

**PROMOTION 2006**  
**Mémoire d'actuariat présenté devant**  
**Le Jury du CEA**  
Pour l'obtention du  
**Diplôme d'Actuaire du CENTRE D'ÉTUDES**  
**ACTUARIELLES**  
**Par Julie GAMONET**

**Sur le Sujet**  
**MODÉLISATION DU RISQUE OPÉRATIONNEL**  
**DANS L'ASSURANCE**  
**Devant un jury composé de**

Liste du Jury

Directeur du mémoire

Thomas BEHAR  
Vincent DAMAS  
Arnaud COHEN  
Gérard CROSET  
Jean-Pierre DIAZ  
Brigitte DUBUS  
Paul ESMEIN  
Michel FROMENTEAU  
Benoît HUGONIN  
Christophe IZART  
Pierre PETAUTON  
Florence PICARD  
Christian Yann ROBERT

Stéphane MENART

CONFIDENTIALITÉ



## REMERCIEMENTS

J'adresse mes remerciements à Stéphane Ménart, directeur de la Direction Technique et des Risques, pour sa disponibilité, son soutien et ses précieux conseils.

Je tiens également à remercier tous les membres de la Direction Technique et des Risques qui m'ont supportée et beaucoup aidée, en particulier la directrice de la Veille Technique Brigitte Dubus et mes collègues Romain Boyer Chammard, Alexandre You et Véronique Girault.

Je voudrais aussi remercier Christian Yann Robert pour sa grande disponibilité au cours de la rédaction de mon mémoire.

Je tiens à remercier Patrick Naïm pour son aide dans le cadre des réseaux bayésiens.

Enfin, je souhaite remercier mon entourage pour m'avoir soutenue et encouragée.



## RÉSUMÉ

La nouvelle norme européenne Solvabilité II sera applicable dès 2012 et remplacera Solvabilité I. Elle impose aux compagnies d'assurance un nouveau cadre réglementaire plus strict mais plus adapté à leurs risques propres. Une des grandes nouveautés est que les compagnies ont l'obligation de mobiliser une partie de leurs fonds propres en couverture de leur exposition aux risques opérationnels. Pour calculer ce montant de capital, le régulateur propose deux approches : 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 le risque correspond réellement à la situation de l'entreprise. Le CEIOPS incite les compagnies à opter pour le modèle interne en rendant l'approche standard beaucoup plus consommatrice en fonds propres. Ce mémoire s'adresse aux compagnies qui souhaitent mettre en place un modèle interne pour le risque opérationnel. Il propose notamment un modèle interne simplifié qui ne nécessite pas forcément de données observées en s'appuyant sur des avis d'experts. Puis il propose un modèle plus complexe dans lequel il distingue les risques de fréquence des risques de gravité qu'il modélise de façon indépendante par les approches Loss Distribution Approach (LDA) et réseaux bayésiens respectivement. Il traite des problèmes de manque de données, de seuil de collectes, d'estimation de probabilités conditionnelles et d'agrégation des risques. Il aborde également la théorie des valeurs extrêmes pour modéliser les risques de gravité.

## ABSTRACT

The new European standard Solvency II will be applicable from 2012 and will replace Solvency I. It imposes to insurance companies a new stricter statutory frame but more adapted to their appropriate risks. One of the big novelties is that the companies have the obligation to mobilize a part of their stockholders' equity on the cover from their exposure to the operational risks. To calculate this amount of capital, the regulator proposes two approaches: a standard approach and an advanced approach. The standard approach is a simplified approach computed as a percentage of the premiums or the reserves. The advanced approach is an internal model where the risk really corresponds to the situation of the company. The CEIOPS incites the companies to opt for the internal model by making the standard approach much more consumer in stockholders' equity. This report addresses the companies which wish to set up an internal model for the operational risk. It proposes in particular a simplified internal model which does not require necessarily observed data by leaning on opinions of experts. Then it proposes a more complex model in which he distinguishes the risks of frequency of the risks of gravity which are modelled in an independent way by the approaches LDA and the Bayesian networks respectively. It treats problems of lack of data, of threshold of collections, estimation of conditional probability and aggregation of the risks. It also approaches the theory of the extreme values to model the risks of gravity.

# TABLE DES MATIÈRES

<b>1 Introduction</b> .....	<b>9</b>
1.1 Qu'est-ce que le risque opérationnel ? .....	9
1.2 Contexte Solvabilité II .....	9
1.3 Enjeux et objectifs pour les compagnies d'assurance .....	10
1.4 Objectifs du mémoire .....	11
<b>2 Présentation du risque opérationnel</b> .....	<b>13</b>
2.1 Exemples de risque opérationnel fréquent .....	14
2.2 Exemples de risque opérationnel grave potentiel .....	14
2.3 Exemples de risque opérationnel grave avéré .....	15
2.4 Comparaison du risque opérationnel sous Solvabilité II et Bâle II .....	15
<b>3 Gestion du risque opérationnel dans une compagnie d'assurance</b> .....	<b>19</b>
3.1 Présentation générale de la démarche .....	19
3.2 Principe du Self Risk Assessment .....	20
3.3 Exemple de quantification de risque dans le Self Risk Assessment .....	23
3.4 Cartographie des risques .....	23
<b>4 Présentation des méthodes de quantification</b> .....	<b>27</b>
4.1 Formule standard .....	27
4.1.1 Formule standard issue des QIS4 .....	27
4.1.2 Formule standard issue du CP53 .....	29
4.2 Méthode de mesures avancées .....	31
4.2.1 Méthode LDA (Loss Distribution Approach) .....	31
4.2.2 Approche par scénarios .....	34
4.2.3 Approche Bayésienne .....	34
4.3 Méthodes de calcul de la charge totale .....	37
4.3.1 Formule récursive de Panjer .....	37
4.3.2 Inversion de la fonction caractéristique .....	41
4.3.3 Méthode basée sur les moments .....	43
4.3.4 Simulation de Monte-Carlo .....	43
4.4 Théorie des Valeurs Extrêmes .....	45
4.5 Mesures de risque .....	51
4.6 Agrégation des risques .....	52
4.6.1 Mesure de concordance .....	53
4.6.2 Définition des copules .....	54
4.6.3 Estimation des paramètres d'une copule .....	58
4.6.4 Méthode des rangs .....	60
<b>5 Modèle théorique du risque opérationnel</b> .....	<b>63</b>
5.1 Modélisation des risques de fréquence .....	64
5.2 Modélisation des risques de gravité .....	68
5.2.1 Apprentissage des connaissances .....	69
5.2.2 Apprentissage des données .....	75
5.3 Définition de la matrice des risques .....	87
<b>6 Modélisation simplifiée du risque opérationnel</b> .....	<b>89</b>

<b>7 Application de la modélisation des risques de fréquence .....</b>	<b>95</b>
7.1 Statistiques descriptives des données .....	99
7.2 Estimation des distributions de sévérité .....	103
7.3 Estimation des distributions de fréquence.....	114
7.4 Application de la méthode Loss Distribution Approach avec seuil .....	117
7.5 Simulation des scénarios de perte totale .....	126
7.6 Agrégation des scénarios.....	128
<b>8 Application de la modélisation des risques de gravité.....</b>	<b>131</b>
8.1 Self Risk Assessment du risque .....	132
8.2 Modèle simplifié .....	133
8.3 Modèle Bayésien .....	138
8.4 Comparaison des résultats et conclusions .....	145
<b>9 Application de la théorie des valeurs extrêmes.....</b>	<b>147</b>
9.1 Modélisation des pertes extrêmes .....	147
9.2 Modélisation des pertes attritionnelles.....	155
9.3 Simulation des pertes et résultat.....	163
<b>10 Capital pour tout le risque opérationnel .....</b>	<b>165</b>
10.1 La formule standard issue des QIS4.....	165
10.2 La formule standard issue du CP53.....	166
10.3 Comparaison des résultats entre formules standards QIS4 vs CP53.....	167
10.4 Modèle interne simplifié .....	168
<b>11 Conclusion générale .....</b>	<b>169</b>
<b>12 Bibliographie.....</b>	<b>171</b>
<b>13 Liste des abréviations .....</b>	<b>173</b>
<b>14 Annexes .....</b>	<b>174</b>
14.1 Annexe A : Référentiel des risques Bâle II.....	174
14.2 Annexe B : Référentiel des processus de Generali .....	176
14.3 Annexe C : Référentiel des risques de Generali.....	177
14.4 Annexe D : Scénarios de place du secteur bancaire.....	180

# 1 INTRODUCTION

## 1.1 QU'EST-CE QUE LE RISQUE OPÉRATIONNEL ?

Le risque opérationnel peut être défini comme le risque de pertes résultant de procédures internes inadaptées ou défaillantes, d'erreurs humaines, de défaillance des systèmes informatiques, d'évènements extérieurs. Ce sont donc des risques qui existent depuis toujours dans les compagnies. Il peut s'agir par exemple, de problèmes informatiques qui ralentissent les salariés pendant quelques heures qui sont des incidents banals et courants mais il peut également s'agir de l'incendie des locaux comme l'incendie du Crédit Lyonnais qui avait engendré de lourdes pertes. Le risque opérationnel est donc un risque qui peut s'avérer très coûteux.

Jusqu'à présent, la réglementation ne prévoyait pas sa prise en compte. Aujourd'hui, c'est chose faite avec la nouvelle réglementation Solvabilité II.

## 1.2 CONTEXTE SOLVABILITÉ II

Depuis 2001, le Comité Européen des Contrôleurs d'Assurance et des Pensions Professionnelles (CEIOPS) prépare un nouveau cadre réglementaire pour les compagnies d'assurances, Solvabilité II, qui s'apparente à celui mis en place dans les banques avec la réforme Bâle II. Ce cadre permettra d'harmoniser et de moderniser les règles de solvabilité en les rendant plus adaptées aux risques de l'entreprise.

Ce projet s'organise en plusieurs niveaux dont le premier niveau est la directive Solvabilité II qui édicte les principes. Adoptée par le Conseil Ecofin<sup>1</sup> et votée le 22 avril 2009 par le Parlement Européen, la nouvelle directive entrera en vigueur en octobre 2012 et remplacera la directive Solvabilité I en vigueur actuellement. Le niveau 2 représente les mesures d'exécution de la directive. Pour les définir, le CEIOPS organise de vastes consultations auprès des acteurs de marché par le biais des papiers consultatifs (CP) et des études quantitatives (QIS). Le niveau 3 représente les recommandations qui seront élaborées ultérieurement.

La réforme Solvabilité II prévoit une structure à trois piliers de type Bâle II :

- le Pilier I définit des règles quantitatives avec principalement une évaluation prospective des provisions techniques, deux niveaux de besoin en capital (MCR<sup>2</sup> et SCR) et des règles de couverture des engagements ;

<sup>1</sup> Conseil pour les Affaires Économiques et Financières

<sup>2</sup> MCR est le capital minimum requis et le SCR est le capital de solvabilité requis

- 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 transparentes et plus harmonisée au niveau européen en fixant les exigences en matière d'information publique, d'information aux assurés et d'information prudentielle.

Une des grandes nouveautés de Solvabilité II est l'introduction du risque opérationnel qui intervient dans le pilier I avec la quantification du besoin en fonds propres mais également dans le pilier II avec la mise en place de leur gestion.

### **1.3 ENJEUX ET OBJECTIFS POUR LES COMPAGNIES D'ASSURANCE**

Actuellement, le régulateur propose aux compagnies 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 semble vouloir inciter les compagnies à opter pour le modèle interne en rendant l'approche standard beaucoup plus consommatrice en fonds propres.

Les risques opérationnels sont par nature très difficiles à quantifier et l'on pourrait penser que les compagnies préféreront choisir la formule standard qu'investir dans un modèle interne.

Le modèle interne ne doit pas seulement fournir un capital au titre du risque opérationnel, il doit également être utilisé dans la gestion d'entreprise c'est-à-dire que les compagnies doivent prouver qu'elles utilisent leur modèle interne comme un outil de pilotage et de management des risques (« use-test »).

De plus, les compagnies doivent mettre en place des moyens de gestion et de contrôle des risques qui soient implantés dans toute l'entreprise, traduits dans l'organisation, les procédures et les systèmes d'information, documentés, à la base des décisions opérationnelles en préservant la réactivité de l'entreprise.

Il s'agit donc de mettre en place un véritable outil centralisé de quantification et de pilotage des risques, assorti d'un dispositif de contrôle interne performant permettant d'identifier, de contrôler et de gérer ces risques.

## 1.4 OBJECTIFS DU MÉMOIRE

Dans ce cadre, ce mémoire va s'attacher à apporter des solutions concrètes en termes de quantification par modèle interne du capital au titre du risque opérationnel.

Nous commencerons par présenter le risque opérationnel dans le chapitre 2, en donnant des exemples pour que le lecteur comprenne bien le caractère diffus des risques opérationnels dans l'entreprise et donc la difficulté de modélisation.

Ensuite, dans le chapitre 3, nous présenterons un exemple de gestion des risques opérationnels dans le cadre d'une première évaluation.

Dans le but de présenter les modèles, nous introduirons, dans le chapitre 4, les différentes méthodes de quantification existantes : approche standard (formule standard des QIS4 et CP53) et approche avancée (LDA ou Loss Distribution Approach, scénarios et réseau bayésien) et la théorie permettant de les appliquer (calcul de la charge globale). Nous présenterons également dans ce chapitre la théorie des valeurs extrêmes, la mesure de risque et l'agrégation des risques par les copules.

Nous proposerons alors, dans le chapitre 5, la démarche de quantification décomposant les risques opérationnels en deux principales catégories : risques de fréquence modélisés par LDA (avec notamment le problème des seuils de collecte des données) et risques de gravité modélisés par les réseaux bayésiens. Ce modèle sera long à mettre en œuvre pour les compagnies et nécessitera beaucoup d'informations.

Aussi face aux manques de données des compagnies qui commencent à s'intéresser au sujet, nous proposerons, dans le chapitre 6, un modèle alternatif simplifié, basé sur les avis d'experts, nous permettant d'avoir une première vision de nos risques.

Le reste du mémoire sera constitué d'applications.

Dans le chapitre 7, nous modéliserons les risques de fréquence en appliquant la méthode LDA sur des données du secteur bancaire car, étant donné la similarité entre les risques, les données d'une banque nous semblaient appropriées. Nous distinguerons deux méthodes : la méthode LDA classique et la méthode LDA sur des données tronquées. Nous verrons également l'impact des copules pour l'agrégation de ces risques.

Puis, dans le chapitre 8, nous prendrons un exemple de risque de gravité sur lequel nous appliquerons le modèle simplifié et les réseaux bayésiens, nous permettant ainsi de les comparer.

Dans le chapitre 9, nous appliquerons la théorie des valeurs extrêmes sur notre base de données bancaires.

Enfin, dans la dernière partie, nous exposerons les résultats issus des formules standards des QIS4 et du CP53 que nous comparerons aux résultats de notre modèle interne simplifié.



## 2 PRÉSENTATION DU RISQUE OPÉRATIONNEL

Le CEIOPS définit le risque opérationnel comme « le risque de pertes résultant de procédures internes inadaptées ou défaillantes, ou de membre du personnel et de systèmes, ou d'évènements extérieurs. Le risque opérationnel ainsi défini comprend les risques juridiques mais non les risques découlant de décisions stratégiques ni les risques de réputation ».

Le risque opérationnel n'est pas un concept nouveau puisqu'il existe depuis toujours dans les compagnies d'assurance du fait même qu'elles aient des clients, qu'elles utilisent des outils informatiques, qu'elles fassent des placements financiers... Le risque opérationnel est partout dans les entreprises. La nouveauté provient de la mise en place d'outils de mesure et de contrôle des risques opérationnels comme pour les risques de marché, de crédit et d'assurance. Nous citerons plus loin des exemples de risques opérationnels.

Chez Generali, les risques opérationnels se divisent en neuf grandes catégories (voir annexe C pour plus de détails):

- Fraude interne (R1)
- Fraude externe (R2)
- Dysfonctionnement de l'activité et des systèmes (R3)
- Dommages aux actifs corporels (R4)
- Clients / Tiers et produits (R5)
- Exécution et gestion des processus (R6)
- Pratiques en matière d'emploi et sécurité sur le lieu de travail (R7)
- Règlementation et conformité (R8)
- Pilotage (cette catégorie est évoquée à indicatif mais elle est exclue dans la Directive)

Ces risques sont appelés risques de niveau 1.

Chaque risque de niveau 1 est subdivisé en plusieurs sous-risques et ainsi de suite jusqu'au niveau 4. Le niveau 4 représente l'évènement de risque.

Exemple de risque :

Niveau 1	Niveau 2	Niveau 3	Niveau 4
Réglementation et conformité	Marché financier et règles pratiques	Non-conformité	Non-conformité de la société de gestion à son agrément (moyens, organisation, activité ...)

Le risque opérationnel peut également être subdivisé en deux catégories :

- risques maîtrisables : ce sont les pertes provenant de défaillances des processus, des erreurs humaines, de défaillances informatiques... Ce type de risques peut être atténué par anticipation notamment grâce à des contrôles préventifs,
- risques non maîtrisables : ce sont les pertes causées par des événements extérieurs non maîtrisables comme des catastrophes naturelles ou d'autres facteurs non contrôlables par les compagnies d'assurances. C'est un type de risques non négligeable mais non maîtrisé par les assureurs. Par exemple, la crue de la Seine n'est pas maîtrisée par les assureurs, il faut prévoir des plans de replis pour continuer l'activité en cas d'inondation des bureaux. Les compagnies auront souvent recours à des couvertures d'assurances et des plans de continuité d'activité pour prendre en compte ce type de risques.

## **2.1 EXEMPLES DE RISQUE OPÉRATIONNEL FRÉQUENT**

Afin d'illustrer la notion de risque opérationnel, voici quelques exemples de risques fréquemment observés :

- Incidents informatiques qui empêchent momentanément les agents de réaliser des affaires ;
- Erreur dans le paramétrage informatique des commissions ;
- Un partenaire nous quitte avec son portefeuille de clients ;
- Des sachants quittent la compagnie sans avoir transmis leur savoir ;
- Non respect des délais de livrable dans la conception d'un produit ;
- Délais de régularisation des écarts ACAV<sup>3</sup>.

## **2.2 EXEMPLES DE RISQUE OPÉRATIONNEL GRAVE POTENTIEL**

Voici quelques exemples de risques opérationnels susceptibles de survenir :

- un incendie du serveur informatique qui paralyse toute la compagnie ;
- une crue centennale de la Seine qui touche les bureaux de la compagnie (les dégâts faits sur les biens des assurés ne sont pas pris en compte ici mais dans le risque assurantiel) ;
- Une pandémie affecte un grand nombre de salariés de l'entreprise et empêche la continuité des activités.

---

<sup>3</sup> Ajustements correspondant à la moins-value potentielle des placements boursiers sur les contrats UC

### **2.3 EXEMPLES DE RISQUE OPÉRATIONNEL GRAVE AVÉRÉ**

Dans les banques et compagnies d'assurances du monde entier, les plus grosses pertes opérationnelles observées sont liées à des placements financiers risqués ou frauduleux. Ces pertes sont liées à la déficience en matière de contrôle interne : non séparation des pouvoirs, mauvaise supervision des employés ou non respect des process.

L'exemple de la Société Générale : La Société Générale a enregistré en 2008 des pertes énormes résultant de prises de positions frauduleuses d'un de ces traders M. Jérôme Kerviel. Il aurait exposé la banque à un risque de marché alors que ce n'était pas dans ses attributions : il aurait accumulé des positions acheteuses sur les contrats à terme portant sur indice et dissimulé ces opérations faites sur le marché en introduisant dans le système informatique de la Société générale des opérations inverses fictives les compensant. Les causes de ces importantes pertes sont la fraude interne et une défaillance dans les contrôles.

L'exemple du Crédit Lyonnais : Le Crédit Lyonnais a été victime de l'incendie spectaculaire de son siège parisien, le 5 mai 1996. L'incendie couvert par les assureurs, a occasionné des dégâts très importants qui ont coûté plus de 1 milliard de francs.

L'exemple de la banque Barings : l'effondrement de Barings a constitué la faillite la plus spectaculaire au monde car elle était la banque la plus ancienne du Royaume Uni. La cause de cette faillite provient des risques pris par son trader M. Nicolas Leeson. En effet, ce dernier a pris d'importantes positions à découvert sur l'indice Nikkei en dépassant les montants autorisés. Le trader était à la fois responsable du Back office et du trading ce qui lui permettait de faire ce qu'il voulait. La chute de l'indice due au tremblement de terre de Kobe a engendré d'énormes pertes pour Barings, ce qui l'a conduit à la faillite en 1995. Les causes ont été un risque humain volontaire associé à un manque de contrôles et une mauvaise séparation des pouvoirs.

### **2.4 COMPARAISON DU RISQUE OPÉRATIONNEL SOUS SOLVABILITÉ II ET BÂLE II**

Différents contextes réglementaires :

- Dans les banques, il existe une réglementation forte en matière de contrôle interne depuis 1997 (CRBF 97-02 : ce texte a marqué une étape importante dans le fonctionnement de l'audit interne dans la mesure où il introduit les notions de risque et de contrôle continu),

- En assurance, l'introduction des exigences de dispositif de contrôle interne est récente (décret du 13 mars 2006). Les exigences à satisfaire sont moins précises que dans la réglementation Bâle II.

#### Objectifs comparables pour le risque opérationnel :

L'objectif des réglementations bancaire et assurantielle est d'introduire une contrainte de surveillance du risque opérationnel pour éviter la survenance de faillites liées à des déficiences opérationnelles (i.e. Barings, incendie du siège du Crédit Agricole,...).

#### Définition similaire du risque opérationnel :

La définition du risque opérationnel est identique dans Solvabilité II et Bâle II :

- Bâle II : risque de pertes résultant d'une inadaptation ou d'une défaillance imputable à des procédures, personnels et systèmes internes, ou à des événements extérieurs. Inclut les événements de faible probabilité d'occurrence, mais à risque de perte élevée. Le risque opérationnel ainsi défini inclut le risque juridique, mais exclut les risques stratégiques et de réputation (Source : art. 4-1 de l'arrêté transposant Bâle II en France) ;
- Solvabilité II : 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 (Source : directive cadre du 10 juillet 2007, articles 13 et 100).

#### Poids du risque opérationnel dans le capital réglementaire :

- Dans le secteur bancaire, le poids du risque opérationnel est inférieur à celui des autres risques dans les exigences en capital : entre 12% et 20% des exigences en fonds propres pour une banque généraliste.
- Avec Solvabilité II, il semblerait que le poids soit revu à la hausse dans la deuxième vague de papiers consultatifs mais ce n'est pas encore définitif. Il représente pour Generali un des trois plus gros risques.

#### Méthodes de quantification :

Deux méthodes sont proposées dans Solvabilité II contre trois dans Bâle II

- Modèles internes :

- Bâle II : VaR à 99,9% et prise en compte des effets d'assurances (i.e. remboursement des pertes) avec un plafond maximal de 20% des pertes sur base individuelle ;
- Solvabilité II : VaR à 99,5% pour le calcul du SCR (art 100, p3), la possibilité de prendre en compte des couvertures d'assurance n'est pas précisée ;
- Formules standard :
  - Bâle II : Indicateurs de base et standard, construits sur la base du PNB et d'un coefficient multiplicateur de 15% pour la méthode de base et entre 12% et 18% pour la méthode standard ;
  - Solvabilité II : Méthode standard construite sur la base des primes et provisions par activité, avec application d'un plafond à 60% du BSCR.

Nomenclature des lignes de métiers et risques :

- Ligne de métiers et catégories de risques normalisées dans Bâle II (ie matrice 8 x 7),
- mais non précisées dans Solvabilité II.



### **3 GESTION DU RISQUE OPÉRATIONNEL DANS UNE COMPAGNIE D'ASSURANCE**

Dans ce chapitre, nous allons exposer la démarche de gestion des risques opérationnels dans le cadre de la modélisation : le Self Risk Assessment (SRA) et les travaux de cartographie avec leurs objectifs quantitatifs et qualitatifs.

#### **3.1 PRÉSENTATION GÉNÉRALE DE LA DÉMARCHE**

Dans le cadre de la mise en œuvre du dispositif de calcul du capital économique, le Groupe GENERALI a mené des analyses qualitatives et quantitatives de ses risques opérationnels. Les données quantitatives issues de ces analyses permettront à terme d'alimenter le modèle de calcul.

##### 1) Approche « Top down » :

Cette approche consiste à partir d'une vision générale du risque par les dirigeants puis de faire analyser ces risques par des experts.

Les fonctions opérationnelles de Generali sont réparties en 3 directions par marché. Chaque direction a été interrogée pour faire ressortir ses 10 risques opérationnels majeurs.

Finalement, 25 risques ont été retenus et étudiés. L'analyse de chacun de ces risques par les causes, les dispositifs de contrôle et les conséquences associés a permis d'estimer des paramètres de quantification du risque et de réaliser une évaluation de chaque risque. Cette approche est appelée Self Risk Assessment. Cette démarche a été déployée en ateliers de travail avec des experts opérationnels internes désignés par les Comités de Direction.

Cette approche dite « Top down » nous permet d'obtenir dans un premier temps une visibilité sur les principaux risques auxquels est exposée la Compagnie. Elle constitue par ailleurs un vecteur important d'information et de sensibilisation aux enjeux du risque opérationnel.

##### 2) Approche « Bottom-up » :

Depuis Août 2008, Generali mène des analyses plus exhaustives des risques. C'est l'approche par la cartographie des risques dite « Bottom-up ». Elle consiste à analyser de façon exhaustive et systématique les risques opérationnels et les contrôles liés à chaque processus identifié de Generali. C'est un outil majeur du pilotage du risque opérationnel pour le management, les contrôleurs internes et les opérationnels eux-mêmes.

Cette approche repose sur deux éléments :

- Trouver un niveau de granularité approprié. Une maille trop fine impliquerait un chantier trop lourd et la cartographie qui en résulterait serait ingérable au quotidien. Au contraire, une maille trop large ne permettrait pas le pilotage du risque opérationnel. Generali a donc sélectionné un certain niveau de granularité qui sera exposé plus loin.
- S'assurer de l'implication des acteurs de la gestion opérationnelle dans l'élaboration de la cartographie. Les propriétaires des processus et des contrôles, les contrôleurs internes, les managers et les gestionnaires des risques doivent s'impliquer dans la construction dans un premier temps, puis dans la mise à jour de l'outil dans un second temps.

3) Mise en place d'une base d'incidents :

Parallèlement à la cartographie, Generali va mettre en place une base de données historisées des événements de risques opérationnels survenus et des pertes associées.

La constitution de cette base permettra :

- de corroborer ou non les évaluations données par les deux premières approches (basées sur un Self Assessment) et participera au pilotage des risques opérationnels,
- de suivre l'évolution des risques et l'efficacité des contrôles mis en place,
- d'alimenter le modèle interne pour le calcul du capital économique.

### **3.2 PRINCIPE DU SELF RISK ASSESSMENT**

Les risques majeurs sélectionnés par le Comité ont été traités de la façon suivante : pour chaque risque, une analyse de scénarios a été faite selon 3 hypothèses : basse, moyenne et haute. Cette analyse est à la fois qualitative (causes / conséquences / dispositif de contrôle) et quantitative (fréquence / sévérité).

Plus précisément, la démarche méthodologique de Self Risk Assessment des risques opérationnels se décompose en quatre étapes :

Étape 1 : définition du périmètre d'analyse du risque

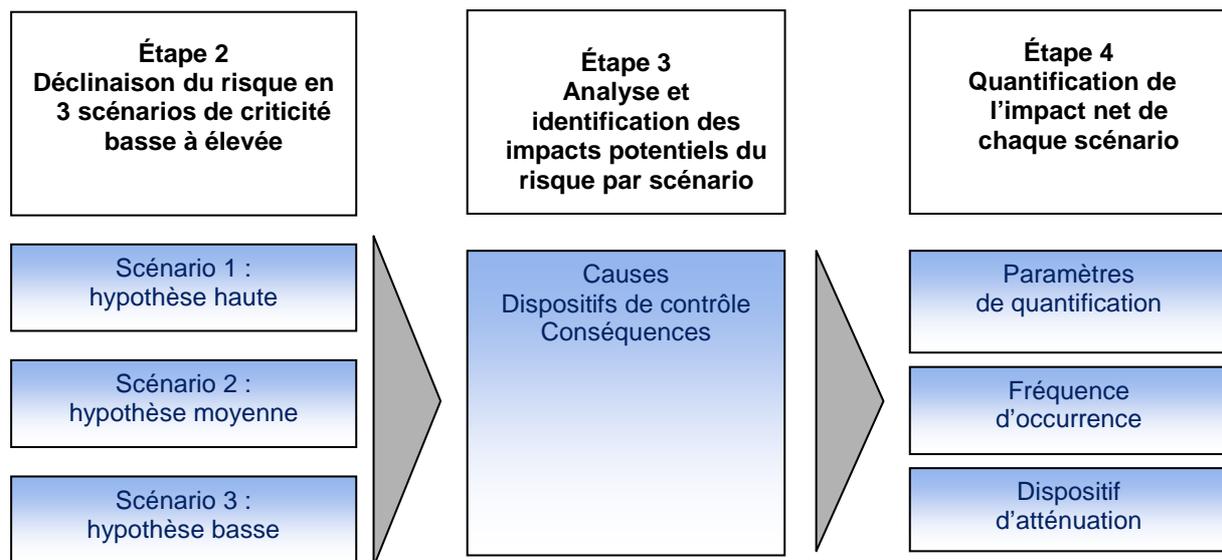
Étape 2 : détermination du contexte et des hypothèses d'évaluation

Étape 3 : identification des facteurs d'analyse qualitative

Étape 4 : identification des paramètres d'analyse quantitative

Étape 1 : l'expert définit et décrit précisément le risque. Cela permet de délimiter le périmètre d'analyse. Il décline tous les contextes potentiels d'occurrence du risque.

Parmi tous les événements de risque identifiés, il sélectionne celui qui lui semble le plus significatif.



Étape 2 : l'expert détermine ensuite des hypothèses d'analyse qui correspondent chacune à un scénario de gravité et de rareté croissantes : l'hypothèse basse est la plus faible en terme d'impact et de fréquence ; l'hypothèse haute correspond à une situation sévère. Ces hypothèses doivent être décrites précisément pour permettre une analyse qualitative et quantitative du risque.

Étape 3 : Analyse qualitative du risque fondée sur l'identification et la description des facteurs suivants : cause du risque, dispositif de contrôle, conséquences.

La cause du risque constitue l'élément générateur direct de l'évènement de risque analysé.

Le dispositif de contrôle désigne tous les contrôles, procédures, systèmes, indicateurs de suivi tant internes qu'externes dont dispose Generali afin de réduire l'occurrence de l'évènement de risque et de limiter les impacts en cas d'occurrence.

La conséquence est un effet consécutif à l'occurrence d'un évènement de risque. Elle peut être financière ou qualitative. Elle peut être directe ou indirecte : on considère qu'elle est indirecte dès lors que les pertes sont consécutives à des actions correctrices menées par l'entité.

Pour aider l'expert dans sa réflexion, le risk manager lui présente des typologies de causes et de conséquences.

Une fois tous ces facteurs de risques identifiés, il s'agit de construire des combinaisons « Causes / Conséquences / Dispositif de contrôle » associées à chaque hypothèse définie afin de permettre la quantification des impacts et fréquence.

Étape 4 : la démarche d'analyse quantitative consiste à évaluer les conséquences identifiées pour chacune des hypothèses définies préalablement et la fréquence d'occurrence associée.

Les conséquences qualitatives sont recensées pour information et pourront faire l'objet de quantification dès lors qu'elles donnent lieu à des coûts indirects.

Pour chaque conséquence quantitative, l'expert devra déterminer une formule précise de calcul de l'impact.

La somme de toutes ces quantifications par hypothèse, donnera l'impact financier global de chaque scénario.

Il est important de prendre en compte pour chaque hypothèse, les diverses mesures d'atténuation et de récupérations afin de déterminer l'impact net.

La détermination de la fréquence d'occurrence de chaque scénario est laissée à l'appréciation des experts en fonction de leur connaissance de l'activité et de leur propre sensibilité au risque sans qu'aucune assistance à la décision ne leur soit proposée.

Voici la liste des risques étudiés dans le cadre du SRA :

1	Non respect des procédures TRACFIN
2	Évolution réglementaire impactant les produits et l'activité
3	Non-respect du taux de PB annoncé
4	Vente forcée, défaut de conseil
5	Abus de marché (délict d'initié)
6	Non respect de la confidentialité
7	Indisponibilité temporaire des systèmes d'information
8	Paramétrage informatique
9	Virus informatique, malveillance
10	Inadéquation des ressources (dimensionnement équipe)
11	Perte de capital humain
12	Détournement de fonds de clients (fraude interne)
13	Fraude de la part des clients
14	Retard dans le traitement des actes de gestion des contrats
15	Défaillance du processus décisionnel (stratégie d'investissement)
16	Non respect des délégations commerciales
17	Mauvaise appréciation du risque à la souscription
18	Actes de vente inadaptés (couple risque/produit)
19	Non respect par un fournisseur de son contrat de prestation
20	Désadossement actif-passif
21	Erreurs dans les comptes consolidés
22	Fiabilité des données comptables (piste d'audit)
23	Risques de planification
24	Erreur d'exécution sur les marchés financiers
25	Sinistres sur immeubles d'exploitation

### 3.3 EXEMPLE DE QUANTIFICATION DE RISQUE DANS LE SELF RISK ASSESSMENT

Nous présentons ici un résumé de fiche technique pour le risque d'erreur dans le passage d'ordre :

Risque d'erreur interne sur le sens, la quantité ou le code valeur à la saisie d'un ordre			
	Hypothèse Basse	Hypothèse Moyenne	Hypothèse Haute
<b>Description</b>	1% de décalage par rapport au marché Moins de 1% de la VL moyenne du portefeuille Durée de détection de l'erreur (entraînant risque d'erreur de VL et de fait un risque d'indemnisation) : quelques heures Montant de l'ordre moyen fourchette basse en portefeuille (5M€)  Périmère : Generali Investments France	2% de décalage par rapport au marché Entre 1% et 2% de la VL moyenne du portefeuille Durée de détection de l'erreur (entraînant risque d'erreur de VL et de fait un risque d'indemnisation) : un jour Montant de l'ordre moyen fourchette moyenne en portefeuille (50M€)  Périmère : Generali Investments France	5% de décalage par rapport au marché Au-delà de 2% de la VL moyenne du portefeuille Durée de détection de l'erreur (entraînant risque d'erreur de VL et de fait un risque d'indemnisation) : plusieurs mois Impact sur les marchés de l'erreur Montant de l'ordre moyen fourchette haute en portefeuille (250M€)  Périmère : Generali Investments France
<b>Fréquence</b>	<b>12 fois tous les ans</b>	<b>6 fois tous les ans</b>	<b>1 fois tous les 5 ans</b>
<b>Conséquences financières (K€)</b>	<b>175</b>	<b>825</b>	<b>15 500</b>
Perte marge technique brute			
Perte financière directe	175	825	15 250
Coûts de communication / formation			
Coûts de rectification de l'erreur			
Gestes co. / Réclamations clts			
Sanctions réglementaires			250
Pertes juridiques			
Autres pertes directes			
<b>Conséquences qualitatives</b>			
Satisfaction client / partenaire			
Assignment			
Réglementaire			Fort
Incidence sur la continuité			
Santé et Sécurité			
Climat social			
Image / Réputation Grand Public			Fort

VL = Valeur liquidative

### 3.4 CARTOGRAPHIE DES RISQUES

L'obligation de suivi et de maîtrise de nos risques passe par l'étape incontournable de la cartographie des risques opérationnels. Cette étape est d'autant plus importante pour l'entreprise qu'elle conditionne les actions de réduction des risques et le système de reporting incidents, qu'elle contribue au dimensionnement du contrôle interne et qu'elle fait partie du modèle interne soumis à la validation de l'autorité de contrôle.

De façon générale, la cartographie consiste à décrire tous les processus et à y rattacher des risques notamment pour permettre au contrôle interne de cibler ses actions et de préparer des plans d'actions.

La cartographie requiert au préalable le recensement de tous les processus existants répertoriés dans un référentiel de processus (voir annexe B) et celui de tous les risques pouvant affecter une compagnie d'assurance, répertoriés dans un référentiel de risques (voir annexe C).

Exemple de processus :

Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Processus métier	Gérer les réseaux	Administrer les réseaux agents, courtiers, apporteurs	Agents	Sélectionner un Apporteur

Une fois ce travail effectué, la cartographie consiste à identifier pour chaque processus les risques opérationnels pouvant entraîner une perte sur le processus et à quantifier les impacts moyens de ces risques.

Les processus seront déclinés par niveau comme les risques. Le référentiel des processus se décompose en cinq niveaux de granularité :

- Niveau 1 : métier (IARD, Vie, etc.)
- Niveau 2 : domaine (habitat, auto, etc.)
- Niveau 3 : processus (souscription, sinistre, etc.)
- Niveau 4 : opération (enregistrement, règlement, etc.)
- Niveau 5 : tâche élémentaire (envoi du chèque, etc.)

Mais à quel niveau de processus la cartographie doit-elle être faite ?

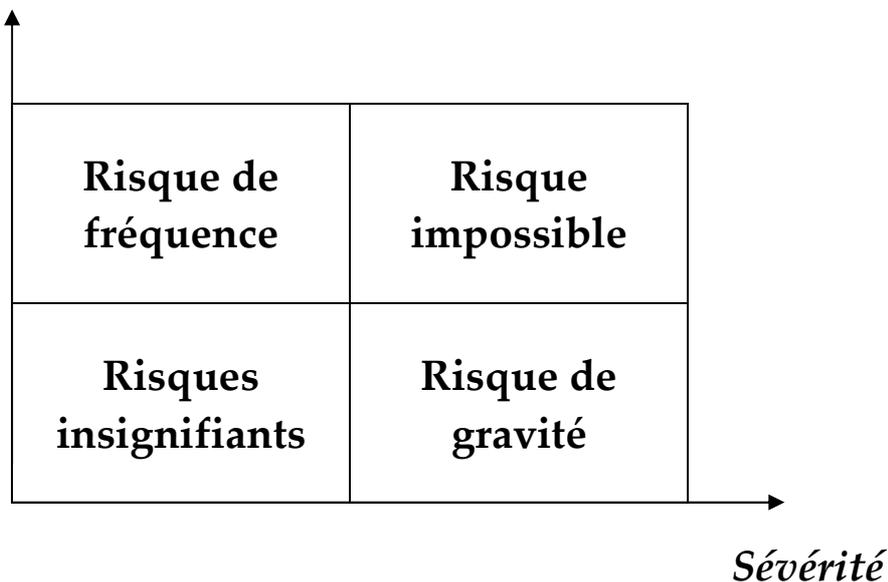
Le niveau médian « processus » semble être le niveau pertinent de la cartographie. Il constitue d'une part le meilleur compromis entre le temps passé à l'élaboration de la cartographie et le degré de pertinence de la vision des risques. D'autre part, c'est le niveau d'équilibre permettant de faire converger et de relier les éléments obtenus à partir de chacune des deux principales méthodes (Top down et Bottom up).

Il semble également que ce niveau soit pertinent dans la construction et le suivi d'une base incidents.

Dans le cadre du calcul du besoin en fonds propres, la cartographie intervient surtout pour mettre en exergue les risques à modéliser. Pour cela, les experts détermineront la fréquence moyenne et la sévérité moyenne de chaque risque qui seront ensuite utilisées pour classer les risques dans la matrice de Prouty (voir [1]).

Voici une représentation théorique de la matrice de Prouty qui croise fréquence moyenne et sévérité moyenne de chaque risque :

*Fréquence*



Il ressort quatre grandes catégories de risque de cette matrice :

- risque faible : fréquence faible, sévérité faible. Ce type de risque est négligeable ;
- risque de fréquence : fréquence forte et sévérité faible. Le sinistre a un coup unitaire faible mais survient souvent. Sa réduction relève essentiellement de la prévention ;
- risque de gravité : fréquence faible et sévérité forte. Le sinistre est rare mais a un coût important voire très important. Sa réduction relève essentiellement de la protection et parfois du plan de continuité d'activité ;
- risque impossible : fréquence forte et sévérité forte. Il doit être hors du champ des possibles pour éviter la disparition rapide de l'entreprise.

Cette classification va nous permettre d'identifier les risques de fréquence et de gravité et de les modéliser indépendamment.



## 4 PRÉSENTATION DES MÉTHODES DE QUANTIFICATION

### 4.1 FORMULE STANDARD

Désormais, Solvabilité II nous impose de détenir du capital au titre du risque opérationnel. Ce montant peut être calculé soit par une formule standard donnée par le régulateur, soit par un modèle interne qui sera validé par les autorités de contrôle.

Pour les entreprises qui auraient du mal à mettre en place un modèle pour le risque opérationnel, la formule standard permet de déterminer facilement le besoin en capital du risque opérationnel.

Le CEIOPS fait tester les formules standard auprès des assureurs du marché par le biais de questionnaires appelés quantitative impact studies (QIS) et fait évoluer les formules en fonction des remarques et difficultés rencontrées par les entreprises.

#### 4.1.1 FORMULE STANDARD ISSUE DES QIS4

La dernière formule standard qui a été testée est celle issue des QIS 4, était de la forme suivante :

$$SCR_{op} = \min \left\{ Op_{load} \cdot BSCR; \max \left\{ \begin{array}{l} 0.03 \cdot Earn_{life} + 0.02 \cdot Earn_{nl} + 0.02 \cdot Earn_h; \\ 0.003 \cdot TP_{life} + 0.02 \cdot TP_{nl} + 0.002 \cdot TP_h \end{array} \right\} \right\} + Op_{unitlinked}$$

$TP_{life}$	=	provisions techniques vie (brutes de réassurance)
$TP_{nl}$	=	provisions techniques non vie (brutes de réassurance)
$TP_h$	=	provisions techniques d'assurance-santé (brutes de réassurance)
$Earn_{life}$	=	primes acquises vie (brutes de réassurance)
$Earn_{nl}$	=	primes acquises non vie (brutes de réassurance)
$Earn_h$	=	primes acquises santé (brutes de réassurance)
$BSCR$	=	SCR avant ajustement
$SCR_{op}$	=	Le montant de capital au titre du risqué opérationnel
$Op_{load}$	=	Coefficient prédéfini (entre 25 et 50%)
$Op_{unitlinked}$	=	Coefficient prédéfini à 1% des provisions techniques pour l'entité liée

Cette formule limite le risque opérationnel par un plafond qui est un pourcentage du SCR. Elle décompose le risque en vie, non-vie et santé d'une part et en primes et provisions d'autre part. Le coefficient attribué aux primes en vie est plus élevé que pour la non vie et la santé car le risque semble plus élevé en vie du fait des placements financiers qui n'existent pas ou peu en non vie et santé. En ce qui concerne les provisions, cette fois, c'est le coefficient non vie qui est le plus important.

Cette formule est trop simpliste pour être réaliste, elle ne prend pas en compte les événements extérieurs ni la façon de gérer les risques : 2 compagnies peuvent avoir le même montant de primes et gérer leurs risques totalement différemment, elle ne tient compte que des primes et des provisions. D'autres facteurs sont très importants pour évaluer ce risque.

Voici le retour sur les QIS4 (source CEIOPS):

Les compagnies qui ont répondu au QIS4 sont partagées sur le fait que la formule standard soit adéquate ou non (47% pensent qu'elle l'est contre 53%). En général, les assureurs non vie ont une opinion plus positive de la charge du risque opérationnel que les assureurs vie.

De nombreuses compagnies ont déclaré que la formule standard avait besoin de quelques améliorations. Les principaux problèmes remontés par les compagnies au sujet de la formule sont :

- La corrélation à 100% avec les autres risques ;
- Le plafond des 30% du SCR est trop élevé : en moyenne, la charge au titre du risque opérationnel tourne autour de 6% du SCR mais dans 8 états certaines grosses compagnies ont dépassé le plafond des 30% ;
- Le manque de sensibilité au risque : la segmentation du risque opérationnel dans la formule standard n'est pas suffisamment fine pour réellement capter les risques.

Certaines compagnies pensent que le risque opérationnel ne peut être appréhendé que via un modèle interne et repose essentiellement sur le pilier 2.

Les propositions de solutions faites par les compagnies :

- Modifier la formule standard en prenant un pourcentage du BSCR ou du SCR ;
- Le capital doit être sensible au management des risques opérationnels ;
- Prendre en compte la diversification des risques.

Il ressort également que l'application de la formule standard à de nombreuses compagnies pourraient conduire certaines compagnies à détenir plus de capital que leur modèle interne ne le suggère et d'autres à n'en pas détenir assez.

D'autre part, le rapport du CEIOPS sur le QIS4 montre les résultats suivants : le modèle interne aurait des résultats bien supérieurs à ceux obtenus par la formule standard avec un ratio médian de 133%.

Le CEIOPS interprète ce résultat par le fait que le capital de la formule standard n'est pas assez élevé. C'est pourquoi le calibrage de la formule standard du SCR a été revu dans la deuxième vague de « consultation papers » (CP). Les nouveaux coefficients ont été calibrés par le CEIOPS à partir de données collectées dans cinq pays européens et 32 compagnies au total de façon à faire converger l'approche formule standard et modèle interne. Cette étude consistait à analyser les ratios des charges issues des modèles internes sur les provisions techniques, primes acquises en vie et non-vie en excluant les unités de comptes. Le CEIOPS a choisi de retenir les coefficients qui correspondaient au quantile à 60%. Pour les unités de compte, le CEIOPS a supposé qu'elles évoluaient de la même façon que la vie et a donc appliqué aux paramètres d'unités de compte la même évolution que les paramètres de la vie.

#### 4.1.2 FORMULE STANDARD ISSUE DU CP53

Il résulte de l'étude du CEIOPS de nouveaux facteurs beaucoup plus pénalisant, le plafond du BSCR a été revu à la hausse en passant de 30% à 60%. Il pénalise également la croissance des entreprises en ajoutant aux provisions et aux primes leurs variations d'une année sur l'autre lorsqu'elles sont supérieures à 10%.

De plus, le CEIOPS a intégré un nouveau terme pour pénaliser la gestion financière déléguée. Ceci pénalise en particulier les compagnies qui auraient une délégation concentrée chez un même dépositaire.

Comparaison des facteurs :

	Facteurs QIS4	Facteurs CP53
Provisions techniques Vie	0,3%	1,0%
Provisions techniques Non Vie	2,0%	4,4%
Primes acquises Vie	3,0%	7,6%
Primes acquises Non Vie	2,0%	4,1%
Facteur Unité de compte	25,0%	50,0%
Plafond BSCR	30,0%	60,0%
Dépositaire		0,5%

Voici la nouvelle formule issue des derniers CP :

$$SCR_{op} = \text{Min} \{ 0,60 * BSCR; Op_{all\_non\_ul} \} + 0,5 * Exp_{ul} + 0,005 * Invest_{Outsourc}$$

$$\text{Où } Op_{all\_non\_unit\_linked} = \text{Max}(Op_{premiums}; Op_{provisions})$$

Avec

$$Op_{premiums} = 0,076 * (Earn_{life} + Earn_{SLTHealth} - Earn_{life-ul}) +$$

$$0,041 * (Earn_{non-life} + Earn_{NonSLTHealth}) +$$

$$Max(0; 0,076 * (\Delta Earn_{life} - \Delta Earn_{life-ul})) +$$

$$Max(0; 0,041 * \Delta Earn_{non-life})$$

Dans le cas où

$$Op_{provisions} = 0,009 * (TP_{life} + TP_{SLTHealth} - TP_{life-ul}) +$$

$$0,044 * (TP_{non-life} + TP_{NonSLTHealth}) +$$

$$Max(0; 0,009 * (\Delta TP_{life} - \Delta TP_{life-ul})) +$$

$$Max(0; 0,044 * \Delta TP_{non-life})$$

Dans le cas où

$$Op_{provisions} = 0,01 * (TP_{life} + TP_{SLTHealth} - TP_{life-ul}) +$$

$$0,044 * (TP_{non-life} + TP_{NonSLTHealth}) +$$

$$Max(0; 0,01 * (\Delta TP_{life} - \Delta TP_{life-ul})) +$$

$$Max(0; 0,044 * \Delta TP_{non-life})$$

$\Delta TP$	=	Variation dans les provisions entre t-1 et t
$\Delta Earn$	=	Variation dans les primes acquises entre t-1 et t
$TP_{life}$	=	provisions techniques vie (brutes de réassurance)
$TP_{nl}$	=	provisions techniques non vie (brutes de réassurance)
$TP_h$	=	provisions techniques d'assurance-santé (brutes de réassurance)
$Earn_{life}$	=	primes acquises vie (brutes de réassurance)
$Earn_{nl}$	=	primes acquises non vie (brutes de réassurance)
$Earn_h$	=	primes acquises santé (brutes de réassurance)
BSCR	=	SCR avant ajustement
$SCR_{op}$	=	Le montant de capital au titre du risqué opérationnel
$Op_{unitlinked}$	=	Coefficient prédéfini à 1% des provisions techniques pour l'entité liée
$Exp_{ul}$	=	Montant des dépenses annuelles (brutes de réassurance)
$Invest_{Outsourc}$	=	Montant correspondant au plus important investissement financier déposé ou externalisé avec une seule contrepartie

## 4.2 MÉTHODE DE MESURES AVANCÉES

Les entreprises qui le souhaitent, ont la possibilité de mettre en place un modèle interne en utilisant des méthodes de mesures avancées. Ces mesures s'avèrent très intéressantes pour les entreprises car elles permettent une analyse plus précise de leur risque et une éventuelle réduction de l'allocation en fonds propres mais pas toujours.

Actuellement, trois méthodes sont principalement utilisées : la méthode LDA, l'approche par scénario et l'approche Bayésienne présentées en détail plus bas.

### 4.2.1 MÉTHODE LDA (LOSS DISTRIBUTION APPROACH)

C'est l'une des approches les plus utilisées dans le secteur bancaire pour la quantification du risque opérationnel : il s'agit d'ajuster des lois statistiques à des données de pertes, plus précisément de modéliser d'une part la fréquence des incidents opérationnels et d'autre part leur sévérité, puis de les combiner pour obtenir la distribution de pertes totales. Cette approche est bien connue des assureurs qui l'utilisent depuis longtemps pour modéliser la charge totale des sinistres dans le cadre de la tarification, du provisionnement, de l'optimisation de la réassurance...

La condition essentielle pour appliquer cette méthode sera la disponibilité de données de pertes historiques afin de calibrer le modèle.

Étape 1 : choix et calibrage des lois de fréquence et de sévérité

La modélisation de la distribution des pertes se fera pour chaque croisement de risque  $k$  et de processus  $i$  :

$$S_{i,k} = \sum_{j=1}^{N_{i,k}} X_{i,k}^{(j)}$$

où  $N_{i,k}$  est la variable aléatoire représentant le nombre de pertes pour le croisement  $(i,k)$

$X_{i,k}^{(j)}$  est la variable aléatoire représentant le coût de la perte  $j$  pour le croisement  $(i,k)$

$S_{i,k}$  est la somme des pertes du croisement  $(i,k)$

La démarche doit être reproduite pour chaque croisement  $(i,k)$ . Elle consiste à ajuster des lois de probabilité sur les données de fréquence des événements de pertes d'une part et les données de sévérité de ces mêmes événements d'autre part, sur un intervalle de temps donné.

Pour ce faire il faut trier les événements de pertes par fréquence d'une part, et par coût d'autre part, puis représenter le résultat sous forme graphique. Pour chacune des distributions empiriques obtenues, il s'agit de rechercher la distribution

théorique qui rend le mieux compte de la forme de la courbe. Les principales lois utilisées pour la fréquence sont Poisson, Binomiale et Binomiale Négative.

Les deux principales méthodes utilisées pour calibrer les paramètres de lois testées sont :

- la méthode des moments généralisés qui est relativement facile à implémenter mais dont les résultats sont de moindre qualité ; la méthode des moments est une méthode statistique intuitive qui consiste à estimer les paramètres d'une loi en égalisant certains moments théoriques avec les moments empiriques, basée sur la loi des grands nombres.
- la méthode du maximum de vraisemblance qui est plus difficile à mettre en œuvre mais dont les résultats sont de meilleure qualité. La méthode du maximum de vraisemblance est une méthode statistique courante permettant d'estimer les paramètres d'une loi en maximisant sa fonction de vraisemblance :

Soit  $f(x; \theta)$  la densité de la loi étudiée dont l'ensemble des paramètres est noté  $\theta$

Soit  $(x_1, x_2, \dots, x_n)$  un échantillon donné

La fonction de vraisemblance est définie par :

$$L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

$$\text{Soit : } \ln(L(x_1, x_2, \dots, x_n; \theta)) = \sum_{i=1}^n \ln(f(x_i; \theta))$$

Il s'agit ensuite de résoudre l'équation :

$$\frac{\delta \ln(L(x_1, x_2, \dots, x_n; \theta))}{\delta \theta} = 0$$

Il y a un quasi consensus de place pour utiliser la méthode du Maximum de vraisemblance afin d'estimer les paramètres des lois.

Pour valider le choix d'un modèle mathématique, le premier test d'adéquation est un test graphique qui consiste à placer sur un même graphique les données réelles observées et les réalisations simulées par le modèle. Si les deux courbes se superposent, le modèle semble fiable mais il est essentiel de valider cette impression visuelle par des calculs. En effet, les tests statistiques sont plus précis et permettent de valider le choix des modèles. Pour cela, nous disposons des tests de Kolmogorov-Smirnov, Anderson-Darling, Cramer Von Mises et du Khi-Deux.

Remarque : nous exposons ici la méthode LDA en supposant qu'il n'y aura pas de seuil de collectes, ce qui n'est pas très réaliste. Le seuil de collecte va induire un biais impactant principalement le corps de la distribution puisque de manière générale, les

incidents de faible intensité qui ne seront pas remontés sont les plus fréquents et constituent le corps de la distribution.

Les travaux de la place sont encore en cours pour compenser ce biais mais la technique principalement utilisée jusqu'à présent est la prise en compte du seuil suivant une logique de probabilité conditionnelle qui sera étudiée dans la suite du mémoire.

Etape 2 : Construction de la distribution de pertes totales

Remarque : Cette modélisation ne conduit à une expression analytique de la fonction de répartition  $F_S$  de  $S$  que dans le cas où  $N$  suit une loi géométrique et  $X$  suit une loi exponentielle. Cela nous permettra de faire des comparaisons entre la vraie distribution et celle qui est estimée pour apprécier la qualité des estimations.

Il est possible d'obtenir des approximations de  $F_S$  avec les méthodes suivantes que nous détaillerons plus loin :

- algorithme de Panjer qui donne même les valeurs exactes dans le cas où  $X$  est arithmétique
- Inversion numérique de la fonction caractéristique de  $S$
- Utilisation des premiers moments de  $S$  et d'une distribution approximante, l'approximation Normal Power
- Simulation de Monte-Carlo.

Pour chaque croisement risque processus, nous obtenons une courbe agrégée de distribution des pertes pour un horizon de temps donné (dans Solvabilité II il s'agit d'une année).

