

جامعة باجي مختار - عنابة

BADJI MOKHTAR - ANNABA UNIVERSITY  
UNIVERSITÉ BADJI MOKHTAR - ANNABA

كلية العلوم

FACULTY OF SCIENCES  
FACULTE DES SCIENCES

قسم الرياضيات

DEPARTMENT OF MATHEMATICS  
DEPARTEMENT DE MATHEMATIQUES



# THÈSE

Présentée en vue de l'obtention du diplôme de

Doctorat

Option

Modélisation Mathématique-ACTUARIAT

## TITRE

EVALUATION DES PROVISIONS TECHNIQUES BASEE SUR  
LA SURDISPERSION DES SINISTRES

Par

BELHAMRA THARA

**DIRECTEUR DE THÈSE :** REMITA Mohamed Riad Prof. U.B.M. Annaba  
**CO-ENCADREUR :** ZEGHDOUDI Halim M.C.A U.B.M. Annaba

Devant le jury

**PRESIDENT :** BOUTABIA Hacène Prof U.B.M. Annaba

**Examineurs :** SEDDIK-AMEUR Nacera Prof U.B.M. Annaba

ARRAR Nawel M.C.A U.B.M. Annaba

BENCHAABANE Abbes M.C.A U.8 mai 45, Guelma

# Remercîment

Je tiens à remercier vivement le professeur Mohamed Riad REMITA, pour la confiance qu'il m'a témoignée en acceptant la direction scientifique de mes travaux. Je lui suis reconnaissante de m'avoir fait bénéficier tout au long de ce travail de sa grande compétence, de sa rigueur intellectuelle, de son dynamisme, et de son efficacité certaine que je n'oublierai jamais. Soyez assuré de mon attachement et de ma profonde gratitude.

Je suis très honoré à remercier de la présence à mon jury de thèse et je tiens à remercier :

Monsieur Halim ZAGHDOUDI, docteur à l'université Badji Mokhtar Annaba, d'avoir accepté de faire partie du jury de cette thèse. Je le remercie pour les conseils scientifiques qu'il a apportés en qualité de Co-encadreur, pour les suggestions et les remarques judicieuses qu'il m'a indiquées, ainsi que pour son immense aide pour mener à bien ces travaux.

Monsieur, BOUTABIA Hacène, qui a bien voulu juger une grande partie de ce travail. Je le remercie pour le temps consacré à la lecture de ce travail ainsi que pour les commentaires m'ayant permis de l'améliorer.

Madame Nawel REMITA pour l'honneur qu'elle m'a fait pour sa participation à mon jury de thèse en qualité d'examinatrice, pour le temps consacré à la lecture de cette thèse, et pour toutes remarques intéressantes qu'elle m'a faites.

J'associe à ces remerciements à madame SEDDIK AMEUR Nacira, pour avoir accepté d'examiner mon travail.

Monsieur BENCHAAABANE Abbes, pour l'intérêt qu'il a manifesté en participant en qualité de membre invité à ce jury.

Je tiens à remercier sincèrement la famille pédagogique du département de

mathématiques, qui s'est toujours montrée à l'écoute et très disponible tout au long de mon cursus.

A titre plus personnel, Je remercie chaleureusement mon ami et père, Ali pour la grande patience, l'encouragement et la confiance qu'il m'a témoigné dont il a fait preuve à la relecture de mon manuscrit. Je tiens à le remercier surtout pour son soutien moral ininterrompu et ses nombreux conseils tout le long de ma thèse.

Bien sûr, atteindre ces objectifs n'aurait pas été possible sans l'aide de mes deux adorables frères, Amir et Amjed, je tiens à les remercier profondément pour leur soutien et leurs encouragements, leur patience tout au long de la thèse.

Je souhaite remercier spécialement mes amies Yaya, Lili, Narimane, Chahla et Zaineb pour leur aide et leur bonne humeur.

# Résumé

En assurance non-vie, les provisions pour sinistres à payer sont souvent les plus importantes en termes de montant et nécessitent, par conséquent, une évaluation très précise. Celles-ci correspondent aux sinistres survenus à la date d'inventaire mais qui n'ont pas encore été réglées. De plus, Les recherches actuelles dans le domaine d'évaluation des provisions techniques reposent principalement sur la théorie. De ce fait, cette thèse traite cette thématique à travers une approche graphique pour la méthode de Chain-Ladder.

En s'appuyant sur une analyse comparative des résultats des estimations des réserves du modèle proposé avec des méthodes paramétriques et non-paramétriques dans le but de mettre en lumière l'estimation des provisions techniques au sein des compagnies d'assurances pour ainsi développer un aspect géométrique de la théorie standard de provisionnement en utilisant des paiements incrémentaux.

**Mots clés :** Provisionnement non-vie, IBNR, Chain-Ladder, Approche graphique, Approche incrémentale, les GLMs, Facteur de développement graphique.

# Abstract

In non-life insurance, claims reserves are often the most important in terms of amount and require, therefore, a very accurate assessment. These correspond to claims occurring on the date inventory but have not yet been settled. In addition, current researches in the field of valuation of technical provisions are mainly based on the theory. Thus, this thesis deals with this topic through a graphical approach for the Chain-Ladder method. Based on a comparative analysis of the results of reserves estimates of the proposed model with parametric and non-parametric methods in order to highlight the estimation of technical provisions within insurance companies to develop a geometry aspect of the standard theory of provisioning using incremental payments.

**Key words :** Non-life provisioning, IBNR, Chain-Ladder, Graphical approach, Incremental approach, The GLMs, Graphical development factor

## الملخص

في التأمين على غير الحياة ، غالباً ما تكون إحتياجات المطالبات هي الأكثر أهمية من حيث المبلغ و بالتالي تتطلب تقييماً محدداً للغاية.

هذه المطالبات هي التي تحدث في تاريخ المخزون و لكن لم تتم تسويتها بعد.

بالإضافة الى ذلك ، يستند البحث الحالي في مجال تقييم الأحكام التقنية أساساً على النظرية. و نتيجة لذلك ، تتناول هذه الرسالة هذا الموضوع من خلال مقارنة بيانبة لطريقة Chain-Ladder .

استناداً على تحليل مقارن لنتائج تقديرات الإحتياطيات للنموذج المقترح مع نتائج طرق برامترية و غير برامترية من أجل تسليط الضوء على المميزات التقنية داخل شركات التأمين لإعطاء جانب هندسي للنظرية القياسية للتمويل باستخدام المدفوعات الإضافية.

**الكلمات المفتاحية:** توفير غير - الحياة ، Chain-Ladder ، الطريقة الهندسية ، الطريقة التدريجية GLMs ، عامل تطوير الرسومات.

# Table des matières

<b>Introduction</b>	<b>8</b>
<b>1 Méthodes de provisionnement</b>	<b>15</b>
1.1 La méthode Chain-Ladder . . . . .	16
1.1.1 Les link-ratios . . . . .	16
1.1.2 L'estimation des provisions . . . . .	17
1.2 Le modèle de Mack . . . . .	18
1.2.1 Description du modèle . . . . .	18
1.2.2 Lien avec la méthode de Chain-Ladder . . . . .	18
1.2.3 Version stochastique de Chain-Ladder . . . . .	19
1.2.4 Propriétés des estimateurs de Chain-Ladder . . . . .	19
1.2.5 L'erreur quadratique moyenne de prédiction . . . . .	21
1.3 La méthode bootstrap . . . . .	23
1.3.1 Définition de la méthode . . . . .	23
1.3.2 Méthodologie de la procédure Bootstrap . . . . .	24
1.4 L'approche incrémentale . . . . .	26
1.4.1 Présentation de l'approche . . . . .	26
1.4.2 Hypothèses des modifications du modèle de Mack . . . . .	27
1.4.3 Estimation des paramètres . . . . .	27
1.4.4 Calcul de l'erreur standard . . . . .	31
1.5 Comparaison entre le modèle de Mack et l'approche incrémentale . . . . .	32
<b>2 Les modèles factoriels stochastiques</b>	<b>33</b>
2.1 Les modèles linéaires généralisés . . . . .	34
2.1.1 Introduction et motivation . . . . .	34
2.1.2 Familles exponentielles . . . . .	39
2.1.3 Modèle GLM . . . . .	41
2.2 Les GLM et le provisionnement . . . . .	45
2.2.1 Le modèle de poisson . . . . .	45
2.2.2 Le modèle ODP (Over-dispersed poisson) . . . . .	49
2.2.3 Le modèle Gamma . . . . .	55

<b>3 Sur l'approche géométrique de la méthode Chain-Ladder : Propriétés et Application</b>	<b>57</b>
3.1 Approche graphique pour le calcul des réserves . . . . .	58
3.1.1 Théorie et développement du modèle . . . . .	58
3.1.2 Facteurs de développement graphique : . . . . .	59
3.1.3 Réserves de sinistres . . . . .	62
3.1.4 Illustration numérique . . . . .	62
3.2 Application et analyse critique . . . . .	66
<b>Conclusion</b>	<b>73</b>
<b>Perspectives de recherches</b>	<b>75</b>
<b>Bibliographie</b>	<b>77</b>



# Introduction

## Contexte

Le monde de l'assurance connaît un changement réglementaire considérable. En effet, les assureurs sont considérés comme les institutions financières ayant un impact économique non négligeable au niveau des pays et même de par le monde. Cette interdépendance nécessite la protection des différents protagonistes en particulier les assurés pour éviter une crise financière semblable à celle des Suprimes. Elle est le résultat de la surévaluation des capacités financières des assurés au niveau des banques ; ainsi les banques se sont retrouvées avec un parc immobilier considérable et la chute des prix est amorcée et la crise financière s'est installée [5].

Les banques, pour éviter ces crises, doivent faire recours aux assurances pour couvrir et prévenir ce genre de risques car parmi les prérogatives des assurances c'est de veiller aux intérêts de l'assuré selon ces ressources.

Le point faible du système d'assurance se trouve dans le fait que le cycle de production soit inversé, c'est-à-dire : l'assureur ne connaît son prix de revient qu'après avoir payé les sinistres ; Les primes étant déjà encaissées, il est important pour lui d'avoir suffisamment de provisions pour pouvoir payer les sinistres futurs et donc rembourser ses dettes envers les assurés [7].

Une première réglementation, Solvabilité I, permet d'imposer les contours du marché de l'assurance, mais elle montre aussi ses faiblesses face à l'environnement financier peu stable. Dès lors une nouvelle version, Solvabilité II, s'est imposée. Elle va permettre de corriger certains défauts tout en se fondant sur trois piliers :

- Les exigences quantitatives.
- Le processus de contrôle prudentiel.
- Information et discipline de marché [5].

Le paiement de sinistre ne s'effectue pas en général dans l'année même de sa survenance ; c'est particulièrement vrai dans certaines branches de l'assurance comme par exemple la responsabilité civile automobile en cas de sinistre accompagnés de dommages corporels.

Le règlement des sinistres s'étale au fil du temps il est donc nécessaire de constituer des réserves pour honorer les dettes futures, comme le montant qui sera au final payé pour couvrir le sinistre est inconnu au départ, la somme à mettre en réserve est également inconnue donc il faut l'estimer et c'est le but de notre travail.

Parallèlement, les provisions techniques représentent la grande partie du passif d'une compagnie d'assurance (non-vie). Comme il est nécessaire de savoir qu'une mauvaise estimation des fonds propres basé sur une évaluation erronée des provisions techniques pourrait mener à des décisions peut appropriées à l'état de santé financier d'un assureur [7].

D'autre part, certaines catégories d'assurance impliquent un retard important dans l'événement générant une réclamation et sa gestion. Pendant cet intervalle, il peut y avoir une grande incertitude quant au montant du règlement final [7]. La réserve de perte consiste à estimer les paiements futurs des créances survenues sur un portefeuille d'assurances. Dans ce cas, Les paiements futurs qui seront effectués sur ces créances sont un passif pour l'assureur et la plupart des assureurs sont tenus d'estimer la taille de ces passifs pour inclusion dans leurs états financiers [23]. Ces estimations peuvent être calculées par la soi-disant théorie IBNR (Incurred But Not Reported), qui est basé sur les paiements de réclamations passées pour estimer son développement futur.

Plusieurs approches ont été développées pour donner des estimations raisonnables. Mack [17] a développé pour la première fois un modèle stochastique sous-jacent à la méthode de Chain-Ladder qui est une méthode déterministe et probablement la méthode la plus populaire pour estimer les réserves de sinistres IBNR [5].

Après cela, Renshaw et Verral (1998) [12] ont présenté un modèle statistique sous-jacent à la technique Chain-Ladder, qui se présente sous la forme d'un modèle linéaire généralisé, pour ça ils ont utilisé une approche quasi-vraisemblance. Mack et Venter (2000) [10] fournissent une comparaison détaillée des modèles stochastiques qui reproduisent l'estimation des réserves de Chain-Ladder et qui pense que la question de quel modèle Sous-tend cet algorithme ne peut pas être répondu par une demande historique ou même par une déduction logique stricte et pour évaluer

son erreur de prédiction, un modèle stochastique sous-jacent est requis.

Les modèles comparés dans Mack [17] et England et Verral [20] qui sont le modèle Chain-Ladder distribution free (DFCL) et Poisson sur-dispersé (ODP) sont les seuls qui ont conduit à la même estimation pour  $D_{i,n}$  que l'algorithme de Chain-Ladder et ils ont conclu que seul le modèle DFCL peut être considéré comme sous-jacent à l'algorithme de Chain-Ladder. Ensuite, Verral [33] a fait une enquête sur les modèles de calcul des réserves stochastiques et la technique de Chain-Ladder dans laquelle il a exploré en détail la relation entre les modèles décrits par Renshaw et Verral [28] et Mack [19], il a suggéré dans son article un nouveau modèle et de nouvelles façons de tenir compte des paiements incrémentaux négatifs pour la technique de Chain-Ladder [17], puis England and Verral [11,12] ont montré comment les distributions prédictives du passif en suspens en assurance générale peuvent être obtenus en utilisant bootstrap ou les techniques bayésienne pour les modèles statistiques clairement définis. Et ils donnent un moyen d'obtenir une distribution prédictive à partir des modèles de réclamation de réserve récursive  $y$  compris le modèle introduit par Mack.

Chorfi et Remita [8] ont proposé une nouvelle méthode de calcul de réserves en utilisant une approche incrémentale stochastique, en modifiant le modèle de Mack et en évitant une étape de calcul, ils n'ont utilisé que les paiements incrémentaux pour estimer les réserves de sinistres.

A la lecture de l'état de l'art, il existe plusieurs techniques d'évaluation et d'estimation des provisions parmi lesquelles on retient celle dite IBNR (Incurred But Not Reported) basé sur l'évolution passée du coût des sinistres pour estimer son développement futur. En pratique ces techniques sont utilisées soit pour estimer les réserves purement IBNR, soit les réserves de sinistres totales. Cette étude a permis dans un premier moment de réaliser un article scientifique appelé Approche graphique pour le calcul des réserves [4], dans lequel nous avons proposé une nouvelle vision pour le problème du provisionnement pour les compagnies d'assurances.

Notre thèse est répartie en trois chapitres dont les détails (notre thèse est organisée comme suit)

On introduit dans un premier lieu, les méthodes classiques de provisionnement, pour cela nous avons détaillés la méthode de Chain-Ladder et sa version stochastique qui est le modèle de Mack, ensuite, nous avons parlé de l'approche incrémentale et pour en déduire la distribution des réserves nous avons défini la méthode Bootstrap. Dans un second lieu, on s'intéresse aux modèles factoriels stochastiques pour la modélisation des provisions techniques dans les compagnies d'assurances,

pour cela nous avons détaillés le provisionnement par les modèles linéaires généralisées. Dans un troisième lieu, nous avons présenté notre travail de recherche qui se résume dans une approche graphique pour le calcul des réserves. Enfin, on s'appuie sur des données en assurance non-vie pour illustrer, comprendre et comparer les notions introduites tout au long de cette thèse.

L'ensemble des résultats présentés dans cette thèse ont été obtenus à travers des outils développés sous le logiciel R, Matlab, Excel, AutoCAD (logiciel d'illustration graphique).

## Cadre mathématique et Notations

Dans cette section, nous introduisons le cadre mathématique pour les réserves de sinistres et les principales notations utilisées dans cette thèse. Dans la plupart des cas, les passifs de sinistres en suspens (les indemnités de perte) sont estimés dans des triangles de développement des sinistres qui séparent les sinistres selon deux axes temporels : les années de survenance  $i \in \{1, \dots, I\}$  et les années de développement  $j \in \{1, \dots, J\}$ . Nous supposons que la dernière période de développement est donnée par  $J$ , i.e.  $X_{i,j} = 0$  pour  $j > J$ . De plus, pour simplifier, nous supposons que  $I = J$ . Bien sûr, toutes les formules sont vraies pour  $I > J$  (trapèzes de développement). Nous considérons la matrice des variables aléatoires

$$X := (X_{i,j})_{1 \leq i,j \leq I} \quad (1)$$

Pour certains  $I \leq 2$ . Nous supposons que la variable  $X_{i,j}$  désigne les paiements incrémentaux de l'année de développement  $j$  pour les sinistres survenus au cours de l'année d'accident  $i$ . Dans un triangle de développement de sinistres, les années d'accident sont généralement sur la ligne verticale alors que les périodes de développement sont sur la ligne horizontale. Habituellement, les tables de développement des pertes se divisent en deux parties : le triangle supérieur où nous avons des observations  $\hat{X}_{i,j} \in R$  et le triangle inférieur où nous voulons estimer les paiements impayés. Sur les diagonales, nous voyons toujours les années comptables. Les données de sinistres ont donc la structure suivante ( $I = J$ ) :

Les données peuvent être présentées sous forme cumulative ou sous forme non-cumulative (incrémentale). Nous considérons une matrice de variables aléatoires

$$C := (C_{i,j})_{1 \leq i,j \leq I} \quad (2)$$

Pour certains  $I \geq 2$ . La variable  $C_{i,j}$  est interprétée comme le montant des sinistres cumulés de l'année d'accident  $i$  jusqu'à l'année de développement  $j$ , les données cumulatives sont données par

$$C_{i,j} = \sum_{l=1}^j X_{i,l} \quad (3)$$

Dans la théorie standard des réserves de sinistres, également appelée la théorie IBNR, on suppose que les paires  $(i, j)$  d'années d'accident et de développement avec  $i + j \leq I + 1$  décrivent les années passées pour lesquelles le triangle run-off des observations réelles  $\hat{C}_{i,j} \in R$  sont disponible.

$$\hat{C} := (\hat{C}_{i,j})_{1 \leq i, j \leq I+1} \quad (4)$$

Les paires  $(i, j)$  avec  $i + j > I + 1$  se réfèrent aux années futures avec des résultats inconnus pour les montants de sinistres correspondants.

