وزارة التعليم العالي والبحث العلمي

BADJI MOKHTAR –ANNABA UNIVERSITY
UNIVERSITE BADJI MOKHTAR ANNABA

جامعة باجي مختار
- عنابة -

Faculté des Sciences
Département de Mathématiques
Laboratoire Probabilités et Statistique LaPS

THÈSE

Présentée en vue de l'obtention du diplôme de Doctorat en mathématiques

Prime de crédibilité par la méthode de l'entropie maximale en diversifiant la fonction des données observées

Filière : Mathématiques Appliquées

Spécialité : Actuariat

Présentée Par : Naim BOUDJELIDA

Devant le jury

PRESIDENT	Mohamed Cherif BOURAS	Prof.	U.B.M. ANNABA
DIRECTEUR DE THÈSE	Mohamed Riad REMITA	Prof.	ENSIA. Alger
EXAMINATEUR	Abdelali EZZEBSA	MCA.	U. 08 MAI 1945. Guelma
EXAMINATEUR	Farouk METIRI	MCA.	U.B.M. ANNABA
EXAMINATRICE	Ahlem DJEBAR	MCA.	U.B.M. ANNABA

Année : 2023 / 2024

Remerciements

Dieu merci de m'avoir donner le courage et la volonté d'achever ce travail.

Je tiens à exprimer ma reconnaissance et ma gratitude à mon encadreur, le Prof. REMITA Mohamed Riad, qui a bien voulu accepter de m'accorder ce privilège d'encadrement et d'avoir consacré beaucoup de temps pour les orientations et le suivi jusqu'à la rédaction de la thèse.

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

Je remercie également le Prof. BOURAS Mohamed Cherif d'avoir accepté la présidence du jury.

Je remercie également les membres de jury d'avoir accepter juger mon travaille : Dr. EZZEBSA Abdelali, Dr. METIRI Farouk, Dr. DJEBAR Ahlem

Je tiens à remercier les membres de laboratoire Prababilités et Statistique LaPS notemment le directeur de laboratoire Pr. ZEGHDOUDI Halim

Je voudrais également exprimer ma gratitude envers mes amis, collègues et enseignants de l'université d'Annaba, notamment Prof. Haiour Mohammed, Dr. ABBAS Abderrahmane, Dr. SAOULI Nabil, Dr. SADOUN Ahmed et Dr. HADDARI Alla pour leurs soutient et encouragement

Je tiens à remercier mes parents pour leurs encouragements tout au long de mes années d'études.

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

Dédicace

Je dédie ce modeste travail à vous, qui êtes les piliers de ma vie et les étoiles qui illuminent mon chemin.

À mes parents :

Maman, papa, votre amour inconditionnel et vos encouragements ont été ma force tout au long de ce travail. Vos sacrifices et votre bienveillance ont façonné ma persévérance. Chaque ligne écrite ici porte un peu de votre sagesse et de votre soutien.

À mes enfants :

Mes trésors : Sirine et Mohamed Islem, ce travail est un témoignage de mon engagement envers l'apprentissage et la croissance. Puissiez-vous trouver dans ces pages l'inspiration pour poursuivre vos propres rêves. Que chaque formule, chaque idée, soit un héritage que je vous lègue avec amour.

À vous tous :

Que cette dédicace soit le reflet de ma gratitude éternelle à tous ceux qui m'ont aidé de près ou de loin.

Résumé

Dans le domaine de l'assurance, repérer la prime de crédibilité la plus efficace est un grand défi, essentiellement lorsque la distribution des sinistres $X_i(i = 1, 2, ..., n)$ reste inconnue. Cette étude utilise une méthodologie bayésienne pour aborder cette question. Dans ce cadre, les hypothèses initiales concernant les paramètres non reconnus du processus de sinistres sont présentées au moyen d'une fonction de distribution $f(x|\theta)$. L'objectif est de trouver la meilleure prime de crédibilité en maximisant l'utilité attendue de la compagnie d'assurance tout en respectant des contraintes spécifiques.

En appliquant la méthode de l'entropie maximale, il est possible de calculer la distribution postérieure du processus de pertes en utilisant les données observées à notre disposition. Cette approche est avantageuse car elle offre une flexibilité de modélisation et permet des estimations mises à jour progressivement à mesure que de nouvelles données deviennent disponibles. Elle est également particulièrement utile dans les situations où les données sont limitées ou incomplètes, comme c'est souvent le cas dans le secteur de l'assurance.

L'étude met en évidence les caractéristiques et les avantages de cette méthode bayésienne. Elle souligne comment l'approche gère l'incertitude et intègre les connaissances antérieures dans le processus d'estimation. De plus, une illustration numérique des données réelles est incluse pour démontrer l'utilité d'appliquer de la méthode et son impact sur la détermination de la crédibilité de la prime par rapport à celle obtenue par **Gomez-Deniz (2006)** et **Alla et al. (2021)**.

En résumé, cette recherche offre une perspective innovante sur la gestion des risques dans le secteur de l'assurance. Elle propose une méthode qui non seulement adhère à l'approche ludique de la gérance des risques potentielles, mais aussi à l'approche de la prime de crédibilité.

Mots clés : *Prime de crédibilité, entropie maximale, fonction de perte, processus de sinistre.*

Abstract

In the intricate field of insurance, finding the credibility of the most well planned premium is a major challenge, especially when the distribution of claims $X_i(i = 1, 2, ..., n)$ remains unknown. This study uses a Bayesian methodology to address this issue. In this framework, the initial assumptions concerning the unidentified parameters of the claims process are expressed by means of a distribution function $f(x|\theta)$. The objective is to establish premium credibility by maximizing the insurance company's expected utility while respecting specific constraints. By applying the maximum entropy method, we can derive the posterior distribution of the loss process from the observed data available. This approach is advantageous because it offers modeling flexibility and enables estimates to be progressively updated as new data becomes available. It is also particularly useful in situations where data is limited or incomplete, as is often the case in the sector of insurance. The study spotlight the characteristics and advantages of this Bayesian method. It highlights how the approach handles uncertainty and integrates prior knowledge into the estimation process. In addition, numerical illustration to real data is included to demonstrate the practical application of the method and its impact on determining premium credibility comparing to premium obtained of **Gomez-Deniz(2006)** and **Alla et al.(2021)**. In summary, this research offers an innovative perspective on risk management in the insurance sector. It proposes a me-

thod that not only adheres to the playful approach to risk management, but also to the premium credibility approach.

Key words : *Premium credibility, maximum entropy, loss function, claim process.*

ملخص

في مجال التأمين المعقد، تحديد مصداقية القسط الأكثر فعالية هو تحدي كبير، خاصةً عندما يظل توزيع الخسائر غير معروفٍ. تستخدم هذه الدراسة منهجية بايزية للتعامل مع هذا السؤال. في هذا السياق، يتم التعبير عن الافتراضات الأولية المتعلقة بالمعلمات غير المعروفة لعملية المطالبات من خلال وظيفة توزيع. الهدف هو تحديد مصداقية القسط من خلال تحقيق أقصى فائدة متوقعة لشركة التأمين مع احترام القيود المحددة. من خلال تطبيق طريقة الانتروبيا القصوى، يمكننا استنتاج التوزيع اللاحق لعملية الخسائر من البيانات المتاحة. هذا النهج مفيد لأنه يوفر مرونة في النمذجة ويسمح بتحديث التقديرات تدريجيًا مع توفر بيانات جديدة. كما أنه مفيد بشكل خاص في الحالات التي تكون فيها البيانات محدودة أو غير مكتملة، كما هو الحال غالبًا في قطاع التأمين. تسلط الدراسة الضوء على ميزات وفوائد هذا النهج البايزي. وتوضح كيف يدير هذا النهج عدم اليقين ويدمج المعرفة السابقة في عملية التقدير. بالإضافة إلى ذلك، يتم تضمين عدة أمثلة رقمية لتوضيح التطبيق العملي للنهج وتأثيره على تحديد مصداقية القسط. إجمالًا، تقدم هذه الدراسة منظورًا مبتكرًا حول إدارة المخاطر في قطاع التأمين. وتقترح نهجًا يندرج ليس فقط في نهج اللعب في إدارة المخاطر، ولكن أيضًا في نهج مصداقية القسط العربية

Table des matières

Introduction

L'assurance est un domaine complexe et essentiel pour gérer des risques financiers. Elle permet aux souscripteurs (individus et entreprises) de céder une partie de leurs risques à une compagnie d'assurance en échange du paiement de primes. L'objectif principal de l'assurance est de protéger contre les pertes financières imprévues, qu'il s'agisse de dommages matériels, de maladies, d'accidents ou d'autres événements. Dans ce contexte, le rôle de l'actuaire est crucial. Ce dernier est chargé d'estimer les réserves nécessaires pour faire face aux dépenses futures liées aux sinistres, de calculer les risques associés aux différentes polices d'assurance et de tarifer ces contrats de manière à maximiser les bénéfices de la compagnie tout en minimisant les imprévus.

Les techniques actuarielles, qui reposent sur la théorie des probabilités et de la statistique, sont employées par l'actuaire pour évaluer et gérer les risques dans le domaine de l'assurance. Ces dernières sont listées ci-dessous :

1. **Système bonus-malus :** Ce système ajuste les primes en fonction du comportement de l'assuré. Si l'assuré a peu de sinistres, il bénéficie d'un bonus (réduction de prime). En revanche, s'il a de nombreux sinistres, il subit un malus (augmentation de prime).

2. **Modèles linéaires généralisés (GLM) :** Les GLM sont des outils statistiques qui permettent de simuler le lien entre une variable de réponse (par exemple, le coût des sinistres) et des variables explicatives (comme l'âge, le sexe, etc.). Ils sont largement utilisés pour prévoir les coûts des sinistres.

3. **Théorie de la crédibilité :** La théorie de la crédibilité est un domaine important de l'actuariat. Elle vise à estimer les risques lorsque les données sont

minimes. Pour calculer une prime de manière optimale, l'actuaire associe l'expérience individuelle (données de l'assuré) aux informations collectives (données agrégées). Cette approche tient compte des biais potentiels dans les données et permet d'ajuster la prime en conséquence.

Cette théorie a été développée par **Mowbray (1914)**, suivi par **Whitney (1918)**, qui a suggéré de fusionner l'expérience individuelle et la prime collective en utilisant un facteur de crédibilité M (avec $0 \leq M \leq 1$) pour calculer une prime unique. **Bailey (1950)** a démontré que l'estimateur résultant, obtenu en minimisant l'erreur quadratique dans un contexte bayésien, correspond précisément à la prime de crédibilité. Par la suite, le modèle de **Bühlmann (1967)** est révélé, avec l'idée de forcer la prime bayésienne à être linéaire. Le modèle de Bühlmann repose essentiellement sur la formule suivante :

$$P_c = M\bar{X} + (1 - M)\eta$$

où P_c est la prime de crédibilité, μ est la prime collective, M est le facteur de crédibilité (avec $0 \leq M \leq 1$), $M = \frac{n\phi^2}{n\phi^2+\rho^2}$, et $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Les termes $\rho^2 = E[V(X_i|\theta)]$, $\phi^2 = V[E(X_i|\theta)]$, et $\eta(\theta) = E(X_i|\theta)$ sont également définis.

Des poids naturels ont été introduit par Bühlmann et Erwin Straub pour développer leur célèbre modèle qui porte leurs noms (**Bühlmann et Straub, 1970**). Nombreux autres modèles de crédibilité ont vu le jour dans les années 70, tels que celui de **Hachemeister (1975)**, le modèle hiérarchique de **Jewel (1975)** et le modèle de **De Vylder (1976)**.

Dans la pratique, les actuaires ont constaté que les compagnies d'assurance risquaient la faillite si elles se contentaient de percevoir la prime nette des assureurs.

Pour résoudre ce problème et améliorer l'estimation des risques futurs, certains auteurs ont changé la fonction de perte quadratique classique par une fonction pondérée nommée la fonction de perte équilibrée pondérée :

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

où $\delta_0(x)$ est une fonction des données observées choisie par l'actuaire, et $0 \leq \omega \leq 1$ est un facteur de pondération déterminé par l'actuaire. En utilisant la fonction de perte équilibrée pondérée (**WBLF**), on peut obtenir des primes plus flexibles, permettant aux différentes compagnies d'assurance d'adopter des fonctions de pondération positives sur les marchés. Par exemple, pour $h(x) = 1$ et $y(x) = \exp(rx)$ (avec $r > 0$), en effectuant une minimisation, nous obtenons respectivement les primes de crédibilité équilibrées nettes et équilibrées d'Esscher. L'utilisation de la fonction $y(x) = \exp(rx)$ est restreinte au cas où l'actuaire souhaite 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, **Gomez-Déniz (2008)** a élargi son usage dans un contexte actuariel, où P est la prime de crédibilité.

Dans la théorie classique, pour trouver la prime de crédibilité il est important de faire certaines hypothèses sur la distribution des sinistres X_i (pour $i = 1, 2, \ldots, n$). Dans cette thèse, notre but est de déterminer des primes de crédibilité sans nécessité de fixer la fonction de distribution des sinistres et se basant sur la méthode de l'entropie maximale.

La méthode de l'entropie maximale est basée sur l'hypothèse du choix de la distribution qui a le plus d'incertitude, ou l'entropie la plus élevée, parmi un ensemble de distributions possibles, en fonction de contraintes données.

Cette thèse est structurée en trois parties. Le premier chapitre offre un aperçu sur les fonctions de pertes en actuariat, suivi d'une démonstration des différentes fonctions de perte fréquemment utilisées. Le deuxième chapitre introduit les concepts clés de la méthode de l'entropie maximale, illustrés par des exemples d'applications. Le troisième chapitre se base sur l'estimation de la prime de crédibilité dans le cas où la distribution de probabilités des sinistres est non communiquée $X_i (i = 1, 2, \ldots, n)$, à part les données observées. Pour gérer ce cas, nous employons la méthode de l'entropie maximale avec une fonction de perte équilibrée. Ce chapitre se conclut par une application aux données, où nous comparons la prime de crédibilité obtenue avec celle de **Gomez-Deniz (2006)** et celle obtenue par **Alla et al (2021)**. Les conclusions et les perspectives futures sont présentées après le dernier chapitre.

Chapitre 1

Fonctions de Perte et Primes de Crédibilité en Actuariat

1.1 Fondation de la Théorie de la Décision

1.1.1 Évaluation des Estimateurs

Dans le domaine de l'inférence statistique, l'objectif ultime est souvent de guider le statisticien, ou son client, vers une décision éclairée. Il est donc primordial de disposer d'un critère d'évaluation pour évaluer les procédures de décision. Les conséquences de chaque décision sont prises en considération par ce critère en fonction des paramètres du modèle, reflétant ainsi la réalité de la population étudiée.

