibrage d'une loi de Pareto Généralisée

2.3.1 Les données

La base de données du CRED permet de construire l'historique des mortalités annuelles dues à des séismes depuis 1900. Le nombre exact d'observations est de $n=108$. L'échantillon est noté :

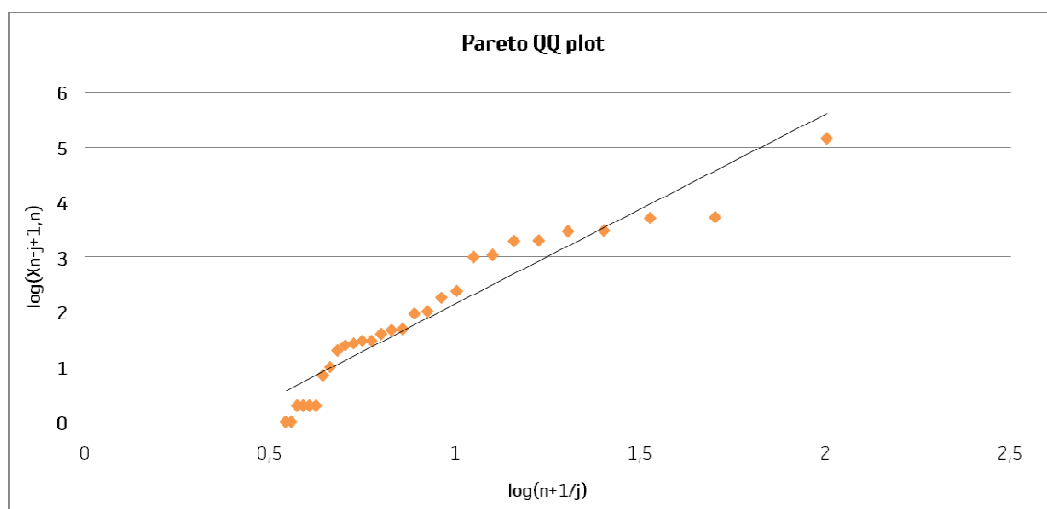
$$X_{1,108} \leq \dots \leq X_{108,108}$$

Les observations sont considérées comme indépendantes et identiquement distribuées.

2.3.2 Signe de l'indice des valeurs extrêmes : recours au Pareto QQ plot

La base de données du CRED permet de construire l'historique des séismes depuis 1900.

Un Pareto QQplot consiste à porter en ordonnée $\log(X_{n-j+1}, n)$ et en abscisse $\log(n+1/j)$.



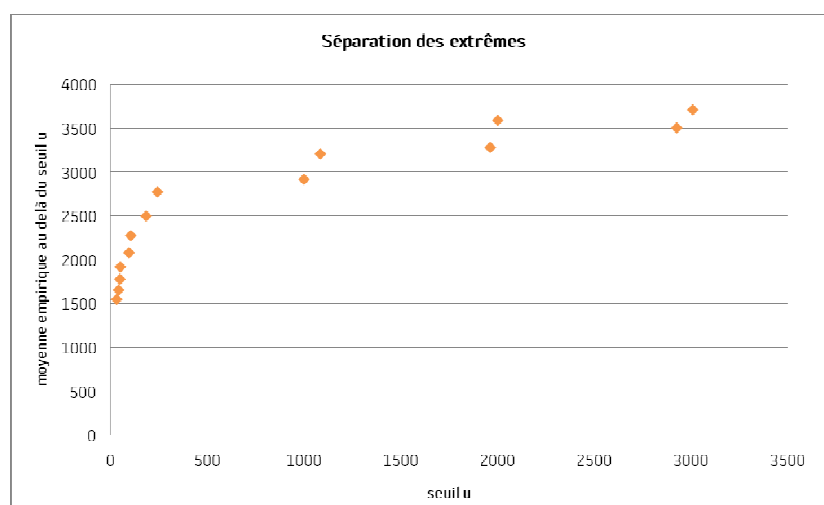
Si les points s'alignent sur une droite, alors F se situe dans le domaine d'attraction de Fréchet. C'est-à-dire dans le domaine d'attraction d'une loi GEV de paramètre $\xi > 0$. Le graphique permet de conclure à un alignement. Cette phase de détection de l'indice des valeurs extrêmes est important puisque la façon dont ξ est estimé par la suite est valable seulement si ξ est positif, c'est-à-dire seulement si F appartient au domaine d'attraction de Fréchet.

2.3.3 Séparation des extrêmes

Le jugement est uniquement visuel, basé sur le graphique qui porte la moyenne empirique au-delà du seuil u en ordonnée et le seuil u en abscisse.

valeur de u	moyenne empirique au-delà du seuil u
3008	3710
2925	3504
2000	3590
1961	3280
1083	3207
998	2915
241	2770
182	2496
103	2272
94	2076
49	1915
47	1771
40	1649
30	1542

Ce tableau se traduit graphiquement :



Il apparaît que le seuil de mortalité à partir duquel les points s'alignent sur une droite est le nombre de 241 victimes. La loi GPD sera donc estimée à partir des excès au-delà de ce seuil.

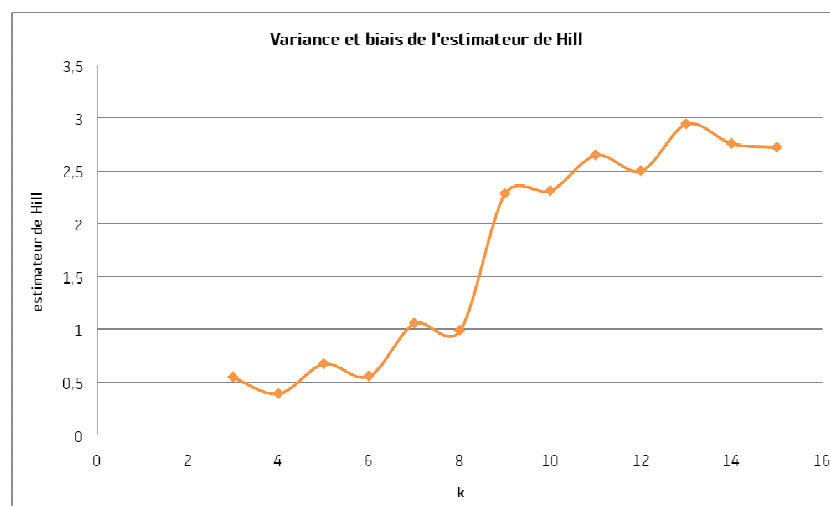
2.3.4 Estimation de l'indice des valeurs extrêmes

La première étape dans le calibrage d'une loi GDP sur les excès retenus dans le paragraphe précédent est d'estimer ξ .

Pour cela, une gamme d'estimateurs de Hill est calculée pour une gamme de nombres k de mortalités extrêmes retenues. Si $k = 1$, une seule mortalité, la plus extrême, est retenue. Si $k = 2$, seul $X_{108,108}$ et $X_{107,108}$ sont retenus, etc.

k	estimateur de Hill
3	0,550
4	0,395
5	0,676
6	0,561
7	1,061
8	0,991
9	2,288
10	2,315
11	2,652
12	2,503
13	2,946
14	2,761
15	2,725

Ce tableau se traduit graphiquement :



L'utilité est de constater si le seuil $u=241$ équivalent à $k=9$ se trouve bien dans la zone optimale du graphique de $\hat{\xi}_{k,108}^H$ en fonction de k .

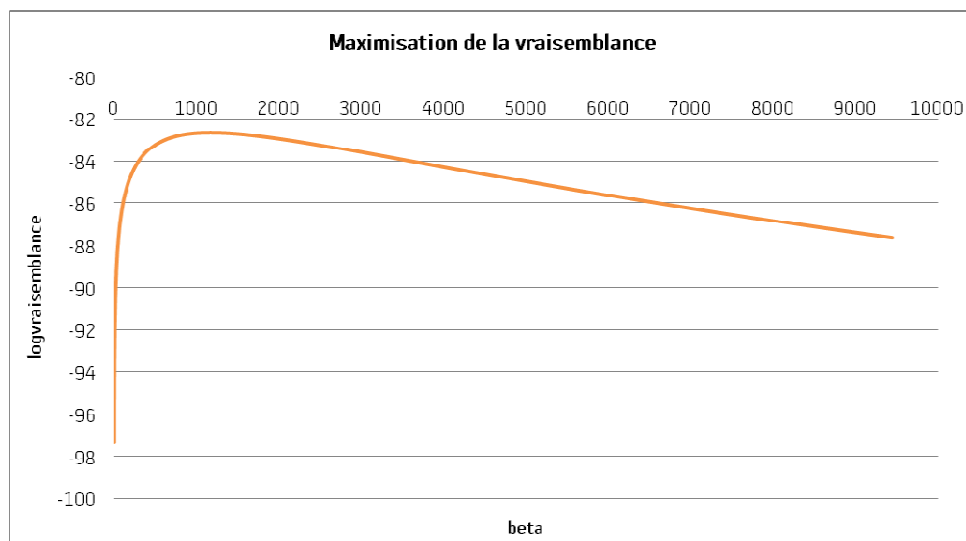
2.3.5 Calibrage de la loi GPD

La seconde étape dans l'ajustement de la loi de Pareto Généralisée est l'estimation de β , que l'on note

$$\hat{\beta}_{MV}(\hat{\xi}_{k,n}^H)$$

Il s'agit de maximiser la vraisemblance des excès observés en fonction du paramètre β , seul paramètre variable car ξ a déjà été estimé.

Le logarithme de la vraisemblance porté en abscisse et la variable β en ordonnée donne :



Le solveur d'Excel permet de trouver le β tel que la logvraisemblance est maximisée :

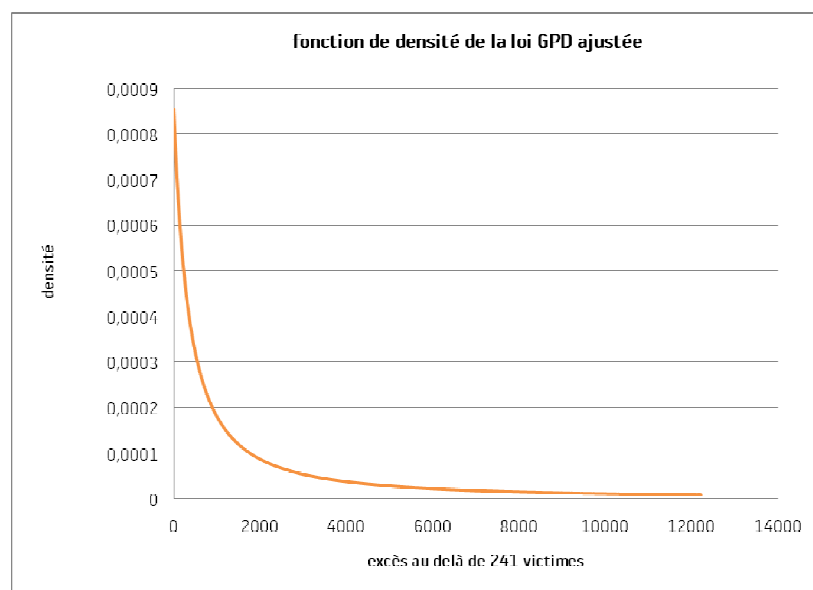
$$l(x_1, \dots, x_n; \beta) = -n \ln \beta - \left(\frac{1}{\hat{\xi}_{k,n}^H} + 1 \right) \sum_{i=1}^n \ln \left(1 + \frac{\hat{\xi}_{k,n}^H}{\beta} y_i \right)$$

Finalement la loi de Pareto Généralisé est complètement estimée et les excès de mortalité caractérisés.

En résumé, les paramètres obtenus sont les suivants :

beta	estimateur de Hill
1169	2,29

La fonction de densité estimée peut être représentée graphiquement.



2.3.6 Conclusion

La variable aléatoire prend manifestement des valeurs que l'historique des mortalités ne donnent pas. C'est précisément le but d'une modélisation stochastique : permettre, par des simulations, de générer des événements probables et qui ne sont jamais parvenus par le passé.

Une loi d'intensité définitive est construite en redimensionnant la loi d'intensité estimée à l'intervalle borné : $[0 ; 5057]$. Ce choix est motivé par le fait qu'il n'est pas envisageable, compte tenu des progrès parasismiques réalisés, que les pires mortalité connues soit dépassées.

La loi des mortalités non excessives, c'est-à-dire inférieures à 241 victimes est obtenue à l'aide de l'historique, pour les mortalités comprises entre 1 et 103 victimes, et suivant la loi de Zipf, pour les mortalités comprises entre 103 et 241 victimes.

La loi est donc au final une loi discrète. Elle est représentée par le tableau suivant :

loi d'intensité		
événement	probabilité	probabilité cumulée
entre 1 et 50 victimes	0,84000	0,8400
[49;94]	0,03600	0,8760
[95;103]	0,00400	0,8800
[104;182]	0,04001	0,9200
[183;241]	0,01494	0,9349
[242;998]	0,03291	0,9679
[999;1083]	0,00189	0,9697
[1084;1961]	0,01288	0,9826
[1962;2000]	0,00040	0,9830
[2001;2925]	0,00726	0,9903
[2926;3008]	0,00050	0,9908
[3009;5131]	0,00874	0,9995
[5132;5297]	0,00048	1

Afin de caractériser complètement le risque de séismes au Japon, une loi de fréquence est estimée.

3 *Loi de fréquence de survenance de séismes meurtriers*

3.1 La loi binomiale

Supposons que l'on veuille modéliser la fréquence d'apparition d'un séisme meurtrier sur 30 ans. Cette période est par exemple le nombre d'années couramment utilisé pour le calcul de la valeur intrinsèque d'un portefeuille.

Une réalisation x d'une variable X est générée suivant une loi Binomiale(30,0.25). Cela donne un nombre d'années touchées par un séisme meurtrier.

Cela est équivalent à dire que, chaque année et pendant trente ans, un séisme meurtrier se réalise avec une probabilité de 0.25, les réalisations étant indépendantes d'une année sur l'autre.

Le paramètre 0,25 est issu de l'historique en constatant que 25 années sur 100 sont marquées par un sinistre faisant au moins 1 victime.

3.2 Ajustement d'un processus de Poisson sur le processus de survenance des séismes meurtriers au Japon

Le temps d'attente entre deux séismes meurtriers est modélisé par une loi exponentielle car le processus de renouvellement $(N(t); t \geq 0)$ est supposé être un processus de Poisson, avec $N(t)$ le nombre de séismes entre la date d'aujourd'hui et la date t .

Le paramètre λ est estimé en maximisant la vraisemblance des observations d'attente. Elles varient entre 1 et 19 ans, sur l'historique de données centennal.

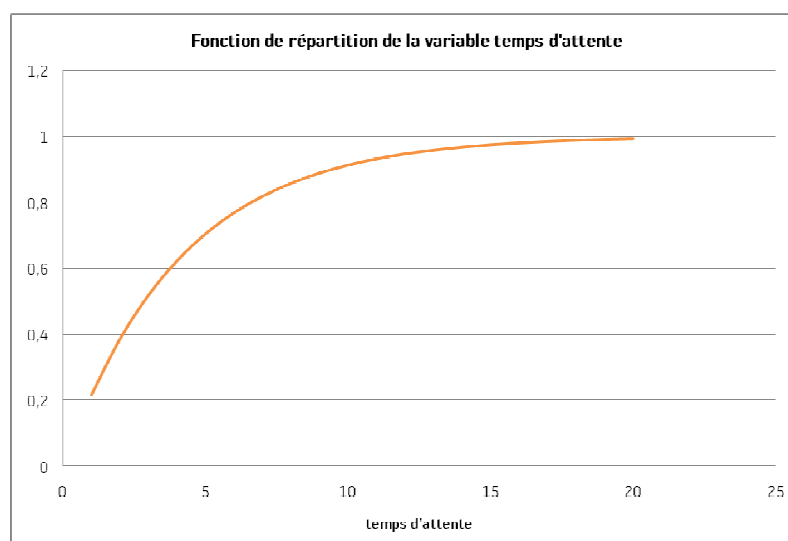
Un calcul simple permet de montrer que l'estimateur du maximum de vraisemblance $\hat{\lambda}_{MV}$ de λ s'écrit

$$\hat{\lambda}_{MV} = \frac{1}{\frac{1}{N} \sum_i T_i}$$

Avec T_i le i ème temps d'attente et N le nombre de séismes meurtriers observé sur l'historique centennal.

Numériquement $\hat{\lambda}_{MV}=0,243$.

La fonction de répartition estimée est la suivante, en supposant que le temps d'attente est supérieur ou égal à 1.



Des tests simples permettent de vérifier que les variables T_i proviennent bien de la loi exponentielle estimée. Les tests utilisés sont les suivants. Ils sont réalisés par le logiciel XLStat, au seuil alpha de 0,05 :

Le test de Kolmogorov-Smirnov qui teste H_0 : l'échantillon suit une loi exponentielle de paramètre $\hat{\lambda}_{MV}=0,243$

Le test bilatéral de Kolmogorov-Smirnov sur deux échantillons qui teste H_0 : les échantillons formés des quantiles de la loi exponentielle et des quantiles observés n'ont pas une distribution significativement différente

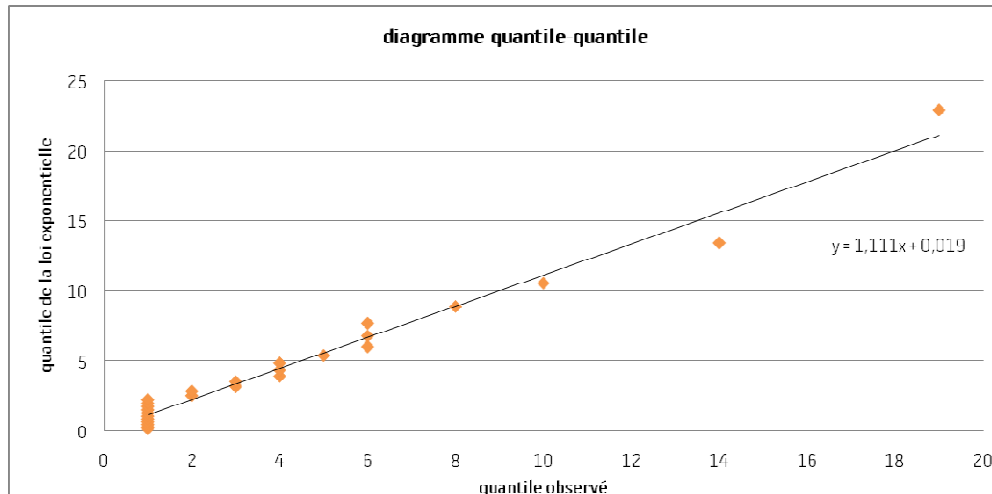
Le test bilatéral de Mann-Whitney qui teste H_0 : la différence de position entre les deux échantillons n'est pas significativement différente de 0. Autrement dit, il s'agit de tester si les échantillons formés des quantiles de la loi exponentielle et des quantiles observés sont issus d'une même et unique distribution.

Les résultats des tests donnent des P-value suffisamment grande pour ne pas rejeter H_0 .

Les tests effectués sont issus des données suivantes :

rang	quantiles de la loi exponentielle	temps d'attente ordonnés
1	0,161	1
2	0,329	1
3	0,505	1
4	0,687	1
5	0,879	1
6	1,080	1
7	1,291	1
8	1,513	1
9	1,749	1
10	1,998	1
11	2,264	1
12	2,548	2
13	2,853	2
14	3,182	3
15	3,540	3
16	3,932	4
17	4,366	4
18	4,851	4
19	5,400	5
20	6,035	6
21	6,785	6
22	7,703	6
23	8,887	8
24	10,556	10
25	13,408	14
26	22,884	19

La représentation graphique de cette bonne adéquation est le diagramme quantile-quantile de la loi exponentielle (de paramètre $\hat{\lambda}_{MV}=0,243$).



