

Mélange de distributions des valeurs extrêmes généralisées

Pascal Alain Dkengne Sielenou

Travail en collaboration avec Stéphane Girard (INRIA)

Réunion de synchronisation du 12 janvier 2024

Rappels : Loi des valeurs extremes généralisées

- Soit X une variable aléatoire de fonction de répartition F .
- Soient X_1, \dots, X_n des copies indépendantes de X .
- Soit $M_n = \max\{X_1, \dots, X_n\}$, la suite des maximums de X .

Theorem

S'il existent des suites de constantes $a_n > 0$ et $b_n \in \mathbb{R}$ telles que

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left\{ \frac{M_n - b_n}{a_n} \leq x \right\} = \lim_{n \rightarrow +\infty} F^n(a_n x + b_n) = G(x) \quad (1)$$

pour une fonction non dégénérée G , alors G appartient à la famille des lois des valeurs extrêmes généralisées (GEV)

$$G(x) = G(x; \gamma, \sigma, \mu) = \exp \left\{ - \left[1 + \gamma \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}} \right\}, \quad (2)$$

définies sur $\{x \in \mathbb{R} : 1 + \gamma \left(\frac{x - \mu}{\sigma} \right) > 0\}$, où $\gamma \neq 0$, $\mu \in \mathbb{R}$, $\sigma > 0$.

Si F vérifie (1), on dit qu'il appartient au domaine d'attraction de la fonction G .

Definition (Mélange des distributions de probabilités)

Une loi de probabilités est dite **loi de mélange** si sa fonction de répartition est une **moyenne pondérée algébrique**, une **moyenne pondérée géométrique** ou une **moyenne pondérée harmonique** de plusieurs fonctions de répartition.

Exemple

Considérons une suite de p fonctions de répartition $F_j, j = 1, \dots, p$ et un vecteur $\omega = (\omega_1, \dots, \omega_p) \in [0, 1]^p$ tel que $\sum_{j=1}^p \omega_j = 1$. Les lois de mélange moyennes pondérées algébriquement, géométriquement et harmoniquement ont respectivement les formes (3), (4) et (5) ci-dessous

$$F_S(x; \omega) = \sum_{j=1}^p \omega_j F_j(x), \quad (3)$$

$$F_P(x; \omega) = \prod_{j=1}^p F_j^{\omega_j}(x), \quad (4)$$

$$F_H(x; \omega) = \left(\sum_{j=1}^p \frac{\omega_j}{F_j(x)} \right)^{-1}. \quad (5)$$

Definition (Mélange de distributions GEV)

- Désignons par $\omega = (\omega_1, \dots, \omega_p) \in [0, 1]^p$ un vecteur tel que $\sum_{j=1}^p \omega_j = 1$.
- Désignons par G_j une fonction de répartition de la loi GEV définie par (2).
- Désignons par $\Theta = (\Theta_j, j = 1, \dots, p)$ où $\Theta_j = (\gamma_j, \sigma_j, \mu_j)$ un vecteur des paramètres de la distribution GEV nommée G_j .

On définit les modèles de mélange G_S , G_P et G_H des lois GEV nommées G_j par

$$\begin{aligned} G_S(x; \omega, \Theta) &= \sum_{j=1}^p \omega_j G_j(x; \Theta_j), \\ G_P(x; \omega, \Theta) &= \prod_{j=1}^p G_j^{\omega_j}(x; \Theta_j), \\ G_H(x; \omega, \Theta) &= \left(\sum_{j=1}^p \frac{\omega_j}{G_j(x; \Theta_j)} \right)^{-1}. \end{aligned} \tag{6}$$

Definition

Considérons une suite de p fonctions de répartition G_j de la loi GEV définie par

$$G_j(x) = G(x; \gamma_j, \sigma_j, \mu_j) = \exp \left\{ - \left[1 + \gamma_j \left(\frac{x - \mu_j}{\sigma_j} \right) \right]^{-\frac{1}{\gamma_j}} \right\}, \quad (7)$$

sur l'ensemble $\left\{ x \in \mathbb{R} : 1 + \gamma_j \left(\frac{x - \mu_j}{\sigma_j} \right) > 0 \right\}$, où $\gamma_j \neq 0$, $\mu_j \in \mathbb{R}$, $\sigma_j > 0$.

On définit les modèles de mélange G_{\min} , G_{\max} , G_{\inf} , et G_{\sup} par

$$G_{\min}(x) = \min_{1 \leq j \leq p} G_j(x), \quad (8)$$

$$G_{\max}(x) = \max_{1 \leq j \leq p} G_j(x), \quad (9)$$

$$G_{\inf}(x) = G \left(x; \max_{1 \leq j \leq p} \gamma_j, \max_{1 \leq j \leq p} \sigma_j, \max_{1 \leq j \leq p} \mu_j \right), \quad (10)$$

$$G_{\sup}(x) = G \left(x; \min_{1 \leq j \leq p} \gamma_j, \min_{1 \leq j \leq p} \sigma_j, \min_{1 \leq j \leq p} \mu_j \right). \quad (11)$$

Theorem (Inégalités entre les modèles des distributions GEV)

Les inégalités suivantes sont satisfaites

$$G_{\inf}(x) \leq G_{\min}(x) \leq G_H(x; \omega, \Theta) \leq G_P(x; \omega, \Theta) \leq G_S(x; \omega, \Theta) \leq G_{\max}(x) \leq G_{\sup}(x). \quad (12)$$

La preuve de ce résultat exploite les relations entre les moyennes arithmétiques, géométriques et harmoniques disponibles sur les pages web suivantes

https://en.wikipedia.org/wiki/Generalized_mean

Theorem (Stabilité de la famille $\{G_{\mathbb{P}}(\cdot; \omega, \Theta)\}$)

Pour tout entier positif m et pour tout réel x , la propriété suivante est satisfaite

$$G_{\mathbb{P}}^m(x; \omega, \Theta) = \prod_{j=1}^p [G_j(x; \Theta_j(m))]^{\omega_j} = G_{\mathbb{P}}(x; \omega, \Theta(m)). \quad (13)$$

Ici, $\Theta(m) = (\Theta_j(m), j = 1, \dots, p)$ où $\Theta_j(m) = (\gamma_j(m), \sigma_j(m), \mu_j(m))$

avec $\gamma_j(m) = \gamma_j$, $\sigma_j(m) = \sigma_j m^{\gamma_j}$, $\mu_j(m) = \mu_j + \sigma_j \left(\frac{m^{\gamma_j} - 1}{\gamma_j} \right)$.

