

- 1 Introduction
 - En l'absence de covariable
 - En présence de covariable

- 2 Estimation des courbes de niveaux extrêmes
 - Modèle et méthode d'estimation
 - Hypothèses et résultats asymptotiques
 - Application à l'estimation de l'indice de queue conditionnel
 - Application à l'estimation des quantiles conditionnels très extrêmes

- 3 Expériences numérique et illustration sur des données réelles
 - Expériences numérique
 - Estimation des niveaux de retour de pluies dans les Cévennes

- 4 Conclusions et perspectives
 - Conclusions
 - Perspectives

Plan : 2 - Introduction

- 1 Introduction
 - En l'absence de covariable
 - En présence de covariable
- 2 Estimation des courbes de niveaux extrêmes
- 3 Expériences numérique et illustration sur des données réelles
- 4 Conclusions et perspectives

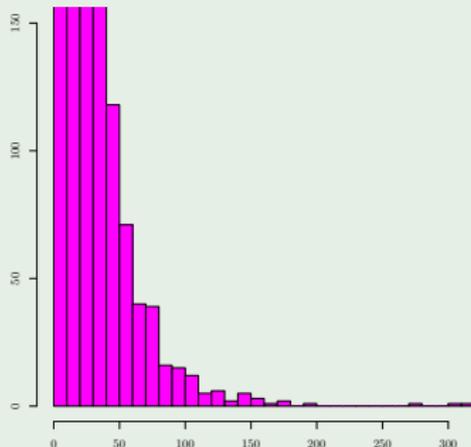


Paris Match (Djakarta, mars 2009) : La rupture d'un barrage dans la banlieue de la capitale indonésienne dans la nuit de jeudi à vendredi a entraîné l'inondation d'un quartier résidentiel très peuplé. Plus de 500 maisons dans les zones résidentielles de Cireundeu et de Tangerang, dans le sud de l'agglomération de Djakarta ont été prises sous les eaux Selon le chef de la cellule de crise du ministère de la Santé, Rustam Pakaya, le bilan est d'au moins 50 morts et plusieurs dizaines de disparus.

Hypothèse

Les précipitations notées Y_1, \dots, Y_n sont des réalisations de variables aléatoires indépendantes et de même loi.

Trouver la hauteur des digues dont la probabilité d'être dépassée par les inondations est proche de zéro ?



Loi empirique des précipitations journalières (mm) à Montpellier entre 1952 — 2000.

Problématique : en l'absence de covariable

Partant d'un échantillon d'observations indépendantes $\{Y_i, i = 1, \dots, n\}$.

- Estimer les **quantités extrêmes** q_{α_n} associées à une v.a $Y \in \mathbb{R}$ définies par

$$\mathbb{P}(Y > q_{\alpha_n}) = \alpha_n,$$

quand $\alpha_n \rightarrow 0$ lorsque $n \rightarrow \infty$.

- Calculer la **probabilité** α_n d'observer une quantité extrême y_n définie par

$$\alpha_n \stackrel{\text{def}}{=} \mathbb{P}(Y > y_n),$$

quand $y_n \rightarrow \infty$ lorsque $n \rightarrow \infty$.

Vocabulaire

- Y est appelée la **variable d'intérêt**,
- q_{α_n} est le **quantile** d'ordre α_n ,
- si $\alpha_n \rightarrow 0$ quand $n \rightarrow \infty$ alors, on parle de **quantile extrême**.

Résultat principal de la théorie des valeurs extrêmes

Théorème [Fisher and Tippett, 1928, Gnedenko, 1943]

Sous certaines conditions de régularité sur la fonction de répartition F , il existe un paramètre réel γ et deux suites $(a_n)_{n \geq 1} > 0$ et $(b_n)_{n \geq 1} \in \mathbb{R}$ tels que pour tout $y \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{\max(Y_1, \dots, Y_n) - b_n}{a_n} \leq y \right] = \Lambda_\gamma(y),$$

avec

$$\Lambda_\gamma(y) = \begin{cases} \exp \left[-(1 + \gamma y)^{-1/\gamma} \right] & \text{si } \gamma \neq 0 \\ \exp \left[-\exp(-y) \right] & \text{si } \gamma = 0 \end{cases} \quad \text{et pour tout } y \text{ tel que } 1 + \gamma y > 0.$$

et Λ_γ la fonction de répartition de la loi des valeurs extrêmes généralisée.

Vocabulaire et notation

- γ est appelé l'indice de queue ou l'indice des valeurs extrêmes.

3 domaines d'attraction

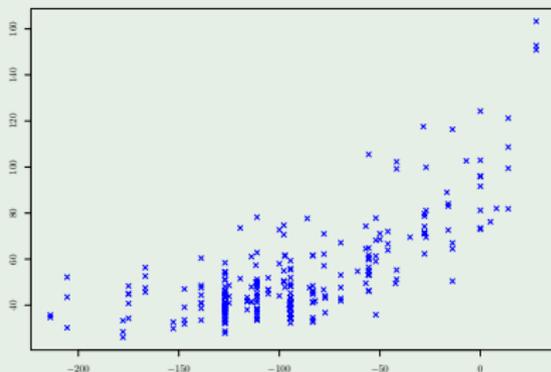
Si F vérifie le Théorème [Fisher and Tippet, 1928, Gnedenko, 1943], alors on dit alors que F appartient au domaine d'attraction de Λ_γ .

Fréchet ($\gamma > 0$)	Weibull ($\gamma < 0$)	Gumbel ($\gamma = 0$)
Cauchy	Uniforme	Normale
Pareto	Beta	Exponentielle
Chi-deux	Weibull	Log-normale
Student		Gamma
Burr		Weibull _m
Fréchet		

Plan : 2 - Introduction

- 1 Introduction
 - En l'absence de covariable
 - En présence de covariable
- 2 Estimation des courbes de niveaux extrêmes
- 3 Expériences numérique et illustration sur des données réelles
- 4 Conclusions et perspectives

Exemple 1 : modélisation de données de fiabilité de réacteurs nucléaires



La ténacité de la cuve (verticalement) et la température (horizontalement).

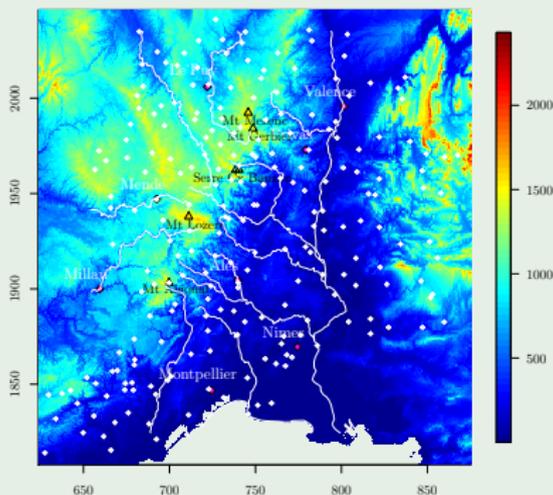
Objectifs :

- évaluer la probabilité (très faible) d'apparition de micro fissures dans ses cuves,
- évaluer des bas-fractiles de la ténacité en fonction de la température.

- quantité d'intérêt → ténacité,
- covariable (**unidimensionnelle**) → température.

Source : données fournies par le Laboratoire de Conduite et Fiabilité des Réacteurs (LCFR) du CEA Cadarache.

Exemple 2 : Carte des niveaux de retour moyen des pluies dans une région



Objectifs :

- évaluer la hauteur de pluies horaires ou journalières pouvant être dépassée une fois toutes les N années ;
- variable d'intérêt → hauteur de pluies,
- covariable (**[bi, tri] dimensionnelle**) → position géographique.