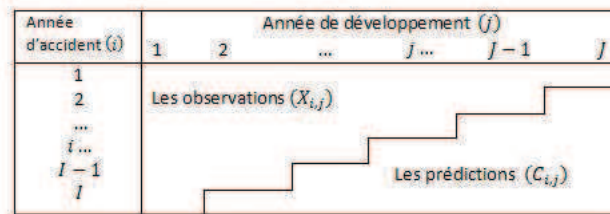


FIGURE 1 – Triangle de liquidation

**Remarque 1.** *La périodicité des données considérées est annuelle.*

### Stage pratiques et séminaires :

La thèse a été valorisée par la participation aux

- Journée jeunes chercheurs. Annaba, Algérie (Octobre 2014).
- Journées nationales sur les mathématiques appliquées "JNMA'14". Skikda, Algérie. (Novembre 2014).
- Journées Scientifiques LAROMAD'III. Tizi-Ouzou, Algérie. (Octobre 2015).
- First international conference on the "Evolution of Contemporary Mathematics and their Impact in Sciences and Technology" (ECMI6SciTech2017). Constantine, Algérie (Octobre 2017).

— Stage pratique du 09/05/2015 au 06/08/2015 dans le laboratoire de recherche Paul Painlevé à l'université de Lille1, France. Sous la direction du Professeur Nadji RAHMANIA.

# Chapitre 1

## Méthodes de provisionnement

Un rôle important, et central, de l'actuaire non-vie, est le calcul des provisions. En particulier, il existe différentes méthodes techniques de provisionnement, que nous allons rappeler dans ce chapitre. Ces provisions techniques sont destinées à permettre le règlement intégral des engagements pris envers les assurés et bénéficiaires de contrats. Elles sont liées à la techniques même de l'assurance, et imposé par la réglementation, pour tenir compte de " l'inversion du cycle de production", expliqué dans la partie introduction. Les provisions techniques mesurent ainsi les engagements restant à honorer par l'assureur, et représentent, en moyenne en assurance non-vie, 75% du total du bilan. Cette notion comptable nécessite néanmoins un modèle probabiliste sous-jacent, puisqu'il s'agit de prévoir une charge finale. La méthode Chan-Ladder est probablement la méthode la plus simple et la plus utilisée par les compagnies d'assurance au siècle dernier. Toutefois, si elle propose un estimateur du montant de provision, il est difficile d'en extraire davantage d'information. D'autres méthodes seront également proposées par la suite, introduisant une approche stochastique, permettant d'obtenir une marge d'erreur sur le montant des provisions.

## 1.1 La méthode Chain-Ladder

La méthode Chain-Ladder est une méthodes déterministe qui repose sur l'hypothèse de stabilité du délai s'écoulant entre la survenance d'un sinistre et le(s) règlement(s), quel que soit l'exercice de survenance, en absence d'inflation, de changement de structure de portefeuille, des garanties des contrats, des franchises, et plus généralement de la gestion des sinistres. Cette méthode compte parmi les plus populaires, car elle est facile à mettre en œuvre, et facile à comprendre. L'idée est que le déroulement des paiements est gouverné par des facteurs de développement strictement positifs (noté  $f_j$ ) qui ne dépendent que de l'année de développement  $j$ .

Les paramètres intervenant dans la méthode Chain-Ladder présentent l'avantage d'avoir une interprétation claire, et sont facilement estimables. L'inconvénient étant que cette estimation est relativement peu robuste. Cette méthode ne fait aucune hypothèse quant à la loi que peuvent suivre les coûts des sinistres, ou leur fréquence.

### 1.1.1 Les link-ratios

La méthode Chain-Ladder est fondée sur l'utilisation de link-ratios, encore appelés coefficients de passage, coefficients de déroulement, entre les différentes années de développement. Les hypothèses sous-jacentes sont :

- \*  $H1$  : Les paiements cumulés  $C_{i,j}$  des années de survenances  $i$  sont indépendants.
- \*  $H2$  : Les années de développement sont les variables explicatives du comportement des sinistres futurs.

La méthode standard de Chain-Ladder consiste à supposer que les  $(C_{i,j})_{j=1,\dots,J}$  sont liés par un modèle de la forme

$$C_{i,j+1} = f_j C_{i,j} \text{ pour tout } 1 \leq i \leq I \text{ et } 1 \leq j \leq J \quad (1.1)$$

Tel que pour tout  $i \in 1, \dots, I$  et  $j \in 1, \dots, J$ , on a :

$$E[C_{i,j} | C_{i,1}, \dots, C_{i,j-1}] = E[C_{i,j} | C_{i,j-1}] = f_{j-1} C_{i,j-1} \quad (1.2)$$



Les coefficients  $f_j$  de Chain-Ladder sont appelés facteurs de développement, ils peuvent être estimés, à l'aide des observations, par le rapport des totaux relatifs aux éléments communs de deux colonnes successives, c'est-à-dire  $\forall j = 0, \dots, J-1$  on a

$$\hat{f}_j = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}} = \sum_{i=0}^{I-j-1} \frac{C_{i,j}}{\sum_{k=0}^{I-j-1} C_{k,j}} \frac{C_{i,j+1}}{C_{i,j}} \quad (1.3)$$

**Lemme 1.** *Sous les hypothèses (H1) et (H2), nous avons :  $\forall 1 \leq i \leq I$*

$$E [C_{i,j}|D_I] = E [C_{i,J}|C_{i,I-i}] = C_{i,I-i} f_{I-i} \dots f_{J-1} \quad (1.4)$$

*Pour les coefficients de Chain-Ladder  $f_j$  connus, les montants de provisions pour une année d'accident  $i$  suivant les informations disponibles au moment de l'estimation des provisions  $D_I$  avec  $D_I = \{C_{i,j}|i+j \leq I+1\}$  sont prédites par*

$$E [C_{i,J}|D_I] - C_{i,I-i} = C_{i,I-i} (f_{I-i} \dots f_{J-1} - 1) \quad (1.5)$$

*Cette formule correspond aux réserves best estimate par année d'accident  $i$  en date  $I$ , conditionnellement aux observations  $D_I$  [13, 31].*

**Proposition 1.** *L'estimateur de Chain-Ladder pour  $E [C_{i,j}|D_I]$  est donné par :*

$$\forall i, j > I, \hat{C}_{i,j}^{CL} = \hat{E} [C_{i,j}|D_I] = C_{i,I-i} \hat{f}_{I-i} \dots \hat{f}_{J-1} \quad (1.6)$$

### 1.1.2 L'estimation des provisions

L'estimation des provision annuelle  $\hat{R}_i$  par année de survenance  $i$  pour  $i \in 1, \dots, I$  est définie par la relation suivante

$$\hat{R}_i = \hat{C}_{i,J} - \hat{C}_{i,J-i+1} \quad (1.7)$$

La provision totale  $\hat{R}$  à constituer pour l'ensemble des années de survenance est égale à la somme des provisions pour chaque année de survenance

$$\hat{R} = \sum_{i=1}^I \hat{R}_i \quad (1.8)$$

[34, 14, 9]

## 1.2 Le modèle de Mack

Une des finalités du provisionnement est de mesurer l'incertitude des projections de nos réserves à partir du triangle de liquidation. Les méthodes déterministes ne pouvant aboutir, il est apparu des méthodes dites stochastiques où les constituants  $C_{i,j}$  du triangle supérieur sont interprétés comme des variables aléatoires. Dès lors, il devient possible d'émettre des hypothèses sur le comportement de ces variables  $C_{i,j}$  qui seront plus ou moins validées, d'étudier la variabilité de la provision prévue par le modèle, d'obtenir des estimations et des intervalles de confiance pour des paramètres d'intérêt tels que la Value-at-Risk ou la probabilité d'insuffisance ou encore de simuler la sinistralité des exercices futurs.

La très grande simplicité du modèle de Chain-Ladder lui vaut un engouement particulier à tel point que les modèles stochastiques tentent de correspondre au mieux à ses résultats. Certains remarquent que la charge ultime de  $CL$  peut être obtenue en maximisant la vraisemblance d'un modèle de Poisson (Hachemeister et Stanard en 1975, Kremer en 1985 et Mack en 1991). D'autres ont associé la méthode de Chain-Ladder à un modèle linéaire généralisé (Renshaw en 1989 et Renshaw et Verall en 1998).

Tous les modèles sont différents mais tous aboutissent à des résultats proches de Chain-Ladder. Le modèle de Mack en 1993 donne exactement les mêmes résultats.

### 1.2.1 Description du modèle

L'idée de Mack est de considérer les lignes de règlements cumulés comme des processus markoviens vérifiant certaines propriétés. Le modèle de Mack est la version stochastique de Chain-Ladder, en effet le montant des provisions estimé est identique. Il a cependant l'avantage d'estimer une erreur de prédiction des provisions [13,34].

### 1.2.2 Lien avec la méthode de Chain-Ladder

La méthode Chain-Ladder consiste à supposer que l'on a un modèle de la forme

$$C_{i,j+1} = f_j C_{i,j} \quad \text{pour tout } 0 \leq i \leq I, 0 \leq j \leq J \quad (1.9)$$

La version stochastique de cette relation consiste à écrire

$$E[C_{i,j+1}] = f_j E[C_{i,j}] \quad \text{pour tout } 0 \leq i \leq I, 0 \leq j \leq J \quad (1.10)$$

### 1.2.3 Version stochastique de Chain-Ladder

Mack (1993) a proposé un modèle non paramétrique, conditionnel à la réalisation du triangle, permettant d'estimer les erreurs commises lors de l'évaluation des réserves. Pour cela, il suppose l'indépendance ligne par ligne, i.e.

(H1) Les paiements cumulés  $(C_{i,j})_{j=1,\dots,J}$  et  $(C_{i',j})_{j=1,\dots,J}$  sont indépendants pour  $i \neq i'$ . Et qu'il est possible de lier l'espérance conditionnelle de  $C_{i,j+1}$ , sachant le passé  $C_{i,1}, \dots, C_{i,j}$ , à la dernière observation  $C_{i,j}$ , à un facteur multiplicatif près, correspondant à un link ratio, i.e.

(H2) Les  $(C_{i,j})_{j=1,\dots,J}$  forment une chaîne de Markov et il existe des constantes  $f_j \geq 0$  et  $\sigma_j^2 \geq 0$  tel que, pour tout  $0 \leq i \leq I$ ,  $0 \leq j \leq I-1$  on a :

$$E[C_{i,j+1}|C_{i,1}, \dots, C_{i,j}] = f_j C_{i,j} \quad (1.11)$$

Et

$$Var[C_{i,j+1}|C_{i,1}, \dots, C_{i,j}] = \sigma_j^2 C_{i,j} \quad (1.12)$$

Les coefficients  $(f, \sigma^2) := (f_j, \sigma_j^2)_{j=1,\dots,I-1}$  sont appelés facteur de développement de CL et paramètres de variance respectivement [9, 17].

### 1.2.4 Propriétés des estimateurs de Chain-Ladder

La spécification d'un modèle stochastique grâce à (H1) et (H2) permet d'étudier les propriétés des estimateurs des facteurs de développement de la méthode Chain-Ladder. Mack a prouvé dans l'article [17] les résultats suivant :

**Théorème 1.** *Pour  $D_I$  l'ensemble de données observées et sous les hypothèses (H1) et (H2) nous avons*

$$E(C_{i,I}|D_I) = C_{i,I-i+1} \prod_{j=I-i+1}^{I-1} f_j \quad (1.13)$$

*Démonstration.* En utilisant l'abréviation

$$E(X) = E(X|C_{i,I} \dots C_{i,I-i+1})$$

D'après (H2) et en répétant l'application de (H1) on obtient

$$\begin{aligned} E(C_{i,I}|D) &= E(C_{i,I}) = E(E(C_{i,I}|C_{i,1} \dots C_{i,I-i+1})) \\ &= E(C_{i,I-1}|f_{I-1}) = E(C_{i,I-1})f_{I-1} = etc \\ &= E(C_{i,I-i+1})f_{I-i+1} * \dots * f_{I-1} = C_{i,I-i+1}f_{I-i+1} * \dots * f_{(I-1)} \end{aligned}$$

□

Ce théorème montre que l'estimateur  $\hat{C}_{i,I}$  a la même forme que  $E(C_{i,I}|D)$  qui est la meilleur provision de  $C_{i,j}$  basé sur les observations  $D_I$ .  
Le prochain théorème montre que l'estimation des  $f_{I-i+1} * \dots * f_{I-1}$  Par  $\hat{f}_{I-i+1} * \dots * \hat{f}_{I-1}$  est en effet une procédure raisonnable [9].

**Théorème 2.** *Les estimateurs du facteur de développement de Chain-Ladder, i.e.*

$$\hat{f}_j = \frac{\sum_{i=1}^{I-j} C_{i,j+1}}{\sum_{i=1}^{I-j} C_{i,j}} \text{ pour } j = 1, \dots, I-1 \quad (1.14)$$

*Sont sans biais et non corrélées*

*Démonstration.* Soit  $\beta_k = \{C_{i,j} | j \leq k, i + j \leq I + 1\}$ ,  $1 \leq k \leq I$ . D'après (H2) et (H1) on obtient

$$E(C_{i,k+1} | \beta_k) = E(C_{i,k+1} | C_{i,1} \dots C_{i,k}) = C_{i,k} f_k$$

Par conséquent, nous avons

$$E(\hat{f}_k | \beta_k) = \frac{\sum_{j=1}^{I-k} E(C_{i,k+1} | \beta_k)}{\sum_{j=1}^{I-k} C_{i,k}} = f_k$$

$$E(\hat{f}_k) = E(E(\hat{f}_k | \beta_k)) = f_k \quad , \quad 1 \leq k \leq I-1$$

Ce qui montre immédiatement que les paramètres  $\hat{f}_k$  sont sans biais [17].  
Montrons aussi que les  $\hat{f}_k$  sont non corrélés pour  $j < k$

$$\begin{aligned} E(\hat{f}_j \hat{f}_k) &= E(E(\hat{f}_j \hat{f}_k | \beta_k)) \\ &= E(\hat{f}_j E(\hat{f}_k | \beta_k)) \\ &= E(\hat{f}_j) f_k = E(\hat{f}_j) E(\hat{f}_k) \end{aligned}$$

Ce qui achève la démonstration □

Comme le note Mack [17], cette absence de corrélation entre les estimateurs peut paraître surprenante car les estimateurs reposent sur les même observations. Cette non corrélation est néanmoins centrale puisqu'elle permet d'écrire

$$E[\hat{f}_{I-i+1} * \dots * \hat{f}_{I-1}] = f_{I-i+1} * \dots * f_{I-1} \quad (1.15)$$

Donc, le montant des réserves annuelles estimé  $\hat{R}_i = \hat{C}_{i,I} - C_{i,I-i+1}$  est un estimateur non biaisé du vrai montant des réserves annuelles  $\hat{R}_i = C_{i,I} - C_{i,I-i+1}$ .

## 1.2.5 L'erreur quadratique moyenne de prédiction

Pour mesurer l'incertitude autour de l'estimation des réserves, nous devons définir l'erreur quadratique moyenne de prédiction  $mse(\hat{C}_{i,I})$  de l'estimateur  $\hat{C}_{i,I}$  de  $C_{i,I}$  défini par

$$\begin{aligned} mse(\hat{C}_{i,I}) &= E((C_{i,I} - \hat{C}_{i,I})^2 | D_I) \\ &= \underbrace{var(C_{i,I} | D_I)}_{\text{processus de variance}} + \underbrace{(E[C_{i,I} | D_I] - \hat{C}_{i,I})^2}_{\text{l'erreur d'estimation}} \end{aligned} \quad (1.16)$$

Cette écriture décompose la MSE en deux termes. Le premier terme représente l'erreur stochastique (processus de variance), il décrit la variation due au modèle stochastique, il est purement aléatoire et ne peut pas être supprimé. Le second terme de l'équation est l'erreur d'estimation des facteurs de développement de Chain-Ladder  $f_j$  qui traduit l'incertitude provenant de l'estimation des coefficients de Chain-Ladder  $f_j$  par  $\hat{f}_j$  et de leur espérance conditionnelle. Plus l'échantillon initial est important, moins l'erreur d'estimation sera importante. Pour pouvoir calculer la MSE nous avons besoin de la formule de la variance des  $C_{i,j}$ .

Nous pouvons introduire que  $var(C_{i,j+1} | C_{i,j}, C_{i,1}, \dots, C_{i,j})$  qui doit être inversement proportionnel ou équivalente à  $C_{i,j}$ .

$$(H3) \quad var\left(\frac{C_{i,j+1}}{C_{i,j}} | C_{i,1}, \dots, C_{i,j}\right) = C_{i,j} \sigma_j^2 \quad 1 \leq i \leq I \quad , \quad 1 \leq j \leq I-1 \quad (1.17)$$

Avec les paramètres inconnus  $\sigma_j^2$ ,  $1 \leq j \leq I-1$ . C'est l'hypothèse de la variance qui sous-tend implicitement la méthode de Chain-Ladder [17].

**Théorème 3.** *L'estimateur de  $\sigma_j^2$  pour  $1 \leq j \leq I-2$  est donné par*

$$\hat{\sigma}_j^2 = \frac{1}{I-j-1} \sum_{i=1}^{I-j} C_{i,j} \cdot \left( \frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^2 \quad (1.18)$$

$\hat{\sigma}_j^2$  est un estimateur sans biais de  $\sigma_j^2$ , avec de plus

$$\hat{\sigma}_{I-1}^2 = \min\left(\frac{\hat{\sigma}_{I-2}^4}{\hat{\sigma}_{I-3}^2}, \min(\hat{\sigma}_{I-3}^2, \hat{\sigma}_{I-2}^2)\right) \quad (1.19)$$

Maintenant, nous sommes en mesure d'énoncer le résultat principal

**Théoreme 4.** *sous les hypothèses (H1) , (H2) et (H3) , l'erreur quadratique moyenne de prédiction  $MSEP(\hat{R}_i)$  peut être estimé par*

$$M\hat{S}EP(\hat{R}_i) = \hat{C}_{i,I}^2 \sum_{j=I-i+1}^{I-i} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2} \left[ \frac{1}{\hat{C}_{i,j}} + \frac{1}{\sum_{k=1}^{I-j} C_{k,j}} \right] \quad (1.20)$$

*Cette expression est la somme de deux principaux termes, où le premier correspond au processus de variance et le second à l'erreur d'estimation [17].*

**Corolaire 1.** *Sous les hypothèses et les notations du théorème (4) l'erreur quadratique moyenne de prédiction de l'estimateur de la réserve totale  $\hat{R} = \hat{R}_2 + \dots + \hat{R}_I$  peut être estimé par*

$$M\hat{S}EP(\hat{R}) = \sum_{i=2}^I \left[ MSEP(\hat{R}_i) + 2\hat{C}_{i,I} \left( \sum_{K=i+1}^I \hat{C}_{k,I} \right) \sum_{j=I-i+1}^{I-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2 \sum_{l=1}^{I-j} \hat{C}_{l,j}} \right] \quad (1.21)$$

Ainsi, les modèles stochastiques nécessitent au préalable de poser correctement certaines définitions. La qualité des estimateurs ou des prédicteurs des réserves à l'ultime est donnée par les deux premiers moments : le coefficient de variation (rapport de l'écart-type sur la moyenne) et l'erreur moyenne quadratique de prédiction. En pratique, il n'est pas possible de déterminer analytiquement la distribution prédictive des réserves et nous devons faire appel à des algorithmes numériques de type Bootstrap pour simuler la distribution. Avec ces distributions simulées, nous sommes alors capables non seulement de déterminer les deux premiers moments des réserves mais aussi certaines caractéristiques telles les intervalles de confiance ou des mesures comme la Value-at-Risk. Dans la pratique et dans les politiques de solvabilité aujourd'hui, les estimations des premier et deuxième moments sont suffisantes pour déterminer une distribution de l'ultime [18,19].

