

INTRODUCTION À L'ESTIMATION NON-PARAMÉTRIQUE DES QUANTILES EXTRÊMES CONDITIONNELS

Alexandre LEKINA

Docteurant, Équipe Mistis, INRIA Grenoble-Rhône-Alpes & LJK
<http://mistis.inrialpes.fr/people/lekina>

3 septembre 2009

Troisièmes Rencontres des Jeunes Statisticiens

- 1 Motivation
- 2 Cadre d'étude
- 3 Méthode d'estimation
- 4 Estimateurs et lois asymptotiques
- 5 Illustration
- 6 Du modèle à plan fixe au modèle à plan aléatoire

Hydrologie

En fonction de la localisation géographique x , les précipitations sont modélisées par une variable aléatoire Y . On dispose de Y_1, \dots, Y_n hauteurs de précipitations annuelles respectivement en n coordonnées x_1, \dots, x_n déterministes.

Hydrologie

En fonction de la localisation géographique x , les précipitations sont modélisées par une variable aléatoire Y . On dispose de Y_1, \dots, Y_n hauteurs de précipitations annuelles respectivement en n coordonnées x_1, \dots, x_n déterministes.

Problème :

- Calculer pour une probabilité $p < 1/n$, une hauteur de précipitation h extrême en un point x vérifiant $\mathbb{P}(Y > h|x) = p$.

Hydrologie

En fonction de la localisation géographique x , les précipitations sont modélisées par une variable aléatoire Y . On dispose de Y_1, \dots, Y_n hauteurs de précipitations annuelles respectivement en n coordonnées x_1, \dots, x_n déterministes.

Problème :

- Calculer pour une probabilité $p < 1/n$, une hauteur de précipitation h extrême en un point x vérifiant $\mathbb{P}(Y > h|x) = p$.

Difficultés :

- La fonction de survie conditionnelle $\mathbb{P}(Y > h|x)$ est inconnue et difficile à estimer au-delà de l'observation maximale.
- La hauteur de précipitation h est une fonction de x .

Modèle à plan fixe (design fixe)

- $Y \in \mathbb{R}$ est une v.a associée à une covariable non-aléatoire $x \in E$.
- E désigne un espace métrique muni d'une distance d .
- E peut être **de dimension infinie**.

Modèle à plan fixe (design fixe)

- $Y \in \mathbb{R}$ est une v.a associée à une covariable non-aléatoire $x \in E$.
- E désigne un espace métrique muni d'une distance d .
- E peut être **de dimension infinie**.

But

Estimer pour tout $x \in E$ et tout $\alpha \rightarrow 0$ le réel $q(\alpha, x)$ défini par

$$\mathbb{P}(Y > q(\alpha, x) | x) = \alpha,$$

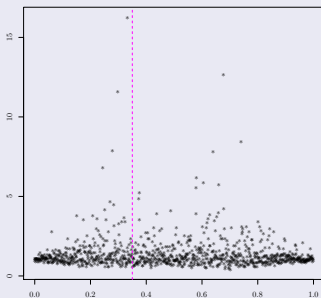
lorsque la loi conditionnelle est à **queue lourde**, i.e pour tout $\lambda > 0$, (se référer à Bingham, Goldie et Teugels (1987))

$$\lim_{\alpha \rightarrow 0} \frac{q(\lambda\alpha, x)}{q(\alpha, x)} = \lambda^{-\gamma(x)}, \quad (1)$$

- $\gamma(x) > 0$ est une fonction inconnue de la covariable x appelée "**indice des valeurs extrêmes conditionnel**".
- $q(\alpha, x)$: "**quantile extrême conditionnel**" d'ordre $1 - \alpha$.

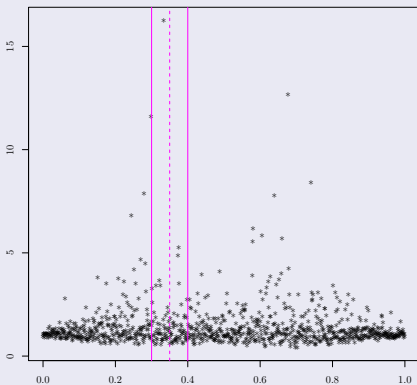
Méthode

- On dispose d'un échantillon $\{(Y_i, x_i), i = 1, \dots, n\}$ d'observations indépendantes.
- On veut construire en un point $x \in E$ un estimateur du quantile conditionnel $q(\alpha, x)$ lorsque $\alpha \rightarrow 0$.
- Comme l'illustre ce graphique, en ordonnée on a la variable d'intérêt et en abscisse la covariable. Ici $E = [0, 1]$ et $x = 0.37$ (en trait interrompu).



