

Projet de Langage R en Actuariat

Marie GANON, Daniel NKAMENI, Florian SALAUN

1 Partie 1 - Agrégation simple des risques

Avant toute chose, nous stockons les valeurs des paramètres des lois normales et des copules.

Puis nous créons deux fonctions afin de calculer le *Best Estimate* (BE) et le *Solvency Capital Requirement* (SCR) d'une variable aléatoire X . Ces deux grandeurs sont définies de la manière suivante:

$$BE(X) = \mathbb{E}(X) \text{ et } SCR(X) = VaR_{99,5\%}(X) - BE(X)$$

où $VaR_{99,5\%}(X)$ est la *Value at Risk* au niveau 99,5% correspond au quantile de niveau 99,5%.

1.1 Modélisation avec copule gaussienne

1.1.1 Formule standard

On définit la copule gaussienne à partir des paramètres introduits plus haut, et on génère $n = 10^7$ simulations de (S_1, S_2) .

Afin d'obtenir un intervalle de confiance de $SCR(S_1 + S_2)$, il nous faut tout d'abord calculer les SCR des deux risques individuels. On peut le faire de manière computationnellement efficace en utilisant les fonctions `colMeans` (pour le calcul du *Best Estimate*) et `colQuantiles` (nécessitant l'importation de la librairie `MatrixStats`).

[5581957, 5585333]

La difficulté consiste dans l'estimation du coefficient de corrélation entre S_1 et S_2 . L'énoncé nous rappelle que :

$$\sqrt{n}(\hat{r}_n - r) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}(0, 1), \quad \hat{r}_n = \frac{1}{2} \ln \left(\frac{1 + \hat{\rho}_n}{1 - \hat{\rho}_n} \right) = \text{arctanh}(\hat{\rho}_n), \quad r = \text{arctanh}(\rho).$$

Comme la fonction $x \mapsto \tanh(x)$ est dérivable sur \mathbb{R} , d'après le théorème de la méthode Delta :

$$\begin{aligned} \sqrt{n}(\hat{\rho}_n - \rho) &= \sqrt{n}(\tanh(\hat{r}_n) - \tanh(r)) \\ &\xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}(0, (\tanh'(r))^2) \end{aligned}$$

avec $\tanh'(r) = 1 - \tanh(r)^2 = 1 - \rho^2$. Ainsi :

$$\sqrt{n} \frac{\hat{\rho}_n - \rho}{1 - \rho^2} \xrightarrow[n \rightarrow +\infty]{(d)} Y$$

avec $Y \sim \mathcal{N}(0, 1)$. Dans cette formule, deux termes dépendent du coefficient inconnu ρ , ce qui ne nous permet pas de déduire directement des intervalles de confiance. Or, on sait que $\hat{\rho}_n \xrightarrow[n \rightarrow +\infty]{p.s.} \rho$ car

$$\widehat{\text{Cov}}(X, Y) \xrightarrow[n \rightarrow +\infty]{p.s.} \text{Cov}(X, Y), \quad \widehat{\text{Var}}(X) \xrightarrow[n \rightarrow +\infty]{p.s.} \text{Var}(X), \quad \widehat{\text{Var}}(Y) \xrightarrow[n \rightarrow +\infty]{p.s.} \text{Var}(Y)$$

et car la fonction $(x, y, z) \mapsto \frac{x}{\sqrt{y}\sqrt{z}}$ est continue sur $\mathbb{R} \times (\mathbb{R}_+^*)^2$. Par le lemme de Slutsky, on en déduit que :

$$\left(\sqrt{n} \frac{\hat{\rho}_n - \rho}{1 - \rho^2}, \hat{\rho}_n \right) \xrightarrow[n \rightarrow +\infty]{(d)} (Y, \rho).$$

Or, la fonction $(u, v) \mapsto u \frac{1-\rho^2}{1-v^2}$ est continue sur $\mathbb{R} \times]-1, 1[$, d'où :

$$\sqrt{n} \frac{\hat{\rho}_n - \rho}{1 - \hat{\rho}_n^2} \xrightarrow[n \rightarrow +\infty]{(d)} Y.$$

Pour que ce résultat soit rigoureusement valide, il faudrait s'assurer que $\mathbb{P}(\hat{\rho}_n = 1) = 0$. Sinon, il faudrait rajouter un terme $\varepsilon > 0$ arbitrairement petit devant le terme $1 - \hat{\rho}_n^2$, ce qui aurait une influence arbitrairement petite sur l'intervalle de confiance obtenu. On décide donc d'omettre cette vérification.

