1. Analyse des extrêmes univariés dans un cadre statistique

1.1 Lois limites pour le maximum

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

$\mathbb{P}(M_n \leq x) = [F(x)]^n$

- $F(x) < 1$ $\qquad \mathbb{P}(M_n \leq x) \xrightarrow[n\to\infty]{} 0$
- $x^F = \sup\{x : F(x) < 1\}$ $\qquad M_n \xrightarrow[n\to\infty]{\mathbb{P}} x^F$

↳ La distribution de M_n est dégénérée $\quad \mathbb{P}(M_n \leq x) = \begin{cases} 0 & x \leq x^F \\ 1 & x > x^F \end{cases}$

Théorie des sommes de variables aléatoires

$\mathbb{E}(X_i) = m$ et $\mathbb{V}(X_i) = \sigma^2 < \infty$

- LGN : $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n\to\infty]{\mathbb{P}} m$

↳ La distribution de $\bar{X}_n$ est dégénérée $\quad \mathbb{P}(\bar{X}_n \leq x) = \begin{cases} 0 & x \leq \mathbb{E}(X) \\ 1 & x > \mathbb{E}(X) \end{cases}$

$\mathbb{E}(\bar{X}_n) = m$

- TCL : $\mathbb{P}(\bar{X}_n \leq \sigma_n \cdot x + m_n) \xrightarrow[n\to\infty]{} \Phi(x)$

$\sigma_n = \sqrt{\mathbb{V}(\bar{X}_n)} = \sqrt{\sigma^2/n}$

→ même distribution limite quelque soit F

→ normalisation dépend des 2 premiers moments de F

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

> Quelle condition sur F assure que la limite $\mathbb{P}(M_n \leq u_n)$ existe pour une suite appropriée u_n ?

- $\mathbb{P}(M_n \leq u_n) \xrightarrow[n\to\infty]{} e^{-\tau} \iff \underbrace{n\bar{F}(u_n)}_{\text{nb dépassements moyen au-delà du seuil } u_n} \xrightarrow[n\to\infty]{} \tau$

$\mathbb{P}(M_n \leq u_n) = [F(u_n)]^n = [1 - \bar{F}(u_n)]^n = \exp\{n \cdot \ln\{1 - \bar{F}(u_n)\}\}$

$F(x_F) = 1 \Rightarrow \bar{F}(u_n) = 1 - F(u_n) \xrightarrow[n\to\infty]{} 0$ car $u_n \xrightarrow[n\to\infty]{} x^F$

Rappel : $\ln(1+u) \underset{u\to 0}{\sim} u$ (développement limité)

$\Rightarrow \mathbb{P}(M_n \leq u_n) \xrightarrow[n\to\infty]{} \exp\{-\underbrace{n\bar{F}(u_n)}_{\xrightarrow[n\to\infty]{} \tau}\}$

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

i.e. si les sauts de $\overline{F}$ ne décroissent pas suffisamment vite, alors il n'existe pas de distribution limite non dégénérée pour M_n.

Ex : $X \sim Geom(q)$ $\quad \mathbb{P}(X = k) = q(1-q)^k, \ k \geqslant 1$

- $x \in \mathbb{R} \setminus \mathbb{N}^* \Rightarrow \frac{\overline{F}(x)}{\overline{F}(x-)} = 1$

- $x = k \in \mathbb{N}^*$

$\overline{F}(x) = \mathbb{P}(X > k) = \sum_{j=k+1}^{\infty} \mathbb{P}(X=j)$

$= \sum_{j=k+1}^{\infty} q(1-q)^{j-1} = q(1-q)^k \sum_{j=0}^{\infty} (1-q)^j$

$= q(1-q)^k \frac{1}{1-(1-q)} = (1-q)^k$

$\frac{\overline{F}(x)}{\overline{F}(x-)} = \frac{\mathbb{P}(X>k)}{\mathbb{P}(X \geqslant k)} = \frac{(1-q)^k}{(1-q)^{k-1}} = 1-q < 1$

$\lim_{x \to x^F} \frac{\overline{F}(x)}{\overline{F}(x-)} \neq 1$

> Quelle conséquence ce choix de u_n a-t-il sur le nombre de dépassements de ce seuil par $X_1, \dots, X_n$?

$N_n = \sum_{i=1}^{n} \mathbb{1}_{\{X_i > u_n\}}$: nb de dépassement du seuil u_n par $X_1, \dots, X_n$

- Si $\exists (u_n)$ tq $n\overline{F}(u_n) \xrightarrow[n\to\infty]{} \tau$

Alors $N_n \xrightarrow[n\to\infty]{\mathcal{L}} N \sim Poisson(\tau)$ $\quad \mathbb{P}(N=n) = e^{-\tau} \frac{\tau^n}{n!} \quad n \geqslant 0$

"loi des évènements rares"

Les transformées de Laplace :

- $L_N(t) = \mathbb{E}(e^{-tN}) = \sum_{n=0}^{\infty} e^{-tn} \, \mathbb{P}(N=n)$

$= \sum_{n=0}^{\infty} e^{-tn} \cdot e^{-\tau} \frac{\tau^n}{n!} = e^{-\tau} \sum_{n=0}^{\infty} \frac{(e^{-t} \cdot \tau)^n}{n!}$

$= e^{-\tau} [e^{(e^{-t} \cdot \tau)}]^{-1} \sum_{n=0}^{\infty} e^{(e^{-t} \cdot \tau)} \frac{(e^{-t} \cdot \tau)^n}{n!} = e^{-\tau(1-e^{-t})}$

- $L_{N_n}(t) = \mathbb{E}(e^{-tN_n}) = \mathbb{E}(e^{-t \cdot \sum_{i=1}^{n} \mathbb{1}_{\{X_i > u_n\}}}) \underset{X_i \text{ iid}}{=} \left(\mathbb{E}[e^{-t \cdot \mathbb{1}_{\{X_i > u_n\}}}]\right)^n$

$\mathbb{1}_{\{X_i > u_n\}} = \begin{cases} 0 & \text{si } X_i \leqslant u_n \text{ i.e. avec proba } 1 - \overline{F}(u_n) \\ 1 & \text{si } X_i > u_n \text{ i.e. avec proba } \overline{F}(u_n) \end{cases}$

$$L_{N_n}(t) = \left(e^{-t\cdot 0}\,[1-\bar{F}(u_n)] + e^{-t\cdot 1}\,\bar{F}(u_n)\right)^n$$
$$= \left(1-\bar{F}(u_n) + e^{-t}\bar{F}(u_n)\right)^n$$
$$= \left(1-\bar{F}(u_n)(1-e^{-t})\right)^n$$
$$= \exp\{n\,\ln\{1-\bar{F}(u_n)(1-e^{-t})\}\}$$

