

# 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`).

```
## [1] 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  $\rho$ , 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 :

```
## [1] 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 :

```
## [1] 0.2487491 0.2504138
```

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

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

```
## [1] "SCR" "8827039.96358392"
```

```
## [1] "Intervalle de confiance" "8824099.54901394"
```

```
## [3] "8829979.39898544"
```

### 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:

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

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

```
## [1] 47047371
```

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

```
## [1] 55703292
```

Le  $\text{SCR}(S)$  est donc égal à :

## [1] 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  $q_\alpha$ . 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  $q_\alpha$ . 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 dans le graphique ci-dessous:

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

## [1] 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 :

## [1] 2.807034

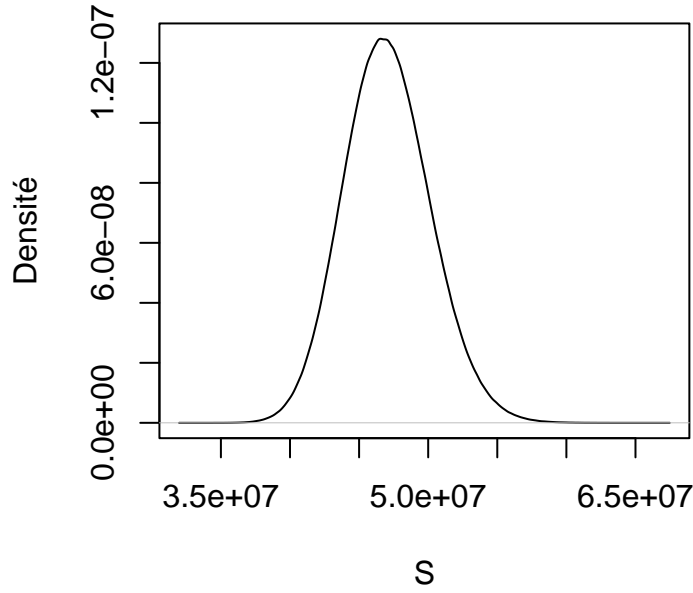


Figure 1: Densité de S - Copule gaussienne

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  $VaR_{99,5\%}$  :

## [1] 55703291 55703293

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

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

## [1] 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 :

## [1] 5577021 5584137

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

```
## [1] 0.2526801 0.2543413
```

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

```
## [1] "SCR" "8836053.54070242"
```

```
## [1] "Intervalle de confiance" "8833125.53949727"
```

```
## [3] "8838980.57197777"
```

### 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 :

```
## [1] 47049130
```

La  $\text{VaR}_{99,5\%}(S)$  est égale à :

```
## [1] 56666747
```

Le  $\text{SCR}(S)$  est donc égal à :

```
## [1] 9617616
```

L'estimation de la densité de  $S$  est représentée dans le graphique ci-dessous :

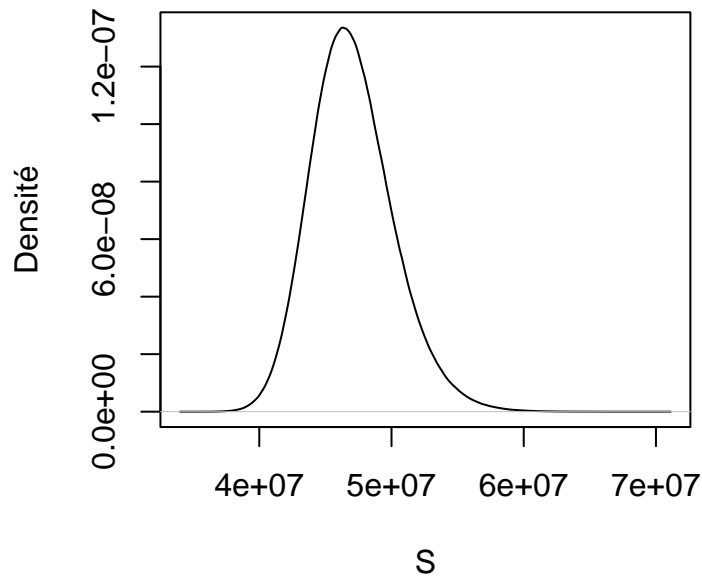


Figure 2: Densité de  $S$  - Copule de Clayton

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

```
## [1] 3.025291e-09
```

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

```
## [1] 56666746 56666748
```

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

```
## [1] 9617616 9617617
```

### 1.3 Comparaison des méthodes

Afin de comparer les deux méthodes, nous présentons dans la table suivante l'ensemble de nos résultats.

