

**Mémoire présenté le :  
pour l'obtention du diplôme  
de Statisticien Mention Actuariat  
et l'admission à l'Institut des Actuares**

Par : Monsieur REICHARDT Léo

**Titre du mémoire :**

*Évaluation de l'incertitude sur paramètres et intégration de données externes  
dans le calcul du risque de primes*

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

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

Membres présents du jury de la  
filière :

Signature :

Entreprise :

Nom : *GROUPAMA ASSURANCES  
MUTUELLES*

Signature : *[Signature]*

Directeur de mémoire en  
entreprise

Membres présents du jury de  
l'Institut des Actuares :

Signature :

Nom : *MANSOUR YOUVA*

Signature : *[Signature]*

Invité :

Nom :

Signature :

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

Signature du responsable  
entreprise :

*[Signature]*

Signature du candidat :

*[Signature]*

# Résumé

L'activité d'assurance se caractérise par l'inversion de son cycle de production, un phénomène intrinsèquement lié à la nature des engagements contractés par les compagnies d'assurance. En effet, contrairement à de nombreux secteurs, les assureurs s'engagent sur des périodes potentiellement longues, conditionnant leurs prestations à des événements futurs incertains. Dès lors, les compagnies se trouvent contraintes de constituer des réserves à la hauteur de leurs engagements, ce qui implique de les quantifier le plus finement possible. Traditionnellement, les actuaires ont eu recours à des méthodes statistiques basées sur l'analyse d'historiques de sinistralités, restituant l'inférence sous forme de paramètres. Deux questions cruciales émergent dans ce contexte : ces paramètres sont-ils réellement pertinents pour anticiper les futures obligations financières des assureurs ? Est-il possible d'enrichir les calibrages afin d'améliorer la précision des évaluations du risque ? La réunion de ces deux problématiques dans un unique mémoire s'explique par la prééminence, sur certains aspects, de la statistique bayésienne par rapport à la statistique classique généralement employée dans le monde de l'assurance. L'objet de la réflexion exposée dans ce mémoire vise donc à exploiter le potentiel de la statistique bayésienne pour ouvrir la voie à une meilleure compréhension et gestion des risques dans le secteur assurantiel.

Cette étude s'inscrit dans un contexte de calcul de Capital de Solvabilité Requis et se focalise plus particulièrement sur l'évaluation du risque de primes en assurance non-vie.

**Mots Clés** : Incertitude sur paramètres, Intégration de données externes, Statistiques bayésiennes, Risque de primes, Modèle interne, Prior informatif, Prior non-informatif, Jugement d'expert, SCR, Solvabilité II

# Abstract

Insurance activity is characterized by the inversion of its production cycle, a phenomenon intrinsically linked to the nature of commitments made by insurance companies. Unlike many sectors, insurers commit to potentially long periods, conditioning their services on uncertain future events. Consequently, companies are compelled to establish reserves commensurate with their commitments, requiring precise quantification. Traditionally, actuaries have employed statistical methods based on the analysis of claims histories, presenting inferences in parameter form. Two crucial questions arise in this context : are these parameters genuinely relevant for anticipating insurers' future financial obligations? Is it possible to enhance calibrations to improve the accuracy of risk assessments? The convergence of these two issues in a single thesis is explained by the pre-eminence, in certain aspects, of Bayesian statistics over classical statistics commonly used in the insurance industry. The objective of the reflection presented in this thesis is to harness the potential of Bayesian statistics to pave the way for a better understanding and management of risks in the insurance sector.

This study is conducted in the context of calculating the Solvency Capital Requirement and focuses more specifically on the assessment of non-life insurance premium risk.

**Keywords** : Parameter Uncertainty, Parameter Risk, Integration of External Data, Bayesian Statistics, Premium Risk, Internal Model, Informative Prior, Non-informative Prior, Expert Judgment, SCR (Solvency Capital Requirement), Solvency II.

# Note de Synthèse

L'activité d'assurance se caractérise par l'inversion de son cycle de production, un phénomène intrinsèquement lié à la nature des engagements contractés par les compagnies d'assurance. En effet, contrairement à de nombreux secteurs, les assureurs s'engagent sur des périodes potentiellement longues, conditionnant leurs prestations à des événements futurs incertains. Dès lors, les compagnies se trouvent contraintes de constituer des réserves à la hauteur de leurs engagements, ce qui implique de les quantifier le plus finement possible. Traditionnellement, les actuaires ont eu recours à des méthodes statistiques basées sur l'analyse d'historiques de sinistralités, restituant l'inférence sous forme de paramètres. Deux questions cruciales émergent dans ce contexte : ces paramètres sont-ils réellement pertinents pour anticiper les futures obligations financières des assureurs ? Et, est-il possible d'enrichir les calibrages afin d'améliorer la précision des évaluations du risque ? La réunion de ces deux problématiques dans un unique mémoire s'explique par la supériorité, sur certains aspects, de la statistique bayésienne par rapport à la statistique classique généralement employée dans le monde de l'assurance. L'objet de la réflexion exposée dans ce mémoire vise donc à exploiter le potentiel de la statistique bayésienne pour ouvrir la voie à une meilleure compréhension et gestion des risques dans le secteur assurantiel.

Cette étude s'inscrit dans un contexte de calcul de Capital de Solvabilité Requis et se focalise plus particulièrement sur l'évaluation du risque des primes en assurance non-vie. Mathématiquement, l'objectif des travaux consiste à élaborer une distribution de probabilité qui, en raison de contraintes techniques, doit être renseignée sous la forme d'une série de quantiles.

## **Incertitude sur paramètres**

L'incertitude, ou risque, sur paramètres correspond à l'incertitude concernant l'aptitude des valeurs estimées des paramètres à retranscrire le comportement du phénomène modélisé. Elle est inhérente aux techniques actuarielles de modélisation et ne peut être complètement effacée. Une particularité de ce risque réside dans le fait qu'il ne se diversifie pas au sein d'une même branche d'assurance. Aucune mutualisation n'est possible entre les différents assurés, ce qui prive les assureurs de leurs outils usuels de gestion des risques.

Dans ce contexte, l'article 230 du règlement délégué 2.A de Solvabilité II impose aux assureurs d'explicitier et de justifier chaque hypothèse retenue, notamment les choix de paramètres, en considération de leur importance et de l'incertitude qui leur est inhérente.

Se pose alors la question de l'évaluation de cette incertitude. Pour ce faire, l'approche adoptée est similaire à celle d'une mesure d'effet. Le coût de l'incertitude sur paramètres est défini comme la différence entre le résultat du calibrage fréquentiel classique de Groupama et le résultat d'un calibrage prenant en compte cette incertitude.

L'objectif de cette étude n'est pas d'ajouter une couche aux Fonds Propres réglementaires, mais plutôt de garantir la maîtrise pleine et entière de la matérialité de l'hypothèse retenue quant à la valeur et la constance des paramètres.

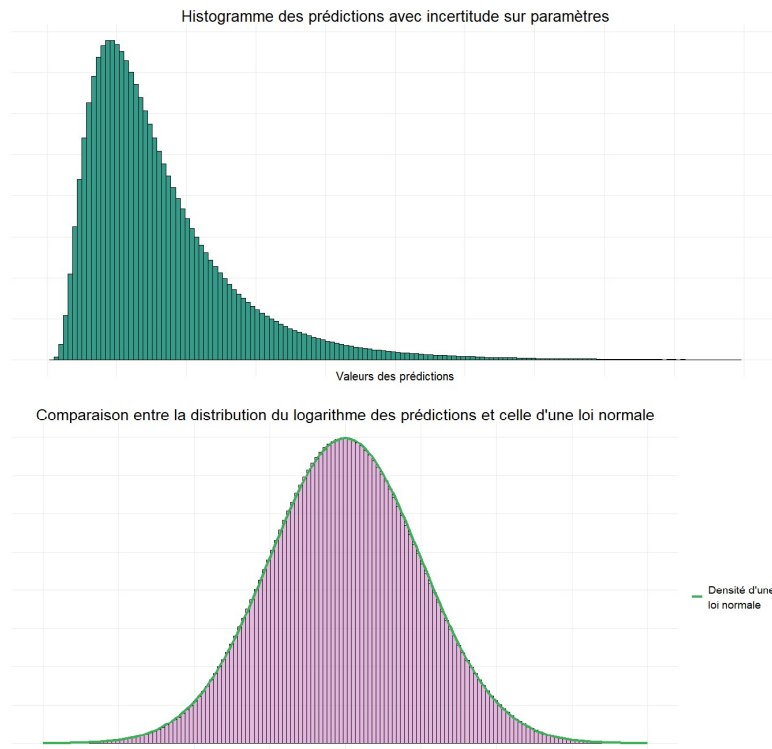


FIGURE 1 – Distribution prédictive avec prise en compte de l’incertitude sur paramètres. Le logarithme des prédictions est distribué selon une loi normale.

Historiquement, les équipes de Groupama utilisaient une méthode consistant à prendre comme loi des paramètres la loi asymptotique de leurs estimateurs, sous l’hypothèse que les paramètres sont fixes. Cette approche a produit des résultats incohérents sur les données de Groupama.

A contrario, la méthodologie présente dans ce mémoire est entièrement basée sur l’hypothèse que les paramètres sont des quantités incertaines sur lesquelles les observations apportent des informations. Or, la formalisation de ce cadre d’hypothèse correspond à la théorie des statistiques bayésiennes.

Dès lors, il revient à l’actuaire de définir une loi a priori  $\pi$  pour ses paramètres, de la mettre à jour à partir de l’observation de l’échantillon, puis de déduire, au moyen d’une fonction de coût, une valeur pour la quantité d’intérêt.

Étant donné que l’objectif est de comparer des quantiles obtenus par un calibrage fréquentiel basé uniquement sur l’historique de données, le prior retenu appartient à la catégorie des priors dits ”non-informatifs”.

Une fois la distribution a posteriori  $\pi(\theta|X)$  obtenue, il ne reste plus qu’à en déduire des valeurs de quantiles pour la distribution prédictive. L’approche retenue consiste à aborder directement le problème comme une estimation de quantile dans un cadre bayésien. Sous cet angle, le quantile de niveau  $\alpha$  de la sinistralité est défini à partir de la fonction quantile de la vraisemblance en la moyennant selon la loi a posteriori.

Cette méthodologie présente l’avantage que la vraisemblance attribuée aux données se retrouve dans la loi prédictive. De plus, une réinterprétation de l’évaluation du risque sur paramètres sous la forme d’une inégalité de Jensen permet de fournir des arguments supplémentaires en faveur de cette méthode. En effet, dans le cas

de l'étude, la fonction quantile de la vraisemblance notée  $q_\alpha$  est celle d'une loi log-normale. Elle est donc convexe et l'inégalité de Jensen donne  $\mathbb{E}_{\pi(\theta|X)}[q_\alpha(\theta)] \geq q_\alpha(\mathbb{E}_{\pi(\theta|X)}[\theta])$  avec un écart d'autant plus large que le posterior est dispersé. Or,  $\mathbb{E}_{\pi(\theta|X)}[\theta]$  est en pratique proche du paramétrage fréquentiel pour un prior non informatif, donc  $q_\alpha(\mathbb{E}_{\pi(\theta|X)}[\theta])$  est proche du quantile sans incertitudes. La différence entre les deux côtés de l'inégalité représente alors la matérialité de l'incertitude sur paramètres. En résumé, dans le cadre de l'étude, la méthode proposée dernière jouit alors des propriétés suivantes :

- la prise en compte du risque sur paramètres ne peut que faire augmenter l'appréciation du risque ;
- la matérialité est d'autant plus grande que le posterior est diffus, c'est-à-dire que les incertitudes sur la valeurs des paramètres est grande ;
- l'incertitude est considérée comme nulle. uniquement si le posterior est concentré en un point ;
- la distribution prédictive est de même nature que la vraisemblance ;
- de manière générale, l'incertitude sur paramètres est d'autant plus grande que la fonction quantile de la vraisemblance est courbée, ie. que les incertitudes de calibrages ne se compensent pas entre elles.

Finalement, une application sur les données de Groupama conduit à la conclusion que ce risque est négligeable dans le calcul de son risque de primes.

### **Intégration de données externes**

Dans le monde de l'assurance, la disponibilité croissante de données provenant de sources variées offre de nouvelles opportunités pour éclairer les prises de décision. L'exploitation stratégique de ces ressources pourrait conférer un avantage compétitif significatif en améliorant la compréhension et la modélisation des risques au sein d'une compagnie.

Les données externes englobent tout ce qui ne figure pas dans les historiques traditionnels, provenant aussi bien de sources externes qu'internes, telles que les résultats d'études menées par d'autres directions. Ces dernières peuvent se révéler particulièrement précieuses du fait qu'elles se rapportent directement aux mêmes garanties et portefeuilles d'assurés.

Dans ce contexte, la statistique bayésienne se démarque par sa capacité à orienter l'inférence à travers l'utilisation d'une loi a priori. Cependant, le prior devant être une distribution de probabilité sur l'espaces des paramètres, le défi réside dans la conversion efficace d'une information externe à ce format. L'étude se restreint au cas d'un jugement chiffré de la forme  $f(\text{Sinistralité}) = u$ , devant représenter le résultat d'une étude antérieure.

La démarche proposée se divise en deux phases. Tout d'abord, la relation fonctionnelle sur la sinistralité est transformée en une relation sur sa loi, et par conséquent sur ses paramètres, représentée par l'équation  $f(\theta) = u$ . Ensuite, le prior est construit de manière à ce que ses lignes de niveaux coïncident avec celles de la fonction  $f$ . Cette approche repose sur la logique selon laquelle deux paramètres ayant la même image devraient être chargés de la manière lorsque la seule information disponible est la

valeur d'une fonction évaluée sur les paramètres. De plus, entre deux paramétrages possibles, celui dont l'image se rapproche le plus du niveau spécifié par l'expert devrait être considéré comme plus probable. A l'issue d'un raisonnement logique, ces observations fixent la forme du prior à la composition d'une distance entre  $f(\theta)$  et  $u$  avec une fonction positive décroissante.

Dans le cas de l'étude, la fonction décroissante et la distance sont choisies par analogie avec une loi normale, donnant ainsi comme loi a priori  $\pi(\theta) \propto \exp\left(-\frac{(f(\theta)-u)^2}{\tau}\right)$  où  $\tau$  est un paramètre à fixer par le modélisateur. Un exemple de calibrage est donné dans le cas d'un jugement concernant un coût de réassurance.

L'analyse démontre que l'utilisation du prior défini précédemment permet d'obtenir plusieurs résultats :

- la distribution a priori de  $f(\textit{Sinistralité})$  est centrée autour du niveau  $u$  spécifié par l'expert ;
- le paramètre d'incertitude  $\tau$  de contrôle dans quelle mesure des paramètres peu conformes au jugement d'expert peuvent être pris en compte ;
- en même temps, le paramètre  $\tau$  régule le poids relatif de l'information a priori par rapport à celle provenant des données dans la formation du posterior ;
- les distributions prédictives a posteriori sont cohérentes dans leur forme avec celles obtenues par des méthodes fréquentielles, ce qui atteste de leur réalisme. De plus, l'ajout d'un prior tend à mettre la distribution prédictive en adéquation avec le jugement de l'expert.

La méthode proposée permet donc de traduire efficacement une information extérieure sous la forme d'une distribution des paramètres. La lisibilité de son unique paramètre à fixer la rend de plus intuitive et maniable.

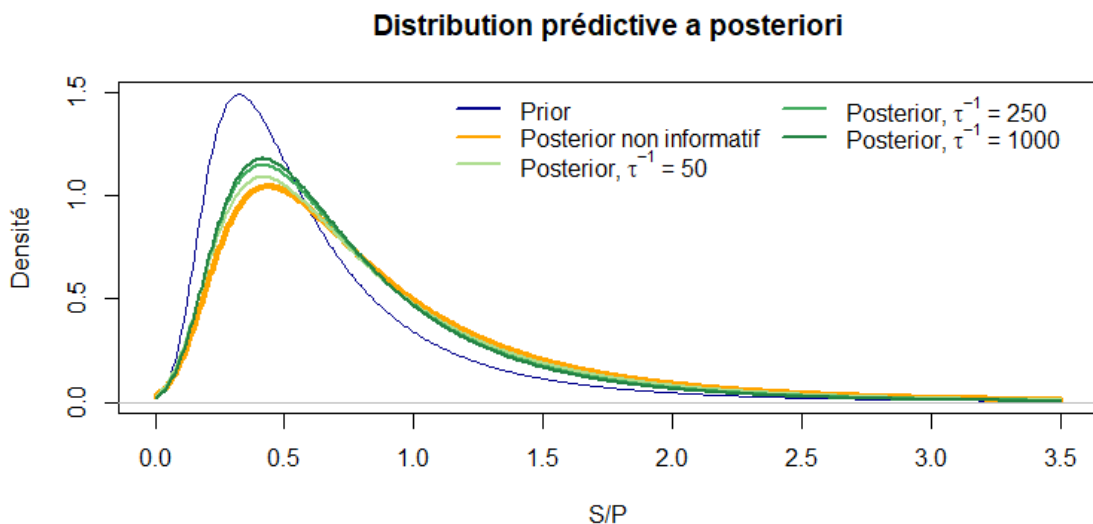


FIGURE 2 – Exemples de distributions prédictives avec intégration de données externes.

Cette approche présente toutefois plusieurs limites. Le choix basé sur une analogie avec une loi normale est arbitraire, il serait intéressant de mener une réflexion sur les manières de le guider. De plus, l'intégration de données externes met sur

un même plan un historique de données et des hypothèses, ce qui peut susciter des réticences dans des calculs aussi critiques et réglementés que ceux du Capital de Solvabilité Requis. Une ébauche de réflexion sur la gouvernance nécessaire pour encadrer cette intégration est proposée à la fin de l'étude.

Toutefois, le procédé décrit dans ce mémoire pourrait trouver une application pertinente dans un cadre de souscription, notamment dans des domaines tels que la santé prévoyance collective. En effet, il pourrait permettre de concilier les résultats du modèle tarifaire de l'entreprise avec les données comptables de sinistralité historique du client lors de la tarification d'un nouveau contrat.

En conclusion, cette étude met en avant deux contributions possibles de la statistique bayésienne à l'actuariat. Premièrement, elle fournit une analyse des méthodes envisageables pour évaluer les impacts de l'incertitude sur les paramètres, illustrée par une application sur les données de Groupama qui démontre la maîtrise de ce risque dans le calcul de son risque de primes. Deuxièmement, l'étude présente une méthode permettant d'intégrer efficacement des données externes dans la modélisation actuarielle, ouvrant des perspectives intéressantes malgré les enjeux de gouvernance associés. En résumé, la statistique bayésienne se positionne comme un outil puissant pour rehausser la précision des modèles actuariels. Elle offre des applications prometteuses, par exemple dans le domaine de la souscription, et peut s'avérer particulièrement pertinente dans la quantification de risques émergent tels que le risque climatique ou cyber, où les historiques traditionnels de sinistralité se révèlent insuffisants.



# Executive Summary

The insurance activity is characterized by the inversion of its production cycle, a phenomenon intrinsically linked to the nature of commitments made by insurance companies. Indeed, unlike many sectors, insurers commit to potentially long periods, conditioning their services on uncertain future events. Consequently, companies are compelled to establish reserves commensurate with their commitments, requiring precise quantification. Traditionally, actuaries have employed statistical methods based on the analysis of claims histories, presenting inferences in parametric form. Two crucial questions arise in this context : are these parameters genuinely relevant for anticipating insurers' future financial obligations? And, is it possible to enhance calibrations to improve the accuracy of risk assessments? The convergence of these two issues in a single thesis is explained by the superiority, in certain aspects, of Bayesian statistics over classical statistics commonly used in the insurance industry. The objective of the reflection presented in this thesis is to harness the potential of Bayesian statistics to pave the way for a better understanding and management of risks in the insurance sector.

This study is conducted in the context of calculating the Solvency Capital Requirement and focuses more specifically on the assessment of non-life insurance premium risk. Mathematically, the goal of the work is to develop a probability distribution that, due to technical constraints, must be specified in the form of a series of quantiles.

## Parameter Uncertainty

Parameter uncertainty, or risk, corresponds to uncertainty regarding the ability of the estimated parameter values to reflect the behavior of the modeled phenomenon. It is inherent in actuarial modeling techniques and cannot be completely eliminated. A particular characteristic of this risk is that it does not diversify within the same branch of insurance. No mutualization is possible among different policyholders, depriving insurers of their usual risk management tools.

In this context, Article 230 of Delegated Regulation 2.A of Solvency II requires insurers to explain and justify each assumption made, especially the choices of parameters, considering their importance and the inherent uncertainty.

The question then arises of the evaluation of this uncertainty. To do this, the approach adopted is similar to that of an effect measure. The cost of parameter uncertainty is defined as the difference between the result of Groupama's classical calibration and the result of a calibration taking this uncertainty into account.

The aim of this study is not to add a layer to the Solvency Capital Requirement but rather to ensure full mastery of the materiality of the assumption regarding the value and constancy of the parameters.

Historically, Groupama's teams used a method consisting of assuming the asymptotic law of their estimators as the law of the parameters, under the assumption that the parameters are fixed. This approach produced inconsistent results on Groupama's data.

Conversely, the methodology presented in this thesis is entirely based on the assumption that parameters are uncertain quantities on which observations provide

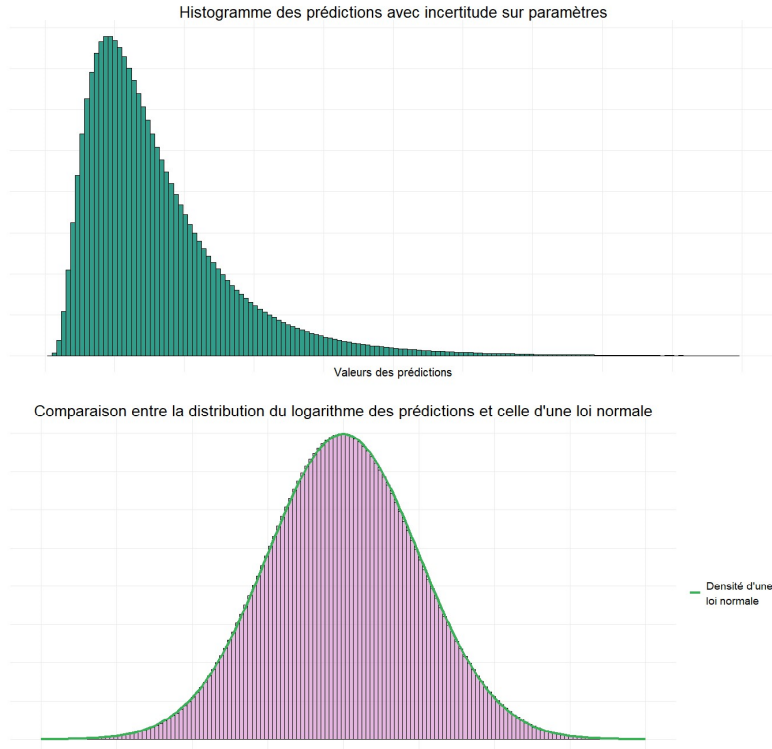


FIGURE 3 – Predictive distribution with consideration of parameter uncertainty. The logarithm of predictions follows a normal distribution.

information. The formalization of this framework corresponds to the theory of Bayesian statistics.

Therefore, it is the responsibility of the actuary to define a prior distribution  $\pi$  for their parameters, update it based on the observation of the sample, and then deduce, using a cost function, a value for the quantity of interest. Given that the objective is to compare quantiles obtained by a classical calibration based solely on the historical data, the chosen prior belongs to the category of so-called "non-informative" priors.

Once the posterior distribution  $\pi(\theta|X)$  is obtained, it only remains to deduce quantile values for the predictive distribution. The approach chosen is to directly address the problem as a quantile estimation in a Bayesian framework. From this perspective, the  $\alpha$ -level quantile of claims is defined based on the quantile function of the likelihood averaged according to the posterior distribution.

This methodology has the advantage that the likelihood attributed to the data is reflected in the predictive distribution. Furthermore, a reinterpretation of the risk assessment on parameters in the form of a Jensen's inequality provides additional arguments in favor of this method. Indeed, in the case of the study, the quantile function of the likelihood denoted  $q_\alpha$  is that of a log-normal distribution. It is thus convex, and Jensen's inequality yields  $\mathbb{E}\pi(\theta|X)[q_\alpha(\theta)] \geq q_\alpha(\mathbb{E}\pi(\theta|X)[\theta])$  with a larger gap as the posterior is more dispersed. However,  $\mathbb{E}\pi(\theta|X)[\theta]$  is practically close to the frequentist parameterization for a non-informative prior, so  $q_\alpha(\mathbb{E}\pi(\theta|X)[\theta])$  is close to the uncertainty-free quantile. The difference between the two sides of the

inequality represents the materiality of the uncertainty on parameters. In summary, in the context of the study, the proposed method enjoys the following properties :

- Considering parameter uncertainty can only increase the risk assessment.
- Materiality is greater when the posterior is diffuse, i.e., when uncertainties in parameter values are high.
- Uncertainty is considered zero only if the posterior is concentrated at a single point.
- The predictive distribution is of the same nature as the likelihood.
- Generally, parameter uncertainty is greater when the quantile function of the likelihood is curved, i.e., when calibration uncertainties do not offset each other.

Finally, an application to Groupama’s data leads to the conclusion that this risk is negligible in the calculation of its premium risk.

### **Integration of External Data**

In the insurance world, the increasing availability of data from various sources offers new opportunities to inform decision-making. Strategically harnessing these resources could confer a significant competitive advantage by improving the understanding and modeling of risks within a company.

External data encompasses everything not found in traditional histories, coming from both external and internal sources, such as the results of studies conducted by other departments. The latter can be particularly valuable as they directly relate to the same coverage and policyholder portfolios.

In this context, Bayesian statistics stand out for their ability to guide inference through the use of a prior distribution. However, since the prior must be a probability distribution over the parameter space, the challenge lies in the efficient conversion of external information to this format. The study is limited to the case of a quantitative judgment of the form  $f(Claims) = u$ , representing the result of a previous study.

The proposed approach is divided into two phases. First, the functional relationship on claims is transformed into a relationship on its distribution and, consequently, on its parameters, represented by  $f(\theta) = u$ . Then, the prior is constructed so that its level curves coincide with those of the function  $f$ . This approach is based on the logic that two parameters with the same image should be weighted similarly when the only available information is the value of a function evaluated at the parameters. Moreover, between two possible parameterizations, the one whose image is closest to the level specified by the expert should be considered more probable. Following logical reasoning, these observations determine the shape of the prior as a composition of a distance between  $f(\theta)$  and  $u$  with a decreasing positive function.

In the case of the study, the decreasing function and the distance are chosen by analogy with a normal distribution, thus giving the prior distribution as  $\pi(\theta) \propto \exp\left(-\frac{(f(\theta)-u)^2}{\tau}\right)$  where  $\tau$  is a parameter to be set by the modeler. An example calibration is given for a judgment concerning reinsurance cost.

The analysis demonstrates that the use of the previously defined prior yields several results :

- The prior distribution of  $f(Claims)$  is centered around the expert-specified level  $u$ .
- The uncertainty parameter  $\tau$  controls to what extent parameters that deviate from the expert judgment can be taken into account.
- Simultaneously, the parameter  $\tau$  regulates the relative weight of prior information compared to data in forming the posterior.
- The posterior predictive distributions are consistent in form with those obtained by frequentist methods, attesting to their realism. Furthermore, adding a prior tends to align the predictive distribution with the expert’s judgment.

The proposed method effectively translates external information into the form of a parameter distribution. Its single parameter to set makes it more intuitive and manageable.

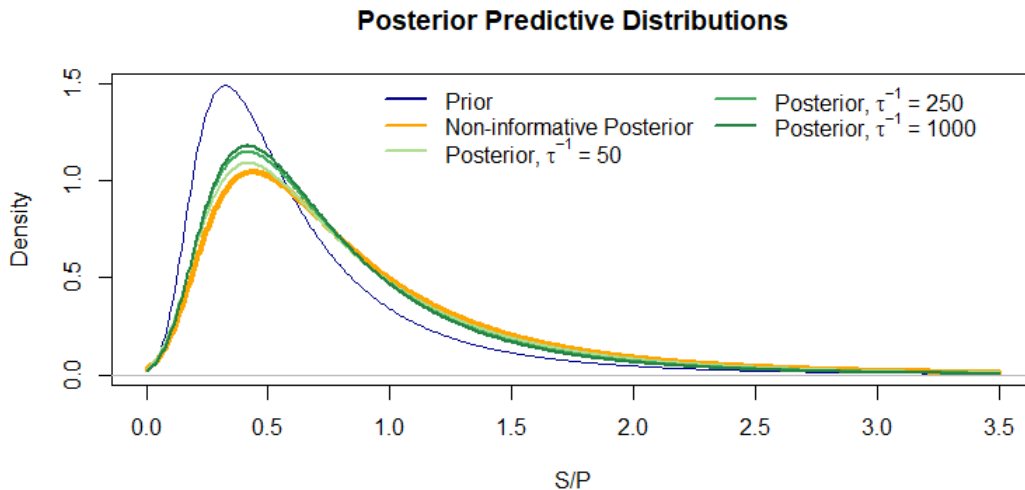


FIGURE 4 – Examples of predictive distributions with external data integration.

However, this approach has several limitations. The choice based on an analogy with a normal distribution is arbitrary, and it would be interesting to reflect on ways to guide it more rationally. Moreover, integrating external data places historical data and assumptions on the same level, which may raise concerns in calculations as critical and regulated as those of the Solvency Capital Requirement. A preliminary reflection on the necessary governance to oversee this integration is proposed at the end of the study.

Nevertheless, the process described in this thesis could find a relevant application in an underwriting context, especially in areas such as collective health insurance. Indeed, it could reconcile the results of the company’s pricing model with the client’s historical claims accounting data when pricing a new contract.

In conclusion, this study highlights two possible contributions of Bayesian statistics to actuarial science. Firstly, it provides an analysis of possible methods to assess the impacts of uncertainty on parameters, illustrated by an application to

Groupama's data that demonstrates the control of this risk in the calculation of its premium risk. Secondly, the study presents a method for effectively integrating external data into actuarial modeling, opening interesting perspectives despite the associated governance challenges. In summary, Bayesian statistics emerge as a powerful tool to enhance the accuracy of actuarial models. It offers promising applications, for example in the underwriting field, and can be particularly relevant in quantifying emerging risks such as climate or cyber risk, where traditional claims histories prove insufficient.

# Remerciements

Je tiens à exprimer mes remerciements les plus sincères à toutes les personnes qui ont joué un rôle essentiel dans la création de ce mémoire d'actuariat.

En premier lieu, un immense merci à Youva MANSOUR, mon tuteur en entreprise, dont l'expertise, les conseils avisés, la disponibilité et l'implication dans mon encadrement ont été des piliers tout au long de cette expérience professionnelle. Au delà de ses compétences professionnelles, je souhaiterais saluer ses qualités humaines d'écoute et de bienveillance.

Mes remerciements vont également à Nicolas BOUSQUET, mon tuteur académique, pour ses retours constructifs et les nombreuses ressources bibliographiques qu'il m'a fournies. Ma compréhension de la statistique bayésienne doit beaucoup à la qualité de son cours de M2.

Je souhaite exprimer ma reconnaissance envers Renaud BRUNEL, le directeur de la Direction des Risques Opérationnels et Contrôle Permanent pour son soutien, ses encouragements et l'intérêt qu'il porte aux travaux de ses équipes. Son expertise et ses conseils éclairés ont été une source d'inspiration tout au long de mon alternance.

Un remerciement particulier à tous les membres de l'équipe de Validation du Modèle Interne Non-Vie pour leurs nombreuses relectures mais aussi pour le climat de bonne ambiance qu'ils ont réussi à instaurer. Mon stage et mon alternance n'auraient pas été les mêmes sans les excellentes pâtisseries de Valentine NAPOLEON, les histoires rocambolesques de Bertrand BILONGI et la bonne humeur de Marie HERR. La qualité de vos travaux et votre professionnalisme ont rendu cette année passée à vos côtés riches en apprentissages.

À ma famille, je suis reconnaissant pour leur soutien indéfectible tout au long de mon parcours académique. Leurs encouragements constants et leur confiance inébranlable ont été une force motrice essentielle.

Enfin, ce mémoire est dédié à tous les acteurs du monde de l'assurance, qui travaillent chaque jour pour offrir une protection à ceux qui en ont besoin. Je suis fier de faire partie de cette communauté et je remercie tous ceux qui oeuvrent en permanence pour améliorer le quotidien des assurés.

# Table des matières

Table des matières	xv
Table des figures	xviii
Liste des tableaux	xx
<b>I Contextes de l'étude</b>	<b>3</b>
<b>1 Contexte réglementaire et professionnel</b>	<b>4</b>
1.1 Le modèle interne et l'équipe de validation	4
1.2 Méthodologie de calcul du risque de primes et du risque de souscription non-vie	6
1.2.1 Segmentation en groupes de risques homogènes	7
1.2.1.1 Segmentation par risque	7
1.2.1.2 Segmentation par gravité	8
1.2.2 Mise en as-if	9
1.2.2.1 Mise en as-if des sinistres graves	9
1.2.2.2 Mise en as-if de la sinistralité attritionnelle	11
1.2.3 Agrégation et simulations	11
<b>2 Contexte actuariel</b>	<b>14</b>
2.1 Incertitude sur paramètres	14
2.1.1 Définition	14
2.1.2 Non-diversification du risque de paramètres	15
2.1.2.1 Diversification des risques sans incertitude sur paramètres	16
2.1.2.2 Diversification des risques avec incertitude sur paramètres	16
2.2 Intégration de données externes	18
2.2.1 Sources de données	19
2.2.1.1 Données internes	19
2.2.1.2 Données externes	20
2.2.2 Méthodes d'incorporation de ces données	20
<b>3 Outils théoriques : rappels de statistique et comparaison de paradigmes</b>	<b>22</b>
3.1 La statistique fréquentielle	22
3.1.1 Démarche générale	22

3.1.2	Limites du paradigme fréquentiel . . . . .	23
3.2	Rappels de statistique bayésienne . . . . .	24
3.2.1	Une conception différente de la probabilité . . . . .	24
3.2.1.1	Principe d'actualisation . . . . .	24
3.2.1.2	L'inférence bayésienne . . . . .	25
3.2.2	La définition d'une fonction de coût . . . . .	26
3.2.2.1	Quantile de la loi prédictive d'un point de vue bayésien . . . . .	27
3.2.2.2	Quantiles prédictifs en raisonnant à partir de la vraisemblance . . . . .	28
3.2.3	L'élicitation d'une mesure a priori . . . . .	29
3.2.3.1	Règles de construction de mesures a priori objectives . . . . .	30
3.2.3.2	Construction d'a priori informatifs . . . . .	32
<b>II Corps de l'étude</b>		<b>34</b>
<b>4</b>	<b>Modélisation actuelle</b>	<b>35</b>
4.1	Processus de modélisation . . . . .	35
4.1.1	Méthode de calibrage . . . . .	35
4.1.2	Évaluation de l'incertitude sur paramètres . . . . .	37
4.2	Limites de la méthode . . . . .	37
4.2.1	Choix des lois . . . . .	37
4.2.1.1	Choix de la famille de distribution . . . . .	37
4.2.1.2	Choix des paramètres . . . . .	42
4.2.2	Incertainitude sur paramètres . . . . .	45
<b>5</b>	<b>Méthodologies alternatives</b>	<b>47</b>
5.1	Démarche générale . . . . .	47
5.2	Choix des lois a priori . . . . .	48
5.2.1	Encodage de jugements d'expert en a priori informatifs . . . . .	48
5.2.1.1	Typologie des jugements . . . . .	49
5.2.1.2	Quantiles donnés par l'expert . . . . .	49
5.2.1.3	Cas général, exemple des coûts de réassurance . . . . .	57
5.2.2	Loi non-informative . . . . .	60
5.3	Obtention de la loi prédictive a posteriori . . . . .	60
5.3.1	Algorithme de simulation . . . . .	61
5.3.1.1	Présentation de l'algorithme . . . . .	61
5.3.2	Simulation selon la loi prédictive a posteriori . . . . .	62
5.3.2.1	Metropolis-Hastings direct . . . . .	63
5.3.2.2	Simulations par composition . . . . .	63
<b>6</b>	<b>Mise en pratique</b>	<b>65</b>
6.1	Données disponibles . . . . .	65
6.1.1	Historiques utilisés . . . . .	65
6.1.2	Jugements d'expert . . . . .	66
6.2	Modélisation bayésienne avec a priori objectifs . . . . .	66
6.2.1	Obtention de la loi prédictive . . . . .	66
6.2.2	Évaluation de l'impact de l'incertitude sur paramètres . . . . .	69
6.2.2.1	Loi prédictive bayésienne . . . . .	69



6.2.2.2	Quantiles prédictifs par la vraisemblance . . . . .	73
6.3	Modélisation avec a priori informatifs . . . . .	77
6.3.1	Paire de quantiles . . . . .	77
6.3.1.1	Troncature de l'ensemble des paramètres . . . . .	77
6.3.1.2	Validation des propriétés du prior . . . . .	77
6.3.1.3	Sensibilités aux paramètres $\lambda_1$ et $\lambda_2$ . . . . .	78
6.3.2	Méthode générale . . . . .	80
6.3.2.1	Troncature de l'espace des paramètres . . . . .	80
6.3.2.2	Validation des propriétés du prior . . . . .	81
6.3.2.3	Sensibilité à l'hyperparamètre $\tau$ . . . . .	82
6.3.2.4	Conclusion . . . . .	87
	<b>Conclusion</b>	<b>89</b>
	<b>Appendices</b>	<b>92</b>
	<b>A Annexe calculatoire</b>	<b>92</b>
A.1	Calcul de la fonction de coût . . . . .	92
A.2	Calcul du gain de diversification . . . . .	93
A.2.1	Sans incertitude sur paramètres . . . . .	93
A.2.2	Avec incertitude sur paramètres . . . . .	93
A.3	Équations des iso-quantile pour une loi log-normale . . . . .	94
A.4	Calcul du gap de Jensen . . . . .	95
	<b>B Définitions mathématiques</b>	<b>96</b>
B.1	Définitions de lois . . . . .	96
B.1.1	Loi log-normale . . . . .	96
B.1.2	Loi gamma . . . . .	96
B.1.3	Loi log-normale . . . . .	96
B.1.4	Loi normale . . . . .	96
B.1.5	Loi de Pareto . . . . .	96
B.2	Définitions de tests statistiques . . . . .	97
B.2.1	Test de Kolmogorov-Smirnov . . . . .	97
B.2.2	Test d'Anderson-Darling . . . . .	97
B.2.3	Test de Kuiper . . . . .	97
	<b>C Article 230 du règlement délégué 2.A de Solvabilité II</b>	<b>98</b>
	<b>Bibliographie</b>	<b>99</b>

# Table des figures

1	Distribution prédictive avec prise en compte de l'incertitude sur paramètres. Le logarithme des prédictions est distribué selon une loi normale. . . . .	v
2	Exemples de distributions prédictives avec intégration de données externes. . . . .	vii
3	Predictive distribution with consideration of parameter uncertainty. The logarithm of predictions follows a normal distribution. . . . .	x
4	Examples of predictive distributions with external data integration. . . . .	xii
1.1	Modules de risque couverts dans le modèle interne partiel de Groupama	5
1.2	Position du risque de réserve et de primes sur un triangle Survenance x Développement . . . . .	6
1.3	Liste des lignes d'activités de la formule standard de Solvabilité II . . . . .	7
1.4	Mise en as-if des paiements d'un sinistre . . . . .	10
1.5	Mise en as-if d'un sinistre . . . . .	10
1.6	Processus de calibrage du risque de souscription non-vie . . . . .	13
2.1	CoV de la sinistralité en fonction du volume . . . . .	18
4.1	Résultats des tests d'adéquations avec des lois aux paramètres fixes ou calibrés sur les données . . . . .	39
4.2	Adéquation des différentes lois selon le test de Kolmogorov-Smirnov . . . . .	40
4.3	Résultats des tests d'adéquations pour des données générées selon une loi log-normale . . . . .	41
4.4	Résultats des tests d'adéquations pour des données générées selon une loi de Pareto . . . . .	42
4.5	Calibrage des paramètres, n=15, 20 et 25 . . . . .	44
4.6	Répartition empirique des écarts relatifs calibrage/simulation, n=25 . . . . .	45
5.1	Courbes iso-quantiles pour loi log-normale . . . . .	50
5.2	Répartition de la densité . . . . .	52
5.3	Longueur des demi-axes en fonction des valeurs propres . . . . .	53
5.4	Propositions de prior basés sur les quantiles de l'expert . . . . .	56
5.5	Coût de la réassurance en fonction des paramètres . . . . .	59
5.6	Log-densité du prior en fonction de $\tau$ . . . . .	60
6.1	Vraisemblance et a priori . . . . .	67
6.2	Représentation de la loi a posteriori . . . . .	67
6.3	Diagnostic de Gelman-Rubin à trois chaînes . . . . .	68
6.4	Autocorrélogramme des simulations . . . . .	69

6.5	Distribution des prédictions avec incertitudes sur paramètres, méthode 1 . . . . .	70
6.6	Comparaison de la distribution du log des prédictions avec et sans incertitude sur paramètres, méthode 1 . . . . .	70
6.7	Distributions des prédictions avec incertitudes sur paramètres, méthode 1 . . . . .	71
6.8	Écarts relatifs des quantiles entre la méthode paramétrique et la méthode 1 . . . . .	72
6.9	Distribution des prédictions avec incertitudes sur paramètres, méthode 2 . . . . .	73
6.10	Distribution des prédictions avec incertitudes sur paramètres . . . . .	74
6.11	Distribution des prédictions avec incertitudes sur paramètres . . . . .	74
6.12	Écarts relatifs des quantiles entre la méthode paramétrique et la méthode 2 . . . . .	75
6.13	Distributions des quantiles d'intérêts représentées sous forme de boîtes à moustache . . . . .	78
6.14	Distributions des quantiles d'intérêts représentées sous forme d'histogrammes . . . . .	79
6.15	Coûts de la réassurance et avis de l'expert . . . . .	80
6.16	Coûts de la réassurance associés aux paramètres simulés . . . . .	81
6.17	Moyenne et volatilité des coûts de réassurance en fonction de $\tau$ . . . . .	82
6.18	Distribution des quantiles prédictifs, prior informatif . . . . .	83
6.19	Posterior en fonction du paramètre $\tau$ . . . . .	84
6.20	Taux d'acceptation en fonction du paramètre $\tau$ . . . . .	85
6.21	Distributions prédictives a posteriori en fonction du paramètre $\tau$ . . . . .	86

# Liste des tableaux

6.1	Valeur des kurtosis normalisés pour plusieurs segments . . . . .	74
6.2	Probabilité d'acceptation d'un saut en fonction de $\tau$ . . . . .	85
6.3	Coût prédictif a posteriori de réassurance pour différents prior . . . . .	87

# Introduction

L'étymologie du mot "assurance", dérivé du latin "securus", renvoie à l'idée fondamentale de sécurité. Depuis l'Antiquité tardive, ce terme a évolué pour désigner de nos jours la compagnie d'assurance, qui est le prestataire de services couvrant, moyennant le paiement préalable d'une prime, son client contre la réalisation d'un risque futur pesant sur sa personne. Ainsi, l'assureur est bien celui qui vend de la sécurité à son assuré.