La propriété (13) montre que si la loi d'une v.a. X appartient à la famille des lois de probabilités $\{G_{\mathbb{P}}(\cdot; \omega, \Theta)\}$, alors la loi du maximum de m copies indépendantes de X appartient également à cette même famille des lois de probabilités.

Propriétés 3

Theorem (Equivalence des queues de distributions)

On considère les distributions de probabilités F_S , F_P et F_H définies resp. par (3), (4) et (5). S'ils existent des constantes $\tau_j > 0$ telles que pour tout $j = 1, \dots, p$ nous avons

$$\lim_{x \rightarrow x_{F_j}} \frac{1 - F_j(x)}{1 - G_j(x; \Theta_j)} = \tau_j, \quad (14)$$

où $x_{F_j} = \sup\{x \in \mathbb{R} : F_j(x) < 1\}$ est la borne supérieure du support de la distribution F_j , alors les limites suivantes sont satisfaites

$$\lim_{x \rightarrow x^*} \frac{1 - F_H(x; \omega)}{1 - G_P(x; \omega, \Theta(\tau))} = 1, \quad \lim_{x \rightarrow x^*} \frac{1 - F_S(x; \omega)}{1 - G_P(x; \omega, \Theta(\tau))} = 1, \quad \lim_{x \rightarrow x^*} \frac{1 - F_P(x; \omega)}{1 - G_P(x; \omega, \Theta(\tau))} = 1, \quad (15)$$

où $x^* = \max\{x_{F_j}, j = 1, \dots, p\}$ et $\tau = (\tau_1, \dots, \tau_p)$.

Ici, $\Theta(\tau) = (\Theta_j(\tau_j), j = 1, \dots, p)$ où $\Theta_j(\tau_j) = (\gamma_j(\tau_j), \sigma_j(\tau_j), \mu_j(\tau_j))$

avec $\gamma_j(\tau_j) = \gamma_j$, $\sigma_j(\tau_j) = \sigma_j \tau_j^{\gamma_j}$, $\mu_j(\tau_j) = \mu_j + \sigma_j \left(\frac{\tau_j^{\gamma_j - 1}}{\gamma_j} \right)$.

Preuve de la Propriété 3 (1/3)

Quand x tend vers x^* , on peut effectuer les équivalents suivants

$$\begin{aligned} 1 - F_S(x) &= 1 - \sum_{j=1}^p \omega_j F_j(x) = \sum_{j=1}^p \omega_j (1 - F_j(x)) \\ &\sim \sum_{j=1}^p \omega_j \tau_j (1 - G_j(x; \Theta_j)), \quad (\text{faisant usage de (14)}) \\ &\sim - \sum_{j=1}^p \omega_j \log G_j^{\tau_j}(x; \Theta_j) = - \sum_{j=1}^p \log G_j^{\omega_j}(x; \Theta_j(\tau_j)) \\ &= - \log \prod_{j=1}^p G_j^{\omega_j}(x; \Theta_j(\tau_j)) \\ &\sim 1 - \prod_{j=1}^p G_j^{\omega_j}(x; \Theta_j(\tau_j)) \\ &= 1 - G_P(x; \omega, \Theta(\tau)). \end{aligned}$$

Preuve de la Propriété 3 (2/3)

Quand x tend vers x^* , on peut effectuer les équivalents suivants

$$\begin{aligned} 1 - F_{\mathbb{P}}(x) &\sim -\log F_{\mathbb{P}}(x) = -\log \prod_{j=1}^p F_j^{\omega_j}(x) \\ &= \sum_{j=1}^p \omega_j (-\log F_j(x)) \sim \sum_{j=1}^p \omega_j (1 - F_j(x)) \\ &\sim \sum_{j=1}^p \omega_j \tau_j (1 - G_j(x; \Theta_j)), \quad (\text{faisant usage de (14)}) \\ &\sim -\sum_{j=1}^p \omega_j \log G_j^{\tau_j}(x; \Theta_j) = -\sum_{j=1}^p \log G_j^{\omega_j}(x; \Theta_j(\tau_j)) \\ &= -\log \prod_{j=1}^p G_j^{\omega_j}(x; \Theta_j(\tau_j)) \sim 1 - \prod_{j=1}^p G_j^{\omega_j}(x; \Theta_j(\tau_j)) \\ &= 1 - G_{\mathbb{P}}(x; \omega, \Theta(\tau)). \end{aligned}$$

Preuve de la Propriété 3 (3/3)

Quand x tend vers x^* , on peut effectuer les équivalents suivants

$$\begin{aligned} 1 - F_H(x) &\sim -\log F_H(x) = \log \sum_{j=1}^p \frac{\omega_j}{F_j(x)} \sim \sum_{j=1}^p \omega_j \left(\frac{1}{F_j(x)} - 1 \right) \\ &\sim \sum_{j=1}^p \omega_j (-\log F_j(x)) \sim \sum_{j=1}^p \omega_j (1 - F_j(x)) \\ &\sim \sum_{j=1}^p \omega_j \tau_j (1 - G_j(x; \Theta_j)), \quad (\text{faisant usage de (14)}) \\ &\sim -\sum_{j=1}^p \omega_j \log G_j^{\tau_j}(x; \Theta_j) = -\sum_{j=1}^p \log G_j^{\omega_j}(x; \Theta_j(\tau_j)) \\ &= -\log \prod_{j=1}^p G_j^{\omega_j}(x; \Theta_j(\tau_j)) \sim 1 - \prod_{j=1}^p G_j^{\omega_j}(x; \Theta_j(\tau_j)) \\ &= 1 - G_P(x; \omega, \Theta(\tau)). \end{aligned}$$

Theorem

- On considère une suite de p fonctions de répartition $F_j, j = 1, \dots, p$.
- On considère un vecteur $\omega = (\omega_1, \dots, \omega_p) \in [0, 1]^p$ tel que $\sum_{j=1}^p \omega_j = 1$.
- On désigne par $x^* = \max \{x_{F_1}, x_{F_2}, \dots, x_{F_p}\}$, où $x_{F_j} \leq +\infty$ la borne supérieure du support de la distribution F_j .

