Mémoire présenté le :

pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA
et l'admission à l'Institut des Actuaires

Par : Jemuel Dimitri KINGNE WAFFO

Titre : Modélisation d'une provision agrégée sur un portefeuille IARD

Confidentialité : ☐ NON ☒ OUI (Durée : ☐ 1 an ☒ 2 ans)

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Membres présents du jury de l'Institut des Actuaires

Membres présents du jury de l'ISFA

Entreprise : Cardif IARD

Nom :

Signature :

Directeur de mémoire en entreprise :

Nom : Manal BOURACHID

Signature :

Invité :

Nom :

Signature :

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

Signature du responsable entreprise

Signature du candidat

SOMMAIRE

I CONTEXTE ET REGLEMENTATION **10**

1 CARDIF IARD **11**
1 Création . . . 11
2 Activité . . . 11

2 Assurance IARD, Risques et Sinistres **13**
1 L'assurance IARD . . . 13
2 Risques et Sinistres . . . 13
3 Provisionnement non-vie : Réglementation et natures de provisions . . . 14
4 Données et outils à disposition . . . 16
4.1 Données sélectionnées . . . 16
4.2 Outils utilisés . . . 16

II Les méthodes de provisionnement **17**

3 La méthode de Chain-Ladder **20**
1 Présentation de la méthode . . . 20
2 Validation empirique de la méthode . . . 20
2.1 Validation par les C-C Plot . . . 21
2.2 Validation par le D-triangle . . . 21
3 Limites de la méthode . . . 21
4 Application de Chain-Ladder sur la lob 5 . . . 22

4 Le modèlede Mack **24**
1 Présentation de la méthode . . . 24
2 Estimations par le modèle . . . 25
3 Application du modèle de Mack . . . 26

5 Les modèles stochastiques factoriels **27**
1 Contexte d'utilisation . . . 27
2 Régression log-normale . . . 28
3 Modélisation GLM . . . 29
3.1 Résidus du modèle . . . 30
3.2 Validation du modèle . . . 31
4 Modèle Glm-Bootstrap . . . 31
4.1 Application . . . 32

III Dépendance entre les risques **34**

6 Introduction à la dépendance **35**
1 Distributions multivariées . . . 35
1.1 Fonction de répartition jointe . . . 35

1.2 Fonction de densité jointe 35
2 Dépendance 36
2.1 Dépendance vs Corrélation 36
2.2 Outils de mesure de dépendance 36
2.3 Les types de dépendance 38

7 Méthodes de modélisation de la dépendance 40
1 L'approche Top-Down 40
2 L'approche Bottom-Up 40

IV Introduction à la notion de Copule 42

8 Présentation et propriétés 43
1 Définition d'une copule 43
1.1 Exemple de Copule 43
1.2 Propriétés d'une copule 43
2 Lien entre copules et mesures de dépendance 45
2.1 Lien avec le rho de Spearman 45
2.2 Lien avec le Tau de Kendall 45
3 Les Familles de Copules 45
3.1 Les Copules archimédiennes 45
3.2 Les Copules Elliptiques 48

9 Simulation des Copules 51
1 Algorithme de Marshall-Olkin 51
2 Méthode des distributions 52
2.1 Pour la copule normale 52
2.2 Pour la copule de Student 52
3 Méthode des distributions conditionnelles 53

10 Application de la théorie des copules 54
1 Ajustement des distributions marginales sur chacune des lobs 54
1.1 Densités empiriques 54
1.2 Recherche de lois 54
2 Structure de dépendance entre les 03 lob 55
2.1 Calcul des coefficients de corrélation entre les 03 lobs 56
2.2 Dépendogrammes deux à deux 58
2.3 Estimation des paramètres des Copules 58

Table des annexes 63

A Explication générale du maximum de vraisemblance 64

B Récapitulatif des résultats du Chain-Ladder sur l'ensemble des 03 lob 65

C Analyse rétrospective de Chain-Ladder sur l'ensemble des 03 lob 66

D Test de Shapiro-Wilk 68

E Test du signe des différences 69

F Analyse rétrospective des méthodes factorielles 70

G Illustration graphique de la normalité des résidus de Pearson des méthodes factorielles **72**

H Illustration de la fonction carré **74**

I Illustration de la densité empirique sur chacune des lob **75**

J Densité des lois utilisées en assurance non-vie **76**

K Illiusttration graphique de l'ajustement de la distribution empirique des provisions aux lois paramétriques par des qq-plot **77**

L Représentation graphique des résidus de Pearson **79**

Table des figures

2.1 Triangle de règlements cumulés 18
2.2 Triangle de règlements incrémentaux 18

3.1 Illustration du schéma de développement de dernière survenance 21
3.2 C(i,1) en fonction de C(i,2)sur la lob 5 23
3.3 Facteurs de de développement individuels pour j = 1 23

5.1 Quelques fonctions liens classiques 29
5.2 Quelques fonctions de lien canonique et de variance pour des distributions classiques 30
5.3 Calcul de déviance pour quelques distributions 31

7.1 Illustration de l'approche top-down 40
7.2 Illustration de l'approche Bottom-Up 41

8.1 Illustration du théorème d'invariance des copules 44
8.2 Densité et dépendogramme pour une copule de Frank de paramètre 3 47
8.3 Densité et dépendogramme pour une copule de Clayton de paramètre 3 47
8.4 Densité et dépendogramme pour une copule de Gumbel de paramètre 3 47
8.5 Caractéristiques des copules archimédiennes 48
8.6 Densité et dépendogramme pour une copule de Gaussienne de paramètre $\rho = 0.5$ 50
8.7 Densité et dépendogramme pour une copule de Gaussienne de paramètre $\rho = 0.5$, $\nu = 4$ 50

10.1 Dépendogrammes deux à deux entre les différentes lobs 58
10.2 Dépendogrammes de la Copule la mieux ajustée à nos données 59
10.3 Densité de la provision agrégée : meilleure Copule + meilleures lois et «Cas individuel» 60

F.1 Analyse rétrospective de la régression lognormale sur la lob4 70
F.2 Analyse rétrospective de la régression gamma sur la lob4 71

G.1 QQ-plot des résidus de Pearson : GLM de loi gamma 72
G.2 QQ-plot des résidus de Pearson : GLM de loi log normale 73

H.1 Illustration de la dépendance entre X et X^2 à travers la fonction carré 74

I.1 Illustration de la densité empirique sur chacune des lob 75

J.1 Illustration de la densité des lois utilisées en assurance non-vie 76

K.1 QQ plot d'ajustement de la lob 4 aux lois paramétriques 77
K.2 QQ plot d'ajustement de la lob 7 aux lois paramétriques 78
K.3 QQ plot d'ajustement de la lob 5 aux lois paramétriques 78

L.1 Représentation graphique des résidus de Pearson 79

Liste des tableaux

3.1 Résultats de Chain-Ladder sur la lob 5 22

4.1 Récapitulatifs des résultats obtenus pour le modèle de Mack 26

5.1 Résultats de la régression Log-Normale sur nos lobs 32
5.2 Résultats de la régression Gamma sur nos lobs 33

9.1 Générateurs et inverses de transformés de Laplace -Stieltjes des Copules archimédiennes 51

10.1 Test du caractère iid de nos données 55
10.2 Paramètres et statistiques du test d'ajustement des lois aux données 55
10.3 Marice de Corrélation de Pearson entre les lobs 57
10.4 Marice de Corrélation de Kendall entre les lobs 57
10.5 Marice de Corrélation de Spearman entre les lobs 57
10.6 Paramètres des copules associées à nos lobs 58
10.7 Synthèse globale des provisions agrégées par différentes Copules 59
10.8 Etude comparative entre la meilleure Copule et le Cas individuel 60

RESUME

Ce mémoire a pour objectif le calcul d'une provision consolidée en prenant en compte la dépendance entre différentes branches (lobs) du portefeuille d'un assureur non-vie en utilisant les copules.

Dans un contexte de concurrence croissante et d'émergence de nouvelles sources de risques, les compagnies d'assurance doivent adapter leurs méthodes de modélisation des provisions consolidées. La modélisation actuelle utilise souvent un coefficient de corrélation linéaire, ce qui présente des limites car elle ne tient compte que des dépendances linéaires et suppose une distribution normale pour les branches étudiées. Cela conduit souvent à des hypothèses simplificatrices pour la modélisation des dépendances entre les branches ou les garanties en assurance. Pour surmonter ces limites, une alternative est l'utilisation de copules, qui permettent de modéliser les dépendances multivariées en proposant différentes structures de dépendance. Après avoir examiné les méthodes de provisionnement classiques, notre objectif sera de déterminer les lois de distribution des provisions individuelles pour chaque branche, de déterminer la structure de dépendance entre elles, de choisir la meilleure copule en utilisant des tests d'adéquation et d'estimer ses paramètres pour simuler la provision consolidée sur l'ensemble des branches.

MOTS CLEFS :

Lob : line of business , Copule, Dépendance, corrélation linéaire

ABSTRACT

This thesis aims to use copulas to calculate a consolidated reserve for a non-life insurer's portfolio, taking into account the dependence between different lines of business (LOBs). In a highly competitive industry with the emergence of new sources of risk, insurance companies must adapt their methods for modeling consolidated reserves. The traditional method of using linear correlation coefficients has weaknesses, as it assumes normal distribution for the LOBs and only considers linear dependencies. As a result, modeling dependencies between LOBs in insurance has often relied on simplifying assumptions. Copulas offer an alternative to classical modeling, providing answers regarding multivariate modeling of dependence by proposing other dependence structures.

After presenting a review of traditional reserving methods, this thesis will determine the different distributions of individual provisions for LOBs, establish the structure of dependence between them, use adequacy tests to determine the best copula for modeling the structure and its parameters, and simulate it to obtain a consolidated reserve for all LOBs.

KEYWORDS :

Lob : line of business, Copula, Dependence, linear correlation

REMERCIEMENTS

Je remercie toutes les personnes qui m'ont aidé, conseillé et soutenu pour la rédaction de ce mémoire.

Je remercie tout particulièrement Manal Bourachid et Joachim Lemaire qui m'ont encadré depuis le début de mon stage jusqu'à la rédaction de ce mémoire. Je remercie également Bassirou Baldé pour la relecture de mes travaux et toutes les équipes de Cardif IARD avec lesquelles j'ai eu le plaisir de travailler.

Dernièrement,je tiens à remercier l'ensemble du Corps enseignant de l'ISFA, et en particulier Monsieur Denys Pommeret pour l'encadrement académique, sa disponibilité et sa réactivité dans le suivi de ce mémoire.

INTRODUCTION

Les provisions d'une compagnie d'assurance correspondent à toutes les réserves constituées par celle-ci, dans le but de respecter les engagements pris envers les assurés. La principale provision estimée au bilan d'un assureur non-vie est la PSAP (provision pour sinistres à payer) correspondant à la somme à mettre en réserve par la compagnie d'assurance pour faire face à la sinistralité survenue. Etant donné qu'il existe différents risques au sein du portefeuille d'un assureur, il serait intéressant de calculer cette provision en prenant en compte les structures de dépendance entre ces différents risques ou branches d'assurance : on obtiendra donc une provision consolidée.

Dans le domaine de l'assurance non-vie en général et IARD en particulier, l'une des possibilités s'offrant à l'assureur est l'utilisation du coefficient de corrélation de Pearson pour modéliser cette provision consolidée ; mais cette méthode présente un inconvénient dû au fait qu'elle ne modélise que la corrélation linéaire, elle ne permet donc pas de prendre en compte toute la structure de dépendance entre les branches étudiées. Une autre méthode qui s'offre à l'assureur est l'ajustement de lois multivariées sur l'ensemble des branches étudiées, mais cette méthode est contraignante car, elle impose la même loi de probabilité pour toutes les branches ce qui est rarement le cas dans la pratique.

Pour lever ces contraintes, les actuaires se sont récemment penchés sur de nouvelles méthodes de modélisation de lois multivariées. Parmi celles-ci, la théorie des Copules, qui remonte aux années 50 avec les travaux de M. Fréchet, permet de construire des lois multivariées sur la base de lois marginales choisies et une structure de dépendance donnée. L'application de cette théorie dans le cadre assurantiel est assez récente et les écrits sur le sujet ont notamment débuté avec Klugman (1998). Depuis, de nombreux articles et études traitent de ce sujet et montrent l'intérêt de la mise en œuvre de cette théorie que ce soit dans le domaine de la finance ou de l'assurance.

Dans le cadre de ce mémoire, nous allons utiliser la théorie des Copules afin de modéliser une provision consolidée sur l'ensemble des branches (lobs) d'un assureur non-vie, le but étant d'analyser l'impact de la prise en compte des dépendances entre branches d'assurance sur la provision globale de l'assureur.

Nous allons dans la suite de nos travaux, après une présentation des principales méthodes de provisionnement, faire une introduction à la notion de dépendance et à la théorie des copules, puis utiliser ces outils pour pouvoir dans le cadre d'une application sur nos données, estimer une provision consolidée. Notre analyse se basera donc sur une analyse comparative entre le cas où l'assureur tient compte de la structure de dépendance entre les branches pour avoir une provision consolidée et le cas où le calcul des provisions est fait individuellement sur chaque branches puis obtenu par somme de provisions individuelles par branches.

Première partie

CONTEXTE ET REGLEMENTATION

Chapitre 1

CARDIF IARD

1 Création

Cardif IARD est une coentreprise, spécialisée en assurance dommages, créée suite à la signature du protocole d'accord entre BNP Paribas Cardif et Matmut en 2016. Elle est détenue à 34% par Matmut et 66% par BNP Paribas Cardif.

BNP Paribas Cardif est la filiale assurance vie et dommages de BNP Paribas. Elle conçoit des produits et services dans le domaine de l'épargne et de la prévoyance et les commercialise par l'intermédiaire de multiples canaux de distribution. Présente dans 35 pays, ses activités sont géographiquement diversifiées avec des positions fortes en Europe, en Amérique Latine et en Asie.

Matmut est un acteur mutualiste de référence sur le marché français de l'assurance. A ce jour avec plus de 3,8 millions de sociétaires et 7,2 millions de contrats d'assurance gérés, elle offre aux particuliers, professionnels, entreprises, associations une gamme complète de produits d'assurance des personnes et des biens (auto, moto, bateau, habitation, protection de la famille, santé, protection juridique et assistance) et de services financiers et d'épargne (crédits auto, projet, assurance emprunteur, livret d'épargne, assurance vie).

2 Activité

Cardif IARD a été lancée en mai 2018, son démarrage constitue un projet d'envergure particulièrement remarquable sur le marché de l'assurance dommages, extrêmement concurrentiel en France. Son activité principale est de concevoir, promouvoir, distribuer, gérer et maintenir une offre de produits d'assurance IARD compétitive au travers des canaux de distribution de BNP Paribas (Banque de Détail France et Personal Finance), sous les marques du groupe BNP Paribas ou sous des marques propres de Cardif IARD et potentiellement d'autres canaux de distribution dans un modèle omnicanal et multidistributeur.

Au 31 décembre 2022, Cardif IARD propose les contrats d'assurance IARD suivants :

— Contrat Auto
— Contrat Moto
— Contrat Habitation
— Contrat Assurance Scolaire
— Contrat Habitation Etudiant
— Contrat Assurance Multirisque Professionnelle
— Contrat Mobilité Electrique (EDPM)
— Contrat Risques Immobiliers

En plus de son propre portefeuille, Cardif IARD détient depuis janvier 2018 les portefeuilles Natio et Avanssur suite à un transfert depuis la société NATIO Assurance. Celle-ci était le fruit d'un partenariat entre les groupes BNP Paribas et AXA France IARD qui a été rompu en 2018. La cessation du partenariat s'est traduite par le rachat de la part du Groupe AXA dans la coentreprise Natio Assurance. Et depuis cela, Natio Assurance a été mise en run-off avec pour conséquence l'arrêt de la souscription d'affaires nouvelles, à l'exclusion du renouvellement de polices existantes et des devis émis à date qui restent à honorer ultérieurement.

Chapitre 2

Assurance IARD, Risques et Sinistres

1 L'assurance IARD

L'assurance IARD a pour but de couvrir les préjudices subis par des biens ou des personnes couvertes par un contrat d'assurance. Les contrats IARD ont donc pour rôle la protection des biens ou des personnes, et sont donc indispensables pour la protection des personnes, d'équipements, de locaux, de stocks ou encore des véhicules contre les accidents, vols ou toutes autres causes occasionnant la dégradation de ces derniers. Ces préjudices peuvent donc être des accidents, des incidents, des vols, des inondations ou des détériorations à la suite de toutes causes non intentionnelles de la part de l'assuré. Pour pouvoir apporter une couverture optimale à l'assuré, l'assureur IARD met en place différents contrats d'assurance, dont :

- Les contrats auto : qui couvrent les véhicules des particuliers ou encore les flottes automobiles des entreprises, mais aussi les automobilistes
- Les contrats habitation ou locaux professionnels : qui permettent de protéger les habitations et leurs occupants, des locaux commerciaux et les autres biens matériels. Les contrats habitations intègrent généralement la responsabilité civile vie privée destinée à couvrir les dommages causés à des tiers
- Les contrats multirisques professionnels : qui couvrent les marchandises, stocks, la perte d'exploitation mais aussi la responsabilité civile.

2 Risques et Sinistres

Notion de sinistre

La survenance d'un sinistre est la circonstance non souhaitée mais envisagée par le contrat d'assurance.
Le sinistre est traditionnellement conçu comme la survenance de l'événement dont l'éventualité a été prise en considération lors de la souscription du contrat.
Dans les faits, il n'existe pas une définition stricte de la notion de sinistre dans le domaine assurantiel car, chaque sinistre dans chaque type d'assurance répond à sa propre définition. Ainsi, au lieu de donner une définition générale de ce qu'est un sinistre, le législateur donne des définitions particulières en fonction de la lob d'assurance. De ce fait, en responsabilité civile, selon l'article L.121-1-1 : « constitue un sinistre tout dommage ou ensemble de dommages causés à des tiers, engageant la responsabilité de l'assuré, résultant d'un fait dommageable et ayant donné lieu à une ou plusieurs réclamations. Le fait dommageable est celui qui constitue la cause génératrice du dommage ».

Ainsi, trois dates sont importantes à prendre en compte dans le traitement d'un sinistre :

— La date de survenance : qui est la date à laquelle a lieu le sinistre.

— La date de déclaration : C'est la date à laquelle l'assureur est informé du sinistre (par l'assuré, par l'assureur adverse, par un tiers). Cette date est nécessairement postérieure à la date de survenance (l'écart peut être de quelques heures comme plusieurs années). L'assureur réalise une 1ère évaluation du sinistre à la date de déclaration.

— La date de règlement : elle correspond à la date à laquelle l'assuré est remboursé, il peut il y en avoir plusieurs, dépendant de la complexité dans l'évaluation réelle du préjudice subi par l'assuré ou la victime. Cette date est nécessairement postérieure à la date de déclaration. Sur les petits sinistres, peut être payé en une seule fois. Sur les sinistres plus importants souvent en plusieurs fois (en fonction de la connaissance / évolution précise du sinistre, des expertises, des factures, etc).

Les catégories de risque

De façon générale, l'actuariat inventaire catégorise ses risques en fonction du délai entre la survenance et le temps qu'il met à régler totalement le sinistre.

De cette façon, on distingue donc deux grandes catégories de risques : les risques à déroulement long et les risques à déroulement court qui ont respectivement des durées de règlement longues et courtes comme indiquées par leurs noms. A titre illustratif, les sinistres de dommages aux biens (vols, incendies, dégâts des eaux...) seront déclarés et réglés assez rapidement (dans un délai de zéro à trois ans) ; par contre, les sinistres de responsabilité civile peuvent être déclarés beaucoup plus tard et réglés pendant dix voire vingt ans. C'est le cas par exemple d'un sinistre en responsabilité civile corporelle automobile pendant lequel, dans le cas d'une potentielle invalidité d'un des protagonistes, le règlement sera sous forme de rente qui sera payée jusqu'à la fin de vie de celui-ci.

3 Provisionnement non-vie : Réglementation et natures de provisions

Les types de provisions en assurance non-vie

Pour pouvoir respecter ses engagements vis à vis de ses assurés, l'assureur doit constituer des provisions suffisantes. Parmi ces provisions, les principales sont :

— La provision pour sinistres à payer (PSAP) : elle a pour but de couvrir les coûts liés aux sinistres déjà survenus. Elle concerne aussi bien les sinistres déjà déclarés que les tardifs.

— La provision Mathématiques (PM) : en assurance non-vie, dans le cas de sinistres corporels graves donnant lieu à une invalidité de la victime, l'assureur peut être appelé à verser une rente pour invalidité, et ce jusqu'au décès de la personne couverte. Il faut donc dans ce cas calculer une provision mathématique correspondant à la somme des rentes futures actualisées pondérées par les probabilités de survie de la personne couverte.

— La provision pour primes non-acquises (PPNA) : c'est la part de prime que l'assureur doit conserver pour faire face aux risques à venir. Il s'agit en réalité de la part de prime pondérée par la période de couverture restante. Elle prend en compte pour son calcul :

  — Les primes émises (PE) : C'est la prime facturée puis payée par l'assuré pour une période de couverture donnée.

— Les primes acquises (PA) : La prime acquise en assurance représente la partie de la prime annuelle qui correspond à la période pendant laquelle l'assuré a été effectivement couvert par l'assurance. Par exemple, si l'assuré a souscrit une assurance pour une année entière et que la prime annuelle est de 1200€, la prime acquise au bout de six mois sera de 600€.

Segmentation de portefeuille

La segmentation d'un portefeuille d'assurance IARD se fait généralement soit :

— Par catégories ministérielles : il s'agit d'une segmentation proche du couple (Produit, Garantie).
Les catégories ministérielles sont listées à l'article A.344.2. du code des assurances. C'est la segmentation utilisée dans le cadre de la norme French Gaap.
— Par lobs : proche de la segmentation par catégories ministérielles, c'est la segmentation de référence dans le cadre de la norme solvabilité 2.

Dans la suite de nos travaux, nous allons nous intéresser à la modélisation de la provision pour sinistres à payer.

4 Données et outils à disposition

4.1 Données sélectionnées

Nous utilisons l'ensemble des données de paiement des portefeuilles Cardif IARD, Natio et Avanssur que nous allons considérer comme étant un portefeuille unique dans le but de pouvoir travailler avec un historique de données conséquent.

Ainsi, nous avons décidé pour nos travaux de nous attarder sur 3 des lobs les plus importantes du portefeuille combiné et qui contiennent à leur tour plusieurs groupes de risques homogène (GRH). On peut donc citer :

- La lob 4 qui comprend les GRH : Auto - RCM (responsabilité civile matérielle), Auto - RCC (responsabilité civile corporelle)
- La lob 5 qui comprend les GRH : Auto - Evenements climatiques, Auto-Dommage accidents , Auto-Vol Incendie, Auto Bris de glace et Catnat.
- La lob 7 qui comprend les GRH : dommage aux biens, dégats des eaux, incendie et de bris de glace en général (sauf sur l'auto)

Concernant les survenances des sinistres, nous sommes sur une période allant de 2007 pour les survenances les plus anciennes jusqu'à 2021 pour les survenances de sinistres les plus récentes.

4.2 Outils utilisés

Nous utiliserons principalement :

- Pour la modélisation des méthodes de Chain-Ladder : le logciel IBNRS.
- Pour la modélisation des méthodes stochastiques factorielles et des copules le logicel R
- Une fois les résultats obtenus les calculs supplémentaires et certaines mises en forme se feront sous EXCEL.

Deuxième partie

Les méthodes de provisionnement

Au-delà des estimations réalisées sinistre par sinistre par les gestionnaires, les méthodes de provisionnement (d'IBNR) sont basées la plupart du temps sur des données de sinistres agrégées, représentées dans un triangle de liquidation en fonction des périodes de survenance et des périodes de développement, permettant de refléter la dynamique de vie des sinistres. Il est à noter que d'autres méthodes moins utilisées fonctionnent avec des données de sinistres individuelles mais nous ne nous intéresserons pas à ces dernières.