Le R^2 de la régression linéaire est très proche de 1. Ce qui montre la qualité de l'ajustement.

3.3 Simulation de la loi de fréquence sur 30 ans

Simuler un processus de Poisson passe par la simulation d'une variable exponentielle. Pour cela, la fonction de répartition de la loi exponentielle de paramètre $\hat{\lambda}_{MV}=0,243$ est inversée, en prenant la précaution de ne générer que des réalisations supérieures ou égales à 1. Cette restriction permet de ne pas générer deux temps d'attente successifs inférieurs à 1, c'est-à-dire de ne pas générer de scénarios où une même année serait marquée par deux séismes meurtriers.

Il est possible de trier les scénarios par gravité. Un scénario est considéré comme plus grave qu'un autre si le nombre de séismes sur 30 ans du premier est plus grand que le nombre du second.

4 Simulations : Intérêts et possibilités

Les deux parties précédentes ont permis d'estimer deux lois :

- La loi d'intensité (caractérisant le nombre de tués au Japon) donnée par le tableau suivant. Cette loi est celle qui sera utilisée dans les applications sur le portefeuille.

loi d'intensité		
événement	probabilité	probabilité cumulée
entre 1 et 50 victimes	0,84000	0,8400
[49;94]	0,03600	0,8760
[95;103]	0,00400	0,8800
[104;182]	0,04001	0,9200
[183;241]	0,01494	0,9349
[242;998]	0,03291	0,9679
[999;1083]	0,00189	0,9697
[1084;1961]	0,01288	0,9826
[1962;2000]	0,00040	0,9830
[2001;2925]	0,00726	0,9903
[2926;3008]	0,00050	0,9908
[3009;5131]	0,00874	0,9995
[5132;5297]	0,00048	1

- Le processus de survenance de séismes est un processus de Poisson dont le paramètre a été estimé en maximisant la vraisemblance des observations.

$$\hat{\lambda}_{MV}=0,243.$$

Les estimations des ces lois sont très critiquables. En effet, elles sont réalisées à partir d'un échantillon de très petite taille. Une amélioration consisterait à disposer de données mensuelles. Le nombre d'observations serait en effet multiplié par 12.

De plus, les lois estimées valent pour le pays entier et non pour le portefeuille client. Des hypothèses fortes devront être donc formulées notamment sur la dispersion des assurés sur le territoire japonais, afin de justifier les prorata qui seront effectués dans le cadre des applications sur les données du portefeuille.

Dans le cadre d'une modélisation interne du risque de séismes, les estimations qui viennent d'être réalisées doivent donc être accompagnées de multiples précautions, sans quoi ces calibrages seraient irréalistes et irrecevables.

Si des précautions suffisantes sont prises, des simulations de Monte Carlo de ces lois peuvent être envisagées. Ces simulations consistent, sur une durée déterminée de plusieurs années, à simuler des scénarios probables. Elles peuvent par exemple être décomposées en trois étapes :

- Générer un scénario de survenance (en simulant le processus de Poisson), c'est-à-dire déterminer les années touchées par un séisme meurtrier, sur un horizon prédéterminé.
- Pour les années touchées, simuler la loi de mortalité.
- Répéter les deux premiers points plusieurs milliers de fois.

Ce protocole n'est évidemment pas figé. Il est simplement un exemple d'utilisation des lois estimées. Dans le chapitre 5, un protocole différent est proposé. Il met également en évidence des possibilités d'applications pratiques intéressantes, notamment dans le cadre du calcul de la valeur intrinsèque.

V Le risque de canicule en France

La gestion du risque de canicule, d'une manière analogue au risque de séisme, passe d'abord par l'estimation de deux lois :

- La loi de surmortalité lors d'une canicule.
- La loi de fréquence de survenance des vagues de chaleur, pour une région ou un pays donnée (ici la France)

Les parties qui suivent tentent d'approcher ces deux lois. Dans la première (Travaux préliminaires), nous nous demandons si les canicules sont bien des événements réellement catastrophiques. Dans les deux parties suivantes, les estimations proprement dites sont effectuées.

1 Travaux préliminaires

1.1 Problématique

Les questions qui se posent sont les suivantes :

- Comment définir une canicule ?
- Une vague de chaleur seule entraîne-t-elle une surmortalité ? Doit-elle être accompagnée d'une forte pollution, d'une faible humidité ?

- La culture du risque dans les pays tempérés (qui connaissent depuis le début du siècle des épisodes caniculaires longs et intenses) sera-t-elle suffisante à l'avenir, malgré la répétition annoncée d'événements caniculaires à cause du changement climatique?
- Les victimes sont-elles des personnes qui seraient décédées de toute façon dans les semaines ou mois à venir. Doit-on conclure à une « illusion de surmortalité » ?
- Comment prévoir les canicules? Autrement dit, comment les modéliser?

1.2 Définition

Une difficulté préalable réside dans l'ambiguïté de la notion de vague de chaleur. Ni les dictionnaires usuels, ni l'Organisation Météorologique Mondiale (OMM) ne sont d'un grand secours en précisant qu'il s'agit d'un réchauffement important de l'air, d'une période caractérisée par des températures anormalement élevées ou d'une invasion d'air chaud.

Les météorologues français et américains définissent une canicule en termes absolus, en fixant un seuil à 30°C et 32°C.

Les Britanniques définissent une « Heat Wave » en termes relatifs, lorsque le thermomètre indique au moins 4°C au dessus de la moyenne trentennale du lis et du mois.

Afin de disposer de notre propre définition, il est nécessaire de disposer de sources qui d'une part proposent des définitions mais qui d'autre part les justifient. Les études de référence utilisées sont :

- Institut national de la santé et de la recherche médical (INSERM) : surmortalité liée à la canicule d'août 2003 – Rapport d'étape (25 septembre 2003)
- Jean Pierre Besancenot, Centre Universitaire d'épidémiologie de population de Dijon : vagues de chaleur et mortalité dans les grandes agglomérations urbaines.

L'INSERM propose la définition suivante : un jour de grande chaleur est un jour au cours duquel la température maximale est supérieure ou égale à 35°C et la température minimale est supérieure ou égale à 20°C.

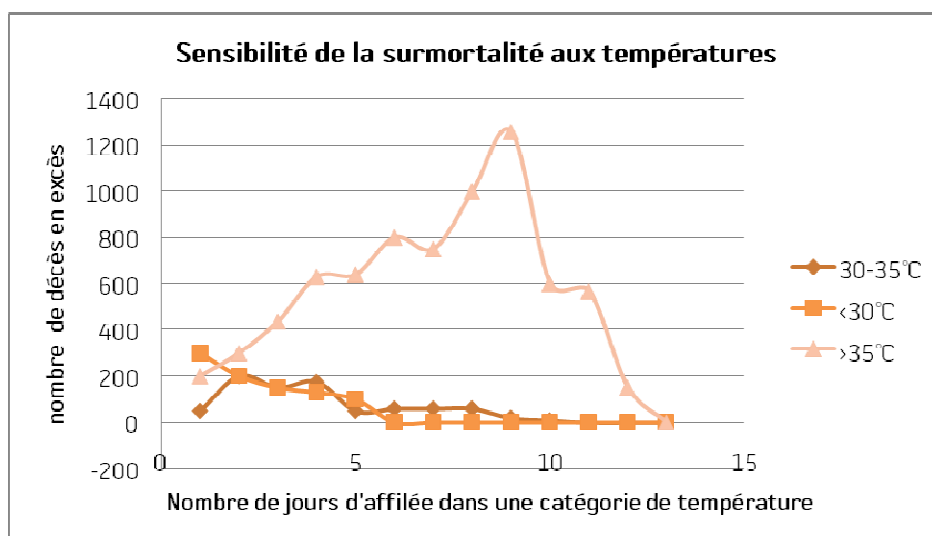
La durée peut varier. Plus elle est grande, plus les classes d'âge touchées sont jeunes. De ce fait, la loi de surmortalité proposée plus loin sera fonction de la durée (en jours) de la canicule.

Le seuil de 35°C trouve une explication simple. En effet, une surmortalité significative ne sera pas observée si le seuil est trop bas. L'INSERM a comparé les surmortalités observées chez les sujets de 75 ans ou plus (c'est-à-dire les plus sensibles), lors de la canicule d'août 2003, en fonction du nombre de jours d'affilée où le département a connu une température de plus de 35°C, de 30 à 35°C, ou de moins de 30°C. Il s'avère qu'une surmortalité significative est observée seulement si le département connaît une température de plus de 35°C. Le choix d'un seuil minimum s'explique par le fait qu'une canicule doit obligatoirement être caractérisée par une absence de « repos » : les températures restent hautes et les nuits ne sont pas suffisamment longues pour que le sol ait le temps de se refroidir suffisamment. A titre d'exemple, les fortes chaleurs de 1983 dans le Midi de la France où seules les régions marquées par une absence de rémission nocturne (températures minimales élevées) ont été réellement endeuillées.

Le graphique ci-dessous donne les surmortalités chez les sujets de 75 ans ou plus, pour la canicule de 2003, en fonction du nombre de jours d'affilée avec des températures maximales supérieures à 35°C, entre 30°C et 35°C et inférieur à 30°C.

Pour cela, trois départements ont été choisis, correspondant à chacune des courbes :

- Un département où la température maximale a dépassé 35°C pendant 13 jours d'affilée
- Un département où la température maximale n'a pas dépassé 30°C pendant 13 jours d'affilée
- Un département où la température maximale était comprise entre 30°C et 35°C.



Finalement, la définition choisie reste dans le cadre national, car Météo France définit également les vagues de chaleur de cette manière, mais en ne fixant pas les mêmes seuils suivant les départements. Les régions du Sud de la France connaissent une canicule si les températures dépassent un certain seuil qui est évidemment plus élevé que le seuil fixé pour les régions du nord, compte tenu de l'habitude et des prédispositions des habitants du Sud à subir de fortes chaleurs. De plus, Météo France fixe généralement une durée minimum, qui varie également suivant les départements. (3 jours en moyenne).

1.3 Caractère meurtrier d'une vague de chaleur

La question peut également se poser de la manière suivante : Comment identifier les canicules qui sont susceptibles d'accroître la mortalité ? Une hausse de température engendre-t-elle inévitablement une surmortalité, ou doit-elle survenir conjointement avec d'autres événements, tels qu'un pic de pollution ou une forte baisse de l'humidité ?

La température n'est en effet pas toujours un indicateur suffisamment fidèle des efforts imposés à l'organisme et des risques d'accidents pathologiques qui s'ensuivent. D'autres éléments du climat interfèrent pour rendre la chaleur plus ou moins « supportable » :

- Vitesse du vent
- Humidité de l'air
- Niveau de la pollution atmosphérique

L'exemple des vagues de chaleur ayant frappées Athènes en 1987 et 1988 permet d'illustrer le fait que la surmortalité peut être vraiment indépendante des températures enregistrées, et même des indices de confort. En effet, les températures apparentes enregistrées durant le premier été n'étaient jamais alarmistes (température chiffrée). En effet, le DI n'a jamais dépassé 32°C et laissait supposer qu'aucune mesure particulière de protection civile ne s'imposait. Toutefois, 2 000 personnes ont périés pendant l'épisode. L'explication se trouve sans doute dans l'interférence d'une chaleur forte et d'une pollution atmosphérique majeure car l'année suivante, seulement 28 décès ont été enregistrés, alors que la vague de chaleur était « pire » que l'année précédente mais la pollution réduite de manière draconienne par les pouvoirs publics.

Il faut donc connaître, s'il survient une vague de chaleur de grande ampleur, la sensibilité de la surmortalité à la qualité de l'air et au niveau d'humidité.

L'expérience de 2003 permet de trancher, dans le cadre de notre modélisation. Une vague de chaleur seule, c'est-à-dire indépendamment des facteurs pollution et humidité, peut engendrer une surmortalité significative, tandis que le niveau de pollution accentue vraisemblablement le caractère meurtrier. C'est ce qui expliquerait le caractère particulièrement meurtrier de la vague de chaleur de 2003 en Ile de France.

Une surmortalité est considérée comme significative lorsque le ratio O/E (mortalité observée sur la mortalité attendue) est supérieur ou égal à 1,1.

Des régions comme la Corse ou l'Auvergne, peu urbanisées et donc relativement peu polluées, affichent des surmortalités très importantes pendant le mois d'août 2003 (ratio O/E entre 1,1 et 2). De plus, dans le cadre d'une modélisation d'une canicule, il semble judicieux de ne pas trop paramétrer le modèle. Se contenter uniquement de données de températures n'est pas, en se référant en chiffres de l'INED, extrêmement choquant.

1.4 La surmortalité engendrée par les canicules n'est-elle qu'une illusion ?

Une canicule emporte-t-elle des vies destinées à durer ou hâte-t-elle la mort de personnes en fin de vie ?

- Dans le premier cas, une canicule est réellement une catastrophe. Du point de vue de l'assureur, la survenance d'un tel évènement engendre une sur-sinistralité réelle.
- Dans le second cas, l'assureur ne voit pas de catastrophe dans le sens où il n'y a pas de sur-sinistralité. Les prestations nouvelles engendrées par le décès de personnes assurées auraient de toute façon été déclenchées quelques semaines ou mois après la canicule.

L'Institut National d'Études Démographiques (INED) répond à cette question pour le cas de la canicule d'août 2003. La conclusion de cet organisme sera généralisée pour les futures canicules qui surviendront potentiellement dans les décennies à venir, en France principalement.

L'INED répond « sans l'ombre d'un doute » sur le caractère réellement catastrophique de la canicule de 2003. Elle a emporté des vies destinées à durer encore plusieurs années, conformément à l'espérance de vie de chaque classe d'âge.

Nous ne nous étendrons pas sur les explications données par l'INED. Elles sont considérées comme sérieuses, compte tenu du statut de cet organisme.

2 Proposition d'une loi de surmortalité

Cette loi ne résultera d'aucune simulation. Elle met en liaison la durée de la canicule, les classes d'âges et la surmortalité exprimée par le rapport O/E. Elle a été mise en évidence par l'INSERM suite à la canicule de 2003. Une tentative de projection sera ensuite effectuée pour rendre compte des possibilités d'aggravation de cette loi.

2.1 Loi de surmortalité de 2003

Le tableau suivant donne la loi de mortalité en mettant en relation la surmortalité et le nombre de jours où la température a été au dessus de 35°C et la température minimale d'au moins 20°C (du 1^{er} au 20 août).

en 2003 : loi de surmortalité													
	inférieur à 45 ans			45-74 ans			75 ans et plus			Tous âges			
	Nombre de départements	O	E	O/E	O	E	O/E	O	E	O/E	O	E	O/E
inférieure ou égal à 1 jour	38	678	646	1,05	4034	3543	1,14	9414	6360	1,48	14126	10548	1,30
entre 2 et 5 jours	26	493	441	1,12	3018	2323	1,30	7527	4493	1,68	11038	7257	1,50
supérieure ou égal à 5 jours	31	677	620	1,09	4189	2925	1,43	11591	5468	2,12	16457	9013	1,80

2.2 Projection de la loi de surmortalité

Il n'est pas évident que cette loi puisse être unanimement considérée comme une loi de référence pour les futures canicules qui surviendront potentiellement dans les décennies à venir, en France principalement.

Il convient d'énumérer les facteurs qui pourraient soit aggraver la surmortalité soit la réduire, afin de conclure sur la vraisemblance de cette loi, à l'avenir (dans le cas de la France pour être prudent).

Le facteur aggravant est le changement climatique actuel. D'après le Groupe d'Expert Intergouvernemental sur l'Evolution du Climat (GIEC), il se caractérise, et se caractérisera de plus en plus, vraisemblablement, dans les pays tempérés, par une multiplication de conditions et événements météorologiques extrêmes, notamment les vagues de chaleur.

Le deuxième facteur aggravant est le vieillissement de la population française et européenne. Par conséquent, le nombre de personnes à risques va en grandissant.

Les facteurs réducteurs sont :

- La technologie : amélioration des systèmes de climatisation
- L'acclimatation
- La politique de culture du risque : mise en place de mesures de prévention pour diminuer sensiblement l'exposition de personnes à risques aux fortes chaleurs.
- Les progrès de la médecine

L'acclimatation ne joue pas un rôle prépondérant compte tenu du fait qu'elle nécessite un temps relativement long, pour les régions nouvellement exposées à ce type d'évènement. Il semble inconcevable, dans une optique de prévisions sur 30 ans, que ces régions en question puissent rapidement s'acclimater.

Par contre, la culture du risque et la technologie peuvent jouer un grand rôle dans l'évolution des tendances de surmortalité due aux canicules.

3 Prédiction d'événements caniculaires par les séries chronologiques (loi de fréquence)

3.1 Principes et étapes de la modélisation de la loi de fréquence

La loi de surmortalité estimée peut être considérée comme des lois d'intensité des canicules, mais elle ne renseigne en aucun cas sur leur survenance. De ce fait un modèle de fréquence doit être établi. Ce modèle devra donner une loi de survenance sur 30 ans, d'une part, mais devra d'autre part évaluer la gravité de la canicule en termes de nombre de jours successifs de grande chaleur.

« Un jour de grande chaleur est un jour au cours duquel une température maximale supérieure ou égale à 35°C et une température minimale supérieure ou égale à 20°C sont constatées ».

La méthode choisie est basée sur la théorie des séries chronologique qu'il s'agit d'appliquer sur des historiques centennaux de températures journalières. Ce sont des historiques de taille très

importante composés de plusieurs dizaines de milliers de températures journalières (30 000 environ).

Des bases de données de ce type sont disponibles sur le site de l'European climate Assessment&Dataset (ECA&D).

Deux types d'historiques sont requis :

- La série chronologique des températures journalières minimales
- La série chronologique des températures journalières maximales

La méthodologie consiste à :

- Calibrer un modèle de type ARIMA à l'aide du logiciel Minitab.
- Simuler le modèle obtenu sur 30 ans.
- Détecter des scénarios caniculaires. Si un jour de grande chaleur est détecté, la loi de surmortalité projetée à l'année détectée est appliquée (par simulation). Si plusieurs jours consécutifs de grande chaleur sont détectés, la loi de surmortalité est utilisée de la même façon

3.2 Choix des données

Le site internet ECA&D (European Climate Assessment&Dataset) met à disposition les deux historiques nécessaires :

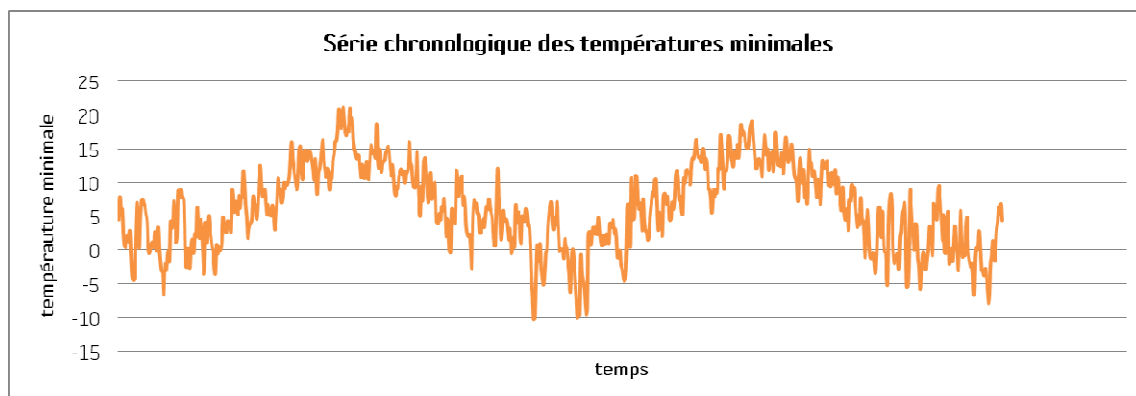
- L'historique des températures journalières maximales depuis le 1^{er} janvier 1900
- L'historique des températures journalières minimales depuis le 1^{er} janvier 1900

Seul le traitement de l'historique des températures minimales est décrit, l'historique des températures maximales étant traité d'une manière analogue. Les parties 6 et 7 feront intervenir les deux historiques.

