

---

---

RÉSUMÉ DE CE QUI A ÉTÉ FAIT DEPUIS VOTRE DÉPART.

---

---

SOUS LA SUPERVISION DE  
ÉTIENNE MARCEAU

PRÉPARÉ PAR  
ALEXANDRE LEPAGE,  
DIAMILATOU N'DIAYE,

LE 1<sup>er</sup> AOÛT 2019



UNIVERSITÉ  
**LAVAL**

FACULTÉ DES SCIENCES ET DE GÉNIE  
ÉCOLE D'ACTUARIAT  
UNIVERSITÉ LAVAL



## Table des matières

<b>1</b>	<b>Estimation de paramètres avec une copule archimédienne à hautes dimensions.</b>	<b>1</b>
<b>2</b>	<b>Revue de littérature sur le modèle collectif du risque avec structure de dépendance.</b>	<b>2</b>
2.1	Modèle de probabilités conditionnelles ( <i>Two-steps frequency-severity model</i> ) . . . . .	2
2.2	Copules . . . . .	3
2.3	L'effet aléatoire partagé ( <i>Shared random effect</i> ) . . . . .	3
<b>3</b>	<b>Conclusion</b>	<b>4</b>

## Liste des illustrations

1	Influence de $\alpha_0$ dans l'estimation de $\alpha_1$ par inversion du tau de Kendall . . . . .	1
2	Classification des articles lus sur le modèle collectif du risque avec dépendance. . . . .	2

# 1 Estimation de paramètres avec une copule archimédienne à hautes dimensions.

Dans les étapes précédentes du projet, nous avons trouvé que la méthode du maximum de vraisemblance complète donnait d'excellents résultats pour estimer les paramètres d'un échantillon simulé. Cependant, le principal désavantage de cette méthode sont les temps de dérivation et d'optimisation numérique. Nous avons donc testé la méthode du maximum de vraisemblance composite. Cette dernière estimait précisément les paramètres des lois marginales ainsi que le paramètre de dépendance de la loi mère de la copule archimédienne hiérarchique. Par contre, lorsque venait le temps de calculer le paramètre de la loi enfant, nous étions confronté au même problème qu'avec la méthode complète puisque les montants de sinistres sont tous regroupés ensemble dans une même copule archimédienne. Dans un contexte d'assurance, il n'est pas impossible de voir un nombre de sinistre supérieur à 10. Il faut donc trouver une façon alternative d'estimer le paramètre de dépendance de la loi enfant. Plusieurs solutions sont présentées dans [Hofert et al., 2013].

Pour débiter, la méthode des moments en inversant le tau de Kendall n'est pas une solution envisageable puisque, lorsque l'on utilise cette méthode, l'estimation du paramètre de dépendance de la loi enfant est fortement influencée par le niveau de dépendance induit par la loi mère comme le démontre l'illustration 1.

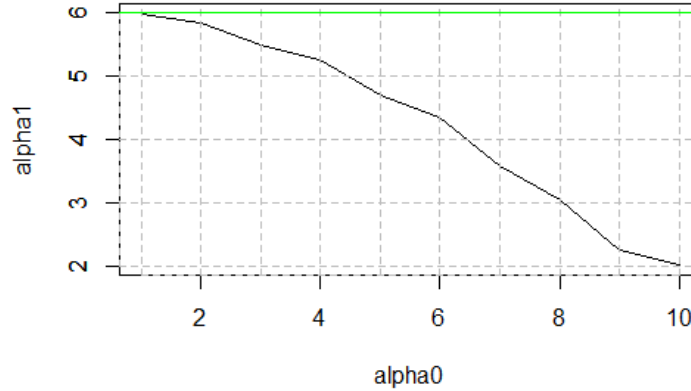


Illustration 1 – Influence de  $\alpha_0$  dans l'estimation de  $\alpha_1$  par inversion du tau de Kendall : Cette illustration est générée en simulant  $10 \times 10\,000$  observations du modèle collectif du risque présenté dans [Cossette et al., 2019] pour  $\alpha_0 = 1, \dots, 10$ ,  $\alpha_1$  demeure constant à 6 et la copule est logarithmique-logarithmique. La ligne verte correspond au vrai paramètre  $\alpha_1$  définit pour les simulations.