Le triangle de liquidation propose une lecture à deux dimensions :

— Les lignes correspondent aux exercices de survenance des sinistres, représentées par l'indice i = 1,...,n
— Les colonnes correspondent aux exercices de développement des sinistres, représentées par l'indice j = 1,...,m

Ainsi nous noterons donc $X_{i,j}$ comme étant la donnée du triangle incrémental : c'est le montant des sinistres survenus à l'exercice i et réglés durant l'exercice i+j (où j exercices après la survenance i).

Nous noterons $C_{i,j}$ comme étant le montant cumulé pour chaque exercice de survenance i et par exercice de développement j : $C_{i,j} = \sum_{k=1}^{j} X_{i,k}$. Les triangles de liquidation avec des règlements cumulés sont donc sous la forme :

Année de survenance i	Année de développement j				
	1	2	...	n-1	n
1	$C_{1,1}$	$C_{1,2}$	...	$C_{1,n-1}$	$C_{1,n}$
2	$C_{2,1}$	$C_{2,2}$	...	$C_{2,n-1}$	
.	.	.			
.	.	.			
.	.	.			
n-1	$C_{n-1,1}$	$C_{n-1,2}$			
n	$C_{n,1}$				

FIGURE 2.1 – Triangle de règlements cumulés

Les triangles de liquidation avec des règlements incrémentaux sont donc de la forme :

Année de survenance i	Année de développement j				
	1	2	...	n-1	n
1	$X_{1,1}$	$X_{1,2}$	...	$X_{1,n-1}$	$X_{1,n}$
2	$X_{2,1}$	$X_{2,2}$	...	$X_{2,n-1}$	
.	.	.			
.	.	.			
.	.	.			
n-1	$X_{n-1,1}$	$X_{n-1,2}$			
n	$X_{n,1}$				

FIGURE 2.2 – Triangle de règlements incrémentaux

L'exercice de provisionnement consistera donc à remplir le triangle inférieur droit en estimant les valeurs correspondantes aux cellules vides.
Dans le cas du triangle de règlements incrémentaux, le montant de la provision globale est la somme des estimations du triangle inférieur, dans le cas du triangle de règlements cumulés, ce montant équivaut à la somme des différences par exercice de survenance entre le dernier montant estimé et le dernier montant connu.

Nous avons donc :

$$\hat{R} = \sum_{i=1}^{n} = \hat{R}_i = \sum_{i=1}^{n} \hat{C}_{i,n} - \hat{C}_{i,n+1-i} = \sum_{(i,j)\in D_n} \hat{X}_{i,j}$$

Avec :

$$D_n = (i,j)\ ,\ i+j \geq n+2\ ,\ i \leq n \text{ et } j \leq n$$

C'est le domaine qui constitue le triangle inférieur : le triangle dont on estime les montants. Nous allons donc présenter les principales méthodes de provisionnement, que nous allons mettre en place au niveau de l'application sur nos données.

Chapitre 3

La méthode de Chain-Ladder

1 Présentation de la méthode

C'est l'une des méthodes de provisionnement les plus anciennes et des plus répandues dans le secteur de la non-vie. Cela se justifie par sa simplicité et sa facilité de mise en oeuvre.
Cette méthode se base sur le principe selon lequel sur un même segment, les sinistres survenus à des exercices différents se développent de la même manière. Ainsi, elle repose sur les deux hypothèses suivantes :

— (H1) : Les années de survenance sont indépendantes entre elles, c'est-à-dire que pour i $\neq k$, (C$_{i,1}$,..., $C_{i,n}$) et ($C_{k,1}$,..., $C_{k,n}$) sont indépendants.

— (H2) : Les règlements des sinistres sont stables dans le temps.

Ainsi, sous ces hypothèses nous avons pour j = 0,...,n-1 :

$$\frac{C_{0,j+1}}{C_{0,j}} = \frac{C_{1,j+1}}{C_{1,j}} = ... = \frac{C_{n-j-1,j+1}}{C_{n-j-1,j}}$$

Et un facteur de développement unique par exercice de développement j qui est :

$$\hat{f}_j = \frac{\sum_{i=0}^{n-j-1} C_{i,j+1}}{\sum_{i=0}^{n-j-1} C_{i,j}} \text{ , avec j = 0,...,n-1}$$

Une fois les facteurs de développement estimés, la charge ultime par année de survenance qui est donnée par :

$$\hat{C}_{i,n} = (\hat{f}_{n-i}.....\hat{f}_{n-1}) * C_{i,n-i}$$

Ainsi que la provision à constituer par exercice de survenance qui sera :

$$\hat{R}_i = \hat{C}_{i,n} - C_{i,n-i} \text{ , avec i = 1,...,n}$$

C'est en réalité la différence entre le règlement ultime (estimé) et le dernier règlement connu du triangle cumulé.
La provision globale à constituer sera donc de :

$$\hat{R} = \sum_{i=1}^{n} \hat{R}_i$$

2 Validation empirique de la méthode

Il s'agit des méthodes de vérification du respect des hypothèses de la méthode par les données en présence.

2.1 Validation par les C-C Plot

Il s'agit ici de vérifier pour un même exercice de développement l'alignement sur une droite des points de cordonnées :

$$(C_{i,j},\, C_{i,j+1})\ ,\ \text{i} = 0,...,\text{n-j-1}.$$

Cette méthode part du principe selon lequel, si il existe un facteur f_j tel que :

$$C_{i,j+1} = f_j\ C_{i,j}\ ,$$

alors à j fixé , on s'attend à avoir des points sensiblement alignés sur une droite passant par l'origine.

2.2 Validation par le D-triangle

Il s'agit du triangle des facteurs de développement individuels :

$$f_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}\ ,\ \text{pour i+j} \leq \text{n-1}$$

On doit pouvoir observer une constance des facteurs de développement individuels à exercice de développement fixé.

3 Limites de la méthode

Bien qu'étant une méthode de provisionnement éprouvée, malgré sa simplicité et son efficacité, elle présente tout de même certaines limites :

Le fait que le schéma de développement soit le même pour tout exercice de survenance n'est pas toujours vrai en pratique : si on se place dans le cadre d'une société qui pratique une segmentation mensuelle de ses triangles (ou alors pour une lob longue), en considérant le dernier exercice de survenance, le premier montant réglé pourrait être anormalement bas et considérer un même schéma de développement pour cette survenance pourrait engendrer un sous-provisionnement sur celui-ci comme l'illustre l'image ci-dessous :

					R_N	
Annéé de Développement	1	2	3	4	5	
Dernière survenance (N)	500	650	910	1092	1201,2	701,2
coefficients de développement	1,3	1,4	1,2	1,1	1	

					R_N	
Année de Développement	1	2	3	4	5	
Dernière survenance (N)	1000	1300	1820	2184	2402,4	1402,4
coefficients de développement	1,3	1,4	1,2	1,1	1	

FIGURE 3.1 – Illustration du schéma de développement de dernière survenance

On observe donc que l'unique montant connu sur la dernière année de survenance du triangle peut avoir un très gros impact sur la provision. Dans notre exemple on observe que pour des montants observés de 500 et 1000 la provision est respectivement de 701 et 1402.

On peut aussi remarquer que, plus l'exercice de développement est élevé, moins il est fiable. A titre d'exemple, le dernier facteur de développement est calculé à partir de deux valeurs seulement mais il sera généralisé à tous les exercices de survenance lors du provisionnement.

Etant une méthode qui repose sur l'hypothèse d'une cadence de règlement constatée par

le passé et identique aux exercices suivants, elle ne prend donc pas en compte les différents changements susceptibles de se produire au cours du temps.

4 Application de Chain-Ladder sur la lob 5

On supposera que notre triangle est fermé ; c'est à dire qu'à tout exercice de survenance, le dernier paiement survient au plus tard au dernier exercice de développement de notre triangle. Au vu de la cadence de paiement sur notre triangle, cette hypothèse n'aura vraisemblablement pas un impact important sur les provisions totales car, pour chaque exercice de survenance, les incréments correspondant aux développements de plus d'une dizaines d'années, peuvent être considérés comme négligeables (Pour des raisons de confidentialité, nous n'allons pas présenter le triangle en question dans le cadre de ce mémoire).

Ci-dessous un récapitulatif des provisions par années de survenance sur la lob 5 :

Années de Survenance	Ultimes projetés	Provisions totales
01/2007	10 448 763	0
01/2008	11 497 722	11
01/2009	13 915 707	26
01/2010	14 107 314	38
01/2011	12 849 164	533
01/2012	13 110 857	898
01/2013	14 614 242	1069
01/2014	17 384 351	1415
01/2015	15 242 404	2267
01/2016	14 226 787	24 457
01/2017	12 609 323	56 957
01/2018	14 381 127	138 151
01/2019	17 514 024	211 959
01/2020	16 489 808	292 177
01/2021	23 435 723	5 731 828
Total	221 827 315	6 461 787

TABLE 3.1 – Résultats de Chain-Ladder sur la lob 5

Effectuons donc une vérification des hypothèses du modèle :

1. Alignement des couples $(C_{i,j}; C_{i,j+1})_{i=0,\ldots,n-j-1}$: à j fixé, ayant supposé l'existence d'un unique coefficient f_j tel que $C_{i,j+1} = f_j C_{i,j}$, les couples $(C_{i,j}; C_{i,j+1})_{i=0,\ldots,n-j-1}$ doivent être sensiblement alignés sur une droite passant par l'origine.
2. Examen des facteurs de développement individuels : l'hypothèse est que à j fixé on ait les facteurs de développement individuels : $f_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$ pour j = 0,1,...,n-1 sensiblement constants.

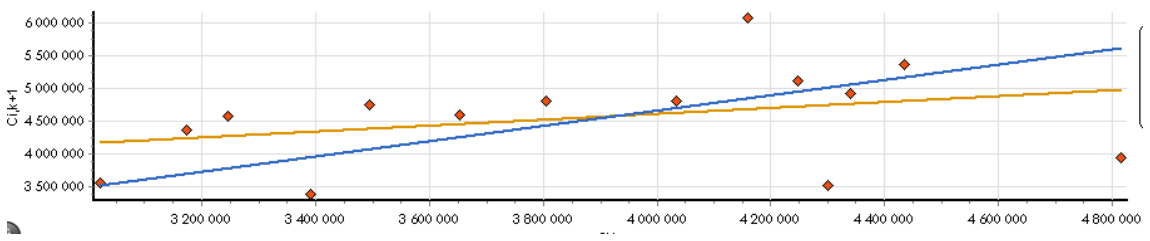


FIGURE 3.2 – C(i,1) en fonction de C(i,2)sur la lob 5

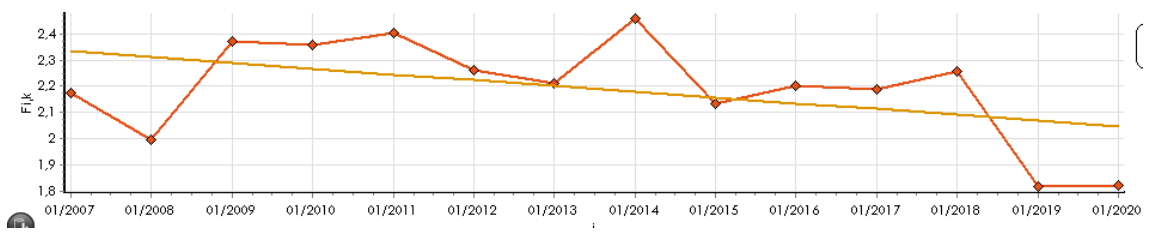


FIGURE 3.3 – Facteurs de de développement individuels pour j = 1

On observe que les hypothèses de la méthode ne sont pas totalement respectées par les données sur la lob 5. Cette non-vérification d'hypothèses de la méthode de Chain-Ladder est généralisée sur la plupart des exercices de développement récents de notre triangle, ce qui implique que nos facteurs de développement sont largement perfectibles (Constats observés sur les deux autres lobs de façon similaire).
Le détail des résultats de la méthode de Chain-Ladder sur les 3 lob figure en annexe B du mémoire.

Chapitre 4

Le modèlede Mack

Nous considerons D^n comme étant l'ensemble des informations apportées par le triange supérieur.

$$D^n = \{C_{i,k}, i + k \leq n\}$$

1 Présentation de la méthode

Le modèle de Mack est une version stochastique de la méthode classique de Chain Ladder, introduite par T.Mack en 1993. Le but de cette méthode est de pouvoir quantifier la variabilité des estimations des provisions et ainsi y associer des intervalles de confiances. Au vu de celà, nous nous attendons donc à obtenir les mêmes valeurs de provisions espérées qu'obtenues précédement dans la méthode de chain-ladder.

Hypothèses du modèle :

— (H1) : Les années de survenance sont indépendantes entre elles, c'est-à-dire que pour i $\neq k$, (C$_{i,1}$,..., $C_{i,n}$) et ($C_{k,1}$,..., $C_{k,n}$) sont indépendants.

— (H2) : Il existe une constante f_j strictement supérieur à 0, telle que $\forall$ j = 1,...,n-1 et $\forall$ i = 1,..,n :

$$\mathrm{E}(C_{i,j+1} \ / \ C_{i,1},...,C_{i,n}) = f_j \ C_{i,j}$$

Sous (H1) et (H2), les facteurs de développement $\hat{f}_j = \frac{\sum_{i=1}^{n-j} C_{i,j+1}}{\sum_{i=1}^{n-j} C_{i,j}}$

obtenus par la méthode de Chain Ladder classique, sont des estimateurs sans biais de f_j et non corrélés.

— (H3) : Pour j = 1,...,n-1 , il existe une constante σ_j^2 telle que :

$$\mathrm{V}(C_{i,j+1} \ / \ D^n) = \sigma_j^2 \ C_{i,j}$$

$\widehat{\sigma}$ étant la volatilité par période de développement, et :

$$\widehat{\sigma}_j^2 = \frac{1}{n-j-1} \sum_{i=1}^{n-j} \hat{C}_{i,j} (\frac{\hat{C}_{i,j+1}}{\hat{C}_{i,j}} - \hat{f}_j)^2$$

est un estimateur sans biais de σ_j^2 pour $1 \leq j \leq n-2$

Pour j = n-1, nous avons : $\sigma_j^2 = \min\ (\frac{\sigma_{j-1}^4}{\sigma_{j-2}^2}\ ,\ \sigma_{j-1}^2\ )$

2 Estimations par le modèle

Soit $\hat{C}_{i,m} = \hat{C}_{i,m-i} \prod_{l=m-i}^{m-1} f_l$, c'est un estimateur sans biais du montant de règlement espéré après m exercices de développement pour chaque exercice de survennance i, sa valeur est de :

$$\mathrm{E}(C_{i,m}\ /\ D^n) = C_{i,m-i} \prod_{l=m-i}^{m-1} f_l$$

Ainsi, on obtient donc la valeur d'un estimateur sans biais du règlement espéré à l'ultime comme étant :

$$\hat{C}_{i,n} = \mathrm{E}(C_{i,n}/D^n) = C_{i,n-i} \prod_{l=n-i}^{n-1} f_l$$

La montant total des réserves espérées correspondant à l'exercice de survenance i est :

$\mathrm{E}(R_i = C_{i,n} - C_{i,n-i}/D^n)$ et est estimé par :

$$\hat{R}_i = \hat{R}_{i,n} - C_{i,n-i}$$

Le réserve globale est estimée par : $\hat{R} = \sum_{i=1}^{n} \hat{R}_i$

Etant dans un cadre où les données de notre triangle sont des variables aléatoires, sous les hypothèses précédentes, les résultats obtenus de nos provisions présenteront donc des incertitudes que nous devrons évaluer dans le modèle.
Le modèle de Mack indique que l'erreur présente dans l'estimation de R_i par $\hat{R}_i$ est l'erreur quadratique moyenne (mean squared error : m.s.e). Elle est estimée par :

$$m.\hat{s}.e\ (\ \hat{R}_i) = \hat{C}_{i,n}^2 \sum_{k=n+1-i}^{n-1} \frac{\widehat{\sigma}_k^2}{f_k^2}\ (\ \frac{1}{\hat{C}_{i,k}} + \frac{1}{\sum_{j=1}^{n-k} C_{j,k}})$$

Avec comme convention : $\hat{C}_{i,n-i+1} = C_{i,n-i+1}$ et $\hat{C}_{i,k}$ étant les valeurs du triangle inférieur que nous avons estimées.
Il indique également l'incertitude présente dans l'estimation de la provision agrégée comme étant :

$$m.\hat{s}.e\ (\hat{R}) = \sum_{i=2}^{n} m.\hat{s}.e\ (\ \hat{R}_i) + \hat{C}_{i,n}\ (\ \sum_{k=i+1}^{n} \hat{C}_{k,n}) \sum_{j=n-i+1}^{n-1} \frac{2\widehat{\sigma}_j^2}{\hat{f}_j^2 \sum_{h=1}^{n-j} C_{h,j}}.$$

Partant du principe selon lequel $\hat{R}_i = \hat{C}_{i,n} - \hat{C}_{i,n-i}$, l'aléa au niveau de la provision par exercice de survenance provient donc de l'estimation du règlement à l'ultime. On a donc : $m.\hat{s}.e\ (\ \hat{R}_i) = m.\hat{s}.e\ (\hat{C}_{i,n})$. Ainsi, le modèle de Mack a donc donné une estimation de l'erreur présente dans l'estimation du montant des règlements cumulés du triangle inférieur $C_{i,j}$ par $\hat{C}_{i,j}$ avec (i+j $\geq$ n+2) à l'aide de la formule récursive suivante :

$$m.\hat{s}.e\ (\ \hat{C}_{i,j+1}) = \hat{C}_{i,j}^2 \frac{\widehat{\sigma}_j^2}{\hat{f}_j^2}(\frac{1}{\hat{C}_{i,k}^2} + \frac{1}{\sum_{k=1}^{n-j} \hat{C}_{k,j}^2})$$

.
En prenant comme valeur initiale $m.\hat{s}.e\ (\ \hat{C}_{i,n+1-i}) = 0$, car $C_{i,n+1-i}$ est en réalité une constante, c'est la derniere valeur connue par exercice de survenance.

3 Application du modèle de Mack

Le tableau ci-dessous, nous donne un récapitulatif des résultats obtenus sur les différentes lobs étudiées en appliquant le modèle de Mack.

	lob 4	lob 5	lob 7	Total
Espérance	28 345 916	6 461 787	56 979 761	91 787 464
Incertitude de prédiction (MSEP)	2 905 916	870 063	6 240 573	10 016 552
MSEP (relative)	10%	13%	11%	11%
Var90 (Loi Normale)	32 069 621	7 576 818	64 912 193	104 558 632
Var95 (Loi Normale)	33 124 777	7 892 914	67 244 271	108 261 962
Var99.5 (Loi Normale)	35 829 579	8 702 921	73 054 412	117 589 912
Var90 (Loi Log-Normale)	32 145 371	7 604 235	65 148 867	104 898 473
Var95 (Loi Log-Normale)	33 361 689	7 983 712	67 785 355	109 130 756
Var99.5 (Loi Log-Normale)	36 692 788	9 044 838	75 038 797	120 776 423

TABLE 4.1 – Récapitulatifs des résultats obtenus pour le modèle de Mack

Nous pouvons nous rendre compte (comme nous l'avons dit précédemment) que nous obtenons la même espérance de provision avec Mack et Chain-Ladder sur l'ensemble de nos lobs.

Le modèle de Mack utilise l'hypothèse selon laquelle les incréments du triangle sont indépendants, suivent une loi normale d'espérance la valeur de l'estimation et de variance le MSEP. A partir de ces hypothèses, il est donc aisé d'obtenir les paramètres de la provision sachant que la provision suivra aussi une loi normale. Mais on préfèrera généralement la loi log-normale pour cette hypothèse car elle évitera d'avoir des valeurs simulées de provisions négatives (Si ln(X) suit une loi $N(m, \sigma^2)$ alors X suit une loi log-normale avec : $E(X) = exp(m + \frac{\sigma^2}{2})$ et $V(X) = (exp(2m + \sigma^2))(exp\sigma^2 - 1))$.
Ces données ne prennent pas en compte la corrélation entre les lobs étudiées donc les *value-at-risk* totales sont sommées.

Par la suite, nous essayerons de visualiser une autre analyse de fiabilité de la méthode de Mack sur nos données à travers une analyse rétrospectives sur les données réelles. Le principe ici est qu'à partir des coefficients de développement calculés par notre modèle, on puisse faire une ré-estimation des derniers incréments connus par années de survenance et observer les écarts : les résultats des écarts observés sont donnés en annexe C du mémoire.

Au vu des résultats, on constate que toutes les hypothèses du modèle ne sont pas rigoureusement vérifiées et donc, la fiabilité des coefficients de développement est fortement remise en cause au vu des écarts entre les incréments estimés et les incréments connus sur nos triangles (analyse rétrospective en annexe C du mémoire). Nous allons donc nous intéresser à des méthodes qui étudient la cadence de paiement de manière plus fine : Les modèles stochastiques factoriels et qui sont moins exigeantes que la méthode de Chain-Ladder en termes d'hypothèses à vérifier.

Chapitre 5

Les modèles stochastiques factoriels

1 Contexte d'utilisation

En supposant que les données du triangle de règlement incrémental $(x_{i,j})_{i+j\leq n}$ sont des réalisations de variables aléatoires $(X_{i,j})_{i+j\leq n}$ indépendantes, on cherche donc à modéliser $\mu_{i,j} = E(X_{i,j})$ au moyen de certaines variables explicatives pertinentes.

Pour ce faire, on retiendra donc les trois variables explicatives suivantes :

— La variable "exercice de survenance" : elle est qualitative ordinale et prend les modalités 0,1,...,n en fonction du nombre d'exercices de survenance. Elle permet d'expliquer l'effet "volume" du portefeuille, c'est à dire l'évolution des primes d'assurances au fil du temps. Elle est remplacée par n variables indicatrices dont les paramètres sont α_1 ,α_2,....,α_n, ($\alpha_0 = 0$, car il n'y a pas d'estimation à faire sur la première ligne).

— La variable "exercice de développement" : On peut la considérer soit comme variable quantitative discrète soit comme variable factorielle. Elle sera donc représentative de l'effet "colonne" du triangle, c'est à dire de la cadence de paiement des sinistres. Elle est remplacée par m variables indicatrices dont les paramètres sont β_1 ,β_2,....,β_n, ($\beta_0 = 0$, car il n'y a pas d'estimation à faire sur la première colonne).

— La variable "exercice calendaire" : permettant d'avoir une situation datée du règlement et peut donc ainsi permettre de capter l'effet de l'inflation. Dans notre cas on supposera cet effet constant dans le temps pour simplifier le modèle. Elle fait donc entrer en jeu 2n paramètres μ_{i+j} et donc pour simplifier, on considèrera $\forall i, j = 0, 1, ..., n$, on pose $\mu_{i+j} = \mu$.