La démarche même de l'assurance, c'est-à-dire de payer d'avance une prime pour se prémunir contre la réalisation d'un risque futur, inscrit ce secteur à contre-courant du reste de l'industrie : son cycle de production est inversé, la vente a lieu avant de connaître la nature exacte de la prestation concernée. L'incertitude pesant sur les risques portés par les assureurs est donc intrinsèque à leurs activités. Peu importe le travail des actuaires ou des autres corps de métiers, ils ne pourront jamais la faire disparaître, le hasard est inhérent à l'assurance.

Émerge alors la nécessité pour les assureurs d'évaluer le plus finement possible leurs engagements afin de constituer les réserves nécessaires pour tenir leurs obligations. Cette évaluation quantitative n'est pas à l'entière discrétion des compagnies. En France, comme dans le monde entier, les compagnies d'assurances sont soumises à une réglementation stricte visant à protéger à la fois les assurés et l'ensemble du secteur économique, que la faillite d'une compagnie d'assurance mettrait en difficulté. Parmi ces textes de loi se trouve la réforme Solvabilité 2, en vigueur sur le marché européen, qui encadre la gestion des risques des assureurs en institutionnalisant, entre autres, la constitution d'une réserve de fonds propres devant permettre de faire face à un événement bicentenaire.

L'estimation de ce montant de fonds propres représente un défi majeur pour les actuaires européens et s'inscrit dans la continuité de leurs autres missions d'évaluation et de suivi des risques. Pour ce faire, la réglementation laisse plusieurs options aux assureurs. Soit ils utilisent une formule standardisée sur l'ensemble du marché européen ; soit, comme c'est le cas dans ce mémoire, ils utilisent des méthodes qui leur sont propres au sein d'un modèle dit interne. Dans tous les cas, les actuaires se sont traditionnellement appuyés sur des méthodes statistiques dites classiques. Celles-ci reposent sur un historique d'observations du phénomène modélisé pour produire des valeurs numériques, appelées paramètres, censées décrire les différentes caractéristiques du risque. Par la suite, les paramètres sont utilisés tels quels, comme s'ils décrivaient parfaitement le phénomène, même lorsque la profondeur de l'historique ne permet pas une inférence robuste.

Deux questions apparaissent alors assez naturellement : l'incertitude d'estimation de ces paramètres pourrait-elle avoir une incidence matérielle sur le montant de fonds propres réglementaires à thésauriser ? Par ailleurs, serait-il possible d'enrichir les historiques traditionnels avec d'autres données, afin d'améliorer le processus d'inférence et, par conséquent, réduire l'incertitude pesant sur les paramètres.

Pour répondre à ces deux problématiques, ce mémoire se basera sur une approche dite bayésienne des statistiques. En effet, le cadre bayésien est tout indiqué pour

intégrer des informations extérieures aux historiques traditionnels, notamment par le biais de l'élicitation d'une loi a priori. De plus, la restitution de l'inférence non pas sous la forme d'un unique paramétrage, mais sous la forme d'une loi de probabilité sur l'ensemble des paramétrages possibles, ouvre la voie à une estimation rationnelle de l'impact des incertitudes de calibrage.

Après avoir exposé plus en détail l'incertitude pesant sur les paramètres résultants des processus d'inférence classiques, il s'agira, dans un premier temps, de développer des méthodes permettant d'évaluer l'impact de cette incertitude sur la distribution de la sinistralité. Ensuite, dans un second temps, une réflexion sera menée sur la typologie des données pouvant venir compléter les historiques traditionnels ainsi que sur les manières de les incorporer aux analyses actuarielles.

L'étude se focalise sur l'assurance non-vie, en particulier sur l'évaluation du risque de primes, associé aux engagements de l'année en cours dont l'assureur n'a donc pas encore une connaissance détaillée.

Première partie  
Contextes de l'étude

# Chapitre 1

## Contexte réglementaire et professionnel

### 1.1 Le modèle interne et l'équipe de validation

Les compagnies d'assurance en France comme dans le monde sont soumises à des réglementations strictes visant à encadrer leur activité dans le but de protéger leurs assurés. Groupama, une mutuelle française, est notamment soumise à Solvabilité II, un ensemble de normes et de mesures mises en place par l'Union Européenne et visant principalement à affermir la solvabilité des acteurs du secteur de l'assurance européenne. Pour ce faire, ce règlement rassemblant de nombreuses exigences, aussi bien qualitatives que quantitatives, impose aux compagnies d'assurance d'avoir un niveau de fonds propres suffisant pour respecter leurs engagements envers les assurés, même en cas d'imprévu. Ce niveau de fonds propres, nommé SCR d'après Solvency Capital Requirement, doit permettre aux assureurs d'avoir moins d'une chance sur 200 de faire faillite dans l'année.

La réglementation propose une formule dite "standard" pour déterminer la valeur du SCR mais cette formule a été calibrée sur l'ensemble du marché européen et de ce fait n'a pas de raison d'être pleinement adaptée aux spécificités du profil de risque de chaque assureur. Pour cette raison, il est permis aux sociétés d'assurance de proposer un modèle interne propre à l'entreprise pour le calcul de leur SCR. Ce modèle doit être préalablement homologué par le superviseur, l'Autorité de Contrôle Prudentiel et de Résolution (ACPR) en France, et faire l'objet de contrôles réguliers.

Groupama a fait le choix d'adopter un modèle interne partiel (MIP), c'est-à-dire d'utiliser la formule standard complétée par un modèle interne dans les branches où la formule standard ne permet pas d'appréhender au mieux son profil de risque. Les modules ainsi modélisés (voir figure 1.1) sont ceux relatifs à l'activité d'assurance non-vie et d'assurance santé non similaire à de la vie (*Health non Similar to Life Techniques* en anglais).



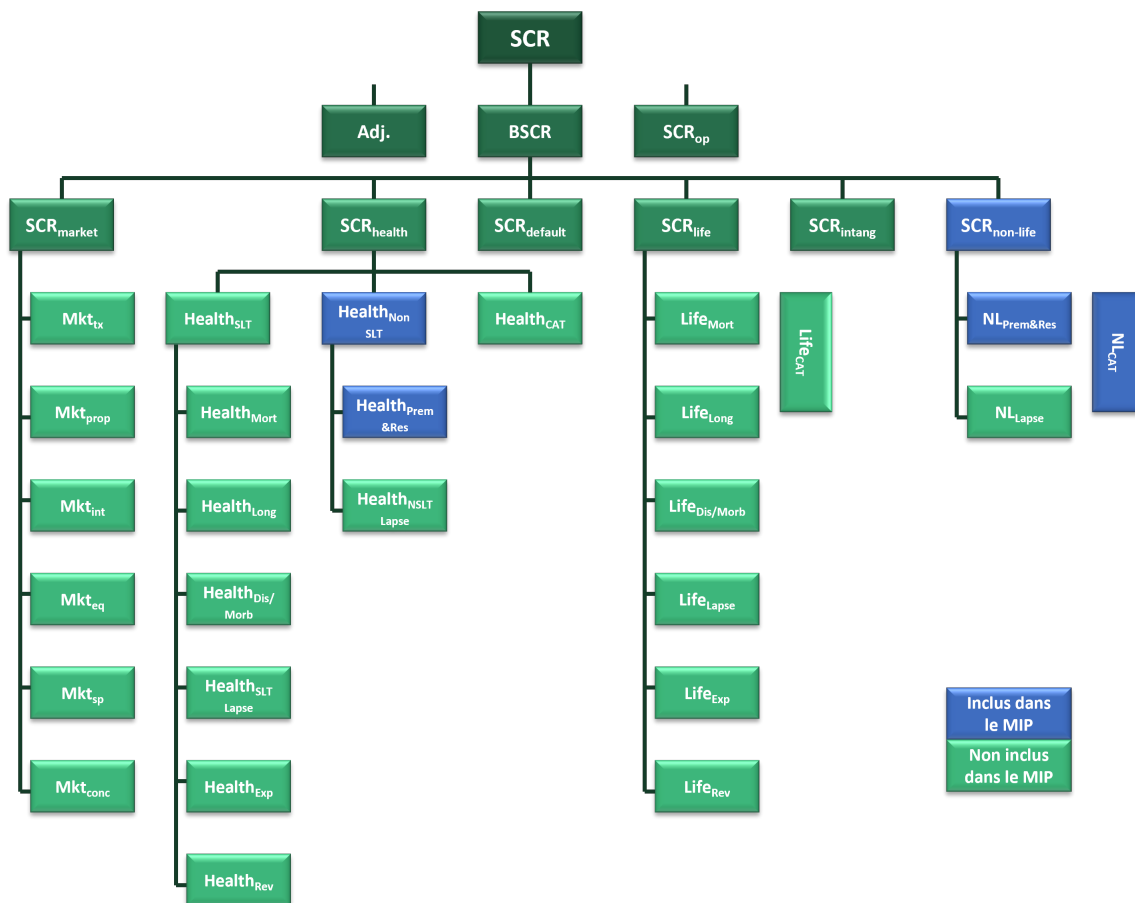


FIGURE 1.1 – Modules de risque couverts dans le modèle interne partiel de Groupama. Les modules de risque inclus dans le modèle interne partiel de Groupama sont figurés en bleu.

Le rôle de l'équipe de validation du modèle interne non-vie dans laquelle a été rédigé ce mémoire est de contrôler les travaux des autres équipes travaillant sur le MIP afin de veiller à sa pertinence, sa rigueur et sa bonne justification tout en prenant en compte les recommandations émises par l'ACPR lors de ses contrôles. Contrairement à l'équipe de modèle interne non-vie rattachée à la Direction Actuariat Groupe (DAG), l'équipe de validation est rattachée à la Direction Risques Contrôle et Conformité Groupe (DRCCG) et plus particulièrement à la sous-direction des Risques Opérationnels et Contrôle Permanent (DROCP). L'équipe de validation est ainsi pleinement indépendante de la direction actuariat, condition nécessaire pour mener à bien ses missions.

Le calcul du SCR n'est pas directement réalisé pour l'ensemble des activités du groupe, chaque ligne d'activité (LoB pour « *line of business* » en anglais) est calculée séparément avant d'être agrégée au calcul groupe. Sont distinguées, au sein de chaque LoB, deux composantes au calcul du SCR : le risque de réserve et le risque de prime. Le risque CAT est quant à lui incorporé au risque de primes chez Groupama. Le risque de réserve concerne le risque lié aux sinistres des années de survenance antérieures. Il traite en particulier de l'incertitude quant à liquidation, ie. des dates et des montants des paiements futurs, des sinistres déjà survenus. Le risque de primes, auquel l'essentiel de ce mémoire est consacré, traite du risque relatif aux sinistres non encore survenus, aux primes devant être émises au cours de

la période, et aux risques en cours sur la période résiduelle des contrats existants. Il inclut également, dans le MIP de Groupama, l'ensemble des risques catastrophes (CAT).

Sur un triangle de charge classique "Année de survenance x Année de développement", la partie bleue de la nouvelle diagonale entre dans le risque de réserve tandis que la partie verte appartient au risque de primes :

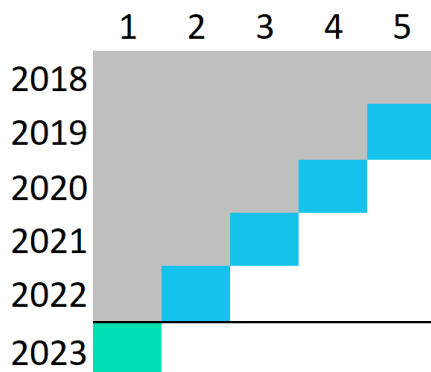


FIGURE 1.2 – Position du risque de réserve et de primes sur un triangle Survenance x Développement

La sinistralité associée au risque de primes n'étant pas encore observée, il existe une inconnue sur l'application des traités de réassurance qui empêche de passer simplement d'une distribution globale de la sinistralité brute à une distribution du résultat technique. Par conséquent, une approche par simulation a été privilégiée. Il est alors nécessaire de recourir à des lois de probabilité pour modéliser plusieurs éléments de la sinistralité future, incluant la charge, l'inflation et l'interdépendance entre les risques. D'autres éléments ayant un impact sur le résultat technique comme les primes ou les frais sont modélisés de manière déterministe. La méthode générale du calcul de SCR de primes est décrite de manière détaillée dans une documentation interne dont les grandes lignes seront résumées dans la section suivante.

## 1.2 Méthodologie de calcul du risque de primes et du risque de souscription non-vie

La détermination du montant de capital de solvabilité requis peut se décomposer en plusieurs processus successifs, regroupant principalement le calibrage de sous-groupes de risques homogènes, suivi par l'agrégation de ces modules. Bien que seul le risque de primes, et donc qu'une partie des sous-groupe homogène, soit étudié dans ce mémoire, les enjeux de son calcul ne peuvent se comprendre qu'en le replaçant dans le cadre plus large du calcul risque de souscription non-vie (bloc "SCR non-life" de la figure 1.1).

Cette section n'a pas pour objectif de présenter les modèles actuariels employés mais de contextualiser le calcul du risque de primes pour en comprendre tant les objectifs que les contraintes. Ainsi, après avoir exposé les grands principes méthodologiques

gouvernant le calibrage du risque de souscription non-vie dans toute sa généralité, ce mémoire se concentrera sur le premier processus exposé, celui du calibrage des distributions marginales, et plus spécifiquement celles liées au risque de primes.

## 1.2.1 Segmentation en groupes de risques homogènes

L'ensemble des risques assurés par une mutuelle comme Groupama est varié. Afin de pouvoir proposer une modélisation pertinente de chacune des activités de la compagnie, il est nécessaire de regrouper ces différentes branches en des familles cohérentes.

### 1.2.1.1 Segmentation par risque

Ces ensembles sont en premier lieu basés sur la typologie du risque assuré. Dans sa formule standard, la norme Solvabilité II propose la maille en ligne d'activité, dite maille "LoB", pour le calcul du SCR de souscription non-vie. Une liste de ces lignes d'activités est donnée en la figure 1.3.

Ses éléments les plus importants sont :

- FIRE : LoB contenant l'ensemble des dégâts aux biens comme les habitations ou les cultures agricoles ;
- MOTOR : LoB réunissant les assurances dommages aux véhicules ;
- MTPL : Lob regroupant les garanties de Responsabilité Civile Automobile.

Comme énoncé dans la section 1.1, chaque LoB est à son tour découpée en une composante de risque de réserve et une de risque de primes.

01	Medical expenses
02	Income protection
03	Worker's compensation
04	Motor Third Party Liability
05	Motor other
06	Marine, aviation and transport
07	Fire and other property
08	Third-party liability
09	Credit and surety ship
10	Legal expenses
11	Assistance
12	Miscellaneous

FIGURE 1.3 – Liste des lignes d'activités de la formule standard de Solvabilité II

Cependant, le découpage en lignes d'activités Solvabilité II reste encore une segmentation trop grossière pour pouvoir offrir une modélisation satisfaisante de chacun de ses segments. Pour certaines LoB, la sinistralité est encore subdivisée en fonction de la nature du risque assuré.

Par exemple, la sinistralité associée aux assurances habitation est traitée de manière distincte de celle des contrats multirisques climatiques, couvrant les exploitations

agricoles, bien que toutes deux relèvent de la LoB FIRE. En outre, pour un même objet assuré, la modélisation peut dépendre du fait générateur : les dommages dus aux tempêtes sur un bâtiment et ceux résultant de la grêle sur un bâtiment sont considérés comme des segments distincts.

### 1.2.1.2 Segmentation par gravité

Dans une pratique courante du monde de l'actuariat, les sinistres sont ensuite séparés en fonction de leur fréquence et de leur sévérité. Chaque groupement peut alors être caractérisé par des paramètres de fréquence et de sévérité pertinents, ouvrant ainsi la voie à une meilleure compréhension et anticipation des pertes potentielles futures. De plus, les contraintes liées à l'application des traités de réassurance imposent également un maillage des engagements de l'entreprise en fonction de leurs coûts. Les sinistres sont répartis en trois catégories :

- les sinistres attritionnels caractérisés par leur faible montant et leur fréquence élevée ;
- les sinistres graves, moins fréquents mais plus coûteux ;
- les sinistres dits "catastrophes", exceptionnels tant par la rareté de leur survenance que par l'importance de leur coût.

Chacune de ces catégories dispose de modélisations et de typologies de données qui lui sont propres.

Pour des besoins d'application de la réassurance, les sinistres graves et catastrophes font habituellement l'objet de simulations individuelles via un modèle "Fréquence x Coût". Un premier tirage, selon une variable de fréquence, donne un nombre  $N$  qui correspond au nombre de simulations à faire selon la loi du coût. Les  $N$  tirages subséquents, réalisés conformément à la loi de coût, représentent chacun un sinistre grave survenu au cours de la période d'exposition modélisée.

Cependant, en raison de la rareté, voire de l'absence d'historique, ou encore des contraintes liées à l'application de la réassurance, la sinistralité catastrophe peut être amenée à être calibrée à l'aide de jugements d'expert.

La sinistralité attritionnelle n'est pas directement modélisée, mais est implicitement captée à travers le ratio charge attritionnelle totale à primes noté  $S/P$ . L'avantage de traiter ces deux quantités de manière agrégée est de pouvoir capter à la fois la variabilité de la charge de sinistre attritionnelle et la variabilité des primes historiques.

Si le maillage fin des engagements a pour finalité d'aboutir à des ensembles de risques homogènes, encore faut-il être en mesure de comparer, au sein d'une même maille, les différents points de l'historique. Généralement, les données disponibles ne satisfont pas à cette exigence sous leur forme brute car chacune est rattachée à une certaine année de survenance disposant de caractéristiques qui lui sont propres. Il convient alors d'opérer une normalisation de ces spécificités de sorte à disposer d'un échantillon de données comparables, ce qui constitue une hypothèse de base des modèles actuariels les plus couramment utilisés.

Le processus de mise en conformité par un retraitement des années d'historique est couramment désigné sous le terme de "mise en as-if", dont le sens va être approfondi dans la section suivante.

## 1.2.2 Mise en as-if

Divers facteurs, tels qu'une évolution dans la tarification, la souscription ou l'environnement économique, qui dépendent de l'année et sont connus dans une certaine mesure au moment de la modélisation, peuvent impacter les charges de sinistralité. Plus spécifiquement, ces facteurs peuvent inclure, une augmentation des franchises, une modification des garanties, une inflation générale ou sectorielle, une évolution dans la population assurée ou encore des variations dans les coûts d'indemnisation.

Les retraitements, ou leur absence, à appliquer aux données en amont de la modélisation doivent être soigneusement considérés, car ils peuvent impacter lourdement la qualité et la précision des prévisions en sortie du modèle. Par ailleurs, il est important de reconnaître que ces facteurs peuvent être interdépendants et donc affecter simultanément la sinistralité. Il s'impose alors d'adopter une approche globale et cohérente dans la sélection de ces facteurs, ainsi que dans le choix des méthodes à retenir pour les retranscrire lors du processus de mise en as-if.

Cependant, il faut également noter que dans certains cas, il est préférable de ne pas effectuer de retraitement plutôt que de retraiter partiellement les données, au risque de déformer arbitrairement l'historique. L'essentiel est donc d'analyser attentivement les données et leur contexte historique avant de décider s'il convient d'appliquer un retraitement et, le cas échéant, dans quelle mesure l'appliquer. Cette analyse doit être menée avec rigueur afin d'éviter des erreurs pouvant mener à des biais significatifs résultant d'une mauvaise exploitation des données.

Ces études, dont seules les conclusions seront énoncées dans ce mémoire, ont été conduites avec attention par les équipes de la Direction Actuariat Groupe de Groupama et validées par celle de Validation. Il a été établi, sans surprise, qu'il était opportun d'opérer des retraitements différents en fonction du segment modélisé, en distinguant en fonction du péril et de la gravité des sinistres.

### 1.2.2.1 Mise en as-if des sinistres graves

Classiquement, les sinistres graves sont redressés d'un indice d'inflation, de sorte à adapter leurs montants à l'environnement de prédiction considéré. La méthode retenue par Groupama est dite rétrospective, elle traite séparément les charges (somme des paiements et des provisions) et les paiements. Dans un premier temps, un indice déterministe d'inflation future est appliqué aux paiements, de sorte à les translater dans le temps conformément à la figure 1.4.

La somme des paiements retraités permet alors d'obtenir la dernière charge connue lorsque le sinistre est clos. S'il est encore en cours, les provisions à date sont assimilées à un paiement effectué l'année suivante. Dans un second temps, les provisions sont obtenues à partir de la dernière charge connue : seule la dernière vision de la charge est retraitée de l'inflation, les autres s'obtiennent en répliquant les boni-mali historiques.

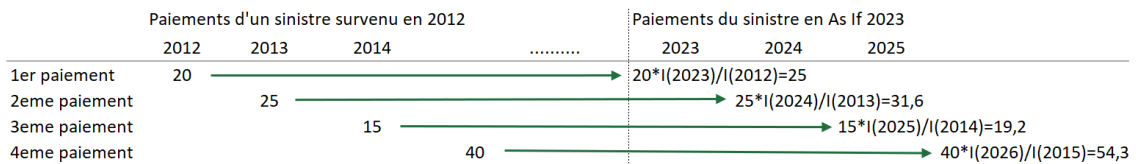


FIGURE 1.4 – Mise en as-if des paiements d'un sinistre

Concrètement, pour un sinistre survenu en 2012 et clôturé en 2015, la mise en as-if à la clôture 2022 se déroulera selon les deux étapes décrite dans la figure 1.5.

Ainsi :

- Chaque paiement est vieilli comme si le sinistre était survenu en 2023.
- Seule la dernière charge est retraitée de l'inflation, en fonction de l'année de survenance du sinistre et de l'année de vision de la dernière charge connue. Par charge il est entendu la somme des paiements et des provisions.
- Les boni/mali constatés historiquement sont répliqués en partant de la dernière charge inflatée et en remontant jusqu'au premier montant connu, de proche en proche il est possible de reconstruire un historique complet du sinistre. Il est à noter que les mali sont répliqués multiplicativement : si la charge d'un sinistre a historiquement été multiplié par 1,2 entre l'année  $N$  et  $N - 1$ , alors la charge  $N - 1$  en as-if s'obtient en divisant la charge  $N$  mise en as-if par ce même facteur.

Etape 1		Année de développement	1	2	3	4
Sinistre original	Année	2012	2013	2014	2015	
	Paiement	20	25	15	40	
	Provision	70	54	50	0	
	Charge	90	99	110	100	
	Charge N-1/N	//	91%	90%	110%	
Sinistre mis en as-if	Année	2023	2024	2025	2026	
	Paiement	25	31,6	19,2	54,3	
	Provision					
	Charge				130,1	

↓

Etape 2		Année de développement	1	2	3	4
Sinistre original	Année	2012	2013	2014	2015	
	Paiement	20	25	15	40	
	Provision	70	54	50	0	
	Charge	90	99	110	100	
	Charge N-1/N	//	91%	90%	110%	
Sinistre mis en as-if	Année	2023	2024	2025	2026	
	Paiement	25	31,6	19,2	54,3	
	Provision	92,1	72,2	67,3	0	
	Charge	117,1	128,8	143,1	130,1	

FIGURE 1.5 – Exemple de mise en as-if d'un sinistre

L'avantage de cette méthode réside dans le fait qu'elle ne crée pas implicitement de boni ou de mali non observé sur les sinistres historiques : tout boni/mali constaté historiquement est conservé sur les données retraitées.

### 1.2.2.2 Mise en as-if de la sinistralité attritionnelle

La mise en as-if de la sinistralité attritionnelle est plus délicate à mettre en oeuvre. Cette dernière étant modélisée à partir du ratio  $S/P$ , il faut produire des règles de retraitement à la fois pour les primes et pour les paiements.

Plusieurs procédures de mise en as-if peuvent être envisagées :

- Retraiter l'échantillon de  $S/P$  des variations historiques d'expositions en s'appuyant potentiellement sur un indicateur comme le nombre de contrat pouvant être pertinent pour des risques de masses. Cela permettrait de retraiter l'exposition tout en évitant de capter des effets liés aux revalorisations tarifaires.
- Retraiter l'historique par un indice l'inflation. Bien qu'il soit tentant de penser que les revalorisations du numérateur et dénominateur peuvent se compenser, cette approche est pertinente lorsque les sinistres et les primes ne suivent pas le même indice comme c'est généralement le cas.
- Corriger les autres biais pouvant impacter la chronique des  $S/P$  tels que des modifications de franchises, de politique de souscription ou de taux de commissionnement.

En pratique, dans une démarche prudente, les ratios ne sont généralement pas retraités, en raison du manque de données de qualité suffisante et de la complexité de construire des indices reflétant de manière exhaustive et appropriée les facteurs, ainsi que leurs effets croisés, influant sur la chronique des  $S/P$ .

Une fois les données et les périmètres de calcul clairement définis, il est possible d'entreprendre les calibrages pour chaque segment. Les résultats de ces calculs actuariels se présentent sous la forme de distributions propres à chaque sous-bloc de risque. Cependant, le travail de l'actuaire ne s'arrête pas ici : il lui reste encore à consolider les distributions marginales entre elles, afin d'obtenir une distribution complète du résultat de l'entreprise.

La partie suivante traitera de cette dernière étape essentielle au calcul du risque de souscription non-vie sous Solvabilité II.

### 1.2.3 Agrégation et simulations

A l'issue de l'étape préliminaire et fondamentale du calibrage des sous blocs de risques homogènes, la question de l'agrégation de toutes les lois marginales se pose nécessairement. Par ailleurs, ces distributions sont calibrées en brut de réassurance, il est donc indispensable de modéliser l'application des traités de réassurance ce qui constitue une étape critique dans les processus de modélisation des risques financiers complexes.

Chez Groupama, ces deux tâches, l'agrégation et l'application de la réassurance, sont effectuées en plusieurs étapes au sein d'un même logiciel, *Risk explorer* qui sera abrégé *Rex*.

Comme mentionné en introduction, les traités de réassurances nécessitent parfois d'avoir le détail des événements sinistre par sinistre sur une année. Du fait de la complexité de leur application, cette étape est réalisée à l'aide de simulation de scénario de sinistralité.

L'agrégation, et donc la corrélation des blocs, se fait en premier lieu. En effet, les contrats de réassurances couvrant généralement plusieurs risques, la réalisation d'un évènement sur un certain segment peut influencer les cessions d'une autre branche. Il est donc nécessaire de disposer d'une distribution globale avant de vouloir prendre en compte la réassurance.

En entrée de ce processus d'agrégation se trouve naturellement la sortie du processus précédent : les distributions marginales. Le logiciel *Rex* accepte plusieurs types de données d'entrée, les lois peuvent être caractérisées soit sous une forme paramétrique par un type de loi avec son jeu de paramètres associé ou soit d'une manière non paramétrique à travers une fonction quantile évaluée sur un échantillon suffisamment large de points. Les corrélations entre les risques, en brut de réassurance, font également l'objet d'un calibrage et sont donc renseignées à cette étape. Ainsi, le logiciel a tous les ingrédients pour produire les simulations corrélées et en brut de réassurance.

La sortie de ce processus peut se voir comme une matrice de scénario de sinistralités brutes de réassurance à la même maille que celle définie lors des calibrages de la section 1.2.1.

Une fois ces tirages obtenus, l'étape de passage au net de réassurance peut débuter. En effet, pour un scénario fixé, il est possible de déterminer le reste à charge pour l'entreprise en appliquant les traités à la sinistralité tirée. L'application de ce processus à chaque scénario simulé abouti à une distribution des risques corrélés et nets de réassurance.

S'ensuit une dernière phase, essentiellement d'habillage des résultats, visant à mettre en forme les retours du logiciel afin d'en obtenir une meilleure lisibilité. Les processus de calibrage du risque de souscription non-vie sont schématisés sur la figure 1.6.

Comme annoncé précédemment, seul le processus de modélisation des sous-groupes homogènes sera abordé dans cette étude. Cette étape implique le déploiement de diverses méthodes de calculs statistiques dans le but de modéliser de la manière la plus adéquate possible les risques et les incertitudes portées par l'entreprise.



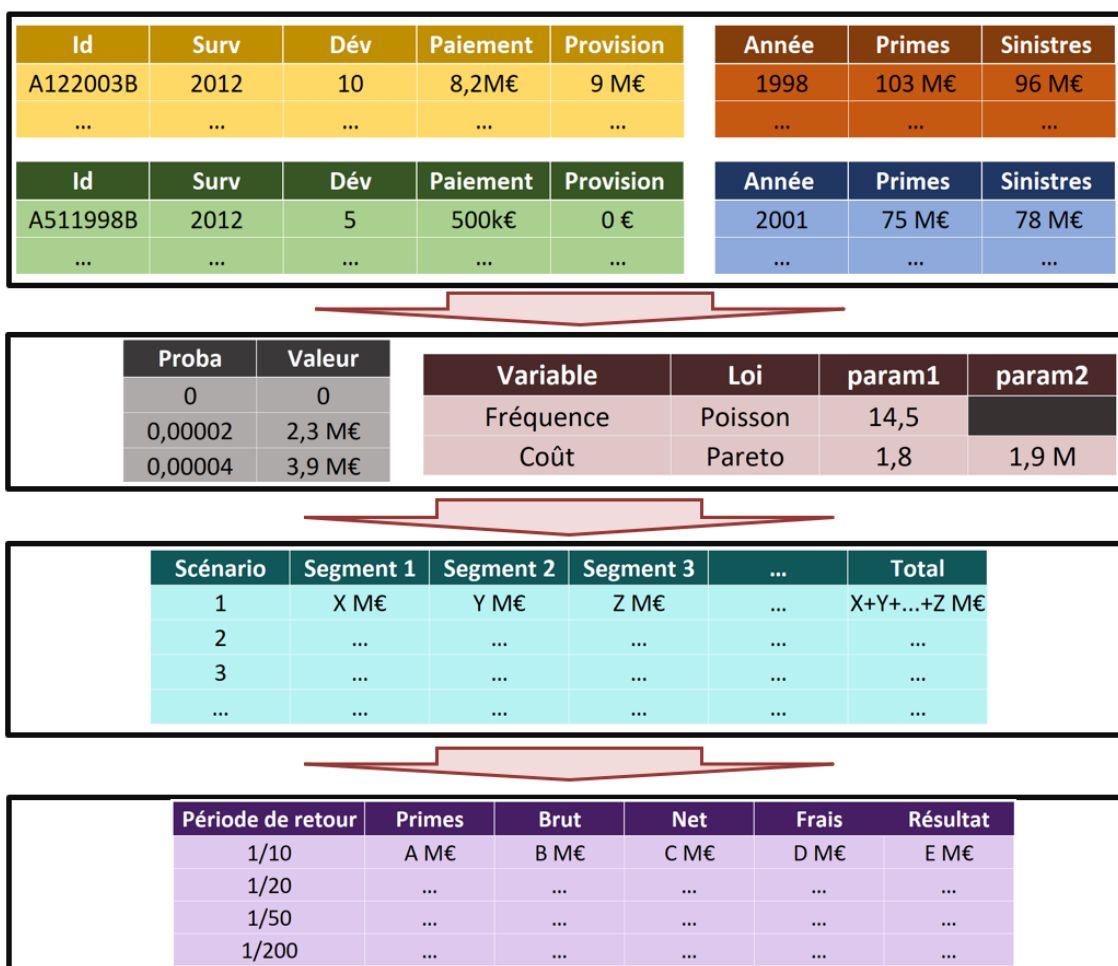


FIGURE 1.6 – Processus de calibrage du risque de souscription non-vie.

Le premier rectangle du graphique représente les historiques de sinistralité. À gauche, se trouve les historiques de sinistralités graves, comprenant l'identifiant du sinistre, sa date de survenance et son développement. À droite, se trouvent les historiques de sinistralités attritionnelles, où les données sont agrégées par années.

Par la suite, diverses méthodes statistiques sont appliquées aux historiques afin d'obtenir des distributions de ces risques. Ces distributions peuvent se présenter sous une forme paramétrique, telle que présentée dans le tableau brun, ou prendre la forme d'une fonction quantile, comme illustrée dans le tableau gris.

Une fois les distributions calibrées, des scénarios de sinistralités sont simulés en prenant en compte les corrélations entre les différents risques.

Pour finir, les contrats de réassurance sont appliqués aux simulations brutes. Divers éléments tels que les primes ou encore les frais sont alors intégrés, permettant ainsi d'obtenir une distribution du résultat net de l'entreprise.

Plusieurs manières d'appréhender l'incertitude entourant les résultats de la compagnie existent. Le cadre usuellement employé dans le monde de l'assurance est un cadre de statistique dit « fréquentiel », le cadre sur lequel cette étude se concentrera est quant à lui dit « bayésien ». Avant d'étudier de plus près les spécificités et les modes opératoires de ces deux paradigmes, le prochain chapitre visera à introduire et à formaliser les contextes actuariels des sujets abordés dans ce mémoire. Cela permettra de mettre en évidence les contributions potentielles de la statistique bayésienne à ces différents domaines de l'actuariat.

# Chapitre 2

## Contexte actuariel

En tant qu'actuaire, il est essentiel de comprendre et de relever les défis et les problématiques spécifiques auxquels l'industrie de l'assurance est confrontée. Deux de ces enjeux seront introduits et formalisés dans ce chapitre, fournissant ainsi les bases nécessaires pour les aborder de manière approfondie dans le corps de ce mémoire. Le premier défi concernera l'incertitude sur paramètres, tandis que le second relèvera de la question plus large de l'incorporation d'informations extérieures aux données.

### 2.1 Incertitude sur paramètres

À de rares exceptions près, la quasi-totalité des études de modélisation actuarielles menées sur le marché passent par la définition d'un modèle dont les paramètres doivent ensuite être estimés. Ne pas prendre en compte les incertitudes autour de ces estimations dans le processus de modélisation pourrait conduire à une mauvaise évaluation des risques, voire à une surestimation du gain de diversification au sein d'une même ligne d'assurance.

L'incertitude sur paramètres constitue un défi majeur dans le domaine de l'actuaire, car elle influence directement les résultats et les décisions prises dans le cadre de la tarification, du provisionnement, de l'évaluation de la marge de solvabilité et dans la gestion des risques en général. Sa maîtrise dans un cadre de prédiction, et donc plus spécifiquement au sein d'un modèle interne, fait même l'objet d'une obligation réglementaire à travers l'article 230 du règlement délégué 2.A de Solvabilité II disponible en Annexe C.

#### 2.1.1 Définition

L'incertitude ou risque sur paramètres correspond à l'incertitude quant à l'aptitude des valeurs estimées des paramètres à retranscrire le comportement du phénomène modélisé.

Il émane en grande partie de deux biais inhérents aux méthodes actuarielles. Le premier est le biais d'échantillonnage : les paramètres sont issus d'une estimation sur un échantillon fini qui ne contient pas toutes les caractéristiques de la population. Le deuxième biais est celui d'observation : l'échantillon provient d'une période d'exposition passée qui, malgré les divers retraitements, reste une époque différente

qui n'est pas pleinement représentative de l'environnement actuel de prédiction et ne peut contenir toutes les informations à son sujet.

L'incertitude sur paramètres ne comprend pas les risques liés à un mauvais choix de modèle qui sont eux affiliés aux risques de modèle. Contrairement au risque lié à la nature foncièrement aléatoire d'une compagnie d'assurance, qui sera nommé risque de contrat, le risque sur paramètres ne se diversifie pas avec le volume de souscription. Cette propriété essentielle et intuitive sera développée dans la section suivante.

### 2.1.2 Non-diversification du risque de paramètres

Au sein d'une ligne d'assurance donnée, il est avantageux pour un assureur de souscrire un grand nombre de contrats en raison des bénéfices associés à la mutualisation des risques, qui constitue une forme de diversification. En effet, souscrire un grand nombre de contrats permet à la compagnie de réduire la volatilité de son résultat en répartissant les risques sur un ensemble large et varié de souscripteurs.

Cette diversification contribue à atténuer l'impact potentiel des sinistres sur l'entreprise, car les pertes subies par certains assurés peuvent être compensées par les primes perçues des autres assurés. Deux phénomènes sont alors en concurrence : d'une part, l'augmentation du nombre de contrats induit une augmentation du nombre de sinistres à indemniser, tandis que d'autre part, la hausse des souscriptions permet de collecter davantage de primes pour les couvrir. Si le résultat espéré en moyenne ne varie pas lors d'un changement de volume, il s'avère que sous des hypothèses communes la dispersion autour de ce dernier diminue quand la souscription augmente.

La mutualisation est donc un outil fondamental à la disposition des compagnies d'assurances dans la gestion de leurs risques, permettant de réduire l'incertitude inhérente à leurs activités.

La manière selon laquelle ce gain est traduit dans un modèle peut être mise en lumière à l'aide du calcul de l'incertitude de la charge de sinistralité. Pour cela, il faut d'abord traduire la notion d'incertitude. Deux approches permettent d'aboutir au résultat, une consiste à utiliser la variance de la charge puis de la ramener à un nombre unitaire de contrats, l'autre à traduire mathématiquement la notion d'incertitude par un coefficient de variation qui est la mesure de dispersion relative de référence. Au final, ces deux approches sont équivalentes, seule celle basée sur le coefficient de variation sera présentée.

Ces calculs sont en grande partie inspiré de l'article *A note on Parameter Risk* de VENTER et SAHASRABUDDHE (2012)[5], à la différence près que le calcul du gain de diversification avec incertitude sur paramètre a été adapté à un contexte d'estimation de paramètres. Dans leur article, Venter et Sahasrabuddhe s'intéressait au cas particulier d'une perturbation multiplicative aléatoire symbolisant l'existence d'une tendance inconnu de la sinistralité. Leur calcul ne prenait donc pas directement en compte la volatilité des estimateurs utilisés dans les modèles de prédiction tels que ceux employés par Groupama.

### 2.1.2.1 Diversification des risques sans incertitude sur paramètres

L'objectif va être de déterminer le coefficient de variation de la variable aléatoire  $S$  représentant le montant total de sinistres. Cette dernière est définie à l'aide d'une variable  $N$  traduisant le nombre inconnu de sinistres et de  $X$  la variable du coût individuel d'un sinistre de la manière suivante :

$$S = \sum_{i=1}^N X_i$$

Ce modèle, de la forme fréquence x coût, est extrêmement commun dans le monde de l'assurance. Les  $X_i$  sont mutuellement indépendants, il est en outre supposé qu'ils sont indépendants de la variable  $N$ .

L'espérance de  $S$  est donnée par le calcul suivant :

$$\mathbb{E}[S] = \mathbb{E}[\sum_{i=1}^N X_i] = \mathbb{E}[\mathbb{E}[\sum_{i=1}^N X_i | N]] = \mathbb{E}[N\mathbb{E}[X]] = \mathbb{E}[N]\mathbb{E}[X]$$

La variance se détermine à partir du moment d'ordre 2, dont le calcul est donné en Annexe A.2.1, l'expression finale est :

$$Var[S] = \mathbb{E}[N]Var[X] + \mathbb{E}[X]^2Var[N]$$

Ce qui donne finalement :

$$CoV^2[S] = \frac{1}{\mathbb{E}[N]}CoV^2[X] + \frac{1}{\mathbb{E}[N]} \frac{Var[N]}{\mathbb{E}[N]}$$

Le coefficient de variation de la variable  $X$  ne dépend pas du nombre de contrats et le terme  $\frac{Var[N]}{\mathbb{E}[N]}$  qui représente le coefficient de dispersion est généralement constant pour une loi de comptage. Par exemple, pour une loi de Poisson, ce ratio est égal à 1, tandis que pour une loi binomiale négative, il vaut  $1/p$ . Pour rappel, la loi binomiale négative est la loi de comptage du nombre d'échecs avant d'obtenir  $n$  succès pour une famille de lois de Bernoulli indépendantes et identiquement distribuées, avec une probabilité de succès  $p$ .

