Statistics and Probability Letters 109 (2016) 78–88

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

# Bias reduced tail estimation for censored Pareto type distributions

J. Beirlant[a,d,*], A. Bardoutsos[b], T. de Wet[c], I. Gijbels[a]

[a] Department of Mathematics and Leuven Statistics Research Center, KU Leuven, Belgium
[b] Actuarial Research group, KU Leuven, Belgium
[c] Department of Statistics and Actuarial Science, University of Stellenbosch, South Africa
[d] Department of Mathematical Statistics and Actuarial Science, Free State University, South Africa

ARTICLE INFO

*Article history:*
Received 9 September 2015
Accepted 27 October 2015
Available online 12 November 2015

*Keywords:*
Pareto type model
Random right censoring
Bias reduction
Extended Pareto distribution

ABSTRACT

We consider bias reduced estimators for the tail index and tail probabilities under random right censoring in case of Pareto-type distributions. The solution is based on second-order refined peaks-over-threshold modelling as developed in Beirlant et al. (2009).


## 1. Introduction

In certain long-tailed insurance products, such as car liability insurance, long developments of claims are encountered. At evaluation of the portfolio a large proportion of the claims are not fully dealt with yet, and hence the real cumulative payments are unknown and are censored by the payments up to evaluation. Other applications of extreme value analysis under right censoring can be found in Beirlant et al. (2007) and Einmahl et al. (2008).

To formalize such setting, let $X$ be a variable of interest with distribution function $F$ belonging to the max-domain of attraction of an extreme value distribution with Extreme Value Index (EVI) $\gamma_1 > 0$. Next let $C$ be a censoring distribution with distribution function $G$ which also belongs to the max-domain of attraction of an extreme value distribution with EVI $\gamma_2 > 0$. Define the variable $Z = \min(X, C)$ and an indicator $\delta = 1_{(X \le C)}$ which equals 1 if the observation $Z$ is non-censored. Under the classical random right censoring model one further assumes the independence of $X$ and $C$. We assume here also that the densities of $X$ and $C$ exist.

In this paper we discuss the estimation of the EVI $\gamma_1$ and tail probabilities $1 - F(u) = P(X > u)$ for large $u$, based on independent and identically distributed observations $(Z_i, \delta_i)$ $(i = 1, \ldots, n)$, when both the distributions of $X$ and $C$ are of Pareto-type:

$$1 - F(x) = x^{-1/\gamma_1}\ell_1(x) \quad \text{and} \quad 1 - G(y) = y^{-1/\gamma_2}\ell_2(y), \quad x, y > 1, \tag{1}$$

where both $\ell_1, \ell_2$ are slowly varying at infinity:

$$\ell_j(tx)/\ell_j(t) \to_{t\to\infty} 1, \quad \text{for every } x > 1 \ (j = 1, 2). \tag{2}$$

Denoting the distribution function of $Z$ with $H$, we obtain

$$1 - H(z) = P(Z > z) = (1 - F)(1 - G)(z)$$

* Correspondence to: KU Leuven, Department of Mathematics and LStat, Celestijnenlaan 200B, 3001 Heverlee, Belgium. Tel.: +32 16379023.
*E-mail address:* jan.beirlant@wis.kuleuven.be (J. Beirlant).

http://dx.doi.org/10.1016/j.spl.2015.10.016
0167-7152/

from which it follows that $Z$ also belongs to the Fréchet max-domain of attraction with extreme value index $\gamma = \gamma_1\gamma_2/(\gamma_1+\gamma_2)$ and slowly varying function $\ell_Z = \ell_1\ell_2$:

$$1 - H(z) = z^{-1/\gamma_1 - 1/\gamma_2}\ell_Z(z), \quad z > 1.$$

In Beirlant et al. (2007) and Einmahl et al. (2008) estimators of the type

$$\hat{\gamma}_{1,k}^{(c,.)} = \frac{\hat{\gamma}_{Z,k}^{(.)}}{\hat{p}_k}$$

with $\hat{\gamma}_{Z,k}^{(.)}$ an estimator of the EVI $\gamma$ based on $k$ largest order statistics $Z_{n,n} \geq \cdots \geq Z_{n-j+1,n} \geq \cdots \geq Z_{n-k,n}$ $(1 \leq j \leq k \leq n)$, and

