République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Abderrahmane Mira de Béjaïa

Faculté des Sciences Exactes
Département de Recherche Operationelles

En vue de l'obtention du diplôme de Master

Option : Mathématiques financières

Thème

Développement de quelques techniques d'estimation du coût de cotisation : Cas Société Algérienne d'Assurance (S.A.A)

Présenté par :

BENMOUHOUB ZAKARI et MAIOU SAID

Soutenu le 01/07/2024 devant le jury :

Présidente	Mme. DJERROUD Lamia	M.C.A	U. A/Mira Béjaïa
Promotrice	Mme. TAKHEDMIT Baya	M.C.A	U. A/Mira Béjaïa
Co-promoteur	M. ABBAS Karim	Proffeseur	U. A/Mira Béjaïa
Examinatrice	Mme. ABBACI Leila	M.C.A	U. A/Mira Béjaïa
Examinateur	M. SOUFIT Massinissa	M.C.B	U. A/Mira Béjaïa
Invité	M. BOURGHADEN Hamza	Président	S.A.A

2023 – 2024

Remerciements

Au terme de ce parcours universitaire, nous tenons à exprimer notre profonde gratitude envers ceux qui ont rendu cette aventure enrichissante et mémorable.

Tout d'abord, nous remercions chaleureusement nos familles, piliers inébranlables de soutien et de compréhension. Votre encouragement constant et votre amour inconditionnel ont été les sources d'inspiration qui ont éclairé notre chemin.

À nos amis et collègues, merci d'avoir partagé cette aventure avec nous. Vos encouragements, échanges d'idées et moments de détente ont apporté une dimension humaine à notre travail acharné.

Un remerciement spécial à nos professeurs, guides éclairés qui ont partagé leur expertise, stimulé notre réflexion et façonné notre apprentissage. Votre dévouement à l'éducation a laissé une empreinte indélébile sur notre parcours académique.

Enfin, à tous ceux qui ont contribué de près ou de loin à la réalisation de ce mémoire, nous vous exprimons notre reconnaissance sincère. Chacun de vous a joué un rôle essentiel dans notre réussite.

Que ces mots reflètent notre profonde gratitude envers une communauté qui a rendu possible la concrétisation de ce travail. Merci pour votre soutien indéfectible.

Bien à vous,

DÉDICACE

À **mes parents**, **frères** et **soeurs**,

Vous qui avez été ma source constante de soutien, d'amour et d'inspiration. Ce modeste travail est dédié à vous, pour toutes les valeurs que vous m'avez inculquées, pour votre encouragement sans faille et pour les sacrifices que vous avez consentis pour mon éducation. Votre influence positive est gravée dans chaque page de ce mémoire. Merci pour tout ce que vous faites et pour être les piliers de ma réussite.

Et spécialement pour mon oncle **Bouzid** et ma chère tante **Karima** qui étaient toujours avec moi durant tout mon parcours.

Finalement, a tous mes amis **Lounis Laib , Abdelouahab Adel . . .**
Avec tout mon amour et ma gratitude,

Said Maiou

DÉDICACE

À ma famille,

Ce mémoire est le fruit de vos sacrifices silencieux, de votre soutien indéfectible et de votre amour inconditionnel. Chaque ligne écrite ici porte l'empreinte de vos conseils sages et de votre exemple inspirant. Vous m'avez montré la voie avec patience et dévouement, nourrissant mes rêves et éclairant mon chemin. Que ces pages vous rendent hommage, car elles sont le reflet de votre dévouement et de votre confiance en moi.

Je dedie ce travail a mes soeurs **Cherifa , Halima, Kenza, Salma** et mes frères **Younes**, et mes amis **Redha Atmani, Lounis Laib**.

Avec une reconnaissance infinie et un amour éternel.

Benmouhoub Zakari

Table des Matières

Table des figures

Liste des tableaux

Introduction Générale

L'assurance, pilier essentiel de la gestion des risques, a connu une évolution remarquable au fil des siècles, s'adaptant constamment aux besoins changeants des sociétés et des individus. En Algérie, cette histoire débute en 1845 avec l'implantation de l'Union Incendie, une compagnie française pionnière dans le pays. Le véritable essor du secteur des assurances algérien s'amorce dans les années 1950. Un tournant majeur survient en 1958 avec l'instauration de l'assurance obligatoire, notamment dans le domaine automobile. Cette mesure transforme profondément le paysage de la protection financière et juridique en Algérie. Au fil du temps, l'assurance s'est imposée comme une nécessité première, tant pour les particuliers que pour les entreprises. Elle offre aujourd'hui un large éventail de garanties couvrant une multitude de risques. Les compagnies d'assurance jouent désormais un rôle crucial dans l'économie et la société algériennes. Elles apportent une sécurité financière indispensable face aux aléas de la vie, permettant aux individus et aux organisations de se prémunir efficacement contre des pertes potentiellement dévastatrices [2].

Cette étude vise à optimiser les techniques d'estimation des coûts de cotisation pour la Société Algérienne d'Assurance (S.A.A). Elle cherche à développer des méthodes plus précises, adaptées au marché algérien, pour assurer la viabilité financière de la compagnie tout en proposant des primes équitables. Nous analyserons les méthodes actuelles de la S.A.A., identifierons leurs limites, et testerons de nouvelles approches statistiques et de modélisation. L'objectif est de trouver un équilibre entre rentabilité et accessibilité des assurances, en intégrant les particularités du marché algérien et en proposant des solutions concrètes pour améliorer les pratiques de gestion des risques dans le secteur des assurances en Algérie.

Le mémoire est structuré en trois parties principales. La première partie présente les généralités sur les assurances, incluant un historique de l'assurance en Algérie, les concepts généraux, et une présentation de la S.A.A. La deuxième partie est dédiée aux modèles de risque en assurance, où nous explorerons la théorie de la ruine et les différents modèles utilisés pour évaluer le risque. Enfin, la troisième partie se concentre sur l'application pratique des techniques d'estimation des coûts de cotisation, avec des études de cas et des analyses basées sur des données réelles.

En conclusion, cette étude vise à contribuer à une meilleure compréhension des mécanismes d'estimation des coûts dans le secteur des assurances en Algérie, et à proposer des outils pratiques pour les compagnies d'assurance, afin d'optimiser leurs stratégies de tarification et de gestion des risques.

1

Généralités sur l'actuariat

Partie I :Généralités sur les assurances

1.1 Introduction

L'assurance est un mécanisme de couverture de risques qui a évolué au cours des siècles pour répondre aux besoins des sociétés et des individus. En Algérie, les premiers signes d'une activité d'assurance apparaissent en 1845 avec l'arrivée de la société française Union Incendie. Le marché algérien de l'assurance a commencé à se développer à partir de 1950. En 1958, l'introduction de l'assurance obligatoire, notamment l'assurance automobile obligatoire, a marqué un tournant significatif dans le paysage de la protection financière et juridique des individus. Aujourd'hui, l'assurance est devenue un bien de consommation courante, voire de première nécessité pour les particuliers et les entreprises [2].

1.2 Historique de l'assurance en Algérie

Pendant la période coloniale en Algérie, les besoins en assurance étaient considérés comme négligeables en raison des faibles revenus et de la situation socioculturelle des Algériens. Après l'indépendance, le développement de l'assurance a débuté, mais la plupart des opérations étaient entre les mains d'entreprises françaises, avec une grande partie des primes collectées envoyées à l'étranger. Ce n'est qu'en 1966 que l'état a repris le monopole de l'assurance en Algérie, confiant cette activité aux entreprises nationales nouvellement créées dans les années 1963-1964. En 1976, les quatre principales compagnies d'assurance nationales - CAAR (compagnie algérienne d'assurance et de réassurance), SAA (la société algérienne d'assurance), MAATEC (mutuelle assurance algérienne des travailleurs de l'éducation et de la culture) et la CNMA (caisse nationale de mutualité agricole) – se sont spécialisées dans la couverture de différents types de risques. En 1995, une ordonnance a mis fin au monopole de l'état sur le marché de l'assurance, permettant la création de compagnies d'assurance privées. Cette ouverture a permis aux assureurs de proposer une gamme plus diversifiée de garanties, au-delà de la simple responsabilité civile, contribuant ainsi à la rentabilité du secteur [2].

1.3 Concepts généraux d'assurance

Compagnie d'assurance

Une compagnie d'assurance est une entreprise qui fournit des services d'assurance à des clients qui deviennent des assurés. Les prestations d'assurance sont effectuées soit par des sociétés, soit par des mutuelles [16].

Contrat d'assurance

L'article 2 de l'ordonnance 95/07 du 25/01/1995 dit que l'assurance est, au sens de l'article 619 du code civil, un contrat par lequel l'assureur s'oblige, moyennant des primes ou autres versements pécuniaires, à fournir à l'assuré ou au tiers bénéficiaire au profit duquel l'assurance est souscrite, une somme d'argent, une rente ou une autre prestation pécuniaire, en cas de réalisation du risque prévu au contrat [12].

Réassurance

La réassurance est un mécanisme essentiel dans le domaine de l'assurance, souvent décrit comme « l'assurance des assureurs ». Ce transfert de risques permet à la compagnie d'assurance de renforcer sa trésorerie, d'accepter plus de garanties et de mieux gérer son exposition aux risques. La réassurance peut être facultative, laissant le choix au réassureur de souscrire ou non un contrat, ou obligatoire, imposant aux parties de s'engager. Elle peut être proportionnelle, où chaque partie prend une quote-part des cotisations et des indemnités, ou non proportionnelle, où le réassureur intervient selon des seuils prédéfinis [4].

Co-assurance

La coassurance est un mécanisme dans le domaine de l'assurance où plusieurs assureurs partagent la prise en charge d'un même risque au sein d'un seul contrat d'assurance. En cas de sinistre, chaque assureur indemnise l'assuré proportionnellement à sa part de couverture du risque [10].

Sinistre

C'est la réalisation d'un événement prévu au contrat et susceptible d'entraîner la prise.En charge financière du dommage par l'assureur. La définition du sinistre est en fonction des risques couverts, par exemple, la survenance d'un incendie entraînant des dommages constituant un sinistre, la réclamation du tiers en assurance responsabilité civile constitue un sinistre, etc [16].

Réclamation

La réclamation, dit aussi perte, est une somme déboursée pour les sinistres [12].

Réserves

Les réserves mentionnées par le Plan comptable sectoriel des assurances (les réserves obligataires) [19] :

- Les réserves pour remboursement de l'emprunt pour fonds d'établissement : Constituées lorsque le remboursement n'obéit pas à l'annuité constante.
- Les réserves pour fluctuation de change : Destinées à prendre en charge les plus-values nettes des moins-values sur conversion des devises lors de l'inventaire.

• Les réserves de capitalisation : destinées à compenser la dépréciation des valeurs mobilières comprises dans l'actif de l'entreprise.

Risque

Le risque est généralement défini comme la possibilité qu'un événement assuré se produise, entraînant une perte financière pour l'assuré ou l'assureur.

Prime

C'est la contrepartie que donne l'assuré à l'assureur en échange de la garantie qui lui est accordée. En d'autres termes, c'est le prix de la garantie. Elle peut être constante, comme elle peut être variable d'un exercice à l'autre.
Il existe trois types de primes : la prime pure, nette et totale [12].

Prime pure C'est la prime permettant à l'assureur de régler les sinistres frappant la mutualité des assurés. Elle est également appelée (prime de risque) ou encore (prime d'équilibre) [11]. La prime pure est égale à la fréquence du risque multipliée par le coût moyen d'un sinistre.

Prime nette C'est la prime figurant sur les tarifs des sociétés d'assurance. Elle est parfois appelée prime commerciale. Elle est égale à l'addition de la prime pure et des chargements permettant de couvrir les frais des acquisitions et de gestion du contrat [11].

Prime totale C'est la prime payée par l'assuré. Elle est égale à l'addition de la prime nette, des frais accessoires et des taxes [11].

Coût de police

Il correspond aux frais engagés pour la réalisation du contrat, c'est un chiffre d'affaires pour la société.

Taxe sur la valeur ajoutée (TVA)

Elle est de 17% du montant (prime nette + coût de police) et est reversée.

