



HAL
open science

Estimation de l'indice des valeurs extrêmes en présence de covariables

Antoine Schorgen

► **To cite this version:**

Antoine Schorgen. Estimation de l'indice des valeurs extrêmes en présence de covariables. 1ères Rencontres R, Jul 2012, Bordeaux, France. hal-00717531

HAL Id: hal-00717531

<https://hal.science/hal-00717531>

Submitted on 13 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimation de l'indice des valeurs extrêmes en présence de covariables

Antoine Schorgen^a

^aDépartement de Mathématiques
Université de Strasbourg et CNRS
IRMA, UMR 7501
7 Rue René Descartes
67084 Strasbourg Cedex
schorgen@unistra.fr

Mots clefs : Valeurs Extrêmes, Estimateur à noyau, Covariables.

Cet exposé sera basé sur l'article [1] écrit en collaboration avec Laurent Gardes et Armelle Guillou et portera sur l'estimation de l'indice des valeurs extrêmes en présence de covariables (aussi appelé indice de queue conditionnel).

Nous nous plaçons dans le domaine de Fréchet où les queues de distribution sont lourdes : la fonction de distribution conditionnelle de Y sachant x se modélise alors de la façon suivante :

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

où x est fixée dans un espace métrique muni d'une distance d , et $L(\cdot, x)$ est une fonction à variations lentes.

Supposons que l'on dispose d'une suite $\{(Y_i, x_i), i = 1, \dots, n\}$ de couples indépendants provenant du modèle (1), la classe d'estimateurs proposés est basée sur un estimateur à noyau des quantiles conditionnels $\hat{q}_n(\alpha, x) := \hat{F}_n^{\leftarrow}(\alpha, x)$ où

$$\hat{F}_n^{\leftarrow}(y, x) = \sum_{i=1}^n H\left(\frac{d(x, x_i)}{h_{1,n}}\right) K\left(\frac{y - Y_i}{h_{2,n}}\right) \Bigg/ \sum_{i=1}^n H\left(\frac{d(x, x_i)}{h_{1,n}}\right),$$

avec $h_{1,n}$ et $h_{2,n}$ deux suites positives non aléatoires, $H(\cdot)$ un noyau asymétrique sur $[0, 1]$ et $K(\cdot)$ un noyau intégré. Cet estimateur a été introduit par Ferraty et Vieu (2006).

Pour chaque x fixé, les observations utilisées sont celles dont les covariables se trouvent dans le voisinage de x , noté $B(x, h_{1,n})$ et désignons par m_x le nombre de ces observations.

L'estimateur proposé de notre indice peut s'écrire sous la forme suivante :

$$\tilde{\gamma}_\theta(x) = \int_0^{k_x/m_x} \Psi_\theta(\alpha, k_x/m_x, x) \log \hat{q}_n(\alpha, x) d\alpha,$$

avec $k_x \in (1, m_x)$ et Ψ_θ une fonction convenablement choisie. Sous des hypothèses classiques en théorie des valeurs extrêmes, nous avons établi la normalité asymptotique de notre estimateur.

Afin d'étudier son comportement à distance finie, nous avons effectué des simulations en choisissant explicitement les noyaux H et K ainsi que la fonction Ψ_θ . Toutes les simulations ont été effectuées sous R, avec plusieurs difficultés à surmonter. Tout d'abord la présence de covariables nous incite à travailler localement : obtenir des tailles de voisinages acceptables implique une

taille d'échantillon totale conséquente ($n = 5000$). De plus, la présence de multiples facteurs à optimiser par validation croisée (taille de voisinage m_x , fenêtre pour les noyaux, nombre de valeurs extrêmes k_x) rend l'algorithme coûteux en temps. Pour réduire les temps de calcul, il a été nécessaire de programmer les éléments clés à l'aide du langage C et de découper les estimations en plusieurs blocs parallèles. La présentation abordera plus particulièrement ces aspects pratiques liés à l'optimisation de la programmation des simulations.

À titre d'exemple, nous illustrons sur la Figure 1, après optimisation des différents paramètres, le comportement de notre estimateur pour une fonction $\gamma(x)$ fixée (courbe en trait plein) et une distribution conditionnelle de type Burr pour les observations.

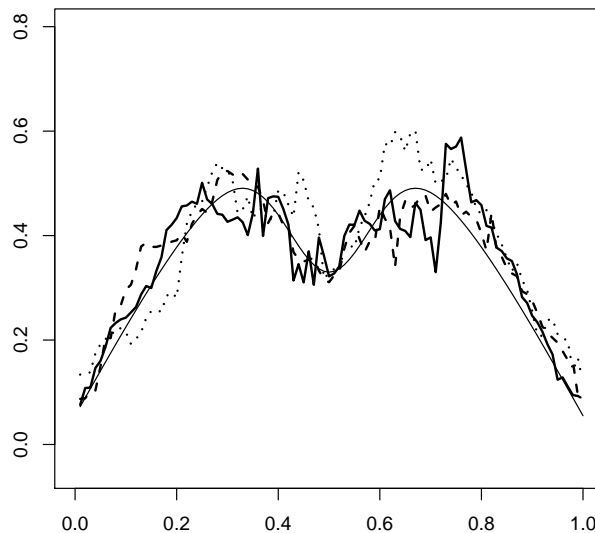


FIGURE 1 – Estimation de $\gamma(\cdot)$: Médiane (plein), Quantiles de niveaux 10% (tirets) et 90% (pointillés) du carré de l'erreur L^2 et vraie fonction (—) pour 100 échantillons de taille 5000.

Références

- [1] Gardes L., Guillou A., Schorgen A. (2012). Estimating the conditional tail index by integrating a kernel conditional quantile estimator, à paraître dans *Journal of Statistical Planning and Inference*.
- [2] Daouia, A., Gardes, L., Girard, S., Lekina, A. (2010). Kernel estimators of extreme level curves, *Test*, **20**(2).
- [3] Ferraty, F., Vieu, P. (2006). *Nonparametric Functional Data Analysis : Theory and Practice*, Springer Series in Statistics, Springer.