Les décisions influencées par ce critère peuvent varier grandement, allant de l'investissement en bourse basé sur les rendements futurs estimés θ, à la décision de conclure une étude biologique évaluant l'efficacité θ d'un vaccin viral innovant. Il est crucial d'évaluer l'impact de ces décisions, c'est-à-dire de vérifier la cohérence d'une théorie avec les données expérimentales disponibles. Sans un tel critère, il devient difficile de faire la distinction entre les procédures de décision efficaces et les approches non fondées.

La théorie de la décision fournit un contexte structuré pour cette analyse critique. Toutefois, le choix du critère d'évaluation est sujet à débat en raison de ses implications significatives sur les décisions finales. Cette complexité a même poussé certains statisticiens à rejeter la théorie de la décision, arguant que le critère idéal est souvent impraticable ou indéterminable.

En règle générale, ce critère est symbolisé par la notion de perte, définie dans toutes les décisions envisageables. L'espace de décision, souvent désigné par D, se concentre généralement sur le cas $D = \Theta$, qui illustre le contexte standard d'estimation.

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

Où $L(d, \theta)$ symbolise l'erreur ou la sanction liée à la décision d lorsque le paramètre véritable est θ. Dans le contexte traditionnel de l'estimation des paramètres, la fonction de perte $L(d, \theta)$ mesure l'erreur découlant de l'estimation de θ par d. Déterminer avec précision la fonction de perte est fréquemment un challenge en pratique, particulièrement lorsqu'on envisage des populations de grande envergure ou infinies. Dans les modèles qualitatifs, l'évaluation des impacts de chaque décision peut se révéler complexe.

La subjectivité inhérente à la fonction de perte rend sa détermination complexe. Les statisticiens se tournent alors fréquemment vers des fonctions de perte classiques, choisies pour leur simplicité et leur facilité de manipulation mathématique. Ces fonctions sont aussi essentielles pour l'élaboration théorique de procédures optimales, en l'absence de justification pratique pour une fonction de perte spécifique.

Les fonctions de perte classiques trouvent leurs origines chez **Laplace (1773)** avec la perte d'erreur absolue, et **Gauss (1810)** pour la perte quadratique. À cette époque, Il était fréquent de faire la confusion entre le biais des estimateurs et celui des variances. Néanmoins, l'utilisation répandue de ces fonctions de perte ne doit pas être interprétée comme une validation de leur pertinence. En pratique, il est souvent préférable de prendre une décision basée sur un critère approximatif dans un délai raisonnable, plutôt que de chercher indéfiniment la fonction de perte parfaite.

L'inférence bayésienne repose sur trois piliers fondamentaux qui doivent être spécifiés de manière précise :

— La loi de probabilité qui présente les observations $f(x|\theta)$;

— La distribution initiale des paramètres $\pi(\theta)$;

— La fonction de coût liée aux décisions prises, $L(\delta, \theta)$.

Ces éléments sont souvent le résultat de jugements subjectifs, notamment le prior et la fonction de coût. Les critiques fréquentistes pointent du doigt les bayésiens qui ne prennent pas en compte la complexité de construire une fonction de coût, qui est aussi ardue que l'établissement d'une distribution préalable. **Lindley** en **(1985)** a mis l'accent sur le fait que la fonction de coût et les priors sont étroitement liés et nécessitent une analyse conjointe.

Exemple : Estimation de la Moyenne d'une Distribution Normale Prenons l'exemple de l'estimation de la moyenne θ d'une distribution normale $X \sim \mathcal{N}_n(\theta, \Sigma)$, où Σ est une matrice diagonale connue avec des éléments ρ_i^2 sur la diagonale. En l'absence d'informations supplémentaires, il est judicieux de choisir une fonction de coût qui traite chaque composante de manière équitable, soit une fonction de la forme :

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

où L atteint son minimum à zéro. Cette méthode garantit que les composantes avec des variances plus élevées n'affectent pas excessivement le choix de l'estimateur. Autrement dit, les composantes avec une grande variance ne sont pas sur-représentées lorsque les erreurs d'estimation sont normalisées par ρ_i. La fonction de coût quadratique $L(\phi) = \phi^2$ est souvent choisie, signifiant que l'erreur globale d'estimation est la somme des carrés des erreurs individuelles. La théorie de la décision en actuariat repose sur l'évaluation des estimateurs, qui vise à évaluer la performance des méthodes statistiques utilisées pour estimer les paramètres d'intérêt. Un estimateur est jugé selon son biais, sa variance, et son erreur quadratique moyenne (EQM), définie par :

$$EQM(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] \tag{1.1}$$

où $\hat{\theta}$ est l'estimateur et θ est le paramètre réel.

1.1.2 Utilité et Perte

Dans le contexte statistique, la théorie de la décision nous informe que le modèle statistique est composé de trois espaces : X, l'espace d'observation, Θ, l'espace des paramètres, et D, l'espace des décisions. L'inférence statistique vise à prendre une décision $d \in D$ liée au paramètre $\theta \in \Theta$ basée sur l'observation $x \in X$, où x et θ sont liés par la fonction de distribution $f(x|\theta)$. Dans la plupart des cas, la décision sera d'estimer une fonction de θ, $h(\theta)$, de la manière la plus appropriée possible.

La théorie de la décision ajoute que chaque action d peut être évaluée et donc associée à une récompense ϕ avec une utilité $U(\phi)$ (qui existe sous l'hypothèse de la rationalité des décideurs). Cette utilité est représentée sous la forme $U(\phi, \theta)$ pour montrer l'influence de ces deux facteurs. Lorsque d'autres facteurs aléatoires ϕ sont impliqués dans U, l'utilité devient :

$$u(\phi, \theta) = E[U(\phi)].$$

Cela implique que $U(\theta, d)$ peut être considéré comme une mesure de la proximité entre l'estimation proposée d et la valeur réelle $y(\theta)$. Il suffit de construire ou d'approximer la fonction d'utilité pour définir la fonction de perte correspondante :

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

En général, la fonction de perte est supposée non négative, ce qui renvoie à $U(\theta, d) \leq 0$, cela signifie qu'il n'y a aucune décision avec une utilité infinie. L'existence d'une limite inférieure sur L peut être critiquée comme étant stricte. Du point de vue statistique, la fonction de perte L représente la perte (ou l'erreur) due à une mauvaise évaluation de la fonction θ qui nous intéresse, et donc la meilleure évaluation de cette fonction est une perte nulle lorsque θ est connue. Dans le cas contraire, il y aurait une discontinuité de la fonction de perte qui pourrait même empêcher le choix d'une procédure de décision.

Évidemment, à l'exception des paramètres les plus ordinaires (ou simples), il est généralement impossible de minimiser uniformément en d la fonction de perte lorsque θ est inconnu. Pour dériver un critère de comparaison efficace de la fonction de perte, Selon l'approche fréquentiste, il est préférable de prendre en compte la perte moyenne.

(ou le risque fréquentiste) :

$$R(\theta, \delta) = E_\theta[L(\theta, \delta(x))] = \int_x L(\theta, \delta(x)) f(x|\theta) dx, \tag{1.2}$$

où $\delta(x)$ est la règle de décision, c'est-à-dire assigner une décision à chaque résultat $x \sim f(x|\theta)$ de l'expérience aléatoire. La fonction δ, de X en D, est nommée estimateur (alors que la valeur $\delta(x)$ est nommée estimation de θ). Lorsqu'il n'y a pas de risque de confusion, on note par D l'ensemble des estimateurs. La notion fréquentiste s'appuie sur ce critère pour comparer les estimateurs et, si possible, choisir 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. Cependant, cette approche présente plusieurs difficultés :

— La perte (l'erreur) est faite en moyenne sur les différentes valeurs de x proportionnellement à la 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 convaincante pour un client, qui demande des résultats optimaux par rapport à ses données x et pas celles 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, $R(\theta, d)$ est approximativement la perte moyenne sur les répétitions indépendantes et identiquement distribuées de la même expérience, selon la loi des grands nombres.

— Pour une procédure δ, le risque $R(\theta, \delta)$ est une fonction du paramètre θ. Il est difficile de comparer entre les procédures de décision ou même impossible de comparer par l'approche fréquentiste, car deux fonctions de risque croisées empêchent la comparaison entre les estimateurs correspondants. Il est préférable de trouver s'il existe une décision δ_0 qui minimise $R(\theta, \delta)$ uniformément en θ. De tels cas se produisent rarement, seulement si l'espace des procédures est limité. Les meilleures procédures ne peuvent être obtenues qu'en limitant assez artificiellement l'ensemble des procédures autorisées.

1.1.3 L'Approche Bayésienne : Une Alternative Éclairée

L'approche bayésienne diffère fondamentalement de l'approche fréquentiste en intégrant sur l'espace Θ, traitant θ comme inconnu, plutôt que sur l'espace X, puisque x est connu. Elle repose sur l'espérance de perte a posteriori :

$$\rho(\pi, d|x) = \mathbb{E}_x[L(\theta, d)|x] = \int_\Theta L(\theta, d)\mu(\theta|x)d\theta \tag{1.3}$$

où $\mu(\theta|x)$ est la loi a posteriori continue de θ, reflétant la moyenne de l'erreur en fonction de la distribution postérieure de θ, compte tenu de l'observation x.

Avec une distribution a priori μ, on peut également définir le risque intégral, qui est la moyenne du risque fréquentiste sur les valeurs de θ selon leur distribution a priori :

$$r(\mu, \delta) = \mathbb{E}^*[R(\theta, \delta)] = \int_\Theta \int_X L(\theta, \delta(x))f(x|\theta)dx\mu(\theta)d\theta$$

Ce concept attribue à chaque estimateur un nombre réel, permettant une comparaison directe des estimateurs. Ainsi, l'approche bayésienne, en tenant compte de l'information a priori, permet une prise de décision plus compatible et efficace, et les deux notions sont équivalentes en ce qu'elles conduisent à la même décision.

Théorème 1.1.1. *Un estimateur qui minimise le risque intégral $r(\mu, \delta)$ peut être obtenu en choisissant, pour chaque $x \in \mathcal{X}$, la valeur de $\delta(x)$ qui minimise la perte a posteriori $\rho(\mu(\theta), \delta|x)$, car :*

$$r(\mu, \rho) = \int_{\mathcal{X}} \rho(\mu, \delta(x)|x) f(x)\, dx \tag{1.4}$$

où

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

.

Preuve. En appliquant le théorème de Fubini pour intervertir l'ordre des intégrations, nous obtenons :

$$r(\mu, \delta) = \int_{\Theta} \int_{\mathcal{X}} L(\theta, \delta(x)) f(x|\theta)\mu(\theta)\, dx \mu(\theta)\, d\theta$$

$$= \int_{\mathcal{X}} \int_{\Theta} L(\theta, \delta(x)) f(x|\theta)\mu(\theta)\, d\theta\, dx$$

$$= \int_{\mathcal{X}} \int_{\Theta} L(\theta, \delta(x))\mu(\theta|x)\, d\theta f(\theta)\, dx$$

□

Définition 1.1.2. *Un estimateur bayésien associé à une distribution a priori μ et une fonction de perte L est un estimateur $\delta^{\mu}(x)$ qui minimise $r(\mu, \delta)$. Pour chaque $x \in \mathcal{X}$,* *il est donné par $\delta_\mu(x)$, le minimum selon d de $\rho(\mu, d|x)$. La valeur obtenue est alors appelée le risque de Bayes.*

Le théorème ci-dessus fournit une méthode pour déterminer les estimateurs bayésiens. Dans la perspective bayésienne, seule l'espérance de perte a posteriori $\rho(\mu, \delta|x)$ est pertinente, car elle est basée sur une approche conditionnelle. Il n'est pas optimal de moyenniser sur toutes les valeurs possibles de x lorsque la valeur observée de x est connue. Cependant, l'équivalence établie dans le théorème est cruciale car elle montre que l'approche conditionnelle n'est pas aussi fondamentale que les fréquentistes pourraient le prétendre. L'approche bayésienne, tout en fonctionnant sous la condition de l'observation actuelle x, intègre également les propriétés probabilistes de la distribution de l'observation $f(x|\theta)$. De plus, cette équivalence crée un pont entre les résultats classiques de la théorie des jeux et l'approche axiomatique bayésienne, basée sur la distribution a posteriori, et explique l'importance des estimateurs bayésiens dans les critères d'optimalité fréquentistes.

Ce résultat est valable pour les distributions a priori correctes et incorrectes, tant que le risque de Bayes $r(\mu)$ est fini. Dans le cas contraire, nous définissons un estimateur bayésien généralisé, qui minimise la perte a posteriori attendue pour chaque x. Du point de vue de l'optimalité fréquentiste, la distinction entre les distributions a priori correctes et incorrectes est moins importante que celle entre les estimateurs bayésiens réguliers et généralisés, car ces derniers sont admissibles. Il est à noter que les estimateurs bayésiens sont uniques si les fonctions de perte sont strictement convexes.

1.1.4 Critères d'Optimalité

Dans le cadre de la théorie de la décision fréquentiste, deux critères d'optimalité se distinguent comme fondamentaux, introduits respectivement par **Wald (1950)** et par **Neyman and Pearson (1933)**. À l'opposé de la perspective bayésienne, la méthodologie fréquentiste ne garantit pas systématiquement la conduite vers un estimateur optimal. Néanmoins, il est impératif de considérer et d'examiner minutieusement l'approche fréquentiste, même en l'absence de données préalables, puisqu'elle révèle fréquemment que les estimateurs bayésiens coïncident avec les critères d'optimalité fréquentistes. En d'autres termes, l'exploitation efficace de la distribution a priori dans la dérivation d'estimateurs performants nous autorise à remettre en question la pertinence du paradigme bayésien et à négliger l'importance de cette distribution. Par conséquent, l'approche bayésienne s'avère être d'une importance équivalente à celle de la fréquentiste, en nous dotant d'un mécanisme robuste pour l'élaboration d'estimateurs optimaux. Des illustrations de cette assertion ont été fournies par **Brown (1971,2000)**, **Strawderman (1974)**, ainsi que par **Berger (1985)** et **Berger and Robert (1990)**.

Les critères d'optimalité, tels que la minimaxité et l'admissibilité, sont des concepts clés pour choisir entre différents estimateurs ou décisions. Un estimateur est dit minimax s'il minimise la perte maximale possible, tandis qu'un estimateur est admissible s'il n'existe pas d'autre estimateur qui soit toujours meilleur (ou au moins aussi bon) pour tous les paramètres possibles.

Dans l'analyse de la fonction d'utilité, il est impératif d'élargir l'espace de décision pour inclure tous les estimateurs stochastiques qui prennent des valeurs dans D^*, qui représente l'ensemble des distributions de probabilité sur D. L'emploi d'un estimateur stochastique δ^* implique que l'action est choisie conformément à la distribution ayant la densité de probabilité $\delta^*(x, \cdot)$, suite à l'acquisition de l'observation x.

