CONDITIONAL TAIL INDEX ESTIMATION FOR RANDOM FIELDS

Aladji Bassène [1], Sophie Dabo-Niang [1,2] & Aliou Diop [3]

[1] *Laboratoire LEM, Université Charles-De-Gaulle, Lille3, Maison de la Recherche, domaine universitaire du Pont de Bois, BP 60149, 59653 Villeneuve d'Ascq cedex, France. Email: aladji.bassene@etu.univ-lille3.fr.*
[2] *MODAL TEAM, INRIA-LILLE NORD Europe, France. Email: sophie.dabo@univ-lille3.fr.*
[3] *Laboratoire LERSTAD, UFR SAT, BP 234, université Gaston-Berger, Saint-Louis, Sénégal. Email: aliou.diop@ugb.edu.sn.*

Résumé. Estimation de l'indice de queue conditionnel pour des champs aléatoires. Nous traitons l'estimation de l'indice de queue d'une distribution à queue lourde en présence de covariables pour les processus spatiaux en utilisant l'estimateur de Hill. Soit $\left\{Z_{\mathbf{i}} = (Y_{\mathbf{i}}, x_{\mathbf{i}}) \in \mathbb{R} \times \mathbb{R}^d, \mathbf{i} \in \mathbb{Z}^N\right\}$ un processus spatial strictement stationnaire, nous étudions une estimation de l'indice de queues lourdes de la fonction de distribution conditionnelle spatiale de la variable réponse $Y_{\mathbf{i}}$ étant donnée la variable explicative $x_{\mathbf{i}}$. Notre estimateur est construit sur la base de l'estimateur bien connu de Hill tout en combinant une approche de fenêtre mobile pour capter l'information des covariables. La consistance de l'estimateur de Hill est obtenue lorsque l'échantillon considéré est une suite $\alpha-$*mélangeante.*

Mots-clés. Processus spatiaux; Estimation de l'indice de queue; Consistance.

Abstract. We deal with the estimation of the tail index of a heavy-tailed distribution of a spatial process using Hill's estimator when a covariate is recorded simultaneously with the quantity of interest. Given a stationary multidimensional spatial process $\left\{Z_{\mathbf{i}} = (Y_{\mathbf{i}}, x_{\mathbf{i}}) \in \mathbb{R} \times \mathbb{R}^d, \mathbf{i} \in \mathbb{Z}^N\right\}$, we investigate a heavy-tail index estimate of the spatial conditional distribution function of $Y_{\mathbf{i}}$ given $x_{\mathbf{i}}$. Consistency of Hill's estimator is obtained when the sample considered is $\alpha-$*mixing.*

Keywords. Spatial processes , Tail index estimate , Consistency.

Introduction

In this paper, we are interested in nonparametric conditional tail index estimation for spatial data. Let $\mathbb{Z}^N; N \geq 1$, denotes the integer lattice points in the $N-$dimensional Euclidean space and $\left(Z_{\mathbf{i}} = (Y_{\mathbf{i}}, x_{\mathbf{i}}), \mathbf{i} \in \mathbb{Z}^N\right)$ be an $\mathbb{R} \times \mathbb{R}^d-$valued measurable strictly stationary spatial process, with $x_{\mathbf{i}}$ a fixed design, $Y_{\mathbf{i}}$ has same distribution as Y defined on the probability space (Ω, A, P). Let $\mathbb{R}^d$ be a metric space associated to a metric d. We assume that the conditions of regularly varying tail probabilities Y given $x \in \mathbb{R}^d$ is, for all y,

$$\mathbb{P}\Big(|Y| > \ y \mid x\Big) = y^{-\frac{1}{\gamma(x)}} L\left(y, x\right), \tag{1}$$

where $\gamma(.) > 0$ is an unknown positive function of the covariate x and for x fixed, $L\left(., x\right)$ is a slowly varying function at infinity that is, for $\lambda > 0$, $\lim\limits_{y\to\infty} \frac{L(\lambda y, x)}{L(y,x)} = 1$. We also assume that the tail balancing conditions holds. That is

$$\lim_{y\to\infty} \frac{\mathbb{P}\left(Y > y|x\right)}{\mathbb{P}\left(|Y| > y|x\right)} = \pi_0, \quad \lim_{y\to\infty} \frac{\mathbb{P}\left(Y < -y|x\right)}{\mathbb{P}\left(|Y| > y|x\right)} = 1 - \pi_0, \tag{2}$$

where $0 < \pi_0 \leq 1$. We are interested in nonparametric estimation of the conditional tail-index $\gamma(.)$ for spatial data. Given a sample $Z_{\mathbf{i}} = (Y_{\mathbf{i}}, x_{\mathbf{i}})$ of observations from (1) over a rectangular domain