$\bar{F}(u_n) \xrightarrow[n\to\infty]{} 0$ car $u_n \xrightarrow[n\to\infty]{} x^F \Rightarrow \ln\{1-\bar{F}(u_n)(1-e^{-t})\} \xrightarrow[n\to\infty]{} -\bar{F}(u_n)(1-e^{-t})$

$\Rightarrow L_{N_n}(t) \xrightarrow[n\to\infty]{} \exp\{-n\,\bar{F}(u_n)(1-e^{-t})\} \xrightarrow[n\to\infty]{} \exp\{-\tau(1-e^{-t})\} = L_N(t)$

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

Ex : $X \sim \mathrm{Exp}(1)$ $\quad F(x) = 1-\exp(-x) \quad x>0$

On cherche $u_n = a_n x + b_n$ tq $\mathbb{P}(M_n \leq u_n) \longrightarrow g$

- $a_n = 1$
- $b_n = \log n$

$\mathbb{P}\left(\frac{M_n - a_n}{b_n} \leq x\right)$

$$\Rightarrow \mathbb{P}(M_n \leq u_n) = \mathbb{P}(M_n \leq x + \log n)$$
$$= [F(x+\log n)]^n$$
$$= [1-\exp\{-x-\log(n)\}]^n$$
$$= [1-\exp\{-x\}\cdot \exp\{\underbrace{-\log(n)}_{\log(1/n)}\}]^n$$
$$= \left[1-\frac{1}{n}\exp\{-x\}\right]^n$$
$$= \exp\left\{n\,\ln\left\{1-\underbrace{\frac{1}{n}e^{-x}}_{\xrightarrow[n\to\infty]{}0}\right\}\right\} \qquad \ln(1+u) \xrightarrow[u\to 0]{} u$$
$$= \exp\left\{n\times\left(-\frac{1}{n}e^{-x}\right)\right\}$$
$$= \exp\{-e^{-x}\} \leftarrow \text{C'est la loi de Gumbell}$$

Classes d'équivalence de distributions limites

$F_2(ax+b) = F_1(x) \quad \forall x$: F_1, F_2 de même type
i.e. des distributions de même type sont les même à des facteurs de position (b) et d'échelle (a) près.

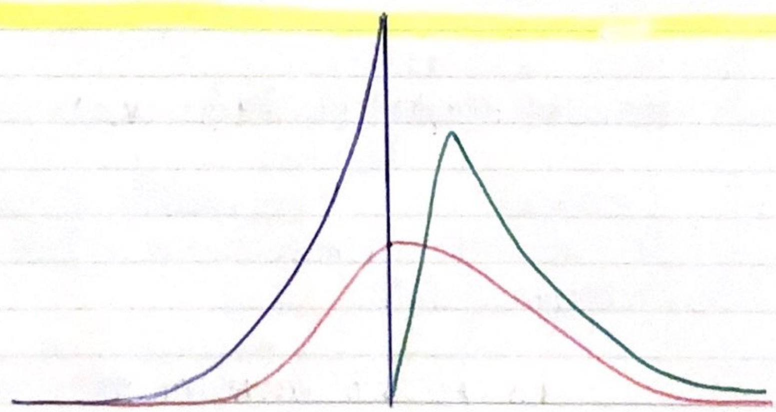
Ex $\mathcal{N}(\mu_1, \sigma_1^2)$; $\mathcal{N}(\mu_2, \sigma_2^2)$
$\Gamma(\alpha, \beta_1)$; $\Gamma(\alpha, \beta_2)$

- Fisher - Tippett

Si $\exists c_n > 0, d_n$ tq $\mathbb{P}\left(\frac{M_n - d_n}{c_n} \leq x\right) \xrightarrow[n \to \infty]{} G(x)$ non dégénérée

Alors G est du même type que l'une des distributions suivantes :

- Fréchet : $\Phi_\alpha(x) = \begin{cases} 0 & x \leq 0 \\ \exp\{-x^{-\alpha}\} & x > 0 \end{cases} \quad \alpha > 0$
- Weibull : $\Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^{\alpha}\} & x \leq 0 \\ 1 & x > 0 \end{cases} \quad \alpha > 0$
- Gumbel : $\Lambda(x) = \exp\{-e^{-x}\} \quad x \in \mathbb{R}$



(i) $X, X_1, \dots, X_n \overset{iid}{\sim} \Phi_\alpha \Rightarrow n^{-1/\alpha} \max(X_1, \dots, X_n) \overset{\mathcal{L}}{=} X$

$M_n \overset{\mathcal{L}}{=} n^{1/\alpha} X$

(ii) $X, X_1, \dots, X_n \overset{iid}{\sim} \Psi_\alpha \Rightarrow n^{-1/\alpha} \max(X_1, \dots, X_n) \overset{\mathcal{L}}{=} X$

(iii) $X, X_1, \dots, X_n \overset{iid}{\sim} \Lambda \Rightarrow \max(X_1, \dots, X_n) - \ln n \overset{\mathcal{L}}{=} X$

$M_n \overset{\mathcal{L}}{=} X + \ln n \to$ tend moins vite vers ∞ que Fréchet

↳ Distribution "max-stables"

(i) $M_n = \max(X_1, \dots, X_n) \overset{\mathcal{L}}{=} n^{1/\alpha} X \Rightarrow \mathbb{E}(M_n) = n^{1/\alpha}\mathbb{E}(X)$ si $\mathbb{E}(X) < +\infty$

$\sigma(M_n) = n^{1/\alpha}\sigma(X)$ si $\mathbb{E}(X) < +\infty$ et $\mathbb{E}(X^2) < +\infty$

$$\mathbb{E}(X^k) = \int_0^\infty x^{k-1}\, \mathbb{P}(X > x)\, dx$$

$$= \int_0^\infty x^{k-1}\left[1 - \exp\{-x^{-\alpha}\}\right] dx$$

<u>Rappel</u> : $e^u \underset{u \to 0}{\longrightarrow} 1 + u$ (developpement limité)

$$\mathbb{E}(X^k) \sim \int_0^\infty x^{k-1}\left[1 - (1 - x^{-\alpha})\right] dx = \int_0^\infty x^{-\alpha+k-1}\, dx$$

$\Rightarrow \mathbb{E}(X) < +\infty \Leftrightarrow \alpha > 1$ et $\mathbb{E}(X^2) < +\infty \Leftrightarrow \alpha > 2$

$$\mathbb{P}(n^{-1/\alpha} M_n \leq x) = \mathbb{P}(M_n \leq n^{1/\alpha} x) = \left[\Phi_\alpha(n^{1/\alpha} x)\right]^n$$
$$= \left[\exp\{-(n^{1/\alpha} x)^{-\alpha}\}\right]^n = \exp\{-x^{-\alpha}\}$$
$$= \Phi_\alpha(x)$$

(ii) $M_n \overset{\mathcal{L}}{=} X + \ln n \Rightarrow \mathbb{E}(M_n) = \mathbb{E}(X) + \ln n$