Contrairement à l'espérance du nombre de sinistres, le coefficient de dispersion de la distribution varie généralement peu avec le nombre d'assurés, ou du moins il ne croit pas linéairement.

En conclusion, le coefficient de variation peut se réécrire  $CoV^2[S] = \frac{1}{\mathbb{E}[N]} \left( CoV^2[X] + \frac{Var[N]}{\mathbb{E}[N]} \right)$  et, puisque l'espérance du nombre de sinistres est proportionnelle au nombre de polices, le coefficient de variation va lui être inversement proportionnel au nombre d'assurés dans ce modèle. En outre, la branche d'assurance est alors asymptotiquement sans risque car  $CoV(S) \rightarrow 0$ . Tout le risque peut être diversifié.

Les résultats d'une telle modélisation ne semblent pas adéquats à la situation d'un assureur. En effet, la volatilité relative modélisée d'un grand portefeuille peut devenir illusoirement faible et inciter la compagnie à souscrire davantage de contrats sans qu'elle n'en mesure pleinement les conséquences.

### 2.1.2.2 Diversification des risques avec incertitude sur paramètres

En réalité, le modèle précédent ne prend en compte que le risque de contrat et néglige dans ses distributions prédictives d'autres types d'aléa dont notamment l'incertitude sur paramètres. Or, ce risque affecte tous les contrats à la fois, il n'a donc pas de raison d'être diversifiable avec le volume. Se tromper sur la fréquence moyenne

de survenance d'un sinistre n'est pas moins grave lorsque les volumes sont grands, bien au contraire même. Il est possible, par le même type d'exemple, d'illustrer l'impact de la prise en compte du risque sur paramètres de la manière suivante :

En admettant les deux conditions sur la variable de fréquence  $N$  :

$$\begin{cases} \mathbb{E}[N] \propto N_{ass} \\ \frac{\mathbb{V}[N]}{\mathbb{E}[N]} \propto 1 \end{cases}$$

il est possible de réécrire la variable  $N$  sous la forme :  $N = f \cdot N_{ass} + \eta \cdot \sqrt{N_{ass}} \cdot \epsilon$  avec  $f$  et  $\eta$  des constantes,  $\epsilon$  une variable aléatoire centrée réduite et  $N_{ass}$  le nombre d'assurés.

La première condition signifie que le nombre de sinistre moyen est proportionnel au nombre d'assurés. La deuxième est une hypothèse vérifiée par les principales lois de comptage utilisée en assurance (voir section précédente).

Lors de son inférence, un actuare estime les paramètres  $f$  et  $\eta$  par  $\hat{f}$  et  $\hat{\eta}$ , puis choisit la forme de la distribution de  $\epsilon$ . Il dispose donc dans sa modélisation d'une variable  $N' = \hat{f} \cdot N_{ass} + \hat{\eta} \cdot \sqrt{N_{ass}} \cdot \epsilon$

Or, en gardant l'incertitude sur les paramètres  $f$  et  $\eta$ , le coefficient de variation de la variable  $S' = \sum_{i=1}^{N'} X_i$  s'en retrouve modifié. Tous les estimateurs en jeu sont supposés ici sans biais comme c'est généralement le cas dans le monde de l'assurance.

Si l'espérance de  $S'$  est égale à celle de  $S$  car  $\mathbb{E}[N']\mathbb{E}[X] = \mathbb{E}[N]\mathbb{E}[X]$ , leurs variances n'ont plus la même expression (voir Annexe A.2.1 pour les détails) :

$$Var[S'] = \mathbb{E}[N]Var[X] + \mathbb{E}[X^2] \left[ Var[N] + N_{ass}^2 Var[\hat{f}] + N_{ass} Var[\hat{\eta}] \right]$$

Le coefficient de variation de  $S'$  vaut alors :

$$CoV^2[S'] = CoV^2[S] + \frac{1}{\hat{f}^2} Var[\hat{f}] + \frac{1}{N_{ass}} Var[\hat{\eta}]$$

Le coefficient de variation de la sinistralité conserve une composante inversement proportionnelle au nombre d'assurés ( $CoV^2[S]$ ), reflétant le bénéfice de la diversification sur le risque de contrat. Néanmoins, il admet en complément une composante indépendante du nombre d'assurés, dépendant de l'incertitude sur les paramètres du modèle.

Ainsi, le risque sur paramètres ne se diversifie pas avec le volume et le  $CoV$  est minorée par  $CoV^2[S'] \xrightarrow{N_{ass} \rightarrow \infty} \frac{1}{\hat{f}^2} Var[\hat{f}] = CoV^2[\hat{f}]$ .

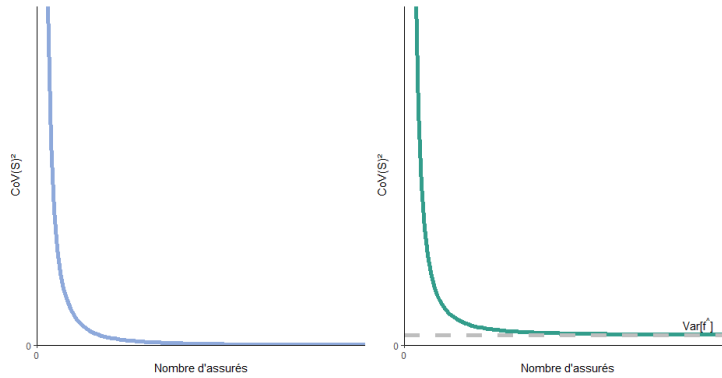


FIGURE 2.1 – CoV de la sinistralité en fonction du volume, la figure de gauche correspond à une modélisation sans incertitude sur paramètres tandis que celle de droite correspond à une modélisation incluant ce risque.

Plusieurs objections pourraient être formulées :

- l'incertitude sur paramètres décroît également avec le nombre d'assurés, asymptotiquement  $Var[\hat{f}]$  est nulle donc le risque tend encore vers 0 ;
- ces résultats sont asymptotiques et n'ont donc qu'une portée théorique, rien ne garantit qu'ils aient un quelconque impact dans le monde de l'assurance.

La première remarque est plus facile à traiter que la seconde. L'idée sous-jacente est qu'avec une grande population assurée, les variances des estimateurs devraient tendre vers 0, comme c'est souvent le cas dans un cadre fréquentiel où le théorème limite central s'applique. Premièrement, une grande population assurée l'année  $N + 1$  ne signifie pas nécessairement un grand historique disponible. Secondement, si le biais d'échantillonnage peut raisonnablement être considéré comme asymptotiquement nul, il ne faut pas oublier que l'incertitude sur paramètres comprend également le biais d'observation. Ce dernier est issu du fait que la période d'exposition de l'échantillon n'est pas, malgré la mise en as-if, nécessairement représentative de l'environnement de prédiction ; il n'a donc aucune raison de se diversifier.

Concernant la deuxième remarque, elle semble parfaitement recevable. L'un des objectifs de ce mémoire sera de quantifier la composante de risque liée à l'incertitude sur paramètres afin de conclure quant à la pertinence de son incorporation dans un modèle interne tel que celui de Groupama.

Pour ce faire, des solutions pour gérer ce risque seront proposées, la matérialité du risque de paramètres sera alors déduite à partir de comparaison.

La section suivante abordera la question de l'intégration de données externes à l'historique dans le calibrage de la sinistralité. Bien que le lien entre cette problématique et celle de l'incertitude sur paramètres puisse sembler flou à ce stade, la relation entre ces deux sujets sera explicitée. Une réponse commune à leur adresser sera par la suite proposée dans le corps de l'étude.

## 2.2 Intégration de données externes

Dans le monde de l'assurance, la disponibilité croissante de données provenant d'origines variées offre de nouvelles opportunités pour informer les prises de décision.

Une exploitation stratégique de ces diverses ressources de données conférerait un avantage compétitif majeur à une compagnie en lui ouvrant la voie à une meilleure compréhension, modélisation et gestion des risques dans un environnement en constante évolution.

Face à la multiplicité des provenances possibles d'informations, il convient de consacrer une section à leur présentation.

## **2.2.1 Sources de données**

Au-delà des historiques de sinistres traditionnels, l'usage d'une combinaison de données provenant de sources à la fois internes et externes à l'entreprise est aujourd'hui une nécessité. Parmi ces sources, les données dites "internes" représentent un intérêt tout particulier.

### **2.2.1.1 Données internes**

Le modèle interne d'une compagnie d'assurance non-vie peut grandement bénéficier de l'alimentation de cette famille de données qui englobe notamment les résultats d'études et d'expertises menées par différentes directions au sein de l'entreprise. Ces apports se réfèrent donc spécifiquement aux mêmes portefeuilles d'assurés et aux mêmes garanties. Ils sont donc pleinement pertinents et représentent une source particulièrement adaptée d'informations et de connaissances supplémentaires.

De plus, ces différentes informations se complètent en éclairant la compréhension de plusieurs typologies de risques concomitants. La direction de la réassurance peut fournir des évaluations détaillées des risques cumulatifs et des scénarios de pertes majeures, tandis que la direction de la tarification peut apporter des connaissances sur les profils de risque des assurés, les risques de masse et les comportements de souscription. En combinant ces informations avec les données déjà présentes par ailleurs dans le modèle interne, l'assureur peut améliorer l'évaluation des besoins de capitaux nécessaires pour faire face aux scénarios les plus adverses et, en retour, optimiser la tarification des primes et renforcer la gestion de sa réassurance.

L'alimentation du modèle interne par les travaux des autres directions présente également des avantages en termes de collaboration et d'échange d'expertise au sein de l'entreprise. Cette approche favorise une analyse actuarielle complémentaire et donc plus complète, en intégrant différentes perspectives et en exploitant les connaissances spécifiques de chaque direction. Ces échanges permettent aussi une meilleure communication interne à l'entreprise et favorise également l'échange d'expertise au sein du Groupe. Finalement, en tirant parti des connaissances et des expertises disponibles, l'assureur conforte son aptitude à anticiper et à gérer les risques de manière proactive, contribuant ainsi à une meilleure performance globale et à une gestion plus efficace des risques.

Les opportunités de compléter sa base de données ne se limitent pas à l'ajout d'informations internes à l'entreprise, mais s'étendent également à l'utilisation de données externes et plus particulièrement de données dites en "opendata".

### 2.2.1.2 Données externes

L'opendata, terme désignant les données accessibles au public et partagées de manière ouverte, connu une expansion significative au cours des deux dernières décennies. Elle présente des avantages substantiels pour les assureurs. Tout d'abord, son accès est généralement gratuit et sa collecte s'effectue de manière rapide et aisée, ce qui en fait une ressource économiquement attrayante pour les compagnies d'assurance.

Par ailleurs, l'opendata comprend une variété de sources, y compris des données gouvernementales et institutionnelles, qui sont généralement considérées à juste titre comme fiables. Les gouvernements collectent et mettent à disposition un large éventail de données dans des domaines tels que la démographie, l'économie, l'environnement et la santé, qui peuvent être pertinentes pour l'analyse des risques dans le contexte assurantiel. De plus, ces sources gouvernementales sont généralement considérées comme pérennes et offrent une base pour l'exploration de nouvelles perspectives. Par exemple, dans le cas d'une assurance récolte, le ministère de l'agriculture publie chaque année les chiffres de rendements des agriculteurs en donnant l'information de la culture, du département et de la superficie de l'exploitation.

Cependant, il est important de noter que l'utilisation de l'opendata nécessite une évaluation attentive de la qualité des données. Étant donné que les assureurs ont un contrôle limité sur la collecte et la maintenance de ces données, il est essentiel de s'assurer de la fiabilité de la source et de la validité des informations fournies. Les actuaires doivent également être conscients que la qualité et la pertinence des données peuvent varier d'une source à l'autre, ce qui peut influencer la précision des résultats obtenus.

De plus, bien que l'opendata puisse fournir des informations utiles, il convient de souligner que ces données ne sont pas spécifiques aux assurés d'une compagnie d'assurance particulière. Elles reflètent généralement des tendances et des statistiques à un niveau agrégé, ce qui peut limiter leur applicabilité directe aux risques spécifiques d'une entreprise. Les sociétés d'assurance doivent donc prendre en compte cette limitation et compléter les informations de l'opendata par des données internes plus spécifiques à leur portefeuille d'assurés.

En résumé, l'opendata offre une opportunité intéressante pour les compagnie d'assurance d'enrichir leurs analyses en intégrant des sources externes gratuites et rapidement accessibles. Cependant, la qualité des données et leur pertinence doivent être évaluées avec soin, compte tenu de la nature variable des sources et de la nécessité de les adapter aux besoins spécifiques de l'assureur. En utilisant judicieusement l'opendata en complément de leurs bases de données internes, les assureurs peuvent bénéficier d'une vision plus globale des risques, renforçant ainsi leurs capacités d'analyse et de prise de décision.

## 2.2.2 Méthodes d'incorporation de ces données

Si trouver et rendre exploitables ces diverses sources d'informations représente déjà un défi en soi, leur exploitation représente encore une tâche qui n'est pas



sans difficultés. L'utilisation de ces données nécessite une approche rigoureuse et cohérente pour pouvoir en extraire les informations utiles afin de prendre des décisions éclairées. Il est donc tout aussi crucial de développer des méthodes statistiques appropriées pour leur utilisation au sein d'un modèle interne que de trouver des données externes.

Dans le cadre des statistiques fréquentielles, quelques méthodes peuvent être appliquées pour être en mesure de se servir de ces nouvelles sources de données. Cela peut passer par l'usage des méthodes classiques de régression ou d'estimation de paramètres qui sont déjà largement mises en oeuvre dans l'industrie de l'assurance pour analyser les données historiques et évaluer les risques futurs. Cependant, elles peuvent rapidement présenter des limites lorsqu'il s'agit de combiner plusieurs sources d'informations de manière rigoureuse dans le but d'incorporer efficacement des connaissances externes.

C'est également dans ce domaine que le paradigme bayésien trouve son intérêt. En calibrant des *a priori* d'une manière appropriée, les actuaires peuvent intégrer les connaissances provenant de différentes sources, y compris les données internes et externes, pour améliorer les prédictions et leurs estimations de risques. L'approche bayésienne, à travers sa construction basée sur l'actualisation d'une connaissance, permet de prendre en compte les incertitudes associées à chaque source d'information et de les combiner de manière cohérente et efficace.

Ainsi, bien que l'exploitation de nouvelles sources de données puisse présenter un défi statistique, le paradigme bayésien offre une approche prometteuse pour combiner et exploiter efficacement ces informations. En facilitant les croisements entre plusieurs sources d'informations, l'approche bayésienne renforce la capacité des assureurs à évaluer les risques de manière plus précise et à prendre des décisions éclairées. De plus, par la construction et l'usage même de la loi *a posteriori*, l'incertitude sur les paramètres est déjà incluse au sein des prédictions du modèle. L'utilisation judicieuse de l'approche permettrait donc de répondre aux deux problématiques soulevées dans ce chapitre et d'améliorer de ce fait la gestion des risques, l'évaluation des besoins en capitaux et la prise de décision stratégique au sein de l'industrie de l'assurance.

Avant de pouvoir aller plus en avant et de définir formellement une méthodologie à suivre, il est essentiel de revenir aux fondements de la statistique et d'explorer les idées clés qui sous-tendent les paradigmes classique et bayésien. L'objectif de ces rappels n'est pas de fournir un cours de mathématiques, mais plutôt de présenter une vue d'ensemble des principes fondamentaux et d'introduire des outils qui serviront dans les analyses futures. Mettre l'accent sur les concepts clés servira à mieux appréhender les différences entre les approches classiques et bayésienne et permettra de comprendre leur pertinence dans le contexte actuariel. Finalement, des outils théoriques issus de la littérature seront présentés en fin de chapitre afin de préparer leur usage dans ce mémoire.

# Chapitre 3

## Outils théoriques : rappels de statistique et comparaison de paradigmes

De manière générale, la statistique est un ensemble d'outils visant à permettre l'étude d'un phénomène incertain à partir d'un nombre limité d'informations. Usuellement, ces informations sont sous la forme d'observations de ce qu'auraient pu produire ou ont produit les causes du phénomène étudié. L'objectif est alors, à partir des informations disponibles, de tirer, tout en les évaluant, des conclusions quant aux mécanismes générateurs.

Plus spécifiquement, dans le cadre de ce mémoire, la finalité de l'usage de méthodes statistiques sera de produire des prédictions quant à la valeur d'une quantité d'intérêt issue d'un ensemble de causes.

Il est important de noter que le phénomène étudié est considéré comme intrinsèquement aléatoire, une part de l'incertitude est « incompressible » : même avec un nombre infini d'observations et en connaissant exactement le mécanisme générateur  $P$ , il n'est pas possible de prédire exactement la quantité d'intérêt  $X$ . Seuls des modèles dit paramétriques seront étudiés, c'est-à-dire des modèles où l'ensemble des causes peuvent être résumées à un vecteur de paramètres noté  $\theta$ .

### 3.1 La statistique fréquentielle

#### 3.1.1 Démarche générale

Le cadre général de la statistique paramétrique fréquentielle est le suivant : l'actuaire dispose d'un jeu de données qui, quitte à être retraitées, sont supposées comme provenant d'un phénomène pouvant être modélisé par une distribution de probabilité  $P$ . Cette distribution dépend d'un paramètre  $\theta$  supposé inconnu mais fixe et appartenant à un ensemble  $\Theta$  d'états du monde possibles. L'objectif de l'actuaire va être, à partir des observations, d'inférer la « véritable » valeur  $\theta_0$  du paramètre  $\theta$ .

Habituellement, l'estimation est basée sur l'optimisation d'un critère comme celui de la vraisemblance, des moindres carrés ou encore, comme dans le cadre du modèle

interne de Groupama notamment, par une méthode des moments. Dans tous les cas, la valeur inférée de  $\theta$  dépend de l'échantillon et la fonction permettant de passer des observations à l'estimation du paramètre est appelée estimateur.

Le rôle de l'actuaire est alors, en premier lieu, de spécifier le modèle envisagé  $\{P_\theta, \theta \in \Theta\}$  puis, dans un second temps, de mener les calculs nécessaires (biais, vitesse de convergence) au choix de l'estimateur.

Une fois ce travail effectué, l'actuaire peut évaluer la robustesse de l'estimation de diverses manières, les méthodes les plus fréquemment utilisées recourent à la distribution de l'estimateur. En effet, ce dernier étant une fonction d'observations modélisées par des variables aléatoires, il est lui-même une variable aléatoire disposant d'une distribution qui lui est propre. Des calculs probabilistes sur cette loi conduisent alors à la construction d'intervalles de confiance prenant en compte l'incertitude sur l'estimation de  $\theta$ .

Cependant, la loi de l'estimateur est généralement trop complexe à manipuler. En pratique, cette difficulté est contournée à l'aide de théorèmes comme le théorème limite central ou la méthode delta donnant des approximations robustes des distributions d'intérêt, pour peu que l'échantillon comporte suffisamment de données.

### 3.1.2 Limites du paradigme fréquentiel

Malgré la richesse théorique et l'abondance d'outils du cadre fréquentiel, ce dernier souffre de plusieurs limites majeures.

La pertinence de la conception fréquentielle des probabilités, fondée sur la répétabilité des expériences, peut parfois être remise en question car elle fournit des informations sur les valeurs attendues d'un grand nombre de réalisations indépendantes plutôt que sur une réalisation spécifique. Cela soulève une problématique lors du calcul du risque de primes, où l'intérêt se porte sur le résultat à 1 an qui ne sera observé qu'une seule fois.

Dans le contexte fréquentiel, cette conception des probabilités peut également présenter des limites dans l'analyse des incertitudes de calibrage. En effet, puisque l'échantillon est fixé, l'estimation prend également une valeur fixe : conditionnellement aux données, l'estimateur est entièrement déterminé et ne contient plus d'aléa. Il en est alors de même de l'intervalle de confiance, ce qui conduit à considérer la probabilité d'appartenance d'un paramètre  $\theta_0$ , une constante, à un intervalle dont les bornes sont également constantes.

Dans ce cadre, l'interprétation d'un intervalle de niveau  $\alpha$  est qu'en reproduisant l'expérience un grand nombre de fois,  $\theta_0$  appartiendra à l'intervalle de confiance  $\alpha$  dans  $\alpha\%$  des cas. Cette interprétation est à nouveau intrinsèquement liée à la répétition d'expériences indépendantes, ce qui peut ne pas être adéquat dans des situations où l'observation n'aura lieu qu'une fois.

Par ailleurs, la majorité des méthodes et des résultats issus du paradigme fréquentiel reposent sur un cadre asymptotique qui ne correspond pas toujours à la réalité du monde de l'assurance. Cela se révèle particulièrement problématique pour les

prédictions S/P, où l'historique se compose généralement d'un nombre limité de points, souvent autour de la vingtaine. Cette faible profondeur d'historique remet en question la validité des bornes des intervalles de confiance, ainsi que les garanties théoriques associées à l'utilisation des estimateurs, telles que la consistance ou l'absence asymptotique de biais.

L'approche classique, bien que largement utilisée, est donc contrainte par des difficultés à la fois pratiques et conceptuelles qui devraient essentiellement restreindre son usage à un cadre d'asymptotisme. En réalité, cette formalisation de la statistique est une approximation du paradigme plus large de la statistique bayésienne dont la validité théorique ne se borne pas à des situations de répétabilité des expériences et permet donc ainsi une meilleure gestion de l'incertain.

## 3.2 Rappels de statistique bayésienne

### 3.2.1 Une conception différente de la probabilité

La spécificité principale du paradigme bayésien par rapport au cadre fréquentiel réside dans le sens donné aux probabilités. Dans la conception bayésienne, les probabilités sont considérées comme des degrés de confiance, elles ne sont donc plus vues comme des fréquences ou des proportions. Concrètement, cette différence conceptuelle majeure se traduit par le fait que le paramètre d'intérêt  $\theta$  n'est plus supposé fixe mais est maintenant considéré comme étant lui-même aléatoire.

Pour ce faire, il est donné à  $\Theta$  la structure d'un espace mesuré, dont l'élicitation de la mesure, nommée mesure ou loi a priori, constitue l'une des principales difficultés à la mise en place d'un cadre bayésien. Cependant, bien que ce choix représente une charge de travail supplémentaire pour l'actuaire, la probabilisation du paramètre d'intérêt constitue une manière plus flexible et expressive de traduire la présence, voire même l'absence, d'informations extérieures à l'historique d'observations.

#### 3.2.1.1 Principe d'actualisation

Ainsi, dans le paradigme bayésien, l'état des causes du phénomène incertain est lui-même représenté par une distribution de probabilité de sorte à pondérer l'ensemble des possibles par son degré de confiance.

L'inférence bayésienne consiste alors, non plus à estimer une valeur ponctuelle du paramètre, mais à mettre à jour sa distribution à partir de l'information apportée par les observations.

La notion de conditionnement est centrale au cadre bayésien. L'objectif n'est pas de déterminer la distribution paramètres, qui n'est qu'un simple outil pour décrire une connaissance partielle, mais la loi des paramètres conditionnellement aux données, qui est nommée loi a posteriori. Lorsque les quantités considérées ont des lois à densité, comme ce sera le cas dans ce mémoire, le principe d'actualisation<sup>1</sup> s'écrit :

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

Avec :

---

1. Aussi connu sous le nom de théorème de Bayes.

- $\pi$  la densité de la loi a priori sur  $\theta$  ;
- $f(x|\theta)$  la densité de loi  $X|\theta$ , également appelée "vraisemblance" ;
- $\pi(\theta|x)$  la loi de  $\theta$  a posteriori de l'observation de  $x$ .

Le travail de l'actuaire commence alors par fixer un espace de paramètres  $\Theta$  mesuré par une mesure a priori  $\pi$  et une fonction de vraisemblance  $f(\cdot|\theta)$  représentant la densité de l'échantillon conditionnellement aux paramètres. Cependant, sa tâche ne se résume pas simplement au choix de ce triplet.

En effet, bien que le théorème de Bayes permette d'obtenir immédiatement une distribution des paramètres actualisée par l'observation d'un échantillon, la loi a posteriori à elle seule ne permet généralement pas de conclure. La finalité d'une étude statistique, que ce soit dans le cadre d'une prédiction, d'une description d'un phénomène ou plus largement d'un test d'hypothèse, peut être définie, de manière formelle, comme une prise de décision. Certes, cette décision nécessite généralement d'avoir une vision éclairée du mécanisme générateur mais ce dernier n'est qu'un moyen de traduire l'information utile et disponible.

Afin d'obtenir un cadre rigoureux et épistémologiquement valide, il est alors nécessaire de se donner un moyen d'évaluer la qualité d'une décision qui réponde à quelques axiomes de rationalité<sup>2</sup>.

### 3.2.1.2 L'inférence bayésienne

Le critère d'évaluation d'une décision est appelé fonction de coût. De ce dernier découlera les diverses notions de risques qui jouent un rôle central dans le calcul bayésien.

Généralement, la quantité d'intérêt dépend des observations, c'est-à-dire des effets. L'examen des causes, les paramètres du modèle, n'est mené que dans la mesure où il peut fournir des informations sur le comportement des effets. Dans le cadre bayésien, le passage d'une distribution des causes à une décision concernant la quantité d'intérêt s'effectue par la minimisation d'un risque défini au moyen d'une fonction de coût.

Pour être en mesure de quantifier les conséquences d'une décision, la fonction de coût doit prendre en entrée le « véritable état » des causes, ie. le paramètre  $\theta$ . De ce fait, elle est définie comme une fonction qui associe un réel à un couple (décision ; état réel des causes). Par commodité d'usage, le réel est généralement pris comme étant positif, il doit traduire le coût de la prise d'une décision pour un certain état du monde. Mathématiquement, la fonction de coût est notée :

$$L : \Theta \times \mathcal{D} \longrightarrow \mathbb{R}^+$$

Dans le cadre de ce mémoire, l'espace  $\mathcal{D}$  des décisions possibles sera le même espace  $\mathcal{X}$  que celui de la variable d'intérêt, l'objectif étant d'en estimer ses quantiles. De cette notion de fonction de coût découle divers concepts de risques dont notamment celui de risque intégré défini par :

---

2. Ces axiomes, issus de la théorie de la décision, ne seront pas listés dans ce mémoire. Ils peuvent être consultés dans le cours de statistique de BOUSQUET (2023) [1].

$$R_B(\delta|\pi) = \int_{\Theta} \int_{\Omega} L(\theta, \delta(\mathbf{X})) f(\mathbf{X}|\theta) d\mathbf{X} \pi(\theta) d\theta$$

L'intégrale portant à la fois sur  $\theta$  et sur l'échantillon, ce risque associe à une règle de décision un réel et induit donc une relation d'ordre totale, permettant ainsi la définition de l'estimateur de Bayes en tant que règle de décision,  $\delta_B$ , minimisant ce risque.

Un résultat théorique fréquemment utilisé permet d'obtenir une autre vision de cet estimateur. Par une interversion d'intégrales et un peu d'effort théorique, il est possible de montrer qu'un estimateur est l'estimateur de Bayes si et seulement si il minimise le risque a priori pour tous les échantillons observables :

$$\delta_B \in \underset{\delta \in \mathbb{D}}{\operatorname{argmin}} R_B(\delta|\pi) \iff \forall \mathbf{X}, \delta_B \in \underset{\delta \in \mathbb{D}}{\operatorname{argmin}} R_P(d = \delta(\mathbf{X})|\pi, \mathbf{X})$$

Le risque a posteriori étant défini pour un échantillon et un a priori par :

$$R_P(d = \delta(\mathbf{X})|\pi, \mathbf{X}) = \int_{\Theta} L(\theta, \delta(\mathbf{X})) \pi(\theta|\mathbf{X}) d\theta$$

L'équivalence est un théorème central de la statistique bayésienne qui sera utilisé par après.

Plusieurs résultats concernant les différentes notions de risque viennent alors légitimer l'usage de l'estimateur de Bayes et c'est en partie sur ces résultats que se base la légitimité de l'approche développée dans ce mémoire. Ils peuvent se résumer au fait qu'aucun estimateur ne domine l'estimateur de Bayes, c'est-à-dire que tout estimateur ne fera jamais qu'aussi bien que l'estimateur de Bayes.

Malgré ses attraits théoriques et conceptuels, la statistique bayésienne reste rarement employée dans le monde de l'actuariat et est quasiment absente du modèle interne partiel de Groupama. Cette absence s'explique principalement par la technicité de sa mise en place qui nécessite, entre autres, la définition d'une fonction de coût et, point généralement bloquant, d'une mesure a priori.

Les prochaines sections seront donc consacrées au choix de ces deux fonctions, l'objectif sera d'apporter plus de détails sur les procédures existantes pouvant guider ces choix cruciaux et d'en exposer les idées sous-jacentes avec un focus particulier sur le cas de cette étude.

### 3.2.2 La définition d'une fonction de coût

L'estimateur de Bayes dépend intrinsèquement d'une fonction de coût à travers le concept de risque bayésien. Il représente l'estimateur le plus naturel et le plus rationnel du point de vue de la théorie de la décision pour le problème d'inférence. Ainsi, contrairement au cadre fréquentiel, il ne s'agit pas de choisir un type d'estimateur, mais de construire une fonction de coût en adéquation avec le problème considéré.

Dans le cas particulier où une certaine quantité est recherchée, la fonction doit être conçue de manière à ce que l'estimateur de Bayes associé représente précisément cette quantité d'intérêt.

Afin d'illustrer ce concept et de pouvoir mener à bien l'étude, le cas de l'estimation d'un quantile de la loi prédictive a posteriori va être donné. Pour cela deux approches ont été retenues, chacune se basant sur des raisonnements différents.

### 3.2.2.1 Quantile de la loi prédictive d'un point de vue bayésien

Cette méthode est la méthode la plus classique à envisager lors d'une inférence bayésienne. En effet, les lois  $X|\theta$  et de  $\theta$  fixées dans le modèle bayésien définissent directement une loi prédictive à posteriori  $X_{n+1}|(X_1, \dots, X_n)$  dont le quantile est bien une quantité fixée.

Le problème s'énonce alors de la manière suivante : l'objectif est d'obtenir le quantile  $q_{\alpha_i}$  d'ordre  $\alpha_i$  de cette loi prédictive  $X_{n+1}|(X_1, \dots, X_n)$ . C'est à dire le réel  $q_{\alpha_i}$  tel que  $\mathbb{P}(X_{n+1} < q_{\alpha_i} | X_1, \dots, X_n) = \alpha_i$ .

L'élicitation de la fonction de coût repose sur un résultat énoncé dans la section précédente, selon lequel un estimateur minimisant le risque a posteriori pour tous les échantillons est l'estimateur de Bayes.

Il s'agit donc de trouver une fonction  $L_{\alpha_i} : \Theta \times \mathcal{D} \rightarrow \mathbb{R}^+$  telle que :

$$\forall X, q_{\alpha_i} = \operatorname{argmin}_{y \in \mathcal{D}} \int_{\Theta} L_{\alpha_i}(\theta, y) \pi(\theta | X) d\theta$$

La résolution de ce problème passe par l'usage de la fonction  $L^1$  sous la condition  $\frac{c_1}{c_1+c_2} = \alpha_i$ . Cette fonction pénalise l'écart en norme 1 entre ses deux arguments, tout en pondérant l'erreur en fonction de sa direction.

$$L^1 : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}_+$$

$$(y, x) \mapsto \begin{cases} c_1(y - x) & \text{si } y > x \\ c_2(x - y) & \text{si } y \leq x \end{cases}$$

La fonction de coût à considérer est alors :

$$L_{\alpha_i} : \Theta \times \mathcal{D} \mapsto \mathbb{R}_+$$

$$(\theta, y) \mapsto \mathbb{E}_{x_{n+1}|\theta}[L^1(y, x_{n+1})]$$

En effet, un calcul d'intégrales donne :

$$\begin{aligned} \int_{\Theta} L_{\alpha_i}(\theta, y) \pi(\theta | X) d\theta &= \int_{\Theta} \mathbb{E}_{x_{n+1}|\theta}[L^1(y, x_{n+1})] \pi(\theta | X) d\theta \\ &= \int_{\Theta} \int_{\mathcal{X}} L^1(y, x_{n+1}) p(x_{n+1}|\theta) dx_{n+1} \pi(\theta | X) d\theta \\ &= \int_{\mathcal{X}} \int_{\Theta} L^1(y, x_{n+1}) p(x_{n+1}|\theta) \pi(\theta | X) d\theta dx_{n+1} \\ &= \int_{\mathcal{X}} L^1(y, x_{n+1}) \int_{\Theta} p(x_{n+1}|\theta) \pi(\theta | X) d\theta dx_{n+1} \\ &= \int_{\mathcal{X}} L^1(y, x_{n+1}) p(x_{n+1}|X) dx_{n+1} \\ \int_{\Theta} L_{\alpha_i}(\theta, y) \pi(\theta | X) d\theta &= \mathbb{E}_{x_{n+1}|X}[L^1(y, x_{n+1})] \end{aligned}$$

Les intégrales étant intervertibles du fait que toutes les grandeurs soient positives. Un raisonnement donné en annexe [A.1](#) montre que pour toute variable aléatoire  $Z$ , le minimiseur  $y^*$  de la fonction  $y \rightarrow \mathbb{E}_Z[L(y, X)]$  est le fractile  $\frac{c_1}{c_1+c_2}$  de la loi de  $Z$ . La solution du problème  $\operatorname{argmin}_{y \in \mathcal{D}} \int_{\Theta} L_{\alpha_i}(\theta, y) \pi(\theta | X) d\theta$  est donc bien  $q_{\alpha_i}$  ce qui était

le résultat souhaité.

Ainsi, en pondérant les erreurs de prédiction selon leur direction, l'optimum sera toujours le quantile souhaité de la loi prédictive a posteriori, quel que soit l'échantillon considéré. Une pondération plus élevée accordée à une erreur de surestimation dans la fonction  $L_1$  conduira à l'estimation d'un quantile plus bas, permettant ainsi de définir un estimateur pour chaque fractile de la loi prédictive a posteriori.

L'usage de cet estimateur de Bayes pour évaluer l'incertitude sur paramètre dans des modèles de prédiction actuarielle se trouve, sans les détails concernant la fonction de coût, dans les travaux de WACEK (2005)[7], PAPACHRISTOU (2009)[4] ainsi que ceux de VENTER et SAHASRABUDDH (2012)[5].

Le possible inconvénient de cette méthode est que la loi prédictive n'a pas de raison de correspondre à la vraisemblance, ce qui est à première vue tout à fait naturel. Effectivement, l'hypothèse de vraisemblance ne stipulait qu'une loi conditionnellement à un paramètre  $\theta$ . Il n'y a aucune raison de retomber sur la même loi en conditionnant par rapport à autre chose qu'un paramètre fixe. Néanmoins, la structure de la vraisemblance a pu être choisie car elle correspond supposément bien au comportement de la quantité d'intérêt. Dans ce cas, le modélisateur peut vouloir souhaiter d'essayer de garder une forme similaire dans les résultats de son inférence.

### 3.2.2.2 Quantiles prédictifs en raisonnant à partir de la vraisemblance

Dans cette section, la fonction de coût considérée est  $L(\theta, y) = d(q(\theta) - y)$  où  $d$  est une norme et  $q$  la fonction qui a un paramètre  $\theta$  associe la quantile de niveau  $\alpha$  de la loi  $X|\theta$ . Ainsi, pour un état du monde fixé, le risque est croissant en fonction de la distance entre le quantile estimé et le quantile réel.

L'estimateur de Bayes est alors le minimiseur de la quantité suivante :

$$\begin{aligned} \int_{\Theta} L(\theta, y)\pi(\theta|X)d\theta &= \int_{\Theta} d(q(\theta) - y)\pi(\theta|X)d\theta \\ \int_{\Theta} L(\theta, y)\pi(\theta|X)d\theta &= \mathbb{E}_{\pi(\theta|X)}[d(q(\theta) - y)] \end{aligned}$$

L'expression de l'estimateur va alors dépendre de la distance  $d$  choisie :

- pour la distance en valeur absolue, le résultat sera la médiane de la loi de  $q(\theta)$  avec  $\theta \sim \pi(\cdot|X)$  ;
- si la distance en norme absolue est pondérée en fonction de la direction, le résultat sera un quantile d'après la démonstration de la section précédente ;
- pour la distance quadratique, le résultat sera la moyenne de la loi de  $q(\theta)$  avec  $\theta \sim \pi(\cdot|X)$ .

L'intérêt de cette méthode réside dans le fait que le quantile s'obtienne à partir de la fonction  $q(\theta)$ , ce qui permet de mettre l'accent sur la forme de la vraisemblance choisie car cette dernière détermine entièrement la fonction  $q$ .

La distance quadratique sera choisie dans ce mémoire car elle permet de pénaliser plus durement les écarts élevés jugés aberrants. Ce choix d'estimateur de Bayes se retrouve également chez PAPACHRISTOU (2009)[4]. Cependant, l'auteur ne le justifie pas par un calcul bayésien, ni par d'autres arguments théoriques. Il réserve par



ailleurs l'utilisation de cet estimateur à une situation où le portefeuille est supposé être composé d'une multitude de polices hétérogènes. Au contraire, dans le cas où les polices sont considérées comme homogènes, il recommande l'estimateur mentionné dans la sous-section précédente.

La dérivation de l'estimateur par un calcul bayésien permet de légitimer son utilisation, quelles que soient les hypothèses sur l'hétérogénéité du portefeuille, et de le réinterpréter comme la solution optimale d'un problème de la théorie de la décision.

Finalement, la fonction de coût  $L_\alpha$  et la fonction de coût  $(\theta, y) \rightarrow (q(\theta) - y)^2$  ainsi que les deux estimateurs de Bayes associés seront utilisés dans ce mémoire. Pour plus de clarté dans la suite de l'énoncé, les noms des deux méthodes associées à chaque fonction de coût vont être fixés. La méthode basée sur la fonction de coût  $L_\alpha$  et ayant pour estimateur de Bayes le quantile de niveau  $\alpha$  de la loi  $X_{n+1} | X_1, \dots, X_n$  sera nommée "méthode bayésienne prédictive a posteriori". La méthode basée sur la fonction de coût  $(\theta, y) \rightarrow (q(\theta) - y)^2$  sera la "méthode prédictive par la vraisemblance".

Le cas de l'étude est ici un peu particulier. La construction de la fonction de coût est généralement imposée par les contraintes du problème. Habituellement, le modélisateur commence l'étude avec une fonction de coût à minimiser, puis recherche ensuite la décision optimale minimisant ce coût.

Dans ce mémoire, le processus a été inversé, l'objectif était dès le départ de trouver un quantile prédictif de la sinistralité. Les fonctions de coûts ont cette fois-ci été sélectionnées de sorte à ce que l'estimateur de Bayes puisse s'assimiler à la quantité d'intérêt.

Étant donné que deux fonctions ont finalement été retenues, il conviendra alors de comparer la performance des deux estimateur qui en découlent tout en reliant les résultats observés aux propriétés des fonctions dont ils sont issus.

Une fois la fonction de coût explicitée, il reste encore à choisir la mesure a priori, une étape cruciale de l'inférence qui s'avère régulièrement être la plus complexe. En effet, la mesure a priori doit refléter les éventuelles informations disponibles et qui ne sont pas toujours facile à exprimer sous la forme d'une distribution de probabilité du paramètre du modèle.

### 3.2.3 L'élicitation d'une mesure a priori

Tout d'abord, peu importe le paradigme employé, il est utile de rappeler qu'une connaissance a priori est toujours incluse dans l'étude et ce a minima lors du choix du modèle : la structure de l'ensemble  $\Theta$  exprime déjà une connaissance, même partielle, sur la structure des causes. La statistique bayésienne ne permet que de donner plus de possibilités pour retranscrire les éventuelles informations à la disposition de l'actuaire.

Ces informations peuvent être issues de plusieurs sources : des avis d'expert, des résultats d'essais autres que les données, des contraintes physiques, des connaissances liées à d'anciens produits, des résultats d'autres directions, des prévisions comptables, voire même ne pas exister.

Contrairement à une conception fréquemment rencontrée, la définition d'une loi

a priori n'a pas nécessairement pour finalité d'orienter l'inférence vers des états de causes jugés plus conformes à une connaissances préalables. En effet, l'intérêt de la statistique bayésienne réside non seulement dans la fusion de sources d'informations, mais également dans la définition d'un cadre formalisant solidement la prise de décision par un statisticien à la lumière d'observations, tout en proposant une formulation rigoureuse et fonctionnelle des probabilités qui ne repose pas sur la répétabilité de l'expérience.

Par ailleurs, il peut arriver que les connaissances a priori soient difficilement transposables à une mesure, que ce soit par difficulté d'encodage, de faible fiabilité des sources externes ou encore d'absence pure et simple. De plus, l'apport apparemment subjectif d'information d'une loi a priori peut-être mal considéré, surtout dans un domaine lourdement réglementé comme celui de l'assurance. Il reste néanmoins possible, même dans ces cas a priori défavorables, de jouir du cadre riche de la statistique bayésienne. Des procédures plus ou moins automatisées permettent de construire des priors dit «objectifs» ou «faiblement informatifs» par opposition aux mesures habituelles dites «informatives» ou encore «subjectives».

