

**Mémoire présenté devant le jury de l'ENSAE en vue de l'obtention du
Diplôme d'Actuaire ENSAE
et de l'admission à l'Institut des Actuaires**

le 04 juillet 2013

Par : Nicolas LECLERCQ et Jules MICHEL

Titre : Modélisation bayésienne du risque opérationnel

Confidentialité : oui, un an. *Les signataires s'engagent à respecter la confidentialité indiquée.*

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

Entreprise

Crédit Agricole S.A.

Directeur de mémoire en entreprise

Vincent LEHERISSE et Sophie LAVAUD

Signature :

Signature :

Membres présents du jury de l'ENSAE

Xavier MILHAUD

Invité

Signature :

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

Signature des responsables entreprise

Signature des candidats

Bibliothèque :

Secretariat :

Mémoire d'actuariat

Modélisation bayésienne du risque opérationnel

Nicolas LECLERCQ, *Jules* MICHEL



Tuteurs :

Option : Actuariat

ENSAE :
Xavier MILHAUD

Entreprise :
Vincent LEHERISSE

Sophie LAVAUD

Résumé. Les risques opérationnels recouvrent les risques de pertes résultant de l'inadaptation ou de la défaillance de procédures, personnes, systèmes internes ou résultant d'événements externes (catastrophe, incendie, fraude, etc.). Ils sont quantifiés à des fins réglementaires via des modèles internes sévérité-fréquence au sein de cellules de risque prédéfinies, permettant de construire une distribution de perte (via la méthode de Monte Carlo) à un an et d'en déterminer la VaR à 99.9%. Les distributions peuvent être fondées soit sur des données historiques collectées par la banque, soit sur des analyses de scénarios établies à dire d'experts. En conséquence, afin de caractériser complètement un type de risque, il est intéressant pour la banque d'être en mesure de construire des distributions de pertes intégrant à la fois des données internes et des analyses de scénarios, ce qui permet notamment de déterminer un calcul de capital unique. Actuellement, le Crédit Agricole adopte une méthodologie de calcul qui ne combine pas ces données. Ce mémoire traite des différentes méthodes bayésiennes qui pourraient être appliquées au contexte du risque opérationnel. Quatre types de méthodes sont envisagées : une méthode bayésienne pure qui se fonde sur la détermination d'une loi a priori et d'une loi a posteriori via des lois conjuguées en développant la théorie de Bayes. La deuxième méthode repose sur le principe de la crédibilité. La troisième méthode appelée "cascade bayésienne" se base sur une théorie récente qui utilise l'approche bayésienne pure combinée à la théorie des valeurs extrêmes (TVE). La dernière méthode s'affranchit du recours à des lois conjuguées en utilisant les propriétés des chaînes de Markov. Enfin, une synthèse des méthodes est présentée pour dégager les avantages et inconvénients de chacune, proposer des alternatives et donner des pistes de réflexion.

Mots-clés. Statistique bayésienne - Risque opérationnel - Loss Distribution Approach (LDA) - Actuariat non-vie - Théorie de la crédibilité - méthode de Monte Carlo - Théorie des Valeurs Extrêmes - MCMC

Abstract. Operational risks include the risks of losses resulting from inadequate or failed processes, people, internal systems or from external events (disaster, fire, fraud, etc.). They are quantified for regulatory purposes via internal frequency-severity models within predefined risk cells in order to build a loss distribution (via the Monte Carlo method) at one year and so as to determine the Value-at-Risk at 99.9%. Distributions can be based either on historical data collected by the bank, either scenario analysis relied on expert opinions. Therefore, in order to fully characterize a risk type, it is interesting for the bank to be able to build loss distributions incorporating both internal data and scenario analysis, which allows to determine a unique calculation of the capital. Currently, The Crédit Agricole takes a calculation methodology which does not combine these data. This thesis deals with different bayesian methods that could be applied to the context of operational risk. Four kinds of methods are considered : a pure bayesian approach based on the determination of a prior distribution and a posterior distribution via combined distributions by developing and adapting the Bayes theory. The second method is based on credibility theory. The third method called "cascade bayesian" relies on a recent theory which combines the pure bayesian approach with the Extreme Value Theory (EVT).The last method overcomes the combined distributions using the properties of Markov chains. Finally, a summary of the methods is presented to identify the advantages and drawbacks of each one, to suggest alternatives and give potential improvements.

Keywords. Bayesian theory - Operation Risk - Loss Distribution Approach (LDA) - Non-life insurance - Credibility theory - Monte Carlo method - Extreme Value Theory (EVT) - MCMC

Remerciements

Nous tenons à remercier nos encadrants au Crédit Agricole S.A. Vincent LEHERISSE et Sophie LAVAUD pour leur disponibilité et leurs conseils avisés. Nous souhaitons également remercier notre entourage pour nous avoir soutenus et encouragés.

Table des matières

Introduction	10
1 Le risque opérationnel : présentation du contexte	12
1.1 Définition	12
1.2 Bâle II et le risque opérationnel	13
1.2.1 Présentation de Bâle II	13
1.2.2 Les méthodes pour quantifier le risque opérationnel	14
1.3 Rapprochement avec Solvabilité II	16
1.3.1 La formule standard	17
1.3.2 Les modèles internes	17
1.4 Les différents modèles AMA	17
1.4.1 Les approches proposées par le Comité de Bâle	18
1.4.2 L'approche top-down	19
1.4.3 L'approche bottom-up	19
2 Le modèle LDA : Loss Distribution Approach	22
2.1 Description générale du modèle	22
2.2 Calcul du capital réglementaire basé sur les données historiques	23
2.2.1 Estimation de la distribution de sévérité	23
2.2.2 Estimation de la distribution de fréquence	25
2.2.3 Construction de la perte annuelle	25
2.3 Calcul du capital réglementaire basé sur l'analyse de scénario	28
2.3.1 Construction et utilisation des scénarios	28
2.3.2 Prise en compte des assurances	30
2.4 Agrégation des charges en capital	31
2.4.1 Méthodologie d'agrégation	32
2.5 Application numérique à nos données	33
2.5.1 Présentation des données	33
2.5.2 Détermination des capitaux réglementaires à l'aide des données historiques	34
2.5.3 Détermination des capitaux réglementaires à l'aide des scénarios	36
2.5.4 Agrégation des capitaux	37
2.6 Objectifs du mémoire	38
3 Approche bayésienne du modèle LDA	40
3.1 Les principes de la statistique bayésienne	40
3.2 Estimation de la loi a priori	41
3.2.1 Les distributions conjuguées	41
3.2.2 Les lois non informatives	43
3.3 Difficultés préliminaires et choix de modélisation	44
3.3.1 La définition de la loi a priori	44
3.3.2 Le calcul de la loi a posteriori	44

3.4	Mise en place de la méthode 1 : utilisation des lois conjuguées	45
3.4.1	La loi de fréquence des sinistres : loi de Poisson	45
3.4.2	La loi de sévérité des pertes : loi log-normale $\mathcal{LN}(\mu, \sigma)$	51
3.5	Calcul des charges en capital CaR et sensibilité	59
3.5.1	Calcul des charges en capital	59
3.5.2	Sensibilités aux paramètres d'entrée du modèle	61
3.6	Discussion sur la méthode	63
4	Application de la théorie de la crédibilité	64
4.1	Principes théoriques	64
4.1.1	Notion de crédibilité	64
4.1.2	Le modèle de Bühlmann	65
4.1.3	Le modèle de Bühlmann-Straub	66
4.2	Adaptation du modèle de Bühlmann-Straub au modèle LDA	67
4.2.1	Modélisation de la fréquence des pertes	67
4.2.2	Modélisation de la sévérité des pertes	69
4.2.3	Application numérique aux données	70
4.2.4	Calcul des charges en capital	73
4.3	Adaptation de la méthode à une loi log-normale	75
4.3.1	Nouvelle modélisation	75
4.3.2	Application numérique aux données	78
4.3.3	Calcul des charges en capital	79
4.3.4	Sensibilités aux paramètres d'entrée du modèle	80
4.4	Discussion sur la méthode	81
5	Approfondissement de la théorie bayésienne	84
5.1	Présentation de la cascade bayésienne	84
5.1.1	Détermination des seuils L_j	85
5.1.2	Les bases de la théorie des valeurs extrêmes (TVE)	86
5.1.3	Méthodes d'estimation de L	87
5.1.4	Application aux données du Crédit Agricole	88
5.1.5	Implémentation de la cascade bayésienne	89
5.1.6	Application numérique aux données	90
5.1.7	Sensibilités aux paramètres d'entrée du modèle	92
5.1.8	Discussion sur la cascade bayésienne	93
5.2	Une alternative aux lois conjuguées : application des méthodes de Monte Carlo par chaînes de Markov	93
5.2.1	Principe des méthodes MCMC	94
5.2.2	Algorithmes utilisés pour la MCMC	94
5.2.3	Application aux données du Crédit Agricole	96
5.2.4	Sensibilités aux paramètres	99
5.2.5	Discussion sur la méthode MCMC	99
6	Synthèse des méthodes et perspectives	102
6.1	Comparaison des méthodes	102
6.1.1	Informations et données utilisées pour la mise en œuvre des méthodes	102
6.1.2	Résultats numériques	103
6.2	Les critiques du mémoire	104
6.3	Prolongements du mémoire	105
6.3.1	Une alternative non bayésienne : les lois de mélange	105
6.3.2	Une seconde alternative : les réseaux bayésiens	106
6.3.3	Amélioration de la modélisation LDA	107

Conclusion	110
A Quelques caractéristiques sur les lois usuelles utilisées	113
A.1 Les lois discrètes	113
A.1.1 La loi de Poisson	113
A.2 Les lois continues	114
A.2.1 La loi uniforme	114
A.2.2 La loi Gamma	115
A.2.3 La loi Inverse Gamma	115
A.2.4 La loi normale	116
A.2.5 La loi log-normale	116
A.2.6 La loi de Pareto	117
A.2.7 La loi GPD	118
B Quelques éléments sur la théorie des valeurs extrêmes	119
B.1 Méthode de calcul du seuil L avec la fonction <i>Mean excess</i>	119

Table des figures

1.1	Répartition des fonds propres à allouer dans une banque	13
1.2	Répartition du capital alloué au risque opérationnel sous AMA [29]	16
2.1	Comparaison des fonctions de répartition empirique et théorique	36
3.1	Représentation schématique de la méthode 1 pour la fréquence des pertes	49
3.2	Comparaison des densités de la fréquence pour le cas 2	50
3.3	Représentation schématique de la méthode 1 pour la sévérité dans le cas σ connu	54
3.4	Représentation schématique de la méthode 1 pour la sévérité dans le cas où μ et σ sont inconnus	57
3.5	Evolution de la distribution de sévérité	59
3.6	Cas de calcul de la CaR pour la méthode 1	59
3.7	Evolution de la CaR en fonction de $Vco(\Lambda)$	61
3.8	Evolution de la CaR en fonction de $Vco(\Theta_\mu)$	62
3.9	Evolution de la CaR en fonction de $Vco(\Theta_\sigma^2)$	62
4.1	Représentation schématique des données utilisées pour la crédibilité	70
4.2	Evolution des poids de crédibilité en fonction du seuil L	80
4.3	Evolution de la CaR_{CLN} en fonction du seuil L	81
5.1	Répartition des différents types de pertes (source :[18])	85
5.2	ME-plot respectivement pour CACIB EL6, CACF EL6, AMUNDI EL6	88
5.3	Hill-plot respectivement pour CACIB EL6, CACF EL6, AMUNDI EL6	89
5.4	Représentation schématique de la cascade bayésienne	90
5.5	Influence des coefficients $Vco(\Theta_\mu)$ et $Vco(\Theta_\sigma^2)$ sur la $CaR_{cascade}$ pour le cas 1	92
5.6	Influence du seuil L_1 sur la $CaR_{cascade}$ pour les cas 1 et 2	93
5.7	Allure de la chaîne de Markov $\lambda^{(m)}$ pour le cas 2	98
5.8	Allure des chaînes de Markov $\mu^{(m)}$ (en bleu) et $\sigma^{(m)}$ (en rouge) pour le cas 2	98
A.1	Densité de la loi de Poisson pour différentes valeurs de λ	114
A.2	Densité de la loi lognormale pour différentes valeurs de μ ($\sigma = 1$ à gauche) et de σ ($\mu = 1$ à droite)	116
A.3	Densité de la loi Pareto pour différentes valeurs de α	117

Liste des tableaux

1.1	Répartition des β suivant les lignes de métiers	15
2.1	Présentation des données par entité et type de risque selon que la cellule de risque dispose de données internes (\bullet) ou de scénarios (\times)	33
2.2	Cellules de risques choisies pour notre première étude	34
2.3	Présentation des données traitées	35
2.4	Statistiques descriptives	35
2.5	Estimation des paramètres de loi à partir des données internes	35
2.6	Charges en capital obtenues à partir des données internes	36
2.7	Couples (x, d) fournis par les experts métiers	36
2.8	Estimation des paramètres de loi à partir des scénarios	37
2.9	Charges en capital obtenues à partir des scénarios	37
2.10	Méta-matrice de corrélation pour les entités AMUNDI, CACF et CACIB	37
2.11	Charges en capital agrégées	38
3.1	Les lois conjuguées classiques	42
3.2	Valeurs des paramètres a priori pour la fréquence	49
3.3	Valeurs des paramètres a posteriori pour la fréquence	50
3.4	Valeurs finales du paramètre de fréquence $\hat{\lambda}_{posteriori}$	50
3.5	Valeurs des paramètres a priori pour la sévérité	58
3.6	Valeurs des paramètres a posteriori pour la sévérité	58
3.7	Valeurs finales des paramètres de sévérité $\hat{\mu}_{posteriori}$ et $\hat{\sigma}_{posteriori}$	58
3.8	Montants des différentes charges en capital calculés selon la méthode 1 des lois conjuguées	60
4.1	Les paramètres à prendre en compte dans le modèle de Bühlmann-Straub	66
4.2	Calcul des CaR_{CLN}	79
5.1	Récapitulatif des données internes et externes utilisées pour la cascade bayésienne	89
5.2	Paramètres structuraux des lois a posteriori 1	90
5.3	Paramètres structuraux des lois a posteriori 2	91
5.4	Paramètres de la loi de sévérité obtenus avec la cascade bayésienne	91
5.5	Charges en capital obtenues avec la cascade bayésienne	91
6.1	Récapitulatif des montants de charges en capital obtenus avec les différentes approches	103

Introduction

L'industrie financière, en particulier les banques et les assurances, subit aujourd'hui des changements majeurs. L'instabilité chronique des marchés financiers de ces dernières années en atteste. La chute désormais célèbre de la banque d'affaires Lehman Brothers en 2008, qui était pourtant jugée "too big to fail" a mis en évidence la fragilité du système financier. Le besoin d'une meilleure couverture des risques bancaires s'est donc fait sentir. L'objectif visé était de limiter les faillites des banques qui peuvent avoir un effet de contagion et menacer la stabilité du système financier dans son ensemble.

Suite à cette préoccupation grandissante pour la gestion des risques, une réglementation spécifique a été mise en place, qui est représentée aujourd'hui par l'accord de Bâle II en ce qui concerne les banques. Celles-ci ont été contraintes de mettre en place des mesures visant à quantifier de manière appropriée leur exposition au risque et à assurer son pilotage. C'est dans ce contexte que le risque opérationnel a commencé à être considéré comme un risque majeur. En effet, le risque opérationnel a pris une importance croissante ces dernières années, principalement en raison des modifications du cadre d'exercice et de la conduite des activités bancaires. Cet essor est notamment lié à la banalisation de la gestion en temps réel des opérations, engendrant un risque de règlement, dans un contexte de globalisation et d'internationalisation des activités. Il est par ailleurs le résultat de la sophistication de ces activités, tant dans la conception de nouveaux produits auxquels sont associés des risques juridiques, que dans la mise en place de systèmes d'information de plus en plus complexes.

Cette importance croissante s'est largement concrétisée ces dernières années. Les pertes subies par les établissements au titre du risque opérationnel sont en effet généralement évaluées à plus de 200 milliards d'euros sur la période 1980-2000. Plus récemment, l'exercice de collecte de pertes réalisé en 2002 par le groupe *Risk Management* du Comité de Bâle révèle que 89 banques ont connu sur le seul exercice 2001 plus de 47 000 événements de pertes pour un montant cumulé de pertes opérationnelles s'élevant à près de 7,8 milliards d'euros. Bien que des mesures aient été instaurées à partir de 2001, l'exemple le plus célèbre de perte opérationnelle est intervenu en janvier 2008. Un trader de la Société Générale, Jérôme Kerviel, a été l'auteur d'opérations frauduleuses sur les marchés qui ont engendré une perte d'un montant total de 4,9 milliards d'euros.

Comme l'illustre cet exemple, la grande difficulté qui réside dans la gestion des risques est la capacité à anticiper des événements rares ayant des conséquences extrêmement lourdes pour les banques. C'est dans l'objectif de couvrir ces pertes à caractère imprévisible qu'elles ont développé de nombreuses méthodologies et modèles statistiques dans leurs modèles internes. Les banques veulent adopter une approche plus pragmatique en terme de risque opérationnel en rééquilibrant le dispositif vers la gestion des risques plutôt que vers leur seule mesure. C'est pourquoi les banques exploitent non seulement leurs données historiques mais elles prennent en compte également des jugements d'experts qui induisent des critères qualitatifs. Les deux sources de modélisation (quantitative et qualitative) permettent alors d'offrir une vision complète d'un

type de risque et gagnent donc à être combinées pour former une distribution de pertes unique. La traduction quantitative des jugements d'experts est cependant très ardue car ils ne s'inscrivent pas dans une démarche bien structurée et homogène. Ces données sont présentées actuellement sous la forme de scénarios.

La statistique bayésienne propose des méthodes permettant de combiner les informations issues de plusieurs sources de données. L'objet de ce mémoire est d'étudier les différentes modélisations issues de la théorie bayésienne qu'il existe dans la littérature et de les adapter au contexte du risque opérationnel. Ensuite, l'objectif est de les comparer en soutirant de ces méthodes leurs avantages et inconvénients et de proposer en fin du mémoire une réflexion d'une part sur l'utilisation du bayésien dans le calcul réglementaire du capital et d'autre part sur les perspectives potentielles qui permettraient de poursuivre notre travail.

Ce rapport est organisé en six parties. La partie 1 est entièrement consacrée à la définition du contexte en présentant de manière détaillée le risque opérationnel et les différentes approches pour le modéliser. On fait également le rapprochement dans cette partie avec le risque opérationnel en assurance avec en particulier l'introduction de la réforme européenne Solvabilité II. La partie 2 traite de la méthode actuelle employée au sein du Crédit Agricole pour calculer le capital nécessaire à couvrir le risque opérationnel ; dans cette partie sont par ailleurs présentées les données sur lesquelles on travaillera par la suite. Une fois ces contextes bien présentés, on redéfinira précisément les objectifs de notre étude. La partie 3 expose notre étude complète sur la théorie de Bayes qu'on appellera approche bayésienne pure. La partie 4 regroupe notre travail sur l'adaptation de la théorie de la crédibilité. Dans la partie 5, on approfondit les méthodes mentionnées précédemment sous la forme de deux autres méthodes : d'une part la "cascade bayésienne" qui se base essentiellement sur l'approche bayésienne pure et utilise la théorie des valeurs extrêmes ; d'autre part une méthode qui permet de s'affranchir de certaines hypothèses faites dans l'approche bayésienne pure. Enfin, la dernière partie dresse un bilan de l'étude ainsi que des préconisations.

Partie 1

Le risque opérationnel : présentation du contexte

Comme l'explique le Comité de Bâle dans [11], la réglementation nécessite l'estimation de risques complexes et dans le contexte de mondialisation des services financiers et de sophistication grandissante des techniques financières, cette estimation est de plus en plus complexe. Parmi ces risques nouveaux, on peut citer l'automatisation des techniques qui peut transformer les risques d'erreurs humaines en risques de plus grande portée ou encore les techniques d'atténuation du risque (dérivés de crédit, titrisation,...) qui peuvent engendrer des risques juridiques.

A ce titre, le Comité de Bâle a défini le risque opérationnel pour traiter ces pratiques.

1.1 Définition

Le Comité de Bâle a défini de manière précise le risque opérationnel dans [12] :

"Le risque opérationnel se définit comme le risque de pertes résultant de carences ou de défauts attribuables à des procédures, personnels et systèmes internes ou à des événements extérieurs. La définition inclut le risque juridique, mais exclut les risques stratégique et de réputation."

Le risque juridique inclut, entre autres, l'exposition à des amendes, pénalités et dommages pour faute résultant de l'exercice de surveillance prudentielle ainsi que des transactions privées. Parmi les types d'incidents de nature opérationnelle susceptibles d'occasionner de lourdes pertes, le Comité a identifié sept typologies de risques :

- **Fraude interne** : cela regroupe par exemple les informations inexactes sur les positions, les vols commis par un employé et délits d'initié d'un employé opérant pour son propre compte ;
- **Fraude externe** : par exemple, un hold-up, un faux en écriture, des dommages dus au piratage informatique ;
- **Pratiques en matière d'emploi et sécurité sur le lieu de travail** : on distingue ici les demandes d'indemnisation des travailleurs, les violations des règles de santé et de sécurité

des employés, les activités syndicales, ou encore les plaintes pour discrimination et responsabilité civile en général ;

- **Pratiques concernant les clients, les produits et l'activité commerciale** : tout ce qui relève de l'utilisation frauduleuse d'informations confidentielles sur la clientèle, les opérations boursières malhonnêtes pour le compte de la banque, le blanchiment d'argent ou la vente de produits non autorisés ;
- **Domage aux biens physiques** : par exemple, les actes de terrorisme, de vandalisme, ou les séismes, incendies et inondations ;
- **Interruption d'activité et pannes de systèmes** : cette catégorie rassemble toutes les pannes de matériel et de logiciel informatiques ainsi que tous les problèmes dus aux télécommunications et aux pannes d'électricité ;
- **Exécution des opérations, livraisons et processus** : enfin, on distingue ici les erreurs d'enregistrement des données, les défaillances dans la gestion des sûretés, les lacunes dans la documentation juridique, les erreurs d'accès aux comptes de la clientèle et les défaillances des fournisseurs ou conflits avec eux.

Les dernières réglementations à suivre par les banques concernant le risque opérationnel sont dans les textes de la réglementation Bâle II mise en service en 2004 (voir Comité de Bâle [10]).

1.2 Bâle II et le risque opérationnel

1.2.1 Présentation de Bâle II

Les régulateurs exigent des banques d'allouer un certain montant de fonds propres pour prendre connaissance des pertes potentielles et celles non attendues en particulier. Les différents risques à considérer sont le risque de crédit, le risque opérationnel ainsi que le risque de marché, comme le montre le graphique suivant (illustration provenant de ONB [23]) :

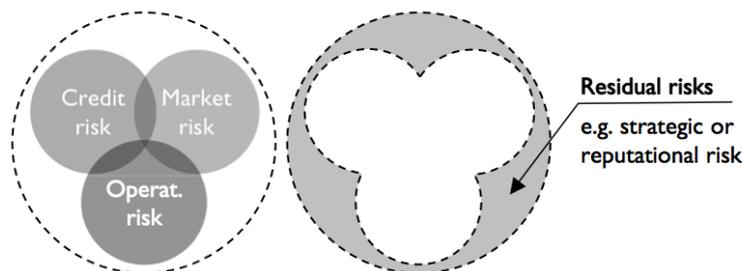


FIGURE 1.1 – Répartition des fonds propres à allouer dans une banque

Le risque opérationnel est significatif dans la plupart des institutions financières : il représente environ une part de 15 à 25 % du capital total dans la plupart des grosses banques. C'est le deuxième risque le plus important après le risque de crédit. L'environnement défini par Bâle II repose sur trois piliers :

- **Pilier I.** C'est un pilier quantitatif qui exige le calcul explicite d'un niveau de fonds propres minimum pour couvrir l'ensemble du risque opérationnel. Les différentes méthodes possibles pour le calculer sont présentées dans la section suivante ;
- **Pilier II.** C'est un pilier qualitatif qui se concentre sur la supervision des systèmes bancaires et sur la validation des fonds propres par les autorités réglementaires ;
- **Pilier III.** Il s'agit d'un pilier qualitatif dont l'objectif est d'établir une certaine discipline par le biais de définition de mesures de risques et de diffusion d'articles sur la gestion du risque opérationnel.

Une banque doit disposer d'une fonction de gestion de risque opérationnel bien identifiée, responsable de la conception et de la mise en œuvre du dispositif de mesure et de gestion de ce risque. Ce dispositif doit être intégré à la gestion quotidienne des risques de l'établissement et le risque encouru doit faire l'objet de comptes rendus adéquats. Il doit aussi faire l'objet d'un examen périodique des auditeurs. Par ailleurs, Bâle II reprend la cartographie des risques présentés dans la section précédente, rappelant ainsi l'importance tant de l'implication de l'organe exécutif dans la mise en place d'un tel dispositif que de l'identification des risques.

Dans la suite de notre mémoire, nous nous concentrerons uniquement sur l'aspect quantitatif du risque opérationnel, à savoir le pilier I.

1.2.2 Les méthodes pour quantifier le risque opérationnel

Le dispositif d'adéquation des fonds propres offre aux banques la possibilité de mesurer leur exposition au risque opérationnel selon trois approches, par ordre croissant de complexité et de sensibilité au risque : l'approche de l'indicateur de base (*Basic Indicator Approach* ou **BIA**), l'approche standard (*The Standardised Approach* ou **TSA**) et l'approche des mesures avancées (*Advanced Measurement Approach* ou **AMA**). Le dispositif incite par ailleurs les banques à passer des approches les plus simples (BIA et TSA) aux plus avancées (AMA) et à développer ainsi des systèmes et des pratiques plus sensibles aux risques, adaptés à leur profil de risque et à la complexité de leurs activités. Banque de France [3] détaille ces différentes approches tout en faisant un bilan de la situation actuelle.

L'approche de l'indicateur de base (BIA)

Les banques utilisant l'approche BIA doivent détenir des fonds propres correspondant à un pourcentage fixe α de leur produit brut bancaire moyen PB sur les trois dernières années :

$$FP_{BIA} = \alpha PB$$

Le produit net bancaire correspond aux produits d'intérêts nets et autres produits d'exploitation. Il doit être positif ; s'il s'avérait être nul ou négatif une année, on fait la moyenne des produits bancaires sur l'ensemble des années où il est strictement positif.

L'approche standard (TSA)

Les banques utilisant l'approche standard doivent détenir des fonds propres correspondant pour chacune de leurs huit lignes de métier à un pourcentage fixe (β_i) de leur produit brut

bancaire moyen sur les trois dernières années :

$$FP_{TSA} = \sum_{i=1}^8 \beta_i PB_i$$

où les β_i sont distribués de la manière suivante :

N°	Lignes de métiers	β
1	Financement des entreprises	18%
2	Négociation et vente	18%
3	Banque de détail	12%
4	Banque commerciale	15%
5	Paieement et règlement	18%
6	Fonctions d'agent	15%
7	Gestion d'actifs	12%
8	Courtage de détail	12%

TABLE 1.1 – Répartition des β suivant les lignes de métiers

De la même manière qu'avec l'approche basique, $PB > 0$. Cependant, les exigences de fonds propres négatives dans une ligne de métier peuvent compenser les exigences de fonds positives sur d'autres lignes. Ainsi si la somme annuelle des PB_i est négative, le montant ne sera pas pris en compte dans le calcul final du capital.

L'approche des mesures avancées (AMA)

La banque calcule l'exigence en fonds propres nécessaire en utilisant son propre modèle interne qui doit être approuvé par l'autorité de régulation. Le modèle interne doit inclure à la fois des données historiques (données de pertes internes et externes) et des données prospectives (analyse de scénarios d'événements potentiels, facteurs d'environnement et de contrôle interne). Quel que soit le modèle développé, la mesure du risque opérationnel doit être d'une fiabilité comparable à celle du risque de crédit (période de détention d'un an et intervalle de confiance de 99,9%).

Ces trois approches ont pour but de quantifier le risque opérationnel avec une sensibilité variable et donc de contribuer à une meilleure surveillance de ce dernier. En effet, un risque ne peut être correctement maîtrisé que s'il est identifié, mesuré, évalué et géré. C'est pourquoi, le Comité de Bâle a développé deux autres piliers qui permettent de développer les principes nécessaires à une bonne maîtrise des risques opérationnels, en soumettant l'utilisation des approches de calcul des fonds propres au respect de critères qualitatifs, notamment en matière de gouvernance, d'audit et de contrôle interne. D'ailleurs, les fonds propres réservés pour le risque opérationnel doivent respecter le ratio de MacDonough :

$$\frac{FPrélementaires}{Risque_{crédit} + Risque_{marché} + Risque_{opérationnel}} \geq 8\%$$

Ici, on ne s'intéresse qu'à l'approche AMA. Pour cela, les informations nécessaires à l'élaboration d'un modèle interne sont les données internes, les données externes pertinentes, les analyses

de scénarios et les facteurs reflétant l'activité qui doivent être adaptés aux systèmes internes de contrôle. La mesure de risque utilisée dans le capital réglementaire doit correspondre à un niveau de confiance de 99.9% pour une année. Les bénéfices induits par la diversification sont acceptés à condition que le modèle de dépendance soit approuvé par le régulateur. La réduction de capital grâce aux assurances ne peut dépasser 20%.

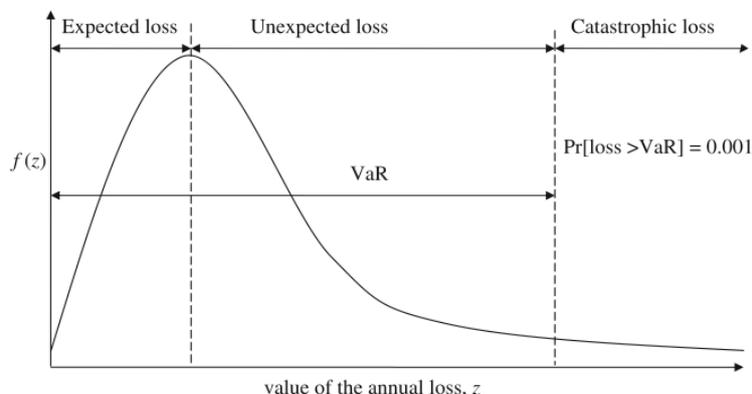


FIGURE 1.2 – Répartition du capital alloué au risque opérationnel sous AMA [29]

Des études quantitatives d'impact (QIS) ont été fournies aux banques pour que les autorités de contrôle aient pu définir les différentes directives de Bâle II et estimer les tendances générales au sein des lignes d'activités.

1.3 Rapprochement avec Solvabilité II

La réglementation Solvabilité II, initialisée par le CEIOPS (Comité Européen des Contrôleurs d'Assurance et des Pensions Professionnelles) s'apparente à celui mis en place dans les banques avec la réforme Bâle II. De la même manière, Solvabilité II est un dispositif à trois piliers :

- Le **Pilier I** définit des règles quantitatives avec principalement une évaluation prospective des provisions techniques, deux niveaux de besoin en capital (le minimum de capital requis MCR (Minimum Capital Requirement) et le capital de solvabilité requis SCR (Solvency Capital Requirement)) et des règles de couverture des engagements ;
- Le **Pilier II** impose un processus de contrôle prudentiel avec un renforcement des contrôles en matière de gestion des risques, contrôle interne, gouvernance, processus de supervision et pouvoirs d'action du superviseur ;
- Le **Pilier III** a vocation à rendre la communication externe des sociétés plus transparente et plus harmonisée au niveau européen en fixant les exigences en matière d'information publique, d'information aux assurés et d'information prudentielle.

De la même manière que Bâle II, Solvabilité II introduit le risque opérationnel qui intervient dans le pilier I avec la quantification du besoin en fonds propres et dans le pilier II avec la mise en place de leur gestion. Actuellement, le régulateur propose aux compagnies d'assurance deux approches pour quantifier leur risque opérationnel : une approche standard et une approche avancée. L'approche standard est une approche simplifiée calculée comme un pourcentage des

primes ou des provisions. L'approche avancée est un modèle interne où les risques correspondent réellement à la situation de l'entreprise. Le régulateur incite également les compagnies à opter pour le modèle interne en rendant l'approche standard beaucoup plus consommatrice en fonds propres. Ces spécifications sont détaillées dans CEIOPS [8] et EIOPA [14].

La définition du risque opérationnel sous Solvabilité II est identique :

"Le risque opérationnel est le risque de pertes résultant de procédures internes inadaptées ou défaillantes, ou de membre du personnel et de systèmes, ou d'événements extérieurs. Le risque opérationnel ainsi défini comprend les risques juridiques mais non les risques découlant de décisions stratégiques ni les risques de réputation." (*Directive cadre du 10 juillet 2007, article 13 et 100.*)

Pour les grosses assurances, le risque opérationnel représente beaucoup de poids dans le calcul du capital réglementaire : c'est l'un des trois plus gros risques. Dans Solvabilité II, les lignes de métier et la typologie des risques n'est pas précisée.

1.3.1 La formule standard

Afin de mesurer l'impact de Solvabilité II sur les compagnies d'assurance, la Commission Européenne a chargé le CEIOPS de réaliser des QIS. La dernière réalisée est le QIS5 : cette étude permet de définir le $SCR_{opérationnel}$ comme le minimum entre 30% du $BSCR$ (le SCR de base) et le montant de capital Op pour couvrir le risque opérationnel auquel s'ajoute des ajustements liés aux effets absorbants du risque des provisions techniques et de certaines taxes. Le BSCR correspond au capital de solvabilité avant ajustements permettant de couvrir les risques de souscription en vie, en non-vie, en santé, le risque de marché, le risque de contrepartie ainsi que le risque associé aux actifs intangibles. On obtient ainsi la formule suivante :

$$SCR_{op} = \min(0.3BSCR, Op) + 0.25Exp_{ul}$$

avec $Op = \text{Max}(Op_{primes}, Op_{provisions})$. Op est un capital qui se déduit comme un maximum entre une pondération des primes et une pondération des provisions de l'assureur (liées aux contrats en euros) et Exp_{ul} correspond au montant annuel des frais supportés pendant les 12 derniers mois pour ce qui concerne les obligations d'assurance-vie où le risque d'investissement est porté par le souscripteur (contrats *unit-linked*).

1.3.2 Les modèles internes

Le $SCR_{opérationnel}$ est calculé sur un intervalle de confiance de 99.5%, à horizon d'un an. La possibilité de prendre en compte des couvertures d'assurance n'est pas précisée. Pour une étude approfondie des modèles internes du $SCR_{opérationnel}$, de nombreux articles ont été publiés, dont Metametrus [21] et J. Gamonet [17].

1.4 Les différents modèles AMA

L'AMA proposée par le Comité de Bâle suppose l'utilisation par les établissements bancaires de quatre types de données : des données de pertes internes, de pertes externes, des analyses

de scénarios et d'événements potentiels et des facteurs d'environnement et de contrôle interne. Le Comité de Bâle propose trois méthodes différentes mais également d'autres approches sont utilisées à l'intérieur des banques. Banque de France [3] et P.V. Shevchenko [29] reprennent ces méthodes.

1.4.1 Les approches proposées par le Comité de Bâle

L'Internal Measurement Approach (IMA)

On suppose qu'il existe une relation (linéaire ou non) entre la mesure de risque et la perte espérée. Pour une catégorie d'activité i et un type de risque j , la charge en capital est définie par exemple de la façon suivante :

$$FP(i, j) = EL(i, j) \times \gamma(i, j)$$

où $EL(i, j)$ représente la perte moyenne et γ est un facteur d'échelle. Notons que l'on peut calculer la perte moyenne comme le produit de trois termes :

$$EL(i, j) = IE(i, j) \times PE(i, j) \times LGE(i, j)$$

où $IE(i, j)$ est l'indicateur d'exposition, $PE(i, j)$ est la probabilité d'occurrence d'une perte unitaire et $LGE(i, j)$ est le montant de la perte unitaire. La grande difficulté est l'estimation du paramètre γ .

L'approche Scorecard

Cette méthode consiste à se donner pour un type de risque ou pour une ligne de métier une valeur initiale de l'exigence en capital pour couvrir le risque opérationnel, puis de modifier cette valeur à l'aide de méthodes de notations (dites de *scoring*). Ces notations anticipent l'évolution de la gestion des risques au sein des lignes de métier (par exemple), et donc de pouvoir donner une meilleure exigence en capital pour les années futures. Le Comité de Bâle n'a aucune formulation mathématique pour cette approche. Néanmoins, des formules circulaient à l'époque à l'intérieur des banques. Elles étaient de la forme :

$$FP(i, j) = IE(i, j) \times \omega(i, j) \times RS(i, j)$$

avec IE l'indicateur d'exposition, RS le score de risque et ω un facteur d'échelle.

La Loss Distribution Approach (LDA)

La dernière méthode est celle qui a été retenue par le Crédit Agricole et développée par le GRO ainsi que par la majorité des grands groupes bancaires. Elle permet une meilleure appréciation du risque opérationnel et présente l'avantage de réduire l'exigence en fonds propres de la banque. Elle a été validée au sein de Crédit Agricole S.A. par la commission bancaire en 2008. Cette démarche s'appuie sur des modèles statistiques appliqués aux données internes et aux scénarios d'experts : elle sera détaillée par la suite.

1.4.2 L'approche top-down

Certains établissements ont développé un modèle essentiellement statistique de calcul de fonds propres en s'appuyant sur des données de pertes internes et externes et en utilisant une méthode du type *Value-at-Risk* (horizon d'un an et intervalle de confiance de 99.9%). Pour ces derniers, en complément au modèle théorique, la mise en œuvre d'une AMA est tout d'abord l'occasion de développer et d'améliorer un système informatique de collecte systématique de pertes. L'utilisation de données internes de pertes pose des questions théoriques et pratiques en particulier relatives au niveau à partir duquel toute perte doit être collectée et la façon dont celle-ci doit être capturée pour assurer une remontée correcte des données recherchées et une distribution crédible des pertes.

L'utilisation de données de pertes internes comme source première du modèle interne relève d'une approche **top-down**, où les risques opérationnels sont identifiés et mesurés sur une base consolidée à partir des effets (les pertes) et où les fonds propres sont ensuite alloués aux différentes lignes de métier. L'utilisation de données externes soulève de la part des établissements des interrogations relatives à la nécessaire correction de biais statistiques et aux modalités d'adaptation de ces données à leur situation interne (problèmes de *scaling*).