$\sigma(M_n) = \sigma(X)$

$$\mathbb{P}(M_n - \ln n \leq x) = \mathbb{P}(M_n \leq x + \ln n) = \left[\Lambda(x + \ln n)\right]^n$$
$$= \left[\exp\{e^{-(x+\ln n)}\}\right]^n = \exp\{\cancel{n}(e^{-x} \cdot \tfrac{1}{\cancel{n}})\} = \exp\{e^{-x}\}$$
$$= \Lambda(x)$$

$X \sim \Phi_\alpha \Leftrightarrow -X^{-1} \sim \Psi_\alpha \Leftrightarrow \ln X^\alpha \sim \Lambda$

$$\mathbb{P}(\ln X^\alpha \leq x) = \mathbb{P}(\alpha \ln X \leq x) = \mathbb{P}(X \leq e^{x/\alpha}) = \Phi_\alpha(e^{x/\alpha})$$
$$= \exp\{-(e^{x/\alpha})^{-\alpha}\} = \exp\{-e^{-x}\} = \Lambda(x)$$

Distributions des extrêmes généralisées (GEV)

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

- ξ paramètre de forme
- μ paramètre de position
- σ paramètre d'échelle

$$D = \begin{cases}]-\frac{1}{\xi}; +\infty[\\ ]-\infty; \frac{1}{\xi}[\\ \mathbb{R} \end{cases}$$

indice de queue
- $\xi > 0 \to$ épaisse
- $\xi < 0 \to$ finie
- $\xi = 0 \to$ intermédiaire

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

- Fréchet : $GEV(1, \alpha^{-1}, \alpha^{-1})$
- Weibull : $GEV(-1, \alpha^{-1}, -\alpha^{-1})$
- Gumbel : $GEV(0, 1, 0)$

$$GEV\left(\mu = 1, \sigma = \frac{1}{\alpha}, \xi = \frac{1}{\alpha}\right) = \exp\left\{-\left[1+\frac{1}{\alpha}\left(\frac{x-1}{1/\alpha}\right)\right]_+^{-\alpha}\right\} = \exp\{-x_+^{-\alpha}\}$$
$$= \exp\{-x^{-\alpha}\} \mathbb{1}_{\{x \in \mathbb{R}^+\}}$$
$$= \Phi_\alpha(x)$$

- $X, X_1, \dots, X_n \overset{iid}{\sim} GEV(0,1,\xi) \Rightarrow n^{-\xi}\left(\max(X_1,\dots,X_n) - \frac{n^\xi - 1}{\xi}\right) \overset{\mathcal{L}}{=} X$

- Fisher-Tippet

Si $\exists c_n > 0, d_n$ tq $\mathbb{P}\left(\frac{M_n - d_n}{c_n} \leq x\right) \xrightarrow[n\to\infty]{} G(x)$ non dégénérée

Alors G est du même type que $G_\xi(x)$

→ pas de garantie sur l'existence d'une limite non-dégénérée
→ pas de précision sur ξ
→ on ne sait pas comment choisir c_n, d_n

Si $\sigma^2 = \infty$ — Sum-Stable-Law dist. ↳ en général, pas d'expression analytique (contre exemple : loi de Cauchy)

Alors $\exists\, c > 1/2$ tq

$$\mathbb{P}\left(n^{1-c}(\bar{X}_n - a) \leq x\right) \xrightarrow[n\to\infty]{} SSL_c(x)$$

↳ $\begin{cases} m & c \in (1/2\,;\,1) \\ 0 & \text{sinon} \end{cases}$

↳ les sommes de va ont des lois limites paramétrées par c

- la constance c est liée à l'épaisseur de la queue de F
  si F symétrique et loi limite maximum non dégénérée : $c = \xi$

- en général, lien fort entre théorie des lois limites des sommes et celle des maximums, lorsque les queues sont épaisses

Pseudos preuves : si une loi limite non dégénérée existe, elle doit être de la même forme que donnée précédement

• TCL

Simplifications : $\mathbb{E}(X_i) = 0$ et $\mathbb{V}(X_i) = 1$
$S_n = n^{1/2} \bar{X}_n \Rightarrow \mathbb{E}(S_n) = 0$ et $\mathbb{V}(S_n) = 1$

On suppose $S_n \xrightarrow[n\to\infty]{\mathcal{L}} Y$ de distribution connue
$\Rightarrow \sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{\mathcal{L}} n^{1/2} Y$

$k \in \mathbb{N}^*$, $r_n = [n/k]$
$\sum_{i=1}^{n} X_i \approx \sum_{j=1}^{k} \underbrace{\sum_{i=r_n(j-1)+1}^{r_n j} X_i}$
les sommes partielles sont indépendantes comme sommes de va indépendantes

n grand $\Rightarrow r_n$ grand $\Rightarrow$ chaque somme partielle $\xrightarrow[n\to\infty]{\mathcal{L}} r_n^{1/2} Y$

$Y_1, \dots, Y_k \overset{\mathcal{L}}{=} Y$
$n^{1/2} Y = r_n^{1/2} Y_1 + \dots + r_n^{1/2} Y_k$
$k^{1/2} Y = Y_1 + \dots + Y_k$

La loi normale est la seule à posséder cette propriété
$\Rightarrow$ la loi limite est la loi normale lorsque la variance des X_i est finie

• Loi limite des maximums

$(M_n - b_n)/a_n \xrightarrow[n\to\infty]{\mathcal{L}} Y$ de distribution connue
$M_n \xrightarrow[n\to\infty]{\mathcal{L}} a_n Y + b_n$

$k \in \mathbb{N}^*$, $r_n = [n/k]$. $M_n \approx \max\limits_{j=1,\dots,k} \underbrace{\max(X_{r_n(j-1)+1}, \dots, X_{r_n j})}$
les max partiels sont independants comme max de va indépendantes

n grand $\Rightarrow r_n$ grand $\Rightarrow$ chaque max partiel $\xrightarrow[n\to\infty]{\mathcal{L}} a_{r_n} Y + b_{r_n}$

$Y_1, \dots, Y_k \overset{\mathcal{L}}{=} Y$: $a_n Y + b_n = a_{r_n} \max(Y_1, \dots, Y_k) + b_{r_n}$
$\dfrac{a_n Y + b_n - b_{r_n}}{a_{r_n}} = \max(Y_1, \dots, Y_k)$

Condition : $G(A_k x + B_k) = [G(x)]^k \longrightarrow$ GEV

Domaines d'attraction

- Théorème de convergence de la loi limite du maximum

Si $\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = [F(a_n x + b_n)]^n \xrightarrow[n\to\infty]{} G(x)$ ↳ non dégénérée

Alors G a une distribution des extrêmes généralisée (GEV)

- $F \in D(G)$ si $\exists (a_n), (b_n)$ tq convergence

Comprendre les conditions sur F qui conduisent à $G(\xi)$:
- trouver a_n et b_n pour F
- déterminer les situations où $(a_n), (b_n)$ n'existent pas
- identifier les interconnexions avec les lois stables par addition
- analyser les taux de convergence de M_n normalisé vers une dist. GEV

<u>Rappel</u> : stabilité par maximum

$$X, X_1, \dots, X_n \overset{iid}{\sim} GEV(0,1,\xi) \Rightarrow \frac{M_n - b_n}{a_n} \overset{\mathcal{L}}{=} X \text{ avec } a_n = n^{\xi}, \quad b_n = (n^{\xi} - 1)/\xi$$

Ici, pas de convergence, la limite est vraie $\forall n$.

