

**Mémoire présenté le :**  
**pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA**  
**et l'admission à l'Institut des Actuaires**

Par : Pierre-Olivier Goffard

Titre : Echantillonneurs de Monte Carlo séquentiels pour la calibration et la sélection d'un modèle composite pour les pertes en assurance non-vie.

Confidentialité :  NON       (Durée :  1 an     2 ans)

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

*Membres présents du jury de Signature  
l'Institut des Actuaires*

Nicolas Zec  
.....  
.....  
.....

*Membres présents du jury de  
l'ISFA*

Stéphane Loisel  
.....  
.....  
.....

*Entreprise :*

*Nom :*

*Signature :*

*Directeur de mémoire en entre-  
prise :*

*Nom :*

*Signature :*

*Invité :*

*Nom :*

*Signature :*

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

Signature du responsable entreprise

Signature du candidat

*Goffard*

## Résumé

La distribution des montants indemnisés en assurance non vie est caractérisée par une forte occurrence de petits montants (attritionnels) et une occurrence plus faible sans être négligeable de montants importants (atypiques ou extrêmes). Des modèles "composites" se sont développés dans la littérature actuarielle pour prendre en compte cette particularité. Ces modèles combinent deux lois de probabilités, la première modélise la partie attritionnelle tandis que la deuxième se focalise sur la partie atypique ou extrême de la distribution des montants indemnisés. Les paramètres de ces modèles synthétisent la sinistralité, l'un des paramètres joue le rôle de seuil permettant de distinguer les sinistres graves des autres sinistres. La calibration de ces modèles s'effectue généralement en utilisant la technique du maximum de vraisemblance. Ce travail de recherche propose d'utiliser une approche bayésienne reposant sur un algorithme d'échantillonnage de Monte Carlo séquentiel. La méthode est validée sur des données simulées et illustrée sur des jeux de données réelles.

**Mots clés:** Distribution des pertes; Modèles composites; Seuil de valeurs extrêmes; Statistique bayésienne; Monte Carlo séquentiel.

## Abstract

The insurance claim sizes distribution is characterized by a high frequency of small amounts and a lower, but not insignificant, occurrence of large claim amounts. Composite models, which link two probability distributions, one for the “belly” and the other for the “tail” of the loss distribution, have emerged in the actuarial literature to take this specificity into account. The parameters of these models summarize the distribution of the losses. One of them corresponds to the breaking point between small and large claim amounts. The composite models are usually fitted using maximum likelihood estimation. A Bayesian approach is considered in this work. A sequential Monte Carlo sampler is designed to sample from the posterior distribution to fit and compare the composite models. The method is validated via a simulation study and illustrated on insurance loss datasets.

**Keywords:** Loss distribution; Composite models; Extreme value threshold; Bayesian statistics; Sequential Monte Carlo.

## **Remerciements**

Ce travail a été initié lors de ma prise de fonction à l'Institut de Science Financière et d'Assurances en tant que maître de conférence des universités en septembre 2018. Je tiens à remercier les personnes que j'ai eu la chance de rencontrer et avec qui j'ai pu collaborer lors de ces trois dernières années. Je veux remercier les enseignant-e-s, les personnel·les administratifs, les chercheur-e-s post doctoraux, les doctorant-e-s et bien entendu les étudiant-e-s des parcours ACTUARIAT et ECO-STAT.

## Table des matières

<b>Résumé</b>	2
<b>Abstract</b>	3
<b>Remerciements</b>	4
<b>1 Introduction</b>	6
<b>2 Modélisation paramétrique de la loi des montants de sinistres</b>	8
2.1 Les lois classiques . . . . .	8
2.2 Les modèles composites . . . . .	11
<b>3 Inférence bayésienne des modèles composites</b>	14
3.1 La marche aléatoire de Metropolis-Hasting . . . . .	15
3.2 L'échantillonneur de Gibbs . . . . .	18
3.3 Application aux modèles composites . . . . .	18
<b>4 Sélection Bayésienne d'un modèle composite</b>	20
4.1 Echantillonnage préférentiel . . . . .	20
4.2 Algorithme de Monte Carlo séquentiel . . . . .	22
4.3 Les critères d'informations Bayésiens . . . . .	25
4.4 Application aux modèles composites . . . . .	26
<b>5 Etude de simulation</b>	28
5.1 Convergence des estimations a posteriori . . . . .	29
5.2 Cohérence des méthodes de sélection de modèle . . . . .	30
<b>6 Application sur un jeu de données réelles</b>	33
6.1 Analyse globale de la sinistralité . . . . .	35
6.2 Analyse annuelle de la sinistralité . . . . .	38
6.2.1 Analyse annuelle de la sinistralité sans le modèle lognormal-Pareto . . . . .	41
6.2.2 Analyse annuelle de la sinistralité en remplaçant le modèle Lognormal-Pareto par un modèle lognormal . . . . .	42
6.3 Analyse mensuelle de la sinistralité . . . . .	43
6.3.1 Analyse mensuelle de la sinistralité sans le modèle Lognormal-Pareto . . . . .	45
6.3.2 Analyse mensuelle de la sinistralité en remplaçant le modèle Lognormal-Pareto par un modèle lognormal . . . . .	46
6.4 Discussion des résultats de l'analyse sur données réelles . . . . .	47
<b>7 Conclusion et perspectives</b>	50

## 1 Introduction

La distribution des montants indemnisés en assurance non-vie est caractérisée par une forte occurrence de petits montants (attritionnels) et une occurrence plus faible sans être négligeable de montants importants (atypiques ou extrêmes). La survenance de valeurs extrêmes rend très volatile l'évaluation du coût moyen ce qui entraîne une augmentation de la prime pure. L'inclusion d'un chargement dans la prime pour compenser ces pertes extrêmes passe par la définition d'un seuil de perte extrême  $\gamma$  qui discrimine la partie attritionnelle et la partie exceptionnelle du risque associé au portefeuille. Une approche classique par coût moyen/fréquence moyenne, par exemple via un modèle de Regression Linéaire Généralisé (GLM), est appliquée sur les sinistres de faible intensité. Les sinistres de forte intensité sont modélisés à part via un modèle adapté aux valeurs extrêmes, par exemple au moyen de la loi de Pareto. Le coût des sinistres extrêmes est estimé puis repartis sur les primes de l'ensemble des assurés en portefeuille. Cette pratique est présentée dans les mémoires d'actuariat de [Naji \[2016\]](#); [Raoult \[2016\]](#); [Vegni \[2011\]](#). La détermination du seuil permet aussi la construction d'une base de données d'entraînement pour calibrer un algorithme d'apprentissage statistique afin d'estimer la probabilité qu'un sinistre soit grave ou qu'un assuré subisse un sinistre grave, c'est l'objet du mémoire de [Bouché \[2014\]](#).

L'inadéquation entre les primes et les sinistres de forte intensité doit être prise en compte dans la gestion des risques et le calcul des réserves de capital de solvabilité à un an. L'évaluation sur le périmètre non vie peut-être effectuée en approchant les quantiles de la distribution d'un modèle collectif dont les montants de sinistres exceptionnels suivent une loi de probabilité à queue lourde. C'est l'approche utilisée dans le travail de [Baarani \[2017\]](#) pour évaluer les *Short Term Economic Capital* (STEC) associés aux risques de pertes futures. Ce risque couvre les écarts potentiels entre le montant estimé et le montant réel des sinistres. Il participe à la détermination du niveau de fonds propre requis, aussi appelé *Solvency Capital Requirements* (SCR), sur le périmètre non-vie.

La survenance de sinistres extrêmes concerne de nombreuses branches d'activité du périmètre Incendie, Accidents et Risques Divers (IARD) incluant l'assurance habitation [Naji \[2016\]](#), auto [Baarani \[2017\]](#); [Bouché \[2014\]](#); [Vegni \[2011\]](#), santé [Xu \[2017\]](#), et risques industriels [Raoult \[2016\]](#). La détermination du seuil de perte extrême emploie dans la plupart de ces travaux les outils de la théorie des valeurs extrêmes [Beirlant et al. \[2004\]](#). Ces outils prennent souvent une forme graphique. Les quantiles empiriques des montants et du logarithme des montants sont comparées aux quantiles de la loi exponentielle via des diagrammes quantile-quantile. La courbe des excédents moyens (Mean Excess Plot), la courbe de [Hill \[1975\]](#) et la courbe de [Gerstengarbe and Werner \[1989\]](#) sont souvent étudiées pour décider du meilleur seuil. Ce mémoire propose de définir un modèle composite comme mélange de deux modèles l'un pour les sinistres de faible montant (loi gamma, Weibull ou lognormal) et l'autre pour les montants plus importants (loi de Pareto). Ce modèle inclut un paramètre de seuil  $\gamma$ . Ce paramètre peut-être estimé via les outils graphiques mentionnés précédemment, les paramètres des deux parties sont ensuite inférés respectivement sur la base des observations inférieures et supérieures au seuil. Une autre méthode, introduite par [Cooray and Ananda \[2005\]](#), consiste à estimer tous les paramètres simultanément par maximum de vraisemblance. Les différentes combinaisons de modèles sont alors comparées via des critères d'information tels que le critère d'Akaike ou le critère Bayésien, voir par exemple l'étude de [Abu Bakar et al.](#)

[2015] ou [Grün and Miljkovic \[2019\]](#). Une revue récente de la modélisation du montant des sinistres via les modèles composites est effectuée dans le travail de [Wang et al. \[2020\]](#).

L'objectif de ce mémoire est de proposer une méthode bayésienne de calibration et de comparaison des modèles composites. La statistique bayésienne définit les paramètres d'un modèle comme des variables aléatoires. Une loi a priori est associée à chacun des paramètres, puis mise à jour grâce à la fonction de vraisemblance pour obtenir la loi à posteriori, voir l'ouvrage de [Gelman et al. \[2013b\]](#). La loi à posteriori est souvent inaccessible analytiquement (sauf dans le cas où les lois a priori et a posteriori sont conjuguées). Les algorithmes Markov Chain Monte Carlo (MCMC) permettent de simuler des réalisations de la loi a posteriori en générant des chaînes de Markov dont la loi stationnaire cible la loi a posteriori. Le résultat est un échantillon de valeurs de paramètres distribuées suivant la loi a posteriori. Des estimateurs ponctuels sont obtenus en prenant par exemple la moyenne empirique de ces échantillons. De tels méthodes d'estimations ont été utilisées dans le cadre des modèles composites, voir par exemple [Cooray and Cheng \[2013\]](#), le modèle composite lognormal-Pareto de [Scollnik \[2007\]](#) fait partie des exemples du logiciel WinBUGS [[Lunn et al., 2000](#), Exemples Vol. III]. Dans ce mémoire, un algorithme de type *Sequential Monte Carlo* (SMC) est implémenté. Cette algorithmes, décrit dans les travaux de [Moral et al. \[2006\]](#), consiste à faire évoluer un nuage de particules pondérées qui *in fine* seront distribuées comme la loi a posteriori. Le poids des particules au fur et à mesure des estimations donne accès à la constante de normalisation de la loi a posteriori permettant la comparaison de modèles via les facteurs de Bayes, voir l'article de référence de [Kass and Raftery \[1995\]](#).

Ce mémoire est organisé de la façon suivante. La [Section 2](#) introduit les modèles paramétriques couramment utilisés pour étudier la distribution du montant des sinistres. La [Section 3](#) rappelle les principes de l'inférence bayésienne et décrit les algorithmes permettant d'échantillonner la loi a posteriori. La [Section 4](#) présente l'algorithme de Monte Carlo séquentiel, principal objet de ce travail de recherche, et montre comment l'utiliser pour sélectionner un modèle. La [Section 5](#) étudie la convergence des estimateurs dans le cadre d'une application à des données générées artificiellement. La [Section 6](#) illustre la méthode sur un échantillon de montants de sinistres automobiles provenant d'un assureur australien.

## 2 Modélisation paramétrique de la loi des montants de sinistres

Les montants de sinistres en assurance sont représentés par une variable aléatoire  $X$  positive dont la loi de probabilité est continue de densité notée  $f(x; \theta)$ , où  $\theta \in \Theta \subset \mathbb{R}^d$  désigne les paramètres du modèle. L'objectif de la statistique paramétrique est de trouver le paramètre  $\hat{\theta}$  qui permet la meilleure adéquation du modèle aux données  $\mathbf{x} = (x_1, \dots, x_n)$  vues comme des réalisations indépendantes et identiquement distribuées (i.i.d.) de  $X$ . Une approche standard consiste à chercher la valeur du paramètre qui maximise la fonction de vraisemblance définie par

$$L(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i; \theta),$$

qui correspond à la probabilité d'obtenir l'échantillon  $\mathbf{x}$  avec le modèle considéré. L'estimateur du maximum de vraisemblance est donné par

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L(\mathbf{x}|\theta).$$

Plusieurs modèles sont comparés au moyen de critères d'information mesurant l'adéquation aux données. La [Section 2.1](#) rappelle quelques lois paramétriques classiques pour modéliser les pertes en assurance et la [Section 2.2](#) introduit les lois composites. Les modèles seront calibrés sur un jeu de données célèbre comprenant des montants indemnisés au Danemark au titre d'une garantie "incendie". Les données sont récupérées via le paquet R `SMPRACTICALS` qui accompagne l'ouvrage de [Davison \[2011\]](#). La [Figure 1](#) montre l'histogramme et la boîte à moustache de la distribution des pertes. La distribution est typique des pertes en assurance avec une

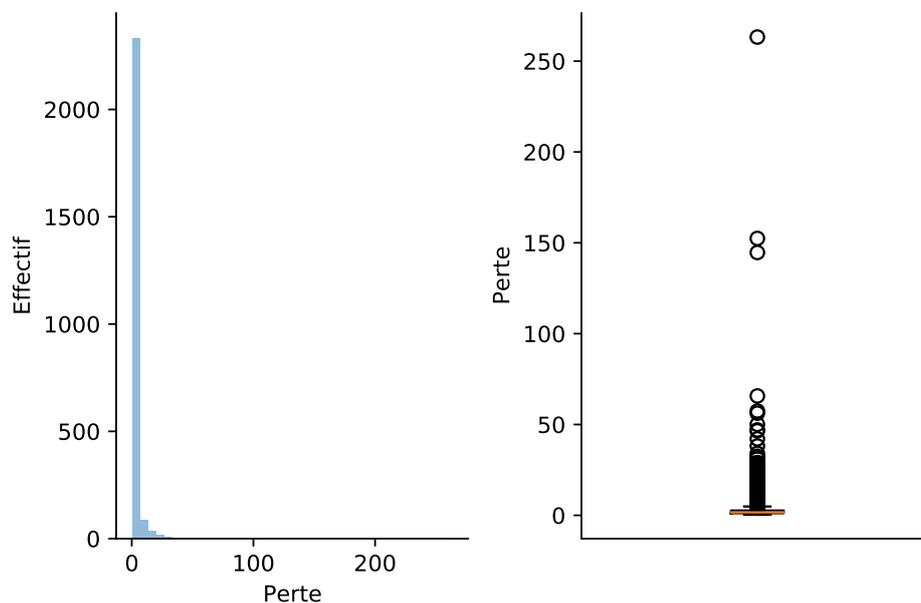


Figure 1 – Histogramme et boîte à moustache des pertes "incendie" danoises.

occurrence forte de faibles montants et une queue de distribution alourdie par des pertes extrêmes.

### 2.1 Les lois classiques

La loi gamma, notée  $\text{Gamma}(r, m)$ , de densité

$$f(x; r, m) = \frac{e^{-x/m} x^{r-1}}{m^r \Gamma(r)}, \quad x > 0, \quad (1)$$

est une loi couramment utilisée en assurance. Il s'agit du modèle sous-jacent à l'estimation du coût moyen par les modèles linéaires généralisés pour la tarification. La queue de la distribution admet une décroissance exponentielle avec

$$\mathbb{P}(X > x) \sim e^{-x/m}, \text{ lorsque } x \rightarrow +\infty,$$

ce qui n'est pas adapté au vu de la distribution empirique des données sur la [Figure 1](#). Les paramètres estimés par maximum de vraisemblance sont donnés par  $\widehat{r} = 1.30$  et  $\widehat{m} = 2.61$ . Le diagramme quantile-quantile sur la [Figure 2](#) montre que la distribution gamma ne permet pas une adéquation satisfaisante. Les quantiles

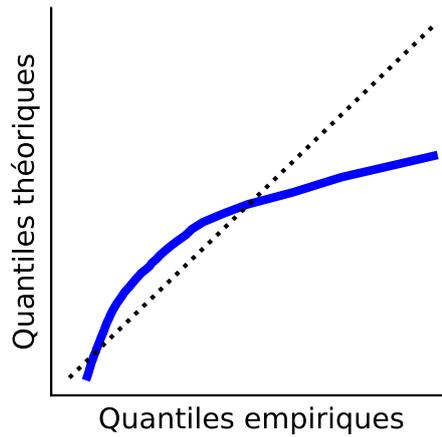


Figure 2 – Diagramme quantile-quantile d'ajustement de la loi gamma  $\text{Gamma}(\widehat{r} = 1.30, \widehat{m} = 2.61)$  aux pertes "incendie" danoises.

empiriques d'ordre élevé sont bien supérieurs à ceux de la loi gamma. Il est nécessaire de proposer des modèles permettant l'occurrence de sinistres plus sévères en termes de montant. C'est le cas de la loi de Weibull, notée  $\text{Weib}(k, \beta)$ , de densité

$$f(x; k, \beta) = \frac{k}{\beta} \left( \frac{x}{\beta} \right)^{k-1} e^{-(x/\beta)^k}, \quad x > 0. \quad (2)$$

La fonction de survie de la loi de Weibull est donnée par

$$\bar{F}(x; k, \beta) = \Pr(X > x) = e^{-(x/\beta)^k}. \quad (3)$$

La queue de la distribution est plus lourde que celle de la loi gamma si le paramètre de forme vérifie  $k < 1$  et plus légère si  $k > 1$ . L'inférence de ce paramètre caractérise la sévérité des sinistres. La loi lognormale, notée  $\text{LogNorm}(\mu, \sigma)$ , de densité

$$f(x; \mu, \sigma) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{[\ln(x)-\mu]^2}{2\sigma^2}}, \quad x > 0. \quad (4)$$

possède une queue de distribution d'autant plus lourde que le paramètre de variance  $\sigma$  augmente, cette distribution est couramment utilisée pour modéliser les pertes opérationnelles, c.f. [Shevchenko and Wüthrich \[2006\]](#). Sur les données "incendie" danoises, les paramètres de la loi de Weibull sont estimés par  $\widehat{k} = 0.96$  et  $\widehat{\beta} = 3.29$  et ceux de la loi lognormale par  $\widehat{\mu} = 0.79$  et  $\widehat{\sigma} = 0.72$ . La [Figure 3](#) montre les diagrammes quantile-quantile associés à chacun des deux modèles. Bien que l'adéquation semble visuellement meilleure, la queue de distribution des modèles Weibull et lognormale est toujours trop légère. Le modèle paramétrique de référence

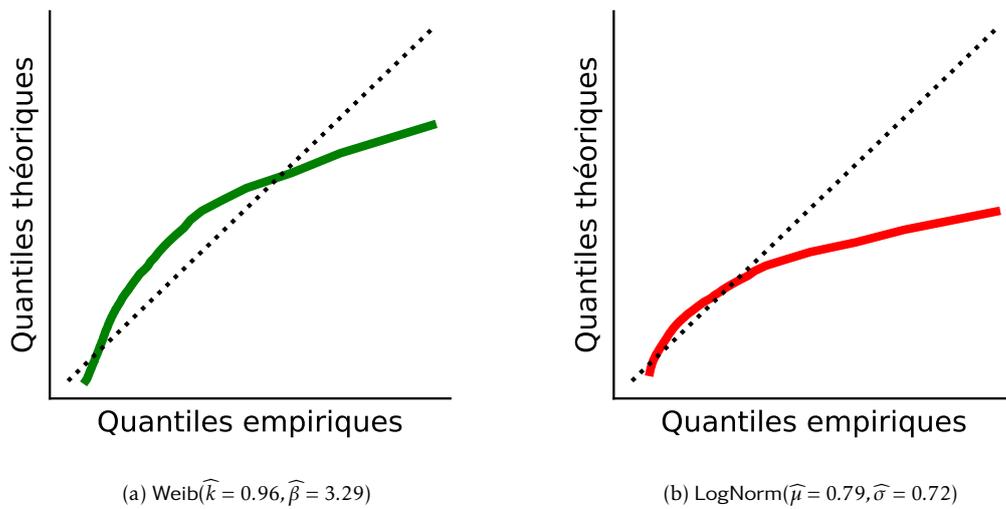


Figure 3 – Diagramme quantile-quantile d'ajustement des lois de Weibull (vert) et lognormale (rouge) aux pertes "incendie" danoises.

en théorie de valeurs extrêmes est la loi de Pareto, notée  $\text{Par}(\alpha, \gamma)$ , de densité

$$f(x; \gamma, \alpha) = \begin{cases} 0, & x \leq \gamma, \\ \frac{\alpha \gamma^\alpha}{x^{\alpha+1}}, & x > \gamma. \end{cases} \quad (5)$$

La fonction de survie de la loi de Pareto est donnée par

$$\bar{F}(x; \gamma, \alpha) = \begin{cases} 1, & x \leq \gamma, \\ \left(\frac{\gamma}{x}\right)^\alpha, & x > \gamma. \end{cases}$$

La queue de distribution, qui décroît de manière polynomiale (bien plus lentement que l'exponentielle), est d'autant plus lourde que le paramètre  $\alpha$  est petit. En particulier, si le paramètre  $\alpha$  est inférieur à 1, la distribution des montants n'admet pas de moments d'ordre 1 ce qui rend l'approche de tarification par coût moyen inutilisable (la moyenne empirique peut toujours être calculée mais l'estimateur est très volatile). Les estimateurs du maximum de vraisemblance du modèle Pareto, pour les données danoises, sont données par  $\hat{\gamma} = 0.31$  et  $\hat{\alpha} = 0.54$ . Le diagramme quantile-quantile est donné sur la [Figure 4](#). Le modèle Pareto admet une queue de distribution plus lourde que la distribution empirique. L'inférence du paramètre  $\alpha$  est influencée par les sinistres de faible intensité, l'écart important entre les sinistres attritionnels et atypiques contribue à alourdir la queue de distribution conduisant à sous-estimer le paramètre  $\alpha$ .

A la vue de ces résultats, il semble naturel d'introduire un modèle plus flexible combinant une distribution adaptée aux sinistres de faible montant (gamma, lognormale ou Weibull) et une distribution à queue lourde (Pareto) pour les sinistres de montant important. C'est l'objet des modèles composites, introduits dans la section suivante.

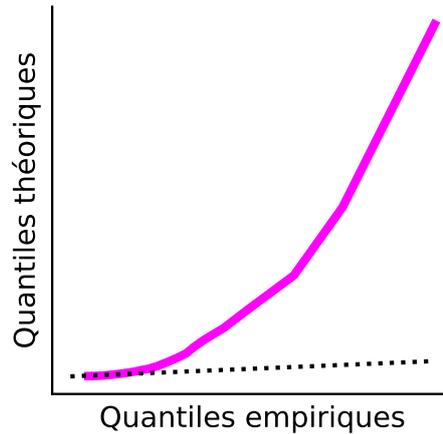


Figure 4 – Diagramme quantile-quantile d'ajustement de la loi Pareto  $\text{Par}(\hat{\alpha} = 0.54, \hat{\gamma} = 0.31)$  aux pertes "incendie" danoises.