Remarques :

Il semblerait plus logique d'appliquer cette méthode uniquement aux risques de gravité faible et de fréquence élevée car les pertes observées seraient en nombre suffisant pour que la modélisation soit valable. C'est ce que nous verrons plus loin.

Cette méthode permet de simuler indépendamment tous les risques et de les agréger en additionnant les distributions empiriques. Cela suppose l'indépendance entre tous les risques opérationnels ce qui n'est vraisemblablement pas le cas.

Dans le cadre de la réglementation Bâle II, le capital requis pour ce risque est la VaR calculée avec une probabilité de 99,9%. Les banquiers utilisent plusieurs méthodes pour agréger leurs risques : utilisation de matrice de corrélation, des copules, simple somme des VaR...). Dans le cadre de Solvabilité II, la méthode d'agrégation des risques n'est pas imposée.

#### **4.2.2 APPROCHE PAR SCÉNARIOS**

L'approche par scénarios consiste à réaliser dans un premier temps une analyse qualitative des risques avec les experts métiers, qui vont construire des scénarios pour chaque risque afin de déterminer les fréquences et sévérités moyennes ainsi que la sévérité extrême du risque. La construction des scénarios s'apparente à l'approche SRA décrite plus haut. Les valeurs seront ensuite utilisées pour estimer les paramètres de lois choisies a priori qui pourront donner lieu à des simulations de Monte-Carlo pour obtenir la charge totale de chaque risque. Nous verrons par la suite des applications de cette approche.

Cette approche constitue un complément intéressant quand les données historiques ne sont pas suffisantes pour appliquer une méthode purement statistique. Elle trouve en particulier son application dans l'évaluation des impacts d'événements de risque de sévère amplitude.

#### **4.2.3 APPROCHE BAYÉSIENNE**

La troisième approche la plus utilisée est l'approche bayésienne.

L'approche bayésienne consiste à réaliser une analyse qualitative des risques à dire d'experts et à la transformer en une analyse quantitative. Cette méthode s'apparente à la méthode par scénarios dans la mesure où elles se composent toutes deux d'une première phase d'analyse qualitative et de constitution de scénarios.

L'approche bayésienne permet d'introduire de la variabilité dans les scénarios étudiés et de simuler les pertes sans nécessiter de base historisée. Elle permet également d'identifier des leviers de réduction de risque dans la mesure où les pertes sont conditionnées à des indicateurs qui peuvent être suivis et modifiés pour réduire le risque. Et enfin elle a l'avantage d'éviter les problèmes de diversification par la construction même du réseau comme nous le verrons plus loin.

L'avantage de l'approche bayésienne est aussi qu'elle permet de modéliser des risques exogènes pour lesquels nous n'avons presque pas de données (par exemple la crue centennale).

Un réseau bayésien est un graphe causal probabiliste représentant la structure de la connaissance d'un certain domaine. Il est constitué de variables aléatoires discrètes reliées par des arcs orientés, ces variables étant appelées des nœuds. Une distribution est rattachée à chaque nœud. Les arcs sont des liens qui représentent une dépendance causale.

Une fois le réseau bayésien construit, il est possible de calculer la distribution d'un nœud conditionnellement à une observation effectuée sur un autre nœud, il s'agit de l'inférence. Cela pourra nous permettre entre autres de détecter les variables leviers, les variables sur lesquelles les contrôleurs internes devront porter plus d'attention.

Nous avons étudié en particulier la méthode « XSG » (exposition, survenance, gravité) qui consiste à définir et modéliser les trois grandeurs caractéristiques du risque à savoir l'exposition, la survenance et la gravité (voir [20]). Ces trois grandeurs sont influencées par des variables appelées indicateurs de risque (ou KRI<sup>4</sup>).

Nous allons présenter les méthodes pour évaluer les différents éléments du réseau bayésien.

#### a) Evaluation de l'exposition

L'exposition est l'ensemble des éléments de l'entreprise qui sont exposés au risque. Elle doit être définie de sorte que le risque ne puisse survenir qu'une seule fois au plus dans l'année.

La quantification du risque opérationnel doit être faite sur l'année à venir, le nombre d'objets exposés au risque doit donc être anticipé pour l'exercice suivant.

En général, l'exposition prévue pour l'année à venir est l'exposition constatée multipliée par un coefficient d'augmentation.

#### b) Evaluation de la survenance

L'objet exposé étant choisi de sorte qu'il ne puisse être frappé qu'au plus une seule fois, la survenance sera par construction une loi Binomiale  $B(n,p)$  où  $n$  est le nombre d'objets exposés et  $p$  la probabilité que nous aurons à estimer.

Il existe plusieurs méthodes pour évaluer cette probabilité :

- Méthode empirique : elle consiste à rapporter le nombre de sinistres constatés au nombre d'objets exposés dans le passé. Si aucun sinistre n'est survenu pendant la période de référence alors on supposera à titre prudentiel qu'un sinistre est survenu.
- Méthode externe : c'est une évaluation externe pour les risques non spécifiques à la compagnie.
- Méthode subjective : à dire d'experts.
- Méthode combinée : il s'agit de combiner l'avis d'expert au résultat empirique.

#### c) Evaluation de la gravité

Il faut se placer dans la situation où la survenance de la perte est avérée et identifier les variables quantifiables (KRI) intervenant dans le calcul de la gravité.

#### d) Evaluation des KRI :

Il existe plusieurs méthodes pour évaluer les KRI selon les informations dont on dispose (nous privilégions toujours les évaluations basées sur des observations)

L'évaluation à dire d'experts se fait autant que possible sous la forme de trois hypothèses : hypothèse basse (cas favorable : la probabilité rattachée à cette

---

<sup>4</sup> KRI = Key Risk Indicator

hypothèse est fixée a priori à 25%), hypothèse moyenne (la probabilité rattachée à cette hypothèse est fixée a priori à 50%), hypothèse haute (cas défavorable : la probabilité rattachée à cette hypothèse est fixée a priori à 25%).

L'évaluation empirique des KRI consiste à construire une distribution empirique sur la base de données observées.

La structure du réseau bayésien sera définie par les experts à travers des scénarios. Les paramètres du réseau bayésien pourront être déterminés de façon empirique ou à dire d'experts. Ce point sera développé en détail plus bas.

Une fois le réseau bayésien construit, il reste à définir l'algorithme de calcul.

Soient  $(X_1, \dots, X_n)$  les objets exposés au risque opérationnel étudié.

Soit  $P_i = P(\text{Exposition} = X_i)$  la probabilité que l'exposition soit les objets  $X_i$ .

Soit  $PS_i = P(\text{Survenance} = \text{"Oui"} \mid \text{Exposition} = X_i)$  la probabilité que le risque survienne sachant que l'exposition est  $X_i$ .

Ces deux probabilités sont connues (elles ont été estimées comme vu précédemment).

Soit  $PG_i = P(\text{Gravité} \mid \text{Survenance} = \text{"Oui"} \text{ et } \text{Exposition} = X_i)$  la distribution de la gravité sachant que le risque s'est produit sur les objets  $X_i$ .

L'algorithme consiste à réaliser successivement les étapes suivantes :

- 1) positionner l'exposition à  $X_i$ , la survenance à Oui dans le réseau bayésien et lire la distribution de la gravité  $PG_i = P(\text{Gravité} \mid \text{Survenance} = \text{"Oui"} \text{ et } \text{Exposition} = X_i)$ .
- 2) échantillonner le nombre de pertes  $F_i$  selon la loi binomiale  $B(nb(X_i); PS_i)$ .
- 3) pour chacun des incidents de 1 à  $F_i$ , échantillonner la gravité suivant la distribution  $PG_i$ .
- 4) sommer les  $F_i$  gravités.

Répéter ces 4 étapes un grand nombre de fois en conservant les sommes des gravités à chaque fois. Nous obtenons ainsi une distribution de pertes totales.

### 4.3 MÉTHODES DE CALCUL DE LA CHARGE TOTALE

Nous décrivons ici les différentes méthodes permettant d'obtenir la distribution de la charge totale d'un risque opérationnel, sachant que nous avons ajusté au préalable des lois de fréquence et de coût.

#### 4.3.1 FORMULE RÉCURSIVE DE PANJER

La formule de Panjer est une formule de récurrence qui fournit la vraie distribution de charge totale si les lois de coûts individuels et de fréquence sont connues, à valeurs entières et bornées. Sinon ce sera une approximation.

Plusieurs conditions sur la loi des coûts et la loi des fréquences doivent être remplies pour appliquer la méthode de Panjer. Il y a notamment une contrainte sur la forme de la loi des coûts individuels qui doit être à valeurs entières et bornées et une contrainte sur la forme de la loi de fréquence. Voici ces conditions et comment adapter nos lois à ces conditions :

1) Rendre la loi du coût individuel à valeurs entières et bornées :

Pour pouvoir appliquer l'algorithme de Panjer, il est nécessaire de transformer la loi des coûts individuels qui est en général continue en une loi à valeurs entières et bornées.

Pour cela, nous avons utilisé la méthode des arrondis que nous allons vous présenter ci-dessous.

Méthode des arrondis :

Cette méthode s'applique à une loi continue pour la rendre à valeurs entières bornées.

Soit  $F$  la fonction de répartition du coût individuel  $X$  que nous avons déterminé au préalable.

Soit  $K$  la fonction de répartition du coût individuel « arithmétisé »  $\tilde{X}$  que nous cherchons à approcher.

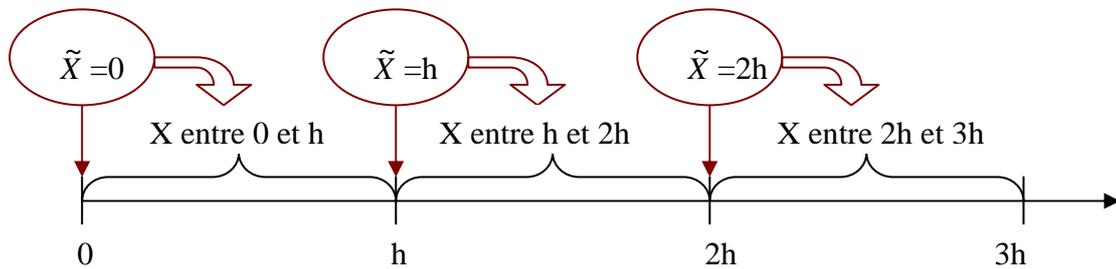
En particulier, nous allons chercher à estimer les probabilités  $P(\tilde{X} = jh)$  pour  $j=0, 1, \dots, n$  et  $h>0$  le pas choisi (pour appliquer Panjer nous prendrons  $h=1$ ).

La méthode des arrondis consiste à estimer ces probabilités selon les trois approches suivantes :

➤ Arrondi à l'unité inférieure :  $P_A$

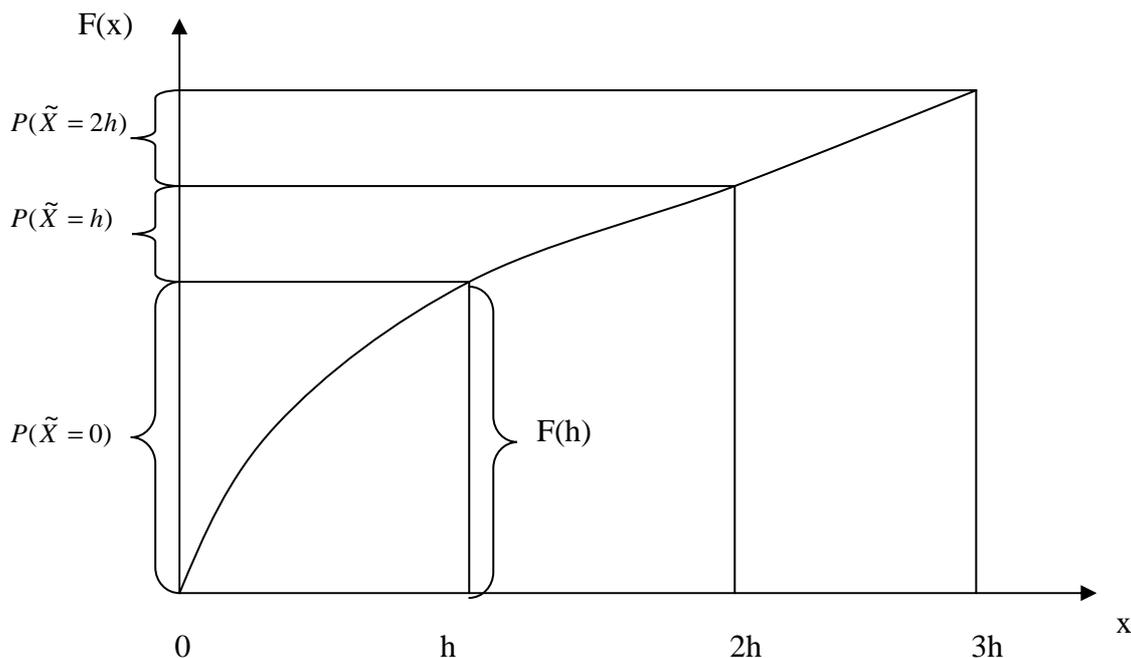
Lorsque  $\tilde{X}$  prend la valeur 0, nous dirons que cela correspond à «  $X$  prend ses valeurs entre 0 et  $h$ , la valeur de  $\tilde{X}$  est la valeur de la borne inférieure de l'intervalle  $[0 ; h]$  » :  $P_A(\tilde{X} = 0) = P(0 \leq X \leq h) = F(h)$

De même, lorsque  $\tilde{X}$  prend la valeur  $h$ , nous dirons que cela correspond à «  $X$  prend ses valeurs entre  $h$  et  $2h$ , la valeur de  $\tilde{X}$  est la valeur de la borne inférieure de l'intervalle  $[h ; 2h]$  » :  $P_A(\tilde{X} = h) = P(h \leq X \leq 2h) = F(2h) - F(h)$



En généralisant : lorsque  $\tilde{X}$  prend la valeur  $jh$ , nous dirons que cela correspond à «  $X$  prend ses valeurs entre  $jh$  et  $(j+1)h$ , la valeur de  $\tilde{X}$  est la valeur de la borne inférieure de l'intervalle  $[jh ; (j+1)h]$  » :

$$P_A(\tilde{X} = jh) = P(jh \leq X \leq (j+1)h) = F((j+1)h) - F(jh) \quad \text{pour } j \in \mathbb{N}$$

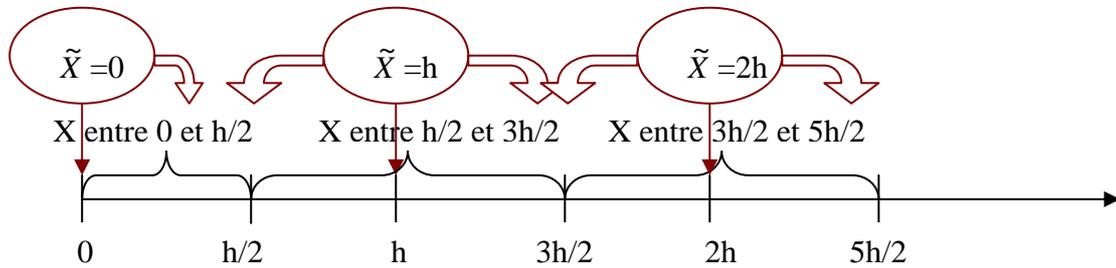


➤ Arrondi à l'unité la plus proche :

Lorsque  $\tilde{X}$  prend la valeur  $0$ , nous dirons que cela correspond à «  $X$  prend ses valeurs entre  $0$  et  $h/2$  » :  $P_B(\tilde{X} = 0) = P\left(0 \leq X \leq \frac{h}{2}\right) = F\left(\frac{h}{2}\right)$

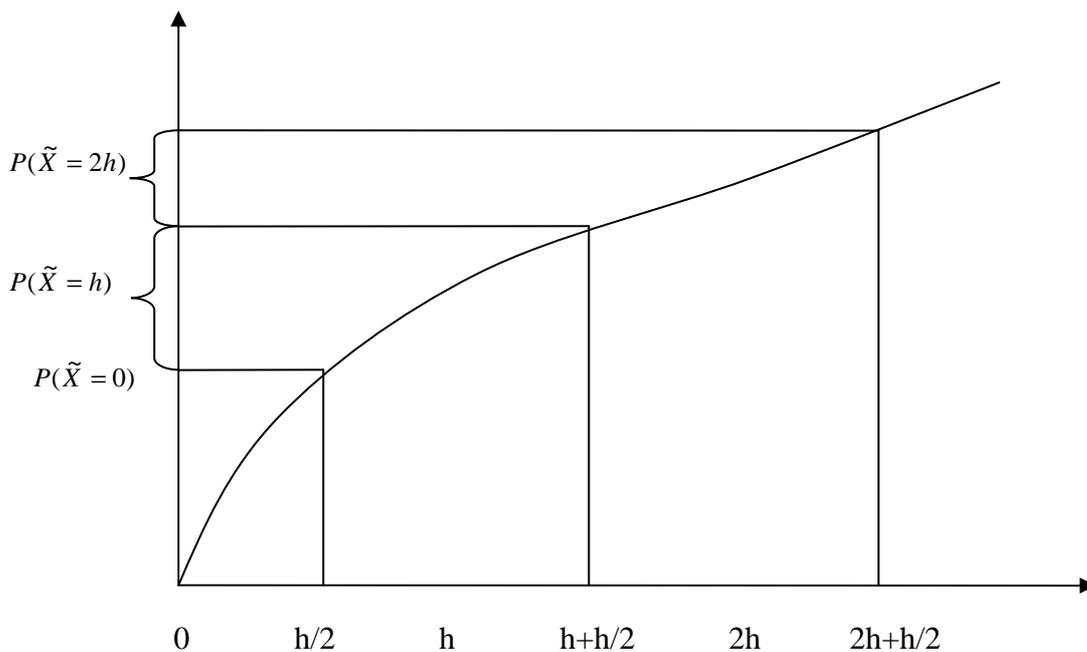
De même, lorsque  $\tilde{X}$  prend la valeur  $h$ , nous dirons que cela correspond à «  $X$  prend ces valeurs entre  $h/2$  et  $h+h/2$ , la valeur de  $\tilde{X}$  est le milieu de l'intervalle  $[h/2 ; 3h/2]$  »:

$$P_B(\tilde{X} = h) = P\left(\frac{h}{2} \leq X \leq \frac{3h}{2}\right) = F\left(\frac{3h}{2}\right) - F\left(\frac{h}{2}\right)$$



En généralisant : lorsque  $\tilde{X}$  prend la valeur  $jh$ , nous dirons que cela correspond à «  $X$  prend ces valeurs entre  $jh-h/2$  et  $jh+h/2$ , la valeur de  $\tilde{X}$  est le milieu de l'intervalle  $[jh-h/2 ; jh+h/2]$  »:

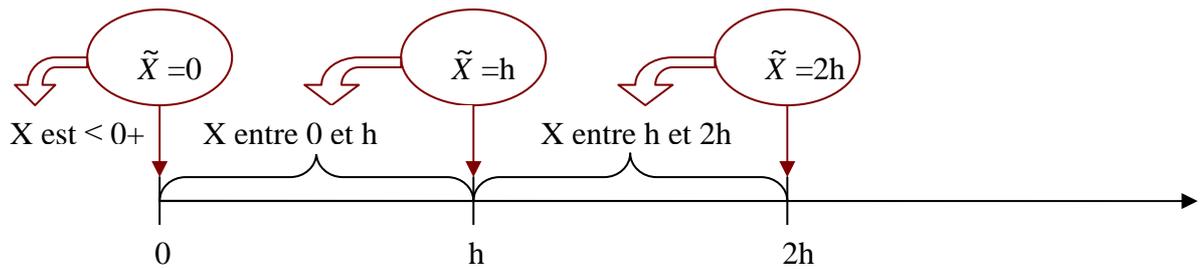
$$P_B(\tilde{X} = jh) = P\left(jh - \frac{h}{2} \leq X \leq jh + \frac{h}{2}\right) = F\left(jh + \frac{h}{2}\right) - F\left(jh - \frac{h}{2}\right) \quad \text{pour } j \in \mathbb{N}^*$$



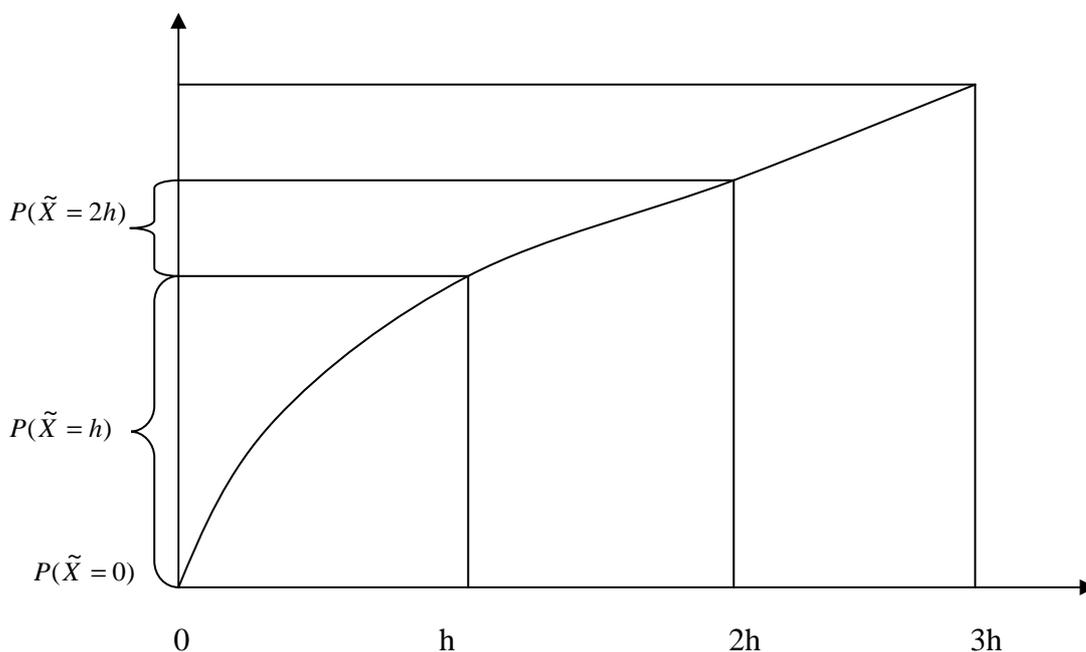
➤ Arrondi à l'unité supérieure :

Lorsque  $\tilde{X}$  prend la valeur 0, nous dirons que cela correspond à «  $X$  prend ces valeurs entre 0 et  $0+$  »:  $P_C(\tilde{X} = 0) = P(X \leq 0+) = F(0^+)$

De même, lorsque  $\tilde{X}$  prend la valeur  $h$ , nous dirons que cela correspond à «  $X$  prend ces valeurs entre 0 et  $h$ , la valeur de  $\tilde{X}$  est la valeur de la borne inférieure de l'intervalle  $[0 ; h]$  »:  $P_C(\tilde{X} = h) = F(h^+) - F(0^+)$



En généralisant : lorsque  $\tilde{X}$  prend la valeur  $jh$ , nous dirons que cela correspond à «  $X$  prend ces valeurs entre  $jh-h$  et  $jh$ , la valeur de  $\tilde{X}$  est le milieu de l'intervalle  $[jh-h ; jh]$  » :  $P_c(\tilde{X} = jh) = P(jh-h \leq X \leq jh) = F(jh^+) - F((j-1)h^+)$  pour  $j \in \mathbb{N}^*$



L'arrondi à l'unité la plus proche donne une bonne approximation de la distribution de la charge totale et les deux autres arrondis forment un encadrement de la distribution.

Par exemple pour « arithmétiser » la loi de Weibull avec un pas de 1, on aurait :

$$P_B(\tilde{X} = 0) = F\left(\frac{1}{2}\right) = 1 - \exp\left(-\left(\frac{1/2}{b}\right)^a\right) \quad (\text{a et b sont définis page 92})$$

$$P_B(\tilde{X} = j) = F\left(j + \frac{1}{2}\right) - F\left(j - \frac{1}{2}\right) = \exp\left(-\left(\frac{j-1/2}{b}\right)^a\right) - \exp\left(-\left(\frac{j+1/2}{b}\right)^a\right) \quad \text{pour } j \in \mathbb{N}^*$$

2) Vérification de la forme de la loi de fréquence :

Une fois la loi du coût est à valeurs entières et bornées, nous devons nous assurer que la loi de la fréquence  $(p_n)_{n \geq 0}$  utilisée vérifie bien :

$$\exists a, b \in \mathfrak{R} \text{ tels que } p_n = \left( a + \frac{b}{n} \right) p_{n-1} \text{ pour } n \geq 1$$

Cette propriété est nécessaire à l'application de la formule de Panjer (voir démonstration [21])

Les distributions Binomiale, Poisson et Binomiale Négative sont les seules lois à vérifier cette propriété avec les a et b indiqués dans le tableau suivant :

Distributions	a	b
$B(\alpha, p)$	$-\frac{p}{q}$	$(\alpha + 1)\frac{p}{q}$
$P(\lambda)$	0	$\lambda$
$BN(r, p)$	$q$	$(r - 1)q$

3-a) Formule récursive de Panjer dans le cas où X à valeurs dans  $\mathbb{N}^*$ :

La formule de Panjer nous permet de reconstituer la distribution de S discrétisée à partir de  $P(S=0) = P(N=0)$  (car toute perte a un coût  $>0$ ) :

$$P(S = s) = \sum_{k=1}^s \left( a + b \frac{k}{s} \right) P(X = k) P(S = s - k) \text{ pour } s \geq 1$$

3-b) Formule récursive de Panjer dans le cas où X à valeurs dans  $\mathbb{N}$ :

La formule s'étend au cas où X est à valeurs dans  $\mathbb{N}$  :

$$P(S = 0) = \sum_{k=0}^{+\infty} P(N = k) P^k(X = 0)$$

$$P(S = s) = \frac{\sum_{k=1}^s \left( a + b \frac{k}{s} \right) P(X = k) P(S = s - k)}{1 - aP(X = 0)} \text{ pour } s \geq 1$$

#### 4.3.2 INVERSION DE LA FONCTION CARACTÉRISTIQUE

Cette méthode consiste à obtenir la fonction caractéristique  $\varphi_s$  de la charge totale S puis de l'inverser par la transformée inverse de Fourier pour obtenir la loi de S.

La fonction caractéristique  $\varphi_S$  de la charge totale S peut être déterminée à partir de la fonction génératrice des probabilités  $g_N$  de la loi de fréquence et la fonction caractéristique  $\varphi_X$  de la loi du coût individuel :

$$\varphi_S(t) = g_N[\varphi_X(t)] \quad \text{pour } t \in \mathfrak{R}$$

La transformation de Fourier fonctionne avec des variables aléatoires à valeurs dans  $k\mathbb{N}$ , ainsi il va falloir « arithmétiser » la loi de X et par conséquent celle de S.

L'inverse de la fonction caractéristique repose sur la propriété suivante :

Si Y est une variable aléatoire arithmétique à valeurs dans  $h\mathbb{N}$  avec  $h > 0$  fixé, on a

$$P(Y = kh) = \frac{1}{2\pi} \int_0^{2\pi} e^{-iik} \varphi_Y\left(\frac{t}{h}\right) dt \quad \text{pour } k \in \mathbb{N}$$

Cette quantité peut ensuite être estimée grâce aux transformées de Fourier, comme nous allons le montrer ici :

L'intervalle  $[0 ; 2\pi]$  est découpé en  $m$  sous intervalles de même longueur égale à  $\frac{2\pi}{m}$ .

L'expression peut alors s'écrire

$$P(Y = kh) = \frac{1}{2\pi} \sum_{j=0}^{m-1} \int_{\frac{2\pi}{m}j}^{\frac{2\pi}{m}(j+1)} e^{-iik} \varphi_Y\left(\frac{t}{h}\right) dt \quad \text{pour } k \in \mathbb{N}$$

L'idée est alors d'approximer les intégrales sur les sous intervalles par des aires de rectangles pour tomber sur des transformées de Fourier :

$$\int_{\frac{2\pi}{m}j}^{\frac{2\pi}{m}(j+1)} e^{-iik} \varphi_Y\left(\frac{t}{h}\right) dt \approx \frac{2\pi}{m} \times e^{-\frac{2\pi}{m}jik} \varphi_Y\left(\frac{2\pi j}{mh}\right) \quad \text{pour } j, k = 0, 1, \dots, m-1$$

$$\text{D'où } P(Y = kh) \approx \frac{1}{m} \sum_{j=0}^{m-1} e^{-\frac{2\pi}{m}jik} \varphi_Y\left(\frac{2\pi j}{mh}\right) \quad \text{pour } k \in \mathbb{N}$$

Les différentes étapes de calcul sont :

- Arithmétiser la loi du coût individuel (vu précédemment)
- Calculer la fonction caractéristique du X arithmétisé
- Calculer  $\varphi_S$  en appliquant la fonction génératrice de N à  $\varphi_X$  selon la propriété vue plus haut
- Calculer la loi de S en appliquant la transformée de Fourier inverse.

### 4.3.3 MÉTHODE BASÉE SUR LES MOMENTS

Cette méthode repose essentiellement sur la loi Normale et les moments de  $S$  qui s'obtiennent facilement à partir des moments de  $N$  et de  $X$ .

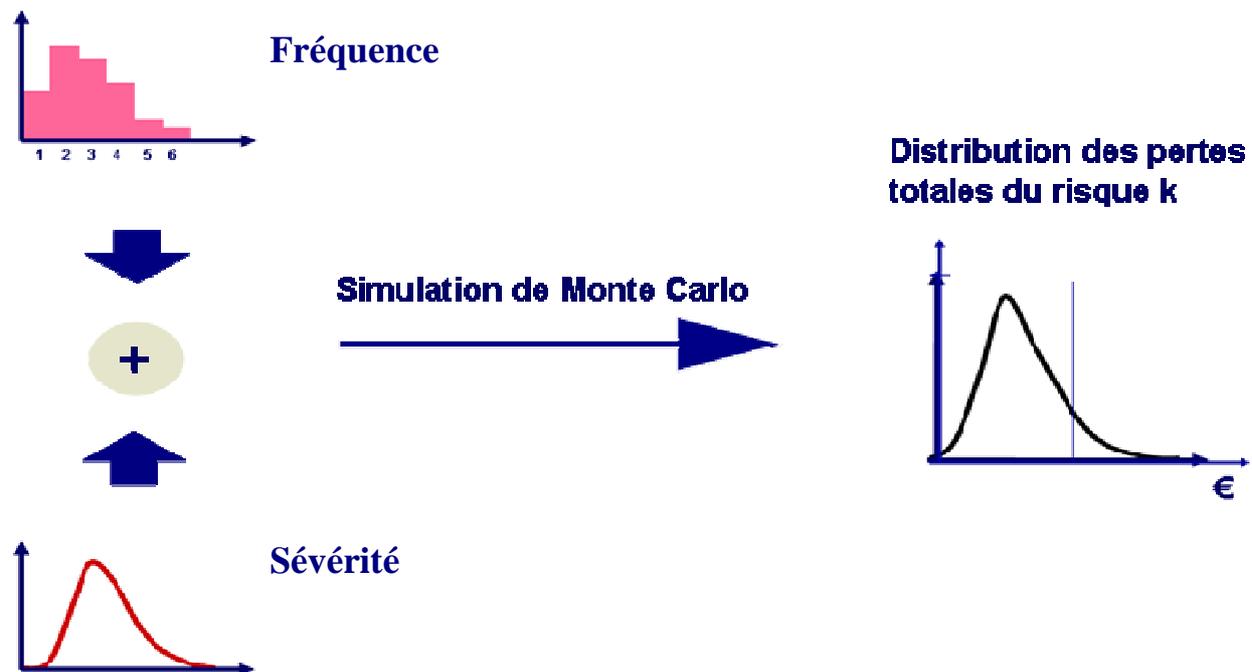
Notons  $\mu = E(S)$ ,  $\sigma = \sqrt{V(S)}$  et  $\gamma_1 = E\left[\left(\frac{S-\mu}{\sigma}\right)^3\right]$  les moments de  $S$ .

Il existe plusieurs approximations basées sur les moments notamment l'approximation Normal Power :

$$F_S(s) \approx \Phi\left[-\frac{3}{\gamma_1} + \sqrt{\frac{9}{\gamma_1^2} + 1 + \frac{6}{\gamma_1}\left(\frac{x-\mu}{\sigma}\right)}\right] \text{ pour } x \geq \mu - \sigma\left[\frac{\gamma_1}{6} + \frac{3}{2\gamma_1}\right] \text{ et } \gamma_1 > 0$$

### 4.3.4 SIMULATION DE MONTE-CARLO

Dans le cadre du calcul de la charge totale, la méthode de Monte-Carlo consiste à simuler un certain nombre de scénarios à partir des lois de fréquence et de coût individuel des pertes comme le montre le schéma suivant.



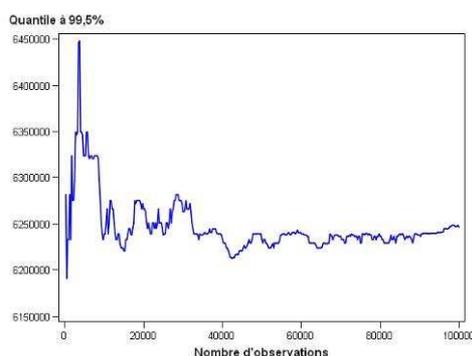
La simulation d'un scénario consiste tout d'abord, à simuler la variable aléatoire de fréquence avec la loi retenue. En fonction de la réalisation obtenue il faut alors simuler le montant des pertes de façon indépendante. Dans l'exemple ci-dessous, le scénario n°1 correspond à la simulation des pertes sur une année. Au cours de cette année, l'entreprise aura subi 135 pertes du risque k dont la perte n°1 aura pour montant 10 000€ et la perte n°135 aura pour montant 150 000€ et le total des pertes

sur un an pour le risque k sera de 680 000€, c'est ce montant que nous conserverons et qui représentera un point de la distribution empirique.

Il suffit alors de répéter cette opération en nombre assez important, de classer tous les scénarios en ordre croissant et nous obtenons la distribution empirique des pertes totales pour le risque k.

N° Scénario	Simulation d'une valeur de fréquence n	Simulation de n coûts de pertes	Calcul de la perte du scénario par somme de $S_i = C_i(1) + \dots + C_i(n_i)$
N° 1	$n_1 = 135$	$\left\{ \begin{array}{l} C_1(1) = 10\ 000\text{€} \\ \dots \\ C_1(135) = 150\ 000\text{€} \end{array} \right.$	$S_1 = 680\ 000\text{€}$
N° 2	$n_2 = 118$	$\left\{ \begin{array}{l} C_2(1) = 9\ 575\text{€} \\ \dots \\ C_2(118) = 22\ 300\text{€} \end{array} \right.$	$S_2 = 364\ 000\text{€}$
•	•	•	•
•	•	•	•
•	•	•	•
N° m	$n_m = 90$	$\left\{ \begin{array}{l} C_m(1) = 5\ 500\text{€} \\ \dots \\ C_m(90) = 18\ 000\text{€} \end{array} \right.$	$S_m = 210\ 000\text{€}$

La convergence se vérifie en représentant graphiquement les valeurs du quantile en fonction du nombre de simulations. Si cela se stabilise cela a toutes les chances de converger.



#### 4.4 THÉORIE DES VALEURS EXTRÊMES

La Théorie des Valeurs Extrêmes (voir [22]) fournit un certain nombre d'outils probabilistes et statistiques pour la modélisation des événements rares qui peut nous être utile pour modéliser les risques opérationnels graves. Nous présentons ici les grandes lignes de cette théorie puis nous verrons comment elle peut s'appliquer au risque opérationnel.

Il existe deux approches principales :

- L'approche des maxima, ou méthode par blocs, qui consiste à observer et modéliser les valeurs de pertes maximales. Si nous raisonnons sur tout notre échantillon, nous n'avons qu'un seul maximum. Pour avoir un certain nombre d'observations de maximum, il faut découper notre échantillon et prendre le maximum pour chaque bloc. Nous obtenons ainsi un nombre de maxima égal au nombre de blocs.

Le théorème de Fisher Tippett nous fournit la loi limite du maximum normalisé. Cette loi est la loi des extrêmes généralisés. Nous utilisons alors nos maxima observés pour la calibrer.

- L'approche dite « Peaks Over Threshold (POT) » qui consiste à considérer les valeurs au-delà d'un seuil et non plus seulement le maximum. Toute la difficulté revient à choisir un seuil approprié qui nous permette d'appliquer le théorème de Pickands. Ce théorème nous permet de modéliser nos excès par une distribution de Pareto Généralisée (GPD). Cette approche est plus couramment utilisée que l'approche des maxima car elle permet d'exploiter plus d'informations fournies par notre échantillon.

Approche des maxima :

Soit  $(x_1, x_2, \dots, x_n)$  un échantillon donné représentant le coût des pertes.

En pratique, le maximum de l'échantillon  $x_{(n)}$  ne représente qu'une seule observation, il n'est pas possible de se baser sur une seule observation pour faire notre modélisation.

L'idée est donc de découper notre échantillon en  $m$  échantillons de taille  $n/m$  pour obtenir  $m$  valeurs de maximum.

Puis le théorème de Fisher-Tippett nous fournit la loi du maximum.

Théorème de Fisher Tippett :

Soit  $(X_1, X_2, \dots, X_n)$  une suite de variables aléatoires indépendantes et identiquement distribuées de fonction de répartition  $F_X$ .

Notons  $X_{n:n} = \max(X_1, X_2, \dots, X_n)$ .

S'il existe deux suites de réels  $(a_n) \in \mathbb{R}$  et  $(b_n) > 0$  et une loi non dégénérée de fonction de répartition  $G$  tels que  $\frac{X_{n:n} - a_n}{b_n} \xrightarrow{loi} G$

alors  $G$  est nécessairement sous l'une des trois formes suivantes :

Fréchet	Weibull	Gumbel
$\Phi_\alpha(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \exp(-x^{-\alpha}) & \text{si } x > 0 \end{cases}$	$\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) & \text{si } x \leq 0 \\ 1 & \text{si } x > 0 \end{cases}$	$\Lambda(x) = \exp(-e^{-x}) \quad \text{pour } x \in \mathfrak{R}$

Ces trois distributions sont le cas particulier de la loi GEV (Generalized Extreme

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

$\xi$  représente l'indice des valeurs extrêmes. Plus il sera élevé en valeur absolue, plus le poids des extrêmes dans la distribution initiale sera important.

$\mu$  le paramètre de localisation. Il indique approximativement le cœur de la distribution.

$\sigma$  le paramètre d'échelle. Il indique l'étalement des extrêmes.

Fréchet	Weibull	Gumbel
$\xi = \alpha^{-1} > 0$ $\mu_n = 0$ $\sigma_n = F^{-1}\left(1 - \frac{1}{n}\right)$	$\xi = -\alpha^{-1} < 0$ $\mu_n = x_F$ $\sigma_n = x_F - F^{-1}\left(1 - \frac{1}{n}\right)$	$\xi = 0$ $\mu_n = F^{-1}\left(1 - \frac{1}{n}\right)$ $\sigma_n = \frac{1}{S(\mu_n)} \int_{\mu_n}^{\infty} S(t) dt$

Où  $S(x) = 1 - F(x)$  et  $x_F = \sup\{x \text{ tel que } F_X(x) < 1\}$

La fonction de répartition  $F_X$  influe sur la distribution limite  $G$ , c'est-à-dire sur les paramètres, elle est dite dans le domaine d'attraction de  $G$ .

La théorie des valeurs extrêmes nous permet de connaître les lois qui appartiennent aux trois domaines d'attractions :

- Les lois de Cauchy, Pareto, Student et LogGamma appartiennent au domaine d'attraction de la loi de Fréchet : ce sont des lois à queues épaisses,

- Les lois Weibull, Normale, Exponentielle, Gamma et Log-Normale appartiennent au domaine d'attraction de la loi de Gumbel : ce sont des lois à queues fines et moyennes,
- Les lois Uniformes et Beta appartiennent au domaine d'attraction de la loi de Weibull : ce sont des lois « à queues finies ».

Ainsi, si nous connaissons la loi des sinistres, nous pouvons en déduire la loi du maximum. Or, en pratique nous ne connaissons pas la loi théorique de notre échantillon et les paramètres doivent être estimés par des méthodes statistiques. Il existe plusieurs méthodes pour estimer les paramètres : nous pouvons distinguer les approches paramétriques et non paramétriques.

Estimation des paramètres par la méthode du maximum de vraisemblance :

Considérons l'échantillon des maxima sur les  $m$  échantillons créés, noté :  $(M_1, M_2, \dots, M_m)$ . La log-vraisemblance peut s'écrire :

Pour  $\xi \neq 0$ ,

$$\begin{aligned} \ln L(M_1, \dots, M_m; \xi, \mu, \sigma) &= \sum_{i=1}^m \ln(h_{\xi, \mu, \sigma}(M_i)) \\ &= -m \ln \sigma - \left(\frac{1+\xi}{\xi}\right) \sum_{i=1}^m \ln\left(1 + \xi \frac{M_i - \mu}{\sigma}\right) - \sum_{i=1}^m \left(1 + \xi \frac{M_i - \mu}{\sigma}\right)^{-1/\xi} \end{aligned}$$

Où  $h_{\xi, \mu, \sigma}$  est la densité de la loi des Valeurs Extrêmes Généralisée (GEV)

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

La résolution de cette équation se fait par des méthodes numériques (algorithme de Quasi-Newton). En pratique, nous utilisons les fonctions prédéfinies sur les logiciels statistiques tels que le logiciel R.

Remarque : Lorsque la taille de l'échantillon est petite, l'estimateur du maximum de vraisemblance fournit une estimation biaisée.

Dans le cas où  $\xi=0$ , la log-vraisemblance s'écrit :

$$\ln L(M_1, \dots, M_m; 0, \mu, \sigma) = -n \ln \sigma - \sum_{i=1}^m \exp\left(-\frac{M_i - \mu}{\sigma}\right) - \sum_{i=1}^m \frac{M_i - \mu}{\sigma}.$$

La résolution se fait là encore par des méthodes numériques.

L'approche des maxima est utilisée dans le cas où nous avons des observations périodiques. Dans le cadre des risques opérationnels, cette méthode ne peut pas être

appliquée dans la mesure où nous n'avons pas forcément d'observations tous les mois et peu de recul.

Méthode des sinistres au-delà d'un seuil :

Cette méthode a pour but de modéliser la loi des observations au-delà d'un seuil par la loi de Pareto Généralisée à travers le théorème suivant :

Théorème de Pickands-Balkema-de Haan :

Soit  $X$  variable aléatoire de fonction de répartition  $F_X$ .

$$\lim_{n \rightarrow \infty} P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = G(x) \Leftrightarrow \lim_{u \rightarrow x_F} \sup_{[0, x_F]} \|F_X^u(x) - G_{\xi, \sigma}(x)\| = 0$$

Avec  $x_F = \sup\{x \text{ tel que } F_X(x) < 1\}$ ,  $F_X^u(x) = P(X - u \leq x \mid X > u)$  et  $G_{\xi, \sigma}$  est la fonction de répartition de la loi de Pareto généralisée :

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

Remarque : l'indice de queue intervenant dans la limite du maximum est identique au paramètre  $\xi$  de la loi GPD.

Pour un seuil  $u$  fixé, la fonction de répartition des excès au-delà de ce seuil peut

$$s'écrire : F_X^u(x) = P(X - u \leq x \mid X > u) = \frac{P(X \leq x + u \cap X > u)}{P(X > u)} = \frac{F_X(x + u) - F_X(u)}{1 - F_X(u)}$$

$u$  doit être suffisamment grand pour que l'on puisse appliquer le résultat précédent, mais  $u$  ne doit pas être trop grand afin d'avoir suffisamment de données pour obtenir des estimateurs de bonne qualité. Généralement,  $u$  est déterminé graphiquement. Pour des échantillons de taille inférieure à 500, le nombre d'excès représentent entre 5 et 10% de l'échantillon.

Méthodes pour déterminer le seuil  $u$  :

Définissons la fonction d'excédent moyen (Mean Excess Function), il s'agit de l'espérance des excès sachant que les pertes dépassent le seuil :

$$e(u) = E(X - u \mid X > u) = \int_u^{x_F} (y - u) \frac{dF_X(y)}{1 - F_X(u)} \text{ avec } x_F = \sup\{x \text{ tel que } F_X(x) < 1\}$$

Cette fonction peut être estimée par la fonction d'excédent moyen empirique :

$$e_n(u) = \frac{\sum_{i=1}^n (X_i - u)^+}{\sum_{i=1}^n I_{\{X_i > u\}}}$$

Pour  $v \geq u$ ,  $E[X - v \mid X > v] = \frac{\sigma(u) + \xi \times (v - u)}{1 - \xi}$  si les excès au-delà du seuil suivent une loi GPD.

L'espérance conditionnelle des excès est une fonction affine de  $v$ , lorsque  $v$  est supérieur au seuil de référence  $u$ . Cela donne un moyen empirique de tester si le seuil  $u$  choisi pour les calculs est suffisamment élevé : les espérances des excès au-delà des seuils plus élevés doivent être alignées sur une droite de pente  $\frac{\xi}{1 - \xi}$ .

Estimateur de Hill : c'est un estimateur de l'indice de queue qui n'est valable que pour  $\xi > 0$ , il s'écrit :

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

Une approche alternative au mean excess plot consiste à calculer l'estimateur de Hill de l'indice de queue pour différents seuils et à chercher à partir de quel seuil cet estimateur est approximativement constant.

Il est également possible de calculer un estimateur de l'indice de queue par la méthode du maximum de vraisemblance pour différents seuils et à chercher à partir de quel seuil cet estimateur est approximativement constant.

Estimation des paramètres :

Lorsque  $u$  est déterminé, les paramètres peuvent être estimés par la méthode du maximum de vraisemblance.

La densité de la loi de Pareto généralisée peut s'écrire :

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

Estimation des paramètres par la méthode du maximum de vraisemblance :

La log-vraisemblance s'écrit  $\ln L(X_1, X_2, \dots, X_{n_u}; \xi, \sigma) = \sum_{i=1}^{n_u} \ln(g_{\xi, \sigma}(X_i))$  avec  $n_u$  la taille de notre échantillon  $X_1, X_2, \dots, X_{n_u}$  qui contient les pertes au dessus du seuil  $u$ .

Pour  $\xi \neq 0$ , la log vraisemblance est égale à :

$$\ln L(X_1, X_2, \dots, X_{n_u}; \xi, \sigma) = -n_u \ln(\sigma) - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^{n_u} \ln\left(1 + \frac{\xi}{\sigma} X_i\right).$$

Dans ce cas, la maximisation de la log-vraisemblance se fait par des méthodes numériques.

Pour  $\xi = 0$ , la log vraisemblance est égale à :

$$\ln L(X_1, X_2, \dots, X_{n_u}; 0, \sigma) = -n_u \ln(\sigma) - \frac{1}{\sigma} \sum_{i=1}^{n_u} X_i$$

Dans ce cas, la maximisation donne une forme analytique de l'estimateur  $\hat{\sigma}_{n_u}$  de  $\sigma$  :

$$\hat{\sigma}_{n_u} = \sum_{i=1}^{n_u} \frac{X_i}{n_u}$$

Il existe d'autres méthodes pour estimer les paramètres d'une loi GPD. Hosking et Wallis (voir [25]) ont montré que, pour des échantillons de taille inférieure à 500, les méthodes des moments et des moments pondérés sont plus efficaces que la méthode du maximum de vraisemblance. Cependant, le problème essentiel est le domaine de validité de  $\xi$  :  $\xi < 1/2$ .

Jean Diebolt, Armelle Guillou et Imen Rached (voir [13]) ont trouvé une méthode permettant d'élargir le domaine de validité à  $\xi < 3/2$ . Il s'agit de la méthode des moments pondérés généralisée.

Pour un seuil suffisamment élevé, la fréquence des excès est modélisée par une loi de Poisson (voir démonstration page 428 [11]).

Cette méthode est très intéressante dans le cadre des risques opérationnels. En effet, l'idée est de trouver un seuil par type de risque et de simuler les pertes en dessous du seuil avec des lois classiques et les pertes au-dessus avec une GPD. Puis de les agréger en supposant que les graves sont indépendants des pertes attritionnelles ce qui est une hypothèse couramment utilisée. Si le nombre de graves est trop faible par risque, nous considérerons notre base et estimerons le seuil des graves sur toute la base. Puis nous simulerons de la même façon que celle décrite précédemment les pertes attritionnelles et les pertes graves.

## 4.5 MESURES DE RISQUE

Dès lors que l'on dispose de la distribution estimée de la perte totale, comment peut-on en déduire le capital économique ? Il faut alors utiliser une mesure de risque qui dans la plupart des cas sera la VaR ou la TVaR que nous définissons plus bas.

Une mesure de risque est, par définition, une fonctionnelle  $f$  qui associe le réel  $f(X)$  à une perte  $X$  du risque.  $f(X)$  sera donc considéré comme le montant de capital dont la compagnie doit disposer pour faire face à une perte de montant  $X$ . Cette mesure doit vérifier un certain nombre propriétés:

➤ Invariance par translation :  $\rho(X + a) = \rho(X) + a$  pour tout  $X$  et  $a \geq 0$

➤ Homogénéité :  $\rho(\lambda X) = \lambda \rho(X)$  pour tout  $X$  et  $\lambda \geq 0$

L'homogénéité peut traduire une invariance par rapport aux unités monétaires.

➤ Monotonie :  $X \leq Y$  implique  $\rho(X) \leq \rho(Y)$  pour tout  $X$  et  $Y$

La monotonie traduit le fait que plus le risque est important, plus le montant de capital sera élevé.

➤ Sous-additivité :  $\rho(X + Y) \leq \rho(X) + \rho(Y)$  pour tout  $X$  et  $Y$

La sous-additivité traduit la réduction de risque par diversification.

➤ Borne supérieure :  $\rho(X) \leq \max(X)$  pour tout  $X$

Si une mesure vérifie toutes les propriétés ci-dessus, elle est dite cohérente.

Les mesures de risque les plus utilisées actuellement sont la Value at Risk (VaR) et la Tail Value at Risk (TVaR).

Définissons la VaR :

Etant donné un risque  $X$  et un niveau de probabilité  $\alpha \in [0; 1]$ , la VaR notée  $VaR_\alpha(X)$  est le quantile d'ordre  $\alpha$  de  $X$ .

$$VaR_\alpha(X) = \inf \{x \in \mathfrak{R} \mid F_X(x) \geq \alpha\}$$

La VaR ne possède pas toutes les propriétés précédentes et n'est donc pas une mesure cohérente. Elle est invariante par translation, homogène et monotone.

La VaR n'est pas sous-additive ce qui implique que l'effet de diversification n'est pas toujours positif : une compagnie pourrait avoir une VaR globale plus élevée que la somme des VaR calculées sur chaque risque. Ceci pousserait les compagnies à ne pas diversifier leurs risques ce qui est contraire à la logique du modèle interne.

De plus, l'inconvénient de la VaR est qu'elle fournit un niveau de pertes qui sera dépassé dans  $1-\alpha$  cas au plus mais elle ne donne aucune information à propos de ce qui se passe lorsque la VaR est dépassée. Par contre, c'est ce que va nous fournir la TVaR.

En effet, la TVaR représente la perte moyenne au-delà de la  $VaR_\alpha(X)$ , ie dans les pires des cas :  $TVaR_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 VaR_\xi(X) d\xi$

La TVaR est cohérente.

Malgré ces qualités, la TVaR est peu utilisée en assurance du fait de la difficulté de calcul et d'interprétation.

Par ailleurs, Solvabilité II recommande d'utiliser la VaR pour déterminer le capital économique.

Calcul de la VaR dans la théorie des extrêmes :

Soit  $n$  le nombre de pertes observées.

$N_u$  représente le nombre de sinistres qui dépassent le seuil  $u$ .

$X$  représente le coût de chaque sinistre.

Une fois les paramètres de la loi de Pareto Généralisée  $\hat{\xi}, \hat{\sigma}$  estimés, il faut inverser la fonction de répartition  $\hat{F}_X(x) = 1 - \frac{N_u}{n} \left( 1 + \hat{\xi} \left( \frac{x-u}{\hat{\sigma}} \right) \right)^{-\frac{1}{\hat{\xi}}}$  pour obtenir la fonction quantile. Cela nous permet ensuite de calculer la VaR d'ordre  $\alpha$  par la formule

suivante:  $VaR_\alpha = u + \frac{\hat{\sigma}}{\hat{\xi}} \left( \left( (1-\alpha) \frac{n}{N_u} \right)^{-\hat{\xi}} - 1 \right)$ .

## 4.6 AGRÉGATION DES RISQUES

Comment modéliser les liens entre les risques ? Comment agréger les scénarios simulés ?

Dans la méthode classique LDA, les VaR des différents risques sont additionnées ce qui correspond comme nous l'avons déjà vu à une dépendance totale positive : les scénarios sont triés par ordre croissant sur chaque risque puis ils sont additionnés. L'emploi d'une matrice de corrélation est très souvent observé, cela permet de prendre en compte la diversification. L'indépendance est également souvent utilisée pour des raisons de simplicité bien qu'elle ne soit pas représentative de la réalité.

Depuis plusieurs années est apparue l'utilisation de la théorie des copules dans la modélisation des dépendances entre les risques. L'avantage des copules par rapport à une matrice de corrélation réside dans la possibilité de lier les risques de façon non linéaire. Nous allons exposer brièvement l'essentiel de la théorie.

Avant d'entamer la théorie des copules, rappelons que les notions de dépendance et corrélation ne sont pas égales. En effet, si deux variables sont indépendantes alors elles ne seront pas corrélées. Par contre, la réciproque n'est pas vraie (excepté dans le

cas gaussien). Intéressons-nous aux mesures de concordance qui sont utiles dans la théorie des copules.

#### 4.6.1 MESURE DE CONCORDANCE

Définissons la concordance :

Soient  $(x, y)$  et  $(\tilde{x}, \tilde{y})$  deux réalisations d'un vecteur aléatoire continu  $(X, Y)$ , alors  $(x, y)$  et  $(\tilde{x}, \tilde{y})$  sont dites concordantes si  $(x - \tilde{x})(y - \tilde{y}) > 0$  et discordantes si  $(x - \tilde{x})(y - \tilde{y}) < 0$ .

Le tau de Kendall et le rhô de Spearman sont deux mesures de concordance couramment utilisées en statistique. Elles donnent une mesure de la corrélation entre les rangs des observations, à la différence du coefficient de corrélation linéaire qui lui apprécie la corrélation entre les valeurs des observations. Elles offrent par ailleurs l'avantage de s'exprimer simplement en fonction de la copule associée au couple de variables aléatoires. Par contre, ces mesures ne permettent d'estimer la dépendance qu'entre 2 variables.

Tau de Kendall :

Pour  $(x, y)$  et  $(\tilde{x}, \tilde{y})$  deux réalisations d'un vecteur aléatoire continu  $(X, Y)$ , le tau de Kendall est défini par  $\tau(X, Y) = \Pr((x - \tilde{x})(y - \tilde{y}) > 0) - \Pr((x - \tilde{x})(y - \tilde{y}) < 0)$ .