(i) $X_i \sim Exp(1) \Rightarrow M_n - \log n \xrightarrow[n\to\infty]{\mathcal{L}} \Lambda$

$X_i \sim \Lambda \Rightarrow M_n - \log n \overset{\mathcal{L}}{=} \Lambda$

$$\Rightarrow \frac{\mathbb{P}_{Exp(1)}(X_i > x)}{\mathbb{P}_\Lambda(X_i > x)} \xrightarrow[x\to\infty]{} 1 \qquad e^u \xrightarrow[u\to 0]{} 1+u$$

$$\frac{\mathbb{P}_{E(1)}(X_i > x)}{\mathbb{P}_\Lambda(X_i > x)} = \frac{\exp(-x)}{1 - \exp(-e^{-x})} \xrightarrow[x\to\infty]{} \frac{\exp(-x)}{1-(1-e^{-x})} = 1$$

(ii) $X_i \sim \mathcal{U}([-1,0]) \Rightarrow nM_n \xrightarrow[n\to\infty]{\mathcal{L}} \Psi_1$

$X_i \sim \Psi_1 \Rightarrow nM_n \overset{\mathcal{L}}{=} \Psi_1$

$$\Rightarrow \frac{\mathbb{P}_{\mathcal{U}([-1,0])}(X_i > x)}{\mathbb{P}_{\Psi_1}(X_i > x)} \xrightarrow[x\to\infty]{} 1$$

$$X \sim \mathcal{U}([-1,0]) \Rightarrow \mathbb{P}(X < x) = \frac{x-(-1)}{0-(-1)} = x + 1$$

$$\mathbb{P}(nM_n \leq x) = \mathbb{P}(M_n \leq x/n) = [(x/n)+1]^n = \exp\left\{n \ln\left\{\frac{x}{n}+1\right\}\right\}$$

$\ln(1+u) \underset{u\to 0}{\to} u$

$$\xrightarrow[n\to\infty]{} \exp\left\{n \cdot \frac{x}{n}\right\} = \exp\{-(-x)^1\} = \Psi_1(x)$$

(iii) $X_i \sim$ Cauchy $\Rightarrow \frac{\pi}{n} M_n \xrightarrow[n\to\infty]{\mathcal{L}} \Phi_1$

$X_i \sim \Phi_1 \quad \Rightarrow \frac{1}{n} M_n \stackrel{\mathcal{L}}{=} \Phi_1$

$\Rightarrow \dfrac{\mathbb{P}_{Cauchy}(X_i > x)}{\mathbb{P}_{\Phi_1}(X_i > x)} \xrightarrow[x\to\infty]{} \dfrac{1}{\pi}$

$X \sim$ Cauchy $\quad f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} \mathbb{1}_{\{x\in\mathbb{R}\}}$

$\bar{F}(x) \underset{x\to x^F}{\sim} \frac{1}{\pi} \cdot \frac{1}{x} \xrightarrow[x\to\infty]{} 0$

L'Hospital

$$\lim_{x\to\infty} \frac{\mathbb{P}_{Cauchy}(X_i > x)}{\mathbb{P}_{\Phi_1}(X_i > x)} = \lim_{x\to\infty} \frac{\frac{1}{\pi}\cdot\frac{1}{x}}{\exp\{-x^{-1}\}} \stackrel{\downarrow}{=} \lim_{x\to\infty} \frac{-\frac{1}{\pi}\cdot\frac{1}{x^2}}{-\frac{1}{x^2}\exp\{-x^{-1}\}} = \frac{1}{\pi}\lim_{x\to\infty} e^{1/x} = \frac{1}{\pi}$$

$$\mathbb{P}\left(\frac{\pi}{n} M_n \leq x\right) = \mathbb{P}\left(M_n \leq \frac{1}{\pi} n x\right) = \left[1 - \bar{F}\left(\frac{1}{\pi} n x\right)\right]^n$$

$$\xrightarrow[n\to\infty]{} \left[1 - \frac{1}{\pi}\cdot\frac{1}{\frac{1}{\pi} n x}\right]^n = \exp\left\{n \ln\left\{1 - \frac{1}{nx}\right\}\right\}$$

$$\xrightarrow[n\to\infty]{} \exp\left\{\cancel{n}\cdot\left(-\frac{1}{\cancel{n}x}\right)\right\} = \exp\{-x^{-1}\} = \Phi_1(x)$$

- F et G équivalentes en termes de queue de distribution si (1) elles ont le même point extrêmal ($x^F = x^G$)
  (2) $\lim_{x\to x^F} \frac{\bar{F}(x)}{\bar{G}(x)} = c \in (0, \infty)$

- Si F et G équivalente en termes de queue de distribution Alors F et G appartiennent au même domaine d'attraction

F, G 2 FdR avec $F \in D(GEV)$

$\Rightarrow \exists\, a_n > 0, b_n$ tq $\mathbb{P}\left(\frac{M_n - a_n}{b_n} \leq x\right) = [F(a_n x + b_n)]^n \xrightarrow[n\to\infty]{} G_\xi(x) = \exp\{-(1+\xi x)_+^{-1/\xi}\}$

$\ln(1+u) \sim u$

$\Rightarrow n \ln[1 - \bar{F}(a_n x + b_n)] \xrightarrow[n\to\infty]{} -n\bar{F}(a_n x + b_n) \xrightarrow[n\to\infty]{} -(1+\xi x)_+^{-1/\xi}$

$\Rightarrow n\bar{F}(a_n x + b_n) \xrightarrow[n\to\infty]{} (1+\xi x)_+^{-1/\xi}$

F et G équivalentes $\Rightarrow \frac{\bar{F}(x)}{\bar{G}(x)} \xrightarrow[x\to x^F]{} c \Rightarrow n\bar{G}(a_n x + b_n) \longrightarrow \frac{1}{c}(1+\xi x)_+^{-1/\xi}$

On veut $a_n' > 0, b_n'$ tq $\mathbb{P}\left(\frac{M_n - a_n'}{b_n'} \leq y\right) \xrightarrow[n\to\infty]{} G_\xi(y) \Leftrightarrow n\bar{G}(a_n' y + b_n') \xrightarrow[n\to\infty]{} (1+\xi y)_+^{-1/\xi}$