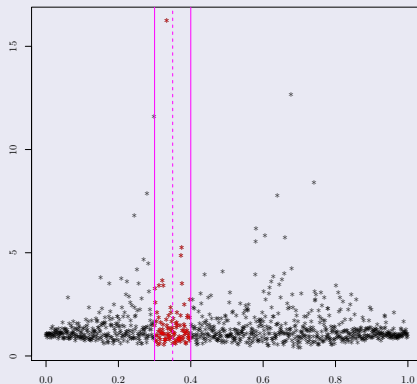
Méthode

- On utilise une méthode de fenêtres mobiles.
- $B(x, r_{n,x})$: le rayon de la boule est tel que $r_{n,x} \rightarrow 0$ quand $n \rightarrow \infty$.



Méthode

- On ne sélectionne que les observations Y_i pour lesquelles $x_i \in B(x, r_{n,x})$.
- On note par $\{Z_i(x), i = 1, \dots, m_{n,x}\}$ ces observations et $Z_{1,m_{n,x}}(x) \leq \dots \leq Z_{m_{n,x},m_{n,x}}(x)$ les statistiques ordonnées correspondantes.

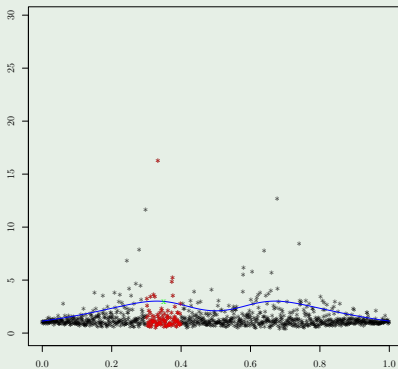


- Trois situations en fonction de la vitesse de convergence de $\alpha_{m_n, x}$ vers 0.

Situation (S.1)

Convergence "lente" de $\alpha_{m_{n,x}}$ vers 0, i.e $\alpha_{m_{n,x}} \rightarrow 0$ et $m_{n,x}\alpha_{m_{n,x}} \rightarrow \infty$.

Quantile théorique d'ordre $\alpha = 10/100 = 0.1$



Situation (S.1) Convergence "lente" : $\alpha_{m_{n,x}} \rightarrow 0$ et $m_{n,x}\alpha_{m_{n,x}} \rightarrow \infty$.

Estimateur

$$\hat{q}_1(\alpha_{m_{n,x}}, x) = Z_{m_{n,x} - \lfloor m_{n,x}\alpha_{m_{n,x}} \rfloor + 1, m_{n,x}}(x). \quad (2)$$

Loi asymptotique (Gardes, Girard & Lekina (2009))

Sous l'hypothèse que $q(\cdot, x)$ soit à variations régulières d'indice $-\gamma(x)$ (cf. équation (1)) et sous des conditions de régularités (du quantile conditionnel)

$$(m_{n,x}\alpha_{m_{n,x}})^{1/2} \left(\frac{\hat{q}_1(\alpha_{m_{n,x}}, x)}{q(\alpha_{m_{n,x}}, x)} - 1 \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \gamma^2(x)).$$

Situation (S.1) Convergence "lente" : $\alpha_{m_{n,x}} \rightarrow 0$ et $m_{n,x}\alpha_{m_{n,x}} \rightarrow \infty$.

Estimateur

$$\hat{q}_1(\alpha_{m_{n,x}}, x) = Z_{m_{n,x} - \lfloor m_{n,x}\alpha_{m_{n,x}} \rfloor + 1, m_{n,x}}(x). \quad (2)$$

Loi asymptotique (Gardes, Girard & Lekina (2009))

Sous l'hypothèse que $q(\cdot, x)$ soit à variations régulières d'indice $-\gamma(x)$ (cf. équation (1)) et sous des conditions de régularités (du quantile conditionnel)

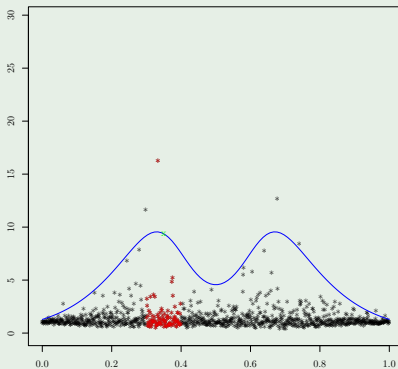
$$(m_{n,x}\alpha_{m_{n,x}})^{1/2} \left(\frac{\hat{q}_1(\alpha_{m_{n,x}}, x)}{q(\alpha_{m_{n,x}}, x)} - 1 \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \gamma^2(x)).$$

- La variance asymptotique est inversement proportionnelle à $\alpha_{m_{n,x}}$.
- La variance asymptotique croît avec l'indice de queue $\gamma(x)$.

Situation (S.2)

Convergence "rapide" de $\alpha_{m_{n,x}}$ vers 0, i.e $\alpha_{m_{n,x}} \rightarrow 0$ et $m_{n,x}\alpha_{m_{n,x}} \rightarrow c \geq 1$.