Dans ces modèles, on n'essaie pas de modéliser les types de risque individuellement. On peut citer comme modèles employés :

- **Les modèles multifacteurs de tarification de l'equity.** Cette approche fait l'hypothèse d'efficience du marché où le prix actuel de l'actif reflète toute l'information pertinente. Le processus qui permet d'avoir les prix est supposé dépendre de facteurs dus au marché, au crédit et autres risques non opérationnels. Le terme résiduel de cette régression est traité comme du risque opérationnel. On peut alors établir des dépendances entre les risques opérationnels et les variables macroéconomiques.
- **Les CAPMs (*Capital Asset Pricing Models*).** Ici, on modélise la prime de risque associée aux actifs. On peut alors traiter les contributions du risque de crédit et du risque de marché, et on traite de la même manière le risque opérationnel comme le résidu.
- **Les modèles à indices de risque.** Ces modèles lient le risque opérationnel et des indices d'exposition comme le volume des transactions, le nombre de personnel, etc.

1.4.3 L'approche bottom-up

D'autres établissements bâtissent leur modèle de mesure en privilégiant davantage des données prospectives, de type analyse de scénarios et indicateur de risque. Dans ce cas, l'approche se veut **bottom-up**, les risques étant cartographiés au niveau de chaque ligne de métier à partir des causes, puis mesurés sur la base de fréquences et de sévérités de pertes estimées par les experts de chaque métier et/ou d'indicateurs de performance, de contrôle et de risque. Certaines banques recourent aux analyses de scénarios pour les seuls événements à faible probabilité et à forte sinistralité, d'autres de manière plus systématique pour tous les types d'événements identifiés. Dans tous les cas, la démarche doit être suffisamment structurée et cohérente dans le groupe afin que les quantifications subjectives des risques au niveau des métiers puissent alimenter correctement le modèle de calcul des fonds propres. L'utilisation d'indicateurs de risque est également privilégiée par certains établissements soucieux de conférer à leur modèle un caractère aussi prédictif

que possible, dans une logique de prévention des risques. En pratique, l'identification des indicateurs de risque s'effectue à partir de risques identifiés lors de la cartographie et par rapport à des indicateurs existants (indicateurs de qualité, de performance,...). Parmi les difficultés qui pourraient émerger de ce type de modélisation, on distingue entre autres la définition de niveaux d'alertes cohérents avec la politique de risques du groupe ainsi que les modalités d'agrégation des indicateurs.

La mise en oeuvre d'une AMA par les banques combinant pleinement données historiques (données de pertes internes et externes) et données prospectives (analyses de scénarios et indicateurs de risque) semble s'accélérer. Ceci résulte de l'insuffisance quantitative et conceptuelle des seules données historiques ainsi que des progrès réalisés dans la structuration des démarches plus qualitatives de type bottom-up.

Il existe deux types de modèles existants pour ce type d'approche : les modèles basés sur des processus et les modèles LDA (*Loss Distribution Approach*) qui sont les seuls modèles auxquels on s'intéressera le long de notre étude. C'est pourquoi on ne présente que dans cette section le premier type de modèle qui se décompose en plusieurs alternatives :

- **Les réseaux causaux.** Ce sont des modèles subjectifs, liés aux approches dites *scorecard*. Pour chaque ligne d'activité de la banque, un arbre d'événements pouvant conduire à une perte opérationnelle est construit. La probabilité de chaque événement est spécifiée par un expert. Typiquement, les réseaux bayésiens sont utilisés pour quantifier la probabilité future de la perte. Ces réseaux permettent également d'établir le lien entre la probabilité et la sévérité des pertes.
- **Les modèles causaux multifacteurs.** Ils s'appuient sur la régression de la perte opérationnelle sur un nombre de variables explicatives comme le nombre de personnel, le niveau de compétence, etc. Ces facteurs sont utilisés pour prédire les pertes futures en supposant qu'ils sont connus durant la période d'estimation.
- **Les modèles de fiabilité.** Ces modèles permettent de quantifier la probabilité qu'un système opérera de manière satisfaisante pendant une certaine période. Ce type de modélisation est pertinent dans certains processus comme la modélisation de la fiabilité des systèmes de transactions.

A présent, nous nous concentrerons uniquement sur le modèle LDA du Crédit Agricole.

Partie 2

Le modèle LDA : Loss Distribution Approach

2.1 Description générale du modèle

L'objectif du modèle LDA est de modéliser la perte liée au risque opérationnel pour une période d'une année afin d'en déduire la valeur en risque. La particularité de cette perte est qu'elle résulte de plusieurs pertes individuelles. Beaucoup de banques utilisent ce modèle qui a été exposé dans de nombreux articles : F. Aue, M. Kalkbrener [2], A. Frachot et al. [15], A. Frachot et al. [16] et P.V. Shevchenko [32].

Pour modéliser la perte totale liée à un type de risque donné, on distingue deux dimensions :

- la **fréquence** des pertes, qui correspond au nombre de pertes individuelles, a priori inconnu pour l'année à venir, et que l'on modélise par un processus de comptage ;
- la **sévérité** des pertes, qui modélise le montant de chacune de ces pertes individuelles.

La perte annuelle est donc la somme aléatoire de ces pertes individuelles. Sa distribution est construite à partir des distributions de fréquence et de sévérité. La valeur en risque est ensuite calculée par une méthode de Monte Carlo. L'objectif est alors d'estimer avec précision l'exigence de fonds propres réglementaires pour une période de détention fixée à un an et un niveau de confiance $\alpha = 99.9\%$. Cela correspond à la couverture d'un risque millénaire.

On considère I catégories d'activités différentes ($i = 1, \dots, I$) correspondant aux *Business Lines* et J types de risques ($j = 1, \dots, J$) correspondant aux *Event Lines*, $X_{i,j}$ est la variable aléatoire représentant le montant d'une perte pour l'activité i et le type de risque j , dont la fonction de répartition est notée $F_{i,j}$. Le nombre d'occurrences sur une année est modélisé par une variable de comptage $N_{i,j}$, de fonction de probabilité $p_{i,j}$. La fonction de répartition de la fréquence des pertes $P_{i,j}$ correspond alors à $P_{i,j}(n) = \sum_{k=1}^n p_{i,j}(k)$.

La perte pour la ligne métier i et le type de risque j sur une année vaut donc :

$$L_{i,j} = \sum_{k=1}^{N_{i,j}} X_{i,j}^k$$

et sa fonction de répartition est donc la distribution composée suivante :

$$F_{L_{i,j}}(x) = \mathbb{P}(L_{i,j} \leq x) = \begin{cases} \sum_{k=0}^{\infty} p_{i,j}(k) F_{i,j}^{k*}(x) & \text{si } x > 0 \\ p_{i,j}(0) & \text{si } x = 0 \end{cases}$$

Où $F_{i,j}^{k*}$ est un produit de convolution d'ordre k qui correspond à la fonction de répartition de $X_{i,j}^1 + \dots + X_{i,j}^k$.

La charge en capital CaR, pour *Capital-at-Risk*, pour l'activité i et le type de risque j correspond alors au quantile α de $F_{L_{i,j}}$:

$$\text{CaR}_{i,j}(\alpha) = F_{L_{i,j}}^{-1}(\alpha) = \inf\{x | F_{L_{i,j}}(x) \geq \alpha\}$$

Les CaRs calculées au sein des différentes cellules (i, j) sont ensuite agrégées pour obtenir la CaR totale.

2.2 Calcul du capital réglementaire basé sur les données historiques

En tant qu'approche par distribution de pertes, la modélisation LDA s'appuie sur l'ensemble des pertes collectées dans les bases de données internes. D'un point de vue statistique, cette approche a été adaptée notamment pour tenir compte de deux contraintes liées au dispositif de collecte des pertes internes du Groupe :

- la définition d'un seuil de collecte : seules les pertes supérieures à un seuil H sont collectées pour conserver une qualité correcte des données ;
- la présence d'agrégats de pertes : pour certaines catégories de risque, les bases de données contiennent souvent des agrégats de pertes, c'est à dire qu'une perte individuelle de montant $X_{i,j}^k$ peut être composée de κ_k événements distincts dont on ignore les montants individuels (supérieurs à H).

2.2.1 Estimation de la distribution de sévérité

La troncature des données liée au seuil de collecte H affecte l'estimation des paramètres de sévérité puisque la distribution empirique (pertes effectivement collectées) est différente de la vraie distribution (celle que l'on doit obtenir si aucune des pertes n'est négligée). Il faut donc relier la vraie distribution à sa version empirique par le biais de la densité conditionnelle suivante :

$$\tilde{f}_{\theta|H}(x) = \frac{f_{\theta}(x)}{\int_H^{+\infty} f_{\theta}(u) du} \mathbb{1}_{\{x \geq H\}}$$

On suppose que l'on a collecté un échantillon de n pertes de montants X_1, \dots, X_n (supposées indépendantes et identiquement distribuées) au-delà du seuil H . La loi de probabilité adoptée pour la modélisation de la sévérité est la loi log-normale qui offre de bonnes qualités d'ajustement et on estime ses paramètres $\theta = (\mu, \sigma)$ par la méthode des moments généralisée (GMM) décrite par la suite.

Méthode des Moments Généralisée

La méthode des moments généralisée consiste à déterminer le vecteur de paramètres θ qui minimise l'écart entre les moments théoriques et les moments empiriques. À l'ordre p , les conditions de moments sont définies par :

$$g_p(\theta) = \frac{1}{n} \sum_{k=1}^n ((X_k)^p - \kappa_k m_p(\theta))$$

avec $m_p(\theta) = \mathbb{E}[X^p | X \geq H] = \int_{-\infty}^{+\infty} x^p \tilde{f}_{\theta|H}(x) dx$

En posant $g(\theta) = (g_1(\theta), \dots, g_p(\theta))'$, l'estimateur GMM est défini par :

$$\hat{\theta}_{GMM} = \arg \min_{\theta} g(\theta)' W^{-1} g(\theta)$$

où W est une matrice symétrique définie positive, optimale en prenant $W = \mathbb{V}[g(\theta)]$

En pratique, il faut au moins autant de conditions de moments que de paramètres à estimer. Dans le cas log-normal, $X \sim \mathcal{LN}(\mu, \sigma)$, cas de sévérité à deux paramètres, on considère ainsi les deux premiers moments conditionnels suivants :

$$\begin{cases} m_1(\theta) = \frac{1 - \Phi\left(\frac{\ln H - (\mu + \sigma^2)}{\sigma}\right)}{1 - \Phi\left(\frac{\ln H - \mu}{\sigma}\right)} e^{(\mu + \sigma^2)/2} \\ m_2(\theta) = \frac{1 - \Phi\left(\frac{\ln H - (\mu + 2\sigma^2)}{\sigma}\right)}{1 - \Phi\left(\frac{\ln H - \mu}{\sigma}\right)} e^{2(\mu + \sigma^2)} \end{cases}$$

où Φ désigne la fonction de répartition d'une loi normale centrée réduite.

Une fois les paramètres estimés par résolution du système précédent, l'ajustement des distributions de pertes via la loi log-normale est mesuré au moyen de quatre tests d'adéquation. Les statistiques de test utilisées prennent en compte la troncature des données au seuil H . On note $z_H = F_{\theta}(H)$ et $z_j = F_{\theta}(y_j)$ où y_j la $j^{\text{ème}}$ perte de l'échantillon.

- Test de Kolmogorov-Smirnov (KS)

Le test de Kolmogorov mesure l'écart maximal (en valeur absolue) entre la distribution empirique et la distribution estimée. Il s'agit donc d'un test local, et la statistique de test a pour expression :

$$\text{KS} = \frac{\sqrt{n}}{1 - z_H} \max\left\{\sup(z_H - z_j + \frac{j}{n}(1 - z_H)), \sup(z_j - z_H + \frac{j-1}{n}(1 - z_H))\right\}$$

- Test de Cramer-Von-Mises (CVM)

Le test de Cramer-Von-Mises est un test quadratique mesurant la moyenne du carré des écarts entre la distribution empirique et la distribution estimée, dans le cas où la pondération ψ vaut 1 quelle que soit l'observation. La statistique de test est définie par :

$$\text{CVM} = \frac{n}{3} + \frac{nz_H}{1 - z_H} + \frac{1}{n(1 - z_H)} \sum_{j=1}^n (1 - 2j)z_j + \frac{1}{(1 - z_H)^2} \sum_{j=1}^n (z_j - z_H)^2$$

○ Test d'Anderson Darling (AD)

Ce test est dérivé du test de Cramer-Von-Mises et correspond au cas où la pondération vaut $\psi(\widehat{F}(x)) = (\widehat{F}(x)(1 - \widehat{F}(x))^{-1})$, ce qui a pour effet d'accorder un poids plus important aux queues gauches et droites des distributions. La statistique de test vaut :

$$AD = -n + 2n \ln(1 - z_H) - \frac{1}{n} \sum_{j=1}^n (1 + 2(n - j)) \ln(1 - z_j) + \frac{1}{n} \sum_{j=1}^n n(1 - 2j) \ln(z_j - z_H)$$

○ Test d'Anderson Darling modifié (AD_{up})

Ce dernier test est basé sur le même principe que le précédent en considérant une pondération égale à $\psi(\widehat{F}(x)) = (\widehat{F}(x)(1 - \widehat{F}(x))^{-2})$, qui attribue un surcroît d'importance aux queues droites des distributions correspondant à la région des pertes les plus sévères :

$$AD_{up} = -2n \ln(1 - z_H) + 2 \sum_{j=1}^n \ln(1 - z_j) + \frac{1 - z_H}{n} \sum_{j=1}^n (1 - 2j) \frac{1 + 2(n - j)}{1 - z_j}$$

2.2.2 Estimation de la distribution de fréquence

La distribution de la fréquence est modélisée par une variable aléatoire N suivant une loi de Poisson $\mathcal{P}(\lambda)$, de fonction de probabilité :

$$\mathbb{P}[N = k] = \frac{\lambda^k}{k!} e^{-\lambda}$$

Cette distribution présente l'avantage de ne nécessiter qu'un seul paramètre λ qui s'estime aisément comme la moyenne empirique du nombre annuel de pertes (estimateur du maximum de vraisemblance). On calibre la distribution de la fréquence après celle de la sévérité afin de prendre en compte la présence du seuil de collecte. En effet, la probabilité qu'une perte soit supérieure au seuil de collecte est égale au ratio du nombre de pertes collectées sur le nombre total "réel" de pertes. On corrige donc l'estimation du paramètre de fréquence que l'on note $\widehat{\lambda}_H$ en le divisant par la proportion de pertes supérieures au seuil H :

$$\widehat{\lambda} = \frac{\widehat{\lambda}_H}{\mathbb{P}[X > H]} = \frac{\widehat{\lambda}_H}{1 - F_{\widehat{\theta}}(H)}$$

En pratique, il suffit de calculer la moyenne empirique du nombre de pertes annuel (qui est une estimation de λ_H) et d'utiliser l'estimateur de θ pour obtenir la vraie valeur du paramètre de la distribution de la fréquence.

2.2.3 Construction de la perte annuelle

Une fois les paramètres $\widehat{\theta}$ et $\widehat{\lambda}$ calibrés, on a alors déterminé les distributions de sévérité, X , et de fréquence, N . Il s'agit donc de construire la distribution de la perte annuelle, définie par :

$$L = \sum_{k=1}^N X_k$$

Trois méthodes peuvent alors être employées pour construire cette perte annuelle et sont évoquées dans [27] et [32], il s'agit de la méthode de Monte Carlo, la méthode de la transformée de Fourier et l'approche récursive de Panjer.

Méthode de la transformée de Fourier

La transformée de Fourier d'une fonction f et sa transformée inverse sont respectivement définies par :

$$\mathcal{F}(f)(v) = g(v) = \int_{-\infty}^{+\infty} f(u)e^{-ivu} du \quad \text{et} \quad \mathcal{F}^{-1}(g)(u) = f(u) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} g(v)e^{iuv} dv$$

Ainsi, pour une variable aléatoire Z , on peut relier la densité f_Z et la fonction caractéristique ϕ_Z via les transformées de Fourier :

$$\phi_Z(u) = \mathbb{E}[e^{iuZ}] = \int_{-\infty}^{+\infty} f_Z(v)e^{iuv} dv = \overline{\int_{-\infty}^{+\infty} f_Z(v)e^{-iuv} dv} = \overline{\mathcal{F}(f_Z)(u)}$$

Or :

$$\overline{\mathcal{F}(f_Z)(u)} = \int_{-\infty}^{+\infty} f_Z(v)e^{iuv} dv = \mathcal{F}(f_Z)(-u)$$

Alors la densité se retrouve en écrivant :

$$\mathcal{F}^{-1}(\phi_Z)(u) = \mathcal{F}^{-1}(\mathcal{F}(f_Z)(-u)) = \overline{\mathcal{F}^{-1}(\mathcal{F}(f_Z)(u))} = \overline{f_Z(u)} = f_Z(u)$$

A partir de la densité de la sévérité f_X , on détermine sa fonction caractéristique par transformée de Fourier : $\phi_X = \mathcal{F}(f_X)$. En utilisant le fait que la fonction caractéristique d'une somme de variables aléatoires indépendantes est le produit des fonctions caractéristiques de ces variables, on a :

$$\phi_{X_1+\dots+X_n}(u) = \prod_{i=1}^n \phi_{X_i}(u) = (\phi_X(u))^n$$

La fonction génératrice d'une variable $N \sim \mathcal{P}(\lambda)$ s'écrit :

$$g_N(t) = \sum_{n=1}^{+\infty} \mathbb{P}[N = n]t^n = \sum_{n=1}^{+\infty} \frac{\lambda^n}{n!} e^{-\lambda} t^n = e^{\lambda(t-1)}$$

La fonction caractéristique de la distribution des pertes annuelles vaut donc (N étant indépendante de X) :

$$\phi_L(u) = g_N \circ \phi_Y(u) = e^{\lambda(\phi_Y(u)-1)}$$

La densité de la distribution des pertes annuelles se déduit par transformée de Fourier inverse :

$$f_L = \mathcal{F}^{-1}(\phi_L)$$

On intègre enfin la densité obtenue pour en déduire la fonction de répartition F_L et on détermine ainsi la charge en capital réglementaire en prenant le quantile à $\alpha = 99.9\%$:

$$\text{CaR} = F_L^{-1}(\alpha)$$

Approche récursive de Panjer dans le cas Poisson

L'algorithme de Panjer est une relation récursive qui permet l'évaluation de la fonction de distribution de la perte annuelle L définie selon la loi composée :

$$L = \sum_{k=1}^N X_k$$

Les pertes X_k sont indépendantes, identiquement distribuées et positives de fonction de répartition F_X continue. La première étape est une étape de discrétisation de la fonction F_X . On définit \tilde{L} :

$$\tilde{L} = \sum_{k=1}^N \tilde{X}_k$$

Les variables aléatoires \tilde{X}_k sont aussi des variables indépendantes et identiquement distribuées et sont concentrées sur le support $\{0, 1h, \dots, rh\}$ où h est le pas de discrétisation et r un nombre arbitrairement choisi. On a alors $\tilde{L} \in \{0, 1h, 2h, \dots\}$. L'idée est d'approximer F_X par $F_{\tilde{X}}$. On note les fonctions de masses de probabilité et de répartition de \tilde{X} par

$$f_{\tilde{X}}(kh) = \mathbb{P}(\tilde{X} = kh), \quad k = 0, 1, \dots, r$$

et

$$F_{\tilde{X}}(kh) = \sum_{j=0}^k f_{\tilde{X}}(jh), \quad k = 0, 1, \dots, r$$

On applique alors l'algorithme de Panjer dans le cas où N suit une loi de Poisson de paramètre λ .

Point de départ

$$\begin{aligned} F_{\tilde{L}}(0) &= \Pr(\tilde{L} = 0) \\ &= M_N\{\ln[f_{\tilde{X}}(0)]\} \\ &= e^{-\lambda(1-f_{\tilde{X}}(0))} \end{aligned}$$

Relation récursive

$$f_{\tilde{L}}(kh) = \sum_{y=1}^{\min(k,r)} \frac{\lambda y}{k} f_{\tilde{X}}(yh) f_{\tilde{L}}(kh - yh)$$

On peut alors calculer la fonction de répartition $F_{\tilde{L}}$ et en déduire une approximation de la charge en capital réglementaire en prenant le quantile à $\alpha = 99.9\%$:

$$\text{CaR} = F_L^{-1}(\alpha) \approx F_{\tilde{L}}^{-1}(\alpha)$$

Méthode de Monte Carlo

La construction de la perte annuelle par la méthode de Monte Carlo consiste à en simuler un grand nombre S de réalisations. Pour chaque simulation $s \in \llbracket 1, S \rrbracket$, on effectue un tirage aléatoire de la variable aléatoire $N \sim \mathcal{P}(\hat{\lambda})$ qui donne une réalisation n^s du nombre annuel de

pertes. On simule ensuite des pertes unitaires en simulant n_s réalisations indépendantes issues d'une loi $\mathcal{LN}(\hat{\mu}, \hat{\sigma})$ et notées $x_1^s, \dots, x_{n_s}^s$. La somme de ces n_s pertes unitaires donne alors la réalisation l_s de la perte annuelle :

$$l_s = \sum_{k=1}^{n_s} x_k^s$$

On obtient ainsi un vecteur de réalisations de la perte annuelle $l = (l_1, \dots, l_S)'$. La charge en capital réglementaire est alors obtenue en prenant le quantile à $\alpha = 99.9\%$, c'est-à-dire la $(\alpha \times S)^{\text{ème}}$ plus grande valeur de l . En pratique, $S = 5\,000\,000$ simulations sont utilisées afin de garantir une précision satisfaisante.

Pour calculer nos montants de capitaux, une fois les paramètres des lois de fréquence et de sévérité obtenus, nous avons utilisé dans le cadre de notre mémoire la méthode de Monte Carlo.

2.3 Calcul du capital réglementaire basé sur l'analyse de scénario

En complément des calculs de charges en capital basés sur les pertes internes, le Crédit Agricole intègre des informations, à travers les analyses de scénario, qui sont liées à certains événements extrêmes ne figurant pas dans les bases de données car ils n'ont pas encore été observés ou qu'ils l'ont été mais n'y figurent pas. Le fait d'intégrer ces nouvelles informations conduit à un nouveau calcul de capital réglementaire.

A ce jour, trois types de scénario interviennent :

1. **Les scénario stand alone** : leur but est de compléter les historiques de pertes lorsque ceux-ci sont jugés incomplets par les experts, ou bien de déterminer la distribution des pertes dans les cas où il n'y a pas de données internes.
2. **Les scénarios internalisés** : certains scénarios "stand alone" sont homogènes au profil de risque des pertes internes, ils peuvent par conséquent être intégrés aux historiques.
3. **Les pertes internes externalisées** : les pertes internes de montant élevé et très atypiques peuvent dans certains cas être externalisées sous forme de scénarios "stand alone".

Dans le cadre du mémoire, nous nous intéresserons principalement aux scénarios "stand alone".

2.3.1 Construction et utilisation des scénarios

Les scénarios sont établis par des experts métiers afin de décrire le risque induit par un type d'événement donné. Plus précisément, l'expert se prononce en formulant un scénario sous la forme suivante :

« une perte supérieure ou égale à x arrive toutes les d années ».

Un scénario est par conséquent constitué d'un couple montant-temps de retour, noté (x, d) . L'enjeu est alors de traduire ce scénario (x, d) en termes de restrictions sur les paramètres du modèle LDA, c'est à dire à travers la fréquence, modélisée par une loi de Poisson $\mathcal{P}(\lambda)$, et la sévérité, par une variable aléatoire continue de paramètre θ , de fonction de répartition notée F_θ .

On cherche tout d'abord à établir une relation entre x et d , en s'intéressant à la durée δ^x entre deux pertes supérieures au montant x . Calculons la distribution de δ^x . Comme cela a été fait dans A. Frachot et al. [16], nous supposons que le nombre de pertes suit un processus de Poisson de paramètre λ . Si l'on note T_i la date à laquelle apparaît la $i^{\text{ème}}$ perte, les durées $e_i = T_i - T_{i-1}$ entre deux pertes consécutives sont indépendantes et identiquement distribuées selon une loi exponentielle de paramètre λ .

On suppose que les pertes X_i sont indépendantes et identiquement distribuées de distribution F . On note δ_j^x les durées entre deux pertes excédant un seuil x . On déduit directement que les durées sont indépendantes et identiquement distribuées. Il suffit à présent de caractériser $\delta^x = \delta_1^x$. On a :

$$\begin{aligned}
 \mathbb{P}(\delta^x \geq t) &= \sum_{i \geq 1} \mathbb{P}[T_i \geq t, X_1 \leq x, \dots, X_{i-1} \leq x, X_i \geq x] \\
 &= \sum_{i \geq 1} \mathbb{P}[T_i \geq t] F(x)^{i-1} (1 - F(x)) \\
 &= \sum_{i \geq 1} (1 - F(x)) F(x)^{i-1} \sum_{k=0}^{i-1} e^{-\lambda t} \frac{\lambda^k t^k}{k!} \\
 &= (1 - F(x)) \sum_{k=0}^{\infty} e^{-\lambda t} \frac{\lambda^k t^k}{k!} \sum_{i=k}^{\infty} F(x)^i \\
 &= e^{-\lambda t} \sum_{k=0}^{\infty} \frac{\lambda^k t^k}{k!} F(x)^k \\
 &= e^{-\lambda(1-F(x))t}
 \end{aligned}$$

On en déduit que δ^x suit une loi exponentielle $\mathcal{E}(\lambda(1 - F_\theta(x)))$. La durée moyenne entre deux pertes supérieures à x est donc la moyenne de δ^x , soit $\frac{1}{\lambda(1-F_\theta(x))}$, et le nombre moyen vaut donc $\lambda(1 - F_\theta(x))$. Il en résulte que, pour un scénario donné (x, d) , les paramètres doivent satisfaire la contrainte suivante :

$$d = \frac{1}{\lambda(1 - F_\theta(x))}$$

Cette relation est ensuite utilisée pour estimer les paramètres du modèle. Pour un scénario donné et dans le cas log-normal, il suffit que l'expert établisse 3 hypothèses (x_k, d_k) (basse, moyenne et haute) du scénario pour estimer les trois paramètres λ, μ et σ . On obtient alors les estimateurs en minimisant un critère quadratique :

$$(\hat{\lambda}, \hat{\mu}, \hat{\sigma}) = \arg \min_{\lambda, \mu, \sigma} \sum_{k=1}^3 w_k \left(d_k - \frac{1}{\lambda(1 - F_{\mu, \sigma}(x_k))} \right)^2$$

où les poids w_k sont optimaux en prenant $w_k = \left(\frac{1}{d_k}\right)^2$.

Les experts ont également la possibilité de se prononcer sur le paramètre de fréquence λ ou encore sur l'expected loss $EL = \lambda e^{\mu + \sigma^2/2}$. Cela se traduit par des contraintes dans l'estimation des paramètres (chaque contrainte réduit le nombre d'hypothèses nécessaires pour estimer les paramètres) :

- Contrainte sur la fréquence $\lambda = \lambda_0$:

$$\begin{cases} \hat{\lambda} = \lambda_0 \\ (\hat{\mu}, \hat{\sigma}) = \arg \min_{\mu, \sigma} \sum_{k=1}^2 w_k \left(d_k - \frac{1}{\lambda_0(1 - F_{\mu, \sigma}(x_k))} \right)^2 \end{cases}$$

- Contrainte sur l'expected loss $\text{EL} = \text{EL}_0$:

$$\begin{cases} \hat{\lambda} = \text{EL}_0 e^{-\hat{\mu} - \hat{\sigma}^2/2} \\ (\hat{\mu}, \hat{\sigma}) = \arg \min_{\mu, \sigma} \sum_{k=1}^2 w_k \left(d_k - \frac{e^{\mu + \sigma^2/2}}{\text{EL}_0(1 - F_{\mu, \sigma}(x_k))} \right)^2 \end{cases}$$

- Contrainte sur la fréquence $\lambda = \lambda_0$ et sur l'expected loss $\text{EL} = \text{EL}_0$:

$$\begin{cases} \hat{\lambda} = \lambda_0 \\ \hat{\mu} = \ln\left(\frac{\text{EL}_0}{\lambda_0}\right) - \frac{\sigma^2}{2} \\ \hat{\sigma} = \arg \min_{\sigma} \left(d_k - \frac{1}{\lambda_0 \left(1 - F_{\ln\left(\frac{\text{EL}_0}{\lambda_0}\right) - \frac{\sigma^2}{2}, \sigma}(x_k) \right)} \right)^2 \end{cases}$$

2.3.2 Prise en compte des assurances

Dans le cadre de l'approche AMA, une banque est autorisée à utiliser l'assurance pour se protéger contre le risque opérationnel, dans la limite de 20% de l'exigence totale de fonds propres au regard du risque opérationnel. Ces polices d'assurances sont contractées pour couvrir des événements faisant l'objet de scénarios.

Pour une police d'assurance définie par une franchise F , un plafond par sinistre P , des taux d'assurabilité t_1 (entre F et P) et t_2 (au-delà de P), et un taux de couverture c , l'effet réducteur lié aux assurances est estimé au sein des simulations Monte Carlo, en réécrivant la perte annuelle

$$L = \sum_{k=1}^N X_k$$

en remplaçant chaque perte individuelle X_k par une perte après "assurance" $X_k^a = X_k \mathbf{1}_{\{u \leq c\}} + X_k \mathbf{1}_{\{u > c\}}$, où u est un tirage d'une loi uniforme $\mathcal{U}_{[0,1]}$, et

$$\begin{aligned} X_k^* &= X_k \mathbf{1}_{\{X_k \leq F\}} + (F + (1 - t_1)(X_k - F)) \mathbf{1}_{\{F < X_k \leq P\}} \\ &\quad + (F + (1 - t_1)(P - F) + (1 - t_2)(X_k - P)) \mathbf{1}_{\{X_k > P\}} \end{aligned}$$

La perte annuelle avec assurance s'écrit donc :

$$L^a = \sum_{k=1}^N X_k^a$$

Si la police comporte un plafond annuel π , la perte annuelle avec assurance et plafond annuel s'écrit :

$$L^{a, \pi} = L^a + (L - L^a - \pi) \mathbf{1}_{\{L - L^a \geq \pi\}}$$

Au final, l'effet total de réduction ne doit pas excéder 20% de la charge en capital après diversification.

2.4 Agrégation des charges en capital

Le Comité de Bâle aborde la prise en compte de la diversification des risques dans le calcul de la charge en capital au titre du risque opérationnel. La banque est en effet autorisée à appliquer des corrélations déterminées en interne à condition de démontrer que ses systèmes de détermination des corrélations sont sains, mis en œuvre avec intégrité et tiennent compte des incertitudes inhérentes à toute estimation de corrélations.

L'agrégation des charges en capital requiert donc de définir tout d'abord les niveaux de corrélation applicables puis la méthode de prise en compte de ces dépendances.

Niveaux de corrélation

Le modèle LDA impose l'hypothèse d'indépendance entre les pertes d'un même type de risque mais il est difficile de supposer à la fois cette indépendance et de supposer une corrélation des pertes entre types de risque. On fait donc l'hypothèse que la corrélation entre les pertes annuelles est une corrélation des fréquences.

Pour deux pertes annuelles $L_1 = \sum_{k=1}^{N_1} X_{1,k}$ et $L_2 = \sum_{k=1}^{N_2} X_{2,k}$, en supposant que N_1 et N_2 sont parfaitement corrélés, c'est-à-dire $N_1 = N_2 = N \sim \mathcal{P}(\lambda)$, la covariance entre L_1 et L_2 vaut :

$$\begin{aligned} \text{Cov}[L_1, L_2] &= \mathbb{E} \left[N^2 \mathbb{E}[X_1] \mathbb{E}[X_2] \right] - \lambda^2 \mathbb{E}[X_1] \mathbb{E}[X_2] \\ &= \left(\mathbb{V}[N] + \mathbb{E}^2[N] - \lambda^2 \right) \mathbb{E}[X_1] \mathbb{E}[X_2] \\ &= \lambda \mathbb{E}[X_1] \mathbb{E}[X_2] \end{aligned}$$

Puisque $\mathbb{V}[L_i] = \lambda \mathbb{E}[X_i^2]$, la corrélation entre L_1 et L_2 ne dépend pas de λ et vaut :

$$\text{Cor}[L_1, L_2] \leq \frac{\mathbb{E}[X_1] \mathbb{E}[X_2]}{\sqrt{\mathbb{E}[X_1^2] \mathbb{E}[X_2^2]}}$$

Avec $X_i \sim \mathcal{LN}(\mu_i, \sigma_i)$, on a $\mathbb{E}[X_i] = e^{\mu_i + \sigma_i^2/2}$ et $\mathbb{E}[X_i^2] = e^{2(\mu_i + \sigma_i^2)}$. Ainsi, pour décrire la corrélation entre deux pertes annuelles L_k et $L_{k'}$, on utilise le coefficient analytique suivant :

$$\varrho_{k,k'}^a = \text{Cor}[L_k, L_{k'}] = \exp \left(-\frac{\sigma_k^2 + \sigma_{k'}^2}{2} \right)$$

Parallèlement à ce calcul, les experts métiers se prononcent également et proposent, pour chaque couple (k, k') , un coefficient expert $\varrho_{k,k'}^e$. Le coefficient de corrélation finalement pris en considération est le maximum $\varrho_{k,k'}$ entre le coefficient établi à dire d'experts et le coefficient analytique :

$$\varrho_{k,k'} = \max(\varrho_{k,k'}^e, \varrho_{k,k'}^a)$$

On peut alors définir une **matrice d'agrégation**. On considère que la banque est composée de I entités et l'entité $i \in \llbracket 1; I \rrbracket$ est composée de $J^{(i)}$ catégories de risque. Pour décrire les corrélations entre les pertes annuelles, on définit par blocs une matrice Ω carrée, de dimension $K = \sum_{i=1}^I J^{(i)}$:

$$\Omega = \begin{pmatrix} M_{1,1} & \cdots & M_{1,I} \\ \vdots & \ddots & \vdots \\ M_{I,1} & \cdots & M_{I,I} \end{pmatrix}$$

où pour $(i, i') \in \llbracket 1, I \rrbracket^2$, $M_{i,i'}$ est de dimension $J^{(i)} \times J^{(i')}$ et vaut :

$$M_{i,i'} = \begin{pmatrix} \rho_{1,1}^{i,i'} & \cdots & \rho_{1,J^{(i')}}^{i,i'} \\ \vdots & \ddots & \vdots \\ \rho_{J^{(i)},1}^{i,i'} & \cdots & \rho_{J^{(i)},J^{(i')}}^{i,i'} \end{pmatrix}$$

où pour $(j, j') \in \llbracket 1, J^{(i)} \rrbracket \times \llbracket 1, J^{(i')} \rrbracket$, $\rho_{j,j'}^{i,i'}$ représente la corrélation entre la catégorie de risque j de l'entité i et la catégorie j' de i' . On peut donc assimiler la matrice à une matrice de corrélation de dimension K en notant $\varrho_{k,k'}$ le coefficient de corrélation entre les unités de mesure k et k' :

$$\Omega = \begin{pmatrix} 1 & \cdots & \varrho_{1,k} & \cdots & \varrho_{1,K} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \varrho_{1,k} & \cdots & 1 & \cdots & \varrho_{k,K} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \varrho_{1,K} & \cdots & \varrho_{k,K} & \cdots & 1 \end{pmatrix}$$

2.4.1 Méthodologie d'agrégation

La méthodologie d'agrégation repose sur l'hypothèse que le vecteur \mathbf{L} des K pertes annuelles a la structure d'un vecteur gaussien :

$$\mathbf{L} = (L_1, \dots, L_K)' \sim \mathcal{N}(\mathbf{m}, \Sigma)$$

avec $\mathbf{m} = (m_1, \dots, m_K)'$ et

$$\Sigma = \begin{pmatrix} s_1^2 & & & & \\ \vdots & \ddots & & & \\ \vdots & & \varrho_{k,k'} s_k s_{k'} & & \\ \vdots & & & \ddots & \\ & & & & s_K^2 \end{pmatrix}$$

En notant $\mathbf{1} = (1, \dots, 1)'$, on a $\sum_{k=1}^K L_k = \mathbf{1}'\mathbf{L} \sim \mathcal{N}(\mathbf{1}'\mathbf{m}, \mathbf{1}'\Sigma\mathbf{1})$, c'est-à-dire :

$$\sum_{k=1}^K L_k \sim \mathcal{N}\left(\sum_{k=1}^K m_k, \sum_{k=1}^K \sum_{k'=1}^K \varrho_{k,k'} s_k s_{k'}\right)$$

Alors, la CaR globale de niveau α s'écrit :

$$\text{CaR}(\alpha) = \sum_{k=1}^K m_k + \Phi^{-1}(\alpha) \sqrt{\sum_{k=1}^K \sum_{k'=1}^K \varrho_{k,k'} s_k s_{k'}}$$

En particulier $l_k \sim \mathcal{N}(m_k, s_k^2)$, on a donc :

$$\frac{l_k - m_k}{s_k} \sim \mathcal{N}(0, 1)$$

Alors, la CaR peut se calculer par une méthode gaussienne :

$$\text{CaR}_k(\alpha) = F_{l_k}^{-1}(\alpha) = m_k + s_k \Phi^{-1}(\alpha)$$

Or, l'*expected loss* vaut $\text{EL}_k = \mathbb{E}[l_k] = m_k$, donc l'*unexpected loss* vaut :

$$\text{UL}_k(\alpha) = \text{CaR}_k(\alpha) - \text{EL}_k = s_k \Phi^{-1}(\alpha)$$

Ainsi, on obtient que :

$$\forall (k, k') \in \llbracket 1, K \rrbracket^2, \quad s_k s_{k'} = \frac{1}{(\Phi^{-1}(\alpha))^2} \text{UL}_k(\alpha) \text{UL}_{k'}(\alpha)$$

Si on remplace dans l'équation générale de la CaR, on arrive à :

$$\text{CaR}(\alpha) = \sum_{k=1}^K \text{EL}_k + \sqrt{\sum_{k=1}^K \sum_{k'=1}^K \varrho_{k,k'} \text{UL}_k(\alpha) \text{UL}_{k'}(\alpha)}$$

2.5 Application numérique à nos données

2.5.1 Présentation des données

On dispose de données relatives à 6 entités du Groupe Crédit Agricole et à 8 types de risques. Nos données regroupent les historiques de pertes pour chacune des cellules de risque ainsi définies et les couples (x, d) fournis par les experts pour chaque scénario. La matrice suivante résume l'ensemble des données dont on dispose ; les cases munies du symbole \times représentent les cellules de risque disposant de scénarios d'expert ; celles avec le symbole \bullet correspondent aux données historiques :

	LCL	AGOS	CACF	AMUNDI	CACIB	CR
EL1	$\bullet \times$	\times	\times	\times	\times	$\bullet \times$
EL2HM	$\bullet \times$	$\bullet \times$	$\bullet \times$	\times	$\bullet \times$	$\bullet \times$
EL2M	\bullet	\bullet	\bullet		\bullet	\bullet
EL3	$\bullet \times$	$\bullet \times$	$\bullet \times$	\times	$\bullet \times$	$\bullet \times$
EL4	$\bullet \times$	\times	\times	$\bullet \times$	$\bullet \times$	$\bullet \times$
EL5	$\bullet \times$	$\bullet \times$	\times	\times	\times	$\bullet \times$
EL6	$\bullet \times$	\times	$\bullet \times$	$\bullet \times$	$\bullet \times$	$\bullet \times$
EL7	\bullet	\bullet	\bullet	\bullet	\bullet	\bullet

TABLE 2.1 – Présentation des données par entité et type de risque selon que la cellule de risque dispose de données internes (\bullet) ou de scénarios (\times)

A première vue, certaines catégories de risque ne disposent pas des deux sources de données. Par exemple, les pertes pour le risque monétaire EL2M ne présentent pas de scénarios car les experts considèrent que le montant de ces pertes n'est pas suffisamment élevé pour qu'il soit nécessaire de créer un scénario qui est censé caractériser une perte extrême. Il en est de même pour le risque EL7. On choisit ici de travailler avec des catégories de risque similaires (on verra par la suite qu'on a choisi ces catégories car elles facilitent une comparaison au niveau des résultats) pour lesquelles on dispose à la fois des données historiques et des scénarios. Considérons les trois cellules suivantes :

	LCL	AGOS	CACF	AMUNDI	CACIB	CR
EL1	• ×	×	×	×	×	• ×
EL2HM	• ×	• ×	• ×	×	• ×	• ×
EL2M	•	•	•		•	•
EL3	• ×	• ×	• ×	×	• ×	• ×
EL4	• ×	×	×	• ×	• ×	• ×
EL5	• ×	• ×	×	×	×	• ×
EL6	• ×	×	• ×	• ×	• ×	• ×
EL7	•	•	•	•	•	•

TABLE 2.2 – Cellules de risques choisies pour notre première étude

En ce qui concerne le choix des lois de fréquence et de sévérité des pertes, on rappelle qu'au sein du modèle actuel du Crédit Agricole, les lois utilisées sont une loi de Poisson pour modéliser la fréquence des pertes, et une loi log-normale pour modéliser la sévérité des pertes.