Par identification, on veut $\frac{1}{c}(1+\xi x)_+^{-1/\xi} = (1+\xi y)_+^{-1/\xi} \quad \Rightarrow \ldots$

$$\dots \Rightarrow (1+\xi x)^{-1/\xi} = c(1+\xi y)^{-1/\xi} \Rightarrow 1+\xi x = c^{-\xi}(1+\xi y) \Rightarrow x = \frac{c^{-\xi}(1+\xi y)-1}{\xi}$$

$$\Rightarrow x = \frac{c^{-\xi}-1}{\xi} + c^{-\xi} y$$

$$\Rightarrow a_n x + b_n = a_n\left(\frac{c^{-\xi}-1}{\xi} + c^{-\xi} y\right) + b_n = \underbrace{a_n \cdot c^{-\xi}}_{a_n'} y + \underbrace{a_n \cdot \frac{c^{-\xi}-1}{\xi} + b_n}_{b_n'}$$

Une intuition sur les conditions du domaine d'attraction

X va continue de densité f

Inverse de la fonction de hasard : $\boxed{h(x) = \dfrac{1-F(x)}{f(x)} \quad x_F < x < x^F}$

On peut mq $\forall u, x \; \exists y$ tq $u \leq y \leq u + x h(u)$

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

On suppose $h'(y) \xrightarrow[y \to x^F]{} \xi$

On définit $b_n = F^{-1}(1-1/n) \Rightarrow \bar{F}(b_n) = 1/n$

$a_n = h(b_n)$

$$\frac{1-F(u+xh(u))}{1-F(u)} = \frac{1-F(a_n x + b_n)}{1/n} = n\bar{F}(a_n x + b_n) = [1+h'(b_n)x]_+^{-1/h'(b_n)}$$

$$\xrightarrow[\substack{n\to\infty \\ (b_n \to x^F)}]{} [1+\xi x]_+^{-1/\xi}$$

$$\exp\{-n\bar{F}(a_n x + b_n)\} \xrightarrow[n\to\infty]{} \exp\left\{-[1+\xi x]_+^{-1/\xi}\right\}$$

$$-n\bar{F}(a_n x + b_n) \underset{n\to\infty}{\sim} n \ln\{1-\bar{F}(a_n x + b_n)\} = \ln\{[F(a_n x + b_n)]^n\}$$

$$\Rightarrow [F(a_n x + b_n)]^n \xrightarrow[n\to\infty]{} \exp\left\{-[1+\xi x]_+^{-1/\xi}\right\}$$

Dans ce cas, conditions d'appartenance au domaine d'attraction $GEV(0,1,\xi)$:
$h'(y) \xrightarrow[y\to x^F]{} \xi$ et on choisit a_n, b_n tq $1-F(b_n) = 1/n$
$a_n = h(b_n)$

$\to \xi = 0 \Rightarrow$ Gumbel i.e. inverse de la fonction de hasard constante.

Ex • $X \sim \mathcal{E}(1)$

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

$f(x) = e^{-x}$

$h(x) = \dfrac{1 - F(x)}{f(x)} = \dfrac{e^{-x}}{e^{-x}} = 1 \Rightarrow h'(x) = 0$

$h'(x) \xrightarrow[x \to x_F]{} \xi \Rightarrow \xi = 0$

$1 - F(b_n) = 1/n \Rightarrow e^{-b_n} = 1/n \Rightarrow b_n = -\ln(1/n) = \ln n$

$a_n = h(b_n) = 1$

$\dfrac{H_n - b_n}{a_n} \xrightarrow[n \to \infty]{\mathcal{L}} GEV(0, 1, \xi) \Rightarrow H - \ln n \xrightarrow{\mathcal{L}} GEV(0,1,0) \overset{\mathcal{L}}{=} \Lambda(x)$

• $X \sim \mathcal{N}(0,1)$

$\phi(x) = \dfrac{1}{\sqrt{2\pi}} \exp\left\{-\dfrac{x^2}{2}\right\} \quad x \in \mathbb{R}$

$\phi'(x) = -x\,\phi(x)$

$-\dfrac{\phi(x)}{\phi'(x)} = \dfrac{1}{x} \Leftrightarrow -\dfrac{\phi(x)}{\phi'(x)/x} = 1$

On cherche $h(x) = \dfrac{\overline{\Phi}(x)}{\phi'(x)}$

$\left(\dfrac{\phi(x)}{x}\right)' = \dfrac{\phi'(x)x - \phi(x)}{x^2} = -\dfrac{\phi(x)x^2 - \phi(x)}{x^2} = -\phi(x)\left(1 + \dfrac{1}{x^2}\right)$

$\dfrac{(\overline{\Phi}(x))'}{(\phi(x)/x)'} = \dfrac{-\phi(x)}{-\phi(x)\left(1 + \frac{1}{x^2}\right)} \xrightarrow[x \to x_F]{} 1$

L'Hopital : $\dfrac{\overline{\Phi}(x)}{\phi(x)/x} \xrightarrow[x \to x_F]{} \dfrac{(\overline{\Phi}(x))'}{(\phi(x)/x)'} \xrightarrow[x \to x_F]{} 1$

$\Rightarrow h(x) = \dfrac{\overline{\Phi}(x)}{\phi'(x)} \sim \dfrac{1}{x}$

$\Rightarrow h'(x) = -\dfrac{1}{x^2} \xrightarrow[x \to x_F]{} 0 \Rightarrow \xi = 0$

- $X \sim \text{Pareto}(c, \alpha)$ $\quad \bar{F}(x) = \dfrac{c}{x^\alpha}$ $\quad f(x) = \dfrac{\alpha c}{x^{\alpha+1}}$

$$h(x) = \frac{\cancel{c}}{\cancel{x^\alpha}} \cdot \frac{x^{\cancel{\alpha+1}}}{\alpha \cancel{c}} = \frac{x}{\alpha}$$

$$h'(x) = \frac{1}{\alpha} = \xi$$

$$\bar{F}(b_n) = 1/n \Leftrightarrow \frac{c}{b_n^\alpha} = \frac{1}{n} \Leftrightarrow b_n^\alpha = c \cdot n \Leftrightarrow b_n = [c \cdot n]^{1/\alpha}$$

$$a_n = h(b_n) = \frac{[c \cdot n]^{1/\alpha}}{\alpha}$$

$$\frac{M_n - b_n}{a_n} \xrightarrow[n \to \infty]{\mathcal{L}} GEV(0, 1, \xi) \Rightarrow \frac{M_n - [c \cdot n]^{1/\alpha}}{[c \cdot n]^{1/\alpha} / \alpha} = \alpha \left[M_n (c \cdot n)^{1/\alpha} - 1 \right] \xrightarrow[n \to \infty]{\mathcal{L}} GEV(0, 1, 1/\alpha$$

Le domaine d'attraction de la distribution de Fréchet

- L fonction mesurable et positive $]0,\infty[$
  $\forall x>0 \quad \frac{L(tx)}{L(t)} \xrightarrow[t\to\infty]{} 1$: L est à variations lentes

- $F \in D(\Phi_\alpha)$ ssi $\bar{F}(x) = x^{-\alpha} L(x)$
  ↳ $d_n = 0$ et $c_n = F^{\leftarrow}(1-n^{-1})$ avec $F^{\leftarrow}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\}$

  Ex : Cauchy, Pareto, Loggamma, ...