On fera donc une hypothèse sur la loi suivie par la variable aléatoire $X_{i,j}$ Ainsi, le modèle est donc le suivant :

$$\mu_{i,j} = E(X_{i,j}) = \mathrm{f}(\mu,\alpha_i,\beta_j)$$

Les paramètres du modèle peuvent être estimés par la méthode du maximum de vraisemblance (expliquée en annexe A du mémoire). Soit $\theta = (\mu,(\alpha_i),(\beta_j))_{i,j=1,...,n}$, le vecteur des paramètres, soit $\widehat{\theta}_{MV} = (\widehat{\mu},\widehat{\alpha}_i,\widehat{\beta}_j)$ l'estimateur du maximum de vraisemblance (EMV) de ce paramètre, nous avons que :

L'EMV de notre variable d'intérêt $\mu_{i,j}$ est :

$$\widehat{\mu_{i,j}} = \mathrm{f}(\widehat{\mu},\widehat{\alpha}_i,\widehat{\beta}_j)$$

Celui de $E(R_i)$ est :

$$(R_i) = \sum_{j=n-i+1}^{n} \widehat{\mu_{i,j}}$$

Celui de E(R) est :

$$(R) = \sum_i \sum_j \widehat{\mu_{i,j}}$$

L'expression de la variance des provisions estimées par exercice de survenance s'écrit :

$$Var((R_i)) = \sum_{j=n-i+1}^{n} Var(\hat{\mu_{i,j}}) + \sum_{j_1=n-i+1}^{n} \sum_{j_2=n-i+1}^{n} cov(\hat{\mu_{i,j_1}}, \hat{\mu_{i,j_2}})$$

$$\text{Avec } j_1 \neq j_2$$

Et celle des provisions estimées totales :

$$Var((R)) = \sum_{i=1}^{n} \sum_{j=n-i+1}^{n} Var(\hat{\mu_{i,j}}) +$$

$$\sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{j_1=n-i+1}^{n} \sum_{j_2=n-i+1}^{n} cov(\hat{\mu_{i,j_1}}, \hat{\mu_{i,j_2}}), (i_1 \neq i_2)$$

2 Régression log-normale

Dans ce modèle, on supposera que $X_{i,j} \sim LN(m_{i,j}, \sigma^2)$. On prendra ici comme variables explicatives l'année de survenance (α_i) et l'année de développement (β_j), on pose donc

$$m_{i,j} = \mu + \alpha_i + \beta_j \text{ Avec } \alpha_0 = \beta_0 = 0.$$

En considérant $Y_{i,j} = ln(X_{i,j})$ nous avons par définition que :

$$Y_{i,j} \sim N(m_{i,j}, \sigma^2) \text{ avec } E(Y_{i,j}) = m_{i,j} = \mu + \alpha_i + \beta_j.$$

On peut donc écrire :

$$Y_{i,j} = \mu + \alpha_i + \beta_j + \epsilon_{i,j} \text{ , avec } \epsilon_{i,j} \sim N(0, \sigma^2).$$

Il en résulte donc que :

$$\mu_{i,j} = E(X_{i,j}) = exp(m_{i,j} + \frac{\sigma^2}{2})$$

$$Var(X_{i,j}) = exp(2m_{i,j} + \sigma^2)(exp(\sigma^2) - 1)$$

Au final, il s'agit d'estimer, à partir des éléments du triangle de liquidation, un modèle de régression linéaire classique qui s'écrit sous forme matricielle :

$$Y = M\theta + \epsilon$$

Avec Y le vecteur colonne, d'ordre $t = \frac{(n+1)(n+2)}{2}$, des éléments du triangles pris ligne à ligne et ϵ le vecteur des erreurs. θ est le vecteur des paramètres du modèle (d'ordre p = 2n+1) et M est la matrice de régression dont la première colonne est le vecteur unitaire, puis les autres colonnes sont les valeurs prises par chacune des variables explicatives du modèle.

On estime donc θ par

$$\hat{\theta} = (M'M)^{-1}M'Y$$

et nous avons donc :

$$\hat{\epsilon_i} = y_i - \hat{y_i} \text{ et}$$

$$\hat{\sigma}^2 = \frac{1}{t-p} \sum_{i=1}^{n} \hat{\epsilon_i}^2$$

Ce qui permet d'avoir :

$$\hat{\mu_{i,j}} = exp(\hat{\mu}) + \hat{\alpha_{i,j}} + \hat{\beta_{i,j}} + \frac{\sigma^2}{2}$$

et au final :

$$\hat{R} = \sum_{i+j>n} \hat{\mu}_{i,j}$$

3 Modélisation GLM

C'est un modèle à trois composantes : la composante aléatoire, la composante systématique et la fonction lien.

— La composante aléatoire : Nous supposons dans le GLM que notre variable aléatoire (variable réponse) suit une loi de probabilité appartenant à la famille exponentielle. Les densités de cette famille sont de la forme :

$$f(x_{i,j};\theta_{i,j};\phi) = exp(\frac{[\theta_{i,j}x_{i,j} - b(\theta_{i,j})]W_{i,j}}{\phi} + c(x_{i,j};\phi))$$

Avec :

$\theta_{i,j}$ le paramètre naturel de la loi
ϕ un paramètre de dispersion strictement positif
$W_{i,j}$ une pondération
b et c des fonctions caractéristiques du modèle, b étant deux fois dérivable à valeur dans R et c à valeur dans R^2

— La composante déterministe : ce sont les composantes déterministes du modèle en plus de l'inflation supposée constante. On note :

$$\eta_{i,j} = \mu + \alpha_i + \beta_j \text{ i,j} = 0,1,...,n$$
$$(\text{avec } \alpha_0 = \beta_0 = 0)$$

— La fonction lien g : (fonction réelle, strictement monotone et dérivable) telle que :

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

A titre illustratif, pour une fonction lien log , on aura :

$$\log(\mu_{i,j}) = \eta_{i,j} \iff \mu_{i,j} = e^{\mu+\alpha_i+\beta_j}$$

Les fonctions liens classiques sont les suivantes :

Fonction identité	g(z) = z
Fonction logarithme	g(z) = Log (z)
Fonction inverse	g(z) =1 / z
Fonction Logit	g(z) = Log (z / (1-z))
Fonction Probit	g(z) = Φ(z)

FIGURE 5.1 – Quelques fonctions liens classiques

Φ étant la fonction de répartition de la loi normale N(0,1)

Ces composantes définies nous permettent donc de dire que dans le cadre du GLM on aura :

$$\mu_{i,j} = E(X_{i,j}) = g^{-1}(\eta_{i,j}) \text{ et } Var(X_{i,j}) = \phi V(\mu_{i,j})$$

$V(\mu_{i,j})$ étant la fonction variance qui dépend de la famille de distribution utilisée dans le GLM.

A titre d'illustration, la fonction variance pour une loi normale N(μ,σ^2) vaut 1 et celle d'une loi gamma G(μ,σ^2) vaut μ^2

A toute loi de probabilité est associée un paramètre canonique. La fonction lien qui utilise ce paramètre est la fonction lien canonique. La plupart des modèles GLM se font à partir de la fonction lien canonique.

Distribution	**Fonction de lien canonique g(μ)**	**Fonction Variance V(μ)**
Normale N(μ,σ²)	μ	1
Poisson P(μ)	Log μ	μ
Gamma G(μ,ν)	-1 / μ	μ²
Binomiale Négative	Log (kμ / (1+ kμ))	μ+kμ²

FIGURE 5.2 – Quelques fonctions de lien canonique et de variance pour des distributions classiques

La variation de la provision estimée totale quant à elle peut être estimée par la méthode delta (méthode complexe à mettre en place et qui ne donne que des expressions asymptotiques de l'estimation de la variance), ou par la méthode bootstrap qui est simple à mettre en place et facile à comprendre.

3.1 Résidus du modèle

Les résidus (ou erreurs) observés sont définis comme étant les outils de mesure des différences ou des distances entre les valeurs observées et les valeurs estimées par un modèle de régression. Ils ont la particularité de représenter la partie non expliquée par l'équation de régression.
Il existe 3 types de résidus dans un GLM :

1. Les résidus lignes : ce sont les plus couramment utilisés. Il s'agit simplement ici de faire la différence entre la valeur observée (valeur réelle) et la valeur estimée par le GLM. Sa valeur sera donc ici : $r_{i,j} = X_{i,j} - \widehat{\mu_{i,j}}$, avec $E(r_{i,j})$ qui doit avoir une valeur proche de 0, $\widehat{\mu_{i,j}}$ étant un estimateur asymptotiquement sans biais de $\mu_{i,j}$.

2. Les résidus de Pearson : Ce sont des résidus lignes qu'on a standardisé, $r_{pi,j} = \frac{X_{i,j}-\widehat{\mu_{i,j}}}{V(\widehat{\mu_{i,j}})}$. Ces résidus sont censés suivre une loi normale centrée réduite et être indépendants.

3. Les résidus de déviance : Ils sont de la forme $r_{Di,j} = \sqrt{d_{i,j}}signe(X_{i,j} - \widehat{\mu_{i,j}})$, avec $d_{i,j}$ qui est la contribution de l'observation d'indices (i,j) à la déviance du modèle.

La déviance normalisée D^* est la comparaison entre le modèle saturé (un modèle ayant autant de paramètres que d'observations et dont les valeurs prédites sont égales aux valeurs réelles) et notre GLM. En notant $\tilde{L}$ la vraisemblance du modèle saturé et L celle du GLM, nous avons $D^* = -2(L - \tilde{L})$. Ainsi, en notant D comme étant la déviance du GLM nous avons que $D = \phi D^*$.

Distribution	φ	Déviance
Normale	σ^2	$\sum_i (y_i - \mu_i)^2$
Binomiale Négative	1	$2\sum_i \left[y_i \log(y_i / \mu_i) - (y_i + 1/k) \log\left(\frac{y_i + 1/k}{\mu_i + 1/k} \right) \right]$
Poisson	1	$2\sum_i \left[y_i \log(y_i / \mu_i) - (y_i - \mu_i) \right]$
Gamma	$1/\nu$	$2\sum_i \left[-\log(y_i / \mu_i) + \frac{y_i - \mu_i}{\mu_i} \right]$

FIGURE 5.3 – Calcul de déviance pour quelques distributions

3.2 Validation du modèle

Elle se fait principalement par l'examen des résidus donnés par le GLM. Dans un bon modèle GLM, les résidus sont censés suivre une loi normale (mais de variances différentes), il sera donc nécessaire d'utiliser des résidus normalisés pour pouvoir les comparer rigoureusement. Pour cela, nous utiliserons donc les résidus de Pearson définis précédemment. Il sera donc nécessaire ici d'effectuer sur ces résidus un test de normalité de type Shapiro-Wilk (Annexe D du mémoire) et un test d'indépendance : test du signe des différences (Annexe E du mémoire).

Dans la méthode de provisionnement par GLM, l'estimation de la variance de la provision globale se fait par la méthode delta, qui est une méthode difficile à mettre en place : nous opterons donc ici à l'estimation de la variance par la méthode de *bootstrap*.

4 Modèle Glm-Bootstrap

En statistiques, les techniques de bootstrap sont des méthodes d'inférence statistique basées sur la réplication multiple des données à partir du jeu de données étudié, selon les techniques de rééchantillonnage. Elles datent de la fin des années 1970, époque où la possibilité de calculs informatiques intensifs devient abordable.
Dans cette méthode il est nécessaire que notre échantillon de départ soit indépendant et identiquement distribué, ce qui n'est en général pas le cas des incréments des triangles de paiements. Dans ce cas, le rééchantillonage se fera donc au moyen des résidus de Pearson.
Algorithme de la méthode :

1. On estime par le GLM les valeurs du triangle inférieur : les valeurs $\widehat{\mu_{i,j}}$ du GLM
2. On calcule par la suite les résidus de Pearson $r_{pi,j} = \frac{X_{i,j} - \widehat{\mu_{i,j}}}{V(\widehat{\mu_{i,j}})}$ et on estime le paramètre de dispersion à l'aide des résidus de Pearson
$$\widehat{\phi} = \frac{1}{N-p} \sum_{i+j \leq n} (r_{pi,j})^2.$$
N étant le nombre d'éléments du triangle supérieur et p le nombre de paramètres inconnus.
3. On repète B fois (b = 1,...,B) les opérations suivantes :
   — On tire au hasard un résidu de Pearson du modèle initial parmi B résidus (tirés avec remise parmi les résidus de Pearson du modèle intitial) qu'on va appeler : $({r_{Pi,j}}^{*b})$
   — On détermine le nouveau triangle d'incréments tenant en compte le résidu en question $({x_{i,j}}^{*b})$ donné par :
$$({x_{i,j}}^{*b}) = \widehat{\mu_{i,j}} + ({r_{Pi,j}}^{*b}) \sqrt{V(\widehat{\mu_{i,j}})}$$

— On calcule les valeurs du nouveau triangle inférieur à partir de celles du nouveau triangle supérieur : $(\widehat{\mu_{i,j}}^{*b})_{i+j\geq n+1}$ (en lui appliquant de nouveau le GLM utilisé initialement) et on obtient donc le provision globale du nouveau triangle

$$\widehat{R}^{*b} = \sum_{i+j\geq n+1} \widehat{\mu_{i,j}}^{*b}$$

— on stocke $\widehat{R}^{*b}$ et on répète.

4. On obtient donc un vecteur de B provisions globales $(\widehat{R}^{*1}, ..., \widehat{R}^{*B})$

Ainsi on obtient les valeurs suivantes :

— $E_{boot}(\widehat{R}) = \frac{1}{B}\sum_{b=1}^{B} \widehat{R}^{*b}$

— Le risque d'estimation : c'est l'erreur provenant de l'application du Bootstrap

$$MSE_{boot}(\widehat{R}) = \frac{1}{B-1}\sum_{b=1}^{B}[\widehat{R}^{*b} - E_{boot}(\widehat{R})]^2$$

— Le risque de processus : c'est l'erreur provenant de l'estimation des incréments du triangle de provisionnement

$$Var_{boot}(R) = \widehat{\phi}\sum_{i+j\geq n+1} V(\widehat{\mu_{i,j}})$$

— L'erreur de prédiction : qui combine les risques d'estimation et de processus

$$sep_{boot}(\widehat{R}) = \sqrt{MSE_{boot}(\widehat{R}) + Var_{boot}(R)}$$

4.1 Application

Nous appliquons dans cette partie à viser comparative le GLM avec la famille Gamma et la régression log-normale sur nos données, combinés à un bootstrap de 10 000 simulations (qui nous donnera plus tard la distribution empirique des provisions sur chaque lob).
Le récapitulatif des résultats obtenus est donné ci-dessous.

Pour la régression Log-Normale nous avons :

	lob 4	lob 5	lob 7	Total
Espérance (régression)	28 417 299	6 470 789	55 737 683	90 625 771
Espérance (régression + bootstrap)	29 563 478	7 525 697	54 958 698	92 047 873
Erreur d'estimation	23 475 658	6 452 987	35 458 976	65 387 621
Erreur de Procéssus	312 290	234 424	965 095	1 511 809
Erreur de prédiction	23 477 735	6 457 244	35 472 107	65 407 086
Erreur de prédiction relative	79%	86%	65%	71%
Var90	58 938 313	15 719 810	99 896 825	174 554 948
Var95	67 735 785	17 989 059	113 395 749	199 120 593
Var99.5	90 774 383	23 597 927	148 295 180	262 667 490
Shapiro-Wilk (test) : p-value	0.052	0.034	0.046	
Signe des différenceq (test) : p-value	0.15	0.15	0.63	
AIC	3758	3334	3674	
R^2	0.648	0.724	0.58	

TABLE 5.1 – Résultats de la régression Log-Normale sur nos lobs

Pour la régression Gamma nous avons :

	lob 4	lob 5	lob 7	Total
Espérance (régression)	26 399 006	4 556 554	49 034 213	79 989 773
Espérance (régression + bootstrap)	26 526 151	4 635 471	48 940 735	80 102 357
Erreur d'estimation	16 947 552	5 327 233	26 584 478	48 859 263
Erreur de Procéssus	3668	2891	4653	10 924
Erreur de prédiction	16 947 553	5 327 234	26 584 479	48 859 266
Erreur de prédiction relative	64%	115%	54%	61%
Var90	42 989 708	8 966 358	83 836 365	135 792 431
Var95	51 797 914	17 917 593	105 510 553	175 226 060
Var99.5	104 084 783	29 300 300	144 550 579	277 935 662
Shapiro-Wilk (test) : p-value	0.06	0.062	0.05	
Signe des différenceq (test) : p-value	0.87	0.87	0.874	
AIC	3464	2713	3572	
R^2	0.69	0.789	0.838	

TABLE 5.2 – Résultats de la régression Gamma sur nos lobs

Une première vérification qui est à faire est celle d'observer que les valeurs de provisions espérées par le GLM tout seul et celles espérées par la combinaison du GLM et du Bootstrap soient assez similaires sur l'ensemble des lobs, ce qui est bien le cas de nos données : en utilisant la méthode GLM toute seule on obtient une provision quasiment similaire au cas où on utiliserait la méthode GLM + Bootsrap.
La deuxième chose que l'on peut constater est qu'en utilisant la distribution gamma sur la lob 5, l'erreur de prédiction est très important(plus grande que l'espérance). Ainsi, nous proposons une analyse rétrospective de la méthode GLM qui consiste à estimer les écarts entre les incréments des triangles supérieurs réels et estimés : les résultats sont donnés en annexe F du mémoire (Les résultats sont présentés uniquement sur la Lob 4).A partir de cette analyse, nous ne sommes pas capable de définir quelle est la loi la plus adaptée pour la modélisation de notre GLM.
Nous analysons donc graphiquement la vérification de l'hypothèse de normalité des résidus de Pearson pour le GLM (log-normale) et le GLM (Gamma) proposés en annexe G du mémoire : les résidus de Pearson sont censés être alignés sur la droite d'équation x = y du repère. On constate donc à travers ces graphiques et par le test de Shapiro-Wilk que dans les deux cas cet hypothèse ne se vérifient pas totalement, mais est plus acceptable dans le cas de la modélisation Glm avec loi gamma que dans celle de la régression log-normale.

Au vu des différents critères de vérification des hypothèses du modèle, de normalité des résidus de Pearson et d'ajustement avec les données observées, nous choisirons pour la suite de nos travaux d'utiliser le GLM avec famille gamma dans l'optique de modélisation de dépendance par les copules.
Nous devons noter que dans la pratique d'un arrêté de comptes, il serait indispensable de comprendre économiquement les écarts enre les différentes espérances obtenues avant d'arbitrer.

Les méthodes de provisionnement appliquées précédemment et les résultats de provision globale obtenus par somme ne considèrent pas de prise en compte de dépendance. Dans la suite de notre travail, nous allons nous intéresser à la modélisation de la dépendance entre les lobs et à l'analyse de l'impact de cette dépendance sur la provision espérée mais surtout sur la *value-at-risk* de la provision globale.
Nous allons aussi supposer dans la suite de notre travail que le risque de processus sur nos lobs peut être négligé (au vu des résultats obtenus) et ainsi faire l'hypothèse selon laquelle la distribution de la variable aléatoire des provisions est donnée par les 10 000 simulations *bootstrapées*(prise en compte uniquement du risque d'estimation qui est plus conséquent), ainsi on pourra faire des tests de lois sur la probable distribution que pourraient suivre nos provisions pour chaque lob (au lieu de faire des hypothèses fortes).

Troisième partie

Dépendance entre les risques

Chapitre 6

Introduction à la dépendance

Cardif IARD étant un assureur multi-branches, est confronté à une problématique de dépendance entre certaines des lobs de son portefeuille. De façon générale dans le domaine de l'assurance, ne pas considérer une quelconque dépendance entre les risques peut être la cause d'une mauvaise estimation du montant des provisions. Dans le cas d'une tempête par exemple, on observe une dépendance entre les lobs dommage auto et dommages aux biens. Il est donc important pour la compagnie d'assurance de pouvoir étudier et modéliser cette dépendance, ainsi que de comparer les hypothèses de dépendance au cas où les estimations sont faites individuellement sur chaque lob pour savoir quel est l'impact de celle-ci sur le montant de ses réserves.

1 Distributions multivariées

Pour pouvoir bien comprendre la notion de dépendance, il est important de comprendre la notion de « distributions multivariées » car cette notion est la base de l'étude de la dépendance entre variables aléatoires. Nous allons nous intéresser ici aux variables aléatoires continues (par opposition aux variables aléatoires discrètes).

1.1 Fonction de répartition jointe

Soient $X_1, ..., X_k$ des variables aléatoires continues (k étant un entier naturel strictement plus grand que 1). Alors la fonction de répartition du k-uplet $(X_1, ..., X_k)$ est définie par :

$$F(x_1, ..., x_k) = P(X_1 \leq x_1, ..., X_k \leq x_k)$$

Les fonctions de répartition marginales quant à elles sont les fonctions de répartitions de chacune des composantes du vecteur prise individuellement. Pour une composante donnée Xi du k-uplet, elle est définie par :

$$F_i(x_i) \rightarrow P(X_i \leq x_i) = F(\infty, \infty, .., x_i..., \infty, \infty)$$

1.2 Fonction de densité jointe

Soient $X_1, X_2, ..., X_k$ des variables aléatoires continues, alors :

$$f(x_1, ..., x_k) = \frac{\partial^n F(x_1,...,x_k)}{\partial x_1, \partial x_2, ..., \partial x_k}$$

est la fonction de densité jointe de $X_1, X_2, ..., X_k$ si :

1. $f(x_1, ..., x_k) \geq 0$
2. $\int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(x_1, ..., x_k) dx_1 ... dx_k = 1$
3. Soit A un ensemble de dimension k $P[(X_1, X_2, ..., X_k) \in A] = \int^A f(x_1, ..., x_k) dx_1 ... dx_k$

2 Dépendance

2.1 Dépendance vs Corrélation

Bien que très souvent confondues, les notions de dépendance et de corrélation sont différentes, mais elles sont tout de même associées.
Soient deux variables aléatoires réelles X et Y : si X et Y sont indépendantes, alors elles sont non corrélées(linéairement), la réciproque est cependant fausse de façon générale (sauf dans le cas où les distributions sont toutes normales).

Montrons que la non-corrélation n'implique pas l'indépendance :

On se donne X suivant une loi normale N(0,1) et on crée la variable aléatoire réelle $Y = X^2$. De toute évidence, X et Y sont deux variables liées fonctionnellement : il suffit de représenter le graphe de la fonction $f(x) = x^2$ pour s'en convaincre (en annexe H du mémoire) car on voit que f(x) est strictement monotone sur R_+ et sur R_-.Nous avonce cependant que $\rho(X, Y) = 0$.
Car :

$$X \sim N(0,1) \Rightarrow \text{E[X]} = 0 \; et \text{Var(X)} = 1.$$

Or,

$$Var(X) = E(X^2) - E(X)^2 = E(X^2) = 1 \; .$$

Aussi,

$$Var(Y) = E(Y^2) - E(Y)^2 = E(X^4) - E(X^2)^2 = E(X^4) - 1 = 3 - 1 = 2$$

(calcul fait avec la fonction de densité de la loi normale : soit g la densité et f une fonction continue,$E[f(X)] = \int f(x)g(x)dx$)
D'où :

$$Cov(X, Y) = E([X - E[X]][Y - E[Y]]) = E[XY] = E[X^3] = 0$$

$$\Rightarrow \rho(X, Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = 0 \; .$$

$\rho(X, Y)$ étant le coefficient de corrélation de Pearson.

Les variables X et Y sont donc non corrélées mais sont tout de même dépendantes. Cet exemple illustre également le fait que la corrélation de Pearson ne capture que les dépendances linéaires. La dépendance quant à elle prend en compte les liaisons linéaires et non-linéaires, elle permet d'expliquer la structure complète de la liaison. L'outil commun de mesure de dépendance est le coefficient de corrélation linéaire, mais il perd de sa pertinence en dehors du cadre gaussien (car il n'est plus capable d'expliquer toute la structure de dépendance en dehors de ce cadre-là) il est donc nécessaire d'étudier quels pourraient être les autres outils de mesure de dépendance à notre disposition.

2.2 Outils de mesure de dépendance

Définition : Une mesure de dépendance est une application qui associe à deux variables aléatoires un réel permettant de quantifier la force de dépendance qui lie ces deux variables aléatoires.