La région des Cévennes-Vivarais et ses alentours.
 Horizontalement : longitude (km), verticalement : latitude (km), échelle des couleurs : altitude (m), villes (roses), montagnes (triangles), fleuves (lignes grises) et stations d'observations (blancs).

Source : Données fournies par le Laboratoire des Transferts en Hydrologie et Environnement (LTHE) de Grenoble.

Problématique : en présence de covariable

Partant d'un échantillon d'observations indépendantes $\{(X_i, Y_i), i = 1, \dots, n\}$ du couple (X, Y) .

- $Y \in \mathbb{R}$ est une variable d'intérêt (**aléatoire**) associée à une covariable X .
- Estimer des **quantiles extrêmes conditionnels** $q(\alpha_n|x)$ définies par

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

quand $\alpha_n \rightarrow 0$ lorsque $n \rightarrow \infty$.

- Évaluer les **petites probabilités conditionnelles** définies par

$$\alpha_n \stackrel{\text{def}}{=} \mathbb{P}(Y > y_n|X = x),$$

quand $y_n \rightarrow \infty$ lorsque $n \rightarrow \infty$.

Problématique : en présence de covariable

Partant d'un échantillon d'observations indépendantes $\{(X_i, Y_i), i = 1, \dots, n\}$ du couple (X, Y) .

- $Y \in \mathbb{R}$ est une variable d'intérêt (**aléatoire**) associée à une covariable X .
- Estimer des **quantiles extrêmes conditionnels** $q(\alpha_n|x)$ définies par

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

quand $\alpha_n \rightarrow 0$ lorsque $n \rightarrow \infty$.

- Évaluer **les petites probabilités conditionnelles** définies par

$$\alpha_n \stackrel{\text{def}}{=} \mathbb{P}(Y > y_n|X = x),$$

quand $y_n \rightarrow \infty$ lorsque $n \rightarrow \infty$.

Difficultés

- $\mathbb{P}(Y > y_n|X = x)$: inconnue, difficile à estimer quand $y_n \rightarrow \infty$.
- $q(\alpha_n|x)$ est une fonction de X .

Plan : 3 - Estimation des courbes de niveaux extrêmes

- 1 Introduction
- 2 Estimation des courbes de niveaux extrêmes
 - **Modèle et méthode d'estimation**
 - Hypothèses et résultats asymptotiques
 - Application à l'estimation de l'indice de queue conditionnel
 - Application à l'estimation des quantiles conditionnels très extrêmes
- 3 Expériences numérique et illustration sur des données réelles
- 4 Conclusions et perspectives

Modèle d'étude

- Soit $\{(X_i, Y_i), i = 1, \dots, n\}$ un échantillon d'observations indépendantes du couple $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$.
- Estimer $\forall x \in \mathbb{R}^p$ et $\forall \alpha_n \rightarrow 0$, les **courbes de niveaux extrêmes** définies comme les graphes de fonctions $x \in \mathbb{R}^p \mapsto q(\alpha_n|x) \in \mathbb{R}$ vérifiant :

$$\bar{F}(q(\alpha_n|x)|x) \stackrel{\text{def}}{=} \mathbb{P}(Y > q(\alpha_n|x)|X = x) = \alpha_n.$$

- On suppose que la loi conditionnelle de $Y|X = x$ appartient au **domaine d'attraction de Fréchet**, i.e.

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

- $\gamma(\cdot)$ une fonction inconnue et positive de la covariable x appelée l'**indice de queue conditionnel**,
- $\ell(\cdot|x)$ une fonction à **variations lentes** (à x fixé) à l'infini, i.e. $\forall \lambda > 0$,

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

Méthode d'estimation et définition des estimateurs

- Estimer le **quantile conditionnel** en inversant la fonction de $\bar{F}(\cdot|x)$, i.e.

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

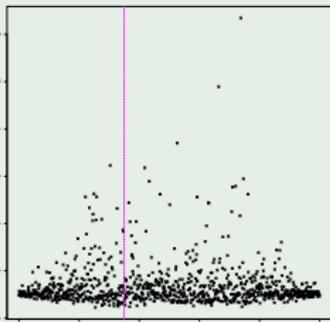
- Nécessite d'estimer la probabilité $\bar{F}(y_n|x)$ lorsque $y_n \rightarrow \infty$ (qd $n \rightarrow \infty$).

Méthode d'estimation et définition des estimateurs

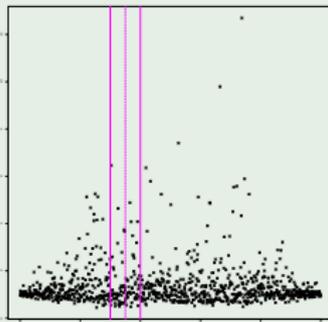
- Estimer le **quantile conditionnel** en inversant la fonction de $\bar{F}(\cdot|x)$, i.e.

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

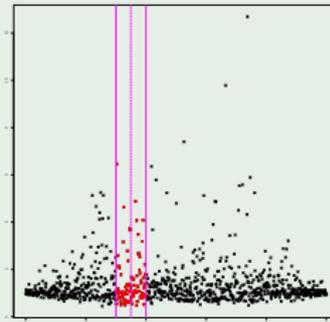
- Nécessite d'estimer la probabilité $\bar{F}(y_n|x)$ lorsque $y_n \rightarrow \infty$ (qd $n \rightarrow \infty$).
- Utiliser les v.a $\{(X_i, Y_i), i = 1, \dots, n\}$ dont les points d'observations X_i sont distants au plus de $h_n > 0$ du point x ,



On se place au point $X = x$



On fixe le paramètre $h_n > 0$



On retient les (X_j, Y_j) dans la $B(x, h_n)$

Méthode d'estimation et définition des estimateurs

- Estimer le **quantile conditionnel** en inversant la fonction de $\bar{F}(\cdot|x)$, i.e.

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

- Nécessite d'estimer la probabilité $\bar{F}(y_n|x)$ lorsque $y_n \rightarrow \infty$ (qd $n \rightarrow \infty$).
- Utiliser les v.a $\{(X_i, Y_i), i = 1, \dots, n\}$ dont les points d'observations X_i sont distants au plus de $h_n > 0$ du point x ,
- Attribuer aux Y_i sélectionnées, des poids qui tiennent compte de la distance de leur X_i au point x ,

Méthode d'estimation et définition des estimateurs

- Estimer le **quantile conditionnel** en inversant la fonction de $\bar{F}(\cdot|x)$, i.e.

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

- Nécessite d'estimer la probabilité $\bar{F}(y_n|x)$ lorsque $y_n \rightarrow \infty$ (qd $n \rightarrow \infty$).
- Utiliser les v.a $\{(X_i, Y_i), i = 1, \dots, n\}$ dont les points d'observations X_i sont distants au plus de $h_n > 0$ du point x ,
- Attribuer aux Y_i sélectionnées, des poids qui tiennent compte de la distance de leur X_i au point x ,
- Estimer α_n par une moyenne pondérée de la variable réponse $\mathbb{1}_{\{Y > y_n\}}$, i.e.

$$\hat{F}_n(y|x) = \frac{\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) \mathbb{1}_{\{Y_i > y_n\}}}{\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)}, \text{ [Collomb, 1980]}$$

- $\mathbb{1}_{\{\cdot\}}$ est la fonction indicatrice.
- K est une fonction positive, bornée, intégrable et à support compact $S \subseteq \mathbb{R}^p$.
- h_n est une suite non aléatoire telle que $h_n \rightarrow 0$ quand $n \rightarrow \infty$.

