

Théorie des Valeurs Extrêmes en Sciences du Climat

Pierre Ribereau
Gwladys Toulemonde, Ali Gannoun
Université Montpellier 2

Mèze : Journées Stat du Sud

24/06/2010

Outline

- 1 Extrêmes univariés
- 2 Extrêmes multivariés
- 3 Approche par processus ponctuel et mesure spectrale

Outline

- 1 Extrêmes univariés
- 2 Extrêmes multivariés
- 3 Approche par processus ponctuel et mesure spectrale

Exemple

On dispose d'un échantillon X_1, X_2, \dots, X_n , de hauteurs d'eau journalières d'une rivière de fonction de répartition F et de fonction de survie associée notée \bar{F} et définie comme $\bar{F}(x) = \mathbb{P}(X > x) = 1 - F(x)$.

On souhaite être capable de répondre aux **deux questions complémentaires** suivantes :

- Calculer le seuil z_T qui ne sera dépassé qu'une seule fois en moyenne au cours des T prochaines années.
- Quelle est la probabilité $p = \frac{1}{T}$, pour une année quelconque, que le max annuel dépasse z_T ?

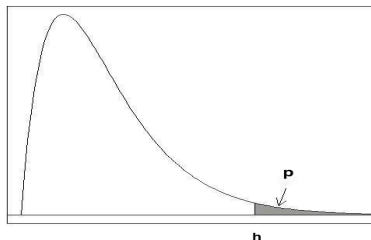
Exemple

On souhaite répondre à l'une de ces deux questions complémentaires avec p très petit ($T = 1/p$ très grand) ou de manière équivalente avec z_T très grand.

↪ **Etude de la queue de distribution**



Difficulté commune : la fonction de survie $\bar{F}(x)$ est inconnue et difficile à estimer au-delà d'un certain seuil h .



Approche paramétrique

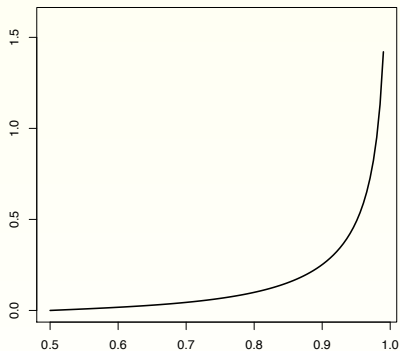
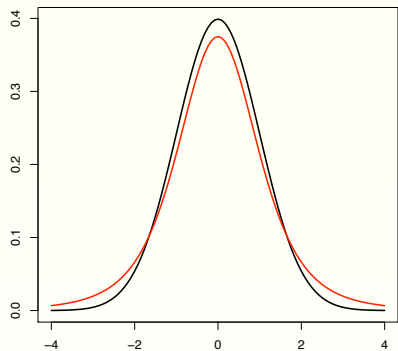
Démarche :

- On suppose un modèle paramétrique *a priori* pour la fonction de survie : $\overline{F} \in \{\overline{F}_\theta, \theta \in \Theta\}$.
- On estime θ par $\hat{\theta}_n$.



Problème : Un bon ajustement sur l'échantillon ne garantit pas une bonne modélisation dans les queues de distribution.

Illustration



A gauche : Distribution Normale et de Student à 4 degrés de liberté.

A droite : Ecart absolu entre les quantiles d'ordre $1 - p$ de la loi Normale et de la loi de Student à 4 degrés de liberté.

Théorème général

Théorème général

Soit un échantillon X_1, X_2, \dots, X_n de variables indépendantes et identiquement distribuées. Si il existe $\{a_n\}, \{b_n > 0\}$ et une distribution G non dégénérée telles que

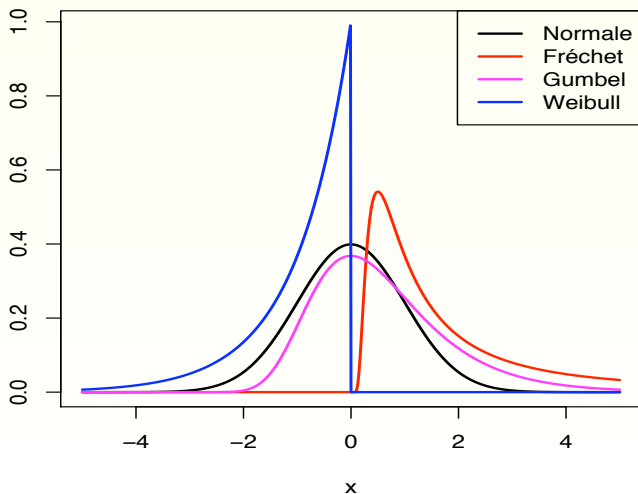
$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\frac{X_{n,n} - a_n}{b_n} \leq x \right) = G_\gamma \left(\frac{x - \mu}{\sigma} \right)$$

alors G_γ appartient à la famille de la loi généralisée des valeurs extrêmes :

$$G_\gamma(x) = \exp \left\{ - [1 + \gamma x]^{-1/\gamma} \right\}$$

où $x \in \{z : 1 + \gamma z > 0\}$, avec $\gamma \in \mathbb{R}$.

Lois des valeurs extrêmes



Application à l'extrapolation

Comme $\mathbb{P}(X_{\mathbf{n},\mathbf{n}} \leq x) = F^n(x)$, on déduit directement du théorème des valeurs extrêmes une **approximation de la queue de la distribution** :

$$\bar{F}(x) \approx 1 - G_\gamma^{1/n} \left(\frac{x - a_n}{b_n} \right)$$

et de son inverse :

$$\begin{aligned} \bar{F}^{-1}(p) &\approx a_n + \frac{b_n}{\gamma} [(-\mathbf{n} \log(1-p))^{-\gamma} - 1] \text{ si } \gamma \neq 0 \\ &\approx a_n - b_n \log(-\mathbf{n} \log(1-p)) \text{ si } \gamma = 0. \end{aligned}$$

Domaine d'attraction

Définition

Soit G_γ une distribution non dégénérée. On dit que F appartient au domaine d'attraction des extrêmes de G_γ (noté $F \in DA(G_\gamma)$) s'il existe $\{a_n\}$ et $\{b_n > 0\}$ tels que

$$\lim_{n \rightarrow +\infty} F^n(a_n + b_n x) = G_\gamma(x).$$

L'appartenance à un domaine d'attraction est liée à la forme de la queue de distribution et plus particulièrement à la vitesse de décroissance de $\bar{F}(x)$ vers 0 quand x devient très grand.

On a donc 3 domaines d'attraction :

- Le domaine d'attraction de **Weibull** ($\gamma < 0$), distributions bornées
- Le domaine d'attraction de **Gumbel** ($\gamma = 0$) décroissance exponentielle
- Le domaine d'attraction de **Fréchet** ($\gamma > 0$)

Domaine d'attraction de Fréchet ($\gamma > 0$)

- Ensemble des lois dites à **queues lourdes**, $\bar{F}(x) \rightarrow 0$ comme une puissance de x lorsque $x \rightarrow +\infty$.
- Plus précisément, on a la représentation

$$\bar{F}(x) = x^{-1/\gamma} \ell(x),$$

où ℓ est une fonction à variation lente.