Soit X_t^1 l'historique des températures minimales et X_t^2 l'historique des températures maximales. L'indice t commence le 1^{er} janvier 1900. Ces historiques sont considérés comme des séries temporelles ou encore séries chronologiques c'est-à-dire une suite finie $(x_1, \dots, x_n)$ de données indexées par le temps.

Un premier retraitement est effectué : les températures enregistrées les 29 février sont éliminées pour obtenir des historiques avec un nombre de jours identiques de 365 pour toutes les années.

Voici l'allure de ces données :



Le but des parties qui suivent est d'estimer les paramètres de la décomposition de X_t suivantes.

$$X_t = T_t + S_t + \varepsilon_t$$

Avec

- $\varepsilon_t - \sum_{i=1}^p \phi_i \varepsilon_{t-i} = Z_t + \sum_{j=1}^q \theta_j Z_{t-j}$ (modèle ARMA(p,q) sur les résidus ε_t), Z_t un bruit que l'on souhaite être un bruit blanc gaussien
- T_t représente la tendance
- S_t représente la saisonnalité

Les calibrages qui suivent permettent d'aboutir à un modèle de type ARIMA (Auto Regressive Integrated Moving Average) de la forme :

$$X_t = T_t + S_t + \varepsilon_t - \sum_{i=1}^p \phi_i \varepsilon_{t-i} = Z_t + \sum_{j=1}^q \theta_j Z_{t-j}$$

3.3 Calibrage d'un modèle ARIMA sur les historiques de températures journalières

3.3.1 Estimation de la tendance et dessaisonalisation

La première étape de la modélisation consiste à écrire X_t sous la forme d'une somme de deux composantes : une tendance T_t et une saisonnalité S_t et d'une composante aléatoire ε_t , c'est-à-dire :

$$X_t = T_t + S_t + \varepsilon_t$$

Les composantes T_t et S_t sont des supposées être des combinaisons linéaires de fonctions connues dans le temps, T_t^i et S_t^j :

$$\begin{cases} T_t = T_t^1 \beta_1 + T_t^2 \beta_2 + \dots + T_t^m \beta_m \\ S_t = S_t^1 \delta_1 + S_t^2 \delta_2 + \dots + S_t^n \delta_n \end{cases}$$

Le but est d'estimer les paramètres $\beta_1, \dots, \beta_m$ et $\delta_1, \dots, \delta_n$ à partir de T observations.

$$X_t = \sum_{i=1}^m T_t^i \beta_i + \sum_{j=1}^n S_t^j \delta_j + \varepsilon_t \text{ pour } t \text{ allant de } 1 \text{ à } T. \quad (1)$$

i Composante saisonnière

Les fonctions S_t^j sont des fonctions indicatrices. Dans le cas de données journalières nous avons :

$$S_t^j = \begin{cases} 0 & \text{si } t = \text{jour } i \\ 1 & \text{si } t \neq \text{jour } j \end{cases}$$

ii Composante tendancielle

Cette composante a une forme simple, reflétant la croissance moyenne. Il existe plusieurs types de composante tendancielle parmi lesquels le type linéaire, le type exponentiel ou encore le type quadratique.

Pour une tendance linéaire, $T_t = \beta_0 + \beta_1 t$ ce qui équivaut à $T_t^1=1$ et $T_t^2=t$ dans le modèle (1).

Le modèle (1) est ajusté sur les historiques de températures. Comme il s'agit de données journalières, il s'agit d'un modèle journalier. La tendance est de type linéaire.

$$X_t = \underbrace{\beta_0 + \beta_1 t}_{T_t} + \underbrace{\sum_{i=1}^{365} \delta_i S_t^i}_{S_t} + \varepsilon_t \quad (2)$$

Matriciellement, ce modèle s'écrit de la manière suivante, en l'appliquant sur les premières données de températures journalières minimales (du 1^{er} janvier 1900 au 7 janvier 1900)

$$\begin{pmatrix} 4,5 \\ 7,9 \\ 6 \\ 6 \\ 2,2 \\ 0,6 \\ 0,4 \\ . \\ . \\ . \\ 3,5 \\ 3,7 \\ 1 \\ -4,2 \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ . \\ . \\ . \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ . \\ . \\ . \\ 366 \\ 367 \\ 368 \\ 369 \end{pmatrix} + \delta_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ . \\ . \\ . \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \dots + \delta_{365} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ . \\ . \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ . \\ . \\ . \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \\ \varepsilon_{11} \end{pmatrix}$$

Le modèle présente donc une régression linéaire multiple de la variable endogène X_t en fonction de 367 variables explicatives (365 variables pour la saisonnalité et 2 pour la tendance). Les 367 paramètres peuvent être estimés par minimisation des Moindres Carrés Ordinaires. Il convient toutefois de spécifier le modèle différemment car la matrice $Y'Y$ n'est pas inversible.

$$Y = \begin{pmatrix} 1 & 1 & 1 & . & . & . & 0 \\ 1 & 2 & 0 & . & . & . & 0 \\ 1 & 3 & 0 & . & . & . & 0 \\ 1 & 4 & 0 & . & . & . & 0 \\ 1 & 5 & 0 & . & . & . & 0 \\ 1 & 6 & 0 & . & . & . & 0 \\ 1 & 7 & 0 & . & . & . & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & 1 \\ 1 & 366 & 1 & . & . & . & 0 \\ 1 & 367 & 0 & . & . & . & 0 \\ 1 & 368 & 0 & . & . & . & 0 \\ 1 & 369 & 0 & . & . & . & 0 \end{pmatrix}$$

En rajoutant la contrainte $\delta_1 + \delta_2 + \dots + \delta_{365} = 0$, une solution générale MCO peut être trouvée. Cette solution est sous forme matricielle.

Il existe également une formule, dite de Buys-Ballot, qui donne une solution du modèle (équation (2) sous la contrainte $\delta_1 + \delta_2 + \dots + \delta_{365} = 0$). Cette solution est non matricielle et facilement implémentable.

$$\hat{\beta}_1 = \frac{12}{m} \frac{\sum_{n=1}^N n \tilde{x}_n - \frac{N(N+1)}{2} \bar{x}}{N(N^2 - 1)}$$

$$\hat{\beta}_0 = \bar{x} - \hat{\beta}_1 \frac{Nm+1}{2}$$

$$\hat{\delta}_j = \tilde{x}_j - \bar{x} - \hat{\beta}_1 \left[j - \frac{m+1}{2} \right]$$

Avec :

- N le nombre d'années entières : 109 années de 1900 à 2008
- $\tilde{x}_n$ la moyenne des X_t relative à l'année n.
- $\tilde{x}_j$ la moyenne des X_t relative à au jour j.
- $\bar{x}$ la moyenne de toutes les observations X_t

Par conséquent, l'estimation des 367 paramètres passe au préalable par l'estimation de N des 108 $\tilde{x}_n$, des 365 $\tilde{x}_j$ et de $\bar{x}$.

Les formules de Buys-Ballot appliquées aux historiques des températures journalières minimales donnent :

n	année	moyenne	n×moyenne
1	1900	7,17	7,17
2	1901	6,19	12,38
3	1902	6,16	18,47
4	1903	6,45	25,79
5	1904	6,63	33,15
	...	...	...
104	2003	9,36	973,30
105	2004	9,09	954,41
106	2005	9,39	995,70
107	2006	9,63	1030,62
108	2007	9,73	1050,56
109	2008	4,61	502,85

mois	jour	moyenne
01	01	1,64
01	02	2,10
01	03	2,10
01	04	1,76
...	...	...
06	04	11,99
06	05	12,17
06	06	12,46
06	07	12,38
12	29	1,90
12	30	1,86
12	31	1,61

Moyenne globale
7,76

Les paramètres proprement dit du modèle peuvent alors être estimés :

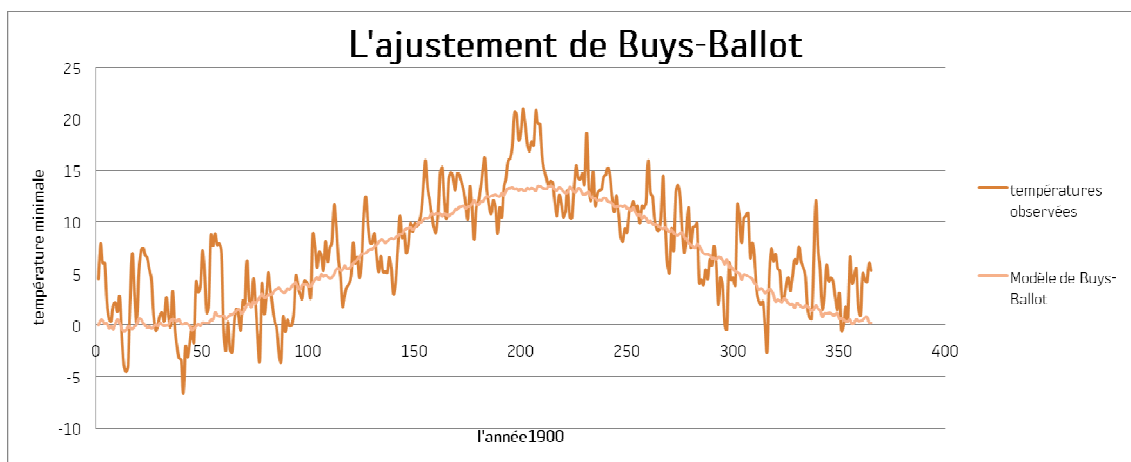
Les paramètres de la tendance :

$\hat{\beta}_1$	$\hat{\beta}_0$
0,0000851	6,09

Les paramètres de la saisonnalité :

$\hat{\delta}_1$	-6,11
$\hat{\delta}_2$	-5,65
$\hat{\delta}_3$	-5,65
$\hat{\delta}_4$	-5,99
$\hat{\delta}_5$	-5,95
$\hat{\delta}_6$	-6,41
$\hat{\delta}_7$	-6,24
...	...
$\hat{\delta}_{360}$	-5,72
$\hat{\delta}_{361}$	-5,66
$\hat{\delta}_{362}$	-5,41
$\hat{\delta}_{363}$	-5,35
$\hat{\delta}_{364}$	-5,88
$\hat{\delta}_{365}$	-5,92

Le modèle ainsi estimé permet de dégager l'architecture de base de la série de températures. Graphiquement, cela se traduit par une oscillation de la série X_t autour de cette base centrale $T_t + S_t$.



Il apparaît clairement la nécessité d'analyser les résidus du modèle, c'est-à-dire la différence entre les températures observées et les températures modélisées par un modèle de Buys-Ballot. Cette analyse doit déboucher sur une décision de modélisation (ou non) de ces résidus.

3.3.2 Identification d'un modèle ARMA sur les résidus

Les résidus ε_t du modèle de Buys-Ballot peuvent également être qualifiés de bruit. Un bruit est intéressant s'il possède des qualités particulières. Elles sont énoncées dans les définitions qui suivent :

i Premières définitions

*Définition 1 : Un processus (X_t) est **stationnaire** si :*

Pour tout t , $E(X_t^2) < +\infty$

Pour tout t , $E(X_t) = \mu$ constante indépendante de t

Pour tout t et pour tout h , $\text{cov}(X_t, X_{t+h}) = E[(X_t - \mu) \cdot (X_{t+h} - \mu)] = \gamma(h)$ indépendante de t

Un processus est stationnaire si les deux premiers moments sont stables. En réalité cette stationnarité est une stationnarité faible ou de second ordre, par opposition à la stationnarité au sens fort qui se définit par une stabilité des moments à tous les ordres (définitions issues de la théorie des chaînes de Markov).

Définition 2 : la fonction γ est appelé fonction d'autocovariance

*Définition 3 : Un **bruit blanc** est un processus stationnaire de second ordre dont les autocovariances sont toutes nulles.*

*Définition 4 : Un **bruit blanc gaussien** est un bruit blanc où les X_t sont normalement distribués.*

ii Comment s'assurer du caractère de bruit blanc gaussien

Il y a peu de chance pour que ε_t soit un bruit blanc gaussien. Il faudrait pour cela que le processus (ε_t) soit stationnaire, que les autocovariances soient nulles et que les ε_t soient normalement distribués. Toutes ces caractéristiques sont rarement réunies sur les résidus résultant d'une élimination de la tendance et de la saisonnalité, quelque soit la série chronologique originelle. Toutefois, il est nécessaire de le prouver.

Le caractère de processus stationnaire est une propriété suffisamment faible pour être aisément vérifiée. Le processus ε_t peut être considéré comme stationnaire car les deux premiers moments sont stables.

Un protocole de vérification peut ensuite être proposé pour vérifier que les autocovariances sont significativement nulles et que les ε_t sont normalement distribués. Mais avant cela, une notion nouvelle doit être introduite :

Définition 5 : la fonction d'autocorrélation d'un processus stationnaire est la fonction

$h \rightarrow \rho(h) = \frac{\gamma(h)}{\gamma(0)}$ pour tout t soit :

$$\rho(h) = \text{corr}(X_t, X_{t-h}) = \frac{\text{cov}(X_t, X_{t-h})}{\sqrt{V(X_t)}\sqrt{V(X_{t-h})}}$$

Pour un échantillon $X_1, \dots, X_T$, ces fonctions peuvent être estimées de la façon suivante :

$$\hat{\gamma}(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} (X_t - \mu)(X_{t-h} - \mu)$$

$$\text{et } \hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

Le protocole peut maintenant être donné :

- Valider le caractère de bruit blanc :

Se fier dans un premier temps au diagramme des autocorrélations, qui place en ordonnée l'autocorrélation et en abscisse l'écart h (appelé lag). Si des autocorrélations sont significativement non nulles de manière visuelle, le caractère de bruit blanc est à l'évidence rejeté.

Si l'autocorrélogramme ne permet pas de décider, effectuer des tests statistiques de bruit blanc (voir le test de Ljung-Box et de Box-Pierce dans la partie 5)

- Valider la distribution normale des résidus :

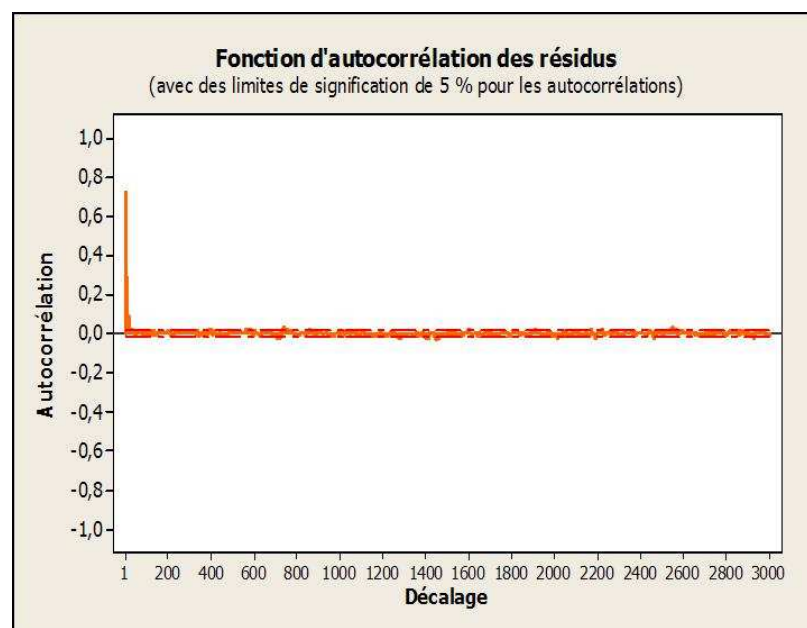
Se faire une impression graphique de la fonction de répartition empirique des résidus. Si visuellement cette courbe colle mal à la fonction de répartition théorique de la loi normale de

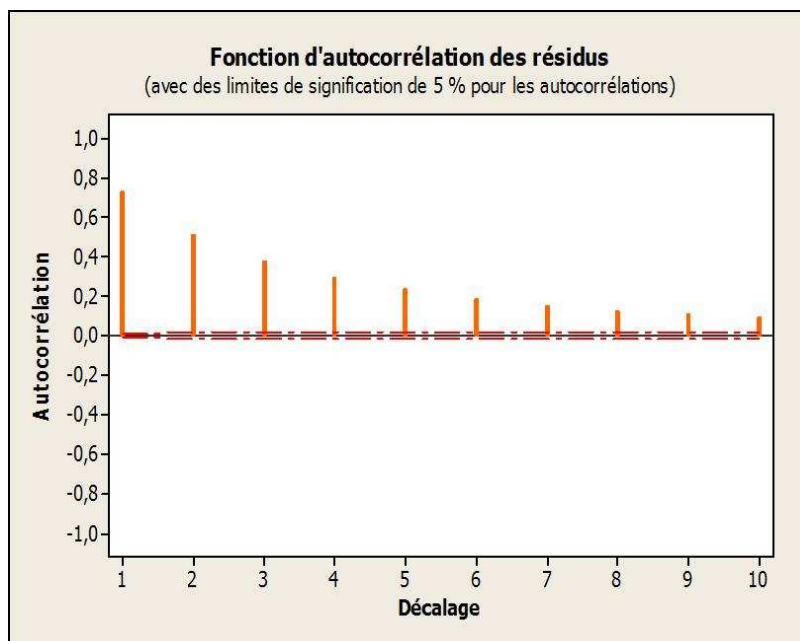
moyenne $\frac{1}{T} \sum_{i=1}^T \varepsilon_i = \bar{\varepsilon}$ et de variance $\frac{1}{T} \sum_{i=1}^T (\varepsilon_i - \bar{\varepsilon})^2$, la distribution normale est rejetée.

Si la courbe colle suffisamment, rien ne peut être conclu car une impression graphique ne permet pas de conclure à une bonne adéquation. La deuxième étape consiste à tracer la droite de Henry. Si l'alignement des points n'est pas évident visuellement, la distribution normale est rejetée

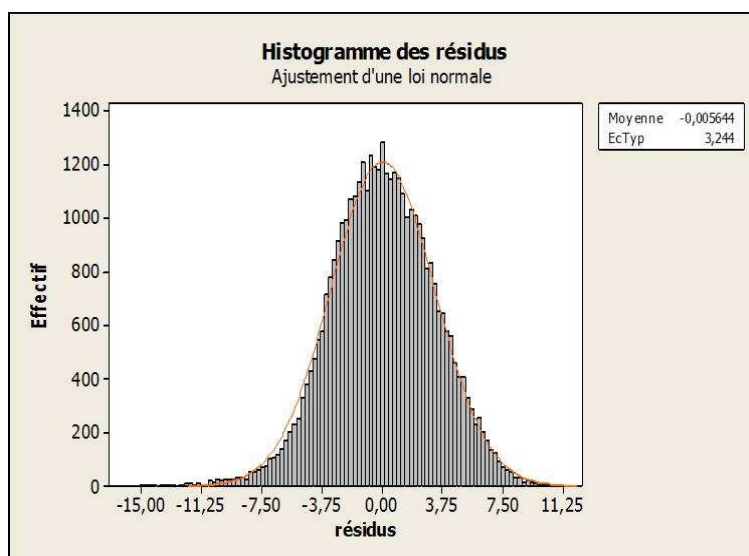
Dans le cas inverse, effectuer un test de normalité pour valider définitivement la distribution normale des résidus. Un test de Kolmogorov conviendrait parfaitement.