De son côté, Diamilatou a trouvé (1) afin de pouvoir dériver la copule en plusieurs étapes afin de réduire le temps de dérivation. D'ailleurs, il se trouve que (1) correspond à l'équation (9) de [Hofert et al., 2013].

$$c(\mathbf{u}) = \psi^{(d)} \left( \sum_{j=1}^d \psi^{-1}(u_j) \right) \times \prod_{j=1}^d (\psi^{-1})'(u_j) \quad (1)$$

Nous avons donc programmé cette équation. Celle-ci donne la possibilité d'imbriquer la fonction `R apply` dans l'expression de la dérivée ; ce qui permet d'augmenter l'efficacité des calculs en réduisant les temps d'optimisation numérique pour estimer les paramètres. Cependant le temps de dérivation demeure problématique puisqu'il faut toujours dériver  $d$  fois la fonction génératrice  $\psi(t) = \mathcal{L}_\Theta(t) = \mathcal{P}_M \circ \mathcal{L}_B(t)$ . Ce qui nous limite, de façon général, à 5 dérivées.

Il reste donc la solution de la dérivation numérique proposée par [Hofert et al., 2013], à la page 38 où il estime  $\psi^{(d)}$ , soit en évaluant numériquement (2) ou en simulant avec (3).

$$(-1)^d \psi^{(d)}(t) = \int_0^\infty \theta^d e^{-\theta t} dF_\Theta(\theta) = \mathbb{E} [\theta^d e^{-\theta t}], \quad t \in (0, \infty) \quad (2)$$

$$(-1)^d \psi^{(d)}(t) \approx \frac{1}{m} \sum_{k=1}^m \theta_k^d e^{-\theta_k t}, \quad t \in (0, \infty), \quad (3)$$

où  $\theta_k \sim F_\Theta, k \in \{1, \dots, m\}$ ,  $m$  est le nombre de simulations nécessaires à l'estimation de la dérivée. Il doit être raisonnablement grand (ex : 10 000). Cependant, lorsque l'on regarde (2), on voit qu'il faut connaître  $F_\Theta$ . À ce stade, nous n'avons pas tenté de trouver les fonctions de répartition de  $\Theta$  pour chacune des lois. Cependant, cela ne devrait pas être trop difficile et pourrait nous permettre de résoudre le problème de temps de dérivation. Par contre, autant pour l'intégration numérique que pour la simulation Monte Carlo, cette méthode augmenterait considérablement les temps de calcul. À ce moment-ci, l'utilisation d'une *GPU* pourrait permettre d'obtenir des temps de calcul raisonnables.

## 2 Revue de littérature sur le modèle collectif du risque avec structure de dépendance.

Après avoir essayé les solutions présentées dans la section 1, nous avons lu plusieurs articles pour prendre du recul et voir ce qui se faisait dans la littérature au sujet du modèle collectif du risque avec structure de dépendance. Pour débiter, le tableau 2 permet de comparer, d'un premier coup d'œil, les différents articles. Par la suite, un court résumé est fait pour chacun d'eux.

Articles	Variables aléatoires		Dépendance				
	N et M *	N et {Y1,...,YN}	Copule archimédiennes	Copules archimédiennes-hiérarchiques	Copule gaussienne/student	Effet aléatoire partagé **	Probabilités conditionnelles ***
Hernandez-Bastida et al., 2009		X				X	
Czado et al., 2012	X				X		
Krämer et al., 2013	X		X		X		
Shi et al., 2015	X		X				X
Baumgartner et al., 2015	X					X	
Garido et al., 2016	X						X
Frees et al., 2016	X				X		
Cossette et al., 2018		X		X			
Oh et al., 2019a		X			X		
Oh et al., 2019b	X				X	X	X

\* M désigne la moyenne des montants de sinistre

\*\* **Effet aléatoire partagé** (*Shared random effect*) : Dans un contexte de calcul de prime bayésienne, la dépendance est induite dans la paramétrisation des lois marginales. On utilise donc une variable aléatoire commune aux deux lois marginales, ou encore, on utilise une copule pour lier les distributions a posteriori.

\*\*\* **Probabilités conditionnelles** (*Two-step frequency-severity*) : On insère le résultat de N comme une variable explicative de M.

Illustration 2 – Classification des articles lus sur le modèle collectif du risque avec dépendance.

