

Théorie des valeurs extrêmes

M2 IMSA, M. BOUTAHAR

Chapitre 1. Approche des maximums par bloc

1.1. Introduction

La théorie des valeurs extrêmes est souvent liée à la gestion des risques qui peuvent être de différentes natures :

- Risque financier (chute brutale d'une valeur boursière, risque de faillite d'une compagnie d'assurance...),
- Risque géologique (tremblement de terre...),
- Risque environnemental (pollution de l'eau de la mer par du pétrole, pollution de l'air par des particules...),
- Risque climatiques (crues, avalanches...)

- En assurance :
 - Calcul du SCR dans le cadre de Solvabilité 2 (quantile d'ordre 99,5%)

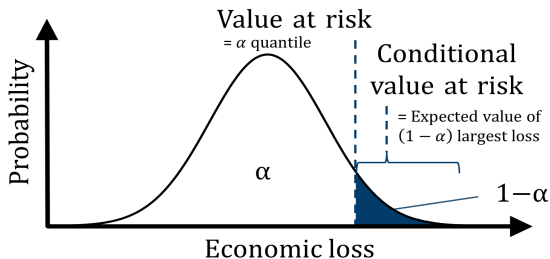


Figure 1 – Calcul du SCR

-Séparer les sinistres attritionnels et les sinistres graves.

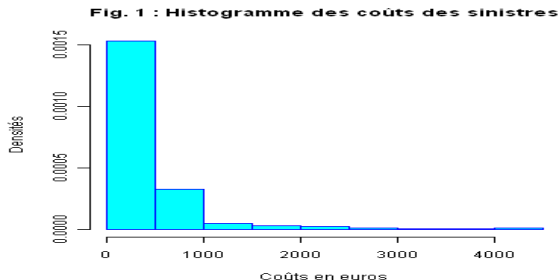


Figure 2 – Calcul du seuil des sinistres graves

Deux approches

- L'estimation de la loi des maximums par bloc.
- L'estimation de la loi des dépassements au delà d'un seuil.

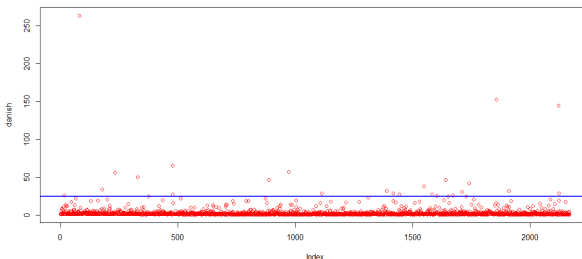


Figure 3 – Dépassements au delà d'un seuil

- L'ouvrage de référence de cette théorie est celui d'Embrechts, Kluppelberg et Mikosch (1997). Nous nous limiterons ici à donner une introduction de cette théorie, illustrée à l'aide des packages "evir", "evmix" et "extRemes".

1.2. Maximum de variables aléatoires

Soit $(X_1, X_2, \dots, X_n)$ une suite de n variables aléatoires (v.a.) indépendantes et identiquement distribuées (i.i.d.) de fonction de répartition F définie par $F(x) = P(X \leq x)$. On commence par modéliser la loi du maximum de ces variables aléatoires :

$$M_n = \max(X_1, X_2, \dots, X_n). \quad (1)$$

Loi limite du maximum

La fonction de répartition de M_n est donnée par

$$\begin{aligned} F_{M_n}(x) &= P(M_n \leq x) \\ &= P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) \\ &= P(X_1 \leq x) \times P(X_2 \leq x) \times \dots \times P(X_n \leq x) \\ &= (F(x))^n, \quad n \in \mathbb{N} \end{aligned} \tag{2}$$

La loi de M_n est très difficile à exploiter en pratique, car souvent F est inconnue, d'où l'intérêt d'étudier le comportement asymptotique du maximum convenablement normalisé. Fisher et Tippett ont donné en 1928 le théorème suivant qui est l'équivalent du théorème centrale limite pour les suites de v.a. i.i.d.

Theorem

(Fisher-Tippett (1928)).

