The stochastic approximation method for recursive kernel estimation of the conditional extreme value index

Fatma BEN KHADHER [1] & Yousri SLAOUI [2]

[1] *fatma.b.khadher@gmail.com*
[2] *Yousri.Slaoui@math.univ-poitiers.fr*

Résumé. L'objectif de ce travail est d'appliquer les méthodes d'approximations stochastiques à l'estimation de la fonction d'indice des valeurs extrêmes. Cette méthode nous permet de construire toute une classe d'estimateurs récursifs à noyau de la fonction d'indice des valeurs extrêmes. Ensuite, nous étudions les différentes propriétés asymptotiques de ces estimateurs afin de comparer la performance de notre estimateur récursive avec celle non-récursive de Goegebeur. Nous montrons que, avec un choix optimal de paramètres, l'estimateur récursif proposé par la méthode d'approximation stochastique est très efficace en terme de gain de temps de calcul. Enfin, nous confirmons ces résultats théoriques à l'aide des simulations.

Mots-clés. Algorithme d'approximation stochastique, estimation non-paramétrique, indice des valeurs extrêmes, distribution de type Pareto.

Abstract. The aim is to apply the stochastic approximation method to define a class of recursive kernel estimator of the conditional extreme value index. Then, we study the properties of this recursive estimator and compare them with the non-recursive Goegebeur's estimator. We show that using some optimal parameters, the proposed recursive estimator defined by the stochastic approximation algorithm, will be very competitive with the non-recursive Goegebeur's estimator. Finally, simulations are done to corroborate the obtained theoretical results.

Keywords. Stochastic approximation algorithm, tail index, extreme value, Pareto-type distribution.

1 Introduction

In this paper we are interested by the estimation of tail index, associated to a random variable Y in the area of extreme value theory. There was an intensive work for estimating this parameter (see Beirlant et al. (2004), De Haan and Fereira (2006), Reiss and Thomas (2007) and Gardes et al. (2010). This parameter, denoted by γ, characterizes the distribution tail heaviness of Y. In the framework of parametric estimation for the tail index, we can list the most well-known estimator of the tail index proposed by Hill

(1975) and in the context of non-parametric estimation, the kernel version of Hill's estimator proposed by Goegebeur et al. (2014). During the last years, data streams have become more and more important in the area of research. The present work concerns a nonparametric estimation of the recursive kernel estimator of the conditional extreme value index defined by the stochastic approximation algorithm.

2 Assumptions and main results:

Let $(X_i, Y_i), i = 1, \cdots, n$, be independent realizations of the random vectors $(X, Y) \in \mathbb{R}^d \times \mathbb{R}_0^+$, where X has a distribution with joint density function g, and the conditional survival function of Y given $X = x$ is denoted by $\overline{F}(y|x) = \mathbb{P}(Y > y|X = x)$. The probability density function of Y given $X = x$ is denoted by $f(y|x) = \mathbb{P}(Y = y|X = x)$. More precisely, we assume that the conditional survival function of Y given $X = x$ satisfies
(C1): $\overline{F}(y|x) = y^{-\frac{1}{\gamma(x)}} l(y|x)$,
where $\gamma(.)$ is an unknown positive function of the covariate x called the tail function and for x fixed, $l(.|x)$ is a slowly varying function at infinity, i.e for all $\lambda > 0$,

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

(C2): $l(.|x)$ is normalized,
without losing generality since slowly-varying functions are of interest only asymptotically. In such a case, the Karamata representation (Theorem 1.3.1 given in Bingham et al. (1987) of the slowly-varying function can be written as

$$l(y|x) = c(x) \exp\left(\int_1^y \frac{\varepsilon(z|x)}{z} dz \right),$$

where $c(.)$ is a positive function and $\varepsilon(z|x) \longrightarrow 0$ as $z \longrightarrow \infty$. Thus, $l(.|x)$ is differentiable and the auxiliary function is given by $\varepsilon(z|x) = z\frac{l'(z|x)}{l(z|x)}$.
(C3): There exists a function $\rho(x) < 0$, a strictly positive function $\gamma(x)$ and a rate function $b(.|x)$ with $b(y|x) \longrightarrow 0$ for $y \longrightarrow \infty$, of constant sign for large values of y, such that for all $\upsilon > 0$

$$\lim_{y \to \infty} \frac{\frac{\overline{F}(\upsilon y|x)}{\overline{F}(y|x)} - \upsilon^{-\frac{1}{\gamma(x)}}}{b(y|x)} = \upsilon^{-\frac{1}{\gamma(x)}} \frac{\upsilon^{\frac{\rho(x)}{\gamma(x)}} - 1}{\frac{\rho(x)}{\gamma(x)}}.$$