$$\hat{p}_k = \frac{1}{k}\sum_{j=1}^{k} \delta_{[n-j+1,n]}$$

the proportion of non-censored observations in the largest $k$ order statistics of $Z$, with $\delta_{[n-j+1,n]}$ denoting the $\delta$-indicator corresponding to $Z_{n-j+1,n}$. Worms and Worms (2014) considered some other estimators of $\gamma_1$ restricting to the Pareto-type case and considered the use of bias reduced versions of the Hill (1975) estimator

$$H_{Z,k} = \frac{1}{k}\sum_{j=1}^{k} \log Z_{n-j+1,n} - \log Z_{n-k,n}.$$

Restricting to the Pareto-type case, note that (1) means that the conditional distributions $1 - F_t(x) = P(X > x|X > t)$ and $1 - G_t(y) = P(Y > y|Y > t)$ satisfy

$$\lim_{t\to\infty}(1 - F_t)(xt) = x^{-1/\gamma_1} \quad \text{and} \quad \lim_{t\to\infty}(1 - G_t)(xt) = x^{-1/\gamma_2}, \quad x > 1, \tag{3}$$

so that for $t \to \infty$

$$\bar{F}_t(x) \sim \left(\bar{G}_t(x)\right)^{\gamma_2/\gamma_1} \sim \left(\bar{H}_t(x)\right)^{p}, \tag{4}$$

with $p = \gamma_2/(\gamma_1+\gamma_2)$. The present model hence corresponds to an asymptotic version of the Koziol–Green model of random censorship, see Csörgő and Horváth (1981).

Estimators of the type $\hat{\gamma}_{1,k}^{(c,.)}$ can now be interpreted as slope estimators in a log–log probability plot, when plotting non-parametric estimates of $-\gamma_1 \log P(Xz_{n-j+1,n}|X > z_{n-k,n}) = \log(z_{n-j+1,n}/z_{n-k,n})$ against the estimates of $p \log \bar{H}_{z_{n-k,n}}(z_{n-j+1,n})$:

$$\left(\log(z_{n-j+1,n}/z_{n-k,n}), \hat{p}_k \log((k+1)/j)\right), \quad 1 \leq j \leq k. \tag{5}$$

Estimating the slope in (5) using the simple ratio of the average of the vertical coordinates against the average of the horizontal coordinates leads to the Hill (1975) version of $\hat{\gamma}_{1,k}^{(c,.)}$:

$$\hat{\gamma}_{1,k}^{(c,H)} = \frac{H_{Z,k}}{\hat{p}_k} = \frac{\frac{1}{k}\sum_{j=1}^{k}\log(Z_{n-j+1,k}/Z_{n-k,n})}{\hat{p}_k}. \tag{6}$$

Beirlant et al. (2007) also observed that the estimator $\hat{\gamma}_{1,k}^{(c,H)}$ can be derived by maximum likelihood on the excesses $R_{j,k} = Z_{n-j+1,n}/Z_{n-k,n}, j = 1, \ldots, k$:

$$\hat{\gamma}_{1,k}^{(c,H)} = \underset{\gamma_1}{\operatorname{argmax}}\, \Pi_{j=1}^{k} \left(\gamma_1^{-1} R_{j,k}^{-\gamma_1^{-1}-1}\right)^{\delta_{[n-j+1,n]}} \left(R_{j,k}^{-\gamma_1^{-1}}\right)^{1-\delta_{[n-j+1,n]}}, \tag{7}$$

and that it can also be interpreted as the slope of the Pareto QQ-plot adapted for censoring making use of the Kaplan–Meier estimator $1 - \hat{F}^{KM}$ of $1 - F$:

$$\left(-\log(1 - \hat{F}^{KM}(Z_{n-j+1,n})), \log Z_{n-j+1,n}\right), \quad 1 \leq j \leq n-1. \tag{8}$$

We will also revisit the tail estimators proposed in Beirlant et al. (2007) and Einmahl et al. (2008):

$$\hat{P}_k^{(KM,.)}(X > u) = \left(1 - \hat{F}^{KM}(Z_{n-k,n})\right)\left(\frac{u}{Z_{n-k,n}}\right)^{-1/\hat{\gamma}_{1,k}^{(c,.)}}. \tag{9}$$

In this paper we discuss bias reduced estimators for $\gamma_1$ and $P(X > u)$, and illustrate the proposed method with the indexed total payments ($X$) data set from a car liability insurance portfolio with records over 20 years. Here $C$ denotes the indexed