

---

---

# ESTIMATION DES PARAMÈTRES PAR LA MÉTHODE COMPOSITE

---

---

SOUS LA SUPERVISION DE  
HÉLÈNE COSSETTE  
ÉTIENNE MARCEAU

PRÉPARÉ PAR  
ALEXANDRE LEPAGE,  
DIAMILATOU N'DIAYE,

LE 03 JUILLET 2019



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



## Table des matières

<b>1</b>	<b>Estimation par la Méthode Composite</b>	<b>1</b>
<b>2</b>	<b>Première Approche : Méthode <i>IFM</i></b>	<b>1</b>
2.1	Comportement Asymptotique . . . . .	2
2.1.1	Comportement asymptotique des estimateurs pour une copule archimédienne hiérarchique Logarithmique-Gamma de paramètres ( $\alpha_0 = 6, \alpha_1 = 4$ ) . . . . .	3
<b>3</b>	<b>Seconde Approche : Méthode composite</b>	<b>4</b>
3.1	Comportement Asymptotique . . . . .	5
<b>4</b>	<b>Impact de la paramétrisation de la loi de fréquence</b>	<b>6</b>
<b>5</b>	<b>Impact du choix de la copule et des paramètres de dépendance</b>	<b>8</b>
<b>6</b>	<b>Conclusion</b>	<b>11</b>

## Liste des illustrations

1	Paramètre $\alpha_0$ . . . . .	2
2	Paramètre $\alpha_1$ . . . . .	3
3	Paramètre $\alpha_0$ . . . . .	3
4	Paramètre $\alpha_1$ . . . . .	4
5	Paramètre $\alpha_0$ . . . . .	5
6	Paramètre $\alpha_1$ . . . . .	6
7	Effet de la paramétrisation de la loi de fréquence sur le calcul du tau de Kendall empirique. . . . .	7
8	Influence de $\alpha_0$ dans l'estimation de $\alpha_1$ par inversion du tau de Kendall . . . .	10

## Liste des tableaux

1	Valeurs estimées de $q$ , $\beta$ , $\alpha_0$ , $\alpha_1$ ainsi que le temps de calcul par la première approche. . . . .	2
2	Valeurs estimées de $q$ , $\beta$ , $\alpha_0$ , $\alpha_1$ ainsi que le temps de calcul par la deuxième approche. Le temps de calcul de $q$ , $\beta$ et $\alpha_0$ est leur temps de calcul commun étant donné qu'ils sont estimés ensembles. . . . .	5
3	Résultats du premier scénario. . . . .	7
4	Résultats du deuxième scénario. . . . .	8
5	Six premières réalisations de l'exemple 1. . . . .	9
6	Sommaire des données simulées pour l'exemple 1. . . . .	9
7	Résultats du scénario 3 . . . . .	10
8	Résultats du scénario 4 . . . . .	11

## 1 Estimation par la Méthode Composite

Le but de ce rapport est de présenter les résultats d'estimation des paramètres d'un modèle collectif du risque tel que décrit dans [Cossette et al., 2019b] avec la méthode composite présentée dans [Cossette et al., 2019a]. Dans les exemples suivants,  $N \sim \text{Binom}(5, q)$ ,  $q = 0.2$  et  $X \sim \text{Exp}(\beta)$ ,  $\beta = 1/100$ . La copule hiérarchique étudiée est construite à partir d'une v.a. mère qui suit une loi géométrique de paramètre  $\alpha_0 = 0.6$  et d'une v.a. enfant qui suit une loi géométrique de paramètre  $\alpha_1 = 0.4$ . Les estimations sont réalisées sur un échantillon simulé de 10 000 observations.

## 2 Première Approche : Méthode IFM

La première approche consiste à estimer séparément tous les paramètres. Dans un premier temps, les paramètres  $q$  et  $\beta$  sont estimés par les estimateurs du maximum de vraisemblance de leur loi respectives.

$$\hat{q} = \frac{\sum_{i=0}^{n_{obs}} n_i}{5 \times n_{obs}},$$

$$\hat{\beta} = \frac{\sum_{i=0}^{n_{obs}} n_i}{\sum_{i=0}^{n_{obs}} \sum_{j=1}^{n_i} x_{i,j}}$$

où  $n_{obs}$  représente le nombre d'observations ( $i$ ) présents dans l'échantillon,  $n_i$  représente la  $i^e$  réalisation de  $N$  et  $x_{i,j}$  correspond au  $j^e$  sinistre de la  $i^e$  observation de la v.a.  $X$ .

