Théorie des valeurs extrêmes - Questions de cours

Question 1 :

$B \sim \mathcal{B}(m,p)$ $\qquad \mathbb{P}(B=k) = \binom{m}{k} p^k (1-p)^{m-k}$

$N \sim \mathcal{P}(\lambda = mp)$ $\qquad \mathbb{P}(N=k) = \dfrac{e^{-\lambda}\lambda^k}{k!}$

Montrer que $\mathbb{P}(B=k) = \mathbb{P}(N=k)\left(\dfrac{(1-\frac{\lambda}{m})^m}{e^{-\lambda}} \dfrac{1}{(1-\frac{\lambda}{m})^k} \prod_{i=1}^{k}(1-\frac{i-1}{m})\right)$

En déduire $B \xrightarrow[\substack{m\to\infty \\ p\to 0}]{\mathcal{L}} N$

Expliquer comment ce résultat peut être utilisé pour caractériser la loi du max et des statistiques d'ordre élevé.

Notons $A(m,p) = \dfrac{(1-\frac{\lambda}{m})^m}{e^{-\lambda}} \times \dfrac{1}{(1-\frac{\lambda}{m})^k} \times \prod_{i=1}^{k}(1-\frac{i-1}{m})$

$$\Rightarrow \mathbb{P}(N=k)\,A(m,p) = \frac{e^{-\lambda}\lambda^k}{k!} \times \frac{(1-\frac{\lambda}{m})^m}{e^{-\lambda}} \times \frac{1}{(1-\frac{\lambda}{m})^k} \prod_{i=1}^{k}\left(1-\frac{i-1}{m}\right)$$

$$= \frac{1}{k!}\lambda^k \left(1-\frac{\lambda}{m}\right)^{m-k} \prod_{i=1}^{k}\left(\frac{m-i+1}{m}\right)$$

$$= \frac{1}{k!}\left(\prod_{i=1}^{k}(m-i+1)\right)\left(\frac{\lambda}{m}\right)^k \left(1-\frac{\lambda}{m}\right)^{m-k}$$

Or $\binom{m}{k} = \dfrac{m!}{k!(m-k)!} = \dfrac{1}{k!} \times \dfrac{\prod_{i=1}^{m} i}{\prod_{i=1}^{m-k} i} = \dfrac{1}{k!}\prod_{i=m-k+1}^{m} i = \dfrac{1}{k!}\prod_{i=1}^{k}(m-i+1)$

$$\Rightarrow \mathbb{P}(N=k)\,A(m,p) = \binom{m}{k}\left(\frac{\lambda}{m}\right)^k\left(1-\frac{\lambda}{m}\right)^{m-k}$$

Comme $\lambda = mp$,

$$\Rightarrow \mathbb{P}(N=k)\,A(m,p) = \binom{m}{k} p^k (1-p)^{m-k}$$
$$= \mathbb{P}(B=k)$$

$$A(m,p) = \frac{(1-p)^m}{e^{-mp}} \frac{1}{(1-p)^k} \prod_{i=1}^{k}\left(1-\frac{i-1}{m}\right) \xrightarrow[\substack{m\to\infty \\ p\to 0}]{} 1$$

$$\mathbb{P}(B=k) = \mathbb{P}(N=k)\,A(m,p) \xrightarrow[\substack{m\to\infty \\ p\to 0}]{} \mathbb{P}(N=k)$$

Si on pose $B_m(x) = \sum_{i=1}^{m} \mathbb{1}_{\{X_i > x\}}$

Alors $B_m(x) \sim \mathcal{B}(m, p = \mathbb{P}(X_i > x)) \xrightarrow[m\to\infty]{\mathcal{L}} \mathcal{P}(mp)$

$X_{(1)} \geq X_{(2)} \geq \dots \geq X_{(m)}$

$$\mathbb{P}(X_{(k)} \leq x) = 1 - \mathbb{P}(X_{(k)} \geq x)$$
$$= 1 - \mathbb{P}(B_m(x) \geq k)$$
$$= 1 - [1 - \mathbb{P}(B_m(x) < k)]$$
$$= \mathbb{P}(B_m(x) < k) \xrightarrow[m\to\infty]{} \mathbb{P}(N < k)$$

Question 2 :

Qu'est-ce que la propriété de max-stabilité ?
$X \sim \mathcal{E}(1)$, est-ce que X ou $-X$ a une loi max-stable ?
Si oui, donner les coeffs permettant de déduire sa max-stabilité.

