



Théorie des Valeurs Extrêmes

ISFA -Université Lyon 1

ISFA M2 Actuariat 2021 - 2022

▷ OBJECTIFS PEDAGOGIQUES

- Introduire les concepts fondamentaux de la Théorie des Valeurs Extrêmes :
 - lois du maximum et des plus grandes valeurs d'un échantillon de variables aléatoires,
 - lois jointes des plus grandes valeurs,
 - lois des dépassements de seuils,
 - lois des arrivées des dépassements de seuils.
- Présenter les outils statistiques pour mettre en application cette théorie à partir de données issues de l'assurance et de la finance.



▷ BIBLIOGRAPHIE

- Beirlant J., Goegebeur Y., Segers J. and Teugels, J. (2004) Statistics of Extremes: Theory and Applications, Wiley Series in Probability and Statistics, John Wiley & Sons Ltd., Chichester.
- Coles S. (2001) An Introduction to Statistical Modeling of Extreme Values. Springer, London.
- Embrechts P., Klüppelberg C. and Mikosch T. (1997) Modelling extremal events for insurance and finance. Berlin: Spring Verlag.
- Robert C. (2018) Support de cours

- ▷ “There is always going to be an element of doubt, as one is extrapolating into areas one doesn’t know about. But what EVT is doing is making the best use of whatever data you have about extreme phenomena” - Richard Smith.
- ▷ “The key message is that EVT cannot do magic - but it can do a whole lot better than empirical curve-fitting and guesswork. My answer to the sceptics is that if people aren’t given well-founded methods like EVT, they’ll just use dubious ones instead” - Jonathan Tawn.
- ▷ “Many real life questions require estimation, but since no data or only few has been observed - as by definition extreme events are rare - essential estimations are more often based on feeling than on fact. EVT is a branch of statistics that deals with such rare situations and that gives a scientific alternative to pure guesswork” - Chavez-Demoulin, Armin Roehrl.

Plan:

1. Analyses des extrêmes univariés dans un cadre statique
 - 1.1 Lois limites pour le maximum
 - 1.2 Les dépassements de seuils
 - 1.3 Estimations des paramètres des lois des extrêmes
2. Analyses des extrêmes dans des cadres dynamique et multivarié
 - 2.1 Lois limites pour le maximum et les dépassements de seuils dans un cadre dynamique
 - 2.2 Lois limites pour les maxima des composantes d'un vecteur dans un cadre statique
 - 2.3 Caractérisation de la dépendance extrêmale

Chapitre 1. Analyses des extrêmes univariés dans un cadre statique

1. Lois limites pour le maximum

Soit X_1, \dots, X_n une suite de variables aléatoires IID de fonction de distribution F et soit $M_n = \max(X_1, \dots, X_n)$. Alors

$$\begin{aligned}\Pr(M_n \leq x) &= \Pr(X_1 \leq x, \dots, X_n \leq x) \\ &= \Pr(X_1 \leq x) \dots \Pr(X_n \leq x) \\ &= [F(x)]^n.\end{aligned}$$

Si F n'est pas connue, cette formule est peu utile. De plus, nous nous intéressons souvent à M_n , lorsque la taille de l'échantillon est importante, et nous souhaitons avoir des approximations asymptotiques.

En particulier, il est important de disposer d'une expression simple pour la loi asymptotique et il serait souhaitable que cette expression ne dépende pas trop de F .

Tout d'abord, notons que

- si $F(x) < 1$, alors $P(M_n \leq x) \rightarrow 0$ quand $n \rightarrow \infty$;
- si x^F est le point extrémal de F , i.e.

$$x^F = \sup \{x : F(x) < 1\},$$

alors $M_n \rightarrow x^F$ en probabilité quand $n \rightarrow \infty$ (et même presque sûrement comme suite croissante). La distribution asymptotique de M_n est donc dégénérée.

Ceci suggère qu'il faut passer par une transformation ou une normalisation. Regardons d'abord ce que l'on peut apprendre de la théorie des sommes de variables aléatoires.

▷ Théorie des sommes de variables aléatoires

Comme précédemment, X_1, \dots, X_n sont des variables aléatoires IID, mais maintenant nous supposons de plus que $\mathbb{E}(X_i) = m$, $Var(X_i) = \sigma^2 < \infty$.

Soit $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ la moyenne empirique. La loi faible (forte) des grands nombres (LGN) dit que

$$\bar{X}_n \rightarrow m \text{ en probabilité (presque sûrement) quand } n \rightarrow \infty.$$

Par conséquent, la distribution de \bar{X}_n est dégénérée.

Le théorème centrale limite (TCL) permet de contourner ce problème en introduisant une normalisation linéaire, telle que pour x fixé

$$\Pr(\bar{X}_n \leq \sigma_n x + m_n) \rightarrow \Phi(x) \text{ quand } n \rightarrow \infty,$$

où $m_n = \mathbb{E}(\bar{X}_n) = \mathbb{E}(X_1)$, $\sigma_n^2 = Var(\bar{X}_n) = n^{-1}Var(X_1)$ et Φ est la fonction de distribution de la loi normale centrée et réduite.

- Remarques sur le TCL:

- la distribution limite est la même quelque soit F (pourvu que $\sigma^2 < \infty$);
- la normalisation dépend des deux premiers moments de F ;
- le TCL suggère l'approximation

$$\bar{X}_n \sim N(m, \sigma^2/n)$$

comme un modèle pour la moyenne empirique quand n est fini et suffisamment grand.

Q : Quelle condition sur F assure que la limite $\Pr(M_n \leq u_n)$ existe pour une suite appropriée u_n ?

Proposition: Pour un $\tau > 0$ et une suite de réels (u_n) , les conditions suivantes sont équivalentes: quand $n \rightarrow \infty$

- 1) $\Pr(M_n \leq u_n) \rightarrow e^{-\tau},$
- 2) $n\bar{F}(u_n) = n(1 - F(u_n)) \rightarrow \tau.$

Théorème: Soit F une fonction de distribution de point extrémal $x^F \leq \infty$ et soit $\tau > 0$. Il existe une suite (u_n) satisfaisant $n\bar{F}(u_n) \rightarrow \tau$ quand $n \rightarrow \infty$ si et seulement si

$$\lim_{x \uparrow x^F} \frac{\bar{F}(x)}{\bar{F}(x-)} = 1,$$

où $\bar{F} = 1 - F$. Si les sauts de \bar{F} ne décroissent pas suffisamment vite, alors il n'existe pas de distribution limite non-dégénérée pour le maximum.

Q : Quelle conséquence ce choix de u_n a-t-il sur le nombre de dépassemens de ce seuil par X_1, \dots, X_n ?

Soit $N_n = \sum_{i=1}^n \mathbb{I}_{\{X_i > u_n\}}$ le nombre de dépassemens du seuil u_n par X_1, \dots, X_n .

Proposition: S'il existe une suite de réels (u_n) , telle que $n\bar{F}(u_n) \rightarrow \tau$ quand $n \rightarrow \infty$, alors N_n converge en loi vers N de loi de Poisson de paramètre τ , c'est-à-dire

$$\Pr(N = n) = e^{-\tau} \frac{\tau^n}{n!}, \quad n \geq 0.$$

Remarque: La loi de Poisson est souvent appelée la loi des évènements rares.

Q : *Est-il possible d'avoir une normalisation linéaire de M_n pour obtenir une distribution limite non-dégénérée?*

Exemple: $X_i \sim \text{Exponential}(1)$, telle que

$$F(x) = 1 - \exp(-x) \text{ pour } x > 0.$$

Puisque $x^F = \infty$, $M_n \rightarrow \infty$, mais

$$\begin{aligned} \Pr(M_n - \log n \leq x) &= \Pr(M_n \leq x + \log n) \\ &= [F(x + \log n)]^n \\ &= [1 - \exp(-x - \log n)]^n \text{ pour } x > -\log n \\ &= [1 - \exp(-x) \exp(-\log n)]^n \\ &= [1 - \exp(-x)/n]^n \\ &\rightarrow \exp[-\exp(-x)] \text{ quand } n \rightarrow \infty \text{ pour } -\infty < x < \infty. \end{aligned}$$

▷ Classes d'équivalence de distributions limites

Nous introduisons la notion de classe d'équivalence de distributions limites (distributions dites de même type).

Si F_1 et F_2 sont deux fonctions de distribution et s'il existe des constantes $a > 0$ et b telles que

$$F_2(ax + b) = F_1(x) \text{ pour tout } x,$$

alors F_1 et F_2 sont de même type.

Ainsi des distributions de même type sont les mêmes à des facteurs de position et d'échelle près.

• Exemples:

- $N(\mu_1, \sigma_1^2)$ et $N(\mu_2; \sigma_2^2)$ sont de même type;
- $\Gamma(\alpha, \beta)$ est la distribution Gamma de facteur d'échelle β et de forme α , alors $\Gamma(\alpha, \beta_1)$ et $\Gamma(\alpha, \beta_2)$ sont de même type, mais $\Gamma(\alpha_1, \beta)$ et $\Gamma(\alpha_2, \beta)$ ne sont pas de même type.

Théorème (Fisher-Tippett): S'il existe des suites de réels $c_n > 0$ et d_n , telles que, quand $n \rightarrow \infty$

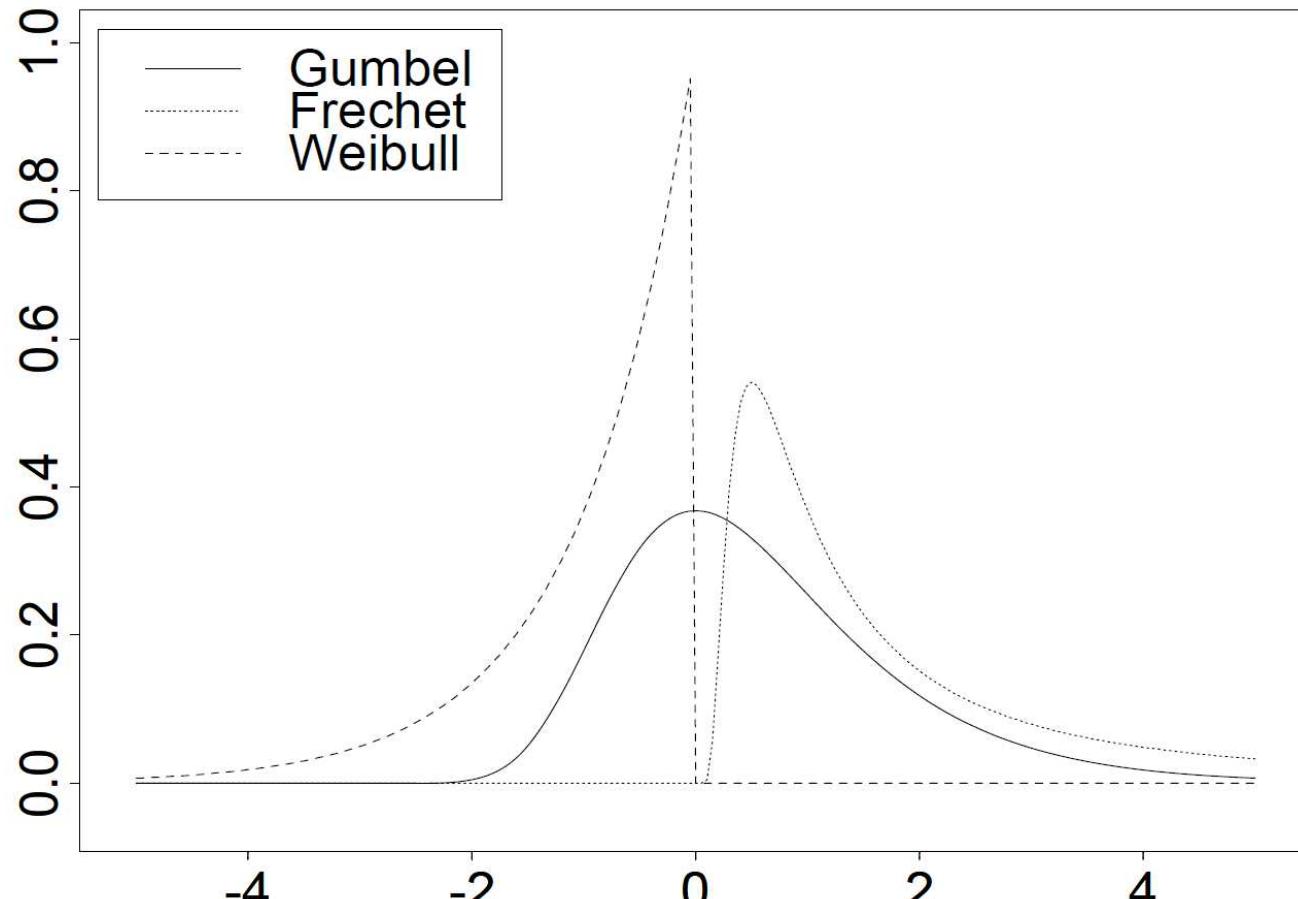
$$\Pr\left(\frac{M_n - d_n}{c_n} \leq x\right) \rightarrow G(x)$$

pour une distribution non-dégénée G , alors G est du même type que l'une des trois distributions suivantes:

$$\text{Fréchet } (\alpha > 0) : \quad \Phi_\alpha(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ \exp\{-x^{-\alpha}\} & \text{si } x > 0. \end{cases}$$

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

$$\text{Gumbel} : \quad \Lambda(x) = \exp\{-e^{-x}\} \quad x \in \mathbb{R}$$



Densités des distributions de Gumbel, Fréchet et Weibull

Proposition:

i) Soient X, X_1, \dots, X_n des variables aléatoires IID de loi de Fréchet Φ_α , alors

$$n^{-1/\alpha} \max(X_1, \dots, X_n) \stackrel{d}{=} X.$$

ii) Soient X, X_1, \dots, X_n des variables aléatoires IID de loi de Weibull Ψ_α , alors

$$n^{1/\alpha} \max(X_1, \dots, X_n) \stackrel{d}{=} X.$$

iii) Soient X, X_1, \dots, X_n des variables aléatoires IID de loi de Gumbel Λ , alors

$$\max(X_1, \dots, X_n) - \ln n \stackrel{d}{=} X.$$

On parle de distributions *max-stables*.

Remarque: Si $X \sim \Phi_\alpha \iff -X^{-1} \sim \Psi_\alpha \iff \ln X^\alpha \sim \Lambda$.

▷ Distributions des extrêmes généralisées (GEV)

Dans certains cas (en particulier pour des problèmes statistiques), il est préférable d'avoir une seule distribution qui unifie les trois précédentes.

Von Mises (1954) et Jenkinson (1955) ont proposé la distribution $GEV(\mu, \sigma, \xi)$

$$G(x) = \begin{cases} \exp\left(-\left[1 + \xi \left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right) & \text{si } \xi \neq 0, \\ \exp\left(-\exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right)\right) & \text{si } \xi = 0, \end{cases}$$

où $x_+ = \max(x, 0)$ et $\sigma > 0$. ξ est le paramètre de forme, μ le paramètre de position, σ le paramètre d'échelle.

On notera

$$G_\xi(x) = \exp\left(-(1 + \xi x)_+^{-1/\xi}\right).$$

- La distribution de Gumbel correspond à $\xi = 0$; $GEV(0, 1, 0) = \text{Gumbel}$;
- La distribution de Fréchet correspond à $\xi > 0$; $GEV(1, \alpha^{-1}, \alpha^{-1}) = \text{Fréchet}(\alpha)$;
- La distribution de Weibull correspond à $\xi < 0$; $GEV(-1, \alpha^{-1}, -\alpha^{-1}) = \text{Weibull}(\alpha)$.

Proposition:

Soient X, X_1, \dots, X_n des variables aléatoires IID de loi $GEV(0, 1, \xi)$, alors

$$n^{-\xi} \left(\max(X_1, \dots, X_n) - \frac{n^\xi - 1}{\xi} \right) \stackrel{d}{=} X.$$

Les distributions des extrêmes généralisées sont des distributions *max-stables*.

Reformulation du théorème de Fisher-Tippett: S'il existe des suites de réels $a_n > 0$ et b_n , telles que, quand $n \rightarrow \infty$,

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) \rightarrow G(x)$$

pour une distribution non-dégénérée G , alors G est du même type que la distribution suivante:

$$G_\xi(x) = \exp\left(-(1 + \xi x)_+^{-1/\xi}\right).$$

- Remarques sur le théorème de convergence du maximum:

- ξ est aussi appelé parfois l'indice de queue et donne une information sur l'épaisseur des queues de distribution;
- $\xi > 0$: "Queue épaisse", $\xi = 0$: "Queue intermédiaire", $\xi < 0$: "Queue fine";
- Le théorème précédent ne garantit pas l'existence d'une limite non-dégénérée, ni ne dit quelle valeur de ξ apparaît dans la loi limite (cela dépend de F);
- Contrairement au TCL, le théorème précédent ne dit pas comment choisir les suites a_n et b_n ;
- Contrairement au TCL (dans le cas $\sigma^2 < \infty$), le théorème propose une famille de lois limites (paramétrées par ξ).

⚠ Si $\sigma^2 = \infty$, alors il existe une constante $c > 1/2$ telle que

$$\Pr\left(n^{1-c}(\bar{X}_n - a) \leq x\right) \rightarrow SSL_c(x) \text{ quand } n \rightarrow \infty$$

où SSL désigne une loi stable par addition (Sum-Stable-Law distribution) et $a = m$ si $c \in (1/2; 1)$ et 0 sinon.

Donc de manière plus générale, les sommes de variables aléatoires ont des lois limites paramétrées par c .

Les lois SSL n'ont en général pas d'expression analytique pour leur densité où leur fonction de répartition (un contre exemple est la distribution de Cauchy).

La constante c est liée à l'épaisseur de la queue de distribution de F . Si F est une distribution symétrique et si la loi limite du maximum est non-dégénérée, alors $c = \xi$.