Soit $y \in \mathbb{R}$. D'après la convergence en loi, pour n assez grand (10^7 convient largement), on a :

$$\mathbb{P} \left(\sqrt{n} \frac{|\hat{\rho}_n - \rho|}{1 - \hat{\rho}_n^2} \leq y \right) \approx \mathbb{P}(|Y| \leq y)$$

soit

$$\mathbb{P} \left(|\hat{\rho}_n - \rho| \leq y \frac{1 - \hat{\rho}_n^2}{\sqrt{n}} \right) \approx \mathbb{P}(|Y| \leq y).$$

Or, $\mathbb{P}(|Y| \leq y) = 2\mathbb{P}(Y \leq y) - 1$ par symétrie de Y . De plus, $2\mathbb{P}(Y \leq y) - 1 = 0.995$ revient à $\mathbb{P}(Y \leq y) = 0.9975$, et on peut en déduire $y = q_{0.9975}$ à partir de la fonction `qnorm` de R :

2.807034

Finalement, avec probabilité 99,5% :

$$\rho \in \left[\hat{\rho}_n - q_{0.9975} \frac{1 - \hat{\rho}_n^2}{\sqrt{n}}, \hat{\rho}_n + q_{0.9975} \frac{1 - \hat{\rho}_n^2}{\sqrt{n}} \right]$$

et on peut déduire l'application numérique :

[0.248749092957888, 0.250413830450627]

ainsi qu'un intervalle de confiance de $\text{SCR}(S_1 + S_2)$ à 99.5% via la formule

$$\text{SCR}(S_1 + S_2) = \sqrt{\text{SCR}(S_1)^2 + \text{SCR}(S_2)^2 + 2\rho(S_1, S_2)\text{SCR}(S_1)\text{SCR}(S_2)}.$$

SCR : 8827039.96358392

Intervalle de confiance : [8824100, 8829979]

1.1.2 Approche exacte

Dans cette partie, nous calculons le SCR de $S = S_1 + S_2$ en utilisant une approche directe. Il sera question de simuler S_1 et S_2 , de calculer S et ensuite d'appliquer la formule:

$$\text{SCR}(S) = \text{VaR}_{99,5\%}(S) - BE(S) \text{ où } BE(S) = \mathbb{E}(S)$$

Le *Best Estimate* (moyenne) de S , $BE(S)$ est égal à:

47047371

La $\text{VaR}_{99,5\%}(S)$ est obtenue en calculant le quantile d'ordre 99,5% de S . Elle est égale à:

55703292

Le $\text{SCR}(S)$ est donc égal à :

8655921

Calculons à présent l'intervalle de confiance à 99,5% de ce SCR. Nous savons que :

$$\sqrt{n}(\hat{q}_n^\alpha - q_\alpha) \xrightarrow[n \rightarrow +\infty]{(d)} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{f_S(q_\alpha)}\right)$$

où $f_S(q_\alpha)$ est la densité de S au point q_α . On peut réécrire cette convergence en loi sous la forme:

$$\sqrt{\frac{nf_S(q_\alpha)}{\alpha(1-\alpha)}}(\hat{q}_n^\alpha - q_\alpha) \xrightarrow[n \rightarrow +\infty]{(d)} Y \text{ où } Y \sim \mathcal{N}(0, 1)$$

Comme à la question précédente, deux termes dépendent du coefficient inconnu q_α . Pour déterminer l'intervalle de confiance, on fera l'hypothèse que $\hat{q}_n \xrightarrow[n \rightarrow +\infty]{p.s.} q_\alpha$. Puisque la fonction f_S est continue sur \mathbb{R} , on a $f_S(\hat{q}_n) \xrightarrow[n \rightarrow +\infty]{p.s.} f_S(q_\alpha)$ et par le Lemme de Slutsky, on en déduit que:

$$\left(\sqrt{\frac{nf_S(q_\alpha)}{\alpha(1-\alpha)}}(\hat{q}_n^\alpha - q_\alpha), f_S(\hat{q}_n^\alpha) \right) \xrightarrow[n \rightarrow +\infty]{(d)} (Y, f_S(q_\alpha))$$