Alors, les lois de mélange moyennes pondérées ci-dessous

$$F_S(x; \omega) = \sum_{j=1}^p \omega_j F_j(x), \quad F_P(x; \omega) = \prod_{j=1}^p F_j^{\omega_j}(x), \quad F_H(x; \omega) = \left(\sum_{j=1}^p \frac{\omega_j}{F_j(x)} \right)^{-1}$$

satisfont la propriété suivante :

$$\lim_{x \rightarrow x^*} F_S(x; \omega) = \lim_{x \rightarrow x^*} F_P(x; \omega) = \lim_{x \rightarrow x^*} F_H(x; \omega). \quad (16)$$

Preuve de la Propriété 4 (1/2)

$$\begin{aligned}\lim_{x \rightarrow x^*} F_S(x; \omega) &= \lim_{x \rightarrow x^*} \sum_{j=1}^p \omega_j F_j(x) = \lim_{x \rightarrow x^*} \exp \left\{ \log \left(\sum_{j=1}^p \omega_j F_j(x) \right) \right\} \\&= \lim_{x \rightarrow x^*} \exp \left\{ \sum_{j=1}^p \omega_j F_j(x) - 1 \right\} = \lim_{x \rightarrow x^*} \exp \left\{ \sum_{j=1}^p \omega_j (F_j(x) - 1) \right\} \\&= \lim_{x \rightarrow x^*} \exp \left\{ \sum_{j=1}^p \omega_j \log F_j(x) \right\} = \lim_{x \rightarrow x^*} \exp \left\{ \log \left(\prod_{j=1}^p F_j^{\omega_j}(x) \right) \right\} \\&= \lim_{x \rightarrow x^*} \prod_{j=1}^p F_j^{\omega_j}(x) = \lim_{x \rightarrow x^*} F_P(x; \omega).\end{aligned}$$

Preuve de la Propriété 4 (2/2)

$$\begin{aligned}\lim_{x \rightarrow x^*} F_H(x; \omega) &= \lim_{x \rightarrow x^*} \left(\sum_{j=1}^p \frac{\omega_j}{F_j(x)} \right)^{-1} = \lim_{x \rightarrow x^*} \exp \left\{ -\log \left(\sum_{j=1}^p \frac{\omega_j}{F_j(x)} \right) \right\} \\&= \lim_{x \rightarrow x^*} \exp \left\{ 1 - \sum_{j=1}^p \frac{\omega_j}{F_j(x)} \right\} = \lim_{x \rightarrow x^*} \exp \left\{ \sum_{j=1}^p \omega_j \left(1 - \frac{1}{F_j(x)} \right) \right\} \\&= \lim_{x \rightarrow x^*} \exp \left\{ \sum_{j=1}^p \omega_j \log F_j(x) \right\} = \lim_{x \rightarrow x^*} \exp \left\{ \log \left(\prod_{j=1}^p F_j^{\omega_j}(x) \right) \right\} \\&= \lim_{x \rightarrow x^*} \prod_{j=1}^p F_j^{\omega_j}(x) = \lim_{x \rightarrow x^*} F_P(x; \omega).\end{aligned}$$

Theorem

On considère les distributions de probabilités F_S et F_P définies resp. par (3) et (4). S'ils existent des suites $a_{n,j} > 0$ et $b_{n,j} \in \mathbb{R}$ telles que pour tout $x \in \mathbb{R}$, nous avons

$$\lim_{n \rightarrow +\infty} n [1 - F_j(a_{n,j}x + b_{n,j})] = -\log G_j(x) \Leftrightarrow \lim_{n \rightarrow +\infty} F_j^n(a_{n,j}x + b_{n,j}) = G_j(x), \quad (17)$$

alors les limites suivantes sont satisfaites pour tout $x \in \mathbb{R}$:

$$\lim_{n \rightarrow +\infty} F_P^n(v_n(x)) \leq G_P(x) \leq \lim_{n \rightarrow +\infty} F_P^n(u_n(x)), \quad (18)$$

$$\lim_{n \rightarrow +\infty} F_S^n(v_n(x)) \leq G_P(x) \leq \lim_{n \rightarrow +\infty} F_S^n(u_n(x)), \quad \lim_{n \rightarrow +\infty} F_H^n(v_n(x)) \leq G_P(x) \leq \lim_{n \rightarrow +\infty} F_H^n(u_n(x)),$$

où les suites $u_n(x)$ et $v_n(x)$ sont définies par

$$u_n(x) = \max_{1 \leq j \leq p} \{a_{n,j}x + b_{n,j}\}, \quad v_n(x) = \min_{1 \leq j \leq p} \{a_{n,j}x + b_{n,j}\}. \quad (19)$$

De plus, il existe trois suites non linéaires $p_n(x), s_n(x), h_n(x) \in [v_n(x), u_n(x)]$ qui sont strictement croissantes et satisfont

$$\lim_{n \rightarrow +\infty} F_P^n(p_n(x)) = G_P(x) = \lim_{n \rightarrow +\infty} F_S^n(s_n(x)) = \lim_{n \rightarrow +\infty} F_H^n(h_n(x)). \quad (20)$$