## 1.3 La méthode bootstrap

Les modèles stochastiques comme celui de Mack estiment les moments d'ordre 1 et 2. Néanmoins, il arrive que nous ne puissions pas calculer ces moments avec une formule explicite. Et nous serions aussi intéressés par la distribution des réserves pour en déduire des caractéristiques particulières comme la value-at-Risk et ses applications au cadre Solvency 2. La méthode Bootstrap est une méthode de rééchantillonnage (avec remise) qui a été introduite par Efron en 1979 et est considérée comme une avancée spectaculaire en statistique dont le principal atout est de fournir une estimation de la distribution des provisions lorsqu'on fait intervenir les modèles linéaires généralisés (GLM). Beaucoup d'informations, comme les quantiles de la distribution, peuvent être obtenues par cette méthode à partir d'un unique échantillon, c'est-à-dire obtenir à partir d'un échantillon observé d'autres échantillons. La méthode Bootstrap, non paramétrique, a des conditions d'application bien moins strictes et nombreuses que les méthodes d'inférences statistiques classiques [10, 11].

### 1.3.1 Définition de la méthode

La méthode bootstrap est généralement associée à un modèle GLM où la distribution du terme d'erreur peut être modélisé par un modèle de poisson sur-dispersé décrit par Renshaw et Verral (1998) (qui sera détaillé dans le chapitre 2), car les réserves estimées par le modèle sur un triangle à incréments positifs et avec un lien logarithmiques, sont identiques à celles données par la méthode Chain-Ladder. Les calculs de réserves et erreur de prédiction deviennent beaucoup plus faciles, les deux autres nécessitant des ajustements compliqués.

Si les paiements incrémentaux  $X_{i,j}$  pour l'année d'accident  $i$  et l'année de développement  $j$  : variables réponses du modèle GLM, suivent une loi de poisson sur-dispersée de paramètre  $\phi$ . Autrement dit  $\phi X_{i,j}$  suit une loi de poisson de paramètre  $\mu_{i,j}$  alors

$$E(X_{i,j}) = \mu_{i,j} \quad \text{et} \quad \text{var}(X_{i,j}) = \phi E(X_{i,j}) = \phi \mu_{i,j} \quad (1.22)$$

La fonction de lien est le logarithme et le prédicteur linéaire s'écrit :  $\forall i \in \{1, \dots, I\}, j \in \{1, \dots, J\}$

$$\vartheta_{i,j} = \log(\mu_{i,j}) = c + \alpha_i + \beta_j \quad \text{avec} \quad \alpha_1 = \beta_1 = 0 \quad (1.23)$$

Le paramètre  $\phi$  est un paramètre d'échelle inconnu et estimé au cours de la procédure d'ajustement [11].

### 1.3.2 Méthodologie de la procédure Bootstrap

Les paiements incrémentaux  $X_{i,j}$  du triangle de liquidation de départ, utilisées par Bootstrap doivent être indépendantes et identiquement distribuées (*i.i.d.*) ce qui n'est souvent pas le cas. En revanche, les résidus sont plus (*i.i.d.*) que les variables originelles. Le Bootstrap sera donc appliqué sur ces derniers. Pour les modèles GLM, les résidus les plus communs sont les résidus de Déviance et les résidus de Pearson définis par :

$$r_D = \text{signe}(c - m) \sqrt{2(c \log(c|m) - c + m)} \quad (1.24)$$

$$r_p = \frac{c - m}{\sqrt{m}} \quad (1.25)$$

Après rééchantillonnage et inversion de la formule des résidus, nous obtenons un nouveau triangle de liquidation. L'inversion par résidus de Déviance n'aboutit pas à une solution analytique et les résidus de Pearson seront préférés. La démarche est donc la suivante :

1. Appliquer le modèle sur le triangle d'incrément. Comme le modèle ODP reproduit celui de Chain-Ladder, cela revient à appliquer cette dernière méthode. Déterminer alors les facteurs de développement  $\hat{f}_{j,j} \in \{1, \dots, j - 1\}$  à partir du triangle de charges cumulées, puis estimer la provision  $\hat{R}$  par la méthode Chain-Ladder décrite précédemment.
2. Estimer le triangle ajusté : calculer les valeurs prédites par le modèle  $\hat{C}_{i,j}$  pour la partie supérieure du triangle à partir de la diagonale connue, à l'aide des facteurs de développement calculés en étape 1.

$$\hat{C}_{i,j} = \frac{\hat{C}_{i,J}}{\prod_{k=j}^{J-1} \hat{f}_{i,k}} \quad \text{pour } i \in \{1, \dots, I\} \quad \text{et } j \leq I - i + 1 \quad (1.26)$$

3. Dédire du triangle précédent les paiements incrémentaux ajustés  $\hat{X}_{i,j}$  pour la partie supérieure du triangle :

$$\hat{X}_{i,j} = \hat{C}_{i,j} - \hat{C}_{i,j-1} \quad (1.27)$$

4. Calculer les résidus de Pearson qui correspondent aux résidus bruts standardisés définis par :

$$r_{i,j}^p = \frac{X_{i,j} - \hat{X}_{i,j}}{\sqrt{\hat{X}_{i,j}}} \quad (1.28)$$

L'erreur standard du bootstrap ne tient pas compte du nombre de paramètres du modèle. Ceci introduit un biais dans la comparaison des estimations analytiques et bootstrap que l'on corrige en ajustant les résidus de



Pearson. On note

$$\begin{cases} n = \frac{I(I+1)}{2} & \text{le nombre de données} \\ p = 2I - 1 & \text{le nombre de degré de libertés} \end{cases}$$

Les résidus ajustés sont alors :

$$r_{i,j}^{adj} = \sqrt{\frac{n}{n-p} r_{i,j}^p} \quad (1.29)$$

Les résidus étant en moyenne proches de zéro, l'ajustement par le nombre de paramètres augmente la variance. Le paramètre d'échelle  $\phi$  est alors estimé comme suit :

$$\hat{\phi} = \sum_{i+j \leq I+1} \left( r_{i,j}^{P_{ajus}} \right)^2 \quad (1.30)$$

5. Les étapes suivantes sont répétées  $N$  fois c'est-à-dire pour chaque simulation :

a- Effectuer un rééchantillonnage aléatoire avec remise des résidus de Pearson ajustés.  
On obtient un nouveau triangle de résidus  $\hat{r}_{i,j}^{(k)}$  pour la  $k^{ieme}$  boucle bootstrap.

b- Reconstitution du pseudo-triangle des incréments  $\hat{X}_{i,j}$  par la formule (1.29) :

$$\hat{X}_{i,j} = r_{i,j}^{adj} \cdot \sqrt{\hat{X}_{i,j}} + \hat{X}_{i,j} \quad (1.31)$$

c- Reconstitution du pseudo-triangle des règlements incrémentaux par la relation suivante :

$$\hat{X}_{i,j}^{(k)} = \hat{X}_{i,j} + \hat{r}_{i,j}^{(k)} \sqrt{\hat{X}_{i,j}} \quad (1.32)$$

d- Calcul des facteurs de développement suivant Chain-Ladder.

e- Réutilisation du modèle pour :

- i. Obtenir le triangle inférieur prévu par le modèle.
- ii. Estimer la provision globale  $\hat{R}^{(k)}$  par Chain Ladder.

f- Stocker le résultat de la provision et réitérer.

6. Obtenir une distribution empirique de l'estimation de la provision ; déterminer l'erreur de prédiction associée à cette estimation [11, 12].

## 1.4 L'approche incrémentale

### 1.4.1 Présentation de l'approche

Dans ce travail, Chorfi et Remita se sont basé uniquement sur les données des paiements incrémentaux  $D_{i,j}$  pour estimer les paramètres. Cela signifie qu'ils n'ont pas besoin de faire un pas en avant pour calculer les paiements cumulés comme dans les techniques traditionnelles de calcul des réserves de sinistres, autrement dit, ils ont montré que nous pouvons trouver les mêmes résultats des techniques traditionnelles de provisionnement basées sur le triangle de développement cumulé, directement par le triangle de développement incrémental et pour cela ils présentent l'approche comme suit.

En premier lieu, une présentation des hypothèses de l'approche en modifiant le modèle de Mack est donné, et montrer que le développement sera conditionné aux données de paiements incrémentaux, en second lieu, les formules d'estimation des paramètres seront établi, et enfin, une comparaison entre les formules de Mack et les formules données par l'approche incrémentale sera faite [8].

Le provisionnement est un problème de prédiction, conditionné par l'information disponible à l'instant  $t = I$ . Pour cela,  $D_{i,j}$  désigne l'ensemble de toutes les données disponibles à l'instant  $t = I$ , plus formellement

$$D_{i,j} = \sigma\{X_{i,j} | i + j \leq I + 1\}$$

Si nous nous concentrons sur l'ensemble des données observées jusqu'à l'année de développement  $j$ , on note

$$D_j = \sigma\{X_{i,l} | i + l \leq I + 1, l \leq j\}$$

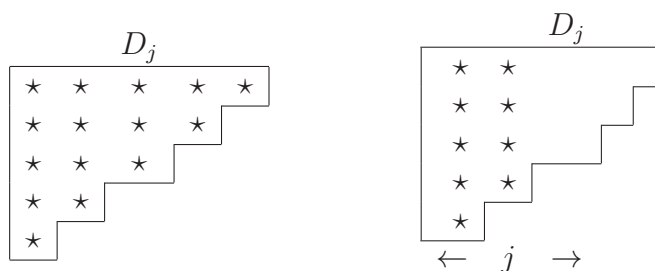


TABLE 1.1 – L'information disponible pour faire la prédiction

## 1.4.2 Hypothèses des modifications du modèle de Mack

Cette nouvelle approche est également basée sur trois hypothèses, qui sont les suivantes :

H1 : Il existe des constantes  $f_j$ , telles que pour  $1 \leq i \leq I, 1 \leq j \leq I - 1$  nous avons

$$E[X_{i,j+1}|X_{i,1}, \dots, X_{i,j}] = S_{i,j} \cdot (f_j - 1), \quad \text{avec} \quad S_{i,j} = \sum_{l=1}^j X_{i,l} \quad (1.33)$$

H2 :  $X_{i,j}$  et  $X_{k,j}$  sont stochastiquement indépendants pour  $i \neq k$ .

$$H3 : \text{var}(X_{i,j+1}|X_{i,1}, \dots, X_{i,j}) = S_{i,j} \cdot \sigma_j^2, \quad 1 \leq i \leq I \text{ et } 1 \leq j \leq J \quad (1.34)$$

## 1.4.3 Estimation des paramètres

### Les facteurs de développement

Pour chaque année de développement  $j \in \{1, \dots, J - 1\}$ , les facteurs de développement (facteurs de Chain-Ladder) sont estimé par

$$\hat{f}_j = \frac{\sum_{i=1}^{I-j} X_{i,j+1}}{\sum_{i=1}^{I-j} S_{i,j}} + 1 \quad (1.35)$$

Les vraies valeurs des facteurs de développement  $f_1, \dots, f_{J-1}$  ne peuvent pas être devinées à partir des données car l'ensemble du triangle n'est pas encore connu au moment  $t = I$ . Elles ne peuvent être estimées qu'en utilisant les montants des paiements incrémentaux  $X_{i,j}$  comme il a été montré dans la formule (1.35), une éminente propriété d'un bon estimateur est que l'estimateur doit être sans biais [8].

**Théoreme 5.** *Sous les hypothèses (H1) et (H2), les estimateurs des facteurs de développement  $\hat{f}_1, \dots, \hat{f}_{J-1}$  définie par (1.35) sont sans biais et non corrélés [8].*

*Démonstration.* Démontrons d'abord que les facteurs de développement sont sans biais i.e.  $E(\hat{f}_j) = f_j$ . En raison de la règle itérative

$$E(\hat{f}_j) = E\left(E\left(\hat{f}_j|D_j\right)\right)$$

Nous avons

$$\begin{aligned} E(X_{i,j+1}|D_j) &= E(X_{i,j+1}|X_{i,1}, \dots, X_{i,j}), \text{ par } H2 \\ &= S_{i,j}(f_j - 1), \text{ par } H1 \end{aligned}$$

Cela implique que

$$\begin{aligned} E\left(\hat{f}_j|D_j\right) &= E\left[\frac{\sum_{i=1}^{I-j} X_{i,j+1}}{\sum_{i=1}^{I-j} S_{i,j}} + 1|D_j\right] \\ &= \left[\frac{\sum_{i=1}^{I-j} E[X_{i,j+1}|D_j]}{\sum_{i=1}^{I-j} S_{i,j}} + 1\right] \quad \text{car } S_{i,j} \text{ est } D_j\text{-mesurable} \\ &= \frac{(f_j - 1) \sum_{i=1}^{I-j} S_{i,j}}{\sum_{i=1}^{I-j} S_{i,j}} + 1 = f_j \end{aligned}$$

$$E(\hat{f}_j) = E\left(E\left(\hat{f}_j|D_j\right)\right) = E(f_j) = f_j.$$

Ce qui montre que les estimateurs  $\hat{f}_j$  sont sans biais [8].

Ensuite, démontrons la non-corrélation entre les estimateurs des facteurs de développement, i.e.  $E(\hat{f}_{j_1} \cdot \hat{f}_{j_2}) = E(\hat{f}_{j_1}) \cdot E(\hat{f}_{j_2})$  pour  $j_1 < j_2$ ,

$$\begin{aligned} E\left(\hat{f}_{j_1} \cdot \hat{f}_{j_2}\right) &= E\left(E\left[\hat{f}_{j_1} \cdot \hat{f}_{j_2}|D_{j_2}\right]\right) \\ &= E\left(\hat{f}_{j_1} E\left[\hat{f}_{j_2}|D_{j_2}\right]\right) \quad \text{car } j_1 < j_2 \\ &= E\left(\hat{f}_{j_1}\right) \cdot \hat{f}_{j_2} \\ &= E\left(\hat{f}_{j_1}\right) \cdot E\left(\hat{f}_{j_2}\right) \quad \text{car } \hat{f}_{j_2} \text{ est sans biais} \end{aligned}$$

Alors les estimateurs des facteurs de développement sont non-corrélés [8]. □

## Le montant de la charge ultime

Le but de la méthode Chain-Ladder et tous les méthodes de provisionnement est l'estimation du montant de la charge ultime  $\hat{Z}_i$  pour les années d'accidents  $i = 1, \dots, I$ . Dans cette approche, Chorfi et Remita ont calculé  $\hat{Z}_i$  en appliquant deux formules, la première en utilisant la somme de tous les paiements incrémentaux

$$\begin{aligned}\hat{Z}_i &= \sum_{j=1}^{I-I+1} X_{i,j} + \sum_{I-i+2}^J \hat{X}_{i,j} \\ &= Z_i^0 + \hat{R}_i\end{aligned}\tag{1.36}$$

Où le premier terme représente la partie inconnue (l'obligation de l'année passée)  $Z_i^0$ , et le deuxième terme correspond à  $\hat{R}_i$ , ou simplement en utilisant la forme classique donnée par

$$\hat{Z}_i = Z_i^0 \left( \hat{f}_{I-i+1} \dots \hat{f}_{J-1} \right)\tag{1.37}$$

**Théorème 6.** *Sous les hypothèses H1 et H2, l'estimation du montant de la revendication ultime  $\hat{Z}_i$  définie par (1.37) sont sans biais.*

*Démonstration.*

$$\begin{aligned}E \left( \hat{Z}_i \right) &= E \left( Z_i^0 \cdot \hat{f}_{I-i+1} \dots \hat{f}_{J-1} \right) \\ &= E \left( E \left( \left[ Z_i^0 \hat{f}_{I-i+1} \dots \hat{f}_{J-1} \middle| D_{J-1} \right] \right) \right) \\ &= E \left( Z_i^0 E \left[ \hat{f}_{I-i+1} \dots \hat{f}_{J-1} \middle| D_{J-1} \right] \right) \text{ car } Z_i^0 \text{ est } D_{J-1}\text{-mesurable} \\ &= E \left( Z_i^0 \hat{f}_{I-i+1} \dots \hat{f}_{J-2} E \left[ \hat{f}_{J-1} \middle| D_{J-1} \right] \right),\end{aligned}$$

En outre  $\hat{f}_{I-i+1} \dots \hat{f}_{J-2}$  est  $D_{J-1}$ -mesurable

$$= E \left( Z_i^0 \hat{f}_{I-i+1} \dots \hat{f}_{J-2} \right) \hat{f}_{J-1}$$

Alors

$$E \left( \hat{Z}_i \right) = Z_i^0 \cdot \hat{f}_{I-i+1} \dots \hat{f}_{J-2} \cdot \hat{f}_{J-1},$$

En répétant l'opération

$$= \hat{Z}_i$$

□

## Les réserves de sinistres annuelles

L'un des avantages de l'approche incrémentale est le faite de calculer les réserves de sinistres annuelles par une simple somme des estimateurs de paiements incrémentaux pour  $i \in \{2, \dots, I\}$

$$\hat{R}_i = \sum_{j=I-i+2}^J \hat{X}_{i,j} \quad (1.38)$$

Les  $\hat{R}_i$  peuvent être obtenu aussi en utilisant le principe classique, i.e. (faire la différence entre les estimateurs du montant de la charge ultime  $\hat{Z}_i$  et les obligations passées  $Z_i^0$  qui a déjà été payé jusqu'à maintenant), et on note pour  $i \in \{2, \dots, I\}$

$$\hat{R}_i = \hat{Z}_i - Z_i^0 = Z_i^0 \cdot \left( \prod_{j=I-i+1}^{J-1} \hat{f}_j - 1 \right) \quad (1.39)$$

Et leur somme donnera

$$\begin{aligned} \hat{R} &= \sum_{i=2}^I \hat{R}_i = \sum_{i=2}^I \left[ \hat{Z}_i - Z_i^0 \right] \\ &= \sum_{i=2}^I Z_i^0 \left[ \prod_{j=I-i+1}^{J-1} \hat{f}_{j-1} \right] \end{aligned} \quad (1.40)$$

On remarque ici qu'on peut obtenir la réserve totale par une simple somme des futurs paiements des incréments dans la partie inférieure du triangle de développement.

$$\hat{R}_i = \sum_{i=2}^I \sum_{j=I-i+2}^J \hat{X}_{i,j} \quad (1.41)$$

**Théoreme 7.** *Sous les hypothèses H1 et H2, l'estimation des réserves annuelles  $\hat{R}_i$  et l'estimation de la réserve totale  $\hat{R}$  sont sans biais [8].*

*Démonstration.*

$$\begin{aligned} E(\hat{R}_i) &= E(\hat{Z}_i - Z_i^0) \\ &= E(\hat{Z}_i) - E(Z_i^0) \quad (C1) \\ &= \hat{Z}_i - Z_i^0 = \hat{R}_i \end{aligned}$$

Et

$$E(\hat{R}) = E\left(\sum_{i=2}^I \hat{R}_i\right) = \sum_{i=2}^I \hat{R}_i = R \quad (C2)$$

□

### 1.4.4 Calcul de l'erreur standard

#### L'estimation de l'erreur moyenne standard

La formule pour estimer l'erreur standard est basée sur le modèle de Mack, et dans cette approche, Chorfi et Remita [8] ont mesuré cette incertitude en utilisant uniquement les paiements incrémentaux.

$$\widehat{MSE}(\hat{R}_i) = \hat{Z}_i^2 \sum_{j=I-i+1}^{J-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2} \left( \frac{1}{\hat{S}_{i,j}} + \frac{1}{\sum_{l=1}^{I-j} S_{l,j}} \right) \quad (1.42)$$

Les paramètres  $\sigma_j^2$  sont inconnus et doivent être estimés

**Propriété 1.** *Sous H1, H2 et H3,*

$$\hat{\sigma}_j^2 = \frac{1}{I-j-1} \sum_{i=1}^{I-j} S_{i,j} \left( \frac{X_{i,j+1}}{S_{i,j}} + 1 - \hat{f}_j \right)^2$$

*Est un estimateur sans biais de  $\sigma_j^2$ .*

#### L'estimation d'erreur de la réserve totale

Il est aussi intéressant de calculer l'erreur de l'estimation de la réserve totale  $\hat{R} = \hat{R}_2 + \dots + \hat{R}_I$ . Nous ne pouvons pas faire la somme des erreurs  $MSE(\hat{R})$  car ils sont corrélés par les mêmes estimateurs  $\hat{f}_j$  et  $\hat{\sigma}_j^2$ , mais on peut utiliser le théorème suivant

**Théorème 8.** *Sous les hypothèses H1, H2 et H3,  $MSE(\hat{R})$  peut être estimé par*

$$\widehat{MSE}(\hat{R}) = \sum_{i=2}^I \left[ MSE(\hat{R}_i) + \hat{Z}_i \left( \sum_{k=i+1}^I \hat{Z}_k \right) \sum_{j=I-i+1}^{J-1} \frac{2 \frac{\hat{\sigma}_j^2}{\hat{f}_j^2}}{\sum_{k=1}^{I-j} S_{k,j}} \right] \quad (1.43)$$

## 1.5 Comparaison entre le modèle de Mack et l'approche incrémentale

A partir de l'approche de Chorfi et Remita [8] on peut remarquer qu'en utilisant uniquement les revendications incrémentaux et en procédant à partir des formules de cette approche, nous pouvons automatiquement arriver au mêmes formules que le modèle de Mack [17].