Droits de timbre de dimension

en assurance font référence aux frais administratifs ou taxes appliqués selon la valeur ou les caractéristiques spécifiques des biens ou des risques assurés.

Droit de timbre gradué

Dédié uniquement à l'assurance automobile et matériel agricole, il est reversé.

Fond de garantie automobile (FGA)

C'est la contribution pour le fond de garantie automobile afin d'alimenter ce dernier pour l'accomplissement de sa mission. Le taux est de l'ordre de 3% de (prime nette RC + coût de police).

1.4 Présentation de la Société Algérienne d'Assurance (SAA)

La Société algérienne d'assurance, plus connue sous le sigle SAA, est l'une des principales institutions du secteur des assurances en Algérie. Créée en 1963, à peine un an après l'indépendance du pays, elle incarne la volonté des autorités de l'époque de bâtir un marché des assurances national fort et régulé.
Son cœur de métier englobe les assurances dommages pour les particuliers et les entreprises, les assurances de personnes ainsi que la réassurance. La SAA dispose d'un maillage territorial dense avec des bureaux dans la plupart des wilayats du pays, employant des milliers de collaborateurs.
En près de 60 ans d'existence, elle est devenue un acteur incontournable de la protection des biens et des personnes en Algérie, contribuant ainsi au développement socio-économique national. Longtemps en situation de quasi-monopole, elle fait désormais face à la concurrence de compagnies privées depuis l'ouverture du marché il y a une dizaine d'années [3].

1.4.1 Présentation de l'Agence 3207 AMIZOUR

Le siège de la société est situé à AMIZOUR. En Algérie, ils ont rangé les directions régionales, 5 au centre et 5 à l'ouest, et pour l'Est, il y a 5 directions . Au niveau de la direction Est, on a 3 directions régionales avec une autre a Sétif . La direction de Sétif possède 51 agences, 50 agences de type A, B, C et une agence 3207 de type A.

1.4.2 Parts de la SAA dans le marché

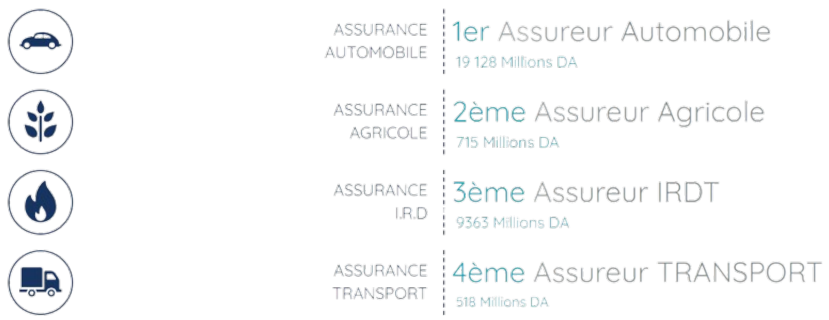

FIGURE 1.1 – Positionnement de la SAA sur le marché d'assurance en 2023 [18].

1.4.3 Différents types d'assurance

Il existe plusieurs types d'assurances, chacun conçu pour couvrir différents risques ou besoins, et tout ça pour répondre aux besoins et aux risques diversifiés, on vas citer quelques types d'assurance les plus connus en Algérie :

1. Assurance automobile, risque simple et les flottes
2. Assurance de personnes.
3. Assurance incendie et éléments naturels.
4. Assurance de responsabilités civiles générales.
5. Assurance transport.

1.4.4 Assurance automobile

L'assurance automobile est une assurance qui couvre les dommages causés "avec"ou "à" un véhicule automobile. Elle est en général obligatoire et est régie par un code des assurances. Son objectif est d'apporter une compensation financière face aux pertes subies par un assuré ou une personne tierce, notamment lors d'un accident de la route où lors de la survenance d'un évènement où le véhicule a été endommagé dans certaines conditions. Les formes de contrats comme les garanties proposées par les compagnies d'assurances sont variées. Chaque contrat souscrit est propre à une situation. Que ce soit le véhicule, les garanties choisies, le souscripteur ou la compagnie d'assurance [22].

1.4.4.1 Garanties proposées par les compagnies d'assurance

Garantie de responsabilité civile (RC)

C'est une assurance obligatoire. La responsabilité civile couvre ce que l'on appelle les « prestations civiles » pour des tiers. D'autre part, l'assureur paie à votre place les frais occasionnés aux autres lors d'un accident dont vous portez la responsabilité. Dans le cadre de cette branche de garantie, ces frais peuvent concerner les lésions corporelles (mort, blessure) ou les dégâts matériels.

En même temps, elle ne remboursera pas les dommages de votre voiture, même si elle est conduite par quelqu'un d'autre. De plus, elle ne couvrira pas vos lésions corporelles si vous conduisez et que vous êtes fautif, sauf en cas d'incapacité supérieure à 50%. Donc, c'est une assurance de dommages à caractère indemnitaire dont l'objet est la dette de responsabilité de l'assuré envers la tierce victime. Cela veut dire que lorsqu'un individu s'assure, il s'engage par l'intermédiaire de la compagnie d'assurance à indemniser une victime [3].

Garantie des dommages causés au véhicule assuré

En cas de collision avec un autre véhicule, de choc contre un corps fixe ou mobile, ou de renversement sans collision préalable, du véhicule assuré, sont garantis [3] :

- L'indemnisation des dommages que cet événement aura causés au véhicule assuré ou aux accessoires ou pièces de rechange prévues dans le catalogue du constructeur.
- Est compris dans la garantie le paiement de la réparation des dommages causés par : hautes eaux, inondations, éboulement de rochers, chutes de pierres, glissement de terrains et grêle à l'exclusion de tout autre cataclysme.

Garantie de dommage et collision

En cas de collision survenant hors des garages, remises ou propriétés, occupés par l'assuré, entre le véhicule assuré et, soit un piéton identifié, soit un véhicule ou un animal domestique appartenant à un tiers identifié, la société garantit, à l'assuré, le paiement [3].

Garantie dommage sans collision (tous risques)

Cette garantie couvre les dommages causés à votre véhicule, même si vous êtes responsable de l'accident. Cela peut inclure des dommages dus à des collisions, au vol, etc [3].

Garantie de bris de glaces

La société garantit, à l'assuré la réparation ou l'indemnisation des dommages causés, au véhicule assuré à la suite d'un bris [3] :

1. du pare brise.
2. de la lunette arrière.
3. de la lunette du toit ouvrant.
4. des glaces latérales.
5. des glaces des rétroviseurs latéraux.

La garantie joue indifféremment que le véhicule soit en mouvement ou à l'arrêt.

Garantie de vole

En cas de vol ou de tentative de vol, et si des dommages en sont résultés, la SAA rembourse tous les frais engagés après l'accord préalable, notamment en cas de disparition ou de détérioration du véhicule [3].

Garantie d'incendie et exposions

Sont garantis les dommages subis par le véhicule assuré et par les accessoires et les pièces de rechange dont le catalogue du constructeur prévoit la livraison en même temps que le véhicule, lorsque ces dommages résultent de l'un des événements suivants : incendie, combustion spontanée, chute de la foudre et explosions, à l'exclusion des dommages occasionnés par tout explosif transporté illégalement ou n'ayant pas été déclaré préalablement à l'assureur [3].

Garantie défense et recours

Cette garantie permet à l'assuré qui l'achète d'obtenir le remboursement des frais d'enquête, d'expertise, et de constitution d'avocat engagé en vue [3] :

1. Soit d'obtenir la réparation des dommages matériels subis par les objets transportés ou de dommages corporels.

2. Soit pour se défendre en cas d'action en justice exercée contre lui.

Partie II : Modèle de risque en assurances

1.5 Introduction

La théorie de la ruine, dans le contexte des assurances, est une branche de la théorie des probabilités appliquée aux modèles financiers pour évaluer les risques de faillite d'une compagnie d'assurance (dans le domaine des sciences actuarielles, on dit que le surplus financier soit négatif), Cette théorie examine les probabilités de ruine financière d'une compagnie d'assurance en raison de la survenance d'événements imprévus ou de sinistres coûteux.
Dans le cadre de l'assurance, pour quantifier le risque associé à un surplus financier, il est nécessaire de modéliser son comportement. Plus précisément, On cherche à analyser la probabilité que ce excédent financier soit déficitaire., événement qu'on appellera « la ruine » [15].

1.6 Théorie de la ruine

La ruine en assurance fait référence à l'état où une compagnie d'assurance ne peut plus faire face à ses obligations financières suite à une accumulation de pertes. Il s'agit d'une situation critique qui peut entraîner la faillite de l'entreprise. La théorie de la ruine en assurance vise à définir les seuils à partir desquels une compagnie est considérée en risque de ruine et à étudier les méthodes pour la prévenir [15].
Le modèle classique de la théorie de ruine ou le modèle de Cramer-Lundberg, représente l'évolution des réserves d'une compagnie d'assurance par un processus Poisson-composé. Explicitement, si on note R_t avec $t > 0$ le processus qui estime les réserves, on a :

$$R_t = u + ct - \sum_{k=1}^{N_t} V_k \tag{1.1}$$

ou,
• u , c sont deux réels supposés positifs, représentant respectivement les réserves initiales de la compagnie et le taux de cotisation demandé aux assurés, qui est supposé constant,
• $(N_t)_{t\geq 0}$ est un processus de Poisson homogène d'intensité λ représentant le nombre de sinistres,
• V_k est une variable aléatoire positive représentant le montant du $k^{\text{ème}}$ sinistre.
• Les $(V_k)_{k\geq 1}$ sont supposées indépendantes, identiquement distribuées de $(N_t)_{t\geq 0}$.

On notera F_X leur fonction de répartition de V_k, f_X leur densité et μ leur moyenne commune. On note par la suite $(S_t)_{t_0 \geq 0}$ le processus qui représente les pertes agrégées de la compagnie avec :

$$S_t = \sum_{k=1}^{N_t} V_k, \tag{1.2}$$

On peut écrire :

$$R_t = u + ct - S_t \tag{1.3}$$

Avec $R_0 = u$

1.7 Modèles utilisés pour évaluer le risque de ruine

Modèle individuel

Le modèle individuel modélise la charge globale générée par les sinistres individu par individu. La charge totale pour un portefeuille comprenant N contrats est définie par [9] :

$$S_{ind} = \sum_{i=1}^{n} X_i \tag{1.4}$$

Où,

• X_i est une variable aléatoire positive qui indique le montant total des sinistres subis par l'assureur sur la période d'observation.

• n est le nombre de contrats.

Modèle collectif

Le modèle collectif modélise la charge totale de réclamations en s'intéressant au nombre aléatoire. de montants de sinistres indépendants et identiquement distribués. On définit la charge totale des sinistrées par la variable aléatoire positive ; En d'autres termes [9] :

$$S_{coll} = \sum_{i=1}^{N} X_i \tag{1.5}$$

Où,

• N est une variable aléatoire qui représente le nombre de sinistres.

• X_i est une variable aléatoire qui représente le montant de la $i^{\text{ème}}$ perte.

Remarque :
En effet, si tous les contrats ont les mêmes caractéristiques, alors le modèle individuel correspond exactement au modèle collectif (le modèle collectif est un cas particulier du modèle individuel) et fait intervenir les distributions composées.

Processus de réserve et de surplus

La réserve d'une compagnie d'assurance représente les fonds disponibles pour couvrir les réclamations de ses clients. Ces fonds sont alimentés par les primes d'assurance versées par les clients, constituant ainsi la seule source de revenus de la compagnie.
Le processus de réserve $\{R(t), t \geq 0\}$ d'une compagnie d'assurance prend la forme suivante :

$$R_t = u + ct - Z_t \tag{1.6}$$

ou,
• R_t représente la réserve de la compagnie d'assurance à l'instant t.
• u est la réserve initiale ou capital initial de la compagnie d'assurance.
• c est le taux auquel les primes d'assurance sont perçues (en unités monétaires par unité de temps).
• Z_t est le processus des réclamations cumulées jusqu'à l'instant t.