Preuve de la Propriété 5

$$\begin{aligned}
 \lim_{n \rightarrow +\infty} F_{\mathbb{P}}^n(v_n(x); \omega) &= \lim_{n \rightarrow +\infty} \left(\prod_{j=1}^p F_j^{\omega_j}(v_n(x)) \right)^n \leq \lim_{n \rightarrow +\infty} \prod_{j=1}^p [F_j^n(a_{nj}x + b_{nj})]^{\omega_j} \\
 &= \prod_{j=1}^p G_j^{\omega_j}(x), \quad (\text{faisant usage de (17)}). \tag{21}
 \end{aligned}$$

$$\begin{aligned}
 \lim_{n \rightarrow +\infty} F_{\mathbb{P}}^n(u_n(x); \omega) &= \lim_{n \rightarrow +\infty} \left(\prod_{j=1}^p F_j^{\omega_j}(u_n(x)) \right)^n \geq \lim_{n \rightarrow +\infty} \prod_{j=1}^p [F_j^n(a_{nj}x + b_{nj})]^{\omega_j} \\
 &= \prod_{j=1}^p G_j^{\omega_j}(x), \quad (\text{faisant usage de (17)}). \tag{22}
 \end{aligned}$$

Faisant usage des égalités (16) et des inégalités (21-22), on peut écrire

$$\lim_{n \rightarrow +\infty} F_S^n(v_n(x); \omega) = \lim_{n \rightarrow +\infty} F_H^n(v_n(x); \omega) = \lim_{n \rightarrow +\infty} F_P^n(v_n(x); \omega) \leq G_P(x; \omega),$$

$$\lim_{n \rightarrow +\infty} F_S^n(u_n(x); \omega) = \lim_{n \rightarrow +\infty} F_H^n(u_n(x); \omega) = \lim_{n \rightarrow +\infty} F_P^n(u_n(x); \omega) \geq G_P(x; \omega).$$

Theorem

- On considère une suite de p fonctions de répartition $F_j, j = 1, \dots, p$.
- On considère un vecteur $\omega = (\omega_1, \dots, \omega_p) \in [0, 1]^p$ tel que $\sum_{j=1}^p \omega_j = 1$.

La variable aléatoire X associée à la fonction de répartition $F_P(x) = \prod_{j=1}^p F_j^{\omega_j}(x)$ est définie par

$$X = \max_{1 \leq j \leq p} \{F_j^{-1}(\exp\{-Y_j\})\}, \quad (23)$$

où Y_j est une variable aléatoire qui suit la loi exponentielle de paramètre ω_j .

Simulation

La formule (23) suggère les deux étapes suivantes pour simuler une observation x de la distribution F_P .

- 1 Générer les valeurs indépendantes $y_j, j = 1, 2, \dots, p$ à l'aide de la loi exponentielle de paramètre ω_j .
- 2 Prendre comme valeur simulée de la distribution F_P la quantité x définie par

$$x = \max_{1 \leq j \leq p} \{F_j^{-1}(\exp\{-y_j\})\}. \quad (24)$$

Preuve de la Propriété 6

Soit X_j la variable aléatoire associée à la distribution $F_j^{\omega_j}$.

La variable aléatoire $U_j = F_j^{\omega_j}(X_j)$ suit la loi uniforme sur l'intervalle $[0, 1]$.

Ainsi $X_j = F_j^{-1}(U_j^{1/\omega_j}) = F_j^{-1}\left(\exp\left\{\frac{1}{\omega_j} \log U_j\right\}\right)$.

La variable aléatoire $Y_j = -\frac{1}{\omega_j} \log U_j$ suit la loi exponentielle de paramètre ω_j .

En effet pour tout réel $y > 0$, on a :

$$\begin{aligned}\mathbb{P}\{Y_j \leq y\} &= \mathbb{P}\left\{-\frac{1}{\omega_j} \log U_j \leq y\right\} \\ &= \mathbb{P}\{U_j > \exp\{-\omega_j y\}\} \\ &= 1 - \exp\{-\omega_j y\}.\end{aligned}$$

Pour tout $x \in \mathbb{R}$, on a :

$$\begin{aligned}\mathbb{P}\{X \leq x\} &= \prod_{j=1}^p \mathbb{P}\{X_j \leq x\} \\ &= \mathbb{P}\{X_1 \leq x, \dots, X_p \leq x\} \\ &= \mathbb{P}\{\max\{X_1, \dots, X_p\} \leq x\}.\end{aligned}$$

Theorem

Soit $\omega = (\omega_1, \dots, \omega_p) \in [0, 1]^p$ un vecteur de poids, c'est-à-dire $\sum_{j=1}^p \omega_j = 1$.
On considère les distributions de probabilités F_S , F_H et F_P définies par

$$F_S(x; \omega) = \sum_{j=1}^p \omega_j F_j(x), \quad F_H(x; \omega) = \left(\sum_{j=1}^p \frac{\omega_j}{F_j(x)} \right)^{-1}, \quad F_P(x; \omega) = \prod_{j=1}^p F_j^{\omega_j}(x).$$

Si $x_j(\alpha)$ désigne le quantile d'ordre $\alpha \in [0, 1]$ de la distribution F_j , i.e. $F_j(x_j(\alpha)) = 1 - \alpha$, alors les quantités $x_1(\alpha) = \min_{1 \leq j \leq p} \{x_j(\alpha)\}$, $x_2(\alpha) = \max_{1 \leq j \leq p} \{x_j(\alpha)\}$ satisfont pour tout vecteur de poids ω les inégalités suivantes :

$$F_S(x_1(\alpha); \omega) \leq 1 - \alpha \leq F_S(x_2(\alpha); \omega) \iff x_1(\alpha) \leq F_S^{-1}(1 - \alpha; \omega) \leq x_2(\alpha),$$

$$F_H(x_1(\alpha); \omega) \leq 1 - \alpha \leq F_H(x_2(\alpha); \omega) \iff x_1(\alpha) \leq F_H^{-1}(1 - \alpha; \omega) \leq x_2(\alpha),$$

$$F_P(x_1(\alpha); \omega) \leq 1 - \alpha \leq F_P(x_2(\alpha); \omega) \iff x_1(\alpha) \leq F_P^{-1}(1 - \alpha; \omega) \leq x_2(\alpha).$$