Si $X, X_1, \dots, X_n \overset{iid}{\sim} F$ et $M_n = \max(X_1, \dots, X_n)$
On dit que X a une loi max-stable si $\exists\, a_n, b_n > 0$ tq $\frac{M_n - b_n}{a_n} \xrightarrow[n\to\infty]{\mathcal{L}} X$

$GEV(\mu, \sigma, \xi)$

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

ξ : param de forme, μ : param de posit°
σ : param d'échelle.

Λ Gumbel : $\xi = 0$, $GEV(0,1,0)$
Φ_α Fréchet : $\xi > 0$, $GEV(1, \alpha^{-1}, \alpha^{-1})$
Ψ_α Weibull : $\xi < 0$, $GEV(-1, \alpha^{-1}, -\alpha^{-1})$

$F \in D(GEV(0,1,\xi))$: $h'(x) \xrightarrow[x \to x^F]{} \xi$ et $1 - F(b_n) = 1/n$
$a_n = h(b_n)$

$$h(x) = \frac{1-F(x)}{f(x)}$$

$X \sim \mathcal{E}(1) \Rightarrow F(x) = 1 - e^{-x}$ et $f(x) = e^{-x}$

$h(x) = \frac{e^{-x}}{e^{-x}} = 1 \Rightarrow h'(x) = 0 = \xi$

$F \in D(GEV(0,1,0)) = D(\Lambda)$ et $e^{-b_n} = 1/n \Rightarrow b_n = -\ln(1/n) = \ln(n)$
$a_n = h(b_n) = 1$

$M_n \le u_n$ si $u_n = a_n x + b_n \Leftrightarrow M_n \le a_n x + b_n$
$\Leftrightarrow \frac{M_n - b_n}{a_n} \le x$

$F \in D(GEV(0,1,\xi))$ si $h'(y) \xrightarrow[y \to x]{} \xi$
$\frac{M_n - b_n}{a_n} \xrightarrow{\mathcal{L}} GEV(0,1,\xi)$

$\Rightarrow b_n = F^{-1}\left(1 - \frac{1}{n}\right)$ $a_n = h(b_n)$

$F_{-X}(x) = P(-X \le x) = P(X \ge -x)$
$= 1 - P(X \le -x)$
$= 1 - [1 - e^{-(-x)}] = e^{x}$

$f_{-X}(x) = e^{x}$

$h_{-X}(x) = \frac{1 - F_X(x)}{f_X(x)} = \frac{1 - e^{x}}{e^{x}} = e^{-x} - 1$

$h'_{-X}(x) = -e^{-x} \xrightarrow[x\to\infty]{} 0 = \xi$

$F \in D(GEV(0,1,0))$ et $1 - e^{-b_n} = 1/n \Rightarrow b_n = \ln(1 - 1/n)$
$a_n = h(b_n) = 1$

Question 3 :

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

Expliquer à quel domaine d'attraction appartient $GPD(\beta, \xi)$

* $\xi = 0$

$F(x) = 1 - e^{-x/\beta}$

$f(x) = -\left(-\frac{1}{\beta}\right)e^{-x/\beta} = \frac{1}{\beta} e^{-x/\beta}$

$h(x) = \frac{1 - F(x)}{f(x)} = \frac{e^{-x/\beta}}{1/\beta\, e^{-x/\beta}} = \beta$

$h'(x) = 0 \Rightarrow \xi = 0$
$\Rightarrow GPD(\beta, 0) \in D(GEV(0,1,0) = \Lambda)$

* $\xi \neq 0$

$F(x) = 1 - [1 + \xi(x/\beta)]_+^{-1/\xi}$

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

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

$$h'(x) = \xi$$

$$\Rightarrow GPD(\beta,\xi) \in \begin{cases} D(\phi) & \text{si } \xi>0 \\ D(\psi) & \text{si } \xi<0 \end{cases}$$

Question 4 :

Expliquer ce que veut dire "F appartient au domaine d'attract° d'une loi max-stable G ($F \in D(G)$)"

$H(x) = F(\alpha x + \beta)$, $\alpha > 0$

A quel domaine d'attract° appartient H ?

