

**Mémoire présenté devant le Centre d'Etudes Actuarielles  
pour l'obtention du diplôme  
du Centre d'Etudes Actuarielles  
et l'admission à l'Institut des Actuaire**  
le : \_\_\_\_\_

Par : Sabrina SAVARRE et Benoît PAYRE

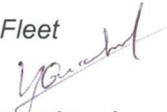
Titre : Charge ultime nette de réassurance en RC corporelle :  
2 modèles stochastiques pour les flottes automobiles.

Confidentialité :  NON  OUI (Durée :  1an  2 ans)  
Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

*Membre présent du jury de  
l'Institut des Actuaire :*

signature

*Entreprise :*

Nom : Covéa Fleet  
Signature : 

*Membres présents du jury du  
Centre d'Etudes Actuarielles :*

*Directeur de mémoire en entreprise :*

Thomas BEHAR

Nom : Gilles MONGTS  
Signature : 

Vincent DAMAS

*Invité :*

Gérard CROSET

Arnaud COHEN

Nom : \_\_\_\_\_  
Signature : \_\_\_\_\_

Jean-Pierre DIAZ

Brigitte DUBUS

Paul ESMEIN

Michel FROMENTEAU

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

Stéphane MENART

Christophe IZART

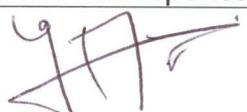
Pierre PETAUTON

Florence PICARD

Olivier LOPEZ

*Secrétariat :*

Signature du responsable entreprise



*Bibliothèque :*

Signature du candidat



## **Présentation et résumé**

L'objectif de ce mémoire est d'estimer la charge ultime nette de réassurance des sinistres de RC corporelle pour les flottes automobiles.

Pour l'atteindre, nous avons sélectionné deux méthodes. La première est connue et éprouvée : c'est la méthode probabiliste via un modèle paramétrique. Pour appréhender la sinistralité des graves avec plus de finesse, nous l'enrichissons par une loi mélange.

La seconde est plus récente (2005) et nous a séduit parce qu'elle tient compte de la diversité de l'évolution individuelle d'un sinistre : c'est la méthode des matrices de transition décrite par Mahon.

Avant de décliner la conception des modèles, nous relevons le besoin de disposer de deux résultats importants : définir les seuils qui séparent des typologies homogènes de sinistres (attritionnel, graves ordinaire, non ordinaire) et élargir l'assiette des sinistres pour pouvoir appliquer Mahon avec robustesse, le périmètre flotte étant restreint. L'utilisation de concepts issus principalement de la théorie des valeurs extrêmes nous permet de déterminer les seuils ; une double démonstration qualitative (analyse descriptive des victimes) et quantitative (test de conformité) nous autorise à enrichir notre base par d'autres sinistres homogènes avec les nôtres et donc à constituer nos matrices.

Les phénomènes décrits par Mahon n'étant pas observés sur nos sinistres, nous construisons une procédure innovante qui conjugue les atouts de la méthode probabiliste (modélisation fine des graves) avec ceux de la méthode des matrices (utilisation de toute l'information disponible). Nous investissons sur un coefficient pondérateur entre les sinistres graves et les très graves que nous optimisons par la distance de Cramér – von Mises.

Toutes les simulations effectuées pour obtenir les applications numériques sont de la forme Monte-Carlo qui permet des tirages aléatoires individuels et donc une application des traités de réassurance, même complexes.

L'objectif est atteint dans les deux cas ; nous argumentons sur le choix d'un modèle pour en sélectionner un.

Mots clés :

Seuil, théorie des valeurs extrêmes, modèle individuel, matrice de transition, loi mélange, tests EDF, provisionnement, base victimes, charge nette de réassurance.

## **Abstract**

The object of this thesis is to assess the ultimate loss burden of bodily injury liability claims after application of reinsurance treaties in motor fleet insurance.

To achieve this, we selected two methods. The first is known, has been tried and tested : it is the probabilistic method using a parametric model. To apprehend severe claims with a better accuracy, we intend to improve it with a mixture model.

The second is more recent (2005) and attracted us because it takes into consideration the diversity of the claim's individual evolution : it is the transition matrix approach described by Mahon.

Before starting to build models, we need to obtain two important results : define the thresholds that separate homogeneous types of claims (attritional, common severe, non common severe) and expand the claims database in order to apply the Mahon method with robustness because the number of fleet claims is limited. The use of concepts mainly derived from the extreme value theory allows us to determine the thresholds ; a twofold demonstration both qualitative (victims descriptive analysis) and quantitative (conformity testing) allows us to expand our base with other claims which are consistent with ours. That enables us to build our matrix.

As the phenomena described by Mahon are not observed on our claims, we create an innovative procedure

that combines the advantages of the probabilistic method (accurate modeling of severe claims) with those of the matrix method (use of all available information). Then we handle a weighting coefficient linking severe claims to very severe ones which we optimize with the Cramér-von Mises distance.

Each simulation we used to obtain numerical applications are based on the Monte Carlo method which allows individual random selections and then an application of reinsurance (complex) treaties.

The objective is achieved in both cases ; we argue on the choice of a model with a view to select one.

Key words :

threshold, extreme value theory, transition matrix, mixture law, EDF test, individual modeling, reserving, casualties database.

## **Avertissement**

Pour des raisons de confidentialité, les valeurs numériques présentes dans ce mémoire ont été modifiées. Les ordres de grandeur sont néanmoins conservés.

## **Remerciements – Auteurs**

Nous adressons nos sincères remerciements à Gilles Mongis pour son encadrement tout au long de ce mémoire. Ses précieux conseils ont été déterminants pour nous aider à avancer. Son intérêt pour notre sujet et les solutions proposées ont été très stimulants pour approfondir nos recherches.

Je remercie tous les membres de la Direction Technique, anciens et nouveaux et plus particulièrement Julien Médard pour leur aide.

J'exprime également ma gratitude à ma famille pour leur patience, leur précieux soutien et leurs encouragements. Je remercie mes parents qui m'ont stimulée et encouragée pendant mes études. Je remercie ma sœur qui a toujours su s'intéresser à mes avancées. Enfin mes pensées vont à mon mari pour son soutien sans faille.

A mes enfants qui ont parfois dû trouver un autre compagnon de jeu mais qui continuent d'aimer les mathématiques,

A mon épouse, pour sa patience et sa confiance et sans qui cette formation n'aurait pas été possible,

A ma famille, mes amis et mes collègues qui m'ont souvent manifesté leur intérêt et leurs encouragements tout au long de ce parcours,

SS

BP

## SOMMAIRE

<b><u>I- INTRODUCTION</u></b>	<b><u>8</u></b>
<b>I - A PRESENTATION DE COVEA FLEET</b>	<b>8</b>
I - A . 1. Le groupe Covéa	8
I - A . 2. Le modele Covéa Fleet	8
I - A . 3. L'assureur Covea Fleet	9
I - A . 4. Les particularités de Covéa Fleet	10
<b>I - B OBJECTIF DU MEMOIRE</b>	<b>11</b>
I - B . 1. Contexte de l'étude	11
I - B . 2. Le seuil des sinistres graves utilisé	12
I - B . 3. La détermination actuelle des ultimes	13
I - B . 4. Systèmes d'information et périmètre des données	13
I - B . 5. L'évolution des normes de solvabilité	14
I - B . 6. L'objectif du mémoire	14
<b><u>II - LES ELEMENTS DE CONTEXTE</u></b>	<b><u>15</u></b>
<b>II - A CONTEXTE DU RISQUE ROUTIER</b>	<b>15</b>
<b>II - B L'ENVIRONNEMENT JURIDIQUE</b>	<b>17</b>
II - B . 1. La responsabilité civile	17
II - B . 2. L'assurance de la responsabilité civile automobile	18
II - B . 3. Les 7 étapes récentes de l'indemnisation du corporel	18
II - B . 4. Prospective sur les travaux en cours	25
II - B . 5. Conclusion	26
<b>II - C LA GESTION DU SINISTRE CORPOREL</b>	<b>27</b>
II - C . 1. Les étapes clés de la gestion d'un sinistre corporel	27
II - C . 2. Les conventions inter-compagnies en corporel	28
II - C . 3. La gestion des sinistres corporels dans le groupe Covéa	29
<b><u>III - ETAPE 1 : L'ANALYSE DE BASES VICTIMES</u></b>	<b><u>32</u></b>
<b>III - A DEMARCHE DE L'ETUDE</b>	<b>32</b>
<b>III - B APPLICATIONS</b>	<b>35</b>
<b>III - C CONCLUSION</b>	<b>40</b>
<b><u>IV - ETAPE 2 : CALCUL D'UN SEUIL SEPARANT LES ATTRITIONNELS DES GRAVES</u></b>	<b><u>41</u></b>
<b>IV - A DE LA THEORIE DES VALEURS EXTREMES A LA METHODE DE DEPASSEMENT DE SEUIL</b>	<b>41</b>
IV - A . 1. Distribution des extrema dans le cas d'un échantillon de taille finie	41
IV - A . 2. Distribution asymptotique du maximum	42
IV - A . 3. Loi des excès et distribution GPD (Generalized Pareto Distribution)	45
IV - A . 4. Estimation de l'indice de queue	46
<b>IV - B METHODES DE DETERMINATION DU SEUIL</b>	<b>48</b>
IV - B . 1. QQ-plot : Quantile-quantile plot	48
IV - B . 2. Mean excess function (fme) - Pente de l'espérance résiduelle	49
IV - B . 3. L'estimateur de Hill	51
IV - B . 4. L'estimateur de Pickands	51
IV - B . 5. L'estimateur DEdh	52
IV - B . 6. L'estimateur du maximum de vraisemblance (EMV)	52
IV - B . 7. Minamse	53
IV - B . 8. Gertensgarbe plot	54

<b>IV - C TRANSFORMATION « AS IF » DES DONNEES</b>	<b>56</b>
IV - C . 1. Quelques notations	56
IV - C . 2. Traitement de l'inflation	56
IV - C . 3. Projection à l'ultime de la charge des sinistres	58
<b>IV - D APPLICATIONS</b>	<b>60</b>
IV - D . 1. Les données	60
IV - D . 2. Mise « as if »	61
IV - D . 3. Détermination du seuil	62
<b>IV - E CONCLUSION</b>	<b>73</b>
<b><u>V - METHODE 1 : MODELE PROBABILISTE DE LA CHARGE RC CORPORELLE</u></b>	<b><u>74</u></b>
<b>V - A MODELISATION DU NOMBRE DE SINISTRES</b>	<b>75</b>
<b>V - B MODELISATION DE LA SEVERITE</b>	<b>76</b>
V - B . 1. Les principales lois	76
V - B . 2. Modélisation de la sévérité des sinistres graves	78
<b>V - C LES TESTS D'ADEQUATION ET DE CONFORMITE</b>	<b>82</b>
<b>V - D APPLICATIONS</b>	<b>84</b>
V - D . 1. Résultats de la modélisation des nombres	84
V - D . 2. Résultats de la modélisation des montants	86
V - D . 3. Résultats des simulations de la charge	92
<b>V - E CONCLUSION</b>	<b>99</b>
<b><u>VI - METHODE 2 : MODELISATION « MARKOVIENNE » DE LA CHARGE RC CORPORELLE</u></b>	<b><u>100</u></b>
<b>VI - A METHODE</b>	<b>100</b>
<b>VI - B APPLICATIONS : APPROCHE MAHON (LOI UNIQUE)</b>	<b>105</b>
VI - B . 1. Mise en place de la méthode : données et matrices	105
VI - B . 2. Vérification du contexte log-normalien	109
VI - B . 3. Estimation des paramètres des lois uniques $\ln(\mu_{i,j}, \sigma_{i,j})$	111
VI - B . 4. Recherche de liens	112
<b>VI - C APPLICATIONS : NOUVELLE APPROCHE (LOI MELANGE)</b>	<b>115</b>
VI - C . 1. Estimation des paramètres avec une méthode de minimisation de distance	115
VI - C . 2. Estimation du paramètre $\lambda$	117
VI - C . 3. Simulation de la charge brute et nette de réassurance	120
<b>VI - D CONCLUSION</b>	<b>124</b>
<b><u>VII - CONCLUSION</u></b>	<b><u>125</u></b>
VII - A . 1. Comparaison	125
VII - A . 2. Conclusion	127
<b><u>VIII - ANNEXES</u></b>	<b><u>129</u></b>
<b>VIII - A ANNEXE - ETUDE COMPARATIVE DU PROJET LEFRAND ET DU LIVRE BLANC DE L'AFA</b>	<b>129</b>
<b>VIII - B ANNEXE - ANALYSE DU FICHIER DES VICTIMES INDEMNISEES (FVI)</b>	<b>130</b>
<b>VIII - C ANNEXE - CONSTITUTION DE LA BASE SINISTRES (SAS)</b>	<b>132</b>
<b>VIII - D ANNEXE - OUTIL EXCEL DE DÉTERMINATION DES CADENCES</b>	<b>133</b>
<b>VIII - E ANNEXE - PROGRAMME DE DÉTERMINATION DU SEUIL (R)</b>	<b>134</b>
<b>VIII - F ANNEXE - PROGRAMME DE DÉTERMINATION DU SEUIL - GERSTENGARBE (SAS)</b>	<b>136</b>
<b>VIII - G ANNEXE - PROGRAMME D'AJUSTEMENT DE LOIS (R)</b>	<b>138</b>
<b>VIII - H ANNEXE - RAPPEL SUR LES MOMENTS</b>	<b>139</b>

<b>VIII - I</b>	<b>ANNEXE - CALCUL DE LA BORNE SUPERIEURE DE LA CLASSE 12</b>	<b>140</b>
<b>VIII - J</b>	<b>METHODE 2 - CONSTRUCTION DES MATRICES INTERMEDIAIRES</b>	<b>141</b>
<b>VIII - K</b>	<b>ANNEXE - OUTIL DE SIMULATION : IGLOO</b>	<b>142</b>
<b>IX -</b>	<b><u>BIBLIOGRAPHIE</u></b>	<b>144</b>
<b>IX - A</b>	<b>INDEMNISATION DOMMAGE CORPOREL</b>	<b>144</b>
<b>IX - B</b>	<b>ETUDES ASSUREURS / REASSUREURS</b>	<b>144</b>
<b>IX - C</b>	<b>AUTRES ETUDES</b>	<b>144</b>
<b>IX - D</b>	<b>MEMOIRES D'ACTUARIAT</b>	<b>144</b>
<b>IX - E</b>	<b>THÉORIE ET CONCEPTS</b>	<b>144</b>
<b>IX - F</b>	<b>LOGICIELS</b>	<b>145</b>
<b>IX - G</b>	<b>SITES INTERNET</b>	<b>145</b>

## I - INTRODUCTION

Le sujet abordé, l'évaluation de la charge ultime des sinistres de responsabilité civile corporelle (RCC) pour les flottes automobiles, nous est familier au travers de quelques sinistres majeurs subissant de fortes variations et plus globalement, par les particularités de sa liquidation dans le temps.

La charge la plus volatile que la compagnie d'assurance doit appréhender est celle de la RCC et notamment celle des sinistres graves. Les enjeux sont importants, nous les présenterons dans ce chapitre.

### I - A PRESENTATION DE COVEA FLEET

---

#### I - A . 1. LE GROUPE COVEA

Covéa (« co » et « via ») signifie « faire la route ensemble ».

En 2001, le président Jean-Claude Seys initie la stratégie Covéa visant à réunir des activités communes d'assurance pour des mutuelles partenaires. Les premières à franchir le pas sont MMA, MAAF et DAS.

Cette stratégie s'incarne autour de différents pôles par les créations suivantes en 2001 :

- Covéa A.I.S. (Assistance Indemnisation Services) pour la gestion de tous les sinistres,
- Covéa Risks, société d'assurance généraliste, point d'entrée du groupe Covéa pour le courtage,
- Covéa Fleet, pour l'assurance des risques :
  - o flottes automobiles,
  - o marchandises transportées,
  - o automobiles pour les grands comptes affinitaires tels que les sociétés de financement et les constructeurs automobiles.

En 2003, Covéa trouve le véhicule juridique adéquat pour concrétiser ses ambitions, la SGAM et devient ainsi la première SGAM créée en France et en Europe.

*La SGAM est une structure juridique permettant à des mutuelles d'assurance de constituer un groupe, de l'afficher et d'organiser en son sein des liens de solidarité importants et durables.*

En 2006, GMF et Azur rejoignent la SGAM Covéa, formant ainsi un groupe qui est aujourd'hui le leader français en assurances de biens et responsabilités.

#### I - A . 2. LE MODELE COVÉA FLEET

Pour Covéa Fleet, à sa création, cette stratégie s'est traduite par l'apport des portefeuilles concernés d'origine MMA, MAAF et DAS.

Le statut juridique retenu de cette nouvelle entité est celui d'une société anonyme à directoire et conseil de surveillance et non celui d'un groupement d'intérêt économique (GIE). C'est donc bien une nouvelle compagnie d'assurance qui apparaît sur le marché.

Le modèle séduit ensuite d'autres compagnies comme SMABTP (2002) et l'Auxiliaire (2005).

Pour ces assureurs partenaires, l'attrait de Covéa Fleet réside en différentes motivations :

- externaliser une activité non stratégique en la confiant à des experts,
- mutualiser les coûts (informatique, gestion sinistres, réassurance...) et les risques (tarif plus cohérent par l'atteinte d'une taille supérieure),
- modifier la nature du risque supporté du fait de la mutation du statut d'assureur vers celui d'actionnaire,
- économiser des fonds propres grâce à la mutualisation des provisions.

Par ricochet, toute croissance des actionnaires induit celle du portefeuille de Covéa Fleet qui intègre alors les risques flottes et marchandises transportées de ces compagnies. C'est le cas par exemple quand MMA rachète Winterthur France en 2003 et fusionne avec Azur en 2006.

Covéa Fleet quitte progressivement son état initial de « patchwork » de portefeuilles portés par plusieurs systèmes d'information, en réalisant pour l'activité flottes d'entreprises :

- un nouveau système d'informations spécialisé (2001 - 2003),
- de nouvelles offres d'assurance hébergées dans un extranet permettant aux apporteurs leur souscription et leur gestion (2003 - 2009),
- les migrations des anciens contrats de ces partenaires vers ces nouvelles offres (2001 - 2009).

A chaque apport, la gestion des sinistres est déléguée à Covéa A.I.S., répartie dans ses différentes entités, fonctionnant dans un premier temps avec les systèmes d'informations d'origine puis sur ses propres systèmes. La politique de règlements s'harmonise ensuite sur les nouveaux sinistres. C'est le symétrique, côté sinistres, de ce que Covéa Fleet fait avec les migrations de contrats.

### **I - A . 3.      L'ASSUREUR COVEA FLEET**

#### **Quels types de clients et que veulent-ils ?**

<b>Segment de clientèle adressée</b>	<b>Principales attentes du client concernant l'assureur</b>	<b>Réponse de l'assureur sous forme de garanties</b>
Les entreprises de transport pour propre compte	Assurer la mobilité des préposés dans le cadre de leur activité professionnelle ; que cette mobilité soit permise par un véhicule de l'entreprise ou du préposé. Protéger leur patrimoine : véhicules, matériels transportés ou expédiés. Disposer de solutions de continuité pour la mobilité des préposés en cas de panne et d'accident. Une couverture de leur responsabilité élargie aux cas de travaux avec des véhicules spéciaux	Responsabilité civile (RC) circulation Auto-mission des préposés Dommages aux véhicules (vol, incendie, bris de glaces, dommages tous accidents) Perte financière Facultés tous types de transport Assistance aux personnes et au véhicule  Responsabilité civile fonctionnement
Les transporteurs routiers de marchandises	Idem entreprises et en plus : Couvrir les marchandises de leurs clients Couvrir leur responsabilité dans les missions effectuées pour leurs clients (manutention, dépositaire,...) Protéger leurs locaux (bureau, entrepôt)	Responsabilité contractuelle et déclaration de valeurs au voyage Responsabilité entreprise  Dommages aux biens
Les transporteurs routiers de voyageurs	Idem entreprises et en plus : Couvrir leur responsabilité dans l'organisation d'un voyage Avoir une continuité pour la mobilité de leurs clients en cas de panne et d'accident Protéger leurs locaux (bureau, entrepôt)	Responsabilité entreprise ciblée  Assistance aux voyageurs  Dommages aux biens
Les loueurs (courte et longue durée)	Idem entreprises (remplacer préposé par locataire) et en plus : Couvrir leur responsabilité dans le cadre de la location Protéger leurs locaux (bureau)	Responsabilité entreprise ciblée  Dommages aux biens
Les constructeurs automobiles, les établissements financiers	Limiter leur risque de résultats en couvrant le financement du véhicule Fidéliser leurs clients par un élargissement de la gamme de services proposés	Pertes financières  Garanties mécaniques

### Quels sont les principaux clients ?

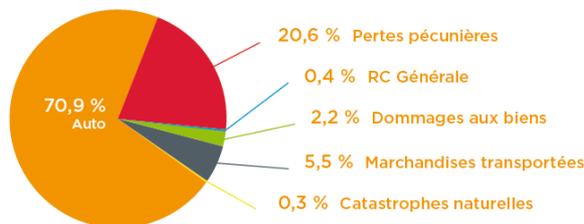
Répartition des cotisations acquises 2010 par type de clients-produits :



Les entreprises de transport pour propre compte représentent la principale clientèle. Avec le transport routier de marchandises, cela représente 1 euro sur 2 de cotisations acquises.

### Quelles sont les garanties souscrites ?

Répartition des cotisations acquises 2010 par type de produits-garanties :



L'auto est l'activité phare en cotisations acquises (70,9%) ; côté provisions, le constat se renforce car 92% d'entre elles sont de l'auto.

### Quelle place tient Covéa Fleet dans le marché des flottes ?

En 2008, le marché auto européen représente 30% du marché non vie européen et 127 milliards d'€ de primes. La France y a une part de marché de 14,1%<sup>1</sup>, soit environ 18 milliards d'€.

A la même époque, l'activité flottes en France est estimée à 1,845 milliard d'€. La part de marché de Covéa Fleet sur l'activité flottes est d'environ 20% ce qui en fait un acteur représentatif.

## **I - A . 4. LES PARTICULARITES DE COVEA FLEET**

On vient de voir que la délégation de **gestion des sinistres** accordée à Covéa A.I.S. induit structurellement une certaine hétérogénéité liée aux spécificités des différents SI et c'est une première particularité. Elle est accrue par une délégation de gestion de sinistres de fréquence accordée à certains réseaux.

Par ailleurs, Covéa Fleet est une entreprise spécialisée surtout sur les risques du **marché des entreprises** (flottes automobiles, marchandises transportées). L'activité affinitaire consistant majoritairement à des ventes aux particuliers via leurs partenaires.

A elle seule, la garantie responsabilité civile (RC) automobile représente 83% des provisions, essentiellement pour le corporel. Le risque principal de Covéa Fleet en terme de provisionnement est donc bien celui de **la RC corporelle automobile**.

On peut également noter que **la spécialisation des activités** de Covéa Fleet lui interdit toute mutualisation de ses résultats avec d'autres risques (RC générale, dommages aux biens, santé, vie, ...) à l'instar des assureurs entreprises et exige une rentabilité de chacune de ses branches. Cette dernière est aussi une

<sup>1</sup> CEA Statistics n°38, The European Motor Insurance Market, February 2010

<sup>2</sup> Le marché de l'assurance automobile en 2008, Analyse des résultats détaillés

garantie de pérennité, argument commercial important pour un assureur spécialiste pouvant être perçu comme plus fragile. Covéa Fleet doit donc obtenir une mutualisation suffisante de bons risques dans son portefeuille. Cela passe par une tarification segmentée, par l'atteinte d'une taille critique et par une sélection permanente des risques acquis.

La distribution commerciale des produits de Covéa Fleet se fait dans **un environnement multi-réseaux** qui est très particulier : on distingue les réseaux mandataires (agents généraux MMA, réseaux salariés MAAF, SMABTP et Auxiliaire) et les courtiers.

## **I - B**      **OBJECTIF DU MEMOIRE**

---

### **I - B . 1.**      **CONTEXTE DE L'ETUDE**

On a vu que le risque principal de Covéa Fleet en matière de provisionnement est celui issu de la garantie RC automobile, plus spécialement porté par la charge des sinistres corporels survenus aux tiers au véhicule. Dans cette partie, avant de présenter l'objectif du mémoire, nous allons rappeler l'enjeu de ce sujet, décrire l'existant, le contexte informatique et les impacts de Solvabilité II.

#### **L'enjeu de la maîtrise de la charge des sinistres corporels**

Pour mesurer cet enjeu, nous regardons les ordres de grandeur de la charge sinistre de Covéa Fleet par garantie pour l'activité flottes.

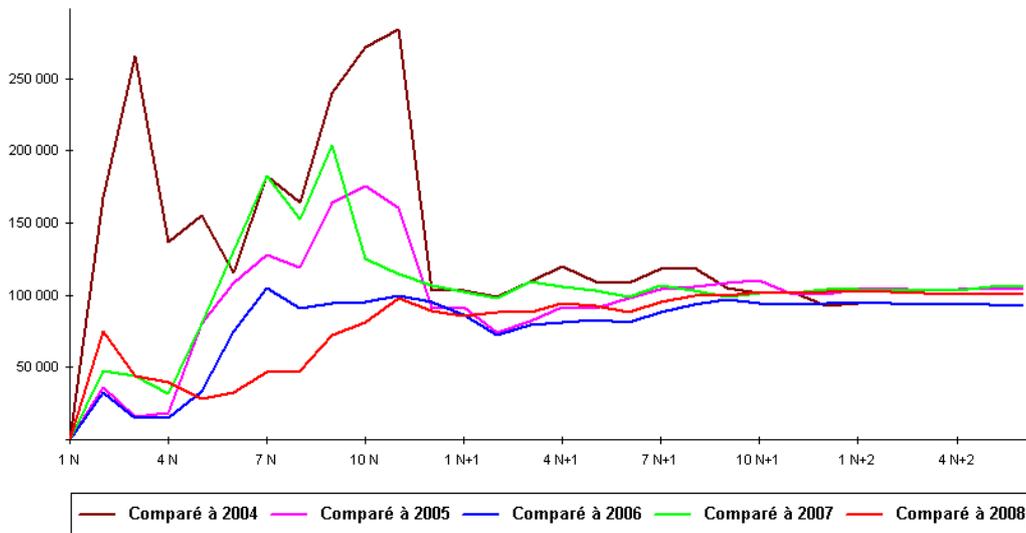
Nous choisissons de prendre en compte la charge brute de conservations. Les conservations étant la part de sinistralité certaine, prise en charge directement par l'assuré au travers d'une franchise annuelle pour certains contrats importants. De plus, nous utilisons le seuil actuel de 75 k€ pour séparer la charge RC corporelle hors grave et grave.

	Poids vs la charge totale ultime			
	2007	2008	2009	2010
Charge RC corp. et mat.	60%	58%	59%	54%
Charge RC corporelle	32%	30%	33%	28%
<i>dont hors graves</i>	8%	8%	7%	6%
<i>dont graves</i>	24%	22%	26%	22%

L'analyse du tableau ci-dessus montre que la charge ultime des graves représente près d'1/4 de la charge brute totale. Au seuil actuel, la charge RC corporelle des graves est donc importante, justifiant des travaux pour mieux la circonscrire.

De plus, cette charge des graves est difficile à appréhender dans le temps. La liquidation mensuelle est différente d'une année sur l'autre ; le graphique ci-dessous représente les estimations de charge 2009 à partir des cadences mensuelles des autres années. La convergence s'observe après un an d'informations :

Estimations de la charge ultime de RC corporelle de 2009 à partir des cadences d'exercices antérieurs



## I - B . 2. LE SEUIL DES SINISTRES GRAVES UTILISE

Actuellement, le seuil pour segmenter les sinistres graves (pesant 3/4 de la charge RC corporelle) et hors graves (1/4) est 75 k€ dans le cadre du pilotage opérationnel et du provisionnement en vision économique.

Le niveau de ce seuil est historique et n'est pas indexé ; à distribution identique, la charge des sinistres graves augmente chaque année du fait de l'inflation des dépenses de santé, de l'évolution de la réglementation et de l'inflation jurisprudentielle (influence des cabinets d'avocats spécialisés).

D'autres seuils sont utilisés chez Covéa Fleet ; pour le tarif des grandes flottes (30 k€), il sert à définir un taux de sur-crête mutualisée et pour la réassurance (2 500k€), il correspond à la priorité.

Dans le groupe Covéa, on constate que les seuils utilisés pour définir le provisionnement comptable en auto sont supérieurs à ceux retenus chez Covéa Fleet : 150 k€ chez 2 mutuelles et 300 k€ chez la dernière.

Dans ses études sur la sinistralité automobile, la FFSA qualifie de graves les sinistres actualisés supérieurs à 750 k€ ou supérieurs à 50% d'AIPP<sup>3</sup>. Le centre européen d'assurance retient 100 k€ mais il y a peut-être un biais lié aux formats de données des 21 pays interrogés.

Enfin, on a vu ici pour l'instant des seuils liés uniquement aux montants de sinistres. Les graves peuvent aussi être définis par rapport à un seuil objectif constaté médicalement (taux d'AIPP) et/ou la typologie du traumatisme subi (paraplégique, traumatisé crânien grave, tétraplégique,...) et/ou la nécessité d'une aide humaine aux côtés de la victime.

Chez Covéa AIS, l'organisation mise en place pour le règlement des sinistres corporels considère les seuils suivants :

- corporels légers (moins de 10% d'AIPP),
- corporels médians (plus de 10% d'AIPP et jusqu'à 50% d'AIPP sans besoin en aide humaine),
- corporels graves (plus de 50% d'AIPP et/ou besoin en aide humaine).

La notion de seuil est complexe et dépend du but recherché : provisionnement, tarification, organisation, réassurance,... Pour notre étude, il nous paraît nécessaire de redéfinir une valeur de seuil robuste.

<sup>3</sup> L'AIPP (Atteinte à l'intégrité physique et psychique) se définit comme « la réduction définitive du potentiel physique, psycho-sensoriel ou intellectuel résultant d'une atteinte à l'intégrité anatomo-physiologique médicalement constatable. S'y ajoutent les phénomènes douloureux et les répercussions psychologiques liés à l'atteinte séquellaire décrite ainsi que les conséquences dans la vie de tous les jours.»

### I - B . 3. LA DETERMINATION ACTUELLE DES ULTIMES

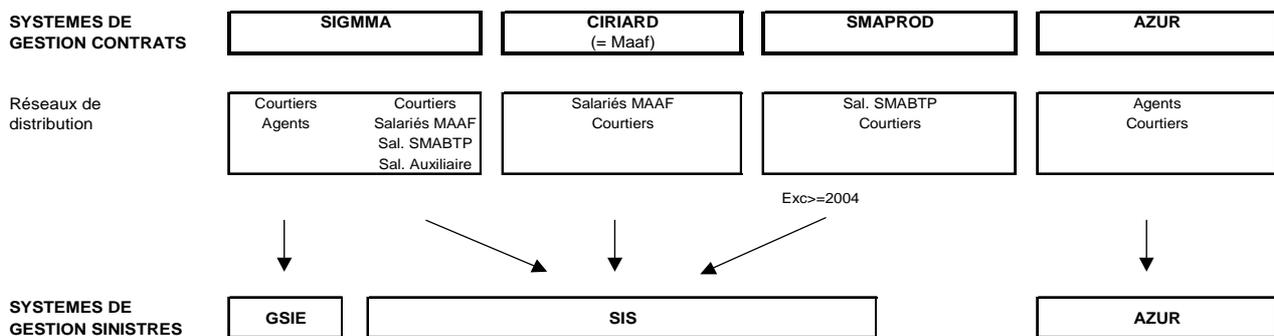
Pour piloter la rentabilité de Covéa Fleet, la direction technique s'est dotée d'un outil déterministe sur la base de cadences mensuelles utilisant les projections Chain-Ladder afin d'obtenir une vision ultime de la charge par branche, par marché, par garantie et par tranche de coût.

L'étude de l'historique montre que la méthode déterministe fonctionne bien pour estimer la charge ultime des sinistres RC corporelle hors graves. En effet, on constate de faibles écarts de prévisions entre l'ultime déterminé à fin N et le constaté vieillissant.

Par contre, on a vu que des cadences n'avaient pas de sens sur les graves au seuil actuel (cf I-B.1). En risque automobile, il est vérifié que la charge ultime RC corporelle des graves est la plus difficile à appréhender du fait de sa forte volatilité. En l'absence de modèle statistiquement fiable sur cette charge de graves, la vision ultime est définie aujourd'hui par une cadence complétée d'une analyse graphique comparant le déroulé de la charge de l'exercice N avec celui des liquidations antérieures. C'est un ultime de charge des graves en brut de réassurance qui est ainsi défini.

### I - B . 4. SYSTEMES D'INFORMATION ET PERIMETRE DES DONNEES

Une cartographie des systèmes d'informations en montre la complexité :



La multitude des systèmes impacte aussi les données disponibles.

Avant d'être convergés dans un système d'informations unique, les contrats flottes comportant des garanties RC automobile étaient hébergés dans quatre systèmes différents ; de même, les sinistres étaient gérés dans quatre systèmes différents.

Une demande complémentaire de données a donc été nécessaire pour travailler sur un périmètre homogène en historique et en structure de données.

Pour exploiter les paramètres techniques du contrat, un travail préalable sur les données dans les quatre systèmes de gestion des contrats a été nécessaire.

Les fichiers de l'inventaire permanent formeront le socle des données utilisées.

Pour enrichir les données analysées et pour disposer d'un modèle le plus complet possible, d'autres bases de données ont été identifiées et seront utilisées dans ce mémoire. Une utilisation de ces bases sera faite avec une recherche de critères permettant d'affiner la modélisation.

Un nouveau projet d'une base victimes unique Covéa se développe ; cette base des victimes corporelles importantes est hébergée dans un logiciel sécurisé et reprend l'historique des sinistres survenus depuis 2007 mais les fonctions de pilotage seront livrées en 2012.

## **I - B . 5.      L'EVOLUTION DES NORMES DE SOLVABILITE**

Solvabilité II va introduire un changement dans la détermination des provisions techniques. En effet, la nécessité de déterminer une "valeur de marché" de ces provisions, alors qu'il n'existe pas de marché des passifs pour pouvoir en obtenir une valeur en lecture directe, va requérir des changements importants dans la démarche de provisionnement avec un impact tant quantitatif que qualitatif.

Dans Solvabilité II, les provisions techniques sont la somme d'un best estimate (noté BE) et d'une marge de risque (cost of capital). La prudence implicite dans les provisions est extériorisée dans les fonds propres.

Le BE est déterminé comme la valeur actuelle probabilisée des cash-flows futurs. Le BE sinistres devrait être calculé en dissociant les éléments bruts de réassurance, les effets de l'inflation, les recours, la provision pour frais de gestion des sinistres et les impacts de la réassurance.

Il faut donc calculer, dans un premier temps, un BE au brut de réassurance, puis passer au net de réassurance. La marge de risque (cost of capital) est théoriquement le montant qu'un repreneur éventuel du passif d'assurance exigerait, au-delà du BE.

Ce contexte renforce l'intérêt de distinguer et d'estimer la charge en brut et net de réassurance.

## **I - B . 6.      L'OBJECTIF DU MEMOIRE**

La charge la plus volatile que la compagnie doit appréhender est celle de la RC corporelle et plus particulièrement celle des sinistres graves. La méthode actuelle peut être améliorée. Les enjeux sont importants : compétitivité tarifaire maîtrisée, rentabilité renforcée, niveau des provisions et des fonds propres à immobiliser plus juste, protection de réassurance affinée, budget de réassurance optimisé. Cela passe par le choix et le calibrage d'un modèle qui calcule la charge ultime brute et nette de réassurance.

Après s'être imprégné du contexte comme la réglementation et la gestion des sinistres corporels, nous étudierons des bases victimes pour enrichir notre analyse. Nous définirons ensuite le sinistre grave selon un nouveau seuil plus homogène et nous modéliserons la charge totale ultime brute et nette de réassurance à l'aide de deux méthodes différentes. Enfin nous choisirons le modèle le plus adapté.

## II - LES ELEMENTS DE CONTEXTE

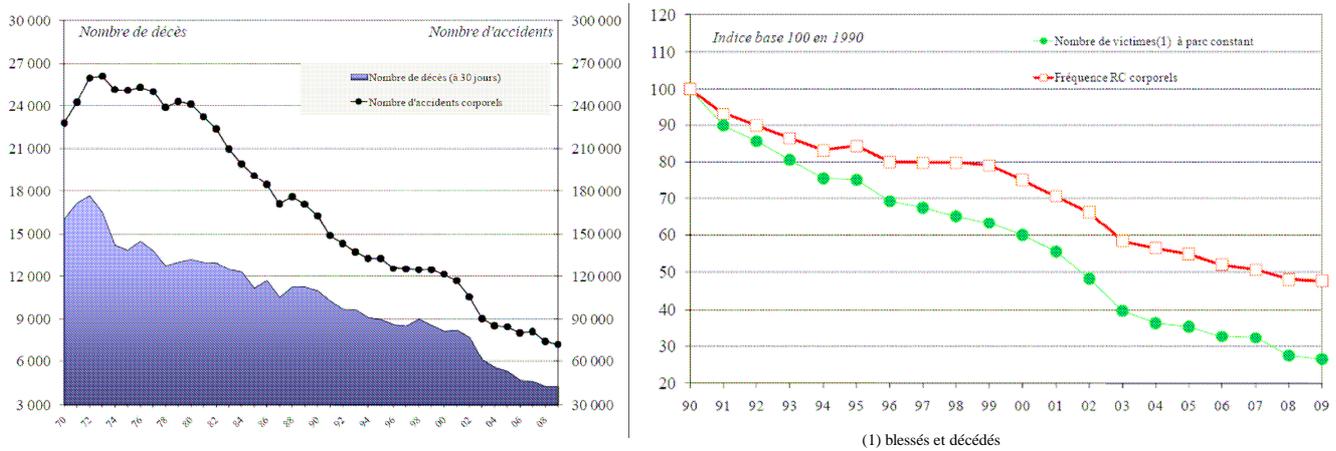
Dans ce chapitre, nous présentons l'indemnisation des accidents corporels en présentant leur risque originel puis leur contexte juridique et les modalités de gestion qui leur sont propres chez Covéa Fleet.

### II - A CONTEXTE DU RISQUE ROUTIER

La sinistralité étudiée est causée par l'activité de circulation routière. Le parc de véhicules ne cesse de croître depuis 30 ans et atteint 37,75 millions d'unités en 2011. Cette croissance est vérifiée quelle que soit la nature du véhicule (particulier, utilitaire ou car-bus, source CCFA<sup>4</sup>).

En France, les principales tendances comportementales sont la baisse des vitesses moyennes chaque année depuis 2002 et l'augmentation du taux de port de la ceinture (avant, arrière) chaque année depuis 2005.

Au niveau de la sinistralité, on relève la baisse continue du nombre de décès et d'accidents corporels de 1970 à 2009<sup>5</sup>.



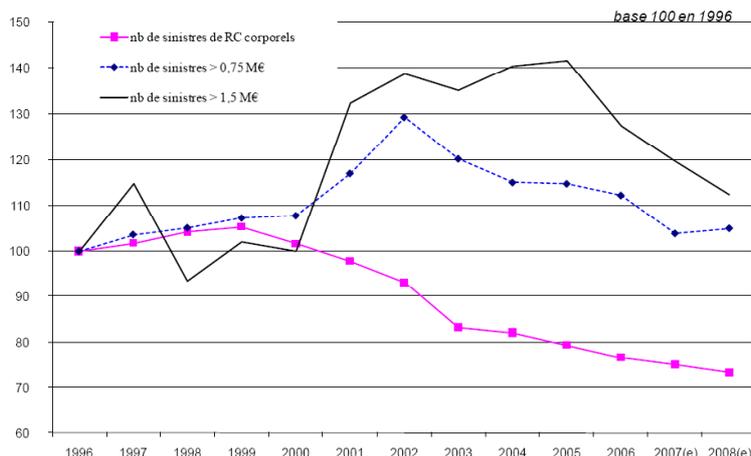
La baisse de victimes relevée par les pouvoirs publics s'inscrit dans une tendance éprouvée (fig. 1). La traduction de cette tendance est moindre chez les assureurs car de petits dommages corporels leur sont déclarés sans apparaître dans les données des forces de l'ordre (fig. 2).

Elle peut s'expliquer par les éléments suivants :

- Les politiques de prévention et de répression (radars).
- L'augmentation des équipements de sécurité passive à l'intérieur des véhicules :
  - o **La structure** du véhicule en est le premier élément. Elle absorbe une partie de l'énergie du choc et fait de l'habitacle une cellule de survie préservant les occupants de l'intrusion de tout élément. Le choix des matériaux est déterminant ainsi que la disposition des organes les plus dangereux.
  - o **La sécurité passive** passe également par les systèmes de sécurité gonflables (airbags) qui se propagent au sein de l'habitacle : passager avant, passagers arrière, airbags latéraux, ... Outre les chocs, la sécurité passive prémunit aussi les passagers contre les dommages liés au feu. Les matériaux de structure ou de revêtement de l'habitacle sont peu inflammables et ne dégagent pas de gaz toxiques en cas de combustion.
  - o La frontière entre **sécurité active** et sécurité passive s'estompe, cette dernière devenant de plus en plus interactive avec les systèmes électroniques de type ESP, ABS, capteurs de roues ... l'enjeu étant d'anticiper les situations extrêmes. Alors que les airbags se déclenchent sur détection d'un choc effectif (décélération brutale), les systèmes les plus avancés modifient le réglage des sièges ou « prétensionnent » les ceintures de sécurité avant le choc.
- La multiplication des sources de communication permettant l'alerte des secours (téléphones portables, bornes d'appel) et la localisation de l'accident (triangulation cellulaire, GPS).
- L'efficacité des secours liée à une professionnalisation accrue (formation, travail avec les forces de l'ordre) et aux moyens actuels de communication (reroutage ou débordement téléphonique) qui accroissent la rapidité d'intervention,
- Les progrès de la médecine urgentiste.

<sup>4</sup> Le CCFA est le Comité Français des Constructeurs Automobiles qui suit l'évolution du secteur automobile.

<sup>5</sup> Le marché de l'assurance automobile en 2009 – Analyse des résultats détaillés FFSA / GEMA – octobre 2010



Le graphe<sup>6</sup> ci-joint montre que le nombre de sinistres corporels graves (> 750 k€) reste lui quasiment constant.

Celui des très graves (> 1 500 k€) est plus volatil.

Concernant les coûts, on observe une inflation<sup>7</sup> entre 3,4% et 5,4% par an sur l'ensemble des sinistres corporels entre 2001 et 2009.

Pour les sinistres graves (AIPP>50%), l'inflation moyenne annuelle constatée depuis 2003 est supérieure (+7,6%), tirée vers le haut par la forte croissance des postes de tierce personne (+11,3%) et des frais médicaux et hospitaliers (+5,8%).

	Coût moyen 2008 (en euros)	Evolution moyenne annuelle depuis 2003
<b>Indemnité totale</b>	<b>988 000 €</b>	<b>+ 7,6 %</b>
Déficit fonctionnel permanent	206 000 €	- 0,2 %
Préjudices personnels	57 000 €	+ 1,4 %
Dépenses de santé actuelles et futures <sup>(1)</sup>	233 000 €	+ 5,8 %
Perte de gains professionnels actuels	33 000 €	+ 3,3 %
Assistance par tierce personne	341 000 €	+ 11,3 %
Autre frais	118 000 €	ns

ns : non significatif  
(1) Marché FFSA

La persistance d'un nombre élevé de graves, en dehors de l'aléa pur, est liée à plusieurs facteurs :

- La nature et la gravité des séquelles se modifient ; certaines victimes sont désormais relevées avec des séquelles nouvelles dont la prise en charge médicale est encore expérimentale et qui nécessitent des soins longs et coûteux.
- L'âge des victimes diminue, l'espérance de vie croît.
- La professionnalisation du recours et le lobbying associatif.
- L'inflation des postes de préjudices (nombre et montant).
- La majoration du coût des tierces personnes, poste qui représente plus du tiers de la charge d'un grave ; ce poste comportant deux paramètres (nombre d'heures allouées et montant horaire) augmente sous la pression des associations de victimes et des avocats recours. La demande de personnes qualifiées notamment pour les victimes traumatisées crâniennes graves concoure aussi à une augmentation des tarifs. La mise en œuvre des 35 heures a également généré des surcoûts liés à l'emploi de davantage de personnes augmentant mécaniquement les temps de recouvrement horaire lié au passage des consignes et générant des nombres d'heures journaliers supérieurs à 24.
- L'impact des barèmes de capitalisation des préjudices économiques futurs.

<sup>6</sup> Assurance automobile, les sinistres corporels en 2008 – FFSA / GEMA – avril 2010

<sup>7</sup> Inflation calculée à partir de l'indice des salaires de l'activité de la santé humaine.

## II - B L'ENVIRONNEMENT JURIDIQUE

---

### II - B . 1. LA RESPONSABILITE CIVILE

Ce principe est de droit commun et s'applique à tous. Les deux articles fondateurs sont :

- Article 1382 du code civil : « **Tout fait quelconque de l'homme, qui cause à autrui un dommage, oblige celui par la faute duquel il est arrivé à le réparer** ».
- Article 1384 du code civil : « **On est responsable non seulement du dommage que l'on cause par son propre fait, mais encore de celui qui est causé par le fait des personnes dont on doit répondre, ou des choses que l'on a sous sa garde...** ».

La RC a un rôle de réparation essentiellement financier. Elle se distingue des trois autres types de responsabilité que sont la responsabilité morale, la responsabilité disciplinaire et la responsabilité pénale.

La RC peut être contractuelle (non exécution d'un élément du contrat), délictuelle (le sinistre est causé par une faute intentionnelle) ou quasi-délictuelle (sinistre causé par une faute non intentionnelle ou par une personne, une chose dont on est responsable). La RC automobile est principalement **quasi-délictuelle**.

A noter que la prescription extinctive des actions en RC contractuelle est de 30 ans (sauf limite contractuelle) alors que celle de la RC délictuelle et quasi-délictuelle est de 10 ans. Pour la RC automobile qui nous intéresse, la prescription contractuelle est biennale pour l'assuré et décennale pour le tiers victime. Le délai est calculé par rapport à la date de survenance (ou de réouverture du dossier en cas d'aggravation).

Pour qu'il y ait mise en œuvre de la RC, il faut qu'il y ait simultanément :

- un **dommage** direct et certain, licite et personnel,
- un **fait générateur** (fait personnel, des choses ou d'autrui),
- un **lien** de causalité entre le dommage et le fait générateur.

Le principe qui prévaut est alors celui de la **réparation intégrale** du dommage. En corporel, cela entraîne un examen de la victime « in concreto » (i.e. prise en compte de la situation de la victime dans son environnement) et non l'application d'un barème moyen.

- Exemple en jurisprudence de RC délictuelle : « le propre de la RC est de rétablir aussi exactement que possible l'équilibre détruit par le dommage et de replacer la victime dans la situation où elle se serait trouvée si l'acte dommageable n'avait pas eu lieu » (Civ.II, 28-10-1954, JCP 1955, II, 8767).
- Exemple en RC contractuelle : « Les dommages et intérêts dus au créancier sont, en général, de la perte qu'il a faite et du gain dont il a été privé... » (article 1149 du code civil).

Principe qui se résume aussi selon « Tout le préjudice, rien que le préjudice » : la victime ne doit pas minimiser son préjudice ni s'enrichir, notamment par un cumul d'indemnités qui est interdit. C'est ce qui explique l'exercice du recours des tiers payeurs<sup>8</sup> sur l'indemnité versée.

Les juges du fond ont beaucoup de latitude pour déterminer le montant des indemnités (d'où une disparité selon les cours d'appel) et leur modalité de règlement (rente viagère ou capital).

---

<sup>8</sup> Est considéré comme tiers payeur, tout organisme versant ou susceptible de verser, par application d'une disposition légale ou conventionnelle, des prestations à une personne qui a été victime d'un accident. Exemples : sécurité sociale, mutuelle santé, assureurs, employeurs...

## **II - B . 2.      L'ASSURANCE DE LA RESPONSABILITE CIVILE AUTOMOBILE**

Cette garantie est définie le 13 juillet 1930 dans le cadre de la loi sur l'assurance automobile et reste alors **facultative**.

- « Art. 12. Les pertes et les dommages occasionnés par des cas fortuits ou causés par la faute de l'assuré sont à la charge de l'assureur, sauf exclusion formelle et limitée contenue dans la police. Toutefois, l'assureur ne répond pas, nonobstant toute convention contraire, des pertes et dommages provenant d'une faute intentionnelle ou dolosive de l'assuré.
- Art. 13. L'assureur est garant des pertes et dommages causés par des personnes dont l'assuré est civilement responsable en vertu de l'article 1384 du Code civil, quelles que soient la nature et la gravité des fautes de ces personnes. »

Elle devient **obligatoire** par la loi du 27 février 1958 (articles L211-1 et A211-1 et suivants du Code des assurances) qui vise à protéger les victimes de dommages corporels et/ou matériels lors d'un sinistre automobile.

La RC automobile est **élargie** par la loi du 7 janvier 1981 stipulant que les **membres de la famille** du conducteur ou de l'assuré sont désormais considérés comme des tiers.

De même, le 9 juin 1983, un décret (repris ensuite dans la loi 93-121 du 27 janvier 1993) **étend la notion de tiers au préposé, au souscripteur, au propriétaire et au gardien** ; seul le conducteur reste exclu.

La loi Badinter du 5 juillet 1985 que l'on étudiera plus en détail ci-après étend le périmètre de la garantie RC, vise à l'amélioration de la situation des victimes d'accidents de la circulation et à l'accélération des procédures d'indemnisation.

Nous présentons ci-dessous un exemple de rédaction de cette garantie dans un contrat flottes récent.

### **>>>>QUI EST ASSURÉ ?**

Le souscripteur, le propriétaire du véhicule, le passager et toute personne dont la responsabilité civile est engagée du fait du véhicule assuré, à l'exception des professionnels de la réparation, de la vente, du contrôle et du dépannage de l'automobile et de leurs préposés en ce qui concerne les véhicules qui leur sont confiés en raison de leurs fonctions.

### **>>>CE QUE NOUS GARANTISSONS**

L'indemnisation des dommages causés à un tiers suite à un accident de circulation, un incendie ou une explosion lorsque le véhicule assuré est impliqué dans la réalisation de ces dommages (article L.211.1 du Code). Cette garantie s'exerce dans les limites des exclusions générales.

Le montant de la garantie est illimité en cas de dommages corporels (code des assurances R211-7) et il est au minimum de 1 200 k€ par véhicule en cas de dommages matériels.

A noter qu'à l'heure actuelle, le conducteur n'est pas considéré comme tiers au véhicule et qu'il existe une assurance facultative, la garantie du conducteur, visant à le couvrir en cas de sinistre responsable et partiellement responsable (exemple : conducteur avec alcoolémie).

## **II - B . 3.      LES 7 ETAPES RECENTES DE L'INDEMNISATION DU CORPOREL**

Nous présentons ici les principales étapes juridiques ou de réflexion qui témoignent du **lent travail de maturation** conduisant à éclaircir les intentions du législateur autour de l'indemnisation des victimes suite à dommages corporels.

Il nous apparaît en effet intéressant de montrer que ce travail est engagé depuis de nombreuses années, qu'il est itératif, qu'il n'est pas achevé et qu'il converge vers une meilleure prise en compte des victimes y compris sur l'aspect financier.

En parcourant les étapes recensées ci-après, on se rend mieux compte de l'enjeu de société que ce sujet représente. Pour l'assureur, les impacts sont directs et concourent à une augmentation des coûts soit directement, soit indirectement via la réassurance.

## **La loi du 5 juillet 1985 dite loi Badinter**

Le texte de loi est consultable sur [www.legifrance.gouv.fr](http://www.legifrance.gouv.fr). Nous nous intéressons d'abord à l'origine de cette loi.

Dans le cadre du régime de responsabilité du fait des choses qui prévaut avant l'application de cette loi, le conducteur peut s'exonérer de toute responsabilité en cas de force majeure (événement extérieur, imprévisible et irrésistible) ou en cas de fait d'un tiers (exemple : faute de la victime).

Devant la difficulté de qualifier l'imprévisibilité et l'irrésistibilité, la jurisprudence rendue génère alors des décisions contradictoires.

Exemples : un chien qui passe devant une voiture est imprévisible<sup>9</sup> mais pas un chat passant devant un scooter<sup>10</sup>, un caillou projeté par un pneumatique pouvant être l'un ou l'autre<sup>11</sup>.

De même, l'appréciation de la faute de la victime permettant au conducteur de s'exonérer totalement ou partiellement est délicate. L'évolution des mentalités tend alors à limiter l'exonération partielle du conducteur au seul cas de la faute inexcusable de la victime qu'on peut définir comme un comportement répréhensible et délibérément choisi.

L'arrêt Desmares de la cour de cassation, le 21 juillet 1982, va au-delà de cette idée en ne reconnaissant au conducteur que le seul cas de force majeure pour s'exonérer, qu'il y ait présence ou non d'une faute de la victime. Cette position est alors critiquée pour son manque d'équité et le risque de dérive. L'équilibre n'est alors pas atteint et une loi est devenue nécessaire pour clarifier.

Regardons maintenant quelles sont les principales avancées de la loi Badinter.

Le législateur précise désormais que **la force majeure ou le fait d'un tiers ne sont plus opposables aux victimes** lorsqu'un véhicule à moteur est impliqué dans un accident. Seule la faute inexcusable de la victime<sup>12</sup>, si cette faute a été la cause exclusive de l'accident, l'empêche d'être indemnisée. Dans la majeure partie des cas, les victimes seront donc indemnisées par l'assureur RC du véhicule terrestre à moteur impliqué.

On peut noter ici que, dans le cas d'un accident entre deux véhicules, les victimes peuvent être des conducteurs et seront alors le plus souvent indemnisées par l'assureur adverse. En cas de faute d'un conducteur, ce droit à indemnisation sera cependant réduit ou exclu.

La deuxième évolution importante de cette loi est la création du statut de « **victimes protégées** » qui seront toujours indemnisées, y compris en cas de faute inexcusable, sauf dans le cas rare de la faute intentionnelle, principalement la tentative de suicide. Ces victimes protégées sont les mineurs de moins de 16 ans, les seniors de plus de 70 ans et les grands invalides.

Ce statut de victime protégée surprend et est alors qualifié par H. Groutel<sup>13</sup>, éminent juriste, de discriminant envers les autres victimes. Cependant, en 2010, 2 arrêts rendus par la cour de cassation mettent fin à une éventuelle question prioritaire de constitutionnalité (procédure d'appel au conseil constitutionnel) sur ce caractère discriminant entre les victimes. Serait également discriminante la différence de traitement entre le conducteur qui ne peut plus invoquer le fait d'un tiers ou la force majeure et la victime. Enfin l'application de la loi elle-même est jugée discriminante, certaines affaires ayant bénéficié de sa rétroactivité.

La troisième évolution est le **formalisme**, touchant le respect des délais et une communication écrite encadrée, que doit respecter l'assureur pour présenter ses offres d'indemnisation, qu'elles soient

<sup>9</sup> Cassation, 10-04-64.

<sup>10</sup> Cassation, 24-03-65.

<sup>11</sup> Cassation, 03-02-66.

<sup>12</sup> « Les fautes inexcusables sont très rares, limitées en pratique au cas du piéton qui franchit une barrière de sécurité ou un muret pour accéder à une voie rapide qu'il traverse, en négligeant les passages protégés souterrains ou aériens existant à proximité ». Jurisprudence automobile, janvier 2011.

<sup>13</sup> Le droit à l'indemnisation des victimes d'un accident de la circulation, H. Groutel, 1987.

provisionnelles ou totales. Cela participe à clarifier les droits de la victime, voire à les élargir. Par exemple, l'offre de transaction dûment acceptée par la victime peut désormais être dénoncée par elle sous 15 jours.

Enfin, cette loi précise les deux derniers éléments importants suivants :

- le droit de la victime à bénéficier de l'**aggravation**, défini comme le droit qu'a la victime d'obtenir un complément d'indemnisation si son état se détériore après la transaction,
- les **recours des organismes tiers payeurs** sont définis de telle sorte que la victime ne les supporte plus sur ses préjudices personnels. De même, ces recours ne peuvent plus se faire qu'à titre subrogatoire<sup>14</sup> auprès de l'assureur et non plus directement vers l'assuré.

Afin d'illustrer ce que cette loi a modifié, il nous apparaît intéressant de présenter des exemples de « sinistres Badinter » indemnisés par Covéa Fleet et qui n'auraient pas donné lieu à indemnisation auparavant au regard des fautes commises par les victimes :

- Le préposé de notre assuré circule la nuit au volant de son ensemble routier. Alors qu'il s'engage sur une route nationale non éclairée, le chauffeur aperçoit au dernier moment un piéton qui marchait sur les pointillés séparant la voie d'accélération de la voie rapide et le renverse.
- Dans le cadre d'une fête de village, au cours d'un "défilé", une personne se trouvant sur une remorque attelée à un tracteur a chuté. Cette personne bien qu'elle soit fortement alcoolisée a obtenu une indemnisation intégrale en sa qualité de passagère (article 3) d'un véhicule terrestre à moteur. Elle a été blessée très grièvement, son état interdisant d'établir son alcoolisme.
- Notre assuré circulait au volant de son véhicule lorsqu'un cycliste n'a pas respecté un stop. Le sociétaire très surpris n'a rien pu faire pour éviter le choc. Là encore le blessé a obtenu une indemnisation (article 3).

En conclusion, on peut remarquer que la loi s'attache davantage à préciser l'accès à l'indemnisation que la quantification de l'indemnisation. Elle a nécessité de nombreuses précisions par la jurisprudence, notamment sur les notions « d'implication », de « véhicule terrestre à moteur » et « d'ouverture de la circulation au public ».

Par ailleurs, sur le fond, on peut faire 2 interprétations de cette loi :

- soit on considère qu'elle se rattache au droit commun de la réparation, venant corriger l'arrêt Desmares et règlementant les causes d'exonération. Le fondement du droit à réparation reste alors l'article 1384 du code civil. Le droit commun prévaut alors toujours sur le droit à la réparation, notamment s'agissant du lien de causalité entre une faute et le dommage.
- soit cette loi crée un système autonome d'indemnisation, distinct du droit commun de la responsabilité civile. En excluant la force majeure et le fait du tiers comme causes d'exonération, le législateur aurait écarté sciemment les raisonnements classiques de la responsabilité civile, fondés notamment sur un lien de causalité. En effet, que reste-t-il de la causalité quand une personne est obligée de réparer un dommage alors qu'elle a démontré que la source unique de celui-ci se situe ailleurs que dans son fait personnel ou dans le fait de la chose dont elle a la garde ? On recherche alors moins un responsable que le débiteur d'une obligation d'indemnisation qui est la compagnie d'assurances d'un conducteur.

---

<sup>14</sup> Les tiers payeurs peuvent invoquer les mêmes textes que la victime pour exercer leur recours qui doit être présenté seulement sur les préjudices que leurs prestations ont contribué à réparer. Pour l'assureur, la méthode de calcul de l'indemnisation à la victime est : [+ préjudices soumis au recours des tiers-payeurs (PR) – leurs créances (C)] \*  $I_{\{PR > C\}}$  + préjudices personnels.

**Le rapport Lambert-Faivre sur l'indemnisation du dommage corporel de juin 2003**

Ce rapport est consultable sur <http://lesrapports.ladocumentationfrancaise.fr>.

Son objectif est la « recherche de modalités d'indemnisation de la victime plus justes et plus transparentes » sous-tendu par l'élaboration d'un référentiel national des différents postes de préjudices et par une définition plus claire de ces derniers. Ce groupe de travail est initié avec le CNAV<sup>15</sup>.

Il y est notamment constaté que la restriction du recours des tiers payeurs défini dans la loi Badinter n'est pas souvent appliquée. Les victimes ne sont pas satisfaites du reliquat d'indemnités versé après que les tiers payeurs se soient « servis » en s'appuyant sur des jurisprudences anciennes.

Une **distinction** terminologique est faite entre dommage corporel (constat médical) et préjudices (constat juridique).

La nécessité d'un **barème médical unique** est rappelée et le principe d'une liste non limitative des postes de préjudices, conformes aux études européennes, est posé.

Ces préjudices sont classés selon qu'ils relèvent du patrimoine ou de la personne.

Pour une même typologie de traumatisme, une indemnisation personnalisée des préjudices personnels et professionnels doit être faite. On épingle ainsi la prévalence de l'ancien système basé sur l'IPP (incapacité permanente partielle), initialement instauré pour qualifier l'incapacité des accidents du travail et qui, abusivement transposé en droit commun, conditionnait l'indemnisation toute entière.

La nomenclature des postes de préjudices préconisée distingue 3 axes : **nature** du préjudice, sa **durée** et la **qualité** de la victime :

	Préjudices économiques patrimoniaux		Préjudices non économiques personnels	
	Temporaire	Permanent	Temporaire	Permanent
Victime directe	Analyse patrimoniale (avoir) : quelles sont les pertes subies ?		Analyse extrapatrimoniale (être) liée à la personne humaine	
Victime(s) par ricochet	quels sont les gains manqués ?			

L'importance de l'**évaluation intuitu personae** est rappelée ainsi que la préférence pour un mode d'évaluation indemnitaire référencée plutôt que barémée ; en lien avec la loi Badinter, cela nécessite des évolutions du fichier<sup>16</sup> géré par l'Agira<sup>17</sup> à qui l'on reproche d'être peu rempli (environ 25% des cas) notamment par les greffes, peu diffusé et sans contrôle de l'Etat.

Le rapport demande l'**unicité du barème** de capitalisation (taux, table), sa mise à jour chaque année et un régime fiscal unique (capital ou rente).

La tierce personne est perçue comme fondamentale pour la sécurité de la victime, évolutive et il est demandé que l'indemnisation en rente de ce poste soit indexée et revue tous les 5 ans.

Par poste de préjudices, une analyse est menée pour illustrer les prestations versées par les organismes sociaux et sert donc à définir leur **nouveau périmètre de recours**.

Enfin, il est demandé que soit exclu tout recours sur les préjudices personnels et les préjudices non indemnisés par le tiers payeur. Pour cela, une rédaction plus claire d'articles de la loi Badinter et du code de la sécurité sociale est proposée.

<sup>15</sup> Conseil national de l'aide aux victimes ; créée en 1999, cette instance de concertation est chargée de faire toute proposition portant notamment sur l'accueil, l'information, la prise en charge et l'indemnisation des victimes.

<sup>16</sup> Fichier des victimes indemnisées (FVI) consultable sur <http://www.victimesindemniees-fvi.fr> qui existe depuis 1988.

<sup>17</sup> Agira : association pour la gestion des informations sur le risque automobile.

## Le rapport de juillet 2005 dit rapport Dintilhac

Le rapport est consultable sur [www.courdecassation.fr](http://www.courdecassation.fr).

Le 29 septembre 2004, le conseil des ministres lance un groupe de travail sur le « droit des victimes de préjudices corporels à une juste indemnisation » afin de procéder à « **l'établissement d'une nomenclature des chefs de préjudice corporel cohérente**, reposant sur une distinction claire entre les préjudices économiques et non économiques, notamment en ce qui concerne l'incapacité permanente partielle (IPP) ».

On veut définir une nomenclature incontestable en respectant les principes d'une indemnisation complète, équitable et d'une égalité de traitement entre les victimes.

Pourquoi ce besoin est-il ainsi exprimé ?

Les postes de préjudices soumis à l'expertise médicale sont correctement définis, les autres sont soumis à interprétations voire à multiplications abusives. Des textes de lois et notamment celui de la loi Badinter du 5 juillet 1985 imposent que l'indemnisation soit calculée et ventilée par poste de préjudice. Il importe de clarifier le contenu de tous ces postes.

De plus, l'insistance sur la nécessité de préciser l'IPP vient du fait que ce poste est encore souvent interprété comme ayant une composante physiologique et une composante économique.

Enfin, il n'existe pas de référentiel national. Il faut donc en créer un pour harmoniser les pratiques.

De nombreux points d'accord sont trouvés avec le rapport Lambert-Faivre :

- le distinguo entre dommage (factuel) et préjudice (droit),
- les trois axes de structure d'une nomenclature sont validés : préjudices patrimoniaux et extra patrimoniaux, temporaires ou permanents et ceux de la victime directe et des victimes indirectes,
- principe de non limitation des postes de préjudices,

Les différentes nomenclatures existantes sont étudiées :

- celle des juristes de l'Office national d'indemnisation des accidents médicaux (ONIAM),
- celle du fonds d'indemnisation des victimes de l'amiante (FIVA),
- celle relative aux missions d'expertise médicale de l'AREDOC,
- celle du rapport Lambert-Faivre.

Le groupe de travail prend en compte des éléments de droit européen et de droit comparé (Espagne, Canada, Allemagne, Italie).

Pour supprimer l'ambiguïté de certains postes de préjudices qui réparaient à la fois des préjudices d'ordre physiologique et professionnel, la nouvelle nomenclature abandonne les notions d'incapacité temporaire de travail (ITT) et d'IPP au profit de 2 nouveaux postes de préjudices distincts :

- le préjudice patrimonial qui répare les pertes de gains professionnels de la victime,
- le préjudice extra patrimonial qui répare le déficit fonctionnel.

Cette nomenclature consacre par ailleurs la création de 3 nouveaux postes de préjudices extra patrimoniaux issus de la jurisprudence :

- le préjudice esthétique temporaire (pour les grands brûlés et les traumatisés de la face),
- les préjudices permanents exceptionnels qui constituent un préjudice extra patrimonial permanent particulier, non indemnisable par un autre biais,
- le préjudice lié à des pathologies évolutives qui cible les affections résultant d'une contamination par virus de l'hépatite C, du VIH ou de la maladie de Creutzfeldt-Jacob et dont la particularité est qu'il existe en dehors de toute consolidation.

Soucieux de l'application de cette nouvelle nomenclature, notamment vis-à-vis des tiers payeurs, le groupe de travail préconise :

- d'indemniser chaque préjudice poste par poste comme la loi Badinter le demandait déjà,
- de supprimer l'automatisme pratiqué à l'époque : préjudices patrimoniaux → recours, préjudices extra patrimoniaux → pas de recours,
- d'autoriser le recours sur les postes justifiant d'une indemnisation versée au titre de ce préjudice.