Aujourd'hui, deux types de charges en capital sont calculés au sein du Groupe Crédit Agricole à l'aide des données historiques, plus communément appelées données internes, et à l'aide des scénarios. Les méthodes utilisées pour les calculer sont celles présentées précédemment. Ces capitaux sont ensuite agrégés pour fournir une charge en capital pour tout le Groupe, tous risques confondus.

On décrit dans la suite les informations contenues dans les cellules, les estimations des paramètres des lois selon les deux méthodes présentées précédemment ainsi que les valeurs de charges en capital réglementaires obtenues.

2.5.2 Détermination des capitaux réglementaires à l'aide des données historiques

On commence par présenter les échantillons de pertes internes des trois cellules qui seront utilisés dans les méthodes de calcul des capitaux réglementaires présentées dans ce mémoire. Chaque cellule comporte l'historique des pertes opérationnelles collectées sur 5 ans entre 2007 à 2011.

Cas	Entité	Catégorie de risque	Pertes	Seuil de collecte
1	AMUNDI	EL6	129	1000
2	CACF	EL6	308	1
3	CACIB	EL6	1307	1000

TABLE 2.3 – Présentation des données traitées

Statistiques descriptives

Etant donnée la particularité des données du risque opérationnel, il est intéressant de mener une étude exploratoire préalable. Les statistiques descriptives sont contenues dans le tableau suivant :

Cas	Moyenne	Ecart-type	Skewness	Kurtosis corrigé
1	10 424	29 812	7	61
2	3 833	18 117	9	90
3	15 092	113 683	26	770

TABLE 2.4 – Statistiques descriptives

Le skewness mesure l'asymétrie de la distribution des données, on voit ici qu'il est toujours positif, c'est à dire que la queue à droite de la distribution est plus longue, et que la masse des montants de perte est concentrée à gauche. Le kurtosis mesure quant à lui l'aplatissement de la distribution c'est à dire l'épaisseur de la queue. Le kurtosis étant toujours supérieur à 3, ici, cela signifie que les distributions sont leptokurtiques¹.

Estimation et calcul de la CaR

En appliquant la méthode pour les données internes (DI) décrites précédemment, on obtient les estimations des paramètres corrigés de fréquence et de sévérité suivant :

Cas	λ_{DI}	μ_{DI}	σ_{DI}
1	53.15	7.56	1.61
2	81	5.21	2.17
3	815.96	6.15	2.24

TABLE 2.5 – Estimation des paramètres de loi à partir des données internes

Ces paramètres permettent d'obtenir l'ajustement de loi de sévérité suivant pour le cas 1 :

1. ayant des queues plus épaisses que la loi Normale

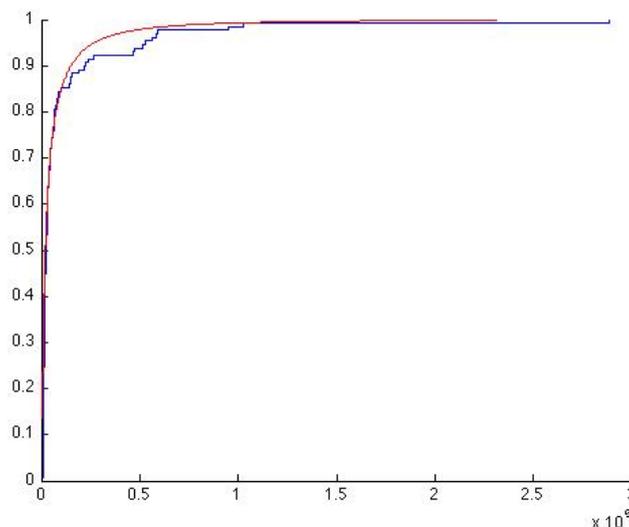


FIGURE 2.1 – Comparaison des fonctions de répartition empirique et théorique

A l'aide de la méthode de simulation de Monte Carlo mise en place pour le calcul de la charge en capital, on obtient les résultats suivants pour les trois cellules de risque choisies :

Cas	CaR _{DI}
1	1 874 733 €
2	1 953 667 €
3	22 722 500 €

TABLE 2.6 – Charges en capital obtenues à partir des données internes

2.5.3 Détermination des capitaux réglementaires à l'aide des scénarios

On commence par présenter les scénarios établis par les experts métiers afin de décrire le risque induit par un type d'événement donné. On rappelle que ces scénarios sont fournis sous la forme de couples montant-temps de retour (x, d) signifiant qu'une perte supérieure ou égale à x arrive toutes les d années. Les informations des experts utilisées dans ce mémoire sont les suivantes (pour chaque entité, le scénario imaginé est une dysfonctionnement informatique) :

Cas	x_1	d_1	x_2	d_2	x_3	d_3
1	6 969 000	50	17 043 000	150		
2	255 000	5	762 000	10	1 474 000	30
3	2 000 000	5	12 000 000	10	25 000 000	40

 TABLE 2.7 – Couples (x, d) fournis par les experts métiers

On constate que pour le cas 1, il n'y a pas de troisième couple fourni. Comme on l'a vu précédemment, les experts ont également la possibilité de se prononcer sur le paramètre de fréquence λ ou encore sur l'expected loss $EL = \lambda \exp(\mu + \sigma^2/2)$. Ces contraintes viennent alors réduire le

nombre nécessaire d'hypothèses pour estimer les paramètres de loi. Dans le cas 1, les experts se sont prononcés sur l'expected loss : $EL_1=315\ 000\ \text{€}$. En appliquant la méthode consistant à minimiser un critère quadratique, on obtient les estimations des paramètres de loi suivantes :

Cas	λ_S	μ_S	σ_S
1	0.02	16.5	0.34
2	0.22	13.48	0.71
3	0.2	16.3	0.64

TABLE 2.8 – Estimation des paramètres de loi à partir des scénarios

A l'aide de la méthode de simulation de Monte Carlo mise en place pour le calcul de la charge en capital, on obtient les résultats suivants pour les trois cellules de risque choisies :

Cas	CaR _S
1	26 191 333 €
2	4 946 000 €
3	68 627 000 €

TABLE 2.9 – Charges en capital obtenues à partir des scénarios

2.5.4 Agrégation des capitaux

Sous l'hypothèse gaussienne, on peut à présent déterminer la CaR pour chaque cellule de risque. En ne considérant que les entités AMUNDI, CACF et CACIB, nous disposons de la méta-matrice de corrélation suivante :

		AMUNDI		CACF		CACIB	
		EL6	Scénario	EL6	Scénario	EL6	Scénario
Entité AMUNDI	EL6	1	0	0.06	0	0.02	0
	Scénario	0	1	0	0	0	0
Entité CACF	EL6	0.06	0	1	0	0	0
	Scénario	0	0	0	1	0	0
Entité CACIB	EL6	0.02	0	0	0	1	0
	Scénario	0	0	0	0	0	1

TABLE 2.10 – Méta-matrice de corrélation pour les entités AMUNDI, CACF et CACIB

On peut alors déterminer les *expected losses* EL, les *unexpected losses* UL, puis les CaR agrégées :

Entité	EL	CaR	UL	CaR agrégée	CaR globale
AMUNDI EL6	372 938 €	1 874 733 €	1 501 795 €	28 066 066 €	81 749 871 €
AMUNDI Scn	310 450 €	26 191 333 €	25 880 883 €		
CACF EL6	156 201 €	1 953 667 €	1 797 466 €	6 899 667 €	
CACF Scn	202 384 €	4 946 000 €	4 743 616 €		
CACIB EL6	4 700 434 €	22 722 500 €	18 022 066 €	91 349 500 €	
CACIB Scn	2 944 242 €	68 627 000 €	65 682 758 €		

TABLE 2.11 – Charges en capital agrégées

On remarque que les charges en capital agrégées pour chaque entité sont la somme des charges venant des deux sources de données. Cela est dû à la forme de la méta-matrice de corrélation.

2.6 Objectifs du mémoire

Aujourd’hui, les scénarios et les données internes sont traités comme deux catégories de risque à part entière, y compris pour un même type de risque. Il est donc difficile de dégager un unique profil de risque pour une entité et un event-type donné. Combiner ces deux sources apporterait plus de lisibilité à l’analyse des profils de risque des entités mais aussi de la banque une fois ces charges agrégées.

Par ailleurs, l’hypothèse gaussienne utilisée aujourd’hui pour agréger les CaR obtenues entre catégories de risque suppose une diversification des risques. Les scénarios sont traités comme des catégories de risque à part entière, et il existe donc une corrélation entre un scénario relatif à la catégorie de risque i et une catégorie de risque (pertes internes ou scénario) j . Naturellement, un scénario d’une catégorie i ne devrait pas être corrélé aux pertes internes d’une catégorie de risque i pour une business line donnée mais certaines exceptions existent en raison d’analyse métiers dans ce sens.

Ainsi, la méta-matrice qui permet d’agréger les charges en capital prend en compte toutes les corrélations entre les scénarios et les données internes de toutes les business lines. Remarquons tout d’abord que dans le cadre de matrices de corrélation par business line, l’absence de corrélation entre un scénario lié à la catégorie de risque i et l’historique de données internes de la catégorie de risque i induit de sommer les charges obtenues par les deux sources de données. De plus, en supposant des alternatives à l’hypothèse gaussienne comme méthode de diversification qui peut sembler discutable (copules par exemple), il peut être nécessaire d’obtenir des matrices de corrélation de dimensions moins grandes.

C’est pourquoi la théorie bayésienne paraît être un cadre propice à l’obtention de matrices par entités de taille restreinte (nombre de catégories de risque) et à la diversification entre scénarios et données internes d’une même catégorie de risque, et enfin à l’obtention de profils de risque uniques pour chaque catégorie de risque considérée.

Partie 3

Approche bayésienne du modèle LDA

3.1 Les principes de la statistique bayésienne

A la différence de la statistique classique, en statistique bayésienne les paramètres sont considérés comme des variables aléatoires auxquelles on affecte une densité de probabilité. La définition du modèle statistique habituel s'en trouve alors modifiée : il faut affecter des densités de probabilité aux paramètres au moment de la définition du modèle pour pouvoir prendre en compte leur nature aléatoire. Cette distribution est la distribution *a priori* des paramètres car elle est définie avant d'observer les résultats (voir [29] et [28]).

Considérons $X = (X_1, X_2, \dots, X_n)'$ un vecteur de variables aléatoires de pertes opérationnelles dont la densité jointe pour un vecteur de paramètres aléatoires $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_n)'$ est $h(x | \theta)$. Alors la densité jointe est :

$$h(x, \theta) = h(x | \theta)\pi(\theta) = \pi(\theta | x)h(x)$$

où :

- $\pi(\theta)$ est la densité de probabilité des paramètres appelée **loi de densité a priori**
- $\pi(\theta | x)$ est la densité de probabilité conditionnelle des paramètres sachant le vecteur X appelée **loi a posteriori**
- $h(x, \theta)$ est la densité jointe des données observées et des paramètres
- $h(x | \theta)$ est la densité des observations pour des paramètres donnés (fonction de vraisemblance)
- $h(x)$ est la densité marginale de X qui peut s'écrire comme $h(x) = \int h(x | \theta)\pi(\theta)d\theta$

Comme on veut modéliser le risque opérationnel à l'aide de l'approche LDA, on sépare la modélisation de la fréquence des pertes et leur sévérité. Pour ces variables, on va donc appliquer une méthode bayésienne : notre objectif final sera donc d'estimer la distribution de probabilité *a posteriori* de la fréquence d'une part et de la sévérité d'autre part. Cette méthode repose principalement sur la formule de Bayes qui est un mécanisme d'apprentissage qui combine l'information *a priori* sur les paramètres et l'information apportée par les données en une information *a posteriori* sur les paramètres. Précisément la formule de Bayes s'applique à la densité conjointe *a priori* des paramètres et renvoie la densité conjointe des paramètres *a posteriori* c'est-à-dire conditionnellement aux informations.

La formule de Bayes appliquée à des densités de probabilité s'écrit :

$$\pi(\theta | x) = \frac{h(x | \theta)\pi(\theta)}{h(x)}$$

L'objectif dans le contexte du risque opérationnel est d'estimer la distribution de probabilité de la future observation X_{n+1} sachant l'information disponible $X = (X_1, X_2, \dots, X_n)'$. La densité conditionnelle de X_{n+1} vaut donc :

$$f(x_{n+1} | x) = \int f(x_{n+1} | \theta, x) \pi(\theta | x) d\theta$$

Ainsi on peut suivre la démarche suivante :

1. déterminer la loi a priori $\pi(\theta)$ par l'intermédiaire de l'analyse de scénario
2. combiner les données historiques et cette loi a priori grâce à l'équation précédente pour déterminer $\pi(\theta | x)$
3. calculer la densité de la loi de X_{n+1} grâce à la formule encore plus haut.

La formule de Bayes et le calcul des différentes lois marginales qui en découle font appel à de nombreux calculs d'intégrales multidimensionnelles pour lesquelles la forme analytique peut être très complexe. On verra par la suite que des méthodes de simulations Monte Carlo permettent alors d'approcher ces intégrales.

Dans l'approche bayésienne plusieurs méthodes coexistent suivant le type de données qu'on utilise. Dans l'**approche bayésienne pure**, la distribution de la loi a priori est spécifiée subjectivement (grâce aux jugements d'experts). Plusieurs méthodes existent qu'on détaille dans la section suivante.

En revanche, sous l'**approche bayésienne empirique**, le paramètre θ est traité comme un échantillon aléatoire venant de la distribution a priori. Ainsi, en utilisant des données collectives pour des risques similaires, les paramètres de la loi a priori sont estimés en utilisant une distribution marginale des observations. Même si ces risques similaires ont des profils différents, ils ont en commun que les Θ_k ont la même densité $\pi(\theta)$. Ainsi, on peut trouver la distribution inconditionnelle des données X et accorder la loi a priori à l'ensemble des données (par exemple par méthode de maximum de vraisemblance ou méthode des moments généralisée).

3.2 Estimation de la loi a priori

La détermination de la loi a priori est la clé de voûte de l'inférence bayésienne et sa détermination est donc l'étape la plus importante dans la mise en œuvre de cette inférence. Il existe deux manières différentes de procéder présentées dans [29] et [28].

3.2.1 Les distributions conjuguées

Le choix de modéliser la distribution *a priori* par une distribution conjuguée avec la vraisemblance permet une interprétation claire de la formule de Bayes, en tant qu'un processus de mise à jour des paramètres de la distribution *a priori* qui donne la distribution *a posteriori*. Une distribution a priori conjuguée est également un choix objectif pour la distribution a priori où seuls les paramètres restent à déterminer. Voici quelques exemples de lois conjuguées :

$h(x \theta)$ vraisemblance	$\pi(\theta)$ a priori	$\pi(\theta x)$ a posteriori
Normale $\mathcal{N}(\theta, \sigma^2)$	Normale $\mathcal{N}(\mu, \tau^2)$	Normale $\mathcal{N}(\rho(\sigma^2\mu + \tau^2x), \rho\sigma^2\tau^2), \rho^{-1} = \sigma^2 + \tau^2$
Binomiale $\mathcal{B}(n, \theta)$	Beta $Beta(\alpha, \beta)$	Beta $Beta(\alpha + x, \beta + n - x)$
Poisson $\mathcal{P}(\theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	Gamma $\mathcal{G}(\alpha + x, \beta + 1)$

TABLE 3.1 – Les lois conjuguées classiques

Le principal problème réside dans le choix des paramètres des lois a priori. La littérature propose certaines solutions.

Approche par histogramme

C'est la méthode la plus appropriée lorsqu'on considère que l'ensemble des paramètres Θ est continu et parcourt un certain intervalle. La méthode consiste à diviser l'ensemble Θ en plusieurs intervalles et à déterminer la probabilité correspondant à chacun d'entre eux. On peut alors en déduire un histogramme de probabilités dans un premier temps, puis calibrer une loi a priori en minimisant l'erreur (optimisation par moindres carrés, choix d'une distance spécifique, etc.).

Approche par détermination de la CDF : *Cumulative Distribution Function*

Cette méthode consiste à déterminer subjectivement la fonction de répartition. Cela peut être réalisé en déterminant différents quantiles $q(\alpha)$ et en représentant le couple $(q(\alpha), \alpha)$ pour différentes valeurs de α . On obtient une fonction de répartition qu'on peut choisir de lisser en reliant tous ces points.

Approche de la vraisemblance relative

L'objectif est de comparer les vraisemblances intuitives de différents points de Θ (en disant par exemple qu'une certaine valeur du paramètre est cinq fois plus vraisemblable qu'une autre) et d'en déduire une densité a priori.

Approche s'appuyant sur une loi a priori donnée

C'est l'approche la plus utilisée. Elle consiste à choisir la loi de la densité a priori et de calculer les paramètres de ces lois grâce aux données dont on dispose. Il est possible de calculer les paramètres a priori en s'appuyant sur les moments. Le principal problème réside dans l'estimation des moments car les queues de distribution peuvent avoir des effets non négligeables sur

ceux-ci. Une seconde méthode consiste à estimer des quantiles de la distribution a priori puis de choisir les paramètres de manière à respecter les contraintes.

Cette approche est la plus attrayante car il existe de nombreuses tables de quantiles pour les lois les plus fréquentes. Compte-tenu de l'information dont on dispose, on se concentrera par la suite sur cette méthode. En outre, on fait en sorte de modéliser la fréquence des pertes opérationnelles par une loi de Poisson et la sévérité de celles-ci par une loi log-normale : à cause des lois conjuguées, le choix de la loi a priori ne restreint qu'à une ou deux possibilités. On se basera donc au début de notre implémentation sur cette option.

3.2.2 Les lois non informatives

La section précédente a montré que les lois conjuguées peuvent être utiles en tant qu'approximation des véritables lois a priori. En revanche, lorsqu'aucune information a priori n'est disponible, leur unique justification est analytique, puisqu'elles donnent des expressions exactes pour quelques quantités a posteriori. Dans de telles situations, il est impossible de justifier le choix d'une loi a priori sur des bases subjectives et les paramètres des lois conjuguées ne peuvent être déterminés qu'arbitrairement. Plutôt que de revenir aux alternatives classiques, comme l'estimation par maximum de vraisemblance, ou d'utiliser les données pour estimer ces hyperparamètres, comme dans une analyse bayésienne empirique, il est préférable de faire appel à des techniques bayésiennes. Dans un tel cas, ces lois a priori particulières doivent être construites à partir de la distribution de l'échantillon, puisque c'est la seule information disponible. Pour ces raisons, de telles lois sont dites *non informatives*.

Les lois a priori de Laplace

Une première méthode consiste à donner la même vraisemblance à chaque valeur du paramètre par le biais d'une loi uniforme. On considère les transformations du problème qui ne doivent pas influencer intuitivement la loi a priori non informative, c'est-à-dire que "si deux problèmes ont la même structure, alors la loi a priori non informative doit être la même". Il s'agit donc de trouver une distribution invariante par une opération donnée. Si la famille de distributions est stable par une classe de transformations, on pourra chercher une distribution a priori invariante par cette famille de transformations. Cependant, des critiques ont été émises sur ce choix. Par exemple, les lois résultantes sont impropres quand l'espace des paramètres n'est pas compact.

La loi a priori de Jeffreys

Jeffreys propose une approche intrinsèque différente fondée sur l'information de Fisher, donnée par :

$$I(\theta) = \mathbb{E}_\theta \left[\left(\frac{\partial \ln f(X | \theta)}{\partial \theta} \right)^2 \right]$$

$I(\theta)$ représente une mesure de la quantité d'information sur θ contenue dans l'observation. Plus $I(\theta)$ est grande, plus l'observation apporte de l'information. Il semble alors naturel de favoriser (au sens de rendre plus probable suivant $\pi(\theta)$), les valeurs de θ pour lesquels $I(\theta)$ est grande : cela minimise l'influence de la loi a priori au profit de l'observation. On prendrait alors pour loi a priori la loi suivante :

$$\pi(\theta) = \sqrt{I(\theta)}$$

La loi a priori coïncidante

Une approche particulière de la modélisation non informative est de s'intéresser aux propriétés fréquentistes de la loi a priori, c'est-à-dire en moyenne sur x plutôt que conditionnellement à x . Il existe des lois a priori donnant des estimateurs optimaux selon des critères fréquentistes comme la minimaxité ou l'admissibilité et on peut souhaiter restreindre le choix de la loi a priori à ces distributions optimales. Cependant, une telle restriction réduit rarement le choix de la loi a priori à une distribution unique. Une approche plus standard est d'imposer que certaines probabilités a posteriori coïncident avec la couverture fréquentiste correspondante, d'où l'appellation de lois a priori coïncidentes.

3.3 Difficultés préliminaires et choix de modélisation

3.3.1 La définition de la loi a priori

Sans conteste, le point le plus critiquable et le plus critiqué de l'analyse bayésienne est le choix de la loi a priori. Une fois que cette loi a priori est connue, l'inférence peut être conduite mécaniquement comme on l'a vu précédemment grâce à la formule de Bayes. Dans la pratique, il est rare que l'information a priori soit suffisamment précise pour conduire à une détermination exacte de la loi a priori et c'est bien entendu le cas dans la modélisation apportée par les experts par le biais de scénarios. On décide dans un premier temps d'utiliser les lois conjuguées et une loi a priori prédéfinie pour modéliser notre risque opérationnel.

Cependant, les experts ne donnent qu'une estimation des pertes extrêmes (au-delà d'un certain seuil) ce qui fait qu'on n'a pas accès à l'ensemble de la distribution a priori. Dans un premier temps, on essaie d'implémenter la méthode la plus simple possible pour en dégager les premiers résultats et examiner sous quelle forme les problèmes se présentent.

Si l'information a priori est riche, la loi a priori sera bien entendue mieux définie que dans le cas non informatif. Dans tous les cas, il est important de s'assurer que l'impact de cette indétermination sur la loi a priori sur les quantités a posteriori soit bien évalué et que la partie arbitraire de l'a priori ne soit pas prédominante. L'étude de ces aspects est appelé étude de sensibilité (ou de robustesse). La littérature sur la robustesse bayésienne s'est considérablement accrue ces dernières années et nous renvoyons les lecteurs à l'article développé par [4].

3.3.2 Le calcul de la loi a posteriori

Dans le cas particulier des lois a priori conjuguées, les espérances a posteriori des paramètres naturels admettent évidemment des expressions explicites : mais c'est pratiquement le seul cas où des expressions analytiques sont disponibles dans une telle généralité. Pour certaines lois a priori, la loi a posteriori ne s'intègre pas si facilement que cela et on est obligé de recourir à des méthodes sophistiquées qui nécessitent beaucoup de calculs. Afin d'aborder l'ensemble des difficultés liées à la modélisation bayésienne, nous décrivons dans cette section l'ensemble des techniques permettant d'obtenir une loi a posteriori : de plus, on sera amené pour les méthodes que l'on présente après à implémenter ce genre de technique.

Pour une fonction de coût L et une loi a priori π données, l'estimation bayésienne associée

à une observation x est la décision d minimisant le coût a posteriori :

$$L(\pi, d | x) = \int_{\Theta} L(\theta, d)\pi(\theta | x)d\theta$$

Dans la pratique, minimiser peut être rendu difficile pour deux raisons :

- le calcul explicite de la loi a posteriori $\pi(\theta | x)$ peut être impossible (via la formule de Bayes)
- même si cette loi a posteriori est connue, cela n'implique pas nécessairement que minimiser l'expression L soit facile.

Des techniques classiques peuvent faciliter les calculs et sont particulièrement adaptées aux exigences de l'approche bayésienne. Il existe des méthodes de Monte Carlo plus générales permettant d'approcher la génération de variables aléatoires d'une loi a posteriori $\pi(\theta | x)$ lorsque cette loi ne peut pas être simulée directement. L'avantage de ces méthodes sur les méthodes de Monte Carlo classiques qu'on a abordé dans la section 2.2.3 est qu'elle ne nécessite pas la construction précise d'une fonction d'importance puisqu'elle prend en compte les caractéristiques de $\pi(\theta | x)$. Cette extension est appelée la **méthode de Monte Carlo par chaînes de Markov** qu'on note en abrégé la MCMC.

3.4 Mise en place de la méthode 1 : utilisation des lois conjuguées

On étudie de manière indépendante les distributions de probabilité de la fréquence et la sévérité des pertes. Nous avons adapté les travaux effectués par P.V. Schevchenko dans [29], [31] et [20].

3.4.1 La loi de fréquence des sinistres : loi de Poisson

Présentation du modèle

On suppose que la loi régissant le nombre de pertes opérationnelles survenant en une année suit une loi de Poisson de paramètre λ . Le paramètre d'intensité λ n'est pas connu et l'approche bayésienne le modélise par une variable aléatoire que l'on note Λ . Ensuite on peut appliquer le modèle suivant pour les années $t = 1, 2, \dots, T, T + 1$ (où $T + 1$ correspond à l'année future).

Hypothèses du modèle

- On suppose que pour $\Lambda = \lambda$ donné, les données N_1, \dots, N_{T+1} sont des variables aléatoires indépendantes et identiquement distribuées de loi de Poisson $\mathcal{P}(\lambda)$:

$$\mathbb{P}(N_t = n | \lambda) = \exp(-\lambda) \frac{\lambda^n}{n!}, \lambda \geq 0.$$

- On suppose que la distribution a priori de Λ est une loi Gamma, $\mathcal{G}(\alpha_0, \beta_0)$. La densité a priori s'écrit :

$$\pi(\lambda) = \frac{1}{\Gamma(\alpha_0)\beta_0} \left(\frac{\lambda}{\beta_0}\right)^{\alpha_0-1} \exp\left(-\frac{\lambda}{\beta_0}\right), \lambda > 0, \alpha_0 > 0, \beta_0 > 0.$$

Calcul de la distribution a posteriori

La fonction de vraisemblance de la loi de Poisson s'écrit :

$$h(\mathbf{n} | \lambda) = \prod_{i=1}^T e^{-\lambda} \frac{\lambda^{n_i}}{n_i!}$$

où $\mathbf{n} = (n_1, \dots, n_T)'$ représente l'historique du nombre de pertes annuelles. On peut alors déterminer la loi a posteriori à l'aide de la formule de Bayes suivante :

$$\pi(\lambda | \mathbf{n}, \alpha_0, \beta_0) = \frac{h(\mathbf{n} | \lambda) \pi(\lambda | \alpha_0, \beta_0)}{h(\mathbf{n})}$$

On obtient alors les relations suivantes :

$$\begin{aligned} \pi(\lambda | \mathbf{n}, \alpha_0, \beta_0) &= \frac{(\lambda/\beta_0)^{\alpha_0-1}}{h(\mathbf{n})\Gamma(\alpha_0)\beta_0} e^{-\lambda/\beta_0} \prod_{i=1}^T e^{-\lambda} \frac{\lambda^{n_i}}{n_i!} \\ &= \frac{\beta_0^{-1+\sum_{i=1}^T n_i} \prod_{i=1}^T \frac{1}{n_i!}}{h(\mathbf{n})\Gamma(\alpha_0)} \left(\frac{\lambda}{\beta_0}\right)^{\alpha_0-1+\sum_{i=1}^T n_i} e^{-\lambda(T+\frac{1}{\beta_0})} \end{aligned}$$

Etant donné que nous sommes dans un cas de lois conjuguées, il nous faut identifier la nouvelle loi Gamma qui se reconnaît facilement en écrivant la relation de proportionnalité suivante :

$$\pi(\lambda | \mathbf{n}, \alpha_0, \beta_0) \propto \left(\frac{\lambda}{\beta_0}\right)^{\alpha_0-1+\sum_{i=1}^T n_i} e^{-\lambda(T+\frac{1}{\beta_0})}$$

Avec les hypothèses précédentes, la densité a posteriori peut être écrite :

$$\pi(\lambda) = \frac{1}{\Gamma(\alpha_T)\beta_T} \left(\frac{\lambda}{\beta_T}\right)^{\alpha_T-1} \exp\left(\frac{-\lambda}{\beta_T}\right)$$

Ceci correspond à une distribution $\mathcal{G}(\alpha_T, \beta_T)$, la même distribution que la distribution a priori mais avec des paramètres différents :

$$\boxed{\alpha_T = \alpha_0 + \sum_{i=1}^T n_i, \quad \beta_T = \frac{\beta_0}{1 + \beta_0 T}}$$

Il est supposé que pour $\Lambda = \lambda$ donné, N_{T+1} et $\mathbf{N} = \{N_1, \dots, N_T\}$ sont indépendants. Le nombre espéré d'événements pour l'année suivante, sachant les observations du passé, $\mathbb{E}[N_{T+1} | \mathbf{N}]$, c'est à dire la moyenne de la loi Gamma $\mathcal{G}(\alpha_T, \beta_T)$ (correspondant à la moyenne de la distribution a posteriori), permet une bonne interprétation comme suit :

$$\mathbb{E}[\lambda | \mathbf{N} = \mathbf{n}] = \alpha_T \times \beta_T = \beta_0 \frac{\alpha_0 + \sum_{i=1}^T n_i}{1 + \beta_0 T} = \omega_T \hat{\lambda}_T^{MLE} + (1 - \omega_T) \lambda_0$$

Avec :

- $\hat{\lambda}_T^{MLE} = \frac{1}{T} \sum_{i=1}^T n_i$ est l'estimateur du maximum de vraisemblance de λ utilisant uniquement les observations ;
- $\lambda_0 = \alpha_0 \times \beta_0$ est l'estimation du λ utilisant uniquement la distribution a priori (spécifiée par les experts) ;
- $\omega_T = \frac{T\beta_0}{T\beta_0+1}$ est le poids de la crédibilité utilisé pour combiner $\hat{\lambda}_T^{MLE}$ et λ_0 .

On voit très bien dans cette formule que ω_T est le poids de crédibilité qu'on donne à l'estimateur du maximum de vraisemblance associé aux données internes par rapport à l'estimation d'un expert.

De cette formule, on peut faire alors plusieurs remarques : plus le nombre d'années observées T est grand, plus le poids de crédibilité est grand, cela veut dire qu'on assigne plus de poids aux observations qu'à l'estimation des experts. De plus, plus la volatilité de l'opinion de l'expert est grande, plus on accorde de poids aux observations.

Estimation de la distribution a priori

On suppose que la distribution de probabilité de la fréquence annuelle N des pertes pour le risque opérationnel est modélisée par une distribution de Poisson $\mathcal{P}(\Lambda = \lambda)$, et la densité a priori $\pi(\lambda)$ pour Λ est la loi Gamma $\mathcal{G}(\alpha_0, \beta_0)$.

On a alors $\mathbb{E}[N \mid \Lambda] = \Lambda$ et $\mathbb{E}[\Lambda] = \alpha_0 \times \beta_0$. L'expert doit donc spécifier la moyenne du nombre de sinistres mais il ne peut être sûr de cette estimation. Cette estimation doit plutôt être considérée comme la meilleure estimation faite par les experts pour le nombre de sinistres attendus correspondant à $\mathbb{E}[\mathbb{E}[N \mid \Lambda]] = \mathbb{E}[\Lambda]$. En plus de cette estimation, il est nécessaire que les experts fournissent une incertitude que le vrai λ pour l'année suivante se situe dans un intervalle $[a, b]$ avec une probabilité $p = \mathbb{P}[a \leq \Lambda \leq b]$ ce qui permet d'obtenir le système à deux équations et deux inconnues suivant :

$$\mathbb{E}[\Lambda] = \alpha_0 \times \beta_0$$

$$\mathbb{P}(a \leq \Lambda \leq b) = p = \int_a^b \pi(\lambda \mid \alpha_0, \beta_0) d\lambda = F_{\alpha_0, \beta_0}^{(G)}(b) - F_{\alpha_0, \beta_0}^{(G)}(a)$$

Ce système permet de calculer les paramètres structuraux α_0 et β_0 de la loi a priori Gamma. Ici $F_{\alpha_0, \beta_0}^{(G)}(\cdot)$ est la distribution Gamma $\mathcal{G}(\alpha_0, \beta_0)$, c'est à dire :

$$F_{\alpha_0, \beta_0}^{(G)}(y) = \int_0^y \frac{x^{\alpha_0-1}}{\Gamma(\alpha_0)\beta_0^{\alpha_0}} \exp\left(-\frac{x}{\beta_0}\right) dx$$

Le problème réside alors dans le fait d'adapter les informations fournies par les experts pour obtenir ces deux équations permettant ainsi d'obtenir les paramètres structuraux α_0 et β_0 .