La propriété de bruit blanc est d'emblée rejetée par l'autocorrélogramme. Les autocorrélations sont très importantes pour des lag faibles mais décroissent rapidement. Les autocorrélogrammes sont obtenus avec le logiciel Minitab :

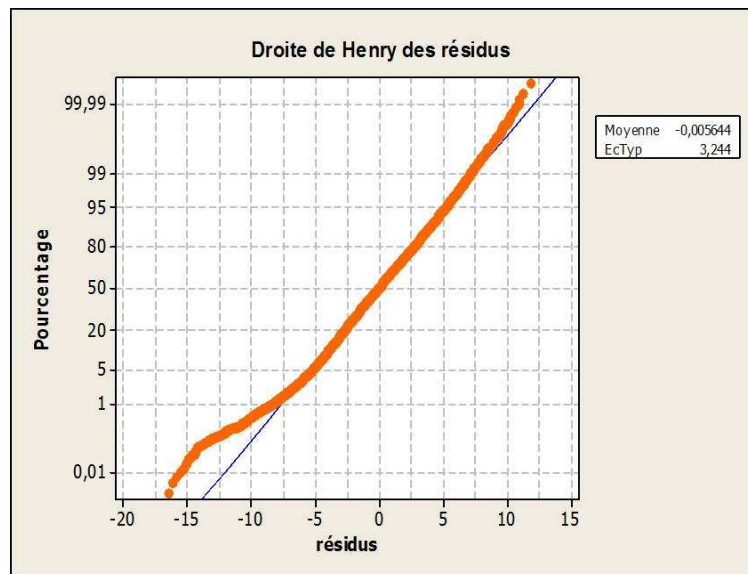




La normalité semble possible au regard de l'histogramme suivant :



Elle est toutefois rejetée en visualisant la droite de Henry.



Le processus (ε_t) est donc bien loin d'être un bruit blanc gaussien. Il est seulement stationnaire. Il faut donc «épurer» du bruit, c'est-à-dire le décomposer en une composante déterministe et une composante aléatoire. Cette composante aléatoire devrait être un bruit blanc gaussien.

iii Modèle ARMA : définition et identification

- Définitions

La notion d'autocorrélation partielle est introduite et les processus ARMA définis.

Définition 6 : L'autocorrélation partielle au lag h d'un processus stationnaire (ε_t) est défini par :

$$\alpha(h) = \text{corr}(\varepsilon_h - E(\varepsilon_h | X_1, \dots, \varepsilon_{h-1}), X_0 - E(X_0 | X_1, \dots, X_{h-1}))$$

$$\alpha(0) = \text{corr}(X_0, X_0) = 1$$

« C'est une corrélation des erreurs de prévision vers le futur et des erreurs de prévision vers le passé »

Pour un processus stationnaire, il est possible d'estimer la fonction d'autocorrélation partielle de manière récursive, en se basant sur les estimations des fonctions d'autocorrélation.

$$\hat{\psi}(h) = \frac{|\hat{R}^*(h)|}{|\hat{R}(h)|} \text{ pour tout } h$$

Avec

$$R(h) = \begin{pmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(h-3) & \rho(h-2) & \rho(h-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(h-4) & \rho(h-3) & \rho(h-2) \\ \rho(2) & \rho(1) & 1 & \dots & \rho(h-5) & \rho(h-4) & \rho(h-3) \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \rho(h-3) & \rho(h-4) & \rho(h-5) & \dots & 1 & \rho(1) & \rho(2) \\ \rho(h-2) & \rho(h-3) & \rho(h-4) & \dots & \rho(1) & 1 & \rho(1) \\ \rho(h-1) & \rho(h-2) & \rho(h-3) & \dots & \rho(2) & \rho(1) & 1 \end{pmatrix}$$

$R^*(h)$ est obtenu en remplaçant la dernière colonne de $R(h)$ par le vecteur $(\rho(1), \dots, \rho(h))'$. $||$ est l'opérateur déterminant.

Le lecteur pourra trouver une preuve dans le cours de Charpentier sur les séries chronologiques.

Définition 7 : On appelle processus ARMA(p,q) un processus stationnaire (ε_t) vérifiant une relation du type :

$$\varepsilon_t - \sum_{i=1}^p \phi_i \varepsilon_{t-i} = Z_t + \sum_{j=1}^q \theta_j Z_{t-j} \text{ pour tout } t \text{ entier relatif.}$$

Avec :

- Les ϕ_i et les θ_j sont des réels
- Le processus (Z_t) est un bruit blanc gaussien.

Un processus AR(p) est un processus ARMA(p,0) et un processus MA(q) est un processus ARMA(0,q). Les entiers p et q constituent l'ordre du modèle.

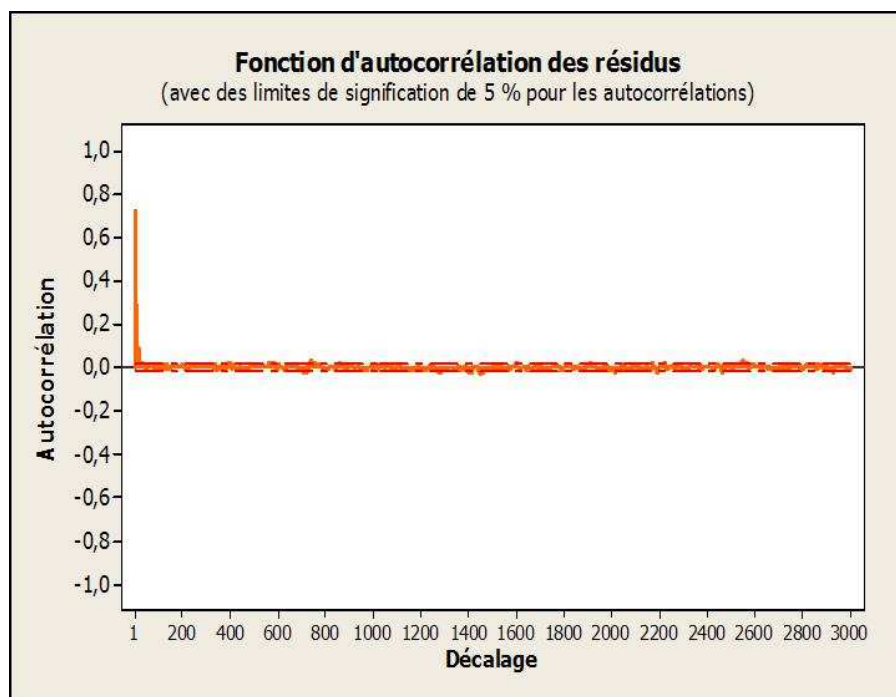
- Identification

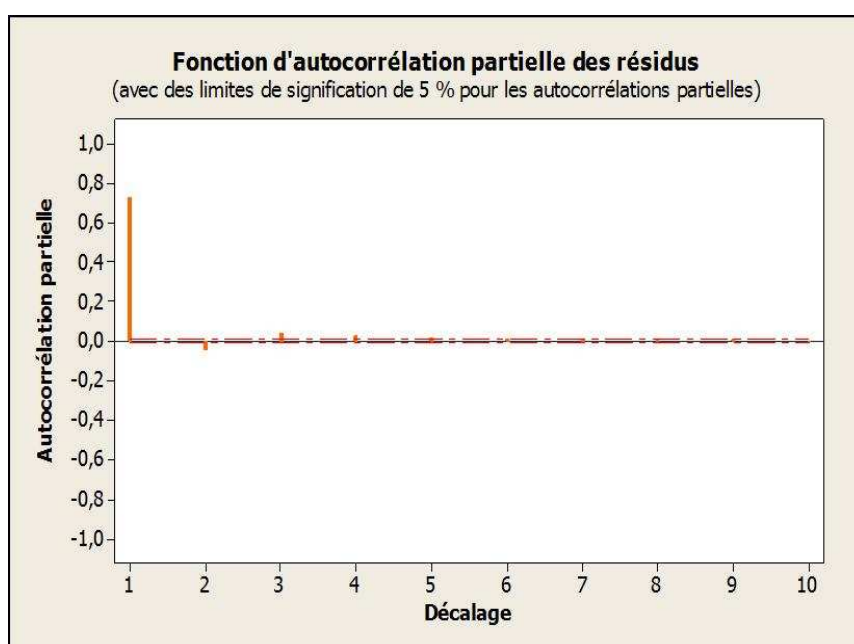
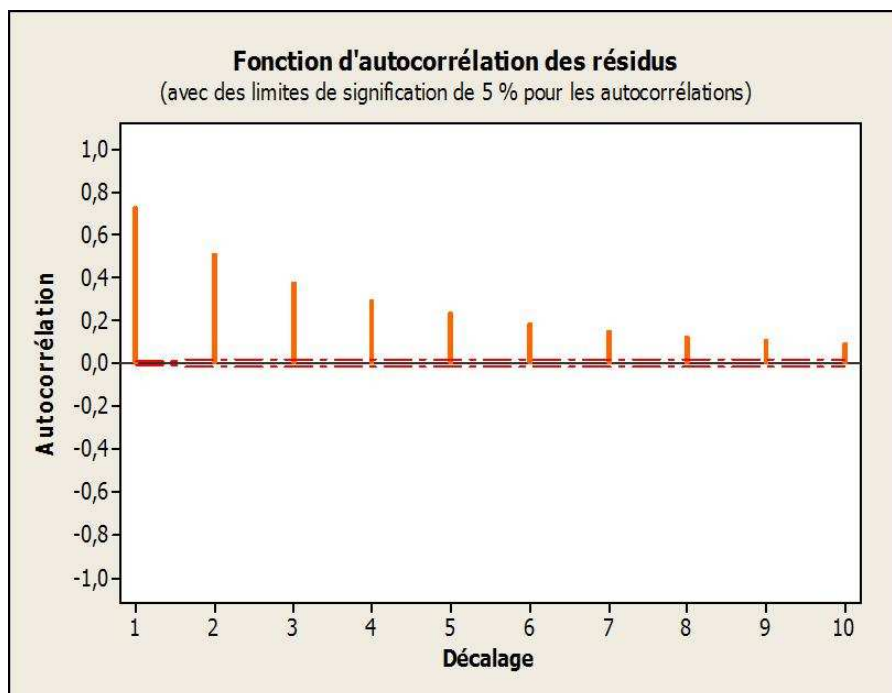
Une règle simple d'identification peut être utilisée. Il s'agit de déterminer l'ordre du modèle.

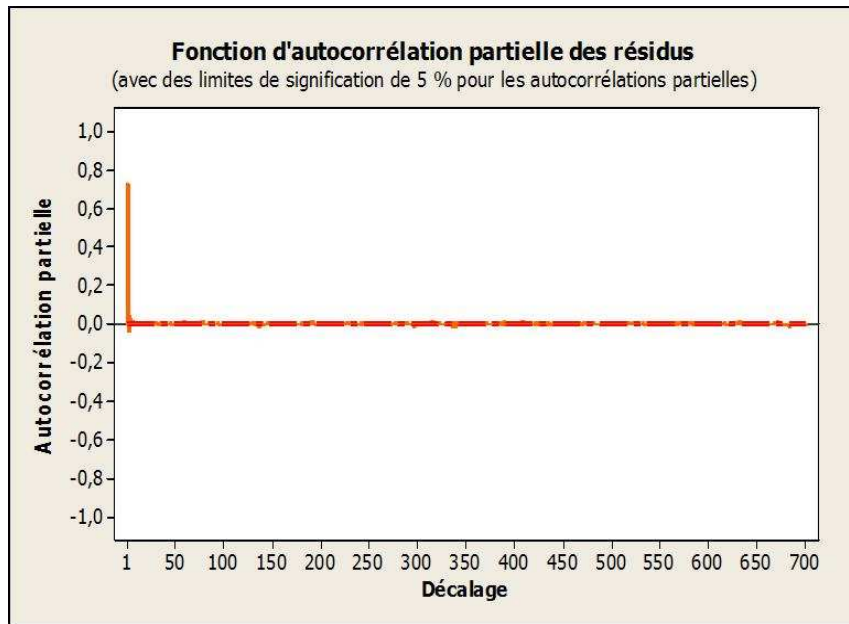
Règle : Pour les processus $AR(p)$, l'autorélogramme partiel s'annule à partir de p . Pour les processus $MA(q)$, l'autocorrélogramme s'annule à partir de q . De ce fait, l'ordre p d'un modèle ARMA est obtenu en déterminant le lag à partir duquel les autocorrélations partielles s'annulent. L'ordre q est obtenu en déterminant le lag à partir duquel les autocorrélations s'annulent.

Cas particulier des modèles AR : les autocorrélations décroissent vers 0 exponentiellement avec h

L'observation des autocorrélogramme et autocorrélogramme partiel des résidus permet d'identifier un modèle **AR(4)**. En effet, les autocorrélations partielles s'annulent à partir de $q=4$. Les autocorrélations décroissent exponentiellement vers 0 en restant positives, ce qui caractérise un ordre p positif (voir le cours de Charpentier). Les sorties Minitab suivantes permettent d'en attester:







La phase d'identification est considérée comme close. Elle est toutefois en réalité écourtée car elle doit en toute rigueur être beaucoup plus complexe. L'idéal aurait été de fixer dans un premier temps des facteurs p^* et q^* par l'identification graphique qui vient d'être réalisée, puis de faire varier les facteurs p et q pour calculer dans chaque cas une quantité permettant de comparer toutes les identifications possibles. Le lecteur intéressé pourra se référer à des cours appropriés.

3.3.3 Estimation des paramètres du modèle identifié

Le modèle identifié sur les ε_t est le suivant :

$$\varepsilon_t - \sum_{i=1}^p \phi_i \varepsilon_{t-i} = Z_t$$

Avec :

- Les ϕ_i sont des réels
- Le processus (Z_t) est un bruit blanc gaussien

Les paramètres ϕ_i d'un modèle AR(p) peuvent être estimés par la méthode des moindres carrés mais peuvent être également l'être par maximisation de la vraisemblance. Minitab utilise la première méthode. C'est pourquoi il convient de rappeler les éléments théoriques.

Un modèle AR(4) s'écrit :

$$\varepsilon_t = c + \phi_1 \varepsilon_{t-1} + \dots + \phi_4 \varepsilon_{t-4} + Z_t = A'_t B + \varepsilon_t \text{ où } A'_t = (1, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-4}) \text{ et } B' = (c, \phi_1, \dots, \phi_4)$$

L'estimation des paramètres du modèle $\varepsilon = A'B + Z$ par les MCO donne :

$$\hat{B} = (AA')^{-1} A\varepsilon$$

Toutefois, cet estimateur est biaisé. Il possède néanmoins la qualité d'être convergent en probabilité vers la vraie valeur de B.

Les résultats donnés par Minitab sont les suivants. Le modèle est spécifié sans la constante c. En effet, la moyenne $\bar{\varepsilon}$ des résidus est quasi-nulle (voir l'histogramme et la droite de Henry plus haut).

ϕ_1	0,755
ϕ_2	-0,072
ϕ_3	0,025
ϕ_4	0,027

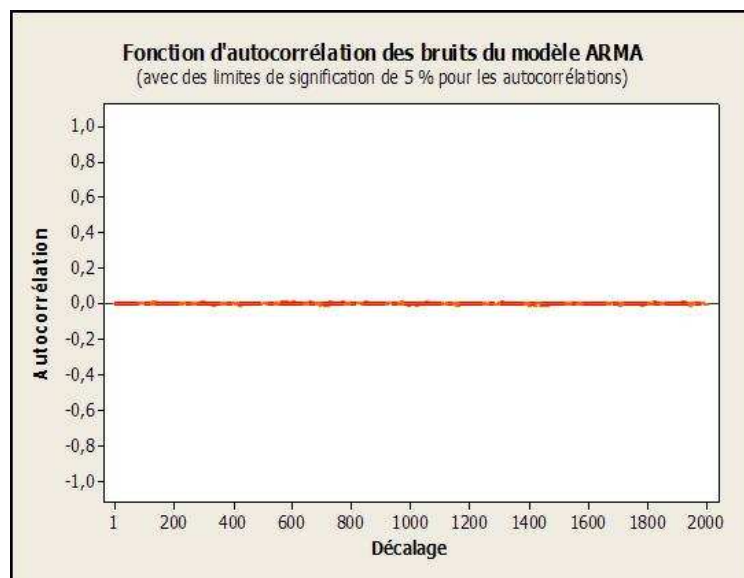
La dernière étape de la modélisation consiste à analyser le bruit Z_t . Le caractère de bruit blanc gaussien est souhaité. Les résultats sont présentés dans la partie 5.

3.3.4 Caractère de bruit blanc du processus (Z_t) et normalité des Z_t : test d'identification

Le protocole d'identification de bruit blanc gaussien peut être repris dans cette partie.

i Bruit blanc

La première étape de l'analyse est d'abord de vérifier si les autocorrélations sont visuellement nulles, sur l'autocorrélogramme. Manifestement, elles le sont, comme le montrent les 2 diagrammes suivants.



Les autocorrélogrammes sont concluants. Toutefois la validation de bruit blanc n'est pas acquise. Seul un test statistique permet de conclure définitivement. Les deux tests suivants peuvent être proposés :

Test n°1 : Le test de Box-Pierce

Il permet de tester l'hypothèse nulle $H_0 : \text{cov}(\varepsilon_t, \varepsilon_{t-h}) = 0$ pour tout h . Il s'écrit donc :

$$\begin{cases} H_0 = \rho(1) = \dots = \rho(h) = 0 \\ H_1 : \text{il existe } i \text{ tel que } \rho(i) \neq 0 \end{cases}$$

La statistique du test est la statistique de Box-Pierce notée Q , donnée par :

$$Q_h = T \sum_{k=1}^h \hat{\rho}_k^2$$

Où

- h est le nombre de retards
- T est le nombre d'observations
- $\hat{\rho}_k$ l'autocorrélation empirique

Sous H_0 , Q_h suit une loi de χ^2 à $h - (p+q)$ degrés de liberté, avec (p,q) l'ordre du modèle ARMA choisi.