Ces mesures a priori objectives peuvent s'interpréter comme une régularisation de la vraisemblance des données. Bien que la littérature à leur sujet ait largement crû depuis la seconde moitié du XXème siècle, la recherche de lois a priori non informatives appropriées reste une question ouverte et active pour la recherche en statistiques bayésiennes. Toutefois, depuis le XVIIIème siècle, plusieurs approches rationnelles et maniables ont été développées.

### 3.2.3.1 Règles de construction de mesures a priori objectives

Par définition des a priori objectifs, pour une information donnée, chaque modélisateur devrait proposer le même a priori objectif. Pour cela, il faut des règles.

Historiquement, la première de ces règles remonte au concept, formulé au tournant du XIXème par Pierre Siméon de Laplace, de raison insuffisante des causes. Bien que ce concept soit aujourd'hui dépassé, il sert utilement d'introduction à la notion d'a priori objectif.

D'après Laplace, en cas d'absence de connaissance, il n'y a pas de raison de donner plus de poids à une hypothèse plutôt qu'à une autre, et donc chaque événement élémentaire devrait être également probable. Les événements élémentaires étant ici les valeurs du paramètre  $\theta$ , le prior objectif à considérer serait donc de la forme  $\pi(\theta) \propto 1$ .

Néanmoins ce prior ne respecte pas nécessairement la règle d'objectivité énoncée en début de section. En effet, le principe de raison insuffisante vise à donner le même poids à tous les états possibles, cependant la fonction de densité qui distribue la masse n'est pas invariante par reparamétrisation. Ainsi, une mesure pouvant apparaître "uniforme" sur l'ensemble  $\Theta$  pour une certaine paramétrisation ne le sera pas forcément pour une autre.

Une méthode de construction alternative proposée par José-Miguel Bernardo et James Orvis Berger à la fin des années 1970 et finalement formalisée à la fin des

années 2000 par Berger et al. (2009)[2] sera utilisée dans ce mémoire.

Intuitivement, un prior peut-être dit "non informatif" s'il ne joue qu'un rôle marginal dans la formation de la loi a posteriori; il doit donc pouvoir s'effacer devant l'information portée par les données. Une manière de formaliser cette définition est de se donner un critère de concordance entre la loi a priori et la loi a posteriori puis de définir une procédure qui, à une vraisemblance fixée, donne la mesure a priori maximisant la distance entre le prior et le posterior. L'objectif est donc de prendre comme prior de référence la mesure maximisant le gain d'information apporté par l'échantillon au posterior.

Selon un usage commun, la mesure de concordance utilisée est la divergence de Kullback-Leibler qui se définit, pour deux lois  $P$  et  $Q$  portant sur le même espace d'évènement  $\mathcal{X}$  et de densité  $p$  et  $q$  contre une mesure  $\mu$ , de la manière suivante :

$$D_{KL}(P||Q) = \int_{\mathcal{X}} p(x) \log\left(\frac{p(x)}{q(x)}\right) \mu(dx)$$

Cette notion issue de la théorie de l'information permet de mesurer la quantité d'information, au sens de l'entropie de Shannon, perdue lorsqu'on utilise la loi  $Q$  pour approximer  $P$ . Elle est donc à ce titre un indicateur cohérent dans la définition des mesure à priori.

L'idée derrière la construction des prior de référence est d'élucider un a priori  $\Pi$  qui soit une mauvaise approximation du posteriori  $\Pi_{|\mathcal{X}}$ . La quantité à maximiser serait donc  $D_{KL}(\Pi_{|\mathcal{X}}||\Pi)$  mais elle nécessite de disposer de la loi a posteriori et donc des données qui n'ont pas encore été observées.

La solution consiste alors à moyennner la quantité précédente selon  $X$  afin de ne plus dépendre de l'échantillon. En toute logique la loi retenue pour l'échantillon est la loi prédictive a priori  $m_{\pi}$ . Lorsque l'échantillon généré est de taille  $n$  le score à maximiser est donc :

$$J_n(\pi) = \int_{\mathcal{X}_n} D_{KL}(\Pi_{|x_n}||\Pi) m_{\pi}(x_n) dx_n$$

Finalement, dans le but de ne pas faire dépendre le résultat de la taille de l'échantillon considéré, le critère final de choix du prior de référence fait intervenir un passage à la limite.

Ainsi, pour une vraisemblance fixée, le prior de référence  $\pi^*$  est définie à partir de la quantité  $J_n$  de la manière suivante :

$$\pi^* = \operatorname{argmin}_{\pi} \lim_{n \rightarrow +\infty} J_n(\pi)$$

Une des propriétés intéressantes de cet a priori est qu'il vérifie la contrainte d'invariance par reparamétrisation et satisfait donc à cette exigence des loi objectives.

Toutefois, bien que conceptuellement et théoriquement fondé, sa détermination explicite peut aboutir à un problème ardu analytiquement. De plus, l'actuaire n'a pas toujours le temps de déployer des méthodes "sophistiquées" dans chacun de ses travaux et doit toujours arbitrer entre technicité de l'étude et réalisme de sa mise en place.

Fort heureusement pour la profession, les priors objectifs font l'objet d'une littérature riche et fournie qu'il est possible d'exploiter. Par exemple, une liste de prior objectifs, établis selon différentes méthodes, a été constituée en 1998 par Ruoyong Yang et Berger dans leur article "*A Catalog of Noninformative Priors*"[3]. Cet inventaire comprend les priors établis via la méthode mentionnées de Jeffreys, Bernardo-Berger ou Laplace ainsi que d'autres qui n'ont pas été abordées dans ce mémoire pour une trentaine de modèles ou de vraisemblances usuelle comme les modèles auto-regressif, linéaire, ou les lois de Pareto, Weibull, normale multivariée, Poisson, binomiale negative, etc.

En plus des expressions des mesures a priori, Yang et Berger donnent également la forme de la loi a posteriori et une vérification que le posterior est bien intégrable.

### 3.2.3.2 Construction d'a priori informatifs

Après avoir traité le cas des prior objectifs et sélectionné une manière de définir des a priori "de référence", il est maintenant temps de s'intéresser aux priors informatifs. Il est important de souligner d'entrée de jeu qu'il n'existe pas de procédure théorique permettant de construire des a priori informatifs comme c'est le cas pour les priors objectifs. En effet, chaque cas est unique. La construction d'un a priori doit prendre en compte les spécificités de la forme, du degré de certitude et du contenu des informations qui visent à être intégrées à l'inférence.

Cette sous-section ne se concentrera donc pas sur la présentation de méthodes standards ou automatisées pour définir des loi informatives, mais proposera plutôt des principes directeurs pour guider la construction d'a priori subjectifs. La démarche explicite utilisée dans ce mémoire ne sera décrite que dans la sous-section 5.2.1.

A l'inverse des mesures non informatives qui sont basées sur des considérations de neutralité subjective, les priors informatifs permettent d'incorporer dans la modélisation des informations spécifiques, qu'elles proviennent d'experts, d'études extérieures ou de connaissances du domaine, pour influencer les résultats de l'inférence.

Parmi les méthodes de conception de loi informative, plusieurs types d'approches se distinguent. Le premier groupe est celui des méthodes basées sur l'expertise, ces approches consistent à solliciter le conseil de professionnels de l'assurance ou d'experts sur un risques spécifiques disposant de connaissances ou d'une expériences approfondies du domaine considéré. Les jugements peuvent être recueillis par le biais d'entrevues, de sondages ou de groupes de discussions qui nécessitent donc de pouvoir mobiliser les différents experts.

Une seconde voie consiste à utiliser des données empiriques extérieures, provenant par exemple de bases de données ouvertes, de données gouvernementales ou même d'autres compagnies d'assurance. De même les conclusions d'études d'autres directions peuvent se relever particulièrement précieuses puisqu'elles ont été réalisées sur les mêmes portefeuilles et les modélisateurs sont généralement encore disponibles, du moins dans une certaine mesure.

Dans tous les cas, il est essentiel d'examiner les résultats obtenus à partir de loi informative afin de s'assurer de leur validité et de leur adéquation aux connaissances de l'assureur. Divers moyens de contrôle existent, la réalisation d'une étude

de sensibilité des résultats quant au paramétrage du prior en est un exemple. Par ailleurs, il est également recommandé de comparer les résultats avec les données observées en prenant en compte les tendances récentes. De même, il est possible de comparer les résultats avec des données externes telles que des conclusions d'études indépendantes.

Ces derniers rappels sur l'élicitation des lois a priori concluent ce chapitre qui visait à poser les bases théoriques indispensables aux études présentées dans ce mémoire. Sans elles, il n'est pas possible d'aborder de manière rigoureuse les problèmes spécifiques du monde de l'assurance, ni d'y apporter des solutions appropriées. Le cadre et le contexte général de l'étude étant maintenant posés, il est temps de passer au corps de l'étude à proprement parler.

Deuxième partie  
Corps de l'étude

# Chapitre 4

## Modélisation actuelle

Ce chapitre se penchera sur les aspects techniques relatifs aux méthodes utilisées par Groupama dans la modélisation de ses sous-groupes de risques homogènes. L'objectif principal est d'acquérir une compréhension détaillée de la situation actuelle afin d'identifier les limites des différentes méthodes employées. L'accent sera mis sur la sinistralité attritionnelle pour des motifs qui seront dévoilés ultérieurement.

### 4.1 Processus de modélisation

Conformément à la section 1.2.1, la modélisation de la sinistralité attritionnelle se fait à travers son ratio sinistres à primes. Il s'agit, dans cette section, d'exposer les méthodes retenues chez Groupama pour déterminer les lois prédictives de ces ratio  $S/P$  et le risque associé aux incertitudes d'estimation de leurs paramètres.

#### 4.1.1 Méthode de calibrage

Le processus de calibrage d'une loi de probabilité se déroule en plusieurs étapes. Tout d'abord, il est nécessaire de déterminer la famille de distributions appropriée pour modéliser le phénomène voulu. Cette étape implique une analyse des caractéristiques du problème assurantiel et de l'historique de données disponibles. Selon la nature des risques et les propriétés des variables étudiées, différentes familles de distributions peuvent être envisagées, telles que les distributions exponentielles, gamma, log-normales, ou encore des distributions à queue lourde telles que la distribution de Pareto. Toutes ces distributions sont définies de manière détaillée dans l'annexe B.1.

Une fois que la famille de distributions a été sélectionnée, la prochaine étape consiste à estimer les paramètres spécifiques de cette distribution, lesquels seront utilisés pour calibrer le modèle. Cette étape repose sur des techniques d'estimation statistique telles que la méthode des moments ou la méthode du maximum de vraisemblance. Ces méthodes permettent d'estimer les paramètres optimaux selon un certain critère en comparant les caractéristiques des données observées avec celles des modèles envisagés.

En pratique, les deux étapes sont réalisées indépendamment et le choix du type de loi précède celui de ses paramètres. Une analyse statistique préliminaire vise à

pré-sélectionner les familles de lois candidates. Ensuite, l'adéquation de ces candidats est évaluée à l'aide de différents tests ou critères classiques comme les tests de Kolmogorov, Anderson-Darling, Kuiper ou les critères d'AIC et d'erreur quadratique. De plus, des visualisations graphiques, des contrôles des moments et quantiles principaux sont effectués afin de valider plus sereinement le choix de la distribution retenue. Cependant, la tenue de ces tests nécessite d'avoir une loi de probabilité, c'est-à-dire une famille de lois et un jeu de paramètres spécifique. Le paramétrage utilisé lors de ces tests est obtenu par maximum de vraisemblance.

Pour rappel, les tests d'adéquation utilisés (mis à part l'AIC) ont pour statistique une mesure de dissimilarité évaluée en la distribution empirique et la distribution candidate. Sous l'hypothèse nulle que les données observées ont été générées selon la loi candidate, la statistique de test a alors une loi connue. Pour clarifier l'énoncé, la mesure de dissimilarité associée à un test  $T$  donné sera notée  $D_T$  et la loi de la statistique (donc de la valeur de la mesure de dissimilarité) aura pour fonction de répartition  $F_{\mathcal{L}}$ .

Avec ces notations, l'algorithme de sélection sur base de test d'adéquation est donné dans l'encadré "Algorithme 1".

La sélection de loi par le critère d'AIC ou des moindres carrés se déroule de la même

---

**Algorithme 1** Sélection d'une famille de distribution pour un test  $T$

---

Fournir  $n$  familles de distribution  $\mathcal{M}_1 = \{P_\theta^1, \theta \in \Theta_1\}, \dots, \mathcal{M}_n$ .

Fournir une distribution empirique  $F_0$ .

**Pour**  $i = 1, \dots, n$  **faire**

Fournir un paramétrage  $\hat{\theta}$  à la distribution  $P^i$

$\delta_i \leftarrow D_T(F_0, P_{\hat{\theta}}^i)$

$q_i \leftarrow F_{\mathcal{L}}^{-1}(\delta_i)$

**fin Pour**

Choisir le modèle  $\mathcal{M}_{i^*}$  tel que  $i^* = \operatorname{argmin}_{1 \leq i \leq n} q_i$ .

---

manière : le critère en question est calculé pour chaque famille de distribution sur la base de distribution calibrée par maximum de vraisemblance et la famille retenue est celle disposant de la valeur minimale du critère.

Une fois la famille de distribution choisie selon les critères statistiques d'adéquation précédemment cités vient le calibrage des paramètres finalement retenus dans le modèle. Ce dernier se fait non pas par maximum de vraisemblance mais par une méthode des moments avec pondération. Les paramètres de moyenne et d'écart type sont donnés par les formules suivantes :

$$\hat{m} = \sum_{i \in \mathcal{H}} \frac{P_i}{\sum_{j \in \mathcal{H}} P_j} \frac{S_i}{P_i} = \frac{\sum_{i \in \mathcal{H}} S_i}{\sum_{i \in \mathcal{H}} P_i}$$

$$\hat{\sigma}^2 = \frac{|\mathcal{H}|}{|\mathcal{H}| - 1} \sum_{i \in \mathcal{H}} \frac{P_i}{\sum_{j \in \mathcal{H}} P_j} \left( \frac{S_i}{P_i} - \hat{m} \right)^2$$

Ici,  $\mathcal{H}$  représente une indexation de l'historique,  $|\mathcal{H}|$  le nombre d'années dans l'historique,  $P_i$  et  $S_i$  les montants de primes collectées et de sinistres survenus liés à l'année de souscription  $i$ .



Ainsi, si la loi retenue est une log-normale, la distribution retenue pour le risque sera  $\mathcal{LN}(\hat{m}, \hat{\sigma})$ .

### 4.1.2 Évaluation de l'incertitude sur paramètres

L'incertitude sur paramètres n'est pas incluse dans les prédictions et le calibrage de Groupama. Le contrôle de sa matérialité est réalisé à travers une comparaison des charges simulées via le Modèle Interne Partiel par rapport à celles obtenues via un calibrage alternatif, devant quant à lui prendre en compte le risque sur paramètres.

Le calibrage alternatif est basé sur un résultat théorique de normalité asymptotique de l'estimateur du maximum de vraisemblance. Ce dernier stipule que sous certaines conditions et lorsque les données ont été générées selon la loi  $f(.|\theta_0)$ , l'estimateur de maximum de vraisemblance  $\hat{\theta}_{EMV}$  associé converge en loi de la manière suivante :  $\sqrt{n}(\theta_0 - \hat{\theta}_{EMV}) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}^{-1}(\theta_0))$ .

En pratique, le terme  $\mathcal{I}^{-1}(\theta_0)$  représentant l'inverse de l'information de Fisher étant inconnu, il est lui-même estimé par  $\mathcal{I}^{-1}(\hat{\theta}_{EMV})$ . Les simulations de sinistralité se font alors de la manière suivante :

- Un paramétrage  $\theta$  est tiré à partir d'une loi  $\mathcal{N}(\hat{\theta}_{EMV}, \mathcal{I}^{-1}(\hat{\theta}_{EMV}))$ .
- La charge de cette simulation est tirée selon la loi  $f(.|\theta)$ .

Par ce procédé, le paramétrage de la sinistralité varie de sorte à prendre en compte les incertitudes liées à son approximation par un estimateur. Cette idée d'attribuer aux paramètres la loi de leurs estimateurs s'inscrit dans la continuité des travaux de Veber (2009)[6] et Wacek (2005)[7].

## 4.2 Limites de la méthode

Maintenant que la méthode actuelle de calibrage des lois de sinistralité attritionnelle a été présentée, il est temps d'en exposer les limites et d'émettre des critiques qui permettront de progresser vers des approches plus avancées.

### 4.2.1 Choix des lois

#### 4.2.1.1 Choix de la famille de distribution

La sélection sur base de tests d'adéquation des familles de distribution est une approche fréquemment rencontrée dans le monde de l'actuariat et des statistiques appliquées en général. Cependant, ces tests ont été construits dans le but de comparer la loi observée empiriquement avec une distribution précise et non pas une famille de distribution complète.

Dans la méthodologie actuelle du modèle interne, les paramètres de la distribution candidate sont estimés à partir des données de sorte à ce que la loi proposée colle le mieux possible à l'échantillon. Cette procédure a pour effet de biaiser le test d'adéquation en faveur d'une meilleure concordance avec l'hypothèse nulle. À partir d'un exemple, il est possible de montrer empiriquement que la puissance du test, c'est-à-dire sa capacité à rejeter l'hypothèse nulle lorsqu'elle est fautive, se dégrade

significativement lorsque la loi candidate dépend des données.

Dans cette optique, 1000 échantillons contenant 15, 20 et 25 points ont été simulés selon une loi log-normale de paramètres fixés. Le test d'adéquation de Kolmogorov Smirnov a alors été mené sur chaque échantillon pour 5 lois candidates calibrées sur les données : les lois gamma, log-normale, normale, Weibull et Pareto ainsi que pour la loi log-normale qui a servi à générer les données et une loi de Weibull aux paramètres fixés dont la densité est "proche"<sup>1</sup> de la log-normale de simulation. Deux objectifs étaient visés ici, le premier était de comparer les taux d'acceptation (au seuil de 5%) entre les lois fixes et les lois calibrées sur l'échantillon, le deuxième était de comparer les p-valeurs obtenues sur les différentes familles de distributions.

Concernant le premier point, il apparaît que les p-valeurs des tests effectués sur des lois dépendant des données sont nettement plus élevées que celles obtenues pour une distribution aux paramètres fixés (voir figure 4.1). Ce résultat s'observe même lorsque la loi fixe est la véritable distribution des données et la loi variable appartient à la famille Weibull. Ainsi, même quand la distribution proposée n'est pas de la même famille que la loi génératrice, si ses paramètres sont autorisés à dépendre des données, alors le test d'adéquation indiquera généralement une bien meilleure concordance qu'avec la véritable loi de l'échantillon.

---

1. Cette distribution a été obtenue en calibrant une Weibull par maximum de vraisemblance sur un grand échantillon généré sous la log-normale qui a servi à générer les données. Ses paramètres étaient donc fixés.

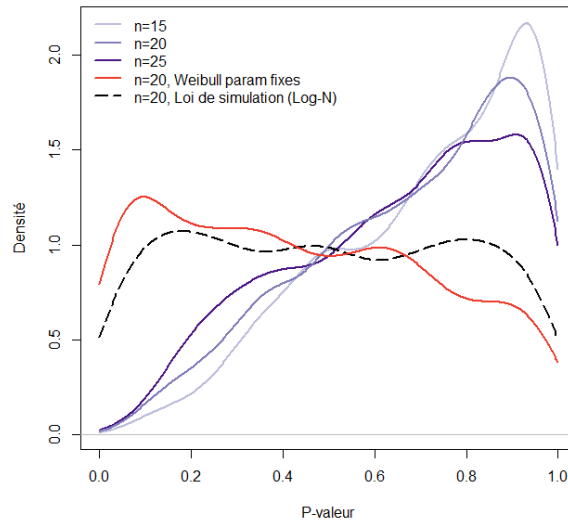


FIGURE 4.1 – Résultats des tests d’adéquations avec des lois aux paramètres fixes ou calibrés sur les données. Chaque courbe représente la densité (estimée par noyaux) des p-valeurs obtenues.

Les trois courbes violettes correspondent aux lois de Weibull calibrées sur l’échantillon. Sur ces dernières, le test ne remet jamais en cause l’adéquation avec les données et la densité est donc nulle à proximité de 0. Lorsque les paramètres sont fixés, la loi de Weibull est légèrement plus rejetée que la loi Log-Normale dont les p-valeurs sont distribuées uniformément, en adéquation avec la théorie.

En termes de taux d’acceptation, au seuil classique de 5%, les lois de la famille de Weibull calibrées sur les données ne sont tout bonnement jamais refusées (la p-valeur minimale sur 1000 échantillon de 25 observation est de 5.4%) tandis que l’hypothèse nulle est refusée dans 4,9% des cas pour la véritable loi de simulation. Techniquement, la puissance est donc empiriquement de 0% lorsque les paramètres sont autorisés à dépendre de l’échantillon.

Si la p-valeur perd son sens d’origine et ne permet plus à elle seule de fournir une appréciation pertinente de l’adéquation d’une loi, elle reste tout de même une mesure de concordance de par sa définition. En effet, cette dernière est définie à partir d’une mesure de divergence entre la loi empirique et la loi proposée, de telle sorte à ce qu’une valeur plus faible implique une divergence moins importante. Ainsi, bien que les p-valeurs n’ont plus de sens prises individuellement, elles peuvent encore en avoir lorsqu’elles sont comparées entre elles.

Sur base de comparaison (voir figure 4.2), le test reste capable de discriminer entre certaines familles de loi. Toujours avec des échantillons simulés selon une loi log-normale, le test de Kolmogorov-Smirnov arrive à se montrer méfiant envers la loi de Pareto tandis qu’il a plus de mal à choisir entre une log-normale, une gamma ou une Weibull.

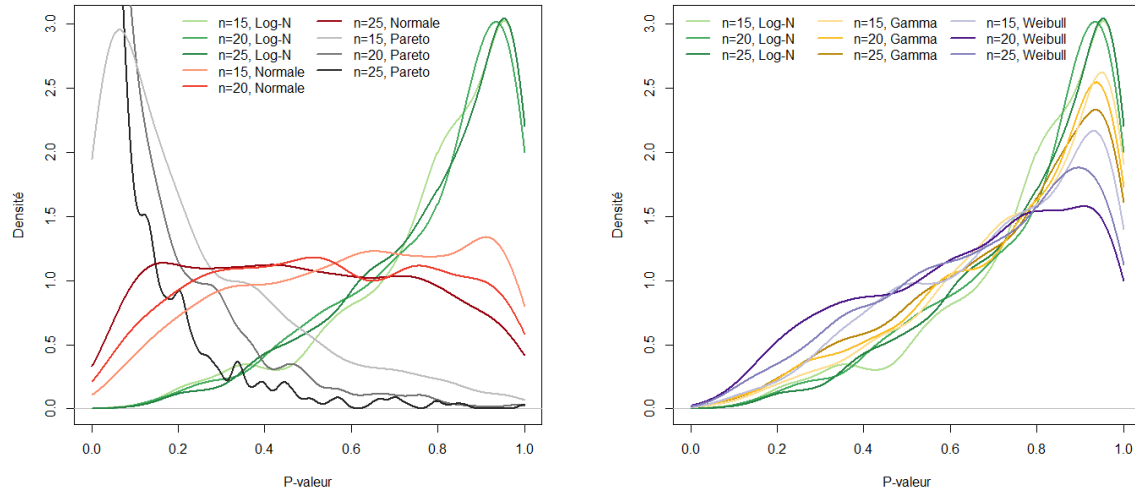


FIGURE 4.2 – Adéquation des différentes lois selon le test de Kolmogorov-Smirnov. Les résultats sont visualisés sous la forme de la densité des p-valeur obtenues. La loi de Pareto obtient des p-valeurs basses (mauvaise adéquation) dès 15 observations, tandis que celles de la loi normale sont distribuées uniformément. Dans les deux cas, la densité est de plus en plus piquée vers 0 au fur et à mesure que l'échantillon augmente. Les lois de Weibull, gamma et log-normale obtiennent des p-valeurs correspondant à une bonne adéquation, leurs courbes densité se chevauchent et la part des cas où le test associe la plus haute p-valeur à la bonne famille de loi n'est pas claire. Les densités correspondant aux lois gamma et de Weibull s'aplanissent également avec la taille de l'échantillon, le test de Kolmogorov-Smirnov a besoin de plus de points pour leur préférer la loi log-normale.

Dans l'exemple précédent, seul le test de Kolmogorov-Smirnov a été testé et ce sur des données simulées selon une loi log-normale. Afin d'évaluer plus globalement les performances des algorithmes de sélection de loi, ces derniers ont ensuite été implémentés pour l'ensemble des mesures d'adéquation retenues par Groupama (tests de Kolmogorov-Smirnov, Anderson-Darling, Kuiper ainsi qu'AIC et distance quadratique) pour des lois simulées selon une loi log-normale et de Pareto.

L'objectif était ici de tester plus largement les méthodes employées en recensant quelle loi était jugée la plus adéquate par quel test. Les résultats, toujours pour 1000 échantillons de 15, 20 et 25 observations, sont disponibles en figure 4.3 et 4.4. Il ressort de cette essai théorique qu'une autre limite de la méthode employée est sa faible capacité à discriminer entre deux lois sur un échantillon réduit. Pour une vingtaine de points, ce qui représente déjà un historique conséquent lors d'une modélisation agrégée par année, il apparaît qu'un unique test est incapable à lui seul de retrouver la loi d'origine. De plus, seulement 40% des échantillons générés mettent tous les tests d'accord. Par ailleurs, la bonne loi obtient la majorité des votes dans à pleine plus de 70% des cas ; cela laisse une chance assez large d'un point de vue actuariel de retenir une mauvaise famille de loi pour la modélisation du risque et donc d'aboutir à des conclusions erronées. Dans tous les cas, les performances sont hautement sensibles au nombre de points.

Cette faible capacité à discriminer entre les différentes lois sur des échantillons réduits est régulièrement observée sur les données de Groupama. Les équipes actuarielles doivent, dans leurs travaux, produire des calibrations pour des risques similaires et sur diverses entités du Groupe, notamment ses caisses régionales. Si les paramètres des lois varient logiquement avec l'entité, il arrive fréquemment que le choix de la loi diffère pour l'une ou l'autre filiale alors même que les engagements et la nature du risque sont similaires. Dans ces cas là, le modélisateur peut choisir de ne pas retenir la meilleure loi pour différentes raisons comme la stabilité par rapport à l'année précédente, la cohérence avec les autres entités ou la prudence.

		Log-Normale(-0,2; 0,5)		
		N=15	N=20	N=25
Kolmogorov-Smirnov	Gamma	12%	13%	15%
	<b>Log-Normale</b>	<b>59%</b>	<b>68%</b>	<b>73%</b>
	Normale	5%	3%	2%
	Weibull	8%	8%	7%
	Pareto	16%	8%	4%
AIC	Gamma	19%	21%	22%
	<b>Log-Normale</b>	<b>42%</b>	<b>52%</b>	<b>61%</b>
	Normale	0%	0%	0%
	Weibull	19%	15%	11%
	Pareto	20%	13%	6%
Anderson Darling	Gamma	0%	0%	0%
	<b>Log-Normale</b>	<b>58%</b>	<b>64%</b>	<b>69%</b>
	Normale	6%	4%	2%
	Weibull	10%	14%	13%
	Pareto	14%	7%	3%
Kuiper	Gamma	20%	21%	20%
	<b>Log-Normale</b>	<b>47%</b>	<b>54%</b>	<b>58%</b>
	Normale	10%	8%	6%
	Weibull	15%	13%	14%
	Pareto	8%	4%	2%
Least square	Gamma	14%	14%	13%
	<b>Log-Normale</b>	<b>64%</b>	<b>72%</b>	<b>76%</b>
	Normale	4%	3%	2%
	Weibull	5%	6%	7%
	Pareto	13%	6%	3%
5/5		<b>27%</b>	<b>34%</b>	<b>43%</b>
Majorité		<b>55%</b>	<b>66%</b>	<b>72%</b>

FIGURE 4.3 – Résultats des tests d'adéquations pour des données générées selon une loi log-normale

		Pareto(0,5; 2)		
		N=15	N=20	N=25
Kolmogorov Smirnov	Gamma	1%	0%	1%
	Log-Normale	15%	11%	5%
	Normale	0%	0%	0%
	Weibull	5%	2%	0%
	<b>Pareto</b>	<b>79%</b>	<b>87%</b>	<b>94%</b>
AIC	Gamma	0%	0%	0%
	Log-Normale	0%	1%	1%
	Normale	0%	0%	0%
	Weibull	0%	0%	0%
	<b>Pareto</b>	<b>100%</b>	<b>99%</b>	<b>99%</b>
Anderson Darling	Gamma	0%	0%	0%
	Log-Normale	34%	27%	14%
	Normale	0%	0%	0%
	Weibull	9%	4%	0%
	<b>Pareto</b>	<b>57%</b>	<b>69%</b>	<b>86%</b>
Kuiper	Gamma	3%	1%	1%
	Log-Normale	18%	6%	7%
	Normale	1%	0%	0%
	Weibull	0%	0%	0%
	<b>Pareto</b>	<b>78%</b>	<b>93%</b>	<b>92%</b>
Least square	Gamma	0%	0%	0%
	Log-Normale	13%	6%	6%
	Normale	0%	0%	0%
	Weibull	1%	0%	0%
	<b>Pareto</b>	<b>86%</b>	<b>94%</b>	<b>94%</b>
5/5		46%	65%	83%
Majorité		85%	94%	94%

FIGURE 4.4 – Résultats des tests d'adéquations pour des données générées selon une loi de Pareto

#### 4.2.1.2 Choix des paramètres

Plusieurs points sont à aborder dans cette section. Premièrement, l'estimateur retenu pour la variance donné en 4.1.1 n'est pas correctement débiaisé. En effet, le biais de l'estimateur de la variance est affecté par la pondération appliquée aux observations, la forme sans biais est la suivante :  $\frac{(\sum_i P_i)^2}{(\sum_i P_i)^2 - \sum_i P_i^2} \sum_{i \in \mathcal{H}} \frac{P_i}{\sum_{j \in \mathcal{H}} P_j} \left( \frac{S_i}{P_i} - \hat{m} \right)^2$

Néanmoins, l'écart entre les deux expressions tient à facteur  $\frac{|\mathcal{H}|-1}{|\mathcal{H}|} \frac{(\sum_i P_i)^2}{(\sum_i P_i)^2 - \sum_i P_i^2}$  qui, dans la pratique, est très proche de 1. Cette correction est donc négligeable.

Deuxièmement, les tests d'adéquation précédemment réalisés ont été effectués sur des lois calibrées par maximum de vraisemblance. Aucun test n'est mené sur les lois finalement retenues qui sont calibrées par méthode des moments pondérés. Ainsi, rien ne garantit que la distribution retenue reste la plus adéquate. En outre, la cohérence des principaux quantiles et dans une moindre mesure des moments n'est pas non plus contrôlée. Il est à noter que la méthode des moments admettant une pondération, l'estimateur retenu diffère toujours de celui du maximum de vraisemblance.

Troisièmement, la pondération par les primes historiques mérite quelques réflexions. L'idée sous-jacente repose sur le fait qu'un volume de souscription important permet de connaître avec moins d'incertitude le ratio sinistres à primes moyen. Le volume de prime est alors utilisé comme une mesure de taille du portefeuille. Étant donné que la modélisation par  $S/P$  concerne la sinistralité attritionnelle, le montant de primes fournit un indicateur relativement bon du nombre d'unités de risque assurées. Néanmoins, ce n'est non pas une prime pure qui est utilisée dans la pondération mais une prime commerciale qui varie en fonction des stratégies de tarification. De plus, conformément à la section 1.2.2 "Mise en as-if", les primes ne sont pas retraits de l'inflation. Un même montant de prime ne devrait donc pas représenter la même quantité d'unité de risque d'une année à l'autre. Bien que ce fait ait tendance à accorder plus de poids aux survenances récentes, la manière dont ce poids est réparti va dépendre d'un indice d'inflation qui n'a aucune raison d'être un indicateur de la pertinence d'une observation.

Finalement, le dernier point concerne la validité même de l'inférence. Dans ce cadre fréquentiel de la statistique, la pertinence des résultats dépend directement de la taille de l'échantillon. Or dans le cas des ratio  $S/P$ , ce dernier ne comporte qu'entre 15 et 25 points dans le meilleur des cas, il est donc extrêmement réduit. Afin d'évaluer la fiabilité des résultats, des échantillons contenant entre 15 à 25 points ont été générés selon une loi aux paramètres fixes et connus. Ensuite, les calibrages obtenus sur chaque échantillon ont été comparés, que ce soit en termes de valeurs des paramètres ou de quantiles des lois correspondantes. Les résultats sont visualisés en figure 4.5. Il est à noter que l'échelle de couleur pour les quantiles en fond est logarithmique : un écart d'une unité représente donc en réalité un facteur d'environ 2,7 tandis qu'un écart d'une bande de couleur (0,1 unité) représente un facteur d'environ 110%.

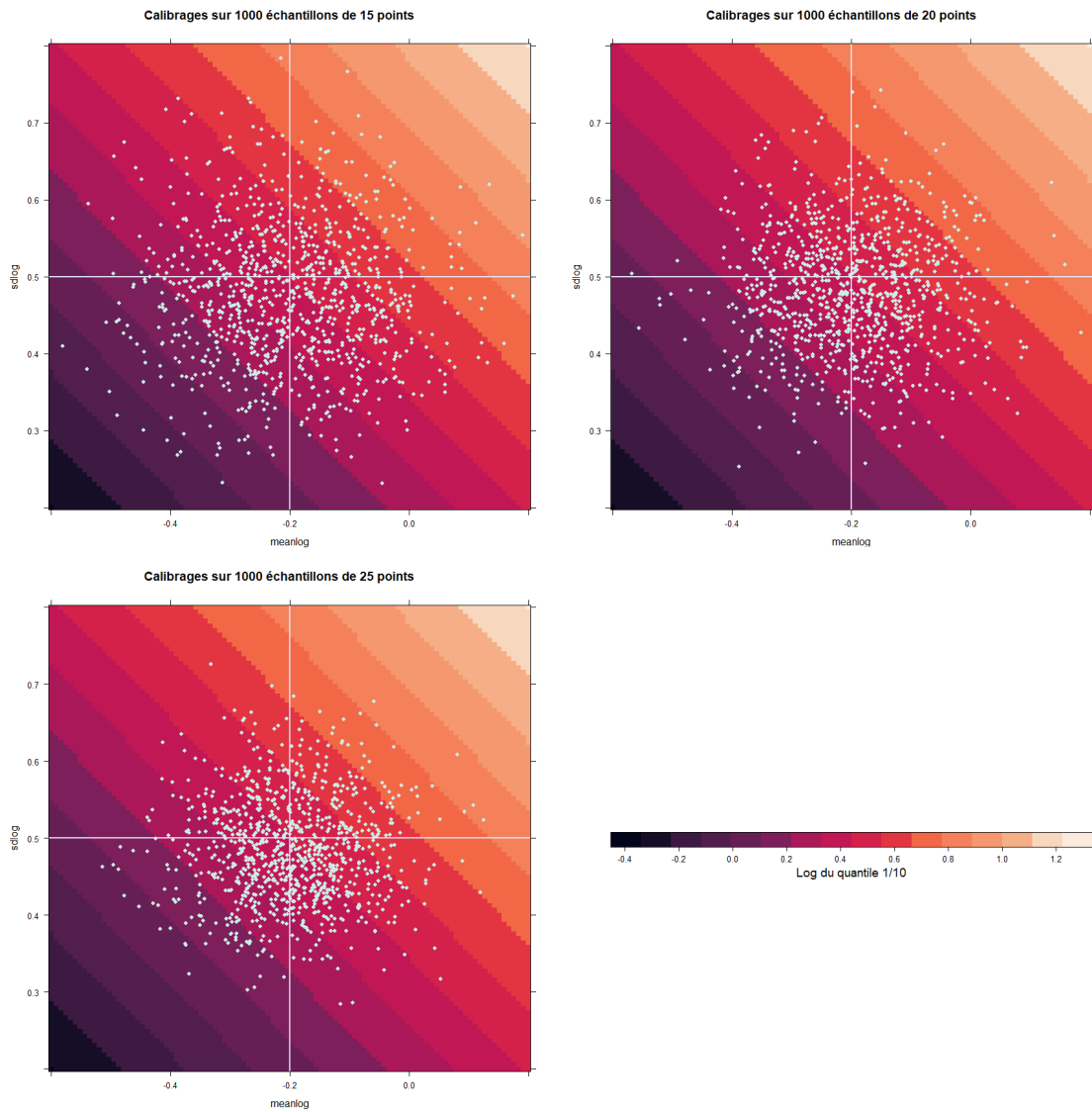


FIGURE 4.5 – Calibrage d’un paramétrage sur 1000 échantillons simulés de 15, 20 et 25 individus. Chaque point représente un couple de paramètres calibré sur un échantillon, les barres horizontales et verticales représentent le paramétrage utilisé pour simuler les données. La couleur de fond donne la valeur du quantile 90% (échelle logarithmique).

La distribution des points est centrée autour du ”vrai” jeu de paramètre et la dispersion des points diminue avec la taille de l’échantillon. Pour 25 points, les quantiles calibrés restent largement disséminés autour de leur valeur théorique.

Afin de mieux visualiser les résultats, la fonction de répartition de l’erreur relative a été tracée en figure 4.6. Cette erreur a été prise sur les valeurs (sans passage au logarithme) des quantiles obtenus avec 25 points et en valeur absolue : son expression mathématique est  $\frac{|\text{quantile}_{\text{calibré}} - \text{quantile}_{\text{théorique}}|}{\text{quantile}_{\text{théorique}}}$ .

Pour rappel, toute l’expérience a été menée sous l’hypothèse très généreuse que la famille de distribution utilisée représente parfaitement la distribution du risque. Avec cette connaissance, d’un point de vue fréquentiel et au seuil de tolérance de 20%, la précision de la méthode actuelle pour estimer un quantile 90% est d’environ 85%. Il y a environ une chance sur 2 pour que le quantile à 10 ans prédit par calibrage soit distant de plus de 10% de sa véritable valeur.



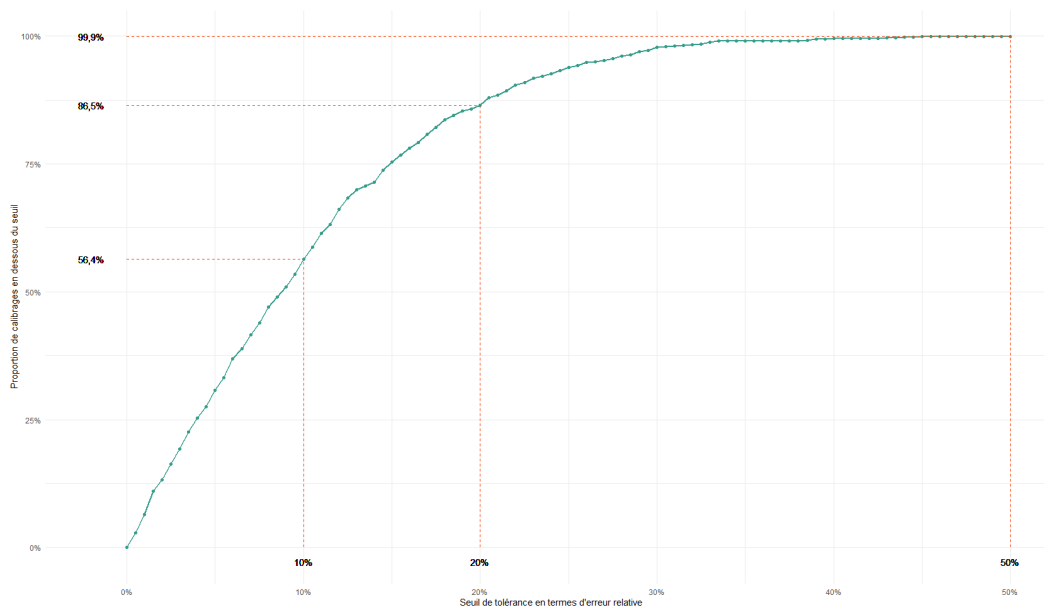


FIGURE 4.6 – Répartition empirique des écarts relatifs entre les quantiles 90% de la loi de simulation et ceux obtenus par calibrage, 1000 échantillons de 25 points ont été considéré. La répartition a été représentée sous forme d’une fonction de répartition : pour un seuil de tolérance donné, le graphique expose la proportion des calibrages ayant abouti à un quantile dans l’intervalle admissible.

Par exemple, 56,4% des calibrages ont abouti à un quantile situé à  $\pm 10\%$  de la véritable valeur, et donc près de 45% ont eu un écart plus important. De même, près d’un sixième des calibrages aboutissent à une erreur supérieure à  $\pm 20\%$ .

## 4.2.2 Incertitude sur paramètres

Quelques critiques peuvent également être soulevées quant à l’évaluation des incertitudes de calibrage. La plus évidente réside dans le fait que la loi théorique retenue pour l’estimateur est basée sur une estimation par maximum de vraisemblance, alors même que le calibrage est effectué par méthode des moments pondérés. Corollaire de ce premier point, l’estimateur du maximum de vraisemblance satisfaisant plusieurs conditions d’optimalité, sa variance asymptotique est plus faible que celle de l’estimateur des moments pondérés. Ainsi, l’approche retenue va avoir tendance à sous-estimer l’incertitude sur paramètre et n’est donc pas prudente.