Le modèle de Mack	L'approche incrémentale
$\hat{f}_j = \frac{\sum_{i=1}^{I-j} C_{i,j+1}}{\sum_{i=1}^{I-j} C_{i,j}}$	$\hat{f}_j = \frac{\sum_{i=1}^{I-j} X_{i,j+1}}{\sum_{i=1}^{I-j} S_{i,j}} + 1$
$\hat{C}_{i,J} = C_{i,I-i+1} \prod_{j=I-i+1}^{J-1} \hat{f}_j$	$\hat{Z}_i = \sum_{j=1}^{I-I+1} X_{i,j} + \sum_{I-i+2}^J \hat{X}_{i,j}$
$\hat{R}_i = \hat{C}_{i,J} - C_{i,I-i+1}$	$\hat{R}_i = \sum_{j=I-i+2}^J \hat{X}_{i,j}$
$\hat{R} = \sum_{i=2}^I \hat{R}_i$	$\hat{R} = \sum_{i=2}^I \hat{R}_i$

TABLE 1.2 – Tableau comparatif du modèle de Mack et l'approche incrémentale



## Chapitre 2

# Les modèles factoriels stochastiques

Les modèles linéaires généralisés se sont avérés être un outil très utile dans l'assurance générale, voir par exemple Brockman et Wright (1992) et Ohlsson et Johansson (2004). Il a été largement utilisé par les actuaires comme instrument d'établissement des prix pendant de nombreuses années et, ces dernières années, il est également devenu un outil puissant dans d'autres secteurs de l'assurance générale. Les avantages des modèles linéaires généralisés sont reconnus dans de nombreux domaines de la pratique actuarielle, plusieurs articles actuariels ont été rédigés au cours des 30 dernières années décrivant l'utilisation du GLM dans la provision pour pertes, l'estimation du coût des sinistres, le coût moyen des sinistres et bien plus encore. Un certain nombre de logiciels statistiques ont été développés pour une utilisation actuarielle et comme les actuaires continuent à se familiariser avec les GLM, nous nous attendons à ce que le nombre augmente.

## 2.1 Les modèles linéaires généralisés

La classe des GLM est une extension des modèles linéaires classiques en termes de loi. Ils représentent une façon mathématique de quantifier la relation entre une variable réponse  $Y$  et un ensemble de variables indépendantes,  $y$  compris une classe générale de modèles statistiques. Le modèle linéaire généralisé développé initialement en 1972 par Nelder et Wedderburn et dont on trouvera des exposés détaillés dans Nelder et Mc Cullagh (1983), Agresti (1990) ou Antoniadis et al. (1992). L'utilisation de GLM fournit un avantage important par rapport au modèle de régression linéaire multiple. L'un des avantages est que nous allons maintenant au-delà de l'hypothèse d'une distribution normale pour les variables réponses et pouvons utiliser n'importe quel membre de la famille de distributions exponentielle. Cela inclut Poisson, Poisson sur-dispersée et gamma, Un autre avantage important est que la relation entre la variable réponse attendue ( $Y$ ) et les variables dépendantes n'a plus besoin d'être linéaire.

### 2.1.1 Introduction et motivation

#### Cadre du problème

**Exemple 1.** *Le modèle linéaire classique*

*Nous cherchons à expliquer une variable  $Y$  par  $P$  variables  $X = (X_1, \dots, X_p)$ . Par exemple une explication de la présence ou l'absence d'un cancer du poumon (variable à expliquer  $Y$ ) par l'âge, la variable fumeur / non-fumeur etc...*

*Nous voulons cependant étendre le modèle linéaire classique  $Y = X + \epsilon$  au cas où :*

- *$Y$  peut être une variable qualitative.*
- *Les erreurs peuvent ne pas avoir la même variance (= hétéroscédasticité)*

*Rappelons qu'en général dans le modèle linéaire classique, les variables  $X$  considérées comme fixes (non-aléatoires). Toute la source de variabilité constatée dans  $Y$  provient d'une erreur aléatoire  $\epsilon$  qui possède la même variance quelle que soit les observations. Le modèle linéaire classique s'écrit comme suit*

$$Y = X\beta + \epsilon = \sum_{j=1}^p X_j\beta_j + \epsilon \quad (2.1)$$

*Avec  $\epsilon \sim N(0; \sigma^2)$ . Cela veut dire que  $Y \sim N(X\beta; \sigma^2)$ . L'espérance de  $Y$  est donc  $X\beta$ .*

*Si maintenant  $X$  est un vecteur aléatoire (à  $P$  composantes) alors on voit que la variable  $Y$  dépend de  $X$  et  $\epsilon$ . Si le vecteur aléatoire prend la valeur  $x$  alors sachant*

cette valeur, la moyenne de  $Y|x\beta$  et la variance est  $\sigma^2$  (pour rester cohérent avec le modèle classique). Cela s'écrit :

$$E(Y|X = x) = x\beta \quad (2.2)$$

$$\text{Var}(Y|X = x) = \sigma^2 \quad (2.3)$$

En fait plus précisément, sachant  $X = x$ , la distribution conditionnelle est une loi normale ( $Y|X = x) \sim N(x\beta; \sigma^2)$ ). Cette présentation ne change rien mais elle va nous permettre de généraliser ce modèle linéaire aisément.

**Exemple 2.** La régression logistique

Nous souhaitons expliquer la variable  $Y$  présence / absence d'une maladie cardiovasculaire (notée aussi  $CHD$ ), par l'âge  $X$  des patients. Le tableau de données est représenté dans le tableau 2.1

La représentation des données est la suivante :

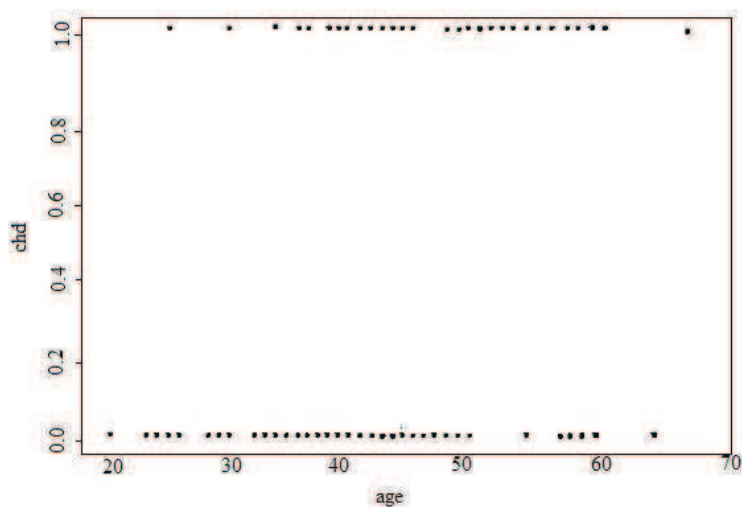


FIGURE 2.1 – Représentation directe de  $Chd$  (notée aussi  $Y$ , la variable à expliquer) en fonction de l'âge  $X$

En regroupons les patients par classe d'âge nous obtenons le tableau suivant :

CHD				
Age	n	Absent	Présent	Moyenne
19+thru 29	10	9	1	0.1
29+thru 34	15	13	2	0.133333
34+thru 39	12	9	3	0.25
39+thru 44	15	10	5	0.333333
44+thru 49	13	7	6	0.461538
49+thru 54	8	3	5	0.625
54+thru 59	17	4	13	0.764706
59+thru 69	10	2	8	0.8

TABLE 2.1 – Données regroupées en classe d'âge

La liaison entre l'âge moyen d'une classe et la proportion de malade dans cette classe apparait nettement sous la forme d'une courbe sigmoïde (i.e. en forme de "S"). Il semblerait donc "naturel" de modéliser cette proportion de malade par classe d'âge en fonction de l'âge par une courbe sigmoïde.

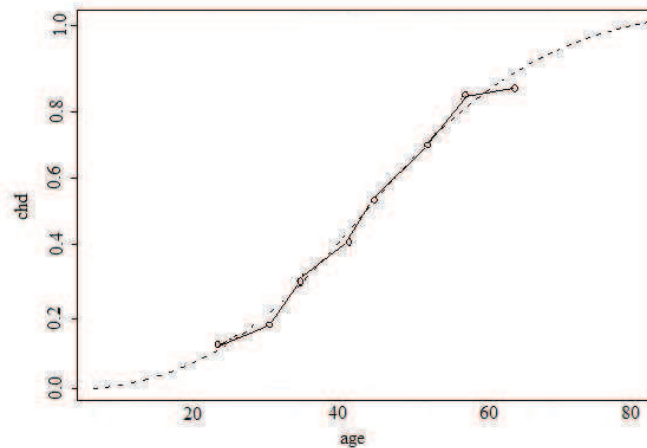


FIGURE 2.2 – Fréquence de Chd par classe d'âge en fonction de l'âge  $X$

Revenons à la terminologie statistique : l'âge noté  $X$  est une variable aléatoire et nous avons remarqué que la proportion de malade dans une classe d'âge  $X \in [x - dx; x + dx]$  suit approximativement une courbe sigmoïde. Cette proportion de malade est en fait la moyenne empirique des observations de cette classe, qui est un estimateur de l'espérance (ou moyenne) dans la "classe" d'âge  $X = x$  (si l'on fait grandir le nombre d'observation dans la classe tout en diminuant la largeur de la classe). Nous pouvons donc proposer un modèle théorique du type l'espérance conditionnelle de  $Y$  sachant  $X = x$  suit une courbe sigmoïde. Bien entendu les fonctions de forme sigmoïde sont nombreuses mais nous pouvons choisir une forme très simple, continue et dérivable de la forme

$$f(x) = h(\beta_0 + \beta_1 x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \quad (2.4)$$

Si on pose  $\eta = \beta_0 + \beta_1 x$  c'est-à-dire  $\eta$  suit un modèle de régression linéaire sans erreur, nous avons :

$$h(\eta) = \frac{\exp(\eta)}{1 + \exp(\eta)}$$

Cette fonction  $h(\cdot)$  est une fonction inversible, nous pouvons alors poser

$$g(\cdot) = h^{(-1)}(\cdot)$$

Et reprendre le modèle théorique sous deux formes

$$E(Y|X = x) = h(\beta_0 + \beta_1 x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \quad (2.5)$$

$$g[E(Y|X = x)] = \beta_0 + \beta_1 x \quad (2.6)$$

Cette dernière formulation est tout simplement la même que pour la régression linéaire (figure1) avec ici  $g(\cdot)$  n'est pas une fonction identité (i.e.  $g(\cdot) = x, \forall x$ ) mais une fonction continue dérivable et inversible.

Nous avons donc généralisé le modèle de régression simple en ajoutant une transformation de la moyenne  $g(\cdot)$ . Cette fonction est appelée "fonction de lien".

Mais notons tout de même que ce modèle n'est pas suffisant. En effet dans le modèle de régression classique nous supposons qu'il existe un bruit additif d'espérance nulle et de variance  $\sigma^2$  ce qui permet de connaître les deux premiers moments centrés de la variable à expliquer  $Y$ . En général on suppose en outre que  $\epsilon$  suit une loi normale ce qui donne la loi de  $Y \sim N(X\beta; \sigma^2)$ .

## Conclusion

- La variable aléatoire  $Y$  est qualitative.
- Nous modélisons une transformation de l'espérance conditionnelle  $E(Y|X = x)$  car  $X$  est aléatoire.
- Cette modélisation (après transformation) est linéaire en  $x$  :

$$g(Y|X = x) = \beta_0 + \beta_1 x$$

- Les erreurs ne sont pas normale mais de loi de Bernoulli.
- La variance n'est pas même et varie selon les valeurs de  $X$  [1, 15, 34]

## La modélisation GLM

La modélisation GLM reprends les principes exposés ci-dessus dans un cadre général qui englobe les 2 exemples précédents et d'autres encore. Les principales caractéristiques sont les suivantes :

- Une transformation de l'espérance conditionnelle par la fonction  $g$  de lien :

$$g(Y|X = x) = \eta(x)$$

- Cette transformation de l'espérance conditionnelle est modélisée par une combinaison linéaire des variables explicatives au point

$$X = x : \eta(x) = \sum \beta_j x_j$$

- $Y|X = x$  suit une loi de la famille exponentielle [35].

## 2.1.2 Familles exponentielles

### Densité de $y$ la variable à expliquer

**Définition 1.** Dans tous les modèles GLM nous supposons que la variable à expliquer  $Y$  possède une densité par rapport à une mesure dominante  $\nu$ . Cette densité s'écrit :

$$f_{\theta,\phi}(y) = \exp \{ y^\theta - b(\theta) | a(\phi) + c(y, \phi) \} \quad (2.7)$$

Où  $a(\cdot)$ ;  $b(\cdot)$  et  $c(\cdot)$  sont des fonctions connues et dérivables,  $b(\cdot)$  est de plus 3 fois dérivable et sa dérivée première  $b(\cdot)$  est inversible (i.e.  $\exists (b')^{-1}$ ). Les paramètres  $(\theta; \phi)$  appartiennent à  $\Omega \in \mathbb{R}^2$  et  $\phi$  est appelé paramètre de dispersion [36].

### Lien avec la famille exponentielle

- Si  $\phi$  est connu (n'est pas donc un paramètre à estimer), la densité de  $y$  est un élément de la famille exponentielle.
- Si  $\phi$  est inconnu, alors pour obtenir un élément de la famille exponentielle  $c(y, \phi)$  doit s'écrire sous la forme suivante :

$$c(y, \phi) = a(\phi)d(y) + s(a(\phi)) + t(y)$$

### Relation entre la moyenne, la variance et les paramètres $\theta$ et $\phi$

Moyenne de  $y$  : Soit  $y$  une variable de densité  $f_{\theta,\phi}$  suivant la définition (1). Puisque  $f_{\theta,\phi}$  est une densité par rapport à  $\nu$  nous avons donc

$$u(\theta, \phi) = \int_y f_{\theta,\phi} d\nu = 1, \quad \forall (\theta, \phi) \in \Omega$$

Cette fonction  $u(\cdot, \cdot)$  est une fonction constante donc dérivable et de dérivée partielle par rapport à  $\theta$  nulle. Nous avons donc grâce à la linéarité de l'intégrale et au lemme de Fatou

$$\begin{aligned} \frac{\delta u}{\delta \theta} &= \lim_{h \rightarrow 0} \frac{u(\theta + h, \phi) - u(\theta, \phi)}{h} = \lim_{h \rightarrow 0} \frac{\int_y f_{\theta+h,\phi} - f_{\theta,\phi} d\nu}{h} \\ &= \lim_{h \rightarrow 0} \inf \frac{\int_y f_{\theta+h,\phi} - f_{\theta,\phi} d\nu}{h} = \lim_{h \rightarrow 0} \sup \frac{\int_y f_{\theta+h,\phi} - f_{\theta,\phi} d\nu}{h} = 0 \\ \int_y \lim_{h \rightarrow 0} \inf \frac{\int_y f_{\theta+h,\phi} - f_{\theta,\phi} d\nu}{h} &\leq 0 \leq \int_y \lim_{h \rightarrow 0} \sup \frac{\int_y f_{\theta+h,\phi} - f_{\theta,\phi} d\nu}{h} \end{aligned}$$

Comme la dérivée partielle de  $f_{\theta,\phi}$  par rapport à  $\theta$  existe, nous avons donc

$$\int_y \frac{\delta f_{\theta,\phi}}{\delta \theta} dv = 0$$

Or si l'on calcul la dérivée partielle par rapport à  $\theta$  selon la définition 1 nous avons alors

$$\frac{\delta f_{\theta,\phi}}{\delta \theta} = \frac{y - b'(\theta)}{a(\phi)} f(y)$$

Et l'équation (2.7) devient

$$\int_y \frac{\delta f_{\theta,\phi}}{\delta \theta} dv = \frac{1}{a(\phi)} \left[ \int_y y f(y) dv - b'(\theta) \int_y f(y) dv \right] = 0$$

Nous avons  $a(\theta) \neq 0$  sinon  $f_{\theta,\phi}$  n'est pas définie alors nous en déduisons la proposition suivante [35, 36].