Test n°2 : Le test de Ljung-Box

Ce test est analogue au précédent sauf que la statistique du test est :

$$Q'(h) = T(T+2) \sum_{i=1}^h \frac{\hat{\rho}_i^2}{T-i}$$

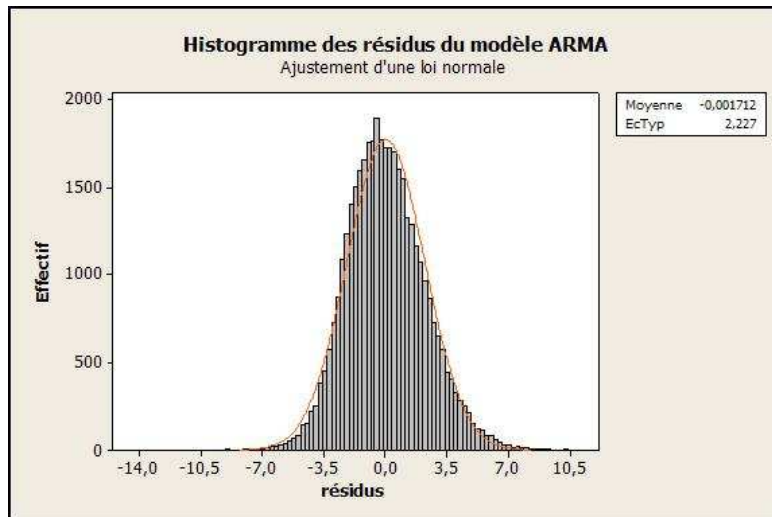
Ce test est celui que Minitab effectue automatiquement lorsque un modèle ARMA est estimé. Les résultats sont les suivants, pour les résidus Z_t :

lag	12	24	36	48
p-valeur	0,004	0,007	0,0027	0,0053

Il apparaît clairement que H_0 est rejetée car les p-value sont faibles. Rigoureusement, le processus (Z_t) n'est pas un bruit blanc. Toutefois, l'impression graphique étant bonne, il sera considéré comme tel par la suite.

ii Distribution normale

Des tests classiques d'adéquation comme le test du khi Deux ou de Kolmogorov permettent de conclure. Toutefois ils ne sont pas nécessaires ici car une simple visualisation de la fonction de répartition empirique permet de constater une dissymétrie évidente. Les deux sorties Minitab qui suivent permettent de le prouver :



iii Conclusion

Rigoureusement les tests et les visualisations graphiques sont clairs : le caractère de bruit blanc gaussien est rejeté.

Toutefois, le but étant de simuler ce processus (Z_t), les résultats précédents sont suffisants. L'ACF permet de conclure à une corrélation nulle entre les résidus, que l'on assimile très abusivement à une indépendance. La loi normale ajustée sur l'histogramme des résidus est considérée comme satisfaisante et permet de conclure, toujours abusivement, à une distribution normale des résidus Z_t , de moyenne -0,001712 et d'écart-type 2,227.

Les Z_t sont donc iid de loi $N(-0,001712 ; 2,227)$. L'approximation qui vient d'être faite est nécessaire pour aller plus loin dans la modélisation. Elle donne le droit de simuler le modèle de type ARIMA qui a été estimé.

3.3.5 Simulation du processus sur 30 ans par génération des résidus

i Principe

La simulation passe par deux étapes :

- Générer des résidus Z_t . Comme le processus de ces résidus est un bruit blanc gaussien, cette étape consiste à générer un Z_t normalement distribué. Comme le processus est un bruit blanc, les Z_t peuvent être générés sans prendre en compte de corrélations.
- Simuler une trajectoire du modèle ARIMA sur 30 ans, pour les températures maximales et les températures minimales

Cette simulation est répétée un grand nombre de fois. Les milliers de scénarios sont classés par ordre de gravité. Un scénario est considéré comme plus grave qu'un autre si le nombre de jours total de grande chaleur est plus grand. Puis le pire ou un des pires des scénarios est analysé.

ii Focus sur la génération des Z_t

Les résidus sont générés à l'aide de la fonction « loi.normale.inverse ». Le problème traditionnel de la discrétance se pose ici. Il aurait fallu recourir à un générateur de réalisations de loi normale meilleur c'est-à-dire qui converge vite et bien vers la loi normale théorique. Toutefois comme les résidus sont générés chaque jour pendant 30 ans, le total de résidus générés est supérieur à 10000. Comme de plus, la version d'Excel est avancée, il est admis sans plus de preuves, que le générateur loi.normale.inverse basé sur la fonction alea, donne des résidus bien répartis, au sens du test du khiDeux et du test de Kolmogorov, tout comme la fonction alea.

3.3.6 Détection de canicule

2000 scénarios sont générés.

Le pire des cas est une canicule avec 11 jours de grande chaleur sur les 30 ans à venir. Ce scénario présente les caractéristiques suivantes :

- Un jour de grande chaleur le 20 août 2020.
- **5 jours de grande chaleur du 11 aout au 15 aout 2023**
- **3 jours de grande chaleur du 17 au 19 aout 2023**
- Un jour de grande chaleur le 11 aout 2025
- Un jour de grande chaleur le 27 aout 2028

Cette détection croisée à loi de surmortalité de 2003 permet de réaliser une projection de la surmortalité en France pour les 30 prochaines années.

Chapitre 4 Application sur des garanties

décès d'un portefeuille prévoyance

I Introduction

Le but de ce chapitre est de comparer, à partir de portefeuilles de grandes tailles, les capitaux cibles retenus par la modélisation interne (déterministe et stochastique) avec la préconisation du QIS4. Deux pays sont choisis et un portefeuille est isolé pour chacun des deux pays.

II Pays Choisis – Sensibilité

Le tableau de sensibilité permet d'établir un ordre de priorité de traitement. Ce tableau est établi en croisant les historiques des nombres de blessés et de tués avec des avis d'experts, pour intégrer une vision prospective (voir chapitre 2).

L'historique est le critère principal dans la mesure de sensibilité. L'inconvénient majeur reste, malgré l'introduction d'avis d'expert (en aggravant ou en réduisant les sensibilités suivant si les avis sont pessimistes ou optimistes), de ne pas prendre en compte des risques qui sont susceptibles de devenir réellement considérables dans les années à venir. Deux exemples peuvent être donnés :

- Séisme dans la région de Nice.
- Attaque terroriste contre une tour à la Défense.

Ces risques ne s'étant jamais réalisés avec une ampleur grande, il ne figure pas dans les historiques. Ils constituent toutefois une menace réelle d'après les avis d'experts.

1 France

Les risques auxquels le portefeuille français est particulièrement soumis sont le risque de canicule et le risque d'accident industriel. Une alternative consisterait à considérer que le scénario

catastrophe envisagé pour le risque industriel peut valoir pour un autre risque auquel la France est sensible comme le risque de séisme ou le risque d'attaque terroriste.

2 Japon

Le Japon est soumis au risque de séismes en premier lieu et aux risques d'inondation et de tempêtes dans un second temps. Ces trois catastrophes naturelles ont été, à de multiples reprises, particulièrement meurtrières dans l'histoire du Japon. L'historique de l'EMDAT fait clairement ressortir ces trois catastrophes comme étant les trois risques catastrophes majeurs.

III Portefeuilles étudiés

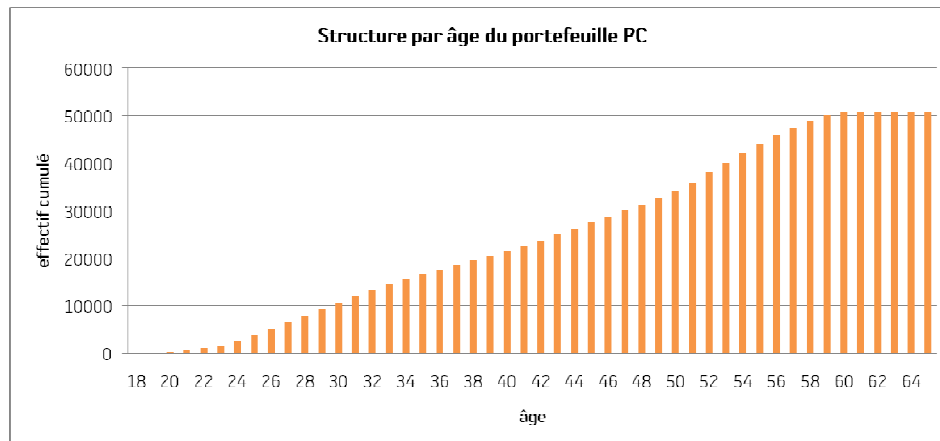
Le périmètre initial du portefeuille est mondial et les contrats sont des contrats d'assurance emprunteur et des contrats classiques de prévoyance individuelle et collective. Toutefois, seulement une partie du périmètre a été choisi et seulement une partie des portefeuilles a été conservé pour mener des études d'impacts (voir chapitre 1 du mémoire).

Les portefeuilles d'assurance emprunteur n'ont pas été retenus et seuls les portefeuilles français et japonais ont été regardés.

1 France : Portefeuille Prévoyance Collective

Le portefeuille contient des contrats de prévoyance collective (noté portefeuille PC par la suite). Ils sont toutefois individualisables. Il peut être sommairement décrit :

- Les garanties sont le décès et l'hospitalisation.
- Le montant total de capitaux sous risques est de plusieurs milliards d'euros.
- Le nombre d'assurés est de plusieurs dizaines de milliers.
- L'histogramme des effectifs cumulés par âge est :



L'âge maximal est de 65 ans, l'impact de la canicule risque donc d'être peu visible et les capitaux cibles relativement faible. Les simulations permettront de chiffrer l'exigence en capital.

L'individualisation semble de plus nécessaire car une étude d'impacts sur un portefeuille collectif doit nécessairement prendre en compte le paramètre de localisation des individus. En effet, les assurés composant un portefeuille d'assurance collective sont regroupés par paquet sur le territoire français. En terme probabiliste, si un assuré est touché par une catastrophe, alors un assuré ayant la même localisation géographique (par exemple dans la même entreprise assurée par un portefeuille collectif) a de fortes chances d'être touché également.

Cette « dépendance » ne sera pas prise en compte dans les simulations qui suivent.

L'hypothèse simplificatrice est faite :

Les assurés du portefeuille PC ne sont pas regroupés par paquet. Ils sont dispersés sur le territoire français. Les caractéristiques de la dispersion des assurés sur le territoire sont les mêmes que les caractéristiques de dispersion de la population totale.

Autrement dit le portefeuille est assimilé à un portefeuille d'assurance individuelle.

Cette hypothèse permet d'affirmer :

- **La probabilité qu'un assuré du portefeuille PC soit touché est la même pour tous les assurés du portefeuille (de la même classe d'âge pour le risque de canicule, quelque soit la classe d'âge pour le risque industriel)**

- Cette probabilité est la même que la probabilité qu'un individu sur le territoire français soit touché, qui est par ailleurs la même quelque soit l'individu (de la même classe d'âge pour le risque de canicule, quelque soit la classe d'âge pour le risque industriel).

2 Japon : Portefeuille global périmètre vie

L'accessibilité à des données détaillées pour des portefeuilles hors-France est plus difficile. Le portefeuille japonais est constitué de 30 collectifs de contrats. Les capitaux sous risques en cas de décès sont isolés et le montant est clairement identifiable pour chaque collectif de contrats.

Pour des raisons de confidentialité, le portefeuille n'est pas affiché.

IV Préconisation du QIS4 sur le risque catastrophe

Une analyse de la version définitive du QIS4 met en évidence quatre formules standards pour les risques catastrophiques. Une doit être clairement prise en compte. Deux sont clairement exclues. La prise en compte de la quatrième est très incertaine.

Les conclusions sont données dans le chapitre 4.

Face à l'incertitude sur la prise en compte des formules du module Santé, deux solutions peuvent être envisagées.

1 Prendre en compte la formule standard du module Santé

La formule standard du module Risque de souscription Vie identifie une charge de capital pour la garantie vie (décès) tandis que les formules du module Santé absorbent les pertes potentielles pouvant être générées par les autres garanties (hospitalisation, chômage, incapacité/invalidité).

Si ce choix de retenir les deux formules est fait, alors la modélisation interne devra elle aussi cibler une charge de capital pour les coûts liés au décès et une charge pour les coûts liés aux autres garanties.

2 *Le choix : retenir uniquement la formule standard pour les activités Vie*

Dans un souci de clarté et de simplification, le choix suivant a été fait :

Les modélisations internes chercheront à calculer uniquement le coût potentiel de la garantie vie. Seule la formule standard du Module Risque de souscription vie SCRlife est prise en compte.

Pour effectuer une comparaison, les modélisations internes qui suivent devront mesurer le capital nécessaire pour absorber les coûts dus au décès des assurés et uniquement au décès (et non dus à l'incapacité, l'invalidité, le chômage ou autre).

La formule standard qui servira de base de comparaison est la suivante :

$$SCR_{vie} = \sum_{police} 0,0015 \cdot \text{capitaux sous risque}$$

Autrement dit, le CEIOPS considère qu'un montant de 1,5 pour mille des capitaux sous risques du portefeuille est adapté pour cibler les pertes probables pouvant être engendrés par les garanties décès (et éventuellement des clauses non vie mineures rattachées à la garantie décès).

V **Modélisation interne : méthodes et résultats**

1 *France*

1.1 Simulation d'une canicule de plus de 5 jours sur la France en 2008

La canicule de 2003 a entraîné le décès de plus de 14 000 personnes en France. La surmortalité constatée par les démographes est la suivante :

		inférieur à 45 ans			45-74 ans			75 ans et plus			Tous âges		
	Nombre de départements	O	E	O/E	O	E	O/E	O	E	O/E	O	E	O/E
inférieure ou égal à 1 jour	38	678	646	1,05	4034	3543	1,14	9414	6360	1,48	14126	10548	1,30
entre 2 et 5 jours	26	493	441	1,12	3018	2323	1,30	7527	4493	1,68	11038	7257	1,50
supérieure ou égal à 5 jours	31	677	620	1,09	4189	2925	1,43	11591	5468	2,12	16457	9013	1,80

Le nombre de décès attendus donné par l'INSERM dans ce tableau contient implicitement une probabilité théorique de survie. Elle contient également implicitement un taux de survie observé. Ces deux quantités seront estimées plus loin. Leur estimation est fondamentale pour mesurer correctement l'aggravation de mortalité due à la canicule (le nombre de décès observés ne suffit pas).

Cette canicule est répétée sur le portefeuille PC, en considérant que la canicule dure au moins 5 jours sur tout le territoire français, c'est-à-dire en aggravant l'événement historique. La méthodologie choisie puis les résultats numériques sont donnés dans les deux paragraphes qui suivent.

1.1.1 Méthodologie

1. Identifier les départements touchés par une canicule de plus 5 jours en 2003.
2. Consulter les bases de données de l'INSEE pour connaître la structure par âge de la population dans ces départements en 2003 (nombres d'habitants par classe d'âge). Les classes d'âges considérées sont les personnes de plus de 75 ans, les personnes ayant entre 45 et 74 ans et les personnes de moins de 45 ans.
3. Calculer les trois probabilités théoriques sous entendus par l'INSERM. La formule est la suivante :

$$\frac{l_x - N_A}{l_x} = \text{Probabilité théorique de survie sous entendue à l'âge } x$$

Avec :

- I_x le nombre d'habitants dans la classe d'âge x des départements touchés par une canicule de plus de 5 jours.
 - N_A le nombre attendu de décès donnés par l'INSERM.
4. Calculer les trois taux de survie réellement observés dans ces régions par la même formule que précédemment en remplaçant N_A par N_O le nombre de décès observés.

Le ratio probabilité théorique de survie/taux de survie réellement observé quantifie l'impact de la canicule de 2003 sur la survie des habitants des départements ayant connus plus de 5 jours consécutifs de grande chaleur.

5. Choisir une table de mortalité et la choquer par le ratio : Probabilité théorique de survie/taux de survie réellement observé. Le choc est évidemment fait en distinguant les trois classes d'âges.

La table choquée est une table qui donne les taux de survie lors d'une année caniculaire.

6. La dernière étape consiste à faire vieillir le portefeuille d'un an en distinguant deux scénarios : aucune canicule ne survient et une canicule de 5 jours au moins s'abat sur toute la France pendant l'année. Les capitaux sous risques d'un assuré qui décède deviennent exigibles par les bénéficiaires et constituent un coût pour l'assureur. Les capitaux sont versés sous forme d'une indemnité en une seule fois aux bénéficiaires. Le coût d'une canicule est calculé en sommant les capitaux sous risques libérés par la canicule et qui n'auraient pas été libérés si la canicule n'était pas survenue.

Une distribution empirique du coût actualisé de la canicule peut aisément être construite, en effectuant 2000 fois l'étape 6.

La méthodologie est appliquée sur le portefeuille. Les résultats sont les suivants.

1.1.2 Les résultats

1. D'après l'INSERM, les départements touchés par une canicule de plus de 5 jours en 2003 sont :

75	Paris
95	Val d'Oise
92	Haut de Seine
94	Val de Marne
93	Seine Saint Denis
78	Yvelines
91	Essonne
72	Sarthe
37	Indre et Loire
86	Vienne
79	Deux Sèvres
36	Indre
18	Cher
89	Yonne
21	Côte d'Or
39	Jura
25	Doubs
63	Puys de Dôme
15	Cantal
69	Rhône
46	Lot
82	Tarn et Garonne
32	Gers
31	Haute Garonne
81	Tarn
30	Gard
84	Vaucluse

2. L'INSEE fournit des données démographiques. Pour 2003, la répartition de la population de ces départements par âges était la suivante :

<45 ans	45-74	75 ans et plus	Total
10,646 millions	7,596 millions	1,404 millions	19,647 millions

3. Les probabilités implicites de survie pour chacune des 3 classes d'âge se déduisent du tableau précédent et de la loi de surmortalité de la canicule de 2003. Ces quantités prévalent sur un mois car les nombres de décès attendus et observés par l'INSERM sont des chiffres enregistrés entre le 1^{er} et le 20 août, ce qui est quasiment une période d'un mois.

- Probabilité mensuelle implicite de survie pour les personnes de moins de 45 ans
- Probabilité mensuelle implicite de survie pour les personnes ayant entre 45 et 74 ans
- Probabilité mensuelle implicite de survie pour les personnes de plus de 75 ans

4. Les taux de survie réellement observés sont

- taux de survie mensuel réellement observé pour les personnes de moins de 45 ans
- taux de survie mensuel réellement observé pour les personnes ayant entre 45 et 74 ans
- taux de survie mensuel réellement observé pour les personnes de plus de 75 ans

Pour obtenir des probabilités de survie qui prévalent sur un an, l'hypothèse suivante est faite :

$$l_{x+t} = l_x^{1-t} \cdot l_{x+1}^t$$

Avec x un âge entier, l_x le nombre de personnes présentes à l'âge x et t un réel strictement supérieur à 0 et strictement inférieur à 1. Cette hypothèse, certes forte et donc peut être peu réaliste, qui implique par ailleurs une force de mortalité constante par année d'âge, implique l'égalité suivante.

$${}_t p_x = (p_x)^t$$

Cette égalité permet de passer directement de taux mensuels de survie à des taux annuels. Les taux qui prévalent sur un an sont les suivants :

- Aucune canicule ne survient : la **probabilité implicite annuelle de survie** est la probabilité mensuelle exposant 12.
- Une canicule survient pendant un mois de l'année : le **taux annuel de survie réellement observé** est la probabilité mensuelle implicite de survie exposant 11 multiplié par le taux de survie mensuel réellement observé pendant le mois de canicule.

Les **ratios de détérioration de la survie** s'en déduisent. Ils sont définis par le rapport entre le **taux annuel de survie réellement observé** et la **probabilité implicite annuelle de survie**.