Ensuite l'estimateur de  $\alpha_0$  est obtenu en maximisant la fonction de vraisemblance (1).

$$L(\alpha_0 | \hat{q}, \hat{\beta}) = \prod_{i=1}^{n_{NX}} \left( \frac{\partial}{\partial x} C(F_N(n_i | \hat{q}), F_X(x_i | \hat{\beta})) - \frac{\partial}{\partial x} C(F_N(n_i - 1 | \hat{q}), F_X(x_i | \hat{\beta})) \right) \quad (1)$$

où  $n_{NX}$  représente le nombre de couples  $(N_i, X_{i,j})$  présents dans l'échantillon.

De plus, avec le modèle collectif du risque défini par  $S = \sum_{j=1}^{\infty} X_j \times \mathbb{1}\{j \leq N\}$ , la copule reliant la v.a.  $N$  avec n'importe lequel des  $X_j$  s'exprime comme

$$C(u_0, u_1) = \mathcal{L}_M(\mathcal{L}_M^{-1}(u_0) + \mathcal{L}_M^{-1}(u_1)). \quad (2)$$

À noter que la copule décrite en (2) est archimédienne.

Et enfin, de la même manière que pour  $\alpha_0$ , l'estimateur de  $\alpha_1$  est obtenu en maximisant la fonction de vraisemblance (3).

$$L(\alpha_1 | \hat{\beta}, \hat{\alpha}_0) = \prod_{i=1}^{n_{XX}} \frac{\partial^2}{\partial x_1 \partial x_2} C(F_X(x_{1i} | \hat{\beta}), F_X(x_{2i} | \hat{\beta}) | \hat{\alpha}_0), \quad (3)$$

où  $n_{XX}$  représente le nombre de couples  $\{(X_{i,k}, X_{i,j}) : k \neq j\}$  présents dans l'échantillon.

$$C(u_1, u_2 | \hat{\alpha}_0) = \mathcal{L}_{\Theta | \hat{\alpha}_0}(\mathcal{L}_{\Theta | \hat{\alpha}_0}^{-1}(u_1) + \mathcal{L}_{\Theta | \hat{\alpha}_0}^{-1}(u_2)),$$

avec

$$\mathcal{L}_{\Theta|\hat{\alpha}_0}(t) = \mathcal{L}_{M|\hat{\alpha}_0}(-\ln(\mathcal{L}_B(t)))$$

et

$$\mathcal{L}_{\Theta|\hat{\alpha}_0}^{-1}(u) = \mathcal{L}_B^{-1}\left(\exp(-\mathcal{L}_{M|\hat{\alpha}_0}^{-1}(u))\right).$$

Le tableau 1 présente les valeurs des estimateurs.

	$q$	$\beta$	$\alpha_0$	$\alpha_1$
Valeur	0.2000	0.0100	0.6000	0.4000
Estimateur	0.1996	0.0087	0.3603	0.8545
Temps	0.0000	0.0000	1.3900	0.7900

Tableau 1 – Valeurs estimées de  $q$ ,  $\beta$ ,  $\alpha_0$ ,  $\alpha_1$  ainsi que le temps de calcul par la première approche.

On voit en regardant le tableau 1 que les résultats de l'estimation sont loins des vrais paramètres.

## 2.1 Comportement Asymptotique

La méthode précédente est répétée 100 fois afin d'observer le comportement asymptotique des estimateurs. Pour  $\alpha_0$ , on obtient une moyenne de 0.3556106 et un intervalle de confiance de niveau  $\alpha = 95\%$  de  $[0.3117377, 0.3994836]$ .

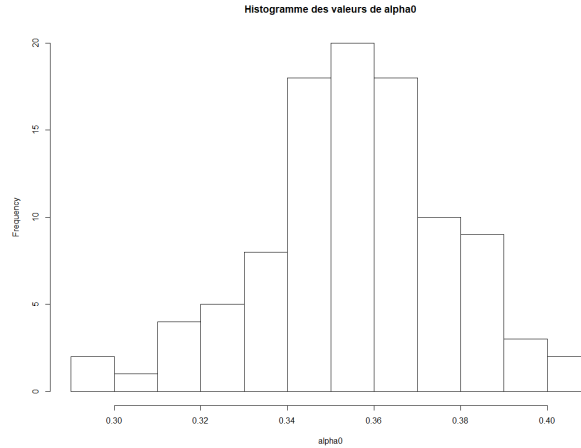


Illustration 1 – Histogramme des valeurs prises par  $\alpha_0$  sur 100 répétitions

Pour  $\alpha_1$  on obtient une moyenne de 0.8201792 et un intervalle de confiance de niveau  $\alpha = 95\%$  de  $[0.7704843, 0.8698740]$ .