Avec une structure de copule  $C$  continue (voir définition copule sur la page suivante), nous pouvons déduire :

$$\tau(X, Y) = 4E[C(U, V)] - 1$$

Rhô de Spearman :

Le Rhô de Spearman est défini par :

$$\rho_s(X, Y) = 3\{\Pr((x - \tilde{x})(y - \tilde{y}) > 0) - \Pr((x - \tilde{x})(y - \tilde{y}) < 0)\}$$

$$\text{Avec une copule continue : } \rho_s(X, Y) = 12 \iint_{[0,1]^2} C(u, v) du dv - 3$$

Le Rhô de Spearman peut aussi être considéré comme un coefficient de corrélation sur les rangs des variables.

$$\rho_s(X, Y) = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{(\sum_{i=1}^n (R_i - \bar{R})^2) \sqrt{(\sum_{i=1}^n (S_i - \bar{S})^2)}}, \text{ où } R_i \text{ et } S_i \text{ sont les rangs de la réalisation de } (X, Y).$$

#### 4.6.2 DÉFINITION DES COPULES

Nous pouvons à présent entrer dans le vif du sujet avec, pour commencer, la définition d'une copule :

Une copule est une fonction de répartition multivariée  $C$  définie sur  $[0;1]^n$  et dont les lois marginales sont uniformes standards :

$$C(u_1, \dots, u_n) = P(U_1 \leq u_1, \dots, U_n \leq u_n)$$

L'outil fondamental de la théorie des copules est le théorème de Sklar (1959) :

Soit  $F$  une fonction de répartition  $n$ -dimensionnelle avec des marginales  $F_1, \dots, F_n$ , alors il existe une  $n$ -copule  $C$  telle que pour tout  $x$  de  $\mathbb{R}^n$

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n))$$

Si de plus les marginales  $F_1, \dots, F_n$  sont continues alors  $C$  est unique.

Ainsi une copule résume toute la structure de dépendance.

Il existe plusieurs types de copules dont les deux principaux sont : les copules elliptiques et les copules archimédiennes.

##### ➤ Copules archimédiennes :

La définition d'une copule archimédienne est la suivante :

Soit  $\varphi$  une fonction convexe, continue, strictement décroissante de  $[0; 1]$  dans  $[0; \infty]$  telle que  $\varphi(1)=0$  et  $\varphi(0)=\infty$  alors  $C(u_1, \dots, u_n) = \varphi^{-1}(\varphi(u_1) + \dots + \varphi(u_n))$  est une copule archimédienne stricte et  $\varphi$  est appelé générateur strict de  $C$ .

Les copules archimédiennes ont l'avantage de décrire des structures de dépendance très diverses. L'inconvénient des copules archimédiennes est qu'elles ne possèdent qu'un seul paramètre donc dans le cas où la dimension sera supérieure à 3 la structure se résumera en un seul paramètre. Dans ce cas, il faudra passer par plusieurs copules archimédiennes bi-variées mais là, le problème sera le conflit entre les dépendances.

Le tableau suivant comporte les caractéristiques des copules les plus connues.

Copule	$\varphi(u)$	$C(u,v)$	Taux de Kendall $\tau_a$
Gumbel	$(-\ln(u))^a, a \geq 1$	$\exp\left(-\left[(-\ln u)^a + (-\ln v)^a\right]^{1/a}\right)$	$1 - 1/a$
Frank	$-\ln\left(\frac{e^{-au} - 1}{e^{-a} - 1}\right), a \neq 0$	$-\frac{1}{a} \ln\left[1 + \frac{(e^{-au} - 1)(e^{-av} - 1)}{e^{-a} - 1}\right]$	$1 - \frac{4(1 - D_1(a))}{a}$
Clayton	$\frac{(u^{-a} - 1)}{a}, a > 0$	$(u^{-a} + v^{-a} - 1)^{-1/a}$	$\frac{a}{a + 2}$

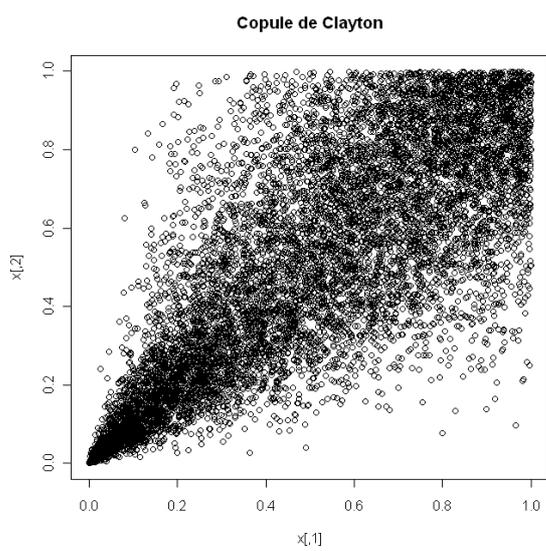
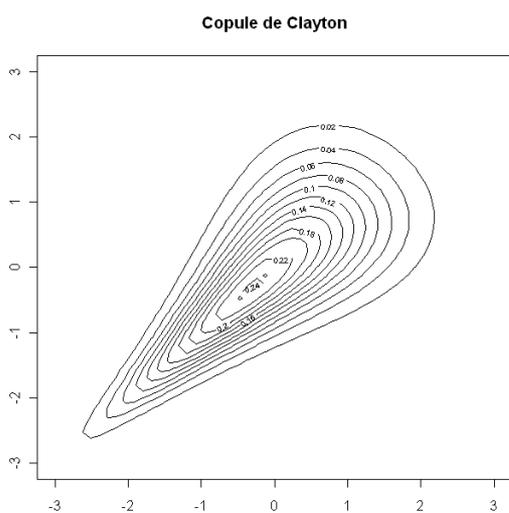
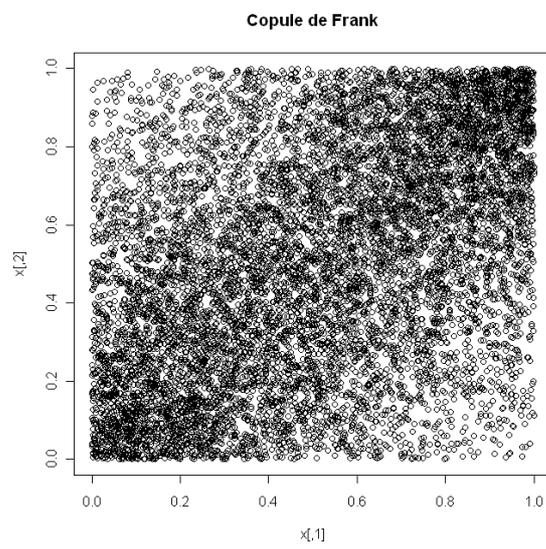
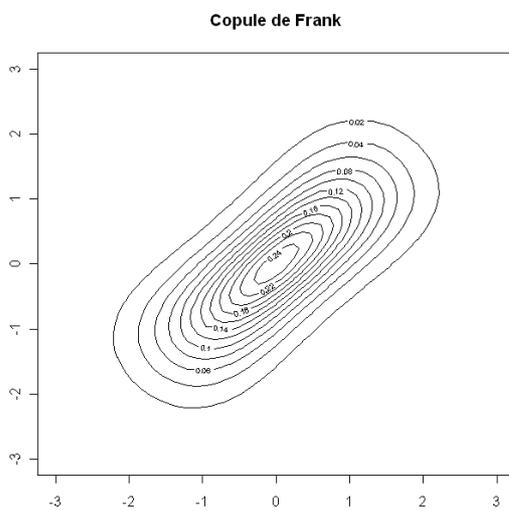
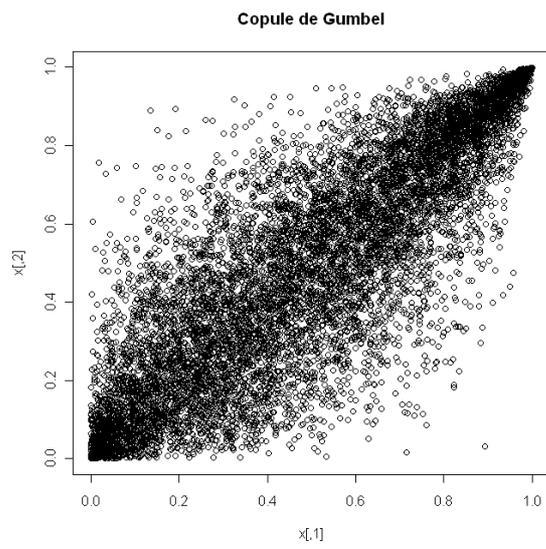
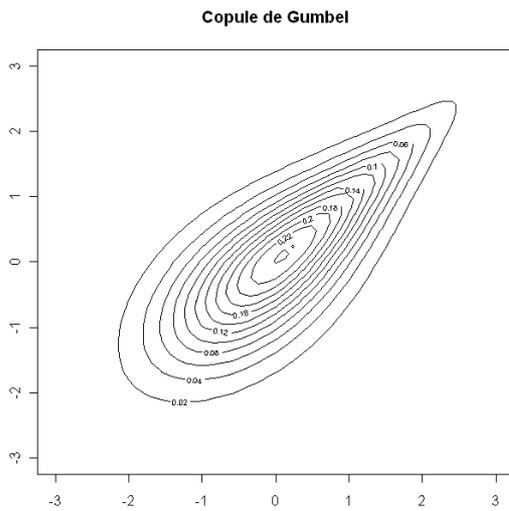
$$\text{Où } D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} dt$$

Le paramètre  $a$  mesure le degré de dépendance entre les risques : plus il est élevé plus la dépendance est forte et une valeur positive de  $a$  indique une dépendance positive.

La copule de Gumbel n'appréhende que des dépendances positives et notamment la queue supérieure. En particulier, elle convient très bien pour modéliser les dépendances entre des risques rares à fort impact.

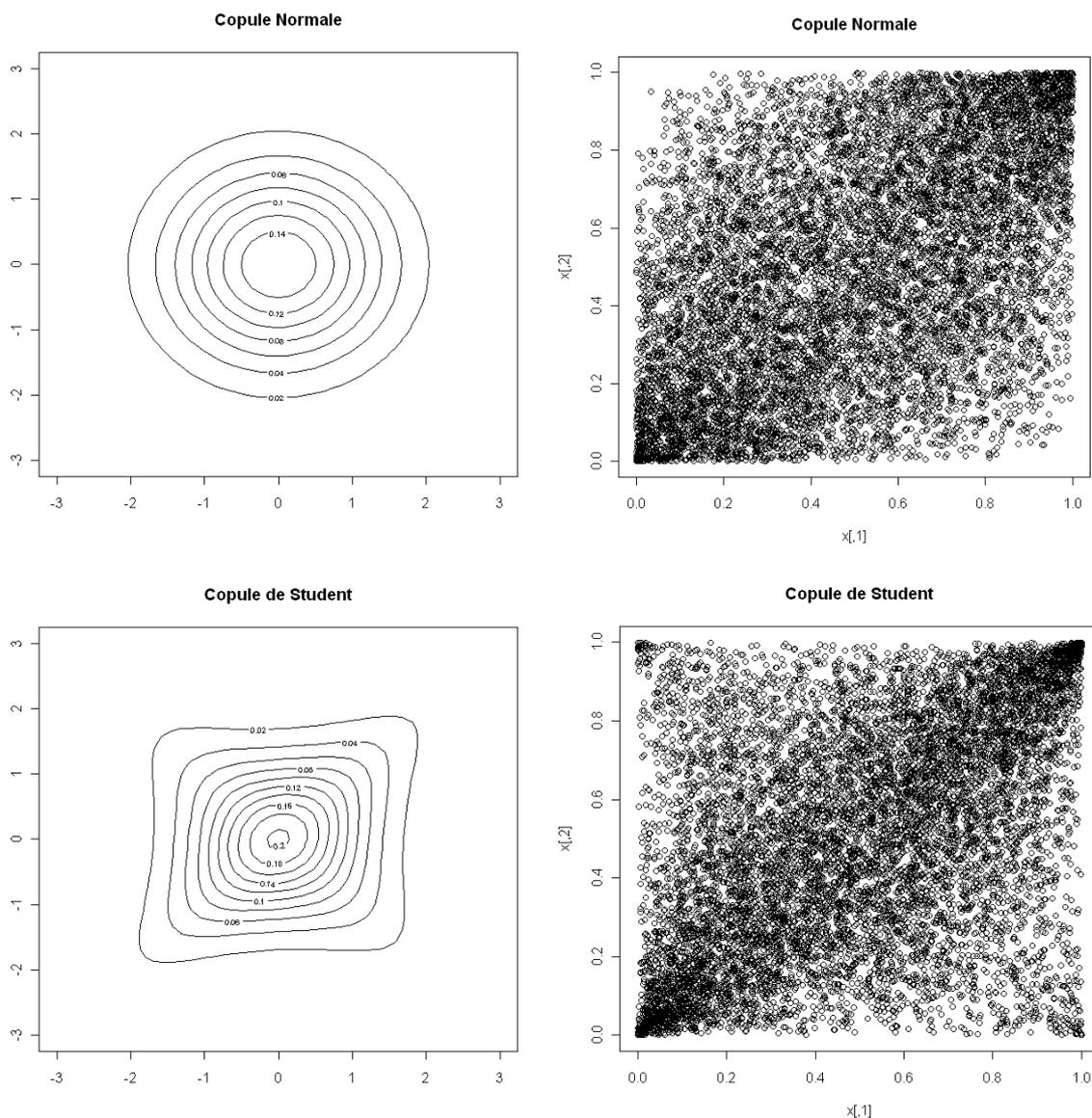
La copule de Frank permet de modéliser les dépendances aussi bien positives que négatives mais il n'existe pas de dépendance de queue pour cette copule.

La copule de Clayton ne permet de modéliser que des dépendances positives comme la copule de Gumbel. Cependant, elle appréhende les dépendances entre risques de fréquence à faible intensité.



➤ Copules elliptiques :

Parmi les copules elliptiques figurent les copules gaussienne et de Student, elles sont peu adaptées à nos problématiques dans la mesure où elles s'appliquent à des distributions symétriques. Cependant, les copules gaussiennes offrent l'avantage d'appréhender des dépendances en dimension  $n$  très simplement à l'aide d'une matrice de corrélation.



La copule gaussienne est définie par :

$$C(u_1, \dots, u_d) = \Phi_{\Sigma}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d))$$

Où  $\Phi(x)$  est la fonction de répartition de la loi normale centrée réduite et  $\Phi_{\Sigma}(x_1, \dots, x_d)$  est la fonction de répartition jointe des distributions normales et  $\Sigma$  la matrice de corrélation.

Le nombre de paramètres de la copule est le nombre de corrélations  $d(d-1)/2$ .

Il n'existe pas de forme simple de la copule. Dans le cas bi dimensionnel (avec un seul élément de corrélation), la copule gaussienne peut s'écrire :

$$C(u_1, u_2) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right) dy dx$$

#### 4.6.3 ESTIMATION DES PARAMÈTRES D'UNE COPULE

Nous allons présenter ici les différentes méthodes permettant d'estimer les paramètres d'une copule.

➤ Méthode des moments :

Cette méthode consiste à utiliser une mesure de concordance qui dépendra du paramètre de la copule et à l'inverser pour en déduire une estimation du paramètre. Cette méthode n'est pas robuste dans la mesure où il y a plusieurs mesures de concordances utilisables. Dans la pratique, c'est le taux de Kendall qui est le plus utilisé en tant que mesure de concordance pour estimer le paramètre car il est facile à calculer.

Pour les copules archimédiennes, nous avons vu plus haut les relations qui existaient entre le tau de Kendall et le paramètre. Il suffit d'« inverser » cette relation. Dans le cas de la copule de Frank, le tau de Kendall s'exprime en fonction d'une intégrale du paramètre, cette résolution sera donc numérique.

Pour les copules elliptiques, il existe un lien simple entre la matrice de forme  $\Sigma$  et le tau de Kendall (ou le rhô de Spearman). Ces deux mesures ne dépendent pas des distributions marginales des variables, mais uniquement de la copule.

$$\tau(X_i, X_j) = \frac{2}{\pi} \arcsin(\rho_{i,j}), \text{ où } \rho_{i,j} \text{ est le coefficient de corrélation entre } X_i \text{ et } X_j.$$

Nous en déduisons  $\Sigma$ , en inversant chaque relation :

$$\Sigma = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \cdots & \rho_k \\ \rho_1 & 1 & \rho_{k+1} & \cdots & \cdots & \rho_j \\ \rho_2 & \rho_{k+1} & \ddots & \rho_{j+1} & \cdots & \rho_{n-3} \\ \vdots & \vdots & \rho_{j+1} & \ddots & \rho_{n-2} & \rho_{n-1} \\ \vdots & \vdots & \vdots & \rho_{n-2} & 1 & \rho_n \\ \rho_k & \rho_j & \rho_{n-3} & \rho_{n-1} & \rho_n & 1 \end{pmatrix}$$

➤ Méthode du maximum de vraisemblance :

La méthode du maximum de vraisemblance consiste à maximiser la vraisemblance de la loi jointe.

Grâce au théorème de SKLAR, nous pouvons exprimer la densité d'un vecteur aléatoire  $(X_1, X_2, \dots, X_n)$  en fonction de la densité de la copule associée  $c$  et de ses marginales.

Soit  $f$  la fonction de densité de la distribution jointe de  $(X_1, X_2, \dots, X_n)$ , alors :

$f(x_1, \dots, x_n) = c(F_1(x_1), \dots, F_n(x_n)) \prod_{i=1}^n f_i(x_i)$ , où les  $f_i$  sont les densités des lois marginales, les  $F_i$  sont les fonctions de répartition des lois marginales et  $c(\cdot)$  est la densité de la copule.

La fonction de log-vraisemblance s'écrit alors :

$L(\theta) = \sum_{t=1}^T \ln c(F_1(x_1^t; \theta_1), \dots, F_n(x_n^t; \theta_n); \alpha) + \sum_{t=1}^T \sum_{i=1}^n \ln f_i(x_i^t; \theta_i)$ , où  $\theta = (\theta_1, \dots, \theta_n, \alpha)$  est le vecteur des paramètres à estimer

Nous pouvons déduire l'estimateur  $\theta$  en maximisant la vraisemblance.

➤ Méthode IFM (Inference Functions for Margins):

Cette méthode propose une estimation séparée des paramètres des distributions marginales et ceux de la copule.

Dans un premier temps, nous devons estimer les paramètres de chaque distribution marginale une à une par la méthode du maximum de vraisemblance :

$\hat{\theta}_i = \arg \max L^i(\theta_i) = \arg \max \sum_{t=1}^T \ln f_i(x_i^t; \theta_i)$ , où  $L^i$  est la fonction de log-vraisemblance de la distribution marginale  $F_i$ .

Nous pouvons alors déduire les paramètres  $\alpha$  de la copule en utilisant l'estimation :

$\hat{\alpha} = \arg \max L^c(\alpha) = \arg \max \sum_{t=1}^T \ln c(F_1(x_1^t; \hat{\theta}_1), \dots, F_n(x_n^t; \hat{\theta}_n); \alpha)$ , où  $L^c$  est la fonction de log-vraisemblance de la copule.

➤ Méthode CML (Canonical Maximum Likelihood):

La différence entre cette méthode et la méthode IFM est que nous n'avons pas besoin d'estimer les paramètres des distributions marginales pour calibrer la copule.

Etape 1 : Nous devons dans un premier temps transformer l'échantillon des observations  $(x_1^t, \dots, x_n^t)$  en un échantillon de fonction de répartition marginale uniforme standard  $(\hat{u}_1^t, \dots, \hat{u}_n^t)$  en utilisant la distribution empirique.

Etape 2 : Les paramètres de la copule sont alors les paramètres qui maximisent la log-vraisemblance de la copule :  $\hat{\alpha} = \arg \max \sum_{t=1}^T \ln c(\hat{u}_1^t, \dots, \hat{u}_n^t; \alpha)$ .

Une fois la copule choisie et ses paramètres estimés, l'agrégation des scénarios est réalisée par la méthode des rangs.

#### 4.6.4 MÉTHODE DES RANGS

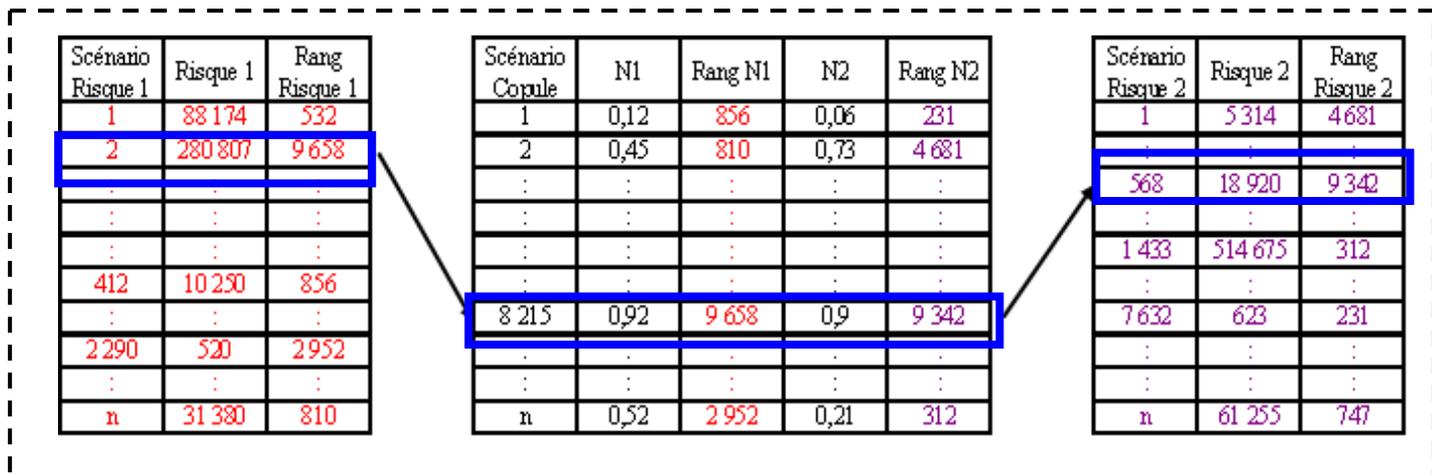
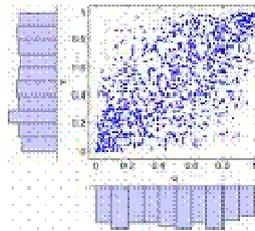
De manière générale, cette méthode consiste à lier les scénarios de différents risques par leur rang à l'aide des copules pour obtenir des scénarios agrégés. Par souci de clarté, la méthode détaillée ci-dessous est appliquée à deux risques uniquement. Cependant, il est possible de la généraliser à plusieurs risques (c'est d'ailleurs ce que nous avons fait dans l'application ci-après).

- 1) Simuler n scénarios de pertes totales pour chacun des deux risques ;
- 2) Pour chaque risque, classer les scénarios obtenus par le rang : la perte la plus élevée aura un rang égal à n et la plus faible, un rang égal à 1 ;
- 3) Simuler n fois un couple de variables aléatoires  $N_1, N_2$  par une copule ;
- 4) Attribuer un rang à chacune des variables  $N_1$  et  $N_2$  ;
- 5) Lier le rang de  $N_1$  avec le rang du premier risque et celui de  $N_2$  avec celui du second risque. Les valeurs du risque 1 sont ainsi reliées aux valeurs du risque 2 ;
- 6) Sommer les deux risques pour obtenir une distribution de pertes totales.

Voici une illustration de cette méthode :

Pour le risque 1, le scénario n° 2 correspond à la perte de rang 9 658.

D'après le scénario n° 8 215 de la copule, le rang 9 658 du risque 1 est lié au rang 9 342 du risque 2, ce qui signifie que le scénario n°2 du risque 1 est lié au scénario n° 568 du risque 2 : la perte du risque 1 de 280 807 euros s'ajoute à la perte du risque 2 de 18 920 euros ce qui donne une perte totale de 299 727 euros.



Scénario Risque 1	Risque 1	Rang Risque 1
1	88 174	532
2	280 807	9 658
...	...	...
412	10 250	856
...	...	...
2 290	520	2 952
...	...	...
n	31 380	810

Scénario Copule	N1	Rang N1	N2	Rang N2
1	0,12	856	0,06	231
2	0,45	810	0,73	4 681
...	...	...	...	...
8 215	0,92	9 658	0,9	9 342
...	...	...	...	...
n	0,52	2 952	0,21	312

Scénario Risque 2	Risque 2	Rang Risque 2
1	5 314	4 681
...	...	...
568	18 920	9 342
...	...	...
1 433	514 675	312
...	...	...
7 632	623	231
...	...	...
n	61 255	747

Scénario Copule	Scénario Risque 1	Risque 1	Scénario Risque 2	Risque 2	Perte Totale
1	412	10 250	7 632	623	10 873
2	n	31 380	1	5 314	36 694
...	...	...	...	...	...
8 215	2	280 807	568	18 920	299 727
...	...	...	...	...	...
n	2 290	520	1 433	514 675	515 195

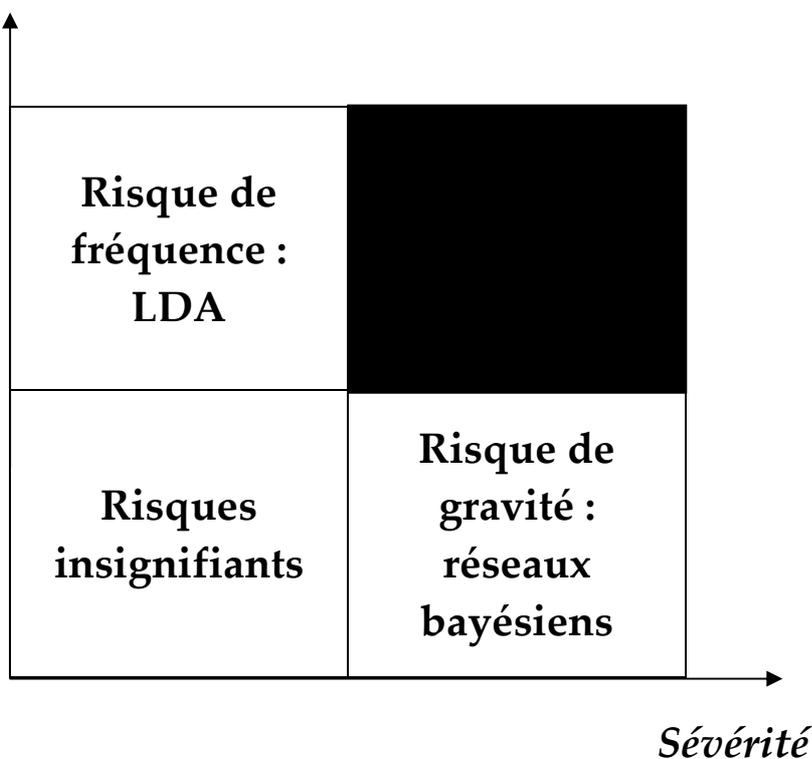


## 5 MODÈLE THÉORIQUE DU RISQUE OPÉRATIONNEL

Ce chapitre expose le modèle interne que Generali envisage de mettre en place à terme. Il nécessite un certain nombre d'informations et distingue deux types de risques opérationnels qui sont modélisés de façon différente :

- les risques de fréquence : associé à des pertes de montants peu élevés mais fréquentes, sont modélisés par la méthode LDA ;
- les risques de gravité : associé à des pertes de montants très importants mais très rares, sont modélisés par les réseaux bayésiens.

*Fréquence*



Pour identifier la méthodologie à appliquer, le type de risque sera identifié lors de la cartographie des risques et devra être revu périodiquement.

Les risques de pertes importantes et fréquentes ne seront pas modélisés car lorsqu'un risque de gravité survient l'entreprise mettra en place des contrôles et autres plans d'actions qui viendront diminuer voire éliminer ce risque (case noire).

Lors de la cartographie, il faudra distinguer les risques qui nécessitent du capital de ceux qui n'en nécessitent pas.

Les raisons possibles pour ne pas détenir de capital sont :

- Le risque n'impacte pas la solvabilité : risque d'opportunité stratégique, risque d'affaires nouvelles ;

- Il existe des actions plus appropriées pour faire face à ce risque, pour certains risques, il vaut mieux :
  - renforcer les contrôles pour réduire ou éliminer le risque (par exemple en installant un nouveau processus de surveillance),
  - transférer le risque à travers des assurances : assurer le risque (lorsqu'il est assurable),
  - avoir un plan de contingence autre que détenir du capital comme par exemple avoir des fonds externes,
  - éviter le risque en stoppant l'activité à risque en la différant ou en changeant de stratégie.

L'entreprise devra documenter en détail les raisons qui l'autorisent à ne pas détenir de capital pour certains risques, y compris les raisons qui lui permettent de croire que les contrôles et actions de réduction sont adéquats.

- L'impact est couvert par ailleurs.

Remarque 1 : Un des problèmes dans la modélisation du risque opérationnel est de garantir l'exhaustivité des risques et d'éviter les doubles comptages. C'est le référentiel de risques qui va nous permettre de résoudre en partie ce problème. C'est pour cela qu'il est essentiel et qu'il doit être bien verrouillé avant de passer à la modélisation.

Remarque 2 : La modélisation séparée des risques de fréquence et de gravité rend la modélisation des risques de fréquence beaucoup moins sensibles aux choix techniques (lois ajustées, agrégation).

## **5.1 MODÉLISATION DES RISQUES DE FRÉQUENCE**

La méthode LDA décrite précédemment sera utilisée pour modéliser les risques de fréquence. C'est ce qui est couramment employé dans le secteur bancaire pour modéliser ce type de risque.

Pour utiliser cette méthode, il est nécessaire de disposer d'une base d'incidents suffisamment bien alimentée pour pouvoir faire des ajustements de lois. Se pose ainsi le problème du seuil de collecte des pertes : à partir de quel niveau doit-on récupérer les pertes sans que cela ne soit trop lourd en terme de charges de travail pour les opérationnels mais pour que le modèle soit suffisamment alimenté en données ? Dans le secteur bancaire, le seuil le plus fréquemment utilisé est de 10 000 euros.

La base de données devra être retraitée des pertes issues des risques de gravité qui sont modélisés par ailleurs.

Quelles sont les informations nécessaires et suffisantes pour mettre en œuvre le modèle ?

La base devra contenir au minimum :

- les catégories de risque aux niveaux 1, 2 et 3 ce qui permettra de rattacher la perte au référentiel de risques,
- le processus impacté (sous forme de codification du processus) ce qui permettra de rattacher la perte au processus impacté,
- le service dans lequel s'est produit l'incident,
- les services impactés,
- la date de survenance,
- la date de détection,
- la date de fin d'incident,
- le coût estimé de la perte,
- les récupérations éventuelles.

Voici un extrait de la base prévisionnelle :

N° Perte	Date de survenance	Date de détection	Date de fin	Processus impacté	Service impacté	Catégorie de risque niveau 1	Catégorie de risque niveau 2	Catégorie de risque niveau 3	Pertes estimées a priori	Pertes réelles	Récupération estimée	Récupérations réelles
1	20070115	20070115	20070115	000015	000104	000010	000010	061	147,95	147,95	0	0
2	20070118	20070118	20070118	000020	000152	000040	000040	096	1405000	576620,22	0	290254,7
3	20070119	20070119	20070119	000015	000105	000010	000010	066	289	501	0	0
4	20061111	20061111	20061111	000015	000118	000010	000010	008	12,2	12,2	0	0
5	20061208	20061208	20061208	000015	000105	000010	000010	086	170	247,38	0	0
6	20061108	20061108	20061108	000015	000105	000010	000010	086	30	45,66	0	0
7	20061229	20061229	20061229	000015	000105	000010	000010	086	1700	2401,21	0	0
8	20061213	20061213	20061213	000015	000105	000020	000020	086	80	181,26	0	0
9	20061127	20061127	20061127	000015	000104	000020	000020	008	545,41	515,41	0	0
10	20061215	20061215	20061215	000015	000105	000020	000020	008	9582,06	3194,02	0	0

A partir de cette base, des analyses de données seront menées pour ajuster les lois de fréquence et de coûts. C'est une méthode classique de l'assurance.

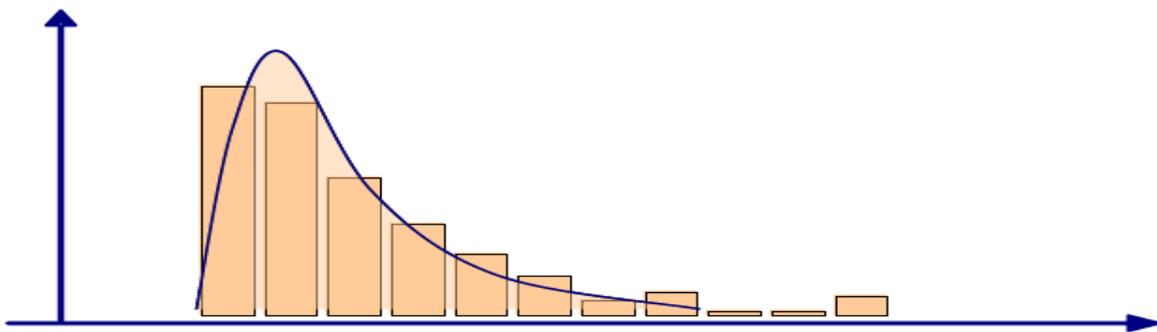
Remarque : Le risque que nous souhaitons modéliser est le risque que nous encouurons dans l'année à venir. L'historique de la base d'incidents ne sera utile que pour les risques dont la structure est stable dans le temps.

Seuil de collecte :

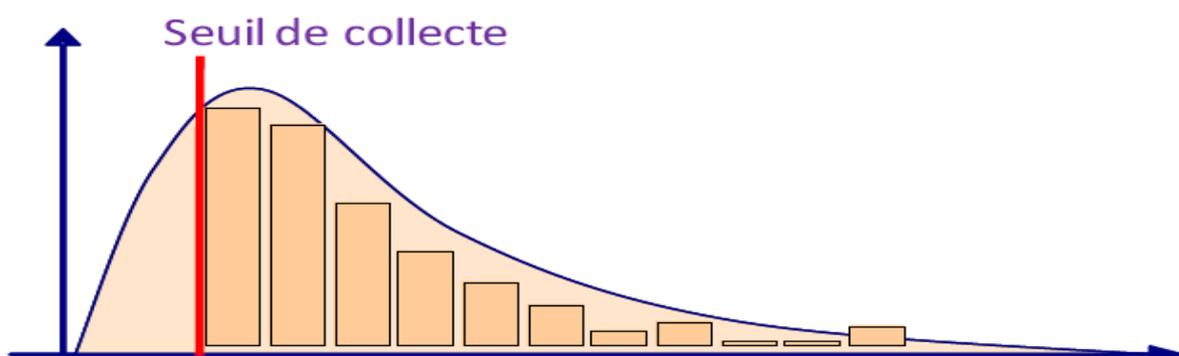
Généralement, les opérationnels ne déclareront pas toutes les pertes. Le seuil de déclaration sera déterminé par le risk management. Les observations dont nous disposerons seront alors les pertes dont le montant dépasse ce seuil. Il faudra ainsi prendre en compte cette troncature dans nos ajustements et modifier l'estimation des paramètres et des tests d'adéquation en conséquence. En effet, si nous ne tenons pas compte du fait que les données sont tronquées, nous allons sous-estimer la fréquence et ajuster la loi des coûts individuels de façon erronée.

## Loi du coût individuel :

Si nous ne tenons pas compte du seuil de troncature, nous allons ajuster à nos données une loi déformée comme nous le prouve le graphique ci-dessous (ce graphique représente le coût des pertes individuelles et l'adéquation de la loi théorique ajustée sur les données tronquées) :



Les données en dessous du seuil ne sont pas remontées mais doivent être prises en compte dans l'estimation des paramètres : il faudra estimer les paramètres de la loi en tenant compte du manque de données à gauche de l'histogramme tel que nous le montre le graphique ci-dessous (ce graphique représente le coût des pertes individuelles et l'adéquation de la loi théorique réajustée pour tenir compte des données tronquées) :



Soit  $U$  le seuil de collecte des pertes.

Soit  $X$  la variable aléatoire représentant les coûts des pertes non tronquées (c'est-à-dire les pertes dont le montant peut être inférieur à  $U$ ) et  $F$  sa fonction de répartition.

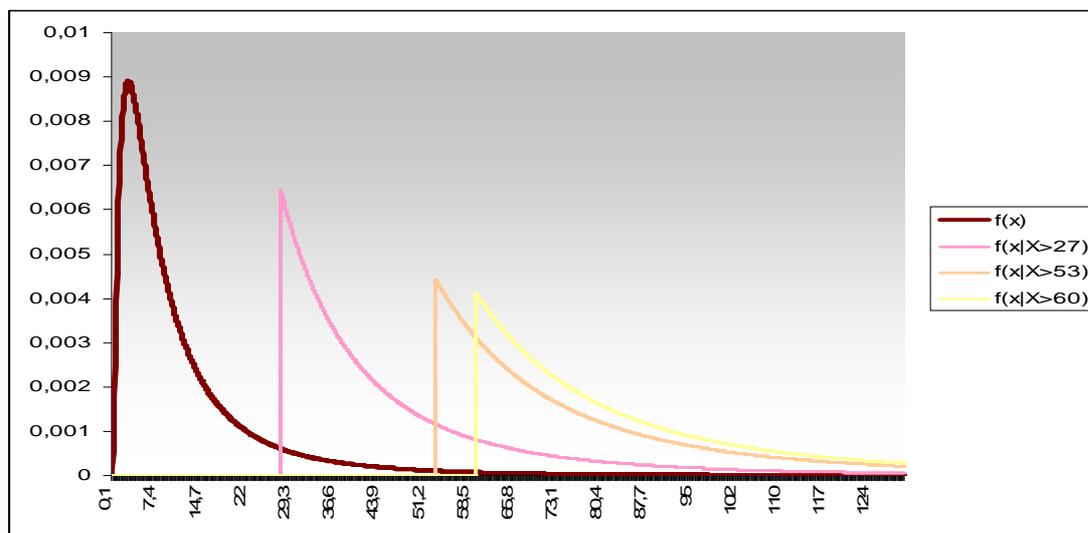
Seule la loi de  $X \mid X \geq U$  produit des observations. Cette loi conditionnelle est la loi de  $X$  tronquée à gauche de  $U$ .

Nous pouvons expliciter la fonction de répartition de la loi tronquée sachant que les observations sont au-delà du seuil  $U$ :

$$P(X < x \mid X \geq U) = \frac{P(U \leq X < x)}{P(X \geq U)} = \frac{F(x) - F(U)}{1 - F(U)} \quad \text{pour tout } x > U$$

$$P(X \leq x \mid X \geq U) = 0 \quad \text{pour tout } x \leq U$$

Le graphique ci-dessous montre la déformation d'une loi Log-normale  $\ln(2; 1)$  en fonction du niveau de seuil de collectes.



Remarque : certaines distributions sont stables par une troncature à gauche comme les lois exponentielle et Pareto.

La méthode du maximum de vraisemblance nous permet de calibrer nos lois mais ne nous donne pas de solution analytique.

Soit  $(x_1, x_2, \dots, x_n)$  un échantillon de pertes dont le montant excède le seuil  $U$ , la vraisemblance conditionnelle au seuil  $U$  peut s'écrire :

$$\prod_{i=1}^n \frac{f(x_i)}{P(X_i \geq U)} = \prod_{i=1}^n \frac{f(x_i)}{1 - F(U)}$$

La log-vraisemblance peut alors s'écrire :

$$\sum_{i=1}^n \ln \left( \frac{f(x_i)}{1 - F(U)} \right) = \sum_{i=1}^n \ln(f(x_i)) - n \ln(1 - F(U))$$

Les paramètres sont estimés de manière classique en minimisant la log-vraisemblance. En général, il n'y a pas de forme analytique des estimateurs des paramètres. Les algorithmes classiques tels que Newton ne sont pas très efficaces. Nous aurons recours à des fonctions du logiciel R telles que « cdens » qui permet de

calculer la densité d'une loi connue tronquée et « fitdistr » qui utilise le maximum de vraisemblance.

Loi de la fréquence :

Lorsqu'il existe un seuil de collecte, la fréquence des pertes observées est sous-estimée et elle doit être réajustée pour tenir compte des pertes non déclarées. Pour chaque période d'observation  $i$ , le nombre de sinistres noté  $n_i$  est augmenté du nombre de sinistres estimés en-dessous du seuil noté  $m_i$  :  $n_i^r = n_i + m_i$ .

Le rapport entre le nombre de pertes en dessous du seuil et le nombre de pertes observées est le même que le rapport entre la partie en dessous de la courbe de la densité du coût individuel située à gauche du seuil, et celle située à droite du seuil :

$$\frac{m_i}{n_i} = \frac{\hat{F}(U)}{1 - \hat{F}(U)} \text{ où } \hat{F} \text{ est la fonction de répartition tronquée de } X, \text{ dont les paramètres}$$

ont été estimés avec la méthode précédente.

$$\text{D'où } n_i^r = n_i + m_i = n_i + \frac{n_i \times \hat{F}(U)}{1 - \hat{F}(U)} = \frac{n_i}{1 - \hat{F}(U)}.$$

## 5.2 MODÉLISATION DES RISQUES DE GRAVITÉ

Contrairement aux pertes des risques de fréquence, les pertes de risques de gravité sont trop rares pour qu'une base de données soit constituée et exploitable. Le modèle va donc reposer sur les scénarios à dire d'experts issus du SRA et des observations de d'indicateurs de risques et sera complété par d'autres scénarios de risques identifiés lors de la cartographie des risques par les experts. Nous devons faire un SRA pour ces nouveaux risques.

Rappelons que dans le SRA nous avons défini pour chaque risque de gravité trois scénarios qui vont être utilisés pour définir un graphe bayésien.

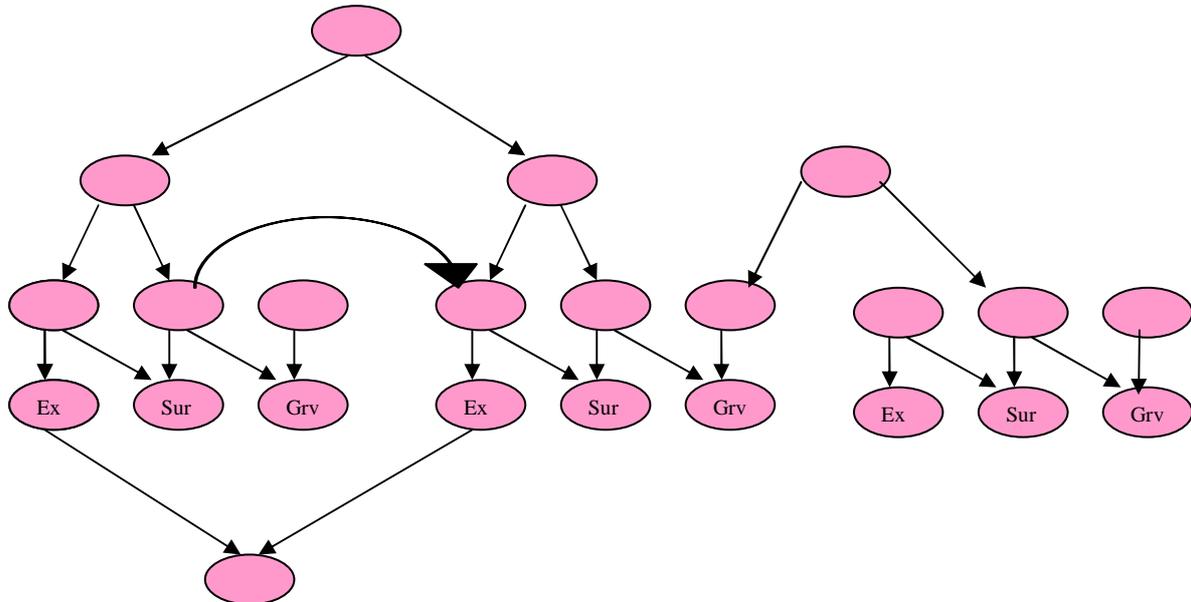
Dans le modèle XSG décrit plus haut, chaque graphe bayésien est composé d'une exposition, d'une fréquence de survenance et d'un coût. Les différentes valeurs des variables et les probabilités associées sont en principe déterminées par l'expert en collaboration avec le risk manager ou issues des observations des indicateurs de risques.

Le SRA nous permet de construire la structure du réseau bayésien. Cependant il n'est pas suffisant pour en déduire tous les paramètres du réseau, notamment les probabilités conditionnelles.

Une fois la structure du réseau fixée, il s'agit de déterminer l'ensemble des paramètres, ce qui est appelé l'apprentissage des paramètres. Pour cela, il y a deux

approches possibles : soit il existe des données et nous appliquons des méthodes statistiques, soit il n'y en a pas et il est nécessaire de réinterroger les experts pour tenter de déterminer les éléments manquants.

Les différents réseaux bayésiens sont assemblés pour former un réseau bayésien global qui nous donne la distribution totale des risques de gravité. Il suffit de connecter les différents indicateurs ensemble et de lancer toutes les simulations en même temps.



*Ex = exposition, Sur = survenance, Grv = gravité*

### 5.2.1 APPRENTISSAGE DES CONNAISSANCES

Cette partie consiste à présenter les méthodes permettant d'estimer les paramètres à dire d'experts lorsqu'aucune information n'est disponible.

Nous commençons par présenter une méthode pour aider les experts à estimer les probabilités : il s'agit de l'échelle de probabilités.

L'échelle de probabilité permet aux experts d'utiliser des informations à la fois textuelles et numériques pour assigner un degré de réalisation à leur affirmation puis de comparer les probabilités des événements pour les modifier.

certain	100
probable	85
attendu	75
moitié - moitié	50
incertain	25
improbable	15
impossible	0

Le problème qui se pose souvent, est que l'expert va devoir estimer des probabilités conditionnelles en nombre très important. Ceci est peu réaliste et nous conduit à envisager d'autres solutions. En général, pour pallier ce problème, nous ne retenons que les variables que nous pouvons quantifier (expertise, base de données, formules). Cependant, il existe plusieurs approches qui permettent de déduire un certain nombre de probabilités conditionnelles à partir d'un nombre réduit de probabilités estimées par les experts : OU-bruité, les facteurs d'interpolation, le modèle log-linéaire... Nous allons étudier les modèles OU-bruité et facteurs d'interpolation.

Les modèles OU-bruité et facteurs d'interpolation consistent à déterminer les probabilités de l'enfant  $Y$  conditionnellement à l'état des parents  $X_i$ . Cela nécessite auparavant que les experts estiment certaines probabilités conditionnelles comme nous le verrons plus loin.

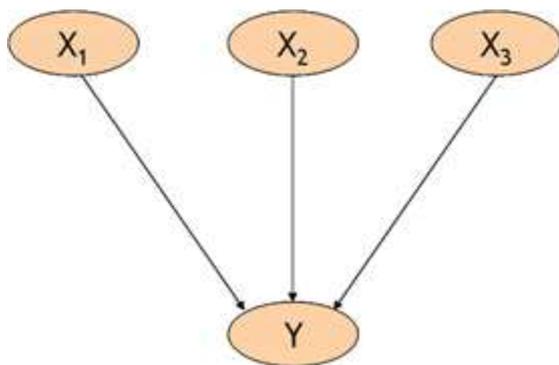
Ces modèles reposent sur l'hypothèse que l'effet de chaque parent  $X_i$  sur l'enfant  $Y$  ne dépend pas de la valeur des autres parents  $X_i$ .

Nous prenons le cas où les parents  $X_i$  et l'enfant  $Y$  sont des variables binaires. Dans le cas des facteurs d'interpolation, cette théorie est généralisable à des variables qui peuvent prendre plus de deux valeurs et même à des variables continues (voir le document de Cain [4]).

Nous noterons  $x_i$  l'état positif du parent  $X_i$  qui va le plus probablement provoquer l'état de succès du nœud enfant  $Y$  ( $\bar{x}_i$  étant l'état négatif).

Nous noterons  $y$  l'état de succès du nœud enfant  $Y$  ( $\bar{y}$  étant l'état d'échec de l'enfant).

Plaçons-nous dans le cas où l'enfant  $Y$  possède trois parents.



L'expert estime dans un premier temps les probabilités marginales de chaque variable :  $P(Y = y)$ ,  $P(X_1 = x_1)$ ,  $P(X_2 = x_2)$  et  $P(X_3 = x_3)$ .

L'expert devrait en théorie estimer  $2^3$  probabilités conditionnelles ( $2^n$  dans le cas de  $n$  parents). Nous allons voir que ce nombre peut être réduit par les modèles.

Modèle OU-bruité :

L'expert est interrogé pour donner les probabilités  $P_{4y}$ ,  $P_{6y}$  et  $P_{7y}$  correspondant au cas où un seul parent est dans son état positif (voir tableau ci-dessous). Le nombre de probabilités conditionnelles à estimer n'est plus que de trois (n probabilités dans le cas de n parents).

Le tableau ci-dessous représente la table des probabilités conditionnelles de l'enfant Y

Numéro de la combinaison	Parent $X_1$	Parent $X_2$	Parent $X_3$	Fils Y
1	$x_1$	$x_2$	$x_3$	$P_{1y} = P(Y = y   X_1 = x_1, X_2 = x_2, X_3 = x_3)$ à estimer
2	$x_1$	$x_2$	$\bar{x}_3$	$P_{2y} = P(Y = y   X_1 = x_1, X_2 = x_2, X_3 = \bar{x}_3)$ à estimer
3	$x_1$	$\bar{x}_2$	$x_3$	$P_{3y} = P(Y = y   X_1 = x_1, X_2 = \bar{x}_2, X_3 = x_3)$ à estimer
4	$x_1$	$\bar{x}_2$	$\bar{x}_3$	$P_{4y} = P(Y = y   X_1 = x_1, X_2 = \bar{x}_2, X_3 = \bar{x}_3)$ estimée par l'expert
5	$\bar{x}_1$	$x_2$	$x_3$	$P_{5y} = P(Y = y   X_1 = \bar{x}_1, X_2 = x_2, X_3 = x_3)$ à estimer
6	$\bar{x}_1$	$x_2$	$\bar{x}_3$	$P_{6y} = P(Y = y   X_1 = \bar{x}_1, X_2 = x_2, X_3 = \bar{x}_3)$ estimée par l'expert
7	$\bar{x}_1$	$\bar{x}_2$	$x_3$	$P_{7y} = P(Y = y   X_1 = \bar{x}_1, X_2 = \bar{x}_2, X_3 = x_3)$ estimée par l'expert
8	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	$P_{8y} = P(Y = y   X_1 = \bar{x}_1, X_2 = \bar{x}_2, X_3 = \bar{x}_3)$ à estimer

L'expert n'a estimé que trois probabilités qui permettent cependant de déduire les probabilités restantes de la façon suivante :

$$P_{1y} = 1 - (1 - P_{4y}) \times (1 - P_{6y}) \times (1 - P_{7y})$$

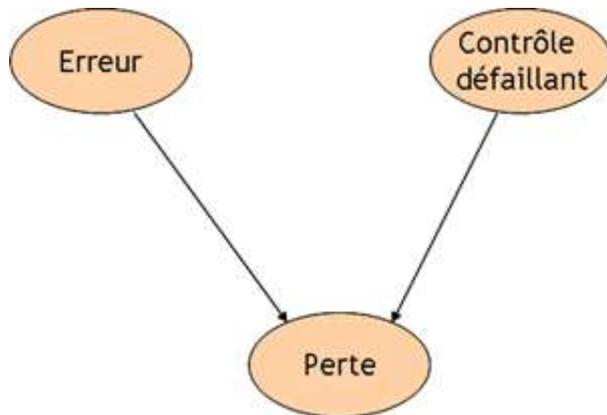
$$P_{2y} = 1 - (1 - P_{4y}) \times (1 - P_{6y})$$

$$P_{3y} = 1 - (1 - P_{4y}) \times (1 - P_{7y})$$

$$P_{5y} = 1 - (1 - P_{6y}) \times (1 - P_{7y})$$

$$P_{8y} = 0$$

Exemple :



$P(\text{Perte} = \text{« Oui »} \mid \text{Erreur} = \text{« Oui »}, \text{Contrôle défaillant} = \text{« Non »}) = 0,1$

$P(\text{Perte} = \text{« Oui »} \mid \text{Erreur} = \text{« Non »}, \text{Contrôle défaillant} = \text{« Oui »}) = 0,99$

Alors nous en déduisons la probabilité :

$$P(\text{Perte} = \text{« Oui »} \mid \text{Erreur} = \text{« Oui »}, \text{Contrôle défaillant} = \text{« Oui »}) = 1 - (1-0,1) \times (1-0,99) = 0,991$$

L'inconvénient de cette méthode est qu'elle estime que  $P(\text{Perte} = \text{« Oui »} \mid \text{Erreur} = \text{« Non »}, \text{Contrôle défaillant} = \text{« Non »}) = 0$ .

#### Modèle des facteurs d'interpolation :

L'expert est interrogé pour donner les probabilités suivantes :

- la première probabilité  $P_{1y}$  correspond au cas où les parents sont tous dans leurs états positifs,
- la deuxième probabilité  $P_{8y}$  correspond au cas où les parents sont tous dans leurs états négatifs,
- les autres probabilités  $P_{2y}$ ,  $P_{3y}$  et  $P_{5y}$  correspondent au cas où un seul parent est dans son état négatif.

Le nombre de valeurs à estimer est de cinq ( $2^n - n$  valeurs dans le cas de  $n$  parents).

Le tableau ci-dessous représente la table des probabilités conditionnelles de l'enfant Y

Numéro de la combinaison	Parent $X_1$	Parent $X_2$	Parent $X_3$	Fils Y
1	$x_1$	$x_2$	$x_3$	$P_{1y} = P(Y = y   X_1 = x_1, X_2 = x_2, X_3 = x_3)$ estimée par l'expert
2	$x_1$	$x_2$	$\bar{x}_3$	$P_{2y} = P(Y = y   X_1 = x_1, X_2 = x_2, X_3 = \bar{x}_3)$ estimée par l'expert
3	$x_1$	$\bar{x}_2$	$x_3$	$P_{3y} = P(Y = y   X_1 = x_1, X_2 = \bar{x}_2, X_3 = x_3)$ estimée par l'expert
4	$x_1$	$\bar{x}_2$	$\bar{x}_3$	$P_{4y} = P(Y = y   X_1 = x_1, X_2 = \bar{x}_2, X_3 = \bar{x}_3)$ à estimer
5	$\bar{x}_1$	$x_2$	$x_3$	$P_{5y} = P(Y = y   X_1 = \bar{x}_1, X_2 = x_2, X_3 = x_3)$ estimée par l'expert
6	$\bar{x}_1$	$x_2$	$\bar{x}_3$	$P_{6y} = P(Y = y   X_1 = \bar{x}_1, X_2 = x_2, X_3 = \bar{x}_3)$ à estimer
7	$\bar{x}_1$	$\bar{x}_2$	$x_3$	$P_{7y} = P(Y = y   X_1 = \bar{x}_1, X_2 = \bar{x}_2, X_3 = x_3)$ à estimer
8	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	$P_{8y} = P(Y = y   X_1 = \bar{x}_1, X_2 = \bar{x}_2, X_3 = \bar{x}_3)$ estimée par l'expert

Les facteurs d'interpolation sont calculés pour chaque commutation dans l'état d'un parent de positif à négatif. Ils sont le rapport de la différence entre la probabilité de succès sachant qu'un seul parent est dans son état négatif et la probabilité que tous les parents soient dans leurs états négatifs et de la différence entre la probabilité la plus haute pour l'état de succès de l'enfant (c'est-à-dire quand tous les parents sont dans leurs états positifs) et la plus basse (quand tous les parents sont dans leurs états négatifs).

