Extreme value Theory. - Théorie des valeurs extrêmes

* Étudier les évènements extrêmes/rares et prédire des quantiles

exemple : - estimer la proba d'une inondation qui survient tous les 100 ans
- en finance, estimer la VAR et sa proba de survenue.

$\{ VaR = \inf(x \in \mathbb{R} : F_L(x) \geq \alpha) \}$

Il existe trois approches d'études des valeurs extrêmes

I L'approche par blocs (Block maxima)

- $(X_i)_{i=1,\dots,m}$ des v.a.r iid.
- $M_m = \max(X_1, \dots, X_m)$
- $M_m^{(k)} = \max(X_1, \dots, X_m)$ le max des m (X_i) dans le bloc (k)

Théorème clé

s'il existe $\{a_m > 0, m \geq 1\}$ et $\{b_m \in \mathbb{R}, m \geq 1\}$ tq

$$P\left(\frac{M_m - b_m}{a_m} \leq x\right) \xrightarrow[m \to +\infty]{} H(x)$$

alors, $H(x) = \exp\left\{-\left(1 + \varepsilon \frac{x - \mu}{\sigma}\right)^{-1/\varepsilon}\right\}, \quad 1 + \varepsilon \frac{x-\mu}{\sigma} > 0$

ou

$= \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}, \quad \varepsilon = 0, \; x \in \mathbb{R}$

* μ est le param de position (valeur type de (X_i))
* ε est le param de forme (agit sur la variance & la position)
* σ est le param d'échelle (agit sur la variance)

$\Rightarrow$ X suit donc une GEV $(\mu, \sigma, \varepsilon)$ (Generalized Extrem value)

Pour determiner les constantes qui assure la limite vers une GEV (si elle existe), on fait appelle aux conditions de von Mises* qui necessite initialement des connaissances sur la loi de X.

En pratique, on ajuste directement le modèle par maximum de vraisemblance, methode des moments ou vraisemblance profilée.

[Conditions de Von Mises.
Mills ratio : $r(x) = (1-F(x))/f(x)$
$b_m = F^{-1}\left(1-\frac{1}{m}\right)$, $a_m = r(b_m)$, $\xi = \lim_{x \to x_+} r'(x)$]

1/ Periode de retour et niveau de retour

Il peut être interessant d'estimer un quantile d'une $GEV(\mu, \sigma, \xi)$ par determiner la valeur à partir de laquelle X depasse cette valeur avec proba p.
$(F(y_p) = 1-p \Leftrightarrow p = 1 - F(y_p))$

$$y_p = \mu - \sigma \frac{1 - \{-\log(1-p)\}^{-\xi}}{\xi}$$

niveau de retour associé à un periode $1/p$. (utilisé pour prediction)

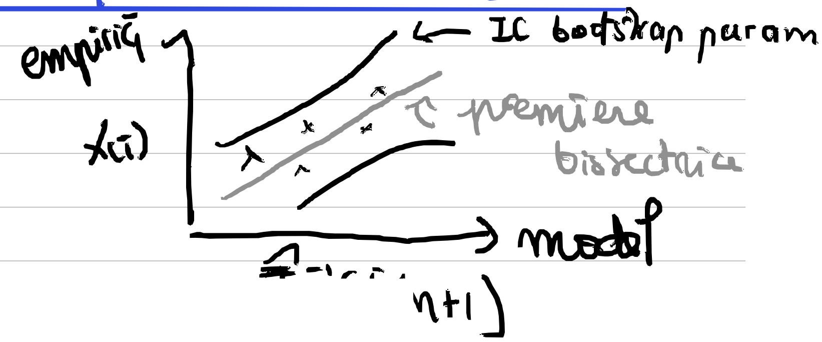
(en finance, y_p est la VaR)

Interprétation

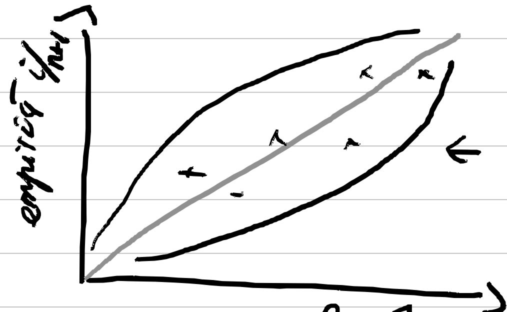
- y_p est en moyenne depassé une fois tous les $1/p$ blocs (saisonalité)
- Dans un bloc, la proba que y_p soit dépassé est $1/p$.