La perte associée à un estimateur stochastique δ^* est alors exprimée en tant que perte espérée :

$$L(\theta, \delta^*(x)) = \int_D L(\theta, a)\delta^*(x, a)da, \tag{1.5}$$

où a désigne un vecteur de paramètres. Une telle extension s'avère cruciale pour aborder les notions de minimaxité et d'admissibilité. Il est manifeste que l'utilisation de ces estimateurs est inappropriée, notamment parce qu'ils violent le principe de probabilité en proposant plusieurs issues potentielles pour une même valeur de x, affectant ainsi $L(\theta|x)$. Il est également paradoxal d'introduire de l'incertitude dans un processus décisionnel déjà sujet à l'incertitude.

Cependant, du point de vue fréquentiste des tests statistiques, les estimateurs stochastiques sont essentiels car ils permettent d'atteindre des niveaux de confiance qui seraient autrement inaccessibles. De ce fait, l'ensemble D^* est considéré comme un complément nécessaire de D. Néanmoins, cette modification de l'espace de décision n'affecte pas les décisions bayésiennes, comme le démontre le résultat suivant, où D^* désigne aussi l'ensemble des fonctions à valeurs dans D^*.

Théorème 1.1.2. *Pour toute distribution μ sur Θ, le risque de Bayes pour l'ensemble des estimateurs aléatoires est identique au risque de Bayes pour l'ensemble des estimateurs déterministes :*

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

Démonstration 1.1.1. *Pour tout $x \in X$ et tout $\delta^* \in D^*$, nous obtenons :*

$$\int_\Theta \int_D L(\theta,\gamma)\delta^*(x,\gamma)d\gamma\mu(\theta|x)d\theta$$

$$= \int_D \int_\Theta L(\theta,\gamma)\mu(\theta|x)d\theta\delta^*(x,\gamma)d\gamma$$

$$\geq \int_D \inf_\gamma \int_\Theta L(\theta,\gamma)\mu(\theta|x)d\theta\delta^*(x,\gamma)d\gamma$$

$$= \rho(\mu,\delta^\mu|x).$$

Minimaxité

Le principe de minimaxité est envisagé comme une stratégie défensive face au scénario le plus défavorable, visant à minimiser la perte espérée dans la situation la moins favorable.

Définition 1.1.3. *Le risque minimax, en lien avec une fonction de perte L, est défini par la valeur :*

$$\bar{R} = \inf_{\delta\in D^*}\sup_{\theta\in\Theta} R(\theta,\delta) = \inf_{\delta\in D^*}\sup_{\theta\in\Theta} \mathbb{E}_\theta[L(\theta,\delta(x))]$$

(c'est-à-dire, le minimum sur δ appartenant à D^* du maximum sur θ de la fonction de risque), et un estimateur minimax est tout estimateur δ_0 (potentiellement aléatoire) tel que :

$$\sup_{\theta\in\Theta} R(\theta,\delta) = \bar{R}.$$

La théorie des jeux soutient cette idée, mettant en scène des adversaires, ici incarnés par le statisticien et la nature, dans une lutte stratégique. Le statisticien établit une procédure, et en réponse, la nature choisit l'état qui augmente au maximum la perte du statisticien. Ce processus est souvent comparable à la sélection d'une distribution a priori μ, ce qui signifie que l'approche bayésienne s'écarte de cette dynamique conflictuelle, car la distribution a priori est supposée être connue. Adopter une telle perspective antagoniste en statistique est généralement contre-productif, car cela conduit à envisager la nature comme un adversaire, favorisant ainsi une préparation pour les scénarios les plus défavorables et négligeant l'exploitation des données disponibles. Pour une discussion sur la minimaxité, voir les travaux de **Brown (1993)** et **Strawderman (2000)**.

La notion de minimaxité reflète la prudence inhérente à l'approche fréquentiste. En rejetant toute supposition sur le paramètre θ, elle doit traiter les cas extrêmes comme étant tout aussi probables, se focalisant ainsi sur le risque le plus élevé. D'un point de vue bayésien, cela revient souvent à considérer une priorité concentrée sur ces cas extrêmes. Cette perspective est fréquemment trop prudente, car certaines valeurs du paramètre sont moins probables que d'autres.

Concernant les règles minimax et la stratégie maximin, il est important de noter qu'un estimateur minimax n'est pas toujours garanti. **Ferguson (1967)** et **Berger (1985, chapitre 5)** ont établi des conditions suffisantes pour leur existence. Notamment, une stratégie minimax est possible lorsque l'ensemble Θ est fini et que la fonction de perte est continue. De manière plus générale, **Brown (1976)**, ainsi que Le **Cam (1986)** et **Strasser (1985)**, envisagent l'espace de décision D comme inclus dans un ensemble plus vaste. Avec des hypothèses additionnelles, il devient alors faisable de déduire des estimateurs minimax pour des fonctions de perte continues. Toutefois, ces généralisations nécessitent des techniques topologiques complexes qui dépassent le cadre de cette thèse. Ainsi, nous nous référons au résultat suivant, dont la preuve est fournie par **Blackwell et Girshick (1954)**.

Théorème 1.1.3. *Pour un ensemble compact $D \subset \mathbb{R}^k$ et une fonction de perte $L(\theta, d)$ continue et convexe en d pour tout $\theta \in \Theta$, il existe un estimateur minimax déterministe.*

La préférence pour les estimateurs déterministes lorsque la fonction de perte est convexe découle de l'inégalité de Jensen, car :

$$L(\theta, \delta^*) = \mathbb{E}^{\delta^*}\left[L(\theta, \mathbb{E}^{\delta^*}(\delta))\right].$$

Cette affirmation est une application spécifique du théorème de **Rao-Blackwell** (référence à **Lehmann & Casella (1998)**). Le lemme suivant met en évidence la connexion entre l'approche bayésienne et le principe minimax.

Lemme 1.1.1. *Le risque bayésien est toujours inférieur au risque minimax :*

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

La première quantité est désignée comme le risque maximin et la distribution μ^* telle que $r(\mu^*) = \underline{R}$ est nommée la distribution la moins favorable, lorsqu'elle existe. Cela implique le risque de Bayes le plus élevé et donc les distributions les moins avantageuses en termes de performance de perte si elles ne sont pas suggérées par les informations préalablement disponibles. Ce constat est cohérent puisque les informations existantes ne peuvent qu'améliorer l'estimation des erreurs, même dans le pire des cas.

Un cas d'intérêt particulier est défini comme suit :

Définition 1.1.4. *On considère qu'un problème d'estimation admet une solution lorsque :*

$$\underline{R} = \overline{R}$$

et donc lorsque :

$$\sup_{\mu} \inf_{\delta \in D} r(\mu, \delta) = \inf_{\delta \in D^*} \sup_{\theta} R(\theta, \delta).$$

Quand le problème possède une valeur, il est possible d'obtenir des estimateurs minimax en calculant les estimateurs de Bayes correspondant aux distributions a priori les moins favorables. Le lemme suivant fournit des conditions suffisantes pour la minimaxité.

Lemme 1.1.2. *Considérons δ_0 comme un estimateur bayésien par rapport à μ_0. Si pour tout θ dans le support de μ_0, nous avons $R(\theta, \delta_0) \leq r(\mu_0)$, alors δ_0 est un estimateur minimax et μ_0 est considérée comme la distribution la moins favorable.*

L'estimateur minimax est souvent assimilé à un estimateur de Bayes généralisé, particulièrement lorsque le risque bayésien est infini. Dans ce cas, l'estimateur de Bayes généralisé est défini comme celui qui minimise la perte a posteriori pour chaque observation x. Pour prouver la minimaxité, il est fréquemment nécessaire de recourir à une approche asymptotique plutôt que de calculer directement le risque bayésien en s'appuyant sur le lemme précédent.

Lemme 1.1.3. *Si nous avons une suite (μ_n) de distributions propres pour lesquelles l'estimateur généralisé de Bayes δ_0 satisfait l'inégalité $R(\theta, \delta_0) \leq \lim_{n\to\infty} r(\mu_n) < +\infty$ pour tout $\theta \in \Theta$, alors δ_0 est minimax.*

Admissibilité

Le critère fréquentiste de l'admissibilité établit un classement partiel parmi les estimateurs δ dans D^* en comparant leurs risques $R(\theta, \delta)$.

Définition 1.1.5. *Un estimateur δ_0 est dit inadmissible s'il est dominé par un autre estimateur δ_1, c'est-à-dire que pour tout θ,*

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

et qu'il existe au moins un θ_0 tel que

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

Dans le cas contraire, δ_0 est considéré comme admissible.

L'admissibilité est un concept réducteur : en théorie, les estimateurs inadmissibles ne devraient pas être retenus car ils sont susceptibles d'amélioration. Cependant, l'admissibilité à elle seule ne suffit pas pour justifier l'utilisation d'un estimateur. Par exemple, un estimateur constant $\delta(x) = \theta_0$ est souvent admissible. Il est donc crucial, du point de vue fréquentiste, de trouver un estimateur qui soit à la fois minimax et admissible.

Proposition 1.1.1. *Si un estimateur minimax est unique, alors il est admissible.*

Démonstration 1.1.2. *Si δ^* est le seul estimateur minimax, alors pour tout autre estimateur $\bar{\delta} \neq \delta^*$,*

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

Par conséquent, $\bar{\delta}$ ne peut pas dominer δ^*.

Lorsque la fonction de perte L est strictement convexe, nous avons la proposition suivante :

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

Démonstration : Pour $\delta_0 \in \Theta$,

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

Si un autre estimateur δ_1 existe tel que

$$\bar{R} \leq \sup_\theta R(\theta, \delta_1) < R(\theta_0, \delta_0),$$

alors δ_0 ne peut pas être admissible.

Nous avons précédemment établi que certains estimateurs minimax sont également des estimateurs bayésiens. L'admissibilité est encore plus étroitement liée au paradigme bayésien, car dans la plupart des problèmes statistiques, les estimateurs bayésiens constituent la classe des estimateurs admissibles.

Proposition 1.1.3. *Si une distribution a priori μ est strictement positive sur Θ, avec un risque bayésien fini, et si la fonction de risque $R(\theta, \delta)$ est continue en θ pour chaque δ, alors l'estimateur bayésien δ^μ est admissible.*

Démonstration 1.1.3. *Supposons que δ^μ soit inadmissible et qu'il existe un estima-*

teur δ' qui le domine. Alors pour tout θ,

$$R(\theta, \delta') \leq R(\theta, \delta^{\mu}),$$

et il existe un ensemble ouvert $C \subset \Theta$ tel que

$$R(\theta, \delta') < R(\theta, \delta^{\mu}).$$

En intégrant cette inégalité, on obtient

$$r(\mu, \delta') < r(\mu, \delta^{\mu}) = \int_{\Theta} R(\theta, \delta^{\mu})\mu(\theta)d\theta,$$

ce qui est impossible car le risque minimax est toujours supérieur au risque bayésien....

1.2 Proposition et démonstration

Considérons que si δ^* représente l'unique estimateur minimax, alors pour tout autre estimateur $\overline{\delta}$ différent de δ^*, nous avons l'inégalité suivante :

$$\sup_{\Theta} R(\Theta, \delta^*) > \sup_{\Theta} R(\Theta, \overline{\delta}).$$

Cela implique que $\overline{\delta}$ ne peut pas dominer δ^*.

Dans le cas où la fonction de perte L est strictement convexe, nous pouvons énoncer la proposition suivante :

Proposition 1.2.1. *Si δ_0 est un estimateur admissible présentant un risque constant,* *alors δ_0 constitue le seul estimateur minimax.*

Démonstration 1.2.1. *Pour un estimateur δ_0 appartenant à Θ, il s'ensuit que :*

$$\sup_{\Theta} R(\Theta, \delta_0) > \sup_{\Theta} R(\Theta_0, \delta_0).$$

Par conséquent, il existe un estimateur δ_1 tel que $\overline{R} \leq \sup_{\Theta} R(\Theta, \delta_1) < R(\Theta_0, \delta_0)$, ce qui signifie que δ_0 ne peut être considéré comme admissible.

Il a été démontré auparavant que certains estimateurs minimax sont également des estimateurs bayésiens. L'admissibilité est étroitement liée au paradigme bayésien, car dans la majorité des problèmes statistiques, les estimateurs bayésiens génèrent la classe des estimateurs admissibles. Ces derniers peuvent être représentés par des estimateurs bayésiens, des estimateurs bayésiens généralisés ou des limites d'estimateurs bayésiens.

Nous présentons ici des propositions significatives :

Proposition 1.2.2. *Si une distribution a priori μ est strictement positive sur Θ, possède un risque bayésien fini et si la fonction de risque $R(\theta, \delta)$ est une fonction de θ pour chaque δ, alors l'estimateur bayésien δ^{μ} est admissible.*

Démonstration 1.2.2. *Supposons que δ^{μ} soit inadmissible et qu'il existe un estimateur δ' qui le domine uniformément. Pour chaque θ, nous avons $R(\theta, \delta') \leq R(\theta, \delta^{\mu})$ et dans un ensemble ouvert C de Θ, $R(\theta, \delta') < R(\theta, \delta^{\mu})$. En intégrant cette inégalité, nous obtenons :*

$$r(\mu, \delta') < r(\mu, \delta^{\mu}) = \int_{\Theta} R(\theta, \delta^{\mu}) \mu(\theta) d\theta,$$

ce qui est contradictoire, car le risque bayésien est toujours inférieur au risque minimax.

Proposition 1.2.3. *Si l'estimateur bayésien associé à la distribution a priori μ est* *unique, alors il est admissible. Même si l'estimateur bayésien n'est pas unique, il est toujours possible de trouver au moins un estimateur bayésien admissible.*

Proposition 1.2.4. *Si un estimateur bayésien δ^μ, associé à une distribution a priori μ, présente un risque bayésien*

$$r(\mu) = \int_\Theta R(\theta, \delta^\mu)\mu(\theta)d\theta$$

fini, alors δ^μ est admissible.

1.3 Les Fonctions de Perte en Actuariat

1.3.1 La Fonction de Perte Quadratique

La fonction de perte quadratique, introduite par **Legendre (1805)** et **Gauss (1810)**, est donnée par :

$$L(\theta, d) = (\theta - d)^2 \tag{1.6}$$

Une variante est la fonction de perte quadratique pondérée :

$$L(\theta, d) = \omega(\theta)(\theta - d)^2 \tag{1.7}$$

Sous l'hypothèse d'un coût quadratique, l'estimateur de Bayes associé à la loi a priori μ est la moyenne a posteriori de θ :

$$\delta^\mu(x) = \mathbb{E}_{\mu(\cdot|x)}(\theta) = \int_{\theta\in\Theta} L(\theta, \delta(x))\mu(\cdot|\theta)d\theta \tag{1.8}$$

Preuve. L'estimateur de Bayes minimise le coût a posteriori, soit $\rho(\mu, \delta) = \mathbb{E}^{\mu(\cdot|\theta)}(L(\theta, \delta(x)))$.

