

Prédire des événements extrêmes le pire est-il déjà arrivé ?

Maud Thomas

50 ans du LJLL - Journées Tarantola
Défis en géosciences
Paris, 6-7 juin 2019

Qu'est-ce qu'un événement extrême ?



Événement rare vs événement extrême ?

- **Événement rare** = probabilité d'occurrence est très petite
 - Penser à une loi bimodale
 - Temps entre deux éruptions d'un geyser, analyse du trafic routier, distribution en eau des habitations
- **Événement extrême** = situé dans la queue de distribution
 - Plus grandes observations d'un échantillon ou celles dépassant un certain seuil
 - Canicules, inondations, tempêtes de vent, ...

Événement rare vs événement extrême ?

- **Événement rare** = probabilité d'occurrence est très petite
 - Penser à une loi bimodale
 - Temps entre deux éruptions d'un geyser, analyse du trafic routier, distribution en eau des habitations
- **Événement extrême** = situé dans la queue de distribution
 - Plus grandes observations d'un échantillon ou celles dépassant un certain seuil
 - Canicules, inondations, tempêtes de vent, ...

Etudier les événements extrêmes, c'est étudier le maximum des observations ou les données au-dessus d'un seuil très élevé

Un peu d'histoire



- A Delft (Pays-Bas) en 1953, une tempête a tué des milliers de personnes et détruit près de 50 000 habitations
- Décision du gouvernement : construire une digue telle qu'il n'y ait pas plus d'une inondation tous les 10 000 ans
- MAIS les données disponibles ne couvrent que 100 ans

Un peu d'histoire



- A Delft (Pays-Bas) en 1953, une tempête a tué des milliers de personnes et détruit près de 50 000 habitations
- Décision du gouvernement : construire une digue telle qu'il n'y ait pas plus d'une inondation tous les 10 000 ans
- MAIS les données disponibles ne couvrent que 100 ans

⇒ Comment déterminer la hauteur de la digue ?

Un peu d'histoire



⇒ Comment déterminer la hauteur de la digue ?

- Déterminer la hauteur de la digue $\iff$ Estimer un quantile
- Prendre la vague la plus haute comme référence $\iff$ Considérer que le pire c'est déjà produit

Un peu d'histoire



But de la théorie des valeurs extrêmes

1. Estimer un quantile extrême
2. Estimer la probabilité d'occurrence d'un événement qui n'a pas (encore) été observé

⇒ Inférence en dehors du support de l'échantillon

Cadre

- $X_1, \dots, X_n$ n v.a. i.i.d. $\sim F$ fonction de répartition **inconnue**
- $X_{(1)} \geq \dots \geq X_{(n)}$ statistiques d'ordre (réarrangement décroissant)

Deux questions

1. Pour $x \gg X_{(1)}$, pour $X^\star$ une nouvelle observation $\mathbb{P}\{X^\star > x\}$?
2. Estimer un quantile d'ordre $1 - 1/(xn)$ avec $x \gg 1$?

Approche non paramétrique

1. Fonction de répartition empirique

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}}$$

$\rightarrow \mathbb{P}\{X^\star > x\}$ estimée par 0

2. **Quantile empirique** $\hat{F}_n^{-1}(p) = X_{(i)}$ pour $p \in](n-i)/n, (n-i+1)/n[$
 $\rightarrow$ Estimer $F^{-1}(1-\alpha)$ par $X_{(1)}$, le pire s'est déjà produit

- $X_1, \dots, X_n$ n v.a. i.i.d. $\sim F$ fonction de répartition **inconnue**
- $X_{(1)} \geq \dots \geq X_{(n)}$ statistiques d'ordre (réarrangement décroissant)

Deux questions

1. Pour $x \gg X_{(1)}$, pour X^* une nouvelle observation $\mathbb{P}\{X^* > x\}$?
2. Estimer un quantile d'ordre $1 - 1/(xn)$ avec $x \gg 1$?