(C4): There exists $c_g > 0$ such that $g(x) - g(y) = c_g d(x, y)$.
(C5): There exists $c_{\overline{F}} > 0$ such that

$$\frac{\ln \overline{F}(y|x)}{\ln \overline{F}(y|z)} - 1 = c_{\overline{F}} d(x, z).$$

(C6): K is a bounded density function on $\mathbb{R}^d$, with support Ω included in the unit hypersphere of $\mathbb{R}^d$ and satisfying $\int_{\mathbb{R}^d} K(z)dz = 1$.

In this paper, we propose the following recursive estimator for the conditional tail index $\gamma(x)$:

$$\widehat{\gamma}_n^H(x) = \frac{\pi_n \sum_{k=1}^n \pi_k^{-1} \gamma_k K_{h_k}(x - X_k)\left[\ln Y_k - \ln t_n\right] \mathbb{1}_{\{Y_k > t_n\}}}{Q_n \sum_{k=1}^n Q_k^{-1} \beta_k K_{h_k}(x - X_k)\, \mathbb{1}_{\{Y_k > t_n\}}}, \tag{1}$$

where $\pi_n = \prod_{k=1}^n (1 - \gamma_k)$ and $Q_n = \prod_{k=1}^n (1 - \beta_k)$.

The recursive property (1) is particularly useful in large sample size since $\widehat{\gamma}_n^H(x)$ can be easily updated with each additional observation.

The aim of this paper is to study the properties of the recursive estimator defined by (1) and to compare them with the kernel version of Hill's estimator of the conditional extreme value index proposed by Goegebeur et al. (2014), and defined as

$$\tilde{\gamma}_n^H(x) = \frac{\frac{1}{n} \sum_{i=1}^n K_{h_i}(x - X_i)\left[\ln Y_i - \ln t_n\right] \mathbb{1}_{\{Y_i > t_n\}}}{\frac{1}{n} \sum_{i=1}^n K_{h_i}(x - X_i)\, \mathbb{1}_{\{Y_i > t_n\}}}. \tag{2}$$

Throughout this paper, we consider stepsizes and bandwidths belonging to the following class of regularly varying sequences.

Definition 1. *Let $u \in \mathrm{R}$ and $(u_n)_{n \geqslant 1}$ be a nonrandom positive sequence. We say that $u_n \in \mathcal{GS}(u)$ if*

$$\lim_{n \to \infty} n \left[1 - \frac{u_{n-1}}{u_n}\right] = u.$$

This condition was introduced by Galambos and Seneta (1973) and the acronym $\mathcal{GS}$ stand for (**G**alambos and **S**eneta).

(C7):

i) $\gamma_n \in \mathcal{GS}(-\alpha)$ with $\alpha \in (1/2, 1]$.

ii) $h_n \in \mathcal{GS}(-p)$ with $p \in (0, \alpha/d)$.

iii) $\lim_{n \to \infty} n\gamma_n \in (\min(p, (\alpha - pd)/2), \infty]$.

iv) $\beta_n \in \mathcal{GS}(-b)$ with $b \in (1/2, 1]$.

v) $\lim_{n \to \infty} n\beta_n \in (\min(p, (b - pd)/2), \infty]$.

vi) $nh_n^{d+2} \ln^2 t_n \underset{n \to \infty}{\longrightarrow} \infty$.

Moreover, we use the following notations:

$$\begin{aligned}
\varepsilon &= \lim_{n\to\infty} (n\gamma_n)^{-1}, \\
\varepsilon_1 &= \lim_{n\to\infty} (n\beta_n)^{-1}, \\
c'_{\overline{F}} &= c_{\overline{F}} \|z\|_2 \text{ such as } z \in \mathbf{B}^*_{\mathbb{R}^d}(0,1), \\
C &= \left(-\frac{1}{\gamma(x)} + o(1)\right) c_{\overline{F}} \|u\|_2 \text{ for all } u \in \Omega. \\
m_n(x) &= \gamma(x)\overline{F}(t_n|x)C_x.
\end{aligned}$$

The following Theorem give the bias and the variance of $\widehat{\gamma}_n^H$.