**Proposition 2.** *L'espérance (ou moyenne) d'une variable  $y$  dont la densité est définie par (2.7) est égale à la dérivée de  $b(\cdot)$  par rapport à  $\theta$ .*

$$E(y) = b'(\theta) \tag{2.8}$$

*Variance de  $Y$  : Si nous recommençons le même raisonnement en utilisant la dérivée partielle seconde par rapport à  $\theta$  nous avons*

$$\frac{\delta^2 u(\theta, \phi)}{\delta \theta^2} = 0 = \int_y \frac{\delta^2 f_{\theta,\phi}}{\delta \theta^2} dv$$

*Et avec*

$$\begin{aligned} \int_y \frac{\delta^2 f_{\theta,\phi}}{\delta \theta^2} &= -\frac{b''(\theta)}{a(\phi)} f_{\theta,\phi}(y) + \frac{y - b'(\theta)}{a(\phi)} \frac{\delta f_{\theta,\phi}}{\delta \theta} \\ &= -\frac{b''(\theta)}{a(\phi)} f_{\theta,\phi}(y) + \frac{1}{a(\phi)^2} \left[ (y - b'(\theta))^2 f_{\theta,\phi}(y) \right] \end{aligned}$$

*Puisque*

$$b'(\theta) = E(y)$$

*Nous obtenons alors*

$$-\frac{b''(\theta)}{a(\phi)} f_{\theta,\phi}(y) + \frac{1}{a(\phi)^2} \int_y (y - b'(\theta))^2 f_{\theta,\phi}(y) = 0$$

$$\text{var}(y) = b''(\theta) a(\phi)$$



**Proposition 3.** La variance d'une variable  $y$  dont la densité est définie par (2.7) est caractérisée par

$$\text{var}(y) = b''(\theta)a(\phi) \quad (2.9)$$

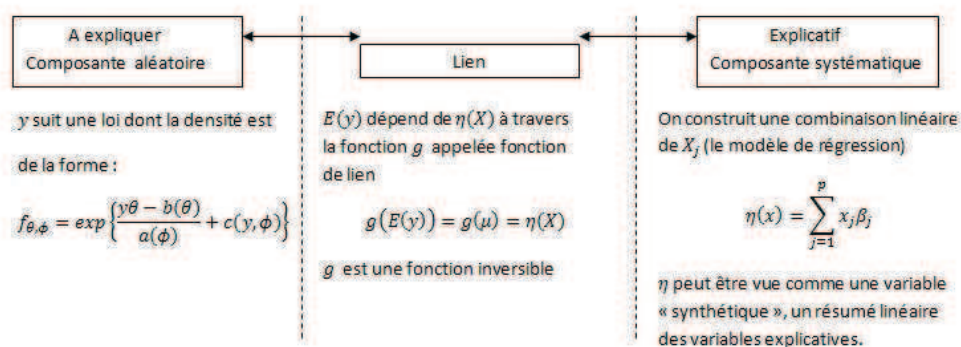
**Remarque 2.**

- Le paramètre  $\theta$  contrôle la moyenne et la variance au moyen de  $b'(\theta)$  et  $b''(\theta)$ , la moyenne est donc liée à la variance.
- Le paramètre  $\phi$  contrôle la variance via  $a(\phi)$ . Le paramètre  $\phi$  est appelé paramètre de dispersion [34, 35].

### 2.1.3 Modèle GLM

**Définition d'un modèle GLM**

Un modèle GLM est un modèle qui tente de relier des variables explicatives  $X_1, X_2, \dots, X_p$  à une variable à expliquer  $y$  selon le schéma suivant :



**Remarque 3.**

Pour choisir un modèle GLM il faut donc :

- Choisir la loi de la variable à expliquer  $y$  dans la famille exponentielle des GLM (définition 1.1), ce qui fixe  $b(\cdot)$  et  $V(\cdot)$ .
- Choisir une fonction de lien inversible  $g$ .

Pour utiliser un modèle GLM il faudra alors :

- Estimer les paramètres  $\beta = (\beta_1, \dots, \beta_p)$ , ce qui une fois réalisé, fixe donc  $\eta(x)$ , ce qui fixe  $\mu = g^{-1}(\eta(x))$ . (qui est la moyenne, donc la prévision par le modèle) et ce qui finalement fixe  $\theta = (b')^{-1}(\mu)$  car estimer  $\beta$  fixe  $\theta$ .
- Estimer le paramètre de dispersion  $\phi$  qui lui ne dépend pas de  $\beta$ .

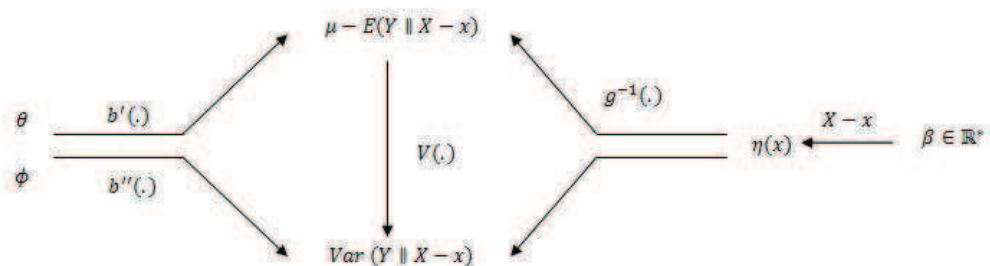


FIGURE 2.3 – Diagramme fonctionnelle d'un modèle GLM

## Fonction de lien

Une fonction quelconque ne peut pas être fonction de lien. Il faut que la fonction de lien soit inversible. Un choix particulier qui simplifie les calculs est le choix de la fonction de lien canonique, à savoir  $g(\cdot) = (b')^{-1}(\cdot)$ . ces fonctions de liens sont rappelés dans le tableau 2.2 [36].

Loi	Nom du lien	Fonction de lien
Bernouilli/Binomiale	lien logit	$g(\mu) = \text{logit}(\mu) = \log(\mu/1 - \mu)$
Poisson	lien log	$g(\mu) = \log(\mu)$
Normale	lien identité	$g(\mu) = \mu$
Gamma	lien réciproque	$g(\mu) = -1/\mu$

TABLE 2.2 – Fonction de lien canonique des modèle usuels

D'autres liens classiques existent et sont récapitulés dans le tableau 2.3

Nom du lien	Fonction de lien
lien identité	$g(\mu) = \mu$
lien log	$g(\mu) = \log(\mu)$
lien cloglog	$g(\mu) = \log(-\log(1 - \mu))$
lien logit	$g(\mu) = \log(\mu/1 - \mu)$
lien probit	$g(\mu) = \Phi^{-1}(\mu)$
lien réciproque	$g(\mu) = -1/\mu$
lien puissance	$g(\mu) = \mu^\lambda, \lambda \neq 0$ $g(\mu) = \log(\mu), \lambda = 0$
Aranda Ordaz	$g(\mu) = \log \frac{(1-\mu)^\lambda - 1}{\lambda}$

TABLE 2.3 – Fonction de lien usuelles

## Comment choisir parmi ces fonctions de lien

En fonction du type de la variable à expliquer  $y$ .

- Si  $y$  est un comptage entre  $[0, m]$  où  $m$  est fixé, dans ce cas seule la proportion est intéressante, de même si  $y$  est binaire. Dans ces cas le type de lien classique est le lien logistique mais il est possible d'utiliser le lien probit ou le lien cloglog (log-log complémentaire).

- Si les données sont des comptages, alors le lien classique est le lien log.
- Par contre si  $y$  est continue, un choix classique est le lien canonique des lois gamma ou normale.

## 2.2 Les GLM et le provisionnement

L'utilisation du GLM dans la science actuarielle est bien développée et largement acceptée. Non seulement le cadre de GLM offre une certaine souplesse dans la sélection des paramètres et des modèles, mais dans certains cas, comme dans le cas de la méthode CL, GLM récupère les méthodes traditionnelles d'estimation des réserves de sinistres. Après Renshaw et Verrall (1994), nous pouvons formuler la plupart des modèles stochastiques de réservation de revendications au moyen d'une famille particulière de modèles linéaires généralisés (voir McCullagh et Nelder (1989)). La structure de ces GLM sera donnée par :

- (i)  $Y_{i,j} \sim f(y; \mu_{i,j}, \phi)$  avec  $Y_{i,j}$  indépendants,  $\mu_{i,j} = E(Y_{i,j})$  et où  $f(\cdot)$  est la fonction de densité de  $Y_{i,j}$  appartient à la famille exponentielle,  $\phi$  est le paramètre d'échelle.
- (ii)  $\eta_{i,j} = g(\mu_{i,j})$
- (iii)  $\eta_{i,j} = c + \alpha_i + \beta_j$  avec  $\alpha_1 = \beta_1 = 0$  pour éviter la sur-paramétrisation.

Il est courant dans l'estimation des réserves de prendre en compte trois distributions possibles pour la variable  $C_{i,j}$  : Lognormal, Gamma ou Poisson. Pour les modèles basés sur les distributions Gamma ou Poisson, les relations (i) - (iii) définissent un GLM avec  $Y_{i,j} = C_{i,j}$  indiquant les montants incrémentaux des sinistres. La fonction de lien est  $\eta_{i,j} = \ln(\mu_{i,j})$  [1, 36].

### 2.2.1 Le modèle de poisson

Lors de la recherche d'un modèle stochastique qui reproduit les estimations de Chain-Ladder, certaines hypothèses doivent être formulées au sujet des réclamations d'assurance. Il est possible soit de préciser la répartition des sinistres d'assurance, soit d'indiquer simplement les deux premiers moments [13].

Il existe un large éventail de modèles de réservation stochastique et ils peuvent être divisés en «type» de Chain-Ladder et en tant qu'extensions de Chain-Ladder. Les modèles "type" de Chain-Ladder peuvent reproduire exactement les résultats de Chain-Ladder ou peuvent avoir une structure similaire à Chain-Ladder sans donner exactement les mêmes résultats. Nous considérons le modèle de Poisson qui reproduit les estimations de réserves données par la technique de Chain-Ladder [13].

(England, Verrall, 2002). Renshaw & Verrall (1998) n'ont pas été les premiers à remarquer le lien entre la technique de Chain-Ladder et la distribution de Pois-

son, mais ont été les premiers à adapter le modèle en utilisant une méthodologie standard dans la modélisation statistique. Déjà en 1975, un modèle stochastique correspondant au modèle de Poisson, qui conduit à la technique de Chain-Ladder a été découvert. Ce modèle fonctionne sur les montants incrémentaux.

$$C_{i,j} = D_{i,j} \text{ si } j = 1 \quad (2.10)$$

$$C_{i,j} = D_{i,j} - D_{i,j-1} \text{ si } j > 1 \quad (2.11)$$

## Hypothèses

Les hypothèses du modèle sont les suivantes :

- (P1)  $E(C_{i,j}) = x_i y_j$  avec  $x_i$  et  $y_j$  des paramètres inconnus.
- (P2) Chaque montant incrémental  $C_{i,j}$  a une distribution de Poisson.
- (P3) Les montants incrémentaux  $C_{i,j}$  sont indépendants [27].

Ici  $x_i$  est le montant des sinistres ultimes attendus (jusqu'à la dernière année de développement observée jusqu'ici) et  $y_j$  est la proportion des sinistres ultimes apparaissant dans chaque année de développement avec la restriction  $\sum_{k=1}^n y_k = 1$ . La restriction découle immédiatement du fait que  $y_j$  est interprété comme la proportion de sinistres déclarés dans l'année de développement  $j$ . Évidemment, la proportion globale pour toutes les périodes doit être 1.

Nous estimons les paramètres inconnus  $x_i$  et  $y_j$  du triangle des données connues (la notation  $\Delta$  est utilisée pour cela) avec la méthode du maximum de vraisemblance. La procédure d'estimation et les résultats sont donnés dans le lemme suivant [27, 28].

**Lemme 2.** *Supposons que tous les  $C_{i,j}$  sont indépendants avec une distribution de Poisson et  $E(C_{i,j}) = x_i y_j$ . Alors, les estimateurs du maximum de vraisemblance pour  $x_i$  et  $y_j$  sont donnés par :*

$$x_i = \frac{\sum_{j \in \Delta} C_{i,j}}{\sum_{j \in \Delta} y_j} \quad (2.12)$$

$$y_j = \frac{\sum_{i \in \Delta} C_{i,j}}{\sum_{i \in \Delta} x_i} \quad (2.13)$$

*Démonstration.* Nous dérivons les estimations du maximum de vraisemblance pour les paramètres inconnus  $x_i$  et  $y_j$  avec la fonction de vraisemblance

$$L = \prod_{i,j \in \Delta} \frac{(x_i y_j)^{C_{i,j}}}{C_{i,j}!} \exp(-x_i y_j).$$

Par conséquent, la fonction de log-vraisemblance est

$$l = \ln(L) = - \sum_{i,j \in \Delta} x_i y_j + \sum_{i,j \in \Delta} C_{i,j} \ln(x_i y_j) - \sum_{i,j \in \Delta} \ln(C_{i,j}!).$$

Où la somme est pour tout  $i, j$  où  $C_{i,j}$  est connu. Les estimateurs du maximum de vraisemblance sont les valeurs  $x_i, y_j$  qui maximisent  $L$  ou de manière équivalente  $\ln(L)$ . Ils sont donnés par les équations

$$0 = \frac{\delta l}{\delta x_i} = - \sum_{j \in \Delta_i} y_j + \sum_{j \in \Delta_i} C_{i,j} \frac{1}{x_i}$$

$$0 = \frac{\delta l}{\delta y_j} = - \sum_{i \in \Delta_j} x_i + \sum_{i \in \Delta_j} C_{i,j} \frac{1}{y_j}$$

Ainsi l'estimateur de vraisemblance pour  $x_i$  et  $y_j$  est donné, respectivement, par

$$x_i = \frac{\sum_{j \in \Delta} C_{i,j}}{\sum_{j \in \Delta} y_j}$$

$$y_j = \frac{\sum_{i \in \Delta} C_{i,j}}{\sum_{i \in \Delta} x_i}$$

□

Le lemme est prouvé [27].

Ainsi, les facteurs de proportion  $y_j$  expriment le rapport de la somme des valeurs incrémentales observées pour certaines années de développement  $j$  par rapport à certaines revendications finales, c'est-à-dire  $y_j$  la proportion de sinistres déclarés dans l'année de développement  $j$ . Les paramètres  $x_i$  se réfèrent au rapport de la somme des valeurs incrémentales observées pour certaines années d'origine  $i$  par rapport aux facteurs proportionnels correspondants, c'est-à-dire que si les montants incrémentaux des sinistres et les proportions respectives sont connus, il est simple de déduire l'année d'origine  $i$ . On peut noter les principales similitudes avec la technique de Chain-Ladder, où les facteurs de développement sont aussi les résultats de certains ratios.

La moyenne donnée par  $E(C_{i,j}) = x_i y_j$  dans l'hypothèse (P1) a une structure multiplicative, c'est-à-dire qu'elle est le produit de l'effet de la ligne et de l'effet de la colonne. Les effets de la ligne et de la colonne ont tous les deux des interprétations spécifiques et il est parfois utile de conserver le modèle sous cette forme.

Néanmoins, à des fins d'estimation, il est parfois préférable de reparamétriser le modèle pour que la moyenne ait une forme linéaire. Le modèle de Poisson peut être moulé sous la forme d'un GLM et linéariser le modèle multiplicatif dont nous avons besoin pour choisir le logarithme en tant que fonction de lien de sorte que

$$E(C_{i,j}) = \exp(\ln x_i + \ln y_j) \quad (2.14)$$

Ou, équivalent,

$$\ln (E(C_{i,j})) = \alpha_i + \beta_j \quad (2.15)$$

Où  $\alpha_i = \ln x_i$  et  $\beta_j = \ln y_j$  et la structure du prédicteur linéaire (2.15) est encore de type Chain-Ladder. Car les paramètres pour chaque ligne  $i$  et chaque colonne  $j$  sont donnés.

Ainsi, la structure (2.15) est définie comme un modèle linéaire généralisé, dans lequel les valeurs incrémentales  $C_{i,j}$  sont modélisées comme des variables aléatoires de Poisson avec une fonction de lien logarithmique et un prédicteur linéaire [28].

$$\eta_{i,j} = c + \alpha_i + \beta_j \quad (2.16)$$

Avec la contrainte

$$\alpha_1 = \beta_1 = 0 \quad (2.17)$$

Est nécessaire pour estimer les paramètres restants du modèle  $c, \alpha_i, \beta_j$  et pour éviter la surparamétrisation. Considérant un seul paiement incrémental  $C_{i,j}$  avec l'année d'origine  $i$  et les paiements de sinistres dans l'année de développement  $j$  (encore à observer), nous obtenons les estimations des paiements futurs à partir des estimations des paramètres en les insérant dans l'équation (2.15) et exponentielles, résultant comme

$$\bar{C}_{i,j} = \bar{x}_i \bar{y}_j = \exp(\bar{\eta}_{i,j}) \quad (2.18)$$

Compte tenu de l'équation (2.18), les estimations de la réserve pour l'année d'origine et les estimations globales peuvent être facilement obtenues par somme :

$$\bar{R}_i = \bar{x}_i \bar{y}_{n+2-i} + \dots + \bar{x}_i \bar{y}_n \quad (2.19)$$

À partir des hypothèses (P1) - (P3), l'estimateur du maximum de vraisemblance (2.19) de la réserve de sinistres pour l'année d'origine  $i$ ,  $R_i = C_{i,I+2-i} + \dots + C_{i,I} = D_{i,I} - D_{i,I+1-i}$ , donne la même prédiction  $\check{D}_{i,I} = D_{i,I+1-i} + \bar{R}_i$  que la méthode de Chain-Ladder. Selon l'hypothèse (P3),  $D_{i,I+1-i} + \bar{R}_i$  est un estimateur de l'espérance conditionnelle  $E(D_{i,I} | D_{i,I+1-i})$  et l'hypothèse (P2) contraint tous les montants incrémentaux  $C_{i,j}$  à être des entiers non négatifs [28].



Dans le modèle de Poisson, pour la provision pour perte, il est supposé que les réclamations incrémentales sont indépendantes et que Poisson est distribué avec les anticipations étant le produit de deux facteurs, selon l'année de survenance et l'année de développement, respectivement. Il est bien connu que l'estimation de la vraisemblance maximale dans le modèle de Poisson donne les estimateurs de Chain-Ladder des revendications globales ultimes attendues (Schmidt, 2002). De plus, Renshaw et Verrall (1998) ont remarqué que cela est également vrai pour les modèles de Poisson sur-dispersés.

Nous rappelons que les seules hypothèses de distribution utilisées dans les GLM sont la relation fonctionnelle entre la variance et la moyenne et le fait que la distribution appartient à la famille exponentielle. Dans le cas de Poisson, la relation mentionnée est  $V(C_{i,j}) = E(C_{i,j})$  et elle peut être généralisée à  $V(C_{i,j}) = E(C_{i,j})$  sans aucun changement de forme et de solution des équations de vraisemblance (Mack, Venter, 1999). Ce genre de généralisation permet plus de dispersion dans les données. Pour la résolution des équations de vraisemblance, il n'est pas nécessaire que les valeurs incrémentales  $C_{i,j}$  soient non-négatives ou entières et ceci conduit à un modèle de Poisson sur-dispersé et à des équations de quasi-vraisemblance, puisque la distribution sous-jacente n'est plus importante [33].

### 2.2.2 Le modèle ODP (Over-dispersed poisson)

La plupart des modèles factoriels possèdent deux paramètres. Cette flexibilité naturelle permet de capter aussi bien la moyenne que la variance des données. Il a été démontré que certaines de ces lois ont une variance qui est fonction de la moyenne.

Cependant, certaines distributions, tel que la loi de poisson ne possède pas suffisamment de paramètres pour modéliser aussi bien la variance que l'espérance puisque l'espérance est égale à la variance. On voudrait trouver une façon de rendre ces modèles plus flexibles. L'idée est d'ajouter artificiellement un paramètre supplémentaire pour forcer la loi à avoir une variance qui reflète mieux les données. Ainsi, nous définissons un paramètre de sur-dispersion, noté  $\phi$ , comme étant le facteur d'ajustement de la variance, lorsque la variance de la loi est fonction de son espérance [35].