Preuve de la Propriété 7 (1/2)

Par hypothèse, pour tout $j = 1, \dots, p$ on a les deux égalités suivantes vraies

$$F_j^{\omega_j}(x_j(\alpha)) = (1 - \alpha)^{\omega_j}, \quad \omega_j F_j(x_j(\alpha)) = \omega_j(1 - \alpha), \quad \frac{\omega_j}{F_j(x_j(\alpha))} = \frac{\omega_j}{1 - \alpha}.$$

Par définition, pour tout $j = 1, \dots, p$ on a les inégalités suivantes :

$$F_j\left(\max_{1 \leq k \leq p} \{x_k(\alpha)\}\right) \geq F_j(x_j(\alpha)), \quad F_j\left(\min_{1 \leq k \leq p} \{x_k(\alpha)\}\right) \leq F_j(x_j(\alpha)).$$

Ce qui implique

$$F_j^{\omega_j}(x_2(\alpha)) \geq (1 - \alpha)^{\omega_j}, \quad \omega_j F_j(x_2(\alpha)) \geq \omega_j(1 - \alpha), \quad \frac{\omega_j}{F_j(x_2(\alpha))} \leq \frac{\omega_j}{1 - \alpha}.$$

$$F_j^{\omega_j}(x_1(\alpha)) \leq (1 - \alpha)^{\omega_j}, \quad \omega_j F_j(x_1(\alpha)) \leq \omega_j(1 - \alpha), \quad \frac{\omega_j}{F_j(x_1(\alpha))} \geq \frac{\omega_j}{1 - \alpha}.$$

Preuve de la Propriété 7 (2/2)

$$F_{\mathbb{P}}(x_2(\alpha)) = \prod_{j=1}^p F_j^{\omega_j}(x_2(\alpha)) \geq \prod_{j=1}^p (1 - \alpha)^{\omega_j} = (1 - \alpha)^{\sum_{j=1}^p \omega_j} = 1 - \alpha.$$

$$F_{\mathbb{P}}(x_1(\alpha)) = \prod_{j=1}^p F_j^{\omega_j}(x_1(\alpha)) \leq \prod_{j=1}^p (1 - \alpha)^{\omega_j} = (1 - \alpha)^{\sum_{j=1}^p \omega_j} = 1 - \alpha.$$

$$F_{\mathbb{S}}(x_2(\alpha)) = \sum_{j=1}^p \omega_j F_j(x_2(\alpha)) \geq (1 - \alpha) \sum_{j=1}^p \omega_j = 1 - \alpha.$$

$$F_{\mathbb{S}}(x_1(\alpha)) = \sum_{j=1}^p \omega_j F_j(x_1(\alpha)) \leq (1 - \alpha) \sum_{j=1}^p \omega_j = 1 - \alpha.$$

$$\frac{1}{F_{\mathbb{H}}(x_2(\alpha))} = \sum_{j=1}^p \frac{\omega_j}{F_j(x_2(\alpha))} \leq \sum_{j=1}^p \frac{\omega_j}{1 - \alpha} = \frac{1}{1 - \alpha}.$$

$$\frac{1}{F_{\mathbb{H}}(x_1(\alpha))} = \sum_{j=1}^p \frac{\omega_j}{F_j(x_1(\alpha))} \geq \sum_{j=1}^p \frac{\omega_j}{1 - \alpha} = \frac{1}{1 - \alpha}.$$

Propriétés 8

Theorem

Soit $\omega = (\omega_1, \dots, \omega_p) \in [0, 1]^p$ un vecteur de poids, c'est-à-dire $\sum_{j=1}^p \omega_j = 1$.
On considère les distributions de probabilités F_S , F_H et F_P définies par

$$F_S(x; \omega) = \sum_{j=1}^p \omega_j F_j(x), \quad F_H(x; \omega) = \left(\sum_{j=1}^p \frac{\omega_j}{F_j(x)} \right)^{-1}, \quad F_P(x; \omega) = \prod_{j=1}^p F_j^{\omega_j}(x)$$

de densités de probabilités respectives $g_S(\cdot; \omega)$, $g_H(\cdot; \omega)$, $g_P(\cdot; \omega)$, où F_j est une fonction de répartition de densité de probabilités f_j .

- Soit $x_j(\alpha) \in \mathbb{R}$ avec $\alpha \in (0, 1)$ tel que $F_j(x_j(\alpha)) = 1 - \alpha$.
- Soit $j_0 = \arg \max_{1 \leq j \leq p} \{x_j(\alpha)\}$, i.e. $x_{j_0}(\alpha) = \max_{1 \leq j \leq p} \{x_j(\alpha)\}$.

Alors la quantité $x(\alpha) \in \mathbb{R}$ définie par

$$x(\alpha) = x_{j_0}(\alpha) + \frac{\alpha}{f_{j_0}(x_{j_0}(\alpha))} - \frac{\alpha}{f_S(x_{j_0}(\alpha); \omega)}$$

est une fonction continue de α qui satisfait les conditions ci-dessous

$$\lim_{\alpha \rightarrow 0} F_S(x(\alpha); \omega) = \lim_{\alpha \rightarrow 0} F_H(x(\alpha); \omega) = \lim_{\alpha \rightarrow 0} F_P(x(\alpha); \omega) = 1 - \alpha.$$

Preuve de la Propriété 8 (1/2)

L'équation

$$F_S(x(\alpha); \omega) = 1 - \alpha$$

peut encore s'écrire de la manière suivante

$$F_S(x(\alpha); \omega) - F_S(x_{j_0}(\alpha); \omega) = 1 - \alpha - F_S(x_{j_0}(\alpha); \omega).$$

A l'aide du développement limité de Taylor à l'ordre 1 on en déduit que

$$[x(\alpha) - x_{j_0}(\alpha)] \cdot f_S(x_{j_0}(\alpha); \omega) \sim 1 - \alpha - F_S(x_{j_0}(\alpha); \omega)$$

au voisinage de $\alpha = 0$ car $\lim_{\alpha \rightarrow 0} x(\alpha) = \lim_{\alpha \rightarrow 0} x_{j_0}(\alpha)$.