Deux points de vue en théorie des valeurs extrêmes

- Etude des lois limites du maximum d'un échantillon
- Etude des excès par rapport à un seuil u

Lois limites des maxima

Etude de la loi du maximum

- Loi de $X_{(1)}$

$$\mathbb{P}\{X_{(1)} \leq x\} = \prod_{i=1}^n \mathbb{P}\{X_i \leq x\} = F^n(x)$$

- Loi limite de $X_{(1)}$

$$F^n(x) \xrightarrow{n \rightarrow \infty} \begin{cases} 0 & \text{si } x < x_F \\ 1 & \text{si } x \geq x_F \end{cases}$$

où $x_F = \sup\{x \in \mathbb{R}, F(x) \leq 1\}$ = point terminal de F

Etude de la loi du maximum

- Loi de $X_{(1)}$

$$\mathbb{P}\{X_{(1)} \leq x\} = \prod_{i=1}^n \mathbb{P}\{X_i \leq x\} = F^n(x)$$

- Loi limite de $X_{(1)}$

$$F^n(x) \xrightarrow{n \rightarrow \infty} \begin{cases} 0 & \text{si } x < x_F \\ 1 & \text{si } x \geq x_F \end{cases}$$

où $x_F = \sup\{x \in \mathbb{R}, F(x) \leq 1\}$ = point terminal de F

$\Rightarrow X_{(1)}$ tend en loi vers δ_{x_F} (loi dégénérée)

Analogie avec les sommes de v.a. i.i.d.

- Suppose $\mathbb{E}X_1 = m \in \mathbb{R}$ et $\text{Var } X_1 = \sigma^2 \in]0, +\infty[$
- **Loi des grands nombres**

$$\overline{X}_n := \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{p.s.} m$$

$\Rightarrow \overline{X}_n$ tend en loi vers δ_m (loi dégénérée)

→ **Question** : Est-il possible de trouver deux suites $a_n > 0$ et b_n telles que

$$\frac{\overline{X}_n - b_n}{a_n}$$

converge en loi vers une loi non-dégénérée ? Si oui, quelle est cette loi limite ?

Analogie avec les sommes de v.a. i.i.d.

- Suppose $\mathbb{E}X_1 = m \in \mathbb{R}$ et $\text{Var } X_1 = \sigma^2 \in]0, +\infty[$
- **Loi des grands nombres**

$$\overline{X}_n := \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{p.s.} m$$

$\Rightarrow \overline{X}_n$ tend en loi vers δ_m (loi dégénérée)

$\rightarrow$ **Question** : Est-il possible de trouver deux suites $a_n > 0$ et b_n telles que

$$\frac{\overline{X}_n - b_n}{a_n}$$

converge en loi vers une loi non-dégénérée ? Si oui, quelle est cette loi limite ? **OUI!**

$\rightarrow$ **Théorème Central Limite** $a_n = \sqrt{\sigma^2/n}$ et $b_n = m$

$$\sqrt{n} \frac{\overline{X}_n - m}{\sigma} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

Retour sur le maximum

Fisher et Tippett (1928), Gnedenko (1953)

S'il existe deux suites $a_n > 0$ et b_n et une loi non-dégénérée G telles que

$$\mathbb{P} \left\{ \frac{X_{(1)} - b_n}{a_n} \leq x \right\} = F^n(a_n x + b_n) \xrightarrow[n \rightarrow \infty]{d} G(x)$$

alors G est *nécessairement* du type

$$G_{\mu, \sigma, \gamma}(x) = \exp \left(- \left(1 + \gamma \frac{x - \mu}{\sigma} \right)_+^{-1/\gamma} \right), x \in \mathbb{R}$$

- Si $\gamma = 0$, le membre de droite se lit $\exp(-\exp(-(x - \mu)/\sigma))$
- Si F vérifie les hypothèses ci-dessus, on dit que F appartient au **max-domaine d'attraction de G_γ** , $F \in \text{DA}(\gamma)$