En général, il y a un lien fort entre la théorie des lois limites des sommes et celles des maximums lorsque les queues de distribution sont épaisses.

▷ Pseudo preuves

Nous essayons d'expliquer que, si une loi limite non-dégénérée existe, elle doit être de la même forme que donnée précédemment.

Preuve pour le TCL

Par simplicité, nous supposons que $\mathbb{E}(X_i) = 0$ et $Var(X_i) = 1$. Soit S_n la moyenne empirique \bar{X}_n normalisée par $n^{1/2}$,

$$S_n = n^{1/2} \bar{X}_n,$$

alors $\mathbb{E}(S_n) = 0$, $Var(S_n) = 1$.

Supposons que S_n converge en distribution vers une variable aléatoire Y de distribution inconnue.

Cela signifie que, lorsque n est grand, la distribution de S_n est bien approximée par celle de Y , ou que celle de $\sum_{i=1}^n X_i$ est bien approximée par celle de $n^{1/2}Y$.

Considérons la somme $\sum_{i=1}^n X_i$ comme une somme de sommes. Plus précisément, soit k un entier et posons $r_n = [n/k]$.

$$\sum_{i=1}^n X_i \simeq \sum_{j=1}^k \sum_{i=r_n(j-1)+1}^{r_n j} X_i.$$

Les sommes partielles sont indépendantes car ce sont des sommes de variables aléatoires indépendantes.

Si n est grand, c'est aussi le cas pour r_n , donc chaque somme partielle a une distribution qui peut être approximée par $r_n^{1/2}Y$.

Si Y_1, \dots, Y_k sont des variables indépendantes de même distribution que Y , on a alors

$$\begin{aligned} n^{1/2}Y &= r_n^{1/2}Y_1 + \dots + r_n^{1/2}Y_k \\ k^{1/2}Y &= Y_1 + \dots + Y_k. \end{aligned}$$

Ainsi la distribution de Y a la propriété que, lorsque l'on somme des variables aléatoires indépendantes de même loi que Y , alors la somme a la même loi que Y à un coefficient d'échelle près.

C'est exactement la propriété de stabilité par addition (convolution) de la distribution Normale.

Puisque la loi Normale est la seule loi (avec une variance finie) à posséder cette propriété, on en déduit que la loi limite est nécessairement la loi Normale lorsque la variance des X_i est finie.

Preuve pour la loi limite des maximums

Supposons que $(M_n - b_n)/a_n$ converge en distribution vers une variable aléatoire Y de distribution inconnue.

Cela signifie que, lorsque n est grand, la distribution de $(M_n - b_n)/a_n$ est bien approximée par celle de Y , ou que celle de M_n est bien approximée par celle de $a_n Y + b_n$.

Considérons le maximum M_n comme le maximum de maximums. Plus précisément, soit k un entier et posons $r_n = [n/k]$.

$$M_n \simeq \max_{j=1,\dots,k} \max(X_{r_n(j-1)+1}, \dots, X_{r_n j}).$$

Les maximums partiels sont indépendants car ce sont des maximums de variables aléatoires indépendantes.

Si n est grand, c'est aussi le cas pour r_n , donc chaque maximum partiel a une distribution qui peut être approximée par $a_{r_n}Y + b_{r_n}$.

Si Y_1, \dots, Y_k sont des variables indépendantes de même distribution que Y , on a alors

$$\begin{aligned} a_nY + b_n &= a_{r_n} \max(Y_1, \dots, Y_k) + b_{r_n} \\ \frac{a_nY + b_n - b_{r_n}}{a_{r_n}} &= \max(Y_1, \dots, Y_k) \end{aligned}$$

Ainsi la distribution de Y a la propriété que, lorsque l'on prend le maximum de variables aléatoires indépendantes de même loi que Y , alors le maximum a la même loi que Y à un coefficient d'échelle près.

Cette propriété peut être écrite en termes de fonction de distribution

$$G(A_kx + B_k) = [G(x)]^k \text{ pour des constantes } A_k > 0 \text{ et } B_k$$

pour tout k . Cette propriété est appelée stabilité par maximum (max-stability). La distribution GEV est la seule distribution à satisfaire la stabilité par maximum.

Ceci explique, pourquoi la distribution limite des maximums (si elle existe) doit être une distribution de type GEV.

▷ Domaines d'attraction

Le théorème de convergence de la loi limite du maximum énonce que si

$$\Pr \left(\frac{M_n - b_n}{a_n} \leq x \right) = [F(a_n x + b_n)]^n \rightarrow G(x)$$

où G est une distribution non-dégénérée, alors G a une distribution des extrêmes généralisée.

Définition: On dit que F appartient au domaine d'attraction de G ($F \in D(G)$) s'il existe deux suites (a_n) et (b_n) telle que la convergence précédente ait lieu.

Nous souhaitons comprendre les conditions sur F qui conduisent à une distribution particulière G , caractérisée par ξ . Cela veut dire:

- trouver a_n et b_n pour une distribution F ;
- déterminer les situations pour lesquelles il n'existe pas de suite de normalisation a_n et b_n ;
- identifier les interconnexions avec les lois stables par addition;
- analyser les taux de convergence de maximum M_n normalisé vers une distribution de type GEV.

Rappel: stabilité par maximum

Si $X_i \sim \text{GEV}(0, 1, \xi)$, i.e.

$$F(x) = \exp\left(-(1 + \xi x)_+^{-1/\xi}\right),$$

alors en choisissant $a_n = n^\xi$ et $b_n = (n^\xi - 1)/\xi$, on a

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = \exp\left(-(1 + \xi x)_+^{-1/\xi}\right),$$

et M_n a une distribution GEV avec paramètre de forme ξ pour tout n . Ceci est une conséquence immédiate de la propriété de stabilité par maximum des distributions GEV.

Il n'y a pas de convergence ici, la limite est vraie pour tout n .

Une remarque préliminaire

i) Si $X_i \sim \text{Exponential}(1)$, alors, quand $n \rightarrow \infty$,

$$M_n - \log n \xrightarrow{d} \Lambda$$

et si $X_i \sim \Lambda$ alors

$$M_n - \log n \stackrel{d}{=} \Lambda.$$

On notera que

$$\lim_{x \rightarrow \infty} \frac{\Pr_{\text{Exp}}(X_i > x)}{\Pr_{\text{Gumbel}}(X_i > x)} = 1.$$

ii) Si $X_i \sim U[-1, 0]$, alors, quand $n \rightarrow \infty$,

$$nM_n \xrightarrow{d} \Psi_1$$

et si $X_i \sim \Psi_1$ alors

$$nM_n \xrightarrow{d} \Psi_1.$$

On notera que

$$\lim_{x \rightarrow 0} \frac{\Pr_{\text{Uni}}(X_i > x)}{\Pr_{\text{Weibull}(1)}(X_i > x)} = 1.$$

iii) Si $X_i \sim \text{Cauchy}$, alors, quand $n \rightarrow \infty$,

$$\frac{\pi}{n} M_n \xrightarrow{d} \Phi_1$$

et si $X_i \sim \Phi_1$ alors

$$\frac{1}{n} M_n \xrightarrow{d} \Phi_1.$$

On notera que

$$\lim_{x \rightarrow \infty} \frac{\Pr_{\text{Cauchy}}(X_i > x)}{\Pr_{\text{Fréchet}(1)}(X_i > x)} = \frac{1}{\pi}.$$

Définition: Deux distributions F et G sont équivalentes en termes de queue de distribution si elles ont le même point extrémal ($x^F = x^G$) et

$$\lim_{x \rightarrow x^F} \frac{\bar{F}(x)}{\bar{G}(x)} = c \in (0, \infty).$$

Proposition: Si F et G sont équivalentes en termes de queue de distribution, elles appartiennent au même domaine d'attraction.

Une intuition sur les conditions des domaines d'attraction

Considérons une variable aléatoire absolument continue et notons f sa densité. On peut définir l'inverse de la fonction de hasard par

$$h(x) = \frac{1 - F(x)}{f(x)} \text{ pour } x_F < x < x^F,$$

où x_F et x^F sont les valeurs qui définissent le support de la variable aléatoire.

Des conditions plus faibles que l'existence d'une densité peuvent être utilisées, mais cela nuit à l'intuition du résultat.

Après une suite de calculs, on peut montrer que pour tout u et x , il existe un y tel que $u \leq y \leq u + xh(u)$ et

$$\frac{1 - F(u + xh(u))}{1 - F(u)} = [1 + h'(y)x]_+^{-1/h'(y)}.$$

Supposons que $h'(y) \rightarrow \xi$ quand $y \rightarrow x^F$ (avec ξ fini).

Définissons

- b_n comme le quantile d'ordre $1 - 1/n$, i.e. $1 - F(b_n) = 1/n$,
- $a_n = h(b_n)$, i.e. la transformée de b_n par l'inverse de la fonction de hasard.

Fixons x et laissons $u = b_n$ tendre vers x^F . Alors on obtient

$$n[1 - F(a_n x + b_n)] \rightarrow (1 + \xi x)_+^{-\frac{1}{\xi}} \text{ quand } n \rightarrow \infty.$$

Comme $-\log(x) = -\log[1 - (1 - x)] \approx 1 - x$ quand $x \uparrow 1$, $n[1 - F(a_n x + b_n)]$ est approximativement égal à $-n \log F(a_n x + b_n)$, on obtient la convergence

$$[F(a_n x + b_n)]^n \rightarrow \exp\left(-(1 + \xi x)_+^{-\frac{1}{\xi}}\right).$$

Ainsi, dans ce cas, la caractérisation des conditions d'appartenance au domaine d'attraction est simple: une distribution GEV($0, 1, \xi$) est obtenue comme limite si

$$h'(y) \rightarrow \xi \text{ quand } y \rightarrow x^F,$$

et nous choisissons alors a_n et b_n tels que:

$$1 - F(b_n) = 1/n \text{ et } a_n = h(b_n).$$

Remarquons que la distribution de Gumbel est obtenue comme limite si $\xi = 0$, i.e. si l'inverse de la fonction de hasard est approximativement constante pour les grandes valeurs de la variable aléatoire (ceci arrive fréquemment en pratique).

Exemples:

- Loi Exponentielle de paramètre 1:

Soit $F(x) = 1 - \exp(-x)$, $f(x) = \exp(-x)$ et ainsi $h(x) = 1$ pour tout $x > 0$ (fonction de hasard constante). Alors

$$h'(x) = 0 \text{ pour tout } x$$

et $\xi = 0$. Donc $\exp(-b_n) = 1/n$, i.e. $b_n = \log n$ et $a_n = h(b_n) = 1$.

Ce sont exactement les coefficients de normalisation trouvés précédemment et de plus nous retrouvons que

$$M_n - \log n$$

converge vers la distributions de Gumbel.

- Distribution Normale

Pour des grandes valeurs de x

$$1 - \Phi(x) \approx (2\pi)^{-1/2} \exp(-x^2/2)(x^{-1} - x^{-3} + 3x^{-5} + \dots).$$

Par conséquent $h(x) = x^{-1} - x^{-3} + \dots$, et

$$h'(x) = -x^{-2} + 3x^{-4} + \dots$$

Donc $\xi = 0$ et a_n et b_n peuvent être caractérisés analytiquement par

$$b_n = (2 \log n)^{1/2} - \frac{1}{2}(2 \log n)^{-1/2} [\log(\log n) + \log(4\pi)] \text{ et } a_n = (2 \log n)^{-1/2}.$$

Remarquons que, puisque $a_n \rightarrow 0$, alors $M_n \sim b_n$, i.e. $M_n \sim (2 \log n)^{1/2}$.

Les maximums de variables aléatoires Gaussiennes augmentent très doucement et de manière déterministe!

- Distributions de type Pareto

Les distributions de type Pareto sont définies par

$$1 - F(x) \sim c/x^\alpha \text{ quand } x \rightarrow \infty$$

pour $c > 0$ et $\alpha > 0$. Des exemples sont les distributions de Pareto, de Cauchy, les distributions t et \mathcal{F} . Si la densité existe, alors

$$f(x) \sim \alpha cx^{-\alpha-1}, h(x) \sim \alpha^{-1}x \text{ et } h'(x) \sim \alpha^{-1} \text{ quand } x \rightarrow \infty.$$

Par conséquent $\xi = \alpha^{-1}$, $b_n = (cn)^{1/\alpha}$ et $a_n = \alpha^{-1}(cn)^{1/\alpha}$. La loi de Fréchet(α) est donc obtenue comme limite.

Remarque: supposons que la distribution est symétrique. Si $\alpha < 2$ la variance n'existe pas,

$$n^{1-1/\alpha} \bar{X}_n \rightarrow SSL_{1/\alpha} \text{ et } \frac{M_n - b_n}{a_n} \rightarrow GEV(0, 1, \alpha^{-1}) \text{ quand } n \rightarrow \infty.$$

On peut montrer que, dans ce cas, \bar{X}_n et M_n sont asymptotiquement dépendants car quelques rares et grandes valeurs de X_1, \dots, X_n dominent les variations de \bar{X}_n .

▷ Le domaine d'attraction de la distribution de Fréchet

Définition: Une fonction L mesurable et positive sur $]0, \infty[$ est à *variations lentes*, si, pour tout $x > 0$, $\lim_{t \rightarrow \infty} L(tx)/L(t) = 1$.

Théorème: La distribution F appartient au domaine d'attraction de la loi de Fréchet Φ_α ($\alpha > 0$), si et seulement si $\bar{F}(x) = x^{-\alpha}L(x)$, pour une fonction à variations lentes L .

Nous pouvons alors choisir $d_n = 0$ et $c_n = F^\leftarrow(1 - n^{-1})$, où

$$F^\leftarrow(t) = \inf \{x \in \mathbb{R} : F(x) \geq t\}.$$

Exemples: Distribution de Cauchy, distribution de Pareto, distribution Loggamma,...

Maximum domain of attraction of the Fréchet distribution.

Fréchet	$\Phi_\alpha(x) = \exp\{-x^{-\alpha}\}, \quad x > 0, \alpha > 0.$
MDA (Φ_α)	$x_F = \infty, \quad \bar{F}(x) = x^{-\alpha}L(x), \quad L \in \mathcal{R}_0.$
Norming constants	$c_n = F^\leftarrow(1 - n^{-1}) = n^{1/\alpha}L_1(n), \quad L_1 \in \mathcal{R}_0, \quad d_n = 0.$
Limit result	$c_n^{-1}M_n \xrightarrow{d} \Phi_\alpha$
Examples	
Cauchy	$f(x) = (\pi(1 + x^2))^{-1}, \quad x \in \mathbb{R}.$ $c_n = n/\pi$
Pareto	$\bar{F}(x) \sim Kx^{-\alpha}, \quad K, \alpha > 0.$
Burr stable with index $\alpha < 2$	$c_n = (Kn)^{1/\alpha}$
Loggamma	$f(x) = \frac{\alpha^\beta}{\Gamma(\beta)}(\ln x)^{\beta-1}x^{-\alpha-1}, \quad x > 1, \quad \alpha, \beta > 0.$ $c_n = ((\Gamma(\beta))^{-1}(\ln n)^{\beta-1}n)^{1/\alpha}$

▷ Le domain d'attraction de la loi de Gumbel

Théorème: La distribution F appartient au domaine d'attraction de la loi de Gumbel Λ , si et seulement si il existe une fonction positive g telle que :

$$\lim_{x \rightarrow x^F} \frac{\bar{F}(x + tg(x))}{\bar{F}(x)} = \exp(-t), \quad t \in \mathbb{R}.$$

Un choix possible pour la fonction g est

$$g(x) = \frac{\int_x^{x^F} \bar{F}(t) dt}{\bar{F}(x)} = \mathbb{E}(X - x | X > x).$$

Les coefficients c_n et d_n ont alors la forme suivante : $d_n = F^\leftarrow(1 - n^{-1})$, $c_n = g(d_n)$.

Exemples: Distribution Normale, distribution Exponentielle...

Maximum domain of attraction of the Gumbel distribution.

Gumbel	$\Lambda(x) = \exp\{-e^{-x}\} , \quad x \in \mathbb{R} .$
MDA(Λ)	$x_F \leq \infty , \quad \bar{F}(x) = c(x) \exp\left\{-\int_{x_0}^x \frac{g(t)}{a(t)} dt\right\} , \quad x_0 < x < x_F ,$ where $c(x) \rightarrow c > 0 , \quad g(x) \rightarrow 1 , \quad a'(x) \rightarrow 0 \quad \text{as } x \uparrow x_F .$
Norming constants	$d_n = F^\leftarrow(1 - n^{-1}) , \quad c_n = a(d_n) .$
Limit result	$c_n^{-1}(M_n - d_n) \xrightarrow{d} \Lambda$
Examples	
Exponential-like	$\bar{F}(x) \sim K e^{-\lambda x} , \quad K, \lambda > 0 .$ $c_n = \lambda^{-1}$ $d_n = \lambda^{-1} \ln(Kn)$
Weibull-like	$\bar{F}(x) \sim K x^\alpha \exp\{-cx^\tau\} , \quad K, c, \tau > 0 , \alpha \in \mathbb{R} .$ $c_n = (c\tau)^{-1} (c^{-1} \ln n)^{1/\tau-1}$ $d_n = (c^{-1} \ln n)^{1/\tau} + \frac{1}{\tau} (c^{-1} \ln n)^{1/\tau-1} \left\{ \frac{\alpha}{c\tau} \ln(c^{-1} \ln n) + \frac{\ln K}{c} \right\}$

Gamma	$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0, \quad \alpha, \beta > 0.$ $c_n = \beta^{-1}$ $d_n = \beta^{-1} (\ln n + (\alpha - 1) \ln \ln n - \ln \Gamma(\alpha))$
Normal	$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}.$ $c_n = (2 \ln n)^{-1/2}$ $d_n = \sqrt{2 \ln n} - \frac{\ln(4\pi) + \ln \ln n}{2(2 \ln n)^{1/2}}$
Lognormal	$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-(\ln x - \mu)^2/2\sigma^2}, \quad x > 0, \quad \mu \in \mathbb{R}, \sigma > 0.$ $c_n = \sigma(2 \ln n)^{-1/2} d_n$ $d_n = \exp \left\{ \mu + \sigma \left(\sqrt{2 \ln n} - \frac{\ln(4\pi) + \ln \ln n}{2(2 \ln n)^{1/2}} \right) \right\}$

▷ Le domaine d'attraction de la loi de Weibull

Théorème: La distribution F appartient au domaine d'attraction de la loi de Weibull Ψ_α , si et seulement si $x^F < \infty$ et $\bar{F}(x^F - x^{-1}) = x^{-\alpha}L(x)$ pour une fonction à variations lentes L .