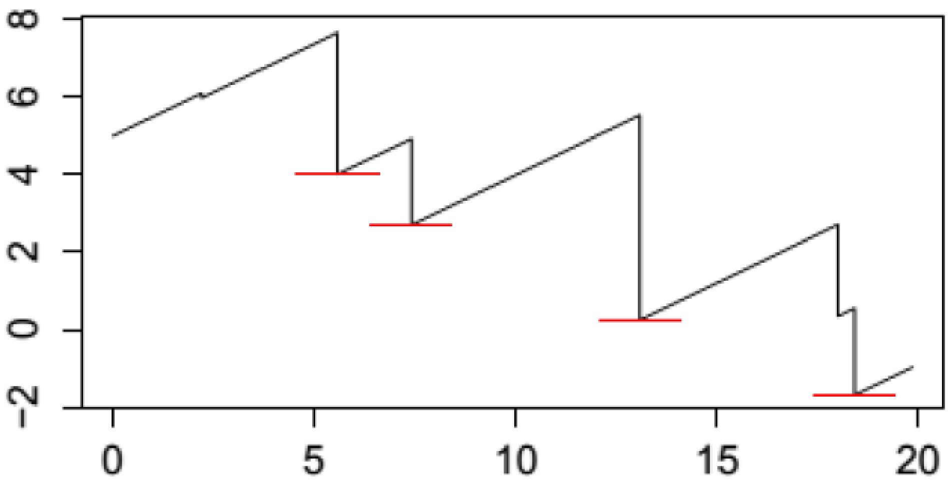

FIGURE 1.2 – Évolution du processus de réserve au cours du temps [5]

Processus de Poisson

Un processus de Poisson est un modèle statistique utilisé pour décrire le nombre d'événements se produisant dans un intervalle de temps, où les intervalles de temps entre chaque événement suivent une distribution exponentielle et sont indépendants les uns des autres.

Définition 2.1 Soit $\{T_i; i \geq 1\}$ une suite de variables aléatoires indépendantes et identiquement distribuées de loi exponentielle de paramètre $\mu > 0$. Notons $\sigma_n = \sum_{i=1}^{n} T_i$. Le processus de comptage $\{Z(t); t \geq 0\}$ défini par [5] :

$$Z(t) = \sum_{n \geq 1} 1_{\{\sigma_n \leq t\}} \tag{1.7}$$

est un processus de Poisson homogène d'intensité μ.

processus de Poisson composé

$$N(t) = \sum_{i=1}^{Z(t)} X_i \tag{1.8}$$

avec,

• $\{Z(t), t \geq 0\}$: est un processus de Poisson d'intensité $\lambda > 0$.

• $\{X_i, i \in \mathbb{N}^*\}$: est une suite de variables aléatoires indépendantes, identiquement distribuées d'espérance $E(X_i) = \mu$, où X_i est le montant du $i^{\text{ème}}$ sinistre qui est indépendant du processus de Poisson $\{Z(t), t \geq 0\}$.

• **Cas particulier : Modèle de Lundberg P/P** Le modéle de Lundberg est un cas particulier du modéle de risque classique appelé aussi P/P. Il se caractérise par la distribution exponentielle des montants des réclamations :

$$H_N(y) = 1 - \exp(-\frac{y}{\mu}) \tag{1.9}$$

où H_N est la fonction de répartition de la variable aléatoire N qui génère les montants des réclamations [7].

Probabilité de ruine

Pour tout modèle de risque d'un surplus financier, la première quantité d'intérêt est la probabilité de ruine. Et pour la définir, il faut d'abord définir le moment ou survient la ruine. τ.

Le moment de la ruine τ

Dénotons par τ la variable aléatoire du moment de la ruine [1].

$$\tau = \inf\{t \mid U(t) < 0\} \tag{1.10}$$

La probabilité de ruine peut ainsi être définie à travers le moment de la ruine. En effet, si on observe la ruine pour le processus $U(t)$, c'est que τ existe et est fini, c'est-à-dire $\tau < \infty$. Par contre, si on n'observe pas de ruine, alors τ n'existe pas ($\tau = \infty$). Ainsi, la probabilité ultime de ruine, notée $\psi(x)$, peut se définir de la manière suivante :

$$\psi(x) = \mathbb{P}(\tau < \infty \mid U(0) = x) \tag{1.11}$$

La probabilité ultime de ruine n'est autre que la probabilité que $\tau < \infty$, pour un surplus initial $U(0) = x$. Si $\tau = \infty$, on parle de survie pour le processus $U(t)$. Le cas échéant, si $\tau < \infty$, on parle de ruine. En cas de survie, la probabilité d'un tel scénario est donnée par la probabilité de survie $\Psi(u) = 1 - \psi(u)$.

Probabilité de survie

La probabilité de survie ou de non ruine, notée ϕ, est définie par [1] :

$$\Phi(u) = 1 - \Psi(u) \tag{1.12}$$

La probabilité de ruine à horizon fini

Soit $\{U(t)\}_{t\geq 0}$ un processus de surplus tel que défini en (1.2.2), dénotons par $\psi_T(u)$ la probabilité de ruine à horizon fini, pour un surplus initial $U(0) = u$, alors

$$\psi_T(u) := P(\tau < T|U(0) = u) \tag{1.13}$$

De plus, il est clair que :

$$\lim_{T\to\infty} \psi_T(u) := \psi(u) \tag{1.14}$$

est beaucoup plus facile à simuler que $\psi(u)$. De plus, grâce à la limite (1.2.7), on peut approcher $\psi(u)$ par $\psi_T(u)$, qui est la probabilité que la ruine survienne à l'intérieur de l'intervalle de temps $[0, T]$ lorsqu'il n'existe pas d'expression analytique fermée pour exprimer $\psi(u)$. Cette probabilité a donc deux utilités remarquables, la première, sa facilité de calcul, qui nous offre une alternative aux obstacles posés par la simulation de $\psi(u)$, et la seconde, d'un ordre plus pratique. En effet, si $\psi(u)$ est la première quantité à laquelle nous nous sommes intéressés pour mesurer le risque d'un modèle de surplus financier, elle n'indique pas le moment où la ruine est susceptible d'arriver. $\psi(u)$ indique simplement que la ruine surviendra, avec une certaine probabilité ; mais cela peut arriver à n'importe quel moment dans le temps. Or pour une compagnie d'assurances qui veut adopter un modèle pour son surplus, il est plus intéressant de fixer un horizon de temps fini pertinent, comme une année d'exercice fiscal, pour essayer d'évaluer le risque encouru vis-à-vis d'une éventuelle ruine, au cours de cet exercice financier. Cependant il faut retenir que $\psi_T(u)$ est bien plus complexe à étudier [1].

Le processus de Lévy

Soit$\{X(i)\}_{t\geq 0}$ un processus stochastique à temps continu. On dit que $\{X(i)\}_{t\geq 0}$ est un processus de Lévy si et seulement sa fonction caractéristique ι(s) est de la forme :

$$\gamma(s) = E[e^{iuX(t)}] = e^{-t/\psi(u)} \tag{1.15}$$

où γ est l'exposant caractéristique dans la caractérisation de Lévy-Kitchine. [5] Les processus de Lévy sont des processus stochastiques dont les propriétés sont trés intéressantespour la construction des modéles de risque [1].

Définition 2.2 Un processus stochastique $\{X_t : t \geq 0\}$ défini sur $\mathbb{R}$ est un processus de Lévy s'il satisfait les conditions suivantes :

1. Pour tout choix de $0 \leq t_0 < t_1 < \ldots < t_n$ les variables aléatoires $X_{t_0}, X_{t_1} - X_{t_0}, X_{t_2} - X_{t_0}, X_{t_3} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ sont indépendantes.
2. $X_0 = 0$
3. La loi de $X_0 = 0$ ne dépend pas de s.
4. (X_t) est continu en probabilité, c'est-à-dire que pour tout $t \geq 0$ et pour tout $\epsilon > 0$
$$\lim_{s\to t} \mathbb{P}[|X_s - X_t| > \epsilon] = 0 \tag{1.16}$$
5. Il existe un ensemble $\Omega_0 \in f$ tel que tel que $\mathbb{P}[\omega_0] = 1$ et pour tout $\omega \in \Omega_0$, $X_t(\omega)$ est continu à droite en tout t $>=$ 0 et a une limite à gauche en tout t $>$ 0 où :
   - Ω est l'ensemble des résultats possibles (ou espace des événements)
   - f est une sigma-algèbre, c'est-à-dire une collection d'événements pour lesquels on peut

définir une probabilité.

• P est une mesure de probabilité définie sur les événements de f.

Un processus aléatoire vérifiant les conditions (1) a (4) est dit processus de Lévy en loi.

Processus de renouvellement

Un processus de renouvellement consiste à compter les occurrences d'un événement donné, avec des intervalles entre ces occurrences qui sont des variables aléatoires indépendantes et identiquement distribuées. Par exemple, cela peut être utilisé pour compter le nombre de pannes d'un équipement électronique en théorie de la fiabilité, où l'équipement est remplacé après chaque panne (d'où le terme "renouvellement"), pour dénombrer les arrivées de clients dans une file d'attente, ou pour enregistrer les occurrences de sinistres pour une compagnie d'assurance [1].

Condition de non ruine (Chargement de sécurité)

On Considère le modèle du risque donnée par la relation suivante [1] :

$$R(t) = u + ct - Z(t) = u + ct - \sum_{i=1}^{N(t)} X_i, \quad \forall t \geq 0 \tag{1.17}$$

Ainsi, avec le théorème de Wald , nous permet d'obtenir :

$$E(R(t)) = u + ct - \lambda\mu t \tag{1.18}$$

Avec :

• μ étant le taux de remboursement.

• λ représentele nombre moyen de sinistres par unité de temps.

Remarque :

Dans le modèle P/P (μ) est la moyenne des montants de réclamations, c'est-à-dire que ($\frac{1}{\mu}$) est le paramètre de la loi exponentielle.

Une condition qu'il est naturel d'imposer est que l'espérance de la fortune d'une compagnie d'assurance soit toujours positive, soit $c \geq \lambda\mu$.

Remarquons que la ruine peut se produire nécessairement à l'occasion d'un sinistre, c'est-à-dire à l'une des dates T_n. On a donc :

$$\psi(u) = P\left\{\exists t \geq 0 : R(t) < 0\right\} = P\left\{\exists n \geq 0 : R_{t_n} < 0\right\} \tag{1.19}$$

Or, on a :

$$R_{T_n} = u + cT_n - Z_{(T_n)} = u - \sum_{i=1}^{n}(X_i - c\tau_i) \tag{1.20}$$

D'ou :

$$\psi(u) = P\left\{\exists n \geq 0 : \sum_{i=1}^{n}(X_i - c\tau_i) > \mu\right\} = 1 - P\left\{\max_{n\geq 1}\sum_{i=1}^{n}(X_i - c\tau_i) \leq \mu\right\} \tag{1.21}$$

Par la loi des grands nombres, on sait que $\frac{1}{n}\sum_{i=1}^{n}(X_i - c\tau_i)$ converge prèsque sûrement vers

$$E(X_1 - c\tau_1) = \mu - \frac{c}{\lambda} \tag{1.22}$$

• La ruine est presque certaine si $\lambda\mu > c$, et elle converge presque sûrement vers $+\infty$.

• La ruine n'est pas prèsque sure, si $\lambda\mu < c$, qui converge prèsque sûrement vers -∞, c'est-à-dire qu'il existe une probabilité positive de non ruine à horizon infini.

• Si $\lambda\mu =$ c, la thèorie des marches alèatoires montre que $\limsup \sum_{i=1}^{n}(x_i - c\tau_i) = +\infty$ présque sûrement et donc la ruine est presque sûre.

La condition de non-ruine est donc :

$$\lambda\mu < c$$

On peut alors définir la charge de sécurité ϱ par :

$$\varrho = \frac{c - \lambda\mu}{\lambda\mu} = \frac{c}{\lambda\mu} - 1 \tag{1.23}$$

Approximations Cramer-Lundberg

Soit $\Psi(u)$, avec $u \geq 0$ la probabilité de ruine du modèle précèdent . On suppose que le chargement de sécurité relatif $\pi = \frac{c-\lambda\mu}{\lambda\mu}$ est strictement positif. On note $\overline{H}(x) = 1 - H(x)$, où H est la fonction de distribution des moments des réclamations.

