Estimation des paramètres d'une copule pour un modèle de risque collectif

Sous la supervision de
Étienne Marceau

Rapport des travaux réalisés

Préparé par

Alexandre Lepage,
Diamilatou N'diaye,
Amedeo Zito

le 06 juin 2019

Faculté des sciences et de génie
École d'actuariat
Université Laval
Automone 2018

Table des matières

Liste des illustrations

Liste des tableaux

1 Notions préliminaires

1.1 Modèle collectif du risque

Dans [Cossette et al., 2019b], on propose un modèle collectif du risque dont les composantes sont dépendants entre elles. Ce modèle s'exprime comme suit :

Soient N, la v.a. du nombre de sinistres, tel que $N \in \mathbb{N}$, et $\{X_i, i = 1, \ldots, n\}$, la v.a. de la sévérité des sinistres, tel que $X_i \sim X \in \mathbb{R}^+$, on a

$$S = \sum_{i=1}^{\infty} X_i \times \mathbb{1}_{\{N \geq i\}}, \tag{1}$$

où N est dépendant de $\underline{X}$ et les X_i sont dépendants entre eux.

Maintenant, on cherche à estimer les paramètres afférents à un tel modèle. Pour y arriver, [Cossette et al., 2019a] propose une approche de calcul de la vraisemblance par décomposition hiérarchique et une autre méthode, dite plus classique, qui prend l'ensemble du modèle. Dans le cas présent, compte tenu que la variable dénombrante est discrète et que les variables de sinistres sont continues, il est plus simple d'utiliser la méthode de vraisemblance complète. Le présent rapport présente donc les résultats d'une telle approche.

1.2 Fonction du maximum de vraisemblance

Soit une copule de Clayton multivariée dénotée C et représentée par

$$C(u_0, \ldots, u_n; \alpha) = (u_0^{-\alpha} + \cdots + u_n^{-\alpha} - n)^{-\frac{1}{\alpha}}. \tag{2}$$

Avec le modèle posé dans la section 1.1, on peut trouver la fonction de densité de la copule C de la manière suivante :

Soient λ et β les paramètres des lois de fréquence et de sévérité respectivement, ainsi que α, le paramètre de la copule de Clayton, alors la densité conjointe des v.a. N et $(X_1, \ldots X_N)$ est donnée par

$$\begin{aligned} f_{N,X_1,\ldots,X_n}(n, x_1, \ldots, x_n; \lambda, \beta, \alpha) = {} & \frac{\partial^n}{\partial x_1 \ldots \partial x_n} C(F_N(n; \lambda), F_{X_1}(x_1; \beta), \ldots, F_{X_n}(x_n; \beta); \alpha) \\ & - \frac{\partial^n}{\partial x_1 \ldots \partial x_n} C(F_N(n-1; \lambda), F_{X_1}(x_1; \beta), \ldots, F_{X_n}(x_n; \beta); \alpha). \end{aligned} \tag{3}$$

Puisque la dérivation en chaîne nécessaire pour trouver (3) est excessivement longue à calculer, on peut utiliser `R` à l'aide de la fonction `Deriv` [1] du package du même nom pour y arriver. Cependant, il est important de noter que, même avec cet outil, si la loi de fréquence admet des valeurs supérieures à 5, le temps de calcul peut grimper très rapidement.

Soient n_k, le nombre d'observations où $N = k$, et $x_{i,k}$ représente le k-ième sinistre de la i-ième observation, la fonction de vraisemblance du modèle énoncé dans la section 1.1 est donné par

$$\begin{aligned} \mathcal{L}(\lambda, \beta, \alpha) = {} & (\Pr(N = 0; \lambda))^{n_0} \\ & \times \prod_{i=1}^{n_1} f_{N,X_1}(1, x_{i,1}; \lambda, \beta, \alpha) \\ & \times \prod_{i=1}^{n_2} f_{N,X_1,X_2}(2, x_{i,1}, x_{i,2}; \lambda, \beta, \alpha) \\ & \ldots \\ & \times \prod_{i=1}^{n_k} f_{N,X_1,\ldots,X_k}(k, x_{i,1}, \ldots, x_{i,k}; \lambda, \beta, \alpha). \end{aligned}$$

1. https ://cran.r-project.org/web/packages/Deriv/Deriv.pdf

Par la suite, afin d'estimer les paramètres, il faut avoir recours à une méthode d'optimisation numérique. Cela implique de trouver le minimum d'une fonction strictement décroissante. Pour cette raison, il faut minimiser la log-vraisemblance négative.