- a_n et b_n associées : $a_n = 0$ et $b_n = \bar{F}^{-1}(1/n) = \inf\{x : 1 - F(x) \leq \frac{1}{n}\}$.
- $\mathbb{E}(X^a) = \infty$ si $a > \frac{1}{\gamma}$.

Théorème

F appartient au domaine d'attraction de Fréchet ($F \in DA(\Phi_{1/\gamma})$) si et seulement si \bar{F} est une fonction à variation régulière d'ordre $-1/\gamma$.

Distribution	$1 - F(x)$	γ
Burr(β, τ, λ), $\beta > 0$, $\tau > 0$, $\lambda > 0$	$\left(\frac{\beta}{\beta + x^\tau}\right)^\lambda$	$\frac{1}{\lambda\tau}$
Fréchet($\frac{1}{\alpha}$), $\alpha > 0$	$1 - \exp(-x^{-\alpha})$	$\frac{1}{\alpha}$
Pareto(α), $\alpha > 0$	$x^{-\alpha}$	$\frac{1}{\alpha}$
Gumbel(μ, β), $\mu \in \mathbb{R}$, $\beta > 0$	$\exp\left(-\exp\left(-\frac{x-\mu}{\beta}\right)\right)$	0
Logistique	$\frac{2}{1+\exp(x)}$	0
Exp(λ), $\lambda > 0$	$\exp(-\lambda x)$	0
ReverseBurr($\beta, \tau, \lambda, \tau_F$), $\beta > 0$, $\tau > 0$, $\lambda > 0$	$\left(\frac{\beta}{\beta + (\tau_F - x)^{-\tau}}\right)^\lambda$	$-\frac{1}{\lambda\tau}$
Uniforme(a, b)	$1 - \frac{x-a}{b-a}$	-1

TABLE 1: Quelques distributions et leur indice extrême associé.

En pratique, dans certains cas...

- On ne dispose pas nécessairement de max dès le départ.
- Il faut pourtant un échantillon de maxima pour utiliser les méthodes étudiées
- Ces maxima sont parfois difficiles à extraire des données initiales, problème du nombre d'observations utilisées...

↔ Approche par blocs

Maxima par bloc

Au départ, on suppose avoir des réalisations indépendantes et de même loi F d'un certain phénomène d'intérêt : X_1, X_2, \dots, X_n .

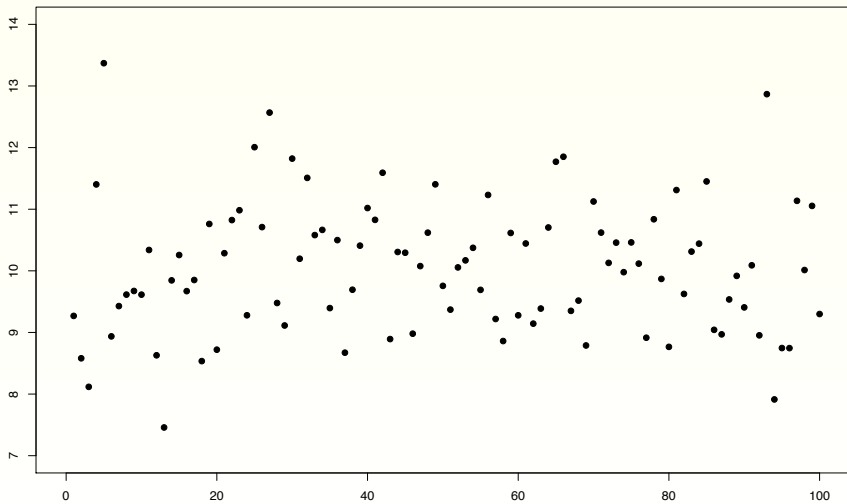
Exemple : Pluie journalière ; niveau de pollution ; hauteur de vagues...

Pour obtenir un échantillon de max, on découpe les données en m blocs de même taille k (dépend du contexte, du nombre de données $n = m \times k, \dots$) :

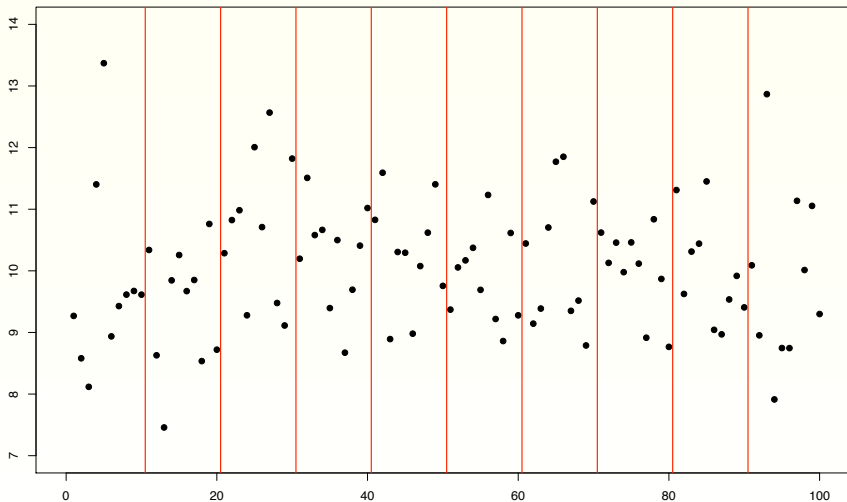
$$\underbrace{X_1, X_2, \dots, X_k}_{Z_1}, \underbrace{X_{k+1}, X_{k+2}, \dots, X_{2k}}_{Z_2}, \dots, \underbrace{X_{(m-1)k+1}, \dots, X_{mk}}_{Z_m}$$

On obtient alors un échantillon de m réalisations de loi GEV : Z_1, Z_2, \dots, Z_m sur lequel on peut mettre en oeuvre les méthodes vues précédemment.

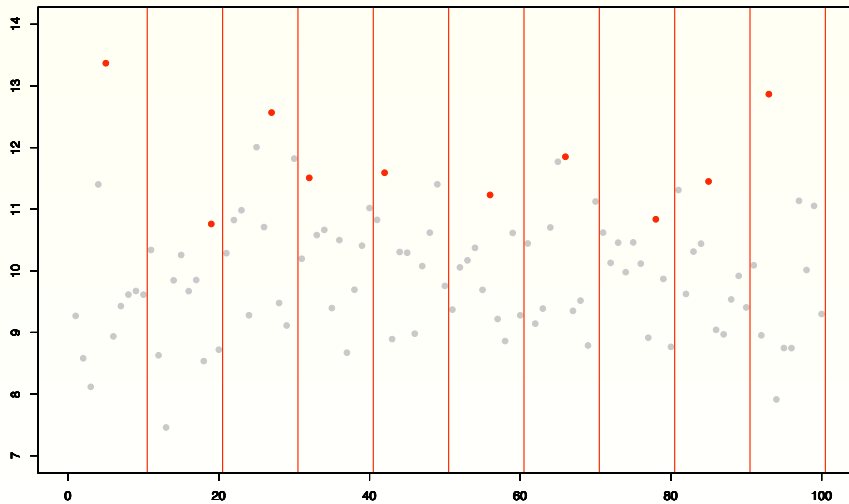
Maxima par bloc : exemple



Maxima par bloc : exemple



Maxima par bloc : exemple



En pratique, dans certains cas...