- ratio de détérioration de la survie pour les personnes de moins de 45 ans
- ratios de détérioration de la survie pour les personnes ayant entre 45 et 74 ans
- ratio de détérioration de la survie pour les personnes de plus de 75 ans

Le tableau suivant donne les résultats des calculs :

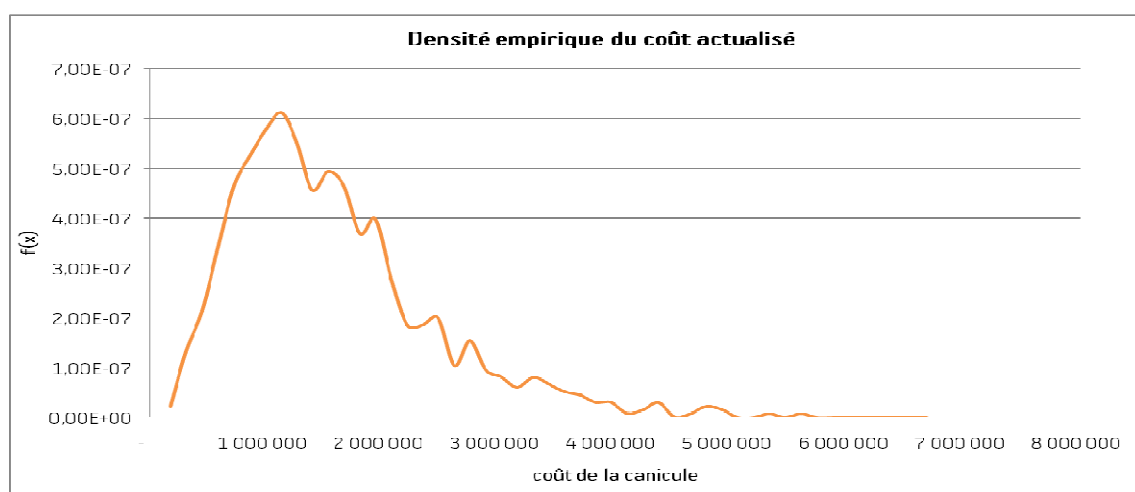
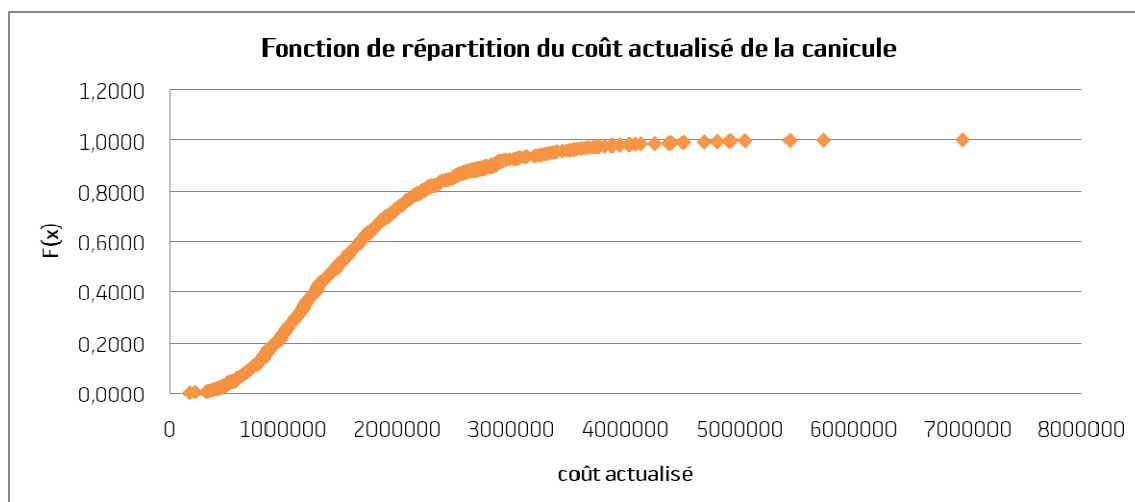
	inférieur à 45 ans	45-74 ans	75 ans et plus
probabilité mensuelle théorique de survie	0,999941757	0,99961498	0,996106221
taux mensuel de survie observé	0,999936412	0,99944854	0,991746618
probabilité annuelle de survie sans canicule	0,999301303	0,99538952	0,954262436
taux annuel de survie avec une canicule	0,999295962	0,99522378	0,950085969
ratio de détérioration de la survie	0,999994655	0,9998335	0,995623356

5. La table de mortalité choisie est la table TH00-002. Les probabilités de survie données par la table sont choquées (par une multiplication) par les ratios de détérioration de la survie. Les probabilités initiales et les probabilités choquées sont données dans le tableau suivant :

Age	Px	Px avec canicule supérieure à 5 jours
21	0,998947581	0,998942242
22	0,998946473	0,998941134
23	0,998955502	0,998950163
...	...	...
44	0,996400651	0,996395325
45	0,996029573	0,991670306
46	0,995654243	0,995488464
...	...	...
74	0,96231752	0,958105798
75	0,958833423	0,95463695
76	0,954937157	0,950757737

6. La simulation de la canicule peut maintenant être réalisée. Le portefeuille est vieilli d'un an dans le cas avec canicule (au moyen des Px avec canicule supérieure à 5 jours) et le cas sans canicule (au moyen des Px).

2000 simulations sont réalisées. La fonction de répartition empirique, la distribution empirique, la moyenne et le quantile à 99,5% du coût actualisé de la canicule sont donnés ci-dessous. La courbe de taux utilisée pour l'actualisation est la courbe de taux du CEIOPS.



Le quantile à 99,5% de la distribution vaut environ 0,7 pour mille des capitaux sous risques.

quantile	ordre	pourcentage des capitaux sous risques
4 682 654	99,5%	0,642

Ce quantile (voir le paragraphe sur les mesures de risque dans le chapitre 1) peut être assimilé à la Value at Risk au niveau 99,5% du coût actualisé.

Dans une optique très proche d'agrégation des risques catastrophiques (voir paragraphe 1.4), il conviendrait de mesurer le risque de pertes dues à une canicule par une mesure de risque cohérente. La TVaR, ou Conditional Tail Expectation (ces deux mesures sont assimilables pour des distributions continues) possèdent cette propriété.

Les pertes potentielles sont mesurées dans le tableau qui suit par la TVaR au niveau 99,5%. En pratique, cette perte potentielle est la moyenne des pertes au-delà du quantile à 99,5%.

TVaR	ordre	pourcentage des capitaux sous risques
5 498 074	99,5%	0,754

La distribution empirique obtenue, et de manière équivalente la chronique des coûts actualisés, est une chronique qui n'est pas forcément valable dans tous les cas pour l'année à venir. Certains événements, ou de manière équivalente certains scénarios, peuvent également survenir : la canicule qui est impactée sur notre portefeuille **ne surviendra pas**, avec un coût actualisé associé nul. Cet événement va considérablement traduire nos coûts actualisés potentiels, et par conséquent les diminuer considérablement. Le quantile à 99,5% sera donc plus faible. Il en sera de même pour la TVaR au même niveau. Remarquons que la plus grande valeur, le quantile à 100 % ne sera pas impactés.

1.2 Ajustement du quantile à 99,5%

Cette section tente de voir comment le quantile à 99,5% est impacté par l'incorporation de scénarios sans canicules, c'est-à-dire des scénarios où le coût actualisé de la canicule est nul.

En toute rigueur, la constitution d'une chronique réaliste passerait par une double simulation :

- Simulation du modèle ARIMA
- Si l'année prochaine est touchée par une canicule de plus de 5 jours, simuler la table choquée et constater un coût non nul. Sinon inscrire un coût nul.
- Répéter les deux premiers points un grand nombre de fois, par exemple 2000 fois.

Cette double simulation déboucherait sur une chronique de coûts actualisés à partir de laquelle le quantile à 99,5% peut être facilement extrait.

Par manque de temps, ces simulations ne sont pas effectuées. Une approximation est faite en constituant une chronique « forcée » de coûts actualisés.

Une méthode pragmatique peut être proposée pour recomposer une chronique réaliste, de taille M au total, de coûts actualisés. Le but est d'obtenir la proportion de chroniques où le coût actualisé de la canicule est nul (parce qu'elle n'est pas survenue).

Pour cela 2000 scénarios sont simulés. Chaque scénario consiste à effectuer le protocole suivant :

- Grâce au modèle ARIMA, simuler une trajectoire de températures sur 30 ans, compter le nombre de canicules d'au moins 5 jours.
- Diviser ce nombre par 30 ans.

En prenant la moyenne, le paramètre p désignant la proportion de chronique (sur M au total) où le coût de la canicule est non nul (précisément car une canicule d'au moins 5 jours est survenue).

La chronique finale est obtenue en ajoutant à la première chronique M-2000 coûts actualisés nuls, c'est-à-dire $\frac{2000}{p} - 2000$ coût actualisé nul. Le quantile à 99,5% est ensuite réestimé.

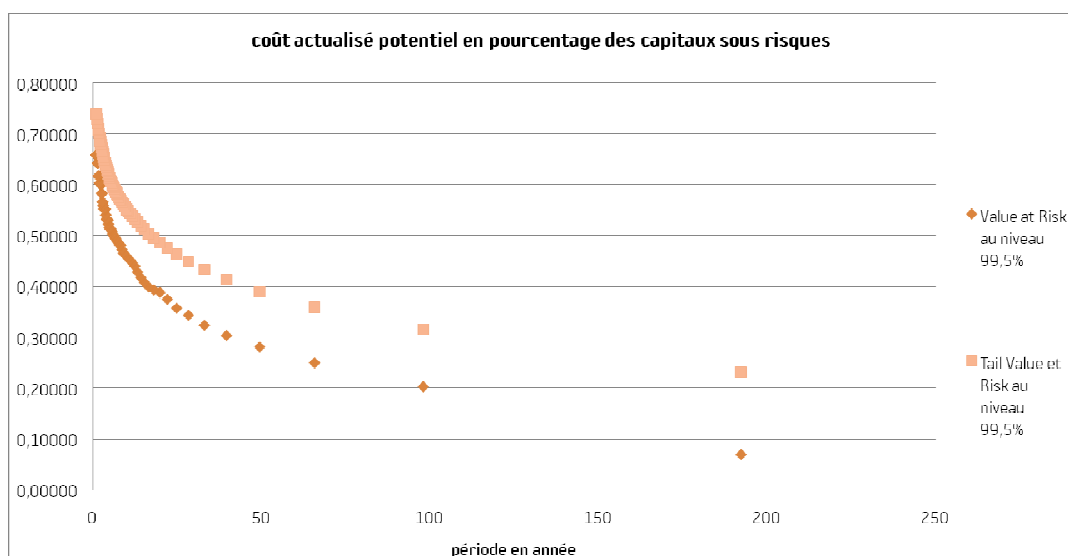
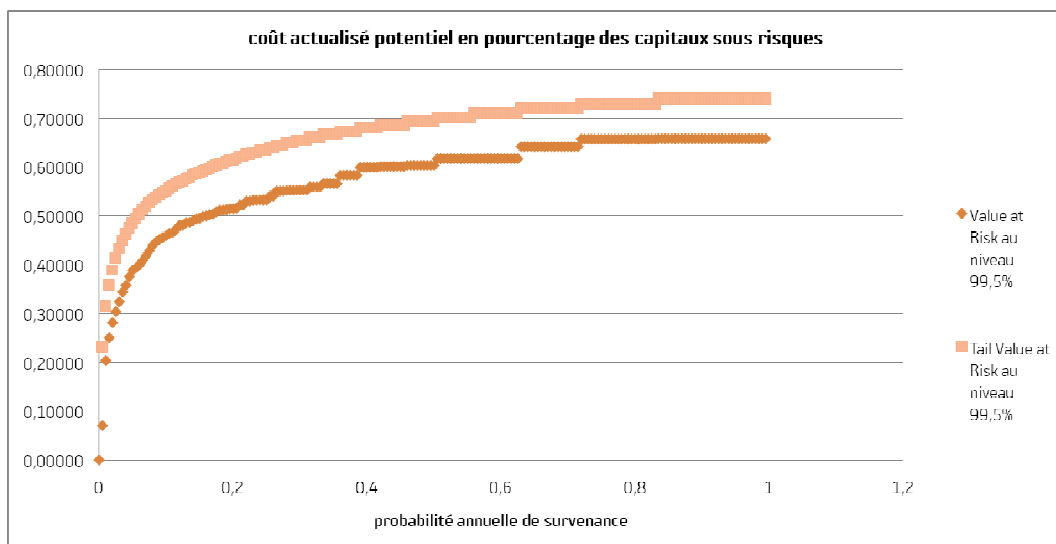
Les 2000 simulations mettent en évidence 12 scénarios avec une canicule d'au moins 5 jours sur les 30 ans. Ce qui signifie que la probabilité d'avoir une canicule d'au moins 5 jours sur 30 ans est de 12/2000, c'est-à-dire 0,006. Ce qui implique que la probabilité d'avoir une canicule de 5 jours pour l'année qui suit est de $0,006/30=0,0002$.

$$p=0,0002$$

A l'évidence, le modèle considère qu'une canicule de 5 jours est un événement extrêmement rare qui survient tous les 5000 ans. Or compte tenu des prévisions des spécialistes sur le changement climatique il paraît invraisemblable de retenir un tel paramètre p. Le quantile à 99,5% serait par ailleurs nul.

L'explication se trouve dans une mauvaise calibration du modèle ARIMA qui manque à l'évidence de réalisme. Il est beaucoup trop callé sur les données de l'historique centennal et ne met donc pas en évidence une tendance notablement croissante. Il aurait fallu pencher la tendance de quelques °C pour détecter plus de canicules longues.

Pour pallier à ce manque de réalisme, il peut être intéressant de représenter le quantile à 99,5% du coût actualisé de la canicule en pourcentage des capitaux sous risques en fonction du paramètre p . Les deux graphiques suivants mettent en relation ces deux variables. Le premier place p en abscisse tandis que le deuxième porte la période $T=1/p$ (en années) de la canicule de 5 jours.

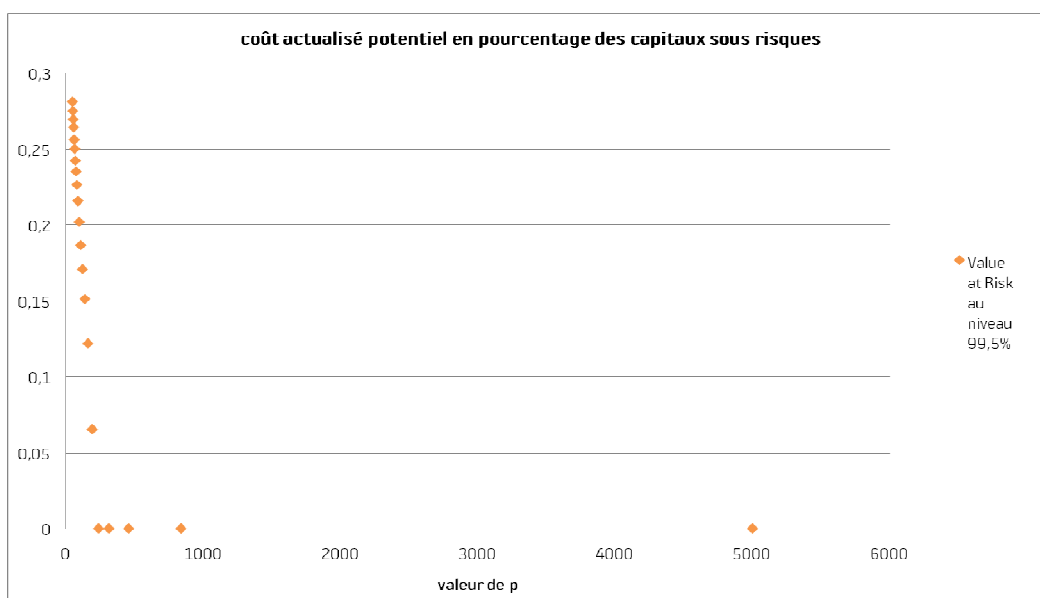


Le premier graphique permet de retrouver le quantile à 99,5% du coût actualisé dans le cas où la canicule survient à coup sûr pour l'année à venir (pour $p=1$).

Il permet également de visualiser le risque correctement mesuré, au moyen de la TVaR au même niveau que les VaR.

Le deuxième montre comment ce quantile et cette TVaR est sensible à la période T de la catastrophe.

Le graphique qui suit représente les quantiles à 99,5% des coûts actualisés en fonction de p très petit.



Trois remarques peuvent être faites sur le dernier graphique :

- Un coût nul (à 99,5%) pour la valeur de p donné par le modèle ARIMA ($p=0,0002$)
- Un coût nul (à 99,5%) pour $p=0,005$, c'est-à-dire en supposant qu'une canicule de 5 jours au moins est un sinistre deux-centennal. C'est le dernier coût nul, par construction de la chronique forcée.
- Un coût non nul (à 99,5%) pour $p=0,01$ c'est-à-dire en supposant qu'une canicule de 5 jours au moins est un sinistre centennal

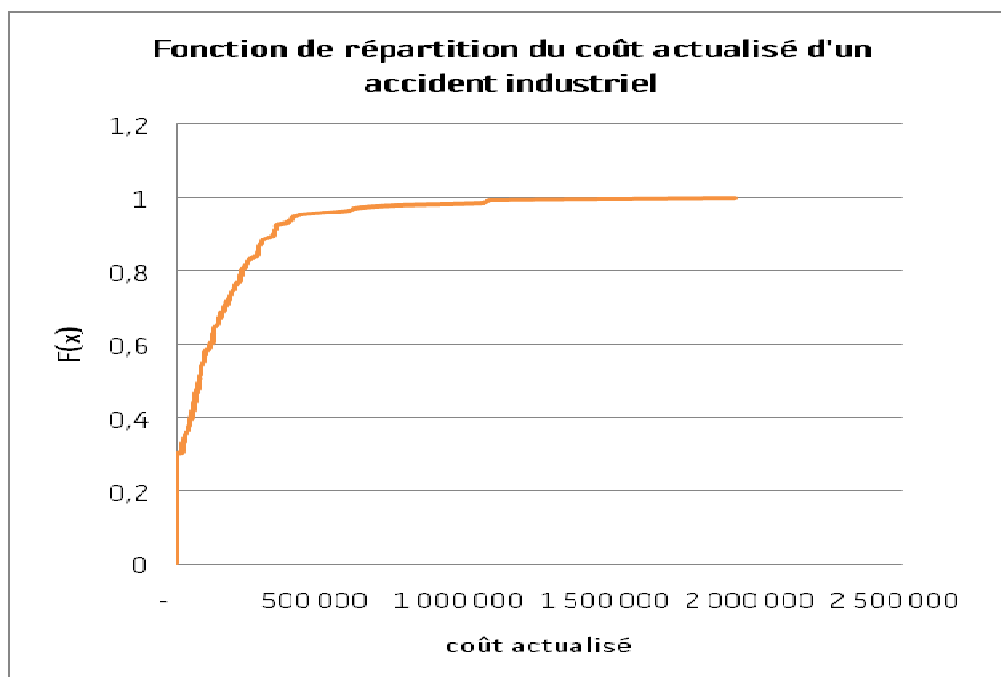
1.3 Le risque industriel : pire des cas répété aujourd'hui

Il n'est pas question d'imaginer un scénario complexe pour en répercuter les impacts sur le portefeuille.

Le principe consiste à estimer une probabilité qu'un individu pris au hasard sur le territoire français soit touché par une catastrophe industrielle localisée faisant 1100 victimes, c'est-à-dire le pire des cas de l'historique centennal (survenu en 1906 d'après le CRED). L'estimation est approximativement de 1100 divisé par 60 millions (nombre d'habitants en métropole).

L'accident industriel est ensuite simulé sur le portefeuille PC. Pour chaque assuré un nombre entre 0 et 1 est généré. S'il est supérieur à la probabilité, l'individu n'est pas touché, sinon il l'est et son montant de capitaux sous risques devient exigible.

Une distribution du coût actualisé de cette canicule peut être obtenue en effectuant 2000 scénarios d'impacts.



