Questions de cours

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
2018 RABAT								
2018 ENSAE								
2017								
2017 S2								
2017 ENSAE								
2016								
2016 ENSAE								
2015								
2014								
2013								
2013 S2								
2012								
2012 S2								

Q1

Soit B une variable aléatoire de distribution Binomiale(n, p)

$$P(B = k) = C_n^k p^k (1-p)^{n-k}$$

et N une variable aléatoire de distribution Poisson(λ). On pose $\lambda = np$. Montrer que

$$P(B = k) = P(N = k)\left(\frac{\left(1-\frac{\lambda}{n}\right)^n}{e^{-\lambda}}\frac{1}{\left(1-\frac{\lambda}{n}\right)^k}\prod_{i=1}^{k}\left(1-\frac{i-1}{n}\right)\right).$$

En déduire que si $n \to \infty$, $p \to 0$ tel que $\lambda = np$, alors B converge en distribution vers N.

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

Q2

Qu'est-ce que la propriété de max-stabilité? Soit X une variable aléatoire de loi exponentielle standard. Est-ce que X ou $-X$ a une loi max-stable? Si oui, donner les coefficients permettant de déduire la max-stabilité.

Q3

On rappelle que la distribution de Pareto généralisée GPD(β, ξ) est définie par

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

Expliquer à quel domaine d'attraction appartient la distribution de Pareto généralisée GPD(β, ξ).

Q4

Expliquer ce que veut dire "F appartient au domaine d'attraction d'un loi max-stable G ($F \in D(G)$)". On considère la variable aléatoire de distribution H telle que

$$H(x) = F(\alpha x + \beta)$$

avec $\alpha > 0$. A quelle domaine d'attraction appartient H. Donner les relations qui existent entre les suites de normalisation pour H et F.

Supposons que F a une densité positive f et un point extrémal infini x^F telle que

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

A quel domaine d'attraction appartient F ?

Q5

Soit $X_1, ..., X_n$ une suite de variables aléatoires IID de fonction de distribution F et soit $M_n = \max(X_1, ..., X_n)$. On rappelle que:

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

$$\begin{array}{ll} 1) & \Pr(M_n \le u_n) \to e^{-\tau}, \\ 2) & n\bar{F}(u_n) = n(1 - F(u_n)) \to \tau. \end{array}$$

ii) Il existe une suite (u_n) satisfaisant $n\bar{F}(u_n) \to \tau$ quand $n \to \infty$ si et seulement si

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

où $x^F = \sup\{x : F(x) < 1\}$. Peut-on trouver une suite (u_n) telle que $\lim_{n\to\infty} \Pr(M_n \le u_n) = e^{-\tau}$ si

- les X_i ont une loi Exponentielle de paramètre λ

$$\Pr(X_i > x) = \exp(-\lambda x), \quad x > 0;$$

- les X_i ont une loi Géométrique de paramètre p

$$\Pr(X_i = n) = p(1-p)^n, \quad n \in \mathbb{N};$$

- les X_i ont une loi de Poisson de paramètre λ

$$\Pr(X_i = n) = e^{-\lambda} \frac{\lambda^n}{n!}, \quad n \in \mathbb{N}.$$

Q6

1. On rappelle que les lois "max-stable" sont définies par

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

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

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

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

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

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

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

Q7

On suppose que deux distributions F et G ont le même point extrêmal ($x^F = x^G$) et

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

Montrer que F et G appartiennent au même domaine d'attraction (disons celui de la GEV$(0,1,\xi)$) et donner le lien entre les constantes de normalisation. On rappelle que la fonction de répartition d'une GEV$(0,1,\xi)$ est

$$\exp\left(-(1+\xi x)_+^{-\frac{1}{\xi}}\right), \quad x \in \mathbb{R}.$$

Réciproquement supposons que F et G ont le même point extrêmal et appartiennent au domaine d'attraction de la Gumbel ($\Lambda(x) = \exp(-(\exp(-x)))$ telles qu'il existe $c_n > 0$ et d_n vérifiant

$$\lim_{n\to\infty} F^n(c_n x + d_n) = \Lambda(x) \quad \text{et} \quad \lim_{n\to\infty} G^n(c_n x + d_n) = \Lambda(x+b).$$