Donner les relations qui existent entre les suites de normalisation pour H et F.

F densité finie positive f et $x^F = \infty$ tq

$$\frac{(1+\beta x)f(x)}{\bar{F}(x)} \xrightarrow[x \to x^F]{} c$$

À quel domaine d'attraction appartient F ?

$X_1, \dots, X_n \overset{iid}{\sim} F$ et $M_n = \max(X_1, \dots, X_n)$

$$F \in D(G) \Rightarrow \mathbb{P}\left(\frac{M_n - a_n}{b_n} \leq x\right) \xrightarrow[n\to\infty]{\mathcal{L}} G$$

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = \mathbb{P}(M_n \leq a_n x + b_n) = \left[F(a_n x + b_n)\right]^n$$
$$= \left[H\left(\frac{a_n x + b_n - \beta}{\alpha}\right)\right]^n \xrightarrow[n\to\infty]{\mathcal{L}} G \quad \text{avec } \tilde{a}_n = \frac{a_n}{\alpha} \text{ et } \tilde{b}_n = \frac{b_n - \beta}{\alpha}$$

$$h(x) = \frac{1-F(x)}{f(x)} = \frac{\bar{F}(x)}{f(x)}$$

$$\lim_{x\to x^F} \frac{(1+\beta x)f(x)}{\bar{F}(x)} = c \iff \lim_{x\to x^F} \frac{1+\beta x}{h(x)} = c$$

règle de l'hospital $\iff \lim_{x\to x^F} \frac{\beta}{h'(x)} = c$

$$\iff \lim_{x\to x^F} h'(x) = \frac{\beta}{c} \Rightarrow \xi = \frac{\beta}{c}$$

$$\Rightarrow F \in \begin{cases} D(\phi) & \text{si } \beta/c > 0 \\ D(\Lambda) & \text{si } \beta/c = 0 \\ D(\psi) & \text{si } \beta/c < 0 \end{cases}$$

Règle de l'Hôpital :

Soient f et g deux fonct° dérivables sur $]a,b[$ et g' ne s'annule pas

* si $\lim_a f = \lim_a g = 0$ et $\lim_a \frac{f'}{g'} = \ell$ alors $\lim_a \frac{f}{g} = \ell$
* si $\lim_a f = \lim_a g = +\infty$ et $\lim_a \frac{f'}{g'} = \ell$ alors $\lim_a \frac{f}{g} = \ell$

Question 5 :

$X_1, \dots, X_n \overset{iid}{\sim} F$ et $M_n = \max(X_1, \dots, X_n)$

Rappels : 1- $\tau > 0$, $(u_n) \in \mathbb{R}$

$$\mathbb{P}(M_n \leq u_n) \xrightarrow[n\to\infty]{} e^{-\tau} \iff n\bar{F}(u_n) \xrightarrow[n\to\infty]{} \tau$$

2- $\exists (u_n)$ tq $n\bar{F}(u_n) \xrightarrow[n\to\infty]{} \tau \iff \frac{\bar{F}(x)}{\bar{F}(x^-)} \xrightarrow[x\to x^F]{} 1$

Peut-on trouver (u_n) tq $\mathbb{P}(M_n \leq u_n) \xrightarrow[n\to\infty]{} e^{-\tau}$ si :

(i) $X_i \sim \mathcal{E}(\lambda)$: $\mathbb{P}(X_i > x) = e^{-\lambda x}$ $x > 0$

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

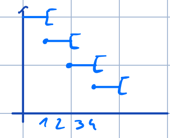
(ii) $X_i \sim \mathcal{G}(p)$: $\mathbb{P}(X_i = m) = p(1-p)^m \quad m \in \mathbb{N}$