Dans le cas de la loi Gamma, nous n'avons pas besoin d'ajouter artificiellement de nouveaux paramètres car la loi possède déjà deux paramètres. Ainsi, le paramètre de sur dispersion est introduit suite à une reparamétrisation de sorte à ce qu'on isole l'effet de la moyenne  $\lambda_{ij}$  et de la sur dispersion  $\phi$ , nous obtenons alors

un modèle où :

$$E(C_{ij}) = \lambda_{ij}, V(C_{ij}) = \phi \lambda_{ij}^2 \quad (2.20)$$

Donc  $\phi$  complète la relation où la variance est fonction du carré de l'espérance.

En ce qui concerne la loi normale, le paramètre de sur dispersion  $\phi$  est exactement le paramètre qui représente la variance de la loi, comme il est démontré dans les équations (2.21) et (2.22).

$$C_{ij} \sim N(\lambda_{ij}, \sigma^2) \quad (2.21)$$

$$\sigma^2 = \phi \quad (2.22)$$

Finalement, pour un modèle de Poisson, l'introduction d'un facteur de sur dispersion doit se faire en utilisant le principe de la quasi-vraisemblance. Pour être bref, ce principe met l'accent sur la relation entre l'espérance et la variance de la loi plutôt que sur la fonction de densité et son domaine. En d'autres mots, il suffit de trouver une statistique qui engendre la relation entre l'espérance et la variance désirée sur un domaine préétabli. Dans le cas de la loi de poisson, le nouveau modèle appelé loi de Poisson sur-dispersée et peut être défini à l'aide du raisonnement ci-dessous.

**Définition 2.**  $X_{ij}$  une variable aléatoire qui suit une loi de Poisson de moyenne et de variance toutes les deux égales à  $\mu_{ij}$  [34, 35].

$$X_{ij} \sim \text{Poisson}(\mu_{ij})$$

Et aussi soit

$$Y_{ij} \sim \text{Poisson}\left(\frac{\mu_{ij}}{\phi}\right), Y \in \{0, 1, \dots\}$$

$$Z_{ij} = \phi Y_{ij}, Z_{ij} \in \{0, \phi, 2\phi, \dots\}$$

Où  $\phi$  est le facteur de sur-dispersion désiré. Nous avons alors :

$$E(Y_{ij}) = \frac{\mu_{ij}}{\phi}, \quad \text{Var}(Y_{ij}) = \frac{\mu_{ij}}{\phi} \quad (2.23)$$

$$E(Z_{ij}) = \mu_{ij}, \quad \text{Var}(Z_{ij}) = \phi \mu_{ij} \quad (2.24)$$

La variable aléatoire possède alors une relation entre l'espérance et la variance. De plus, elle se comporte comme une loi de Poisson, mais qui est définie seulement sur les multiples du facteur de sur-dispersion, ce qui fait en sorte que pour un tel modèle, la distribution des réserves sera en escaliers [34].

## Le modèle de la loi de Poisson sur-dispersée

L'idée principale d'un modèle linéaire généralisé dans cette étude est de capturer la tendance des règlements incrémentaux au fil des années d'origine et des années de développement en supposant que les règlements suivent une loi paramétrique.

Parmi les familles de distribution possibles telles que Poisson sur-dispersée, Lognormal, Gamma etc. nous choisissons de présenter la première car la théorie prouve qu'avec un GLM cette famille reproduit exactement les mêmes prédictions que la méthode de Chain-Ladder déterministe. Avant d'entrer dans les détails, nous rappelons la notion d'une loi de poisson sur-dispersée.

### Définition 3. (Loi de poisson sur-dispersée)

Soient  $\mu, \phi$  deux réels strictement positifs. Une variable aléatoire  $X$  suit la loi de Poisson sur dispersée de paramètre  $(\mu, \phi)$  si et seulement si  $\frac{X}{\phi}$  suit une loi de Poisson de paramètre  $\frac{X}{\phi}$ . Dans ce cas nous notons  $X = ODP(\mu, \phi)$ . Avec le paramètre  $\phi$  en plus, une loi de Poisson sur-dispersée généralise une loi de Poisson habituelle. Elle permet ainsi une relation plus flexible entre la variance et l'espérance de la variable. Concrètement,  $Var(X) = \phi E(X)$ . Par ailleurs, la famille de la loi de Poisson sur-dispersée possède une propriété intéressante : elle est invariante par l'additivité montrée dans le théorème connu suivant :

**Théorème 9.** Soient  $X_1, X_2$  deux variables aléatoires telles que  $X_1 = ODP(\mu_1, \phi)$  et  $X_2 = ODP(\mu_2, \phi)$ . Alors, nous avons  $X_1 + X_2 = ODP(\mu_1 + \mu_2, \phi)$ [34].

Ces propriétés ont un rôle important dans le modèle linéaire généralisée basé sur l'hypothèse suivante :

### Hypothèses : (GLM, Loi de poisson sur-dispersée)

Soient  $(Y_{i,j})_{i,j \geq 1}$  les règlements non cumulés, alors pour tout  $i, j = 1, 2, \dots, n$  :

- Les  $(Y_{i,j})_{i,j}$  sont indépendants.
- $Y_{i,j}$  suit la loi de poisson sur-dispersée de paramètre  $(\mu_{i,j}, \phi)$  avec  $\phi > 0$ .
- Il existe des paramètres réels  $\mu, \alpha_1, \alpha_2, \dots, \alpha_n$  et  $\beta_1, \beta_2, \dots, \beta_n$  tels que :  
$$\mu_{i,j} = \exp(\mu + \alpha_i + \beta_j)$$
- $\alpha_1 = 0$  et  $\beta_1 = 0$ .

Les paramètres  $\alpha_i$  ont pour rôle de capter la tendance des règlements par rapport aux données d'origine. De même façon, les  $\beta_j$  décrivent la tendance des règlements par rapport années de développement alors que  $\mu$  permet un rôle global [34].

L'hypothèse 4 est une manière simple de faire face au problème de surparamétrisation. En effet, si nous nous n'avions pas cette hypothèse, le modèle ne serait pas identifiable car en remplaçant  $\mu$  par  $\mu - \epsilon$  et  $\alpha_i$  par  $\alpha_{i-\epsilon}$ , les  $\mu_{i,j}$  resteraient inchangés. C'est pour cette raison,  $\alpha_1$  et  $\beta_1$  ne font pas partie des paramètres à estimer dans la suite et leurs valeurs sont déjà fixées, par l'hypothèse 4, à 0.

Nous avons à notre disposition la réalisation  $y_{i,j}$  des variables aléatoires  $Y_{i,j}$  avec  $i + j \leq n$ . Cet échantillon servira à estimer le vecteur des paramètres  $\theta = (\mu, \alpha, \beta)$  où  $\alpha = (\alpha_2, \dots, \alpha_n)$  et  $\beta = (\beta_2, \dots, \beta_n)$  par la méthode de maximum de vraisemblance.

Une fois avoir obtenu l'estimateur  $\check{\theta}$  de  $\theta$ , nous remplirons les cases manquantes, c'est à dire les cases  $(i, j)$  où  $n + 1 < i, j < 2n$  par leurs valeurs espérées estimées

$$\check{E}(Y_{i,j}) = (\check{\mu} + \check{\alpha}_i + \check{\beta}_j)$$

### Estimation des paramètres

Etant donné que la variable  $Y_{i,j}$  suit la loi de Poisson sur-dispersée de paramètres  $(\mu_{i,j}, \phi)$ , la densité de probabilité de  $Y_{i,j}$  en  $y_{i,j}$  est :

$$P(Y_{i,j} = y_{i,j}) = \frac{\left(\frac{\mu_{i,j}}{\phi}\right)^{\frac{y_{i,j}}{\phi}} \exp\left(-\frac{\mu_{i,j}}{\phi}\right)}{\left(\frac{y_{i,j}}{\phi}\right)!} \quad (2.25)$$

Par ailleurs, l'indépendance des variables  $Y_{i,j}$  nous permet d'écrire le logarithme de la vraisemblance du modèle comme suit

$$\begin{aligned} l(\Delta, \theta) &= \log \prod_{i+j \leq n} \frac{\left(\frac{\mu_{i,j}}{\phi}\right)^{\frac{y_{i,j}}{\phi}}}{\left(\frac{y_{i,j}}{\phi}\right)!} \\ &= \frac{1}{\phi} \sum_{i+j \leq n} (y_{i,j}(\mu + \alpha_i + \beta_j) - \exp(\mu + \alpha_i + \beta_j)) - f(y_{i,j}, \phi) \end{aligned} \quad (2.26)$$

Avec  $f(y_{i,j}, \phi)$  qui regroupe tous les termes qui ne dépendent que de  $y_{i,j}$  et  $\phi$ . Remarquons que  $f(y_{i,j}, \phi)$  et  $\phi$  ne jouent aucun rôle dans la maximisation de la

vraisemblance par rapport au paramètre  $\theta$ . De fait que, l'estimateur de maximum de vraisemblance du paramètre  $\theta$  est le résultat du programme suivant :

$$\check{\theta} = \arg \max l(\Delta, \theta) = \sum_{i+j \leq n} (y_{i,j}(\mu + \alpha_i + \beta_j) - \exp(\mu + \alpha_i + \beta_j)) \quad (2.27)$$

Pour résoudre ce programme, nous dérivons  $l(\Delta, \theta)$  par rapport à  $\mu$ , on obtient :

$$x_{i,j} - \exp(\mu + \alpha_i + \beta_j) = 0$$

$$\ln x_{i,j} = \check{\mu} + \check{\alpha}_i + \check{\beta}_j$$

### Les hypothèses du modèle ODP

Les sinistres incrémentaux  $X_{i,j}$  sont indépendants et distribués selon la loi de Poisson sur-dispersée ; il existe  $\phi, \mu_i, \gamma_j$  tels que

$$\frac{X_{i,j}}{\phi} \sim \text{Poisson}\left(\frac{\mu_i \gamma_j}{\phi}\right)$$

Et

$$\sum_{j=0}^J \gamma_j = 1$$

Ce modèle exige que les sinistres incrémentaux  $X_{i,j}$  ne soient pas négatifs, ce qui est souvent le cas pour les paiements de sinistres, mais pas pour les pertes subies [13, 34].

### La première propriété du modèle ODP

Nous avons le modèle multiplicatif avec

$$E(X_{i,j}) = \mu_i \gamma_j \text{ et } \text{var}(X_{i,j}) = \phi \mu_i \gamma_j$$

Par ailleurs

$(\gamma_j)_{j=0, \dots, J}$  : est le modèle de développement des sinistres incrémentaux.

$\beta_j = \sum_{k=0}^j \gamma_k$  modèle de développement des sinistres cumulés [34].

## La fonction de log-vraisemblance

La fonction de log-vraisemblance, compte tenu des données  $D_I$  est :

$$l_{D_I}(\mu_i, \gamma_j) \propto \sum_{i+j \leq I} -\frac{\mu_i \gamma_j}{\phi} + \frac{X_{i,j}}{\phi} \log\left(\frac{\mu_i \gamma_j}{\phi}\right)$$

Le but de trouver les estimateurs du maximum de vraisemblance  $\check{\mu}_i$  et  $\check{\gamma}_j$  i.e. maximiser  $l_{D_I}(\mu_i, \gamma_j)$ . Résoudre pour tout  $i$  et  $j$  :

$$\frac{\delta l_{D_I}(\mu_i, \gamma_j)}{\delta \mu_i} = 0 \text{ et } \frac{\delta l_{D_I}(\mu_i, \gamma_j)}{\delta \gamma_j} = 0$$

## Estimateurs du maximum de vraisemblance du modèle ODP

Les emv  $\check{\mu}_i^{emv}$  et  $\check{\gamma}_j^{emv}$  dans le modèle ODP sont obtenus en résolvant :

$$\mu_i \sum_{j=1}^{I-i+1} \gamma_j = \sum_{j=1}^{I-i+1} X_{i,j} = C_{i,I-i} \quad (2.28)$$

$$\gamma_j \sum_{i=1}^{I-i+1} \mu_i = \sum_{i=1}^{I-i+1} X_{i,j} \quad (2.29)$$

Sous la contrainte latérale  $\sum_{j=0}^J \gamma_j = 1$  [34, 35].

**Remarque 4.** Le paramètre de dispersion (constant) s'annule en (2.28)-(2.29).

## La prédiction des sinistres incrémentaux dans le modèle ODP

Les paiements cumulés  $C_{i,J}$  données par  $D_I$  est prédits par l'emv :

$$\check{C}_{i,J} = C_{i,I-i} + \check{\mu}_i^{emv} \sum_{j=I-i+1}^J \check{\gamma}_j^{emv}$$

$$\check{X}_{i,j} = \check{\mu}_i^{emv} \check{\gamma}_j^{emv, i+j \geq I}$$

**Remarque 5.** En générale,  $\check{X}_{i,j}$  n'est pas un estimateur sans biais, i.e.

$$E(\check{X}_{i,j}) \neq E(X_{i,j}) = \mu_i \gamma_j$$

## ODP et le modèle CL (Chain-Ladder)

**Théorème 10.** *Nous avons*

$$\check{C}_{i,J}^{CL} = \check{C}_{i,J} = \check{\mu}_i^{emv} \quad (2.30)$$

*Démonstration.* Hachmeister-Stanard (1975), Kremer (1985), Mack (1991) [18]. □

Les modèles CL et ODP sont très différents d'un point de vue stochastique, mais ils mènent par hasard vers les mêmes réserves pour sinistres. La différence est dans le calcul des moments d'ordre supérieur, comme MSEP [20].

Par conséquent, le prédicteur de CL classique  $\check{C}_{i,J}^{CL}$ , peut être utilisé pour obtenir son emv.

Désormais, (2.30) fournit une solution analytique pour (2.28)-(2.29), par ailleurs

$$\beta_j^{emv} = \sum_{k=0}^j \check{\gamma}_k^{emv} = \prod_{l=j}^{J-1} \frac{1}{\check{f}_l^{CL}}$$

Où  $\check{\beta}_j^{emv}$  est l'estimateur du modèle de développement des sinistres cumulés

$$\beta_j = \sum_{k=0}^j \gamma_k$$

Voir W.Merz (2008), Corollary 2.18 [22].

### 2.2.3 Le modèle Gamma

La modélisation des montants des paiements incrémentaux en tant que variables réponse gamma indépendantes, avec une fonction de lien logarithmique et le prédicteur linéaire de type Chain-Ladder [20] produit exactement les mêmes résultats que ceux obtenus par Mack (1991). La relation entre ce modèle linéaire généralisé et le modèle proposé par Mack a d'abord été mise en évidence par Renshaw et Verrall (1994). Le même modèle peut être ajusté en utilisant le GLM décrit dans le modèle de Poisson sur-dispersé, mais dans lequel les montants incrémentaux sont modélisés comme des variables réponse Gamma indépendantes, avec une fonction de lien logarithmique et le même prédicteur, et juste en remplaçant  $Var(C_{i,j}) = \phi\mu_{i,j}$  par  $Var(C_{i,j}) = \phi\mu_{i,j}^2$ . Comme c'était le cas avec le modèle log-normal, les valeurs prédites fournies par le modèle Gamma sont généralement proches des estimations de Chain-Ladder, mais elles ne peuvent être garanties [13].

**Remarque 6.** (*England, Verrall, 1998*) Le modèle Gamma mis en œuvre sous la forme d'un modèle linéaire généralisé donne exactement les mêmes estimations de réserves que le modèle Gamma mis en œuvre par Mack (1991), ce qui est plutôt réconfortant que surprenant. Pour obtenir des prédictions et des erreurs de prédiction pour le modèle Gamma, il suffit d'un petit changement dans le modèle ODP. Le modèle Gamma est donné avec la moyenne

$$E(C_{i,j}) = \mu_{i,j} \quad (2.31)$$

Avec la variance

$$\text{Var}(C_{i,j}) = \phi(E(C_{i,j}))^2 = \phi\mu_{i,j}^2 \quad (2.32)$$

Donc la variance de ce modèle est proportionnelle à la moyenne quadratique, non proportionnelle à la moyenne comme dans le cas du modèle ODP [12].

**Remarque 7.** Nous devons imposer que chaque valeur incrémentale soit non négative si nous travaillons avec des modèles gamma (Poisson). Cette restriction peut être surmontée en utilisant une approche de quasi-vraisemblance.

En utilisant de la structure de prédicteur de type Chain-Ladder

$$\eta_{i,j} = c + \alpha_i + \beta_j, \alpha_1 = \beta_1 = 0$$

$$\log(\mu_{i,j}) = \eta_{i,j}$$

Il est facile d'obtenir des estimations de paramètres et des valeurs prédites en utilisant les GLM [34, 18].



## Chapitre 3

# Sur l'approche géométrique de la méthode Chain-Ladder : Propriétés et Application

Ce chapitre, avec quelques développements et modifications est le texte d'un article apparu dans le journal "International Journal of Applied mathematics & Statistics " dans lequel un aspect géométrique de la théorie standard des estimations des réserves en utilisant les paiements incrémentaux est donné. De plus, les facteurs de développement ( $k_j$ ) sont estimés par une technique graphique, il est donc nécessaire de calculer les tangentes pour chaque année de développement  $j$ , puis les cercles inférieurs sont estimés en utilisant des cercles et des tangentes observés. Cela conduit à des estimations annuelles et des provisions totales. Afin de valider l'approche proposée, une application numérique est effectuée, quelques exemples sont donnés pour illustrer et montrer l'efficacité de l'approche proposée.

Enfin, une application et une analyse critique des différentes méthodes de provisionnement non-vie qui ont été évoquer dans les chapitres précédent sont proposé.

## 3.1 Approche graphique pour le calcul des réserves

### 3.1.1 Théorie et développement du modèle

Le développement théorique du modèle proposé est basé sur les suppositions suivantes :

Pour certains  $I \geq 2$ , la variable  $X$  est interprétée comme le montant des paiements incrémentaux des années d'accident  $i$  et des années de développement  $j$ . Pour plus de commodité, nous définissons  $X = 0$ , et nous supposons que  $I = J$ .

Dans la théorie standard de la revendication des réserves, également appelée la théorie IBNR, on suppose que les paires  $(i, j)$  des années d'accident et de développement avec  $i + j \leq I + 1$  décrit les dernières années pour lesquelles le « Triangle » run-off des observations réelles  $X \in R$  sont disponibles :

$$\widehat{X} := \left( \widehat{X}_{i,j} \right)_{(i,j) \in \mathfrak{S}} \quad (3.1)$$

Où  $\mathfrak{S} := \{(i, j) \mid 1 \leq i \leq I, 1 \leq j \leq I - i + 1\}$  est l'ensemble d'index du «triangle». Les paires  $(i, j)$  avec  $i + j > I + 1$  se réfère aux années futures avec des résultats inconnus pour les montants de réclamation correspondants. C'est la tâche de la théorie IBNR de fournir des estimations pour les sinistres à payer, c'est-à-dire des estimateurs pour la provision totale aléatoire.