Plan : 3 - Estimation des courbes de niveaux extrêmes

- 1 Introduction
- 2 Estimation des courbes de niveaux extrêmes
 - Modèle et méthode d'estimation
 - Hypothèses et résultats asymptotiques
 - Application à l'estimation de l'indice de queue conditionnel
 - Application à l'estimation des quantiles conditionnels très extrêmes
- 3 Expériences numérique et illustration sur des données réelles
- 4 Conclusions et perspectives

Hypothèses de régularité

- **Notations** : soit g la densité de X , on désigne par $d(x, x')$ la distance entre deux points $(x, x') \in \mathbb{R}^p \times \mathbb{R}^p$.
- **Conditions de Lipschitz** : il existe des constantes $y_0 > 1$, $c_\gamma > 0$, $c_\ell > 0$ et $c_g > 0$ telles que

$$\begin{aligned} \left| \frac{1}{\gamma(x)} - \frac{1}{\gamma(x')} \right| &\leq c_\gamma d(x, x'), \\ \sup_{y \geq y_0} \left| \frac{\log \ell(y|x)}{\log y} - \frac{\log \ell(y|x')}{\log y} \right| &\leq c_\ell d(x, x'), \\ |g(x) - g(x')| &\leq c_g d(x, x'). \end{aligned}$$

Normalité asymptotique de $\bar{F}(y_n|x)$

Théorème

Si de plus,

- $y_n \rightarrow \infty$ tq $nh_n^p \bar{F}(y_n|x) \rightarrow \infty$ et $nh_n^{p+2} \log^2(y_n) \bar{F}(y_n|x) \rightarrow 0$ qd $n \rightarrow \infty$,
- $\{a_j, j = 1, \dots, J\}$ une suite strictement positive et croissante,

alors, pour tout $x \in \mathbb{R}^p$ tel que $g(x) > 0$,

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

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

Normalité asymptotique de $\bar{F}(y_n|x)$

Théorème

Si de plus,

- $y_n \rightarrow \infty$ tq $nh_n^p \bar{F}(y_n|x) \rightarrow \infty$ et $nh_n^{p+2} \log^2(y_n) \bar{F}(y_n|x) \rightarrow 0$ qd $n \rightarrow \infty$,
- $\{a_j, j = 1, \dots, J\}$ une suite strictement positive et croissante,

alors, pour tout $x \in \mathbb{R}^p$ tel que $g(x) > 0$,

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

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

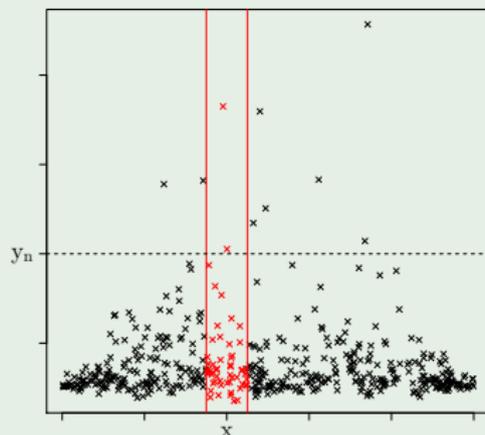
Remarques sur la variance asymptotique

- Inversement proportionnelle à $nh_n^p \bar{F}(y_n|x)$.
- Terme additionnel $1/\bar{F}(y_n|x) \rightarrow \infty$ par rapport au cas $y_n = y$ bornée où [Berinet et al., 2001, Théorème 6.3] trouvent $\frac{\bar{F}(y|x)(1-\bar{F}(y|x))}{nh_n^p} \frac{\|K\|_2^2}{g(x)}$.

Interprétation de la condition $nh_n^p \bar{F}(y_n|x) \rightarrow \infty$

- CNS pour qu'il y ait presque sûrement au moins un point dans la région $B(x, h_n) \times [y_n, +\infty[\in \mathbb{R}^{p+1}$.
- Si y_n est borné, alors on retrouve la condition de normalité asymptotique classique : $nh_n^p \rightarrow \infty$.

Illustration géométrique de $nh_n^p \bar{F}(y_n|x) \rightarrow \infty$, avec $p = 1$



Interprétation de la condition $nh_n^{p+2} \bar{F}_n(y_n|x) \log^2(y_n) \rightarrow 0$

- Condition pour que le carré du biais asymptotique, de l'ordre de

$$(h_n \log y_n)^2,$$

soit négligeable devant la variance asymptotique, de l'ordre de

$$\frac{1}{nh_n^p \bar{F}(y_n|x)}.$$

- Si y_n est borné, alors on retrouve la condition de normalité asymptotique classique : $nh_n^{p+2} \rightarrow 0$.

Hypothèses sur la fonction à variations lentes

- **Représentation de Karamata** : toute fonction à variations lentes peut s'écrire [Bingham et al., 1987, Théorème 1.3.1]

$$\ell(y|x) = c(y|x) \exp \left(\int_1^y \frac{\varepsilon(u|x)}{u} du \right),$$

avec $c(y|x) \rightarrow c(x) > 0$ et $\varepsilon(y|x) \rightarrow 0$ quand $y \rightarrow \infty$

- $\ell(\cdot|x)$ est **normalisée**, i.e. $c(x|y) = c(x)$.

Hypothèses sur la fonction à variations lentes

- **Représentation de Karamata** : toute fonction à variations lentes peut s'écrire [Bingham et al., 1987, Théorème 1.3.1]

$$\ell(y|x) = c(y|x) \exp \left(\int_1^y \frac{\varepsilon(u|x)}{u} du \right),$$

avec $c(y|x) \rightarrow c(x) > 0$ et $\varepsilon(y|x) \rightarrow 0$ quand $y \rightarrow \infty$

- $\ell(\cdot|x)$ est **normalisée**, i.e. $c(x|y) = c(x)$.

Conséquence : $\forall x \in \mathbb{R}^p$, la fonction auxiliaire $\varepsilon(y|x)$ est définie par

$$\varepsilon(y|x) = y \frac{\ell'(y|x)}{\ell(y|x)}.$$

Normalité asymptotique de $\hat{q}_n(\alpha_n|x)$

Théorème

Si de plus,

- $\alpha_n \rightarrow 0$ telle que $nh_n^p \alpha_n \rightarrow \infty$ et $nh_n^{p+2} \alpha_n \log^2(\alpha_n) \rightarrow 0$ quand $n \rightarrow \infty$
- $\{\tau_j, j = 1, \dots, J\}$ une suite strictement positive et décroissante.

Alors, pour tout $x \in \mathbb{R}^p$ tel que $g(x) > 0$,

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

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

Normalité asymptotique de $\hat{q}_n(\alpha_n|x)$

Théorème

Si de plus,

- $\alpha_n \rightarrow 0$ telle que $nh_n^p \alpha_n \rightarrow \infty$ et $nh_n^{p+2} \alpha_n \log^2(\alpha_n) \rightarrow 0$ quand $n \rightarrow \infty$
- $\{\tau_j, j = 1, \dots, J\}$ une suite strictement positive et décroissante.

Alors, pour tout $x \in \mathbb{R}^p$ tel que $g(x) > 0$,

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

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

Remarques sur la variance asymptotique

- Inversement proportionnelle à $nh_n^p \alpha_n$ et proportionnelle à $\gamma^2(x)$.
- Terme additionnel $1/\alpha_n \rightarrow \infty$ par rapport au cas $\alpha_n = \alpha \in]0, 1[$ fixé où [Berlinet et al., 2001, Théorème 6.4] trouvent $\frac{\alpha(1-\alpha)}{nh_n^p} \frac{\|K\|_2^2}{g(x)}$.

Normalité asymptotique de $\hat{q}_n(\alpha_n|x)$

Théorème

Si de plus,

- $\alpha_n \rightarrow 0$ telle que $nh_n^p \alpha_n \rightarrow \infty$ et $nh_n^{p+2} \alpha_n \log^2(\alpha_n) \rightarrow 0$ quand $n \rightarrow \infty$
- $\{\tau_j, j = 1, \dots, J\}$ une suite strictement positive et décroissante.