S'il existe deux suites de constantes de normalisation avec $a_n > 0$, b_n et une loi H non-dégénérée telles que :

$$\frac{M_n - b_n}{a_n} \xrightarrow{\mathcal{L}} H, \quad (3)$$

où $\mathcal{L}$ désigne la convergence en loi lorsque $n \rightarrow \infty$; alors H est forcément la fonction de répartition de l'une des trois lois limites :

Theorem

- la distribution (standard) de Gumbel

$$H_1(x) = \exp(-\exp(-x)), \quad \forall x.$$

- la distribution (standard) de Fréchet ($\alpha > 0$)

$$H_{2,\alpha}(x) = \begin{cases} \exp(-x^{-\alpha}) & \text{si } x \geq 0, \\ 0 & \text{si } x < 0. \end{cases}$$

- la distribution (standard) de Weibull ($\alpha < 0$)

$$H_{3,\alpha}(x) = \begin{cases} \exp(-(-x)^\alpha) & \text{si } x \leq 0, \\ 1 & \text{si } x > 0. \end{cases}$$

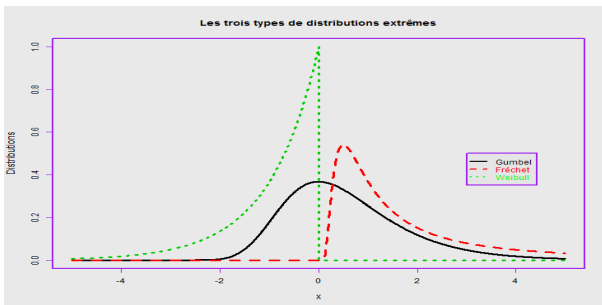


Figure 4 – Les densités de probabilités de Gumbel, Fréchet et Weibull

Même si, comme illustré dans la Figure 4, les trois lois sont différentes, on peut les combiner en une seule connue sous le nom de la distribution généralisée de valeurs extrêmes (GEV).

Loi GEV

Definition

Notons $\mu \in \mathbb{R}$ la localisation et $\sigma > 0$ le paramètre d'échelle. Considérons aussi $\xi \in \mathbb{R}$, appelé indice des valeurs extrêmes (ou indice de queue), qui décrit l'épaisseur de la queue de la loi.

La **distribution généralisée de valeurs extrêmes (GEV)** admet pour fonction de répartition

$$H_{\mu,\sigma,\xi}(x) = \exp \left(- \left(1 + \xi \frac{x - \mu}{\sigma} \right)^{-\frac{1}{\xi}} \right),$$

avec $1 + \xi \left(\frac{x - \mu}{\sigma} \right) \geq 0$, et $\xi \neq 0$.

Remarquons que

- $\xi = 0$ correspond à la loi de Gumbel de paramètres μ et σ . Dans ce cas, la queue de la loi décroît d'une manière exponentielle.
- $\xi > 0$ correspond à la loi de Fréchet de paramètres μ, σ et $\alpha = 1/\xi$.
- $\xi < 0$ correspond à la loi de Weibull de paramètres μ, σ et $\alpha = -1/\xi$.

Domaine d'attraction

Definition

Soit $(X_1, X_2, \dots, X_n)$ une suite de n v.a. i.i.d. de fonction de répartition F .

Soit H une distribution de valeur extrême.

On dit que F appartient au **domaine maximum d'attraction de H** , noté $F \in \text{DMA}(H)$, s'il existe deux suites de constantes de normalisation avec $a_n > 0$, b_n telles que :

$$\frac{M_n - b_n}{a_n} \xrightarrow{\mathcal{L}} H, \quad (4)$$

Proposition

- Si les X_i suivent une loi normale, exponentielle, gamma ou lognormale, alors la loi du maximum converge vers une loi de Gumbel. Ainsi, la loi des X_i est dans le domaine maximum d'attraction de Gumbel $DMA(H_1)$.
- Les lois de Cauchy et de Pareto sont dans le domaine maximum d'attraction de Fréchet $DMA(H_{2,\alpha})$.
- Les lois uniforme ou béta sont dans le domaine maximum d'attraction de Weibull $DMA(H_{3,\alpha})$.

Preuve pour la loi exponentielle

Supposons que X suit une loi exponentielle de paramètre $\lambda > 0$. Sa fonction de répartition est donnée par $F(x) = 1 - e^{-\lambda x}$. On a alors

$$\begin{aligned} P(\lambda M_n - \log(n) \leq x) &= P\left(M_n \leq \frac{x + \log(n)}{\lambda}\right) \\ &= \left(1 - e^{-(x + \log(n))}\right)^n \\ &= \left(1 - \frac{e^{-x}}{n}\right)^n \rightarrow \exp(-e^{-x}) = H_1(x). \end{aligned}$$

La loi limite est donc la loi de Gumbel.

Preuve pour la loi Pareto

Supposons que X suit une loi Pareto de paramètre $c > 0$ et $\alpha > 0$.

Sa fonction de répartition est donnée par $F(x) = 1 - cx^{-\alpha}$.

$$\begin{aligned} P\left(\frac{M_n}{(nc)^{1/\alpha}} \leq x\right) &= P(M_n \leq (nc)^{1/\alpha}x) \\ &= \left(1 - c((nc)^{1/\alpha}x)^{-\alpha}\right)^n \\ &= \left(1 - \frac{x^{-\alpha}}{n}\right)^n \\ &\rightarrow \exp(-x^{-\alpha}) = H_{2,\alpha}(x). \end{aligned}$$

La loi limite est donc la loi de Fréchet.