Avec un coût quadratique, on a :

$$\rho(\mu, \delta) = \mathbb{E}^{\mu(\cdot|\theta)}((\theta - \delta(x))^2) = \mathbb{E}^{\mu(\cdot|\theta)}(\theta)^2 - 2\delta(x)\mathbb{E}^{\mu(\cdot|\theta)}(\theta) + \delta(x)^2$$

Ce qui est un polynôme du second degré en $\delta(x)$ et est minimal pour $\mathbb{E}^{\mu(\cdot|\theta)}(\theta)$. $\square$

1.3.2 La Fonction de perte Linex

La fonction Linex est une alternative asymétrique à la fonction de perte quadratique, utile lorsque les coûts des erreurs de sous-estimation et de surestimation ne sont pas équivalents. Elle est exprimée comme :

$$L(y, \hat{y}) = e^{c(y-\hat{y})} - c(y - \hat{y}) - 1 \tag{1.9}$$

où c est un paramètre qui détermine le degré d'asymétrie.

La fonction de perte Linex, introduite par Varian en 1975, est une fonction de perte asymétrique. Elle est définie par :

$$L(\Omega) \propto e^{\gamma\Omega} - \gamma\Omega - 1$$

où $\Omega = (\theta, \delta(x))$ et $\delta(x)$ est un estimateur de θ.

Pour γ proche de zéro, la fonction de perte Linex se rapproche de la fonction de perte quadratique :

$$\mathbb{E}_\theta(L(\delta(x), \theta)) \propto e^{\gamma\delta(x)}\mathbb{E}_\theta(e^{-\gamma\theta}) - \gamma\delta(x) - \mathbb{E}_\theta(\theta) - 1 \tag{1.10}$$

où $\mathbb{E}_\theta(.)$ signifie l'espérance a posteriori de la densité aposteriori de θ.

Pour trouver l'estimateur de Bayes $\delta_\mu(x)$ qui minimise l'équation (1.10), nous dérivons l'équation (1.10) par rapport à $\delta(x)$. nous obtenons

$$\frac{d}{d\delta(x)} E_\theta(L(\delta(x),\theta)) = \gamma \exp(\gamma\delta(x)) E_\theta(\exp(-\gamma\theta)) - \gamma \tag{1.11}$$

En mettant l'équation (1.11) égale à zéro, nous obtenons

$$\exp(-\gamma\delta(x)) E_\theta(\exp(-\gamma\theta)) = \gamma \tag{1.12}$$

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

$$\delta(x) = -\frac{1}{\gamma} \log \mathbb{E}_\theta(e^{-\gamma\theta})$$

à condition que $\mathbb{E}_\theta(e^{-\gamma\theta})$ existe et soit fini.

La Fonction de Perte d'Entropie

La fonction de perte d'entropie est particulièrement adaptée aux problèmes de classification et de prévision probabiliste. Elle est définie comme :

$$L(p,q) = -\sum_i p_i \log(q_i) \tag{1.13}$$

où p est la distribution de probabilité réelle et q est la distribution de probabilité prédite.

La fonction de perte d'entropie, développée par **Galabria** et **Pulcini** (**1994**), est une variante de la fonction de perte Linex. Elle est définie par :

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

et atteint son minimum quand $d = \theta$.

L'estimateur de Bayes pour un paramètre sous cette fonction de perte est :

$$\delta(x) = \left(\mathbb{E}_\theta\left(\theta^{-p}\right)\right)^{-\frac{1}{p}}$$

— Pour $p = 1$, cet estimateur correspond à l'estimateur de Bayes pour la fonction de perte quadratique pondérée $\frac{(d-\theta)^2}{\theta}$.
— Pour $p = -1$, il s'aligne avec l'estimateur de Bayes pour la fonction de perte quadratique.

D'autres types de fonctions de perte existent, tels que la fonction de perte absolue, la fonction de perte 0-1, la fonction de perte de DeGroot, et bien d'autres. Ces fonctions de perte adoptent toutes une forme classique. Cependant, dans la prochaine section, nous introduirons une variante de ces fonctions, connue sous le nom de fonction de perte pondérée équilibrée. Notre attention se portera principalement sur la fonction de perte quadratique dans un contexte équilibré.

1.4 L'utilisation de la Fonction de Perte Équilibrée Pondérée dans le but d'obtention des primes de crédibilités

Dans le domaine de l'actuariat, l'ajustement des primes de crédibilité est crucial. Une méthode alternative, basée sur une fonction de perte équilibrée pondérée, est explorée ici. Cette fonction est une généralisation qui englobe la fonction de perte quadratique équilibrée, souvent utilisée en actuariat. À partir de cette fonction, il est possible de dériver des primes de crédibilité en fonction d'une distribution de probabilité et d'antécédents spécifiques.

1.4.1 Motivation

La théorie de la crédibilité est un corpus de techniques quantitatives qui permettent aux assureurs d'ajuster les primes futures basées sur les performances passées. Typiquement, la prime de crédibilité est formulée comme une combinaison pondérée de la moyenne de l'échantillon et de la prime collective.

Dans les travaux de **Heilmann (1989)**, diverses primes de crédibilité ont été dérivées en utilisant la théorie de la décision statistique bayésienne et une fonction de perte quadratique d'erreur appropriée (**WLF**), définie par

$$L_1(\gamma, x) = y(x)(x-\gamma)^2$$

et en considérant différentes paires de fonctions de probabilité et de distributions a priori. Selon la forme fonctionnelle choisie pour $y(x)$, différents principes de prime peuvent être appliqués.

Par exemple, avec $y(x) = 1$ et $y(x) = e^{cx}$, $c > 0$, on obtient respectivement les principes de la prime nette et de la prime d'Esscher, comme démontré dans **Heilmann (1989)** et **Gomez et al. (2006).**

Il est reconnu, comme l'a montré **Jewell (1974)**, que pour la famille exponentielle de distributions et ses antécédents conjugués, des primes de crédibilité exactes peuvent être obtenues, y compris pour le couple Poisson-Gamma sous la prime d'Esscher, comme indiqué dans **Heilmann (1989)**. La prime de Bayes peut alors être exprimée sous la forme d'une formule de crédibilité :

$$P_B^{L_1} = M(t)g(\overline{X}) + [1 - M(t)]P_C^{L_1} \tag{1.14}$$

où $P_B^{L_1}$ et $P_C^{L_1}$ représentent respectivement la prime de Bayes et la prime collective selon le cadre de **WLF** (**Heilmann (1989)** et **Gomez et al. (2006).**), et $\varepsilon(\overline{X})$ est une fonction des données observées.

Dans ce travail, les généralisations des primes de crédibilité sont obtenues en utilisant la fonction de perte équilibrée pondérée générale (**WBLF**), introduite comme suit :

$$L_2(\gamma, x) = \omega y(x)(\delta_0(x) - \gamma)^2 + (1 - \omega)y(x)(x - \gamma)^2 \tag{1.15}$$

Ici, ω est un facteur de pondération compris entre 0 et 1, déterminé par l'actuaire, $y(x)$ est une fonction de pondération positive, et $\delta_0(x)$ est une fonction des données observées.

La **WBLF** est une extension de la fonction de perte proposée par **Zellner (1994)**, et apparaît également dans les travaux de **Dey et al. (1999)**, **Farsipour et Asghar-**

zadhe (2004), et **Jafari et al. (2006)**, en prenant $y(x) = 1$. Cette fonction inclut le **WLF** comme cas particulier lorsque ω est égal à 0. De plus, l'utilisation de la **WBLF** permet également de généraliser l'expression de crédibilité de **Bühlmann (1967)** dans une approche non-paramétrique.

1.4.2 Méthodologie

Dans le cadre de cette recherche, nous avons élaboré des formules innovantes de crédibilité basées sur le concept de prime nette. Ces formules ont été développées en appliquant la fonction de Weibull à la loi de probabilité dans l'équation (1.15) . Nous partons du principe que le risque individuel X est caractérisé par une densité de probabilité $f(x|\theta)$, dépendante d'un paramètre θ appartenant à l'ensemble Θ. Ce paramètre θ suit une distribution a priori avec une densité $\mu(\theta)$. Lorsqu'une observation x est réalisée, la densité a posteriori est notée $\mu^x(\theta)$.

En actuariat, pour estimer une prime de risque inconnue $P_R^L \equiv P_R^L(\theta)$, on cherche à minimiser l'espérance de perte $E_f[L(\theta, P)]$ pour une fonction de perte donnée L. En absence de données expérimentales, l'actuaire utilise la prime collective P_C^L, obtenue en minimisant l'espérance du risque, soit $E_\pi[L(P_R^L(\theta), P_C^L)]$.

Lorsque des données expérimentales sont disponibles, la prime de Bayes est calculée en suivant la même méthode que pour la prime collective, mais en remplaçant la distribution a priori par la distribution a posteriori. La proposition suivante étend le Lemme 2 de **Jafari et al. (2006)**, permettant d'obtenir l'estimateur de Bayes pour θ sous l'hypothèse a priori μ.

Proposition 1.4.1. *Sous l'hypothèse de la fonction de Weibull et des distributions a priori μ, les primes de risque et collectives sont définies comme suit :*

$$P_R^{(L_2)} \equiv P_R^{(L_2)}(\theta) = \omega \left(\frac{E_{f(x|\theta)}[\delta_0(x)y(x)|\theta]}{E_{f(x|\theta)}[y(x)|\theta]} \right) + (1-\omega) \left(\frac{E_{f(x|\theta)}[Xy(x)|\theta]}{E_{f(x|\theta)}[y(x)|\theta]} \right) \quad (1.16)$$

$$P_C^{(L_2)} = \omega\delta_0^* + (1-\omega) \left(\frac{E_\mu[P_R^{(L_2)}y(P_R^{(L_2)})]}{E_\mu[y(P_R^{(L_2)})]} \right) \quad (1.17)$$

où δ_0^* est une estimation ciblée de la prime de risque $P_R^{(L_2)}$.

Preuve. La démonstration repose sur la minimisation de $E_{f(x|\theta)}[L_2(\theta, P_R^{(L_2)})]$ et $E_{\mu(\theta)}[L_2(P_R^{(L_2)}, P_C^{(L_2)})]$ par rapport à $P_R^{(L_2)}$ et $P_C^{(L_2)}$ afin d'obtenir respectivement la prime de risque et la prime collective. □

La prime de Bayes $P_B^{(L_2)}$ est ensuite dérivée en substituant dans l'équation (1.17) la distribution $\mu(\theta)$ par $\mu^x(\theta)$.

En posant $\beta = \frac{E_\mu[P_R^{(L_2)}y(P_R^{(L_2)})]}{E_\mu[y(P_R^{(L_2)})]}$, on a :

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

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

Ce résultat reste valable lorsque l'on remplace C par B. Ainsi, l'actuaire peut sélectionner la valeur de δ_0^* pour fixer la prime selon ses préférences.

Considérons la prime nette de crédibilité dans le contexte de la fonction de perte WBLF, où nous posons $y(x) = 1$. La proposition reformulée est la suivante :

Proposition 1.4.2. *Si la prime nette de Bayes calculée avec $L_1(\gamma, x)$ est une formule de crédibilité, alors la prime nette équilibrée de Bayes obtenue avec la WBLF est aussi une formule de crédibilité, exprimée par :*

$$P_B^{(L_2)} = M(t) \cdot l(P_c^{(L_1)}) + [1 - M(t)] \cdot l(\overline{X}),$$

où $M(t) \in [0,1]$ *et*

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

Preuve. En appliquant les équations (1.16), (1.17) avec $h(x) = 1$, nous obtenons :

$$P_R^{(L_2)} = \omega E[\delta_0(x|\theta)] + (1-\omega) E_{f(x|\theta)}(x|\theta),$$

et

$$P_c^{L_2} = \omega \delta_0^* + (1-\omega) E_\pi(\theta) \left[\omega E_{f(x|\theta)} \left[\delta_0(x \mid \theta) \right] \right] + (1-\omega) E_{f(x|\theta)}(x \mid \theta)$$

$$= \omega \delta_0^* + \omega(1-\omega) E_\mu(\theta) \left[\omega E_{f(x|\theta)} \left[\delta_0(x|\theta) \right] \right] + (1-\omega)^2 E_\mu(\theta) \left[E_{f(x|\theta)}(x|\theta) \right].$$

Par conséquent, nous avons :

$$P_B^{(L_2)} = \omega \delta_0^* + \omega(1-\omega) E_\mu(\theta) \left[E_{f(x|\theta)} \left[\delta_0(x|\theta) \right] \right] + (1-\omega)^2 P_B^{(L_1)}.$$

Maintenant, si $P_B^{(L_1)}$ est une formule de crédibilité de la forme

$$P_B^{(L_1)} = M(t) P_c^{(L_1)} + [1 - M(t)] \overline{X},$$

alors nous obtenons :

$$P_B^{L_2} = \delta_0^* + \omega(1-\omega) E_\mu(\theta) \left[E_{f(x|\theta)} \left[\delta_0(x \mid \theta) \right] + (1-\omega)^2 \left(M(t) P_c^{L_1} + [1 - M(t)] \bar{X} \right) \right]$$

$$= M(t) \left[(1-\omega)^2 P_c^{L_1} + \omega(1-\omega) E_\mu^x(\theta) \left[E_{f(x|\theta)} \left[\delta_0(x \mid \theta) \right] \right] \right]$$

$$+(1-M(t))\left[(1-\omega)^2\bar{X}+\omega\delta_0^*+\omega(1-\omega)E_\mu^x(\theta)\left[E_{f(x|\theta)}\left[\delta_0(x\mid\theta)\right]\right]\right]$$

$$=M(t)l(P_c^{L_1})+[1-M(t)]l(\bar{X})$$

□

Corollaire 1.4.1. *Si nous considérons que la famille exponentielle de distribution est la fonction de probabilité et son conjugué antérieur, alors la prime nette équilibrée de Bayes est une formule de crédibilité exacte.*

Preuve. Cette preuve est simplifiée en tenant compte du fait que dans la famille exponentielle des distributions, la prime nette de Bayes obtenue avec la perte $L(x,\gamma)=(x-\gamma)^2$ est une expression de crédibilité semblable à l'équation (1.14). □

Remarque 1.4.1. *Le facteur de crédibilité de cette proposition est identique à celui obtenu avec la fonction de perte $L(x,\gamma)=(x-\gamma)^2$.*

Corollaire 1.4.2. *Les primes nettes équilibrées pondérées de Bayes constituent une formule de crédibilité lorsqu'on utilise les paires Poisson-Gamma, négative binomial-beta, binomial-beta, normal-normal, et gamma-gamma.*

Preuve. Cette preuve est dérivée en appliquant le corollaire précédent aux paires mentionnées. □

1.4.3 Approche Non Paramétrique

Dans l'approche non-paramétrique, il a été établi que la prime de crédibilité linéaire dans le cadre classique correspond à une combinaison linéaire de la sinistralité moyenne observée dans un contrat individuel et celle de l'ensemble du portefeuille, comme l'ont démontré **Bühlmann (1967)** et **Jewell (1974)**. Gomez-Deniz a proposé une méthode innovante consistant à remplacer la prime de crédibilité exacte $H(\mu(\theta)|X_1, X_2, \ldots, X_t)$ par une fonction linéaire de la forme $c_0 + \sum_{s=1}^{t} c_s X_s$, basée sur les sinistres passés X_i, pour $i = 1, 2, \ldots, t$, en appliquant la méthode WBLF. On suppose que les variables $X_1|\theta, X_2|\theta, \ldots, X_t|\theta$ sont indépendantes et identiquement distribuées. Pour simplifier, nous introduisons les notations suivantes :

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

$$b = \omega\delta_0^* + (1-\omega)\{\omega E[\delta_0(x|\theta)] + (1-\omega)E[X_s|\theta]\},$$

$$\phi = \mathrm{Var}[E[X_s|\theta]],$$

$$\rho^2 = E[\mathrm{Var}[X_s|\theta]],$$

conformément aux conventions de la littérature spécialisée (voir **Herzog (1996)** et **Goovaerts et al (1990)**).