Utilisation des informations fournies par les experts

A l'aide des couples (x_i, d_i) associés à un scénario pour $i \in \{1, 2, 3\}$, où x_i représente un montant et d_i le temps de retour associé à ce montant de perte, les experts ont estimé, par minimisation d'un critère quadratique présenté dans la partie 2, les différents paramètres des distributions de fréquence et sévérité. Parmi ces estimations, on retrouve λ_S représentant l'estimation de l'intensité de fréquence λ . On fait alors les choix de modélisation suivants :

1^{er} choix de modélisation : on considère que la meilleure estimation faite par les experts pour le nombre de sinistres attendus est ce λ_S . On peut alors écrire :

$$\mathbb{E}[N] = \lambda_S$$

Or :

$$\mathbb{E}[N] = \mathbb{E}[\mathbb{E}[N | \Lambda]] = \mathbb{E}[\Lambda] = \alpha_0 \times \beta_0$$

On obtient donc la première équation suivante :

$$\lambda_S = \alpha_0 \times \beta_0$$

2nd choix de modélisation : il s'agit ici de trouver une deuxième équation qui permettrait la résolution d'un système d'équations à deux inconnues. Un problème se pose : on ne dispose uniquement de l'espérance du nombre de pertes. Il paraît donc difficile d'obtenir la distribution du paramètre Λ à partir de cette seule donnée. Certaines informations sont particulièrement difficiles à obtenir comme les coefficients a , b , et la probabilité p telle que $\mathbb{P}[a \leq \Lambda \leq b] = p$ comme on l'a présenté pour estimer la distribution a priori. Par manque d'informations données par les experts pour la construction de la deuxième équation par la méthode présentée précédemment, il est aussi possible d'utiliser le coefficient de variation $Vco(\Lambda) = \sqrt{V(\Lambda)}/\mathbb{E}[\Lambda]$ très utilisé dans le monde assurantiel. Le choix a été de considérer un coefficient de variation. Pour une loi Gamma $\mathcal{G}(\alpha_0, \beta_0)$ on a :

$$\mathbb{E}(\Lambda) = \alpha_0 \times \beta_0^2 \quad \text{et} \quad V(\Lambda) = \alpha_0 \times \beta_0$$

La deuxième équation s'écrit donc :

$$Vco(\Lambda) = \frac{1}{\sqrt{\alpha_0}}$$

Résolution du système : Le système de deux équations à deux inconnues précédent se résout de manière explicite et l'on obtient les valeur suivantes pour les paramètres de la loi a priori :

$$\alpha_0 = \left(\frac{1}{Vco(\Lambda)} \right)^2$$

$$\beta_0 = \lambda_S (Vco(\Lambda))^2$$

Une fois ces paramètres a priori déterminés, nous pouvons en déduire les paramètres de la loi a posteriori $\mathcal{G}(\alpha_T, \beta_T)$ qui découlent des formules suivantes, démontrées précédemment :

$$\alpha_T = \alpha_0 + \sum_{i=1}^T n_i \quad \text{et} \quad \beta_T = \frac{\beta_0}{1 + \beta_0 T}$$

Après le calcul des paramètres a posteriori, on choisit d'utiliser la méthode de l'erreur quadratique minimale pour estimer la nouvelle valeur $\lambda_{posteriori}$ du paramètre de la loi $\mathcal{P}(\lambda)$ de la fréquence des sinistres :

$$\hat{\lambda}_{posteriori} = \mathbb{E}[\Lambda | N] = \int \lambda \Pi(\lambda | N) d\lambda$$

Ce qui permet d'écrire :

$$\hat{\lambda}_{posteriori} = \alpha_T \times \beta_T = \omega_T \hat{\lambda}_T^{MLE} + (1 - \omega_T) \lambda_0$$

Cette valeur pourra ensuite être utilisée pour modéliser la loi de fréquence dans le calcul de la charge en capital. On peut alors résumer la méthode sous la forme schématique suivante :

Résumé schématique de la méthode

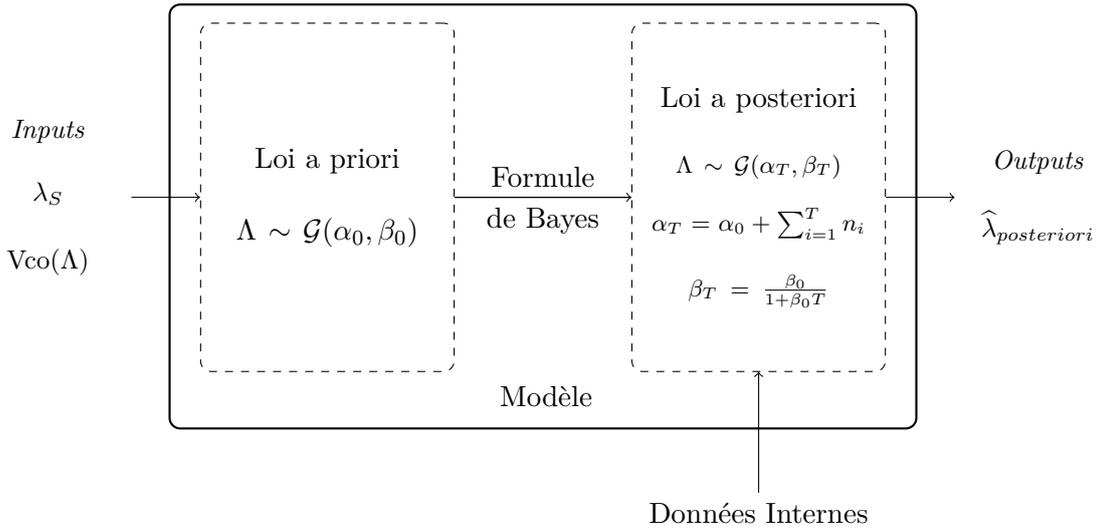


FIGURE 3.1 – Représentation schématique de la méthode 1 pour la fréquence des pertes

Applications numériques à nos données

Concernant les trois cellules de risque choisies, on vient de voir que dans chaque cas la valeur $E[\Lambda]$ correspond à la valeur λ_S fournie par les experts, ce qui permet de déduire les premières équations suivantes :

Cas	$E[\Lambda] = \lambda_S = \alpha_0 \times \beta_0$
1	0.02
2	0.22
3	0.2

Pour obtenir la deuxième équation, on choisit une valeur du $V_{co}(\Lambda)$ égale à 50% ce qui nous donne directement la valeur du coefficient α_0 égale à 4. On peut alors calculer les valeurs de β_0 et en déduire que la loi a priori de Λ est une loi Gamma $\mathcal{G}(\alpha_0, \beta_0)$ dont les valeurs des paramètres structuraux sont affichées dans le tableau ci-dessous :

Cas	α_0	β_0
1	4	0.005
2	4	0.055
3	4	0.05

TABLE 3.2 – Valeurs des paramètres a priori pour la fréquence

On a vu précédemment que si la loi de fréquence est une loi de Poisson de paramètre $\Lambda = \lambda$ et que la loi a priori de Λ est une loi Gamma $\mathcal{G}(\alpha_0, \beta_0)$, alors la loi a posteriori de Λ est une

loi Gamma $\mathcal{G}(\alpha_T, \beta_T)$. Les valeurs des nouveaux paramètres structuraux sont affichées dans le tableau suivant :

Cas	α_T	β_T
1	133	0.0049
2	599	0.0431
3	1311	0.04

TABLE 3.3 – Valeurs des paramètres a posteriori pour la fréquence

Les paramètres de loi a posteriori étant calculés, on obtient la nouvelle valeur $\hat{\lambda}_{posteriori}$ du paramètre de la loi $\mathcal{P}(\lambda)$ de la fréquence des sinistres en utilisant la méthode de l'erreur quadratique minimale. On obtient les valeurs suivantes :

Cas	$\hat{\lambda}_{posteriori}$
1	0.649
2	25.84
3	52.44

TABLE 3.4 – Valeurs finales du paramètre de fréquence $\hat{\lambda}_{posteriori}$

On peut alors comparer graphiquement les densités de fréquence :

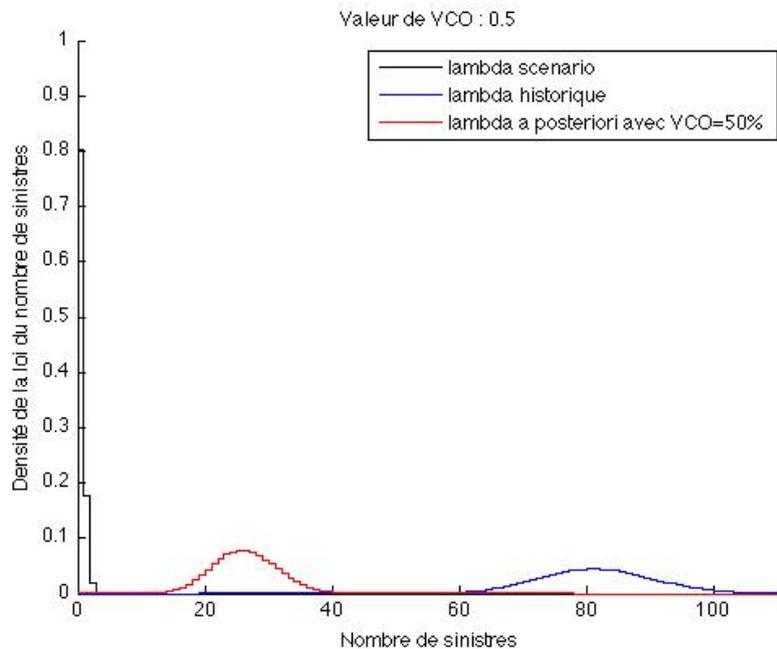


FIGURE 3.2 – Comparaison des densités de la fréquence pour le cas 2

On constate que la densité de la loi de fréquence obtenue se situe entre les deux densités obtenues avec les données internes et les scénarios. Ces résultats seront utilisés afin de déterminer les

charges en capital.

3.4.2 La loi de sévérité des pertes : loi log-normale $\mathcal{LN}(\mu, \sigma)$

Présentation du modèle

Supposons que la sévérité des pertes dans une banque est modélisée comme une variable aléatoire avec une distribution log-normale $\mathcal{LN}(\mu, \sigma)$. La densité se note :

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

Cette distribution très répandue est un bon moyen pour modéliser les pertes opérationnelles. En outre, elle appartient à une classe de distributions à queue épaisse. Les paramètres μ et σ ne sont pas connus et l'approche bayésienne les modélisent par des variables aléatoires Θ_μ et Θ_σ . On suppose que les pertes durant les années $t = 1, 2, \dots, T$ sont connues et doivent être modélisées pour l'année suivante $T + 1$. Pour simplifier les notations, on notera les pertes des T dernières années X_1, \dots, X_n et les pertes futures sont X_{n+1}, \dots .

Les hypothèses du modèle général

- On suppose que pour $\Theta_\sigma = \sigma$ et $\Theta_\mu = \mu$ donnés, X_1, \dots, X_n, \dots sont des variables aléatoires indépendantes de loi $\mathcal{LN}(\mu, \sigma)$. Les variables $Y_i = \ln X_i$ pour $i = 1, 2, \dots$ suivent donc une loi normale $\mathcal{N}(\mu, \sigma)$.
- On suppose que la distribution a priori de Θ_σ^2 est la loi inverse gamma $\mathcal{IG}(\nu/2, \beta/2)$ et que la distribution a priori de Θ_μ est une loi normale $\mathcal{N}(\theta, \sigma/\sqrt{\phi})$. Les densités a priori s'écrivent :

$$\pi(\sigma^2) = \frac{(\sigma^2/\beta)^{-1-\nu/2}}{\Gamma(\nu/2)\beta^{2\nu/2}} \exp\left(-\frac{\beta}{2\sigma^2}\right)$$

$$\pi(\mu | \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2/\phi}} \exp\left(-\frac{(\mu - \theta)^2}{2\sigma^2/\phi}\right)$$

De plus, on note les pertes des années passées sous forme vectorielle de la façon suivante $\mathbf{X} = (X_1, \dots, X_n)'$ et les logarithmes des pertes $\mathbf{Y} = (Y_1, \dots, Y_n)'$.

1er cas : cas particulier où μ est inconnu et σ connu

On considère ici que le paramètre σ ne suit pas une loi inverse-gamma mais est une constante connue et fixée. Dans ce cas, on pose $\mu_0 = \theta$ et $\sigma_0 = \sigma/\sqrt{\phi}$ ce qui nous permet d'écrire que Θ_μ suit une loi normale $\mathcal{N}(\mu_0, \sigma_0)$.

Calcul de la distribution a posteriori de Θ_μ

Sous ces hypothèses, la densité jointe des données des années passées à la position $\mathbf{Y} = \mathbf{y}$, conditionnellement aux valeurs σ et $\Theta_\mu = \mu$, s'écrit :

$$h(\mathbf{y} | \mu, \sigma) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_i - \mu)^2}{2\sigma^2}\right)$$

On peut alors déterminer la loi a posteriori à l'aide de la formule de Bayes suivante :

$$\pi(\mu \mid \mathbf{y}, \mu_0, \sigma_0) = \frac{h(\mathbf{y} \mid \mu, \sigma)\pi(\mu \mid \mu_0, \sigma_0)}{h(\mathbf{y} \mid \sigma)}$$

La densité a posteriori peut alors être écrite, à une constante multiplicative près, de la façon suivante :

$$\pi(\mu \mid \mathbf{y}, \mu_0, \sigma_0) \propto \frac{\exp\left(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right)}{\sigma_0\sqrt{2\pi}} \prod_{i=1}^n \frac{\exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right)}{\sigma\sqrt{2\pi}}$$

Etant donné que nous sommes dans un cas de loi conjuguée, on sait que la loi a posteriori de Θ_μ est une loi normale $\mathcal{N}(\mu_{0,n}, \sigma_{0,n})$:

$$\pi(\mu \mid \mathbf{y}, \mu_0, \sigma_0) = \exp\left(-\frac{(\mu - \mu_{0,n})^2}{2\sigma_{0,n}^2}\right)$$

En identifiant les termes du polynôme en μ se trouvant dans l'exponentielle, on obtient un système de deux équations à deux inconnues permettant de trouver les valeurs des nouveaux paramètres structuraux $\mu_{0,n}$ et $\sigma_{0,n}$:

$$\begin{cases} \text{terme en } \mu^2 : \frac{1}{2\sigma_{0,n}^2} = \frac{1}{2\sigma_0^2} + \frac{n}{2\sigma^2} \\ \text{terme en } \mu : \frac{\mu_{0,n}}{\sigma_{0,n}^2} = \frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^n y_i}{\sigma^2} \end{cases}$$

On peut conclure que Θ_μ suit a posteriori une loi normale $\mathcal{N}(\mu_{0,n}, \sigma_{0,n})$ telle que :

$$\boxed{\mu_{0,n} = \frac{\mu_0 + \omega \sum_{i=1}^n \ln(x_i)}{1 + n\omega} \quad , \quad \sigma_{0,n}^2 = \frac{\sigma_0^2}{1 + n\omega}}$$

avec $\omega = \sigma_0^2/\sigma^2$.

La valeur attendue de Y_{n+1} sachant les observations du passé, $\mathbb{E}[Y_{n+1} \mid \mathbf{Y} = \mathbf{y}]$, se calcule de la façon suivante :

$$\mathbb{E}[Y_{n+1} \mid \mathbf{Y} = \mathbf{y}] = \mathbb{E}[\Theta_\mu \mid \mathbf{Y} = \mathbf{y}] = \mu_{0,n} = \frac{\mu_0 + \omega \sum_{i=1}^n y_i}{1 + n\omega} = \omega_n \bar{y}_n + (1 - \omega_n)\mu_0$$

Avec :

- $\bar{y}_n = \frac{1}{n} \sum_{i=1}^n y_i$ est l'estimateur du maximum de vraisemblance de λ utilisant uniquement les observations ;
- μ_0 est l'estimation du μ utilisant uniquement la distribution a priori (spécifiée par les experts) ;
- $\omega_n = \frac{n}{n+\sigma^2/\sigma_0^2}$ est le poids de la crédibilité utilisé pour combiné \bar{y}_n et μ_0 .

On voit très bien dans cette formule que ω_n est le poids de crédibilité qu'on donne à l'estimateur du maximum de vraisemblance par rapport à l'estimation d'un expert.

De cette formule, on peut faire alors plusieurs remarques : lorsque le nombre d'observations augmente, le poids de crédibilité augmente, cela veut dire qu'on assigne plus de poids aux observations qu'à l'estimation des experts. De plus, plus la volatilité de l'opinion de l'expert est grande, plus on accorde de poids aux observations.

Estimation de la distribution a priori de Θ_μ

On suppose que la distribution de probabilité de la sévérité X des pertes pour le risque opérationnel est modélisée par une distribution log-normale $\mathcal{LN}(\mu, \sigma)$, et la densité a priori $\pi(\mu)$ pour Θ_μ est la loi normale $\mathcal{N}(\mu_0, \sigma_0)$.

La perte attendue s'écrit :

$$\Omega = \mathbb{E}[X \mid \Theta_\mu] = \exp\left(\Theta_\mu + \frac{1}{2}\sigma^2\right)$$

Et le quantile de niveau q s'écrit :

$$\mathcal{Q}_q = \exp\left(\Theta_\mu + \sigma\Phi^{-1}(q)\right)$$

Dans ce cas, Ω a une distribution $\mathcal{LN}(\mu_0 + 1/2\sigma^2, \sigma_0)$. L'expert doit spécifier la meilleure estimation des pertes attendues $E[\Omega]$ mais il ne peut être sûr de cette estimation. Cette estimation doit plutôt être considérée comme la meilleure estimation faite par les experts pour la sévérité des sinistres correspondant à $\mathbb{E}[\mathbb{E}[X \mid \Theta_\mu]] = \mathbb{E}[\Omega]$. En plus de cette estimation, il est nécessaire que les experts fournissent une incertitude que le vrai Ω pour l'année suivante se situe dans un intervalle $[a, b]$ avec une probabilité $\mathbb{P}[a \leq \Omega \leq b] = p$ ce qui permet d'obtenir le système à deux équations et deux inconnues suivant :

$$\begin{aligned} \mathbb{E}[\Omega] &= \exp\left(\mu_0 + \frac{1}{2}\sigma^2 + \frac{1}{2}\sigma_0^2\right) \\ p = \mathbb{P}[a < \Omega < b] &= \Phi\left(\frac{\ln b - \frac{1}{2}\sigma^2 - \mu_0}{\sigma_0}\right) - \Phi\left(\frac{\ln a - \frac{1}{2}\sigma^2 - \mu_0}{\sigma_0}\right) \end{aligned}$$

Le problème résulte alors dans le fait d'adapter les informations fournies par les experts pour obtenir ces deux équations.

Utilisation des informations fournies par les experts

A l'aide des couples fournis par les experts (x_i, d_i) associés à un scénario pour $i \in \{1, 2, 3\}$, où x_i représente un montant et d_i le temps de retour associé à ce montant de pertes, on estime par minimisation d'un critère quadratique les différents paramètres des distributions de fréquence et sévérité. Parmi ces estimations, on retrouve μ_S décrivant une partie de la loi de sévérité spécifiée par les experts. On fait alors les choix de modélisation suivants :

1^{er} choix de modélisation : on considère que la meilleure estimation faite par les experts pour la sévérité des sinistres attendus est l'espérance de la loi log-normale spécifiée par les experts. On est ici dans le cas où σ est connu, mais on constate qu'il n'est pas nécessaire de choisir sa valeur pour déterminer les paramètres a priori de la loi de Θ_μ . La première équation s'écrit :

$$\exp(\mu_S + \frac{1}{2}\sigma^2) = \exp\left(\mu_0 + \frac{1}{2}\sigma^2 + \frac{1}{2}\sigma_0^2\right)$$

2nd choix de modélisation : il s'agit ici de trouver une deuxième équation qui permettrait en quelque sorte de calculer le moment d'ordre 2. Etant confrontés au même problème que pour la fréquence, on a recours ici au coefficient de variation $Vco(\Theta_\mu) = \sqrt{V(\Theta_\mu)}/\mathbb{E}[\Theta_\mu]$. La deuxième équation s'écrit alors :

$$Vco(\Theta_\mu) = \frac{\sigma_0}{\mu_0}$$

Résolution du système : Le système de deux équations à deux inconnues précédent se résout de manière explicite et l'on obtient les valeurs suivantes pour les paramètres de la loi a priori :

$$\mu_0 = \frac{-1 + \sqrt{1 + 2(Vco(\Theta_\mu))^2 \mu_S}}{(Vco(\Theta_\mu))^2}$$

$$\sigma_0 = \frac{-1 + \sqrt{1 + 2(Vco(\Theta_\mu))^2 \mu_S}}{Vco(\Theta_\mu)}$$

Une fois ces paramètres a priori déterminés, on peut en déduire les paramètres de la loi a posteriori $\mathcal{N}(\mu_{0,n}, \sigma_{0,n})$ en utilisant les formules de passages déterminées précédemment. Après le calcul des paramètres a posteriori, on choisit comme pour la fréquence d'utiliser la méthode de l'erreur quadratique minimale pour estimer la nouvelle valeur $\hat{\mu}_{posteriori}$ du paramètre d'espérance de la loi $\mathcal{LN}(\mu, \sigma)$ de sévérité des sinistres. Cette valeur ainsi que celle de σ_{DI} seront ensuite utilisées pour modéliser la loi de sévérité des sinistres dans le calcul des charges en capital. On peut résumer la méthode sous la forme schématique suivante :

Résumé de la méthode

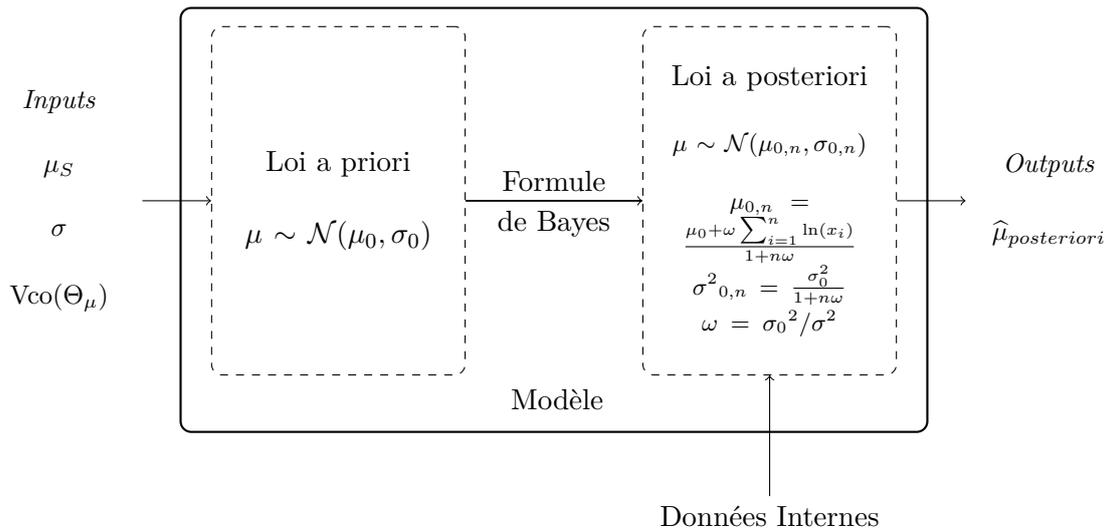


FIGURE 3.3 – Représentation schématique de la méthode 1 pour la sévérité dans le cas σ connu

2nd cas : cas général où μ et σ sont inconnus
Calcul des distribution a posteriori de Θ_μ et Θ_σ^2

Sous les hypothèses du modèle général, la densité jointe a priori est :

$$\begin{aligned}\pi(\mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2/\phi}} \exp\left(-\frac{(\mu - \theta)^2}{2\sigma^2/\phi} - \frac{\beta}{2\sigma^2}\right) \times \frac{2^{-\nu/2}}{\beta\Gamma(\nu/2)} \left(\frac{\sigma^2}{\beta}\right)^{-\frac{\nu}{2}-1} \\ &\propto (\sigma^2)^{-\frac{\nu+1}{2}-1} \exp\left(-\frac{1}{2\sigma^2}(\beta + \phi(\mu - \theta)^2)\right)\end{aligned}$$

On peut voir assez simplement que la distribution marginale a priori de Θ_μ est une distribution de Student translatée à ν degrés de liberté :

$$\begin{aligned}\pi(\mu) &= \int \pi(\mu, \sigma^2) d\sigma^2 \propto \int x^{-\frac{\nu+1}{2}-1} \exp\left(-\frac{1}{2x}[\beta + \phi(\mu - \theta)^2]\right) dx \\ &\propto \int y^{\frac{\nu+1}{2}-1} \exp\left(-\frac{y}{2}[\beta + \phi(\mu - \theta)^2]\right) dy \\ &\propto [\beta + \phi(\mu - \theta)^2]^{-\frac{\nu+1}{2}} \int z^{\frac{\nu+1}{2}-1} \exp(-z) dz \\ &\propto \left(1 + \frac{\phi\nu(\mu - \theta)^2}{\nu\beta}\right)^{-\frac{\nu+1}{2}}\end{aligned}$$

On note $\Psi_\sigma = (\sigma^2)^{-\frac{\nu+1+n}{2}-1}$, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ et $\bar{y}^2 = \frac{1}{n} \sum_{i=1}^n y_i^2$. Sachant $\mathbf{Y} = \mathbf{y}$, la densité jointe a posteriori a la forme suivante :

$$\begin{aligned}\pi(\mu, \sigma^2 | \mathbf{y}) &\propto \Psi_\sigma \exp\left(-\frac{1}{2\sigma^2} \left(\beta + \phi(\mu - \theta)^2 + \sum_{i=1}^n (y_i - \mu)^2\right)\right) \\ &\propto \Psi_\sigma \exp\left(-\frac{1}{2\sigma^2} \left(\beta + (\phi + n)\mu^2 + \phi\theta^2 - 2\mu(\phi\theta + n\bar{y}) + n\bar{y}^2\right)\right) \\ &\propto \Psi_\sigma \exp\left(-\frac{1}{2\sigma^2} \left(\beta + \phi\theta^2 + n\bar{y}^2 - \frac{(\phi\theta + n\bar{y})^2}{\phi + n} + (\phi + n) \left(\mu - \frac{\phi\theta + n\bar{y}}{\phi + n}\right)^2\right)\right) \\ &\propto (\sigma^2)^{-\frac{\nu+n+1}{2}-1} \exp\left(-\frac{1}{2\sigma^2}(\beta_n + \phi_n(\mu - \theta_n)^2)\right)\end{aligned}$$

Cette distribution jointe a posteriori a la même forme que la densité jointe a priori mais avec les nouvelles valeurs des paramètres structuraux suivantes :

$\begin{aligned}\nu_n &= \nu + n \\ \beta_n &= \beta + \phi\theta^2 + n\bar{y}^2 - \frac{(\phi\theta + n\bar{y})^2}{\phi + n} \\ \theta_n &= \frac{\phi\theta + n\bar{y}}{\phi + n} \\ \phi_n &= \phi + n\end{aligned}$

Les lois a posteriori des variables Θ_σ^2 et Θ_μ sont donc respectivement une loi inverse-gamma de paramètres $\nu_n/2$ et $\beta_n/2$ et une loi normale de paramètre θ_n et $\sigma/\sqrt{\phi_n}$. La variance attendue des pertes de l'année future sachant les observations du passé, $\mathbb{E}[\Theta_\sigma^2 \mid \mathbf{Y} = \mathbf{y}]$, s'exprime de la façon suivante :

$$\mathbb{E}[\Theta_\sigma^2 \mid \mathbf{Y} = \mathbf{y}] = \frac{\beta_n}{\nu_n - 2}$$

Estimation des distributions a priori de Θ_μ et Θ_σ

Pour Θ_μ et Θ_σ donnés, la perte attendue s'écrit :

$$\Omega = \mathbb{E}[X \mid \Theta_\mu, \Theta_\sigma] = \exp\left(\Theta_\mu + \frac{1}{2}\Theta_\sigma^2\right)$$

Et le quantile de niveau q s'écrit :

$$Q_q = \exp\left(\Theta_\mu + \Theta_\sigma \Phi^{-1}(q)\right)$$

On peut ensuite trouver Θ_σ grâce à deux quantiles Q_{q_1} et Q_{q_2} de la façon suivante :

$$\Theta_\sigma = \frac{\ln(Q_{q_2}/Q_{q_1})}{\Phi^{-1}(q_2) - \Phi^{-1}(q_1)}$$

Ensuite, on peut essayer de modéliser la distribution a priori de Θ_σ à l'aide des avis d'experts sur les quantités $\mathbb{E}[\ln(Q_{q_2}/Q_{q_1})]$ et $\mathbb{P}[a \leq Q_{q_2}/Q_{q_1} \leq b]$ ou des opinions impliquant plusieurs paires de quantiles. Enfin connaissant la meilleure valeur de σ , la distribution a priori de Θ_μ peut être estimée comme dans le cas où σ est connu.

Le problème résulte alors, encore une fois, dans le fait d'adapter les informations fournies par les experts pour obtenir deux équations qui nous permettront d'estimer les paramètres de la loi a priori de Θ_σ^2 .

Utilisation des informations fournies par les experts

Les informations fournies aujourd'hui par les experts métier du Crédit Agricole ne permettent pas de se prononcer sur les quantités $\mathbb{E}[\ln(Q_{q_2}/Q_{q_1})]$ et $\mathbb{P}[a \leq Q_{q_2}/Q_{q_1} \leq b]$. On dispose uniquement des valeurs μ_S et σ_S décrivant la loi de sévérité, spécifiées à partir des jugements d'experts et issues de la minimisation d'un critère quadratique. On fait alors les choix de modélisation suivants :

1^{er} choix de modélisation : on considère que la meilleure estimation faite par les experts pour le paramètre de variance de la sévérité des sinistres est égale à σ_S^2 . On considère donc, dans un souci de manque de données, que cette dernière valeur est égale à l'espérance de loi de Θ_σ^2 ce qui nous permet d'obtenir une première équation :

$$\sigma_S^2 = \mathbb{E}[\Theta_\sigma^2] = \frac{\beta}{\nu - 2}$$

2nd choix de modélisation : La deuxième équation permettant d'estimer les paramètres de loi a priori de Θ_σ^2 est obtenue en utilisant comme précédemment le coefficient de variation $Vco(\Theta_\sigma^2)$. On obtient la seconde équation suivante :

$$Vco(\Theta_\sigma^2) = \sqrt{\frac{2}{\nu - 4}}$$

Résolution du système : Le système de deux équations à deux inconnues précédent se résout de manière explicite et l'on obtient les valeurs suivantes pour les paramètres de la loi a priori :

$$\nu = \frac{2}{(Vco(\Theta_\sigma^2))^2} + 4$$

$$\beta = 2\sigma_S^2 \left(\frac{1}{(Vco(\Theta_\sigma^2))^2} + 1 \right)$$

Une fois tous les paramètres a priori déterminés, on peut en déduire les paramètres des lois a posteriori $\mathcal{IG}(\nu_n/2, \beta_n/2)$ $\mathcal{N}(\mu_n, \sigma/\phi_n)$ en utilisant les formules de passage déterminées précédemment. Après le calcul des paramètres a posteriori, on choisit comme pour la fréquence d'utiliser la méthode de l'erreur quadratique minimale pour estimer les nouvelles valeurs $\hat{\sigma}_{posteriori}$ $\hat{\mu}_{posteriori}$ des paramètres de la loi $\mathcal{LN}(\mu, \sigma)$ de sévérité des sinistres. On résume la méthode sous la forme schématique suivante :

Résumé de la méthode

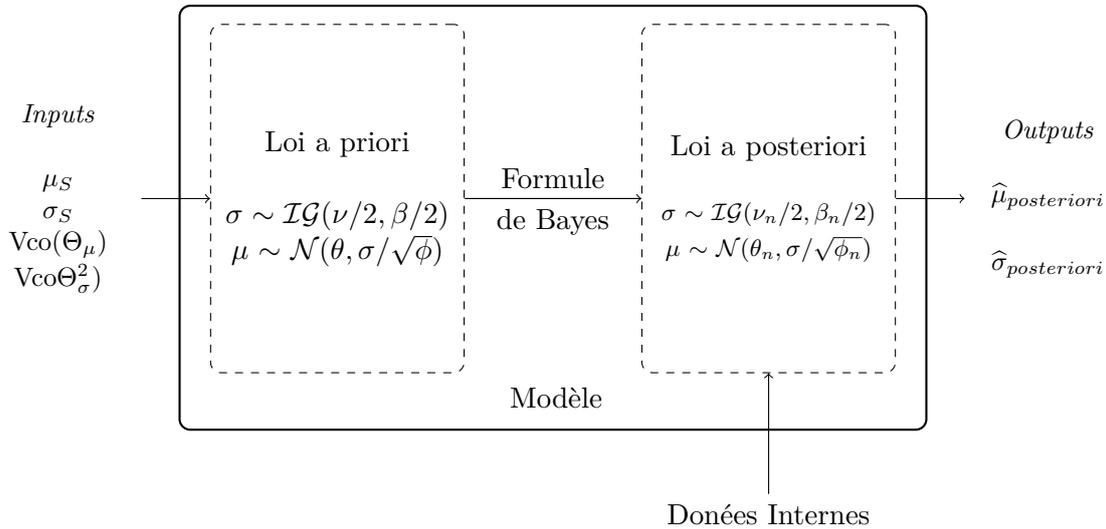


FIGURE 3.4 – Représentation schématique de la méthode 1 pour la sévérité dans le cas où μ et σ sont inconnus

Application numérique à nos données

Concernant les trois cellules de risques choisies, on a vu que les valeurs $\mu_S, \sigma_S, \text{Vco}(\Theta_\mu)$ et $\text{Vco}(\Theta_\sigma^2)$ permettent de déterminer les paramètres structuraux des lois a priori. On choisit une valeur de 50% pour les quantités $\text{Vco}(\Theta_\mu)$ et $\text{Vco}(\Theta_\sigma^2)$. On obtient alors les résultats suivants :

Cas	μ inconnu et σ connu		Cas	μ et σ inconnus			
	μ_0	σ_0		θ	ϕ	ν	β
1	8.17	4.08	1	8.17	0.0069	12	1.16
2	7.13	3.56	2	7.13	0.0397	12	5.04
3	8.1	4.05	3	8.1	0.025	12	4.1

TABLE 3.5 – Valeurs des paramètres a priori pour la sévérité

A l'aide des formules de passage décrites dans cette partie, on peut calculer les paramètres structuraux des lois a posteriori. On affiche ces valeurs dans le tableau suivant :

Cas	μ inconnu et σ connu		Cas	μ et σ inconnus			
	$\mu_{0,n}$	$\sigma_{0,n}$		θ_n	ϕ_n	ν_n	β_n
1	8.15	0.03	1	8.15	129	141	272.44
2	5.76	0.04	2	5.76	308	320	1246.7
3	6.71	0.018	3	6.71	1307	1319	5588.8

TABLE 3.6 – Valeurs des paramètres a posteriori pour la sévérité

Les paramètres de loi a posteriori étant calculés, on a accès aux nouvelles valeurs des paramètres de la loi de sévérité des pertes $\mathcal{LN}(\mu, \sigma)$ en utilisant la méthode de l'erreur quadratique minimale. On obtient les valeurs suivantes :

Cas	μ inconnu et σ connu		Cas	μ et σ inconnus	
	$\mu = \hat{\mu}_{posteriori}$	$\sigma = \hat{\sigma}_{DI}$		$\mu = \hat{\mu}_{posteriori}$	$\sigma = \hat{\sigma}_{posteriori}$
1	8.15	1.61	1	8.15	1.4
2	5.76	2.17	2	5.76	1.98
3	6.71	2.24	3	6.71	2.06

TABLE 3.7 – Valeurs finales des paramètres de sévérité $\hat{\mu}_{posteriori}$ et $\hat{\sigma}_{posteriori}$

On peut alors regarder graphiquement l'évolution de la distribution de la sévérité :

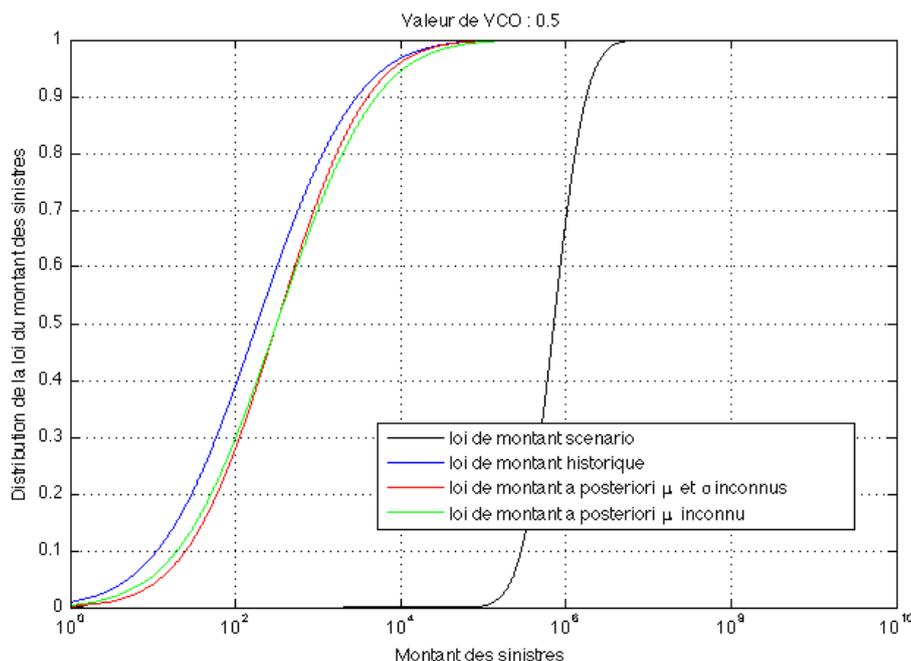


FIGURE 3.5 – Evolution de la distribution de sévérité

Les distributions de sévérité a posteriori se rapprochent de celle de la sévérité associée aux pertes internes.

3.5 Calcul des charges en capital CaR et sensibilité

3.5.1 Calcul des charges en capital

On s’intéresse dans ce qui suit à calculer les charges en capital requis **CaR** dans le cas des trois cellules de risques choisies. On utilise l’approche fréquence/sévérité décrite précédemment en modélisant la fréquence par une loi de Poisson $\mathcal{P}(\lambda)$ et la sévérité par une loi log-normale $\mathcal{LN}(\mu, \sigma)$. Avec ce premier modèle, on envisage le calcul de la CaR pour les trois cellules dans les quatres cas différents ci-dessous :

		Paramètres de sévérité	
		μ inconnu et σ connu	μ et σ inconnus
Paramètre de fréquence	λ inconnu	CaR_{M1}	CaR_{M2}
	λ connu	CaR_{M3}	CaR_{M4}

FIGURE 3.6 – Cas de calcul de la CaR pour la méthode 1

Afin de déterminer les CaR pour les différents cas, il est nécessaire d’avoir calculé au préalable les trois paramètres λ , μ et σ . Pour les cas où λ est inconnu, on utilise la valeur λ_{DI} obtenue par l’utilisation des données internes car cette valeur est vérifiée par les experts et peut être corrigée par une nouvelle valeur lorsque les experts considèrent que l’estimation qui est faite n’est pas représentative. Pour les cas où σ est connu, on utilise la valeur σ_{DI} résultant de

l'estimation faite à partir des données internes et de la méthode des moments généralisée. Pour les paramètres inconnus, la méthode 1 des lois conjuguées est ici utilisée et les valeurs retenues sont celles obtenues par la méthode de l'erreur quadratique minimale. On peut alors à l'aide de l'algorithme de Monte Carlo déduire les CaR correspondant aux VaR(99.9%). Les montants obtenus sont affichés dans le tableau suivant :

Cas	CaR _{M1}	CaR _{M2}	CaR _{M3}	CaR _{M4}
1	423 830 €	223 160 €	3 401 800 €	1 663 800 €
2	1 819 700 €	857 410 €	3 326 500 €	1 529 100 €
3	8 922 100 €	4 303 100 €	39 895 000 €	19 324 000 €

TABLE 3.8 – Montants des différentes charges en capital calculés selon la méthode 1 des lois conjuguées

Les résultats présentés dans ce tableau soulignent **une forte disparité entre les charges en capital obtenues avec les différentes méthodes**. On s'intéresse ici aux modifications qui ont été effectuées pour le calcul des CaR_{M_i}.

Pour la CaR_{M3}, la seule différence avec la CaR_{DI} est l'application de la méthode bayésienne des lois conjuguées pour le paramètre μ . On rappelle que le paramètre μ a posteriori est obtenu grâce à l'égalité suivante :

$$\hat{\mu}_{posteriori} = \mathbb{E}[\Theta_\mu | \mathbf{Y} = \mathbf{y}] = \frac{\mu_0 + \omega \sum_{i=1}^n y_i}{1 + n\omega} = \omega_n \bar{y}_n + (1 - \omega_n) \mu_0$$

Où $\omega_n = \frac{n}{n + \sigma^2/\sigma_0^2}$ avec n le nombre de pertes collectées, σ^2 la variance de la loi log-normale et σ_0^2 la variance de la loi normale a priori. Pour chacune des cellules de risque, on obtient un poids ω_n égale à 99.99%. On n'accorde donc aucune attention à l'information des experts s'exprimant à travers la valeur μ_0 . Le $\hat{\mu}_{posteriori}$ est donc égal à la moyenne des logarithmes des pertes qui correspond également au maximum de vraisemblance. On obtient des valeurs de charge en capital CaR_{M3} supérieures aux valeurs CaR_{DI} car la méthode des moments généralisée donne des valeurs inférieures à \bar{y}_n . On note tout de même que les valeurs CaR_{M3} sont comprises entre la CaR_{DI} et la CaR_S.

Pour la CaR_{M1}, la différence avec la CaR_{M3} est l'application de la méthode bayésienne des lois conjuguées pour le paramètre λ . On rappelle que le paramètre λ a posteriori est obtenu grâce à l'égalité suivante :

$$\hat{\lambda}_{posteriori} = \mathbb{E}[\Lambda | \mathbf{N} = \mathbf{n}] = \omega_T \frac{\sum_{i=1}^T N_i}{T} + (1 - \omega_T) \lambda_0$$

Où $\omega_T = \frac{T}{T + 1/\beta_0}$ avec T le nombre d'années d'historique égale ici à 5 et β_0 la valeur d'un des paramètres a priori pour le paramètre de fréquence. On constate que les valeurs de β_0 estimées a priori sont très faibles ce qui fait que le terme $1/\beta_0$ est supérieur à T . On trouve donc des poids de crédibilité ω_T variant entre 2.5% et 20%. On accorde donc un poids significatif, supérieur à 80% selon les cas, aux données des experts $\lambda_0 = \lambda_S$ qui par conséquent amène à trouver des valeurs de charge en capital plus faibles.

Pour les CaR_{M2} et CaR_{M4}, que l'on peut comparer respectivement avec les CaR_{M1} et CaR_{M3}, la seule différence est l'application de la méthode bayésienne des lois conjuguées pour le paramètre σ . De la même manière que pour le paramètre μ , les valeurs $\hat{\sigma}_{posteriori}$ sont proches des valeurs estimées par maximum de vraisemblance mais différentes de celles obtenues par la

méthode des moments généralisée. En l'occurrence, elles sont plus faibles dans notre étude ce qui donne par conséquent des valeurs des CaR_{M2} et CaR_{M4} inférieures à CaR_{M1} et CaR_{M3} . On note tout de même que les CaR_{M4} restent proche des CaR_{DI} .