- Ces maxima sont parfois difficiles à extraire des données initiales
- Problème du nombre d'observations utilisées...
- Une valeur peu élevée peut correspondre à un maximum sur une période

↔ **Alternative** : L'approche POT

Etude des excès

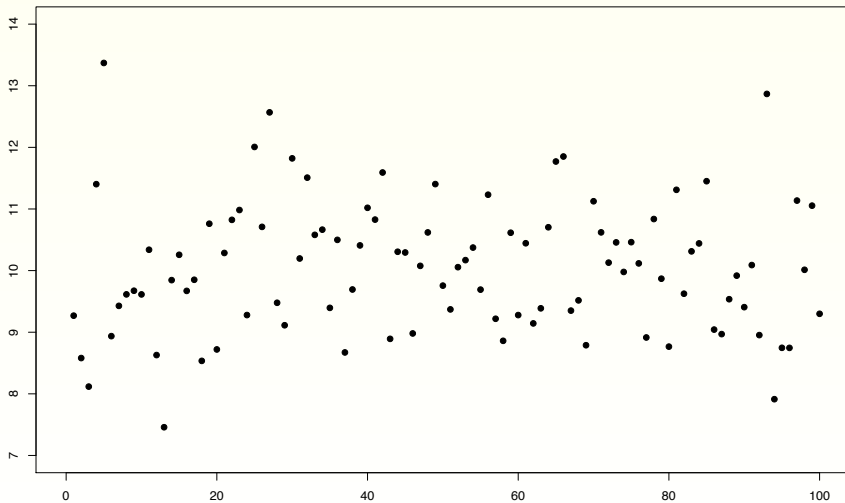
$$X_1, X_2, \dots, X_n \stackrel{i.i.d.}{\sim} F$$

Excès

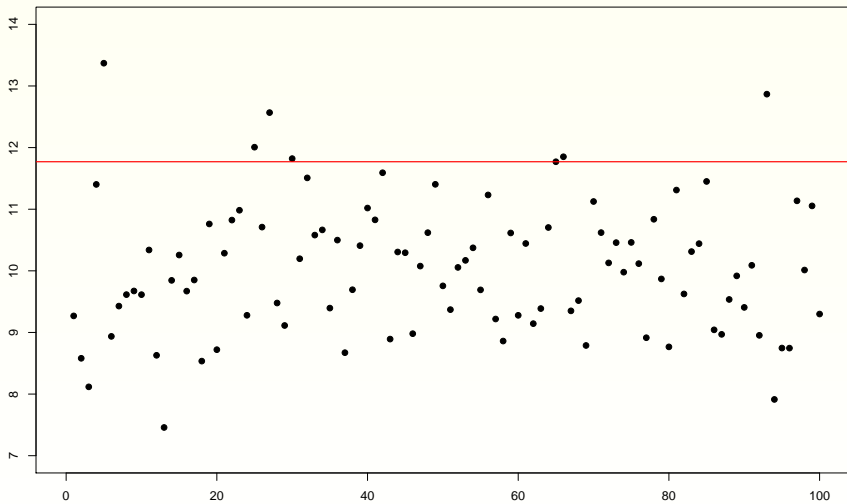
Plutôt que de se focaliser sur le maximum, on étudie les r_u valeurs dépassant un certain seuil donné u .

L'excès Y de la variable X au dessus du seuil u est défini par $X - u$ si $X > u$.

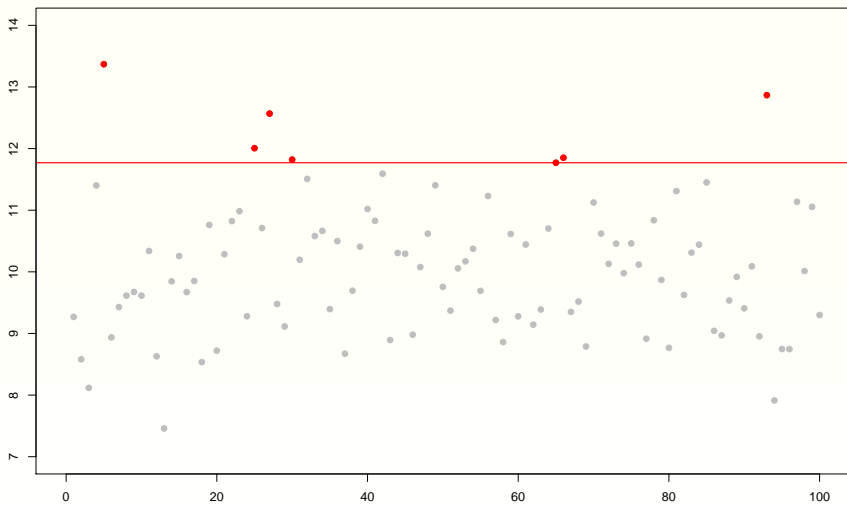
En pratique



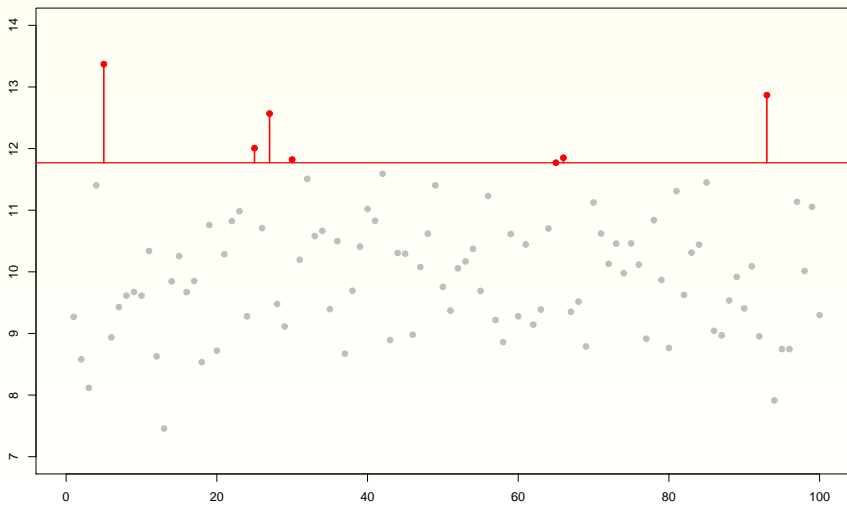
En pratique



En pratique



En pratique



Etude des excès

- La fonction de répartition F_u d'un excès au-dessus d'un seuil u est donnée pour $y > 0$ par (**Bayes !!**)

$$\begin{aligned} F_u(y) &= \mathbb{P}(Y \leq y | X > u) = \mathbb{P}(X - u \leq y | X > u) \\ &= \frac{\mathbb{P}(u < X \leq u + y)}{\mathbb{P}(X > u)} = \frac{F(u + y) - F(u)}{1 - F(u)} \end{aligned}$$

- La fonction de survie \bar{F}_u d'un excès au-dessus d'un seuil u est donnée pour $y > 0$ par

$$\bar{F}_u(y) = 1 - F_u(y) = \frac{1 - F(u + y)}{1 - F(u)}$$

- Lorsque le seuil u est suffisamment grand, on peut approcher cette quantité par la fonction de survie d'une loi de Pareto généralisée.