On obtient donc

$$\begin{aligned}\ell(\lambda, \beta, \alpha) = &-n_0 \times \ln\left(\Pr(N=0;\lambda)\right) \\ &- \sum_{i=1}^{n_1} \ln\left(f_{N,X_1}(1, x_{i,1}; \lambda, \beta, \alpha)\right) \\ &- \sum_{i=1}^{n_2} \ln\left(f_{N,X_1,X_2}(2, x_{i,1}, x_{i,2}; \lambda, \beta, \alpha)\right) \\ &\dots \\ &- \sum_{i=1}^{n_k} \ln\left(f_{N,X_1,\dots,X_k}(k, x_{i,1}, \dots, x_{i,k}; \lambda, \beta, \alpha)\right). \end{aligned} \tag{4}$$

Finalement, on trouve les paramètres avec `R` en utilisant la fonction `constrOptim` [2].

1.3 Copule archimédienne hiérarchique

Comme les copules archimédiennes hiérarchiques offrent une bonne flexibilité dans la modélisation de la dépendance, la présente section se penchera sur ce sujet d'intérêt.

Les copules archimédiennes hiérarchiques avec des distributions multivariées composées sont décrites dans [Cossette et al., 2017].
La représentation graphique du modèle collectif du risque tel que décrit dans la section 1.1 et expliquée dans [Cossette et al., 2019b] est présentée dans l'illustration 1.

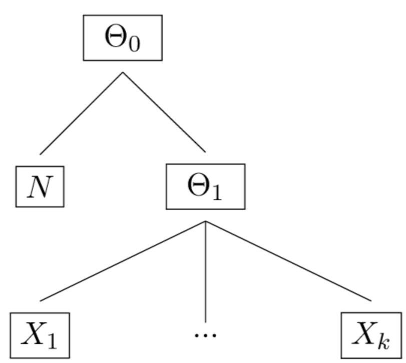

Illustration 1 – Arbre hiérarchique à un niveau.

Ici, Θ_0 représente une variable aléatoire définie sur $\mathbb{N}^+$ qui sert à créer un lien de dépendance entre la variable N et les $\{X_i, i = 1, \dots, k\}$. Ensuite, on a $\theta_1 = \sum_{i=1}^{\theta_0} B_i$ qui sert à créer un lien de dépendance entre les X_i. Les B_i sont i.i.d. et indépendants de θ_0. B_i peut suivre des distributions discrètes comme continues en excluant les valeurs inférieurs ou égales à 0.

Sous cette représentation, on obtient une copule archimédienne hiérarchique s'exprimant comme

$$C(u_0, u_1, \dots, u_n) = \mathscr{L}_{\theta_0}\left(\mathscr{L}_{\theta_0}^{-1}(u_0) - \ln\left(\mathscr{L}_B\left(\sum_{i=1}^{n} \mathscr{L}_B^{-1}\left(\exp\left(-\mathscr{L}_{\theta_0}^{-1}(u_i)\right)\right)\right)\right)\right), \tag{5}$$

où $\mathscr{L}$ correspond à la transformée de Laplace d'une variable aléatoire.

À ce point-ci, afin de trouver les paramètre d'une telle copule, il suffit d'appliquer (3) et (4).

2. https ://stat.ethz.ch/R-manual/R-devel/library/stats/html/constrOptim.html

2 Résultats

La présente section explique les scénarios testés ainsi que les résultats obtenus avec la méthodologie expliquée dans la section 1.2.

2.1 Copule de Clayton

Pour débuter gentiment, prenons un modèle binomial-exponentiel avec une copule de Clayton dont la représentation est exprimée en (2). Soient $N \sim Binom(n, q)$ et $X_i \sim X \sim Exp(\beta)$, les résultats obtenus lors de tests sont exprimés dans le tableau 1.

		n	q	λ	α
Estimateurs		5.00	0.40	0.01	6.44
Vrais paramètres		5.00	0.40	0.01	6.00
Temps de calcul	236 sec.				

Tableau 1 – Résultats de l'estimation des paramètres avec une copule de Clayton, $N \sim Binom(n, q)$ et $X \sim Exp(\beta)$ suite à 10 000 simulations.

