

وزارة التعليم العالي والبحث العلمي

BADJI MOKHTAR -ANNABA
UNIVERSITY
UNIVERSITE BADJI MOKHTAR
ANNABA



جامعة باجي مختار
- عنابة -

Faculté des Sciences

Année : 2020/2021

Département de Mathématiques



THÈSE

Présentée en vue de l'obtention du diplôme de
Doctorat

Prime de crédibilité en assurance non vie

Spécialité
Modélisation Mathématique

Par
HADDARI Allaeddine

DIRECTEUR DE THÈSE : REMITA Mohamed Riad Prof. U.B.M. ANNABA

Devant le jury

PRESIDENT: BOUTABIA Hacene Prof. U.B.M. ANNABA

EXAMINATEUR: BENCHETTAH Azzedine Prof. U.B.M. ANNABA

EXAMINATEUR : EZZEBSA Abdelali M.C.A U. 8 mai 1945 Guelma

Remerciements

Premièrement, je suis extrêmement reconnaissant à Allah qui m'a donné la chance et le courage d'achever ce travail.

Je voudrais exprimer ma reconnaissance et ma gratitude à mon encadreur prof. REMITA Mohamed Riad qui a bien voulu accepter de m'accorder ce privilège et d'avoir consacré beaucoup de temps à me former.

Je salue en lui ses grandes compétences, sa qualité professionnelle et surtout sa gentillesse et son soutien dont il m'a gratifiée tout au long de ce travail.

Je remercie Prof. BOUTABIA Hacene qui a accepté d'être le président de jury.

Je remercie aussi prof. BENCHETTAH Azzedine ainsi que Dr. EZZEBSA Abdelali de l'université de Guelma, pour l'honneur d'avoir accepté de faire partie du jury.

Je me permets de remercier mes parents et mes frères pour leur encouragement tout au long des années de mes études.

Je voudrais également remercier mes amis et collègue de l'Université d'Annaba, Dr. Metiri Farouk, Dr. Sadoun Ahmed ainsi que mes amis MELIANI Mohamed, TINE Nacer Eddine et Khalfalah Mohammed El-Arbi.

Enfin je remercie tous ceux qui m'ont aidé de près ou de loin.

Dédicace

Je dédie ce modeste travail à
la lumière de ma vie, à celle qui s'est tellement sacrifiée pour me voir toujours meilleur : ma
très chère mère.

À l'être le plus cher, à celui qui m'a toujours soutenu et guidé par ses conseils et qui m'a
encouragé à poursuivre mes études : mon père.

À mes chers frères : Iskander, Tedj Eddine, Badis et Monder.

Enfin, à tous ceux qui m'ont aidé de près ou de loin.

Résumé

Dans cette thèse, nous nous focalisons sur un outil populaire dans la théorie de la crédibilité qui est l'estimation de la prime. Pour calculer la prime de crédibilité dans la théorie classique, il faut donner certaines hypothèses sur la distribution des sinistres $X_i (i = 1, 2, \dots, n)$. Pour cette raison, nous essaierons dans cette thèse de déterminer des primes de crédibilité sans avoir besoin de fixer la fonction de distribution des sinistres $f(X | \theta)$.

Pour traiter ce cas, nous appliquerons la méthode de l'entropie maximale sous la fonction de perte équilibrée afin d'obtenir une nouvelle prime de crédibilité. En outre, une simulation numérique et une application aux données réelles sont présentées pour comparer la prime de crédibilité obtenue avec celle de **Gómez-Déniz (2006)** en utilisant l'erreur quadratique moyenne comme critère d'évaluation.

Mots clés : Prime de crédibilité, Estimation, Fonction de perte quadratique équilibrée, La méthode de l'entropie maximale.

Abstract

In this thesis, we focus on a popular tool in credibility theory which is the estimation of the premium. To calculate the credibility premium in the classical theory, we have to give certain assumptions about the distribution of claims $X_i (i = 1, 2, \dots, n)$. For this reason, we will try in this thesis to determine credibility premiums without the need to fix the claims distribution function $f (X | \theta)$.

To treat this case, we apply the Maximum Entropy Method under the balanced loss function to obtain a new credibility premium. Furthermore, a numerical simulation and an application to real data are presented to compare the credibility premium obtained with that of **Gómez-Déniz (2006)** by using mean square Errors as an evaluation criterion.

**Key words : Credibility premium, Estimation, Balanced square loss function ;
The maximum entropy method.**

ملخص

في هذه الأطروحة ، نركز على أداة شائعة في نظرية المصدقية وهي تقدير قسط التأمين . لحساب قسط المصدقية فالنظرية الكلاسيكية؛ يجب إعطاء فرضيات معينة بشأن توزيع المخاطر (الكوارث) $X_i (i = 1, 2, \dots, n)$. لهذا السبب، سنحاول في هذه الأطروحة تحديد أقساط المصدقية دون الحاجة إلى تحديد دالة توزيع المخاطر $f(X | \theta)$.

للتعامل مع هذه الحالة ، نطبق طريقة أقصى إنتروبيا وفق دالة الخسارة المتوازنة من أجل الحصول على قسط مصدقية جديد. وبالإضافة إلى ذلك ، نقدم محاكاة رقمية وتطبيق على بيانات حقيقية لمقارنة قسط المصدقية الذي تحصلنا عليه مع قسط المصدقية الخاص بغوميز (2006) وهذا باستخدام الخطأ التربيعي المتوسط كمعيار للتقييم.

الكلمات المفتاحية: قسط المصدقية ، التقدير ، دالة الخسارة المتوازنة ، طريقة أقصى إنتروبيا .

Table des matières

Introduction	10
1 Les fonctions de perte et les primes de crédibilité	14
1.1 Fondements de la théorie de la décision	14
1.1.1 Évaluation des estimateurs	14
1.1.2 Utilité et perte	18
1.1.3 Les critères d'optimalité (minimaxité et admissibilité)	23
1.2 Les fonctions de perte en actuariat	29
1.2.1 La fonction de perte quadratique	29
1.2.2 La fonction de perte Linex	30
1.2.3 La fonction de perte d'entropie	31
1.3 L'utilisation de la fonction de perte équilibrée pondérée pour obtenir des primes de crédibilité	32
1.3.1 Motivation	32
1.3.2 La méthodologie	34
1.3.3 L'approche non-paramétrique	37
2 La méthode de l'entropie maximale	42
2.1 Entropie (théorie de l'information)	42
2.2 La méthode de l'entropie maximale	46
2.2.1 Historique de la méthode	47
2.2.2 Aperçu général sur la méthode de l'entropie maximale	47
2.2.3 Informations testables	49
2.2.4 Applications	49
2.2.5 Solution générale pour la distribution de l'entropie maximale avec contraintes linéaires	51
2.2.6 Justification de la méthode de l'entropie maximale	57
3 Primes de crédibilité sous la fonction de perte équilibrée en utilisant la méthode de l'entropie maximale	63
3.1 La méthode de l'entropie maximale dans le domaine actuariel	64
3.2 Les principaux résultats	65
3.3 Exemples numériques	80
3.3.1 Simulation numérique	80
3.3.2 Application à des données réelles :	82

Conclusion générale et perspectives	86
Bibliographie	87

Liste des tableaux

2.1	Tableau des distributions de probabilité et des contraintes d'entropie maximale correspondantes	56
3.1	Résultats de la simulation numérique	82
3.2	Montants des primes pour les six contrats (Les primes sont en millions de drachmes :1 <i>Euro</i> = 340.75 <i>drachmas</i>)	84

Table des figures

3.1	Résultats des primes $P_{\mathbf{B}}^{\mathbf{Gomez}}$ et P_{2002}^{MEM} à partir d'un ensemble de données de l'organisme de sécurité sociale en Grèce	84
-----	--	----

Introduction

L'assurance est un moyen de protection contre les pertes financières, c'est une forme de gestion des risques (sinistres) principalement utilisée pour se couvrir contre le risque d'une perte éventuelle ou incertaine. L'entité qui fournit une assurance est connue sous le nom d'assureur, la compagnie d'assurance ou souscripteur. La personne ou l'entité qui achète une assurance est appelée l'assuré.

L'opération d'assurance implique que l'assureur s'engage à exécuter une prestation (indemnisation) au profit de l'assuré en cas de réalisation d'un sinistre, l'assuré reçoit un contrat appelé aussi la police d'assurance qui détaille les conditions et les circonstances dans lesquelles l'assureur indemniserait l'assuré. Le montant facturé par l'assureur à l'assuré pour la couverture prévue par la police d'assurance est appelé la prime ou la cotisation.

Parmi les employés de la compagnie d'assurance, on trouve l'actuaire. Les principales tâches d'un actuaire sont :

- L'estimation des réserves que la compagnie doit établir pour faire face aux dépenses, selon différents contrats signés avec les assurés.
- Calculer les risques et prévoir tous les aléas d'une situation donnée, autrement dit faire une tarification qui maximise les bénéfices de la compagnie d'assurance et qui ne laisse aucune place à l'imprévu.

Le calcul de la prime est un point fondamental en tarification, il a pour but l'évaluation du montant attendu des sinistres pour chaque assuré pour la période d'assurance étudiée. Pour faire ce calcul l'actuaire utilise des techniques issues principalement de la théorie des probabilités et de la statistique comme : le système bonus-malus, les modèles linéaires généralisés et la technique d'expérience rating appliquée aussi théorie de la crédibilité.

La théorie de la crédibilité est une branche de la science actuarielle qui peut être considérée comme l'un des outils quantitatifs permettant aux assureurs d'effectuer une tarification personnalisée, en d'autres termes, ajuster les primes futures en fonction de l'expérience passée. Cette théorie a vu le jour grâce à **Mowbray (1914)**, suivi par **Whitney (1918)** qui a proposé de pondérer l'expérience individuelle et la prime collective par un facteur de crédibilité $Z(0 \leq Z \leq 1)$ en une seule prime. **Bailey (1950)** démontre qu'en minimisant l'erreur quadratique dans un contexte bayésien, l'estimateur qui en résulte est une fonction linéaire des observations qui correspond exactement à la prime de crédibilité. Après cela, le modèle de Bühlmann (1967) est apparu ; son idée est de forcer la prime bayésienne à être linéaire. Le modèle de Bühlmann est essentiellement basé sur la formule suivante :

$$P_c = Z\bar{X} + (1 - Z)\mu, \quad (0.1)$$

où P_c est la prime de crédibilité, μ est la prime collective, $Z = \frac{n\tau^2}{n\tau^2 + \sigma^2}$ est le facteur de crédibilité ($0 \leq Z \leq 1$), et $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$. avec $\sigma^2 = E[V(X_i | \theta)]$, $\tau^2 = V[E(X_i | \theta)]$ et $\mu(\theta) = E(X_i | \theta)$.

Puis, en introduisant un poids naturel, Bühlmann et Erwin Straub ont développé le célèbre modèle qui porte leurs noms (**Bühlmann et Straub (1970)**). Les années 70 ont vu

l'apparition de nombreux modèles de crédibilité comme **Hachemeister (1975)**, le modèle hiérarchique de **Jewel (1975)** et les modèles de **De Vylder (1976)**.

Dans la pratique, les actuaires ont constaté que les compagnies d'assurance risquent la faillite si elles n'obligent les assurés qu'à payer la prime nette. Pour résoudre ce problème et améliorer la qualité de l'estimation des risques futurs, certains auteurs ont remplacé la forme de la fonction de perte quadratique classique par une fonction ayant une forme équilibrée appelée la fonction de perte équilibrée pondérée :

$$L(\delta, \theta) = \omega h(x)(\delta_0(x) - P)^2 + (1 - \omega)h(x)(\theta - P)^2, \quad (0.2)$$

où $\delta_0(x)$ est une fonction des données observées choisie par l'actuaire, $0 \leq \omega \leq 1$ est un facteur de pondération déterminé par l'actuaire. En utilisant la fonction de perte équilibrée pondérée (**WBLF** : weighted balanced loss function), on peut obtenir des primes avec plus de flexibilité de sorte que différentes compagnies d'assurance peuvent prendre des fonctions de pondération positives particulières $h(x)$ pour répondre à leurs besoins et faire en sorte que la prime soit compétitive sur les marchés. Par exemple pour $h(x) = 1$ et $h(x) = \exp(cx)$, $c > 0$, en effectuant une minimisation, nous obtenons les primes de crédibilité équilibrées nettes et équilibrées d'Esscher respectivement. L'utilisation de la fonction $h(x) = \exp(cx)$, $c > 0$ est restreinte au cas où l'actuaire veut faire payer des primes aux assurés ayant des sinistres importants.

La fonction de perte équilibrée pondérée a été proposée pour la première fois par **Zellner (1994)**. Ensuite, **Gómez-Déniz (2008)** a généralisé l'utilisation de cette fonction de perte dans un contexte actuariel où P représente la prime de crédibilité.

Pour calculer la prime de crédibilité dans la théorie classique, il faut donner certaines hypothèses sur la distribution des sinistres $X_i (i = 1, 2, \dots, n)$. Pour cette raison, nous essaierons dans cette thèse de déterminer des primes de crédibilité sans avoir besoin de fixer la fonction de distribution des sinistres $f(X | \theta)$ et ça à l'aide de la méthode de l'entropie maximale qui nous permet d'obtenir une inférence statistique la moins biaisée possible lorsque des informations partielles et insuffisantes sont disponibles, cette méthode a été proposée pour la première fois par **Jaynes (1957)**.

La méthode de l'entropie maximale repose sur l'hypothèse que lorsqu'on estime la distribution de probabilité. On doit sélectionner une distribution à partir d'une famille donnée de distributions de probabilité, ce qui lui laisse la plus grande incertitude restante (c'est-à-dire l'entropie maximale) en fonction de contraintes données.

Cette thèse s'articule autour de trois chapitres, le premier chapitre jette un coup d'oeil sur les concepts des fonctions de pertes en actuariat, ensuite les différentes fonctions de perte usuelles sont présentées. Dans le deuxième chapitre, on découvre les principales notions de la méthode de l'entropie maximale avec des exemples d'application. Le troisième chapitre est dédié à l'estimation de la prime de crédibilité dans le cas où nous ne savons rien sur la distribution de probabilité des sinistres $X_i (i = 1, 2, \dots, n)$ à l'exception des données d'observation. Pour traiter ce cas, nous utilisons la méthode de l'entropie maximale sous une fonction de perte équilibrée. On termine ce chapitre par une simulation numérique et une application aux données réelles dans lesquelles nous comparons la prime de crédibilité obtenue avec celle de **Gómez-Déniz (2006)**. Les conclusions et les perspectives sont données après le dernier chapitre.

Chapitre 1

Les fonctions de perte et les primes de crédibilité

Dans ce chapitre, nous présentons en premier lieu les principes de base de la théorie de la décision. Ensuite nous présentons quelques fonctions de perte ainsi que leurs estimateurs bayésiens associés. On termine ce chapitre par l'utilisation de la fonction de perte équilibrée pondérée pour obtenir des primes de crédibilité.

1.1 Fondements de la théorie de la décision

1.1.1 Évaluation des estimateurs

Considérant que l'objectif général de la plupart des études inférentielles est de fournir au statisticien (ou à un client) une décision, il semble raisonnable de demander un critère d'évaluation des procédures de décision qui évalue les conséquences de chaque décision et dépend des paramètres du modèle, c'est-à-dire de l'état réel du monde (ou de la Nature). Ces décisions peuvent être de diverses natures allant de l'achat d'actions en fonction de leurs rendements futurs θ , à l'arrêt d'une expérience agricole sur la productivité θ d'une nouvelle espèce de culture, à l'estimation de la contribution de l'économie souterraine θ au **PNB** américain. Il s'agit également d'évaluer si une nouvelle théorie scientifique est compatible avec

les preuves expérimentales dont on dispose. Si aucun critère d'évaluation n'est disponible, il est impossible de comparer les différentes procédures de décision et les solutions absurdes, telles que proposer $\hat{\theta} = 3$ pour tout problème d'estimation réel ou plus spectaculaire encore, la réponse que l'on veut imposer ne peuvent être éliminées que par un raisonnement **ad-hoc**. Éviter un tel raisonnement implique une axiomatisation renforcée du cadre d'inférence statistique appelée **Théorie de la décision**. Cette structure théorique augmentée est nécessaire pour que les statistiques atteignent une cohérence autrement irréalisable.

Bien que presque tout le monde s'accorde sur la nécessité d'un tel critère d'évaluation, le choix de ce critère fait l'objet d'une importante controverse car les conséquences sur la décision ne sont pas anodines. Cette difficulté a même conduit certains statisticiens à rejeter totalement la théorie de la décision, au motif qu'une détermination pratique du critère d'évaluation du décideur est totalement impossible dans la plupart des cas.

Ce critère est généralement appelé perte et est défini comme suit, où D désigne l'ensemble des décisions possibles. D est appelé l'espace de décision et la plupart des exemples théoriques se concentrent sur le cas $D = \Theta$, qui représente le cadre d'estimation standard.

Définition 1.1.1 *Une fonction de perte est toute fonction L de $\Theta \times D$ dans $[0, +\infty)$.*

Cette fonction de perte est censée évaluer la pénalité (ou l'erreur) $L(\theta, d)$ associée à la décision d lorsque le paramètre prend la valeur θ . Dans un cadre traditionnel d'estimation des paramètres, lorsque D est Θ ou $h(\Theta)$, la fonction de perte $L(\theta, \delta)$ mesure l'erreur commise dans l'évaluation de $h(\theta)$ par δ .

La détermination réelle de la fonction de perte est souvent délicate en pratique, notamment parce que la détermination des conséquences de chaque action pour chaque valeur de θ est

généralement impossible lorsque D ou Θ sont de grands ensembles, par exemple lorsqu'ils comportent un nombre infini d'éléments. De plus, dans les modèles qualitatifs, il peut être délicat de quantifier les conséquences de chaque décision.

La complexité de la détermination de la fonction de perte subjective du décideur incite souvent le statisticien à utiliser des pertes classiques, choisies en raison de leur simplicité et de leur traçabilité mathématique. Ces pertes sont également nécessaires pour un traitement théorique de la dérivation des procédures optimales, lorsqu'il n'y a pas de motivation pratique pour le choix d'une fonction de perte particulière. Le terme classique est lié à leur longue histoire qui remonte à **Laplace (1773)** pour la perte d'erreur absolue et à **Gauss (1810)** pour la perte quadratique, lorsque les erreurs en termes de performance des estimateurs ou de conséquences des décisions étaient confondues avec les erreurs en termes de variabilité irréductible des variables aléatoires (variance). Mais cet attribut ne doit pas être considéré comme une déclaration de valeur, car une utilisation extensive de ces pertes ne les légitime pas davantage. En fait, le recours à de telles pertes automatiques (ou génériques), bien que souvent justifié dans la pratique, il est toujours préférable de prendre une décision dans un temps limité en utilisant un critère approximatif plutôt que de passer un temps infini à déterminer exactement la fonction de perte appropriée a généré bon nombre des critiques adressées à la théorie de la décision.

Une base fondamentale de la théorie de la décision bayésienne est que l'inférence statistique doit commencer par la détermination rigoureuse de trois facteurs :

- La famille de loi des observations $f(x | \theta)$;
- La distribution a priori des paramètres $\pi(\theta)$;
- La perte associée aux décisions $L(\delta, \theta)$;

la prior et la perte, et même parfois la distribution des échantillons, étant dérivées de considérations en partie subjectives. Les théoriciens de la décision classique ignorent le deuxième point. Les critiques fréquentistes du concept bayésien ne prennent pas en compte le problème de la construction de la fonction de perte, même si cela peut être au moins aussi compliqué que la dérivation de la distribution a priori. En plus, présupposer l'existence d'une fonction de perte implique que certaines informations sur le problème en question soient disponibles. Ces informations pourraient donc être utilisées plus efficacement en établissant une distribution a priori. En fait, **Lindley (1985)** affirme que la perte et les priors sont difficiles à séparer et doivent être analysés simultanément. Dans certains cas, il est possible de réduire la classe des fonctions de perte acceptables par des considérations d'invariance, par exemple lorsque le modèle est invariant sous l'action d'un groupe de transformations. Il est également intéressant de noter que ces motivations d'invariance sont souvent utilisées dans d'autres approches de la théorie de la décision, où une réduction drastique de la classe des procédures inférentielles est nécessaire pour choisir la meilleure solution.

Exemple 1.1.1 *Considérons le problème de l'estimation de la moyenne θ d'un vecteur normal, $X \sim N_n(\theta, \Sigma)$, où Σ est la matrice diagonale connue avec les éléments diagonaux $\sigma_i^2 (1 \leq i \leq n)$. Dans ce cas, $D = \ominus = R^n$ et δ_i représente une évaluation de θ_i . Si aucune information supplémentaire n'est disponible sur le modèle, il semble logique de choisir la fonction de perte de manière à ce qu'elle pondère de manière égale l'estimation de chaque composante, c'est-à-dire d'utiliser une perte de la forme*

$$\sum_{i=1}^n L \left(\frac{\delta_i - \theta_i}{\sigma_i} \right), \quad (1.1)$$

où L prend son minimum à 0. En effet, pour de telles pertes, les composantes ayant des

variances plus importantes ne biaisent pas fortement la sélection de l'estimateur résultant. En d'autres termes, les composantes ayant une variance plus importante ne sont pas surpondérées lorsque les erreurs d'estimation $(\delta_i - \theta_i)$ sont normalisées par σ_i . Le choix habituel de L est la perte quadratique $L(t) = t^2$, c'est-à-dire que l'erreur d'estimation globale est la somme des erreurs quadratiques de la composante.