Au vu de ces résultats, on voit que le choix d'une des 4 méthodes est déterminant pour le résultat de la charge en capital requis. On effectue dans la partie suivante des sensibilités aux paramètres d'entrée du modèle pour voir sur quel paramètre il est le plus judicieux d'utiliser cette méthode bayésienne des lois conjuguées.

3.5.2 Sensibilités aux paramètres d'entrée du modèle

On vient de déterminer les montants de charge en capital requis en s'appuyant sur certaines hypothèses concernant les lois a priori. Afin de compléter ces premiers calculs, il semble intéressant de poursuivre par différentes analyses de sensibilité. On cherche donc à déterminer la sensibilité de la CaR aux variations des paramètres d'entrée du modèle. Ces sensibilités seront réalisées pour le cas 1.

Sensibilités au $V_{co}(\Lambda)$

On rappelle ici que la loi a priori du paramètre de fréquence Λ est une loi gamma $\mathcal{G}(\alpha_0, \beta_0)$ dont les paramètres structuraux α_0 et β_0 sont déterminés à partir d'une valeur de fréquence imposée par les experts λ_S ainsi que le coefficient de variation $V_{co}(\Lambda)$ que l'on a choisi de fixer arbitrairement à 50%. On cherche donc à déterminer la sensibilité de CaR_{M1} et CaR_{M2} aux variations de valeur de $V_{co}(\Lambda)$. On fait varier ce coefficient sur l'intervalle $[0.25, 0.75]$ par pas de 0.02 :

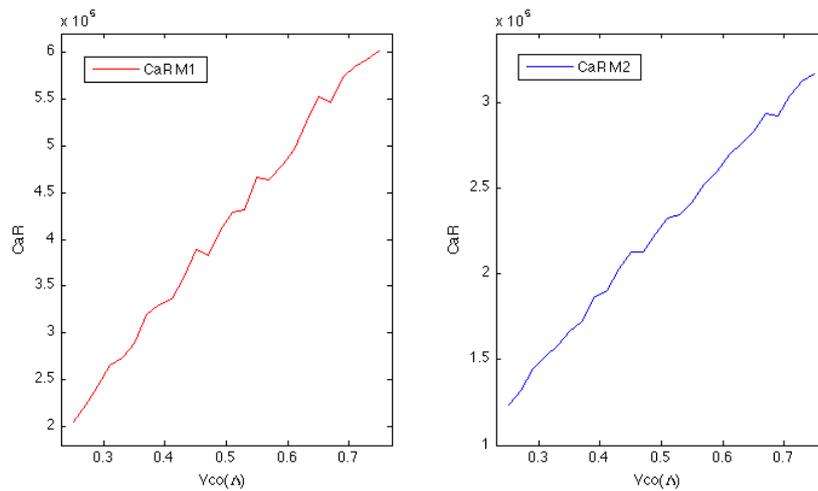


FIGURE 3.7 – Evolution de la CaR en fonction de $V_{co}(\Lambda)$

Au vu de ces résultats, on constate que **la charge en capital est sensible à la valeur de $V_{co}(\Lambda)$ utilisée en entrée du modèle**. Ces graphiques suggèrent que la CaR est une fonction affine croissante du $V_{co}(\Lambda)$, ce qui est dû au fait que la valeur $\hat{\lambda}_{posteriori}$ augmente de manière quasi linéaire avec $V_{co}(\Lambda)$.

Sensibilités au $Vco(\Theta_\mu)$

On rappelle ici que la loi a priori du paramètre de sévérité Θ_μ est une loi normale $\mathcal{N}(\mu_0 = \theta, \sigma_0 = \sigma/\sqrt{\phi})$ dont les paramètres structuraux sont déterminés à partir d'une valeur imposée par les experts μ_S ainsi que le coefficient de variation $Vco(\Theta_\mu)$ que l'on a choisi de fixer arbitrairement à 50%. On cherche donc à déterminer la sensibilité de CaR_{M3} et CaR_{M4} aux variations de valeur de $Vco(\Theta_\mu)$. On fait varier ce coefficient sur l'intervalle $[0.25, 0.75]$ par pas de 0.01 :

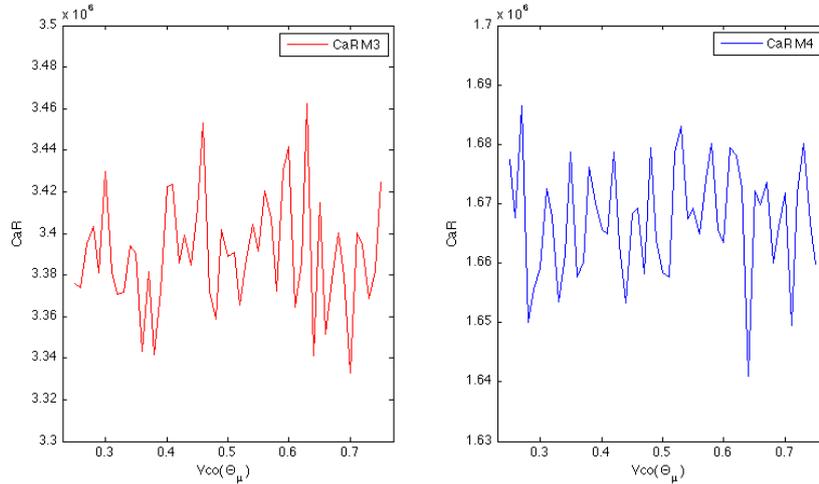


FIGURE 3.8 – Evolution de la CaR en fonction de $Vco(\Theta_\mu)$

Ces résultats suggèrent que **la charge en capital n'est pas sensible aux variations du paramètre $Vco(\Theta_\mu)$** .

Sensibilités au $Vco(\Theta_\sigma^2)$

On rappelle ici que la loi a priori du paramètre de sévérité Θ_σ^2 est une loi inverse-gamma $\mathcal{IG}(\nu/2, \beta/2)$ dont les paramètres structuraux sont déterminés à partir d'une valeur imposée par les experts $\hat{\sigma}_S$ ainsi que le coefficient de variation $Vco(\Theta_\sigma^2)$ que l'on a choisi de fixer arbitrairement à 50%. On cherche donc à déterminer la sensibilité de CaR_{M4} aux variations de valeur de $Vco(\Theta_\sigma^2)$. On fait varier ce coefficient sur l'intervalle $[0.25, 0.75]$ par pas de 0.01 :

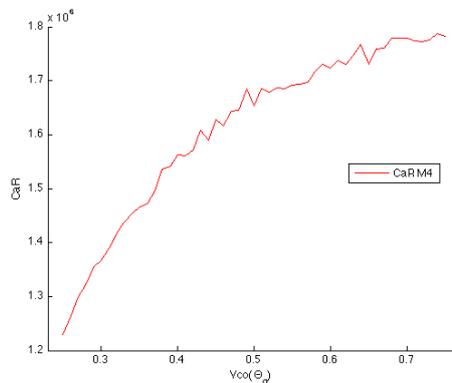


FIGURE 3.9 – Evolution de la CaR en fonction de $Vco(\Theta_\sigma^2)$

De même que pour le $Vco(\Lambda)$, on constate que **la charge en capital est sensible à la valeur de**

Vco(Θ_σ^2) utilisée en entrée du modèle. Ces graphiques suggèrent que la CaR est une fonction croissante du Vco(Θ_σ^2) ce qui est dû au fait que la valeur $\hat{\sigma}_{posteriori}$ augmente avec Vco(Θ_σ^2). La variation de valeur de la CaR, sur l'intervalle d'étude, correspond à un écart relatif de 20% par rapport à la valeur obtenus avec Vco(Θ_σ^2) = 50%. L'approche M3 semble donc être celle la plus adaptée avec cette méthode bayésienne des lois conjuguées.

3.6 Discussion sur la méthode

Dans cette première méthode que l'on peut surnommer approche bayésienne pure, il est possible, à l'aide de formules explicites, d'avoir accès à des paramètres de lois permettant le calcul d'un unique montant de capital réglementaire. Cette méthodologie utilise les principes de l'inférence bayésienne et combine les informations fournies par les experts métier ainsi que les données internes à travers l'utilisation de lois dites "conjuguées". La principale motivation qui pousse à utiliser une telle approche bayésienne est de **faire entrer l'expertise métier dans l'analyse statistique du profil de risque en estimant une distribution jointe unique qui combine les deux sources de données**. Globalement, l'utilisation des méthodes d'inférence bayésienne pour l'estimation des distributions de fréquence et de sévérité des pertes opérationnelles sont prometteuses. La méthode présentée dans cette partie est basée sur la spécification des distributions a priori des loi de fréquence et sévérité à partir des avis d'experts métier. Puis à l'aide d'artifices de calculs numériques assez simples, les paramètres des distributions a priori sont pondérés avec les 5 ans d'observations réelles de la banque pour estimer les paramètres des distributions a posteriori. La méthode est simple à mettre en place du fait de l'utilisation des lois conjuguées qui donnent des formules exactes pour les nouveaux paramètres facilitant ainsi leur calcul.

Compte-tenu du caractère rare des pertes représentées par les scénarios, il est difficile de fournir un paramétrage de distribution fiable et ainsi, **les experts ont un manque de visibilité du risque et ne peuvent donc pas fournir un nombre d'informations suffisant pour estimer convenablement les paramètres des distributions a priori**. En effet, les informations fournies ici ne donnent que trois points à partir desquelles on ajuste une distribution qui apporte les meilleures valeurs espérées des lois a priori. Il paraît donc compliqué d'obtenir des distributions adéquates car les experts ne peuvent pas apporter directement les valeurs des paramètres a priori. Et il est difficile de prévoir avec certitude la survenance d'une perte jamais apparue avant. Compte tenu de ces difficultés liées au manque de données, une méthode alternative consiste à utiliser l'approche par crédibilité qui permet de mieux modéliser les paramètres de fréquence et sévérité liés aux risques à basse fréquence et haute sévérité.

Partie 4

Application de la théorie de la crédibilité

Les données permettant de quantifier la fréquence et la sévérité des distributions de pertes sont très limitées si bien que l'estimation a priori par les méthodes bayésiennes pose beaucoup de problèmes. La théorie de la crédibilité apporte une approche alternative car elle exige moins d'informations. On s'appuie sur la théorie développée par Bühlmann et Straub.

4.1 Principes théoriques

4.1.1 Notion de crédibilité

La théorie de la crédibilité repose sur les principes de l'inférence bayésienne. Cette théorie est très utilisée pour calculer des primes sur les contrats d'assurance et a été adaptée dans le calcul des pertes opérationnelles. Intéressons-nous à une banque et à son risque opérationnel. Supposons qu'elle dispose d'un portefeuille rassemblant ses pertes internes au sein de J cellules de risque et ce sur un historique de 5 ans. Ces pertes sont associées à des variables aléatoires $X_{j,t}$ pour $t = 1, \dots, T$ et $j = 1, \dots, J$ où $X_{j,t}$ représente le montant annuel de pertes associées à une cellule de risque j pour la $t^{\text{ème}}$ année. On fait l'hypothèse que ces montants sont indépendants et identiquement distribués.

L'idée de la théorie de la crédibilité est de prévoir les pertes opérationnelles futures en intégrant l'information disponible qui est présente dans les historiques de pertes internes. A cette information on accorde un certain poids qu'on appelle le **poids de crédibilité**.

Considérons à présent la cellule de risque j à laquelle correspondent les montants annuels de pertes $X_{j,1}, \dots, X_{j,T}$ et le paramètre de risque Θ_j . Ce paramètre de risque n'est pas observable mais pourra être partiellement recomposé à partir des pertes observées. Donnons-nous des lois entièrement spécifiées pour le paramètre de risque et pour les variables de pertes conditionnellement au paramètre de risque :

- la loi marginale de Θ_j admet une densité $\mu(\Theta_j)$
- les variables $X_{j,1}, \dots, X_{j,T}$ sont conditionnellement à Θ_j indépendantes entre elles et distribuées uniquement avec pour densité $f(x_{j,t} | \Theta_j)$

On a vu que l'approche bayésienne consistait en l'application de la formule de Bayes pour la détermination d'une loi a posteriori. Voyons à présent les deux modèles proposés d'une part par Bühlmann et d'autre part par Bühlmann et Straub et repris dans H. Bühlmann et A. Gisler [6].

4.1.2 Le modèle de Bühlmann

On suppose les variables $X_{j,1}, \dots, X_{j,T}$ sont conditionnellement à Θ_j indépendantes entre elles et distribuées identiquement. La proposition suivante nous donne la meilleure approximation de $\mathbb{E}[X_{j,T+1} | \Theta_j]$ par une fonction affine.

Proposition 1 *La meilleure approximation au sens des moindres carrés de $\mathbb{E}[X_{j,T+1} | \Theta_j]$ par une fonction affine des observations $X_{j,1}, \dots, X_{j,T}$ est donnée par :*

$$p(X_{j,1}, \dots, X_{j,T}) = \left(1 - \frac{\text{cov}(X_{j,1}, X_{j,2})}{\mathbb{V}(\bar{X}_{jT})}\right) \mathbb{E}[X_{j,1}] + \frac{\text{cov}(X_{j,1}, X_{j,2})}{\mathbb{V}(\bar{X}_{jT})} \bar{X}_j$$

où $\bar{X}_j = \frac{1}{T} \sum_{t=1}^T X_{j,t}$. De plus, l'erreur d'approximation commise est donnée par :

$$e_T^2 = \mathbb{V}(\mathbb{E}[X_{j,T+1} | \Theta_j] - p(X_{j,1}, \dots, X_{j,T})) = \left(1 - \frac{\text{cov}(X_{j,1}, X_{j,2})}{\mathbb{V}(\bar{X}_{jT})}\right) \text{cov}(\mathbb{E}[X_{j,T+1} | \Theta_j], X_{j,T+1})$$

La démonstration est présentée dans C. Partrat et J-L. Besson [24].

Le modèle de crédibilité de Bühlmann utilise cette approximation linéaire car il propose de modéliser la perte interne à la date $T + 1$ de la forme :

$$p(X_{j1}, \dots, X_{jT}) = c_0 + c_1 X_{j1} + \dots + c_T X_{jT}$$

où les c_t sont choisis de manière à minimiser l'écart quadratique moyen

$$\mathbb{E} \left[(\mathbb{E}[X_{j,T+1} | \Theta_j] - c_0 - c_1 X_{j1} - \dots - c_T X_{jT})^2 \right]$$

En résolvant ce système (voir C. Partrat et J-L. Besson [24]), on remarque que le terme c_t ne dépend pas de t , ce qui signifie que l'ancienneté de la perte n'est pas prise en compte : dans le modèle de Bühlmann, une perte interne aura donc le même poids quelle que soit sa date d'occurrence. Dans le modèle de Bühlmann, on a donc une pondération constante entre la cellule de risque j et l'espérance a priori des pertes historiques :

$$\mathbb{E}[X_{j,T+1} | \theta_j] = \alpha \bar{X}_j + (1 - \alpha) \mathbb{E}[X_{jt}]$$

Pour corriger cette hypothèse peu réaliste, le modèle de Bühlmann-Straub permet d'intégrer des pondérations aux pertes.

4.1.3 Le modèle de Bühlmann-Straub

Le modèle de Bühlmann-Straub est une extension du modèle de Bühlmann qui intègre des pondérations sur les observations. En effet, il apparaît judicieux de ne pas accorder la même importance aux informations les plus récentes qu'aux informations les plus anciennes. Ce modèle repose sur les hypothèses suivantes :

1. Pour $j = 1, \dots, J$, les vecteurs $(\Theta_j, X_{j,1}, \dots, X_{j,T})$ sont indépendants
2. Les variables $\Theta_1, \dots, \Theta_J$ sont indépendantes et identiquement distribuées
3. Conditionnellement à Θ_j , les variables aléatoires $X_{j,1}, \dots, X_{j,T}$ sont indépendantes et identiquement distribuées
4. Les variables aléatoires $X_{j,1}, \dots, X_{j,T}$ sont de carré intégrable, d'espérance conditionnelle $\mathbb{E}[X_{j,t} | \Theta_j] = \mu(\Theta_j)$ et de variance conditionnelle $\mathbb{V}(X_{j,t} | \Theta_j) = \frac{\sigma^2(\Theta_j)}{\omega_{j,t}}$ où $\omega_{j,t}$ traduit l'importance de la perte interne de la cellule j au cours de la $t^{\text{ème}}$ année

On a alors une pondération suivant les cellules de risque :

$$\mathbb{E}[X_{j,T+1} | \Theta_j] = \alpha_j \bar{X}_j + (1 - \alpha_j) \mathbb{E}[X_{j,t}]$$

En résumé, il faut retenir que dans le modèle de Bühlmann-Straub, on fait une pondération entre une cellule de risque j et l'ensemble des cellules :

cellule de risque j	ensemble des cellules
$\mu(\Theta_j) = \mathbb{E}[X_{j,t} \Theta_j]$	$\mu_0 = \mathbb{E}[\mu(\Theta_j)]$
$\sigma^2(\Theta_j) = \omega_{j,t} \mathbb{V}[X_{j,t} \Theta_j]$	$\sigma^2 = \mathbb{E}[\sigma^2(\Theta_j)]$
	$\tau^2 = \mathbb{V}[\mu(\Theta_j)]$
$\mu(\Theta_j)$: espérance de la perte interne de la cellule j	μ_0 : espérance de la perte interne collective
$\sigma^2(\Theta_j)$: variance associée à la perte interne de la cellule j	σ^2 : espérance de la variance des pertes individuelles
	τ^2 : variance au sein des cellules de risque

TABLE 4.1 – Les paramètres à prendre en compte dans le modèle de Bühlmann-Straub

De ce modèle, on peut dégager quelques interprétations utiles pour la suite. Le coefficient de crédibilité α_j vaut :

$$\alpha_j = \frac{\omega_{j\bullet}}{\omega_{j\bullet} + \tau^2/\sigma^2}$$

On peut donc déjà en déduire que :

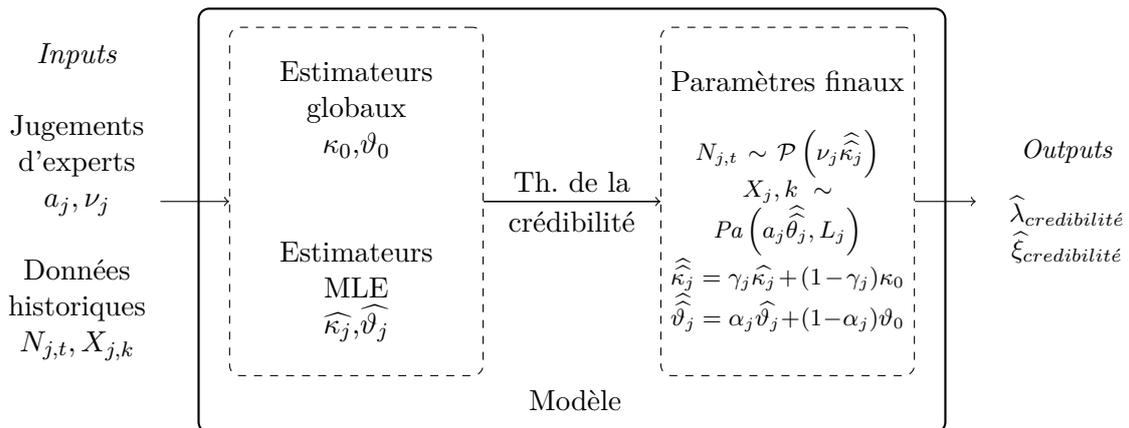
- plus le poids $\omega_{j\bullet}$ est grand, plus α_j est grand, c'est-à-dire que plus on accorde de poids aux données, plus on accordera de crédibilité aux pertes internes de la cellule de risque en question
- plus τ/σ est petit, plus α_j est grand.

4.2 Adaptation du modèle de Bühlmann-Straub au modèle LDA

On cherche à modéliser par une approche fréquence/sévérité les pertes du risque opérationnel liées à chaque cellule de risque j que l'on note $X_{j,k}$, avec $k \in \llbracket 1, K_j \rrbracket$ où K_j est le nombre de pertes qu'on a collectées dans chaque cellule de risque j . On note $N_{j,t}$ le nombre de pertes collectées lors de la $t^{\text{ème}}$ année dans chaque cellule j . On ne modélise que les pertes dépassant un certain seuil L .

Une première approche, décrite dans H. Bühlmann et P.V. Shevchenko [7], modélise respectivement les fréquences et les sévérités des pertes par des lois de Poisson et de Pareto, de paramètre $\lambda_j = \nu_j \kappa_j$ et $\xi_j = a_j \vartheta_j$. Dans cette approche, (λ_j, ϑ_j) sont les paramètres finaux des lois que l'on cherche. (a_j, ν_j) sont des facteurs choisis par les experts qui sont censés représenter les différences relatives de sévérité et de fréquence au sein des cellules de risque.

Dans les deux cas de la fréquence et de la sévérité, les estimateurs sont calculés comme une moyenne pondérée entre deux estimateurs : le premier est l'estimateur du maximum de vraisemblance calculé uniquement à partir des données historiques, le second est un estimateur permettant de prendre en compte l'ensemble des données à savoir dans notre cas les données historiques et les jugements d'experts. On pourrait souhaiter également prendre en compte des données externes à l'entreprise (voir cascade bayésienne présentée dans la section 5.1.).



4.2.1 Modélisation de la fréquence des pertes

On considère une collection de J cellules de risque. Soit $N_{j,t}$ le nombre annuel d'événements dont la perte annuelle excède un certain seuil L dans la $j^{\text{ème}}$ cellule de risque ($j = 1, \dots, J$) dans la $t^{\text{ème}}$ année.

Etant donné κ_j , on suppose que les $N_{j,t}$ sont indépendants et distribués par la loi de Poisson de paramètre $\lambda_j = \nu_j \kappa_j$, c'est-à-dire

$$\mathbb{P}[N_{j,t} = n \mid \lambda_j] = \frac{(\lambda_j)^n}{n!} \exp(-\lambda_j)$$

où ν_j sont déterminés à l'aide du jugement des experts et expriment les différences de fréquence entre les cellules de risque.

Le modèle

En supposant que les $N_{j,t}$, $t = 1, \dots, T$ dans la $j^{\text{ème}}$ cellule de risque sont indépendants conditionnellement à κ_j , on peut alors démontrer que l'estimateur par maximum de vraisemblance de κ_j s'écrit :

$$\hat{\kappa}_j = \frac{1}{\tilde{\nu}_j} \sum_{t=1}^T N_{j,t} \quad \text{où} \quad \tilde{\nu}_j = \nu_j T$$

Avec :

$$\mathbb{E}[\hat{\kappa}_j | \kappa_j] = \kappa_j \quad \text{et} \quad \mathbb{V}(\hat{\kappa}_j | \kappa_j) = \frac{\kappa_j}{\tilde{\nu}_j}$$

Traditionnellement, seules quelques pertes sont observées au sein de certaines catégories de risque si bien que l'on ne peut pas compter sur l'estimateur du maximum de vraisemblance standard $\hat{\lambda}_j = \nu_j \hat{\kappa}_j$. L'idée est alors d'utiliser l'ensemble des pertes collectées de la banque ainsi que les jugements des experts pour mieux estimer les paramètres de la loi de Poisson et construire un estimateur qui prend en compte toutes ces informations.

On suppose que les κ_j sont indépendants et identiquement distribués avec $\mathbb{E}[\kappa_j] = \kappa_0$ et $\mathbb{V}(\kappa_j) = \omega_0^2$. On observe que les fréquences standardisées $F_{j,t} = N_{j,t}/\nu_j$ satisfont :

$$\mathbb{E}[F_{j,t} | \kappa_j] = \kappa_j \quad \text{et} \quad \mathbb{V}(F_{j,t} | \kappa_j) = \frac{\kappa_j}{\tilde{\nu}_j}$$

On constate que $F_{j,t}$ satisfait les hypothèses du modèle de Bühlmann-Straub ce qui permet d'écrire l'estimateur au sens de la crédibilité suivant pour κ_j :

$$\hat{\kappa}_j = \gamma_j \hat{\kappa}_j + (1 - \gamma_j) \kappa_0 \quad \text{où} \quad \gamma_j = \frac{\tilde{\nu}_j}{\tilde{\nu}_j + \frac{\kappa_0}{\omega_0^2}}$$

Les paramètres structuraux λ_0 et ω_0 peuvent être estimés aisément en résolvant un système non linéaire de deux équations (en utilisant par exemple la procédure itérative présentée dans H. Bühlmann et A. Gisler [6]). Ce système correspond aux deux équations formulées dans les hypothèses et s'écrit :

$$\hat{\kappa}_0 = \frac{1}{\tilde{\gamma}} \sum_{j=1}^J \gamma_j \hat{\kappa}_j \quad \text{et} \quad (\hat{\omega}_0)^2 = \max \left[c \times \left\{ R - \frac{J \hat{\kappa}_0}{\nu_0} \right\}, 0 \right]$$

où

$$\nu_0 = \sum_{j=1}^J \tilde{\nu}_j, \quad R = \frac{J}{J-1} \sum_{j=1}^J \frac{\tilde{\nu}_j}{\nu_0} (\hat{\kappa}_j - \bar{F})^2, \quad \tilde{\gamma} = \sum_{j=1}^J \gamma_j,$$

$$\bar{F} = \sum_{j=1}^J \hat{\kappa}_j, \quad c = \frac{J}{J-1} \left\{ \frac{\tilde{\nu}_j}{\nu_0} \sum_{j=1}^J \left(1 - \frac{\tilde{\nu}_j}{\nu_0} \right) \right\}^{-1} \quad \text{et} \quad \gamma_j = \frac{\tilde{\nu}_j}{\tilde{\nu}_j + \frac{\kappa_0}{\omega_0^2}}.$$

Le meilleur estimateur de crédibilité pour le paramètre de fréquence de la $j^{\text{ème}}$ cellule est alors $\hat{\lambda}_j = \nu_j \hat{\kappa}_j$. On suppose que les constantes ν_j sont connues a priori et fixées par les experts. Par exemple, l'expert peut estimer le nombre n_j de pertes attendues excédant le seuil L pour la

$j^{\text{ème}}$ cellule en donnant la valeur \hat{n}_j . En utilisant alors les relations $\nu_j \kappa_j = n_j$ et $\mathbb{E}[\kappa_j] = \kappa_0$, il est possible d'estimer ν_j en prenant la valeur \hat{n}_j/κ_0 . Seules les différences de fréquence jouent un rôle ici, on peut donc poser $\lambda_0 = 1$ ce qui permet d'estimer facilement les ν_j sans perdre de généralité.

4.2.2 Modélisation de la sévérité des pertes

Dans un premier temps, on suppose que les pertes au-delà du seuil L dans la $j^{\text{ème}}$ cellule de risque ($j = 1, \dots, J$) sont indépendantes et identiquement distribuées suivant la loi de Pareto (on suit immédiatement la démarche de P.V. Shevchenko) de paramètres $\xi_j = a_j \vartheta_j$ et L :

$$f(x | \xi_j) = \frac{\xi_j}{L} \left(\frac{x}{L}\right)^{-\xi_j-1} \text{ et } F(x | \xi_j) = 1 - \left(\frac{x}{L}\right)^{-\xi_j}$$

pour $x \geq L$ et $\xi_j > 0$. Les a_j expriment la différence de sévérité entre les différentes cellules de risque et sont fournis par les experts, ce qui sera expliqué par la suite dans la présentation du modèle.

Le modèle

Comme les pertes $X_{j,k}$, $k = 1, \dots, K_j$ dans la $j^{\text{ème}}$ cellule de risque de la banque sont supposées être indépendantes et identiquement distribuées conditionnellement à ϑ_j , l'estimateur du maximum de vraisemblance de ϑ_j peut s'écrire :

$$\widehat{\Psi}_j = \left[\frac{a_j}{K_j} \sum_{k=1}^{K_j} \ln \left(\frac{X_{j,k}}{L} \right) \right]^{-1}$$

On montre alors qu'un estimateur sans biais de ϑ_j est :

$$\widehat{\vartheta}_j = \frac{\tilde{K}_j - 1}{\tilde{K}_j} \widehat{\Psi}_j$$

Avec :

$$\mathbb{E}[\widehat{\vartheta}_j | \vartheta_j] = \vartheta_j \text{ et } \mathbb{V}(\widehat{\vartheta}_j | \vartheta_j) = \frac{\vartheta_j^2}{K_j - 2}$$

Traditionnellement, seules quelques pertes sont observées au sein de certaines catégories de risque si bien que l'on ne peut pas compter sur l'estimateur du maximum de vraisemblance standard $\widehat{\xi}_j = a_j \widehat{\vartheta}_j$. L'idée est alors comme dans le cas de la fréquence d'utiliser l'ensemble des pertes collectées de la banque ainsi que les jugements des experts pour mieux estimer les paramètres de la loi de Pareto et construire un estimateur qui prend en compte toutes ces informations.

L'estimateur du paramètre de queue $\widehat{\xi}_j = a_j \widehat{\vartheta}_j$ peut être amélioré en utilisant toutes les données collectées par la banque. On suppose que les ϑ_j pour $j = 1, \dots, J$ sont des variables indépendantes et identiquement distribuées avec $\mathbb{E}[\vartheta_j] = \vartheta_0$ et $\mathbb{V}(\vartheta_j) = (\tau_0)^2$, où ϑ_0 correspond au profil de risque de la banque entière. On constate que l'estimateur non biaisé $\widehat{\xi}_j$ satisfait les hypothèses du modèle de Bühlmann-Straub ce qui permet d'écrire l'estimateur de crédibilité suivant :

$$\widehat{\vartheta}_j = \alpha_j \widehat{\vartheta}_j + (1 - \alpha_j) \vartheta_0 \text{ où } \alpha_j = \frac{K_j - 2}{K_j - 1 + \left(\frac{\vartheta_0}{\tau_0}\right)^2}$$

De même que pour la modélisation de la fréquence, les paramètres ϑ_0 et τ_0 peuvent être estimés en résolvant le système non linéaire de deux équations suivant :

$$(\tau_0)^2 = \frac{1}{J-1} \sum_{j=1}^J \alpha_j (\widehat{\vartheta}_j - \widehat{\vartheta}_0)$$

$$\widehat{\vartheta}_0 = \frac{1}{W} \sum_{j=1}^J \alpha_j \widehat{\vartheta}_j, \quad W = \sum_{j=1}^J \alpha_j \text{ et } \alpha_j = \frac{K_j - 2}{K_j - 1 + \left(\frac{\widehat{\vartheta}_0}{\tau_0}\right)^2}$$

Le meilleur estimateur de crédibilité pour le paramètre de queue de la $j^{\text{ème}}$ cellule est alors $\widehat{\xi}_j = a_j \widehat{\vartheta}_j$. On suppose que les constantes a_j sont connues a priori et fixées par les experts. Par exemple, l'expert spécifie la probabilité q_j que la perte de la $j^{\text{ème}}$ cellule de risque excède un certain niveau x_j et utilise les relations pour estimer les a_j :

$$a_j \vartheta_j = -\frac{\ln q_j}{\ln \left(\frac{x_j}{L_j}\right)} \text{ et } \mathbb{E}[\vartheta_j] = \vartheta_0$$

On note que ces constantes sont définies à un facteur près. Seules les différences jouent donc un rôle, ce qui permet sans perdre de généralité de poser pour l'estimation des a_j $\vartheta_0 = 1$.

4.2.3 Application numérique aux données

Dans le cadre de ce mémoire, on fait l'hypothèse que les trois cellules de risque choisies qui correspondent respectivement aux entités AMUNDI, CACF et CACIB pour la catégorie de risque EL6, constituent ici une banque fictive comme cela est expliqué dans le schéma ci-dessous :

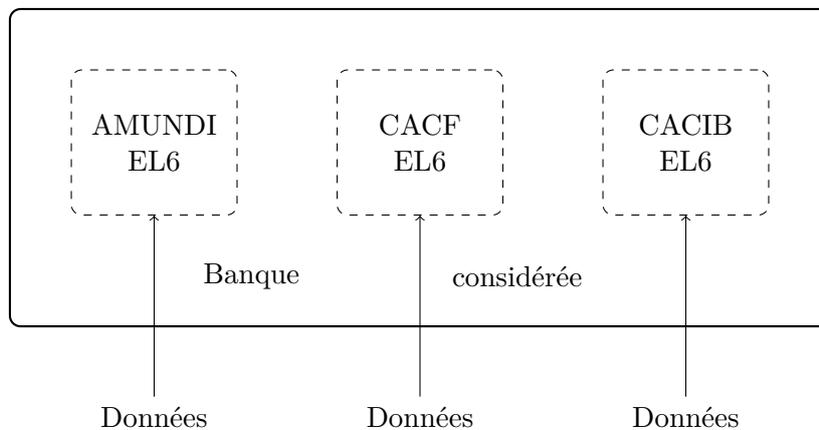


FIGURE 4.1 – Représentation schématique des données utilisées pour la crédibilité

Les cellules ont été choisies de telle manière qu'elles se rapportent à un risque similaire afin que la méthode par crédibilité soit applicable.

Détermination des coefficients ν_j

On doit avoir en entrée des paramètres ν_j qui caractérisent les différences de fréquence au sein des cellules de risque. Avec les données dont on dispose, la meilleure manière de représenter la différence entre deux cellules j et j' est de faire le rapport entre les deux paramètres λ_j et $\lambda_{j'}$ de la loi de Poisson. Ce rapport correspond à un rapport d'espérance du nombre de pertes attendues. L'idée est donc de fixer une cellule de risque j_0 tel que $\nu_{j_0} = 1$ puis pour toutes les autres cellules calculer :

$$\nu_j = \frac{\lambda_j}{\lambda_{j_0}}$$

Pour réaliser les premières applications numériques, on choisit arbitrairement un seuil L de telle sorte que des pertes dépassant ce seuil aient été collectées pour chacune des cellules de risque. On choisit une valeur du seuil égale à 20000 €, ce qui correspond à garder environ 8% des pertes de chaque cellule de risque. On verra dans la partie 5 qu'il existe des méthodes permettant d'aider à déterminer un seuil pertinent et objectif. On obtient les valeurs suivantes des ν_j et des estimateurs du maximum de vraisemblance des κ_j en prenant le cas 1 d'AMUNDI comme cellule de référence :

Cas	ν_j	$\hat{\kappa}_j$
1	1	2.24
2	11	0.2
3	10	2.8

Résultats numériques pour les paramètres de fréquence

Une fois les ν_j estimés et les estimateurs du maximum de vraisemblance des κ_j calculés, on peut résoudre le système non linéaire à deux inconnues afin de déterminer des valeurs de κ_0 et ω_0 et on trouve :

$$\hat{\kappa}_0 = 1.7 \text{ et } \hat{\omega}_0 = 2.3$$

On obtient alors les valeurs suivantes des poids de crédibilité :

Cas	γ_j
1	94,1%
2	99.4%
3	99.4%

On peut alors calculer les $\hat{\kappa}_j$ et par conséquent les meilleurs estimateurs $\hat{\lambda}_{\text{crédibilité}} = \nu \hat{\kappa}$ au sens de la crédibilité :

Cas	$\widehat{\lambda}_{\text{crédibilité}}$
1	2.74
2	2.30
3	22.37

On verra dans la section suivante que les valeurs de ce paramètre peuvent être utilisées pour modéliser les pertes à basse fréquence et haute sévérité.

Détermination des coefficients a_j

Pour rappel, les experts du Crédit Agricole fournissent pour chaque scénario trois montants de pertes potentielles, on choisit de déterminer les a_j en utilisant la méthode des moindres carrés :

$$\forall i \in \{1, 2, 3\}, a_j \vartheta_j = -\frac{\ln q_j^{(i)}}{\ln \frac{x_j^{(i)}}{L}}$$

On cherche donc $x = a_j \vartheta_j$ qui minimise l'expression :

$$\sum_{i=1}^3 \left(x + \frac{\ln q_j^{(i)}}{\ln \frac{x_j^{(i)}}{L}} \right)^2$$

Soit :

$$x = -\frac{1}{3} \sum_{i=1}^3 \frac{\ln q_j^{(i)}}{\ln \frac{x_j^{(i)}}{L}}$$

En combinant $\vartheta_0 = 1 = \mathbb{E}[\vartheta_j]$, on a finalement :

$$a_j = -\frac{1}{3} \sum_{i=1}^3 \frac{\ln q_j^{(i)}}{\ln \frac{x_j^{(i)}}{L}}$$

Certains scénarios ne sont définis que par un seul voire deux couples (x, d) . On adapte facilement la formule précédente au cas où on est sur ce type de scénario. L'application numérique donne les valeurs suivantes des a_j et des estimateurs du maximum de vraisemblance des ϑ_j :

Cas	ν_j	$\widehat{\vartheta}_j$
1	0.71	1.38
2	0.69	1.19
3	0.41	2.23

Résultats numériques pour les paramètres de sévérité

Une fois les a_j estimés et les estimateurs du maximum de vraisemblance des ϑ_j calculés, on peut résoudre le système non linéaire à deux inconnues afin de déterminer les valeurs de ϑ_0 et τ_0 . La résolution donne une solution négative pour τ_0^2 . Comme cela est expliqué dans H. Bühlmann et A. Gisler [6], on décide de poser $\alpha_j = 0$ et $\hat{\theta}_0 = \frac{1}{W} \sum_{j=1}^J w_j \hat{\theta}_j$, avec $w_j = K_j - 2$ et $W = \sum_{j=1}^J w_j$. On obtient alors la valeur de ϑ_0 suivante :

$$\hat{\vartheta}_0 = 2.08$$

On obtient donc les valeurs suivantes des meilleurs estimateurs $\hat{\xi}_{\text{crédibilité}} = a\hat{\vartheta}$ au sens de la crédibilité :

Cas	$\hat{\xi}_{\text{crédibilité}}$
1	1.47
2	1.43
3	0.85

On verra dans la section suivante que les valeurs de ce paramètre peuvent être utilisées pour modéliser les pertes à basse fréquence et haute sévérité.