Alors, pour tout $x \in \mathbb{R}^p$ tel que $g(x) > 0$,

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

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

Remarques sur l'ordre des quantiles extrêmes

- $nh_n^p \alpha_n \rightarrow \infty$ et $nh_n^{p+2} \alpha_n \log^2(\alpha_n) \rightarrow 0$ implique $\alpha_n > \frac{\log^p(n)}{n}$.
- On ne peut pas estimer des **quantiles très extrêmes**.

Plan : 3 - Estimation des courbes de niveaux extrêmes

- 1 Introduction
- 2 Estimation des courbes de niveaux extrêmes
 - Modèle et méthode d'estimation
 - Hypothèses et résultats asymptotiques
 - Application à l'estimation de l'indice de queue conditionnel
 - Application à l'estimation des quantiles conditionnels très extrêmes
- 3 Expériences numérique et illustration sur des données réelles
- 4 Conclusions et perspectives

Application à l'estimation de l'indice de queue conditionnel

Soit $(\beta_n)_{n \geq 1}$ une suite positive telle que $\beta_n \rightarrow 0$.

- **Estimateur à noyau de type Pickands** : adaptation de l'estimateur de [Pickands, 1975] au cas conditionnel.

$$\hat{\gamma}_n^P(x) = \frac{1}{\log 2} \log \left(\frac{\hat{q}_n(\beta_n|x) - \hat{q}_n(2\beta_n|x)}{\hat{q}_n(2\beta_n|x) - \hat{q}_n(4\beta_n|x)} \right).$$

Application à l'estimation de l'indice de queue conditionnel

Soit $(\beta_n)_{n \geq 1}$ une suite positive telle que $\beta_n \rightarrow 0$.

- 1 **Estimateur à noyau de type Pickands** : adaptation de l'estimateur de [Pickands, 1975] au cas conditionnel.

$$\hat{\gamma}_n^P(x) = \frac{1}{\log 2} \log \left(\frac{\hat{q}_n(\beta_n|x) - \hat{q}_n(2\beta_n|x)}{\hat{q}_n(2\beta_n|x) - \hat{q}_n(4\beta_n|x)} \right).$$

- 2 **Estimateur à noyau de type Hill** : adaptation de l'estimateur de [Hill, 1975] au cas conditionnel.

$$\hat{\gamma}_n^H(x) = \frac{\sum_{j=1}^J [\log \hat{q}_n(\tau_j \beta_n|x) - \log \hat{q}_n(\tau_1 \beta_n|x)]}{\sum_{j=1}^J \log(\tau_1/\tau_j)},$$

avec $J > 1$ et $(\tau_j)_{j \geq 1}$ est une suite de poids strictement positive et décroissante.

Hypothèses sur la fonction à variations lentes

Rappel : représentation de Karamata quand $\ell(\cdot|x)$ est normalisée

$$\ell(y|x) = c(x) \exp \left(\int_1^y \frac{\varepsilon(u|x)}{u} du \right).$$

- **Condition du second ordre** : on suppose que $|\varepsilon(\cdot|x)|$ est décroissante à l'infini.
- La fonction $\varepsilon(y|x)$ représente le **terme de biais**. Il contrôle la vitesse de convergence des estimateurs en théorie des valeurs extrêmes.

Normalité asymptotique des estimateurs de l'indice de queue conditionnel

Corollaire

Si de plus $\beta_n \rightarrow 0$ tq $nh_n^p \beta_n \rightarrow \infty$, $nh_n^{p+2} \beta_n \log^2(\beta_n) \rightarrow 0$ et $nh_n^p \beta_n \varepsilon^2(q(2\beta_n|x)|x) \rightarrow 0$ qd $n \rightarrow \infty$, alors pour tout $x \in \mathbb{R}^p$ tq $g(x) > 0$,

$$\sqrt{nh_n^p \beta_n} \left(\hat{\gamma}_n^P(x) - \gamma(x) \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{\gamma^2(x)(2^{2\gamma(x)+1} + 1)}{4(\log 2)^2(2^{\gamma(x)} - 1)^2} \frac{\|K\|_2^2}{g(x)} \right).$$

Normalité asymptotique des estimateurs de l'indice de queue conditionnel

Corollaire

Si de plus $\beta_n \rightarrow 0$ tq $nh_n^p \beta_n \rightarrow \infty$, $nh_n^{p+2} \beta_n \log^2(\beta_n) \rightarrow 0$ et $nh_n^p \beta_n \varepsilon^2(q(2\beta_n|x)|x) \rightarrow 0$ qd $n \rightarrow \infty$, alors pour tout $x \in \mathbb{R}^p$ tq $g(x) > 0$,

$$\sqrt{nh_n^p \beta_n} \left(\hat{\gamma}_n^P(x) - \gamma(x) \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{\gamma^2(x)(2^{2\gamma(x)+1} + 1)}{4(\log 2)^2(2^{\gamma(x)} - 1)^2} \frac{\|K\|_2^2}{g(x)} \right).$$

Corollaire

Si de plus $\beta_n \rightarrow 0$ tq $nh_n^p \beta_n \rightarrow \infty$, $nh_n^{p+2} \beta_n \log^2(\beta_n) \rightarrow 0$ et $nh_n^p \beta_n \varepsilon^2(q(\tau_1 \beta_n|x)|x) \rightarrow 0$ qd $n \rightarrow \infty$, alors pour tout $x \in \mathbb{R}^p$ tq $g(x) > 0$,

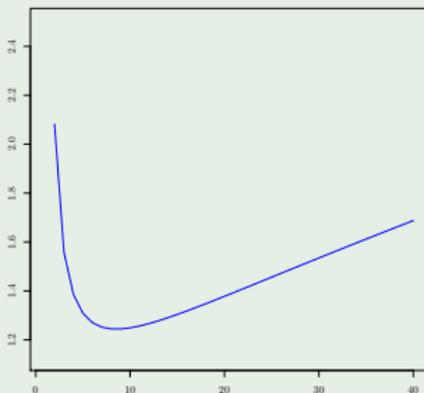
$$\sqrt{nh_n^p \beta_n} \left(\hat{\gamma}_n^H(x) - \gamma(x) \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \gamma^2(x) \frac{\|K\|_2^2}{g(x)} V_J \right),$$

où

$$V_J = \left(\sum_{j=1}^J \frac{2(J-j)+1}{\tau_j} - J^2 \right) / \left(\sum_{j=1}^J \log(\tau_1/\tau_j) \right)^2.$$

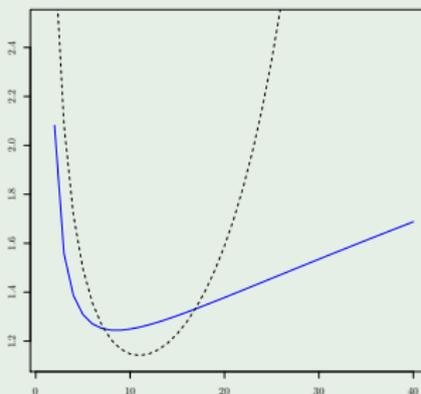
Deux exemples de suites de poids

- La suite de poids harmonique : si $\tau_j^{\text{Ha}} = 1/j$ alors $V_J^{\text{Ha}} = J(J-1)(2J-1)/(6 \log^2(J!))$ est minimum pour $J_{\text{opt}}^{\text{Ha}} = 9$ et $V_9^{\text{Ha}} \simeq 1.245$.



Deux exemples de suites de poids

- La suite de poids harmonique : si $\tau_j^{\text{Ha}} = 1/j$ alors V_J est minimum pour $J_{\text{opt}}^{\text{Ha}} = 9$ et $V_9^{\text{Ha}} \simeq 1.245$.
- La suite géométrique de la balle rebondissante : $\tau_j^{\text{BR}} = b^{2j}$ où $0 < b < 1$. Si $b = 8/9$ alors V_J est minimum pour $J_{\text{opt}}^{\text{BR}} = 11$ et $V_{J_{\text{opt}}}^{\text{BR}} \simeq 1.141$.



