# Estimation de courbes de niveaux extrêmes pour des distributions à queues lourdes

Abdelaati DAOUIA [1], Laurent GARDES [2], Stéphane GIRARD [2] & Alexandre LEKINA [2]

[1] GREMAQ, Université de Toulouse – [2] Équipe-Projet Mistis, INRIA Grenoble-Rhône-Alpes & Laboratoire Jean Kuntzmann

## I. Cadre de l'étude

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

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

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

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

avec $\gamma(.)$ une fonction positive et inconnue de la covariable appelée l'"indice de queue conditionnel" et $\ell(.|x)$ une "fonction à variations lentes", i.e pour tout $\lambda > 0$,

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

## II. Estimateurs

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

$\hat{\bar{F}}_n^{\leftarrow}$ est l'inverse généralisé de l'estimateur à noyau de la fonction de survie conditionnelle

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

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

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

1. Collomb (1976) a étudié le comportement asymptotique de l'estimateur $\hat{g}_n(.)$.
2. Si $\ell(.|x)$ est normalisée et sous des conditions de régularité on établit le comportement asymptotique de l'estimateur $\hat{\psi}_n(y_n, x)$ et on en déduit la loi limite de l'estimateur $\hat{\bar{F}}_n(y_n|x)$ quand la suite $y_n$ converge lentement vers l'infini.
3. De ce qui précède, pour tout $x \in \mathbb{R}^d$ et pour $j = 1, .., J$ tel que $\alpha_{n,j} = \tau_j\alpha_n$, si $\alpha_n \to 0$ et $nh_n^d\alpha_n \to \infty$ quand $n \to \infty$, on déduit que

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

où $\Sigma_{j,j'} = 1/\tau_{j\wedge j'}$ pour $(j, j') \in \{1, \dots, J\}^2$ avec $(\tau_j)$ une suite strictement positive et décroissante.

## IV. Estimation de l'indice de queue conditionnel

- Des résultats précédents, on déduit deux estimateurs de l'indice de queue conditionnel :

$$\begin{cases} \hat{\gamma}_n^P(x) = \dfrac{1}{\log 2}\log\left(\dfrac{\hat{q}_n(\alpha_n|x) - \hat{q}_n(2\alpha_n|x)}{\hat{q}_n(2\alpha_n|x) - \hat{q}_n(4\alpha_n|x)}\right) & \text{estimateur de type Pickands} \\ \hat{\gamma}_n^H(x) = \displaystyle\sum_{j=1}^{J}\log\left(\hat{q}_n(\alpha_{n,j}|x)/\hat{q}_n(\alpha_{n,1}|x)\right) \Bigg/ \sum_{j=1}^{J}\log\left(\tau_1/\tau_j\right) & \text{estimateur de type Hill} \end{cases}$$

avec $J > 1$.
- Condition du second ordre : la valeur absolue de la fonction $\varepsilon(y|x) \stackrel{def}{=} y\dfrac{\ell'(y|x)}{\ell(y|x)}$ est asymptotiquement décroissante et tend vers 0 lorsque $y \to \infty$.
- Cette condition nous permet d'établir la loi limite des estimateurs $\hat{\gamma}_n^P(x)$ et $\hat{\gamma}_n^H(x)$.

## V. Normalité asymptotique de $\hat{\gamma}_n^P(x)$ et $\hat{\gamma}_n^H(x)$

1. Si $\sqrt{nh_n^d\alpha_n}\varepsilon\left(q(2\alpha_n|x)|x\right) \to 0$ quand $n \to \infty$, alors

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

2. Si $\sqrt{nh^d\alpha_n}\varepsilon(q(\alpha_{n,1}|x)|x) \to 0$ quand $n \to \infty$, alors

$$\sqrt{nh^d\alpha_n}\sum_{j=1}^{J}\log\left(\tau_1/\tau_j\right)\left(\hat{\gamma}_n^H(x) - \gamma(x)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{C\|K\|_2^2}{g(x)}\gamma^2(x)\right),$$

où $C = (2J + 1)\sum_{j=1}^{J} 1/\tau_j - 2\sum_{j=1}^{J} j/\tau_j - J^2/\tau_1$.