## 2.2 Les modèles composites

Un modèle composite est obtenu en combinant deux modèles, l'un pour le "ventre" de la distribution et l'autre pour la "queue" la distribution. Sa densité est donnée par

$$f(x) = \begin{cases} p \frac{f_1(x)}{F_1(\gamma)}, & \text{si } x \leq \gamma, \\ (1-p) \frac{f_2(x)}{1-F_2(\gamma)}, & \text{si } x > \gamma, \end{cases} \quad (6)$$

où  $f_1, f_2$  et  $F_1, F_2$  sont respectivement les fonctions de densité et de répartition des modèles pour le "ventre" et la "queue" de la distribution. Le paramètre  $p \in (0, 1)$  est le paramètre de "mélange". Le modèle comprend un paramètre seuil  $\gamma > 0$  permettant la distinction entre les pertes attritionnelles et les pertes extrêmes. Des conditions de continuité et dérivabilité en  $\gamma$  sont couramment imposés sur la densité de probabilité du modèle composite, soient

$$f(\gamma^-) = f(\gamma^+), \text{ et } f'(\gamma^-) = f'(\gamma^+). \quad (7)$$

Cela a pour effet de "fixer" deux paramètres du modèle. Pour le ventre de la distribution, seuls les modèles gamma, Weibull et lognormale seront considérés dans ce travail. Le modèle pour la queue de distribution est la loi de Pareto  $\text{Par}(\alpha, \gamma)$  qui admet par défaut le paramètre de seuil  $\gamma$ . L'effet des conditions de régularité (7) sur les modèles composites  $\text{Gamma}(r, m) - \text{Par}(\alpha, \gamma)$ ,  $\text{Weib}(k, \beta) - \text{Par}(\alpha, \gamma)$  et  $\text{LogNorm}(\mu, \sigma) - \text{Par}(\alpha, \gamma)$  sont détaillés dans les exemples 1, 2 et 3 respectivement.

**Exemple 1** Lorsque  $f_1$  est la densité de la loi  $\text{Gamma}(r, m)$  et que  $f_2$  est la densité de la loi  $\text{Par}(\alpha, \gamma)$ , les conditions de régularité (7) conduisent à exprimer les paramètres  $m$  et  $p$  en fonction des autres paramètres avec

$$m = \frac{\gamma}{k + \alpha}, \quad p = \frac{\alpha \Gamma(k) F_1(\gamma; r, m) e^{k+\alpha} (k + \alpha)^{-k}}{1 + \alpha \Gamma(k) F_1(\gamma; r, m) e^{k+\alpha} (k + \alpha)^{-k}},$$

où  $F_1(\gamma; r, m)$  désigne la fonction de répartition de la distribution gamma.

**Exemple 2** Lorsque  $f_1$  est la densité de la loi  $\text{Weib}(k, \beta)$  et que  $f_2$  est la densité de la loi  $\text{Par}(\alpha, \gamma)$ , les conditions de

régularité (7) conduisent à exprimer les paramètres  $\beta$  et  $p$  en fonction des autres paramètres avec

$$\beta = \left( \frac{k}{k + \alpha} \right)^{1/k} \gamma, \quad p = \frac{\frac{\alpha}{\gamma} \left[ 1 - e^{-\frac{k+\alpha}{k}} \right]}{\frac{\alpha}{\gamma} + \frac{k}{\gamma} e^{-\frac{k+\alpha}{k}}}.$$

Le modèle Weib( $k, \beta$ ) – Par( $\alpha, \gamma$ ) a été traité spécifiquement dans le travail de [Scollnik and Sun \[2012\]](#).

**Exemple 3** Lorsque  $f_1$  est la densité de la loi LogNorm( $\mu, \sigma$ ) et que  $f_2$  est la densité de la loi Par( $\alpha, \gamma$ ), les conditions de régularité (7) conduisent à exprimer les paramètres  $\mu$  et  $p$  en fonction des autres paramètres avec

$$\mu = \ln(\gamma) - \alpha\sigma^2, \quad p = \frac{\alpha\sigma\sqrt{2\pi}\Phi(\alpha\sigma)}{\alpha\sigma\sqrt{2\pi}\Phi(\alpha\sigma) + e^{-\alpha^2\sigma^2/2}},$$

où  $\Phi$  désigne la fonction de répartition de la loi normale centrée et réduite. Le modèle LogNorm( $\mu, \sigma$ ) – Par( $\alpha, \gamma$ ) a été traité spécifiquement dans les travaux de [Cooray and Ananda \[2005\]](#) et [Scollnik \[2007\]](#).

Les paramètres des modèles composites sont usuellement estimés par maximum de vraisemblance. La [Figure 5](#) montre les diagrammes quantile-quantiles des modèles composites Gamma( $r$ ) – Par( $\alpha, \gamma$ ), Weib( $k$ ) – Par( $\alpha, \gamma$ ) et LogNorm( $\sigma$ ) – Par( $\alpha, \gamma$ ) calibrés sur le jeu de données des pertes "incendie"s danoises. L'adéquation des modèles composites est visuellement meilleure que celle des modèles dits classiques. Parmi les modèles composites, il semblerait que le modèle LogNorm( $\sigma$ ) – Par( $\alpha, \gamma$ ) soit le plus adapté, voir [Figure 5c](#). En complément de l'utilisation d'outils graphiques, la sélection d'un modèle se base aussi sur des critères d'information qui reposent sur une fonction de coût appelée déviance et égale à  $D = -2l(\mathbf{x}|\hat{\theta})$ , où  $l(\mathbf{x}|\theta) = \ln[L(\mathbf{x}|\theta)]$  désigne la log-vraisemblance et  $\hat{\theta}$  est l'estimateur du maximum de vraisemblance de  $\theta$ . L'adéquation des modèles aux données augmente avec le nombre de paramètres pouvant mener au sur-apprentissage, c'est pourquoi les critères d'information pénalisent les modèles ayant plus de paramètres. Le critère d'information d'[Akaike \[1998\]](#) (AIC) est donnée par

$$\text{AIC} = 2d - 2l(\mathbf{x}|\theta),$$

et le critère d'information Bayésien de [Schwarz \[1978\]](#) (BIC) est défini par

$$\text{BIC} = d \ln(n) - 2l(\mathbf{x}|\theta),$$

où  $n$  correspond au nombre d'observations et  $d$  est égal au nombre de paramètres du modèle. Le [Tableau 1](#) rappelle la valeur inférée des paramètres et reporte la valeur de chacun des critères d'information pour l'ensemble des modèles évoqués dans les sections 2.1 et 2.2. Le passage aux modèles composites permet un gain notable en termes d'AIC et de BIC. Le modèle préconisé par ces indicateurs est le modèle qui mélange la loi de Weibull avec la loi de Pareto. Cela va à l'encontre de la première intuition graphique fournie par la [Figure 5](#). Le seuil de valeurs extrêmes est différent suivant chacun des modèles composites. Le plus faible est à noter pour le modèle Weib – Par avec  $\hat{\gamma} = 1$ . Le paramètre de la loi de Weibull estimé à  $\hat{k} = 14.03$  indique une queue de distribution très légère pour la partie attritionnelle des montants de sinistres compensée par un indice de queue  $\hat{\alpha} = 1.26$  plus faible que pour les autres modèles. L'intérêt des modèles composites est illustré ici par une meilleure adéquation aux données et une description exhaustive de la distribution du montant des sinistres. Dans la suite de l'étude, le paramètre correspondant à la portion attritionnelle des sinistres sera appelé paramètre de forme, celui correspondant à la queue de la distribution sera appelé paramètre de queue et

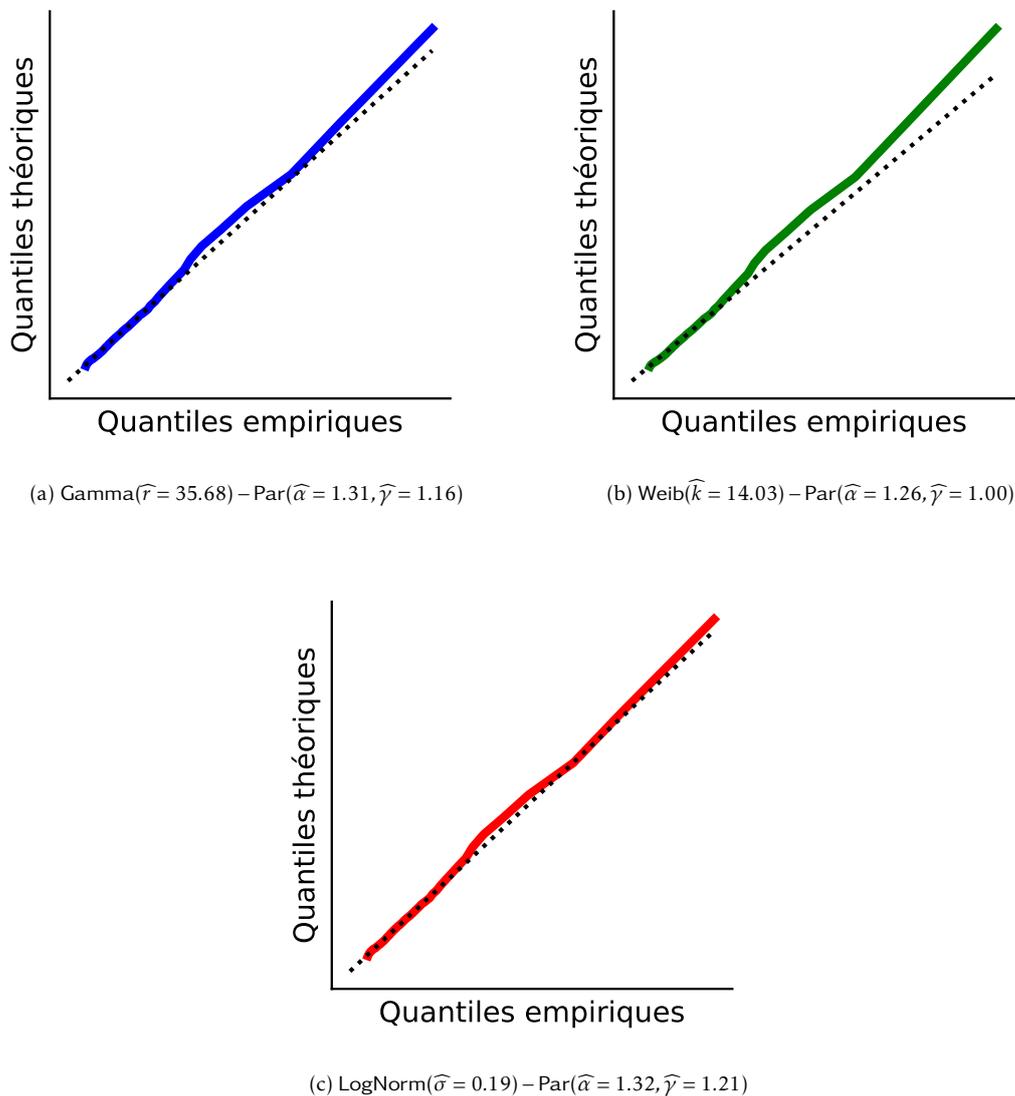


Figure 5 – Diagramme quantile-quantile d'ajustement des modèles composites aux pertes "incendie" danoises.

enfin le paramètre  $\gamma$  sera le paramètre de seuil.

L'objectif de ce travail est d'explorer des approches bayésiennes pour calibrer et choisir le meilleur modèle. Par comparaison avec l'approche fréquentiste du maximum de vraisemblance, l'approche bayésienne permet d'inclure de l'information sur les paramètres à estimer via la loi a priori. Cela peut permettre de compenser par exemple un manque de données, situation courante lorsque des scénarios extrêmes sont considérés. L'autre avantage est l'obtention d'une distribution des valeurs possibles pour les paramètres et donc d'intervalle dit de crédibilité au lieu d'une estimation ponctuelle. La section suivante présente succinctement l'inférence Bayésienne et les méthodes de simulation qui lui sont associées.

Modèle			AIC	BIC
Gamma	$\widehat{r} =$	1.26	10,490.05	10,501.69
	$\widehat{m} =$	2.43		
Weib	$\widehat{k} =$	0.947	10,544.94	10,556.58
	$\widehat{\delta} =$	2.95		
LogNorm	$\widehat{\mu} =$	0.67	8,931.19	8,942.83
	$\widehat{\sigma} =$	0.73		
Par	$\widehat{\alpha} =$	0.54	11,354.19	11,365.83
	$\widehat{\theta} =$	0.31		
Gamma – Par	$\widehat{r} =$	35.68	7,723.68	7,741.14
	$\widehat{\alpha} =$	1.31		
	$\widehat{\theta} =$	1.15		
Weib – Par	$\widehat{k} =$	14.03	7,686.75	7,704.21
	$\widehat{\alpha} =$	1.26		
	$\widehat{\gamma} =$	1.00		
LogNorm – Par	$\widehat{\sigma} =$	0.19	7,737.73	7,755.19
	$\widehat{\alpha} =$	1.32		
	$\widehat{\gamma} =$	1.20		

Table 1 – Paramètres et BIC pour les modèles de pertes présentés dans les Sections 2.1 et 2.2.

### 3 Inférence bayésienne des modèles composites

L'objet d'étude de la statistique bayésienne est la loi a posteriori du paramètre  $\theta$  sachant les données  $\mathbf{x} = (x_1, \dots, x_n)$  définie par

$$\pi(\theta|\mathbf{x}) = \frac{L(\mathbf{x}|\theta)\pi(\theta)}{Z(\mathbf{x})}. \quad (8)$$

Il s'agit d'une simple application du théorème de Bayes pour mettre à jour la loi a posteriori  $\pi(\theta)$  via la fonction de vraisemblance  $L(\mathbf{x}|\theta)$ . Le dénominateur dans l'équation (8) est une constante de normalisation donnée par

$$Z(\mathbf{x}) = \int_{\Theta} L(\mathbf{x}|\theta)\pi(\theta)d\theta, \quad (9)$$

appelée vraisemblance marginale. La loi a posteriori (8) est utilisée à des fins d'inférence. Une estimation ponctuelle du paramètre s'obtient en prenant le mode ou la moyenne de cette distribution. Les méthodes bayésiennes sont le sujet de nombreux ouvrages comme celui de [Gelman et al. \[2013b\]](#). Le principal problème réside dans l'évaluation de  $Z(\mathbf{x})$ . L'intégrale qui définit la constante de normalisation (9) est souvent de grande dimension et n'admet pas de forme explicite. Les rares cas où la loi a posteriori est connue explicitement sont les cas où la loi a priori est dite conjuguée, c'est à dire que les lois a priori et a posteriori sont de même nature, voir l'[Exemple 4](#).

**Exemple 4** Si les montants de sinistres sont de loi exponentielle  $\mathbf{x} \stackrel{i.i.d.}{\sim} \text{Gamma}(1, 1/\delta)$  et que la loi a priori sur le

paramètre est la loi  $\text{Gamma}(a, 1/b)$  alors la loi a posteriori est donnée par

$$\delta|\mathbf{x} \sim \text{Gamma}\left(n + a, \frac{1}{b + \sum_{i=1}^n x_i}\right).$$

La [Figure 6](#) montre la concentration de la loi a posteriori lorsque la taille de l'échantillon augmente. Les hypothèses sont les suivantes:  $\mathbf{x} \sim \text{Gamma}(1, 1/3)$  et  $\delta \sim \text{Gamma}(0.1, 10)$ .

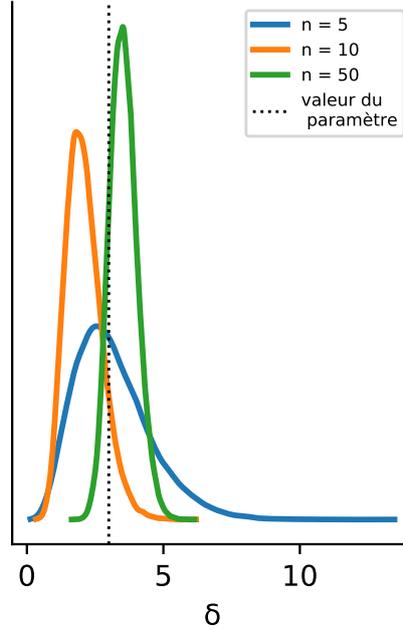


Figure 6 – Histogramme de la loi à posteriori de  $\delta|\mathbf{x}$  lorsque  $\mathbf{x} \sim \text{Gamma}(1, 1/3)$  et  $\delta \sim \text{Gamma}(0.1, 10)$  pour une taille d'échantillon croissante: (bleu)  $n = 5$ , (orange)  $n = 10$ , et (vert)  $n = 50$ .

La situation dans laquelle la loi a posteriori est connue explicitement se limite majoritairement aux lois appartenant aux familles exponentielles naturelles, c.f. [Diaconis and Ylvisaker \[1979\]](#). Dans la pratique, l'échantillonnage de la loi a posteriori implique le recours à des techniques de simulation. Les algorithmes de type *Markov Chain Monte Carlo* (**MCMC**) construisent une chaîne de Markov dont la loi stationnaire est la loi a posteriori. Deux techniques: la marche aléatoire de Metropolis-Hasting et l'échantillonneur de Gibbs, sont détaillées dans la [Section 3.1](#) et la [Section 3.2](#) respectivement avant d'être appliquées aux modèles composites dans la [Section 3.3](#).

### 3.1 La marche aléatoire de Metropolis-Hasting

La marche aléatoire de Metropolis-Hasting construit une suite  $(\theta^i)_{i \geq 0}$  en proposant une nouvelle valeur du paramètre  $\theta^i$  à partir d'une loi de proposition  $K_H(\cdot|\theta^i)$  qui dépend de la valeur courante du paramètre  $\theta^i$  et d'un paramètre  $h > 0$  qui caractérise l'amplitude des mouvements. Une proposition  $\theta^* \sim K_H(\cdot|\theta^i)$  est acceptée avec probabilité

$$\begin{aligned} \alpha(\theta^i, \theta^*) &= \max\left[1, \frac{\pi(\theta^*|\mathbf{x})K_H(\theta^i|\theta^*)}{\pi(\theta^i|\mathbf{x})K_H(\theta^*|\theta^i)}\right] \\ &= \max\left[1, \frac{L(\mathbf{x}|\theta^*)\pi(\theta^*)K_H(\theta^i|\theta^*)}{L(\mathbf{x}|\theta^i)\pi(\theta^i)K_H(\theta^*|\theta^i)}\right], \end{aligned} \quad (10)$$

et dans ce cas  $\theta^i = \theta^*$ , sinon  $\theta^i = \theta^i$ . Cette procédure d'acceptation-rejet génère la suite  $(\theta^i)_{i \geq 0}$  comme la trajectoire d'une chaîne de Markov dont la distribution limite correspond à la loi cible  $\pi(\theta|\mathbf{x})$ . A noter que cette technique s'affranchit de la connaissance de la constante de normalisation qui n'apparaît pas dans l'expression de la probabilité d'acceptation (10). Les paramètres de l'algorithme sont la valeur initiale  $\theta^{(0)}$ , la longueur de la trajectoire  $I$  (soit le nombre d'itération), la longueur de la période d'entraînement  $I^* < I$  (rang en dessous duquel les valeurs de  $(\theta^i)_{i \geq 0}$  ne sont pas considérées, il s'agit d'une période d'attente nécessaire avant que la chaîne de Markov n'atteigne la stationarité) et le paramètre  $h$  de la loi de proposition. Le choix de cette loi se porte généralement vers la loi normale multivariée avec

$$K_H(\cdot|\theta) \sim \text{Normal}(\mu = \theta, \Sigma = H),$$

où  $H$  est une matrice de dimension  $d$  (dimension de  $\theta$ ). La symétrie de la loi normale permet de ré-écrire la probabilité d'acceptation (10) avec

$$\alpha(\theta^i, \theta^*) = \max \left[ 1, \frac{L(\mathbf{x}|\theta^*)\pi(\theta^*)}{L(\mathbf{x}|\theta^i)\pi(\theta^i)} \right],$$

et explique l'appellation "marche aléatoire". La procédure est résumée dans l'Algorithme 1, où  $\text{Unif}(a, b)$  désigne la loi uniforme sur l'intervalle  $[a, b]$ . Le paramètre  $H$ , idéalement, doit refléter la variance et la covariance de la

---

**Algorithme 1** Marche aléatoire de Metropolis-Hasting pour échantillonner  $\pi(\theta|\mathbf{x})$

---

```

1: Initialisation  $\theta^{(0)}$ 
2: for  $i = 0 \rightarrow (I - 1)$  do
3:   Simule  $\theta^* \sim \text{Normal}(\mu = \theta^i, \Sigma = h \cdot \mathbf{I}_d)$ 
4:   Simule  $U \sim \text{Unif}(0, 1)$ 
5:   if  $U \leq \alpha(\theta^i, \theta^*)$  then
6:     Accepte  $\theta^i \leftarrow \theta^*$ 
7:   else
8:     Rejette  $\theta^i \leftarrow \theta^i$ 
9:   end if
10: end for
11: Résultat  $\theta^{(I^*)}, \dots, \theta^I$ 

```

---

loi a posteriori au facteur  $2.38/\sqrt{d}$  près, c.f. Roberts et al. [1997, Corollaire 1.2]. Comme cette information est inconnue, il est possible, en guise d'alternative, de calibrer  $H$  pour avoir un taux d'acceptation asymptotique de 0.234. Ces réglages optimisent la vitesse de convergence de la chaîne de Markov vers sa distribution limite. L'efficacité de l'échantillonnage est souvent diagnostiquée à l'aide d'outils graphiques. La Figure 7 montre la suite des valeurs successives de la chaîne de Markov, et l'histogramme des valeurs échantillonnées avec la densité de la loi a posteriori comme référence pour différentes valeurs de  $h$ . La paramétrisation  $h = 2.8$  conduit à un taux d'acceptation proche de 23.4%. La trace sur la Figure 7b est celle d'une suite stationnaire, quant à l'histogramme il se confond avec la densité de la loi a posteriori. Choisir  $h = 0.01$  conduit à un taux d'acceptation trop important et une convergence lente vers la vraie valeur du paramètre tandis que prendre  $h = 30$  conduit à un taux d'acceptation trop faible avec de grandes variations autour de la valeur du paramètre.

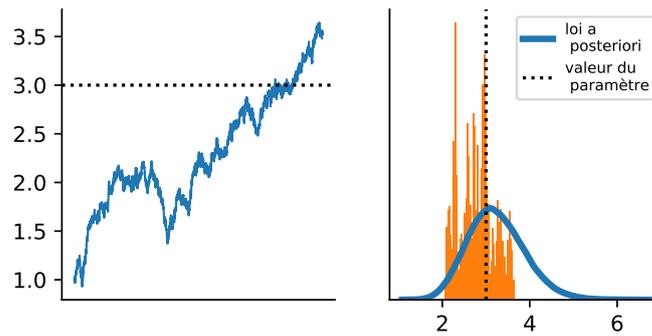
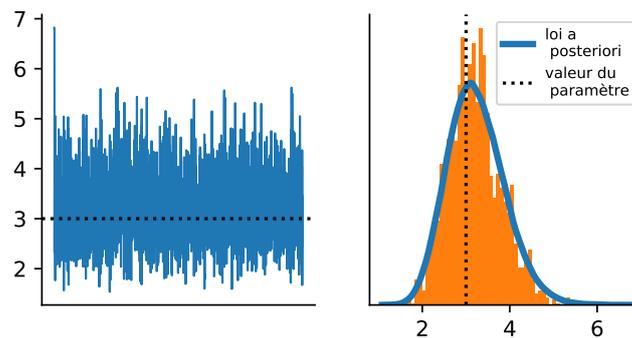
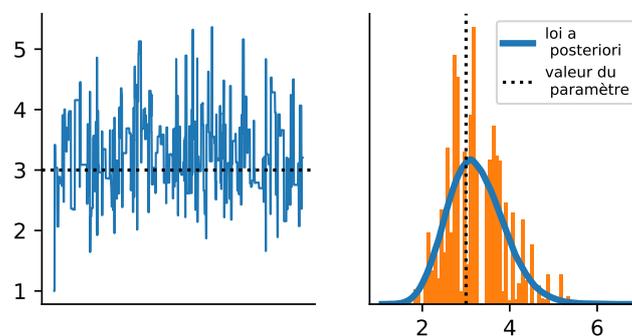
(a)  $h = 0.01$ (b)  $h = 2.8$ (c)  $h = 30$ 

Figure 7 – Trace et histogramme des valeurs du paramètre échantillonnées via la marche aléatoire de Metropolis-Hasting.

L'échantillonnage via l'algorithme 1 perd en efficacité (caractérisée par la vitesse de convergence vers la loi limite) lorsque la dimension  $d$  du paramètre augmente. Cet écueil est contourné par l'utilisation d'une procédure connue sous le nom d'échantillonneur de Gibbs qui fait l'objet de la section suivante.