Les constantes de normalisation peuvent être choisies de la manière suivante

$$d_n = F^\leftarrow(1), c_n = x^F - F^\leftarrow(1 - n^{-1}).$$

Exemples: Distribution Beta, distribution uniforme...

Maximum domain of attraction of the Weibull distribution.

Weibull	$\Psi_\alpha(x) = \exp\{-(x)^{\alpha}\} , \quad x < 0 , \quad \alpha > 0 .$
MDA (Ψ_α)	$x_F < \infty , \quad \bar{F}(x_F - x^{-1}) = x^{-\alpha} L(x) , \quad L \in \mathcal{R}_0 .$
Norming constants	$c_n = x_F - F^\leftarrow(1 - n^{-1}) = n^{-1/\alpha} L_1(n) , \quad L_1 \in \mathcal{R}_0 , \quad d_n = x_F .$
Limit result	$c_n^{-1}(M_n - x_F) \xrightarrow{d} \Psi_\alpha$

Examples

Uniform	$f(x) = 1 , \quad x \in (0, 1) .$ $c_n = n^{-1} , \quad d_n = 1$
Power law behaviour at x_F	$\bar{F}(x) = K(x_F - x)^\alpha , \quad x_F - K^{-1/\alpha} \leq x \leq x_F , \quad K, \alpha > 0 .$ $c_n = (K n)^{-1/\alpha} , \quad d_n = x_F$
Beta	$f(x) = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1 - x)^{b-1} , \quad 0 < x < 1 , \quad a, b > 0 .$ $c_n = \left(n \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b + 1)} \right)^{-1/b} , \quad d_n = 1$

▷ Exemples de distributions qui n'appartiennent pas à un domaine d'attraction

Des exemples de distributions qui n'appartiennent pas à un domaine d'attraction sont:

$$F(x) = 1 - 1/\log x \text{ pour } x > e$$

$$F(x) = 1 - 1/\log(\log x) \text{ for } x > e^e.$$

Pour le premier exemple

$$f(x) = \frac{1}{x(\log x)^2}, \quad h(x) = x \log x \text{ et } h'(x) = 1 + \log x$$

ainsi $h'(x) \rightarrow \infty$ quand $x \rightarrow \infty$, ce qui ne satisfait pas la condition de convergence. La queue de distribution est trop épaisse.

Dans ce cas, il faut envisager une normalisation non-linéaire.

Soit $Y = \log X$. Alors $F_Y(y) = F_X(\exp(y)) = 1 - 1/y$. Ainsi pour Y

$$h(y) = y \text{ et } h'(y) = 1 = \xi$$

et le domaine d'attraction est celui de la loi Fréchet de paramètre $\alpha = 1$.

Ceci suggère l'idée d'utiliser des transformations avant d'essayer d'appliquer la théorie standard (mais c'est une approche aussi bien connue en analyse statistique classique).

▷ Vitesse de convergence

Il est important d'avoir une idée de la vitesse de convergence vers la distribution limite car ceci donne une indication de la possibilité de considérer la distribution limite en pratique (lorsque le nombre de données peut être faible). Soit

$$e_n = \sup_x |[F(a_n x + b_n)]^n - G(x)|.$$

Nous pouvons identifier les facteurs qui influencent la valeur de e_n . Les approximations commises sont les suivantes:

- on a remplacé $h'(y)$ par ξ pour un $y > u = b_n$. Cette erreur peut être quantifiée par $h'(b_n) - \xi$, mais dans certains cas une analyse plus fine est nécessaire;
- on a remplacé $n[1 - F(a_n x + b_n)]$ par $-\log [F(a_n x + b_n)]^n$, ce qui est de l'ordre de $O(n^{-1})$.

Ainsi une approximation raisonnable est donnée par

$$O(e_n) = \max \left\{ O(n^{-1}); O(h'(b_n) - \xi) \right\}.$$

- Loi Exponentielle(1)

Précédemment nous avions trouvé $h'(x) = 0$ et $\xi = 0$, et par conséquent e_n est de l'ordre de $O(n^{-1})$, ce qui est très rapide (dans le cas du TCL avec $\sigma^2 < \infty$, l'erreur est de l'ordre de $O(n^{-1/2})$).

- Loi Normale centrée et réduite $N(0,1)$

Précédemment nous avions trouvé

$$h'(x) = -x^{-2} + 3x^{-4} + \dots \text{ et } \xi = 0.$$

et ainsi e_n est de l'ordre de $O((\log n)^{-1})$, ce qui est très lent.

▷ Convergence des statistiques d'ordre

Théorème: Soit (u_n) une suite de réels telle que $\lim_{n \rightarrow \infty} n\bar{F}(u_n) = \tau$. Alors $N_n = \sum_{i=1}^n \mathbb{I}_{\{X_i > u_n\}}$ converge en loi vers un loi de Poisson de paramètre τ . Si $X_{(k)}$ est la k -ième plus grande valeur parmi X_1, \dots, X_n , alors

$$\lim_{n \rightarrow \infty} \Pr(X_{(k)} \leq u_n) = \lim_{n \rightarrow \infty} \Pr(N_n < k) = e^{-\tau} \sum_{i=0}^{k-1} \frac{\tau^i}{i!}.$$

Si l'on peut choisir de plus u_n sous la forme $u_n = a_n x + b_n$ avec $\tau = (-\ln G(x))$ où G est une distribution GEV, alors

$$\lim_{n \rightarrow \infty} \Pr\left(\frac{X_{(k)} - b_n}{a_n} \leq x\right) = G(x) \sum_{i=0}^{k-1} \frac{(-\ln G(x))^i}{i!}.$$

2. Les dépassements de seuils

Plutôt que de considérer le maximum M_n d'un échantillon X_1, \dots, X_n , on s'intéresse aux $N_n = \sum_{i=1}^n \mathbb{I}_{\{X_i > u_n\}}$ dépassements du seuil u_n , c'est-à-dire aux observations

$$(X_i - u_n)_+$$

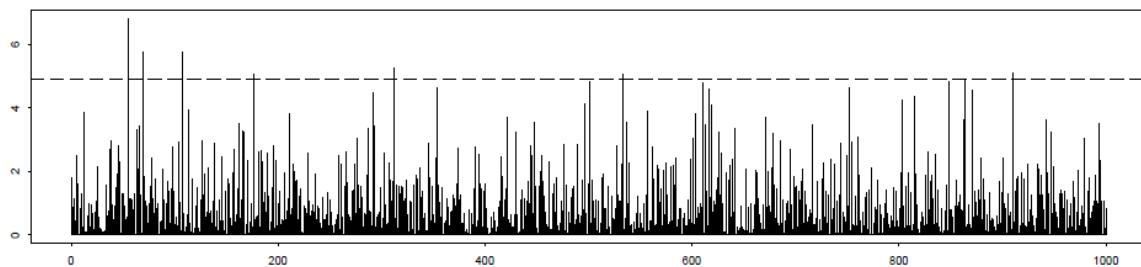
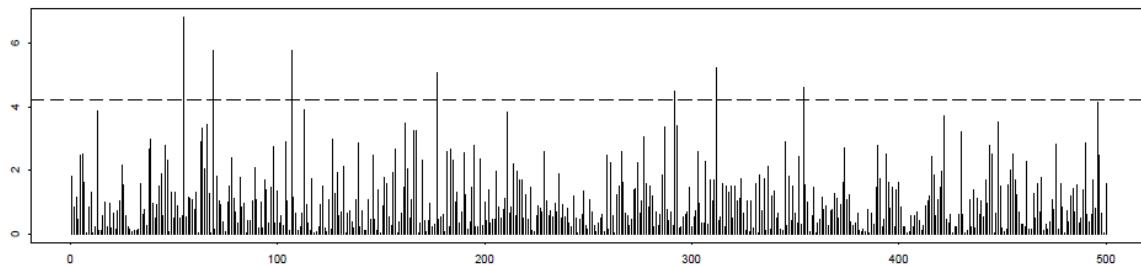
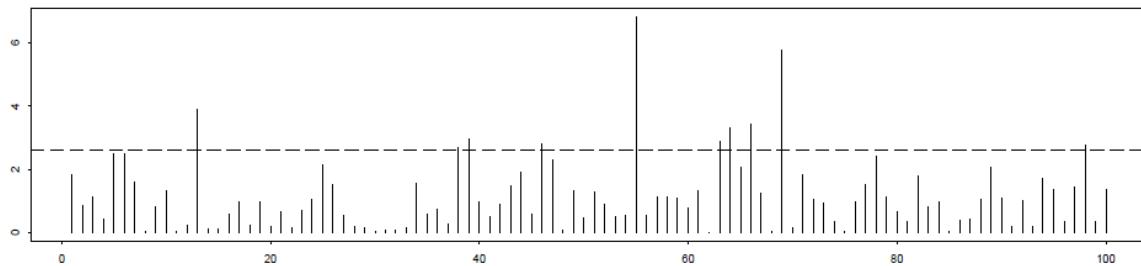
qui sont strictement positives. Elles sont caractérisées par la distribution de Pareto généralisée.

Définition: La distribution de Pareto généralisée GPD(β, ξ) est définie par

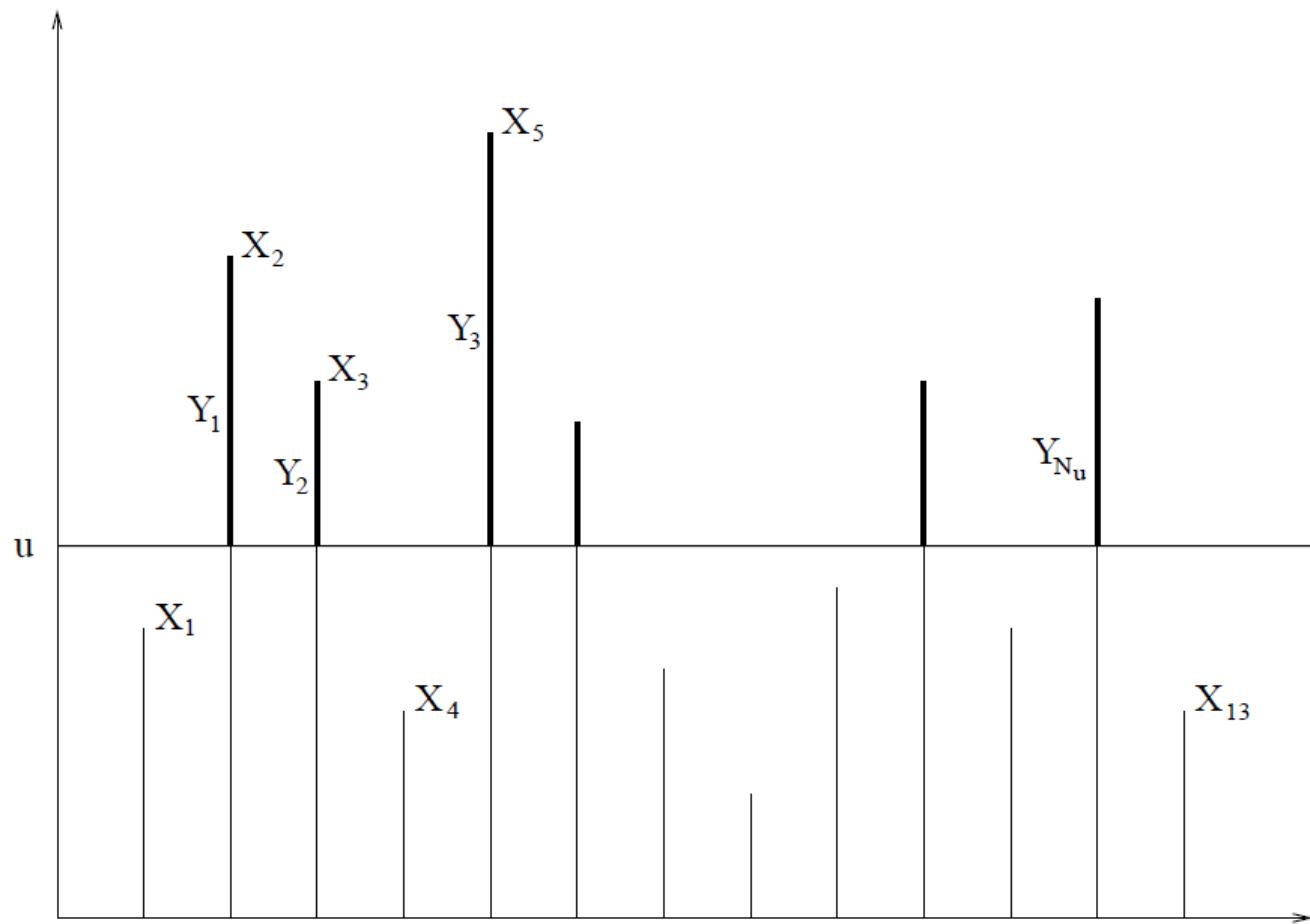
$$G_{\xi, \beta}^p(x) = \begin{cases} 1 - [1 + \xi (x/\beta)]_+^{-1/\xi} & \text{si } \xi \neq 0, \\ 1 - e^{-x/\beta} & \text{si } \xi = 0, \end{cases}$$

où

$$\begin{aligned} x &\geq 0 && \text{si } \xi \geq 0, \\ 0 \leq x &\leq -\beta/\xi && \text{si } \xi < 0. \end{aligned}$$



Dépassements de seuils en fonction de la taille de l'échantillon



Quelques propriétés des GPDs:

- Si $\xi = -1$, $G_{-1,\beta}^p$ est la loi uniforme sur $[0, \beta]$. Si $\xi = 0$, $G_{0,\beta}^p$ est la loi exponentielle de paramètre $1/\beta$.

- La densité de $G_{\xi,\beta}^p$ est donnée par

$$f(x) = \frac{1}{\beta} \left[1 + \xi \left(\frac{x}{\beta} \right) \right]_+^{-1/\xi-1}$$

- Si $U \sim U[0, 1]$, alors

$$Y \sim \beta \left(\frac{U^{-\xi} - 1}{\xi} \right)$$

a une loi $GPD(\beta, \xi)$.

- Si $X \sim GPD(\beta, \xi)$ et $\xi < 1$, alors

$$\mathbb{E}(X) = \frac{\beta}{1 - \xi}.$$

- Si $X \sim \text{GPD}(\beta, \xi)$, alors $X - x | X > x \sim \text{GPD}(\beta + \xi x, \xi)$ et

$$\mathbb{E}(X - x | X > x) = \frac{\beta + \xi x}{1 - \xi}.$$

Quelques notations:

- $\bar{F}_u(x) = \bar{F}(u + x) / \bar{F}(u)$

- $U(t) = F^{\leftarrow}(1 - t^{-1})$ où F^{\leftarrow} est l'inverse généralisée de la fonction de distribution F , i.e. $F^{\leftarrow}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\}$.

Le théorème suivant explique que la convergence du maximum de variables aléatoires IID correctement normalisées est équivalente à la convergence de la distribution des dépassements vers une distribution Pareto Généralisée.

Théorème: Soit $\xi \in \mathbb{R}$, les propositions suivantes sont équivalentes:

- Il existe deux suites a_n et b_n telles que $[F(a_n x + b_n)]^n \rightarrow G_\xi(x)$.

- Il existe une fonction $a(\cdot)$ telle que

$$\lim_{u \rightarrow x^F} \frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \begin{cases} [1 + \xi x]_+^{-1/\xi} & \text{si } \xi \neq 0, \\ e^{-x} & \text{si } \xi = 0. \end{cases}$$

- Pour $x, y > 0$, $y \neq 1$

$$\lim_{s \rightarrow \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} = \begin{cases} (x^\xi - 1) / (y^\xi - 1) & \text{si } \xi \neq 0, \\ \log x / \log y & \text{si } \xi = 0. \end{cases}$$

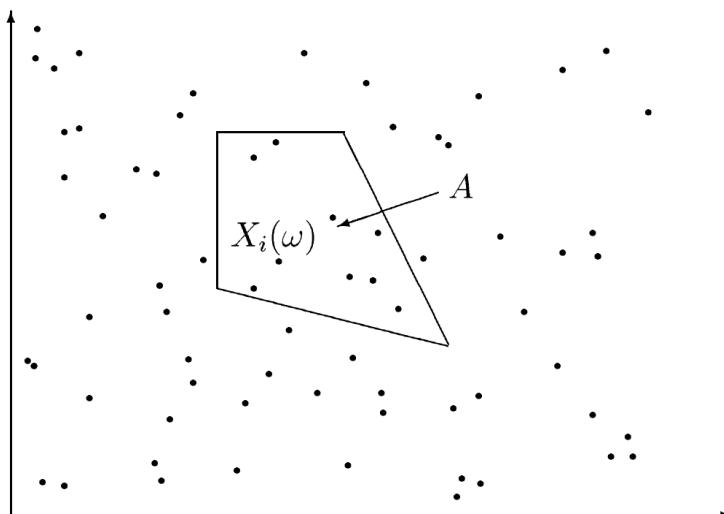
- Il existe une fonction positive $\beta(\cdot)$ telle que

$$\lim_{u \rightarrow x^F} \sup_{0 < x < x^F - u} \left| F_u(x) - G_{\xi, \beta(u)}^p(x) \right| = 0.$$

▷ Dynamique des dépassements de seuils (apparté mathématique)

Quelques notions sur les processus ponctuels

Un processus ponctuel N peut être considéré comme une distribution aléatoire de points X_i dans un espace E . Il est caractérisé par les distributions du nombre de points pour tout ensemble A de l'espace: $N(A)$ compte le nombre de X_i dans A .