Preuve pour la loi uniforme

Supposons que X suit une loi uniforme sur $[0, 1]$. Sa fonction de répartition est donnée par $F(x) = x$. On a alors

$$\begin{aligned}P\left(\frac{M_n - 1}{1/n} \leq x\right) &= P(M_n \leq 1 + x/n) \\&= (1 + x/n)^n \\&\rightarrow e^x = H_{3,-1}(x).\end{aligned}$$

La loi limite est donc la loi de Weibull.

1.3. Nature de la loi extrême

A moins de connaître la loi sous-jacente des variables X_i et d'utiliser les résultats sur les domaines d'attraction, il est souvent difficile d'identifier la nature de la loi extrême.

Cependant on peut utiliser les méthodes graphiques. Il faut en effet remarquer que la distribution de Gumbel a comme support $\mathbb{R}$ tout entier, alors que celle de Fréchet a un support borné à gauche, et celle de Weibull a un support borné à droite. L'histogramme des valeurs observées ou un estimateur non paramétrique de la densité f fournissent une information préliminaire sur la nature de la loi (voir la Figure 5).

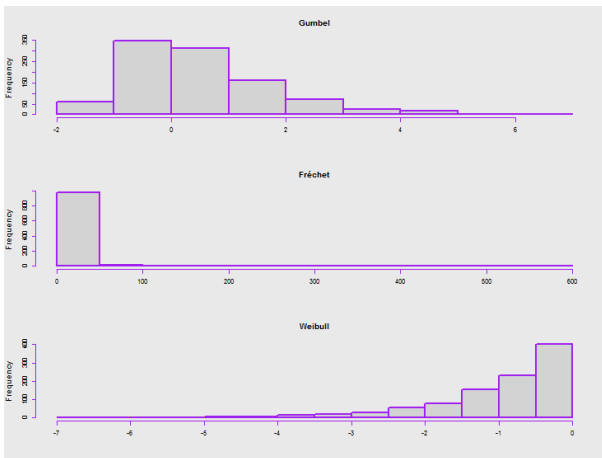


Figure 5 – Histogrammes des 3 lois extrêmes

Exemple. Sinistres incendie

On considère la série "danish" du package "evir" qui représente les indemnités de l'assurance incendie (en millions de Couronnes danoises) au Danemark entre le 3 janvier 1980 et le 31 décembre 1990. Etant donné que les montants renseignés sont les sinistres déclarés, on peut observer plusieurs montants associés à une même date et le temps séparant 2 déclarations est irrégulier. Une représentation classique sous la forme d'une série temporelle n'est donc pas appropriée.

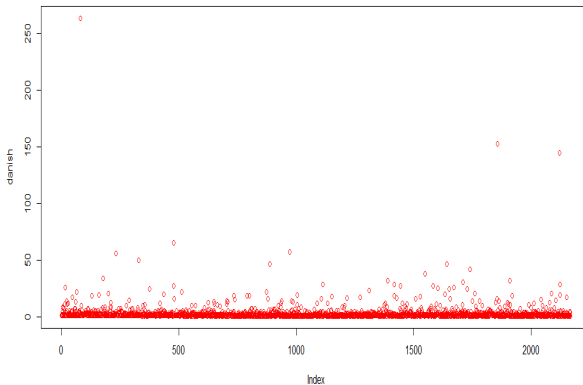


Figure 6 – Les primes payées par l'assurance incendie

On s'intéresse à la série des maximums sur 5 sinistres consécutifs. On propose de modéliser la loi de probabilité de la série par une GEV.

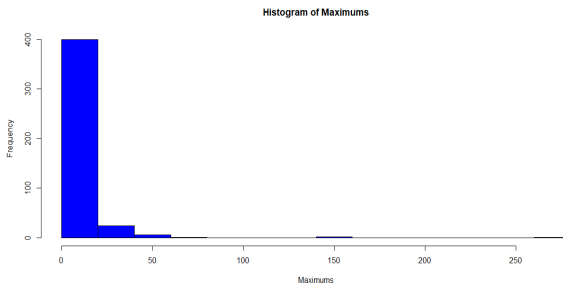


Figure 7 – Maximums des primes incendie

L'histogramme des valeurs observées, tracé dans la Figure 7 indique qu'une modélisation par une loi de Fréchet semble pertinente. On notera qu'en théorie, il faut que n tende vers l'infini pour que la série des maximums observés soit modélisable par une GEV. Ici, choisir $n = 5$ suffit à identifier un comportement de type Fréchet, et permet de disposer d'une série des maximums de taille $n = 433$, ce qui permettra de réaliser des estimations correctes des paramètres.