Cette nomenclature est aujourd'hui **reconnue** et **utilisée** par les assureurs et les magistrats alors même qu'elle n'est pas officielle. Par exemple, dans le fichier FVI des sinistres réglés en 2010, 31 % des indemnisations sont faites sous le format Dintilhac.

**La loi du 21 décembre 2006 qui a imposé le principe du recours subrogatoire des caisses de sécurité sociale poste par poste.**

Cette loi est consultable sur <http://www.textes.justice.gouv.fr>.

Le recours subrogatoire des tiers payeurs est au cœur des réflexions développées dans les 2 groupes de travail de Lambert-Faivre et de Dintilhac.

Cette loi vise à mieux circonscrire les conditions dans lesquelles la créance des tiers payeurs peut venir en diminution de l'indemnisation allouée à la victime.

Le calcul de l'indemnisation du dommage corporel supposait jusqu'à présent d'établir pour chaque chef de préjudice l'indemnité due à la victime et de faire masse de ceux se rapportant aux préjudices à caractère économique, lesquels incluaient le déficit fonctionnel, afin d'en soustraire les créances des tiers payeurs. En outre, un éventuel partage de responsabilité, par lequel la victime se voyait reconnue comme étant à l'origine d'une partie de son dommage, était inopposable aux tiers payeurs.

La combinaison de ces 2 règles aboutissait à un **paiement préférentiel des tiers payeurs**, qui recouvraient leurs créances dans la seule limite de la somme allouée en réparation de l'ensemble des préjudices soumis à recours, y compris les indemnisations réparant des chefs de préjudice pour lesquels les tiers payeurs n'avaient versé aucune prestation.

L'ambiguïté de ces modalités d'exercice du recours au regard du droit de la subrogation amène le législateur à une **clarification des règles** applicables en la matière.

Cette évolution législative entraîne une modification en profondeur de la manière de calculer l'indemnisation du dommage corporel, selon les 2 orientations suivantes :

- une **indemnisation poste de préjudice par poste de préjudice**,
- un **paiement préférentiel de la victime** par rapport aux caisses de sécurité sociale et autres tiers payeurs.

A partir de la probable diminution des recours que cette loi génère, on peut s'interroger sur les impacts économiques pour les finances publiques et sur un éventuel financement qui pourrait se faire via une augmentation du taux de taxe de la garantie RC.

**Le livre blanc sur l'indemnisation du dommage corporel, association française de l'assurance 04-2008**

Ce rapport est consultable sur <http://www.associationfrancaisedelassurance.fr>.

Les assureurs français règlent chaque année à des tiers 186 000 sinistres corporels automobile soit 4,3 milliards d'euros.

S'appuyant sur les résultats d'une étude<sup>18</sup> réalisée pour eux par le Crédoc<sup>19</sup>, ils publient leurs souhaits d'évolutions, constatant que l'insatisfaction des victimes reste grande.

<sup>18</sup> Etude sur la réinsertion sociale et professionnelle des personnes sortant d'établissement ou service sanitaire de médecine physique réadaptation, Credoc, juin 2007.

<sup>19</sup> Centre de recherche pour l'étude et l'observation des conditions de vie.

De cette étude, nous avons retenu les éléments suivants concernant 66 victimes ayant droit à une indemnisation par les compagnies d'assurance :

- 55% des personnes interrogées estiment que la procédure d'indemnisation a été rapide. A contrario, 27% déclarent que cela n'a pas du tout été le cas.
- 47% estiment que le montant a été suffisant, 27% estiment qu'il ne l'a pas du tout été.
- 52% déclarent que le montant de l'indemnisation leur a permis de réorganiser la vie à la maison, 39% pour la vie sociale et 40% pour la vie professionnelle. A l'opposé, 42% que l'indemnisation ne leur a pas du tout permis de réorganiser leur vie à la maison ou sociale et 41% pour la vie professionnelle.

Ainsi, si pour chaque item, environ 50% des interrogés ont une opinion favorable de l'indemnisation, une part non négligeable en ont une opinion très défavorable. L'opinion est d'autant plus défavorable lorsqu'on évoque l'impact de l'indemnisation sur la **réorganisation de la vie** tant quotidienne que professionnelle. Dans ce domaine, un grand nombre (92%) ne voit aucun effet pour l'emploi. Cette opinion ne varie guère en fonction du type d'accident, du montant et de la forme d'indemnisation.

Il nous paraît intéressant de confronter les demandes des assureurs issues du livre blanc avec le projet du législateur actuellement à l'étude au Sénat, celui de Lefrand. L'étude détaillée montre que les attentes des parties sont peu partagées (cf. Annexe VIII-A). On présente ci-dessous ce projet puis une conclusion des écarts.

### **Le projet de loi Lefrand du 5 novembre 2009**

Texte consultable sur <http://www.assemblee-nationale.fr>.

Ce texte veut compléter la loi Badinter du 5 juillet 1985 par la mise en place d'outils communs d'évaluation du préjudice pour lever progressivement les disparités et inégalités importantes dans l'indemnisation des victimes. Ces évolutions souhaitées par la profession de l'assurance dans le cadre du livre blanc devraient apporter la garantie d'une meilleure équité entre toutes les victimes mais aussi d'une meilleure transparence.

Ce projet est issu des 2 autres constats suivants :

- mauvaise application du principe du contradictoire (les deux parties peuvent défendre leur intérêt),
- remise en question de l'organisation de l'expertise médicale actuelle générant des conflits d'intérêt (triple casquette : expert judiciaire, expert de compagnie d'assurances, expert de la victime).

De plus, ce texte étend aux tramways et chemins de fer l'application de la loi Badinter.

### **Conclusion sur les écarts entre le livre blanc et la loi Lefrand**

On peut penser que les assureurs seront satisfaits des points suivants :

- Transparence et égalité de traitement entre les victimes, consensus avec les associations de victimes, complément par rapport à la loi Badinter sur l'indemnisation,
- Barème médical unique, missions types d'expertise, unique nomenclature officielle des postes de préjudices, réactualisation du barème de capitalisation<sup>20</sup> : table unique pour le calcul des préjudices futurs et pour la conversion des rentes.

Et moins de ceux-là :

- Le référentiel national d'indemnisation restant « indicatif », l'instauration d'une rente provisoire jusqu'à la stabilisation puis la révision de la rente après stabilisation,
- La loi et la nomenclature ne concerneraient que les accidents de la circulation, pas les autres accidents corporels,
- L'absence de dynamique autour du projet de vie.

<sup>20</sup> Le barème de capitalisation n'est en effet pas unique. Depuis 2004, les assureurs utilisent le BCIV (barème de capitalisation pour l'indemnisation des victimes) pour régler, en remplacement du barème issu du décret du 06/01/86 devenu obsolète. Le BCIV utilise les tables TH-TF 00-02 et une moyenne du TEC 10 sur les 24 derniers mois. Le barème de la Gazette du Palais, depuis novembre 2004, est basé sur la table 2001 et un taux de 3,20%. Ce dernier qui détermine un capital constitutif de la rente plus important, est surtout utilisé par les avocats de recours.

Les points suivants restent à surveiller :

- la définition qui sera retenue de la consolidation : stabilisation fonctionnelle seule ou avec stabilisation situationnelle permettant une indemnisation plus juste du préjudice professionnel,
- la désignation de la nomenclature Dintilhac comme référente ou pas et si, sous la pression des avocats de recours, elle sera non limitative, ouvrant la porte aux décisions jurisprudentielles,
- la diminution du taux du barème de capitalisation pour prendre en compte l'inflation et si le FGAO reste encore chargé de la revalorisation des rentes,
- l'effet, sûrement inflationniste, de la proposition d'indemnisation pour chaque poste de la nomenclature et sa compréhension par les familles de victimes ; d'autant plus si l'offre est faite nette des recours des tiers payeurs.

#### **Accident du travail et faute inexcusable de l'employeur**

Décision consultable sur <http://www.conseil-constitutionnel.fr>.

En juin 2010, le conseil constitutionnel a élargi le nombre de postes de préjudices auxquelles les victimes d'un accident du travail ont droit, en cas de faute inexcusable de l'employeur.

Désormais les postes suivants présentant des enjeux conséquents pourront être indemnisés : l'aménagement du logement, du véhicule, les frais médicaux non pris en charge ou l'assistance d'une tierce personne.

## **II - B . 4. PROSPECTIVE SUR LES TRAVAUX EN COURS**

### **Le rapport d'information Anziani et Béteille**

Projet consultable sur [www.senat.fr](http://www.senat.fr).

Ce texte repose sur le rapport Catala (2005) et concerne toute la responsabilité civile, pas seulement celle de l'automobile.

Le constat posé est que le droit de la responsabilité civile a fortement évolué et s'est sensiblement complexifié, au prix de nombreuses incohérences et d'une certaine insécurité juridique, sous le double effet du développement de la jurisprudence et de la multiplication, au coup par coup, de régimes particuliers tels que celui des accidents de la circulation.

S'agissant des règles propres aux accidents de la circulation édictées par la loi du 5 juillet 1985, le groupe de travail préconise, d'une part, d'assimiler les accidents de chemin de fer et de tramway<sup>21</sup> aux autres accidents dans lesquels un véhicule terrestre à moteur (VTAM) est impliqué, d'autre part, **d'assimiler le conducteur aux autres victimes** d'un accident de la circulation.

S'attachant à la question de l'évaluation du préjudice, le groupe de travail propose d'obliger le juge à procéder à une évaluation distincte pour chaque chef de préjudice allégué sans pouvoir, sauf pour les dommages de plus faible montant, se limiter à une évaluation in globo, « toutes causes de préjudice confondues ». Il a aussi précisé :

- de prévoir l'adoption d'un barème national d'invalidité qui puisse servir de référence au juge dans son évaluation du dommage,
- de privilégier le versement de la réparation sous forme de capital pour les dommages légers,
- de conforter la possibilité offerte au juge, lorsqu'il décide le versement d'une rente indexée, de déterminer cet indice et de prévoir, le cas échéant, les conditions de révision de la rente.

A noter qu'en voulant aligner le sort des victimes conductrices sur celui des non conducteurs, cette loi nous éloignerait des pratiques européennes, le conducteur fautif étant alors toujours indemnisé (sauf en cas de faute inexcusable) alors qu'en Europe on leur oppose leur faute simple.

---

<sup>21</sup> Le tramway est considéré aujourd'hui comme VTAM seulement sur les carrefours partagés avec les voitures.

### La garantie du conducteur

Désormais les principales victimes restant exclues du droit à l'indemnisation sont les conducteurs fautifs ou sans tiers (environ 20% des cas d'accident<sup>22</sup>). Pour y remédier, 2 options sont aujourd'hui étudiées, avec un impact certain sur le prix et notamment en deux-roues :

- soit l'extension de la loi de 1985 aux conducteurs fautifs ou sans tiers (cf. projet de loi Béteille ci-dessus), ce qui aurait pour effet de renforcer le particularisme français vis-à-vis des autres pays européens et de définir une nouvelle (et étrange) notion de véhicule « responsable » pouvant causer des dommages au conducteur-tiers...
- soit la souscription rendue obligatoire de l'actuelle garantie du conducteur qui serait renforcée, plus indemnitaire que forfaitaire, basée sur les postes de préjudices de droit commun et avec des montants de garanties majorés. Proposition moins coûteuse du fait des plafonds moindres et soutenue par la profession.

### Indemnisation sur les voies non ouvertes à la circulation publique

Une question prioritaire de constitutionnalité a été déposée en juillet 2011 par la cour de cassation pour étendre le droit à indemnisation sur les zones aéroportuaires ou les chantiers selon la loi Badinter alors qu'il relevait jusqu'à présent de l'accident de travail avec des indemnisations moins complètes (cf. faute inexcusable).

Le conseil constitutionnel a toutefois décidé le 23 septembre 2011 que la loi Badinter ne serait pas étendue aux accidents survenus dans un lieu non ouvert à la circulation publique :

- le préposé passager, victime d'un accident de circulation continuera donc à être indemnisé dans le cadre du code de la sécurité sociale ;
- s'il est conducteur, cette indemnisation pourra être complétée par la souscription d'une garantie dommages corporels du conducteur.

## **II - B . 5.      CONCLUSION**

A travers cette étude, nous avons découvert **l'importance de la jurisprudence** qui continue parfois longtemps de préciser le dispositif juridique originel. Le cadrage du droit à l'indemnisation des victimes est un long processus de maturation de l'ensemble des acteurs, non encore abouti en France. Pour l'assureur, la traduction de ce phénomène est une « pollution » des cadences, souvent diagonalement. Cela explique la nécessité du redressement des coûts lors de tout travail préalable de modélisation en posant une hypothèse sur l'inflation des sinistres (cf. IV-C.2).

Pour tous, le besoin de plus de clarté, de transparence, d'équité et de sécurité juridique est mis en avant. On peut toutefois s'interroger sur l'intérêt des conseils spécialisés d'aboutir à un dispositif trop cadré.

La **notion de responsabilité** est très chahutée, voire appauvrie et bientôt peut-être vidée de son contenu si elle devient limitée à la faute inexcusable.

D'ores et déjà, le législateur fait jouer aux assureurs un rôle quasi-social qui alourdit leurs charges et dont les limites restent le maintien de leur propre solvabilité et d'une tarification acceptable par l'assuré, soit un **délicat équilibre** à trouver. En effet, avec des primes trop élevées, le risque est de voir augmenter le nombre de non assurés.

En illustration, si la loi intègre les conducteurs au rang des victimes, le surcoût de la garantie RC serait de 20 à 25%<sup>23</sup> ; on peut se demander combien d'assurés cette mesure découragerait du paiement de leur prime.

Mais les assureurs ne sont pas les seuls à s'interroger : **les réassureurs** auront-ils la capacité de maintenir une garantie illimitée en dommages corporels de l'assurance obligatoire ? Par le passé, ils ont déjà prudemment

<sup>22</sup> Jurisprudence automobile, janvier 2011.

<sup>23</sup> Jurisprudence automobile, janvier 2011.

réduit la voilure<sup>24</sup>.

Beaucoup de décisions seront prises prochainement qui auront des impacts financiers forts pour les assureurs. Et au-delà, les assureurs vont-ils réussir à instaurer une culture du **projet de vie** initiée dans la loi Borloo du 26 juillet 2005 et qui projette la victime dans un avenir dynamique incluant une réinsertion sociale et professionnelle<sup>25</sup> ?

## **II - C LA GESTION DU SINISTRE CORPOREL**

---

Selon qu'on est victime ou pas, il est évident que les étapes clés ne sont pas les mêmes. Pour la victime, parmi les principales, on peut citer l'atténuation des souffrances, la guérison et une autonomie retrouvée quand c'est possible. Pour l'assureur, c'est souvent la phase de consolidation qui sera l'étape clé, lui permettant d'avoir une vision avancée du coût final de l'indemnisation à verser. Pour décrire ces étapes, on se placera ici du point de vue de l'assureur pendant la gestion de l'indemnisation due à la victime.

### **II - C . 1. LES ETAPES CLES DE LA GESTION D'UN SINISTRE CORPOREL**

#### **La survenance**

C'est la date de l'accident. Elle est objectivée par le constat amiable ou le procès-verbal des forces de l'ordre (PV). Elle compte pour valider la garantie et pour les délais d'information à la victime.

#### **La déclaration de l'accident**

Elle peut être faite par l'assuré, son intermédiaire d'assurances ou son conseil. Certains contrats prévoient une déchéance de garantie si la déclaration n'est pas faite dans les 5 jours mais, dans les faits, l'assureur doit alors prouver que ce retard lui a porté préjudice ce qui est rarement le cas. Dans le cadre de l'IRCA (cf. II-C.2), la victime choisit à cette étape qui l'indemniser, sa compagnie ou celle adverse.

#### **L'ouverture du dossier sinistre**

Elle génère la vérification de l'acquisition de la garantie :

- le contrat est en règle : dernier avenant et conditions particulières signés, prime payée,
- la garantie est souscrite,
- les circonstances de l'accident autorisent l'application du contrat.

#### **L'évaluation initiale du sinistre**

C'est souvent un forfait d'ouverture correspondant à un coût moyen ultime ou, si le dossier est déjà bien étayé par un rapport médical, une évaluation plus détaillée. Le PV permet de déterminer les circonstances de l'accident, les responsabilités des acteurs et donc le taux retenu pour l'indemnisation.

#### **Les expertises**

Les principales sont les expertises médicales requises par le médecin chargé d'évaluer le dommage. Plusieurs types de mission existent pour guider l'expert : blessé léger, traumatisé crânien grave, troubles locomoteurs graves, aggravation. Ces missions sont définies par l'AREDOC<sup>26</sup>.

D'autres expertises existent : les études CESVI<sup>27</sup> reconstituant l'accident et les vitesses des véhicules à partir de la déformation des pièces, celles demandées lors de l'instruction pénale (expert pneumatique,...), l'étude d'aménagement du logement par un architecte, ...

---

<sup>24</sup> Suite à l'accident de circulation du tunnel du Mont Blanc en mars 1999, limitation en 2003 de la RC matérielle qui passe de l'illimité à 100 millions d'euros par véhicule et suite à l'accident des tribunes de football à Furiani en mai 1992, limitation des dommages corporels en assurance non obligatoire qui passe de l'illimité à 8 millions d'euros.

<sup>25</sup> Dans ce cas la période moyenne de consolidation étant de 4 ans, il faudra faire évoluer les délais d'offres de la loi Badinter.

<sup>26</sup> Association pour l'étude de la réparation du dommage corporel, organisme professionnel des assureurs.

<sup>27</sup> CESVI France est un centre de formation et de recherche technique, appliquée au domaine de l'assurance automobile.

### **La révision des évaluations**

La cause principale est médicale : évolution ou information accrue sur l'état de santé de la victime. Cela peut être aussi lié à un changement du droit à indemnisation ou du barème de capitalisation (taux ou table).

### **La consolidation médico-légale**

Cette étape importante fait l'objet d'une définition précise<sup>28</sup>. La date est déterminée par le médecin expert, judiciaire ou celui de l'assureur et/ou de la victime. Une offre d'indemnisation des préjudices permanents doit alors être faite<sup>29</sup>. C'est aussi à partir de cette date qu'on solde les préjudices temporaires.

### **Les provisions**

Elles sont régies par la loi Badinter quand le délai de consolidation dépasse 3 mois. L'inspecteur fait le point avec la victime régulièrement pour régler les dépenses passées, majorées d'une avance de trésorerie.

### **L'offre d'indemnisation**

La transaction est le mode majoritaire : 95% des cas quand blessures et 78% quand décès (source FFSA). Quand l'écart de vues entre les 2 parties est trop important (désaccord sur l'analyse médicale, exigence trop élevée du conseil de la victime,...), la transaction n'est pas possible et le recours au juge a lieu.

### **Le règlement**

Il est effectué en part nette des avances, en capital ou rente. Il se déclenche suite au procès-verbal d'accord entre l'inspecteur de la compagnie et le parti de la victime<sup>30</sup> ou suite à la décision judiciaire.

### **La clôture**

Elle intervient quand les règlements de la victime et des tiers-payeurs sont effectués. La rente est mise en place ; chez Covéa, un département particulier suit alors le versement des rentes.

### **L'aggravation**

On distingue 3 types de réouverture pour aggravation

- médicale : de nouvelles séquelles liées à l'accident surgissent qu'il faut évaluer. La difficulté est de constater le lien avec l'accident.
- économique : le sinistre est ancien, a été réglé avec les paramètres économiques et réglementaires de l'époque et ceux-ci ont changé (ex : coût horaire de tierce personne).
- situationnelle/environnementale : l'environnement de la victime évolue (décès d'un parent s'occupant d'elle, changement de structure d'accueil, naissance d'un enfant,...) et nécessite une adaptation.

## **II - C . 2.      LES CONVENTIONS INTER-COMPAGNIES EN CORPOREL**

### **La convention IRCA et le principe de l'indemnisation directe**

Ce principe, établi dans une convention passée entre les sociétés d'assurances et appelée convention d'indemnisation et de recours corporel automobile (IRCA), est applicable dans la plupart des accidents et permet **d'accélérer l'indemnisation des victimes**.

L'IRCA fonctionne pour tous les sinistres corporels et formalise particulièrement les relations entre assureurs pour ceux inférieurs à 5% d'AIPP ; cela concerne presque 90% des accidents corporels de circulation. L'IRCA est en application depuis le 1<sup>er</sup> avril 2002 et concerne les accidents de la circulation survenus en France métropolitaine, dans les départements d'outre-mer ou sur le territoire de la principauté de Monaco.

Le principe est que l'assureur prend le mandat pour le conducteur du véhicule qu'il assure ainsi que pour ses passagers à raison d'un mandat par victime. Les occupants de chacun des véhicules impliqués dans l'accident

<sup>28</sup> C'est le moment où les lésions se fixent et prennent un caractère permanent tel qu'un traitement n'est plus nécessaire, si ce n'est pour éviter une aggravation, et qu'il est possible d'apprécier un certain degré d'incapacité permanente réalisant un préjudice définitif.

<sup>29</sup> 2 cas sont possibles : soit la date de consolidation a lieu dans le trimestre qui suit l'accident ; l'assureur dispose alors de 8 mois à compter de la date de l'accident pour faire une offre définitive. Soit la date de consolidation a lieu après ; l'assureur doit faire une offre provisionnelle dans les 8 mois qui suivent l'accident et une offre définitive dans les 5 mois qui suivent la consolidation.

<sup>30</sup> Si la victime est un mineur ou un majeur sous tutelle, ce procès-verbal est soumis au juge des affaires matrimoniales.

sont indemnisés directement par l'assureur RC du véhicule dans lequel ils se trouvaient. Les recours entre les différents assureurs s'effectuent dans un second temps en fonction des règles de recours adoptées dans cette même convention.

L'IRCA montre la capacité qu'ont les assureurs de s'organiser indépendamment des législateurs pour accélérer les indemnisations, en définissant leur propre protocole de recours (barème).

Les conseils spécialisés que l'IRCA peut priver d'opportunités commerciales, utilisent comme argument le fait que l'assureur de la victime n'a pas d'intérêt à indemniser au-delà du forfait conventionnel. C'est oublier que l'assureur s'appuie sur des experts médicaux indépendants et qu'il arbitre l'économie d'une convention sur un grand nombre de sinistres et pas au dossier (principe de mutualisation).

## **II - C . 3. LA GESTION DES SINISTRES CORPORELS DANS LE GROUPE COVEA**

### **Organisation de la gestion des sinistres corporels**

On a vu en I-B.2. qu'il existe trois entités pour gérer les sinistres corporels, à partir d'un critère médical :

- légers : victimes dont le taux d'AIPP  $\leq 10\%$  ce qui inclue le périmètre de l'IRCA,
- médians : les légers avec incidence professionnelle ou les victimes de taux d'AIPP  $\in ]10\% ; 50\%]$  sans besoin en aide humaine ou les victimes décédées,
- graves : victimes dont le taux d'AIPP  $> 50\%$  ou médians avec besoin en aide humaine.

Les gestionnaires sont garants d'une vision globale du dossier au niveau garantie, procédure et créance des tiers payeurs. Des pouvoirs de règlement incluant l'évaluation, la conduite à tenir sur le dossier et la décision judiciaire sont définis par niveau hiérarchique.

Les inspecteurs prennent contact avec la victime, son avocat et l'assureur adverse et font le lien avec le gestionnaire. Le groupe Covéa a une soixantaine d'inspecteurs corporels dont un tiers est spécialisé sur les victimes graves. Ils assistent aux expertises médicales, versent des provisions et promeuvent l'élaboration d'un projet de vie pour la victime. **Le gestionnaire pilote l'événement et l'inspecteur accompagne la victime.**

### **Une spécialisation pour les graves**

Pour ces sinistres atypiques à fort enjeu financier, il s'avère nécessaire de spécialiser des gestionnaires et des inspecteurs pour maîtriser les arcanes du droit et de la jurisprudence autant que les conseils des victimes, eux-mêmes regroupés en association<sup>31</sup>.

De même, il importe de constituer un réseau externe d'experts : des avocats spécialisés pour une meilleure défense judiciaire, des médecins expérimentés pour l'identification des lésions et des séquelles possibles...

L'argumentation étayée d'un expert renforce la crédibilité, permet le dialogue et participe à trouver des solutions réalistes moins onéreuses (ex : nombre d'heures de TP et coût horaire, aménagement du logement).

### **Grands principes de la politique d'évaluation et de règlement**

#### **Le forfait d'ouverture**

En corporel automobile, c'est encore un unique coût moyen ultime qui est retenu. Le mécanisme de l'IRCA impliquera néanmoins d'en distinguer la valeur selon la gravité présumée.

#### **L'évaluation des sinistres**

L'évaluation au coût réel est recherchée. Une projection de l'inflation des coûts médicaux est faite.

#### **Nomenclature Dintilhac**

Depuis 2007, le FGAO et les assureurs l'utilisent pour les graves. Elle est désormais étendue aux indemnisations des corporels légers et médians, avec un certain effet inflationniste à dire d'expert.

---

<sup>31</sup> L'ANADAVI : association nationale des avocats de victimes de dommages corporels, créée en 2004.

## Le barème

Pour les indemnités transactionnelles, c'est le BCIV qui est utilisé (cf. II-B.3). Il existe aussi un référentiel interne souvent actualisé, validé par le service médical et organisé par typologie de victimes : les traumatisés crâniens < 80% d'AIPP, ceux > 80% d'AIPP, les amputés, les paraplégiques et les tétraplégiques.

## Rente ou capital ?

Dans le cadre d'une transaction, l'assureur peut orienter les modalités de versement de l'indemnité. Quand la décision est judiciaire elle appartient au juge.

L'indemnité en capital a l'avantage psychologique de présenter un montant total important et de rentrer dans la succession en cas de décès de la victime. Les intérêts de l'indemnisation versée en capital sont assujettis à l'impôt, pas le capital. Dans la pratique, il est souvent mal géré par la victime et/ou sa famille. De ce fait, la victime s'expose à ne plus pouvoir couvrir ses frais futurs et, la transaction étant définitive, la victime n'a alors aucun recours. Enfin, le capital subit l'augmentation du coût de la vie et s'érode.

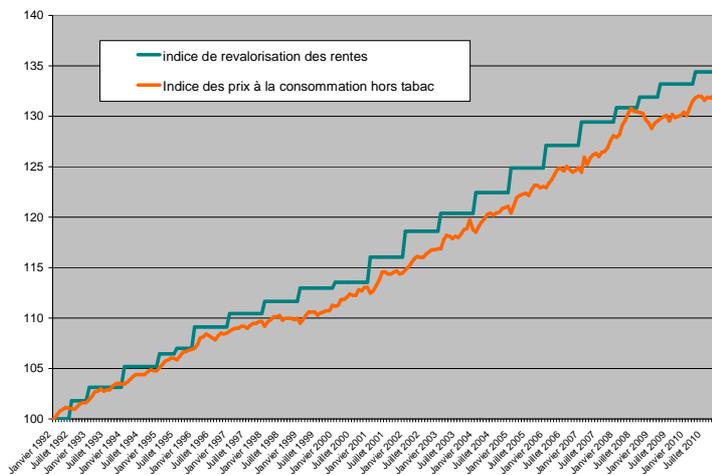
Dans le groupe Covéa et de plus en plus souvent par les juges du fond, le versement en rente (temporaire ou viagère) est préconisé pour garantir à la victime un revenu régulier face aux frais futurs.

La rente servie est protégée de l'inflation par l'intermédiaire du FGAO<sup>32</sup>, auprès duquel, les assureurs se financent pour les revalorisations de rentes indexées sur le taux d'inflation de la Sécurité Sociale.

*Cette mission de revalorisation pourrait être transférée aux assureurs dans l'avenir. Pour les rentes servies et les rentes évaluées, l'assureur attendra alors des provisions FGAO adaptées à ses risques. Pour les rentes futures, l'impact sera un calcul actuariel de ce surcoût de sinistre à intégrer dans la tarification.*

On a vu en II-C.1 que l'aggravation économique peut être activée quand la victime devient perdante face à des évolutions financières extérieures ; la juste prise en compte de l'inflation est alors primordiale.

Evolution des indices (base 100 en janvier 1992)



Après reconstitution de tables d'indices homogènes de janvier 1992 à septembre 2010, on compare les revalorisations de rente pratiquées par le FGAO et l'inflation.

Il apparaît que la victime n'est pas perdante face à l'inflation dans le système actuel.

Pour respecter le principe de réparation du préjudice sans enrichissement de l'assuré, il serait souhaitable qu'en cas d'amélioration de l'état de santé de la victime, on puisse alors adapter le montant de la rente à la baisse. La rente est donc la modalité la plus souple qu'il faut préférer.

*Mais il persiste encore un paradoxe : alors qu'elle est promue par la société, la rente reste généralement soumise à l'impôt sur le revenu (incohérence relevée dans le livre blanc).*

## Le projet de vie

L'accident casse une dynamique personnelle que l'indemnisation ne suffit pas à relancer. La réparation en nature passe par l'élaboration d'un véritable projet de vie. Il est de l'intérêt de la victime ou de sa famille de saisir l'opportunité de bâtir ce projet avec toute une équipe : avocat, ergothérapeute et médecin réadaptateur de l'assureur, architecte, ... Pour promouvoir cela, Covéa s'est doté d'une mission spécifique « projet de vie ».

<sup>32</sup> Fond de garantie de l'assurance obligatoire, créé en 1951 pour assurer une mission d'intérêt général : indemniser les victimes d'accidents de la circulation dont les auteurs n'étaient pas assurés ou pas identifiés. Au fil des années, ses compétences ont été étendues et il est devenu en 2003 le FGAO. Il est financé par les assurés et les assureurs.

## La maison des 4

C'est une innovation Covéa. L'idée repose sur la mutualisation des heures de tierce personne auprès de 4 victimes ; c'est de l'indemnisation « in concreto ». Le placement d'une victime dans une maison des 4 reste cependant difficile car la décision d'un tel placement est une combinaison très complexe entre le choix de la famille de la victime, de son avocat, de la distance géographique, de l'avis des experts médicaux,...

A noter que cette bonne idée qui réduit les coûts du sinistre et donc la prime mutualisée de tous les assurés, est aujourd'hui critiquée par certains avocats payés en pourcentage de l'indemnisation obtenue.

## Outils

Le tableau ci-dessous présente les 3 principaux dispositifs de pilotage des victimes existants :

	Objectif	Moyen	Avantages	Inconvénients
Agira	Piloter les dossiers corporels transmis à l'Agira	Fichier infocentre basé sur le SI sinistres MMA	Fonctionne depuis 2006	Pas d'historisation Pas de vision analogue avec les autres SI sinistres
Rentes potentielles	Calculer l'impact sur le portefeuille d'un changement de taux ou de table	Outil Excel	Rapidité de la mesure d'impact	Echantillon restreint aux rentes
Base Victimes (BVCI)	Piloter les dossiers les plus coûteux	Logiciel ad hoc	Informations très détaillées	Reprise d'historique limitée à 2007 En cours de développement, pas d'accès à l'information possible avant 2012

Enfin, pour communiquer les évaluations des sinistres majeurs corporels et matériels, il existe un support à destination de l'interne (directions techniques, comptabilité) et des réassureurs.

## Conclusion

La finalité recherchée de ce chapitre était de s'immerger dans l'environnement de l'indemnisation des corporels. Avant de « partir en modélisation », il nous semblait en effet nécessaire de mieux comprendre le cadre juridique, le risque routier et la gestion des sinistres ainsi que la dynamique de leurs évolutions respectives.

L'étude juridique nous a permis de mesurer l'élargissement progressif du périmètre de l'assurance RC automobile ; tant l'enrichissement passé que celui qui s'annonce par les projets de lois. Les conséquents impacts financiers attendus viennent confirmer tout le soin que les actuaires doivent continuer à apporter à la mesure de ce risque.

L'évolution favorable du risque routier contrebalance en partie le surenchérissement juridique. Elle est cependant atténuée par la persistance de victimes graves représentant un coût très élevé, lié surtout à l'augmentation du poste de préjudice de tierce personne.

La gestion des sinistres, à l'instar de celle des matériels auto, montre une dichotomie entre les sinistres de masse et les graves. Les premiers permettent une industrialisation via l'IRCA qui, in fine, bénéficie autant à l'assuré (relation unique avec son propre assureur, indemnisation barémée plus rapide) qu'à l'assureur (diminution des frais de gestion, meilleure lisibilité de la charge). Pour les sinistres graves, afin de maîtriser le coût, c'est le domaine de l'expertise qui prime d'un bout à l'autre de la chaîne de gestion : gestionnaire, inspecteur, réseaux (avocats, médecins,...).

Dans un avenir proche et pour renforcer la maîtrise de l'activité, nous relevons deux évolutions à mener :

- ❑ affiner les forfaits d'ouverture pour tenir compte de la convention, en scindant les évaluations IRCA (coût moyen IRCA) des autres (coût moyen hors IRCA).
- ❑ rééquilibrer, en cas de rente, le sort de l'assureur qui subit les cas d'aggravation mais pas d'amélioration de la santé de la victime alors même que la médecine progresse régulièrement.

Des éclairages détaillés ayant été apportés sur les principaux éléments de contexte, nous pouvons désormais nous engager plus avant vers la création de modèles de mesure d'une charge ultime RC corporelle nette.

Les éléments de contexte étant appréhendés, nous dressons alors les protocoles d'élaboration du modèle pour chaque méthode. Deux éléments sont vite repérés et nous semblent suffisamment conséquents pour nécessiter une étude ad hoc préalable : la taille de la base sinistres permettant d'utiliser la méthode 2 et la valeur de seuil définissant un sinistre corporel grave.

Nous choisissons donc d'aborder dès à présent ces deux sujets. En effet, le résultat du premier point conditionne l'application de la méthode 2 et le second est un préalable transverse aux deux méthodes.

### **III - ÉTAPE 1 : L'ANALYSE DE BASES VICTIMES**

#### **III - A DEMARCHE DE L'ETUDE**

---

##### **Les 2 objectifs de l'étude**

Nous cherchons dans ce chapitre à :

- mener l'analyse descriptive de bases victimes, élément peu fréquent dans la littérature assurance, avec néanmoins de sérieuses limites que nous relèverons.
- obtenir la confirmation statistique qu'il n'y a pas de différence entre les sinistres RC corporelle issus de l'assurance automobile grand public et ceux issus des flottes pour un type de véhicule donné.

Nous souhaitons en effet montrer ce résultat pour pouvoir appliquer la méthode 2 nécessitant un grand volume de sinistres. Pour cela, nous chercherons donc à identifier la variable principale (celle qui explique le plus le coût d'un sinistre) sur laquelle nous établirons un test d'homogénéité.

##### **Qu'est-ce que c'est ?**

La première base victimes étudiée est celle que Covéa AIS envoie à l'Agira (cf. II-B.3).

Cette base recense les victimes de dommages corporels causés essentiellement par la circulation. Ces informations sont issues d'un SI de gestion des sinistres utilisé par MMA et Covéa Fleet. Cette base contient environ 62% des victimes Flottes. C'est la seule vision « victime » dont on puisse disposer en interne en 2011. La seconde, celle du marché, sera présentée plus loin.

##### **A quoi sert cette base ?**

Son rôle est principalement de fournir du pilotage à l'encadrement des équipes de gestion des sinistres, notamment sur le nombre de dossiers envoyés à l'Agira. Elle a été créée en 2005 à partir des sinistres ouverts des exercices antérieurs et courant et elle a intégré depuis toutes les victimes connues.

##### **Comment est-elle structurée ?**

Cette base est composée de 5 parties totalisant 139 variables :

1. gestion globale (11 variables),
2. description de la personne physique (16 variables),
3. suivi des expertises médicales (27 variables),
4. données Agira (25 variables),
5. suivi de l'évaluation (60 variables).

Cette base est mise à jour mensuellement ; les données sont remplacées, il n'y a pas de gestion d'un historique.

##### **Constitution de la base**

La base de départ a été enrichie de nombreuses informations comme l'identification du type de véhicule impliqué dans l'accident ou l'actualisation des montants de sinistres. De même, un grand nombre de filtres a été défini pour éliminer les victimes hors périmètre : celles ayant subi un dommage corporel sans lien avec la circulation routière ou celles dont le sinistre est toujours ouvert ou pour circonscrire l'étude aux victimes présentes dans des véhicules particuliers.

### Etude qualitative des victimes

Devant la grande quantité de variables de cette base, nous optons d'aborder l'étude par univers i.e. par regroupement de variables liées à un environnement particulier. Nous en décelons 3 dans la base : « médical », « agira » et « assureur ». Celui qui est le plus fiable après échanges avec les équipes AIS est « agira ». Nous choisissons donc cette partie de la base pour notre étude. Nous veillons à lui ajouter un socle de variables plus classiques liées à la victime (âge, sexe) et à son environnement (nature juridique, nature du contrat, coût de la part RC matérielle).

Dans cette partie, nous allons donc étudier les caractéristiques des victimes particuliers auto ou flottes (CF), observées sur une période représentative (de 2003 à 2008). En fonction des données disponibles, 3 angles ont été privilégiés : AIPP, lésions-séquelles, préjudices.

	Nombre		Montant	
	CF	Auto	CF	Auto
Blessés légers sans AIPP	85%	86%	39%	37%
Blessés avec AIPP	14%	13%	53%	51%
dont : 1 à 5 %	10%	10%	18%	18%
dont : 6 à 20 %	3%	3%	23%	20%
dont : > 20%	0%	0%	13%	13%
Décès	1%	1%	8%	12%

Ventilation des victimes et de leur indemnisation selon leur état médical

La ventilation en nombre est quasi-identique.

Celle en montant diffère : les victimes flottes sont plus indemnisées dans la tranche [6%;20%]. Cette tranche est large ce qui induit des évaluations disparates et le nombre est faible. Idem pour les décès rendant peu significatif l'écart sur les montants.

	Blessés sans AIPP			Blessés avec AIPP			Décédés		
	Marché	CF	Auto	Marché	CF	Auto	Marché	CF	Auto
Moins de 20 ans	19%	14%	21%	11%	8%	12%	19%	11%	17%
20 à 29 ans	24%	23%	25%	18%	17%	20%	17%	11%	21%
30 à 39 ans	19%	20%	17%	17%	20%	17%	9%	14%	10%
40 à 49 ans	15%	18%	14%	18%	17%	16%	10%	19%	9%
50 à 59 ans	12%	14%	11%	16%	18%	15%	9%	19%	13%
60 ans et plus	11%	10%	11%	21%	19%	20%	36%	27%	30%

Ventilation des victimes par tranche d'âge selon leur état médical

Globalement peu d'écarts sauf pour les décédés moins nombreux et donc plus volatils. Les victimes flottes sont plus souvent âgées entre 30 et 60 ans donc plus de personnes actives (impact sur le préjudice professionnel ?). Les victimes auto suivent les proportions du marché.

Nbre de lésions	Blessés sans AIPP		Blessés avec AIPP		% du total	
	CF	Auto	CF	Auto	CF	Auto
Aucune	95,2%	94,8%	69,1%	67,9%	91,4%	91,1%
1 lésion	4,7%	4,8%	25,2%	26,3%	7,7%	7,7%
2 lésions	0,1%	0,3%	3,5%	4,2%	0,6%	0,9%
3 lésions	0,0%	0,1%	2,1%	1,7%	0,3%	0,3%

Ventilation des victimes par nombre de lésions



Nbre de séquelles	Blessés sans AIPP		Blessés avec AIPP		% du total	
	CF	Auto	CF	Auto	CF	Auto
Aucune	95,1%	94,7%	68,4%	67,2%	91,2%	90,9%
1 séquelle	3,4%	3,4%	20,2%	20,2%	5,9%	5,7%
2 séquelles	1,0%	1,1%	4,9%	6,6%	1,6%	1,9%
3 séquelles	0,4%	0,8%	6,5%	6,0%	1,3%	1,5%

Ventilation des victimes par nombre de séquelles

La répartition par nombre de lésions est quasi la même entre l'auto et les flottes. Tant que les volumes restent importants (moins de 2 lésions), avec ou sans AIPP, les résultats sont très proches. Ils deviennent différents quand il y a 2 ou 3 lésions mais les volumes ne sont alors plus significatifs.

Analyse identique. On remarque plus de victimes avec plus de séquelles à la consolidation que lors de l'évaluation des lésions : des lésions bien saisies en évaluation ?

	% du total en nombre	
	CF	Auto
Lésions générales	9%	5%
Crâne, face, cou, système nerveux central	14%	15%
Membre supérieur et ceinture scapulaire	28%	29%
Membre inférieur et ceinture pelvienne	16%	16%
Thorax, cœur, poumons	22%	20%
Abdomen	6%	8%
Appareil genito urinaire	1%	1%
Lésions endocrines	0%	0%
Pas de lésion ou plus de 4 lésions	4%	6%

Ventilation des lésions selon la nomenclature Agira

	% du total en nombre	
	CF	Auto
Séquelles générales	34%	31%
Crâne, face, cou, système nerveux central	6%	6%
Membre supérieur et ceinture scapulaire	25%	24%
Membre inférieur et ceinture pelvienne	10%	12%
Thorax, cœur, poumons	17%	16%
Abdomen	2%	3%
Appareil genito urinaire	0%	0%
Séquelles endocrines	0%	0%
Pas de séquelles ou plus de 4 séquelles	7%	7%

Ventilation des séquelles selon la nomenclature Agira

	CF	Auto
dépenses de santé actuelles et futures	30%	27%
perte de gains professionnels actuels	14%	16%
déficit fonctionnel permanent	14%	14%
préjudices personnels	18%	18%
préjudices d'affection et d'accompagnement	8%	12%
autres préjudices	15%	13%

Ventilation de 100 € d'indemnités en 2008 - toutes victimes

	CF	Auto
dépenses de santé actuelles et futures	39%	30%
perte de gains professionnels actuels	17%	13%
préjudices personnels	35%	41%
autres préjudices	9%	16%

Ventilation de 100 € d'indemnités en 2008 – victimes sans AIPP

	CF	Auto
déficit fonctionnel permanent	24%	23%
préjudices personnels	18%	18%
perte de gains professionnels actuels	19%	22%
dépenses de santé actuelles et futures	39%	34%
assistance par tierce personne	0%	3%
autres indemnités	0%	0%

Ventilation de 100 € d'indemnités en 2008 – victimes avec AIPP

	CF	Auto
frais d'obsèques	4%	6%
préjudices d'affection et d'accompagnement	46%	64%
perte de revenus des proches	50%	30%

Ventilation de 100 € d'indemnités – victimes décédées

Les lésions constatées sont distribuées de manière très proche.

Dans 2 cas sur 3 quand une lésion est identifiée, c'est le haut du corps qui est touché (membre supérieur + thorax + crâne ~ 65%).

Les séquelles sont 5 à 6 fois moins détaillées au moment de l'indemnisation que lors de l'évaluation. Un grand nombre de lésions évoluent vers des séquelles générales.

Les proportions sont très proches entre l'auto et les flottes.

Quelle que soit la gravité des dommages de la victime (avec ou sans AIPP), les ventilations par poste de préjudices sont très proches entre l'auto et les flottes.

Les victimes flottes sans AIPP étant plus souvent des actives (cf. leur âge), leur perte de gains professionnels est plus importante.

On ne retrouve pas le résultat précédent observé sur seulement 14% des victimes. Au vu du poids nul de TP, on voit que peu de contrats graves sont saisis dans cette base Agira.

On voit une part importante de la perte de revenus des proches : lien de nouveau avec le fait que les victimes flottes sont davantage dans une tranche de personnes actives.

## Commentaires

A travers les critères d'AIPP, d'âge, de nombre et de typologie de lésions / séquelles ou de ventilation de l'indemnité par poste de préjudices, on remarque une très forte proximité de résultats entre les caractéristiques d'une victime flottes et auto. On en tire une **présomption de ressemblances** que l'on va étayer statistiquement.

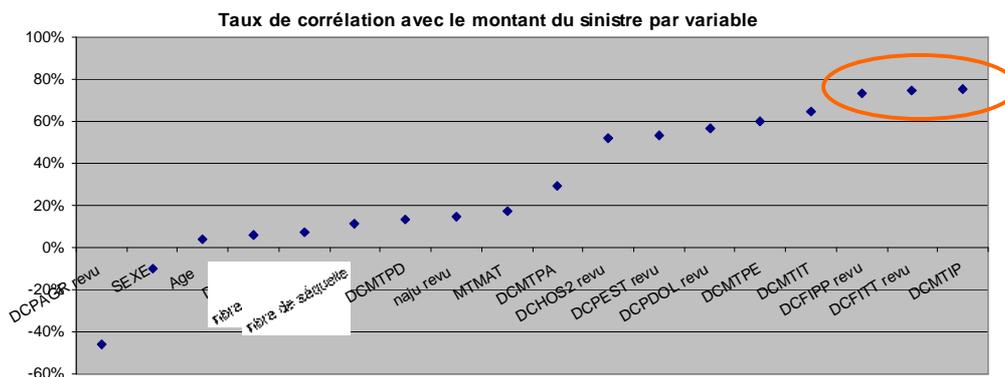
Des biais importants existent cependant, rendant difficile la comparaison avec des données marché et incitant à relativiser les résultats : faible nombre des victimes envoyées à l'Agira (pour mémoire, les mieux documentées) et pas d'alimentation des graves dans la base. Les données sur les victimes décédées étant peu nombreuses, nous choisissons de les écarter par la suite.

## III - B APPLICATIONS

La base étudiée contient environ 4 500 victimes. On l'analysera par les outils suivants :

1. **La matrice de corrélation** permettant de détecter rapidement certaines liaisons, sera faite en amont.
2. Le nombre de variables étant important, les interprétations devenant difficiles, on se tournera vers un type d'analyse factorielle : **l'analyse en composante principale**.
3. **La régression linéaire multiple pas à pas** : on injectera par étape les variables explicatives dans la régression qui les conservera tant qu'elles seront significatives au risque de faire erreur à  $\alpha = 15\%$ .

### La matrice de corrélation



#### Libellé des variables :

DCMTIP : montant d'AIPP

DCFITT revu : nombre de jours ITT

DCFIPP revu : taux d'AIPP

DCMTIT : montant d'ITT

DCMTPE : montant du préj. esthétique

DCDOL revu : taux du praetium doloris

DCPEST revu : taux de préjudice esthétique

DCHOS2 revu : jours d'hospitalisation

DCMTPA : montant du préj. d'agrément

MTMAT : montant du coût matériel

Naju revu : nature de l'indemnisation

DCMTPD : montant praetium doloris

Nbre de séquelle : nombre de séquelle

Nbre de lésion : nombre de lésion

DCFMP : frais médicaux /pharmaceutiques

DCPAGR revu : taux du préjudice d'agrément

Les 3 variables les plus corrélées positivement avec le coût sont par ordre décroissant : le montant d'AIPP, le nombre de jours d'ITT et le taux d'AIPP. Pour la suite, afin d'éviter la colinéarité, la question se pose de retenir le « taux », le « montant » ou les 2 variables. On choisit de retenir le « taux » car cette variable est la plus indépendante ; elle n'est pas polluée par l'utilisation d'un barème comme peut l'être le « montant ». Du coup, les 2 variables corrélées à plus de 70 % avec le coût sont le nombre de jours d'ITT (75%) et le taux d'AIPP (74%).

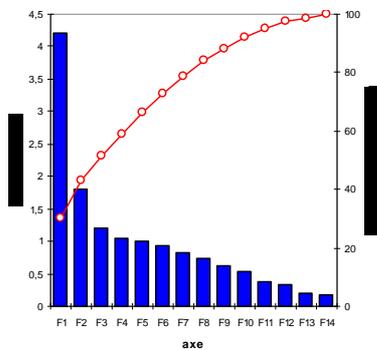
La matrice de corrélation est obtenue par Xlstat ; on vérifie que son rang est bien égal au nombre de variables donc il n'y a pas de détection de colinéarité entre les variables. Le rang est calculé via l'add-in Excel Matrix et la fonction MRANK.

**Analyse en Composantes Principales :**

On définit le nombre d’axes expliquant le maximum d’inertie du système par l’étude des valeurs propres (> 1, critère de Kaiser) et par l’inflexion dans la courbe de l’inertie expliquée cumulée (critère du coude). Si p et n sont le nombre de variables et d’individus, on identifiera les variables contribuant le plus à ces axes (contribution > 1/p ; variables proches du bord du cercle) et les individus contribuant le plus à ces axes (contribution > 1/n) en retirant éventuellement les individus trop contributifs (contribution > 25%).

La variance de chaque variable valant 1 après standardisation, le critère de Kaiser retient les facteurs dont la valeur propre ≥ 1 i.e. ceux qui expliquent plus de variance qu'une variable seule. On retient ainsi les 5 premières valeurs propres qui expliquent 66,1% de l’inertie du nuage.

	F1	F2	F3	F4	F5	F6	F7	F8
Valeur propre	4,2	1,8	1,2	1,0	1,0	0,9	0,8	0,7
Variabilité (%)	30,0	12,9	8,6	7,5	7,1	6,7	5,9	5,3
%cumulé	30,0	42,9	51,5	59,0	66,1	72,8	78,7	83,9

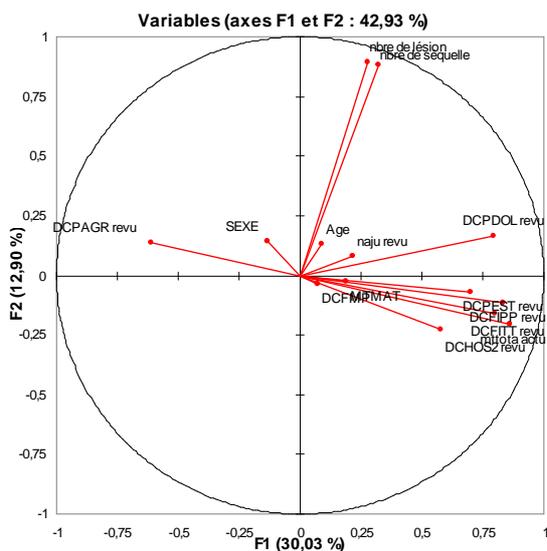


Le critère du coude (scree-test) est défini par l'observation du graphique des valeurs propres : on ne retient que les valeurs qui se trouvent à gauche du point d’inflexion.

On sélectionne ainsi les 3 premières valeurs propres et donc 51,5% de l’inertie cumulée expliquée. C’est trop peu donc on retient le critère de Kaiser. La contrepartie de ce choix est le nombre important de facteurs à appréhender.

Ces 5 facteurs indépendants, notés F<sub>i</sub>, sont des combinaisons linéaires des variables et expliquent plus particulièrement certaines d’entre elles.

F<sub>1</sub> et F<sub>2</sub> expliquent 42,9% de l’inertie du nuage de points et les variables projetées sur (F<sub>1</sub>, F<sub>2</sub>) sont représentées ainsi :



L’analyse des corrélations entre variables et facteurs donne :

F<sub>1</sub> retient les variables médicales analysées et liées à l’enjeu financier : coût du sinistre, nombre de jours d’ITT, taux d’AIPP et praetium doloris...

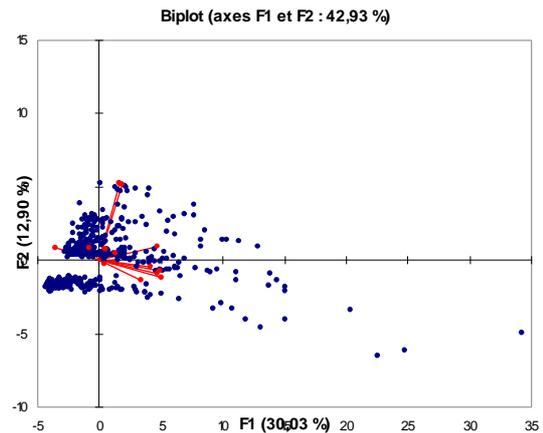
F<sub>2</sub> privilégie les 2 variables médicales objectives brutes : nombre de lésions et nombre de séquelles.

F<sub>3</sub> caractérise la personne victime : âge et sexe.

F<sub>4</sub> surpondère les variables montant matériel et mode de règlement du dossier.

Enfin F<sub>5</sub> concerne surtout les frais médicaux.

Le nuage de points des individus-victimes (ici en bleu un échantillon aléatoire de la base) n'est pas homogène.  $F_1$  et  $F_2$  séparent chacun l'échantillon des individus en 2 : pour  $F_1$  on vérifie que c'est le montant du sinistre qui joue un rôle clé (plus il est important plus on est à droite de l'axe) ; pour  $F_2$ , c'est la présence ou non de lésion / séquelle (quand il n'y en a pas, on est positif). On vérifie qu'aucune observation ne contribue trop.



**La régression linéaire multiple pas à pas (stepwise)**

La **régression linéaire multiple** est une généralisation, à  $p$  variables explicatives, de la régression linéaire simple. Etant donné un échantillon  $(Y_i, X_{i1}, \dots, X_{ip})$ ,  $i = 1, \dots, n$  nous cherchons à expliquer les valeurs prises par  $Y_i$ , dite variable endogène, à partir d'une série de variables explicatives  $X_{i1}, \dots, X_{ip}$ . Le modèle théorique, formulé en termes de variables aléatoires, prend la forme :

$$Y_i = a_0 + a_1 X_{i1} + a_2 X_{i2} + \dots + a_p X_{ip} + \varepsilon_i, i = 1, \dots, n$$

où  $\varepsilon_i$  est l'erreur du modèle et  $a_0, \dots, a_p$  les paramètres à estimer.

Lorsque nous disposons de  $n$  observations  $(y_i, x_{i1}, \dots, x_{ip})$ ,  $i = 1, \dots, n$  qui sont des réalisations des variables aléatoires  $(Y_i, X_{i1}, \dots, X_{ip})$ , l'équation de régression s'écrit :

$$y_i = a_0 + a_1 x_{i,1} + a_2 x_{i,2} + \dots + a_p x_{i,p} + \varepsilon_i, i = 1, \dots, n$$

Les équations suivantes : 
$$\begin{cases} y_1 = a_0 + a_1 x_{1,1} + a_2 x_{1,2} + \dots + a_p x_{1,p} + \varepsilon_1 \\ y_2 = a_0 + a_1 x_{2,1} + a_2 x_{2,2} + \dots + a_p x_{2,p} + \varepsilon_2 \\ \dots \\ y_n = a_0 + a_1 x_{n,1} + a_2 x_{n,2} + \dots + a_p x_{n,p} + \varepsilon_n \end{cases}$$
 peuvent être résumées avec la notation

matricielle : 
$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,p} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{pmatrix} + \begin{pmatrix} \varepsilon_0 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$
. Soit de manière compacte:  $y = Xa + \varepsilon$ .

On rappelle que de nombreuses hypothèses (stochastiques et structurelles) sont nécessaires pour valider la régression.

La régression multiple peut être utilisée à plusieurs fins :

- ❑ trouver la meilleure équation linéaire de prévision (modèle), en évaluer la précision et la signification,
- ❑ estimer la contribution **relative** de 2 ou plusieurs variables explicatives sur la variation d'une variable à expliquer.

Nous retenons principalement ici la seconde.

Quand on dispose de plusieurs variables explicatives, la technique suivante est employée pour sélectionner un nombre réduit de variables qui explique pourtant une quantité raisonnable de variation : la régression "pas à pas" (stepwise) qui consiste à faire entrer les variables l'une après l'autre dans le modèle par sélection progressive et, à chaque étape, à vérifier si celles déjà introduites sont encore significatives (une variable qui ne le serait plus serait rejetée).

On mesurera la qualité du modèle par le  $R^2$  ajusté<sup>33</sup> tenant compte du nombre de paramètres du modèle (constante et variables), la p-valeur du test de Fischer au risque d'erreur  $\alpha$  sur le rejet ou non de l'hypothèse de nullité de tous les coefficients des paramètres. Enfin, on mesurera la qualité détaillée de chaque coefficient de paramètre par la p-valeur du test de Student au risque d'erreur  $\alpha$  sur le rejet ou non de l'hypothèse de nullité de ce coefficient. On vérifiera préalablement les points suivants : nullité de la moyenne des résidus, homoscedasticité des résidus et nullité des covariances entre les résidus et chaque variable explicative.

On présente ci-dessous le résultat de la procédure stepwise obtenue dans SAS.

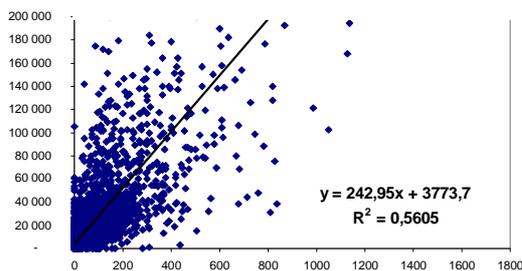
Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	DCFITT revu		DCFITT revu	1	0.5605	0.5605	2588.58	5726.12	<.0001
2	DCFIPP revu		DCFIPP revu	2	0.1253	0.6858	573.170	1790.05	<.0001
3	DCHOS2 revu		DCHOS2 revu	3	0.0199	0.7057	254.703	303.51	<.0001
4	DCPEST revu		DCPEST revu	4	0.0059	0.7116	162.146	91.36	<.0001
5	MTMAT		MTMAT	5	0.0056	0.7172	73.9599	88.84	<.0001
6	DCPAGR revu		DCPAGR revu	6	0.0018	0.7189	47.3401	28.36	<.0001
7	naju revu		naju revu	7	0.0011	0.7201	31.1157	18.13	<.0001
8	Age		Age	8	0.0008	0.7209	19.7547	13.33	0.0003
9	nbre de lésion		nbre de lésion	9	0.0006	0.7215	12.8105	8.94	0.0028
10	DCFMP		DCFMP	10	0.0003	0.7218	9.1866	5.63	0.0177

La variable « nombre de jours d'ITT » est celle qui est sélectionnée en premier ( $R^2$  ajusté<sup>33</sup> = 56,05%) **et que l'on retiendra pour bâtir le test d'homogénéité**. On rappelle que l'ITT, incapacité temporaire totale, correspond à la période pendant laquelle la victime ne pourra accomplir ou subira une gêne importante dans l'accomplissement des actes usuels de la vie (se laver, manger, s'habiller...). Un enfant ou une mère au foyer peut donc avoir de l'ITT.

Le « taux d'AIPP » arrive en second ici. On rappelle que l'AIPP est l'atteinte permanente à l'intégrité physique et/ou psychique et qu'elle concerne la réduction du potentiel physique, psychosensoriel ou intellectuel dont reste atteinte la victime.

Ces 2 variables sont objectives (robustes) et définies par un médecin. Avec ces 2 variables, on a un modèle présentant un  $R^2$  ajusté = 68,58% ; avec les 10 variables sélectionnées, on obtient légèrement plus (72,18%).

On représente ci-dessous la régression du coût du sinistre par la variable « nombre de jours d'ITT ».



Si l'on fait la régression du montant de sinistre par le taux d'AIPP, arrivé en second dans l'analyse pas à pas, on trouve un  $R^2 = 54%$  ; cela montre la proximité dans la qualité d'explication du coût entre ces 2 variables qui sont toutefois bien d'origine différente.

Par ailleurs, on remarque que la variable qualifiant l'origine du contrat de la victime (auto ou flottes), injectée dans la régression, n'est pas retenue : le coût d'un sinistre-victime s'explique donc peu/pas par ce critère, **présomption supplémentaire que les individus de ces 2 populations sont proches**.

<sup>33</sup>  $R^2$  ajusté : le coefficient de détermination ajusté tient compte du nombre de variables. En effet, le principal défaut du  $R^2$  (rapport de la variance expliquée sur la variance totale) est de croître avec le nombre de variables explicatives. Or, un excès de variables produit des modèles à forte variance donc peu robustes. C'est pourquoi on s'intéresse davantage à cet indicateur qu'au  $R^2$ . Avec  $n$  le nombre d'observations et  $k$  le nombre de variables

explicatives, on a 
$$R^2_{ajusté} = R^2 - \frac{k(1-R^2)}{(n-k-1)}$$

De la même manière, nous avons étudié la base victimes du FVI (fichier national des victimes corporelles suite à accident de la circulation et indemnisées en 2010, environ 21 000 individus). Une synthèse des résultats comparés avec ceux de la base Agira est présentée ci-dessous :

	<b>Base Agira</b>	<b>Base FVI</b>
Variables corrélées à plus de 60 % avec le coût	le nombre de jours d'ITT (75%) et le taux d'AIPP (74%)	le taux d'AIPP (64%)
ACP - Inertie du nuage de points expliquée quand le critère de Kaiser est choisi	66,1% (5 facteurs)	59,3% (7 facteurs)
Montant du R <sup>2</sup> ajusté dans le cas de la régression linéaire multiple pas à pas	72,01% (7 variables)	48,69% (11 variables)
Noms des variables sélectionnées dans l'ordre	le nombre de jours d'ITT, le taux d'AIPP, le nombre de jours d'hospitalisation, le taux de préjudice esthétique, l'indication d'un coût matériel, le taux de préjudice d'agrément la modalité de l'indemnisation	le taux d'AIPP, la durée de l'incapacité provisoire de travail l'indication si tierce personne réglée l'indication si dépenses de santé futures l'âge l'indication si préjudice esthétique le taux de préjudice esthétique le taux de souffrances endurées le format de l'indemnisation (Dintilhac/autre) l'indication si accident du travail l'indication si préjudice d'agrément

La base Agira possède des variables plus fortement corrélées avec le coût du sinistre.

Le nuage de points s'explique plus simplement et davantage avec les variables de la base Agira. Les données sont peut-être plus homogènes (on a vu que les graves n'y étaient pas/peu saisis). Cela peut aussi venir du fait que les variables présentes sont plus explicatives que celles du FVI.

Le modèle défini par l'analyse stepwise est de meilleure qualité, au sens du R<sup>2</sup> ajusté, en Agira qu'en FVI.

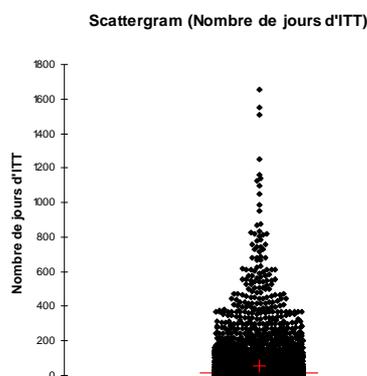
On retrouve bien les 2 mêmes variables en tête dans l'analyse stepwise. Cela vient renforcer le choix opéré de baser notre test sur l'une d'entre elles.

Cette comparaison montre aussi que les variables explicatives peuvent être nombreuses, complexes à définir et à recueillir. Les résultats détaillés de l'étude FVI figurent en annexe VIII-B.

### Test d'homogénéité du $\chi^2$

Les éléments de théorie sous jacents à ce test sont présentés en V-C.

La variable majeure identifiée par la régression pas à pas est donc le nombre de jours d'ITT.



L'étude des quartiles de la distribution de cette variable (représentée ci-contre sous forme de diagramme de dispersion) nous amène à créer les classes homogènes suivantes en nombre de jours d'ITT : 0, entre 1 et 15, entre 16 et 62 et plus de 62.

Le nombre de degrés de liberté est 3.

On échantillonne aléatoirement la population des victimes flottes et auto ; on obtient la mesure de la distance du  $\chi^2 = 6,02 < 7,81$  (valeur de la table au risque 5%).

### III - C CONCLUSION

---

L'analyse qualitative induit une forte ressemblance des populations victimes auto et flottes. C'est confirmé plusieurs fois, sous le prisme de critères différents (taux d'AIPP, âge, typologie des lésions/séquelles).

La matrice de corrélation nous a permis d'identifier les colinéarités (taux et montant).

L'analyse en composantes principales nous invite à l'humilité devant le faible taux d'inertie expliquée et devant le grand nombre de facteurs pourtant nécessaires pour arriver à ce taux.

Enfin, l'analyse de régression multiple pas à pas valide la prédominance de la variable nombre de jours d'ITT sur les autres pour expliquer le coût d'un sinistre. Comparés avec la base FVI, les résultats obtenus nous conforte sur le choix de cette base et de sa qualité.

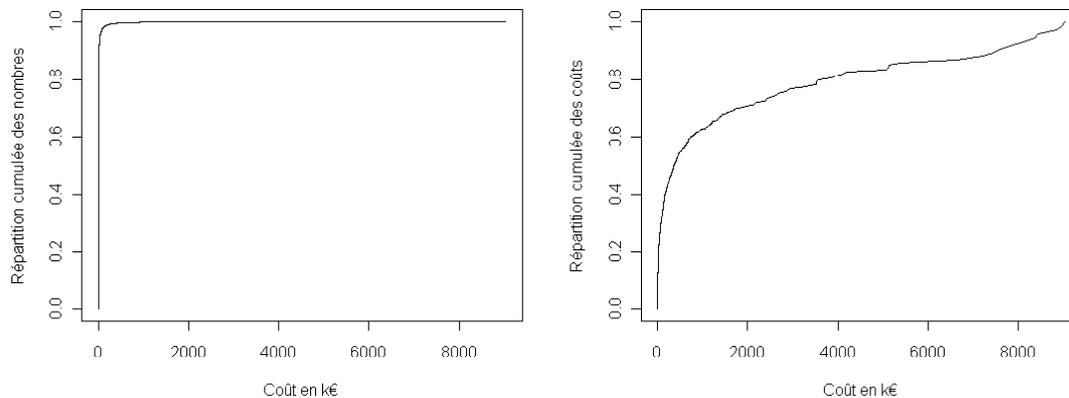
Au vu du test d'homogénéité basé sur un critère sélectif, on ne peut pas rejeter, au niveau du test à 5 %, que les coûts des sinistres auto et flottes sont issus de la même population. Donc, il ne paraît pas déraisonnable **d'assimiler ces 2 populations** ce qui nous permettra de mettre en œuvre la méthode 2. Le résultat est, somme toute, assez intuitif (« les mêmes causes provoquant les mêmes effets ») mais il nous a paru plus sérieux de l'établir scientifiquement. Les résultats de ce mémoire s'appliqueront donc aussi à un portefeuille auto lui conférant une portée plus générale.