1.1.2 Utilité et perte

Revenons à un cadre purement statistique. Du point de vue de la théorie de la décision, le modèle statistique comporte désormais trois espaces : X l'espace d'observation, Θ l'espace de paramètres et D l'espace de décision (ou espace d'action). L'inférence statistique consiste alors à prendre une décision $d \in D$ liée au paramètre $\theta \in \Theta$ basée sur l'observation $x \in X$, x et θ étant liés par la fonction de distribution $f(x|\theta)$. Dans la plupart des cas, la décision sera d'évaluer (ou estimation) une fonction de θ , $h(\theta)$, aussi précisément que possible. La théorie de la décision suppose en outre que chaque action d peut être évaluée (c'est-à-dire que sa précision peut être quantifiée) et conduit à une récompense r avec une utilité $U(r)$ (qui existe sous l'hypothèse de la rationalité des décideurs). Désormais, cette utilité s'écrit $U(\theta, d)$ pour souligner qu'elle ne dépend que de ces deux facteurs. Lorsque d'autres facteurs aléatoires r sont impliqués dans U , nous prenons $U(\theta, d) = E_{\theta, d}[U(r)]$. Par conséquent, $U(\theta, d)$ peut être considéré comme une mesure de la proximité entre l'estimation proposée d et la valeur réelle $h(\theta)$.

Une fois la fonction d'utilité construite (ou approximée), on obtient la fonction de perte correspondante :

$$L(\theta, d) = -U(\theta, d). \quad (1.2)$$

En général, la fonction de perte est censée être non négative, ce qui implique que $U(\theta, d) \leq 0$ et donc qu'il n'y a pas de décision d'une utilité infinie. L'existence d'une limite inférieure pour L peut être critiquée comme étant trop stricte. On peut également affirmer que d'un point de vue statistique, la fonction de perte L représente effectivement la perte (ou l'erreur) due à une mauvaise évaluation de la fonction de θ qui nous intéresse, et donc que même la meilleure évaluation de cette fonction, c'est-à-dire lorsque θ est connue, peut au mieux induire une perte nulle. Dans le cas contraire, il y aurait un manque de continuité de la fonction de perte en $d = \theta$ qui pourrait même empêcher le choix d'une procédure de décision.

De toute évidence, à l'exception des paramètres les plus triviaux, il est généralement impossible de minimiser uniformément (en d) la fonction de perte $L(\theta, d)$ lorsque θ est inconnu. Afin de dériver un critère de comparaison efficace de la fonction de perte, l'approche fréquentiste propose de considérer plutôt la perte moyenne (ou le risque fréquentiste) :

$$R(\theta, \delta) = E_{\theta} [L(\theta, \delta(X))]$$

$$= \int_x L(\theta, \delta(x)) f(x | \theta) dx,$$

où $\delta(x)$ est la règle de décision, c'est-à-dire l'attribution d'une décision à chaque résultat $x \sim f(x|\theta)$ de l'expérience aléatoire. La fonction δ de X en D est généralement appelée estimateur (tandis que la valeur $\delta(x)$ est appelée estimation de θ). Lorsqu'il n'y a pas de risque de confusion, nous désignons également l'ensemble des estimateurs par D .

Le concept de fréquentiste s'appuie sur ce critère pour comparer les estimateurs et si possible pour sélectionner le meilleur estimateur, le raisonnement étant que les estimateurs sont évalués sur leur performance à long terme pour toutes les valeurs possibles du paramètre

θ . Notez cependant que cette approche présente plusieurs difficultés :

- L’erreur (perte) est moyenne sur les différentes valeurs de x proportionnellement à la fonction de densité $f(x|\theta)$. Par conséquent, il semble que l’observation x ne soit plus prise en compte. Le critère de risque évalue les procédures sur leur performance à long terme et non directement pour l’observation donnée x . Une telle évaluation peut être satisfaisante pour le statisticien, mais elle n’est pas si attrayante pour un client, qui veut des résultats optimaux pour ses données x , pas celle d’un autre !.
- L’analyse fréquentiste du problème de décision suppose implicitement que ce problème sera rencontré à plusieurs reprises pour que l’évaluation de la fréquence ait un sens. En effet, $R(\theta, \delta)$ est approximativement la perte moyenne sur les répétitions i.i.d de la même expérience selon la loi des grands nombres. Cependant, pour des raisons à la fois philosophiques et pratiques, la notion même de répétition des expériences fait l’objet de nombreuses controverses (voir **Jeffreys (1961)**). D’une part, si de nouvelles observations arrivent au statisticien, il doit les utiliser, ce qui pourrait modifier la façon dont l’expérience est menée, comme dans les essais médicaux par exemple.
- Pour une procédure δ , le risque $R(\theta, \delta)$ est une fonction du paramètre θ . Par conséquent, l’approche fréquentiste n’induit pas un ordonnancement total sur l’ensemble des procédures. Il est généralement impossible de comparer les procédures de décision avec ce critère, car deux fonctions de risque croisées empêchent la comparaison entre les estimateurs correspondants. Au mieux, on peut espérer une procédure δ_0 qui minimise uniformément $R(\theta, \delta)$, mais de tels cas se produisent rarement, sauf si l’espace des procédures de décision est restreint.

Contrairement à l’approche fréquentiste, l’approche bayésienne de la théorie de la décision

s'intègre sur l'espace Θ puisque θ est inconnu, au lieu de s'intégrer sur l'espace X comme x est connu. Elle s'appuie sur l'espérance de perte a posteriori :

$$\varrho(\pi, d | x) = E^\pi [L(\theta, d) | x] = \int_{\Theta} L(\theta, d) \pi(\theta | x) d\theta, \quad (1.3)$$

où $\pi(\theta | x)$ est la loi a posteriori continue de θ donnée par l'expression :

$$\pi(\theta | x) = \frac{f(x | \theta) \pi(\theta)}{\int_{\Theta} f(x | \theta) \pi(\theta) d\theta},$$

qui prend la moyenne de l'erreur (c'est-à-dire la perte) en fonction de la distribution postérieure du paramètre θ , sous réserve de la valeur observée x . Étant donné x l'erreur moyenne résultant de la décision d est en fait $\varrho(\pi, d|x)$. La perte attendue a posteriori est donc une fonction de x mais cette dépendance n'est pas gênante, contrairement à la dépendance fréquentiste du risque sur le paramètre car x contrairement à θ est connu.

Étant donné une distribution a priori π , il est également possible de définir le risque intégral, qui est le risque fréquentiste moyenné sur les valeurs de θ en fonction de leur distribution a priori :

$$r(\pi, \delta) = E^\pi [R(\theta, \delta)] = \int_{\Theta} \int_x L(\theta, \delta(x)) f(x | \theta) dx \pi(\theta) d\theta. \quad (1.4)$$

Un intérêt particulier de ce deuxième concept est qu'il associe un nombre réel à chaque estimateur, et non une fonction de θ . Il induit donc un ordonnancement total sur l'ensemble des estimateurs, c'est-à-dire permet la comparaison directe des estimateurs. Cela implique que, tout en tenant compte des informations a priori par la distribution a priori, l'approche bayésienne est suffisamment réductrice (dans le sens positif) pour parvenir à une décision efficace. En outre, les deux notions ci-dessus sont équivalentes dans la mesure où elles conduisent à la même décision.

Théorème 1.1.1 *Un estimateur minimisant le risque intégral $r(\pi, \delta)$ peut être obtenu en sélectionnant pour chaque $x \in X$, la valeur $\delta(x)$ qui minimise la perte a posteriori $\varrho(\pi(\theta), \delta | x)$ puisque :*

$$r(\pi, \delta) = \int_x \varrho(\pi, \delta(x) | x) f(x) dx, \quad (1.5)$$

où

$$f(x) = \int_{\ominus} f(x | \theta) \pi(\theta) d\theta.$$

Proof. Le théorème de Fubini nous permet d'interchanger l'ordre des intégrales, nous avons :

$$\begin{aligned} r(\pi, \delta) &= \int_{\ominus} \int_x L(\theta, \delta(x)) f(x | \theta) dx \pi(\theta) d\theta \\ &= \int_x \int_{\ominus} L(\theta, \delta(x)) f(x | \theta) \pi(\theta) d\theta dx \\ &= \int_x \int_{\ominus} L(\theta, \delta(x)) \pi(\theta | x) d\theta f(x) dx. \end{aligned}$$

■

Définition 1.1.2 *Un estimateur bayésien associé à une distribution a priori π et une fonction de perte L est un estimateur $\delta^\pi(x)$ qui minimise $r(\pi, \delta)$. Pour chaque $x \in X$, il est donné par $\delta^\pi(x)$, le minimum d'après le d de $\varrho(\pi, d | x)$. La valeur $r(\pi) = r(\pi, \delta^\pi)$ est alors appelée le risque de Bayes. Le théorème précédent donne alors le moyen de déterminer des estimateurs de Bayes.*

Le théorème ci-dessus fournit donc un outil constructif pour la détermination des estimateurs de Bayes. Remarquez que, d'un point de vue strictement bayésien, seule l'espérance de perte a posteriori $\varrho(\pi, \delta | x)$ est importante, car le concept bayésien est basé sur

l'approche conditionnelle. Faire la moyenne de toutes les valeurs possibles de x , alors que nous connaissons la valeur observée de x , semble être un gaspillage d'informations. Néanmoins, l'équivalence présentée dans le théorème précédent est importante car d'une part, elle montre que l'approche conditionnelle n'est pas nécessairement aussi dangereuse que les fréquentistes peuvent le **dépeindre**. D'autre part, cette équivalence établit un lien entre les résultats classiques de la théorie des jeux et l'approche axiomatique bayésienne, basée sur la distribution postérieure. Elle explique également pourquoi les estimateurs de Bayes jouent un rôle important dans les critères d'optimalité fréquentistes.

Le résultat ci-dessus est valable pour les lois à priori correctes et incorrectes, tant que le risque Bayésien $r(\pi)$ est fini. Dans le cas contraire, la notion d'estimateur de Bayes (théorie de la décision) est affaiblie : nous définissons alors un estimateur de Bayes généralisé comme le minimiseur, pour chaque x de la perte postérieure attendue. En termes d'optimalité fréquentiste, nous verrons que la division entre les priors corrects et incorrects est beaucoup moins importante que la division entre les estimateurs de Bayes réguliers et généralisés, puisque les premiers sont admissibles. A noter que, pour les pertes strictement convexes, les estimateurs de Bayes sont uniques.

1.1.3 Les critères d'optimalité (minimaxité et admissibilité)

Dans cette sous-section on traite les deux notions fondamentales de la théorie de la décision fréquentiste introduite par **Wald (1950)** et **Neyman et Pearson (1933a,b)**. Comme mentionné ci-dessus et contrairement à l'approche bayésienne, le concept fréquentiste n'est pas suffisamment réducteur pour conduire à un seul estimateur optimal. Bien que nous nous intéressions principalement dans ce chapitre aux aspects bayésiens de la théorie de la décision,

il est toujours nécessaire d'étudier ces notions fréquentistes en détail car elles montrent que les estimateurs de Bayes sont souvent optimaux pour les concepts fréquentistes d'optimalité, et doivent donc être pris en compte même lorsque les informations a priori sont ignorées. En d'autres termes, on peut rejeter le concept bayésien et ignorer la signification de la distribution a priori, tout en obtenant de bons estimateurs d'un point de vue fréquentiste lorsqu'on utilise cette distribution a priori. Par conséquent, dans ce sens technique, les fréquentistes devraient également tenir compte de l'approche bayésienne, car elle fournit un outil pour la dérivation d'estimateurs optimaux (voir **Brown (1971, 2000)**, **Strawderman (1974)**, **Berger (1985a)** ou **Berger et Robert (1990)** pour des exemples)..

Estimateurs aléatoires

Comme pour l'étude de la fonction d'utilité, nous devons étendre l'espace de décision à l'ensemble des estimateurs aléatoires, en prenant les valeurs dans D^* , l'espace des distributions de probabilité sur D . Utiliser un estimateur aléatoire δ^* signifie que l'action est générée selon la distribution avec densité de probabilité $\delta^*(x, \cdot)$, une fois que l'observation x a été recueillie. La perte d'un estimateur aléatoire δ^* est alors définie comme la perte moyenne

$$L(\theta, \delta^*(x)) = \int_D L(\theta, a) \delta^*(x, a) da, \quad (1.6)$$

où a est un vecteur de paramètres. Cette extension est nécessaire pour traiter de la minimaxité et de l'admissibilité. Il est évident que de tels estimateurs ne doivent pas être utilisés, ne serait-ce que parce qu'ils contredisent le principe de probabilité, en donnant plusieurs réponses possibles pour la même valeur de x (et donc de $L(\theta|x)$). De plus, il semble assez paradoxal d'ajouter du bruit à un phénomène pour prendre une décision dans l'incertitude !.

Les estimateurs aléatoires sont néanmoins nécessaires d'un point de vue fréquentiste, par

exemple pour la théorie fréquentiste des tests, car ils donnent accès à des niveaux de confiance autrement inatteignables. L'ensemble D^* apparaît donc comme un complément de D . Cependant, cette modification de l'espace de décision ne modifie pas les réponses bayésiennes, comme le montre le résultat suivant (où D^* désigne également l'ensemble des fonctions prenant des valeurs dans D^*).

Théorème 1.1.2 *Pour toute distribution π sur Θ , le risque de Bayes sur l'ensemble des estimateurs aléatoires est le même que le risque de Bayes sur l'ensemble d'estimateurs non-aléatoires :*

$$\inf_{\delta \in D} r(\pi, \delta) = \inf_{\delta^* \in D^*} r(\pi, \delta^*) = r(\pi). \quad (1.7)$$

Proof. Pour tout $x \in X$ et pour tout $\delta^* \in D^*$, nous avons :

$$\begin{aligned} & \int_{\Theta} \int_D L(\theta, a) \delta^*(x, a) da \pi(\theta | x) d\theta \\ &= \int_D \int_{\Theta} L(\theta, a) \pi(\theta | x) d\theta \delta^*(x, a) da \\ &\geq \int_D \inf_a \int_{\Theta} L(\theta, a) \pi(\theta | x) d\theta \delta^*(x, a) da \\ &= \varrho(\pi, \delta^\pi | x) \end{aligned}$$

■

Minimaxité

Le critère minimax que nous introduisons maintenant apparaît comme une assurance contre le pire des cas car il vise à minimiser l'espérance de perte dans le cas le moins favorable. Il représente également un effort fréquentiste pour passer le concept bayésien tout en produisant une commande totale (faible) sur D^* .

En général, il semble malheureux de recourir à une telle perspective antagoniste dans une analyse statistique. En effet, percevoir la Nature (ou la réalité) comme un ennemi implique un biais vers les pires cas et empêche le statisticien d'utiliser les informations disponibles (pour une analyse et une défense de la minimaxité, voir **Brown (1993)** et **Strawderman (2000)**).

La notion de minimaxité illustre bien les aspects conservateurs du concept de fréquentiste. Comme cette approche refuse toute hypothèse sur le paramètre θ , elle doit considérer les pires cas comme également probables, et doit donc se concentrer sur le risque maximal. En fait, d'un point de vue bayésien, il est souvent équivalent de prendre un antécédent concentré sur ces pires cas. Dans la plupart des cas, ce point de vue est donc trop conservateur car certaines valeurs du paramètre sont moins probables que d'autres.

Existence des règles minimax et stratégie maximin

Une difficulté importante liée à la minimaxité est qu'un estimateur de minimax n'existe pas nécessairement. **Ferguson (1967)** et **Berger (1985a, chapitre 5)** donnent des conditions suffisantes. En particulier, il existe une stratégie minimax lorsque Θ est fini et que la fonction de perte est continue. Plus généralement, **Brown (1976)** (voir également **Le Cam (1986)** et **Strasser (1985)**) considère l'espace de décision D comme étant intégré dans un autre espace, de sorte que l'ensemble des fonctions de risque sur D est compact dans cet espace plus large. De ce point de vue et sous des hypothèses supplémentaires, il est alors possible de dériver des estimateurs minimax lorsque la perte est continue. Cependant, ces extensions impliquent des techniques topologiques trop avancées pour être prises en compte dans cette thèse. Par conséquent, nous ne donnons le résultat suivant (voir **Blackwell et Girshick**

(1954) pour une preuve).

Lemme 1.1.1 *Le risque bayésien est toujours plus petit que le risque minimax :*

$$\underline{R} = \sup_{\pi} r(\pi) = \sup_{\pi} \inf_{\delta \in D} r(\pi, \delta) \leq \bar{R} = \inf_{\delta \in D^*} \sup_{\theta} R(\theta, \delta).$$

La première valeur est appelée risque maximin et la distribution π^* telle que $r(\pi^*) = \underline{R}$ est appelée la distribution la moins favorable, quand elle existe. Ces dernières impliquent le plus grand risque de Bayes possible et donc les distributions les moins intéressantes en terme de performance de perte s'ils ne sont pas suggérés par l'information déjà disponible. Ce résultat est logique car l'information disponible ne peut qu'améliorer les erreurs d'estimations même dans le pire des cas.

Admissibilité

Ce deuxième critère fréquentiste induit un ordonnancement partiel sur D^* en comparant les risques fréquentistes des estimateurs, $R(\theta, \delta)$.

Définition 1.1.3 *Un estimateur δ_0 est inadmissible s'il existe un estimateur δ_1 qui domine δ_0 , i. e. pour tout θ ,*

$$R(\theta, \delta_0) \geq R(\theta, \delta_1),$$

et pour au moins une valeur θ_0 du paramètre,

$$R(\theta_0, \delta_0) > R(\theta_0, \delta_1).$$

Sinon, δ_0 est dit admissible.

L'admissibilité est particulièrement intéressante pour son action réductive. En effet, théoriquement, les estimateurs inadmissibles ne devraient pas être considérés car ils peuvent être

uniformément améliorés. L'admissibilité toute seule n'est pas suffisante pour valider l'utilisation d'un estimateur. Par exemple, un estimateur constant $\delta(x) = \theta_0$ est généralement admissible. D'un point de vue fréquentiste, il est donc important de trouver un estimateur qui satisfait les deux critères, minimaxité et admissibilité.

Proposition 1.1.1 *S'il existe un estimateur minimax unique, cet estimateur est admissible.*

Proof. Si δ^* est le seul estimateur minimax, pour tout estimateur $\bar{\delta} \neq \delta^*$,

$$\sup_{\theta} R(\theta, \bar{\delta}) > \sup_{\theta} R(\theta, \delta^*).$$

Donc $\bar{\delta}$ ne peut dominer δ^* . ■

Quand la fonction de perte L est strictement convexe, elle nous permet d'avoir la proposition suivante.

Proposition 1.1.2 *Si δ_0 est admissible avec un risque constant, δ_0 est l'unique estimateur minimax.*

Proof. Pour $\delta_0 \in \Theta$, on a

$$\sup_{\theta} R(\theta, \delta_0) = \sup_{\theta} R(\theta_0, \delta_0).$$

Alors, s'il existe δ_1 tel que $\bar{R} \leq \sup_{\theta} R(\theta, \delta_1) < R(\theta_0, \delta_0)$, δ_0 ne peut être admissible. ■

Nous avons montré précédemment que certains estimateurs minimax sont des estimateurs bayésiens aussi. L'admissibilité est encore plus reliée à l'approche bayésienne dans le sens où dans la plupart des problèmes statistiques, les estimateurs de Bayes « engendrent » la classe des estimateurs admissibles. En effet, ces derniers peuvent être exprimés par des estimateurs de Bayes ou des estimateurs généralisés de Bayes ou par des limites d'estimateurs de Bayes. Nous donnons ici deux propositions importantes.

1.2 Les fonctions de perte en actuariat

En science actuarielle, les fonctions de perte sont utilisées dans le contexte des assurances pour modéliser les prestations versées en sus des primes. Dans cette section nous présentons quelques fonctions de perte et les estimateurs bayésiens associés.

1.2.1 La fonction de perte quadratique

La fonction de perte quadratique a été proposée par **Legendre (1805)** et **Gauss (1810)**, elle est définie par

$$L(\theta, d) = (\theta - d)^2. \quad (1.9)$$

Une variante de cette fonction de perte est une fonction de perte quadratique pondérée qui prend cette forme

$$L(\theta, d) = \omega(\theta)(\theta - d)^2. \quad (1.10)$$

Sous l'hypothèse d'un coût quadratique, l'estimateur de Bayes $\delta^\pi(x)$ de θ associé à la loi a priori π est la moyenne a posteriori de θ

$$\delta^\pi(x) = E^{\pi(\cdot|x)}(\theta) = \int_{\theta \in \Theta} L(\theta, \delta(x)) \pi(\theta) d\theta. \quad (1.11)$$

Proof. Par définition, l'estimateur de Bayes minimise le coût a posteriori i.e $\rho(\pi, \delta) = E^{\pi(\cdot|x)}(L(\theta, \delta(x)))$.

Sous l'hypothèse d'un coût quadratique, on a :

$$\begin{aligned}\rho(\pi, \delta) &= E^{\pi(\cdot|x)}((\theta, \delta(x))^2) \\ &= E^{\pi(\cdot|x)}(\theta^2) - 2\delta(x)E^{\pi(\cdot|x)}(\theta) + \delta^2(x).\end{aligned}$$

Il s'agit d'un polynôme du second degré en $\delta(x)$, il sera minimum en $E^{\pi(\cdot|x)}(\theta)$. ■

1.2.2 La fonction de perte Linex

La fonction de perte Linex (Linear exponential) a été introduite par **Varian (1975)**, c'est une fonction de perte très pratique. Cette fonction de perte est asymétrique, elle croît presque exponentiellement d'un côté de zéro et est approximativement linéaire de l'autre côté. Sous l'hypothèse que la perte minimale est obtenue pour $\widehat{\delta}(x) = \theta$, la fonction de perte linex pour θ s'exprime par

$$L(\Delta) \propto \exp(a(\delta(x) - \theta)) - a(\delta(x) - \theta) - 1, \quad a \neq 0. \quad (1.12)$$

Où $\widehat{\delta}(x)$ est un estimateur du θ . Le signe de a représentant respectivement la direction et le degré de symétrie ($a > 0$: la surestimation est plus grave que la sous-estimation et vice versa). Pour a proche de zéro, la perte Linex est approximativement la fonction de perte quadratique :

$$E_{\theta}(L(\delta(x), \theta)) \propto \exp(a\delta(x))E_{\theta}(\exp(-a\theta)) - a(\delta(x) - E_{\theta}(\theta)) - 1. \quad (1.13)$$

Où $E_{\theta}(\cdot)$ représente l'espérance a posteriori relative à la densité a posteriori de θ .