Le quantile à 99,5% peut être déduit :

quantile	ordre	pourcentage des capitaux sous risques
1 078 908	99,5%	0,148

Le risque d'accident industriel étant un risque catastrophe mineur par rapport au risque de canicule, le coût actualisé en pourmillage des capitaux sous risques est largement inférieur au montant calculé pour les canicules.

De la même façon que pour le risque de canicule, il conviendrait, dans une optique d'agrégation des risques catastrophes, de mesurer autrement le risque d'accident industriel. Une mesure par la TVaR semble meilleure puisque cette mesure de risque est cohérente.

TVaR	ordre	pourmillage des capitaux sous risques
1 501 249	99,5%	0,206

Il conviendrait de nouveau de reconstituer une chronique réaliste des coûts actualisés. L'accident industriel est considéré comme un sinistre centenaire, d'après l'historique du CRED. Cette information est suffisante pour considérer la proportion de coûts actualisés nuls dans la chronique des coûts possibles.

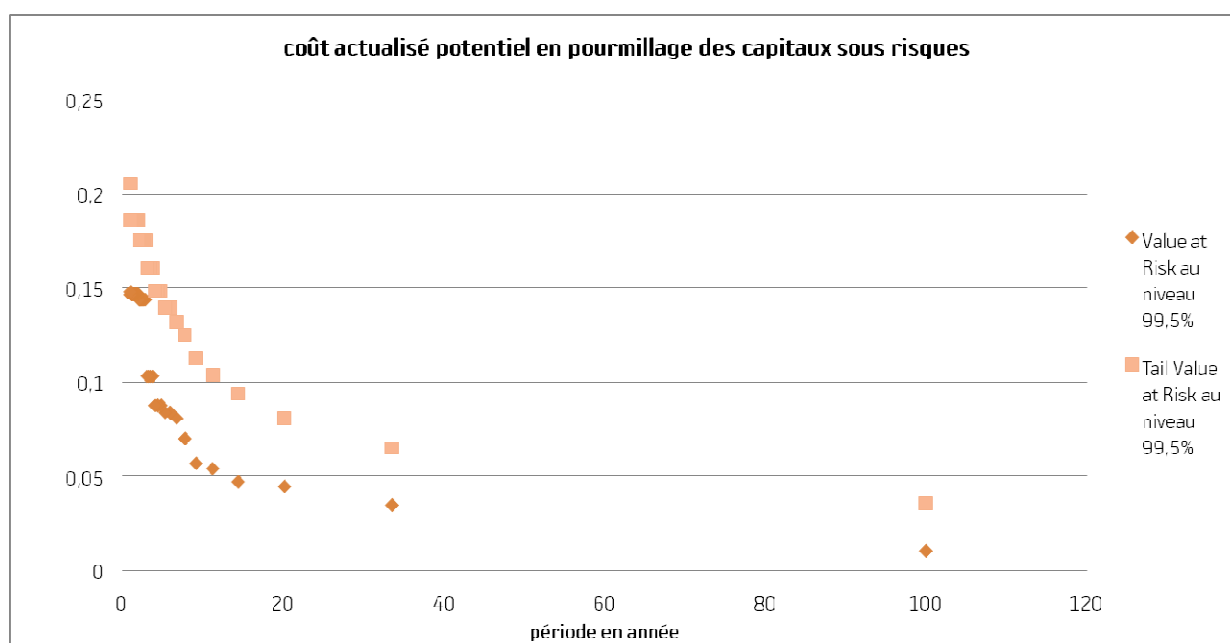
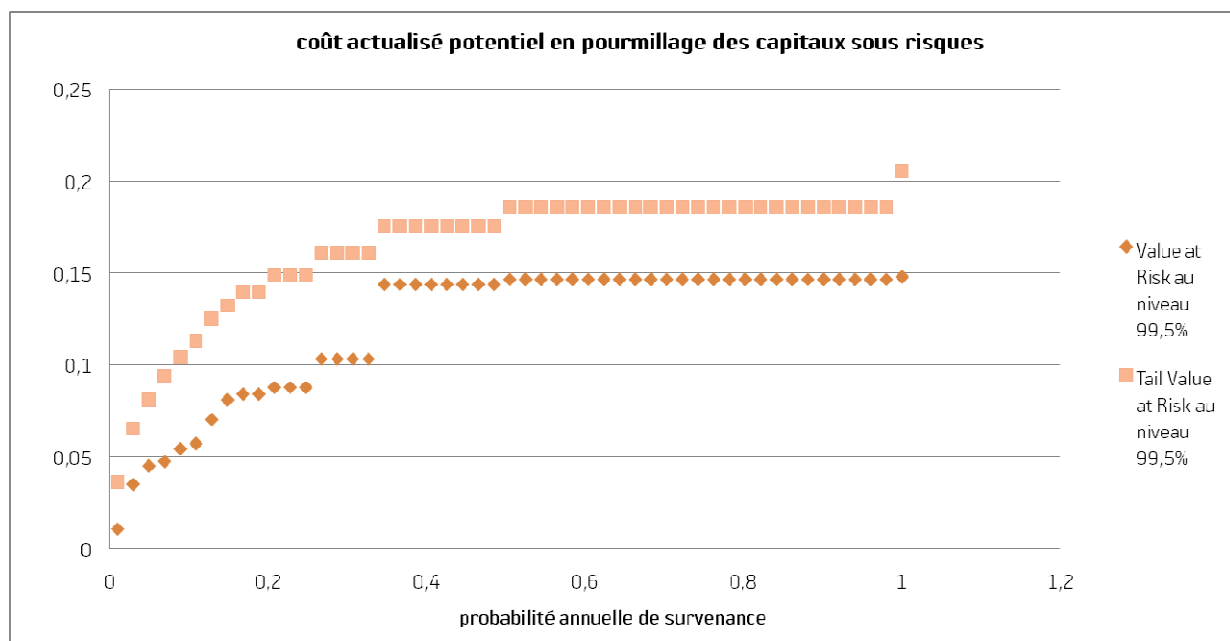
Le paramètre p vaut $1/100 = 0,01$. La chronique est reconstituée et le quantile à 99,5% du coût actualisé devient :

quantile	ordre	pourmillage des capitaux sous risques
76 078	99,5%	0,010

Les graphiques suivants présentent le quantile à 99,5% du coût actualisé de l'accident industriel en pourmillage des capitaux sous risques en fonction du paramètre p d'abord, puis de la fréquence en années ensuite.

Ils présentent également les TVaR, dans l'optique d'une agrégation de risques, qui sera réalisée dans le paragraphe suivant.

Les éventuelles incohérences avec les données dans les tableaux qui précèdent proviennent des approximations liées à la constitution de chronique forcée, et notamment de l'utilisation de l'opérateur « partie entière », dans les manipulations destableaux.



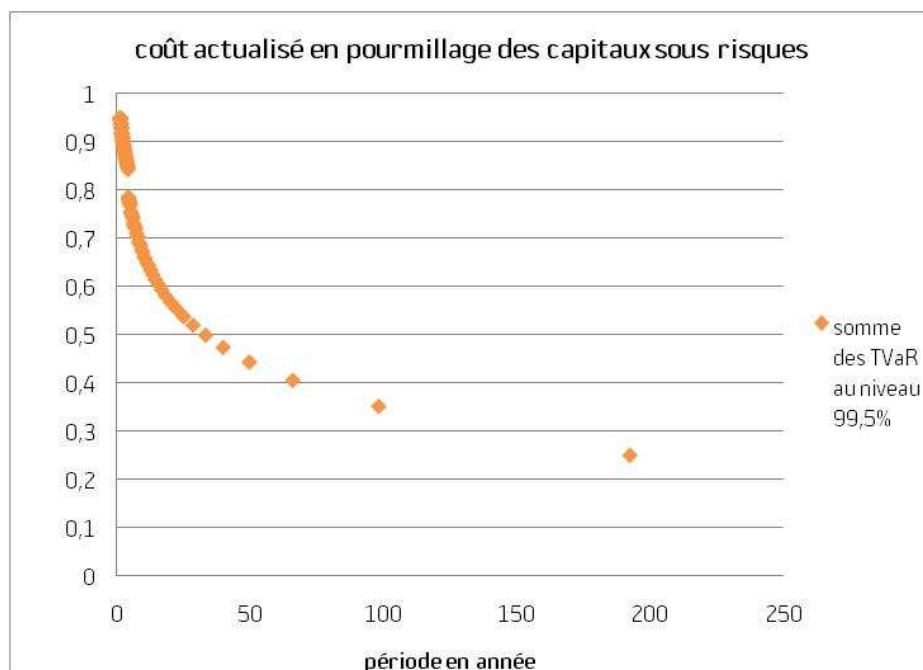
Le paragraphe qui suit calcule un coût global potentiel des catastrophes sur le portefeuille PC, pour l'année à venir. Ce coût sera considéré comme un capital de solvabilité et sera comparé à la préconisation du QIS4.

1.4 Comparaison de la charge retenue avec la préconisation du QIS4.

Les coûts potentiels d'une canicule et d'une catastrophe industrielle ont été mesurés, à un haut niveau de prudence.

La VaR, ou la fonction quantile, n'est pas une mesure cohérente du risque. Par conséquent, elle n'est pas sous additive. Il est donc impossible, en sommant les Value at Risk du risque de canicule et du risque industriel, de savoir où se situe la Value at Risk de la perte globale par rapport à cette somme. Par contre, la TVaR est une mesure de risque cohérente et donc obligatoirement sous-additive. Il est donc possible d'affirmer que la TVaR de la perte globale est inférieure ou égale à la somme des deux TVaR calculées précédemment.

Le tableau suivant donne les la somme des TVaR au niveau 99,5%



La TVaR au niveau 99,5% du risque catastrophe est forcément inférieure à cette somme, par sous additivité de la TVaR, quelque soit la fréquence.

La modélisation interne présenterait donc un certain intérêt pour l'actionnaire puisque les coûts actualisés mesurés par la TVaR sont toujours inférieur à 1,5 pour mille des capitaux sous risques.

Il y a toutefois des limites et des questions qui demeurent :

- Cette conclusion a valeur pour le seul **portefeuille PC**
- **Elle vaut pour les principales catastrophes auxquelles la France est soumise et non pour toutes.** Par exemple un séisme pourrait potentiellement survenir dans la région niçoise or le risque de séisme n'est pas du tout pris en compte jusqu'à présent. D'une manière générale, les pires des cas des autres catastrophes répertoriés par le CRED regroupent 1571 victimes. Dans l'optique où une année serait marquée par tous les pires cas répertoriés par le CRED, le coût potentiel s'élèverait, en effectuant des simulations analogues à celles effectuées pour un accident industriel, à plus de 1 pour mille des capitaux sous risques (environ 1,06).
- **La question du degré de prudence se pose.** Le but étant d'assurer une probabilité de ruine de 0,5% au maximum, la modélisation est-elle assez prudente, pas assez, ou trop prudente ? Le fait de ne retenir que le risque de catastrophe industriel et le risque de canicule permet d'éviter d'être trop prudent et donc d'immobiliser trop de capital. Les simulations des coûts actualisés permettent de dégager des quantiles élevés et d'assurer une certaine prudence. De même, le fait de considérer une probabilité de survenance de 1 est extrêmement prudent. Il se pose un réel problème de quantification de la prudence.
- **Le problème de la fréquence.** Les fréquences trop faibles donnent des quantiles à 99,5% nuls. Pour une chronique de scénarios issue de simulations, le quantile d'ordre 99,5% devient généralement non nul si le paramètre p est supérieur ou égale à 0,005 (1-0,995). En dessous de ce seuil de fréquence, les chroniques des simulations, quelques soient par ailleurs le nombre de simulations, présentent tellement de 0 que le quantile à 99,5% ne se situent pas dans les scénarios correspondant à des coûts actualisés non nuls. Aucune démonstration n'est faite, car cette remarque est purement intuitive.

2 Japon

2.1 Simulation du quantile à 99,5 de la loi d'intensité – des Worst Case – Comparaison

Un séisme est impacté sur le portefeuille pour l'année à venir. La gravité du séisme est mesurée en nombre de victimes.

Il se pose tout d'abord la question du choix du nombre de victimes causé par le séisme. Ce nombre doit refléter un haut niveau de prudence, que l'on fixe à 99,5%. Le module de simulation des séismes permet de dégager une chronique de 20 000 scénarios. Le quantile à 99,5% de cette distribution empirique est de **3854** victimes. Les lois de mortalité avaient été calibrées à partir d'un historique centennal écrêté des nombres annuels de tués. Le pire des cas de cet historique écrêté était de **5297** victimes c'est-à-dire le nombre de tués lors du séisme de Kobe en 1995. L'historique original contient le séisme de 1923 ayant fait plus de **143 000** victimes.

Dans un premier temps un séisme de 3854 victimes est impacté. Puis le nombre de victimes devient une variable et le coût potentiel, exprimé en pourcentage des capitaux sous risques, est analysé comme une fonction du nombre de victimes potentielles qui variera entre 3854 et 143000 victimes.

Dans cette section, le séisme n'est pas simulé sur le portefeuille comme dans le cas des canicules. Seul un rééchelonnement est effectué. Les collectifs de contrats qui composent le portefeuille japonais sont tous touchés dans les mêmes proportions, alors que des simulations auraient permis d'envisager des scénarios où certain collectif auraient été massivement touchés tandis que d'autres n'auraient rien ressentis.

Les coûts actualisés en pourcentage des capitaux sous risques sont donnés pour 3 scénarios :

- Le séisme cause le décès de 3854 victimes, quantile à 99,5% de la loi de mortalité estimée (voir partie sur la modélisation des séismes)
- Le séisme de 1995
- Le séisme de 1923

nombre de victimes	coût actualisé en pourcentage des capitaux sous risques
3 845	0,03
5 897	0,05
143 000	1,10

Les coûts actualisés obtenus ne sont pas réalistes puisque il n'est pas certain qu'un séisme de grande ampleur survienne.

2.2 Constitution d'une chronique réaliste

Le modèle fréquence/coût est utilisé pour obtenir une chronique de coût réaliste pour l'année prochaine, chronique à partir de laquelle le quantile à 99,5% sera réestimé.

Pour obtenir cette chronique plus réaliste, la méthode suivante peut être proposée :

- Générer une chronique d'années touchées.
- Si la **première année** est touchée, générer un coût humain sinon inscrire 0 pour le coût financier actualisé.
- Répéter 2000 fois les 2 premières étapes

Les coûts humains

Ce protocole donne une chronique de 2000 coûts potentiels pour l'année prochaine. Le quantile à 99,5% est considérablement diminué. Il s'élève à **1428**.

Le tableau récapitulatif des coûts actualisés en pourcentage des capitaux sous risques en fonction du nombre de victimes est le suivant

nombre de victimes	coût actualisé en pourcentage des capitaux sous risques
1428	0,01
3845	0,03
5897	0,05
143000	1,10

2.3 Autres catastrophes

Le Japon a connu des catastrophes naturelles de grande ampleur dans son histoire. L'historique du CRED fait apparaître deux catastrophes principales :

- La tempête de 1959 ayant provoqué la mort de 5264 personnes. Le cyclone tropical Vera s'est abattu sur le pays entier au mois de septembre
- Les inondations de 1953 ont été particulièrement meurtrières. Elles ont entraîné la mort de 3967 personnes. Le Japon est un pays particulièrement exposé au risque de crues, pour des raisons topographiques.

Le coût global peut être calculé de la même façon que pour les séismes.

	nombre de victimes	coût actualisé en pourcentage des capitaux sous risques
Inondation	3967	0,03
Tempête	5264	0,04
Total	9231	0,07

De la même façon que pour les canicules, la chronique des coûts doit être rééquilibrée. Les deux catastrophes mentionnées sont des Worst Case centennaux d'après le CRED, les coûts potentiels obtenus plus haut sont des coûts qui ne surviennent qu'une fois sur 100. Les coûts potentiels doivent donc être réestimé.

Conclusion

Les travaux entrepris s'organisent en trois parties :

- Une méthodologie de sélection des catastrophes est proposée. Elle permet de déboucher sur **un tableau de sensibilités** qui montre les expositions majeures du portefeuille étudié aux catastrophes. Ces sensibilités sont segmentées par pays. Le périmètre géographique choisi aurait pu être différent. Il convient de caller le périmètre sur les contraintes pratiques de la compagnie.
- Des modélisations stochastiques ont été proposées. Elles permettent de **rendre compte des réalisations potentielles du risque catastrophe**, du point de vue de la fréquence comme du point de vue de l'intensité. Des mesures de risques peuvent ensuite être effectuées.
- Les études d'impacts ont permis d'établir un rapprochement entre les coûts humains potentiels des catastrophes (estimés par les modélisations du chapitre 3) et les coûts financiers actualisés associés. Elles ont également permis de mettre en évidence une **prudence relativement grande de la formule standard pour le risque catastrophe**. Un montant de capitaux de solvabilité équivalent à 1,5 pour mille des capitaux sous risques semble en effet trop élevé.

L'introduction du mémoire avait introduit une problématique large. L'ambition était relativement grande puisque la question qui était posée était de savoir comment la solvabilité et la richesse d'une compagnie pouvaient être impactées par le risque catastrophe. Cette problématique avait été reprise dans le premier chapitre qui introduisait des possibilités d'études d'impacts sur le capital économique et la valeur intrinsèque d'un portefeuille prévoyance.

A l'issu du stage et de la rédaction du mémoire, cet objectif initial se présente plus comme une tentative de formulation d'une problématique globale en gestion actif/passif. En effet, parvenir à intégrer une modélisation stochastique fine et réaliste des risques catastrophes dans un moteur de

calcul de valeur intrinsèque et de capital économique est un élargissement possible des travaux amorcés dans ce mémoire. Pour que cet élargissement soit possible, il semble absolument incontournable pour toute compagnie d'assurance (réassureur ou courtier) de réunir des spécialistes des risques catastrophes étudiés, ou du moins des avis d'experts détaillés, ainsi que des moyens informatiques importants pour entreprendre les simulations de Monte Carlo nécessaires.

Annexes

Annexe 1 : Tests statistiques

Annexe 2 : Historiques centennaux de l'EMDAT

Annexe 3 : description des garanties non vie

Annexe 4 : Typologie des catastrophes

Annexe 5 : Générer des nombres et simuler en assurance

Annexe 6 : L'approche analytique pour les quantiles extrêmes

Annexe 1 : Tests statistiques

■ Test du Khi Deux

Nous disposons d'un n -échantillon indépendant et identiquement distribué $(X_1, \dots, X_n)$ de X . X est à valeur dans E . P_x est la loi de X .

$(A_1 \dots A_d)$ est une partition de E telle que $\Pi_k = P_0[A_k] > 0$ pour tout k

P_0 est une loi connue sur E .

Soit N_k le nombre de variables aléatoires X_i dans A_k .

Considérons la statistique D définie comme suit :

$$D^2 = \sum_{k=1}^d \frac{(N_k - n\Pi_k)^2}{n\Pi_k}$$

L'hypothèse H_0 est $P_0 = P_x$

Sous H_0 D est asymptotiquement distribué comme une variable χ_{d-1}^2 .