Loi de Pareto généralisée (GPD)

Sa fonction de survie est donnée par

$$\begin{aligned}\bar{H}_{\gamma,\sigma}(y) &= (1 + \gamma \frac{y}{\sigma})^{-1/\gamma} && \text{si } \gamma \neq 0 \\ &= \exp(-\frac{y}{\sigma}) && \text{sinon}\end{aligned}$$

Son ensemble de définition est \mathbb{R}^+ si $\gamma \geq 0$ ou $[0; -\sigma/\gamma[$ si $\gamma < 0$.

Elle dépend de deux paramètres :

- $\sigma > 0$ est un paramètre d'échelle,
- $\gamma \in \mathbb{R}$ est un paramètre de forme.

Deux cas particuliers :

- Le cas $\gamma = 0$ correspond à la loi exponentielle d'espérance σ ,
- $\gamma = -1$ correspond à la loi uniforme sur $[0; \sigma]$.

Approche POT - Théorème

Théorème de Pickands, 1975

Il y a équivalence entre la convergence en loi du maximum vers une GEV et la convergence en loi d'un excès vers une GPD :

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\frac{X_{n,n} - a_n}{b_n} \leq x \right) = G_\gamma(x),$$

si et seulement si

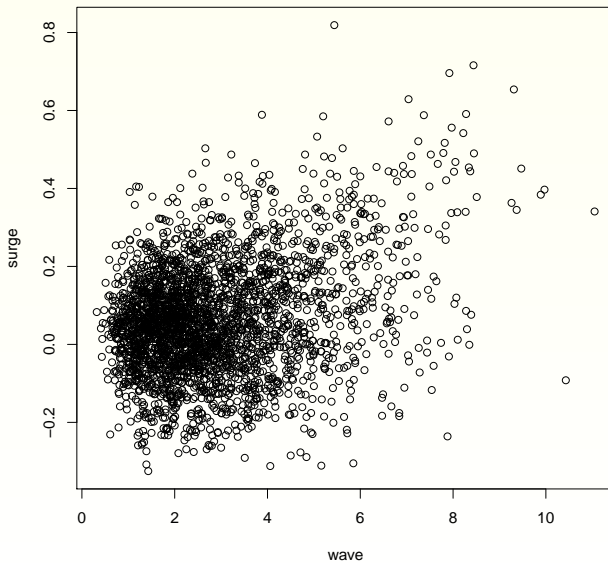
$$\lim_{u \rightarrow x_F} \sup_{y \in [0; x_F - u]} \left| \bar{F}_u(y) - \bar{H}_{\gamma, \sigma(u)}(y) \right| = 0.$$

Remarque : Le paramètre de forme γ est le même pour la GEV et pour la GPD.

Outline

- 1 Extrêmes univariés
- 2 **Extrêmes multivariés**
- 3 Approche par processus ponctuel et mesure spectrale

Données wave/surge Sud de l'Angleterre



Questions

- Qu'est-ce qu'un événement extrême multivarié ? Est-ce qu'il suffit que l'une des composantes soit extrême ? Est-ce que toutes doivent être extrêmes ? Si on a n composantes, est-ce que cela suffit que 2 soient extrêmes ?
- La notion fondamentale est celle de dépendance des extrêmes : comment les extrêmes de différentes variables sont-ils liés ?
- Les structures de dépendance des extrêmes sont-elles soumises à des contraintes particulières ? Sont-elles liées à la dépendance des observations "moyennes" ? Comment estimer la dépendance des extrêmes ?

Maximum par composant

- Soit $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ une suite de vecteurs aléatoires indépendants et identiquement distribués de même distribution que (X, Y) de fonction de répartition $F(x, y)$.
- Par exemple (X, Y) peut être la hauteur de précipitation en deux sites ou alors, comme sur l'exemple précédent deux variables d'un même phénomène comme la hauteur des vagues et la montée du niveau de la mer.
- Comme dans le cadre univarié, la Théorie des Extrêmes multivariés est basée sur la convergence des maxima. Posons :

$$M_{x,n} = \max_{i=1, \dots, n} \{X_i\} \quad \text{et} \quad M_{y,n} = \max_{i=1, \dots, n} \{Y_i\}$$

$$M_n = (M_{x,n}, M_{y,n})$$

M_n est le **vecteur des maxima par composant**. Le i pour lequel on a le maximum sur les X_i n'est pas forcément le même que l'indice pour lequel on a celui des Y_i . M_n **n'est donc pas nécessairement un point observé**.

Convergence

- Si on a les suites $(a_n)_n$ et $(b_n)_n$ telles que

$$F^n(a_n z + b_n) \xrightarrow{d} G(z) \quad \text{quand } n \rightarrow \infty$$

on dit que $a_n^{-1}(M_n - b_n)$ converge vers G , c'est à dire F est dans le domaine d'attraction de G , noté $F \in D(G)$ et G est dite loi des extrêmes bivariée.

⇒ 2 nouvelles questions :

- 1 Quelles sont les G possibles ?
- 2 Pour une F donnée, comment caractériser son domaine d'attraction ?

Marginales

- On peut déjà facilement caractériser les marginales possibles pour G :
comme G est non dégénérée, la convergence de F^n vers G revient à dire que l'on a la convergence pour les marginales correspondantes
(Cvg de vect. aléatoires \Rightarrow Cvg des composantes)
donc :

$$F_i^n(a_{n,i}z_i + b_{n,i}) \rightarrow G_i(z_i) \quad \text{pour } i = 1, 2$$

- Donc chaque G_i est elle même une loi univariée des extrêmes :

$$F_i \in D(G_i)$$

- On a les marginales, il ne manque plus que la structure de dépendance !
Pb : pas de représentation paramétrique comme dans le cas univarié...

Une première représentation (très proba)

- Balkema et Resnick ont montré que pour toute loi G max-infini divisible, il existe une mesure μ telle que :

$$G(z) = \exp(-\mu\{[-\infty, z]^c\})$$

- Cette mesure est la mesure exposant (exponent measure) : elle n'est pas unique en général. En plus ça n'est pas une représentation simple. Est-ce que l'on peut l'écrire différemment ? Oui, mais il faut commencer par se ramener à des marges semblables.

Réduction à des marginales fixées

- Pour étudier la structure de dépendance d'une loi max-stable, il est préférable de se ramener à des marginales semblables.
- Le choix de la distribution n'est pas important mais peut conduire à des résultats plus ou moins intéressants pour l'application. Le choix de marginales exponentielles, uniformes, Gumbel a été beaucoup étudié dans la littérature mais celui qui revient le plus souvent est le choix d'une loi de Fréchet de paramètre 1 :

$$\Psi_1(x) = \mathbb{P}(X_1 \leq z_1) = \exp(-1/z_1) \quad z_1 > 0$$

Proposition

Soit G une distribution des extrêmes bivariée, de marginales G_1 et G_2 . Si $(X, Y) \sim G$, alors :

$$\left(-\frac{1}{\log G_1(x)}, -\frac{1}{\log G_2(x)} \right) \sim G^*$$

où G^* est une loi bivariée extrême de marginales G_1^* et G_2^* Fréchet standard, c'est à dire Ψ_1 et Ψ_2