Lois de valeurs extrêmes généralisées

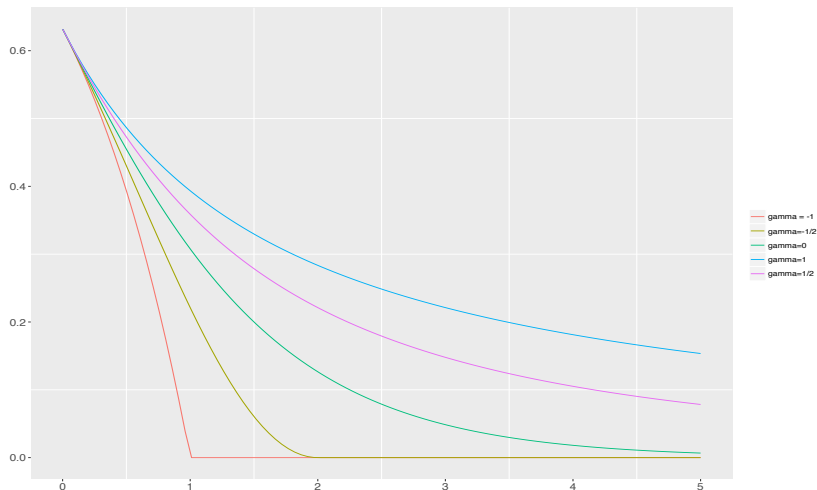


Figure: Fonctions de survie des lois de valeurs extrêmes généralisées.

Quantiles extrêmes

- $\overline{F} = 1 - F$: fonction de survie
- **Quantile d'ordre $1 - \alpha$**

$$q(\alpha) = \inf\{y: F(y) \geq 1 - \alpha\} = \inf\{y: \overline{F}(y) \leq \alpha\}$$

→ $q(1/2)$ = médiane

→ $q(1/4)$ = 3^e quartile

Quantiles extrêmes

Un quantile d'ordre extrême d'ordre $1 - \alpha_n$ est défini par

$$q(\alpha_n) = \inf\{y: F(y) \geq 1 - \alpha_n\} = \inf\{y: \overline{F}(y) \leq \alpha_n\}$$

avec $\alpha_n \rightarrow 0$ quand $n \rightarrow \infty$

- $\alpha_n \rightarrow 0 \Leftrightarrow$ l'information importante pour estimer un quantile extrême est contenue dans la queue de distribution

Estimation de quantiles extrêmes

- Quelle est la probabilité de dépasser le maximum observé ?
- Quelle est la probabilité qu'un quantile soit plus grand que le maximum de l'échantillon ?

$$\mathbb{P}\{X_{(1)} \leq q(\alpha_n)\} \underset{n \rightarrow \infty}{=} \exp(-n\alpha_n(1 + o(1)))$$

- **1^{er} cas : Si $n\alpha_n \rightarrow \infty$, alors $\mathbb{P}\{X_{(1)} \leq q(\alpha_n)\} \rightarrow 0$**
 - Estimer un quantile qui se trouve avec grande probabilité **dans** l'échantillon
 - Requiert une **interpolation** à l'intérieur de l'échantillon
 - Estimateur naturel $X_{(\lfloor n\alpha \rfloor)}$
 - Estimateur asymptotiquement normal

Estimation de quantiles extrêmes

- Quelle est la probabilité de dépasser le maximum observé ?
- Quelle est la probabilité qu'un quantile soit plus grand que le maximum de l'échantillon ?

$$\mathbb{P}\{X_{(1)} \leq q(\alpha_n)\} \underset{n \rightarrow \infty}{=} \exp(-n\alpha_n(1 + o(1)))$$

- **2^e cas : Si $n\alpha_n \rightarrow 0$, alors $\mathbb{P}\{X_{(1)} \leq q(\alpha_n)\} \rightarrow 1$**
 - Estimer un quantile qui se trouve avec grande probabilité **hors** l'échantillon
 - Inverser la fonction de répartition empirique n'est plus une solution
 - Requiert une **extrapolation** à l'extérieur de l'échantillon
- Estimer les quantiles d'une GEV

Quantiles d'une loi de valeurs extrêmes généralisées

- Il suffit d'inverser l'équation de la GEV

$$G_{\mu,\sigma,\gamma}(x) = \exp \left(- \left(1 + \gamma \frac{x - \mu}{\sigma} \right)_+^{-1/\gamma} \right), x \in \mathbb{R}$$