Plus généralement, on peut remarquer qu'on étudie des êtres humains, complexes par nature. Lors d'un accident corporel, cette diversité s'exprime, la victime étant soumise à une chaîne d'interactions multiples (blessures, secours, soins, conseils, soutien,...). Des éléments comme le niveau de protection du véhicule, la qualité des intervenants médicaux, la présence d'un avocat, l'état psychologique de la victime, la qualité de son entourage, la variation des évaluations dans le temps,... pourraient aussi aider à expliquer le montant du sinistre. Autant de variables à recueillir, à tester, pour bâtir des modèles de plus en plus affinés.

Une évolution des mentalités acceptant d'investir dans le recueil d'une information de qualité permettra d'accroître la maîtrise de cette charge. Il faut que l'ensemble de la chaîne de valeurs soit experte, y compris dans le choix et la qualité des informations saisies dans l'outil. La rentabilité et la compétitivité des compagnies se joueront aussi à ce niveau. Pour cela, Covéa se dote d'une telle base qui sera disponible à l'analyse en 2012.

## IV - ETAPE 2 : CALCUL D'UN SEUIL SEPARANT LES ATTRITIONNELS DES GRAVES

Un risque peut survenir avec une probabilité faible, mais non nulle et présenter une perte immense. Pour l'illustrer, on donne ci-dessous la courbe représentant pour chaque seuil le pourcentage de sinistres de coût inférieur au seuil et la proportion de la charge totale représentée par ces sinistres.



La proportion cumulée du nombre de sinistres atteint très rapidement des valeurs proches de 1 tandis que celle des coûts croît bien plus lentement. Tous les sinistres n'ont pas la même gravité et certains sont nettement plus lourds que les autres.

On pourrait donc penser à utiliser une loi à queue très épaisse type Pareto ou log-gamma pour modéliser le coût des sinistres. Cependant, la courbe s'ajustera mal aux faibles coûts qui représentent l'essentiel des données. Il faut donc modéliser séparément les sinistres de faibles coûts (de fréquence), des sinistres graves qui n'ont pas le même comportement.

La question est alors : « à partir de quel montant un sinistre peut-il être qualifié de majeur ? »

Après avoir rappelé les fondements de la théorie des valeurs extrêmes et en particulier la méthode basée sur la distribution de Pareto généralisée, plusieurs méthodes (valeurs record, moyenne des excès, approximation GPD) seront proposées pour la détermination d'un seuil à partir duquel un événement est considéré comme atypique.

### IV - A DE LA THEORIE DES VALEURS EXTREMES A LA METHODE DE DEPASSEMENT DE SEUIL

La théorie des valeurs extrêmes ou EVT a été développée dans le but d'étudier le comportement asymptotique de la fonction de queue. Ce problème est à priori complexe car il s'agit d'étudier des phénomènes rares donc pour lesquels on dispose de peu d'information.

On peut distinguer deux parties dans la théorie des valeurs extrêmes. D'une part le théorème de Fischer-Tippett qui permet de connaître la distribution asymptotique du maximum de  $n$  variables aléatoires. D'autre part l'étude de "l'excess function" (loi des excès) qui fait apparaître la distribution de Pareto généralisée.

#### IV - A . 1. DISTRIBUTION DES EXTREMA DANS LE CAS D'UN ECHANTILLON DE TAILLE FINIE

Considérons  $X_1, \dots, X_n$ , une suite de  $n$  variables aléatoires indépendantes et identiquement distribuées (i.i.d) où  $X_i$  représente les coûts des sinistres de l' $i^{\text{ème}}$  assuré. Leur fonction de répartition  $F_X$  est définie par :

$$F(x) = P(X_i \leq x) \text{ pour } i = 1, \dots, n \quad (1)$$

Pour étudier le comportement extrême des événements, on considère la variable aléatoire  $M_n = \max \{ X_1, \dots, X_n \}$  où  $M_n$  représente le plus important sinistre survenu.

Comme les variables aléatoires sont i.i.d., la fonction de répartition de  $M_n$  est donnée par :

$$F_{M_n}(x) = P(M_n \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = [F(x)]^n \quad (2)$$

La formule (2) présente un intérêt très limité. En effet, notons  $x_F = \sup\{x \in \mathfrak{R} / F(x) < 1\}$  le point terminal à droite de la fonction de répartition F.

Pour tout  $x < x_F$ ,  $\lim_{n \rightarrow \infty} (F(x))^n = 0$  ; et si  $x_F < \infty$  alors pour tout  $x > x_F$ ,  $\lim_{n \rightarrow \infty} (F(x))^n = 1$

Donc la distribution asymptotique du maximum, déterminée en faisant tendre n vers l'infini, donne une loi dégénérée. De plus, la loi d'une variable aléatoire parente X est rarement connue avec précision et, même si la loi de cette variable parente X est connue avec exactitude, la loi du terme maximum n'est pas toujours facilement calculable.

Pour ces raisons, il est intéressant de considérer les comportements asymptotiques du maximum convenablement normalisé.

Un des résultats fondamentaux de la EVT est le théorème suivant établi en 1928 par Fisher et Tippett.

#### IV - A . 2. DISTRIBUTION ASYMPTOTIQUE DU MAXIMUM

**Définition** : lois de même type

On dit que deux variables aléatoires X et Y sont **de même type** s'il existe deux constantes  $a \in \mathfrak{R}$  et  $b > 0$  telles que :  $X \stackrel{loi}{=} aY + b$ .

Autrement dit, les variables de même type ont la même loi, à un facteur de localisation et d'échelle près.

On dit aussi que F appartient au **domaine d'attraction du maximum** de G que l'on note MDA(G) (Maximum Domain of Attraction).

De façon analogue au théorème central limite, peut-on trouver des constantes de normalisation  $a_n$  et  $b_n$  avec  $a_n \in \mathfrak{R}$  et  $b_n > 0$  et une loi non-dégénérée de fonction de répartition G telle que :

$$P \left[ \frac{M_n - a_n}{b_n} \leq x \right] = F(a_n + b_n x)^n \rightarrow G(x) ?$$

Fisher et Tippett trouvent en 1928 une solution à ce problème au moyen d'un théorème<sup>34</sup> qui porte leur nom et qui est l'un des fondements de la théorie des valeurs extrêmes.

**Théorème de Pickands Fisher et Tippett** (ou théorème des 3 types extrêmes)

S'il existe deux suites de constantes de normalisation avec  $a_n \in \mathfrak{R}$  et  $b_n > 0$  et une loi non-dégénérée de fonction de répartition G telle que  $\frac{M_n - a_n}{b_n} \stackrel{loi}{\rightarrow} G$ , alors G est du même type qu'une des trois lois suivantes,

données par leur fonction de répartition standard ( $\mu=0$  et  $\sigma=1$ ) :

Type 1 : Gumbel       $\Lambda(x) = \exp(-\exp(-x)) \quad \forall x \in \mathfrak{R}$

Type 2 : Fréchet

$$\Phi_\alpha(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \exp(-x^{-\alpha}) & \text{si } x > 0, \alpha > 0 \end{cases}$$

<sup>34</sup> Pour la preuve, nous renvoyons à Embrechts et al., 1997

Type 3 : Weibull

$$\Psi_{\alpha}(x) = \begin{cases} 1 & \text{si } x > 0 \\ \exp(-(-x)^{-\alpha}) & \text{si } x \leq 0, \alpha < 0 \end{cases}$$

Ce théorème est l'analogie du théorème central limite (TCL), qui s'intéresse à la loi asymptotique de la somme : alors que le TCL fait apparaître une loi normale, le théorème de Fischer-Tippett met en exergue la **distribution GEV** (Generalized Extreme Value) à la différence près que trois types de loi limite sont possibles.

Jenkinson en a proposé une écriture unifiée :

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

On a les correspondances suivantes:

- Le cas  $\xi = 0$  correspond au type 1 Gumbel.
- Le cas  $\xi > 0$  correspond au type 2 Fréchet de paramètre  $\alpha = \frac{1}{\xi}$ .
- Le cas  $\xi < 0$  correspond au type 3 Weibull de paramètre  $\alpha = -\frac{1}{\xi}$ .

Les paramètres  $\mu$  et  $\sigma$  sont respectivement les limites de  $a_n$  et  $b_n$ .

$\sigma$  joue le rôle d'une variance, c'est donc un paramètre de dispersion tandis que  $\mu$  est un paramètre de localisation.

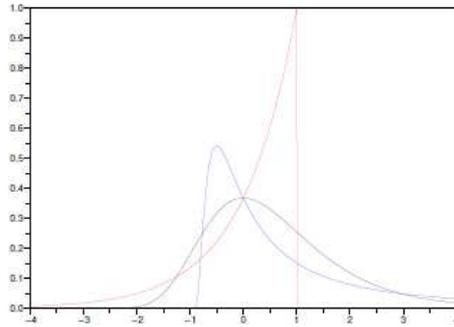
Le paramètre  $\xi$  n'a pas d'équivalent dans le TCL, il contrôle la «lourdeur» de la queue, c'est pourquoi on lui donne généralement le nom d'indice de queue ou d'indice de valeur extrême. Plus cet indice est élevé en valeur absolue, plus le poids des extrêmes dans la distribution initiale est important.

A chaque domaine d'attraction correspondent des lois caractérisées par leur queue. Elles sont indiquées dans le tableau ci-dessous :

Domaine d'attraction	Gumbel $\xi = 0$	Fréchet $\xi > 0$	Weibull $\xi < 0$
Loi	queue fine ou moyenne	queue épaisse	queue finie
Exemples	Normale Exponentielle Log-normale Gamma Weibull	Cauchy Pareto Student	Uniforme Beta

L'épaisseur de la queue de la fonction de répartition est résumée par le paramètre  $\xi$ . Ainsi les lois appartenant au domaine d'attraction de la loi de Fréchet ( $\xi > 0$ ) voient leur queue décroître en fonction puissance tandis que celles appartenant au domaine d'attraction de Gumbel ( $\xi = 0$ ) ont des queues qui décroissent de manière exponentielle.

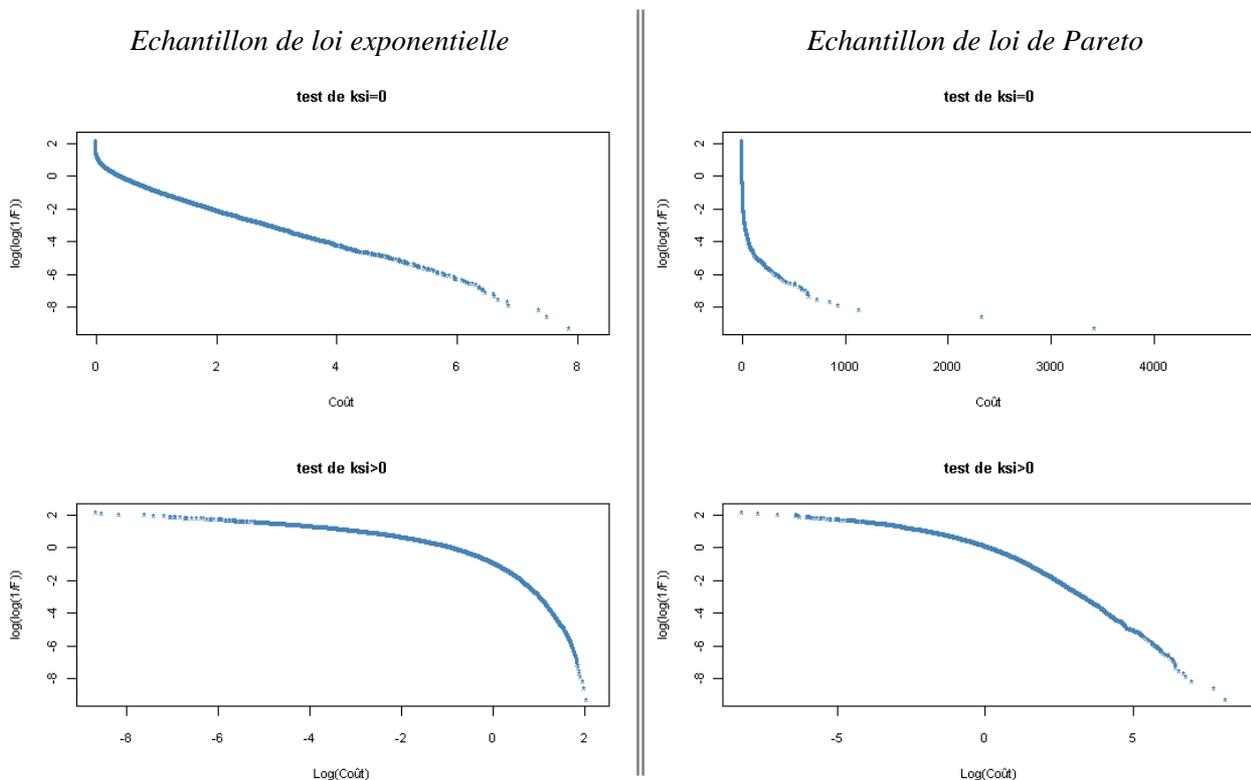
Exemples de densités associées à la loi des valeurs extrêmes ( $\xi = 0$ ,  $\xi = 1$  et  $\xi = -1$ ) :



Afin de savoir si nous sommes dans le cas Gumbel ( $\xi = 0$ ) ou dans le cas Fréchet ( $\xi > 0$ ), nous pouvons appliquer la méthode suivante :

- Nous avons vu précédemment que la fonction de répartition de la loi de Gumbel est :  $\Lambda(x) = \exp(-\exp(-(x-\mu)/\sigma)) \quad \forall x \in \mathbb{R}$ .  
Autrement dit,  $\ln(\ln(1/\Lambda(x)))$  doit être une fonction linéaire en  $x$ . Pour savoir si on est dans le cas  $\xi = 0$ , on regarde donc si les points sont alignés ou non.
- De la même manière, la fonction de répartition de la loi de Fréchet est  $\Phi_\alpha(x) = \exp(-x^{-\alpha})$  si  $x > 0$ ,  $\alpha > 0$ .  
Autrement dit,  $\ln(\ln(1/\Phi_\alpha(x)))$  doit être une fonction linéaire en  $\ln(x)$ . Pour savoir si on est dans le cas  $\xi > 0$ , on regarde donc si les points sont alignés ou non.

Nous indiquons ci-dessous les résultats obtenus pour un échantillon de loi exponentielle et pour un échantillon de loi de Pareto :



Sur les graphiques obtenus nous n'avons pas nettement une droite affine décroissante. Cependant pour la loi exponentielle, la fin du premier graphe est une droite alors que le second est convexe, nous validons bien qu'elle appartient au domaine  $\xi = 0$ . Pour la loi de Pareto, le premier graphe est concave, alors que la fin du deuxième est une droite, nous validons bien qu'elle appartient au domaine  $\xi > 0$ .

L'approche basée sur la GEV a été critiquée dans la mesure où l'utilisation d'un seul maximum conduit à une perte d'information contenue dans les autres grandes valeurs de l'échantillon. Pour pallier à ce problème, la méthode POT (Peaks-over-Threshold) ou méthode des excès au-delà d'un seuil élevé a été introduite par Pickands (1975).

#### IV - A . 3. LOI DES EXCES ET DISTRIBUTION GPD (GENERALIZED PARETO DISTRIBUTION)

Ce second volet de la théorie des valeurs extrêmes, appelée méthode à dépassement de seuil ou méthode POT (Peak-Over-Threshold ou méthode des excès) est une méthode d'estimation des queues de distribution. L'approche basée sur la loi de Pareto généralisée a été introduite par Pickands (1975) et reprise par de Haan et Rootzen (1993).

La méthode consiste à observer non pas le maximum ou les plus grandes valeurs mais toutes les valeurs des réalisations qui excèdent un certain seuil élevé. L'idée de base de cette approche consiste à choisir un seuil suffisamment élevé et à étudier les excès au-delà de ce seuil.

Cette méthode peut être définie comme suit :

Soit  $u$  un réel « suffisamment grand » et inférieur à la valeur maximale, appelé seuil.

La **méthode POT** s'appuie sur une approximation de la loi des excès au-dessus du seuil  $u$  de la variable aléatoire réelle  $X$ , ce qui correspond à la loi conditionnelle de la variable aléatoire  $X-u$  sachant que  $X > u$ .

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

$$F_u(y) = P(X - u \leq y / X > u) = \begin{cases} \frac{F(u+y) - F(u)}{1 - F(u)} & \text{si } y \geq 0 \\ 0 & \text{si } y < 0 \end{cases}$$

Le théorème de Pickands-Balkema-de Haan ci-après donne la forme de la loi limite pour les valeurs extrêmes : sous certaines conditions de convergence, la loi limite est une loi de Pareto généralisée que l'on notera GPD.

##### Théorème de Pickands-Balkema-de Haan (1975).

La fonction de répartition  $F$  appartient au **max-domaine d'attraction** de  $G_\xi$  si et seulement si, il existe une fonction positive  $\sigma(u)$  et un réel  $\xi$  tels que :

$$\lim_{u \rightarrow x_F} \sup_{0 < y < x_F} \left\{ \left| F_u(y) - G_{\xi, \sigma(u)}(y) \right| \right\} = 0$$

où  $F_u(y)$  est la fonction de répartition conditionnelle des excès pour  $u$  élevé,  $x_F$  est le point terminal de  $F$ ,  $x_F = \sup \{x \in \mathfrak{X} : F(x) < 1\}$  et  $G_{\xi, \sigma(u)}(y)$  est la fonction de Pareto généralisée (GPD, Generalized Pareto Distribution) définie par :

$$G_{\xi, \sigma(u)}(y) = \begin{cases} 1 - \left( 1 + \xi \frac{y}{\sigma(u)} \right)^{-1/\xi} & \text{si } \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma(u)}\right) & \text{si } \xi = 0 \end{cases} \quad \text{pour } y \geq 0 \text{ si } \xi \geq 0 \text{ et } 0 \leq y \leq \frac{-\sigma(u)}{\xi} \text{ si } \xi < 0.$$

Cette loi sera également notée dans la suite **GPD( $\xi, \sigma$ )**.

Ce résultat montre que l'indice de queue  $\xi$  intervenant dans la limite du maximum est identique au paramètre  $\xi$  de la loi GPD décrivant le comportement des dépassements d'un niveau  $u$  suffisamment élevé. Ceci permet de montrer l'existence d'une relation étroite entre max-domaine d'attraction et le comportement limite de la loi des excès.

D'autre part, pour la GPD, le cas  $\xi > 0$  correspond aux lois à queues épaisses, pour lesquelles 1-G se comporte comme une puissance  $y^{-\frac{1}{\xi}}$  quand  $y$  est assez élevé.

Si  $\xi = 0$ , on trouve  $1 - \exp\left(-\frac{y}{\sigma(u)}\right)$ , c'est la loi exponentielle de paramètre  $\sigma$ .

Si  $\xi < 0$  c'est la loi de Pareto de type II à support borné. Il est intéressant de noter que le cas  $\xi = -1$  correspond à la loi uniforme sur  $[0, \sigma]$ .

Ces deux façons de caractériser l'appartenance à un domaine d'attraction (approches GEV et POT) ont en commun le problème de l'estimation de l'indice des valeurs extrêmes.

#### IV - A . 4. ESTIMATION DE L'INDICE DE QUEUE

On note diverses méthodes pour estimer l'indice de queue  $\xi$ , notamment en passant par l'estimation des paramètres des lois GEV et GPD.

##### Estimation des paramètres de la loi GEV

L'idée proposée par Gumbel<sup>35</sup> est de considérer des **maxima par blocs** : on subdivise l'échantillon de taille  $n$  en  $m$  blocs de taille  $r = n/m$  :

$$\underbrace{X_1, \dots, X_r}_{1^{er} \text{ bloc}}, \underbrace{X_{r+1}, \dots, X_{2r}}_{2^{ème} \text{ bloc}}, \dots, \underbrace{X_{(i-1)r+1}, \dots, X_{ir}}_{i^{ème} \text{ bloc}}, \dots, \underbrace{X_{(m-1)r}, \dots, X_{mr}}_{m^{ième} \text{ bloc}}$$

On note alors  $Y_i$  le maximum obtenu sur le  $i^{ème}$  bloc :  $Y_i = \text{Max}(X_{(i-1)r+1}, \dots, X_{ir})$ ,  $i=1, \dots, m$ .

En particulier, si  $\xi \neq 0$ , la log-vraisemblance de l'échantillon  $Y_1, \dots, Y_m$  distribué selon une loi GEV( $\xi, \mu, \sigma$ ) est la suivante :

$$\text{Log}L = -m \log \sigma - (1 - \xi^{-1}) \sum_{i=1}^m \log \left( 1 + \xi \frac{Y_i - \mu}{\sigma} \right) - \sum_{i=1}^m \left( 1 + \xi \frac{Y_i - \mu}{\sigma} \right)^{-\frac{1}{\xi}}$$

Et si  $\xi = 0$  :

$$\text{Log}L = -m \log \sigma - \sum_{i=1}^m \exp\left(-\frac{Y_i - \mu}{\sigma}\right) - \sum_{i=1}^m \left(\frac{Y_i - \mu}{\sigma}\right)$$

De plus, Smith en 1985 a montré que si  $\xi > -1/2$ , alors :  $\sqrt{m} \left[ \begin{pmatrix} \hat{\mu}, \hat{\sigma}, \hat{\xi} \\ \text{loi} \end{pmatrix} - (\mu, \sigma, \xi) \right] \rightarrow N(0, V)$  où  $V$  matrice de variance-covariance (on renvoie à cet auteur pour la forme de  $V$ ).

En pratique l'analyse des maxima par bloc implique certaines restrictions. Du fait de l'échantillonnage, il se peut qu'un bloc contienne une ou plusieurs valeurs extrêmes supplémentaires qui sont plus grandes que les maxima d'autres blocs. En prenant dans ce cas un seul maximum par bloc pour décrire le comportement des extrêmes, on perd une quantité d'information qui pourrait contribuer à une meilleure compréhension des extrêmes. Nous ne privilégierons donc pas cette méthode.

<sup>35</sup> Voir l'ouvrage Statistical theory of extreme values and some practical applications, Gumbel, 1958

### Estimation du paramètre $\xi$ de la loi GPD

Dans la littérature de la théorie des valeurs extrêmes on trouve plusieurs techniques semi-paramétriques pour l'estimation de l'indice de queue. On peut citer l'estimateur de Pickands (1975), celui de Hill (1975) valable pour  $\xi > 0$  et celui de Dekkers et al. (1989).

Ces estimateurs sont basés sur la statistique d'ordre  $X_{n,m} \leq \dots \leq X_{n,1}$  et sont définis par :

$$\xi_{n,m}^{Pickands} = \frac{1}{\log 2} \log \frac{X_{n-m:n} - X_{n-2m:n}}{X_{n-2m:n} - X_{n-4m:n}}$$

$$\xi_{n,m}^{Hill} = \frac{1}{m} \sum_{i=1}^m \log X_{n-i+1:n} - \log X_{n-m:n}$$

$$\xi_{n,m}^{DEdH.} = \xi_{n,m}^{H(1)} + 1 - \frac{1}{2} \left[ 1 - \frac{(\xi_{n,m}^{H(1)})^2}{\xi_{n,m}^{H(2)}} \right]^{-1}$$

où :  $\xi_{n,m}^{H(r)} = \frac{1}{m} \sum_{i=1}^{m-1} (\log X_{n-i:n} - \log X_{n-m:n})^r \quad r=1,2,\dots$

Ils sont tous confrontés au choix optimal de  $m$ . En effet, il est indispensable de calculer ces estimateurs sur les vraies queues de distribution. Choisir un  $m$  trop élevé engendre le risque de prendre en compte des valeurs qui ne sont pas extrêmes, nous sortons de la queue de distribution, à l'inverse, un sous-échantillon trop court ne permet pas aux estimateurs d'atteindre leur niveau de stabilité.

Il est possible de montrer que ces estimateurs convergent quand  $n \rightarrow \infty$ ,  $m \rightarrow \infty$  et  $m/n \rightarrow 0$ . On a (sous des hypothèses de régularité et de vitesse de convergence de  $m$  par rapport à  $n$ ) une normalité asymptotique des estimateurs.

$$\sqrt{m}(\xi_{n,m}^{Pickands} - \xi) \xrightarrow[loi]{} N\left(0, \frac{\xi^2(2^{\xi+1} + 1)}{(2(2^\xi - 1) \log 2)^2}\right)$$

$$\sqrt{m}(\xi_{n,m}^{Hill} - \xi) \xrightarrow[loi]{} N(0, \xi^2)$$

$$\sqrt{m}(\xi_{n,m}^{DEdH.} - \xi) \xrightarrow[loi]{} N(0, 1 + \xi^2)$$

L'estimateur de Pickands est valable quelle que soit la distribution des extrêmes (Gumbel, Weibull ou Fréchet). Quand on trace le graphique de cet estimateur en fonction du nombre  $m$  d'observations considérées, le comportement est très volatil au départ, ce qui nuit à la lisibilité du graphique. De plus, l'estimateur est mécaniquement très sensible à la taille de l'échantillon sur lequel on travaille, ce qui le rend peu robuste.

L'estimateur de Hill n'est utilisable que pour les distributions de Fréchet pour lesquelles il fournit un estimateur de l'indice de queue plus efficace que l'estimateur de Pickands. Pour vérifier que l'on est dans le cas de Fréchet, nous appliquerons la méthode vue en page 43.

### Conclusion

La théorie des valeurs extrêmes nous permet d'exposer deux approches possibles pour l'estimation de l'indice de queue  $\xi$  : l'une par la loi des valeurs extrêmes généralisée (GEV) et l'autre par la loi de Pareto généralisée (GPD). Nous ne privilégions pas l'approche GEV (maxima par blocs) car on perdrait une quantité d'information pouvant contribuer à une meilleure compréhension des extrêmes.

Dans la partie suivante, nous utiliserons les estimateurs du paramètre  $\xi$  de la loi GPD (Hill, Pickands, DeDH) afin de proposer des méthodes graphiques de détermination du seuil de sinistralité extrême.

## IV - B METHODES DE DETERMINATION DU SEUIL

Nous allons présenter dans la section suivante le QQ-plot qui nous permettra de vérifier que nos observations suivent bien une loi à queue épaisse.

Ensuite nous verrons quelques approches graphiques permettant d'aborder la détermination du seuil  $u$  au-delà duquel les observations sont extrêmes : fonction moyenne des excès (FME), estimateur de Hill, estimateur de Pickands, estimateur de Dekkers et *al.*, estimateur du maximum de vraisemblance (EMV). Une valeur  $u$  constituera un seuil acceptable lorsque, graphiquement, l'évolution de  $\hat{\xi}$  en fonction du seuil se stabilise. Cette approche visuelle ne permet pas de déterminer une valeur précise mais plutôt un intervalle dans lequel se situerait le seuil. Elle nécessite l'intervention et la prise de décision subjective de l'utilisateur. Pour comparer, nous utiliserons une autre méthode, la Gertensgarbe plot, non issue de la théorie des valeurs extrêmes et qui s'appuie sur le test de Mann-Kendall. Cette méthode présente l'intérêt de déterminer la valeur du seuil  $u$  par l'intersection de deux courbes.

### IV - B . 1. QQ-PLOT : QUANTILE-QUANTILE PLOT

Le QQ-plot est un graphique qui oppose les quantiles de l'échantillon des  $x_i$  ( $i = 1, \dots, n$ ),  $(\hat{F}_n^{-1}(p))$  aux quantiles de la distribution théorique envisagée ( $F^{-1}(p)$ ), pour  $p \in [0,1]$ , où  $F$  est la fonction de répartition de la vraie loi et  $\hat{F}_n(p)$  la fonction de répartition empirique. Si l'échantillon provient bien de cette distribution théorique, alors **le QQ-plot sera linéaire**.

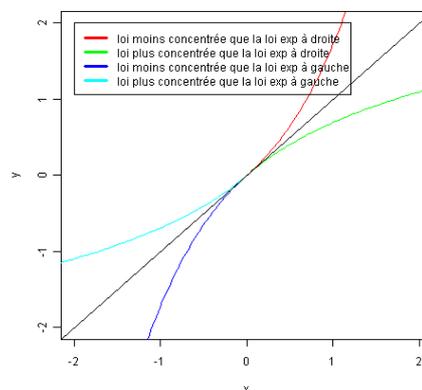
Le graphique du QQ-plot consiste à tracer les points  $\left( F^{-1}\left(\frac{i}{n+1}\right), x_{(i)} \right)$  pour  $i = 1, \dots, n$  où  $x_{(i)}$  est la  $i^{\text{ème}}$  valeur de l'échantillon ordonné.

Dans la théorie des valeurs extrêmes, le QQ-plot est basé sur la distribution exponentielle  $F(x) = 1 - \exp(-x/\lambda)$  où  $x > 0$ . Ce graphique est l'ensemble des points :  $\left( -\ln\left(1 - \frac{i}{n+1}\right), x_{(i)} \right)$ .

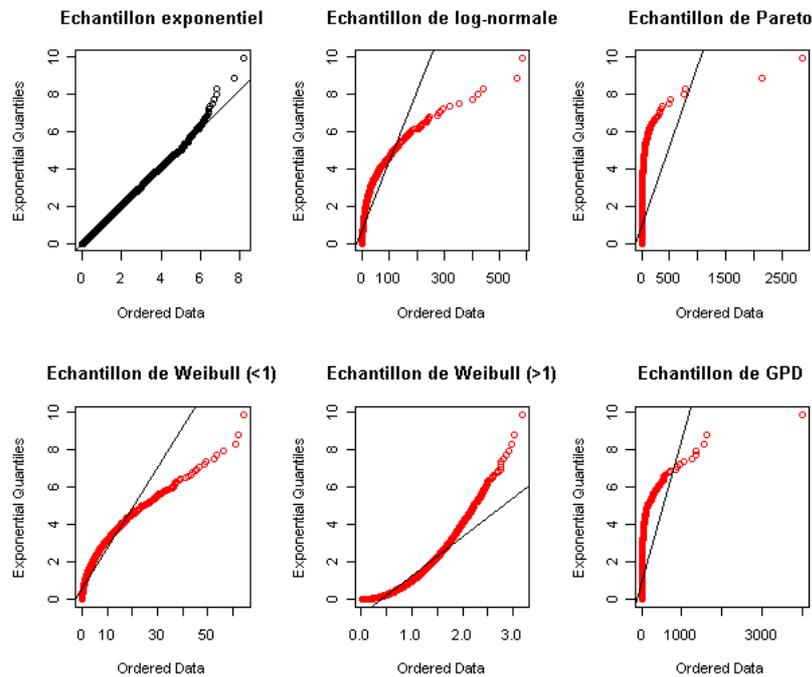
L'intérêt de ce graphique est de nous permettre d'obtenir la forme de la queue de la distribution.

Trois cas de figure sont possibles :

- Les données suivent la loi exponentielle : la distribution présente une queue très légère, les points du graphique présentent une forme linéaire de pente  $1/\lambda$ .
- Les données suivent une distribution à queue plus lourde : le graphique QQ-plot est convexe.
- Les données suivent une distribution à queue plus légère : le graphique QQ-plot a une forme concave.



Exemples de QQ-plot exponentiel pour des échantillons de loi exponentielle, log-normale, de Pareto, Weibull et de GPD :



Par ailleurs, le QQ-plot nous permet de tester graphiquement l'adéquation d'un échantillon à une multitude de lois.

On indique ci-dessous l'ensemble des points du QQ-plot selon quelques lois standards :

Loi	QQ-plot
Pareto	$\left( -\ln\left(1 - \frac{i}{n+1}\right), \ln(x_{(i)}) \right) i=1, \dots, n$
Weibull	$\left( \ln\left( -\ln\left(1 - \frac{i}{n+1}\right) \right), \ln(x_{(i)}) \right) i=1, \dots, n$
Normale	$\left( \Phi^{-1}\left(\frac{i}{n+1}\right), x_{(i)} \right) i=1, \dots, n$ où $\Phi$ fonction de répartition de $N(\mu, \sigma^2)$
Log-normale	$\left( \Phi^{-1}\left(\frac{i}{n+1}\right), \ln(x_{(i)}) \right) i=1, \dots, n$ où $\Phi$ fonction de répartition de $N(\mu, \sigma^2)$

**IV - B . 2. MEAN EXCESS FUNCTION (FME) - PENTE DE L'ESPERANCE RESIDUELLE**

Un deuxième outil pour évaluer l'épaisseur de queue est la FME. Cette méthode nous permettra aussi de décrire la prédiction de dépassement du seuil  $u$  lorsqu'un excès se produit.

Elle est définie par :  $e_n(u) = E[X - u / X > u]$  où  $0 < u \leq x_F$ .

L'estimateur empirique de la fonction d'excès en moyenne est défini comme étant le rapport entre le montant total des excès par rapport au seuil  $u$  et le nombre total de points dépassant le seuil  $u$ .

Il est donné par : 
$$\hat{e}_n(u) = \frac{\sum_{i=1}^n (X_i - u)^+}{\sum_{i=1}^n I_{\{X_i > u\}}}$$
 où  $(X_i - u)^+ = \text{Sup}(X_i - u, 0)$ .

Plus une loi est à queue épaisse, plus son espérance résiduelle croît rapidement. Comme pour le QQ-plot, on

peut distinguer différents cas :

- Si la FME est constante, alors la queue de distribution est similaire a une exponentielle (si  $X \sim \text{Exp}(\lambda)$ , alors  $e(u) = 1/\lambda$  : constante),
- Si la FME est croissante, alors la queue est plus lourde,
- Si la FME est décroissante alors la queue est plus légère.

En particulier la FME est linéaire dans le cas où  $X$  suit une loi de Pareto généralisée (si  $X \sim \text{GPD}(\zeta, \sigma)$ , alors  $e(u) = \sigma/(1-\zeta) + \zeta/(1-\zeta) \cdot u$ ).

L'approche pour la détermination du seuil consiste à tracer le mean excess plot de  $X$ , c'est-à-dire le graphe des points  $(u, \hat{e}_n(u))$  et à choisir  $u$  de manière à ce que  $\hat{e}_n(u)$  soit approximativement linéaire pour tout  $x \geq u$ . Graphiquement, cela se traduit par un changement de la pente de la courbe qui ensuite reste stable.

**Propriété de la GPD (Stabilité) :**

Soit  $Y$  une variable aléatoire distribuée selon une GPD de paramètres  $(\xi, \sigma)$ .

La variable aléatoire  $Y_u = [Y - u/Y > u]$  est distribuée selon une GPD de paramètre  $(\xi, \sigma + \xi u)$ . On en déduit que si  $\xi < 1$ , alors pour tout  $u < y_F = \sup \{y \in \mathbb{R}, F(y) < 1\}$ ,

$$E[Y - u/Y > u] = \frac{\alpha + \xi u}{1 - \xi} \text{ pour } \alpha + \xi u > 0$$

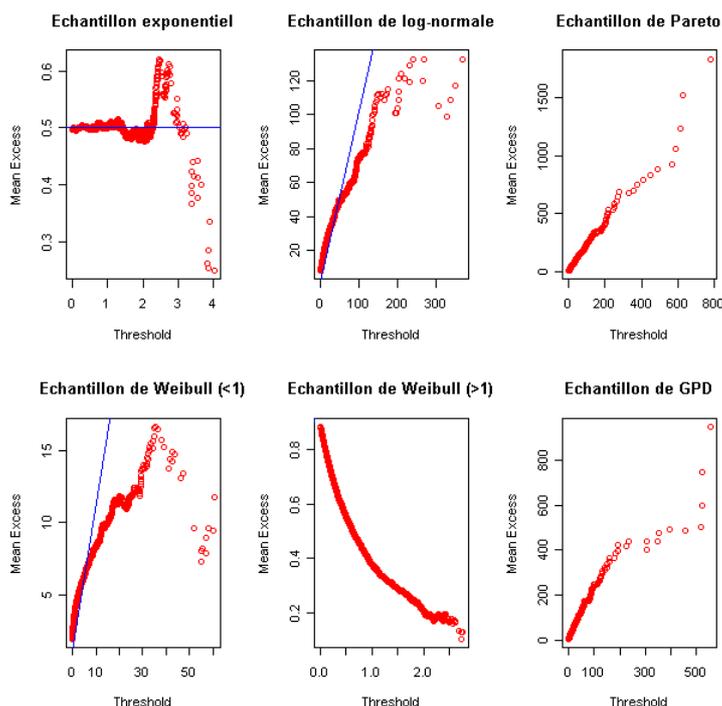
Comme la fonction d'espérance résiduelle d'une loi GPD de paramètre  $\xi < 1$  est linéaire :  $e(v) = \frac{\alpha + \xi v}{1 - \xi}$ ,

on cherchera un seuil  $u$  le plus petit possible, avec la contrainte que l'estimateur empirique de l'espérance résiduelle des excès au delà de  $u$  soit approximativement linéaire.

En pratique, cette méthode est parfois délicate car le graphe de l'estimation de  $e(u)$  peut présenter des aspects irréguliers. Une solution à ce problème consiste à s'intéresser à la pente de l'espérance résiduelle :  $\frac{\xi}{1 - \xi}$ . On conserve tous les sinistres au dessus du seuil  $u$  et on effectue une régression linéaire. Le

coefficient directeur obtenu, noté  $a$ , permet de calculer l'estimateur :  $\hat{\xi}(u) = \frac{a}{1 + a}$  (pour un seuil donné).

*mean excess plot pour des simulations de lois usuelles*



Dans l'exemple ci-dessus, la mean excess function d'un échantillon de loi exponentielle de paramètre 0.5 est

une droite constante de valeur 0.5, hormis pour les dernières observations.

La mean excess function d'un échantillon de Weibull de paramètre  $>1$  est décroissante ; tandis que celle de l'échantillon de log-normale ou de l'échantillon de Weibull de paramètre  $<1$  est croissante.

Concernant les échantillons de lois de Pareto et GPD, nous observons une tendance affine, même si pour la GPD l'alignement est moins bon sur les derniers points.

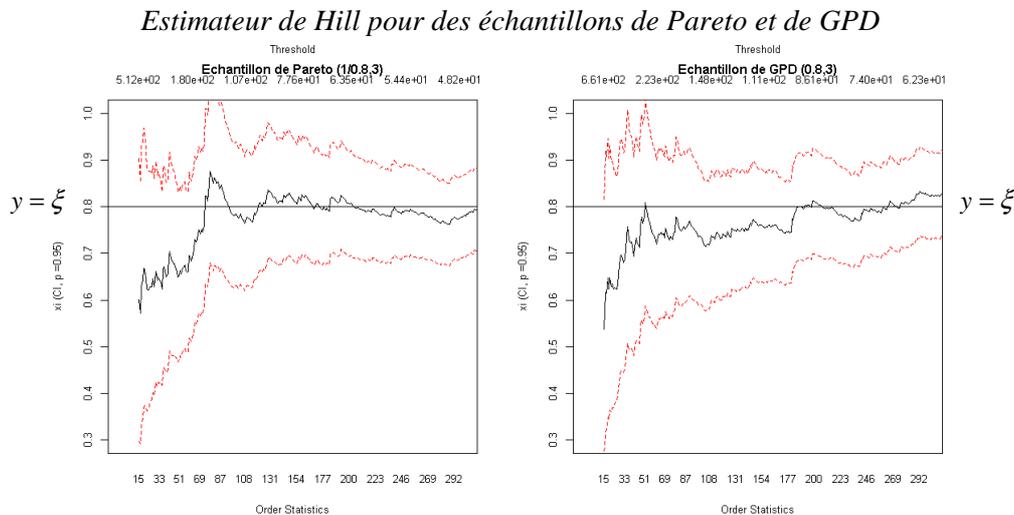
En conclusion, sur données réelles, la mean excess plot est instable notamment sur les dernières observations (maxima). Construire ce graphe est toutefois très intéressant car il nous permet de connaître l'épaisseur de la queue et d'avoir une idée du seuil séparant les valeurs extrêmes (moment où l'on commence à avoir une droite).

### IV - B . 3. L'ESTIMATEUR DE HILL

C'est l'un des estimateurs les plus répandus de l'indice de valeurs extrêmes. Introduit par Hill en 1975, la méthode consiste à ordonner les observations par ordre décroissant  $X_1 > X_2 > \dots > X_n$ , l'indice de queue étant donné par l'équation ci-dessous valable que lorsque  $\xi > 0$  (distributions de Fréchet) :

$$\hat{\xi}^{Hill}(u) = \frac{1}{k} \sum_{i < k} \text{Ln} \left( \frac{X_i}{X_k} \right) \text{ où } k \text{ est le nombre d'observations supérieures au seuil } u.$$

Il s'agit de sélectionner graphiquement le nombre d'excès au-delà duquel la valeur de l'indice de queue  $\xi$  devient stable, c'est-à-dire quand l'estimation devient plus robuste. Selon Dress, de Haan et Resnick (1998), cette méthode est bien adaptée aux distributions d'excès convergeant vers une GPD en assurant un bon équilibre entre biais et variance.



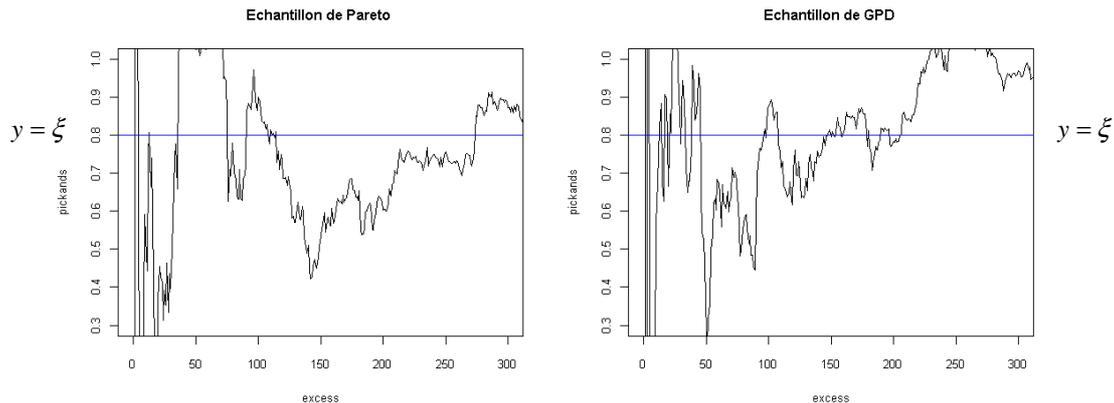
L'estimateur de Hill est volatil quand le nombre d'extrema considérés est trop petit, ensuite il est très près de la valeur de l'indice de queue ( $y = \xi$ ).

### IV - B . 4. L'ESTIMATEUR DE PICKANDS

La méthode est analogue à celle de Hill, sauf que nous utilisons l'estimateur suivant :

$$\hat{\xi}^{Pickands}(u) = \frac{1}{\log 2} \log \frac{X_k - X_{2k}}{X_{2k} - X_{4k}}$$

*Estimateur de Pickands pour des échantillons de Pareto et de GPD*



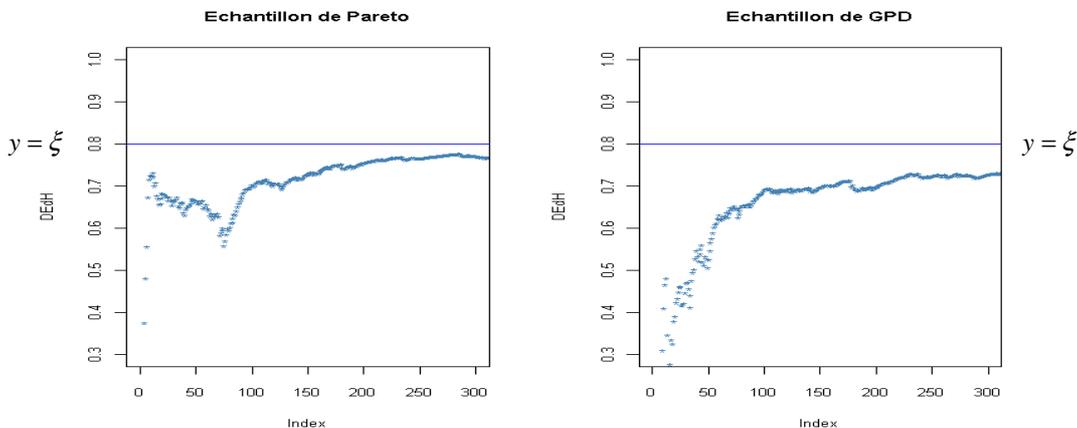
L'estimateur de Pickands est plus volatil du fait qu'il ne tient compte que de 3 observations ( $X_k, X_{2k}$  et  $X_{4k}$ ), il est donc sensible aux sauts. Nous avons utilisé le même échantillon que pour l'exemple de Hill, la convergence vers l'indice de queue est difficile à appréhender.

**IV - B . 5. L'ESTIMATEUR DE DH**

La méthode est encore une fois analogue, l'estimateur utilisé est le suivant :

$$\hat{\xi}^{DEdH}(u) = \hat{\xi}^{Hill(1)}(u) + 1 - \frac{1}{2} \left[ 1 - \frac{(\hat{\xi}^{Hill(1)})^2}{\hat{\xi}^{Hill(2)}} \right]^{-1} \quad \text{où : } \hat{\xi}^{Hill(r)} = \frac{1}{k} \sum_{i < k} \left( \text{Ln} \left( \frac{X_i}{X_k} \right) \right)^r \quad r=1,2$$

*Estimateur DEdH pour des échantillons de Pareto et de GPD*



Cet estimateur paraît donner une estimation de l'indice de queue moins bonne que celle de l'estimateur de Hill.

**IV - B . 6. L'ESTIMATEUR DU MAXIMUM DE VRAISEMBLANCE (EMV)**

On peut ajuster une loi de Pareto sur les coûts en excédent de u en utilisant l'EMV de cette loi.

Soit  $X=(X_1, X_2, \dots, X_k)$ , un échantillon aléatoire de loi de Pareto :  $f(x) = \alpha \frac{x_m^\alpha}{x^{1+\alpha}} \quad x \geq x_m$

Par définition, la fonction de vraisemblance pour  $x \geq x_m$  s'écrit :  $L(\alpha, x_m) = \prod_{i=1}^k \alpha \frac{x_m^\alpha}{x_i^{\alpha+1}} = \alpha^k x_m^{k\alpha} \prod_{i=1}^k \frac{1}{x_i^{\alpha+1}}$

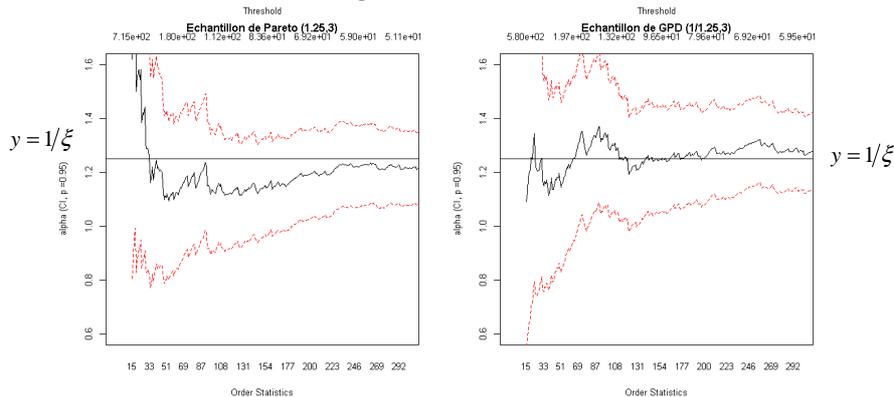
En prenant le logarithme de cette expression on obtient :  $\log L(\alpha, x_m) = k \ln \alpha + k\alpha \ln x_m - (\alpha + 1) \sum_{i=1}^k \ln x_i$

La dérivée par rapport à  $\alpha$  donne :  $\frac{\partial \log L}{\partial \alpha} = \frac{k}{\alpha} + k \ln x_m - \sum_{i=1}^k \ln x_i$

D'où l'estimateur :  $\hat{\alpha} = \frac{k}{\sum_i (\ln x_i - \ln \hat{x}_m)}$

On retrouve l'inverse de l'estimateur de Hill de l'indice de queue  $\xi$ .  
 Notre échantillon étant constitué des coûts supérieurs ou égaux au seuil  $u$  on a :  $u = x_m$ .

*Maximum de vraisemblance pour des échantillons de Pareto et de GPD*



L'erreur statistique attendue est la suivante:  $\sigma = \frac{\hat{\alpha}}{\sqrt{n}}$ .

**IV - B . 7. MINAMSE**

**D'où cela vient-il ?**

Les auteurs Beirlant et al.<sup>36</sup> constatent que la technique actuelle de détermination graphique d'un seuil via une zone de stabilité de la séquence de l'estimateur est parfois difficile à mettre en oeuvre à cause de la grande volatilité des estimateurs.

**Qu'est-ce que c'est ?**

La méthode repose sur une interprétation novatrice des estimateurs kernel comme solutions d'un algorithme de moindres carrés pondérés reposant sur le quantile plot de Pareto des données, avec les poids induits par le kernel. Le but d'une fonction kernel est de pondérer les observations par rapport à un point de référence de sorte que plus une observation est proche de la référence, plus son poids sera grand. C'est donc un algorithme qui estime les paramètres d'échelle et de forme d'une distribution de Pareto par minimisation asymptotique de l'erreur quadratique moyenne de l'estimateur kernel<sup>37</sup>.

Cet algorithme permet en plus de trouver un entier k qui représente le nombre des montants les plus extrêmes de sinistres pris en compte pour minimiser l'AMSE (Asymptotic Mean Squared Error) :

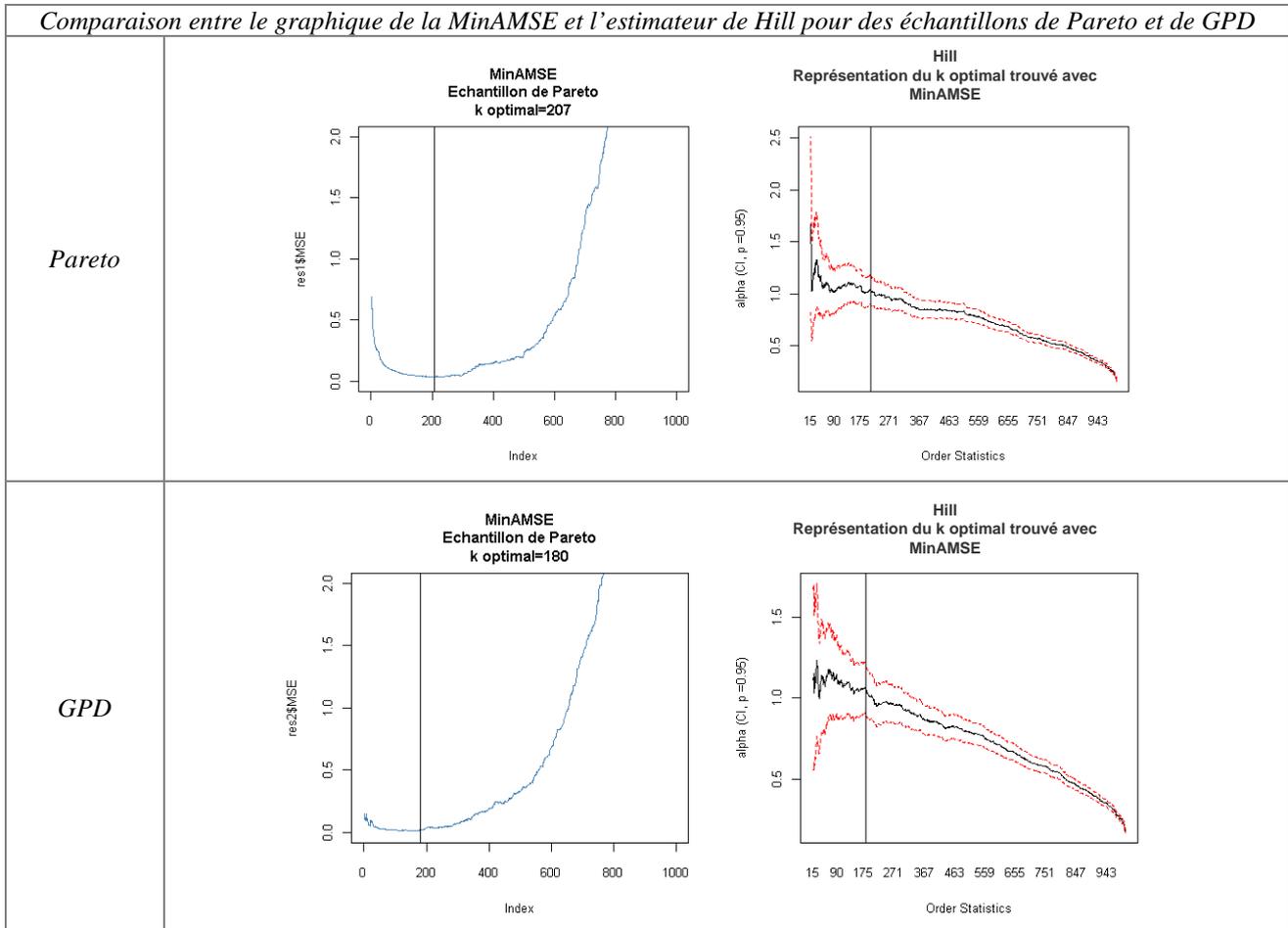
$$MSE_{opt}(k) = \frac{1}{k} \sum_{i=1}^k w_{i,k}^{opt} \left[ \log \frac{X_{n-i+1}}{X_{n-k}} - \frac{1}{\alpha} \log \left( \frac{k+1}{i} \right) \right]^2$$

où  $w_{i,k}^{opt}$  représente le poids qui permet d'optimiser la MSE pour avoir un estimateur consistant qui minimise

<sup>36</sup> Excess functions and estimation of the extreme-value index, J.Beirlant and al., Journal of the American Statistical Association, décembre 1996.  
<sup>37</sup> Kernel goodness-of-fit statistics for Pareto-type behavior and threshold selection for tail index estimation, Y.Goegebeur and al.

asymptotiquement l'erreur quadratique moyenne de Hill.

Sur un  $n$ -échantillon, le  $(n-k)^{ième}$  montant de sinistre est alors le seuil à partir duquel la distribution des montants de sinistres suit une loi de Pareto.  $k$  est en fait défini tel que la droite de régression à travers la  $k^{ième}$  plus grande observation avec une pente induite par l'estimateur particulier des moindres carrés pondérés (par exemple l'estimateur de Hill), s'adapte le mieux possible à la queue du quantile plot de Pareto.



La fonction MinAMSE renvoie bien le nombre d'excès à partir duquel l'estimateur de Hill n'est plus stable.

#### IV - B . 8. GERTENSGARBE PLOT

Cette méthode est issue des travaux de Gerstengarbe et Werner<sup>1</sup> et permet de déterminer le point de départ de la région extrême en donnant une estimation du seuil optimal. Elle s'appuie sur le test de Mann-Kendall utilisé habituellement pour identifier une tendance de saisonnalité. L'idée est qu'à partir de la zone des sinistres extrêmes on aura sans doute une modification du comportement des écarts  $\Delta_i$  (différence de coût entre deux sinistres). Nous cherchons donc à identifier un changement dans une série.

Plus précisément, pour un échantillon  $x_1, \dots, x_n$  de montants de sinistres, on considère la série des écarts  $\Delta_i$  de l'échantillon ordonné noté  $x_{[1]}, \dots, x_{[n]}$  et on a :  $\Delta_i = x_{[i]} - x_{[i-1]}$ , avec  $i = 2$  à  $n$ .

Nous utilisons la version séquentielle du test de Mann-Kendall : pour chaque élément  $\Delta_k$ , on calcule combien d'éléments précédents  $\Delta_2, \dots, \Delta_k$  lui sont inférieurs, soit  $n_k$ . Le test va porter sur la grandeur  $U_i^*$

<sup>1</sup> A Method for the Statistical Definition of Extreme-Value Regions and Their Application to Meteorological Time Series, F.W. Gertensgarbe and P.C Werner, 2009.



## **IV - C TRANSFORMATION « AS IF » DES DONNEES**

Pour appliquer les outils vus précédemment, il est nécessaire d'avoir une base homogène de sinistres, c'est-à-dire dans laquelle les conditions de tous les exercices soient identiques.

Nous effectuerons la mise « as if » des sinistres afin de rendre l'inflation exogène, en effet un sinistre passé a un coût différent du même sinistre qui se serait produit aujourd'hui. Ensuite, tous les sinistres n'ayant pas la même maturité, nous utiliserons la méthode Chain Ladder pour estimer les sinistres à l'ultime.

### **IV - C . 1. QUELQUES NOTATIONS**

Classiquement, on notera :

- $i$  (en ligne) l'année de survenance,  $i = 2000, \dots, 2010$
- $j$  (en colonne) l'année de développement,  $j = 1, \dots, 11$
- $Y_{i,j}$  les paiements de l'année de développement  $j$ , pour les sinistres survenus l'année  $i$
- $C_{i,j}$  les paiements cumulés, au sens où  $C_{i,j} = Y_{i,0} + Y_{i,1} + \dots + Y_{i,j}$ , pour l'année de survenance  $j$
- $N_{i,j}$  le nombre cumulé de sinistres pour l'année de survenance  $i$  vu au bout de  $j$  années
- $SAP_{i,j}$  : le montant restant à payer l'année  $j$  pour un sinistre survenu l'année  $i$
- $Inc_{i,j}$  : le montant évalué l'année  $j$  pour un sinistre survenu l'année  $i$

### **IV - C . 2. TRAITEMENT DE L'INFLATION**

La cadence de développement de la sinistralité en Responsabilité Civile Corporelle automobile (RCC) est très longue, au bout de 11 ans certains sinistres pouvant ne pas être totalement réglés. Ces sinistres subissent l'évolution normale des principaux coûts (salaires hospitaliers, etc.).

Pendant cette période on peut aussi avoir eu des chocs (réglementaires, jurisprudentiels,...) ayant entraîné une modification de l'inflation « normale » (voir II-B.5).

Si l'inflation varie d'une année à l'autre et que l'on n'en tient pas compte, notre analyse de provisionnement sera perturbée : une hausse de l'inflation génèrera un sous-provisionnement et une baisse de l'inflation génèrera un sur-provisionnement. D'autre part, la qualité des estimateurs de lois sera dégradée. Il est donc préférable de scinder les phénomènes de liquidation (hors inflation) et les effets de l'inflation.

Plusieurs méthodes permettent de mettre les sinistres « as if », c'est-à-dire de déterminer le coût du sinistre s'il était survenu l'année étudiée. Dans notre étude, nous cherchons quel taux d'inflation prendre en compte, l'impact de la jurisprudence sur les règlements semblant l'emporter sur celui de l'évolution des prix.

Plusieurs indices peuvent nous permettre de redresser les montants de l'inflation parmi lesquels :

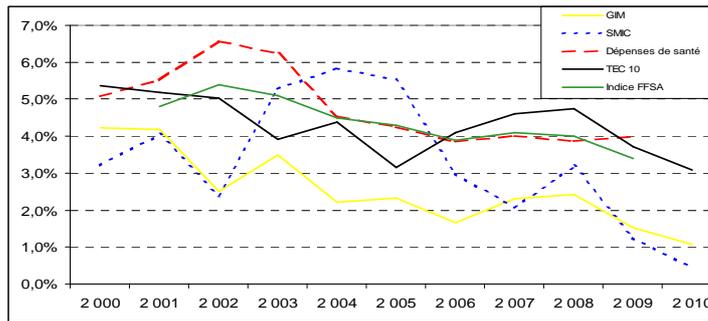
- Indice GIM : coût de main d'œuvre de la métallurgie,
- Evolution des coûts moyens de sinistres,
- Evolution des salaires,
- Indice de santé,
- Indice FFSA de l'inflation des sinistres corporels<sup>38</sup> ( $2,18 \times \Delta$  Indice des salaires de base).

Les indices GIM et salaires nous semblent trop spécialisés. L'indice de santé étant corrélé à l'âge de la population, il nous paraît difficilement exploitable. L'indice FFSA apparaît être le plus intéressant car il dissocie l'effet inflation des autres facteurs comme la gravité de l'accident observée dans les coûts moyens ; seule la variable explicative traduisant l'inflation propre à la garantie est conservée.

On pourrait aussi construire un indice tenant compte de l'évolution des principales composantes du coût moyen des sinistres corporels.

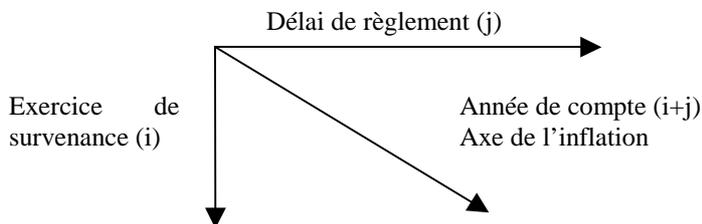
<sup>38</sup> Se reporter à l'annexe 2 de l'enquête FFSA « Le marché de l'assurance automobile en 2009 - Analyse des résultats détaillés »

Evolution des indices :



Ces indices ont des évolutions très différentes même si les tendances sont parallèles. Nous pouvons noter que l'indice GIM est même moins élevé que l'indice des dépenses de santé, ce qui explique que certains réassureurs y ajoutent une sur-inflation.

On prend l'hypothèse que l'inflation se propage selon l'axe des années de comptes (effet diagonal):



On note  $I_k$  le facteur d'inflation de l'année k.

La mise « as if » s'effectue sur les montants payés non cumulés et sur les montants restant à payer. Pour établir les valeurs hors inflation des sinistres de l'année de survenance N, il faut multiplier chaque valeur par le facteur  $I_{(2010+j)}/I_{(N+j)}$  où j est l'année de développement du sinistre (voir schéma de mise « as if » ci-après).

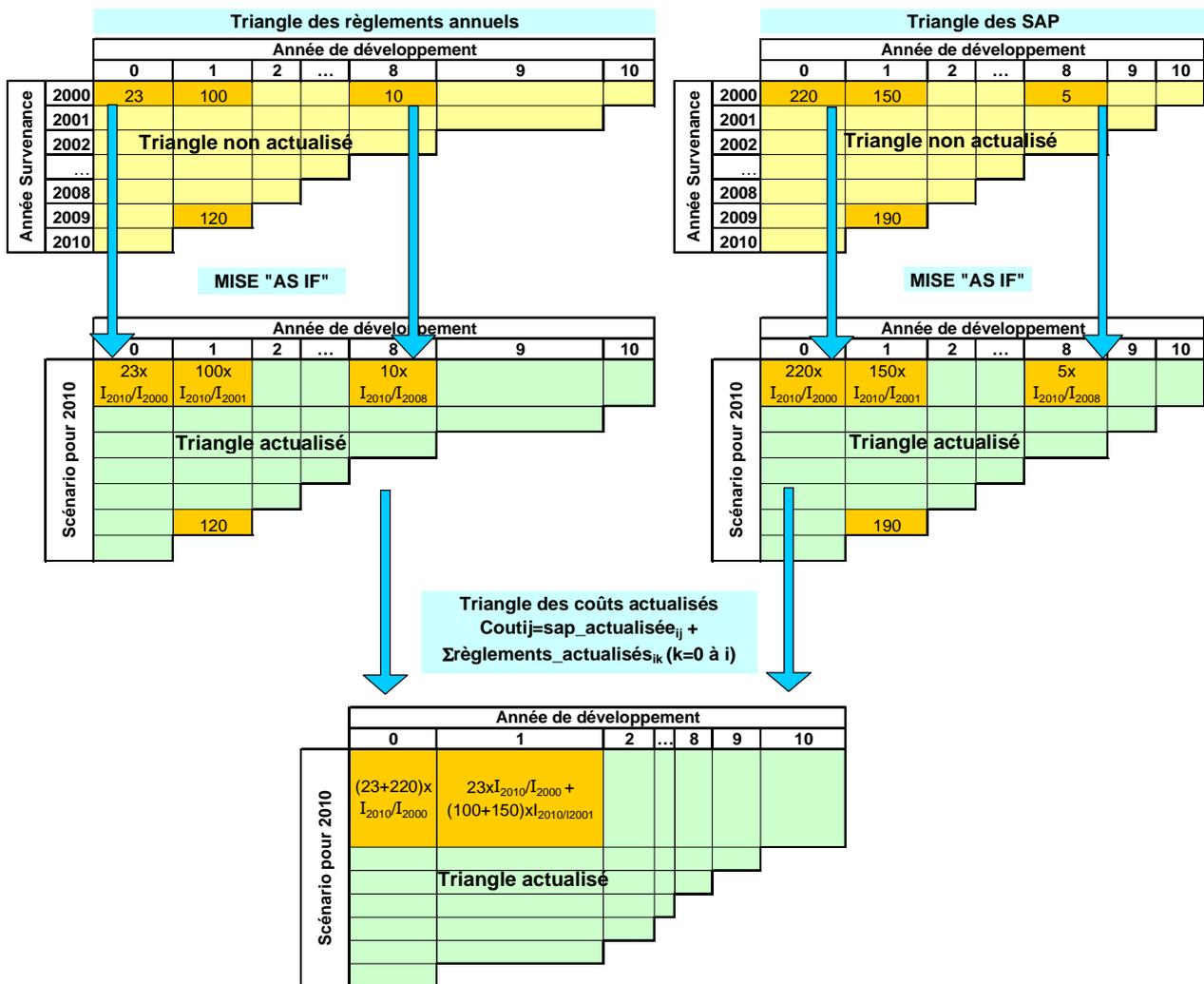
Exemple d'actualisation du sinistre suivant survenu en 2002 :

Année de développement	0	1	2	3	4	5	6	7	8
<b>Païement annuels (<math>Y_{i,j}</math>)</b>	0	86 491	83 267	57 886	599 228	0	0	0	0
<b>Païement Cumulés (<math>C_{i,j}</math>)</b>	0	86 491	169 758	227 644	826 872	826 872	826 872	826 872	826 872
<b>Restant (<math>SAP_{i,j}</math>)</b>	8 000	524 809	444 300	506 212	298 776	298 776	0	0	0
<b>Coût Total (<math>Inc_{i,j}</math>)</b>	8 000	611 300	614 058	733 856	1 125 648	1 125 648	826 872	826 872	826 872

En appliquant la mise « as if » avec une inflation constante de 4%, on obtient :

Année de développement	0	1	2	3	4	5	6	7	8
<b>Païement annuels (<math>Y_{i,j}</math>)</b>	0	118 369	113 957	79 221	820 085	0	0	0	0
<b>Païement Cumulés (<math>C_{i,j}</math>)</b>	0	118 369	232 326	311 547	1 131 631	1 131 631	1 131 631	1 131 631	1 131 631
<b>Restant (<math>SAP_{i,j}</math>)</b>	10 949	718 237	608 055	692 786	408 896	408 896	0	0	0
<b>Coût Total (<math>Inc_{i,j}</math>)</b>	10 949	836 606	840 381	1 004 333	1 540 527	1 540 527	1 131 631	1 131 631	1 131 631

Schéma de mise « as if » :



Remarque : la gestion n'est pas différenciée selon le mode de règlement du sinistre : rente ou capital. Dans le cas d'une rente, le coût du sinistre correspond au capital constitutif. On n'a donc pas le détail des règlements de la rente année par année pour le retraitement de l'inflation.