On obtient alors les estimateurs suivants :

$$\hat{\xi} = 0.713, \hat{\sigma} = 2.390, \hat{\mu} = 3.667.$$

avec les écart-types suivants :

$$\sigma_{\hat{\xi}} = 0.056, \sigma_{\hat{\sigma}} = 0.146, \sigma_{\hat{\mu}} = 0.132.$$

l'indice des valeurs extrême ξ est positif, ce qui confirme bien qu'il s'agit d'une loi de Fréchet.

Remarque. On peut aussi faire un test d'hypothèse

$$H_0 : \xi = 0 \text{ contre } H_1 : \xi > 0.$$

On trouve alors une p-value=0, on rejette H_0 au profit de H_1 .

1.4. Estimation de l'indice de queue ξ

Afin d'identifier la nature de la loi GEV, on peut aussi calculer une estimation du paramètre de queue ξ , à partir de la série des maximums, notée $Y_1, Y_2, \dots, Y_n$. Il existe plusieurs méthodes non paramétriques d'estimation (l'estimateur de Pickands (1975), l'estimateur de Hill (1975), et l'estimateur de Dekkers-Einmahl-De Haan (1989)).

1.4.1. Estimateur de Pickands

Definition

On considère la série des maximums $Y_1, \dots, Y_n$, de longueur n . Pour un entier k vérifiant $1 \leq j \leq n$, on note $Y_{j,n}$ la statistique d'ordre supérieur d'ordre j , c'est à dire que $Y_{n,n} \leq \dots \leq Y_{1,n}$.

Pour $1 \leq k \leq \frac{n}{4}$, on définit comme **estimateur de Pickands** la variable

$$\hat{\xi}_{k,n}^P = \frac{1}{\ln(2)} \ln \left(\frac{Y_{k,n} - Y_{2k,n}}{Y_{2k,n} - Y_{4k,n}} \right).$$

Proposition

Posons $k = k(n)$ tel que $\lim_{n \rightarrow +\infty} k(n) = +\infty$ et $\lim_{n \rightarrow +\infty} \frac{k(n)}{n} = 0$.

Alors l'estimateur $\hat{\xi}_{k(n),n}^P$ est fortement consistant et asymptotiquement normal (voir théorème 6.4.2 d'Embrechts, Kluppelberg, et Mikosch (1997)).

$$\sqrt{k(n)} \left(\hat{\xi}_{k(n),n}^P - \xi \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, v(\xi)), \text{ avec}$$

$$v(\xi) = \frac{\xi^2 (2^{2\xi+1} + 1)}{(2(2^\xi - 1) \ln(2))^2}.$$

Ce résultat permet à la fois de construire des intervalles de confiance, et de tester la nature de la loi extrême, par exemple le test $H_0 : \xi = 0$ permet de tester si la distribution limite H est une loi de Gumbel.

Exemple (Suite)

Reprenons l'étude de la série des maximums de "danish".

L'histogramme des maximums observés, tracé dans la Figure 7, suggérerait une loi de Fréchet.

Utilisons l'estimateur de Pickands pour le confirmer.

L'estimation de ξ dépend de la valeur k . Il est difficile de choisir k de façon arbitraire.

Considérer $k(n) = \sqrt{n}$ est une alternative convenable vu que dans ce cas, $\lim_{n \rightarrow +\infty} k(n) = +\infty$ et $\lim_{n \rightarrow +\infty} \frac{k(n)}{n} = 0$; ce qui est préconisé dans la Proposition 7.

Avec $k(n) = \sqrt{n}$ on trouve

$$\hat{\xi}_{k(n),n}^P = 0.109$$

L'autre alternative consiste à calculer l'estimateur de Pickands pour k variant de $n^{1/4}$ à $n^{3/4}$ par exemple, tout en s'assurant que $k \leq \frac{n}{4}$. On y adjoint les intervalles de confiance, calculés grâce à la Proposition 7. On obtient la figure 8.

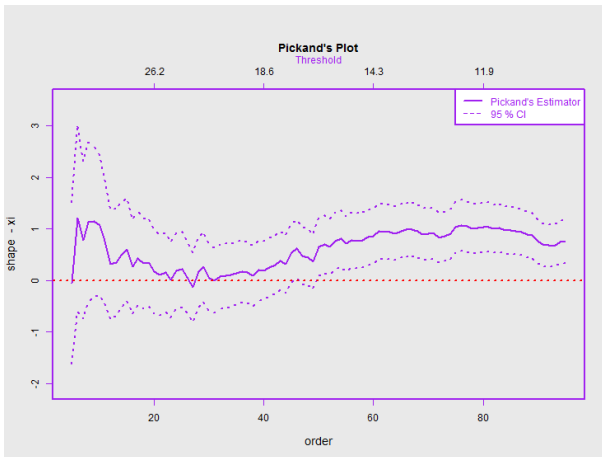


Figure 8 – Estimations de Pickands et les intervalles de confiance

*

Dans la Figure 8, on observe que l'estimation de ξ est

1.4.2. Estimateur de Hill

Definition

On suppose que $F \in DMA(H_{2,\alpha})$, c'est à dire que F appartient au domaine maximum d'attraction de la distribution de Fréchet.