Quantiles extrêmes

$$G_{\mu,\sigma,\gamma}(z_p) = 1 - p \Leftrightarrow z_p = \begin{cases} \mu - \frac{\sigma}{\gamma} \left(1 - (-\log(1 - p))^{-\gamma} \right) & \text{pour } \gamma \neq 0 \\ \mu - \sigma \log(-\log(1 - p)) & \text{pour } \gamma = 0 \end{cases}$$

- Pour estimer les quantiles extrêmes, remplacer μ, σ, γ par leurs estimateurs
- z_p est aussi appelé le **niveau de retour** à $1/p$ années
 - C'est la valeur qui sera dépassée en moyenne 1 fois dans les $1/p$ prochaines années
 - C'est la valeur qui sera dépassée avec probabilité p l'année prochaine

Lois limites des excès

Etude de la loi des excès

- X v.a. de fonction de répartition F (fonction de survie $\overline{F} = 1 - F$)
- u seuil fixé
- Excès de X au dessus de $u =$ v.a. $Y_u = X - u$ définie sur $\{X > u\}$

Loi d'excès

La loi de Y_u est la loi dont la fonction de survie est donnée par

$$\overline{F}_u(x) = \frac{\overline{F}(u+x)}{\overline{F}(u)}$$

Famille de lois limites des excès

Lois de Pareto généralisées (GPD)

Si F appartient au max domaine d'attraction d'une GEV $G_{\mu,\sigma,\gamma}$, alors $\bar{F}_u$ peut être approchée par une loi de fonction de répartition donnée par

$$H_{\sigma,\gamma}(x) = \left(1 + \gamma \frac{x}{\tilde{\sigma}}\right)_+^{-1/\gamma} \quad x > 0$$

avec $\tilde{\sigma} = \sigma + \xi(u - \mu)$

- Si $\gamma = 0$, le membre de droite se lit $\exp(-x/\tilde{\sigma})$
- Si F vérifie les hypothèses ci-dessus, on dit que F appartient au domaine d'attraction des excès de $H_{\sigma,\gamma}$
- Domaine d'attraction des excès = max-domaine d'attraction
- Lois de Pareto généralisées (GPD) = famille des lois limites

Quantiles d'une loi de Pareto généralisée

- Les quantiles extrêmes peuvent être également vus comme des quantiles d'une GPD
- Il suffit d'inverser l'équation de la GPD

$$H_{\sigma,\gamma}(x) = 1 - \left(1 + \gamma \frac{x}{\sigma}\right)_+^{-1/\gamma} \quad x > \mu$$

Quantiles extrêmes

$$H_{\sigma,\gamma}(x_m) = \frac{1}{m} \Leftrightarrow x_m = \begin{cases} u + \frac{\sigma}{\gamma} ((m\zeta_u)^\gamma - 1) & \text{pour } \gamma \neq 0 \\ u + \sigma \log(m\zeta_u) & \text{pour } \gamma = 0 \end{cases}$$

- **Interprétation :** x_m est le seuil dépassé en moyenne toutes les m observations

Méthode « Peaks over Threshold »

- u seuil élevé
- $X_1, \dots, X_n$ n v.a. i.i.d.

$$t_1 = \inf\{j \geq 1 : X_j > u\}$$

$$t_2 = \inf\{j > t_1 : X_j > u\}$$

$$\vdots \quad \quad \vdots$$

$$t_r = \inf\{j > t_{r-1} : X_j > u\}$$

- $(X_{t_r}, r \geq 1)$ = suite des X_i ayant dépassé u
- $(X_{t_r} - u, r \geq 1)$ = suite des excès au-dessus de u
- Ajuster une loi de Pareto généralisée sur cette suite de v.a.

Méthode « Peaks over Threshold »