Caractérisation

Théorème

Soit (X_i, Y_i) des vecteurs aléatoires dont les marges sont Fréchet et $(M_{x,n}^*, M_{y,n}^*)$ défini par

$$M_n^* = (M_{x,n}^*, M_{y,n}^*) = \left(\max_{i=1, \dots, n} \{X_i\}/n, \max_{i=1, \dots, n} \{Y_i\}/n \right)$$

si

$$\mathbb{P}(M_{x,n}^* \leq x, M_{y,n}^* \leq y) \xrightarrow{\mathcal{L}} G(x, y) \text{ (en fait } G^*(x, y))$$

où G est une distribution non dégénérée, alors G est de la forme :

$$G(x, y) = \exp(-V(x, y)), \quad x > 0, y > 0$$

où

$$V(x, y) = 2 \int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) dH(\omega)$$

et H est une distribution sur $[0, 1]$ qui vérifie la contrainte de moyenne suivante :

$$\int_0^1 \omega dH(\omega) = 1/2$$

Remarques

Rappel

$$G(x, y) = \exp(-V(x, y)), \quad V(x, y) = 2 \int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) dH(\omega)$$

$$\int_0^1 \omega dH(\omega) = 1/2$$

- La famille de distributions limites G est la classe des **distributions des extrêmes bivariés**.
- Une conséquence du théorème, est qu'il y a une équivalence entre les distributions bivariées extrêmes et l'ensemble des fonctions H sur $[0, 1]$ qui vérifie la contrainte de moyenne. Si H admet une densité h , on a :

$$V(x, y) = 2 \int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) h(\omega) d\omega$$

Exemple de distribution H

- Si H est une mesure qui met une masse 0,5 en 0 et 1, c'est à dire :

$$H(0) = \frac{1}{2} \text{ et } H(1) = \frac{1}{2}$$

on a bien :

$$\int_0^1 \omega dH(\omega) = 0.H(0) + 1.H(1) = 1/2$$

et on a alors :

$$\begin{aligned} V(x, y) &= 2 \int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) dH(\omega) \\ &= 2 \left[\frac{1}{2} \max\left(0, \frac{1}{y}\right) + \frac{1}{2} \max\left(\frac{1}{x}, 0\right) \right] \\ &= x^{-1} + y^{-1} \end{aligned}$$

- La distribution des extrêmes bivariée est alors :

$$G(x, y) = \exp\left(-(x^{-1} + y^{-1})\right), \quad x > 0, y > 0$$

ou encore :

$$G(x, y) = \exp(-x^{-1}) \exp(-y^{-1}), \quad x > 0, y > 0$$

Autre exemple de distribution H

- Si H est une mesure qui met une masse 1 en 0,5 c'est à dire :

$$H(0,5) = 1$$

on a bien :

$$\int_0^1 \omega dH(\omega) = 0,5.H(0,5) = 1/2$$

et on a alors :

$$\begin{aligned} V(x,y) &= 2 \int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) dH(\omega) \\ &= 2 \max\left(\frac{1}{2x}, \frac{1}{2y}\right) \\ &= \max\left(\frac{1}{x}, \frac{1}{y}\right) \end{aligned}$$

- La distribution des extrêmes bivariée est alors :

$$G(x,y) = \exp\left(\max\left(\frac{1}{x}, \frac{1}{y}\right)\right), \quad x > 0, y > 0$$

donc $G(x,y)$ se réduit à sa marginale et on a $X = Y$ p.s.

Exemple de Modélisation : Logistique (Gumbel, 1960)

- Si on considère la mesure H de densité h :

$$h(\omega) = \frac{1}{2} (\alpha^{-1} - 1) (\omega(1 - \omega))^{-1-1/\alpha} \left(\omega^{-1/\alpha} + (1 - \omega)^{-1/\alpha} \right)^{\alpha-2}$$

pour $0 < \omega < 1$. h est symétrique pour $\omega = 0,5$. On obtient alors :

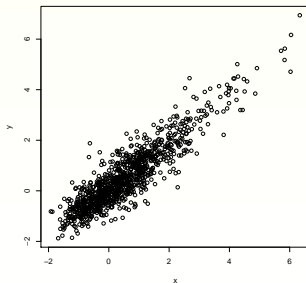
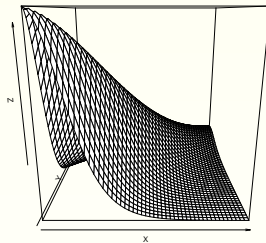
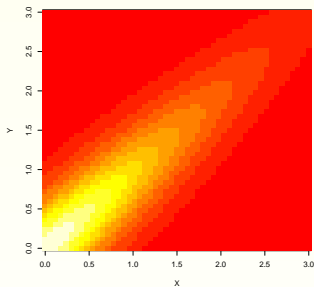
$$G(x, y) = \exp \left(- \left(x^{-1/\alpha} + y^{-1/\alpha} \right)^\alpha \right)$$

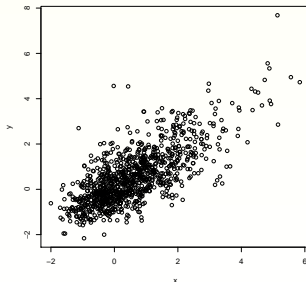
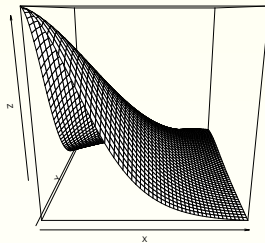
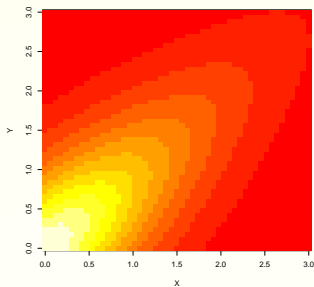
- Modèle apprécié car flexible :

- Quand $\alpha \rightarrow 1$ alors $G(x, y) \rightarrow \exp \left(- \left(\frac{1}{x} + \frac{1}{y} \right) \right)$ donc indépendance.
- Quand $\alpha \rightarrow 0$ alors $G(x, y) \rightarrow \exp \left(- \max \left(\frac{1}{x}, \frac{1}{y} \right) \right)$ donc dépendance parfaite car $G(x, y) = \min (G_1(x), G_2(y))$

⇒ “Couverture” complète de l'échelle des dépendances.

Limitation : X et Y doivent être échangeables car H symétrique.

Densité de G pour H logistique avec $\alpha = 0,3$ 

Densité de G pour H logistique avec $\alpha = 0,5$ 

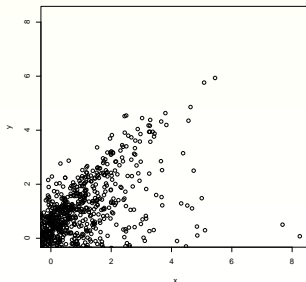
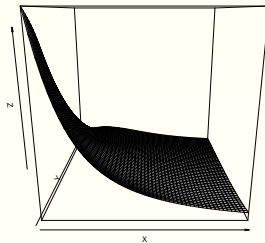
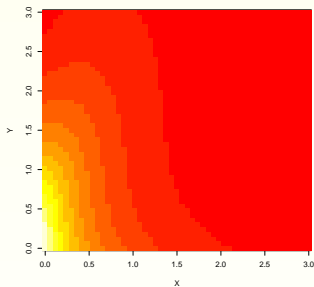
Exemple de Modélisation : Logistique asymétrique (Tawn, 1988)

- Une généralisation du modèle précédent est le modèle logistique asymétrique qui conduit à la forme suivante de loi bivariée des extrêmes :

$$G(x, y) = \exp \left\{ - (1 - t_1) x - (1 - t_2) y - \left[(t_1 x)^{1/\alpha} + (t_2 y)^{1/\alpha} \right]^\alpha \right\}$$

où $0 < r \leq 1$ et $0 \leq t_1, t_2 \leq 1$. Le paramètre α contrôle la dépendance et t_1 et t_2 l'asymétrie.