Le domaine d'attraction de la loi de Gumbel

- $F \in D(\Lambda)$ ssi $\exists g>0 \quad \frac{\bar{F}(x+tg(x))}{\bar{F}(x)} \xrightarrow[x\to x^F]{} \exp(-t)$

  Un choix possible pour g : $g(x) = \frac{\int_x^{x^F} \bar{F}(t)\,dt}{\bar{F}(x)} = \mathbb{E}(X-x \mid X>x)$

  ↳ $d_n = F^{\leftarrow}(1-n^{-1})$ et $c_n = g(d_n)$

  Ex : Normale, Exponentielle, ...

$$\frac{\bar{F}(x+tg(x))}{\bar{F}(x)} = \frac{\mathbb{P}(X>x+tg(x))}{\mathbb{P}(X>x)} = \mathbb{P}(X>x+tg(x) \mid X>x)$$
$$= \mathbb{P}\left(\frac{X-x}{g(x)} > t \;\middle|\; X>x\right) \xrightarrow[x\to x^F]{} e^{-t}$$

- Si $X \in D(\Lambda) \Rightarrow \exists g$ tq $\frac{X-u}{g(u)} \mid X>u \xrightarrow[u\to x^F]{\mathcal{L}} \mathcal{E}(1)$

Le domaine d'attraction de la loi Weibull

- $F \in D(\Psi_\alpha)$ ssi (1) $x^F < \infty$
  (2) $\bar{F}(x^F - x^{-1}) = x^{\alpha} L(x)$

  ↳ on peut choisir $d_n = F^{\leftarrow}(1)$ et $c_n = x^F - F^{\leftarrow}(1-n^{-1})$

  Ex : Beta, uniforme, ...

Exemple de distributions qui n'appartiennent pas à un domaine d'attractio

(1) $F(x) = 1 - 1/\ln x \qquad x > e$
(2) $F(x) = 1 - 1/\ln\{\ln x\} \qquad x > e^e$

(1) $f(x) = + \dfrac{1/x}{[\ln x]^2} = \dfrac{1}{x(\ln x)^2}$

$h(x) = \dfrac{\bar{F}(x)}{f(x)} = \dfrac{1}{\cancel{\ln x}} \cdot x(\ln x)^{\cancel{2}} = x \ln x$

$h'(x) = \ln x + \cancel{x} \cdot \dfrac{1}{\cancel{x}} = 1 + \ln x \xrightarrow[x \to \infty]{} \infty$

La condition de convergence n'est pas satisfaite, la queue est trop épaisse.
Appliquer une normalisation non linéaire
$Y = \ln X$
$\mathbb{P}(Y > y) = \mathbb{P}(\ln X > y) = \mathbb{P}(X > e^y) = \dfrac{1}{\ln e^y} = \dfrac{1}{y}$

$\Rightarrow Y \sim \text{Pareto}(1,1)$ et $X = e^Y$

$f(y) = + \dfrac{1}{y^2}$

$h(y) = \dfrac{1}{\cancel{y}} \cdot y^{\cancel{2}} = y$

$h'(y) = 1 \Rightarrow \xi = 1$
$F_Y \in D(GEV(0,1,1)) = D(\Phi_1)$

Vitesse de convergence

La vitesse de convergence vers la distribution limite donne une indication quant à la possibilité de considérer la distribution limite en pratique (lorsque le nb de données peut être faible).

$$e_n = \sup_{x} \left| [F(a_n x + b_n)]^n - G(x) \right|$$

Approximations

- $h'(y) \sim \xi$ pour $y > u = b_n$ $\rightarrow$ Erreur $= h'(b_n) - \xi$
- $n\,[1 - F(a_n x + b_n)] \sim -\log[F(a_n x + b_n)]^n$ $\rightarrow$ Erreur $\sim O(n^{-1})$

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

- $\mathcal{E}(1)$

  On avait $h'(x) = 0$ et $\xi = 0$ $\Rightarrow O(e_n) = O(1/n)$

Convergence des statistiques d'ordre

- (u_n) tq $n\bar{F}(u_n) \xrightarrow[n\to\infty]{} \tau$

  Alors $N_n = \sum_{i=1}^{n} \mathbb{1}_{\{X_i > u_n\}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{P}(\tau)$

  $$\lim_{n\to\infty} \mathbb{P}(X_{(k)} \leqslant u_n) = \lim_{n\to\infty} \mathbb{P}(N_n \leqslant k) = e^{-\tau} \sum_{i=0}^{k-1} \frac{\tau^i}{i!}$$

  Si $u_n = a_n x + b_n$ et $\tau = -\ln G(x)$, G GEV

  Alors $$\lim_{n\to\infty} \mathbb{P}\left(\frac{X_{(k)} - b_n}{a_n} \leqslant x\right) = G(x) \sum_{i=0}^{k-1} \frac{(-\ln G(x))^i}{i!}$$

1.2 Les dépassements de seuil $N_n = \sum_{i=1}^{n} \mathbb{1}_{\{X_i > u_n\}}$

- Distribution de Pareto généralisée $GPD(\beta, \xi)$

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

où $x \geq 0$ $\quad \xi \geq 0$

$0 \leqslant x \leqslant -\beta/\xi \quad \xi < 0$

· $\xi = -1 \Rightarrow G_{-1,\beta} \overset{\mathcal{L}}{=} \mathcal{U}([0,\beta])$

$\xi = 0 \Rightarrow G_{0,\beta} \overset{\mathcal{L}}{=} \mathcal{E}(1/\beta)$

· $X \sim G_{\xi,\beta} \Rightarrow f(x) = \frac{1}{\beta}\left[1 + \xi\left(\frac{x}{\beta}\right)\right]_+^{-1/\xi - 1}$

· Si $U \sim \mathcal{U}([0,1])$

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

· Si $X \sim GPD(\beta, \xi)$, $\xi < 1$

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

· Si $X \sim GPD(\beta, \xi)$

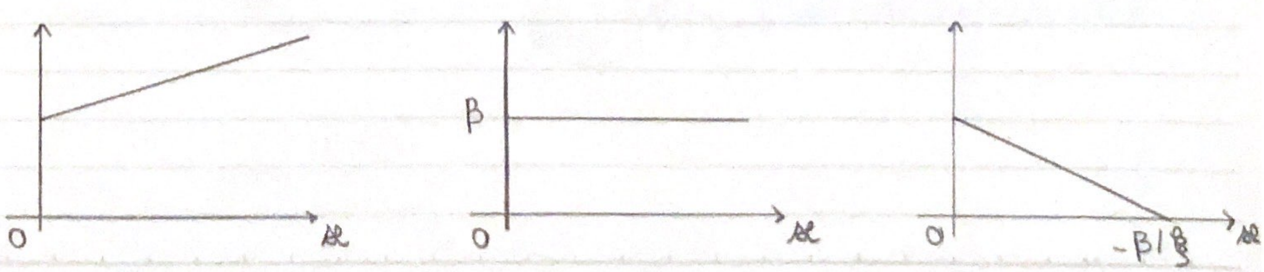
Alors $X - x \mid X > x \sim GPD(\beta + \xi x, \xi)$