(iii) $X_i \sim \mathcal{P}(\lambda)$: $\mathbb{P}(X_i = m) = e^{-\lambda}\dfrac{\lambda^m}{m!} \quad m \in \mathbb{N}$

(i) $\overline{F}(x) = \overline{F}(x^-) \Rightarrow \dfrac{\overline{F}(x)}{\overline{F}(x^-)} \xrightarrow[x \to x^F]{} 1$ donc Ok (cf Rappels)

(ii) Si $x \in \mathbb{R} \setminus \mathbb{N}^*$, $\dfrac{\overline{F}(x)}{\overline{F}(x^-)} = 1$ (sur les marches de l'escalier, c'est C^0)

Si $x \in \mathbb{N}^*$, $\dfrac{\overline{F}(x)}{\overline{F}(x^-)} = \dfrac{\overline{F}(x)}{\overline{F}(x-1)} = 1 - p < 1$



$\Rightarrow \limsup\limits_{x \to x^F} \dfrac{\overline{F}(x)}{\overline{F}(x-1)} = 1 \neq \liminf\limits_{x \to x^F} \dfrac{\overline{F}(x)}{\overline{F}(x-1)} = 1 - p$

$\Rightarrow$ limite n'existe pas $\Rightarrow$ NON.

iii) Si $x \in \mathbb{R} \setminus \mathbb{N}^*$, $\overline{F}(x) = \overline{F}(x^-)$

Si $x = m \in \mathbb{N}^*$, $\overline{F}(m) = 1 - \sum\limits_{k=0}^{m-1} e^{-\lambda}\dfrac{\lambda^k}{k!}$

$\overline{F}(m^-) = \overline{F}(m-1) = 1 - \sum\limits_{k=0}^{m-2} e^{-\lambda}\dfrac{\lambda^k}{k!}$

$$\Rightarrow \frac{\overline{F}(m)}{\overline{F}(m^-)} = \frac{1 - \sum_{k=0}^{m-1}(\lambda^k/k!)e^{-\lambda}}{1 - \sum_{k=0}^{m-2}(\lambda^k/k!)e^{-\lambda}}$$

Or $\dfrac{1 - \sum_{k=0}^{m-1}(\lambda^k/k!)e^{-\lambda}}{1 - \sum_{k=0}^{m-2}(\lambda^k/k!)e^{-\lambda}} \xrightarrow[m \to \infty]{} 1 \Rightarrow \dfrac{\overline{F}(m)}{\overline{F}(m^-)} \xrightarrow[m \to \infty]{} 1$ donc Ok

Question 6 :

1) On rappelle que les lois max-stables sont définies par

Fréchet ($\alpha > 0$) : $\Phi_\alpha(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ e^{-x^{-\alpha}} & \text{si } x > 0 \end{cases}$ queue lourde

Lorsqu'il y a bcp de valeurs dans la queue de distrib.

Weibull ($\alpha > 0$) : $\Psi_\alpha(x) = \begin{cases} e^{-(-x)^\alpha} & \text{si } x \leq 0 \\ 1 & \text{si } x > 0 \end{cases}$ queue normale

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

Montrer que $X \sim \Phi_\alpha \Leftrightarrow \ln(X^\alpha) \sim \Lambda \Leftrightarrow -X^{-1} \sim \Psi_\alpha$

Si $X \sim \Phi_\alpha$, $\mathbb{P}(X \leq x) = e^{-x^{-\alpha}} \mathbb{1}_{\{x > 0\}}$

$\Leftrightarrow \mathbb{P}(\ln(X^\alpha) \leq y) = \mathbb{P}(X \leq (e^y)^{1/\alpha})$

$= \mathbb{P}(X \leq e^{y/\alpha})$

$= e^{-(e^{y/\alpha})^{-\alpha}} \mathbb{1}_{\{e^{y/\alpha} > 0\}} \to$ Vrai $\forall y \in \mathbb{R}$

$= e^{-e^{-y}}$

Donc $X \sim \Phi_\alpha \Leftarrow \ln(X^\alpha) \sim \Lambda$

Si $\ln(X^\alpha) \sim \Lambda \Leftrightarrow \mathbb{P}(\ln(X^\alpha) \leq x) = \Lambda(x) = e^{-e^{-x}}$