Plan : 3 - Estimation des courbes de niveaux extrêmes

- 1 Introduction
- 2 Estimation des courbes de niveaux extrêmes
 - Modèle et méthode d'estimation
 - Hypothèses et résultats asymptotiques
 - Application à l'estimation de l'indice de queue conditionnel
 - Application à l'estimation des quantiles conditionnels très extrêmes
- 3 Expériences numérique et illustration sur des données réelles
- 4 Conclusions et perspectives

Application à l'estimation des quantiles conditionnels très extrêmes

Estimateur à noyau de type Weissman : adaptation de l'estimateur de [Weissman, 1978] au cas conditionnel.

$$\hat{q}_n^W(\alpha_n|x) = \hat{q}_n(\beta_n|x)(\alpha_n/\beta_n)^{-\hat{\gamma}_n(x)},$$

avec

- $\hat{q}_n(\beta_n|x)$ l'estimateur à noyau précédent,
- $\hat{\gamma}_n(x)$ est un estimateur de l'indice de queue conditionnel,
- $(\alpha_n/\beta_n)^{-\hat{\gamma}_n(x)}$ permet d'extrapoler.

Normalité asymptotique de l'estimateur à noyau de Weissman

Théorème

Si de plus

- $\beta_n \rightarrow 0$ tq $nh_n^p \beta_n \rightarrow \infty$ et $nh_n^{p+2} \beta_n \log^2(\beta_n) \rightarrow 0$.
- $\alpha_n / \beta_n \rightarrow 0$.
- $\sqrt{nh_n^p \beta_n} (\hat{\gamma}_n(x) - \gamma(x)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, v^2(x))$ avec $v(x) > 0$.

Alors pour tout $x \in \mathbb{R}^p$ tel que $g(x) > 0$,

$$\frac{\sqrt{nh_n^p \beta_n}}{\log(\beta_n / \alpha_n)} \left(\frac{\hat{q}_n^W(\alpha_n | x)}{q(\alpha_n | x)} - 1 \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, v^2(x)).$$

Au delà du Théorème

La loi limite de $\hat{q}_n^W(\cdot | x)$ peut dépendre du comportement de $\hat{q}_n(\cdot | x)$.

Plan : 4 - Expériences numérique et illustration sur des données réelles

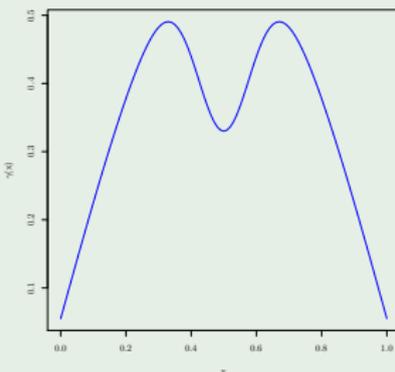
- 1 Introduction
- 2 Estimation des courbes de niveaux extrêmes
- 3 Expériences numérique et illustration sur des données réelles
 - Expériences numérique
 - Estimation des niveaux de retour de pluies dans les Cévennes
- 4 Conclusions et perspectives

Expériences numérique

- On génère $m = 100$ répliques d'un échantillon $\{(X_i, Y_i), i = 1, \dots, n\}$ de taille $n = 1000$ suivant le modèle $X \sim U[0, 1]$ et $Y|X = x$ est distribué selon une loi de **Fréchet**, *i.e.*

$$\bar{F}(y|x) = \exp\left(-y^{-1/\gamma(x)}\right),$$

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



Expériences numérique

- On génère $m = 100$ réplifications d'un échantillon $\{(X_i, Y_i), i = 1, \dots, n\}$ de taille $n = 1000$ suivant le modèle $X \sim U[0, 1]$ et $Y|X = x$ est distribué selon une loi de **Fréchet**, *i.e.*

$$\bar{F}(y|x) = \exp\left(-y^{-1/\gamma(x)}\right),$$

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

- But** : estimer le quantile extrême conditionnel $q(\alpha_n|x) = (-\log(\alpha_n))^{-\gamma(x)}$ d'ordre $\alpha_n \in \{5 \log(n)/n, 1/2n\}$.
- On utilise un **noyau bi-quadratique**

$$K(x) = \frac{15}{16} (1 - x^2)^2 \mathbb{1}_{\{|x| \leq 1\}}.$$

Expériences numérique : l'estimateur $\hat{q}_n(\alpha_n|x)$

Choix du paramètre de lissage

- Validation croisée :

$$\hat{h}_{cv} = \arg \min_{h_n \in \mathcal{H}} \sum_{i=1}^n \sum_{j=1}^n \left\{ \mathbb{1}_{\{Y_i \geq Y_j\}} - \hat{F}_{n,-i}(Y_j|X_i) \right\}^2, \text{ [Yao, 1999]}$$

- $\hat{F}_{n,-i}$ est l'estimateur à noyau calculé sur l'échantillon $\{(X_\ell, Y_\ell), 1 \leq \ell \leq n, \ell \neq i\}$.

Expériences numérique : l'estimateur $\hat{q}_n(\alpha_n|x)$

Choix du paramètre de lissage

- Validation croisée :

$$\hat{h}_{cv} = \arg \min_{h_n \in \mathcal{H}} \sum_{i=1}^n \sum_{j=1}^n \left\{ \mathbb{1}_{\{Y_i \geq Y_j\}} - \hat{F}_{n,-i}(Y_j|X_i) \right\}^2, \text{ [Yao, 1999]}$$

- $\hat{F}_{n,-i}$ est l'estimateur à noyau calculé sur l'échantillon $\{(X_\ell, Y_\ell), 1 \leq \ell \leq n, \ell \neq i\}$.

- Stratégie Oracle :

$$\hat{h}_{oracle} = \arg \min_{h_n \in \mathcal{H}} \mathbb{D}(\hat{q}_n(\alpha_n|\cdot), q(\alpha_n|\cdot)),$$

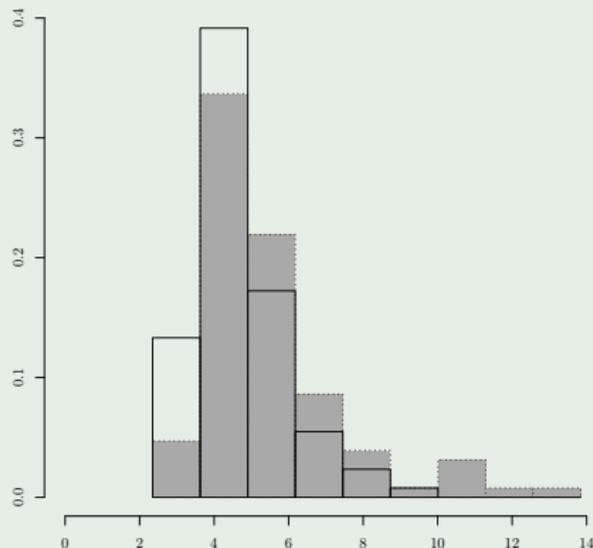
avec

$$\mathbb{D}(u, v) = \left\{ \sum_{\ell=1}^L (u(t_\ell) - v(t_\ell))^2 \right\}^{1/2}.$$

- t_1, \dots, t_L sont des points d'une grille régulière sur $[0, 1]$.

Expériences numérique : l'estimateur $\hat{q}_n(\alpha_n|x)$.

Histogramme des erreurs calculées sur $N = 100$ réplifications.