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

$$\mathbb{P}(X - x > u \mid X > u) = \frac{\mathbb{P}(X > x + u)}{\mathbb{P}(X > u)} = \left[1 + \xi\left(\frac{x+u}{\beta}\right)\right]^{-1/\xi} \cdot \left[1 + \xi\left(\frac{u}{\beta}\right)\right]^{1/\xi}$$

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

$0 < \xi < 1$ $\qquad\qquad \xi = 0$ $\qquad\qquad \xi < 0$



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

$U(t) = F^{\leftarrow}(1-t^{-1}) \quad F^{\leftarrow}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\}$

Théoreme qui explique que la convergence du max de va iid correctement normalisée est équivalente à la convergence de la distribution des dépassements vers une GPD :

- $\xi \in \mathbb{R}$

$$\exists a_n, b_n \text{ tq } [F(a_n x + b_n)]^n \longrightarrow G_\xi(x)$$

$$\Leftrightarrow \exists a(\cdot) \text{ tq } \frac{\bar{F}(u + x a(u))}{\bar{F}(u)} \xrightarrow[u \to x^F]{} \begin{cases} [1+\xi x]_+^{-1/\xi} & \xi \neq 0 \\ e^{-x} & \xi = 0 \end{cases}$$

$$\Leftrightarrow x, y > 0, y \neq 1 \quad \frac{U(sx) - U(s)}{U(sy) - U(s)} \xrightarrow[s \to \infty]{} \begin{cases} (x^\xi - 1)/(y^\xi - 1) & \xi \neq 0 \\ \log x / \log y & \xi = 0 \end{cases}$$

$$\Leftrightarrow \exists \beta(\cdot) > 0 \text{ tq } \sup_{0 < x < x^F - u} |F_u(x) - G_{\xi, \beta(u)}(x)| \xrightarrow[u \to x^F]{} 0$$

$\exists a_n, b_n$ tq $[F(a_n x + b_n)]^n \to G_\xi(x)$

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

$$\Leftrightarrow \frac{M_n - b_n}{a_n} \xrightarrow[n \to \infty]{\mathcal{L}} GEV(0, 1, \xi)$$

$$\frac{\bar{F}(u + x a(u))}{\bar{F}(u)} = \mathbb{P}(X > u + x a(u) \mid X > u) = \mathbb{P}\left(\frac{X-u}{a(u)} > x \,\middle|\, X > u\right)$$

$$\xrightarrow[u \to x^F]{} [1+\xi x]_+^{-1/\xi} \Rightarrow \frac{X-u}{a(u)} \Big| X > u \xrightarrow[u \to x^F]{\mathcal{L}} GPD(1, \xi)$$

Si h existe $\quad b_n = F^{-1}(1 - 1/n)$

$a_n = h(b_n)$

$a(u) = h(u)$

Dynamique des dépassements de seuil

- Processus ponctuel N
  ~ distribution aléatoire de points X_i dans un espace E
  caractérisé par les distributions du nb de points pour tout ensemble A de l'espace
  $N(A)$ compte le nb de X_i dans A

$(x_i)_{i\geq 1}$ suite donnée de E
$m(A) = \sum_{i=1}^{\infty} \mathbb{1}_{\{x_i \in A\}}$: mesure de comptage
ponctuelle si $m(K) < \infty \quad \forall K \subset E$

$M_p(E)$: espace des mesures ponctuelles sur E

Un processus ponctuel sur E est une fonction mesurable
$N : [\Omega, \mathcal{F}, P] \rightarrow [M_p(E), \mathcal{M}_p(E)]$
↳ σ-algèbre appropriée
~ ensemble $(N(A))_{A \in \mathcal{E}}$ où $\mathcal{E}$ σ-algèbre de E générée par les ouverts

Ex • Processus ponctuel des dépassements normalisés
$N_n^{(\tau)}(B) = \sum_{i=1}^{n} \mathbb{1}_{\{i/n \in B, x_i > u_n(\tau)\}}$
$\forall$ Borélien $B \subset E := (0,1]$ où $(u_n(\tau))$ suite de seuils déterministes

• Processus ponctuel des dépassements bi-dimensionnels
$N_n^{(\tau)}(B) = \sum_{i=1}^{n} \mathbb{1}_{\{(i/n, (X_i - b_n)/a_n) \in B\}}$
$\forall$ Borélien $B \subset E := (0,1] \times \mathbb{R}$

Les réalisations de N sont des mesures ponctuelles
→ loi de N caractérisée par la famille des mesures ponctuelles
$P_N(A) = P(N \in A)$, $A \in \mathcal{M}_p(E)$
→ distribution de N déterminée par la distribution des vecteurs
$(N(A_1), \dots, N(A_m)) \quad \forall A_1, \dots, A_m$, $m \geq 1$
→ distribution de N caractérisée par sa transformée de Laplace :
$L_N(g) = E \exp\{-\int_E g \, dN\} = \int_{M_p(E)} \exp\{-\int_E g(x) \, dm(x)\} \, dP_N(m)$

- N est un processus de Poisson de mesure μ
  Si (i) $\forall A \in \mathcal{E}$, $N(A) \sim \mathcal{P}(\mu(A))$
  (ii) si $A_1, \dots, A_m$ ensembles mutuellement disjoints
  alors $N(A_1), \dots, N(A_m)$ sont mutuellement indépendants

Transformée de Laplace d'un processus de Poisson de mesure moyenne μ
$$L_N(g) = \exp\left\{-\int_E (1-e^{-g(x)})\, d\mu(x)\right\}$$

- N est un processus de Poisson composé de mesure μ
  et de taille des groupes π
  Si (i) $\forall A \in \mathcal{E}$, $N(A) \sim \mathcal{P}(\mu(A))$
  et de taille des groupes π
  (ii) "

$$L_N(g) = \exp\left\{-\int_E (1-L(g(x)))\, d\mu(x)\right\}$$
$\uparrow$ transformée de Laplace de π

- $N, N_1, N_2, \dots$ processus ponctuels sur E
  (N_n) converge faiblement vers N dans $M_p(E)$
  Si $\mathbb{P}((N_n(A_1), \dots, N_n(A_m))) \xrightarrow[n\to\infty]{} \mathbb{P}(N(A_1), \dots, N(A_m))$
  $\forall A_1, \dots, A_m$ tq $\mathbb{P}(N(\partial A_i) = 0) = 1$, $i = 1, \dots, m$, $m \geq 1$