### 3.2 L'échantillonneur de Gibbs

L'échantillonneur de Gibbs permet d'échantillonner une loi multivariée  $\pi(\theta|\mathbf{x})$ , avec  $\theta = (\theta_1, \dots, \theta_d)$  en simulant depuis ses lois conditionnelles univariées (dites marginales) définies par

$$\pi(\theta_j|\mathbf{x}, \theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_d), \quad j = 1, \dots, d.$$

Il s'agit encore d'une procédure de type MCMC qui conduit à générer une suite  $(\theta^i)_{i \geq 0}$  comme la trajectoire d'une chaîne de Markov convergeant vers la loi cible  $\pi(\theta|\mathbf{x})$ . Etant donnée la valeur courante du paramètre  $\theta^i$ , la mise à jour menant à  $\theta^{i+1}$  s'effectue composante par composante en commençant par la première avec

$$\theta_1^{i+1} \sim \pi(\cdot|\mathbf{x}, \theta_2^i, \dots, \theta_d^i), \quad (11)$$

puis la deuxième avec

$$\theta_2^{i+1} \sim \pi(\cdot|\mathbf{x}, \theta_1^{i+1}, \theta_3^i, \dots, \theta_d^i), \quad (12)$$

et ainsi de suite. La suite de valeur  $\theta^i = (\theta_1^i, \dots, \theta_d^i)$  pour  $i \geq 0$  est la réalisation d'une chaîne de Markov sur l'espace d'état  $\Theta$  dont la loi limite est la loi a posteriori  $\pi(\theta|\mathbf{x})$ . Les lois conditionnelles marginales de la loi a posteriori (11) et (12) ne sont généralement pas connues explicitement mais à une constante multiplicative près. L'algorithme de simulation de Metropolis-Hasting, présenté dans la section précédente, est alors utilisé (*Metropolis-Hasting within Gibbs*). En effet, si les lois a priori sur chacun des paramètres sont indépendantes (hypothèse très courante) alors la loi conditionnelle de la composante  $j$  s'écrit

$$\pi(\theta_j|\mathbf{x}, \theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_d) = \frac{L(\mathbf{x}|\theta)\pi(\theta_j)}{\int_{\mathbb{R}} L(\mathbf{x}|\theta)\pi(\theta_j)d\theta_j}.$$

La loi a posteriori des paramètres des modèles composites présentés dans la [Section 2.2](#) est obtenue via l'échantillonneur de Gibbs dans la section suivante.

### 3.3 Application aux modèles composites

Les modèles  $\text{Gamma}(r) - \text{Par}(\alpha, \theta)$ ,  $\text{LogNorm}(\sigma) - \text{Par}(\alpha, \theta)$ , et  $\text{Weib}(r) - \text{Par}(\alpha, \theta)$  sont à nouveau calibrés sur les données de pertes danoises en utilisant les méthodes d'inférence Bayésienne. Les hypothèses a priori pour la partie attritionnelle sont données par

$$r \sim \text{Gamma}(0.1, 10), \quad \sigma \sim \text{Gamma}(0.1, 10), \quad k \sim \text{Gamma}(0.1, 10),$$

Les hypothèses a priori pour la partie extrême sont les mêmes pour chacun des modèles

$$\alpha \sim \text{Gamma}(0.1, 10), \quad \theta \sim \text{Gamma}(0.1, 10).$$

Le choix de la loi gamma comme loi a priori est très courant pour des paramètres positifs, le choix des hyperparamètres confère à la loi a priori un caractère "vague" ou "non-informatif" du fait de sa variance importante. La mise à jour de la chaîne de Markov procède par des marches aléatoires de Metropolis-Hasting univariées pour chacun des paramètres. La paramétrisation de  $H$  est délicate lorsque le nombre de paramètres est plus grand que 1, elle est effectuée en ciblant un taux d'acceptation de 23.4%. La [Figure 8](#) montre la densité des lois a posteriori associées aux paramètres des modèles composites  $\text{Gamma}(r) - \text{Par}(\alpha, \theta)$ ,  $\text{LogNorm}(\sigma) - \text{Par}(\alpha, \theta)$ , et  $\text{Weib}(r) - \text{Par}(\alpha, \theta)$ . Les lois a posteriori sont très concentrées et coïncident avec l'estimateur du maximum de

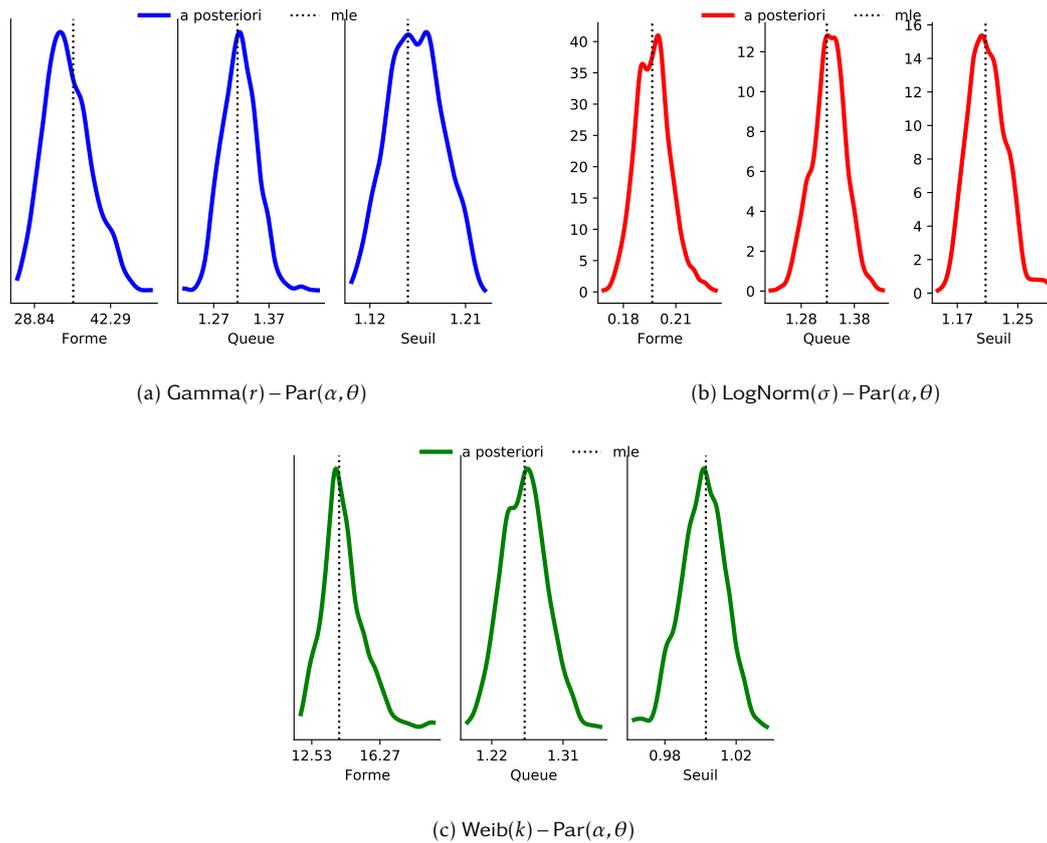


Figure 8 – Loi a posteriori des paramètres des modèles composites calibrés sur les données de pertes "incendie" danoises.

vraisemblance. Cela s'explique par la taille de l'échantillon qui est relativement grande (avec 2,492 observations) en cohérence avec les résultats théoriques obtenus par [Bunke and Milhaud \[1998\]](#). La loi a posteriori, lorsque le modèle est mal spécifié (les données ont très peu de chance d'avoir été générées par un des modèles composites considérés), se concentre sur les régions de  $\Theta$  qui minimise la distance de Kullback-Leibler tout comme l'estimateur du maximum de vraisemblance. L'inférence Bayésienne est plutôt utilisée dans des situations où peu de données sont disponibles. Il est alors intéressant d'avoir une distribution autour de chacun des paramètres pour prendre en compte l'incertitude et compenser ce manque de données par un choix informé de la loi a priori. Dans la situation considérée, la valeur ajoutée d'avoir une distribution plutôt qu'un estimateur ponctuel est assez faible.

La question de la sélection du modèle le plus adapté aux données fait l'objet de la section suivante et nécessite l'évaluation de la constante de normalisation définie dans l'équation (9). L'algorithme de Monte Carlo séquentiel présenté dans la [Section 4](#) permet d'échantillonner la loi a posteriori et de retourner une approximation de la constante de normalisation.

## 4 Sélection Bayésienne d'un modèle composite

Soit un ensemble  $\mathcal{M} = \{m_1, \dots, m_J\}$  de modèle en concurrence. Le choix du meilleur modèle en statistique bayésienne repose sur la définition d'une variable aléatoire  $M$  à valeur dans l'ensemble des modèles  $\mathcal{M}$  associée à une loi de probabilité a priori vérifiant

$$\mathbb{P}(M = m_j) = \pi(m_j) \geq 0, \text{ pour } j = 1, \dots, J, \text{ et } \sum_{j=1}^J \pi(m_j) = 1.$$

Cette loi a priori, souvent choisie telle que

$$\pi(m_j) = 1/J, \quad j = 1, \dots, J,$$

est mise à jour via la fonction de vraisemblance des données  $L(\mathbf{x}|m_j)$  associée à chaque modèle  $m_j \in \mathcal{M}$ , pour obtenir la loi a posteriori de  $M$  donnée par

$$\pi(m_j|\mathbf{x}) = \frac{L(\mathbf{x}|m_j)\pi(m_j)}{\sum_{i=1}^J L(\mathbf{x}|m_i)\pi(m_i)}, \quad j = 1, \dots, J. \quad (13)$$

Chaque modèle  $m$  possède un jeu de paramètres de loi a priori  $\pi(\theta|m)$ . La vraisemblance des données suivant un modèle  $m \in \mathcal{M}$  est obtenue par intégration contre la loi a priori  $\pi(\theta|m)$ , avec

$$L(\mathbf{x}|m) = \int_{\Theta} L(\mathbf{x}|m, \theta)\pi(\theta|m)d\theta,$$

et coïncide exactement avec la constante de normalisation (9). Le modèle associé à la plus grande probabilité a posteriori  $\pi(m|\mathbf{x})$  est ensuite choisi. Si le calcul de la constante de normalisation peut être évité pour l'inférence des paramètres des différents modèles, il est incontournable pour la sélection du meilleur modèle. Cette section présente l'algorithme de Monte Carlo séquentiel qui permet d'inférer les paramètres du modèle tout en approximant la constante de normalisation par simulation.

La [Section 4.1](#) rappelle le principe de l'échantillonnage préférentiel nécessaire à la compréhension de l'échantillonneur de Monte Carlo séquentiel **SMC** présenté dans la [Section 4.2](#). Cet algorithme, au coeur de ce travail de recherche, permet simultanément d'échantillonner la loi a posteriori et de retourner une estimation de la constante de normalisation. La [Section 4.3](#) présente les critères d'information Bayésiens. L'algorithme **SMC** est enfin appliqué au jeu de données des pertes "incendie" danoises dans la [Section 4.4](#).

### 4.1 Echantillonnage préférentiel

En inférence bayésienne, une estimation ponctuelle du paramètre  $\theta$  peut être obtenue en calculant l'espérance

$$\mathbb{E}_{\pi(\theta|\mathbf{x})}(\varphi) = \int_{\Theta} \varphi(\theta)\pi(\theta|\mathbf{x})d\theta, \quad (14)$$

d'une application mesurable  $\varphi$ . La moyenne a posteriori correspond au cas où  $\varphi(\theta) = \theta$ . L'espérance (14) est évaluée en pratique via son approximation de Monte Carlo

$$\mathbb{E}_{\pi(\theta|\mathbf{x})}(\varphi) \approx \frac{1}{N} \sum_{i=1}^N \varphi(\theta_i), \quad (15)$$

où  $\theta_1, \dots, \theta_N$  est un échantillon i.i.d. suivant la loi a posteriori et généré par un algorithme **MCMC** vu dans la [Section 3](#). L'échantillonnage préférentiel consiste à simuler des valeurs suivant une loi de proposition  $g(\theta)$  qui

est plus simple à échantillonner que  $\pi(\theta|\mathbf{x})$  ou qui permet d'améliorer l'estimateur (15) dont l'erreur est donnée par  $\mathbb{V}_{\pi(\theta|\mathbf{x})}(\varphi)/N$ . Le principe de l'échantillonnage préférentiel repose sur l'identité suivante

$$\begin{aligned}\mathbb{E}_{\pi(\theta|\mathbf{x})}(\varphi) &= \int_{\Theta} \varphi(\theta)\pi(\theta|\mathbf{x})d\theta \\ &= \int_{\Theta} \varphi(\theta)\frac{L(\mathbf{x}|\theta)\pi(\theta)}{Z(\mathbf{x})}d\theta \\ &= Z(\mathbf{x})^{-1} \int_{\Theta} \varphi(\theta)\frac{L(\mathbf{x}|\theta)\pi(\theta)}{g(\theta)}g(\theta)d\theta \\ &= Z(\mathbf{x})^{-1} \int_{\Theta} \varphi(\theta)w(\theta)g(\theta)d\theta \\ &= Z(\mathbf{x})^{-1}\mathbb{E}_g(\varphi \cdot w),\end{aligned}$$

où  $w(\theta) = L(\mathbf{x}|\theta)\pi(\theta)/g(\theta)$  désigne une fonction de poids non-normalisée. En prenant  $\varphi(\theta) = 1$ , il est possible de ré-écrire la constante de normalisation comme une espérance

$$Z(\mathbf{x}) = \mathbb{E}_g(w),$$

et de l'approcher par

$$Z(\mathbf{x}) \approx \frac{1}{N} \sum_{i=1}^N w(\tilde{\theta}_i),$$

où  $\tilde{\theta}_1, \dots, \tilde{\theta}_N$  est un échantillon i.i.d. de loi  $g$ . Une approximation alternative à (15) est alors donnée par

$$\mathbb{E}_{\pi(\theta|\mathbf{x})}(\varphi) \approx \sum_{i=1}^N W_i \varphi(\tilde{\theta}_i),$$

où

$$W_i = \frac{w(\tilde{\theta}_i)}{\sum_{j=1}^N w(\tilde{\theta}_j)}, \quad i = 1, \dots, N.$$

Le résultat de la procédure d'échantillonnage préférentiel est un ensemble de points pondérés  $\{(W_i, \tilde{\theta}_i), i = 1, \dots, N\}$ , appelé nuage de particules dans la suite, dont la distribution empirique approche la loi a posteriori au sens où

$$\sum_{i=1}^N W_i \varphi(\tilde{\theta}_i) \rightarrow \mathbb{E}_{\pi(\theta|\mathbf{x})}(\varphi), \quad \text{pour } N \rightarrow \infty,$$

pour toute application mesurable  $\varphi$ . Le problème est alors de choisir la loi de proposition  $g$ . La qualité d'une telle loi pour reproduire la loi cible est mesurée via la taille de l'échantillon effective (*Effective Sample Size*) estimée, c.f. Kong et al. [1994], par

$$\text{ESS} \approx \frac{1}{\sum_{i=1}^N W_i^2}.$$

Il s'agit d'un indicateur compris entre 1 et  $N$  qui correspond à la taille de l'échantillon i.i.d. de même variance que la distribution empirique du nuage de particules  $\{(W_i, \tilde{\theta}_i), i = 1, \dots, N\}$ . Il caractérise la dégénérescence du nuage de particules. Par exemple, un ESS proche de 1 correspond à une situation où la plupart des particules ont un poids proche de 0.

Le problème central de l'échantillonnage préférentiel réside dans le choix d'une loi de proposition  $g$  adaptée. L'algorithme **SMC**, présenté ci-après, s'affranchit de ce choix en construisant une suite de distributions intermédiaires qui converge vers la loi a posteriori tout en maintenant la taille de l'échantillon effective à un certain niveau.

## 4.2 Algorithme de Monte Carlo séquentiel

L'idée de l'algorithme de Monte Carlo séquentiel est de construire une suite de distribution  $\pi_s(\theta|\mathbf{x})$ ,  $s = 0, \dots, t$ . La suite est initialisée par la loi a priori  $\pi_0(\theta|\mathbf{x}) = \pi(\theta)$ , mise à jour progressivement par la construction des distributions intermédiaires pour converger finalement vers la loi cible, la distribution a posteriori  $\pi_t(\theta|\mathbf{x}) = \pi(\theta|\mathbf{x})$ . Deux constructions de la suite  $\pi_s(\theta|\mathbf{x})$ ,  $s = 0, \dots, t$  sont étudiées dans ce travail. La première consiste à intégrer les données au fur et à mesure des itérations avec

$$\pi_s(\theta|\mathbf{x}) = \frac{L(\mathbf{x}_{1:n_s}|\theta)\pi(\theta)}{Z_s}, \quad s = 0, \dots, t, \quad (16)$$

où  $n_s$ ,  $s = 0, \dots, t$  est une suite d'entiers vérifiant  $0 = n_0 < n_1 < \dots < n_t = n$ , la constante de normalisation est donnée par

$$Z_s = \int_{\Theta} L(\mathbf{x}_{1:n_s}|\theta)\pi(\theta)d\theta.$$

et  $\mathbf{x}_{1:n_s} = (x_1, \dots, x_{n_s})$  désigne l'échantillon  $\mathbf{x}$  tronqué à l'indice  $n_s$ . Cette approche a été introduite dans le travail de [Chopin \[2002\]](#). La deuxième approche procède d'une activation progressive de la vraisemblance avec

$$\pi_s(\theta|\mathbf{x}) = \frac{L(\mathbf{x}|\theta)^{\tau_s}\pi(\theta)}{Z_s}, \quad s = 0, \dots, t, \quad (17)$$

où

$$Z_s = \int_{\Theta} L(\mathbf{x}|\theta)^{\tau_s}\pi(\theta)d\theta.$$

et  $\tau_k$ ,  $k = 0, \dots, n$  est une suite numérique (dont les éléments sont appelés "températures") telle que  $0 = \tau_0 < \tau_1 < \dots < \tau_n = 1$ . Cette approche par recuit simulé s'inspire des travaux de [Neal \[2001\]](#). Les distributions intermédiaires  $\pi_s$ ,  $s = 0, \dots, n$  sont approchées par la distribution empirique de systèmes de particules  $\{(W_i^s, \theta_i^s), i = 1, \dots, N\}$  (similaire à ceux introduit pour l'échantillonnage préférentiel dans la [Section 4.1](#)) qui évoluent au fur et à mesure des itérations via un procédé de repondération, sélection et perturbation. Le nuage de particules est initialisé suivant la loi a priori avec

$$\theta_i^{(0)} \stackrel{\text{i.i.d.}}{\sim} \pi(\theta), \text{ et } W_i^{(0)} = \frac{1}{N}, \text{ pour } i = 1, \dots, N.$$

Le nuage de particules  $\{(W_i^{s+1}, \theta_i^{s+1}), i = 1, \dots, N\}$ , est obtenu à partir du nuage de particules précédent  $\{(W_i^s, \theta_i^s), i = 1, \dots, N\}$  en 3 étapes.

1. Repondération: Cette première étape prépare le nuage de particules courant  $\{(W_i^s, \theta_i^s), i = 1, \dots, N\}$  pour cibler la distribution suivante  $\pi_{s+1}$ . les particules sont repondérées par

$$W_i^{s+1} \propto w_i^{s+1} = \frac{\pi_{s+1}(\theta_i^s)}{\pi_s(\theta_i^s)}, \quad i = 1, \dots, N$$

où  $\propto$  signifie "proportionnel à" et les  $w_i^{s+1}$ ,  $i = 1, \dots, N$  désigne les poids non-normalisés qui seront utiles à l'évaluation de la constante de normalisation. Les poids  $W_i^s$  ainsi recalculés correspondent à ceux introduits pour l'échantillonnage préférentiel dans la [Section 4.1](#). Ils doivent donc vérifier

$$\text{ESS} \approx \frac{1}{\sum_{i=1}^N (W_i^{s+1})^2} \geq \rho N,$$

où  $0 < \rho \leq 1$ . La distribution cible  $\pi_{s+1}$  est choisie de façon à vérifier cette condition de non-dégénérescence du nuage de particules. La taille de l'échantillon  $n_{s+1}$  dans l'équation (16) ou la température  $\tau_{s+1}$  dans l'équation (17) sont obtenus via un algorithme de recherche dichotomique. En pratique, le seuil choisi est  $\rho = 1/2$ , c.f. [Jasra et al. \[2010\]](#).

2. Sélection: La deuxième étape consiste en un tirage aléatoire des particules  $\theta_1^s, \dots, \theta_N^s$  suivant leur poids mis à jour  $W_1^{s+1}, \dots, W_N^{s+1}$ . De nouvelles particules  $\tilde{\theta}_1^s, \dots, \tilde{\theta}_N^s$  sont émises telles que

$$\tilde{\theta}_j^s = \theta_i^s \text{ avec probabilité } W_i^s, \text{ pour } 1 \leq i, j \leq N.$$

3. Perturbation: Les nouvelles particules  $\theta_i^{s+1}$ ,  $i = 1, \dots, N$  sont issues de l'application successive de transitions du type Gibbs-Metropolis-Hasting sur les particules sélectionnées  $\tilde{\theta}_i^s$ ,  $i = 1, \dots, N$ . La marche aléatoire de Metropolis-Hasting, décrite en [Section 3.1](#) cible la distribution  $\pi_{s+1}$ , le paramètre  $H$  est donné par  $\widehat{\Sigma} \cdot 2.38/\sqrt{d}$ , où  $\widehat{\Sigma}$  désigne la matrice de variance-covariance empirique du système de particules repondérées  $\{(W_i^{s+1}, \theta_i^s), i = 1, \dots, N\}$ . Le nombre de transitions  $k$  à appliquer est choisi pour s'assurer que les particules sont effectivement perturbées. En pratique, une transition est appliquée à chacune des particules dans le cadre d'un tour d'essai permettant l'estimation de la probabilité d'acceptation  $\widehat{p}_a$ . Le nombre de transitions est ensuite calibré avec

$$k = \max \left\{ k_{\max}, \min \left[ k_{\min}, \frac{\log(1-c)}{\log(1-\widehat{p}_a)} \right] \right\},$$

où  $k_{\min}$  et  $k_{\max}$  sont des nombres de transition minimum et maximum,  $c \in (0, 1)$  est la probabilité que chacune des particules soit perturbée au moins une fois. Les nouvelles particules  $\theta_1^{s+1}, \dots, \theta_N^{s+1}$  sont distribuées suivant  $\pi_{s+1}$  et leur poids respectif est donné par  $W_i^{s+1} = 1/N$  pour  $i = 1, \dots, N$ .