Or, la fonction $(u, v) \mapsto u \frac{\sqrt{v}}{\sqrt{f_S(q_\alpha)}}$ est continue sur $\mathbb{R} \times \mathbb{R}$, d'où :

$$\sqrt{\frac{nf_S(\hat{q}_n^\alpha)}{\alpha(1-\alpha)}}(\hat{q}_n^\alpha - q_\alpha) \xrightarrow[n \rightarrow +\infty]{(d)} Y$$

Pour la suite, il nous faut $f_S(\hat{q}_n^\alpha)$. Nous allons estimer la densité de S , f_S par la méthode des noyaux. Il s'agit d'une méthode non paramétrique permettant d'estimer la densité de probabilité d'une variable aléatoire continue. Cette estimation est donnée par:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left[\frac{x - x_i}{h}\right]$$

Où h est le pas, n la taille de l'échantillon et K est le noyau choisi. Ce noyau est en général gaussien, uniforme ou triangulaire. La fonction `density` de R permet de faire cette estimation sans difficulté. L'estimation de la densité de S est représentée sur la Figure 1.

L'estimation de cette densité au point $\hat{q}_{0,995} = \widehat{VaR}_{99,5\%}$ est obtenue par interpolation grâce à la fonction `approx` de R et est égale à :

3.882129e-09

Soit $y \in \mathbb{R}$. Pour n assez grand, on a :

$$\mathbb{P}\left(\sqrt{\frac{nf_S(\hat{q}_n^\alpha)}{\alpha(1-\alpha)}}|\hat{q}_n^\alpha - q_\alpha| \leq y\right) \approx \mathbb{P}(|Y| \leq y)$$

soit

$$\mathbb{P}\left(|\hat{q}_n^\alpha - q_\alpha| \leq y \sqrt{\frac{\alpha(1-\alpha)}{nf_S(\hat{q}_n^\alpha)}}\right) \approx \mathbb{P}(|Y| \leq y).$$

Or, $\mathbb{P}(|Y| \leq y) = 2\mathbb{P}(Y \leq y) - 1$ par symétrie de Y . De plus, $2\mathbb{P}(Y \leq y) - 1 = 0,995$ revient à $\mathbb{P}(Y \leq y) = 0.9975$, et on peut en déduire $y = q_{0,9975}^{norm}$ à partir de la fonction `qnorm` de R :

2.807034

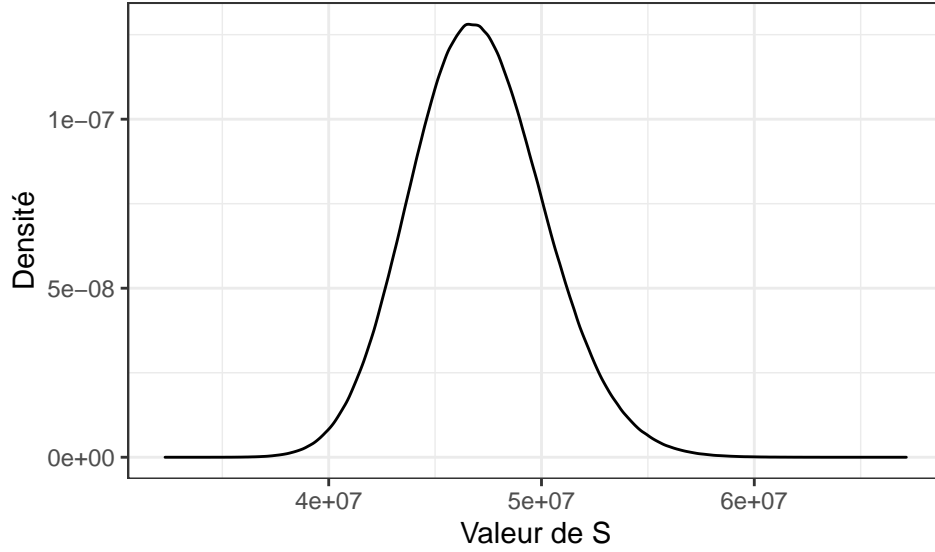


Figure 1: Densité de S - Copule gaussienne

Finalement, avec probabilité 99,5% :