2/ Verification de la qualité du modèle

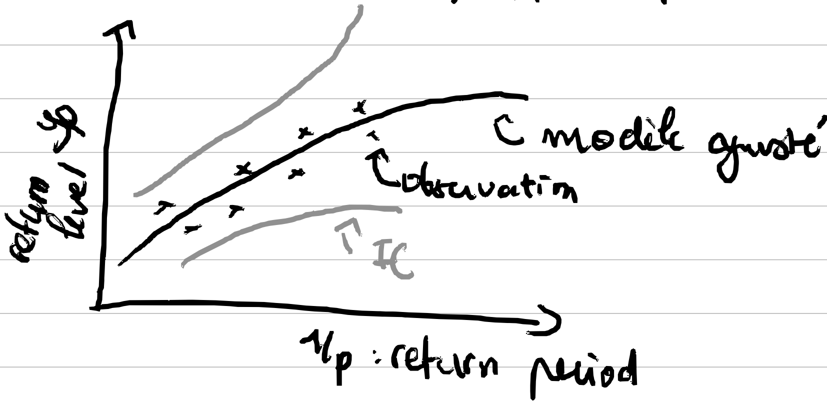
* QQ-Plot :



* PP-plot :



+ return level / return period



En pratique, les logiciels représentent $-1/\log(1-1/T)$
$\approx T$
$T = \frac{1}{p} \to +\infty$

+ Convergence de la méthode d'optimisation

[penser au test de Wald et test de rapport de vraisemblance (puissant) pour simplifier le modèle (annuler certains param ou autre...]

3) Tester l'incertitude des estimateurs

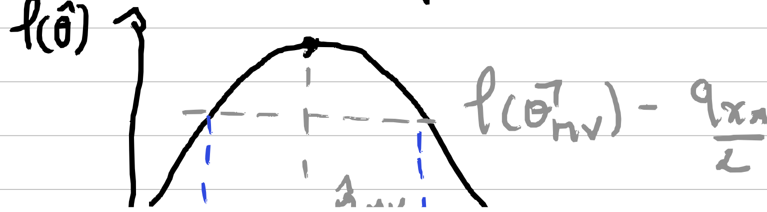
L'intervalle de confiance permet de savoir les valeurs possibles prises par un estimateur

IC (symétrique) de Wald :

$$\hat{\theta} \pm z_{1-(1-\alpha)/2} \times \hat{\sigma}(\hat{\theta})$$

erreur type

Il est qq fois utile de construire des IC non symétriques pour savoir vers quelle borne tend le paramètre d'intérêt. On utilise la vraisemblance profilée.



4. Reparamétrisation

Il est parfois utile de reparamétriser la loi GEV(y_p, σ, ε). De fait,

$$\mu : y_p + \sigma \frac{1 - \{-\log(1-p)\}^{-\varepsilon}}{\varepsilon}$$

→ Cela permet d'avoir un IC pour le quantile
→ change la valeur typique prise par la distribution.

?? intérêt

II- L'approche par excès de seuil (peak over threshold - Threshold exceedances)

On étudie dans ce cas les valeurs qui excèdent un certain seuil.

Théorème clé

Si X converge vers une GEV, alors

$$P(X > u_m(1+x) \mid X > u_m(x)\} \xrightarrow[m\to+\infty]{} 1 - \tilde{H}(x)$$

suite croissante

avec $u_m(x) = a_m x + b_m$, $x \in \mathbb{R}_+$.

$$\tilde{H}(x) = \begin{cases} 1 - (1+\varepsilon x/_\tau)^{-1/\varepsilon}, & \varepsilon \neq 0 \\ 1 - \exp(-x/_\tau), & \varepsilon = 0 \end{cases}$$

où τ est le nouveau paramètre d'échelle
$\tau = \sigma + \varepsilon(u - \mu)$

On dit que $\tilde{H}(\cdot)$ est une GPD(τ, ε), Generalized Pareto Distribution.