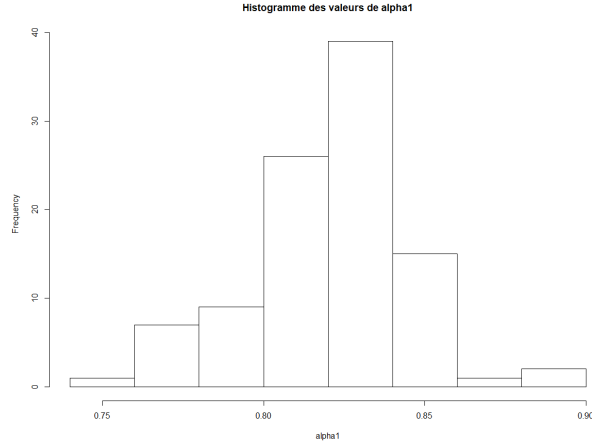


Illustration 2 – Histogramme des valeurs prises par  $\alpha_1$  sur 100 répétitions

### 2.1.1 Comportement asymptotique des estimateurs pour une copule archimédienne hiérarchique Logarithmique-Gamma de paramètres $(\alpha_0 = 6, \alpha_1 = 4)$

L'estimateur de  $\alpha_0$  produit une moyenne de 3.232702 et un intervalle de confiance de niveau  $\alpha = 95\%$  de  $[3.114684, 3.350721]$ .

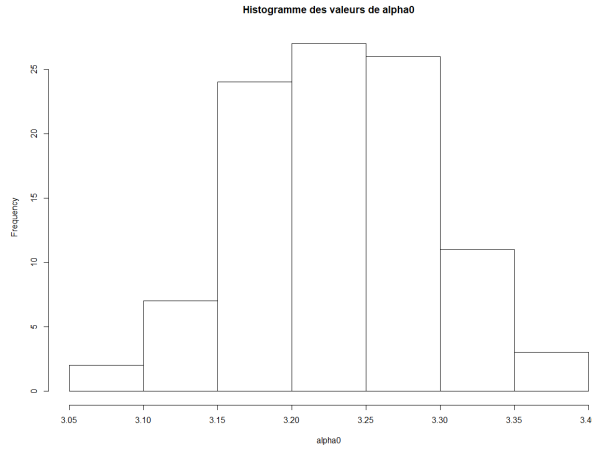


Illustration 3 – Histogramme des valeurs prises par  $\alpha_0$  sur 100 répétitions

L'estimateur de  $\alpha_1$  produit une moyenne de 6.646969 et un intervalle de confiance de niveau  $\alpha = 95\%$  de  $[6.094799, 7.199139]$ .



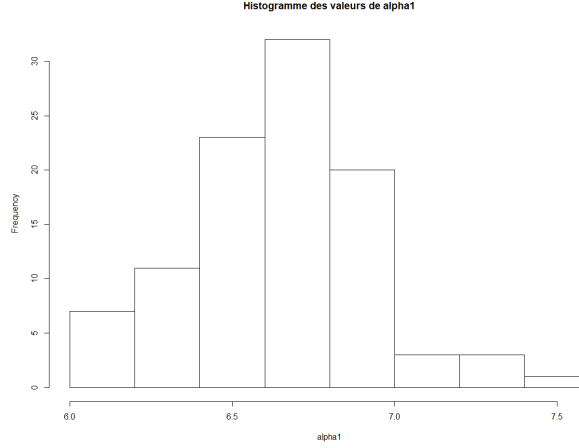


Illustration 4 – Histogramme des valeurs prises par  $\alpha_1$  sur 100 répétitions

Les estimations sont donc assez loin des vrais paramètres.

### 3 Seconde Approche : Méthode composite

La seconde approche consiste à estimer  $q$ ,  $\beta$  et  $\alpha_0$  en même temps puis  $\alpha_1$  par la suite. La fonction de vraisemblance à maximiser est :

$$L(\alpha_0, q, \beta) = \prod_{i=1}^{n_{NX}} \left( \frac{\partial}{\partial x} C(F_N(n_i), F_X(x_i)) - \frac{\partial}{\partial x} C(F_N(n_i - 1), F_X(x_i)) \right)$$

où  $n_{NX}$  représente le nombre de couples  $(N_i, X_{i,j})$  présents dans l'échantillon.