L'estimateur de Bayes $\delta_{\pi}(x)$ qui minimise (1.13). Pour trouver l'estimateur, nous dérivons l'équation (1.13) par rapport à $\delta(x)$, nous obtenons

$$\frac{d}{d\delta(x)}(E_\theta(L(\delta(x), \theta))) = a \exp(a\delta(x))E_\theta(\exp(-a\theta)) - a. \quad (1.14)$$

En mettant l'équation (1.14) égale à zéro, nous obtenons

$$\exp(-a\delta(x))E_\theta(\exp(-a\theta)) = a. \quad (1.15)$$

Alors, l'estimateur de Bayes $\widehat{\delta}_L(x)$ sous la fonction de perte Linex est :

$$\delta(x) = -\frac{1}{a} \log(E_\theta(\exp(-a\theta))), \quad (1.16)$$

étant donné que $\exp(-a\theta)$ existe et est finie.

1.2.3 La fonction de perte d'entropie

La fonction de perte d'entropie a été proposée par **Galabria et Pulcini (1994)**, cette fonction découle de la fonction de perte Linex et elle est définie par

$$L_E(\theta, d) \propto \left(\frac{d}{\theta}\right)^a - a \ln\left(\frac{d}{\theta}\right) - 1, \quad (1.17)$$

qui a un minimum lorsque $d = \theta$.

L'estimateur de Bayes de paramètre sous cette fonction de perte est

$$\delta(x) = (E_\theta(\theta^{-a}))^{-\frac{1}{a}}. \quad (1.18)$$

- Lorsque $a = 1$, l'estimateur de Bayes coïncide avec l'estimateur de Bayes sous la fonction de perte quadratique pondérée $\frac{(d-\theta)^2}{\theta}$.

- Lorsque $a = -1$, l'estimateur de Bayes coïncide avec l'estimateur de Bayes sous la fonction de perte quadratique.

Il existe d'autres fonctions de perte comme : Fonction de perte absolue, fonction de perte 0-1, fonction de perte de DeGroot, ... etc.

Toutes les fonctions de perte mentionnées ci-dessus ont la forme classique. Dans la section suivante, on va présenter une autre forme pour les fonctions de perte qui s'appelle la fonction de perte pondérée équilibrée. Nous nous focalisons sur la fonction de perte quadratique dans le cas équilibré.

1.3 L'utilisation de la fonction de perte équilibrée pondérée pour obtenir des primes de crédibilité

Dans cette section, une alternative à la prime de crédibilité habituelle qui survient pour une fonction de perte équilibrée pondérée est envisagée. Il s'agit d'une fonction de perte généralisée qui inclut comme cas particulier la fonction de perte quadratique équilibrée traditionnellement utilisée en actuariat. De cette fonction, on peut déduire des primes de crédibilité selon une fonction de probabilité et des antécédents appropriés.

1.3.1 Motivation

La théorie de la crédibilité est un ensemble de méthodes quantitatives qui permet à un assureur d'ajuster les primes futures en fonction de l'expérience passée. Généralement, l'expression de crédibilité obtenue est écrite comme une somme pondérée de la moyenne de l'échantillon et de la prime collective.

Dans **Heilmann (1989)**, de nombreuses primes de crédibilité ont été obtenues selon la

théorie de la décision statistique du point de vue bayésien et en utilisant la fonction de perte d'erreur quadratique appropriée (**WLF** dorénavant : weighted squared-error loss function) $L_1(a, x) = h(x)(x - a)^2$ et des paires de fonctions de probabilités et de distributions a priori (généralement connues dans la pratique actuarielle sous le nom de fonction de structure). En utilisant différentes formes fonctionnelles pour $h(x)$ nous avons différents principes de prime.

Par exemple, pour $h(x) = 1$ et $h(x) = \exp\{cx\}$, $c > 0$ nous avons les principes de la prime nette et de la prime Esscher (**Heilmann (1989)** et **Gómez et al. (2006)**; entre autres), respectivement.

Il est bien connu (voir **Jewell (1974)**) que pour la famille exponentielle de distributions et ses antécédents conjugués on obtient des primes nettes exactes de crédibilité et aussi pour la paire Poisson-gamma sous la prime Esscher (voir **Heilmann (1989)**). Dans ce cas, la prime Bayes peut être écrite comme une formule de crédibilité sous la forme :

$$P_B^{L_1} = Z(t) g(\bar{X}) + [1 - Z(t)] P_C^{L_1}, \quad (1.19)$$

où nous avons désigné par $P_B^{L_1}$ et $P_C^{L_1}$ la prime de Bayes et la prime collective obtenue dans le cadre de WLF (voir **Heilmann (1989)** et **Gómez et al. (2006)**) et $g(\bar{X})$ est une fonction des données observées.

Dans cette section, les généralisations des primes de crédibilité sont dérivées en utilisant la fonction de perte équilibrée pondérée générale introduite ici comme suit :

$$L_2(a, x) = \omega h(x)(\delta_0(x) - a)^2 + (1 - \omega)h(x)(x - a)^2, \quad (1.20)$$

où $0 \leq \omega \leq 1$ est un facteur de pondération déterminé par l'actuaire, $h(x)$ est une fonction

de pondération positive et $\delta_0(x)$ est une fonction des données observées.

WBLF est une fonction de perte généralisée introduite dans **Zellner (1994)**, et qui apparaît aussi dans **Dey et al. (1999)**, **Farsipour and Asgharzadhe (2004)** et **Jafari et al. (2006)** en prenant $h(x) = 1$ dans (1.20). Cette perte inclut comme cas particulier du **WLF** lorsque ω est choisi comme égal à 0. De plus, en utilisant **WBLF** on obtient également une généralisation de l'expression de crédibilité de **Bühlmann (1967)** dans le cadre de l'approche non-paramétrique.

1.3.2 La méthodologie

Dans cette partie, de nouvelles formules de crédibilité selon le principe de la prime nette sont dérivées en utilisant **WBLF** dans (1.20). Pour cette raison, nous supposons que le risque individuel X a une fonction de densité $f(x|\theta)$ indexée par un paramètre $\theta \in \Theta$ qui a une distribution a priori avec la densité $\pi(\theta)$. Soit maintenant $\pi^x(\theta)$ la densité postérieure lorsque x est observé.

Dans la littérature actuarielle, la prime de risque inconnue $P_R^L \equiv P_R^L(\theta)$ est obtenue en minimisant la perte espérée $E_f[L(\theta, P)]$ pour certaines fonctions de perte L . Si l'expérience n'est pas disponible, l'actuaire facture la prime collective P_C^L qui est donnée si en minimisant la fonction de risque, c'est-à-dire minimiser $E_\pi[L(P_R^L(\theta), P_C^L)]$.

Enfin, si l'expérience est disponible, l'actuaire calcule la prime de Bayes qui est calculée de la même manière que la prime collective en intervertissant l'antérieur par la distribution postérieure. La proposition suivante est une généralisation du Lemme 2 dans **Jafari et al. (2006)** à partir duquel l'estimateur de Bayes de θ sous π a priori est obtenu.

Proposition 1.3.1 *Dans le cadre du **WBLF** et des π a priori, les primes de risque et les*

primes collectives sont données par :

$$P_R^{L_2} \equiv P_R^{L_2}(\theta) = \frac{E_{f(x|\theta)}[\delta_0(x)h(x) | \theta]}{E_{f(x|\theta)}[h(x) | \theta]} + (1 - \omega) \frac{E_{f(x|\theta)}[Xh(x) | \theta]}{E_{f(x|\theta)}[h(x) | \theta]}, \quad (1.21)$$

$$P_C^{L_2} = \omega \delta_0^* + (1 - \omega) \frac{E_\pi[P_R^{L_2} h(P_R^{L_2})]}{E_\pi[h(P_R^{L_2})]}, \quad (1.22)$$

respectivement et où δ_0^* est un estimateur ciblé pour la prime de risque $P_R^{L_2}$.

Proof. La preuve est simple : minimiser $E_{f(x|\theta)} [L_2(\theta, P_R^{L_2})]$ et $E_{\pi(\theta)} [L_2(P_R^{L_2}, P_C^{L_2})]$ par rapport à $P_R^{L_2}$ et $P_C^{L_2}$ pour obtenir la prime de risque et la prime collective, respectivement.

■

On obtient maintenant la prime de Bayes $P_B^{L_2}$ en remplaçant dans (1.22) $\pi(\theta)$ par $\pi^x(\theta)$.

Nous observons que, en mettant $\gamma = \frac{E_\pi[P_R^{L_2} h(P_R^{L_2})]}{E_\pi[h(P_R^{L_2})]}$,

$$P_C^{L_2} \in (\delta_0^*, \gamma), \text{ si } \delta_0^* < \gamma,$$

$$P_C^{L_2} \in (\gamma, \delta_0^*), \text{ si } \delta_0^* > \gamma,$$

et le même résultat se produit lorsque C est remplacé par B . Par conséquent, l'actuaire peut choisir la valeur de δ_0^* pour obtenir une prime en fonction de ses préférences.

La proposition suivante fournit la prime nette de crédibilité dans le cadre du **WBLF**, c'est-à-dire que nous supposons que $h(x) = 1$.

Proposition 1.3.2 *Si la prime nette de Bayes obtenue sous $L_1(a, x)$ est une prime de crédibilité (i.e. prend une formule linéaire), la prime nette équilibrée de Bayes obtenue sous la **WBLF** est également une formule de crédibilité de la forme :*

$$P_B^{L_2} = Z(t) l(P_C^{L_1}) + [1 - Z(t)] l(\bar{X}),$$

où $Z(t) \in [0, 1]$ et

$$l(x) = (1 - \omega)^2 x + \omega(1 - \omega) E_{\pi^x(\theta)} [E_{f(x|\theta)}(\delta_0(X|\theta))] + \omega \delta_0^*.$$

Proof. En utilisant (1.21) et (1.22) avec $h(x) = 1$, on obtient :

$$P_R^{L2} = \omega E[\delta_0(X|\theta)] + (1 - \omega) E_{f(x|\theta)}(X|\theta),$$

et

$$\begin{aligned} P_C^{L2} &= \omega \delta_0^* + (1 - \omega) E_{\pi(\theta)} [\omega E_{f(x|\theta)}[\delta_0(X|\theta)] + (1 - \omega) E_{f(x|\theta)}(X|\theta)] \\ &= \omega \delta_0^* + \omega(1 - \omega) E_{\pi(\theta)} \{E_{f(x|\theta)}[\delta_0(X|\theta)]\} + (1 - \omega)^2 E_{\pi(\theta)} [E_{f(x|\theta)}\delta_0(X|\theta)]. \end{aligned}$$

Par conséquent

$$P_B^{L2} = \omega \delta_0^* + \omega(1 - \omega) E_{\pi^x(\theta)} \{E_{f(x|\theta)}[\delta_0(X|\theta)] + (1 - \omega)^2 P_B^{L1}\}.$$

Maintenant, si P_B^{L1} est une formule de crédibilité sous la forme

$$P_B^{L1} = Z(t) P_C^{L1} + [1 - Z(t)] \bar{X},$$

alors

$$\begin{aligned} P_B^{L2} &= \omega \delta_0^* + \omega(1 - \omega) E_{\pi^x(\theta)} \{E_{f(x|\theta)}[\delta_0(X|\theta)] + (1 - \omega)^2 \{Z(t) P_C^{L1} + [1 - Z(t)] \bar{X}\}\} \\ &= Z(t) [(1 - \omega)^2 P_C^{L1} + \omega(1 - \omega) E_{\pi^x(\theta)} \{E_{f(x|\theta)}[\delta_0(X|\theta)]\} + [1 - Z(t)] \\ &\quad [(1 - \omega)^2 \bar{X} + \omega \delta_0^* + \omega(1 - \omega) E_{\pi^x(\theta)} \{E_{f(x|\theta)}[\delta_0(X|\theta)]\}]] \\ &= Z(t) l(P_C^{L1}) + [1 - Z(t)] l(\bar{X}). \end{aligned}$$

■

Maintenant, il est simple de dériver le corollaire suivant qui est une généralisation du résultat dans **Jewell (1974)**.

Corollaire 1.3.1 *Si nous supposons que la famille exponentielle de distributions est la fonction de probabilité et son conjugué antérieur, alors la prime nette équilibrée de Bayes est une formule de crédibilité exacte.*

Proof. Elle est simple en tenant compte du fait que dans la famille exponentielle des distributions comme la fonction de probabilité et son conjugué antérieur, la prime nette de Bayes obtenue en utilisant la perte $L(x, a) = (x - a)^2$ est une expression de crédibilité comme dans (1.19). ■

Remarque 1.3.1 *Nous observons que le facteur de crédibilité de la proposition précédente est le même que le facteur de crédibilité obtenu en utilisant la fonction de perte $L(x, a) = (x - a)^2$.*

Corollaire 1.3.2 *Les primes nettes équilibrées pondérées de Bayes sont une formule de crédibilité lorsqu'on utilise les paires : Poisson–gamma, negative binomial–beta, binomial–beta, normal–normal et gamma–gamma.*

Proof. Il est dérivé pour appliquer le Corollaire précédent aux paires mentionnées ci-dessus. ■

1.3.3 L'approche non-paramétrique

Il est bien connu (**Bühlmann, 1967**) que la prime de crédibilité linéarisée dans le modèle classique est une somme pondérée de la sinistralité attendue dans un contrat et de la sinistralité attendue pour l'ensemble du portefeuille. Ce fait coïncide avec le résultat dans **Jewell (1974)**.

L'idée principale de **Gómez-Déniz** consiste à remplacer la prime de crédibilité exacte

$H(\mu(\theta) | X_1, X_2, \dots, X_t)$ par une expression linéaire de la forme $c_0 + \sum_{s=1}^t c_s X_s$ en fonction des sinistres passés $X_i, i = 1, \dots, t$ en utilisant **WBLF**. Nous supposons que les variables $X_1 | \theta, X_2 | \theta, \dots, X_t | \theta$ sont indépendantes et identiquement distribuées. Pour faciliter le travail, nous introduisons les notations suivantes :

$$\mu(\theta) = \omega E[\delta_0(X) | \theta] + (1 - \omega) E(X_s | \theta),$$

$$m = \omega \delta_0^* + (1 - \omega) \{ \omega E[E(\delta_0(X) | \theta)] + (1 - \omega) E[E(X_s | \theta)] \}, \quad (1.23)$$

$$\tau^2 = Var[E(X_s | \theta)],$$

$$s^2 = E[Var(X_s | \theta)],$$

ce qui est traditionnel dans la littérature (voir **Herzog (1996)** et **Goovaerts et al. (1990)**).

Ensuite, les coefficients $c_i, i = 1, \dots, t$ doivent être déterminés en minimisant

$$\min_{c_i} E[\omega(\delta_0(X) - c_0 - \sum_{s=1}^t c_s X_s)^2 + (1 - \omega)(\mu(\theta) - c_0 - \sum_{s=1}^t c_s X_s)^2].$$

En fixant la dérivée par rapport à c_i égale à zéro, on obtient le système d'équations

$$c_0 = \omega E[\delta_0(X)] + (1 - \omega) E[\mu(\theta)] - \sum_{s=1}^t c_s E(X_s),$$

$$c_0 E(X_r) = \omega E(X_r \delta_0(X)) + (1 - \omega) E[X_r \mu(\theta)] - \sum_{s=1}^t c_s E(X_r X_s), \quad r = 1, \dots, t,$$

qui est équivalent au système

$$\sum_{s=1}^t c_s \text{Cov}(X_r, X_s) = (1 - \omega) \text{Cov}[X_r, \mu(\theta)] + \omega \text{Cov}[X_r, \delta_0(X)], \quad r = 1, \dots, t. \quad (1.24)$$

Maintenant, compte tenu du fait que

$$\begin{aligned} \text{Cov}(X_r, X_s) &= E[\text{Cov}(X_r, X_s | \theta) + \text{Cov}[E(X_r | \theta), E(X_s | \theta)]] \\ &= E(0) + \text{Var}\left[\frac{\mu(\theta) - E(\delta_0(X))}{1 - \omega}\right] \\ &= \text{Var}[\mu^*(\theta)] = (\tau^*)^2, \end{aligned}$$

où

$$\mu^*(\theta) = \frac{\mu(\theta) - E(\delta_0(X))}{1 - \omega},$$

$$\text{Var}(X_r) = E[\text{Var}(X_r | \theta)] + \text{Var}[E(X_r | \theta)] = s^2 + \tau^2,$$

et

$$\begin{aligned} \text{Cov}[X_r, \mu(\theta)] &= E[\text{Cov}(X_r, \mu(\theta) | \theta) + \text{Cov}[E(X_r | \theta), E(\mu(\theta) | \theta)]] \\ &= E(0) + \text{Cov}[E(X_r | \theta), E(\mu(\theta) | \theta)] \\ &= \frac{\text{Var}[\mu(\theta)] - \omega \text{Cov}[\delta_0(X), \mu(\theta)]}{1 - \omega}, \end{aligned}$$

le système en (1.24) est réduit à

$$cs^2 + (\tau^*)^2 ct = Var[\mu(\theta)] - \omega Cov[\delta_0(X), \mu(\theta)] + \omega Cov[X_r, \delta_0(X)],$$

dont nous obtenons

$$c = \frac{Var[\mu(\theta)] + \omega\{\omega Cov[X_r, \delta_0^*] - \omega Cov[\delta_0(X), \mu(\theta)]\}}{s^2 + (\tau^*)^2 t},$$

et par conséquent

$$\begin{aligned} H(\mu(\theta) | X_1, X_2, \dots, X_t) &= c_0 + ct\bar{X} = \omega E[\delta_0(X)] + (1 - \omega)E[\mu(\theta)] - ctE(X_s) + ct\bar{X} \\ &= \omega E[\delta_0(X)] + (1 - \omega)E[\omega E(\delta_0(X) | \theta) + (1 - \omega)E(X | \theta)] \\ &\quad - ctE(X) + ct\bar{X}. \end{aligned} \tag{1.25}$$

Maintenant, en supposant que $\delta_0^* = E[\delta_0(X)]$ de (1.23) l'expression (1.25) après avoir mis

$c^* = \frac{c}{1-\omega}$ peut être réécrite comme suit

$$\begin{aligned} H(\mu(\theta) | X_1, X_2, \dots, X_t) &= m - \left\{ \frac{m - \omega\delta_0^*}{(1 - \omega)^2} - \frac{\omega ct E[E(\delta_0(X | \theta))]}{(1 - \omega)} \right\} + ct\bar{X} \\ &= m - c^*(m - \omega\delta_0^*)t + \omega(1 - \omega)E[E(\delta_0(X | \theta))]tc^* + c^*t(1 - \omega)^2\bar{X} \\ &= (1 - c^*t)m + \{\omega\delta_0^* + \omega(1 - \omega)E[E(\delta_0(X | \theta))] + (1 - \omega)^2\bar{X}\}c^*t. \end{aligned}$$

Enfin, en désignant c^*t par $Z(t)$, nous trouvons que

$$H(\mu(\theta) | X_1, X_2, \dots, X_t) = Z(t)l(\bar{X}) + [1 - Z(t)]l(P_C^{L^2}).$$

Remarque 1.3.2 *On peut constater que si $\omega = 0$ alors $\mu^*(\theta) = \mu(\theta)$, $(\tau^*)^2 = \tau^2$ et par conséquent $Z(t) = \frac{\tau^2}{\tau^2(s^2 + \tau^2 t)}$ qui correspond au facteur de crédibilité traditionnel obtenu en utilisant **WLF**.*

Chapitre 2

La méthode de l'entropie maximale

Ce chapitre se concentre sur la méthode de l'entropie maximale, cette méthode utilise une mesure de l'entropie du degré d'incertitude d'un événement dans la théorie de l'information. Pour cela, nous introduisons d'abord l'entropie puis nous présentons la méthode de l'entropie maximale.

2.1 Entropie (théorie de l'information)

Dans la théorie de l'information, l'entropie d'une variable aléatoire est le niveau moyen d'information, de surprise ou d'incertitude inhérent aux résultats possibles de la variable. Le concept d'entropie de l'information a été introduit par Claude Shannon dans son article de 1948 «A Mathematical Theory of Communication» **Shannon (1948)**, et est parfois appelée l'entropie de Shannon en son honneur. A titre d'exemple, considérons une pièce de monnaie faussée avec une probabilité p de tomber sur pile et une probabilité $1 - p$ de tomber sur face. La surprise maximale est pour $p = 1/2$, lorsqu'il n'y a aucune raison d'attendre un résultat plutôt qu'un autre, et dans ce cas, un tirage au sort a une entropie d'un bit. La surprise minimale est lorsque $p = 0$ ou $p = 1$, lorsque l'événement est connu et que l'entropie est de zéro bit. D'autres valeurs de p donnent des entropies différentes entre zéro et un bit.

Étant donné une variable aléatoire discrète X , avec des résultats possibles x_1, x_2, \dots, x_n , qui se produisent avec une probabilité de $P(x_1), P(x_2), \dots, P(x_n)$, l'entropie de X est formellement définie par :

$$H(X) = - \sum_{i=1}^n P(x_i) \log P(x_i) \quad (2.1)$$

où \sum indique la somme des valeurs possibles de la variable et \log est le logarithme, le choix de la base varie entre les différentes applications. La base 2 donne l'unité de bits (ou shannons), tandis que la base e donne les «unités naturelles» nat et la base 10 donne une unité appelée dits, bans ou hartleys. Une définition équivalente de l'entropie est la valeur espérée de l'auto-information d'une variable (**Pathria (2011)**).