Coefficient de Pearson

Soient deux variables aléatoires réelles X et Y, soient X_i, Y_i (i = 1,..,n) leurs réalisations respectives. Le coefficient de corrélation de Pearson entre X et Y s'écrit :

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

Avec :

— $Cov(X,Y) = E([X - E[X]][Y - E[Y]]) = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$

— $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

— $Var(X) = E[X^2] - E[X]^2 = Cov(X,X)$

Nous avons alors que :

— X et Y sont corrélées linéairement $\Leftrightarrow \rho(X,Y) \neq 0$

— X et Y sont non corrélées linéairement $\Leftrightarrow \rho(X,Y) = 0$

Comme mentionné précédemment, c'est un coefficient qui n'explique bien que la structure de dépendance linéaire (qui est la structure complète uniquement dans le cas de marginales toutes gaussiennes). Cependant, la dépendance entre variables aléatoires n'est pas que linéaire, nous verrons donc par la suite les autres outils de mesure de dépendance entre les risques qui prennent en compte le rang de nos observations et basées sur des notions de concordance et discordance car ils ont une interprétation différente de la structure de dépendance que celle du coefficient de Pearson.

Notions de concordance et discordance

Soit un couple de variable aléatoire (X,Y), soient (x_1, x_2) et (y_1, x_2) deux observations du vecteur (X,Y). Ces observations sont dites concordantes si $(x_1 - x_2)(y_1 - y_2) > 0$ et discordantes sinon.

Rang des observations

Au sein d'une série d'observations, le rang d'une série de données, le rang d'une observation correspond au nombre d'observations inférieures ou égales à celle-ci.
Soit $(X_1, ..., X_n)$ un vecteur aléatoire réel, alors :

$$Rang(X_i) = Cardinal\ \{j \neq i : X_j \leq X_i\}\, \forall i, j = 1, .., n$$

Il est à noter que la concordance des observations et la concordance des rangs est équivalente :
En notant $R_i = Rang(X_i)$ et $S_j = Rang(Y_j)$:

$$(R_1 - R_2)(S_1 - S_2) > 0 \Leftrightarrow (x_1 - x_2)(y_1 - y_2) > 0$$

le "Rho" de Spearman

Le "Rho" de Spearman ρ_s se définit comme étant la corrélation entre les rangs des observations. Soient $X = (X_1, .., X_n)$ et $Y = (Y_1, .., Y_n)$ deux vecteurs aléatoires réels et $R = (R_1, .., R_n)$ et $S = (S_1, .., S_n)$ les vecteurs aléatoires des rangs respectivement de X et Y, alors :

$$\rho_s(X,Y) = \rho_s(R,S) = \frac{Cov(R,S)}{\sqrt{Var(R)Var(S)}} = \frac{\sum_{i=1}^{n}(R_i-\bar{R})(S_i-\bar{S})}{\sqrt{\sum_{i=1}^{n}(R_i-\bar{R})^2 \times \sum_{i=1}^{n}(S_i-\bar{S})^2}}$$

ρ_s vérifie les propriétés suivantes :

1. $\rho_s(X,Y) = \rho_s(Y,X) \forall (X,Y)$

2. $|\rho_s(X,Y)| \leq 1 \forall (X,Y)$

3. X et Y co-monotones $\Leftrightarrow \rho_s(X,Y) = 1$

   X et Y anti-monotones $\Leftrightarrow \rho_s(X,Y) = -1$

4. Si g est une fonction strictement croissante alors :

   $\rho_s(g(X), g(Y)) = \rho_s(X,Y)$

Enfin comme dans la mesure de Pearson il n'y a toujours pas équivalence entre indépendance et non-corrélation.

Le "Tau" de Kendall

Soit (X,Y) un vecteur aléatoire et (X',Y') une copie de ce vecteur. Le tau de Kendall τ est défini comme étant la différence entre les probabilités de concordance et de discordance entre (X,Y) et (X',Y') :

$$\tau(X,Y) = P[(X-X')(Y-Y') > 0] - P[(X-X')(Y-Y') < 0].$$

τ vérifie également les 4 propriétés précédentes (du "Rho" de Spearman) et aussi, il n'y a pas d'équivalence entre indépendance et non corrélation : X et Y indépendants $\Rightarrow \tau(X,Y) = 0$.

2.3 Les types de dépendance

Bien que deux variables aléatoires n'aient qu'une seule manière d'être indépendante, la dépendance quant à elle peut être de plusieurs formes. Ces différentes formes constituent la structure de dépendance entre les deux variables en question. On distinguera ici deux grands types de dépendance : la dépendance de queue et la dépendance parfaite.

La dépendance parfaite

On parle de dépendance parfaite lorsque deux risques peuvent s'écrire comme des fonctions croissantes ou décroissantes d'une même variable aléatoire sous-jacente. Il existe essentiellement deux types de dépendance parfaite : la comonotonicité et la contre-comonotonicité.

— **La comonotonicité :** C'est une notion qui fait principalement référence à la dépendance positive parfaite entre les composantes d'un vecteur aléatoire, en disant qu'elles peuvent être représentées comme des fonctions croissantes d'une seule variable aléatoire.
Le couple (X_1, X_2) est dit comonotone s'il existe des fonctions non-décroissantes g_1 et g_2 et une variable aléatoire Z telles que (X_1, X_2) et $(g_1(Z), g_2(Z))$ aient la même loi. Dans ce cas on parle de dépendance totale positive

— **La Contre-comonotonicité :** Dans ce cas, contrairement à la comonotonicité, la dépendance est parfaitement négative.
Le couple (X_1, X_2) est dit contre-comonotone ou antimonotone s'il existe une fonction non-décroissante g_1, une fonction non-croissante g_2 et une variable aléatoire Z telles que (X_1, X_2) et $(g_1(Z), g_2(Z))$ aient la même loi.

La dépendance de queue

C'est une mesure de dépendance qui est spécifique à l'étude des dépendances sur les valeurs extrêmes. Le coefficient de dépendance de queue mesure la probabilité conditionnelle que des valeurs extrêmes se réalisent simultanément au niveau des queues de distribution.
Pour illustrer cela mathématiquement, prenons deux variables aléatoires X et Y, qui représentent chacune une perte potentielle dans un portefeuille d'assurance. Si X et Y sont indépendantes, alors la probabilité d'une perte importante dans X n'affecte pas la probabilité d'une perte importante dans Y. Cependant, si X et Y ont une dépendance de queue, alors la probabilité d'une perte importante dans X augmente la probabilité d'une perte importante dans Y. Le coefficient de dépendance de queue supérieure de deux variables aléatoires continues X et Y ayant pour fonction de distribution respectives F_X et F_Y, est défini par la limite (si elle existe)

$$\lambda_U = \lim_{u \to 1} P(X > F_X^{(-1)}(u) / Y > F_Y^{(-1)}(u))$$

Avec $F^{(-1)}(t) = inf\,\{x/F(x) \geq t\}$
.

Donc λ_U représente la probabilité qu'une variable aléatoire dépasse une valeur u, sachant que l'autre variable est également au-dessus de u, dans la limite de u proche de 1.

Et de la même manière, on définit le coefficient de dépendance de queue inférieure dans la limite des quantiles extrêmes proche de 0 :

$$\lambda_L = \lim_{u \to 0} P(X \leq F_X^{(-1)}(u) / Y \leq F_Y^{(-1)}(u))$$

On dit que X et Y sont asymptotiquement dépendantes au niveau supérieur de la queue de distribution si $u \in [0, 1]$ et asymptotiquement indépendantes au niveau supérieur de la queue de distribution si u = 0. On peut exprimer λ_U et λ_L en terme de la copule C.

Plusieurs outils de mesure de dépendance ont été mis sur pieds mais malheureusement, ils ne permettent que d'avoir une "indication" sur le type et le sens de la dépendance, ils ne peuvent pas à eux seules être utilisés pour une modélisation de toute la structure de dépendance entre variables aléatoires car pour aucun d'entre eux il n'y a d'équivalence entre indépendance et nullité du coefficient de dépendance. Un coefficient de mesure de dépendance nul (dans le cas Pearson, Spearman ou Kendall), ne nous donnera jamais d'information complète sur la structure de dépendance par exemple. Plusieurs méthodes ont donc été développées afin de pouvoir pallier ce problème et pouvoir avoir la structure complète de dépendance entre variables.

Chapitre 7

Méthodes de modélisation de la dépendance

Les différentes problématiques posées par les outils de modélisation de dépendance présentés précédemment ont donné lieu ces dernières années au développement de lois multivariées caractérisées soit par la loi jointe soit par les lois marginales et la structure de dépendance entre les marginales. Ainsi, on pourra donc distinguer deux approches de construction de modélisation de loi multivariée : à partir de la loi jointe ou à partir des marginales.

1 L'approche Top-Down

L'idée de cette méthode est de pouvoir déduire de la loi jointe $F_{X,Y}$ la structure de dépendance entre les marginales et aussi les lois F_X et F_Y de celles-ci. Les modèles de choc commun par exemple sont construits à partir de cette méhode.

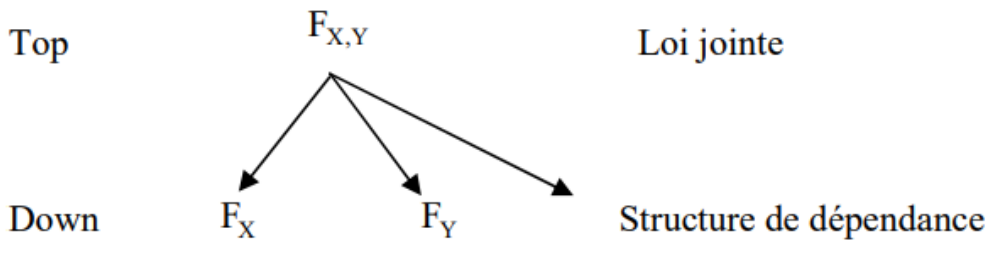


FIGURE 7.1 – Illustration de l'approche top-down

Pour plus d'éléments sur les modèles à choc commun, le lecteur pourra se référer au mémoire de Géraldine Krauth (en bibliographie de mémoire).

Le principal inconvénient de cette méthode est qu'elle est difficilement extensible à une dimension supérieure à 2 et aussi le fait que les marginales doivent être de même nature ce qui n'est pas toujours le cas en assurance non-vie car dépendant de la nature des lobs, les provisions peuvent suivre des lois différentes.

Nous ne privilégions pas cette approche ; notre but étant d'avoir une provision consolidée sur les différents lobs, nous allons donc nous intéresser à l'approche inverse de celle-ci : l'approche Bottom-Up.

2 L'approche Bottom-Up

Dans cette approche, on construit le modèle à partir de la loi des marginales et de la structure de dépendance à priori.

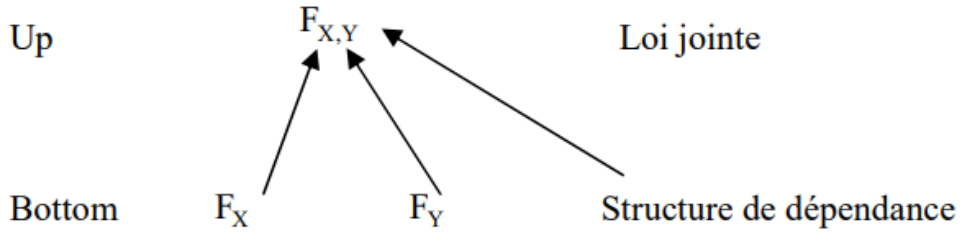


FIGURE 7.2 – Illustration de l'approche Bottom-Up

Il est donc nécessaire d'avoir que :

— $F_X(x) = P(X < x) = \lim_{y \to +\infty} F_{X,Y}(x, y)$
— $F_Y(y) = P(Y < y) = \lim_{x \to +\infty} F_{X,Y}(x, y)$
— La dépendance induite par $F_{X,Y}$ soit la même que celle donnée à priori.

La mise en place de cette approche est donc possible grâce à un outil appelé Copule. Le grand avantage ici est que non seulement toute le structure de dépendance pourra être expliquée, mais aussi que nos marginales ont la liberté d'être différentes les unes des autres, ce qui en fait un outil de choix dans la construction de ce type de modèle.

Quatrième partie

Introduction à la notion de Copule

Chapitre 8

Présentation et propriétés

L'apparition du concept des copules est l'œuvre de Sklar en 1959 suite à un problème énoncé par Maurice Fréchet dans le cadre des espaces métriques aléatoires. Cependant, l'étude de cette notion de manière plus approfondie a débuté au milieu des années 1980 avec Christian Genest et son équipe. Son intérêt est qu'elle permet une modélisation flexible de la dépendance entre deux ou plusieurs variables aléatoires (asymétrie, dépendance de queue etc).

1 Définition d'une copule

Une Copule est une fonction liant des marginales pour en générer une loi jointe. L'idée ici est de pouvoir séparer la structure de dépendance (représentée par la Copule) des marginales. C'est donc une fonction de répartition multivariée ayant des lois marginales uniformes sur [0,1].

$C(u_1, ..., u_n) = P(U_1 \leq u_1, ..., U_n \leq u_n)$ définie de $[0,1]^n$ vers $[0,1]$, et vérifiant :

1. $\forall i \in (1, .., n), C(u_1, ...u_n)$ fonction croissante en u_i

2. $\forall i \in (1, .., n), C(1, ..., 1, u_i, 1, ...1) = u_i$

3. $\forall (a_1, ...a_n), (b_1, ...b_n) \in [0,1]^2$ et $a_i \leq b_i$,

   $\sum_{i_1=1}^{2} ... \sum_{i_n=1}^{2} (-1)^{i_1+...+i_n} C(u_{1i_1}...u_{ni_n}) \geq 0$ avec $u_{j_1} = a_j$ et $u_{j_2} = b_j$, $j \in (1, ..., n)$

1.1 Exemple de Copule

Soit $u, v \in I = [0,1]$, alors C(u,v) = min(u,v) est une copule de dimension 2 et nous avons :

- $\forall u, v \in I, min(u, 0) = min(0, u) = 0$
- $\forall u, v \in I, min(u, 1) = min(1, u) = u$
- $min(v_1, v_2) - min(u_1, v_2) - min(v_1, u_2) + min(u_1, u_2) \geq 0, \forall (u_1, u_2) \in [0,1]^2, (v_1, v_2) \in [0,1]^2$ tq $u_1 \leq u_2$, $v_1 \leq v_2$, $u_1 \leq v_1$, $u_2 \leq v_2$. (Car $min(u_2, v_2) \geq min(u_2, v_1)$ et $min(u_1, v_2) \geq min(u_1, v_1)$)

Ainsi, les 3 propriétés de la définition d'une copule sont respectées pour la fonction C(u,v) = min(u,v) ce qui fait d'elle une copule.

1.2 Propriétés d'une copule

Théorème de Sklar :

C'est un théorème fondamental dans la théorie des copules, il permet d'établir le lien crée par la copule entre les fonctions de répartition marginales et la fonction de répartition multivariée.

Il s'énonce comme suit :
Soit F une fonction de répartition N-dimensionnelle dont les lois marginales sont F_1,.....,F_N. Il existe une copule C de dimension N telle que pour (x_1,..,x_N) $\in \bar{R}^N$ on ait :

$$F(x_1, ...x_N) = C(F_1(x_1), ..., F_N(x_N))$$

Si les lois marginales F_1,.....,F_N sont continues, la copule C est unique, sinon elle est déterminée de manière unique sur $Im(\mathrm{F}_1)^1 \times ... \times Im(\mathrm{F}_N)^1$. Dans le cas de lois marginales continues, pour tout $\underline{\mathrm{u}} \in [0\,;1]^N$.

Théorème d'invariance :

Soient N variables aléatoires continues X_1,.....,X_N de fonctions de répartitions marginales respectives F_1,.....,F_N. Soit F la fonction de répartition du vecteur aléatoire (X_1,.....,X_N) et $C_{X_1,..,X_N}$ la copule associée. En considérant h_1.... h_n comme étant N fonctions strictement croissantes sur $Im(X_1)^1$ ....$Im(X_n)^1$ respectivement, alors :

$$\forall u_1....u_n \in [0,1], C_{h_1(X_1),.....,h_n(X_n)}(u_1, ..., u_n) = C_{X_1,...,X_n}(u_1, ..., u_n)$$

La copule est donc invariante par transformations strictement croissantes des variables aléatoires.

Ce théorème nous permet d'avoir de la flexibilité dans la modélisation de la dépendance par copules dans le sens où,il n'est pas obligatoire de modéliser la dépendance par copules de travailler directement avec les variables aléatoires en présence, il est possible d'utiliser leurs transformations. Le dépendogramme entre X et Y est donc le même que celui de g(X) et g(Y).
[1]

Pour illustrer, soit A suivant une loi gamma(3,0.5) et Y suivant une loi Exp(2). Soit C = A^2 et D = exp(B) on simule 1000 fois les différentes variables dont on calcule les $pseudo - observations$[2] pour obtenir les dépendogrammes suivants :
[2]

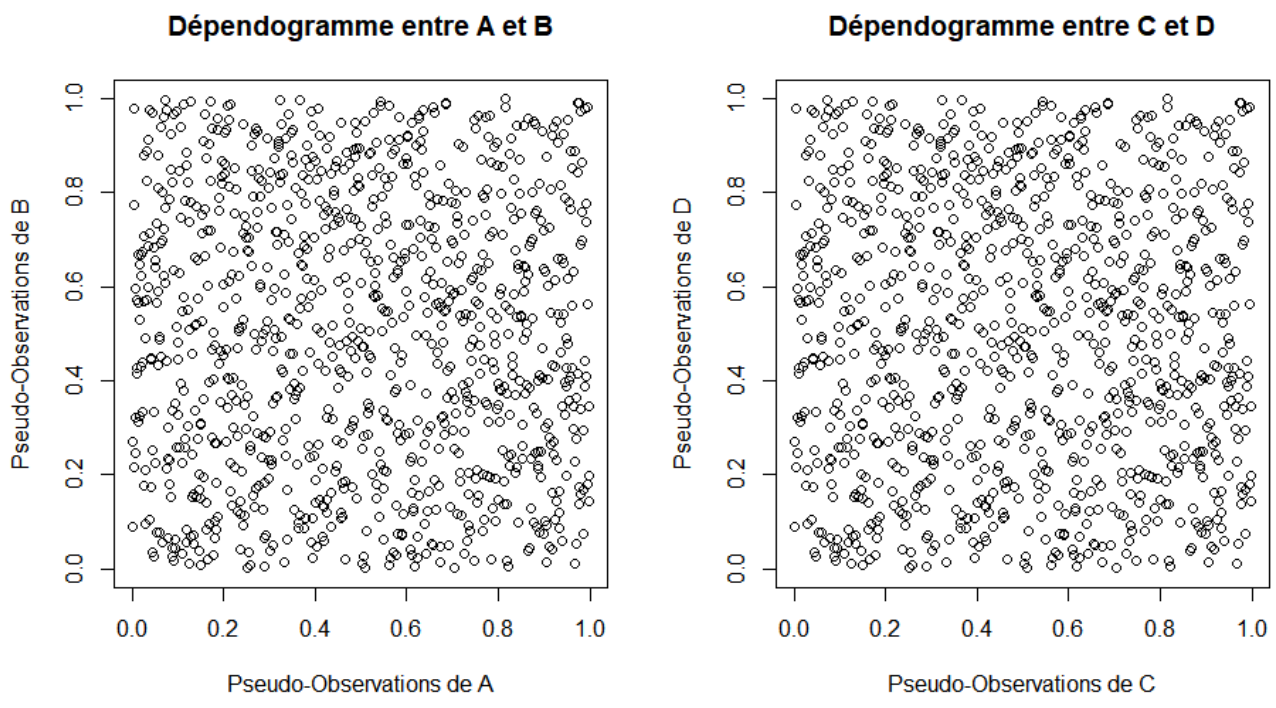


FIGURE 8.1 – Illustration du théorème d'invariance des copules

On se rend donc compte que les dépendogrammes sont identiques.

1. $Im(X) = \{y/\exists\omega \in \Omega : y = X(\omega)\}$
2. Les pseudo-Observations : soient n réalisations d'une variable aléatoire réelle X :($\mathrm{x}_1, ..., x_n$), $lespseudo - observations\mathrm{u}_i, i \in [1, n]$ désignent le rang de l'observation divisé par le nombre d'observations présentes, nous avons donc $\forall i \in [1, n], u_i \in [0, 1]$

2 Lien entre copules et mesures de dépendance

Il existe un lien entre les copules et les mesures de dépendances (tau de Kendall et rho de Spearman). Ce lien permet d'effectuer des tests d'indépendance entre les variables aléatoires étudiées.

2.1 Lien avec le rho de Spearman

$\hat{\rho_s}$ est un estimateur sans biais de :

$$\rho_s = 12 \int_{[0,1]^2} u.v \, \mathrm{d}C(u,v) - 3$$

Sous l'hypothèse nulle d'indépendance entre deux variables aléatoires réelles X et Y, la distribution de ρ_s est proche de $N(0, \frac{1}{n-1})$ et on rejette H_0 au niveau $\alpha = 5\%$ si $\sqrt{n-1}|\rho_s| > z_{\frac{\alpha}{2}} = 1.96$, avec $P(X \leq z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$ quand X suit la loi N(0,1).

2.2 Lien avec le Tau de Kendall

$\hat{\tau}$ est un estimateur sans biais de :

$$\tau = 4 \int_{[0,1]^2} C(u,v) \, \mathrm{d}C(u,v) - 1$$

Sous l'hypothèse nulle d'indépendance entre deux variables aléatoires réelles X et Y, la distribution de τ est proche de $N(0, \frac{2(2n+5)}{(9n(n-1))})$.On rejette H_0 au niveau $\alpha = 5\%$ si $\sqrt{\frac{9n(n-1)}{(2(2n+5))}}|\tau| > z_{\frac{\alpha}{2}} = 1.96$

3 Les Familles de Copules

Il est possible de construire une multitude de copules à condition de respecter les propriétés énoncées dans la partie précédente. De ce fait, il existe des sous familles de copules ayant des similitudes plus ou moins fortes parmi lesquelles : les copules archimédiennes et les copules elliptiques qui sont les familles de copules les plus couramment utilisées.

3.1 Les Copules archimédiennes

Ce sont des copules qui permettent de décrire des structures de dépendance diverses dont la dépendance asymétrique : dépendance où les coefficients de queue inférieure et supérieure sont différents.
Soit ϕ une fonction décroissante convexe sur $[0,1] \to [0,+\infty]$ telle que $\phi(1) = 0$ et $\phi(0) = +\infty$.
On appelle copule archimédienne stricte de générateur ϕ la copule définie par :

$$C(u,v) = \phi^{-1}(\phi(u) + \phi(v)), u, v \in [0,1]$$

Dans le cas mutidimensionnel, la copule Archimédienne n-dimensionnelle de générateur ϕ est alors définie par :

$$C(u_1 ..., u_n) = \phi^{-1}(\sum_{l=1}^{n} \phi(u_l))$$

Il en existe 3 : Les copules de Frank,Clayton et Gumbel

Ayant une définition des copules Archimédiennes multidimensionnelle, nous allons pour plus de simplicité dans les formules présenter les différentes Copules Archimédiennes ci-dessous en dimension 2.

Copule de Franck

Soit $\phi(x) = -log(\frac{e^{-\theta x}-1}{e^{-\theta}-1})$ avec $\theta \neq 0$ considéré comme étant le générateur de la copule de Frank. Alors, la copule de Frank est définie comme suit :

$$C_\theta(u,v) = \frac{-1}{\theta} log(1 \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1})$$

Avec :

— $\lim\limits_{\theta \to +\infty} C_\theta = C^+$

— $\lim\limits_{\theta \to 0} C_\theta = \Pi$

— $\lim\limits_{\theta \to -\infty} C_\theta = C^-$

Copule de Clayton

Soit $\phi(t) = \theta(t^{\frac{-1}{\theta}} - 1), \theta > 0$, considéré comme étant le générateur de la copule de Clayton. Alors, La copule de Clayton est définie comme étant :

$$C_\theta(u,v) = (u^{\frac{-1}{\theta}} + v^{\frac{-1}{\theta}} - 1)^{-\theta}$$

Avec :

— $\lim\limits_{\theta \to +\infty} C_\theta = C^-$

— $\lim\limits_{\theta \to 0} C_\theta = \Pi$

Copule de Gumbel

Soit $\phi(t) = (-log(t))^\theta, \theta > 1$, considéré comme étant le générateur de la copule de Gumbel. Alors, la copule de Gumbel est définie comme étant :

$$C_\theta(u,v) = exp(-(-log(u))^\theta + (-log(v))^\theta)^{\frac{1}{\theta}}$$

Avec :

— $\lim\limits_{\theta \to +\infty} C_\theta = C^+$

— $\lim\limits_{\theta \to 1} C_\theta = \Pi$

Nous allons illustrer graphiquement la dépendance induite par les trois copules pour un paramètre $\theta = 3$ en visualisant leurs densités et les dépendogrammes associés.