Quand un des parents passe d'un état positif à un état négatif, la probabilité que l'enfant soit dans son état de succès est réduite. Le facteur d'interpolation évalue simplement cette réduction pour la commutation de chaque parent.

Posons  $IF_i = \frac{P_{iy} - P_{8y}}{P_{1y} - P_{8y}}$  pour  $i$  dans  $\{2;3;5\}$

Il n'est pas nécessaire de calculer l'un d'entre eux pour trouver les probabilités restantes :

$$P_{4y} = (P_{3y} - P_{8y}) \times IF_2 + P_{8y}$$

$$P_{6y} = (P_{5y} - P_{8y}) \times IF_2 + P_{8y}$$

$$P_{7y} = (P_{5y} - P_{8y}) \times IF_3 + P_{8y}$$

Nous n'avons pas besoin de calculer des facteurs d'interpolation pour les probabilités d'échec. Une fois que nous avons utilisé les facteurs d'interpolation pour calculer les probabilités de succès de chaque combinaison, la probabilité d'échec s'en déduit simplement par la formule suivante :  $P_{iy} = 1 - P_{iy}$ .

#### Test de cohérence des estimations :

Le test de cohérence permet de vérifier la cohérence des estimations des experts.

Le test consiste à comparer la probabilité estimée par l'expert  $P(Y)$  avec la formule des probabilités totales :

$$\begin{aligned} P(Y = y) &= P(Y = y | X_1 = x_1, X_2 = \bar{x}_2, X_3 = \bar{x}_3) \times P(X_1 = x_1, X_2 = \bar{x}_2, X_3 = \bar{x}_3) \\ &+ P(Y = y | X_1 = \bar{x}_1, X_2 = x_2, X_3 = x_3) \times P(X_1 = \bar{x}_1, X_2 = x_2, X_3 = x_3) \\ &+ P(Y = y | X_1 = \bar{x}_1, X_2 = x_2, X_3 = \bar{x}_3) \times P(X_1 = \bar{x}_1, X_2 = x_2, X_3 = \bar{x}_3) \\ &+ P(Y = y | X_1 = x_1, X_2 = \bar{x}_2, X_3 = x_3) \times P(X_1 = x_1, X_2 = \bar{x}_2, X_3 = x_3) \\ &+ P(Y = y | X_1 = \bar{x}_1, X_2 = \bar{x}_2, X_3 = x_3) \times P(X_1 = \bar{x}_1, X_2 = \bar{x}_2, X_3 = x_3) \\ &+ P(Y = y | X_1 = x_1, X_2 = x_2, X_3 = \bar{x}_3) \times P(X_1 = x_1, X_2 = x_2, X_3 = \bar{x}_3) \\ &+ P(Y = y | X_1 = \bar{x}_1, X_2 = \bar{x}_2, X_3 = \bar{x}_3) \times P(X_1 = \bar{x}_1, X_2 = \bar{x}_2, X_3 = \bar{x}_3) \\ &+ P(Y = y | X_1 = x_1, X_2 = x_2, X_3 = x_3) \times P(X_1 = x_1, X_2 = x_2, X_3 = x_3) \end{aligned}$$

Si le résultat est incohérent, il faut alors corriger :

- soit manuellement : l'expert réestime les probabilités,
- soit automatiquement : en modifiant les probabilités en respectant les proportions pour que l'égalité soit vraie.

Reprenons l'exemple du modèle OU-bruité :

$$\begin{aligned}
 P(Y = y) &= P(Y = y | X_1 = x_1, X_2 = \bar{x}_2) \times P(X_1 = x_1, X_2 = \bar{x}_2) \\
 &\quad + P(Y = y | X_1 = \bar{x}_1, X_2 = x_2) \times P(X_1 = \bar{x}_1, X_2 = x_2) \\
 &\quad + P(Y = y | X_1 = x_1, X_2 = x_2) \times P(X_1 = x_1, X_2 = x_2) \\
 &\quad + P(Y = y | X_1 = \bar{x}_1, X_2 = \bar{x}_2) \times P(X_1 = \bar{x}_1, X_2 = \bar{x}_2)
 \end{aligned}$$

$$\begin{aligned}
 P(Y = y) &= 0,1 \times 0,001 \times 0,9999 + 0,99 \times 0,999 \times 0,0001 + 0,991 \times 0,001 \times 0,0001 + 0 \\
 &= 0,00129
 \end{aligned}$$

Les experts avaient estimé cette probabilité :  $P(Y = y) = 0,012$  ce qui est très proche et serait validé. L'inconvénient d'estimer nulle la probabilité  $P(Y = y | X_1 = \bar{x}_1, X_2 = \bar{x}_2)$  peut être très problématique et fournir une estimation de  $P(Y = y)$  totalement fausse.

Cette méthode ne fonctionne donc que pour des valeurs de  $P(Y = y | X_1 = \bar{x}_1, X_2 = \bar{x}_2)$  proche de 0.

### 5.2.2 APPRENTISSAGE DES DONNÉES

Là encore il existe plusieurs approches selon que les données soient complètes ou non.

#### ➤ Cas des données complètes :

Étudions le cas de données complètes :

Nous rappelons que le but est de déterminer des probabilités conditionnelles.

Notons  $P_{1y} = P(Y = y | X_1 = x_1, X_2 = x_2, X_3 = x_3)$  les probabilités à estimer.

#### Apprentissage statistique :

La méthode la plus utilisée dans le cas de données complètes est la méthode du maximum de vraisemblance qui est assez intuitive et qui consiste à calculer la probabilité d'un événement par sa fréquence d'apparition dans la base de données.

Notons  $x_j$  la configuration numéro  $j$  des parents et  $x_k$  l'état du nœud  $X_i$ .

Soit  $pa(X_i)$  le parent de  $X_i$ .

Pour estimer la probabilité conditionnelle que  $X_i$  soit dans l'état  $x_k$  sachant que ses parents sont dans la configuration  $x_j$ ,  $P(X_i = x_k | pa(X_i) = x_j)$ , nous pouvons utiliser la formule :

$$\hat{P}(X_i = x_k | pa(X_i) = x_j) = \frac{N_{i,j,k}}{\sum_k N_{i,j,k}}$$

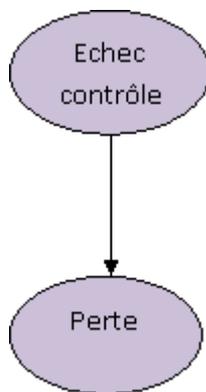
Où  $N_{i,j,k}$  représente le nombre de fois où  $X_i$  apparaît dans l'état  $x_k$  alors que ses parents sont dans la configuration  $x_j$ .

Vous trouverez la démonstration de cette formule dans [20] page 118.

Notons  $\hat{\theta}_{i,j,k}^{MV}$  l'estimateur par la méthode du maximum de vraisemblance de  $P(X_i = x_k \mid pa(X_i) = x_j)$ .

Pour économiser du temps et de l'espace nous noterons  $\theta_{i,j,k}$  la probabilité conditionnelle  $P(X_i = x_k \mid pa(X_i) = x_j)$  qui est une notation classique dans le domaine bayésien.

Exemple : considérons le réseau bayésien suivant.



Le nœud Échec contrôle est le parent du nœud Perte. Tous deux sont des variables aléatoires binaires qui ont pour modalités Oui=O et Non=N.

Le nœud Échec contrôle sera noté E et le nœud Perte sera noté P. E sera la variable  $X_1$  et P la variable  $X_2$ .

Si  $E = O$  alors cela signifiera que le contrôle a échoué. Cela correspond à son état 1.

Si  $E = N$  alors cela signifiera que le contrôle a fonctionné. Cela correspond à son état 2.

Si  $P = O$  alors cela signifiera qu'une perte s'est réalisée. Cela correspond à son état 1.

Si  $P = N$  alors cela signifiera qu'aucune perte n'a été enregistrée. Cela correspond à son état 2.

Supposons à présent que nous ayons les observations suivantes concernant ce réseau :

Numéro scénario	Échec du contrôle	Perte
1	O	O
2	O	O
3	O	O
4	O	N
5	O	N
6	O	N
7	N	O
8	N	O
9	N	O
10	N	N
11	N	N
12	N	N
13	N	N
14	N	N
15	N	N

Les paramètres à estimer sont les probabilités :

$$P(E = O) \text{ (nous en déduirons } P(E = N) = 1 - P(E = O))$$

$$P(P = O \mid E = O) \text{ (nous en déduirons } P(P = N \mid E = O) = 1 - P(P = O \mid E = O))$$

$$P(P = O \mid E = N) \text{ (nous en déduirons } P(P = N \mid E = N) = 1 - P(P = O \mid E = N))$$

Intéressons-nous en particulier à  $P(P = O \mid E = O)$  et  $P(P = O \mid E = N)$ , que nous noterons respectivement  $\theta_{P=O \mid E=O}$  et  $\theta_{P=O \mid E=N}$ .

La formule du maximum de vraisemblance nous donne :

$$\hat{\theta}_{P=O|E=O}^{MV} = \frac{N_{2,1,1}}{N_{2,1,1} + N_{2,1,2}}$$

ceci correspond au rapport du nombre de fois où P est dans son état 1 sachant que son parent E est dans son état 1 par le nombre de fois où P est dans ses états 1 et 2 sachant que son parent E est dans son état 1.

Voici une explication des indices :

$$\hat{\theta}_{P=O|E=O}^{MV} = \frac{N_{2,1,1}}{N_{2,1,1} + N_{2,1,2}}$$

Les observations nous permettent de quantifier l'estimateur :

$$\hat{\theta}_{P=O|E=O} = \frac{N_{2,1,1}}{N_{2,1,1} + N_{2,1,2}} = \frac{3}{3+3} = \frac{3}{6} = 0,5$$

$$\text{Nous en déduisons } \hat{\theta}_{P=N|E=O} = 1 - \hat{\theta}_{P=O|E=O} = 1 - \frac{1}{2} = \frac{1}{2} = 0,5$$

$$\hat{\theta}_{P=O|E=N} = \frac{N_{2,2,1}}{N_{2,2,1} + N_{2,2,2}} = \frac{3}{3+6} = \frac{3}{9} \approx 0,333$$

$$\text{Nous en déduisons } \hat{\theta}_{P=N|E=N} = 1 - \hat{\theta}_{P=O|E=N} = 1 - \frac{1}{3} = \frac{2}{3} \approx 0,667$$

### Apprentissage bayésien :

Cette approche consiste tout d'abord à appliquer le théorème de Bayes.

Soient  $(X_1, \dots, X_n) := X$  les variables ou nœuds composant notre réseau bayésien.

Soit  $d^{(l)} = (x_{k_1}^{(l)}, \dots, x_{k_n}^{(l)})$  un scénario des états des nœuds de notre base de données.

$$\text{On a : } P(X = d^{(l)} | \theta) = P(X_1 = x_{k_1}^{(l)}, \dots, X_n = x_{k_n}^{(l)} | \theta)$$

Soit  $D = \{d^{(l)} \text{ pour } l \in [1; N]\}$  l'ensemble des N scénarios de notre base de données.

Le théorème de Bayes nous donne :

$$P(\theta | X = D) \propto P(X = D | \theta) P(\theta)$$

Cette approche utilise un a priori sur la distribution des paramètres, notamment la distribution de Dirichlet qui est souvent employée pour certains avantages que nous étudierons plus loin :

$P(\theta) \propto \prod_{i=1}^n \prod_{j=1}^{q_i} \prod_{k=1}^{r_j} (\theta_{i,j,k})^{\alpha_{i,j,k}-1}$  où les paramètres  $\alpha_{i,j,k}$  sont déterminés a priori par les experts.

L'approche du maximum a posteriori nous donne l'estimateur suivant :

$$\hat{P}(X_i = x_k \mid pa(X_i) = x_j) = \frac{N_{i,j,k} + \alpha_{i,j,k} - 1}{\sum_k (N_{i,j,k} + \alpha_{i,j,k} - 1)}$$

Où  $\alpha_{i,j,k}$  sont les paramètres de la distribution de Dirichlet associée à la loi a priori

$$P(X_i = x_k \mid pa(X_i) = x_j).$$

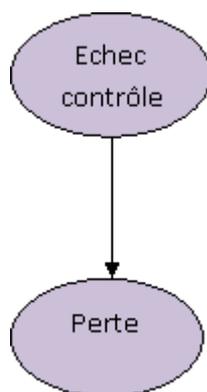
Notons  $\hat{\theta}_{i,j,k}^{MAP}$  l'estimateur du maximum a posteriori de  $P(X_i = x_k \mid pa(X_i) = x_j)$ .

La deuxième approche bayésienne :

$$\hat{P}(X_i = x_k \mid pa(X_i) = x_j) = \frac{N_{i,j,k} + \alpha_{i,j,k}}{\sum_k (N_{i,j,k} + \alpha_{i,j,k})}$$

Notons  $\hat{\theta}_{i,j,k}^{EAP}$  l'estimateur de l'espérance a posteriori de  $P(X_i = x_k \mid pa(X_i) = x_j)$ .

Exemple : reprenons notre exemple précédent.



Supposons que les paramètres  $\theta_{P=O \mid E=N}$  et  $\theta_{P=N \mid E=N}$  suivent une distribution a priori de Dirichlet avec  $\alpha_{i,j,k} = 1$  pour tous  $i, j, k \in \{1,2\}$ .

Maximum a posteriori :

$$\hat{\theta}_{P=O|E=O}^{MAP} = \frac{N_{2,1,1} + \alpha_{2,1,1} - 1}{N_{2,1,1} + \alpha_{2,1,1} - 1 + N_{2,1,2} + \alpha_{2,1,2} - 1} = \frac{3+1-1}{3+1-1+3+1-1} = \frac{1}{2} = 0,5$$

$$\hat{\theta}_{P=N|E=O}^{MAP} = 1 - \hat{\theta}_{P=O|E=O}^{MAP} = 1 - \frac{1}{2} = 0,5$$

$$\hat{\theta}_{P=O|E=N}^{MAP} = \frac{N_{2,2,1} + \alpha_{2,2,1} - 1}{N_{2,2,1} + \alpha_{2,2,1} - 1 + N_{2,2,2} + \alpha_{2,2,2} - 1} = \frac{3+1-1}{3+1-1+6+1-1} = \frac{1}{3} \approx 0,333$$

$$\hat{\theta}_{P=N|E=N}^{MAP} = 1 - \hat{\theta}_{P=O|E=N}^{MAP} = 1 - \frac{1}{3} \approx 0,667$$

Espérance a posteriori :

$$\hat{\theta}_{P=O|E=O}^{EAP} = \frac{N_{2,1,1} + \alpha_{2,1,1}}{N_{2,1,1} + \alpha_{2,1,1} + N_{2,1,2} + \alpha_{2,1,2}} = \frac{3+1}{3+1+3+1} = \frac{4}{8} = 0,5$$

$$\hat{\theta}_{P=N|E=O}^{EAP} = 1 - \hat{\theta}_{P=O|E=O}^{EAP} = 1 - \frac{1}{2} = \frac{1}{2} = 0,5$$

$$\hat{\theta}_{P=O|E=N}^{EAP} = \frac{N_{2,2,1} + \alpha_{2,2,1}}{N_{2,2,1} + \alpha_{2,2,1} + N_{2,2,2} + \alpha_{2,2,2}} = \frac{3+1}{3+1+6+1} = \frac{4}{11} \approx 0,364$$

$$\hat{\theta}_{P=N|E=N}^{EAP} = 1 - \hat{\theta}_{P=O|E=N}^{EAP} = 1 - \frac{4}{11} \approx 0,636$$

➤ Cas des données incomplètes :

Les méthodes que nous venons de présenter sont uniquement valables dans le cas où toutes les données sont observées. Si tel n'est pas le cas, voici d'autres méthodes permettant d'estimer les probabilités conditionnelles.

Apprentissage statistique EM :

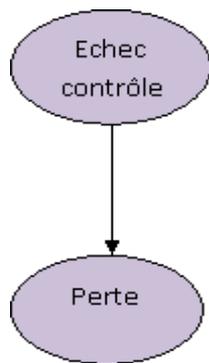
Il s'agit de l'apprentissage statistique combiné à l'algorithme « expectation maximisation » (EM).

Soit  $D_o$  l'ensemble des données observées et  $D_m$  l'ensemble des données manquantes.

$$N_{i,j,k}^* = E_{\theta} [N_{i,j,k}] = N \times P(X_i = x_k, pa(X_i) = x_j | \theta^*)$$

$$\theta_{i,j,k}^{(t+1)} = \frac{N_{i,j,k}^*}{\sum_k N_{i,j,k}^*}$$

Exemple :



Supposons que nous disposions uniquement des informations suivantes :

Numéro scénario	Echec du contrôle	Perte
1	O	?
2	O	O
3	O	O
4	O	N
5	O	N
6	O	N
7	N	O
8	N	O
9	N	O
10	N	?
11	N	?
12	N	N
13	N	N
14	N	N
15	N	N

Intéressons-nous en particulier à  $P(P=O | E=O)$  et  $P(P=O | E=N)$ , que nous noterons respectivement  $\theta_{P=O | E=O}$  et  $\theta_{P=O | E=N}$

L'algorithme EM fonctionne avec des paramètres initiaux qui peuvent être issus d'un avis d'expert :

$$\theta_{P=O|E=O}^{(0)} = 0,6 \text{ et } \theta_{P=O|E=N}^{(0)} = 0,3$$

1<sup>ière</sup> itération :

Etape E :

Numéro scénario	Echec du contrôle	Perte	E = O		E = N	
			P = O	P = N	P = O	P = N
1	O	?	0,6	0,4	0	0
2	O	O	1	0	0	0
3	O	O	1	0	0	0
4	O	N	0	1	0	0
5	O	N	0	1	0	0
6	O	N	0	1	0	0
7	N	O	0	0	1	0
8	N	O	0	0	1	0
9	N	O	0	0	1	0
10	N	?	0	0	0,3	0,7
11	N	?	0	0	0,3	0,7
12	N	N	0	0	0	1
13	N	N	0	0	0	1
14	N	N	0	0	0	1
15	N	N	0	0	0	1
N*			2,6	3,4	3,6	5,4

L'étape M donne  $\theta_{P=O|E=O}^{(1)} = \frac{2,6}{2,6+3,4} = \frac{2,6}{6} \approx 0,433$ ,  $\theta_{P=N|E=O}^{(1)} = 1 - 0,433 = 0,567$  et

$$\theta_{P=O|E=N}^{(1)} = \frac{3,6}{3,6+5,4} = \frac{3,6}{9} = 0,4, \quad \theta_{P=N|E=N}^{(1)} = 1 - 0,4 = 0,6.$$

2<sup>ème</sup> itération :

Numéro scénario	Echec du contrôle	Perte	E = O		E = N	
			P = O	P = N	P = O	P = N
1	O	?	0,433	0,567	0	0
2	O	O	1	0	0	0
3	O	O	1	0	0	0
4	O	N	0	1	0	0
5	O	N	0	1	0	0
6	O	N	0	1	0	0
7	N	O	0	0	1	0
8	N	O	0	0	1	0
9	N	O	0	0	1	0
10	N	?	0	0	0,4	0,6
11	N	?	0	0	0,4	0,6
12	N	N	0	0	0	1
13	N	N	0	0	0	1
14	N	N	0	0	0	1
15	N	N	0	0	0	1
N*			2,433	3,567	3,8	5,2

L'étape M donne

$$\theta_{P=O|E=O}^{(2)} = \frac{2,433}{2,433 + 3,567} = \frac{2,433}{6} \approx 0,406, \quad \theta_{P=N|E=O}^{(2)} = 1 - 0,406 = 0,5945$$

$$\text{et } \theta_{P=O|E=N}^{(2)} = \frac{3,8}{3,8 + 5,2} = \frac{3,8}{9} \approx 0,422, \quad \theta_{P=N|E=N}^{(2)} = 1 - 0,422 = 0,578.$$

Répetons les étapes E et M encore deux fois pour converger vers les estimateurs :

$$\hat{\theta}_{P=O|E=O}^{MV+EM} = 0,4, \quad \hat{\theta}_{P=N|E=O}^{MV+EM} = 0,6 \quad \text{et} \quad \hat{\theta}_{P=O|E=N}^{MV+EM} = 0,428, \quad \hat{\theta}_{P=N|E=N}^{MV+EM} = 0,572.$$

### Apprentissage bayésien EM:

Il s'agit de l'apprentissage bayésien combiné à l'algorithme « expectation maximisation » (EM).

Soit  $D_o$  l'ensemble des données observées et  $D_m$  l'ensemble des données manquantes.

$$N_{i,j,k}^* = E_{\theta} [N_{i,j,k}] = N \times P(X_i = x_k, pa(X_i) = x_j | \theta^*)$$

$$\theta_{i,j,k}^{(t+1)} = \frac{N_{i,j,k}^*}{\sum_k N_{i,j,k}^*}$$

Exemple : reprenons l'exemple précédent dans le quel nous souhaitons estimer les probabilités  $P(P=O | E=O)$  et  $P(P=O | E=N)$ , que nous noterons respectivement

$$\theta_{P=O|E=O} \quad \text{et} \quad \theta_{P=O|E=N}.$$

Nous supposons un a priori sur la distribution des paramètres : prenons une distribution de Dirichlet avec  $\alpha_{i,j,k} = 1$  pour tous  $i, j, k \in \{1,2\}$ .

L'algorithme EM fonctionne avec des paramètres initiaux qui peuvent être issus d'un avis d'expert :

$$\theta_{P=O|E=O}^{(0)} = 0,6 \quad \text{et} \quad \theta_{P=O|E=N}^{(0)} = 0,3.$$

1<sup>ère</sup> itération :

Étape E :

Numéro scénario	Echec du contrôle	Perte	E = O		E = N	
			P = O	P = N	P = O	P = N
1	O	?	0,6	0,4	0	0
2	O	O	1	0	0	0
3	O	O	1	0	0	0
4	O	N	0	1	0	0
5	O	N	0	1	0	0
6	O	N	0	1	0	0
7	N	O	0	0	1	0
8	N	O	0	0	1	0
9	N	O	0	0	1	0
10	N	?	0	0	0,3	0,7
11	N	?	0	0	0,3	0,7
12	N	N	0	0	0	1
13	N	N	0	0	0	1
14	N	N	0	0	0	1
15	N	N	0	0	0	1
N*			2,6	3,4	3,6	5,4

L'étape M donne  $\theta_{P=O|E=O}^{(1)} = \frac{2,6+1}{2,6+1+3,4+1} = \frac{3,6}{8} = 0,45$ ,  $\theta_{P=N|E=O}^{(1)} = 1 - 0,45 = 0,55$  et

$$\theta_{P=O|E=N}^{(1)} = \frac{3,6+1}{3,6+1+5,4+1} = \frac{4,6}{11} \approx 0,418, \theta_{P=N|E=N}^{(1)} = 1 - 0,418 = 0,582.$$

2<sup>ème</sup> itération :

Numéro scénario	Echec du contrôle	Perte	E = O		E = N	
			P = O	P = N	P = O	P = N
1	O	?	0,45	0,55	0	0
2	O	O	1	0	0	0
3	O	O	1	0	0	0
4	O	N	0	1	0	0
5	O	N	0	1	0	0
6	O	N	0	1	0	0
7	N	O	0	0	1	0
8	N	O	0	0	1	0
9	N	O	0	0	1	0
10	N	?	0	0	0,418	0,582
11	N	?	0	0	0,418	0,582
12	N	N	0	0	0	1
13	N	N	0	0	0	1
14	N	N	0	0	0	1
15	N	N	0	0	0	1
N*			2,45	3,55	3,836	5,164

L'étape M donne

$$\theta_{P=O|E=O}^{(2)} = \frac{2,45+1}{2,45+1+3,55+1} = \frac{3,45}{8} \approx 0,431, \theta_{P=N|E=O}^{(2)} = 1-0,431 = 0,569$$

$$\text{et } \theta_{P=O|E=N}^{(2)} = \frac{3,836+1}{3,836+1+5,164+1} = \frac{4,836}{11} \approx 0,44, \theta_{P=N|E=N}^{(2)} = 1-0,44 = 0,56.$$

Répetons les étapes E et M encore un certain nombre de fois pour converger vers les estimateurs :

$$\hat{\theta}_{P=O|E=O}^{EAV+EM} = 0,429, \hat{\theta}_{P=N|E=O}^{EAV+EM} = 0,571 \text{ et } \hat{\theta}_{P=O|E=N}^{EAV+EM} = 0,444, \hat{\theta}_{P=N|E=N}^{EAV+EM} = 0,556.$$

### 5.3 DÉFINITION DE LA MATRICE DES RISQUES

Dans le secteur bancaire, le Comité de Bâle a établi une matrice 8\*7 de cellules de risques opérationnels (8 lignes métiers et 7 colonnes de types de risques). Dans chacune de ces 56 cellules, les institutions financières déterminent le montant de fonds propres puis les agrègent via une matrice de corrélation.

Dans Solvabilité II, rien n'est recommandé pour le moment mais il nous semble judicieux d'utiliser une telle matrice pour modéliser nos risques opérationnels de façon exhaustive. Il reste ensuite à définir la maille et le niveau de granularité de la matrice.

Nous proposons une matrice croisant les risques de niveau 1 avec les activités de l'entreprise. La base d'incidents étant définie sur plusieurs axes, dont les risques de niveau 1 et les activités, il sera alors facile de récupérer des informations par cellule.

Le travail de modélisation consistera à déterminer la distribution de perte totale pour chaque cellule et à les agréger.

Ce découpage doit permettre de modéliser des risques homogènes. Aussi il arrivera que nous soyons amenés à subdiviser une cellule qui ne contiendrait pas que des risques homogènes. La matrice croise les risques avec les activités. Chaque cellule peut contenir des risques de fréquence et de gravité. La matrice permet d'avoir une vision globale du risque et de pouvoir zoomer sur certains risques ou certaines activités. Pour obtenir le capital total de la matrice, les risques de gravité sont sommés (ils ont été simulés en même temps par construction) et les risques de fréquence sont agrégés par une matrice de corrélation ou par des copules.

#### Construction de la matrice :

Nous souhaitons construire une matrice par risque de niveau 1 et activités.

Pour les risques de fréquence, nous récupérerons les pertes observées dans chaque case.

Il est possible qu'il n'y ait pas assez de données dans une cellule, auquel cas nous devons agréger cette case avec une autre case.

Si les données d'une cellule sont en nombre suffisant, nous regarderons si elles sont homogènes par le biais de représentation graphique par exemple.

Lorsque les données d'une cellule sont homogènes, nous ajusterons une loi de fréquence et une loi de coût à nos données. Puis nous calculerons la distribution de charge totale par une des méthodes décrites plus haut. Nous pourrions alors calculer un montant de capital par case au titre des risques de fréquence.

Si les données d'une case ne sont pas homogènes, nous diviserons la case à un niveau de risque plus fin (c'est-à-dire au niveau 2 ici) et ainsi de suite si cela est nécessaire, jusqu'à obtenir des risques homogènes pour que nos adéquations de lois soient bien adaptées. Les distributions de pertes ajustées à des niveaux inférieurs seront ensuite agrégées jusqu'au niveau 1 en supposant l'indépendance des risques de niveau inférieur.

L'agrégation des cases permettra d'obtenir un capital global pour les risques de fréquence. Elle pourra être réalisée via des copules ou une matrice de corrélation.

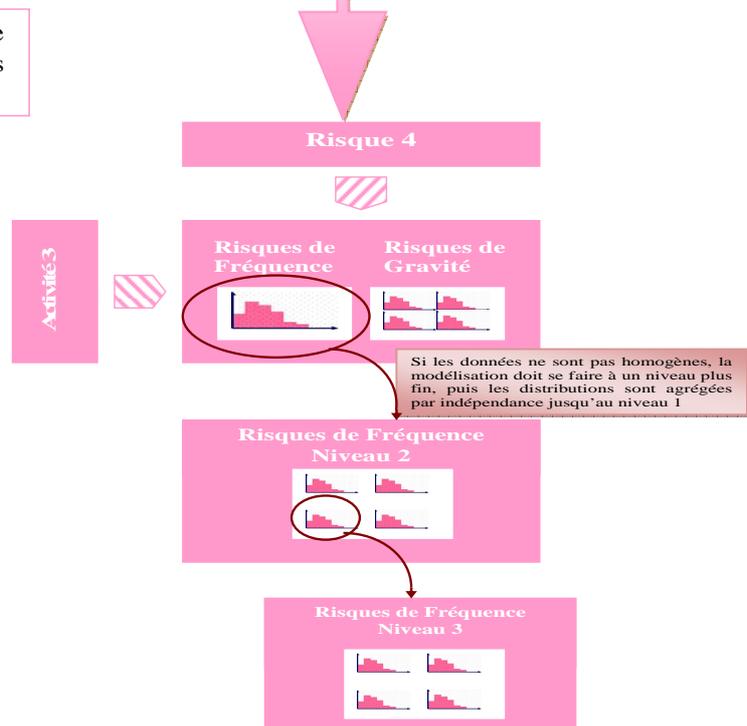
Les risques de gravité doivent également être modélisés de façon à attribuer une part de capital à chaque case au titre des risques de gravité.

Les risques de fréquence sont supposés indépendants des risques de gravité.

L'avantage d'une matrice par activité est la mise en place de « use tests » tels que des ratios de rentabilité par risque et par activité. La base d'incidents possède tous les axes d'analyse permettant de descendre au niveau d'analyse le plus fin dans la matrice.

		Risques de niveau 1							
		Risque 1	Risque 2	Risque 3	Risque 4	Risque 5	Risque 6	Risque 7	Risque 8
Activités	A1								
	A2								
	A3								
	A4								
	A5								
	A6								

Pas assez de données dans une case => regroupement avec des risques du même type



## 6 MODÉLISATION SIMPLIFIÉE DU RISQUE OPÉRATIONNEL

Modéliser le risque opérationnel n'est pas chose facile d'autant que cela nécessite une quantité très importante d'informations. Or, pour l'instant, nous ne disposons que de peu d'information. Nous allons proposer ici un modèle très simplifié et provisoire basé sur le SRA.

Chez Generali, les seules informations disponibles pour l'instant sont les avis d'experts sur les vingt-cinq risques opérationnels les plus importants issus du SRA (hors risques externes de type tempête, crue de la seine...).

Au vu de ce manque d'informations, le modèle proposé ici permettra d'avoir une première idée du besoin en capital en attendant de pouvoir récupérer d'avantages d'informations sur les différents indicateurs composant les scénarios.

L'idée générale du modèle est d'utiliser les trois scénarios de chacun des risques majeurs qui fournissent chacun une fréquence et un montant de pertes. Il s'agit d'ajuster avec ces six points une loi de fréquence et une loi de coût et de simuler la distribution de pertes de chacun de ces risques.

Le modèle consiste, pour chaque risque, à :

- d'une part, choisir une loi a priori du coût unitaire de la perte en estimant les paramètres par la méthode des quantiles et sélectionner une d'entre elles à partir d'un critère de décision ;
- et d'autre part, utiliser la loi de Poisson pour simuler la fréquence des coûts en estimant le paramètre par la fréquence moyenne des trois scénarios du SRA.

Notations utilisées dans le modèle :

Pour le scénario  $i$  et le risque  $k$  avec  $i \in \{B, M, H\}$  où B correspond à l'hypothèse basse, M à l'hypothèse moyenne et H à l'hypothèse haute et  $k \in [1; 25]$ , notons :

$\lambda_{k,i}$  fréquence issue du SRA

$c_{k,i}$  coût issu du SRA

$N_{k,i}$  variable aléatoire représentant le nombre de pertes

$X_{k,i,j}$  variable aléatoire représentant le coût de la perte  $j$

$S_{k,i}$  variable aléatoire représentant la charge totale des pertes

$U_k$  variable aléatoire de loi uniforme  $[0 ; 1]$

	Risque 1			...	Risque 25		
	Scénario Hypothèse Basse	Scénario Hypothèse Moyenne	Scénario Hypothèse Haute		Scénario Hypothèse Basse	Scénario Hypothèse Moyenne	Scénario Hypothèse Haute
Fréquence	$\lambda_{1,B}$	$\lambda_{1,M}$	$\lambda_{1,H}$	...	$\lambda_{25,B}$	$\lambda_{25,M}$	$\lambda_{25,H}$
Coût	$c_{1,B}$	$c_{1,M}$	$c_{1,H}$	...	$c_{25,B}$	$c_{25,M}$	$c_{25,H}$

1<sup>ière</sup> approche : Avant de développer le modèle, nous proposons une approche empirique qui consiste à ajuster une loi de fréquence mais pas de loi pour le coût.

Nous choisissons une loi de fréquence a priori en estimant ses paramètres à l'aide du SRA, et en utilisant les valeurs de pertes pour simuler le coût (le coût est constant).

Tous les scénarios d'un même risque peuvent survenir en même temps.

La loi de fréquence que nous avons choisie est la loi de Poisson de paramètre  $\lambda$  égal à la fréquence indiquée dans le scénario.

Pour tout  $i, k$ , on a :

$N_{k,i}$  suit la loi de Poisson  $P(\lambda_{k,i})$

$X_{k,i,j} = c_{k,i}$  pour tout  $j$

La charge totale  $S_{k,i}$  du risque  $k$  dans le scénario  $i$  est alors  $S_{k,i} = c_{k,i} \times N_{k,i}$

La loi de  $S_{k,i}$  est à valeurs dans  $c_{k,i} N$ , on a alors :

$$P(S_{k,i} = n \times c_{k,i}) = P(N_{k,i} \times c_{k,i} = n \times c_{k,i}) = P(N_{k,i} = n) = \frac{\lambda_{k,i}^n}{n!} \times \exp(-\lambda_{k,i}) \quad \text{pour } n \in \mathbb{N}$$

2<sup>nde</sup> approche : nous choisissons une loi de sévérité a priori en estimant les paramètres avec la méthode des quantiles (décrite plus bas) en supposant que :

➤ la perte issue du scénario moyen donne le quantile à  $f_M := \frac{\lambda_{k,B} + \lambda_{k,M}}{2}$  :

$$f_M := \frac{\lambda_{k,B} + \lambda_{k,M}}{2} : \sum_{i \in \{B, M, H\}} \lambda_{k,i}$$

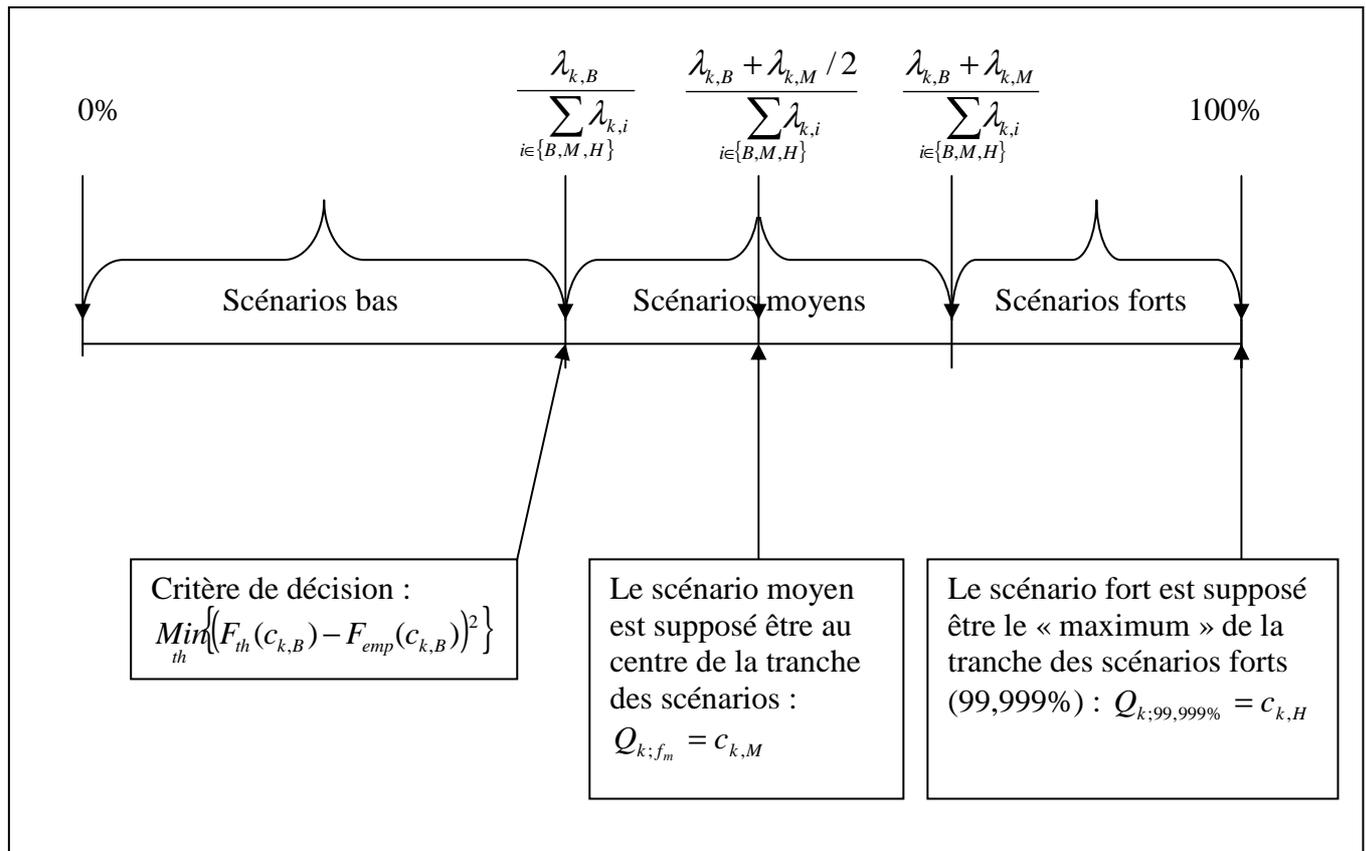
$$Q_{k;f_m} = c_{k,M}$$

➤ la perte issue du scénario fort donne le quantile à 99,999% :  $Q_{k;99,999\%} = c_{k,H}$ .

Puis, nous utilisons un critère basé sur le scénario hypothèse basse pour départager les lois :

$$\text{Min}_{th} \left\{ \left( F_{th}(c_{k,B}) - F_{emp}(c_{k,B}) \right)^2 \right\} \quad \text{où} \quad F_{emp}(c_{k,B}) = \frac{\lambda_{k,B}}{\sum_{i \in \{B, M, H\}} \lambda_{k,i}} \quad \text{et} \quad F_{th} \quad \text{est la fonction de}$$

répartition empirique testée.



La fréquence est modélisée avec une loi de Poisson dont le paramètre est la somme des fréquences du risque issues du SRA :

$$\lambda_k = \sum_i \lambda_{k,i}$$

Nous nous ramenons ainsi à un problème classique en assurance de détermination de la charge totale et nous pouvons utiliser les méthodes décrites précédemment.

Remarque : cette méthode est très sensible à la précision choisie après la virgule pour le quantile extrême à 99,999%. Nous avons réalisé des tests de sensibilités des résultats.

Décrivons la méthode des quantiles qui nous permet de calibrer notre loi de coût.

Cette méthode consiste à estimer les paramètres d'une loi à l'aide de quantiles car ceux-ci s'expriment en fonction des paramètres de la loi.

Soient  $n$  réalisations de  $X$  qui suivent la même loi.

Posons  $F_\theta(x) = F(x|\theta)$  la fonction de répartition de cette loi avec  $\theta^T = (\theta_1, \theta_2, \dots, \theta_p)$

Où  $\theta^T$  est le transposé de  $\theta$  qui est un vecteur colonne contenant  $p$  paramètres à estimer.

Posons  $Q_g(\theta)$  le quantile à  $g\%$  de la loi  $F_\theta$ .

Nous avons par conséquent :  $F[Q_g(\theta)|\theta] = g$  (si la fonction de répartition est continue il y a au moins une solution à cette équation).

Posons  $\hat{Q}_g$  un estimateur du quantile à  $g\%$ .

Un estimateur par la méthode des quantiles de  $\theta$  est une solution de  $p$  équations :

$$Q_{g_k}(\theta) = \hat{Q}_{g_k}, k = 1, 2, \dots, p$$

Où  $g_1, g_2, \dots, g_p$  sont  $p$  quantiles choisis arbitrairement.

Par définition des quantiles, les équations peuvent encore être écrites de la façon suivante :

$$F(\hat{Q}_{g_k} | \theta) = g_k, k = 1, 2, \dots, p$$

Les limites de cette méthode est qu'il n'y a aucune garantie de l'existence et de l'unicité d'une solution.

Nous avons étudié 4 lois de sévérité : Exponentielle, Log-normale, Weibull et Pareto.

#### Loi Exponentielle :

La fonction de répartition de la loi Exponentielle est

$$F(x) = (1 - \exp(-\lambda x)) \times 1_{\mathbb{R}^+}(x) \quad \text{pour } \lambda > 0$$

$$Q_g(\lambda), \text{ le quantile à } g\%, \text{ est égal à : } Q_g(\lambda) = -\frac{1}{\lambda} \times \ln(1 - g)$$

Nous pouvons en déduire un estimateur du paramètre en inversant la formule et en remplaçant  $Q_g(\lambda)$  par  $\hat{Q}_g$  le quantile estimé :

$$\hat{\lambda} = -\frac{1}{\hat{Q}_g} \times \ln(1 - g)$$

#### Loi de Weibull :

La fonction densité d'une loi de Weibull se présente sous la forme :

$$f(x) = \frac{a}{b} \times \left(\frac{x}{b}\right)^{a-1} \times \exp\left(-\left(\frac{x}{b}\right)^a\right) \quad \text{pour } a \text{ et } b > 0$$

$$\text{Et sa fonction de répartition : } F(x) = 1 - \exp\left(-\left(\frac{x}{b}\right)^a\right)$$

Nous en déduisons le quantile théorique à  $g\%$  :  $Q_g = b \times (-\ln(1 - g))^{\frac{1}{a}}$

Pour estimer les deux paramètres  $a$  et  $b$ , nous devons donc utiliser au moins deux équations :

$$Q_{g_1} = b \times (-\ln(1 - g_1))^{\frac{1}{a}} \text{ et } Q_{g_2} = b \times (-\ln(1 - g_2))^{\frac{1}{a}}, \quad g_1 \neq 1, g_2 \neq 1$$

$$\text{Soit } \frac{Q_{g_1}}{Q_{g_2}} = \left( \frac{\ln(1 - g_1)}{\ln(1 - g_2)} \right)^{\frac{1}{a}} \text{ et } \ln\left(\frac{Q_{g_1}}{Q_{g_2}}\right) = \frac{1}{a} \times \ln\left(\frac{\ln(1 - g_1)}{\ln(1 - g_2)}\right)$$

Les estimateurs des paramètres  $a$  et  $b$  s'en déduisent :

$$\hat{a} = \frac{\ln\left(\frac{\ln(1 - g_1)}{\ln(1 - g_2)}\right)}{\ln\left(\frac{\hat{Q}_{g_1}}{\hat{Q}_{g_2}}\right)} \text{ et } \hat{b} = \frac{\hat{Q}_{g_1}}{(-\ln(1 - g_1))^{\frac{1}{\hat{a}}}}$$

### Loi Log-normale :

La fonction de répartition d'une loi Log-normale de paramètres  $\mu$  et  $\sigma$  est :

$$F(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right) \text{ pour } \sigma > 0 \text{ et où } \Phi(x) = \int_{-\infty}^x \exp\left(-\frac{y^2}{2}\right) \frac{dy}{\sqrt{2\pi}}$$

Nous en déduisons une expression du logarithme du quantile :

$$\ln(Q_g) = \Phi^{-1}(g) \times \sigma + \mu$$

Pour estimer les deux paramètres  $\mu$  et  $\sigma$ , nous devons utiliser au moins deux équations :

$$\ln(Q_{g_1}) = \Phi^{-1}(g_1) \times \sigma + \mu \text{ et } \ln(Q_{g_2}) = \Phi^{-1}(g_2) \times \sigma + \mu$$

$$\ln(Q_{g_1}) - \ln(Q_{g_2}) = \sigma \times (\Phi^{-1}(g_1) - \Phi^{-1}(g_2))$$

Les estimateurs des paramètres  $\mu$  et  $\sigma$  s'en déduisent :

$$\hat{\sigma} = \frac{\ln(\hat{Q}_{g_1}) - \ln(\hat{Q}_{g_2})}{\Phi^{-1}(g_1) - \Phi^{-1}(g_2)} \text{ et } \hat{\mu} = \ln(\hat{Q}_{g_1}) - \Phi^{-1}(g_1)$$

### Loi Pareto :

La fonction de répartition d'une loi Pareto de paramètres  $\lambda$  et  $\alpha$  s'écrit :

$$F(x) = 1 - \left(\frac{\lambda}{\lambda + x}\right)^\alpha$$

Nous en déduisons l'expression analytique des quantiles à  $g_1\%$  et  $g_2\%$  :

$$Q_{g_1} = \lambda \times \left( (1 - g_1)^{-\frac{1}{\alpha}} - 1 \right) \text{ et } Q_{g_2} = \lambda \times \left( (1 - g_2)^{-\frac{1}{\alpha}} - 1 \right)$$

Les estimateurs des paramètres  $\lambda$  et  $\alpha$  s'en déduisent :

$$\frac{\hat{Q}_{g_1}}{\hat{Q}_{g_2}} = \frac{(1 - g_1)^{-\frac{1}{\hat{\alpha}}} - 1}{(1 - g_2)^{-\frac{1}{\hat{\alpha}}} - 1} \text{ et } \hat{\lambda} = \frac{\hat{Q}_{g_1}}{(1 - g_1)^{-\frac{1}{\hat{\alpha}}} - 1}$$

## 7 APPLICATION DE LA MODÉLISATION DES RISQUES DE FRÉQUENCE

L'objectif de ce chapitre est de montrer la méthodologie à adopter pour modéliser des risques à partir d'une base de données. En particulier, nous utiliserons la base pour montrer comment s'applique la méthode LDA classique et comment gérer le problème de seuil de collectes. Nous montrerons également comment utiliser la théorie des valeurs extrêmes pour modéliser les pertes extrêmes. Nous étudierons enfin l'impact d'une agrégation par les copules.

Nous rappelons que nous avons récupéré une base issue du secteur bancaire où les risques sont assez similaires à ceux de l'assurance.

Il s'agit d'une partie de l'historique d'une banque française sur 4 ans. Cette base comporte les variables suivantes : le numéro de la perte, le début de la perte (date à laquelle la perte est survenue), la fin de la perte (date de clôture de la perte), la constatation de la perte (date de déclaration de la perte), la catégorie de risque (niveau 1 du référentiel) et les pertes réelles (montant de perte payé).

Voici un extrait de la base utilisée :

Numéro de la perte	Début de la perte	Fin de la perte	Constatation de la perte	Catégorie de risque	Pertes réelles
0000000000000330	20070115	20070115	20070115	PROCES	147,95
0000000000000338	20070118	0	20070118	PROCES	576 620,22
0000000000000342	20070119	20070119	20070119	PROCES	501,00
0000000000000345	20061111	20070123	20061111	PROCES	12,20
0000000000000347	20061208	20061214	20061208	PROCES	247,38
0000000000000348	20061108	20061108	20061108	PROCES	45,66
0000000000000350	20061229	20061229	20061229	PROCES	2 401,21
0000000000000352	20061213	20061214	20061213	PROCES	181,26
0000000000000353	20061127	20061208	20061127	PROCES	515,41
0000000000000354	20061215	20061219	20061215	PROCES	3 194,02
0000000000000355	20060407	20060407	20060407	PROCES	233,59
0000000000000356	20061213	20061214	20061213	CLIENT	5 795,84
0000000000000358	20061115	20061115	20061115	PROCES	611,91
0000000000000359	20061214	20061214	20061214	PROCES	78,25
0000000000000360	20061220	20061229	20061220	PROCES	44,65
0000000000000361	20061229	20061229	20061229	SYSTEM	872,70
0000000000000367	20070118	20070125	20070118	PROCES	86,46
0000000000000369	20070122	20070123	20070122	PROCES	256,00
0000000000000380	20070110	20070110	20070110	PROCES	2 196,72

Les montants sont exprimés en euros. L'établissement bancaire remonte ses pertes dès le premier euro, afin de modéliser de manière fine sa fréquence. Les pertes dont les montants sont très petits (inférieur à 5 euros) correspondent pour l'essentiel à des

incidents de process de traitement d'ordres, qui sont détectés et remontés automatiquement.

#### Composition des risques :

Cette base fournit la catégorie de risque au niveau 1 et ne donne pas le détail des différentes sous catégories de risques. Cela va impliquer une certaine hétérogénéité dans nos données. L'idéal aurait été de disposer d'un niveau plus fin pour modéliser des risques homogènes mais encore aurait-il fallu avoir assez de données dans chaque sous-catégorie. Tous les risques de niveau 1 ne ressortent pas dans cette base, en effet il n'apparaît que 4 risques parmi les 7 risques du référentiel de Bâle 2 : le risque clients, le risque de dysfonctionnement de l'activité et des systèmes, le risque de gestion des processus et le risque de fraude externe.

Les risques auraient également dû être décomposés par ligne métier afin de les modéliser de façon homogène, là encore nous ne disposons pas de l'information.

Cette étude a été réalisée sous le logiciel R avec les packages « vcd » pour LDA classique, « truncgof » pour LDA avec seuil de collectes, « evir » pour la théorie des valeurs extrêmes et « copula » pour les copules.

#### Données as if :

La modélisation LDA s'appuie sur l'analyse des pertes observées dans le passé. Si la période d'observation est longue, il y a lieu d'actualiser les montants observés pour qu'ils correspondent à des montants qui pourraient être observés aujourd'hui. Les risques opérationnels sont des risques sur les objets exposés de l'entreprise, ils sont donc différents des risques assurantiels que nous avons l'habitude de voir. La vision ultime de la perte est soumise à une incertitude plus grande.

De plus, nous ne connaissons pas les sous risques qui composent notre base. Par conséquent, appliquer un coefficient d'inflation générale aux montants de pertes historisées serait inadéquat. En effet, l'actualisation doit être spécifique à chaque catégorie de risque et nous ne disposons pas d'informations suffisantes sur le type de risque pour lui appliquer le coefficient d'inflation adéquat.

Nous choisissons donc de ne pas rendre nos données « as if ».

#### Hypothèses LDA :

Nous souhaitons appliquer ici la méthode LDA. Cette méthode repose sur l'hypothèse que les variables de coût individuel sont indépendantes et identiquement distribuées, i.e. chaque perte survient indépendamment des précédentes et elles suivent la même loi. De plus, elle suppose que la fréquence et le coût sont indépendants.

Il semble raisonnable de vérifier avant exploitation des données que l'hypothèse « iid » n'est pas grossièrement fautive. Pour cela, nous pouvons simplement examiner visuellement la suite des coûts individuels ou faire appel à des tests statistiques.

Les tests d'indépendance entre la fréquence et le coût individuel sont peu robustes en raison du faible nombre d'observations annuelles. Il faut alors se ramener à des observations mensuelles du coût moyen des sinistres et de la fréquence.

Tests d'échantillon iid:

Soit  $(X_1, X_2, \dots, X_n)$  une suite finie de variables aléatoires réelles telles que

$$P(X_i = X_j) = 0 \quad \text{pour } i \neq j$$

Pour un niveau de confiance donné, les hypothèses testées sont les suivantes :

$$H_0 : X_1, X_2, \dots, X_n \text{ v.a.r. i.i.d. contre } H_1 : X_1, X_2, \dots, X_n \text{ v.a.r. non i.i.d.}$$

Dans une suite chronologique de  $n$  observations  $(x_1, x_2, \dots, x_n)$ , nous nous intéressons à la différence entre deux observations successives :  $x_i - x_{i-1}$ .

Nous avons ainsi  $n-1$  différences :  $(x_2 - x_1, x_3 - x_2, \dots, x_{n-1} - x_n)$

Test des signes (voir page 155 [15]) : ce test s'intéresse au nombre de différences positives successives de notre échantillon.

Soit  $S$  le nombre de différences positives :  $S = \sum_{i=1}^{n-1} 1_{\mathbb{R}^+}(X_i - X_{i+1})$

Sous l'hypothèse  $H_0$ , nous avons :

$$\frac{S - E(S)}{\sigma(S)} \xrightarrow{n \rightarrow +\infty} N(0,1) \quad \text{avec } E(S) = \frac{n-1}{2} \text{ et } \sigma^2(S) = \frac{n+1}{12}$$

L'approximation normale est valable dès que  $n \geq 12$ .

Test du nombre de points de retournement (voir page 159 [15]) : ce test s'intéresse au nombre de points de retournement de notre échantillon c'est-à-dire aux points  $X_i$  tels que  $X_{i-1}$  et  $X_{i+1}$  sont tous deux inférieurs ou supérieurs à  $X_i$ .

$$\text{Soit } Z_i = \begin{cases} 1 & \text{si } (X_{i+1} - X_i)(X_i - X_{i-1}) < 0 \\ 0 & \text{sin on} \end{cases} \quad \text{pour } i \in \{2, \dots, n-1\}$$

Soit  $R$  la variable aléatoire réelle représentant le nombre de points de retournement :

$$R = \sum_{i=2}^{n-1} Z_i$$

Sous l'hypothèse  $H_0$ , nous avons :

$$\frac{R - E(R)}{\sigma(R)} \xrightarrow{n \rightarrow +\infty} N(0,1) \quad \text{avec } E(R) = \frac{2}{3}(n-2) \text{ et } \sigma^2(R) = \frac{16n-29}{90}$$

L'approximation normale est valable dès que  $n \geq 50$ .

Test d'indépendance entre la fréquence et le coût individuel :

Soient  $F_1, F_2, \dots, F_n$  et  $X_1, X_2, \dots, X_n$  les mesures de deux variables F et X sur n individus.

Supposons que les  $F_i$  sont iid de loi  $P_F$  et les  $X_i$  sont iid de loi  $P_X$ .

Test des rangs de Spearman :

Ce test est basé sur la corrélation entre les rangs de F et ceux de X.

Notons  $R_i$  le rang de  $F_i$  et  $S_i$  le rang de  $X_i$ .

Les hypothèses testées sont  $H_0$  : F et X sont indépendants

$H_1$  : F et X sont liées

Soit  $D_i = S_i - R_i$  l'écart entre les rangs, la statistique du test est alors

$$T = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n D_i^2$$

T est une statistique tabulée sous  $H_0$ .

## 7.1 STATISTIQUES DESCRIPTIVES DES DONNÉES

Dans un premier temps, nous allons étudier les données de façon à avoir une première idée de la forme des distributions empiriques de la fréquence et du coût individuel des pertes.