Oracle : transparent – Validation croisée : gris

Expériences numérique : l'estimateur $\hat{q}_n^W(\alpha_n|x)$

Choix des paramètres h_n et β_n

Dans la suite de ces expériences,

$$\hat{q}_n^W(\alpha_n|x) = \hat{q}_n(\beta_n|x) (\alpha_n/\beta_n)^{-\hat{\gamma}_n^H(x)}.$$

- Approche préconisée :

- (1) Choisir h_n par le critère de validation croisée précédent,
- (2) Estimer β_n en mesurant la similarité entre deux estimateurs de quantiles, *i.e.*

$$\hat{\beta}_{1,2} = \arg \min_{\beta_n \in]0,1]} \mathbb{D} \left(\hat{q}_{n,1}^W(\alpha_n|\cdot), \hat{q}_{n,2}^W(\alpha_n|\cdot) \right),$$

- $\hat{q}_{n,1}^W(\alpha_n|\cdot)$ construit avec $\tau_j^{\text{Ha}} = 1/j$ où $J_{\text{opt}}^{\text{Ha}} = 9$.
- $\hat{q}_{n,2}^W(\alpha_n|\cdot)$ construit avec $\tau_j^{\text{BR}} = (8/9)^{2j}$ où $J_{\text{opt}}^{\text{BR}} = 11$.

Expériences numérique : l'estimateur $\hat{q}_n^W(\alpha_n|x)$

Choix des paramètres h_n et β_n

Dans la suite de ces expériences,

$$\hat{q}_n^W(\alpha_n|x) = \hat{q}_n(\beta_n|x) (\alpha_n/\beta_n)^{-\hat{\gamma}_n^H(x)}.$$

- Approche préconisée :

- (1) Choisir h_n par le critère de validation croisée précédent,
- (2) Estimer β_n en mesurant la similarité entre deux estimateurs de quantiles, i.e.

$$\hat{\beta}_{1,2} = \arg \min_{\beta_n \in]0,1]} \mathbb{D} \left(\hat{q}_{n,1}^W(\alpha_n|\cdot), \hat{q}_{n,2}^W(\alpha_n|\cdot) \right),$$

- $\hat{q}_{n,1}^W(\alpha_n|\cdot)$ construit avec $\tau_j^{\text{Ha}} = 1/j$ où $J_{\text{opt}}^{\text{Ha}} = 9$.
- $\hat{q}_{n,2}^W(\alpha_n|\cdot)$ construit avec $\tau_j^{\text{BR}} = (8/9)^{2j}$ où $J_{\text{opt}}^{\text{BR}} = 11$.

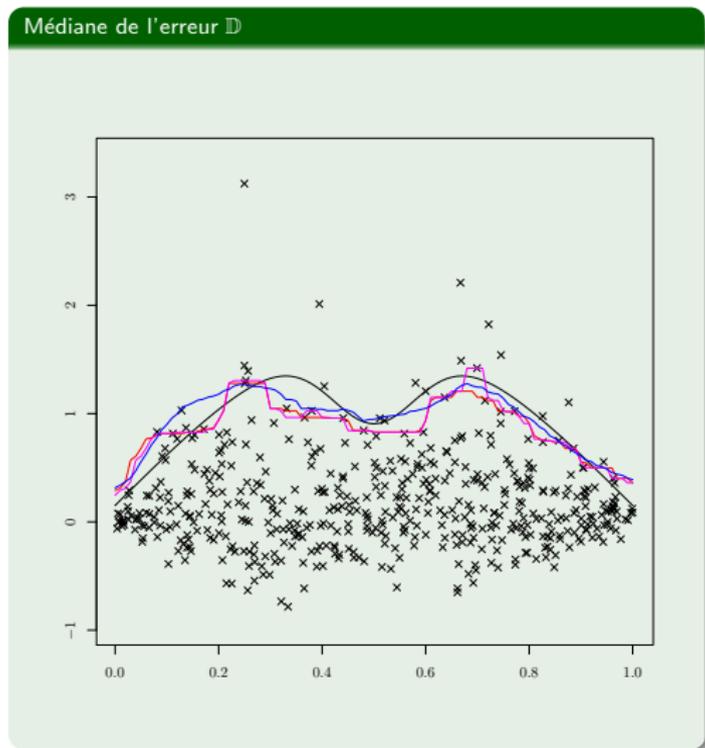
- Stratégie Oracle :

$$\left(\hat{h}_{i,\text{oracle}}, \hat{\beta}_{i,\text{oracle}} \right) = \arg \min_{h_n \in \mathcal{H}, \beta_n \in]0,1]} \mathbb{D} \left(\hat{q}_{n,i}^W(\alpha_n|\cdot), q(\alpha_n|\cdot) \right),$$

avec $i = 1, 2$. Dans la suite $i = 1$.

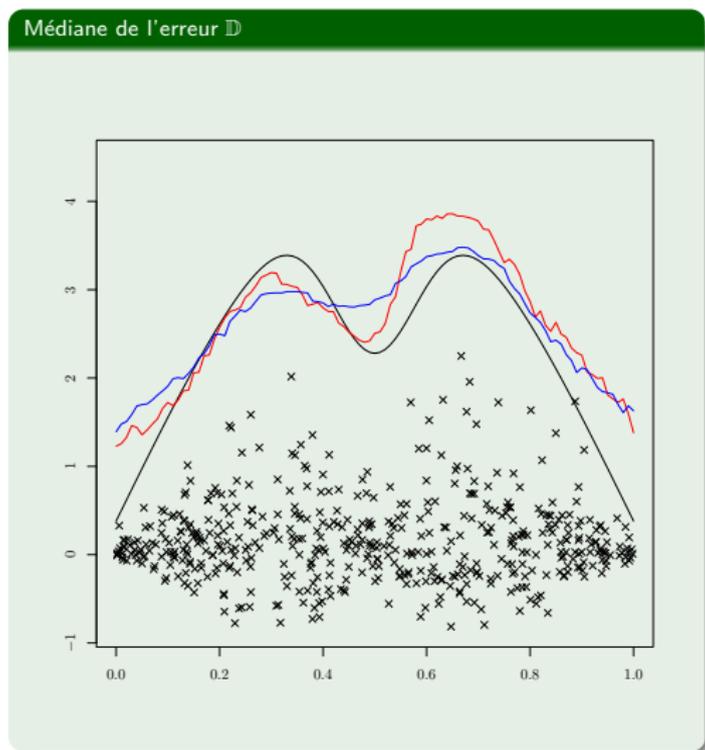
Expériences numérique : estimation du quantile d'ordre $\alpha_n = 5 \log(n)/n$.

— $\hat{q}_n(\alpha_n|\cdot)$: validation croisée — $\hat{q}_n(\alpha_n|\cdot)$: oracle — $\hat{q}_{n,1}^W(\alpha_n|\cdot)$: oracle



Expériences numérique : estimation du quantile d'ordre $\alpha_n = 1/2n$.

— $\hat{q}_{n,1}^W(\alpha_n|\cdot)$: validation croisée + similarité — $\hat{q}_{n,1}^W(\alpha_n|\cdot)$: oracle



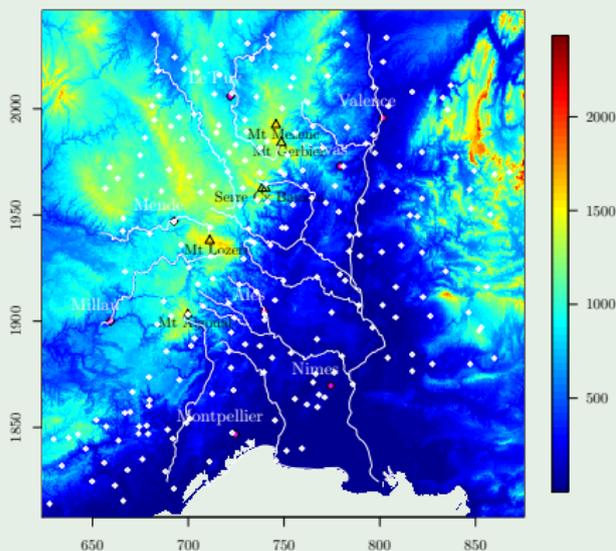
Plan : 4 - Expériences numérique et illustration sur des données réelles

- 1 Introduction
- 2 Estimation des courbes de niveaux extrêmes
- 3 Expériences numérique et illustration sur des données réelles
 - Expériences numérique
 - Estimation des niveaux de retour de pluies dans les Cévennes
- 4 Conclusions et perspectives

Estimation des niveaux retour de pluies (en mm) dans les Cévennes

Horizontalement : longitude (km), verticalement : latitude (km), échelle des couleurs : altitude (m), losanges roses : villes, triangles : montagnes, lignes grises : cours d'eaux, losanges blancs : 225 stations de Météo France.