Ainsi au voisinage de $\alpha = 0$ on a

$$x(\alpha) \sim x_{j_0}(\alpha) + \frac{1 - \alpha - F_S(x_{j_0}(\alpha); \omega)}{f_S(x_{j_0}(\alpha); \omega)} = x_{j_0}(\alpha) + \frac{1 - F_S(x_{j_0}(\alpha); \omega)}{f_S(x_{j_0}(\alpha); \omega)} - \frac{\alpha}{f_S(x_{j_0}(\alpha); \omega)}.$$

Preuve de la Propriété 8 (2/2)

$$\begin{aligned}
 \frac{1 - F_S(x_{j_0}(\alpha); \omega)}{f_S(x_{j_0}(\alpha); \omega)} &= \frac{\sum_{j=1}^p \omega_j [1 - F_j(x_{j_0}(\alpha); \omega)]}{\sum_{j=1}^p \omega_j f_j(x_{j_0}(\alpha); \omega)} \\
 &= \frac{1 - F_{j_0}(x_{j_0}(\alpha); \omega)}{f_{j_0}(x_{j_0}(\alpha); \omega)} \cdot \frac{\omega_{j_0} + \sum_{j \neq j_0}^p \omega_j \frac{1 - F_j(x_{j_0}(\alpha); \omega)}{1 - F_{j_0}(x_{j_0}(\alpha); \omega)}}{\omega_{j_0} + \sum_{j \neq j_0}^p \omega_j \frac{f_j(x_{j_0}(\alpha); \omega)}{f_{j_0}(x_{j_0}(\alpha); \omega)}} \\
 \text{(Hospital rule)} \quad &\sim \frac{1 - F_{j_0}(x_{j_0}(\alpha); \omega)}{f_{j_0}(x_{j_0}(\alpha); \omega)} \cdot \frac{\omega_{j_0} + \sum_{j \neq j_0}^p \omega_j \left(\frac{dx_{j_0}(\alpha)}{d\alpha} \right) \cdot \left(\frac{dx_{j_0}(\alpha)}{d\alpha} \right)^{-1} \cdot \frac{f_j(x_{j_0}(\alpha); \omega)}{f_{j_0}(x_{j_0}(\alpha); \omega)}}{\omega_{j_0} + \sum_{j \neq j_0}^p \omega_j \frac{f_j(x_{j_0}(\alpha); \omega)}{f_{j_0}(x_{j_0}(\alpha); \omega)}} \\
 &\sim \frac{1 - F_{j_0}(x_{j_0}(\alpha); \omega)}{f_{j_0}(x_{j_0}(\alpha); \omega)}.
 \end{aligned}$$

Corollary

Soit $\omega = (\omega_1, \dots, \omega_p) \in [0, 1]^p$ un vecteur de poids, c'est-à-dire $\sum_{j=1}^p \omega_j = 1$.

On considère les distributions de probabilités G_S , G_H et G_P définies par

$$G_S(x; \omega) = \sum_{j=1}^p \omega_j G_j(x), \quad G_H(x; \omega) = \left(\sum_{j=1}^p \frac{\omega_j}{G_j(x)} \right)^{-1}, \quad G_P(x; \omega) = \prod_{j=1}^p G_j^{\omega_j}(x)$$

de densités de probabilités respectives $g_S(\cdot; \omega)$, $g_H(\cdot; \omega)$, $g_P(\cdot; \omega)$, où G_j est une distribution GEV de paramètre $(\gamma_j, \sigma_j, \mu_j) \in \mathbb{R}^3$ et de densité g_j .

Soit $x_j(\alpha) = \mu_j + \sigma_j \left\{ \frac{[-\log(1-\alpha)]^{-\gamma_j} - 1}{\gamma_j} \right\} \in \mathbb{R}$ avec $\alpha \in (0, 1)$, i.e. $G_j(x_j(\alpha)) = 1 - \alpha$.

Soit $j_0 = \arg \max_{1 \leq j \leq p} \{x_j(\alpha)\}$, i.e. $x_{j_0}(\alpha) = \max_{1 \leq j \leq p} \{x_j(\alpha)\}$.

Alors la fonction $x(\alpha) \in \mathbb{R}$ définie par (25) croît simultanément avec γ_{j_0} , σ_{j_0} et μ_{j_0}

$$x(\alpha) = x_{j_0}(\alpha) + \frac{\alpha}{g_{j_0}(x_{j_0}(\alpha))} - \frac{\alpha}{g_S(x_{j_0}(\alpha); \omega)}. \quad (25)$$

De plus, les conditions ci-dessous sont satisfaites

$$\lim_{\alpha \rightarrow 0} G_S(x(\alpha); \omega) = \lim_{\alpha \rightarrow 0} G_H(x(\alpha); \omega) = \lim_{\alpha \rightarrow 0} G_P(x(\alpha); \omega) = 1 - \alpha.$$

Estimation de la distribution de mélange $F_{\mathbb{P}}$ (1/2)

Cadre de travail

Soit $\omega = (\omega_1, \dots, \omega_p) \in [0, 1]^p$ un vecteur de poids, c'est-à-dire $\sum_{j=1}^p \omega_j = 1$.
On considère la distribution de probabilités $F_{\mathbb{P}}$ définie pour tout $x \in \mathbb{R}$ par

$$F_{\mathbb{P}}(x; \omega) = \prod_{j=1}^p F_j^{\omega_j}(x) \iff -\log F_{\mathbb{P}}(x; \omega) = \sum_{j=1}^p \omega_j (-\log F_j(x)).$$

Soit $X_{\mathbb{P}}$ la variable aléatoire associée à la distribution $F_{\mathbb{P}}$.