Le choix adaptatif de la distribution cible lors de l'étape 1, l'utilisation de la matrice de variance-covariance pour calibrer la marche aléatoire de Metropolis-Hasting et la calibration du nombre de transitions à effectuer dans l'étape 3 sont des astuces algorithmiques standards des échantillonneurs de Monte Carlo séquentiels comme celui décrit dans l'article de [South et al. \[2019\]](#) ou implémenté dans le module `pymc` [Salvatier et al. \[2016\]](#) de python. La procédure d'échantillonnage par Monte Carlo séquentiel est résumée dans l'algorithme 2. L'échantillonneur de Monte Carlo séquentiel a plusieurs avantages sur l'échantillonneur de Gibbs. L'algorithme **SMC** ne nécessite pas une initialisation arbitraire. L'échantillonneur de Gibbs peut-être vu comme un algorithme d'optimisation stochastique dont la convergence vers un optimum global est fonction de l'initialisation. Par exemple l'exploration d'une distribution a posteriori multimodale est difficile si l'algorithme se concentre sur un seul des modes. L'algorithme **SMC** exploite l'information contenue dans les nuages de particules représentant les distributions intermédiaires pour calibrer de manière optimale la marche aléatoire de Metropolis-Hasting. L'étape de perturbation de l'algorithme **SMC** est très facile à paralléliser ce qui permet d'améliorer les temps de calcul. Enfin, la transition par des distributions intermédiaires permet une estimation de la constante de normalisation justifiée par la théorie de l'échantillonnage préférentiel. En effet, les poids non-normalisés

$$w_i^s = \frac{\pi_s(\theta_i^{(s)})}{\pi_{s-1}(\theta_i^{(s)})}, \quad i = 1, \dots, N.$$

de l'itération  $s = 1, \dots, t$  fournissent une estimation du ratio des constantes de normalisations  $Z_s/Z_{s-1}$  avec

$$\frac{Z_s}{Z_{s-1}} \approx \frac{1}{N} \sum_{i=1}^N w_i^s.$$

La constante de normalisation de la loi a posteriori s'en déduit via

$$Z(\mathbf{x}) = Z_t = \prod_{s=1}^t \frac{Z_s}{Z_{s-1}} \approx \prod_{s=1}^t \left( \frac{1}{N} \sum_{i=1}^N w_i^s \right).$$

**Algorithme 2** Monte Carlo Séquentiel pour échantillonner  $\pi(\theta|\mathbf{x})$ 

- 
- 1: Paramètres  $\rho \in (0, 1)$ ;  $k_{\min} \in \mathbb{N}$ ;  $k_{\max} \in \mathbb{N}$ ;  $c \in (0, 1)$
  - 2: Initialisation  $s \leftarrow 0$ ;  $\pi_0(\theta) \leftarrow \pi(\theta)$ ;
  - 3: **for**  $i = 1 \rightarrow N$  **do**
  - 4:    $\theta_i^0 \sim \pi(\theta)$ ;  $W_i^0 \leftarrow 1/N$
  - 5: **end for**
  - 6: **while**  $\pi_s(\theta) \neq \pi(\theta|\mathbf{x})$  **do**
  - 7:   Recherche  $\pi_{s+1}$  telle que
 
$$\frac{1}{\sum_{i=1}^N (W_i^{s+1})^2} \geq \rho N, \text{ avec } W_i^{s+1} \propto w_i^{s+1} = \pi_{s+1}(\theta_i^s) / \pi_s(\theta_i^s), i = 1, \dots, N$$
  - 8:   Calcule  $\widehat{\Sigma} = \text{Cov}(\{(W_i^{s+1}, \theta_i^s), i = 1, \dots, N\})$
  - 9:   **for**  $i = 1 \rightarrow N$  **do**
  - 10:     Sélectionne  $\tilde{\theta}_i \sim \{\theta_1^{(s)}, \dots, \theta_N^{(s)}\}$  avec probabilité  $W_j^{s+1}$ , pour  $1 \leq j \leq N$
  - 11:   **end for**
  - 12:   **for**  $i = 1 \rightarrow N$  **do**
  - 13:      $\tilde{\theta}_i^* \leftarrow K_H(\tilde{\theta}_i, \cdot)$  où  $K_H(\tilde{\theta}_i, \cdot)$  correspond à une perturbation du type Gibbs-Metropolis-Hasting avec
 
$$H = \frac{2.38}{\sqrt{d}} \cdot \widehat{\Sigma}$$
  - 14:   **end for**
  - 15:   Calcule  $p_a = N^{-1} \sum_{i=1}^N \mathbb{I}_{\tilde{\theta}_i^* = \tilde{\theta}_i}$ ;  $k = \max\{k_{\max}, \min\left[k_{\min}, \frac{\log(1-c)}{\log(1-p_a)}\right]\}$
  - 16:   **for**  $i = 1 \rightarrow N$  **do**
  - 17:      $\theta_i^{s+1} \leftarrow K_H^{*(k-1)}(\tilde{\theta}_i^*, \cdot)$  où  $K_H^{*(k-1)}(\tilde{\theta}_i^*, \cdot)$  correspond à  $k-1$  perturbations du type Gibbs-Metropolis-Hasting
  - 18:      $W_i^{s+1} \leftarrow 1/N$
  - 19:   **end for**
  - 20: **end while**
  - 21: Résultat  $(W_1^t, \theta_1^t), \dots, (W_N^t, \theta_N^t)$
- 

Dans l'exemple 4, où les données suivent une loi exponentielle  $\mathbf{x} \stackrel{\text{i.i.d.}}{\sim} \text{Gamma}(1, 1/\delta)$  et la loi a priori sur le paramètre est la loi  $\text{Gamma}(a, 1/b)$ , la constante de normalisation est donnée par

$$Z(\mathbf{x}) = \frac{b^a}{\left(\sum_{i=1}^n x_i + b\right)^{a+n}} \prod_{i=1}^n (a+i). \quad (18)$$

Cet exemple permet d'apprécier la précision de l'estimateur de la constante de normalisation par l'échantillonneur de Monte Carlo séquentiel. Les paramètres de l'algorithme sont les suivants:

$$\rho = \frac{1}{2}; k_{\min} = 2; k_{\max} = 25; c = 0.99.$$

Les données et les paramètres de la loi a priori sont similaires à ceux considérés dans la Section 3 avec  $\mathbf{x} \sim \text{Gamma}(1, 1/3)$  et  $\delta \sim \text{Gamma}(0.1, 10)$ . La précision est mesurée pour un nombre croissant de particules  $N \in \{500, 2000, 5000\}$ . La constante normalisation est évaluée 100 fois par l'algorithme SMC pour chacune

des méthodes de définitions des distributions intermédiaires (16) et (17). La Figure 9 montre l'histogramme des estimateurs de la constante de normalisation pour 100 répétitions de la procédure d'approximation en fonction du nombre de particules et du mode de définition des distribution intermédiaires utilisé. Le tableau

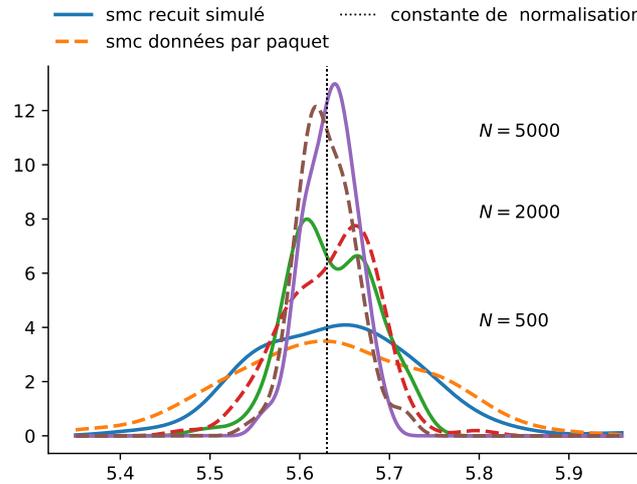


Figure 9 – Histogramme de l'estimateur par Monte Carlo Séquentielle de la constante de normalisation pour des données de loi exponentielle  $x \sim \text{Gamma}(1, 1/3)$  et de loi a priori  $\delta \sim \text{Gamma}(0.1, 10)$  en fonction du nombre de particules et de la méthode utilisée.

2 reporte l'écart-type des estimateurs de la constante de normalisation pour 100 répétitions de la procédure d'approximation en fonction du nombre de particules et du mode de définition des distribution intermédiaires utilisé. La précision de l'estimateur augmente avec le nombre de particules qui est en lien direct avec le temps

	Recuit simulé	Données par paquet
N=500	0.085348	0.098087
N=2000	0.042403	0.047473
N=5000	0.031853	0.031469

Table 2 – Ecart-types des estimateurs par Monte Carlo Séquentielle de la constante de normalisation pour des données de loi exponentielle  $x \sim \text{Gamma}(1, 1/3)$  et de loi a priori  $\delta \sim \text{Gamma}(0.1, 10)$  en fonction du nombre de particules et de la méthode utilisée.

de calcul. L'idée est donc de choisir le nombre de particules maximum pour un budget de calcul donné. La section suivante présente les critères d'information en statistique bayésienne en complément de la probabilité a posteriori des modèles.

### 4.3 Les critères d'informations Bayésiens

Deux critiques sont adressées à la sélection de modèle basée sur la probabilité a posteriori. Le premier reproche est la sensibilité de la constante de normalisation au choix de la loi a priori. Le critère BIC, introduit dans la Section 2.2, est une approximation du log de la vraisemblance marginale valide pour une taille d'échantillon très grande. Il repose cependant sur l'estimateur du maximum de vraisemblance dont la présence est justifiée

par la convergence des lois a posteriori (sous réserve de vérifier certaines conditions de régularité) vers une loi normale multivariée centrée en l'estimateur du maximum de vraisemblance lorsque la taille de l'échantillon grandit et que le modèle est bien spécifié, voir [Hong and Martin \[2017\]](#). Il présente l'avantage de gommer l'effet de la loi a priori dans le choix du modèle. L'équivalent bayésien de l'AIC est le *Deviance Information Criterion*, introduit par [Spiegelhalter et al. \[2002\]](#), car il repose sur un arbitrage entre adéquation et complexité du modèle (correspondant au nombre de paramètres). Il est défini par

$$DIC = -2[l(\mathbf{x}|\tilde{\theta}) - p_{DIC}],$$

où  $\tilde{\theta} = \mathbb{E}_{\pi(\theta|\mathbf{x})}$  est la moyenne a posteriori et  $p_{DIC}$  est un terme de correction qui tend vers le nombre de paramètres du modèle lorsque la taille de l'échantillon augmente. Le terme de correction  $p_{DIC}$  est donné par

$$p_{DIC} = \mathbb{E}_{\pi(\theta|\mathbf{x})}[l(\mathbf{x}|\theta)] - l(\mathbf{x}|\tilde{\theta}) \approx \frac{1}{N} \sum_{i=1}^N l(\mathbf{x}|\theta_i) - l(\mathbf{x}|\tilde{\theta}),$$

où  $\theta_1, \dots, \theta_N$  résulte de l'échantillonnage de la loi a posteriori. Le second reproche est que l'adéquation des modèles soit mesurée sur les données qui ont servi à leur calibration. Il serait préférable de mesurer l'adéquation du modèle sur de nouvelles données, notées  $\mathbf{x}^* = (x_1^*, \dots, x_n^*)$ , et de calculer la vraisemblance de ces données sous la loi a posteriori par exemple via

$$\text{lppd} = \sum_{j=1}^n \log \mathbb{E}_{\pi(\theta|\mathbf{x})}[L(x_j^*|\theta)] \approx \sum_{j=1}^n \log \left[ \frac{1}{N} \sum_{i=1}^N L(x_j^*|\theta_i) \right].$$

où lppd signifie *log pointwise predictive density*. Cette quantité pourrait être estimée via une procédure de validation-croisée mais la ré-estimation du modèle plusieurs fois avec des méthodes bayésiennes est trop coûteuse en temps de calcul. L'idée du *Widely Applicable Information Criterion* (WAIC) est alors de corriger la lppd calculée sur l'échantillon observé. Le WAIC est donné par

$$WAIC = -2 \left\{ \sum_{j=1}^n \log \mathbb{E}_{\pi(\theta|\mathbf{x})}[L(x_j|\theta)] - p_{WAIC} \right\},$$

où la correction  $p_{WAIC}$  est donnée par

$$p_{WAIC} = \mathbb{V}_{\pi(\theta|\mathbf{x})}[l(\mathbf{x}|\theta)],$$

et correspond à la variance sous la loi a posteriori de la log vraisemblance des données. [Watanabe \[2010\]](#) établit l'équivalence asymptotique (pour une grande taille d'échantillon) entre ce critère d'information et l'estimation de la lppd avec une procédure de validation croisée de type *leave-one-out*. [Gelman et al. \[2013a\]](#) propose une discussion sur l'interprétation des critères d'information. Dans le cadre de la présente étude, le DIC et le WAIC seront calculés en complément de la probabilité a posteriori pour comparer les différents modèles.

#### 4.4 Application aux modèles composites

L'application de l'algorithme SMC sur les données danoises représente un défi numérique eu égard au grand nombre d'observations. Le nombre de particules est fixé à 1,000 et les paramètres de l'algorithme sont les mêmes que dans la section précédente avec

$$\rho = \frac{1}{2}; k_{\min} = 2; k_{\max} = 25; c = 0.99.$$

Les hypothèses a priori sont les mêmes que dans la [Section 3.3](#), soit

$$r \sim \text{Gamma}([0.1, 10]), \sigma \sim \text{Gamma}([0.1, 10]), k \sim \text{Gamma}([0.1, 10]),$$

pour la partie attritionnelle, et

$$\alpha \sim \text{Gamma}([0.1, 10]), \theta \sim \text{Gamma}([0.1, 10]),$$

pour la partie extrême. Les lois a posteriori des paramètres pour chacun des trois modèles composites sont données sur la [Figure 10](#). L'estimation des paramètres est similaire à celle obtenue via le maximum de

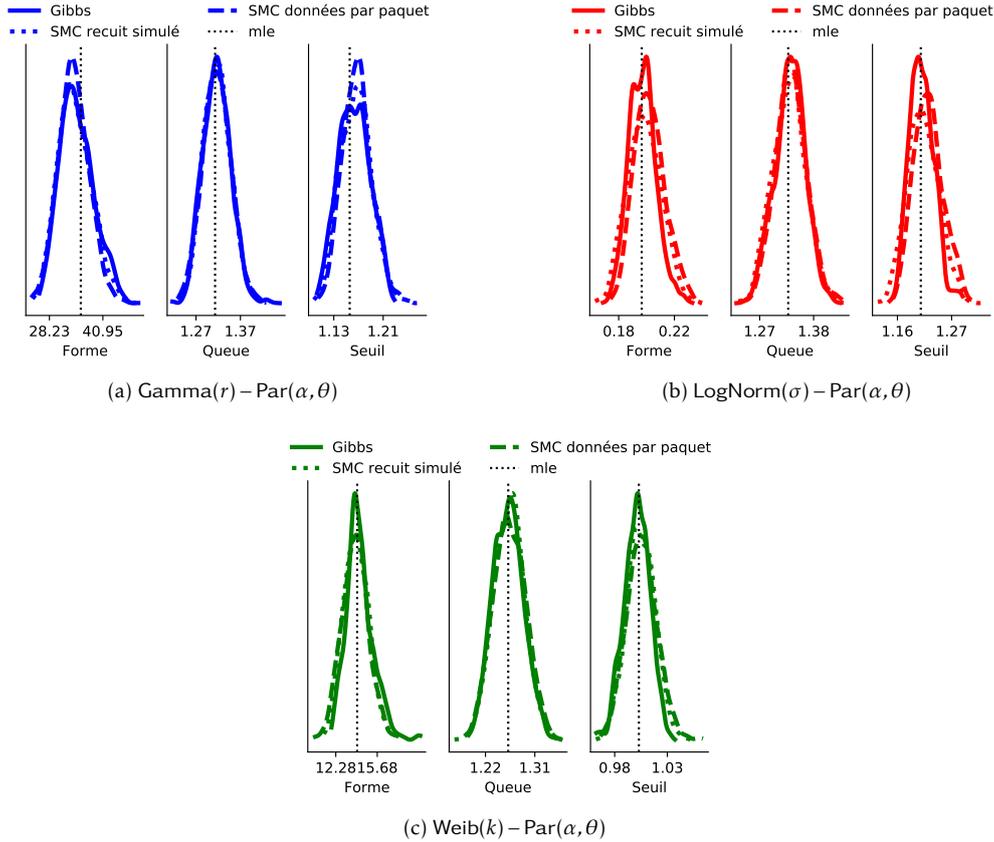


Figure 10 – Loi a posteriori des paramètres des modèles composites calibrés sur les données de pertes "incendie" danoises obtenu via l'échantillonneur de Gibbs (ligne continue) et les échantillonneurs de Monte Carlo séquentiels via les approches d'introduction progressive de la vraisemblance (ligne discontinue) et des données (pointillés).

vraisemblance dans la [Section 2.2](#) et l'échantillonneur de Gibbs dans la [Section 3.3](#). L'algorithme **SMC** fournit une estimation des constantes de normalisation qui permettent l'évaluation des probabilités a posteriori de chacun des modèles en concurrence avec

$$\pi(m_j|\mathbf{x}) = \frac{L(\mathbf{x}|m_j)\pi(m_j)}{\sum_{i=1}^J L(\mathbf{x}|m_i)\pi(m_i)}, \quad j = 1, \dots, J.$$

Le logarithme des constantes de normalisation, les probabilités a posteriori associées à une probabilité a priori égale à 1/3 pour chacun des modèles, et les critères d'information DIC et WAIC sont reportés dans le [tableau 3](#). Des temps de calculs plus importants sont à noter lorsque la définition des distributions intermédiaires repose

Méthode	Modèles	$\log Z(\mathbf{x})$	$\pi(m \mathbf{x})$	DIC	WAIC	Temps
Recuit simulé	LogNorm – Par	–3882.53	0	7725.72	7743.71	242s
	Weib – Par	–3858.50	1	7674.48	7689.55	
	Gamma – Par	–3878.20	0	7711.70	7730.08	
Données par paquet	LogNorm – Par	–3897.10	0	7725.99	7744.63	1156s
	Weib – Par	–3857.55	1	7674.45	7689.83	
	Gamma – Par	–3877.91	0	7711.60	7729.29	

Table 3 – Probabilité à posteriori et critères d’information Bayésiens pour les modèles composites ajustés sur les données de pertes "incendie" danoises.

sur l’intégration progressive des données. La méthode par activation progressive de la vraisemblance (17) se révèle être plus efficace. Son avantage réside dans la nature continue du paramètre de température  $\tau$  qui permet d’atteindre précisément la taille d’échantillon effective cible. L’interprétation des critères d’information bayésien conduit à la conclusion que le modèle composite Weibull-Pareto doit être privilégié, ce qui est cohérent avec les résultats de la Section 2. La section suivante propose d’appliquer l’algorithme de Monte-Carlo séquentiel sur des données simulées.

## 5 Etude de simulation

L’algorithme de Monte Carlo séquentiel est appliqué sur des données simulées suivant le modèle composite LogNorm( $\sigma = 1/2$ ) – Par( $\alpha = 1, \gamma = 5$ ). Dans la Section 5.1, les modèles composites lognormale-Pareto, gamma-Pareto et Weibull-Pareto sont ajustés sur des échantillons de taille 50, 100 et 250. L’objectif est de vérifier la concentration des lois a posteriori autour de la vraie valeur des paramètres pour le modèle lognormale-Pareto et autour du maximum de vraisemblance pour les modèles gamma-Pareto et Weibull-Pareto. Dans la Section 5.2, 1,000 échantillons de taille 50, 100, 250 et 500 sont générés. L’objectif est de mesurer la capacité des critères d’information à choisir le bon modèle, ici le modèle lognormale-Pareto. Pour cette étude, l’algorithme SMC par activation progressive de la vraisemblance est utilisé avec une population de 1,000 particules. Ce choix s’explique par les temps de calcul plus importants constatés pour la méthode d’introduction progressive des données.

Dans le cadre de l’étude de la distribution des montants de sinistres en assurances, une attention particulière est portée à l’estimation des quantiles d’ordre élevé qui caractérise le risque du portefeuille. Les lois de probabilité a posteriori pour les paramètres permettent de définir des lois a posteriori pour des quantités d’intérêt telles que les quantiles 95% et 99%. Soit  $\Delta$  la quantité d’intérêt, le quantile à 95% par exemple. La loi a posteriori de  $\Delta$  pour un modèle  $m_j$ ,  $j = 1, \dots, J$  paramétré par  $\theta$  est donnée par

$$\pi(\Delta|\mathbf{x}, m_j) = \int \pi(\Delta|\mathbf{x}, m_j, \theta) \pi(\theta|\mathbf{x}, m_j). \quad (19)$$

L’échantillonnage de la loi a posteriori des paramètres  $\pi(\theta|\mathbf{x}, m_j)$  via l’algorithme SMC permet d’approcher la loi a posteriori  $\pi(\Delta|\mathbf{x}, m_j)$  de  $\Delta$  par une distribution empirique  $\Delta_{i,j}$ ,  $i = 1, \dots, N$ , où  $\Delta_{i,j}$  est la quantité  $\Delta$  suivant

le modèle  $m_j$  paramétré par  $\theta_i$ ,  $i = 1, \dots, N$ . La distribution a posteriori des quantiles d'ordre 95% et 99% est étudiée dans la [Section 5.1](#).

Une estimation ponctuelle de  $\Delta$  par le modèle  $m_j$  est obtenue en prenant la moyenne sous la loi (19) avec

$$\widehat{\Delta}_j := \mathbb{E}(\Delta|\mathbf{x}, m_j) \approx \frac{1}{N} \sum_{i=1}^N \Delta_{i,j}, j = 1, \dots, J. \quad (20)$$

L'incertitude vis-à-vis des modèles en concurrence est caractérisée par les probabilités a posteriori  $\pi(m_j|\mathbf{x})$ ,  $j = 1, \dots, J$ . Cette incertitude peut être prise en compte pour l'estimation de  $\Delta$  en exprimant sa loi a posteriori par

$$\pi(\Delta|\mathbf{x}) = \sum_{j=1}^J \pi(\Delta|\mathbf{x}, m_j) \pi(m_j|\mathbf{x}). \quad (21)$$

Une estimation ponctuelle, qui tient compte du risque de modèle, est obtenue en combinant les estimateurs (20) des différents modèles avec

$$\widehat{\Delta} := \mathbb{E}(\Delta|\mathbf{x}) \approx \sum_{j=1}^J \widehat{\Delta}_j \pi(m_j|\mathbf{x}), j = 1, \dots, J. \quad (22)$$

Une description détaillée de cette procédure, appelée *Bayesian Model Averaging* (BMA), est donnée dans le travail de [Hoeting et al. \[1999\]](#). Cette méthode est examinée dans la [Section 5.2](#).

## 5.1 Convergence des estimations a posteriori

Les modèles  $\text{Gamma}(r)\text{-Par}(\alpha, \theta)$ ,  $\text{LogNorm}(\sigma)\text{-Par}(\alpha, \theta)$ , et  $\text{Weib}(r)\text{-Par}(\alpha, \theta)$  sont ajustés sur trois échantillons comprenant 50, 100, et 250 observations simulées suivant un modèle  $\text{LogNorm}(\sigma = 1/2)\text{-Par}(\alpha = 1, \gamma = 5)$ . Les hypothèses a priori pour la partie attritionnelle sont données par

$$r \sim \text{Gamma}(0.1, 10), \sigma \sim \text{Gamma}(0.1, 10), k \sim \text{Gamma}(0.1, 10),$$

Les hypothèses a priori pour la partie extrême sont identiques pour chacun des modèles avec

$$\alpha \sim \text{Gamma}(0.1, 10), \theta \sim \text{Gamma}(0.1, 10).$$

Les lois a posteriori des différents paramètres sont données sur la [Figure 11](#).

Les lois a posteriori se rapprochent des valeurs attendues au fur et à mesure que la taille de l'échantillon augmente. A noter que les estimateurs du maximum de vraisemblance pour les modèles Weibull-Pareto et gamma-Pareto sur les [Figures 11c](#) et [11b](#) ont été calculés sur des échantillons de taille 10,000 et correspondent donc à des estimateurs pseudo-asymptotiques. Les quantiles à 95% et 99% sont estimés avec chacun des modèles, leur distribution a posteriori est donnée sur la [Figure 12](#) sous la forme de boîtes à moustache par soucis de visibilité.

Les quantiles du modèle  $\text{LogNorm} - \text{Par}(\sigma = 1/2, \alpha = 1, \gamma = 5)$ , générateur des données, sont recouvert par les distributions a posteriori pour toutes les tailles d'échantillon et pour chacun des modèles. L'estimation des quantiles demeure fiable même si le modèle est mal spécifié ce qui montre la flexibilité de ces modèles composites.