$$C(u_0, u_1) = \mathcal{L}_M(\mathcal{L}_M^{-1}(u_0) + \mathcal{L}_M^{-1}(u_1)).$$

Pour ce qui est de  $\alpha_1$ , la même estimation que la section précédente est réalisée.

$$L(\alpha_1, \beta, \alpha_0) = \prod_{i=1}^{n_{XX}} \frac{\partial^2}{\partial x_1 \partial x_2} C(F_X(x_{1i}), F_X(x_{2i})), \quad (4)$$

où

$$C(u_1, u_2) = \mathcal{L}_\Theta(\mathcal{L}_\Theta^{-1}(u_1) + \mathcal{L}_\Theta^{-1}(u_2)),$$

avec

$$\mathcal{L}_\Theta(t) = \mathcal{L}_M(-\ln(\mathcal{L}_B(t)))$$

et  $n_{XX}$  représente le nombre de couples  $(X_{i,k}, X_{i,j})$  présents dans l'échantillon. Le tableau 1 présente les valeurs des estimateurs ;

	$q$	$\beta$	$\alpha_0$	$\alpha_1$
Valeur	0.2000	0.0100	0.6000	0.4000
Estimateur	0.3589	0.0087	0.4972	0.8000
Temps	45.6000	45.6000	45.6000	1.4100

Tableau 2 – Valeurs estimées de  $q$ ,  $\beta$ ,  $\alpha_0$ ,  $\alpha_1$  ainsi que le temps de calcul par la deuxième approche. Le temps de calcul de  $q$ ,  $\beta$  et  $\alpha_0$  est leur temps de calcul commun étant donné qu'ils sont estimés ensembles.

Dans le tableau 2, on voit que le niveau de précision est légèrement meilleur que pour la première approche lorsque l'on regarde  $\alpha_0$  et  $\alpha_1$ , bien que ce ne soit toujours pas satisfaisant. Tandis que pour  $\hat{q}$ , les résultats se détériorent.

### 3.1 Comportement Asymptotique

L'estimation est ensuite répétée 100 fois afin d'observer le comportement asymptotique des estimateurs. L'estimateur de  $\alpha_0$  produit une moyenne de 0.4856781 et un intervalle de confiance de niveau  $\alpha = 95\%$  de  $[0.3888275, 0.5825286]$ .

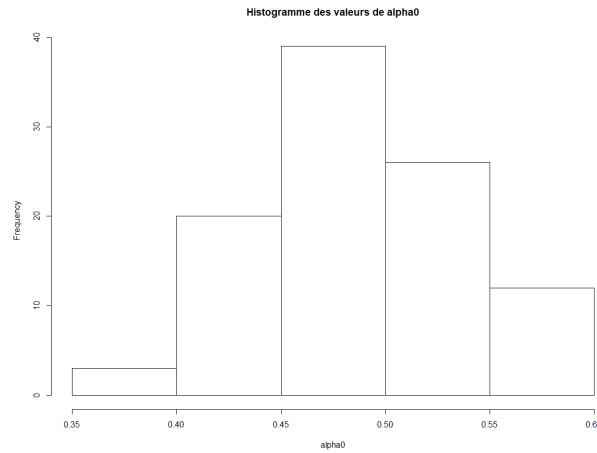


Illustration 5 – Histogramme des valeurs prises par  $\alpha_0$  sur 100 répétitions

L'estimateur de  $\alpha_1$  produit une moyenne de 0.7690707 et un intervalle de confiance de niveau  $\alpha = 95\%$  de  $[0.7103108, 0.8278306]$ .

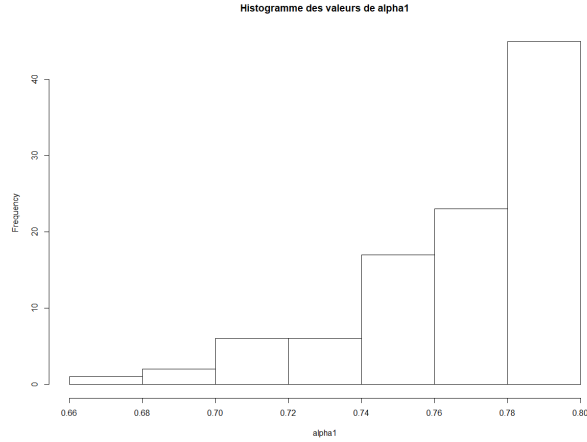


Illustration 6 – Histogramme des valeurs prises par  $\alpha_1$  sur 100 répétitions

## 4 Impact de la paramétrisation de la loi de fréquence

Dans [Nikoloulopoulos and Karlis, 2009], il est fait mention que la distribution marginale des variables aléatoires discrètes a un impact sur le calcul des taus de Kendall à cause de l'existence de valeurs identiques. Dans le scénario présenté dans la section 1, cette problématique se manifeste comme il est possible de le voir dans l'illustration 7. En effet, on y voit que plus le paramètre de probabilité  $q$  est petit, plus le tau de Kendall est sous-estimé dans le calcul empirique. Cela est assimilable au grand nombre d'observations de  $N$  qui sont égales à zéro. Ce problème se manifeste non seulement dans cette mesure de dépendance, mais aussi dans l'estimation des paramètres de la copule.