L'entropie a été créée à l'origine par Shannon dans le cadre de sa théorie de la communication selon laquelle un système de communication de données est composé de trois éléments : une source de données, un canal de communication et un récepteur. Dans la théorie de Shannon, le problème fondamental de la communication comme l'exprime Shannon est que le récepteur doit être capable d'identifier quelles données ont été générées par la source en fonction du signal qu'il reçoit par le canal (**Shannon (1948)**). Shannon a étudié différentes façons d'encoder, de compresser et de transmettre des messages à partir d'une source de données, et il a prouvé dans son célèbre théorème de codage des sources que l'entropie représente une limite mathématique absolue sur la façon dont les données de la source peuvent être compressées sans perte sur un canal parfaitement silencieux. Shannon a considérablement renforcé ce résultat pour les canaux bruyants dans son théorème de codage des canaux bruyants.

L'entropie dans la théorie de l'information est directement analogue à l'entropie dans la thermodynamique statistique. L'entropie présente un intérêt pour d'autres domaines des mathématiques tels que la combinatoire. La définition peut être dérivée d'un ensemble d'axiomes établissant que l'entropie doit être une mesure de la surprise du résultat moyen d'une variable. Pour une variable aléatoire continue, l'entropie différentielle est analogue à l'entropie.

Définitions : Nommé d'après le théorème H de Boltzmann, Shannon a défini l'entropie H (lettre majuscule grecque eta) d'une variable aléatoire discrète X avec des valeurs possibles $\{x_1, \dots, x_n\}$ et la fonction de masse de probabilité $P(x)$ par :

$$H(X) = E[I(X)] = E[-\log(P(x))]. \quad (2.2)$$

Ici, E est l'opérateur de valeur espérée et I est le contenu d'information de X (**Han et al (2002)**), (**Borda (2011)**). $I(X)$ est elle-même une variable aléatoire.

L'entropie peut s'écrire explicitement comme suit :

$$H(X) = - \sum_{i=1}^n P(x_i) \log_b P(x_i) \quad (2.3)$$

où b est la base du logarithme utilisé. Les valeurs communes de b sont 2 : le nombre d'Euler e et 10, et les unités d'entropie correspondantes sont les bits pour $b = 2$, nats pour $b = e$ et bans pour $b = 10$ (**Schneider (2007)**).

Dans le cas où $P(x_i) = 0$ pour quelque i , la valeur de la somme correspondante $0 \log_b(0)$ est considérée comme 0, ce qui est cohérent avec la limite :

$$\lim_{p \rightarrow 0^+} p \log(p) = 0.$$

On peut également définir l'entropie conditionnelle de deux événements X et Y en prenant respectivement les valeurs x_i et y_j comme suit :

$$H(X|Y) = - \sum_{i,j} P(x_i, y_j) \log\left(\frac{P(x_i, y_j)}{P(y_j)}\right) \quad (2.4)$$

où $P(x_i, y_j)$ est la probabilité que $X = x_i$ et $Y = y_j$. Cette quantité doit être comprise comme le degré de caractère aléatoire de la variable aléatoire X par rapport à la variable aléatoire Y (**Thomas et Joy (1991)**).

Exemple 2.1.1 *Envisagez de lancer une pièce de monnaie avec des probabilités connues, pas nécessairement équitables de tomber pile ou face ; cela peut être modélisé comme un processus de Bernoulli.*

L'entropie du résultat inconnu du prochain lancer de la pièce est maximisée si la pièce est juste (c'est-à-dire si pile et face ont toutes deux une probabilité égale à $1/2$). C'est la situation d'incertitude maximale car il est très difficile de prévoir le résultat du prochain tirage au sort ; le résultat de chaque tirage au sort fournit une information complète. C'est parce que :

$$\begin{aligned} H(X) &= - \sum_{i=1}^n P(x_i) \log_b P(x_i) \\ &= - \sum_{i=1}^2 \frac{1}{2} \log_2\left(\frac{1}{2}\right) \\ &= - \sum_{i=1}^2 \frac{1}{2} \cdot (-1) \\ &= 1 \end{aligned}$$

Cependant, si nous savons que la pièce n'est pas équilibrée, mais qu'elle se trouve à pile ou

face avec des probabilités p et q où $p \neq q$, alors il y a moins d'incertitude.

A chaque fois qu'il est lancé, un côté a plus de chances de remonter que l'autre. L'incertitude réduite est quantifiée dans une entropie plus faible : en moyenne, chaque tirage au sort fournit moins d'une information complète. Par exemple, si $p = 0.8$, alors

$$\begin{aligned} H(X) &= -p \log_2(p) - q \log_2(q) \\ &= -0.8 \log_2(0.8) - 0.2 \log_2(0.2) \\ &\approx -0.8 \cdot (-0.322) - 0.2 \cdot (-2.322) \\ &= 0.722 < 1 \end{aligned}$$

Une probabilité uniforme donne une incertitude maximale et donc une entropie maximale. L'entropie ne peut donc que diminuer à partir de la valeur associée à une probabilité uniforme. Le cas extrême est celui d'une pièce à deux faces qui n'aboutit jamais à une pile, ou d'une pièce à deux piles qui n'aboutit jamais à une face. Il n'y a donc aucune incertitude. L'entropie est égale à zéro : chaque tirage au sort ne fournit aucune nouvelle information car le résultat de chaque tirage au sort est toujours certain (**Thomas et Joy(1991)**).

L'entropie peut être normalisée en la divisant par la longueur de l'information. Ce rapport est appelé entropie métrique et est une mesure du caractère aléatoire de l'information.

2.2 La méthode de l'entropie maximale

La méthode de l'entropie maximale stipule que la distribution de probabilité qui représente le mieux l'état actuel des connaissances est celle qui présente la plus grande entropie, dans le contexte de données a priori précisément énoncées (comme une proposition qui exprime des

informations testables).

Une autre façon de le dire : Prenez des données a priori précises ou des informations testables sur une fonction de distribution de probabilité. Considérons l'ensemble de toutes les distributions de probabilités d'essai qui coderaient les données à priori. Selon cette méthode, la distribution avec une entropie maximale de l'information est le meilleur choix.

Puisque la distribution avec l'entropie maximale est celle qui fait le moins d'hypothèses sur la véritable distribution des données, la méthode de l'entropie maximale peut être considérée comme une application du Rasoir d'Ockham (le principe de résolution des problèmes selon lequel «les entités ne doivent pas être multipliées sans nécessité»).

2.2.1 Historique de la méthode

La méthode de l'entropie maximale a été exposée pour la première fois par E. T. Jaynes dans deux articles en 1957 (**Jaynes(1957)**) où il a souligné une correspondance naturelle entre la mécanique statistique et la théorie de l'information.

En particulier, Jaynes a présenté une nouvelle justification très générale du fonctionnement de la méthode de mécanique statistique de Gibbsian. Il a fait valoir que l'entropie de la mécanique statistique et l'entropie de l'information de la théorie de l'information sont fondamentalement la même chose. Par conséquent, la mécanique statistique doit être considérée comme une application particulière d'un outil général d'inférence logique et de la théorie de l'information.

2.2.2 Aperçu général sur la méthode de l'entropie maximale

Dans la plupart des cas pratiques, les données a priori où les informations testables sont données par un ensemble de quantités conservées (valeurs moyennes de certaines fonctions de

moment) associées à la distribution de probabilité en question. C'est ainsi que la méthode de l'entropie maximale est le plus souvent utilisée en thermodynamique statistique. Une autre possibilité consiste à prescrire certaines symétries de la distribution de probabilité. L'équivalence entre les quantités conservées et les groupes de symétrie correspondants implique une équivalence similaire pour ces deux façons de spécifier l'information testable dans la méthode de l'entropie maximale.

La méthode de l'entropie maximale est également nécessaire pour garantir l'unicité et la cohérence des assignations de probabilité obtenues par différentes méthodes, la mécanique statistique et l'inférence logique en particulier.

La méthode de l'entropie maximale rend explicite notre liberté d'utiliser différentes formes de données a priori. Dans un cas particulier, une densité de probabilité a priori uniforme (principe d'indifférence de Laplace, parfois appelé principe de raison insuffisante) peut être adoptée. Ainsi, la méthode de l'entropie maximale n'est pas seulement une façon alternative d'envisager les méthodes habituelles d'inférence des statistiques classiques, mais représente une généralisation conceptuelle importante de ces méthodes.

Toutefois, ces déclarations n'impliquent pas que les systèmes thermodynamiques n'ont pas besoin d'être démontrés comme étant ergodiques pour justifier un traitement en tant qu'ensemble statistique.

En langage ordinaire, on peut dire que la méthode d'entropie maximale exprime une revendication de modestie épistémique ou d'ignorance maximale.

La distribution sélectionnée est celle qui prétend le moins être informée au-delà des données a priori énoncées, c'est-à-dire celle qui admet le plus d'ignorance au-delà des données a priori énoncées.

2.2.3 Informations testables

La méthode de l'entropie maximale n'est explicitement utile que lorsqu'elle est appliquée à des informations testables. Une information testable est une donnée sur une distribution de probabilité dont la vérité ou la fausseté est bien définie. Par exemple les données suivantes :

l'espérance de la variable x est 2,87

et

$$p_2 + p_3 > 0,6$$

(où p_2 et p_3 sont des probabilités d'événements) sont des données des informations testables.

Étant donné l'information testable, la procédure d'entropie maximale consiste à rechercher la distribution de probabilité qui maximise l'entropie de l'information sous réserve des contraintes de l'information. Ce problème d'optimisation sous contrainte est généralement résolu en utilisant la méthode des multiplicateurs de Lagrange.

La maximisation de l'entropie sans information testable respecte la "contrainte" universelle selon laquelle la somme des probabilités est égale à un. Sous cette contrainte, la distribution de probabilité discrète de l'entropie maximale est la distribution uniforme

$$p_i = \frac{1}{n} \text{ pour tous } i \in \{1, 2, \dots, n\}.$$

2.2.4 Applications

La méthode de l'entropie maximale est communément appliquée de deux façons aux problèmes d'inférence :

Probabilités a priori

La méthode de l'entropie maximale est souvent utilisée pour obtenir des distributions de probabilité a priori pour l'inférence bayésienne. Jaynes était un des plus grands défenseur de cette approche, affirmant que la distribution à entropie maximale représentait la distribution la moins informative (**Jaynes (1968)**). Une grande partie de la littérature est maintenant consacrée à l'élicitation des priors d'entropie maximale et des liens avec le codage des canaux (**Soofi (2000)**, **Clarke (2006)**, **Bousquet (2008)**, **Palmieri (2013)**).

Probabilités postérieures

L'entropie maximale est une règle de mise à jour suffisante pour un probabilisme radical. La cinématique de probabilité de Richard Jeffrey est un cas particulier d'inférence d'entropie maximale. Toutefois, l'entropie maximale n'est pas une généralisation de toutes ces règles de mise à jour suffisantes **Skyrms(1987)**.

Modèles de l'entropie maximale

Alternativement, la méthode est souvent invoquée pour la spécification du modèle : dans ce cas, les données observées elles-mêmes sont supposées être les informations testables. Ces modèles sont largement utilisés dans le traitement du langage naturel. Un exemple d'un tel modèle est la régression logistique qui correspond au classificateur d'entropie maximale pour les observations indépendantes.

Estimation de la densité de probabilité

L'une des principales applications de la méthode de l'entropie maximale est l'estimation de la densité discrète et continue (**Botev et Kroese (2008,2011)**). Comme pour les estima-

teurs de machines vectorielles de support, la méthode de l'entropie maximale peut nécessiter la solution d'un problème de programmation quadratique et donc fournir un modèle de mélange dispersé comme estimateur de densité optimale. Un avantage important de la méthode est sa capacité à incorporer des informations à priori dans l'estimation de la densité (**Kesavan et Kapur (1990)**).

2.2.5 Solution générale pour la distribution de l'entropie maximale avec contraintes linéaires

Cas discret

Nous disposons d'informations testables sur une quantité X prenant des valeurs dans $\{X_1, X_2, \dots, X_n\}$. Nous supposons que ces informations ont la forme de m contraintes sur les espérances des fonctions f_k , c'est-à-dire que nous avons besoin de notre distribution de probabilité pour satisfaire les contraintes d'inégalité-égalité du moment :

$$\sum_{i=1}^n Pr(x_i) f_k(x_i) \geq F_k \quad k = 1, \dots, m,$$

où les F_k sont observables,

et

$$\sum_{i=1}^n Pr(x_i) = 1. \tag{2.5}$$

La distribution de probabilité avec une entropie maximale de l'information soumise à ces contraintes d'inégalité-égalité est de la forme :

$$Pr(x_i) = \frac{1}{Z(\lambda_1, \dots, \lambda_m)} \exp[\lambda_1 f_1(x_i) + \dots + \lambda_m f_m(x_i)] \quad (2.6)$$

pour certaines $\lambda_1, \dots, \lambda_m$. On l'appelle parfois la distribution de Gibbs. La constante de normalisation est déterminée par :

$$Z(\lambda_1, \dots, \lambda_m) = \sum_{i=1}^n \exp[\lambda_1 f_1(x_i) + \dots + \lambda_m f_m(x_i)], \quad (2.7)$$

et elle est conventionnellement appelée la fonction de partition. (Le théorème de Pitman-Koopman stipule que la condition nécessaire et suffisante pour qu'une distribution d'échantillonnage admette des statistiques suffisantes de dimension limitée est qu'elle ait la forme générale d'une distribution d'entropie maximale).

Les paramètres λ_k sont des multiplicateurs de Lagrange. Dans le cas des contraintes d'égalité, leurs valeurs sont déterminées à partir de la solution des équations non linéaires

$$F_k = \frac{\partial}{\partial \lambda_k} \log Z(\lambda_1, \dots, \lambda_m).$$

Dans le cas des contraintes d'inégalité, les multiplicateurs de Lagrange sont déterminés à partir de la solution d'un programme d'optimisation convexe avec des contraintes linéaires (**Botev et Kroese (2008)**). Dans les deux cas, il n'existe pas de solution sous forme fermée et le calcul des multiplicateurs de Lagrange nécessite généralement des méthodes numériques.

Cas continu

Pour les distributions continues, l'entropie de Shannon ne peut pas être utilisée, car elle n'est définie que pour des espaces de probabilité discrets. **Edwin Jaynes (1963, 1968,**

2003) a plutôt donné la formule suivante qui est fortement liée à l'entropie relative (voir aussi l'entropie différentielle).

$$H_c = - \int p(x) \log \frac{p(x)}{q(x)} dx \quad (2.8)$$

où $q(x)$ que Jaynes a appelé la "mesure invariante" est proportionnelle à la densité limite des points discrets. Pour l'instant, nous supposons que q est connu ; nous en discuterons plus en détail après que les équations de la solution auront été données.

Une quantité étroitement liée, l'entropie relative est généralement défini comme la divergence Kullback-Leibler de p par rapport à q (bien qu'elle soit parfois confusément définie comme la négative de ceci). Le principe de l'inférence qui consiste à minimiser cela en raison de Kullback est connu sous le nom de principe de l'information sur la discrimination minimale.

Nous disposons d'informations testables sur une quantité X qui prend des valeurs dans un certain intervalle des nombres réels (toutes les intégrales ci-dessous sont sur cet intervalle). Nous supposons que ces informations ont la forme de m contraintes sur les espérances des fonctions f_k , c'est-à-dire que nous avons besoin de notre fonction de densité de probabilité pour satisfaire les contraintes de moment d'inégalité (ou purement d'égalité) :

$$\int P(x) f_k(x) dx \geq F_k \quad k = 1, \dots, m,$$

où F_k sont observables et

$$\int P(x) dx = 1. \quad (2.9)$$

La fonction de densité de probabilité avec un maximum de H_c soumis à ces contraintes est :

$$P(x) = \frac{1}{Z(\lambda_1, \dots, \lambda_m)} q(x) \exp[\lambda_1 f_1(x) + \dots + \lambda_m f_m(x)], \quad (2.10)$$

avec la fonction de partition déterminée par

$$Z(\lambda_1, \dots, \lambda_m) = \int q(x) \exp[\lambda_1 f_1(x) + \dots + \lambda_m f_m(x)] dx. \quad (2.11)$$

Comme dans le cas discret, dans le cas où toutes les contraintes de moment sont égales, les valeurs des paramètres λ_k sont déterminées par le système d'équations non linéaires :

$$F_k = \frac{\partial}{\partial \lambda_k} \log Z(\lambda_1, \dots, \lambda_m).$$

Dans le cas des contraintes de moment d'inégalité, les multiplicateurs de Lagrange sont déterminés à partir de la solution d'un programme d'optimisation convexe (**Botev et Kroese (2011)**).

La fonction de mesure invariante $q(x)$ peut être mieux comprise en supposant que x est connu pour ne prendre des valeurs que dans l'intervalle borné (a, b) et qu'aucune autre information n'est donnée.

Alors la fonction de densité de probabilité d'entropie maximale est :

$$P(x) = A \cdot q(x), \quad a < x < b \quad (2.12)$$

où A est une constante de normalisation. La fonction de mesure invariante est en fait la fonction de densité a priori codant le "manque d'informations pertinentes". Elle ne peut

être déterminée par le principe de l'entropie maximale et elle doit être déterminée par une autre méthode logique comme le principe des groupes de transformation ou la théorie de la marginalisation.

Exemples :

Dans le tableau ci-dessous, chaque distribution énumérée maximise l'entropie pour un ensemble particulier de contraintes fonctionnelles listées dans la troisième colonne et la contrainte que x soit inclus dans le support de la densité de probabilité qui est listée dans la quatrième colonne (**Lisman et Zuylen (1972)**, **Jaynes (1988)**). Plusieurs exemples (Bernoulli, géométrique, exponentielle, Laplace, Pareto) listés sont trivialement vrais car les contraintes qui leur sont associées sont équivalentes à l'assignation de leur entropie. Ils sont inclus de toute façon parce que leur contrainte est liée à une quantité commune ou facilement mesurable.

Pour référence :

$$\Gamma(x) = \int_0^{\infty} \exp(-t) t^{x-1} dt$$

est la fonction gamma,

$$\Psi(x) = \frac{d}{dx} \ln \Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$$

est la fonction digamma,

$$B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$$

est la fonction bêta, γ_E est la constante d'Euler-Mascheroni.

TAB. 2.1 – Tableau des distributions de probabilité et des contraintes d'entropie maximale correspondantes

Nom de la distribution	La densité de probabilité la fonction de masse	Contrainte d'entropie maximale	Support
Uniforme (discrete)	$\frac{1}{b-a+1}$	Aucune	$\{a, a+1, \dots, b-1, b\}$
Uniforme (continue)	$\frac{1}{b-a}$	Aucune	$[a, b]$
Bernoulli	$p^k(1-p)^{1-k}$	$E(k) = p$	$\{0, 1\}$
géométrique	$(1-p)^{k-1}p$	$E(k) = \frac{1}{p}$	$\mathbb{N} - \{0\} = \{1, 2, 3\}$
Exponentielle	$\exp(-\lambda x)$	$E(x) = \frac{1}{\lambda}$	$[0, \infty)$
Laplace	$\frac{1}{2b} \exp(-\frac{ x-\mu }{b})$	$E(x-\mu) = b$	$(-\infty, \infty)$
Laplace asymétrique	$\frac{\lambda \exp(-(x-m)\lambda s k^s)}{k+1/k}$	$E((x-m)sk^2) = \frac{1}{\lambda}$	$(-\infty, \infty)$
Pareto	$\frac{\alpha x_m^\alpha}{x^{\alpha+1}}$	$E(\ln(x)) = \frac{1}{\sigma} + \ln(x_m)$	$[x_m, \infty)$
Normale	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$	$E(x) = \mu, E((x-\mu)^2) = \sigma^2$	$(-\infty, \infty)$
von Mises	$\frac{1}{2\pi I_0(k)} \exp(k \cos(\theta - \mu))$	$E(\cos \theta) = \frac{I_1(k)}{I_0(k)} \cos \mu,$ $E(\sin \theta) = \frac{I_1(k)}{I_0(k)} \sin \mu$	$[0, 2\pi)$
Rayleigh	$\frac{x}{\sigma^2} \exp(-\frac{x^2}{2\sigma^2})$	$E(x^2) = 2\sigma^2,$ $E(\ln(x)) = \frac{\ln(2\sigma^2) - \gamma_E}{2}$	$[0, \infty)$
Beta	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$ pour $0 \leq x \leq 1$	$E(\ln(x)) = \Psi(\alpha) - \Psi(\alpha + \beta),$ $E(\ln(1-x)) = \Psi(\beta) - \Psi(\alpha + \beta)$	$[0, 1]$
Cauchy	$\frac{1}{\pi(1+x^2)}$	$E(\ln(1+x^2)) = 2 \ln 2$	$(-\infty, \infty)$
khi	$\frac{1}{2^{k/2}\Gamma(k/2)} x^{k-1} \exp(-\frac{x^2}{2})$	$E(x^2) = k,$ $E(\ln(x)) = \frac{1}{2}[\Psi(\frac{k}{2}) + \ln(2)]$	$[0, \infty)$
khi carré	$\frac{1}{2^{k/2}\Gamma(k/2)} x^{\frac{k}{2}-1} \exp(-\frac{x}{2})$	$E(x) = k,$ $E(\ln(x)) = \Psi(\frac{k}{2}) + \ln(2)$	$[0, \infty)$
Erlang	$\frac{\lambda^k}{(k-1)!} x^{k-1} \exp(-\lambda x)$	$E(x) = \frac{k}{\lambda},$ $E(\ln(x)) = \Psi(k) - \ln(\lambda)$	$[0, \infty)$

Gamma	$f(x) = \frac{x^{k-1} \exp(-\frac{x}{\theta})}{\theta^k \Gamma(k)}$	$E(x) = k\theta,$ $E(\ln(x)) = \Psi(k) + \ln(\theta)$	$[0, \infty)$
log-normale	$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp(-\frac{(\ln(x)-\mu)^2}{2\sigma^2})$	$E(\ln(x)) = \mu,$ $E((\ln(x) - \mu)^2) = \sigma^2$	$[0, \infty)$
distribution de vitesses de Maxwell	$\frac{1}{a^3} \sqrt{\frac{2}{\pi}} x^2 \exp(-\frac{x^2}{2a^2})$	$E(x^2) = 3a^2,$ $E(\ln(x)) = 1 + \ln(\frac{a}{\sqrt{2}})$ $-\frac{\gamma_E}{2}$	$[0, \infty)$
Weibull	$\frac{k}{\lambda^k} x^{k-1} \exp(-\frac{x^k}{\lambda^k})$	$E(x^k) = \lambda^k,$ $E(\ln(x)) = \ln(\lambda) - \frac{\gamma_E}{k}$	$[0, \infty)$
normale multidimensionnelle	$\frac{\exp(-\frac{1}{2}(\vec{x}-\vec{\mu})^\top \Sigma^{-1} \cdot (\vec{x}-\vec{\mu}))}{(2\pi)^{N/2} \Sigma ^{1/2}}$	$E(\vec{x}) = \vec{\mu},$ $E((\vec{x} - \vec{\mu})(\vec{x} - \vec{\mu})^\top) = \Sigma$	\mathbb{R}^n
binomiale	$\binom{n}{k} p^k (1-p)^{n-k}$	$E(x) = \mu,$ $f \in$ n-distribution binomiale généralisée	$\{0, \dots, n\}$
Poisson	$\frac{\lambda^k \exp(-\lambda)}{k!}$	$E(x) = \lambda,$ $f \in$ n-distribution binomiale généralisée	$\mathbb{N} \cup \{0\}$

2.2.6 Justification de la méthode de l'entropie maximale

Les défenseurs la méthode de l'entropie maximale justifient son utilisation dans l'attribution des probabilités de plusieurs façons, notamment par les deux arguments suivants. Ces arguments considèrent l'utilisation de la probabilité bayésienne comme donnée et sont donc soumis aux mêmes postulats.