4.2.4 Calcul des charges en capital

Le modèle de crédibilité présenté ici propose de modéliser les pertes à basse fréquence et haute sévérité au-dessus d'un certain seuil L . Bien évidemment, la modélisation des pertes à haute fréquence et basse sévérité correspondant aux pertes collectées en dessous de ce seuil doit être ajoutée au modèle avant d'estimer la charge en capital pour le risque opérationnel. Généralement les pertes à basse fréquence et haute sévérité sont celles qui contribuent le plus à la charge en capital finale. On a choisi d'adapter la méthode présentée dans H. Bühlmann et P.V. Shevchenko [7] : on suppose ici que le modèle LDA final pour la perte annuelle Z_j dans la cellule de risque j s'écrit :

$$Z_j = Z_j^{BF} + Z_j^{HF} = \sum_{n=1}^{N_j^{BF}} X_{j,n}^{BF} + \sum_{k=1}^{N_j^{HF}} X_{j,k}^{HF}$$

Ici, N_j^{BF} et $X_{j,n}^{BF}$, $n = 1, \dots, N_j^{BF}$ sont respectivement le nombre annuel de pertes et les montants indépendants et identiquement distribués des pertes à basse fréquence et haute sévérité distribués respectivement selon les fonctions de répartition $P_j^{BF}(\cdot)$ et $F_j^{BF}(\cdot)$. N_j^{HF} et $X_{j,n}^{HF}$, $n = 1, \dots, N_j^{HF}$ sont respectivement le nombre annuel de pertes et les montants indépendants et identiquement distribués des pertes à basse fréquence et haute sévérité distribués respectivement selon les fonctions de répartition $P_j^{HF}(\cdot)$ et $F_j^{HF}(\cdot)$. Il est raisonnable de supposer que les pertes à basse fréquence et haute sévérité et celles à haute fréquence et basse sévérité sont indépendantes, i.e : $X_{j,n}^{BF}$ et $X_{j,n}^{HF}$ sont indépendants ainsi que N_j^{BF} et N_j^{HF} . La distribution de la perte annuelle pour chaque cellule de risque j peut ensuite être calculée en utilisant la procédure de Monte Carlo suivante :

Étape 1 : Pour chaque cellule de risque j , on simule les nombres de pertes annuelles N_j^{BF} et N_j^{HF} à partir des distributions $P_j^{BF}(\cdot)$ et $P_j^{HF}(\cdot)$.

Étape 2 : Avec les valeurs de N_j^{BF} et N_j^{HF} données par l'étape 1, on simule les montants des pertes $X_{j,n}^{BF}$, $n = 1, \dots, N_j^{BF}$ et $X_{j,k}^{HF}$, $n = 1, \dots, N_j^{HF}$ à partir des distributions $F_j^{BF}(\cdot)$ et $F_j^{HF}(\cdot)$.

Étape 3 : Le calcul de la perte annuelle s'effectue à l'aide de la formule suivante :

$$Z_j = \sum_{n=1}^{N_j^{BF}} X_{j,n}^{BF} + \sum_{k=1}^{N_j^{HF}} X_{j,k}^{HF}$$

Étape 4 : On répète les étapes 1 à 3 un grand nombre de fois pour obtenir un échantillon de pertes annuelles. Il est alors possible de calculer le quantile à 99.9% de manière usuelle à partir d'un échantillon simulé.

Une approche intuitive

Une première démarche assez intuitive, pour calculer les charges en capital pour chaque cellule de risque j avec la décomposition décrite précédemment, est de considérer que les variables aléatoires N_j^{BF} et $X_{j,n}^{BF}$, $n = 1, \dots, N_j^{BF}$ sont distribuées selon les lois issues des données internes et les variables N_j^{HF} et $X_{j,n}^{HF}$, $n = 1, \dots, N_j^{HF}$ distribuées selon les lois issues des scénarios. On obtient les résultats suivants :

Cas	CaR _{naïve}
1	26 523 000 €
2	5 226 200 €
3	73 913 000 €

Au vu de ces résultats, on constate que les valeurs sont supérieures aux deux valeurs de charges en capital CaR_{DI} et CaR_S. On s'intéresse maintenant au calcul des charges en capital faisant intervenir les meilleurs estimateurs au sens de la crédibilité calculés précédemment.

Approche par crédibilité

Dans cette approche, les lois de fréquence et sévérité utilisées pour modéliser les variables N_j^{BF} et $X_{j,n}^{BF}$, $n = 1, \dots, N_j^{BF}$ ne sont plus les lois issues des scénarios mais les lois dont les paramètres ont été estimés précédemment à l'aide de la méthode par crédibilité. La fréquence est donc modélisée par une loi de Poisson de paramètre $\hat{\lambda}_{\text{crédibilité}}$ et la sévérité par une loi de Pareto de paramètre $\hat{\xi}_{\text{crédibilité}}$ et L . On calcule les charges en capital que l'on note CaR_{CPa} et on obtient alors les résultats suivants :

Cas	CaR _{CPa}
1	5 015 500 €
2	5 394 100 €
3	2 548 500 000 €

Le constat que l'on peut faire en regardant ces résultats est que la charge en capital explose pour le cas 3. Ceci est dû au fait que l'estimation du paramètre $\widehat{\xi}_{\text{crédibilité}}$, pour ce cas, renvoie une valeur inférieure à 1 qui rend la loi de Pareto non intégrable :

$$\mathbb{E} \left[X_{3,n}^{BF} \right] = +\infty$$

Par conséquent, la valeur du quantile à 99.9% de la distribution de la perte annuelle est très largement surestimée. La loi de Pareto est une loi dangereuse qui a tendance à surévaluer les queues de distribution, on se doit donc de l'utiliser avec parcimonie et précaution. Dans la partie suivante, on cherche à adapter la méthode par crédibilité avec une loi log-normale pour modéliser la sévérité.

4.3 Adaptation de la méthode à une loi log-normale

4.3.1 Nouvelle modélisation

Dans cette partie, on cherche à adapter la méthode de crédibilité présentée précédemment en utilisant cette fois-ci également une loi log-normale pour les pertes à basse fréquence et haute sévérité.

On suppose cette fois-ci que les pertes au-delà du seuil L dans la $j^{\text{ème}}$ cellule de risque ($j = 1, \dots, J$) sont indépendantes et identiquement distribuées suivant la loi log-normale tronquée en dessous du seuil L et de paramètre $\mu_j = a_j \delta_j$ et $\sigma_j = b_j \eta_j$. La densité de cette loi s'écrit :

$$f(x \mid (\mu_j, \sigma_j)) = \frac{\frac{1}{x\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(\ln x - \mu_j)^2}{2\sigma_j^2}\right)}{1 - \Phi\left(\frac{\ln L - \mu_j}{\sigma_j}\right)}$$

et

$$F(x \mid (\mu_j, \sigma_j)) = \frac{\Phi\left(\frac{\ln x - \mu_j}{\sigma_j}\right)}{1 - \Phi\left(\frac{\ln L - \mu_j}{\sigma_j}\right)}$$

pour $x \geq L$. Les a_j et b_j expriment, pour la sévérité, les différences d'espérance et de variance entre les cellules de risque et sont fournis par les experts métier. Ceci sera expliqué par la suite dans la présentation de cette adaptation du modèle de crédibilité proposé par P.V. Shevchenko [29].

Étant donné que les pertes $X_{j,k}$, $k = 1, \dots, K_j$ dans la $j^{\text{ème}}$ cellule de risque de la banque sont supposées être indépendantes et identiquement distribuées conditionnellement à $\Theta_j = (\delta_j, \eta_j)$, la vraisemblance s'écrit :

$$L(x_1, \dots, x_{K_j} \mid \delta_j, \eta_j) = \prod_{k=1}^{K_j} \frac{\frac{1}{x_k \sqrt{2\pi} b_j \eta_j} \exp\left(-\frac{(\ln x_k - a_j \delta_j)^2}{2b_j^2 \eta_j^2}\right)}{1 - \Phi\left(\frac{\ln L - a_j \delta_j}{b_j \eta_j}\right)}$$

Et la log-vraisemblance s'écrit :

$$l(x_1, \dots, x_{K_j} \mid \delta_j, \eta_j) = - \sum_{k=1}^{K_j} \left(\frac{(\ln x_k - a_j \delta_j)^2}{2b_j^2 \eta_j^2} + \ln x_k \right) - K_j \ln \left(\sqrt{2\pi} b_j \eta_j \left(1 - \Phi\left(\frac{\ln L - a_j \delta_j}{b_j \eta_j}\right) \right) \right)$$

Les estimateurs du maximum de vraisemblance de δ_j et η_j sont alors les solutions du système suivant :

$$\begin{cases} \frac{\partial l}{\partial \delta_j} = \frac{a_j K_j}{b_j^2 \eta_j^2} (\overline{\ln x} - a_j \delta_j) - \frac{a_j K_j}{b_j \eta_j} g(L, \delta_j, \eta_j) = 0 \\ \frac{\partial l}{\partial \eta_j} = \frac{1}{b_j^2 \eta_j^3} \left(\sum_{k=1}^{K_j} (\ln x_k - a_j \delta_j)^2 \right) - \frac{K_j}{\eta_j} - \frac{K_j}{\eta_j^2} \left(\frac{\ln L - a_j \delta_j}{b_j} \right) g(L, \delta_j, \eta_j) = 0 \end{cases}$$

Avec

$$g(L, \delta_j, \eta_j) = \frac{\varphi \left(\frac{\ln L - a_j \delta_j}{b_j \eta_j} \right)}{1 - \Phi \left(\frac{\ln L - a_j \delta_j}{b_j \eta_j} \right)}$$

On ne peut pas résoudre ce système de manière explicite, néanmoins on trouve que les solutions obtenues numériquement sont identiques si l'on prend la loi log-normale tronquée et la loi log-normale non tronquée. Ceci est dû au fait que la fonction g est négligeable devant les autres termes de l'équation. Pour la suite des calculs nous faisons donc l'hypothèse que nous sommes en présence d'une loi log-normale non tronquée. Cela revient à résoudre le système suivant :

$$\begin{cases} \overline{\ln x} - a_j \delta_j = 0 \\ \frac{1}{b_j^2} \left(\sum_{k=1}^{K_j} (\ln x_k - a_j \delta_j)^2 \right) - \eta_j^2 = 0 \end{cases}$$

Les estimateurs du maximum de vraisemblance de δ_j et η_j ont alors les expressions suivantes :

$$\begin{cases} \hat{\delta}_j = \frac{1}{a_j} \overline{\ln x} \\ \hat{\eta}_j^2 = \frac{\sum_{k=1}^{K_j} (\ln x_k - a_j \hat{\delta}_j)^2}{b_j^2 n_j} \end{cases}$$

On voit assez facilement que l'estimateur du maximum de vraisemblance de δ_j est un estimateur sans biais ce qui n'est pas le cas de celui de η_j . Il est alors aisé de montrer qu'un estimateur sans biais de η_j est :

$$\hat{\eta}_j^2 = \frac{\sum_{k=1}^{K_j} (\ln x_k - a_j \hat{\delta}_j)^2}{b_j^2 (K_j - 1)}$$

On peut alors déduire de ces formules l'espérance et la variance de ces estimateurs ce qui nous donne :

$$\mathbb{E} [\hat{\delta}_j | \delta_j] = \delta_j \text{ et } \mathbb{V} (\hat{\delta}_j | \delta_j) = \frac{b_j^2 \eta_j^2}{a_j^2 n_j}$$

$$\mathbb{E} [\hat{\eta}_j^2 | \eta_j^2] = \eta_j^2 \text{ et } \mathbb{V} (\hat{\eta}_j^2 | \eta_j^2) = \frac{2b_j^2 \eta_j^4}{n_j}$$

De la même façon que pour le cas de loi de Pareto, seulement quelques pertes sont observées au-delà du seuil L choisi si bien que l'on ne peut pas compter sur les estimateurs du maximum de vraisemblance standards $\hat{\mu}_j = a \hat{\delta}_j$ et $\hat{\sigma}_j^2 = b^2 \hat{\eta}_j^2$. L'idée reste la même que précédemment et consiste à utiliser l'ensemble des pertes collectées de la banque ainsi que les jugements des experts pour mieux estimer les paramètres de la loi log-normale pour les pertes basse fréquence et haute sévérité et ainsi construire des estimateurs qui prendront en compte toutes ces informations.

Les deux estimateurs précédents peuvent être améliorés en utilisant la méthode par crédibilité. On suppose que les δ_j et les η_j pour $j = 1, \dots, J$ sont des variables indépendantes et identiquement distribuées avec :

$$\mathbb{E}[\delta_j] = \delta_0, \quad \mathbb{V}(\delta_j) = \omega_0^2, \quad \mathbb{E}[\eta_j^2] = \eta_0, \quad \mathbb{V}(\eta_j^2) = \rho_0^2$$

où δ_0 et η_0 correspondent aux profils de risque de la banque entière. On constate que les estimateurs non biaisés $\widehat{\delta}_j$ et $\widehat{\eta}_j^2$ satisfont les hypothèses du modèle de Bühlmann-Straub avec les paramètres structuraux suivants :

Pour le paramètre δ_j	Pour le paramètre η_j^2
$\delta_0 = \mathbb{E}[\delta_j]$	$\eta_0^2 = \mathbb{E}[\eta_j^2]$
$\sigma^2 = \mathbb{E}[\sigma^2(\delta_j)] = \mathbb{E}[\eta_j^2] = \eta_0^2$	$\sigma^2 = \mathbb{E}[\sigma^2(\eta_j^2)] = \mathbb{E}[(\eta_j^2)^2] = \rho_0^2 + (\eta_0^2)^2$
$\tau^2 = \mathbb{V}(\delta_j) = \omega_0^2$	$\tau^2 = \mathbb{V}(\eta_j^2) = \rho_0^2$

Avec ces expressions des paramètres structuraux, cela permet d'écrire les estimateurs de crédibilité suivants :

$$\widehat{\delta}_j = \alpha_j \widehat{\delta}_j + (1 - \alpha_j) \delta_0 \quad \text{où} \quad \alpha_j = \frac{\frac{n_j a_j^2}{b_j^2}}{\frac{n_j a_j^2}{b_j^2} + \left(\frac{\eta_0}{\omega_0}\right)^2}$$

et

$$\widehat{\eta}_j^2 = \beta_j \widehat{\eta}_j^2 + (1 - \beta_j) \eta_0^2 \quad \text{où} \quad \beta_j = \frac{\frac{n_j}{(2b_j^2)}}{\frac{n_j}{(2b_j^2)} + 1 + \left(\frac{\eta_0^2}{\rho_0}\right)^2}$$

Les paramètres η_0^2 et ρ_0^2 peuvent être estimés en résolvant dans un premier temps le système non linéaire de deux équations suivant :

$$\widehat{\rho}_0^2 = \frac{1}{J-1} \sum_{j=1}^J \beta_j (\widehat{\eta}_j^2 - \widehat{\eta}_0^2)^2$$

$$\widehat{\eta}_0^2 = \frac{1}{W} \sum_{j=1}^J \beta_j \widehat{\eta}_j^2, \quad W = \sum_{j=1}^J \beta_j \quad \text{et} \quad \beta_j = \frac{\frac{n_j}{2b_j^2}}{\frac{n_j}{2b_j^2} + 1 + \left(\frac{\widehat{\eta}_0^2}{\rho_0}\right)^2}$$

Une fois ce système résolu, les paramètres $\widehat{\eta}_0^2$ et $\widehat{\rho}_0^2$ sont estimés ce qui donne accès au meilleur estimateur de crédibilité pour le paramètre σ de la loi log-normale de la $j^{\text{ème}}$ cellule de risque qui est alors $\widehat{\sigma}_j^2 = b_j^2 \widehat{\eta}_j^2$. Dans un second temps, pour estimer le paramètre d'espérance, on résout le second système linéaire suivant :

$$\widehat{\omega}_0^2 = \frac{1}{J-1} \sum_{j=1}^J \alpha_j (\widehat{\delta}_j - \widehat{\delta}_0)^2$$

$$\widehat{\delta}_0 = \frac{1}{X} \sum_{j=1}^J \alpha_j \widehat{\delta}_j, \quad X = \sum_{j=1}^J \alpha_j \quad \text{et} \quad \alpha_j = \frac{\frac{n_j a_j^2}{b_j^2}}{\frac{n_j a_j^2}{b_j^2} + \left(\frac{\widehat{\eta}_0}{\widehat{\omega}_0}\right)^2}$$

Pour résoudre ce système non linéaire à deux inconnues, il est nécessaire d'avoir résolu au préalable le premier système pour estimer la valeur $\widehat{\eta}_0^2$ que l'on voit apparaître dans l'expression du poids de crédibilité α_j . La résolution de ce second système permet d'obtenir l'estimation des paramètres $\widehat{\delta}_0$ et $\widehat{\omega}_0^2$. Le meilleur estimateur de crédibilité pour le paramètre d'espérance de la loi log-normale de la $j^{\text{ème}}$ cellule de risque est alors $\widehat{\mu}_j = a_j \widehat{\delta}_j$. On suppose que les constantes a_j et b_j sont connues a priori et fixées par les experts. Dans la section suivante, on présente le choix de modélisation qui a été fait.

4.3.2 Application numérique aux données

De la même manière que pour la première approche avec la loi de Pareto, on fait l'hypothèse que les trois cellules de risque choisies correspondent à une seule et même banque. On commence par déterminer les constantes à l'aide des informations des experts.

Détermination des coefficients a_j et b_j

On voit apparaître dans les expressions des paramètres de la loi log-normale des coefficients a_j et b_j qui sont connus a priori et fixés par les experts comme pour la méthode avec la loi de Pareto. Comme les logarithmes des sévérités suivent une loi normale, les a_j et b_j représentent ici respectivement les différences d'espérance et de variance des lois normales associées aux lois log-normales des cellules de risque. Avec les données dont on dispose, la meilleure manière de représenter la différence entre deux cellules de risque j et j' est de faire le rapport entre les paramètres de la loi log-normale fournis par les experts en choisissant une des cellules de risque j_0 comme la cellule de risque de référence puis pour toutes les autres on aurait :

$$a_j = \frac{\mu_j}{\mu_{j_0}} \text{ et } b_j = \frac{\sigma_j}{\sigma_{j_0}}$$

On obtient les valeurs suivantes des a_j , b_j ainsi que des estimateurs du maximum de vraisemblance des δ_j et η_j^2 en prenant le cas 1 d'AMUNDI comme cellule de référence :

Cas	a_j	$\widehat{\delta}_j$	b_j	$\widehat{\eta}_j^2$
1	1	9.1014	1	2.06
2	0.82	11.3010	1.88	0.47
3	0.98	9.3231	2.09	1.15

Résultats numériques pour les paramètres de la loi de sévérité

Une fois les a_j et b_j estimés ainsi que les estimateurs du maximum de vraisemblance des δ_j et η_j^2 calculés, on peut résoudre le premier système non linéaire à deux inconnues afin de déterminer les valeurs de η_0^2 et ρ_0 . On trouve :

$$\eta_0^2 = 1.34 \text{ et } \rho_0 = 0.66$$

Ces deux valeurs permettent d'avoir accès aux poids de crédibilité et de trouver les valeurs des $\widehat{\eta}_j$. Par conséquent, on peut calculer les meilleurs estimateurs $\widehat{\sigma}_{\text{crédibilité}} = b_j \widehat{\eta}_j^2$ au sens de la crédibilité :

Cas	β_j	$\hat{\sigma}_{\text{crédibilité}}$
1	85.2%	0.68
2	50.9%	0.96
3	92.9%	0.99

Il reste alors à estimer le paramètre $\hat{\mu}_{\text{crédibilité}}$. Puisque que l'on connaît déjà la valeur de $\hat{\eta}_0$, il est possible de résoudre le second système non linéaire à deux inconnues afin de déterminer les valeurs de δ_0^2 et ω_0 . On trouve :

$$\delta_0 = 9.87 \text{ et } \omega_0 = 2.35$$

Ces deux valeurs permettent d'avoir accès aux poids de crédibilité et de trouver les valeurs des $\hat{\delta}_j$. Par conséquent, on peut calculer les meilleurs estimateurs $\hat{\mu}_{\text{crédibilité}} = a\hat{\delta}$ au sens de la crédibilité :

Cas	α_j	$\hat{\mu}_{\text{crédibilité}}$
1	98.9%	10.89
2	91.3%	10.87
3	99.5%	10.99

On utilise dans la section suivante les valeurs de ces paramètres de crédibilité pour calculer les charges en capital pour les différentes cellules de risque.

4.3.3 Calcul des charges en capital

De façon identique à la première approche avec la loi de Pareto pour modéliser la sévérité des pertes liées au risque opérationnel, on suppose que le modèle LDA final pour la perte annuelle Z_j dans la cellule de risque j s'écrit avec les mêmes hypothèses de la façon suivante :

$$Z_j = Z_j^{BF} + Z_j^{HF} = \sum_{n=1}^{N_j^{BF}} X_{j,n}^{BF} + \sum_{k=1}^{N_j^{HF}} X_{j,k}^{HF}$$

Dans cette nouvelle méthode, les distributions $P_j^{BF}(\cdot)$, $P_j^{HF}(\cdot)$ et $F_j^{HF}(\cdot)$ restent inchangées. Seule la distribution $F_j^{BF}(\cdot)$ n'est plus une distribution de Pareto mais une loi log-normale tronquée de paramètre $\hat{\mu}_{\text{crédibilité}}$ et $\hat{\sigma}_{\text{crédibilité}}$. On calcule les nouvelles charges en capital que l'on note CaR_{CLN} et on obtient alors les résultats suivants :

Cas	CaR_{CLN}
1	2 401 600 €
2	3 002 000 €
3	27 540 000 €

TABLE 4.2 – Calcul des CaR_{CLN}

Les résultats présentés dans ce tableau sont, dans chacun des cas, compris entre les valeurs CaR_{DI} et CaR_S calculées initialement. Ces montants correspondent à une majoration de la charge en capital CaR_{DI} d'un pourcentage allant de 20 à 50% selon les cas. Le constat est que plus le nombre de pertes observées au-dessus du seuil L est faible, plus la majoration de la CaR_{DI} sera grande. Ceci répond à l'intuition de départ que l'on va accorder le plus de confiance au résultat lorsque l'échantillon de pertes est large.

4.3.4 Sensibilités aux paramètres d'entrée du modèle

On vient de déterminer les montants de charge en capital requis en s'appuyant sur certaines hypothèses de choix d'un seuil L en entrée du modèle. Afin de compléter cette méthode par crédibilité, il semble intéressant de déterminer les sensibilités des différents paramètres, calculés au sein du modèle, aux variations de ce seuil L . On réalise ces sensibilités pour le cas 2. On fera varier la valeur du seuil dans l'intervalle $[15\ 000, 25000]$ avec un pas de 100.

Sensibilités des poids de crédibilité

On rappelle ici que les pertes basse fréquence et haute sévérité sont modélisées par une loi log-normale tronquée de paramètres $\hat{\mu}_{\text{crédibilité}}$ et $\hat{\sigma}_{\text{crédibilité}}$ pour la sévérité et une loi de Poisson de paramètre $\hat{\lambda}_{\text{crédibilité}}$ pour la fréquence. Ces différents paramètres de loi sont obtenus par pondération entre un paramètre global pour la banque, représentée ici par les trois cellules choisies, et un paramètre estimé par maximum de vraisemblance à partir des données collectées au-dessus du seuil L . On cherche donc à déterminer la sensibilité des poids de crédibilité des différents paramètres aux variations du seuil L :

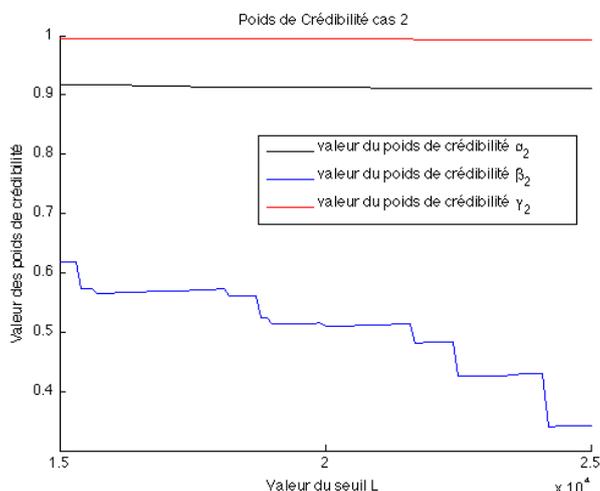


FIGURE 4.2 – Evolution des poids de crédibilité en fonction du seuil L

Ces résultats suggèrent que **les poids de crédibilité γ et α associés aux paramètres d'espérance des lois de fréquence et de sévérité des pertes ne sont pas sensibles aux variations du seuil effectuées**. A contrario, **le poids β lié à la variance de la loi de sévérité des pertes est lui assez sensible**. En effet, on constate la présence de sauts faisant diminuer fortement le poids β lorsque le seuil augmente. Augmenter le seuil revient à enlever au fur et à mesure le nombre de pertes collectées ce qui fait mécaniquement diminuer la confiance que l'on a en l'estimateur du maximum de vraisemblance. Pour le poids α lié à la variance de la loi de sévérité des pertes, on

constate que les estimateurs du maximum de vraisemblance restent proches de la valeur globale de la banque lorsque l'on fait varier le seuil, les poids ne varient donc quasiment pas. Enfin **pour le cas du poids γ lié à la fréquence, chaque cellule dispose du même nombre d'observations, ici égal à 5, quelle que soit la valeur du seuil L , on n'observe donc pas de sensibilité de ce poids aux variations de L .**

Sensibilités de la CaR_{CLN}

On cherche ici à déterminer la sensibilité de la CaR_{CLN} aux variations de valeur du seuil L . On obtient les résultats suivants :

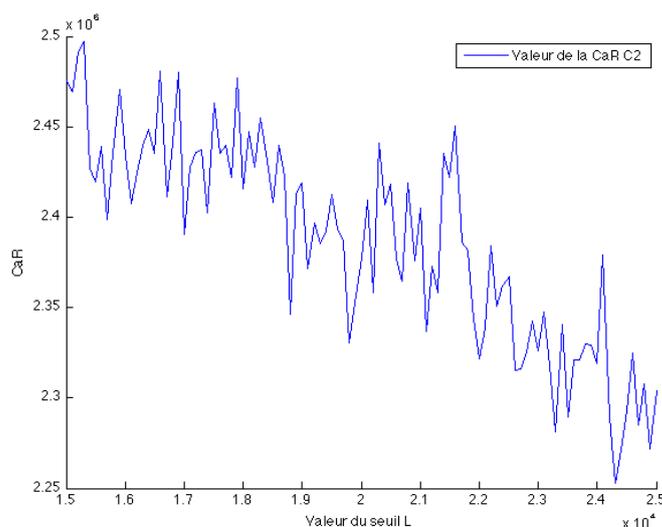


FIGURE 4.3 – Evolution de la CaR_{CLN} en fonction du seuil L

Au vu de ce résultat, on constate que **la CaR_{CLN} est sensible aux variations du seuil L et diminue lorsqu'il augmente**. Ceci peut s'expliquer car lorsque l'on regarde les variations des paramètres finaux de l'approche par crédibilité $\hat{\lambda}_{\text{crédibilité}}$, $\hat{\mu}_{\text{crédibilité}}$ et $\hat{\sigma}_{\text{crédibilité}}$, on remarque que les variations relatives des paramètres de la loi de sévérité sont faibles comparées à celles du paramètre de la loi de fréquence. Lorsque L augmente, $\hat{\lambda}_{\text{crédibilité}}$ décroît ce qui a pour conséquence de faire diminuer la CaR_{CLN}

4.4 Discussion sur la méthode

Cette méthode permet de calculer un unique montant de charge en capital en partant de deux distributions : l'une représentant les pertes haute fréquence et basse sévérité et l'autre les pertes basse fréquence et haute sévérité. Le principe de cette méthode est de considérer que la perte annuelle est décomposée sous la forme de deux sommes aléatoires indépendantes. La première somme correspond aux pertes haute fréquence et basse sévérité avec l'hypothèse qu'elles sont distribuées selon la distribution des données internes. La seconde somme représente les pertes basse fréquence et haute sévérité pour lesquelles on applique l'approche par crédibilité dans un souci de manque d'observation. Le fait de regrouper les données de plusieurs cellules de risque similaires permet d'avoir une plus grande confiance dans l'estimation des paramètres de loi des pertes extrêmes. Malgré une théorie assez difficile au premier abord, cette méthode a l'avantage de conduire à un montant de charge en capital transparent.

Bien que très intéressante, cette méthode possède quelques inconvénients. Tout d'abord, le choix des cellules de risque est une première étape très importante à ne pas négliger : il est primordial que les cellules choisies se rapportent à un profil de risque similaire. En effet, si des disparités sont présentes dans les données observées de chacune des cellules, certains paramètres risquent d'être surévalués et il en résultera des montants de charges en capital inexploitable. Enfin, le choix d'un seuil pour définir les pertes extrêmes est un paramètre d'entrée dans cette modélisation et ne peut être choisi arbitrairement. On présente dans la partie suivante des techniques pour déterminer un seuil de façon objective.

Partie 5

Approfondissement de la théorie bayésienne

Dans cette section sont présentées deux nouvelles méthodes qui se basent sur la théorie bayésienne détaillée dans la partie 3. La première approche appelée cascade bayésienne est une idée nouvelle qui permettrait de s'affranchir du problème d'hétérogénéité des données. La deuxième met en exergue une technique d'implémentation basée sur les chaînes de Markov et permet d'utiliser des lois non informatives.

5.1 Présentation de la cascade bayésienne

Récemment, une méthodologie innovante a été proposée pour combiner les données internes/externes et les scénarios des experts dans l'approche bayésienne du modèle LDA. L'intégration de ces trois types de sources est effectuée en deux étapes pour assurer que le modèle se conforme bien aux données internes et ainsi obtenir une meilleure caractérisation de la sévérité des pertes opérationnelles. Dans un premier temps, les scénarios des experts sont utilisés pour déduire la distribution de la loi a priori et de la même manière que précédemment on obtient une loi a posteriori initiale à partir des données externes. Dans un second temps, la loi a posteriori initiale est utilisée en tant que loi a priori et on obtient une seconde loi a posteriori à partir des données internes de l'entreprise.

Dans B.K. Hassani et A. Renaudin [18], on utilise trois types de données : les données internes, les données externes et les jugements des experts. Généralement, les données internes représentent la grosse partie de la distribution, les jugements des experts sont interprétés comme la queue de la distribution et les données externes sont considérées comme intermédiaires entre les deux types de données précédents. On le voit très bien sur la figure suivante :

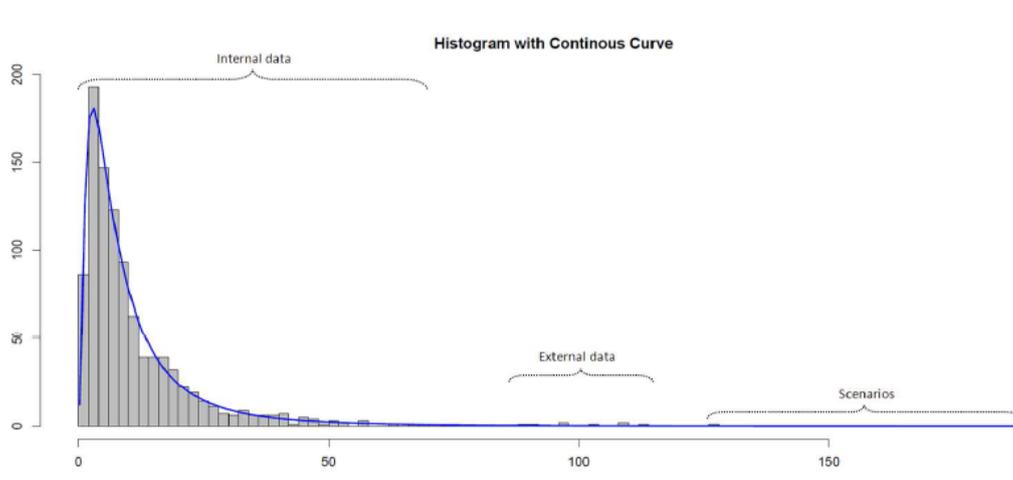


FIGURE 5.1 – Répartition des différents types de pertes (source :[18])

Etant données les difficultés identifiées pour combiner les deux sources de données (scénarios et données historiques) car elles représentaient de part et d'autre la tête et la queue de la distribution de pertes, l'idée est d'utiliser la cascade bayésienne à l'aide des pertes externes qui vont permettre de marquer une transition entre les données historiques et les scénarios. La structure de la méthode est de combiner dans un premier temps les scénarios avec ces pertes externes, puis de combiner dans un second temps la distribution obtenue avec les pertes internes qui seront censées être deux distributions plus homogènes.

Comme nous ne disposons pas de jeu de données externes, il va falloir les construire : dans la continuité de l'approche par crédibilité, on a l'idée de prendre comme données externes les plus grandes pertes historiques de chacune des trois cellules de risque choisies. Pour chaque cas, il est donc nécessaire de déterminer des seuils L_j , $j \in \{1, 2, 3\}$, au-dessus desquels les données internes pourront être considérées comme faisant partie des données externes. Le principal problème consiste donc à déterminer les valeurs de ces seuils L_j .

5.1.1 Détermination des seuils L_j

Contrairement à l'approche par crédibilité où l'on avait choisi un seuil de façon arbitraire, on s'intéresse ici à déterminer des seuils de façon objective en utilisant des éléments de la théorie des valeurs extrêmes. La théorie des valeurs extrêmes, selon l'approche retenue, propose différentes méthodes pour estimer un seuil à partir duquel une observation sera considérée comme valeur extrême. Une bonne détermination des seuils L_j est primordiale. En effet, un seuil trop faible conduit à une mauvaise adéquation entre la fonction de survie et la loi GPD et les estimateurs sont biaisés. Inversement, un seuil trop élevé entraîne une faible taille d'échantillon de dépassements et donc des estimateurs volatiles. Nous allons voir dans la suite diverses méthodes étudiées par le Crédit Agricole et efficaces dans l'estimation de ce seuil. Auparavant exposons quelques rappels sur la théorie des valeurs extrêmes¹. Dans la suite on notera L les seuils recherchés dans chacun des cas pour simplifier les écritures.

1. Des articles traitent de l'application de la théorie des valeurs extrêmes au risque opérationnel (voir V. Chavez-Demoulin et P. Embrechts [9] et M. Degen et al. [13]).

5.1.2 Les bases de la théorie des valeurs extrêmes (TVE)

La modélisation par la TVE repose sur deux théorèmes fondamentaux. Par la suite, notons X_1, \dots, X_n n variables aléatoires i.i.d. de montants de pertes de fonction de répartition F inconnue et x_1, \dots, x_n la réalisation de ces n variables aléatoires que l'on cherche à modéliser.

Théorème de Fisher-Tippet

On note $X_{(n)} = \max_{1 \leq i \leq n} X_i$, alors : $\mathbb{P}(X_{(n)} \leq x) = \mathbb{P}(X_1 \leq x, \dots, X_n \leq x) = (F(x))^n$. Le théorème fondamental de la Théorie des Valeurs Extrêmes dit :

Théorème 1 *Supposons qu'il existe des constantes normalisatrices $a_n \in \mathbb{R}$ et $b_n > 0$ et une loi non dégénérée H telles que*

$$b_n^{-1}(X_{(n)} - a_n) \xrightarrow{\mathcal{L}} H$$

Alors H est d'une même type qu'une des trois lois suivantes (données par leur fonction de répartition) avec $\gamma \geq 0$:

1. *loi de Fréchet, $H_{1,\gamma}(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\gamma}) & x > 0 \end{cases}$*
2. *loi de Weibull, $H_{2,\gamma}(x) = \begin{cases} \exp(-(-x)^\gamma) & x \leq 0 \\ 1 & x > 0 \end{cases}$*
3. *loi de Gumbel, $H_{3,0}(x) = \exp(-e^{-x})$*

On dit alors que F appartient au domaine d'attraction de la loi H (noté $DA(H)$).

Ces trois lois traduisent chacune un comportement différent du maximum de la loi. Si la probabilité que X prenne de grandes valeurs décroît exponentiellement alors H sera du type Gumbel. Si cette décroissance est plus lente (i.e. de type puissance) et que le support de la loi est borné alors H sera du type Weibull. Enfin si la décroissance est lente et que le support de la loi est infini alors H sera du type Fréchet. C'est souvent cette dernière loi qui semble la plus adaptée pour modéliser le risque opérationnel. En effet, dans notre contexte, les queues de distribution sont particulièrement épaisses et les pertes extrêmes peuvent prendre des valeurs "infinies". Une généralisation de ces 3 lois peut être faite à l'aide du modèle appelé GEV (Generalised Extreme Value) qui regroupe la famille des lois définies par les fonctions de répartition :

$$H_\xi(x) = \begin{cases} \exp(-(1 + \xi x)^{-1/\xi}) & \text{si } \xi \neq 0 \\ \exp(-e^{-x}) & \text{si } \xi = 0 \end{cases} \quad \text{où } 1 + \xi x > 0$$

On retrouve les 3 distributions précédemment introduites suivant le signe de ξ : si $\xi > 0$, c'est la distribution de Fréchet ; si $\xi < 0$, il s'agit de la distribution de Weibull, enfin $\xi = 0$ donne la distribution de Gumbel. ξ est appelé indice de valeur extrême. Plus cet indice est élevé en valeur absolue, plus la probabilité des extrêmes est élevée. On dit dans ce cas que la queue de distribution est épaisse.

Théorème de Balkema-de Haan-Pickands

Soit L_∞ la limite supérieure finie ou infinie du support de la loi F . Soit N_L le nombre d'observations qui excèdent un seuil L , $Z_i - L$ qui forment l'échantillon des excès Z_1, \dots, Z_{N_L} supposés i.i.d. On s'intéresse maintenant à la loi des observations au-dessus d'un certain seuil L qu'on appelle $F_L(x)$. La loi essentielle à la modélisation des excès pour des valeurs extrêmes du seuil est la loi de Pareto généralisée (en anglais GPD (General Pareto Distribution)). On appelle loi GPD une loi dont la fonction de répartition s'écrit :

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \frac{\xi x}{\beta})^{-1/\xi} & \text{si } \xi \neq 0 \\ 1 - \exp(-\frac{x}{\beta}) & \text{si } \xi = 0 \end{cases}$$

Le théorème qui définit la loi des excès au-dessus d'un seuil est le théorème de Balkema-de Haan-Pickands.

Théorème 2 Pour $\xi \in \mathbb{R}$, les affirmations suivantes sont équivalentes :

1. $F \in DA(H_\xi)$
2. il existe une fonction mesurable positive $\beta(\cdot)$ telle que, pour $1 + \xi x > 0$,

$$\lim_{L \rightarrow L_\infty} \left(\sup_{0 \leq x \leq L_\infty - L} |F_L(x) - G_{\xi,\beta(L)}(x)| \right) = 0$$

La GPD est donc la distribution limite de la distribution des excès lorsque le seuil tend vers L_∞ . Les théorèmes de Fisher-Tippett et de Balkema-de Haan-Pickands sont étroitement liés puisque l'indice de valeur extrême de la loi GPD est le même que celui de la loi GEV.