Pour un entier k allant de 1 à la longueur n de la série, l'estimateur de Hill de ξ est donné par

$$\hat{\xi}_{k,n}^H = \frac{1}{k} \sum_{j=1}^k \ln(Y_{j,n}) - \ln(Y_{k,n}).$$

Proposition

Posons $k = k(n)$ tel que $\lim_{n \rightarrow +\infty} k(n) = +\infty$ et $\lim_{n \rightarrow +\infty} \frac{k(n)}{n} = 0$.

L'estimateur $\hat{\xi}_{k,n}^H$ est fortement consistant et asymptotiquement normal (voir théorème 6.4.6 d'Embrechts, Kluppelberg, et Mikosch (1997))

$$\sqrt{k(n)} (\hat{\xi}_{k(n),n}^H - \xi) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \xi^2).$$

Exemple (Suite)

Utilisons l'estimateur de Hill sur la série des maximums des sinistres d'incendie.

L'estimation de ξ dépend de la valeur k . Considérons $k(n) = \sqrt{n}$. On obtient alors

$$\hat{\xi}_{k,n}^H = 0.603.$$

On trace aussi l'estimation de ξ pour k variant de $n^{1/4}$ à $n^{3/4}$, avec $k \leq \frac{n}{4}$. On y adjoint les intervalles de confiance, calculés grâce à la Proposition 9. On obtient la Figure 9.

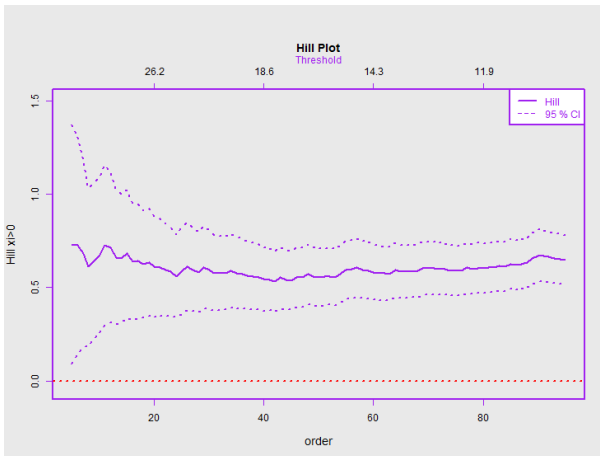


Figure 9 – Estimations de Hill et les intervalles de confiance

Dans la Figure 9, on observe que l'estimation de ξ est strictement positive pour tous les ordres k . Ceci correspond

1.5. Estimation des paramètres de la GEV

Il existe plusieurs méthodes pour estimer l'ensemble des paramètres des lois GEV. La plus populaire est celle du maximum de vraisemblance (J.A. Smith (1987)). Il existe aussi les méthodes des moments (N. Christopeit (1994), J.A. Greenwood, J.M. Landwehr, N.C. Matalas, et J.R. Wallis (1979)) ou encore les méthodes bayésiennes (L.M. Lye, K.P. Hapuarachchi, et S. Ryan (1993)).

Dans ce document on présente seulement les deux premières méthodes.

1.5.1. Estimateur du maximum de vraisemblance

Soit $\theta = (\xi, \sigma, \mu)$ les paramètres à estimer de la GEV.

L'estimateur du maximum de vraisemblance de θ est donné par

$$\hat{\theta} = \arg \max_{\theta} L(\theta, Y),$$

où

$$L(\theta, Y) = \prod_{i=1}^n h(\theta, Y_i) \mathbf{1}_{\left\{1 + \xi \frac{Y_i - \mu}{\sigma} > 0\right\}} \quad \text{est la vraisemblance et}$$

$$h(\theta, x) = \exp\left(-\left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-\frac{1}{\xi}}\right) \times \left(\left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-\frac{1}{\xi-1}}\right)$$

est la densité de la GEV.