Les coefficients c_i, pour $i = 1, \ldots, t$, sont déterminés en minimisant l'expression suivante :

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

En annulant la dérivée par rapport à c_i, on obtient le système d'équations suivant :

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

$$c_0E[X_r] = \omega E[X_r\delta_0(x)] + (1-\omega)E[X_r\eta(\theta)] - \sum_{s=1}^{t} c_s E[X_rX_s], \quad \text{pour } r = 1, \ldots, t.$$

qui est équivalent au système

$$\sum_{s=1}^{t} c_s \text{cov}(X_r, X_s) = (1-\omega)\text{cov}[X_r, \eta(\theta)] + \omega\text{cov}[X_r, \delta_0(X)], \quad r = 1, \ldots, t. \tag{1.18}$$

En tenant compte du fait que :

$$\text{cov}(X_rX_s) = E[\text{cov}(X_r, X_s|\theta)] + \text{cov}(E[X_r|\theta], E[X_s|\theta])$$

$$= E(0) + \text{Var}\left[\frac{\eta(\theta) - E[\delta_0(x)]}{1-\omega}\right]$$

$$= \text{Var}[\eta^*(\theta)] = \phi^*,$$

où

$$\eta^*(\theta) = \frac{\eta(\theta) - E[\delta_0(x)]}{1-\omega},$$

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

et

$$\text{cov}(X_r, \eta(\theta)) = \text{cov}(X_r, \eta(\theta)|\theta) + \text{cov}(E[X_r|\theta], E[X_r, \eta(\theta)|\theta])$$

$$= E(0) + \text{cov}\left(E[X_r \mid \theta], E[X_r, \eta(\theta) \mid \theta]\right)$$

$$= \frac{\text{Var}[\eta(\theta)] - \omega\text{cov}(\delta_0(X), \eta(\theta))}{1-\omega}.$$

Le système en (1.18) se réduit à :

$$c\rho^2 + \phi^* c_t = \text{Var}[\eta(\theta)] - \omega\text{cov}(\delta_0(X), \eta(\theta)) + \omega\text{cov}(X_r, \delta_0(X)),$$

d'où l'on déduit :

$$c = \frac{\mathrm{Var}[\eta(\theta)] + \omega\{\omega \mathrm{cov}(X_r, \delta_0^*(X)) - \omega \mathrm{cov}(\delta_0(X), \eta(\theta))\}}{\rho^2 + \phi^* c_t},$$

et par conséquent :

$$H(\eta(\theta)|X_1, X_2, \ldots, X_t) = c_0 + c_t\bar{X} = \omega E[\delta_0(x)] + (1-\omega)E[\eta(\theta)] - c_t E[X_s] + c_t\bar{X}.$$

$$= \omega E[\delta_0(x)] + (1-\omega)E[\omega E[\delta_0(x)](1-\omega)E[X \mid \theta] - ctE[X] + ct\overline{X}] \tag{1.19}$$

En supposant que $\delta_0^* = E[\delta_0(x)]$, l'expression (1.18), l'expression (1.19) peut être réécrite comme suit, avec $c^* = \frac{c}{1-\omega}$:

$$H(\eta(\theta)|X_1, X_2, \ldots, X_t) = b - \frac{b - \omega\delta_0^*}{(1-\omega)^2} - \frac{\omega c_t E[E[\delta_0(x|\theta)]]}{1-\omega} + c_t\bar{X},$$

$$= b - c^*(b - \omega\delta_0^*)t + \omega(1-\omega)E[E[\delta_0(x|\theta)]]tc^* + tc^*(1-\omega)^2\bar{X},$$

$$= (1 - c^*t)b + \{\omega\delta_0^* + \omega(1-\omega)E[E[\delta_0(x|\theta)]] + (1-\omega)^2\bar{X}\}c^*t.$$

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

$$H(\eta(\theta)|X_1, X_2, \ldots, X_t) = M(t)l(\bar{X}) + (1 - M(t))l(P_c^{L_2}).$$

Remarque 1.4.2. *Si $\omega = 0$, alors $\eta^*(\theta) = \eta(\theta)$, $\phi^* = \phi$ et $M(t) = \frac{\phi}{\phi(\rho^2+\phi t)}$, ce qui correspond au facteur de crédibilité traditionnel obtenu avec la méthode **WLF**.*

Ce chapitre a exploré les fondements théoriques et pratiques des fonctions de perte et leur application dans le calcul des primes de crédibilité en actuariat. Ces outils permettent aux actuaires de prendre des décisions éclairées et de développer des stratégies de tarification qui sont à la fois équitables et économiquement viables.

Chapitre 2

La Méthode de l'Entropie Maximale

Ce chapitre se consacre à l'examen approfondi de la méthode de l'entropie maximale, un élément fondamental de la théorie de l'information qui évalue l'incertitude associée à un événement. Nous commençons par explorer les complexités de l'entropie, puis nous passons à une explication détaillée de la méthode de l'entropie maximale.

2.1 Entropie : Le Cœur de la Théorie de l'Information

L'entropie, dans l'univers de la théorie de l'information, est la mesure moyenne d'information, de surprise ou d'incertitude qui émane des issues potentielles d'une variable aléatoire. Cette notion a été révélée au monde par Claude Shannon dans son œuvre révolutionnaire, "A Mathematical Theory of Communication" (**Shannon (1948)**), et est souvent honorée du nom d'entropie de Shannon.

Imaginons une pièce de monnaie truquée avec une probabilité p de montrer pile et $1-p$ de montrer face. L'incertitude atteint son apogée lorsque $p = \frac{1}{2}$, car aucun résultat n'est favorisé, conférant à chaque lancer une entropie d'un bit. À l'inverse, l'incertitude est nul lorsque $p = 0$ ou $p = 1$, car l'issue est certaine et l'entropie tombe à zéro. Des valeurs intermédiaires de p produisent des entropies variant entre ces deux extrêmes.

Formellement, pour une variable aléatoire discrète X, avec des issues $x_1, x_2, \ldots, x_n$ ayant des probabilités $P(x_1), P(x_2), \ldots, P(x_n)$, l'entropie de X est définie comme suit :

$$Y(X) = -\sum_{i=1}^{n} P(x_i) \log_b P(x_i) \tag{2.1}$$

La base du logarithme (b) peut varier selon l'application. La base 2 est mesurée en bits (ou Shannons), la base e en nats, et la base 10 en 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 est également utilisée dans la théorie de l'information de Shannon, où elle représente une limite mathématique sur la compression sans perte des données d'une source sur un canal parfaitement silencieux.

Dans le contexte de la théorie de l'information, l'entropie est un concept qui trouve son analogue dans l'entropie de la thermodynamique statistique. L'entropie est également pertinente dans d'autres branches des mathématiques, comme la combinatoire. On peut dériver sa définition à partir d'un ensemble d'axiomes qui stipulent que l'entropie doit quantifier l'incertitude moyenne résultant d'une variable.

Pour une variable aléatoire continue, l'entropie différentielle est un concept similaire à l'entropie.

Définition

En s'inspirant du théorème de Boltzmann, Shannon a introduit l'entropie Y d'une variable aléatoire discrète X avec des valeurs possibles $X_1, X_2, ..., X_n$ et la fonction de masse de loi $P(x)$ comme suit :

$$Y(X) = E[I(X)] = E[-\log(P(x))]$$

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

L'entropie peut être explicitement exprimée comme :

$$Y(X) = -\sum_{i=1}^{n} P(x_i) \log_z P(x_i)$$

où z est la base du logarithme utilisé. Les bases couramment utilisées sont 2, le nombre d'Euler e et 10, et les unités d'entropie correspondantes sont les bits pour $z = 2$, les *nats* pour $z = e$ et les *bans* pour $z = 10$ **(Schneider, 2007)**.

Dans le cas où $P(x_i) = 0$ pour un certain i, la valeur de la somme correspondante $0 \log_{b(0)}$ est considérée comme 0, ce qui est cohérent avec la limite :

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

2.2 Entropie Conditionnelle et Normalisée

L'entropie conditionnelle de deux événements X et T est définie comme le degré d'aléatoire de X par rapport à T :

$$H(X|T) = -\sum_{i,j} P(x_i, t_j) \log \left(\frac{P(x_i, t_j)}{P(t_j)} \right)$$

où $P(x_i, t_j)$ est la probabilité conjointe que $X = x_i$ et $T = t_i$.

L'entropie peut être normalisée par la longueur de l'information, donnant lieu à l'entropie métrique, un indicateur du caractère aléatoire de l'information.

En résumé, l'entropie est un concept fondamental dans la théorie de l'information, et elle joue un rôle essentiel dans la compréhension de l'incertitude et de la variabilité des résultats possibles d'une variable aléatoire.

Considérons l'exemple d'un lancer de pièce avec des probabilités connues pour obtenir pile ou face, qui peut être représenté par un processus de Bernoulli. L'incertitude quant au résultat d'un lancer est à son maximum lorsque la pièce est équilibrée, c'est-à-dire que les probabilités pour pile et face sont égales à $\frac{1}{2}$. Cette situation représente le degré d'incertitude le plus élevé, rendant la prédiction du résultat suivant particulièrement difficile. Chaque lancer fournit alors une information complète. Mathématiquement, cela s'exprime par l'entropie $Y(X)$:

$$Y(X) = -\sum_{i=1}^{n} P(x_i) \log_z P(x_i) = -\sum_{i=1}^{2} P\left(\frac{1}{2}\right) \log_2 P\left(\frac{1}{2}\right) = -\sum_{i=1}^{2} \left(\frac{1}{2}\right) \log_2 \left(\frac{1}{2}\right) = 1$$

Si la pièce est déséquilibrée avec des probabilités p et q pour pile et face respectivement, où $p \neq q$, l'incertitude diminue. L'entropie, qui quantifie cette incertitude réduite, est alors inférieure à 1. Par exemple, si $p = 0.7$, l'entropie $Y(x)$ est calculée comme suit :

$$Y(x) = -p \log_2(p) - q \log_2(q) \approx -0.7 \log_2(0.7) - 0.3 \log_2(0.3) \approx 0.8816 < 1$$

Une distribution uniforme des probabilités entraîne une incertitude maximale et, par conséquent, une entropie maximale. L'entropie ne peut que diminuer à partir de la valeur associée à une distribution uniforme. Dans le cas extrême où une pièce ne montre jamais pile ou jamais face, il n'y a aucune incertitude et l'entropie est nulle, chaque lancer produisant un résultat certain.

L'entropie peut être normalisée en la divisant par la longueur de l'information, ce ratio étant appelé entropie métrique, qui mesure le degré d'aléatoire de l'information.

2.3 La Méthode de l'Entropie Maximale

La méthode de l'entropie maximale est un principe statistique qui postule que, parmi toutes les distributions de probabilité possibles, celle qui reflète le mieux notre état de connaissance sans introduire d'informations supplémentaires est celle qui maximise l'entropie. Cette méthode est particulièrement pertinente lorsque les données a priori sont exprimées sous forme de contraintes précises, comme des informations testables ou des propositions vérifiables.

2.3.1 Historique et Justification Théorique

Introduite par **E.T. Jaynes (1957)**, la méthode de l'entropie maximale établit un lien entre la mécanique statistique et la théorie de l'information. Jaynes a avancé que l'entropie en mécanique statistique et l'entropie en théorie de l'information sont conceptuellement identiques, ce qui implique que la mécanique statistique est une application spécifique d'un cadre général d'inférence logique basé sur la théorie de l'information.

2.3.2 Aperçu Général

Dans le domaine de la thermodynamique statistique, il est courant de se baser sur des données ou des informations vérifiables qui prennent la forme d'un ensemble de quantités conservées, telles que les valeurs moyennes de certaines fonctions de moment, pour décrire une distribution de probabilité donnée. La méthode de l'entropie maximale, qui utilise ces données pour déterminer la distribution la plus probable, est fréquemment appliquée dans ce contexte. Parfois, on peut également définir la distribution de probabilité en imposant certaines symétries. Lorsque les quantités conservées et les symétries correspondantes sont équivalentes, cela entraîne une équivalence dans la spécification des informations vérifiables dans le cadre de cette méthode.

L'application de la méthode de l'entropie maximale est essentielle pour assurer que les probabilités attribuées soient uniques et cohérentes, peu importe la méthode utilisée, que ce soit la mécanique statistique ou l'inférence logique. Cette méthode souligne également la possibilité de choisir parmi différentes formes de données a priori. Par exemple, on peut opter pour une densité de probabilité a priori uniforme, suivant le principe d'indifférence de Laplace, aussi connu sous le nom de principe de raison insuffisante. Ainsi, la méthode de l'entropie maximale ne se contente pas d'offrir une alternative aux méthodes traditionnelles d'inférence statistique, mais elle en constitue une extension conceptuelle significative.

Cependant, cela ne signifie pas que les systèmes thermodynamiques peuvent se passer de la démonstration de leur caractère ergodique pour justifier leur analyse en tant qu'ensemble statistique. En termes simples, la méthode de l'entropie maximale adopte une position de modestie épistémologique, ou d'ignorance maximale. La distribution choisie est celle qui, tout en respectant les données a priori, prétend le moins possible à une connaissance supplémentaire, admettant ainsi le plus haut degré d'ignorance au-delà des informations fournies.