Par ailleurs, la méthode repose une fois de plus sur une hypothèse d’asymptotisme non vérifiée dans le monde de l’assurance.

Néanmoins, il est nécessaire de contrebalancer ces points précédents. L’écart entre les deux méthodes d’estimation et l’impact de la pondération par les primes reste généralement assez faible. De plus, même pour une vingtaine de point, la loi asymptotique théorique (loi normale) fournit une approximation relativement robuste de la loi théorique de l’estimateur, c’est plutôt l’estimation de ses paramètres qui laisse à désirer. En outre, les questions de qualité de données (absence totale de retraitement) doivent probablement peser plus lourd que d’éventuels écarts à une loi asymptotique.

Le choix de centrer la méthode autour de l’estimateur du maximum de vraisemblance s’explique essentiellement par des questions de simplicité d’implémentation.

La littérature concernant cet estimateur est plus fournie, notamment du fait de ses propriétés théoriques intéressantes ; en conséquence, son implémentation sur des logiciels est plus répandue et il est donc plus simple de se reposer sur des travaux déjà existants. Par exemple, chez Groupama les calibrages de l'estimateur du maximum de vraisemblance et de sa variance asymptotique sont complètement automatisés dans un logiciel métier.

Toutefois, la méthode détaillée en section 4.1.2 de prise en compte de l'incertitude sur paramètre est également critiquable sur sa démarche même.

L'objectif de la méthode, et de l'évaluation de l'incertitude sur paramètre en général, est de construire une distribution des valeurs probables du paramètre. Cependant, dans les méthodes de la statistique classique, les causes sont supposées fixe et déterminées, elles sont représentées par un vecteur  $\theta_0$  simple. L'idée derrière la méthode de Groupama est de remarquer que l'estimateur dépendant des données, il est lui même une variable aléatoire et il dispose donc d'une loi.

Comme l'estimateur est supposé être un bon substitut aux véritables paramètres, sa distribution est prise comme étant celle des causes.

Un problème émerge assez naturellement : la loi de l'estimateur dépendant de celle des données, elle dépend donc elle-même du véritable paramètre.

En termes plus mathématiques, l'estimateur  $\hat{\theta}$  est une fonction des données, il est donc possible d'écrire  $\hat{\theta}$  sous la forme  $f(X_1, \dots, X_n)$  et il est clair que la loi de  $\hat{\theta}$  dépend de celle de  $X$ . Mis à part les deux premières remarques, la mauvaise fonction  $f$  (celle correspondant à l'estimateur du maximum de vraisemblance) et la "mauvaise" loi  $f(X_1, \dots, X_n)|\theta$  (loi asymptotique alors que l'échantillon est réduit) sont considérées chez Groupama, il reste un problème de taille. En effet, la loi de l'estimateur est elle même inconnue et dépend du paramètre dont la loi est recherchée.

Il y a alors deux raisons qui rendent l'usage de la méthode actuelle discutable. Premièrement la loi retenue pour les paramètres est celle de son estimateur, c'est à dire celle de  $\hat{\theta}|\theta$  ou encore  $f(X_1, \dots, X_n)|\theta$ . Deuxièmement, le paramètre  $\theta$  étant inconnu, il est remplacé par une estimation ponctuelle via un simple plug-in. Par conséquent, la méthode d'évaluation de l'incertitude sur paramètres repose elle-même sur une estimation ponctuelle qui ne prend donc aucunement en compte l'incertitude sur paramètres à laquelle elle est assujettie.

# Chapitre 5

## Méthodologies alternatives

### 5.1 Démarche générale

Bien qu'il soit tentant de prendre la loi  $\hat{\theta}|\theta$  comme distribution des paramètres, cela reviendrait à ne pas prendre en compte l'asymétrie de la relation de conditionnement. L'idée derrière cette méthode vient de ce que, dans un cadre fréquentiel, l'estimateur est utilisé comme substitut du paramètre. Si des résultats théoriques permettent d'assurer une certaine robustesse au fait de retenir une estimation ponctuelle comme état des causes, aucune théorie ne justifie de donner aux paramètres du modèle la distribution asymptotique de leurs estimateurs.

Certains résultats peuvent mener à l'obtention d'une loi pour une fonction à valeurs ponctuelles des paramètres d'intérêts, c'est le cas notamment de la delta méthode. Néanmoins, aucune ne permet de manière satisfaisante d'inclure l'incertitude de calibrage sur une distribution complète de loi des causes. Par exemple, la delta méthode permet d'avoir une distribution incluant l'incertitude d'estimation d'un quantile précis de la loi prédictive mais elle ne permet pas d'avoir une loi prédictive complète qui incorpore les incertitudes de calibrage.

De manière plus générale, le paradigme fréquentiel étant basé sur l'hypothèse de constance des paramètres, il ne peut que difficilement être à même d'incorporer une incertitude sur l'état réel du mécanisme générateur. D'un point de vue théorique, la méthode actuelle porte une part d'incohérence du fait qu'elle ignore le sens de la relation de conditionnement : la loi finalement retenue est celle de  $\hat{\theta}|\theta$  alors que l'objectif était d'obtenir une loi pour  $\theta$  conditionnellement aux données.

Une suite logique de la procédure serait d'inverser la relation de conditionnement, pour aboutir à la loi de  $\theta|\hat{\theta}$ . Or, cette inversion n'est autre qu'une formulation du théorème de Bayes, qui nécessite cependant une loi a priori du paramètre  $\theta$ .

Dans un modèle paramétrique où la distribution de la variable aléatoire  $X$  dépend des paramètres  $\theta$ , la prise en compte d'une incertitude sur paramètres conduit naturellement à considérer une loi de probabilité sur ces derniers. Lorsque l'objectif est aussi de prendre en compte un historique de données observées, qui sont distribuées selon la même loi conditionnelle  $X|\theta$ , l'actuaire se trouve face à la nécessité logique de considérer une loi conditionnelle sur les paramètres compte tenu de l'observation de l'échantillon.

En d'autres termes, une prise en compte de l'incertitude sur paramètres devant

aboutir à une distribution des causes et exploitant l'information contenue dans les observations ne peut se faire que selon un mécanisme de conditionnement, d'actualisation, c'est-à-dire dans un cadre bayésien.

La démarche suivie dans ce mémoire sera donc basée sur une vision bayésienne des statistiques. Elle se fera en trois étapes. La première sera d'éliciter une loi a priori portant sur les paramètres. La deuxième étape sera de mener une inférence bayésienne dans le but de d'obtenir une loi a posteriori puis d'en déduire une loi prédictive. Pour finir, la dernière étape consistera à interpréter les résultats obtenus en menant une analyse des propriétés de la distribution prédictive. En fonction du choix du type du prior retenu, l'objectif visé sera soit d'estimer l'incertitude sur paramètres, soit d'intégrer des données externes.

## 5.2 Choix des lois a priori

Deux familles de loi a priori seront exploitées dans ce mémoire. Sans surprise, ces deux familles seront celles des lois informatives et des lois objectives. Concernant les lois objectives, également dites non-informatives, la règle de construction retenue sera celle de Bernardo-Berger. Du côté des a priori informatifs, deux tâches principales seront à accomplir. La première sera de réunir des sources de données susceptibles d'apporter des informations complémentaires sur les risques modélisés ; la seconde sera d'encoder ces informations sous la forme d'une loi a priori des paramètres du modèle.

La transformation d'une information concernant la variable en une distribution de probabilité sur ses paramètres est une aire de la statistique bayésienne qui reste encore lacunaire. Contrairement à l'élicitation de lois objectives, la théorie n'a pas encore fait émerger des méthodes robustes et maniables permettant de coder les informations extérieures.

Dans ce mémoire, l'intérêt sera porté sur un calibrage avec une vraisemblance log-normale du fait de la qualité de son adéquation aux données. Des propositions d'encodage de données externes adaptées au monde de l'assurance seront proposées dans les sections suivantes puis implémentées par la suite.

### 5.2.1 Encodage de jugements d'expert en a priori informatifs

L'objectif principal de cette section est de proposer une méthodologie permettant de passer d'une information sous la forme de jugement d'expert, terme ayant ici pour sens "information extérieure aux données", à une distribution des paramètres. Les jugements utilisés dans cette partie seront de la forme  $f(S/P) = u$  où  $f$  sera une fonction quelconque, par exemple la fonction "quantile à 90%". L'idée générale, derrière le processus d'encodage et commune à tous ces types d'informations, va être d'essayer de charger les régions de  $\Theta$  dont les paramètres permettent de vérifier la contrainte spécifiée par l'expert.

La construction d'une telle loi se fera alors en plusieurs étapes. La première sera

de trouver les régions de  $\Theta$  les plus intéressantes, en explicitant les lignes de niveaux de la fonction  $f$  par exemple. Il faudra ensuite, dans un second temps, construire une loi chargeant préférentiellement les zones de l'espace les plus compatibles avec le jugement d'expert.

L'avis d'expert contenant lui même une marge d'incertitude, il est important de souligner que la loi retenue doit toutefois garder une certaine part de flexibilité.

### 5.2.1.1 Typologie des jugements

Avant d'aller plus loin et de traiter des cas concrets où la fonction  $f$  est spécifiée, il reste nécessaire d'apporter plus de précisions concernant la nature des différents jugements considérés.

Plusieurs types de jugement doivent être distingués en fonction de leur caractère. Certains sont restrictifs et pris comme sûrs, ils correspondent à des contraintes logiques ou physiques. Par exemple, pour un S/P d'un portefeuille historiquement à l'équilibre, il peut être considéré que le niveau médian se situe en dessous de 4000%. Une telle contrainte concerne la possibilité, elle est binaire : soit un paramétrage est possible, soit il ne l'est pas. Une traduction en terme de densité a priori serait donc de donner une valeur de 0 aux paramétrages "inadmissibles" et une valeur constante sur le reste, le jugement statuant uniquement sur la possibilité et non pas sur la probabilité.

Au contraire, d'autres jugements doivent être considérés comme incertains. Pour une raison ou pour une autre, un expert peut juger que le S/P moyen devrait être aux alentours de 95%. Ce jugement n'est pas certain, et l'actuaire aurait tort d'exclure tous les paramétrages n'aboutissant pas à des lois dont la moyenne serait 0,95. Une telle information, dite de probabilité ou permissive, renseigne sur les probabilité relative des calibrages les uns par rapport aux autres, en apportant l'information que les jeux de paramètres associés à des moyennes de 95% sont plus vraisemblables que les autres.

Les sous-sections suivantes traiterons des cas de jugements permissifs. La première présentera un exemple simple de fonction  $f$ , mais qui revêt néanmoins un intérêt potentiel dans le monde de l'assurance. La méthodologie qui y sera proposée ne sera pas la plus aboutie de ce mémoire, toutefois elle servira d'illustration et d'introduction utile aux raisonnements sous-tendant la seconde méthode.

### 5.2.1.2 Quantiles donnés par l'expert

Dans cette partie, la fonction  $f$  est la fonction quantile : la contrainte  $f(S/P) = u$  se lit donc "le quantile de niveau  $\alpha$  du  $S/P$  vaut  $u$ ". Sous une hypothèse de log-normalité, la détermination d'une courbe de niveau de la fonction quantile est une tâche relativement aisée, la solution étant une simple droite (voir figure 5.1).

En effet, pour un quantile  $\alpha_i$  de valeur  $q_{\alpha_i}$ , un rapide calcul donné en annexe A.3 montre que la courbe des paramétrages admissibles a pour équation  $\mu = \log(q_\alpha) - \sigma\Phi^{-1}(\alpha)$ , où  $\Phi$  est la fonction de répartition d'une loi normale centrée réduite.

Ainsi, pour une seule condition sur un seul quantile, toute une bande de paramétrage est valide. De manière générale, une unique condition donnera le plus

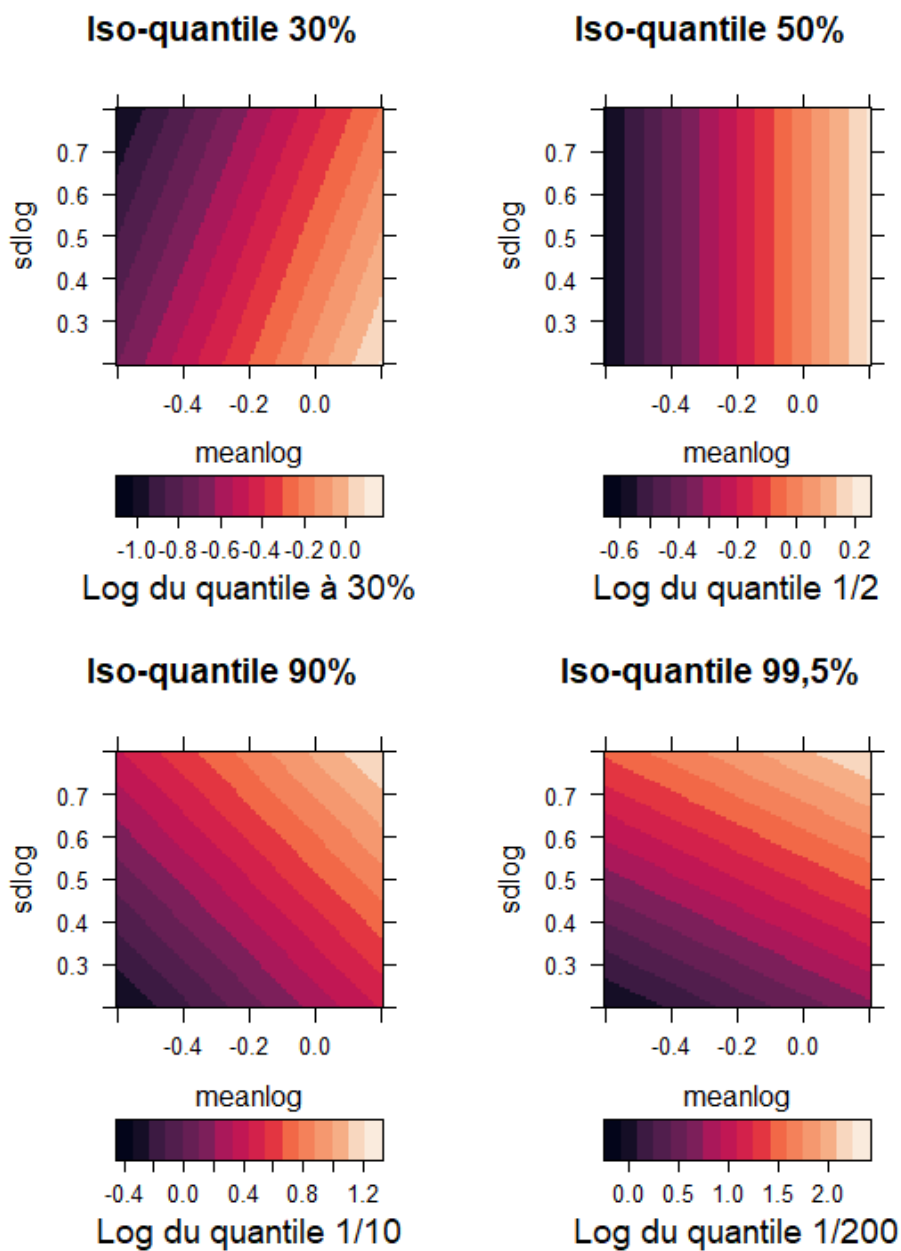


FIGURE 5.1 – Courbes iso-quantiles pour loi log-normale. La valeur du quantile pour un jeu de paramètre (meanlog ; sdlog) fixé est donnée en couleur : les lignes de niveaux sont des droites parallèles dont la pente dépend du niveau du quantile.

souvent un ensemble infini de jeux de paramètres possibles. Cela peut s'expliquer par le fait que l'ensemble  $\Theta$  possède 2 degrés de liberté et qu'un jugement d'expert n'en fixe usuellement qu'un seul. Si deux quantiles sont spécifiés, un unique jeu de paramètres satisfait aux contraintes. Lorsque plus de 3 quantiles sont fixés, rien ne garantit qu'un jeu de paramètres convienne. Cela n'est pas gênant en pratique, d'une part il est rare d'avoir jusqu'à trois avis fiables sur une même loi et d'autre part il existe des méthodes permettant d'aboutir à un consensus entre les experts. Dans le pire des cas, l'actuaire peut toujours choisir de garder les deux les plus cohérents.

L'enjeu est donc de trouver une loi qui permette de charger les bonnes bandes, tout en gardant suffisamment de flexibilité pour laisser les données s'exprimer. Lorsque l'information est pauvre et qu'un seul niveau de quantile est spécifié, une méthodologie plus générale proposée dans la section suivante peut être employée.

Reste à traiter un cas plus intéressant, celui où deux quantiles sont disponibles. Généralement, les quantiles correspondent à un niveau plutôt moyen et à un niveau plutôt "extrême". Un défi technique est alors de réussir à réunir les deux jugements dans une seule distribution, il faut savoir les combiner tout en les pondérant par un degré de confiance respectif.

L'objectif est alors de réussir à éliciter un prior chargeant préférentiellement les paramètres se situant sur les deux bandes qui, n'étant pas parallèles, forment une croix n'ayant qu'un seul point d'intersection. Assez logiquement, le centre de la croix correspondant à l'unique point vérifiant les deux contraintes, il doit se voir attribuer la densité la plus grande. De même, la masse doit décroître le long des branches de la croix car si les paramétrages vérifient bien l'une des deux contraintes, ils s'éloignent mécaniquement de l'autre (voir la figure 5.2 pour une illustration).

La méthode proposée ici se base sur le calibrage d'ellipses ayant pour axes principaux les iso-quantiles conseillées par l'expert, puis d'utiliser ces mêmes ellipses pour construire les lignes de niveaux de la densité du prior.

Le choix de lignes de niveaux elliptiques est à rapprocher des distributions normales multivariées qui représentent une référence en termes de modélisation d'erreurs de mesure. Toutefois, ce choix est débatable à plusieurs niveaux et reste arbitraire.

La méthodologie se décline alors en plusieurs étapes :

- trouver le centre  $M = (\mu ; \sigma)$  des ellipses ;
- construire une famille d'ellipses par jugement ;
- combiner les deux familles pour construire une mesure a priori.

La première étape revient à résoudre un système linéaire dont, en conservant les

même notations, les deux équations sont : 
$$\begin{cases} \mu = \log(q_{\alpha_1}) - \sigma\Phi^{-1}(\alpha_1) \\ \mu = \log(q_{\alpha_2}) - \sigma\Phi^{-1}(\alpha_2) \end{cases}$$

La solution est : 
$$\begin{cases} \mu^* = \frac{\log(q_{\alpha_1})\Phi^{-1}(\alpha_2) - \log(q_{\alpha_2})\Phi^{-1}(\alpha_1)}{\Phi^{-1}(\alpha_2) - \Phi^{-1}(\alpha_1)} \\ \sigma^* = \frac{\log(q_{\alpha_1}) - \log(q_{\alpha_2})}{\Phi^{-1}(\alpha_2) - \Phi^{-1}(\alpha_1)} \end{cases}$$

La construction des ellipses se base sur l'observation suivante : pour une loi normale multivariée  $\mathcal{N}(M, \Sigma)$ , les courbes de niveaux sont des ellipses centrées en  $M$  dont les axes principaux sont les vecteurs propres de la matrice  $M$ , le grand axe étant associé à la plus grande valeur propre. La démarche consiste alors à construire une matrice  $\Sigma$ , symétrique réelle et définie positive, dont le vecteur propre principal

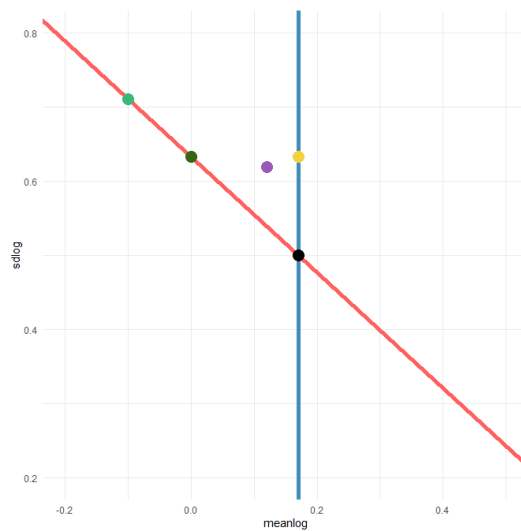


FIGURE 5.2 – Espace  $\Theta$  et positionnement de différents points. La droite bleue représente l'ensemble des paramétrages compatibles avec le premier jugement et la droite rouge ceux compatibles avec le deuxième. Le point noir étant le seul à se conformer aux deux jugements, il devrait se voir attribuer la densité la plus grande. De même, le point vert foncé devrait peser plus lourd que le point vert clair car s'ils sont à égalité vis à vis de la courbe rouge, le vert foncé est plus proche de la courbe bleue.

Reste à déterminer comment attribuer la masse entre des calibrages qui respectent un jugement mais l'autre (point jaune contre point vert) ou un point "proche" des deux par rapport à un qui se conforme à l'un mais est éloigné de l'autre (point violet contre point jaune ou vert).



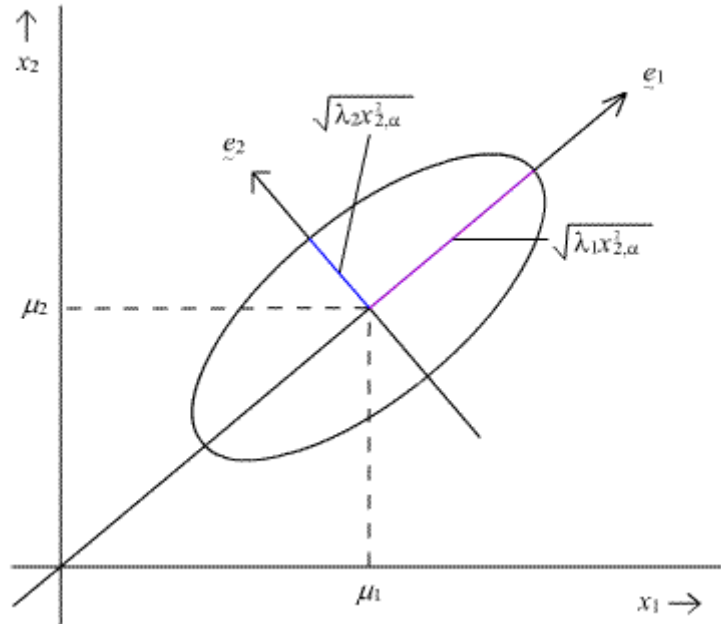


FIGURE 5.3 – Longueur des demi-axes en fonction des valeurs propres[8].

est le vecteur directeur de la droite iso-quantile.

Toute matrice symétrique réelle ayant des sous-espaces propres orthogonaux, le deuxième vecteur propre est alors nécessairement un orthogonal du premier vecteur propre. De plus, la matrice  $\Sigma$  s'écrit  $\Sigma = O \text{diag}(\lambda_1, \lambda_2) O^T$  où  $\lambda_i$  correspond à la valeur propre associée au  $i$ ème espace propre dont un vecteur directeur normé compose la  $i$ ème colonne de la matrice orthonormée  $O$ .<sup>1</sup>

Concrètement, la densité retenue, admettant des ellipses comme lignes de niveaux, est alors celle d'une loi normale multivariée de paramètres  $M = (\mu^*; \sigma^*)$  et de matrice de variance-covariance  $\Sigma = \begin{pmatrix} e_1 & e_1^\perp \\ e_2 & e_2^\perp \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} e_1 & e_2 \\ e_1^\perp & e_2^\perp \end{pmatrix}$ .

Une expression du vecteur directeur normalisé est  $(e_1; e_2) = \frac{1}{\sqrt{\Phi^{-1}(\alpha_i)^2 + 1}} (-\Phi^{-1}(\alpha_i); 1)$ , le vecteur  $(e_1^\perp; e_2^\perp)$  étant un vecteur normé et orthogonal à  $(e_1; e_2)$ . Chaque loi ainsi obtenue correspond à une distribution de la masse le long d'une région qui coïncide avec l'un des deux jugements d'experts.

Le choix des valeurs propres permet de réguler la taille et la forme de l'ellipse (voir figure 5.3). Le rapport  $\sqrt{\lambda_1/\lambda_2}$  permet de déterminer l'épaisseur de l'ellipse, c'est-à-dire à quel degré la masse doit être concentrée le long de l'axe, tandis que la valeur de  $\sqrt{\lambda_1}$  contrôle la dimension de l'ellipse, c'est-à-dire sur quelle étendue la masse est distribuée le long de l'axe. Ainsi, un haut degré de confiance sera associé à un ratio  $\sqrt{\lambda_1/\lambda_2}$  élevé : la masse ne s'écartera que peu de l'axe fourni par l'expert. Si le modélisateur souhaite accorder plus de crédit à un des deux jugements, il pourra lui attribuer une valeur propre  $\sqrt{\lambda_1}$  élevée de sorte à balayer plus largement la ligne de niveau fournie par l'expert. Une analyse plus approfondie quant à l'influence de ces deux paramètres sur la modélisation sera menée dans le chapitre de mise en pratique.

1. Ce résultat classique d'algèbre linéaire porte le nom de théorème spectral.

Cette procédure aboutit à l'élaboration de deux a priori,  $\pi_1$  et  $\pi_2$  de densités respectives  $\pi_i(\mu, \sigma) \propto \frac{\exp(-\frac{1}{2}((\mu; \sigma) - M)^T \Sigma_i ((\mu; \sigma) - M))}{\sqrt{(2\pi)^k |\Sigma_i|}}$ .

Il ne reste alors plus que la dernière étape, celle consistant à réunir ces deux lois dans un seul prior. La manière de les combiner n'est pas unique, elle dépend de la forme de densité que le modélisateur souhaite obtenir. Dans tous les cas, l'idée proposée ici est de fixer une fonction  $\phi$ , puis de garder comme prior  $\pi(\mu; \sigma) = \phi(\pi_1(\mu; \sigma); \pi_2(\mu; \sigma))$ .

La méthode d'élicitation présentée peut se résumer sous la forme d'un algorithme prenant en entrée les deux jugements d'experts, des coefficients sous forme de valeurs propres dont le sens a été donné précédemment et une fonction  $\phi$ .

---

**Algorithme 2** Elicitation d'un prior informatif à partir de jugements d'experts sur des quantiles

---

Fournir deux jugements  $(\alpha_1, q_{\alpha_1})$  et  $(\alpha_2, q_{\alpha_2})$ .

$$\mu^* \leftarrow \frac{\log(q_{\alpha_1})\Phi^{-1}(\alpha_2) - \log(q_{\alpha_2})\Phi^{-1}(\alpha_1)}{\Phi^{-1}(\alpha_2) - \Phi^{-1}(\alpha_1)}$$

$$\sigma^* \leftarrow \frac{\log(q_{\alpha_1}) - \log(q_{\alpha_2})}{\Phi^{-1}(\alpha_2) - \Phi^{-1}(\alpha_1)}$$

**Pour**  $i \in \{1, 2\}$  **faire**

Fournir  $\lambda_{1,i}, \lambda_{2,i}$

$$(e_{1,i}, e_{2,i}) \leftarrow (-\Phi^{-1}(\alpha_i); 1) / \sqrt{\Phi^{-1}(\alpha_i)^2 + 1}$$

$$(e_{1,i}^\perp, e_{2,i}^\perp) \leftarrow (1; -\Phi^{-1}(\alpha_i)) / \sqrt{\Phi^{-1}(\alpha_i)^2 + 1}$$

$$\Sigma_i \leftarrow \begin{pmatrix} e_{1,i} & e_{1,i}^\perp \\ e_{2,i} & e_{2,i}^\perp \end{pmatrix} \begin{pmatrix} \lambda_{1,i} & 0 \\ 0 & \lambda_{2,i} \end{pmatrix} \begin{pmatrix} e_{1,i} & e_{2,i} \\ e_{1,i}^\perp & e_{2,i}^\perp \end{pmatrix}$$

$$\pi_i(\mu, \sigma) \leftarrow \frac{\exp(-\frac{1}{2}((\mu; \sigma) - (\mu^*; \sigma^*))^T \Sigma_i ((\mu; \sigma) - (\mu^*; \sigma^*)))}{\sqrt{(2\pi)^k |\Sigma_i|}}$$

**fin Pour**

Poser  $\pi(\mu, \sigma) = \phi(\pi_1(\mu, \sigma); \pi_2(\mu, \sigma))$

---

Concernant cette fonction, plusieurs possibilités sont envisageables, l'important pour l'actuaire est de savoir à quelle région il souhaite attribuer la masse du prior et surtout pour quelles raisons.

L'objectif peut être de privilégier les points se conformant à au moins un des deux jugements, les lignes de niveaux auront alors des formes de croix. Si un équilibre des deux jugements doit être respecté, il est possible de se baser sur des ellipses chargeant des zones entre les deux droites admissibles. Un choix intermédiaire, plus nuancé, pourrait conduire à équilibrer les deux en pénalisant l'écart au point central tout en prenant en compte la conformité à un jugement.

La figure 5.4 donne une illustration de ces trois exemples possibles, chacun pouvant se justifier en fonction des objectifs et des contraintes de conformité aux jugements souhaités.

Toutes les lois sont représentées à travers leur log-vraisemblance, les graphiques ont la même échelle de couleur et les mêmes axes.

Les figures (a) et (b) représentent les ellipses calibrées marginalement sur chaque jugement d'expert, elles chargent bien préférentiellement l'axe iso-quantile et sont centrées sur le point d'intersection des deux bandes proposées par l'expert. Chaque loi couvre un large ensemble de paramètres, ces priors sont diffus.

Le graphique (c) représente une combinaison qui privilégie la conformité aux deux conditions à la fois, la loi est nettement plus informative : sa densité est plus piquée. Cette log vraisemblance a été obtenue en moyennant celle de (a) et celle de (b), ainsi  $\phi(x, y) = \frac{x+y}{2}$ .

La vraisemblance (d) favorise l'adéquation à au moins un des jugements, elle s'obtient en prenant le max des densité (a) et (b), ie.  $\phi(x, y) = \max(x, y)$ . Cet a priori est lui aussi particulièrement diffus.

Finalement, la log vraisemblance (e) a été obtenue avec la fonction  $\phi(x, y) = 0.5 * \max(x, y) + 0.5 * \frac{x+y}{2}$  ce qui revient à faire la moyenne de la log-vraisemblance (c) et (d). Cela correspond à un choix médian, le prior reste relativement diffus et charge une vaste gamme de paramétrage.

Il est à noter que ces lois sont sensibles aux choix des valeurs propres. La longueur et la largeur des ellipses sont réglables à souhait.

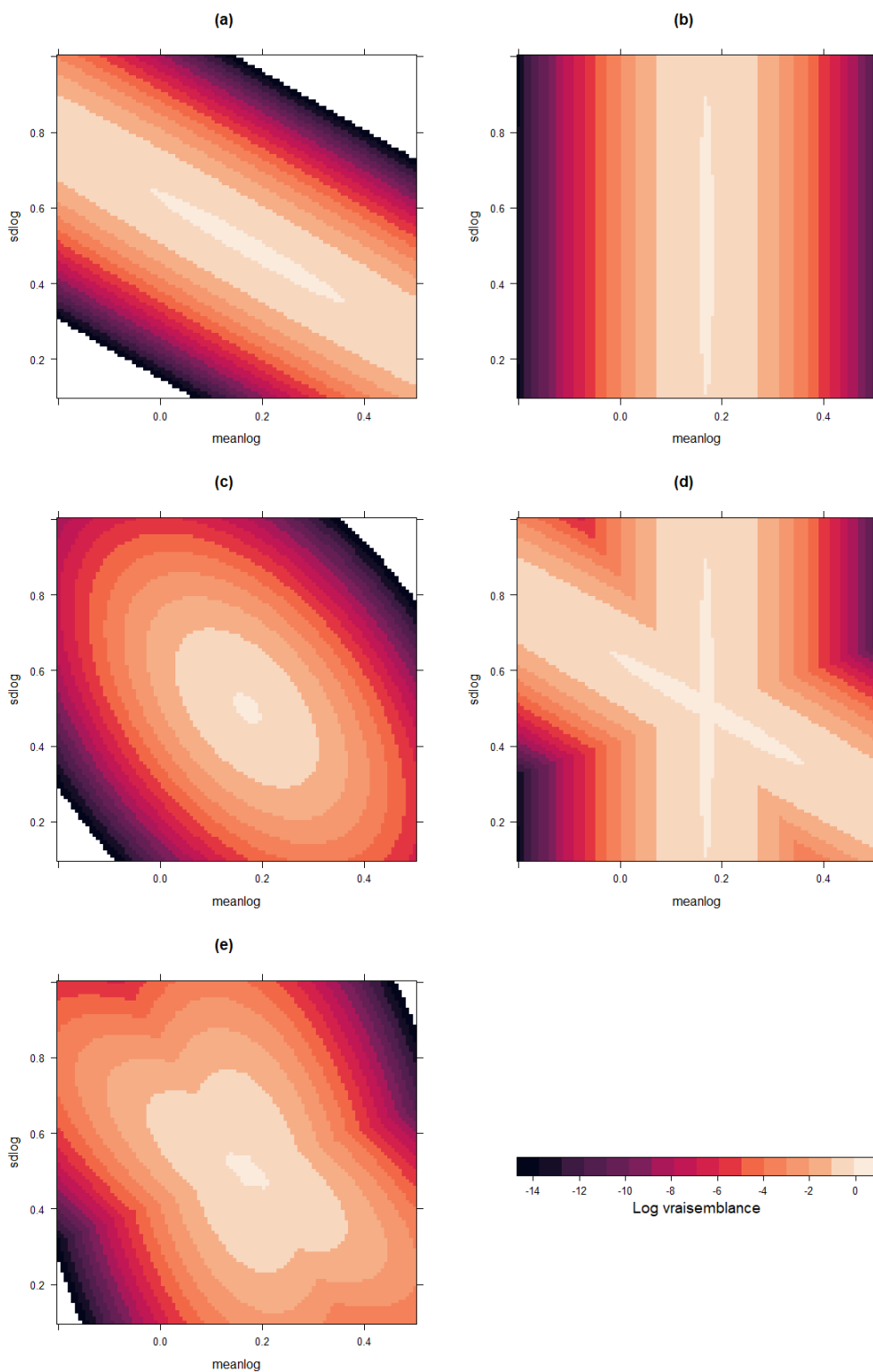


FIGURE 5.4 – Propositions de prior à partir des quantiles donnés par l’expert.

Cette méthode constitue une première approche qui vise à charger les régions données par l’expert. L’idée consiste à trouver des distributions simples qui charge principalement une zone précise, puis de les combiner pour réussir à charger du mieux possible la région visée. Néanmoins, il paraît clair que la méthode reste archaïque et ne répond pas à un critère rationnel et objectif.

Si la finalité est de réussir à charger l’espace des paramètres en fonction de la

conformité au jugement, il est alors nécessaire que deux points vérifiant de la même manière la contrainte se retrouvent chargés de la même façon. Dès lors, il faut se munir d'un critère pour quantifier le fait que deux points "vérifient la contrainte de la même manière".

La contrainte concernant la valeur prise par une fonction  $f$ , la matière du critère recherché devrait rationnellement être les images prises par la fonction sur les deux points considérés.

Par suite, un point l'image est plus proche du niveau spécifié se doit d'être plus chargé qu'un autre dont l'image serait plus éloignée.

La méthode géométrique présentée dans cette section se basait sur la distance dans l'espace  $\Theta$  entre un point et le seul point vérifiant toutes les contraintes pour quantifier la proximité aux jugements des experts. C'était donc la position dans l'espace qui était évaluée et non pas la valeur prise par la fonction  $f$ . La méthode faisait implicitement l'hypothèse qu'un point plus proche dans l'espace des paramètres était également plus proche en termes d'image par la fonction  $f$ , ce qui n'a pas de raison d'être vérifié.

Dans la section suivante, l'a priori en un point dépendra directement de la valeur de la fonction  $f$ , évaluée en ce même point. De plus, la valeur du prior sera d'autant plus grande que la distance entre  $f(\mu, \sigma)$  et  $u$  est petite. Cette méthode a l'avantage de ne pas nécessiter d'explicitement les lignes de niveaux et d'être plus respectueuse du jugement fourni par l'expert.

### 5.2.1.3 Cas général, exemple des coûts de réassurance

Alors que l'obtention de la valeur d'un quantile de la charge puisse ne pas être une priorité opérationnelle et par conséquent limiter la finesse et la pertinence des connaissances à son propos, les coûts de réassurance sont généralement mieux établis et documentés. Les équipes spécialisées dans le domaine de la réassurance investissent des ressources significatives pour développer et maintenir des modèles actuariels fiables permettant d'estimer les primes pures avec précision. De plus, la direction de la réassurance entretient également des contacts et interagit lors du processus de tarification avec d'autres acteurs du secteur qui disposent de données supplémentaires, ce qui leur permet d'avoir une idée des opinions et des évaluations sur un plus grand historique des coûts appropriés.

Étant donné l'enjeu financier crucial lié à ces estimations, il est raisonnable de s'attendre à ce que les primes pures de réassurance bénéficient d'un niveau de fiabilité supérieur à celui de quantiles pouvant être issus de prévisions comptables tenant généralement plus d'un objectif que d'un réel attendu.

Bien que la prime pure d'un contrat de réassurance puisse ne pas être directement communiquée par la contrepartie, il est toujours possible de l'estimer à partir d'une évaluation du retour sur investissement souhaité par le réassureur ou la cédante.

La méthodologie au sein de cette partie sera toujours la même, déterminer les zones de  $\Theta$  conformes à l'avis d'expert puis construire une loi les chargeant préférentiellement. Ainsi, avant de continuer, il est nécessaire de pouvoir évaluer l'adéquation d'un jeu de paramètres avec le jugement.

Pour cela, il faut commencer par trouver la fonction  $f$  telle que l'avis d'expert puisse se mettre sous la forme  $f(S/P) = u$ .

Les types et les modalités de contrats de réassurance sont multiples et prennent diverses formes, il n'y a donc évidemment pas de fonction universelle valable dans tous les cas. Dans cette section, l'intérêt sera porté sur l'exemple d'un traité communément appelé Stop-Loss où le réassureur n'intervient qu'à partir un certain seuil de sinistralité et prend en charge l'intégralité des pertes au-delà. La raison derrière ce choix réside dans le fait que de nombreux contrats, tels que les Quote-Part ou les XS, peuvent se démembrer en une combinaison linéaire de Stop Loss.

D'un point de vue mathématique, la variable à considérer est  $(X - T_1)\mathbb{1}_{X > T_1}$  où  $X$  représente le  $S/P$  et  $T_1$  le seuil d'intervention de la réassurance. La traduction d'une valeur de prime pure en tant que jugement d'expert est alors  $\mathbb{E}[(X - T_1)\mathbb{1}_{X > T_1}] = u$  et la fonction est donc  $f(\cdot) = \mathbb{E}[(\cdot - T_1)\mathbb{1}_{\cdot > T_1}]$ . Pour une vraisemblance  $X|\theta$  log-normale et un paramétrage classique  $\theta = (\mu, \sigma)$ , le coût d'un traité de tranche basse  $T_1$  est  $f(\mu, \sigma) = \exp(\mu + \frac{1}{2}\sigma^2) \left[ 1 - \Phi\left(\frac{\log(T_1) - \mu - \sigma^2}{\sigma}\right) \right] - T_1 \cdot \Phi\left(\frac{\log(T_1) - \mu}{\sigma}\right)$ . Trouver une ligne de niveau de cette fonction revient à trouver  $\psi$  telle que  $f(\psi(\sigma), \sigma) = K$ . En dérivant selon  $\sigma$ , il est possible d'aboutir à une équation différentielle que doit vérifier  $\psi$  et donc dont la résolution analytique permet d'obtenir une équation des courbes de niveaux.

A titre illustratif, l'équation de la forme  $\psi'(x) = g(\psi(x), x)$  que doit vérifier  $\psi$  ici est :

$$\psi'(\sigma) = \frac{\sigma\sqrt{2\pi}\exp(\kappa_1(\sigma))[\Phi(\kappa_2(\sigma)) - 1] + \left(\frac{\kappa_3(\sigma)}{\sigma} - 1\right)\exp\left(-\frac{1}{2}\kappa_2(\sigma)^2 + \kappa_1(\sigma)\right) - \frac{T_1\kappa_3(\sigma)}{\sigma}\exp\left(-\frac{1}{2}\kappa_3(\sigma)^2\right)}{\sqrt{2\pi}\exp(\kappa_1(\sigma))[1 - \Phi(\kappa_2(\sigma))] + \frac{1}{\sigma}\exp\left(-\frac{1}{2}\kappa_2(\sigma)^2 + \kappa_1(\sigma)\right) + \frac{T_1}{\sigma}\exp\left(-\frac{1}{2}\kappa_3(\sigma)^2\right)}$$

Avec :

- $\kappa_1(\sigma) = \psi(\sigma) + \frac{1}{2}\sigma^2$
- $\kappa_2(\sigma) = \frac{\log(T_1) - \psi(\sigma) - \sigma^2}{\sigma}$
- $\kappa_3(\sigma) = \frac{\log(T_1) - \psi(\sigma)}{\sigma}$

En pratique, il devient très vite irréaliste d'essayer de résoudre analytiquement l'équation différentielle liée aux courbes de niveau. En outre, la visualisation des lignes de niveaux peut se faire plus simplement en traçant la fonction  $f$  sur une grille de couples  $(\mu, \sigma)$  comme dans la figure 5.5. Néanmoins, même avec une visualisation de ces courbes, leur forme est souvent si erratique qu'il est difficile de construire visuellement un prior les chargeant comme cela a été fait dans la section précédente. Le besoin d'une méthodologie plus générale et surtout plus automatique se fait donc rapidement ressentir.