En utilisant les arguments de renouvellement et en conditionnant par rapport au temps et au montant de la première réclamation, on a la probabilité de ruine qui vérifie l'équation intégrale suivante [1] :

$$\Psi(u) = \frac{\lambda}{c}\int_u^{\infty}\overline{H}(y)dy + \frac{\lambda}{c}\int_u^{\infty}\Psi(u-y)\overline{H}(y)dy. \tag{1.24}$$

En général, il est très difficile de dériver des expressions explicites de la probabilité de ruine. Cependant, sous certaines conditions convenables, on peut obtenir quelques approximations de cette quantité. Les premiers travaux sur ces approximations ont été réalisés par Cramer-Lundberg dès 1930. La condition Cramer-Lundberg stipule l'existence d'une constante $k > 0$ satisfaisant l'équation de Lundberg :

$$\int_0^{\infty} e^{kx}\overline{H}(x)dx = \frac{c}{\lambda}, \tag{1.25}$$

qui est équivalente à :

$$\int_0^{\infty} e^{kx}dG(x) = 1 + \pi. \tag{1.26}$$

Où :

$G(x) = \frac{1}{\mu}\int_0^x se^{kx}\overline{H}(y)dy$ est la distribution équilibrée de F.

Supposons que l'équation (Inégalité de Markov) est vérifiée. La formule asymptotique de la probabilité de ruine est donnée comme suit :

Si :

$\int_0^{\infty} e^{kx}dG(x) < \infty$, alors,

$$\Psi(u) \sim \frac{\pi u}{k\int_0^{\infty} ye^{ky}\overline{H}(y)dy}e^{-ku} \quad \text{quand} \quad u \to \infty, \tag{1.27}$$

avec $a(x) \sim b(x) \quad \text{quand} \quad x \to \infty \iff \lim_{u\to\infty}\frac{a(x)}{b(x)} = 1$.

Si $\int_0^{\infty} e^{kx}dG(x) = \infty$, alors,

$$\Psi(u) = O(e^{-kx}) \quad \text{quand} \quad u \to \infty.$$

Ainsi, on a l'inégalité de Lundberg :

$$\Psi(u) \leq e^{-ku}, \quad u \geq 0. \tag{1.28}$$

Approximations de Cramer-Lundberg dans le modèle P/P

La distribution des montants de réclamations dans le modèle de Lundberg est exponentielle, d'où : $\overline{H}(x) = \exp\left(-\frac{x}{\mu}\right)$ $x \geq 0$, on aura la formule de la probabilité de ruine [1] :

$$\Psi(u) = \frac{1}{1+\pi}\exp\left(-\frac{\pi u}{\mu(1+\pi)}\right) \quad u \geq 0. \tag{1.29}$$

Borne de Lundberg

Lorsque le coefficient d'ajustement π existe, l'inégalité de Lundberg garantit que la probabilité de ruine à horizon infini $\Psi(u)$ est bornée par une fonction qui décroît de façon exponentielle en fonction du capital initial u.

Théorème 1.1 [15] Supposons que le coefficient d'ajustement (la charge de sécurité) $\pi > 0$

existe. $\forall u \geq 0$:

$$a^- \exp\{-\pi u\} \leq \Psi(u) \leq a^+ \exp\{-\pi u\}, \tag{1.30}$$

où

$$a^- = \inf_{x \in [0,x_0)} \frac{e^{\pi x} \int_x^\infty (1 - H_Z(y))dy}{\int_x^\infty e^{\pi y}(1 - H_Z(y))dy} \tag{1.31}$$

et

$$a^+ = \sup_{x \in [0,x_0)} \frac{e^{\pi x} \int_x^\infty (1 - H_Z(y))dy}{\int_x^\infty e^{\pi y}(1 - H_Z(y))dy}. \tag{1.32}$$

Formule de Pollaczek-Khinchine

Théorème 1.2 [5] Pour tout $u \geq 0$:

$$\Psi(u) = \left(1 - \frac{\lambda\mu}{c}\right) \sum_{n=1}^{\infty} \left(\frac{\lambda\mu}{c}\right)^n (H_X^s)^{*n}(u), \tag{1.33}$$

où $(H_X^s)^{*n}(u) = 1-(H_X^s)^{*n}$ et $(H_X^s)^{*n}$ est la n-ième convolution de la fonction de distribution complémentaire H_X^s telle que H_X^s est la fonction de répartition de X définie par :

$$H_X^s(x) = \frac{1}{\mu} \int_0^x (1 - H_X(y))dy, \quad x \geq 0. \tag{1.34}$$

La formule de la théorie de la ruine est appelée formule de Pollaczek-Khinchine ou encore formule de Beehan.

La représentation en série infinie donnée dans La formule de la théorie de la ruine est particulièrement utile pour des considérations théoriques. Toutefois, il est également utile d'utiliser des approximations numériques de la probabilité de ruine $\Psi(u)$, telle que l'algorithme de Panjer

.

Formule de Pollaczek-Khinchine dans le modèle P/P

En utilisant la formule de P-K pour des montants de réclamations de distribution exponentielle et de moyenne μ, nous allons déduire l'expression exacte de la probabilité de ruine :

$$\Psi(u) = \left(1 - \frac{\lambda\mu}{c}\right) \sum_{n=1}^{\infty} \left(\frac{\lambda\mu}{c}\right)^n \overline{(H_Z^s)^{*n}}(u). \tag{1.35}$$

Pour des montants de réclamations exponentiels de paramètre $\frac{1}{\mu}$, nous avons

$$H_X(u) = \begin{cases} 1 - e^{-\frac{u}{\mu}} & u \geq 0 \\ 0 & \text{Ailleurs} \end{cases} \tag{1.36}$$

Calculons

$$H_X^s(u) = \frac{1}{\mu}\int_0^u (1 - H_X(y))dy, \quad u \geq 0 \tag{1.37}$$

$$H_X^s(u) = \frac{1}{\mu}\int_0^u e^{-\frac{y}{\mu}}dy = \frac{1}{\mu}\left(-\frac{1}{\frac{1}{\mu}}\right)\left[e^{-\frac{y}{\mu}}\right]_0^u = 1 - e^{-\frac{u}{\mu}}. \tag{1.38}$$

Ainsi, $H_X^s(u) = H_X(u)$, $u \in \mathbb{R}$, $(F_Z^s)^{*n}$; $u \in \mathbb{R}$, $(H_Z^s)^{*n}$ représente la n-ème convolution de (H_X^s). Puisque nous avons l'indépendance des n variables aléatoires X_i, $i = 1, \ldots, n$, de distribution commune $Exp\left(\frac{1}{\mu}\right)$ et que $H_X^s = H_X^s$, alors $(H_X^s)^{*n}$ est la fonction de répartition de la somme des n variables aléatoires X_i, $i = 1, \ldots, n$. Nous utiliserons les transformées de Laplace afin de déterminer $(H_X^s)^{*n}$.

$$\hat{L}((H_X)^{*n})(x) = [\hat{L}(H_X)(x)]^n, \quad \text{ou encore} \quad \hat{L}((h_X)^{*n})(x) = [\hat{L}(h_X)(x)]^n, \tag{1.39}$$

où f_X est la densité de probabilité des montants de réclamations X_i, $i = 1, \ldots, n$. Ainsi,

$$\hat{L}_{h_X}(x) = \int_0^{\infty} h_X(t)e^{-xt}\,dt = \int_0^{\infty} \frac{1}{\mu}e^{-\left(\frac{1}{\mu}\right)t}\,dt = \frac{1}{1+\mu x} \tag{1.40}$$

D'où

$$\hat{L}_{(h_X)^{*n}}(x) = \left[\frac{1}{1+\mu x}\right]. \tag{1.41}$$

En utilisant la table des transformées de Laplace , nous trouvons que

$$(h_X)^{*n}(x) = \frac{\frac{1}{\mu}\left(\frac{x}{\mu}\right)^{n-1}}{(n-1)!}e^{-\frac{x}{\mu}}, \quad x \geq 0 \tag{1.42}$$

qui correspond à la densité de probabilité de la loi d'*Erlang*$(\frac{1}{\mu}, n)$ dont la fonction de répartition est donnée par

$$(H_X)^n(x) = (H_X^s)^n(x) = \frac{\Gamma(n, \frac{1}{\mu}x)}{(n-1)!} = 1 - \sum_{k=0}^{n-1} \frac{(\frac{1}{\mu}x)^k e^{-\frac{1}{\mu}x}}{k!}. \tag{1.43}$$

Ce résultat signifie que la somme de n variables aléatoires indépendantes de distributions exponentielles de même paramètre $\frac{1}{\mu}$ est une loi d'*Erlang*$(\frac{1}{\mu}, n)$. Alors

$$(\bar{H}_X)^n(x) = 1 - (H_X^s)^n(x) = \sum_{k=0}^{n-1} e^{\frac{1}{\mu}x} \frac{(\frac{1}{\mu}x)^k}{k!}. \tag{1.44}$$

Ainsi,

$$\Psi(u) = (1-\rho) \sum_{n=1}^{\infty} \rho^n \sum_{k=0}^{n-1} \frac{(\frac{1}{\mu}u)^k e^{\frac{1}{\mu}u}}{k!} \tag{1.45}$$

$$= (1-\rho) e^{\frac{u}{\mu}} \sum_{n=1}^{\infty} \rho^n \sum_{k=0}^{n-1} \frac{(\frac{1}{\mu}u)^k}{k!} \tag{1.46}$$

$$= (1-\rho) e^{\frac{u}{\mu}} \sum_{k=0}^{\infty} \frac{(\frac{1}{\mu}u)^k}{k!} \sum_{n=k+1}^{\infty} \rho^n \tag{1.47}$$

$$= (1-\rho) e^{\frac{u}{\mu}} \sum_{k=0}^{\infty} \frac{(\frac{1}{\mu}u)^k}{k!} \frac{\rho^{k+1}}{1-\rho} \tag{1.48}$$

$$\rho e^{-\frac{u}{\mu}} \sum_{k=0}^{\infty} \frac{\left(\frac{\rho u}{\mu}\right)^k}{k!} = \rho e^{-\frac{u}{\mu}} e^{\rho \frac{u}{\mu}} = \rho e^{-\left(\frac{1}{\mu} - \frac{\lambda}{c}\right)u} \tag{1.49}$$

Finalement,

$$\psi(u) = \frac{\lambda\mu}{c} \exp\left\{-u\left(\frac{c-\lambda\mu}{c\mu}\right)\right\}. \tag{1.50}$$

Autres approches

En plus de l'approche stochastique pour l'évaluation de la probabilité de ruine, qui possède de large champs d'application dans les modèles de risques, il existe plusieurs autres approches. Ces approches permettent une meilleure considération des faits, car certains faits ignorés dans la modélisation stochastique se retrouvent dans d'autres domaines. C'est le cas des réactions des assureurs et des assurés dans la théorie des jeux. En général, les solutions proposées pour estimer la probabilité de ruine sont basées sur : les théorèmes limites des marches aléatoires , les représentations matricielles avec modèles markoviens , la théorie des martingales et intégrales de probabilités , les méthodes d'optimisation ,les transformations analytiques , et la théorie de distributions .

1.8 Conclusion

Dans ce chapitre, nous avons présenté les principaux résultats de la théorie de ruine, des expressions exactes, des approximations et des bornes de la probabilité de ruine pour le modèle de risque classique.

Une attention particulière est portée sur le modèle de Cramér-Lundberg (modèle classique). En combinant les différentes théories de la théorie du risque déjà à plusieurs études, et de nombreux résultats pour la probabilité de ruine existent déjà pour ce modèle.

2

Approches d'estimation dû coût de cotisation

2.1 Introduction

L'objectif de ce chapitre est de décrire quelques techniques méthodologiques utilisées pour l'évaluation des risques d'assurance.

2.2 Régression linéaire

La régression linéaire est l'une des techniques statistiques les plus fondamentales et largement utilisées en science des données et en recherche. Elle permet de modéliser et d'analyser la relation entre une variable dépendante et une ou plusieurs variables indépendantes. À travers une équation linéaire, elle cherche à estimer comment les changements dans les variables indépendantes influencent la variable dépendante.