$\Leftrightarrow \mathbb{P}(X \leq e^{x/\alpha}) = e^{-e^{-x}}$

$$\Leftrightarrow \mathbb{P}(-X^{-1} \leq -e^{-x/\alpha}) = e^{-e^{-x}}$$
$$\Leftrightarrow \mathbb{P}(-X^{-1} \leq y) = e^{-e^{-\alpha h(y)}}$$
$$\Leftrightarrow \mathbb{P}(-X^{-1} \leq y) = e^{-(-y)^{\alpha}}$$

Donc $X \sim \Phi_\alpha \Leftrightarrow h(X^{-1}) \sim \Lambda \Leftrightarrow -X^{-1} \sim \Psi_\alpha$

2) Qu'est-ce que le domaine d'attraction d'une loi max stable ? Quelles sont les conditions qui définissent le domaine d'attract° de la loi de Frechet.

Soit G une loi max stable,

$$D(G) = \left\{ F \text{ f.d.r tq } \exists\, a_m > 0, b_m \text{ des suites tq } \mathbb{P}\left(\frac{M_m - b_m}{a_m} \leq x\right) \xrightarrow[m\to\infty]{} G(x), \text{ où } M_m = \max(X_1, \ldots, X_m),\ X_i \overset{iid}{\sim} F \right\}$$

$$F \in D(\Phi_\alpha) \Leftrightarrow \bar{F}(x) = x^{-\alpha} L(x) \quad \text{où } L \text{ à variations lentes, i.e } \forall x,\ \frac{L(tx)}{L(t)} \xrightarrow[t\to\infty]{} 1$$

3) Quel lien relie les distribut° GEV (Generalised Extreme Value distributions) et les distribut° GPD (Generalised Pareto distribut°) ?

$$(X_i) \text{ iid de f.d.r } F \text{ et } F \in D(GEV(\mu, \sigma, \xi)) \Leftrightarrow B_m(x) = \sum_{i=1}^{m} \mathbb{1}_{\{X_i > x\}} \sim GPD(\beta, \xi)$$

4) Quelles sont les procédures statistiques à utiliser pour estimer les paramètres des distributions GEV ? (discuter suivant la nature des observations)

Estimation des params GEV : Cette distribution est utilisée lorsqu'on dispose de données du type maximum par blocs

* Choix du modèle probabiliste $\Rightarrow \mathbb{P}(M_n \leq x) \sim GEV(\mu, \sigma, \xi)$
* Estimation par maximum de vraisemblance des paramètres
* Par exploration graphique Q-Q plot
* Considération des niveaux de retour et des périodes de retour
* Estimation du niveau de retour de période $\frac{1}{p}$ et de quantile $1-p$ $0<p<1$.
* Méthode des moments pondérés qui consiste à égaliser moments théorique de la distribution GEV avec les moments empiriques conduits à partir des données.

GPD : elle est utilisée pour modéliser les observat° des dépassements de seuil

$$\mathbb{P}(Y_u < y \mid Y_u > u) = 1 - \left[1 + \xi\left(\frac{y}{\sigma_u}\right)\right]_+^{-1/\xi}$$

Il faut donc choisir méticuleusement le seuil u (si peu élevé l'approximat° est mauvaise m̂ si ça permet d'↗ le nb de donnée et trop élevé limite le nb de données et les estimateurs sont moins précis $\Rightarrow$ Utilisat° des prop de la loi GPD pour déterminer un seuil. Stabilité par seuil, fonct° de dépassat° moyen, graphique de dépassement moyen.

Question 7 :

1) $x^F = x^G$ et $\frac{\bar{F}(x)}{\bar{G}(x)} \xrightarrow[x \to x^F]{} c \in (0, \infty)$

Mq F et G appartiennent au même domaine d'attraction $(GEV(0,1,\xi))$ et donner le lien entre les ctes de normalisat°.

$GEV(0,1,\xi) = \exp(-(1+\xi x)_+^{-1/\xi})$ $x \in \mathbb{R}$.