Si l'objectif est d'obtenir des iso-densités proches des courbes de niveau de la fonction du jugement, de sorte à ce qu'une même adéquation à l'avis d'expert aboutisse à une même densité, alors une solution directe serait de faire dépendre la densité de la fonction  $f$ .

La solution proposée dans cette section est basée sur cette idée, le prior sera de la forme  $\pi(\mu, \sigma) \propto \phi(d(f(\mu, \sigma), K))$  où  $d$  est une fonction de distance et  $\phi$  une fonction décroissante. Ainsi, plus  $f(\mu, \sigma)$  est proche du niveau  $K$  spécifié par l'expert, plus la densité sera élevée.

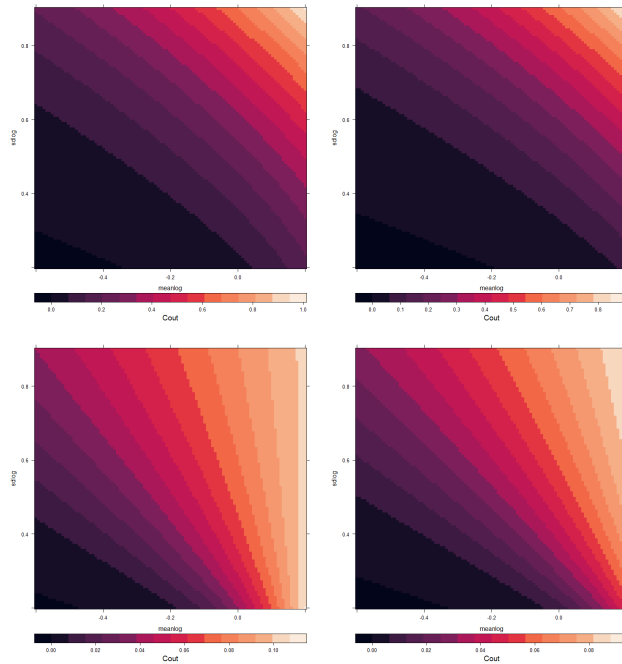


FIGURE 5.5 – Coût de la réassurance en fonction des paramètres. Les axes sont les mêmes mais chaque figure dispose d'une échelle de couleur appropriée. Les contrats sont, dans le sens de la lecture, un Stop-Loss 110%, un Stop-Loss 130%, un XS 110%-130% et un XS 130%-150%.

Il faut alors choisir une fonction de distance, pour cela plusieurs options sont envisageables. Par analogie à une loi normale, une distance quadratique et la fonction  $\phi(\cdot) \propto \exp(-\frac{\cdot}{\tau})$  peuvent être choisies. L'hyperparamètre  $\tau$  permet alors de fixer la dispersion de l'a priori : le prior sera d'autant plus diffus que  $\tau$  est grand. Des représentations de log-densités pour diverses valeurs de  $\tau$  sont données en figure 5.6. Il est important de garder à l'esprit que les courbes de niveaux de tous ces priors sont, par construction, exactement les mêmes que celles de la fonction  $f$  : si  $f(\mu, \sigma)$  est constante sur un ensemble alors  $\exp\left(-\frac{f((\mu, \sigma) - K)^2}{\tau}\right)$  le sera aussi. Ce que permettent de fixer  $\phi$ ,  $d$  et  $\tau$  est tout simplement les valeurs prises sur chacune de ces lignes de niveau.

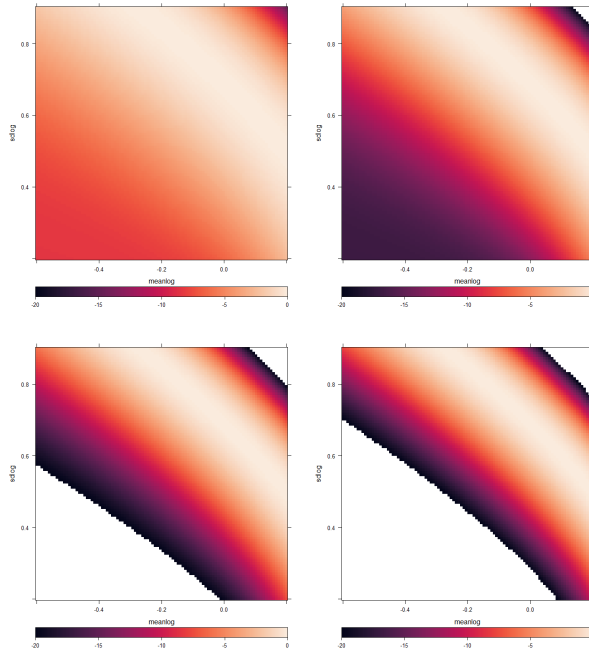


FIGURE 5.6 – Log-densité de l’a priori en fonction de l’hyperparamètre  $\tau$ . Dans le sens de la lecture  $\tau = \frac{1}{50}, \frac{1}{100}, \frac{1}{150}$  et  $\frac{1}{200}$ . Plus  $\tau$  est grand, plus le prior sera diffus.

## 5.2.2 Loi non-informative

Conformément à l’introduction de la section 5.2, la méthode de construction de loi non-informative adoptée ici est celle proposée par Bernardo et Berger.

Dans le cas présent, deux paramètres,  $\mu$  et  $\sigma$ , sont inconnus. L’hypothèse de log-normalité signifie tout simplement que le logarithme des données est distribué selon une loi normale, les paramètres d’une loi log-normale étant effectivement la moyenne et l’écart-type du logarithme des données.

Ainsi, cela revient au même d’inférer un couple  $(\mu, \sigma)$  sur une série de données  $(x_1, \dots, x_n)$  suivant une loi log-normale que d’inférer  $(\mu, \sigma)$  à partir des données  $(\log(x_1), \dots, \log(x_n))$  distribuées normalement. Cette observation permet de faciliter l’éllicitation du prior de référence, en autorisant à utiliser celui de la loi normale et en réalisant l’inférence sur le logarithme des données. D’après (Yang et Berger, 1996)[3], le prior de référence dans ce cas-ci est  $\pi(\mu, \sigma) = \frac{1}{\sigma}$ .

## 5.3 Obtention de la loi prédictive a posteriori

Pour rappel, un calcul bayésien se fait selon les étapes suivantes :

- Choix d’un modèle  $\{\Theta, \pi, f(\cdot|\theta)\}$  et de l’espace des décisions  $\mathcal{D}$  ;
- Choix d’une fonction de coût  $L$  adaptée au problème et quantifiant la pertinence d’une décision pour un état des causes fixés ;
- Observation d’un échantillon et actualisation de la loi a priori en une loi a posteriori via le théorème de Bayes ;
- Recherche de la décision optimale qui minimise le risque a posteriori pour l’échantillon observé.



Le choix de la fonction de coût a été traité en section 3.2.2, et celui de la loi a priori dans la section 5.2. Le théorème de Bayes ayant déjà été mentionné, il ne reste plus qu'à définir une méthode permettant d'obtenir la décision optimale.

Une première voie attrayante est alors celle du calcul littéral. Le modélisateur dispose en effet de toutes les expressions dont il a besoin et, s'il venait à être couronné de succès, il verrait alors se dévoiler devant lui une solution à la fois rapide d'implémentation et exacte.

Toutefois, une telle entreprise représente une route sinueuse dans laquelle il est bien souvent vain de s'aventurer, particulièrement dans un cadre bayésien.

A titre illustratif, la densité de la loi prédictive a posteriori se définit de la manière suivante :

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

Une méthode pour obtenir les quantiles souhaités serait de commencer par intégrer cette densité pour obtenir une fonction de répartition, puis de l'inverser pour aboutir à une fonction quantile. Outre la difficulté qu'entraîneraient de tels calculs analytiques, la densité de la loi a posteriori n'est généralement connue qu'à une constante près. Bien que des méthodes d'intégrations numériques successives puissent être déployées, la complexité et la succession d'approximations d'une telle démarche pousse à se tourner vers une approche par simulation. La minimisation de la seconde fonction de coût par des moyens théoriques n'est pas tellement plus aisée.

Pour ces diverses raisons, la recherche de la décision optimale se fera à l'aide de l'algorithme de Metropolis-Hastings dont les principales propriétés vont être rappelées dans la section suivante.

## 5.3.1 Algorithme de simulation

### 5.3.1.1 Présentation de l'algorithme

L'algorithme de Metropolis-Hastings (abrégé MH), inventé par Nicholas Metropolis et publié par William Hastings dans les années 1970, est une méthode permettant de simuler selon une loi cible de densité  $\nu$  connue lorsque l'échantillonnage direct n'est pas possible.

L'idée fondamentale de cet algorithme est de considérer une loi, dite loi instrumentale, pour laquelle une procédure d'échantillonnage rapide et peu coûteuse est connue. L'algorithme consiste alors à générer une suite de données sous la loi instrumentale et de n'en accepter que certains éléments, de sorte à retomber sur un échantillon distribué selon la loi cible.

Pour ce faire, chaque candidat proposé selon la loi instrumentale va voir ses chances d'avoir été générés selon  $\nu$  évaluées à l'aide d'un "ratio d'acceptation". Ensuite, l'étape finale de l'itération est de l'accepter, auquel cas le nouveau point serait ajouté au bout d'une liste appelée chaîne, ou de le refuser en fonction du résultat d'un tirage aléatoire dont la probabilité des issues dépend du ratio d'acceptation. D'après la théorie, la chaîne ainsi produite est une chaîne de Markov ayant pour loi

stationnaire  $\nu$  : au bout d'un certain nombre d'itérations, les éléments de la chaîne sont distribués selon la loi souhaitée.

---

**Algorithme 3** Algorithme de Metropolis-Hastings

---

Fournir un point de départ  $x_0$ , une taille d'échantillon  $N$  souhaitée, une loi de transition instrumentale  $h(\cdot|\cdot)$  et une fonction  $f \propto \nu$ .

$n \leftarrow 0$

$x \leftarrow x_0$

$C = \{\}$

**tant que**  $n < N$  **faire**

    Tirer  $x' \sim h(\cdot|x)$

    Calculer le ratio d'acceptation  $r \leftarrow \frac{\nu(x')}{\nu(x)}$

    Tirer  $B \sim \mathcal{B}(r)$

**si**  $B = 1$  **alors**

$x \leftarrow x'$  et  $C = C \cup \{x'\}$

**fin si**

**fin tant que**

**retourne**  $C$

---

L'une des spécificités de l'algorithme réside dans son mécanisme de génération des données : la loi instrumentale évolue à chaque itération, en fonction de l'état actuel de la chaîne. En conséquence, un état initial doit être fourni en entrée de l'algorithme et, plus important encore, les éléments de la chaîne ne sont pas indépendants. En effet, la loi du candidat dépendant de l'état actuel, le prochain maillon dépendra alors du dernier élément de la chaîne. Il est également à noter que la densité cible  $\nu$  n'a besoin d'être connue qu'à une constante près, la constante de proportionnalité s'annulant dans le ratio d'acceptation.

Lorsque l'objectif est d'obtenir un échantillon *iid*, il s'impose alors d'utiliser un algorithme de décorrélation qui consiste généralement à ne conserver que des points espacés d'une certaine distance afin d'obtenir une certaine forme d'indépendance. L'idée est qu'à partir d'un certain nombre d'étapes, la chaîne a pu parcourir suffisamment de chemin pour que le nouveau point ne dépende plus du dernier état retenu. Une autre option serait de considérer des chaînes simulées séparément. En pratique, les deux options sont souvent exploitées simultanément.

Maintenant qu'il a été montré qu'il était possible de simuler selon une loi à partir d'une connaissance de sa densité à une constante près, des méthodes pour simuler selon la loi prédictive a posteriori vont être données.

### 5.3.2 Simulation selon la loi prédictive a posteriori

L'objectif est ici d'obtenir un échantillon simulé sous la loi prédictive au sens bayésien  $m_{\pi|X}$ . Toutefois cette loi n'est en pratique pas connue, pas même à une

constante près :

$$\begin{aligned}
 m_{\rho|X}(x) &= \int_{\Theta} f(x|\theta)\pi(\theta|X)d\theta \\
 &= \int_{\Theta} f(x|\theta) \frac{\pi(\theta)f(X|\theta)}{\int_{\Theta} \pi(\theta)f(X|\theta)d\theta} d\theta \\
 &\propto \int_{\Theta} f(x|\theta)\pi(\theta)f(X|\theta)d\theta
 \end{aligned}$$

Et la résolution de l'intégrale n'est généralement pas possible à la seule force du stylo.

### 5.3.2.1 Metropolis-Hastings direct

Une façon de contourner le problème consiste réécrire la densité prédictive sous la forme d'une espérance, puis de l'estimer à l'aide d'un estimateur de Monte Carlo :

$$\begin{aligned}
 m_{\rho|X}(x) &= \int_{\Theta} f(x|\theta)\pi(\theta|X)d\theta \\
 m_{\rho|X}(x) &= \mathbb{E}_{\pi(\theta|X)} [f(x|\theta)] \\
 \implies m_{\rho|X}(x) &\approx \frac{1}{N} \sum_{i=1}^N f(x|\theta_i) \quad \text{où } \theta_i \stackrel{iid}{\sim} \pi(\theta|X)
 \end{aligned}$$

Une condition nécessaire pour appliquer cette méthode est de disposer d'un échantillon de réalisations indépendantes  $(\theta_i)_i$  identiquement distribuées selon le posterior  $\pi(\theta|X)$ . Or, c'est justement ce que permet de faire l'algorithme de Metropolis-Hastings, à condition de le coupler avec une méthode de décorrélation.

Ainsi, une variante de l'algorithme de Metropolis-Hastings où le ratio d'acceptation  $\frac{m_{\rho|X}(x')}{m_{\rho|X}(x)}$  serait calculé à l'aide d'un échantillon  $(\theta_i)_i$ , lui-même simulé par Metropolis-Hastings, permettrait potentiellement d'atteindre le but fixé.

Toutefois, le fait que la densité utilisée dans le ratio d'acceptation ne soit qu'une approximation de la densité souhaitée fait manquer de robustesse à la méthode. De plus, elle suppose de stocker un ensemble important  $(\theta_i)_i$  et l'évaluation de  $\frac{1}{N} \sum_{i=1}^N f(x|\theta_i)$  peut être coûteuse computationnellement.

### 5.3.2.2 Simulations par composition

Une manière plus rapide de procéder est de considérer la variable aléatoire  $Z$  définie comme le résultat de l'expérience suivante :

- Tirer  $\theta_i$  selon la loi  $\pi(\theta|X)$ .
- Tirer  $Z_i$  selon la loi  $f(x|\theta_i)$ .

La variable  $Z$  est alors distribuée selon la loi prédictive a posteriori. En effet,

pour  $K \in \mathcal{X}$  :

$$\begin{aligned}
\mathcal{P}(Z < K) &= \int_{\Theta} \mathcal{P}(Z < K|\theta) f_{\theta_i}(\theta) d\theta \\
&= \int_{\Theta} \mathcal{P}(Z < K|\theta) \pi(\theta|X) d\theta \\
&= \int_{\Theta} \int_{z < K} f(z|\theta) dz \pi(\theta|X) d\theta \\
&= \int_{\Theta} \int_{z < K} f(z|\theta) \pi(\theta|X) dz d\theta \\
&= \int_{z < K} \int_{\Theta} f(z|\theta) \pi(\theta|X) d\theta dz \\
\mathcal{P}(Z < K) &= \int_{x < K} m_{\rho|X}(x) dx
\end{aligned}$$

Avec  $f_{\theta_i}$  la densité de la variable  $\theta_i$ , c'est à dire la fonction  $\pi(\theta|X)$ . Une fois de plus, l'interversion des intégrales est possible car toutes les quantités en jeu sont positives.

Ainsi, cette procédure en deux étapes permet de simuler directement selon la loi prédictive a posteriori sans avoir à évaluer sa densité. De plus, étant donné que la simulation selon  $f(\cdot|\theta)$  est directe, la méthode proposée est à peine plus coûteuse que la simulation selon  $\pi(\theta|X)$ . Pour toutes ces raisons, lorsque l'inférence sera basé sur une densité prédictive a posteriori (méthode prédictive bayésienne d'après les conventions de la fin du chapitre 3), les charges de sinistres seront simulées selon l'algorithme présenté en début de section 5.3.2.2. Lorsque l'inférence utilisera une fonction quantile associée à une vraisemblance fixée (méthode prédictive par la vraisemblance d'après les conventions de la fin du chapitre 3), un simple algorithme de Metropolis-Hasting sur le posterior suffira.

Dans tous les cas, des estimateurs de Monte Carlo seront ensuite utilisés pour obtenir la valeur de la décision optimale qui sera soit la médiane soit la moyenne de la loi simulée.

# Chapitre 6

## Mise en pratique

Ce chapitre traitera de la mise en pratique de la méthodologie développée dans le chapitre précédent et contiendra trois sections. La première traitera des données, c'est-à-dire des historiques et des jugements utilisés dans ce mémoire, tandis que les deux suivantes sépareront les calibrages réalisés avec des a priori non-informatifs de ceux obtenus à l'aide d'a priori informatifs.

### 6.1 Données disponibles

#### 6.1.1 Historiques utilisés

Dans cette étude, les historiques analysés portent sur les ratios S/P de différents segments inclus dans les principales lignes d'activités de Groupama : les LoB MTPL, Motor et Fire. Les segments considérés sont tous composés de sinistralité attritionnelle.

Bien que cette dernière soit généralement moins volatile que la sinistralité grave, et qu'elle contribue de ce fait moins à l'atteinte de quantiles extrêmes du résultat technique, elle joue toutefois un rôle prépondérant dans la sinistralité "habituelle" de la plupart des scénarios.

Conformément aux directives des régulateurs, l'utilisation d'un modèle interne au sein d'une entreprise ne doit pas se borner à la seule détermination du capital de solvabilité requis. Au contraire, il est attendu qu'il alimente également le processus de pilotage et la prise de décision stratégique de la compagnie. Dans cette perspective, une modélisation robuste de la sinistralité attritionnelle revêt une importance cruciale qu'il convient de ne pas sous-estimer.

Or, il est paradoxal de constater que, malgré leur importance dans le nombre total de sinistres, ces segments attritionnels présentent généralement des échantillons réduits, comptant à peine plus d'une vingtaine de points dans le meilleur des cas. Cette situation s'explique par le fait que les données historiques relatives à ces segments sont agrégées entre elles, de sorte à ne former qu'un seul ratio sinistres à primes par année. Cette pauvreté de l'historique engendre une sensibilité accrue de la sinistralité attritionnelle à l'incertitude sur paramètres. Par conséquent, les segments contrôlés en premier dans cette étude sont ceux qui sont les plus susceptibles de fournir des écarts significatifs et qui, de ce fait, constituent une borne haute dans l'évaluation de l'incertitude sur paramètres.

## 6.1.2 Jugements d'expert

Dans l'idée de préserver la confidentialité des données de l'entreprise, ainsi que pour conserver la possibilité de remettre ce mémoire sans clause de confidentialité, les jugements d'experts utilisés dans ce mémoire ne correspondent pas à des véritables données de l'entreprise. L'objectif de ce chapitre est de donner un exemple de traitement de jugements et bien que l'étude perde légèrement en vraisemblance, la mise en pratique des méthodes ne nécessite pas de données exactes. Toutefois, pour maintenir une cohérence dans les résultats énoncés, les ordres de grandeur seront tout de même choisis de sorte à rester crédibles.

En outre, le choix de ne pas utiliser directement les données de l'entreprise a également été motivé par la difficulté d'obtenir des jugements pertinents sur un nombre suffisant de survenances passées. Sans cet historique, il est impossible de mener à bien un *backtesting* satisfaisant et ainsi de pouvoir valider une quelconque source de données externes de manière rigoureuse.

## 6.2 Modélisation bayésienne avec a priori objectifs

L'intérêt principal de cette section est d'estimer l'impact d'une prise en compte de l'incertitude portant sur les valeurs des paramètres, cette incertitude étant traduite sous la forme d'une distribution de probabilité des causes.

Ici, aucune information externe ne vient enrichir la modélisation, l'objectif étant simplement de se plonger dans un cadre où le paramétrage n'est plus fixe mais où toutes ses valeurs probables sont prises en compte.

### 6.2.1 Obtention de la loi prédictive

Conformément aux chapitres précédents, l'a priori employé sera celui de Bernardo-Berger. L'inférence se fait alors en plusieurs étapes. En premier lieu, la loi a posteriori est obtenue à partir du principe d'actualisation sur les log-densités :

$$\log(\pi(\theta|X)) = \log(f(X|\theta)) + \log(\pi(\theta)) + Const$$

Une connaissance de la densité à une constante multiplicative près étant suffisante à l'application de l'algorithme de Metropolis-Hastings, la connaissance de la log-densité à une constante additive près est suffisante pour procéder à des simulations. L'égalité à une constante additive près sera notée " $\equiv$ ".

La modélisation se faisant sur le logarithme des données avec une vraisemblance normale, le log-posterior a pour expression :

$$\log(\pi(\theta|X)) \equiv -\log(\sigma) - \frac{1}{2} \sum \left( \frac{x_i - \mu}{\sigma} \right)^2 - \log(\sigma)$$

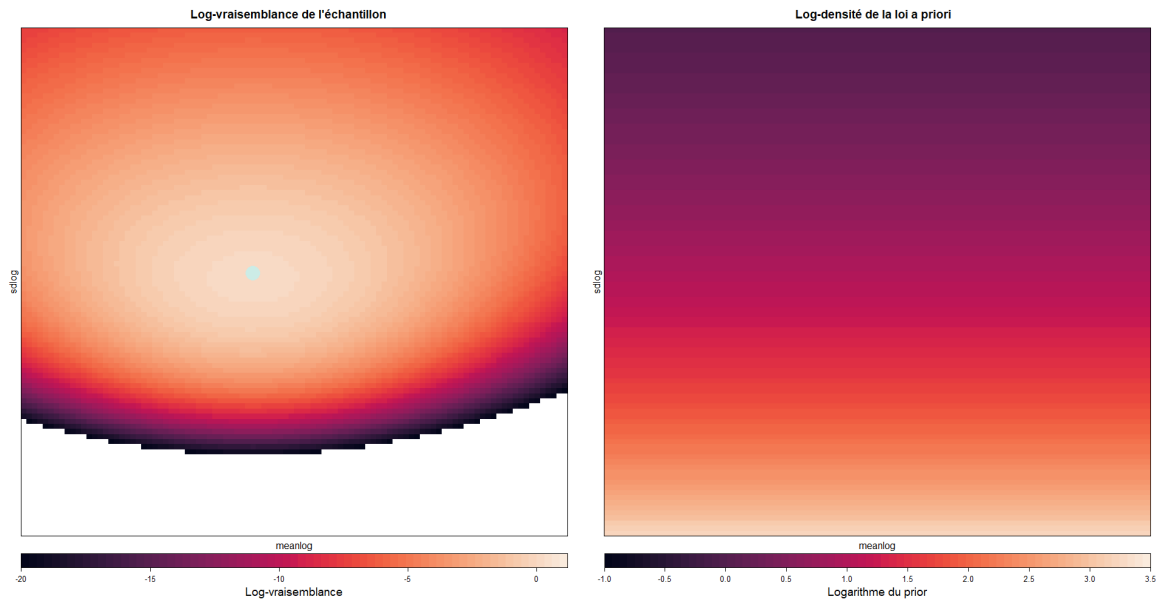


FIGURE 6.1 – Représentation de la log-vraisemblance et du logarithme du prior. Le prior  $(\mu; \sigma) \rightarrow \frac{1}{\sigma}$  est bien indépendant de  $\mu$  et décroissant en  $\sigma$ . La log-vraisemblance est unimodale et concentrée autour d'un point représentant le paramétrage par maximum de vraisemblance (en bleu sur le graphique). La différence entre les échelles des variations du prior et le vraisemblance implique que la loi a posteriori sera essentiellement formée par la vraisemblance.

Le prior et la vraisemblance sont représentés en figure 6.1. L'échelle de valeur étant nettement plus élevée pour la vraisemblance, cette dernière est prépondérante dans la formation du posterior (figure 6.2).

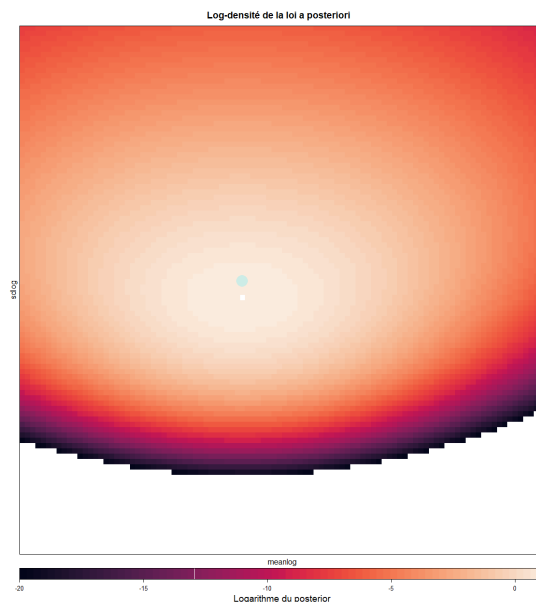


FIGURE 6.2 – Représentation du posterior (échelle logarithmique). Ce dernier ressemble effectivement à la vraisemblance à la différence que l'ensemble de la distribution a été légèrement déplacé vers le bas. En particulier le mode de la distribution à posteriori (point blanc) est légèrement en dessous de l'estimateur du maximum de vraisemblance.

Par la suite, l'algorithme détaillé en section 5.3.1 est employé pour simuler trois chaînes parallèles qui seront retraitées pour obtenir l'échantillon final selon les étapes suivantes :

- simulation de trois chaînes indépendantes ;
- retrait des premières itérations correspondant au temps de chauffe ;
- échantillonnage au sein de chaque chaîne avec un pas permettant d'avoir une corrélation non significative entre deux simulations consécutives ;
- échantillonnage parmi les trois chaînes parallèles.

Le diagnostic de Gelman Rubin (figure 6.3) est utilisé dans un premier temps pour obtenir le temps de convergence de la loi de simulation vers le posterior et contrôler la stationnarité à partir de ce point. Par la suite, des tests d'auto-corrélation (figure 6.4) sont utilisés pour déterminer le pas permettant de décorréliser deux simulations successives. Finalement, une représentation graphique de l'échantillon permet de vérifier que toutes les zones chargées par le posterior ont été visitées et sont représentées dans l'échantillon final.

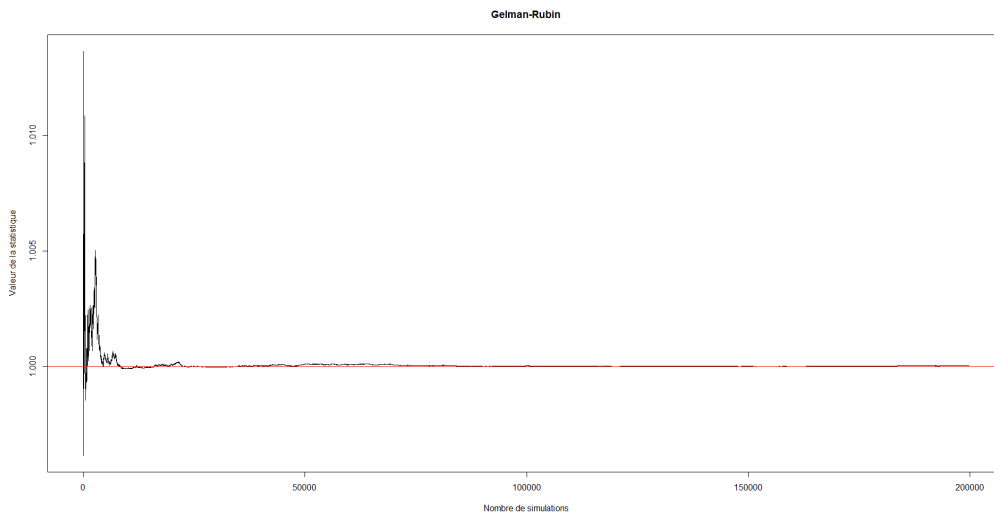


FIGURE 6.3 – Diagnostic de Gelman-Rubin à trois chaînes, seules les 200 000 premières simulations ont été représentées.

Les chaînes ont convergé entre elles vers la loi cible lorsque la valeur de la statistique a convergé vers 1. Ici cette convergence a lieu assez rapidement, vers 10 000 simulations.



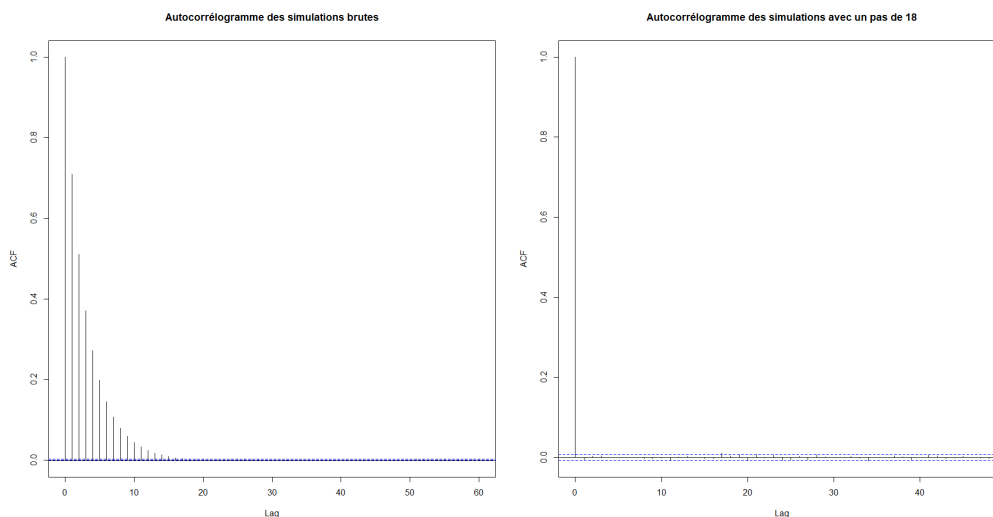


FIGURE 6.4 – Autocorrélogramme des simulations. Un pas de l'ordre de la vingtaine est nécessaire pour obtenir des simulations successives décorréélées.

Toutes ces étapes aboutissent à un échantillon de paramétrages supposés *iid* et distribués selon le posterior. Il reste alors à faire le passage vers des quantiles de loi prédictives en déterminant les estimateur de Bayes.

## 6.2.2 Évaluation de l'impact de l'incertitude sur paramètres

Dans cette section, les deux types de lois prédictives retenues, prédictive bayésienne et prédictive par la vraisemblance, seront d'abord traitées séparément avant d'être confrontées l'une à l'autre.

### 6.2.2.1 Loi prédictive bayésienne

En premier lieu, l'échantillon de prédiction est obtenu à partir de l'algorithme 5.3.2.2 en simulant des charges à partir de paramètres tirés selon le posterior. Un histogramme de l'échantillon de charge ainsi simulé est donné en figure 6.5.

Afin de mieux visualiser la différence avec la loi sans incertitude sur paramètres (qui est une log-normale), un histogramme du logarithme de l'échantillon est également donné en figure 6.6. La comparaison avec la densité obtenue sans incertitude met en évidence la globale proximité entre deux lois : chacune est symétrique, unimodale et rappelle une courbe en cloche gaussienne. Cependant, une différence notable se manifeste par la légère leptokurtose de la loi prenant en compte le risque sur paramètres : ses queues sont quelque peu plus lourdes.

Ainsi, bien que l'histogramme rappelle une courbe en cloche caractéristique d'une loi normale, l'excès de kurtosis observé (voir table 6.1) exclut cette distribution pour l'échantillon, les prédictions ne peuvent pas suivre une loi log-normale.

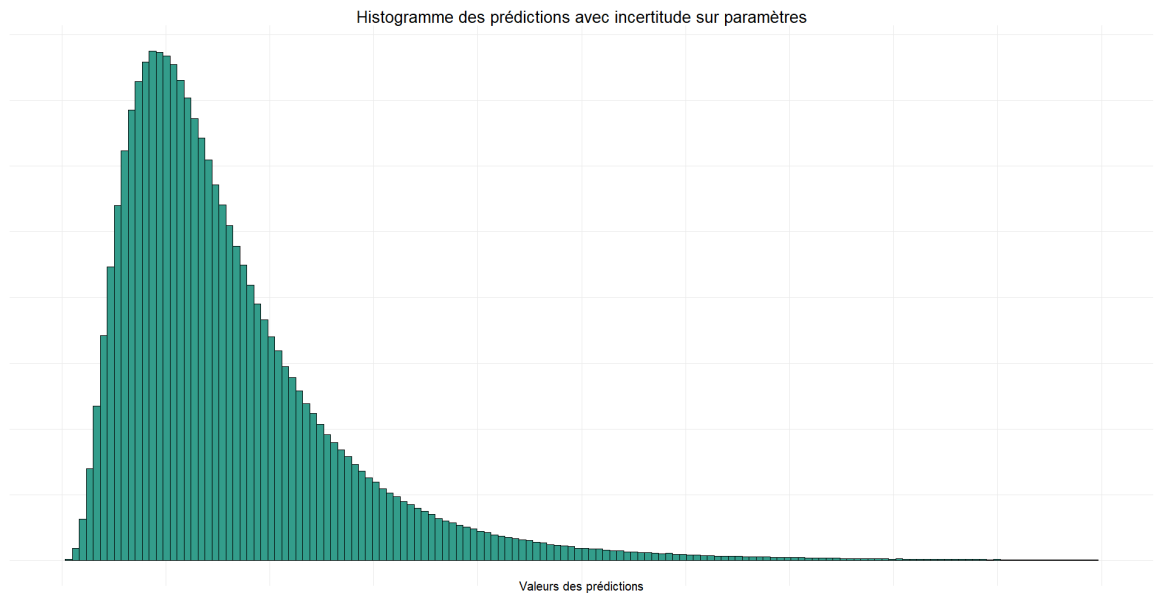


FIGURE 6.5 – Distribution des prédictions avec incertitudes sur paramètres, méthode 1. La répartition obtenue est proche de celle d’une loi log-normale.

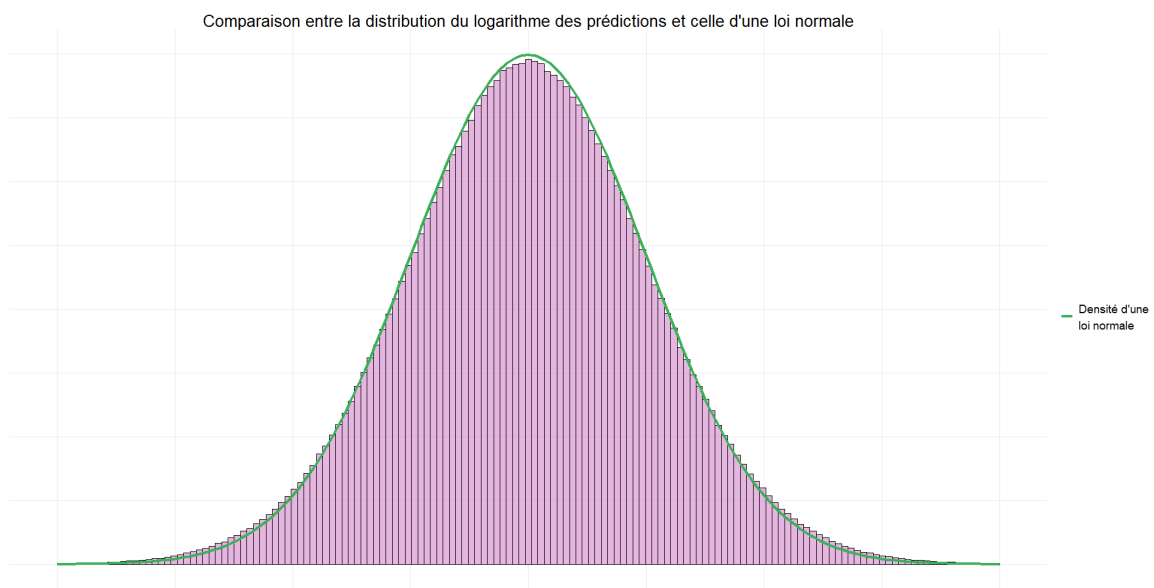


FIGURE 6.6 – Comparaison de la distribution du log des prédictions avec (histogramme rose) et sans (ligne verte) incertitude sur paramètres, méthode 1. Par comparaison, la log-distribution sans incertitude, qui est une loi normale, charge moins les queues de la distribution et charge plus son mode.

Un candidat naturel qui est à la fois symétrique, unimodal, proche de la loi normale mais légèrement plus leptokurtique est la loi de Student. Il s’avère en effet que l’histogramme du logarithme des prédictions se conforme très bien à cette loi (figure 6.7), la loi prédictive a posteriori serait donc une log-Student.

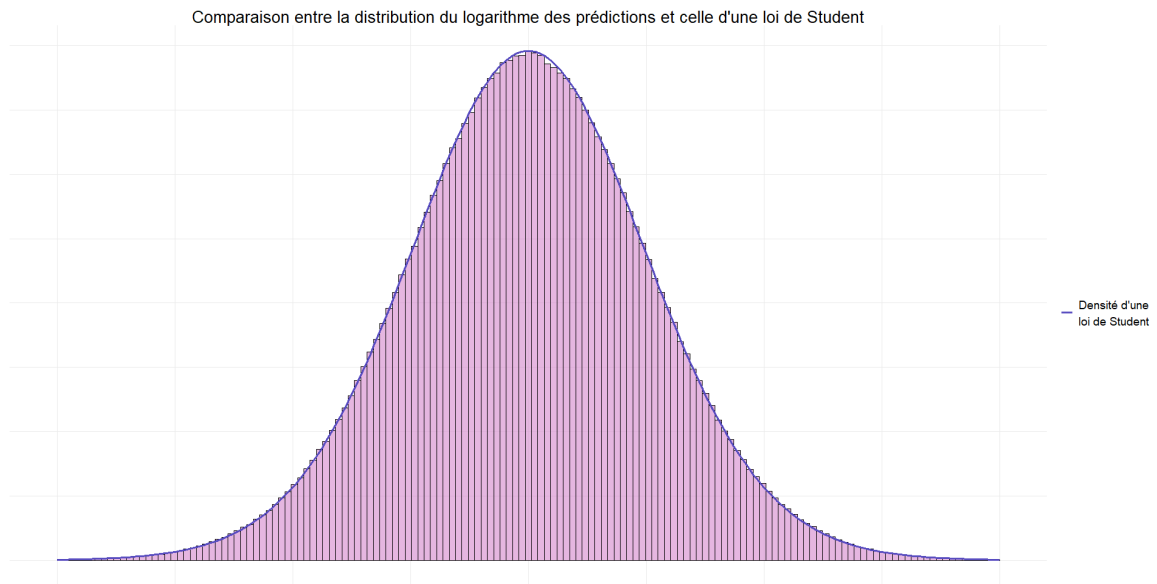


FIGURE 6.7 – Log-distributions prédictives avec incertitudes sur paramètres, méthode 1. La répartition obtenue correspond de manière satisfaisante avec celle d’une loi de Student.

Ce fait n’est pas très surprenant, sachant qu’une vraisemblance log-normale associée à une loi inverse-gamma sur  $\sigma$  aboutit à une log-Student et qu’avec un prior non informatif, le posterior sur  $\sigma$  suit effectivement une loi inverse-gamma. De plus, il est à noter que le nombre de degrés de liberté de la log-Student est théoriquement relié au nombre d’observations dans l’échantillon, par conséquent la loi prédictive converge bien asymptotiquement vers une log-normale.

La figure 6.8 donne l’écart relatif entre le quantile prédictif bayésien et celui obtenu sans incertitude sur paramètres en fonction du niveau du quantile. La courbe semble relativement symétrique en valeur absolue autour de  $1/2$ , où elle est presque plate, et devient très pentue dans les extrêmes. Conformément au graphique 6.5, les quantiles bas seraient sur-estimés et les quantiles hauts sous-estimés lorsque l’incertitude sur paramètres n’est pas prise en compte. L’écart vaut 2,5% pour un quantile à 10 ans et monte jusqu’à 9,5% pour un quantile à 200 ans, l’incertitude sur paramètres n’est alors pas négligeable selon les résultats de cette méthode.