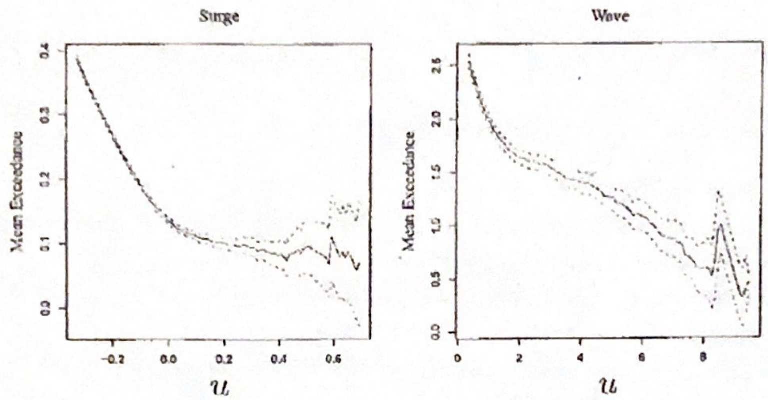
Montrer que

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

et caractériser 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?



Questions de cours

$C_n^k = \binom{n}{k} = \frac{n!}{(n-k)!\,k!}$

Q1) $B \sim Bin(n,p)$ $\quad \mathbb{P}(B=k) = C_n^k\, p^k (1-p)^{n-k}$

$N \sim \mathcal{P}(\lambda = np)$

Mq $\quad \mathbb{P}(B=k) = \mathbb{P}(N=k) \times \underbrace{\left(\frac{(1-\frac{\lambda}{n})^n}{e^{-\lambda}} \, \frac{1}{(1-\frac{\lambda}{n})^k} \prod_{i=1}^{k} \left(1 - \frac{i-1}{n}\right)\right)}_{A(n,p)}$

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

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

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

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

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

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

$$\frac{n!}{(n-k)!} = \frac{\prod_{i=1}^{n} i}{\prod_{i=1}^{n-k} i} = \frac{\cancel{\prod_{i=1}^{n-k} i} \times \prod_{i=n-k+1}^{n} i}{\cancel{\prod_{i=1}^{n-k} i}} = \prod_{i=n-k+1}^{n} i = \prod_{j=1}^{k} n-j+1$$

$$\Rightarrow \mathbb{P}(N=k) = \frac{1}{k!}\cdot\frac{n!}{(n-k)!}\left(\frac{\cancel{n}p}{\cancel{n}}\right)^k\left(1-\frac{\cancel{n}p}{\cancel{n}}\right)^{n-k} = C_n^k\, p^k (1-p)^{n-k} \quad //$$

$$A(n,p) = \underbrace{\overbrace{\frac{(1-p)^n}{e^{-np}}}^{\to 1}}_{\to 1} \; \underbrace{\frac{1}{(1-p)^k}}_{\to 1} \; \underbrace{\prod_{i=1}^{k}\left(1-\frac{i-1}{n}\right)}_{\to 1} \xrightarrow[\substack{n\to\infty \\ p\to 0}]{} 1$$

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

Si on pose $\quad B_n(x) = \sum_{i=1}^{n} \mathbb{1}\{X_i > x\}$

Alors $\quad B_n(x) \sim Bin(n, p = \mathbb{P}(X_i > x)) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{P}(np)$

$X_{(1)} \geq X_{(2)} \geq \ldots \geq X_{(n)}$

$$\mathbb{P}(X_{(k)} \leq x) = 1 - \mathbb{P}(X_{(k)} > x) = 1 - \mathbb{P}(B_n(x) \geq k) = 1 - [1 - \mathbb{P}(B_n(x) < k)]$$

$$= \mathbb{P}(B_n(x) < k) \xrightarrow[n\to\infty]{} \mathbb{P}(N < k) \quad //$$

Q2) Qu'est-ce que la propriété de max-stabilité ?
$X \sim Exp(1)$, est-ce que X ou $-X$ a une loi max-stable ?
Si oui, donner les coefficients 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
$\frac{M_n - b_n}{a_n} \xrightarrow[n \to \infty]{\mathcal{L}} X$ //