- (N_n) converge faiblement vers N dans $M_p(E)$
  ssi la suite de transformées de Laplace converge $\forall g$ à support compact
  $L_{N_n}(g) \longrightarrow L_N(g)$

- $(u_n(\tau))$ tq $n\bar{F}(u_n(\tau)) \xrightarrow[n\to\infty]{} \tau$
  Alors $N_n^{(\tau)}$ converge faiblement vers un processus de Poisson ponctuel homogène N sur $(0,1]$ d'intensité $\tau\, |\cdot|$
  $\uparrow$ mesure de Lebesgue

- Soit $B = [0;1]$; $N_n([0,1]) \xrightarrow{\mathcal{L}} \mathcal{P}(\tau)$

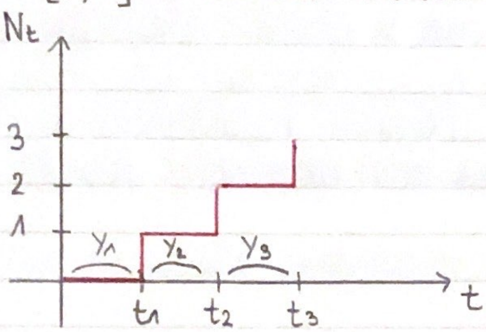
- Soit $B_1 = \left[0;\frac{1}{2}\right]$ et $B_2 = \left[\frac{1}{2};1\right]$

  $N_n(B_1) \xrightarrow{\mathcal{L}} \mathcal{P}(\tau/2)$ et $N_n(B_2) \xrightarrow{\mathcal{L}} \mathcal{P}(\tau/2)$

  ⇒ la proba. d'occurence dépend de l'intervalle

  $N_n(B_1)$ indépendant de $N_n(B_2)$

- $B_t = [0,t]$ $N_n(B_t) \xrightarrow{\mathcal{L}} \mathcal{P}(t)$



$Y_i = t_i - t_{i-1}$

$Y_i \overset{iid}{\sim} \mathcal{E}(\tau)$

1.3 Estimations des paramètres des lois des extrêmes

Le choix des méthodologies d'estimation dépend du type de données :
- max par blocs
- donnée de seuils
- record
- plus haute statistique d'ordre
- toutes les données

- Modèles statistiques compatibles avec le processus physique qui a généré les données
- Utilisation de statistiques capable d'évaluer l'ajustement du modèle
- Distributions des valeurs extrêmes bien approchées par les formes limites

Analyse exploratoire des données

- Probabilités et graphiques quantiles-quantiles (QQ plots)

$$(F(X_{(i)}))_{i=1,\dots,n} \overset{\mathcal{L}}{=} (U_{(i)})_{i=1,\dots,n}$$

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

→ comparaison empirique / théorique
courbe ~ linéaire (m̃ avec une transformation linéaire)

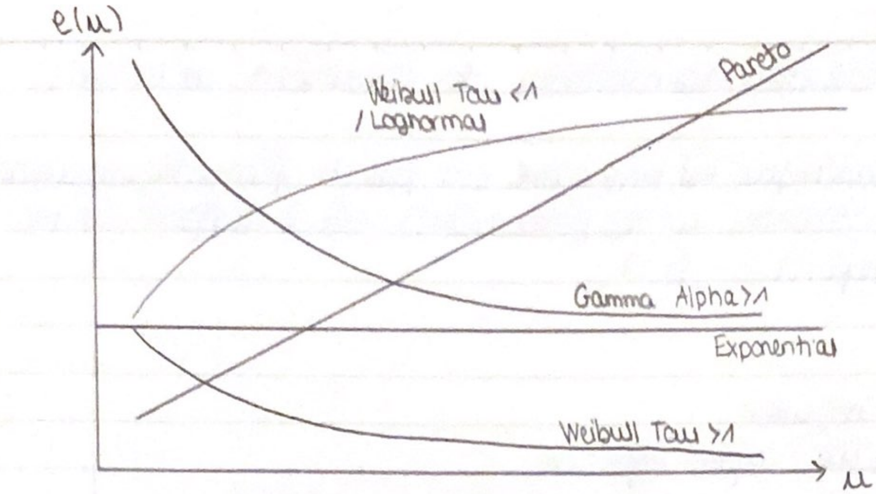
→ queues de distribution
queue épaisse (tendance à avoir des valeurs plus grandes pour une proba donnée)
↳ courbe incurvée vers le bas à gauche et/ou vers le haut à droite

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

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

$$e_n(u) = \frac{\sum_{i\in\Delta_n(u)} (X_i-u)}{\text{card}\,\Delta_n(u)} \qquad \Delta_n(u) = \{i : i=1,\dots,n,\ X_i>u\}$$

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



Estimation des paramètres de la distribution GEV

GEV : lorsqu'on dispose de données du type max par blocs

- Choix du modèle probabiliste

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

$\mathbb{P}(M_n \leq x) \sim GEV(\mu, \sigma, \xi)$

Si $F \in D(GEV(\mu, \sigma, \xi))$, $\exists\, a_n, b_n$ tq

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = \exp\left\{-\left[1 + \xi x\right]_+^{-1/\xi}\right\}$$

$(a_n), (b_n)$ dépendent de F mais à n fixé ce sont des constantes de normalisation :

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

$$y = a_n x + b_n \quad , \quad \sigma = a_n \quad , \quad \mu = b_n$$

Hypothèse : loi limite adaptée pour modéliser la loi à distance finie

→ choix de n : les blocs de max doivent être de taille suffisament grande

→ $n' \sim$ constante : ça dépend de F

- Estimation par max de vraisemblance des paramètre de la GEV

- une maximisation numérique est nécessaire car pas de formule analytique
- pour certains cas, contraintes sur les paramètres car le support de la distribution peut dépendre de θ.

Smith :

- $\xi > -1/2$: EMV régulier
- $-1 < \xi < -1/2$: EMV super-efficace
- $\xi < -1$: estimateur du point extrêmal = max(obs)

- Niveau et période de retour

Quelle est la probabilité pour un processus donné de dépasser un niveau x dans une période de temps future ?
Quel est le niveau x tel que cette probabilité soit suffisamment petite ?

- Période de retour pour un niveau x : temps d'attente moyen pour que le niveau x soit à nouveau dépassé.
- Niveau de retour de période T : niveau pour lequel le temps d'attente moyen de dépassement est T en années

Données iid :

- niveau de retour = quantiles
- période de retour = inverse des proba. de dépassement

$1 - F(x_p) = p$: le niveau de retour x_p a une période de retour de p^{-1} obs

- 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 GEV.

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

$V(\hat{z}_p) = \nabla z_p^{\prime}\, V\, \nabla z_p$ 	V : matrice de variance-covariance de $\hat{\theta}$

∇z_p : vecteur des dérivés // μ, σ, ξ, évalué pour $\theta = \hat{\theta}$