- Modèle flexible :
 - Quand $t_1 = t_2 = 1$, on retombe sur le modèle logistique.
 - Quand $\alpha = 1$ et $t_1 = t_2 = 0$ on tombe sur l'indépendance.
 - Quand $t_1 = t_2 = 1$ et $\alpha \rightarrow 0$, on va la dépendance parfaite.
 - On obtient différentes limites quand t_1 et t_2 sont fixés et $\alpha \rightarrow 0$.

Densité de G logistique asymétrique avec $\alpha = 0,3$, $t_1 = 0,5$ et $t_2 = 1$ 

Outline

- 1 Extrêmes univariés
- 2 Extrêmes multivariés
- 3 Approche par processus ponctuel et mesure spectrale

Processus ponctuel spatial

L'approche par processus ponctuel inclut une interprétation de la mesure H :

Théorème

Soit $(x_1, y_1), (x_2, y_2), \dots$ une suite d'observations indépendantes bivariées provenant d'une distribution admettant des marginales Fréchet standard et satisfaisant la convergence du maximum par composantes :

$$\mathbb{P}(M_{x,n}^* \leq x, M_{y,n}^* \leq y) \longrightarrow G^*(x, y).$$

Soit $\{N_n\}$ la suite de processus ponctuels définis par :

$$N_n = \{(n^{-1}x_1, n^{-1}y_1), (n^{-1}x_2, n^{-1}y_2), \dots, (n^{-1}x_n, n^{-1}y_n)\}$$

Alors

$$N_n \xrightarrow{d} N$$

sur les régions bornées (inférieurement) par l'origine $(0, 0)$ et N est un processus de Poisson non-homogène sur $(0, \infty) \times (0, \infty)$

Processus ponctuel spatial

Théorème (suite)

Si en plus on pose

$$r = x + y \quad \text{et} \quad \omega = \frac{x}{x + y}$$

l'intensité du processus N est :

$$\lambda(r, \omega) = 2 \frac{dH(\omega)}{r^2},$$

où H et G sont liées par le premier théorème de caractérisation.

Pourquoi ça sert ?

- Pour comprendre pourquoi c'est utile, il faut d'abord comprendre la transformation :

$$r = x + y \quad \text{et} \quad \omega = \frac{x}{x + y}$$

qui est une transformation de coordonnées cartésiennes vers des coordonnées pseudo-polaires, où r donne la distance du point à l'origine et ω mesure l'angle sur une échelle $[0, 1]$. En particulier $\omega = 0$ correspond à l'axe des x et $\omega = 1$ à celui des y .

- L'équation sur l'intensité du processus :

$$\lambda(r, \omega) = 2 \frac{dH(\omega)}{r^2},$$

nous dit que cette intensité du processus limite N sépare les composantes radiales et angulaires.

Pourquoi ça sert ?

- En d'autres termes, la partie angulaire des points de N est déterminée par H et est indépendante de la distance radiale.
- Du coup ça explique beaucoup de la mystérieuse formule :

$$V(x, y) = 2 \int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) dH(\omega)$$

H est là pour expliquer la part angulaire des points dans le processus ponctuel limite.

- L'interprétation est plus facile si H est dérivable. Si ω mesure l'angle, h détermine la fréquence relative des événements dans cette direction. En particulier si les extrêmes sont presque indépendants, on s'attend à ce que les grandes valeurs de x/n apparaissent avec des petites de y/n et vice-versa.

Pourquoi ça sert ?

- Dans ce cas de la presque indépendance, on s'attend à ce que $h(\omega)$ soit grand pour les valeurs de ω proches de 0 et de 1 et très petite partout ailleurs.
- Au contraire, si les extrêmes sont presque dépendants, on s'attend à ce que x/n et y/n soient très similaires. Dans ce cas, on s'attend à ce que $h(\omega)$ soit grand pour ω proche de 1/2 et petit partout ailleurs.
- C'est pour ces deux cas en particulier que l'on garde la formule générale

$$\lambda(r, \omega) = 2 \frac{dH(\omega)}{r^2},$$

plutôt que supposer que H admet une densité : dans le cas de la dépendance complète $H(0, 5) = 1$ et dans le cas de l'indépendance complète $H(0) = H(1) = 1/2$.

Remarque

- En fait, on peut utiliser n'importe quelle transformation bijective $\mathbb{R}_+^{*2} \rightarrow (0, \infty) \times [0, C]$ pour un certain $c > 0$ telle que

$$\begin{cases} r = r(x, y) \\ d = d(x, y) \end{cases}$$

qui vérifie la propriété pour $a, x, y > 0$:

$$\begin{cases} r(ax, ay) = ar(x, y) \\ d(ax, ay) = d(x, y) \end{cases}$$

(c'est à dire l'angle ne varie pas si on multiplie le point par a). Un autre exemple de transformation est celle en coordonnées polaires :

$$\begin{cases} r(x, y) = \sqrt{x^2 + y^2} \\ d(x, y) = \arctan \frac{x}{y} \end{cases}$$

Dans ce cas là, on peut réécrire G sous la forme :

$$G(x, y) = \exp \left(- \int_0^{\pi/2} \max \left(\frac{\cos \theta}{x}, \frac{\sin \theta}{y} \right) \psi(d\theta) \right)$$

Remarque

- ψ est définie différemment : Soit

$$B_{r,\theta} = \left\{ (x, y) \in \mathbb{R}_+^{*2} : \sqrt{x^2 + y^2} > r \text{ et } \arctan \frac{y}{x} \leq \theta \right\}$$

Il apparaît que

$$B_{r,\theta} = r B_{1,\theta}$$

et donc

$$V(B_{r,\theta}) = r^{-1} V(B_{1,\theta})$$

Comme dans la précédente transformation, on obtient une mesure produit.
Pour $0 \leq \theta \leq \frac{\pi}{2}$, on définit ψ par :

$$\psi(\theta) = V(B_{1,\theta})$$

En pratique

- En pratique, on suppose qu'on a une bonne approximation du processus de Poisson limite par le Processus N_n sur des régions appropriées. En particulier, on sait que la convergence est assurée sur les régions bornées par l'origine. Les choses sont beaucoup plus simples si on considère des régions du type :

$$A = \{(x, y) : x/n + y/n > r_0\}$$

pour un certain r_0 .

- Dans ce cas :

$$\begin{aligned}\Lambda(A) &= 2 \int_A \frac{dr}{r^2} dH(\omega) = 2 \int_{r=r_0}^{\infty} \frac{dr}{r^2} \int_{\omega=0}^1 \\ &= DH(\omega) = 2/r_0,\end{aligned}$$

qui est constant par rapport aux paramètres de H .