	Formule.standard	Modèle.exact
Copule gaussienne	8827039.96	8655921.40
Copule de Clayton	8836053.54	9617616.36

Table 1: Récapitulatif des SCR calculés

## 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  $X_n^2 = k + \frac{s}{\xi}(U^{-\xi} - 1)$ .
4. Enfin nous sommes 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 ressort que l'usage de C permet effectivement d'accélérer le code, via un gain moyen d'environ 20 millisecondes. Nos expériences ont montré que cet écart s'accroît lorsque le nombre de simulations  $n$  fixé est plus important, ce qui tend à renforcer l'intérêt de notre démarche d'optimisation pour  $n = 10^5$  simulations.

```
## Unit: milliseconds
##
## S1 <- sapply(nb_S1, function(x) sum(rlnorm(x, meanlog = mu_log,          sdlog = sigma_log)))      expr
##                                     S1 <- rsum(nb_S1, mu_log, sigma_log)
##      min      lq      mean  median      uq      max neval
## 140.7980 146.1596 151.7588 149.7533 154.4202 205.7563   100
## 131.3847 133.7331 136.7560 136.2595 139.2933 150.5007   100
```

### 2.1 Modélisation avec copule gaussienne

Dans cette partie, nous appliquons une structure de copule gaussienne de paramètre  $\rho_C$ .

#### 2.1.1 Formule standard

Pour appliquer la formule standard, nous avons tout d'abord besoin de calculer le coefficient de corrélation linéaire  $\rho$  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:

```
## [1] 12232788435
```

### 2.1.2 Modèle agrégé

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

```
## [1] 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.

*Ajout calculs*

Nous devons donc approximer la densité de  $S$ . Par la méthode des noyaux, nous obtenons la fonction suivante.

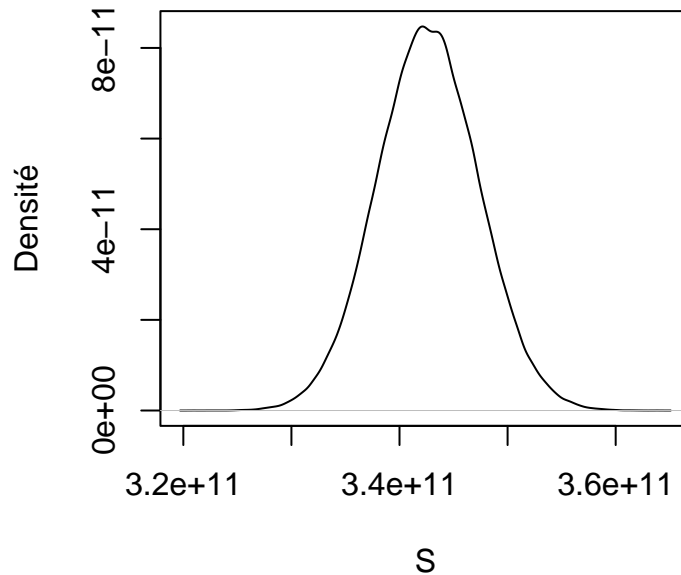


Figure 3: Densité de  $S$  - Copule gaussienne

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%.

```
##          99.5%          99.5%
## 354877236376 354877237111
```

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%.

```
##          99.5%          99.5%
## 12231709499 12231710234
```

## 2.2 Modélisation avec copule de Clayton

Dans cette partie, nous appliquons une structure de copule de Clayton inversée de paramètre  $\alpha_C$ .

### 2.2.1 Formule standard

La formule standard donne le SCR suivant:

```
## [1] 8590335
```

### 2.2.2 Modèle agrégé

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

```
## [1] 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.

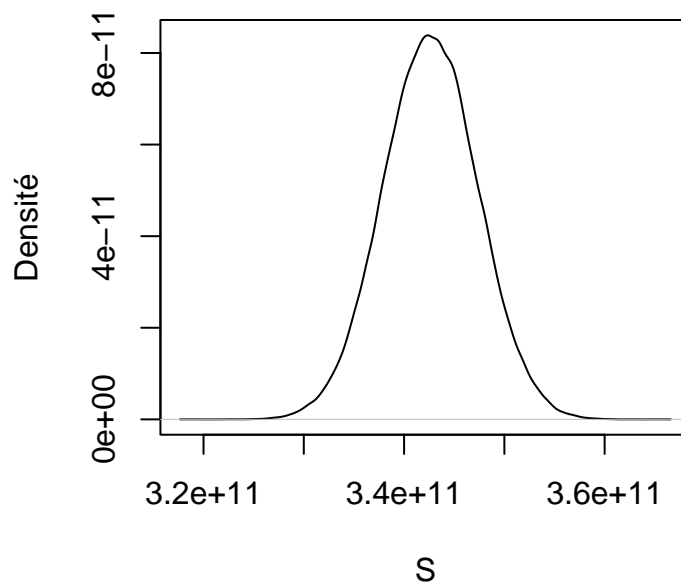


Figure 4: Densité de  $S$  - Copule de Clayton

Nous évaluons ensuite 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%.

```
##          99.5%          99.5%  
## 354660708321 354660709023
```

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 suivante l'ensemble de nos résultats.



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

Table 2: Récapitulatif des SCR calculés