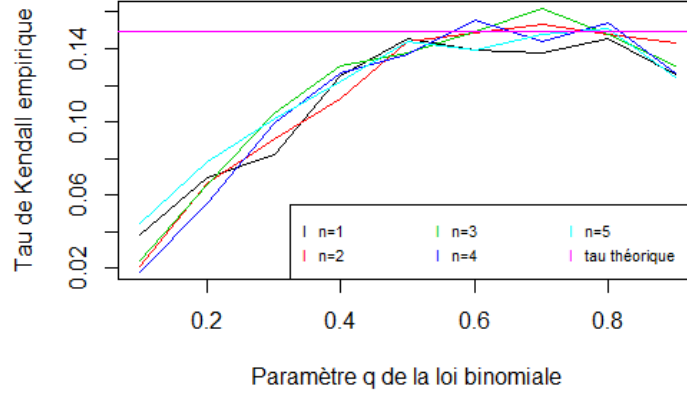


Illustration 7 – Effet de la paramétrisation de la loi de fréquence sur le calcul du tau de Kendall empirique : les graphiques sont réalisés avec  $30 \times 5 \times 10$  simulations de 100 observations d’une copule AMH avec une première variable aléatoire qui suit une loi binomiale( $n, q$ ) et une seconde qui suit une loi exponentielle de moyenne 100, où  $n = 1, \dots, 5$  et  $q = \frac{1}{10}, \frac{2}{10}, \dots, \frac{9}{10}$ . Le tau théorique, présenté sous forme de ligne droite rose, est calculé avec une loi binomiale( $n = 5, q = 3/5$ ). Ce dernier varie très peu en fonction de  $q$ , mais beaucoup en fonction de  $n$ .

Posons maintenant deux scénarios pour illustrer cette problématique sur l’estimation des paramètres dans le contexte d’un modèle collectif du risque décrit dans [Cossette et al., 2019b].

**Scénario 1.**  $N \sim \text{Binomiale}(5, 0.2)$

Pour le premier scénario, reprenons la paramétrisation décrite dans la section 1. On obtient donc le tableau de résultats présenté dans le tableau 3<sup>1</sup>.

	$q$	$\beta$	$\alpha_0$	$\alpha_1$
Vrais paramètres	0.2000	0.0100	0.6000	0.4000
Méthode <i>IFM</i>	0.1980	0.0090	0.4241	0.8407
Méthode composite	0.2972	0.0090	0.5487	0.7911
Méthode des moments	0.1980	0.0087	0.2752	0.4339

Tableau 3 – Résultats du premier scénario : La méthode dite *IFM* consiste à trouver les paramètres des distributions marginales, puis à estimer  $\hat{\alpha}_0$  sachant ces paramètres. De la même façon, on trouve  $\hat{\alpha}_1$  en fixant  $\beta$  et  $\alpha_0$  comme connus. La méthode dite *composite*, quant à elle, consiste à trouver tous les paramètres simultanément en utilisant la méthode décrite dans la section 3. Finalement, la méthode des moments, dans les cas de  $\hat{\alpha}_0$  et  $\hat{\alpha}_1$ , consiste à inverser le tau de Kendall empirique.

Dans le tableau 3, on voit que ni la méthode des moments ni la méthode *IFM* ne donne de bons résultats pour estimer  $\alpha_0$ . De ce fait, on ne peut obtenir de bons paramètres de départ

1. À noter que les résultats diffèrent légèrement de ceux présentés dans les tableaux 1 et 2 du fait que les simulations n’ont pas le même encrage (*seed*) et que le code utilisé est différent.

pour l'optimisation numérique de la méthode composite. Malgré tout, cette dernière donne un résultat relativement prêt du vrai paramètre. Cependant, lorsque l'on regarde l'estimation de  $q$  sous cette méthode, celle-ci s'éloigne du vrai paramètre.