1. Le choix du seuil u.

/!\ On choisit u, on ne l'estime pas

Il est habituel que $u = \bar{F}(0,95)$, cependant, on peut utiliser des méthodes permettant d'avoir une idée sur le u (optimal).

(/!\ attention au compromis biais (u bas), variance (u élevé))

Soit $X-u_0 \mid X>u_0 \sim GPD(\tau,\varepsilon)$, alors $\forall u \geq u_0$

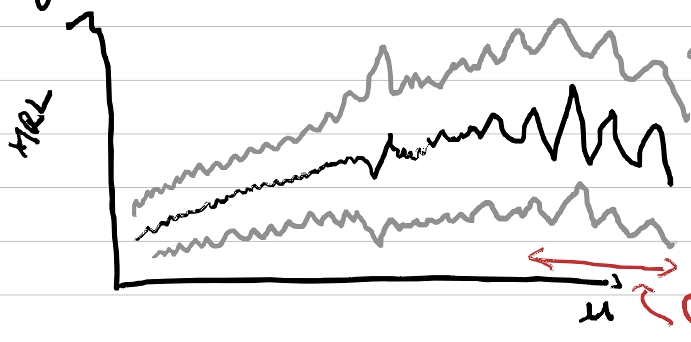
$$X-u \mid X>u \sim GPD(\bar{\tau},\varepsilon),\ \bar{\tau} = \tau + \varepsilon(u-u_0)$$

(lorsque u est bien sélectionné, lorsqu'on augmente le seuil, la loi conditionnelle suit toujours une GPD avec une modification de paramètre d'échelle.

a/ approche MRL

Si $X-u_0 \mid X>u_0 \sim GPD(\tau,\varepsilon)$, $\varepsilon<1$, alors $\forall u \geq u_0$,

$MRL(u) = \mathbb{E}(X-u \mid X>u) = \dfrac{\tau(u_0)+\varepsilon u}{1-\varepsilon}$ est une fonction linéaire de u.



← IC de Wald (symétrique)

on cherche le u à partir duquel on arrive à tracer une droite

comportement normal quand u grand car on fait une moyenne sur peu d'obs.

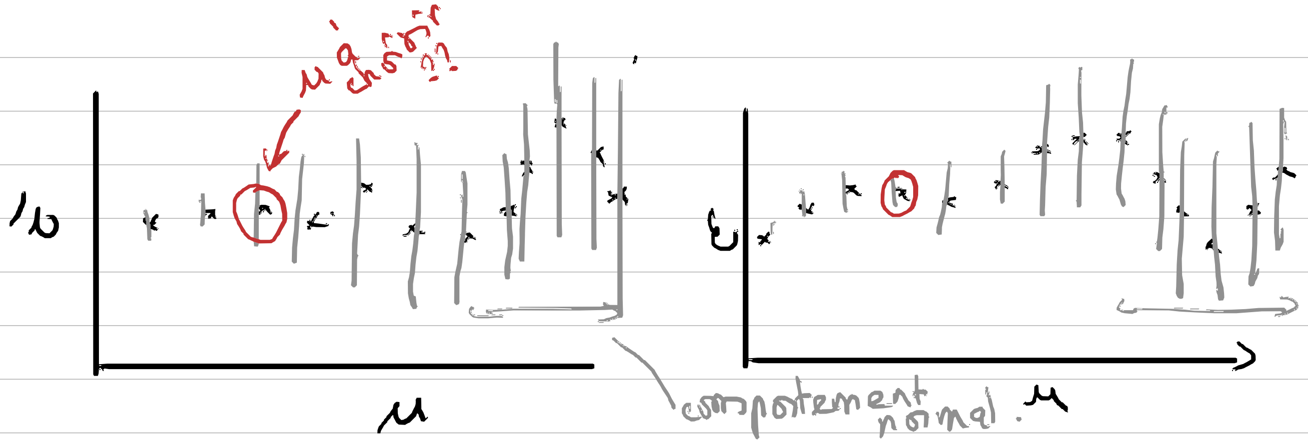
b/ approche "Parameters stability"