2.3.3 Informations Testables

La méthode de l'entropie maximale s'avère particulièrement pertinente lorsqu'elle est appliquée à des données empiriquement vérifiables. Ces données, ou informations testables, sont des éléments spécifiques d'une distribution de probabilité qui peuvent être clairement validés ou invalidés. Par exemple, des informations telles que "l'espérance de la variable X est 2.87" et "la somme des probabilités $p_2 + p_3$ est supérieure à 0.6" (où p_2 et p_3 représentent les probabilités de certains événements) constituent des informations testables.

Dans le cadre de ces informations, la méthode de l'entropie maximale vise à identifier la distribution de probabilité qui maximise l'entropie tout en respectant les contraintes imposées par les informations testables. Ce problème d'optimisation avec contraintes est couramment résolu via la méthode des multiplicateurs de Lagrange.

Lorsqu'il n'y a pas d'informations testables, la maximisation de l'entropie se conforme à la contrainte fondamentale que la somme des probabilités doit être égale à un. Dans ce cas, la distribution de probabilité discrète qui maximise l'entropie est la distribution uniforme, où chaque probabilité p_i est égale à $\frac{1}{n}$ pour tout i appartenant à l'ensemble $\{1, 2, \ldots, n\}$.

2.3.4 Applications

La méthode de l'entropie maximale, essentielle dans l'inférence statistique, est principalement utilisée de deux manières : pour établir des probabilités a priori dans l'inférence bayésienne, comme le préconisait **Jaynes (1968)**, et comme principe de mise à jour pour les probabilités postérieures. Cette méthode est considérée comme produisant la distribution la moins informative, une idée largement développée dans les travaux de **Soofi (2000)**, **Clarke (2006)**, **Bousquet (2008)** et **Palmieri (2013)**.

En tant que règle de mise à jour, elle est compatible avec des approches telles que la cinématique de probabilité de Jeffrey, bien que ne généralisant pas toutes les méthodes de mise à jour, comme le souligne **Skyrms (1987)**. Pour la spécification de modèles, en particulier dans le traitement du langage naturel, elle permet de considérer les données observées comme des informations testables, avec des applications notables comme la régression logistique.

L'estimation de la densité de probabilité, tant discrète que continue, est une autre application significative de cette méthode. Elle peut impliquer la résolution de problèmes de programmation quadratique et favorise l'intégration d'informations a priori dans l'estimation, offrant ainsi un modèle de mélange comme estimateur de densité optimal, une approche soutenue par les travaux de **Botev** et **Kroese (2008, 2011)** et **Kesavan et Kapur (1990)**..

2.3.5 Approche Générale pour Optimiser l'Entropie Maximale sous Contraintes Linéaires

Pour une série de contraintes linéaires, la distribution de l'entropie maximale peut être trouvée en utilisant la méthode des multiplicateurs de Lagrange, qui conduit à une distribution de type exponentiel.

Cas discret Nous examinons un ensemble de données testables sur une variable aléatoire X qui peut prendre des valeurs dans l'ensemble $\{x_1, x_2, \ldots, x_n\}$. Ces données sont représentées par m contraintes sur les valeurs attendues des fonctions f_k, impliquant que notre distribution de probabilité doit satisfaire les contraintes suivantes :

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

où les F_k sont des valeurs observées. De plus, nous imposons que la somme des probabilités soit unitaire, ce qui équivaut à une contrainte sur la fonction identité avec une valeur observée égale à 1 :

$$\sum_{i=1}^{n} P_r(x_i) = 1$$

La distribution de probabilité qui maximise l'entropie de l'information tout en respectant ces contraintes est donnée par :

$$P_r(x_i) = \frac{1}{M(\lambda_1, \ldots, \lambda_m)} \exp\left[\lambda_1 f_1(x_i) + \cdots + \lambda_m f_m(x_i)\right]$$

pour certains $\lambda_1, \ldots, \lambda_m$. Cette distribution est souvent désignée sous le nom de distribution de Gibbs. La constante de normalisation, ou fonction de partition, est définie par :

$$M(\lambda_1, \ldots, \lambda_m) = \exp\left[\lambda_1 f_1(x_i) + \cdots + \lambda_m f_m(x_i)\right],$$

et est essentielle pour assurer que la distribution reste normalisée.

Les λ_k sont connus comme les multiplicateurs de Lagrange. Pour les contraintes d'égalité, leurs valeurs sont obtenues en résolvant les équations non linéaires suivantes :

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

Dans le cas de contraintes d'inégalité, les multiplicateurs de Lagrange sont déterminés en résolvant un problème d'optimisation convexe avec des contraintes linéaires, comme décrit par **Botev et Kroese (2008)**. Dans les deux situations, il n'y a pas de solutions analytiques et le calcul des multiplicateurs de Lagrange nécessite l'utilisation de méthodes numériques.

Cas continue Dans le contexte des distributions continues, l'entropie de Shannon n'est pas applicable car elle est spécifique aux espaces de probabilité discrets. Edwin

Jaynes a proposé une formule alternative liée à l'entropie relative, aussi connue sous le nom d'entropie différentielle :

$$Y_c = -\int p(x) \log \left(\frac{p(x)}{q(x)} \right) dx$$

où $q(x)$, désignée par Jaynes comme la "mesure invariante", est proportionnelle à la densité limite des points discrets. Nous supposerons que q est connue et nous approfondirons cette discussion après avoir présenté les équations de la solution.

L'entropie relative est souvent définie comme la divergence de Kullback-Leibler entre p et q, bien que parfois elle soit définie de manière confuse comme l'opposé de cela. Le principe d'inférence qui consiste à minimiser cette divergence est connu sous le nom de principe de minimisation de l'information discriminante.

Considérons des informations testables sur une variable X prenant des valeurs dans un intervalle donné des nombres réels. Ces informations se manifestent sous forme de m contraintes sur les moments des fonctions f_k, c'est-à-dire que notre fonction de densité de probabilité doit satisfaire les contraintes de moment suivantes :

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

où les F_k sont des valeurs observables. La densité de probabilité doit également s'intégrer à un, ce qui peut être vu comme une contrainte sur la fonction identité avec une valeur observable égale à 1 :

$$\int P(x) dx = 1$$

La fonction de densité de probabilité qui maximise Y_c tout en respectant ces contraintes est :

$$P(x) = \frac{1}{M(\lambda_1, \ldots, \lambda_m)} q(x) \exp\left[\lambda_1 f_1(x) + \cdots + \lambda_m f_m(x)\right]$$

La fonction de partition $M(\lambda_1, \ldots, \lambda_m)$ est déterminée par :

$$M(\lambda_1, \ldots, \lambda_m) = \int q(x) \exp\left[\lambda_1 f_1(x) + \cdots + \lambda_m f_m(x)\right] dx$$

Dans le cas où toutes les contraintes de moment sont des égalités, les valeurs des paramètres λ_k sont déterminées par le système d'équations non linéaires suivant :

$$F_k = \frac{\partial}{\partial \lambda_k} \log M(\lambda_1, \ldots, \lambda_m)$$

Pour les contraintes de moments d'inégalité, les multiplicateurs de Lagrange sont obtenus en résolvant un problème d'optimisation convexe, comme décrit par Botev et Kroese (2011).

La mesure invariante $q(x)$ peut être interprétée en supposant que x ne prend des valeurs que dans un intervalle borné (a, b) sans autre information préalable. La fonction de densité de probabilité d'entropie maximale est alors :

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

où A est une constante de normalisation. La mesure invariante représente la densité a priori qui encode le "manque d'informations pertinentes". Elle ne peut être déterminée par le principe de l'entropie maximale et doit être établie par une autre méthode logique, telle que le principe des groupes de transformation ou la théorie de la marginalisation.

Pour référence, les fonctions spéciales suivantes sont utilisées :

- La fonction gamma $\Gamma(x)$ est définie comme :

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

- La fonction digamma $\Psi(x)$, qui est la dérivée logarithmique de la fonction gamma :

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

- La fonction bêta $B(p, q)$ est exprimée en termes de la fonction gamma :

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

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, ..., 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	$(1-p)^{k-1}p$	$E(k) = \frac{1}{P}$	$N - 0 = 1, 2, 3$
Exponentielle	$exp(-\lambda x)$	$E(x) = \frac{1}{\lambda}$	$[0, \infty)$
Laplace	$\frac{1}{2b} exp(-\frac{\lvert x-\mu \rvert}{b})$	$E(\lvert x - \mu \rvert) = b$	$(-\infty, \infty)$
Laplace asymétrique	$\frac{\lambda exp(-(x-m)\lambda s k^2)}{\frac{k+1}{k}}$	$E((x-m)sk^2) = \frac{1}{\lambda}$	$(-\infty, \infty)$
Pareto	$\frac{a x^a m}{x^{a+1}}$	$E(ln(x)) = \frac{1}{\sigma} + ln(x_m)$	$[x_m, \infty)$
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	$\frac{x}{\sigma}^2 exp(-\frac{x^2}{2\sigma^2})$	$E(x^2) = \sigma^2$, $E(ln(x)) = \frac{ln(2\sigma^2) - \gamma_E}{2}$	$[0, \infty)$
Beta	$\frac{x^{a-1}(1-x)^{\beta-1}}{B(a,\beta)}$ pour $0 \leq x \leq 1$	$Eln(x)) = \Psi(a) - \Psi(a+\beta)$, $E(ln(1-x)) = \Psi(\beta) - \Psi(a+\beta)$	$[0, 1]$
Cauchy	$\frac{1}{\pi(1+x)^2}$	$E(ln(1+x^2)) = 2ln2$	$(-\infty, \infty)$
Khi	$\frac{1}{2^{(\frac{k}{2})}\Gamma(\frac{k}{2})} x^{k-1} exp(-\frac{x^2}{2})$	$E(x^2) + k, E(ln(x))$ $= \frac{1}{2}[\Psi(\frac{k}{2}) + ln(2)]$	$[0, \infty)$
Khi carré	$\frac{1}{2^{(\frac{k}{2})}\Gamma(\frac{k}{2})} x^{k-1} exp(-\frac{x^2}{2})$	$E(x^2) + k$, $E(ln(x)) = [\Psi(\frac{k}{2}) + ln(2)]$	$[0, \infty)$
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	$\frac{exp(-\frac{1}{2}(\vec{x} - \vec{\mu}^T)\Sigma^{-1}(\vec{x} - \vec{\mu}))}{(2\pi)^{\frac{N}{2}} \lvert\Sigma\rvert^{\frac{1}{2}}}$	$E(\vec{x}) = \vec{\mu}$, $E((\vec{x} - \vec{\mu})(\vec{x} - \vec{\mu})^{\Upsilon})$ $= \Sigma$	R^n
Binomiale	$\binom{n}{k} p^k (1-p)^{n-k}$	$E(x) = \mu$, $f \in n-$distribution bionomiale généralisée	$0, ..., n$
Poisson	$\frac{\lambda^k exp(-\lambda)}{k!}$	$E(x) = \lambda$, $f \in \infty-$ distribution bionomiale généralisée	$N \cup 0$

2.3.6 Rationalisation de la méthode d'entropie maximale

Pour diverses raisons, les partisans de la méthode d'entropie maximale l'adoptent, en se reposant sur deux arguments principaux qui reposent sur les mêmes postulats que ceux de l'utilisation de la probabilité bayésienne.

L'entropie de l'information comme indicateur de l'absence d'information

Considérons une distribution de probabilité discrète qui comprend m options qui s'excluent mutuellement. La distribution qui contient le plus d'informations est celle où une seule option est identifiée comme correcte, ce qui entraîne une entropie de l'information de zéro. Par contre, la distribution qui contient le moins d'informations est celle où aucune option n'est favorisée par rapport aux autres. Dans cette situation, la seule distribution de probabilité rationnelle serait uniforme, et l'entropie de l'information atteindrait alors son maximum possible, qui est $\log m$.

L'entropie de l'information peut donc être interprétée comme une échelle numérique qui indique le degré d'absence d'informations dans une distribution de probabilité donnée, allant de zéro (complètement informative) à $\log m$ (complètement sans informations).

En choisissant la distribution qui a l'entropie maximale autorisée par nos données, l'idée est que nous optons pour la distribution qui est la moins informative. Sélectionner une distribution avec une entropie plus basse serait comme supposer que nous avons des informations supplémentaires que nous n'avons pas réellement. Par conséquent, la distribution avec l'entropie maximale est la seule distribution logique. Néanmoins, la dépendance de la solution à la mesure prédominante, symbolisée par $m(x)$, est un point critique de cette méthode, car cette mesure prédominante est finalement arbitraire (**Druilhet et Marin, 2007**).

La supposition de Wallis

L'argumentation qui va suivre est basée sur une proposition faite par **Graham Wallis** à **E.T. Jaynes** en **1962**. C'est essentiellement le même raisonnement mathématique qui est utilisé pour les statistiques de Maxwell-Boltzmann en mécanique statistique, bien que l'approche conceptuelle soit légèrement différente. Ce raisonnement a la particularité d'être entièrement basé sur la combinatoire, sans faire appel à l'entropie de l'information comme indicateur d'incertitude, de manque d'information ou tout autre concept vaguement défini. La fonction d'entropie de l'information n'est pas supposée au départ, mais elle apparaît plutôt au fil de l'argumentation. De plus, l'argumentation conduit naturellement à la procédure de maximisation de l'entropie de l'information plutôt qu'à son traitement de toute autre manière.

Présumons qu'une personne tente d'attribuer une probabilité parmi m options qui s'excluent mutuellement. Elle possède des informations qui peuvent être vérifiées, mais elle ne sait pas comment les incorporer dans son évaluation des probabilités (chacune étant de $\frac{1}{N}$) de manière aléatoire parmi les m options (on pourrait envisager qu'elle lance N balles dans m seaux tout en ayant les yeux bandés).
Dans un souci d'impartialité maximale, chaque lancer doit être indépendant des autres et chaque seau doit être de taille égale. Une fois l'expérience achevée, il vérifiera si la probabilité attribuée correspond à ses informations. (Pour que cette étape soit couronnée de succès, l'information doit être une contrainte définie par un ensemble ouvert

dans l'espace des mesures de probabilité). Si elle est incohérente, il la rejettera et recommencera le processus.

Si elle est juste, son estimation sera

$$p_i = n_i/N$$

où p_i est la probabilité associée à la i^{me} proposition, et n_i représente le nombre de quanta attribués à la i^{me} proposition (autrement dit, le nombre de balles qui ont atterri dans le seau i).

