Statistics and Probability Letters 107 (2015) 378–384

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

On the asymptotic normality of the extreme value index for right-truncated data

Souad Benchaira, Djamel Meraghni, Abdelhakim Necir *

Laboratory of Applied Mathematics, Mohamed Khider University, Biskra, Algeria

ARTICLE INFO

Article history:
Received 29 August 2015
Received in revised form 31 August 2015
Accepted 1 September 2015
Available online 16 September 2015

MSC:
60G70
60F17
62G30

Keywords:
Bivariate extremes
Extreme values
Random truncation
Tail dependence

ABSTRACT

Recently, Gardes and Stupfler (2015) introduced an estimator of the extreme value index under random truncation based on two distinct sample fractions of extremes from truncated and truncation data. In this paper, we make use of the weighted tail-copula processes to complete their work in the case of equal fractions.

1. Introduction

Let $(\mathbf{X}_i, \mathbf{Y}_i)$, $1 \leq i \leq N$, be $N \geq 1$ independent copies from a couple $(\mathbf{X}, \mathbf{Y})$ of independent positive random variables (rv's) defined over some probability space $(\Omega, \mathcal{A}, \mathbf{P})$, with continuous marginal distribution functions (df's) $\mathbf{F}$ and $\mathbf{G}$ respectively. Suppose that $\mathbf{X}$ is right-truncated by $\mathbf{Y}$, in the sense that $\mathbf{X}_i$ is only observed when $\mathbf{X}_i \leq \mathbf{Y}_i$. We assume that both survival functions $\overline{\mathbf{F}} := 1 - \mathbf{F}$ and $\overline{\mathbf{G}} := 1 - \mathbf{G}$ are regularly varying at infinity with respective negative indices $-1/\gamma_1$ and $-1/\gamma_2$. That is, for any $s > 0$

$$\lim_{x\to\infty} \frac{\overline{\mathbf{F}}(sx)}{\overline{\mathbf{F}}(x)} = s^{-1/\gamma_1} \quad \text{and} \quad \lim_{y\to\infty} \frac{\overline{\mathbf{G}}(sy)}{\overline{\mathbf{G}}(y)} = s^{-1/\gamma_2}. \tag{1.1}$$

Being characterized by their heavy tails, these distributions play a prominent role in extreme value theory. They include distributions such as Pareto, Burr, Fréchet, stable and log-gamma, known to be appropriate models for fitting large insurance claims, log-returns, large fluctuations, etc. (see, e.g., Resnick, 2006). The truncation phenomenon may occur in many fields, for instance, in astronomy, economics (see, e.g., Woodroofe, 1985), medicine (see, e.g., Wang, 1989) and reliability (see, e.g., Gardes and Stupfler, 2015 for the analysis of lifetimes of automobile brake pads as an application of randomly truncated heavy-tailed models).

Let us now denote (X_i, Y_i), $i = 1, \ldots, n$, to be the observed data, as copies of a couple of rv's (X, Y) with joint df H, corresponding to the truncated sample $(\mathbf{X}_i, \mathbf{Y}_i)$, $i = 1, \ldots, N$, where $n = n_N$ is a sequence of discrete rv's. By the law of

* Corresponding author.
E-mail addresses: benchaira.s@hotmail.fr (S. Benchaira), djmeraghni@yahoo.com (D. Meraghni), necirabdelhakim@yahoo.fr (A. Necir).

http://dx.doi.org/10.1016/j.spl.2015.08.031
0167-7152/

large numbers, we have $n_N/N \xrightarrow{\mathbf{P}} p := \mathbf{P}(\mathbf{X} \leq \mathbf{Y})$, as $N \to \infty$. For convenience, we use, throughout the paper, the notation $n \to \infty$ to say that $n \xrightarrow{\mathbf{P}} \infty$. For $x, y \geq 0$, we have

$$\begin{aligned} H(x,y) &:= \mathbf{P}(X \leq x, Y \leq y) \\ &= \mathbf{P}(\mathbf{X} \leq \min(x, \mathbf{Y}), \mathbf{Y} \leq y \mid \mathbf{X} \leq \mathbf{Y}) = p^{-1} \int_0^y \mathbf{F}(x,z)\, d\mathbf{G}(z). \end{aligned}$$

Note that, conditionally on n, the observed data are still independent. The marginal distributions of the observed $X's$ and $Y's$, respectively denoted by F and G, are equal to

$$F(x) = p^{-1} \int_0^x \overline{\mathbf{G}}(z)\, d\mathbf{F}(z) \quad \text{and} \quad G(y) = p^{-1} \int_0^y \mathbf{F}(z)\, d\mathbf{G}(z),$$

it follows that the corresponding tails are

$$\overline{F}(x) = -p^{-1} \int_x^\infty \overline{\mathbf{G}}(z)\, d\overline{\mathbf{F}}(z) \quad \text{and} \quad \overline{G}(y) = -p^{-1} \int_y^\infty \mathbf{F}(z)\, d\overline{\mathbf{G}}(z).$$