Theorem 1. *(Bias and variance of $\widehat{\gamma}_n^H$)*
*Let Assumptions **(C1)-(C7)** hold, and suppose that the stepsize $(\beta_n) = (n^{-1})$.*

1. *If $p \in (0, \alpha/(d+2)]$, then*

$$\begin{aligned}
\mathbb{E}\left(\widehat{\gamma}_n^H(x)\right) - \gamma(x) &= -\left(\frac{C}{1-p\varepsilon} + \frac{C}{1-p}\right)\gamma(x)\, h_n \ln t_n \\
&\quad + o\left(h_n \ln t_n\right). \qquad (3)
\end{aligned}$$

   If $p \in (\alpha/(d+2)), 1/d)$, then

$$\mathbb{E}\left(\widehat{\gamma}_n^H(x)\right) - \gamma(x) = o\left(\sqrt{\frac{\gamma_n}{h_n^d}}\right). \qquad (4)$$

2. *If $p \in (0, \alpha/(d+2))$, then*

$$\mathbb{V}ar\left(\widehat{\gamma}_n^H(x)\right) = o\left(h_n^2 \ln^2 t_n\right). \qquad (5)$$

   If $p \in [\alpha/(d+2), 1/d)$, then

$$\mathbb{V}ar\left(\widehat{\gamma}_n^H(x)\right) = \frac{1}{b^2(x)}\frac{1}{C_x}\frac{6}{2-(\alpha-pd)\varepsilon}\|K^2\|_1 m_n(x) g(x)\,\gamma(x)\frac{\gamma_n}{h_n^d} + o\left(\frac{\gamma_n}{h_n^d}\right). \qquad (6)$$

Remark 1. *The bias and the variance of the estimator $\widehat{\gamma}_n^H$ depend on the choice of the stepsizes (γ_n) and (β_n).*

Let us state the following Theorem, which gives the weak convergence rate of the proposed recursive estimator $\widehat{\gamma}_n^H$ defined in 1 in the special case of $(\beta_n) = (n^{-1})$.

Theorem 2. *Let Assumptions **(C1)-(C7)** hold, and suppose that $(\beta_n) = (n^{-1})$.*

1. *If there exists $r > 0$ such that $\gamma_n^{-1} h_n^{d+2} \ln^2 t_n \underset{n\to\infty}{\longrightarrow} r$ then*

$$\sqrt{\gamma_n^{-1} h_n^d}\left(\widehat{\gamma}_n^H(x) - \gamma(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\sqrt{r}B(x), V(x)\right),$$

where

$$\begin{aligned} B(x) &= -\left(\frac{C}{1-p\varepsilon} + \frac{C}{1-p}\right)\gamma(x), \\ V(x) &= \frac{1}{b^2(x)}\frac{1}{C_x}\frac{6}{2-(\alpha-pd)\varepsilon}\|K^2\|_1 m_n(x) g(x)\gamma(x). \end{aligned}$$

2. *If $\gamma_n^{-1} h_n^{d+2} \ln^2 t_n \underset{n\to\infty}{\longrightarrow} \infty$, then*

$$\frac{1}{h_n \ln t_n}\left(\widehat{\gamma}_n^H(x) - \gamma(x)\right) \xrightarrow{\mathbb{P}} B(x),$$

where $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution, $\mathcal{N}$ the Gaussian-distribution and $\xrightarrow{\mathbb{P}}$ the convergence in probability.

3 Simulation study

We consider the following simulation design, we simulate $N = 500$ samples of size $n = 250$ of independent replicates (X_i, Y_i) where X is uniformly distributed on $[0, 1]$ and the conditional distribution of Y_i given $X_i = x$ is Pareto with parameter $\gamma(x) = 0.5(0.1 + sin(\pi x) \times (1.1 - 0.5\exp(-64(x - 0.5)^2)))$. For each of the N simulated samples, we estimate $\gamma(.)$ at $x = (0.1, 0.2, 0.4 0.5, 0.7, 0.8)$. In order to calculate our estimator, we consider a biquadratic kernel $K(x) = \frac{15}{16}(1 - x^2)^2 \mathbb{1}_{[-1,1]}$, we select the bandwidth h using cross-validation criterion and the threshold t_n was chosen the $(n - k)$ th order statistic $Y_{(n-k)}$. For each configuration, we calculate the average IAE (Integrated Absolute Error), the average ISE (Integrated Squared Error) and L_∞ of the estimators over $N = 500$ trials. We consider the stepsizes (γ_n, β_n) equal respectively to (n^{-1}, n^{-1}), $((2/3)\, n^{-1}, n^{-1})$, $(n^{-1}, ((2/3)\, n^{-1})$ and $(((2/3)\, n^{-1}, ((2/3)\, n^{-1})$. These four choices of parameters of the recursive estimators are referred to as $R1$, $R2$, $R3$ and $R4$ respectively. Results are given in Table 1.