$$q_{0,995} = VaR_{99,5\%} \in \left[\widehat{VaR}_{99,5\%} - q_{0,9975}^{norm} \sqrt{\frac{\alpha(1-\alpha)}{nf_S(\hat{q}_\alpha)}}, \widehat{VaR}_{99,5\%} + q_{0,9975}^{norm} \sqrt{\frac{\alpha(1-\alpha)}{nf_S(\hat{q}_\alpha)}} \right]$$

Où $q_{0,9975}^{norm}$ est le quantile d'ordre 0,9975 de la loi normale centrée réduite. On peut déduire l'intervalle de confiance à 99,5% de $VaR_{99,5\%}$:

[55703291, 55703293]

ainsi qu'un intervalle de confiance de $SCR(S_1 + S_2)$ à 99.5% via la formule

$$SCR(S) = VaR_{99,5\%}(S) - BE(S)$$

[8655920, 8655922]

1.2 Modélisation avec copule de Clayton

On définit la copule de Clayton à partir des paramètres introduits plus haut, et on génère $n = 10^7$ simulations de (S_1, S_2) .

1.2.1 Formule standard

On répète les étapes effectuées avec une modélisation par copule gaussienne. On calcule d'abord les SCR individuels :

5577021 5584137

On détermine ensuite un intervalle de confiance à 99.5% du coefficient de corrélation de la même manière que précédemment :

[0.252680079969758, 0.254341308062386]

Enfin, on peut en déduire un intervalle de confiance à 99.5% de $SCR(S_1 + S_2)$:

SCR : 8836053.54070242

Intervalle de confiance : [8833126, 8838981]

1.2.2 Approche exacte

La démarche à suivre dans cette partie est identique à celle utilisée avec la copule gaussienne. En simulant S_1 et S_2 grâce à la copule de Clayton, le *Best Estimate* de S , $BE(S)$ est égal à:

47049130

La $VaR_{99,5\%}(S)$ est égale à:

56666747

Le $SCR(S)$ est donc égal à :

9617616

L'estimation de la densité de S est représentée sur la Figure 2.

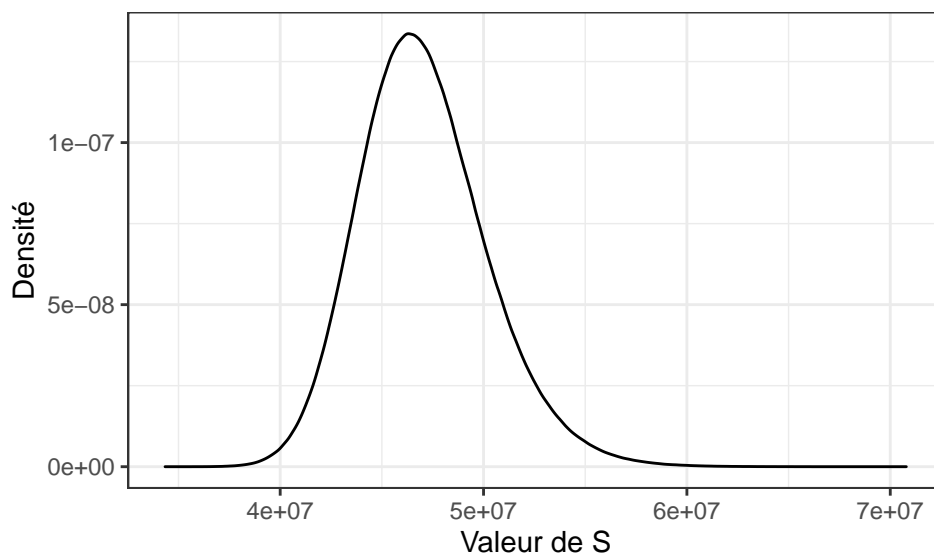


Figure 2: Densité de S - Copule de Clayton

L'estimation de cette densité au point $\hat{q}_{0,995} = \widehat{VaR}_{99,5\%}$ est égale à :