Exemple(Suite)

On a vu que l'estimation de ξ par les estimateurs de Pickands et de Hill est strictement positive, ce qui correspond effectivement à une loi de Fréchet. Calculons maintenant l'estimation globale des paramètres par maximum de vraisemblance :

Avec le package "evir" on obtient les estimateurs suivants :

$$\hat{\xi} = 0.713, \hat{\sigma} = 2.390, \hat{\mu} = 3.667.$$

Avec le package "extRemes" on obtient les estimateurs suivants :

$$\hat{\xi} = 0.713, \hat{\sigma} = 2.389, \hat{\mu} = 3.666.$$

Les deux packages fournissent des estimations proches des paramètres μ , σ et ξ . Le paramètre de queue $\xi = 0.713 > 0$ indique bien qu'il s'agit d'une loi de Fréchet.

Chapitre 2. Approche du dépassement d'un seuil

On s'intéresse maintenant à la queue de la distribution afin d'identifier les quantiles extrêmes.

2.1. Loi de Pareto généralisée

Definition

Notons $\beta > 0$ et $\xi \in \mathbb{R}$ l'indice de queue.

La **distribution généralisée de Pareto (DGP)** admet pour fonction de répartition

$$G_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-\frac{1}{\xi}} & \text{si } \xi \neq 0 \\ 1 - e^{-\frac{x}{\beta}} & \text{si } \xi = 0 \end{cases} \quad (5)$$

$$\text{avec } \begin{cases} x \geq 0 & \text{si } \xi \geq 0 \\ 0 \leq \frac{x}{\beta} \leq -\frac{1}{\xi} & \text{si } \xi < 0. \end{cases}$$

Remarquons que

- $\xi = 0$ correspond à la loi exponentielle de paramètre $1/\beta$.
- $\xi > 0$ correspond à une loi à la queue épaisse. Plus exactement, elle se comporte comme la puissance $x^{-\frac{1}{\xi}}$ pour les grandes valeurs de x .
- $\xi < 0$ correspond à la loi de Pareto.

2.2. Loi des excès au-delà d'un seuil

La loi des excès au-delà d'un seuil, connue sous le nom "*Peak Over Threshold, POT*", décrit le comportement des observations qui dépassent un seuil donné u .

Proposition

Soit X v.a. de fonction de répartition F . On note $Y = X - u$ les excès, quand $X > u$.

On note F_u la fonction de répartition conditionnelle des excès. Autrement dit, $F_u(y) = P(Y \leq y | X > u)$.

Alors,

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)}.$$

Exemples.

1. Loi exponentielle

Supposons que X suit une loi exponentielle de paramètre $\lambda > 0$. Sa fonction de répartition est donnée par

$F(x) = 1 - e^{-\lambda x}$. On a alors

$$F_u(y) = \frac{e^{-\lambda u} - e^{-\lambda(u+y)}}{e^{-\lambda u}} = 1 - e^{-\lambda y}$$

Donc $Y \mid (X > u)$ suit une loi exponentielle de paramètre $\lambda > 0$.

2. Loi Pareto

Supposons que X suit une loi Pareto de paramètre $c > 0$ et $\alpha > 0$.

Sa fonction de répartition est donnée par $F(x) = 1 - cx^{-\alpha}$.

On a alors

$$F_u(y) = \frac{u^{-\alpha} - (y + u)^{-\alpha}}{u^{-\alpha}} = 1 - \left(1 + \frac{y}{u}\right)^{-\alpha}$$

Donc $Y \mid (X > u)$ suit une loi de Pareto généralisée $G_{\xi, \beta(u)}(y)$ avec

$$\xi = \frac{1}{\alpha}, \beta(u) = \xi u.$$

3. Loi uniforme

Supposons que X suit une loi uniforme sur $[0, 1]$. Sa fonction de répartition est donnée par $F(x) = x$. On a alors

$$F_u(y) = \frac{(u + y) - u}{1 - u} = \frac{y}{1 - u}$$

Donc $Y \mid (X > u)$ suit une loi uniforme sur $[0, 1 - u]$.

Soit $x_F = \sup \{x \in \mathbb{R} : F(x) < 1\}$, le point terminal de la distribution F . Le théorème de Pickands-Balkema-de Haan suivant montre que la fonction de répartition conditionnelle des excès peut être approchée, lorsque le seuil u tend vers le point terminal x_F , par une loi de Pareto généralisée.

Theorem

(de Pickands-Balkema-de Haan, voir Embrechts, Kluppelberg, et Mikosch (1997), page 165)).

Une fonction de répartition F appartient au domaine d'attraction maximale de G si et seulement si, il existe une fonction positive $\beta(u)$ telle que

$$\lim_{u \rightarrow x_F} \sup_{0 < y < x_F - u} |F_u(y) - G_{\xi, \beta(u)}(y)| = 0. \quad (6)$$

2.3. Choix du seuil u

Le choix du seuil u est un problème difficile. On doit en effet choisir u tel que :

- L'approximation de $F_u(y)$ par $G_{\xi, \beta(u)}(y)$ soit valide. Ainsi, u doit être proche x_F , donc assez grand.
- L'estimation de la loi $G_{\xi, \beta(u)}(y)$, basée sur les dépassements, soit possible. Ainsi, u ne doit pas être trop grand afin que le nombre des dépassements ne soit pas réduit à zéro.