Quantile théorique d'ordre $\alpha = 1/100 = 0.01$



Situation (S.2) Convergence "rapide" : $\alpha_{m_{n,x}} \rightarrow 0$ et $m_{n,x}\alpha_{m_{n,x}} \rightarrow c \geq 1$.

Estimateur

$$\hat{q}_1(\alpha_{m_{n,x}}, x) = Z_{m_{n,x} - \lfloor m_{n,x}\alpha_{m_{n,x}} \rfloor + 1, m_{n,x}}(x).$$

Loi asymptotique (Gardes, Girard & Lekina (2009))

Sous l'hypothèse que $q(\cdot, x)$ soit à variations régulières d'indice $-\gamma(x)$ et sous des conditions de régularités (du quantile conditionnel)

$$\left(\frac{\hat{q}_1(\alpha_{m_{n,x}}, x)}{q(\alpha_{m_{n,x}}, x)} - 1 \right) \xrightarrow{\mathcal{L}} \mathcal{E}(c, \gamma(x))$$

où $\mathcal{E}(c, \gamma(x))$ est une loi non dégénérée et **non gaussienne**.

Situation (S.2) Convergence "rapide" : $\alpha_{m_{n,x}} \rightarrow 0$ et $m_{n,x}\alpha_{m_{n,x}} \rightarrow c \geq 1$.

Estimateur

$$\hat{q}_1(\alpha_{m_{n,x}}, x) = Z_{m_{n,x} - \lfloor m_{n,x}\alpha_{m_{n,x}} \rfloor + 1, m_{n,x}}(x).$$

Loi asymptotique (Gardes, Girard & Lekina (2009))

Sous l'hypothèse que $q(\cdot, x)$ soit à variations régulières d'indice $-\gamma(x)$ et sous des conditions de régularités (du quantile conditionnel)

$$\left(\frac{\hat{q}_1(\alpha_{m_{n,x}}, x)}{q(\alpha_{m_{n,x}}, x)} - 1 \right) \xrightarrow{\mathcal{L}} \mathcal{E}(c, \gamma(x))$$

où $\mathcal{E}(c, \gamma(x))$ est une loi non dégénérée et **non gaussienne**.

- L'estimateur $\hat{q}_1(\alpha_{m_{n,x}}, x)$ n'est pas **constant** i.e

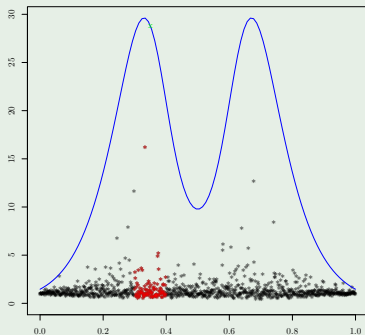
$$\frac{\hat{q}_1(\alpha_{m_{n,x}}, x)}{q(\alpha_{m_{n,x}}, x)} \text{ ne converge pas en probabilité vers } 1 \text{ [consistance faible].}$$

Situation (S.3)

Convergence "très rapide" de $\alpha_{m_{n,x}}$ vers 0, i.e

$$\alpha_{m_{n,x}} \rightarrow 0 \text{ et } m_{n,x} \alpha_{m_{n,x}} \rightarrow c \in [0, 1].$$

Quantile théorique d'ordre $\alpha = 0.1/100 = 0.001$



Situation **(S.3)** Convergence “très rapide” de $\alpha_{m_n, x}$ vers 0 :

$$\alpha_{m_n, x} \rightarrow 0 \text{ et } m_{n, x} \alpha_{m_n, x} \rightarrow c \in [0, 1[.$$

Estimateur (Weissman (1978))

$$\hat{q}_2(\alpha_{m_n, x}, x) = \hat{q}_1(\beta_{m_n, x}, x) \left(\frac{\beta_{m_n, x}}{\alpha_{m_n, x}} \right)^{\hat{\gamma}_n(x)}, \quad (3)$$

- $\beta_{m_n, x}$ satisfait **(S.1)**.
- $\hat{\gamma}_n(x)$ est un estimateur de l'indice des valeurs extrêmes conditionnel, (Beirlant et Goegebeur (2002) & Gardes et Girard (2008)).

Situation **(S.3)** Convergence “très rapide” de $\alpha_{m_n, x}$ vers 0 :

$$\alpha_{m_n, x} \rightarrow 0 \text{ et } m_{n, x} \alpha_{m_n, x} \rightarrow c \in [0, 1[.$$

Estimateur (Weissman (1978))

$$\hat{q}_2(\alpha_{m_n, x}, x) = \hat{q}_1(\beta_{m_n, x}, x) \left(\frac{\beta_{m_n, x}}{\alpha_{m_n, x}} \right)^{\hat{\gamma}_n(x)}, \quad (3)$$

- $\beta_{m_n, x}$ satisfait **(S.1)**.
- $\hat{\gamma}_n(x)$ est un estimateur de l'indice des valeurs extrêmes conditionnel, (Beirlant et Goegebeur (2002) & Gardes et Girard (2008)).
- L'estimateur $\hat{q}_2(\cdot, x)$ dépend non seulement de la statistique d'ordre $\hat{q}_1(\cdot, x)$ mais aussi de l'estimateur d'indice de queue $\hat{\gamma}_n(x)$.