Dans ce contexte (où on a une loi de fréquence qui a une masse à zéro élevée), l'approche recommandée est de trouver le paramètre de la distribution de la loi de fréquence avec sa fonction de vraisemblance marginale (en ignorant les autres paramètres), puis d'utiliser la méthode composite pour estimer les autres paramètres.

**Scénario 2.**  $N \sim \text{Binomiale}(5, 0.6)$

Pour le second scénario, on reprend la même paramétrisation que pour le premier, mais on remplace  $q = 0.2$  pour  $q = 0.6$ . On obtient alors les résultats présentés dans le tableau 4.

	$q$	$\beta$	$\alpha_0$	$\alpha_1$
Vrais paramètres	0.6000	0.0100	0.6000	0.4000
Méthode <i>IFM</i>	0.5984	0.0099	0.5980	0.6026
Méthode composite	0.6037	0.0100	0.5825	0.6117
Méthode des moments	0.5984	0.0094	0.5789	0.4427

Tableau 4 – Résultats du deuxième scénario : La méthode dite *IFM* consiste à trouver les paramètres des distributions marginales, puis à estimer  $\hat{\alpha}_0$  sachant ces paramètres. De la même façon, on trouve  $\hat{\alpha}_1$  en fixant  $\beta$  et  $\alpha_0$  comme connus. La méthode dite *composite*, quant à elle, consiste à trouver tous les paramètres simultanément en utilisant la méthode décrite dans la section 3. Finalement, la méthode des moments, dans les cas de  $\hat{\alpha}_0$  et  $\hat{\alpha}_1$ , consiste à inverser le tau de Kendall empirique.

Ce que l'on remarque dans le tableau 4, c'est que toutes les estimations sont plus près des vrais paramètres, même pour  $\alpha_1$ . Cela est explicable du fait qu'il y a plus d'observations pour  $X$ ; ce qui permet un meilleur ajustement pour  $\hat{\beta}$ . Cela a des répercussions sur le calcul du maximum de vraisemblance des autres paramètres. Par la suite, on a des meilleures valeurs de départ pour la méthode composite puisque le tau de Kendall est plus crédible. Finalement, l'estimation de  $\alpha_1$  s'est amélioré étant donnée que les autres paramètres selon lesquels sa vraisemblance dépend sont mieux ajustés. Dans ce contexte, la méthode composite est l'approche la plus adéquate pour estimer les paramètres du modèle. Cependant, on voit que la méthode de calcul de la vraisemblance proposé en (3) et en (4) ne donne pas de bons résultats. Il faudra donc trouver une méthode alternative pour estimer adéquatement  $\alpha_1$ .

## 5 Impact du choix de la copule et des paramètres de dépendance

Dans le cas d'un modèle où la loi mère est géométrique, comme le niveau de dépendance n'est pas très élevé, même avec un paramètre  $\alpha_0$  avoisinant 1, on n'observe pas de biais dans l'estimation de  $\beta$ . Cependant, lorsque la loi mère de la copule archimédienne hiérarchique est logarithmique, la dépendance est plus forte. Il s'ensuit que les valeurs que peut prendre la v.a.  $X_{i,j}$  seront davantage influencées par les valeurs que prendra la v.a.  $N_i$ ,  $i=1, \dots, n_{obs}$ , comme l'illustre l'exemple 1.

**Exemple 1.** On a une copule archimédienne hiérarchique logarithmique-logarithmique avec  $\alpha_0 = 10$  et  $\alpha_1 = 10$ . Les variables aléatoires sous-jacentes sont  $N \sim \text{binomiale}(n = 5, q = 3/5)$  et  $X \sim \text{exponentielle}(1/100)$ . On simule 10 000 réalisations d'un tel modèle. Le tableau 5 présente les six premiers résultats de l'exemple 1.

i	N	X1	X2	X3	X4	X5
1	2	8.14	9.45			
2	5	314.10	627.06	205.52	487.95	257.58
3	4	192.08	189.38	251.15	173.13	
4	2	69.24	82.65			
5	3	100.02	85.79	204.63		
6	4	166.59	159.45	272.40	217.62	

Tableau 5 – Six premières réalisations de l'exemple 1.

Dans le tableau 5, on voit que, pour de grands  $N_i$ , les valeurs de  $X_{i,j}$  sont également élevées. Conséquemment, on se retrouve donc avec un nombre plus élevé de grands  $X$  comme l'illustre le tableau 6. De cette façon, si on calcule le paramètre de la loi exponentielle avec la méthode du maximum de vraisemblance, on a  $\hat{\beta} = 1/\bar{x} = 1/125.0237 \neq 1/100$ .