Pour une suite donnée $(x_i)_{i \geq 1}$ de E ,

$$m(A) = \sum_{i=1}^{\infty} \mathbb{I}_{\{x_i \in A\}}$$

définit une mesure de comptage qui est appelée mesure ponctuelle si $m(K) < \infty$ pour tout ensemble compact $K \subset E$.

Soit $M_p(E)$ l'espace des mesures ponctuelles sur E (équipée d'une σ -algèbre $\mathcal{M}_p(E)$ appropriée).

Définition: Un processus ponctuel sur E est une fonction mesurable

$$N : [\Omega, \mathcal{F}, P] \rightarrow [M_p(E), \mathcal{M}_p(E)]$$

Il est possible de considérer un processus ponctuel comme un ensemble $(N(A))_{A \in \mathcal{E}}$ où \mathcal{E} est la σ -algèbre de E générée par les ouverts.

Exemples:

- Le processus ponctuel des dépassements normalisés $N_n^{(\tau)}(\cdot)$ est défini par

$$N_n^{(\tau)}(B) = \sum_{i=1}^n \mathbb{I}_{\{i/n \in B, X_i > u_n(\tau)\}},$$

pour tout Borelien $B \subset E := (0, 1]$, où $(u_n(\tau))$ est une suite de seuils déterministes.

- Le processus ponctuel des dépassements bi-dimensionnel $N_n^{(\tau)}(\cdot)$ est défini par

$$N_n^{(\tau)}(B) = \sum_{i=1}^n \mathbb{I}_{\{(i/n, (X_i - b_n)/a_n) \in B\}},$$

pour tout Borélien $B \subset E := (0, 1] \times \mathbb{R}$.

Les réalisations d'un processus ponctuel N sont des mesures ponctuelles. Par conséquent la loi de N est caractérisée par la famille des mesures ponctuelles

$$P_N(A) = P(N \in A), \quad A \in \mathcal{M}_p(E).$$

La distribution de N est déterminée par la distribution des vecteurs (de dimension finie)

$$(N(A_1), \dots, N(A_m))$$

pour n'importe quel choix de A_1, \dots, A_m et de $m \geq 1$.

La distribution de N est aussi caractérisée par sa transformée de Laplace

$$L_N(g) = E \exp \left\{ - \int_E g dN \right\} = \int_{M_p(E)} \exp \left\{ - \int_E g(x) dm(x) \right\} dP_N(m)$$

définie pour toute fonction positive et mesurable g .

Définition: Un processus ponctuel N est un processus de Poisson (ou une mesure aléatoire de Poisson) de mesure μ si les deux conditions suivantes sont réalisées:

- i) pour $A \in \mathcal{E}$, $N(A)$ a une distribution de Poisson de paramètre $\mu(A)$;
- ii) si A_1, \dots, A_m sont des ensembles mutuellement disjoints alors $N(A_1), \dots, N(A_m)$ sont mutuellement indépendants.

La transformée de Laplace d'une mesure aléatoire de Poisson de mesure moyenne μ est donnée par

$$L_N(g) = \exp \left\{ - \int_E \left(1 - e^{-g(x)} \right) d\mu(x) \right\}.$$

Définition: Un processus ponctuel N est un processus Poisson composé de mesure μ et de taille des groupes π si les deux conditions suivantes sont réalisées:

- i) pour $A \in \mathcal{E}$, $N(A)$ a une distribution Poisson composée de paramètre $\mu(A)$ et de taille des groupes π ;
- ii) si A_1, \dots, A_m sont des ensembles mutuellement disjoints alors $N(A_1), \dots, N(A_m)$ sont mutuellement indépendants.

La transformée de Laplace d'une mesure aléatoire Poisson composée de mesure moyenne μ et de taille des groupes π est donnée par

$$L_N(g) = \exp \left\{ - \int_E (1 - L(g(x))) d\mu(x) \right\}$$

où L est la transformée de Laplace de π .

Définition: Soient N, N_1, N_2, \dots des processus ponctuels sur E . On dit que (N_n) converge faiblement vers N dans $M_p(E)$ si

$$\Pr((N_n(A_1), \dots, N_n(A_m))) \rightarrow \Pr((N(A_1), \dots, N(A_m)))$$

pour n'importe quel choix A_1, \dots, A_m satisfaisant $\Pr(N(\partial A_i) = 0) = 1$, $i = 1, \dots, m$ et $m \geq 1$.

Théorème: La suite de processus ponctuels (N_n) converge faiblement vers N dans $M_p(E)$ si et seulement si la suite des transformées de Laplace converge pour n'importe quelle fonction positive g à support compact

$$L_{N_n}(g) \rightarrow L_N(g).$$

Théorème: Soit $(u_n(\tau))$ une suite telle que $\lim_{n \rightarrow \infty} n\bar{F}(u_n(\tau)) = \tau$ avec $\bar{F} := 1 - F$. Alors $N_n^{(\tau)}$ converge faiblement vers un processus de Poisson ponctuel homogène N sur $(0, 1]$ d'intensité $\tau |\cdot|$, où $|\cdot|$ est la mesure de Lebesgue.

3. Estimations des paramètres des lois des extrêmes

Nous avons étudié les modèles probabilistes qui décrivent les comportements des extrêmes de variables aléatoires IID.

Dans cette section, nous présentons les méthodologies statistiques basées sur ces modèles et discutons des avantages et inconvénients de chacune de ces méthodologies d'estimation.

Le choix de l'utilisation d'une méthode plutôt que d'une autre dépendra surtout du type de données à disposition:

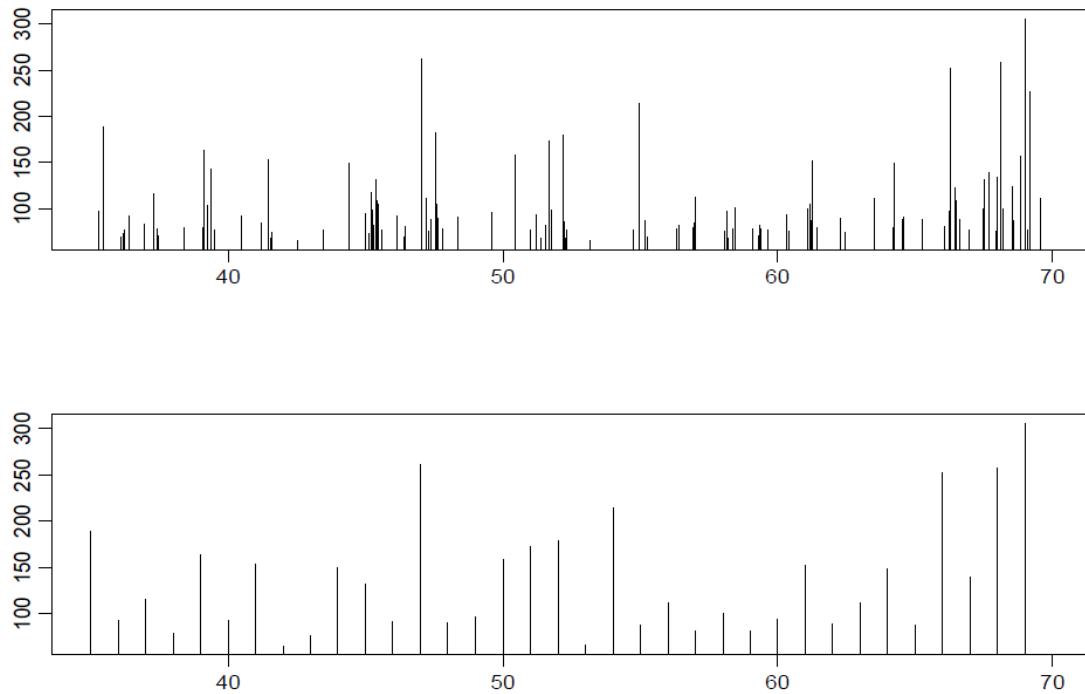
- maximums par blocs,
- données de seuils,
- records,
- plus hautes statistiques d'ordre,
- toutes les données.

Tout au long de l'analyse de données, il faut essayer de choisir des modèles statistiques qui sont compatibles avec le processus physique qui a généré les données.

Il est important d'utiliser des statistiques capables d'évaluer l'ajustement du modèle et ainsi faciliter la procédure de sélection du modèle.

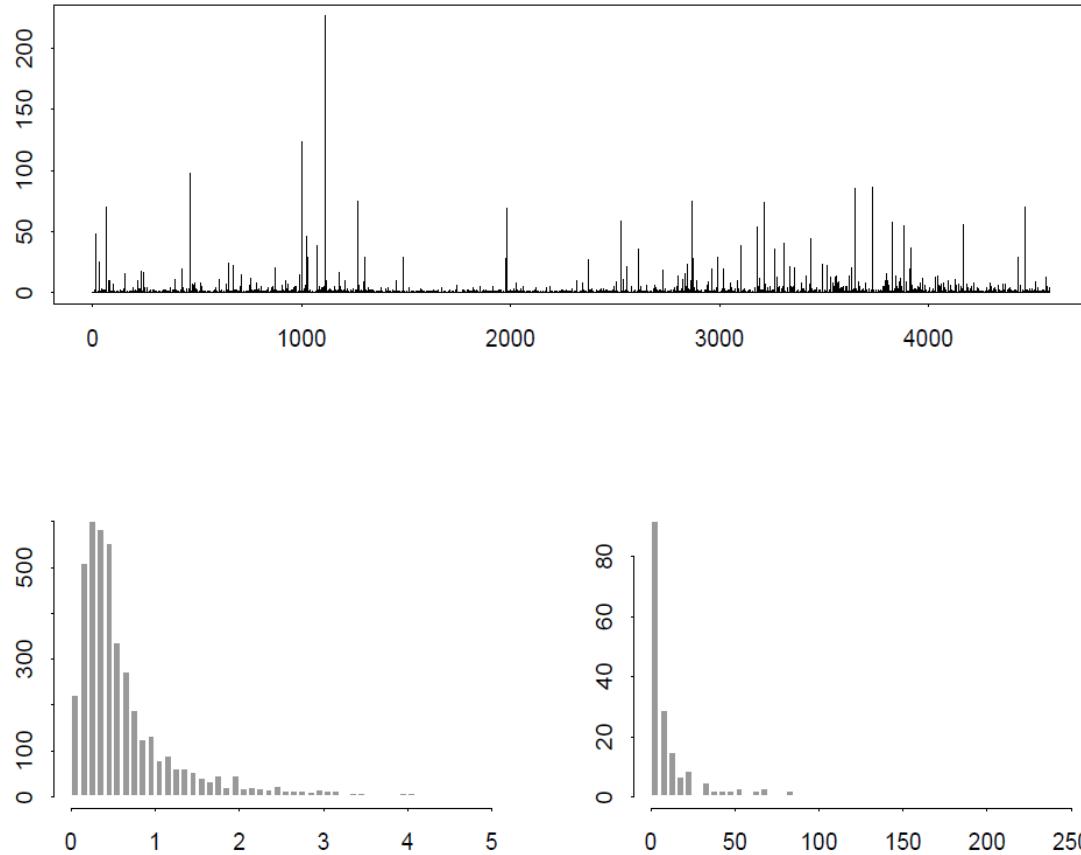
Le point de départ de toute analyse statistique des extrêmes est la condition que les distributions des valeurs extrêmes sont bien approchées par les formes limites théoriques. Il est donc nécessaire de savoir si l'approximation par la forme limite est suffisamment bonne.

Nous allons illustrer ces points avec plusieurs exemples.

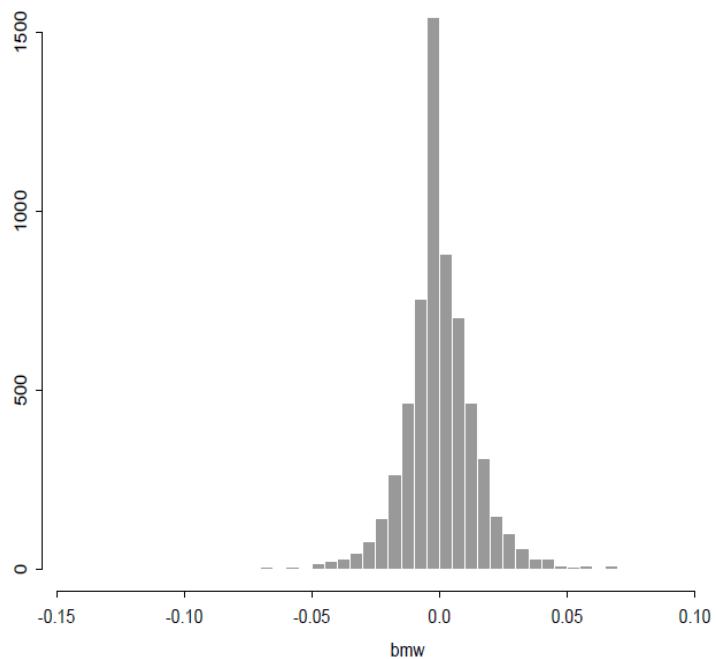
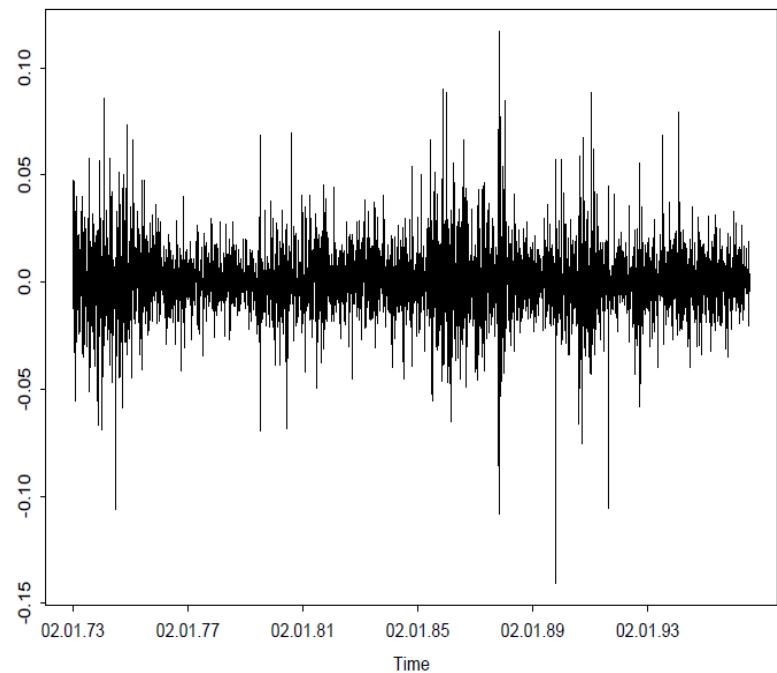


The river Nidd data 1934–1970 (top) and the corresponding annual maxima (bottom). The data are measured in CUMECS.

These data represent high river levels of the River Nidd in Yorkshire above a threshold value of 65.



These data represent claims from a fire insurance portfolio. The values are multiples of 1000 SFr. The corresponding histogram of the claims less than 5000 SFr (left) and of the remaining claims exceeding 5000 SFr (right).



Daily log returns of BMW share prices for the period January 2, 1973 - July 23, 1996 ($n=6146$)
together with a histogram of the data

▷ Analyse exploratoire des données

Il est important d'examiner les données avant de s'engager dans une analyse statistique détaillée. A notre époque où la puissance de calcul des ordinateurs est quasi-illimitée, cette exploration graphique est de plus en plus fondamentale.

- Probabilités et graphiques quantiles-quantiles (QQ plots)

Soit $X_{(1)} \geq \dots \geq X_{(n)}$ la statistique d'ordre associée à notre échantillon. Si F est continue, alors les variables aléatoires $U_i = F(X_i)$ sont IID et loi uniforme sur $(0, 1)$. De plus

$$(F(X_{(i)}))_{i=1,\dots,n} \stackrel{d}{=} (U_{(i)})_{i=1,\dots,n}$$

ou de manière équivalente

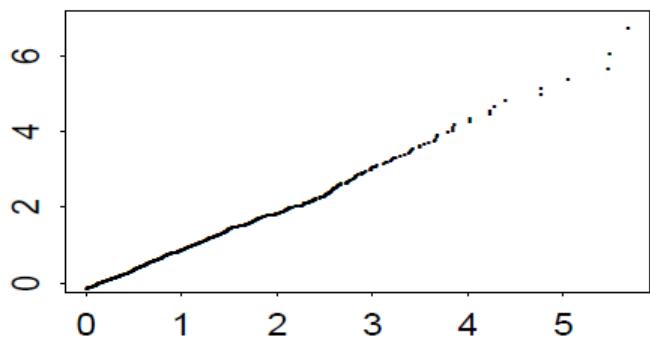
$$(X_{(i)})_{i=1,\dots,n} \stackrel{d}{=} (F^{-1}(U_{(i)}))_{i=1,\dots,n}.$$

Le graphique

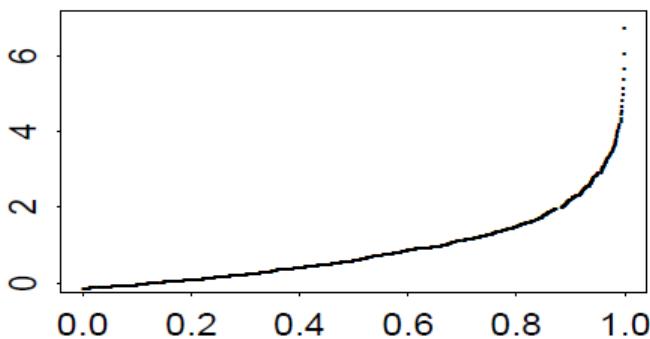
$$\left\{ X_{(i)}, F^{-1}(1 - i/n) : i = 1, \dots, n \right\}$$

est appelé graphique quantiles-quantiles ou QQ plot.