On suppose $G \in D(GEV(0,1,\xi))$

$$\mathbb{P}(M_n \leq a_n x + b_n) \xrightarrow[n\to\infty]{} \exp(-\overbrace{(1+\xi x)_+^{-1/\xi}}^{\tau_1})$$

$$\Leftrightarrow n\,\bar{G}(a_n x + b_n) \xrightarrow[n\to\infty]{} \tau_1$$

$$\frac{n\,\bar{F}(x)}{n\,\bar{G}(x)} \xrightarrow[x\to x^F]{} c \Rightarrow n\bar{F}(x) \xrightarrow[x\to x^F]{} c\tau_1 \Rightarrow F \in D(GEV(0,1,\xi)^c)$$

$$(GEV(0,1,\xi))^c = \exp(-c[1+\xi(a_n x + b_n)]^{-1/\xi})$$
$$= \exp(-[c^{-\xi}(1+\xi(a_n x+b_n))]^{-1/\xi})$$
$$= \exp(-[\xi c^{-\xi} a_n x + \xi c^{-\xi} b_n + c^{-\xi}]^{-1/\xi})$$

Identification à une $GEV(0,1,\xi)$ i.e $\exp(-[1+\xi(c_n x + d_n)]^{-1/\xi})$

$$\Rightarrow \xi c_n = \xi c^{-\xi} a_n \text{ et } 1+\xi d_n = \xi c^{-\xi} b_n + c^{-\xi}$$

2) Réciproquement, supposons que $x^F = x^G$ et $F, G \in D(\Lambda)$

$$\Lambda(x) = e^{-e^{-x}}$$

$$F^n(c_n x + d_n) \xrightarrow[n\to\infty]{} \Lambda(x) \text{ et } G^n(c_n x + d_n) = \Lambda(x+b)$$

Mq $\frac{\bar{F}(x)}{\bar{G}(x)} \xrightarrow[x\to x^F]{} c \in (0,\infty)$ et caractériser c.

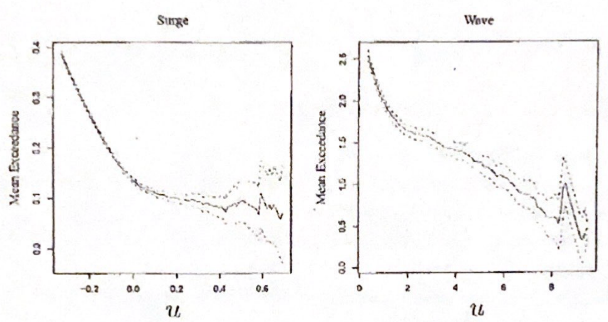
$$F^n(c_n x + d_n) \xrightarrow[n\to\infty]{} e^{-e^{-x}} \Leftrightarrow n\bar{F}(c_n x + d_n) \to \tau_1 = e^{-x}$$

$$G^n(c_n x + d_n) \xrightarrow[n\to\infty]{} e^{-e^{-(x+b)}} \Leftrightarrow n\,\bar{G}(c_n x + d_n) \to \tau_2 = e^{-(x+b)}$$

$$\lim_{x\to x^F} \frac{\bar{F}(x)}{\bar{G}(x)} = \lim_{n\to\infty} \frac{n\bar{F}(c_n x + d_n)}{n\,\bar{G}(c_n x + d_n)} = \frac{\tau_1}{\tau_2} = \frac{e^{-x}}{e^{-(x+b)}} = e^{b} = c$$

Q8

Quelle utilisation fait-on de la fonction de dépassement moyen empirique? Vous trouverez ci-dessous deux représentations graphiques de la fonction de dépassement moyen empirique pour deux grandeurs physiques (Surge et Wave) en fonction d'un seuil u. Quel comportement de la fonction de dépassement est attendu pour les grandes valeurs de u?



La fonct° de dépassement moyen empirique sert à identifier le seuil u à partir duquel les dépassements suivent une GPD
Lorsque la v.a suit une GPD de seuil u alors le graphique devient approximativement linéaire au-delà de ce seuil.

Pour grandes valeurs de u $\Longrightarrow$ très volatile car peu de valeurs extrêmes.