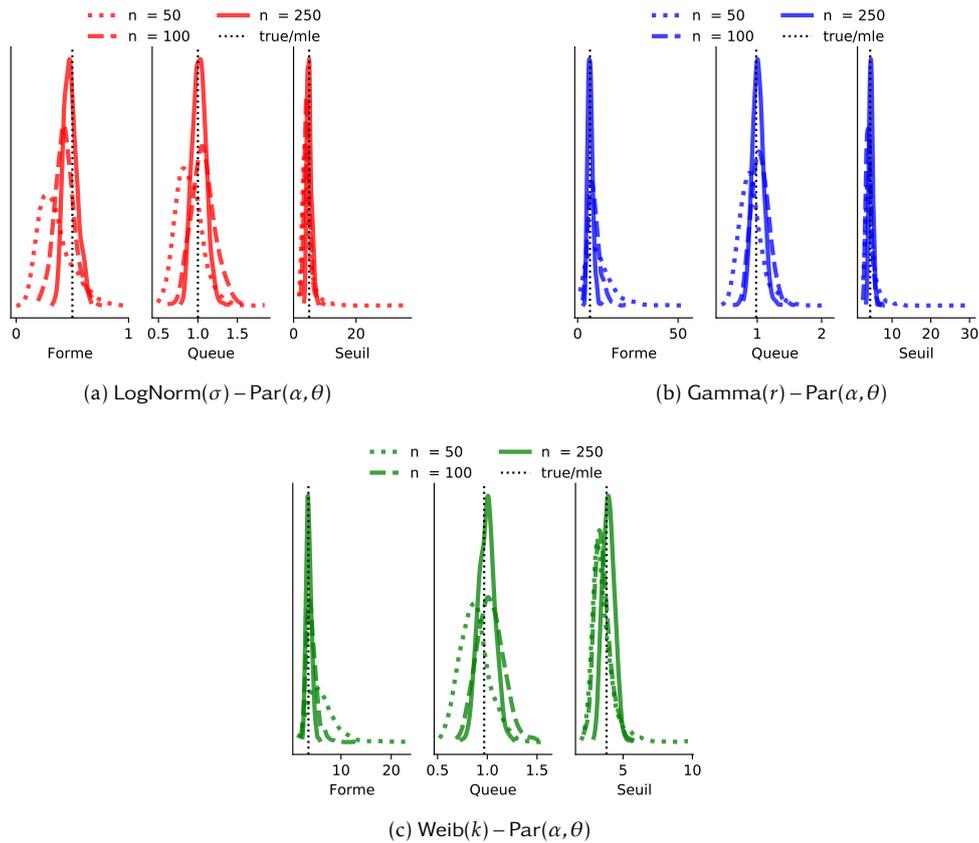


Figure 11 – Lois a posteriori des paramètres des modèles composites calibrés sur des échantillons de données simulées suivant le modèle  $\text{LogNorm}(\sigma = 1/2) - \text{Par}(\alpha = 1, \gamma = 5)$  comprenant 50 observations (pointillés), 100 observations (ligne discontinue) et 250 observations (ligne continue).

## 5.2 Cohérence des méthodes de sélection de modèle

L'expérience de simulation menée dans la section précédente est répétée 1,000. Le premier objectif est d'étudier le comportement des critères d'information permettant la sélection du modèle en terme d'adéquation aux données en fonction de la taille de l'échantillon. La [Figure 13](#) montre la distribution du log de la constante de normalisation et de la probabilité a posteriori de chaque modèle.

Les différences entre les logarithmes des constantes de normalisation sont à peine perceptibles. Le passage à l'exponentielle pour calculer les probabilités a posteriori révèle de manière plus claire quel modèle est privilégié. Les modèles lognormale-pareto et gamma-Pareto ont une plus forte probabilité a posteriori, la méthode favorise le bon modèle (lognormale-Pareto) pour les tailles d'échantillons 250 et 500. Si un seul modèle doit être conservé pour l'analyse, il semble naturel de choisir celui associé à la plus grande probabilité a posteriori. Les critères d'information DIC et WAIC, introduits dans la [Section 4.3](#) peuvent être utilisés de la même façon en sélectionnant le modèle associé à la plus petite valeur de DIC ou de WAIC. L'efficacité des critères d'information est alors mesurée en comptant le nombre de fois que le modèle lognormale-Pareto est choisi pour chacune des tailles d'échantillon et chacun des critères de sélection. La [Figure 14](#) reporte la fréquence de sélection de chacun des modèles composites suivant les différents critères d'information.

Le choix via la probabilité a posteriori procède de la maximisation du log de la constante de normalisation, ce

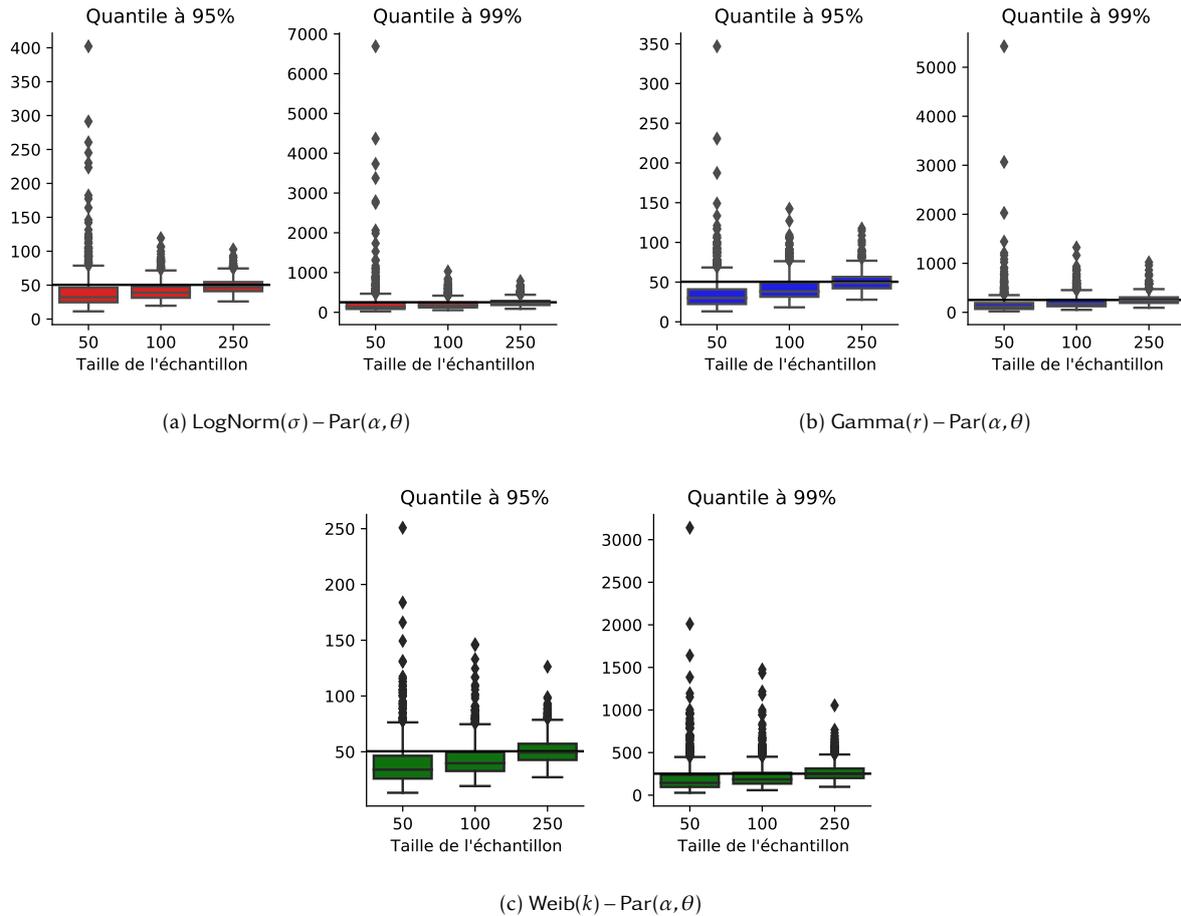


Figure 12 – Distribution des quantiles estimés a posteriori par chacun des modèles composites calibrés sur des échantillons de données simulées suivant le modèle LogNorm – Par( $\sigma = 1/2, \alpha = 1, \gamma = 5$ ) comprenant 50, 100, et 250 observations. La ligne continue indique le niveau des quantiles d'ordre 95% et 99% du modèle LogNorm( $\sigma = 1/2$ ) – Par( $\alpha = 1, \gamma = 5$ ).

critère est appelé LMD pour *Log Marginal Deviance*. Le critère DIC favorise nettement le modèle lognormale-Pareto pour toutes les tailles d'échantillon. Les critères LMD et WAIC sont plus hésitants entre les modèles lognormale-Pareto et gamma-Pareto pour les petites tailles d'échantillon  $n \in \{50, 100\}$ . Les critères favorisent tous le bon modèle (lognormal-Pareto) pour les échantillons de taille  $n \in \{250, 500\}$ . Comme mentionné précédemment, l'estimation des quantiles est importante d'un point de vue pratique. La précision de l'estimation des quantiles d'ordre 95% et 99% par les modèles composites est comparée à l'estimation non paramétrique via le quantile empirique. L'étude considère également l'estimation via BMA qui combine les estimations obtenues via les modèles composites (22). Les écarts absolus moyens, calculés sur les 1,000 simulations, pour les quantiles d'ordre 95% et 99% sont reportés dans le tableau 4.

Pour les petites tailles d'échantillon  $n \in \{50, 100\}$  le modèle gamma-Pareto renvoie les meilleures estimations. Le modèle lognormale-Pareto est plus précis pour les plus grandes tailles d'échantillon  $n \in \{250, 500\}$ . Les erreurs d'estimations pour les modèles composites sont très similaires, ce qui est cohérent avec les résultats obtenus dans la section Section 5.1. L'estimation consolidée par BMA ne permet pas d'améliorer la précision. L'estimation via le quantile empirique est de bonne facture pour le quantile d'ordre 95%, elle est bien moins

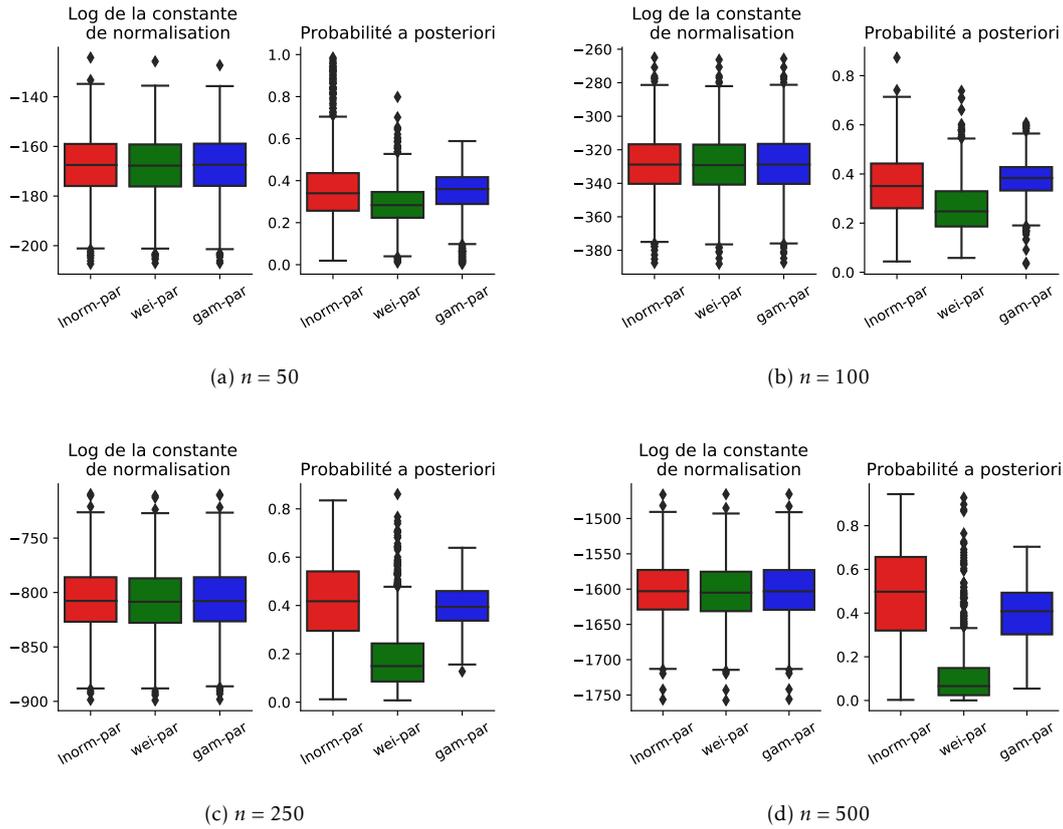


Figure 13 – Log de la constante de normalisation et probabilités a posteriori des modèles composites pour 1,000 échantillons de données simulées suivant le modèle LogNorm – Par( $\sigma = 1/2, \alpha = 1, \gamma = 5$ ) comprenant  $n \in \{50, 100, 250, 500\}$  observations.

modèle	$n = 50$		$n = 100$		$n = 250$		$n = 500$	
	95%	99%	95%	99%	95%	99%	95%	99%
Empirique	23.48	531.88	17.31	205.45	11.33	126.41	7.72	85.76
LogNorm – Par	24.27	240.95	<b>15.50</b>	136.56	<b>9.30</b>	<b>77.80</b>	<b>6.16</b>	<b>50.69</b>
Weib – Par	23.30	221.10	16.49	148.49	10.09	88.59	6.90	60.46
Gamma – Par	<b>21.68</b>	<b>195.71</b>	<b>15.50</b>	<b>134.61</b>	9.59	80.87	6.35	53.18
BMA	23.57	228.57	15.75	139.02	9.54	80.69	6.28	52.28

Table 4 – Ecart absolu moyen dans l'estimation des quantiles à 95% et 99% avec les modèles composites calculés sur 1,000 échantillons de données simulées suivant le modèle LogNorm – Par( $\sigma = 1/2, \alpha = 1, \gamma = 5$ ) comprenant  $n \in \{50, 100, 250, 500\}$  observations.

précise que celle des modèles composites pour le quantile d'ordre 99%. Cela démontre la supériorité de l'approche paramétrique dans l'estimation des quantiles extrêmes pour les nombres d'observations considérés dans cette étude. En complément du Tableau 4, la distribution des erreurs absolues est fournie par la Figure 15. L'erreur d'estimation par le quantile empirique est plus importante pour les tailles d'échantillon  $n \in \{250, 500\}$ , voir les Figures 15c et 15d. La différence est notable surtout pour le quantile d'ordre 99%. La distribution des

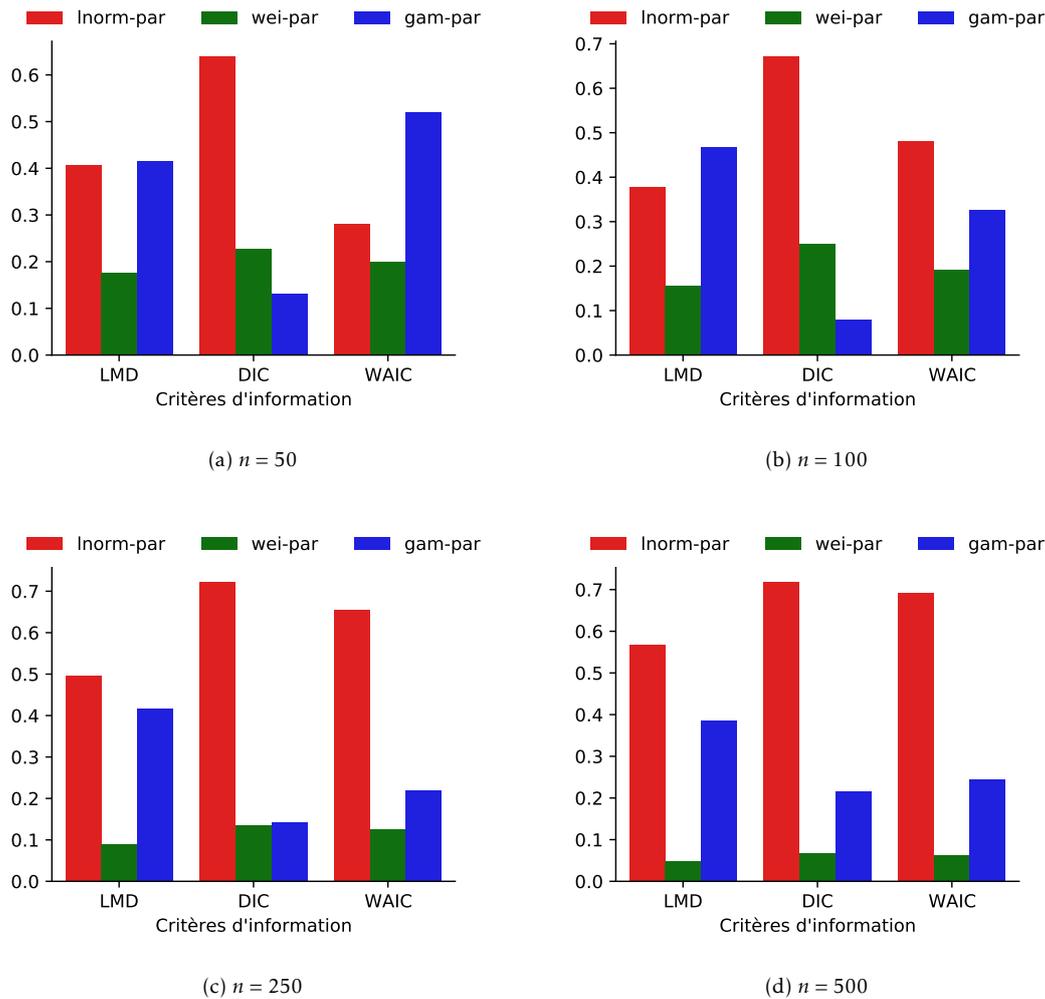


Figure 14 – Fréquence de sélection des modèles composites en concurrence pour des échantillons de données simulées suivant le modèle LogNorm-Par( $\sigma = 1/2, \alpha = 1, \gamma = 5$ ) comprenant  $n \in \{50, 100, 250, 500\}$  observations.

erreurs est très similaire pour tous les modèles composites. L'utilisation de la méthode BMA ne permet pas une réduction significative de la variance des erreurs. L'algorithme SMC est appliqué sur un jeu de données réelles dans la section suivante.

## 6 Application sur un jeu de données réelles

Le jeu de données considéré comprend des données de sinistres automobiles impliquant des dommages corporels en Australie. Le jeu de données ausautoBI8999 appartient au package R CASData sets maintenu par Dutang and Charpentier [2016] qui accompagne l'ouvrage de Charpentier [2014]. La base de données comprend 22,036 montants de sinistres renseignés dans la variable AggClaim. La date de clôture des sinistres est donnée par la variable FinDate. Un extrait du jeu de données est fourni dans le Tableau 5. La variable FinMth est un indice correspondant au mois de clôture du sinistre. Les statistiques descriptives sont données dans le tableau 6. Les montants de sinistres montrent une grande disparité avec une valeur maximale supérieure à 4 millions de dollars australiens et un écart-type très supérieur à la moyenne. L'histogramme et la boîte à moustache de la

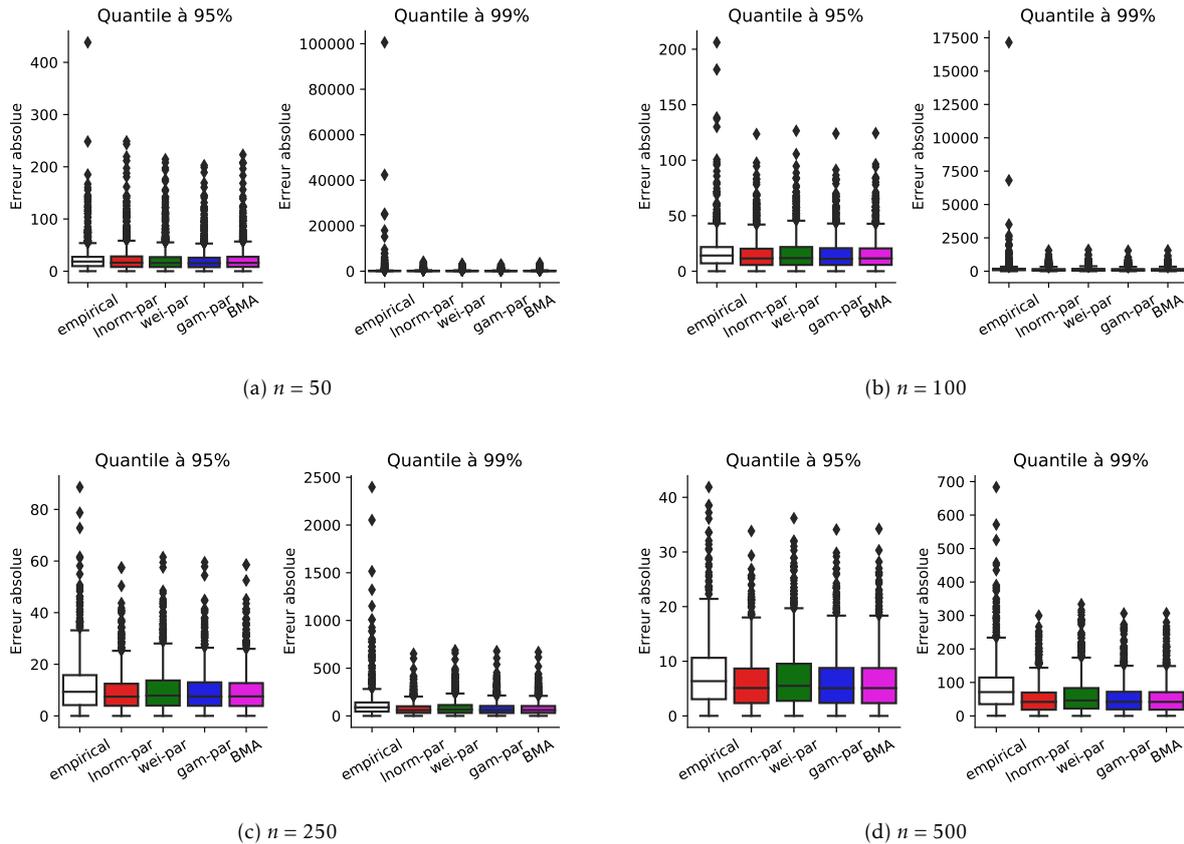


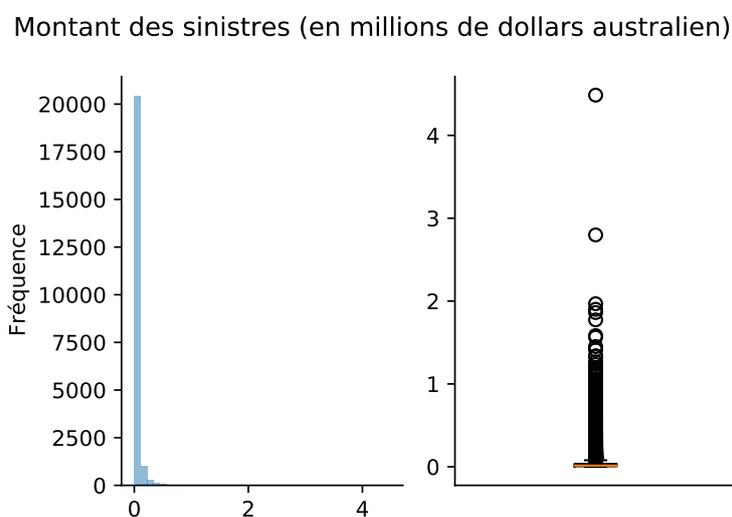
Figure 15 – Erreurs en valeur absolue sur les quantiles de la distribution LogNorm – Par( $\sigma = 1/2, \alpha = 1, \gamma = 5$ ) estimés a posteriori par chacun des modèles composites calibrés sur 1,000 échantillons de données simulées comprenant  $n \in \{50, 100, 250, 500\}$  observations.

FinDate	FinMth	AggClaim
1993-10-01	52	87.75
1994-02-01	56	353.62
1994-02-01	56	688.83
1994-05-01	59	172.80
1994-09-01	63	43.29

Table 5 – Extrait du jeu de données ausautoBI8999 contenant des montants de sinistre automobile.

distribution des montants de sinistre est donnée par la Figure 16. La distribution des montants de sinistres admet une forte occurrence de sinistres de faible intensité et une occurrence plus faible mais non négligeable de sinistres de très forte intensité. Les modèles composites sont a priori adaptés à la situation. L'algorithme SMC est appliqué à l'ensemble des données dans la Section 6.1 puis aux données segmentées annuellement dans la Section 6.2 et mensuellement dans la Section 6.3. L'objectif est d'étudier l'évolution de la sinistralité au cours des périodes d'exercice via les paramètres des modèles composites. Ces modèles à trois paramètres résument parfaitement la distribution des montants de sinistre. Le paramètre de forme décrit la partie attritionnelle, le

	AggClaim
Nombre d'observations	22,036.00
Moyenne	38,367.22
Ecart-type	90,981.11
Minimum	9.96
Quantile à 25%	6,296.97
Quantile à 50%	13,853.87
Quantile à 75%	35,123.42
Maximum	4,485,797.21

Table 6 – Statistiques descriptives de la variable `AggClaim` du jeu de données `ausautoBI8999`.Figure 16 – Histogramme et boîte à moustache de la distribution des montants de sinistre du jeu de données `ausautoBI8999`.

paramètre de queue traduit l'épaisseur de la queue de distribution et enfin le seuil discrimine les sinistres de faible et de forte intensité. L'estimation des quantiles via les modèles composites est comparée aux quantiles empirique d'ordre 95% et 99%. La qualité des estimateurs et l'adéquation des modèles sont discutées. La segmentation des données par mois et par an permet de comparer le comportement des estimateurs suivant deux régimes de données en termes de taille d'échantillon. La [Figure 17](#) donne le nombre de sinistres cloturés par an ([Figure 17a](#)) et par mois ([Figure 17b](#)). Le nombre de sinistres par mois varie entre 100 et 600 tandis que le nombre de sinistres par an se situe entre 1000 et 5000. Dans la suite de l'étude, les montants de sinistres sont divisés par  $10^6$  et exprimés en millions de dollars australiens.