### 2.1 Modèle de probabilités conditionnelles (*Two-steps frequency-severity model*)

Les modèles de probabilités conditionnelles ou, communément appelés dans la littérature, *The two-steps frequency-severity models* sont, de loin, les modèles les plus simples autant dans l'interprétation que dans la paramétrisation puisqu'ils se résument simplement à utiliser la réalisation de la v.a.  $N$  comme une variable explicative du modèle linéaire généralisé (glm) de la v.a.  $M$ , où  $M = \frac{1}{n} \sum_{j=1}^n Y_j$ .

D'une part, [Garido et al., 2016] explique comment cette approche est mise en place. D'autres part, [Shi et al., 2015] compare ce type de modèle avec celui utilisant des copules. Finalement, dans ce dernier article, l'auteur vient à la conclusion que les modèles utilisant des copules sont beaucoup plus précis

du fait qu’il permette d’induire des relations de dépendance autres que linéaire et peuvent donc mieux s’adapter aux données.

## 2.2 Copules

En ce qui a trait aux copules, [Krämer et al., 2013] explique vraiment bien comment créer un modèle basé sur des copules. Il propose également une méthodologie pour choisir la copule à adopter en fonction du ratio de vraisemblance. Par la suite, [Czado et al., 2012] propose une méthode alternative à la maximisation de la vraisemblance standard pour estimer les paramètres du modèle. Cette méthode se fonde sur le score de Fisher et est utilisée par la fonction `R glm` pour estimer les paramètres des modèles linéaires généralisés. L’avantage de cette approche est qu’elle permet d’estimer un très grand quantité de paramètres à la fois en utilisant un nombre d’itération minimal.

Par la suite, [Oh et al., 2019a] présente une méthode alternative à [Cossette et al., 2019] pour modéliser la dépendance entre  $N$  et  $\{Y_1, \dots, Y_n\}$  en construisant une copule gaussienne avec une matrice de corrélation utilisant deux paramètres. Cette méthode est intéressante du fait qu’elle ne nécessite pas un long processus de dérivation pour trouver la densité de la copule. De plus, il est possible de travailler avec des réalisations de  $N$  qui peuvent prendre des valeurs élevées. Dans l’étude de cas présenté dans l’article, il y avait jusqu’à 12 sinistres pour un même assuré.

## 2.3 L’effet aléatoire partagé (*Shared random effect*)

La méthode de l’effet aléatoire partagé repose sur le calcul de la prime d’assurance en utilisant l’approche de crédibilité bayésienne. Se faisant, on peut induire de la dépendance entre les v.a. de fréquence et de sévérité à l’aide de leurs classes de risques respectives que l’on peut dénoter  $\Theta_N$  et  $\Theta_Y$ . Pour ce qui est de l’estimation des paramètres il s’agit d’avoir recours à la méthode de Monte Carlo pour simuler des chaînes de Markov. Dans la littérature, cela est appelé la méthode MCMC (Voir [Brooks et al., 2011] et [Roberts et al., 2004]).

Dans [Hernández-Bastida et al., 2009], on utilise la famille de distributions bivariées de Samarnov-Lee, dont la copule FGM fait partie, pour créer une structure de dépendance entre  $\Theta_N$  et  $\Theta_Y$ . De cette façon, les réalisations de  $Y$  sont liées entre elles puisqu’elles partagent la même classe de risque et un lien de dépendance est créé entre les v.a.  $N$  et  $Y$  grâce à la copule qui lie leurs distributions a posteriori. De plus, l’avantage d’induire la dépendance à travers la classes de risque permet de présumer l’indépendance des montants de sinistre conditionnellement à cette classe. Ainsi, on peut utiliser la relation  $\mathbb{E}[S|\Theta_N = \theta_N, \Theta_Y = \theta_Y] = \mathbb{E}[N|\Theta_N = \theta_N] \times \mathbb{E}[Y|\Theta_Y = \theta_Y]$  sans perdre de précision en agrégeant les réalisation de  $Y$ .

Par la suite, [Baumgartner et al., 2015] présente une façon alternative d’utiliser l’effet aléatoire partagé. En effet, dans l’article, plutôt que d’utiliser une copule pour créer la relation entre  $N$  et  $Y$ , on ajoute une variable explicative commune aux modèles linéaires généralisés des distributions marginales qui servira exclusivement à insérer de la dépendance entre les variables aléatoires. Cette façon de faire permet une simplification du modèle présenté par [Hernández-Bastida et al., 2009] de même que la possibilité d’avoir plusieurs niveaux de dépendance différents selon la classe de risque. Finalement, lors d’une analyse de cas, l’auteur compare son modèle à celui présenté par [Krämer et al., 2013] pour venir à la conclusion que le sien était plus précis et plus flexible.

