Projet de Langage R en Actuariat

Marie GANON, Daniel NKAMENI, Florian SALAUN

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%.

Modélisation avec copule gaussienne

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 :

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

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

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

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

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

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 $\hat{\rho}_n \xrightarrow[n\to+\infty]{p.s.} \rho$ car

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

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 :

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

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

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

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 $\mathrm{SCR}(S_1+S_2)$ à 99.5% via la formule

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

```
## [1] 8824100 8829979
```

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 caluler S et ensuite d'appliquer la formule:

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

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

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

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

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

Le SCR(S) est donc égal à :

```
## [1] 8655921
```

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

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

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

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

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

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

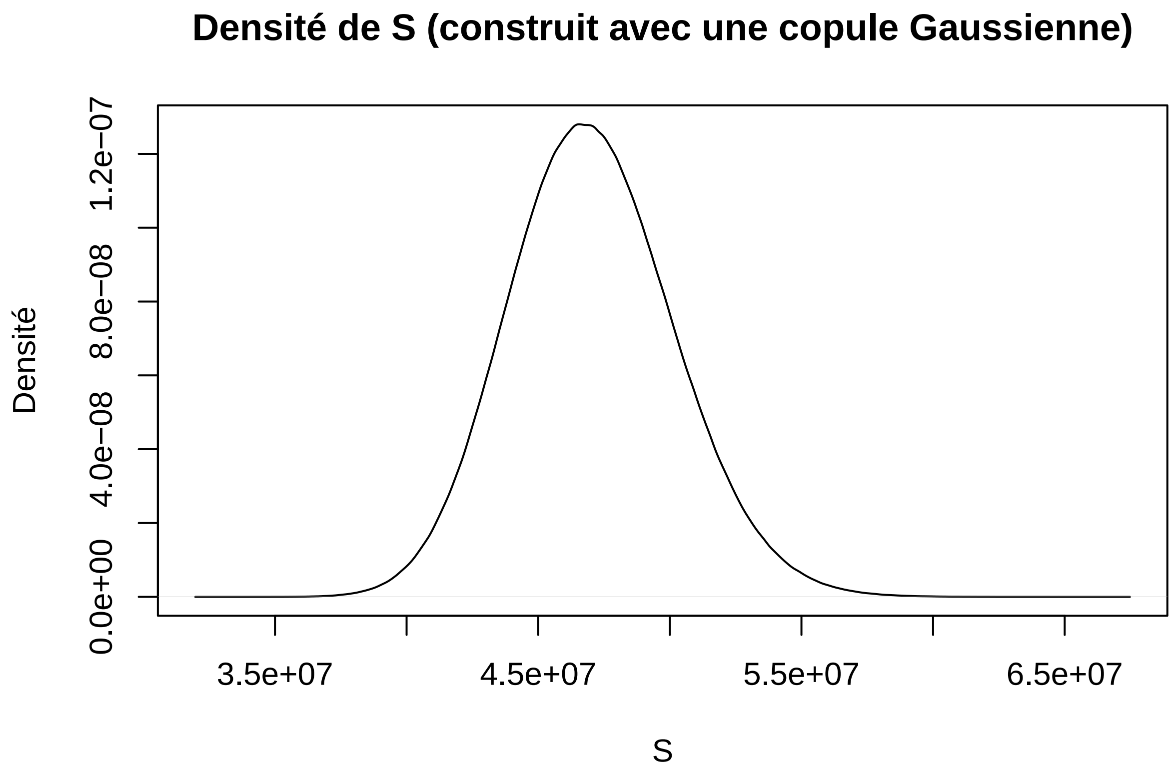
Or, la fonction $(u,v) \mapsto u\frac{\sqrt{v}}{\sqrt{f_S(q_\alpha)}}$ est continue sur $\mathbb{R} \times \mathbb{R}$, d'où :

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

Pour la suite, il nous faut $f_S(\hat{q}_\alpha)$. 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 $\hat{q}_{0,995} = V\hat{a}R_{99,5\%}$ 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 :

