Journal of Statistical Planning and Inference 168 (2016) 20–37

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Nonparametric estimation of the conditional extreme-value index with random covariates and censoring

Pathé Ndao [a], Aliou Diop [a], Jean-François Dupuy [b,*]

[a] LERSTAD, CEA-MITIC, Gaston Berger University, Saint Louis, Senegal
[b] IRMAR-INSA, Rennes, France

ARTICLE INFO

Article history:
Received 5 September 2014
Received in revised form 21 June 2015
Accepted 21 June 2015
Available online 2 July 2015

Keywords:
Conditional extreme-value index
Conditional extreme quantile
Conditional Kaplan–Meier estimator
Kernel estimator
Simulations

ABSTRACT

Estimation of the extreme-value index of a heavy-tailed distribution is addressed when some random covariate information is available and the data are randomly right-censored. A weighted kernel version of Hill's estimator of the extreme-value index is proposed and its asymptotic normality is established. Based on this, a Weissman-type estimator of conditional extreme quantiles is constructed. A simulation study is conducted to assess the finite-sample behavior of the proposed estimators.

1. Introduction

Estimation of extreme quantiles has become a crucial issue in many fields, such as hydrology, insurance, medicine and reliability. Let Y be a random variable with cumulative distribution function F and let $Y_1, \ldots, Y_n$ be independent copies of Y. Extreme quantiles of F are defined as quantities of the form

$$F^{\leftarrow}(1-\alpha) = \inf\{y : F(y) \geq 1-\alpha\},$$

where α is so small that this quantile falls beyond the range of the observed $Y_1, \ldots, Y_n$. This problem is closely related to estimation of the extreme-value index of Y. The extreme-value index drives the behavior of F in its right tail and thus plays a central role in the analysis of extremes. Recent monographs on extreme value theory and in particular on estimation of the extreme-value index and extreme quantiles include Embrechts et al. (1997), Beirlant et al. (2004), Reiss and Thomas (2007) and Novak (2012).

When some covariate information X is available and the distribution of Y depends on X, the problem is to estimate the conditional extreme-value index and conditional extreme quantiles $F^{\leftarrow}(1-\alpha|x) = \inf\{y : F(y|x) \geq 1-\alpha\}$ of the distribution $F(\cdot|x)$ of Y given $X = x$. Motivating examples include estimation of extreme rainfalls given the geographical location (Gardes and Girard, 2010), analysis of extreme temperatures given topological parameters (Ferrez et al., 2011) and the study of extreme earthquakes given the location (Pisarenko and Sornette, 2003). Estimation of the conditional extreme-value index and conditional extreme quantiles with fixed (or non-random) covariates has been investigated rather extensively in the recent extreme value literature. We refer to Beirlant et al. (2004), Gardes and Girard (2008), Gardes et al. (2010), Stupfler (2013) and the references therein for an overview of the available methodology, including the case where the covariate is

* Corresponding author.
E-mail addresses: ndao.pathe@yahoo.fr (P. Ndao), aliou.diop@ugb.edu.sn (A. Diop), Jean-Francois.Dupuy@insa-rennes.fr (J.-F. Dupuy).

http://dx.doi.org/10.1016/j.jspi.2015.06.004
0378-3758/

functional (Gardes and Girard, 2012). To date, less attention has been paid to the random covariate case, despite its practical interest. Gardes and Stupfler (2014) and Goegebeur et al. (2014b) adapt Hill's estimator of the extreme-value index of a heavy-tailed distribution to the presence of a random covariate. The moment estimator introduced by Dekkers et al. (1989) is adapted to the presence of random covariates by Goegebeur et al. (2014a). Daouia et al. (2011) proposed a kernel-based estimator of conditional extreme quantiles with random covariates.

In this paper, we address estimation of the conditional extreme-value index and conditional extreme quantiles with random covariates when moreover, the observations $Y_1, \ldots, Y_n$ are randomly right-censored. Censoring commonly occurs in the analysis of event time data. For example, Y may represent the duration until the occurrence of some event of interest (such as death of a patient, ruin of a company...). If censoring is present, the observations consist of triplets (Z_i, δ_i, X_i), $i = 1, \ldots, n$, where $Z_i = \min(Y_i, C_i)$, $\delta_i = 1_{\{Y_i \leq C_i\}}$, $1_{\{\cdot\}}$ is the indicator function and C_i is a random censoring time which provides a lower bound on Y_i if $\delta_i = 0$. When there is no covariate information, estimation of the extreme-value index from censored data is considered by Delafosse and Guillou (2002), Beirlant et al. (2010), Gomes and Neves (2011), Brahimi et al. (2013) and Worms and Worms (2014). Matthys et al. (2004), Beirlant et al. (2007) and Einmahl et al. (2008) additionally address estimation of extreme quantiles. Ndao et al. (2014) address estimation of the conditional extreme-value index and conditional extreme quantiles with fixed covariates and censoring.