Situation **(S.3)** Convergence “très rapide” de $\alpha_{m_{n,x}}$ vers 0 :

$$\alpha_{m_{n,x}} \rightarrow 0 \text{ et } m_{n,x} \alpha_{m_{n,x}} \rightarrow c \in [0, 1[.$$

Loi asymptotique (Gardes, Girard & Lekina (2009))

Si la suite $\beta_{m_{n,x}}$ satisfait **(S.1)** et s'il existe une suite positive $v_n(x)$ et une loi \mathcal{D} telle que

$$v_n(x) (\hat{\gamma}_n(x) - \gamma(x)) \xrightarrow{\mathcal{L}} \mathcal{D},$$

alors, deux situations se présentent :

(i) Soit la loi asymptotique est héritée de celle de $\hat{q}_1(\beta_{m_{n,x}}, x)$ et nous avons

$$(m_{n,x} \alpha_{m_{n,x}})^{1/2} \left(\frac{\hat{q}_2(\alpha_{m_{n,x}}, x)}{q(\alpha_{m_{n,x}}, x)} - 1 \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \gamma^2(x) \right).$$

(ii) Sinon, elle provient de $\hat{\gamma}_n(x)$ et nous avons

$$\frac{v_n(x)}{\log(\beta_{m_{n,x}}/\alpha_{m_{n,x}})} \left(\frac{\hat{q}_2(\alpha_{m_{n,x}}, x)}{q(\alpha_{m_{n,x}}, x)} - 1 \right) \xrightarrow{\mathcal{L}} \mathcal{D}.$$

Simulations

- La famille d'estimateurs proposée par Gardes et Girard (2008)

$$\hat{\gamma}_n(x, W) = \sum_{i=1}^{k_{n,x}} i \log \left(\frac{Z_{m_{n,x}-i+1, m_{n,x}}}{Z_{m_{n,x}-i, m_{n,x}}} \right) W(i/k_{n,x}, x) \Big/ \sum_{i=1}^{k_{n,x}} W(i/k_{n,x}, x),$$

- $W(\cdot, x)$ est une fonction de poids définie sur $]0, 1[$ dont l'intégrale vaut 1.
- $k_{n,x} = m_{n,x} \beta_{m_{n,x}}$.

Simulations

- La famille d'estimateurs proposée par Gardes et Girard (2008)

$$\hat{\gamma}_n(x, W) = \sum_{i=1}^{k_{n,x}} i \log \left(\frac{Z_{m_{n,x}-i+1, m_{n,x}}}{Z_{m_{n,x}-i, m_{n,x}}} \right) W(i/k_{n,x}, x) \Big/ \sum_{i=1}^{k_{n,x}} W(i/k_{n,x}, x),$$

- $W(\cdot, x)$ est une fonction de poids définie sur $]0, 1[$ dont l'intégrale vaut 1.
- $k_{n,x} = m_{n,x} \beta_{m_{n,x}}$.

Remarques

- Le choix des paramètres $r_{n,x}$ et $\beta_{m_{n,x}}$ est un problème délicat.

Simulations

- La famille d'estimateurs proposée par Gardes et Girard (2008)

$$\hat{\gamma}_n(x, W) = \sum_{i=1}^{k_{n,x}} i \log \left(\frac{Z_{m_{n,x}-i+1, m_{n,x}}}{Z_{m_{n,x}-i, m_{n,x}}} \right) W(i/k_{n,x}, x) \Big/ \sum_{i=1}^{k_{n,x}} W(i/k_{n,x}, x),$$

- $W(\cdot, x)$ est une fonction de poids définie sur $]0, 1[$ dont l'intégrale vaut 1.
- $k_{n,x} = m_{n,x} \beta_{m_{n,x}}$.

Remarques

- Le choix des paramètres $r_{n,x}$ et $\beta_{m_{n,x}}$ est un problème délicat.
- Si $W(\cdot, x) = 1$, alors

$$\hat{\gamma}_n(x) := \hat{\gamma}_n(x, 1) = \frac{1}{k_{n,x}} \sum_{i=1}^{k_{n,x}} i \log (Z_{m_{n,x}-i+1, m_{n,x}} / Z_{m_{n,x}-i, m_{n,x}}),$$

c'est l'estimateur de Hill (1975) adapté au cas conditionnel.

Simulations

- La famille d'estimateurs proposée par Gardes et Girard (2008)

$$\hat{\gamma}_n(x, W) = \frac{\sum_{i=1}^{k_{n,x}} i \log \left(\frac{Z_{m_{n,x}-i+1, m_{n,x}}}{Z_{m_{n,x}-i, m_{n,x}}} \right) W(i/k_{n,x}, x)}{\sum_{i=1}^{k_{n,x}} W(i/k_{n,x}, x)},$$

- $W(\cdot, x)$ est une fonction de poids définie sur $]0, 1[$ dont l'intégrale vaut 1.
- $k_{n,x} = m_{n,x} \beta_{m_{n,x}}$.