$$I_{\mathbf{n}} = \{\mathbf{i} = (i_1, ..., i_N) \in \mathbb{Z}^N, 1 \leq i_k \leq n_k, k = 1, ..., N\},$$

where $\mathbf{n} = (n_1, ..., n_N) \in \mathbb{Z}^N$, our aim is to build a point-wise estimator of the function γ. More precisely, for a given $t \in \mathbb{R}^d$, we want to estimate $\gamma(t)$, focusing on the case where the design points $(x_{\mathbf{i}})$ are nonrandom. A point $\mathbf{i} = (i_1, ..., i_N) \in \mathbb{Z}^N$ will be referred to as a site. We will write $\mathbf{n} \to \infty$ if $\mathbf{min}\{n_k\} \to \infty$ and $\left|\frac{n_j}{n_k}\right| < C$ for a constant C such that $0 < C < \infty$ for all j, k such that $1 \leq j, k \leq N$. In the sequel, all the limits are considered when $\mathbf{n} \to \infty$. For $\mathbf{n} = (n_1, ..., n_N) \in \mathbb{Z}^N$, we set $\widehat{\mathbf{n}} = n_1...n_N$. One can also consider either one, two,...or all of n_i, $i = 1, ..., N$ as increasing with $\widehat{\mathbf{n}}$.

Let $B(t, r)$ be the ball centered at point t with radius r:

$$B(t, r) = \{\omega \in \mathbb{R}^d, d(\omega, t) \leq r\}.$$

Let $r_{\mathbf{n},t}$ be a positive sequence tending to zero as $\mathbf{n}$ goes to infinity. The proposed estimator uses the moving window approach as in Gardes and Girard (2008) since is based on the response variables $Y_{\mathbf{i}}$'s for which the associated covariates $x_{\mathbf{i}}$'s belong to the ball $B(t, r_{\mathbf{n},t})$. The proportion of such design points is thus defined by:

$$\phi(r_{\mathbf{n},t}) = \frac{1}{\widehat{\mathbf{n}}} \sum_{\mathbf{i} \in I_{\mathbf{n}}} \mathbb{I}\left\{x_{\mathbf{i}} \in B(t, r_{\mathbf{n},t})\right\},$$

and plays an important role in the following. It describes how the design points are concentrated in the neighborhood of t when $r_{\mathbf{n},t}$ goes to zero. Thus, the nonrandom number of observations $(Y_{\mathbf{i}}, x_{\mathbf{i}})$ in $\mathbb{R} \times B(t, r_{\mathbf{n},t})$ are given by $m_{\mathbf{n},t} = \widehat{\mathbf{n}}\phi(r_{\mathbf{n},t})$. In order to tackle with the subject, we recall some main notations on random fields indexed over $\mathbb{Z}^N$.

In the sequel, the random fields are indexed over $\mathbb{Z}^N$, with $N \geq 2$. We assume that $\mathbb{Z}^N$ is endowed with lexicographic order. For $\mathbf{i}, \mathbf{j} \in \mathbb{Z}^N$ such that $\mathbf{i} \leq \mathbf{j}$ and $\mathbf{i} \neq \mathbf{j}$ the following indexing subsets in $\mathbb{Z}^N$ will be considered:

$$\mathbf{S}\left[\mathbf{i}, \mathbf{j}\right] = \{\mathbf{l} \in \mathbb{Z}^N, \mathbf{i} \leq \mathbf{l} \leq \mathbf{j}\}; \quad \mathbf{S}\left[\mathbf{i}, \infty\right[= \{\mathbf{l} \in \mathbb{Z}^N, \mathbf{i} \leq \mathbf{l}\}.$$