L'entropie de l'information comme mesure du manque d'information

Considérons une distribution de probabilité discrète parmi m propositions mutuellement exclusives. La distribution la plus informative se fera lorsque l'une des propositions sera connue comme étant vraie. Dans ce cas, l'entropie de l'information serait égale à zéro. La distribution la moins informative aurait lieu lorsqu'il n'y a aucune raison de favoriser l'une des propositions par rapport aux autres. Dans ce cas, la seule distribution de probabilité raisonnable serait uniforme et l'entropie de l'information serait alors égale à sa valeur maximale

possible $\log m$.

L'entropie de l'information peut donc être considérée comme une mesure numérique qui décrit dans quelle mesure une distribution de probabilité particulière est non informative allant de zéro (complètement informative) à $\log m$ (complètement non informative).

En choisissant d'utiliser la distribution ayant l'entropie maximale permise par nos informations, l'argument est que nous choisissons la distribution la moins informative possible. Choisir une distribution avec une entropie plus faible reviendrait à supposer que nous ne possédons pas d'information. Ainsi, la distribution de l'entropie maximale est la seule distribution raisonnable. La dépendance de la solution à la mesure dominante représentée par $m(x)$ est cependant une source de critiques de l'approche puisque cette mesure dominante est en fin arbitraire (**Druilhet et Marin (2007)**).

La dérivation de Wallis

L'argument suivant est le résultat d'une suggestion faite par **Graham Wallis à E. T. Jaynes en 1962 (Jaynes (2003))**. Il s'agit essentiellement du même argument mathématique utilisé pour les statistiques de Maxwell-Boltzmann en mécanique statistique bien que l'accent conceptuel soit tout à fait différent. Elle a l'avantage d'être de nature strictement combinatoire, ne faisant aucune référence à l'entropie de l'information comme mesure d'incertitude, de non-informativité ou de tout autre concept défini de manière imprécise. La fonction d'entropie de l'information n'est pas assumée a priori mais se trouve plutôt dans le cours de l'argument, et l'argument conduit naturellement à la procédure de maximisation de l'entropie de l'information plutôt que de la traiter d'une autre manière.

Supposons qu'un individu souhaite faire une assignation de probabilité parmi m proposi-

tions mutuellement exclusives. Il dispose de certaines informations testables, mais il ne sait pas comment les inclure dans son évaluation des probabilités. Il conçoit donc l'expérience aléatoire suivante. Il distribuera N quanta de probabilité (chacun vaut $1/N$) au hasard parmi les m possibilités. (On pourrait imaginer qu'il lancera N balles dans m seaux les yeux bandés. Afin d'être aussi équitable que possible, chaque lancer doit être indépendant des autres et chaque seau doit avoir la même taille). Une fois l'expérience terminée, il vérifiera si l'assignation de probabilité ainsi obtenue est cohérente avec ses informations. (Pour que cette étape soit réussie, l'information doit être une contrainte donnée par un ensemble ouvert dans l'espace des mesures de probabilité). Si elle est incohérente, il la rejettera et essaiera à nouveau. Si elle est cohérente, son évaluation sera

$$p_i = \frac{n_i}{N}$$

où p_i est la probabilité de la i ème proposition, tandis que n_i est le nombre de quanta qui ont été attribués à la i ème proposition (c'est-à-dire le nombre de balles qui se sont retrouvées dans le seau i).

Maintenant, afin de réduire le 'grain' de l'assignation de probabilité, il sera nécessaire d'utiliser un nombre assez important de quanta de probabilité. Plutôt que de réaliser réellement, et éventuellement de devoir répéter l'expérience aléatoire assez longue, le joueur décide de calculer et d'utiliser simplement le résultat le plus probable. La probabilité d'un résultat particulier est la distribution multinomiale,

$$Pr(p) = W \cdot m^{-N}$$

où

$$W = \frac{N!}{n_1!n_2!\dots n_m!}$$

est parfois connue comme la multiplicité du résultat.

Le résultat le plus probable est celui qui maximise la multiplicité W . Plutôt que de maximiser directement W , le joueur pourrait maximiser de manière équivalente toute fonction monotone croissante de W . Il décide de maximiser

$$\begin{aligned} \frac{1}{N} \log W &= \frac{1}{N} \log \frac{N!}{n_1!n_2!\dots n_m!} \\ &= \frac{1}{N} \log \frac{N!}{(Np_1)!(Np_2)!\dots (Np_m)!} \\ &= \frac{1}{N} (\log N! - \sum_{i=1}^m \log((Np_i)!)) \end{aligned}$$

A ce stade, afin de simplifier l'expression, le joueur prend la limite pour $N \rightarrow \infty$, c.-à-d. que les niveaux de probabilité passent de valeurs discrètes granuleuses à des valeurs continues lisses.

En utilisant l'approximation de Stirling, il trouve

$$\begin{aligned}
\lim_{N \rightarrow \infty} \left(\frac{1}{N} \log W \right) &= \frac{1}{N} (N \log N - \sum_{i=1}^m N p_i \log(N p_i)) \\
&= \log N - \sum_{i=1}^m p_i \log(N p_i) \\
&= \log N - \log N \sum_{i=1}^m p_i - \sum_{i=1}^m p_i \log(p_i) \\
&= (1 - \sum_{i=1}^m p_i) \log N - \sum_{i=1}^m p_i \log(p_i) \\
&= - \sum_{i=1}^m p_i \log(p_i) \\
&= H(p).
\end{aligned}$$

Il ne reste plus au joueur qu'à maximiser l'entropie sous la contrainte de ses informations testables. Il a constaté que la distribution d'entropie maximale est la plus probable de toutes les distributions aléatoires «équitables» dans la limite où les niveaux de probabilité passent du discret au continu.

Compatibilité avec le théorème de Bayes

Giffin et Caticha (2007) affirment que le théorème de Bayes et la méthode de l'entropie maximale sont totalement compatibles et peuvent être considérés comme des cas particuliers de la "méthode de l'entropie relative maximale". Ils affirment que cette méthode reproduit tous les aspects des méthodes d'inférence bayésiennes orthodoxes. En outre, cette nouvelle méthode ouvre la porte à la résolution de problèmes qui ne pourraient être résolus ni par la méthode de l'entropie maximale ni par les méthodes bayésiennes orthodoxes individuellement. De plus, des contributions récentes (**Lazar (2003)** et **Schennach (2005)**) montrent que les approches d'inférence basées sur l'entropie relative fréquentiste (telles que la vraisemblance

empirique et la vraisemblance empirique à inclinaison exponentielle - voir par exemple **Owen (2001)** et **Kitamura (2006)**) peut être combiné avec des informations a priori pour effectuer une analyse postérieure bayésienne.

Jaynes a déclaré que le théorème de Bayes était un moyen de calculer une probabilité, tandis que l'entropie maximale était un moyen d'attribuer une distribution de probabilité a priori (**Jaynes (1988)**).

Il est cependant possible, dans le concept de résoudre une distribution postérieure directement à partir d'une distribution a priori déterminée en utilisant la méthode de l'entropie croisée minimale. (ou la méthode d'entropie maximale étant un cas particulier d'utilisation d'une distribution uniforme comme préalable donné), indépendamment de toute considération bayésienne en traitant formellement le problème comme un problème d'optimisation contraint, la fonction entropie étant la fonction objectif. Dans le cas de valeurs moyennes données comme informations testables (moyenne sur la distribution de probabilité recherchée), la distribution recherchée est formellement la distribution de Gibbs (ou de Boltzmann) dont les paramètres doivent être résolus pour obtenir une entropie croisée minimale et satisfaire les informations testables données.

Chapitre 3

Primes de crédibilité sous la fonction de perte équilibrée en utilisant la méthode de l'entropie maximale

Dans ce chapitre, nous nous concentrons sur l'estimation de la prime de crédibilité dans le cas où nous ne savons rien de la distribution de probabilité des sinistres $X_i (i = 1, 2, \dots, n)$ à l'exception des données d'observation. Pour traiter ce cas, Nous essaierons dans ce chapitre de faire une extension du travail effectué par **HU Yingying et al (2016)** où ils ont utilisés la méthode (**MEM**) pour obtenir la prime de crédibilité, l'originalité de notre travail réside dans le fait qu'on va obtenir une prime de crédibilité sous une fonction de perte équilibrée contrairement aux travaux précédents qui se contentaient des fonctions de perte classiques. L'avantage de la méthode (**MEM**) est que nous pouvons calculer la prime même si nous ne savons rien sur la distribution des sinistres $X_i (i = 1, 2, \dots, n)$. En outre, une simulation numérique et une application aux données réelles sont présentées pour comparer la prime de crédibilité obtenue dans la section 3 de ce chapitre avec celle de **Gómez-Déniz (2006)** en utilisant les erreurs quadratiques moyennes comme critère d'évaluation.

3.1 La méthode de l'entropie maximale dans le domaine actuariel

Dans le chapitre précédent, nous avons présenté la méthode de l'entropie maximale d'une façon générale. dans cette section, nous nous concentrons sur l'utilisation de cette méthode dans le domaine actuariel.

La méthode de l'entropie maximale est un outil permettant d'obtenir l'inférence statistique la moins biaisée lorsqu'elle est partielle et que les informations disponibles sont insuffisantes, cette méthode a été proposée par **Jaynes (1957)**. La **MEM** est plus utilisée en biotechnologie, en thermodynamique, en théorie de l'information, etc. Cependant, dans le domaine de l'actuariat elle n'a pas reçu beaucoup d'attention de la part des actuaires à l'exception de certains travaux. Par exemple, **Landsman et Makov (1999,2000)** ont utilisé la **MEM** pour estimer le facteur de crédibilité et la distribution des paramètres d'échelle dans une famille exponentielle dispersée, **Daroonch (2004,2006)** a utilisé la **MEM** pour dériver une fonction d'utilité par rapport à une assurance non-vie et **Payandeh Najafabadi et al (2011)** ont introduit une nouvelle technique pour approximer les estimateurs de Bayes en utilisant la **MEM**.

La méthode de l'entropie maximale utilise une mesure de l'entropie du degré d'incertitude d'un événement dans la théorie de l'information. **Shannon (1948)** a introduit la plus populaire mesure d'entropie pour un événement avec n résultats distincts qui est proportionnelle à $-\sum_{i=1}^n p_i \ln(p_i)$.

Dans le cas actuariel, pour déterminer la distribution de probabilité nous utilisons la **MEM** sous certaines conditions :

- La probabilité satisfait les conditions axiomatiques : $\sum_{i=1}^n P_i = 1$ pour $P_i \geq 0$ ($i = 1, 2, \dots, n$).
- Les données d'observation peuvent être représentées par leurs moments statistiques (tels que la moyenne, la variance, etc.).

Ainsi, sous certaines contraintes, la distribution de probabilité est déterminée en utilisant le multiplicateur de lagrange pour résoudre le problème des valeurs extrêmes conditionnelles.

Définition 3.1.1 *Supposons qu'une variable aléatoire discrète X prenne des valeurs distinctes X_1, \dots, X_n avec les probabilités correspondantes p_1, \dots, p_n et les contraintes $g_1(\cdot), \dots, g_m(\cdot)$, où $\sum_{i=1}^n g_r(x_i) = a_r$ pour $r = 1, \dots, m$. Alors, le multiplicateur de lagrange sous réserve des contraintes ci-dessus est :*

$$L = - \sum_{i=1}^n p_i \ln(p_i) - \sum_{r=1}^m \lambda_r \left[\sum_{i=1}^n p_i g_r(x_i) - a_r \right]. \quad (3.1)$$

3.2 Les principaux résultats

Dans cette section, nous présentons la principale contribution de notre travail dans le théorème ci-dessous qui montre comment obtenir une nouvelle prime de crédibilité dans le cadre de la fonction de perte équilibrée en appliquant la méthode de l'entropie maximale et en résolvant un problème de minimisation.

Théorème 3.2.1 *Supposons que les sinistres d'un contrat au cours des n dernières années sont X_1, X_2, \dots, X_n alors un estimateur linéaire non homogène optimal du prochain sinistre X_{n+1} est donné par*

$$P_{n+1}^{MEM} = \alpha_0^* + \sum_{k=1}^n \alpha_k^* X_k \quad (3.2)$$

qui satisfait aux conditions suivantes :

(i) minimiser la fonction de perte carrée équilibrée espérée de

$$E \left[\omega \left(\delta_0(X) - \alpha_0 - \sum_{i=1}^n \alpha_i X_i \right)^2 + (1 - \omega) \left(X_{n+1} - \alpha_0 - \sum_{i=1}^n \alpha_i X_i \right)^2 \right],$$

(ii) maximiser l'entropie

avec

$$\alpha_0^* = \omega E[\delta_0(X)] + \frac{(1 - \omega) E(X_{n+1})}{\left\{ 1 + \sum_{i=1}^n \exp \left[\lambda \sum_{j=1}^n \text{Cov}(X_i, X_j) \frac{(1 - \omega) E(X_{n+1})}{E(X_i)} \right] \right\}},$$

$$\alpha_k^* = \frac{(\alpha_0^* - \omega E[\delta_0(X)]) \exp \left[\lambda \sum_{j=1}^n \text{Cov}(X_k, X_j) \frac{(1 - \omega) E(X_{n+1})}{E(X_k)} \right]}{E(X_k)},$$

où λ est la solution de l'équation suivante,

$$\sum_{i=1}^n \exp \left[\lambda \sum_{j=1}^n \text{Cov}(X_i, X_j) \frac{(1 - \omega) E(X_{n+1})}{E(X_i)} \right] \sum_{j=1}^n \text{Cov}(X_i, X_j) \frac{(1 - \omega) E(X_{n+1})}{E(X_i)}$$

$$= \left[1 + \sum_{i=1}^n \exp \left[\lambda \sum_{j=1}^n \text{Cov}(X_i, X_j) \frac{(1 - \omega) E(X_{n+1})}{E(X_i)} \right] \right] \omega \sum_{j=1}^n \text{Cov}(\delta_0(X), X_j).$$

Proof. La première étape consiste à résoudre le problème de la minimisation, c'est-à-dire,

$$\min E \left[\omega \left(\delta_0(X) - \alpha_0 - \sum_{i=1}^n \alpha_i X_i \right)^2 + (1 - \omega) \left(X_{n+1} - \alpha_0 - \sum_{i=1}^n \alpha_i X_i \right)^2 \right].$$

Pour simplifier, on met

$$H = E \left[\omega \left(\delta_0(X) - \alpha_0 - \sum_{i=1}^n \alpha_i X_i \right)^2 + (1 - \omega) \left(X_{n+1} - \alpha_0 - \sum_{i=1}^n \alpha_i X_i \right)^2 \right]. \quad (3.3)$$

On dérive l'équation (3.3) par rapport à α_0 et on met la formule résultante égale à 0,

$$\frac{\partial H}{\partial \alpha_0} = \frac{\partial}{\partial \alpha_0} \left(E \left[\omega \left(\delta_0(X) - \alpha_0 - \sum_{i=1}^n \alpha_i X_i \right)^2 + (1 - \omega) \left(X_{n+1} - \alpha_0 - \sum_{i=1}^n \alpha_i X_i \right)^2 \right] \right) = 0$$

$$\begin{aligned}
&\Rightarrow E \left[-2 \omega \left(\delta_0(X) - \alpha_0 - \sum_{i=1}^n \alpha_i X_i \right) - 2(1-\omega) \left(X_{n+1} - \alpha_0 - \sum_{i=1}^n \alpha_i X_i \right) \right] = 0 \\
&\Rightarrow -2E \left[\omega \left(\delta_0(X) - \alpha_0 - \sum_{i=1}^n \alpha_i X_i \right) + (1-\omega) \left(X_{n+1} - \alpha_0 - \sum_{i=1}^n \alpha_i X_i \right) \right] = 0 \\
&\Rightarrow E \left[\omega \left(\delta_0(X) - \alpha_0 - \sum_{i=1}^n \alpha_i X_i \right) + (1-\omega) \left(X_{n+1} - \alpha_0 - \sum_{i=1}^n \alpha_i X_i \right) \right] = 0 \\
&\Rightarrow \left[\omega \left(E \left(\delta_0(X) - \alpha_0 - \sum_{i=1}^n \alpha_i X_i \right) \right) + (1-\omega) \left(E \left(X_{n+1} - \alpha_0 - \sum_{i=1}^n \alpha_i X_i \right) \right) \right] = 0 \\
&\Rightarrow \omega E(\delta_0(X)) - \omega E(\alpha_0) - \omega E\left(\sum_{i=1}^n \alpha_i X_i\right) + (1-\omega) E(X_{n+1}) - (1-\omega) E(\alpha_0) - (1-\omega) E\left(\sum_{i=1}^n \alpha_i X_i\right) = 0 \\
&\Rightarrow \omega E(\delta_0(X)) - \omega \alpha_0 - \omega \sum_{i=1}^n \alpha_i E(X_i) + (1-\omega) E(X_{n+1}) - \alpha_0 + \omega \alpha_0 - \sum_{i=1}^n \alpha_i E(X_i) + \omega \sum_{i=1}^n \alpha_i E(X_i) = 0 \\
&\Rightarrow \omega E(\delta_0(X)) + (1-\omega) E(X_{n+1}) - \alpha_0 - \sum_{i=1}^n \alpha_i E(X_i) = 0,
\end{aligned}$$

on trouve

$$\alpha_0 = \omega E[\delta_0(X)] + (1-\omega) E(X_{n+1}) - \sum_{i=1}^n \alpha_i E(X_i),$$

qui peut être écrite comme suit,

$$\begin{aligned}
f_1(\underline{\alpha}) &\triangleq \frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega) E(X_{n+1})} + \frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1-\omega) E(X_{n+1})} - \frac{(1-\omega) E(X_{n+1})}{(1-\omega) E(X_{n+1})} = 0, \\
f_1(\underline{\alpha}) &\triangleq \frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega) E(X_{n+1})} + \frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1-\omega) E(X_{n+1})} - 1 = 0,
\end{aligned}$$

de la forme

$$\sum_{i=1}^n P_i = 1,$$

comme $n = 1, 2$ on considère

$$\begin{aligned}
P_1 &= \frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega) E(X_{n+1})}, \\
P_2 &= \frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1-\omega) E(X_{n+1})},
\end{aligned}$$

et $\underline{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_n)$.