La relation suivante

$$-\log F_{\mathbb{P}}(X_{\mathbb{P}}; \omega) = \sum_{j=1}^p \omega_j (-\log F_j(X_{\mathbb{P}}))$$

montre que la variable aléatoire $Y_{\mathbb{P}} = -\log F_{\mathbb{P}}(X_{\mathbb{P}}; \omega)$ est à la fois la loi exponentielle standard et une distribution de mélange arithmétique de la loi exponentielle standard $Y_j = -\log F_j(X_{\mathbb{P}})$.

Cette remarque suggère la stratégie suivante pour l'estimation des paramètres de la distribution $F_{\mathbb{P}}$ à partir d'un échantillon $\mathcal{X} = \{x_1, \dots, x_n\}$ de fonction de répartition empirique F_n .

Estimation de la distribution de mélange $F_{\mathbb{P}}$ (2/2)

Stratégie

- 1 Utiliser par exemple l'algorithme appelé **kmeans** pour partitionner l'échantillon $\mathcal{E} = \{-\log F_n(x_1), \dots, -\log F_n(x_n)\}$ en p clusters nommés $\mathcal{E}_j, j = 1, \dots, p$.
- 2 Estimer les poids $\omega_j = \frac{n_j}{\sum_{k=1}^p n_k}$, où n_j est le nombre d'éléments dans le cluster \mathcal{E}_j .
- 3 Estimer les paramètres des distributions F_j à l'aide des observations du cluster $\mathcal{X}_j = \{x_i \in \mathcal{X} : -\log F_n(x_i) \in \mathcal{E}_j\}$.

Cas des valeurs extrêmes

Pour modéliser les valeurs extrêmes d'une variable aléatoire X dont on connaît un échantillon $\mathcal{X} = \{x_1, \dots, x_n\}$, on peut appliquer la stratégie ci-dessus en utilisant :

- 1 la distribution $G_{\mathbb{P}}(x; \omega) = \prod_{j=1}^p G_j^{\omega_j}(x)$ avec $x \in \mathbb{R}$, où G_j est une distribution GEV,
- 2 l'échantillon $\mathcal{X}(\alpha) = \{x_i \in \mathcal{X} : x_i > x(\alpha)\}$, où $x(\alpha) = F_n^{-1}(\alpha)$ avec $\alpha \geq 0.5$, F_n étant la fonction de répartition empirique estimée sur l'échantillon \mathcal{X} .

Distribution des extrêmes et mélange des lois GEV

- Soit X une v.a. de fonction de répartition F et de borne supérieure x_F .
- Soient b_1, \dots, b_p une suite de p entiers positifs suffisamment grands.
- On suppose que pour tout $j = 1, \dots, p$ et pour toute grande valeur $x \in \mathbb{R}$, l'approximation suivante est satisfaite

$$F(x) = \mathbb{P}\{X \leq x\} \sim G_j^{1/b_j}(x; \Theta_j) = G_j(x; \Theta_j(b_j)), \quad (26)$$

où G_j est une distribution GEV de paramètre $\Theta_j = (\gamma_j, \sigma_j, \mu_j) \in \mathbb{R}^3$.

Alors, quel que soit le vecteur $\omega = (\omega_1, \dots, \omega_p) \in [0, 1]^p$ tel que $\sum_{j=1}^p \omega_j = 1$ et pour toute grande valeur $x \in \mathbb{R}$, on peut faire les approximations suivantes

$$F(x) = \mathbb{P}\{X \leq x\}^{\sum_{j=1}^p \omega_j} = \prod_{j=1}^p (\mathbb{P}\{X \leq x\})^{\omega_j} \sim \prod_{j=1}^p [G_j(x; \Theta_j(b_j))]^{\omega_j} = G_P(x; \omega, \Theta(b)).$$

$$F(x) = \mathbb{P}\{X \leq x\} \sum_{j=1}^p \omega_j = \sum_{j=1}^p \omega_j \mathbb{P}\{X \leq x\} \sim \sum_{j=1}^p \omega_j G_j(x; \Theta_j(b_j)) = G_S(x; \omega, \Theta(b)).$$

$$\frac{1}{F(x)} = \frac{\sum_{j=1}^p \omega_j}{\mathbb{P}\{X \leq x\}} = \sum_{j=1}^p \frac{\omega_j}{\mathbb{P}\{X \leq x\}} \sim \sum_{j=1}^p \frac{\omega_j}{G_j(x; \Theta_j(b_j))} = \frac{1}{G_H(x; \omega, \Theta(b))}.$$

Estimation du paramètre Θ : Cas IID

Soit $\mathcal{X} = (x_1, \dots, x_n)$ un échantillon d'une variable aléatoire X ayant une distribution de probabilité **inconnue**.

- Soit $b = \{b_j \in \mathbb{N}^*, j = 1, \dots, p\}$ un ensemble de p tailles de blocs assez grandes.
- Soit $(\gamma_j, \sigma_j, \mu_j)$ le vecteur des paramètres de la loi GEV nommée $G_j(\cdot)$ caractérisant la distribution des maximums de b_j obs. consécutives de la v.a. X .

Nous avons adoptés les notations ci-dessous pour désigner l'ensemble des paramètres qui figurent dans les modèles de mélange qui précèdent

- $b = (b_1, \dots, b_p),$
- $\Theta(b) = (\Theta_j(b_j), j = 1, \dots, p),$
- $\Theta_j(b_j) = (\gamma_j(b_j), \sigma_j(b_j), \mu_j(b_j)),$
- $\gamma_j(b_j) = \gamma_j, \quad \sigma_j(b_j) = \sigma_j b_j^{-\gamma_j}, \quad \mu_j(b_j) = \mu_j + \sigma_j \left(\frac{b_j^{-\gamma_j} - 1}{\gamma_j} \right).$

Estimation du paramètre Θ : Cas Stationnaire

Soit $X_t, t = 1, 2, \dots$ une série temporelle **stationnaire** de processus aléatoire **inconnu**.