5.1.3 Méthodes d'estimation de L

Méthode 1 : utilisation de la fonction moyenne des excès

Afin de déterminer ce seuil, nous pouvons faire appel à la fonction moyenne des excès (plus connue sous le nom de *Mean excess function* dans la littérature) :

$$e(L) = \mathbb{E}(X - L \mid X \geq L)$$

L'estimateur empirique de cette *mean excess function* est la moyenne des excès au-dessus du seuil :

$$e_n(L) = \frac{\sum_{i=1}^n (X_i - L)^+}{\sum_{i=1}^n \mathbb{1}_{\{X_i \geq L\}}} = \frac{1}{N_u} \sum_{i=1}^n (X_i - L)^+$$

Or, la mean excess function de la loi de Pareto généralisée est exactement (démonstration faite en annexe B) :

$$e_{GPD}(L) = \frac{\beta}{1 - \xi} + L \left(\frac{\xi}{1 - \xi} \right)$$

C'est une fonction linéaire de L de pente $\frac{\xi}{1 - \xi}$. Un critère graphique de sélection du seuil consiste donc à sélectionner la valeur de L à partir de laquelle la fonction e_{GPD} semble avoir un comportement linéaire.

Méthode 2 : utilisation de l'estimateur de Hill

Un autre outil graphique utilisé pour appuyer le choix du seuil est le Hill-plot. Il consiste à tracer graphiquement l'estimateur de Hill $\hat{\xi}_k^{(H)}$ de l'indice de valeur extrême en fonction du nombre de valeurs au-dessus du seuil $(\hat{\xi}_k, k)$. L'estimateur de Hill de ξ est défini par :

$$\hat{\xi}_k^{(H)} = \frac{1}{k} \sum_{i=0}^{k-1} \ln X_{(n-i)} - \ln X_{n-k}$$

C'est l'estimateur le plus efficace et il possède de bonnes propriétés théoriques. Lorsque $k \rightarrow \infty$ et $\frac{k}{n} \xrightarrow{n \rightarrow \infty} 0$, alors $\hat{\xi}_k^{(H)} \xrightarrow{\mathcal{P}} \xi$ et il est asymptotiquement normal. k permet de représenter les k valeurs les plus grandes (ou les plus petites) de X ; l'idée est d'avoir k le plus grand possible lorsque n est très grand, mais sans prendre trop de valeurs de l'échantillon, ce qui conduit à imposer $\frac{k}{n} \rightarrow 0$. Incidemment, cela implique de se poser la question du choix optimal de k . Choisir un k trop élevé engendre le risque de prendre en compte des valeurs qui ne sont pas extrêmes, à l'inverse, un sous-échantillon trop petit ne permet pas aux estimateurs d'atteindre leur niveau de stabilité.

L'estimateur de Hill n'est utilisable que pour les distributions de Fréchet ce qui ne nous dérange pas car ce sont elles qui modélisent le mieux les pertes liées au risque opérationnel.

L'idée est donc de calculer l'estimateur de Hill de ξ et de chercher à partir de quel seuil cet indice est approximativement constant. En effet, si une loi de Pareto généralisée est une modélisation raisonnable pour des excès au-dessus d'un certain seuil L , alors les excès au-dessus d'un seuil plus élevé L' suivent aussi une loi GPD de même paramètre ξ . Si en augmentant le seuil, l'estimation de ξ reste constante, c'est qu'on peut raisonnablement modéliser les excès par une loi GPD.

5.1.4 Application aux données du Crédit Agricole

La méthode ME-plot (*Mean Excess plot*)

On a appliqué la première méthode aux trois cellules de risque : CACIB EL6, CACF EL6 et AMUNDI EL6. Voici les résultats lorsqu'on trace la *mean excess function* en fonction du seuil L :

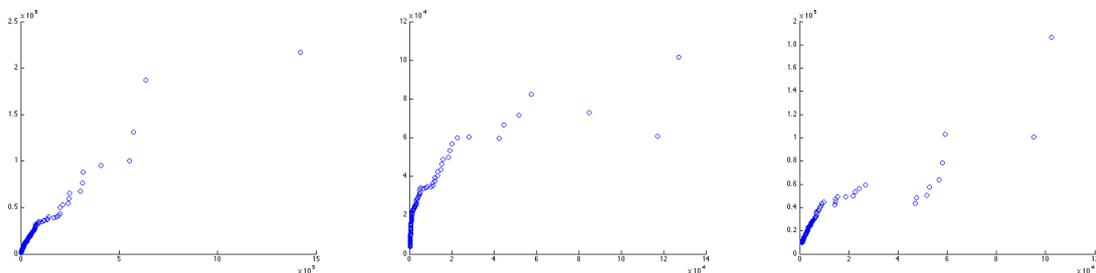


FIGURE 5.2 – ME-plot respectivement pour CACIB EL6, CACF EL6, AMUNDI EL6

Au vu des graphiques, cette méthode de sélection du seuil demeure imprécise, le ME-plot étant difficilement interprétable graphiquement. Elle fait appel à la subjectivité de l'analyste

et n'apparaît pas rigoureuse d'un point de vue purement statistique. Cela est dû au manque de données extrêmes dans les pertes historiques, ce qui rend ainsi l'estimation de la droite très hasardeuse.

Le Hill-plot

Pour l'estimateur de Hill, on retrouve les propriétés évoquées à la section précédente : on a une meilleure estimation de ξ lorsque n augmente (ce qui est le cas pour la branche CACIB).

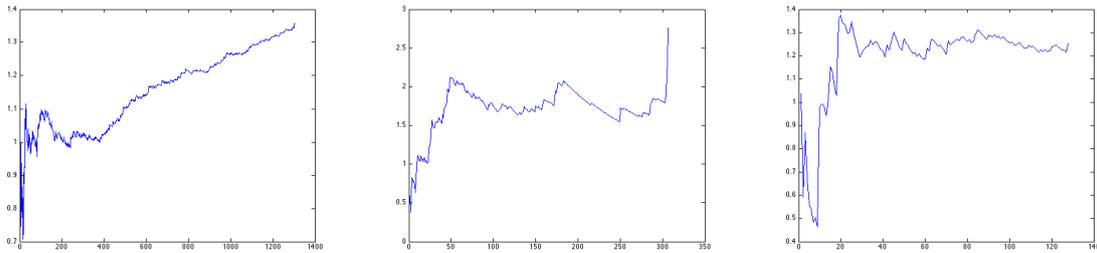


FIGURE 5.3 – Hill-plot respectivement pour CACIB EL6, CACF EL6, AMUNDI EL6

On voit que l'estimateur de Hill pour l'entité CACIB se stabilise entre la 200^{ème} et la 300^{ème} valeur de k , le seuil L correspondrait donc à la 200^{ème} valeur la plus grande du tableau de pertes internes. Dans le cas CACIB, on prendrait $L = 12000$.

Ces méthodes nous donnent d'emblée une bonne estimation de l'endroit où se situe le seuil et on retient ce seuil lorsqu'on programme la cascade bayésienne. On détermine par la suite une sensibilité au paramètre L .

5.1.5 Implémentation de la cascade bayésienne

Les méthodes de détermination des seuils L_j nous permettent de séparer les données historiques en deux types : le corps de la distribution qui sera à présent considéré comme les seules données internes de taille n_{DI} pour chaque cas ; les queues de distribution des 3 cas qui représenteront les données externes de taille n_{DE} . Faisons un bilan des données avec lesquelles nous allons travailler pour chaque entité : AMUNDI (1), CACF (2), CACIB (3) pour la catégorie de risque EL6.

Cas	L	n_{DI}	n_{DE}
1	9740	109	270
2	1080	258	
3	12200	1107	

TABLE 5.1 – Récapitulatif des données internes et externes utilisées pour la cascade bayésienne

Une fois les 3 sources de données à disposition, on peut implémenter la cascade bayésienne. La stratégie est basée sur la construction successive de deux distributions a posteriori. On rappelle que dans un premier temps les scénarios sont utilisés pour obtenir une distribution a priori des paramètres de sévérité. Puis les pertes externes sont utilisées dans la formule de Bayes

ce qui mène à l'obtention d'une première distribution a posteriori. Dans une seconde étape, cette distribution sert de loi a priori et donne accès à la distribution finale a posteriori en utilisant à nouveau la formule de Bayes avec les données internes. On peut résumer cette méthode de la façon suivante en reprenant certaines notations de l'approche bayésienne des lois conjuguées :

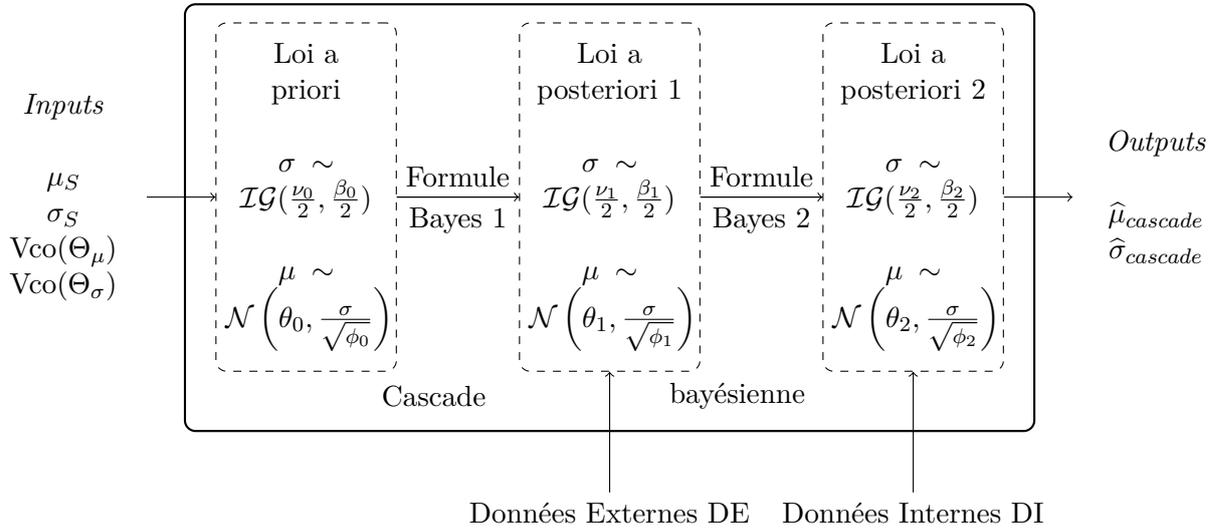


FIGURE 5.4 – Représentation schématique de la cascade bayésienne

5.1.6 Application numérique aux données

Concernant les trois cellules de risque choisies, on a vu dans la méthode bayésienne des lois conjuguées que les valeurs μ_S , σ_S , $Vco(\Theta_\mu)$ et $Vco(\Theta_{\sigma_2})$ permettent de déterminer les paramètres structuraux des lois a priori. On choisit une valeur de 50% pour les quantités $Vco(\Theta_\mu)$ et $Vco(\Theta_{\sigma_2})$. Les résultats a priori sont donc les mêmes que ceux que nous avons obtenus à la section 3.4.2. A l'aide des formules de passage décrites section 3.4.2, on calcule les paramètres structuraux de la première loi a posteriori à l'aide des données externes :

Cas	θ_1	ϕ_1	ν_1	β_1
1	10.17	261	273	369.3
2	10.16	261	273	373.5
3	10.17	261	273	372.3

TABLE 5.2 – Paramètres structuraux des lois a posteriori 1

Les paramètres des premières lois a posteriori étant calculés, on applique avec les données internes spécifiques à chacune des cellules les formules de passage pour calculer les paramètres structuraux finaux.

Cas	θ_2	ϕ_2	ν_2	β_2
1	8.95	370	382	662.1
2	7.12	519	531	1733.1
3	8.34	1377	1389	3914.7

TABLE 5.3 – Paramètres structuraux des lois a posteriori 2

Enfin en utilisant la méthode de l'erreur quadratique minimale, on a accès aux nouvelles valeurs des paramètres de la loi de sévérité des pertes opérationnelles $\mathcal{LN}(\mu, \sigma)$. Les résultats sont affichés dans le tableau ci-dessous :

Cas	$\mu = \hat{\mu}_{cascade}$	$\sigma = \hat{\sigma}_{cascade}$
1	8.95	1.32
2	7.12	1.81
3	8.34	1.68

TABLE 5.4 – Paramètres de la loi de sévérité obtenus avec la cascade bayésienne

Avec ces nouvelles valeurs des paramètres, on s'intéresse dans ce qui suit à calculer les charges en capital dans le cas des trois cellules de risque choisies. On utilise l'approche fréquence/sévérité décrite précédemment en modélisant la fréquence par une loi de Poisson $\mathcal{P}(\lambda)$ où l'on choisit la valeur de λ égale à celle obtenue grâce aux données internes λ_{DI} (on rappelle que cette valeur est vérifiée et peut être corrigée lorsque les experts considèrent que l'estimation qui est faite n'est pas représentative) et la sévérité par une loi log-normale $\mathcal{LN}(\mu, \sigma)$ avec les valeurs calculées au moyen de la cascade bayésienne. On obtient les résultats suivants :

Cas	$\text{CaR}_{cascade}$
1	2 943 800 €
2	3 140 900 €
3	26 321 000 €

TABLE 5.5 – Charges en capital obtenues avec la cascade bayésienne

Les résultats présentés dans ce tableau sont compris entre les valeurs CaR_{DI} et CaR_S calculées aux sections 2.5.2. et 2.5.3. Le constat est que ces montants sont très proches des valeurs CaR_{DI} . Ceci peut s'expliquer car les informations des experts sont noyées dès la première étape ; en effet, quelle que soit la cellule j choisie, les paramètres de la première loi a posteriori sont quasiment identiques. Les différences sont apportées lors de la deuxième étape avec les données internes. L'intérêt peut donc être considéré comme limité car les informations des experts ne sont pas complètement exploitées.

5.1.7 Sensibilités aux paramètres d'entrée du modèle

On vient de déterminer les montants de charge en capital en s'appuyant sur certaines hypothèses concernant les lois a priori ainsi que la détermination de seuil L_j , $j \in \{1, 2, 3\}$ permettant de créer un jeu de données externes. Afin de compléter ces premiers calculs, il semble intéressant de réaliser des sensibilités à la fois à ces seuils L_j mais aussi aux coefficients de variation de la même façon que dans la section 3.5.2.

Sensibilités aux $Vco(\Theta_\mu)$ et $Vco(\Theta_\sigma^2)$

On rappelle ici que les lois a priori des paramètres de sévérité Θ_μ et Θ_σ sont respectivement une loi normale $\mathcal{N}(\theta_0, \sigma\sqrt{\phi_0})$ et une loi inverse-gamma $\mathcal{IG}(\nu_0/2, \beta_0/2)$. Les paramètres structuraux de ces lois sont déterminés à partir des valeurs μ_S et σ_S ainsi que des coefficients de variation $Vco(\Theta_\mu)$ et $Vco(\Theta_\sigma^2)$ que l'on a choisi de fixer à 50%. On cherche donc à déterminer la sensibilité de $CaR_{cascade}$ aux variations respectives des Vco . On fait varier ces coefficients sur l'intervalle $[0.25, 0.75]$ par pas de 0.01 pour le cas 1 :

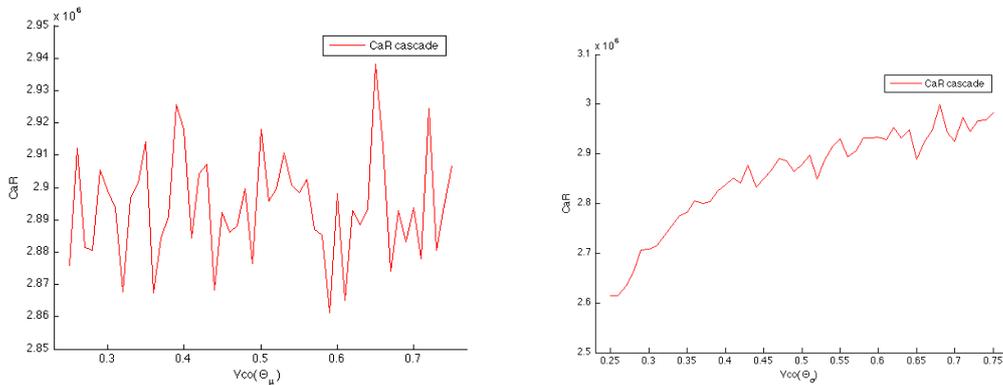
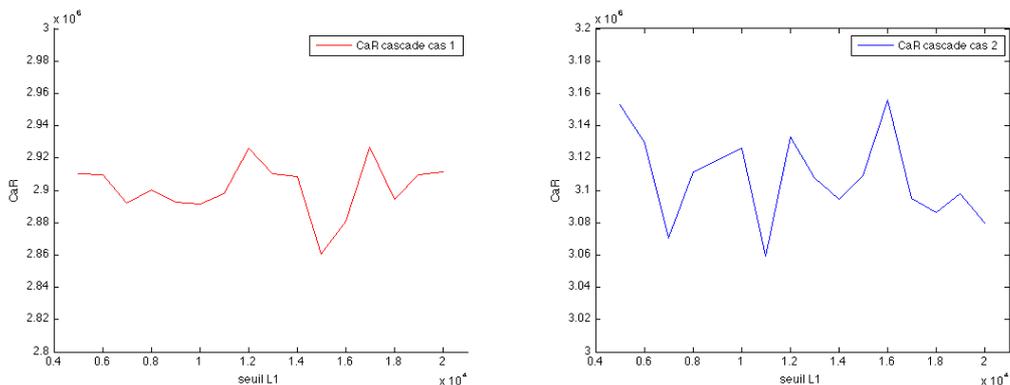


FIGURE 5.5 – Influence des coefficients $Vco(\Theta_\mu)$ et $Vco(\Theta_\sigma^2)$ sur la $CaR_{cascade}$ pour le cas 1

Comme pour la méthode bayésienne des lois conjuguées présentée section 3.5.2, **la charge en capital n'est pas sensible aux variations du paramètre $Vco(\Theta_\mu)$** . De plus, le résultat est toujours **sensible aux valeurs du paramètre $Vco(\Theta_\sigma^2)$** , cependant la variation de valeur de la CaR a diminué avec un écart relatif de 13% à comparer avec le 20% de la première méthode.

Sensibilités aux seuils L_j

On a vu dans la section précédente que la détermination du seuil restait imprécise ; on veut savoir si faire varier les seuils autour d'un petit intervalle impacte significativement les charges en capital. On fait varier le seuil L_1 de 5 000 à 20 000 € par pas de 1 000 €, et on regarde l'impact sur les charges en capital pour le cas 1 et le cas 2 :


 FIGURE 5.6 – Influence du seuil L_1 sur la $CaR_{cascade}$ pour les cas 1 et 2

Au vu de ces résultats, on constate que les charges en capital ne sont pas sensibles à de petites variations des seuils obtenus grâce à l’estimateur de Hill. **La méthode implémentée est donc plutôt robuste.**

5.1.8 Discussion sur la cascade bayésienne

Cette méthode de cascade bayésienne a l’avantage de régler une partie des problèmes d’hétérogénéité entre les informations des experts métier et les données internes en terme de sévérité que nous avons notés avec la première méthode bayésienne des lois conjuguées. Cette approche ne mélange plus deux sources complètement distinctes, mais procède en deux étapes en combinant, à l’aide de données externes, des données moins hétérogènes. Lorsque les trois sources de données sont à disposition, la cascade bayésienne est très simple à mettre en place puisqu’elle repose sur des calculs numériques simples, notamment grâce à l’utilisation de lois conjuguées permettant d’obtenir des formules explicites pour l’actualisation des paramètres structuraux lors des différentes étapes.

Malgré ces avantages, la mise en place de cette approche nécessite toujours un coût d’entrée non négligeable car les experts métier sont toujours censés fournir des intervalles et les probabilités que les paramètres de la loi de sévérité soient dans ces intervalles. Cette information peut être également matérialisée par le coefficient de variation que nous avons choisi d’utiliser. Compte tenu de ces difficultés, il semble intéressant d’étudier une autre méthode bayésienne faisant recours à des lois a priori non informatives permettant d’échapper aux difficultés que l’on vient de rencontrer. On présente cette nouvelle approche dans la section suivante.

5.2 Une alternative aux lois conjuguées : application des méthodes de Monte Carlo par chaînes de Markov

Lorsque les lois ne sont pas conjuguées ou les lois a priori sont non informatives, le calcul des paramètres a posteriori nécessite d’autres méthodes. Nous considérons dans cette section une méthode de Monte Carlo plus générale, permettant d’approcher la génération de variables aléatoires d’une loi a posteriori $\pi(\theta | x)$ lorsque cette loi ne peut pas être simulée directement. L’avantage de cette méthode sur les méthodes de Monte Carlo classiques décrites dans la section

2.2.3 est qu'elle ne nécessite pas la construction précise d'une loi a priori puisqu'elle prend en compte les caractéristiques de $\pi(\theta | x)$.

5.2.1 Principe des méthodes MCMC

Cette extension appelée *Monte Carlo par chaînes de Markov* (MCMC) tire son nom de l'idée que, pour produire des approximations acceptables d'intégrales et d'autres fonctions dépendant d'une loi d'intérêt, il suffit de générer une *chaîne de Markov* $(\theta_m)_m$ de loi limite la loi d'intérêt (voir C.P. Robert [28] et J.O. Berger [5]).

Une chaîne de Markov est une séquence de variables aléatoires X_1, X_2, \dots, X_n dépendantes telle que la distribution de X_i conditionnellement aux observations passées ne dépend que de X_{i-1} :

$$f(x_i | x_{i-1}, \dots, x_1) = f(x_i | x_{i-1})$$

Une chaîne de Markov est donc caractérisée par deux composantes : sa distribution initiale $f(x_1)$ ainsi que les distributions conditionnelles de $X_i | X_{i-1}$ qu'on appelle fréquemment kernels de transition notés $K(x_{i-1}, x_i)$. On choisit de construire des chaînes de Markov car elles possèdent deux propriétés très utiles pour notre problème qui sont l'invariance et stationnarité : si X_i est distribuée suivant une loi f alors X_{i+1} est distribuée suivant cette même loi f . Les autres propriétés des chaînes de Markov comme leur irréductibilité sont également importantes mais nous ne les mentionnerons pas pour simplifier la compréhension du problème.

On cherche à construire une chaîne de Markov pour le paramètre $\theta = (\theta_1, \dots, \theta_m)$. La chaîne de Markov est construite de manière à avoir la distribution a posteriori $\pi(\theta | x)$. La transition d'un élément dans la chaîne au suivant s'effectue grâce aux kernels de transition qui satisfont la propriété suivante qu'on appelle également *condition d'équilibre ponctuel* :

$$\pi(\theta_{i-1} | x)K(\theta_{i-1}, \theta_i) = \pi(\theta_i | x)K(\theta_i, \theta_{i-1})$$

On peut alors approcher $\mathbb{E}[\pi(\theta | x)]$ à l'aide du théorème ergodique :

$$\frac{1}{m} \sum_{i=1}^m \pi(\theta_i | x) \xrightarrow{m \rightarrow +\infty} \mathbb{E}[\pi(\theta | x)]$$

5.2.2 Algorithmes utilisés pour la MCMC

Deux algorithmes coexistent aujourd'hui d'après W.K. Hastings [19].

L'algorithme de Metropolis-Hastings

Pour une densité donnée π (connue à un facteur de normalisation près), et une densité conditionnelle $K(\theta_i, \theta_{i+1})$, l'algorithme génère la chaîne $(\theta_m)_m$ comme suit :

1. **Itération 0** : Initialiser avec une valeur arbitraire θ_0
2. **Itération m** : Mettre à jour θ_m par θ_{m+1} de la manière suivante :
 - (a) Générer $\xi \sim K(\theta_{m+1}, \theta_m)$
 - (b) Calculer la probabilité d'acceptation :

$$\alpha = \min \left(1, \frac{\pi(\xi | x)K(\theta_m, \xi)}{\pi(\theta_m | x)K(\xi, \theta_m)} \right)$$

- (c) Générer une variable aléatoire $U \sim \mathcal{U}[0, 1]$
- (d) Si $u \leq \alpha$ alors $\theta_{m+1} = \xi$, sinon $\theta_{m+1} = \theta_m$

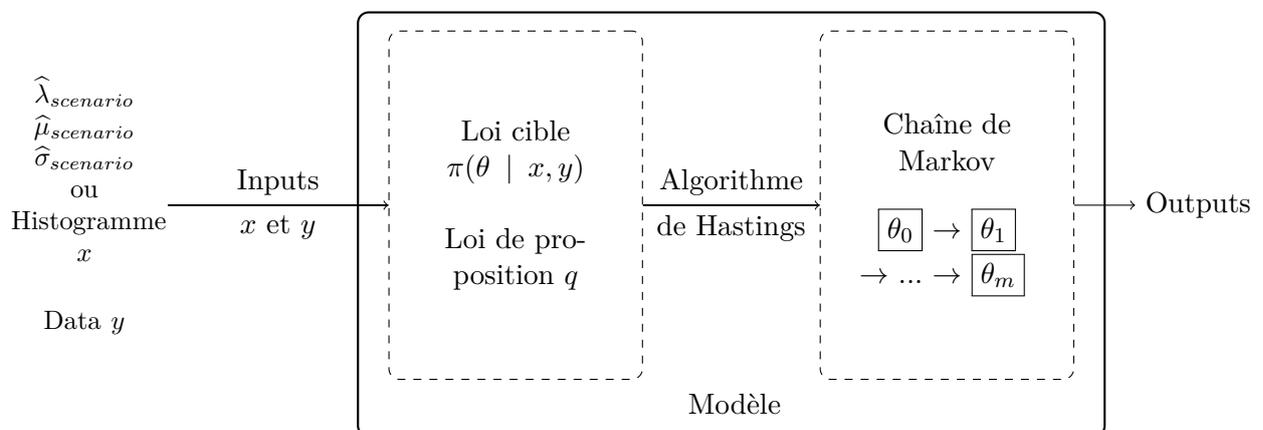
Le kernel de transition contient une loi de densité q à partir de laquelle est déterminée la valeur suivante dans la chaîne et une probabilité d'acceptation α qui détermine si la valeur proposée est acceptée ou non. Cette loi est quelconque : on la choisit donc pour qu'elle soit simulable rapidement (loi normale ou uniforme par exemple). Le choix le plus courant d'après Hastings pour q est de choisir une marche aléatoire. L'idée est de perturber aléatoirement la valeur courante de la chaîne tout en restant aux alentours de ce point et de voir si la nouvelle valeur (dans l'algorithme, il s'agit de ξ) est vraisemblable pour la loi d'intérêt.

La loi de densité π est appelée *loi cible* et la loi de densité q *loi candidate* ou *loi de proposition*.

L'échantillonnage de Gibbs

La technique de Metropolis-Hastings présentée est attrayante mais peut susciter quelques problèmes au niveau de la convergence à cause du manque de connexion entre q et π . L'approche de l'échantillonnage de Gibbs est plutôt fondée sur la loi π . On ne détaillera pas cette méthode.

Résumé schématique de la méthode



5.2.3 Application aux données du Crédit Agricole

On applique l'algorithme de Metropolis-Hastings à notre situation. Les paramètres à déterminer sont toujours le couple (λ, μ, σ) . L'algorithme nécessite comme on l'a vu précédemment de déterminer la bonne loi cible π et la bonne loi de proposition q .

Définition de la loi cible

On cherche une loi cible pour $\theta = (\lambda, \mu, \sigma)$. Si on appelle x les données des experts et y les données historiques, elle vaut :

$$\begin{aligned}\pi(\theta | x, y) &\propto \pi(\theta | x)L_{DI}(y, \theta) \\ &\propto \pi(\theta)L_S(x, \theta)L_{DI}(\theta, y)\end{aligned}$$

La loi cible est donc caractérisée par le produit des vraisemblances de chaque source de données par une loi a priori du paramètre θ . On choisit comme loi a priori une loi non informative : la loi a priori de Jeffreys. On est amené à calculer l'information de Fisher qui se définit de la manière suivante :

$$I(\theta) = \mathbb{E} \left[\left(\frac{\partial \ln f(X | \theta)}{\partial \theta} \right)^2 \right]$$

avec f la vraisemblance associée aux observations. Sous certaines conditions de régularité (voir J.O. Berger [5]), on peut écrire directement :

$$I(\theta) = -\mathbb{E} \left[\frac{\partial^2 \ln f(X | \theta)}{\partial \theta^2} \right]$$

On a vu précédemment qu'on avait alors : $\pi(\theta) \propto |I(\theta)|^{1/2}$.

Pour la sévérité $\mathcal{LN}(\mu, \sigma^2)$, nous devons prendre en compte que les observations disposent d'un seuil de collecte L , on dispose donc de 3 paramètres : $(\mu, \sigma, L) = \Theta$.

$$f(X = x | \Theta) = \frac{1}{(x - L)\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(\ln(x - L) - \mu)^2\right) \mathbb{1}_{]L; +\infty[}$$

Après avoir calculé les dérivées par rapport aux paramètres, on détermine les moyennes théoriques en X de ces quantités en effectuant des changements de variable dans les intégrales qui nous ramènent à un calcul de moments de lois normales. On en déduit :

$$I(\Theta) = \begin{pmatrix} (1 + \frac{1}{\sigma^2})e^{2(\sigma^2 - \mu)} & (1 + \frac{2}{\sigma^2})e^{\frac{\sigma^2}{2} - \mu} & -2(1 + \frac{1}{\sigma^2})e^{2(\sigma^2 - \mu)} \\ (1 + \frac{2}{\sigma^2})e^{\frac{\sigma^2}{2} - \mu} & \frac{1}{\sigma^2} & 0 \\ -2(1 + \frac{1}{\sigma^2})e^{2(\sigma^2 - \mu)} & 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

L'information de Fisher ne dépend pas du seuil de collecte. Par souci de clarté, on n'affiche pas la loi a priori obtenue.

Pour la fréquence $\mathcal{P}(\lambda)$, le calcul est plus simple car il n'y a qu'un seul paramètre et donc une seule dérivée à déterminer. On a :

$$\begin{aligned} \mathbb{P}(X = k) &= e^{-\lambda} \frac{\lambda^k}{k!}, k \in \mathbb{N} \\ \ln \mathbb{P}(X = k) &= -\lambda + k \ln \lambda - \ln(k!) \\ \frac{\partial \ln \mathbb{P}(X = k)}{\partial \lambda}(\lambda) &= -1 + \frac{k}{\lambda} \\ \frac{\partial^2 \ln \mathbb{P}(X = k)}{\partial \lambda^2}(\lambda) &= -\frac{k}{\lambda^2} \\ I(\lambda) &= \frac{1}{\lambda} \text{ car } \mathbb{E}[X] = \lambda \end{aligned}$$

Ainsi, on en déduit :

$$\begin{aligned} \pi(\lambda) &\propto \sqrt{I(\lambda)} \\ &\propto \boxed{\frac{1}{\sqrt{\lambda}}} \end{aligned}$$

En ce qui concerne les vraisemblances des deux sources de données, elles sont faciles à calculer car on connaît les lois. Pour les données internes, il s'agit d'une loi de Poisson composée de loi $\mathcal{LN}(\mu, \sigma^2)$ pour la sévérité et de loi $\mathcal{P}(\lambda)$ pour la fréquence. Pour les jugements d'experts, on distingue plusieurs cas dans notre approche :

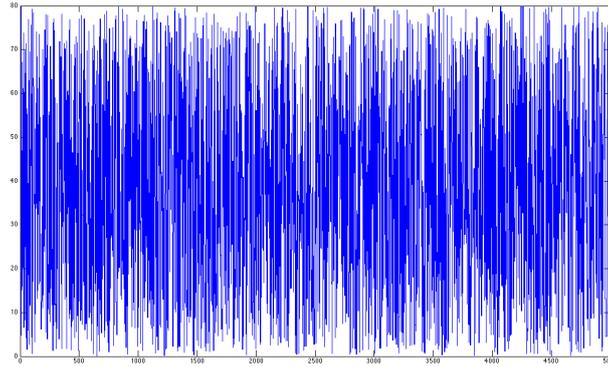
- soit on se repose sur les paramètres $(\lambda_S, \mu_S, \sigma_S)$ estimés par les experts et on détermine la vraisemblance comme pour les données historiques ;
- soit on considère les couples (x, d) et on représente tout ceci à l'aide d'un histogramme, la loi serait alors une combinaison d'uniformes.

Définition de la loi candidate

Nous avons implémenté l'algorithme de Metropolis-Hastings avec plusieurs lois candidates : dans le cas de la fréquence, on peut par exemple choisir une loi uniforme $U \sim \mathcal{U}[0, 10]$ ou encore pour l'étude de la sévérité une marche aléatoire où $\xi^{(m+1)} = \xi^{(m)} + \epsilon$ avec $\epsilon \sim \mathcal{N}(0, \sigma_{DI}^2)$.

Résultats obtenus

Résultats pour la fréquence. Ayant à disposition les paramètres λ_S et λ_{DI} définis respectivement par les experts et par maximum de vraisemblance pour les données historiques, on calcule la vraisemblance associée à chacune des sources de données en simulant un échantillon de $N = 100$ données de lois de Poisson $\mathcal{P}(\lambda_S)$ et $\mathcal{P}(\lambda_{DI})$. L'algorithme est initialisé à partir d'une valeur de $\lambda_0 = \lambda_{DI}$ et on considère comme loi de transition la loi uniforme $\mathcal{U}[\lambda_S, \lambda_{DI}]$. Pour l'entité CACF et la catégorie de risque EL6, on obtient par exemple la chaîne de Markov suivante :

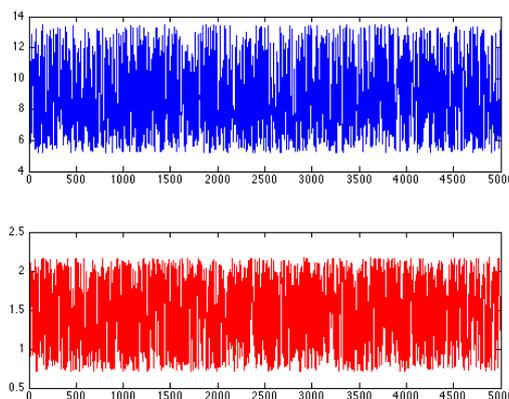
FIGURE 5.7 – Allure de la chaîne de Markov $\lambda^{(m)}$ pour le cas 2

On voit que la chaîne oscille fortement entre λ_S et λ_{DI} : si l'on prend l'espérance $\mathbb{E}[\lambda^{(m)}]$, on obtient la valeur finale du paramètre $\hat{\lambda}_{posteriori} = 36.97$. Voici un tableau récapitulatif pour chacune des entités AMUNDI (1), CACF (2), CACIB (3) :

Cas	$\hat{\lambda}_{posteriori}$
1	24.29
2	36.97
3	381.85

La chaîne est très instable car les données que l'on essaie de combiner sont très hétérogènes.

Résultats pour la sévérité. De la même manière que pour la fréquence, on dispose des paramètres μ_S , μ_{DI} , σ_S et σ_{DI} . On peut alors simuler un échantillon de $N = 100$ données de lois log-normales $\mathcal{LN}(\mu_S, \sigma_S)$ et $\mathcal{LN}(\mu_{DI}, \sigma_{DI})$. L'algorithme est initialisé à partir d'une valeur du couple $(\mu_0, \sigma_0) = (\mu_{DI}, \sigma_{DI})$ et on considère comme lois de transition les lois uniformes $\mathcal{U}[\mu_{DI}, \mu_S]$ et $\mathcal{U}[\sigma_{DI}, \sigma_S]$. Pour l'entité CACFEL6, on obtient par exemple les chaînes de Markov suivantes :

FIGURE 5.8 – Allure des chaînes de Markov $\mu^{(m)}$ (en bleu) et $\sigma^{(m)}$ (en rouge) pour le cas 2

La chaîne est beaucoup plus stable car les données sont plus homogènes que pour la fréquence. On résume les paramètres finaux (en prenant comme pour la fréquence l'espérance des valeurs prises par $\mu^{(m)}$ et $\sigma^{(m)}$) dans le tableau suivant :

Cas	$\hat{\mu}_{posteriori}$	$\hat{\sigma}_{posteriori}$
1	10.90	0.89
2	8.59	1.46
3	10.21	1.45

Une fois ces valeurs trouvées, on peut déterminer aisément la charge en capital :

Cas	CaR
1	4 447 000 €
2	2 620 200 €
3	52 079 000 €

On constate que pour chaque entité, les charges en capital sont comprises entre la charge en capital CaR_{DI} et la charge CaR_S . Cela est dû à la construction de la méthode : pour chaque vecteur de paramètres (λ, μ, σ) , **on fait évoluer la chaîne de Markov de telle sorte que les paramètres restent entre les valeurs historiques et les valeurs associées aux experts** (par l'intermédiaire de la loi candidate). On peut toutefois détecter certaines tendances : on observe que la charge en capital est plus proche des charges historiques pour les entités AMUNDI et CACF, alors qu'elle est plus proche de la charge en capital expert. Cela s'explique par le fait que pour les deux premières entités, **les évolutions des paramètres se compensent entre elles**, tandis que pour l'entité CACIB, la différence de fréquence entre les deux sources de données est si forte ($\lambda_{DI} = 816$ contre $\lambda_S = 0.2$) qu'à la fin de l'algorithme, l'évolution du λ est exagérée comparée aux évolutions des autres paramètres, et donc la charge en capital se rapproche plus d'une charge en capital extrême représentée par CaR_S .

5.2.4 Sensibilités aux paramètres

Comme on l'a dit précédemment, on a le choix entre plusieurs lois candidates. On peut donc étudier dans un premier temps les modifications apportées à la charge en capital lorsque la loi candidate n'est plus une loi uniforme mais une marche aléatoire ; on utilise dans ce cas une loi symétrique centrée autour du paramètre en question, par exemple une loi normale. Les résultats étant similaires, on décide de ne pas les présenter.

On pourrait également influencer sur la loi cible mais il faudrait alors avoir à disposition plus de précision dans les données ce qui aboutirait à une meilleure estimation des paramètres de chaque source.

5.2.5 Discussion sur la méthode MCMC

En résumé, la Markov Chain Monte Carlo est une méthode pertinente car elle permet de s'affranchir de l'utilisation des lois conjuguées. De plus, au lieu d'utiliser une loi a priori, on

intègre l'information des experts d'une autre manière : de ce fait, le manque d'information fournie par les experts ne devient pas une contrainte comme on a pu le voir pour les lois a priori conjuguées qui manquaient de précision.

Toutefois, l'utilisation de la MCMC rend le calcul des paramètres moins intuitif et rend ainsi la méthode plus opaque. Comparé aux méthodes faisant intervenir des lois conjuguées, on ne peut pas introduire de sensibilités sur les paramètres d'une loi a priori mais seulement influencer sur les lois cibles et candidates de manière arbitraire. Enfin, si les experts métier nous fournissaient des scénarios davantage munis de couples (x, d) , on pourrait imaginer construire une vraisemblance associée aux scénarios plus adéquate que celle qu'on a implémentée (vraisemblance calculée à partir du couple $(\lambda_S, \mu_S, \sigma_S)$) et ainsi améliorer la définition de la loi cible.