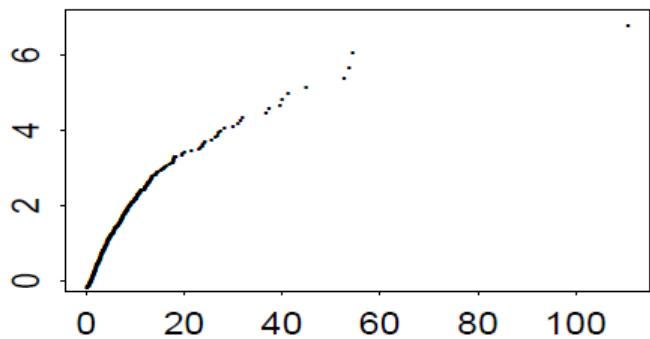
- Comparaison des distributions empirique et théorique: si les données ont été générées à partir d'un échantillon aléatoire de la distribution de référence, la courbe devrait être à peu près linéaire. Cela reste vrai si les données proviennent d'une transformation linéaire de la distribution (dans ce cas, il est possible de déterminer les coefficients de la transformation linéaire à partir de la pente et de l'ordonnée à l'origine de la droite observée).
- Queues de distribution : les différences dans les formes des queues de distribution peuvent être déduites de la courbe. Par exemple, si la distribution de référence a des queues plus épaisses (tendance à avoir des valeurs plus grandes pour une probabilité donnée), la courbe aura tendance à s'incurver vers le bas à gauche et/ou vers le haut à droite.



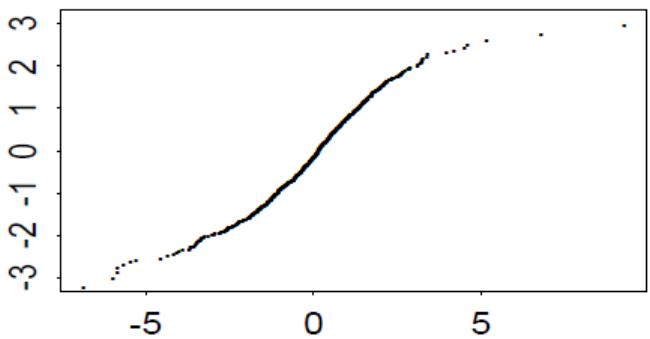
(a)



(b)

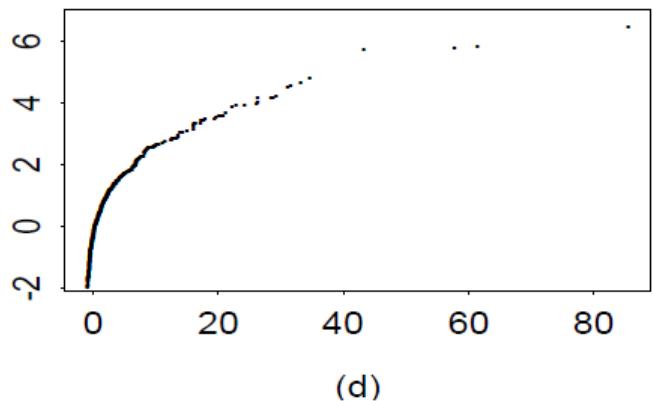
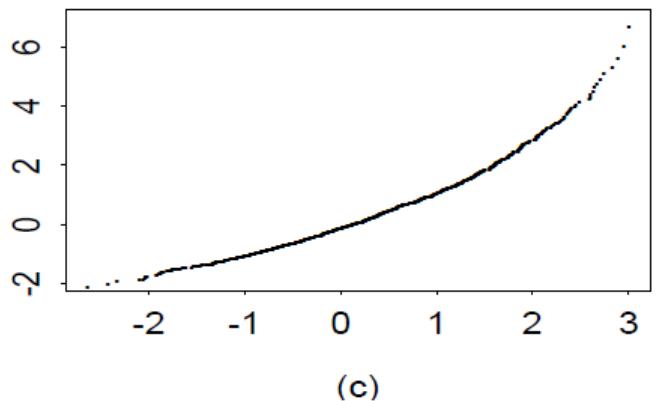
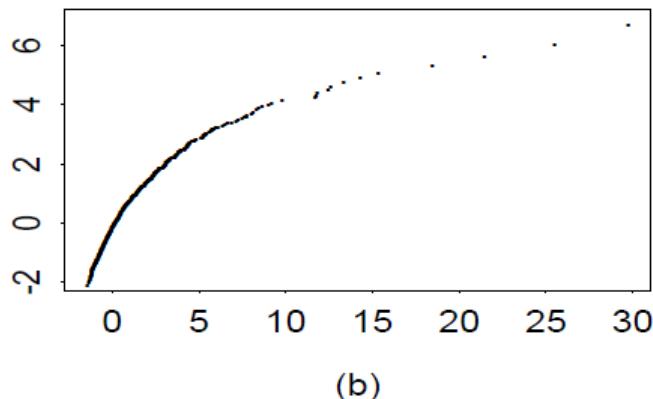
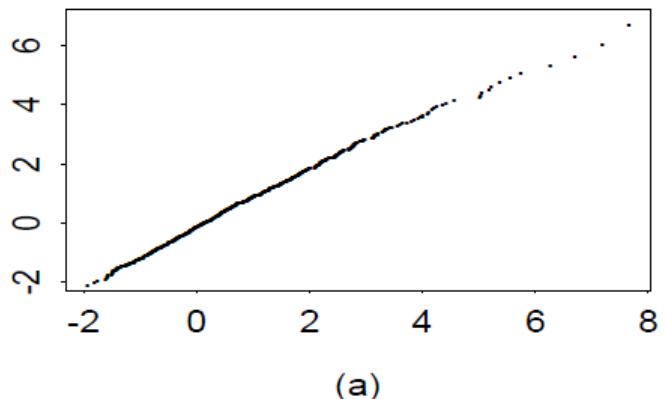


(c)



(d)

QQ-plot of exponentially (a), uniformly (b), lognormally (c) distributed simulated data versus the exponential distribution. In (d) a QQ-plot of t_4 -distributed data versus the standard normal distribution is given.



QQ-plots: (a) Gumbel distributed simulated data versus Gumbel distribution. GEV distributed data with parameters (b): $\xi = 0.3$, (c): $\xi = -0.3$, (d): $\xi = 0.7$, versus Gumbel.

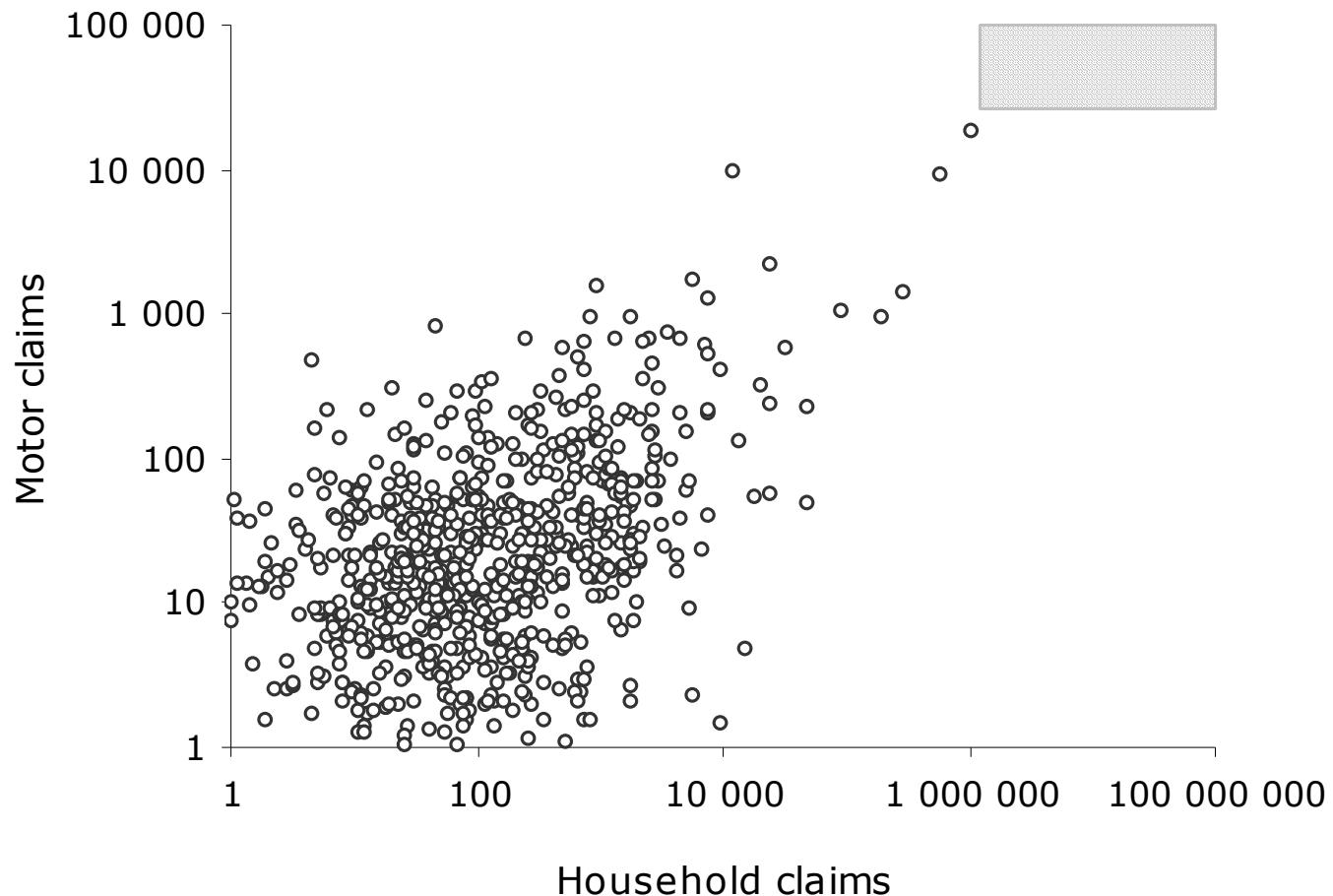
Exemples:

Distribution	QQ plot
Normal	$\{X_{(i)}, \Phi^{-1}(1 - i/n)\}$
Log-Normal	$\{\log X_{(i)}, \Phi^{-1}(1 - i/n)\}$
Exponentiel	$\{X_{(i)}, -\log(i/n)\}$
Pareto	$\{\log X_{(i)}, -\log(i/n)\}$
Weibull	$\{\log X_{(i)}, \log(-\log(i/n))\}$

Illustrons par l'analyse d'un exemple tiré de l'assurance tempête.

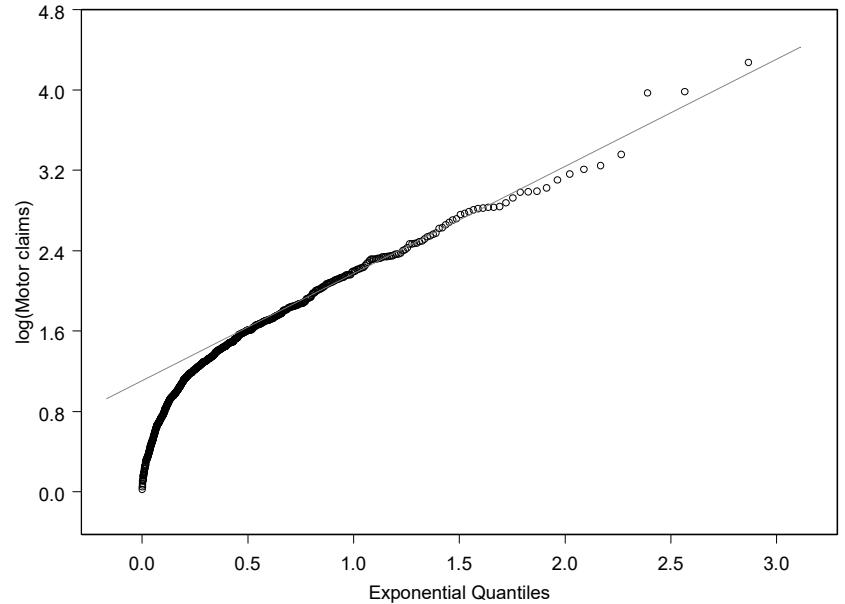
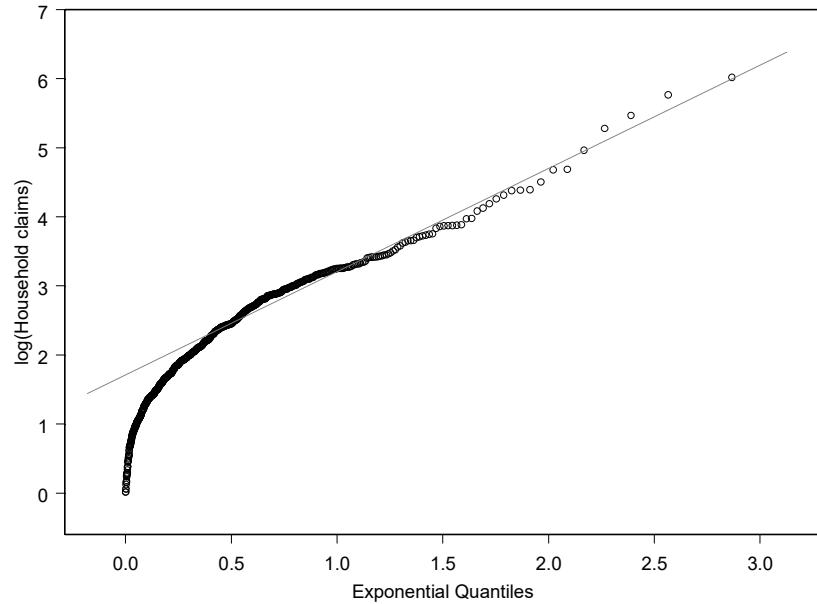
La figure suivante présente les montants des sinistres automobile et MRH ménages dans un portefeuille d'un assureur français. Les sinistres sont liés aux dommages causés par les vents violents, les tornades, les éclairs, les grêles, les pluies torrentielles et le poids de la neige. Les montants ont été construits à partir de 150 000 réclamations d'assurance habitation liées à la météo pour une période de onze ans, du 1er janvier 1990 au 31 décembre 2000. Une échelle de 72 heures et un seuil de vent maximal égal à 100 km / h ont été choisis pour obtenir 736 dégâts de tempête.

- Les deux valeurs aberrantes correspondent aux tempêtes Lothar et Martin sont passées sur la France en décembre 1999.
- Les données démontrent une tendance apparente à des montants de sinistres élevés dans les deux branches. Cela suggère que l'intensité des événements de tempête induit une dépendance dans la série entre les deux branches et peut être considérée comme un facteur commun.



Storm damages (both variables are on a logarithmic scale).

- Pareto-type distributions are graphically detected through Pareto-quantile plots
 $(-\log(1 - k/(n+1)), \log(X_{(k),l}))_{k=1,\dots,n}$



- La fonction de dépassement moyen (mean excess function) ou de durée de vie résiduelle

Un autre outil graphique important pour étudier les extrêmes est la fonction de dépassement moyen définie par

$$e(u) = \mathbb{E}(X - u | X > u).$$

Dans un contexte d'assurance, $e(u)$ peut aussi s'interpréter comme la durée de vie résiduelle d'un assuré au delà de l'âge u ou le coût moyen d'un excédant de sinistre de priorité u et de portée illimitée.