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

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

$X \sim Exp(1) \Rightarrow F(x) = 1 - e^{-x}$ 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 $e^{-b_n} = 1/n \Rightarrow b_n = -\ln(1/n) = \ln n$
$a_n = h(b_n) = 1$ //

$F_{-X}(x) = \mathbb{P}(-X \leq x) = \mathbb{P}(X \geq -x) = 1 - \mathbb{P}(X \leq -x) = 1 - [1 - e^{-(-x)}] = e^{x}$
$f_{-X}(x) = e^{x}$
$h_{-X}(x) = \frac{1 - F_X(x)}{f_{-X}(x)} = \frac{1 - e^{x}}{e^{x}} = e^{-x} - 1$
$h'_{-X}(x) = -e^{-x} \xrightarrow[x \to \infty]{} -1 = \xi$

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

Q3) $GPD(\beta, \xi)$: $G^{p}_{\xi,\beta}(x) = \begin{cases} 1 - [1 + \xi(x/\beta)]_+^{-1/\xi} & \text{si } \xi \neq 0 \\ 1 - e^{-x/\beta} & \text{si } \xi = 0 \end{cases}$

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

- $\xi = 0$

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

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

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

$h'(x) = 0 \Rightarrow \xi = 0$

$\Rightarrow GPD(\beta, 0) \in D(GEV(0,1,0) = \Lambda)$ //

- $\xi \neq 0$

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

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

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

$= \beta + \xi x$

$h'(x) = \xi$

$\Rightarrow GPD(\beta, \xi) \in \begin{cases} D(\Phi) & \xi > 0 \\ D(\Psi) & \xi < 0 \end{cases}$ //

Q4) Expliquer ce que veut dire "F appartient au domaine d'attraction d'une loi max-stable G ($F \in D(G)$)".
$H(x) = F(\alpha x + \beta)$, $\alpha > 0$
A quel domaine d'attraction 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
$$\frac{(1+\beta x) f(x)}{\bar{F}(x)} \xrightarrow[x \to x^F]{} c$$
A quel domaine d'attraction appartient F?

$X_1, \dots, X_n \overset{iid}{\sim} F$ et $M_n = \max(X_1, \dots, X_n)$
$$\mathbb{P}\left(\frac{M_n - b_n}{a_n}\right) \xrightarrow[n \to \infty]{\mathcal{L}} G$$
$$Y, Y_1, \dots, Y_n \overset{iid}{\sim} G \Rightarrow \mathbb{P}\left(\frac{Y_n - c_n}{a_n}\right) \xrightarrow[n \to \infty]{\mathcal{L}} G \quad //$$

$X_1, \dots, X_n \overset{iid}{\sim} F$ et $M_n = \max(X_1, \dots, X_n)$
$$F \in D(G) \Rightarrow \mathbb{P}\left(\frac{M_n - a_n}{b_n} \leq x\right) \xrightarrow[n \to \infty]{\mathcal{L}} G$$
$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = \mathbb{P}(M_n \leq a_n x + b_n) = [F(a_n x + b_n)]^n = \left[H\left(\frac{a_n x + b_n - \beta}{\sigma}\right)\right]^n$$
$$\xrightarrow[n \to \infty]{\mathcal{L}} G \; // \quad \text{avec} \quad \tilde{a}_n = \frac{a_n}{\sigma} \quad \text{et} \quad \tilde{b}_n = \frac{b_n - \beta}{\sigma} \; //$$

$$h(x) = \frac{1 - F(x)}{f(x)} = \frac{\bar{F}(x)}{f(x)}$$
$$\lim_{x \to x^F} \frac{(1+\beta x) f(x)}{\bar{F}(x)} = c \Leftrightarrow \lim_{x \to x^F} \frac{1+\beta x}{h(x)} = c$$
Hospital
$$\Leftrightarrow \lim_{x \to x^F} \frac{\beta}{h'(x)} = c$$
$$\Leftrightarrow \lim_{x \to x^F} h'(x) = \frac{\beta}{c} \Rightarrow \xi = \frac{\beta}{c}$$

$$\Rightarrow F \in \begin{cases} D(\Phi) & \beta/c > 0 \\ D(\Psi) & \beta/c < 0 \\ D(\Lambda) & \beta/c = 0 \end{cases} \; //$$