#### IV - C . 3. PROJECTION A L'ULTIME DE LA CHARGE DES SINISTRES

La méthode Chain Ladder est une méthode déterministe qui s'applique sur des triangles de liquidation cumulés. Elle est basée sur l'utilisation de facteurs de développement (cadence), implicitement supposés constants pour toutes les années d'origine.

Les montants  $C_{i,j}$  sont connus pour  $i + j \leq n$  ; on cherche à estimer les valeurs futures  $C_{i,j}$  pour  $i + j > n$  et en particulier les valeurs ultimes  $C_{i,n}$ .

La méthode s'appuie sur l'hypothèse (forte) suivante :

Pour  $j=0, \dots, n-1$ , (où  $n$  : durée de développement) les ratios des facteurs adjacents  $\frac{C_{i,j+1}}{C_{i,j}}$  sont indépendants de l'année d'origine  $i$ .

$$\text{D'où pour } j=0, \dots, n-1 : \frac{C_{0,j+1}}{C_{0,j}} = \frac{C_{1,j+1}}{C_{1,j}} = \dots = \frac{C_{i,j+1}}{C_{i,j}} = \dots = \frac{C_{10,j+1}}{C_{10,j}}$$

Dans la pratique ces égalités n'étant, au mieux, qu'approximativement vérifiées, il est naturel de choisir comme facteur commun :

$$f_j = \frac{\sum_{i=0}^{n-j-1} C_{i,j+1}}{\sum_{i=0}^{n-j-1} C_{i,j}} \quad (0 \leq j \leq n-1)$$

On en déduit le triangle complet de liquidation par :

		développement					
		0	1	...	j	...	n
survenance	0	$C_{0,0}$	$C_{0,1}$		$C_{0,j}$		$C_{0,n}$
	1	$C_{1,0}$	$C_{1,1}$		$C_{1,j}$		$C_{1,n}$
	...						
	i	$C_{i,0}$	$C_{i,1}$		$C_{i,j}$		$C_{i,n}$
	...						
	n	$C_{n,0}$	$C_{n,1}$		$C_{n,j}$		$C_{n,n}$

$$C_{i,j} = C_{i,n-i} f_{n-i} \dots f_{j-1} = C_{i,n-i} \prod_{k=n-i}^{j-1} f_k$$

Cette méthode se base donc sur l'hypothèse que la cadence historique de l'évolution des sinistres se répète.

En particulier l'évaluation de la charge ultime de l'année i est :

$$S_i = C_{i,n} = C_{i,n-i} f_{n-i} \dots f_{n-1} = C_{i,n-i} \prod_{k=n-i}^{n-1} f_k$$

On calcule enfin l'estimation Chain Ladder de la provision par année de survenance i :

$$R_i = S_i - C_{i,n-i}$$

La provision globale est :  $R = \sum_{i=1}^n R_i$

La méthode Chain Ladder qui a l'avantage d'être simple, est probablement la méthode la plus populaire parmi les méthodes d'estimation des PSAP. Cependant elle présente des défauts non négligeables. Tout d'abord, l'hypothèse fondamentale de Chain Ladder, l'existence de coefficients de passage entre les  $C_{i,j}$  et les  $C_{i,j+1}$ , est une hypothèse très forte qu'il faut vérifier au moins graphiquement en s'assurant que, pour chaque année de développement, les points  $(C_{i,j}, C_{i,j+1})_{i=0, \dots, n-j-1}$  (C-C plot) sont sensiblement alignés sur une droite passant par l'origine.

Il faut également prendre en compte le nombre de paramètres du modèle. On estime n paramètres, les  $f_j$  ( $0 \leq j \leq n-1$ ) avec seulement  $(n+1)(n+2)/2$  données ce qui peut être source d'erreurs. En particulier le dernier paramètre  $f_n$  n'est estimé qu'à l'aide des deux données  $C_{0,n-1}$  et  $C_{0,n}$ . Cette source d'erreur peut vite devenir problématique car le modèle est multiplicatif. De plus la méthode est déterministe et ne permet pas de mesurer ce risque.

## IV - D APPLICATIONS

Pour appliquer les outils vus précédemment, nous avons besoin de constituer une base de données sinistres.

### IV - D . 1. LES DONNEES

Le périmètre que nous utilisons correspond aux sinistres RCC automobile et flottes de 1<sup>ère</sup> catégorie survenus depuis 2000.

Avant d'être convergés dans un système d'information unique (SI), les sinistres étaient gérés dans quatre SI (cf. I-B.4). Les historiques disponibles sont variables d'un SI à l'autre :

Année de vision									
2009	2008	2007	2006	2005	2004	2003	2002	2001	2000
GSIE Flottes									
SIS Flottes									
					SBTPSINI				
AZUR Flottes									
Auto particuliers (GSIE et Azur)									

Information disponible  
 Information non disponible

Les informations disponibles dans l'inventaire permanent sont les suivantes :

- Numéro de dossier sinistre
- Garantie (RC corporelle)
- Numéro de contrat
- Numéro de produit
- Date d'ouverture du sinistre
- Date de survenance du sinistre
- Date de classement du sinistre s'il est clos
- Marché (flottes / auto particuliers)
- Sous-Marché (TPPC, TPM, TPV, LCD)
- Réseau (agents, courtage, MAAF, SMABTP, Auxiliaire)
- Règlement, évaluation, recours encaissés

Nous précisons ici la notion de sinistre qui peut inclure plusieurs victimes (gestion à l'événement). Nous n'avons donc plus le détail individuellement par victime, contrairement aux données de la base victimes.

Afin de disposer de 11 années complètes sur la totalité du périmètre, une reconstitution sinistres du système SIS est nécessaire entre 2000 et 2004. Un enrichissement de ce périmètre nous a été fourni par le service informatique. Nous avons ensuite effectué plusieurs retraitements des données pour les rendre exploitables (format, vision annuelle, traitement des annulations de recours et règlements, suppression des erreurs notamment un sinistre évalué à 19M€).

Par ailleurs, nous souhaitons distinguer les véhicules de 1<sup>ère</sup> catégorie sur lesquels se portera notre étude. Cette information n'est pas disponible directement dans les fichiers sinistres et, selon les SI ou les générations de produits, elle n'est pas identifiable de la même manière. Nous avons conçu un algorithme permettant de retrouver la catégorie du véhicule pour chaque sinistre.

Cette étape s'est décomposée en plusieurs séquences :

1/ sinistres GSIE et SIS :

- a/ lien avec le fichier « véhicules » dont nous avons un historique depuis 2005,
- b/ pour les véhicules dont l'information n'a pas été trouvée (erreur dans la saisie de l'immatriculation ou remorque), étude des garanties pour repérer les véhicules de +/- 3,5 tonnes,

- c/ les véhicules non qualifiés lors des étapes précédentes et appartenant à un contrat composé de plus de 80% de 1<sup>ère</sup> catégorie, sont affectés en 1<sup>ère</sup> catégorie.

2/ sinistres AZUR : l'historique véhicule n'a pas été suffisamment conservé.

- a/ identification des produits ou familles dédiés aux flottes composées de véhicules de 1<sup>ère</sup> catégorie,
- b/ pour les véhicules dont l'information n'a pas été trouvée, un lien avec le fichier « véhicules » est fait permettant d'en retrouver 10%,
- c/ idem étape c précédente.

3/ nous avons vérifié manuellement les 100 plus gros sinistres afin de valider la catégorie trouvée.

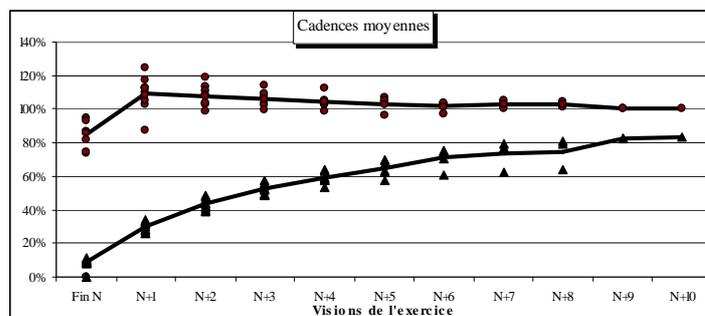
Pour chaque sinistre, nous avons la vision à fin d'année des règlements, recours et évaluations. Enfin, nous avons travaillé en charge brute de conservation et avons exclu certains sinistres reconstitués (effet des convergences) pour lesquels nous n'avons pas toutes les visions historiques.

#### IV - D . 2. MISE « AS IF »

Dans le cadre de cette étude pour la mise « as if » des sinistres nous allons retenir l'hypothèse forte d'une inflation constante (passée et future) de 4%, qui correspond à la moyenne des indices cités en IV-C.2. Cette hypothèse facilite notre étude mais n'obère en rien la portée des méthodes choisies.

Nous appliquons ensuite la méthode Chain Ladder sur les sinistres actualisés pour obtenir les cadences. Une hypothèse préalable est nécessaire : comme nous avons seulement 11 années d'historique sinistres, nous considérerons que la durée de développement est de 11 ans. Il existe cependant des méthodes de prolongement au delà (exemple courbe d'Hoerl).

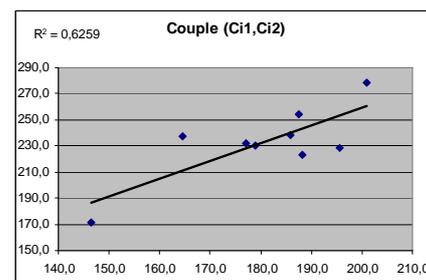
Les sinistres n'étant pas gérés de la même manière selon les équipes gestionnaires, l'estimation des cadences a été faite par SI. Pour l'illustrer, le graphique ci-dessous illustre les cadences moyennes d'évaluation et de règlement par année tous SI confondus. Les points représentent les coefficients de passage entre la vision considérée et l'ultime.



On observe une volatilité autour de la moyenne. D'autre part, l'ajustement sur les coefficients de passage entre les  $C_{i,j}$  et les  $C_{i,j+1}$  ne donne pas de bons résultats.

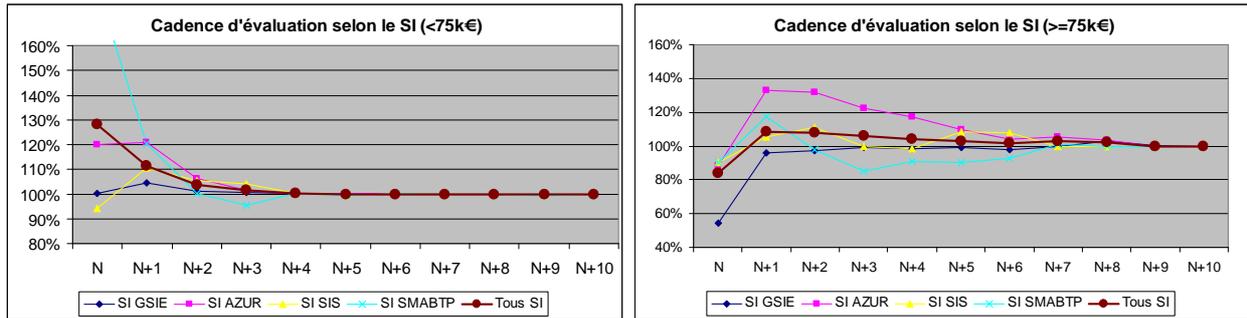
Par exemple pour  $j=1$ , le coefficient de détermination n'est que de 0,63.

Tous les points ne sont pas alignés, l'hypothèse sous-jacente à Chain Ladder n'est pas vérifiée.



Enfin les sinistres graves ont un déroulement plus long que les sinistres non graves et n'ont pas la même évolution dans le temps. Nous distinguerons donc aussi les cadences selon le coût du sinistre (inférieur ou supérieur à 75k€, seuil utilisé actuellement).

Les cadences retenues de coûts par SI sont les suivantes :



Nous constatons encore une fois que le SI et plus particulièrement la gestion ont une influence sur la cadence de liquidation d'un sinistre, même sur ceux de montant inférieur à 75k€. Nous utiliserons ces cadences sur les sinistres pour obtenir leur coût ultime.

### IV - D . 3. DETERMINATION DU SEUIL

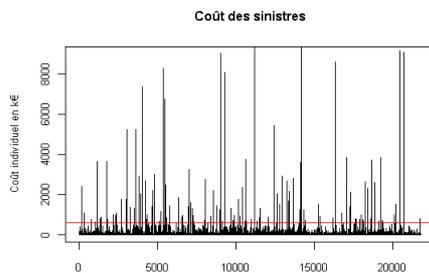
Maintenant nous allons appliquer les méthodes de détermination du seuil sur le coût ultime 2020 calculé précédemment. Pour cela nous avons utilisé les logiciels R et SAS.

Le périmètre utilisé correspond aux sinistres survenus de 2007 à 2009, hors sinistres IRCA à 1 490 € (23% des sinistres). Ils ont été exclus de la base car ils avaient un fort impact sur la méthode de la Gertensgarbe Plot. Concernant les autres méthodes, il n'y a pas d'effet car on s'intéresse au nombre d'excès.

#### □ Etude préalable des données

Nbre obs.	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
21 843	1	1 430	3 540	27 190	10 000	9 605 000

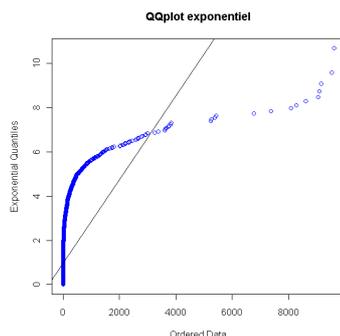
Les coûts des sinistres s'étalent de 1€ à 9,6M€, la médiane est très nettement inférieure à la moyenne, ce qui est typique d'une distribution ayant beaucoup de petits sinistres et quelques uns très importants. Le quantile à 75% est très faible par rapport au sinistre maximum.



On a ordonné les sinistres selon leur date d'ouverture. Le graphique des coûts individuels ci-contre nous permet de vérifier que l'on n'a pas de point atypique (sinistre de coût nettement supérieur aux autres), la majorité des sinistres a un coût inférieur à 600k€.

#### □ Vérification de la présence d'une queue épaisse

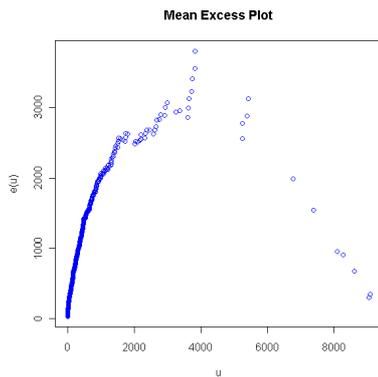
Pour analyser le comportement de nos données, nous traçons le QQ-plot exponentiel.



Graphique QQ-plot Exponentiel (coûts en k€)

On remarque que le QQ-plot exponentiel obtenu a une forme convexe, ce qui signifie que la distribution est à queue plus épaisse que la loi exponentielle. **Les données suivent donc une loi à queue épaisse** comme la Pareto, la log-normale ou la GPD.

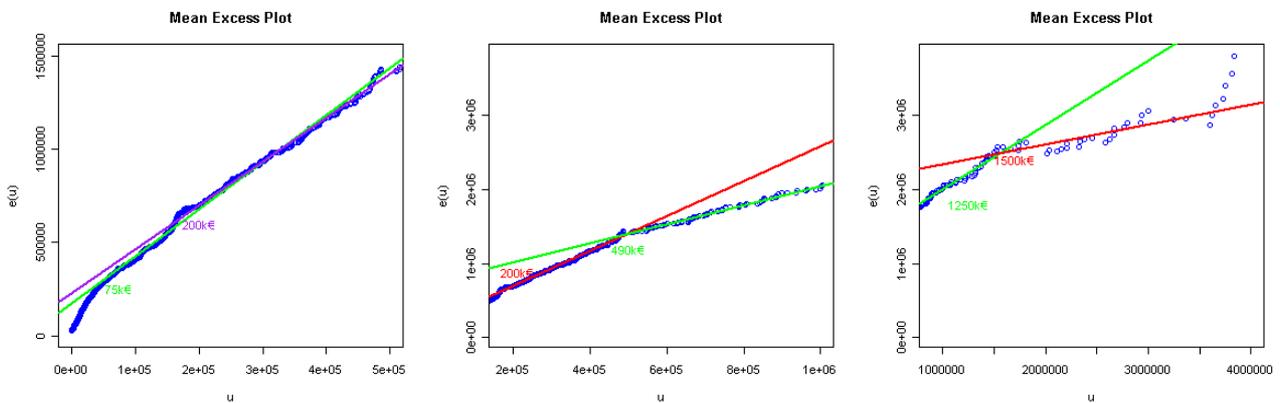
On va maintenant utiliser le Meplot (FME) pour identifier de quelle loi il s'agit.



Graphique Meplot (en k€)

Le mean excess plot est croissant, hormis pour les dernières observations où l'on observe une certaine instabilité liée à la méthode. Cela confirme que la distribution est à **queue épaisse**.

De plus ce graphique est intéressant, car il nous permet d'avoir une idée sur le seuil d'entrée dans la région extrême. En effet, la valeur où la courbe commence à se comporter comme une droite correspond au seuil  $u$ . Aussi on va faire un zoom sur différentes parties :

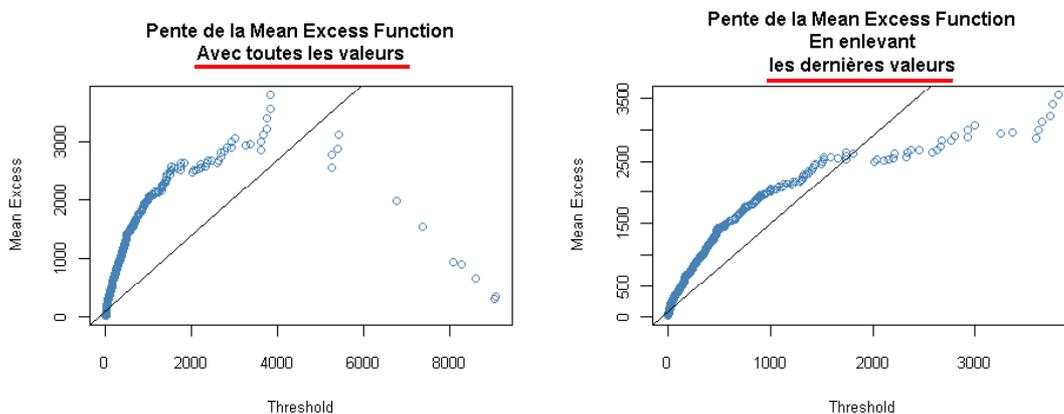


Il ne faut pas attendre de cette technique d'obtenir la valeur exacte de  $u$ , elle fournit néanmoins une aide précieuse. En pratique, plusieurs valeurs de  $u$  doivent être testées. Nous observons plusieurs plages de linéarité, qui pourraient correspondre à des seuils : 75k€, 200k€, 490k€, 1,25M€ et 1,5M€.

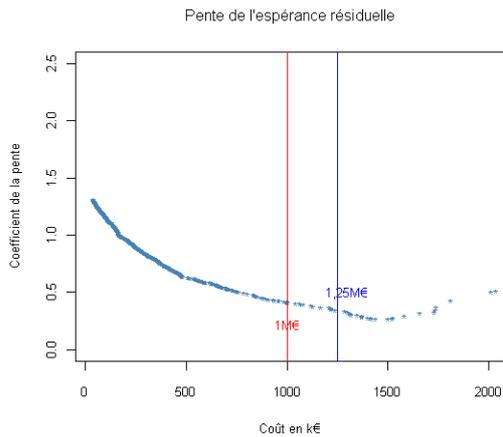
□ **Méthodes de détermination du seuil**

a. **Pente de l'espérance résiduelle**

Si on prend toutes les données, les dernières valeurs vont avoir un impact sur la droite de régression :



Pour étudier l'évolution de la pente de l'espérance résiduelle, nous allons donc omettre les dernières valeurs.



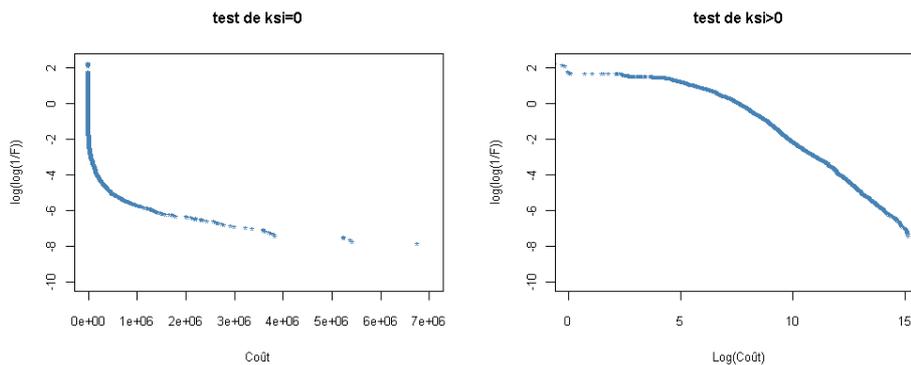
Sur le graphique de la pente résiduelle ci-contre, nous observons une entrée dans la zone de stabilité vers 1M€.

Cette méthode n'est pas optimale et a l'inconvénient d'être influencée par les dernières valeurs, d'ailleurs le coefficient de la pente (0,4) nous semble assez faible. On voit en effet par la suite que l'indice de queue se situe vers 0,9 et nous aurions aimé que le coefficient de la pente soit proche de cette valeur.

Le mean excess plot est instable sur les dernières observations, il est donc nécessaire d'omettre les valeurs que l'on considère comme atypiques, et ce, avec discernement. Notre choix a un impact direct sur le coefficient de la pente, par contre cela ne modifie pas l'entrée en zone de stabilité observée.

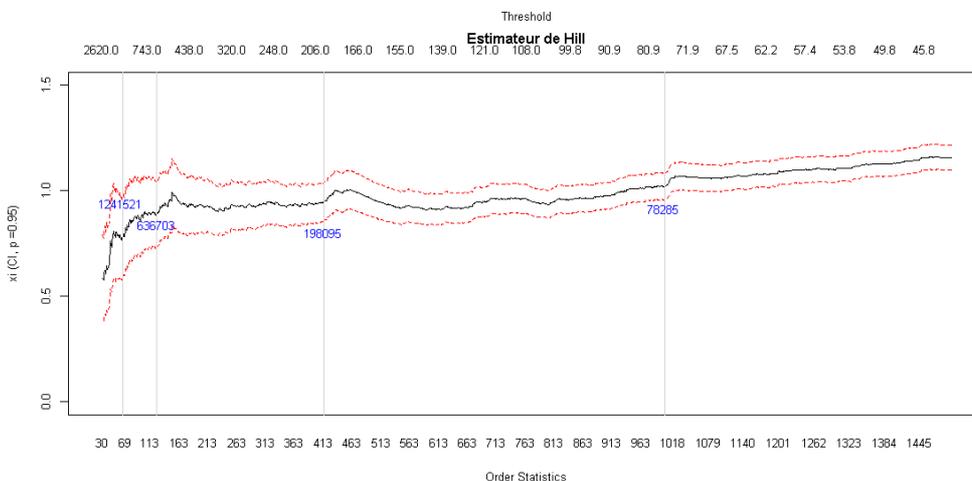
### b. Estimateur de Hill

Pour pouvoir utiliser l'estimateur de Hill, il est indispensable de savoir si l'on est dans le cas où  $\zeta=0$  (cas Gumbel) ou  $\zeta>0$  (cas Fréchet). Nous traçons donc les graphiques suivants :



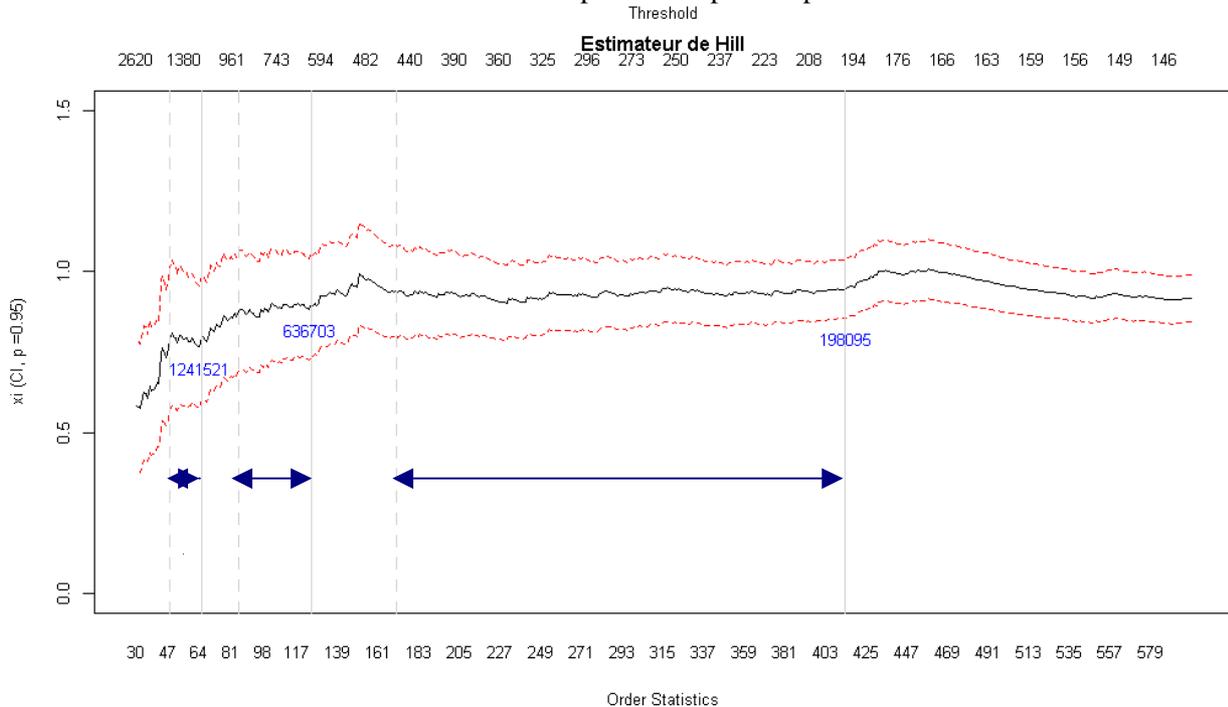
Ces graphiques ne permettent pas de différencier nettement le cas  $\zeta>0$  du cas  $\zeta=0$ . La première courbe est concave, ce qui nous permet de **rejeter l'hypothèse  $\zeta=0$** . D'autre part, la fin de la deuxième courbe peut être assimilée à une droite, **l'hypothèse  $\zeta>0$  semble donc légitime**.

Les données appartenant au domaine de Fréchet, nous pouvons utiliser l'estimateur de Hill. Le graphique suivant représente cet estimateur selon le nombre d'excès considérés. Nous l'avons tracé à partir du 30<sup>ème</sup> excès, afin d'enlever les plus grandes valeurs et ainsi avoir une échelle nous permettant de mieux analyser le comportement de l'estimateur.



L'estimateur de Hill est très volatil au début de la courbe car le nombre d'excès considéré est insuffisant. Il commence à être plus stable après le 45<sup>ème</sup> excès et à être très stable à partir d'environ 170 excès. Après 1000 excès l'estimateur semble changer de comportement ; sur la partie inférieure il n'est pas évident d'identifier précisément la zone de stabilité. Dans un premier temps nous avons identifié différents seuils : 78k€, 198k€, 637k€ et 1,25M€.

Nous allons faire un zoom sur le début des excès pour avoir plus de précision :



Nous observons plusieurs zones de stabilité, une première très courte entre 1,25M€ et 1,6M€, mais elle comprend très peu de sinistres. Nous en avons ensuite une autre entre 640k€ et 910k€ et une dernière très stable entre 200k€ et 450k€.

Les valeurs de l'estimateur de Hill pour différents excès sont les suivantes :

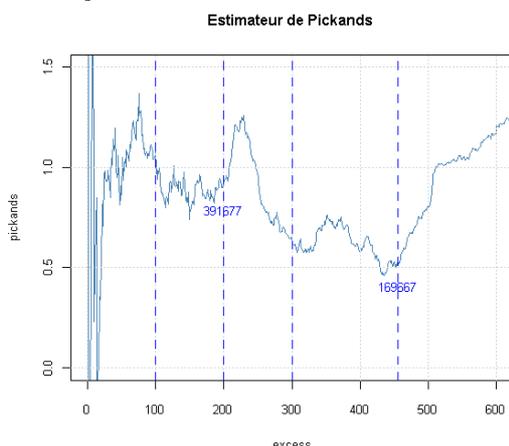
Nombre d'excès	65	100	124	150	275	413	800	1 004	1 500
Montant	1,24M€	787k€	637k€	485k€	295k€	198k€	106k€	78k€	44k€
$\xi$	0,79	0,88	0,90	0,99	0,92	0,94	0,94	1,02	1,16

Les valeurs  $\xi$  à différents excès sont en majorité comprises entre 0,88 et 0,94.

Nous retenons que les valeurs de seuil les plus intéressantes seraient 78k€, 198k€, 637k€ et 1,24M€.

### c. Estimateur de Pickands

Nous traçons maintenant l'estimateur de Pickands :



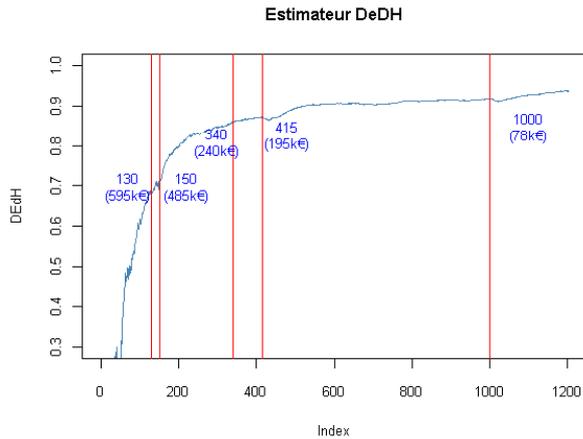
L'estimateur de Pickands est plus volatil que l'estimateur de Hill car il s'appuie sur moins d'observations ce qui le rend plus sensible aux variations des valeurs.

Il tend à être un peu moins volatil entre 100 et 200 excès (400k€ à 700k€) puis entre 300 et 450 excès (170k€ à 270k€). Par ailleurs l'estimation de l'indice de queue paraît moins robuste que celle obtenue avec Hill (par exemple :  $\xi=1$  à 100 excès ;  $\xi=0,5$  à 450 excès).

Nous n'utiliserons donc pas le résultat obtenu avec cette méthode.

**d. Estimateur de Dekkers et al. (estimateur des moments)**

Nous traçons maintenant l'estimateur des moments de Dekkers et al.

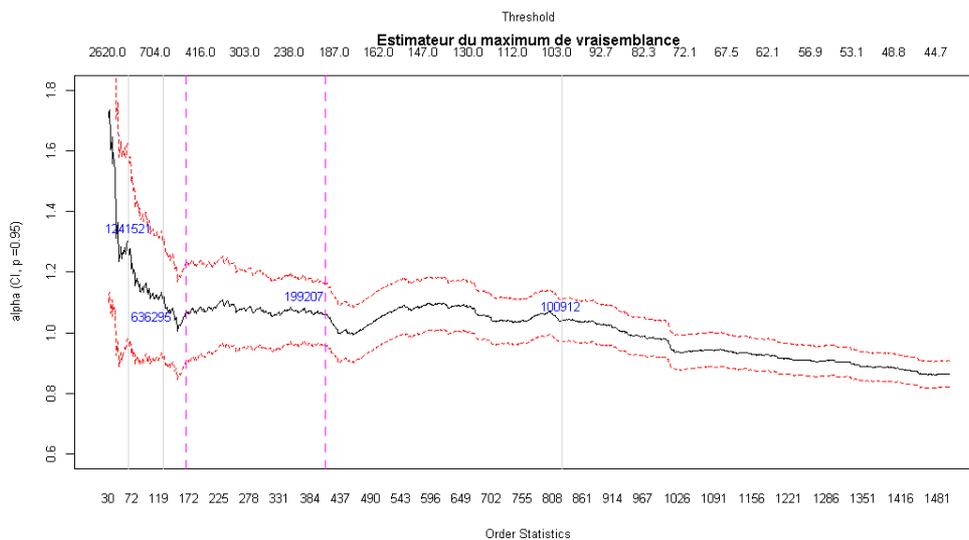


Nous avons une petite zone de stabilité entre les excès 130 et 150, mais compte-tenu du faible nombre de valeurs qu'elle contient, nous ne la retiendrons pas.

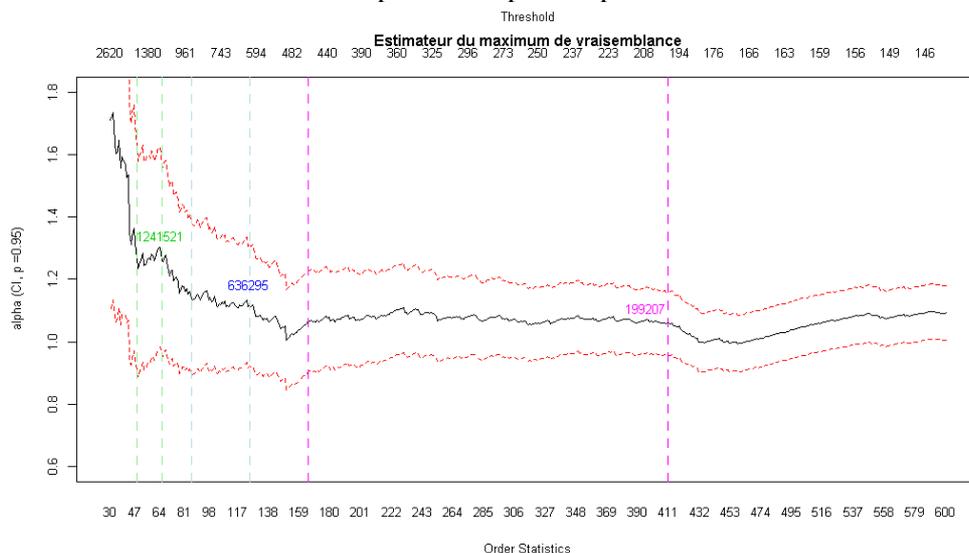
Comme précédemment, nous retrouvons une zone de stabilité avant 415 excès (195 k€) et une autre avant 1 000 excès (78 k€).

**e. Estimateur du maximum de vraisemblance**

Le graphique suivant représente l'estimateur du maximum de vraisemblance selon le nombre d'excès considérés.



Nous faisons un zoom sur le début des excès pour avoir plus de précision :



Globalement, nous faisons les mêmes observations que pour l'estimateur de Hill, ce qui est logique (voir IV-

B.6.). Nous remarquons une très grande volatilité de l'estimateur lorsque le nombre d'excès considérés est insuffisant. Nous observons plusieurs zones de stabilité, une première très courte entre 1,24M€ et 1,6M€, mais elle comprend très peu de sinistres. Nous en avons ensuite une autre entre 636k€ et 900k€ et une autre plus longue entre 200k€ et 470k€. Après 410 excès (100k€), l'estimateur change de comportement et diminue.

Les valeurs de l'estimateur de  $\alpha$  pour différents excès sont les suivantes :

Nombre d'excès	48	65	85	125	165	275	410	1 000
Montant	1,6M€	1,24M€	912k€	636k€	467k€	295k€	199k€	78k€
$\alpha$	1,27	1,26	1,14	1,12	1,07	1,08	1,06	0,98
$\xi=1/\alpha$	0,79	0,79	0,88	0,89	0,94	0,92	0,94	1,02

Nous retenons que les valeurs de seuil les plus intéressantes seraient 100k€, 199k€, 637k€ et 1,24M€.

### f. Fonction minAMSE de R

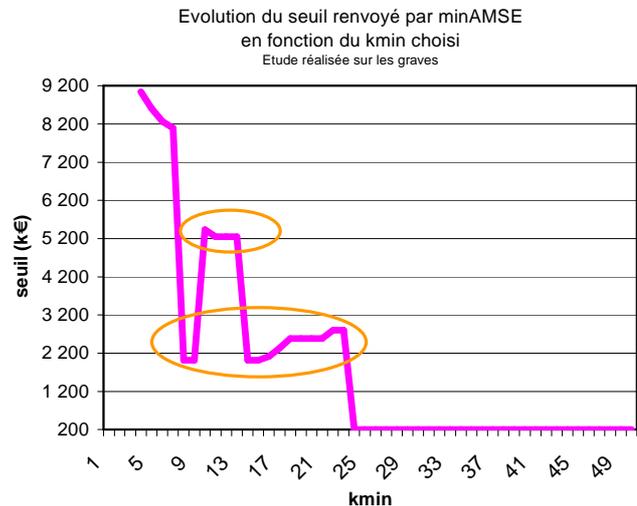
On choisit d'utiliser cette fonction **en détection** (recherche d'un seuil sans a priori) et **en contrôle** (recherche d'un seuil dans une zone ciblée par la méthode de l'estimateur de Hill), sur les **données empiriques** et sur des **données simulées** en réutilisant la loi de Pareto tronquée du §V.D.2.

Avec 2 entiers  $k_{min}$  et  $k_{max}$  où  $\frac{n}{100} < k_{min} < k_{max} < n$

désignant le nombre de sinistres, on définit l'intervalle  $[n-k_{max} ; n-k_{min}]$  dans lequel l'algorithme va chercher le seuil.

En faisant varier  $k_{min}$ , on constate qu'on "n'accroche" pas un seuil unique. L'exemple ci-joint pour les sinistres supérieurs à 200 k€ le montre bien.

Sauf à considérer que les 2 zones plateau (en orange) sont des seuils ; en retenant 5 000 k€ (priorité que la réassurance nous propose pour l'avenir) et 2 500 k€ (priorité actuelle), on sera alors en terrain connu.



La valeur  $k_{min}$  étant influente et sensible, nous choisissons l'ajustement suivant : retenir un seuil quand son rang est différent de  $k_{min}$ .

#### Exemples :

*minAMSE(graves,,7,,) Optimal k : 7 Scale parameter : 8 089 548. On refuse le seuil car  $k_{min} = \text{optimal } k$*

*minAMSE(graves,,8,,) Optimal k : 41 Scale parameter : 2 015 375. On accepte le seuil car  $k_{min} << \text{optimal } k$*

L'arbitrage consiste alors à exclure le moins de sinistres extrêmes possibles tout en vérifiant que la fonction renvoie un résultat cohérent.

Les résultats obtenus sont rassemblés dans le tableau ci-dessous :

	Echantillon	Seuil renvoyé
1	Tous les sinistres hors IRCA	1
2	Sinistres > 75 k€ (seuil actuel)	2 015 375
3	Sinistres > 150 k€ (seuil des graves du groupe)	2 112 629
4	Sinistres > 200 k€ (seuil des graves Hill)	2 015 375
5	Sinistres > 650 k€ (seuil des graves Hill)	1 527 175
6	Sinistres > 1 250 k€ (seuil des graves Hill)	2 327 647
7	Moyenne des simulations de la Pareto tronquée <sup>39</sup>	6 426 115

Distribution des seuils renvoyés lors des simulations

**Analyse :**

On constate l'échec de la méthode sur l'échantillon 1. Sur les échantillons 2, 3 et 4 une relative stabilité des résultats laisse à penser que les sinistres entre 75 k€ et 200 k€ sont relativement homogènes. L'échantillon 5 renvoie un seuil bien différent des précédents ; les sinistres entre 200 k€ et 650 k€ qui ne sont alors plus pris en compte devaient jouer un rôle modérateur favorisant la convergence du résultat. L'échantillon 5 doit donc être hétérogène ; un seuil remarquable doit se trouver entre 200 et 650 k€. Pour l'échantillon 6, le phénomène est similaire au 5 mais plus atténué ; cet échantillon doit être plus homogène. La valeur 1 250 k€ peut donc être un seuil significatif. Quant à l'échantillon 7, on doit constater l'échec que l'on pense lié à l'absence d'ajustements et à l'utilisation d'une loi de Pareto tronquée.

**Conclusion :**

On obtient des résultats relativement convergents quand on procède à des ajustements. On peut retenir le seuil arrondi de 2 000 k€. L'ajustement vient d'ailleurs amplifier le biais de la méthode qui s'affranchit déjà de 1% des sinistres les plus graves pour fonctionner.

Ce seuil trouvé est très élevé ; seuls 42 sinistres (soit 0,19% en nombre) y sont supérieurs, représentant 31% en montant. On peut néanmoins se demander si ce faible nombre est suffisant pour figer un seuil significatif.

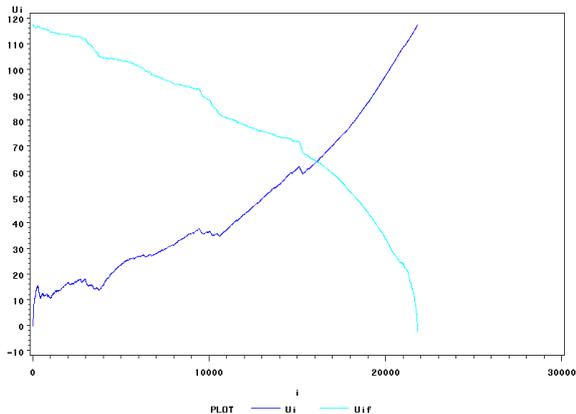
**Ce montant de 2 000k€ est compatible avec la priorité actuelle de cession en réassurance.** C'est d'ailleurs cohérent avec le support parétien de la fonction minAMSE qui étudie les valeurs extrêmes. En creux, cette étude nous a permis de nous conforter sur les seuils pressentis (200 k€ et 1 250 k€) et de découvrir le logiciel R qui deviendra par la suite notre principal outil de travail.

<sup>39</sup> 1000 simulations ont été effectuées, chacune comportant 1000 sinistres bornés à 20 000 k€, tirés dans la Pareto tronquée de paramètre estimé 1,0611 et de seuil 200 k€. Le détail du programme créé sous R pour la simulation figure en annexe VIII E.

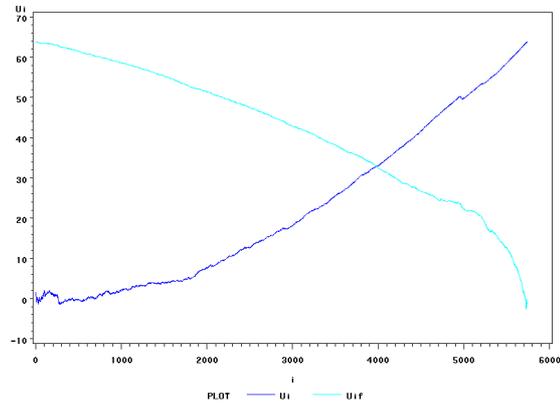
**g. Gertensgarbe Plot**

En appliquant cette méthode sur les données y compris les sinistres IRCA (23% des sinistres), on obtenait un seuil assez bas de 3 529€. Donc un nombre important de sinistres ayant le même montant impacte la méthode. En enlevant ces sinistres IRCA, nous obtenons alors un seuil plus élevé de 9 534€.

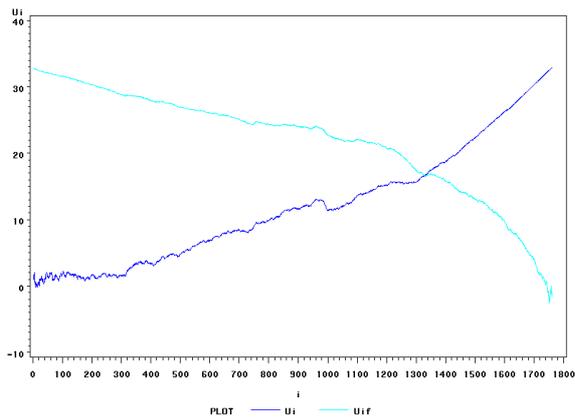
Toutefois, ce seuil semble encore assez bas. Par itération nous allons réappliquer la même méthode aux données supérieures au seuil. Nous obtenons les graphiques suivants :



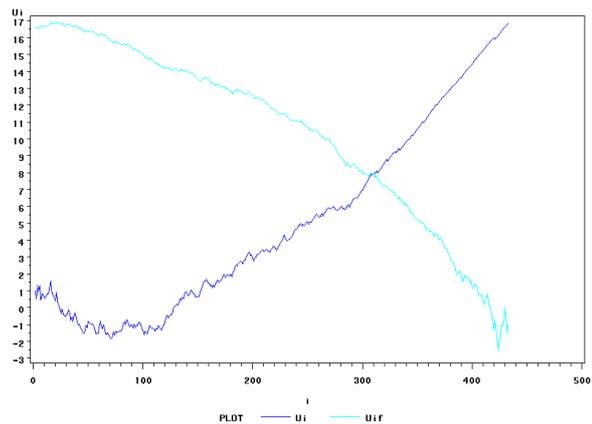
Sur tous les sinistres hors IRCA - Seuil = 9 534€



Sur les sinistres > 9 534€ - Seuil = 35 900€



Sur les sinistres > 35 900€ - Seuil = 187 477€

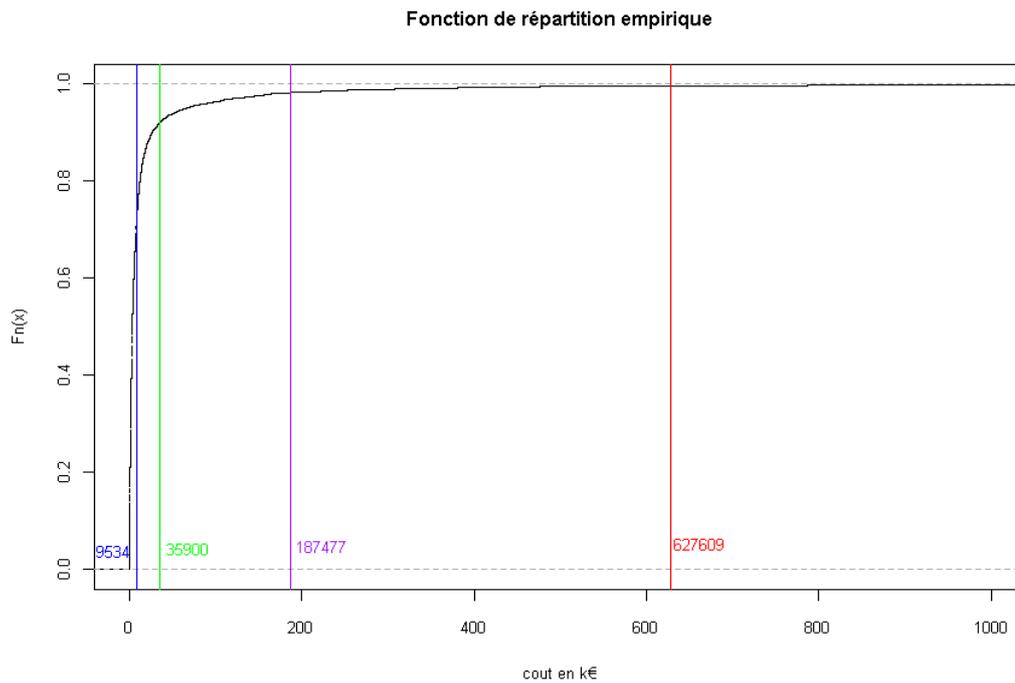


Sur les sinistres > 187 477€ - Seuil = 627 609€

Nous compilons les résultats :

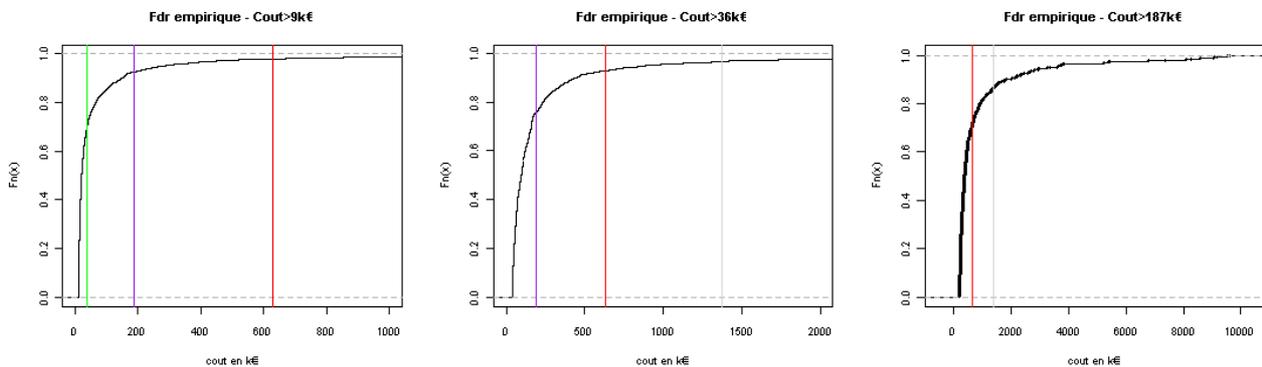
Tous les sinistres 2007-2009	Avec suppression du forfait IRCA (2007-2009)
3 529 €	9 534 €
21 743 €	35 900 €
100 782 €	187 477 €
362 671 €	627 609 €
889 552 €	1 375 075 €
2 040 522 €	1 735 684 €
2 93 0638€	2 327 646 €
5 24 8300€	2 663 077 €
	2 671 024 €
	3 753 025 €

Sur la f. de r., nous représentons les seuils obtenus avec suppression du forfait IRCA :



Les seuils définis par cette méthode traduisent un changement de comportement de la fonction empirique.

Nous représentons ci-dessous les seuils sur la f. de r. des données sélectionnées :



Nous observons encore une fois que les seuils renvoyés par la méthode correspondent à des changements de comportement.

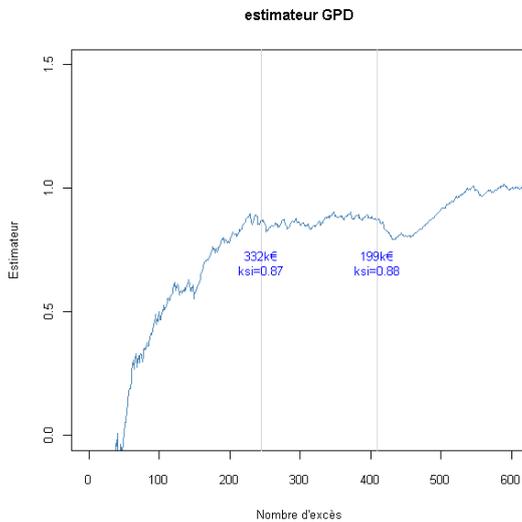
Pour évaluer la robustesse de cette méthode nous avons regardé les résultats selon des seuils fixés :

Coût >	Seuil obtenu
36 000€	195 205€
37 000€	202 475€
39 000€	210 816€
160 000€	453 357€
170 000€	588 089€
175 000€	581 769€
200 000€	591 247€

Les résultats de la méthode Gertensgarbe sont assez variables, cependant ils nous donnent une idée des **endroits de rupture**.

### h. Estimateur GPD

Nous pouvons aussi estimer les paramètres de la GPD sur la queue de distribution. A partir d'un certain nombre d'excès, si la queue de distribution suit une GPD, nous devrions avoir une stabilité des paramètres de la loi<sup>40</sup>.



Nous observons une stabilité de l'estimateur de GPD entre 200k€ et 330k€, ce qui est assez concordant avec nos précédentes analyses.

### i. Conclusion

Les seuils retenus à partir des différentes méthodes sont rassemblés dans le tableau suivant :

Mean Excess Function	Pente	Hill	Pickands	DEdh	EMV	MinAMSE	Gertensgarbe	Estimateur GPD
							9,5k€	
							36k€	
75k€		78k€		78k€				
200k€		198k€		195k€	100k€			
490k€		637 k€			637k€		628k€	
1,25M€	1M€	1,24M€			1,24M€			
1,5M€	1,25M€					2 M€	1,38M€	

Nous retrouvons pour quasiment toutes les méthodes le seuil de 200k€, aussi c'est celui-ci que nous retiendrons pour distinguer les sinistres attritionnels des sinistres graves. Pour séparer les sinistres graves ordinaires et non ordinaires, nous utiliserons le seuil de 1,25M€.

Avec ces nouveaux seuils, la répartition des sinistres est la suivante :

Tranche de coût	% Charge	% Nbre
<200 000 €	38%	98,13%
>200 000 et <1,25M€	25%	1,57%
>1,25M€	37%	0,29%

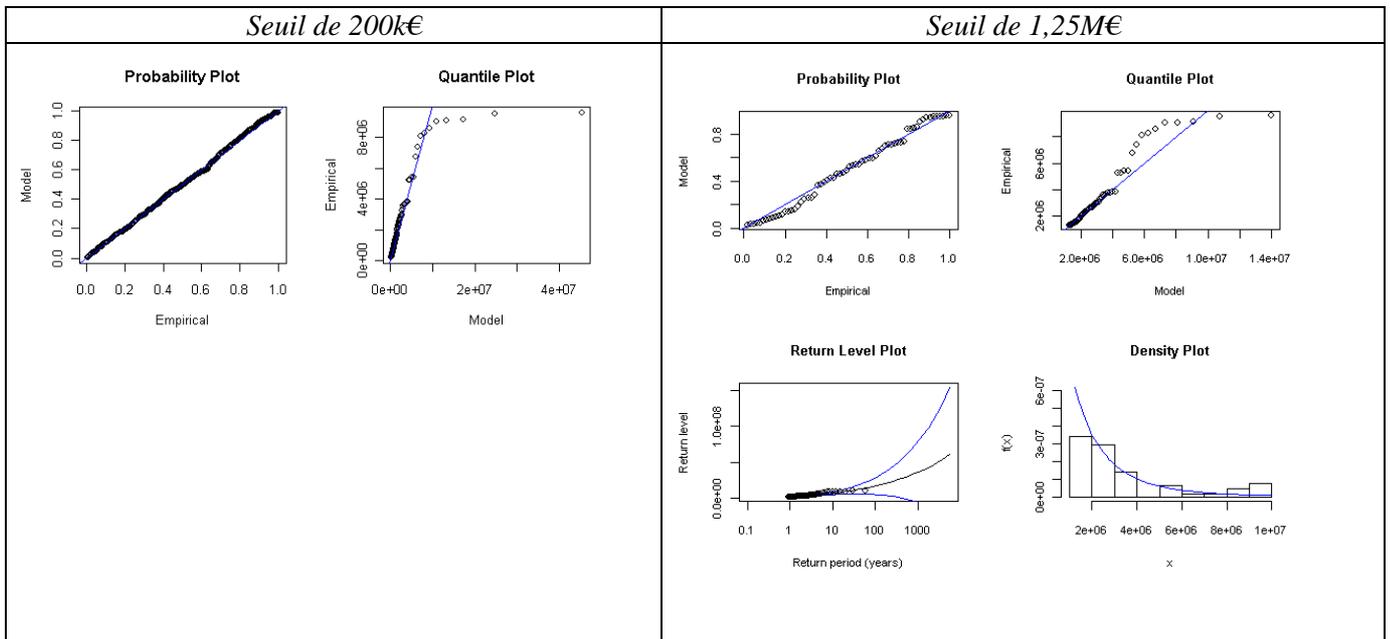
*sinistres 2007-2009 (hors sans-suite et forfaits)*

On constate que les 2 seuils retenus séparent la charge de manière équilibrée. Pour remarque, les sinistres inférieurs au seuil actuellement utilisé de 75 k€ représentent 95% des nombres et 26% de la charge.

<sup>40</sup> Nous avons utilisé la fonction `gpd.fit` de R (Maximum de vraisemblance avec optimisation par la méthode de Nelder-Mead).

**j. Confirmation des seuils par l'analyse GPD :**

Nous allons tester avec le module extRemes de R si une loi GPD de seuil  $u = 200 \text{ k€}$  ou  $u = 1,25 \text{ M€}$  s'adapte à nos données. Nous obtenons les PP-Plot et QQ-plot suivants :



Pour le seuil  $u = 200 \text{ k€}$ , sur le graphique PP-Plot, tous les points se situent sur la bissectrice donc l'adéquation est bonne. C'est aussi le cas du graphique QQ-plot où seuls les derniers points ne sont pas sur la droite. **Nous pouvons donc considérer notre seuil correct.**

Pour le seuil  $u = 1,25 \text{ M€}$ , sur le graphique PP-Plot nous avons quelques points en dessous de la bissectrice, mais ils sont toutefois assez proches. Sur le graphique QQ-plot, nous observons que les données supérieures à  $6 \text{ M€}$  (10 observations) se décrochent de la droite. **Nous pouvons donc considérer ce seuil acceptable.**

## IV - E CONCLUSION

---

L'approche retenue repose sur un certain nombre d'hypothèses fortes pour la mise « as if » des sinistres, à savoir :

- l'hypothèse d'une inflation constante. Nous aurions pu utiliser un modèle d'évolution de l'inflation comme celui de Wilkie, ou encore séparer l'inflation pour les sinistres attritionnels (évolution indice des prix) et pour les sinistres graves (évolution jurisprudence).
- l'hypothèse de développement des sinistres sur 11 ans liée au manque d'historique. Cette limite aurait pu être résolue en utilisant des méthodes de prolongement (courbe d'Hoerl).
- pour les rentes, le retraitement de l'inflation sur le capital constitutif et non selon les cash-flow de règlements.

La recherche du seuil grâce aux estimateurs vus précédemment se fait sur des bases théoriques fondées. Il n'en reste pas moins que son choix par une observation visuelle présente des limites et que sa valeur exacte reste difficile à déterminer. Nous avons comparé les différentes méthodes et avons essayé de trouver un seuil cohérent entre elles. Nous avons retenus 200 k€ pour distinguer les sinistres attritionnels des sinistres graves et 1,25 M€ pour séparer les sinistres graves ordinaires des non ordinaires.

D'autres méthodes<sup>41</sup> non étudiées ici existent pour trouver le seuil optimal. Par exemple, Resnick et Starica (1997) ont mis au point une méthode pour effectuer une sorte d'agrandissement via un changement d'échelle logarithmique du démarrage du graphe de l'estimateur de Hill. On peut aussi utiliser la méthode du bootstrap qui permet d'estimer la précision des estimateurs de paramètres, comme la variance, sans connaître la vraie valeur de cette précision et même une expression de cette précision. Cette méthode procède par ré-échantillonnage (Efron et Tibshirani, 1993, Ardilly, 1994).

Maintenant que l'on connaît les seuils, nous allons voir dans la partie suivante comment modéliser la sinistralité attritionnelle et la sinistralité extrême.

---

<sup>41</sup> Voir pour l'application les mémoires « Événement extrêmes sur les spreads de crédit » (Claire Gauthier, Nathalie Pistre) et « Les sinistres graves en assurance automobile : Une nouvelle approche par la théorie des valeurs extrêmes Nouredine » (Benlagha, Michel Grun-Réhomme, Olga Vasechko).

Les résultats des deux études précédentes sont concluants : nous pourrions utiliser une base sinistres élargie pour la méthode 2 et nous avons observé une convergence de valeurs pour les seuils nous permettant d'affiner nos définitions de graves ordinaires et non ordinaires.

Ayant éclairé les principaux éléments de contexte (juridique, risque routier, gestion des sinistres) et réuni les outils nécessaires, il nous est désormais possible d'élaborer les modèles de charge. Nous commençons par la méthode probabiliste et poursuivrons par une méthode individuelle de matrices de transition.

## V - METHODE 1 : MODELE PROBABILISTE DE LA CHARGE RC CORPORELLE

### Introduction

Nous utiliserons dans cette partie la méthode probabiliste via un modèle paramétrique et résumons ci-dessous les différentes étapes permettant de calculer la charge ultime RC corporelle nette de réassurance.

Type de sinistres	Modélisation		Simulation <sup>(1)</sup>	
	Loi de fréquence	Loi de sévérité	Charge brute de réassurance	Charge nette de réassurance
attritionnels	V-E.1	V-E.2	V-E.3 <sup>(2)</sup>	
graves ordinaires		V-E.2	V-E.3	V-E.3
graves non ordinaires		Loi mélange		
tous			V-E.3	V-E.3

(1) méthode Monte-Carlo voir pages 91 à 96.

(2) les attritionnels ne faisant pas l'objet de cession en réassurance, leur charge brute égale la nette.

Les particularités du modèle recherché résident dans :

- la reconstitution de la charge totale, elle-même étant la résultante de trois modélisations,
- la proximité de ces résultats avec les données réelles (qualité),
- une simulation composée par les tirages aléatoires de nombres et de leurs montants associés,
- la prise en compte d'un programme de réassurance.

Les étapes de modélisation s'appuient sur la détermination de lois et de leur(s) paramètre(s), estimé(s) selon la méthode la plus universelle, celle du maximum de vraisemblance.

Les étapes de simulation reposent principalement sur la théorie du modèle de risque collectif.

Nous présentons donc en introduction un rappel de ces deux fondamentaux de l'assurance non vie.

### Maximum de vraisemblance

On se place dans le cas d'un modèle paramétrique.

Soit un échantillon aléatoire  $(X_1, X_2, \dots, X_n)$  dont la loi mère appartient à une famille paramétrique de densités ou de fonctions de probabilité  $\{f(x; \theta), \theta \in \Theta\}$  où  $\Theta \subseteq \mathcal{R}^k$ . On appelle **fonction de vraisemblance** de  $\theta$  pour une réalisation donnée  $(x_1, x_2, \dots, x_n)$  de l'échantillon, la fonction de  $\theta$  :

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

On choisira pour  $\theta$  la valeur la plus vraisemblable notée  $\hat{\theta}^{MV}$  telle que la loi  $f(x; \hat{\theta}^{MV})$  correspondante confère la plus forte probabilité ou densité de probabilité aux observations relevées.

En pratique, on a plutôt intérêt à maximiser  $\ln L(\theta)$  appelée **log-vraisemblance**.

L'estimation par le maximum de vraisemblance vérifie :  $\frac{\partial}{\partial \theta} \ln L(\theta) = 0$  ou  $\frac{\partial}{\partial \theta} \ln \left[ \prod_{i=1}^n f(x_i; \theta) \right] = 0$  ou

$$\sum_{i=1}^n \frac{\partial}{\partial \theta} \ln f(x_i; \theta) = 0 \text{ (équation de vraisemblance).}$$

L'estimateur du maximum de vraisemblance est défini, unique et calculable pour les lois habituelles. Il est démontré qu'il est asymptotiquement sans biais et convergent.

### Le modèle collectif

Le modèle de risque collectif considère le montant total des sinistres d'un portefeuille composé de plusieurs polices homogènes ; une police pouvant générer plusieurs sinistres.

Soit  $S$  la variable aléatoire réelle qui représente la charge totale des sinistres du portefeuille sur la période étudiée.  $S$  est déterminée par le nombre total de sinistres et par les montants de chacun de ces sinistres.

Ainsi, on pose :

- $N$  : la variable aléatoire réelle à valeurs entières qui représente le nombre des sinistres survenant sur le portefeuille dans la période étudiée,
- $(X_i)_{i \geq 1}$  : la suite de variables aléatoires réelles représentant chacune le montant d'un sinistre.

$$\text{On a alors : } S = \sum_{i=1}^N X_i$$

Nous ferons également les deux hypothèses suivantes :

- **indépendance des nombres et des montants** : les variables aléatoires réelles  $N$  et  $(X_i)_{i \geq 1}$  sont indépendantes.
- **indépendance et équidistribution des montants de sinistre** : les coûts des sinistres successifs  $X_1, X_2, \dots, X_i, \dots$  sont indépendants et de même loi que la variable aléatoire réelle parente  $X$ .

Ainsi, en supposant que  $N$  et  $X$  admettent des moments d'ordre 1, nous avons  $E(S) = E(N)E(X)$

## V - A MODELISATION DU NOMBRE DE SINISTRES

Nous rappelons ici les principales caractéristiques des 3 distributions usuelles servant à la modélisation du nombre ; voir l'annexe VIII-A pour un rappel sur le calcul des moments.

### Distribution binômiale $B(n, p)$

Elle a pour paramètres  $n \in \mathbb{N}^*$  et  $p \in ]0,1[$  ; on posera  $q = 1 - p$ .

Loi :  $P(N = k) = C_n^k p^k q^{n-k}$  pour  $k = 0, 1, \dots, n$  et  $P(N = k) = 0$  sinon.

Moments :  $\mu_{(k)} = n(n-1)\dots(n-k+1)p^k$  ce qui donne  $m = E(N) = np$ ,  $\sigma^2 = V(N) = npq < E(N)$ .

### Distribution de Poisson $P(\lambda)$

Elle a pour paramètre  $\lambda > 0$ .

Loi :  $P(N = k) = e^{-\lambda} \frac{\lambda^k}{k!}$  pour  $k \in \mathbb{N}$ .

Moments :  $\mu_{(k)} = \lambda^k$  d'où pour les premiers moments  $m = E(N) = \lambda$ ,  $\sigma^2 = V(N) = \lambda = E(N)$ .

### Distribution binômiale négative $BN(r, p)$

Elle a pour paramètres  $r > 0$  et  $p \in ]0,1[$  ; nous poserons  $q = 1 - p$ .

Loi :  $P(N = n) = \frac{\Gamma(r+n)}{\Gamma(r)n!} p^r q^n$ , pour  $n \in \mathbb{N}$  avec  $\Gamma$  la fonction Gamma définie par  $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$

vérifiant :  $\frac{\Gamma(r+n)}{\Gamma(r)} = (r+n-1)\dots r$  si  $n \geq 1$  et 1 si  $n = 0$ .

Moments :  $\mu_{(k)} = r(r+1)\dots(r+k-1)\left(\frac{q}{p}\right)^k$  d'où  $m = E(N) = r\frac{q}{p}$   $V(N) = r\frac{q}{p^2} > E(N)$ .