Remarques

- Le choix des paramètres $r_{n,x}$ et $\beta_{m_{n,x}}$ est un problème délicat.
- Si $W(\cdot, x) = 1$, alors

$$\hat{\gamma}_n(x) := \hat{\gamma}_n(x, 1) = \frac{1}{k_{n,x}} \sum_{i=1}^{k_{n,x}} i \log \left(Z_{m_{n,x}-i+1, m_{n,x}} / Z_{m_{n,x}-i, m_{n,x}} \right),$$

c'est l'estimateur de Hill (1975) adapté au cas conditionnel.

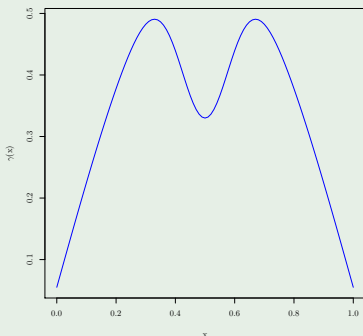
- En posant $W(s, x) = -\log(s)$, on retrouve l'estimateur de Zipf (Schultze et Steinebach (1996) & Kratz et Resnick (1996)) adapté au cas conditionnel.

Simulations

- Soit $E = [0, 1]$, on définit la fonction

$$x \in E \mapsto \gamma(x) = \frac{1}{2} \left(\frac{1}{10} + \sin(\pi x) \right) \left(\frac{11}{10} - \frac{1}{2} \exp(-64(x - 1/2)^2) \right).$$

Allure de la fonction $\gamma(x)$



Simulations

- On génère un **échantillon** $\{(Y_i, x_i), i = 1, \dots, n\}$ de taille $n = 1000$ selon le quantile conditionnel de **loi de Fréchet** :

$$q(\alpha, x) = \left\{ \log \left(\frac{1}{1 - \alpha} \right) \right\}^{-\gamma(x)}$$

Simulations

- On génère **un échantillon** $\{(Y_i, x_i), i = 1, \dots, n\}$ de taille $n = 1000$ selon le quantile conditionnel de **loi de Fréchet** :

$$q(\alpha, x) = \left\{ \log \left(\frac{1}{1 - \alpha} \right) \right\}^{-\gamma(x)}$$

- On choisit arbitrairement les paramètres $\beta_{m_n, x}$ et $r_{n, x}$:

$$\beta_{m_n, x} = 0.25 \text{ et } r_{n, x} = 0.1.$$

Simulations

- On génère **un échantillon** $\{(Y_i, x_i), i = 1, \dots, n\}$ de taille $n = 1000$ selon le quantile conditionnel de **loi de Fréchet** :

$$q(\alpha, x) = \left\{ \log \left(\frac{1}{1 - \alpha} \right) \right\}^{-\gamma(x)}$$

- On choisit arbitrairement les paramètres $\beta_{m_{n,x}}$ et $r_{n,x}$:

$$\beta_{m_{n,x}} = 0.25 \text{ et } r_{n,x} = 0.1.$$

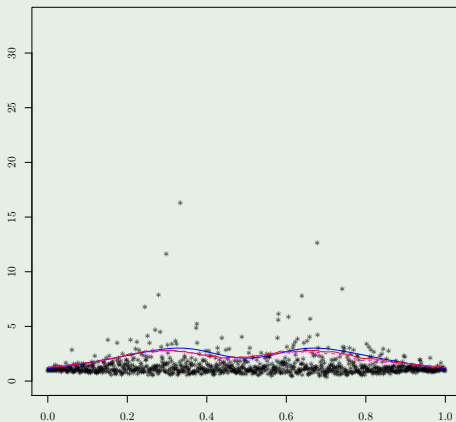
- On utilise l'estimateur

$$\hat{q}_2(\alpha_{m_{n,x}}, x) = \hat{q}_1(\beta_{m_{n,x}}, x) \left(\frac{\beta_{m_{n,x}}}{\alpha_{m_{n,x}}} \right)^{\hat{\gamma}_n(x)}.$$

- $W(s, x) = -\log(s)$ (en magenta) et $W(s, x) = 1$ (en rouge).

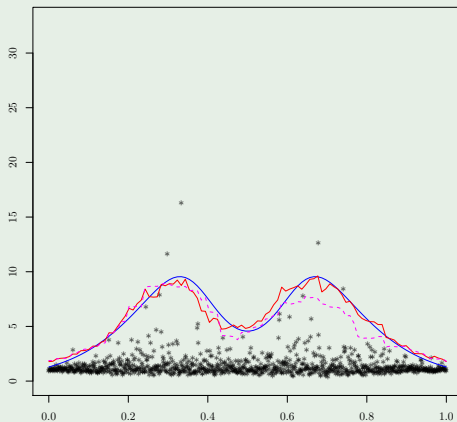
(S.1) Estimation du quantile théorique d'ordre $\alpha = 10/100 = 0.1$ avec $\hat{q}_2(\alpha, x)$

— $\hat{q}_2(\alpha, \cdot)$ avec $W(s, x) = -\log(s)$ et — $\hat{q}_2(\alpha, \cdot)$ avec $W(s, x) = 1$.