## 6.1 Analyse globale de la sinistralité

Les modèles  $\text{Gamma}(r) - \text{Par}(\alpha, \theta)$ ,  $\text{LogNorm}(\sigma) - \text{Par}(\alpha, \theta)$ , et  $\text{Weib}(r) - \text{Par}(\alpha, \theta)$  sont ajustés sur les données via l'algorithme **SMC**. Les hypothèses a priori pour la partie attritionnelle sont données par

$$r \sim \text{Gamma}(0.1, 10), \sigma \sim \text{Gamma}(0.1, 10), k \sim \text{Gamma}(0.1, 10),$$

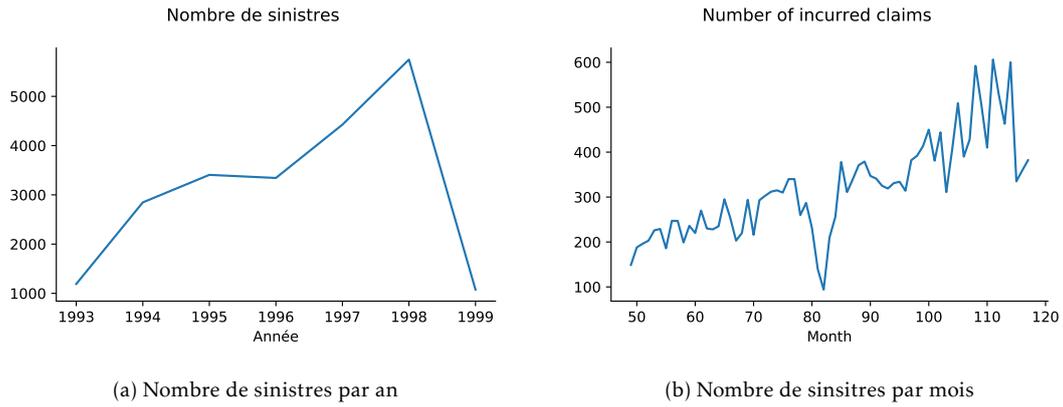


Figure 17 – Nombre de sinistres par période d'exercice.

Les hypothèses a priori pour la partie extrême sont identiques pour chacun des modèles

$$\alpha \sim \text{Gamma}(0.1, 10), \theta \sim \text{Gamma}(0.1, 10).$$

La distribution a posteriori des paramètres de chacun des modèles est donnée par la Figure 18. La loi a

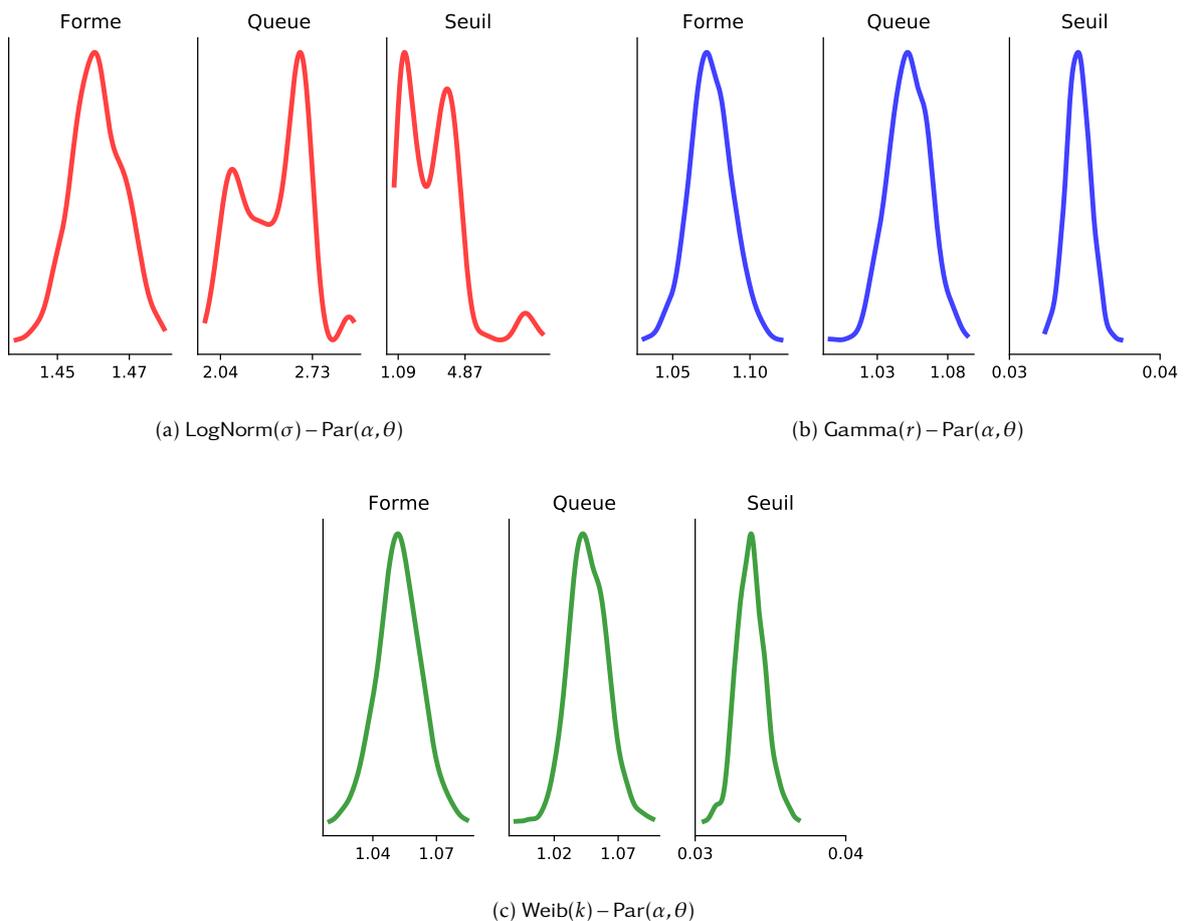


Figure 18 – Loi a posteriori des paramètres des modèles composites pour les données de pertes automobiles australiennes.

posteriori du modèle lognormale-Pareto n'admet pas la forme attendue, le paramètre de seuil prend des valeurs trop élevées, proche du maximum observé dans les données. En poussant la valeur du seuil au maximum, l'algorithme manque de données pour ajuster la queue de la distribution. Le résultat est une loi a posteriori inexploitable pour le paramètre de queue. Ce phénomène s'explique par une bonne adéquation de la loi lognormale aux données. La loi lognormale  $\text{LogNorm}(\mu, \sigma)$  est ajustée aux données avec les hypothèses a priori suivantes:

$$\mu \sim \text{Normal}(0, 10), \sigma \sim \text{Gamma}(0.1, 10).$$

le choix de la loi normale pour le paramètre  $\mu$  s'explique par le fait qu'il est à valeur dans  $\mathbb{R}$ . Les diagrammes quantiles-quantiles des modèles lognormale, lognormale-Pareto, gamma-Pareto et Weibull-Pareto sont données par la Figure 19. Les paramètres de chacun des modèles sont estimés en prenant la moyenne des lois a posteriori. L'ajustement du modèle lognormal est remarquable. Les modèles composites ajustent bien la partie

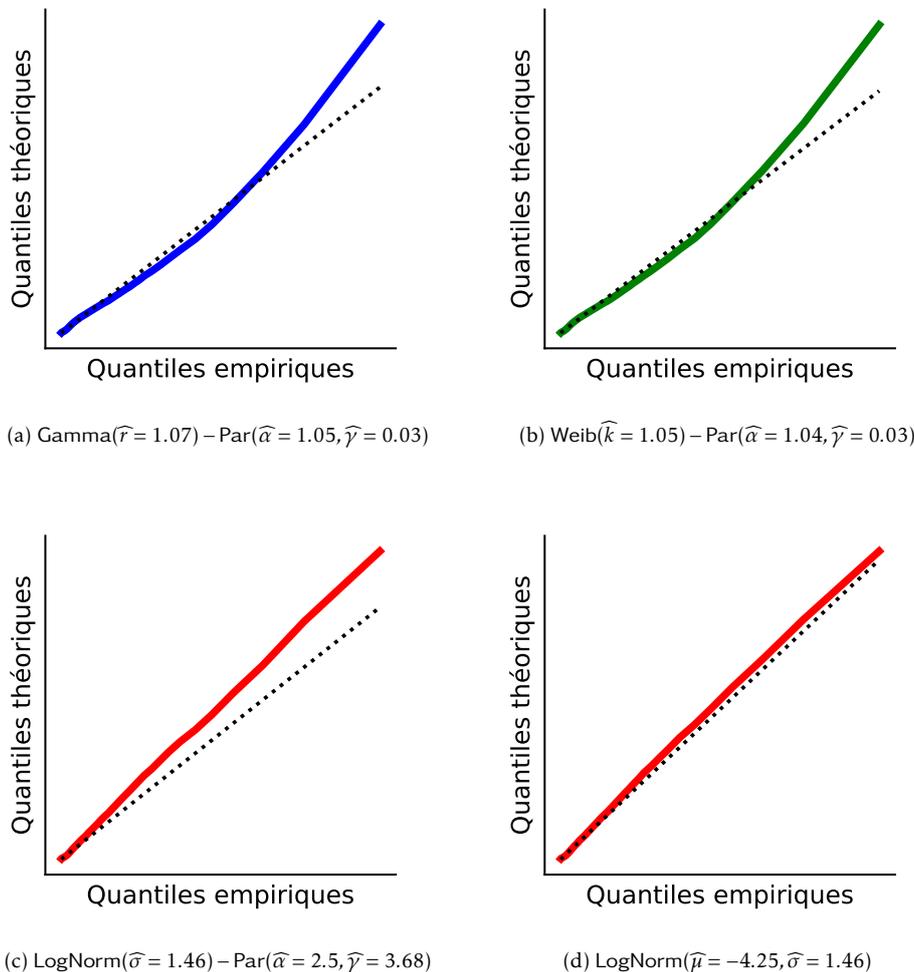


Figure 19 – Diagramme quantile-quantile d'ajustement des modèles composites aux pertes automobiles australiennes.

attritionnelle mais surestime l'épaisseur de la queue de distribution. Les valeurs de la log vraisemblance marginale et des probabilités a posteriori sont reportées dans le Tableau 7. Malgré la bonne adéquation aux quantiles empiriques du modèle lognormale, la probabilité a posteriori des modèles composites gamma-Pareto

Modèle	log de la constante de normalisation	Probabilité a posteriori	DIC	WAIC
LogNorm – Par	54,193.14	0.00	-107,797.02	-108,415.71
Weib – Par	54,558.40	0.02	-109,163.09	-109,149.82
Gamma – Par	54,562.08	0.98	-109,170.88	-109,157.51
LogNorm	54,196.78	0.00	-108,423.77	-108,420.43

Table 7 – Constante de normalisation et probabilité a posteriori des modèles composites et du modèle lognormale calculés sur les données de pertes automobiles australiennes.

et Weibull-Pareto est plus élevée. La bonne adéquation de la partie attritionnelle suffit à augmenter suffisamment la constante de normalisation malgré une surestimation des quantiles d'ordre élevé. Cette différence entre adéquation visuelle aux quantiles empiriques et adéquation en termes de vraisemblance avait déjà été notée dans le cadre de l'application aux données pertes "incendie" danoises. La fonction de vraisemblance est calculée sur l'ensemble des données, l'occurrence plus fréquente de sinistres de montant faible leur donne plus d'importance. Les modèles sont ajustés aux données segmentées par année dans la section suivante.

## 6.2 Analyse annuelle de la sinistralité

La [Figure 20](#) montre l'évolution de la distribution des montants de sinistres au fur et à mesure des années. Les quantiles à 50%, 75% et 95% sont stables d'une année à l'autre, le quantile à 99% semble augmenter. Les

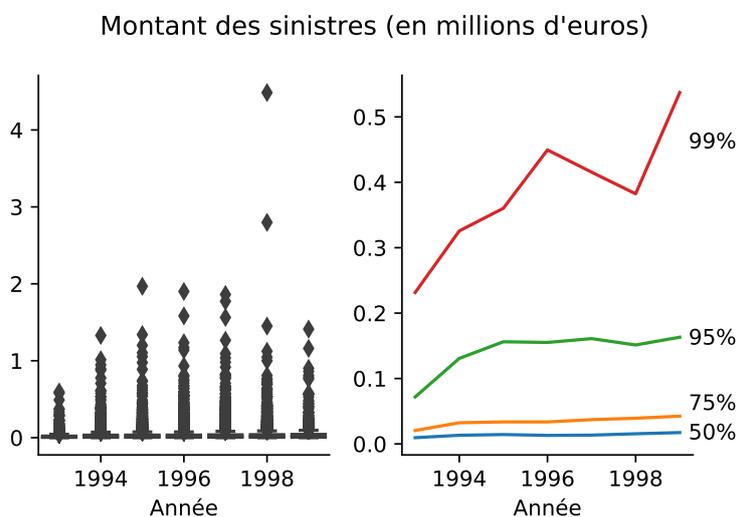


Figure 20 – Quantiles empiriques à 50%, 75%, 95% et 99% mesurés chaque année.

montants de sinistres sont plus faibles pour l'année 1993. Les modèles  $\text{Gamma}(r) - \text{Par}(\alpha, \theta)$ ,  $\text{LogNorm}(\sigma) - \text{Par}(\alpha, \theta)$ , et  $\text{Weib}(r) - \text{Par}(\alpha, \theta)$  sont ajustés sur les données via l'algorithme **SMC**. Les hypothèses a priori pour la partie attritionnelle sont données par

$$r \sim \text{Gamma}(0.1, 10), \sigma \sim \text{Gamma}(0.1, 10), k \sim \text{Gamma}(0.1, 10),$$

Les hypothèses a priori pour la partie extrême sont identiques pour chacun des modèles

$$\alpha \sim \text{Gamma}(0.1, 10), \theta \sim \text{Gamma}(0.1, 10).$$

La distribution a posteriori des paramètres de chacun des modèles pour chaque année est donnée sous la forme de boîtes à moustache par la [Figure 21](#). Les lois a posteriori des paramètres pour les modèles Weib – Par et

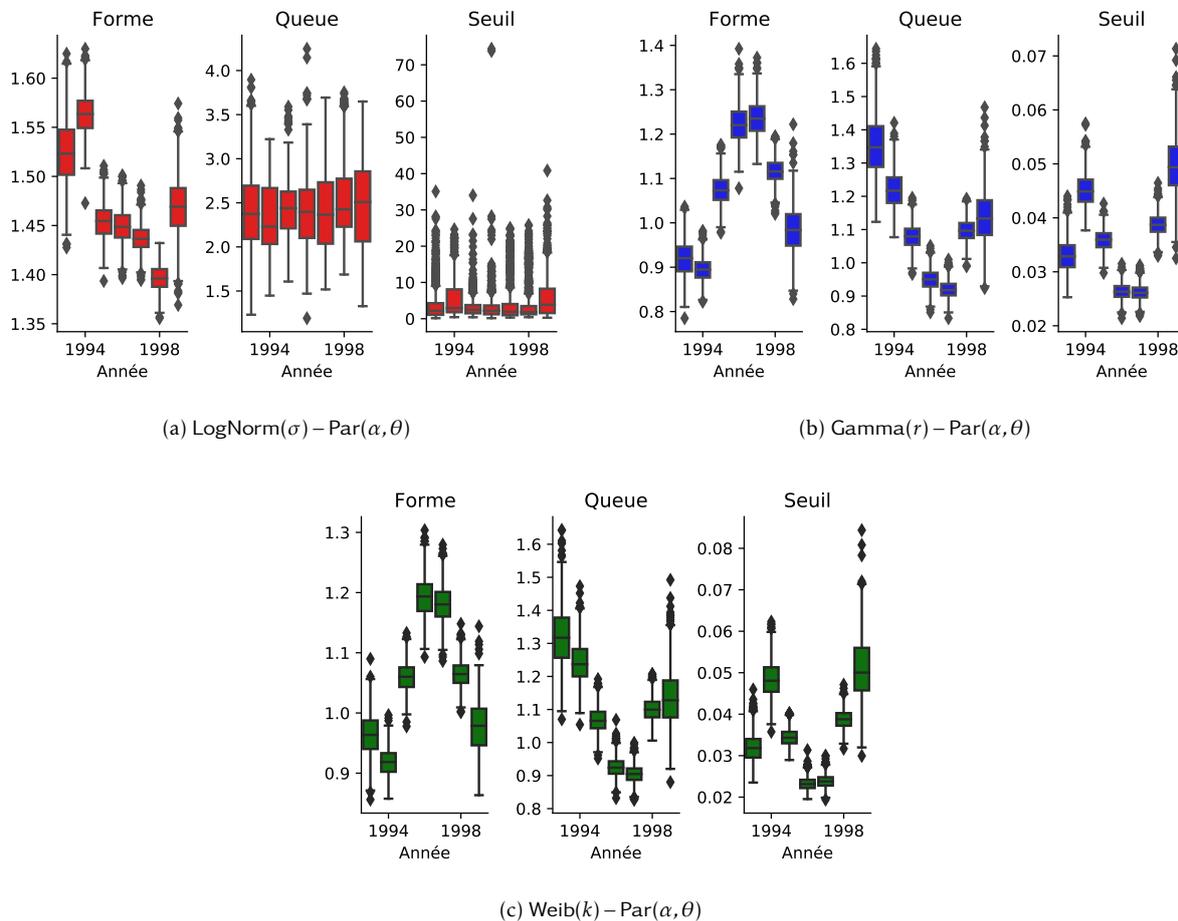


Figure 21 – Loi a posteriori des paramètres des modèles composites pour les données de pertes automobiles australienne segmentées annuellement.

Gamma – Par ont une évolution très similaire, c.f. [Figure 21b](#) et [Figure 21c](#). Le paramètre de forme pour la partie attritionnelle est proche de 1 ce qui indique l'adéquation avec la loi exponentielle pour les montants de sinistres de faible intensité. cette hypothèse semble valide sur l'ensemble des périodes d'exercice. La queue de distribution semble cependant s'épaissir avec les années. Le paramètre de queue est plus faible pour les années entre 1995 et 1997. Il augmente pour les années 1998 et 1999 mais cette augmentation (qui signifie une queue moins épaisse) est compensée par l'augmentation du paramètre seuil. Les résultats obtenus pour le modèle LogNorm – Par posent problèmes. La loi a posteriori des paramètres de queue et de seuil admet une grande variance, c.f. [Figure 21a](#). Les valeurs prises par le paramètre seuil sont très grandes allant jusqu'au maximum des données. Comme dans l'étude menée sur l'ensemble des données dans la section 6.1, l'algorithme modélise l'ensemble des données par la loi lognormale en augmentant autant que possible le paramètre seuil. Le résultat est une loi a posteriori peu fiable. Les probabilités a posteriori des modèles composites pour chaque

année sont reportées sur la [Figure 22](#). La méthode de sélection basée sur la probabilité a posteriori des modèles

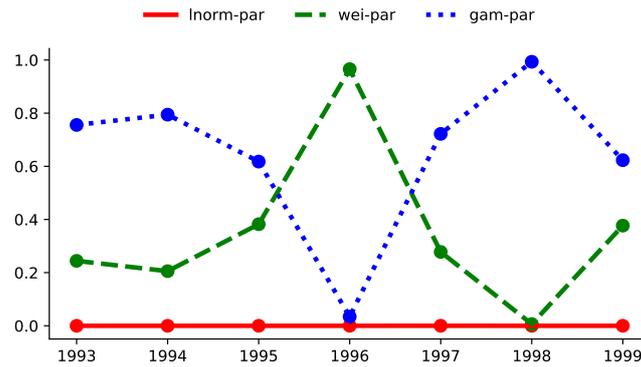


Figure 22 – Probabilités a posteriori des modèles  $\text{Gamma}(r) - \text{Par}(\alpha, \theta)$ ,  $\text{LogNorm}(\sigma) - \text{Par}(\alpha, \theta)$ , et  $\text{Weib}(r) - \text{Par}(\alpha, \theta)$  pour chaque année.

conduit à écarter systématiquement le modèle  $\text{LogNorm} - \text{Par}$ . Le modèle  $\text{Gamma} - \text{Par}$  semble être privilégié. L'estimation des quantiles à 95% et à 99% par les modèles composites ainsi que l'estimation combinant les trois modèles via la méthode *Bayesian Model Averaging* (BMA) est comparée aux quantiles empiriques sur la [Figure 23](#). L'estimation empirique du quantile à 95% est similaire à celle des modèles composites. Une différence très

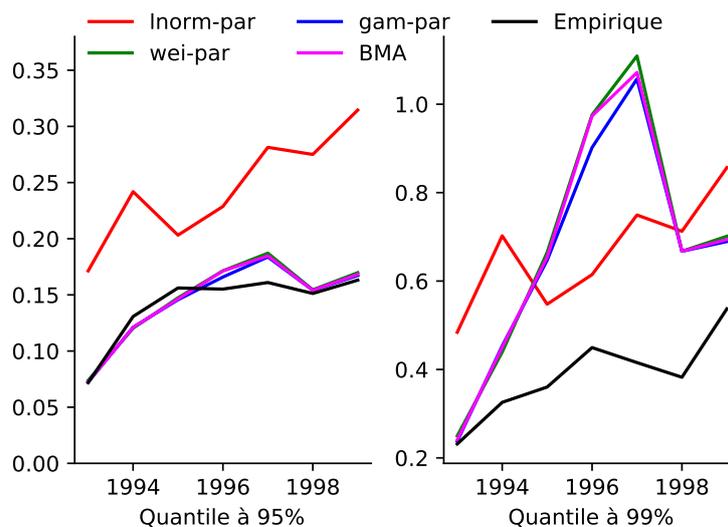


Figure 23 – Quantiles à 95% et 99% estimés empiriquement et via les modèles  $\text{Gamma}(r) - \text{Par}(\alpha, \theta)$ ,  $\text{LogNorm}(\sigma) - \text{Par}(\alpha, \theta)$ , et  $\text{Weib}(r) - \text{Par}(\alpha, \theta)$  pour chaque année.

nette est à noter pour le quantile à 99% dont la valeur est plus élevée pour les modèles composites. L'estimation suivant le modèle  $\text{LogNorm} - \text{Par}$  doit être ignorée au vu de la variabilité des lois a posteriori obtenues, c.f. [Figure 21a](#). Pour pallier au problème posé par le modèle  $\text{LogNorm} - \text{Par}$ , deux solutions sont proposées. La première consiste à écarter le modèle  $\text{LogNorm} - \text{Par}$ , ce qui est fait dans la [Section 6.2.1](#). La deuxième est de remplacer le modèle  $\text{LogNorm} - \text{Par}$  par un modèle  $\text{LogNorm}(\mu, \sigma)$ , ce qui est fait dans [Section 6.2.2](#).