Si $X-u_0 \mid X>u_0 \sim GPD(\tau,\varepsilon)$, alors $\forall u \geq u_0$,

$$X-u \mid X>u \sim GPD(\bar{\tau},\varepsilon),\ \bar{\tau} = \tau+\varepsilon(u-u_0)$$

$$\Rightarrow \tau_*(u) = \bar{\tau} - \varepsilon u = \tau + \varepsilon(u-u_0) - \varepsilon u = \tau - \varepsilon u_0.$$

les paramètres $\bar{\tau}$ et ε doivent être constants pour les seuils u valides.



Lorsque le seuil u est bon, lorsqu'on l'augmente les résultats sont assez stables. De fait le graph. de return level est sensiblement le même.

2) Return period et return level

$$y_p = \inf\{ y \in \mathbb{R} : F(y_p) \geq 1-p \}$$

$$\Rightarrow P(X \leq y_p) = 1-p$$

$$\Leftrightarrow P(X > y_p) = \underbrace{P(X > y_p \mid X > u)}_{\text{GPD}} \underbrace{P(X > u)}_{\text{estimé empiriquement } p(u)} = p$$

$$\Rightarrow y_p = u + \sigma \frac{(p/p(u))^{-\xi} - 1}{\xi} \quad \text{(valeur brute)}$$

lorsqu'il n'y a pas u, y_p n'est qu'un excès

Interprétation : y_p est une valeur qui est dépassée une fois tous les $1/p$ observations (cela dépend de leur unité : jour, trimestre, ...). Il est possible de le mettre à une autre échelle : $1/p \cdot n_y$.

exemple : lorsque les obs. sont en jours, y_p est supposé être dépassé une fois tous les $1/p$ jours. Cependant, on peut travailler sur une échelle annuelle. Il y a $n_y = 365$ jours dans une année donc y_p est dépassé une fois tous les $1/(p \times n_y)$ années.

On peut interpréter la période de retour comme : "dans une période, y_p est excédé avec proba $1/T$"

III - L'approche par processus ponctuel de poisson

Processus ponctuel

Un processus stochastique est une réalisation de points qui arrive aléatoirement dans un espace. On appelle mesure de comptage

$$N(A) = \sum_{i \in I} \delta_{x_i}(A) = \sum_{i \in I} \delta(x_i \in A)$$

où A est un ensemble borélien et δ un dirac.

On mesure mesure d'intensité

$$\Lambda : A \longrightarrow \underbrace{E[N(A)]}_{\text{nombre moyen de points tombé dans l'espace } A}$$

Processus ponctuel de Poisson

Un processus ponctuel avec une mesure d'intensité Λ est un processus de poisson si $\forall\, k \geq 1$ et tous ensembles disjoints boréliens $A, A_1, \dots, A_k \in \mathcal{X}$

i) $N(A) \sim \text{Poisson}(\Lambda(A))$

ii) $N(A_1), \dots, N(A_k)$ sont des v.a independants.

si Λ est regulière, $\Lambda(A) = \int_A \lambda(s)\, ds$

λ : fonction d'intensité

Proposition :

Soit $\{x_i\}$ une réalisation d'un processus de poisson sur $\mathcal{X}$ avec une mesure d'intensité Λ. Alors

$$\mathcal{L}(x_i, \Lambda) = \exp[-\Lambda(\mathcal{X})] \prod_{i=1}^{n} \lambda(x_i)$$

Théorème clé

Soit M_n converge vers une GEV, la sequence de processus ponctuel dans $\mathcal{X} = [0,1] \times \mathbb{R}$ ($\mathbb{R}^d$)

$$\{P_m\}_{m \geq 1} = \left\{ \left(\frac{i}{m+1}, \frac{x_i - b_m}{a_m} \right) : i = 1, \dots, m \right\}$$

converge vers un PPP(μ, σ, ξ) sur $[0,1] \times C$ avec

$$\Lambda\{[a,b] \times [z, \infty)\} = \underbrace{(b-a)}_{*} \left(1 + \xi \frac{z-\mu}{\sigma}\right)^{-1/\xi}$$

où $C = \{x \in \mathbb{R} : 1 + \xi(x-\mu)/\sigma > 0\}$

$* \Rightarrow$ les extrêmes arrivent uniformément dans le temps.