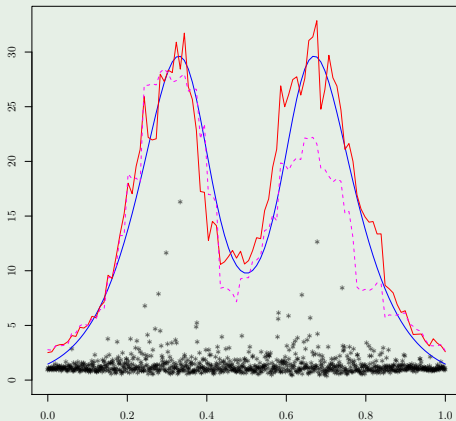
(S.2) Estimation du quantile théorique d'ordre $\alpha = 1/100 = 0.01$ avec $\hat{q}_2(\alpha, x)$

— $\hat{q}_2(\alpha, \cdot)$ avec $W(s, x) = -\log(s)$ et — $\hat{q}_2(\alpha, \cdot)$ avec $W(s, x) = 1$.



(S.3) Estimation du quantile théorique d'ordre $\alpha = 0.001$ avec $\hat{q}_2(\alpha, x)$

— $\hat{q}_2(\alpha, \cdot)$ avec $W(s, x) = -\log(s)$ et — $\hat{q}_2(\alpha, \cdot)$ avec $W(s, x) = 1$.



Introduction au modèle à plan aléatoire (design aléatoire)

But

- Soient (X_i, Y_i) , $i = 1, \dots, n$ des copies indépendantes du couple aléatoire $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$.
- Estimer pour tout $x \in \mathbb{R}^d$ et pour tout $\alpha_n \rightarrow 0$, les courbes de niveaux extrêmes définies comme les graphes de fonctions $x \in \mathbb{R}^d \rightarrow q(\alpha_n|x) \in \mathbb{R}$ vérifiant

$$\mathbb{P}(Y > q(\alpha_n|x)|X = x) = \alpha_n,$$

lorsque la fonction de répartition conditionnelle de Y sachant $X = x$ est à **variations à régulières** d'indice $-1/\gamma(x)$, i.e

$$q(\alpha_n|x) = \alpha_n^{-\gamma(x)} \ell(1/\alpha_n|x),$$

- $\gamma(\cdot) > 0$: “**indice de queue conditionnel**”, fonction inconnue de la covariable.
- $\ell(\cdot|x)$: “**fonction à variations lentes**”, i.e pour tout $\lambda > 0$,

$$\lim_{y \rightarrow \infty} \frac{\ell(\lambda y|x)}{\ell(y|x)} = 1.$$

- $\ell(\cdot|x)$ est normalisée.

Estimateur : l'inverse d'un estimateur de la fonction de survie conditionnelle

$$\hat{q}_n(\alpha_n|x) = \hat{F}_n^{\leftarrow}(\alpha_n|x) = \inf \left\{ t, \hat{F}_n(t|x) \leq \alpha_n \right\}.$$

Nécessite d'estimer la probabilité

$$\bar{F}(y_n|x) \text{ quand } y_n \rightarrow \infty \text{ avec } n \rightarrow \infty.$$

Estimateur à noyau de \bar{F} , (Collomb (1976))

$$\hat{\bar{F}}_n(y|x) = \frac{\frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) \mathbf{1}\{Y_i \geq y\}}{\frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)} = \frac{\hat{\psi}_n(y, x)}{\hat{g}_n(x)},$$

- La fonction $K(\cdot)$ est positive, bornée, intégrable et à support compact.
- Le paramètre de lissage $h_n \rightarrow 0$ quand $n \rightarrow \infty$.
- La fonction $\hat{g}_n(\cdot)$ est l'estimateur à noyau classique de la densité $g(\cdot)$ de X .
- La fonction $\hat{\psi}_n(y, x)$ est un estimateur de $\psi(y, x) = \bar{F}(y|x)g(x)$.