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
N	0.00	2.00	3.00	2.99	4.00	5.00	0.00
X1	0.01	29.25	69.02	100.62	138.48	1108.40	106.00
X2	0.00	37.07	78.76	108.29	147.52	1011.26	880.00
X3	0.12	63.14	105.86	134.07	177.06	1056.53	3196.00
X4	13.14	118.05	170.15	194.77	246.33	832.55	6686.00

Tableau 6 – Sommaire des données simulées pour l'exemple 1.

Une méthode pour remédier à la situation présentée dans l'exemple 1 consiste à prendre un  $X_{i,j}$  au hasard pour chaque observation ( $i$ ), puis de répéter cette expérience plusieurs fois (30 à 100 fois) afin de créer un vecteur de montants de sinistres où aucun quantile ne serait plus représenté qu'un autre. Finalement, on trouve la moyenne de ce nouveau vecteur pour obtenir un paramètre  $\hat{\beta}$  plus représentatif des données.

Voyons maintenant avec deux autres scénarios comment cela se traduit dans l'estimation de tous les paramètres d'un modèle collectif du risque.

**Scénario 3.** On prend tous les  $X_{i,j}$  pour calculer l'estimateur de  $\beta$ .

Posons un modèle collectif du risque où la copule archimédienne hiérarchique est logarithmique-logarithmique avec  $\alpha_0 = 8$  et  $\alpha_1 = 6$ . Pour le scénario 3, on prend l'approche classique où  $\hat{\beta} = 1/\bar{x}$  et que  $\bar{x} = \sum_{i=1}^{n_{obs}} \sum_{j=1}^{N_i} x_{i,j}$ . On prend alors en compte tous les  $X_{i,j}$ . Avec 1 000 observations simulées, on obtient les résultats présentés dans le tableau 7

	$q$	$\beta$	$\alpha_0$	$\alpha_1$
Vrais paramètres	0.6000	0.0100	8.0000	6.0000
Méthode <i>IFM</i>	0.6128	0.0078	7.3331	7.9825
Méthode composite	0.6132	0.0095	8.2209	7.2317
Méthode des moments	0.6128	0.0078	7.8535	4.2966

Tableau 7 – Résultats du scénario 3 : La méthode dite *IFM* consiste à trouver les paramètres des distributions marginales, puis à estimer  $\hat{\alpha}_0$  sachant ces paramètres. De la même façon, on trouve  $\hat{\alpha}_1$  en fixant  $\beta$  et  $\alpha_0$  comme connus. La méthode dite *composite*, quant à elle, consiste à trouver tous les paramètres simultanément en utilisant la méthode décrite dans la section 3. Finalement, la méthode des moments, dans les cas de  $\hat{\alpha}_0$  et  $\hat{\alpha}_1$ , consiste à inverser le tau de Kendall empirique.

Avec le tableau 7 on voit que la méthode *IFM* ne donne pas de bons résultats du fait que  $\hat{\beta}$  est biaisé et qu’il affecte les autres estimateurs. Par la suite, on voit que la méthode composite offre de bons résultats malgré un mauvais point de départ pour l’optimisation de  $\hat{\beta}$ . Finalement, on remarque que l’inversion du tau de Kendall pour  $\hat{\alpha}_1$  ne donne pas un bon résultat non plus. Ce phénomène se manifeste lorsque la loi mère insère une forte dépendance entre les éléments de la copule comme on peut l’observer dans l’illustration 8.

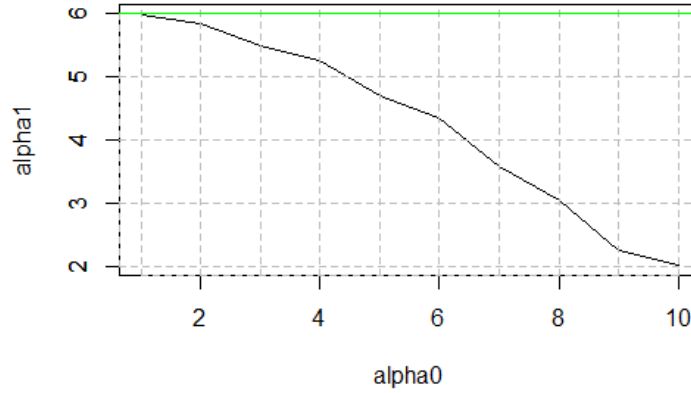


Illustration 8 – Influence de  $\alpha_0$  dans l’estimation de  $\alpha_1$  par inversion du tau de Kendall : Cette illustration est générée en simulant  $10 \times 10\,000$  observations du modèle collectif du risque présenté dans [Cossette et al., 2019b] pour  $\alpha_0 = 1, \dots, 10$ ,  $\alpha_1$  demeure constant à 6 et la copule est logarithmique-logarithmique. La ligne verte correspond au vrai paramètre  $\alpha_1$  définit pour les simulations.