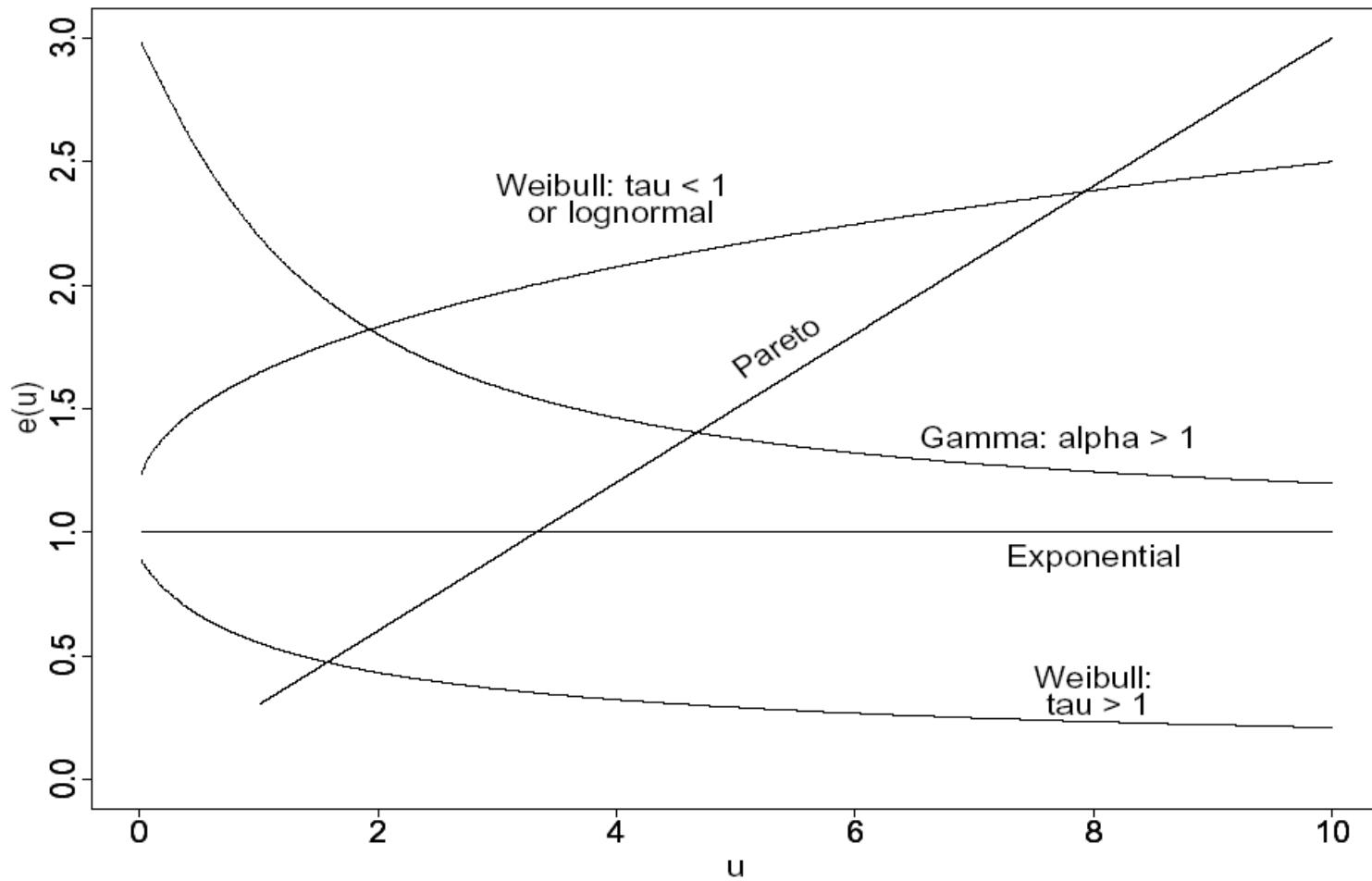
Soit $\Delta_n(u) = \{i : i = 1, \dots, n, X_i > u\}$. La fonction de dépassement moyen empirique est définie par

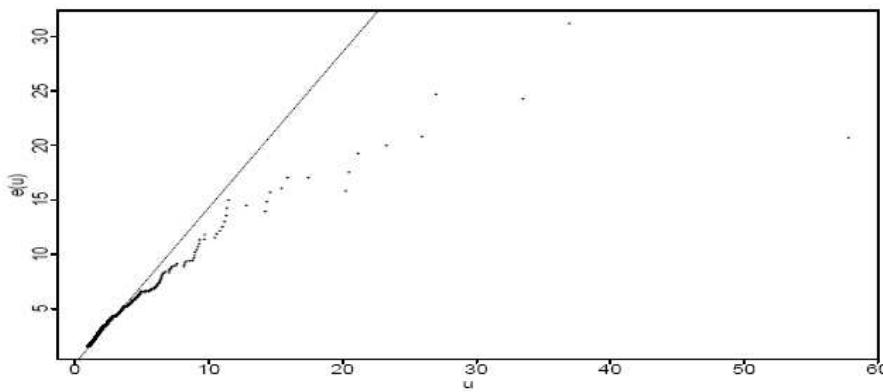
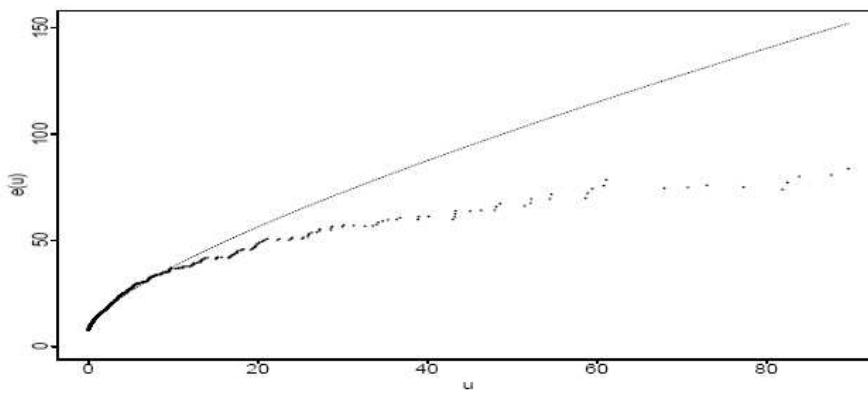
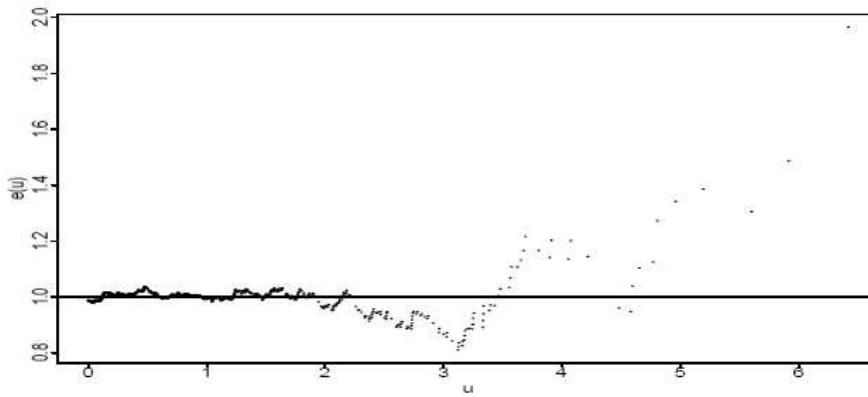
$$e_n(u) = \frac{\sum_{i \in \Delta_n(u)} (X_i - u)}{\text{card} \Delta_n(u)}.$$

Le graphique

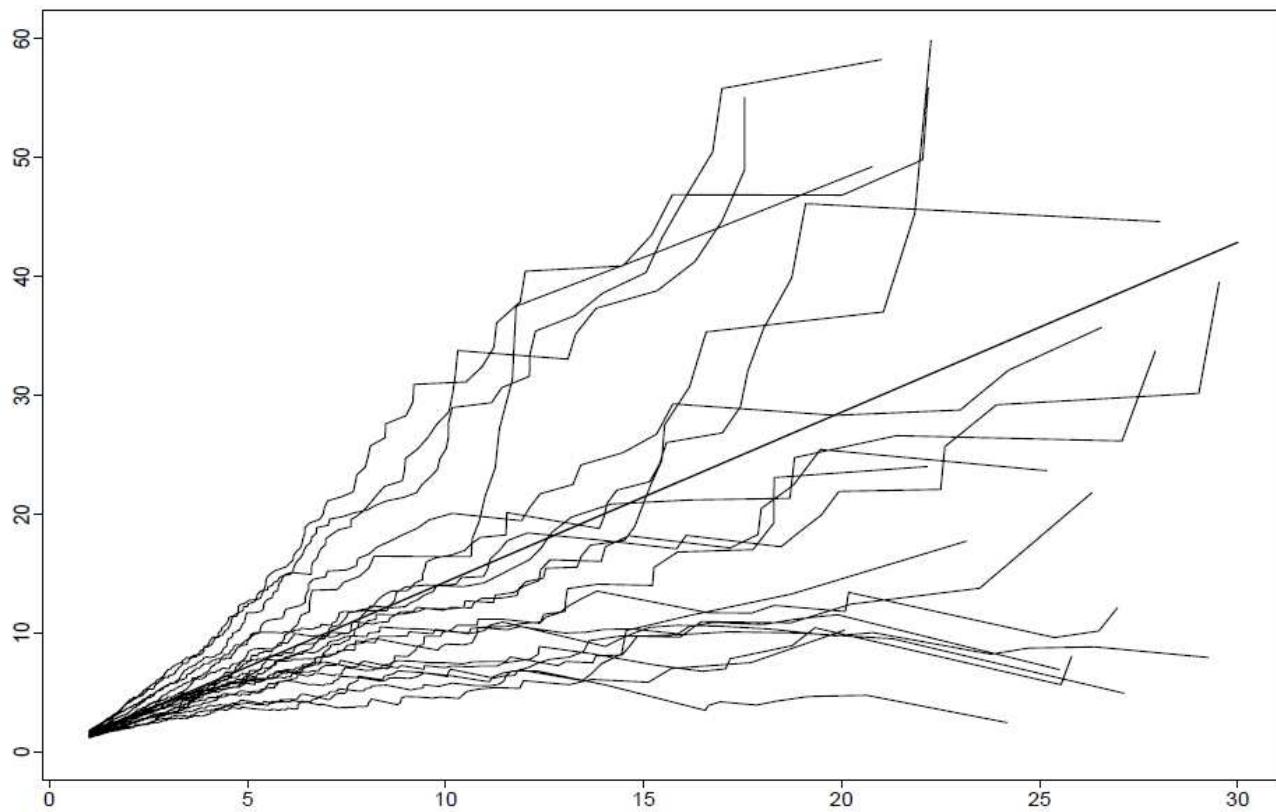
$$\{X_{(i)}, e_n(X_{(i)}) : i = 1, \dots, n\}$$

est appelé graphique de dépassement moyen (mean excess plot).

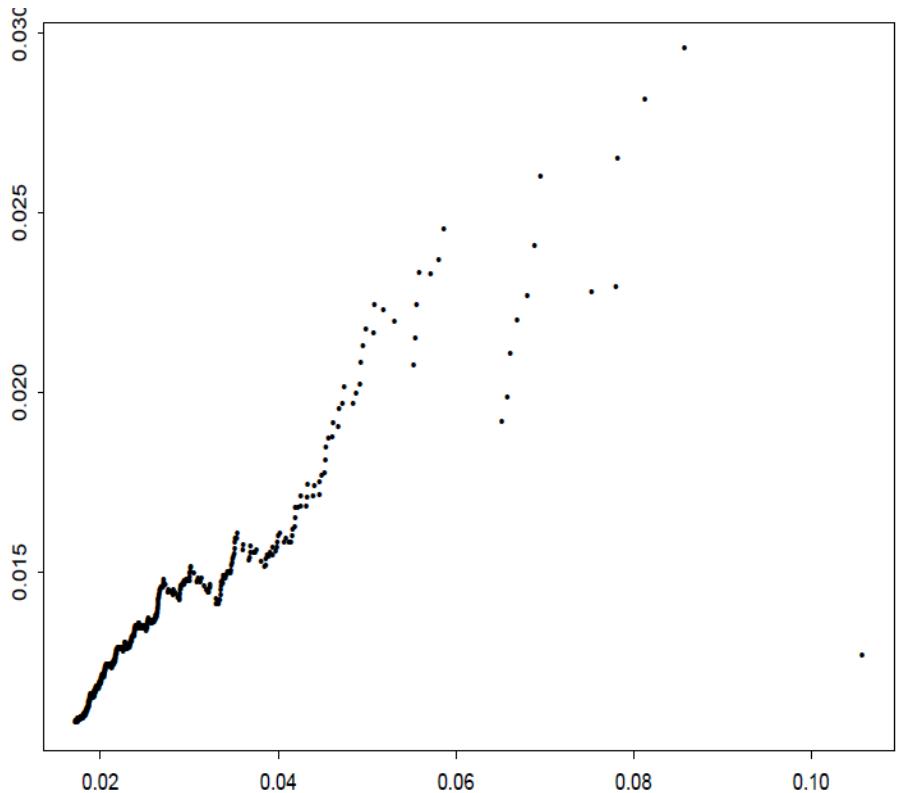
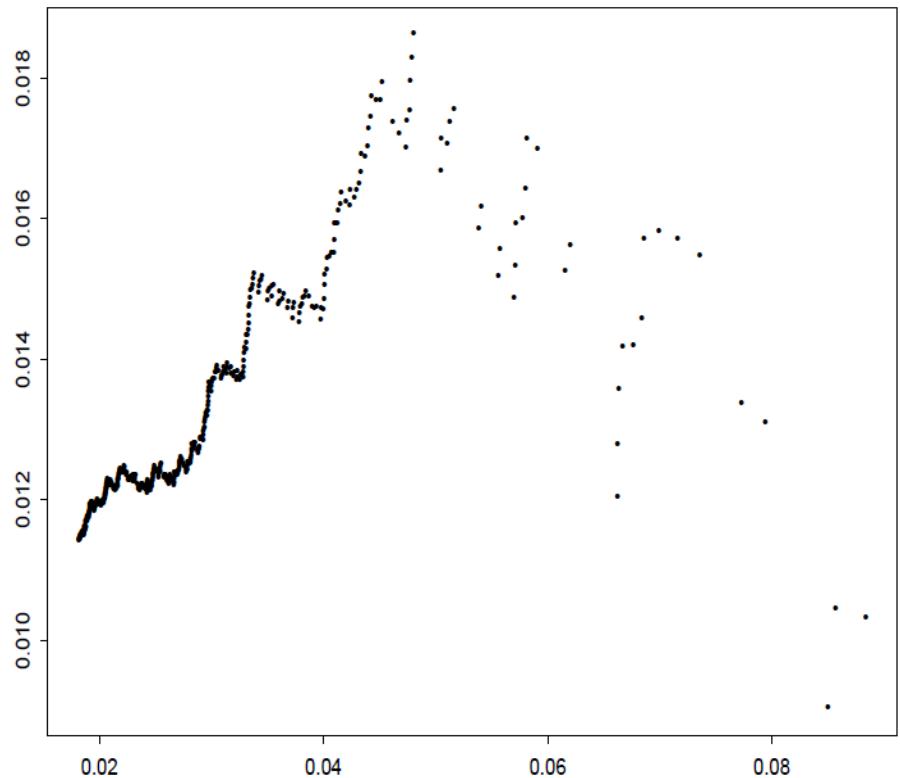




The empirical mean excess function $e_n(u)$ of simulated data ($n = 1000$) compared with the corresponding theoretical mean excess function $e(u)$ (dashed line): standard exponential (top), lognormal (middle) with $\ln X \sim N(0, 4)$, Pareto (bottom) with tail index 1.7.



The mean excess function of the Pareto distribution $\bar{F}(x) = x^{-1.7}$, $x \geq 1$, together with 20 empirical mean excess functions $e_n(u)$ each based on simulated data ($n = 1\,000$) from the above distribution.



Exploratory data analysis of BMW share prices.

▷ Estimation des paramètres de la distribution GEV

La distribution GEV est utilisée lorsque l'on dispose de données du type maximums par blocs.

- Choix du modèle probabiliste

Soit X_1, X_2, \dots une suite de variables aléatoires IID de distribution commune F , et $M_n = \max(X_1, \dots, X_n)$.

D'après la première partie, on peut faire l'approximation suivante

$$\Pr(M_n \leq x) \sim \text{GEV}(\mu, \sigma, \xi).$$

En effet, si F est dans le domaine d'attraction d'une loi limite, il existe deux suites (a_n) et (b_n) telles que

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = \exp\left(-(1 + \xi x)_+^{-1/\xi}\right).$$

Bien que les suites (a_n) et (b_n) dépendent de la distribution sous-jacente F , pour n fixé ce sont simplement des constantes de normalisation, et on peut donc écrire

$$\Pr(M_n \leq y) = \exp\left(-\left(1 + \xi \left(\frac{y - \mu}{\sigma}\right)\right)_+^{-1/\xi}\right)$$

avec $y = a_n x + b_n$, $\sigma = a_n > 0$ et $\mu = b_n$.

L'hypothèse importante ici est que la loi limite est adaptée pour modéliser la loi à distance finie. La validité de cette hypothèse repose sur:

- le choix de n : il est important que les blocs qui ont été utilisés pour calculer les maximums soient de taille suffisamment grande;
- le fait que h' est proche d'une constante: ceci dépend de la distribution F .

- Estimation par maximum de vraisemblance des paramètres de la distribution GEV

La maximisation de la vraisemblance sous l'hypothèse d'une loi GEV de paramètre $\theta = (\mu, \sigma, \xi)'$ donne l'estimateur du maximum de vraisemblance pour l'ensemble des distributions de la classe des distributions GEV. On notera cependant que:

- une maximisation numérique est nécessaire puisqu'il n'y a pas de formule analytique;
- il est nécessaire de mettre des contraintes sur les paramètres dans certains cas car le support de la distribution peut dépendre du paramètre θ .

Des difficultés théoriques et pratiques surviennent lorsque les conditions de régularité usuelles de l'estimateur du maximum de vraisemblance ne sont pas satisfaites.

En particulier lorsque la distribution a un point extrémal fini, les résultats classiques ne sont pas toujours obtenus.

Smith (1985) a donné les résultats suivants:

- si $\xi > -1/2$, l'estimateur du maximum de vraisemblance est régulier (les lois limites sont les lois usuelles);
- si $-1 < \xi < -1/2$, l'estimateur est super-efficace;
- si $\xi < -1$, l'estimateur du point extrémal est donné par la plus grande des observations.

Quand $\xi < -1/2$, les distributions sont à support borné avec des queues de distribution très fines. Ceci est rarement observé en pratique et donc les limites théoriques de l'estimateur du maximum de vraisemblance conduisent rarement à des difficultés d'estimation.