**Commentaires sur l'utilisation de ces distributions**

La distribution binômiale  $B(n, p)$  est appliquée le plus souvent lorsque le paramètre  $n$  qui borne la v.a.r. nombre de sinistres est donné. La condition  $E(N) > V(N)$  est peu rencontrée pour modéliser les nombres de sinistres dans les données empiriques en assurance non vie.

La loi de Poisson  $P(\lambda)$  est la loi fondamentale de l'assurance non vie, bien adaptée aux classes homogènes de risques et présente les avantages suivants : unique paramètre facilement compréhensible et estimable, loi à la base des distributions Poisson-mélange et de l'élaboration d'un système bonus-malus, ...  
Cependant, son unique paramètre la rend moins flexible, elle présente une queue de distribution fine et l'égalité  $E(N) = V(N)$  peut être contraignante.

Enfin, la distribution binomiale négative  $BN(r, p)$  a la particularité  $V(N) > E(N)$  qui la rend souvent utilisable en non vie. Plus flexible que la loi de Poisson grâce à ses 2 paramètres pourtant difficilement interprétables, elle est surtout utilisée pour modéliser la distribution du nombre de sinistres d'un contrat d'une classe de risques dont l'hétérogénéité est encore élevée malgré la segmentation du portefeuille.

**V - B MODELISATION DE LA SEVERITE**

---

Dans la suite nous appelons  $C$  la v.a.r. positive représentant le coût de sinistre.

**V - B . 1. LES PRINCIPALES LOIS**

**Distribution exponentielle  $\mathcal{E}(\theta)$**

Elle dépend d'un unique paramètre  $\theta > 0$ .

Densité et fonction de répartition :  $f(x) = \theta.e^{-\theta x}$  et  $F(x) = 1 - e^{-\theta x}$

Moments :  $m_k = \frac{k!}{\theta^k}$  d'où  $m = E(C) = \frac{1}{\theta}$ ,  $V(C) = \frac{1}{\theta^2}$ ,  $\gamma_1 = 2$ ,  $\gamma_2 = 6$

**Distribution log-normale  $LN(\mu, \sigma^2)$**

Elle a pour paramètres  $\mu \in \mathfrak{R}$  et  $\sigma > 0$ .

La v.a.r. coût  $C$  suit une loi log-normale  $LN(\mu, \sigma^2)$  si et seulement si la v.a.r.  $\ln C$  suit la loi normale  $N(\mu, \sigma^2)$ .

Densité et fonction de répartition :  $f(x) = \frac{1}{\sqrt{2\pi\sigma x}} e^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2}$  et  $F(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$  avec  $\Phi$  fonction de répartition de la loi normale centrée réduite.

Moments :  $m_k = e^{\frac{k\mu + \frac{k}{2}\sigma^2}{}}$  ainsi, en posant  $\omega = e^{\sigma^2}$   $m = E(C) = e^{\mu + \frac{\sigma^2}{2}} = e^\mu \omega^{1/2}$ ,  $V(C) = e^{2\mu} \omega(\omega - 1)$ ,  $\gamma_1 = (\omega - 1)^{1/2}(\omega + 2)$ ,  $\gamma_2 = \omega^4 + 2\omega^3 + 3\omega^2 - 6$

**Distribution Gamma  $\gamma(\alpha, \beta)$**

Elle a pour paramètres  $\alpha$  (la forme) et  $\beta$  (l'échelle) strictement positifs.

Densité et fonction de répartition :  $f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$  et  $F(x) = \frac{\int_0^x \beta^\alpha x^{\alpha-1} e^{-\beta x} dx}{\Gamma(\alpha)}$  avec la fonction

Gamma  $\Gamma$  définie par  $\Gamma(\nu) = \int_0^{+\infty} e^{-t} t^{\nu-1} dt$  pour  $\nu > 0$ .

Moments :  $m_k = \frac{\alpha(\alpha+1)\dots(\alpha+k-1)}{\beta^k}$ , ainsi  $m = E(C) = \frac{\alpha}{\beta}$ ,  $V(C) = \frac{\alpha}{\beta^2}$ ,  $\gamma_1 = \frac{2}{\sqrt{\alpha}}$ ,  $\gamma_2 = \frac{6}{\alpha}$

### Distribution Weibull $W(\tau, \alpha)$

Elle a pour paramètres  $\tau$  (la forme) et  $\alpha$  (l'échelle) strictement positifs.

Densité et fonction de répartition :  $f(x) = \frac{\tau}{\alpha^\tau} x^{\tau-1} e^{-\left(\frac{x}{\alpha}\right)^\tau}$  et  $F(x) = 1 - e^{-\left(\frac{x}{\alpha}\right)^\tau}$

Moments :  $m_k = \alpha^k \Gamma\left(\frac{k}{\tau} + 1\right)$ , ainsi  $m = E(C) = \alpha \Gamma\left(\frac{1}{\tau} + 1\right)$ ,  $V(C) = \alpha^2 \left[ \Gamma\left(\frac{2}{\tau} + 1\right) - \Gamma^2\left(\frac{1}{\tau} + 1\right) \right]$

### Distribution Pareto $P(a, \alpha)$

Elle a pour paramètres  $a$  (souvent donné, niveau d'une franchise par exemple) et  $\alpha$  strictement positifs.

Densité et fonction de répartition :  $f(x) = \frac{\alpha a^\alpha}{x^{\alpha+1}} \mathbf{1}(x)_{]a, +\infty[}$  et  $F(x) = \left[ 1 - \left(\frac{a}{x}\right)^\alpha \right] \mathbf{1}(x)_{]a, +\infty[}$

Moments :  $m_k = \frac{\alpha a^k}{\alpha - k}$  si  $\alpha > k$  ( $m_k = +\infty$  sinon)  $m = \frac{\alpha a}{\alpha - 1}$  si  $\alpha > 1$ ,  
 $V(C) = \frac{\alpha a^2}{(\alpha - 1)^2 (\alpha - 2)}$  si  $\alpha > 2$ ,  $\gamma_1 = \frac{2(\alpha + 1)}{\alpha - 3} \sqrt{\frac{\alpha - 2}{\alpha}}$  si  $\alpha > 3$

### Commentaires sur l'utilisation des différentes lois

La loi exponentielle  $\mathcal{E}(\theta)$  est simple d'utilisation mais présente une queue de distribution très fine souvent inadaptée à la modélisation des coûts en assurance non vie.

La distribution log-normale  $LN(\mu, \sigma^2)$  est la plus utilisée en assurance non vie et en réassurance car elle est faite pour des données positives acceptant une distribution à queue épaisse.

Une autre distribution classique en assurance non vie est la Gamma  $\gamma(\alpha, \beta)$  qui s'adapte efficacement à la modélisation des lois dont la queue de distribution n'est pas trop épaisse.

La loi Weibull  $W(\tau, \alpha)$  est en revanche moins utilisée en assurance non vie à cause des expressions des moments utilisant la fonction  $\Gamma$  et une estimation des paramètres qui s'avère plus difficile.

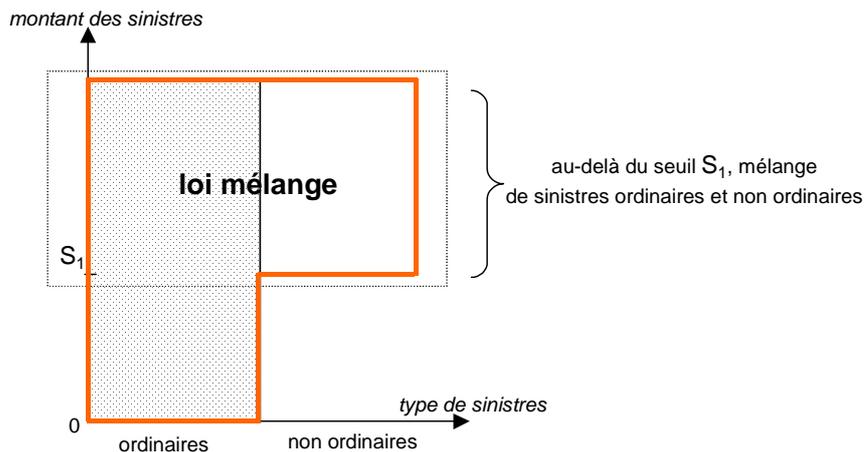
La distribution de Pareto  $P(a, \alpha)$  est la loi privilégiée en réassurance en raison de sa queue de distribution épaisse. Cependant l'existence des moments n'est pas certaine.

## V - B . 2. MODELISATION DE LA SEVERITE DES SINISTRES GRAVES

Pour affiner la modélisation, nous pouvons supposer que parmi les graves, on peut distinguer deux types de sinistres : ceux qui sont « ordinaires » et ceux qui sont « non ordinaires ».

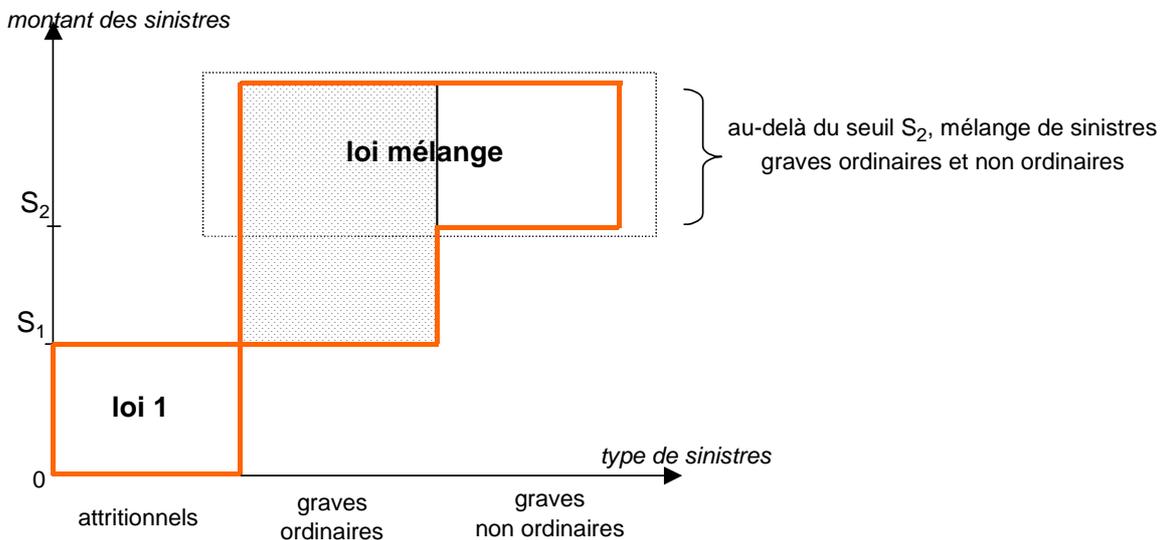
Nous modéliserons donc le coût de ces sinistres par une loi mélange de deux distributions dont la première correspond aux sinistres graves ordinaires et la deuxième aux sinistres graves non ordinaires.

Cette approche repose sur une méthode<sup>42</sup> utilisée pour modéliser l'ensemble des montants de sinistres d'un portefeuille scindé en grands sinistres (non ordinaires) et sinistres ordinaires. Le dessin ci-dessous l'illustre :



Objectif de Benabbou-Partrat : modéliser le mélange des 2 populations de sinistres (zone orange)

Nous appliquerons cette méthode à un portefeuille de sinistres graves segmentés en attritionnels et, pour les graves, en ordinaires et non ordinaires. Le dessin ci-dessous l'illustre.



Notre objectif : modéliser les 3 populations de sinistres (zone orange), les graves par une loi mélange d'après Benabbou-Partrat

<sup>42</sup> « Grands » sinistres et lois mélanges, Z. Benabbou et C. Partrat, 1994.

## Les 5 principales étapes de la méthode de Benabbou-Partrat

### 1 – La translation

Pour modéliser une variable aléatoire réelle  $C$ , montant d'un sinistre au-delà d'un certain seuil  $S_1$ , par une loi dont la densité est  $f_C$ , on modélise  $C-S_1$  (dont le support a pour origine 0) par la loi de densité  $f$  et nous avons  $f_C(x) = f(x - S_1) \cdot 1_{\{x > S_1\}}$

Nous appliquerons la translation :

- pour modéliser par une loi mélange les sinistres graves de montant supérieur à  $S_1$ ,
- au sein même de cette loi mélange, pour modéliser les montants des sinistres graves ordinaires et des graves non ordinaires.
- pour le calcul des statistiques des tests d'adéquation.

### 2 - Le modèle

La variable aléatoire réelle  $C$ , montant d'un sinistre, a pour fonction de répartition  $F(x, \lambda, \theta_1, \theta_2) = \lambda F_1(x, \theta_1) + (1 - \lambda) F_2(x, \theta_2)$  où  $\forall (q_1, q_2) \in \mathbb{N}^2$  :

- $F_1$  est la distribution des sinistres ordinaires dépendant d'un paramètre  $\theta_1 \in \Theta_1$  ouvert de  $\mathfrak{R}^{q_1}$ ,
- $F_2$  est la distribution des sinistres non ordinaires dépendant d'un paramètre  $\theta_2 \in \Theta_2$  ouvert de  $\mathfrak{R}^{q_2}$  vérifiant, pour tout  $\theta_2$ ,  $F_2(x, \theta_2) = 0$  pour  $x \leq x_0$  ( $x_0$  seuil donné),

On note  $\theta = (\theta_1, \theta_2) \in \Theta = \Theta_1 \times \Theta_2$  et  $\lambda \in ]0;1]$  qui est la probabilité qu'un sinistre soit ordinaire.

### 3 - Estimation des paramètres

Nous voulons déterminer  $(\hat{\lambda}_n, \hat{\theta}_{1n}, \hat{\theta}_{2n})$  l'estimateur du maximum de vraisemblance (emv) de  $(\lambda, \theta_1, \theta_2)$  basé sur la réalisation  $(x_i)_{i=1, \dots, n}$  d'un échantillon  $(C_i)_{i=1, \dots, n}$  de la variable aléatoire réelle  $C$  de densité  $f(x, \lambda, \theta) = \lambda f_1(x, \theta_1) + (1 - \lambda) f_2(x, \theta_2)$

Soit la log-vraisemblance associée  $\ln L = \sum_{i=1}^n \ln f(x_i, \lambda, \theta)$ , alors  $(\hat{\lambda}_n, \hat{\theta}_n)$ , sous réserve d'existence, est solution du système :

$$(S) \quad \begin{cases} \frac{\partial \ln L}{\partial \lambda} = \sum_{i=1}^n \frac{f_1(x_i, \theta_1) - f_2(x_i, \theta_2)}{f(x_i, \lambda, \theta)} = 0 \\ \frac{\partial \ln L}{\partial \theta_1^{(k)}} = \lambda \cdot \sum_{i=1}^n \frac{\partial f_1(x_i, \theta_1) / \partial \theta_1^{(k)}}{f(x_i, \lambda, \theta)} = 0 \quad k = 1, \dots, q_1 \\ \frac{\partial \ln L}{\partial \theta_2^{(l)}} = (1 - \lambda) \cdot \sum_{i=1}^n \frac{\partial f_2(x_i, \theta_2) / \partial \theta_2^{(l)}}{f(x_i, \lambda, \theta)} = 0 \quad l = 1, \dots, q_2 \end{cases}$$

L'algorithme E.M. (Expectation Maximisation) permet de déterminer la solution de ce système.

En supposant  $q_1 = q_2 = 1$  pour alléger les notations, on pose :

$$t_1(x) = \frac{\lambda \cdot f_1(x, \theta_1)}{f(x, \lambda, \theta)} \quad \text{et} \quad t_2(x) = \frac{(1 - \lambda) \cdot f_2(x, \theta_2)}{f(x, \lambda, \theta)} = 1 - t_1(x)$$

Le système (S) devenant alors :

$$(S') \begin{cases} \lambda = \frac{1}{n} \sum_{i=1}^n t_1(x_i) \\ \sum_{i=1}^n t_1(x_i) \frac{\partial \ln f_1(x_i, \theta_1)}{\partial \theta_1} = 0 \\ \sum_{i=1}^n t_2(x_i) \frac{\partial \ln f_2(x_i, \theta_2)}{\partial \theta_2} = 0 \end{cases}$$

Une fois la valeur initiale  $(\lambda^{(0)}, \theta^{(0)})$  déterminée, le processus est itératif.

Etape E (expectation) :

$$t_1^{(m)}(x) = \frac{\lambda^{(m)} f_1(x, \theta_1^{(m)})}{f(x, \lambda^{(m)}, \theta^{(m)})} \quad \text{et} \quad t_2^{(m)}(x) = \frac{(1 - \lambda^{(m)}) f_2(x, \theta_2^{(m)})}{f(x, \lambda^{(m)}, \theta^{(m)})}$$

Etape M (maximisation) :

$$\lambda^{(m+1)} = \frac{1}{n} \sum_{i=1}^n t_1^{(m)}(x_i)$$

$$\theta_1^{(m+1)} \text{ est solution de } \sum_{i=1}^n t_1^{(m)}(x_i) \frac{\partial \ln f_1(x_i, \theta_1)}{\partial \theta_1} = 0$$

$$\theta_2^{(m+1)} \text{ est solution de } \sum_{i=1}^n t_2^{(m)}(x_i) \frac{\partial \ln f_2(x_i, \theta_2)}{\partial \theta_2} = 0$$

#### 4 - Estimation de la valeur initiale $(\lambda^{(0)}, \theta^{(0)})$

Un bon choix de la valeur initiale s'avère nécessaire pour que la suite  $(\lambda^{(n)}, \theta^{(n)})$  converge vers  $(\hat{\lambda}_n, \hat{\theta}_n)$ . Sa détermination est basée sur l'identité  $F(x, \lambda, \theta) = \lambda F_1(x, \theta)$  pour  $x \leq x_0$ . Ainsi la loi  $F$  tronquée sur  $[0; x_0]$  coïncide avec la loi  $F_1$  tronquée sur  $[0; x_0]$  :

$$F_{[0; x_0]}(x, \lambda, \theta) = \frac{F(x, \lambda, \theta)}{F(x_0, \lambda, \theta)} = \frac{F_1(x, \theta_1)}{F_1(x_0, \theta_1)} = F_{1[0; x_0]}(x, \theta_1)$$

Dans la pratique, nous modéliserons les sinistres graves par un mélange de lois log-normales. Aussi nous détaillons ci-dessous les formules permettant de déterminer les paramètres correspondants. Elles sont exprimées ici comme dans l'article (sinistres ordinaires et non ordinaires) et seront adaptées aux graves lors de l'application pour prendre en compte les translations.

Les deux composantes de la loi mélange  $LN(\mu_1, \sigma_1^2)$  et  $LN(\mu_2, \sigma_2^2)$  ont pour densité :

$$f_1(x, \mu_1, \sigma_1) = \frac{1}{\sqrt{2\pi}\sigma_1 x} \exp\left(-\frac{1}{2\sigma_1^2} (\ln x - \mu_1)^2\right)$$

$$f_2(x, \mu_2, \sigma_2) = \frac{1}{\sqrt{2\pi}\sigma_2(x-x_0)} \exp\left(-\frac{1}{2\sigma_2^2} (\ln(x-x_0) - \mu_2)^2\right) \mathbf{1}_{]x_0, +\infty[}(x)$$

Une estimation  $\tilde{\theta}_{1n}$  de  $\theta_1$  est déterminée selon les procédures classiques d'estimation par le maximum de

vraisemblance de paramètres de lois tronquées à droite (Cohen, 1991<sup>43</sup>).

Si  $n_0$  est le nombre de sinistres dont chaque montant  $x_i \leq x_0$ , l'emv de  $(\mu_1, \sigma_1^2)$  basé sur la réalisation  $(x_i)_{i=1, \dots, n_0}$  de la loi  $LN(\mu_1, \sigma_1^2)$  tronquée à droite de  $x_0$  s'obtient ainsi : posons  $\bar{y} = \frac{1}{n_0} \sum_{i=1}^{n_0} \ln x_i$ ,

$$s^2 = \frac{1}{n_0} \sum_{i=1}^{n_0} (\ln x_i - \bar{y})^2 \text{ et } w = \frac{s^2}{(\ln x_0 - \bar{y})^2}.$$

Soient les fonctions  $W(x) = \frac{\varphi(x)}{\phi(x)}$ , où  $\varphi$  et  $\phi$  sont respectivement la densité et la fonction de répartition de

la loi normale centrée réduite et  $Q(w) = \frac{W(\hat{u})}{W(\hat{u}) + \hat{u}}$  où  $\hat{u}$  est l'unique solution de l'équation en  $u$  :

$$\frac{1 - W(u)[W(u) + u]}{[W(u) + u]^2} = w \quad \text{alors nous avons : } \tilde{\mu}_{1n} = \bar{y} + Q(w)(\ln x_0 - \bar{y}) \text{ et } \tilde{\sigma}_{1n}^2 = s^2 + Q(w)(\ln x_0 - \bar{y})^2$$

Pour  $w < 0,57081$ , nous avons l'approximation rationnelle suivante (Schneider, 1986<sup>44</sup>) :

$$Q(w) = \begin{cases} 0 & \text{si } w \leq 0,06246 \\ \frac{P_4(w)}{P_3(w)} & \text{si } 0,06246 < w < 0,57081 \end{cases}$$

avec :

$$P_4(w) = -0,00374615 + 0,17462558.w - 2,87168509.w^2 + 17,48932655.w^3 - 11,91716546.w^4$$

$$P_3(w) = 1 + 5,74050101.w - 13,53427037.w^2 + 6,88665552.w^3$$

Une fois déterminée  $(\tilde{\mu}_{1n}, \tilde{\sigma}_{1n}^2)$  l'estimation initiale de  $(\mu_1, \sigma_1^2)$ , nous en déduisons les estimations initiales  $\tilde{\lambda}_n$  et  $(\tilde{\mu}_{2n}, \tilde{\sigma}_{2n}^2)$  de  $\lambda$  et  $(\mu_2, \sigma_2^2)$ .

$\tilde{\lambda}_n = \frac{F_n^*(x_0)}{F_1(x_0, \tilde{\mu}_{1n}, \tilde{\sigma}_{1n}^2)}$  où  $F_n^*$  est la fonction de répartition empirique associée à l'échantillon  $(X_i)_{i=1, \dots, n}$  :

$$F_n^*(x) = \frac{1}{n} \sum_{i=1}^n 1_{[0; x]}(X_i)$$

$(\tilde{\mu}_{2n}, \tilde{\sigma}_{2n}^2)$  sont solutions de l'équation suivante en  $(\mu_2, \sigma_2^2)$ , où  $v$  est une valeur réelle choisie dans  $]x_0; +\infty[$  :  $F_2(v, \mu_2, \sigma_2^2) = \frac{F_n^*(v) - \tilde{\lambda}_n \cdot F_1(v, \tilde{\mu}_{1n}, \tilde{\sigma}_{1n}^2)}{1 - \tilde{\lambda}_n}$

### 5 - Estimation de la valeur $(\lambda^{(m+1)}, \theta^{(m+1)})$

A l'étape  $(m+1)$ , les paramètres des deux lois log-normales sont obtenus par les relations :

$$\mu_1^{(m+1)} = \frac{\sum_{i=1}^n t_1^{(m)}(x_i) \ln x_i}{\sum_{i=1}^n t_1^{(m)}(x_i)}, \quad \sigma_1^{(m+1)2} = \frac{\sum_{i=1}^n t_1^{(m)}(x_i) (\ln x_i - \mu_1^{(m)})^2}{\sum_{i=1}^n t_1^{(m)}(x_i)}$$

<sup>43</sup> Truncated and Censored Samples, A. Clifford Cohen, 1991.

<sup>44</sup> Truncated and Censored Samples from Normal Populations, H. Schneider, 1986.

$$\mu_2^{(m+1)} = \frac{\sum_{\substack{i=1 \\ x_i > x_0}}^n t_2^{(m)}(x_i) (\ln(x_i - x_0))}{\sum_{\substack{i=1 \\ x_i > x_0}}^n t_2^{(m)}(x_i)}, \quad \sigma_2^{(m+1)^2} = \frac{\sum_{\substack{i=1 \\ x_i > x_0}}^n t_2^{(m)}(x_i) (\ln(x_i - x_0) - \mu_2^{(m)})^2}{\sum_{\substack{i=1 \\ x_i > x_0}}^n t_2^{(m)}(x_i)}$$

## V - C LES TESTS D'ADEQUATION ET DE CONFORMITE

Nous présentons ici les tests usuels adaptés à nos travaux et qui seront mis en oeuvre.

### Les tests d'adéquation

Le modèle choisi, le(s) paramètre(s) estimé(s), pour évaluer la qualité de l'ajustement aux données, plusieurs tests d'adéquation peuvent être utilisés.

#### □ Test du $\chi^2$

Ce test s'applique généralement à une variable discrète et nous l'utiliserons pour évaluer l'adéquation des modèles de nombres de sinistres. Appliqué à une variable continue, ce test nécessite un regroupement des données en classes et donc une perte d'informations.

Dans le cas d'une variable discrète, la statistique  $\chi^2$  est définie selon :  $\chi^2(K-1) = \sum_{i=1}^K \frac{(N_i - E_i)^2}{E_i}$

avec  $K$  : le nombre de modalités,  $N_i$  : le nombre de valeurs observées de la  $i^{\text{ème}}$  modalité et  $E_i$  : le nombre de valeurs attendues de la  $i^{\text{ème}}$  modalité.

On détermine la valeur de la loi du  $\chi^2$  à  $K-1$  degrés de liberté ayant 5% de probabilité d'être dépassée. Dans l'hypothèse  $H_0$  où le phénomène étudié suit la loi théorique  $P$ , il y a 95% de probabilité cumulée que la variable  $\chi^2(K-1)$  prenne une valeur inférieure à celle donnée par la loi du  $\chi^2$ . Si la valeur de la loi du  $\chi^2$  obtenu à partir de l'échantillon prélevé est inférieure à celle correspondant aux 95% de probabilité cumulée, nous accepterons l'hypothèse selon laquelle le phénomène suit la loi  $P$ .

#### □ Test de Kolmogorov Smirnov

Pour une v.a.r. continue telle que le coût des sinistres, ce test est plus approprié que le test du  $\chi^2$  ; il fait partie de la famille des tests EDF (empirical distribution function) issus de l'idée que la fonction de répartition empirique est une bonne approximation de la fonction de répartition théorique.

Soit  $(C_1, \dots, C_n)$  un  $n$ -échantillon de la v.a.r. coût  $C$  de fonction de répartition (f.de r.)  $F$  et  $F_n$  sa fonction de répartition empirique. Par définition de  $F_n$ , on a :

$$F_n(x) = \frac{i}{n} \text{ pour } C_{(i)} < x \leq C_{(i+1)} \text{ avec les } (C_{(i)})_{i=1, \dots, n} \text{ ordonnés.}$$

Nous souhaitons tester l'adéquation de la f.de r.  $F$  de la v.a.r.  $C$  à une f.de r.  $G$  continue donnée. Comme  $F$  est inconnue, nous lui substituons la f.de r. empirique  $F_n$ .

La distance de Kolmogorov-Smirnov est définie en  $G$  par :

$$D_n(G) = \sup_x |F_n(x) - G(x)| = \max[D_n^+(G), D_n^-(G)] \text{ où } D_n^+(G) = \sup_x [F_n(x) - G(x)] \text{ et } D_n^-(G) = \sup_x [G(x) - F_n(x)]$$

En pratique nous utilisons la proposition suivante :

$$D_n^+(G) = \max_{1 \leq i \leq n} \left[ \frac{i}{n} - G(C_{(i)}) \right] \text{ et } D_n^-(G) = \max_{1 \leq i \leq n} \left[ G(C_{(i)}) - \frac{i-1}{n} \right]$$

L'avantage de ce test sur celui du  $\chi^2$  est qu'il ne perd pas d'informations (pas de regroupement en classes) et qu'il est précis au niveau de l'écart maximal entre deux distributions. Son inconvénient est de ne pas bien évaluer les éventuelles divergences en queue de distribution.

□ Test d'Anderson Darling

C'est une déclinaison du test précédent qui donne plus de poids à la queue de distribution ; le choix de cette distance est pertinent dans les cas où la queue de distribution doit être bien prise en compte ; ce test fait aussi partie de la famille des tests EDF. Nous l'utiliserons pour mesurer l'ajustement des distributions des sinistres graves.

L'expression opérationnelle de  $A_n^2$  est : 
$$A_n^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) \{ \ln F(C_{(i)}) + \ln [1 - F(C_{(n-i+1)})] \}$$

où  $C_{(i)}$  désigne la  $i^{\text{ème}}$  plus petite réalisation de la v.a.r. coût de sinistre  $C$  dans l'échantillon.

□ Tests d'adéquation EDF et estimation de paramètre

On sait grâce à Stephens<sup>45</sup> que les statistiques EDF habituelles ne sont pas adaptées au test d'adéquation entre les observations d'un échantillon et une famille de lois dont le paramètre est lui-même estimé à partir de cet échantillon.

Pour certaines familles de lois (loi exponentielle, loi normale), des tables de quantiles existent.

Dans les autres cas, nous utiliserons la méthode du demi-échantillon :

- à partir de l'échantillon principal noté EP, tirage aléatoire d'un demi-échantillon noté DE,
- estimation du paramètre  $a$  de la loi sur DE notée  $\hat{a}$ ,
- calcul de la statistique EDF sur EP avec la loi et  $\hat{a}$ . Cette statistique admet pour loi limite celle tabulée pour le cas où le paramètre est défini sans estimation.

□ Tests d'adéquation EDF et loi tronquée

Ce cas est rencontré lors de la modélisation des sinistres attritionnels tronqués au seuil  $S_1$ .

En lien avec les travaux de Hogg et Klugman<sup>46</sup>, on aménage la fonction de répartition empirique pour tenir compte de cette situation.

Soit  $X$  une v.a.r. de fonction de répartition  $F_X$  dont on cherche la loi de distribution. Soit  $Y$  une v.a.r. de fonction de répartition  $F_Y$  qui suit la loi de  $X$  tronquée à gauche en une valeur connue  $p$  et à droite en une valeur connue  $g$ . Nous disposons d'un échantillon  $Y_1, \dots, Y_n$  et des valeurs  $p$  et  $g$ . Par définition de la variable  $Y$ , nous pouvons écrire :

$$F_Y(x) = P(X \leq x / p < X < g) = \frac{P(p < X \leq x)}{P(p < X < g)} \text{ et } F_Y(x) = \begin{cases} 1 & \text{si } x > g \\ \frac{F_X(x) - F_X(p)}{F_X(g) - F_X(p)} & \text{si } p < x \leq g \\ 0 & \text{si } x \leq p \end{cases}$$

<sup>45</sup> EDF statistics for goodness-of-fit and some comparisons, Journal of the American Statistical Association, M.A. Stephens, 1974.

<sup>46</sup> Loss distribution, R. Hogg et S. Klugman, 1984.

## Le test de conformité

### Position du problème

Dans la population P, un caractère peut prendre k valeurs (ou être réparti en k classes), notées  $C_1, C_2, \dots, C_k$ . Exemple de caractère : le montant d'un sinistre, l'âge d'une victime,...

On dispose des échantillons  $E_1, E_2, \dots, E_l$  susceptibles de provenir d'une population P. Quel que soit i élément de  $[1, k]$  et quel que soit j élément de  $[1, l]$ , on note  $O_{ij}$  l'effectif observé de la valeur  $C_i$  dans

l'échantillon  $E_j$ . L'effectif total est  $N = \sum_{j=1}^l \sum_{i=1}^k O_{ij}$ .

### Hypothèse à tester

Nous choisissons pour hypothèse de travail notée ( $H_0$ ) que les différences observées entre les différents échantillons sont dues aux fluctuations d'échantillonnage, c'est-à-dire que nous supposons ces différents échantillons extraits de la même population.

### Loi théorique

Nous ignorons la répartition du caractère étudié dans la population P. Nous ne pouvons que l'estimer à partir des échantillons que nous réunissons en un seul de taille N.

Alors, pour tout i de  $[1, k]$ , la probabilité de  $C_i$  est  $p_i = \frac{1}{N} \left( \sum_{j=1}^l O_{ij} \right)$  et l'effectif du caractère  $C_i$  pour

l'échantillon  $E_j$  est  $T_{ij} = p_i \left( \sum_{i=1}^k O_{ij} \right)$ . En général, on exige que  $T_{ij} \geq 5$  sinon on effectue des regroupements

de classes.

Sous l'hypothèse ( $H_0$ ), la variable aléatoire S prenant pour chaque échantillon de taille N la valeur

$S = \sum_{i=1}^k \sum_{j=1}^l \frac{(O_{ij} - T_{ij})^2}{T_{ij}}$  suit la loi du  $\chi^2$  à  $\nu = (k-1) \cdot (l-1)$  degrés de liberté.

## V - D APPLICATIONS

Dans ce chapitre, nous présentons le modèle paramétrique obtenu par la méthode probabiliste ainsi que les simulations de charges brutes et nettes qu'il permet.

La base de données sélectionnée est celle « as if » utilisée au chapitre IV.

Les seuils retenus sont ceux définis au chapitre IV :  $S_1 = 200$  k€ pour séparer les attritionnels des graves et  $S_2 = 1\,250$  k€ pour séparer les graves ordinaires des graves non ordinaires.

Ces sinistres sont considérés comme homogènes (i.i.d.) et assez nombreux pour se situer dans le cadre de l'application de la loi des grands nombres : convergence de l'espérance de la charge sinistres vers sa moyenne.

### V - D . 1. RESULTATS DE LA MODELISATION DES NOMBRES

Nous définissons une unique loi (Poisson) pour tous les sinistres ; nous utiliserons ensuite des lois binomiales pour répartir le nombre de sinistres en différentes catégories (IRCA, attritionnels, graves).

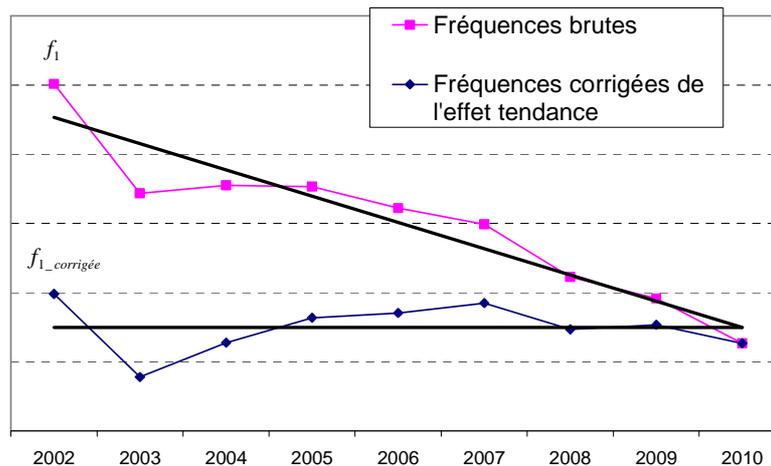
Pour cela, nous constituons le triangle des nombres de sinistres sur 11 ans, calculons les cadences via Chain-Ladder et obtenons des nombres à l'ultime.

Nous retraitions de l'effet portefeuille par passage des nombres aux fréquences. Toutefois, le parc de

véhicules du portefeuille n'est connu que sur une période courte : conséquence des convergences et de la perte d'historique inhérente.

Pour augmenter le périmètre historique des données et avoir un modèle de nombres plus robuste, nous appliquons l'approximation suivante : sur les quelques années où le parc est connu, nous calculons les fréquences ; nous étudions alors leur corrélation avec celles de l'auto. Les liens étant forts ( $R^2 > 91\%$ ), nous extrapolons les fréquences flottes en s'appuyant sur la tendance des fréquences auto.

Nous retraits ensuite de l'effet tendance en déterminant la droite de régression des fréquences et en calculant les écarts des observations par rapport à cette droite. On supprime alors la tendance (pente de la droite = 0 et  $y = \text{fréquence } 2010$ ) et calculons des nouvelles ordonnées des points « fréquence » par ajout des écarts avec la droite.

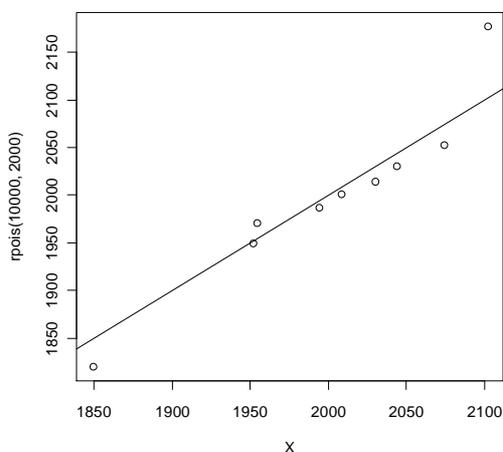


Nous arrivons ainsi à reconstituer une série d'observations homogènes nous permettant de calculer le paramètre de la loi de Poisson ( $\lambda = \text{moyenne}(f_{i\_corrigée} \times \text{parc}_{2010}) = 2001$ ) que nous privilégions.

Nous vérifions la bonne adéquation au sens du test du  $\chi^2$  du modèle trouvé en zoomant sur la répartition 2010 du nombre de sinistres par véhicule.

Test du  $\chi^2$  - Poisson

Distance observée	6,06
Distance tabulée	7,81
Décision	Non rejet



Le QQ-plot de cette loi de Poisson avec les données est correct sauf aux extrémités et confirme l'ajustement.

En plus d'être un modèle satisfaisant au sens du test du  $\chi^2$ , le choix de la loi de Poisson nous satisfait pour sa simplicité d'utilisation, son paramètre explicite, son application éprouvée à un portefeuille homogène et sa queue fine.

## V - D . 2. RESULTATS DE LA MODELISATION DES MONTANTS

Afin d'obtenir une meilleure modélisation, la base est retraitée des individus à forte fréquence.

Nous décelons 3 natures de sinistres sur-représentés :

- les forfaits conventionnels IRCA (1490€) et leurs principaux multiples (50% et 200%),
- les forfaits d'ouverture (8k€),
- des montants moins explicites. Après investigation, ils sont liés aux pratiques de gestion : ouverture d'une garantie corporelle pour régler l'ensemble du sinistre, corporel et matériel. Nous identifions les montants des forfaits IRSA à 100% (1204€) et à 50 % (602€) pouvant être mélangés à du corporel.
  - Exemple : le montant « 1347 € » est retrouvé 1588 fois et provient du calcul suivant : (forfait IRSA de 1204 € + forfait IRCA de 1490 €) \*taux de responsabilité de 50% ; dans ce cas le montant retraité devient 745€.

Pour cette modélisation, nous choisissons d'écarter les sinistres sans suite et ceux de montant égal au forfait IRCA à 100% ; les autres forfaits sont moins nombreux et totalisent des montants négligeables.

Pour les montants, nous avons cantonné l'étude à une période de 3 ans (2007-2009) pour bénéficier de pratiques homogènes en gestion.

### Contexte de modélisation de la sévérité

On gère 2 seuils  $S_1$  et  $S_2$  avec  $S_1 < S_2$ .

Sur  $[0, S_1]$ , la distribution est tronquée à droite et nécessite donc une normalisation.

Sur  $]S_1, +\infty[$  et  $[S_2, +\infty[$ , la distribution mélange n'est pas tronquée mais est la combinaison linéaire de 2 distributions translatées.

Hormis le cas des sinistres IRCA qui sera traité au moment du calcul de la charge, on veut trouver une loi à support dans  $[0, +\infty[$  admettant  $F$  comme fonction de répartition et telle que :

$$F(x) = \lambda_{attri} \cdot F_1^{tronquée}(x) \cdot 1_{\{0 \leq x \leq S_1\}} + (1 - \lambda_{attri}) \cdot F^{mélange}(x) \cdot 1_{\{S_1 < x\}}$$

où  $F^{mélange}(x) = \lambda \cdot F_2(x - S_1) \cdot 1_{\{S_1 < x\}} + (1 - \lambda) \cdot F_3(x - S_2) \cdot 1_{\{x > S_2\}}$

et avec :

- $\lambda_{attri}$  : la proportion de sinistres attritionnels,
- $\lambda$  : la proportion de sinistres graves ordinaires, supérieurs à  $S_1$ ,
- $F_1^{tronquée}$  : la f. de r. normalisée telle que  $F_1^{tronquée} = \frac{F_1(x) - F(0)}{F_1(S_1) - F(0)}$ ,  $F(0) = 0$  et  $F_1$  la f. de r. d'une

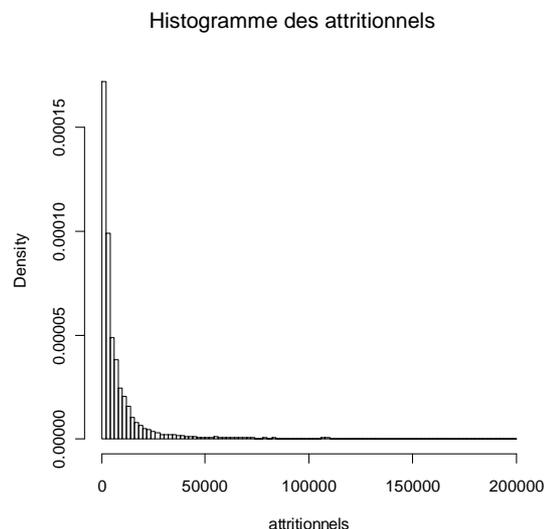
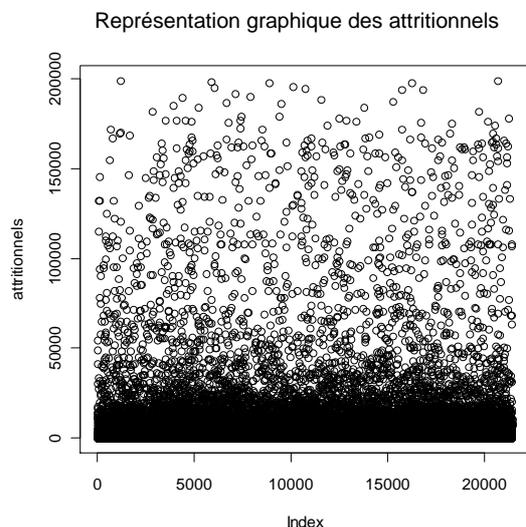
loi log-normale sur  $[0, S_1]$ ,

- $F_2$  : la f. de r. d'une loi log-normale sur  $[S_1, +\infty[$ ,
- $F_3$  : la f. de r. d'une loi log-normale sur  $[S_2, +\infty[$ .

## Modélisation des montants attritionnels

### Statistiques descriptives de cette population :

Moyenne	Ecart-type	Skewness	Kurtosis	1 <sup>er</sup> quartile	Médiane	3 <sup>ème</sup> quartile	4 <sup>ème</sup> quartile
10 650	22 666	4,6	26,9	1 395	3 400	9 363	199 256



### Analyse :

La moyenne est proche du triple de la médiane ce qui indique la présence de valeurs importantes.  
 L'écart-type est proche du double de la moyenne : il y a de forts écarts dans les valeurs de l'échantillon.  
 Le kurtosis<sup>47</sup> étant significativement différent de 3, la distribution est leptokurtique ou pointue.  
 Le skewness<sup>50</sup> étant significativement différent de 0, la distribution est asymétrique à droite.

L'utilisation d'un modèle log-normal apparaît cohérente.

L'estimation des paramètres de loi est faite par la méthode du maximum de vraisemblance.

Les paramètres estimés de  $F_1$  sont  $\mu = 8,14$  et  $\sigma = 1,6$ .

### Calcul des statistiques des tests EDF :

On calcule les estimateurs du maximum de vraisemblance des paramètres de ce modèle sur un demi-échantillon aléatoire ; on trouve  $\mu = 8,15$  et  $\sigma = 1,61$ .

Puis on calcule les statistiques des tests basées sur l'échantillon entier (limité à 500 sinistres aléatoirement<sup>48</sup>), pour la loi définie par les paramètres estimés sur le demi-échantillon et en incluant la normalisation liée à la troncature à droite :

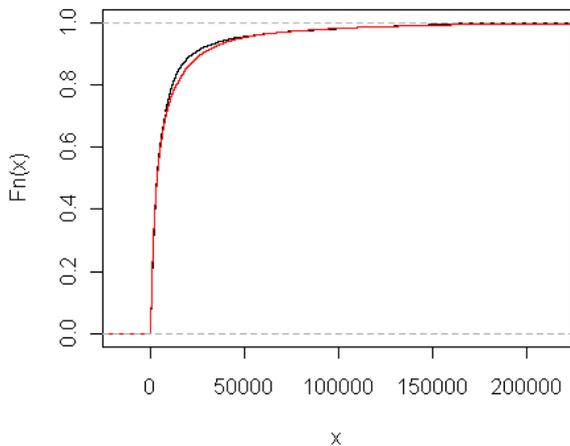
Risque 5%	K-S	A-D
Stat. calculée	1,174	2,126
Stat. tabulée	1,358	2,492
Décision	Non rejet	Non rejet

Les valeurs tabulées de ces 2 statistiques EDF, au risque d'erreur  $\alpha=5\%$ , étant supérieures aux valeurs calculées, **nous ne rejetons pas l'hypothèse nulle qui suppose que les 2 distributions sont identiques.**

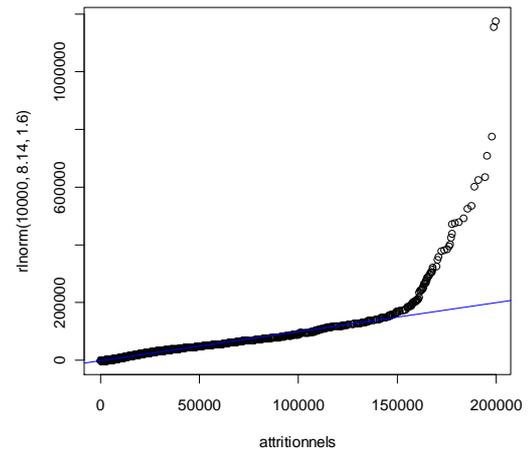
<sup>47</sup> Voir calcul en annexe VIII-A.

<sup>48</sup> Assurance non-vie, modélisation, simulation, C. Partrat & J-L. Besson, 2005.

Courbes des fonctions de répartition empirique (noir) et de la loi log-normale (8,14 ; 1,6) (rouge).



QQ-plot des attritionnels avec la log-normale (8,14 ; 1,6)



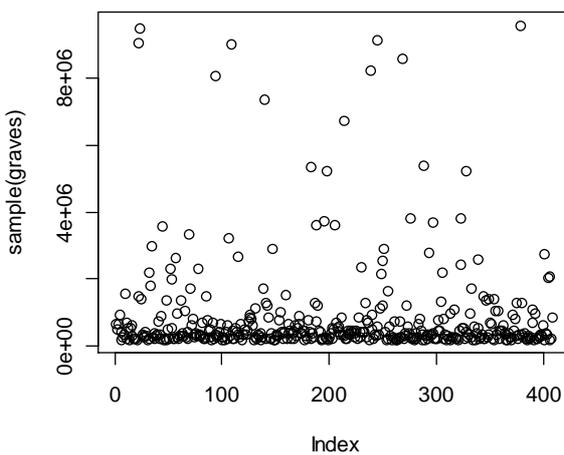
Le QQ-plot confirme que la loi log-normale est bien ajustée à la distribution empirique pour plus de 97% des sinistres. En effet, l'alignement des sinistres inférieurs à 150 k€ sur la bissectrice bleue est bien visible. Les sinistres supérieurs à 150 k€ semblent régis par une autre loi. C'est cohérent avec nos travaux précédents et la présence d'un seuil dans cette zone.

**Modélisation des montants graves**

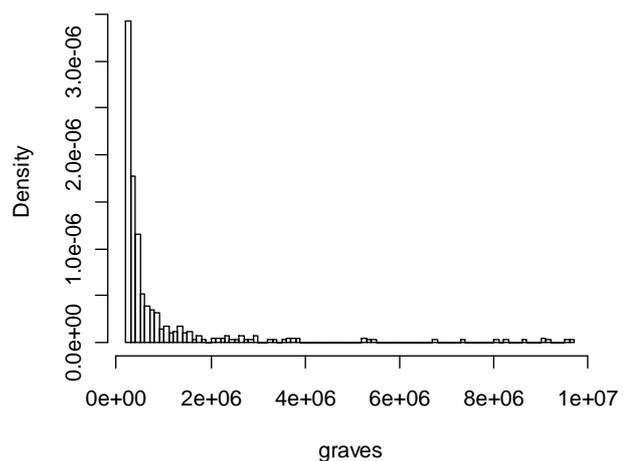
**⊕ Statistiques descriptives des sinistres graves au-delà du seuil  $S_1$  :**

Moyenne	Ecart-type	Skewness	Kurtosis	1 <sup>er</sup> quartile	Médiane	3 <sup>ème</sup> quartile	4 <sup>ème</sup> quartile
896 209	1 481 144	4	20	264 375	388 595	754 571	9 604 969

Représentation graphique des graves ordinaires



Histogramme des graves ordinaires



**Analyse :**

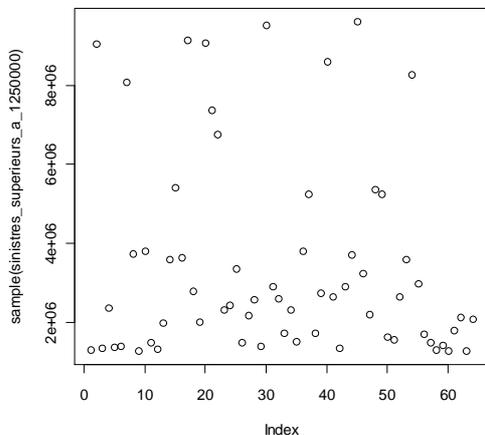
La moyenne est supérieure au double de la médiane ce qui traduit une forte dispersion des valeurs. C'est confirmé par la valeur de l'écart-type, proche de 2 fois la moyenne et par les quartiles. Le kurtosis étant significativement différent de 3, la distribution est leptokurtique ou pointue. Le skewness étant significativement différent de 0, la distribution est asymétrique à droite.

Le choix d'une loi log-normale semble donc pertinent sur  $]S_1, +\infty[$ .

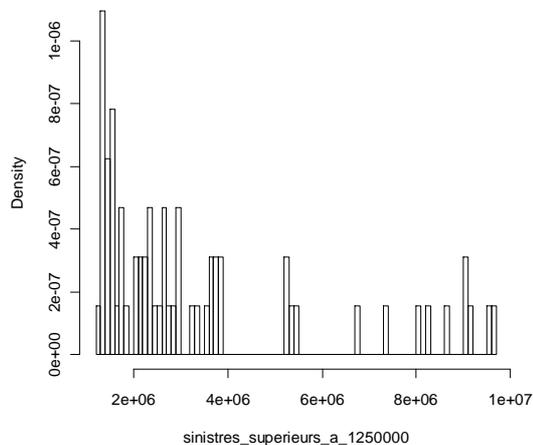
⊕ **Statistiques descriptives des sinistres graves au-delà du seuil  $S_2$  :**

Moyenne	Ecart-type	Skewness	Kurtosis	1 <sup>er</sup> quartile	Médiane	3 <sup>ème</sup> quartile	4 <sup>ème</sup> quartile
3 426 474	2 480 641	1,4	3,6	1 571 890	2 519 046	3 768 348	9 604 969

Représentation graphique des graves non ordinaires



Histogramme des graves non ordinaires



**Analyse :** on constate que la distribution est pointue. Le skewness étant supérieur à 0, la distribution est asymétrique à droite.

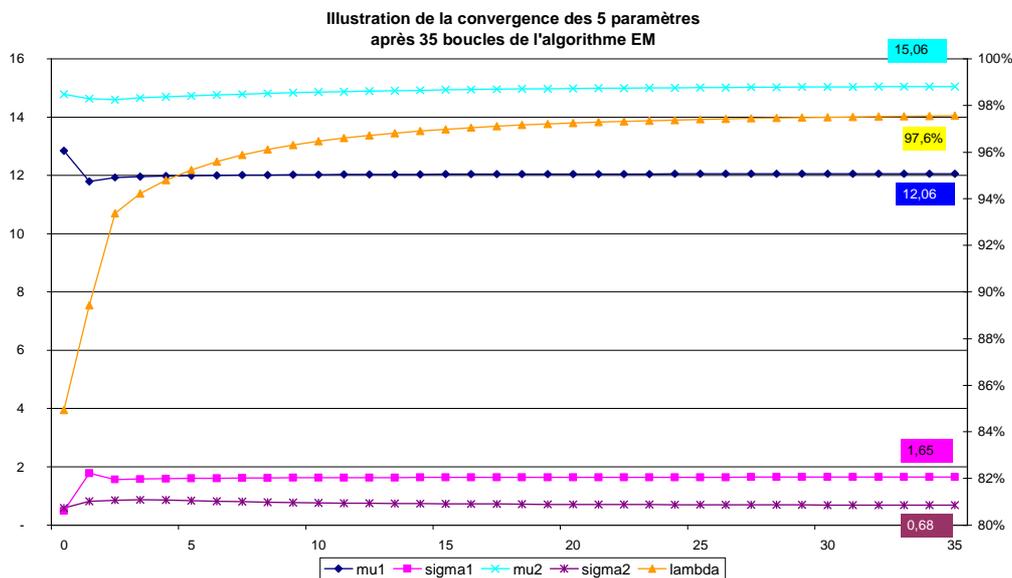
Le choix d'une loi log-normale semble possible sur  $[S_2, +\infty[$ .

⊕ **Conclusion :**

Tant pour les graves ordinaires ( $>S_1$ ) que non ordinaires ( $>S_2$ ), on a vérifié qu'on se situait bien dans un **contexte log-normalien** ; on peut alors estimer les paramètres de la loi mélange par la méthode de Partrat-Benabbou appliquée à 2 log-normales translatées, cette loi étant notée « mélange P-B ».

Nous obtenons les paramètres suivants de la loi mélange P-B :  
 $\lambda = 97,6\%$      $\mu_1 = 12,06$      $\sigma_1 = 1,65$      $\mu_2 = 15,06$      $\sigma_2 = 0,68$

Dans Excel, nous créons un modèle déclinant la méthode et qui permet de définir les 5 paramètres de la loi mélange P-B. Nous vérifions son bon fonctionnement, notamment en ce qui concerne la convergence des estimations des paramètres (voir figure ci-après).



⊕ **Calcul des statistiques des tests EDF :**

On applique la méthode du demi-échantillon à la loi mélange P-B. Les nouveaux paramètres obtenus sont :

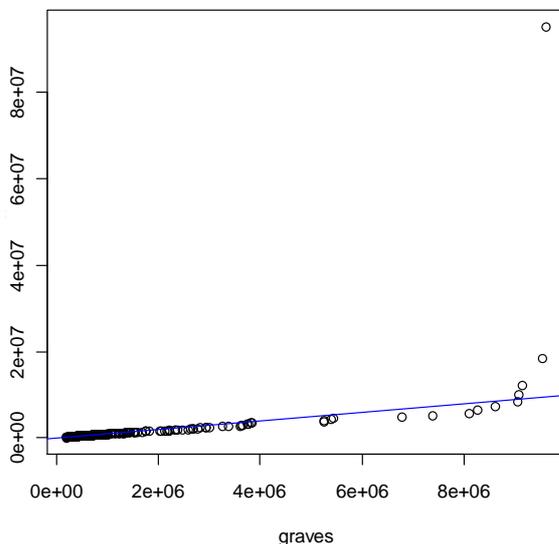
$$\lambda = 94,9\% \quad \mu_1 = 11,93 \quad \sigma_1 = 1,54 \quad \mu_2 = 15,06 \quad \sigma_2 = 0,63$$

Sur l'échantillon complet et avec cette nouvelle loi mélange, on obtient les statistiques des tests EDF suivantes qui incluent les corrections liées aux 2 translations :

Risque 5%	K-S	A-D
Stat. calculée	0,649	0,527
Stat. tabulée	1,358	2,492
Décision	Non rejet	Non rejet

Les valeurs tabulées des statistiques EDF, au risque d'erreur  $\alpha=5\%$ , étant supérieures aux valeurs calculées, **nous ne rejetons pas l'hypothèse nulle qui suppose que les 2 distributions sont identiques.**

QQ-plot des graves avec la mélange P-B



Le QQ-plot confirme ce choix jusqu'à 4 000 k€, soit pour plus de 96% des sinistres.

Seuls quelques sinistres les plus élevés s'écartent de la droite bissectrice  $y=x$  (en bleu).

⊕ **Contrôle :**

On veut vérifier l'existence d'un gain lié à l'utilisation de la loi mélange P-B comparée à une approche par une seule loi. Pour cela, on va tester 2 scenarii utilisant une seule loi sur  $[S_1, +\infty[$ . On retient les modèles log-normale et Pareto pour leur aptitude à reproduire les queues épaisses des distributions. Leurs paramètres e.m.v. sont :

Log-normale		Pareto	
$\mu$	$\sigma$	a	$\alpha$
12,14	1,71	$S_1$	1,0611

S'agissant de lois dont les paramètres sont estimés, on utilise la méthode du demi-échantillon pour obtenir les 2 statistiques EDF, en appliquant une translation pour la distribution log-normale. On présente ci-dessous les paramètres des lois ainsi estimés :

Log-normale		Pareto	
$\mu$	$\sigma$	a	$\alpha$
12,15	1,77	$S_1$	1,0348

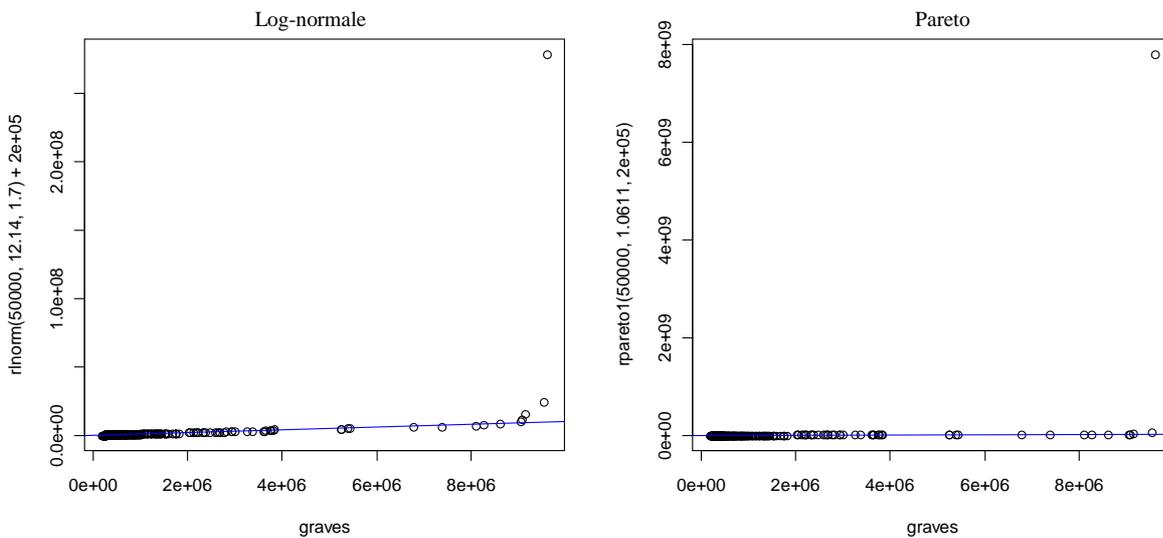
Ainsi que les résultats des calculs des statistiques de tests EDF :

Test EDF	Kolmogorov-Smirnov			Anderson-Darling		
	Risque 5%	Stat. calculée	Stat. tabulée	Décision	Stat. calculée	Stat. tabulée
Log-normal	0,874	1,358	Non rejet	0,605	2,492	Non rejet
Pareto	0,798			0,622		

1<sup>er</sup> constat :

les valeurs tabulées de ces statistiques EDF au risque d'erreur  $\alpha=5\%$  , étant toutes supérieures aux valeurs calculées, **nous ne rejetons pas les hypothèses nulles qui supposent que les 2 distributions testées sont identiques à l'empirique**. Ces 2 scénarii constituent donc, eux aussi, des modèles très corrects.

Les QQ-plots des 2 modèles étudiés et présentés ci-dessous confirment leur qualité.



2<sup>ème</sup> constat :

Stat. EDF	K-S	A-D
Mélange P-B	0,649	0,527
Log-normale	0,874	0,605
Pareto	0,798	0,622

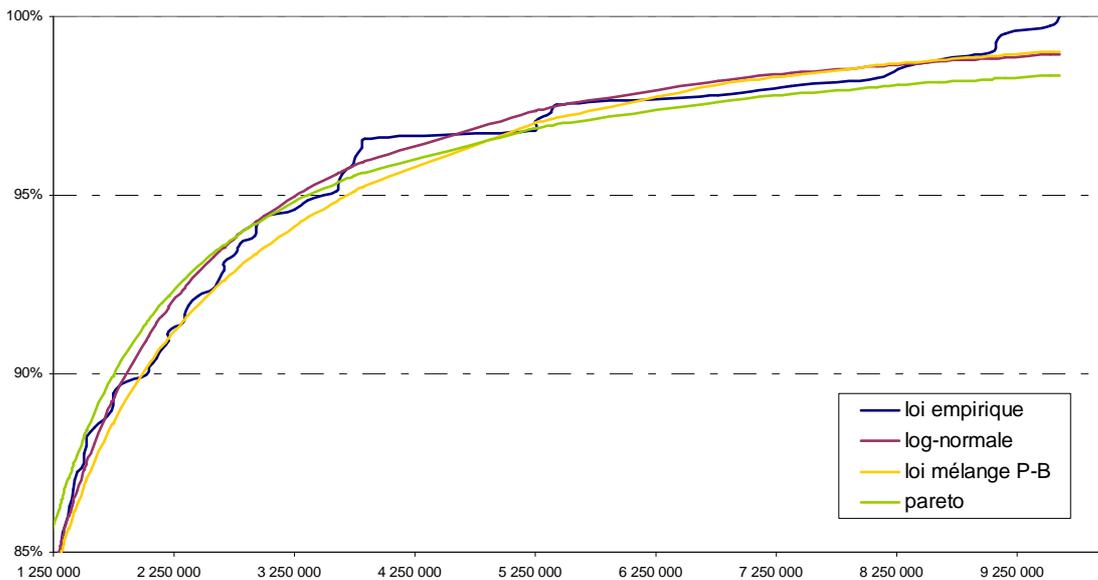
Au vu du tableau de synthèse ci-dessus, les 2 statistiques des tests EDF de la distribution mélange P-B sont inférieures à celles des lois log-normale et Pareto. Le modèle mélange P-B semble donc plus performant au sens des distances de ces tests.

3<sup>ème</sup> constat :

Le graphe ci-après confirme que le modèle Pareto transcrit la queue de distribution la plus épaisse pour les sinistres supérieurs à 5 000 k€.

Dans cette zone, la mélange P-B atténue favorablement cet effet et, à partir de 8 000 k€, elle offre une queue plus proche de l'empirique que la loi log-normale.

**Zoom sur l'ajustement de la distribution empirique sur [1.250;10.000] k€**



**Conclusion**

Les tables des quantiles EDF à notre disposition s’arrêtant à 25%, elles ne nous permettent pas de quantifier par les p-valeurs le gain apporté par la mélange P-B sur les autres lois. Les statistiques des 2 tests EDF et surtout celles d’Anderson-Darling, montrent que **cette loi mélange P-B offre la meilleure adéquation à la distribution empirique.**

Par ailleurs, on remarque le bon ajustement de la distribution empirique par une loi de Pareto au seuil  $S_I$ , ce qui est cohérent avec nos travaux précédents sur la recherche de ce seuil. Il serait intéressant de mesurer l’apport d’une loi mélange entre une loi log-normale (entre  $S_1$  et  $S_2$ ) et une loi de Pareto (au-delà de  $S_2$ ).

Nous avons identifié une amélioration possible de la méthode du demi-échantillon. En effet, elle peut donner des résultats assez volatils quand les données sont peu nombreuses. Une méthode type bootstrap consistant à tirer plusieurs demi-échantillons et à mesurer la moyenne des statistiques EDF obtenues pourrait apporter davantage de stabilité.

**V - D . 3.      RESULTATS DES SIMULATIONS DE LA CHARGE**

Après avoir modélisé la loi des nombres et des montants, comment obtenir la charge ultime ? Nous allons travailler par simulation sur le principe d’une méthode Monte-Carlo.

**Sinistres attritionnels**

On présente les 5 étapes permettant de simuler **la charge brute des attritionnels.**

Etape 1 : on fait  $n$  tirages aléatoires de nombres de sinistres selon la loi de Poisson retenue. On note  $n_1, n_2, \dots, n_n$  les  $n$  nombres tirés. Soit  $\lambda_{IRCA}$  la proportion de sinistres dont le montant est égal au forfait IRCA à 100%, soit  $\lambda_{attri}$  la proportion de sinistres attritionnels et soit  $\lambda_{graves}$  la proportion de sinistres graves. On a :  $\lambda_{IRCA} + \lambda_{attri} + \lambda_{graves} = 1$ .

Etape 2 :  $\forall i \in [1, n]$ , soit  $n_i^{attri}$  le nombre de sinistres attritionnels tiré dans la loi binomiale  $B(n_i, \lambda_{attri})$ .

**Etape 3 :**  $\forall i \in [1, n]$ , pour tout nombre  $n_i^{attri}$  obtenu, on effectue  $n_i^{attri}$  tirages aléatoires de montants de sinistres dans la loi log-normale LN (8,14 ; 1,6) que l'on notera  $X_1^{n_i^{attri}}, X_2^{n_i^{attri}}, \dots, X_{n_i^{attri}}^{n_i^{attri}}$

**Etape 4 :**  $\forall i \in [1, n]$ , on calcule la charge totale des sinistres attritionnels de la  $i^{ème}$  simulation notée :

$$C_i = \sum_{k=1}^{n_i^{attri}} X_k^{n_i^{attri}} + n_i^{IRCA} * IRCA_{100\%}$$

où  $n_i^{IRCA}$  désigne le nombre de sinistres IRCA tiré dans la loi binomiale  $B(n_i - n_i^{attri}; \frac{\lambda_{IRCA}}{\lambda_{IRCA} + \lambda_{graves}})$  et  $IRCA_{100\%}$  désigne le montant du forfait IRCA à 100%, soit 1 490 € en 2010.