Plusieurs méthodes ont été développées afin de déterminer un u plus au moins optimal, nous pouvons citer :

- Méthode graphique (Embrechts, Kluppelberg, et Mikosch (1997)), page 355) :

On détermine u en exploitant graphiquement la linéarité de l'espérance de vie résiduelle

$$e(u) = E(X - u \mid X > u).$$

- Méthode basée sur le MSE (J. Beirlant, P. Vynckier, and J.L. Teugels (1996)) : On choisit u qui minimise l'erreur quadratique moyenne asymptotique de l'estimateur de Hill.

- Méthode basée sur le bootstrap (J. Danielsson, L. De Haan, L. Peng, and C.G. De Vries (2001)) :
Utilisation du bootstrap pour trouver le seuil optimal u .

Exemple (Suite)

Nous illustrons ici la méthode de l'espérance de vie résiduelle (MRL), proposée par les packages "fExtremes" et "evmix".

Pour une v.a. $Y = X - u$ qui suit une $G_{\xi,\beta}(y)$ de paramètre $\xi < 1$, on montre que $e(u)$ est linéaire en u :

$$e(u) = E(X - u \mid X > u) = \frac{\beta + \xi u}{1 - \xi}, (\xi < 1).$$

Avec le package "evmix" on obtient la Figure 10.

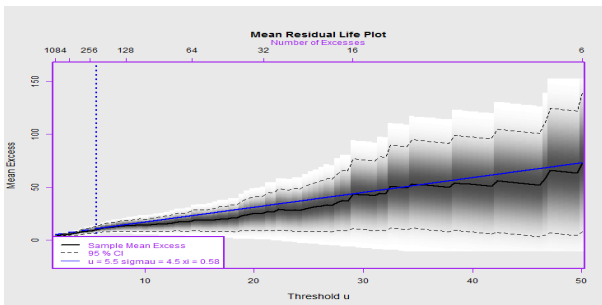


Figure 10 – L'espérance de vie résiduelle. (package : "evmix")

La Figure 10 révèle que le seuil u peut être choisi égal à 5.5. Elle donne aussi une estimation de ξ et $\beta(u)$, (noté sigmau).

2.4. Estimation de la DGP

On ne détaille ici que l'estimation des paramètres $\theta = (\xi, \beta)$ de la DGP par maximum de vraisemblance. L'estimateur du maximum de vraisemblance de θ est donné par

$$\hat{\theta} = \arg \max_{\theta} L(\theta, X),$$

où

$$L(\theta, X) = \prod_{i=1}^{N_u} h(\theta, Y_i) \mathbf{1}_{0 \leq Y_i \leq -\frac{\beta}{\xi}} \quad \text{est la vraisemblance, avec}$$

$Y_i = X_i - u, \quad 1 \leq i \leq N_u$ les dépassements du seuil u , et

$$h(\theta, y) = \frac{1}{\beta} \left((1 + \xi y / \beta)^{-1/\xi - 1} \right) \quad \text{la densité de la GP.}$$

$$N_u = \text{card}\{i, X_i \mid X_i > u\}.$$

Exemple (Suite)

Reprenons l'étude de la série "danish". Si on souhaite modéliser la loi des sinistres dont le coût dépasse 25 millions de Couronne danoise, il suffit d'estimer une DGP avec un seuil $u = 25$.

Remarquons que sur 2167 sinistres, il y a 24 sinistres qui dépassent la somme de 25 millions, les dépassements sont illustrés dans la figure 11.

Avec le package "evmix on obtient

$$\hat{\xi} = 0.823, \quad \hat{\beta}(u) = 10.402.$$

Si on choisit le $u = 5.5$ (choix basé la méthode de l'espérance de vie résiduelle) on obtient

$$\hat{\xi} = 0.603, \quad \hat{\beta}(u) = 4.336.$$

Il y a 221 sinistres qui dépassent la somme de 5.5 millions.