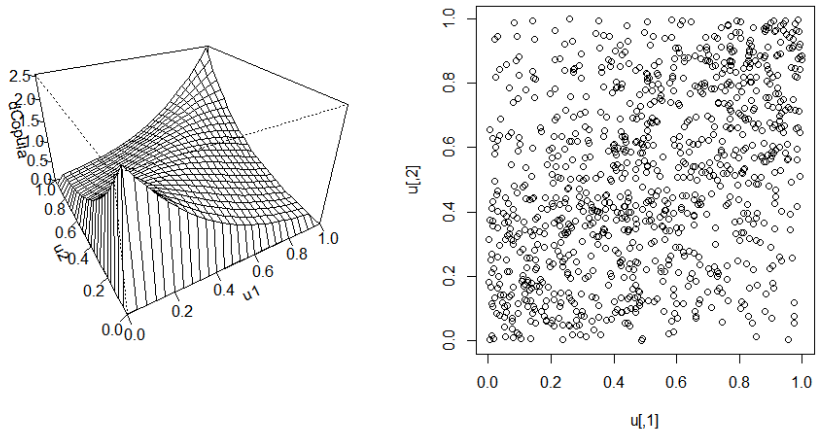


FIGURE 8.2 – Densité et dépendogramme pour une copule de Frank de paramètre 3

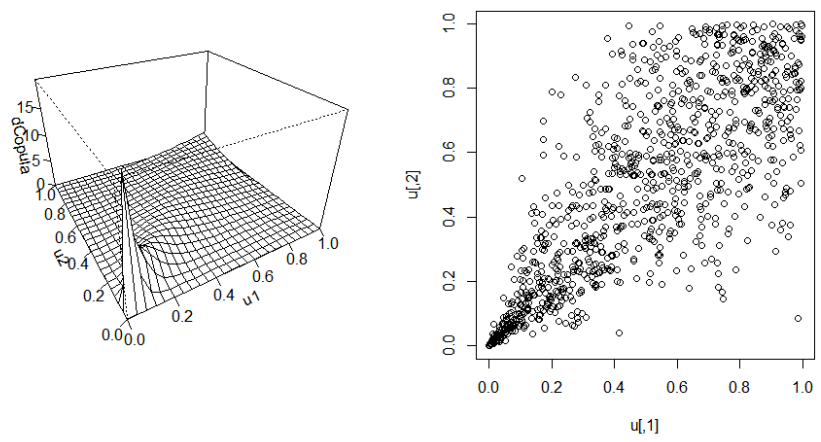


FIGURE 8.3 – Densité et dépendogramme pour une copule de Clayton de paramètre 3

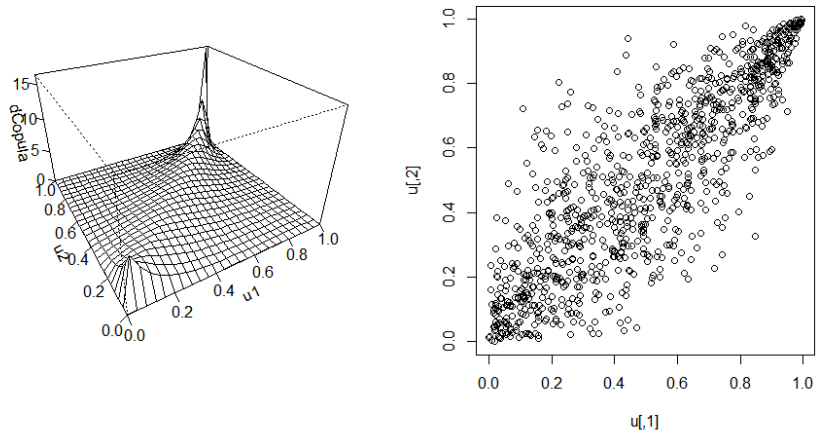


FIGURE 8.4 – Densité et dépendogramme pour une copule de Gumbel de paramètre 3

Les graphiques de densité et dépendogrammes montrent clairement que le type de dépendance varie en fonction de la copule choisie à paramètre égal.
Pour la copule de Clayton, on observe que la queue inférieure est accentuée, pour la copule de Gumbel c'est la queue supérieure qui est accentuée avec une grande densité aux extrêmes, et pour la copule de Frank nous avons une densité élevée aux extrêmes sans accentuation sur les queues.
Le tableau ci-dessous est un récapitulatif des caractéristiques des copules archimédiennes pour une valeur de paramètre θ donnée, de leurs générateurs, formes analytiques (en dimension 2), leurs coefficients de dépendance de queue inférieur et supérieur, leurs taux de Kendall et le type de dépendance modélisée.

	Gumbel	**Frank**	**Clayton**
$\varphi(u)$	$(-\ln u)^{\theta}, \theta \geq 1$	$-\ln(\frac{e^{-\theta u}-1}{e^{-\theta}-1}), \theta \neq 0$	$\frac{u^{-\theta}-1}{\theta}, \theta > 0$
$C(u,v)$	$\exp\left\{-\left[(-\ln u)^{\theta}+(-\ln v)^{\theta}\right]^{\frac{1}{\theta}}\right\}$	$-\frac{1}{\theta}\ln\left[1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right]$	$(u^{-\theta}+v^{-\theta}-1)^{-\frac{1}{\theta}}$
λ_L	0	0	$2^{-\frac{1}{\theta}}$
λ_U	$2-2^{\frac{1}{\theta}}$	0	0
τ_θ	$1-\frac{1}{\theta}$	$1-\frac{4\times(1-D_1(\theta))}{\theta}$ *	$\frac{\theta}{\theta+2}$
Type de dépendance	positive et accentuée sur la queue supérieure	positive ou négative	positive et accentuée sur la queue inférieure

FIGURE 8.5 – Caractéristiques des copules archimédiennes

3.2 Les Copules Elliptiques

La différence avec ces copules est qu'elles ne se construisent pas à partir d'un générateur mais plutôt à partir de lois dites elliptiques (loi Normale et loi de Student), leur particularité est qu'elles représentent des dépendances symétriques c'est à dire que les coefficients de queue inférieure et supérieure sont sensiblement égaux, et rendent moins bien compte de la dépendance aux extrêmes.

Définition

Une Copule est dite elliptique si elle est la Copule d'une loi elliptique.

Ainsi, une loi est dite elliptique de paramètre de position μ et de matrice de forme symétrique définie positive $\sum$ si sa densité f peut s'écrire :

$$f(x) = |\sum|^{\frac{-1}{2}}.g((x-\mu)^T \sum^{-1}(x-\mu))$$

q étant une fonction scalaire à valeur positive telle que

$$\int_{R^n} q(\|y\|^2)\,\mathrm{d}y = 1$$

Copule Gaussienne

soit $\rho \in]-1,1[$ le coefficient de corrélation, Φ la fonction de répartition de la loi N(0,1) et Φ_ρ la distribution bi-variée de matrice de variance covariance $\sum$ (matrice de corrélation dans notre cas) associé à ρ, la copule gaussienne C_ρ est définie par :

3. *Le tau de Kendall s'exprime ici en fonction de θ et par la fonction de Deby définie par : $D_{k(x)=\frac{k}{x^k}\int_0^x \frac{t^k}{e^t-1}\,\mathrm{d}t}$

$$C_\rho(u,v) = \Phi_\rho(\phi^{-1}(u), \phi^{-1}(v))$$

Avec :

$$\Phi_\rho(u,v) = \int_{-\infty}^{u}\int_{-\infty}^{v} \frac{1}{2\pi\sqrt{(1-\rho^2)}} exp(-\frac{s^2+t^2-2\rho st}{2(1-\rho^2)})\,\mathrm{d}s\mathrm{d}t$$

La copule gaussienne est donc paramétrée par le coefficient de correlation ρ.

Remarques :

— $\rho = 0 \to C_0 = \Pi$

— $\rho = -1 \to C_{-1} = C^-$

— $\rho = 1 \to C_1 = C^+$

— la copule Gaussienne ne possède pas de dépendance de queue : $\lambda_L = \lambda_U = 0$

Copules de Student

La copule de Student est une copule paramétrée par le coefficient de correlation ρ et de le dégré de liberté ν.Cette copule est donnée par :

$$C_{\rho,\nu}(u,v) = \int_{-\infty}^{t_\nu^{-1}(u)}\int_{-\infty}^{t_\nu^{-1}(v)} \frac{\Gamma(\frac{\nu+2}{2})}{2\pi\Gamma(\frac{\nu}{2})\sqrt{1-\rho^2}}(1+\frac{s^2+t^2-2\rho st}{\nu(1-\rho^2)})^{-\frac{\nu+2}{2}}\,\mathrm{d}s\mathrm{d}t$$

Avec t_ν la fonction de répartition de la loi de student à ν dégrés de libertés.

Remarques

si $\nu \to +\infty$ alors la copule de student converge vers la copule gaussienne.
Aussi, contrairement à la copule Gaussienne, elle possède une dépendance de queue, et comme elle est symétrique, les coefficients de dépendance de queue inférieur et supérieur sont égaux :

$$\lambda_L = \lambda_U = 2\bar{t_{\nu+1}}(\sqrt{\nu+1}\sqrt{\frac{1-\rho}{1+\rho}})$$

Avec $\bar{t_{\nu+1}}$ la fonction de survie d'une variable de Student à $\nu+1$ degré de liberté.

Nous illustrons donc la dépendance induite par ces deux copules par des graphiques de leurs densités et dépendogrammes respectifs.

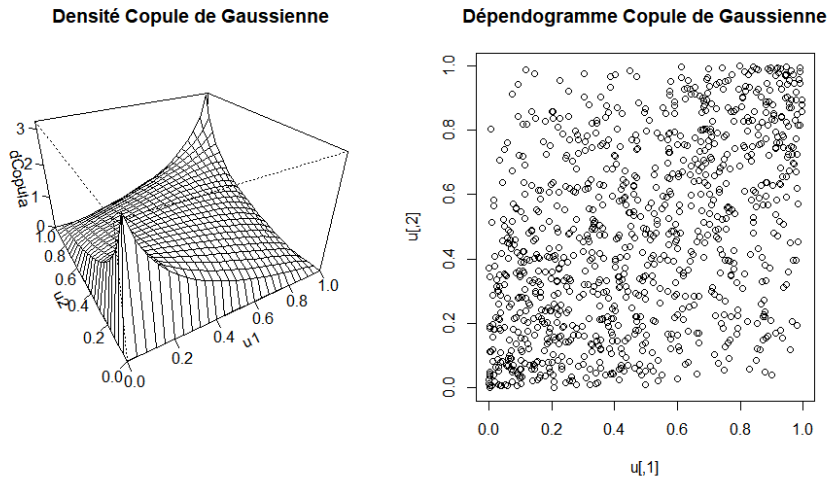


FIGURE 8.6 – Densité et dépendogramme pour une copule de Gaussienne de paramètre $\rho = 0.5$

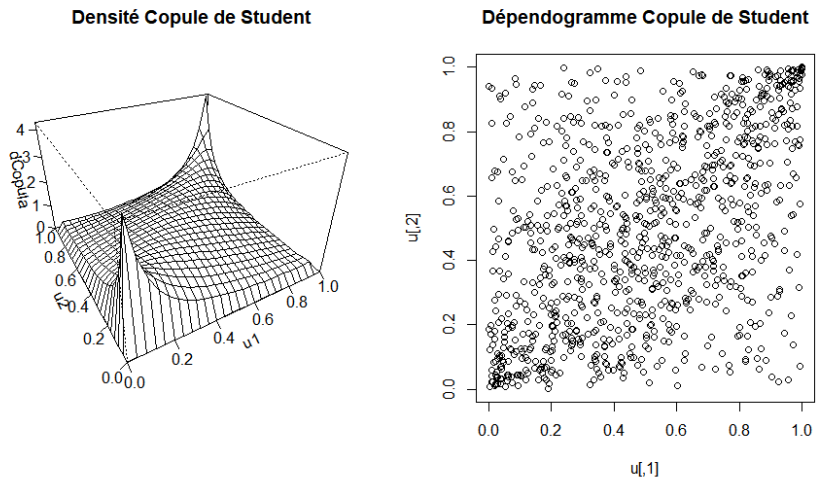


FIGURE 8.7 – Densité et dépendogramme pour une copule de Gaussienne de paramètre $\rho = 0.5$, $\nu = 4$

On peut donc remarquer une plus grande concentration des points (en haut à droite et en bas à gauche) pour la copule de Student par rapport à la Gaussienne, donc la copule de Student modélise mieux la dépendance des extrêmes que la Gaussienne.

Chapitre 9

Simulation des Copules

Dans ce chapitre, nous allons présenter les différentes méthodes utilisées pour générer les montants de provisions agrégées suivant les différentes copules choisies.
Nous présenterons donc ici trois méthodes : l'algorithme de Marshall-Olkin et la méthode des distributions conditionnelles qui permettent de simuler des copules archimédiennes et la méthode des distributions qui permet de simuler des copules elliptiques. Nous allons appliquer sur nos données l'algorithme de Marshall-Olkin pour les copules archimédiennes et la méthode des distributions pour les copules elliptiques.

1 Algorithme de Marshall-Olkin

Le principal outil permettant de simuler les copules archimédiennes est l'algorithme de Marshall-Olkin proposé par Marshall et Olkin en 1988 qui utilise pour cela l'inverse de la transformée de Laplace-Stieltjes du générateur ϕ de la copule en question.
La transformée de Laplace d'une variable positive Y est définie par :

$$\Psi(s) = E[e^{-sY}] = \int_0^\infty e^{-st}\,\mathrm{d}F_Y(t),$$

F_Y étant la fonction de répartition de Y.

Marshall et Olkin montrent que la fonction de répartition jointe F peut s'écrire de la manière suivante :

$$F(x_1, x_2, ..., x_d) = \Psi(\Psi^{-1}(F_1(x_1)) + .. + \Psi^{-1}(F_d(x_d)))$$

F_i étant la fonction de répartition marginale de la fonction de répartition jointe ,$\forall i \in (1, ..d)$.

Pour les copules de Gumbel,Frank et Clayton, les inverses des transformées de Laplace-Stieltjes sont données dans le tableau suivant :

Copule	Générateur $\phi(t)$	$\mathrm{LS}^{-1}(\phi)$
Gumbel	$(-lnt)^\theta, \theta \geq 1$	$S(\frac{1}{\theta}, 1, (cos(\frac{\pi}{2\theta}))^\theta, 0, 1)*$
Frank	$-ln(\frac{e^{-\theta t}-1}{e^{-\theta}-1}), \theta \neq 0$	$(y_k = \frac{(1-e-\theta)^k}{k\theta}, k \in N)*$
Clayton	$\frac{t^{-\theta}-1}{\theta}, \theta > 0$	$\Gamma(\frac{1}{\theta}, 1)$

TABLE 9.1 – Générateurs et inverses de transformés de Laplace -Stieltjes des Copules archimédiennes

Il est important de noter que pour la copule de Frank, l'inverse de la transformée de Laplace-Stieltjes est obtenue par la méthode d'inversion discrète en utilisant la fonction *rlgser* sous le logiciel R.

1

1. * S est la distribution de la loi Stable $S(\alpha, \beta, \lambda, \gamma)$ avec $\alpha \in [0, 2]$, $\beta \in [-1, 1]$, $\lambda \in [0, +\infty]$ et $\gamma \in R$

Les générateurs des copules de Clayton, Frank et Gumbel étant des transformées Laplace de variables aléatoires positives, nous en déduisons l'algorithme suivant :

Soient $X_1; .....; X_n$, n variables aléatoires réelles de fonctions de répartition respectives F_1 et F_n et dont on cherche la distribution jointe. Ayant les données ci-dessus, l'algorithme est le suivant :

1. On simule une variable V suivant une loi $LS^{-1}(\phi)$: dans le cas d'une copule de Gumbel par exemple, on simulera une variable V de loi $S(\frac{1}{\theta}, 1, (cos(\frac{\pi}{2\theta}))^{\theta}, 0, 1)$

2. On simule les variables i.i.d $X_i = U[0,1], i = 1; 2, ...; n$

3. On retourne les valeurs $(U_1\ ;\ ....\ ;\ U_n\ )$, $U_i = \phi(-\frac{ln(X_i)}{V})$, i = 1 ;2..... ;n

4. Finalement, la distribution jointe recherchée $(X_1; .....; X_n)$ est obtenue par :

$$(X_1; .....; X_n) = (F_1^{-1}(U_1); .....; F_n^{-1}(U_n))$$

On obtient donc ainsi la distribution multivariée des variables $X_1; .....; X_n$

2 Méthode des distributions

C'est la méthode utilisée pour générer des copules elliptiques.
Généralement, nous avons :

2.1 Pour la copule normale

Soient X et Y deux variables aléatoires réelles de fonctions de répartition respectives F_X et F_Y et dont on cherche la distribution jointe.
Soit R la matrice de corrélation de Pearson de nos deux variables aléatoires. Ayant les données ci-dessus, l'algorithme est le suivant (en dimension 2 et se généralise en dimension supérieure) :

— Trouver la décomposition L de Cholesky de la matrice de corrélation R

— Simuler un vecteur $z = (z_1, z_2)^T$ de variables aléatoires indépendantes suivant la loi N(0,1)

— Poser $x = (x_1, x_2) = Lz$

— On obtient : $(U_1, U_2) = (\Phi(x_1), \Phi(x_2))$ avec Φ la fonction de répartition de la loi normale centrée réduite.

— Finalement : $(X, Y) = (F_X^{-1}(U_1), F_Y^{-1}(U_1)$

2.2 Pour la copule de Student

Soient X et Y deux variables aléatoires réelles de fonctions de répartition respectives F_X et F_Y et dont on cherche la distribution jointe.
Soit R la matrice de corrélation de Pearson de nos deux variables aléatoires. Ayant les données ci-dessus, l'algorithme est le suivant (en dimension 2 et se généralise en dimension supérieure) :

— Trouver la décomposition L de Cholesky de la matrice de corrélation R

— Simuler un vecteur $z = (z_1, z_2)^T$ de variables aléatoires indépendantes suivant la loi N(0,1)

— Simuler une variable aléatoire s indépendante de z suivant une loi χ^2 avec ν degrés de liberté.

— poser $x = (x_1, x_2) = (\nu/s)^{1/2} Lz$

— On obtient : $(U_1, U_2) = (T_\nu(x_1), T_\nu(x_2))$ avec T_ν la fonction de répartition de la loi de Student avec ν degrés de liberté.

— Finalement : $(X, Y) = (F_X^{-1}(U_1), F_Y^{-1}(U_1)$

3 Méthode des distributions conditionnelles

Cette méthode utilise les distributions conditionnelles. Si on veut simuler d variables aléatoires $U_1, ..., U_d$ de distribution jointe C, alors la distribution conditionnelle U_i sachant $U_1, ..., U_{i-1}$ est donnée par :

$$C(u_i/u_1, ..., u_{i-1}) = P(U_i \leq u_i/U_1 = u_1, ..., U_{i-1} = u_{i-1}) =$$

$$\frac{\partial u_1 ... \partial u_{i-1} C_i(u_1,...,u_i)}{\partial u_1 ... \partial u_{i-1} C_{i-1}(u_1,...,u_i)}$$

Ainsi, pour nos deux variables aléatoires réelles (comme dans les cas précédents), l'algorithme est le suivant :

— Simuler une variable aléatoire u_1 suivant une loi U(0,1)

— Simuler une variable aléatoire u_2 suivant une loi $C_2(./u_1)$ (En dimension d on fera des simulations jusque u_d suivant une loi $C_2(./u_1, ..., u_d)$)

— Finalement, nous avons que $(U_1; U_2; ...; U_d) = (F_{U_1}^{-1}(u_1); .....; F_{U_d}^{-1}(u_d))$

Chapitre 10

Application de la théorie des copules

Le but de ce chapitre est de pouvoir calculer la provision agrégée sur nos 03 lobs en prenant en compte toute la structure de dépendance qui existe entre ces lobs. Cette partie utilisera donc les résultats obtenus dans la méthode de régression GLM de (famille Gamma) utilisée précédemment. Nous allons donc comme expliqué précédemment, utiliser comme données de départ les simulations obtenues par la méthode du GLM-Bootstrap : ces valeurs nous permettrons d'obtenir la meilleure loi suivie par les provisions de chaque lob. Concernant la dépendance entre les lobs, nous allons la modéliser grâce aux résidus de Pearson obtenus sur le GLM.

Pour pouvoir modéliser l'espérance de la provision agrégée sur nos 03 lobs, il est nécessaire de connaître sa distribution empirique et de la simuler par des copules. Pour y parvenir, nous aurons donc besoin des différentes marginales qui seront donc ici les lois suivies par les différentes lobs. Tout l'intérêt de la modélisation de la dépendance par les copules réside dans le fait que nous pouvons librement ajuster des lois sur les différentes lobs, il n'est donc pas nécessaire que les lois suivent une distribution normale par exemple ou bien qu'elles soient de même nature.

1 Ajustement des distributions marginales sur chacune des lobs

1.1 Densités empiriques

Nous pouvons définir la densité empirique comme étant le graphique montrant la proportion des observations prenant chacune des valeurs des données de départ.
Dans le but d'avoir une visualisation graphique des distributions de nos 03 lobs, nous avons décidé de tracer les fonctions de densité empirique de chacune d'elles à l'aide de la fonction *density* du logiciel R. Les graphiques sont donnés en annexe I du mémoire.
Nous notons que les distributions des 03 lobs sont toutes pointues et à queue épaisse à droite.

1.2 Recherche de lois

Pour modéliser les provisions et coûts des sinistres en assurance non-vie, on utilise généralement des lois telles que : la loi normale, la loi log-normale, la loi Gamma, la loi exponentielle et la loi de Pareto. Une illustration du graphique des densités de chacune des lois est donnée en annexe J du mémoire. De cela, nous voyons qu'il n'est pas pertinent sur nos données de tester les lois de Pareto et Exponentielle au regard de la forme pointue de nos densités empiriques (forme qu'on ne retrouve pas dans le cadre des deux lois précédentes)1 . On va donc essayer de réaliser des tests d'ajustement sur les 03 lois restantes.