On remplace α_0 dans l'équation (3.3), puis on la dérive par rapport à α_i et on met la formule résultante égale à 0,

$$H = E \left[\omega \left(\delta_0(X) - \alpha_0 - \sum_{i=1}^n \alpha_i X_i \right)^2 + (1 - \omega) \left(X_{n+1} - \alpha_0 - \sum_{i=1}^n \alpha_i X_i \right)^2 \right],$$

et

$$\alpha_0 = \omega E[\delta_0(X)] + (1 - \omega) E(X_{n+1}) - \sum_{i=1}^n \alpha_i E(X_i).$$

Alors

$$\begin{aligned} H &= E \left[\begin{aligned} &\omega (\delta_0(X) - (\omega E[\delta_0(X)] + (1 - \omega) E(X_{n+1}) - \sum_{i=1}^n \alpha_i E(X_i)) - \sum_{i=1}^n \alpha_i X_i)^2 \\ &+ (1 - \omega) (X_{n+1} - (\omega E[\delta_0(X)] + (1 - \omega) E(X_{n+1}) - \sum_{i=1}^n \alpha_i E(X_i)) - \sum_{i=1}^n \alpha_i X_i)^2 \end{aligned} \right] \\ &= E \left[\begin{aligned} &\omega (\delta_0(X) - \omega E[\delta_0(X)] - (1 - \omega) E(X_{n+1}) + \sum_{i=1}^n \alpha_i E(X_i) - \sum_{i=1}^n \alpha_i X_i)^2 \\ &+ (1 - \omega) (-\omega E[\delta_0(X)] + \omega E(X_{n+1}) + \sum_{i=1}^n \alpha_i E(X_i) - \sum_{i=1}^n \alpha_i X_i)^2 \end{aligned} \right], \\ \frac{\partial H}{\partial \alpha_0} &= \frac{\partial}{\partial \alpha_0} E \left[\begin{aligned} &\omega (\delta_0(X) - \omega E[\delta_0(X)] - (1 - \omega) E(X_{n+1}) + \sum_{i=1}^n \alpha_i E(X_i) - \sum_{i=1}^n \alpha_i X_i)^2 \\ &+ (1 - \omega) (-\omega E[\delta_0(X)] + \omega E(X_{n+1}) + \sum_{i=1}^n \alpha_i E(X_i) - \sum_{i=1}^n \alpha_i X_i)^2 \end{aligned} \right] = 0 \\ \Rightarrow E \left[\begin{aligned} &2\omega (E(X_i) - X_i) (\delta_0(X) - \omega E[\delta_0(X)] - (1 - \omega) E(X_{n+1}) + \sum_{i=1}^n \alpha_i E(X_i) - \sum_{i=1}^n \alpha_i X_i) \\ &+ 2(1 - \omega) (E(X_i) - X_i) (-\omega E[\delta_0(X)] + \omega E(X_{n+1}) + \sum_{i=1}^n \alpha_i E(X_i) - \sum_{i=1}^n \alpha_i X_i) \end{aligned} \right] = 0 \\ \Rightarrow E \left[\begin{aligned} &-2\omega (X_i - E(X_i)) (\delta_0(X) - \omega E[\delta_0(X)] - (1 - \omega) E(X_{n+1}) + \sum_{i=1}^n \alpha_i E(X_i) - \sum_{i=1}^n \alpha_i X_i) \\ &- 2(1 - \omega) (X_i - E(X_i)) (-\omega E[\delta_0(X)] + \omega E(X_{n+1}) + \sum_{i=1}^n \alpha_i E(X_i) - \sum_{i=1}^n \alpha_i X_i) \end{aligned} \right] = 0 \\ \Rightarrow -2E \left[\begin{aligned} &\omega (X_j - E(X_j)) (\delta_0(X) - \omega E[\delta_0(X)] - (1 - \omega) E(X_{n+1}) + \sum_{i=1}^n \alpha_i E(X_i) - \sum_{i=1}^n \alpha_i X_i) \\ &+ (1 - \omega) (X_j - E(X_j)) (-\omega E[\delta_0(X)] + \omega E(X_{n+1}) + \sum_{i=1}^n \alpha_i E(X_i) - \sum_{i=1}^n \alpha_i X_i) \end{aligned} \right] = 0 \\ \Rightarrow E \left[\begin{aligned} &\omega (X_j - E(X_j)) \left(\begin{aligned} &\delta_0(X) - E[\delta_0(X)] + (1 - \omega) E[\delta_0(X)] - (1 - \omega) E(X_{n+1}) \\ &- \sum_{i=1}^n \alpha_i (X_i - E(X_i)) \end{aligned} \right) \\ &+ (1 - \omega) (X_j - E(X_j)) (-\omega E[\delta_0(X)] + \omega E(X_{n+1}) - \sum_{i=1}^n \alpha_i (X_i - E(X_i))) \end{aligned} \right] = 0 \\ \Rightarrow &\omega E[(X_i - E(X_i))(\delta_0(X) - E[\delta_0(X)])] + \omega(1 - \omega) E[(X_i - E(X_i))E[\delta_0(X)]] \\ &- \omega(1 - \omega) E[(X_i - E(X_i))E(X_{n+1})] - \omega \sum_{i=1}^n \alpha_i E[(X_j - E(X_j))(X_i - E(X_i))] \\ &+ \omega(1 - \omega) E[(X_i - E(X_i))E(X_{n+1})] - \omega(1 - \omega) E[(X_i - E(X_i))E[\delta_0(X)]] \\ &- (1 - \omega) \sum_{i=1}^n \alpha_i E[(X_j - E(X_j))(X_i - E(X_i))] = 0 \end{aligned}$$

$$\begin{aligned} &\Rightarrow \omega \operatorname{cov}(\delta_0(X), X_j) - \omega \sum_{i=1}^n \alpha_i \operatorname{cov}(X_i, X_j) - (1-\omega) \sum_{i=1}^n \alpha_i \operatorname{cov}(X_i, X_j) = 0 \\ &\Rightarrow \omega \operatorname{cov}(\delta_0(X), X_j) - \omega \sum_{i=1}^n \alpha_i \operatorname{cov}(X_i, X_j) - \sum_{i=1}^n \alpha_i \operatorname{cov}(X_i, X_j) + \omega \sum_{i=1}^n \alpha_i \operatorname{cov}(X_i, X_j) = 0. \end{aligned}$$

Alors

$$\omega \operatorname{cov}(\delta_0(X), X_j) - \sum_{i=1}^n \alpha_i \operatorname{cov}(X_i, X_j) = 0,$$

qui est équivalente à

$$f_2(\underline{\alpha}) \triangleq \omega \sum_{j=1}^n \operatorname{Cov}(\delta_0(X), X_j) - \sum_{j=1}^n \sum_{i=1}^n \alpha_i \operatorname{Cov}(X_i, X_j) = 0.$$

La formule du multiplicateur de lagrange correspondante peut être exprimée de la manière suivante,

$$\begin{aligned} L &= - \sum_{i=1}^n p_i \ln(p_i) - \lambda_0 f_1(\underline{\alpha}) - \lambda f_2(\underline{\alpha}) \\ &= - \left(\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega)E(X_{n+1})} \right) \ln \left(\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega)E(X_{n+1})} \right) - \left(\frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1-\omega)E(X_{n+1})} \right) \ln \left(\frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1-\omega)E(X_{n+1})} \right) \\ &\quad - \lambda_0 \left(\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega)E(X_{n+1})} + \frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1-\omega)E(X_{n+1})} - 1 \right) \\ &\quad - \lambda \left(\omega \sum_{j=1}^n \operatorname{Cov}(\delta_0(X), X_j) - \sum_{j=1}^n \sum_{i=1}^n \alpha_i \operatorname{Cov}(X_i, X_j) \right). \end{aligned} \quad (3.4)$$

On dérive l'équation (3.4) par rapport à α_0 ,

$$\begin{aligned} \frac{\partial L}{\partial \alpha_0} &= \frac{\partial}{\partial \alpha_0} \left[- \left(\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega)E(X_{n+1})} \right) \ln \left(\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega)E(X_{n+1})} \right) - \left(\frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1-\omega)E(X_{n+1})} \right) \ln \left(\frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1-\omega)E(X_{n+1})} \right) \right. \\ &\quad \left. - \lambda_0 \left(\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega)E(X_{n+1})} + \frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1-\omega)E(X_{n+1})} - 1 \right) \right. \\ &\quad \left. - \lambda \left(\omega \sum_{j=1}^n \operatorname{Cov}(\delta_0(X), X_j) - \sum_{j=1}^n \sum_{i=1}^n \alpha_i \operatorname{Cov}(X_i, X_j) \right) \right] = 0 \\ &\Rightarrow - \left[\frac{\partial}{\partial \alpha_0} \left(\left(\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega)E(X_{n+1})} \right) \right) \ln \left(\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega)E(X_{n+1})} \right) + \left(\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega)E(X_{n+1})} \right) \right. \\ &\quad \left. \frac{\partial}{\partial \alpha_0} \left(\ln \left(\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega)E(X_{n+1})} \right) \right) \right] \\ &\quad - \lambda_0 \left(\frac{\partial}{\partial \alpha_0} \left(\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega)E(X_{n+1})} \right) \right) = 0 \end{aligned}$$

$$\begin{aligned}
&\Rightarrow - \left[\frac{1}{(1-\omega) E(X_{n+1})} \ln \left(\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega) E(X_{n+1})} \right) + \left(\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega) E(X_{n+1})} \right)^{\frac{1}{\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega) E(X_{n+1})}}} \right] \\
&\quad - \lambda_0 \left(\frac{1}{(1-\omega) E(X_{n+1})} \right) = 0 \\
&\Rightarrow - \frac{1}{(1-\omega) E(X_{n+1})} \left[\ln \left(\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega) E(X_{n+1})} \right) + 1 + \lambda_0 \right] = 0 \\
&\quad \Rightarrow \ln \left(\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega) E(X_{n+1})} \right) + 1 + \lambda_0 = 0 \\
&\quad \Rightarrow \exp \left(\ln \left(\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega) E(X_{n+1})} \right) \right) = \exp(-1 - \lambda_0) \\
&\quad \Rightarrow \left(\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega) E(X_{n+1})} \right) = \exp(-1 - \lambda_0) \\
&\quad \Rightarrow \alpha_0 - \omega E[\delta_0(X)] = (1-\omega) E(X_{n+1}) \exp(-1 - \lambda_0),
\end{aligned}$$

on trouve

$$\alpha_0 = \omega E[\delta_0(X)] + (1-\omega) E(X_{n+1}) \exp(-1 - \lambda_0).$$

On remplace α_0 dans l'équation (3.4)

$$\begin{aligned}
L &= - \left(\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega) E(X_{n+1})} \right) \ln \left(\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega) E(X_{n+1})} \right) - \left(\frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1-\omega) E(X_{n+1})} \right) \ln \left(\frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1-\omega) E(X_{n+1})} \right) \\
&\quad - \lambda_0 \left(\frac{\alpha_0 - \omega E[\delta_0(X)]}{(1-\omega) E(X_{n+1})} + \frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1-\omega) E(X_{n+1})} - 1 \right) \\
&\quad - \lambda \left(\omega \sum_{j=1}^n Cov(\delta_0(X), X_j) - \sum_{j=1}^n \sum_{i=1}^n \alpha_i Cov(X_i, X_j) \right). \\
&= - \left(\frac{\omega E[\delta_0(X)] + (1-\omega) E(X_{n+1}) \exp(-1 - \lambda_0) - \omega E[\delta_0(X)]}{(1-\omega) E(X_{n+1})} \right) \\
&\quad \ln \left(\frac{\omega E[\delta_0(X)] + (1-\omega) E(X_{n+1}) \exp(-1 - \lambda_0) - \omega E[\delta_0(X)]}{(1-\omega) E(X_{n+1})} \right) \\
&\quad - \left(\frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1-\omega) E(X_{n+1})} \right) \ln \left(\frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1-\omega) E(X_{n+1})} \right) \\
&\quad - \lambda_0 \left(\frac{\omega E[\delta_0(X)] + (1-\omega) E(X_{n+1}) \exp(-1 - \lambda_0) - \omega E[\delta_0(X)]}{(1-\omega) E(X_{n+1})} + \frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1-\omega) E(X_{n+1})} - 1 \right) \\
&\quad - \lambda \left(\omega \sum_{j=1}^n Cov(\delta_0(X), X_j) - \sum_{j=1}^n \sum_{i=1}^n \alpha_i Cov(X_i, X_j) \right),
\end{aligned}$$

on obtient

$$L = (1 + \lambda_0) \exp(-1 - \lambda_0) - \left(\frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1 - \omega) E(X_{n+1})} \right) \ln \left(\frac{\alpha_i E(X_i)}{(1 - \omega) E(X_{n+1})} \right) - \lambda_0 \exp(-1 - \lambda_0) \\ - \lambda_0 \left(\frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1 - \omega) E(X_{n+1})} \right) + \lambda_0 - \lambda \left(\omega \sum_{j=1}^n Cov(\delta_0(X), X_j) - \sum_{j=1}^n \sum_{i=1}^n \alpha_i Cov(X_i, X_j) \right). \quad (3.5)$$

Comme ci-dessus, on dérive l'équation (3.5) par rapport à α_k et on met la formule résultante égale à 0,

$$\frac{\partial L}{\partial \alpha_k} = \frac{\partial}{\partial \alpha_k} \left[(1 + \lambda_0) \exp(-1 - \lambda_0) - \left(\frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1 - \omega) E(X_{n+1})} \right) \ln \left(\frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1 - \omega) E(X_{n+1})} \right) - \lambda_0 \exp(-1 - \lambda_0) \right. \\ \left. - \lambda_0 \left(\frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1 - \omega) E(X_{n+1})} \right) + \lambda_0 - \lambda \left(\omega \sum_{j=1}^n Cov(\delta_0(X), X_j) - \sum_{j=1}^n \sum_{i=1}^n \alpha_i Cov(X_i, X_j) \right) \right] \\ - \frac{\partial}{\partial \alpha_k} \left[\left(\frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1 - \omega) E(X_{n+1})} \right) \ln \left(\frac{\alpha_i E(X_i)}{(1 - \omega) E(X_{n+1})} \right) \right] - \frac{\partial}{\partial \alpha_k} \left[\lambda_0 \left(\frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1 - \omega) E(X_{n+1})} \right) \right] \\ - \lambda \frac{\partial}{\partial \alpha_k} \left[- \sum_{j=1}^n \sum_{i=1}^n \alpha_i Cov(X_i, X_j) \right] = 0 \\ \Rightarrow - \left[\left(\frac{E(X_k)}{(1 - \omega) E(X_{n+1})} \right) \ln \left(\frac{\alpha_k E(X_k)}{(1 - \omega) E(X_{n+1})} \right) + \left(\frac{\sum_{i=1}^n \alpha_k E(X_k)}{(1 - \omega) E(X_{n+1})} \right) \frac{\frac{E(X_k)}{(1 - \omega) E(X_{n+1})}}{\sum_{i=1}^n \alpha_k E(X_k)} \right] \\ - \lambda_0 \frac{E(X_k)}{(1 - \omega) E(X_{n+1})} + \lambda Cov(X_k, n\bar{X}) = 0 \\ \Rightarrow - \frac{E(X_k)}{(1 - \omega) E(X_{n+1})} \left[\ln \left(\frac{\alpha_k E(X_k)}{(1 - \omega) E(X_{n+1})} \right) + 1 + \lambda_0 - \left(\frac{(1 - \omega) E(X_{n+1})}{E(X_k)} \lambda Cov(X_k, n\bar{X}) \right) \right] = 0 \\ \Rightarrow \ln \left(\frac{\alpha_k E(X_k)}{(1 - \omega) E(X_{n+1})} \right) = -1 - \lambda_0 + \left(\frac{(1 - \omega) E(X_{n+1})}{E(X_k)} \lambda Cov(X_k, n\bar{X}) \right) \\ \Rightarrow \exp \left[\ln \left(\frac{\alpha_k E(X_k)}{(1 - \omega) E(X_{n+1})} \right) \right] = \exp \left[\lambda Cov(X_k, n\bar{X}) \frac{(1 - \omega) E(X_{n+1})}{E(X_k)} - 1 - \lambda_0 \right] \\ \Rightarrow \left(\frac{\alpha_k E(X_k)}{(1 - \omega) E(X_{n+1})} \right) = \exp \left[\lambda Cov(X_k, n\bar{X}) \frac{(1 - \omega) E(X_{n+1})}{E(X_k)} - 1 - \lambda_0 \right]$$

on obtient

$$\alpha_k = \frac{(1 - \omega) E(X_{n+1})}{E(X_k)} \exp \left(\lambda Cov(X_k, n\bar{X}) \frac{(1 - \omega) E(X_{n+1})}{E(X_k)} - 1 - \lambda_0 \right).$$

On remplace α_k dans l'équation (3.5),

$$\begin{aligned}
L &= (1 + \lambda_0) \exp(-1 - \lambda_0) - \left(\frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1 - \omega) E(X_{n+1})} \right) \ln \left(\frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1 - \omega) E(X_{n+1})} \right) - \lambda_0 \exp(-1 - \lambda_0) \\
&\quad - \lambda_0 \left(\frac{\sum_{i=1}^n \alpha_i E(X_i)}{(1 - \omega) E(X_{n+1})} \right) + \lambda_0 - \lambda \left(\omega \sum_{j=1}^n \text{Cov}(\delta_0(X), X_j) - \sum_{j=1}^n \sum_{i=1}^n \alpha_i \text{Cov}(X_i, X_j) \right) \\
&\quad \alpha_k = \frac{(1 - \omega) E(X_{n+1})}{E(X_i)} \exp \left(\lambda \text{Cov}(X_i, n\bar{X}) \frac{(1 - \omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right). \\
\Rightarrow L &= (1 + \lambda_0) \exp(-1 - \lambda_0) - \left(\frac{\sum_{i=1}^n \frac{(1 - \omega) E(X_{n+1})}{E(X_i)} \exp \left(\lambda \text{Cov}(X_i, n\bar{X}) \frac{(1 - \omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right) E(X_i)}{(1 - \omega) E(X_{n+1})} \right) \\
&\quad \ln \left(\frac{\frac{(1 - \omega) E(X_{n+1})}{E(X_i)} \exp \left(\lambda \text{Cov}(X_i, n\bar{X}) \frac{(1 - \omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right) E(X_i)}{(1 - \omega) E(X_{n+1})} \right) - \lambda_0 \exp(-1 - \lambda_0) \\
&\quad - \lambda_0 \left(\frac{\sum_{i=1}^n \frac{(1 - \omega) E(X_{n+1})}{E(X_i)} \exp \left(\lambda \text{Cov}(X_i, n\bar{X}) \frac{(1 - \omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right) E(X_i)}{(1 - \omega) E(X_{n+1})} \right) + \lambda_0 \\
&\quad - \lambda \left(\begin{aligned} &\omega \sum_{j=1}^n \text{Cov}(\delta_0(X), X_j) \\ &- \sum_{j=1}^n \sum_{i=1}^n \frac{(1 - \omega) E(X_{n+1})}{E(X_i)} \exp \left(\lambda \text{Cov}(X_i, n\bar{X}) \frac{(1 - \omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right) \text{Cov}(X_i, X_j) \end{aligned} \right) \\
\Rightarrow L &= \exp(-1 - \lambda_0) + \lambda_0 \exp(-1 - \lambda_0) - \sum_{i=1}^n \exp \left(\lambda \text{Cov}(X_i, n\bar{X}) \frac{(1 - \omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right) \\
&\quad \left(\lambda \text{Cov}(X_i, n\bar{X}) \frac{(1 - \omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right) - \lambda_0 \exp(-1 - \lambda_0) \\
&\quad - \lambda_0 \left(\sum_{i=1}^n \exp \left(\lambda \text{Cov}(X_i, n\bar{X}) \frac{(1 - \omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right) \right) + \lambda_0 - \lambda \omega \text{Cov}(\delta_0(X), n\bar{X}) \\
&\quad + \lambda \sum_{i=1}^n \frac{(1 - \omega) E(X_{n+1})}{E(X_i)} \exp \left(\lambda \text{Cov}(X_i, n\bar{X}) \frac{(1 - \omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right) \text{Cov}(X_i, n\bar{X}) \\
\Rightarrow L &= \exp(-1 - \lambda_0) - \left[\sum_{i=1}^n \exp \left(\lambda \text{Cov}(X_i, n\bar{X}) \frac{(1 - \omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right) \right. \\
&\quad \left. \left(\lambda \text{Cov}(X_i, n\bar{X}) \frac{(1 - \omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right) \right] \\
&\quad - \lambda_0 \left(\sum_{i=1}^n \exp \left(\lambda \text{Cov}(X_i, n\bar{X}) \frac{(1 - \omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right) \right) + \lambda_0 - \lambda \omega \text{Cov}(\delta_0(X), n\bar{X})
\end{aligned}$$

$$\begin{aligned}
& +\lambda \sum_{i=1}^n \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \exp \left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right) Cov(X_i, n\bar{X}) \\
\Rightarrow L & = \exp(-1 - \lambda_0) - \lambda \sum_{i=1}^n \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \exp \left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right) \\
& \quad Cov(X_i, n\bar{X}) + \sum_{i=1}^n \exp \left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right) \\
& \quad + \lambda_0 \sum_{i=1}^n \exp \left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right) \\
& - \lambda_0 \left(\sum_{i=1}^n \exp \left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right) \right) + \lambda_0 - \lambda \omega Cov(\delta_0(X), n\bar{X}) \\
& + \lambda \sum_{i=1}^n \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \exp \left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right) Cov(X_i, n\bar{X}).
\end{aligned}$$

Alors

$$\begin{aligned}
L & = \exp(-1 - \lambda_0) + \lambda_0 - \lambda \omega Cov(\delta_0(X), n\bar{X}) \\
& \quad + \sum_{i=1}^n \exp \left[\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right]. \tag{3.6}
\end{aligned}$$

En traitant l'équation (3.6) par rapport à λ_0 avec les mêmes opérations que ci-dessus,

$$\begin{aligned}
L & = \exp(-1 - \lambda_0) + \lambda_0 - \lambda \omega Cov(\delta_0(X), n\bar{X}) + \sum_{i=1}^n \exp \left[\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right] \\
L & = \exp(-1 - \lambda_0) + \lambda_0 - \lambda \omega Cov(\delta_0(X), n\bar{X}) + \exp(-1 - \lambda_0) \sum_{i=1}^n \exp \left[\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \right] \\
L & = \exp(-1 - \lambda_0) \left[1 + \sum_{i=1}^n \exp \left[\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \right] \right] + \lambda_0 - \lambda \omega Cov(\delta_0(X), n\bar{X}) \\
\frac{\partial L}{\partial \lambda_0} & = \frac{\partial}{\partial \lambda_0} \left\{ \exp(-1 - \lambda_0) \left[1 + \sum_{i=1}^n \exp \left[\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \right] \right] + \lambda_0 - \lambda \omega Cov(\delta_0(X), n\bar{X}) \right\} \\
& \quad \frac{\partial}{\partial \lambda_0} \left\{ \exp(-1 - \lambda_0) \left[1 + \sum_{i=1}^n \exp \left[\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \right] \right] \right\} + 1 = 0 \\
& \Rightarrow -\exp(-1 - \lambda_0) \left[1 + \sum_{i=1}^n \exp \left[\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \right] \right] + 1 = 0
\end{aligned}$$

$$\begin{aligned}
&\Rightarrow \exp(-1 - \lambda_0) \left[1 + \sum_{i=1}^n \exp \left[\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \right] \right] = 1 \\
&\Rightarrow \exp(-1 - \lambda_0) = \frac{1}{\left[1 + \sum_{i=1}^n \exp \left[\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \right] \right]} \\
&\Rightarrow \ln(\exp(-1 - \lambda_0)) = \ln \left(\frac{1}{\left[1 + \sum_{i=1}^n \exp \left[\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \right] \right]} \right) \\
&\Rightarrow -1 - \lambda_0 = \ln(1) - \ln \left(1 + \sum_{i=1}^n \exp \left[\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \right] \right),
\end{aligned}$$

on trouve

$$\lambda_0 = \ln \left[1 + \sum_{i=1}^n \exp \left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \right) \right] - 1.$$