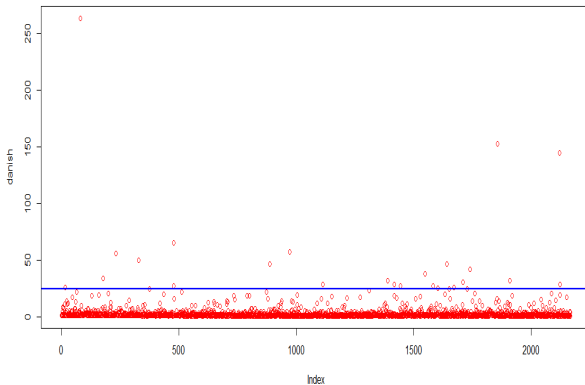


Figure 11 – sinistres dépassant $u = 25$ millions de Couronne danoise

2.5. Estimation des quantiles extrêmes

Soit X une variable aléatoire, pour $p > 0$ on cherche à estimer le quantile x_p , c'est à dire

$$P(X \leq x_p) = p.$$

Rappelons que la loi des excès au-delà de u suffisamment grand est :

$$\begin{aligned} F_u(y) &= \Pr(X - u \leq y \mid X > u) \\ &= \frac{F(u + y) - F(u)}{1 - F(u)} \\ &= \frac{F(u + y) - F(u)}{\bar{F}(u)} \end{aligned}$$

où $\bar{F} = 1 - F$ est la fonction de survie.

Ceci peut être réécrit comme

$$\begin{aligned}F_u(y)\overline{F}(u) &= F(u+y) - F(u) \\&= -1 + F(u+y) + 1 - F(u) \\&= \overline{F}(u) - \overline{F}(u+y)\end{aligned}$$

Donc

$$\begin{aligned}\overline{F}(u+y) &= \overline{F}(u) - F_u(y)\overline{F}(u) \\ &= \overline{F}(u)\overline{F}_u(y)\end{aligned}\tag{7}$$

Rappelons que l'équation (6) du théorème de Balkema et Haan (1974) et Pickands (1975) donne un résultat limite pour $F_u(y)$. Ceci nous amène à proposer l'approximation suivante, pour u proche de x_F :

$$F_u(y) \approx G_{\xi, \beta(u)}(y), \text{ lorsque } u \rightarrow x_F \text{ avec } y > 0 \quad (8)$$

ou encore

$$\bar{F}_u(y) \approx \bar{G}_{\xi, \beta(u)}(y),$$

où $G_{\xi, \beta(u)}(y)$ est la DGP donnée par l'équation (5).

La relation (7) montre que l'estimation de la queue de F pourrait être obtenue en estimant $\bar{F}(u)$ et $\bar{F}_y(u)$ séparément.

Un estimateur naturel pour $\bar{F}(u)$ est la survie empirique

$$\begin{aligned}\hat{\bar{F}}(u) &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i > u\}} \\ &= \frac{N_u}{n}\end{aligned}$$

où $\mathbf{1}_{\{X_i > u\}}$ est la fonction indicatrice qui vaut 1 si $X_i > u$, 0 sinon.

D'autre part, l'approximation (8) permet de proposer un estimateur pour $\overline{F}_u(y)$, soit

$$\hat{\overline{F}}_u(y) = \left(1 + \hat{\xi} \frac{y}{\hat{\beta}(u)} \right)^{-\frac{1}{\hat{\xi}}}$$

où $\hat{\xi}$ et $\hat{\beta}(u)$ sont les estimateurs des paramètres ξ et $\beta(u)$ de la DGP.

L'estimateur résultant pour la queue $\bar{F}(u+y)$ (pour $y > 0$)
a donc la forme

$$\begin{aligned}\hat{\bar{F}}(u+y) &= \hat{\bar{F}}(u)\hat{\bar{F}}_y(y) \\ &= \frac{N_u}{n} \left(1 + \hat{\xi} \frac{y}{\hat{\beta}(u)} \right)^{-\frac{1}{\hat{\xi}}}\end{aligned}\quad (9)$$

Posons

$$x_p = u + y_p.$$

En égalisant $\hat{F}(u + y_p) = \bar{F}(u + y_p) = 1 - p$ dans l'équation (9) on obtient l'estimateur du quantile extrême

$$\hat{x}_p = u + \frac{\hat{\beta}(u)}{\hat{\xi}} \left\{ \left(\frac{n}{N_u} (1 - p) \right)^{-\hat{\xi}} - 1 \right\}, \text{ pour } p > F(u). \quad (10)$$

Calcul du quantile à 99.5%.

En utilisant le package "evmix" on obtient :

-Pour $u = 25$, $x_p = 36.681$

-Pour $u = 5.5$, $x_p = 42.585$

Le quantile empirique est égal à $x_p = 34.141$.

Le quantile correspondant à $u = 5.5$ donne une valeur extrême par rapport aux deux autres.

Remarque.

- Choix du seuil u : Ecrêtement des sinistres graves dans le cadre de la tarification (et du provisionnement) en assurance non vie.
- Estimation des quantiles extrêmes : Calcul du SCR en solvabilité voir : F. Planchet (2017) Solvabilité et valeurs extrêmes. Enjeux en termes de modélisation.