Règle de décision

Si La p-valeur de ce test est strictement supérieure à 0,05 alors on rejette H_0 au seuil de 5%

■ Test de Kolmogorov Smirnov

Soit F_n^* la fonction de répartition empirique d'un n-échantillon d'une variable aléatoire de fonction de répartition F , alors la variable $D_n = \sup |F_n^*(x) - F(x)|$ est asymptotiquement distribuée comme suit :

$$P[\sqrt{n}.D_n < y] \rightarrow \sum_{-\infty}^{+\infty} (-1)^k \cdot \exp(-2k^2.y^2)$$

La statistique D_n est indépendante de F et le résultat précédent permet de tester :

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

Règle de décision

Si La p-valeur de ce test est strictement supérieure à 0,05 alors on rejette H_0 au seuil de 5%

▪ Test de Ljung Box

Le test de Ljung et Box (1978) permet de tester l'hypothèse nulle qu'il n'y a pas d'autocorrélations jusqu'à l'ordre k . La statistique de test est :

$$Q(k) = T(T+2) \sum_{j=1}^k \frac{\tau_j^2}{T-j}$$

Où τ_j est la j -ème autocorrélation des résidus du modèle, et T est le nombre d'observations.

Sous l'hypothèse nulle, la Q-statistique est asymptotiquement distribuée selon une loi du χ^2 , c'est-à-dire $Q(k)$ suit une loi

$$\chi_{1-\alpha}^2(k - p - q)$$

Où p et q sont les ordres d'un modèle ARMA(p, q)

Règle de décision

Si la p -value $< 0,05$ alors on rejette H_0 au seuil de 5% et donc les résidus sont autocorrélés à l'ordre k .

Annexe 2 : Historique centennal

L'historique qui suit donne les nombres de tués au Japon dans les tremblements de terre et les séismes. L'historique est fourni par l'EMDAT.

nombre de tués	Tremblement de terre	Tempêtes
1900	0	0
1901	0	0
1902	0	0
1907	0	0
1910	0	0
1912	0	1000
1914	94	0
1917	0	4000
1920	0	0
1923	143000	3000
1926	0	0
1927	2925	0
1930	0	0
1933	3008	0
1934	0	3006
1938	0	0
1941	0	0
1943	1083	0
1944	998	0
1945	1961	4197
1946	2000	0
1947	0	1930
1948	5131	868
1949	0	617
1950	0	509
1951	0	943
1952	30	176
1953	0	279
1954	0	2253
1955	0	270
1956	0	90
1957	0	109
1958	0	1950
1959	0	5264
1960	182	41
1961	0	296
1962	0	0
1963	0	133
1964	25	42
1965	0	147
1966	0	358
1967	0	290
1968	47	124
1969	0	72
1970	0	0
1971	0	135
1972	0	50
1973	0	101
1974	30	284
1975	0	68
1976	0	169
1977	0	32
1978	49	6
1979	27	71
1980	2	7
1981	0	163
1982	0	218
1983	103	44
1984	20	131
1985	0	49
1986	0	17
1987	2	8
1988	0	0
1989	0	15
1990	0	40
1991	0	101
1992	0	11
1993	241	113
1994	7	3
1995	5297	3
1996	0	6
1997	0	7
1998	0	41
1999	0	26
2000	1	13
2001	2	22
2002	0	5
2003	2	21
2004	40	199
2005	1	123
2006	0	120
2007	10	13
Total	166318	34399

Annexe 3 : Description des garanties non vies

- Perte Totale et Irréversible d'Autonomie (PTIA)
 - Anciennement Invalidité Absolue et Définitive (IAD)
 - Nécessité d'une tierce personne pour exercer au moins 3 des 4 actes de la vie quotidienne (se laver, se nourrir, se vêtir, se déplacer)
 - Inaptitude à tout travail à la suite d'une maladie ou d'un accident, ou à exercer toute activité, même de surveillance ou de direction, susceptible de procurer un salaire ou un gain
- Invalidité Permanente Totale (IPT)
 - Inaptitude, après consolidation de l'état de l'assuré, à tout travail à la suite d'une maladie ou d'un accident, ou à exercer toute activité, même de surveillance ou de direction susceptible de procurer un salaire ou un gain
 - Taux d'invalidité supérieur à 66%
 - Souvent accordée par la Sécurité Sociale après 3 ans d'arrêt de travail
- Invalidité Permanente Partielle (IPP)
 - Impossibilité définitive et permanente d'exercer une activité professionnelle à temps plein
 - Taux d'invalidité compris entre 33% et 66%
- Incapacité temporaire de travail (ITT)
 - Impossibilité temporaire d'exercer toute activité professionnelle rémunérée
 - Perte d'emploi
 - Couvre les licenciements sur les CDI pour lesquels les Assedic ou l'Etat verse des prestations

Annexe 4 : Typologie des catastrophes

- *Epidémies / pandémies* : augmentation inhabituelle d'une maladie infectieuse, qui existait déjà sur la zone ou qui est brusquement apparue.
- *Afflux de parasites ou d'insectes* : afflux et développement de parasites ou d'insectes qui touche animaux, bétails et êtres humains.
- *Feux de forêts* : par forêts, on entend les forêts et les formations subforestières, c'est-à-dire des formations d'arbres feuillus ou de broussailles appelées maquis (formation végétale basse, fermée et dense, poussant sur des sols silicieux) ou garrigue (formation végétale basse mais plutôt ouverte et poussant sur des sols calcaires). Ce type de catastrophe peut être engendré par l'Homme ou par une cause naturelle (éruption volcanique).
- *Tempêtes* : une tempête correspond à l'évolution d'une perturbation atmosphérique ou dépression, où se confrontent deux masses d'air aux caractéristiques bien distinctes (température, humidité...). Inclut les cyclones, tempêtes tropicales, tempêtes tempérées, typhons, tempêtes d'hiver, tornades, grêles ;
- *Sécheresse* ; période longue d'insuffisance d'eau dans le sol. Insuffisance pour la survie des plantes, animaux et Hommes.
- *Montée des eaux, crues de rivières et lacs* : submersion (rapide ou lente) d'une zone pouvant être habitée (inondations).
- *Mouvement de terrain /Avalanche* : Ensemble des déplacements du sol ou du sous-sol, qu'ils soient d'origine naturelle ou anthropique (occasionnés par l'Homme). Sont inclus les effondrements et affaissements, éboulements et chutes de pierres, glissements de terrains et coulées de boue.
- *Tremblement de terre* : Fracturation des roches en profondeur ; celle-ci est due à l'accumulation d'une grande énergie qui se libère, créant des failles, au moment où le seuil de rupture mécanique des roches est atteint (cause tectonique, volcanique ou humaine)
- *Eruption volcanique* : Ouverture mettant en relation la surface du globe avec les profondeurs, permettant à des matériaux terrestres de venir s'épancher en surface (sous forme de laves, gaz...).
- *Accident industriel* : Événement accidentel se produisant sur un site industriel dangereux et entraînant des conséquences immédiates graves pour le personnel, les populations

avoisinentes, les biens ou l'environnement. La loi de 1976 sur les Installations Classées pour la Protection de l'Environnement (I.C.P.E) donne les sites industriels dangereux en France. Sont inclus les fuites de produits chimiques, les explosions, les effondrements, les fuites radioactives (donc les accidents nucléaires), les fuites de gaz et de pétrole, la pollution des sols, nappes phréatiques et de l'atmosphère et enfin les feux. Les sites miniers peuvent également être inclus.

- Les accidents technologiques non-industriels : Accidents provenant de sites non industriels (une propriété privée par exemple). Sont également inclus les explosions, les effondrements et les feux.

Les sites incluent les hôtels et grands magasins, buildings, barrages.

- Catastrophes des transports : Il s'agit des transports aériens et spatiaux, maritimes (cargos, paquebots, pétroliers, plateformes pétrolières) et terrestres (ferroviaires et routiers). Sont inclus les explosions et collisions. Cette catégorie correspond principalement au transport de matières dangereuses.

Annexe 5 : Générer des nombres et simuler en assurance

- *Générer des nombres entre 0 et 1*

Les générateurs de nombres aléatoires entre 0 et 1 sont de plusieurs types et peuvent être classés selon un critère de qualité de répartition (appelée discrétion) de la suite générée sur l'intervalle. Plus la suite couvre vite et bien l'intervalle, meilleur est le générateur.

Une brève classification peut être faite. Pour de plus amples explications, on pourra se référer à Planchet, Jacquemin et Thérond dans « L'utilisation de méthodes de simulation en assurance ».

On distingue d'abord les **générateurs pseudo-aléatoires** des **générateurs quasi-aléatoires**. La discrétion des premiers générateurs est bien moins bonne que celle des deuxièmes. En pratique, cela signifie que pour les premiers, les P-value des tests usuels sont, pour un nombre donné de réalisations, bien plus petites que les P-value des générateurs quasi-aléatoires. De plus, la convergence des P-value vers 1 est beaucoup plus rapide pour ces générateurs quasi-aléatoires.

Le générateur *rnd* de Visual Basic (équivalent à *aléa* d'Excel) est dans la pratique l'exemple type de générateur pseudo aléatoire.

Les générateurs du Tore et du Tore mélangé sont deux générateurs quasi-aléatoires qui possèdent la qualité d'efficacité propre à ce type d'estimateurs et une grande facilité d'implémentation.

- *Simuler une variable de fonction de répartition F*

La génération de nombres aléatoires entre 0 et 1 étant effectuée, la simulation d'autres variables continues en découle, en se basant sur l'inversion de la fonction de répartition de la variable à simuler.

Deux théorèmes fondamentaux permettent de justifier cette pratique : le théorème de la bijection, et une condition suffisante d'existence d'une fonction réciproque.

Théorème de la bijection : Soit f une fonction continue de $\mathbb{R}$ dans $\mathbb{R}$ et monotone sur un intervalle I . Alors la fonction réalise une bijection de l'intervalle I sur l'intervalle $J=f(I)$.

Condition suffisante d'existence : Soit f une fonction de $\mathbb{R}$ dans $\mathbb{R}$ réalisant une bijection de l'intervalle I sur l'intervalle $J=f(I)$. Alors il existe une fonction g appelé fonction réciproque de f notée f^{-1} définie sur J à valeurs dans I .

Pour les fonctions de répartition satisfaisant les hypothèses des deux précédents théorèmes, il est possible de déterminer la fonction de répartition inverse. Ces fonctions sont définies sur un intervalle inclus dans $[0,1]$ car toute fonction de répartition prend des valeurs entre 0 et 1. L'image d'un nombre fixé entre 0 et 1 par la fonction de répartition inverse est, par construction, la réalisation de la variable aléatoire pour laquelle la probabilité d'être en dessous de cette réalisation est le nombre entre 0 et 1.

Ainsi, pour simuler un n -échantillon i.i.d. d'une loi ayant pour fonction de répartition F , il suffit de simuler n réalisations indépendantes d'une v.a.r. de loi uniforme sur l'intervalle $[0;1]$, puis d'appliquer l'inverse de la fonction de répartition à chacune de ces valeurs.

Ceci montre par ailleurs tout l'enjeu d'une simulation "optimale" des réalisations d'une loi uniforme, puisque les seules approximations résident dans cette opération.

Dans certains cas, il est possible de procéder à une inversion analytique de la fonction de répartition. C'est le cas d'une variable distribuée selon une loi exponentielle. Parfois, cette fonction inverse n'est pas aisément calculable. C'est le cas de la loi normale. Pour ce type de fonction, il faut recourir à des méthodes numériques (algorithmes) qui approchent le résultat souhaité. Dans le cas de la loi normale, il existe un grand nombre d'algorithme de simulation. L'utilisation de l'algorithme de Box-Muller ou l'algorithme de DeMoro est très fréquente.

La qualité de génération de variable aléatoire impacte directement la qualité de toute simulation de processus stochastique

- *Simuler un processus stochastique*

Simuler un processus stochastique s'appelle aussi simulation de Monte Carlo.

Une simulation de Monte Carlo est plus un vocable qu'une méthode strictement définie. Tandis qu'il est possible de formaliser parfaitement le terme d'estimateur de Monte Carlo, la simulation de Monte Carlo a un sens plus large aux buts multiples.

Nous disposons de la définition de l'espérance mathématique d'une fonction g de variable aléatoire X , selon laquelle

$$E(g(X)) = \int_{\Omega} g(x) f_X(x) dx$$

f_X est la fonction de densité de X .

Nous disposons également d'un N -échantillon i.i.d $(X_1, X_2, \dots, X_N)$ de la loi de X .

L'estimateur de Monte Carlo de l'espérance est :

$$\tilde{g}_N(x) = \frac{1}{N} \sum_{i=1}^N g(x_i)$$

Cet estimateur est un estimateur sans biais de l'espérance et converge vers cette espérance en probabilité, d'après la loi faible des grands nombres.

Le vocable de simulation de Monte Carlo est adapté aux simulations de processus stochastique.

Vocabulaire : Effectuer une simulation de Monte Carlo de ce processus consiste à générer un très grand nombre de chemins du processus sur une ou plusieurs périodes.

L'hypothèse d'indépendance entre les périodes est classiquement faite.

Cette tâche se résume à générer des réalisations de ε . L'algorithme de DeMoro peut par exemple être utilisé. Pour une échéance donnée, l'estimateur de Monte Carlo peut être aisément calculé, ainsi que tous les quantiles désirés.

Un modèle ARIMA, dont les paramètres ont été au préalable estimés, peut également être simulé de cette manière.

Annexe 6 : L'approche analytique pour les quantiles extrêmes

Supposons que l'on dispose d'un échantillon $(X_1, \dots, X_n)$ où les X_i sont indépendants et identiquement distribués. De plus, F appartient au domaine d'attraction de la loi GEV H_ξ .

On considère ensuite un seuil u et l'on note :

$$N_u = \text{card}\{i : i = 1, \dots, n, X_i > u\} = \text{card}\Delta_n(u)$$

Le nombre d'excès au-delà de u . Les excès sont notés $Y_1, \dots, Y_{N_u}$. La loi des excès est donnée par :

$$F_u(y) = P(X - u \leq y | X > u) = P(Y \leq y | X > u), \quad y \geq 0$$

D'après le théorème de Pickand donné en 1.3, pour u assez grand, on a la relation suivante :

$$F_u(x) \approx G_{\xi, \beta(u)}(x)$$

Les paramètres ξ et β sont estimés par $\hat{\xi}_{k,n}^H$ et $\hat{\beta}_{MV}(\hat{\xi}_{k,n}^H)$ pour le u choisi au préalable.

Le but de cette annexe est de donner une formule analytique pour un quantile d'ordre élevé de F . La formule découle immédiatement de la décomposition suivante de $S=1-F$:

$$S(u+y) = S(u) \cdot S_u(y)$$

Avec $S_u(y) = 1 - F_u(y)$

En notant $\hat{F}$ et $\hat{S}$ respectivement la fonction de répartition estimée et la fonction de survie estimée de X_i , l'égalité précédente donne :

$$\hat{S}(u+y) = \frac{N_u}{n} \left(1 + \hat{\xi}_{k,n}^H \frac{y}{\hat{\beta}_{MV}(\hat{\xi}_{k,n}^H)} \right)^{-1/\hat{\xi}_{k,n}^H}$$

Un estimateur $\hat{x}_p$ du quantile extrême d'ordre $1-p$ (avec p petit, c'est-à-dire $1-p$ proche de 1) s'en déduit, par inversion de $\hat{F}$:

$$\hat{x}_p = u + \frac{\hat{\beta}_{MV}(\hat{\xi}_{k,n}^H)}{\hat{\xi}_{k,n}^H} \left(\left(\frac{n}{N_u} (1-p)^{-\hat{\xi}_{k,n}^H} \right) - 1 \right)$$

Bibliographie

- Ouvrages

- De Grossi Patricia, Howard Kunreuther, Chandu C. Patel : *Catastrophe Modeling A New Approach to Managing Risk*; Springer (2005)
- Embrechts Paul, Klüppelberg Claudia, Mikosch Thomas : *Modelling Extremal Events for Insurance and Finance* ; Springer (2004)
- Gordon Woo : *The Mathematics of natural disaster* ; World Scientific Book Shop (1999)

- Publications

- CEIOPS : *QIS 4 technical specifications*
- INSEE (Institut National de la Statistique et des Etudes Economique) : *Les risques majeurs naturels et technologiques* ; Courrier des statistiques n° 120 (2007)
- OCDE : *Les risques émergents du XXIème siècle* (2003)
- Planchet, Jacquemin, Thérond : *Utilisation de méthodes de simulation en assurance* (2003)
- Planchet Frédéric : *Utilisation de la théorie des valeurs extrêmes dans le contexte solvabilité 2* (2007)
- SWISS RE : *Publications/brochures du département « périls naturels »*

- Mémoires et thèses
 - Henge Frédérique : *Rapprochement des concepts de valeur intrinsèque et de capital économique* ; Mémoire ULP Strasbourg (2006)
 - Meister Vincent : *Solvabilité II : contexte, valorisation et impacts sur l'exigence en capital* ; Mémoire ULP Strasbourg (2007)
 - Roustant Olivier : *Produits dérivés climatiques : aspects économétriques et financiers* ; Thèse (2003)
 - Sibaï David : *Modélisation dynamique des catastrophes naturelles* ; Mémoire ENSAE (2003)
- Cours
 - Charpentier Arthur : *Cours de séries temporelles*
 - Planchet Frédéric : *Utilisation de la théorie des valeurs extrêmes dans le contexte solvabilité 2* ; Cours ISFA
- Site Internet
 - www.emdat.be : EMDAT (The international Emergency Disasters Database) du Centre of Research and Epidemiology of Disaster (CRED) : *Base de données*
 - www.developpement-durable.gouv.fr : Ministère de l'Environnement et du Développement Durable (MEDD)
 - www.irma-grenoble.com : L'Institut des Risques Majeurs de Grenoble (IRMa)

Liste des abréviations

- ACF : Auto Correlation Function
- AMO : Oscillation Atlantique Multidécennale
- ARMA : Auto Regressive Moving Average
- ARIMA : Auto Regressive Integrated Moving Average
- CEIOPS : Committee of European Insurance and Occupational Pensions Supervisor
- CRED : Centre of Research and Epidemiology of Disaster
- CTE : Conditional Tail Expectation
- ECA&D : l'European Climate Assessment&Dataset
- GEV : Generalized Extreme Value
- GIEC : Groupe d'Expert Intergouvernemental sur l'Evolution du Climat
- GPD : Generalized Pareto Distribution
- EMDAT : Emergency Disasters Database
- IAA : International Actuarial Association
- IAD : Invalidité Absolue et Définitive
- IAD : Invalidité Absolue et Définitive
- IARD : Incendie Accident Risques Divers
- INED : L'Institut National d'Etudes Démographiques
- INSEE : Institut National de la Statistique et des Etudes Economiques
- INSERM : Institut National de la Santé Et de la Recherche Médicale
- IPP : Invalidité Permanente Partielle
- IPT : Invalidité Permanente Totale
- IRMa : Institut des Risques Majeurs de Grenoble
- ITT : Incapacité temporaire de travail
- LOB : Lign Of Business
- MCO : Moindres Carrés Ordinaires
- MEDD : Ministère de l'Environnement et du Développement Durable
- OCDE : Organisation de Coopération et de Développement Economique
- OMM : l'Organisation Météorologique Mondiale
- USGS : United States Geological Surey

- PC : Prévoyance Collective
- PM : Provision Mathématique
- POT : Peak Over the Treshold
- PACF : Partial Auto Correlation Function
- PRC : Provision pour Risques Croissants
- PSAP : Provision Pour Sinistres à Payer
- PTIA : Perte Totale et Irréversible d'Autonomie
- QIS : Quantitative Impact Studies
- QQPlot : Quantile Quantile Plot
- SOA : Society of Actuaries
- SCR : Solvency Capital Requirement
- TVaR : Tail Value at Risk
- USGS : United States Geological Surey
- VaR : Value at Risk
- VIF : Valeur de l'I