En pratique

- En supposant que H admette un densité h , la vraisemblance s'écrit :

$$\begin{aligned} L(\theta; (x_1, y_1), \dots, (x_n, y_n)) &= \exp(-\Lambda(A)) \prod_{i=1}^{N_A} \lambda(x_{(i)}/n, y_{(i)}/n) \\ &\propto \prod_{i=1}^{N_A} h(\omega_i), \end{aligned}$$

où $\omega_i = x_{(i)}/(x_{(i)} + y_{(i)})$ pour les N_A points $(x_{(i)}, y_{(i)})$ qui sont dans A .

- On a supposé que les marginales étaient Fréchet mais on peut faire l'estimation conjointement (dépendance-marginales) mais ça rend la vraisemblance beaucoup plus compliquée.

Autre point de vue

- Au lieu de regarder la distribution limite, on peut directement une distribution Multivariate Regularly Varying .
- Soit X un vecteur bidimensionnel. Pour simplifier, on suppose que X admet des marginales Fréchet unitaire. On dit que X est multivarié à variations régulières si il existe une mesure de Radon non nulle ν sur \mathcal{B} , l'ensemble des boréliens de $[-\infty, \infty]^2 \setminus \{0\}$ telle que $\nu((-\infty, \infty)^2 \setminus \{0\}) > 0$ et une fonction normalisante a avec $a(t) \rightarrow \infty$ tq

$$t\mathbb{P}(X/a(t) \in \cdot) \xrightarrow{v} \nu(\cdot)$$

quand $n \rightarrow \infty$ et \xrightarrow{v} est la convergence vague.

Une autre écriture de la mesure spectrale

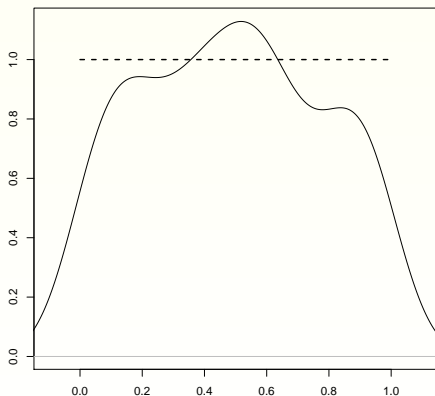
- Une autre écriture fait apparaître le rôle de la mesure spectrale :
- X est à variations régulières si on peut écrire :

$$\lim_{r \rightarrow \infty} \frac{P(\|\mathbf{X}\| > y, \mathbf{X} / \|\mathbf{X}\| \in \cdot)}{P(\|\mathbf{X}\| > r)} = y^{-\alpha} H(\cdot)$$

ce qui nous laisse l'opportunité d'estimer cette mesure spectrale : on se fixe un rayon suffisamment élevé, on récupère les angles pour les observations au delà du rayon et on estime la densité de ces angles.

Interlude : estimation de la densité sur support fini

- Si on dispose de X_1, \dots, X_n des v.a.i.i.d. de loi uniforme sur $[0, 1]$. Si on veut estimer la densité par le noyau, on va avoir un souci ($n = 100$) :



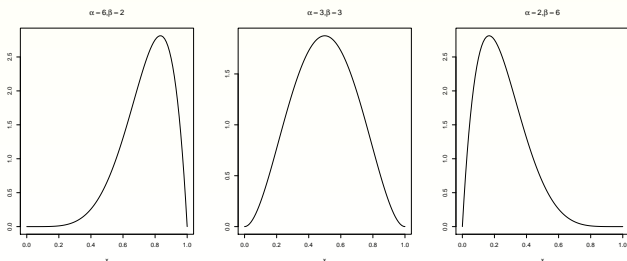
Solution : Les Beta Kernel

- Chen (1999) propose d'utiliser les Beta Kernel. On estime f par

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n k_{\beta_1}(Y_i, b, y)$$

où $b \rightarrow 0$ et $k_{\beta_1}(Y_i, b, y)$ est la densité d'une loi Beta "traffiquée" :

$$\begin{aligned} k_{\beta_1}(u, b, t) &= k_{\frac{t}{b}+1, \frac{1-t}{b}+1}(u) \\ &= \frac{u^{t/b}(1-u)^{(1-t)/b}}{B\left(\frac{t}{b}+1, \frac{1-t}{b}+1\right)} \mathbb{1}_{\{0 \leq u \leq 1\}} \end{aligned}$$



Ou les Beta Kernel modifié

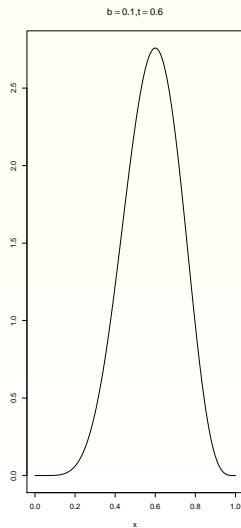
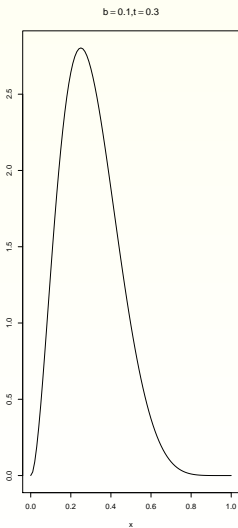
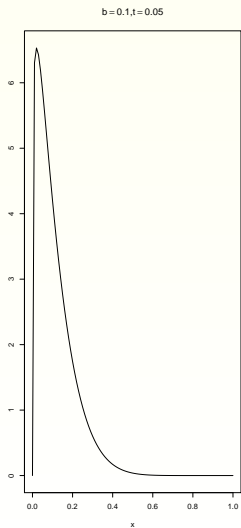
- Pour réduire le biais au bord, on peut utiliser les Beta Kernel modifié :

$$k_{\beta_2}(u, b, t) = \begin{cases} k_{\frac{t}{b}, \frac{1-t}{b}}(u), & \text{si } t \in [2b, 1 - 2b] \\ k_{\rho_b(t), \frac{1-t}{b}}(u), & \text{si } t \in [0, 2b) \\ k_{\frac{t}{b}, \rho_b(1-t)}(u), & \text{si } t \in (1 - 2b, 1] \end{cases}$$

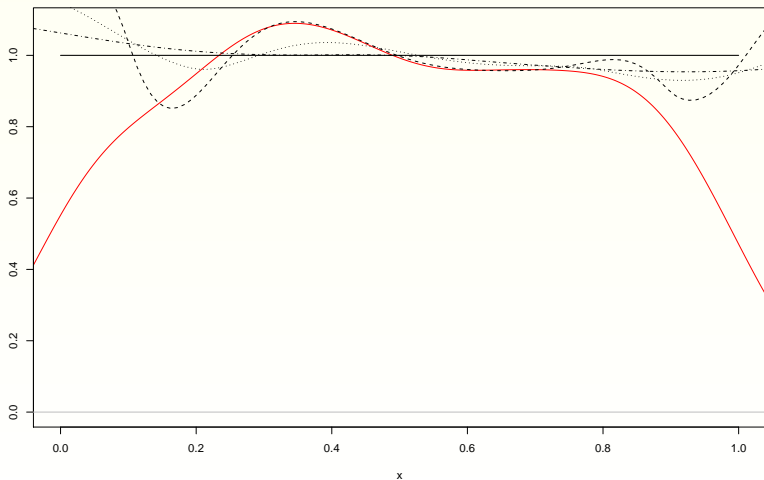
avec $\rho_b(t) = 2b^2 + 2.5 - \sqrt{4b^4 + 6b^2 + 2.25 - t^2 - t/b}$.