Dans le modèle proposée, la variable du montant des paiements incrémentaux  $X_{i,j}$  par la surface des cercles  $S_{i,j}$  avec :

$$S_{i,j} = \pi R_S^2, \quad (3.2)$$

$$\sum_{i=1}^{I-j} S_{i,j+1} = \pi R_{\Sigma S}^2 \quad (3.3)$$

Les montants des paiements cumulées  $C_{i,j}$  sont représentés par la somme des cercles  $D_{i,j}$  avec :

$$D_{i,j} = \pi R_D^2, \quad (3.4)$$

$$\sum_{i=1}^{I-j} D_{i,j} = \pi R_{\Sigma D}^2 \quad (3.5)$$

## Hypothèses

L'approche développée est basé sur deux hypothèses données comme suit :

**H1** : Pour chaque  $1 \leq i \leq I$ ,  $1 \leq j \leq I - 1$  Il existe des constantes  $k_j$  avec :

$$\tilde{S}_{i,j+1} = \sum_{l=1}^j S_{i,l} \cdot \tilde{k}_j$$

**H2** : : Les surfaces  $\{S_{i,1}, \dots, S_{i,I}\}$  et  $\{S_{k,1}, \dots, S_{k,I}\}$  des différentes années d'accident  $i \neq k$  sont indépendant.

### 3.1.2 Facteurs de développement graphique :

Pour chaque année de développement  $j \in \{1, \dots, I - 1\}$  les coefficients  $(k) := (k_j)$  sont appelés **les facteurs de développement graphiques**, i.e. **FDG** et sont estimés par :

$$\tilde{k}_j = \left(1 - \frac{H \cdot \tan \alpha}{R_{\Sigma D}}\right)^2 \quad (3.6)$$

Nous supposons que  $H$  une unité de mesure, où  $H$  est la distance entre les deux cercles  $R_{\Sigma S}$  et  $R_{\Sigma D}$ .

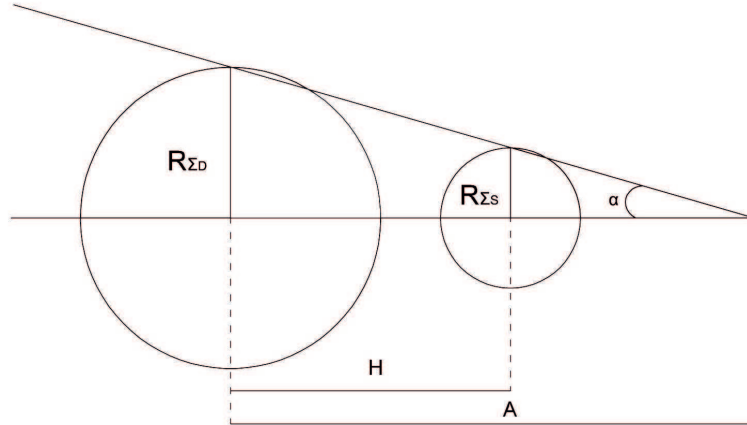


FIGURE 3.1 – La représentation du facteur de développement graphique

Nous ne pouvons pas deviner les vraies valeurs des facteurs de développement graphique  $k_1, \dots, k_{I-1}$  car les surfaces des cercles ne sont pas encore connues à  $t = I$ . Ils ne peuvent être estimés qu'en utilisant les surfaces des paiements incrémentaux  $S_{i,j}$  comme nous l'avons montré dans la formule 3.6.

*Démonstration.* Preuve : Nous démontrons comment nous avons obtenu les FDG,

$$\tan \alpha = \frac{R_{\Sigma D}}{A} = \frac{R_{\Sigma S}}{A - H}$$

En utilisant 7, 3 et 5, nous obtenons

$$\tilde{k} = \left( \frac{A - H}{A} \right)^2$$

En remplaçant  $A$  dans la dernière équation nous obtenons

$$\tilde{k} = \left( \frac{\frac{R_{\Sigma D}}{\tan \alpha} - H}{\frac{R_{\Sigma D}}{\tan \alpha}} \right)^2 = \left( 1 - \frac{H \cdot \tan \alpha}{R_{\Sigma D}} \right)^2$$

□

Nous exposons ici une autre méthode pour calculer les facteurs de développement graphique **GDF**.

$$\tilde{k}_j = \frac{\sum_{i=1}^{I-j} S_{i,j+1}}{\sum_{i=1}^{I-j} D_{i,j}} \quad (3.7)$$

Une propriété éminente d'un bon estimateur est que l'estimateur doit être sans biais.

**Théoreme 11.** *Sous les hypothèses  $H_1$  et  $H_2$ , les estimateurs des facteurs de développement graphique  $\tilde{k}_1, \dots, \tilde{k}_{I-1}$  défini par 7 sont sans biais et non corrélés.*

*Démonstration.* D'abord, nous démontrons que les facteurs de développement graphique sont  $\mathbb{E}[\tilde{k}_j] = k_j$  non biaisés i.e.  $\mathcal{F}_j := \{x_{i,l} | i + l \leq I + 1, l \leq j\}$  La signification mathématique de la première hypothèse  $H_1$  est donnée par,

$$\mathbb{E}(S_{i,j+1} | S_{i,1}, \dots, S_{i,j}) = D_{i,j} \cdot k_j \quad , 1 \leq i \leq I, 1 \leq j \leq I - 1$$

En raison de la règle itérative des attentes,

$$\mathbb{E}[\tilde{k}_j] = \mathbb{E}[\mathbb{E}[\tilde{k}_j | \mathcal{F}_j]]$$

Nous avons

$$\begin{aligned}\mathbb{E}[S_{i,j+1} \mid \mathcal{F}_j] &= \mathbb{E}[S_{i,j+1} \mid S_{i,1}, \dots, S_{i,j}] \quad , \text{ by } H_2 \\ &= \sum_{l=1}^j S_{i,l} \cdot k_j = D_{i,j} \cdot k_j \quad , \text{ by } H_1\end{aligned}$$

Cela implique que

$$\begin{aligned}\mathbb{E}[\tilde{k}_j \mid \mathcal{F}_j] &= \mathbb{E}\left[\frac{\sum_{i=1}^{I-j} S_{i,j+1}}{\sum_{i=1}^{I-j} D_{i,j}} \mid \mathcal{F}_j\right] \\ &= \frac{\sum_{i=1}^{I-j} \mathbb{E}[S_{i,j+1} \mid \mathcal{F}_j]}{\sum_{i=1}^{I-j} D_{i,j}} \quad , \text{ car } D_{i,j} \text{ is } \mathcal{F}_j\text{-measurable} \\ &= \frac{k_j \cdot \sum_{i=1}^{I-j} D_{i,j}}{\sum_{i=1}^{I-j} D_{i,j}} = k_j\end{aligned}$$

$\mathbb{E}[\tilde{k}_j] = \mathbb{E}[k_j]$  Ce qui montre que les estimateurs  $\tilde{k}_j$  ne sont pas biaisés.

Nous passons maintenant à la non-corrélation entre les estimateurs des facteurs de développement,  $\mathbb{E}[\tilde{k}_{j_1} \cdot \tilde{k}_{j_2}] = \mathbb{E}[\tilde{k}_{j_1}] \cdot \mathbb{E}[\tilde{k}_{j_2}]$  for  $j_1 < j_2$ .

$$\begin{aligned}\mathbb{E}[\tilde{k}_{j_1} \cdot \tilde{k}_{j_2}] &= \mathbb{E}\left[\mathbb{E}[\tilde{k}_{j_1} \cdot \tilde{k}_{j_2} \mid \mathcal{F}_{j_2}]\right] \\ &= \mathbb{E}\left[\tilde{k}_{j_1} \cdot \mathbb{E}[\tilde{k}_{j_2} \mid \mathcal{F}_{j_2}]\right] \quad \text{car } j_1 < j_2 \\ &= \mathbb{E}[\tilde{k}_{j_1}] \cdot k_{j_2} \\ &= \mathbb{E}[\tilde{k}_{j_1}] \cdot \mathbb{E}[\tilde{k}_{j_2}] \quad , \text{ car } \tilde{k}_j \text{ ne sont pas biaisés}\end{aligned}$$

□

### 3.1.3 Réserves de sinistres

L'objectif de la méthode de réservation des sinistres est l'estimation des réserves pour cela nous devons estimer les surfaces des cercles futurs, en utilisant l'hypothèse  $H1$ , Pour tout  $2 \leq i \leq I$ ,  $1 \leq j \leq I - 1$

$$\tilde{S}_{i,j+1} = \sum_{l=1}^j S_{i,l} \cdot \tilde{k}_j$$

En additionnant la surface des cercles de chaque ligne, nous obtenons les réserves de sinistres annuelles de l'année d'accident pour  $i = 2, \dots, I$

$$\tilde{R}_i = \sum_{j=I-i+2}^I \tilde{S}_{i,j} \quad (3.8)$$

Et l'ensemble des surfaces estimées donne de la réserve totale :

$$\tilde{R} = \sum_{i=2}^I \tilde{R}_i \quad (3.9)$$

### 3.1.4 Illustration numérique

Pour notre exemple numérique, nous utilisons l'ensemble de données dans la figure 2. Le tableau contient des paiements incrémentaux pour les années d'accident  $i \in \{1, \dots, 6\}$

Nous appliquerons notre approche graphique pour les réclamations de réserves sur les données de paiements incrémentaux  $X_{i,j}$ .

Accident year i	Development year j					
	1	2	3	4	5	6
1	5947	3721,2	895,7	207,8	206,7	62,1
2	6346,8	3246,4	723,2	151,8	67,8	
3	6269,1	2976,2	847,1	262,8		
4	5863	2683,2	722,5			
5	5778,9	2745,2				
6	6184,8					

FIGURE 3.2 – Triangle des paiements incrémentaux, en Euro 1000

Tout d'abord, nous déduisons le rayon des cercles  $S_{i,j}$  de chaque paiement incrémental  $X_{i,j}$ , puis nous les représentons comme dans la figure 3.3.

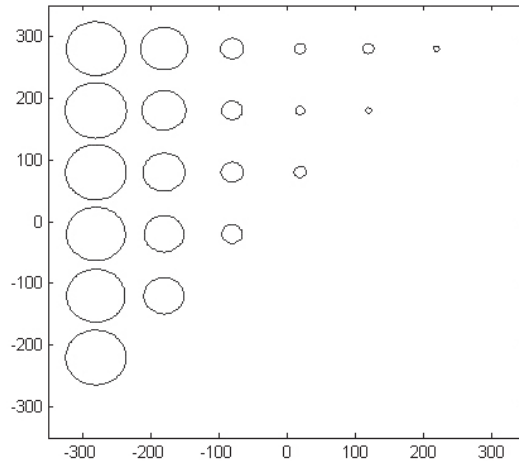


FIGURE 3.3 – Représentation des surfaces des incréments par des cercles.

En utilisant la formule 3.6, nous pouvons estimer les facteurs de développement graphique GDF pour chaque  $j \in \{1, \dots, I - 1\}$ , les rayons des cercles et la distance  $A$  nous permettent de calculer les tangentes comme nous l'avons montré dans la figure 3.1.

$\tilde{k}_j$	0,50893	0,08605	0,0201	0,01292	0,00566
---------------	---------	---------	--------	---------	---------

TABLE 3.1 – Estimation des facteurs de développement graphique.

Ensuite, nous estimons les cercles inférieurs en utilisant la première hypothèse  $H1$  et enfin nous calculons les réserves comme surfaces des cercles estimés, les résultats sont donnés dans la figure 3.4.

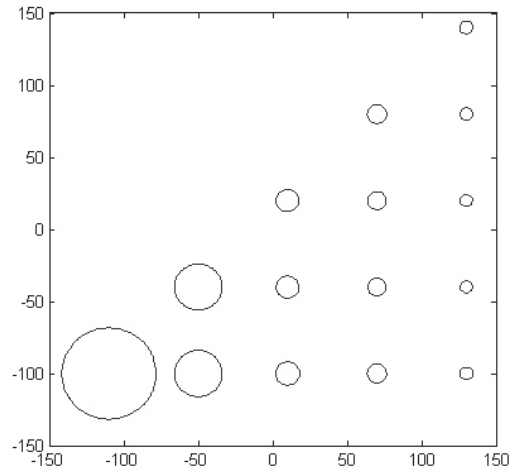


FIGURE 3.4 – Représentation des surfaces des paiements futurs par des cercles.

### Résultats et discussions

Les résultats obtenus sont exactement les mêmes que ceux calculés par la méthode de Chain-Ladder [17] et l'approche incrémentale stochastique [8]. Les réserves estimées par l'approche graphique [4] donnent pour la première fois un aspect géométrique à la théorie standard de provisionnement.

Le tableau ci-dessous compare les résultats de l'approche proposé (modèle graphique) [4], le modèle de Mack [17] et l'approche incrémentale stochastique [8].



i	Modèle de Mack	Approache Graphique	Approache Stochastique incrémentale
1	0	0	0
2	59.6	59.6	59.6
3	193.2	193.2	193.2
4	362.6	362.6	362.6
5	1095.7	1095.7	1095.7
6	4347.3	4347.3	4347.3
Total	6058.4	6058.4	6058.4

TABLE 3.2 – Comparaison des provisions des différents modèles

## Conclusion

Ce travail est axé sur un aspect géométrique de la théorie standard de provisionnement en utilisant des paiements incrémentaux. Une approche graphique est proposée pour le calcul des provisions pour sinistres en assurance non-vie, appelée approche graphique pour calcul des réserves de sinistres en utilisant des règlements incrémentaux. Une application numérique est effectuée et les résultats obtenus sont en accord avec les résultats de la méthode de Chain-Ldder [17] et l'approche incrémentale stochastique [8]. Les recherches actuelles dans le domaine des provisions techniques reposent principalement sur la théorie ; Cependant, dans ce travail, une approche graphique est développée.

## 3.2 Application et analyse critique

Cette partie est consacrée pour une application numérique et analyse critique des différentes méthodes de provisionnement en assurance non-vie dans le cadre de Solvabilité 2 et qui sont évoquées précédemment.

Notre étude ayant pour but l'estimation des paiements futurs concernant les sinistres survenus avant l'année  $I$ , c'est-à-dire la partie supérieure du tableau. Il existe de nombreuses méthodes permettant d'estimer le triangle inférieur. Nous avons retenus les méthodes classiques qui sont le modèle de Mack [17, 19] qui est la version stochastique de la méthode de Chain Ladder [9, 18] et l'approche incrémentale [8] détaillés dans le chapitre 1, ainsi que les modèles linéaire généralisée évoqué dans le chapitre 2 et l'approche graphique proposée dans le chapitre 3.

Nous analysons l'ensemble de données standard de Wüthrich, Merz [34], voir le tableau 3 ci-dessous. Le tableau contient des paiements incrémentaux pour les années d'accident  $i \in \{0, \dots, 9\}$ , d'abord nous appliquerons le modèle de Mack pour estimer les provisions de sinistres qui sera un axe de comparaison pour l'approche incrémentale et l'approche graphique ainsi que la distribution de Poisson sur-dispersée et Gamma par la suite.

$i/j$	0	1	2	3	4	5	6	7	8	9
0	5947,1	3721,2	895,7	207,8	206,7	62,1	65,8	14,8	11,1	15,8
1	6346,8	3246,4	723,2	151,8	67,8	36,6	52,8	11,2	11,6	
2	6269,1	2976,2	847,1	262,8	152,7	65,4	53,5	8,9		
3	5863	2683,2	722,5	190,7	133	88,3	43,3			
4	5778,9	2745,2	653,9	273,4	230,3	105,2				
5	6184,8	2828,3	572,8	244,9	105					
6	5600,2	2893,2	563,1	225,5						
7	5288,1	2440,1	528							
8	5290,8	2357,9								
9	5675,6									

FIGURE 3.5 – les paiements incrémentaux observées  $X_{i,j}$  en Euro (1000).

D'abord, nous utilisons la formule (1.11) pour avoir les estimations des paiements pour les années futures, les résultats sont donnés dans le tableau 3.6.

i/j	0	1	2	3	4	5	6	7	8	9
0										
1										10663,31
2									10646,79	10661,9
3								9734,48	9744,64	9758,47
4							9837,25	9847,86	9858,13	9872,12
5						10005,06	10056,53	10067,38	10077,88	10092,19
6					9419,76	9485,42	9534,22	9544,51	9554,46	9568,03
7				8445,05	8570,39	8630,14	8674,54	8683,9	8692,96	8705,29
8			8243,46	8432,02	8557,17	8616,82	8661,15	8670,5	8679,54	8691,86
9		8471	9129,7	9338,54	9477,14	9543,21	9592,3	9602,65	9612,67	9626,31

FIGURE 3.6 – Les estimations des paiements pour les années futures.

Les résultats de l'estimation des provisions des sinistres par la méthode de Mack sont donnés dans le tableau 3.7.

i	Années de développement										Ri
	0	1	2	3	4	5	6	7	8	9	
0	5947	9668,2	10563,9	10771,7	10978,4	11040,5	11106,3	11121,1	11132,2	11148	0
1	6346,8	9593,2	10316,4	10468,2	10536	10572,6	10625,4	10636,6	10648,2		15,1
2	6269,1	9245,3	10092,4	10355,2	10507,9	10573,3	10626,8	10635,7			26,3
3	5863	8546,2	9268,7	9459,4	9592,4	9680,7	9724				34,5
4	5778,9	8524,1	9178	9451,4	9681,7	9786,9					85,3
5	6184,8	9013,1	9585,9	9830,8	9935,8						156,5
6	5600,2	8493,4	9056,5	9282							286,1
7	5288,1	7728,2	8256,2								449,2
8	5290,8	7648,7									1043,2
9	5675,6										3950,8
fj	1,6257	1,0926	1,0197	1,0192	1,0057	1,006	1,0013	1,001	1,0014		6047,1

FIGURE 3.7 – Les réserves calculées par le modèle de Mack [17] en Euro (1000).

Estimer les réserves par l'approche incrémentale [8] nous permet d'avoir le tableau suivant :

i/j	0	1	2	3	4	5	6	7	8	9	Ri
0	5947,1	3721,2	895,7	207,8	206,7	62,1	65,8	14,8	11,1	15,8	0
1	6346,8	3246,4	723,2	151,8	67,8	36,6	52,8	11,2	11,6		15,1
2	6269,1	2976,2	847,1	262,8	152,7	65,4	53,5	8,9			26,3
3	5863	2683,2	722,5	190,7	133	88,3	43,3				34,5
4	5778,9	2745,2	653,9	273,4	230,3	105,2					85,3
5	6184,8	2828,3	572,8	244,9	105						156,5
6	5600,2	2893,2	563,1	225,5							286,1
7	5288,1	2440,1	528								449,2
8	5290,8	2357,9									1043,2
9	5675,6										3950,8
fj	1,6257	1,0926	1,0197	1,0192	1,0057	1,006	1,0013	1,001	1,0014		6047,1

FIGURE 3.8 – Triangle des paiements incrémentaux en Euro (1000).

L'estimation des facteurs de développement en utilisant la formule (3.7) permet d'avoir le tableau suivant :

$\tilde{k}_j$	0,6257	0,0926	0,0197	0,0192	0,0057	0,006	0,0013	0,001	0,0014

TABLE 3.3 – Les estimations des facteurs de développement graphique [4].