Suite aux résultats concluants du premier exemple, un modèle avec $N \sim Pois(\lambda)$ peut ajouter un défi en terme de temps de calcul puisque cette loi a un support pouvant atteindre de grands nombres. Soient $N \sim Binom(n, q)$ et $X_i \sim X \sim Exp(\beta)$, les résultats obtenus sont présentés dans le tableau 2.

		λ	β	α
Estimateurs		1.00	0.01	5.77
Vrais paramètres		1.00	0.01	6.00
Temps de calcul	82.47 sec.			

Tableau 2 – Résultats de l'estimation des paramètres avec une copule de Clayton, $N \sim Pois(\lambda)$ et $X \sim Exp(\beta)$ suite à 10 000 simulations.

Lorsque $N \sim Poisson(2)$, le temps de calcul dépasse trente minutes puisque le 99,9999 percentile de cette loi est de 12. Ainsi la méthode du maximum de vraisemblance nécessite de générer une liste de douze fonctions qui sont dérivées jusqu'à douze fois. Même avec la fonction `Deriv` de `R`, ce processus est extrêmement long.

Ce que l'on peut observer avec ces deux résultats, c'est, d'une part, qu'avec 10 000 simulations, on obtient des résultats très adéquats. Cependant, il faut soulever que les valeurs de départ pour l'optimisation de la fonction de vraisemblance sont les véritables valeurs ; ce qui peut faire en sorte que les valeurs sont plus précises qu'elles ne l'auraient été si les valeurs de départ avaient été autre. D'autre part, on peut observer que le temps de calcul augmente significativement si N peut prendre des valeurs supérieures à 5 et si le nombre de paramètres à estimer est grand.

2.2 Copule archimédienne hiérarchique

Désormais, prenons la copule qui nous intéresse vraiment : la copule archimédienne hiérarchique. Avec $\theta_0 \sim logarithmique\,(\gamma = 1 - \exp(-\alpha_0))$, $B \sim Gamma(1/\alpha_1, 1)$, $N \sim Binom(5, q)$ et $X_i \sim X \sim Exp(\beta)$ on obtient les résultats présentés dans le tableau 3.

		α_0	α_1	β	q
Estimateurs		0.50	5.17	0.01	0.40
Vrais Paramètres		0.50	5.00	0.01	0.40
Temps de calcul	821.4 sec.				

Tableau 3 – Résultats de l'estimation des paramètres avec une copule archimédienne hiérarchique, $N \sim Binomiale(5, q)$ et $X \sim Exp(\beta)$ suite à 10 000 simulations.

Dans cet exemple, on note que le temps de calcul est significativement supérieur à celui présenté dans le tableau 1. Ce phénomène est explicable du fait qu'il y a plus de paramètres à estimer et que la copule contient plusieurs fonctions imbriquées qui nécessitent plus d'étapes dans le processus de dérivation en chaîne.

3 Conclusion

Pour conclure, le modèle collectif du risque présenté dans [Cossette et al., 2019b] présente un défi dans l'estimation des paramètres par la méthode du maximum de vraisemblance tel que présentée dans la section 1.2 puisque la dérivation en chaîne sur un grand nombre de variables pose un problème de temps de calcul.

À cet effet, une solution envisageable pourrait être d'utiliser la vraisemblance par décomposition hiérarchique proposé dans [Cossette et al., 2019a] afin de trouver le paramètre de dépendance entre N et X_1, puis de se limiter à un nombre restreint de X_i (disons 5) afin d'estimer le paramètre de dépendance entre les X_i. Cependant, avec cette méthode le fait de travailler avec des variables continues et discrètes peut causer un problème. Pour ce qui est de trouver les paramètres des lois de N et X, la méthode du maximum de vraisemblance classique (de façon univariée) pourrait être envisagée.

Références

[Cossette et al., 2017] Cossette, H., Gadoury, S.-P., Marceau, E., and Mtalai, I. (2017). Hiearchical archimedean copulas trough multivariate compound distributions. *Insurance : Mathematics and Economics*, 76.

[Cossette et al., 2019a] Cossette, H., Gadoury, S.-P., Marceau, E., and Robert, C. Y. (2019a). Composite likelyhood estimation method for hierarchical archimedean copulas defined with multivariate compound distributions. *Journal of Multivariate Analysis*, 172.

[Cossette et al., 2019b] Cossette, H., Marceau, E., and Mtalai, I. (2019b). Collective risk models with dependance.