### 6.2.1 Analyse annuelle de la sinistralité sans le modèle lognormal-Pareto

Ecarter le modèle LogNorm-Par n'a pas d'effet sur les lois a posteriori reportées sur la Figure 21. Les probabilités a posteriori varient peu car la probabilité a posteriori du modèle LogNorm-Par était faible, voir la Figure 24. De la même façon l'estimation des quantiles à 95% et 99% n'est pas affectée par le fait d'écarter le modèle

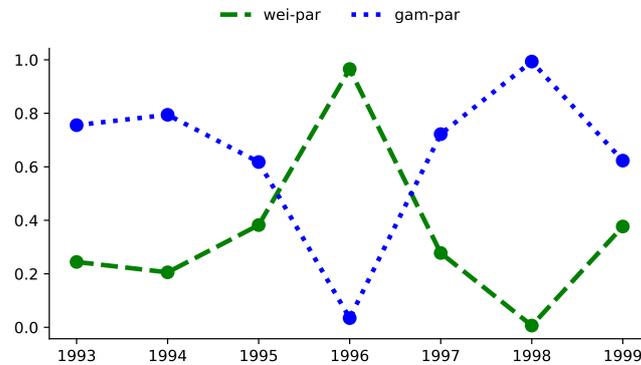


Figure 24 – Probabilités a posteriori des modèles  $\text{Gamma}(r) - \text{Par}(\alpha, \theta)$ , et  $\text{Weib}(r) - \text{Par}(\alpha, \theta)$  pour chaque année.

LogNorm-Par, comme le montre la Figure 25. La méthode d'inférence des paramètres utilisée semble privilégier

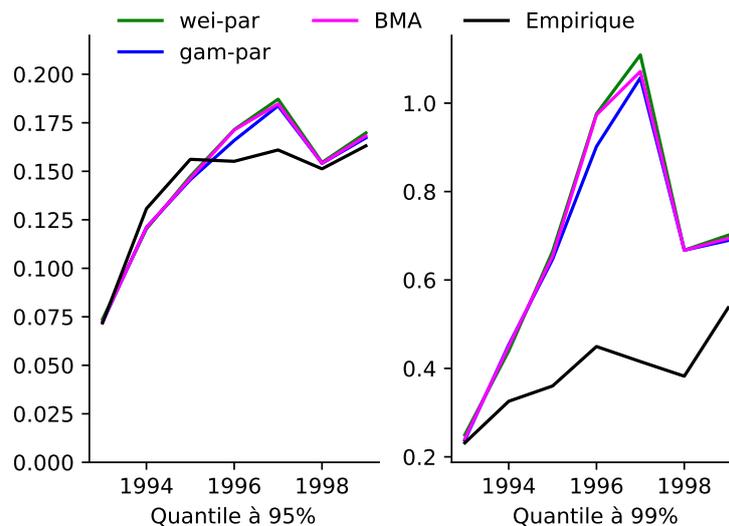


Figure 25 – Quantiles à 95% et 99% estimés empiriquement et via les modèles  $\text{Gamma}(r) - \text{Par}(\alpha, \theta)$ , et  $\text{Weib}(r) - \text{Par}(\alpha, \theta)$  pour chaque année.

l'adéquation de la partie attritionnelle des montants de sinistre. Cela s'explique par le plus grand nombre d'observations disponibles pour les montants de sinistres de faible intensité. Il faut noter également que l'estimation empirique de quantile d'ordre important (par exemple 99%) est très volatile. Le fait que les modèles composites renvoient une valeur plus importante indique que ce quantile extrême est potentiellement sous-estimé empiriquement.

### 6.2.2 Analyse annuelle de la sinistralité en remplaçant le modèle Lognormal-Pareto par un modèle log-normal

Lorsque des modèles offrent un ajustement aux données similaires, il est naturel de privilégier le modèle le plus simple, c'est à dire celui ayant le moins de paramètres. L'inadéquation du modèle LogNorm – Par s'explique en partie par la bonne adéquation de la loi lognormale aux données. Il est logique de comparer l'ajustement du modèle LogNorm( $\mu, \sigma$ ) à celui des modèles composites Gamma – Par et Weib – Par. Les hypothèses a priori pour le modèle modèle LogNorm( $\mu, \sigma$ ) sont les suivantes:

$$\mu \sim \text{Normal}(0, 10), \sigma \sim \text{Gamma}(0.1, 10).$$

La loi a posteriori des paramètres  $\mu$  et  $\sigma$  du modèle lognormal est donnée sous la forme de boîte à moustache sur la Figure Figure 26. La valeur des paramètres du modèle lognormal est stable au cours des années. Les

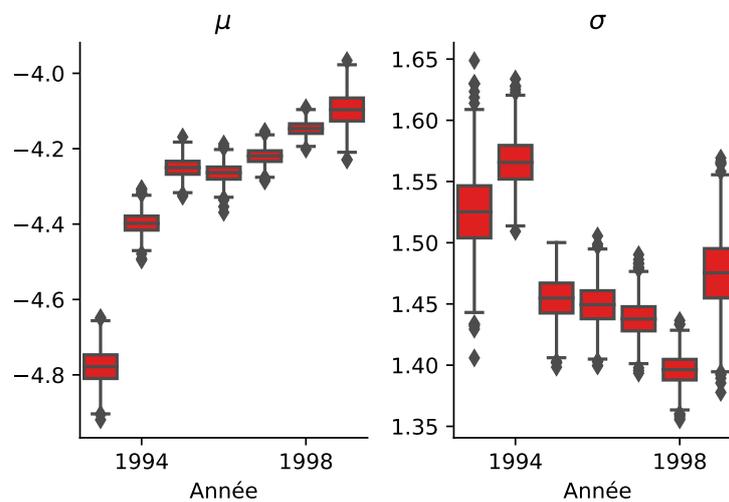


Figure 26 – Quantiles à 95% et 99% estimés empiriquement et via les modèles Gamma( $r$ ) – Par( $\alpha, \theta$ ), et Weib( $r$ ) – Par( $\alpha, \theta$ ) pour chaque année.

probabilités a posteriori des modèles LogNorm, Gamma – Par et Weib – Par sont reportées sur la Figure 27. La

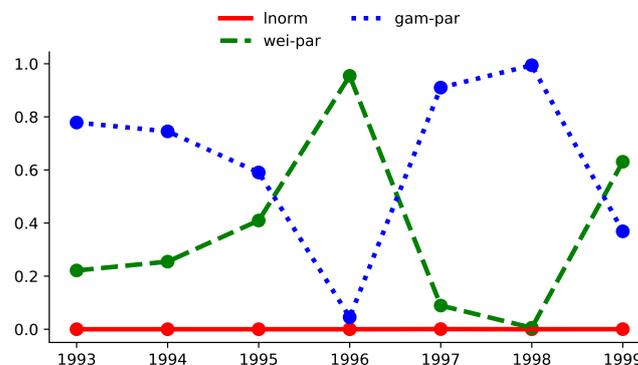


Figure 27 – Probabilités a posteriori des modèles LogNorm( $\mu, \sigma$ ), Gamma( $r$ ) – Par( $\alpha, \theta$ ), et Weib( $r$ ) – Par( $\alpha, \theta$ ) pour chaque année.

méthode de sélection de modèle tend à privilégier les modèles composites au détriment du modèle lognormal.

L'estimation des quantiles d'ordre 95% et 99% via les modèles lognormal, gamma-Pareto et Weibull-Pareto est comparée à l'estimation empirique sur la [Figure 28](#). Les quantiles du modèle lognormal se rapprochent

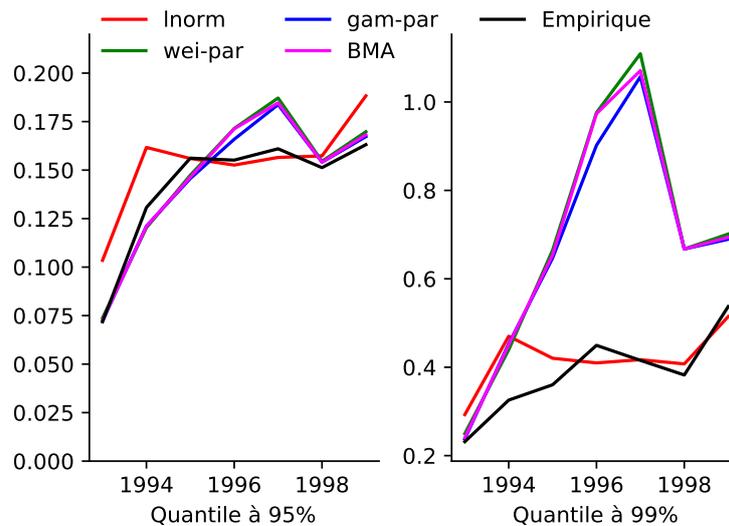


Figure 28 – Quantiles à 95% et 99% estimés empiriquement et via les modèles  $\text{LogNorm}(\mu, \sigma)$ ,  $\text{Gamma}(r) - \text{Par}(\alpha, \theta)$ , et  $\text{Weib}(r) - \text{Par}(\alpha, \theta)$  pour chaque année.

d'avantage des quantiles empiriques. La faible probabilité a posteriori du modèle lognormal ne permet pas de répercuter cela dans l'estimation des quantiles via la méthode BMA. L'objectif de la section suivante est de regarder si les observations faites pour les données segmentées annuellement sont aussi valables pour une segmentation mensuelle.

### 6.3 Analyse mensuelle de la sinistralité

La [Figure 29](#) montre l'évolution de la distribution des montants de sinistres au fur et à mesure des mois. Les quantiles à 50%, 75%, 95% et 99% sont stables d'un mois à l'autre, leur variabilité est en revanche plus

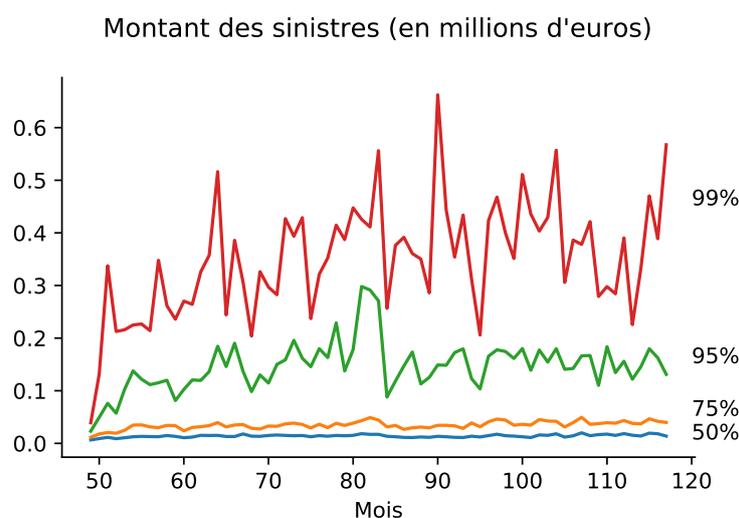


Figure 29 – Quantiles empiriques à 50%, 75%, 95% et 99% mesurés chaque mois.

importante comparée à celle des quantiles empirique mesurés annuellement et reportés sur la [Figure 29](#). Les

modèles  $\text{Gamma}(r) - \text{Par}(\alpha, \theta)$ ,  $\text{LogNorm}(\sigma) - \text{Par}(\alpha, \theta)$ , et  $\text{Weib}(r) - \text{Par}(\alpha, \theta)$  sont ajustés sur les données via l'algorithme **SMC**. Les hypothèses a priori pour la partie attritionnelle sont données par

$$r \sim \text{Gamma}(0.1, 10), \sigma \sim \text{Gamma}(0.1, 10), k \sim \text{Gamma}(0.1, 10),$$

Les hypothèses a priori pour la partie extrême sont identiques pour chacun des modèles

$$\alpha \sim \text{Gamma}(0.1, 10), \theta \sim \text{Gamma}(0.1, 10).$$

La moyenne a posteriori des paramètres de chacun des modèles pour chaque mois, entouré des quantiles de la loi a posteriori d'ordre 5% et 95%, est donnée sur la [Figure 30](#). Les lois a posteriori des paramètres pour les

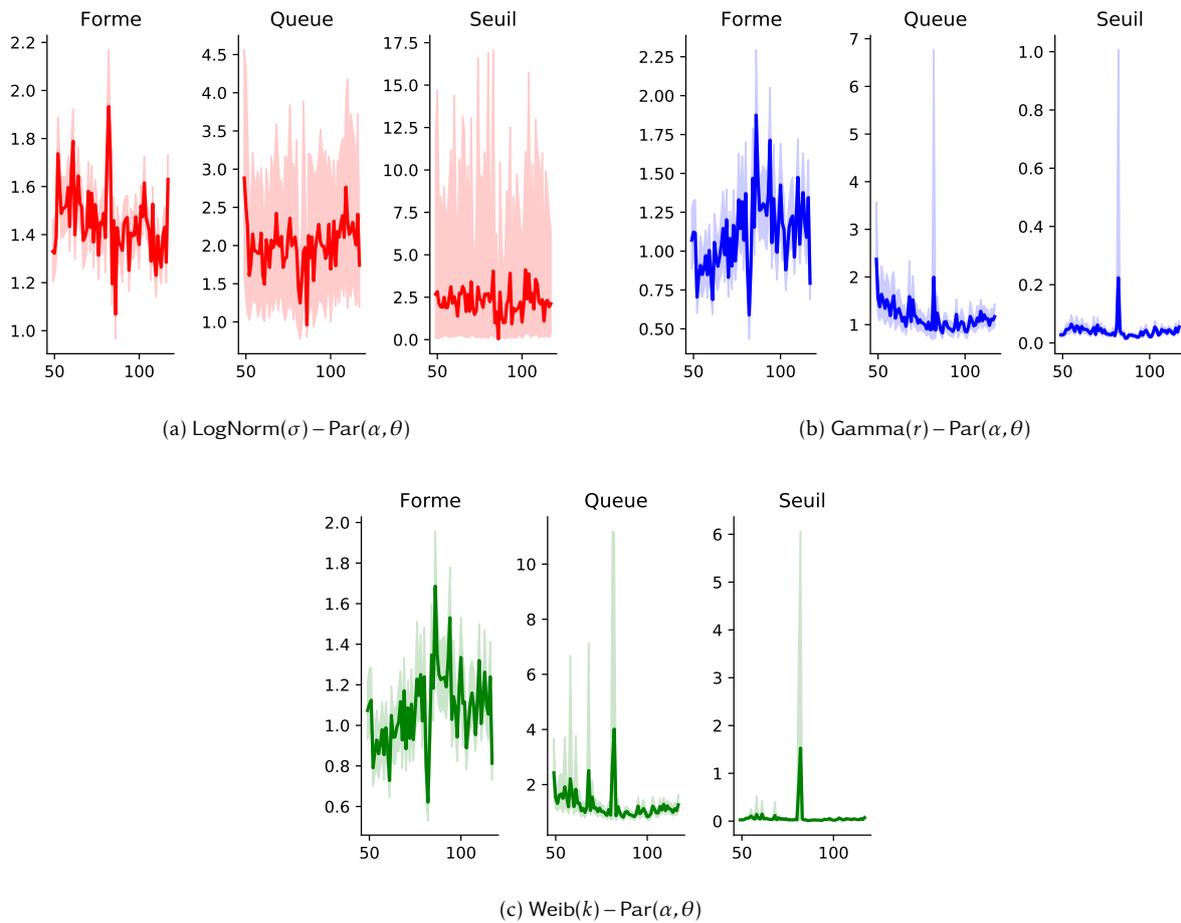


Figure 30 – Moyenne et quantiles d'ordre 5% et 95% de la loi a posteriori des paramètres des modèles composites pour chaque mois.

modèles  $\text{Weib} - \text{Par}$  et  $\text{Gamma} - \text{Par}$  ont une évolution très similaires, c.f. [Figure 30b](#) et [Figure 30c](#). Pour chacun des deux modèles, le mois 82 a posé problème au vu de la grande valeur prise par le quantile d'ordre 95%. La loi a posteriori des paramètres de seuil et de queue pour le modèle  $\text{LogNorm} - \text{Par}$  admet une trop grande variance pour être prise en compte. Les résultats sont comparables à ceux obtenus sur les données ségmentées annuellement dans la [Section 6.2](#). Les probabilités a posteriori des modèles composites pour chaque mois sont reportées sur la [Figure 31](#). Les probabilités a posteriori des modèles n'écartent pas systématiquement le modèle lognormal-Pareto ce qui risque d'être problématique dans l'estimation des quantiles via la méthode

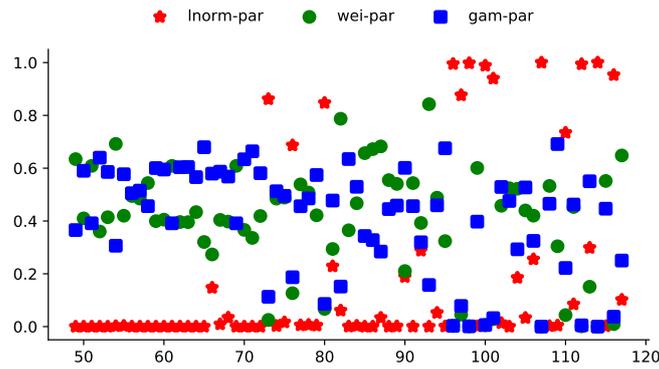


Figure 31 – Probabilités a posteriori des modèles  $\text{Gamma}(r) - \text{Par}(\alpha, \theta)$ ,  $\text{LogNorm}(\sigma) - \text{Par}(\alpha, \theta)$ , et  $\text{Weib}(r) - \text{Par}(\alpha, \theta)$  pour chaque mois.

BMA. L'estimation des quantiles à 95% et à 99% par les modèles composites ainsi que l'estimation combinant les trois modèles via la méthode *Bayesian Model Averaging* (BMA) est comparée aux quantiles empiriques sur la Figure 32. L'estimation des quantiles à 95% et 99% via le modèle lognormal-Pareto est très erratique. Cela se

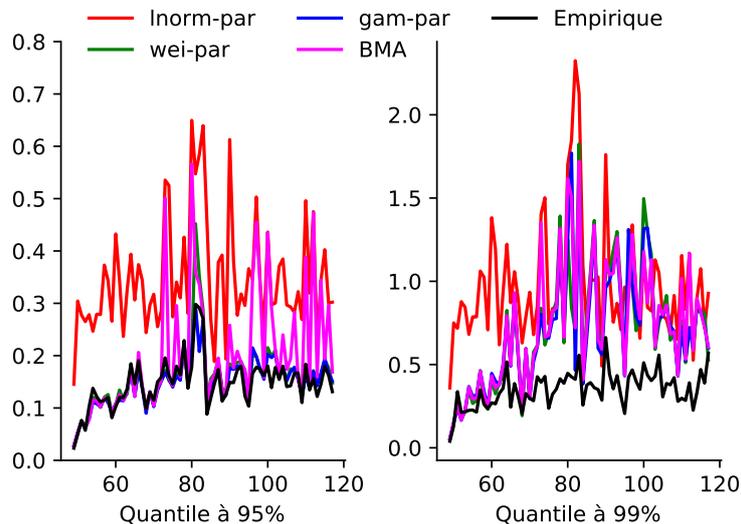


Figure 32 – Quantiles à 95% et 99% estimés empiriquement et via les modèles  $\text{Gamma}(r) - \text{Par}(\alpha, \theta)$ ,  $\text{LogNorm}(\sigma) - \text{Par}(\alpha, \theta)$ , et  $\text{Weib}(r) - \text{Par}(\alpha, \theta)$  pour chaque mois.

répercute sur l'estimation résultant de la combinaison des modèles composites par l'approche BMA. Comme pour les données annuelles, les modèles composites sur-évaluent le quantile à 99% par rapport au quantile empirique. Le modèle lognormal-Pareto est retiré de l'étude menée dans la Section 6.3.1 et remplacé par le modèle  $\text{LogNorm}(\mu, \sigma)$  dans la Section 6.3.2.

### 6.3.1 Analyse mensuelle de la sinistralité sans le modèle Lognormal-Pareto

Ecarter le modèle  $\text{LogNorm} - \text{Par}$  a bien plus d'impact sur les probabilités a posteriori des modèles  $\text{Gamma} - \text{Par}$  et  $\text{Weib} - \text{Par}$  par rapport à l'effet constaté dans la Section 6.2.1. Les probabilités a posteriori des modèles  $\text{Gamma} - \text{Par}$  et  $\text{Weib} - \text{Par}$  pour chaque mois sont reportées sur la Figure 30. La méthode de sélection de modèles ne parvient pas à départager clairement les modèles  $\text{Gamma} - \text{Par}$  et  $\text{Weib} - \text{Par}$ . L'estimation du

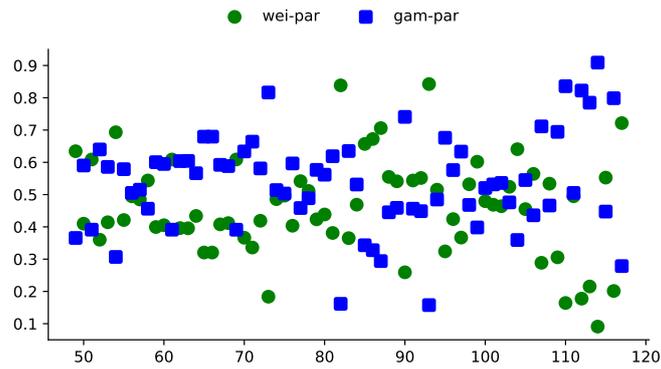


Figure 33 – Probabilités a posteriori des modèles  $\text{Gamma}(r) - \text{Par}(\alpha, \theta)$ , et  $\text{Weib}(r) - \text{Par}(\alpha, \theta)$  pour chaque mois.

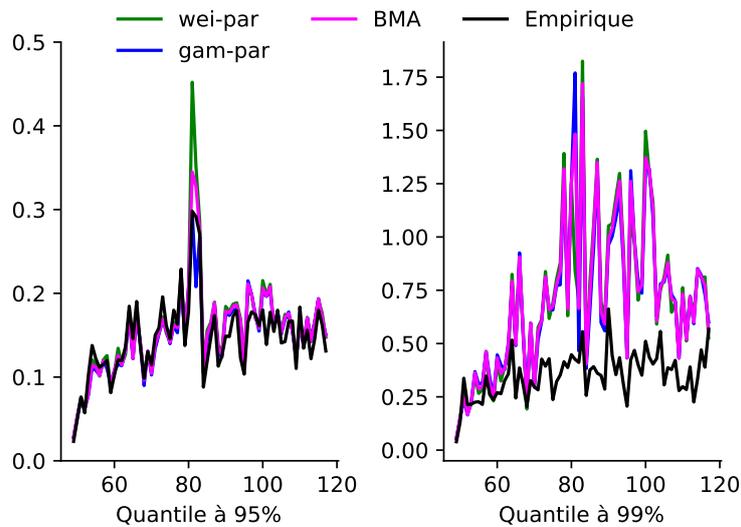


Figure 34 – Quantiles à 95% et 99% estimés empiriquement et via les modèles  $\text{Gamma}(r) - \text{Par}(\alpha, \theta)$ , et  $\text{Weib}(r) - \text{Par}(\alpha, \theta)$  pour chaque mois.

quantile à 95% par les modèles composites est très proche de celle obtenue empiriquement. L'estimation du quantile à 99% par les modèles composites est plutôt variable et toujours plus grande que le quantile empirique.

### 6.3.2 Analyse mensuelle de la sinistralité en remplaçant le modèle Lognormal-Pareto par un modèle log-normal

Dans cette section, le modèle  $\text{LogNorm} - \text{Par}$  est remplacé par le modèle  $\text{LogNorm}(\mu, \sigma)$ . Les hypothèses a priori pour le modèle  $\text{LogNorm}(\mu, \sigma)$  sont les suivantes:

$$\mu \sim \text{Normal}(0, 10), \sigma \sim \text{Gamma}(0.1, 10).$$

La moyenne a posteriori des paramètres  $\mu$  et  $\sigma$  du modèle  $\text{LogNorm}$  est reportée sur la Figure 26 avec les quantiles de la loi a posteriori d'ordre 5% et 95%. La valeur des paramètres du modèle  $\text{LogNorm}$  est stable au cours des mois. Un pic est à noter pour le paramètre de variance  $\sigma$  au mois 82. Les probabilités a posteriori des modèles  $\text{LogNorm}$ ,  $\text{Gamma} - \text{Par}$  et  $\text{Weib} - \text{Par}$  sont reportées sur la Figure 36. La méthode de sélection choisit le modèle lognormal sur certains mois contrairement aux résultats de la Section 6.2.2. L'estimation des quantiles d'ordre 95% et 99% via les modèles  $\text{LogNorm}$ ,  $\text{Gamma} - \text{Par}$ , et  $\text{Weib} - \text{Par}$  est comparée à l'estimation

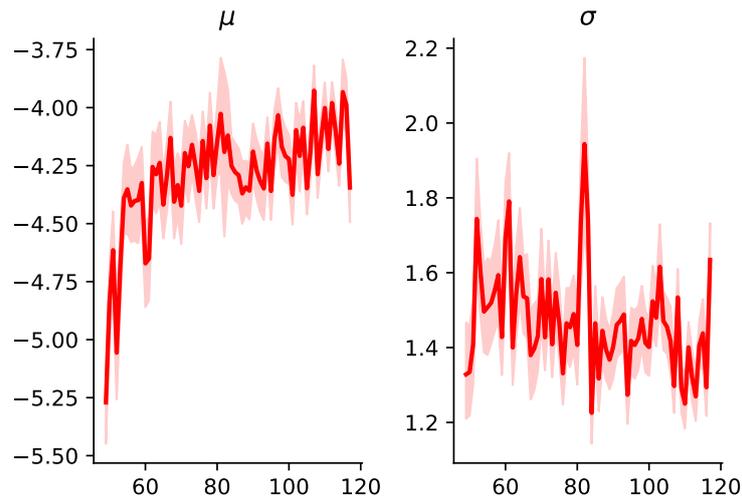


Figure 35 – Moyenne et quantiles d’ordre 5% et 95% de la loi a posteriori des paramètres du modèle  $\text{LogNorm}(\mu, \sigma)$  pour chaque mois.