Voici les statistiques descriptives du montant des pertes (les montants sont exprimés en euros) :

	Risque Client	Risque de processus	Risque de système	Risque de fraude externe
Minimum	2,12	0,12	0,44	8
1 <sup>er</sup> quartile	71,02	31,92	29,32	965,1
Médiane	673,1	190,8	304,9	1 520
3 <sup>ème</sup> quartile	3 455	860,8	612,5	2 632
Maximum	1 000 000	576 600	261 200	27 850
Moyenne	30 270	2 494	9 782	2 642
Ecart-type	154 118,6	26 783,12	45 187,38	5 131,6
Coefficient d'aplatissement	37,42	434,38	27,47	20,82
Coefficient d'asymétrie	5,94	20,41	4,97	4,20
Nombre de pertes	42	487	35	28

Statistiques descriptives du nombre de pertes par mois de survenance :

	Risque Client	Risque de processus	Risque de système	Risque de fraude externe
Moyenne	0,875	10,146	0,729	0,583
Médiane	1	8	0	0
Variance	1,133	51,957	1,095	1,014
Minimum	0	0	0	0
Maximum	5	36	3	6

Statistiques descriptives du nombre de pertes par année :

	Risque Client	Risque de processus	Risque de système	Risque de fraude externe
2004	4	45	4	7
2005	8	100	6	11
2006	12	130	9	6
2007	18	212	16	4
Moyenne	10,5	121,75	8,75	7
Médiane	10	115	7,5	6,5
Variance	35,67	4 858,92	27,58	8,67
Minimum	4	45	4	4
Maximum	18	212	16	11

Nous remarquons que la moyenne la plus élevée de 30 270€ est celle correspondant au risque clients avec une fréquence moyenne de 10 pertes annuelles : en effet, ces pertes sont généralement dues à des poursuites juridiques faites par les clients et peuvent s'avérer très élevées. Il s'agit d'un risque classifié comme étant à faible fréquence et fort impact. Alors que la perte moyenne la plus faible est celle du risque processus qui est de l'ordre de 2 500€ avec une fréquence moyenne de plus de 120

pertes annuelles. Ainsi, le risque processus constitue un risque à fréquence élevée et sévérité faible.

Ce constat confirme le fait que les montants de pertes par type de risque présentent des caractéristiques différentes justifiant la modélisation par type de risque. Nous pouvons même dire qu'il faudrait subdiviser la modélisation en sous-catégorie mais ce n'est pas possible ici car nous ne disposons pas de l'information.

Par ailleurs, nous constatons que les données de pertes sont loin de présenter une distribution symétrique. En effet, une simple comparaison entre la moyenne et la médiane nous indique que la médiane est, dans tous les cas, beaucoup moins élevée que la moyenne, montrant ainsi des distributions asymétriques. Ce constat est confirmé par le calcul du coefficient d'asymétrie qui est positif dans tous les cas favorisant ainsi des queues plus épaisses. De plus, des coefficients d'aplatissement très élevés montrent l'existence de distribution leptokurtique<sup>5</sup> à queues épaisses. Citons l'exemple du type de risque processus où le coefficient d'aplatissement est de l'ordre de 434 et le coefficient d'asymétrie est évalué à 20,41. En effet, nous remarquons que 97,5 % des pertes sont inférieures à 10 000€.

Comme le montrent ces statistiques, les pertes opérationnelles ne peuvent en aucun cas être modélisées par des distributions normales et il y a lieu de tester d'autres distributions plus adaptées.

Application des tests à notre échantillon :

Test des signes :

	Client	Processus	Fraude	Système
S	17	237	18	16
E(S)	20,5	243	13,5	17
V(S)	3,6	40,7	2,4	3,0
S-E(S)-1/2	-1,8	-6,5	2,9	-0,6
S-E(S)-1/2	1,8	6,5	2,9	0,6
sigma(S)*N(0,1)	3,7	12,5	3,0	3,4
<b>Test</b>	<b>hors zone critique</b>	<b>hors zone critique</b>	<b>hors zone critique</b>	<b>hors zone critique</b>

<sup>5</sup> Correspond à un coefficient d'aplatissement (ou Kurtosis) >3

Test du nombre de points de retournement :

	Client	Processus	Fraude	Système
R	21	240	16	18
E(R)	26	322,7	16,7	21,3
V(R)	7,1	86,3	4,7	5,9
$R-E(R)-1/2$	-5,5	-83,2	-1,2	-3,8
$c'=\sigma(R)*N(0,1)$	-4,4	-15,3	-3,5	-4,0
$c=c'+E(R)+1/2$	22,1	307,9	13,6	17,8
<b>Test</b>	<b>zone critique</b>	<b>zone critique</b>	<b>hors zone critique</b>	<b>hors zone critique</b>

Test de corrélation de Spearman :

	Client	Processus	Fraude	Système
n	48	48	48	48
Somme( $D_i^2$ )	2196	14118	1415	719
rs	0,9	0,2	0,9	1,0
$\sigma(rs)*N(0,1)$	0,3	0,3	0,3	0,3
<b>Test</b>	<b>zone critique</b>	<b>hors zone critique</b>	<b>zone critique</b>	<b>zone critique</b>

Le rejet de l'hypothèse  $H_0$  pourrait nous conduire à subdiviser les risques en sous-catégories mais nous ne pouvons rien faire puisque nous n'avons pas l'information. Par contre, c'est un élément qui sera certainement utilisé sur notre base de données assurantielles pour déterminer si les risques sont bien homogènes.

## 7.2 ESTIMATION DES DISTRIBUTIONS DE SÉVÉRITÉ

Nous souhaitons ajuster une loi à nos échantillons de coût. Pour cela, nous sommes limités à quatre distributions qui seraient susceptibles de convenir a priori. Nous aurions pu tester bien d'autres lois mais notre objectif est de présenter la méthodologie. Pour chaque risque nous avons estimé les paramètres des différentes lois via la fonction « fitdistr » de R qui utilise soit la méthode du maximum de vraisemblance soit la méthode des moments selon le paramétrage.

Pour avoir une première idée sur le niveau d'ajustement, nous avons réalisé des tests graphiques tels que le graphique des quantiles (connu également sous le nom « Q-Q plot ») et le graphique des fonctions de répartition empirique et théorique. Les tests graphiques sont ensuite complétés par les tests statistiques de Kolmogorov-Smirnov et d'Anderson-Darling.

### Q-Q plot :

Le graphique Q-Q plot permet de comparer visuellement un échantillon et une distribution théorique. Ce graphique représente les quantiles de la distribution théorique en fonction des quantiles de notre échantillon.

Soient  $(x_1, x_2, \dots, x_n)$  un échantillon et  $(x_{(1)}, x_{(2)}, \dots, x_{(n)})$  l'échantillon ordonné  $(x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)})$ .

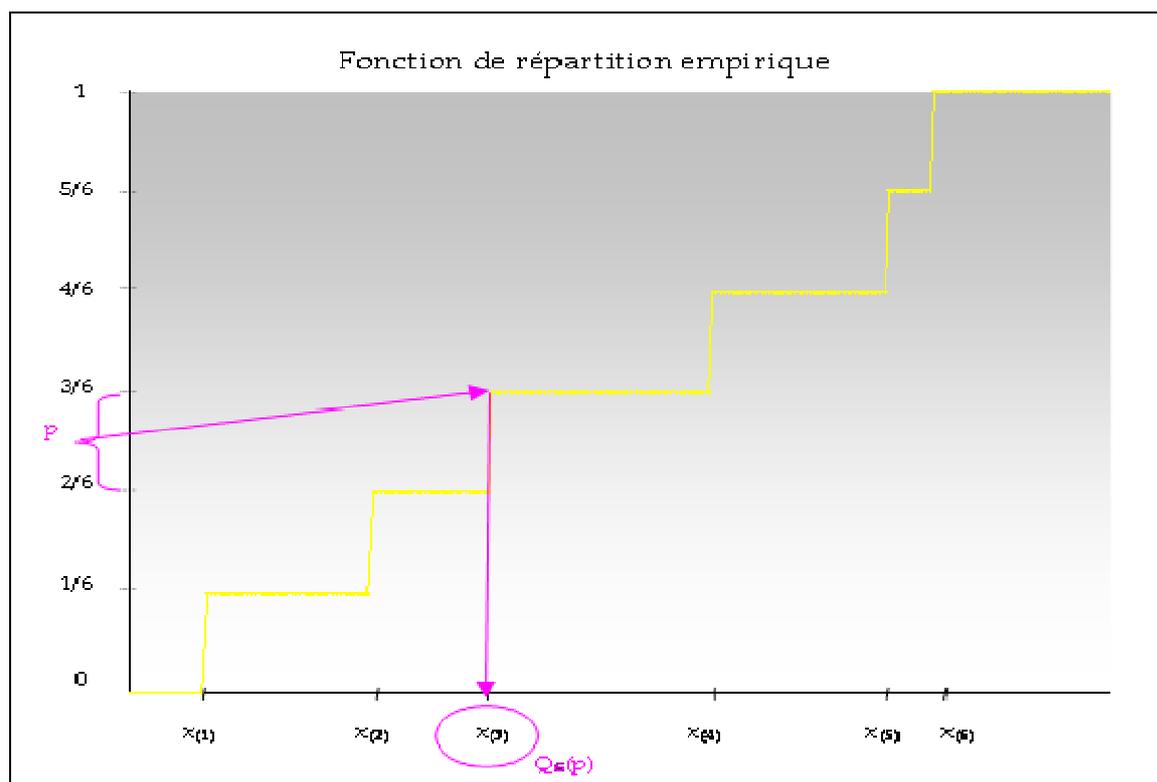
La fonction de répartition empirique  $F_n$  est définie par :

$$F_n(x_{(i)}) = \frac{i}{n} \quad \text{pour tout } i \in \{1, \dots, n\}$$

Dans le cas général, la fonction quantile  $Q$  est définie comme l'inverse de la fonction de répartition  $F$  :  $Q(p) = \inf\{x \text{ tel que } F(x) \geq p\}$  pour tout  $p \in ]0; 1[$

La fonction quantile empirique  $Q_n$  correspond à la fonction quantile de la fonction de répartition empirique.

$$\text{pour tout } p \in \left] \frac{i-1}{n}; \frac{i}{n} \right] \quad Q_n(p) = x_{(i)}$$



Le graphique ci-dessus représente la fonction de répartition empirique d'un échantillon de taille 6 et donne le quantile empirique d'ordre  $p$  lorsque  $p \in \left] \frac{2}{6}; \frac{3}{6} \right]$ .

Au vu de cette définition, le Q-Q plot représente les quantiles de la distribution théorique en fonction de notre échantillon ordonné c'est-à-dire qu'il représente les points de coordonnées  $\left( x_{(i)}, Q\left(\frac{i}{n}\right) \right)$  pour tout  $i \in \{1, \dots, n\}$ .

En particulier, pour la loi exponentielle nous placerons les points  $\left( x_{(i)}, -\ln\left(1 - \frac{i}{n}\right) \right)$  pour tout  $i \in \{1, \dots, n\}$

Pour la loi Log-normale :  $\left( \ln(x_{(i)}), \Phi^{-1}\left(\frac{i}{n}\right) \right)$  pour tout  $i \in \{1, \dots, n\}$

Pour la loi Gamma :  $\left( x_{(i)}, \chi_{2i}^2\left(\frac{i}{n}\right) \right)$  pour tout  $i \in \{1, \dots, n\}$

Pour la loi Weibull :  $\left( \ln(x_{(i)}), \ln\left[-\ln\left(1 - \frac{i}{n}\right)\right] \right)$  pour tout  $i \in \{1, \dots, n\}$

Il existe sous le logiciel R la fonction « qqplot » que nous avons utilisée pour réaliser nos graphes Q-Q plot.

Test de Kolmogorov-Smirnov :

Le test de Kolmogorov-Smirnov est un test d'hypothèse utilisé pour déterminer si un échantillon suit bien une loi continue donnée.

Soit  $(x_1, \dots, x_n)$  notre échantillon supposé indépendant et identiquement distribué de loi inconnue  $P_X$ .

Soit  $F_0$  est la fonction de répartition d'une loi continue donnée.

L'hypothèse nulle sera alors :

$H_0$  : La loi  $P_X$  a pour fonction de répartition  $F_0$

Il faut d'abord calculer la distance maximale notée  $KS_{obs}$  entre l'échantillon et la loi donnée  $F_0$  via la statistique de Kolmogorov-Smirnov KS.

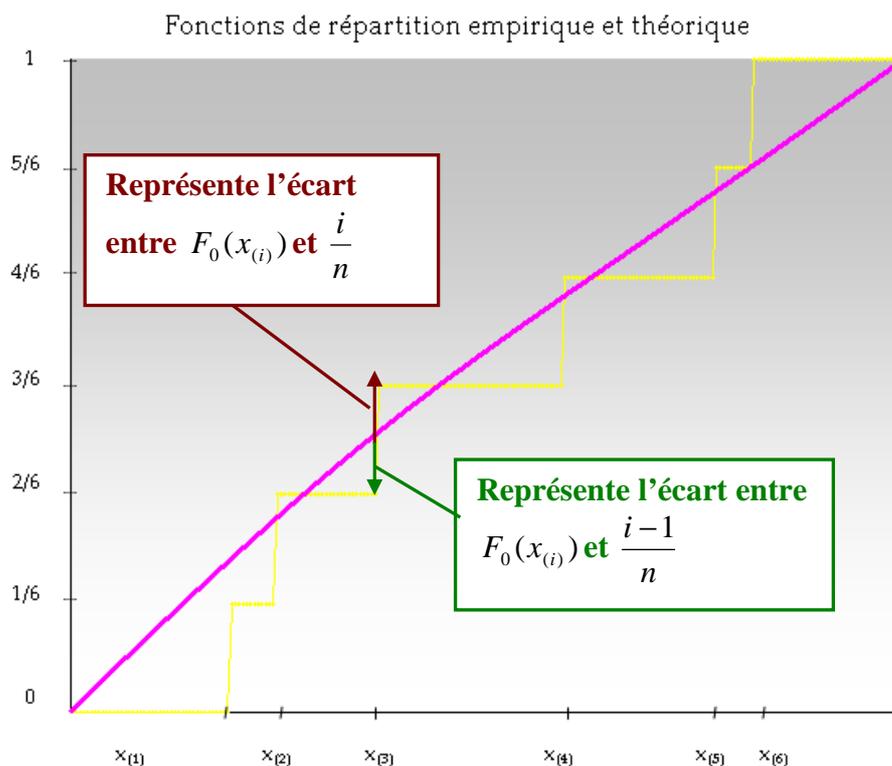
$KS = \sup_x |F_0(x) - F_n(x)|$  Où  $F_n$  est la fonction de répartition empirique pour un échantillon donné

Pour calculer  $KS_{obs}$ , il faut remplacer  $F_n$  par la fonction de répartition de notre

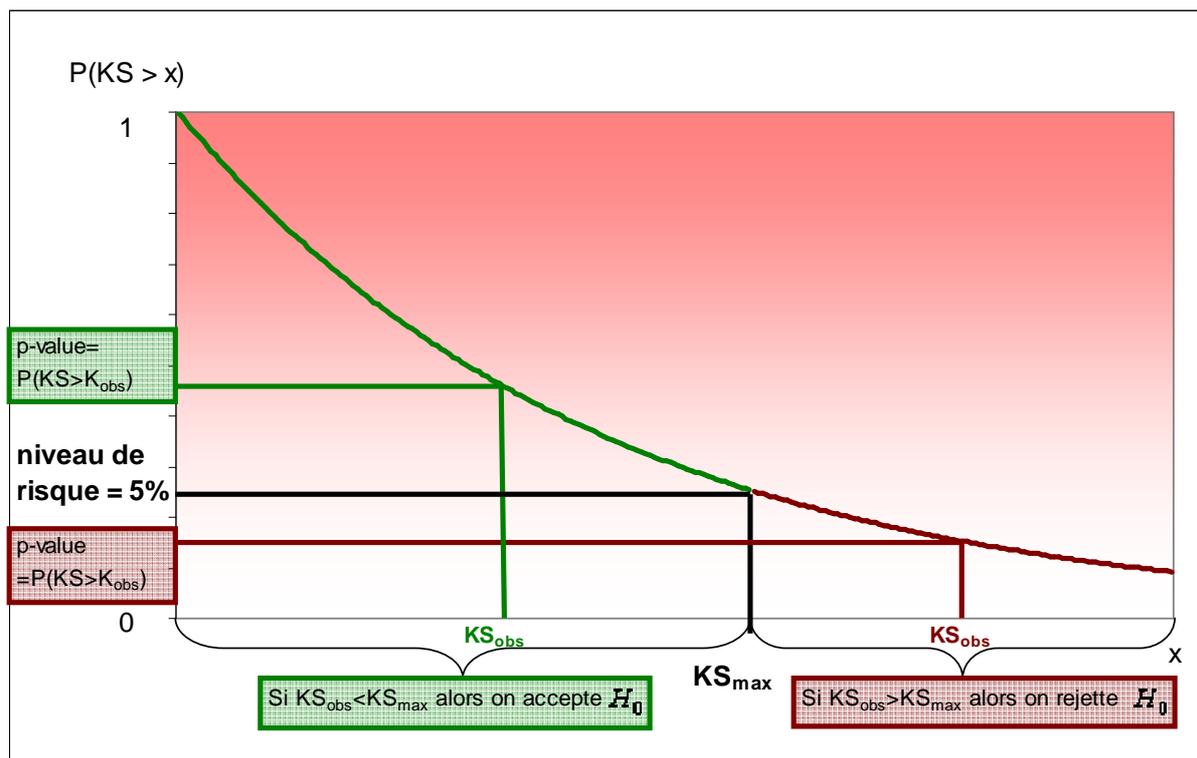
échantillon  $F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{x_{(i)} \leq x}$

Ce qui donne :  $KS_{obs} = \max_{i=1, \dots, n} \left\{ \left| F_0(x_{(i)}) - \frac{i}{n} \right|, \left| F_0(x_{(i)}) - \frac{i-1}{n} \right| \right\}$

Le graphique ci-dessous représente le calcul de  $KS_{obs}$



Le graphique ci-dessous représente la fonction de survie de la statistique KS



Il est nécessaire de se fixer un niveau de risque de première espèce qui est le risque de rejeter à tort l'hypothèse nulle lorsqu'elle est vraie. Souvent, un risque de 5% est considéré comme acceptable (c'est-à-dire que dans 5% des cas quand  $H_0$  est vraie, elle sera rejetée).

En supposant que  $H_0$  est vraie, il faut ensuite calculer la p-value qui est la probabilité d'obtenir une valeur de la variable de décision  $KS_{max}$  au moins aussi grande que la valeur de la statistique que l'on a obtenue avec notre échantillon  $KS_{obs}$ .

On a  $p\text{-value} = P(KS > KS_{obs})$ .

Cette p-value peut se lire dans la table de la statistique KS.

Si la p-value est plus grande que le niveau de risque, l'hypothèse  $H_0$  est acceptée. Si la p-value est plus petite que le niveau de risque alors elle est rejetée.

Remarque : La probabilité pour que  $H_0$  soit acceptée alors qu'elle est fautive est le risque de deuxième espèce. C'est le risque de ne pas rejeter  $H_0$  quand on devrait la rejeter. Sa valeur dépend du contexte, et est très difficilement évaluable (voire impossible à évaluer), c'est pourquoi seul le risque de première espèce est utilisé comme critère de décision.

Le test de Kolmogorov est facilement applicable sous le logiciel R avec la fonction « ks.test ».

Test d'Anderson-Darling :

Ce test a exactement le même principe que le test de Kolmogorov-Smirnov mise à part qu'il s'appuie sur la statistique d'Anderson-Darling :

$$AD = \left\{ - \left[ \sum_{i=1}^n (2i-1) \{ \ln(F_0(x_{(i)})) + \ln(1 - F_0(x_{(n+1-i)})) \} \right] / n \right\} - n$$

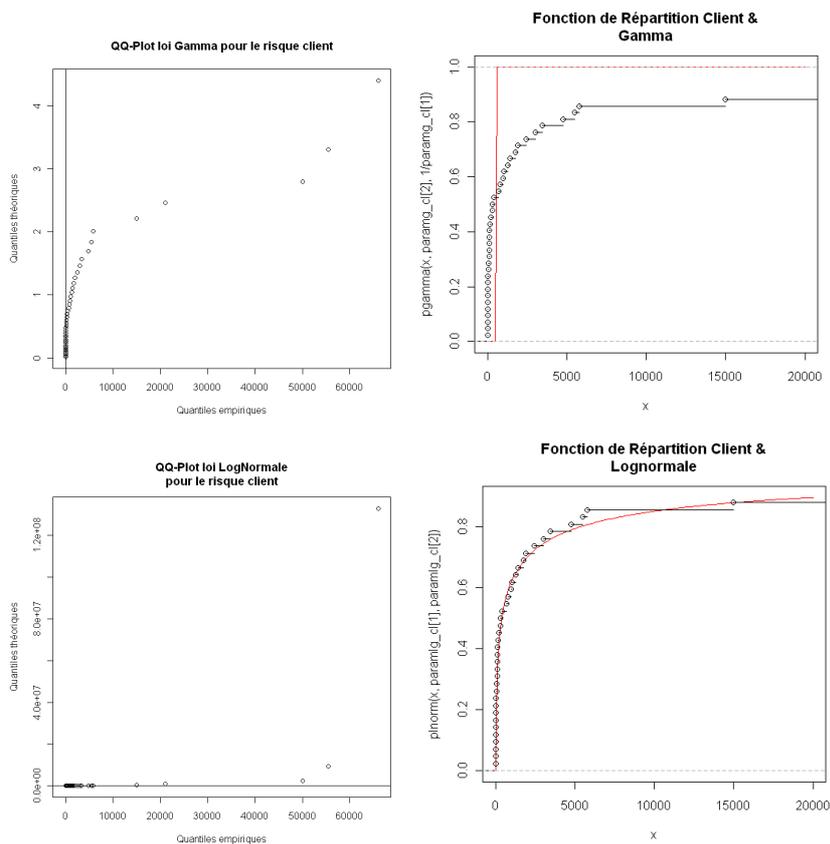
Nous choisissons un niveau de confiance de 95%.

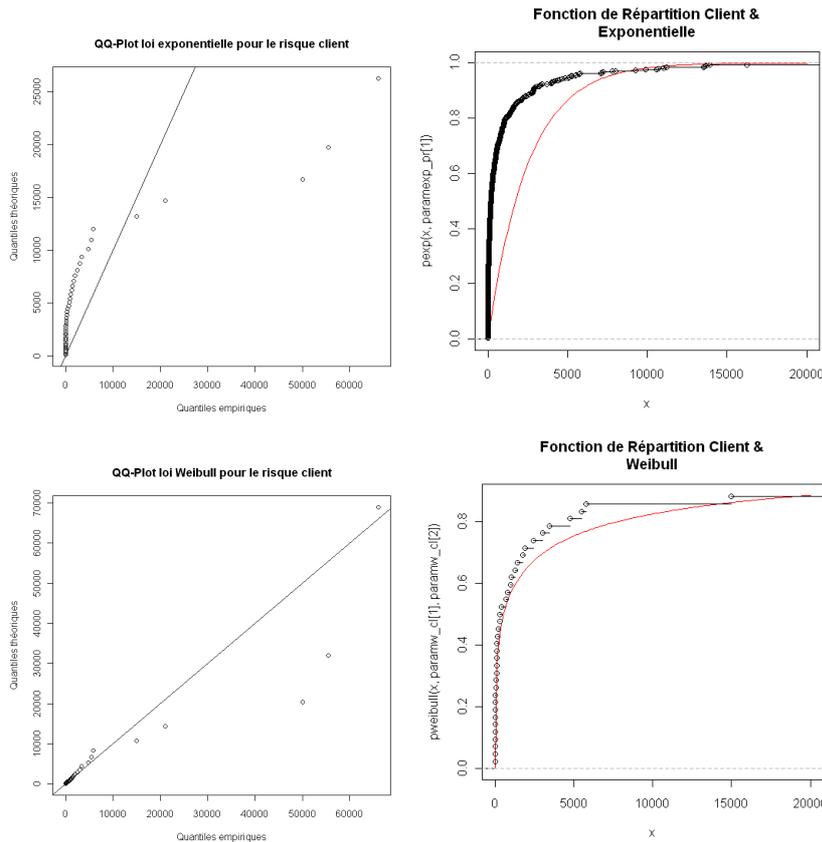
Modélisation du coût individuel des pertes :

Etudions la loi du coût individuel des pertes des 4 risques.

## ➤ Risque Clients :

Estimation de la distribution de sévérité pour le Risque Client :



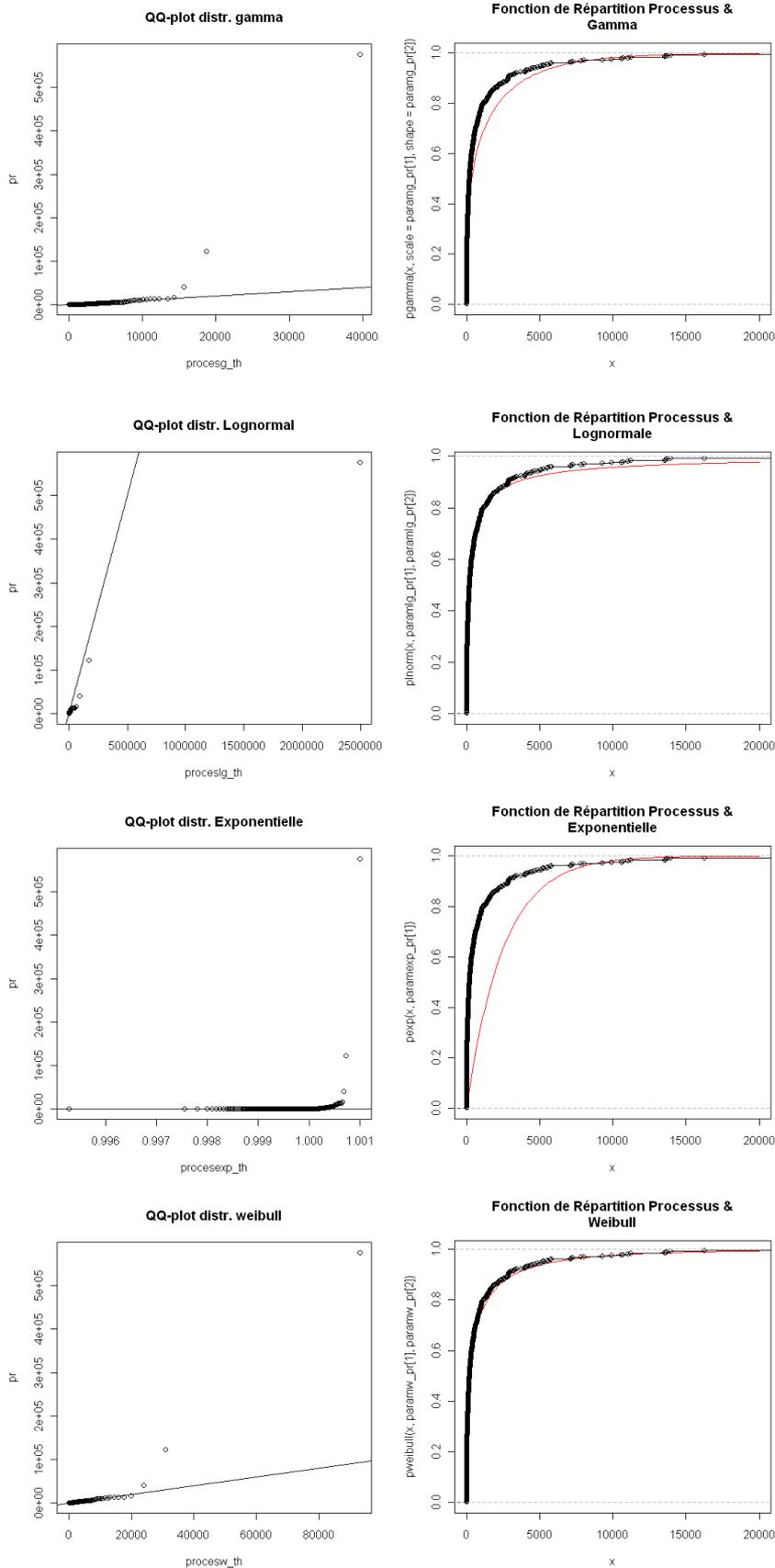


	Gamma	Log-normale	Exponentielle	Weibull
Paramètres	Scale = 801 623 Shape = 0,037	Meanlog = 5,946 Sdlog = 3,126	Rate = 0,000034	Scale = 1 727,41 Shape = 0,315
Test de KS	KS = 0,639 p-value = 2,554e-15	KS = 0,087 p-value = 0,907	KS = 1 p-value < 2,2e-16	KS = 0,116 p-value = 0,623
Test d'Anderson-Darling	AD = Inf p-value = 1429e-05	AD = 0,280 p-value = 0,952	AD = Inf p-value = 1,429e-05	AD = 0,656 p-value = 0,596

Les résultats présentés dans le tableau ci-dessus montrent que les distributions Log-normale et Weibull s'ajustent le mieux à nos données. En particulier, nous retenons la loi Log-normale comme loi de sévérité du risque client du fait de sa p-value largement supérieure au seuil choisi de 5% pour les deux tests effectués.

➤ Risque de Processus :

Estimation des distributions de sévérité pour le Risque Processus :

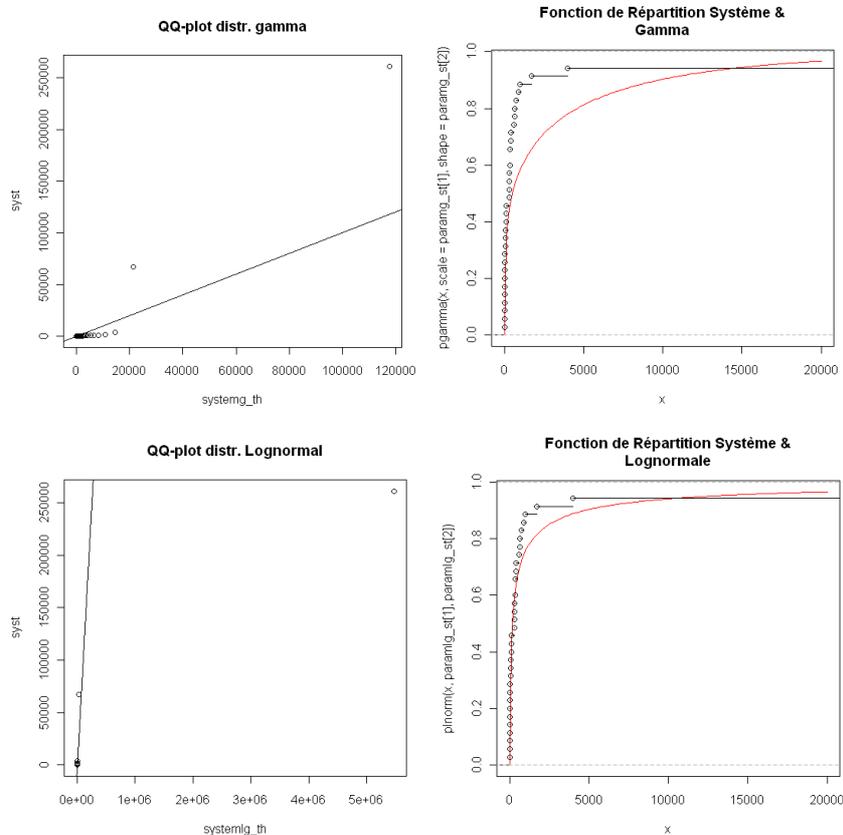


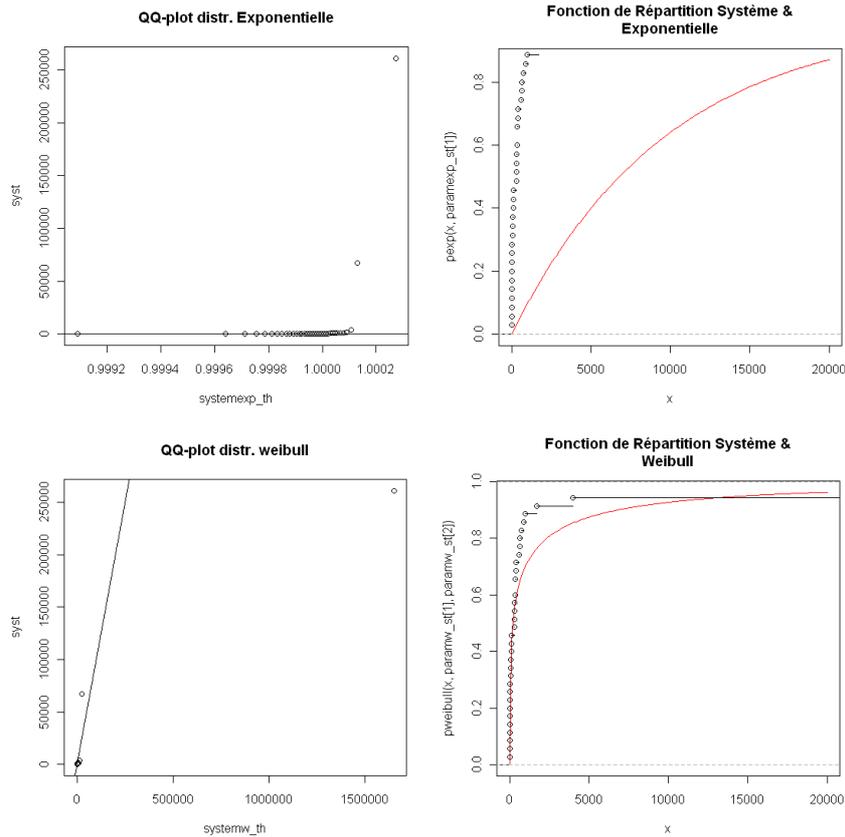
	Gamma	Log-normale	Exponentielle	Weibull
Paramètres	Scale = 4 624,679 Shape = 0,307	Meanlog = 5,019 Sdlog = 2,432	Rate = 0,0004	Shape = 482,505 Scale = 0,435
Test de Kolmogorov-Smirnov	KS = 0,119 p-value = 2,569e-06	KS = 0,059 p-value = 0,068	KS = 0,969 p-value < 2,2e-16	KS = 0,041 p-value = 0,397
Test d'Anderson-Darling	AD = Inf p-value = 1,232e-06	AD = 1,984 p-value = 0,094	AD = Inf p-value = 1,232e-06	AD = 2,037 p-value = 0,088

Les résultats présentés dans le tableau ci-dessus montrent que seules les distributions Log-normale et Weibull « passent » les tests avec un seuil de 5%. En particulier, nous retenons la loi de Weibull comme loi de sévérité du risque processus.

➤ Risque de Système :

Estimation des distributions de sévérité pour le Risque Système :



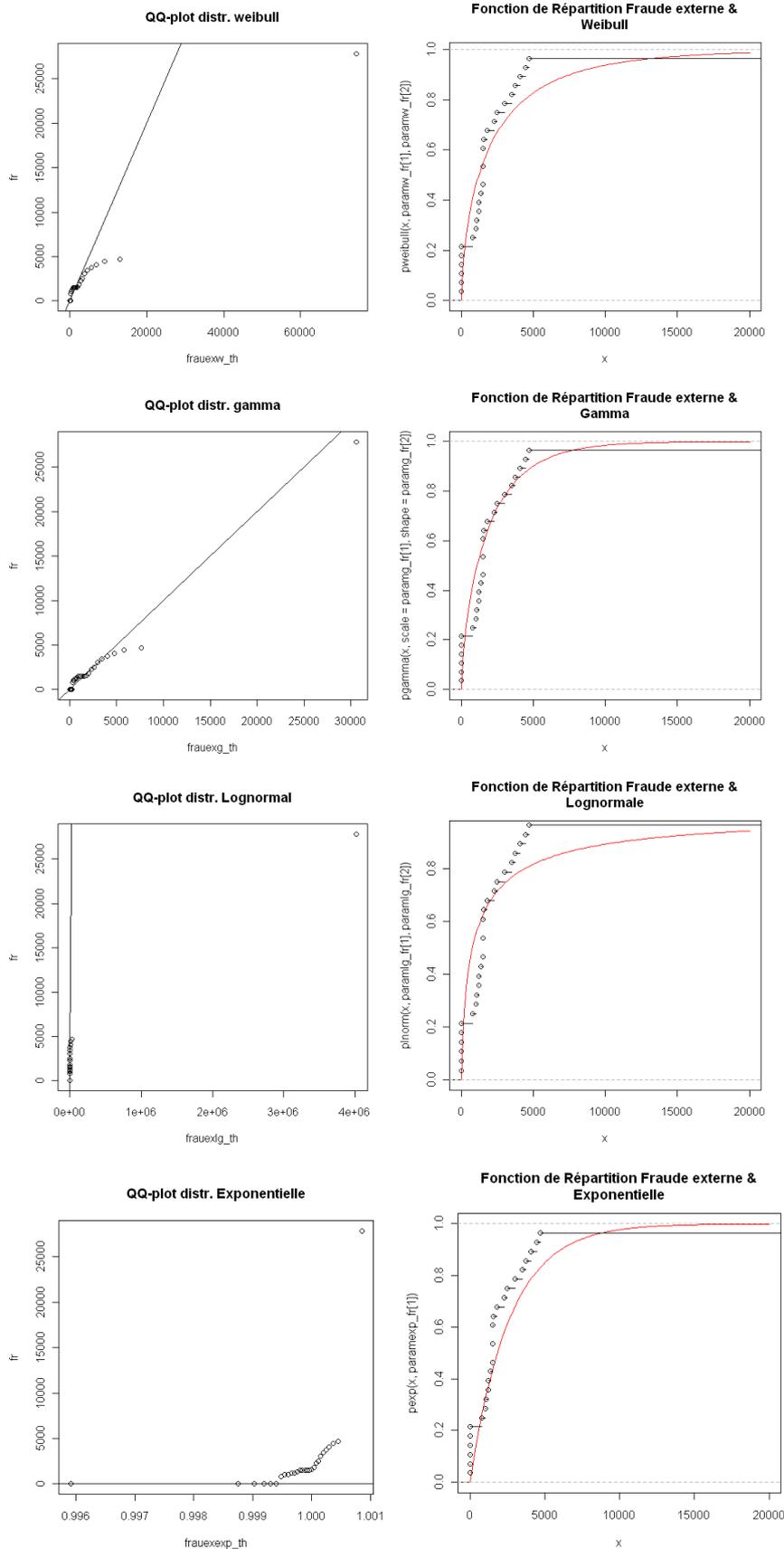


	Gamma	Log-normale	Exponentielle	Weibull
Paramètres	Scale = 13 630,6 Shape = 0,237	Meanlog = 5,055 Sdlog = 2,665	Rate = 0,0001	Scale = 569,214 Shape = 0,336
Test de KS	KS = 0,299 p-value = 0,0038	KS = 0,140 p-value = 0,497	KS = 0,971 p-value < 2,2e-16	KS = 0,183 p-value = 0,190
Test d'Anderson-Darling	AD = Inf p-value = 1,232e-06	AD = 2,603 p-value = 0,044	AD = Inf p-value = 1,232e-06	AD = 13,583 p-value = 1,232e-06

Les résultats présentés dans le tableau ci-dessus montrent que seules les distributions Log-normale et Weibull « passent » le test de KS mais pas d'AD (les p-valeurs sont supérieures au seuil acceptable de 5%). En particulier, nous retenons la loi Log-normale comme loi de sévérité du risque système.

➤ Risque de Fraude externe :

Estimation des distributions de sévérité pour le Risque Fraude externe :



	Gamma	Log-normale	Exponentielle	Weibull
Paramètres	Scale = 2 948,751 Shape = 0,664	Meanlog = 6,661 Sdlog= 2,065	Rate = 0,00038	Scale = 2 130,157 Shape = 0,661
Test de KS	KS = 0,227 p-value = 0,111	KS = 0,297 p-value = 0,014	KS = 1 p-value < 2,2e-16	KS = 0,208 p-value = 0,1757
Test d'Anderson-Darling	AD = Inf p-value = 1,232e-06	AD = 131,4956 p-value = 1,232e-06	AD = Inf p-value = 1,232e-06	AD = Inf p-value = 1,232e-06

Les résultats présentés dans le tableau ci-dessus montrent que seules les distributions Gamma et Weibull « passent » le test de KS avec un seuil de 5% (seul le test de Kolmogorov-Smirnov passe). En particulier, nous retenons la loi Weibull comme loi de sévérité du risque fraude externe.

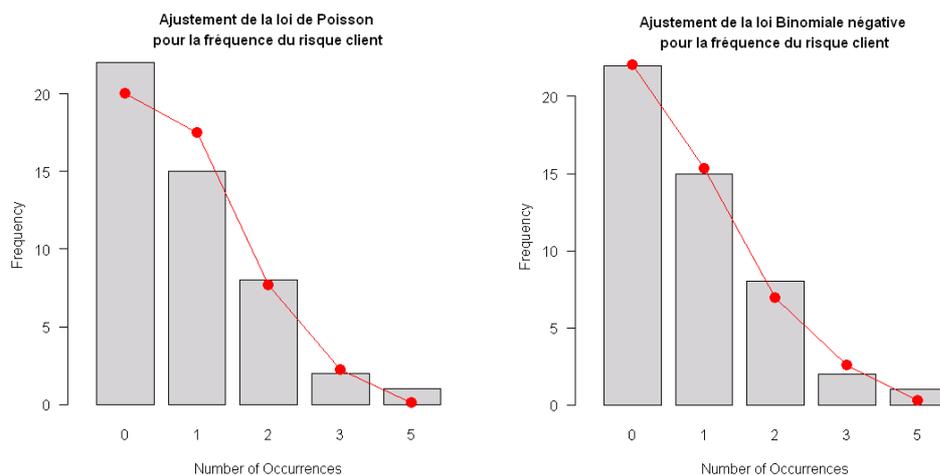
### 7.3 ESTIMATION DES DISTRIBUTIONS DE FRÉQUENCE

Nous proposons d'estimer le nombre de pertes opérationnelles avec les distributions Poisson et Binomiale Négative. De nombreuses études ont déjà modélisé les fréquences des pertes opérationnelles avec la loi de Poisson alors que peu d'entre elles l'ont fait avec le modèle Binomial Négatif.

Nous estimons les paramètres de chacune des deux distributions par la méthode du maximum de vraisemblance via la fonction « goodfit » de R, et ce, pour chacun des risques. Les données utilisées sont les nombres de pertes observées par mois. Nous avons ainsi étudié 48 nombres de pertes par risque.

Il est important de tester le degré d'ajustement des distributions estimées pour chacun des risques. Comme nous sommes en présence de données discrètes, nous mettons en place un test du Chi-deux.

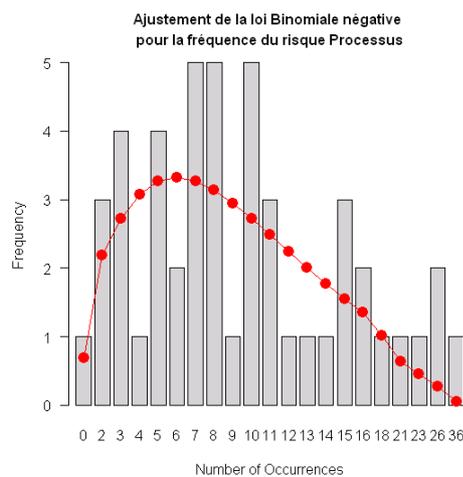
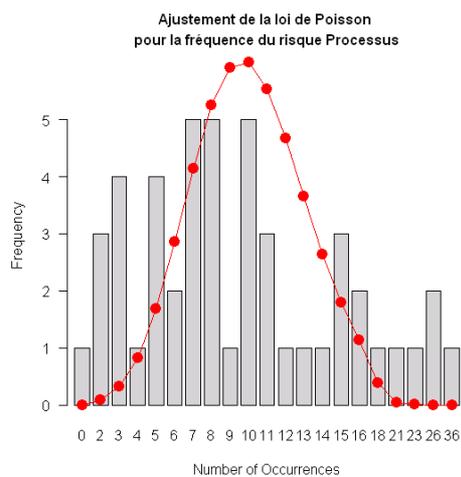
#### ➤ Risque Clients :



	Poisson	Binomiale Négative
Paramètres	Lambda = 0,875	size = 3,419119 prob = 0,7962304
test du Chi-deux	X-squared = 9,5924 p-value = 0,6482	X-squared = 8,301 p-value = 0,7256

Les graphiques montrent que la loi Binomiale Négative semble la plus adéquate. Les résultats présentés dans le tableau ci-dessus le confirment.

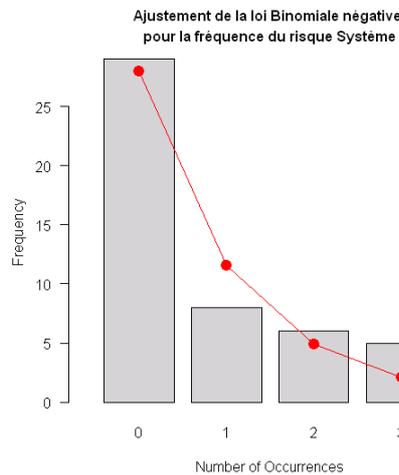
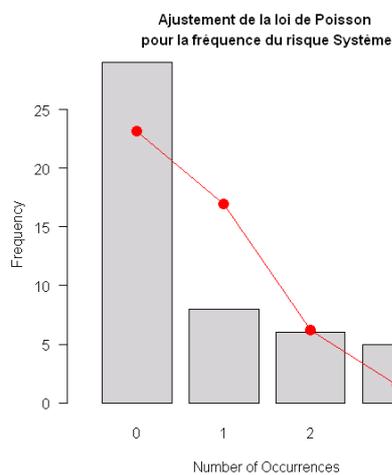
➤ Risque de Processus :



	Poisson	Binomiale Négative
Paramètres	Lambda = 10,14583	Size = 2,752463 prob = 0,2133974
test du Chi-deux	X-squared = 256,38 p-value = 0,8201	X-squared = 325,4533 p-value = 0,91

A partir des graphiques aucune conclusion ne peut être tirée. Les résultats présentés dans le tableau ci-dessus montrent que la Binomiale Négative peut être retenue.

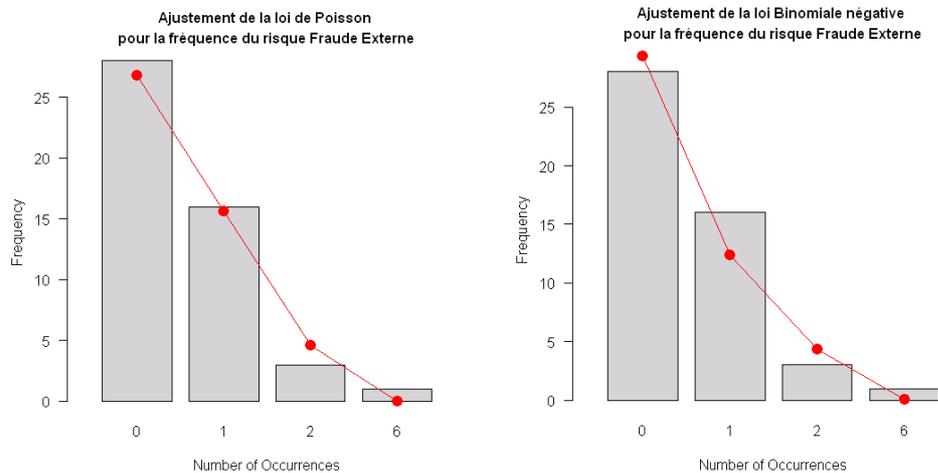
➤ Risque de Système :



	Poisson	Binomiale Négative
Paramètres	Lambda = 0,7291667	Size = 0,9482954 Prob = 0,5653157
test du Chi-deux	X-squared = 14,0443 p-value = 0,1274	X-squared = 19,5246 p-value = 0,2359

Les graphiques montrent que la loi Binomiale Négative semble la plus adéquate. Les résultats présentés dans le tableau ci-dessus le confirment.

➤ Risque de Fraude externe :



	Poisson	Binomiale Négative
Paramètres	Lambda = 0,583	Size = 1,512 prob = 0,722
test du Chi-deux	X-squared = 5,794 p-value = 0,654	X-squared = 15,517 p-value = 0,123

Les graphiques montrent que la loi de Poisson semble la plus adéquate. Les résultats présentés dans le tableau ci-dessus le confirment.

Voici le tableau récapitulatif des lois de coût et de fréquence retenues :

	Risque Clients	Risque de Processus	Risque de Système	Risque de Fraude externe
Loi retenue pour le coût	LN(5,95 ; 3,13)	W(482,5 ; 0,44)	LN(5,056 ; 2,67)	W(2130,16 ; 0,66)
Loi retenue pour la fréquence	BN(3,42 ; 0,8)	BN(2,75 ; 0,21)	BN(0,95 ; 0,57)	P(0,583)

#### 7.4 APPLICATION DE LA MÉTHODE LOSS DISTRIBUTION APPROACH AVEC SEUIL

L'idée est de valider la méthode que nous avons présentée dans le cas où les données sont collectées au-delà d'un seuil. Pour cela, nous avons utilisé nos données bancaires et nous avons appliqué un seuil de 1500€ sur toutes les pertes.

Nous avons utilisé la méthode du maximum de vraisemblance pour estimer nos paramètres. Nous avons maximisé la vraisemblance sous R et sous excel avec le solveur.

Test de Kolmogorov adapté à des données tronquées à gauche :

Dans le cas de données collectées au-delà d'un seuil, le test de Kolmogorov-Smirnov doit être adapté. L'hypothèse devient :

$H_0$  : la loi  $P_X$  a pour fonction de répartition  $F_0^*$ , où  $F_0^* = \frac{F_0(x) - F_0(U)}{1 - F_0(U)}$

Notons à présent  $y_j = F_0(x_j)$  et  $y_U = F_0(U)$

$KS_{obs}^*$  devient :

$$KS_{obs}^* = \frac{\sqrt{n}}{1 - y_U} \max \left\{ \sup_j \left( y_U + \frac{j}{n} (1 - y_U) - y_j \right), \sup_j \left( y_j - \left( y_U + \frac{j-1}{n} (1 - y_U) \right) \right) \right\}$$

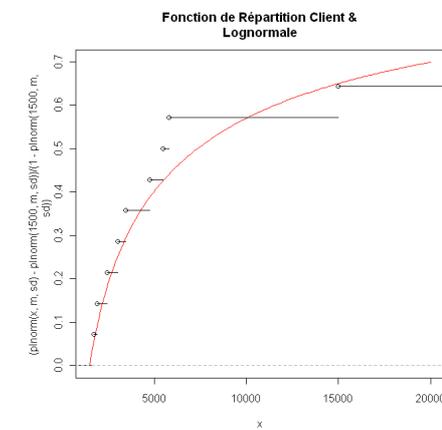
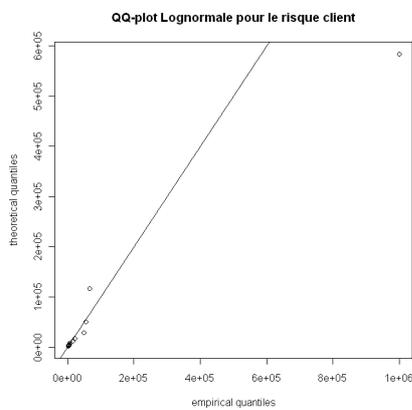
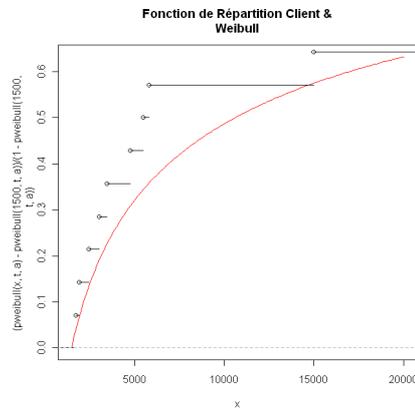
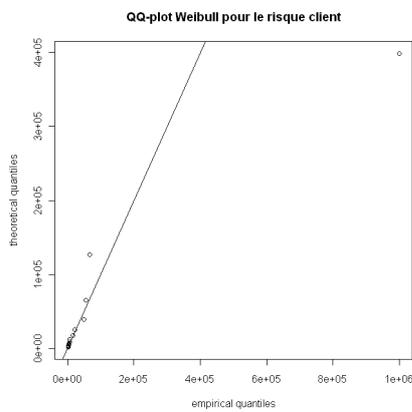
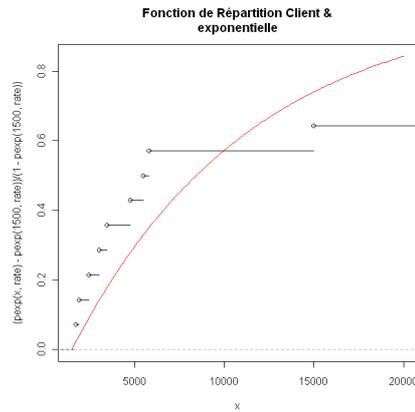
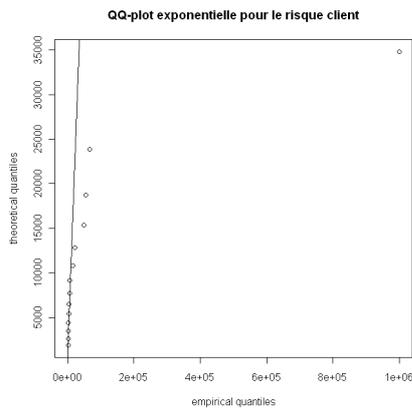
La p-value associée se calcule par simulation de Monte-Carlo de la façon suivante :

- 1) nous créons un grand nombre d'échantillons issus de la distribution ajustée, de la même taille que l'échantillon empirique.
- 2) Pour chacun de ces échantillons, nous calculons la valeur de la statistique  $KS_{obs}^*$ .
- 3) Puis nous estimons la p-value comme la proportion de fois où la valeur de la statistique étudiée est supérieure à la valeur estimée sur l'échantillon empirique. Nous rejetons l'hypothèse si la p-value est inférieure à un seuil donné et nous l'acceptons si la p-value est supérieure.

Nous choisissons un niveau de confiance de 99,5%.