- Quand on estime une densité sur support connu (copules...), il vaut mieux utiliser les Beta Kernel modifié même si pour les petites tailles d'échantillons ($n = 30, 50$), la masse totale peut être assez éloigné de 1 donc utiliser des versions normalisées.

Beta Kernel Modifié



Au niveau de l'estimation de la densité

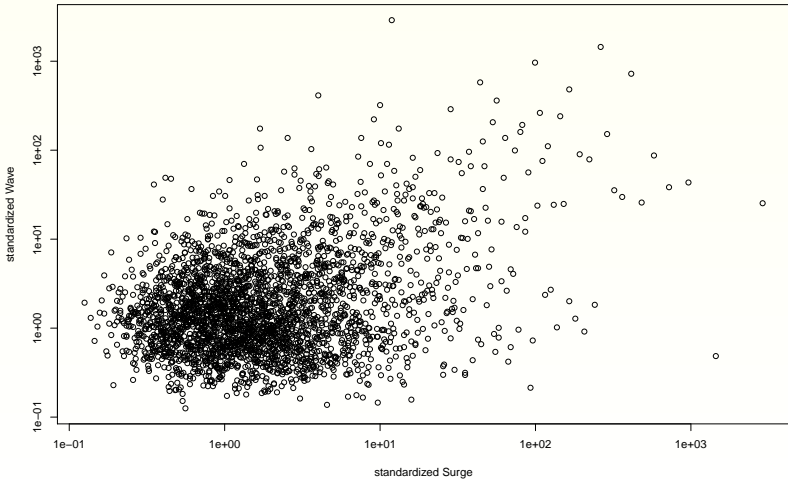


Retournons à la densité spectrale

- Si on a un échantillon X_1, \dots, X_n de distribution MRV de marginales Fréchet (à une transformation près), on commence par transformer tout le monde en coordonnées polaires.
- On ne garde que les observations avec un rayon suffisamment élevé, i.e. $R_i > r$ ou on ne garde que les k observations avec le plus grand rayon.
- On construit l'estimateur par Beta Kernel modifié de la densité angles ayant les plus grands rayons : $\hat{h}(\cdot)$.
- Si k le nombre d'observations conservées avec $k \rightarrow \infty$ et $n/k \rightarrow \infty$, et les conditions classiques sur la suite b_n , l'estimateur $\hat{h}(\cdot)$ est un estimateur consistant de $h(\cdot)$

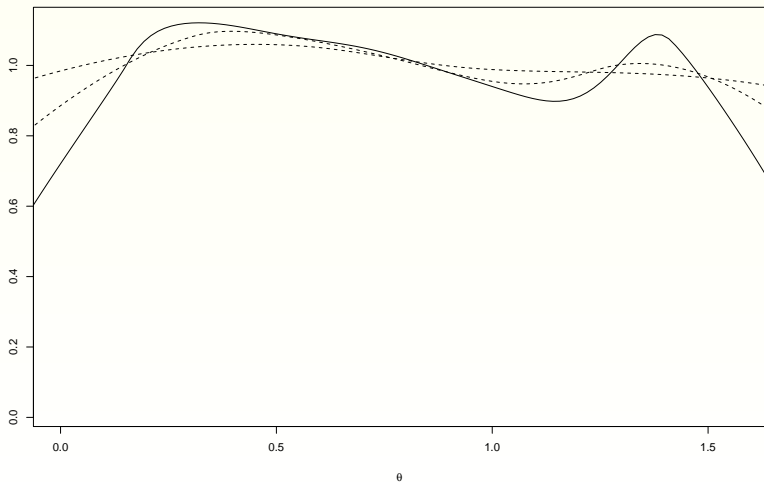
Données wave-surge

On transforme les données :



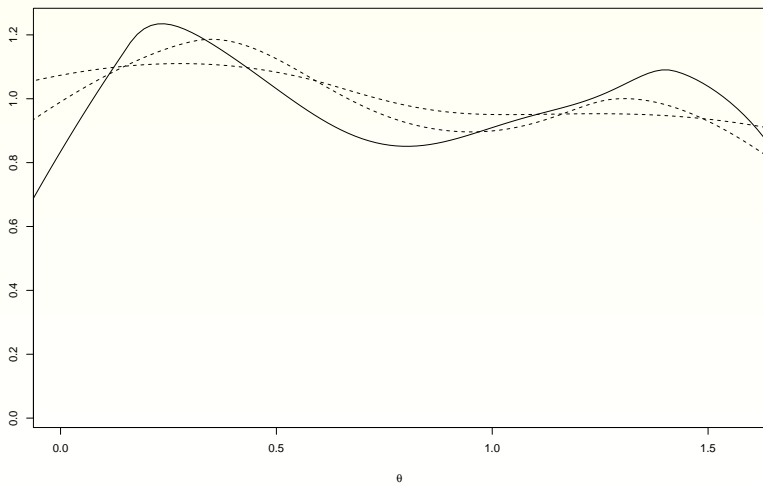
Estimation de la densité spectrale

On retient les 400 plus grands rayons.



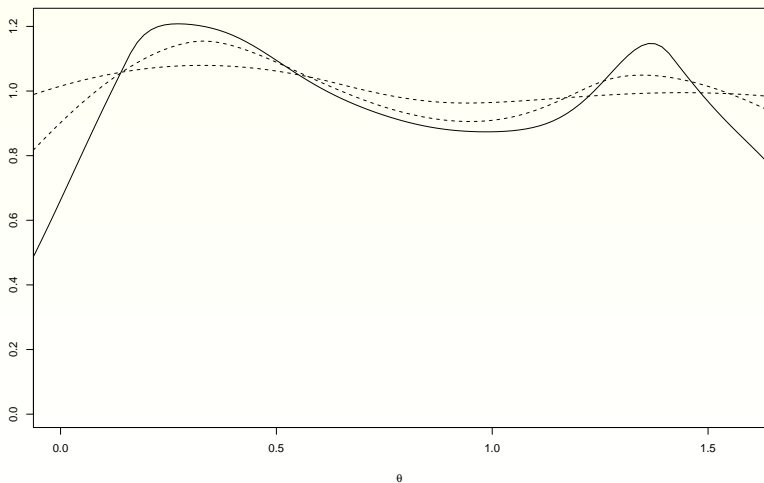
Estimation de la densité spectrale

On retient les 200 plus grands rayons.



Estimation de la densité spectrale

On retient les 100 plus grands rayons.



Conclusions

- Encore du travail : Normalité asymptotique, condition sur k et sur la fenêtre....
- Construire des tests d'adéquation
- Ondelettes ?

Publicité

- Projet MIRACCLE (Mesures et Indicateurs de Risques Adaptés au Changement CLimatiqueE : Thèse sur la définition de ces mesures de risques avec étude l'assurabilité des risques climatiques et l'impact sur les marchés financiers.
- Young researchers school on analyzing extreme events distributions du 5 au 8 Septembre 2010 à Aussois.
- EVA Lyon 2011.