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

soit

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

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
```

Finalement, avec probabilité 99,5% :

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

Où $q_{0.9975}^{norm}$ est le quantile d'ordre $0,9975$ de la loi normale centrée reduite. On peut déduire l'intervalle de confiance à $99,5\%$ de $VaR_{99,5\%}$:

```
## [1] 55703292 55703293
```

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

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

```
## [1] 8655921 8655922
```

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).

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 $\mathrm{SCR}(S_1 + S_2)$:

```
## [1] 8833126 8838981
```

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
```

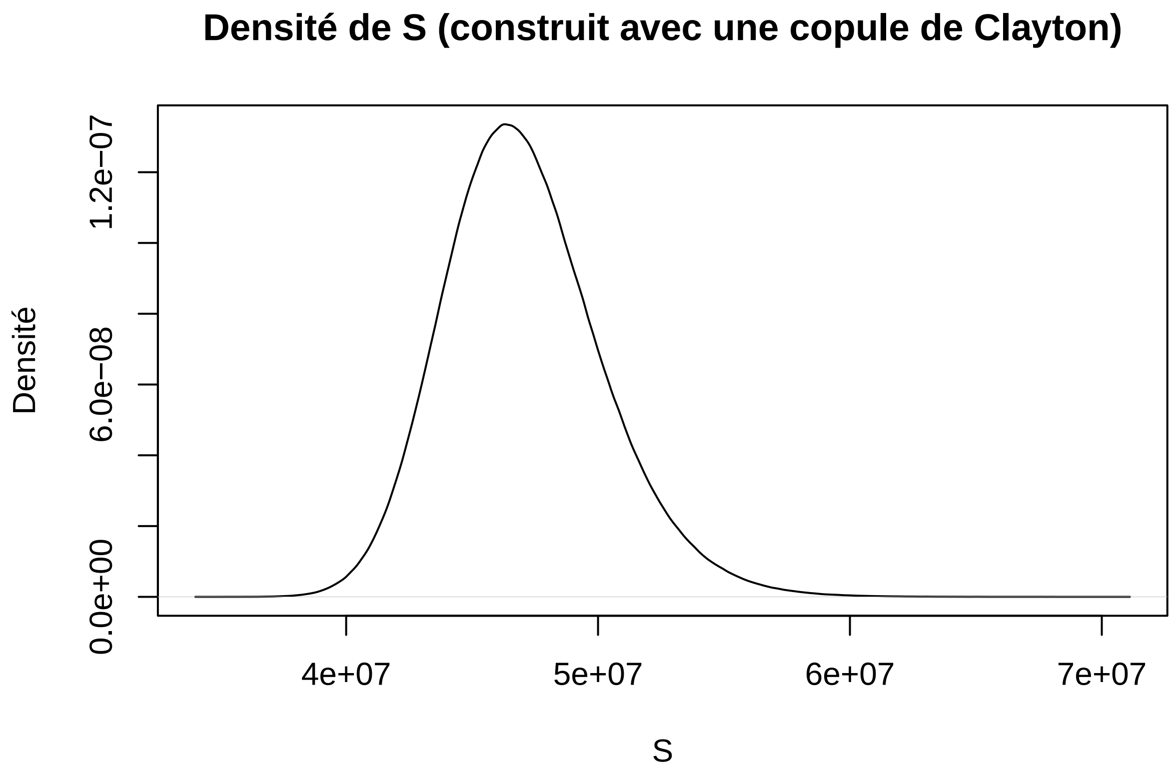
La $VaR_{99,5\%}(S)$ est égale à:

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

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

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

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



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

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

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

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

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

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

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 sommons respectivement les X_n^1 et les X_n^2 pour obtenir S_1 et S_2.

Modélisation avec copule gaussienne

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

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:

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

La formule standard donne le SCR suivant:

```
## [1] 12272368830
```

Modèle agrégé

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

```
## [1] 12271324346
```

Comparaison

Modélisation avec copule de Clayton

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

Formule standard

La formule standard donne le SCR suivant:

```
## [1] 12186778928
```

Modèle agrégé

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

```
## [1] 12186027696
```

Comparaison