Modélisation du coût individuel :

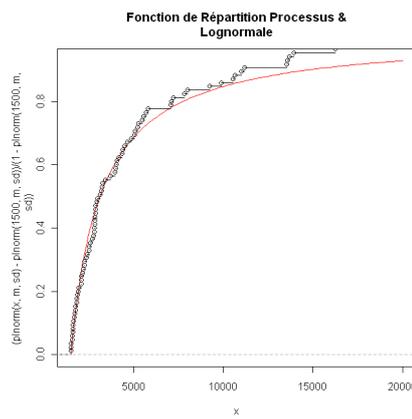
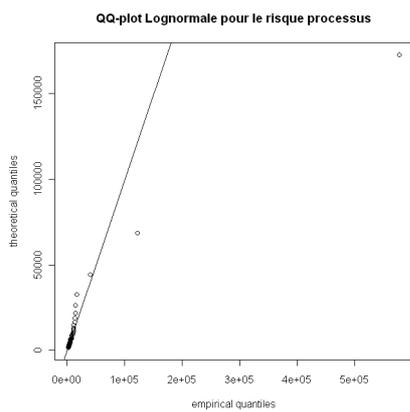
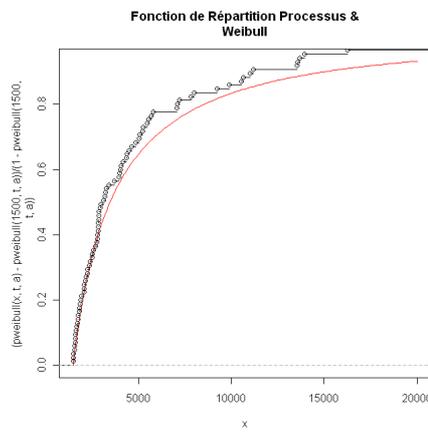
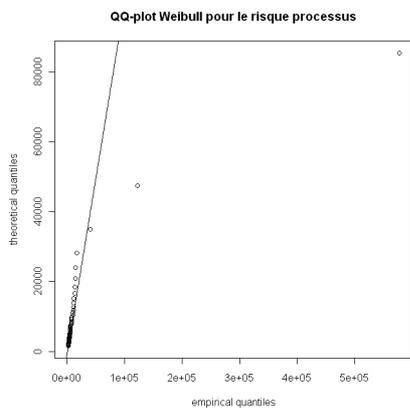
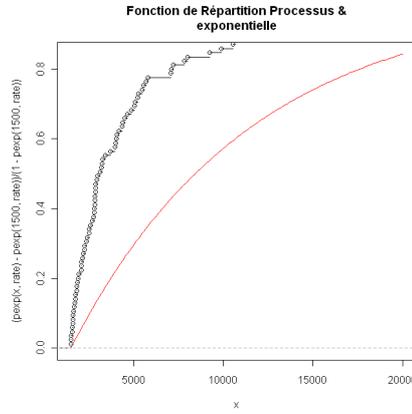
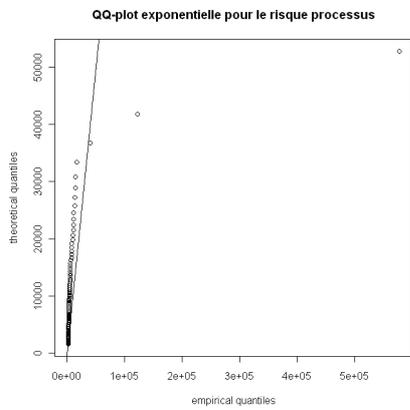
➤ Risque Clients :



	Log-normale	Exponentielle	Weibull
Paramètres	Meanlog = 3,287 Sdlog = 3,924	Rate = 0,0001	Scale = 999,894 Shape = 0,249
Test de KS	KS = 0,489 p-value = 0,74	KS = 1,04 p-value = 0,128	KS = 0,8006 p-value = 0,37

Les résultats ci-dessus indiquent que la loi Log-normale s'ajuste le mieux à nos données.

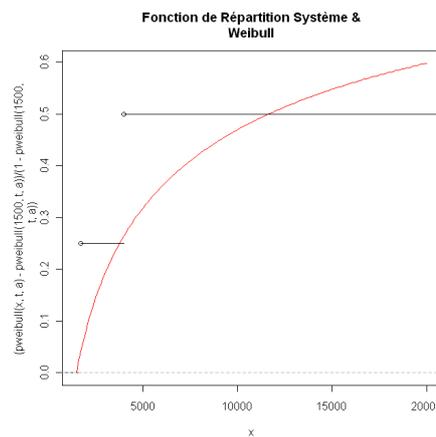
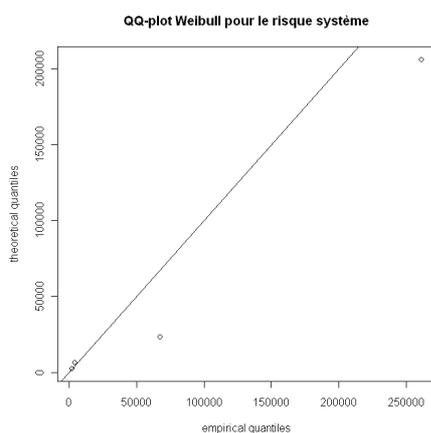
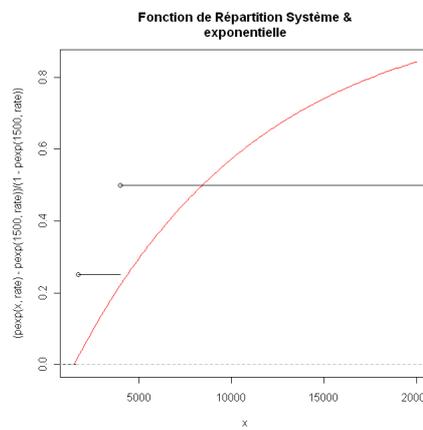
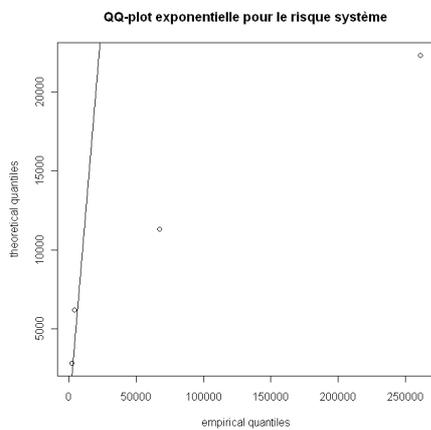
➤ Risque de Processus :

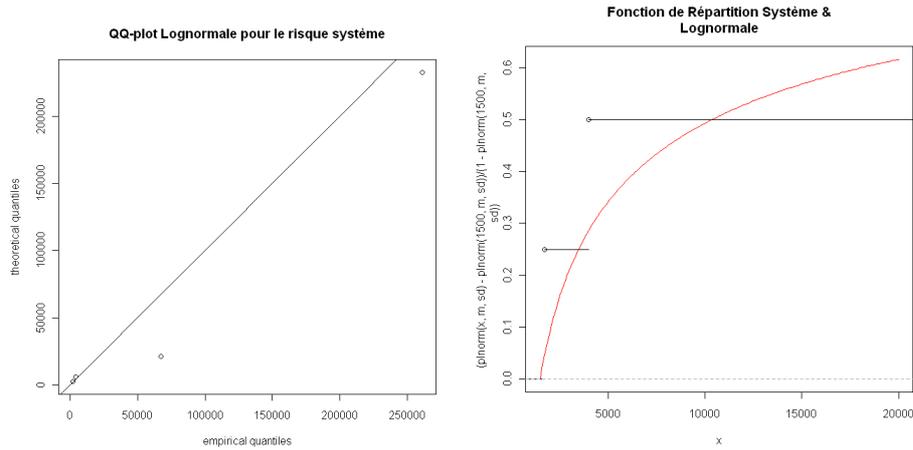


	Log-normale	Exponentielle	Weibull
Paramètres	Meanlog = -2,685 Sdlog = 3,469	Rate = 0,0001	Scale = 17,6 Shape = 0,247
Test de KS	KS = 0,712 p-value = 0,289	KS = 3,939 p-value < 2,2e-16	KS = 0,7504 p-value = 0,33

Les résultats ci-dessus indiquent que la loi Weibull s'ajuste le mieux à nos données.

➤ Risque de Système :

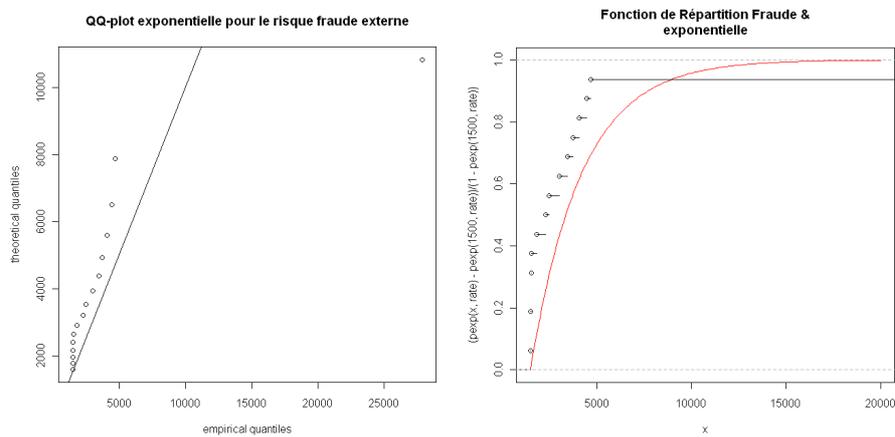


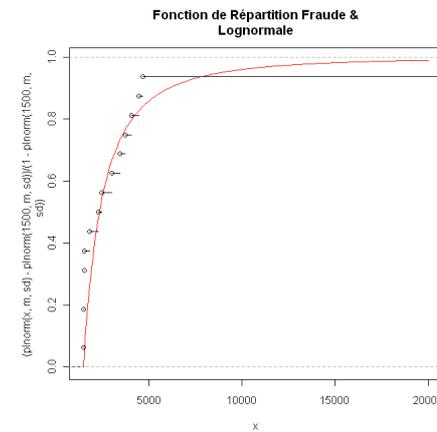
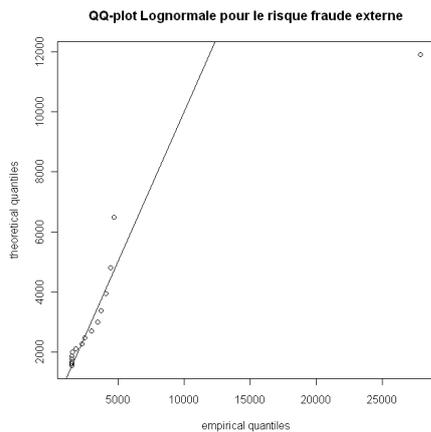
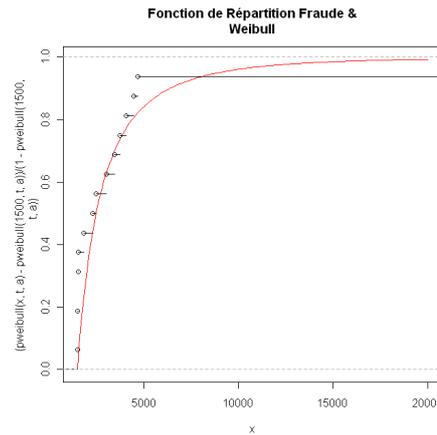
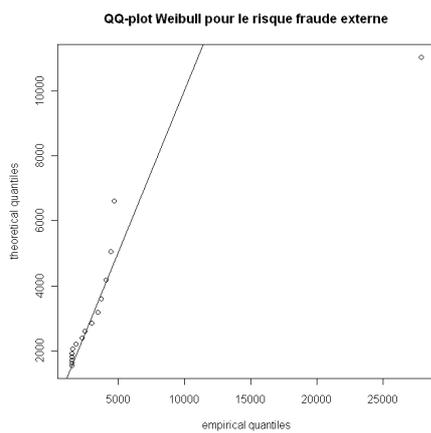


	Log-normale	Exponentielle	Weibull
Paramètres	Meanlog = 1,635 Sdlog = 5,002	Rate = 0,0001	Scale = 10,006 Shape = 0,143
Test de KS	KS = 0,547 p-value = 0,518	KS = 0,9972 p-value = 0,152	KS = 0,534 p-value = 0,77

Les résultats ci-dessus indiquent que la loi Weibull s'ajuste le mieux à nos données.

➤ Risque de Fraude externe:





	Log-normale	Exponentielle	Weibull
Paramètres	Meanlog = 2 Sdlog = 2	Rate = 0,00037	Scale = 12,3 Shape = 0,3
Test de KS	KS = 1,26 p-value = 0,024	KS = 1,407 p-value = 0,01	KS = 1,294 p-value < 2,2e-16

Les résultats ci-dessus ne permettent pas vraiment de conclure, cependant nous choisissons la loi dont la p-value est la plus importante, i.e. la loi Log-normale.

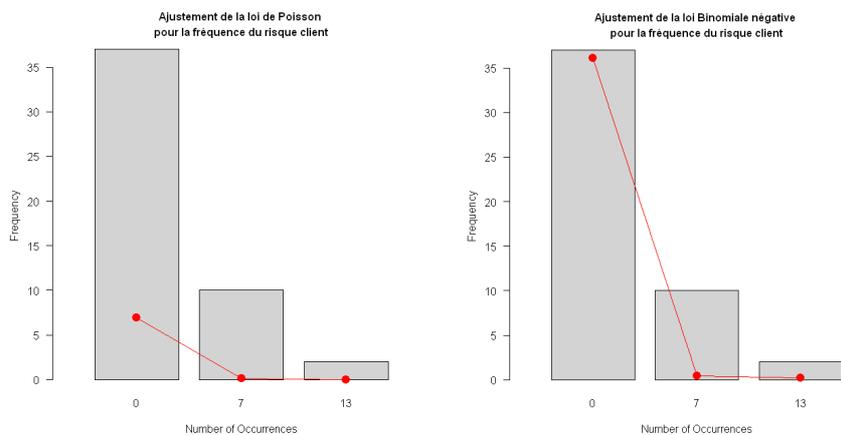
Remarques : l'adéquation à nos données tronquées ne nous donne pas forcément les mêmes résultats que dans notre précédente étude.

L'algorithme de maximisation de la vraisemblance est très sensible aux paramètres initiaux.

Vous remarquerez que nous n'avons pas affiché les tests de la loi Gamma qui ne sont pas concluant pour des problèmes d'optimisation.

## Modélisation de la fréquence :

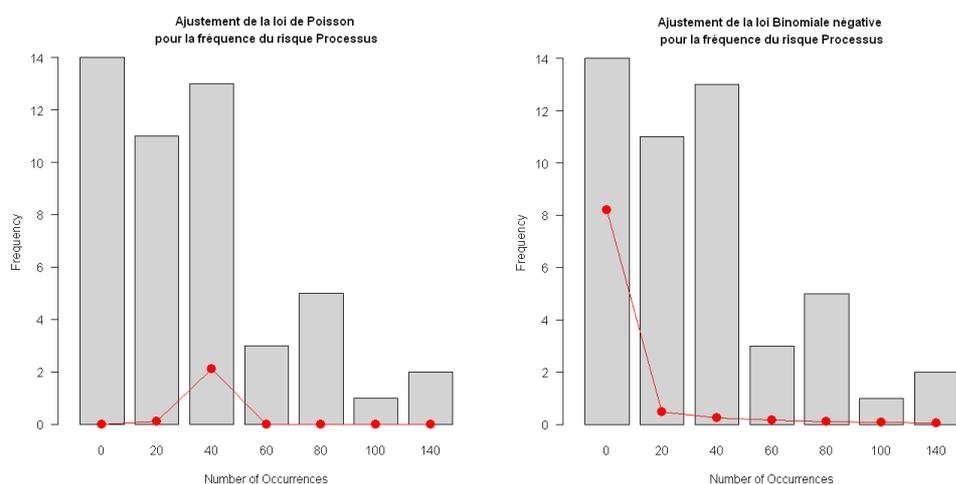
## ➤ Risque Clients :



	Poisson	Binomiale Négative
Paramètres	Lambda = 1,96	size = 0,10 prob = 0,049
test du Chi-deux	X-squared = 6,7 p-value = 0,68	X-squared = 17,48 p-value = 0,46

Nous ne pouvons pas tirer de conclusion à partir des graphiques. Les résultats présentés dans le tableau ci-dessus montrent que la loi de Poisson semble plus adéquate.

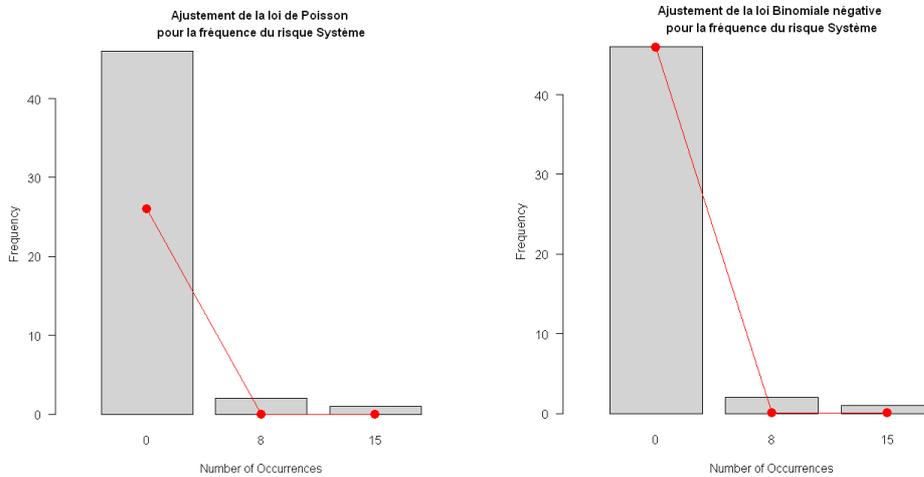
## ➤ Risque de Processus :



	Poisson	Binomiale Négative
Paramètres	Lambda = 34,69	Size = 0,399 prob = 0,011
test du Chi-deux	X-squared = 96,45 p-value = 0,765	X-squared = 103,168 p-value = 0,886

A partir des graphiques aucune conclusion ne peut être tirée. Les résultats présentés dans le tableau ci-dessus montrent que la Binomiale Négative peut être retenue.

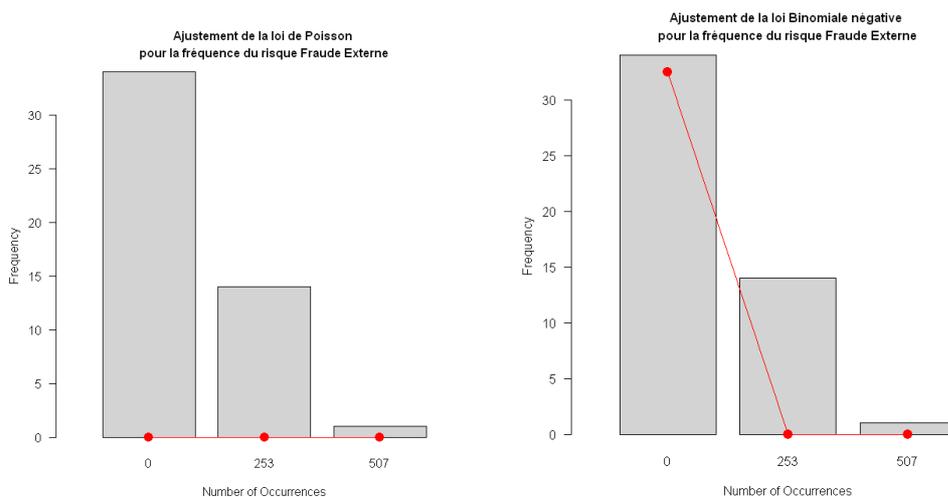
➤ Risque de Système :



	Poisson	Binomiale Négative
Paramètres	Lambda = 0,633	Size = 0,018 Prob = 0,027
test du Chi-deux	X-squared = 6,591 p-value = 0,223	X-squared = 2,813 p-value = 0,72

Les graphiques montrent que la loi Binomiale Négative semble la plus adéquate. Les résultats présentés dans le tableau ci-dessus le confirment.

➤ Risque de Fraude externe :



	Poisson	Binomiale Négative
Paramètres	Lambda = 82,63	Size = 0,056 prob = 0,0007
test du Chi-deux	X-squared = 78,24 p-value = 0,125	X-squared = 5,86 p-value = 0,504

Les graphiques montrent que la loi Binomiale Négative semble la plus adéquate. Les résultats présentés dans le tableau ci-dessus le confirment.

Voici le tableau récapitulatif des lois de sévérité et de fréquence retenues pour la modélisation :

	Risque Clients	Risque de Processus	Risque de Système	Risque de Fraude externe
Loi retenue pour le coût	LN(3,29; 3,92)	W(0,25 ;17,6)	W(0,14 ; 0,006)	LN(2 ; 2)
Loi retenue pour la fréquence	P(1,96)	BN(0,4 ; 0,01)	BN(0,018; 0,027)	BN(0,056 ;0,0007)

## 7.5 SIMULATION DES SCÉNARIOS DE PERTE TOTALE

Nous venons de déterminer les distributions de sévérité et de fréquence des pertes opérationnelles pour chacun des types de risque, sur les données complètes et les données tronquées. Il y a lieu, à ce stade de l'analyse, de combiner les deux distributions pour déterminer la distribution agrégée des pertes annuelles. Comme nous l'avons vu précédemment, il existe plusieurs méthodes d'agrégation et nous choisissons la méthode de simulation Monte Carlo pour déterminer la distribution totale des pertes annuelles de chaque risque. Cette méthode n'est pas la plus rapide mais elle a l'avantage d'être assez précise pour un nombre suffisant de simulations. Nous rappelons que cette méthode consiste à :

1. Générer un nombre  $n$  de pertes par an (on génère 12 nombres de pertes mensuelles indépendamment que l'on somme) ;
2. Générer  $n$  montants de pertes  $X_i$  selon la distribution de sévérité estimée ;
3. Répéter les étapes 1 et 2 un grand nombre de fois (100 000 fois par exemple) pour avoir la distribution empirique des pertes annuelles ;
4. La VaR est déterminée comme étant le quantile d'ordre 99,5 % de la distribution agrégée empirique.

Nous avons fait 100 000 simulations et avons obtenu les résultats suivants :

Résultats sur les pertes complètes :

Risque	VaR 99%	VaR 99,5%	VaR 99,9%
Client	6 654 501	12 171 741	50 357 163
Processus	292 762	317 708	359 160
Système	563 295	951 215	2 497 366
Fraude externe	65 931	75 635	96 297

Nous constatons que le risque Clients est bien un risque grave, l'exigence en fonds propres au titre de ce risque est très élevé par rapport aux autres risques.

Résultats sur les pertes observées au-delà du seuil de 1500€ :

Risque	VaR 99%	VaR 99,5%	VaR 99,9%
Client	10 656 793	19 731 221	58 412 951
Processus	580 510	672 104	857 810
Système	5 781 145	12 147 207	40 941 962
Fraude externe	96 847	111 881	190 095

La méthode a tendance à surestimer le capital. En particulier, le risque de Système fournit des résultats complètement incohérents dans la mesure où le nombre de données utilisées est très faible. Ceci est dû à une surestimation de la fréquence. Il faudrait améliorer la méthodologie permettant de modéliser la fréquence des pertes tronquées.

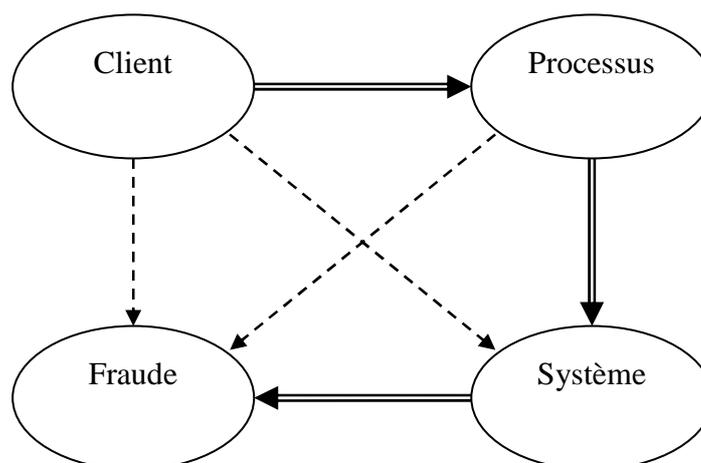
## 7.6 AGRÉGATION DES SCÉNARIOS

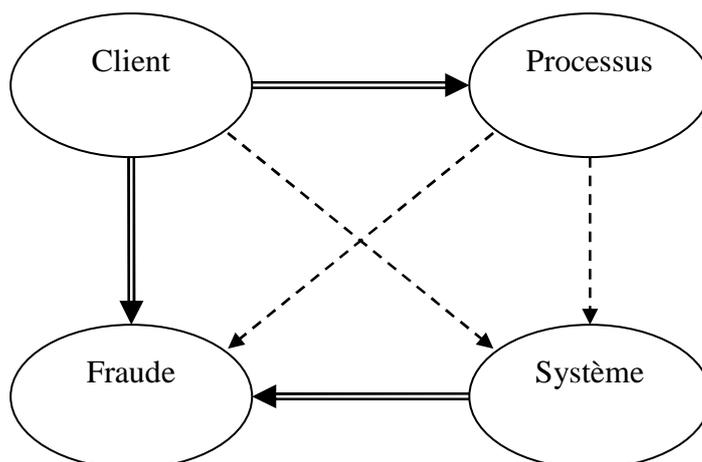
La dépendance entre les risques peut être appréhendée par différentes mesures introduites précédemment. Nous allons en particulier, évaluer pour chaque couple de risque, le coefficient de Pearson, le taux de Kendall et le rho de Spearman. Pour effectuer ces calculs, nous avons regroupé les pertes par mois de survenance pour avoir le même nombre d'observations.

	Pearson	Kendall	Spearman
Client - Processus	0,592	0,138	0,186
Client - Système	0,496	0,204	0,263
Client - Fraude	0,619	-0,019	-0,028
Processus - Système	0,587	0,078	0,108
Processus - Fraude	0,732	-0,028	-0,050
Système - Fraude	0,621	-0,021	-0,020

Il y a plusieurs possibilités pour agréger les risques :

- Il est possible de les agréger deux à deux mais il faut alors choisir 3 relations à modéliser pour éviter les conflits de dépendances entre variables. Les graphes ci-dessous représentent deux exemples de liens de dépendance entre les variables : les doubles flèches suffisent à définir la dépendance de l'ensemble des risques, mais nous aurions très bien pu choisir d'autres sélections pour le définir. Les liens en pointillés ne rendraient peut-être pas la même dépendance que les doubles flèches. L'inconvénient de cette méthode est donc qu'il faut choisir les dépendances à modéliser parmi toutes les possibilités.





Nous choisissons d'estimer des copules bi-variées pour Client-Processus, Processus-Système et Système-Fraude.

Nous estimons les paramètres des copules en utilisant la méthode des moments avec la mesure de Kendall. Nous avons utilisé la fonction « calibKendallsTau » sous le logiciel R, elle est surtout utile pour trouver le paramètre de la copule de Frank car nous n'avons pas de formule analytique du paramètre.

Le tableau ci-dessous montre les estimations obtenues :

	paramètre copule Gumbel	Log-vrais Gumbel	paramètre copule Frank	Log-vrais Frank	paramètre copule Clayton	Log-vrais Clayton	paramètre copule Normale	Log-vrais Normale
Client - Processus	1,088	0,231	1,096	0,636	0,431	1,104	0,179	0,439
Processus - Système	1,008	0,002	0,481	0,113	0,371	0,791	0,099	0,110
Système - Fraude	0,857	0,528	-0,777	0,210			-0,252	0,404

- Nous pouvons également agréger les risques « d'un seul coup » par une copule normale multi-variée pour Client-Processus-Système-Fraude.

	Coefficients Copule Normale
Client -Processus	0,215
Client - Système	0,315
Client - Fraude	-0,030
Processus - Système	0,122
Processus - Fraude	-0,044
Système - Fraude	-0,033

Nous avons également testé la dépendance totale et l'indépendance entre les risques.

Voici le tableau des VaR à 99,5% obtenues par les différentes méthodes :

Risques indépendants	Risques Totalemt Corrélés	Client - Processus: Clayton	Processus - Système: Clayton	Système - Fraude: Normale	Client-Processus-Système-Fraude: Normale
12 430 975	13 516 299	12 566 782			12 341 511

Nous constatons que le résultat pour les risques indépendants est très proche de celui des risques agrégés par les copules. Globalement les résultats sont assez proches dans notre exemple car le risque client est prédominant. Mais en réalité ce ne sera pas toujours le cas. Le problème d'agrégation doit s'appréhender en sachant qu'il ne concerne que les risques de fréquence. En effet, dans notre modèle les risques de gravité seront modélisés par les réseaux bayésiens : la dépendance entre ces risques est prise en compte dans les réseaux par construction. Notre problème d'agrégation s'avère donc relatif et l'impact du choix de la méthode d'agrégation sera limité.

## 8 APPLICATION DE LA MODÉLISATION DES RISQUES DE GRAVITÉ

Ce chapitre concerne une application de la modélisation des risques de gravité. Nous allons appliquer le modèle simplifié et le modèle bayésien puis comparer les résultats.

Le risque que nous avons étudié est le risque d'erreur interne sur le sens, la quantité à la saisie d'un ordre d'investissement/désinvestissement sur les produits actions dans un contexte de décalage des marchés. C'est un risque de gravité qui peut entraîner des pertes significatives.

Nous noterons ici ce risque, le risque n°24.

Dans un premier temps, nous avons appliqué le modèle utilisé à court terme avec la méthode des quantiles en testant plusieurs lois.

Puis, nous avons appliqué le modèle théorique bayésien en testant les probabilités.

Enfin, nous avons comparé les résultats pour valider notre approche à court terme.

### 8.1 SELF RISK ASSESSMENT DU RISQUE

Voici le résultat du SRA sous forme de fiche technique :

<b>Risk</b>		<b>Risque d'exécution</b>																	
<b>Risk event 1</b>		<b>Risque d'erreur interne sur le sens, la quantité ou le code</b>																	
<b>Risk event de</b>		<b>Problématique d'exactitude lors de l'exécution des ordres sur les produits actions</b>																	
<b>Périmètre:</b>		<b>placements pour le compte des sociétés d'assurance du Groupe</b>																	
Hypotheses		Basse			Moyenne			Haute											
Description: Risque d'erreur interne sur la quantité d'un ordre d'investissement / désinvestissement exécuté à la saisie dans un contexte de décalage des marchés (actions)		1% de décalage par rapport au marché Moins de 1% de la VL moyenne du portefeuille Durée de détection de l'erreur (entraînant risque d'erreur de VL et de fait un risque d'indemnisation) : quelques heures Montant de l'ordre moyen fourchette basse en portefeuille (5M€)			2% de décalage par rapport au marché Entre 1% et 2% de la VL moyenne du portefeuille Durée de détection de l'erreur (entraînant risque d'erreur de VL et de fait un risque d'indemnisation) : un jour Montant de l'ordre moyen fourchette moyenne en portefeuille (15M€)			5% de décalage par rapport au marché Au-delà de 2% de la VL moyenne du portefeuille Durée de détection de l'erreur (entraînant risque d'erreur de VL et de fait un risque d'indemnisation) : plusieurs mois Impact sur les marchés de l'erreur Montant de l'ordre moyen fourchette haute en portefeuille (50M€)											
Causes																			
Niveau 1	Niveau 2	Commentaires / Description			Sélection			Sélection			Sélection								
Interne	Acteur	Erreur de saisie d'un acteur interne Générali			✓			✓			✓								
	Processus																		
	SI	SI inadapté (absence d'alertes ou de blocages au-delà d'un certain seuil, si montant de la vente > 5% de l'actif, si vente d'un titre non présent en portefeuille, etc.)						✓			✓								
Externe	Environnement (naturel / technologique / culturel / Réglementaire)	Environnement technologique de migration Sous-effectif conjoncturel			✓						✓								
	Tiers / Fournisseurs / Prestataires																		
	Autres																		
Dispositif de contrôle																			
Description		Sélection			Sélection			Sélection											
Formation des collaborateurs		✓																	
Contrôle de la délégation (SI)								✓											
Contrôle de la bonne application de la procédure de passation des ordres pour chaque type d'instrument financier (dont écoute des bandes enregistrées)					✓			✓											
Mise en place d'alertes ou blocages dans le SI								✓											
Séparation des tâches (contrôles effectués par le Middle Office et divers rapprochements)					✓			✓											
Contrôle du calcul de la VL (analyse des variations atypiques)					✓			✓											
Conséquences																			
Niveau 1	Niveau 2	Sélection			Description			Formule de calcul			Montant								
Conséquence financière directe	Pertes de Chiffre d'affaires (sous-performance, diminution des actifs sous gestion)																		
	Autres pertes directes	✓			Indemnisation du fonds Ordre moyen actions dans fourchette basse en portefeuille 5M€ x delta de cours) + 0,5% de l'ordre moyen (coûts de transaction)  Delta de cours = spread moyen 1% + volatilité moyenne sur le période de non détection de l'erreur soit quelques heures 2% = 3%			175 000 €			Indemnisation du fonds Ordre moyen actions dans fourchette moyenne en portefeuille 15M€ x delta de cours) + 0,5% de l'ordre moyen (coûts de transaction)  Delta de cours = spread moyen 3% + volatilité moyenne sur le période de non détection de l'erreur soit une journée 2% = 5%			825 000 €					
Conséquence qualitative	Satisfaction client (Fort / Moyen / Faible)																		
	Réglementaire	✓									Sanction Réglementaire Montant forfaitaire des pénalités réglementaires			250 000 €					
	Climat social																		
	Image et réputation (Fort / Moyen / Faible)	✓									Impact réputation fort								
		Total			175 000 €			Total			825 000 €			Total			15 500 000 €		
Atténuation / Mitigation / Récupération																			
Description		Sélection			Quantification			Montant			Sélection			Quantification			Montant		
Pas d'actions d'atténuation / mitigation / récupération								€						€			€		
								€						€			€		
								0,00 €						0,00 €			0,00 €		
Fréquence																			
Fréquence d'occurrence		12 fois tous les ans			6 fois tous les ans			1 fois tous les 5 ans											
Quantification (1)		175 000,00 €			825 000,00 €			15 500 000,00 €											
Atténuation / Mitigation / Récupération (2)		0,00 €			0,00 €			0,00 €											
Quantification finale (=1-2)		175 000,00 €			825 000,00 €			15 500 000,00 €											

## 8.2 MODÈLE SIMPLIFIÉ

Nous appliquons ici le modèle simplifié présenté plus haut pour modéliser le risque n°24.

Reprenons les résultats obtenus lors du SRA :

	Risque 24		
	Scénario Hypothèse Basse (B)	Scénario Hypothèse Moyenne (M)	Scénario Hypothèse Haute (H)
Fréquence	$\lambda_{24,B}=12$	$\lambda_{24,M}=6$	$\lambda_{24,H}=1/5$
Coût (M€)	$c_{24,B}=0,175$	$c_{24,M}=0,825$	$c_{24,H}=15,5$

1<sup>ère</sup> approche : Approche empirique

La loi de la charge totale s'obtient par la formule :

$$P(S_{24,i} = n \times c_{24,i}) = \frac{\lambda_{24,i}^n}{n!} \times \exp(-\lambda_{24,i}) \quad \text{pour } n \in \mathbb{N} \text{ et } i \in \{B, M, H\}$$

Où B représente l'hypothèse basse, M l'hypothèse moyenne et H l'hypothèse haute.

Ce qui donne pour les trois scénarios :

$$\text{Pour } i=B, \text{ on a } P(S_{24,B} = 0,175n) = \frac{12^n}{n!} \times \exp(-12) \quad \text{pour } n \in \mathbb{N}$$

$$\text{Pour } i=M, \text{ on a } P(S_{24,M} = 0,825n) = \frac{6^n}{n!} \times \exp(-6) \quad \text{pour } n \in \mathbb{N}$$

$$\text{Pour } i=H, \text{ on a } P(S_{24,H} = 15,5n) = \frac{0,2^n}{n!} \times \exp(-0,2) \quad \text{pour } n \in \mathbb{N}$$

Nous avons testé deux hypothèses d'agrégation : soit les scénarios sont totalement corrélés, soit ils sont indépendants.

Hypothèse 1 : les trois scénarios sont totalement corrélés

Pour chaque scénario, nous calculons numériquement le quantile empirique à 99,5% à partir de la fonction de répartition :

n	n*0,175	n*0,825	n*15,5	F(S(24,B)<=n)	F(S(24,M)<=n)	F(S(24,H)<=n)
0	0	0	0	6,14421E-06	0,002478752	0,818730753
1	0,175	0,825	15,5	7,98748E-05	0,017351265	0,982476904
2	0,35	1,65	31	0,000522258	0,061968804	0,998851519
3	0,525	2,475	46,5	0,002291791	0,151203883	0,99994316
4	0,7	3,3	62	0,007600391	0,2850565	0,999997742
5	0,875	4,125	77,5	0,020341029	0,445679641	0,999999925
6	1,05	4,95	93	0,045822307	0,606302782	0,999999998
7	1,225	5,775	108,5	0,089504497	0,74397976	1
8	1,4	6,6	124	0,155027782	0,847237494	1
9	1,575	7,425	139,5	0,242392162	0,916075983	1
10	1,75	8,25	155	0,347229418	0,957379076	1
11	1,925	9,075	170,5	0,461597333	0,979908036	1
12	2,1	9,9	186	0,575965249	0,991172516	1
13	2,275	10,725	201,5	0,681535632	0,996371507	1
14	2,45	11,55	217	0,772024532	0,998599646	1
15	2,625	12,375	232,5	0,844415652	0,999490902	1
16	2,8	13,2	248	0,898708993	0,999825123	1
17	2,975	14,025	263,5	0,937033703	0,999943083	1
18	3,15	14,85	279	0,96258351	0,999982403	1
19	3,325	15,675	294,5	0,978720231	0,99999482	1
20	3,5	16,5	310	0,988402263	0,999998545	1
21	3,675	17,325	325,5	0,993934853	0,999999609	1
22	3,85	18,15	341	0,996952629	0,999999899	1
23	4,025	18,975	356,5	0,998527121	0,999999975	1
24	4,2	19,8	372	0,999314367	0,999999994	1
25	4,375	20,625	387,5	0,999692245	0,999999999	1
26	4,55	21,45	403	0,99986665	1	1
27	4,725	22,275	418,5	0,999944164	1	1

Nous lisons :  $Q_{24,B;99,5\%} = 3,85M \text{ €}$ ,  $Q_{24,M;99,5\%} = 10,725M \text{ €}$  et  $Q_{24,H;99,5\%} = 31M \text{ €}$ .

D'où  $Q_{24;99,5\%} = 3,85 + 10,725 + 31 = 45,575 M \text{ €}$ .

Remarque : nous aurions pu améliorer l'estimation du quantile par une interpolation linéaire.

Hypothèse 2 : les trois scénarios sont indépendants

Pour chaque scénario, nous simulons la perte totale un grand nombre de fois (ici 1 000 000 fois).

Puis nous additionnons les différents scénarios sans ordonner les valeurs (ce qui reviendrait à une corrélation totale) pour obtenir les valeurs de la charge totale du risque 24.

Ce calcul a été réalisé sous le logiciel R (avec les fonctions « rpois » et « quantile ») et conduit à  $Q_{24,99,5\%} = 39,575 M \text{ €}$ .

2<sup>de</sup> approche : Nous modélisons la fréquence par une loi de Poisson et nous testons plusieurs lois de sévérité.

L'estimateur du paramètre de fréquence est égal à:  $\lambda_{24} = \sum_{i \in \{B,M,H\}} \lambda_{24,i} = 18,2$ .

Dans un premier temps, nous choisissons quatre lois a priori pour modéliser la sévérité d'une perte : Exponentielle, Log-normale, Weibull et Pareto et nous estimons les paramètres par la méthode des quantiles en supposant que :

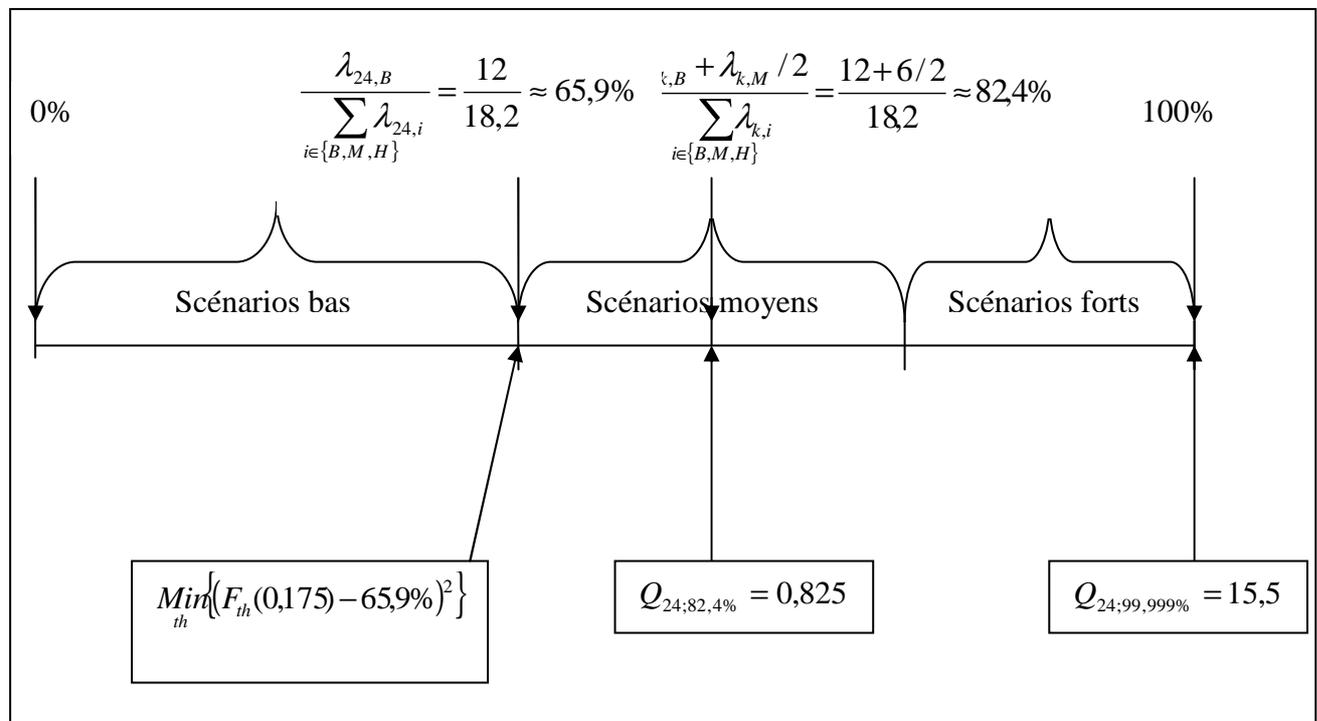
- la perte issue du scénario moyen donne le quantile à

$$\frac{\lambda_{24,B} + \lambda_{24,M} / 2}{\sum_{i \in \{B,M,H\}} \lambda_{24,i}} = \frac{15}{18,2} \approx 82,4\% : Q_{24;82,4\%} = 0,825$$

- La perte issue du scénario fort donne le quantile à 99,999% :  $Q_{24;99,999\%} = 15,50$

Puis, nous utilisons un critère basé sur le scénario hypothèse basse :

$$\text{Min}_{th} \left\{ \left( F_{th}(c_{24,B}) - F_{emp}(c_{24,B}) \right)^2 \right\} \text{ où } F_{emp}(c_{24,B}) = \frac{\lambda_{24,B}}{\sum_{i \in \{B,M,H\}} \lambda_{24,i}}$$



Voici les résultats de la sélection de la loi de coût :

Risque	Ecart sur le scénario bas exponentielle	Ecart sur le scénario bas Weibull	Ecart sur le scénario bas Lognormale	Ecart sur le scénario bas Pareto	Ecart min	Loi retenue
24	0,29	0,04	0,21	0,10	0,04	Weibull

Le critère nous permet de retenir la loi Weibull comme loi de coût.

Pour modéliser la charge totale, nous avons appliqué les quatre méthodes présentées plus haut : Panjer, inversion de la fonction caractéristique, Normal-Power et Monte-Carlo. Nous avons regardé les résultats pour toutes les lois pour étudier les écarts liés au choix de la loi. Les résultats sont donnés dans le tableau ci-dessous :

	Estimation des paramètres	Panjer	Inversion fonction caractéristique	Normal-Power	Monte-Carlo
Loi Exponentielle	$\hat{\lambda}_{24} = -\frac{1}{Q_{24,g}} \times \ln(1-g)$ $= -\frac{\ln(1-99,999\%)}{15,5} \approx 0,74$	25M€	27M€	27,15M€	26,87M€
Loi Weibull	$\hat{a} = \frac{\ln\left(\frac{\ln(1-p_1)}{\ln(1-p_2)}\right)}{\ln\left(\frac{Q_{p_1}}{Q_{p_2}}\right)} \approx 0,65$ $\hat{b} = \frac{Q_{p_1}}{(-\ln(1-p_1))^{\frac{1}{\hat{a}}}} \approx 0,35$	20M€	22M€	22,04M€	21,94M€
Loi Log-normale	$\hat{\sigma} = \frac{\ln(Q_{g_1}) - \ln(Q_{g_2})}{\Phi^{-1}(g_1) - \Phi^{-1}(g_2)}$ $= \frac{\ln(0,825) - \ln(15,5)}{\Phi^{-1}(0,82) - \Phi^{-1}(0,999)} \approx 0,88$ $\hat{\mu} = \ln(Q_{g_1}) - \Phi^{-1}(g_1)$ $= \ln(0,825) - \Phi^{-1}(0,82) \approx -1$	18M€	20M€	20,77M€	20,97M€

Loi de Pareto	$\frac{Q_{g_1}}{Q_{g_2}} = \frac{(1-g_1)^{-\frac{1}{\hat{\alpha}}}-1}{(1-g_2)^{-\frac{1}{\hat{\alpha}}}-1}$ $\hat{\alpha} \approx 5,51$ $\hat{\lambda} = \frac{Q_{g_1}}{(1-g_1)^{-\frac{1}{\hat{\alpha}}}-1} \approx 2,22$	17M€	19M€	18,75M€	20,18M€
---------------	--	------	------	---------	---------

Nous pouvons remarquer que les résultats ont des ordres de grandeur assez comparables.

La loi Weibull retenue nous donne un résultat de 21,94M€ en utilisant la méthode de Monte-Carlo.

Le point sensible de la méthode est le choix du quantile extrême à 99,999%. C'est un choix totalement arbitraire, aussi avons-nous testé la sensibilité à ce paramètre.

Test de sensibilité au quantile extrême : nous avons regardé comment évoluer le résultat en fonction du nombre de neufs ajoutés après la virgule.

	99,9999%	99,99%	99,9%
Loi retenue	Weibull(0,71; 0,38)	Weibull(0,57 ; 031)	Weibull(0,47 ; 0,25)
VaR	20,41	25,93	36,18
Ecart par rapport à 99,999%	-7%	+18%	+65%

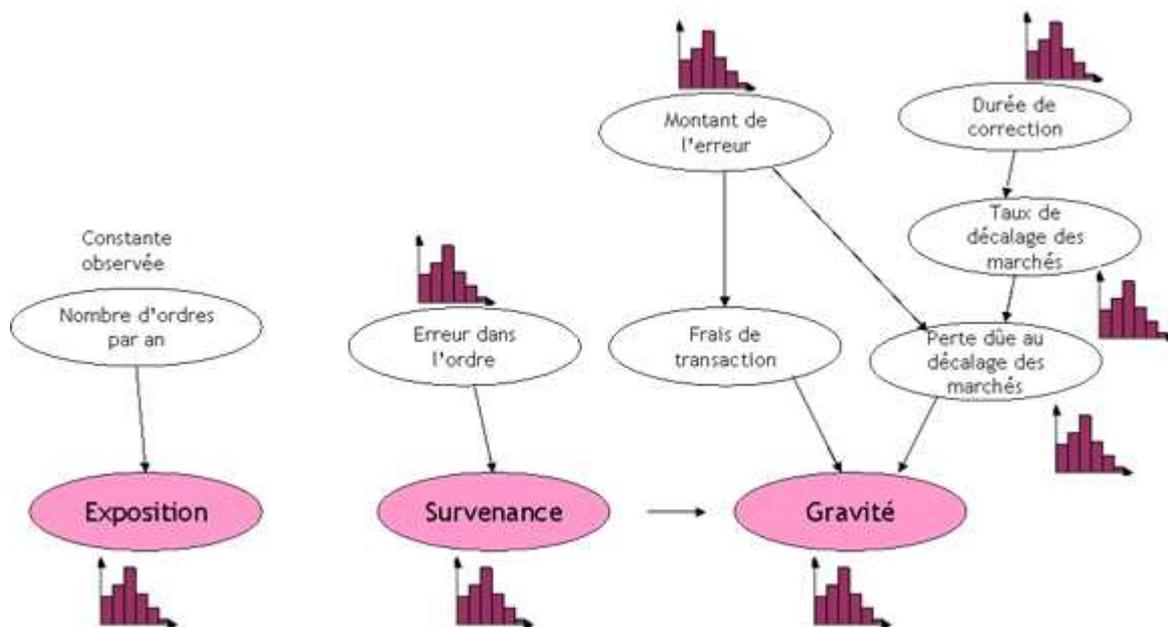
Nous constatons en effet, comme nous le redoutions, que le résultat est très sensible au choix du nombre de « 9 » après la virgule. Nous devons donc interpréter les résultats avec beaucoup de précaution.

### 8.3 MODÈLE BAYÉSIEN

Nous appliquons à présent le modèle bayésien à notre risque n°24. Ce risque est traité de façon très simple car le but est de montrer comment s'applique l'approche Bayésienne.

La première étape dans l'approche bayésienne consiste à créer le réseau à l'aide d'un graphe et à définir les distributions correspondant à chaque facteur. Ces paramètres pourront être revus dans une phase de « back-testing ».

Une analyse du risque nous a permis de construire le graphe suivant :



Remarque : nous avons négligé les sanctions réglementaires.

Rappelons que l'approche bayésienne proposée pour le risque opérationnel consiste à définir trois objets : l'exposition, la survenance et la gravité.

L'exposition : l'exposition doit correspondre aux objets exposés de l'entreprise qui ne peuvent être touchés par le risque qu'une seule fois dans la période. Nous avons choisis les ordres, en effet un ordre ne peut être erroné qu'une seule fois. Nous n'anticipons pas d'augmentation dans le nombre d'ordres pour l'année à venir et le nombre d'ordres observés est en moyenne de 25 000 par an.

La survenance : la survenance est définie comme l'erreur sur un ordre. En moyenne, le SRA nous fournit un nombre de pertes annuelles moyen de 18,2 soit une probabilité d'erreur sur un ordre de  $18,2/25\ 000 = 0,0728\%$ .

La gravité : la gravité est définie comme la somme de la perte due au décalage des marchés et des frais de transaction. Nous faisons les hypothèses suivantes sur les facteurs intervenant dans le calcul de la gravité :

- le montant de l'erreur (exprimé en millions d'euros) :

Montant de l'erreur (en millions d'euros)	
5	66%
15	18%
50	16%

- la durée de correction de l'erreur : c'est la période allant de la date de survenance de l'erreur à sa date de correction par le passage d'un nouvel ordre :

Durée de correction (en jours)	
0,125	66%
1	33%
90	1%

- les frais de transaction : nous supposons que les frais de transaction sont égaux à 0,25% du montant de la transaction. Le montant d'erreur est passé une première fois puis une seconde fois lors de la correction. Les frais seront ici calculés en multipliant le montant de l'erreur par 0,5%.
- le taux de décalage des marchés : il s'agit de la variation des taux observée pendant la durée de correction de l'erreur.

Taux de décalage des marchés	Durée de correction (en jours)		
	0,125	1	90
2%	66%	62%	60%
5%	34%	37,99%	28%
30%	0%	0,01%	12%

- La perte due au décalage des marchés : elle est calculée en multipliant le montant de l'erreur par le taux de décalage.

Ce réseau bayésien est ensuite construit sous un logiciel spécifique, ici nous utilisons la version d'essai d'Hugin. Ce logiciel nous permet d'obtenir un élément important de l'algorithme ci-dessous.

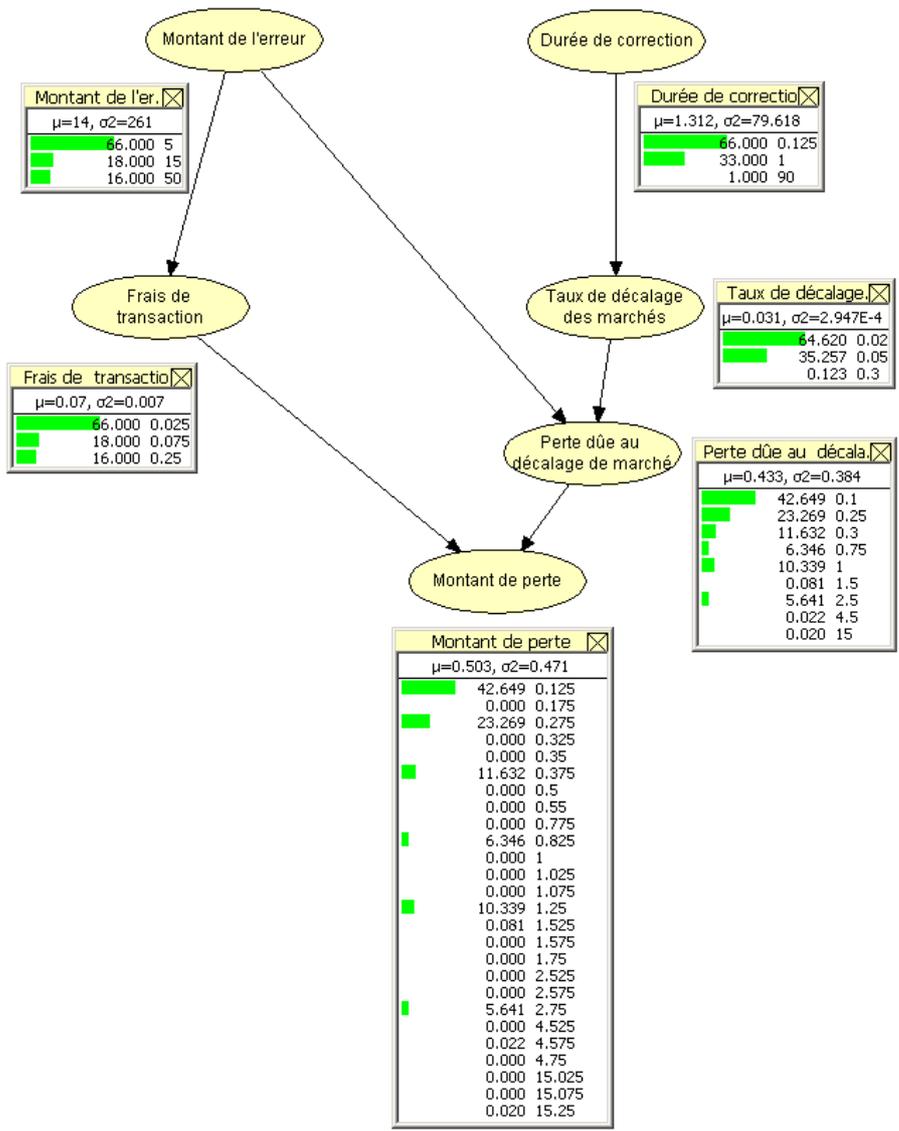
Lançons-nous dans l'algorithme décrit dans la partie théorique :

Ici, les objets exposés sont du même type, ce sont des ordres de même nature :  $X_i = X$ , ce qui implique que  $P_i = P(\text{Exposition} = X_i) = 1$ .