To our knowledge, estimation of the conditional extreme-value index and extreme quantiles with random covariates and censoring has not yet been addressed. This is the topic of the present paper. We first construct an estimator of the conditional extreme-value index and we establish its asymptotic normality. Our proposal combines a kernel version of Hill's estimator of the extreme-value index (such as developed in Goegebeur et al. (2014b) in the uncensored case) with a weighting term whose role is to correct for censoring (such as in Einmahl et al. (2008) and Brahimi et al. (2013), who estimate the unconditional extreme-value index with censoring). Then, we propose a Weissman-type estimator of conditional extreme quantiles under censoring. The finite-sample performance of the proposed estimators are assessed via simulations.

The remainder of this paper is organized as follows. In Section 2, we construct our estimator of the conditional extreme-value index and we establish its asymptotic normality. In Section 3, we propose an estimator of conditional extreme quantiles under censoring. In Section 4, we assess via simulations the finite sample behavior of our estimators. Some perspectives are given in Section 5. All proofs are deferred to an Appendix.

2. Construction of the estimator and asymptotic properties

2.1. The proposed estimator

Let (X_i, Y_i), $i = 1, \ldots, n$, be independent copies of the random pair (X, Y) where Y is a non-negative random variable and $X \in \mathcal{X}$ (with $\mathcal{X}$ some bounded set of $\mathbb{R}^p$) is a p-dimensional covariate with probability density function g. We assume that Y can be right-censored by a non-negative random variable C. Thus we really observe independent triplets (X_i, δ_i, Z_i), $i = 1, \ldots, n$, where $Z_i = \min(Y_i, C_i)$, $\delta_i = 1_{\{Y_i \leq C_i\}}$ and $1_{\{A\}}$ is the indicator function of the event A. The random variable C is defined on the same probability space $(\Omega, \mathcal{C}, \mathbb{P})$ as Y. We assume that $C_1, \ldots, C_n$ are independent of each other and that Y and C are independent given X. Let $F(\cdot|x)$ and $G(\cdot|x)$ denote the conditional cumulative distribution functions of Y and C given $X = x$, respectively. Let also $\bar{F}(\cdot|x) = 1 - F(\cdot|x)$ and $\bar{G}(\cdot|x) = 1 - G(\cdot|x)$ be the conditional survival functions of Y and C given $X = x$.

In this paper, we focus on heavy tails. Precisely, we assume that the conditional survival functions of Y and C given $X = x$ satisfy

(C1) $\bar{F}(u|x) = u^{-1/\gamma_1(x)} L_1(u|x)$ and $\bar{G}(u|x) = u^{-1/\gamma_2(x)} L_2(u|x)$,

where $\gamma_1(\cdot)$ and $\gamma_2(\cdot)$ are unknown positive continuous functions of the covariate x and for x fixed, $L_1(\cdot|x)$ and $L_2(\cdot|x)$ are slowly varying functions at infinity, that is, for all $\lambda > 0$,

$$\lim_{u \to \infty} \frac{L_i(\lambda u|x)}{L_i(u|x)} = 1, \quad i = 1, 2.$$

This amounts to saying that $\bar{F}(\cdot|x)$ and $\bar{G}(\cdot|x)$ are regularly varying functions at infinity with index $-1/\gamma_1(x)$ and $-1/\gamma_2(x)$ respectively. Condition (C1) also amounts to assuming that the conditional distributions of Y and C given $X = x$ are in the Fréchet maximum domain of attraction. In what follows, we further assume that $L_1(\cdot|x)$ and $L_2(\cdot|x)$ belong to the Hall class of slowly-varying functions. The functions $\gamma_1(\cdot)$ and $\gamma_2(\cdot)$ are referred to as conditional extreme-value index functions.

Remark 1. By conditional independence of Y and C, the conditional cumulative distribution function $H(\cdot|x)$ of Z given $X = x$ is also heavy-tailed, with conditional extreme-value index $\gamma(x) = \gamma_1(x)\gamma_2(x)/(\gamma_1(x)+\gamma_2(x))$. To see this, note that for every u and x,

$$\begin{aligned}\bar{H}(u|x) := 1 - H(u|x) &= \bar{F}(u|x)\bar{G}(u|x) \\ &= u^{-1/\gamma_1(x)} L_1(u|x) u^{-1/\gamma_2(x)} L_2(u|x) \\ &= u^{-1/\gamma(x)} L(u|x),\end{aligned}$$