Pour affiner l'attribution de probabilité, il sera nécessaire d'utiliser un grand nombre de quanta de probabilité. Au lieu de réaliser réellement, et potentiellement de devoir répéter l'expérience aléatoire qui peut être longue, l'individu décide de calculer et d'utiliser simplement le résultat le plus probable. La probabilité d'un résultat spécifique est la distribution multinomiale,

$$P_r(p) = \omega.m^{-N}$$

où

$$\omega = \frac{N!}{n_1!n_2!...n_m!}$$

est parfois appelée la multiplicité du résultat.

L'issue la plus susceptible se présente lorsqu'on maximise la multiplicité ω. De manière équivalente, l'actuaire a la possibilité d'optimiser n'importe quelle fonction croissante monotone de ω. il choisi de maximiser

$$\frac{1}{N}\log\omega = \frac{1}{N}\log\frac{N!}{(n_1!n_2!\ldots n_m!)}$$

$$= \frac{1}{N} \log \frac{N!}{(Np_1)!(Np_2)!\ldots(Np_m)!}$$

$$= \frac{1}{N}(\log N! - \sum_{i=1}^{m} \log(Np_i)!)$$

A ce niveau, pour donner une expression plus simple, le joueur choisi la limite pour $N \longrightarrow \infty$, c'est-à-dire que les niveaux de probabilité passent de valeurs discrètes granuleuses à des valeurs continues lisses.

En utilisant l'approximation de Stirling, on trouve :

$$\lim_{n\to\infty} \frac{1}{N} \log \omega = \frac{1}{N}(N \log N - \sum_{i=1}^{m} Np_i \log Np_i)$$

$$= \log N - \sum_{i=1}^{m} Np_i \log Np_i$$

$$= \log N - \log N \sum_{i=1}^{m} p_i - \sum_{i=1}^{m} p_i \log p_i$$

$$= (1 - \sum_{i=1}^{m} p_i) \log N - \sum_{i=1}^{m} p_i \log p_i$$

$$= - \sum_{i=1}^{m} p_i \log p_i$$

$$= Y(p)$$

Au final, il ne reste à l'actuaire qu'à maximiser l'entropie sous la contrainte de ses informations testables. Il a observé que parmi toutes les distributions aléatoires 'équitables', celle qui présente l'entropie maximale est la plus susceptible d'être retenu, particulièrement lorsque les niveaux de probabilité évoluent d'un état discret à un état continu.

Relation avec le Théorème de Bayes

D'après **Giffin** et **Caticha (2007)**, il existe une compatibilité totale entre le théorème de Bayes et la méthode d'entropie maximale, ces deux concepts pouvant être considérés comme des cas particuliers de la "méthode d'entropie relative maximale". Ils affirment que cette approche reproduit intégralement les aspects des méthodes d'inférence bayésienne classiques. En outre, cette méthode innovante permet de résoudre des problèmes qui ne pourraient être abordés ni par la méthode d'entropie maximale ni par les méthodes bayésiennes traditionnelles seules. Des études récentes (**Lazar (2003)**, **Schennach (2005)**) montrent que les approches d'inférence basées sur l'entropie relative fréquentiste (comme la vraisemblance empirique et la vraisemblance empirique à inclinaison exponentielle - voir par exemple **Owen (2001)** et **Kitamura (2006)**) peuvent être fusionnées avec des informations a priori pour réaliser une analyse postérieure bayésienne.

Selon Jaynes, le théorème de Bayes sert à déterminer une probabilité, alors que le principe d'entropie maximale permet d'établir une distribution de probabilité a priori (**Interprétation de Jaynes, 1988**).

Il est possible de déterminer directement une distribution postérieure à partir d'une distribution a priori spécifiée en utilisant la méthode de minimisation de l'entropie croisée (ou la méthode de maximisation de l'entropie comme un cas particulier lorsqu'une distribution uniforme est utilisée comme a priori), sans tenir compte des considérations bayésiennes et en traitant le problème comme un problème d'optimisation avec contraintes, où la fonction d'entropie est la fonction objectif. Lorsque des valeurs moyennes sont fournies comme informations vérifiables (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 minimiser l'entropie croisée et satisfaire les informations vérifiables fournies.

Chapitre 3

Sur la prime de crédibilité bayésienne : l'entropie maximale préalable pour le processus de distributions inconnues et exemple numérique

Dans ce chapitre, nous examinons la question de savoir comment sélectionner les primes d'assurance de façon juste et précise, surtout quand le risque est inconnu ou incertain. Notre but est de trouver une méthode qui permette à la compagnie d'assurance d'estimer les primes de la manière la plus avantageuse possible, tout en respectant certaines règles.

L'approche que nous adoptons pour l'analyse de nos données et l'amélioration progressive de nos prévisions est basée sur le principe de l'entropie maximale. Cette technique se révèle particulièrement efficace dans le domaine de l'assurance, où il est fréquent de rencontrer des données partielles ou absentes.

Ce chapitre tire parti des recherches précédentes de **Hu Yingying et al. (2006)** ainsi que de **Alla.H (2021)**, qui ont mis en œuvre cette approche pour déterminer les primes d'assurance. Nous allons plus loin en élargissant ces recherches pour élaborer une méthode innovante de calcul des primes, même lorsque la distribution des sinistres $X_i(i = 1, 2, ..., n)$ est inconnue. De plus, nous présentons une application à des données réelles pour comparer la prime de crédibilité obtenue dans la section 3 de ce chapitre avec celle de **Gomez-Deniz (2006)** et celle de **Alla.H (2021)**, afin de renforcer ce travail.

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

Précédemment dans le chapitre 2, nous avons présenté de manière générale la méthode de l'entropie maximale. Dans cette partie, nous nous penchons sur son utilisation spécifique dans le secteur actuariel.

L'entropie maximale est un outil statistique qui permet d'effectuer une inférence impartiale lorsque les informations disponibles sont incomplètes et insuffisantes. Cette technique a été introduite par **Jaynes (1957)**. Bien qu'elle soit couramment utilisée en biotechnologie, en thermodynamique et en théorie de l'information, elle n'a pas été autant explorée dans le domaine actuariel, à l'exception de quelques études spécifiques.

Par exemple, **Landsman et Makov (1999, 2000)** ont fait appel à l'entropie maximale pour estimer le facteur de crédibilité et la distribution des paramètres d'échelle dans une famille exponentielle dispersée. **Darooneh (2004, 2006)** a utilisé cette méthode pour élaborer une fonction d'utilité associée à l'assurance non-vie, tandis que

Payandeh Najafabadi et al. (2011) ont proposé une nouvelle approche pour approximer les estimateurs de Bayes en utilisant l'entropie maximale.

L'entropie maximale est basée sur l'évaluation de l'incertitude d'un événement dans le cadre de la théorie de l'information. **Shannon (1948)** a mis en place la mesure d'entropie la plus répandue pour un événement présentant n résultats différents, qui est proportionnelle à :

$$Y = -\sum_{i=1}^{n} p_i \ln(p_i)$$

Dans le contexte actuariel, la distribution de probabilité est définie en utilisant l'entropie maximale sous certaines conditions :

— 1. La probabilité doit répondre aux conditions axiomatiques : $\sum_{i=1}^{n} p_i = 1$ pour $p_i \geq 0$ (pour $i = 1, 2, \ldots, n$).

— 2. Les données d'observation peuvent être représentées par leurs moments statistiques (tels que la moyenne, la variance, etc...).

Pour résoudre le problème des valeurs extrêmes conditionnelles, on utilise le multiplicateur de Lagrange pour calculer la distribution de probabilité en respectant ces conditions.

Définition 3.1.1. *Supposons qu'une variable aléatoire X prenne des valeurs distinctes $x_1, x_2, ..., x_n$ avec des probabilité correspondantes $P_1, P_2, ..., P_n$ et les contraintes $\varepsilon_1(.), ..., \varepsilon_m(.)$, où $\sum_{i=1}^{n} \varepsilon_r(x_i) = \gamma_r$ pour $r = 1, ..., m$. Alors, le multiplicateur de Lagrange sous réserve des contraintes ci-dessus est :*

$$L = -\sum_{i=1}^{n} P_i \log(P_i) - \sum_{r=1}^{m} \lambda_r (\sum_{i=1}^{n} P_i \varepsilon_r (x_i - \gamma_r)) \tag{3.1}$$

3.2 Les principaux résultats

Dans cette section, nous présentons la principale avancée de notre recherche, illustrée par le théorème ci-dessous. Celui-ci démontre comment obtenir une nouvelle mesure de crédibilité dans le contexte de la fonction de perte équilibrée. Pour ce faire, nous résolvons un problème de minimisation et nous appliquons la méthode de l'entropie maximale.

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

$$P_{n+1}^{MEM} = \gamma_0^* + \sum \gamma_k^* X_k, \tag{3.2}$$

qui répond aux critères suivants :

— Minimisation de la fonction de perte carrée équilibrée espérée de :

$$E[\omega(\delta_0 - \gamma_0 - \sum_{i=1}^{n} \gamma_i x_i)^2 + (1-\omega)(X_{n+1} - \gamma_0 - \sum_{i=1}^{n} \gamma_i x_i)^2]$$

— Maximisation de l'entropie avec :

$$\gamma_0^* = \omega E[\delta_0(x)] + \frac{(1-\omega)E[X_{n+1}]}{(1 + \sum_{i=1}^{n} exp^{(\frac{(1-\omega)E[X_{n+1}]}{E[X_k]}\lambda cov(X_i, n\bar{X}))})}$$

$$\gamma_k^* = \frac{(\gamma_0^* - \omega E[\delta_0(x)])exp^{(\frac{(1-\omega)E[X_{n+1}]}{E[X_k]}\lambda cov(X_k, n\bar{X}))}}{E[X_k]}$$

où λ est la solution de l'équation suivante :

$$(\frac{(1-\omega)E[X_{n+1}]}{E[X_i]}cov(X_i,n\bar{X}))\sum_{i=1}^{n}exp^{(\frac{(1-\omega)E[X_{n+1}]}{E[X_i]}\lambda cov(X_i,n\bar{X})} \tag{3.3}$$

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

Preuve. Pour commencer, il faut résoudre le problème de minimisation :

$$minE[\omega(\delta_0-\gamma_0-\sum_{i=1}^{n}\gamma_i x_i)^2+(1-\omega)(X_{n+1}-\gamma_0-\sum_{i=1}^{n}\gamma_i x_i)^2]$$

pour simplification, on note :

$$Y=E[\omega(\delta_0-\gamma_0-\sum_{i=1}^{n}\gamma_i x_i)^2+(1-\omega)(X_{n+1}-\gamma_0-\sum_{i=1}^{n}\gamma_i x_i)^2] \tag{3.4}$$

on dérive l'équation (3.4) par rapport à γ_0 et on met la formule résultante égale à 0, on trouve

$$\gamma_0=\omega E[(\delta_0(x))]+(1-\omega)E[X_{n+1}]-\sum_{i=1}^{n}\gamma_i E[X_i]$$

qui peut s'écrire comme suit :

$$f_1(\underline{\gamma})\underline{\underline{\Omega}}\frac{\gamma_0}{(1-\omega)E[X_{n+1}]}-\frac{\omega E[\delta_0(x)]}{1-\omega)E[X_{n+1}]}+\frac{\sum_{i=1}^{n}E[X_i]}{(1-\omega)E[X_{n+1}]}-1=0$$

$$f_1(\underline{\gamma})\underline{\underline{\Omega}}\frac{\gamma_0-\omega E[\delta_0(x)]}{(1-\omega)E[X_{n+1}]}+\frac{\sum_{i=1}^{n}E[X_i]}{(1-\omega)E[X_{n+1}]}-1=0$$

avec : $\underline{\gamma}=(\gamma_0,\gamma_1,...,\gamma_n)$. On remplace γ_0 dans l'équation (3.4), puis on dérive par rapport à γ_i et on met la formule résultante égale à 0, nous aurons

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

ce qui est équivalent à :

$$f_2(\underline{\gamma}) \underline{\underline{\Omega}} \omega \sum_{j=1}^{n} cov(\delta_0(x), X_j) - \sum_{j=1}^{n} \sum_{i=1}^{n} \gamma_i cov(X_i, X_j) = 0$$

avec : $\underline{\gamma} = (\gamma_0, \gamma_1, ..., \gamma_n)$

La formule associée au multiplicateur de Lagrange peut être représentée comme suit :

$$L = -\sum_{i=1}^{n} P_i Ln(P_i) - \lambda_0 f_1(\underline{\gamma}) - \lambda f_2(\underline{\gamma})$$

$$= -(\frac{\gamma_0 - \omega E(\delta_0(x))}{(1-\omega)E[X_{n+1}]}) ln(\frac{\gamma_0 - \omega E(\delta_0(x))}{(1-\omega)E[X_{n+1}]}) - (\frac{\sum_{i=1}^{n} \gamma_i E[X_i]}{(1-\omega)E[X_{n+1}]}) ln(\frac{\sum_{i=1}^{n} \gamma_i E[X_i]}{(1-\omega)E[X_{n+1}]})$$

$$-\lambda_0(\frac{\gamma_0 - \omega E[\delta_0(x)]}{(1-\omega)E[X_{n+1}]} + \frac{\sum_{i=1}^{n} \gamma_i E[X_i]}{(1-\omega)E[X_{n+1}]} - 1) - \lambda(\omega \sum_{j=1}^{n} cov(\delta_0(x), X_j) - \sum_{j=1}^{n} \sum_{i=1}^{n} \gamma_i cov(X_i, X_j)) \tag{3.5}$$

on dérive l'équation (3.5) par rapport à γ_0, on trouve

$$\gamma_0 = \omega E[\delta_0(x)] + (1-\omega)E[X_{n+1}] exp(-1-\lambda_0)$$

on remplace γ_0 dans l'équation (3.5), on obtient

$$L = (1+\lambda_0) exp^{(-1-\lambda_0)} - (\frac{\sum_{i=1}^{n} \gamma_i E[X_i]}{(1-\omega)E[X_{n+1}]}) ln(\frac{\sum_{i=1}^{n} \gamma_i E[X_i]}{(1-\omega)E[X_{n+1}]})$$

$$-\lambda_0 exp^{(-1-\lambda_0)} - \lambda_0(\frac{\sum_{i=1}^{n}\gamma_i E[X_i]}{(1-\omega)E[X_{n+1}]})+$$

$$\lambda_0 - \lambda(\omega\sum_{j=1}^{n} cov(\delta_0(x), X_j) - \sum_{j=1}^{n}\sum_{i=1}^{n}\gamma_i cov(X_i, X_j)) \qquad (3.6)$$

comme ci-dessus, on dérive l'équation (3.6) par rapport à γ_k et on met la formule résultante égale à 0, on obtient

$$\gamma_k = \frac{(1-\omega)E[X_{n+1}]}{E[X_k]} exp^{(-1-\lambda_0+\frac{(1-\omega)E[X_{n+1}]}{E[X_k]}\lambda cov(X_k, n\bar{X}))}$$