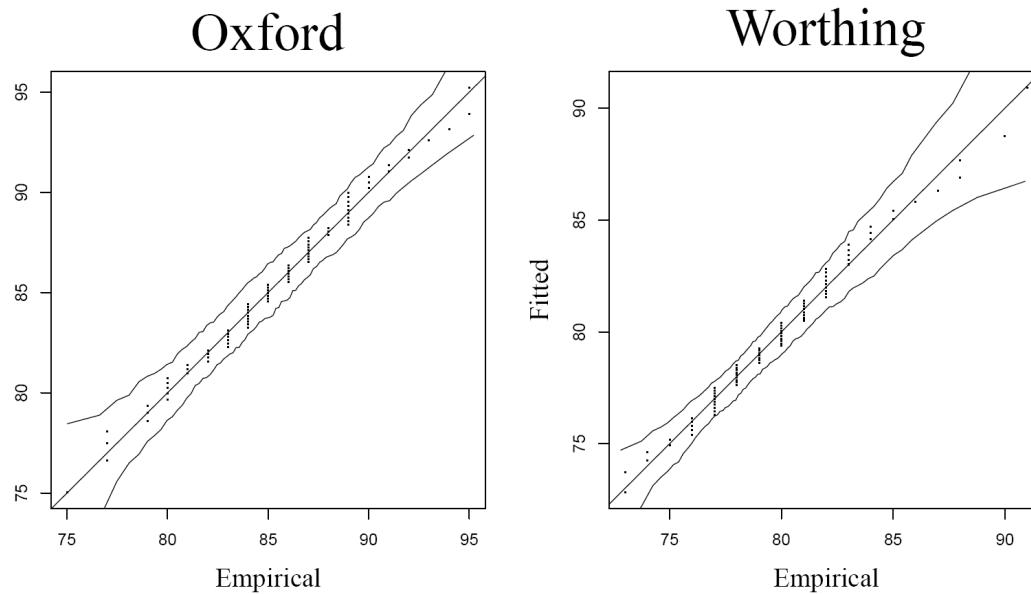
Quand $\xi > -1/2$,

$$\hat{\theta} \sim MVN \left\{ \theta, \mathcal{I}^{-1} (\hat{\theta}) \right\}$$

où $\mathcal{I}^{-1}(\hat{\theta})$ est l'inverse de la matrice d'information de Fisher évaluée dans l'estimateur du maximum de vraisemblance. $\mathcal{I}^{-1}(\hat{\theta})$ peut être calculé analytiquement, mais il est plus facile de recourir à des techniques numériques d'approximation par différence.

Les intervalles de confiance pour les paramètres et des transformations de ces paramètres se déduisent à partir de l'approximation par la loi Normale de l'estimateur du maximum de vraisemblance.

- Exemple: Estimation de la distribution GEV pour des maximums de températures (QQ plots)



La distribution GEV semble appropriée pour modéliser les maximums annuels des températures à Oxford et Worthing: les courbes des QQ-plot sont très proches de la première bissectrice.

Les valeurs des paramètres estimés sont données dans le tableau suivant:

	Oxford	Worthing
$\hat{\mu}$	83.8 (0.52)	78.5 (0.39)
$\hat{\sigma}$	4.3 (0.36)	3.1 (0.27)
$\hat{\xi}$	-0.29 (0.07)	-0.11 (0.07)
Corr($\hat{\theta}$)	$\begin{pmatrix} 1.00 & 0.00 & -0.37 \\ 0.00 & 1.00 & -0.57 \\ -0.37 & -0.57 & 1.00 \end{pmatrix}$	$\begin{pmatrix} 1.00 & 0.24 & -0.35 \\ 0.24 & 1.00 & -0.38 \\ -0.35 & -0.38 & 1.00 \end{pmatrix}$

Des intervalles de confiance à 95% pour le paramètre de forme ξ basés sur l'approximation Normale sont:

- Oxford: (-0.43, -0.15)

- Worthing: (-0.25, 0.03)

- Niveaux et périodes de retour

La question pratique qui est souvent posée dans la plupart des applications est la suivante: quelle est la probabilité pour un processus donné de dépasser un niveau x dans une période de temps future?

Ou, de manière équivalente, quel est le niveau x tel que cette probabilité soit suffisamment petite?

Une des façons de répondre à la question est de considérer les niveaux de retour et les périodes de retour:

- *Période de retour pour un niveau x :* le temps d'attente moyen pour que le niveau x soit à nouveau dépassé;

- *Niveau de retour de période T :* le niveau pour lequel le temps d'attente moyen de dépassement est de T années.

Dans les applications, T est souvent égal à 100 années alors que les données ne couvrent qu'une période de 30 à 50 ans. Pour des données IID, les niveaux de retour

et les périodes de retour correspondent aux quantiles et aux inverses des probabilités de dépassement respectivement. Si $1 - F(x_p) = p$, alors le niveau de retour x_p a une période de retour de p^{-1} observations.

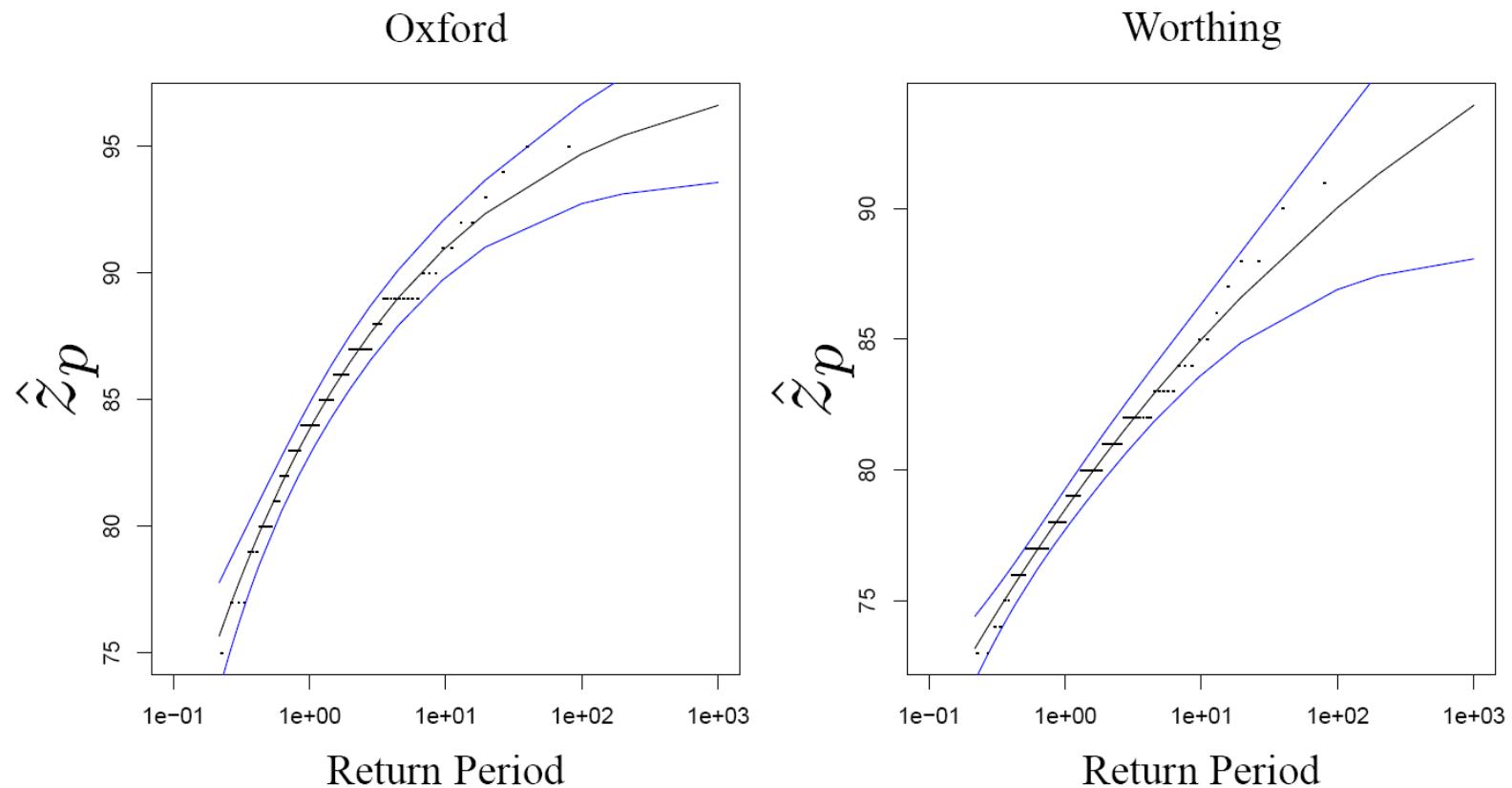
- Estimation du niveau de retour pour la distribution GEV

Le niveau de retour de période p^{-1} est le quantile d'ordre $1 - p$ de la distribution GEV pour $0 < p < 1$.

En substituant les valeurs estimées par maximum de vraisemblance aux paramètres théoriques dans l'expression des quantiles de la distribution GEV pour un retour de période p^{-1} , on obtient:

$$\hat{z}_p = \begin{cases} \hat{\mu} - \hat{\sigma} \hat{\xi}^{-1} [1 - (-\log(1-p))^{-\hat{\xi}}] & \text{pour } \hat{\xi} \neq 0, \\ \hat{\mu} - \hat{\sigma} \log(-\log(1-p)) & \text{pour } \hat{\xi} = 0. \end{cases}$$

La méthode delta permet d'obtenir la variance asymptotique de l'estimateur $Var(\hat{z}_p) = \nabla z'_p V \nabla z_p$ où V est la matrice de variance-covariance de $\hat{\theta}$ et ∇z_p est le vecteur des dérivés de z_p par rapport à μ , σ et ξ respectivement, évalué pour $\theta = \hat{\theta}$.



Cet estimateur de la variance peut être utilisé pour construire des intervalles de confiance basés sur l'approximation à l'aide de la loi Normale.

Ces intervalles de confiance sont symétriques par construction.

- Les graphiques précédents donnent \hat{z}_p en fonction de la période de retour avec des intervalles de confiance à 95%.
- Les niveaux de retour empiriques sont donnés également pour vérifier la qualité du modèle.
- Il y a 80 observations pour chaque échantillon, donc la valeur la plus grande donne une idée du niveau de retour à 80 ans.
- Le fait que le graphique obtenu pour Worthing est moins courbé est dû à une valeur plus forte du paramètre de forme, ξ .
- Si ξ avait été égal à 0 alors la courbe des niveaux de retour aurait été la fonction logarithme.

"The big advantage of maximum likelihood procedures is that they can be generalized, with very little change in the basic methodology, to much more complicated models in which trends or other effects may be present" - R. Smith

- La méthode des moments pondérés

Cette méthode consiste à égaliser moments théoriques de la distribution GEV avec les moments empiriques construits à partir des données.

"This method is simple to apply and performs well in simulation studies. However, until there is some convincing theoretical explanation of its properties, it is unlikely to be universally accepted. There is also the disadvantage that, at present at least, it does not extend to more complicated situations such as regression models based on extreme value distribution" - EKM

▷ Estimation du paramètre de forme ξ à partir de l'ensemble des données

L'idée de base de la construction des estimateurs consiste à trouver des conditions équivalentes d'appartenance à un domaine d'attraction et qui s'expriment à partir du paramètre de forme.

Pour $\xi \in \mathbb{R}$, les estimateurs les plus connus sont: l'estimateur de Pickands et l'estimateur de Dekkers, Einmahl et de Haan

$$\begin{aligned}\xi_{k,n}^{Pickands} &= \left[\ln \frac{X_{(k)} - X_{(2k)}}{X_{(2k)} - X_{(4k)}} \right] / \ln 2, \\ \xi_{k,n}^{DEdH} &= \xi_{k,n}^{H(1)} + 1 - \frac{1}{2} \left[1 - \frac{\left(\xi_{(k)}^{H(1)} \right)^2}{\xi_{(k)}^{H(2)}} \right]^{-1},\end{aligned}$$

$$\text{with } \xi_{k,n}^{H(r)} = \frac{1}{k} \sum_{i=1}^{k-1} \left[\ln X_{(i)} - \ln X_{(k)} \right]^r, \quad r = 1, 2, \dots$$

Pour $\xi > 0$, l'estimateur le plus connu est celui de Hill

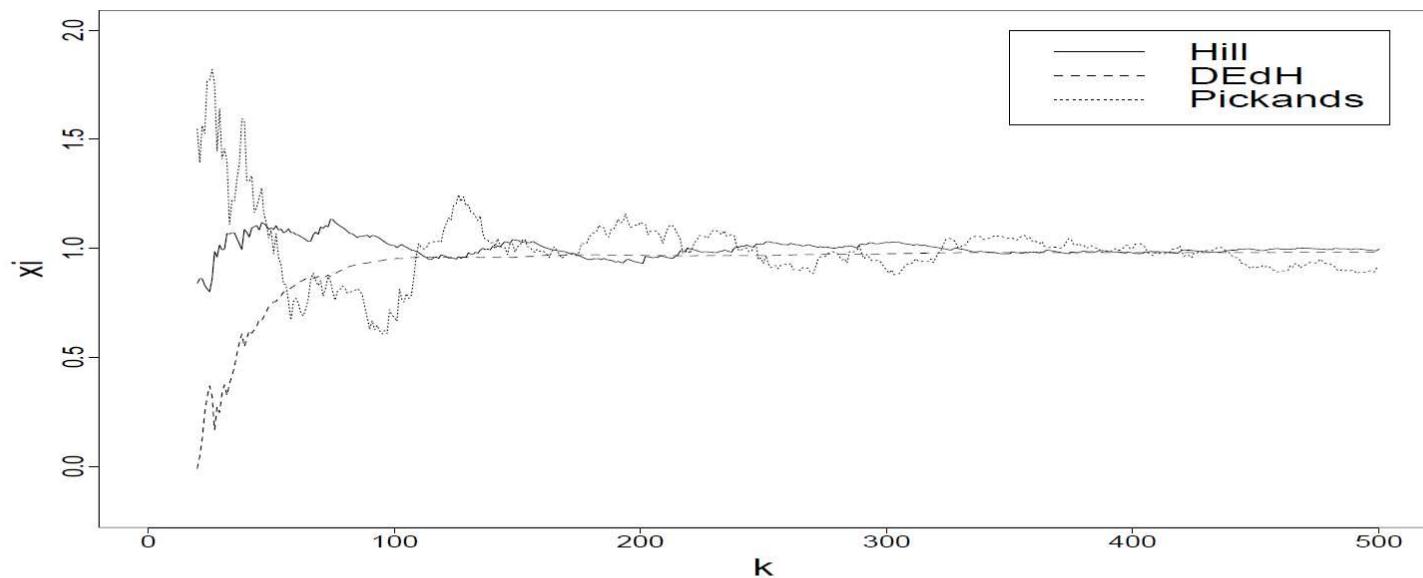
$$\xi_{k,n}^{Hill} = \frac{1}{k} \sum_{i=1}^k \ln X_{(i)} - \ln X_{(k+1)},$$

Si $k \rightarrow \infty$, $k/n \rightarrow 0$ as $n \rightarrow \infty$, les estimateurs sont faiblement convergents. Sous des hypothèses supplémentaires sur k et F , ces estimateurs sont aussi asymptotiquement Gaussiens

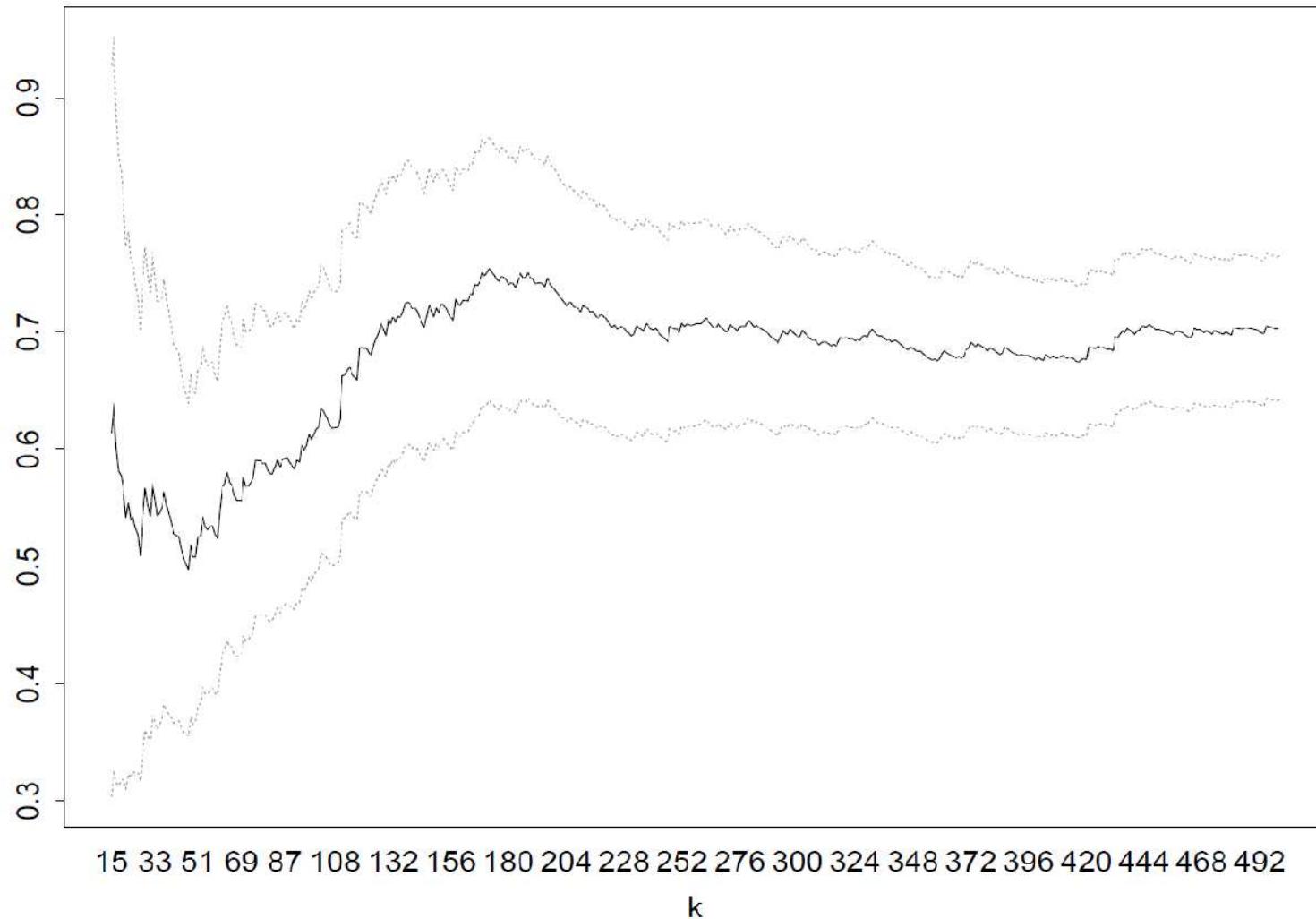
$$k^{1/2} (\xi_{k,n}^P - \xi) \xrightarrow{L} N \left(0, \xi^2 \frac{(2^{2\xi+1} + 1)}{(2(2^\xi - 1) \ln 2)^2} \right),$$

$$k^{1/2} (\xi_{k,n}^{DEdH} - \xi) \xrightarrow{L} N (0, 1 + \xi^2), \quad \xi \geq 0,$$