On remplace λ_0 dans l'équation (3.6)

$$\begin{aligned}
L &= \exp(-1 - \lambda_0) + \lambda_0 - \lambda \omega Cov(\delta_0(X), n\bar{X}) + \sum_{i=1}^n \exp \left[\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} - 1 - \lambda_0 \right] \\
&= \exp(-1 - \lambda_0) \left[1 + \sum_{i=1}^n \exp \left[\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \right] \right] + \lambda_0 - \lambda \omega Cov(\delta_0(X), n\bar{X}) \\
&\quad \lambda_0 = \ln \left[1 + \sum_{i=1}^n \exp \left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \right) \right] - 1 \\
&\Rightarrow L = \exp \left(-1 - \ln \left[1 + \sum_{i=1}^n \exp \left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \right) \right] + 1 \right) \\
&\quad \left[1 + \sum_{i=1}^n \exp \left[\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \right] \right] \\
&\ln \left[1 + \sum_{i=1}^n \exp \left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \right) \right] - 1 - \lambda \omega Cov(\delta_0(X), n\bar{X}) \\
&\Rightarrow L = \exp \left(\ln \left[1 + \sum_{i=1}^n \exp \left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \right) \right]^{-1} \right) \\
&\quad \left[1 + \sum_{i=1}^n \exp \left[\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \right] \right] \\
&\ln \left[1 + \sum_{i=1}^n \exp \left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)} \right) \right] - 1 - \lambda \omega Cov(\delta_0(X), n\bar{X})
\end{aligned}$$

$$\Rightarrow L = \frac{1}{\left[1 + \sum_{i=1}^n \exp\left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)}\right)\right]} \left[1 + \sum_{i=1}^n \exp\left[\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)}\right]\right] \\ + \ln \left[1 + \sum_{i=1}^n \exp\left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)}\right)\right] - 1 - \lambda \omega Cov(\delta_0(X), n\bar{X}),$$

on trouve

$$L = \ln \left[1 + \sum_{i=1}^n \exp\left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)}\right)\right] - \lambda \omega Cov(\delta_0(X), n\bar{X}).$$

On dérive l'équation précédente par rapport à λ et on met la formule résultante égale à 0,

après une série de calculs, une équation à propos de λ peut être présentée comme suit,

$$\frac{\partial L}{\partial \lambda} = \frac{\partial}{\partial \lambda} \left[\ln \left[1 + \sum_{i=1}^n \exp\left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)}\right)\right] - \lambda \omega Cov(\delta_0(X), n\bar{X}) \right] \\ = \frac{\partial}{\partial \lambda} \left[1 + \sum_{i=1}^n \exp\left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)}\right)\right] - \frac{\partial}{\partial \lambda} [\lambda \omega Cov(\delta_0(X), n\bar{X})] = 0 \\ \Rightarrow \frac{\left(Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)}\right) \sum_{i=1}^n \exp\left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)}\right)}{\left[1 + \sum_{i=1}^n \exp\left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)}\right)\right]} - \omega Cov(\delta_0(X), n\bar{X}) = 0 \\ \Rightarrow \frac{\sum_{i=1}^n \exp\left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)}\right) Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)}}{\left[1 + \sum_{i=1}^n \exp\left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)}\right)\right]} = \omega Cov(\delta_0(X), n\bar{X}) \\ \Rightarrow \sum_{i=1}^n \exp\left[\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)}\right] Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)} \\ = \left[1 + \sum_{i=1}^n \exp\left[\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)}\right]\right] \omega Cov(\delta_0(X), n\bar{X}). \quad (3.7)$$

Nous considérons λ comme la solution de l'équation (3.7) et nous déterminons α_0 et α_k

respectivement

$$\alpha_0 = \omega E[\delta_0(X)] + (1-\omega)E(X_{n+1}) \exp(-1 - \lambda_0) \\ \lambda_0 = \ln \left[1 + \sum_{i=1}^n \exp\left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)}\right)\right] - 1,$$

alors

$$\begin{aligned}
\alpha_0^* &= \omega E[\delta_0(X)] + (1-\omega) E(X_{n+1}) \exp\left(-1 - \left(\ln\left[1 + \sum_{i=1}^n \exp\left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)}\right)\right] - 1\right)\right) \\
\Rightarrow \alpha_0^* &= \omega E[\delta_0(X)] + (1-\omega) E(X_{n+1}) \exp\left(-\ln\left[1 + \sum_{i=1}^n \exp\left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)}\right)\right]\right) \\
\Rightarrow \alpha_0^* &= \omega E[\delta_0(X)] + (1-\omega) E(X_{n+1}) \exp\left(\ln\left[1 + \sum_{i=1}^n \exp\left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)}\right)\right]^{-1}\right) \\
\Rightarrow \alpha_0^* &= \omega E[\delta_0(X)] + (1-\omega) E(X_{n+1}) \left[1 + \sum_{i=1}^n \exp\left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)}\right)\right]^{-1},
\end{aligned}$$

donc

$$\alpha_0^* = \omega E[\delta_0(X)] + \frac{(1-\omega) E(X_{n+1})}{\left[1 + \sum_{i=1}^n \exp\left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)}\right)\right]},$$

puis

$$\begin{aligned}
\alpha_k &= \frac{(1-\omega) E(X_{n+1})}{E(X_k)} \exp\left(\lambda Cov(X_k, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_k)} - 1 - \lambda_0\right) \\
&= \frac{(1-\omega) E(X_{n+1})}{E(X_k)} \exp(-1 - \lambda_0) \exp\left(\lambda Cov(X_k, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_k)}\right) \\
\lambda_0 &= \ln\left[1 + \sum_{i=1}^n \exp\left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)}\right)\right] - 1,
\end{aligned}$$

alors

$$\begin{aligned}
\alpha_k^* &= \frac{(1-\omega) E(X_{n+1})}{E(X_k)} \exp\left(-1 - \left(\ln\left[1 + \sum_{i=1}^n \exp\left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)}\right)\right] - 1\right)\right) \\
&\quad \exp\left(\lambda Cov(X_k, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_k)}\right) \\
\Rightarrow \alpha_k^* &= \frac{(1-\omega) E(X_{n+1})}{E(X_k)} \exp\left(-\ln\left[1 + \sum_{i=1}^n \exp\left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)}\right)\right]\right) \\
&\quad \exp\left(\lambda Cov(X_k, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_k)}\right) \\
\Rightarrow \alpha_k^* &= \frac{(1-\omega) E(X_{n+1})}{E(X_k)} \exp\left(\ln\left[1 + \sum_{i=1}^n \exp\left(\lambda Cov(X_i, n\bar{X}) \frac{(1-\omega) E(X_{n+1})}{E(X_i)}\right)\right]^{-1}\right)
\end{aligned}$$

$$\begin{aligned}
& \exp \left(\lambda \text{Cov} (X_k, n\bar{X}) \frac{(1-\omega) E (X_{n+1})}{E (X_k)} \right) \\
\Rightarrow \alpha_k^* &= \frac{(1-\omega) E (X_{n+1})}{E (X_k)} \left[1 + \sum_{i=1}^n \exp \left(\lambda \text{Cov} (X_i, n\bar{X}) \frac{(1-\omega) E (X_{n+1})}{E (X_i)} \right) \right]^{-1} \\
& \exp \left(\lambda \text{Cov} (X_k, n\bar{X}) \frac{(1-\omega) E (X_{n+1})}{E (X_k)} \right) \\
\Rightarrow \alpha_k^* &= \frac{\frac{(1-\omega)E(X_{n+1})}{E(X_k)} \exp \left(\lambda \text{Cov} (X_k, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_k)} \right)}{\left[1 + \sum_{i=1}^n \exp \left(\lambda \text{Cov} (X_i, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_i)} \right) \right]},
\end{aligned}$$

donc

$$\alpha_k^* = \frac{(\alpha_0^* - \omega E [\delta_0 (X)]) \exp \left(\lambda \text{Cov} (X_k, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_k)} \right)}{E (X_k)}.$$

Par conséquent,

$$P_{n+1}^{MEM} = \alpha_0^* + \sum_{k=1}^n \alpha_k^* X_k,$$

comme voulu. ■

Corollaire 3.2.1 *L'estimateur linéaire non-homogène optimale P_{n+1}^{MEM} peut être écrit comme un compromis entre l'expérience individuelle et l'expérience collective :*

$$P_{n+1}^{MEM} = Z^* \bar{X} + (1 - Z^*) m, \tag{3.8}$$

avec $Z^* = \frac{\omega n V(\bar{X})}{(n\tau^2 + \sigma^2)(1-\omega)}$, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ et $m = \omega \bar{X} (2 - \omega) + (1 - \omega)^2 \mu$.

Proof. En premier lieu, nous devrions trouver la solution de λ . En supposant que

$$\text{Cov} (X_i, n\bar{X}) = n\tau^2 + \sigma^2, \quad \text{Cov} (\delta_0 (x), n\bar{X}) = nV (\bar{X}) \text{ et } E (X_{n+1}) = \frac{m - \omega \bar{X}}{(1-\omega)},$$

et en les remplaçant dans l'équation (3.7),

$$\sum_{i=1}^n \exp \left[\lambda \text{Cov} (X_i, n\bar{X}) \frac{(1-\omega) E (X_{n+1})}{E (X_i)} \right] \text{Cov} (X_i, n\bar{X}) \frac{(1-\omega) E (X_{n+1})}{E (X_i)}$$

$$\begin{aligned}
&= \left[1 + \sum_{i=1}^n \exp \left[\lambda Cov (X_i, n\bar{X}) \frac{(1-\omega) E (X_{n+1})}{E (X_i)} \right] \right] \omega Cov (\delta_0 (X), n\bar{X}) \\
\Rightarrow \sum_{i=1}^n \exp \left[\lambda Cov (X_i, n\bar{X}) \frac{(1-\omega) E (X_{n+1})}{E (X_i)} \right] \left[Cov (X_i, n\bar{X}) \frac{(1-\omega) E (X_{n+1})}{E (X_i)} - \omega Cov (\delta_0 (X), n\bar{X}) \right] \\
&= \omega Cov (\delta_0 (X), n\bar{X}) \\
\Rightarrow n \exp \left[\lambda (n\tau^2 + \sigma^2) \frac{(1-\omega) \left(\frac{m-\omega\bar{X}}{(1-\omega)} \right)}{\left(\frac{m-\omega\bar{X}}{(1-\omega)} \right)} \right] \left[(n\tau^2 + \sigma^2) \frac{(1-\omega) \left(\frac{m-\omega\bar{X}}{(1-\omega)} \right)}{\left(\frac{m-\omega\bar{X}}{(1-\omega)} \right)} - \omega nV (\bar{X}) \right] = \omega nV (\bar{X}) \\
\Rightarrow \exp \left[\lambda (n\tau^2 + \sigma^2) (1-\omega) \right] \left[(n\tau^2 + \sigma^2) (1-\omega) - \omega nV (\bar{X}) \right] = \omega V (\bar{X}) \\
\Rightarrow \ln \left(\exp \left[\lambda (n\tau^2 + \sigma^2) (1-\omega) \right] \left[(n\tau^2 + \sigma^2) (1-\omega) - \omega nV (\bar{X}) \right] \right) = \ln (\omega V (\bar{X})) \\
\Rightarrow \left[\lambda (n\tau^2 + \sigma^2) (1-\omega) \right] + \ln \left((n\tau^2 + \sigma^2) (1-\omega) - \omega nV (\bar{X}) \right) = \ln (\omega V (\bar{X})) \\
\Rightarrow \lambda (n\tau^2 + \sigma^2) (1-\omega) = \ln (\omega V (\bar{X})) - \ln \left((n\tau^2 + \sigma^2) (1-\omega) - \omega nV (\bar{X}) \right) \\
\Rightarrow \lambda (n\tau^2 + \sigma^2) (1-\omega) = \ln \left(\frac{\omega V (\bar{X})}{(n\tau^2 + \sigma^2) (1-\omega) - \omega nV (\bar{X})} \right),
\end{aligned}$$

on trouve

$$\lambda = \frac{1}{(n\tau^2 + \sigma^2) (1-\omega)} \ln \left(\frac{\omega V (\bar{X})}{(n\tau^2 + \sigma^2) (1-\omega) - \omega nV (\bar{X})} \right).$$

Ensuite, on peut obtenir

$$\begin{aligned}
\alpha_0^* &= \omega E [\delta_0 (X)] + \frac{(1-\omega) E (X_{n+1})}{\left[1 + \sum_{i=1}^n \exp \left(\lambda Cov (X_i, n\bar{X}) \frac{(1-\omega) E (X_{n+1})}{E (X_i)} \right) \right]} \\
\Rightarrow \alpha_0^* &= \omega (\bar{X}) + \frac{(1-\omega) \left(\frac{m-\omega\bar{X}}{(1-\omega)} \right)}{\left[1 + n \exp \left(\left(\frac{1}{(n\tau^2 + \sigma^2) (1-\omega)} \ln \left(\frac{\omega V (\bar{X})}{(n\tau^2 + \sigma^2) (1-\omega) - \omega nV (\bar{X})} \right) \right) (n\tau^2 + \sigma^2) \frac{(1-\omega) \left(\frac{m-\omega\bar{X}}{(1-\omega)} \right)}{\left(\frac{m-\omega\bar{X}}{(1-\omega)} \right)} \right) \right]} \\
&\Rightarrow \alpha_0^* = \omega \bar{X} + \left[\frac{m - \omega \bar{X}}{1 + \frac{\omega nV (\bar{X})}{(n\tau^2 + \sigma^2) (1-\omega) - \omega nV (\bar{X})}} \right] \\
&\Rightarrow \alpha_0^* = \omega \bar{X} + \frac{(m - \omega \bar{X})}{\left[\frac{(n\tau^2 + \sigma^2) (1-\omega)}{(n\tau^2 + \sigma^2) (1-\omega) - \omega nV (\bar{X})} \right]},
\end{aligned}$$

puis

$$\alpha_k^* = \frac{(\alpha_0^* - \omega E[\delta_0(X)]) \exp \left[\lambda \text{Cov}(X_k, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_k)} \right]}{E(X_k)}$$

$$\alpha_0^* - \omega E[\delta_0(X)] = \alpha_0^* - \omega E[\bar{X}] = \alpha_0^* - \omega \bar{X} = \omega \bar{X} + \frac{(m - \omega \bar{X})}{\left[\frac{(n\tau^2 + \sigma^2)(1-\omega)}{(n\tau^2 + \sigma^2)(1-\omega) - \omega nV(\bar{X})} \right]} - \omega \bar{X}$$

$$= \frac{(m - \omega \bar{X}) [(n\tau^2 + \sigma^2)(1-\omega) - \omega nV(\bar{X})]}{(n\tau^2 + \sigma^2)(1-\omega)} \exp \left[\lambda \text{Cov}(X_k, n\bar{X}) \frac{(1-\omega)E(X_{n+1})}{E(X_k)} \right]$$

$$= \exp \left[\left(\frac{1}{(n\tau^2 + \sigma^2)(1-\omega)} \ln \left(\frac{\omega V(\bar{X})}{(n\tau^2 + \sigma^2)(1-\omega) - \omega nV(\bar{X})} \right) \right) (n\tau^2 + \sigma^2) \frac{(1-\omega) \left(\left(\frac{m - \omega \bar{X}}{(1-\omega)} \right) \right)}{\left(\left(\frac{m - \omega \bar{X}}{(1-\omega)} \right) \right)} \right]$$

$$\frac{\omega V(\bar{X})}{(n\tau^2 + \sigma^2)(1-\omega) - \omega nV(\bar{X})},$$

donc

$$\alpha_k^* = \frac{\frac{(m - \omega \bar{X}) [(n\tau^2 + \sigma^2)(1-\omega) - \omega nV(\bar{X})]}{(n\tau^2 + \sigma^2)(1-\omega)} \frac{\omega V(\bar{X})}{(n\tau^2 + \sigma^2)(1-\omega) - \omega nV(\bar{X})}}{\left(\frac{m - \omega \bar{X}}{(1-\omega)} \right)}$$

$$\Rightarrow \alpha_k^* = \frac{(m - \omega \bar{X}) \omega V(\bar{X}) (1-\omega)}{(n\tau^2 + \sigma^2)(1-\omega) (m - \omega \bar{X})}$$

$$\Rightarrow \alpha_k^* = \frac{\omega V(\bar{X})}{(n\tau^2 + \sigma^2)}.$$

Enfin

$$P_{n+1}^{MEM} = \alpha_0^* + \sum_{k=1}^n \alpha_k^* X_k$$

$$\Rightarrow P_{n+1}^{MEM} = \omega \bar{X} + \frac{(m - \omega \bar{X})}{\left[\frac{(n\tau^2 + \sigma^2)(1-\omega)}{(n\tau^2 + \sigma^2)(1-\omega) - \omega nV(\bar{X})} \right]} + \sum_{k=1}^n \frac{\omega V(\bar{X})}{(n\tau^2 + \sigma^2)} X_k$$

$$\Rightarrow P_{n+1}^{MEM} = \omega \bar{X} + \left[\frac{(m - \omega \bar{X}) [(n\tau^2 + \sigma^2)(1-\omega) - \omega nV(\bar{X})]}{(n\tau^2 + \sigma^2)(1-\omega)} \right] + \frac{\omega nV(\bar{X}) \bar{X}}{(n\tau^2 + \sigma^2)}$$

$$\Rightarrow P_{n+1}^{MEM} = \omega \bar{X} - \left[\frac{\omega \bar{X} [(n\tau^2 + \sigma^2)(1-\omega) - \omega nV(\bar{X})]}{(n\tau^2 + \sigma^2)(1-\omega)} \right] + \frac{\omega nV(\bar{X}) \bar{X}}{(n\tau^2 + \sigma^2)} + \left[\frac{m [(n\tau^2 + \sigma^2)(1-\omega) - \omega nV(\bar{X})]}{(n\tau^2 + \sigma^2)(1-\omega)} \right]$$

$$\begin{aligned}
\Rightarrow P_{n+1}^{MEM} &= \bar{X} \left[\omega - \frac{\omega [(n\tau^2 + \sigma^2)(1 - \omega) - \omega nV(\bar{X})]}{(n\tau^2 + \sigma^2)(1 - \omega)} + \frac{\omega nV(\bar{X})}{(n\tau^2 + \sigma^2)} \right] + \left[m - \frac{m\omega nV(\bar{X})}{(n\tau^2 + \sigma^2)(1 - \omega)} \right] \\
\Rightarrow P_{n+1}^{MEM} &= \bar{X} \left[\frac{\omega (n\tau^2 + \sigma^2)(1 - \omega) - \omega [(n\tau^2 + \sigma^2)(1 - \omega) - \omega nV(\bar{X})] + (1 - \omega)\omega nV(\bar{X})}{(n\tau^2 + \sigma^2)(1 - \omega)} \right] \\
&\quad + \left[1 - \frac{\omega nV(\bar{X})}{(n\tau^2 + \sigma^2)(1 - \omega)} \right] m \\
\Rightarrow P_{n+1}^{MEM} &= \bar{X} \left[\frac{\omega^2 nV(\bar{X}) + \omega nV(\bar{X}) - \omega^2 nV(\bar{X})}{(n\tau^2 + \sigma^2)(1 - \omega)} \right] + \left[1 - \frac{\omega nV(\bar{X})}{(n\tau^2 + \sigma^2)(1 - \omega)} \right] m \\
\Rightarrow P_{n+1}^{MEM} &= \left[\frac{\omega nV(\bar{X})}{(n\tau^2 + \sigma^2)(1 - \omega)} \right] \bar{X} + \left[1 - \frac{\omega nV(\bar{X})}{(n\tau^2 + \sigma^2)(1 - \omega)} \right] m \\
&\Rightarrow P_{n+1}^{MEM} = Z^* \bar{X} + (1 - Z^*) m.
\end{aligned}$$

Avec $Z^* = \frac{\omega nV(\bar{X})}{(n\tau^2 + \sigma^2)(1 - \omega)}$, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ et $m = \omega \bar{X}(2 - \omega) + (1 - \omega)^2 \mu$. ■

3.3 Exemples numériques

Dans cette section, pour montrer l'efficacité de nos résultats théoriques, nous présentons une simulation numérique et une application aux données dans lesquels nous comparons la prime de crédibilité obtenue dans la section précédente avec celle de **Gómez-Déniz (2006)** en utilisant l'erreur quadratique moyenne comme critère d'évaluation.

3.3.1 Simulation numérique

La simulation numérique est effectuée pour la comparaison entre notre prime de crédibilité et celle de **Gómez-Déniz (2006)**.

Selon le théorème (5.1) de **Gómez-Déniz (2006)**, en utilisant la famille exponentielle des distributions, si nous obtenons une prime de bayes linéaire sous la fonction de perte classique, nous pouvons obtenir cette linéarité également sous la fonction de perte équilibrée.

La prime bayésienne de **Gómez-Déniz** $P_B^{\text{Gómez}}$, peut être écrite comme suit :

$$\begin{aligned}
P_{\mathbf{B}}^{\text{Gómez}} &= Z' m' + (1 - Z') [(1 - \omega)^2 + \omega \delta_0 (2 - \omega)] \\
&= \omega \delta_0 (2 - \omega) + (1 - \omega)^2 (Z' \bar{X} + (1 - Z') m').
\end{aligned} \tag{3.9}$$

Où $Z' = \frac{n\tau^2}{(n\tau^2 + \sigma^2)}$ est le facteur de crédibilité et m' la prime collective qui représente l'espérance de la prime individuelle sous la fonction de perte classique. Pour obtenir $P_{\mathbf{B}}^{\text{Gómez}}$ de la formule (3.9) et P_{n+1}^{MEM} , nous supposons que : $(\alpha, \beta) = ((9.1, 12.3), (9.1, 8.1), (8.2, 2.2))$, $\omega = (0.1, 0.6)$, $\delta_0 = 0.3$, $\sigma^2 = 8.1$, $\tau^2 = (0.3, 0.5, 0.98)$ et $\mu = (0.75, 3.75)$.