Estimateur à noyau \bar{F} , (Collomb (1976))

$$\hat{\bar{F}}_n(y|x) = \frac{\frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) \mathbf{1}\{Y_i \geq y\}}{\frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)} = \frac{\hat{\psi}_n(y, x)}{\hat{g}_n(x)},$$

Remarque

$$\frac{\hat{\bar{F}}_n(y_n|x)}{\bar{F}(y_n|x)} - 1 = \left[\frac{\hat{\psi}_n(y_n, x) - \mathbb{E}[\hat{\psi}_n(y_n, x)]}{\hat{g}_n(x)\psi(y_n, x)} \right] + \left[\frac{\mathbb{E}[\hat{\psi}_n(y_n, x)] - \psi(y_n, x)}{\hat{g}_n(x)\psi(y_n, x)} \right] \\ - \left[\frac{\hat{g}_n(x) - \mathbb{E}[\hat{g}_n(x)]}{\hat{g}_n(x)} \right] - \left[\frac{\mathbb{E}[\hat{g}_n(x)] - g(x)}{\hat{g}_n(x)} \right].$$

- La loi asymptotique de $\hat{\bar{F}}_n(y_n|x)$ dépend du comportement limite des termes aléatoires $\left[\frac{\hat{\psi}_n(y_n, x) - \mathbb{E}[\hat{\psi}_n(y_n, x)]}{\hat{g}_n(x)\psi(y_n, x)} \right]$ et $\left[\frac{\hat{g}_n(x) - \mathbb{E}[\hat{g}_n(x)]}{\hat{g}_n(x)} \right]$ lorsque $n \rightarrow \infty$.

Comportement asymptotique de l'estimateur $\hat{g}_n(\cdot)$

Si $nh_n^d \rightarrow \infty$ quand $n \rightarrow \infty$, alors, pour tout $x \in \mathbb{R}^d$,

• Biais asymptotique

$$\mathbb{E}[\hat{g}_n(x)] - g(x) = O(h_n).$$

• Distribution asymptotique

$$\left(nh_n^d\right)^{1/2} (\hat{g}_n(x) - \mathbb{E}[\hat{g}_n(x)]) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, g(x) \|K\|_2^2\right).$$

(voir Collomb (1976), Prop. 2.1 and 2.2).

Propriétés de $\hat{\psi}_n(\cdot, \cdot)$ (Daouia, Gardes, Girard & Lekina (2009))

Si

$y_n \rightarrow \infty$ tel que $h_n \log y_n \rightarrow 0$ et $nh_n^d \bar{F}(y_n|x) \rightarrow \infty$ lorsque $n \rightarrow \infty$,

alors,

pour tout $x \in \mathbb{R}^d$ et pour tout $j = 1, \dots, J$ tel que $y_{n,j} = a_j y_n (1 + o(1))$,

1. Biais asymptotique

$$\left\{ \mathbb{E} \left[\hat{\psi}_n(y_{n,j}, x) \right] \right\}_{\{j=1, \dots, J\}} = \left\{ \psi(y_{n,j}, x) (1 + O(h_n \log y_n)) \right\}_{\{j=1, \dots, J\}}.$$

2. Distribution asymptotique

$$\left\{ \sqrt{nh_n^d \psi(y_n|x)} \left(\frac{\hat{\psi}_n(y_{n,j}, x) - \mathbb{E} \left[\hat{\psi}_n(y_{n,j}, x) \right]}{\psi(y_{n,j}, x)} \right) \right\}_j \xrightarrow{\mathcal{L}} \mathcal{N} \left(\mathbf{0}_{\mathbb{R}^J}, \|K\|_2^2 C(x) \right)$$

où $C_{j,j'}(x) = a_{j \wedge j'}^{1/\gamma(x)} \forall (j, j') \in \{1, \dots, J\}^2$ avec $a_j > 0$.

Rappel

$$\frac{\hat{F}_n(y_n|x)}{\bar{F}(y_n|x)} - 1 = \left[\frac{\hat{\psi}_n(y_n, x) - \mathbb{E}[\hat{\psi}_n(y_n, x)]}{\hat{g}_n(x)\psi(y_n, x)} \right] + \left[\frac{\mathbb{E}[\hat{\psi}_n(y_n, x)] - \psi(y_n, x)}{\hat{g}_n(x)\psi(y_n, x)} \right] \\ - \left[\frac{\hat{g}_n(x) - \mathbb{E}[\hat{g}_n(x)]}{\hat{g}_n(x)} \right] - \left[\frac{\mathbb{E}[\hat{g}_n(x)] - g(x)}{\hat{g}_n(x)} \right].$$

Loi asymptotique de $\hat{F}_n(y_n|x)$ (Daouia, Gardes, Girard & Lekina (2009))

Si

$y_n \rightarrow \infty$ tel que $h_n \log y_n \rightarrow 0$ et $nh_n^d \bar{F}(y_n|x) \rightarrow \infty$ lorsque $n \rightarrow \infty$,

alors,

pour tout $x \in \mathbb{R}^d$ et pour tout $j = 1, \dots, J$ tel que $y_{n,j} = a_j y_n (1 + o(1))$,

$$\left\{ \sqrt{nh_n^d \bar{F}(y_n|x)} \left(\frac{\hat{F}_n(y_{n,j}|x)}{\bar{F}(y_{n,j}|x)} - 1 \right) \right\}_{\{j=1, \dots, J\}} \xrightarrow{\mathcal{L}} \mathcal{N} \left(\mathbf{0}_{\mathbb{R}^J}, \frac{\|K\|_2^2}{g(x)} C(x) \right).$$