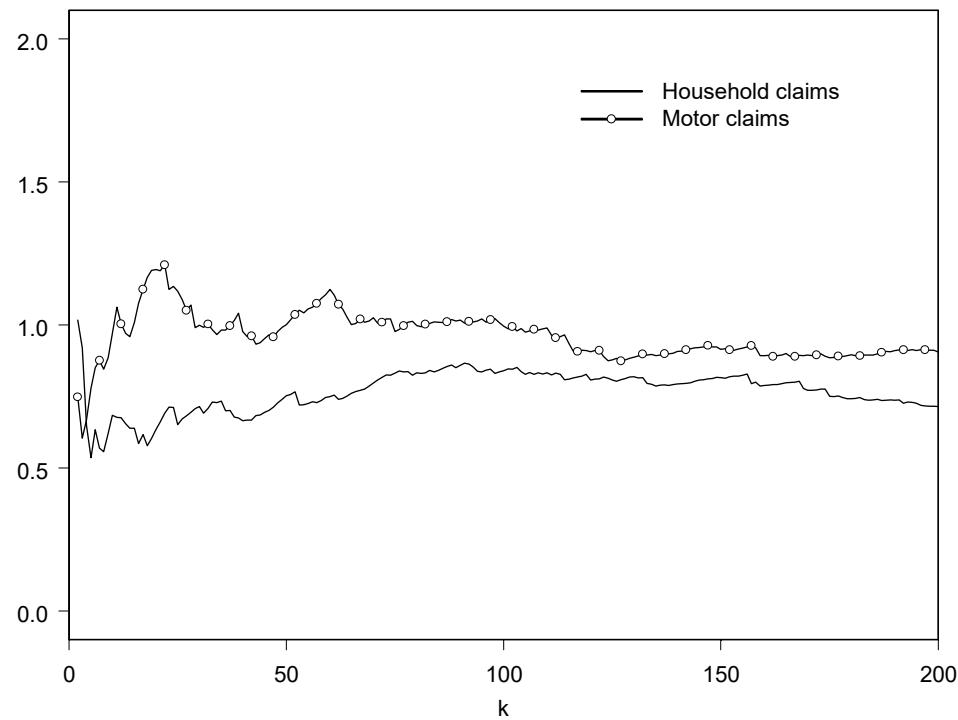
$$k^{1/2} (\xi_{k,n}^H - \xi) \xrightarrow{L} N (0, \xi^2).$$



Pickands-, Hill- and DEdH-plots for 2 000 simulated iid data with df given by $\bar{F}(x) = x^{-1}$, $x \geq 1$.



Hill estimator for the Danish insurance claims



Hill estimators for Household claim amounts and Motor claim amounts

▷ Estimation des paramètres de la loi de Pareto généralisée - POT method

La distribution de Pareto généralisée (Generalised Pareto distribution - GPD) est utilisée pour modéliser les observations des dépassements au delà d'un seuil.

Nous supposons que les dépassements $Y_{u,1}, \dots, Y_{u,n}$ au delà du seuil u suivent une loi $\text{GPD}(\sigma_u, \xi)$ telle que

$$\Pr(Y_u < y | Y_u > 0) = 1 - \left[1 + \xi \left(\frac{y}{\sigma_u} \right) \right]_+^{-1/\xi}.$$

Dans la pratique, l'étape la plus sensible est celle du choix du seuil u pour définir les dépassements:

- utiliser un seuil peu élevé permet d'augmenter le nombre de données, mais l'approximation par la loi Pareto généralisée est mauvaise;
- utiliser un seuil trop élevé limite le nombre de données et les estimateurs sont moins précis.

Un choix du seuil est donc important et doit trouver un équilibre entre ces deux extrêmes. Nous pouvons utiliser les propriétés de la loi GPD pour déterminer un seuil.

- Propriétés des lois GPD

- Stabilité par seuil

Si Y_u satisfait

$$Y_u | Y_u > 0 \sim \text{GPD}(\sigma_u, \xi) \text{ pour un seuil } u$$

alors pour tout seuil $v \geq u$

$$Y_v | Y_v > 0 \sim \text{GPD}(\sigma_u + \xi(v - u), \xi).$$

Notons que le paramètre de forme ξ ne dépend pas du seuil et que le nouveau paramètre d'échelle satisfait $\sigma_v = \sigma_u + \xi(v - u)$.

- La fonction de dépassement moyen

Si $X = Y_u + u$ où $Y_u|Y_u > 0 \sim \text{GPD}(\sigma_u, \xi)$, alors la fonction de dépassement au delà de $v > u$ est

$$E(X - v|X > v) = \frac{\sigma_u + \xi(v - u)}{1 - \xi}.$$

$E(X - v|X > v)$ est linéaire en v avec une pente $\xi/(1 - \xi)$.

- Procédure de sélection d'un seuil

- Le graphique de dépassement moyen.

On trace la fonction de dépassement moyen des excés pour plusieurs seuils.

Si la variable aléatoire suit une loi de Pareto Généralisée pour le seuil u , alors le graphique doit être approximativement linéaire au delà de ce seuil.

- Graphique de stabilité du paramètre d'échelle

Si l'on définit:

$$\sigma^* = \sigma_v - \xi v.$$

alors σ^* ne dépend plus de $v \geq u$ si la variable aléatoire suit une loi de Pareto Généralisée pour le seuil u .

Les paramètres estimés $\hat{\sigma}^*$ et $\hat{\xi}$ sont alors représentés sur un graphique pour différents seuils.

Si la variable aléatoire suit une loi de Pareto Généralisée pour le seuil u , alors les paramètres (σ^*, ξ) seront approximativement constants pour des seuils v plus grands.

Le seuil à choisir est alors défini comme le seuil minimal pour lequel la propriété de stabilité est satisfaite.

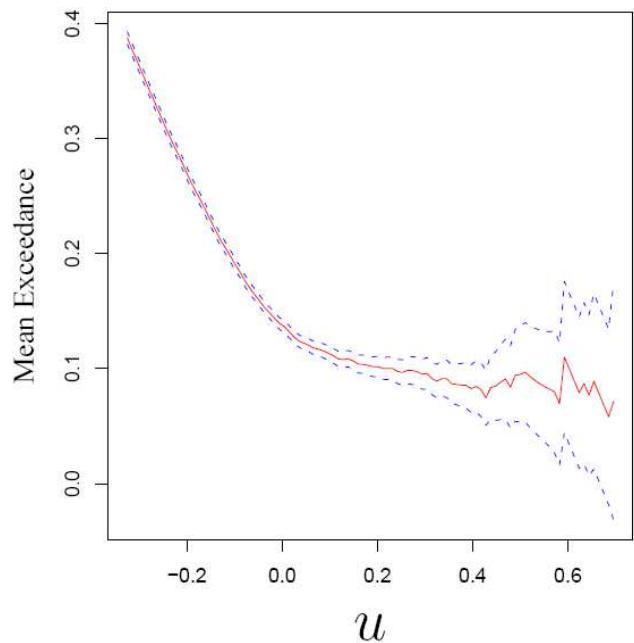
Il est intéressant également de représenter les intervalles de confiance des paramètres pour se rendre compte de l'incertitude qu'il y a sur les paramètres.

- Exemple: Estimation d'une loi GPD pour des données maritimes

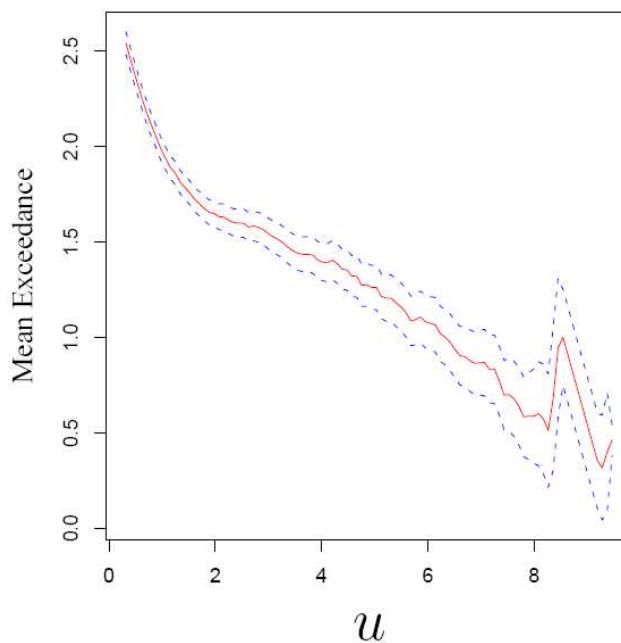
Nous disposons de données relatives à la hauteur de la mer (surge) et à la taille des vagues (wave) à Newlyn (Royaume-Uni).

Pour estimer les paramètres de la loi GPD, nous recherchons d'abord le seuil le plus approprié.

Surge

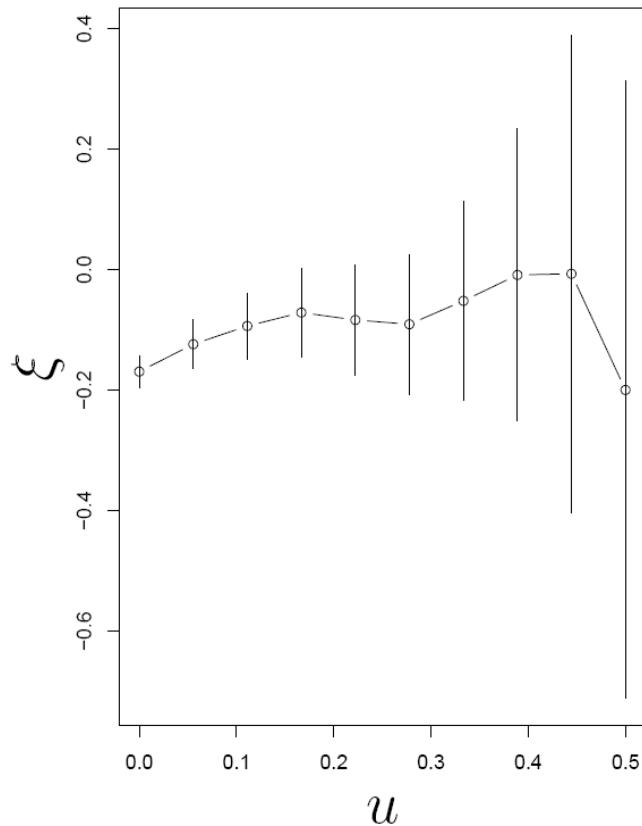
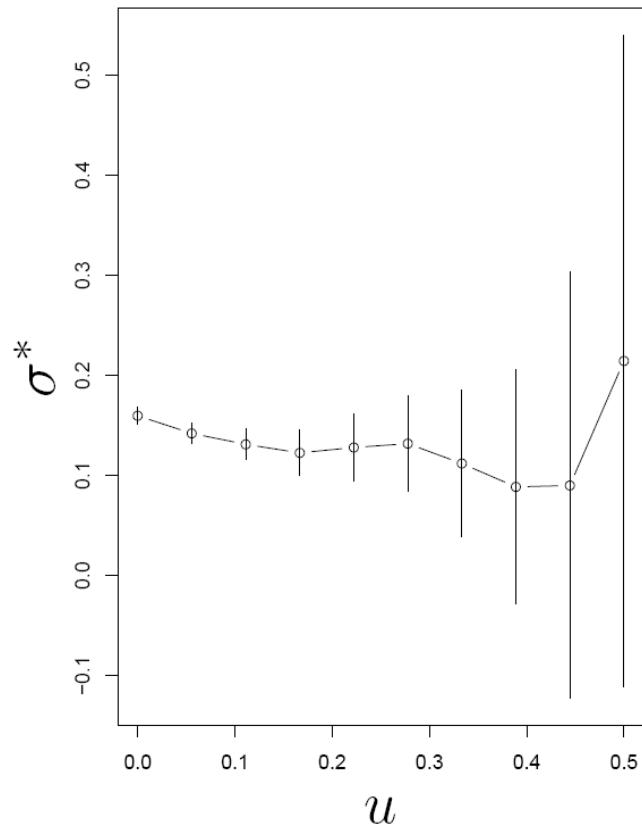


Wave

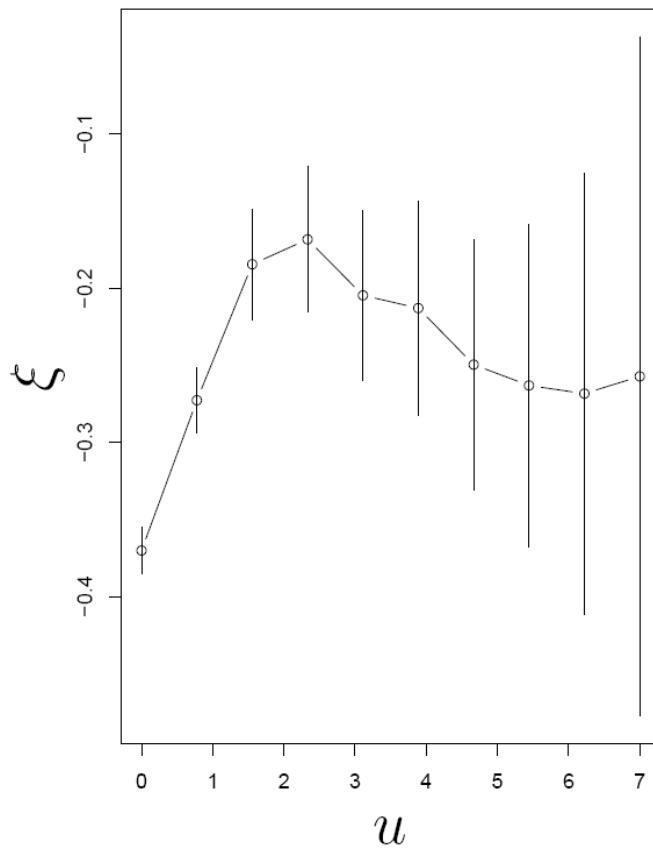
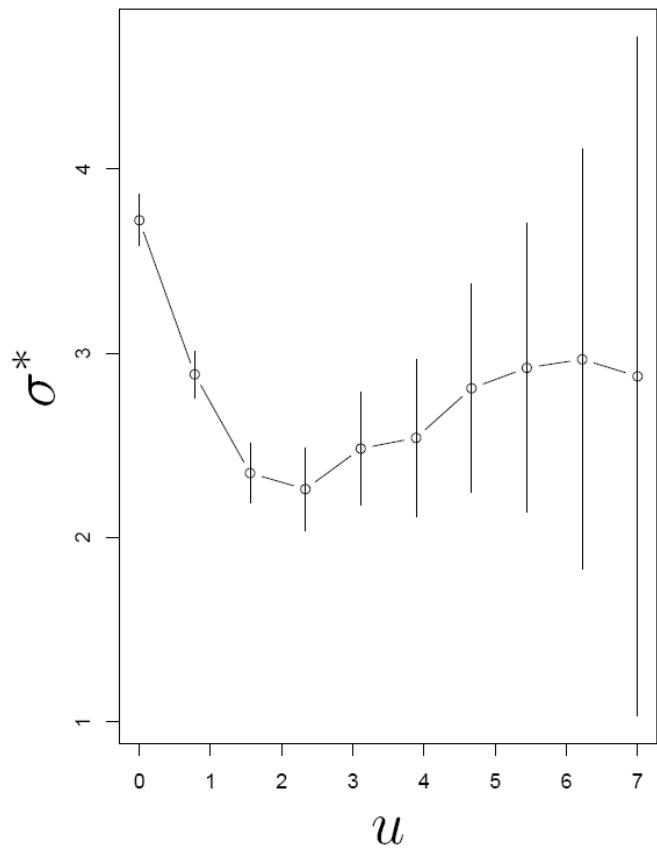


L'analyse des graphiques de la fonction de dépassement moyen suggère un seuil supérieur ou égal à 0.1 pour la hauteur de mer (surge) et de 2 pour la hauteur des vagues.

Pour la hauteur de la mer: Les graphiques de stabilité suggèrent des seuils aux alentours de 0.1 – 0.15.



Pour la hauteur des vagues: Les graphiques de stabilité suggèrent des seuils aux alentours de 4 – 5.



Nous sélectionnons les seuils de 0.1 et de 4 la hauteur de la mer et la hauteur des vagues respectivement.

Cela correspond aux quantiles d'ordre 0.64 et 0.65 pour ces variables.

	Surge	Wave
$\hat{\sigma}$	0.12 (0.005)	1.91 (0.061)
$\hat{\xi}$	-0.10 (0.026)	-0.165 (0.022)
Corr($\hat{\sigma}, \hat{\xi}$)	-0.73	-0.80

Valeurs des paramètres estimés des distributions GPD

Ceci laisse donc à penser que les distributions ont des points extrêmaux finis.