## VI. Illustration

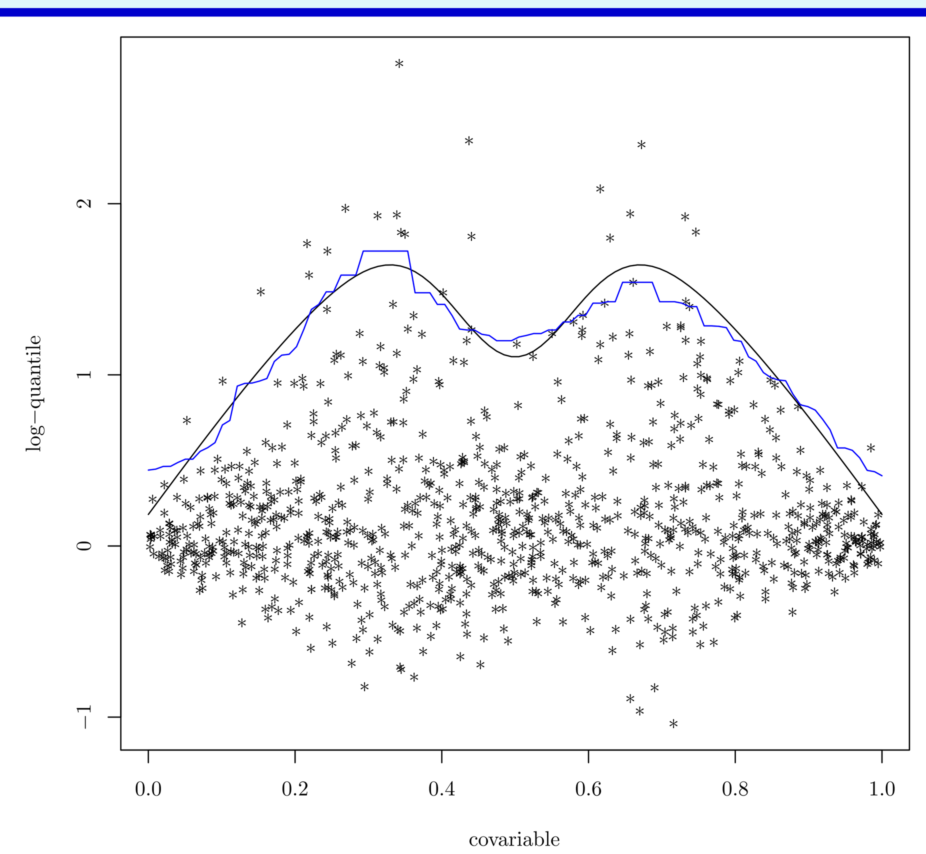
- On génère un échantillon $\{(Y_i, x_i), i = 1, \dots, n\}$ de taille $n = 1000$ dont le quantile conditionnel est donné par

$$q(\alpha|x) = \left\{-\log(1 - \alpha)\right\}^{-\gamma(x)}, \text{ (loi de Fréchet)}$$