**Etape 5 :** on calcule alors l'espérance, un estimateur de la variance et les principaux quantiles de ces n charges.

Si C désigne la charge des sinistres attritionnels, il semble raisonnable de retenir comme estimateur de

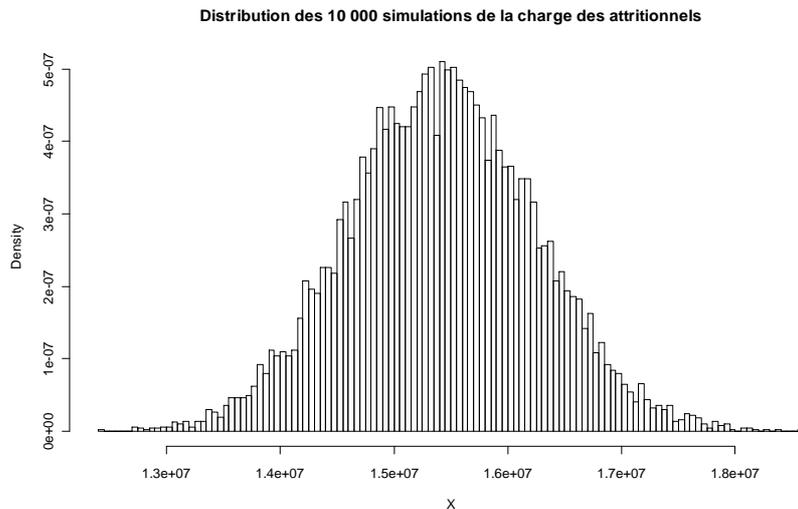
l'espérance :  $\hat{m} = \frac{1}{n} \sum_{i=1}^n C_i$ . De même, pour la variance empirique, on utilise l'estimateur suivant :

$$\hat{s}^2 = \frac{1}{n-1} * \sum_{i=1}^n (C_i - \frac{1}{n} \sum_{i=1}^n C_i)^2$$

A noter que cette charge est aussi nette de réassurance car le seuil  $S_I$  majorant les attritionnels est inférieur à la priorité du traité 1. Le logiciel utilisé est Igloo.

Pour  $n = 10\ 000$  et  $\lambda_{attri} = 71,31\%$ , on obtient **la charge brute des attritionnels** :

Attritionnels <200k€	
Espérance	15 432 656 €
Ecart-type	848 469 €
Q25,0%	14 853 412 €
Q50,0%	15 433 665 €
Q75,0%	15 995 962 €
Q95,0%	16 834 574 €
Q97,5%	17 136 413 €
Q99,0%	17 452 058 €
Q99,5%	17 645 609 €



On note un écart-type faible donc une bonne prévisibilité de cette charge.

On peut lister les éléments exogènes qui seraient de nature à perturber cette prévision : ceux liés à la sinistralité (évolution des fréquences, répression routière renforcée, météorologie, prix du carburant, contexte économique...) et ceux liés à la gestion (changement des forfaits d'ouverture, évolution des cadences de gestion,...).

Avant de présenter la simulation des sinistres graves en brut et en net de réassurance, on rappelle les principaux éléments du programme de réassurance que l'on va intégrer dans la modélisation.

### Présentation de la structure de réassurance que l'on va modéliser

Il s'agit de 2 traités en excédent de sinistres (XS). En effet, il est utile de rappeler que la garantie RC corporelle étant illimitée, il n'est pas envisagé de recourir ni à un traité proportionnel ni à un traité en excédent de perte qui comporte toujours une limite de garantie.

#### **1<sup>er</sup> traité :**

La compagnie réassurée, appelée *cédante*, transfère au réassureur la charge du sinistre au-delà d'un seuil, la *priorité 1*, et à concurrence d'un montant de garantie, la *portée*.

En-deçà de cette priorité, la charge du sinistre est appelée *réretention* et est conservée par la cédante. Sur l'ensemble de la charge cédée, le réassureur n'assure qu'un montant maximal défini appelé *limite de garantie*.

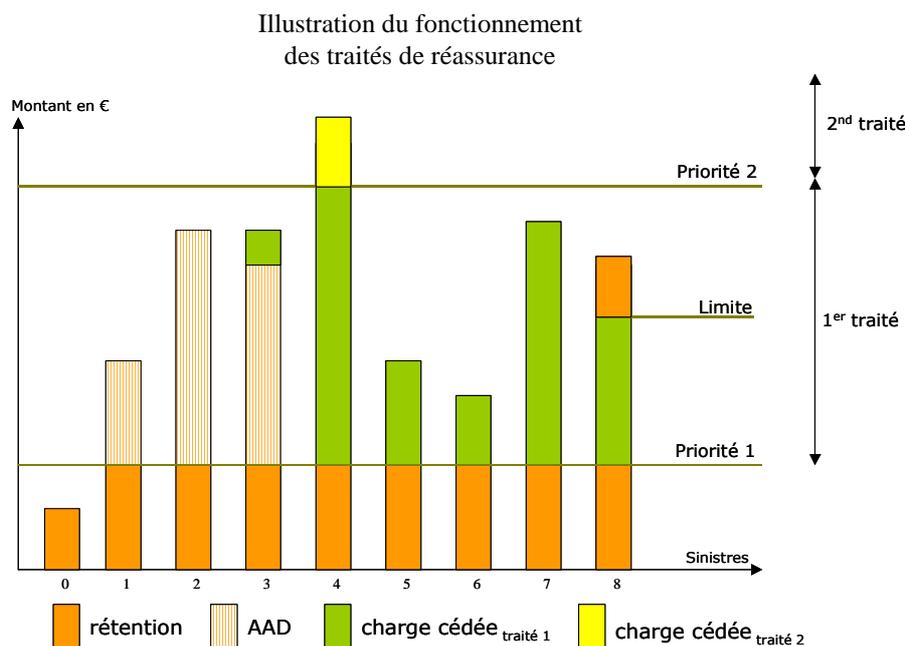
Enfin, le traité s'active seulement lorsque la charge cédée atteint un certain seuil. C'est le principe d'une franchise dont le terme usuel est *AAD* (Aggregate Annual Deductible).

#### **2<sup>nd</sup> traité :**

Pour les sinistres de montant élevé, un autre traité se déclenche à partir d'un seuil qu'on appellera *priorité 2*. Il est ici sans limite de garantie et sans AAD.

Avoir plusieurs traités, appelés aussi *sous-jacentes*, donne plus de souplesse et permet d'ajuster la couverture au profil du portefeuille. C'est particulièrement utile quand le portefeuille subit de fortes variations liées à la politique de souscription et à la surveillance.

Le dessin ci-après illustre le mécanisme de ces 2 traités.



Enfin, plus généralement, on peut rappeler que le choix des paramètres d'un traité de réassurance est issu d'un compromis entre plusieurs facteurs parmi lesquels :

1. La capacité de la cédante à absorber dans ses résultats la charge résiduelle qui est la somme des rétentions et, éventuellement, la charge au-delà de la limite,
2. L'aversion au risque de la cédante pouvant être mesurée par le niveau de quantile retenu,
3. Le prix du programme qui dépend en partie de la perception par le réassureur du risque transféré,
4. Dans l'optique de Solvabilité 2, la volonté de se réassurer plus pour économiser des fonds propres,
5. Le placement du programme : il faut trouver un réassureur qui l'accepte et qui a les qualités requises par la politique de cessions de la compagnie (rating financier, pérennité, expérience).

**Sinistres graves - Charge brute et nette de réassurance**

On détaille ci-dessous les 7 étapes de la procédure appliquée pour calculer **la charge nette de réassurance des graves** vue comme la différence entre la charge brute et la charge cédée au titre des 2 traités.

On notera  $Pr_1$  la priorité,  $Lim$  la limite,  $AAD$  la franchise pour les paramètres du traité 1 et  $Pr_2$  la priorité du traité 2. Dans la continuité du calcul de la charge des attritionnels, on a :

Etape 1 :  $\forall i \in [1, n]$ , soit  $n_i^{graves}$  les  $n$  nombres de sinistres graves tels que  $n_i^{graves} = n_i - n_i^{attri} - n_i^{IRCA}$ .

Etape 2 :  $\forall i \in [1, n]$ , pour tout nombre  $n_i^{graves}$  obtenu, on effectue  $n_i^{graves}$  tirages aléatoires de montants de sinistres selon la loi mélange calculée que l'on notera  $x_1^{n_i^{graves}}, x_2^{n_i^{graves}}, \dots, x_{n_i^{graves}}^{n_i^{graves}}$ .

Etape 3 :  $\forall i \in [1, n]$ ,  $\forall k \in [1, n_i^{graves}]$ , soit la suite  $(y_k^{n_i^{graves}})$  qui, pour toute simulation  $i$ , correspond à la sélection des sinistres touchant le 1<sup>er</sup> traité et au calcul de la charge pouvant être cédée ; la somme des cessions pour la simulation  $i$  étant notée  $Y^i$  :

$$y_k^{n_i^{graves}} = 1_{\left\{ \begin{matrix} Pr_1 \leq x_k^{n_i^{graves}} \\ \leq Pr_2 \end{matrix} \right\}} \cdot (x_k^{n_i^{graves}} - Pr_1) \text{ et } Y^i = \sum_{k=1}^{n_i^{graves}} y_k^{n_i^{graves}}$$

Etape 4 :  $\forall i \in [1, n]$ , on identifie alors si la charge cédée de la simulation  $i$  dépasse l'AAD. Si c'est le cas, on lui applique cette franchise pour obtenir la charge  $Z^i$  nette d'AAD :

$$Z^i = 1_{\{Y^i \geq AAD\}} \cdot (Y^i - AAD)$$

Etape 5 :  $\forall i \in [1, n]$ , on vérifie que la charge cédée nette d'AAD de la simulation  $i$  au titre du traité 1 ne dépasse pas la limite  $Lim$ . Enfin, pour avoir la totalité des cessions notée  $W^i$  de la simulation  $i$ , on lui ajoute les cessions au titre du traité 2.

$$W^i = 1_{\{Z^i \leq Lim\}} \cdot Z^i + 1_{\{Z^i > Lim\}} \cdot Lim + \sum_{k=1}^{n_i^{graves}} 1_{\left\{ \begin{matrix} Pr_2 \leq x_k^{n_i^{graves}} \end{matrix} \right\}} \cdot (x_k^{n_i^{graves}} - Pr_2)$$

$W^i$  est donc la charge cédée en réassurance pour la  $i^{\text{ème}}$  simulation de la charge brute notée

$$X^i = \sum_{k=1}^{n_i^{graves}} x_k^{n_i^{graves}}$$

Etape 6 : la charge nette de réassurance de la  $i^{\text{ème}}$  simulation est donc :  $V^i = X^i - W^i$

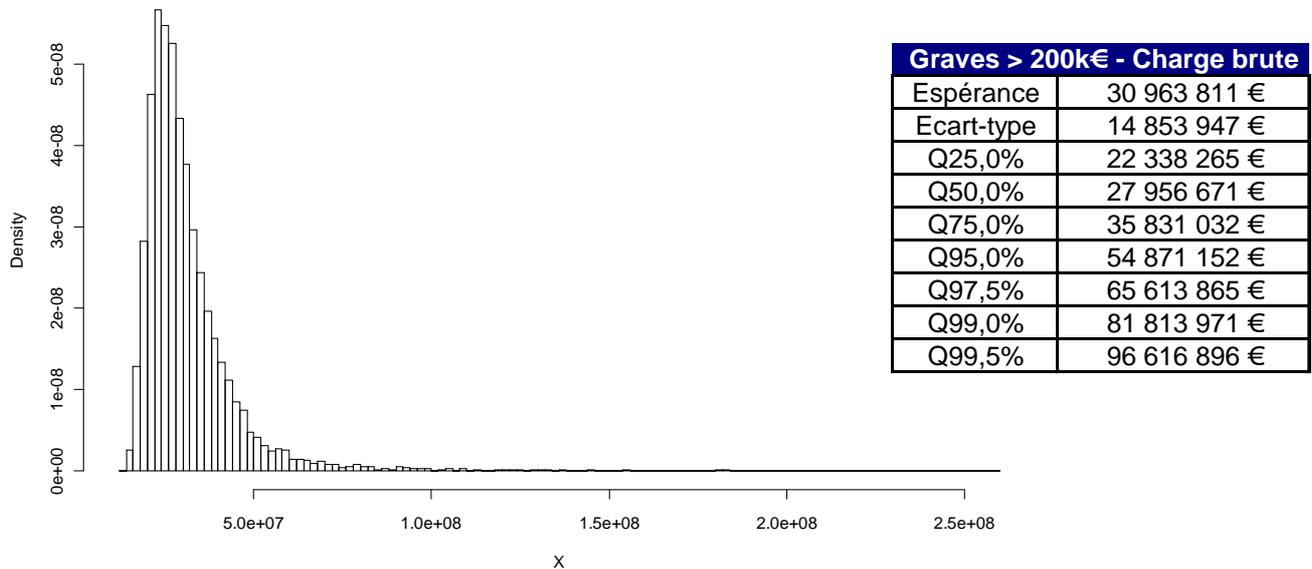
Etape 7 : un estimateur de l'espérance de la charge nette de réassurance des  $n$  simulations est :

$\hat{m} = \frac{1}{n} \sum_{i=1}^n V^i$  et un estimateur de la variance empirique de la charge nette de réassurance des  $n$  simulations

$$\text{est : } \hat{s}^2 = \frac{1}{n-1} \times \sum_{i=1}^n (V^i - \frac{1}{n} \sum_{i=1}^n V^i)^2$$

On obtient la charge brute de réassurance des sinistres graves, avec  $n = 10\ 000$  et  $\lambda_{graves} = 1,56\%$  :

Distribution des 10 000 simulations de la charge brute des graves



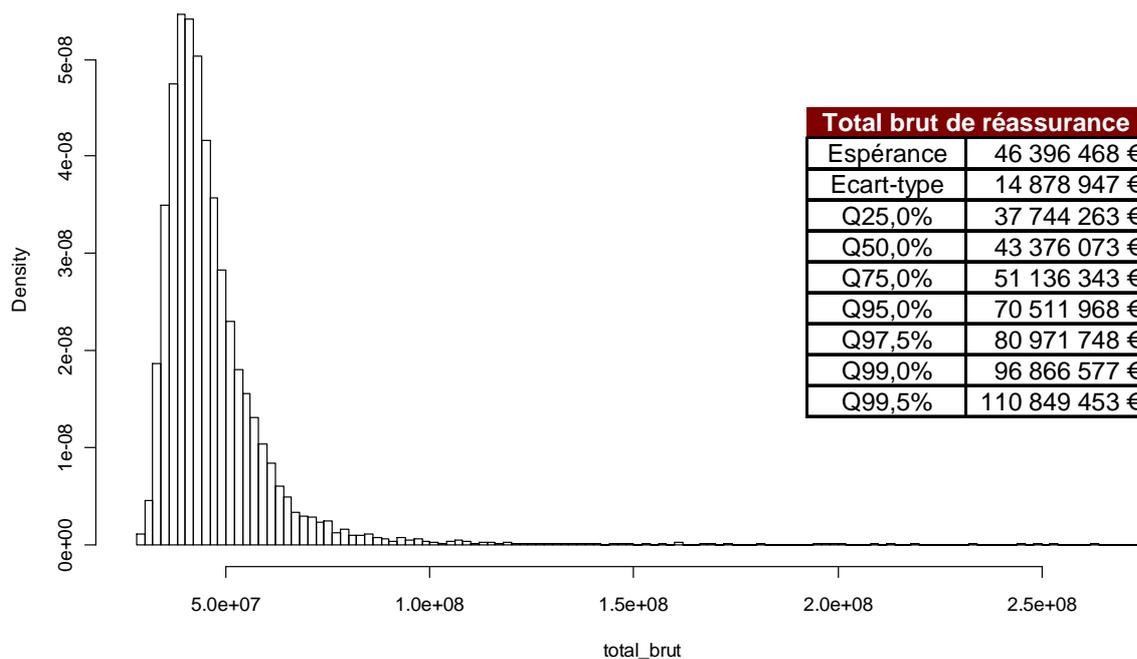
On note une forte volatilité conduisant à des probabilités importantes d'avoir une charge conséquente. Par exemple, tous les 100 ans ( $Q_{99\%}$ ), la charge peut être le triple de l'espérance.

Pour ces sinistres graves, il est donc légitime d'avoir recours à la réassurance XS pour diminuer le risque de dérive de la sinistralité.

### Charge totale brute et nette de réassurance

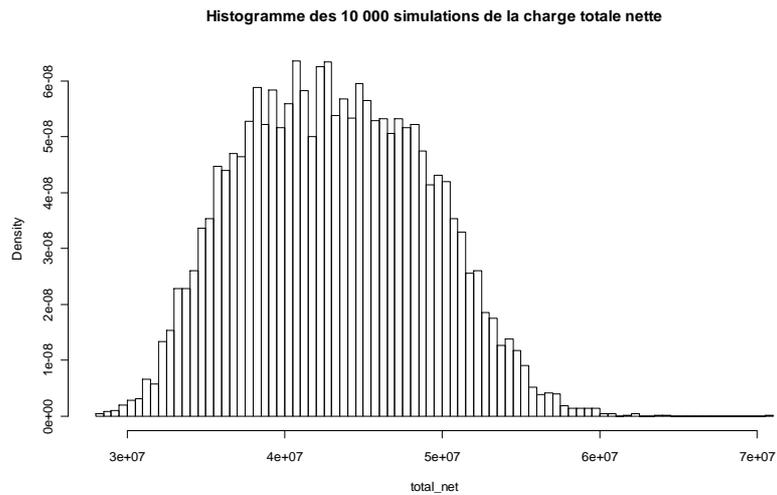
On observe sur la distribution des 10 000 simulations de charge totale brute de réassurance, le même phénomène que pour la charge brute des graves : un écart-type important et une queue de distribution épaisse. Cela induit un besoin de réassurance pour réduire ce risque de volatilité.

Histogramme des 10 000 simulations de la charge totale brute

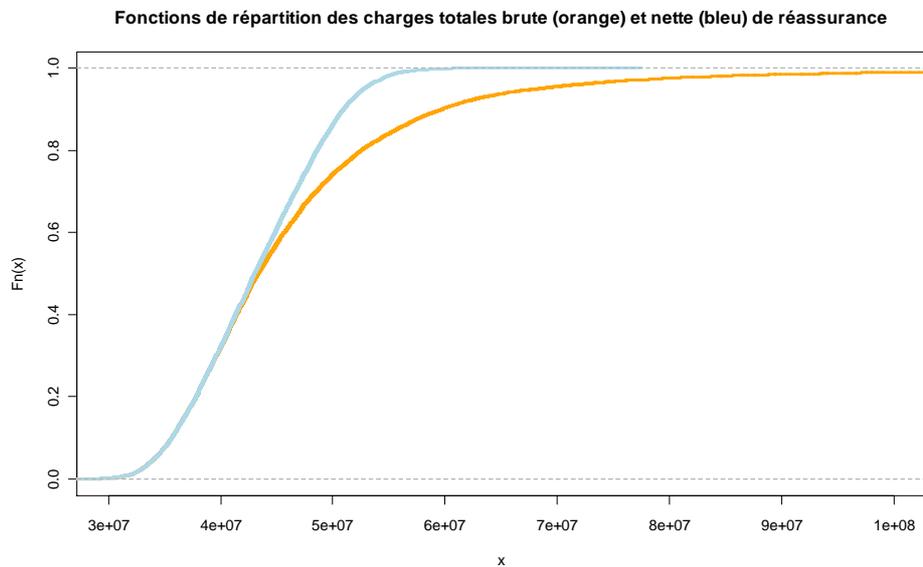


La distribution des 10 000 simulations de la **charge totale nette de réassurance** est visiblement bien concentrée ; l'écart-type est divisé par 2 entre la vision brute et la nette.

Total net de réassurance	
Espérance	43 141 155 €
Ecart-type	7 440 732 €
Q25,0%	37 658 163 €
Q50,0%	42 796 194 €
Q75,0%	48 352 986 €
Q95,0%	55 686 391 €
Q97,5%	57 942 441 €
Q99,0%	60 976 132 €
Q99,5%	62 709 247 €



Vu autrement, le graphe ci-dessous montre l'effet bénéfique de la réassurance XS qui cantonne la charge totale nette probable à des intervalles restreints alors que la charge brute a une queue de distribution épaisse.



### Charge totale brute et nette de réassurance avec une loi unique

Pour comparer, nous avons simulé la **charge totale nette** en utilisant la loi log-normale unique pour les graves à la place de la loi mélange P-B.

Total net de réassurance	
Espérance	41 594 785 €
Ecart-type	7 514 915 €
Q25,0%	36 142 655 €
Q50,0%	41 507 723 €
Q75,0%	46 891 408 €
Q95,0%	53 961 411 €
Q97,5%	56 194 920 €
Q99,0%	59 200 703 €
Q99,5%	62 050 291 €

Nous obtenons pour la loi unique une espérance inférieure, une variance supérieure et des quantiles inférieurs.

En accordant moins de poids à la queue de distribution, la loi unique minimise la charge totale nette et accentue la volatilité du résultat.

**Impacts des paramètres du traité sur la distribution de la charge totale nette**

Ayant modélisé un programme de réassurance complexe, nous souhaitons illustrer l'importance de certains paramètres en les faisant varier : priorité 1, priorité 2 et AAD ; la limite restant fixée. Le scénario 0 (noté sc. 0) est celui de **référence** (base 100). Les paramètres qui varient sont en gras.

Scenario	Paramètres du traité			Probabilité de			Statistiques de la charge nette			
	Priorité 1	AAD	Priorité 2	franchir l'AAD	céder traité 1	céder traité 2	Espérance		Quantile 99,5%	Variance
0	100	100	100	27,0%	25,3%	23,5%	43 141 155 €		62 709 247 €	5E+13
1	100	<b>87,5</b>	100	30,7%	29,0%	23,5%	42 914 935 €	-0,5%	61 899 941 €	5,2E+13
2	100	<b>112,5</b>	100	21,5%	19,8%	23,5%	43 319 702 €	0,4%	63 651 999 €	5,8E+13
3	<b>80</b>	100	100	32,2%	29,5%	23,5%	42 723 007 €	-1,0%	61 270 164 €	4E+13
4	<b>120</b>	100	100	23,4%	22,2%	23,5%	43 410 185 €	0,6%	64 442 295 €	6E+13
5	100	100	<b>89</b>	24,1%	22,8%	27,9%	43 020 399 €	-0,3%	62 452 241 €	5,4E+13
6	100	100	<b>111</b>	27,0%	24,6%	20,0%	43 224 566 €	0,2%	63 365 127 €	5,6E+13
7	<b>200</b>	<b>0</b>	100	49,8%	49,5%	23,5%	42 155 909 €	-2,3%	62 814 560 €	5,0E+13

Variation de l'AAD (sc. 1 et 2)

On retrouve bien le résultat attendu : plus l'AAD est petit, plus vite il est atteint, plus la probabilité de céder au traité 1 est grande et plus faibles sont les estimations de l'espérance et de la variance de la charge nette. Les variations d'AAD (+/-12,5 %) génèrent des variations significatives de variance de resp. +5,6% / -6,2%.

Variation de la priorité 1 (sc. 3 et 4)

On retrouve bien le résultat attendu : plus la priorité 1 est basse, plus la probabilité de céder au traité 1 est grande et plus faibles sont les estimations de l'espérance et de la variance de la charge nette. Les variations de priorité 1 (+/- 20 %) génèrent les plus fortes variations de variance de resp. +9,3% / -11,9%.

Variation de la priorité 2 (sc. 5 et 6)

On retrouve bien le résultat attendu : plus la priorité 2 est grande, plus vite l'AAD est atteint, plus la probabilité de céder au traité 1 est élevée ; cependant, la probabilité de céder au traité 2 étant plus faible, les estimations de l'espérance et de la variance de charge nette sont un peu plus élevées. Les variations de priorité 2 (+/- 11,1 %) génèrent de faibles variations de variance de resp. +1,5% / -1,7%.

Les paramètres du sc. 7 sont issus d'une proposition de notre département de réassurance ; les estimations de l'espérance et de la variance de la charge nette sont proches du sc. 3 montrant ainsi que ce sc. 7 offre une possible alternative.

Globalement, cette étude de cas montre peu d'impact sur l'espérance de la charge nette (entre -2,3% et +0,6%) ; en revanche, il y en a beaucoup sur sa variance (entre -11,9% et +9,3%).

Pour identifier le paramètre qui influence le plus la variance, nous utilisons une **mesure d'élasticité** définie

ainsi : soit y la variance et x un paramètre du traité, alors  $\epsilon(y, x) = \frac{\Delta y}{y} \times \frac{x}{\Delta x}$  est appelée élasticité de y relativement à x.

Scenario	1	2	3	4	5	6
Elasticité	49%	45%	60%	46%	15%	13%

On remarque que la baisse du paramètre (AAD, priorité 1, priorité 2) génère une élasticité supérieure à son augmentation. Ayant les élasticités les plus élevées, les sc.3 et 4 sortent du lot. **Le paramètre priorité 1 est donc le plus élastique à la variance.** Il est suivi de près par l'AAD (sc.1 et 2) ; en revanche la priorité 2 (sc.5 et 6) est le paramètre le moins élastique à la variance. Les spécificités du sc. 7 n'ont pas permis le calcul de l'élasticité.

Cette simulation en net de réassurance a permis de construire un modèle intéressant permettant de mesurer l'efficacité d'un traité. **Ce n'est pas l'objet de cette étude** mais en y rajoutant une méthode de tarification et de l'optimisation, on obtiendrait un modèle délivrant les paramètres de traité les plus adaptés à la sinistralité étudiée.

## V - E CONCLUSION

---

Dans cette partie, nous avons donc utilisé les théories du modèle collectif, de l'estimation du maximum de vraisemblance, des tests EDF, des lois tronquées et des lois mélanges. Ces méthodes éprouvées, devenues des classiques, nécessitent de la rigueur lors de leur mise en oeuvre, un travail conséquent en amont sur les données ainsi qu'en sortie pour générer les simulations.

Nos objectifs sont atteints in fine : les résultats obtenus par la méthode de simulation pour obtenir une charge nette de réassurance prennent bien en compte le double aléa sur les nombres et les coûts ainsi que la structure d'un programme de réassurance complexe.

Les principaux biais identifiés sont liés au traitement de masse des sinistres et des projections de leurs montants cumulés, à l'impossibilité de retraiter les sinistres majeurs et à l'unique taux d'inflation considéré.

Enfin, les calculs des charges nécessitent d'avoir un outil puissant de simulation dont la diffusion se généralise désormais avec SII.

A la relecture de notre démarche, nous listons ici les pistes d'amélioration qui nous semblent pertinentes :

- une modélisation affinée de l'inflation,
- une comparaison de ce modèle avec une loi mélange de log-normale et de Pareto,
- une évolution de la technique du demi-échantillon pour les statistiques EDF vers le bootstrap,
- une identification des changements en gestion pour dépolluer les cadences : accélération, coût horaire de tierce personne, taux de rente, table, forfait d'ouverture,
- en gestion, la mise en place de forfaits d'ouverture distincts selon le type de sinistre.

## VI - METHODE 2 : MODELISATION « MARKOVIENNE » DE LA CHARGE RC CORPORELLE

Nous nous intéressons ici à une autre méthode que l'on rencontre parfois sous le vocable suivant : matrices de transition markovienne. Une illustration de mise en œuvre est donnée dans l'article de Mahon<sup>49</sup>.

Cette méthode nous plaît a priori car elle permet une estimation précise de l'évolution des montants de sinistres individuels ce qui devrait conduire à estimer plus justement la loi qu'ils suivent.

Présenter cette seconde approche sur les graves permettra une comparaison avec la méthode 1, d'être plus critique sur les avantages et inconvénients de chacune et de mesurer que ce sujet complexe est loin d'être circonscrit par une unique approche.

En remarque préalable, nous pouvons noter que dans la méthode 1, en appliquant la cadence des sinistres agrégés (comportement d'ensemble) aux sinistres individuels puis, en considérant le résultat comme le montant ultime de ces sinistres, le **comportement des sinistres individuels est négligé**. En effet, figer un sinistre à un montant ultime  $x\%$  supérieur au montant actuel néglige l'information contenue dans l'évaluation cas par cas faite par les experts. **Le résultat est sous-optimal car n'intègre pas toute l'information disponible.**

### VI - A METHODE

Dans cette approche, nous utilisons une interprétation directe des matrices de transition : les vecteurs produits sont utilisés pour la **modélisation personnalisée des sinistres à l'ultime à partir d'un montant et d'une date d'évaluation**. Nous pouvons suivre les différents spectres d'évaluations annuelles probables que chaque sinistre corporel est susceptible de supporter pendant sa gestion et jusqu'à sa clôture.

Contrairement à la méthode 1, pour estimer le sinistre à l'ultime, nous utilisons sa **connaissance individuelle** (expertise gestionnaire) et donc de manière indirecte toute l'**information qualitative disponible**.

De plus, la théorie des matrices de transition rend compte des **différentes issues possibles**, nous l'illustrons dans le schéma suivant :

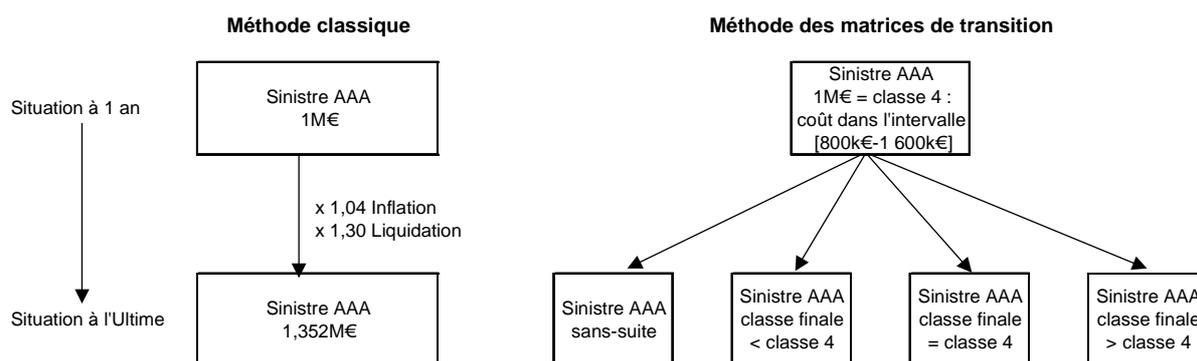


Figure 1 : Etats possibles d'évolution d'un sinistre de 12 mois à l'ultime selon la méthode utilisée

Le périmètre choisi concerne les sinistres RCC flottes de 1<sup>ère</sup> catégorie survenus entre 2000 et 2009 et dont l'évaluation de coût « As If » a franchi au moins une fois le seuil  $S_1 = 200$  k€ au cours de leur gestion. Le chapitre III-C nous autorise à ajouter les sinistres auto afin d'avoir plus de volume.

La liquidation est directement donnée par les probabilités basées sur les observations. Avec le recul que l'on a (profondeur d'historique) et l'agrégat des visions identiques (cf. schéma V-B.1), les observations nous semblent suffisamment stables et nombreuses pour être robustes.

<sup>49</sup> Transition Matrix Theory And Individual Claim Loss Development John B. Mahon, 2005.

## Principes

La théorie repose sur la chaîne de Markov (1906) qui est un processus de Markov à temps discret.

Un processus de **Markov à temps discret** est une séquence  $X_0, X_1, X_2, X_3, \dots$  de variables aléatoires à valeurs dans l'espace des états noté  $E$ . La valeur  $X_n$  étant l'état du processus à l'instant  $n$ .

C'est un processus sans mémoire, toute l'information utile pour la prédiction du futur étant contenue dans l'état présent du processus et qui s'écrit :  $\forall n \geq 0, \forall (i_0, \dots, i_{n-1}, i, j) \in E^{n+2}$ ,

$$P(X_{n+1} = j / X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = i) = P(X_{n+1} = j / X_n = i)$$

Le nombre  $P(X_1 = j / X_0 = i)$  noté  $p_{i,j}$  est la probabilité de transition de l'état  $i$  à l'état  $j$ . La famille de ces nombres  $(p_{i,j})_{(i,j) \in E^2}$  forme la **matrice de transition de la chaîne de Markov**.

Cette matrice est carrée, chacun de ses éléments est un réel compris entre 0 et 1 et elle est stochastique (la somme des éléments de chaque ligne vaut 1).

Le produit de 2 matrices de transition  $A = (a_{i,j})_{(i,j) \in E^2}$  et  $B = (b_{i,j})_{(i,j) \in E^2}$  est alors défini par :  $(AB)_{i,j} = \sum_{l \in E} a_{i,l} b_{l,j}$ .

La **loi de la chaîne de Markov**  $X = (X_n)_{n \geq 0}$  est caractérisée par le couple constitué de sa matrice de transition  $A = (a_{i,j})_{(i,j) \in E^2}$  et de sa loi initiale (la loi de  $X_0$ ) : pour tout  $n \geq 1$  la loi jointe de  $(X_0, X_1, X_2, X_3, \dots, X_n)$  est donnée par :

$$P(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = i_n) = P(X_0 = i_0) p_{i_0, i_1} p_{i_1, i_2} \dots p_{i_{n-1}, i_n}$$

## Construction des matrices

Une matrice de transition d'une chaîne de Markov est une matrice carrée contenant les probabilités de passer d'un **état** à un autre. Les états seront constitués de deux catégories : les sinistres « clos » et « en cours », ainsi que de leur montant.

Ensuite, on répartit les sinistres dans une matrice selon leur **classe** d'origine et leur classe finale.

La valeur des extrémités des classes de montants sont déterminées exponentiellement (elles augmentent d'un facteur constant) et chacune doit contenir un nombre significatif de sinistres.

Ainsi, les bornes des classes entre 2 valeurs  $M_1$  et  $M_2$  sont définies selon la formule suivante :

$$borne_{sup} = r * borne_{inf} \text{ où } r = \sqrt[p]{\frac{M_2}{M_1}} (> 1) \text{ et } p \text{ le nombre de classes entre } M_1 \text{ et } M_2.$$

Ces valeurs pourraient être choisies de façon arbitraire, mais le choix d'une croissance exponentielle fournit des informations supplémentaires sur les résultats obtenus.

En dehors de cet intervalle  $[M_1 ; M_2]$ , les classes sont définies à dire d'expert.

Dans l'exemple ci-dessous, on a choisi 14 classes dont 7 concernent les sinistres ouverts ( $O_0, O_1, O_2, O_3, O_4, O_5, O_6$ ) et 7 les sinistres clos ( $C_0, C_1, C_2, C_3, C_4, C_5, C_6$ ). Si on considère le sinistre AAA évalué à 1M€ à la fin de l'année de sa survenance, puis réévalué à 2M€ un an plus tard, on aurait alors :

Vision à 1 an				
Classe	Situation	Intervalle		Comptage
		Borne basse	Borne haute	
$O_0$	En cours	0	0	
$O_1$	En cours	0	200 000	
$O_2$	En cours	200 000	400 000	
$O_3$	En cours	400 000	800 000	
$O_4$	En cours	800 000	1 600 000	1
$O_5$	En cours	1 600 000	3 200 000	
$O_6$	En cours	3 200 000	6 400 000	
$C_0$	Clos	0	0	
$C_1$	Clos	0	200 000	
$C_2$	Clos	200 000	400 000	
$C_3$	Clos	400 000	800 000	
$C_4$	Clos	800 000	1 600 000	
$C_5$	Clos	1 600 000	3 200 000	
$C_6$	Clos	3 200 000	6 400 000	

Figure 2 : Représentation de tous les états possibles, désignés par un numéro de classe. Dans comptage nous avons indiqué notre sinistre AAA d'un montant 1M€ lors de son évaluation à 1 an.

Vision à 2 ans				
Classe	Situation	Intervalle		Comptage
		Borne basse	Borne haute	
$O_0$	En cours	0	0	
$O_1$	En cours	0	200 000	
$O_2$	En cours	200 000	400 000	
$O_3$	En cours	400 000	800 000	
$O_4$	En cours	800 000	1 600 000	
$O_5$	En cours	1 600 000	3 200 000	1
$O_6$	En cours	3 200 000	6 400 000	
$C_0$	Clos	0	0	
$C_1$	Clos	0	200 000	
$C_2$	Clos	200 000	400 000	
$C_3$	Clos	400 000	800 000	
$C_4$	Clos	800 000	1 600 000	
$C_5$	Clos	1 600 000	3 200 000	
$C_6$	Clos	3 200 000	6 400 000	

Figure 3 : Etat du sinistre AAA, mais avec un montant atteignant désormais 2M€ à 2 ans. Il est maintenant dans la classe  $O_5$ .

Une matrice correspond alors à l'observation dynamique de la variation de l'évaluation des sinistres pendant une période annuelle ; ex : de 1 à 2 ans ou de 4 à 5 ans. Ensuite, pour obtenir une matrice de Markov, on divise chaque nombre par le total des sinistres de la ligne. Il y a donc n-1 matrices pour n années étudiées.

Transition de 1 à 2 ans							
Classe Initiale	Classe Finale						
	$O_0$	...	$O_5$	$O_6$	$C_0$	...	$C_6$
$O_0$							
⋮							
$O_4$			1				
$O_5$							
$O_6$							
$C_0$							
⋮							
$C_6$							

Figure 4 : Matrice de transition pour le sinistre AAA donné à titre d'exemple, appartenant à la classe  $O_4$  en maturité 1 an et étant passé dans la classe  $O_5$  à 2 ans.

Soient  $(n_{i,j})_{0 \leq i \leq 13, 0 \leq j \leq 13}$  désignant les nombres de sinistres passant de la classe  $O_k$  ou  $C_k$  vers la classe  $O_l$  ou  $C_l$  avec  $(k,l) \in (0,1,2,3,4,5,6)^2$ . A gauche, la matrice concerne la transition des sinistres en nombre entre leur état initial (vision à 1 an) et leur état final (vision à 2 ans). A droite, la matrice devient une matrice markovienne par passage des nombres aux probabilités. On a  $n_i = \sum_{j=0}^{13} n_{ij}$ .

Transition de 1 à 2 ans							
Classe Initiale	Classe Finale						
	$O_0$	...	$O_6$	$C_0$	...	$C_6$	Somme
$O_0$	$n_{0,0}$	...	$n_{0,6}$	$n_{0,7}$	...	$n_{0,13}$	$n_0$
$O_1$	$n_{1,0}$	...	$n_{1,6}$	$n_{1,7}$	...	$n_{1,13}$	$n_1$
⋮	⋮	...	⋮	⋮	...	⋮	⋮
$C_0$	$n_{7,0}$	...	$n_{7,6}$	$n_{7,7}$	...	$n_{7,13}$	$n_7$
$C_1$	$n_{8,0}$	...	$n_{8,6}$	$n_{8,7}$	...	$n_{8,13}$	$n_8$
⋮	⋮	...	⋮	⋮	...	⋮	⋮
$C_6$	$n_{13,0}$	...	$n_{13,6}$	$n_{13,7}$	...	$n_{13,13}$	$n_{13}$

Figure 5 : Matrice des nombres de 1 à 2 ans.

Transition de 1 à 2 ans							
Classe Initiale	Classe Finale						
	$O_0$	...	$O_6$	$C_0$	...	$C_6$	
$O_0$	$n_{0,0}/n_0$	...	$n_{0,6}/n_0$	$n_{0,7}/n_0$	...	$n_{0,13}/n_0$	
$O_1$	$n_{1,0}/n_1$	...	$n_{1,6}/n_1$	$n_{1,7}/n_1$	...	$n_{1,13}/n_1$	
⋮	⋮	...	⋮	⋮	...	⋮	
$C_0$	$n_{7,0}/n_7$	...	$n_{7,6}/n_7$	$n_{7,7}/n_7$	...	$n_{7,13}/n_7$	
$C_1$	$n_{8,0}/n_8$	...	$n_{8,6}/n_8$	$n_{8,7}/n_8$	...	$n_{8,13}/n_8$	
⋮	⋮	...	⋮	⋮	...	⋮	
$C_6$	$n_{13,0}/n_{13}$	...	$n_{13,6}/n_{13}$	$n_{13,7}/n_{13}$	...	$n_{13,13}/n_{13}$	

Figure 6 : Matrice des probabilités de 1 à 2 ans.

La méthode prend bien en compte tous les cas de figure :

Classe Initiale	Classe Finale					
	O <sub>0</sub>	...	O <sub>6</sub>	C <sub>0</sub>	...	C <sub>6</sub>
O <sub>0</sub>	n <sub>0,0</sub> /n <sub>0</sub>	...	n <sub>0,6</sub> /n <sub>0</sub>	n <sub>0,7</sub> /n <sub>0</sub>	...	n <sub>0,13</sub> /n <sub>0</sub>
O <sub>1</sub>	n <sub>1,0</sub> /n <sub>1</sub>	...	n <sub>1,6</sub> /n <sub>1</sub>	n <sub>1,7</sub> /n <sub>1</sub>	...	n <sub>1,13</sub> /n <sub>1</sub>
⋮	⋮	⋮	⋮	⋮	⋮	⋮
C <sub>0</sub>	n <sub>7,0</sub> /n <sub>7</sub>	...	n <sub>7,6</sub> /n <sub>7</sub>	n <sub>7,7</sub> /n <sub>7</sub>	...	n <sub>7,13</sub> /n <sub>7</sub>
C <sub>1</sub>	n <sub>8,0</sub> /n <sub>8</sub>	...	n <sub>8,6</sub> /n <sub>8</sub>	n <sub>8,7</sub> /n <sub>8</sub>	...	n <sub>8,13</sub> /n <sub>8</sub>
⋮	⋮	⋮	⋮	⋮	⋮	⋮
C <sub>6</sub>	n <sub>13,0</sub> /n <sub>13</sub>	...	n <sub>13,6</sub> /n <sub>13</sub>	n <sub>13,7</sub> /n <sub>13</sub>	...	n <sub>13,13</sub> /n <sub>13</sub>

Légende

initial	→	final
ouvert	→	ouvert
ouvert	→	clos
clos	→	ouvert
clos	→	clos

On obtient ensuite les matrices ultimes par multiplication des matrices entre elles, de la première jusqu'à la dernière (celle assimilée à l'ultime) :

Classe Initiale	Classe Finale					
	O <sub>0</sub>	...	O <sub>6</sub>	C <sub>0</sub>	...	C <sub>6</sub>
O <sub>0</sub>	p <sub>0,0</sub>	...	p <sub>0,6</sub>	p <sub>0,7</sub>	...	p <sub>0,13</sub>
O <sub>1</sub>	p <sub>1,0</sub>	...	p <sub>1,6</sub>	p <sub>1,7</sub>	...	p <sub>1,13</sub>
⋮	⋮	⋮	⋮	⋮	⋮	⋮
C <sub>0</sub>	p <sub>7,0</sub>	...	p <sub>7,6</sub>	p <sub>7,7</sub>	...	p <sub>7,13</sub>
C <sub>1</sub>	p <sub>8,0</sub>	...	p <sub>8,6</sub>	p <sub>8,7</sub>	...	p <sub>8,13</sub>
⋮	⋮	⋮	⋮	⋮	⋮	⋮
C <sub>6</sub>	p <sub>13,0</sub>	...	p <sub>13,6</sub>	p <sub>13,7</sub>	...	p <sub>13,13</sub>

=

Classe Initiale	Classe Finale					
	O <sub>0</sub>	...	O <sub>6</sub>	C <sub>0</sub>	...	C <sub>6</sub>
O <sub>0</sub>	n <sub>0,0</sub> /n <sub>0</sub>	...	n <sub>0,6</sub> /n <sub>0</sub>	n <sub>0,7</sub> /n <sub>0</sub>	...	n <sub>0,13</sub> /n <sub>0</sub>
O <sub>1</sub>	n <sub>1,0</sub> /n <sub>1</sub>	...	n <sub>1,6</sub> /n <sub>1</sub>	n <sub>1,7</sub> /n <sub>1</sub>	...	n <sub>1,13</sub> /n <sub>1</sub>
⋮	⋮	⋮	⋮	⋮	⋮	⋮
C <sub>0</sub>	n <sub>7,0</sub> /n <sub>7</sub>	...	n <sub>7,6</sub> /n <sub>7</sub>	n <sub>7,7</sub> /n <sub>7</sub>	...	n <sub>7,13</sub> /n <sub>7</sub>
C <sub>1</sub>	n <sub>8,0</sub> /n <sub>8</sub>	...	n <sub>8,6</sub> /n <sub>8</sub>	n <sub>8,7</sub> /n <sub>8</sub>	...	n <sub>8,13</sub> /n <sub>8</sub>
⋮	⋮	⋮	⋮	⋮	⋮	⋮
C <sub>6</sub>	n <sub>13,0</sub> /n <sub>13</sub>	...	n <sub>13,6</sub> /n <sub>13</sub>	n <sub>13,7</sub> /n <sub>13</sub>	...	n <sub>13,13</sub> /n <sub>13</sub>

x ... x

Classe Initiale	Classe Finale					
	O <sub>0</sub>	...	O <sub>6</sub>	C <sub>0</sub>	...	C <sub>6</sub>
O <sub>0</sub>	n' <sub>0,0</sub> /n' <sub>0</sub>	...	n' <sub>0,6</sub> /n' <sub>0</sub>	n' <sub>0,7</sub> /n' <sub>0</sub>	...	n' <sub>0,13</sub> /n' <sub>0</sub>
O <sub>1</sub>	n' <sub>1,0</sub> /n' <sub>1</sub>	...	n' <sub>1,6</sub> /n' <sub>1</sub>	n' <sub>1,7</sub> /n' <sub>1</sub>	...	n' <sub>1,13</sub> /n' <sub>1</sub>
⋮	⋮	⋮	⋮	⋮	⋮	⋮
C <sub>0</sub>	n' <sub>7,0</sub> /n' <sub>7</sub>	...	n' <sub>7,6</sub> /n' <sub>7</sub>	n' <sub>7,7</sub> /n' <sub>7</sub>	...	n' <sub>7,13</sub> /n' <sub>7</sub>
C <sub>1</sub>	n' <sub>8,0</sub> /n' <sub>8</sub>	...	n' <sub>8,6</sub> /n' <sub>8</sub>	n' <sub>8,7</sub> /n' <sub>8</sub>	...	n' <sub>8,13</sub> /n' <sub>8</sub>
⋮	⋮	⋮	⋮	⋮	⋮	⋮
C <sub>6</sub>	n' <sub>13,0</sub> /n' <sub>13</sub>	...	n' <sub>13,6</sub> /n' <sub>13</sub>	n' <sub>13,7</sub> /n' <sub>13</sub>	...	n' <sub>13,13</sub> /n' <sub>13</sub>

Chaque matrice peut être vue comme une série de vecteurs unidimensionnels : une ligne représente un vecteur. Chaque vecteur prédit le montant final en fonction du montant initial et nous permettra de décrire l'évolution d'un sinistre individuel en cours.

Règles à appliquer pour un bon fonctionnement de la méthode

Nous avons identifié 3 règles utiles :

1. Pour aboutir à un état ultime où tous les sinistres sont clos, la matrice relatant les transitions les plus anciennes (10 à 11 ans dans notre cas) doit être **corrigée** en forçant la fermeture des sinistres encore ouverts à cette date. L'hypothèse posée est de les clôturer dans la classe qu'ils occupaient alors. Le biais généré est du même ordre que pour la méthode 1 (IV-D.2.), quand on force la clôture des sinistres dans les triangles de montants (durée de développement de 11 ans).
2. De plus cette même matrice de transition entre 10 et 11 ans, revêt un caractère tout particulier : s'appuyant sur un unique exercice (2000), elle est donc la seule à relier les transitions de 10 à 11 ans. Le nombre de sinistres étudié étant faible, il nous paraît important de **renforcer** la loi de transition de cet état vers l'ultime en élargissant le périmètre à des exercices plus anciens.
3. Enfin, chaque matrice de transition est **auditée** pour permettre le produit matriciel. Par exemple, pour les matrices de jeune maturité, nous n'avons pas énormément de sinistres dans les classes « clos » ; à l'inverse, dans les dernières matrices, ce phénomène se produit dans les classes « ouvert ». Lorsqu'on n'a pas de sinistre dans une classe (le total de la ligne est nul), on regarde les

résultats des classes immédiatement supérieure et inférieure. Si elles ne sont pas nulles, on prolonge par continuité la classe étudiée.

### **Orientation vers un modèle log-normalien**

L'étude des histogrammes des distributions pour chaque classe et chaque projection ( $1 \sim +\infty, 2 \sim +\infty, \dots$ ) ainsi que le calcul des moments (skewness et kurtosis) orientent Mahon vers un modèle log-normal.

Il calcule alors les estimations des paramètres des lois de chaque classe et pour chaque projection ( $1 \sim +\infty, 2 \sim +\infty, \dots$ ) par la méthode des moments. Il observe qu'elles sont liées entre elles par une relation linéaire en fonction de la maturité ( $1 \sim +\infty, 2 \sim +\infty, \dots$ ) et de la classe initiale du sinistre.

On note qu'il n'y a pas de mesure de vérification du modèle par le calcul de statistiques de test. De plus, pour le calcul des estimations des paramètres des lois, il s'appuie sur le milieu de chaque classe. D'autres choix seraient possibles tels que l'utilisation de la moyenne ou de la médiane de chaque classe.

### **Calcul de la charge**

Mahon ne le détaillant pas, nous proposerons une méthode dans la partie applications.

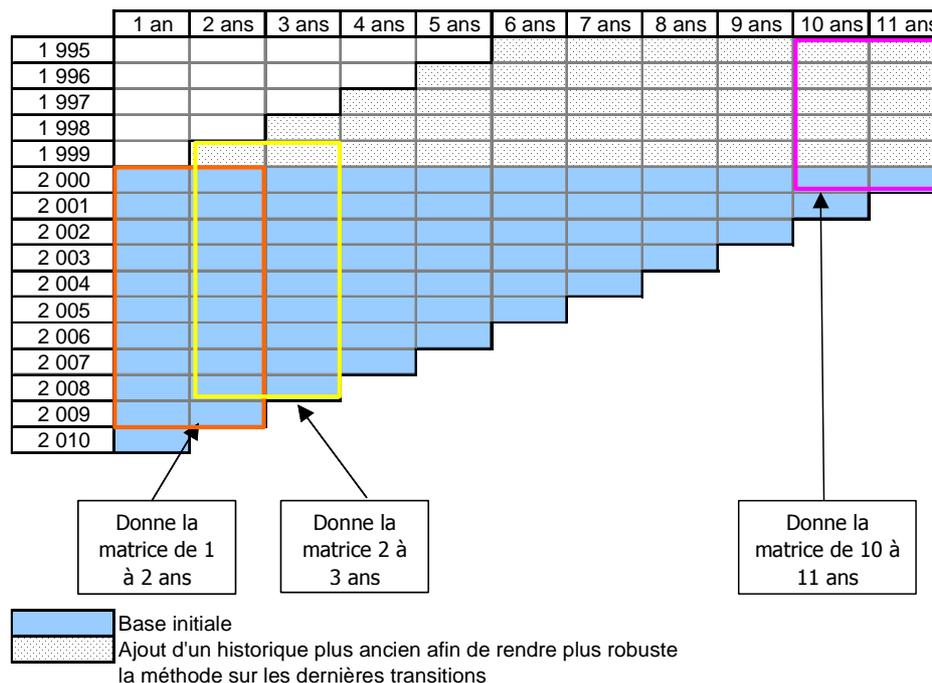
## VI - B APPLICATIONS : APPROCHE MAHON (LOI UNIOUE)

### VI - B . 1. MISE EN PLACE DE LA METHODE : DONNEES ET MATRICES

#### Les données :

Au préalable les sinistres ont été revalorisés en « As If » 2010 (IV-D.2.). Le périmètre choisi concerne les sinistres RC corporelle auto et flottes de 1ère catégorie survenus entre 1995 et 2009 et dont l'évaluation a franchi au moins une fois le seuil  $S_1 = 200$  k€ au cours de leur gestion. Il représente 2579 sinistres. **Sans l'apport des sinistres auto nous n'aurions que 509 sinistres.**

Nous indiquons ci-dessous visuellement le périmètre des données selon les matrices de transition :



Sur le portefeuille considéré, la quasi-totalité des sinistres est connue après 3 ans. De plus, quand on crée les matrices de transition de « 1 à 2 ans » et « 2 à 3 ans », on prend toute l'information disponible, soit tous les sinistres de tous les exercices ayant vécu ces transitions. Ainsi, les nombres que l'on perd au titre des IBNR des années les plus récentes peuvent être considérés comme marginaux et donc pas de nature à modifier les ventilations des nombres par classe. C'est pour cette raison qu'on n'a pas à gérer une particularité IBNR sur les années récentes.

#### Détermination des classes

Nous voulons qu'une borne repose sur le seuil  $S_1$  des graves.

Pour déterminer une charge nette de réassurance, nous souhaitons faire coïncider ces bornes avec la priorité  $Pr_1$  du traité le plus travaillant.

Nous retiendrons  $p = 4$  classes entre  $S_1$  (200 k€) et une valeur compatible avec  $Pr_i$ , ce qui nous permet d'accrocher aussi la valeur  $S_2$  (1,25 M€). On a alors  $r = 1,88$ .

La borne supérieure de chaque classe est égale à 1,88 fois la borne inférieure. Nous avons décidé de regrouper les classes entre 1 € et 30 k€ pour en limiter le nombre.

Pour définir la borne supérieure de la dernière classe, nous pourrions prendre  $9 \text{ M€} * 1,88 = 16,9 \text{ M€}$ . Seul un sinistre serait alors concerné. Nous ne retenons donc pas la croissance exponentielle pour cette dernière borne et nous nous orientons vers son calcul via l'espérance conditionnelle.

L'espérance conditionnelle de X sachant que X est supérieur à z = 9M€ est :

$$E(X / X > z) = \frac{e^{\frac{\mu + \sigma^2}{2}} \cdot [1 - \phi(\frac{\ln z - \mu}{\sigma} - \sigma)]}{1 - \phi(\frac{\ln z - \mu}{\sigma})} \quad (\text{voir démonstration en annexe VIII-I})$$

En utilisant la loi log-normale LN(13,79 ; 1,46) qui ajuste le mieux au sens du test de KS la distribution des sinistres supérieurs à 1 250 k€ et z = 9 000 k€ - 1250 k€ = 7 750 k€ (effet de la translation), on obtient l'application numérique suivante : E(X / X > 7 750 k€) = 18 767 k€ + 1 250 k€ = 20 017 k€. On retient donc la valeur arrondie à 20 000 k€ comme borne supérieure de la dernière classe.

Nous obtenons alors les classes suivantes :

Classe	Intervalle
Ibnr	IBNR
1	0
2	]0;30k€ [
3	[30k€;57k€[
4	[57k€;106k€[
5	[106k€;200k€[
6	[200k€;376k€[
7	[376k€;707k€[
8	[707k€;1 250k€[
9	[1 250k€;2 500k€[
10	[2 500k€;4 700k€[
11	[4 700k€;9 000k€[
12	[9 000k€;20 000k€[

Par ailleurs par rapport à la méthode utilisée par Mahon, nous avons ajouté une classe supplémentaire « Inbr » correspondant aux IBNR. Cette classe nous sera utile lors de la simulation de la charge.

**Construction des matrices intermédiaires de probabilités 1~2 ans, 2~3 ans, ..., 10 ans~∞**

Nous construisons dans un premier temps les matrices de transition intermédiaires. Par exemple pour le développement de 1 à 2 ans, nous obtenons les 3 étapes suivantes :

**Etape 1 : Répartition des nombres de sinistres<sup>50</sup>**

		2 ans													TOTAL						
		Ibnr	1O	...	12O	1C	2C	3C	4C	5C	6C	7C	8C	9C		10C	11C	12C			
1 an	Ibnr	19	0		0	0	0	0	0	0	2	0	0	0	0	0	0	168			
	1O	0	37		0	0	0	0	0	0	0	0	0	0	0	0	0	77			
	2O	0	7		2	0	0	0	1	1	1	0	0	0	0	0	0	530			
	3O	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	66			
	4O	0	0		0	0	0	0	0	0	1	0	0	0	0	0	0	144			
	5O	0	1		0	0	0	0	0	0	2	0	0	0	0	0	0	281			
	6O	0	0		0	4	0	1	1	8	4	0	0	0	0	0	0	367			
	7O	0	0		0	0	1	0	1	3	2	0	0	0	0	0	0	162			
	8O	0	1		0	0	0	0	0	0	0	0	0	0	0	0	0	86			
	9O	0	0		0	1	0	0	0	0	0	0	0	0	0	0	0	67			
	10O	0	0		1	0	0	0	0	0	0	0	0	0	0	0	0	36			
	11O	0	0		2	0	0	0	0	0	0	0	0	0	0	0	0	32			
	12O	0	0		4	0	0	0	0	0	0	0	0	0	0	0	0	4			
1C	0				0	0	0	0	0	0	0	0	0	0	0	0	0				
2C					0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
3C					0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
⋮																					4
12C					0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
															2 025						

<sup>50</sup> Les résultats détaillés figurent en annexe VIII-J

Nous précisons que la vision 1 an correspond à la vision à la fin de l'année de survenance du sinistre, aussi pour un sinistre survenu en mars 2009, la maturité 1 an correspond à la vision de ce sinistre à fin 2009 ; il en est de même pour les autres maturités.

La matrice 10~11 ans est, quant à elle, modifiée afin de forcer les sinistres encore ouverts dans l'état « clos » de la classe où ils se situent. Ainsi, nous obtenons des sinistres uniquement dans la partie « clos » de l'état final.

Au bout de 11 ans plus de 86% des sinistres sont clos et l'on connaît 97% de la charge, il ne s'agit donc pas d'une hypothèse invraisemblable.

Etape 2 : Matrice de probabilités<sup>51</sup>

		2 ans																			
		Ibnr	1O	...	12O	1C	2C	3C	4C	5C	6C	7C	8C	9C	10C	11C	12C	TOTAL			
1 an	Ibnr	11%	0%		0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	1			
	1O	0%	48%		0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1			
	2O	0%	1%		0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1			
	3O	0%	0%		0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1			
	4O	0%	0%		0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	1			
	5O	0%	0%		0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	1			
	6O	0%	0%		0%	1%	0%	0%	0%	2%	1%	0%	0%	0%	0%	0%	0%	1			
	7O	0%	0%		0%	0%	1%	0%	1%	2%	1%	0%	0%	0%	0%	0%	0%	1			
	8O	0%	1%		0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1			
	9O	0%	0%		0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1			
	10O	0%	0%		3%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1			
	11O	0%	0%		6%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1			
	12O	0%	0%		100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1			
1C	0%				0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0			
2C					0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1	
3C					0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0	
⋮																					
12C					0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0

Etape 3 : Correction<sup>51</sup>

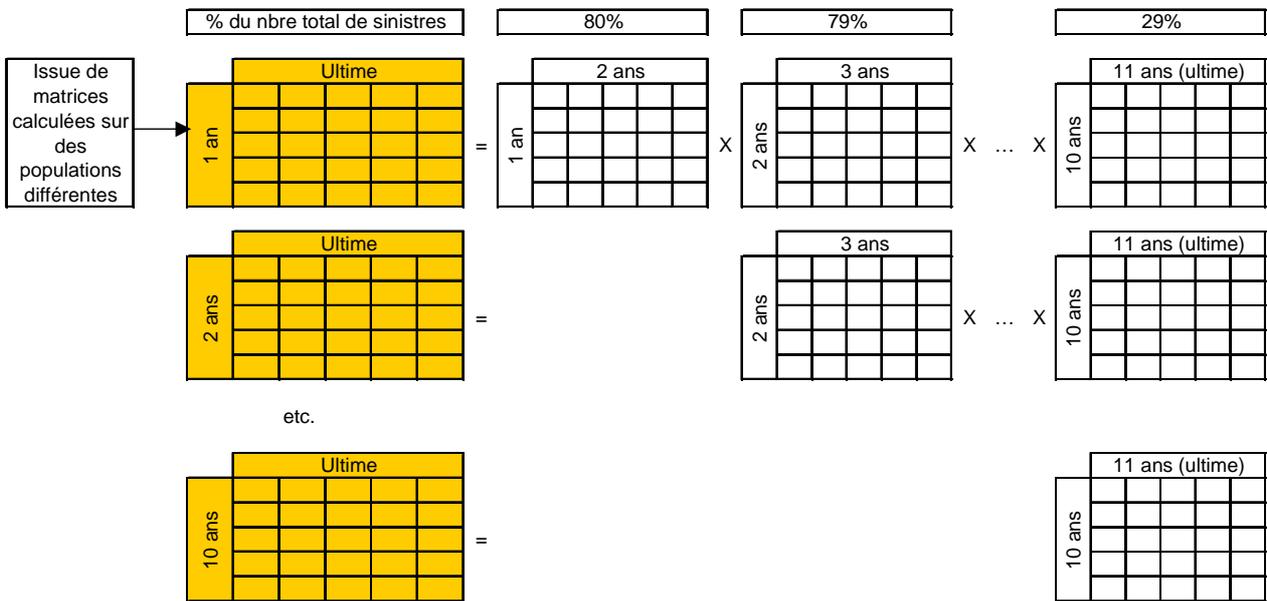
Nous remarquons que certaines classes sont vides, par exemple la classe « 1C ». Nous effectuons la correction permettant le produit matriciel.

		2 ans																			
		Ibnr	1O	...	12O	1C	2C	3C	4C	5C	6C	7C	8C	9C	10C	11C	12C	TOTAL			
1 an	Ibnr	11%	0%		0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	1			
	1O	0%	48%		0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1			
	2O	0%	1%		0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1			
	3O	0%	0%		0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1			
	4O	0%	0%		0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	1			
	5O	0%	0%		0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	1			
	6O	0%	0%		0%	1%	0%	0%	0%	2%	1%	0%	0%	0%	0%	0%	0%	1			
	7O	0%	0%		0%	0%	1%	0%	1%	2%	1%	0%	0%	0%	0%	0%	0%	1			
	8O	0%	1%		0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1			
	9O	0%	0%		0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1			
	10O	0%	0%		3%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1			
	11O	0%	0%		6%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1			
	12O	0%	0%		100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1			
1C	0%				100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1			
2C					0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1	
3C					0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1	
⋮																					
12C					0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	1

<sup>51</sup> Les résultats détaillés figurent en annexe VIII-J

**Construction des matrices de probabilités ultimes 1 an ~ ∞, 2 ans ~ ∞, ..., 10 ans ~ ∞**

Pour obtenir les matrices de N à l'ultime, nous multiplions entre-elles les matrices intermédiaires jusqu'à ce que toutes les matrices de transition soient utilisées :



Il est important de noter qu'une matrice k ans ~ ∞ n'est pas construite à partir du même périmètre de sinistres : multiplication de matrices dont les périmètres sont différents (plus ou moins d'exercices).

Après multiplication des matrices de transition entre elles, nous obtenons **la matrice ultime 1 an ~ ∞** :

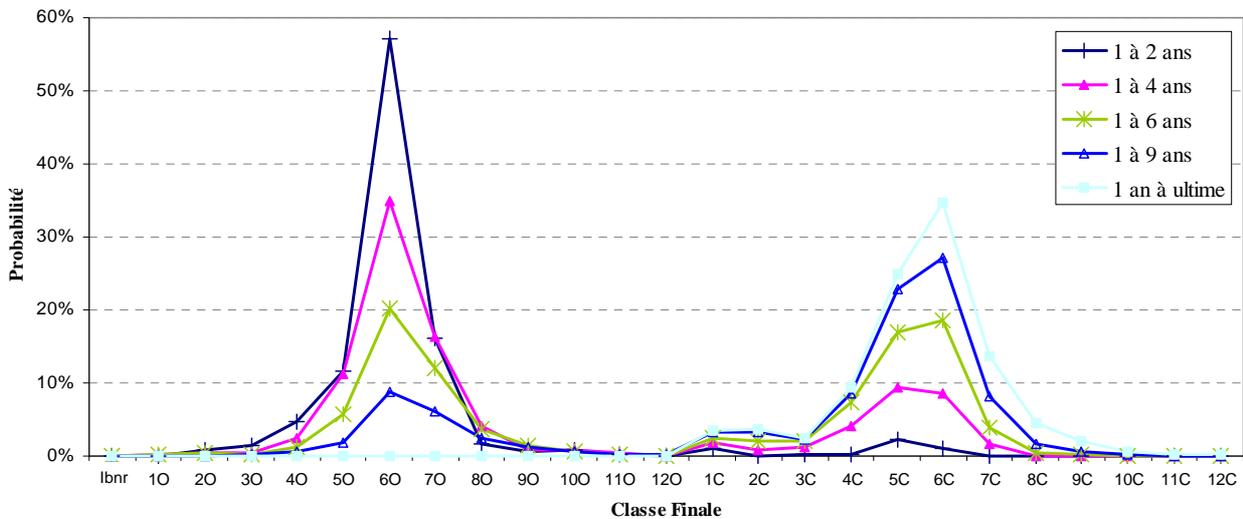
		Ultime																
		Ibnr	1O	...	12O	1C	2C	3C	4C	5C	6C	7C	8C	9C	10C	11C	12C	TOTAL
1 an	Ibnr	0%				2%	5%	2%	9%	20%	31%	15%	7%	4%	1%	1%	1%	1
	1O		2%	7%	2%	7%	17%	28%	18%	10%	5%	2%	1%	1%	1%	1		
	2O		2%	5%	3%	9%	20%	30%	15%	7%	5%	2%	1%	1%	1%	1		
	3O		2%	4%	3%	10%	24%	33%	13%	5%	4%	2%	1%	0%	1			
	4O		3%	4%	3%	11%	24%	34%	14%	5%	2%	1%	0%	0%	1			
	5O		2%	4%	2%	9%	25%	36%	14%	5%	2%	1%	0%	0%	1			
	6O		3%	4%	2%	9%	25%	35%	14%	5%	2%	1%	0%	0%	1			
	7O		2%	4%	2%	7%	17%	30%	25%	9%	3%	1%	0%	0%	1			
	8O		2%	3%	1%	4%	10%	20%	29%	20%	8%	2%	0%	0%	1			
	9O		3%	4%	2%	3%	7%	12%	20%	18%	22%	6%	2%	1%	1			
	10O		1%	5%	3%	2%	5%	8%	12%	10%	22%	16%	10%	5%	1			
	11O		0%	4%	3%	1%	2%	4%	10%	10%	14%	13%	26%	14%	1			
12O	0%	1%	3%	0%	1%	3%	5%	1%	2%	16%	5%	63%	1					
1C		0%	$I_{12}$															1
2C	1																	
...	1																	
12C	1																	

$I_{12}$  : Matrice identité de taille 12

Par ailleurs, cette méthode est intéressante car elle nous permet de décrire l'évolution d'un sinistre individuel en cours.

