Mélange de distributions des valeurs extrêmes généralisées

Pascal Alain Dkengne Sielenou

Travail en collaboration avec Stéphane Girard (INRIA)

Réunion de synchronisation du 08 septembre 2023

Definition (Melange des distributions de probabilités)

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

Example

Considérons une suite de p fonctions de répartition $F_j, j = 1, \cdots, p$ et un vecteur $\omega = (\omega_1, \cdots, \omega_p) \in [0, 1]^p$ tel que $\sum_{j=1}^{p} \omega_j = 1$. Les lois de mélange moyennes pondérées algébriquement et géométriquement ont respectivement les formes (1) et (2) ci-dessous

$$F_{\mathbb{S}}(x; \omega) = \sum_{j=1}^{p} \omega_j F_j(x), \tag{1}$$

$$F_{\mathbb{P}}(x; \omega) = \prod_{j=1}^{p} F_j^{\omega_j}(x). \tag{2}$$

Definition (Mélange des distributions GEV)

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

On définit le modèle de mélange $G_{\mathbb{P}}$ des lois GEV nommées G_j par

$$G_{\mathbb{P}}(x; \omega, \Theta) = \prod_{j=1}^{p} G_j^{\omega_j}(x; \Theta_j). \tag{3}$$

Les fonctions G_j sont explicitement définies par

$$G_j(x) = G(x; \gamma_j, \sigma_j, \mu_j) = \exp\left\{ -\left[1 + \gamma_j \left(\frac{x - \mu_j}{\sigma_j} \right) \right]^{-\frac{1}{\gamma_j}} \right\}, \tag{4}$$

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

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

$$G_{\mathbb{P}}^{m}(x;\omega,\Theta) = \prod_{j=1}^{p} \left[G_j(x;\Theta_j(m))\right]^{\omega_j} = G_{\mathbb{P}}(x;\omega,\Theta(m)). \tag{5}$$

Ici, $\Theta(m) = (\Theta_j(m), j = 1, \cdots, p)$ *où* $\Theta_j(m) = (\gamma_j(m), \sigma_j(m), \mu_j(m))$
avec $\gamma_j(m) = \gamma_j, \quad \sigma_j(m) = \sigma_j\, m^{\gamma_j}, \quad \mu_j(m) = \mu_j + \sigma_j \left(\frac{m^{\gamma_j} - 1}{\gamma_j}\right).$

La propriété (5) montre que si la loi d'une v.a. X appartient à la famille des probabilités $\{G_{\mathbb{P}}(\cdot;\omega,\Theta)\}$, alors la loi du maximum de m copies indépendantes de X appartient également à cette même famille de probabilités.

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 $b_1, \cdots, b_p$ une suite de p entiers positifs suffisamment grands.
- On suppose que pour tout $j = 1, \cdots, p$ et pour toute grande valeur $x \in \mathbb{R}$, l'équivalence suivante est satisfaite

$$(\mathbb{P}\{X \leq x\})^{b_j} = (F(x))^{b_j} \sim G_j(x; \Theta_j), \tag{6}$$

  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 $\omega = (\omega_1, \cdots, \omega_p) \in [0, 1]^p$ tel que $\sum_{j=1}^p \omega_j = 1$ et pour toute grande valeur $x \in \mathbb{R}$, on peut faire l'approximation suivante

$$\mathbb{P}\{X \leq x\} = F(x) \sim \prod_{j=1}^{p} \left[G_j\left(x; \Theta_j(b_j)\right)\right]^{\omega_j} = G_{\mathbb{P}}\left(x; \omega, \Theta(b)\right). \tag{7}$$

Ici, $b = (b_1, \cdots, b_p)$, $\Theta(b) = (\Theta_j(b_j), j = 1, \cdots, p)$ où $\Theta_j(b_j) = (\gamma_j(b_j), \sigma_j(b_j), \mu_j(b_j))$

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

Modélisation des valeurs extrêmes (1/2)

Soit $\mathcal{X} = (x_1, \ldots, x_n)$ un échantillon d'une v.a. X de distribution de probabilités F.