La région des Cévennes-Vivarais et ses alentours



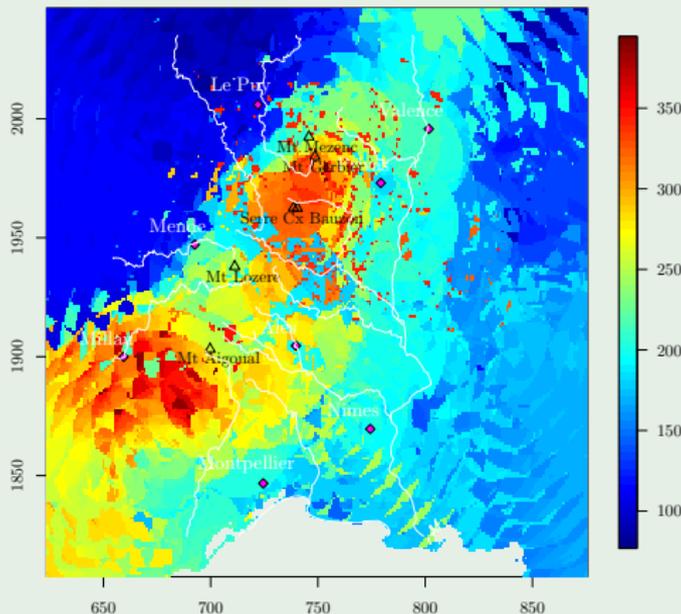
Estimer le **niveau de retour** sur 100 ans, *i.e.* le quantile des précipitations journalier d'ordre $1/(365 \times 100)$.

Pourquoi ?

- Construire des digues d'une hauteur appropriée,
- opérations de nettoyage des fleuves.

Estimation des niveaux retour de pluies (en mm) dans les Cévennes

Niveaux de retour journalier (mm) sur 100 ans en fonction de la longitude et de la latitude



Voir [Molinié et al., 2008] pour une comparaison.

Plan : 5 - Conclusions et perspectives

- 1 Introduction
- 2 Estimation des courbes de niveaux extrêmes
- 3 Expériences numérique et illustration sur des données réelles
- 4 Conclusions et perspectives
 - Conclusions
 - Perspectives

Conclusions

Trois contributions

Lorsque la loi conditionnelle de Y en un point x de la covariable X appartient au domaine d'attraction de Fréchet.

(1/3) $X \in \mathbb{R}^p$ aléatoire, voir la présentation.

Contribution : Kernel estimators of extreme level curves. *TEST*. À paraître. En collaboration avec [A. Daouia](#), [L. Gardes](#) et [S. Girard](#).

(2/3) X déterministe et de dimension non-nécessairement finie.

Contribution : Functional nonparametric estimation of conditional extreme quantiles. *Journal of Multivariate Analysis*, 101, 419–433, 2010. En collaboration avec [Laurent Gardes](#) et [Stéphane Girard](#).

Lorsque la loi de Y appartient au domaine d'attraction de Fréchet.

(3/3) Un nouvel estimateur des quantiles extrêmes.

Contribution : non publiée, voir chapitre 2 de la thèse.

Plan : 5 - Conclusions et perspectives

- 1 Introduction
- 2 Estimation des courbes de niveaux extrêmes
- 3 Expériences numérique et illustration sur des données réelles
- 4 Conclusions et perspectives
 - Conclusions
 - Perspectives

Perspectives

À court terme

- 1 Normalité asymptotique de nos estimateurs dans le cas des données α -mélangeantes.
- 2 Étendre ces résultats aux autres domaines d'attractions.
- 3 Adapter d'autres estimateurs existants de l'indice de queue au cadre conditionnel.
- 4 Proposer de nouveaux estimateurs à noyau (ou double noyau) de l'indice de queue et des quantiles extrêmes conditionnels dans le cadre fonctionnel.

Perspectives

À court terme

- 1 Normalité asymptotique de nos estimateurs dans le cas des données α -mélangeantes.
- 2 Étendre ces résultats aux autres domaines d'attractions.
- 3 Adapter d'autres estimateurs existants de l'indice de queue au cadre conditionnel.
- 4 Proposer de nouveaux estimateurs à noyau (ou double noyau) de l'indice de queue et des quantiles extrêmes conditionnels dans le cadre fonctionnel.

À moyen terme

- 1 Affiner les critères de sélection des paramètres β_n et h_n .

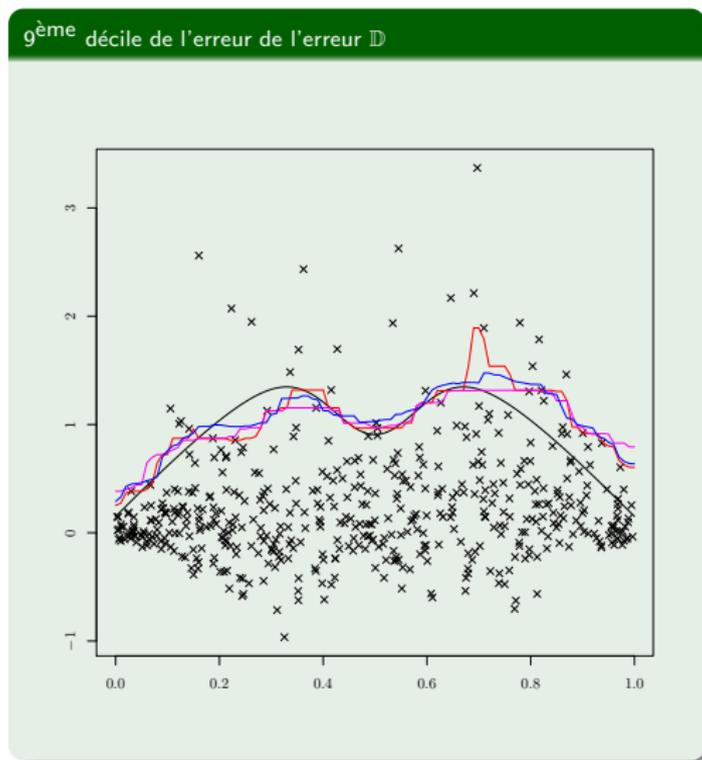
À long terme

- 1 Proposer des nouveaux modèles pour pallier les problèmes dépendance spatio-temporelle assez récurrente en hydrologie.

Annexe

Expériences numérique : estimation du quantile d'ordre $\alpha_n = 5 \log(n)/n$.