Après la représentation graphique des surfaces des paiements incrémentaux par des cercles dans le logiciel statistique Matlab, et en utilisant la formule (3.6) pour avoir les surfaces des paiements futures, et en appliquant la formule (3.9) pour obtenir les réserves des sinistres annuelles et enfin la formule (3.10) nous permet d'avoir la réserve totale, les résultats sont mentionnés dans le tableau suivant :

$i/i$	0	1	2	3	4	5	6	7	8	9	$\hat{R}$
$\hat{R}_i^{Graphique}$	0	15.1	26.3	34.5	85.3	156.5	286.1	449.2	1043.2	3950.8	6047.1

TABLE 3.4 – les réserves de sinistres par l’approche graphique [4]

**Proposition 4.** *Comme noté auparavant, la fonction de densité (2.7) détermine à quelle famille la distribution exponentielle appartient. De même on peut spécifier la structure de la fonction de variance unitaire sous-jacente,  $V(\cdot)$ , définie comme :*

$$V(\mu) = b''((b')^{-1}(\mu))$$

*Nous nous concentrons sur les fonctions de variance unitaire d’une variété puissante, à savoir  $V(\mu) = \mu^p$  pour  $p \in (-\infty, 0] \cup [1, \infty)$ . Les valeurs spécifiques de  $p$  correspondent à des distributions spécifiques, par exemple quand  $p = 0, 1, 2, 3$  on retrouve les distributions Gaussienne, Poisson, Gamma, Gaussienne inverse respectivement [1].*

*Nous centrons nos exemples autour de  $p = 1$  et  $p = 2$ , correspondant au Poisson sur-dispersé et la distribution de Gamma. Les réserves de sinistres pour ces modèles sont facilement accessibles avec la plupart des logiciels statistiques standards.*

### Estimer avec $p=1$ la distribution de Poisson sur-dispersée

En utilisant le logiciel statistique  $R$  nous obtenons les estimateurs de maximum de vraisemblance EMV des paramètres sous-jacent  $\mu_i$  et  $\gamma_j$  de la distribution de Poisson sur-dispersée.

**Remarque 8.** *Notez que pour  $p = 1$ , les estimations des paramètres sous-jacents peuvent être obtenues en utilisant la méthode de Chain-Ladder classique; voir le corolaire 2.18 de Wuthrich et Merz [34].*

$i/j$	$\hat{\mu}_i$	$\hat{\gamma}_i$
0	1.000	6572,76
1	0.957	3237,32
2	0.956	762,83
3	0.875	241,836
4	0.886	160,501
5	0.905	76,540
6	0.858	56,870
7	0.781	12,002
8	0.780	11,641
9	0.863	15,813

TABLE 3.5 – Les estimations des paramètres  $\mu_i$  et  $\lambda_j$  pour le cas  $p=1$

Les réserves de sinistres annuelles et totales obtenues sous les hypothèses de la distribution de Poisson sur-dispersée sont mentionnées dans le tableau 3.6.

$i/i$	0	1	2	3	4	5	6	7	8	9	$\hat{R}$
$\hat{R}_i^{ODP}$	0	15.1	26.3	34.5	85.3	156.5	286.1	449.2	1043.2	3950.8	6047.1

TABLE 3.6 – Les réserves de sinistres annuelles et totales du modèle ODP.

### Estimer avec $p=2$ ; la distribution de Gamma :

Les estimateurs de maximum de vraisemblance des paramètres sous-jacent de la distribution de Gamma sont présentés dans le Tableau 3.7.

$i/j$	$\mu_i^{\hat{e}mv}$	$\gamma_i^{\hat{e}mv}$
0	1	6999.574
1	0.760	3426.601
2	0.900	800.954
3	0.850	252.086
4	1.052	161.788
5	0.809	77.394
6	0.811	61.418
7	0.709	13.159
8	0.722	13.226
9	0.811	15.813

TABLE 3.7 – Les estimations des paramètres  $\mu_i$  et  $\lambda_j$  de la distribution de Gamma.

Les réserves de sinistres annuelles et totales obtenues par la distribution de Gamma sont présentées dans le Tableau 3.8.

$i/i$	0	1	2	3	4	5	6	7	8	9	$\hat{R}$
$\hat{R}_i^{Gamma}$	0.00	13.1	24.5	31.6	73.2	130.4	274.3	430	1013.1	3956.9	5947.1

TABLE 3.8 – les réserves de sinistres annuelles et totales du modèle Gamma.

Pour l'évaluation des provisions techniques annuelles et totales, les différents modèles donnent les résultats suivants :

AA/AD	Distributions non paramétriques			Distributions paramétriques	
$i/j$	$\hat{R}_i^{Mack}$	$\hat{R}_i^{Incrementale}$	$\hat{R}_i^{Graphique}$	$\hat{R}_i^{ODP}$	$\hat{R}_i^{Gamma}$
0	0.00	0.00	0.00	0.00	0.00
1	15.1	15.1	15.1	15.1	13.1
2	26.3	26.3	26.3	26.3	24.5
3	34.5	34.5	34.5	34.5	31.6
4	85.3	85.3	85.3	85.3	73.2
5	156.5	156.5	156.5	156.5	130.4
6	286.1	286.1	286.1	286.1	274.3
7	449.2	449.2	449.2	449.2	430
8	1043.2	1043.2	1043.2	1043.2	1013.1
9	3950.8	3950.8	3950.8	3950.8	3956.9
$\hat{R}$	6047.1	6047.1	6047.1	6047.1	5947.1

### **Analyse des résultats :**

D'après les résultats de calcul des provisions de sinistres incrémentaux étalés sur la période de dix ans on constate une coïncidence entre les réserves annuelles et totales obtenus par la méthode de Mack, l'approche incrémentale et l'approche graphique et la distribution de Poisson sur-dispersée malgré la différence des hypothèses des modèles, ainsi que le modèle de Poisson sur-dispersée qui s'appuie sur les paramètres contrairement au reste des méthodes

Les résultats obtenus par l'approche incrémentale sont identique à ceux calculés par le modèle de Mack en utilisant les paiements cumulés, alors nous pouvons calculer les réserves simplement avec les paiements incrémentaux sans avoir besoin des paiements cumulés, cela nous permet d'éviter une étape de calcul ce qui nous fait gagner du temps.

Les résultats obtenus par l'approche graphique sont exactement les mêmes que ceux calculés par le modèle de Mack et l'approche incrémentale stochastique. Les réserves estimées par l'approche graphique donnent pour la première fois un aspect géométrique à la théorie standard de provisionnement et nous pouvons avoir une image clair des informations que nous avons dans le triangle.

L'exactitude de réserves obtenues par le modèle de poisson sur-dispersée malgré la différence des modèles donnent dans un premier temps une diversité dans le choix des modèles.

La légère différence issue du modèle de Gamma prouve l'exactitude dans l'application des méthodes.

De manière générale, le modèle Log Gamma donne des distributions à queue plus épaisse que le modèle Poisson Sur-dispersé.

### **Conclusion de l'application numérique :**

Cette étude s'appuie tout d'abord sur l'application pratique de méthodes classiques de calcul de provisions techniques pour établir leur adéquation aux données.

Nous avons vu que la plupart des méthodes d'estimation des réserves convergent vers la méthode de Chain-Ladder qui est incontestablement la méthode la plus utilisée en assurance non-vie ; elle est devenue une référence pour la plupart des assureurs qui voient en elle une méthode simple et relativement efficace par rapport à ces critères de performance. D'ailleurs, les modèles de provisionnement relative-



ment innovateurs sont souvent appliqués à la méthode Chain-Ladder.

Les résultats obtenus par les méthodes non paramétriques pour le calcul des provisions techniques et la méthode paramétrique du modèle de Poisson surdispersée donnent dans un premier temps une diversité de choix pour l'estimation des réserves de sinistres pour les compagnies d'assurances, et prouvent un travail compétant des actuaires dans le domaine d'assurance non vie.

### Critiques de la méthode Chain-Ladder

- Cette méthode relativement, simple, pose malgré tout quelques problèmes :
- Le schéma de développement est identique pour toutes les années de survéances, c'est-à-dire que le coût du sinistre au bout de  $j$  années de déroulé est proportionnel au coût de l'année précédente, ou même de n'importe quelle année  $< j$ , et ce coefficient de proportionnalité ne change pas (et correspond à  $f_{j-1}f_{j-2} \dots f_i$ ) Ceci n'est généralement pas le cas en pratique dans plusieurs situations :
  - Changement de jurisprudence : dans ce cas-là, il peut  $y$  avoir un saut au niveau des paiements, qui seront alors beaucoup plus élevés si les paiements ont lieu aujourd'hui que s'ils avaient eu lieu l'année passée ;
  - Changement de management (au niveau de la gestion des sinistres, ou de la souscription) : pour diverses raisons, une entreprise, qui pensait qu'il pouvait être intéressant de faire traîner les sinistres dans la durée (et aller au tribunal plutôt que payer rapidement) peut penser, par la suite, qu'il peut être intéressant de payer très rapidement les sinistres (en particulier pour des raisons de couts internes).
  - Pour les années récentes, l'incertitude est très importante : le coefficient multiplicatif de la dernière année est le produit de  $n - 1$  estimations de coefficients de proportionnalité. Cette incertitude est d'autant plus grande pour les risques longs, où les premiers paiements commencent au bout de quelques années : si les paiements de la première année représentent de l'ordre de 1% du montant total, avoir payé 0.8% ou 1.2% va faire varier le montant total de provisions pour cette année de 50%! l'utilisation des triangles de paiements dans ce cas peut s'avérer hasardeux.
  - Pour reprendre le point précédent, cette méthode déterministe ne permet pas d'obtenir de mesure de précision (ou plutôt d'imprécision) sur les estimations [9]

# Conclusion

Depuis les années 1990, de nombreuses avancées ont été faites dans la modélisation des provisions pour sinistres à payer et dans la compréhension des modèles. Traditionnellement, les actuaires ont utilisé la méthode Chain-Ladder. Depuis, les méthodes économétriques reposant sur des facteurs ligne ou colonne ont été approfondies (et offrent aujourd'hui une très grande richesse). Parallèlement, les méthodes paramétriques ont bénéficié des développements récents des logiciels. Les actuaires commencent seulement à mieux comprendre ces modèles, souvent complexes, à mieux comprendre comment intégrer un avis d'expert, comment prendre en compte des ruptures dans la gestion des sinistres ou de l'impact des traités de réassurance. Malheureusement, ces avancées sur la modélisation, qui s'avèrent maintenant beaucoup plus délicates, n'ont pas été accompagnées d'une communication adaptée.

Les méthodes de provisionnement évoluent à la fois pour répondre à cette nouvelle exigence réglementaire et en vue de satisfaire également à certains critères. En effet, de nombreuses méthodes ont vu le jour pour s'adapter à ce que J.M. Nessi 2 nomme « critères de performance » : données fiables ou nombreuses, branche peu volatile, etc. Les méthodes progressent pour faire face à la non-réalisation d'un ou plusieurs de ces critères. Chain Ladder est incontestablement la méthode la plus utilisée en assurance non-vie ; elle est devenue une référence pour la plupart des assureurs qui voient en elle une méthode simple et relativement efficace par rapport à ces critères de performance. D'ailleurs, les modèles de provisionnement relativement innovateurs sont souvent appliqués à la méthode Chain Ladder dans un premier temps.

Toutefois, on est en droit de se demander si la méthode Chain-Ladder est la plus appropriée dans la plupart des cas. Le provisionnement est souvent affaire de choix et, selon Hans Bühlmann 3, « le provisionnement n'est pas un problème de modélisation complexe et sophistiqué, mais plutôt un exercice de choix de modèle ». En effet, lorsque l'actuaire retient une méthode d'évaluation, cela a nécessairement une conséquence importante sur le niveau des provisions à constituer. C'est

ainsi que le savoir et l'expérience de l'actuaire interviennent dans le processus d'évaluation des provisions techniques. La méthode de Chain-Ladder permet de surcroît d'introduire ce jugement dans le calcul même des provisions.

Les compagnies d'assurance doivent désormais répondre à de nouvelles exigences réglementaires introduites par la directive Solvabilité II. En assurance non-vie, les provisions pour sinistres à payer sont souvent les plus importantes en montant et nécessitent, par conséquent, une évaluation très précise. Celles-ci correspondent aux sinistres survenus à la date d'inventaire mais qui n'ont pas encore été réglés.

C'est le temps pour faire le bilan : résumons ce que nous avons parcouru dans cette thèse. Nous avons revisité la méthode Chain-Ladder à la fois pour montrer ses points faibles et pour nous inspirer sur notre approche géométrique. Nous avons construit un cadre théorique standard avec la notion centrale des mesures de provision cohérente, où nous avons proposé pour la première fois une approche graphique pour le calcul des réserves, Nous avons ensuite étudié les modèles linéaires généralisés en provisionnement et enfin, nous avons fait une application numérique et une analyse critique dans laquelle nous avons appliqué plusieurs méthodes qui ont été détaillé tout au long de cette thèse pour évaluer les provisions techniques au sein d'une compagnie d'assurance.

# Perspectives de recherches

Comme perspectives de travail, il serait intéressant de traiter les sinistres par mois de développement et non pas par année de développement et comparer les résultats des estimations des réserves. Il serait aussi intéressant d'étudier les sinistres incrémentaux négatifs au sein des compagnies d'assurances, un problème qui a été évoqué par R.J. Verral dans l'article [33]. Voilà une belle question de recherche pour de futures investigations.

# Bibliographie

- [1] Alai D, H. Wüthrich M. V, *Model Uncertainty within the Tweedie Exponential Dispersion Family*. Department of Mathematics, ETH Zurich, 8092 Zurich, Switzerland January 13, 2009
- [2] Alai D, H. Wüthrich M. V, *Taylor approximations for Model uncertainty within the tweedy exponential dispersion family*. ASTIN Bulletin : The Journal of the IAA. 2009, 453–477.
- [3] Alai D, H., Merz, M. Wüthrich M. V *Mean Square Error of Prediction in the Bornhuetter-Ferguson Claims Reserving Method*. Annals of Actuarial Science. 2009, 7–31
- [4] Belhamra, T., Remita, M. R., Chorfi, I. *On Geometrical approach for the Chain Ladder method : Properties and Application*. International Journal of Applied mathematics & Statistics. 2018, vol.57
- [5] BENETEAU, B. *Modèle de provisionnement sur données détaillées en assurance non-vie*. ENSAE Paristech. 2004.
- [6] Cabe Chadick, Wes Campbell, Finn Knox-Seith, *COMPARISON OF INCURRED BUT NOT REPORTED (IBNR) METHODS*. Society of Actuaries Health Section, Tech. Rep. 2009.
- [7] Charpentier, A., L. Devineau J-M Nessi : *Mesurer le risque lors du calcul des provisions pour sinistres à payer*. Revue Risques. 2010.
- [8] Chorfi, I., and Remita, M. R. *Stochastic incremental approach for modelling the claims reserves*. International Mathematical Forum 8 (2013), 807-828.
- [9] Denuit, M., et Charpentier, A. *Mathématique de l'assurance non-vie. Tome II : Tarification et provisionnement*. 2005. 596p
- [10] Efron, B., and Tibshirani, R. J. *An introduction to the bootstrap*. Chapman & Hall, 1993.
- [11] England, P. Addendum to 'analytic and bootstrap estimates of prediction errors in claims reserving'. Insurance : Mathematics and Economics 31 (2002), 461-466.

- [12] England, P., and Verrall, R. *Analytic and bootstrap estimates of prediction errors in claims reserving*. Insurance : Mathematics and Economics 25 (1999), 281-293.
- [13] England, P., and Verrall, R. *Stochastic claims reserving in general insurance*. British Actuarial Journal 8 (2002), 443-544.
- [14] England, P., and Verrall, R. *Predictive distributions of outstanding liabilities in general insurance*. Annals of Actuarial Science 1 (2006), 221-270.
- [15] Hoedemakers, T. Beirlant, J. Goovaerts, M. Dhaene, J. *On the Distribution of Discounted Loss Reserves Using Generalized Linear Models*.Scandinavian Actuarial Journal. 2005, 25-45.
- [16] Kaishev V. K. D. S. Dimitrova S. Haberman R. Verrall *Geometrically designed, variable knot regression splines* . Computational Statistics. 2016, 1079-1105.
- [17] Mack, T. *Distribution-free calculation of the standard error of chain-ladder reserve estimates*. ASTIN Bulletin 23 (1993), 213-225.
- [18] Mack, T. *A simple parametric model for rating automobile insurance or estimating ibnr claims reserves*. ASTIN Bull 21 (1991), 93-109.
- [19] Mack, T. *Which stochastic model is underlying the chain-ladder method* Insurance Math. Econom 15 (1994), 133-138.
- [20] Mack, T., and Venter, G. *A comparison of stochastic models that reproduce chain ladder reserve estimates*. Insurance : Mathematics and Economics 26 (2000), 101-107.
- [21] Merz, M., and Wüthrich, M. V. *Prediction error of the expected claims development result in the chain-ladder method*. Bulletin of Swiss Association of Actuaries 1 (2007), 117-137.
- [22] Merz, M., and Wüthrich, M. V. *Modelling the claims development result for solvency purposes*. Casualty Actuarial Society E-Forum (2008), 542-568.
- [23] Mulquiney p : *Artificial Neural Networks in Insurance Loss Reserving*. JCIS. 2006.
- [24] Nicolino Ettore D’Ortona Giuseppe Melisi *Comparison of the claims reserves methods by analyzing the run-off error*. INSURANCE MARKETS AND COMPANIES : ANALYSES AND ACTUARIAL COMPUTATIONS. 2016, 11-a.
- [25] Pinheiro, P., Andrade e Silve, J., and Centeno, M. *Bootstrap methodology in claim reserving*. Journal of Risk and Insurance 70 (2003), 701-714.
- [26] Prakash Narayan, *Anatomy of Actuarial Methods of Loss Reserving*.2013.
- [27] Renshaw, A., and Verrall, R. *The stochastic model underlying the chain-ladder technique*. Actuar. Res. Paper No. 63. (1994).

- [28] Renshaw, A., and Verrall, R. *A stochastic model underlying the chain-ladder technique*. British Actuarial journal 4 (1998), 903-923.
- [29] Schmidt, K. D. *A bibliography on loss reserving*. see : <http://www.math.tu-dresden.de/sto/schmidt> (2013).
- [30] Taylor, G. *Loss Reserving : An Actuarial Perspective*. Kluwer Academic Press, 2000.
- [31] Taylor, G. *The Chain Ladder and Tweedie Distributed Claims Data*. Centre for Actuarial Studies, Department of Economics, University of Melbourne. 2007.
- [32] Ting Yan : *The Prediction of Non-Life Claim Reserves under Inflation*. Open Journal of Statistics. 2016, 320.
- [33] Verrall R.J *An investigation into stochastic claims reserving models and the chain-ladder technique*. Insurance : Mathematics and Economics 26 (2000) 91-99
- [34] Wüthrich, M. V., and Merz, M. *Stochastic Claims Reserving Methods in Insurance*. Wiley Finance, 2008.
- [35] Wüthrich, M. V., Merz, M *Stochastic Claims Reserving Methods in Non-Life Insurance*. 2006.
- [36] Zhou, J., Garrido, J. *A Loss Reserving Model within the framework of Generalized Linear Models*. NSERC Discovery Grant. 2006, 36860-06.