Estimation du paramètre Θ

1. Soit $b = \{b_j \in \mathbb{N}^\star, j = 1, \cdots, p\}$ un ensemble de tailles de blocs assez grandes.
2. Pour chaque taille de blocs $b_j \in b$, partitionner l'échantillon $\mathcal{X}$ en $n\left(b_j\right) = \lfloor n/b_j \rfloor$ blocs disjoints contenant b_j observations consécutives.
3. Désignons par $z_{b_j} = \left(z_{b_j,1}, \cdots, z_{b_j,n\left(b_j\right)}\right)$ l'échantillon des maximums où $z_{b_j,i}$ est le maximum des observations du i-th bloc de taille b_j.
4. Soit $\widehat{\Theta}_j = (\widehat{\gamma}_j, \widehat{\sigma}_j, \widehat{\mu}_j)$ les paramètres de la loi GEV nommée G_j estimés sur l'échantillon des maximums z_{b_j}.
5. Pour des grandes valeurs de $x \in \mathbb{R}$, la formule (7) permet de faire l'approximation $\mathbb{P}\{X \leq x\} \approx G_{\mathbb{P}}\left(x; \omega, \widehat{\Theta}(b)\right) = \prod_{j=1}^{p} \left[G_j\left(x; \widehat{\Theta}_j(b_j)\right)\right]^{\omega_j}$, où les composantes du vecteur $\widehat{\Theta}_j(b_j) = (\widehat{\gamma}_j(b_j), \widehat{\sigma}_j(b_j), \widehat{\mu}_j(b_j))$ constituant le paramètre $\widehat{\Theta}(b)$ s'écrivent $\widehat{\gamma}_j(b_j) = \widehat{\gamma}_j, \quad \widehat{\sigma}_j(b_j) = \widehat{\sigma}_j\, b_j^{\widehat{-\gamma_j}}, \quad \mu_j(b_j) = \widehat{\mu}_j + \widehat{\sigma}_j \left(\frac{b_j^{\widehat{-\gamma_j}} - 1}{\widehat{\gamma}_j}\right)$.

Modélisation des valeurs extrêmes (2/2)

Estimation du paramètre ω

Le vecteur ω des poids de la loi des extrêmes $G_{\mathbb{P}}\left(x; \omega, \widehat{\Theta}(b)\right)$ peut être estimé en résolvant le problème d'optimisation suivant

$$\widehat{\omega} = \arg\min_{\omega} \left\{ \sum_{x \in \mathcal{X}, x > x(\alpha)} \left[F_{X,n}(x) - G_{\mathbb{P}}\left(x; \omega, \widehat{\Theta}(b)\right) \right]^2 \right\}, \tag{8}$$

où $F_{X,n}$ est la fonction de répartition empirique de la v.a. X estimée sur l'échantillon $\mathcal{X}$ de même que le quantile empirique $x(\alpha)$ d'ordre $\alpha > 0.5$.

Estimation des quantiles extrêmes

Soit $\alpha \in [0, 1]$ tel que α tend vers 0.
Le quantile extrême $x(\alpha)$ défini par $\mathbb{P}\{X > x(\alpha)\} = \alpha$ peut être estimé par une quantité $\widehat{x}(\alpha)$ qui est solution numérique de l'équation $G_{\mathbb{P}}\left(x; \widehat{\omega}, \widehat{\Theta}(b)\right) = 1 - \alpha$.

Conclusion : *Ce travail explique comment combiner plusieurs modèles GEV pour obtenir un modèle assez précis dans le calcul des quantiles extrêmes d'une v.a.*

Modèles de mélange des distributions GEV

Modèles

$$G_{\mathbb{P}}(x) = \prod_{j=1}^{p} G_j^{\omega_j}(x;\gamma_j,\sigma_j,\mu_j). \tag{9}$$

$$G_{\mathbb{M}}(x) = G\left(x;\sum_{j=1}^{p}\omega_{\gamma,j}\,\gamma_j,\sum_{j=1}^{p}\omega_{\sigma,j}\,\sigma_j,\sum_{j=1}^{p}\omega_{\mu,j}\,\mu_j\right). \tag{10}$$

Estimation des poids

$$\widehat{\omega} = \arg\min_{\omega}\left\{\sum_{x\in\mathcal{X},x>x(\alpha)}\left[F_{X,n}(x) - G_{\mathbb{P}}\left(x;\omega\right)\right]^2\right\}, \tag{11}$$

$$\left(\widehat{\omega}_\gamma,\widehat{\omega}_\sigma,\widehat{\omega}_\mu\right) = \arg\min_{\omega_\gamma,\omega_\sigma,\omega_\mu}\left\{\sum_{x\in\mathcal{X},x>x(\alpha)}\left[F_{X,n}(x) - G_{\mathbb{M}}\left(x;\omega_\gamma,\omega_\sigma,\omega_\mu\right)\right]^2\right\}, \tag{12}$$

Modèles de mélange des distributions GEV

En pratique

- $\gamma = \gamma(b, \theta)$
- $\sigma = \sigma(b, \theta)$
- $\mu = \mu(b, \theta)$

où b est une taille de blocs et $\theta \in [0, 1]$ est l'**indice extremal**.

Indice extremal

L'indice extremal θ quantifie le degré de dépendance entre l'occurrence des valeurs extrêmes consécutives.

- θ tend vers 0 signifie que cette dépendance est forte.
- θ tend vers 1 signifie que cette dépendance est faible.