Nous en déduisons la probabilité que le risque survienne sachant que l'exposition est  $X_i$  :  $PS_i = P(\text{Survenance} = \text{"Oui"} \mid \text{Exposition} = X_i) = P(\text{Survenance} = \text{"Oui"}) = 0,0728\%$

L'algorithme consiste à réaliser successivement les étapes suivantes :

1) positionner l'exposition à  $X_i$ , la survenance à Oui dans le réseau bayésien et lire la distribution de la gravité  $PG_i = P(\text{Gravité} \mid \text{Survenance} = \text{"Oui"} \text{ et } \text{Exposition} = X_i)$  : cette étape est réalisée avec le logiciel Hugin qui nous permet d'avoir très rapidement la distribution de la gravité conditionnellement à la survenance.



2) Échantillonner le nombre de pertes  $F_i$  selon la loi Binomiale  $B(nb(X_i); PS_i) = B(25\,000; 0,0728\%)$ . En effet, les ordres sont indépendants et chaque ordre suit une loi de Bernoulli.

3) Pour chacun des incidents de 1 à  $F_i$ , échantillonner la gravité suivant la distribution  $PG_i$

4) Sommer les  $F_i$  gravités.

Nous répétons ces 4 étapes 10 000 fois en conservant les sommes des gravités à chaque fois. Nous obtenons ainsi une distribution de pertes totales et nous en déduisons la VaR à 99,5% en ordonnant les pertes et en prenant la 50<sup>ième</sup> perte en partant du bas.

Nous avons réalisé des tests de sensibilité aux différents paramètres du réseau bayésien : à chaque jeu de tests, nous faisons varier les paramètres d'un facteur en fixant les paramètres de tous les autres facteurs.

Le tableau suivant représente les tests de sensibilité de la fréquence d'occurrence. Le premier scénario est le scénario central.

Tests de sensibilité de la fréquence d'occurrence										
Simulation n°1										
Occurrence	Montant de l'erreur		Durée de correction		Décalage des marchés	Durée de correction			Perte totale	
oui	0,0728%	5	66%	0,125		66%	0,125	1		90
non	99,9272%	15	18%	1	33%	2%	66%	62%	60%	
		50	16%	90	1%	5%	34%	37,99%	28%	
						30%	0%	0,01%	12%	
Simulation n°2										
Occurrence	Montant de l'erreur		Durée de correction		Décalage des marchés	Durée de correction			Perte totale	
oui	0,0500%	5	66%	0,125		66%	0,125	1		90
non	99,9500%	15	18%	1	33%	2%	66%	62%	60%	
		50	16%	90	1%	5%	34%	37,99%	28%	
						30%	0%	0,01%	12%	
Simulation n°3										
Occurrence	Montant de l'erreur		Durée de correction		Décalage des marchés	Durée de correction			Perte totale	
oui	0,10%	5	66%	0,125		66%	0,125	1		90
non	99,90%	15	18%	1	33%	2%	66%	62%	60%	
		50	16%	90	1%	5%	34%	37,99%	28%	
						30%	0%	0,01%	12%	
Simulation n°4										
Occurrence	Montant de l'erreur		Durée de correction		Décalage des marchés	Durée de correction			Perte totale	
oui	0,20%	5	66%	0,125		66%	0,125	1		90
non	99,80%	15	18%	1	33%	2%	66%	62%	60%	
		50	16%	90	1%	5%	34%	37,99%	28%	
						30%	0%	0,01%	12%	

Le tableau suivant représente les tests de sensibilité du montant d'erreur.

Tests de sensibilité du montant d'erreur										
Simulation n°1										
Occurrence		Montant de l'erreur		Durée de correction		Décalage des marchés	Durée de correction			Perte totale
oui	0,07%	5	60,0%	0,125	66%		0,125	1	90	
non	99,93%	15	20,0%	1	33%	2%	66%	62%	60%	
		50	20,0%	90	1%	5%	34%	38%	28%	
						30%	0%	0%	12%	
Simulation n°2										
oui	0,07%	5	50,0%	0,125	66%	0,125	1	90	24,175	
non	99,93%	15	30,0%	1	33%	2%	66%	62%		60%
		50	20,0%	90	1%	5%	34%	38%		28%
						30%	0%	0%	12%	
Simulation n°3										
oui	0,07%	5	40,0%	0,125	66%	0,125	1	90	30,45	
non	99,93%	15	30,0%	1	33%	2%	66%	62%		60%
		50	30,0%	90	1%	5%	34%	38%		28%
						30%	0%	0%	12%	
Simulation n°4										
oui	0,07%	5	30,0%	0,125	66%	0,125	1	90	31,375	
non	99,93%	15	40,0%	1	33%	2%	66%	62%		60%
		50	30,0%	90	1%	5%	34%	38%		28%
						30%	0%	0%	12%	
Simulation n°5										
oui	0,07%	5	30,0%	0,125	66%	0,125	1	90	35,15	
non	99,93%	15	30,0%	1	33%	2%	66%	62%		60%
		50	40,0%	90	1%	5%	34%	38%		28%
						30%	0%	0%	12%	

Le tableau suivant représente les tests de sensibilité de la durée de correction.

Tests de sensibilité de la durée de correction										
<b>Simulation n°1</b>										
Occurrence	Montant de l'erreur		Durée de correction		Décalage des marchés	Durée de correction			Perte totale	
oui 0,07%	5	66%	0,125	60%		0,125	1	90		
non 99,93%	15	18%	1	30%	2%	66%	62%	60%	19,575	
	50	16%	90	10%	5%	34%	37,99%	28%		
					30%	0%	0,01%	12%		
<b>Simulation n°2</b>										
Occurrence	Montant de l'erreur		Durée de correction		Décalage des marchés	Durée de correction			Perte totale	
oui 0,07%	5	66%	0,125	50%		0,125	1	90		
non 99,93%	15	18%	1	40%	2%	66%	62%	60%	28,78	
	50	16%	90	10%	5%	34%	37,99%	28%		
					30%	0%	0,01%	12%		
<b>Simulation n°3</b>										
Occurrence	Montant de l'erreur		Durée de correction		Décalage des marchés	Durée de correction			Perte totale	
oui 0,07%	5	66%	0,125	50%		0,125	1	90		
non 99,93%	15	18%	1	30%	2%	66%	62%	60%	37,20	
	50	16%	90	20%	5%	34%	37,99%	28%		
					30%	0%	0,01%	12%		
<b>Simulation n°4</b>										
Occurrence	Montant de l'erreur		Durée de correction		Décalage des marchés	Durée de correction			Perte totale	
oui 0,07%	5	66%	0,125	40%		0,125	1	90		
non 99,93%	15	18%	1	40%	2%	66%	62%	60%	32,35	
	50	16%	90	20%	5%	34%	37,99%	28%		
					30%	0%	0,01%	12%		

Le tableau suivant représente les tests de sensibilité de la variation des actifs.

Tests de sensibilité du taux de décalage des marchés										
Simulation n°1										
Occurrence		Montant de l'erreur		Durée de correction		Décalage des marchés	Durée de correction			Perte totale
oui	0,07%	5	66%	0,125	66%		0,125	1	90	
non	99,93%	15	18%	1	33%	2%	66,00%	62,00%	60,00%	21,375
		50	16%	90	1%	5%	33,99%	37,90%	27,00%	
						30%	0,01%	0,10%	13,00%	
Simulation n°2										
Occurrence		Montant de l'erreur		Durée de correction		Décalage des marchés	Durée de correction			Perte totale
oui	0,07%	5	66%	0,125	60%		0,125	1	90	
non	99,93%	15	18%	1	30%	2%	66,00%	62,00%	60,00%	22,65
		50	16%	90	10%	5%	33,90%	37,80%	27,00%	
						30%	0,10%	0,20%	13,00%	
Simulation n°3										
Occurrence		Montant de l'erreur		Durée de correction		Décalage des marchés	Durée de correction			Perte totale
oui	0,07%	5	66%	0,125	50%		0,125	1	90	
non	99,93%	15	18%	1	40%	2%	65,00%	62,00%	60,00%	23,875
		50	16%	90	10%	5%	34,90%	37,50%	26,00%	
						30%	0,10%	0,50%	14,00%	
Simulation n°4										
Occurrence		Montant de l'erreur		Durée de correction		Décalage des marchés	Durée de correction			Perte totale
oui	0,07%	5	66%	0,125	50%		0,125	1	90	
non	99,93%	15	18%	1	30%	2%	60,00%	60,00%	50,00%	28,83
		50	16%	90	20%	5%	39,00%	39,00%	34,00%	
						30%	1,00%	1,00%	16,00%	

Remarque : les paramètres du modèle ont été fixés par les experts mais nous aurions très bien pu prendre des observations et utiliser les méthodes statistiques d'apprentissage pour définir les distributions des différents facteurs.

#### **8.4 COMPARAISON DES RÉSULTATS ET CONCLUSIONS**

Les résultats des deux modèles sont très comparables : le modèle simplifié nous donne 21,94M€ et le modèle bayésien 21,275M€. Cependant nous avons vu que le modèle simplifié était très sensible au choix du quantile extrême. L'approche simplifiée nous donne une première idée du résultat mais nous ne pouvons pas nous y limiter. Elle doit absolument être complétée par une autre approche.

Le modèle bayésien est une approche beaucoup plus transparente. Il permet de prendre en compte à la fois des facteurs quantitatifs mais aussi des facteurs qualitatifs, ce que ne font pas la plupart des modèles. L'approche bayésienne est très intéressante dans le cadre des risques opérationnels car elle permet de détecter des facteurs de réduction du risque grâce à l'inférence. Elle peut être utilisée à d'autres fins que l'évaluation des fonds propres : le risk management peut l'utiliser pour mettre en place des plans d'actions, constater l'efficacité de ces derniers, voir l'impact des aggravations de certains facteurs sur les risques.



## 9 APPLICATION DE LA THÉORIE DES VALEURS EXTRÊMES

L'idée est de découper notre échantillon en deux sous ensembles : le premier pour les pertes attritionnelles et le second pour les pertes exceptionnelles.

Pour cela, nous utilisons la méthode « POT » décrite plus haut appliquée aux pertes exceptionnelles et nous modélisons nos pertes attritionnelles de façon classique avec la méthode LDA.

### 9.1 MODÉLISATION DES PERTES EXTRÊMES

Nous rappelons que la méthode « POT » consiste à déterminer un seuil de graves au-delà duquel les excès suivent une loi GDP.

Dans un premier temps, nous allons faire des tests graphiques pour regarder si l'espérance et la variance de nos risques sont finies ou infinies : cela nous donnera une première idée de la valeur de l'indice de queue. En effet, si  $X$  est une variable aléatoire de loi GPD, alors :

$E(X)$  finie  $\Leftrightarrow$  indice de queue  $< 1$

$V(X)$  finie  $\Leftrightarrow$  indice de queue  $< 1/2$

Pour cela, considérons  $S_n(r) = \frac{\max(X_1^r, \dots, X_n^r)}{\sum_{i=1}^n X_i^r}$  pour  $r > 0$ .

D'après Embrecht, nous avons la propriété suivante :

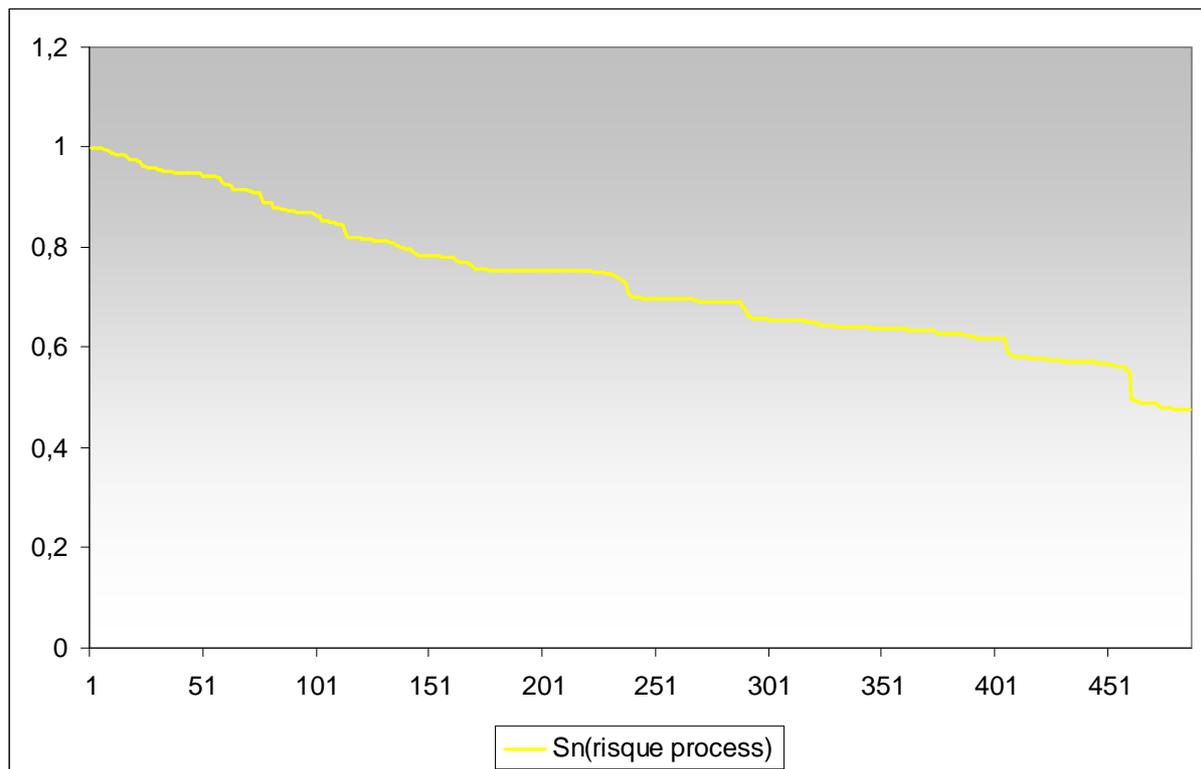
$$S_n(r) \xrightarrow{n \rightarrow +\infty} 0 \text{ ps} \Leftrightarrow E(X^r) < +\infty$$

Pour  $r=1$ , nous pourrions savoir si l'espérance est finie ou infinie en traçant  $S_n(r)$  en fonction de  $n$  : si  $S_n(r)$  tend vers 0 quand  $n$  tend vers l'infini, alors l'espérance sera finie. De même pour  $r=2$ , nous regarderons si la variance est finie ou infinie.

Sur nos échantillons qui sont de taille limitée, il sera difficile de conclure mais cela peut déjà nous donner une idée.

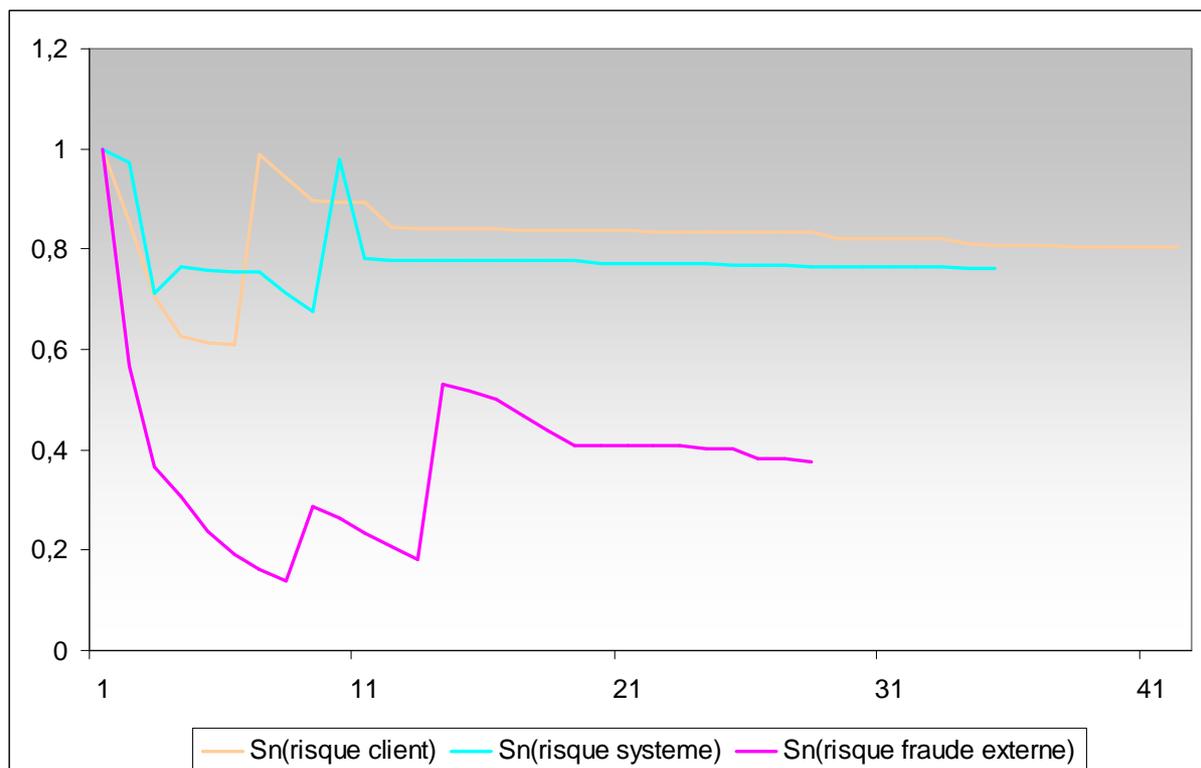
Représentons  $S_n(r)$  pour  $r = 1$  en fonction de  $n$  pour les quatre risques étudiés.

Le graphique ci-dessous représente  $S_n(1)$  pour le risque de processus :



Nous observons que  $S_n(1)$  décroît rapidement, mais nous ne pouvons pas assurer qu'il tende vers 0. Nous pouvons présumer que l'espérance du risque de processus est finie avec des réserves.

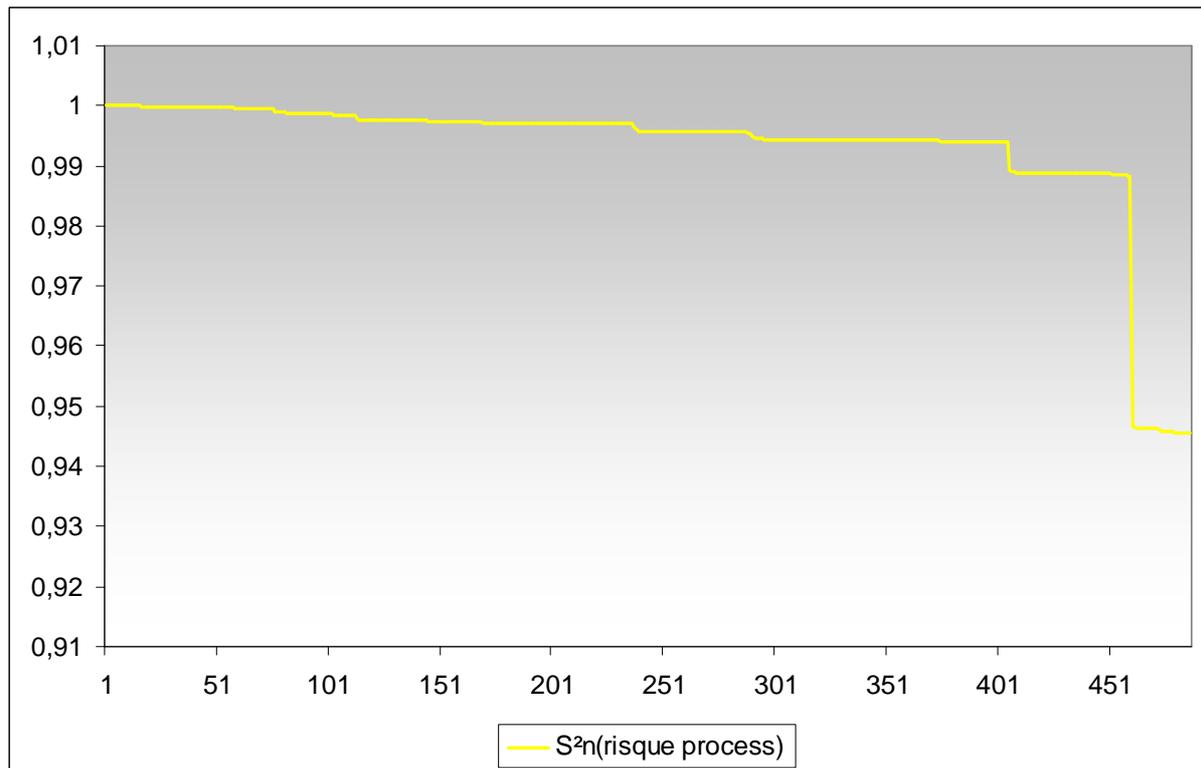
Le graphique ci-dessous représente  $S_n(1)$  pour les risques client, système et fraude externe :



En observant ce graphique, nous pourrions penser que  $S_n(1)$  restera au même niveau et ne décroîtra pas vers zéro pour les trois risques mais, étant donné le faible nombre de données, nous ne pouvons pas nous prononcer.

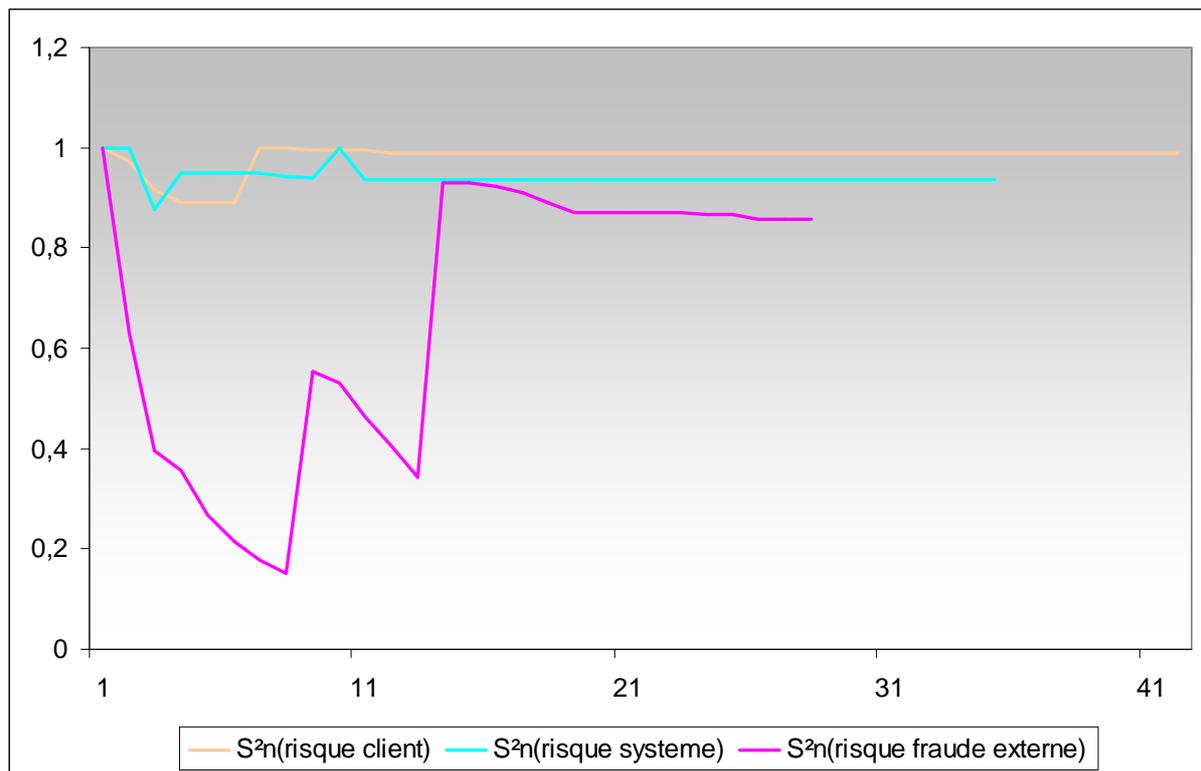
Représentons  $S_n(r)$  pour  $r = 2$  en fonction de  $n$  pour les quatre risques étudiés.

Le graphique ci-dessous représente  $S_n(2)$  pour le risque de processus :



$S_n(2)$  décroît lentement, nous pourrions penser qu'il tende vers 0, ce qui pourrait s'interpréter par le fait que la variance du risque de processus est finie mais nous émettons des réserves sur cette interprétation.

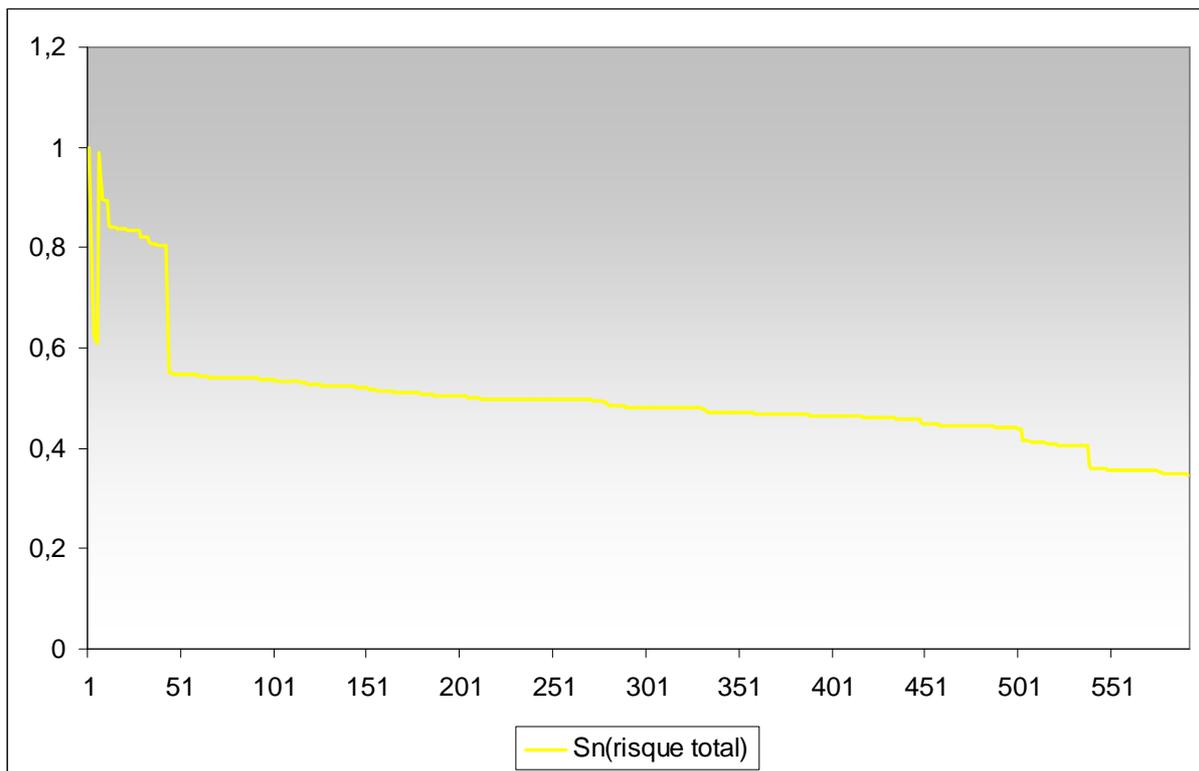
Le graphique ci-dessous représente  $S_n(2)$  pour les risques client, système et fraude externe :



Au premier abord, il semblerait que  $S_n(2)$  ne tende pas vers 0, ce qui pourrait s'interpréter par le fait que les variances sont infinies mais nous émettons des réserves sur cette interprétation vu le faible nombre de données.

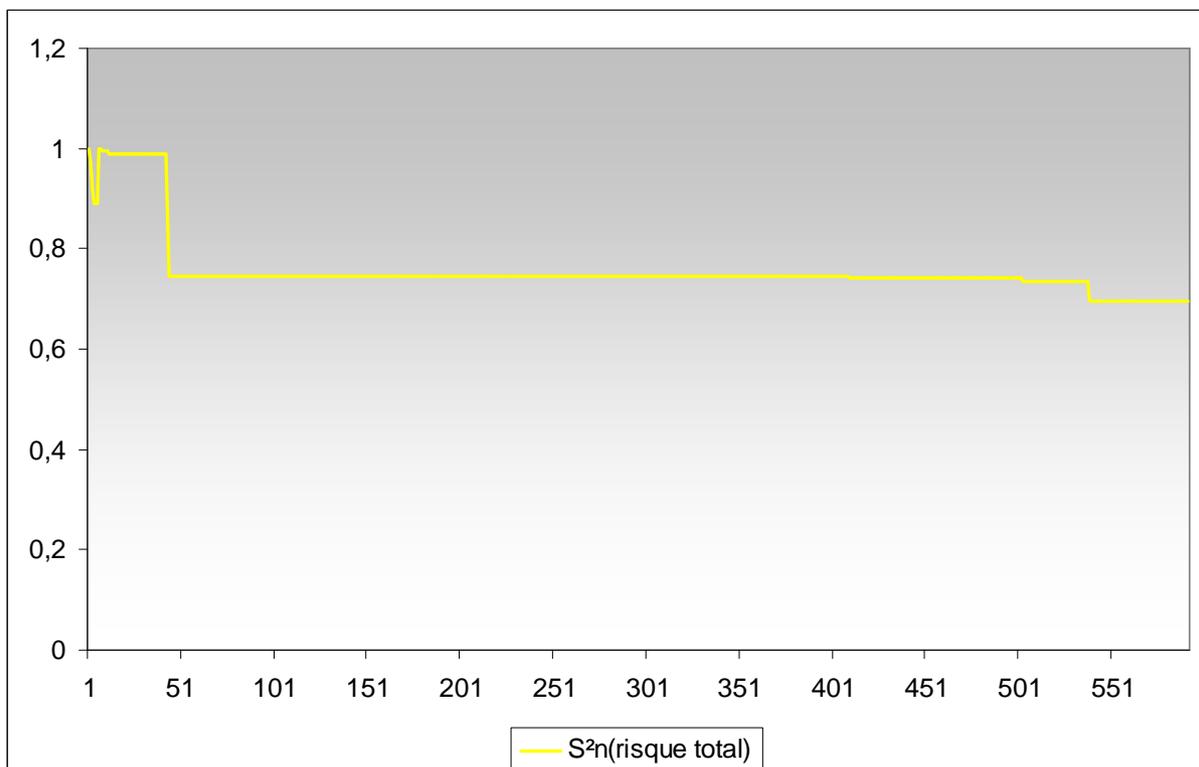
Étant donné le faible nombre d'observations pour les risques client, système et fraude externe, nous choisissons de regrouper toutes nos pertes et d'ajuster une loi GPD sur l'ensemble de nos excès.

Le graphique ci-dessous représente  $S_n(1)$  pour l'ensemble des risques :



Nous observons que  $S_n(1)$  décroît rapidement, mais nous ne pouvons pas assurer qu'il tende vers 0. Nous pourrions penser que l'espérance de l'ensemble des risques est finie et donc que l'indice de queue est inférieur à 1, avec de fortes réserves.

Le graphique ci-dessous représente  $S_n(2)$  pour l'ensemble des risques :



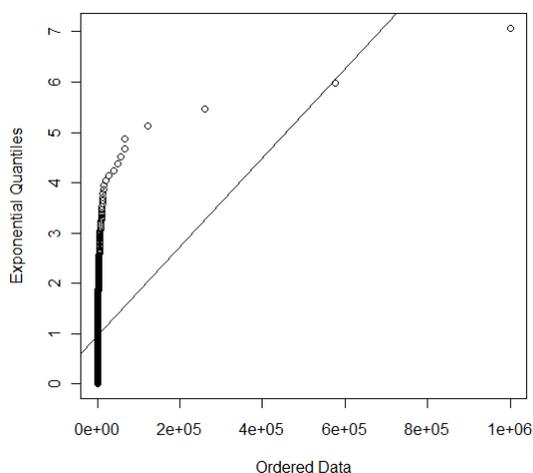
Nous observons que  $S_n(2)$  décroît très lentement, mais nous ne pouvons pas assurer qu'il tende vers 0. Nous pourrions penser qu'il ne tend pas vers zéro et donc que la variance est infinie, ce qui signifierait que l'indice de queue est supérieur à  $\frac{1}{2}$ , mais nous émettons de fortes réserves.

Ce test donne un résultat indicatif qui doit être pris simplement comme une première idée de la finitude de l'espérance et de la variance. De plus, nous avons, ici, très peu de données pour réellement conclure que l'indice de queue est compris entre  $\frac{1}{2}$  et 1, en général ce genre de test se construit sur plusieurs milliers d'observations.

A présent, nous cherchons le seuil des excès pour l'ensemble des risques. Attention à ne pas confondre ce seuil avec celui de collectes.

Nous commençons par tracer le Q-Q plot qui nous fournit une indication de l'appartenance probable à un des domaines d'attraction possibles de la distribution des pertes. Son principe est de représenter graphiquement les quantiles de la distribution exponentielle versus les valeurs ordonnées de notre échantillon.

Nous utilisons la fonction « `qplot` » avec `xi=0` sous le logiciel R.



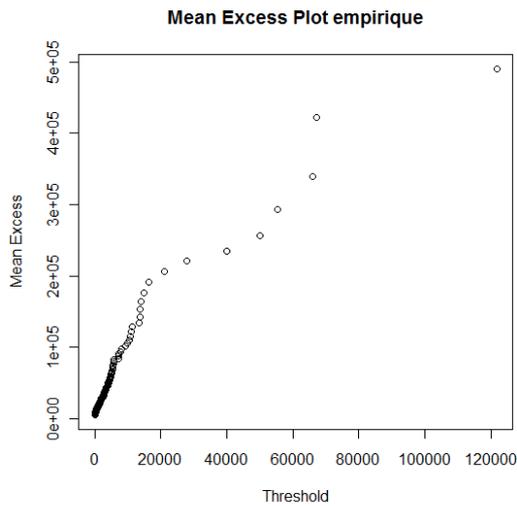
Si la loi appartient au domaine d'attraction de Gumbel alors les points sont approximativement alignés.

Si la loi appartient au domaine d'attraction de Fréchet alors les points ont tendance à se disperser vers le bas.

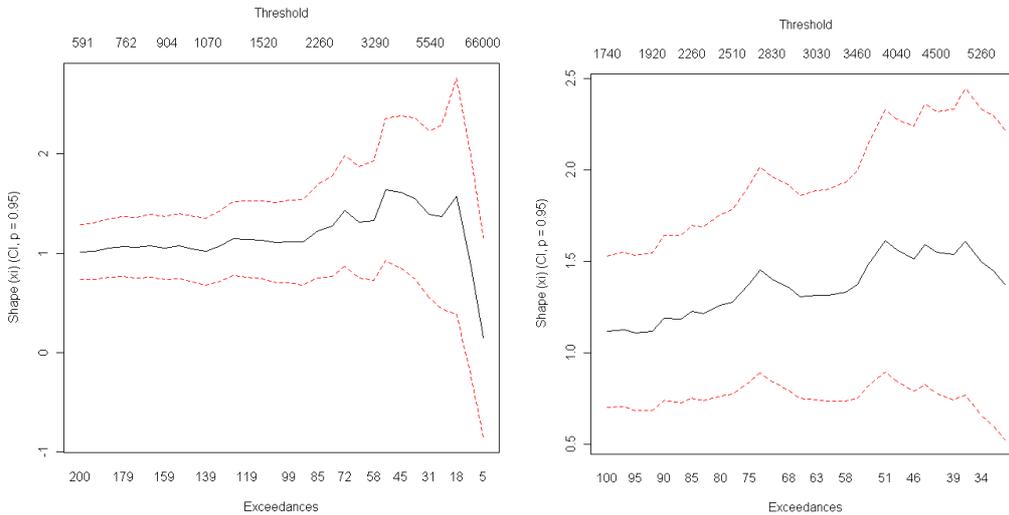
Si la loi appartient au domaine d'attraction de Weibull alors les points ont tendance à se disperser vers le haut.

Le Q-Q plot nous montre que la loi appartient au domaine d'attraction de Fréchet.

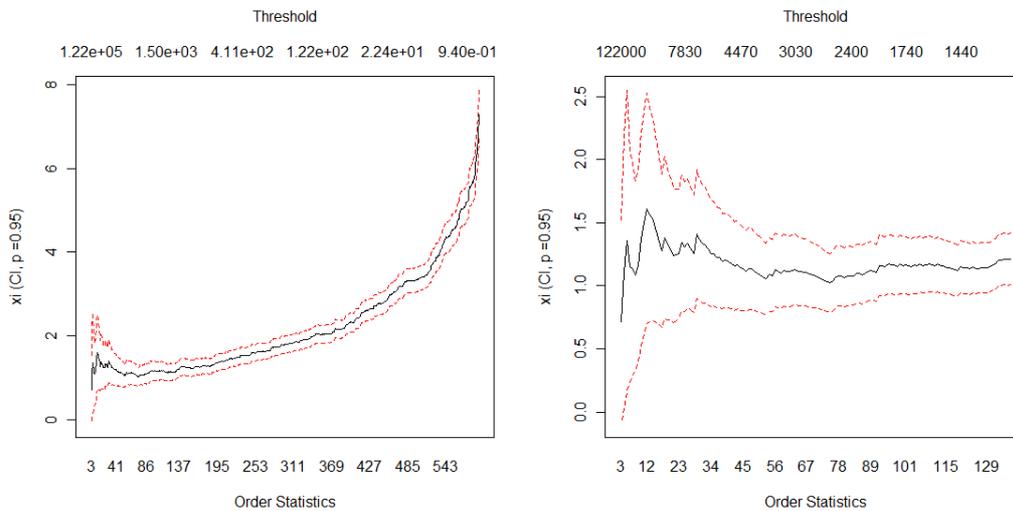
Puis nous traçons le mean excess plot empirique en fonction du seuil  $u$ .



Nous ne pouvons pas tirer de réelle conclusion de ce graphique ce qui nous amène à représenter l'indice de queue en fonction du seuil :



Puis, nous traçons l'estimateur de Hill (défini page 49) en fonction du seuil :



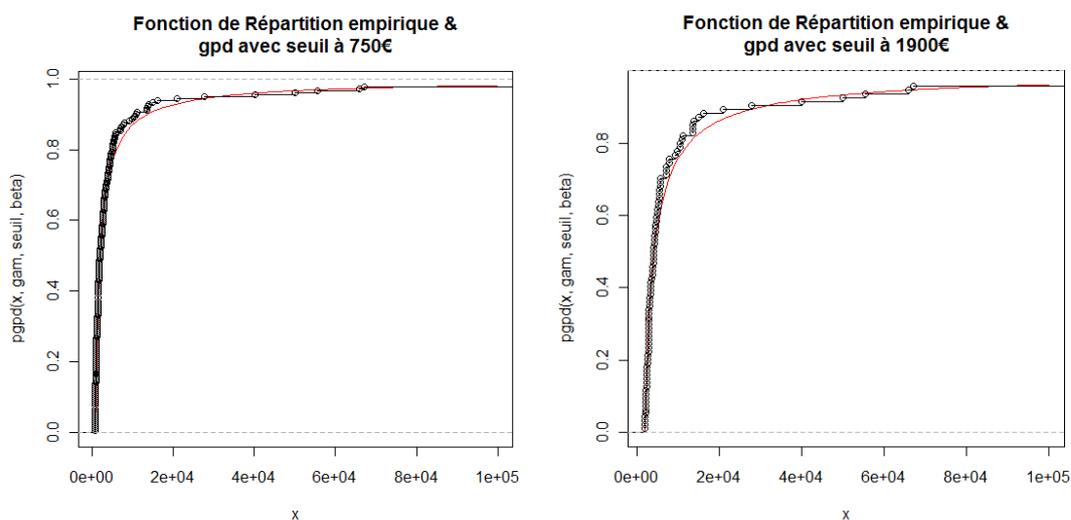
Nous recherchons le seuil autour duquel les courbes deviennent plates car l'indice de queue doit être stable ou du moins pas trop volatile par rapport au seuil.

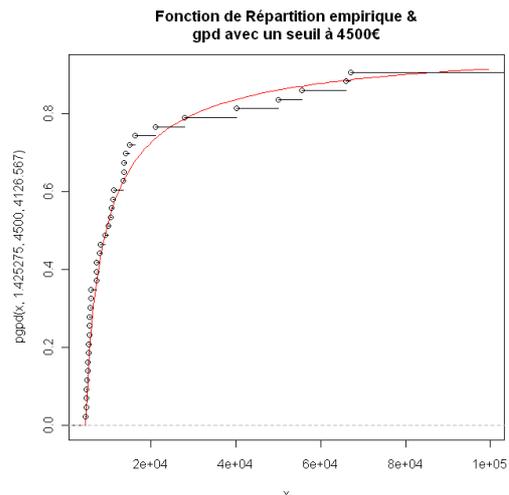
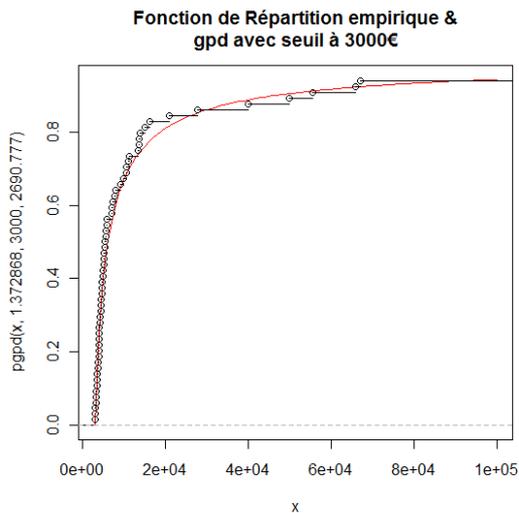
En règle générale, pour des échantillons de taille inférieure à 500, les pertes dépassant le seuil sont de l'ordre de 5 à 10%. Nous choisissons de tester plusieurs seuils qui donnent des indices de queue plus grands que 1, nous aurons donc une espérance et une variance infinie. La plupart des méthodes ne sont alors pas valides pour estimer les paramètres et nous avons eu recours à la méthode des moments pondérés généralisée qui permet une estimation pour un indice de queue  $< 3/2$ .

Nous retenons plusieurs seuils que nous allons tester. Dans le tableau ci-dessous figurent le seuil des pertes graves, le nombre des excès au-delà du seuil et les paramètres de la GPD estimés par la méthode des moments pondérés généralisée :

	Test n°1	Test n°2	Test n°3	Test n°4
Seuil	750	1 900	3 000	4 500
Nombre d'excès	183	94	64	43
Shape	1,18	1,26	1,37	1,43
Beta	1 089,61	2 061,01	2 690,78	4 126,57

Les graphiques ci-dessous représentent l'adéquation de la GPD à notre échantillon pour les différentes valeurs de seuil testées.





Remarque : l'indice de queue supérieur à 1 implique une espérance et une variance infinies. Cela va poser des problèmes lors de la simulation des excès.

## 9.2 MODÉLISATION DES PERTES ATTRITIONNELLES

Nous modélisons classiquement chaque risque pour les pertes inférieures au seuil par la méthode LDA. Ces calculs étant chronophages, nous avons tenté d'utiliser la méthode de Wilson Hilferty qui est une méthode d'approximation basée sur les moments des coûts et de la fréquence qui simule directement la charge totale. Mais elle n'a pas fonctionné dans la mesure où sur certains risques elle a fourni des pertes cumulées négatives. Dans le cadre du risque opérationnel, nous ne devons pas considérer de pertes négatives.

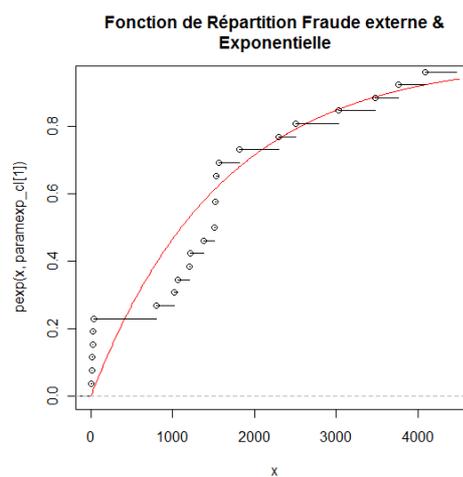
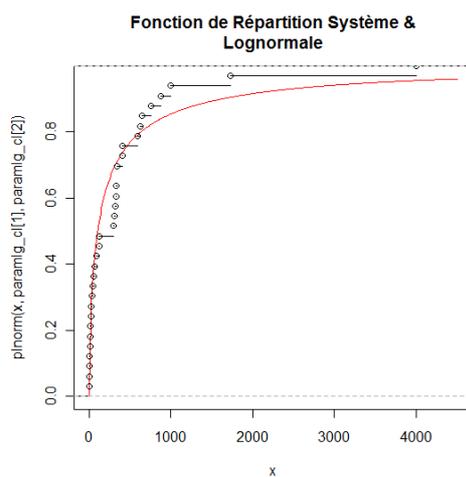
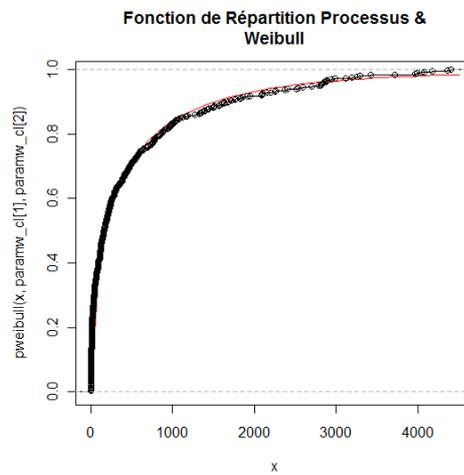
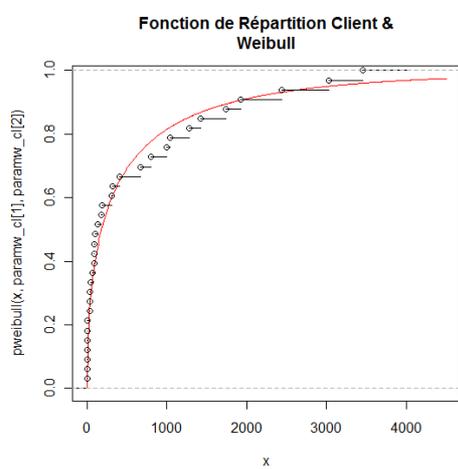
Nous avons appliqué la méthode LDA pour les différents seuils.

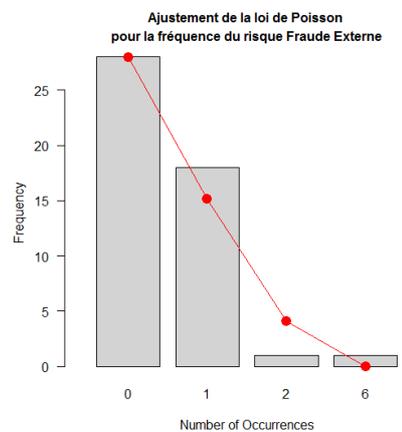
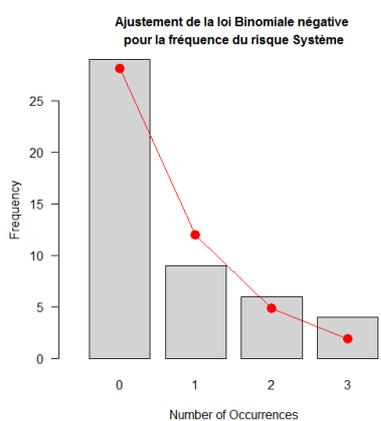
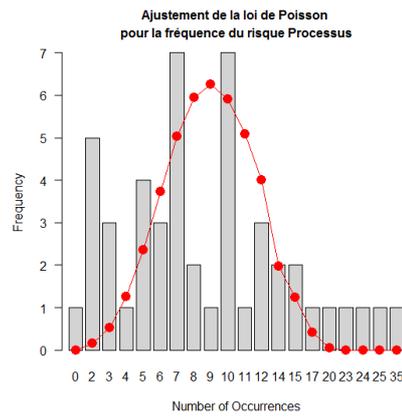
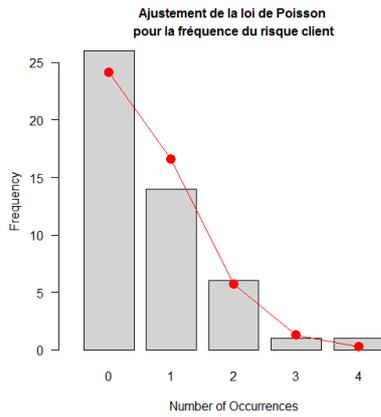
Voici les résultats obtenus pour chaque seuil ainsi que les graphiques d'adéquation :

➤ Seuil des extrêmes : 4 500€

	Risque Clients	Risque de Processus	Risque de Système	Risque de Fraude externe
Loi des coûts retenue	Weibull	Weibull	Log-normale	Exponentielle
Paramètres	Shape=0,52 Scale=367,37	Shape=0,55 Scale=332,23	Meanlog=4,65 Sdlog=2,14	Rate=0,00063
Test de KS	D = 0,12 p-value = 0,72	D = 0,045 p-value = 0,30	D = 0,21 p-value = 0,12	D = 0,21 p-value = 0,20

Loi de fréquence retenue	Poisson	Poisson	Binomiale Négative	Poisson
Paramètres	Lambda=0,69	Lambda=9,46	Size=1,11 Prob=0,62	Lambda=0,54
Test du Khi-deux	X-squared = 4,2 p-value = 0,985	X-squared = 24187 p-value = 0,82	X-squared = 9,29 p-value = 0,94	X-squared = 4,96 p-value = 0,93

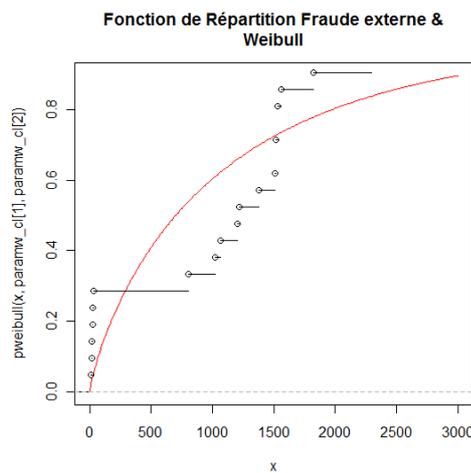
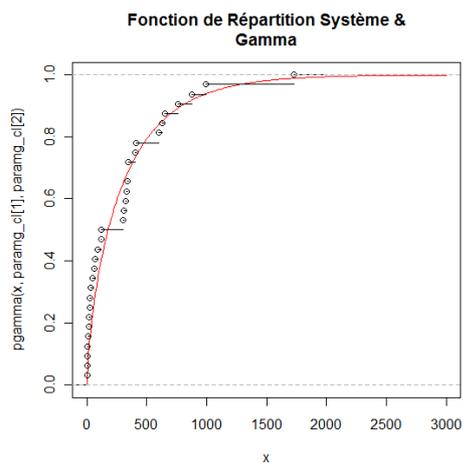
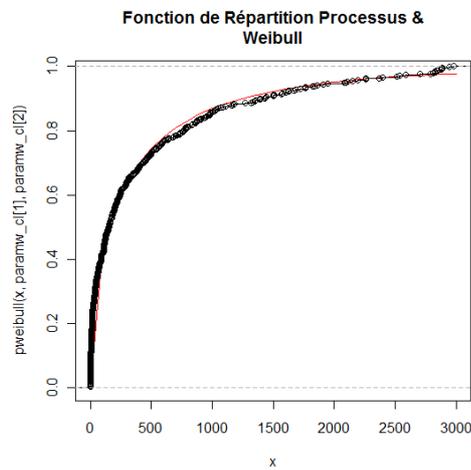
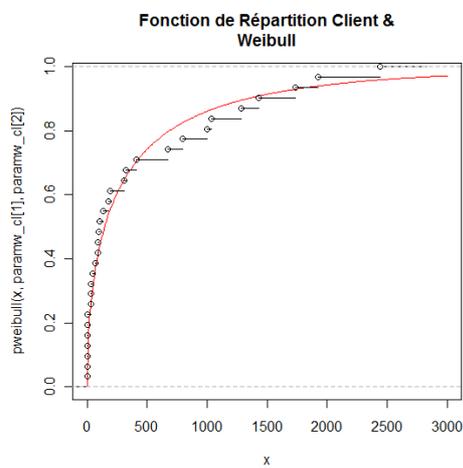


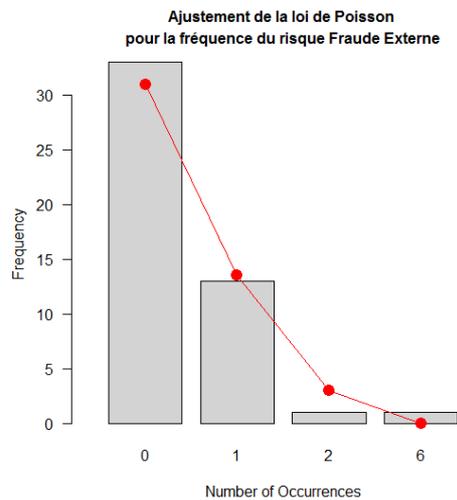
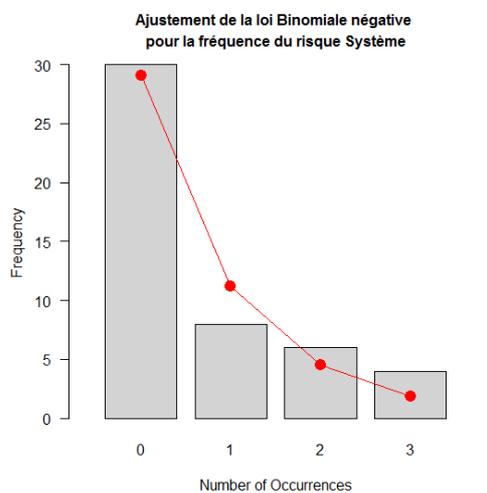
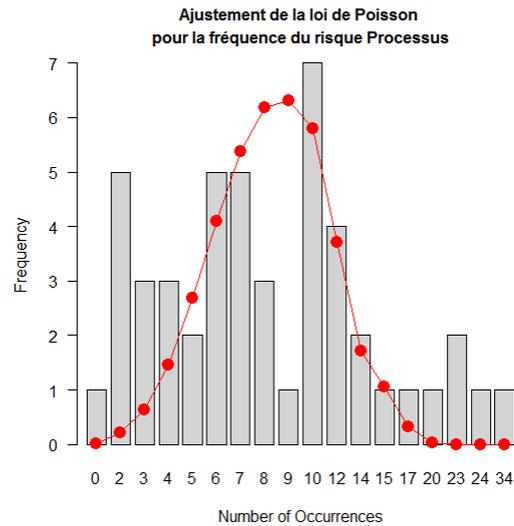
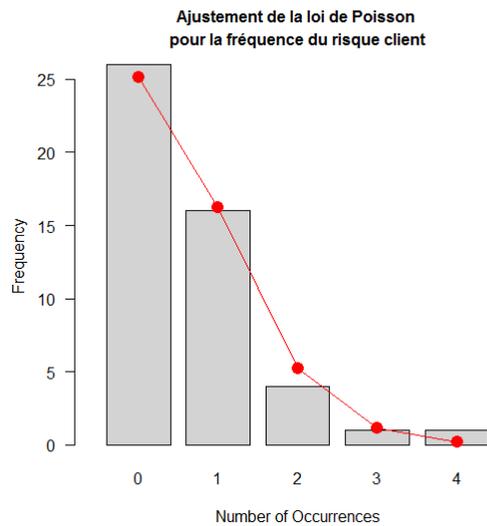