Il en résulte que plus la dépendance insérée par la loi mère est forte, moins l’inversion du tau de Kendall constitue un estimateur adéquat de  $\alpha_1$ .

**Scénario 4.** On sélectionne aléatoirement un  $X_{i,j}$  par observation ( $i$ ) afin de calculer  $\hat{\beta}$ .

Reprenons le scénario 3, mais on définit

$$\hat{\beta} = \frac{(n_{obs})^{n_{sim}}}{\sum_{i=1}^{n_{obs}} \dots \sum_{i=1}^{n_{obs}} x_{i,j}}$$

où  $j$  est sélectionné aléatoirement entre  $[1, N_i]$  pour chacune des simulations,  $n_{sim}$  représente le nombre de simulations et  $n_{obs}$  correspond au nombre d'observations dans le jeu de données. Les résultats sont présentés dans le tableau 8. On voit alors que la méthode *IFM* arrive à des résultats très similaires à ceux de la méthode composite hormis que les temps de calcul sont bien moins longs (11 secondes pour la méthode *IFM* contre 6 minutes pour la composite avec 1 000 observations).

	$q$	$\beta$	$\alpha_0$	$\alpha_1$
Vrais paramètres	0.6000	0.0100	8.0000	6.0000
Méthode <i>IFM</i>	0.6128	0.0096	8.1509	7.1790
Méthode composite	0.6130	0.0095	8.1448	7.2790
Méthode des moments	0.6128	0.0078	7.8535	4.2966

Tableau 8 – Résultats du scénario 3 : La méthode dite *IFM* consiste à trouver les paramètres des distributions marginales, puis à estimer  $\hat{\alpha}_0$  sachant ces paramètres. De la même façon, on trouve  $\hat{\alpha}_1$  en fixant  $\beta$  et  $\alpha_0$  comme connus. La méthode dite *composite*, quant à elle, consiste à trouver tous les paramètres simultanément en utilisant la méthode décrite dans la section 3. Finalement, la méthode des moments, dans les cas de  $\hat{\alpha}_0$  et  $\hat{\alpha}_1$ , consiste à inverser le tau de Kendall empirique.

## 6 Conclusion

À la suite des scénarios présentés dans ce rapport, on a pu voir que la loi de fréquence avait un impact significatif sur l'estimation des paramètres du modèle collectif du risque. Également, plus la dépendance intégrée par la v.a. mère de la copule archimédienne hiérarchique est forte, plus cela vient biaiser l'estimation du paramètre de la v.a. enfant lorsque l'on utilise la méthode d'inversion du tau de Kendall. Également, on a vu que les fonctions de vraisemblance présentées en (3) et (4) ne sont pas adéquates pour estimer le paramètre de dépendance de la loi enfant.

Dans l'ensemble, peu importe la structure du jeu de données, la méthode composite offre de meilleurs performances que les deux autres approches. Et, surtout, les résultats de cette méthode sont moins influencée par la paramétrisation de  $N$  excepté que l'estimation du paramètre de cette dernière variable aléatoire peut être biaisé.

La prochaine étape de l'étude sera donc de trouver une approche alternative pour estimer le paramètre de dépendance de la v.a. enfant d'une copule archimédienne hiérarchique. Pour récapituler, le temps de dérivation pour obtenir la fonction de vraisemblance complète prend un temps phénoménal à calculer lorsque la v.a.  $N$  peut prendre des valeurs supérieures à 5. Par la suite la méthode d'inversion du tau de Kendall est fortement influencée par le niveau de dépendance induite via la v.a. mère ce qui nous amène à l'exclure aussi. Finalement, la



méthode décrite en (3) et en (4) ne constitue pas une bonne façon d'estimer le paramètre  $\alpha_1$  comme il a été illustré dans le présent rapport.

## Références

- [Cossette et al., 2019a] Cossette, H., Gadoury, S.-P., Marceau, E., and Robert, C. Y. (2019a). Composite likelihood 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.
- [Nikoloulopoulos and Karlis, 2009] Nikoloulopoulos, A. K. and Karlis, D. (2009). Modeling multivariate count data using copulas. *Communications in Statistics-Simulation and Computation*, 39(1) :172–187.