3.025291e-09

L'intervalle de confiance à 99,5% de $VaR_{99,5\%}$ est :

[56666746, 56666748]

Et l'intervalle de confiance à 99,5% du $SCR(S_1 + S_2)$ est :

[9617616, 9617617]

1.3 Comparaison des méthodes

Afin de comparer les deux méthodes, nous présentons dans la Table 1 l'ensemble de nos résultats de SCR.

	Formule.standard	Approche.exacte
Copule gaussienne	8827039.96	8655921.40
Copule de Clayton	8836053.54	9617616.36

Table 1: Récapitulatif des SCR calculés

Lorsque l'on applique la formule standard pour calculer le SCR, le choix de la copule n'a pas énormément d'importance. Toutefois, ce dernier devient déterminant dans le cadre de l'approche exacte. En effet, choisir une copule de Clayton pour modéliser la dépendance entre les deux risques conduit à surestimer fortement le SCR, par rapport au calcul effectué en formule standard. A l'inverse, la copule gaussienne diminue légèrement l'estimation du SCR par rapport au calcul effectué en formule standard.

Nous pouvons également comparer les intervalles de confiance pour chacune des approches et chaque structure de copule utilisée. Dans les tables suivantes, nous appellerons donc $\min(\text{SCR})$ et $\max(\text{SCR})$ les limites des intervalles de confiance à 99,5%.

	Formule.standard	Approche.exacte
$\min(\text{SCR})$	8824099.55	8655920.39
SCR	8827039.96	8655921.40
$\max(\text{SCR})$	8829979.40	8655922.40

Table 2: Récapitulatif des SCR calculés pour la copule gaussienne

	Formule.standard	Approche.exacte
$\min(\text{SCR})$	8833125.54	9617615.70
SCR	8836053.54	9617616.36
$\max(\text{SCR})$	8838980.57	9617617.03

Table 3: Récapitulatif des SCR calculés pour la copule de Clayton

Quelle que soit la copule utilisée, nous notons que les intervalles de confiance obtenus en modèle agrégé sont bien plus resserrés autour du SCR moyen que dans le cas de la formule standard. Cependant, si nous nous intéressons uniquement au calcul du SCR avec la formule standard, son écart-type est légèrement plus faible lorsque l'on choisit une copule gaussienne.

2 Partie 2 - Agrégation des risques par somme aléatoire

Nous commençons par définir tous les paramètres qui nous seront utiles par la suite.

Pour simuler S_1 et S_2 nous procédons en plusieurs étapes. Nous simulons dans un premier temps les deux nombres de sinistres, avec des lois marginales négatives binomiales et la structure de copule choisie (gaussienne ou Clayton). Puis, pour chaque simulation :

1. Nous générons les X_n^1 selon la loi $\mathcal{LN}(\mu_{log}, \sigma_{log})$.
2. Nous générons les U_n selon la loi $\mathcal{U}([0, 1])$.
3. Nous calculons $X_n^2 = k + \frac{s}{\xi}(U_n^{-\xi} - 1)$.
4. Enfin nous sommons respectivement les X_n^1 et les X_n^2 pour obtenir S_1 et S_2 .

Plusieurs approches ont été tentées afin d'optimiser le temps d'exécution ainsi que l'usage de mémoire vive. Une telle démarche s'avère nécessaire au vu du grand nombre de simulations à générer. Tout d'abord, nous avons testé une fonction R sans boucle utilisant l'instruction `sapply`. Ensuite, nous avons utilisé un code C (inspiré du cours) contenant une boucle afin d'accélérer l'exécution. Nous avons lancé un *microbenchmark* permettant de répéter 100 fois la génération de $n_0 = 10^2$ simulations de S_1 à partir de ces deux méthodes. Il en résulte la Figure 3.