Par exemple, nous représentons ci-dessous l'évolution d'un sinistre individuel initialement situé dans la classe 6. Nous nous sommes limité à 5 transitions pour la lisibilité du graphique.

Evolution d'un sinistre initialement situé dans la classe 6



Une caractéristique de ce graphique est la présence de deux pics distincts, l'un situé dans la partie des sinistres « ouverts » et le second situé dans la partie des sinistres « clos ». A mesure que le sinistre évolue, le pic de gauche diminue au profit du pic de droite, ce qui correspond à une diminution des sinistres ouverts et une augmentation des sinistres clos. La courbe ultime 1 an  $\sim \infty$  nous indique que les sinistres sont clos majoritairement dans la classe 6 et nous n'avons plus de sinistre ouvert (probabilité nulle sur les classes Ibncr à 12O). Le sinistre possède une certaine probabilité d'être sans-suite (1C), et le reste de la probabilité est répartie selon la distribution de montants indiquée.

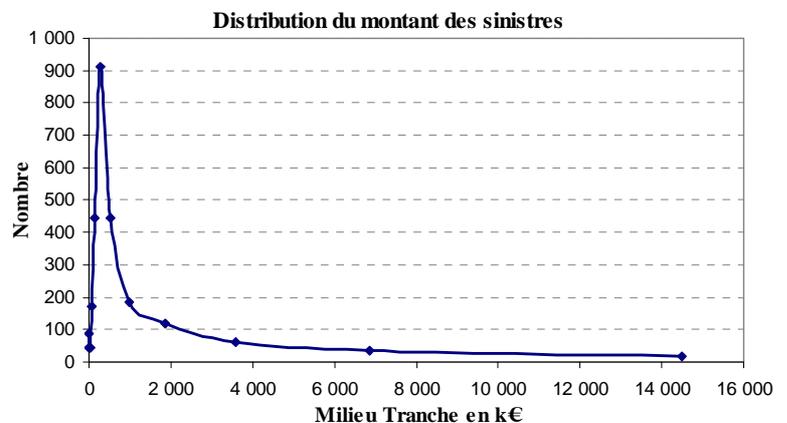
Désormais nous utiliserons uniquement la partie de la matrice ultime 1 an  $\sim \infty$  correspondant aux classes ouvertes à 1 an et closes à l'ultime (partie orange de la matrice vue page précédente).

## VI - B . 2. VERIFICATION DU CONTEXTE LOG-NORMALIEN

### Profil du montant des sinistres

La figure ci-dessous représente la courbe du profil des montants des sinistres. Le nombre de sinistres dans chaque intervalle de montants est représenté en fonction du montant moyen de l'intervalle.

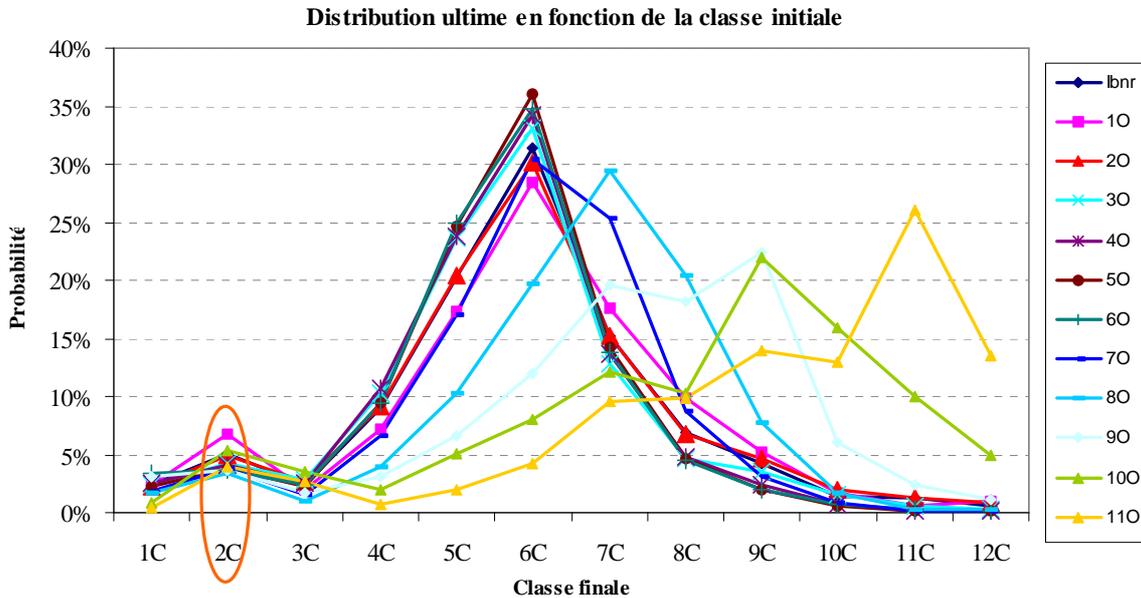
CLASSE	INTERVALLE	Milieu tranche (k€)	Nbr sinistres
1	0	0	44
2	]0;30€[	15	87
3	]30k€;57k€[	44	44
4	]57k€;106k€[	82	174
5	]106k€;200k€[	153	447
6	]200k€;376k€[	288	914
7	]376k€;707k€[	542	447
8	]707k€;1 250k€[	979	186
9	]1 250k€;2 500k€[	1 875	120
10	]2 500k€;4 700k€[	3 600	63
11	]4 700k€;9 000k€[	6 850	37
12	]9 000k€;20 000k€[	14 500	16
			2 579



La courbe présente une forte asymétrie vers la droite, ce qui nous indique qu'il s'agit d'une distribution à queue épaisse, mais cela nous fournit peu d'informations supplémentaires.

**Distributions ultimes des sinistres**

Nous allons étudier les probabilités de passage dans les classes finales selon la classe initiale. Pour cela, nous nous restreignons à la matrice 1 an  $\sim \infty$  :

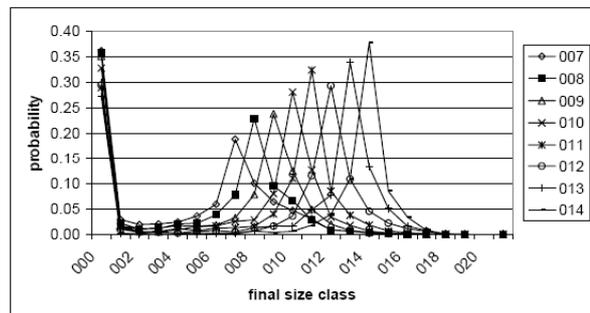


Cette figure représente toutes les distributions ultimes pour la transition entre 1 an et l’ultime. Chaque courbe correspond à une classe de montant initial différente (lbnr, 10, ..., 120). Elles présentent presque toutes un aspect de courbe en « cloche » avec un peu d’asymétrie, ce qui suggère une distribution log-normale.

La classe 2C a des probabilités un peu surélevées par rapport aux classes 1C et 3C. Cela est sûrement dû au fait qu’elle ne suit pas la progression exponentielle des autres intervalles. En effet, elle contient tous les sinistres dont les montants sont compris dans ]0 ; 30 k€[ et qui auraient été répartis dans plusieurs classes si une subdivision plus fine et exponentielle avait été définie.

Les classes inférieures à la classe initiale 6O (200 k€) se retrouvent majoritairement à l’ultime dans la classe 6C [200 k€ ; 376 k€], ce qui est logique compte tenu de notre méthode de sélection des sinistres. Pour les classes supérieures, nous n’observons pas le même phénomène que celui décrit dans l’article de Mahon.

Dans ce dernier, les courbes correspondant à des classes élevées sont d’aspect identique et sont décalées l’une par rapport à l’autre d’une quantité constante. Mahon indique alors que si la forme des courbes ne change pas en fonction de la classe initiale, cela suggère que les paramètres régissant l’étalement sont semblables pour toutes les courbes. Mahon constate aussi que, plus le sinistre est ouvert dans une classe de montant élevé, plus il a de chance de rester dans cette classe et donc que la volatilité diminue.



Dans notre cas, **la probabilité la plus élevée n’est pas toujours identique à la classe initiale** :

Classe initiale	6O	7O	8O	9O	10O	11O
Classe finale ayant la probabilité la plus élevée	6C	6C	7C	9C	9C	11C

Ce tableau montre une liquidation plutôt en boni.

D'autre part, on constate que **plus le sinistre est dans une classe initiale haute et plus la volatilité augmente** (aspect d'aplatissement). Il semble qu'on s'éloigne du modèle de Mahon où toutes les matrices sont régies par la même famille de loi (log-normale) et où l'on trouve une relation entre les estimations des paramètres de chaque loi.

Afin de valider ce constat, nous continuons les étapes de l'étude de Mahon et estimons maintenant les paramètres des lois  $LN(\mu_{\text{ClasseInitiale}, \text{maturité}}, \sigma_{\text{ClasseInitiale}, \text{maturité}})$  pour chaque classe et à chaque maturité.

### VI - B . 3. ESTIMATION DES PARAMETRES DES LOIS UNIQUES $LN(\mu_{i,t}, \sigma_{i,t})$

Nous rappelons que  $Y \sim LN(\mu, \sigma^2)$  si  $\ln(Y) \sim N(\mu, \sigma^2)$ . Nous utiliserons donc les logarithmes pour estimer les paramètres de la loi par la méthode des moments :

Soit  $X$  une variable aléatoire continue suivant une loi normale  $N(\mu, \sigma^2)$  dont la valeur des paramètres n'est pas connue. En utilisant la méthode des moments nous obtenons comme estimateurs sans biais :

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \text{ et } \hat{\sigma}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Remarques :

- Nous retirons les sinistres nuls (sans-suite) en renormalisant les probabilités afin d'ajuster une distribution. Pour cela, nous multiplions les probabilités par le ratio :  $a = 1 + P[\text{coût} = 0] / (1 - P[\text{coût} = 0])$ .

- Nous calculons les coefficients de skewness et de kurtosis qui sont des indicateurs numériques de la forme de la distribution observée : ils donnent un indice de l'éloignement ou du rapprochement de la série de valeurs de celle d'une distribution de loi normale.

Par exemple pour la classe 6 et la maturité  $1 \sim \infty$ , nous obtenons le tableau de synthèse suivant :

borne inf	borne sup	milieu intervalle	X=ln (milieu)	proba	proba normalisée	X*prob	X <sup>2</sup> * prob	3ème moment (X-mu) <sup>3</sup> x prob_norm	4ème moment (X-mu) <sup>4</sup> x prob_norm
IBNR									
-999 999	0	0		3%					
0	30 000	15 000	9,62	4%	4%	0,358	3,440	-0,755	2,060
30 000	57 000	43 500	10,68	2%	3%	0,272	2,906	-0,117	0,195
57 000	106 000	81 500	11,31	9%	10%	1,110	12,555	-0,109	0,113
106 000	200 000	153 000	11,94	25%	26%	3,083	36,810	-0,017	0,007
200 000	376 000	288 000	12,57	35%	36%	4,520	56,815	0,004	0,001
376 000	707 000	541 500	13,20	14%	14%	1,882	24,848	0,090	0,077
707 000	1 250 000	978 500	13,79	5%	5%	0,649	8,950	0,143	0,208
1 250 000	2 500 000	1 875 000	14,44	2%	2%	0,298	4,308	0,191	0,402
2 500 000	4 700 000	3 600 000	15,10	1%	1%	0,104	1,563	0,143	0,394
4 700 000	9 000 000	6 850 000	15,74	0%	0%	0,041	0,653	0,103	0,351
9 000 000	20 000 000	14 500 000	16,49	0%	0%	0,026	0,432	0,113	0,469
<b>Normalisation de la proba:</b>				<b>1,0352</b>					
						$\bar{X}$	$\bar{X}^2$	3ème moment	4ème moment
						12,34	153,28	-0,21	4,28
						S <sup>2</sup>	S	skewness	kurtosis
						0,92	0,96	-0,24	5,07

La colonne encadrée correspond au vecteur unidimensionnel de la classe 6 de la matrice  $1 \sim \infty$  (VI-B.1). Dans ce cas, nous obtenons  $\mu = 12,34$ ,  $\sigma = 0,96$ , le skewness = - 0,24 et le kurtosis = 5,07. Le coefficient d'asymétrie proche de 0 nous indique que la distribution est symétrique, quant au kurtosis il nous indique que la distribution est « pointue » ; ce qui nous suggère que les montants ultimes possèdent bien une distribution log-normale.

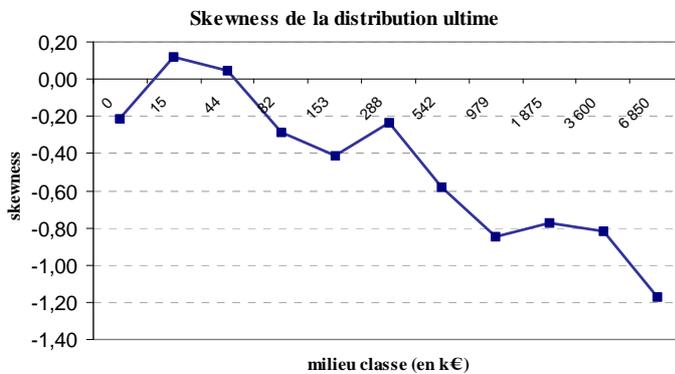
De la même manière, nous estimons les paramètres  $\mu$  et  $\sigma$  de chaque loi à partir des données empiriques : pour toutes les maturités  $k \sim \infty$ ,  $k \in (1, \dots, 10)$  et pour toutes les classes initiales (Ibnr, C11, ..., C112). Soit  $10 \times 13 = 130$  lois.

Par exemple, pour la maturité 1 an  $\sim \infty$ , nous obtenons pour chaque classe initiale les résultats suivants :

De 1 an à l'ultime								
Classe initiale	Milieu Classe	log (milieu classe)	Paramètres de la distribution ultime				$\mu / \sigma$ log(milieu)	$\mu / \sigma$
			skewness	kurtosis	$\mu$	$\sigma$		
IBNR			0,01	4,35	12,49	1,18		10,60
1	0		-0,21	4,03	12,56	1,26		9,95
2	15 000	9,62	0,12	4,24	12,51	1,22	1,30	10,29
3	43 500	10,68	0,04	4,58	12,38	1,08	1,16	11,45
4	81 500	11,31	-0,28	4,68	12,33	0,99	1,09	12,43
5	153 000	11,94	-0,41	4,99	12,34	0,94	1,03	13,15
6	288 000	12,57	-0,24	5,07	12,34	0,96	0,98	12,88
7	541 500	13,20	-0,58	4,82	12,60	1,02	0,95	12,38
8	978 500	13,79	-0,84	4,89	12,99	1,08	0,94	12,00
9	1 875 000	14,44	-0,77	4,01	13,41	1,33	0,93	10,09
10	3 600 000	15,10	-0,82	3,24	13,83	1,69	0,92	8,19
11	6 850 000	15,74	-1,17	4,03	14,52	1,70	0,92	8,53
12	14 500 000	16,49	-1,93	6,15	15,58	1,55	0,94	10,07

Les 2 colonnes de droite serviront ensuite.

Nous représentons graphiquement ci-dessous le skewness selon le milieu de la classe :



Le coefficient d'asymétrie est positif pour les faibles montants (15 k€ et 43,5 k€) et négatif pour les autres montants.

Pour les classes de faible montant, cela s'explique par le fait que, le sinistre étant déclaré avec un faible montant, on attendra un règlement final supérieur.

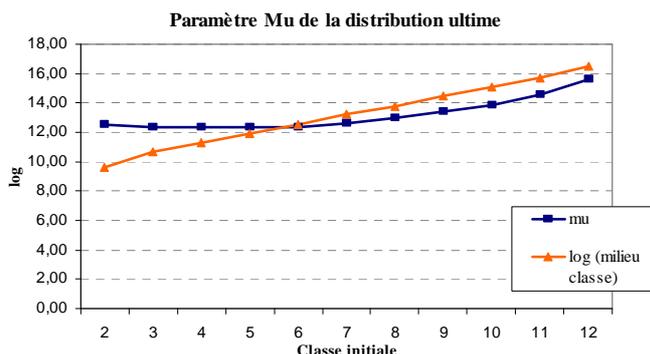
Maintenant que nous avons estimé un très grand nombre de paramètres (130 lois donc 260 paramètres), nous allons chercher des liens simplificateurs entre eux afin de les utiliser plus facilement. Notre intérêt est aussi d'atténuer l'impact de certains paramètres que l'on pourrait considérer comme « mal » estimés (par exemple : pour la dernière maturité 10 ans  $\sim \infty$  contenant peu de sinistres, l'estimation des paramètres peut aboutir à des valeurs peu fiables).

#### VI - B . 4. RECHERCHE DE LIENS

Nous commençons par chercher des liens entre les paramètres des lois  $LN(\mu_{i,1}, \sigma_{i,1})$  de la maturité 1  $\sim \infty$  où  $i$  correspond à la classe initiale  $i \in (IBNR, 1O, \dots, 12O)$ .

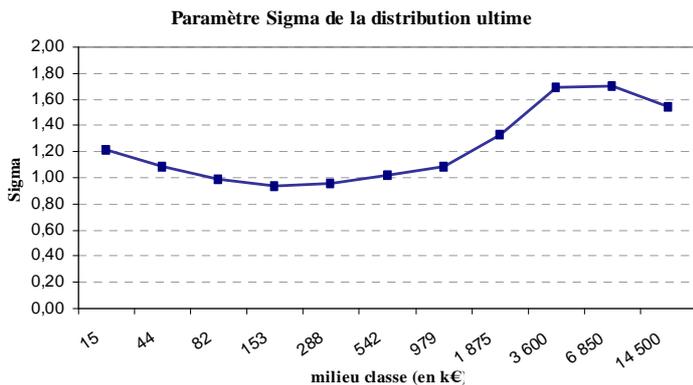
##### Maturité 1 $\sim \infty$

Nous observons dans le tableau précédent que  $\mu$  est très proche du logarithme du milieu de coût de la classe initiale ; en effet, le rapport de ces deux quantités est proche de 1.



Pour poursuivre la même idée, nous représentons ci-contre les valeurs de  $\mu$  et du logarithme du milieu de coût de la classe initiale. Il semble exister une relation entre ces deux quantités.

De même que pour  $\mu$ , nous cherchons un lien entre  $\sigma$  et le milieu de coût de la classe initiale :



Contrairement à Mahon, nous n'observons pas de décroissance progressive de  $\sigma$  quand le montant des sinistres augmente.

Nous remarquons une décroissance jusqu'à la classe 6 (288 k€) puis une augmentation de  $\sigma$ .

Nous confirmons notre remarque vue en VI-B.2 : **il ne semble pas exister de relation entre ces deux quantités.**

Nous avons étudié le comportement de l'évolution des sinistres pour une maturité donnée (1 an à l'ultime) et avons pu montrer qu'il existait une relation entre les  $\mu$ . Cependant pour créer un modèle sur des données réelles, nous devons pouvoir décrire ce comportement quelle que soit la maturité.

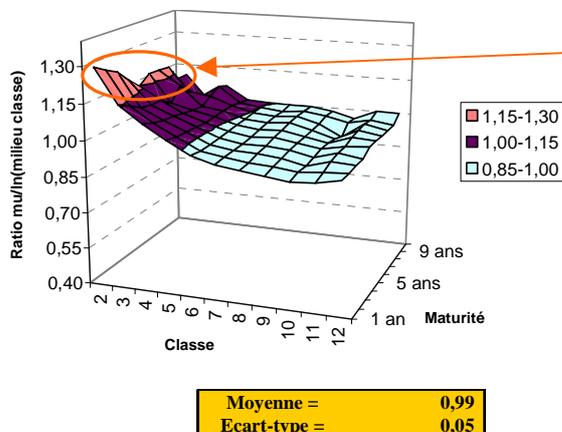
Toutes maturités 1 an ~  $\infty$ , 2 ans ~  $\infty$ , ..., 10 ans ~  $\infty$

Ajustement de la surface des  $\mu$

Le tableau suivant indique les résultats obtenus pour l'estimation du paramètre  $\mu$  selon la maturité et la classe initiale :

Classe	Milieu Classe	log (milieu classe)	Ratio Mu / ln(moyenne de la classe) selon la maturité initiale												
			1 an	2 ans	3 ans	4 ans	5 ans	6 ans	7 ans	8 ans	9 ans	10 ans			
IBNR															
1	0														
2	15 000	9,62	1,30	1,27	1,25	1,19	1,15	1,18	1,18	1,17	1,10	1,10			
3	43 500	10,68	1,16	1,13	1,10	1,09	1,09	1,06	1,12	1,07	1,02	1,00			
4	81 500	11,31	1,09	1,07	1,06	1,06	1,04	1,04	1,03	1,06	1,08	1,06			
5	153 000	11,94	1,03	1,02	1,02	1,02	1,01	1,02	1,02	1,02	1,01	1,01			
6	288 000	12,57	0,98	0,98	0,98	0,98	0,98	0,99	0,99	1,00	1,00	1,00			
7	541 500	13,20	0,95	0,96	0,96	0,96	0,97	0,98	0,98	0,99	0,99	1,00			
8	978 500	13,79	0,94	0,95	0,96	0,96	0,97	0,98	0,99	1,00	1,00	1,00			
9	1 875 000	14,44	0,93	0,94	0,95	0,96	0,96	0,96	0,97	0,97	0,98	0,98			
10	3 600 000	15,10	0,92	0,93	0,93	0,93	0,94	0,94	0,94	0,92	0,96	0,96			
11	6 850 000	15,74	0,92	0,94	0,95	0,96	0,96	0,98	0,98	0,99	0,98	1,00			
12	15 000 000	16,52	0,94	0,94	0,95	0,95	0,97	0,97	0,97	0,97	0,97	1,00			

Nous représentons graphiquement la surface du ratio  $\mu/\ln(\text{milieu de la classe})$  selon la classe et la maturité :



La classe 2 est atypique (bord relevé), qui peut s'expliquer par le périmètre des données choisies. En effet nous sélectionnons uniquement les sinistres dépassant 200k€ ce qui entraîne un biais : les sinistres de la CL2 auront une évolution vers les montants élevés (phénomène déjà vu à travers le skewness).

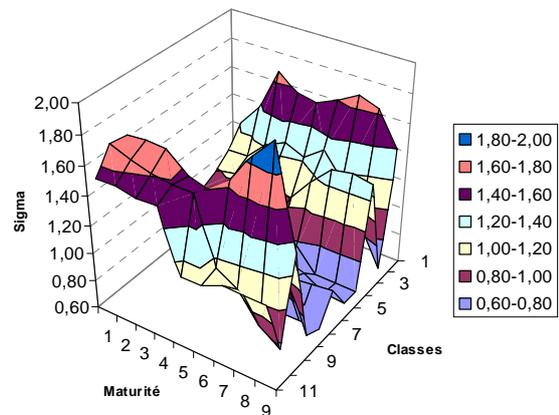
Le reste du graphique est une surface plane (faible écart-type), on en conclut que **le logarithme du milieu du coût de la classe initiale prédit bien  $\mu$ .**

On calcule alors la moyenne et l'écart-type de la distribution du ratio sur toutes les classes et toutes les maturités, hors la classe 2 (atypique). **Notre modèle d'estimation de  $\mu$  est alors :  $\mu = 0,99.\ln(x)$** , où x est le milieu de la classe du sinistre.

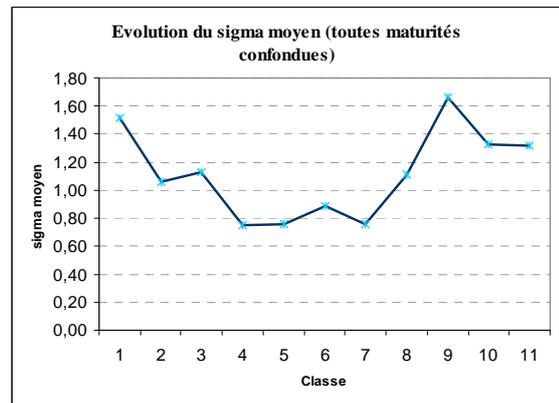
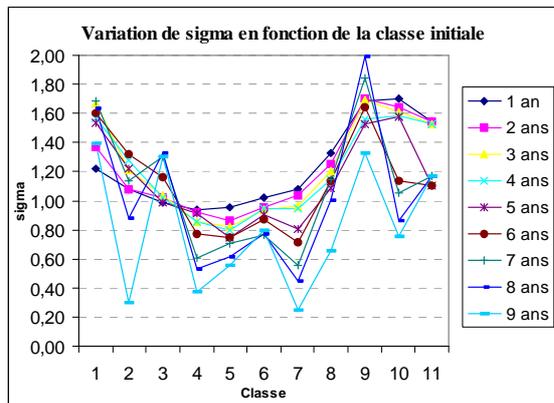
Ajustement de la surface des  $\sigma$

Le tableau de gauche indique les résultats obtenus pour l'estimation du paramètre  $\sigma$  selon la maturité et la classe initiale. Nous représentons sur le graphique de droite la surface formée par les  $\sigma$  des distributions ultimes en fonction de la maturité et de la classe initiale.

Classe	Milieu Classe	ln (milieu classe)	Sigma selon la maturité initiale										
			1 an	2 ans	3 ans	4 ans	5 ans	6 ans	7 ans	8 ans	9 ans		
IBNR			1,18	1,29	1,44	1,25	1,35	1,30					
0	0		1,26	1,27	1,10	1,09	0,97	0,84	1,99	0,80			
1	15 000	9,62	1,22	1,37	1,67	1,57	1,54	1,60	1,69	1,63	1,39		
2	43 500	10,68	1,08	1,08	1,21	1,28	1,22	1,32	1,14	0,88	0,30		
3	81 500	11,31	0,99	1,02	1,04	1,02	0,99	1,16	1,30	1,33	1,30		
4	153 000	11,94	0,94	0,92	0,86	0,85	0,91	0,77	0,60	0,53	0,37		
5	288 000	12,57	0,96	0,87	0,83	0,80	0,75	0,75	0,71	0,62	0,56		
6	541 500	13,20	1,02	0,96	0,94	0,95	0,90	0,87	0,76	0,77	0,80		
7	978 500	13,79	1,08	1,04	0,96	0,95	0,80	0,72	0,56	0,45	0,25		
8	1 875 000	14,44	1,33	1,25	1,21	1,15	1,09	1,14	1,15	1,00	0,66		
9	3 600 000	15,10	1,69	1,70	1,70	1,56	1,53	1,65	1,85	1,99	1,33		
10	6 850 000	15,74	1,70	1,64	1,61	1,59	1,58	1,14	1,05	0,86	0,76		
11	15 000 000	16,52	1,55	1,55	1,52	1,52	1,11	1,11	1,17	1,17	1,17		



En lien avec nos remarques précédentes, nous n'observons pas de surface plane et décroissante selon les classes. Pour compléter l'observation, nous changeons de perspective :



- Sur le graphique de gauche, quelle que soit la maturité, nous faisons les mêmes constats que sur la maturité 1 : contrairement à Mahon, nous n'observons pas de décroissance progressive de  $\sigma$  quand le montant de sinistres augmente.

La maturité 9 comprenant peu de sinistres est assez volatile, ce qui explique les soubresauts constatés.

- Sur le graphique de droite, nous observons une évolution décroissante du sigma moyen des classes 1 à 4, une stabilité entre les classes 4 et 7 et enfin une croissance.

Si nos résultats étaient proches de ceux de Mahon, le sigma devrait être décroissant, et ce, quelle que soit la maturité. En effet, on pourrait penser qu'un sinistre évalué à un montant élevé doit moins varier qu'un sinistre de plus faible montant et avoir plus de chance de se clôturer dans la classe initiale où il se situait. Ce n'est en effet pas sans raison médicale qu'il a été ouvert dans une classe haute.

Ce constat différent peut s'expliquer par les hypothèses suivantes :

- La gestion des sinistres : les équipes de gestionnaires n'ayant pas les mêmes pratiques ou les mêmes méthodes d'évaluation introduisent un biais.
- Un volume pas assez significatif de sinistres dans les classes hautes (139 sinistres pour les classes de 9 à 12 soit 7% du total) entraîne un biais dans la matrice de transition ultime.
- Des distributions ultimes par classe qui seraient mal prises en compte par une loi unique.
- La méthode de détermination des estimateurs des paramètres de la loi : on utilise le milieu de la classe, ce qui entraîne une sur-estimation des variances et des coûts à l'arrivée, constat fait par Mahon lui-même.

**Conclusion :**

En lien avec les travaux de Mahon, nous retrouvons bien une proportionnalité entre  $\mu$  et le logarithme du milieu de la classe initiale.

Cependant, nous constatons que la variance ne décroît pas dans les classes élevées quand on applique la méthode avec une unique loi log-normale. Nous ne pouvons donc pas appliquer de régression linéaire pour exprimer  $\sigma$  en fonction de la maturité et du logarithme du milieu de coût de la classe initiale.

Le modèle de loi unique n'étant pas satisfaisant, **nous proposons d'utiliser une loi mélange de deux lois log-normales** pour mieux appréhender la distribution des classes élevées afin de diminuer cette volatilité. Nous sommes alors en cohérence avec la méthode 1 dans laquelle les sinistres graves sont bien modélisés par un mélange de lois.

**VI - C APPLICATIONS : NOUVELLE APPROCHE (LOI MELANGE)**

Pour chaque classe initiale en maturité 1, nous allons estimer les paramètres de la loi mélange.

A l'image de la méthode 1, nous aurions souhaité appliquer la méthode de Benabbou-Partrat. Mais, nos données étant groupées, c'est-à-dire que nous ne connaissons pas les observations individuelles, force est de constater que nous ne pouvons l'utiliser.

Aussi, nous testerons d'autres méthodes d'estimation des paramètres s'appuyant sur la minimisation de distance. Pour les appliquer sur les graves, nous nous limiterons aux sinistres ayant un coût ultime supérieur à 200 k€ (classes finales 6 à 12).

**VI - C . 1. ESTIMATION DES PARAMÈTRES AVEC UNE MÉTHODE DE MINIMISATION DE DISTANCE**

Nos données sinistres étant groupées au sein d'une même classe, nous ne pouvons donc mesurer la distance qu'entre les bornes supérieures de chaque classe. Nous allons minimiser la distance suivante pondérée par le nombre de sinistres :

$$D = \sum_{i=1}^k n_i [F_n(c_i) - F(c_i)]^2$$

où  $n_i$  est l'effectif de la classe  $i$ ,  $c_i$  la borne supérieure de la classe  $i$ ,  $F_n$  la fonction de distribution empirique et  $F$  la fonction de distribution de la loi mélange (LN-LN).

L'intérêt de pondérer par  $n_i$  est de mieux estimer les paramètres de la loi pour les classes où se situe la plupart des sinistres.

Pour minimiser cette distance  $D$ , nous utilisons le solveur d'Excel. Par exemple, pour la classe 6, nous obtenons les résultats suivants :

Classe Finale	borne inf	borne sup	proba normalisée	Fn(ci)	F(ci)	Poids	Ecart
6	200 000	376 000	62%	62%	62%	1 596	0,0000
7	376 000	707 000	25%	86%	86%	633	0,0000
8	707 000	1 250 000	8%	95%	95%	209	0,0002
9	1 250 000	2 500 000	4%	98%	98%	92	0,0001
10	2 500 000	4 700 000	1%	99%	99%	30	0,0000
11	4 700 000	9 000 000	0%	100%	100%	12	0,0000
12	9 000 000	20 000 000	0%	100%	100%	7	0,0000
<b>Minimisation</b>							<b>0,0004</b>

Résolution Solveur =>

$\mu_1$	11,67
$\sigma_1$	1,29
$\lambda$	0,99
$\mu_2$	14,85
$\sigma_2$	1,65

Du fait du peu de classes (7) et parce que l'on cherche à estimer 5 valeurs ( $\mu_1$ ,  $\sigma_1$ ,  $\mu_2$ ,  $\sigma_2$  et  $\lambda$ ), le solveur ne

trouve pas les mêmes résultats selon les paramètres initiaux fournis. Il n’y a donc pas de convergence.

**Initialisation des paramètres avec la méthode Benabbou-Partrat**

Pour remédier à la divergence des estimations de paramètres, on veut initialiser le calcul du solveur par des valeurs proches du modèle P-B.

Les données étant groupées, il nous faut adapter P-B donc nous créons une liste de sinistres qui respecte la distribution empirique constatée, les montants étant forcés à la borne supérieure de la classe. Sur cet échantillon, nous obtenons alors les résultats suivants :

Résultats	Classe d'origine d'ouverture du sinistre												
	IBNR	1	2	3	4	5	6	7	8	9	10	11	12
$\mu_1$	12,63	12,96	12,63	12,52	12,52	12,51	12,51	échec					13,47
$\sigma_1$	0,69	0,95	0,69	0,63	0,63	0,62	0,62						1,11
$\lambda$	89%	98%	86%	90%	94%	96%	95%						14%
$\mu_2$	14,77	16,49	14,82	14,63	14,44	14,45	14,58						16,39
$\sigma_2$	0,91	0,41	0,92	0,79	0,69	0,72	0,80						0,66
écart-type	114%	103%	115%	101%	93%	95%	102%						129%

**Le paramètre qui subit le plus de transformation entre les classes 6 et 12 est  $\lambda$  : il varie de 95% à 14%.**

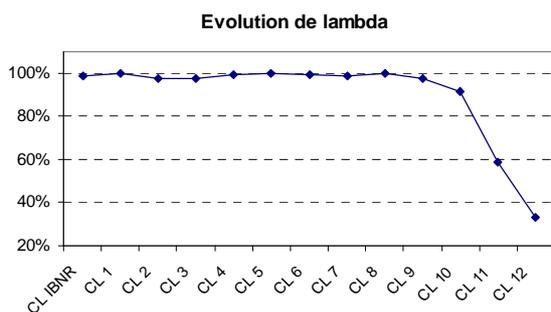
Le modèle définit une loi mélange pour toutes les classes sauf celles de 7 à 11 donc, globalement, pour les graves. Après audit, deux phénomènes causent le rejet : soit un  $\lambda$  initial  $>1$  (biais introduit par nos sinistres « en escalier »), soit un blocage lié à une normalisation trop importante du phénomène de troncature.

**Résultats**

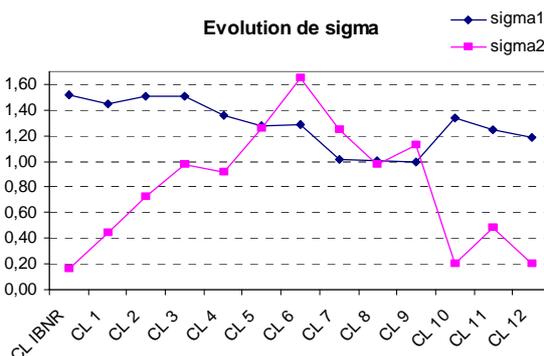
Avec l’initialisation des paramètres par ceux trouvés précédemment, nous obtenons les résultats suivants :

	CL IBNR	CL 1	CL 2	CL 3	CL 4	CL 5	CL 6	CL 7	CL 8	CL 9	CL 10	CL 11	CL 12
$\mu_1$	11,99	12,29	12,04	11,70	11,69	11,66	11,67	12,21	12,80	13,43	14,04	13,98	14,08
$\sigma_1$	1,52	1,45	1,51	1,51	1,35	1,28	1,29	1,02	1,01	1,00	1,34	1,25	1,19
$\lambda$	99%	100%	97%	98%	99%	100%	99%	99%	100%	97%	92%	59%	33%
$\mu_2$	15,81	16,49	15,47	14,59	14,67	14,98	14,85	14,72	15,20	15,26	14,94	15,59	16,19
$\sigma_2$	0,16	0,44	0,73	0,97	0,91	1,25	1,65	1,25	0,98	1,12	0,21	0,48	0,20

Nous représentons ci-dessous graphiquement les évolutions de  $\sigma$  et  $\lambda$  :



Nous constatons, comme précédemment, que  $\lambda$  est le paramètre qui subit le plus de transformation sur les sinistres ouverts dans les classes 8 à 12, il varie de 100% à 33%.



Nous n’obtenons toujours pas de tendance décroissante de  $\sigma_1$  ;  $\sigma_2$  quant à lui est instable, il augmente fortement jusqu’à la classe 6 et diminue rapidement après. La valeur  $\sigma_2$  de la classe 10 paraît assez faible et nous considérons que la méthode ne marche pas pour cette classe.

Devant la difficulté d'estimer les paramètres des lois, **nous décidons de figer ces derniers et de construire un programme d'optimisation pondéré pour estimer  $\lambda$** . Ce choix nous semble intéressant car en plus d'être le pondérateur des deux univers de sinistres graves ordinaires et non ordinaires, nous avons constaté que c'est le paramètre qui subit le plus de transformation.

Nous fixons alors les estimations des paramètres des lois par les valeurs « générales » obtenues dans la méthode 1 ( $\mu_1 = 12,06$ ,  $\sigma_1 = 1,65$ ,  $\mu_2 = 15,06$  et  $\sigma_2 = 0,68$ ) **hormis  $\lambda$  que l'on cherchera à estimer**.

L'objectif est d'affiner la méthode 1 en tenant compte de la gravité initiale du sinistre, c'est-à-dire de toute sa connaissance individuelle.

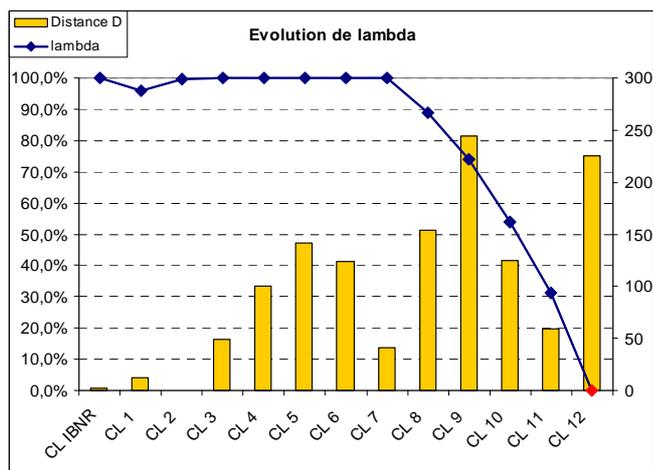
## VI - C . 2. ESTIMATION DU PARAMÈTRE $\lambda$

Dans la loi mélange,  $\lambda$  est le paramètre pondérant les deux typologies de sinistres (graves ordinaires et graves non ordinaires). Nous allons tester trois méthodes de minimisation de distance pour l'estimer.

### Minimisation de la distance D

Nous reprenons la même méthode qu'en VI-C.2. et cherchons à minimiser D à partir de  $\lambda$ .

Nous représentons les valeurs de  $\lambda$  et D par classe :



Pour la classe 12, nous obtenons une valeur  $\lambda = 0$  : la méthode de minimisation ne semble pas marcher.

Il est intéressant de noter que  $\lambda$  commence à diminuer à partir de la classe 8 qui par ailleurs correspond à notre seuil  $S_2$ , la frontière des graves non ordinaires.

Sur les dernières classes (CL8, CL9, CL10, CL12) où l'on souhaite obtenir une meilleure estimation, la distance est plus élevée, **ce qui nous invite alors à tester une autre méthode**.

### Minimisation de la distance du $\chi^2$

Si la fréquence de la  $i^{\text{ème}}$  classe est :  $f_i = n[F_n(c_i) - F_n(c_{i-1})]$ , alors la statistique d'ajustement du chi-deux<sup>52</sup> est :

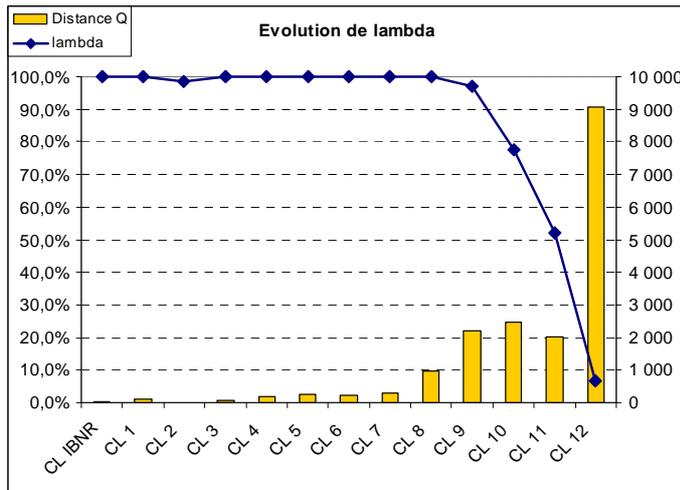
$$Q = \sum_{i=1}^k \frac{f_i - n[F(c_i) - F(c_{i-1})]^2}{n[F(c_i) - F(c_{i-1})]}$$

où  $n$  est le nombre de sinistres,  $k$  le nombre de classes,  $c_i$  la borne supérieure de la classe  $i$ ,  $F_n$  la fonction de distribution empirique et  $F$  la fonction de distribution du modèle (loi mélange LN-LN).

On sélectionne alors la valeur de  $\lambda$  qui minimise Q dans la fonction de répartition F.

<sup>52</sup> Chi-Square Tests, Vol. 19. Mathematical Association of America, Washington, D.C., Moore (1978).

Nous représentons les valeurs de  $\lambda$  et Q par classe :



La distance de la classe 12 est très élevée, ce qui suppose une mauvaise adéquation.

Nous constatons, comme précédemment, que la distance s'élève à partir de la classe 8 qui par ailleurs correspond à la borne du 2<sup>ème</sup> seuil  $S_2$  à 1,25M€.

Les distances étant trop élevées sur les classes sensibles des graves (idem remarque distance D), **cette méthode ne nous satisfait pas**. Nous allons utiliser une dernière méthode prenant mieux en compte la queue de distribution.

**Minimisation de la distance de Cramér-von Mises**

La forme de la statistique de Cramér-von Mises<sup>53</sup> est :

$$\frac{1}{n} \sum_{i=1}^n [F_n(x_i) - F(x_i; \theta)]^2$$

où  $x_1, x_2, \dots, x_n$  sont les observations,  $n_i$  est l'effectif de la classe  $i$ ,  $F_n$  la fonction de distribution empirique et  $F$  la fonction de distribution du modèle (loi mélange LN-LN).

Cependant, dans le cas de données groupées, nous ne connaissons pas les observations individuelles et faisons notre comparaison sur les bornes hautes des classes. Nous obtenons donc :

$$\sum_{i=1}^k [F_n(c_i) - F(c_i; \theta)]^2$$

La distance estimée de Cramér-von Mises est trouvée en minimisant cette expression sous contrainte de  $\theta$ . De plus, il est assez facile d'introduire une fonction de poids, si l'on souhaite mettre l'accent sur la queue de distribution. Nous pouvons donc minimiser la distance :

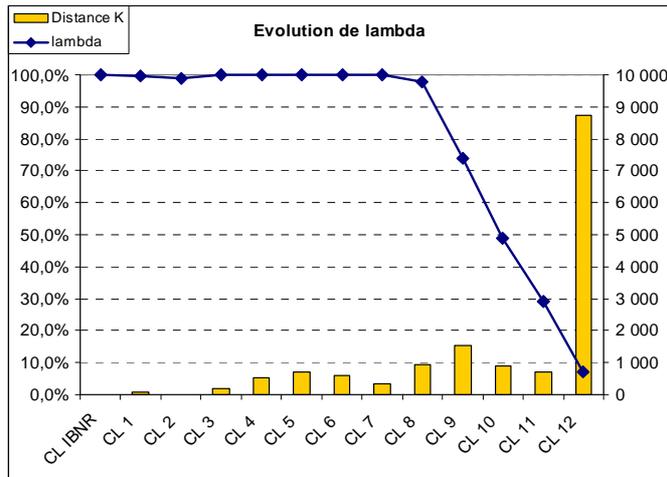
$$K = \sum_{i=1}^k w(c_i) [F(c_i) - F_n(c_i)]^2 \text{ où } w(c_i) \text{ est une fonction arbitraire de pondération.}$$

Un choix naturel de fonction de pondération est :

$$w(x) = n / (F(x) \cdot [1 - F(x)])$$

<sup>53</sup> Loss distributions, Robert V. Hogg et Stuart A. Klugman, 1984

Nous représentons les valeurs de  $\lambda$  et K par classe :



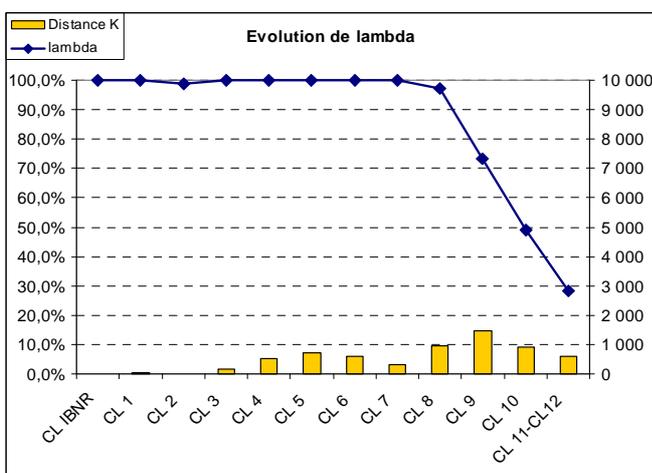
Nous observons que les écarts de distance sont davantage lissés entre toutes les classes.

Seule la classe 12 déroge à la règle du fait d'un faible nombre de sinistres dans cette classe (4 en maturité 1) et donc d'un manque de robustesse de la fonction de répartition empirique.

La méthode de minimisation de la distance de Cramér-von Mises a l'intérêt par rapport aux méthodes précédentes d'accorder plus de poids à la queue de distribution et ainsi de lisser les écarts de distance entre les classes.

Pour résoudre l'écart constaté sur la classe 12, nous l'agrégeons à la classe 11 pour avoir plus de volume. Nous refaisons donc les étapes de calcul de la matrice  $1 \sim \infty$  puis de minimisation de la distance de Cramér-von Mises.

Nous obtenons alors les valeurs suivantes de  $\lambda$  et K par classe :



On constate avec satisfaction que la distance sur la classe « 11-12 » est maintenant du même ordre de grandeur que pour les autres classes.

La valeur de  $\lambda$  (28%) est maintenant un peu inférieure à celle que l'on avait sur la classe 11 (29%). Ce résultat s'explique car en agrégeant les deux classes, on a plus de sinistres graves non ordinaires, donnant ainsi moins de poids à la loi 1.

Nous observons en classe 3 un lambda à 99% que nous corrigerons par cohérence à 100%. Les valeurs retenues sont alors les suivantes :

	CL IBNR	CL 1	CL 2	CL 3	CL 4	CL 5	CL 6	CL 7	CL 8	CL 9	CL 10	CL 11-CL12
$\mu_1$	12,06											
$\sigma_1$	1,65											
$\lambda$	100%	100%	100%	100%	100%	100%	100%	100%	97,2%	73,1%	48,8%	28,2%
$\mu_2$	15,06											
$\sigma_2$	0,68											

La distance K avec regroupement des classes 11 et 12 est celle qui permet le lissage le plus conforme à nos attentes. A partir de la classe 6, nous observons des écarts homogènes sur toutes les classes jusqu'aux plus élevées. La queue de distribution est ainsi prise en compte sans excès. **La distance de Cramér - von Mises et le regroupement des deux dernières classes aboutissent à un résultat satisfaisant.**

### VI - C . 3. SIMULATION DE LA CHARGE BRUTE ET NETTE DE REASSURANCE

#### Simulation de la charge brute

Les étapes conduisant à l'obtention de la charge brute sont très proches de celles vues en V-D-3, avec en plus un modèle détaillé pour chaque classe de sinistres graves. Nous simulons selon la méthode Monte-Carlo.

La matrice de Markov de l'état 1 an à l'état ultime permet de calculer la charge des graves. Pour la partie des sinistres attritionnels nous utilisons les résultats de la simulation via la méthode 1.

Pour calculer la k<sup>ème</sup> simulation d'une charge brute, on procède en 5 étapes présentées ci-dessous :

On pose :

- $\lambda$  : le paramètre de la loi de Poisson,
  - $\lambda_{attri}$  : la proportion de sinistres attritionnels hors IRCA,
  - $\lambda_{irca}$  : la proportion de sinistres IRCA,
  - $\lambda_{graves}$  : la proportion de sinistres graves supérieurs à S<sub>1</sub>,
  - $\lambda_{graves\_Clj}$  : la proportion de sinistres de classe j dans les sinistres graves supérieurs à S<sub>1</sub>.
- } Idem méthode 1
- } Nouveauté méthode 2

Etape 1 : on fait un tirage aléatoire du nombre de sinistres selon la loi retenue pour les nombres  $P(\lambda)$ . On note  $n_k$  le nombre tiré.

Etape 2 :

- Dans  $n_k$  on tire un nombre de sinistres graves  $n_k^{graves}$  selon la loi  $B(n_k, \lambda_{graves})$
- Dans  $n_k - n_k^{graves}$  on tire un nombre de sinistres attritionnels  $n_{k\_attri}$  selon la loi  $B(n_k - n_k^{graves}, \frac{\lambda_{attri}}{\lambda_{IRCA} + \lambda_{attri}})$
- On obtient  $n_k - n_k^{graves} - n_{k\_attri}$  sinistres IRCA noté  $n_k^{IRCA}$
- On répartit les  $n_k^{graves}$  sinistres graves dans les classes initiales CL<sub>IBNR</sub>, CL<sub>1</sub> à CL<sub>11\_12</sub> en utilisant récursivement une loi binomiale de la même manière que précédemment. Nous obtenons  $n_k^{graves\_Clj}$  nombre de sinistres graves dans la j<sup>ème</sup> classe.

Etape 3 : Charge brute des attritionnels (issue de la méthode 1)

On effectue  $n_k^{attri}$  tirages aléatoires de montants de sinistres dans la loi log-normale LN (8,14 ; 1,6) déjà utilisée en V-E-3 et que l'on notera  $\mathcal{X}_1^{n_k^{attri}}, \mathcal{X}_2^{n_k^{attri}}, \dots, \mathcal{X}_{n_k^{attri}}^{n_k^{attri}}$ .

La charge totale des sinistres attritionnels de la k<sup>ème</sup> simulation notée  $C_k$  est :

$$C_k = \sum_{i=1}^{n_k^{attri}} \mathcal{X}_i^{n_k^{attri}} + n_k^{IRCA} * IRCA_{100\%} \text{ où } IRCA_{100\%} \text{ désigne le montant du forfait IRCA à 100\%, soit 1490 € en 2010.}$$

Etape 4 : Charge brute des graves

Pour chaque classe j où  $j \in \{ibnr, 1, \dots, 11\}$ , on effectue  $n_k^{graves\_Clj}$  tirages aléatoires de montants de sinistres

dans la loi mélange calculée en VI-C.3 et que l'on notera  $\mathcal{X}_1^{n_k^{graves\_Clj}}, \mathcal{X}_2^{n_k^{graves\_Clj}}, \dots, \mathcal{X}_{n_k^{graves\_Clj}}^{n_k^{graves\_Clj}}$ .

La charge totale des sinistres graves de la  $k^{\text{ème}}$  simulation notée  $C'_k$  est :  $C'_k = \sum_{j \in \{ibnr, 1, \dots, 11\}} \sum_{i=1}^{n_k^{\text{graves\_Cl}_j}} x_i^{n_k^{\text{graves\_Cl}_j}}$

**Etape 5 : Espérance et variance**

Un estimateur de l'espérance de la charge nette de réassurance des n simulations est :  $\hat{m} = \frac{1}{n} \sum_{i=1}^n C'_i$

Un estimateur de la variance empirique de la charge nette de réassurance des n simulations est :

$$\hat{s}^2 = \frac{1}{n-1} \times \sum_{i=1}^n (C'_i - \frac{1}{n} \sum_{i=1}^n C'_i)^2.$$

**Simulation de la charge nette de réassurance des sinistres graves**

Les étapes conduisant à l'obtention de la charge nette sont celles détaillées en V-D.3 : vérification de l'atteinte de l'AAD par la charge cédée en XS entre la priorité  $Pr_1$  et  $Pr_2$ , plafonnement de la charge cédée par la limite Lim et cession systématique de la charge en XS de  $Pr_2$ , et ce, pour chaque simulation effectuée.

**Résultats**

Nous utilisons l'outil Igloo et faisons  $n = 10\,000$  simulations.

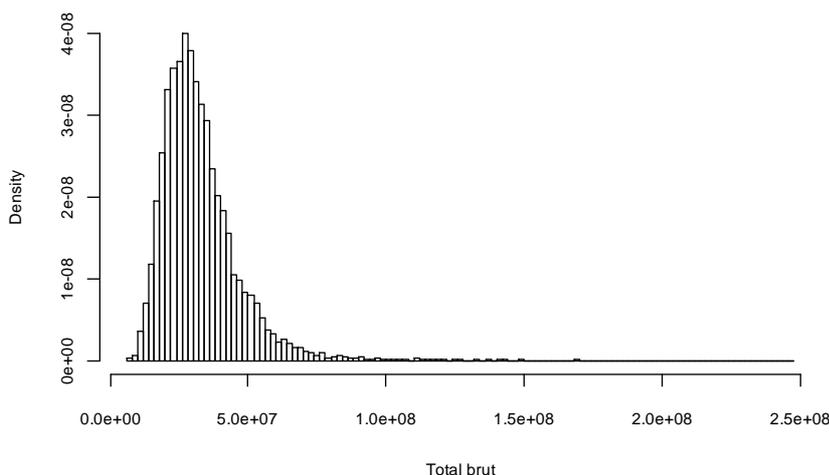
La méthode de simulation de la charge brute des sinistres attritionnels étant identique à la méthode 1, nous conservons ici ces résultats.

1. Charge brute de réassurance des sinistres graves :

Nous indiquons la valeur des proportions de sinistres par classe initiale :

$\lambda_{\text{graves\_CLibr}}$	$\lambda_{\text{graves\_CL1}}$	$\lambda_{\text{graves\_CL2}}$	$\lambda_{\text{graves\_CL3}}$	$\lambda_{\text{graves\_CL4}}$	$\lambda_{\text{graves\_CL5}}$	$\lambda_{\text{graves\_CL6}}$	$\lambda_{\text{graves\_CL7}}$	$\lambda_{\text{graves\_CL8}}$	$\lambda_{\text{graves\_CL9}}$	$\lambda_{\text{graves\_CL10}}$	$\lambda_{\text{graves\_CL11\_12}}$
9,76%	9,70%	19,80%	2,67%	6,72%	9,40%	16,39%	12,35%	6,00%	2,82%	1,83%	2,56%

**Histogramme des 10 000 simulations de la charge brute des graves**



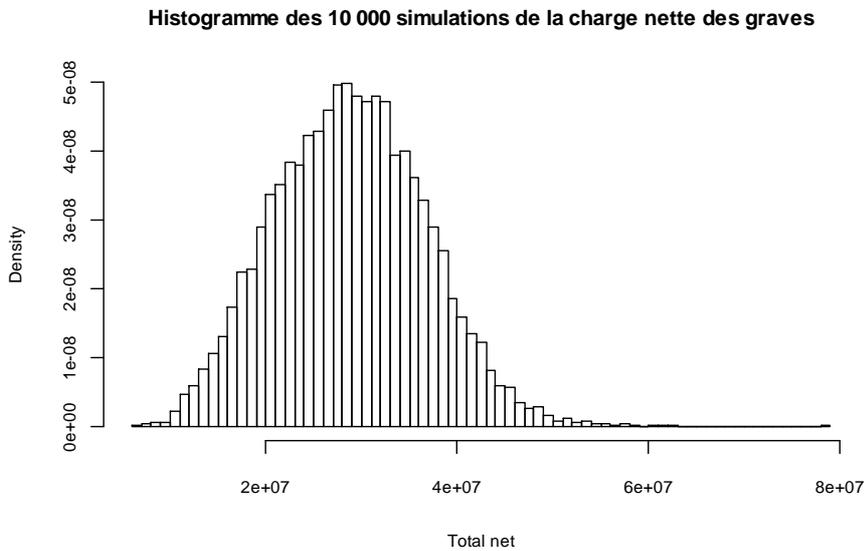
Total graves brut de réassurance	
Espérance	32 733 657 €
Ecart-type	14 778 059 €
Q25,0%	23 365 589 €
Q50,0%	29 921 018 €
Q75,0%	38 458 914 €
Q95,0%	57 294 460 €
Q97,5%	67 188 808 €
Q99,0%	85 023 986 €
Q99,5%	105 779 734 €

On note une grande volatilité conduisant à des probabilités importantes d'avoir une charge conséquente. Par exemple, tous les 40 ans ( $Q_{97,5\%}$ ), la charge peut être le double de l'espérance.

Pour ces sinistres graves, il est donc légitime d'avoir recours à la réassurance XS pour diminuer le risque de dérive de la sinistralité.

## 2. Charge nette de réassurance des sinistres graves :

On représente ci-dessous les distributions des 10 000 simulations de charges nettes de réassurance :

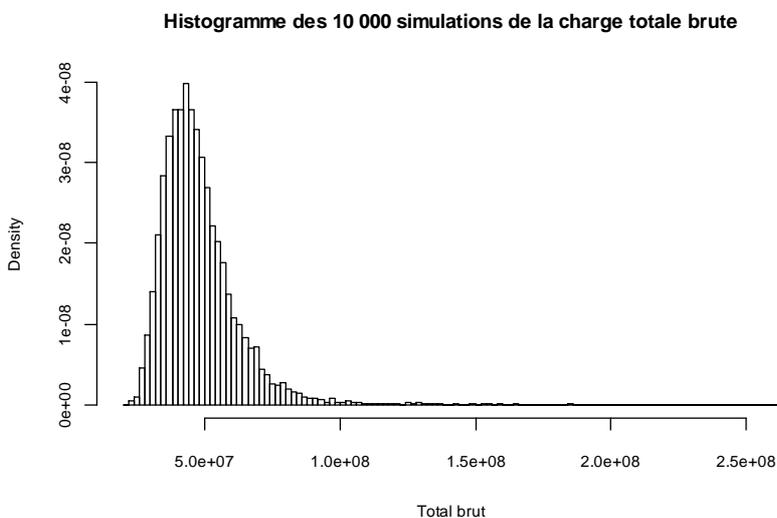


Total graves net de réassurance	
Espérance	28 835 682 €
Ecart-type	7 863 190 €
Q25,0%	23 124 915 €
Q50,0%	28 764 610 €
Q75,0%	34 244 068 €
Q95,0%	41 785 941 €
Q97,5%	44 397 509 €
Q99,0%	47 835 222 €
Q99,5%	50 552 820 €

La réassurance joue son rôle et diminue la volatilité de moitié.

## 3. Charge totale brute de réassurance

On représente ci-dessous les distributions des 10 000 simulations de charges brutes de réassurance :

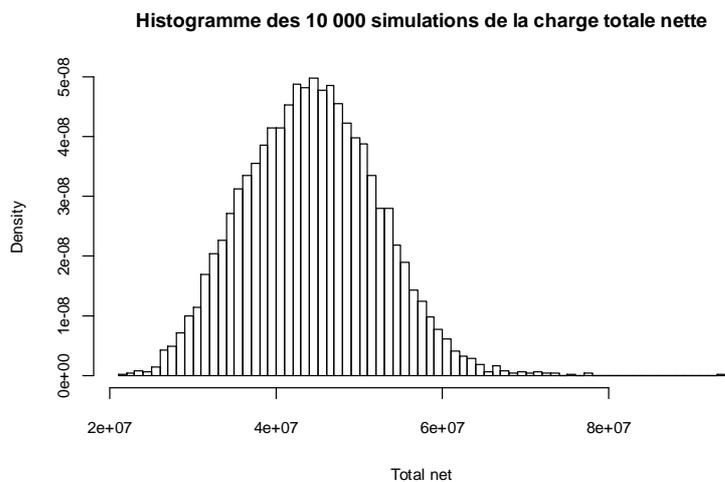


Total brut de réassurance	
Espérance	48 166 313 €
Ecart-type	14 803 206 €
Q25,0%	38 788 044 €
Q50,0%	45 364 608 €
Q75,0%	54 044 313 €
Q95,0%	72 577 306 €
Q97,5%	82 665 546 €
Q99,0%	99 798 290 €
Q99,5%	121 320 645 €

On remarque le même phénomène que pour la charge brute des graves : un écart-type important et une queue de distribution épaisse. Cela induit un besoin de réassurance pour réduire ce risque de volatilité.

#### 4. Charge totale nette de réassurance

On représente ci-dessous les distributions des 10 000 simulations de charges nettes de réassurance :



Total net de réassurance	
Espérance	44 268 338 €
Ecart-type	7 911 355 €
Q25,0%	38 601 898 €
Q50,0%	44 178 088 €
Q75,0%	49 686 582 €
Q95,0%	57 326 678 €
Q97,5%	59 874 190 €
Q99,0%	63 320 409 €
Q99,5%	66 422 914 €

La charge totale nette de réassurance est visiblement bien concentrée ; l'écart-type est presque réduit de moitié entre la vision brute et la nette.

## VI - D CONCLUSION

---

De nombreuses alertes nous ont été données quant à l'utilisation de cette méthode avec des nombres insuffisants de sinistres. C'est pourquoi, nous avons significativement enrichi notre base par les sinistres auto, homogènes aux nôtres.

L'article de Mahon nous a servi de support pour cette approche intéressante qui utilise l'évolution de la connaissance individuelle du sinistre (expertise gestionnaire) et donc de manière indirecte l'information qualitative pour l'estimation à l'ultime. Nous y avons apporté des améliorations en ajoutant une classe de sinistres IBNR nous permettant de gérer cette particularité directement dans le modèle et en proposant une méthode de simulation de la charge. Cependant, contraints par nos données, nous nous sommes éloignés de sa démarche.

Pour mieux appréhender la distribution des classes élevées et ainsi diminuer la volatilité, nous expérimentons alors une nouvelle approche cohérente avec la méthode 1, en ne cherchant non plus une loi unique, mais une loi mélange par classe. Pour ce faire, nous avons utilisé des outils complémentaires d'estimation des paramètres (Benabbou-Partrat et minimisation de distance). Toutefois nous avons abouti à un échec lié à la difficulté d'estimer les paramètres des lois et à la variance ne diminuant pas dans les classes élevées.

Nous avons poursuivi notre réflexion avec  $\lambda$  ; pondérateur des deux univers de sinistres graves ordinaires et non ordinaires, c'est le paramètre qui subit le plus de déformation. Aussi, nous avons figé les paramètres de la loi mélange hormis  $\lambda$ . L'objectif est alors d'obtenir un  $\lambda$  calculable, décroissant, conditionné par la classe initiale et qui définit la loi mélange qui s'adapte le mieux aux données. Deux méthodes de minimisation de distance ont été expérimentées (distance quadratique pondérée et  $\chi^2$ ). Cependant nous obtenions des distances trop élevées sur les classes sensibles des graves là même où l'on souhaitait obtenir une meilleure adéquation. Nous avons alors utilisé la distance de Cramér-von Mises donnant plus de poids à la queue de distribution. Les écarts de distance entre les classes ont ainsi été lissés, hormis sur la classe 12 agrégée à la 11 pour avoir plus de volume.

En conclusion, nous avons sélectionné l'apport des classes issu de la méthode utilisée par Mahon, que l'on a enrichi par l'apport de la loi mélange de la méthode 1. Notre objectif de simuler la charge brute et nette de réassurance a donc pu être atteint. De plus, ce modèle devrait être plus performant car les sinistres graves sont modélisés par 12 lois mélanges correspondant à nos classes et non par une seule loi mélange.

A la relecture de notre démarche, nous listons ici les pistes d'amélioration qui nous semblent pertinentes :

- une modélisation affinée et/ou segmentée (attritionnels, graves) de l'inflation,
- une prise en compte d'autres critères médicaux permettant de connaître l'état de la victime,
- une constitution de matrices semestrielles, trimestrielles ou mensuelles pour une utilisation opérationnelle (prévisionnels, clôture),
- un modèle simplificateur permettant de déduire une valeur de  $\lambda$  pour chaque classe.

Pour une application ordinaire de Mahon sur un jeu de données permettant la déclinaison de la méthode, nous conseillons d'étudier l'influence sur la charge de l'utilisation du milieu de la classe versus la médiane et de renforcer les résultats sur le choix d'une famille de loi par des tests d'adéquation.

## VII - CONCLUSION

Les théories ayant été exposées et les applications ayant délivré leurs résultats, nous cherchons à poser une analyse objective afin d'éclairer notre choix pour l'un des deux modèles.

### VII - A . 1. COMPARAISON

Nous regroupons dans le tableau ci-dessous les principaux axes d'analyse retenus et nos remarques. Pour chaque item, nous prenons position pour retenir la méthode la plus intéressante.

Critère	Méthode 1	Méthode 2	Notre choix
Constitution de la base de données	Nécessite des risques homogènes	Nécessite des risques homogènes et un grand nombre de sinistres pour renforcer la robustesse des distributions par classe.	Méthode 1
Mise en oeuvre	Méthode éprouvée.  Les sinistres attritionnels sont modélisés par cette méthode.  Tests EDF validés.	Méthode peu diffusée et nécessitant une adaptation selon les données. Implémentation longue. Un très grand nombre de paramètres à estimer. Risque d'erreur.  L'ajout des attritionnels dans la modélisation augmenterait cette complexité.  Pas de tests chez Mahon.	Méthode 1
Réalisation des simulations			Idem avec notre outil
Précision	Méthode collective.  On retient une unique loi mélange pour la distribution des graves.	Méthode individuelle.  Chaque classe de graves est modélisée par une loi mélange différente, adaptée à sa distribution.	Méthode 2
Utilisation opérationnelle	Pas d'utilisation des informations sur l'année en cours.	Possibilité de créer des matrices de $x$ mois à $+\infty$ .	Méthode 2
Evolutivité		La méthode doit pouvoir être affinée en retenant : - un nombre différent de classes, - des bornes de classes déterminées autrement, - un conditionnement sachant la trajectoire de l'évaluation (en hausse / baisse, ie sa mémoire) ou selon un critère qualitatif.	Méthode 2

On étudie ensuite les charges des graves obtenues :

Charge des graves	Méthode 1			Méthode 2			Ecart 2 vs 1	
	Espérance	Variance	Ecart-type / espérance	Espérance	Variance	Ecart-type / espérance	Espérance	Variance
Brute	30 963 811	2,21E+14	48%	32 733 657	2,18E+14	45%	6%	-1%
Nette	27 708 499	5,46E+13	27%	28 835 682	6,18E+13	27%	4%	13%
Cédée	3 255 312	1,19E+14	335%	3 897 975	1,03E+14	261%	20%	-13%

On se donne comme indicateur relatif (proportionnel à la largeur relative de l'intervalle de confiance) celui donné par le rapport écart-type/espérance.