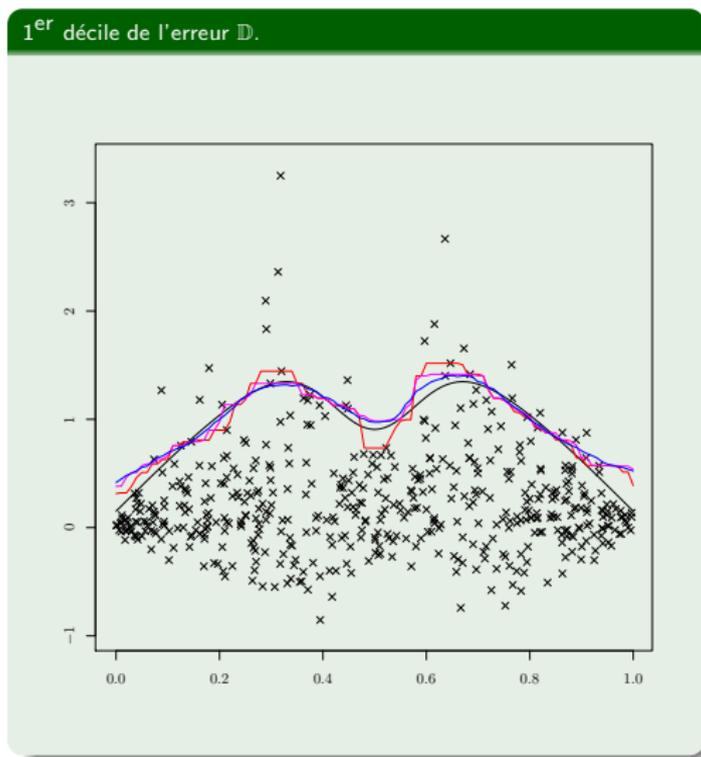
— $\hat{q}_n(\alpha_n|\cdot)$: validation croisée — $\hat{q}_n(\alpha_n|\cdot)$: oracle — $\hat{q}_{n,1}^W(\alpha_n|\cdot)$: oracle



Annexe

Expériences numérique : estimation du quantile d'ordre $\alpha_n = 5 \log(n)/n$.

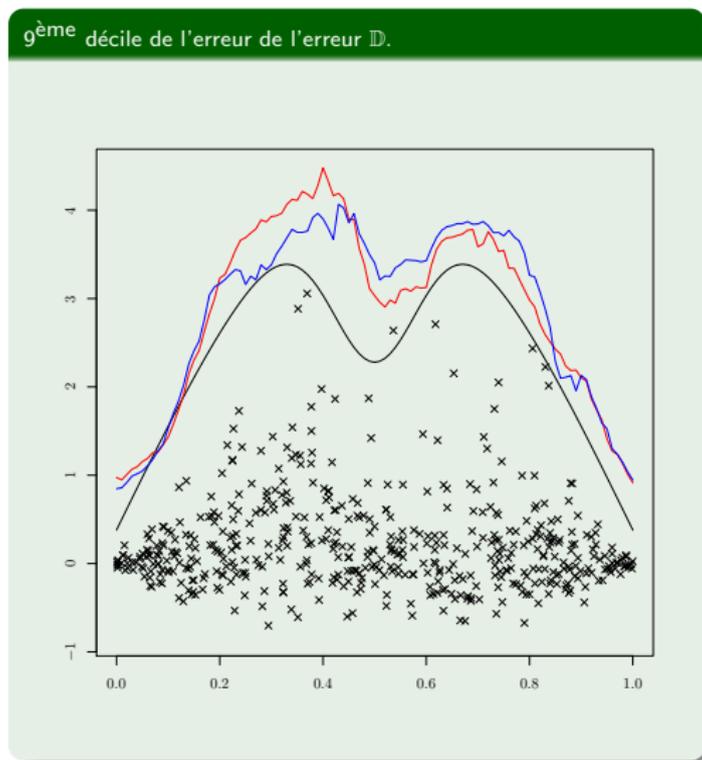
— $\hat{q}_n(\alpha_n|\cdot)$: validation croisée — $\hat{q}_n(\alpha_n|\cdot)$: oracle — $\hat{q}_{n,1}^W(\alpha_n|\cdot)$: oracle



Annexe

Expériences numérique : estimation du quantile d'ordre $\alpha_n = 1/2n$.

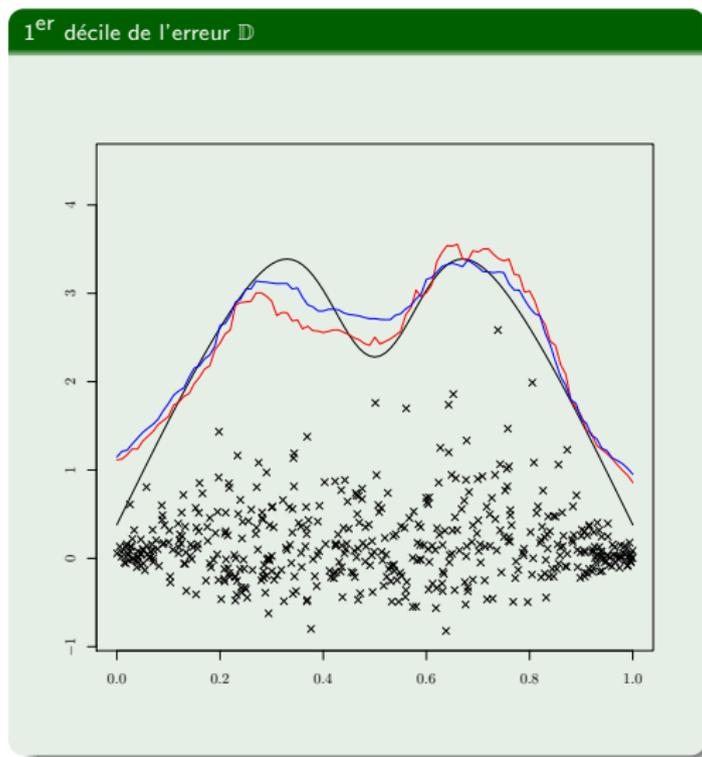
— $\hat{q}_{n,1}^W(\alpha_n|\cdot)$: validation croisée + similarité — $\hat{q}_{n,1}^W(\alpha_n|\cdot)$: oracle



Annexe

Expériences numérique : estimation du quantile d'ordre $\alpha_n = 1/2n$.

— $\hat{q}_{n,1}^W(\alpha_n|\cdot)$: validation croisée + similarité — $\hat{q}_{n,1}^W(\alpha_n|\cdot)$: oracle



Bibliographie I

-  Berlinet, A., Gannoun, A., and Matzner-Lober, E. (2001).
Asymptotic normality of convergent estimates of conditional quantiles.
Statistics, 18 :1400–1415.
-  Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1987).
Regular Variation, volume 27.
Encyclopedia of Mathematics and its Applications, Cambridge University Press.
-  Collomb, G. (1980).
Estimation non paramétrique de probabilités conditionnelles.
Comptes Rendus de l'Académie des Sciences, 291 :427–430.
-  Fisher, R. and Tippett, L. (1928).
Limiting forms of the frequency distribution of the largest or smallest member of a sample.
Proceedings of the Cambridge Philosophical Society, 24 :180–190.
-  Gnedenko, B. (1943).
Sur la distribution limite du terme maximum d'une série aléatoire.
Annals of Mathematics, 44(3) :423–453.

Bibliographie II



Hill, B. (1975).

A simple general approach to inference about the tail of a distribution.
Annals of Statistics, 3(5) :1163–1174.



Molinié, G., Yates, E., Ceresetti, D., Anquetin, S., Boudevillain, B.,
Creutin, J., and Bois, P. (2008).

Rainfall regimes in a mountainous mediterranean region : Statistical
analysis at short time steps.

Article soumis.



Pickands, J. (1975).

Statistical inference using extreme order statistics.
Annals of Statistics, 3(1) :119–131.



Weissman, I. (1978).

Estimation of parameters and large quantiles based on the k -largest
observations.

Journal of the American Statistical Association, 73(364) :812–815.

Bibliographie III



Yao, Q. (1999).

Conditional predictive regions for stochastic processes.
Technical report, University of Kent at Canterbury.