2.2.1 Régression linéaire simple

Détermination de la droite de régression

Soit un couple (X, Y) où X est une variable aléatoire indépendante et Y une variable aléatoire dépendante pour lequel, on recherche une relation du type : [17]

$$Y = \alpha + \beta X + \varepsilon$$

On suppose donc une relation du premier degré entre X et Y représentée par une droite du plan.Le coefficient de régression β est la pente de la droite .La manière d'écrire l'équation suppose que X est la variable explicative et Y est la variable expliquée.

Notons que la relation entre X et Y est une relation statistique. Rien ne permet d'affirmer une relation de cause à effet (causalité) de X sur Y [17]. Le terme $\alpha+\beta X$, constitue l'ajustement de Y en fonction de X tandis que le terme résiduel ε constitue le résidu de l'ajustement, c-à-d la partie non expliquée par la régression . Il faut trouver les valeurs de α et β (estimateurs). Ensuite, il convient d'apprécier la qualité de l'ajustement ainsi réalisé.

Remarque2.1. Notre connaissance a priori sur le modèle aurait pu être $Y = \alpha_1 + \beta_1 e^X + \varepsilon$ ou encore $Y = \alpha_2 + \beta_2 \sqrt{X} + \varepsilon$. Ce sont toujours des modèles de régression linéaire. D'autres modèles non linéaires se ramènent plus facilement au modèle linéaire par de simples transformations. Ainsi le modèle :

$$Y = \alpha X^\beta + \varepsilon, \tag{2.1}$$

Très utilisé en économétrie, il devient un modèle linéaire en passant aux logarithmes . En effet,

$$\log Y = \log(\alpha X^\beta + \varepsilon) = \log \alpha + \beta \log X + \log \varepsilon \tag{2.2}$$

$$Y' = \alpha_1 + \beta X' + \varepsilon' \tag{2.3}$$

Il en est de même pour le modèle à croissance exponontielle

$$Y = \alpha e^{\beta X} + \varepsilon \tag{2.4}$$

En passant aux logarithmes,

$$\log Y = \log(\alpha e^{\beta X} + \varepsilon) = \log \alpha + \beta X \log e + \log \varepsilon \tag{2.5}$$

$$Y' = \alpha_1 + \beta X + \varepsilon' \tag{2.6}$$

Cependant, le modèle $Y = \alpha + e^{\beta X} + \epsilon$, n'est pas linéarisable.

Remarque 2.2 La méthode qui va être développée, s'applique encore si la variable X n'est pas aléatoire, mais contrôlée par l'expérimentateur. On parle alors de modèle linéaire plutôt que de régression linéaire. C'est parce que les propriétés de la méthode des moindres carrés utilisée ne dépendent que des lois conditionnelles à X fixé que l'on peut traiter indifféremment la régression linéaire et le modèle linéaire par les mêmes techniques. [13]

Estimation des paramètres

Pour l'estimation des paramètres α et β, il suffit de réaliser n expériences en $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ qui ne sont autre que la réalisation d'un n-échantillon $Y_1, Y_2, \ldots, Y_n$ aux points $x_1, x_2, \ldots, x_n$ fixes. On suppose que les Y_i sont indépendants et de même variance σ^2. [13]

Pour procéder à des tests et à la détermination d'intervalles de confiance, il est nécessaire de supposer que les variables Y_i soient gaussiennes. [17]

$$Y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, 2, \ldots, n$$

$$E(Y_i) = \alpha + \beta x_i \quad \text{et} \quad V(Y_i) = \sigma^2, \quad i = 1, 2, \ldots, n$$

On suppose que ϵ_i suit une loi normale $\mathcal{N}(0, \sigma)$ telle que $E(\epsilon_i) = 0 \, \forall \, i \neq j$ (ϵ_i indépendants).

Le principe des moindres carrés (loi de Gauss), consiste à minimiser la somme des carrés des résidus en fonction de a et b : [21]

$$Q(\alpha, \beta) = \sum_{i=1}^{n} (Y_i - (\alpha + \beta x_i))^2 = \sum_{i=1}^{n} \epsilon^2$$

Pour obtenir le minimum de cette fonction, on exprime que les dérivées partielles premières de $Q(a, b)$ par rapport à a d'une part et par rapport à b d'autre part sont nulles :

$$\frac{\partial Q(\alpha, \beta)}{\partial \alpha} = 0 \quad \text{et} \quad \frac{\partial Q(\alpha, \beta)}{\partial \beta} = 0$$

$$\frac{\partial Q(\alpha, \beta)}{\partial \alpha} = \sum_{i=1}^{n} \frac{\partial}{\partial \alpha} (Y_i - (\alpha + \beta x_i))^2 = \sum_{i=1}^{n} 2 (Y_i - \alpha - \beta x_i) \left(\frac{\partial}{\partial \alpha} (Y_i - \alpha - \beta x_i) \right)$$

$$= \sum_{i=1}^{n} 2 (Y_i - \alpha - \beta x_i) (-1) = \sum_{i=1}^{n} (Y_i - \alpha - \beta x_i) = 0 \quad \Rightarrow \quad \alpha = \bar{Y} - b\bar{x}$$

En procédant de la même manière pour la dérivée partielle de $Q(\alpha, \beta)$ par rapport à β, on obtient :

$$\frac{\partial Q(\alpha,\beta)}{\partial \beta} = \sum_{i=1}^{n} \frac{\partial}{\partial \beta} \left(Y_i - (\alpha + \beta x_i)\right)^2 = \sum_{i=1}^{n} 2\left(Y_i - \alpha - \beta x_i\right) \left(\frac{\partial}{\partial \beta}(Y_i - \alpha - \beta x_i)\right)$$

$$\begin{aligned} &\sum_{i=1}^{n}[2(Y_i - \alpha - \beta x_i)(-x_i)] = -2\sum_{i=1}^{n} x_i(Y_i - \alpha x_i - \beta x_i^2) = 0 \\ &\Longrightarrow \sum_{i=1}^{n} x_i Y_i - \sum_{i=1}^{n} \alpha x_i - \beta \sum_{i=1}^{n} x_i^2 = 0. \end{aligned}$$

Remplaçons dans cette équation α par $\bar{y} - \beta\bar{x}$, ce qui donne l'équation en β suivante

$$\sum_{i=1}^{n} x_i Y_i - \bar{y}\sum_{i=1}^{n} x_i - \beta \sum_{i=1}^{n}\left(x_i^2 - \bar{x}\sum_{i=1}^{n} x_i\right) = 0.$$

On obtient :

$$\beta = \frac{\frac{1}{n}\left(\sum_{i=1}^{n} x_i Y_i\right) - \bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^{n}(x_i^2 - \bar{x}^2)} = \frac{S_{xy}}{S_x^2}$$

Par conséquent,

$$\hat{\alpha} = \bar{Y} - \left(\frac{S_{xy}}{S_x^2}\right)\bar{x}.$$

Et

$$\hat{Y} = \hat{\alpha} + \hat{\beta}x = (\bar{Y} - \bar{\beta}\bar{x}) + \bar{\beta}x = \bar{Y} + \bar{\beta}(x - \bar{x}) = \bar{Y} + \frac{S_{xy}}{S_x^2}(x - \bar{x}).$$

Les formules pour calculer la pente β et l'ordonnée à l'origine α de la droite de régression par la méthode des moindres carrés ont été établies. Une condition essentielle est que le dénominateur de $\hat{\beta}$ ne doit pas être nul, ce qui correspond à la variance des valeurs de la variable explicative X. Cette variance est nulle uniquement si toutes les valeurs de X sont identiques, ce qui rendrait le problème insensé. La solution au problème est unique, et bien que l'on n'ait pas démontré que l'extremum de la fonction $Q(\alpha, \beta)$ est un minimum, cela est évident car la variance est toujours positive. [13]

2.2.2 Régression linéaire multiple

C'est la généralisation de la régression linéaire simple, qui est un cas particulier avec 2 variables explicatives où l'une des variables est la constante . La régression multiple s'écrit [8] :

$$Y = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \epsilon$$

Les coefficients $\beta_1, \beta_2, \ldots, \beta_k$ s'appellent les coefficients de régression.

Matriciellement, si on pose :

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}, \quad b = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \quad \text{et} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Le modèle s'écrit alors.

$$Y = X\beta + \epsilon$$

Remarque2.3. Soit Y la variable à expliquer et X une variable explicative, mais dont l'effet n'est pas linéaire. Le modèle suivant :

$$Y = a + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \beta_4 X^4 + \epsilon$$

C'est un modèle de régression multiple avec 4 régresseurs qui sont les puissances successives de la variable explicative. Il s'écrit sous forme matricielle :

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 \\ 1 & x_2 & x_2^2 & x_2^3 & x_2^4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & x_n^4 \end{bmatrix} \begin{bmatrix} a \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Ce modèle est polynominal en X, mais linéaire en les paramètres inconnus $a, \beta_1, \beta_2, \beta_3, \beta_4$. C'est un modèle linéaire statistique [8].

Estimation des paramètres

L'estimation des paramètres peut se barer sur le principe des moindres carrés qui consiste à minimiser [14] :

$$Q(\beta_1, \beta_2, \ldots, \beta_k) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} [y_i - (\beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki})]^2$$

En égalant à zéro les dérivées partielles de Q par rapport à $\beta_1, \beta_2, \ldots, \beta_k$, on obtient un système d'équations linéaires en $\beta_1, \beta_2, \ldots, \beta_k$ de la forme :

$$\begin{cases} \beta_1 \sum_{i=1}^{n} x_{1_i}^2 + \beta_2 \sum_{i=1}^{n} x_{1_i} x_{2_i} + \cdots + \beta_k \sum_{i=1}^{n} x_{1_i} x_{k_i} = \sum_{i=1}^{n} x_{1_i} y_i \\ \beta_1 \sum_{i=1}^{n} x_{1_i} x_{2_i} + \beta_2 \sum_{i=1}^{n} x_{2_i}^2 + \cdots + \beta_k \sum_{i=1}^{n} x_{2_i} x_{k_i} = \sum_{i=1}^{n} x_{2_i} y_i \\ \vdots \\ \beta_1 \sum_{i=1}^{n} x_{1_i} x_{k_i} + \beta_2 \sum_{i=1}^{n} x_{2_i} x_{k_i} + \cdots + \beta_k \sum_{i=1}^{n} x_{k_i}^2 = \sum_{i=1}^{n} x_{k_i} y_i \end{cases}$$

Dans le cas de la régression avec terme indépendant β_0, en supposant que la dernière colonne de X soit remplie de 1, le dernière équation devient :

$$\beta_1 \sum_{i=1}^{n} x_{1_i} + \beta_2 \sum_{i=1}^{n} x_{2_i} + \cdots + n\beta_0 = \sum_{i=1}^{n} y_i$$

C'est-à-dire

$$\beta_1 \bar{x}_1 + \beta_2 \bar{x}_2 + \cdots + \beta_0 = \bar{y}$$

Ce système d'équations linéaires peut se mettre sous la forme matricielle $M \cdot \beta = m$, où

$$M = \begin{pmatrix} \sum_{i=1}^{n} x_{1_i}^2 & \sum_{i=1}^{n} x_{1_i} x_{2_i} & \cdots & \sum_{i=1}^{n} x_{1_i} \\ \sum_{i=1}^{n} x_{2_i} x_{1_i} & \sum_{i=1}^{n} x_{2_i}^2 & \cdots & \sum_{i=1}^{n} x_{2_i} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} x_i x_{1_i} & \sum_{i=1}^{n} x_i x_{2_i} & \cdots & n \end{pmatrix}$$

et

$$m = \begin{pmatrix} \sum_{i=1}^{n} x_{1_i} y_i \\ \sum_{i=1}^{n} x_{2_i} y_i \\ \vdots \\ \sum_{i=1}^{n} y_i \end{pmatrix}$$