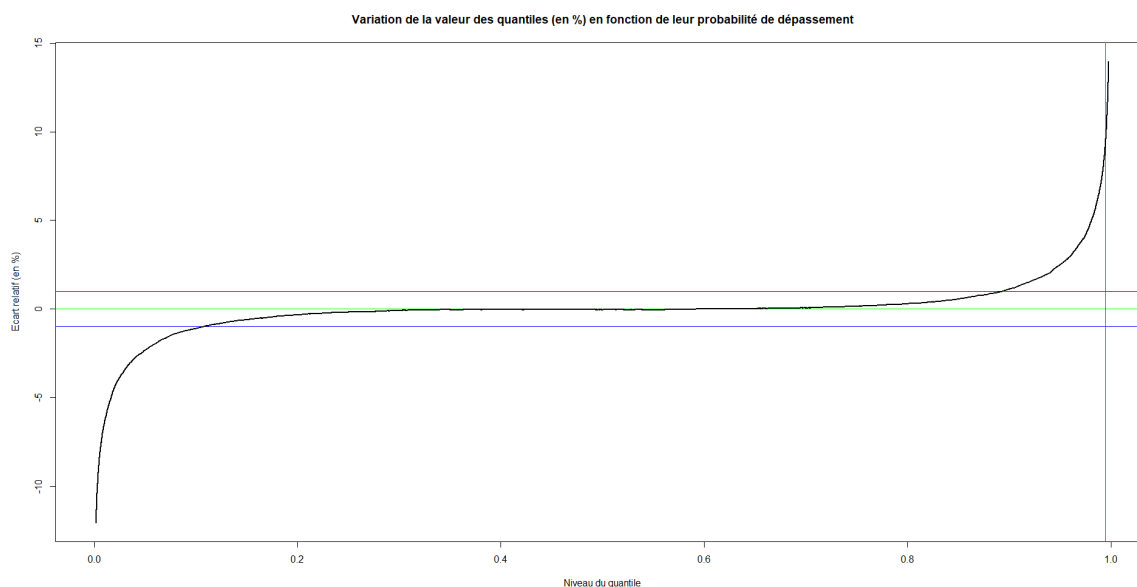


FIGURE 6.8 – Écarts relatifs sur les quantiles entre la méthode paramétrique et la méthode 1. La courbe noir représente, pour un niveau de quantile fixé, l'écart relatif entre le quantile de ce niveau estimé sans incertitude sur paramètres et celui estimé selon la méthode 1.

Les trait bleus représentent la barre des  $\pm 1\%$  d'écart : tous les quantiles dont le niveau est situé approximativement entre 10% et 90% (période de retour entre 1,11 et 10 années) se situent entre ces seuils.

Cependant, l'écart croit significativement en dehors de cet intervalle pour atteindre jusqu'à près de 10% au niveau du quantile 99,5%.

Faudrait-il pour autant provisionner 9,5% de capital en plus ? Sans même parler des effets de diversification qui viendraient atténuer ce constat, le résultat de l'étude interroge. Une des prémices de la modélisation était la supposition que le ratio  $S/P$  pouvait correctement, ou du moins de manière satisfaisante, s'assimiler à une variable suivant une loi log-normale ayant des paramètres fixés. C'est alors tout naturellement que la densité d'une log-normale fut prise comme vraisemblance du modèle. Un prior non informatif y a ensuite été associé afin de pouvoir former un posterior représentant la connaissance disponible quant à la vraie valeur du paramétrage.

Bien qu'il soit habituel dans un cadre bayésien qu'une loi prédictive (la mixture de la vraisemblance et du posterior) ne partage pas les même propriétés que la vraisemblance, ce phénomène se traduit dans cette étude par une vraisemblance log-normale et une loi prédictive log-Student. Certes, il semble cohérent que la prise en compte de l'incertitude sur paramètres aboutisse à des quantiles plus élevés, néanmoins il semble contre-intuitif que la prise en compte de l'incertitude sur paramètres change les caractéristiques de la sinistralité alors même que sa nature, et en particulier le "véritable" montant de solvabilité à provisionner théoriquement, est indépendante de la modélisation que l'actuaire en fait.

Le fait d'aboutir à une loi ayant une structure sensiblement différente, notamment au niveau de la lourdeur de ses queues, est un signe de la trop grande prudence

de cette méthode et constitue un argument en sa défaveur.

### 6.2.2.2 Quantiles prédictifs par la vraisemblance

Comme précédemment, un histogramme des prédictions est donné en figure 6.9. La distribution est proche de celle d'une log-normale et de celle obtenue via les quantiles prédictifs bayésiens.

Afin de pouvoir mieux contrôler la nature de la distribution prédictive, un histogramme des log-prédictions est également représenté en figure 6.10 et 6.11. Il ressort de ces représentations que la loi prédictive obtenue dans cette section correspond effectivement à une distribution log-normale. Son kurtosis normalisé est d'ailleurs nul (-0.009 empiriquement). En formant les quantiles à partir d'une vraisemblance log-normale, la distribution obtenue via un critère d'optimalité bayésien a conservé la structure de la loi souhaitée.

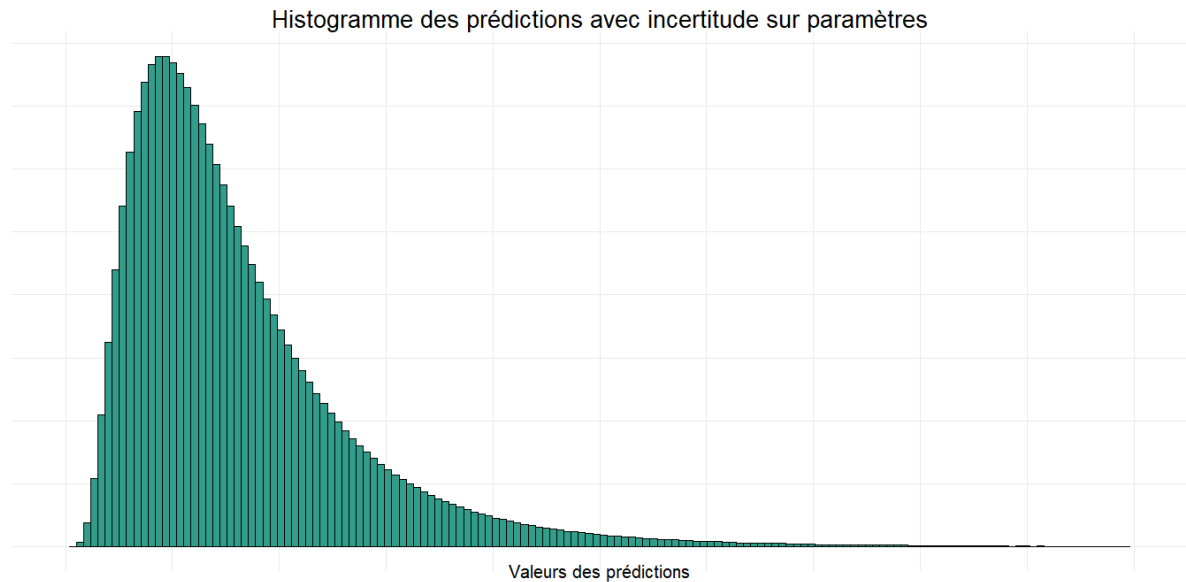


FIGURE 6.9 – Distribution des prédictions avec incertitudes sur paramètres, méthode 2. La répartition obtenue est proche de celle d'une loi log-normale.

Les deux méthodes de prise en compte de l'incertitude sur paramètres ont ensuite été testées sur une série de segments de Groupama dans le but de contrôler le type de la distribution prédictive. Les résultats furent similaires aux observations précédentes : toutes les log-distributions étaient unimodales, symétriques et ressemblant à une gaussienne, seule la lourdeur des queues variait entre les deux méthodes. Le tableau 6.1 recense les kurtosis normalisés de chacune des distributions.

Comparaison entre la distribution du logarithme des prédictions et celle d'une loi normale

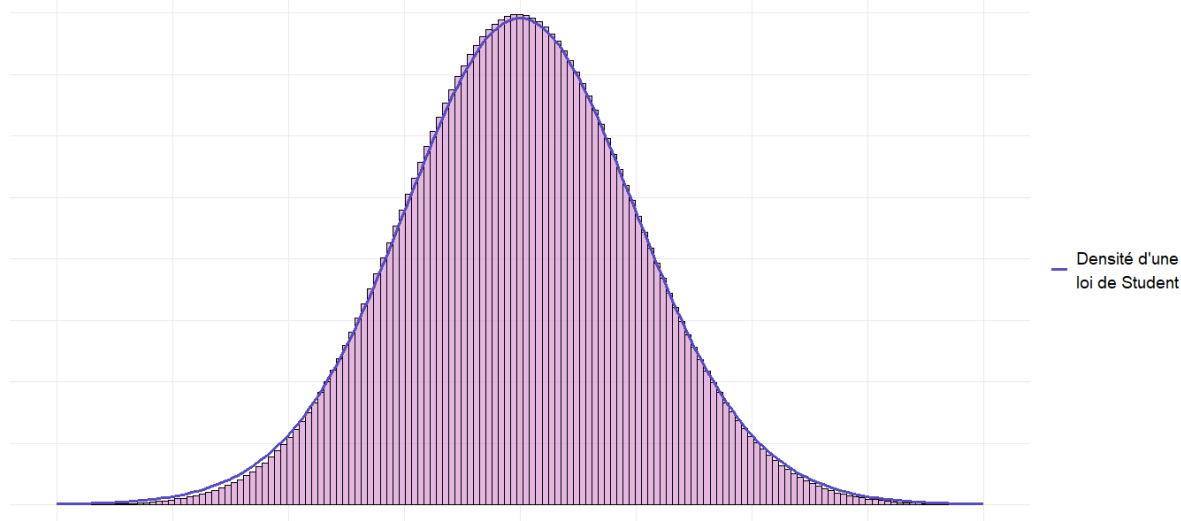


FIGURE 6.10 – Distribution des prédictions avec incertitudes sur paramètres. La loi de Student surcharge les queues et sous-représente le centre de masse

Comparaison entre la distribution du logarithme des prédictions et celle d'une loi normale

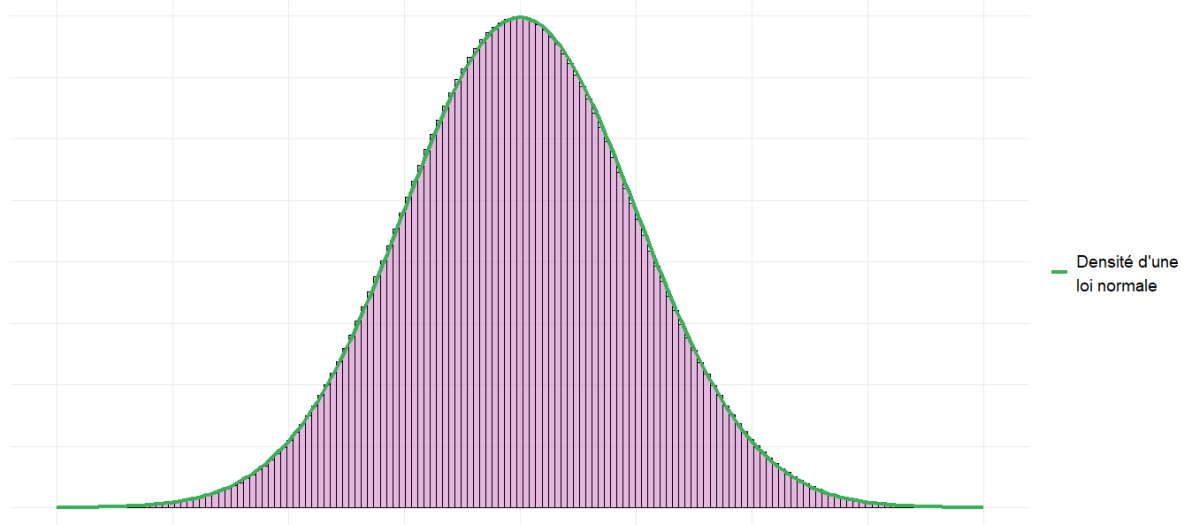


FIGURE 6.11 – Distribution des prédictions avec incertitudes sur paramètres. La densité de la loi normale colle parfaitement à l’histogramme obtenu, cela semble indiquer que la distribution prédictive serait log-normale.

Segment	1	2	3	4	5	6	7	8	9
Méthode 1	18,43%	27,76%	13,73%	21,89%	20,97%	24,88%	21,98%	25,05%	18,5%
Méthode 2	-1,77%	-1,59%	-1,2%	-1,72%	-1,03%	-1,45%	-3,58%	-1,70%	-1,40%

TABLE 6.1 – Valeur des kurtosis normalisés pour plusieurs segments. La méthode 1 correspond aux quantiles prédictifs bayésien tandis que la méthode 2 correspond à la méthode basée sur la fonction quantile de la vraisemblance.

Cette fois ci, les écarts relatifs avec la méthode paramétrique (figure 6.12) sont tous minorés par 0,5%. Cela signifie que les quantiles obtenus avec incertitude sur paramètres sont toujours plus élevés : la prise en compte de l'incertitude sur paramètres ne fait qu'augmenter l'estimation du risque, même si ce n'est que faiblement. La courbe est cette fois symétrique même sans passage à la valeur absolue, elle présente une forme en U. De plus, une différence majeure avec la méthode bayésienne réside dans les niveaux atteints par l'écart : moins de 3,5% pour le quantile 99,5%, une différence modérée pour un test de sensibilité.

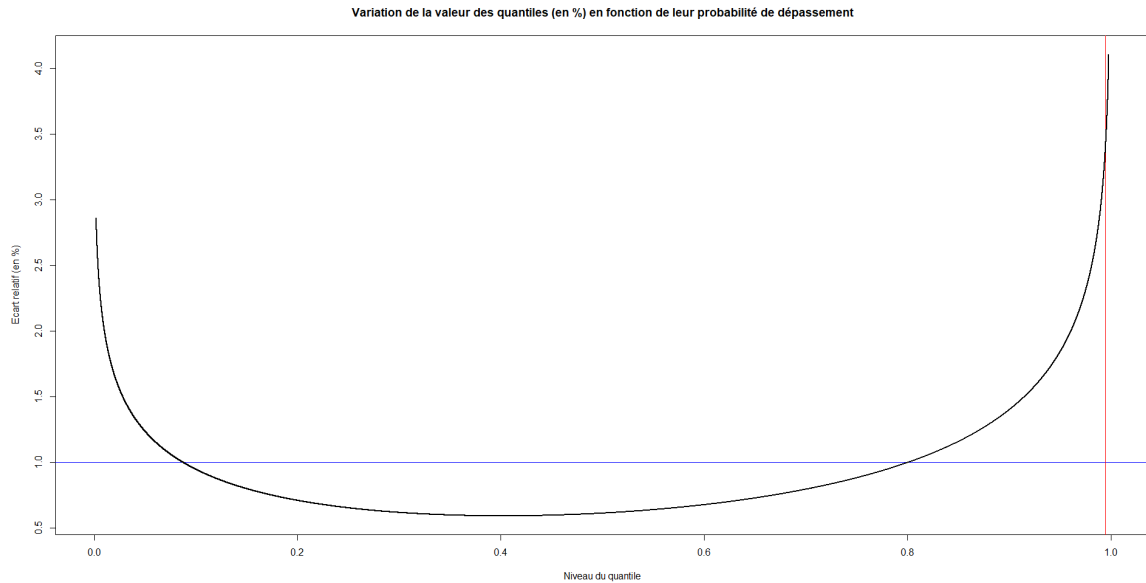


FIGURE 6.12 – Écarts relatifs sur les quantiles entre la méthode paramétrique et la méthode 2. La courbe noire représente, pour un niveau de quantile fixé, l'écart relatif entre le quantile de ce niveau estimé sans incertitude sur paramètres et celui estimé selon la méthode 2.

Le trait bleu représente la barre des 1% d'écart : tous les quantiles dont le niveau est situé approximativement entre 10% et 80% (période de retour entre 1,11 et 5 années) se situent sous ce seuil.

Cependant, l'écart croît plutôt rapidement en dehors de cet intervalle et atteint 3,4% au quantile 99,5%.

Ainsi, si la première méthode présente la limite d'aboutir à une distribution prédictive différente, il s'avère que la seconde méthode permet elle de conserver la loi souhaitée.

De plus, le détail des différentes étapes du calcul d'un quantile sous la méthode basée sur la vraisemblance permet de donner une série d'arguments en faveur de cette méthode. A la fin de l'étape de simulations selon la loi à posteriori, l'actuaire se retrouve avec un échantillon de paramétrages  $(\theta_i)_{1 \leq i \leq N} = (\mu_i; \sigma_i)_{1 \leq i \leq N}$ .

L'objectif est alors d'obtenir l'espérance de la variable  $q_\alpha(\theta)$ , qui admet pour estimateur de Monte-Carlo  $\frac{1}{N} \sum_{i=1}^N q_\alpha(\mu_i, \sigma_i) = \frac{1}{N} \sum_{i=1}^N \exp(\sigma_i * \Phi^{-1}(\alpha) + \mu_i)$ .

Le choix de prendre la moyenne de cette variable peut sembler surprenant : comment être sûr que l'incertitude sur paramètres est correctement préservée en ne conservant

qu'une moyenne qui ne dépend pas de la dispersion des valeurs ?

L'inégalité de Jensen permet d'apporter une réponse à cette interrogation. En effet, la fonction exponentielle étant convexe, l'inégalité s'applique et donne :

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \exp(\sigma_i * \Phi^{-1}(\alpha) + \mu_i) &\geq \exp\left(\frac{1}{N} \sum_{i=1}^N (\sigma_i * \Phi^{-1}(\alpha) + \mu_i)\right) \\ \frac{1}{N} \sum_{i=1}^N \exp(\sigma_i * \Phi^{-1}(\alpha) + \mu_i) &\geq \exp\left(\frac{1}{N} \sum_{i=1}^N \sigma_i * \Phi^{-1}(\alpha) + \frac{1}{N} \sum_{i=1}^N \mu_i\right) \\ \frac{1}{N} \sum_{i=1}^N \exp(\sigma_i * \Phi^{-1}(\alpha) + \mu_i) &\geq \exp(\bar{\sigma} * \Phi^{-1}(\alpha) + \bar{\mu}) \end{aligned}$$

Où  $(\bar{\mu}; \bar{\sigma})$  représente la moyenne empirique de l'échantillon simulé qui, par la loi des grands nombre et le théorème central limite, est proche de l'espérance de la loi a posteriori.

Ainsi, les quantiles obtenus par la deuxième méthode sont toujours plus grands que des quantiles qui seraient obtenus par une estimation ponctuelle. Or, cette estimation ponctuelle est basée sur la moyenne de la loi a posteriori qui est en pratique peu éloignée du mode a posteriori d'après le résultat de convergence mentionné dans la section 4.1.2 ce qui signifie  $(\bar{\mu}; \bar{\sigma}) \approx (\mu^{EMV}; \sigma^{EMV})$ .

L'application de l'inégalité de Jensen permet alors d'avancer trois arguments :

- La prise en compte de l'incertitude sur paramètres ne fait qu'ajouter un "chargement de sécurité" par rapport à des quantiles obtenus par une estimation ponctuelle. Ce chargement est en adéquation avec la méthodologie retenue de mesure d'effet.
- D'après le cas d'égalité de l'inégalité de Jensen, les deux quantiles sont égaux si et seulement si tous les  $\theta_i$  sont égaux, c'est à dire si la loi a posteriori attribue toute sa masse à un unique point, auquel cas il n'y a effectivement aucune incertitude sur la valeur des paramètres.
- L'inégalité est d'autant plus large que les  $\theta_i$  sont dispersés. Ainsi, plus le posterior est diffus (ie. plus l'incertitude sur les paramètres est grande), plus l'évaluation de l'incertitude sur paramètres sera élevée.

En conclusion, la prise en compte de l'incertitude sur paramètres par la seconde méthode se justifie par deux séries d'arguments. La première vient de la construction même de la méthode à partir d'un cadre bayésien : l'évaluation de l'incertitude sur paramètres s'interprète comme le fruit de la décision optimale du point de vue d'une connaissance du phénomène (la loi a posteriori) et de la minimisation d'une fonction de coût rationnelle (minimisation de l'erreur quadratique entre le véritable quantile et le quantile estimé).

La seconde série d'arguments vient de la réinterprétation de l'incertitude sur paramètres comme l'écart dans une inégalité de Jensen portant sur la loi a posteriori. La largeur de cette inégalité permet en effet de quantifier la dispersion de la loi a posteriori et donc l'incertitude sur paramètres.

Une évaluation analytique du gap de Jensen est d'ailleurs possible grâce au développement en série entière de la fonction exponentielle. Pour plus de clarté, le



terme  $\sigma_i * \Phi^{-1}(\alpha) + \mu_i$  sera simplement noté  $x_i$ . Un raisonnement donné en annexe A.4 donne :

$$G = \frac{1}{N} \sum_{i=1}^N \exp(x_i) - \exp(\bar{x})$$

$$G = \frac{\bar{x}^2 - \bar{x}^2}{2} + \frac{\bar{x}^3 - \bar{x}^3}{6} + \sum_{j=4}^{+\infty} \frac{\bar{x}^j - \bar{x}^j}{j!}$$

## 6.3 Modélisation avec a priori informatifs

Cette section sera divisée en deux sous-parties, chacune traitant d'une méthode abordée au chapitre 5.

L'idée est de donner des exemples d'élicitations puis de contrôler les résultats produits par l'inférence.

### 6.3.1 Paire de quantiles

Les quantiles spécifiés dans cette partie sont la médiane et le quantile à 10 ans<sup>1</sup>. Ils correspondent respectivement à une valeur jugée "centrale" et une valeur jugée "extrême".

Avant même d'actualiser la loi a priori et d'en déduire la valeur de l'estimateur de Bayes, l'actuaire doit mener à bien une succession de tâches.

#### 6.3.1.1 Troncature de l'ensemble des paramètres

La première donnée à fixer dans ce travail d'inférence est l'ensemble  $\Theta$  contenant les paramètres. Les prior produits sont basés sur des courbes de niveaux elliptiques, ils admettent donc un mode autour duquel le gros de la masse se concentre. En pratique, la probabilité de s'éloigner significativement de ce mode est suffisamment faible pour être considérée comme nulle et ne pas s'hasarder à tronquer  $\Theta$  de manière subjective. La seule exclusion faite sera relatives aux valeurs de  $\sigma$  négatives, du fait qu'elles ne représentent pas des paramétrages valides pour une loi log-normale. Finalement, l'ensemble des possibles envisagés sera l'ensemble des paramétrages valides pour la loi retenue, à savoir l'ensemble  $\mathbb{R} \times \mathbb{R}_+^*$ .

#### 6.3.1.2 Validation des propriétés du prior

Par la suite, le modélisateur se doit d'étudier les propriétés du prior qu'il a sélectionné. Dans le cas présent, cela va consister dans un premier temps à contrôler la conformité des quantiles prédictifs produits avec le jugement d'expert.

Pour ce faire, deux lois a priori ont été choisies, chacune disposant d'un hyper-paramétrage propre. Pour plus de clarté, ces paramétrages sont restitués de manière synthétique sous la forme suivante :  $\pi_1 = (3; \frac{1}{2}; 1; 0.1)$  signifie que le prior  $\pi_1$  est construit à partir d'une ellipse ayant pour demi-axe 3 et  $\frac{1}{2}$  et d'une autre ellipse

---

1. C'est-à-dire le quantile de niveau 90%.

ayant pour demi-axe 1 et 0,1. Sous cette convention, les deux lois retenues sont respectivement  $\pi_1 = (3; \frac{1}{2}; 1; 0.1)$  et  $\pi_2 = (3; 0.1; 10; 0.1)$

Une étude des distributions prédictives des quantiles 50% et 90% associées à ces deux a priori est effectuée à l'aide de boîtes à moustaches (voir graphique 6.13). En ce qui concerne les résultats, il est pertinent de noter que les distributions des quantiles prédictifs issues des deux a priori présentent des médianes en accord avec les niveaux préalablement spécifiés par l'expert. Cette observation suggère que l'objectif initial d'alignement entre la distribution prédictive et les avis de l'expert est atteint, renforçant ainsi la validité de la méthode de construction proposée pour la prise en compte d'informations externes.

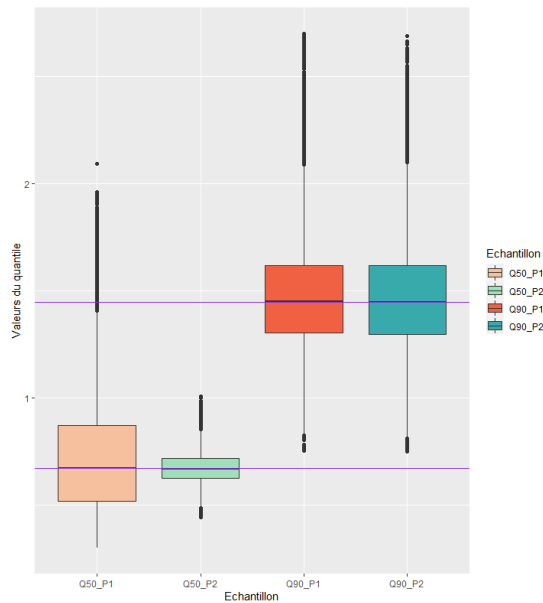


FIGURE 6.13 – Distributions des quantiles d'intérêts représentées sous forme de boîtes à moustache. Les prior considérés sont les suivants  $\pi_1 = (3; \frac{1}{2}; 1; 0.1)$  et  $\pi_2 = (3; 0.1; 10; 0.1)$ . Les barres violettes représentent les niveaux de médiane et de quantile à 10 ans spécifiés par l'expert.

### 6.3.1.3 Sensibilités aux paramètres $\lambda_1$ et $\lambda_2$

Si les statistiques centrales des distributions de quantiles coïncident avec les valeurs spécifiées par l'expert, il reste encore à étudier leur répartition dans sa globalité.

La figure 6.14 représente la distribution des quantiles 50% et 90% pour les deux mêmes prior que précédemment. Il apparaît d'emblée que les égalités de médianes cachent des disparités importantes. Les valeurs de  $\lambda_1$  et de  $\lambda_2$  influent non seulement la dispersion, mais également la morphologie de la distribution des quantiles.

En effet, si le prior  $\pi_2$  aboutit à une distribution relativement symétrique du quantile 50%, le prior  $\pi_1$  donne une distribution du quantile 50% largement décalée à gauche de sa médiane (et donc une queue de distribution étalée vers la droite). De son côté, la distribution du quantile à 10 ans ne varie que très peu entre les

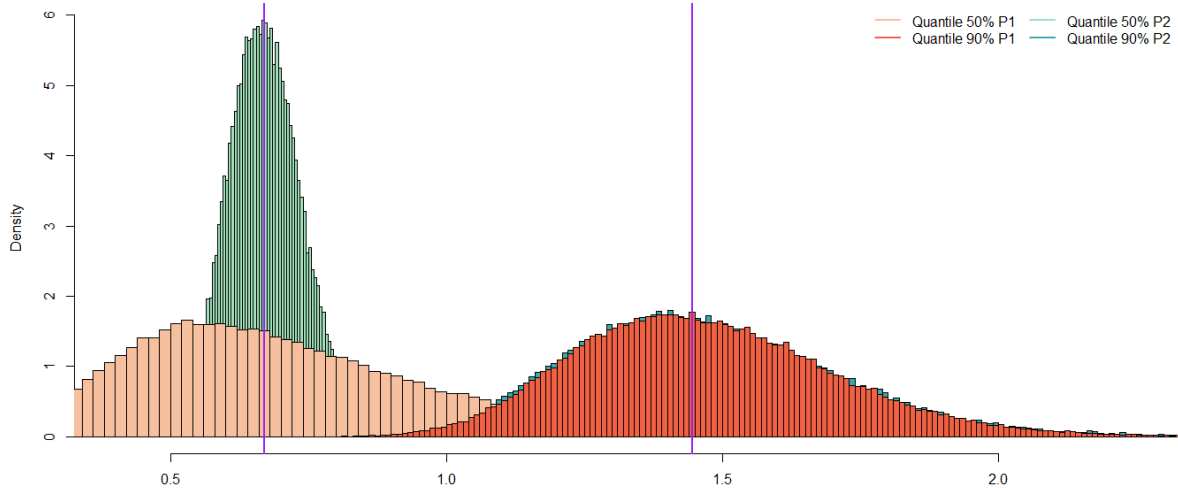


FIGURE 6.14 – Distributions des quantiles d'intérêts représentées sous forme d'histogrammes. Les prior considérés sont les suivants  $\pi_1 = (3; \frac{1}{2}; 1; 0.1)$  et  $\pi_2 = (3; 0.1; 10; 0.1)$ . Les barres violettes représentent les niveaux de médiane et de quantile à 10 ans spécifiés par l'expert.

deux a priori. A première vue, un changement de valeur de la seconde valeur propre (c'est-à-dire celle associée au petit demi-axe) aurait plus d'impact sur l'apparence finale de la distribution.

Cependant, l'impact marginal ou croisé de chacun des  $\lambda$  est difficile à évaluer. Bien qu'a priori, les deux premières valeurs propres concernent l'ellipse associée à la médiane et les deux autres au quantile 90%, il est possible et même probable que chaque valeur influe sur l'ensemble de la distribution, et donc même sur l'autre quantile. Toutefois cet effet n'a pas été observé dans l'exemple précédent : ni le changement du premier groupe de valeurs, ni le changement du second groupe n'ont affecté la distribution du quantile à 10 ans.

Il conviendrait de pousser plus loin l'analyse pour pouvoir appréhender correctement les dépendances en chacune des valeurs propres. Néanmoins, en raison des limitations de la méthodologie proposée, l'étude ne sera pas approfondie davantage. Les principales raisons de cette décision sont les suivantes :

- la méthode est complexe à mettre en œuvre car elle exige d'explicitier les lignes de niveaux ;
- l'impact marginal et croisé des quatre valeurs propres à fixer n'est pas évident à quantifier, ce qui rend leur choix difficile à arrêter ;
- la construction du prior consiste à couvrir les lignes de niveau cibles à l'aide de formes géométriques simples, non seulement la méthode devient vite inutilisable pour des jugements plus complexes mais en plus elle manque de rigueur et d'objectivité ;
- deux jugements sont nécessaires afin de pouvoir fixer un point central aux ellipses ;
- une autre méthodologie, jugée plus fondée, est également proposée.

La présence de cette méthode géométrique dans ce rapport s'explique par le fait qu'elle fut la première méthode envisagée et fait ainsi partie des travaux menés dans le cadre de ce mémoire. Elle a par la suite été supplantée par la seconde méthode qui répond d'avantage à un critère objectif.

### 6.3.2 Méthode générale

Le jugement employé dans cette section concerne un coût de réassurance, le jugement est donc de la forme  $f(\mu, \sigma) = u$  avec  $f$  la fonction de pricing du contrat et  $u$  son coût spécifié par l'expert.

Une courbe entière (voir graphique 6.15) vérifiant exactement la contrainte de l'expert, il devient nécessaire de porter plus d'attention à la troncature de l'espace  $\Theta$ . En effet, la distribution a priori ne se concentre plus exclusivement autour d'un unique mode, mais s'étire le long d'une courbe, englobant par conséquent des régions peu probables du point de vue actuariel, notamment en ce qui concerne la médiane et le coefficient de variation.

**Coût de réassurance et avis de l'expert**

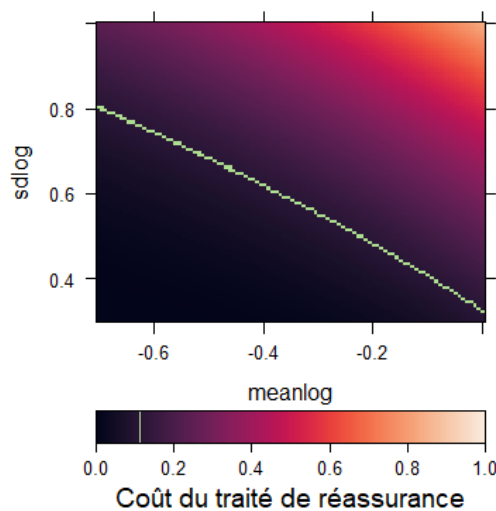


FIGURE 6.15 – Représentation graphique des lignes de niveaux des coûts de réassurances en fonction des paramètres. L'ensemble des paramétrages en accord avec le jugement de l'expert est représenté en vert, il prend la forme du courbe qui traverse un vaste domaine.

#### 6.3.2.1 Troncature de l'espace des paramètres

Les décisions concernant la troncature de l'espace prises dans cette partie vont dans le sens de la prudence en minorant, très grossièrement, la médiane et le coefficient de volatilité de la sinistralité :

- $\mu$  est minoré par  $-1,5$  ce qui équivaut à minorer la médiane par  $\approx 22,3\%$  ;
- $\sigma$  est minoré par  $0,2$  ce qui équivaut à minorer le coefficient de variation par  $\approx 22,1\%$ .

L'approche adoptée en ce qui concerne la réduction de l'espace  $\Theta$  peut sembler légère à une première vue. Cependant, il convient de rappeler que la distribution des paramètres finalement retenue pour les prédictions sera le posterior, résultant de la combinaison du prior et de la vraisemblance. Or, la vraisemblance est elle-même fortement concentrée autour de son mode, ce qui limite très fortement le risque de charger des zones incohérentes d'un point de vue actuariel.

### 6.3.2.2 Validation des propriétés du prior

Sur la base d'un même jugement, deux séries de paramétrages ont été générées pour des valeurs de  $\tau$  différentes.

La figure 6.16 expose la distribution des coûts de réassurance associés à chaque série, c'est-à-dire la distribution empirique des  $(F(\mu_i^j, \sigma_i^j))_i$  où  $(\mu_i^j, \sigma_i^j)_i$  représente la série  $j$  de paramètres simulés.

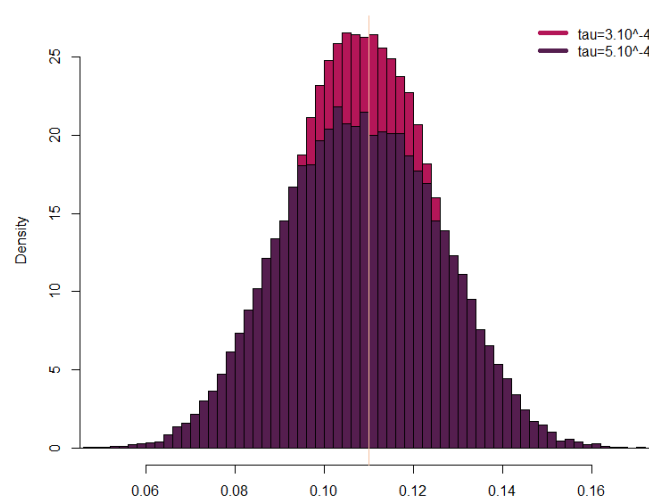


FIGURE 6.16 – Répartition des coûts de la réassurance associés aux paramètres simulés, chaque couple de paramètre  $(\mu, \sigma)$  est utilisé pour déterminer une prime pure. Plus  $\tau$  est petit, donc plus  $\frac{1}{\tau}$  est grand, plus la distribution est piquée. La barre jaune représente le coût de réassurance proposé par l'expert.

Il apparaît que les distributions de coût de réassurance sont à la fois unimodales et symétriques, mais également légèrement biaisées. Néanmoins, ce biais est léger et tend à se résorber lorsque la valeur du paramètre  $\tau$  diminue (voir sous-section suivante).

Bien que les distributions présentées sur la figure 6.16 ressemblent à des lois normales, un calcul de kurtosis (kurtosis normalisé de  $-0,3$ ) ainsi qu'un test de Shapiro-Wilk ( $p_{value} < 10^{-3}$ ) permettent de rejeter cette hypothèse. En conclusion, l'objectif d'obtenir un a priori chargeant prioritairement des paramétrages en accord avec le jugement de l'expert est atteint. En effet, la distribution des paramètres correspond bien à une distribution des coûts de réassurance globalement centrée autour de la valeur cible.

### 6.3.2.3 Sensibilité à l'hyperparamètre $\tau$

Le paramètre  $\tau$  permet de quantifier l'incertitude portant sur le jugement de l'expert. Son influence se déploie à divers niveaux au sein du processus d'inférence et il convient ici de distinguer ses répercussions sur le prior des effets sur le posterior.

#### Effets sur la loi a priori

Dans un premier temps, concernant les impacts sur la loi a priori, le biais mis en lumière précédemment tend à s'alléger avec la valeur de  $\frac{1}{\tau}$  (figure 6.17). En outre,  $\tau$  a également une influence sur la variance des coûts de réassurances prédictifs : plus  $\frac{1}{\tau}$  est grand et moins les coûts de réassurances seront dispersés (figure 6.16).

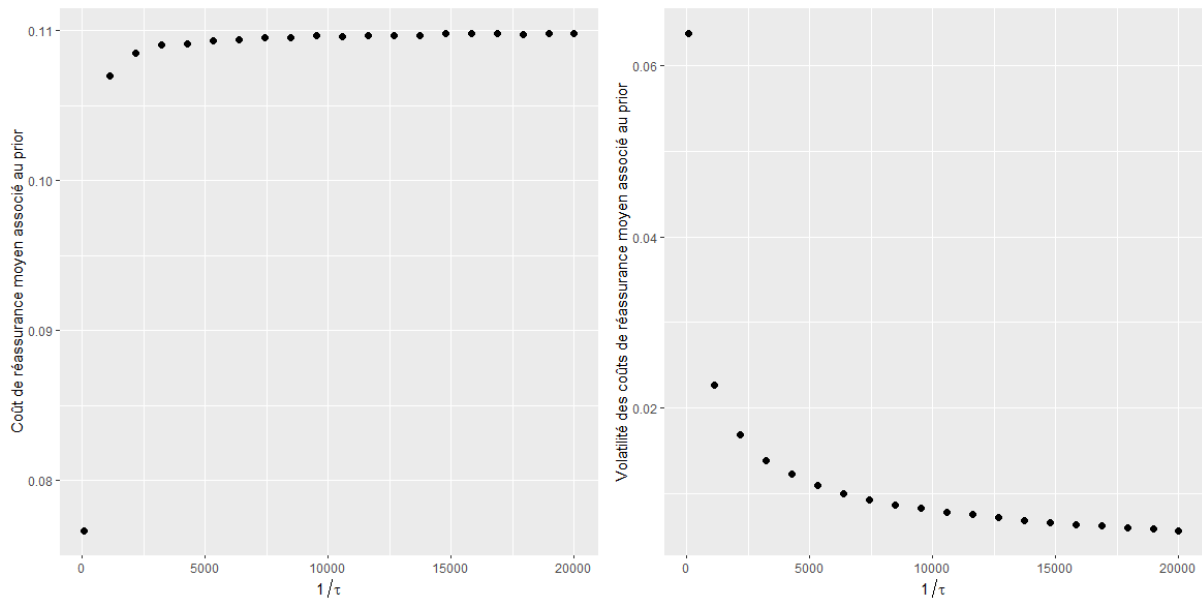


FIGURE 6.17 – Moyenne et volatilité des coûts de réassurance en fonction de  $\frac{1}{\tau}$ . La moyenne converge par le bas vers 0,11, la valeur spécifiée par l'expert, tandis que la volatilité décroît avec  $\frac{1}{\tau}$ .

Deux points sont donc à retenir du côté de l'a priori :

- il existe un léger biais entre le niveau fourni par l'expert et la moyenne du coût de réassurance issu de la loi a priori mais ce biais est faible, de l'ordre de quelques points, et il diminue au fur et à mesure que  $\tau$  diminue ;
- la volatilité des coûts de réassurance associés aux paramétrages décroît quand  $\tau$  diminue.

#### Effets sur les distributions des quantiles prédictifs

Dans un second temps, il faut s'intéresser, non pas à la distribution des coûts de réassurance, mais aux quantiles prédictifs qui constituent les véritables objectifs de l'inférence.

Pour ce faire, les distributions prédictives de quantiles de plusieurs niveaux, à savoir 10%, 30%, 50%, 70% et 90%, ont été tracées pour plusieurs valeurs de  $\tau$ .

De même que pour les coûts de réassurance, le niveau moyen des quantiles ne varie que très légèrement, c'est essentiellement la dispersion autour de ces niveaux qui change. Ainsi, le paramètre d'incertitude du prior, qui se traduit par un paramètre

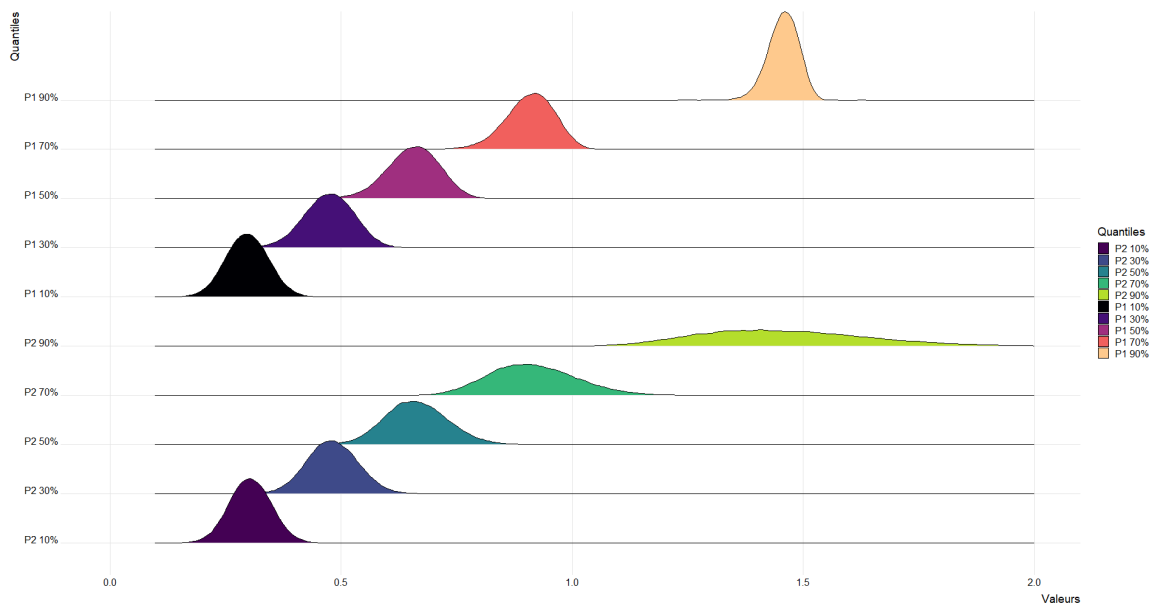


FIGURE 6.18 – P1 correspond à  $\tau$  valant  $5.10^{-5}$  et P2 à un  $\tau$  de  $5.10^{-2}$ . La position des différents quantiles est globalement la même, toutefois les quantiles du second prior sont nettement plus étalés.

de dispersion sur les coûts de réassurance, joue également un rôle de paramètre de dispersion sur les valeurs des quantiles prédictifs. Ce résultat, qui n'avait rien d'évident en soi, permet utilement de fixer l'incertitude sur la valeur du jugement d'expert en même temps que l'incertitude sur les distributions a priori des quantiles. Ainsi, un jugement peu fiable ne figurera pas les valeurs des quantiles prédictifs et laissera donc les données s'exprimer, tout en orientant subtilement l'inférence.