- Soit $b = \{b_j \in \mathbb{N}^*, j = 1, \dots, p\}$ un ensemble de p tailles de blocs assez grandes.
- Soit $(\gamma_j, \sigma_j, \mu_j)$ le vecteur des paramètres de la loi GEV nommée $G_j(\cdot)$ caractérisant la distribution des maximums de b_j obs. consécutives de la série X_t .
- Soit $v_j \in [0, 1]$ l'**indice extremal** associé au seuil défini par le quantile empirique $x_{n,1/b_j}$ d'ordre $1/b_j$ estimé sur une séquence $X_n = \{x_1, \dots, x_n\}$ de la série X_t .

Indice extremal

Un indice extremal v quantifie le degré de dépendance entre l'occurrence des valeurs extrêmes consécutives. Cette dépendance est **forte** lorsque v tend vers **0** et **faible** lorsque v tend vers **1**.

Nous avons adopté les notations ci-dessous pour désigner l'ensemble des paramètres qui figurent dans les modèles de mélange qui précèdent

- $b = (b_1, \dots, b_p), \Theta(b) = (\Theta_j(b_j), j = 1, \dots, p), \Theta_j(b_j) = (\gamma_j(b_j), \sigma_j(b_j), \mu_j(b_j)),$
- $\gamma_j(b_j) = \gamma_j, \quad \sigma_j(b_j) = \sigma_j (v_j b_j)^{-\gamma_j}, \quad \mu_j(b_j) = \mu_j + \sigma_j \left(\frac{(v_j b_j)^{-\gamma_j} - 1}{\gamma_j} \right).$

Estimation du paramètre Θ : Cas Non-Stationnaire (1/2)

Soit $X_t, t = 1, 2, \dots$ une série temporelle **non-stationnaire** de processus **inconnu**.

Soit $Y_t = (Y_{1,t}, \dots, Y_{q,t})$ une série temporelle de q covariables pour la série X_t .

Soit x_1, \dots, x_n une séquence de n obs. de la série X_t .

On suppose que chaque obs. x_ℓ est associée à un vecteur de q cov. $y_\ell = (y_{1,\ell}, \dots, y_{q,\ell})$.

- Soit $b = \{b_j \in \mathbb{N}^*, j = 1, \dots, p\}$ un ensemble de p tailles de blocs assez grandes.
- Soit $(\gamma_j(y_t), \sigma_j(y_t), \mu_j(y_t))$ le vecteur des paramètres de la loi GEV nommée $G_j(\cdot | Y_t = y_t)$ caractérisant la distribution conditionnelle des maximums de b_j obs. consécutives de la série X_t .
- Soit $v_j \in [0, 1]$ l'**indice extremal** associé au seuil défini par le quantile empirique $x_{n,1/b_j}$ d'ordre $1/b_j$ estimé sur une séquence $X_n = \{x_1, \dots, x_n\}$ de la série X_t .

Structure des paramètres

- $\mu_j(y_t) = \mu_{0,j} + \mu_{1,j} f_1(y_t) + \dots + \mu_{q,j} f_q(y_t),$
- $\sigma_j(y_t) = \exp \{ \phi_{0,j} + \phi_{1,j} g_1(y_t) + \dots + \phi_{q,j} g_q(y_t) \},$
- $\gamma_j(y_t) = \gamma_{0,j} + \gamma_{1,j} h_1(y_t) + \dots + \gamma_{q,j} h_q(y_t),$

où f_ℓ, g_ℓ, h_ℓ sont des fonctions continues de supports dans \mathbb{R}^q et à valeurs dans \mathbb{R} .

Estimation du paramètre Θ : Cas Non-Stationnaire (2/2)

Nous avons adoptés les notations ci-dessous pour désigner l'ensemble des paramètres qui figurent dans les modèles de mélange qui précèdent

- $b = (b_1, \dots, b_p),$
- $\Theta(b) = (\Theta_j(b_j, y_t), j = 1, \dots, p, t = 1, \dots, n),$
- $\Theta_j(b_j, y_t) = (\gamma_j(b_j, y_t), \sigma_j(b_j, y_t), \mu_j(b_j, y_t)),$
- $\gamma_j(b_j, y_t) = \gamma_j(y_t),$
- $\sigma_j(b_j, y_t) = \sigma_j(y_t) \cdot (v_j b_j)^{-\gamma_j(y_t)},$
- $\mu_j(b_j, y_t) = \mu_j(y_t) + \sigma_j(y_t) \cdot \left(\frac{(v_j b_j)^{-\gamma_j(y_t)} - 1}{\gamma_j(y_t)} \right).$

Estimation du paramètre ω

Estimation des poids

- 1 On désigne par G_j le modèle GEV ayant pour paramètres $(\gamma_j(b_j), \sigma_j(b_j), \mu_j(b_j))$ dans le cas stationnaire et $(\gamma_j(b_j, y_t), \sigma_j(b_j, y_t), \mu_j(b_j, y_t))$ dans le cas non-stationnaire.
- 2 On se donne une probabilité a priori $\mathbb{P}\{G_j\}$ pour qu'une observation quelconque x_i soit issue du modèle G_j . On peut prendre par exemple $\mathbb{P}\{G_j\} = 1/p$.
- 3 La probabilité a posteriori pour qu'une (**grande**) observation $x_i > x_0$ soit issue du modèle G_j est (d'après le théorème de Bayes)

$$\mathbb{P}\{G_j | x_i\} = \frac{\mathbb{P}\{G_j\} \mathbb{P}\{x_i | G_j\}}{\mathbb{P}\{x_i\}} = \frac{\mathbb{P}\{G_j\} g_j(x_i)}{\sum_{k=1}^p \mathbb{P}\{G_k\} g_k(x_i)},$$

où g_j est la densité de probabilité associée au modèle G_j .

- 4 On estime le nombre total d'observations n_j issues du modèle G_j par

$$n_j = \sum_{x_i > x_0} \mathbf{1}\left(\mathbb{P}\{G_j | x_i\} \geq \max_{1 \leq k \leq p} \mathbb{P}\{G_k | x_i\}\right).$$

- 5 On estime le poids ω_j du modèle G_j par $\omega_j = \frac{n_j}{\sum_{k=1}^p n_k}$.