Loi asymptotique de $\hat{q}_n(\alpha_n|x)$ (Daouia, Gardes, Girard & Lekina (2009))

Si,

$$\alpha_n \rightarrow 0 \text{ et } nh_n^d \alpha_n \rightarrow \infty \text{ quand } n \rightarrow \infty,$$

alors,

pour tout $x \in \mathbb{R}^d$ et pour $j = 1, \dots, J$ tel que $\alpha_{n,j} = \tau_j \alpha_n (1 + o(1))$,

$$\left\{ \sqrt{nh_n^d \alpha_n} \left(\frac{\hat{q}_n(\alpha_{n,j}|x)}{q(\alpha_{n,j}|x)} - 1 \right) \right\}_{\{j=1, \dots, J\}} \xrightarrow{\mathcal{L}} \mathcal{N} \left(\mathbf{0}_{\mathbb{R}^J}, \gamma^2(x) \frac{\|K\|_2^2}{g(x)} \Sigma \right),$$

où $\Sigma_{j,j'}(x) = 1/\tau_{j \wedge j'}$ pour $(j, j') \in \{1, \dots, J\}^2$ avec $\tau_j > 0$.

Remarques : La variance asymptotique

- est inversement proportionnelle à $nh_n^d \alpha_n$, l'estimation est d'autant plus stable qu'on s'éloigne de la frontière de l'échantillon ;
- est proportionnelle à $\gamma(x) \Rightarrow$ l'estimation de $q(\alpha_n|x)$ plus difficile pour les grandes valeurs de l'indice de queue conditionnel.

Applications et perspectives








- Afin de construire des intervalles de confiance du quantile extrême conditionnel $\hat{q}_n(\alpha_n|x)$, nous avons proposé un “**estimateur lissé de Pickands**” défini par

$$\hat{\gamma}_n(x) = \frac{1}{\log 2} \log \left(\frac{\hat{q}_n(k_n/n|x) - \hat{q}_n(2k_n/n|x)}{\hat{q}_n(2k_n/n|x) - \hat{q}_n(4k_n/n|x)} \right),$$








où la suite k_n satisfait la condition : $k_n \rightarrow \infty$ et $k_n/n \rightarrow 0$ qd $n \rightarrow \infty$.

- Nous avons établi la loi asymptotique de cet estimateur.
- De nouveaux estimateurs de $\hat{\gamma}_n(\cdot)$ sont en cours d'étude.
- Étude des cas (S.2) et (S.3).
- **Quelques problèmes ouverts** :
 - 1 design aléatoire : le choix du paramètre de lissage h_n et de la suite k_n ;
 - 2 design fixe : le choix du rayon de la boule $r_{n,x}$ et du seuil $k_{n,x}$.

Bibliographie

-  A. Daouia, L. Gardes, S. Girard and A. Lekina. Kernel estimators of extreme level curves. <http://hal.inria.fr/inria-00393588/fr/>, 2009.
-  J. Beirlant and Y. Goegebeur. Regression with response distributions of Pareto-type. *Computational Statistics and Data Analysis*, 42 :595–619, 2003.
-  A. Berlinet, A. Gannoun and E. Matzner-Løber. Asymptotic normality of convergent estimates of conditional quantiles. *Statistics*, 35 :139–169, 2001.
-  N.H. Bingham, C.M. Goldie and J.L. Teugels. *Regular Variation*, Cambridge University Press, 1987
-  G. Collomb. *Estimation non paramétrique de la régression par la méthode du noyau*. PhD thesis, Université Paul Sabatier de Toulouse, 1976.
-  L. Gardes and S. Girard. Moving window approach for nonparametric estimation of the conditional tail index, *Journal of Multivariate Analysis*, 99 :2368–2388, 2008.
-  L. Gardes, S. Girard and A. Lekina. Functional nonparametric estimation of conditional extreme quantiles, *Journal of Multivariate Analysis*, to appear, 2009.

Bibliographie (suite et fin)

-  B.M. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3 : 1163–1174, 1975.
-  M. Kratz and S. Resnick. The QQ-estimator and heavy tails. *Stochastic Models*, 12 :699–724, 1996.
-  J. Schultze and J. Steinebach. On least squares estimates of an exponential tail coefficient. *Statistics and Decisions*, 14 :353–372, 1996.
-  J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3 :119–131, 1975.
-  C.J. Stone. Consistent nonparametric regression (with discussion). *The Annals of Statistics*, 5 :595–645, 1977.
-  W. Stute. Conditional empirical processes. *The Annals of Statistics*, 14 :638–647, 1986.
-  I. Weissman. Estimation of parameters and large quantiles based on the k -largest observations, *Journal of the American Statistical Association*, 73 :812–815, 1978.