➤ Seuil des extrêmes : 3 000€

	Risque Clients	Risque de Processus	Risque de Système	Risque de Fraude externe
Loi des coûts retenue	Weibull	Weibull	Gamma	Weibull
Paramètres	Shape=0,54 Scale=284,88	Shape=0,57 Scale =289,54	Shape =0,66 Rate=0,0021	Shape=0,81 Scale=1 100,93
Test de KS	D = 0,125 p-value = 0,67	D = 0,049 p-value = 0,24	D = 0,15 p-value = 0,44	D = 0,28 p-value = 0,08

Loi de fréquence retenue	Poisson	Poisson	Binomiale Négative	Poisson
Paramètres	Lambda=0,65	Lambda=9,19	Size=0,91 Prob=0,58	Lambda=0,44
Test du Khi-deux	X-squared = 11,81 p-value = 0,3	X-squared = 193,58 p-value = 0,66	X-squared = 9,70 p-value = 0,62	X-squared = 6 p-value = 0,68

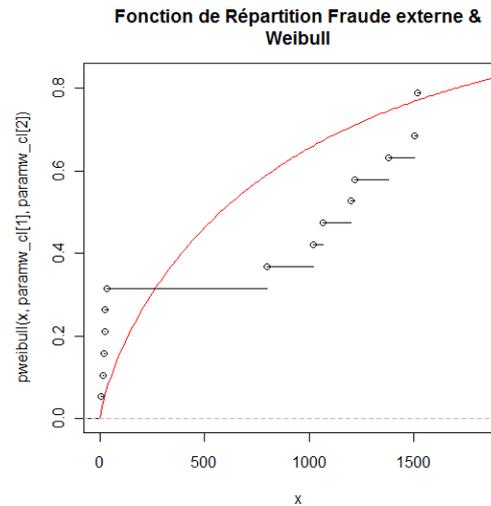
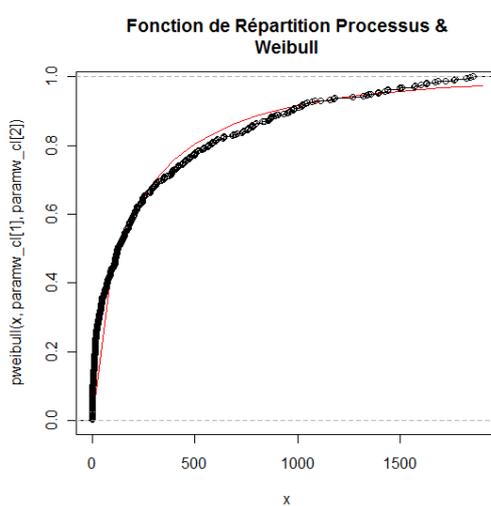
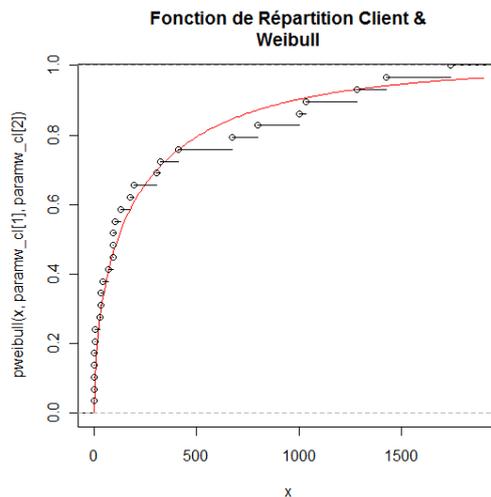
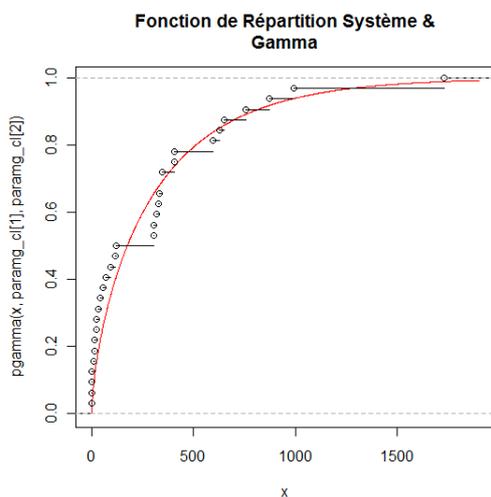


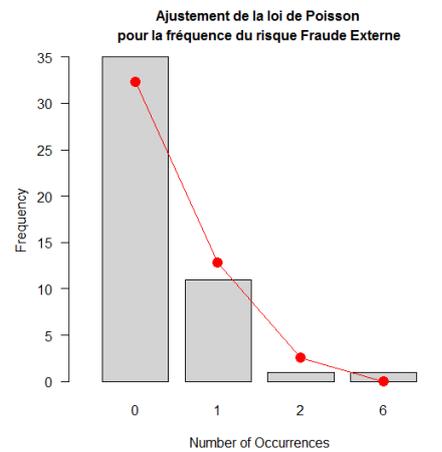
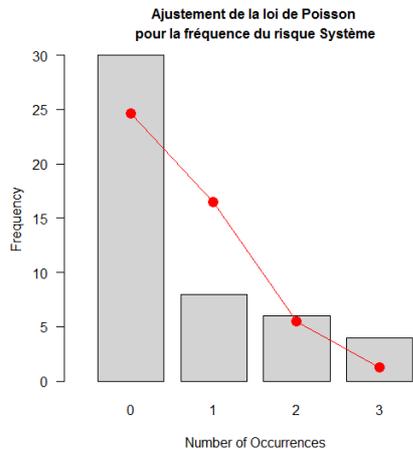
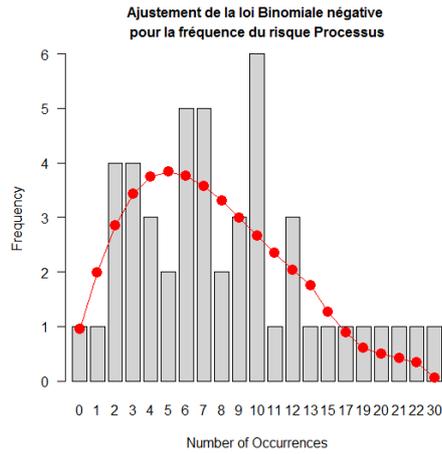
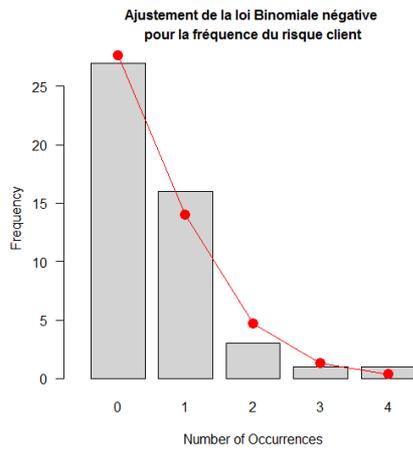


➤ Seuil des extrêmes : 1 900€

	Risque Clients	Risque de Processus	Risque de Système	Risque de Fraude externe
Loi des coûts retenue	Weibull	Weibull	Gamma	Weibull
Paramètres	Shape=0,56 Scale=220,91	Shape=0,61 Scale=225,67	Shape=0,66 Rate=0,0021	Shape=0,78 Scale=925,06
Test de KS	D = 0,13 p-value = 0,63	D = 0,056 p-value = 0,16	D = 0,15 p-value = 0,44	D = 0,29 p-value = 0,079

Loi de fréquence retenue	Binomiale Négative	Binomiale Négative	Poisson	Poisson
Paramètres	Size=3,16 Prob=0,84	Size=2,77 Prob=0,24	Lambda=0,67	Lambda=0,4
Test du Khi-deux	X-squared = 14,14 p-value = 0,51	X-squared = 422,15 p-value = 0,77	X-squared = 13,28 p-value = 0,39	X-squared = 2,55 p-value = 0,91

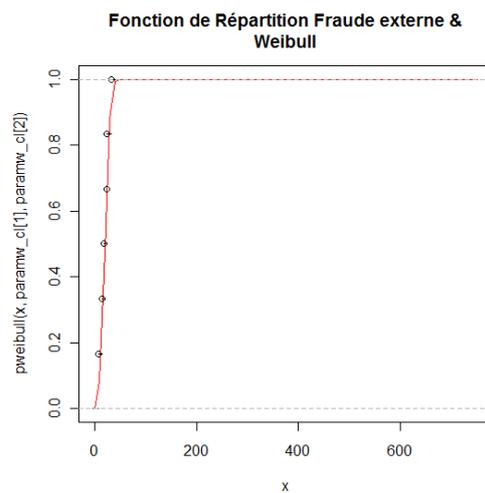
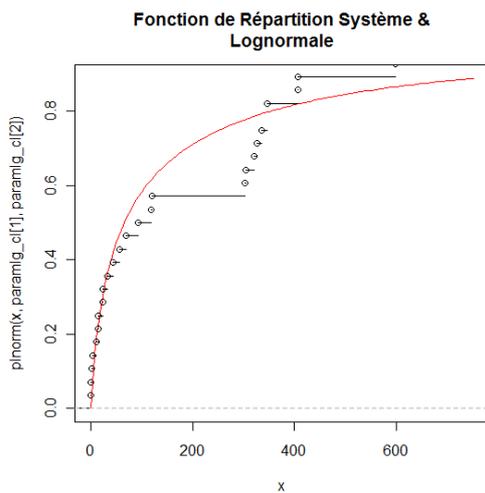
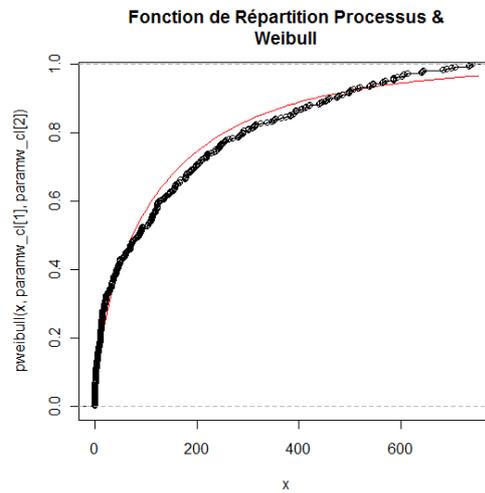
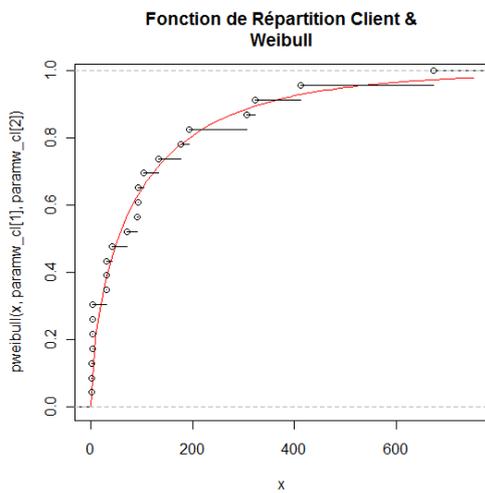


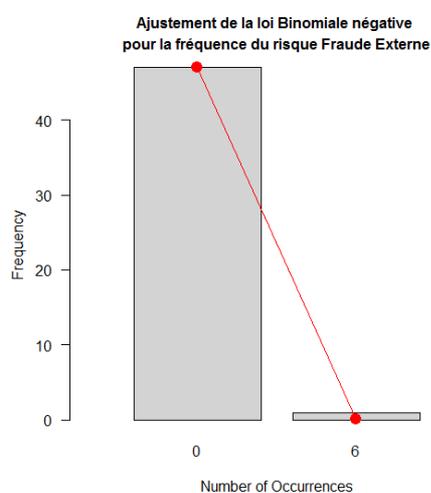
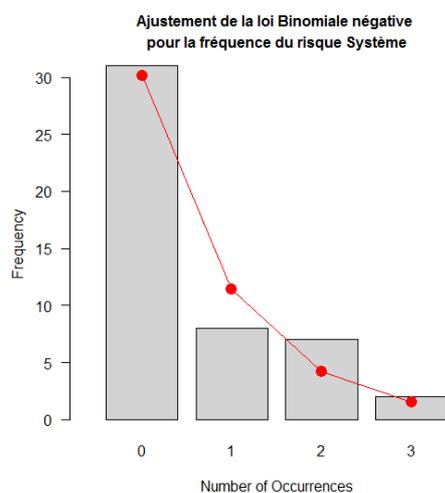
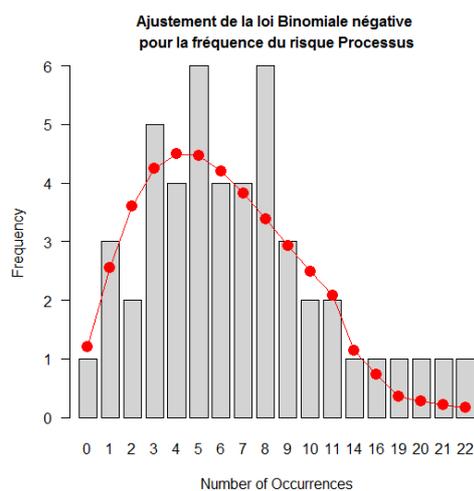
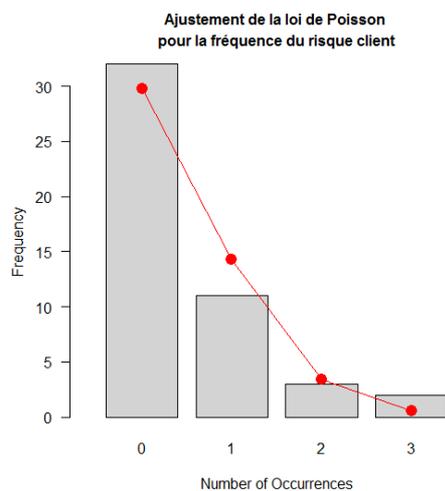


➤ Seuil des extrêmes : 750€

	Risque Clients	Risque de Processus	Risque de Système	Risque de Fraude externe
Loi des coûts retenue	Weibull	Weibull	Log-normale	Weibull
Paramètres	Shape=0,66 Scale=93,6	Shape=0,68 Scale=126,21	Meanlog=4,19 Sdlog=1,99	Shape=2,92 Scale=23,21
Test de KS	D = 0,18 p-value = 0,42	D = 0,062 p-value = 0,13	D = 0,21 p-value = 0,18	D = 0,168 p-value = 0,98

Loi de fréquence retenue	Poisson	Binomiale Négative	Binomiale Négative	Binomiale Négative
Paramètres	Lambda=0,48	Size=2,98 Prob=0,29	Size=1,07 Prob=0,65	Size=0,007 Prob=0,055
Test du Khi-deux	X-squared = 9,79 p-value = 0,29	X-squared = 386,73 p-value = 0,2	X-squared = 12,69 p-value = 0,34	X-squared = 2,48 p-value = 0,46





Pour chaque seuil, nous avons retenu des lois pour la fréquence et la sévérité de chaque risque. Cette « paramétrisation » va nous permettre par la suite de simuler les pertes attritionnelles pour chaque seuil et comparer les résultats par seuil.

### 9.3 SIMULATION DES PERTES ET RÉSULTAT

La simulation des pertes attritionnelles ne pose aucun problème technique puisqu'il s'agit d'appliquer un Monte-Carlo classique. En revanche, la simulation des pertes exceptionnelles va être plus compliquée. En effet, un indice de queue supérieur à 1 implique que la convergence de Monte-Carlo n'est plus valable puisque l'espérance est infinie. Nous pouvons alors introduire un plafond appelé sinistre maximum probable ou SMP en assurance. Nous choisissons un SMP global de 12M€ défini sur avis d'expert. La fréquence des pertes extrêmes est modélisée par une loi de Poisson dont nous affichons le lambda dans le tableau ci-dessous.

Nous avons simulé les pertes attritionnelles et extrêmes pour les différents seuils choisis afin de pouvoir comparer les résultats.

	Test n°1	Test n°2	Test n°3	Test n°4
Seuil	750	1 900	3 000	4 500
Nombre d'excès	183	94	64	43
Shape	1,18	1,26	1,37	1,43
Beta	1 089,61	2 061,01	2 690,78	4 126,57
Lambda	45,75	23,5	16	10,75
VaR(99,5%) pertes extrêmes	12 607 818	12 920 160	13 445 422	13 623 875
VaR(99,5%) Total pertes	12 623 697	12 972 241	13 496 723	13 754 789

Les résultats ne sont pas très sensibles au changement de seuil. Ils sont comparables à ceux obtenus par la méthode classique LDA. Cependant, il faut noter que le résultat est très sensible au choix du SMP, étant donné que l'espérance et la variance sont infinies.

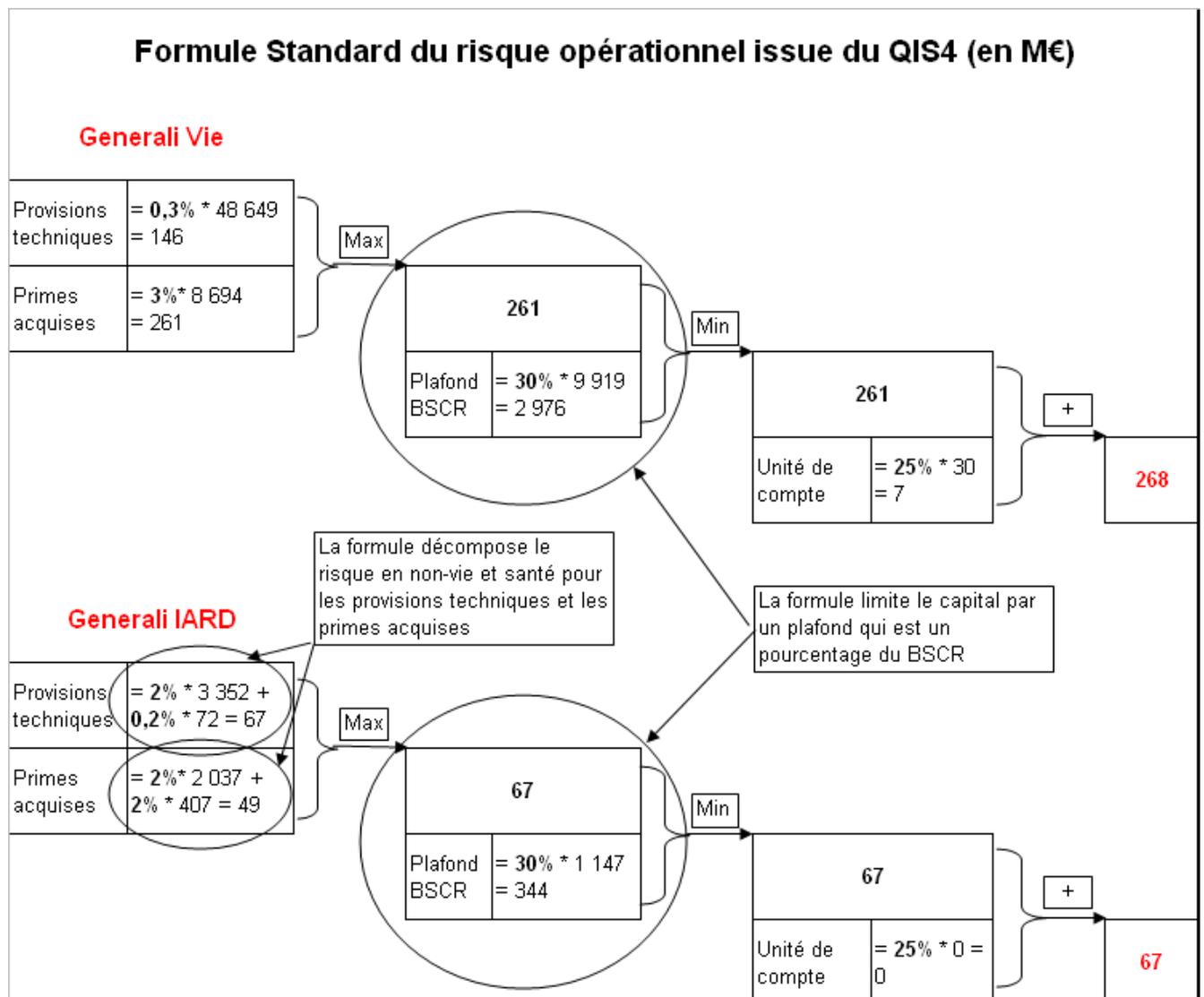
La limite de cet exemple est que nous ne connaissons pas exactement la nature des pertes reportées et par conséquent nous ne pouvons pas apporter un avis d'expert à l'analyse.

## 10 CAPITAL POUR TOUT LE RISQUE OPÉRATIONNEL

Dans cette partie, nous nous intéressons au calcul du capital pour tous les risques opérationnels. Nous présentons les résultats des deux formules standard (QIS4 et CP53) puis nous les comparons. Ensuite nous présentons les résultats du modèle simplifié en faisant l'hypothèse d'une totale corrélation entre les risques.

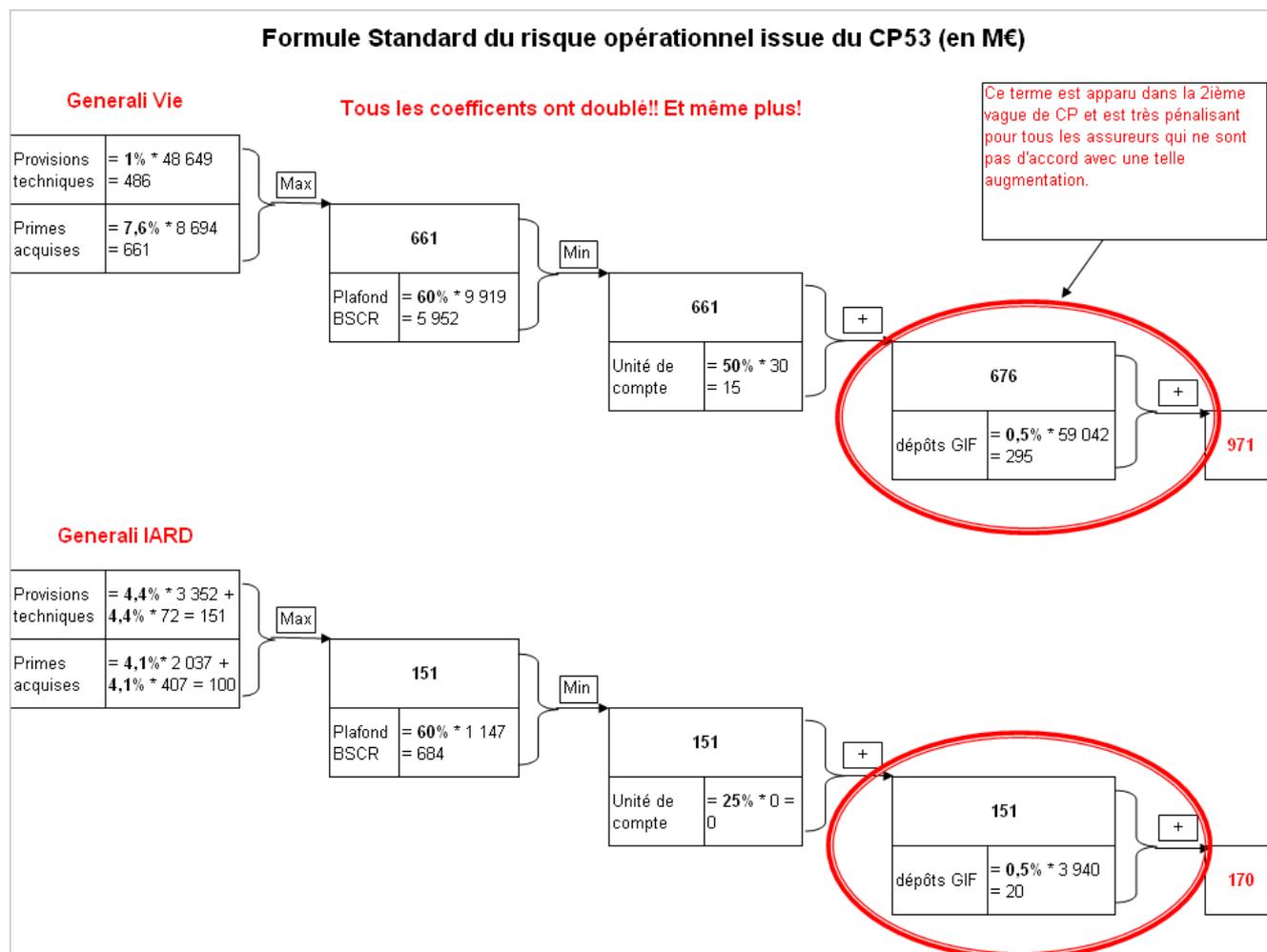
### 10.1 LA FORMULE STANDARD ISSUE DES QIS4

A l'heure actuelle, la dernière formule testée par les compagnies d'assurance est celle des QIS4. Nous présentons ci-dessous les résultats de Generali à fin 2007 :



### 10.2 LA FORMULE STANDARD ISSUE DU CP53

Nous appliquons la formule standard des CP53 sur la même vision que les QIS4 pour pouvoir faire une comparaison. Voici les résultats obtenus :



### **10.3 COMPARAISON DES RÉSULTATS ENTRE FORMULES STANDARDS QIS4 VS CP53**

A fin 2007, le besoin en capital au titre du risque opérationnel est multiplié par 3,6 pour Generali Vie et 2,5 pour Generali IARD en passant de la formule standard issue des QIS4 à la formule standard du CP53 :

<b>Comparaison des SCR risque opérationnel QIS4 et CP53 (en M€)</b>			
	QIS4	CP53	Ecart en %
Generali Vie	268	971	262%
Generali IARD	67	170	154%

## 10.4 MODÈLE INTERNE SIMPLIFIÉ

Nous appliquons la deuxième approche du modèle simplifié, en faisant l'hypothèse d'une totale corrélation entre les risques.

Voici le tableau représentant les lois retenues pour la modélisation dans la deuxième approche du modèle simplifié :

Risque	Ecart sur le scénario bas exponentielle	Ecart sur le scénario bas Weibull	Ecart sur le scénario bas Lognormale	Ecart sur le scénario bas Pareto	Ecart Min	Loi retenue
1,00	0,31	0,34	0,14	0,39	0,14	Log-normale
2,00	0,13	0,14	0,14	0,14	0,13	Exponentielle
3,00	0,15	0,02	0,35	0,11	0,02	Weibull
4,00	0,42	0,07	0,06	0,33	0,06	Log-normale
5,00	0,71	0,05	0,02	0,00	0,00	Pareto
6,00	0,08	0,10	0,01	0,03	0,01	Log-normale
7,00	0,15	0,00	0,17	0,04	0,00	Weibull
8,00	0,00	0,00	0,29	0,21	0,00	Weibull
9,00	0,82	0,08	0,00	0,57	0,00	Log-normale
10,00	0,26	0,00	0,27	0,23	0,00	Weibull
11,00	0,81	0,12	0,00	0,39	0,00	Log-normale
12,00	0,16	0,16	0,00	0,51	0,00	Log-normale
13,00	0,10	0,07	0,04	0,10	0,04	Log-normale
14,00	0,07	0,01	0,40	0,40	0,01	Weibull
15,00	0,38	0,39	0,14	0,14	0,14	Log-normale
16,00	0,74	0,06	0,02	0,41	0,02	Log-normale
17,00	0,02	0,58	0,03	0,69	0,02	Exponentielle
18,00	0,18	0,14	0,21	0,21	0,14	Weibull
19,00	0,22	0,01	0,25	0,00	0,00	Pareto
20,00	0,04	0,14	0,40	0,40	0,04	Exponentielle
21,00	0,27	0,36	0,16	0,16	0,16	Log-normale
22,00						
23,00	0,07	0,73	0,73	0,29	0,07	Exponentielle
24,00	0,29	0,04	0,21	0,10	0,01	Weibull
25,00	0,54	0,04	0,65	0,43	0,04	Weibull
<b>Total</b>						

Pour des raisons de confidentialité, nous ne donnerons pas le résultat du modèle interne simplifié mais il renvoie un SCR risque opérationnel relativement proche de celui de la formule standard des QIS4. Nous sommes confortés dans l'idée que la formule standard issue du CP53 n'est pas représentative de nos risques et donne des résultats beaucoup trop élevés.

Toutefois, ce modèle simplifié est contestable dans la mesure où il ne prend pas en compte les risques externes du type crue centennale.

## 11 CONCLUSION GÉNÉRALE

Le risque opérationnel n'est pas un risque négligeable au sens de Solvabilité II, bien au contraire. En effet, nous avons vu que la formule standard issue de la deuxième vague de CP, qui risque d'être utilisée dans les QIS5, a multiplié par plus de 3 le besoin en capital du risque opérationnel par rapport à l'ancienne formule des QIS4. La formule standard des CP est très consommatrice en fonds propres et ne reflète pas forcément nos risques. C'est pourquoi nous nous sommes intéressés à la mise en place d'un modèle interne qui quantifierait les risques plus précisément.

Nous avons proposé un modèle décomposant les risques opérationnels en deux catégories : les risques de fréquence et les risques de gravité. Les risques de fréquence ont été modélisés par une méthode Loss Distribution Approach (LDA) modifiée pour tenir compte du seuil de collecte des données. En effet, toutes les pertes ne sont pas reportées par les opérationnels, uniquement celles qui dépassent un certain seuil. Pour valider cette approche, nous avons appliqué cette méthode à des données dépassant un seuil fixé à 1 500€ pour tous les risques et nous l'avons comparé à la méthode LDA classique appliquée à nos données complètes. Nous avons constaté que l'approche LDA modifiée est plus pessimiste car elle surestime le nombre de pertes. Il faudrait notamment améliorer l'approximation de la fréquence en développant une nouvelle méthode qui ne surestime pas la fréquence.

Nous avons testé l'agrégation par les copules contre l'agrégation avec une totale corrélation. Cette application nous a montré que l'utilisation des copules pouvait faire diminuer le besoin en capital par rapport à la somme des VaR mais ce n'est pas toujours le cas. Nous devons faire des tests sur nos propres données. Par ailleurs, l'agrégation ne concerne que les risques de fréquence qui sont modélisés par des lois à queues fines. L'impact du choix de la méthode d'agrégation est ainsi limité.

Les risques de gravité ont été modélisés par des réseaux bayésiens. Ce sont des outils qui présentent de nombreux avantages : ils permettent de visualiser les liaisons de causalité entre les variables, ils peuvent prendre en compte des variables qualitatives, ils permettent de gérer les risques en identifiant des variables leviers. Ce modèle va donc être utile au risk management pour piloter les risques. Le risk management pourra l'utiliser pour mettre en place des plans d'actions, constater l'efficacité des plans d'actions, simuler l'impact d'aggravation de certains facteurs sur les risques. L'inconvénient majeur des réseaux bayésiens est qu'ils sont longs à mettre en place. Une autre solution plus rapide pour modéliser les risques de gravité serait les lois extrêmes.

Nous avons ainsi exploré la piste de la théorie des valeurs extrêmes pour modéliser nos risques de gravité. Pour cela nous avons déterminé plusieurs seuils de graves à partir de tests graphiques classiques puis nous avons séparé nos pertes au-dessus et au-dessous du seuil. Les pertes au-dessus du seuil ont été regroupées par manque de

données pour modéliser l'excédent par une loi Pareto généralisée GPD. Cette pratique peut être critiquée dans la mesure où les risques ne sont pas forcément homogènes. Nous privilégierons de décomposer par type de risques cette analyse. L'indice de queue trouvé pour les différents seuils est toujours supérieur à 1 ce qui pose un problème lors de la simulation car l'espérance et la variance sont infinies. Ce problème a été résolu en introduisant un plafond appelé Sinistre Maximum Probable (SMP). Les pertes en-dessous du seuil ont été modélisées par risque avec la méthode LDA. Les différents seuils testés donnent des résultats plutôt comparables avec des écarts de moins de 9%. La limite de cette approche est la sensibilité au choix du SMP. Cependant elle présente l'avantage d'utiliser la base de données sans avoir à réinterroger les experts, ce qui est très consommateur en ressources et qui serait long à refaire.

De plus, il faudrait envisager d'exploiter des bases de données externes type FIRST ou ORIC pour modéliser nos risques de gravité. Cela requiert un travail préalable de retraitement des données avec la mise à l'échelle de l'entreprise, la compréhension des événements, l'appropriation des risques, le traitement des seuils de collecte...

Nous avons proposé un modèle simplifié faisant appel aux avis d'expert pour avoir un premier aperçu de notre capital. Ce modèle s'est appuyé sur la méthode des quantiles et un critère de sélection des lois a priori. Les limites de ce modèle sont qu'il ne modélise pas les risques externes et qu'il est sensible au paramétrage. Nous avons appliqué ce modèle à tous les risques identifiés comme importants par la méthode top-down et étudiés dans le Self-Risk Assessment. Nous avons obtenu un capital proche de celui de la formule standard des QIS4, ce qui semble signifier que la formule standard des CP53 ne reflète pas nos risques. Certes le capital de notre modèle simplifié est à prendre avec précaution car il ne prend pas en compte les risques externes. Mais les risques externes peuvent éventuellement faire l'objet de couverture d'assurance et ne pas rentrer dans le modèle.

Nous avons proposé de modéliser les risques via une matrice de type Bâle II qui découpe les risques par type de risques de niveau 1 et par activité. Elle permet d'obtenir un capital par type de risques et activité. Cela permettra d'utiliser le modèle pour le pilotage des activités : nous pourrions imaginer de calculer et suivre périodiquement des ratios de rentabilité du risque opérationnel par activité et de fournir des objectifs de ratios cibles aux différentes directions. Mais ceci reste encore à définir.

Ce modèle sera utilisé dans le modèle interne pour évaluer un risque global. Quels sont alors les liens avec les autres risques et comment les modéliser ? Par exemple, le risque de pandémie peut à la fois toucher les assurés et les salariés de l'entreprise, il relève donc du risque de mortalité, du risque santé et c'est aussi un risque opérationnel.

La nouvelle formule standard utilisée dans les QIS5 ne vient en rien modifier les conclusions de cette étude.

**12 BIBLIOGRAPHIE**

- [1] Abram T, The Hidden Values of IT Risk Management, Isaca Journal, Volume 2, 2009.
- [2] Baud N, A, Frachot, T, Roncalli, An Internal Model for Operational Risk Computation.
- [3] Cadoux D, J-M Loizeau, Copules et dépendances : application pratique à la détermination du besoin en fonds propres d'un assureur non vie.
- [4] Cain J, Planning improvements in natural resources management, Centre for Ecology and Hydrology, UK, 2004.
- [5] Chernobai A, S Rachev, F Fabozzi, Composite Goodness-of-Fit Tests for Left-Truncated Loss Samples, Juin 2005.
- [6] Condamine L, J-P, Louisot, P, Naïm, Risk Quantification, 2006.
- [7] Cruz MG, Modelling, Measuring and Hedging Operational Risk, John Wiley & Sons, LTD, Chichester, 2002.
- [8] Dahan H, La quantification du risque opérationnel des institutions bancaires, thèse 2006.
- [9] Deloitte, The Forgotten Risk, Novembre 2005.
- [10] Denuit M, A Charpentier, Mathématiques de l'Assurance Non-Vie Tome1, 2004.
- [11] Denuit M, A Charpentier, Mathématiques de l'Assurance Non-Vie Tome2, 2005.
- [12] Dexter, N, C, Ford, P, Jakhria, P, Kelliher, D, McCall, C, Mills, A, Probyn, P, Raddall, J, Ryan, Quantifying Operational Risk in Life Insurance Companies, Mai 2006.
- [13] Diebolt J, A Guillou et I Rached, Généralisation de la méthode des moments pondérés : Application à la loi des excès.
- [14] Elseware, Guide méthodologique : Modélisation des risques opérationnels courants – Approche LDA, Juin 2009.
- [15] Lecoutre J-P, P Tassi, Statistique non paramétrique et robustesse, 1987.
- [16] Leray P, Réseaux bayésiens: apprentissage et modélisation de systèmes complexes, novembre 2006.
- [17] Jimenez C, P, Merlier, D, Chelly, Risques Opérationnels, Février 2008.
- [18] Jobst A, Consistent Quantitative Operational Risk Measurement and Regulation: Challenges of Model Specification, Data Collection, and Loss Reporting, Avril 2004.
- [19] McNeil A J, R Frey, P Embrechts, Quantitative Risk Management, 2005.

- 
- [20] Naïm P, P-H, Wuillemin, P, Leray, O, Pourret et A, Becker, Réseaux Bayésiens, 2007.
- [21] Partrat C, JL Besson, Assurance Non-Vie Modélisation, Simulation, 2005.
- [22] Planchet F, Utilisation de la théorie des valeurs extrêmes dans le contexte solvabilité 2, support de cours, Avril 2008.
- [23] Roncalli T, Introduction à la Gestion des Risques, Octobre 2001.
- [24] Tripp M H, H L, Bradley, R, Devitt, G C, Orros, G L, Overton, L M, Pryor, R A, Shaw, Quantifying Operational Risk in general Insurance Companies, Mars 2004.
- [25] Wallis JR, Hosking JRM, Parameter and quantile estimation for the Generalized Pareto Distribution, Technometrics, 1987.

**13 LISTE DES ABRÉVIATIONS**

LDA = Loss Distribution Approach

SCR = Solvency Capital Requirement

MCR = Minimum Capital Requirement

CP = Consultation paper

QIS = Quantitative impact studies

SRA = Self Risk Assessment

GPD = Loi de Pareto généralisée

VaR = Value at Risk

XSG = Modèle Exposition Survenance Gravité

SMP = Sinistre maximum probable

**14 ANNEXES****14.1 ANNEXE A : RÉFÉRENTIEL DES RISQUES BÂLE II**

Catégories d'événements (Niveau 1)	Définition	Sous-catégories (Niveau 2)	Exemples (Niveau 3)
Fraude interne	Pertes dues à des actes visant à frauder, détourner des biens ou à tourner des règlements, la législation ou la politique de l'entreprise (à l'exception des atteintes à l'égalité et des actes de discrimination) impliquant au moins une partie interne à l'entreprise	Activité non autorisée	Transactions non notifiées (intentionnellement) Transactions non autorisées (avec perte financière) Évaluation erronée d'une position (intentionnellement)
		Vol et fraude	Fraude/fraude au crédit/absence de provisions Vol/extorsion/détournement de fonds/vol qualifié Détournement de biens Destruction malveillante de biens Contrefaçon Falsification de chèques Contrebande Usurpation de compte/d'identité/etc. Fraude/évasion fiscale (délibérée) Corruption/commissions occultes Délit d'initié (pas au nom de l'entreprise)
Fraude externe	Pertes dues à des actes visant à frauder, détourner des biens ou contourner la législation de la part d'un tiers	Vol et fraude	Vol/vol qualifié Contrefaçon Falsification de chèques
		Sécurité des systèmes	Dommages dus au piratage informatique Vol d'informations (avec perte financière)
Pratiques en matière d'emploi et sécurité sur le lieu de travail	Pertes résultant d'actes non conformes à la législation ou aux conventions relatives à l'emploi, la santé ou la sécurité, de demandes d'indemnisation au titre d'un dommage personnel ou d'atteintes à l'égalité/actes de discrimination	Relations de travail	Questions liées aux rémunérations, avantages, à la résiliation d'un contrat Activité syndicale
		Sécurité du lieu de travail	Responsabilité civile (chute, etc.) Événements liés à la réglementation sur la santé et la sécurité du personnel Rémunération du personnel
		Égalité et discrimination	Tous types de discrimination
Clients, produits et pratiques commerciales	Pertes résultant d'un manquement, non intentionnel ou dû à la négligence, à une obligation professionnelle envers des clients spécifiques (y compris exigences en matière de fiducie et de conformité) ou de la nature ou conception d'un produit	Conformité, diffusion d'informations et devoir fiduciaire	Violation du devoir fiduciaire/de recommandations Conformité/diffusion d'informations (connaissance de la clientèle, etc.) Violation de la confidentialité de la clientèle Atteinte à la vie privée Vente agressive Opérations fictives Utilisation abusive d'informations confidentielles Responsabilité du prêteur

Catégories d'événements (Niveau 1)	Définition	Sous-catégories (Niveau 2)	Exemples (Niveau 3)
		Pratiques commerciales/de place incorrectes	Législation antitrust Pratiques incorrectes Manipulation du marché Délit d'initié (au nom de l'entreprise) Activité sans agrément Blanchiment d'argent
		Défauts de production	Vices de production (absence d'agrément, etc.) Erreurs de modèle
		Sélection, parrainage et exposition	Insuffisance de l'analyse clientèle Dépassement des limites d'exposition d'un client
		Services-conseil	Conflits sur l'efficacité des prestations
Domages aux actifs corporels	Destruction ou dommages résultant d'une catastrophe naturelle ou d'autres sinistres	Catastrophes et autres sinistres	Pertes résultant d'une catastrophe naturelle Pertes humaines dues à des causes externes (terrorisme, vandalisme)
Dysfonctionnements de l'activité et des systèmes	Pertes résultant de dysfonctionnements ou de l'activité ou des systèmes	Systèmes	Matériel Logiciel Télécommunications Interruptions/perturbations d'un service public
Exécution, livraison et gestion des processus	Pertes résultant d'un problème dans le traitement d'une transaction ou dans la gestion des processus ou des relations avec les contreparties commerciales et fournisseurs	Saisie, exécution et suivi des transactions	Problèmes de communication Erreurs dans la saisie, le suivi ou le chargement Non-respect de délais ou d'obligations Erreur de manipulation du modèle/système Erreur comptable/d'affectation d'une entité Autres erreurs d'exécution Problèmes de livraison Fautes dans la gestion des sûretés Mauvais suivi des données de référence
		Surveillance et notification financière	Manquement à l'obligation de notification Inexactitudes dans les rapports externes (pertes)
		Admission et documentation clientèle	Absence d'autorisations/renoncations clientèle Documents juridiques absents/incomplets
		Gestion des comptes clients	Accès non autorisé aux comptes Données clients incorrectes (pertes) Actifs clients perdus ou endommagés par négligence
		Contreparties commerciales	Faute d'une contrepartie hors clientèle Divers conflits avec une contrepartie hors clientèle
		Fournisseurs	Sous-traitance Conflits avec les fournisseurs

**14.2 ANNEXE B : RÉFÉRENTIEL DES PROCESSUS DE GENERALI**

Liste des processus de niveau 2 :

<b>EX 1 - Définir et décliner la stratégie</b>
<b>EX 2 - Lever et allouer le capital</b>
<b>EX 3 - Gérer les risques</b>
<b>EX 4 - Mettre en place des règles de gouvernance</b>
<b>EX 5 - Gouverner l'entreprise</b>
<b>EX 6 - Piloter l'activité</b>
<b>EX 7 - Gérer la communication</b>
<b>MT 1 - Concevoir et suivre les produits</b>
<b>MT 10 - Assurer le contrôle de gestion et le pilotage opérationnel</b>
<b>MT 11 - Gérer le juridique opérationnel</b>
<b>MT 2 - Gérer et suivre la relation client</b>
<b>MT 3 - Distribuer les produits</b>
<b>MT 4 - Administrer les contrats</b>
<b>MT 5 - Gérer les réseaux (hors réseau salarié)</b>
<b>MT 6 - Gérer les prestations</b>
<b>MT 7 - Gérer les cotisations</b>
<b>MT 8 - Gérer la réassurance</b>
<b>MT 9 - Gérer les placements</b>
<b>SU 1 - Gérer les Ressources Humaines</b>
<b>SU 2 - Gérer la réglementation juridique et fiscale</b>
<b>SU 3 - Gérer les Systèmes d'Information</b>
<b>SU 4 - Gérer les achats</b>
<b>SU 5 - Gérer les moyens généraux</b>
<b>SU 6 - Assurer la sécurité des biens et des personnes</b>
<b>SU 7 - Gérer la comptabilité et les finances</b>

**14.3 ANNEXE C : RÉFÉRENTIEL DES RISQUES DE GENERALI**

Catégorie de risque (niveau 1)	Définition (niveau 1)	Sous catégorie de risque (niveau 2)	Définition (niveau 2)
Fraude interne	Pertes dues à un acte intentionnel de fraude, de détournement de biens, d'enfreintes à la législation ou aux règles de l'entreprise qui implique au moins une personne en interne	Vol et fraude (interne)	Actes frauduleux qui impliquent au moins une partie interne (salarié ou intermédiaire lié) résultant d'un détournement d'actifs et / ou de documentation, ayant pour conséquence des pertes pour la Société, voire pour les clients.
		Activité non autorisée d'assurance	Actes frauduleux liés au processus d'assurance (souscription, gestion de portefeuille, prestations), qui impliquent au moins une partie interne (salarié ou intermédiaire lié), résultant d'activités non autorisées ou de manipulations de données internes, impliquant des pertes pour la compagnie, voire pour les clients.
		Autre activité non autorisée hors assurance	Activités frauduleuses qui ne sont pas liées au processus assurance qui impliquent au moins une partie interne (salarié ou intermédiaire lié), résultant d'activités non autorisées ou de manipulation de données internes, impliquant des pertes pour la compagnie, voire pour les clients.
		Sécurité des systèmes (fraude interne)	Actes frauduleux qui impliquent au moins une partie interne (salarié ou intermédiaire lié) résultant d'intrusion du système et de non disponibilité / vol / détérioration des données ayant pour conséquence des pertes pour la Société, voire pour les clients.
Fraude externe	Pertes dues à un acte intentionnel de fraude, de détournement de biens, d'enfreintes à la législation ou aux règles par une tierce partie	Vol et fraude (externe)	Actes frauduleux qui impliquent seulement un tiers dans un détournement d'actifs et / ou de la documentation avec pour conséquence des pertes pour la Société, voire pour les clients.
		Activité non autorisée d'assurance	Actes frauduleux en rapport avec le processus d'assurance (souscription, gestion de portefeuille, de traitement des créances), qui impliquent seulement un tiers, résultant d'activités non autorisées ou internes de manipulation de données impliquant des pertes pour la Société, voire pour les clients.
		Autre activité non autorisée hors assurance	Actes frauduleux non liés au processus assurance qui impliquent seulement un tiers, résultant d'activités non autorisées ou internes de manipulation de données, impliquant des pertes pour la Société, voire pour les clients.
		Sécurité des systèmes (fraude externe)	Actes frauduleux, qui impliquent seulement un tiers, résultant d'intrusion du système et non disponibilité / vol / détérioration des données avec pour conséquence des pertes pour la Société, voire pour les clients.
Pratiques en matière d'emploi et sécurité sur le lieu de travail	Pertes résultant d'actes incompatibles au regard de la loi en matière d'emploi, de législation relative à la santé ou à la sécurité, du paiement d'indemnités ou de discrimination sociale	Organisation et gestion du personnel	Pertes imprévues découlant d'une structure organisationnelle inadaptée qui ne fournit pas une définition claire et objective des rapports hiérarchiques et des responsabilités, y compris une séparation appropriée des tâches, en prenant en considération la taille et la nature des activités de l'entreprise, des stratégies, des objectifs et des besoins.
		Rémunération des salariés	Pertes imprévues découlant de la démotivation ou l'insatisfaction liées au niveau de rémunération ou à l'évaluation des performances.
		Formation et compétences	Pertes imprévues découlant de recrutements inadaptés ou d'un niveau de formation inadapté
		Turnover	Pertes inattendues résultant d'un niveau élevé de départs
		Sécurité du lieu de travail	Pertes imprévues résultant de la santé ou du paiement de préjudices corporels subis sur le lieu de travail.
		Relations sociales	Pertes d'efficacité découlant de la détérioration de l'environnement de travail et les relations avec les employés (ex. grève,...)
Clients / Tiers et produits	Pertes résultant d'un acte non intentionnel ou d'une négligence dans l'exercice d'une obligation professionnelle face au client (incluant les exigences en matière fiduciaire et de conformité)	Défauts produits	Pertes de business imprévues découlant de l'absence de produits compétitifs, ou de défauts de l'offre produits (cannibalisation, etc.)
		Adéquation des produits et services client	Pertes de business imprévues, en termes d'insatisfaction du client et / ou érosion du portefeuille, résultant d'une inadéquation de la qualité du service (y/c agents) ou de l'adaptation des produits de la compagnie à la clientèle
		Sélection et exposition du client	Pertes de business imprévues découlant d'une sélection et d'une gestion inappropriée des contreparties (clients) avec une attention particulière à leur solvabilité

Dommages aux actifs corporels	Pertes résultant de la perte ou du dommage sur un actif corporel à la suite d'une catastrophe naturelle ou d'un autre sinistre	Environnement et catastrophes naturelles	Pertes imprévues ou dommages aux actifs matériels et / ou aux ressources humaines de la compagnie découlant de catastrophes naturelles (inondations, tremblements de terre, incendies, pandémie, etc.)
		Catastrophes du fait de l'Homme	Pertes ou dommages aux biens matériels et / ou aux ressources humaines de la compagnie découlant de l'action humaine (guerre, insurrection, terrorisme, vandalisme)
Dysfonctionnements de l'activité et des systèmes	Pertes résultant d'interruptions de l'activité ou de dysfonctionnement des systèmes	Sélection des SI	Pertes imprévues résultant d'une inadéquation des investissements entraînant une baisse d'efficacité du système
		Développement des SI	Pertes imprévues résultant de défaillances du système en raison d'erreurs et de déficiences de projets ou de l'obsolescence technologique entraînant une baisse d'efficacité du système
		Maintenance des SI	Pertes imprévues résultant d'une mise à niveau tardive ou inadaptée du système (en rapport avec l'évolution des besoins de l'entreprise), provoquant une baisse de l'efficacité du système
		Disponibilité des SI et des données	Pertes imprévues résultant de l'interruption d'activité due à une indisponibilité du système informatique causée par des facteurs externes (coupure de courant, perturbations des télécommunications, etc.)
		Intégrité des données	Pertes imprévues résultant de la perte / indisponibilité des données dans le cas d'une indisponibilité des systèmes, issue de dommages à des bases de données ou altération données causées par des erreurs accidentelles par les utilisateurs
Exécution et gestion des processus	Pertes résultant d'un problème dans le traitement d'une transaction ou dans la gestion des processus ou pertes subies avec les contreparties commerciales et les fournisseurs	Complexité et formalisation des processus	Pertes imprévues résultant d'une formalisation / configuration inadéquate des processus / procédures, ou de leur complexité excessive ou d'un degré d'automatisation insuffisant provoquant une baisse de l'efficacité
		Saisie, exécution et suivi des transactions	Pertes imprévues résultant des erreurs, retards dans le traitement des transactions (entrée des données incorrectes, activités retardées, transactions incorrectes, etc.) ou du non respect des procédures provoquant une baisse de l'efficacité
		Surveillance et rapports	Pertes imprévues résultant de reportings internes basés sur une information incomplète ou inadaptée
		Documentation	Pertes imprévues résultant d'erreurs et d'omissions dans la collecte et la conservation de documents, ayant des conséquences sur la validité des contrats
		Intermédiaires et exposition	Pertes imprévues résultant d'une sélection et d'un suivi inadapté des intermédiaires, avec une attention particulière à leur solvabilité, leur comportement et leur capacité, qui peuvent avoir un effet adverse sur la fidélisation du client et la part de marché.
		Vendeurs et fournisseurs	Pertes inattendues résultant la mauvaise exécution des prestations fournies par les vendeurs / fournisseurs de biens et services
		Externalisation	Pertes imprévues résultant de services externalisés inadaptés par rapport aux accords fixés avec le prestataire.
		Modèle	Pertes imprévues issues du risque q'un modèle quantitatif (utilisé pour des besoins financiers ou d'assurance) ne donne pas une description adaptée du sous jacent, ou soit basé sur des paramètres non réalistes / réels.

Réglementation et conformité	Pertes imprévues (légal	Litiges	Pertes imprévues dues à des actions en justice
		Juridique	Pertes imprévues liées à une contractualisation insuffisante
		Règles en matière d'assurance (absence de conformité)	
		Règles actuarielles	
		Marché financier et règles pratiques	Non respect de la réglementation AMF
		Règles comptables	
		Règles en terme de reporting externes	
		Loi italienne 262 (FARG)	
		Réglementation du personnel	
		Règles de gouvernance	
		Loi italienne 230	
		Autres règles	

**14.4 ANNEXE D : SCÉNARIOS DE PLACE DU SECTEUR BANCAIRE**

Défaillance d'un dépositaire externe
Grève généralisée d'un ou des centres techniques ou informatiques
Erreur de saisie d'une instruction VGM par un collaborateur
Non recouvrement REMDOC/CREDOC
Erreur sur le paramétrage des taux conduisant à souscrire des prêts à taux usurier de façon massive
Détournement de fonds VGM par fraude externe
Détournement de fonds VGM par fraude interne
Falsification externe d'un ordre de virement papier au débit d'un client entreprise
Falsification d'un chèque-effet import ou export
Vol de vignettes de chèques dans un centre de stockage et rejeu (représentation) de ces chèques
Monétique - Compromission de clé d'une autorité de certification (fraude interne chez ACCB, VISA, MCI)
Monétique - Compromission externe sur pistes ISO2 (fraude interne chez un commerçant ou un intermédiaire du réseau acquéreur de l'établissement)
Monétique - Compromission interne de clé (fraude interne d'un ou plusieurs officiers de sécurité monétique)
Monétique - Compromission interne sur pistes ISO2 (fraude interne d'une personne ayant accès à des volumes importants de données porteur)
Monétique - Fraude courante externe (vol de cartes ou de pistes)
Monétique - Fraude informatique retrait - fraude interne d'un personnel ou fraude externe d'un prestataire le conduisant à pouvoir vider facilement des GAB
Monétique - Fraude informatique paiement - fraude interne d'un personnel ou fraude externe d'un prestataire le conduisant à court-circuiter l'autorisation
Destruction d'une agence, d'un siège ou d'un bâtiment technique par incendie ou autre cause
Cyber-attaque rendant indisponible les sites internet du groupe
Destruction de gros systèmes informatiques dans des centres de production informatique bancaire
Perte d'intégrité de base de données de production bancaire
Indisponibilité de la plate-forme de paiement / règlement de la banque
Divulgence d'informations confidentielles
Phishing ou autre usurpation d'identité lors d'un accès Banque à Distance
Assignations (actions individuelles, actions collectives) pour différents motifs dont défaut de conseil, soutien ou rupture abusifs.
Place - Crue centennale de la Seine ou autre fleuve
Place - Grève générale des transports
Place - Multiples attentats (Explosif/ Nucléaire, Radiologique, Biologique et chimique)
Place - Pandémie grippale ou autre
Place - Panne générale EDF rendant indisponible plus de 48h plusieurs sites névralgiques
Blanchiment / financement du terrorisme, non respect de la réglementation TRACFIN
Déclaration fiscale ou sociale erronée pour un établissement bancaire
Déclaration fiscale ou sociale établie par la banque pour ses clients erronée (cas de l'IFU)
Délit d'initié
Non respect de la réglementation AMF (Autorité des marchés financiers)
Non respect de la réglementation fiscale, sociale, CNIL, etc.
Départ conflictuel de salariés ou d'équipes de salariés
Départ de salarié clé ou d'équipes de salariés clés à la concurrence
Grève interne
Rogue trading (trader véreux ou courtier ripoux),