1/ Interprétation statistique.

Dans un PPP, on travaille sur $x > u$. On ajuste ainsi un PPP ayant une mesure d'intensité :

$$\Lambda\{(a,b)\times(x,\infty)\} = (b-a)\times\left(1+\xi\,\frac{x-\mu}{\sigma}\right)^{-1/\xi},\quad x>u,\ (a,b)\subset[0,1]$$

Pour travailler à l'échelle annuelle,

$$\Lambda\{(a,b)\times(x,\infty)\} = n_{year}(b-a)\times\left(1-\xi\,\frac{x-\mu}{\sigma}\right)^{-1/\xi}$$

$\Rightarrow$ le nombre d'excès ($>x$) dans une année en moyenne est :

$$E[N\{(0,1/n_{year})\times(x,\infty)\}] = \Lambda\{(0,n_{year}^{-1})\times(x,\infty)\}$$
$$= \left(1+\xi\,\frac{x-\mu}{\sigma}\right)^{-1/\xi}$$

Le niveau y_p qui est dépassé avec proba $1/T$ satisfait :

$$T\left(1+\xi\,\frac{y_p-\mu}{\sigma}\right)^{-1/\xi} \Leftrightarrow y_p = \mu+\sigma\,\frac{p^{\xi}-1}{\xi}$$

/!\ y_p est uniquement calculé sur $x>u$.

IV - Cas Non Stationnaire.

En EVT, on fait l'hypothèse que $\{X_i\}$ iid. Cependant il se peut que l'hypothèse d'indépendance (i), ou celle de la distribution identique (id) ne soient pas respectés.

Dans le contexte de série non stationnaire, i.e $E[X_t] \neq \mu$ (non constante), $Var(X_t) \neq \sigma^2$ (non constante) $Cov(X_i, X_{t+h}) = f(h)$ (non corrélation, dépend que de h), l'hypothèse id n'est pas respecté (saisonalité, tendance)

Il existe donc trois approches :

1/ Stationnariser la série (retrait de saisonnalité, tendance & stabilisation de la variance)

test de nonstationnarité : Dicker Fuller / ADF
Désaisonnalisation : Différenciation, moyenne mobile

2/ Restreindre le périmètre d'étude (saisons extrêmes etc.) → réfléchir à la construction de blocs.

3/ Intégrer une saisonnalité aux paramètres GEV, GPD et PPP.

<u>exemple</u> $\mu(t) = \beta_0 + \beta_1 t$

on peut modéliser $\sigma(t)$ (difficile à estimer) et $\xi(t)$ (très difficile à estimer).

Sélection de modèles.

Il est possible de tester la nullité des coefficients ⇒ Test de Wald, rapport de vraisemblance (modèle emboîté)

test du r.v.

$$W = -2\log\left(\frac{\overbrace{\mathcal{L}(\theta_0)}^{\text{modèle contraint}}}{\underbrace{\mathcal{L}(\hat{\theta})}_{\text{modèle complet}}}\right) \underset{n\to+\infty}{\sim} \chi^2(\underbrace{p}_{\text{nbre de contraintes.}})$$

Test de Wald

$$\sqrt{n}\,(\tilde{\vec{\theta}} - \vec{\theta}_0) \xrightarrow[n\to+\infty]{\mathcal{L}} N(0,\sigma^2)$$

sous H_0, $\sqrt{n}\,\dfrac{\vec{\theta} - \vec{\theta}_0}{\sigma} \longrightarrow N(0,1)$

AIC $\underbrace{-2\ell(\hat{\vec{\theta}})}_{\text{attache aux données}} + 2p$

BIC $-2\ell(\vec{\theta}) + \log(n) \times p$ (plus parcimonieux)

asymptotiquement, l'AIC a une grande proba de sélectionner un mauvais modèle