On remplace γ_k dans l'équation (3.6), la formule devient

$$L = exp^{(-1-\lambda_0)} + \lambda_0 - \lambda\omega cov(\delta_0(x), n\bar{X}) + \sum_{i=1}^{n} exp^{(-1-\lambda_0+\frac{(1-\omega)E[X_{n+1}]}{E[X_k]}\lambda cov(X_i, n\bar{X}))} \qquad (3.7)$$

on traitant l'équation (3.9) par rapport à λ_0 avec les mêmes opérations que ci-dessus, on trouve,

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

puis

$$L = ln(1 + \sum_{i=1}^{n} exp^{(\frac{(1-\omega)E[X_{n+1}]}{E[X_k]}\lambda cov(X_i, n\bar{X}))}) - \lambda\omega cov(\delta_0(x), n\bar{X}) \qquad (3.8)$$

On dérive l'équation précédente par rapport à λ et ont met la formule résultante égale à 0, après une série de calcul simples, une équation à propos de λ peut être présentée comme suit

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

$$= (1 + \sum_{i=1}^{n} exp^{(\frac{(1-\omega)E[X_{n+1}]}{E[X_k]}\lambda cov(X_i, n\bar{X}))})\omega cov(\delta_0(x), n\bar{X})) \tag{3.9}$$

nous considérons λ comme la solution de l'équation (3.9) et nous résolvons γ_0 et γ_k respectivement

$$\gamma_0^* = \omega E[\delta_0(x)] + \frac{(1-\omega)E[X_{n+1}]}{(1 + \sum_{i=1}^{n} exp^{(\frac{(1-\omega)E[X_{n+1}]}{E[X_k]}\lambda cov(X_i, n\bar{X}))})}$$

$$\gamma_k^* = \frac{(\alpha_0^* - \omega E[\delta_0(x)])exp^{(\frac{(1-\omega)E[X_{n+1}]}{E[X_k]}\lambda cov(X_k, n\bar{X}))}}{E[X_k]}$$

par conséquent

$$P_{n+1}^{MEM} = \gamma_0^* + \sum_{k=1}^{n} \gamma_k^* X_k$$

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 :*

$$P_{n+1}^{MEM} = M^*\bar{X} + (1 - M^*)b, \tag{3.10}$$

avec

$$M^* = \frac{\omega n V(\bar{X})}{(n\phi^2 + \rho^2)(1-\omega)},$$

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

$$b = \omega\bar{X}(2-\omega) + (1-\omega)^2\eta.$$

Proposition 3.2.1. *Tout d'abord, nous devrions trouver la solution de λ. En supposant* *que*

- $cov(X_i, n\bar{X}) = n\phi^2 + \rho^2$
- $cov(\delta_0(X), n\bar{X}) = cov(X_{med}, n\bar{X}) = d$, *avec :*

  $X_{med} = X_{(\frac{n+1}{2})} = \alpha\bar{X}$ *if* n *est impair,* $\alpha \in \Re$.

  $X_{med} = X_{(\frac{n}{2})} = \alpha\bar{X}$ *if* n *est pair,* $\alpha \in \Re$.
- $E(X_{n+1}) = \frac{b-\omega\bar{X}}{1-\omega}$

et on les remplaçant dans l'équation (3.9)*, on trouve*

$$\lambda = \frac{1}{(1-\omega)(n\phi^2+\rho^2)} ln(\frac{\omega d}{n((1-\omega)(n\phi^2+\rho^2)-\omega d)})$$

Par suite, on obtient

$$\gamma_0^* = \omega E[\delta_0(X)] + \frac{(m-\omega\bar{X})}{\frac{(1-\omega)(n\phi^2+\rho^2)}{(1-\omega)(n\phi^2+\rho^2)-\omega d}}$$

$$\gamma_k^* = \frac{\omega d}{n(n\phi^2+\rho^2)}$$

Enfin,

$$P_{n+1}^{MEM} = \frac{\omega}{(1-\omega)}(\frac{d}{(n\phi^2+\rho^2)} - 1)\bar{X} + (1 - \frac{\omega}{(1-\omega)}(\frac{d}{(n\phi^2+\rho^2)} - 1))b$$

D'ou :

$$P_{n+1}^{MEM} = M^*\bar{X} + (1-M^*)b$$

avec

$$M^* = \frac{\omega}{(1-\omega)}(\frac{d}{(n\phi^2+\rho^2)} - 1)$$

$$d = cov(\delta_0(X), n\bar{X}) = cov(X_{med}, n\bar{X})$$

$$b = \omega\bar{X}(2-\omega) + (1-\omega)^2\eta.$$

3.3 Simulation numérique et comparaison des primes de crédibilité

Une étude numérique est effectuée pour comparer la prime de crédibilité dérivée dans ce travail avec celle de **Gomez-Deniz (2006)** et celle de **Alla. H (2021)**. Conformément au théorème (3.2.1) présenté dans **Gomez-Deniz (2006)**, en exploitant l'ensemble exponentiel de distributions, nous pouvons obtenir une prime bayésienne linéaire non seulement sous la fonction de perte traditionnelle, mais aussi sous la fonction de perte équilibrée.

La prime bayésienne proposée par Gomez-Deniz, notée P^{Gomez}, peut être formulée de la manière suivante :

$$P^{Gomez} = M^{'}b^{'} + (1 - M^{'})[(1-\omega)^2 + \omega\delta_0(2-\omega)]$$

$$= \omega\delta_0(2-\omega) + (1-\omega)^2(M^{'}\bar{X} + (1-M^{'})b^{'}) \tag{3.11}$$

où, $M^{'} = n\phi^2/(n\phi^2 + \rho^2)$ est le facteur de crédibilité, $b^{'}$ représente la prime individuelle sous la fonction de perte classique.

La prime bayésienne proposée par Alla.H notée $P^{Alla.H}$, peut être présentée de la façon suivante :

$$P^{Alla.H} = M^{*}\bar{X} + (1 - M^{*}b)$$

où $M^{*} = \frac{\omega n Var(\bar{X})}{(n\phi^2+\rho^2)(1-\omega)}$ est le facteur de crédibilité, $b = \omega\bar{X}(2-\omega) + (1-\omega)^2\eta$ représente la prime individuelle sous la fonction de perte équilibrée quadratique.

Afin de valider l'efficacité de nos conclusions théoriques, nous avons recours à un ensemble de données provenant de la plus grande institution de sécurité sociale en Grèce. Cet ensemble de données englobe un total de 5,530,000 individus employés en tant que travailleurs et salariés, bénéficiant de soins médicaux, de médicaments prescrits et de services hospitaliers. Pour plus d'informations, se référer à **Pitselis (2013)**. Les 6 contrats correspondent aux 6 catégories distinctes de prestations fournies par l'institution de sécurité sociale grecque. Ces prestations sont assurées sur une période de 22 ans, précisément de 1980 à 2001. Les six catégories distinctes (contrats) sont :

— $Sick_A$ = Allocation de maladie : les retraités encore en activité qui sont directement assurés ont accès à une indemnité de maladie.

— $Accid_A$ = allocation d'accident : si l'incapacité de travailler résulte d'une plainte ou d'une blessure causée par un accident survenu au cours du travail.

— $Matern_A$ = prestation de maternité : Inclut l'indemnité de naissance. Les bénéficiaires sont des femmes directement assurées, des femmes percevant une pension ou les épouses de maris directement assurés ou de retraités.

— $Funer_{Exp}$ = Frais funéraires : Ils sont payés en cas de décès d'un assuré direct ou d'un retraité en raison de la vieillesse, de l'invalidité ou de la mort.

— $Other_A$ = Autres allocations : comprennent : (1- les soins dentaires/ 2- les soins complémentaires/ 3- le tourisme thérapeutique et une prestation de cure thermale pour les retraités à faible revenu/ 4- la médecine préventive).

— $Manag_{Exp}$ = Frais de gestion : Les frais administratifs de l'organisme de sécurité sociale sont inclus.

Pour calculer la prime de crédibilité pour l'année 2002, nous utilisons les valeurs des paramètres de structure (b', ϕ^2 et ρ^2) qui sont présentées dans le tableau 1 dans **Pitselis (2013)**. Considérons que δ_0 est la médiane des montants des réclamations pour la période 1980-2001 : $\delta_0 = X_{med}$, Le tableau 1 montre les valeurs de P^{Gomez}, P^{Alla} et P^{New}_{2002} obtenues à l'aide de données réelles, Le poids a été pris comme $\omega = 0.53$.

Contrats	P^{Gomez}	P^{Alla}	P^{New}_{2002}	X_j
$Sick_A$	8267.092	19032.284	17531.736	17527.5
$Accid_A$	4931.723	1488.725	1792.833	1970.00
$Matern_A$	5845.365	6293.636	5337.521	6230.96
$Funer_{Exp}$	5158.146	2680.153	2168.865	3026.55
$Other_A$	5158.366	2680.661	2967.473	3027.00
$Manag_{Exp}$	5070.284	2264.51	1665.571	2657.96

TABLE 3.1 – Montants des primes pour les six contrats(les primes sont en millions de drachmes : 1Euro=340.75 drachmes)

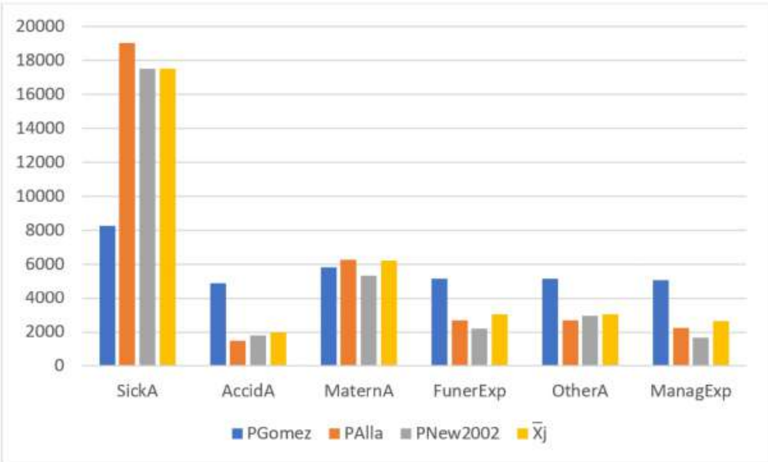

FIGURE 3.1 – Résultats des primes à partie d'un ensemble de données de l'organisme de sécurité sociale en grèce

En analysant les résultats du tableau 3.1 et en examinant la figure 3.3, nous constatons que la prime calculée selon notre approche P^{New}_{2002} est généralement plus proche de l'expérience individuelle de chaque contrat $\bar{X}_j$. Il est essentiel de vérifier ce critère de proximité pour s'assurer que la prime facturée reflète fidèlement l'historique des sinistres des assurés. Après cet exemple numérique, nous pouvons affirmer que notre prime peut être utilisée par des actuaires à la recherche de primes compétitives, attirant ainsi les assurés et préservant la stabilité financière de leur entreprise.

Conclusion

Cette recherche emploie une méthode peu conventionnelle, rarement rencontrée dans les écrits actuariels, pour déterminer les primes de crédibilité au sein d'une fonction de perte équilibrée. Cette stratégie produit des primes de crédibilité qui s'accroissent linéairement, grâce à des calculs simples et directs. Par conséquent, nous avons la possibilité d'élargir l'idée présentée dans cette recherche en intégrant une modification quadratique qui prend en compte des mesures statistiques d'ordre supérieur et incorpore les valeurs carrées des observations antérieures. De plus, notre recherche a examiné spécifiquement le choix de $\delta_0(x)$ comme estimateur cible. Dans les recherches futures, nous envisageons d'utiliser l'estimation gaussienne pour déterminer les primes de crédibilité en utilisant des intervalles de confiance pour divers quantiles.

Bibliographie

[1] Alla Eddine Haddari, Remita Mohamed Riad, Metiri farouk and Sadoun Ahmed. 2021. Credibility premiums under the balanced loss function using the maximum entropy method. Journal of Applied Probability and Statistics 16 (1), 83-94.

[2] Bailey, A., 1950. A generalized theory of credibility. Proceedings of the Casualty Actuarial Society 13, 13-20.

[3] Buhlmann, H., 1967. Experience rating and credibility. Astin Bulletin 4 (3), 199-207.

[4] Buhlmann, H., Straub, E., 1970. Glaubwudigkeit fur Schadensze. Bulletin of the Swiss Association of Actuaries 70, 111-133.

[5] Darooneh, A.H., 2004. Non-life insurance pricing : multi agent model. The European Physical Journal B-Condensed Matter and Complex Systems 42, 119-122.

[6] Darooneh, A.H., 2006. Utility function from maximum entropy principle. Entropy 8, 18-24.

[7] De Vylder, F., 1976. Geometrical credibility. Scandinavian Actuarial Journal, 121-149.

[8] De Vylder, F., 1976. Optimal semilinear credibility. Bulletin of swiss ass. Act., 78, 27-40.

[9] Gomez-Deniz, 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.

[10] Gomez-Deniz, E., 2008. A generalization of the credibility theory obtained by using the weighted balanced loss function. Insurance : Mathematics and Economics 42, 850-854.

[11] 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.

[12] 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.

[13] Jaynes, E.T., 1957. Information theory and statistical mechanics. Physical Reviews 106, 620-630.

[14] Jeweel, W. S., 1975. The use of collateral data in credibility theory : A hierarchicalmodel. Giornale dell ?Istituto Italiano degli Attuari 38, 1-16.

[15] Landsman, Z., Makov, U. E., 1999. Credibility evaluation for the exponential dispersion family. Insurance : Mathematics and Economics 24, 23-29.

[16] Landsman, Z., Makov, U.E., 2000. On credibility evaluation and the tail area of the exponential dispersion family. Insurance : Mathematics and Economics 27, 277-283.

[17] Mowbray, A., 1914. How extensive a payroll is necessary to give dependable pure premium. Proceedings of the Casualty Actuarial Society 1, 24-30.

[18] Naim BOUDJELIDA, Mohamed Riad REMITA, Halim ZEGHDOUDI. 2024. On Bayesian Premium Credibility : Maximum Entropy Prior for Unknown Claims

Process and Numerical Example. Studies in engineering and exact sciences,5(1), 1888-1903.

[19] Payandeh Najafabadi, A. T., Hatami H., Omidi Najafabadi M., 2012. A maximum entropy approach to the linear credibility formula. Mathematics and Economics, 51(1),216-221.

[20] Pitselis, G. (2013). Quantile credibility models. Insurance : Mathematics and Economics, 52, 477-489.

[21] Shannon, C.E., 1948. A mathematical theory of communication. Bell System Technical Journal 27, 379-423. 623-659.

[22] Whitney, A., 1918. The theory of experience rating. Proceedings of the Casualty Actuarial Society 4, 274-292.

[23] 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.