Partie 6

Synthèse des méthodes et perspectives

Dans ce dernier chapitre, nous commençons par faire une comparaison des méthodes implémentées, à savoir :

- la méthode actuelle employée au Crédit Agricole ;
- la méthode bayésienne pure qui fait intervenir les lois conjuguées implémentées dans la partie 3 ; dans cette méthode, on a étudié 4 méthodes différentes ;
- la méthode dite par crédibilité étudiée dans la partie 4 qui se décompose en deux études suivant qu'on ait utilisé la loi de Pareto ou une loi log-normale tronquée ;
- la cascade bayésienne à laquelle on a appliqué la même approche que la méthode bayésienne pure (partie 5) ;
- la méthode MCMC qui permet de s'affranchir du recours aux lois conjuguées indispensables pour la première méthode (partie 5).

Une fois cette comparaison faite, nous allons comparer l'ensemble des résultats à une autre méthode non bayésienne qui fait intervenir des lois de mélange et qui semble la plus naturelle pour combiner des sources différentes de données. D'autres pistes de réflexion seront proposées enfin pour améliorer nos méthodes.

6.1 Comparaison des méthodes

Dans cette section, on compare les méthodes présentées dans ce mémoire à la fois de manière qualitative en s'intéressant aux informations nécessaires pour leur mise en œuvre mais aussi de manière quantitative en s'intéressant aux différents résultats obtenus.

6.1.1 Informations et données utilisées pour la mise en œuvre des méthodes

Aujourd'hui, le Crédit Agricole utilise 5 années d'historique de pertes internes liées au risque opérationnel pour calculer la charge en capital au titre des données internes. Etant donné que les pertes collectées dans la queue de distribution sont peu nombreuses ou qu'il manque certains types d'événements, les charges réelles en capital ne suffisent pas nécessairement à couvrir l'ensemble du risque ; c'est pourquoi dans un souci d'exhaustivité les experts métier fournissent des informations supplémentaires sous forme de couples (montant, temps de retour) donnant lieu à l'estimation de nouveaux paramètres de loi. Les méthodes présentées précédemment utilisent toutes, chacune à leur manière, ces différentes informations en entrée des modèles.

Les différences notables entre les méthodes concernent les informations supplémentaires qu'il est nécessaire de détenir pour leur mise en œuvre. Chacune des méthodes nécessite des paramètres d'entrée qui lui sont spécifiques. Dans la méthode bayésienne des lois conjuguées, les experts doivent fournir, pour chacun des paramètres de loi, un intervalle et la probabilité que ce paramètre soit dans cet intervalle. Cette donnée peut s'exprimer sous la forme du coefficient de variation V_{co} que l'on a décidé d'utiliser. L'approche par crédibilité nécessite de choisir des cellules de risque similaires ainsi qu'un seuil pour modéliser les pertes à basse fréquence et haute sévérité. La cascade bayésienne telle que nous l'avons implémentée nécessite les informations des deux méthodes précédentes. Enfin la méthode Markov Chain Monte Carlo utilise une loi non informative permettant de s'affranchir de l'utilisation des lois conjuguées et donc de l'estimation des paramètres a priori. Cette dernière méthode ne demande pas d'informations supplémentaires.

6.1.2 Résultats numériques

Au cours de chacune des approches présentées dans ce mémoire, nous avons pu calculer une distribution finale de pertes agrégées au titre du risque opérationnel. Ces distributions nous ont menés au calcul des charges en capital imposées par le cadre réglementaire Bâle II et correspondant à la Value-at-Risk à 99.9%.

Méthode		CaR _{AMUNDI}	CaR _{CACF}	CaR _{CACIB}	
Méthode actuelle du C.A.		28.1 M€	6.9 M€	91.3 M€	
Méthode des données internes		1.9 M€	1.9 M€	22.7 M€	
Méthode des scénarios		26.2 M€	4.9 M€	68.6 M€	
Bayésienne pure	λ connu	σ connu	3.4 M€	3.3 M€	39.9 M€
		σ inconnu	1.7 M€	1.5 M€	19.3 M€
	inconnu	λ σ connu	0.4 M€	1.8 M€	8.9 M€
		σ inconnu	0.2 M€	0.9 M€	4.3 M€
Crédibilité	Loi de Pareto		5.0 M€	5.4 M€	2 548.5 M€
	Loi log-normale		2.4 M€	3.0 M€	27.5 M€
Cascade bayésienne		2.9 M€	3.1 M€	26.3 M€	
MCMC		4.4 M€	2.6 M€	52.1 M€	

TABLE 6.1 – Récapitulatif des montants de charges en capital obtenus avec les différentes approches

Le constat que l'on peut faire est que les résultats sont hétérogènes. Avec certaines méthodes, les montants de charge en capital CaR se situent entre les valeurs obtenues avec la méthode des données internes CaR_{DI} et celle des scénarios CaR_S ce qui répond favorablement à une intuition de départ. Les autres valeurs très inférieures à CaR_{DI} ou très supérieures à CaR_S (comme dans le cas de la loi de Pareto pour lequel le paramètre de queue estimé entraîne la divergence de la

charge en capital) ne paraissent pas exploitables. La méthode actuelle du Crédit Agricole est une méthode très prudente qui revient à sommer les valeurs CaR_{DI} et CaR_S . **Au vu de ces résultats, les quatre méthodes qui paraissent intéressantes en termes de charges en capital sont : les méthodes bayésiennes pures des lois conjuguées où λ est connu, la crédibilité utilisant la loi log-normale, la cascade bayésienne ainsi que la méthode MCMC.** Pour ces quatre approches, les montants obtenus sont soit entre les valeurs CaR_{DI} et CaR_S , soit proches de la valeur CaR_{DI} ce qui est une bonne chose dans l'optique de ne pas surévaluer la charge en capital. Il est tout de même important de noter que parmi ces méthodes, certains résultats sont obtenus de manière arbitraire et résultent d'un choix sur les jeux de paramètres utilisés en entrée des modèles. En effet, lors des études de sensibilité des charges en capital aux différents paramètres initiaux, nous avons constaté que certaines méthodes étaient plus robustes que d'autres.

- Dans l'approche par crédibilité, le résultat est sensible et dépend du seuil qui est directement lié au nombre de données internes utilisées pour modéliser les pertes extrêmes.
- Dans les méthodes bayésiennes pures des lois conjuguées et de la cascade bayésienne où le paramètre σ est inconnu, le constat est que le choix de la valeur du coefficient de variation agit directement et de façon importante sur la valeur finale de la charge en capital.
- Dans la méthode bayésienne pure avec λ et σ connus, les paramètres a priori peuvent être particulièrement différents en fonction de la méthode retenue pour les déterminer mais n'ont pas d'impact sur le montant de charge en capital calculé.
- Enfin, dans l'approche MCMC il n'y a pas vraiment de paramétrage et donc pas de sensibilité.

6.2 Les critiques du mémoire

Pour avoir une bonne gestion du risque opérationnel, il est important et primordial que les différents acteurs communiquent pour avoir une démarche cohérente et pertinente. Comme cela a été présenté tout au long de ce mémoire, l'historique de données collectées et utilisées aujourd'hui n'est pas suffisant pour avoir une vision complète du risque opérationnel. En effet, les données internes ne mentionnent pas certains événements extrêmes survenus ailleurs (par exemple l'affaire Kerviel). C'est à cet endroit qu'interviennent les experts métier qui fournissent des informations supplémentaires en visant à anticiper les pertes extrêmes non répertoriées à ce jour. L'étape suivante est alors la combinaison de ces deux sources d'informations, l'une quantitative et l'autre qualitative.

Les méthodes qui ont été étudiées dans ce mémoire ont tenté de résoudre ce problème à l'aide de la statistique bayésienne. La section précédente dans laquelle on retrouve l'ensemble des résultats obtenus précise que dans certains cas, les résultats sont obtenus de façon arbitraire et résultent directement du choix de paramétrage ce qui n'est pas envisageable et présentable auprès du régulateur. Pour la mise en place de ces méthodes, le format de données des experts n'est pas suffisant ; des inputs supplémentaires sont requis comme les coefficients de variation pour la méthode bayésienne pure des lois conjuguées. Le souci est que certains paramètres sont difficiles à appréhender par les experts métier. On comprend donc ici **toute l'importance d'une bonne communication entre ces deux acteurs du risque opérationnel** que sont l'équipe de gestion et modélisation ainsi que les experts métier.

De plus, tout au long de ce mémoire, nous avons été confrontés à des problèmes d'hétérogénéité entre les deux sources de données ce qui a complexifié leur combinaison. A ce titre, nous proposons plusieurs prolongements, bayésiens ou non, qui pourraient outrepasser ces difficultés.

6.3 Prolongements du mémoire

6.3.1 Une alternative non bayésienne : les lois de mélange

Pour pouvoir calculer l'exigence réglementaire en fonds propres pour le risque opérationnel par la méthode LDA, il a été nécessaire de connaître les distributions associées à chaque cellule de risque. Pour chacune d'entre elles, on a eu besoin de générer une distribution unique qui a imposé de calibrer un modèle prenant en compte des pertes de faibles montants et à haute fréquence, ainsi que des pertes extrêmes à montants élevés et petite fréquence. L'objectif de ce mémoire était de combiner deux sources de données pour en tirer une distribution unique ; toutefois, on pouvait imaginer à la base une modélisation très naturelle et très pertinente basée sur un mélange de lois. Plusieurs extensions peuvent alors être imaginées.

Comme on l'a vu, il est très difficile de modéliser correctement à la fois le corps et la queue de distribution en utilisant une loi unique. L'idée est de créer un modèle composite qui permet de modéliser la sévérité des pertes par une combinaison de deux densités, qui consiste en une distribution à queue fine ou moyenne jusqu'à un certain seuil puis d'une distribution à queue épaisse au-delà, par l'intermédiaire de la théorie des valeurs extrêmes qui est adaptée pour cela.

Décomposition de la distribution de sévérité

Soit L un seuil qui permet de délimiter le corps et la queue de la distribution (référence à la section 5) et X la variable aléatoire associée à la sévérité de la perte. Alors la fonction de répartition de la sévérité peut se décomposer de la manière suivante :

$$\begin{aligned} F(x) &= \mathbb{P}(X \leq x) = \mathbb{P}(X \leq x \mid X \leq L)\mathbb{P}(X \leq L) + \mathbb{P}(X \leq x \mid X > L)\mathbb{P}(X > L) \\ &= \underbrace{\mathbb{P}(X \leq x \mid X \leq L)P_L}_{\text{corps}} + \underbrace{\mathbb{P}(X \leq x \mid X > L)(1 - P_L)}_{\text{queue}} \end{aligned}$$

Pour $x \leq L$, on a : $F(x) = \mathbb{P}(X \leq x \mid X \leq L)P_L = F^c(x)P_L$ où F^c a pour seule contrainte que la loi soit tronquée à droite en L , c'est-à-dire $F^c(L) = 1$.

Pour $x > L$, l'expression devient $F(x) = P_L + F_L(x - L)(1 - P_L)$, où F_L peut être approximée par une loi GPD d'après les théorèmes énoncés dans la partie 5.

On peut alors considérer des modèles du type :

$$F(x) = \begin{cases} P_L \frac{F^c(x)}{F^c(L)} & \text{si } 0 < x \leq L \\ P_L + (1 - P_L)G(x) & \text{si } L < x < \infty \end{cases}$$

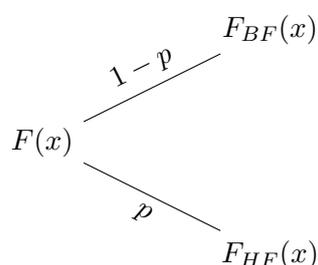
On peut également s'arranger pour prendre en considération le seuil de collecte H et on aboutit à la fonction de répartition suivante :

$$F(x) = \begin{cases} 0 & \text{si } 0 < x < H \\ P_L \frac{F^c(x)}{F^c(L) - F^c(H)} & \text{si } H \leq x \leq L \\ P_L + (1 - P_L)G(x) & \text{si } L < x < \infty \end{cases}$$

La prise en compte du seuil de collecte se fait par l'intermédiaire du corps de la distribution.

Simulation de la loi de mélange

Dans ce cas, il paraît intuitif de combiner dans une loi de mélange les pertes haute fréquence basse sévérité que l'on indicera HF et les pertes basse fréquence haute sévérité que l'on indicera BF .



$p = P_L$ désigne la probabilité qu'on tire les pertes dans les hautes fréquences. F_{BF} désigne la fonction de répartition des pertes BF , il s'agit d'une loi de Poisson composée. Plusieurs choix sont possibles pour paramétrer la fréquence et la sévérité :

- pour la fréquence, on choisit une loi de Poisson de paramètre λ fourni par les experts ;
- pour la sévérité, on a plusieurs possibilités :
 - soit on simule la sévérité par une log-normale de paramètres μ et σ fournis par les experts ;
 - soit on utilise les données historiques et la fonction de répartition vue précédemment en utilisant la loi GPD G ;
 - on pourrait également imaginer de combiner ces deux sources de données en utilisant l'approche bayésienne de la partie 3 ; on utiliserait seulement dans ce cas les données historiques supérieures à un certain seuil L .

En ce qui concerne F_{HF} , on utiliserait uniquement les données historiques qui sont seules explicatives du corps de la distribution, on prendrait alors la fonction de répartition F^c définie précédemment.

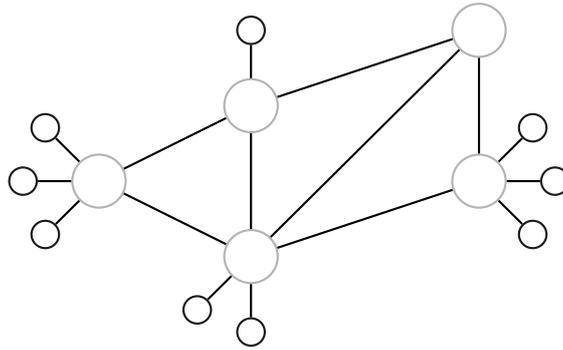
La difficulté restante est le choix de la probabilité p . On pourrait alors se baser sur les couples (x, d) fournis par les experts en définissant la probabilité $p = \frac{1}{d}$.

6.3.2 Une seconde alternative : les réseaux bayésiens

Une méthode abordée rapidement dans la partie 1 mentionnait l'utilisation de l'approche bayésienne par l'intermédiaire de réseaux. Cette méthode n'a pas été traitée car la méthode

nécessite de trop nombreux renseignements fournis par les experts métier. De nombreux articles font intervenir cette méthode (voir V. Aquararo et al. [1] et M. Neil et al. [22] pour plus d'amples références).

Les réseaux bayésiens font appel à la théorie des graphes et à la théorie des probabilités. Graphiquement, les variables aléatoires sont représentées sous forme de nœud(s) et les dépendances entre les variables sous formes de liens (ces dépendances sont quantifiées par des probabilités conditionnelles) :



En associant à chacun de ces nœuds une table de probabilité, il devient possible de représenter la propagation et l'influence d'une variable sur un phénomène complexe.

La modélisation des risques opérationnels passe alors par la définition de scénarios qui servent à identifier les vulnérabilités (pour chaque cellule de risque). Ensuite, ces scénarios sont quantifiés en définissant pour chacun une exposition, une survenance (fréquence) et une gravité (sévérité). C'est dans cette quantification qu'interviennent les données historiques. Par exemple, si la survenance d'un scénario est très fréquente et stable dans le temps, une estimation empirique sera suffisante et pour cela, on utilisera l'historique du nombre de pertes constatées. Chaque scénario a alors son propre réseau bayésien et des techniques sont alors employées pour rassembler ces scénarios et définir une distribution de perte annuelle.

L'avantage de cette modélisation est qu'elle permet de comprendre les processus générateurs de risques et d'identifier des leviers de réduction de ces risques. Toutefois, cette méthode exige beaucoup d'informations car elle repose sur la définition de scénarios précis et nécessite une bonne compréhension des mécanismes d'interaction entre les différentes vulnérabilités.

6.3.3 Amélioration de la modélisation LDA

Enfin, nous concluons ces perspectives en proposant quelques possibilités d'amélioration de nos modèles. Schevchenko et Temnov dans P.V. Shevchenko et G. Temnov [30] évoquent la possibilité d'introduire un seuil variant au cours du temps dans la modélisation du risque opérationnel. Durant tout ce mémoire, nous avons en effet pris en compte des seuils de collecte H et des seuils de valeur extrême L et nous avons fait l'hypothèse de les considérer constants. Cependant, une banque d'une année à l'autre peut changer ses niveaux de reporting et ainsi le seuil peut varier au cours du temps. On pourrait alors imaginer une modélisation en prenant en compte cette caractéristique.

De plus, on a mentionné dans la partie 2 la possibilité de s'assurer contre le risque opérationnel. Le Comité de Bâle autorise les banques à s'assurer au maximum contre 20 % des pertes

opérationnelles. L'introduction d'une assurance permettrait de réduire la charge en capital. Pour en savoir plus sur le sujet, nous conseillons l'article G.W. Peters et al. [25] qui discute de la nécessité ou non de s'assurer contre ce type de pertes.

Notre étude s'est concentrée sur les problématiques de combinaison des sources de données au niveau d'une cellule de risque. Elle pourrait être complétée par les questions d'agrégation. Nous renvoyons à l'article G.W. Peters et al. [26] pour l'introduction de copules dans le contexte d'inférence bayésienne.

Conclusion

Avec sa rigueur mathématique et son cadre relativement simplifié de mise en œuvre, l'approche LDA paraît très adaptée aujourd'hui pour mesurer le risque opérationnel. Cependant, son fondement statistique l'expose aux exigences des données de pertes. Ne disposant pas nécessairement de données internes sur un horizon lointain et vu le caractère dynamique de l'environnement associé au risque opérationnel, il faut trouver des outils permettant d'intégrer toutes les données de l'industrie financière pour mener à bien cette opération : les données internes, les données externes si il y a et les jugements des experts métier. En plus d'intégrer ces autres sources, il faut également pouvoir les combiner ; l'approche actuelle du Crédit Agricole ne le faisant pas à proprement parler, c'est dans ce contexte que ce mémoire a été rédigé.

L'analyse bayésienne fournit un moyen de combiner ces différentes sources de données. L'approche bayésienne a permis d'incorporer une information a priori à l'estimation des paramètres de fréquence et de sévérité pour calculer a posteriori une distribution qui prenait en compte l'ensemble de l'information disponible. Cependant, malgré tous les avantages des méthodes que nous avons pu aborder dans ce mémoire, toutes ont mis en avant la nécessité d'avoir une information plus précise pour être d'autant plus efficaces. Par exemple, au cours de l'implémentation de notre première méthode, qualifiée d'approche bayésienne pure, le principal obstacle a été la définition de la loi a priori. Face à des informations fournies par les experts métier qui ne caractérisaient que les pertes opérationnelles sous la forme de scénarios extrêmes, il a été difficile d'appréhender et de calculer une quelconque loi a posteriori.

Pour remédier à cette difficulté, nous avons mis au point plusieurs méthodes qui permettaient d'intégrer une information a priori moindre. L'approche par crédibilité ne nécessite que l'apport de différences de fréquence et de sévérité entre cellules de risque, mais en contrepartie ces dernières doivent avoir un profil similaire au risque de surestimer les paramètres des distributions a posteriori. La cascade bayésienne permet de réduire l'hétérogénéité des deux sources de données en intégrant des données externes qui se place à mi-distance entre les jugements d'experts et les données historiques. Enfin, la Markov Chain Monte Carlo permet de s'affranchir des lois conjuguées : ainsi, le manque d'information venant des experts n'induit pas une contrainte mais seulement un manque de précision dans nos résultats finaux.

Finalement, toutes les méthodes que nous avons pu implémenter ou les alternatives que nous avons pu évoquer en fin de mémoire ont mis en avant la nécessité d'une bonne communication pour mettre au point une méthode quantitative prenant en compte une expertise qualitative. Le Comité de Bâle impose une modélisation du capital réglementaire transparente et rigoureuse. Avant tout choix de modélisation, il est indispensable de considérer l'ensemble des acteurs et de s'entretenir avec eux pour avoir une démarche cohérente et pertinente.

Bibliographie

- [1] V. AQUARO, M. BARDOSCIA, R. BELLOTTI, A. CONSIGLIO, F. DE CARLO et G. FERRI : *A bayesian networks approach to operational risk*, 2009.
- [2] F. AUE et M. KALKBRENER : *LDA at Work*, 2007.
- [3] BANQUE DE FRANCE : *Le risque opérationnel, pratiques et perspectives réglementaires*, 2003.
- [4] J. BERGER et M. BERLINER : *Robust Bayes and Empirical Bayes analysis with contaminated priors*, 1986.
- [5] James O. BERGER : *Statistical Decision Theory and Bayesian Analysis*. Springer, 1985.
- [6] H. BÜHLMANN et A. GISLER : *A course in Credibility Theory*. Springer, 2005.
- [7] H. BÜHLMANN et P.V. SHEVCHENKO : *A "toy" model for operational risk quantification using credibility theory*, 2007.
- [8] CEIOPS : *QIS5 Technical Specifications*, 2010.
- [9] V. CHAVEZ-DEMOULIN et P. EMBRECHTS : *Advanced extremal models for operational risk*, 2004.
- [10] Comité de BÂLE : *Working paper on the regulatory treatment of operational risk*, 2001.
- [11] Comité de BÂLE : *Saines pratiques pour la gestion et la surveillance du risque opérationnel*, 2003.
- [12] Comité de BÂLE : *Convergence internationale de la mesure et des normes de fonds propres*, 2006.
- [13] M. DEGEN, P. EMBRECHTS et D.D. LAMBRIGGER : *The Quantitative modelling of operational risk : between g-and-h and EVT*, 2007.
- [14] EIOPA : *EIOPA Report on the fifth Quantitative Impact Study (QIS5) for Solvency II*, 2011.
- [15] A. FRACHOT, P. GEORGES et T. RONCALLI : *Loss Distribution Approach for operational risk*. 2001.
- [16] A. FRACHOT, P. GEORGES et T. RONCALLI : *Loss Distribution Approach in practice*. 2003.
- [17] J. GAMONET : *Modélisation du risque opérationnel dans l'assurance*, 2006.
- [18] B.K. HASSANI et A. RENAUDIN : *The Cascade Bayesian Approach for a controlled integration of internal data, external data and scenarios*, 2013.
- [19] W.K. HASTINGS : *Monte Carlo sampling methods using Markov chains and their applications*. Biometrika, 1970.
- [20] D.D. LAMBRIGGER, P.V. SHEVCHENKO et M.V. WÜTHRICH : *The quantification of operational risk using internal data, relevant external data and expert opinions*, 2007.
- [21] METAMETRIS : *Modélisation des risques opérationnels dans le cadre de Solvency II*, 2010.
- [22] M. NEIL, N. FENTON et M. TAILOR : *Using bayesian networks to model expected and unexpected operational losses*, 2005.

- [23] ONB : *Guidelines on Operational Risk Management*, 2006.
- [24] C. PARTRAT et J-L. BESSON : *Assurance non-vie : modélisation et simulation*. Economica, 1970.
- [25] G.W. PETERS, A.D. BYRNES et P.V. SCHEVCHENKO : *Impact of insurance for operational risk : is it worthwhile to insure or be insured for severe losses ?*, 2010.
- [26] G.W. PETERS, P.V. SCHEVCHENKO et M.V. WÜTHRICH : *Dynamic operational risk : modelling dependence and combining differentes sources of information*, 2009.
- [27] A. RENAUDIN : *Modèle de capital économique pour le risque opérationnel bancaire : estimation, diversification*, 2010.
- [28] C.P. ROBERT : *Le choix bayésien : Principes et pratique*. Springer, 2006.
- [29] P.V. SCHEVCHENKO : *Modelling operational risk using bayesian inference*. Springer, 2004.
- [30] P.V. SCHEVCHENKO et G. TEMNOV : *Modeling operational risk data reported above a time-varying threshold*, 2009.
- [31] P.V. SCHEVCHENKO et M.V. WÜTHRICH : *The structural modelling of operational risk via bayesian inference : Combining loss data with expert opinions*, the journal of operational risk, 2006.
- [32] P.V. SHEVCHENKO : *Implementing Loss Distribution Approach for Operational Risk*, 2009.

Annexe A

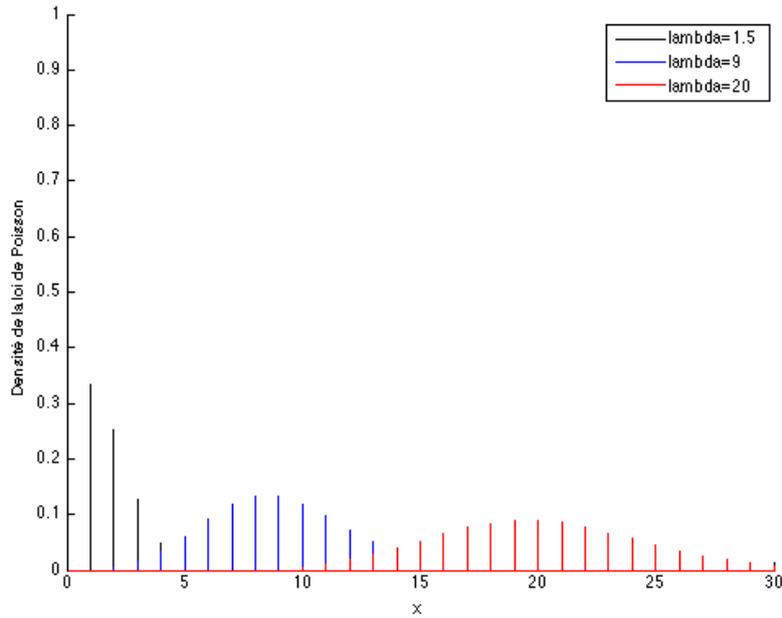
Quelques caractéristiques sur les lois usuelles utilisées

Dans cette annexe sont exposés les principaux résultats sur les lois que l'on utilise tout au long de ce mémoire.

A.1 Les lois discrètes

A.1.1 La loi de Poisson

- Notation : $N \sim \mathcal{P}(\lambda)$
- Paramètres : $\lambda \in \mathbb{R}^+$
- Support : $k \in \mathbb{N}$
- Fonction de densité : $p(k) = \mathbb{P}[N = k] = \frac{\lambda^k e^{-\lambda}}{k!}, k \geq 0$
- Espérance : $\mathbb{E}[N] = \lambda$
- Variance : $\mathbb{V}(N) = \lambda$
- Coefficient de variation : $\text{Vco}(N) = \frac{1}{\sqrt{\lambda}}$
- Fonction génératrice des moments : $M_N(t) = \exp(\lambda(e^t - 1))$

FIGURE A.1 – Densité de la loi de Poisson pour différentes valeurs de λ

A.2 Les lois continues

A.2.1 La loi uniforme

- Notation : $X \sim \mathcal{U}(a, b)$
- Paramètres : $-\infty < a < b < +\infty$
- Support : $x \in [a, b]$
- Fonction de densité : $f(x) = \frac{1}{b-a}$ pour $x \in [a, b]$
- Fonction de répartition : $F(x) = \frac{x-a}{b-a}$ pour $x \in [a, b]$
- Espérance : $\mathbb{E}[X] = \frac{a+b}{2}$
- Variance : $\mathbb{V}(X) = \frac{(b-a)^2}{12}$
- Coefficient de variation : $\text{Vco}(X) = \frac{b-a}{\sqrt{3}(a+b)}$
- Fonction génératrice des moments : $M_X(t) = \frac{e^{bt} - e^{at}}{(b-a)t}$

A.2.2 La loi Gamma

- Notation : $X \sim \mathcal{G}(\alpha, \beta)$
- Paramètres : $\alpha > 0, \beta > 0$
- Support : $x \in [0, +\infty[$
- Fonction de densité : $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, x > 0, \alpha > 0, \beta > 0$
- Fonction de répartition : notée $H(x, \alpha, \beta)$, forme non explicite pour $\alpha \notin \mathbb{N}^+$
- Espérance : $\mathbb{E}[X] = \frac{\alpha}{\beta}$
- Variance : $\mathbb{V}(X) = \frac{\alpha}{\beta^2}$
- Coefficient de variation : $\text{Vco}(X) = \frac{1}{\sqrt{\alpha}}$
- Fonction génératrice des moments : $M_X(t) = \left(\frac{\beta}{\beta-t}\right)^\alpha$

A.2.3 La loi Inverse Gamma

- Notation : $X \sim \mathcal{IG}(\alpha, \beta)$
- Paramètres : $\alpha > 0, \beta > 0$
- Support : $x \in [0, +\infty[$
- Fonction de densité : $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-1-\alpha} \exp\left(-\frac{\beta}{x}\right)$
- Fonction de répartition : $\frac{\Gamma(\alpha, \frac{\beta}{x})}{\Gamma(\alpha)}$
- Espérance : $\mathbb{E}[X] = \frac{\beta}{\alpha-1}$ pour $\alpha > 1$
- Variance : $\mathbb{V}(X) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$
- Coefficient de variation : $\text{Vco}(X) = \frac{1}{\sqrt{\alpha-2}}$

A.2.4 La loi normale

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

- Support : $x \in \mathbb{R}$
- Paramètres : $-\infty < \mu < \infty, \sigma^2 > 0$
- Fonction de densité : $f(x) = \frac{1}{x\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Fonction de répartition : $F(x) = \Phi \frac{x-\mu}{\sigma}$
- Espérance : $\mathbb{E}[X] = \mu$
- Variance : $\mathbb{V}(X) = \sigma^2$
- Coefficient de variation : $\text{Vco}(X) = \frac{\sigma}{\mu}$
- Fonction génératrice des moments : $M_X(t) = e^{t\mu + t^2 \frac{\sigma^2}{2}}$

A.2.5 La loi log-normale

$$X \sim \mathcal{LN}(\mu, \sigma^2)$$

- Support : $x \in \mathbb{R}^+$
- Paramètres : $-\infty < \mu < \infty, \sigma^2 > 0$
- Fonction de densité : $f(x) = \frac{1}{x\sqrt{2\pi\sigma}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$
- Fonction de répartition : $F(x) = \Phi \frac{\ln x - \mu}{\sigma}$
- Espérance : $\mathbb{E}[X] = e^{\mu + \frac{\sigma^2}{2}}$
- Variance : $\mathbb{V}(X) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$
- Coefficient de variation : $\text{Vco}(X) = \sqrt{e^{\sigma^2} - 1}$
- Fonction génératrice des moments : forme non analytique

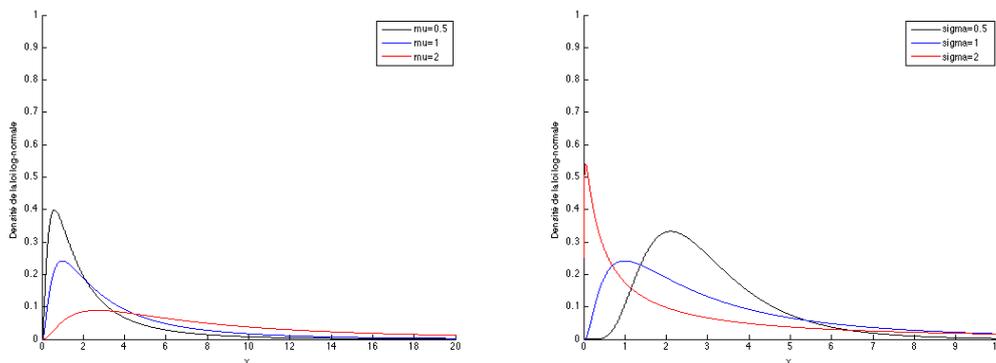


FIGURE A.2 – Densité de la loi lognormale pour différentes valeurs de μ ($\sigma = 1$ à gauche) et de σ ($\mu = 1$ à droite)

A.2.6 La loi de Pareto

$$X \sim Pa(\alpha, \lambda)$$

- Support : $x \in \mathbb{R}^+$
- Paramètres : $\alpha > 0, \lambda > 0$
- Fonction de densité : $f(x) = \frac{\alpha \lambda^\alpha}{(\lambda+x)^{\alpha+1}}$
- Fonction de répartition : $F(x) = 1 - \left(\frac{\lambda}{\lambda+x}\right)^\alpha$
- Espérance : $\mathbb{E}[X] = \frac{\lambda}{\alpha-1}$, si $\alpha > 1$
- Variance : $\mathbb{V}(X) = \frac{\alpha \lambda^2}{(\alpha-1)^2(\alpha-2)}$, si $\alpha > 2$
- Coefficient de variation : $Vco(X) = \sqrt{\frac{\alpha}{\alpha-2}}$, si $\alpha > 2$
- Fonction génératrice des moments : n'existe pas

Le paramètre λ de la loi représentant un seuil, on ne s'attarde que sur l'influence du paramètre α .

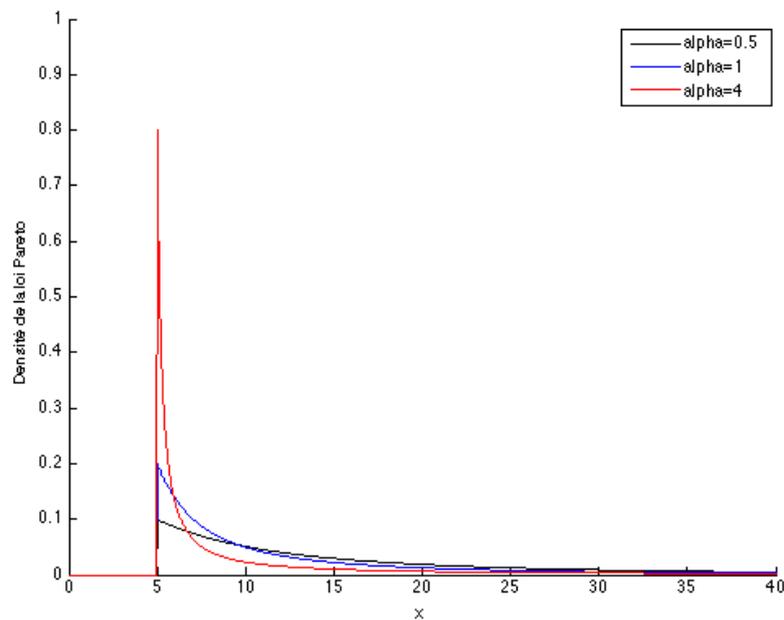


FIGURE A.3 – Densité de la loi Pareto pour différentes valeurs de α

A.2.7 La loi GPD

$$X \sim GPD(\xi, \beta)$$

- Support : $x \in \mathbb{R}^+$ lorsque $\xi \in \mathbb{R}^+$, $x \in \left[0; -\frac{\beta}{\xi}\right]$ lorsque $\xi < 0$
- Paramètres : $\xi \in \mathbb{R}$, $\beta \in \mathbb{R}^*$
- Fonction de densité :

$$f(x) = \begin{cases} \frac{1}{\beta} \left(1 + \frac{\xi x}{\beta}\right)^{-\frac{1}{\xi}-1} & \text{si } \xi \neq 0 \\ \frac{1}{\beta} \exp\left(-\frac{x}{\beta}\right) & \text{si } \xi = 0 \end{cases}$$

- Fonction de répartition :

$$F(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi} & \text{si } \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\beta}\right) & \text{si } \xi = 0 \end{cases}$$

- Espérance : $\mathbb{E}[X] = \frac{\beta}{1-\xi}$, si $\xi < 1$
- Variance : $\mathbb{V}(X) = \frac{\beta^2}{(1-\xi)^2(1-2\xi)}$, si $\alpha > 2$
- Coefficient de variation : $\text{Vco}(X) = \frac{1}{\sqrt{1-2\xi}}$, si $\xi < \frac{1}{2}$

Annexe B

Quelques éléments sur la théorie des valeurs extrêmes

B.1 Méthode de calcul du seuil L avec la fonction *Mean excess*

La fonction moyenne des excès s'exprime de la manière suivante :

$$e(L) = \mathbb{E}(X - L \mid X > L)$$

On peut réécrire :

$$\begin{aligned} e(L) &= \int_0^{x_F - L} x \frac{dF(x + L)}{1 - F(L)} \\ &= \frac{1}{\bar{F}(L)} \int_L^{x_F} (y - L) dF(y) \\ &= -\frac{1}{\bar{F}(L)} \int_L^{x_F} (y - L) d\bar{F}(y) \\ &= -\frac{1}{\bar{F}(L)} ([(y - L)\bar{F}(y)]_L^{x_F} - \int_L^{x_F} \bar{F}(y) dy) \\ &= \frac{1}{\bar{F}(L)} \int_L^{x_F} \bar{F}(x) dx \end{aligned}$$

Dans le cas de la loi de Pareto généralisée (GPD), on a $G_{\xi, \beta} = (1 + \frac{\xi L}{\beta})^{-\frac{1}{\xi}}$. Ainsi :

$$e(L) = (1 + \frac{\xi L}{\beta})^{-\frac{1}{\xi}} \int_L^{x_F} (1 + \frac{\xi x}{\beta})^{-\frac{1}{\xi}} dx$$

Dans le cas $\xi < 1$ et $\xi \neq 0$:

$$\begin{aligned} e(L) &= (1 + \frac{\xi L}{\beta})^{\frac{1}{\xi}} \left[\frac{\beta}{\xi - 1} (1 + \frac{\xi x}{\beta})^{\frac{\xi - 1}{\xi}} \right]_L^{x_F} \\ &= -(1 + \frac{\xi L}{\beta})^{\frac{1}{\xi}} \frac{\beta}{\xi - 1} (1 + \frac{\xi L}{\beta})^{\frac{\xi - 1}{\xi}} \\ &= -\frac{\beta}{\xi - 1} (1 + \frac{\xi L}{\beta}) \\ &= \frac{\beta + \xi L}{1 - \xi} \end{aligned}$$

Si $\xi \geq 1$, l'intégrale ci-dessus est divergente et $e(L)$ n'existe pas : c'est le domaine de "fuite vers l'infini", propre aux distributions à queues épaisses.

Enfin, si $\xi = 0$, nous avons :

$$\begin{aligned} e(L) &= e^{\frac{L}{\beta}} \int_L^{+\infty} e^{-\frac{x}{\beta}} dx \\ &= e^{\frac{L}{\beta}} [-\beta e^{-\frac{x}{\beta}}]_L^{+\infty} \\ &= \beta \end{aligned}$$

Dans le cas $\xi = 0$, la fonction moyenne des excès est une constante, c'est une fonction linéaire en L . Nous pouvons ainsi en déduire le critère de sélection du seuil : on cherche l tel que la *mean excess function* soit linéaire à partir de ce seuil. En effet, cela signifie que la fonction F_L est proche de la GPD et donc que l'hypothèse est raisonnable.