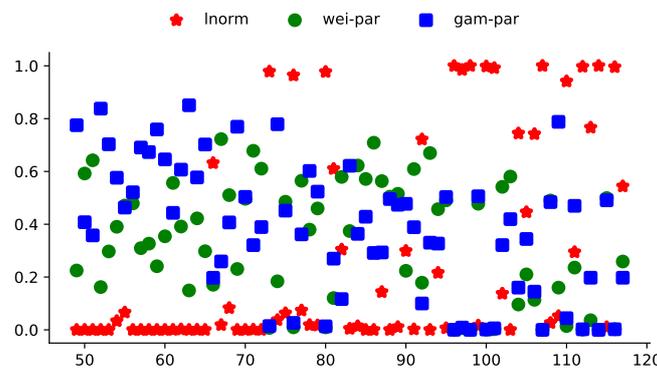


Figure 36 – Probabilités a posteriori des modèles  $\text{LogNorm}(\mu, \sigma)$ ,  $\text{Gamma}(r) - \text{Par}(\alpha, \theta)$ , et  $\text{Weib}(r) - \text{Par}(\alpha, \theta)$  pour chaque mois.

des quantiles empiriques sur la [Figure 37](#). L’estimation des quantiles par le modèle  $\text{LogNorm}$  se rapproche d’avantage des quantiles empiriques. La probabilité a posteriori du modèle  $\text{LogNorm}$  permet de rapprocher l’estimation des quantiles via la méthode BMA des quantiles empiriques.

#### 6.4 Discussion des résultats de l’analyse sur données réelles

L’utilisation d’une loi de Pareto pour modéliser la partie atypique de la distribution des montants de sinistres ne semble pas adaptée aux données étudiées dans cette section. Cette hypothèse conduit à spécifier une queue de distribution trop épaisse. Une solution consiste à considérer d’autres lois telles que la loi de Burr ou de Stoppa pour la queue de de la distribution, voir les travaux de [Abu Bakar et al. \[2015\]](#), [Calderín-Ojeda and Kwok \[2015\]](#) et [Grün and Miljkovic \[2019\]](#).

L’utilisation de la fonction de vraisemblance pour calibrer et sélectionner les modèles donne trop d’importance au ventre de la distribution des montants de sinistres. Si l’objectif final est une bonne estimation des quantiles

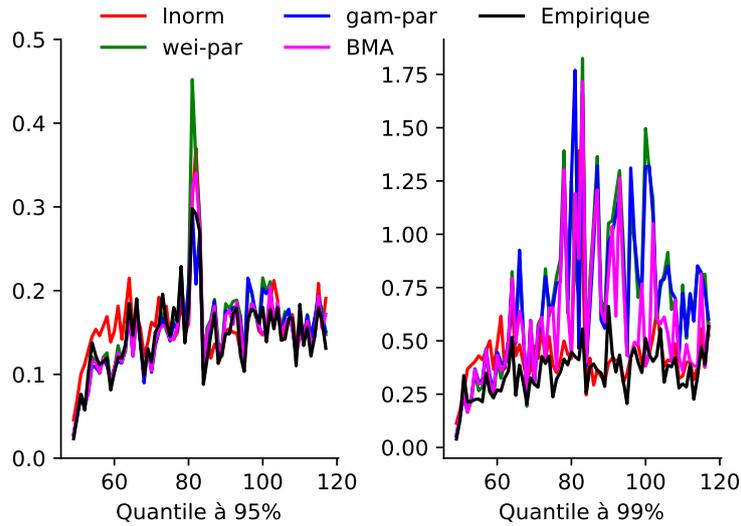


Figure 37 – Quantiles à 95% et 99% estimés empiriquement et via les modèles  $\text{LogNorm}(\mu, \sigma)$ ,  $\text{Gamma}(r) - \text{Par}(\alpha, \theta)$ , et  $\text{Weib}(r) - \text{Par}(\alpha, \theta)$  pour chaque mois.

d'ordre élevé, une méthode visant à minimiser les écarts entre les quantiles des modèles composites et les quantiles empiriques serait plus adaptée. De telles méthodes existent dans la littérature statistique, les travaux de [Bernton et al. \[2019a\]](#) considère un méthode d'estimation paramétrique visant à minimiser la distance de Wasserstein. Dans le cas de données i.i.d. et univariées, cette distance est exactement une distance interquantile. Il est possible d'adopter une approche Bayésienne en utilisant un algorithme de type *Approximate Bayesian Computation* (ABC), cette méthode est devenue très populaire au cours de ces vingt dernières années du fait de leur simplicité et de leur large champ d'application. Les algorithmes ABC sont l'objet de l'ouvrage de [Sisson et al. \[2018\]](#). La distance de Wasserstein est incontournable dans les problèmes de transport optimal et est aussi de plus en plus utilisée en apprentissage statistique. L'emploi de la distance de Wasserstein dans le cadre des algorithmes ABC a été proposé par [Bernton et al. \[2019b\]](#), une illustration en assurance est proposée dans [Goffard and Laub \[2021\]](#).

Parmi les modèles considérés, le plus adapté semble être le modèle lognormal. Ce fait révèle une limitation des modèles composites car l'adéquation des données au modèle lognormale rend l'estimation du modèle lognormale-Pareto difficile. Le paramètre seuil prend des valeurs dépassant le maximum des données ce qui rend les lois a posteriori des paramètres de seuil et de queue inexploitables. Ce résultat peut-être vérifié empiriquement en calibrant les modèles composites sur des données générées par une loi lognormale. Les modèles  $\text{Gamma}(r) - \text{Par}(\alpha, \theta)$ ,  $\text{LogNorm}(\sigma) - \text{Par}(\alpha, \theta)$ , et  $\text{Weib}(r) - \text{Par}(\alpha, \theta)$  sont ajustés sur des données générées par un modèle  $\text{LogNorm}(\mu = 0, \sigma = 1)$  via l'algorithme **SMC**. Les hypothèses a priori pour la partie attritionnelle sont données par

$$r \sim \text{Gamma}(0.1, 10), \sigma \sim \text{Gamma}(0.1, 10), k \sim \text{Gamma}(0.1, 10),$$

Les hypothèses a priori pour la partie extrême sont identiques pour chacun des modèles

$$\alpha \sim \text{Gamma}(0.1, 10), \theta \sim \text{Gamma}(0.1, 10).$$

La [Figure 38](#) montre les lois a posteriori obtenues pour les paramètres de chacun des modèles composites

pour un nombre d'observations égal à 500. Les estimateurs du maximum de vraisemblance (en pointillé sur la

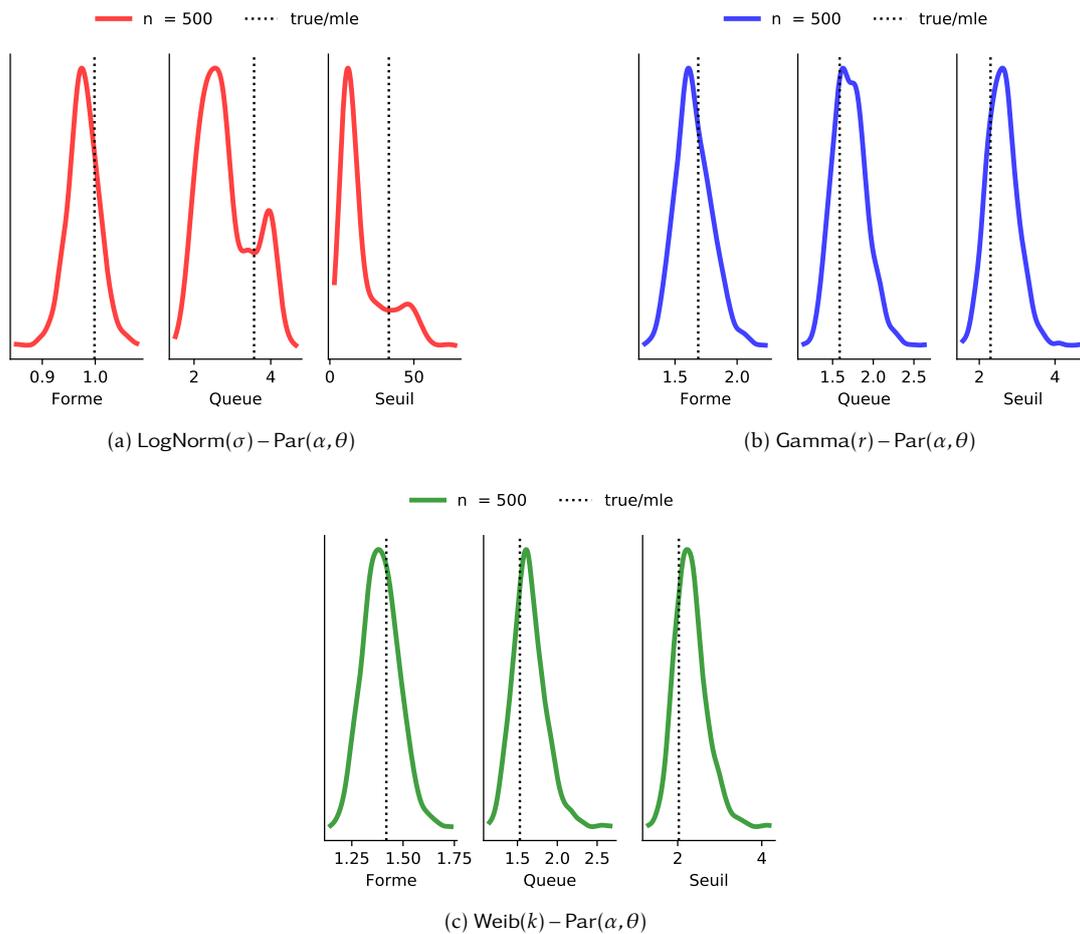


Figure 38 – Loi a posteriori des paramètres des modèles composites pour les données de pertes automobiles australiennes.

Figure 38) ont été calculés sur un échantillon de taille 10,000 et jouent le rôle de valeur de référence. Les lois a posteriori sont de bonne facture pour les modèles gamma-Pareto (Figure 38b) et Weibull-Pareto (Figure 38c). La loi a posteriori pour le modèle lognormal-Pareto (Figure 38a) ressemble à celle obtenue lors de l'application aux données de pertes australiennes, voir la Figure 18a.

Les lois a posteriori associées aux paramètres des modèles composites sont fiables lorsque l'algorithme SMC est appliqué à l'ensemble des données et aux données segmentées annuellement. Le nombre d'observations par mois est parfois insuffisant et mène à des lois a posteriori avec une grande variance. L'efficacité de l'échantillonnage peut-être améliorée en recherchant des lois a priori conjuguées pour les modèle composite, à l'image du travail réalisé par Cooray and Cheng [2013]. Le problème est qu'une étude au cas par cas est nécessaire. La procédure implémentée dans le cadre de ce mémoire s'adapte à n'importe quel modèle composite sous réserve de connaître sa fonction de vraisemblance.

Ces différentes pistes d'amélioration sont reprises dans la section suivante dédiée aux conclusions et perspectives de l'étude.

## 7 Conclusion et perspectives

Ce travail décrit et implémente un algorithme de Monte Carlo séquentiel adapté à l'échantillonnage suivant la loi a posteriori des modèles composites utilisés pour la modélisation du montant des sinistres en assurance non vie. Cet algorithme permet la calibration et la sélection du modèle le mieux adapté aux données dans un cadre bayésien. Le code python est disponible sur le répertoire github suivant <https://github.com/LaGauffre/SMCCompoMo>.

L'approche bayésienne, par rapport à l'approche fréquentiste du maximum de vraisemblance utilisée habituellement pour ces modèles, permet la prise en compte de l'incertitude autour de l'estimation du paramètre et l'incorporation d'informations a priori. L'algorithme de Monte Carlo Séquentiel admet trois avantages par rapport à l'algorithme Monte Carlo par chaine de Markov standard. (1) Il s'affranchit du choix délicat du paramètre de variance de la marche aléatoire qui gouverne la précision de l'échantillonnage. (2) Il fournit une approximation de la constante de normalisation de la loi a posteriori très pratique pour sélectionner ou combiner les modèles en fonction de leur adéquation aux données. (3) Il est très facile à paralléliser pour tirer meilleur parti des capacités des processeurs multi-coeurs des ordinateurs actuels et des serveurs de calcul.

L'étude sur des données simulées a montré que l'algorithme permet d'identifier le modèle générateur des données ainsi que ses paramètres.

Les modèles composites permettent de résumer la distribution des montants de sinistres et d'analyser l'évolution de la sinistralité au cours du temps. L'utilisation de la loi de Pareto pour modéliser la queue de la distribution a conduit à surestimer les quantiles d'ordre élevé (99%) sur les deux jeux de données réelles considérés. Des distributions alternatives sont envisageables comme la loi de Burr ou la loi de Stoppa, voir les travaux de [Abu Bakar et al. \[2015\]](#), [Calderín-Ojeda and Kwok \[2015\]](#) et [Grün and Miljkovic \[2019\]](#). Il suffit d'encoder les fonctions de vraisemblance et d'utiliser les méthodes Python implémentées dans le cadre de ce mémoire.

Les modèles paramétriques ont le potentiel de fiabiliser l'estimation des quantiles d'ordre élevé lorsque peu d'observations sont disponibles. La multiplication des modèles et leur combinaison pour produire un estimateur unique via une méthode d'agrégation telle que la méthode *Bayesian Model Averaging* est une piste intéressante. Il est possible de considérer des méthodes d'agrégation d'estimateurs mises au point récemment et reposant sur la validation croisée, voir les travaux de [Yao et al. \[2018\]](#).

Si l'objectif final est l'estimation des quantiles, alors une procédure d'estimation basée sur la minimisation des écarts aux quantiles empiriques pourrait s'avérer plus adaptée. Les propriétés théoriques de tels estimateurs ont été investiguées dans les travaux de [Bernton et al. \[2019a\]](#) et [Bernton et al. \[2019b\]](#), un futur travail de recherche visera à leur application sur des données de pertes en assurance.

## References

- S. A. Abu Bakar, N. A. Hamzah, M. Maghsoudi, and S. Nadarajah. Modeling loss data using composite models. *Insurance: Mathematics and Economics*, 61:146 – 154, 2015. ISSN 0167-6687. doi: <https://doi.org/10.1016/j.insmatheco.2014.08.008>. URL <http://www.sciencedirect.com/science/article/pii/S0167668714001024>.
- H. Akaike. Information theory and an extension of the maximum likelihood principle. In *Springer Series in Statistics*, pages 199–213. Springer New York, 1998. doi: 10.1007/978-1-4612-1694-0\_15.
- M. El Baarani. Modélisation du risque atypique pour la responsabilité civile automobile: nouvelle approche pour la détermination du seuil. Master’s thesis, EURIA, 2017.
- G. Beirlant, Segers J. L., and Teugels D. *Statistics of Extremes*. John Wiley & Sons, 2004. ISBN 0471976474. URL [https://www.ebook.de/de/product/3611778/beirlant\\_goegebeur\\_seggers\\_jozef\\_1\\_teugels\\_daniel\\_de\\_waal\\_statistics\\_of\\_extremes.html](https://www.ebook.de/de/product/3611778/beirlant_goegebeur_seggers_jozef_1_teugels_daniel_de_waal_statistics_of_extremes.html).
- E. Bernton, P. E. Jacob, M. Gerber, and C. P. Robert. On parameter estimation with the wasserstein distance. *Information and Inference: A Journal of the IMA*, 8(4):657–676, oct 2019a. doi: 10.1093/imaiai/iaz003.
- E. Bernton, P. E. Jacob, M. Gerber, and C. P. Robert. Approximate bayesian computation with the wasserstein distance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(2):235–269, feb 2019b. doi: 10.1111/rssb.12312.
- T. Bouché. Modèle de propension des assurés par rapport aux risques de sinistres corporels graves en assurance automobile. Master’s thesis, EURIA, 2014.
- O. Bunke and X. Milhaud. Asymptotic behavior of bayes estimates under possibly incorrect models. *The Annals of Statistics*, 26(2), apr 1998. doi: 10.1214/aos/1028144851.
- E. Calderín-Ojeda and C. F. Kwok. Modeling claims data with composite stoppa models. *Scandinavian Actuarial Journal*, 2016(9):817–836, apr 2015. doi: 10.1080/03461238.2015.1034763.
- A. Charpentier. *Computational Actuarial Science with R*. Taylor & Francis Ltd., August 2014. URL [https://www.ebook.de/de/product/25483298/computational\\_actuarial\\_science\\_with\\_r.html](https://www.ebook.de/de/product/25483298/computational_actuarial_science_with_r.html).
- N. Chopin. A sequential particle filter method for static models. *Biometrika*, 89(3):539–552, aug 2002. doi: 10.1093/biomet/89.3.539.
- K. Cooray and M. Ananda. Modeling actuarial data with a composite lognormal-pareto model. *Scandinavian Actuarial Journal*, 2005(5):321–334, sep 2005. doi: 10.1080/03461230510009763.
- K. Cooray and C.-I. Cheng. Bayesian estimators of the lognormal–pareto composite distribution. *Scandinavian Actuarial Journal*, 2015(6):500–515, dec 2013. doi: 10.1080/03461238.2013.853368.
- A. C. Davison. *Statistical Models*. Cambridge University Press, October 2011. ISBN 0521773393. URL [https://www.ebook.de/de/product/4229672/a\\_c\\_davison\\_statistical\\_models.html](https://www.ebook.de/de/product/4229672/a_c_davison_statistical_models.html).

- P. Diaconis and D.S Ylvisaker. Conjugate priors for exponential families. *The Annals of Statistics*, 7(2):269–281, mar 1979. doi: 10.1214/aos/1176344611.
- C. Dutang and A. Charpentier. Casdatasets. *R package version*, pages 1–0, 2016.
- A. Gelman, J. Hwang, and A. Vehtari. Understanding predictive information criteria for bayesian models. *Statistics and Computing*, 24(6):997–1016, aug 2013a. doi: 10.1007/s11222-013-9416-2.
- A. J. B. Gelman, H. S. Carlin, S. D. B. Dunson, and A. Vehtari. *Bayesian Data Analysis*. Taylor & Francis Ltd, 2013b. ISBN 1439840954. URL <http://www.stat.columbia.edu/~gelman/book/>.
- F W Gerstengarbe and P C Werner. A method for the statistical definition of extreme-value regions and their application to meteorological time series. *Zeitschrift fuer Meteorologie; (German Democratic Republic)*, January 1989.
- P.-O. Goffard and P. Laub. Approximate Bayesian Computations to fit and compare insurance loss models. working paper or preprint, April 2021. URL <https://hal.archives-ouvertes.fr/hal-02891046>.
- B. Grün and T. Miljkovic. Extending composite loss models using a general framework of advanced computational tools. *Scandinavian Actuarial Journal*, 2019(8):642–660, apr 2019. doi: 10.1080/03461238.2019.1596151.
- B. M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5): 1163–1174, 1975. ISSN 00905364. URL <http://www.jstor.org/stable/2958370>.
- J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. Bayesian model averaging: A tutorial. *Statistical Science*, 14(4):382–401, 1999. ISSN 08834237. URL <http://www.jstor.org/stable/2676803>.
- L. Hong and R. Martin. A review of bayesian asymptotics in general insurance applications. *European Actuarial Journal*, 7(1):231–255, apr 2017. doi: 10.1007/s13385-017-0151-5.
- A. Jasra, D. A. Stephens, A. Doucet, and T. Tsagaris. Inference for Lévy-driven stochastic volatility models via adaptive sequential monte carlo. *Scandinavian Journal of Statistics*, 38(1):1–22, dec 2010. doi: 10.1111/j.1467-9469.2010.00723.x.
- R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, jun 1995. doi: 10.1080/01621459.1995.10476572.
- A. Kong, J. S. Liu, and W. H. Wong. Sequential imputations and bayesian missing data problems. *Journal of the American Statistical Association*, 89(425):278–288, mar 1994. doi: 10.1080/01621459.1994.10476469.
- D. J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter. Winbugs-a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, 10(4):325–337, 2000.
- P. Del Moral, A. Doucet, and A. Jasra. Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, jun 2006. doi: 10.1111/j.1467-9868.2006.00553.x.
- F.-Z. Naji. Nouvelle modélisation du risque extrême dans la tarification de la garantie incendie en assurance multirisques habitation. Master’s thesis, ENSAE, 2016.

- R. M. Neal. Annealed importance sampling. *Statistics and Computing*, 11(2):125–139, 2001. doi: 10.1023/a:1008923215028.
- A. Raoult. Modéliser la sinistralité en risques industriels : l’apport de la théorie des valeurs extrêmes. Master’s thesis, ISUP, 2016.
- G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120, feb 1997. doi: 10.1214/aoap/1034625254.
- J. Salvatier, T. V. Wiecki, and C. Fonnesbeck. Probabilistic programming in python using PyMC3. *PeerJ Computer Science*, 2:e55, apr 2016. doi: 10.7717/peerj-cs.55.
- G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2), mar 1978. doi: 10.1214/aos/1176344136.
- D. P. M. Scollnik. On composite lognormal-pareto models. *Scandinavian Actuarial Journal*, 2007(1):20–33, mar 2007. doi: 10.1080/03461230601110447.
- D. P. M. Scollnik and C. Chenchen Sun. Modeling with weibull-pareto models. *North American Actuarial Journal*, 16(2):260–272, apr 2012. doi: 10.1080/10920277.2012.10590640.
- P. Shevchenko and M. Wüthrich. The structural modeling of operational risk via bayesian inference: combining loss data with expert opinions. *The Journal of Operational Risk*, 1(3):3–26, 2006. doi: 10.21314/jop.2006.016.
- S. A. Sisson, Y. Fan, and M. A. Beaumont, editors. *Handbook of Approximate Bayesian Computation*. Chapman and Hall/CRC, sep 2018. doi: 10.1201/9781315117195.
- L. F. South, A. N. Pettitt, and C. C. Drovandi. Sequential monte carlo samplers with independent markov chain monte carlo proposals. *Bayesian Analysis*, 14(3):753–776, sep 2019. doi: 10.1214/18-ba1129.
- D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, oct 2002. doi: 10.1111/1467-9868.00353.
- M. Vegni. Modélisation du coût des sinistres extrêmes en assurance automobile. Master’s thesis, ISUP, 2011.
- Y. Wang, I. Hobæk Haff, and A. Huseby. Modelling extreme claims via composite models and threshold selection methods. *Insurance: Mathematics and Economics*, 91:257–268, mar 2020. doi: 10.1016/j.insmatheco.2020.02.009.
- S. Watanabe. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(116):3571–3594, 2010. URL <http://jmlr.org/papers/v11/watanabe10a.html>.
- Y. Xu. Modélisation des sinistres graves dans l’assurance santé internationale. Master’s thesis, ESSEC, 2017.
- Y. Yao, A. Vehtari, D. Simpson, and A. Gelman. Using stacking to average bayesian predictive distributions (with discussion). *Bayesian Analysis*, 13(3), sep 2018. doi: 10.1214/17-ba1091.