Notons que m peut se mettre sous la forme d'un produit matriciel

$$m = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{i_n} \\ x_{21} & x_{22} & \cdots & x_{2_n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{k_1} & x_{k_2} & \cdots & xk_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = X^tY,$$

où X^t désigne la matrice transposée de X. De même, M peut aussi s'écrire $M = X^tX$. Les coefficients de régression s'obtiennent donc comme solution du système [14] :

$$(X^tX)\beta = X^tY.$$

Si ce système possède une solution unique, la matrice M carrée $k \times k$ est inversible, c'est-à-dire qu'il existe une matrice $k \times k$ qui comporte partout des zéros sauf sur la diagonale principale qui est remplie de 1. La solution de ce système par la méthode des moindres carrés sera :

$$\hat{\beta} = (X^tX)^{-1}X^tY.$$

$$E(\hat{\beta}) = E((X^tX)^{-1}X^tY) = (X^tX)^{-1}X^tE(Y).$$

Or $E(Y) = E(X\beta + \epsilon) = X\beta + E(\epsilon) = X\beta$. Donc

$$E(\hat{\beta}) = (X^tX)^{-1}X^tX\beta = \beta.beta$$

$\hat{\beta}$ est donc un estimateur sans biais de b.

$$V(\hat{\beta}) = V((X^tX)^{-1}X^tY) = (X^tX)^{-1}X^tV(Y)(X^tX)^{-1}.$$

Or $V(Y) = V(X\beta + \epsilon) = V(\epsilon) = \sigma^2 I_n$. Donc

$$V(\hat{\beta}) = (X^tX)^{-1}X^t\sigma^2 I_n X(X^tX)^{-1} = \sigma^2(X^tX)^{-1}.$$

On peut maintenant calculer les valeurs ajustées $\hat{y}_i$, $i = 1, \ldots, n$ et les résidus ϵ_i par :

$$\hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 x_{11} + \hat{\beta}_2 x_{21} + \ldots + \hat{\beta}_k x_{k1} \\ \hat{\beta}_1 x_{12} + \hat{\beta}_2 x_{22} + \ldots + \hat{\beta}_k x_{k2} \\ \vdots \\ \hat{\beta}_1 x_{1n} + \hat{\beta}_2 x_{2n} + \ldots + \hat{\beta}_k x_{kn} \end{bmatrix} = \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix}$$

Ces formules peuvent être réécrites sous forme matricielle. En effet, en notant X_i la i-ème ligne de X, on a [14] :

$$\hat{y}_i = \hat{\beta}_1 x_{1_i} + \hat{\beta}_2 x_{2_i} + \ldots + \hat{\beta}_k x_{k_i} = [x_{1_i}, x_{2_i}, \ldots, x_{k_i}] \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_n \end{bmatrix} = X_i \hat{\beta}$$

D'où

$$\hat{y} = X\hat{\beta} \quad \text{et} \quad \epsilon = y - \hat{y} = y - X\hat{\beta}$$

2.2.3 Objectif de la régression linéaire

La regression linéaire est une technique de modéliser la relation entre une variable dépendante et plusieurs variables indépendantes, ajustant une équation linéaire à la donnée disponible. Il cherche à trouver les coefficients des variables indépendantes qui minimisent les écarts entre les valeurs observées et les valeurs prédites par le modèle. La régression linéaire a pour simplifier la situation de n'une seule variable indépendante et à trouver le meilleur hyperplan, permettant de comprendre et quantifier la relation entre les variables.

2.2.4 Avantages et limites de la régression linéaire

Avantages

- Facile à interpréter et à expliquer.
- Efficace pour modéliser des relations linéaires simples.
- Utilisation répandue et mise en œuvre relativement simple.
- Offre des outils de diagnostic pour évaluer la qualité du modèle.
- Donne des coefficients qui mesurent l'impact des variables prédictives sur la variable cible.

2.2.5 Limites

- Ne capture pas les relations non linéaires entre les variables.
- Sensible aux valeurs aberrantes dans les données.
- Repose sur des hypothèses strictes (linéarité, normalité des résidus, etc.).
- Moins adaptée aux variables catégoriques ou aux interactions complexes.

• Peut souffrir de surajustement ou de sous-ajustement si le modèle n'est pas correctement spécifié.

2.3 Simulation

2.3.1 Introduction

La simulation est une méthode très utilisée en statistique, en économétrie et dans diverses autres disciplines scientifiques. Elle repose sur l'utilisation de modèles mathématiques pour produire des données artificielles et analyser le comportement de systèmes complexes ou de modèles théoriques. En offrant la possibilité de créer des scénarios contrôlés, elle facilite l'étude approfondie de ces systèmes [20].

2.3.2 Simulation de Monte Carlo

Serve à évaluer l'influence de l'incertitude sur les modèles de prévision. Son principe repose sur la génération de nombres aléatoires pour modéliser des phénomènes aléatoires ou incertains dans un système donné [20].

2.3.3 Étapes de la Simulation

• **Définition du Modèle** : Choisir et définir clairement le modèle mathématique ou statistique à utiliser.
• **Identification des Variables Aléatoires** : Déterminer quelles variables sont aléatoires et spécifier leurs distributions.
• **Génération des Variables Aléatoires** : Utiliser des générateurs de nombres pseudo-aléatoires pour simuler des valeurs pour les variables aléatoires.
• **Répétition des Simulations** : Répéter les calculs un grand nombre de fois (généralement des milliers ou des millions) pour obtenir une distribution des résultats.
• **Analyse des résultats** : Analyser les résultats des simulations pour en tirer des conclusions statistiques et probabilistes.

2.3.4 Avantages et limites

Avantages

• Permet de modéliser des systèmes complexes et réalistes.
• Peut fournir des estimations robustes même en présence d'incertitudes importantes.
• Peut être utilisée pour évaluer des stratégies ou des scénarios alternatifs sans avoir besoin de solutions analytiques exactes.

Limites

• La précision des résultats dépend de la qualité des données aléatoires générées et de la taille de l'échantillon.
• Peut être coûteuse en termes de temps de calcul pour des modèles très complexes ou nécessitant un grand nombre de simulations.

2.4 Conclusion

En résumé, ce chapitre a présenté les principales méthodologies utilisées dans cette étude. Chaque méthode a ses propres avantages et limites, et leur combinaison permet une évaluation plus robuste et précise. Ces techniques offrent des perspectives intéressantes pour des recherches futures et des améliorations méthodologiques.

3

Application des techniques de modélisation

3.1 Introduction

Dans ce chapitre, nous allons mettre en pratique les concepts théoriques abordés dans les chapitres précédents en appliquant des techniques de modélisation statistique sur des données réelles issues du secteur de l'assurance. L'objectif principal est d'estimer les coefficients de différents facteurs influençant une variable d'intérêt, notamment les primes d'assurance.

Nous commencerons par une description détaillée des données utilisées, en présentant leur origine, leur structure et leurs caractéristiques principales. Ensuite, nous introduirons le modèle de base qui prend en compte un seul facteur explicatif, avant de proposer l'ajout d'un second facteur pour enrichir notre modèle.

3.2 Description des données

Les données utilisées dans cette étude proviennent d'une compagnie d'assurance (Société Algérienne d'Assurance) et couvrent une période d'une année. Elles sont fournies sous forme de tableau et comprennent diverses informations sur les conducteurs et leurs primes d'assurance :

- **Num :**Identifiant unique de chaque enregistrement.
- **ID :** Identifiant unique de chaque assuré.
- **Nom :** Nom complet de l'assuré.
- **Effet déchéance (e.ech) :** Date à laquelle la police d'assurance prend effet et expire.
- **Garantie RC (Responsabilité Civile) :** Montant de la garantie de responsabilité civile.
- **Garantie DC (Dommages et Collisions) :** Montant de la garantie de dommages et collisions.
- **BDG :** Montant constant pour bon de garantie.
- **Assistance véhicule :** Montant constant pour l'assistance véhicule.
- **RachVet :** Montant constant pour le rachat de vétusté.
- **PEA :**Montant constant pour une prime d'événements accidentels.
- **Prime nette à payer(REC) :** Montant final de la prime que l'assuré doit payer.

NUM	ID	NOM	E.ECH	RC	DC	BDG	ASSI VEH	RachVet	PEA	REC
1	111	P_1	02/09/24	2590.69	5569.66	850	1150	300	1000	10190.35
2	222	P_2	03/09/24	2076.40	4930.56	850	1150	300	1000	10734.96
3	333	P_3	02/09/24	2606.28	5117.47	850	1150	300	1000	9889.75

Table 3.1 – Tableau de données

3.3 Estimation des Coûts de cotisation

3.3.1 Indisponibilité des Données Confidentielles

Lors de la préparation de cette étude, nous avons initialement cherché à obtenir les données réelles de sinistres et de kilométrage directement auprès de la Société Algérienne d'Assurance (initial). Ces données étaient essentielles pour calculer avec précision les paramètres de sinistralité nécessaires à l'ajustement des primes d'assurance. Toutefois, en raison de la nature hautement confidentielle des informations relatives aux sinistres et aux données clients, la initial n'a pas été en mesure de nous fournir ces données. La protection des données personnelles et la confidentialité des informations internes sont des priorités strictes pour les compagnies d'assurance, rendant l'accès à ces données impossible pour des études externes. Face à cette indisponibilité, nous avons pris la décision de simuler un ensemble de données représentatives basé sur des paramètres réels. Cette simulation s'est basée sur des hypothèses réalistes et des scénarios plausibles, en utilisant des distributions statistiques qui reflètent les tendances observées dans des études similaires. En générant des données simulées, nous avons pu procéder à l'estimation des paramètres α_1 et α_2 et à l'application de notre modèle de régression linéaire, tout en respectant les contraintes de confidentialité imposées par la initial. Cette démarche nous a permis de mener à bien notre étude, tout en soulignant l'importance de la confidentialité des données dans le secteur de l'assurance.

3.3.2 Modèle de base

Dans le cadre de la gestion des risques et de la tarification des primes d'assurance, il est crucial pour les compagnies d'assurance de disposer de modèles précis et fiables pour évaluer les risques et fixer les primes adéquates. L'une des principales difficultés rencontrées par les assureurs est de déterminer une prime juste et équitable qui reflète précisément le niveau de risque associé à chaque assuré. Cette tâche devient encore plus complexe lorsque des facteurs dynamiques et variables entrent en jeu. Le modèle de base considéré est une régression linéaire simple, où nous cherchons à modéliser le montant de la prime d'assurance Y en fonction d'une seule variable explicative X_1.

Le modèle est défini comme suit :

$$Y = \alpha_1 X_1 + c + \epsilon \tag{3.1}$$

où :

• Y est le montant de la prime d'assurance à payer.

• X_1 est le prix du véhicule, choisi comme variable explicative principale pour ce modèle.

• α_1 est le coefficient associé à X_1, qui représente le coût de la cotisation d'assurance par rapport au prix du véhicule.

• c est l'intercept du modèle.

• ϵ est le terme d'erreur, qui capture les variations non expliquées par le modèle.

3.3.3 Introduction d'un nouveau facteur

Nous proposons d'enrichir notre modèle initial en intégrant un nouveau facteur explicatif pour améliorer la précision des estimations des primes d'assurance. Le modèle initial est le suivant :

$$Y = \alpha_1 X_1 + c \tag{3.2}$$

où :

- Y est la variable dépendante représentant la prime d'assurance,
- X_1 est le prix du véhicule, α_1 est le coefficient associé à $_1$ et c est la constante.

Objectifs de l'étude

Face à ces limitations, l'objectif de cette étude est de proposer une nouvelle estimation pour le calcul des primes payées par les assurés, en intégrant le kilométrage parcouru comme facteur clé. Cette approche vise à améliorer la précision et l'équité de la tarification des primes en fonction de l'usage réel des véhicules. Plus spécifiquement, cette étude vise à : Calculer un nouveau paramètre α_2 , représentant le taux de sinistralité par kilomètre, en utilisant des données simulées. Comparer les valeurs de α_2 avec le taux de sinistralité existant α_1 fourni par initial.

Justification de l'Ajout du Facteur

Pour améliorer le modèle, nous proposons d'ajouter une nouvelle variable explicative X_2, représentant le kilométrage annuel du véhicule. Cette décision est motivée par plusieurs raisons fondamentales qui rendent ce paramètre particulièrement pertinent pour évaluer le risque d'assurance. Le choix du kilométrage est justifié par rapport à plusieurs critères :

Corrélation directe avec le risque : Le kilométrage est directement corrélé au risque de sinistre. Plus un véhicule est utilisé, plus il est exposé à des conditions de circulation variées et à des situations potentiellement dangereuses. Cela signifie que chaque kilomètre parcouru augmente la probabilité de rencontrer un événement de sinistre.