### Effets sur le posterior

Si les effets du paramètre  $\tau$  sur le prior ont été traités, il reste encore à étudier l'influence qu'il a sur le posterior. Comme attendu, plus sa valeur est petite, ie. plus  $\frac{1}{\tau}$  est grand, plus la forme du posterior est déterminée par celle du prior (figure 6.19) et donc moins le résultat de l'inférence dépendra des données.

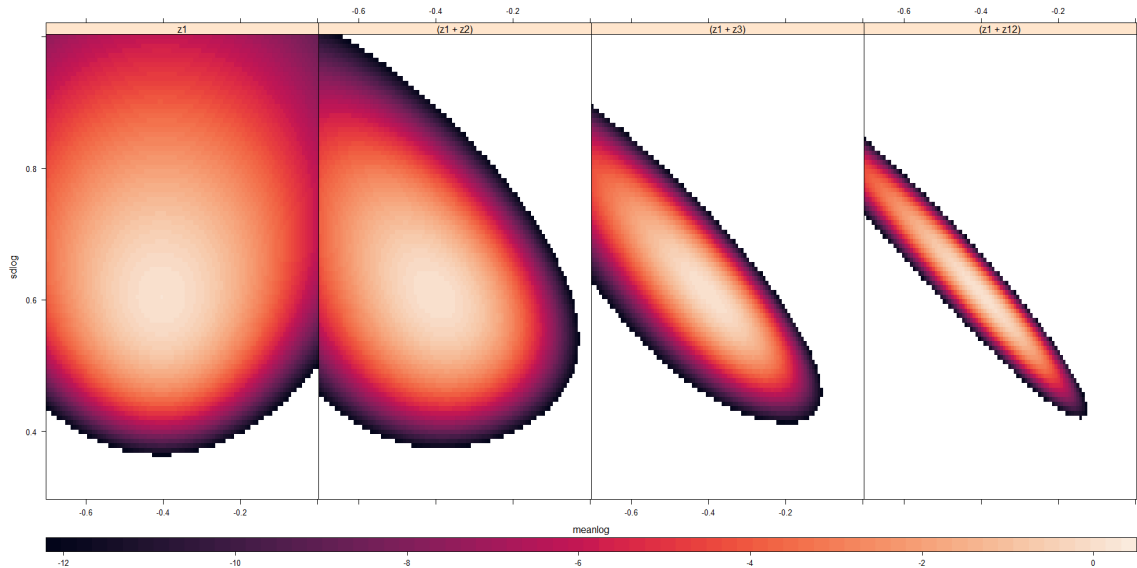


FIGURE 6.19 – Forme du posterior en fonction du paramètres  $\tau$ . Le graphique de gauche représente la vraisemblance de l'échantillon étudié, les trois suivants sont des posterior avec, de gauche à droite,  $\tau^2 = 1/200$ ,  $\tau^2 = 1/2000$ ,  $\tau^2 = 1/10000$ . Plus  $\tau$  diminue, plus le posterior se rapproche du prior.

### Effets sur l'algorithme de Metropolis Hastings

Une autre manière d'appréhender l'influence du paramètre  $\tau$  est de s'intéresser à la valeur du ratio d'acceptation d'un saut dans l'algorithme de Metropolis-Hastings. Cette information est donnée en figure 6.20. Les taux d'acceptation sont pris pour un saut partant d'un point  $(\mu_1, \sigma_1)$  vers un point  $(\mu_2, \sigma_2)$  dont l'écart relatif entre  $f(\mu_2, \sigma_2)$  et la condition de l'expert est donné en abscisse. L'écart relatif entre  $f(\mu_2, \sigma_2)$  est la valeur de l'expert, renseignée par les pointillés noirs.

Ainsi, le graphique du haut suppose que la chaîne est actuellement sur un point vérifiant exactement la condition de l'expert. La probabilité que la chaîne accepte un saut vers un point  $(\mu_2, \sigma_2)$ , tel que  $f(\mu_2, \sigma_2) = K \pm x\%K$ , est donnée en fonction de  $x$  et pour plusieurs valeurs de  $\tau$ .

Le graphique du bas suppose que la chaîne est actuellement sur un point  $(\mu_1, \sigma_1)$  tel que  $|\frac{f(\mu_1, \sigma_1) - K}{K}| = 10\%$ . Pour rappel, la probabilité qu'un saut vers un point vérifiant mieux la contrainte (donc dont l'écart à la condition est inférieur à 10%) soit acceptée est toujours de 1, peu importe la valeur de  $\tau$ . En effet, les probabilités d'acceptation sont les ratios de la densité du prior entre le point d'arrivée et de départ (le ratio étant majoré par 1).



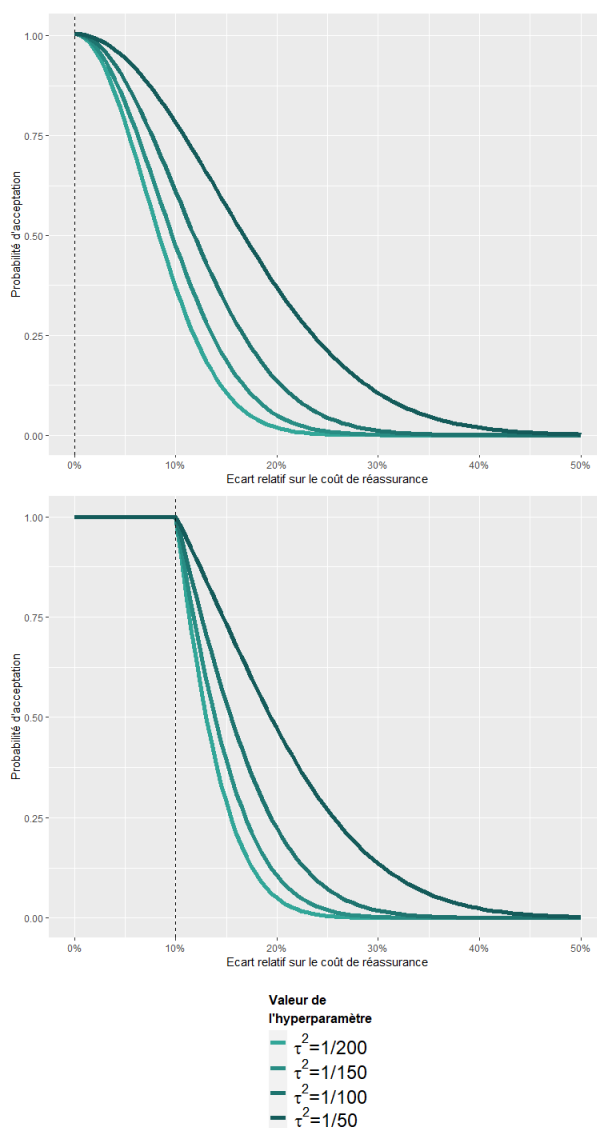


FIGURE 6.20 – Taux d’acceptation de l’algorithme de Metropolis-Hasting en fonction du paramètre  $\tau$ .

Le tableau 6.2 recense des valeurs de ratio d’acceptation pour plusieurs sauts fixés.

La décroissance, pour un saut fixé, du ratio d’acceptation en fonction de  $\tau$  est une autre manière de quantifier et d’appréhender le fait que plus  $\tau$  est petit, plus l’a priori est piqué.

Valeur de $\tau^2$	1/50	1/100	1/150	1/200
Probabilité d’un saut 0% $\rightarrow$ $\pm 10\%$	77,88%	60,65%	47,24%	36,79%
Probabilité d’un saut $\pm 10\% \rightarrow \pm 20\%$	47,23%	22,31%	10,54%	4,98%

TABLE 6.2 – Probabilité d’acceptation d’un saut en fonction de  $\tau$

Finalement, il ne reste plus qu’à se pencher sur les quantiles prédictifs a posteriori obtenus avec cette méthode.

### Effets sur la distribution prédictive a posteriori

Pour les paramètres retenus, les lois prédictives basées sur des prior informatifs sont relativement proches de la loi prédictive basée sur un prior non informatif

(courbe orange), l'ajout d'information a priori n'a pas changé substantiellement la distribution. En d'autres termes, l'information issue des données conserve son rôle prépondérant dans la formation du posterior, la connaissance a priori n'infléchit que légèrement le résultat. Même s'il est théoriquement possible de modifier le poids de l'information a priori, il paraît raisonnable de laisser la plus belle part aux données, à plus forte raison lorsque l'historique comptabilise plus d'une vingtaine de points. Par ailleurs, l'ajout d'information n'a pas déformé la morphologie de la distribution, l'hypothèse de log-normalité reste respectée (figure 6.21).

Néanmoins, l'ajout d'information n'a pas été complètement sans effet. D'une part, les posterior informatifs sont légèrement plus piqués que leur contrepartie non informative : la mutualisation des sources d'information a donc permis d'obtenir une inférence plus fine de la sinistralité en réduisant la volatilité de la distribution prédictive. D'autre part, si les distributions ne sont que faiblement translatées vers la distribution prédictive a priori (courbe bleue), les coûts de réassurance prédictifs se sont eux rapprochés du niveau spécifié par l'expert (voir tableau 6.3).

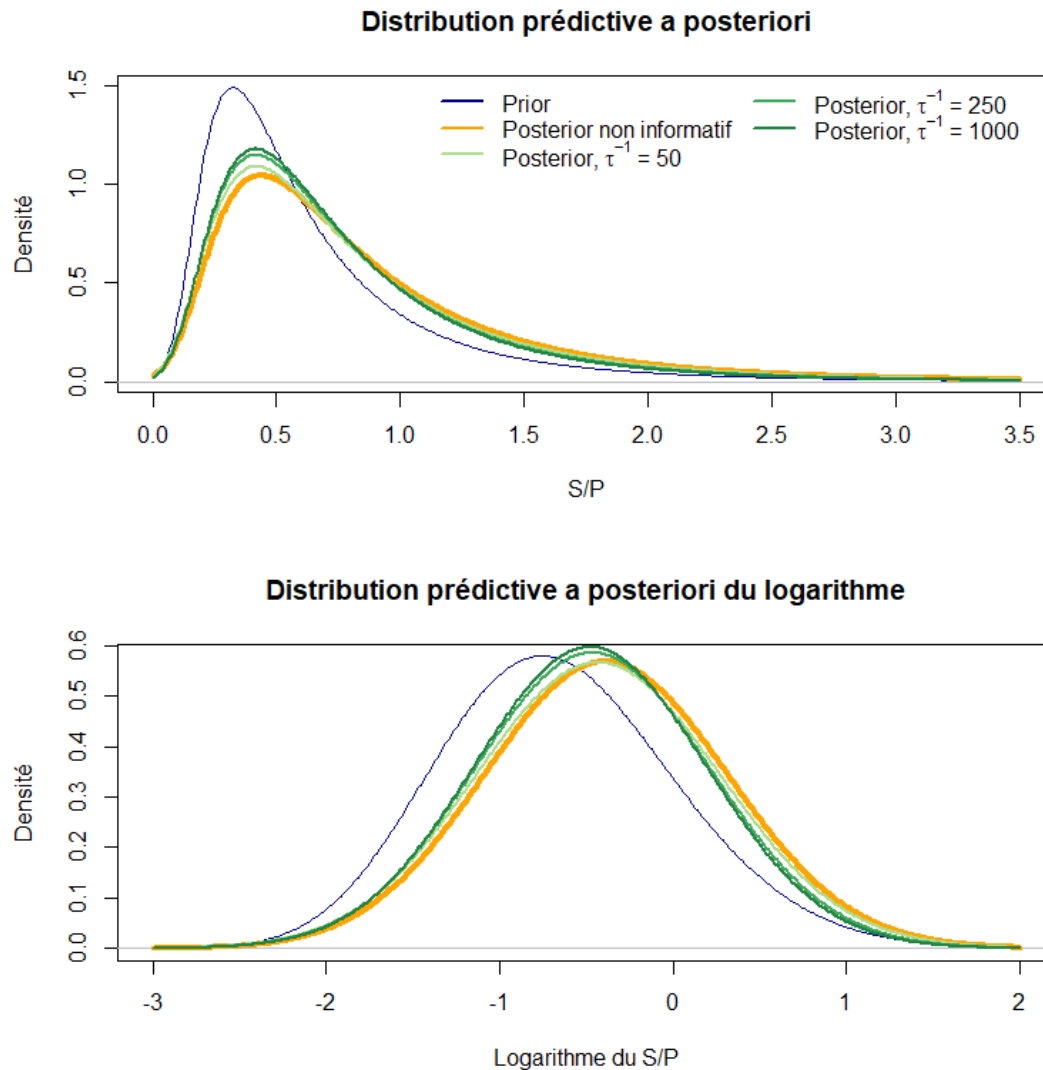


FIGURE 6.21 – Distributions prédictives a posteriori en fonction du paramètre  $\tau$ . La courbe bleue correspond au prior prédictif pour une valeur spécifique de  $\tau$ .

Prior	Coût de réassurance
Prior non informatif	17,1%
Prior informatif $\tau^{-1} = 50$	15,3%
Prior informatif $\tau^{-1} = 250$	12,6%
Prior informatif $\tau^{-1} = 1000$	11,4%

TABLE 6.3 – Coût prédictif a posteriori de réassurance pour différents prior. Les coûts prédictifs de réassurance déterminés à partir d'a priori informatifs se rapprochent des 11% spécifiés par l'expert.

### 6.3.2.4 Conclusion

Ainsi, la méthode proposée combine une série d'avantages :

- Elle est rationnelle d'un point de vue probabiliste : si la seule information donnée par l'expert concerne la valeur d'une certaine caractéristique de la sinistralité, alors il est souhaitable que deux paramétrages qui aboutissent à la même caractéristique se voient attribuer la même densité. Cela revient à imposer que les lignes de niveaux de la fonction permettant de déterminer la valeur de la caractéristique à partir des paramètres soient les mêmes que les lignes de niveaux de la loi a priori.
- La distribution a priori des paramètres ainsi construites correspond bien à une distribution prédictive a priori de la caractéristique centrée sur la valeur donnée par l'expert.
- La présence d'un paramètre d'incertitude permet d'adapter la forme du prior au niveau de confiance en l'expert. Il contrôle notamment la dispersion du prior et permet de charger des paramétrages plus ou moins en adéquation avec l'avis de l'expert.
- Ce paramètre d'incertitude contrôle également le poids que prendra la connaissance a priori dans la formation du posterior.
- Elle permet d'aboutir à des distributions a posteriori proches de la vraisemblance mais dont la valeur de la caractéristique se rapproche aussi du jugement de l'expert. Ainsi, l'information contenue dans les données est mise en adéquation avec l'information extérieure apportée par le prior.

Toutefois, il reste des axes d'améliorations notables, notamment en ce qui concerne les choix de la fonction décroissante et de la distance qui étaient, dans ce mémoire, arbitraires. Pour rappel, la forme générale du prior est  $\pi(\theta) = \phi(d(f(\theta), u))$  avec  $\phi$  une fonction décroissante et  $d$  une distance. Dans l'exemple de ce mémoire, la fonction et la distance retenues étaient, par analogie avec une loi normale, respectivement  $\phi(\cdot) = \exp(-\frac{\cdot}{\tau})$  et  $d(x, y) = (x - y)^2$ , cette sélection a été arbitraire et aurait pu être différente. Il pourrait être intéressant d'envisager l'usage de fonctions et de distances alternatives, et plus particulièrement de baser leur choix sur des critères objectifs, tels que la minimisation d'un certain coût ou la satisfaction de contraintes spécifiques.

De plus, il est important de garder à l'esprit que les calculs des provisions et des indicateurs de solvabilité, notamment le SCR, sous la réforme Solvabilité II,

sont soumis à un cadre réglementaire strict. Outre les difficultés d'un point de vue statistique du modélisateur liées à la recherche d'informations pertinentes, qu'elles viennent d'étude internes ou externes, à la reformulation de ces informations sous la forme d'un jugement d'expert compatible avec la méthode proposée et au choix des fonctions constitutives du prior et de son paramètre d'incertitude, l'actuaire doit aussi prendre en compte les justifications à apporter d'un point de vue réglementaire et prudentiel.

L'utilisation d'hypothèses sous forme de jugement d'expert en complément des données traditionnelles ne doit pas se traduire par une perte de pertinence et de fiabilité des résultats. L'ajout d'une loi a priori peut peser lourdement sur l'inférence et doit répondre aux mêmes exigences d'adéquation, de pertinence et d'exhaustivité que les historiques classiques. Une gouvernance claire doit être mise en place concernant les jugements d'experts. Cette dernière doit notamment permettre de garantir la traçabilité des études, la possibilité de pérenniser l'usage de données externes et de formaliser la validation des hypothèses, des données, et de la bonne tenue de l'étude à partir de laquelle a été formé le jugement d'expert.

# Conclusion

Les objectifs de ce mémoire étaient doubles : d'une part, élaborer une méthodologie pour quantifier les risques associés à l'incertitude sur les paramètres, et d'autre part, proposer une approche pour intégrer des données externes à des études actuarielles en complément des historiques traditionnels. La réunion de ces deux sujets au sein d'un même mémoire trouve son fondement dans les avantages inhérents au paradigme bayésien par rapport à l'approche classique des probabilités.

Après s'être intéressés à la présence et à la portée de l'incertitude sur paramètres dans les processus d'inférence classiques utilisés par Groupama, l'attention s'est tournée sur les manières d'évaluer cette incertitude.

Dans un premier temps, des limites de la méthode employée par Groupama ont été exposées. Dans un second temps, deux méthodes issues de la statistique bayésienne et usant de prior non informatifs ont été proposées en lien avec la bibliographie. Tout l'enjeu aura été ensuite de réussir à formuler adéquatement le problème bayésien, de sorte à respecter les hypothèses de modélisation. Il a été montré qu'une de ces méthodes permettait de respecter la vraisemblance supposée par l'actuaire tout en justifiant l'évaluation de l'incertitude sur paramètres à partir d'arguments issus de la statistique bayésienne. Une réinterprétation du résultat sous la forme d'un écart dans une inégalité de Jensen a permis de dégager et de valider des propriétés intéressantes, à savoir :

- la prise en compte du risque sur paramètre ne peut que faire augmenter l'appréciation du risque
- La matérialité est d'autant plus grande que le posterior est diffus, c'est-à-dire que les incertitudes sur la valeurs des paramètres est grande ;
- l'incertitude est considérée comme nulle uniquement si le posterior est concentré sur un point, ie. qu'il n'y à aucune incertitude sur la valeur des paramètres ;
- l'incertitude sur paramètres est d'autant plus grande que la fonction quantile de vraisemblance est courbée, ie. que les incertitudes de calibrages ne se compensent pas entre elles ;
- la distribution prédictive est similaire à la vraisemblance.

Finalement, une application sur les données de sinistralités attritionnelles de Groupama a permis de conclure que ce risque était négligeable dans le calcul du SCR du Groupe.

La seconde partie de ce mémoire traitait de l'intégration de données externes, externes signifiant extérieures aux historiques traditionnels de sinistralité. Initialement, la question a été de connaître la nature des informations à prendre en considération. Sur ce point, les résultats d'études menées par d'autres directions représentent un

intérêt tout particulier car ils portent sur les mêmes portefeuilles et les mêmes garanties. Cependant, il convient de garder une certaine mesure et de rester critique vis à vis de ces études comme dans le cas des prévisions comptables qui ne reflètent pas toujours pleinement la réalité.

Par la suite, l'étude s'est recentrée sur des aspects plus techniques en explorant la manière dont le paradigme bayésien pouvait être exploité pour intégrer ces données externes. L'hypothèse de base était que ces informations pouvaient être formulées sous la forme d'un jugement d'expert chiffré, à savoir : "Une caractéristique de la sinistralité vaut une valeur  $u$ .", puis de le traduire sous forme mathématique faisant appel à une fonction  $f$  portant sur les paramètres. Deux méthodes ont été proposées, chacune sous-tendue par la volonté de charger préférentiellement les lignes de niveaux de la fonction  $f$  associées à la valeur  $u$  donnée par l'expert.

La première méthode, consistant à couvrir les lignes de niveaux à l'aide de distributions usuelles, a rapidement été abandonnée au profit de la seconde du fait de la complexité de sa mise en place, des nombreux paramètres faiblement lisibles à fixer, et de son côté approximatif. La seconde méthode a été, dans un premier temps, proposée sous une forme assez générale. Elle jouissait d'un caractère plus "exact" dans le sens où ses lignes de niveaux sont exactement les mêmes que celles de la fonction  $f$  ce qui permet de justifier son usage par un raisonnement logique. De plus, son implémentation est nettement plus simple et elle ne dispose que d'un unique paramètre à fixer, qui est d'ailleurs bien plus lisible que ceux de la première méthode.

Par la suite, un exemple d'inférence a été proposé afin de vérifier les propriétés du prior. Il a été établi que la méthode de construction vérifiait plusieurs propriétés intéressantes et qu'elle permettait notamment d'aboutir à des distributions prédictives a posteriori cohérentes avec l'information contenue dans les données tout en mettant en adéquation ces distributions avec le jugement de l'expert.

Toutefois, la démarche présente quelques limites. Les choix de certaines quantités ont été guidé par une analogie avec la loi normale plutôt que par un critère purement rationnel. De plus, l'intégration de données externes s'assimile à une réduction du poids de l'historique dans l'inférence au profit d'hypothèses et peut donc être difficilement accepté par le superviseur dans un calcul. Dans ce contexte, une gouvernance à la fois transparente, responsable et rigoureuse doit être mise en place pour garantir la qualité des informations externes à inclure dans les modélisations.

Malgré ces limites, la méthodologie proposée pourrait trouver son utilité dans un cadre de souscription, où il est courant que l'actuaire dispose à la fois des résultats du modèle tarifaire de son entreprise et de données de sinistralités historiques du client, ou encore dans la modélisation des risques émergent, tels que le risque cyber ou climatique, pour lesquels les historiques traditionnels ne sont pas suffisants.

# Appendices

# Annexe A

## Annexe calculatoire

### A.1 Calcul de la fonction de coût

Soit :

$$L : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_+$$
$$(y, x) \mapsto \begin{cases} c_1(y - x) & \text{si } y > x \\ c_2(x - y) & \text{si } y \leq x \end{cases}$$

L'objectif est de trouver :

$$\delta^* = \operatorname{argmin}_{\delta \in \mathcal{X}} \mathbb{E}_X[L(\delta, X)]$$

La première étape est de poser  $\Psi : \delta \rightarrow \mathbb{E}_X[L(\delta, X)]$ , puis de développer l'espérance.

$$\begin{aligned} \Psi(\delta) &= \mathbb{E}_X[L(\delta, X)] \\ &= \int_{-\infty}^{+\infty} c_1(x - \delta)f(x)dx \cdot \mathbf{1}_{x > \delta} + \int_{-\infty}^{+\infty} c_2(\delta - x)f(x)dx \cdot \mathbf{1}_{x \leq \delta} \\ &= (c_1 \int_{\delta}^{+\infty} (x - \delta)f(x)dx - c_2 \int_{-\infty}^{\delta} (x - \delta)f(x)dx) \\ &= c_1 \int_{\delta}^{+\infty} xf(x)dx - \delta c_1 \int_{\delta}^{+\infty} f(x)dx - c_2 \int_{-\infty}^{\delta} xf(x)dx + \delta c_2 \int_{-\infty}^{\delta} f(x)dx \\ \Psi(\delta) &= \delta [c_1(1 - F(\delta)) - c_2F(\delta)] + c_2 [H(x)]_{\delta}^{+\infty} - c_1 [H(x)]_{-\infty}^{\delta} \end{aligned}$$

Où la fonction  $H$  est une primitive de la fonction  $x \rightarrow xf(x)$  et  $F$  la fonction de répartition de la variable  $X$ . L'optimum de la fonction  $\Psi$  va être trouvé en annulant sa dérivée :

$$\begin{aligned} \frac{\partial \Psi(\delta)}{\partial \delta} &= [c_1(1 - F(\delta)) - c_2F(\delta)] + \delta [c_1f(\delta) + c_2(f(\delta))] - c_2\delta f(\delta) - c_1\delta f(\delta) \\ \frac{\partial \Psi(\delta)}{\partial \delta} &= c_1(1 - F(\delta)) - c_2F(\delta) \end{aligned}$$



D'où

$$\begin{aligned}\frac{\partial \Psi(\delta^*)}{\partial \delta} &= 0 \\ \implies F(\delta^*) &= \frac{c_1}{c_1 + c_2} \\ \implies \delta^* &= F^{-1}\left(\frac{c_1}{c_1 + c_2}\right)\end{aligned}$$

Ainsi, l'optimum est atteint en  $\delta^*$ , le fractile  $\frac{c_2}{c_1 + c_2}$  de la loi de  $X$ .

## A.2 Calcul du gain de diversification

### A.2.1 Sans incertitude sur paramètres

Le moment d'ordre 2 de la variable  $S = \sum_{i=1}^N X_i$  s'obtient de la manière suivante :

$$\begin{aligned}\mathbb{E}[S^2] &= \mathbb{E}\left[\left(\sum_{i=1}^N X_i\right)^2\right] \\ &= \mathbb{E}\left[\sum_{i=1}^N X_i^2 + \sum_{1 \leq i < j \leq N} 2X_i X_j\right] \\ &= \mathbb{E}\left[\sum_{i=1}^N X_i^2\right] + 2\mathbb{E}\left[\sum_{1 \leq i < j \leq N} X_i X_j\right] \\ &= \mathbb{E}[N]\mathbb{E}[X^2] + 2\mathbb{E}\left[\sum_{k=1}^{\frac{N(N-1)}{2}} Z_k\right] \\ &= \mathbb{E}[N]\mathbb{E}[X^2] + \mathbb{E}[N^2]\mathbb{E}[X]^2 - \mathbb{E}[N]\mathbb{E}[X]^2 \\ \mathbb{E}[S^2] &= \mathbb{E}[N]Var[X] + \mathbb{E}[N^2]\mathbb{E}[X]^2\end{aligned}$$

Où la variable  $Z_k$  est définie par  $Z_k = X_{g_1(k)}X_{g_2(k)}$ , les fonctions  $g_1$  et  $g_2$  servant à passer d'une énumération à deux indice sur le triangle  $1 \leq i < j \leq N$  à une énumération à un indice sur  $1 \leq k \leq \frac{N(N-1)}{2}$ .

Un exemple serait de considérer la fonction  $\psi(k) = \frac{k(k+1)}{2}$  et  $\psi^{-1}(y) = \frac{-1 + \sqrt{1+8y}}{2}$  puis de prendre  $g_1(k) = \lceil \psi^{-1}(k) \rceil - \psi(\lceil \psi^{-1}(k) \rceil) + k$  et  $g_2(k) = \psi(\lceil \psi^{-1}(k) \rceil) - k + 1$ .

En soustrayant à l'expression de  $\mathbb{E}[S^2]$  obtenue précédemment un terme  $\mathbb{E}[S]^2$ , la variance a alors pour expression :

$$Var[S] = \mathbb{E}[N]Var[X] + \mathbb{E}[N^2]\mathbb{E}[X]^2 - \mathbb{E}[N]^2\mathbb{E}[X]^2 = \mathbb{E}[N]Var[X] + \mathbb{E}[X]^2Var[N]$$

### A.2.2 Avec incertitude sur paramètres

Le calcul de l'expression du moment d'ordre 2 de la sinistralité totale reste valable d'où :

$$\mathbb{E}[S'^2] = \mathbb{E}[N']Var[X] + \mathbb{E}[N'^2]\mathbb{E}[X]^2$$

Il reste à obtenir le moment d'ordre 2 de la variable  $N'$  à partir de son expression. Il peut d'ailleurs être utile de justifier la forme de  $N'$ . Pour rappel, les deux seules conditions posées sur  $N$  sont :

$$\begin{cases} \mathbb{E}[N] \propto N_{Ass} \\ \frac{\text{Var}[N]}{\mathbb{E}[N]} \propto 1 \end{cases}$$

En notant  $f$  le facteur de proportionnalité de la première condition, il est possible de réécrire la variable  $N$  comme  $N = W + N_{Ass}f$  avec  $W$  une variable centrée. La condition sur la variance s'applique également à  $W$  en tant que translation de  $N$ . En notant  $\eta$  le coefficient de proportionnalité il alors est possible de réécrire  $W$  comme  $W = \sqrt{N_{Ass}}\eta\epsilon$  où  $\epsilon$  est une variable centrée réduite. Ainsi,  $N$  peut bien se mettre sous la forme  $N = N_{Ass}f + \sqrt{N_{Ass}}\eta\epsilon$ .

Il est maintenant temps d'en déduire son moment d'ordre 2 :

$$\begin{aligned} \mathbb{E}[N'^2] &= \mathbb{E}[(N_{Ass}\hat{f} + \sqrt{N_{Ass}}\hat{\eta}\epsilon)^2] \\ &= \mathbb{E}[N_{Ass}^2\hat{f}^2] + \mathbb{E}[2N_{Ass}\hat{f}\sqrt{N_{Ass}}\hat{\eta}\epsilon] + \mathbb{E}[N_{Ass}\hat{\eta}^2\epsilon^2] \\ &= N_{Ass}^2 (\text{Var}[\hat{f}] + \mathbb{E}[\hat{f}^2]) + 2N_{Ass}\sqrt{N_{Ass}}\mathbb{E}[\hat{f}\hat{\eta}]\mathbb{E}[\epsilon] + N_{Ass}\mathbb{E}[\hat{\eta}^2]\mathbb{E}[\epsilon^2] \\ &= N_{Ass}^2 \text{Var}[\hat{f}] + (N_{Ass}f)^2 + N_{Ass} (\text{Var}[\hat{\eta}] + \mathbb{E}[\hat{\eta}^2]) \\ &= N_{Ass}^2 \text{Var}[\hat{f}] + N_{Ass} \text{Var}[\hat{\eta}] + (N_{Ass}f)^2 + N_{Ass}\eta^2 \\ \mathbb{E}[N'^2] &= N_{Ass}^2 \text{Var}[\hat{f}] + N_{Ass} \text{Var}[\hat{\eta}] + \mathbb{E}[N]^2 + \text{Var}[N] \end{aligned}$$

De cette expression découle le moment d'ordre 2 de la variable  $S'$ . En lui soustrayant  $\mathbb{E}[S']^2 = \mathbb{E}[N]^2\mathbb{E}[X]^2$  il vient :

$$\text{Var}[S'] = \mathbb{E}[N]\text{Var}[X] + \mathbb{E}[X]^2 (\text{Var}[N] + N_{Ass}^2 \text{Var}[\hat{f}] + N_{Ass}\text{Var}[\hat{\eta}])$$

### A.3 Équations des iso-quantile pour une loi log-normale

Soit  $X \sim \mathcal{LN}(\mu, \sigma^2)$ , une première étape est d'obtenir l'expression de son quantile  $q_\alpha$  de niveau  $\alpha$ . En remarquant que la variable  $\log(X)$  suit une loi normale de paramètres  $\mu$  et  $\sigma^2$ , il vient :

$$\begin{aligned} \alpha &= \mathbb{P}(X < q_\alpha) \\ \implies \alpha &= \mathbb{P}(\log(X) < \log(q_\alpha)) \\ \implies \alpha &= \mathbb{P}\left(\frac{\log(X) - \mu}{\sigma} < \frac{\log(q_\alpha) - \mu}{\sigma}\right) \\ \implies \Phi^{-1}(\alpha) &= \frac{\log(q_\alpha) - \mu}{\sigma} \\ \implies \log(q_\alpha) &= \sigma\Phi^{-1}(\alpha) + \mu \end{aligned}$$

Ainsi, tous les couples  $(\mu, \sigma)$  situés sur la droite d'équation  $\mu = \log(q_\alpha) - \sigma\Phi^{-1}(\alpha)$  sont associés à des distributions ayant  $q_\alpha$  pour quantile de niveau  $\alpha$ .

## A.4 Calcul du gap de Jensen

Les notations du corps du documents vont être reprises, le terme  $\sigma_i * \Phi^{-1}(\alpha) + \mu_i$  sera simplement noté  $x_i$ .

En outre, l'opérateur  $\frac{1}{N} \sum_{i=1}^N \cdot$  est noté  $\bar{\cdot}$ . Ainsi, le gap, noté  $G$ , a pour expression  $G = \frac{1}{N} \sum_{i=1}^N \exp(x_i) - \exp(\bar{x}) = \exp(\bar{x})(\exp(\bar{x}^{-1}x) - \exp(\bar{x}))$ . Il vient alors :

$$\begin{aligned}
 G &= \frac{1}{N} \sum_{i=1}^N \exp(x_i) - \exp(\bar{x}) \\
 G &= \frac{1}{N} \sum_{i=1}^N \sum_{j=0}^{+\infty} \frac{x_i^j}{j!} - \sum_{j=0}^{+\infty} \frac{\bar{x}^j}{j!} \\
 G &= \sum_{j=0}^{+\infty} \frac{1}{N} \sum_{i=1}^N \frac{x_i^j}{j!} - \sum_{j=0}^{+\infty} \frac{\bar{x}^j}{j!} \\
 G &= \sum_{j=0}^{+\infty} \frac{\bar{x}^j}{j!} - \sum_{j=0}^{+\infty} \frac{\bar{x}^j}{j!} \\
 G &= \frac{\bar{x}^0 - \bar{x}^0}{0!} + \frac{\bar{x}^1 - \bar{x}^1}{1!} + \frac{\bar{x}^2 - \bar{x}^2}{2!} + \frac{\bar{x}^3 - \bar{x}^3}{3!} + \sum_{j=4}^{+\infty} \frac{\bar{x}^j - \bar{x}^j}{j!} \\
 G &= \frac{\bar{x}^2 - \bar{x}^2}{2} + \frac{\bar{x}^3 - \bar{x}^3}{6} + \sum_{j=4}^{+\infty} \frac{\bar{x}^j - \bar{x}^j}{j!}
 \end{aligned}$$

Les sommes pouvant être interverties car toutes les séries convergent normalement. En effet,  $\bar{x}^j$  et  $\bar{x}^j$  sont tous deux majorer par  $\max_i |x_i|^j$  et  $\sum_{j=0}^{+\infty} \frac{\max_i |x_i|^j}{j!} = \exp(\max_i |x_i|) < \infty$ .

# Annexe B

## Définitions mathématiques

### B.1 Définitions de lois

#### B.1.1 Loi log-normale

Une variable  $X$  est dite suivre une loi log normale de paramètres  $(\mu; \sigma)$  si et seulement son logarithme, la variable  $\log(X)$ , suit une loi normale d'espérance  $\mu$  et d'écart type  $\sigma$ . La loi de  $X$  se note alors  $X \sim \mathcal{LN}(\mu, \sigma)$  et sa densité  $X$  a pour expression :  $f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$ .

#### B.1.2 Loi gamma

Une variable  $X$  est dite suivre une loi gamma de paramètres de forme  $(\alpha)$  et d'échelle  $(\beta)$  lorsque sa densité a pour expression  $f(x; \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$ . La loi de  $X$  se note alors  $X \sim \mathcal{G}(\alpha, \beta)$ .

#### B.1.3 Loi log-normale

Une variable  $X$  est dite suivre une loi de Weibull de paramètres de forme  $(\lambda)$  et d'échelle  $(\kappa)$  lorsque sa densité a pour expression  $f(x; \lambda, \kappa) = \frac{\lambda}{\kappa} \left(\frac{x}{\kappa}\right)^{\lambda-1} e^{-(x/\kappa)^\lambda}$ . La loi de  $X$  se note alors  $X \sim \mathcal{W}(\lambda, \kappa)$ .

#### B.1.4 Loi normale

Une variable  $X$  est dite suivre une loi normale de paramètres de moyenne  $(\mu)$  et d'écart-type  $(\sigma)$  lorsque sa densité a pour expression  $f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ . La loi de  $X$  se note alors  $X \sim \mathcal{N}(\mu, \sigma)$ .

#### B.1.5 Loi de Pareto

Une variable  $X$  est dite suivre une loi de Pareto de paramètres de forme  $(\alpha)$  et d'échelle  $(\beta)$  lorsque sa densité a pour expression  $f(x; \alpha, \beta) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}}$ . La loi de  $X$  se note alors  $X \sim \mathcal{P}(\alpha, \beta)$ .

## B.2 Définitions de tests statistiques

Pour chaque test statistique les notations suivantes seront reprises :

- $F$  est la fonction de répartition théorique.
- $(X_1, \dots, X_n)$  est l'échantillon, avec  $n$  sa taille.
- $X_{(i)}$  est la  $i$ -ème valeur la plus grande de l'échantillon.
- $F_n$  est la fonction de répartition empirique de l'échantillon.

### B.2.1 Test de Kolmogorov-Smirnov

Le test a pour hypothèse nulle que l'échantillon suit la distribution théorique. Sa statistique a pour expression  $D = \max |F_n(x) - F(x)|$ . Elle suit approximativement une loi de Kolmogorov-Smirnov sous l'hypothèse nulle.

### B.2.2 Test d'Anderson-Darling

Le test a pour hypothèse nulle que l'échantillon suit la distribution théorique. Sa statistique a pour expression  $A^2 = -n - \sum_{i=1}^n \frac{2i-1}{n} (\ln(F(X_{(i)})) + \ln(1 - F(X_{(n+1-i)})))$ . Elle suit une distribution spécifique sous l'hypothèse nulle.

### B.2.3 Test de Kuiper

Le test a pour hypothèse nulle que l'échantillon suit la distribution théorique. Sa statistique a pour expression  $[V = D^+ + D^-]$  avec  $[D^+ = \max[F_n(x) - F(x)]]$  et  $[D^- = \max[F(x) - F_n(x)]]$ . Sous l'hypothèse nulle  $\sqrt{n}V$  suit une loi de Kuiper.

# Annexe C

## Article 230 du règlement délégué 2.A de Solvabilité II

Informations et hypothèses sous-tendant les méthodes utilisées pour calculer la distribution de probabilité prévisionnelle

1. Des informations ne sont considérées comme crédibles, aux fins de l'article 121, paragraphe 2, de la directive 2009/138/CE, que lorsque l'entreprise d'assurance ou de réassurance fournit la preuve de leur cohérence et de leur objectivité, de la fiabilité de la source dont elles proviennent et de la transparence de la méthode par laquelle elles ont été générées et traitées.

2. Des hypothèses ne sont considérées comme réalistes, aux fins de l'article 121, paragraphe 2, de la directive 2009/138/CE, que lorsqu'elles remplissent l'ensemble des conditions suivantes :

(a) l'entreprise d'assurance ou de réassurance est en mesure d'expliquer et de justifier chacune des hypothèses retenues, compte tenu de l'importance de l'hypothèse considérée, de l'incertitude qui lui est liée et des raisons pour lesquelles elle n'utilise pas les hypothèses alternatives pertinentes ;

(b) les circonstances dans lesquelles les hypothèses retenues seraient considérées comme fausses peuvent être clairement identifiées ;

(c) l'entreprise d'assurance ou de réassurance établit et conserve une explication écrite de la méthode qu'elle utilise pour formuler ces hypothèses.

# Bibliographie

- [1] N. BOUSQUET. *Modélisation et statistique bayésienne computationnelle*. Notes de cours de Master 2, Sorbonne Université/ISUP, 2023.
- [2] J.O. BERGER. J.M. BERNARDO et D. SUN. *The formal definition of reference priors*. T. 37. The Annals of Statistics. Institute of Mathematical Statistics, 2009.
- [3] R. Yang et J.O. BERGER. *A Catalog of Noninformative Priors*. Institute of Statistics et Decision Sciences, Duke University, 1996.
- [4] D. PAPACHRISTOU. *Theoretical and Practical Aspects of Parameter and Model Uncertainty*. 36th ANNUAL GIRO CONVENTION, 2009.
- [5] G. VENTER et R. SAHASRABUDDHE. *A Note on Parameter Risk*. Casualty Actuarial Society E-Forum, Summer 2012, 41, 2012.
- [6] A. VEBER. *Modélisation de la sinistralité non vie, aléa sur la détermination des paramètres des lois de modélisation et impact sur la VaR*. Mémoire d'actuariat, Université Paris Dauphine, 2009.
- [7] M.G. WACEK. *Parameter Uncertainty in Loss Ratio Distributions and its Implications*. ASTIN Bulletin, 6, 195-202, 2005.
- [8] Page personnelle de Y. CORTIAL. *Illustration de la longueur des demi-axes d'une ellipse en fonction des valeurs propres*. URL : <https://web.cortial.net/optim/annexe2.html>.