Q5) $X_1, \dots, X_n \overset{iid}{\sim} F$ et $M_n = \max(X_1, \dots, X_n)$

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

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

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

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

(i) $X_i \sim Exp(\lambda)$: $\mathbb{P}(X_i > x) = \exp(-\lambda x)$, $x > 0$
(ii) $X_i \sim Geom(p)$: $\mathbb{P}(X_i = n) = p(1-p)^n$, $n \in \mathbb{N}$
(iii) $X_i \sim \mathcal{P}(\lambda)$: $\mathbb{P}(X_i = n) = e^{-\lambda}\dfrac{\lambda^n}{n!}$, $n \in \mathbb{N}$

(i) $\overline{F}(x) = \overline{F}(x^-) \Rightarrow \dfrac{\overline{F}(x)}{\overline{F}(x^-)} \xrightarrow[x\to x_F]{} 1$ //

(ii) $$\frac{\mathbb{P}(X_i = n)}{\mathbb{P}(X_i = n+1)} = \frac{\cancel{p(1-p)^n}}{\cancel{p}(1-p)^{\cancel{n}-1}} = 1-p \neq 0 \ //$$

(iii) $$\frac{\mathbb{P}(X_i = n)}{\mathbb{P}(X_i = n-1)} = \frac{\cancel{e^{-\lambda}}\,\lambda^{\cancel{n}}/n!}{\cancel{e^{-\lambda}}\,\lambda^{\cancel{n-1}}/(n-1)!} = \frac{\lambda}{n-1} \xrightarrow[n\to\infty]{} 0 \ //$$

Q6) 1. et 2. déjà traitées.

3. Quel lien relie les distributions GEV et GPD

$X \sim F$ avec $F \in D(GEV(\mu, \sigma, \beta))$

$B_n(x) = \sum_{i=1}^{n} \mathbb{1}_{\{X_i > x\}} \sim GPD(\beta, \xi)$

4. Cf cours

Q7) (1) $x^F = x^G$ et $\dfrac{\bar{F}(x)}{\bar{G}(x)} \xrightarrow[x \to x^F]{} c \in (0, \infty)$

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

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

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

$\Lambda(x) = \exp\{-\exp\{-x\}\}$

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

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

et caractériser c.

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

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

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

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

$[1+\xi(a_n x + b_n)]^{-1/\xi} = c\,[1+\xi(\tilde{a}_n x + \tilde{b}_n)]^{-1/\xi}$

$1+\xi(a_n x + b_n) = c^{-\xi}[1+\xi(\tilde{a}_n x + \tilde{b}_n)]$

$= [c^{-\xi} + \xi\tilde{b}_n] + \xi\,\tilde{a}_n c^{-\xi} x$

Identification $c^{-\xi} + \xi\tilde{b}_n = 1 \Rightarrow \tilde{b}_n = 1/\xi\,(1 - c^{-\xi})$

$\tilde{a}_n c^{-\xi} = a_n \Rightarrow \tilde{a}_n = c^{\xi} a_n$

//

12) $F^n(c_n x + d_n) \underset{n\to\infty}{\longrightarrow} \exp\{-\underbrace{\exp\{-x\}}_{\tau_1}\} \Leftrightarrow n\bar{F}(c_n x + d_n) \longrightarrow \tau_1 = \exp\{-x\}$

$G^n(c_n x + d_n) \underset{n\to\infty}{\longrightarrow} \exp\{-\exp\underbrace{\{-(x+b)\}}_{\tau_2}\} \Leftrightarrow n\bar{G}(c_n x + d_n) \longrightarrow \tau_2 = \exp\{-x-b\}$

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

Q8) La fonction de dépassement moyen empirique sert à identifier le seuil u à partir duquel les dépassements suivent une GPD. Lorsque la va suit une GPD de seuil u alors le graphique devient approximativement linéaire au-delà de ce seuil.