Nous considérons $n = 5, 10, 20$ respectivement et nous effectuons la simulation 10000 fois, les résultats sont présentés dans le tableau suivant :

TAB. 3.1 – Résultats de la simulation numérique

	La vouvelle prime			La prime de Gómez		
	\bar{X}	P_{n+1}^{MEM}	MSE	\bar{X}	$P_{\mathbf{B}}^{Gómez}$	MSE
5	0.7517415	0.7500368	$5.862563e - 04$	0.7017	0.7224318	$1.005855e - 02$
10	0.7531170	0.7497777	$1.612320e - 04$	0.7452	0.7587531	$9.510528e - 03$
20	0.7558048	0.7500193	$7.020930e - 05$	0.7644	0.7797584	$1.410505e - 03$
5	1.069021	1.067912	$3.778032e - 04$	1.1717	0.9224101	$5.412860e - 03$
10	1.073601	1.072102	$9.493073e - 04$	1.1956	1.0716982	$3.592476e - 04$
20	1.092397	1.073209	$6.08365e - 04$	1.2017	1.1002456	$1.4795834e - 03$
5	3.751431	3.749946	$5.848684e - 04$	3.6895	3.670213	$1.4712664e - 04$
10	3.754490	3.749961	$1.690337e - 04$	3.7052	3.7001661	$6.183221e - 02$
20	3.761044	3.754955	$7.473012e - 05$	3.7213	3.7254786	$5.942074e - 02$

D'après les résultats présentés dans le tableau ci-dessus, nous observons qu'il existe une forte corrélation positive entre les années d'assurance et l'expérience individuelle de l'assuré.

Ainsi, pour $P_{\mathbf{B}}^{Gómez}$, nous observons que chaque fois que β diminue il est accompagné d'une nette augmentation de la valeur de l'expérience individuelle \bar{X} c.-à-d. ajouter un poids à cette expérience qui explique la relation inverse entre β et \bar{X} .

En outre, contrairement à la prime de **Gómez**, nous observons que P_{n+1}^{MEM} converge toujours et rapidement vers l'expérience individuelle. À partir des valeurs de l'erreur quadratique moyenne, il semble que l'utilisation de la méthode de l'entropie maximale dans le cas équilibré soit efficace et donne une plus grande flexibilité à l'actuaire qui préfère travailler dans le cas empirique où toute hypothèse subjective est faite au sujet des sinistres.

3.3.2 Application à des données réelles :

Afin de montrer l'efficacité de nos résultats théoriques, nous utilisons un ensemble de données obtenues auprès du plus grand organisme de sécurité sociale en Grèce qui couvre 5 530 000 travailleurs et employés avec des soins médicaux, des médicaments et des soins hospitaliers (voir **Pitselis (2013)**). Les 6 contrats correspondent aux 6 différentes classes

(contrats) d'indemnités (dépenses) toutes couvertes par l'organisme de sécurité sociale de Grèce pour 22 ans d'expérience en matière de sinistres de 1980 à 2001. Les 6 différentes classes (contrats) sont :

$Sick_A$ = Allocation de maladie : les retraités actifs directement assurés ont droit à l'allocation de maladie.

$Accid_A$ = l'indemnité d'accident : si l'incapacité de travail est due à une plainte ou à une blessure due à un accident survenu pendant le travail.

$Matern_A$ = l'allocation de maternité : comprend l'allocation de confinement de la grossesse. Les bénéficiaires sont les femmes directement assurées, celles qui perçoivent une pension ou les épouses des maris ou des retraités directement assurés.

$Funer_{Exp}$ = Frais de funérailles : ils sont payés en cas de décès d'une personne directement assurée ou d'un pensionné pour cause de vieillesse, d'invalidité ou de décès.

$Other_A$ = Autres allocations : comprend (1) les soins dentaires, (2) les soins complémentaires, (3) le tourisme thérapeutique et une allocation de cure thermale pour les retraités à faibles revenus et (4) la médecine préventive.

$Manag_{Exp}$ = Frais de gestion : Comprend toutes les dépenses de gestion effectuées par l'organisme de sécurité sociale.

Pour calculer la prime de sinistre pour l'année 2002, nous utilisons les valeurs des paramètres de structure (m' , τ^2 , et σ^2) qui sont présentées dans le *Tableau 2* de **Pitselis (2013)**.

Considérons que δ_0 est la moyenne totale des montants des sinistres pour la période 1980–2001 : ($\delta_0 = \frac{\sum_{j=1}^6 \bar{X}_j}{6}$), le tableau 3-2 montre les valeurs de $P_{\mathbf{B}}^{\mathbf{Gómez}}$ et P_{2002}^{MEM} obtenus à partir de données réelles, le poids a été pris à $\omega = 0.49$.

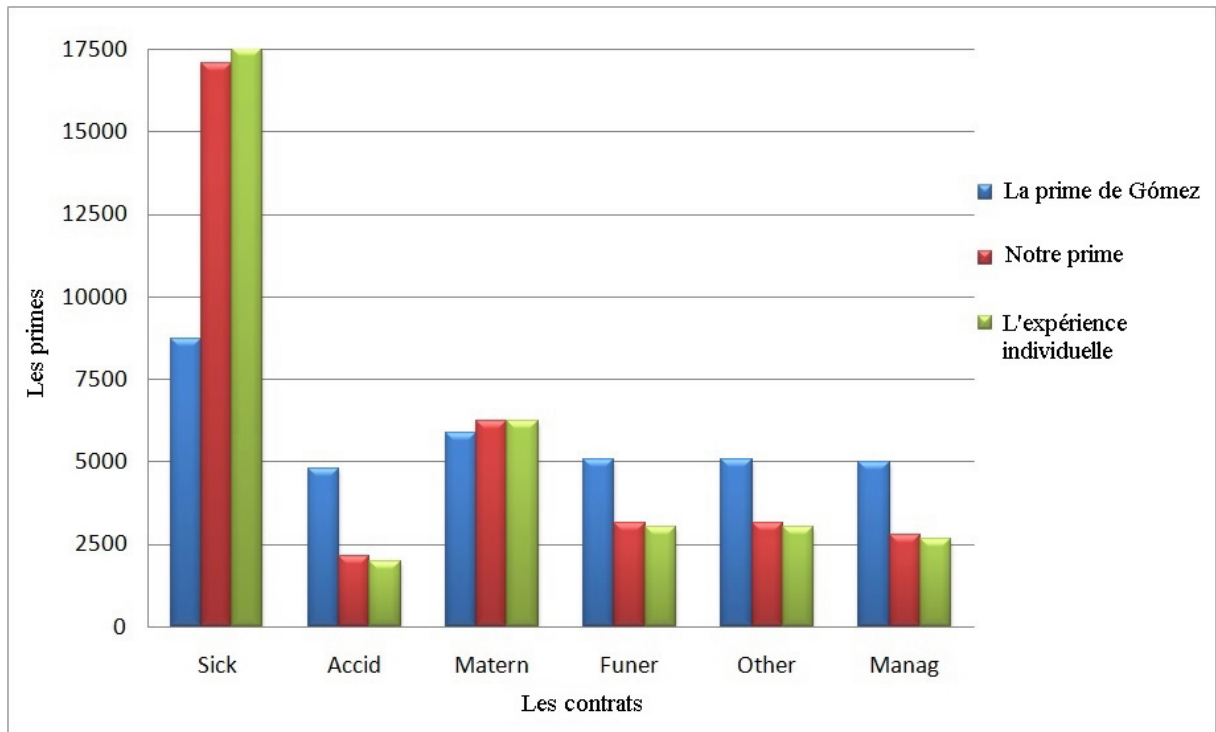


FIG. 3.1 – Résultats des primes $P_B^{Gómez}$ et P_{2002}^{MEM} à partir d'un ensemble de données de l'organisme de sécurité sociale en Grèce

TAB. 3.2 – Montants des primes pour les six contrats (Les primes sont en millions de drachmes :1 *Euro* = 340.75 *drachmas*)

Contrats	$P_B^{Gómez}$	P_{2002}^{MEM}	\bar{X}_j
Sick _A	8715.54	17065.242	17527.5
Accid _A	4788.29	2117.844	1970
Matern _A	5846.064	6211.706	6230.96
Funer _{Exp}	5054.893	3132.96	3026.55
Other _A	5055.153	3133.393	3027
Manag _{Exp}	4962.037	2778.825	2657.96

Selon les résultats obtenus dans le tableau 3-2 qui sont présentés dans la figure 3-1, nous pouvons voir clairement que la prime obtenue dans le cadre de notre approche P_{2002}^{MEM} est plus proche de l'expérience individuelle de chaque contrat \bar{X}_j . Le critère de proximité doit toujours être vérifié pour montrer que la prime à facturer reflète réellement l'historique des sinistres des assurés. Nous pouvons affirmer après deux exemples numériques que notre prime peut être utilisée par des actuaires qui recherchent toujours des primes compétitives qui attirent les assurés et empêchent leur entreprise de tomber en faillite.

Conclusion générale

Dans cette thèse, nous avons utilisé une technique très simple pour obtenir des primes de crédibilité dans le cadre d'une fonction de perte équilibrée. Cette méthodologie qui n'est pas très connue dans la littérature actuarielle nous donne une prime de crédibilité linéaire après un calcul simple. Par conséquent, nous pouvons étendre l'idée établie dans ce travail en effectuant un ajustement quadratique qui considère une fonction d'observations un ordre supérieur de moments et prend en compte les valeurs carrées des observations passées.

En outre, nous avons limité notre attention dans cette étude au choix d'un estimateur ciblé. Nous pourrions considérer dans de futures recherches diverses fonctions pour $\delta_0(x)$ comme par exemple la médiane des observations du portefeuille.

Bibliographie

- [1] Bailey, A., 1950. A generalized theory of credibility. Proceedings of the Casualty Actuarial Society 13, 13-20.
- [2] Berger, J.O. 1985 a. Statistical Decision Theory and Bayesian Analysis (2nd edition). Springer-Verlag, New York.
- [3] Berger, J.O. and Robert, C.P. 1990. Subjective hierarchical Bayes estimation of a multivariate normal mean : on the frequentist interface. Ann. Statist. 18, 617–651.
- [4] Blackwell, D. and Girshick, M.A. 1954. Theory of Games and Statistical Decisions. J. Wiley, New York.
- [5] Borda, Monica. 2011. Fundamentals in Information Theory and Coding. Springer.
- [6] Botev, Z. I. ; Kroese, D. P. 2008. Non-asymptotic Bandwidth Selection for Density Estimation of Discrete Data. Methodology and Computing in Applied Probability. 10 (3) : 435.
- [7] Botev, Z. I. ; Kroese, D. P. 2011. The Generalized Cross Entropy Method, with Applications to Probability Density Estimation. Methodology and Computing in Applied Probability. 13 (1) : 1–27
- [8] Bousquet, N. 2008. Eliciting vague but proper maximal entropy priors in Bayesian experiments. Statistical Papers. 51 (3) : 613–628.

-
- [9] Brown, L.D. 1971. Admissible estimators, recurrent diffusions, and insoluble boundary-value problems. *Ann. Math. Statist.* 42, 855–903.
- [10] Brown, L.D. 1976. Notes on Statistical Decision Theory. Unpublished lecture notes, Ithaca, New York.
- [11] Brown, L.D. 1993. Minimaxity, more or less. In *Statistical Decision Theory and Related Topics 5*, S.S. Gupta and J.O. Berger (eds.), 1–18. Springer-Verlag, New York.
- [12] Brown, L.D. 2000. An essay on Statistical Decision Theory. *J. Amer. Statist. Assoc.* 95, 1277–1282.
- [13] Bühlmann, H., 1967. Experience rating and credibility. *Astin Bulletin* 4 (3), 199-207.
- [14] Bühlmann, H., Straub, E., 1970. Glaubwürdigkeit für Schadensze. *Bulletin of the Swiss Association of Actuaries* 70, 111-133.
- [15] Clarke, B. 2006. Information optimality and Bayesian modelling. *Journal of Econometrics*. 138 (2) : 405–429.
- [16] Darooneh, A.H., 2004. Non-life insurance pricing : multi agent model. *The European Physical Journal B-Condensed Matter and Complex Systems* 42, 119-122.
- [17] Darooneh, A.H., 2006. Utility function from maximum entropy principle. *Entropy* 8, 18-24.
- [18] De Vylder, F., 1976. Geometrical credibility. *Scandinavian Actuarial Journal*, 121-149.
- [19] De Vylder, F., 1976. Optimal semilinear credibility. *Bulletin of swiss ass. Act.*, 78, 27-40.
- [20] Dey, D.G., Ghosh, M., Strawderman, W., 1999. On estimation with balanced loss functions. *Statistics & Probability Letters* 45, 97–101.

-
- [21] Druilhet, Pierre ; Marin, Jean-Michel (2007). Invariant {HPD} credible sets and {MAP} estimators. *Bayesian Anal.* 2 : 681–691.
- [22] Farsipour, N.S., Asgharzadhe, A., 2004. Estimation of a normal mean relative to balanced loss functions. *Statistical Papers* 45, 279–286.
- [23] Ferguson, T.S. 1967. *Mathematical Statistics : a Decision-Theoretic Approach*. Academic Press, New York.
- [24] Gauss, CF. 1810. Least squares method for combinations of observations, Mallet-Bacheleir-Paris.
- [25] Giffin, A. and Caticha, A., 2007, *Updating Probabilities with Data and Moments*.
- [26] Gómez-Déniz, E., 2006. On the use of the weighted balanced loss function to obtain credibility premiums. *International Conference on Mathematical and Statistical Modeling in Honor of Enrique Castillo*. June 28-30, 1-12.
- [27] Gómez-Déniz, E., 2008. A generalization of the credibility theory obtained by using the weighted balanced loss function. *Insurance : Mathematics & Economics* 42, 850-854.
- [28] Gómez, E., Pérez, J.M., Vázquez, F.J., 2006. On the use of posterior regret Γ -minimax actions to obtain credibility premiums. *Insurance : Mathematics and Economics* 39, 115–121.
- [29] Goovaerts, M.J., Kaas, R., Heerwaarden, A.E., Bauwelinckx, T., 1990. *Effective Actuarial Methods*. In : *Insurance Series*, vol. 3. North Holland.
- [30] Hachemeister, C. A., 1975. *Credibility for regression models with application to trend*. In : *Credibility, theory and applications*. Proceedings of the Berkeley actuarial research conference on credibility. Academic Press, New York.

-
- [31] Han, Te Sun & Kobayashi, Kingo. 2002. Mathematics of Information and Coding. American Mathematical Society.
- [32] Harremös, Peter. 2001. Binomial and Poisson distributions as maximum entropy distributions, IEEE Transactions on Information Theory, 47 (5) : 2039–2041.
- [33] Heilmann, W., 1989. Decision theoretic foundations of credibility theory. Insurance : Mathematics and Economics 8 (1), 77–95.
- [34] Herzog, T.N., 1996. Introduction to Credibility Theory, 2nd ed. ACTEX Publications, Winsted.
- [35] HU, Y., WU, L., SUN, Y., 2016. The Maximum entropy method to the credibility estimation.. Chinese Journal of Applied Probability and Statistics, 32(5), 463-475.
- [36] Jafari, M., Marchand, E., Parsian, A., 2006. On estimation with weighted balanced-type loss function. Statistics & Probability Letters 76, 773–780.
- [37] Jaynes, E.T., 1957. Information theory and statistical mechanics. Physical Reviews 106, 620-630.
- [38] Jaynes, E. T. 1957. Information Theory and Statistical Mechanics II. Physical Review. Series II. 108 (2) : 171–190.
- [39] Jaynes, E. T. 1968. Prior Probabilities. IEEE Transactions on Systems Science and Cybernetics. 4 (3) : 227–241.
- [40] Jaynes, E. T. 1988. The Relation of Bayesian and Maximum Entropy Methods, in Maximum-Entropy and Bayesian Methods in Science and Engineering (Vol. 1), Kluwer Academic Publishers, p. 25-29.

-
- [41] Jaynes, E. T. 2003. *Probability Theory : The Logic of Science*, Cambridge University Press, p. 351-355.
- [42] Jeffreys, H. 1961. *Theory of Probability* (third edition). Oxford University Press, London.
- [43] Jeweell, W.S. 1974. Credible means are exact Bayesian for exponential families. *Astin Bulletin* 8 (1), 77–90.
- [44] Jeweell, W. S. 1975. The use of collateral data in credibility theory : A hierarchical model. *Giornale dell'Istituto Italiano degli Attuari* 38, 1-16.
- [45] Kesavan, H. K. ; Kapur, J. N. 1990. Maximum Entropy and Minimum Cross-Entropy Principles. In Fougère, P. F. (ed.). *Maximum Entropy and Bayesian Methods*. pp. 419–432.
- [46] Kitamura, Y., 2006, *Empirical Likelihood Methods in Econometrics : Theory and Practice*, Cowles Foundation Discussion Papers 1569, Cowles Foundation, Yale University.
- [47] Landsman, Z., Makov, U. E., 1999. Credibility evaluation for the exponential dispersion family. *Insurance : Mathematics & Economics* 24, 23-29.
- [48] Landsman, Z., Makov, U.E., 2000. On credibility evaluation and the tail area of the exponential dispersion family. *Insurance : Mathematics & Economics* 27, 277–283.
- [49] Laplace, P.S. 1773. Mémoire sur la probabilité des causes par les événements. *Mémoires de l'Académie Royale des Sciences présentés par divers savans* 6, 621–656.
- [50] Lazar, N. 2003. Bayesian empirical likelihood. *Biometrika*. 90 (2) : 319–326.
- [51] Le Cam, L. 1990. Maximum likelihood : an introduction. *Int. Statist. Rev.* 58, 153–172.
- [52] Legendre, A. 1805. *Nouvelles Méthodes pour la Détermination des Orbites des Comètes*. Courcier, Paris.

-
- [53] Lehmann, E.L. and Casella, G. 1998. Theory of Point Estimation (second edition). Springer–Verlag, New York.
- [54] Lindley, D.V. 1980. Approximate Bayesian Method. *Trabajos de Estadística*. 31, 223-237.
- [55] Lisman, J. H. C. ; van Zuylen, M. C. A. 1972. Note on the generation of most probable frequency distributions. *Statistica Neerlandica*. 26 (1) : 19–23.
- [56] Mowbray, A., 1914. How extensive a payroll is necessary to give dependable pure premium ?. *Proceedings of the Casualty Actuarial Society* 1, 24-30.
- [57] Neyman, J. and Pearson, E.S. 1933 a. On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. Royal Soc. Ser. A* 231, 289–337.
- [58] Neyman, J. and Pearson, E.S. 1933 b. The testing of statistical hypotheses in relation to probabilities a priori. *Proc. Cambridge Philos. Soc.* 24, 492–510.
- [59] Owen, A. B., 2001, *Empirical Likelihood*, Chapman and Hall/CRC.
- [60] P.Robert, C, 2001. *The Bayesian Choice*. New York : Second Edition, SpringerVerlang.
- [61] Palmieri, Francesco A. N. ; Ciunozzo, Domenico 2013. Objective priors from maximum entropy in data classification. *Information Fusion*. 14 (2) : 186–198.
- [62] Park, Sung Y. ; Bera, Anil K. 2009. Maximum entropy autoregressive conditional heteroskedasticity model. *Journal of Econometrics*. 150 (2) : 219–230.
- [63] Pathria, R. K. ; Beale, Paul. 2011. *Statistical Mechanics (Third ed.)*. Academic Press. p. 51.
- [64] Payandeh Najafabadi, A. T., Hatami H., Omidi Najafabadi M., 2012. A maximum-entropy approach to the linear credibility formula. *Mathematics & Economics*, 51(1), 216-221.

-
- [65] Pitselis, G. 2013. Quantile credibility models. *Insurance : Mathematics and Economics*, 52, 477-489.
- [66] Schennach, S. M. 2005. Bayesian exponentially tilted empirical likelihood. *Biometrika*. 92 (1) : 31–46.
- [67] Schneider, T. D. 2007. Information theory primer with an appendix on logarithms, National Cancer Institute.
- [68] Shannon, C.E., 1948. A mathematical theory of communication. *Bell System Technical Journal* 27, 379-423. 623-659.
- [69] Skyrms, B. 1987. Updating, supposing and MAXENT. *Theory and Decision*. 22 (3) : 225–46.
- [70] Soofi, E.S. 2000. Principal Information Theoretic Approaches. *Journal of the American Statistical Association*. 95 (452) : 1349–1353.
- [71] Strasser, H. 1985. *Mathematical Theory of Statistics*. W. de Gruyter, Berlin.
- [72] Strawderman, W.E. 1974. Minimax estimation of location parameters for certain spherically symmetric distributions. *J. Multivariate Anal.* 4, 255–264.
- [73] Strawderman, W.E. 2000. Minimaxity. *J. Amer. Statist. Assoc.* 95, 1364–1368.
- [74] Thomas M. Cover ; Joy A. Thomas (1991). *Elements of Information Theory*. Hoboken, New Jersey : Wiley.
- [75] Varian, HR. A. 1975. Bayesain approach to real estate assement, Amsterdam, North Hol-land ,195–208.
- [76] Wald, A. 1950. *Statistical Decision Functions*. J. Wiley, New York.

-
- [77] Whitney, A., 1918. The theory of experience rating. Proceedings of the Casualty Actuarial Society 4, 274-292.
- [78] Zellner, A., 1994. Bayesian and non-Bayesian estimation using balanced loss function. In : Gupta, S. S., Berger, J. O. (Eds.), Statistical Decision Theory and Related Topics. Springer, New York, pp. 371-390.