Let $\left\{\widetilde{Z}_{\mathbf{i}}(t), \mathbf{i} \in S\left[\mathbf{1}\ , \mathbf{m}_{\mathbf{n},t}\right]\right\}$ be the selected variables $Y_{\mathbf{i}}$'s for which the associate covariates $x_{\mathbf{i}}$'s belong to the ball $B(t, r_{\mathbf{n},t})$, where $S\left[\mathbf{1}\ , \mathbf{m}_{\mathbf{n},t}\right]$ is the spatial domain where the selected variables are observed. For convenient we treat the spatial sample as triangular arrays, that is $(Y_{\mathbf{i}}, x_{\mathbf{i}})_{\mathbf{i} \in I_{\mathbf{n}}}$ are written $(Y_{i,n}, x_{i,n})$ for $1 \leq i \leq n = \widehat{\mathbf{n}}$ (see Robinsson (2011)), n is the sample size. We can identify each of the indices $i = 1, ..., n$ with a location $\mathbf{i}$ in the space $I_{\mathbf{n}}$. More generally, let g be a continuous bijective function such that: $g : \mathbb{N}^N \longrightarrow \mathbb{N}, \quad (i_1, ..., i_N) \longrightarrow i$. For instance when we have a 2-dimensional regularly-spaced lattice ($N = 2$), where both the number n_1 of rows and the number n_2 of columns increase with $n = n_1 * n_2$, the spatial points $\mathbf{i} = (i_1, i_2)$, for $\mathbf{i} \in I_{\mathbf{n}}$ can be indexed by $i = n_2(i_1 - 1) + i_2$. Let $\mathbf{1}$ (resp. $\mathbf{m}_{\mathbf{n},t}$) be the element of $\mathbb{Z}^N$ whose all components are equal to 1 (resp. $[m_{\mathbf{n},t}]$). The set $\left\{\widetilde{Z}_{\mathbf{i}}(t), \mathbf{i} \in S\left[\mathbf{1}\ , \mathbf{m}_{\mathbf{n},t}\right]\right\}$ can be rewritten as follows:

$$\left\{\widetilde{Z}_i(t), i \in J_{\mathbf{m}_{\mathbf{n},t}} = g\left(S\left[\mathbf{1}\ , \mathbf{m}_{\mathbf{n},t}\right]\right)\right\}.$$

Let us denote by $\widetilde{Z}_{(1),m_{\mathbf{n},t}} \leq \widetilde{Z}_{(2),m_{\mathbf{n},t}} \leq \cdots \leq \widetilde{Z}_{(m_{\mathbf{n},t}),m_{\mathbf{n},t}}$ the order statistics associated to the $m_{\mathbf{n},t}$ variables $\widetilde{Z}_i(t)$ of $J_{\mathbf{m}_{\mathbf{n},t}}$. Let $\mathbf{k}_{\mathbf{n},t}$ be a sequence of elements in $\mathbb{N}^N$, whose all components are equal to $\hat{k}_{\mathbf{n},t} \in \mathbb{N}$ and such that $\mathbf{1} \leq \mathbf{k}_{\mathbf{n},t} \leq \mathbf{m}_{\mathbf{n},t}$. We shall assume that $\mathbf{k}_{\mathbf{n},t}$ is an intermediate spatial sequence, which means that $\hat{k}_{\mathbf{n},t}$ is an intermediate sequence of integers, and:

$$\hat{k} = \hat{k}_{\mathbf{n},t} \to \infty \ ; \ \hat{k}_{\mathbf{n},t} = o\left(m_{\mathbf{n},t}\right) \ \text{ as } \mathbf{n} \to \infty. \tag{3}$$

The conditional tail index can be estimated by the following extended version of Hill estimator:

$$\gamma_{\mathbf{n}}(t) = \frac{1}{\hat{k}_{\mathbf{n},t}} \sum_{i=1}^{\hat{k}_{\mathbf{n},t}} \log\left(\widetilde{Z}_{(i),m_{\mathbf{n},t}} / \widetilde{Z}_{(\hat{k}_{\mathbf{n},t}+1),m_{\mathbf{n},t}}\right) \tag{4}$$

We base inference on the $\hat{k}_{\mathbf{n},t}$ top-order statistics, and as in semi-parametric estimation of parameters of extreme events, see for example Resnick & Stărică (1995).
Consistency results of $\gamma_{\mathbf{n}}$ have been investigated. Some numerical illustrations will also be given.

Bibliography

[1] Gardes, L. and Girard, S. (2008), *A moving window approach for nonparametric estimation of the conditional tail index.* J. Mult. Anal., 99, 2368-2388.

[2] Kallenberg, O. (1983), *Random Measures*, 3rd ed. Akademie, Berlin.

[3] Resnick, S. and Stărică, C. (1995), *Consistency of hills estimator for dependent data*, Journal of Applied probability, 139-167.

[4] Coles, S. G. and Tawn, J. A. (1996), *Modelling extremes of the areal rainfall process*, Journal of the Royal Statistical Society, Series B (Methodological), 329-347.

[5] Robinson, P. M. (2011), *Asymptotic theory for nonparametric regression with spatial data*, Journal of Econometrics, 165, 5-19.

[6] Tran, L. T. (1990), *Kernel density estimation on random fields*, Journal of Multivariate Analysis, 34, 37-53.

[7] Ripley, B. D. (1991), *Statistical inference for spatial processes*, Cambridge university press.