Nous avons constaté, sur une machine équipée d'un processeur Intel Core i7-10510U (4 coeurs, 1.8 GhZ) et d'une mémoire vive de 8 Go, que le code C permettait de réduire le temps d'exécution d'environ un quart par rapport à un code R bien optimisé. En outre, cette solution est optimale en termes d'utilisation de mémoire vive. Par comparaison, la proposition qui consistait à construire sur R une matrice de taille $n \times \max(N)$ n'a pas pu aboutir sur nos machines pour $n = 10^5$, car elle conduisait à libérer plus de 10 Go de mémoire vive.

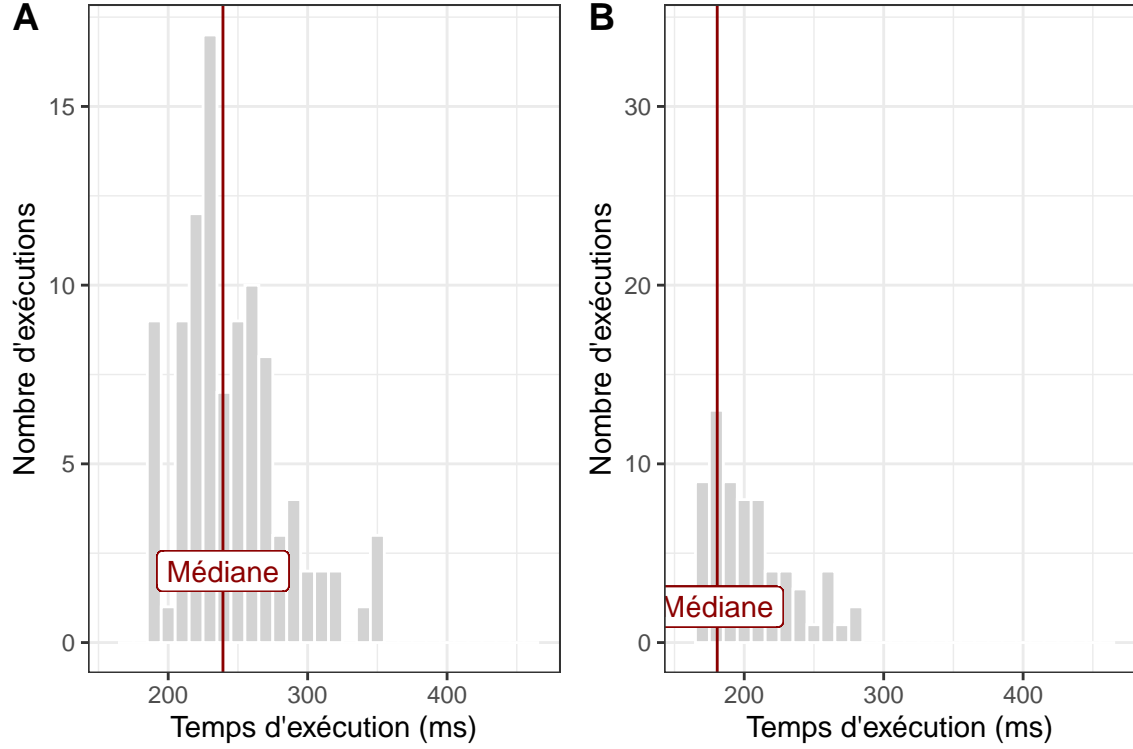


Figure 3: Résultats des microbenchmarks en (A) avec un apply sur R et en (B) avec du code C

2.1 Modélisation avec copule gaussienne

Dans cette partie, nous appliquons une structure de copule gaussienne de paramètre ρ_C .

2.1.1 Formule standard

Pour appliquer la formule standard, nous avons tout d'abord besoin de calculer le coefficient de corrélation linéaire ρ entre S_1 et S_2 . Celui-ci se calcule empiriquement de la manière suivante:

$$\hat{\rho}_n = \frac{\sum_{i=1}^n (S_1^i - \bar{S}_1)(S_2^i - \bar{S}_2)}{\sqrt{\sum_{i=1}^n (S_1^i - \bar{S}_1)^2} \sqrt{\sum_{i=1}^n (S_2^i - \bar{S}_2)^2}}$$

La formule standard donne le SCR suivant:

12232788435

2.1.2 Modèle agrégé

En calculant le SCR de manière agrégée, en posant $S = S_1 + S_2$, nous obtenons:

12231709866

Nous cherchons également à déterminer un intervalle de confiance pour le SCR. Pour ce faire, nous reprenons les calculs effectués dans la partie 1 pour l'approche exacte. Nous estimons donc la *Value at Risk*, ou quantile d'ordre $\alpha=99,5\%$ via le théorème central limite suivant:

$$\sqrt{\frac{nf_S(q_\alpha)}{\alpha(1-\alpha)}}(\hat{q}_n^\alpha - q_\alpha) \xrightarrow[n \rightarrow +\infty]{(d)} Y \text{ où } Y \sim \mathcal{N}(0,1)$$

Nous devons donc approximer la densité de S . Par la méthode des noyaux, nous obtenons la fonction tracée sur la Figure 4.

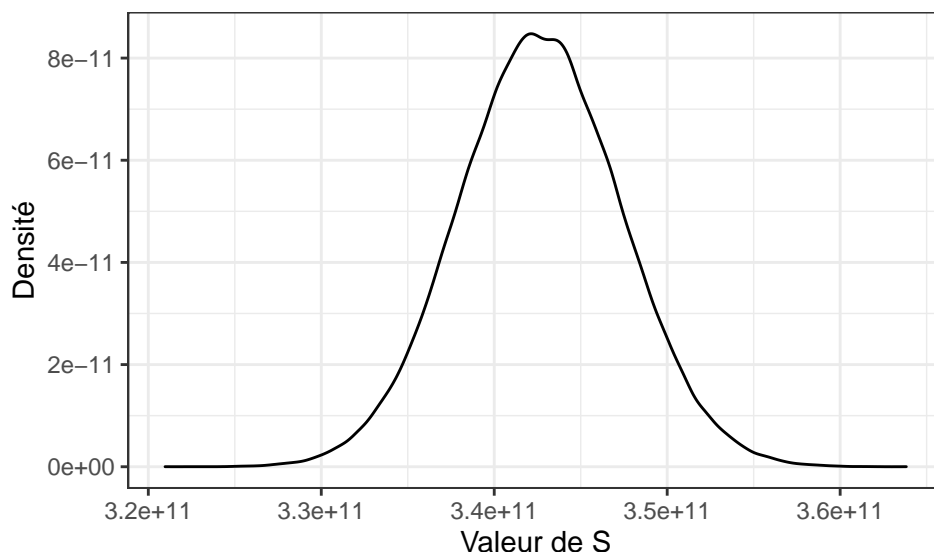


Figure 4: Densité de S - Copule gaussienne

On doit ensuite approximer cette fonction en la valeur du quantile d'ordre 99,5%.

Puis nous calculons l'intervalle de confiance sur le quantile d'ordre 99,5%.

[354877236382, 354877237105]

Une fois l'intervalle de confiance calculé pour le quantile d'ordre 99,5% de S , soit $Var_{99,5\%}$, nous pouvons alors retrancher le Best estimate pour obtenir un intervalle de confiance du SCR au niveau 99,5%.

[12231709504, 12231710228]

2.2 Modélisation avec copule de Clayton

Dans cette partie, nous appliquons une structure de copule de Clayton inversée de paramètre α_C .

2.2.1 Formule standard

La formule standard donne le SCR suivant:

8590335

2.2.2 Modèle agrégé

En calculant le SCR de manière agrégée, en posant $S = S_1 + S_2$, nous obtenons:

12026619384

Nous cherchons également à déterminer un intervalle de confiance pour le SCR. Comme précédemment, nous approximations la densité de S par la méthode des noyaux, puis nous la traçons sur la Figure 5.

Nous évaluons ensuite cette fonction en la valeur du quantile d'ordre 99,5%:

354660708672

Puis nous calculons l'intervalle de confiance sur le quantile d'ordre 99,5%.