La première chose à faire avant d'ajuster une loi sur nos données empiriques est de vérifier le caractère indépendant et identiquement distribué de nos simulations de provisions pour chaque lob (soit H_0 l'hypothèse selon laquelle les provisions sont indépendantes et identiquement distribuées). Pour ce faire, nous aurons donc recours au test du signe des différences sous

R (présenté en annexe E du mémoire). Les résultats de ce test sont stockés dans le tableau suivant :

	Statistique du test	P-value du test	Décision du test
lob_4	-1.182	0.237	Non rejet de H_0
lob_5	0.730	0.465	Non rejet de H_0
lob_7	0.801	0.465	Non rejet de H_0

TABLE 10.1 – Test du caractère iid de nos données

De ces chiffres nous déduisons qu'au seuil de 5% nous ne rejetons pas l'hypothèse H_0 de départ.
Nous allons par la suite en supposant que nos provisions pour chaque lob suivent les 03 lois que nous allons tester, estimer leurs paramètres pour chacune de ces lois. Nous faisons cela par la méthode du maximum de vraisemblance décrite en annexe A du mémoire. Nous faisons ensuite un test d'adéquation de la loi paramétrique aux données par le test de Kolmogorov-Smirnov (Valérie LAGENEBRE, *Prise en compte des dépendances entre risques par la théorie des Copules : Impact sur le besoin en fonds propres de l'assureur*, 2009), dont le principe est de tester l'écart entre la distribution paramétrique et la distribution empirique. Les résultats sont donnés dans les tableaux suivants (au seuil de 5%) :

lob_4			
Loi	Paramètres	P-value du test	Déviance
Loi Normale	$\mu = 26247551, \sigma = 16078177$	$< 3e-16$	0.1386
Loi log-normale	$\mu = 16.923, \sigma = 0.564$	$< 3e-16$	0.1927
Loi Gamma	$s = 2.685, r = 1.015e-07$	$< 3e-16$	0.1695
lob_7			
Loi	Paramètres	P-value du test	Déviance
Loi Normale	$\mu = 49027271, \sigma = 26599216$	$< 3e-16$	0.1782
Loi log-normale	$\mu = 17.578, \sigma = 0.507$	$< 3e-16$	0.1719
Loi Gamma	$s = 3.39, r = 6.92e-08$	$< 3e-16$	0.1583
lob_5			
Loi	Paramètres	P-value du test	Déviance
Loi Normale	$\mu = 4592503, \sigma = 5405032$	$< 3e-16$	0.2419
Loi log-normale	$\mu = 14.90, \sigma = 0.93$	$< 3e-16$	0.2262
Loi Gamma	$s = 7.219e-01, r = 1.572e-07$	$< 3e-16$	0.1465

TABLE 10.2 – Paramètres et statistiques du test d'ajustement des lois aux données

Nous constatons que les p-values sont très basses.De ce fait, il n'est pas judicieux de choisir les lois les plus adaptées en ce basant sur ce critère. Nous allons donc choisir les lois les plus adaptées en nous basant sur le critère de la déviance et on considèrera que la loi s'ajuste le mieux si elle a la déviance la plus basse. Ainsi,pour les lob 7 et 5 la meilleure loi au sens de la déviance est la loi Gamma, pour la lob 4 c'est la loi normale. C'est donc ici que se révèle l'intérêt de l'utilisation des copules car, on peut utiliser des marginales différentes pour la modélisation de la provision agrégée. Une confirmation par illustration graphique de cet ajustement est donnée en annexe K pour chaque lob associée aux différentes lois.

2 Structure de dépendance entre les 03 lob

Cette partie vise à analyser la structure de dépendance existante entre les 03 lobs sur lesquelles porte notre étude. Nous ferons une analyse des dépendances par couple de lob car, bien que ne nous permettant pas de modéliser la corrélation entre les 03 lob, cela nous donnera

une information sur les lobs présentant des corrélations entre elles et d'en étudier l'influence sur le montant et la variabilité des provisions. De plus, la matrice de corrélation de Pearson par exemple nous permettra d'ajuster les copules de Student et Gaussiennes.

2.1 Calcul des coefficients de corrélation entre les 03 lobs

Nous avons eu comme première idée de modéliser la corrélation deux à deux entre les lobs en utilisant le triangle incrémental de règlement (ce qui est plutôt intuitif dans notre cas , la variable aléatoire étudiée étant la provision). Nous avons remarqué par la suite que cette méthode peut présenter un gros biais sur l'estimation des coefficients de corrélation en ce sens que les coefficients estimés seront un reflet de la similarité dans la liquidation (les triangles ayant la même vitesse de liquidation auront généralement des coefficients de corrélation élevés). Ainsi, nous nous appuierons donc sur les résidus de Pearson (obtenus dans notre GLM-Gamma) comme variable d'entrée pour la modélisation de la corrélation entre les 03 lobs. Ci-dessous une preuve que nous pouvons effectivement utiliser les résidus de Pearson comme variable pour la modélisation de la corrélation :

Nous voulons déterminer la corrélation entre deux variables aléatoires Y_1 et Y_2 qui sont expliquées par un GLM avec comme variables explicatives X représentant les années de survenance et de développement. Soient g_1 et g_2 les fonctions liens respectives dans le cadre de ce GLM, nous avons :

$$m_1 = E(Y_1) = g_1(X_1\beta_1)$$

$$m_2 = E(Y_2) = g_1(X_2\beta_2)$$

avec β_1 et β_2 les paramètres du GLM pour Y_1 et Y_2 respectivement.

Nous avons deux estimateurs du maximum de vraisemblance de m_1 et m_2 notés :

$$\hat{m_1} = g_1(X_1\hat{\beta_1}) \text{ et } \hat{m_2} = g_2(X_2\hat{\beta_2})$$

Les résidus sont donnés par :

$$r_1 = Y_1 - \hat{m_1} \text{ et } r_2 = Y_2 - \hat{m_2}$$

Regardons la corrélation entre r_1 et r_2 :

$$E((r_1 - E(r_1))(r_2 - E(r_2)))$$

$$= E((Y_1 - \hat{m_1} - E(Y_1) - E(\hat{m_1}))(Y_2 - \hat{m_2} - E(Y_2) - E(\hat{m_2})))$$

$$= E((Y_1 - E(Y_1) - (\hat{m_1} - E(\hat{m_1})(Y_2 - E(Y_2) - (\hat{m_2} - E(\hat{m_2}))))$$

$$= E((Y_1 - E(Y_1))(Y_2 - E(Y_2))) - E((\hat{m_1})) \ (1)$$

$$\text{- } E((\hat{m_1}))(Y_2 - m_2 - (\hat{m_2} - E(\hat{m_2})))) \ (2)$$

$$\text{- } E((Y_1 - m_1 - (\hat{m_1} - E(\hat{m_1})(\hat{m_2} - E(\hat{m_2})) \ (3)$$

On voit que l'expression (1) est la corrélation entre Y_1 et Y_2. Les expressions (2) et (3) quant à eux tendent vers 0 (par l'inégalité de Minkowski) quand la taille de l'échantillon du GLM

tend vers l'infini.
Au final nous avons que :

$$E((r_1 - E(r_1)(r_2 - E(r_2)) \longrightarrow E((Y_1 - E(Y_1))(Y_2 - E(Y_2)))$$

Donc étudier la corrélation entre les résidus (résidus de Pearson à un facteur multiplicateur près : l'inverse des variances des résidus) revient à étudier la corrélation entre les incréments qui revient ensuite à étudier la corrélation entre les provisions.

Ainsi nous avons ci-dessous les différents coefficients de corrélation, ainsi que la significativité de ceux-ci en utilisant la fonction *cor.test* du logiciel R

La matrice de corrélation de Pearson :

Corrélation de Pearson			
	lob_4	lob_7	lob_5
lob_4	r = 1	r = -0.0349 , p-value = 0.704	r = 0.0669 , p-value = 0.467
lob_7		r = 1	r = 0.0693 , p-value = 0.451
lob_5			r = 1

TABLE 10.3 – Marice de Corrélation de Pearson entre les lobs

La matrice des coefficients de Kendall :

Corrélation de Kendall			
	lob_4	lob_7	lob_5
lob_4	$\tau = 1$	$\tau = -0.0168$, p-value = 0.785	$\tau = 0.119$, p-value = 0.053
lob_7		$\tau = 1$	$\tau = 0.0635$, p-value = 0.303
lob_5			$\tau = 1$

TABLE 10.4 – Marice de Corrélation de Kendall entre les lobs

La matrice des coefficients de Spearman :

Corrélation de Spearman			
	lob_4	lob_7	lob_5
lob_4	$\rho = 1$	$\rho = -0.003$, p-value = 0.968	$\rho = 0.15$, p-value = 0.085
lob_7		$\rho = 1$	$\rho = 0.081$, p-value = 0.376
lob_5			$\rho = 1$

TABLE 10.5 – Marice de Corrélation de Spearman entre les lobs

Nous voyons qu'à un seuil de 5% dans tous les cas de figures, les coefficients de corrélation sont significatifs. On peut en conclure qu'il est pertinent d'évaluer un montant de provision agrégé sur nos 03 lobs. Une illustration graphique de l'aperçu des résidus de Pearson sur nos lobs est donnée en annexe M du mémoire.

2.2 Dépendogrammes deux à deux

Ce type de graphique nous permet d'avoir le type de dépendance que la Copule devrait modéliser.

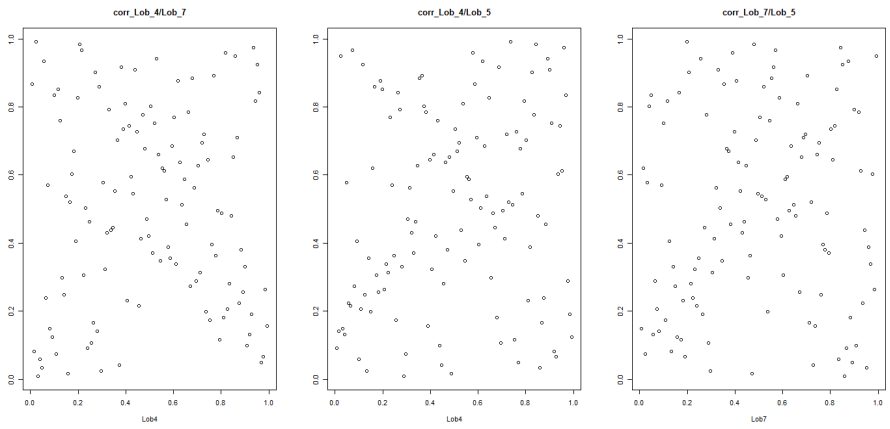


FIGURE 10.1 – Dépendogrammes deux à deux entre les différentes lobs

A partir du dépendogramme ci-dessus, on s'attend à ce que la meilleure Copule choisie modélise :

— une dépendance positive entre les couples de lobs (7,5)

— une dépendance négative entre les couples de lob (4,7)

— Une dépendance de queue en haut à droite et en bas à gauche sur les couples de lob (4,7) ,4,5) et (7,5). Cela signifie qu'il y'a une réalisation simultanée de fortes provisions sur toutes les lobs et une réalisation simultanée de faibles provisions sur toutes les lobs.

2.3 Estimation des paramètres des Copules

Le tableau des paramètres des Copules estimés par la méthode MPL (Maximum pseudo-likelihood estimator), nous notons df comme étant le degré de liberté de la Copule de Student.

Copule	Paramètre estimé (méthode MPL)	Maximum de Log-Vraissemblance
Gumbel	$\theta = 1.036$	0.4148
Clayton	$\theta = 0.1846$	3.563
Frank	$\theta = 0.4546$	3.563
Gaussienne	$\rho = 0.07666$	0.872
Student	$\nu = 0.0957$, df $= 2.7726$	10.37

TABLE 10.6 – Paramètres des copules associées à nos lobs

Plus la log-vraisemblance est élevée, meilleure est l'adaptation des copules à nos données. Selon ce critère donc, la meilleure copule sera donc la copule de Student.
On représente ci-dessous le dépendogramme induit par la Copule de Student que nous considérons comme étant la meilleure Copule. La Copule a été simulée 10000 fois :

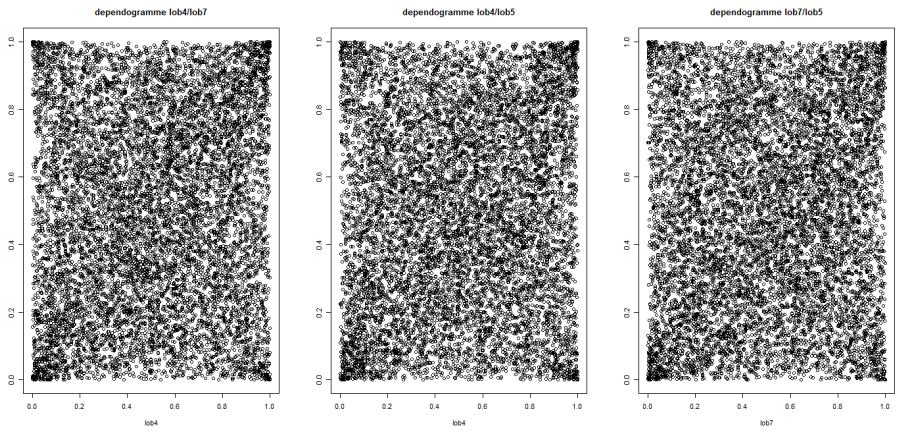


FIGURE 10.2 – Dépendogrammes de la Copule la mieux ajustée à nos données

On voit donc que le dépendogramme induit par la Copule sélectionnée respecte la dépendance observée sur nos données.
Dans la suite du mémoire, nous allons tout de même comparer les distributions agrégées induites par toutes les Copules avec celle où le calcul de la provision est obtenu en sommant les provisions individuelles sur chaque lob.

Copule	Var 99.5%	Var 95%	Var 90%	Espérance
Clayton(Normale)	183 705 356	153 463 256	139 739 724	79 955 441
Clayton(Log-normale)	237 583 573	164 030 045	139 261 853	79 877 325
Clayton(Gamma)	218 810 248	164 610 138	143 042 447	79 906 259
Clayton(*)	210 244 243	160 912 267	141 573 960	79 928 587
Gumbel(Normale)	169 430 332	134 365 870	121 911 649	79 939 022
Gumbel(Log-normale)	212 548 115	141 209 226	121 546 369	79 840 112
Gumbel(Gamma)	196 302 934	140 655 965	123 316 911	79 890 280
Gumbel(*)	200 969 203	140 915 798	122 895 966	79 856 685
Frank(Normale)	165 080 546	135 170 574	123 048 978	79 799 826
Frank(Log-normale)	204 889 608	141 601 271	122 365 762	79 717 378
Frank(Gamma)	191 155 821	141 241 456	124 353 164	79 767 453
Frank(*)	185 077 735	139 770 534	124 221 091	79 721 487
Gaussienne(Normale)	161 649 729	132 608 893	121 472 822	79 819 214
Gaussienne(Log-normale)	197 136 729	139 630 233	120 927 742	79 842 987
Gaussienne(Gamma)	185 155 370	138 243 344	122 540 598	79 844 355
Gaussienne(*)	182 354 083	136 830 227	121 782 655	79 837 110
Student(Normale)	165 475 317	132 102 490	120 222 946	79 269 579
Student(Log-normale)	212 445 938	139 785 517	121 344 219	79 730 885
Student(Gamma)	192 422 749	138 554 713	122 639 938	79 614 000
Student(*)	192 681 123	136 626 314	121 327 615	79 520 150
Cas individuel	187 692 392	146 104 696	123 399 425	79 757 099

TABLE 10.7 – Synthèse globale des provisions agrégées par différentes Copules

Les valeurs *Copule(Loi)* correspondent aux cas où on considère une même loi sur toutes les 03 lobs Les valeurs *Copule(*)* correspondent aux cas où on considère les meilleures lois (les lois s'ajustant le mieux) pour chaque lob : dans notre cas, la loi Normale pour la lob4 et la loi Gamma pour la lob 7 et 5. La valeur *Cas individuel* correspond au cas où la provision globale est obtenue en sommant les provisions individuelles sur chaque lob.
Ainsi, nous avons donc simulé les différentes valeurs de *value-at-risk* et l'espérance attendue de la provision agrégée pour les différentes copules et les marginales considérées.
Il est à noter que pour le «Cas individuel» , nous avons juste utilisé les simulations *bootstrapées* de notre *glm-bootsrap* précédent pour chaque lob, nous avons fait la somme de ces simulations et avons ensuite déduit les différentes *value-at-risk* et l'espérance associée en faisant des sommes de valeurs de chaque lob.

On remarque de prime abord que dans tous les cas de figures la provision espérée est quasiment identique pour toutes les copules mais aussi pour le «Cas individuel». C'est un

résultat auquel on s'attend normalement : cela découle de la loi faible des grands nombres qui assure que la moyenne empirique est un estimateur convergent de l'espérance et de la propriété de linéarité de l'espérance. Nous avonsdonc une espérance des provisions agrégées égale à environ 79.8 Millions d'euros.
Nous remarquons également que la différence entre le «Cas individuel» et la «meilleure copule» : Student(*) s'observe significativement sur la Var 95% et la Var 99.5% mais ce n'est pas le cas sur la Var 90%. Celà montre que le fait pour l'entreprise de considérer la dépendance entre les 03 lob en question ne serait pertinent que si elle considère sa marge de risque associée à une Var 95% ou 99.5%.
Nous analysons de plus près la différence entre le «Cas individuel» et la meilleure Copule (associée aux meilleures lois), en donnant ci-dessous les densités de la provision agrégée pour chaque cas :

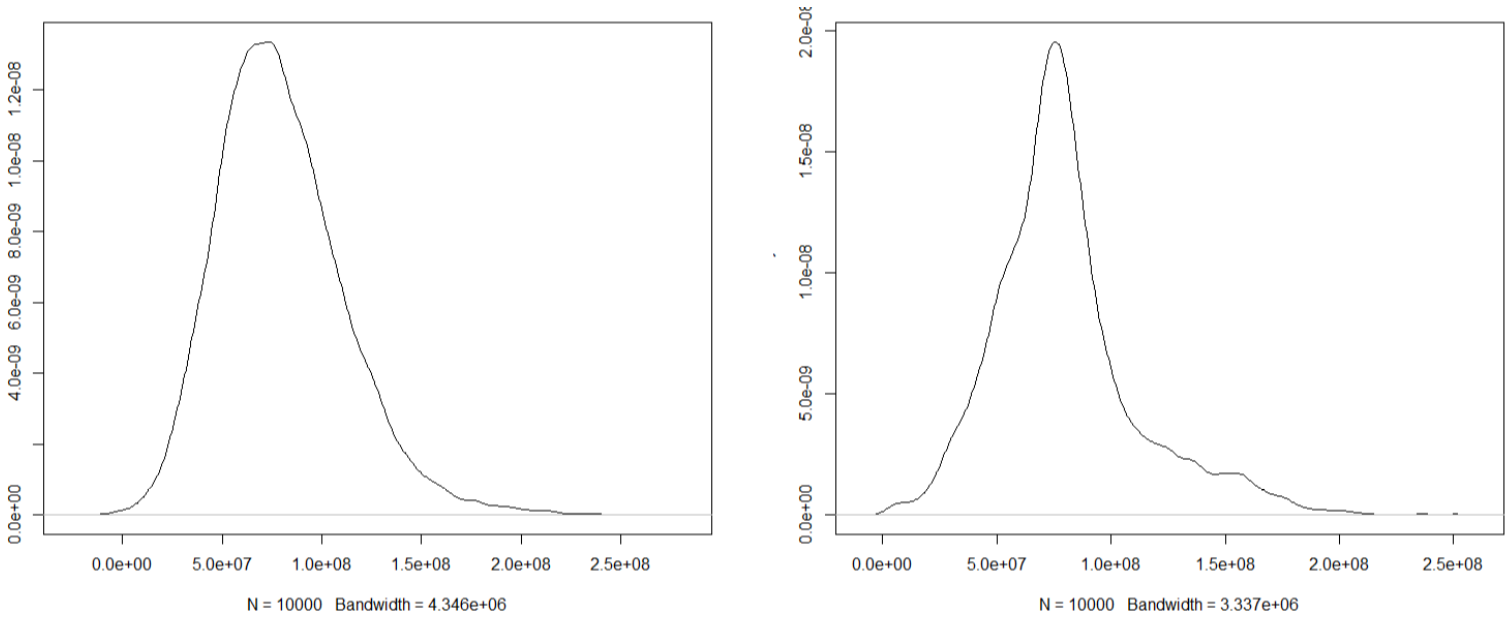


FIGURE 10.3 – Densité de la provision agrégée : meilleure Copule + meilleures lois et «Cas individuel»

On remarque donc que la densité pour le «Cas individuel» est assez pointue comparée à celle de la meilleure Copule associée aux meilleures lois et que les principales différences se trouvent au niveau des queues supérieures. Ce qui nous amène donc à donner un détail plus fin des *value-at-risk* dans les deux cas pour pouvoir étudier la pertinence de la prise en compte de la dépendance, on peut également obtenir le pourcentage de variation entre la valeur de la *value-at-risk* pour le cas de la meilleure Copule et la valeur pour le «Cas individuel». On obtient donc le tableau ci-dessous :

Copule	Var 50%	Var 75%	Var 90%	Var 95%	Var 97.5%	Var 99.5%
Cas individuel	75 857 741	91 619 048	123 399 425	146 104 696	160 380 428	187 692 392
Student(*)	76 154 822	98 150 768	121 327 615	136 626 314	152 947 952	192 681 123
% variation	0.39%	7.12%	−1.67%	−6.48%	−4.3%	2.65%

TABLE 10.8 – Etude comparative entre la meilleure Copule et le Cas individuel

Cela confirme bien ce qui a été vu sur le graphique des densités précédent, mais ne nous permet pas de faire une généralisation car dépendant du quantile choisi, le «Cas individuel» peut donner une valeur plus ou moins élevée que la meilleure Copule. Il faut noter que l'on observe une différence conséquente entre le «Cas individuel» et la meilleure Copule qu'à partir d'un certain niveau de quantile élevé.

CONCLUSION

Notre travail a porté sur la prise en compte de la corrélation entre les lobs dans l'optique de calculer une provision agrégée sur l'ensemble des lobs étudiées.

Notre approche a consisté dans un premier temps à avoir une approche individuelle du calcul des provisions nécessaires sur chaque lob : nous avons donc appliqué en premier lieu les méthodes de Chain-Ladder déterministe et stochastique, qui nous ont permis d'avoir une espérance sur l'ensemble des lobs et l'erreur associée à celle-ci.Nous nous sommes ensuite tournés vers les méthodes stochastiques factorielles qui sont plus flexibles et mieux adaptées à nos données. De cela, nous avons donc pu utiliser la théorie des Copules afin de pouvoir étudier l'impact de la prise en compte de la dépendance sur les provisions agrégées de nos lobs ; que nous avons ensuite comparé au cas où les provisions sont calculées individuellement sur chaque lobs.

Dans un premier temps, il a fallu simuler 10 000 valeurs ***bootstrapées*** de nos provisions pour chaque lob à partir d'un Glm avec pour famille la distribution Gamma. Nous avons ensuite ajusté pour chacune des 10 000 simulations obtenues des lois de probabilités et avons réalisé des test d'ajustement pour déterminer la plus adaptée selon le paramètre de déviance et ce dépendant de la lob, par la même occasion nous avons estimé des paramètres en fonction des différentes lois testées. Nous avons constaté que, la loi Normale est la plus adaptée pour la lob 4 et la loi Gamma pour les lobs 4 et 5.

La structure de dépendance étant nécessaire à a simulation des Copules, nous avons donc réalisé une étude de cette dernière sur nos 03 lobs dans le but de pouvoir tester l'ajustement des Copules à cette structure par le critère du maximum de log-vraisemblance et aussi estimer le paramètre de la Copule testée sous-jacent aux lobs étudiées. Nous avons aussi démontré que nous pouvons étudier la structure de dépendance des provisions en passant par les résidus de Pearson obtenus dans le Glm. Il en est ressorti que la Copule de Student décrit mieux la structure de dépendance entre nos lobs. Nous avons tout de même décidé dans un but comparatif de simuler toutes les copules et toutes les lois testées sur nos données, au regard de la flexibilité d'utilisation de marginales que nous donne les Copules.

La simulation des Copules nous a donc fait remarquer que peu importe la Copule et peu importe la loi utilisée, l'espérance agrégée sur nos lobs est sensiblement constante ce qui est un résultat intuitif et auquel on devrait s'attendre. Aussi, nous remarquons que, la Copule de Clayton de façon générale aura tendance à donner les valeurs les plus élevées de *value-at-risk*. Mais l'observation la plus intéressante est la comparaison des valeurs obtenues entre le «Cas individuel» et la meilleure Copule associée aux meilleures lois ; on voit que le bénéfice de diversification est logiquement plus important sur des quantiles élevés ; ainsi il devient important de ne pas négliger la dépendance pour estimer ces derniers.