Moreover, Table 1 shows that our proposed recursive estimators can gives better results in some specific situation and very close in general to the reference values, which prove the effectiveness of our proposed recursive estimators.

Bibliographie

Beirlant, J., Goegebeur, Y. and Teugels, J. (2004), *Statistics of extremes-Theory and applications*, Wiley, Chichester.

$\gamma(0.1) = 0.225$					$\gamma(0.2) = 0.3777$				
NR	$R1$	$R2$	$R3$	$R4$	NR	$R1$	$R2$	$R3$	$R4$
0.3189	0.3166	**0.2971**	0.3312	0.3108	0.3355	0.3115	**0.3391**	0.3	0.3253
0.0036	0.0036	0.0042	0.0042	0.0038	**0.0037**	0.0038	0.0043	0.0042	0.004
0.0422	0.0425	0.0457	0.0453	0.0438	**0.0427**	0.0435	0.0465	0.0457	0.0447
0.088	0.0883	0.1025	0.1016	0.0957	0.0995	0.1039	**0.0926**	0.1069	0.1237
$\gamma(0.4) = 0.4395$					$\gamma(0.5) = 0.33$				
NR	$R1$	$R2$	$R3$	$R4$	NR	$R1$	$R2$	$R3$	$R4$
0.4269	0.5024	0.4624	0.523	0.4814	0.382	0.359	**0.346**	0.3736	0.3605
0.0036	0.0037	0.0043	0.0043	0.004	**0.0038**	0.0039	0.0045	0.0043	0.004
0.0425	0.0428	0.0462	0.0458	0.0449	**0.0432**	0.0435	0.0472	0.046	0.0448
0.0952	0.0895	**0.0941**	0.1173	0.1039	**0.0925**	0.1099	0.1283	0.1006	0.1125
$\gamma(0.7) = 0.4824$					$\gamma(0.8) = 0.3777$				
NR	$R1$	$R2$	$R3$	$R4$	NR	$R1$	$R2$	$R3$	$R4$
0.4431	**0.4751**	0.5599	0.4351	0.5128	**0.3708**	0.3571	0.405	0.3071	0.3483
0.0038	0.0039	0.0045	0.0044	0.004	**0.3708**	0.3571	0.405	0.3071	0.3483
0.0433	0.0436	0.0469	0.0462	0.0449	**0.0426**	0.043	0.0464	0.0453	0.0441
0.092	0.1177	0.1152	0.12	0.0994	0.0994	**0.0959**	0.1482	0.0986	0.1484

Table 1: Simulation results for $\gamma(x)$. For each configuration of the simulation parameters (n, γ_n, β_n, x), the first line gives the value of each estimator in the indicated point x , the second, the third and the fourth lines give Average $IAEs$, average $ISEs$ and L_∞ respectively (approximated using $N = 500$ trials) of five kernel density estimators.

Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987), *Regular Variation*, Cambridge University Press, Cambridge.

Brilhante, M. F., Gomes, M. I. and Pestana, D. (2013). A simple generalisation of the Hill estimator, *Computational Statistics & Data Analysis*, 57, pp. 518-535.

De Haan, L. and Fereira, A. (2006), *Extreme Value Theory-An Introduction*, Springer, New York.

Galambos, J. and Seneta, E. (1973). Regularly Varying Sequences, *Proceedings of the American Mathematical Society*, 41, pp. 110-116.

Gardes, L., Girard, S. and Lekina, A. (2010). Functional nonparametric estimation of conditional extreme quantiles, *Journal of Multivariate Analysis*, 101, pp. 419-433.

Goegebeur, Y., Guillou, A. and Schorgen, A. (2014). Nonparametric regression estimation of conditional tails: the random covariate case, *Statistics*, 48, pp. 732-755.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution, *The Annals of Statistics*, 3, pp. 1163-1174.

Reiss, R.D. and Thomas, M. (2007), *Statistical Analysis of Extreme Values*, Birkhauser, Basel, Switzerland.