### Charge brute

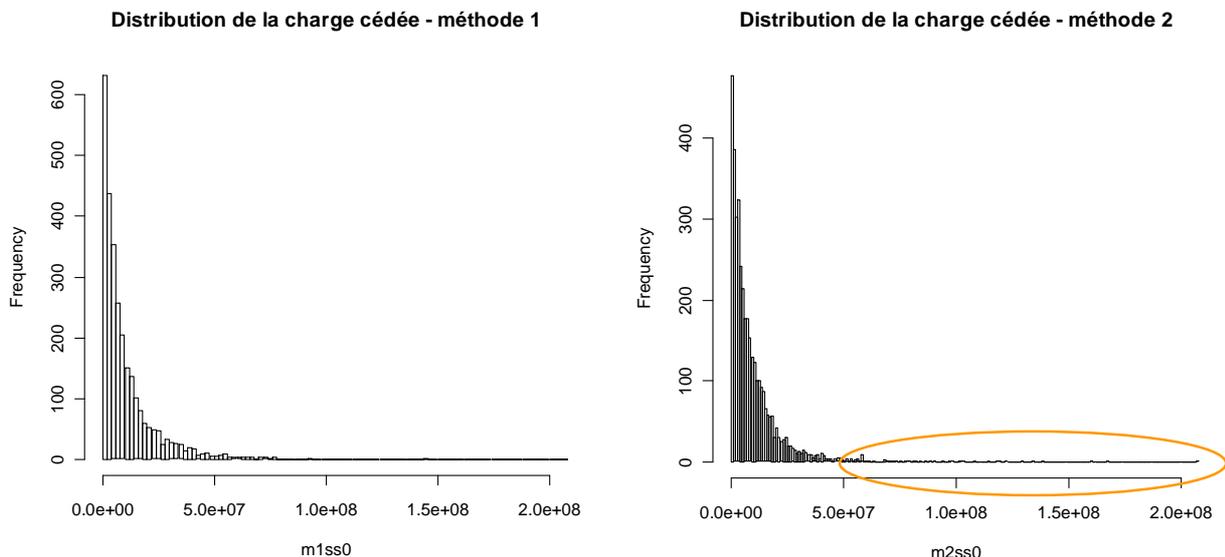
Nous observons que **l'espérance de la charge brute 2 est supérieure à celle de 1 (+6%)** ; la prise en compte de la queue de distribution étant plus précise, cela majore la charge. La méthode 2 est plus prudente. Les écarts de variance des charges brutes sont très faibles ; on les négligera. L'indicateur relatif diminue légèrement avec 2.

### Charge nette

L'espérance de la charge nette 2 reste supérieure à celle de 1 mais dans des proportions moindres (+4%). On peut l'expliquer par une proportion plus grande de graves cédée en 2. En revanche, **la variance 2 est bien supérieure à celle de 1**. On rappelle qu'il y a cession quand le sinistre est supérieur à 2 500 k€ (et que l'AAD est atteint), soit pour de rares et importants sinistres. La volatilité créée par ces forts écarts d'intensité de coûts est accentuée par la méthode qui génère davantage de très graves comme c'est le cas de la méthode 2. On le constate aussi en regardant l'écart entre les 2 espérances de charges cédées qui est de +20 % pour 2 vs 1 et en observant que les probabilités de céder au réassureur sont plus élevées en 2 (38,25%) qu'en 1 (28,64%).

### Charge cédée

En lien avec l'analyse de la charge nette, les schémas ci-dessous permettent d'illustrer la distribution plus étendue de la charge cédée dans la méthode 2 vs 1 à partir de 5 000 k€.



Toutefois, notre indicateur relatif s'améliore avec la méthode 2 passant de 335% à 261% traduisant un bénéfice de risque relatif pour le réassureur.

Suite aux comparaisons (qualitative et quantitative) précédentes, **nous entérinons notre préférence pour la méthode 2**. Nous expliquons l'augmentation de variance de sa charge nette par les 2 phénomènes suivants :

- une importance plus grande donnée à la loi des graves non ordinaires dont la variance est supérieure,
- un traité de réassurance moins adapté pour un modèle accordant plus de poids à une queue épaisse.

## VII - A . 2. CONCLUSION

Le sujet abordé, l'évaluation de la charge ultime des sinistres de responsabilité civile corporelle, nous était familier au travers de quelques sinistres majeurs défrayant la chronique et par les particularités de sa liquidation dans le temps.

Le premier chapitre consacré au contexte a permis de renforcer nos connaissances et de mieux appréhender :

- les impacts de ce risque pour la compagnie,
- l'évolution de l'environnement juridique de cette prestation,
- la dichotomie du risque routier s'atténuant en nombre mais pas en intensité,
- la gestion des sinistres et sa nécessaire expertise sur l'ensemble de la chaîne des acteurs.

Les deux outils préalables que nous avons identifiés ensuite pour atteindre notre objectif, à savoir le calcul de seuil et l'analyse de base victimes, nous ont demandé le maniement d'un large éventail de techniques statistiques telles que l'analyse de données (ACP), la régression multiple pas-à-pas, le test de conformité, des estimateurs particuliers (Hill, Pickands, Dekkers, GPD), la pratique de R (minAMSE) et même une méthode issue de la météorologie (Gertensgarbe).

L'analyse de base victimes nous a permis de démontrer que les profils de victimes auto et flottes, pour une catégorie de véhicules donnée, ne sont pas différents, donnant ainsi accès à l'utilisation pour le risque flottes, plus confidentiel en volume, à de nouvelles méthodes, hier réservées à la masse. Cela induira un pilotage plus fin permettant à la branche flottes d'être plus rentable et d'éviter les accidents industriels qui surviennent de façon récurrente dans cette branche.

De plus, une base victimes riche permettra de trouver de nouveaux critères plus explicatifs permettant d'élaborer des modèles plus robustes et pouvant peut-être à terme faire évoluer la méthode 2 ; à titre d'illustration, on a montré que la base interne « surperforme » la base marché, montrant ainsi le bénéfice de la qualification. On pourra envisager de bâtir avec les réassureurs des traités ayant comme priorité une valeur de la variable la plus représentative (ex : un traité XS basé sur le nombre de jours d'ITT :  $\infty$  XS 500 jours ou sur le pourcentage d'AIPP : 100% XS 50%).

Définir un seuil reste un exercice délicat. Nous avons multiplié les approches disponibles et retenu des valeurs de seuils convergentes. Cette étude met en évidence 3 familles de sinistres (attritionnel, grave ordinaire et grave non ordinaire) à partir des 2 seuils définis. Ces 3 familles pèsent en charge sinistres un poids significatif (chacune > 25% de la charge totale), elles segmentent donc bien la charge.

De plus, nous proposons d'indexer chaque année les seuils selon le type de sinistres : attritionnels (par l'indice des prix) et graves (par l'indice FFSA).

Pour cette étude sur les seuils et pour les 2 suivantes, un travail particulièrement fastidieux aura été mené afin de constituer la base de données, action rendue plus complexe du fait du contexte « patchwork / mille feuilles » de Covéa Fleet.

La modélisation d'une charge brute de sinistres par un système paramétrique (nombre, montant) est un exercice fréquent pour l'actuaire. Dans le cadre de la méthode 1, nous avons choisi d'ajouter un aléa supplémentaire sur les nombres (binomiale) pour tenir compte des 3 familles de sinistres. De même, l'apport de la loi mélange P-B appliquée au périmètre des graves permet une meilleure prise en compte de la queue de distribution qu'une loi unique. Enfin, la simulation de la charge nette de réassurance, rendue possible par la méthode Monte-Carlo, intègre la modélisation d'une protection XS complexe.

Pour la méthode 2, nous avons vite posé le constat d'échec de la méthode Mahon sur nos données et recherché une solution en lien avec la méthode 1 dont nous venions de mesurer l'apport. Nous avons alors complété la méthode et nos matrices par la création d'une classe d'IBNR. Ensuite, par optimisation du poids  $\lambda$  entre les deux lois du mélange P-B de la méthode 1, nous avons défini pour chaque classe la mélange la plus adaptée au sens de la distance de Cramér-von Mises. Les simulations des charges en brut et en net de réassurance ont de nouveau demandé une programmation importante dans l'outil.

Les méthodes propres utilisées pour ces deux modèles nous ont permis de bâtir une grille d'analyse afin de distinguer les forces et faiblesses de chacun. Etayée par de nombreux arguments, notre préférence va au modèle 2. Elle permet en effet de capter une information qualitative, à dire d'expert et au bout d'un an que ne fait pas la méthode 1.

Pour une utilisation opérationnelle en tarification et en pilotage des résultats techniques, il s'agira de rendre dynamique ce modèle ; par exemple, en le rendant capable de projeter les charges annuelles brutes et nettes à partir du constat de l'année en cours, à chaque quadrimestre pour les prévisionnels de comptes.

La méthode 2 prenant davantage en compte la queue de distribution, on observe :

- ❑ pour la cédante, une augmentation de la charge nette et plus globalement du risque. A fonds propres égaux, **la cédante aura donc intérêt à optimiser son traité de réassurance.**
- ❑ pour le réassureur, sa prime pure étant supérieure, il majorera son tarif. De plus, la charge cédée est moins risquée relativement au chiffre d'affaires et lui permet d'améliorer son ratio « fonds propres nécessaires / chiffre d'affaires ».

Enfin dans le contexte de Solvabilité II, nous disposons désormais d'un socle plus robuste (modèle individuel permettant d'estimer la charge brute et nette de réassurance) pour décliner le calcul du best estimate.

## VIII - ANNEXES

### VIII - A ANNEXE - ETUDE COMPARATIVE DU PROJET LEFRAND ET DU LIVRE BLANC DE L'AFA

Projet de loi Lefrand	Demandes exprimées dans le livre blanc
Création d'une base de données de tous les accidents corporels pour mettre à disposition du public les transactions et les décisions judiciaires (transparence).	
Des missions type d'expertise médicale et un barème médical unique d'évaluation des atteintes à l'AIPP. 2 barèmes sont actuellement utilisés selon qu'on se place en droit commun (assureurs) ou en accident du travail (sécurité sociale).	Oui pour l'unicité de ces 2 éléments : la mission d'expertise et le barème médical.
Une nomenclature des postes de préjudices suite à dommage corporel non limitative.	Oui pour plus de transparence. Pas de position sur la notion de « non limitative ».
Un barème de capitalisation officiel concernant les préjudices futurs et applicable, en droit commun mais aussi pour le calcul des prestations futures des tiers payeurs.	La clarification des règles de capitalisation des indemnités : un barème de capitalisation commun.
Une obligation renforcée du formalisme de l'information due à la victime.	
Passage d'un mode d'expertise unilatéral (le conseil de la victime est « muet ») à celui du contradictoire (le conseil de la victime donne un avis) : incitation à renforcer la présence d'un médecin conseil aux côtés de la victime.	
Indépendance du médecin conseil	
Une assistance financière plus rapide : des avances versées sous 1 mois sitôt les constatations médicales faites sur la nécessité de l'aménagement du logement, du véhicule ou de la présence d'une tierce personne.	
Un délai allongé de rétractation de la transaction par la victime (de 15 à 30 jours).	
	Les assureurs demandent des référentiels indemnitaires pour les préjudices non économiques (souffrances endurées, préjudice esthétique,...).
	La clarification du calcul des postes de préjudices économiques : des méthodes de calcul officielles.
	Le recours des organismes sociaux : une refonte nécessaire ; le protocole actuel en vigueur avec les assureurs date de 2003 et n'intègre pas la nomenclature Dintilhac.
	La nécessaire participation de la victime à l'amélioration de sa situation.
	L'accompagnement de la victime dans l'élaboration d'un projet de vie, une démarche nécessaire qui s'inscrit dans le long terme qui inclus : <ul style="list-style-type: none"> <li>- une mise en oeuvre concertée de la réinsertion socio-professionnelle,</li> <li>- un développement des structures d'accueil adaptées aux besoins des victimes.</li> </ul>
	Le suivi des besoins en aides humaines de la victime incluant : <ul style="list-style-type: none"> <li>- Le versement de rentes pour répondre à la réalité des besoins,</li> <li>- Le traitement fiscal des rentes, une réforme jugée indispensable.</li> </ul>

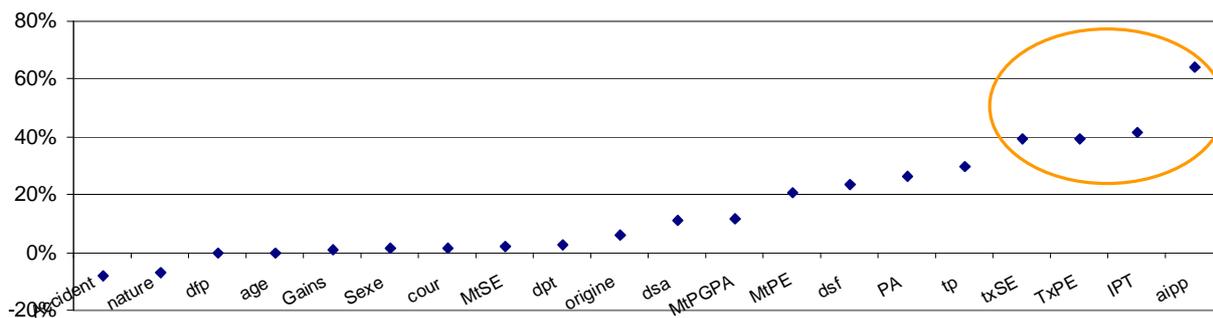
## VIII - B ANNEXE - ANALYSE DU FICHIER DES VICTIMES INDEMNISEES (FVI)

Nous appliquons la méthode précédente en retenant les variables décrivant une mesure (taux, nombre de jours) ou l'indication que le préjudice est indemnisé, pas le montant.

Les commentaires des notes de bas de page [34] et [35] qui formalisent les précautions à prendre (contribution des variables, des individus en ACP et sur les p-valeurs et résidus en régression) s'appliquent également à cette étude.

### Matrice de corrélation du FVI

Taux de corrélation avec le montant du sinistre par variable



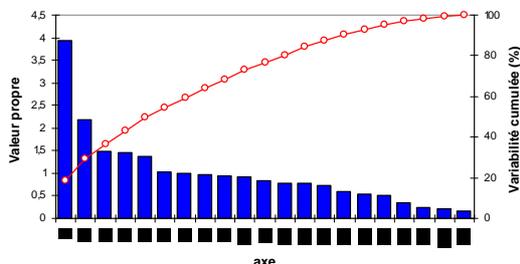
La matrice de corrélation est obtenue par Xlstat ; on vérifie que son rang est bien égal au nombre de variables donc il n'y a pas de détection de colinéarité entre les variables. Le rang est calculé via l'add-in Excel Matrix et la fonction MRANK.

On observe une vraie dichotomie entre les 16 variables pas ou peu corrélées avec le coût et les 4 autres. 1 variable a un taux de corrélation avec le coût égal à 64% : le taux d'AIPP. Les 3 suivantes sont IPT, le nombre de jours d'incapacité provisoire de travail, TxPE le taux de préjudice esthétique et txSE le taux de souffrances endurées.

### Analyse en Composantes Principales du FVI

Via le critère de Kaiser, on retient les 7 premières valeurs propres qui expliquent 59,3% de l'inertie du nuage.

	F1	F2	F3	F4	F5	F6	F7	F8
Valeur propre	3,95	2,17	1,49	1,44	1,36	1,03	1,00	0,98
Variabilité (%)	18,8	10,4	7,1	6,9	6,5	4,9	4,8	4,7
% cumulé	18,8	29,2	36,2	43,1	49,6	54,5	59,3	63,9



Le critère « du coude » sélectionne les 5 premières valeurs propres et 49,6% de l'inertie cumulée expliquée. C'est trop peu donc on retient le critère de Kaiser.

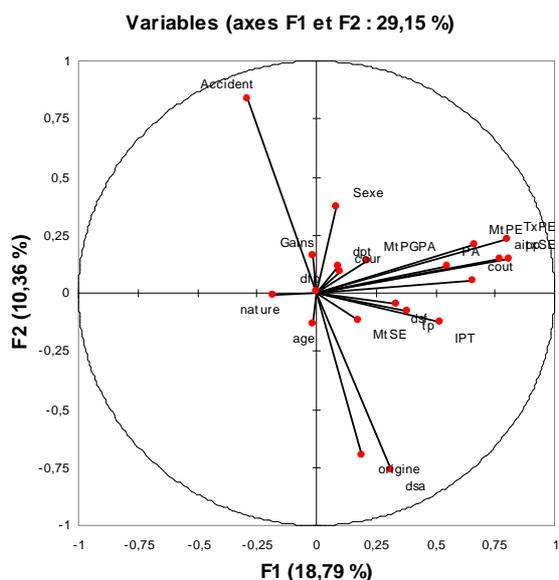
La contrepartie de ce choix est le nombre important de facteurs à appréhender.

Ces 7 facteurs indépendants, notés  $F_i$ , sont des combinaisons linéaires des variables et expliquent plus particulièrement certaines d'entre elles.

$F_1$  et  $F_2$  n'expliquent que 29,2% de l'inertie du nuage de points et les variables projetées sur ( $F_1$ ,  $F_2$ ) sont représentées ainsi :

L'analyse des corrélations « variables – facteurs » donne :

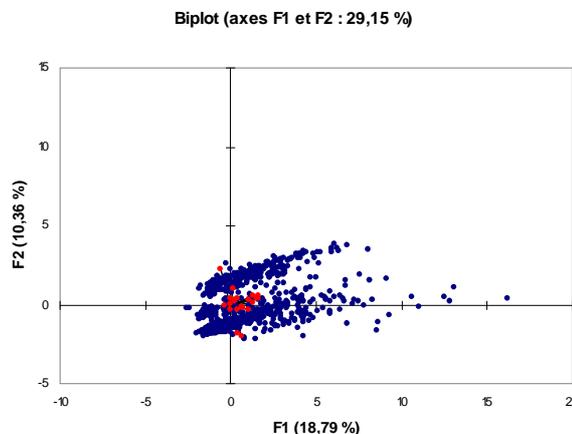
- $F_1$  retient les variables médicales à enjeu financier : les taux de souffrances endurées, de préjudice esthétique et d'AIPP, coût du sinistre et indication si préjudice d'agrément.
- $F_2$  privilégie la variable qualifiant l'accident de travail ou non et en opposition, l'indication de dépenses de santé actuelles et la variable sur la forme des données (Dintilhac ou non).
- $F_3$  met en relief les variables lieu de la cour d'appel et département de l'assuré.
- $F_4$  est surtout lié à l'indication de PGPA (perte de gains professionnels actuels).
- $F_5$  concerne l'âge et la présence de TP.
- $F_6$  prend en compte s'il y a transaction ou non et s'il y a des souffrances endurées.
- Enfin,  $F_7$  est corrélé à la présence de DFP (déficit fonctionnel permanent).



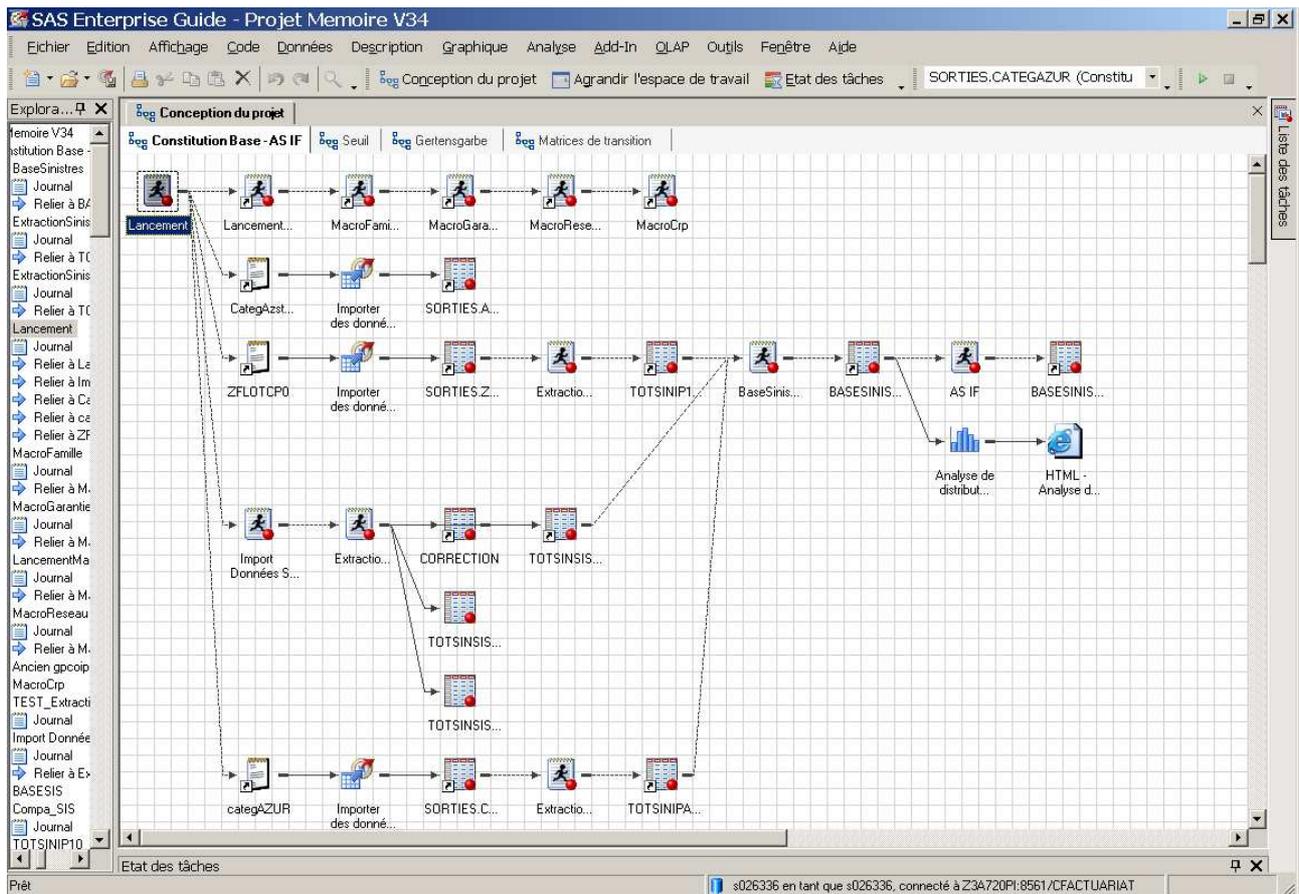
On distingue les variables bien représentées par leur proximité avec le bord du cercle. Les autres auront un facteur « pour elles » :  $F_3$ ,  $F_4$ ,  $F_5$ ,  $F_6$ ,  $F_7$ .

Le nuage de points des individus-victimes (ici en bleu un échantillon aléatoire de la base) n'est pas homogène.  $F_1$  et  $F_2$  séparent chacun l'échantillon des individus en 2 : pour  $F_1$  on vérifie que c'est le montant du sinistre qui joue un rôle clé (plus il est important plus on est à droite de l'axe) ; pour  $F_2$ , c'est la présence ou non d'un accident de travail (quand il n'y en a pas, on est positif).

On vérifie qu'aucune observation ne contribue trop.

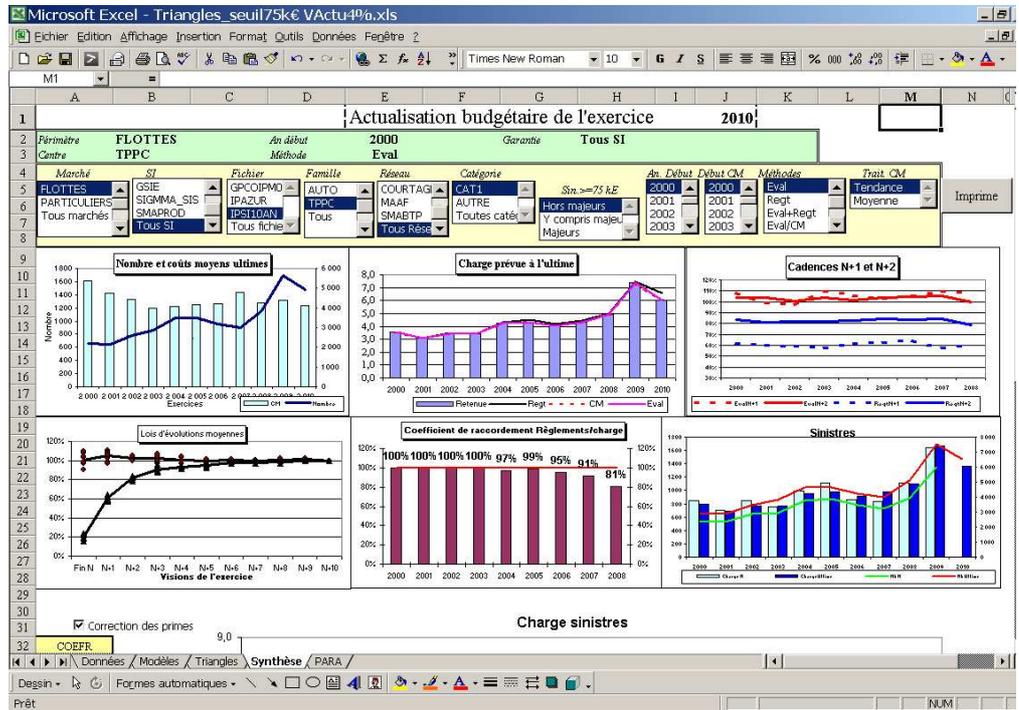


## VIII - C ANNEXE - CONSTITUTION DE LA BASE SINISTRES (SAS)

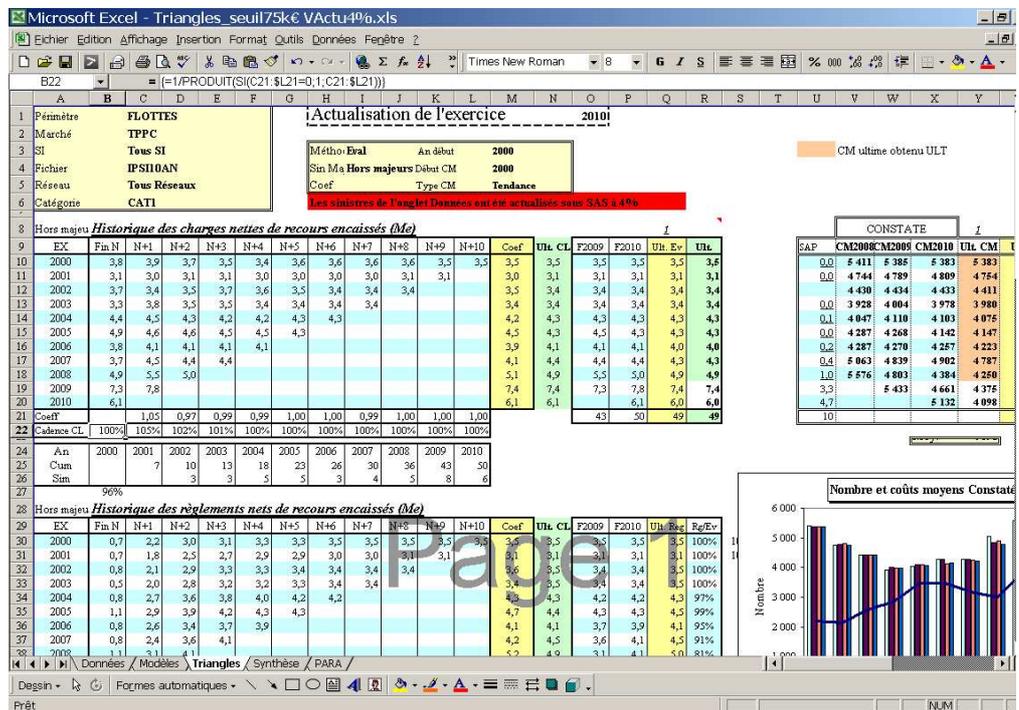


### VIII - D ANNEXE - OUTIL EXCEL DE DÉTERMINATION DES CADENCES

Pour la détermination des cadences de développement IV-D.2., nous avons mis en place un outil Excel que nous détaillons ci-dessous.



L'onglet « synthèse » : permet de sélectionner le périmètre sur lequel on souhaite déterminer la cadence et la méthode utilisée.



L'onglet « Triangles » : donne les triangles cumulés de charges et de règlements.

## VIII - E ANNEXE - PROGRAMME DE DETERMINATION DU SEUIL (R)

### ## lecture des données ##

```
donnees<-read.table("C:\\R\\Etude\\Données\\donneesseuil.txt",header=TRUE)
cout<-donnees[, "ultime"]
coutk<-donnees[, "ultimek"] # cout en k€
```

### ## QQ-Plot ##

```
qqplot(sort(cout/1000), main="QQplot exponentiel", col='blue')
```

### ## Mean Excess Plot ##

```
## Pour ajouter la droite de régression ##
```

```
AjoutReg<-function(inf,sup,couleur){
  res=data.frame(m$x,m$y)
  mbis<-res[res$m.x>=inf & res$m.x<=sup,]
  lm<-lm(mbis$m.y~mbis$m.x)
  abline(lm$coef[1],lm$coef[2],col=couleur,lwd=2)
  list(lm$coef)}
```

```
m<-meplot(cout,xlim=c(0,500000),main="Mean Excess Plot", col="blue",labels=FALSE,ylim=c(0,1500000))
title(ylab="e(u)",xlab="u")
AjoutReg(75000,500000,"green")
AjoutReg(200000,500000,"purple")
text(75000,250000,labels="75k€",col="green")
text(200000,600000,labels="200k€",col="purple")
```

### ## Pente de l'espérance résiduelle ##

```
## Sur toutes les données ##
```

```
m<-meplot(cout/1000,main=paste("Pente de la Mean Excess Function","Avec toutes les valeurs",sep="\n"),col="steelblue")
lm<-lm(m$y~m$x)
abline(lm$coef[1],lm$coef[2])
lm$coef
```

```
## En enlevant les dernières valeurs ##
```

```
m<-meplot(cout/1000,omit=12+3,main=paste("Pente de la Mean Excess Function", "En enlevant les dernières valeurs",sep="\n"),col="steelblue")
lm<-lm(m$y~m$x)
abline(lm$coef[1],lm$coef[2])
lm$coef
```

```
## Graphique de la pente de l'espérance résiduelle ##
```

```
penteespresiduelle <- fonction (fichier,omi,couleur,liminf,limsup){
  t = sort(fichier) # vecteur des abscisses
  n = length(t)-2-(omi+3)
  coeffpente = rep(0,n-1) # vecteur qui contiendra les valeurs de la pente
  coeffb = rep(0,n)
  khi = rep(0,n)
  seuil = rep(0,n)
  for (i in 1:n) {
    m<-meplot(fichier[fichier>=t[i]],omit=omi+3,col=couleur)
    lm<-lm(m$y~m$x)
    abline(lm$coef[1],lm$coef[2])
    seuil[i]=t[i]
    coeffpente[i] =lm$coef[2]
    coeffb[i] =lm$coef[1]
    khi[i] = lm$coef[2]/(1+lm$coef[2]) }
}
```

```
plot(t[t<=t[n]],coeffpente,ylab="Coefficient de la pente",xlab="Coût en k€",font.main=1,
main="Pente de l'espérance résiduelle",pch="*",col = "steelblue"
,xlim=c(liminf,limsup),ylim=c(0,2.5)) # on affiche
}
penteespresiduelle(coutk,20, "red",30,2000)
```

```
## Pour tester si ksi=0 ou ksi>0 ##
```

```
testksi0 <- fonction(X)
{t = sort(X) # vecteur des abscisses
n = length(t)
```

```

F2t = rep(0,n) # vecteur qui contiendra les valeurs de Fx(t)
for (i in 1:length(t))
  F2t[i] = log(log(1/(sum(t<=t[i])/n))) # on calcule Fx(t)
plot(t,F2t,xlab="Coût",ylab="log(log(1/F))", main="test de ksi=0",xlim=c(0,7000000),pch="*",col = "steelblue") # on affiche
}
testksisup0 <- fonction(X)
{ t = sort(X) # vecteur des abscisses
n = length(t)
F2t = rep(0,n) # vecteur qui contiendra les valeurs de Fx(t)
for (i in 1:length(t))
  F2t[i] = log(log(1/(sum(t<=t[i])/n))) # on calcule Fx(t)
plot(log(t),F2t,xlab="Log(Coût)",ylab="log(log(1/F))",main="test de ksi>0",xlim=c(0,15.2),pch="*",col = "steelblue") # on affiche
}

## Estimateur de Hill ##
testksi0(cout)
testksisup0(cout)

temp<-hill(cout/1000, main="Estimateur de Hill",
           option = "xi", start = 30,
           end = 1500, reverse = FALSE, p = NA, ci = 0.95,
           auto.scale = FALSE, labels = TRUE,ylim=c(0,1.5))
temp<-temp[order(temp[,temp$x],decreasing=FALSE),]

## Estimateur de Pickands ##
pgmPickands <- fonction (fichier){
  x<-sort(fichier)
  n<-length(x)
  maxi<- round(n%/4)-1
  pickands<-rep(NA,(maxi))
  excess<-rep(NA,(maxi))
  for(j in seq(from=1,to=maxi)){
    pickands[j]<- (1/log(2))*log((x[n-j]-x[(n-j*2)])/(x[(n-j*2)]-x[(n-j*4)]))
    excess[j]=j
  }
  par(ps=10)
  plot(excess,pickands,main="Estimateur de Pickands ",type="l",xlim=c(0,600),ylim=c(0,1.5)
       ,col = "steelblue")
  grid()
}
pgmPickands(cout)

## Estimateur des Moments DEdH ##
estDEdH <- fonction (x){
  x<-sort(x)
  n<-length(x)
  DEdH<-rep(NA,(n-1))
  temp<-rep(NA,(n-1))
  hill<-rep(NA,(n-1))
  S<-rep(NA,(n-1))
  for(j in seq(from=1,to=(n-1))){
    S[j]<- ((sum ( (log(x[(n-j+1):n])) - (log(x[n-j])) )^2) )/j
    hill[j]<-((sum ( (log(x[(n-j+1):n])) - (log(x[n-j])) ) ) /j
    temp[j]<- (1-(hill[j]^2)/S[j])^(-1)
    DEdH[j]<- (hill[j]+1-0.5*temp[j])
  }
  par(ps=10)
  plot(DEdH,main="Estimateur DeDH",type="l",col = "steelblue",ylim=c(0.3,1),xlim=c(0,1200))
  res=data.frame(S,DEdH,temp,hill)
  res
}
estDEdH(cout)

## Estimateur du maximum de vraisemblance ##
hill(cout/1000, main="Estimateur du maximum de vraisemblance", option = "alpha", start = 30,end = 1500, reverse = FALSE, p =
NA, ci = 0.95,auto.scale = FALSE, labels = TRUE,ylim=c(0.6,1.8))

```

OU

```

pgmEMV <- fonction (fichier){
  x<-sort(fichier)
  n<-length(x)
  EMV=as.vector(1:(n-1))
  t=as.vector(1:(n-1))
  nbexces=as.vector(1:(n-1))
  denominateur=as.vector(1:(n-1))
  for(j in seq(from=1,to=(n-1),by=1)){
    nbexces[j]=j
    denominateur[j]=sum(log(x[(n-j+1):n]))-j*log(x[n-j])
    EMV[j]<- nbexces[j]/denominateur[j]
    t[j]=x[n-j]
  }
  res=data.frame(EMV,t,nbexces,denominateur)
  res
}
res=pgmEMV(coutk)

## MinAMSE ##
sum=0 # initialisation de la variable de stockage
for (i in 1:1000){ # boucle de 1 000 simulations
  x<-rPareto1(1000,1.0611,200000) # création d'un 1 000-échantillon de sinistres tirés dans la Pareto
  y<-x[x<20000000] # sélection des sinistres inférieurs à 20 000 000€
  resy<-minAMSE(y) # lancement de la fonction minAMSE
  sum[i+1]<-sum[i]+resy*x0} # stockage cumulé des valeurs de seuils définies par minAMSE

## Estimateur GPD ##
pgmGPD <- fonction (fichier){
  x<-sort(fichier)
  n<-length(x)
  ksi=as.vector(1:(n-1))
  t=as.vector(1:(n-1))
  nbexces=as.vector(1:(n-1))
  for(j in seq(from=1,to=(n-1),by=1)){
    nbexces[j]=j
    fit<-gpd.fit(x,x[n-j])
    ksi[j]<- fit$mle[2]
    t[j]=x[n-j]
  }
  resgpd=data.frame(ksi,nbexces,t)
  resgpd
}
res=pgmGPD(coutk)
plot(res$nbexces,res$ksi,type="l",xlim=c(0,600),ylim=c(0,1.5),main="estimateur GPD",xlab="Nombre d'exces", ylab="Estimateur",
cex=0.001, pch="*", col = "steelblue")

```

## VIII - F ANNEXE - PROGRAMME DE DÉTERMINATION DU SEUIL - GERSTENGARBE (SAS)

```

/* ***** Série croissante - Série décroissante ***** */
%macro tritable (s_seuil);
/* Série croissante des sinistres sup au seuil */
proc sort data=sorties.seuil (where=(cout>&s_seuil. and cout2010 not in (1490))) /* On enlève les IRCA */
  out=sericroissante (keep=cout) ;
  by cout;
run;
proc sort data=sorties.seuil (where=(cout>&s_seuil. and cout2010 not in (1490))) /* On enlève les IRCA */
  out=seriedecroissante (keep=cout) ;
  by descending cout;
run;
%MEND tritable;

/* ***** Mann - Kendall ***** */
%macro MannKendall (table_entr,s_seuil,table_res);
/* On indice et on calcule Delta */
data temp1 (keep = cout delta_i n );
  set &table_entr.;

```

```

        retain n 0; /* pour mettre en mémoire n*/
        n + 1;
        delta_i = cout - lag(cout); /*couti-couti-1 */
        if n=1 then delta_i=0;
run;

/* Dans nobis on met le nombre d'observations */
proc sql noprint;
select count(*) into :nbobs from &table_entr.;
quit;

/* Calcul de nk */
data temp2 (keep = n cout delta_i nk);
    array W{&nbobs.} _temporary_;
    if _n_ = 1 then do I= 1 to &nbobs.;
        set temp1;
        W{I} = delta_i;
    end;
    set temp1;
    nk = 0;
    do ligne= 1 to (_n_ - 1);
        if W{ligne}<=delta_i then do;
            nk =nk +1;
        end; end;
run;

%MannKendall (seriecroissante,&par_seuil.,sorties.seried);
%MannKendall (seriedecroissante,&par_seuil.,sorties.seried);

proc sql noprint;
select count(*)+2 into :nbobs from sorties.seried;
quit;
data sorties.seried; set sorties.seried (drop=n); retain n &nbobs.; n= n - 1; run;
data sorties.seried; set sorties.seried; if _N_=1 then n=1; run;

proc sql;
    create table &table_res. as
    select a.n as i, a.cout,a.Ui, b.Ui as Uif, abs(a.Ui-b.Ui) AS diff
    from sorties.seried as a
    left join sorties.seried as b
    on (a.n=b.n);
run;

/* On somme les nk = Ui* */
data temp3; set temp2;
retain nktot 0;
nktot = nktot +nk;
run;

/* On calcule Ui */
data &table_res.;
set temp3;
if _n_=1 then do;
    num=0;
    denom=0;
    Ui=0;
end;
if _n_ > 1 then do;
    num = nktot - ((n * (n-1))/4);
    denom = sqrt((n*(n-1)*(2*n+5))/72);
    Ui = num / denom;
end;
run;
%MEND MannKendall;

/* Gertengarbe Plot - Recherche de l'intersection des courbes */
%MACRO Gertensgarbe (par_seuil,table_res);

```

```

%tritable(&par_seuil);
proc sort data=&table_res. (where=(i ne 1)) out=temp; by diff; run;
data seuil_final; set temp; if _N_ = 1 then output; run;

proc sql noprint; select floor(cout) into :valseuil from seuil_final; quit;

data &table_res.; set &table_res.; seuil = &valseuil.; run;

filename orig "&chemin.&table_res.&valseuil.html";
goptions reset=all; goptions device=gif gsfname=orig; title " Gertensgarbe Plot - Seuil obtenu = &valseuil." ;
SYMBOL1 i = join c = blue v = none ;
SYMBOL2 i = join c = cyan v = none ;

Proc GPLOT data = &table_res. (where=(i gt 1)) ;
plot (Ui Uif) * i / OVERLAY LEGEND ;
run ; quit ;
filename orig clear;
%MEND Gertensgarbe;

%macro trouve_seuil (nb);
%Gertensgarbe (0, sorties.gertensgarbe_0)
%do i=2 %to &nb.;
proc sql noprint;
select floor(cout) into :seuilsuivant from seuil_final;
quit;
%Gertensgarbe (&seuilsuivant., sorties.gertensgarbe_2);
%end;
%MEND trouve_seuil;

%trouve_seuil(5);

```

## VIII - G ANNEXE - PROGRAMME D'AJUSTEMENT DE LOIS (R)

fitdistr : pour estimer les paramètres d'une loi

ks.test : pour tester l'adéquation entre les données empiriques et la loi estimée

```

X<-cout[cout>seuil1 & cout<seuil2]-seuil1
a<-fitdistr(X,"lognormal")
ks.test(unique(X),"plnorm", a$estimate[1],a$estimate[2] )

X<-cout[cout>seuil2]-seuil2
b<-fitdistr(X,"lognormal")
ks.test(X,"plnorm", b$estimate[1],b$estimate[2])

p1<-sum(cout>seuil1 & cout<seuil2)/sum(cout>seuil1)
p2<-sum(cout>seuil2)/sum(cout>seuil1)

plotFsup<-function(borneinf,bornesup) {
  x<-seq(borneinf,bornesup,1000)
  y<-x[x>=seuil1 & x<seuil2]
  plot(y,
        plnorm(y-seuil1,a$estimate[1],a$estimate[2])/plnorm(seuil2,a$estimate[1],a$estimate[2])*p1,type="l",col="red",lwd=3,
        main="Ajustement d'une loi mélange",ylim=c(0,1),xlim=range(x))

  plot(ecdf(cout[cout>seuil1]),add=TRUE,col="grey",pch=".",lwd=3)

  y<-x[x>=seuil2]
  lines(y, plnorm(y-seuil2,b$estimate[1],b$estimate[2])*p2+p1,lwd=3,col="blue")
}
plotFsup(seuil1,2000000)

```

## VIII - H ANNEXE - RAPPEL SUR LES MOMENTS

Pour une v.a.r.  $X$  à valeurs entières, le moment factoriel d'ordre  $k$  est défini par  $\mu_{(k)} = E[X(X-1)\dots(X-k+1)]$

Le moment simple d'ordre  $k$  d'une v.a.r.  $X$  est noté  $m_k = E(X^k)$ .

Les moments simples s'expriment en fonction des moments factoriels selon :

$$m_1 = \mu_{(1)}$$

$$m_2 = \mu_{(2)} + \mu_{(1)}$$

$$m_3 = \mu_{(3)} + 3\mu_{(2)} + \mu_{(1)}$$

$$m_4 = \mu_{(4)} + 6\mu_{(3)} + 7\mu_{(2)} + \mu_{(1)}$$

Le moment centré d'ordre  $k$  de la v.a.r.  $X$  noté  $\mu_k$  est tel que  $\mu_k = E[(X - E(X))^k]$

Ils se déduisent des moments simples par les expressions suivantes :

$m_1$  est la moyenne,

$\mu_2 = m_2 - m_1^2$  est la variance,

$$\mu_3 = m_3 - 3m_1m_2 + 2m_1^3$$

$$\mu_4 = m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4$$

$\mu_2, \mu_3$  et  $\mu_4$  étant utilisés pour caractériser la forme d'une distribution.

Nous définissons :

- le coefficient d'asymétrie (skewness) :  $\gamma_1 = \frac{\mu_3}{(\mu_2)^{\frac{3}{2}}}$ 
  - o  $\gamma_1 > 0$  : la distribution est étalée vers la gauche et la queue de distribution vers la droite.
  - o  $\gamma_1 < 0$  : la distribution est étalée vers la droite et la queue de distribution vers la gauche.
  - o  $\gamma_1 = 0$  : la distribution est symétrique, normale.
- l'excès de coefficient d'aplatissement (excès de kurtosis) :  $\gamma_2 = \frac{\mu_4}{(\mu_2)^2} - 3$ .
  - o  $\gamma_2 > 0$  : la distribution est *pointue*,
  - o  $\gamma_2 < 0$  : la distribution est *aplatie*,

$\gamma_2 = 0$  la distribution est *quasi-normale*, d'autant mieux qu'elle sera symétrique.

### VIII - I ANNEXE - CALCUL DE LA BORNE SUPERIEURE DE LA CLASSE 12

Calcul de  $E(X / X > z)$  :

$$E(X / X > z) = \frac{E(X \cdot 1_{\{X > z\}})}{P(X > z)} = \frac{\int_z^{+\infty} x \cdot f(x) \cdot dx}{\sigma \sqrt{2\pi}} \int_z^{+\infty} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} dx$$

En posant  $y = \frac{\ln x - \mu}{\sigma}$ , on a  $x = e^{\sigma y + \mu}$  et  $dx = \sigma \cdot e^{\sigma y + \mu} \cdot dy$  d'où :

$$E(X \cdot 1_{\{X > z\}}) = \frac{1}{\sqrt{2\pi}} \int_{\frac{\ln z - \mu}{\sigma}}^{+\infty} e^{-\frac{y^2}{2}} e^{\sigma y + \mu} dy = \frac{1}{\sqrt{2\pi}} \int_{\frac{\ln z - \mu}{\sigma}}^{+\infty} e^{-\frac{(y - \sigma)^2}{2}} e^{\mu + \frac{\sigma^2}{2}} dy$$

et en posant  $u = y - \sigma$

$$E(X \cdot 1_{\{X > z\}}) = \frac{e^{\mu + \frac{\sigma^2}{2}}}{\sqrt{2\pi}} \int_{\frac{\ln z - \mu}{\sigma} - \sigma}^{+\infty} e^{-\frac{u^2}{2}} du = e^{\mu + \frac{\sigma^2}{2}} \cdot [1 - \phi\left(\frac{\ln z - \mu}{\sigma} - \sigma\right)]$$

avec  $\phi$  la fonction de répartition de la loi normale centrée réduite.

$$\text{D'autre part, on a : } P(X > z) = P(\ln X > \ln z) = P\left[\frac{\ln X - \mu}{\sigma} > \frac{\ln z - \mu}{\sigma}\right] = 1 - \phi\left(\frac{\ln z - \mu}{\sigma}\right)$$

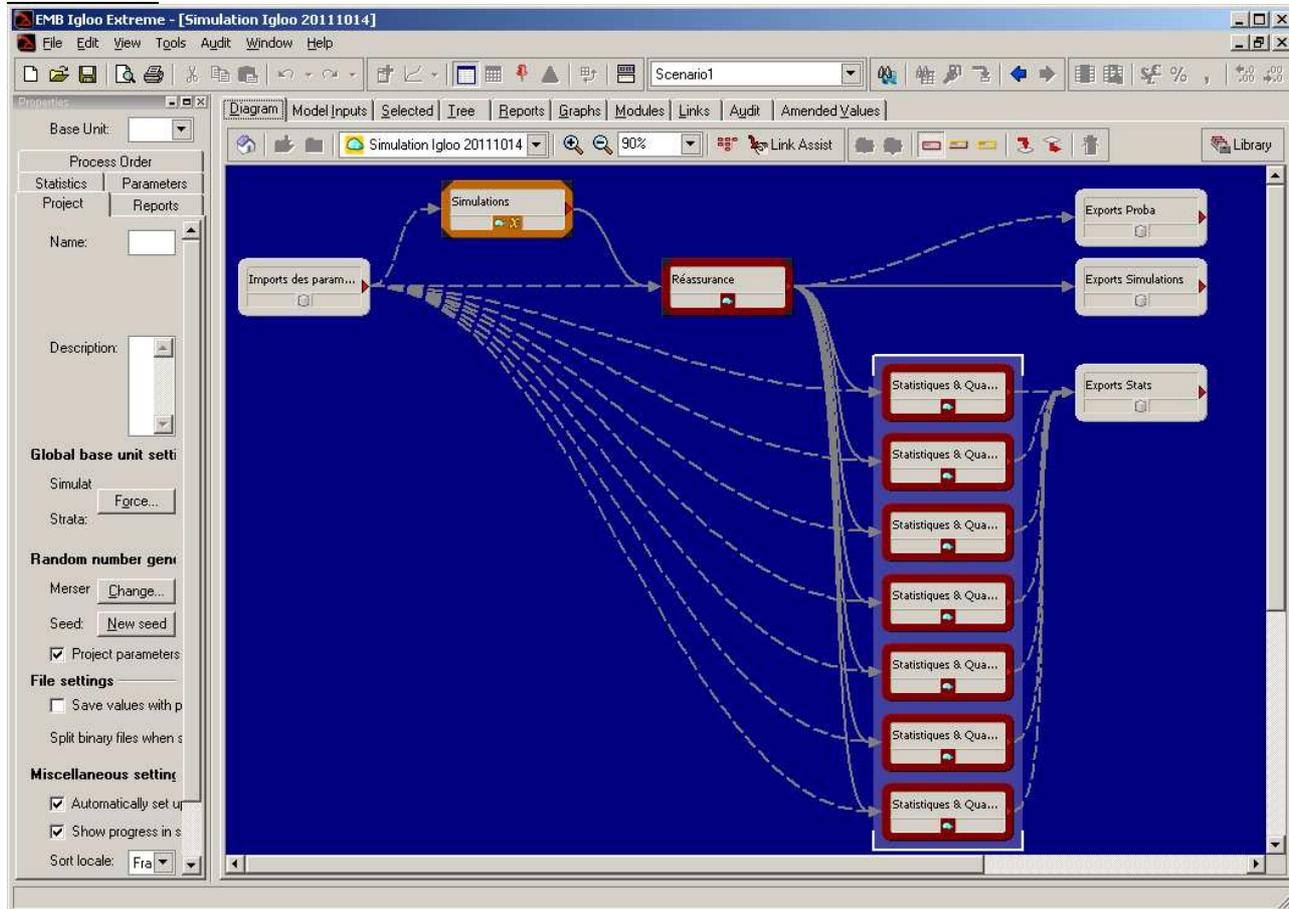
$$\text{D'où : } E(X / X > z) = \frac{e^{\mu + \frac{\sigma^2}{2}} \cdot [1 - \phi\left(\frac{\ln z - \mu}{\sigma} - \sigma\right)]}{1 - \phi\left(\frac{\ln z - \mu}{\sigma}\right)}$$



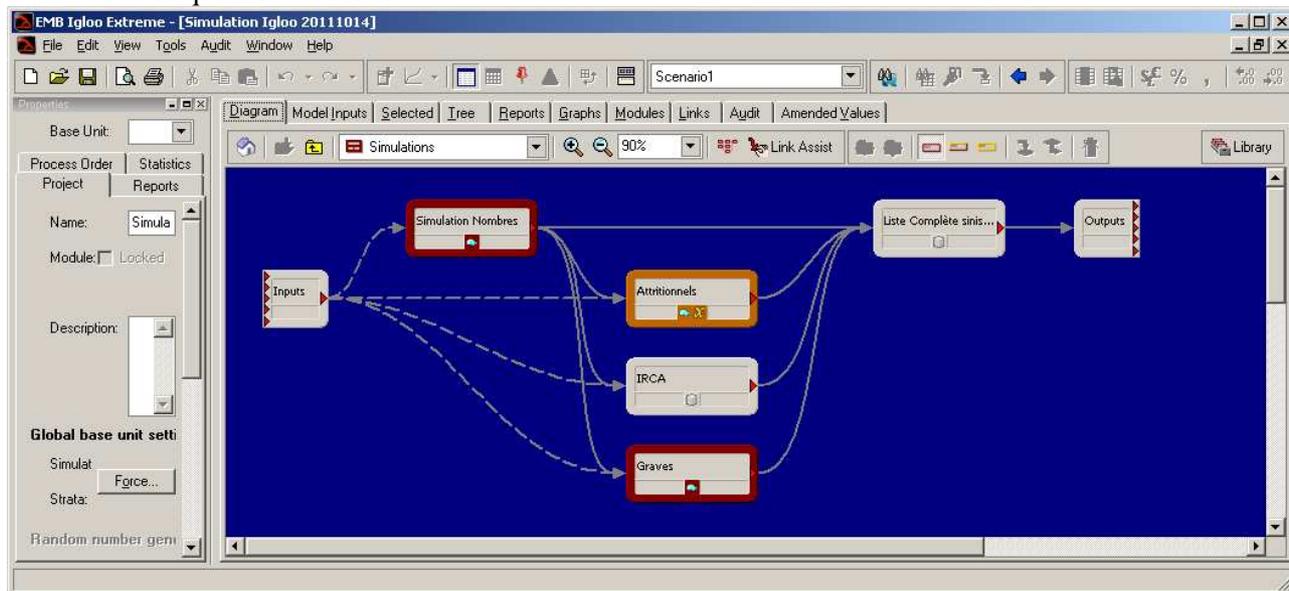
## VIII - K ANNEXE - OUTIL DE SIMULATION : IGLOO

La partie pratique de ce mémoire a donné lieu au développement d'un modèle sous Igloo nous permettant la mise en œuvre des méthodes 1 et 2 présentées dans les parties V-D.3. et VI-C.3 de ce mémoire. Nous indiquons plus précisément les fonctionnalités de l'outil ainsi que le détail des briques mises en place.

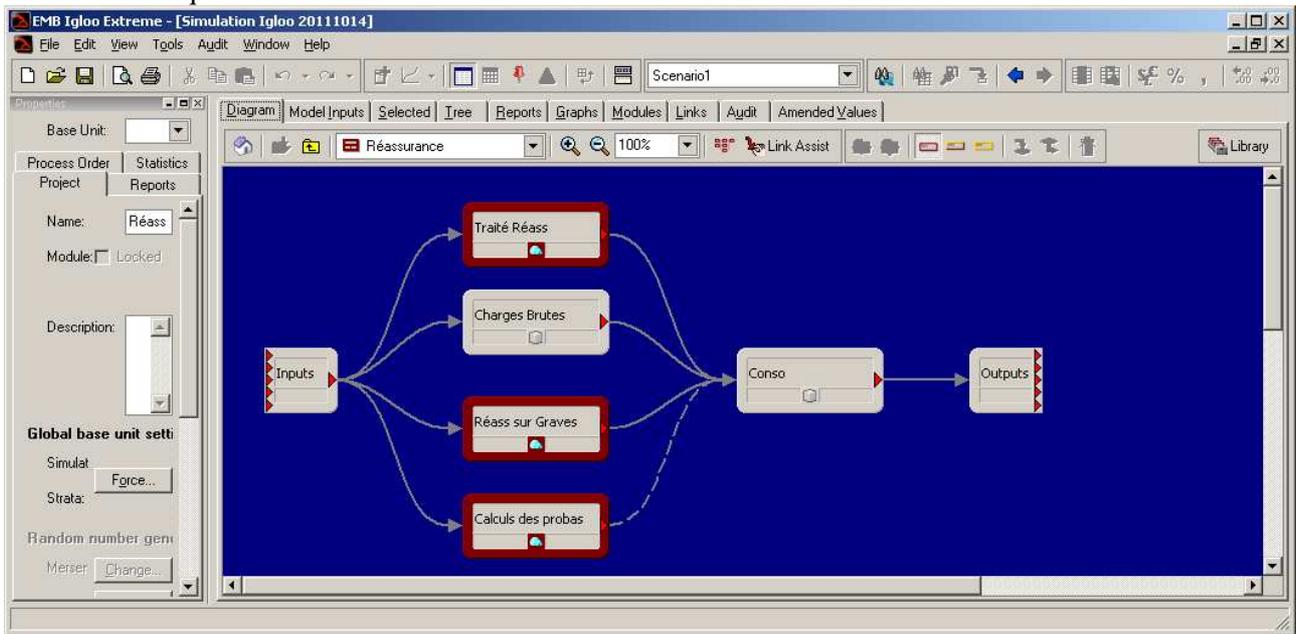
### Méthode 1 :



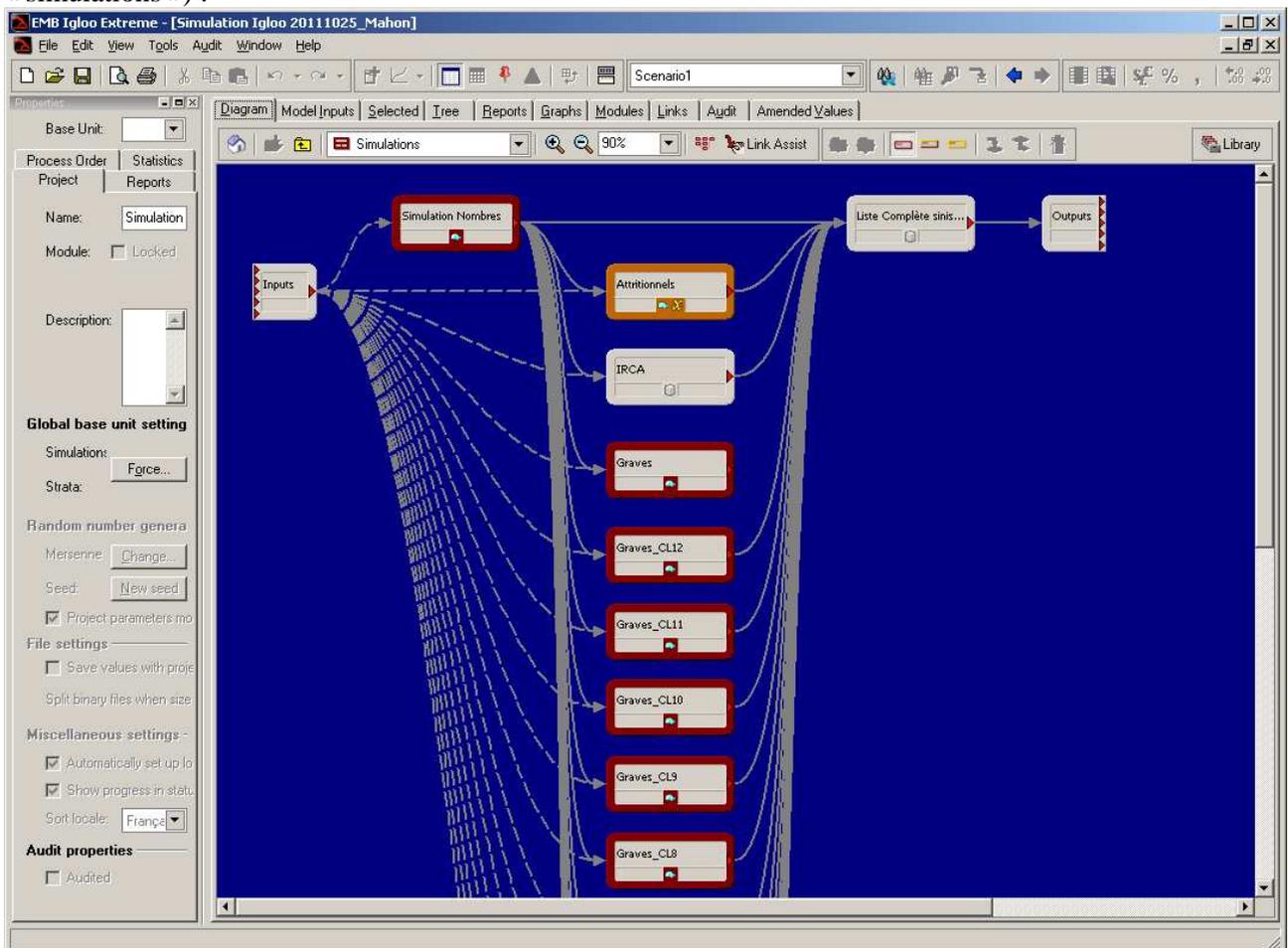
### Détail de la brique « simulations » :



Détail de la brique « Réassurance » :



Pour la méthode 2, nous avons uniquement modifié la simulation des sinistres graves (brique « simulations ») :



## **IX - BIBLIOGRAPHIE**

### **IX - A INDEMNISATION DOMMAGE CORPOREL**

- [1] Le droit à l'indemnisation des victimes d'un accident de la circulation, H. Groutel, 1987.
- [2] Précis d'évaluation du dommage corporel, H. Béjui-Hugues & I. Bessières-Roques & C. Fournier, 2005.
- [3] Rapport du groupe de travail chargé d'élaborer une nomenclature des préjudices corporels, J.-P. Dintilhac juillet 2005.
- [4] Livre blanc sur l'indemnisation du dommage corporel de l'association française de l'assurance, avril 2008.
- [5] Jurisprudence automobile, janvier 2011.

### **IX - B ETUDES ASSUREURS / REASSUREURS**

- [1] Assurance automobile : les sinistres corporels en 2008, avril 2010, FFSA / GEMA.
- [2] Assurance automobile, rapport 2009, FFSA.
- [3] Le marché de l'assurance automobile en 2009, analyse des résultats détaillés, octobre 2010 FFSA / GEMA.
- [4] Les sinistres graves au 31 décembre 2008, février 2011, FFSA et GEMA.
- [5] The european motor insurance market, C.E.A. (european insurance and reinsurance federation), statistic n°38, february 2010.
- [6] Etude des grands sinistres corporels : inflation et surinflation, Secura Re, 2007.

### **IX - C AUTRES ETUDES**

- [1] Etude sur la réinsertion sociale et professionnelle des personnes sortant d'établissement ou service sanitaire de médecine physique réadaptation, CREDOC, juin 2007.
- [2] Fichier des indemnités allouées aux victimes d'accidents de la circulation, Rapport annuel, Dossiers réglés en 2009, Agira, décembre 2010.
- [3] La sécurité routière en France, bilan 2009, ONISR, 2010.

### **IX - D MEMOIRES D'ACTUARIAT**

- [1] Flottes automobiles : un nouveau modèle de tarification. Impact de la conservation sur la distribution du ratio sinistres à primes, T. Nguyen, 2008.
- [2] Processus de surveillance et de majoration des contrats flottes d'entreprise d'Axa France, R. Boyer Chammard, 2008.
- [3] Problématique de seuil dans la modélisation de la sinistralité en réassurance non vie, E. Deprey & A. Godzinski, 2007.

### **IX - E THÉORIE ET CONCEPTS**

- [1] A method for the statistical definition of extreme-value regions and their application to meteorological time series, F.W. Gertensgarbe & P.C Werner, 1989.
- [2] Assurance, comptabilité, réglementation, actuariat, A. Tosetti & T. Béhar & M. Fromenteau & S. Ménart, 2011.
- [3] Assurance non-vie, modélisation, simulation, C. Partrat & J-L. Besson, 2005.
- [4] EDF statistics for goodness-of-fit and some comparisons, Journal of the American Statistical Association, M.A. Stephens, 1974.
- [5] Excess functions and estimation of the extreme-value index, J.Beirlant and al., Journal of the American Statistical Association, décembre 1996.
- [6] Grands sinistres et lois mélanges, Z. Benabbou & C. Partrat, 1994.
- [7] Kernel goodness-of-fit statistics for Pareto-type behavior and threshold selection for tail index estimation, Y.Goegebeur and al.
- [8] Itinéraires en statistiques et probabilités, H. Carnec & J-M. Dagoury & R. Seroux & M. Thomas, 2000.
- [9] Les sinistres graves en assurance automobile : une nouvelle approche par la théorie des valeurs extrêmes, N. Benlaha, M. Grun-Réhomme, O. Vasehko, 2009.
- [10] Loss distribution, R. Hogg & S. Klugman, 1984.
- [11] Mathématiques de l'assurance non-vie – Tome II : Tarification et provisionnement, M. Denuit et A.

Charpentier, *Economica*, 2005.

- [12] Premiers pas en régression linéaire avec SAS®, J. Confais & M. Le Guen.
- [13] Provisionnement technique en assurance non-vie – Perspectives actuarielles modernes, C. Partrat, E. Lecoœur, JM. Nessi, E. Nisipasu et O. Reiz, *Economica*, 2007.
- [14] Statistical theory of extreme values and some practical applications, Gumbel, 1958
- [15] Statistics of extremes, theory and applications, J. Beirlant, Y. Goegebeur, J. Segers, J. Teugels, 2010.
- [16] Statistique, la théorie et ses applications, M. Lejeune, 2005.
- [17] Tail index estimation, Pareto quantile plots and regression diagnostics, J. Beirlant and al., 1996.
- [18] Transition Matrix Theory And Individual Claim Loss Development, John B. Mahon, 2005.
- [19] Truncated and censored samples, A. Clifford Cohen, 1991.
- [20] Truncated and censored samples from normal populations, H. Schneider, 1986.

## **IX - F LOGICIELS**

- [1] Le logiciel R, Maîtriser le langage, effectuer des analyses statistiques, P. Lafaye de Micheaux, R. Drouilhet, B. Liqueur, 2010.
- [2] Introduction to scientific programming and simulation using R, O. Jones, R. Maillardet, A. Robinson, 2009.

## **IX - G SITES INTERNET**

- [3] FFSA : <http://www.ffsa.fr/>
- [4] INSEE : <http://www.insee.fr/>
- [5] A. Charpentier : <http://freakonometrics.blog.free.fr/>
- [6] R : <http://cran.r-project.org/>
- [7] Mémoires : <http://www.ressources-actuarielles.net/>
- [8] "ASTIN" : publications de l'ASTIN & de l'AFIR