Cardif IARD exigeant dans son modèle de provisionnement le choix d'une *value-at-risk* sur les provisions (c'est à dire l'intégration de la marge de risque), ce choix aura donc tendance à faire diminuer le niveau de provision à constituer par rapport au cas où les *value-at-risk* sont estimées individuellement sur chacune des lobs. Pour le niveau de quantile retenu par CARDIF IARD, nous constatons que la prise en compte de la corrélation entre les lobs nous permet d'observer un bénéfice de diversification au niveau de la provision à constituer, mais son impact est tout de même limitée car de l'odre de 1.67 % . Avec une *value-at-risk* de l'ordre

de 95% cet impact est en revanche plus conséquent (de l'ordre de 6.48%).

Enfin, en essayant d'opposer la méthode de Mack à la méthode des copules (associé au glm), nous ne pouvons pas dire à priori qu'une méthode est préférable à l'autre dans tous les cas. La méthode de Mack est une méthode moins volatile que la méthode glm associée aux copules, mais l'inconvénient ici est qu'elle ne nous permettra pas d'étudier aisément la structure complète de dépendance entre les lobs comme avec le modèle glm associé aux copules. Nous allons donc privilégier le modèle de Mack si nous remarquons que l'étude de la corrélation entre les lobs ne nous donne pas de bénéfice de diversification conséquent (ce qui n'est pas le cas dans la plupart des quantiles extrêmes).

Une autre problématique reste encore en suspens pour Cardif IARD pour améliorer son modèle de provisionnement. Il s'agirait pour l'entreprise, en dehors du fait de pouvoir estimer une provision corrélée sur l'ensemble de ses branches, de pouvoir modéliser et prendre en compte le critère d'inflation et son évolution dans le temps et de l'associer à la modélisation de sa provision agrégée.

Annexes

Annexe A

Explication générale du maximum de vraisemblance

Soit X une variable aléatoire réelle, de loi D_θ , de paramètre θ inconnu.Soit f sa fonction d densité ou de probabilité selon que la variable soit continue ou discrète.

On appelle vraisemblance de θ au vu des observations $(x_1, \ldots, x_n)$ d'un n-échantillon indépendant et identiquement distribué selon la loi $\mathcal{D}_\theta$, le nombre :

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

A $(x_1, \ldots, x_n)$ fixé, on cherche à trouver le maximum de cette vraisemblance pour que les probabilités des réalisations observées soient aussi maximum. Ceci est un problème optimisation. Ainsi, un estimateur du maximum de vraisemblance est tout estimateur $\widehat{\theta}$ de vérifiant :

$$\sup_{\theta} L(x_1, \ldots, x_n; \theta) = L(x_1, \ldots, x_n; \hat{\theta}).$$

Annexe B

Récapitulatif des résultats du Chain-Ladder sur l'ensemble des 03 lob

Années de Survenance	Ri lob 4	Ri lob 5	Ri lob 7	Total
01/2007	-	-	-	-
01/2008	593	11	1731	2384
01/20 9	12 749	26	3 799	2 334
01/2010	16 739	38	23 7 68	40 545
01/2011	112 209	533	176 333	289 076
01/2012	161 220	898	249 038	411 155
01/2013	330 275	1 069	411 741	743 085
01/2014	523 390	1 414	613 396	1 138 201
01/2015	674 465	2 267	865 038	1 541 769
01/2016	1 306 856	24 457	1 219 107	2 550 420
01/2017	2 004 698	56 957	1 492 968	3 554 420
01/2018	2 304 795	138 151	3 051 894	5 494 840
01/2019	3 347 918	211 959	5 077 528	8 637 405
01/2020	4 712 575	292 177	8 866 674	13 871 426
01/2021	12 837 435	5 731 828	34 926 746	53 496 009
Total	28 345 916	6 461 787	56 979 761	91 787 464

Annexe C

Analyse rétrospective de Chain-Ladder sur l'ensemble des 03 lob

Analyse rétrospective sur la lob 4 :

	écart absolu	écart relatif
01/2007	0	0%
01/2008	10 565	2641%
01/2009	295	55%
01/2010	- 254 188	-69%
01/2011	- 10 438	-21%
01/2012	17 515	19%
01/2013	18 116	13%
01/2014	- 229 171	-46%
01/2015	- 204 235	-21%
01/2016	73 224	11%
01/2017	150 968	17%
01/2018	411 019	30%
01/2019	411 019	30%
01/2020	1 540 496	44%
Total	1 465 479	16%

Analyse rétrospective sur la lob 5 :

	écart absolu	écart relatif
01/2007	0	0%
01/2008	0	0%
01/2009	2	16%
01/2010	538	5375%
01/2011	143	71%
01/2012	51	512%
01/2013	111	1111%
01/2014	-4113	-78%
01/2015	23 656	7450%
01/2016	39 909	399087%
01/2017	-145 589	-70%
01/2018	- 6 752	−16%
01/2019	98 918	989180%
01/2020	1 786 580	75%
Total	1 793 453	68%

Analyse rétrospective sur la lob 7 :

	écart absolu	écart relatif
01/2007	0	0%
01/2008	1 168	11680%
01/2009	16 293	273%
01/2010	72 641	96%
01/2011	- 58 405	-53%
01/2012	75 430	79%
01/2013	- 64 382	-33%
01/2014	- 26 287	-7%
01/2015	129 925	93%
01/2016	- 475 288	-50%
01/2017	-459 894	-35%
01/2018	- 1 331 004	-39%
01/2019	- 843 005	-18%
01/2020	3 385 081	20%
Total	422 271	1%

Annexe D

Test de Shapiro-Wilk

Il teste l'hypothèse nulle selon laquelle un échantillon $x_1, \ldots, x_n$ suit une loi normale. La statistique du test est :

$$W = \frac{(Zx_{(i)})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Avec :

- $x_{(i)}$ le i-ème plus petit nombre de l'échantillon
- $\bar{x} = \frac{1}{n}(x_1 + ... + x_n)$ la moyenne de l'échantillon
- $(a_1, \ldots, a_n) = \frac{m^T V^{-1}}{\sqrt{m^T V^{-1} V^{-1} m}}$ où $m = (m_1, \ldots, m_n)^T$ et $(m_1, \ldots, m_n)^T$ sont des statistiques d'ordre d'un échantillon de variables iid suivant une loi normale, et V est la matrice de variance-covariance de ces statistiques d'ordre.

En partant du principe que l'hypothèse nulle (H_0) est que la population étudiée suit une loi normale :
On rejette (H_0) au niveau α si $W > q_{\frac{\alpha}{2}}$.Sinon on ne rejette pas(mais cela ne voudrait pas forcément dire que l'hypothèse de normalité est forcément vérifiée mais on pourra supposer que c'est le cas)

Annexe E

Test du signe des différences

Il permet de tester l'indépendance d'un échantillon de données à travers le signe donné par la différence des observations consécutives de la distribution.
Soit $Y_1, \ldots, Y_n$ une suite de variable aléatoire réelle de fonction de répartition continue, telles que :

$$P(Y_i = Y_j) = 0 \text{ , pour } i \neq j$$

Se basant sur les réalisations $y_1, \ldots, y_n$, on teste pour un seuil $\alpha > 0$ l'hypothèse selon laquelle la suite de variable aléatoire précédente soit indépendante et identiquement distribuée.

$$\text{Soit } Z_i = 1_{Y_{i+1} > Y_i} \text{ , pour i = 1,...,n-1. Soit } S_n = \sum_{i=1}^{n-1} Z_i$$

La statistique du test est $T_n = \frac{S_n - E[S_n]}{\sqrt{Var(S_n)}}$
On calcule donc $T_n(\omega)$ pour les valeurs $y_1, \ldots, y_n$ observées et on rejette (H_0) au niveau α si $T_n(\omega) > q_{\frac{\alpha}{2}}$. Avec $q_{\frac{\alpha}{2}}$ le quantile d'ordre $\frac{\alpha}{2}$ de la distribution de S_n .Sinon on ne rejette pas(mais cela ne voudrait pas forcément dire que l'hypothèse de d'indépendance est forcément vérifiée, mais on pourra supposer que c'est le cas).

Annexe F

Analyse rétrospective des méthodes factorielles

Surv	Dev	Ecart absolu	Ecart relatif	Surv	Dec	Ecart absolu	Ecart relatif	Surv	Dev	Ecart absolu	Ecart relatif
N14	1	-132773	-4%	N12	11	4988	15%	N8	5	203870	36%
N14	2	-129520	-4%	N12	12	108108	5124%	N8	6	253960	53%
N14	3	440948	51%	N12	13	285	54%	N8	7	134802	40%
N14	4	155478	25%	N11	1	306491	1%	N8	8	88270	50%
N14	5	291343	142%	N11	2	-245057	-5%	N8	9	20741	15%
N14	6	-98029	-17%	N11	3	37993	2%	N7	1	507914	12%
N14	7	54707	22%	N11	4	191005	23%	N7	2	539269	9%
N14	8	-61782	-27%	N11	5	-203 223	-24%	N7	3	462480	28%
N14	9	97486	1919%	N11	6	413806	72%	N7	4	132139	12%
N14	10	73285	1279%	N11	7	-284392	-42%	N7	5	305215	28%
N14	11	18,662	0%	N11	8	-143 605	-39%	N7	6	260580	26%
N14	12	32782	327820%	N11	9	-351170	-72%	N7	7	50769	9%
N14	13	304	91%	N11	10	45031	7%	N7	8	226594	45%
N14	14	-10462	-51%	N11	11	41196	5336%	N6	1	212295	5%
N14	15	0	0%	N11	12	-245269	-67%	N6	2	26201	1%
N13	1	-367 685	-11%	N10	1	341916	11%	N6	3	352428	23%
N13	2	202118	6%	N10	2	-312036	-7%	N6	4	304950	38%
N13	3	284385	26%	N10	3	-142523	-8%	N6	5	157749	29%
N13	4	-56387	-6%	N10	4	-59962	-6%	N6	6	124614	16%
N13	5	168377	48%	N10	5	249960	68%	N6	7	46754	10%
N13	6	-231625	-32%	N10	6	199383	52%	N5	1	186314	4%
N13	7	186037	144%	N10	7	308593	468%	N5	2	151814	3%
N13	8	39105	10%	N10	8	23051	12%	N5	3	252739	11%
N13	9	-40600	-27%	N10	9	122159	2335%	N5	4	480493	3%
N13	10	-63570	-42%	N10	10	-172766	-62%	N5	5	465894	38%
N13	11	-61340	-174%	N10	11	11259	2%	N5	6	259360	26%
N13	12	37039	370390%	N9	1	55363	1%	N4	1	256756	6%
N13	13	-613	-46%	N9	2	-234692	-5%	N4	2	281721	6%
N13	14	3997	999%	N9	3	126317	8%	N4	3	1148733	37%
N12	1	256645	8%	N9	4	347293	50%	N4	4	292609	20%
N12	2	-295653	-7%	N9	5	-27218	-4%	N4	5	42106	6%
N12	3	176083	13%	N9	6	527731	519%	N3	1	155817	4%
N12	4	-33249	-9%	N9	7	183030	31%	N3	2	79545	2%
N12	5	107153	22%	N9	8	87433	62%	N3	3	223012	11%
N12	6	-57054	-9%	N9	9	109904	404%	N3	4	145123	16%
N12	7	-63761	-15%	N9	10	22129	24%	N2	1	944047	20%
N12	8	202120	2021200%	N8	1	84322	2%	N2	2	654608	17%
N12	9	64062	111%	N8	2	-13361	0%	N2	3	373464	27%
N12	10	95112	1621%	N8	3	-225437	-10%	N1	1	775058	18%
				N8	4	-234326	-16%	N1	2	654339	19%
								N	1	0	0%

FIGURE F.1 – Analyse rétrospective de la régression lognormale sur la lob4

Surv	Dev	Ecart absolu	Ecart relatif	Surv	Dev	Ecart absolu	Ecart relatif	Surv	Dev	Ecart absolu	Ecart relatif
N14	1	-636411	-21%	N12	11	5741	17%	N8	5	96014	17%
N14	2	-933722	-26%	N12	12	90423	4285%	N8	6	166922	35%
N14	3	498149	58%	N12	13	218	41%	N8	7	60	0%
N14	4	269433	43%	N11	1	646880	19%	N8	8	37476	21%
N14	5	383908	187%	N11	2	148,965	3%	N8	9	-14195	-10%
N14	6	5064	1%	N11	3	118282	7%	N7	1	848832	20%
N14	7	121113	49%	N11	4	227302	27%	N7	2	88151	1%
N14	8	-25546	-11%	N11	5	-198696	-23%	N7	3	297 587	18%
N14	9	117325	2310%	N11	6	61311	11%	N7	4	-15766	-1%
N14	10	93419	1631%	N11	7	-287541	-42%	N7	5	-431033	-39%
N14	11	25250	191%	N11	8	-153298	-42%	N7	6	-364687	-36%
N14	12	91653	916530%	N11	9	-361031	-74%	N7	7	-137723	-26%
N14	13	366	95%	N11	10	34069	51%	N7	8	-285360	-57%
N14	14	-10001	-49%	N11	11	37969	4918%	N6	1	-266737	-6%
N14	15	0	0%	N11	12	-274213	-75%	N6	2	-114198	-2%
N13	1	-6273874	-185%	N10	1	111414	3%	N6	3	272347	18%
N13	2	-299095	-9%	N10	2	-725004	-16%	N6	4	248741	31%
N13	3	396361	37%	N10	3	-131115	-7%	N6	5	104385	19%
N13	4	69504	8%	N10	4	-25945	-3%	N6	6	-159949	-20%
N13	5	258451	73%	N10	5	267698	73%	N6	7	-84590	-18%
N13	6	-130902	-18%	N10	6	231510	60%	N5	1	738146	17%
N13	7	246346	190%	N10	7	319224	484%	N5	2	1021719	20%
N13	8	119897	135%	N10	8	23406	12%	N5	3	-315577	-14%
N13	9	-24668	-17%	N10	9	119019	2275%	N5	4	-566557	-34%
N13	10	-52824	-35%	N10	10	-178000	-64%	N5	5	-551945	-45%
N13	11	-56751	-60%	N10	11	-12209	-24%	N5	6	-325787	-33%
N13	12	92137	921370%	N9	1	-204374	-5%	N4	1	669507	17%
N13	13	-583	-44%	N9	2	-649205	-15%	N4	2	901064	19%
N13	14	10001	2500%	N9	3	69743	4%	N4	3	-1191134	-39%
N12	1	-17466	-1%	N9	4	328176	47%	N4	4	-354671	-24%
N12	2	-782403	-18%	N9	5	-48338	-7%	N4	5	-24766	-4%
N12	3	209442	15%	N9	6	523011	514%	N3	1	175716	5%
N12	4	24415	3%	N9	7	-197965	-34%	N3	2	-126133	-3%
N12	5	144321	30%	N9	8	72563	52%	N3	3	-207416	-11%
N12	6	-6494	-1%	N9	9	97353	358%	N3	4	157833	18%
N12	7	39555	9%	N9	10	9037	10%	N2	1	-950837	-20%
N12	8	210862	2108620%	N8	1	185888	4%	N2	2	582986	15%
N12	9	66196	115%	N8	2	214192	4%	N2	3	367852	27%
N12	10	94299	73%	N8	3	-384522	-17%	N1	1	-672682	-16%
				N8	4	-362466	-25%	N1	2	672682	19%
								N	1	0	0%

FIGURE F.2 – Analyse rétrospective de la régression gamma sur la lob4

Annexe G

Illustration graphique de la normalité des résidus de Pearson des méthodes factorielles

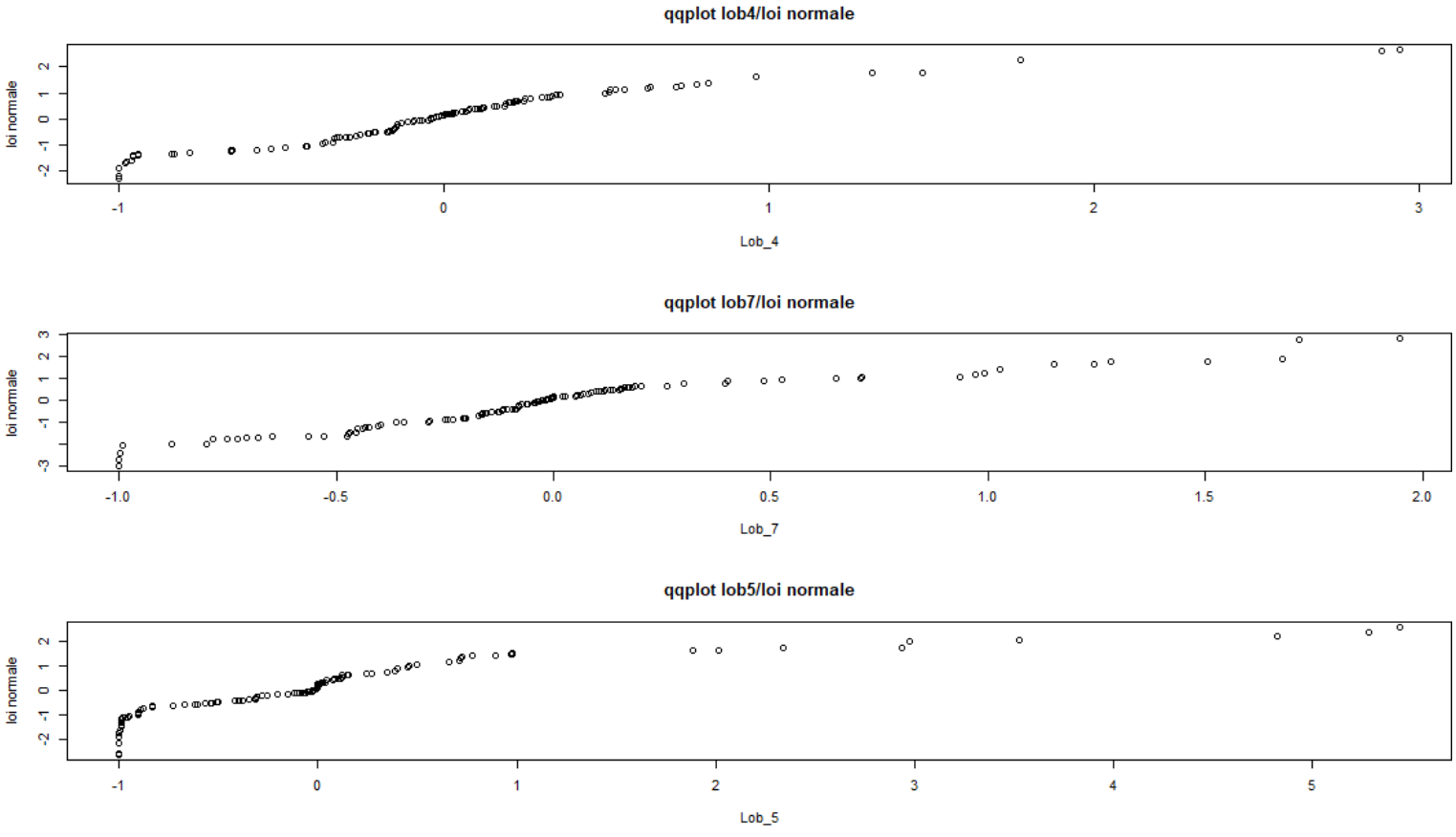


FIGURE G.1 – QQ-plot des résidus de Pearson : GLM de loi gamma

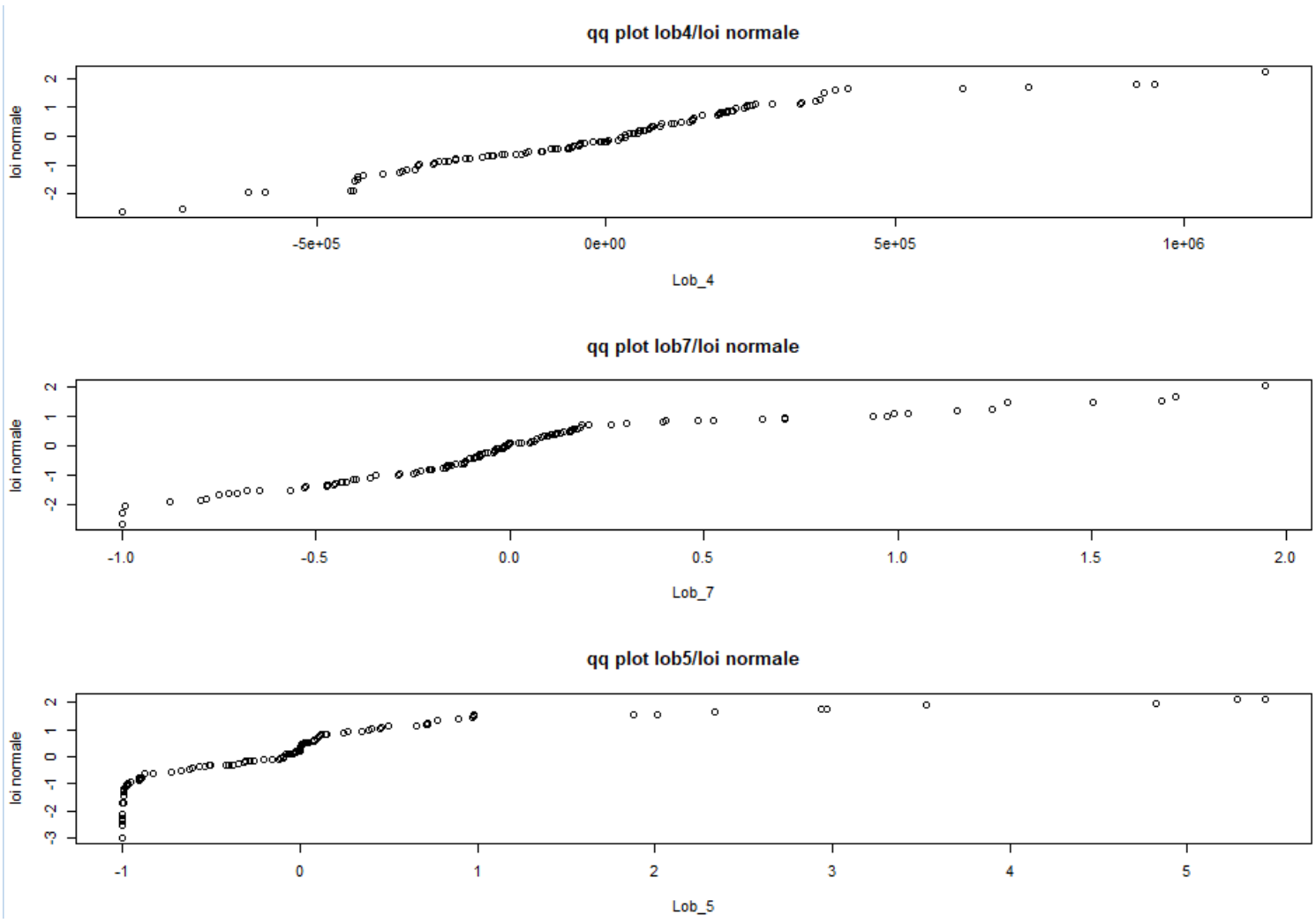


FIGURE G.2 – QQ-plot des résidus de Pearson : GLM de loi log normale

Annexe H

Illustration de la fonction carré

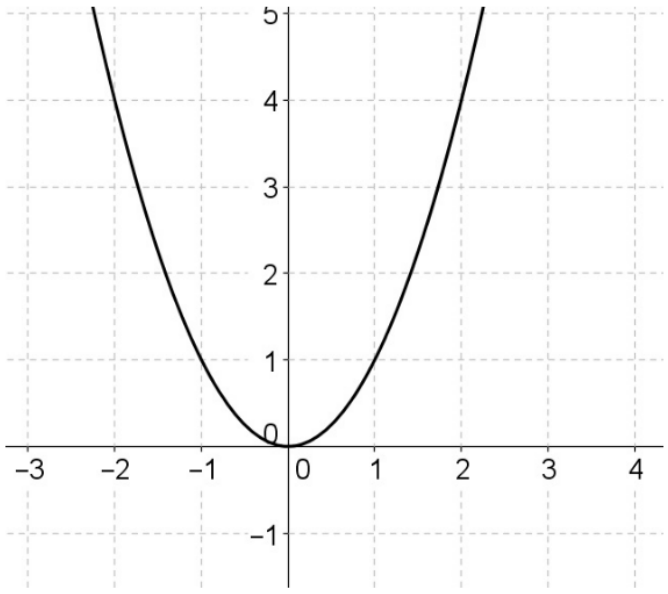


FIGURE H.1 – Illustration de la dépendance entre X et X^2 à travers la fonction carré