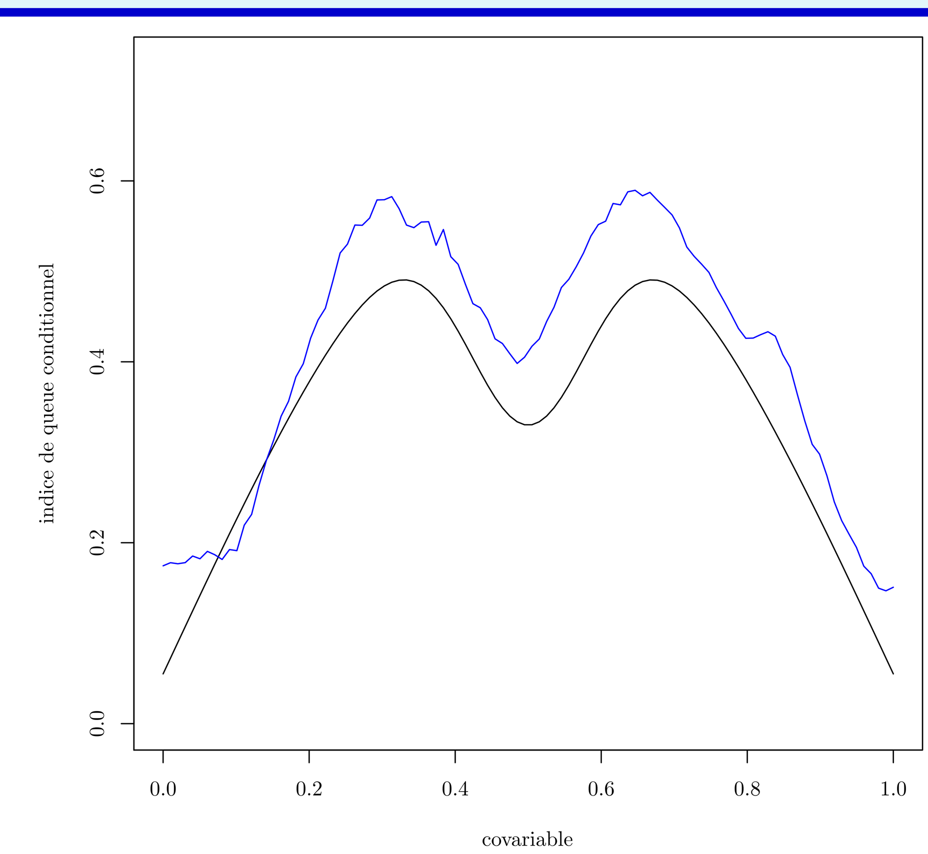
et où la fonction indice de queue conditionnel est définie par

$$x \in [0, 1] \mapsto \gamma(x) = \frac{1}{2}\left(\frac{1}{10} + \sin(\pi x)\right)\left(\frac{11}{10} - \frac{1}{2}\exp\left(-64\left(x - 1/2\right)^2\right)\right).$$

- On choisit le paramètre de lissage par validation croisée (critère de Yao (1999)).
- On fixe $\alpha_n = 5\log(n)/n$ et on représente l'estimateur de quantile $\hat{q}_n(\alpha_n|.)$ correspondant.
- On pose $\tau_j = \tau_0/j$ puis on représente l'estimateur $\hat{\gamma}_n^H(.)$ de variance minimale ($J = 9$).
- Le paramètre de lissage trouvé par le critère vaut $h_{cv} = 0.164$.




Le quantile extrême conditionnel théorique en noir et son estimateur $\hat{q}_n(\alpha_n|.)$ en bleu.




La fonction indice de queue conditionnel en noir et son estimateur $\hat{\gamma}_n^H(.)$ en bleu.

## Références

[1] A. Daouia, L. Gardes, S. Girard et A. Lekina. Kernel estimators of extreme level curves. `http ://hal.inria.fr/inria-00393588/fr/`, 2009.

[2] G. Collomb. *Estimation non paramétrique de la régression par la méthode du noyau.* PhD thesis, Université Paul Sabatier de Toulouse, 1976.

[3] B.M. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3, 1163–1174, 1975.

[4] J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3 :119–131, 1975.

[5] Q. Yao. Conditional predictive regions for stochastic processes. *Technical report*, University of Kent at Canterbury, 1999.

Contact : Alexandre.Lekina@inria.fr
`http ://mistis.inrialpes.fr/people/lekina`