Données disponibles et fiables : Les informations sur le kilométrage sont couramment enregistrées et disponibles via les relevés de compteurs kilométriques des véhicules. Cela permet aux assureurs de collecter facilement des données historiques et actuelles pour évaluer et ajuster les primes.

Le nouveau modèle proposé devient :

$$y = \alpha_1 X_1 + \alpha_2 X_2 + c \tag{3.3}$$

ou α_2 est le coefficient associé à la nouvelle variable X_2 qui est le kilométrage.

3.4 Calcul des coefficients α_1 et α_2

Calcul de α_1

Nous avons recueilli les cotisations de trois personnes, notées (α_1), représentées chacune par sa valeur. Pour obtenir une estimation plus précise, nous avons calculé la moyenne des cotisations de chaque personne. Les résultats sont présentés dans le tableau ci-dessous : Pour

Personne	α_1
Personne 1	**0.002590**
Personne 2	**0.002595**
Personne 3	**0.002606**

TABLE 3.2 – Tableau des α_1 réels

obtenir une estimation plus précise, nous pouvons calculer la moyenne des coups de cotisation de chaque personne :

$$\bar{\alpha_1} = \frac{\sum_{i=1}^{n} \alpha_i}{n}(i = 1, 2, 3) \tag{3.4}$$

	$\bar{\alpha_1}$
Moyenne	0.002597

TABLE 3.3 – Tableau de moyenne des α_1

La moyenne globale des cotisations des trois personnes, basée sur la valeur de α_1, est estimée à $\alpha_1 = 0.002597$. Cette estimation nous donne une indication plus précise des cotisations attendues.

Calcul de α_2

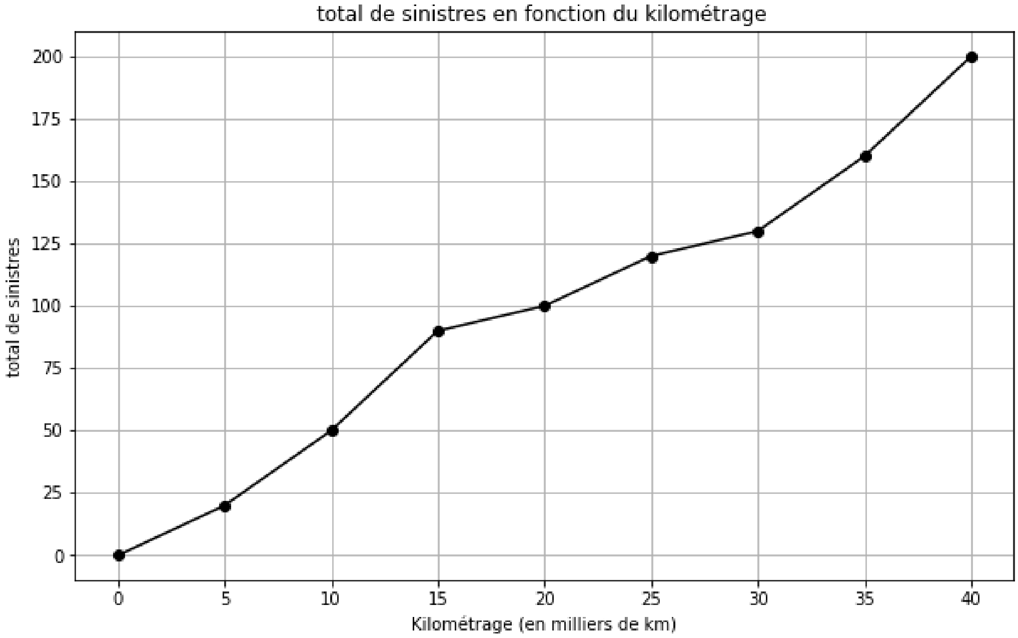

FIGURE 3.1 – total de sinistres en fonction du kilométrage [6]

$$\alpha_2 = \frac{S_t}{K_t} \tag{3.5}$$

• α_2 = Taux de sinistralité par kilomètre

• S_t =total de sinistres

• K_t = Nombre Kilométrage total

Calculer le kilométrage total et le nombre total de sinistres

• Nombre Kilométrage = 0+5000+10000+15000+20000+25000+30000+35000+40000 = 180000

• total de sinistres= 0+20+50+90+100+120+130+160+200= 870

Alors :

$$\alpha_2 = \frac{S_t}{K_t} = \frac{870}{180000} = 0.004833 \tag{3.6}$$

3.5 Estimation des coefficients par régression linéaire

3.5.1 Algorithme utilisé

Algorithm 1 Estimation des coefficients par régression linéaire avec visualisation

1: **Importer les bibliothèques :** `Numpy, matplotlib.pyplot, sklearn.linearmodel`
2: **Fixer la graine aléatoire** pour la reproductibilité.
3: **Générer des données simulées :**
- Générer 1000 prix de véhicules entre 500 000 et 1 000 000
- générer 1000 kilométrages entre 0 et 180000.

4: **Définir les coefficients réels :** α_1=0.002597, α_2=0.004833, constante=2630
5: **Calculer la variable dépendante :** $Y = \alpha_1 * X_1 + \alpha_2 * X_2 + c + \epsilon$
6: **Créer la matrice des caractéristiques :** Combiner prix et kilométrage
7: **Régression linéaire :** Utiliser LinearRegression pour estimer α_1 et α_2
8: **Extraction des coefficients estimés**
9: **Affichage des résultats :** Afficher les coefficients estimés et l'intercept
10: **Visualisation des résultats :**
- Créer un graphique avec deux sous-plots 3D.
- Afficher les données simulées dans le premier sous-plot.
- Afficher la régression linéaire dans le second sous-plot.

11: **Afficher le graphique**

3.5.2 Résultats obtenus

Les résultats obtenus à partir de l'estimation des coefficients du modèle de régression linéaire sont les suivants :

	α_1	α_2	c
Valeurs estimés	0.0024481136986726913	0.004106983687198785	2783.436083093525

TABLE 3.4 – Valeurs estimées par la régression linéaire

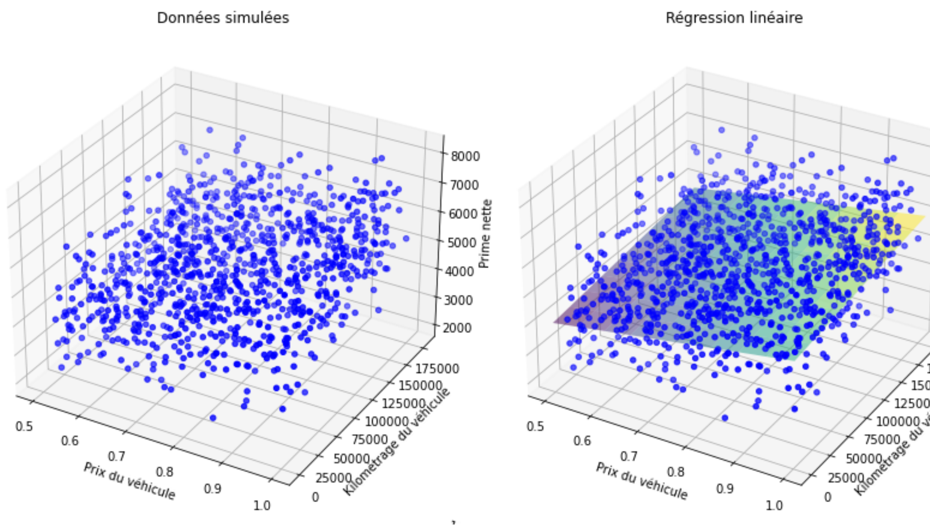

FIGURE 3.2 – Représentation graphique d'estimation des coefficients par régression linéaire

3.5.3 Interprétation des Données

Données Simulées

Sur le graphique, intitulé "Données simulées", chaque point bleu représente un ensemble de valeurs pour le prix du véhicule, le kilométrage du véhicule, et la prime nette. La distribution semble assez aléatoire, ce qui est attendu étant donné que les données ont été générées de manière aléatoire avec un bruit ajouté pour simuler des erreurs.

Régression Linéaire

Sur le graphique, intitulé "Régression linéaire", nous voyons à nouveau les points de données simulées (en bleu). En plus de ces points, une surface colorée est ajoutée pour représenter le modèle de régression linéaire ajusté aux données. Cette surface montre la relation estimée entre le prix du véhicule, le kilométrage, et la prime nette selon le modèle de régression linéaire.

Détail des Coefficients

Les coefficients estimés de modèle de régression linéaire sont les suivants :

- Coefficient pour α_1 (prix du véhicule) : 0.0024481136986726913
- Coefficient pour α_2 (kilométrage du véhicule) : 0.004106983687198785
- c= 2783.436083093525

Ces coefficients indiquent comment chaque variable indépendante (prix du véhicule et kilométrage) influence la variable dépendante (prime nette) dans le modèle.

α_1 (0.002448)

Cela signifie que pour chaque augmentation de 1 unité du prix du véhicule, la prime nette augmente en moyenne de 0.002448 unités, en maintenant le kilométrage constant. La différence entre le coefficient réel et le coefficient estimé pour α_1 est de 0.000149, ce qui indique que l'estimation est proche de la valeur réelle, bien qu'il y ait une légère sous-estimation.

α_2 (0.004107)

Cela signifie que pour chaque augmentation de 1 unité du kilométrage du véhicule, la prime nette augmente en moyenne de 0.004107 unités, en maintenant le prix constant. La différence entre le coefficient réel et le coefficient estimé pour α_2 est de 0.000726, ce qui indique une sous-estimation par le modèle.

Intercept (2783.436)

L'intercept représente la valeur de la prime nette lorsque le prix du véhicule et le kilométrage sont tous les deux égaux à zéro. Bien que cela ne soit pas réaliste dans un contexte pratique (car un véhicule ne peut avoir un prix de 0 et un kilométrage de 0), cela donne une base à partir de laquelle les effets des autres variables sont mesurés. La différence entre l'intercept réel et l'intercept estimé est de 153.436, ce qui montre une légère sur-estimation dans l'estimation de l'intercept par le modèle.

3.6 Algorithme de Calcul des Primes

Algorithm 2 Estimation des coefficients par régression linéaire avec visualisation

1: **Importer les bibliothèques :** `Numpy, matplotlib.pyplot, sklearn.linear_model`

2: **Fixer la graine aléatoire** pour la reproductibilité : `np.random.seed(0)`

3: **Générer des données simulées :**

- Générer 1000 prix de véhicules entre 500 000 et 1 000 000 : `prix = np.random.uniform(500000, 1000000, n_samples)`
- Générer 1000 kilométrages entre 0 et 180 000 :`kilometrage = np.random.uniform(0, 180000, n_samples)`

4: **Définir les coefficients estimés et constants :**

- Coefficient estimé α_1 : `alpha_1_estimé`
- Coefficient estimé α_2 : `alpha_2_estimé`
- Constante estimée : `constante_estimée`

5: **Définir les coefficients réels et constants :**

- Coefficient réel α_1 : `alpha_1_reel`
- Coefficient réel α_2 : `alpha_2_reel`
- Constante réelle : `constante_reelle`

6: **Calculer la variable dépendante** avec les coefficients estimés :

- `erreur = np.random.normal(0, 1000, n_samples)`
- `prime_nette_simulée = alpha_1_estime * prix + alpha_2_estime * kilometrage + constante_estime + erreur`

7: **Calculer les primes nettes réelles** :

- `prime_nette_reelle = alpha_1_reel * prix + constante_reelle`

8: **Visualisation des résultats** :

- Créer un graphique avec deux sous-plots 3D. • Afficher les données simulées dans le premier sous-plot :
- Afficher les primes nettes réelles dans le second sous-plot :

9: **Afficher le graphique** :