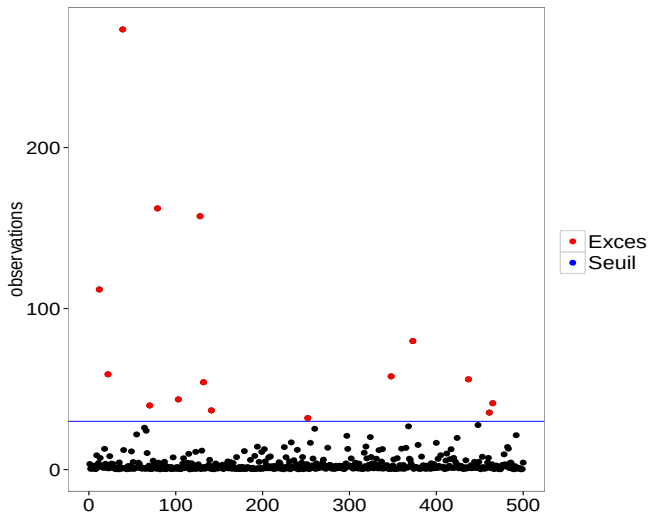
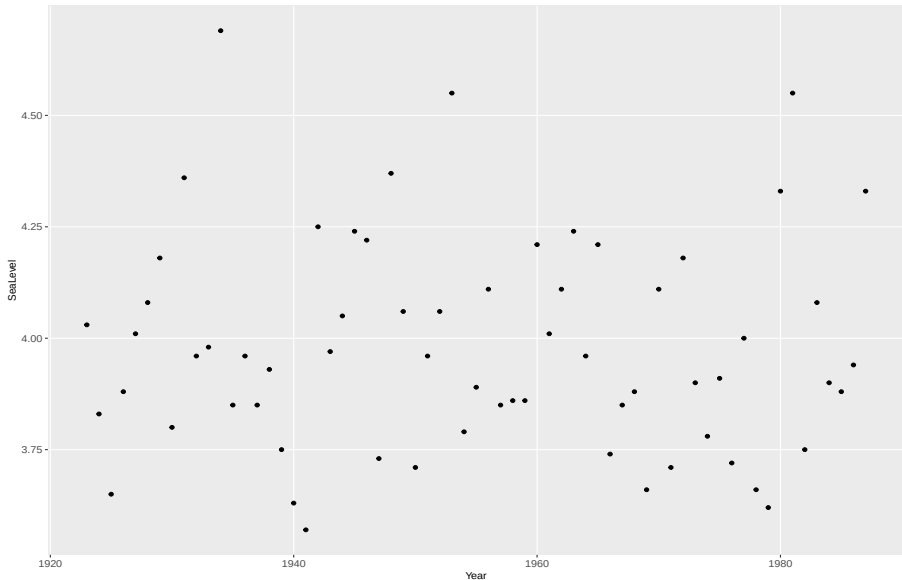


Figure: Illustration de la méthode POT

Exemple : Niveaux de la mer annuels maximaux at Port Pirie de 1923 à 1987



Exemple : Niveaux de la mer annuels maximaux at Port Pirie de 1923 à 1987

- Estimation des paramètres de la GEV

$$\hat{\mu} = 3.87 \quad \hat{\sigma} = 0.20 \quad \hat{\gamma} = -0.05$$

→ $\gamma = 0$?

Exemple : Niveaux de la mer annuels maximaux at Port Pirie de 1923 à 1987

- Estimation des paramètres de la GEV

$$\hat{\mu} = 3.87 \quad \hat{\sigma} = 0.20 \quad \hat{\gamma} = -0.05$$

→ $\gamma = 0$?

→ Il faut regarder l'intervalle de confiance $[-0.242, 0.142]$

→ On peut considérer que $\gamma = 0$ (domaine de Gumbel)

Exemple : Niveaux de la mer annuels maximaux at Port Pirie de 1923 à 1987

- Estimation des paramètres de la GEV

$$\hat{\mu} = 3.87 \quad \hat{\sigma} = 0.20 \quad \hat{\gamma} = -0.05$$

→ $\gamma = 0$?

→ Il faut regarder l'intervalle de confiance $[-0.242, 0.142]$

→ On peut considérer que $\gamma = 0$ (domaine de Gumbel)

- Estimations de niveaux de retour

→ Avec probabilité $p = 0.1$, $\hat{z}_{0.1} = 4.30$

→ Avec probabilité $p = 0.01$, $\hat{z}_{0.01} = 4.70$

- Probabilité de dépasser le maximum observé

→ Maximum observé = 4.69

→ Probabilité que l'on dépasse cette valeur l'an prochain = 0.016

Merci de votre attention



« Il est impossible que l'improbable n'arrive jamais », E.J. Gumbel (1958)