En ce qui a trait à [Oh et al., 2019b], non seulement on reprend un modèle similaire à [Hernández-Bastida et al., 2009] dans le sens où on utilise une copule gaussienne pour créer un lien de dépendance entre les classes de risque de  $N$  et  $M$ , mais, en plus, on ajoute la réalisation de  $N$  comme variable explicative dans le glm de la distribution a priori de  $M$ . Autrement dit, l’auteur a combiné deux approches pour intégrer une structure de dépendance entre les composantes de son modèle. Est-ce que cette méthode accorde trop de poids à la dépendance? Il faudrait faire le test en comparant avec l’approche de [Hernández-Bastida et al., 2009] ou de [Baumgartner et al., 2015].

### 3 Conclusion

Dans l'ensemble, les modèles avec effet aléatoire partagé semblent constituer l'approche la plus prometteuse en termes de précision. Cependant, elle est aussi la plus compliquée à mettre en place. Le modèle proposé dans [Cossette et al., 2019] offre une approche alternative aux modèles traditionnels impliquant des copules du fait qu'elle considère un lien de dépendance possible entre les réalisations de  $Y$ . Cependant, cette approche ne considère pas la possibilité que la dépendance peut varier selon les groupes de risques des assurés et elle rencontre des problèmes dans l'estimation des paramètres lorsque la v.a.  $N$  peut prendre de grandes valeurs. Pour cette raison, il semblerait qu'il serait préférable d'adopter un modèle basé sur l'effet aléatoire partagé. Il reste cependant à voir si cette approche peut aussi bien servir à calculer la réserve actuarielle qu'elle le fait pour déterminer les primes ou bien si elle est trop complexe à mettre en place ; auquel cas, l'approche proposée par [Oh et al., 2019a] constituerait une alternative intéressante.

## Références

- [Baumgartner et al., 2015] Baumgartner, C., Gruber, L. F., and Czado, C. (2015). Bayesian total loss estimation using shared random effects. *Insurance : Mathematics and Economics*, 62 :194–201.
- [Brooks et al., 2011] Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). *Handbook of markov chain monte carlo*. CRC press.
- [Cossette et al., 2019] Cossette, H., Marceau, E., and Mtalai, I. (2019). Collective risk models with dependance.
- [Czado et al., 2012] Czado, C., Kastenmeier, R., Brechmann, E. C., and Min, A. (2012). A mixed copula model for insurance claims and claim sizes. *Scandinavian Actuarial Journal*, 2012(4) :278–305.
- [Garrido et al., 2016] Garrido, J., Genest, C., and Schulz, J. (2016). Generalized linear models for dependent frequency and severity of insurance claims. *Insurance : Mathematics and Economics*, 70 :205–215.
- [Hernández-Bastida et al., 2009] Hernández-Bastida, A., Fernández-Sánchez, M., and Gómez-Déniz, E. (2009). The net bayes premium with dependence between the risk profiles. *Insurance : Mathematics and Economics*, 45(2) :247–254.
- [Hofert et al., 2013] Hofert, M., Mächler, M., and McNeil, A. J. (2013). Archimedean copulas in high dimensions : Estimators and numerical challenges motivated by financial applications. *Journal de la Société Française de Statistique*, 154(1) :25–63.
- [Krämer et al., 2013] Krämer, N., Brechmann, E. C., Silvestrini, D., and Czado, C. (2013). Total loss estimation using copula-based regression models. *Insurance : Mathematics and Economics*, 53(3) :829–839.
- [Oh et al., 2019a] Oh, R., Ahn, J. Y., and Lee, W. (2019a). On copula-based collective risk models. *arXiv preprint arXiv :1906.03604*.
- [Oh et al., 2019b] Oh, R., Shi, P., and Ahn, J. Y. (2019b). Implementation of frequency-severity association in bms ratemaking. *arXiv preprint arXiv :1903.05851*.
- [Roberts et al., 2004] Roberts, G. O., Rosenthal, J. S., et al. (2004). General state space markov chains and mcmc algorithms. *Probability surveys*, 1 :20–71.
- [Shi et al., 2015] Shi, P., Feng, X., and Ivantsova, A. (2015). Dependent frequency–severity modeling of insurance claims. *Insurance : Mathematics and Economics*, 64 :417–428.