3.6.1 Résultats obtenus

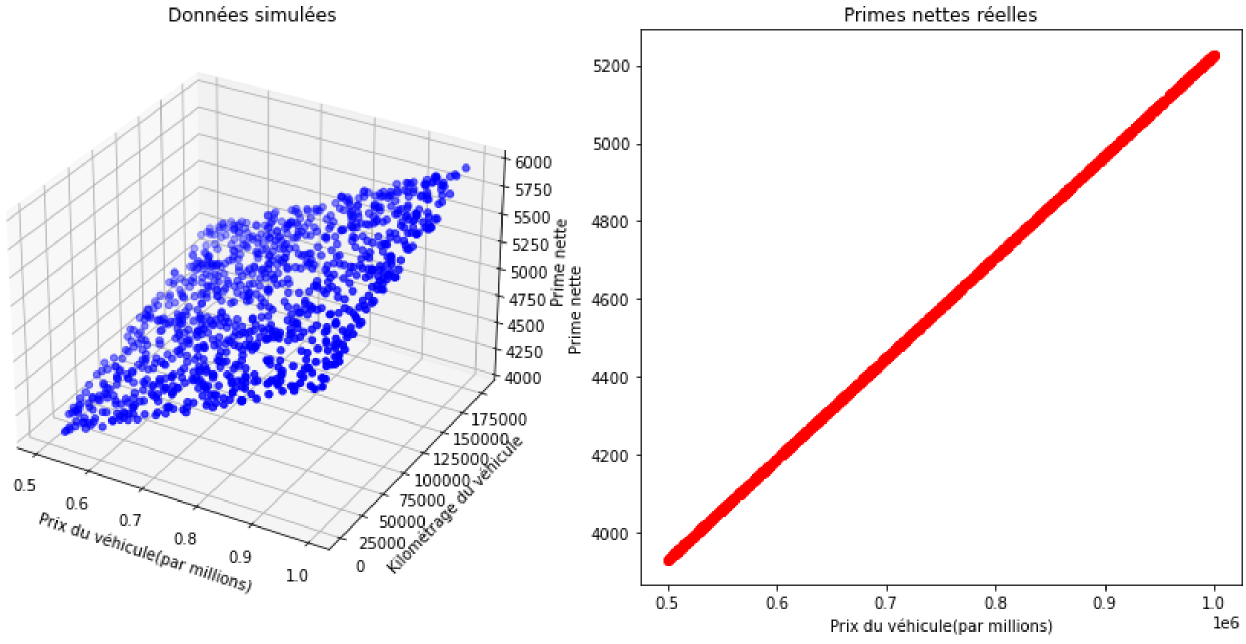

FIGURE 3.3 – Résultats obtenus par l'algorithme

3.6.1.1 Analyse des résultats

Dans cette analyse, nous allons examiner et comparer deux modèles de tarification des primes d'assurance automobile : le modèle initial et un nouveau modèle ajusté en fonction du kilométrage. L'objectif est d'identifier les forces et les faiblesses de chaque modèle, ainsi que leur impact sur les assurés selon différents niveaux de kilométrage du véhicule.

Contexte

• **Modèle initial :** Ce modèle applique une prime fixe basée uniquement sur le prix du véhicule, indépendamment du kilométrage.

• **Nouveau Modèle :**Ce modèle ajuste la prime en fonction du kilométrage, offrant une tarification plus précise et équitable basée sur l'usage réel du véhicule.

Analyse des graphes

Prix du Véhicule [0.5, 1.0] par Kilométrage

Kilométrage [0, 25000] :

Le modèle initial applique une prime fixe basée uniquement sur le prix du véhicule. Cela signifie que pour des véhicules peu utilisés (faible kilométrage), la prime peut être relativement élevée sans refléter le risque réellement réduit. Le nouveau modèle ajuste la prime en fonction du kilométrage, ce qui permet de réduire la prime pour les véhicules avec un faible kilométrage. Cela offre une tarification plus précise et équitable en fonction de l'usage réel du véhicule.

Kilométrage [25000, 50000] :
Pour ce segment, le modèle initial continue d'appliquer une prime fixe, simplifiant ainsi la tarification mais ne prenant pas en compte l'augmentation modérée du risque lié à un kilométrage légèrement plus élevé. En revanche, le nouveau modèle ajuste légèrement la prime pour refléter ce kilométrage accru, ce qui peut être considéré comme plus juste. Cependant, certains préfèrent les primes constantes du modèle initial.

Kilométrage [50000, 100000] :
Le modèle initial maintient une prime fixe, avantageuse pour les véhicules utilisés de manière intensive. Les assurés bénéficient ainsi d'une stabilité dans les primes malgré un usage plus fréquent. En comparaison, le nouveau modèle augmente davantage la prime pour ce niveau de kilométrage, reflétant plus précisément le risque accru. Cela pourrait rendre les primes moins attrayantes pour les assurés avec des véhicules très utilisés.

Kilométrage [100000, 175000] :
Pour les véhicules à très haut kilométrage, le modèle initial maintient une prime fixe, évitant ainsi des augmentations significatives malgré le risque accru lié à l'usage intensif. En contraste, le nouveau modèle ajuste fortement la prime à la hausse pour ces véhicules, reflétant de manière précise le risque élevé. Cela peut rendre ce modèle moins attractif pour les assurés ayant des véhicules très utilisés.

3.6.1.2 Avantages et limites

Modèle initial :

• Avantages : Simple à comprendre et à appliquer.
• Inconvénients : Ne prend pas en compte le kilométrage, ce qui peut entraîner des primes inéquitables pour les véhicules peu utilisés ou très utilisés.

Nouveau Modèle Proposé :

• Avantages : Plus nuancé, prend en compte le kilométrage, permettant une tarification plus précise et équitable.

• Inconvénients : l'augmentation de la prime en fonction du kilométrage peut entraîner la perte de clients au profit d'autres compagnies d'assurance, concurrentes sur le marché.

3.7 conclusion

L'étude réalisée dans ce chapitre montre une comparaison entre le modèle initial utilisé par la Société Algérienne d'Assurance (SAA) et le nouveau modèle proposé, qui intègre des facteurs supplémentaires, notamment le kilométrage des véhicules. Les résultats obtenus montrent que le modèle initial (SAA) offre une approche simple et directe pour l'estimation des primes d'assurance, en se basant principalement sur le prix des véhicules. Cependant, cette méthode présente certaines limites, notamment une tarification potentiellement inéquitable pour les véhicules peu utilisés ou très utilisés, car elle ne prend pas en compte le kilométrage. Le nouveau modèle proposé, en incorporant le kilométrage comme un facteur supplémentaire, permet une tarification plus précise et équitable.

Conclusion générale et perspectives

L'étude menée dans ce mémoire se concentre sur l'optimisation des techniques d'estimation des coûts de cotisation pour la Société Algérienne d'Assurance (S.A.A). Elle souligne l'importance de développer une approche précise et équitable pour le calcul des primes d'assurance, en tenant particulièrement compte des spécificités du marché algérien. En comparant le modèle initial de la S.A.A, qui se base principalement sur le prix des véhicules, avec un nouveau modèle intégrant des facteurs supplémentaires tels que le kilométrage des véhicules, l'étude démontre que ce nouveau modèle permet une tarification plus précise et plus juste. Bien que cette approche puisse entraîner des ajustements significatifs pour certains assurés, elle promet une meilleure adéquation entre les primes payées et les risques réels.

Pour l'avenir, plusieurs pistes d'amélioration et de recherche peuvent être envisagées. Tout d'abord, l'intégration de nouveaux facteurs de risque pourrait enrichir les modèles de tarification. En plus du kilométrage, des facteurs comme l'historique de conduite, l'utilisation du véhicule (professionnelle ou personnelle), et les données démographiques des assurés devraient être considérés.

Ensuite, l'utilisation de techniques avancées de modélisation, telles que le machine learning et l'intelligence artificielle, pourrait encore améliorer la précision des estimations des primes. Ces techniques permettraient de détecter des schémas complexes et d'optimiser les modèles de prédiction.

De plus, l'analyse des données en temps réel, rendue possible par les véhicules connectés, offrirait des insights précieux pour ajuster les primes de manière dynamique en fonction du comportement de conduite actuel des assurés.

Il est également crucial d'évaluer l'impact des nouveaux modèles de tarification sur la satisfaction et la fidélité des clients. Une communication transparente et des explications claires sur la tarification aideraient à maintenir la confiance des assurés.

En conclusion, cette étude offre une base pour l'optimisation des techniques d'estimation des coûts de cotisation dans le secteur des assurances en Algérie. Les perspectives d'avenir visent à intégrer des approches plus innovantes et à continuer d'adapter les méthodes aux besoins évolutifs du marché et des assurés. Cette démarche contribuera à améliorer globalement les pratiques de gestion des risques dans le secteur.

Bibliographie

[1] Hans Gerber. Risk Theory : The Stochastic Basis of Insurance. *Springer*, 1979.

[2] Bouaziz Cheikh. *L'histoire de l'assurance en Algérie.* Assurances et gestion des risques, 2013.

[3] Compagnies d'assurance SAA. *Guide des garanties proposées.* SAA, 2023.

[4] Dupont jean. Principes de la réassurance. *Presses de l'Institut d'Assurance.*, 2020.

[5] Fatah Cheurfa. Analyse de sensibilité et quantification de l'incertitude épistémique dans les modèles de risques classiques. *Pour l'obtention du grade de DOCTEUR EN SCIENCES*, 2021.

[6] Gilbert Thiry. Les risques d'accidents de la route. *Economie et Statistique*, 2013.

[7] Jacques Janssen et Raimondo Manca. Semi-Markov risk models for finance, insurance and reliability. *Springer*, 2017.

[8] Jean Dupont et Paul Martin . Introduction aux méthodes statistiques. *Éditions XYZ*, 2019.

[9] Klugman, S.A., Panjer, H.H., and Willmot, G.E. *Loss Models : From Data to Decisions.* 4th Edition. Wiley., 2012.

[10] Lefebvre, M. Gestion des sinistres en coassurance. *Éditions Financières*, 2019.

[11] M. Djallil and K. Kiroumani. Clasifications inductives et transductives pour la tarification en RC Automobile. *Mémoire d'Ingénieur d'Etat en Recherche Opérationnelle. Université de Béjaia*, 2009.

[12] Ministère des finances. Droit des assurances. *Alger*, 1995.

[13] Norman Richard Draper Smith. Applied Regression Analysis. *Wiley.*, 1998.

[14] Paul Durand et Jean Lefèvre. Méthodes statistiques avancées. *Éditions Mathématiques*, 2020.

[15] Paul Embrechts, Claudia Klüppelberg, Thomas Mikosch. *Théorie du Risque et de la Ruine.* Springer, 1997.

[16] Philippe Trainar et Patrick Thourot. *Techniques d'assurances.* Dunod, 2018.

[17] Sanford Weisberg. Applied linear regression. *Wiley Series in Probability and Statistics.*, 2005.

[18] Société Algerienne d'Assurance. Chiffre saa - plan budget. *Société Nationale d'assurance.*, 2023.

[19] Société Nationale d'Assurance. Règles de gestion techniques et financières. *Alger*, 2010.

[20] Verill M. Law, David Kelton. Simulation modeling and analysis. *Fifth Edition*, 2014.

[21] William H. Greene. Econometric Analysis. *Pearson*, 2018.

[22] Édouard Théron. Le modèle collecti. *Institut de Science Financière et d'Assurances.Lyon. page (5 7)*, 2004.

Résumé

Résumé

Cette thèse porte sur l'estimation des techniques des coûts de cotisation pour la Société Algérienne d'Assurance (SAA). Il est structuré en trois parties : une présentation générale des assurances et de l'historique de l'assurance en Algérie, une analyse des modèles de risque en assurance avec un focus sur la théorie de la ruine, et une application pratique des techniques d'estimation des coûts de cotisation à travers des études de cas. Visent à développer des méthodes plus précises et adaptées au contexte algérien, en intégrant des approches statistiques et de modélisation innovantes. L'objectif est d'assurer la viabilité financière de la SAA tout en proposant des primes équitables et abordables pour les assurés.

Abstract

This dissertation focuses on the estimation of contribution cost techniques for the Algerian Insurance Company (SAA). It is structured in three parts : a general presentation of insurance and the history of insurance in Algeria, an analysis of insurance risk models with a focus on the theory of ruin, and a practical application of insurance techniques. estimation of contribution costs through case studies. Aim to develop more precise methods adapted to the Algerian context, by integrating innovative statistical and modeling approaches. The objective is to ensure the financial viability of SAA while offering fair and affordable premiums for policyholders.