Annexe I

Illustration de la densité empirique sur chacune des lob

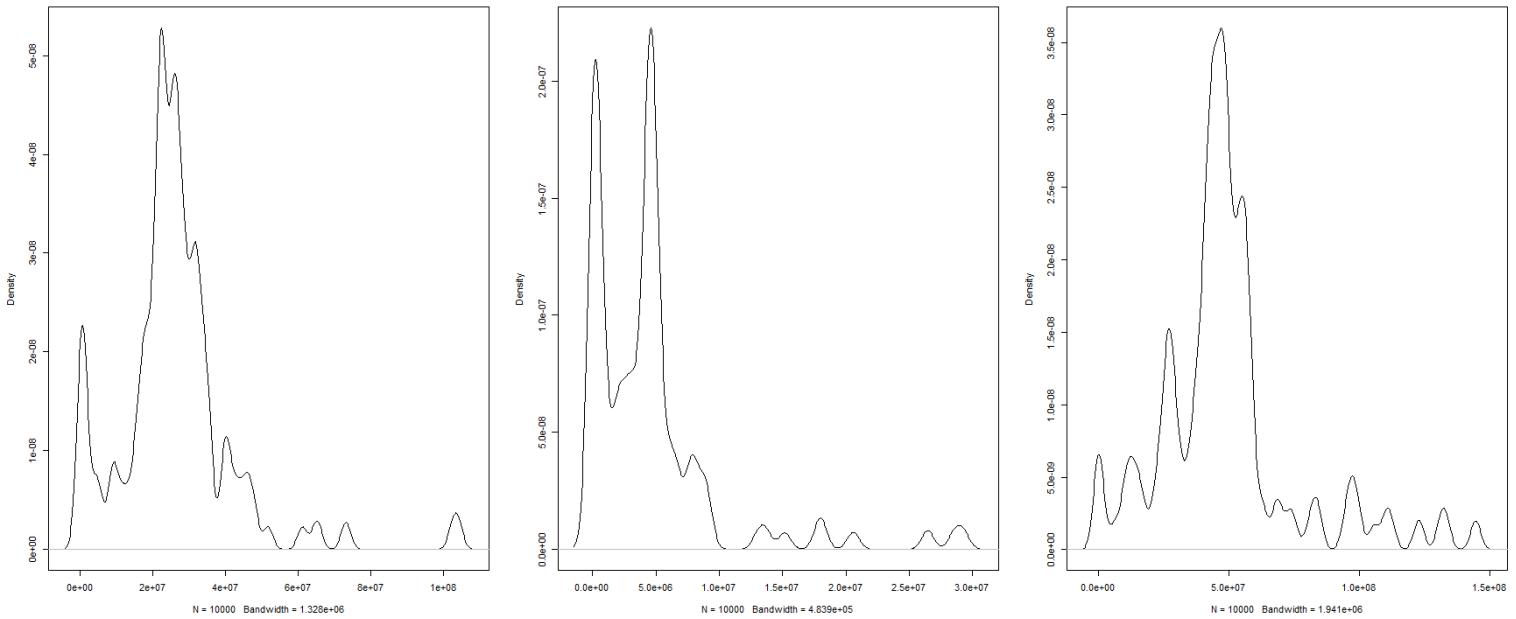
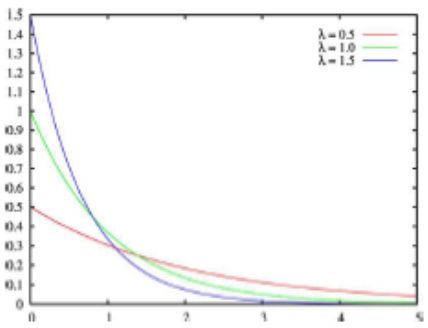


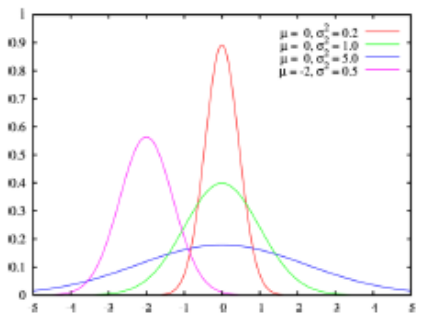
FIGURE I.1 – Illustration de la densité empirique sur chacune des lob

Annexe J

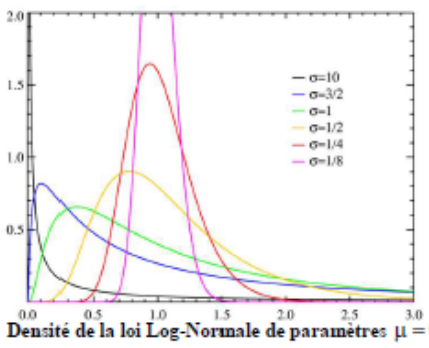
Densité des lois utilisées en assurance non-vie



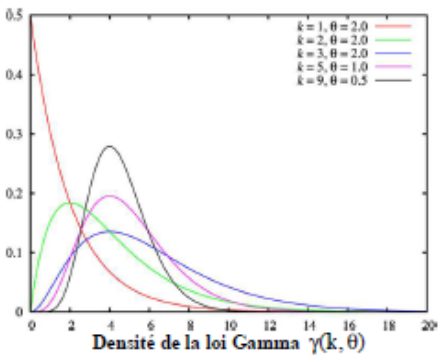
Densité de la loi Exponentielle $\varepsilon(\lambda)$



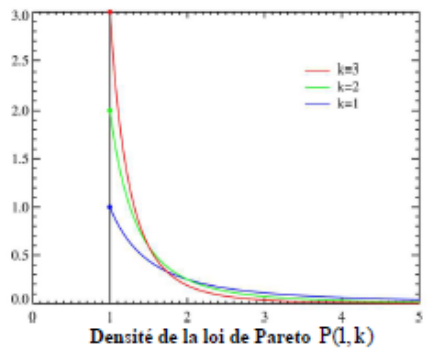
Densité de la loi Normale $N(\mu, \sigma^2)$



Densité de la loi Log-Normale de paramètres $\mu = 0$



Densité de la loi Gamma $\gamma(k, \theta)$



Densité de la loi de Pareto $P(1, k)$

FIGURE J.1 – Illustration de la densité des lois utilisées en assurance non-vie

Annexe K

Illiusttration graphique de l'ajustement de la distribution empirique des provisions aux lois paramétriques par des qq-plot

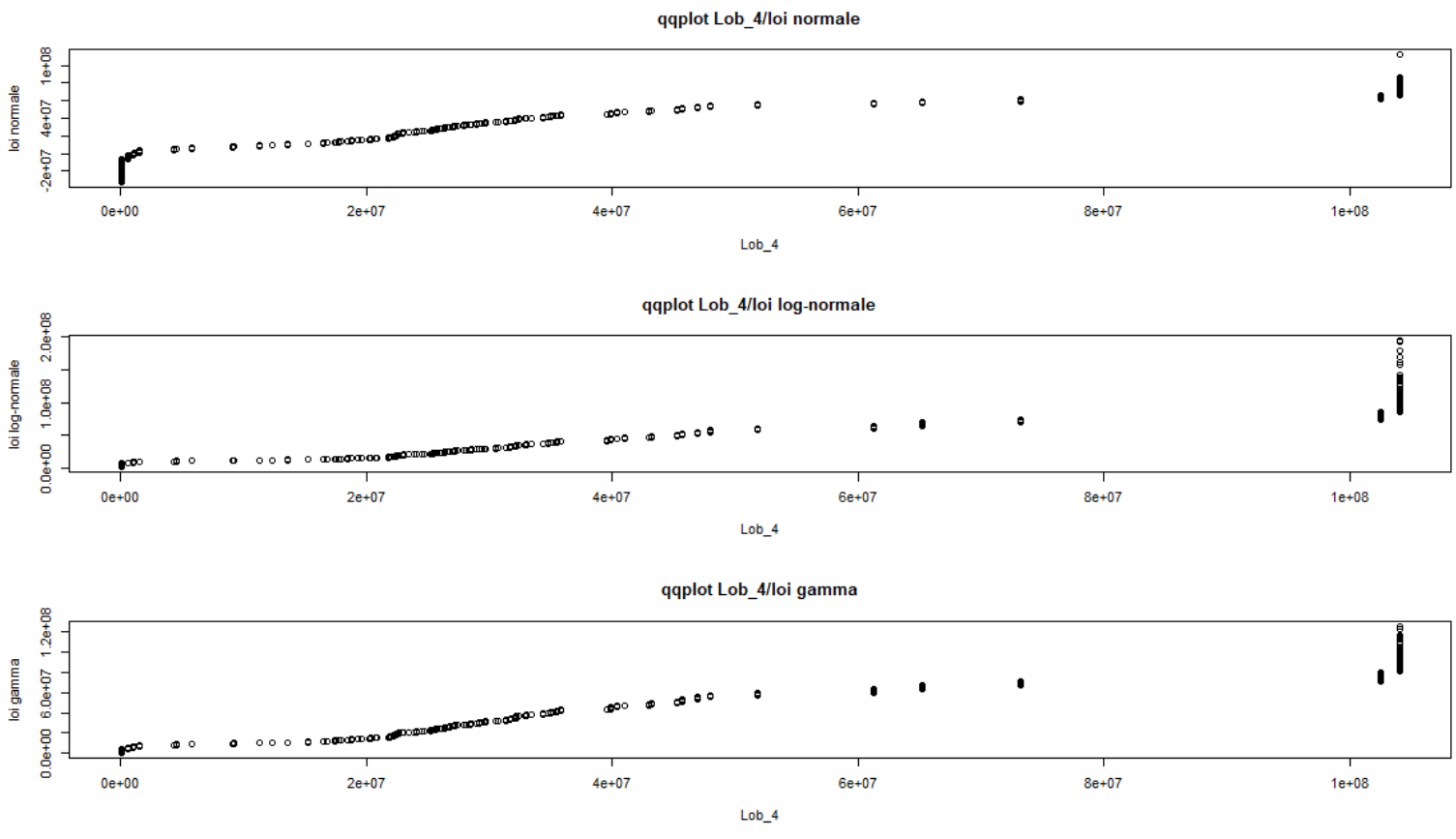


FIGURE K.1 – QQ plot d'ajustement de la lob 4 aux lois paramétriques

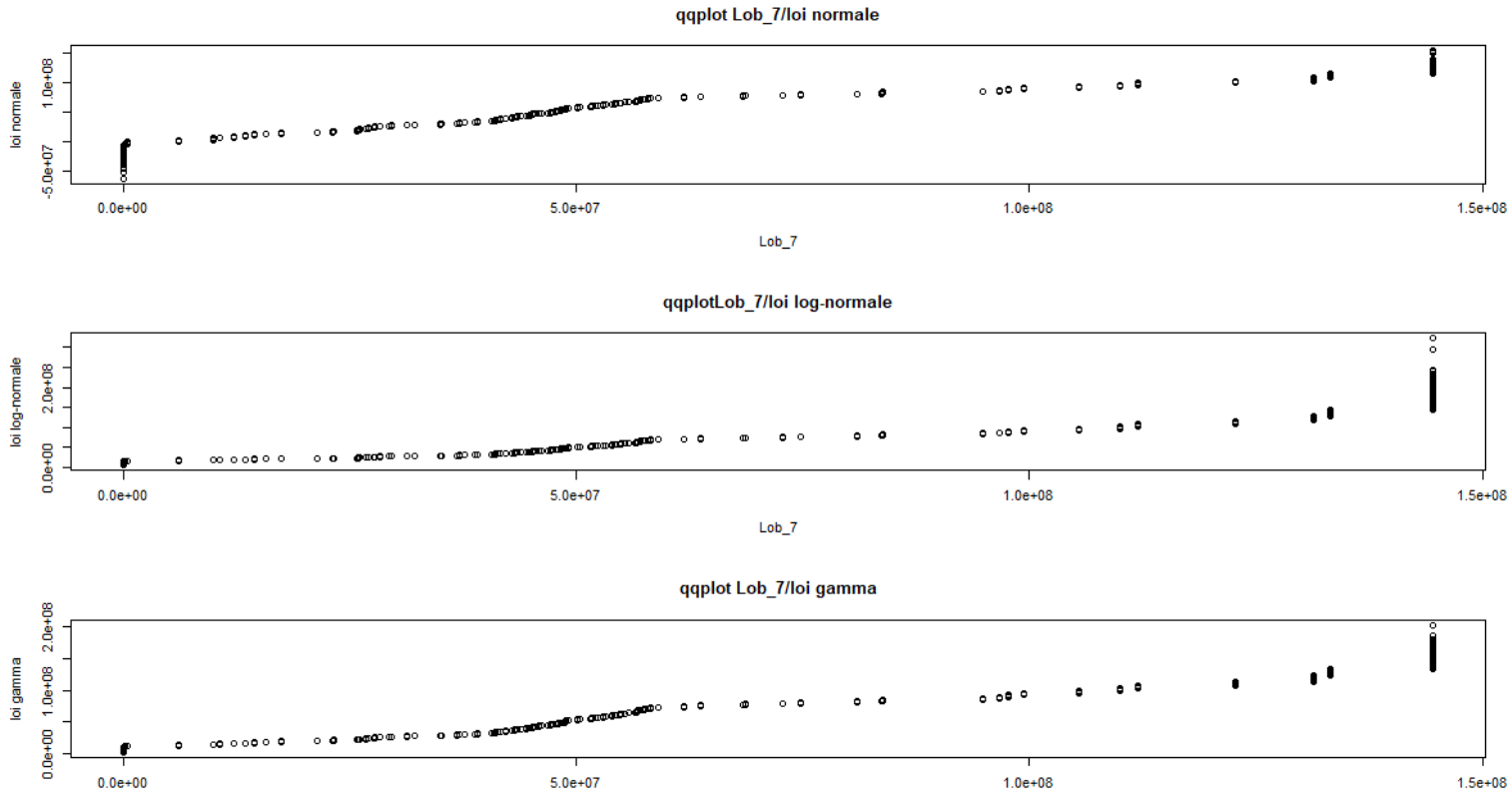


FIGURE K.2 – QQ plot d'ajustement de la lob 7 aux lois paramétriques

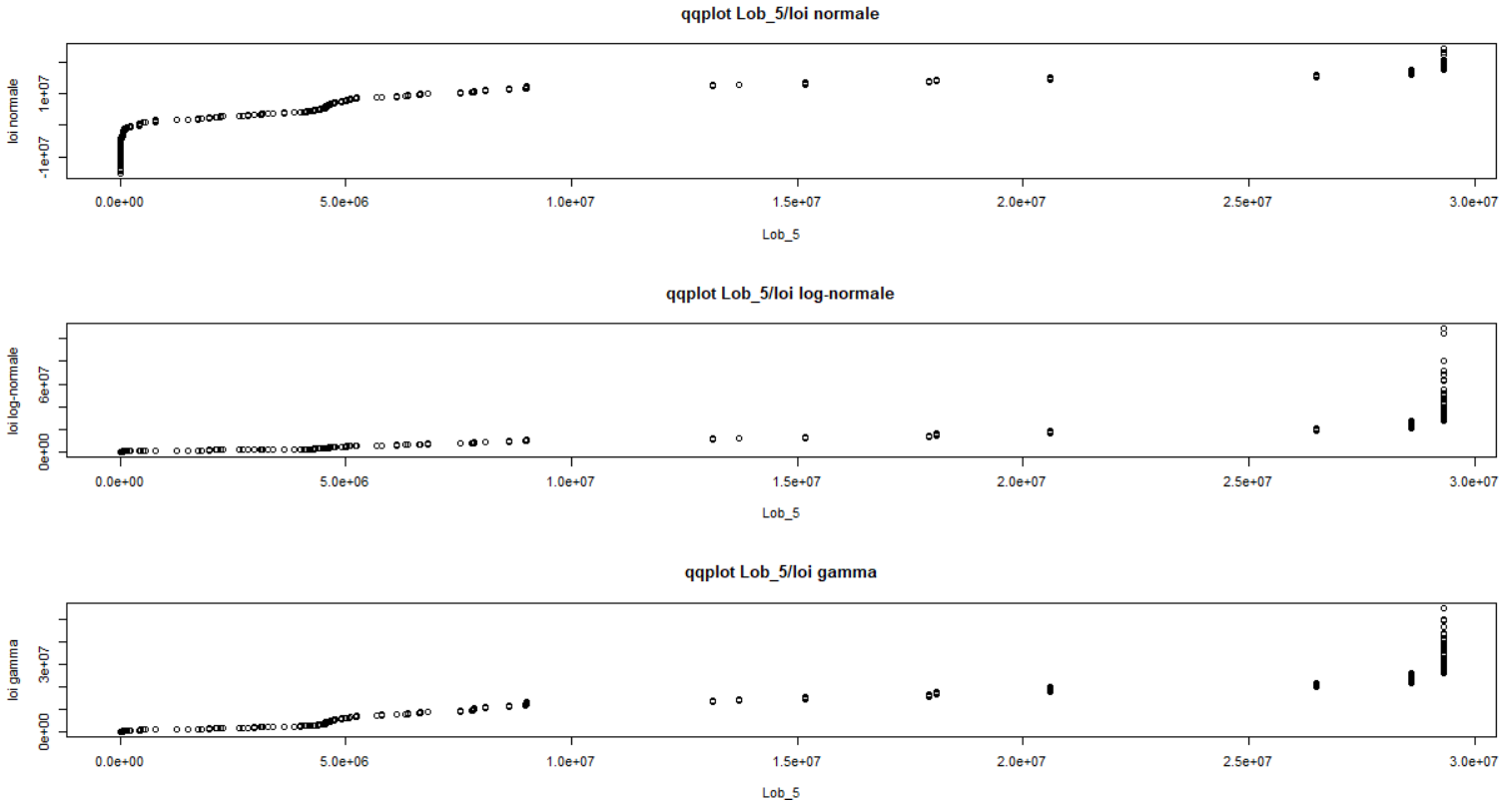


FIGURE K.3 – QQ plot d'ajustement de la lob 5 aux lois paramétriques

Annexe L

Représentation graphique des résidus de Pearson

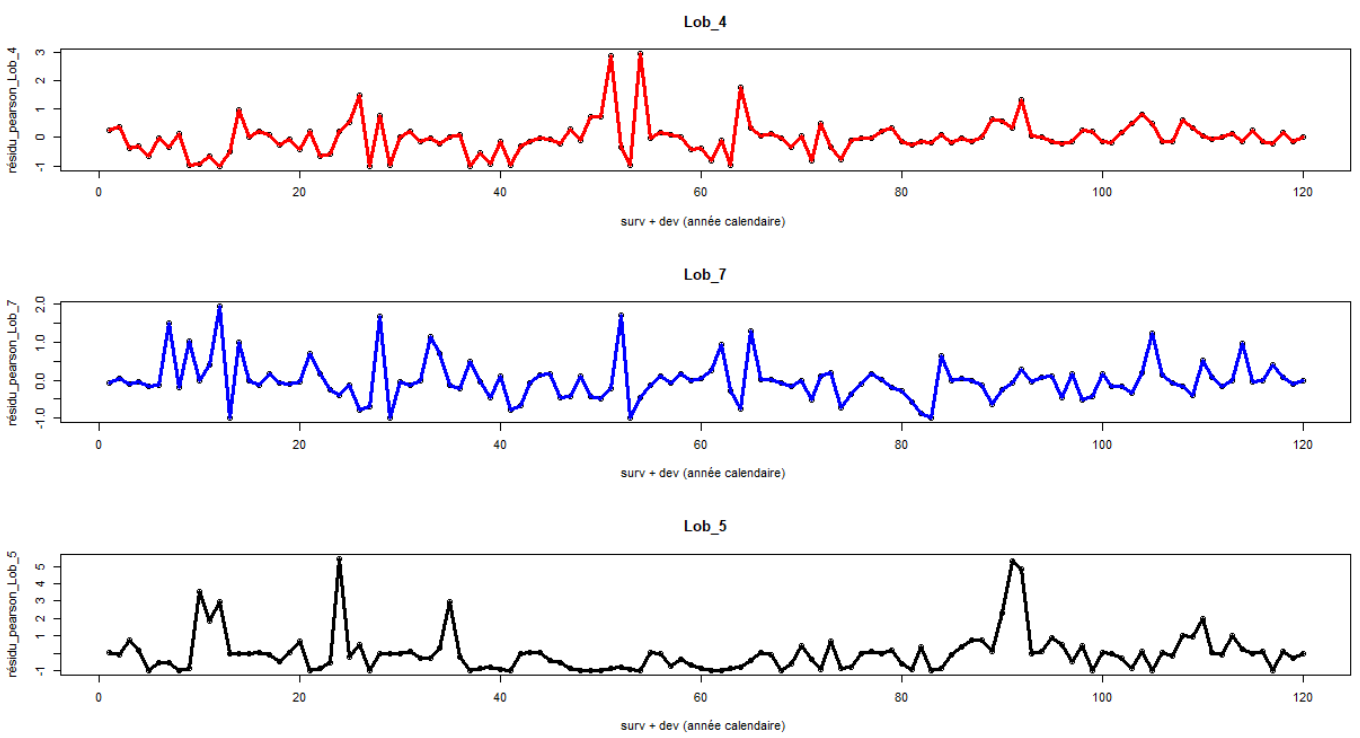


FIGURE L.1 – Représentation graphique des résidus de Pearson

Bibliographie

[1] G. KRAUTH, *Provisionnement et corrélation entre branches*, Mémoire d'actuariat CNAM, 2007

[2] C. PARTRAT, *Corrélation entre risques*, Séminaire Caritat , 2005

[3] E. Masiello, *Méthodes stochastiques de calcul des réserves en assurance non-vie*, ISFA LYON, 2020

[4] B. Abakarim, *Evaluation d'options sur plusieurs sous-jacents par des modèles de copules*, HEC MONTREAL , 2005

[5] K.Mandhouj, *Analyse du Risque de Provisionnement en Assurance IARD et Enjeux dans le Cadre du futur Référentiel « Solvabilité 2 »* , Mémoire d'actuariat ISFA LYON , 2008

[6] K.Mandhouj, *Analyse du Risque de Provisionnement en Assurance IARD et Enjeux dans le Cadre du futur Référentiel « Solvabilité 2 »* , Mémoire d'actuariat ISFA LYON , 2008

[7] C.Genest, J.Mackay , *The joy of copulas : Bivariate distributions with uniform marginals*, American Statistical Association, 1993

[8] C.Hess, *Modélisation de la dépendance entre risques*, Cours de Théorie du risque et réassurance, Dauphine, 2009

[9] A.Miseray, *Analyse du coût des sinistres dans le cadre de la modélisation ligne à ligne d'un portefeuille d'assurance automobile* ,Mémoire d'actuariat ISFA LYON, 2015

[10] M.Denuit, A.Charpentier *Mathématiques de l'assurance non-vie* ,Collection Economie et Statistiques Avancées, Economica, 2005

[11] H.Compain,*Analyse du risque de povisionnement non-vie dans le cadre de la réforme Solvabilité 2* ,Mémoire d'actuariat université Paris Dauphine, 2010

[12] V.Lagenebre,Prise en compte de la dépendance entre risques par la théorie des copules : impact sur le besoin en fonds propres de l'assureur, Mémoire d'actuariat EURIA, 2009