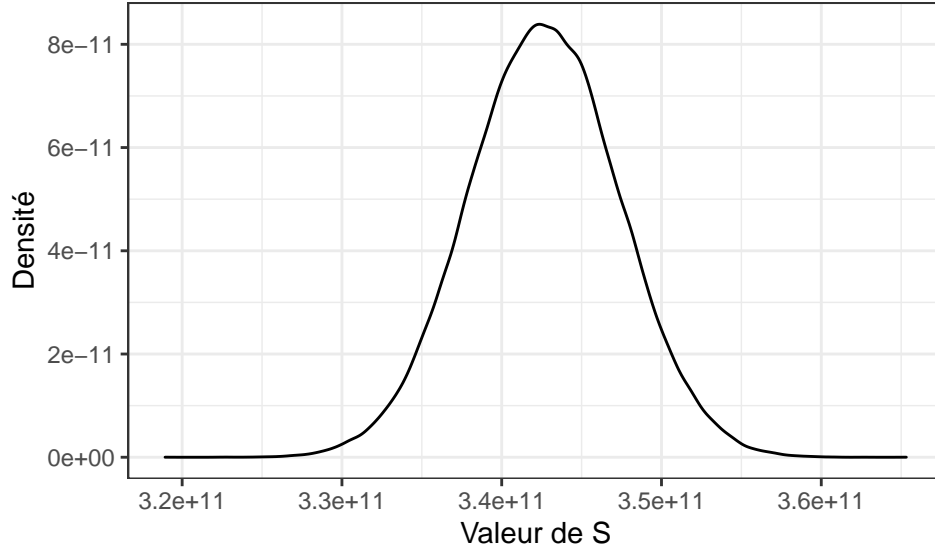


Figure 5: Densité de S - Copule de Clayton

[354660708329, 354660709015]

Une fois l'intervalle de confiance calculé pour le quantile d'ordre 99,5% de S , soit $Var_{99,5\%}$, nous pouvons alors retrancher le *Best Estimate* pour obtenir un intervalle de confiance du SCR au niveau 99,5%.

2.3 Comparaison des méthodes

Afin de comparer les deux méthodes, nous présentons dans la Table 4 l'ensemble de nos résultats de SCR.

	Formule.standard	Modèle.agrégé
Copule gaussienne	12232788435.21	12231709866.18
Copule de Clayton	8590335.01	12026619383.80

Table 4: Récapitulatif des SCR calculés

Lorsque la dépendance entre les nombres de sinistres est modélisée par une copule de Clayton, le SCR calculé avec la formule standard est très éloigné de celui calculé avec le modèle agrégé. Nous remarquons également que le SCR obtenu grâce à la formule standard dans le cas où la copule gaussienne est utilisée est très proche des SCR calculés avec le modèle agrégé peu importe la copule utilisée. Une explication possible à cet écart est que dans notre cas, le coefficient de corrélation linéaire de Pearson utilisé dans la formule standard capte mal le lien induit entre S_1 et S_2 par la copule de Clayton. La vérification de cette hypothèse nécessite une analyse approfondie de la structure de la copule de Clayton. Une telle analyse va au-delà du cadre de ce projet.

Comparons désormais les intervalles de confiance obtenus via le modèle agrégé. Nous les affichons sur la Table 5 ci-dessous.

	Copule.gaussienne	Copule.de.Clayton
min(SCR)	12231709504.36	12026619041.13
SCR	12231709866.18	12026619383.80
max(SCR)	12231710227.99	12026619726.47

Table 5: Intervalles de confiance obtenus pour le SCR

Les deux intervalles de confiance sont très resserrés autour des SCR estimés. En effet, la largeur de l'intervalle

de confiance est de 723.63 dans le cas où une copule gaussienne est utilisée et de 685.34 dans le cas où une copule de Clayton est utilisée. Ces largeurs sont très faibles par rapport aux SCR estimés. Ceci démontre la précision des estimations faites dans les deux cas. Par ailleurs, les valeurs du SCR calculés avec la formule standard et le modèle agrégé sont assez proches dans le cas où la copule gaussienne est utilisée (cf. Table 4). Ceci démontre la robustesse des estimations obtenues en modélisant la dépendance par une copule gaussienne. Pour avoir un mélange de précision et de robustesse des estimations du SCR pour des portefeuilles similaires au nôtre, nous recommandons de modéliser la dépendance entre les nombres de sinistres par une copule gaussienne.