It is clear that the asymptotic behavior of $\overline{F}$ simultaneously depends on $\overline{\mathbf{G}}$ and $\overline{\mathbf{F}}$ while that of $\overline{G}$ only relies on $\overline{\mathbf{G}}$. Making use of Proposition B.1.10 in de Haan and Ferreira (2006), for the regularly varying functions $\overline{\mathbf{F}}$ and $\overline{\mathbf{G}}$, we may readily show that both $\overline{G}$ and $\overline{F}$ are regularly varying at infinity as well, with respective indices γ_2 and $\gamma := \gamma_1\gamma_2/(\gamma_1+\gamma_2)$. That is, we have, for any $s > 0$,

$$\lim_{x\to\infty} \frac{\overline{F}(sx)}{\overline{F}(x)} = s^{-1/\gamma} \quad \text{and} \quad \lim_{y\to\infty} \frac{\overline{G}(sy)}{\overline{G}(y)} = s^{-1/\gamma_2}. \tag{1.2}$$

Recently Gardes and Stupfler (2015) addressed the estimation of the extreme value index γ_1 under random truncation. They used the definition of γ to derive the following consistent estimator:

$$\widehat{\gamma}_1(k,k') := \frac{\widehat{\gamma}(k)\,\widehat{\gamma}_2(k')}{\widehat{\gamma}_2(k') - \widehat{\gamma}(k)},$$

where

$$\widehat{\gamma}(k) := \frac{1}{k}\sum_{i=1}^{k} \log \frac{X_{n-i+1:n}}{X_{n-k:n}} \quad \text{and} \quad \widehat{\gamma}_2(k') := \frac{1}{k'}\sum_{i=1}^{k'} \log \frac{Y_{n-i+1:n}}{Y_{n-k':n}}, \tag{1.3}$$

are the well-known (Hill, 1975) estimators of γ and γ_2, with $X_{1:n} \leq \cdots \leq X_{n:n}$ and $Y_{1:n} \leq \cdots \leq Y_{n:n}$ being the order statistics pertaining to the samples $(X_1,\ldots,X_n)$ and $(Y_1,\ldots,Y_n)$ respectively. The two sequences $k = k_n$ and $k' = k'_n$ of integer rv's, which satisfy

$$1 < k, k' < n, \quad \min(k,k') \to \infty \quad \text{and} \quad \max(k/n, k'/n) \to 0 \quad \text{as } n \to \infty,$$

respectively represent the numbers of top observations from truncated and truncation data. By considering the two situations $k/k' \to 0$ and $k'/k \to 0$ as $n \to \infty$, the authors established the asymptotic normality of $\widehat{\gamma}_1(k,k')$, but when $k/k' \to 1$, they only showed, in Theorem 3, that $\sqrt{\min(k,k')}\left(\widehat{\gamma}_1(k,k') - \gamma_1\right) = O_{\mathbf{p}}(1)$, as $n \to \infty$. It is obvious that an accurate computation of the estimate $\widehat{\gamma}_1(k,k')$ requires good choices of both k and k'. However from a practical point of view, it is rather unusual in extreme value analysis to handle two distinct sample fractions simultaneously, which is mentioned by Gardes and Stupfler (2015) in their conclusion as well. In the present work, we consider the situation when $k = k'$ (rather than $k/k' \to 1$), to obtain an estimator

$$\widehat{\gamma}_1 := \widehat{\gamma}_1(k) = k^{-1} \frac{\sum_{i=1}^{k} \log \frac{X_{n-i+1:n}}{X_{n-k:n}} \sum_{i=1}^{k} \log \frac{Y_{n-i+1:n}}{Y_{n-k:n}}}{\sum_{i=1}^{k} \log \frac{X_{n-k:n}Y_{n-i+1:n}}{Y_{n-k:n}X_{n-i+1:n}}}, \tag{1.4}$$

of simpler form, expressed in terms of a single sample fraction k of truncated and truncation observations. Thereby, the number of extreme values used to compute the optimal estimate value $\widehat{\gamma}_1$ may be obtained by applying one of the various heuristic methods available in the literature such that, for instance, the algorithm of page 137 in Reiss and Thomas (2007). This estimator is used by Gardes and Stupfler (2015) in their simulation study (to evaluate the performance high quantile estimators) where they took $k = k'$ as it is mentioned in their conclusion. The task of establishing the asymptotic normality of $\widehat{\gamma}_1$ is a bit delicate as one has to take into account the dependence structure of X and Y. The authors of Gardes and Stupfler (2015) handled this issue by putting conditions on the sample fractions k and k'. In our case we require that the joint df H have a stable tail dependence function ℓ (see Huang, 1992 and Drees and Huang, 1998), in the sense that the following limit exists:

$$\lim_{t\downarrow 0} t^{-1}\mathbf{P}\left(\overline{F}(X) \leq tx \ \text{ or } \overline{G}(Y) \leq ty\right) =: \ell(x,y), \tag{1.5}$$