Pak.j.stat.oper.res. Vol.17 No.1 2021 pp 235-245 DOI: http://dx.doi.org/10.18187/pjsor.v17i1.2735

Pakistan Journal of Statistics and Operation Research

Robust estimation of the extreme value index of Pareto-type distributions under random truncation with applications

Zahnit Abida[1], Brahimi Brahim[2*], Yahia Djabrane[3]

*Corresponding author

1. Laboratory of Applied Mathematics, Biskra University, Algeria, abida_stat@yahoo.fr
2. Laboratory of Applied Mathematics, Biskra University, Algeria, b.brahimi@univ-biskra.dz
3. Laboratory of Applied Mathematics, Biskra University, Algeria, yahia.djabrane@univ-biskra.dz

Abstract

In this paper, we introduce a new robust estimator for the extreme value index of Pareto-type distributions under randomly right-truncated data and establish its consistency and asymptotic normality. Our considerations are based on the Lynden-Bell integral and a useful huberized M-functional and M-estimators of the tail index. A simulation study is carried out to evaluate the robustness and the finite sample behavior of the proposed estimator. Moreover, an extreme quantiles estimation was also derived and applied to real data-set of lifetimes of automobile brake pads.

Key Words: Extreme value index; Extreme quantiles; Random right-truncation; Robust estimation; Small sample.

Mathematical Subject Classification: 60E05, 62E15, 62G30, 62G32.

1. Introduction

Let (X_j, Y_j), $1 \leq j \leq N$, denote a sample of bivariate positive and independent random variables (rv's) defined over some probability space $(\Omega, \mathscr{A}, P)$, with continuous marginal cumulative distribution functions (cdf's) F and G respectively. Suppose that X is right-truncated by Y, in the sense that the rv of interest X_j is only observed when $X_j \leq Y_j$. We assume that both survival functions $\overline{F} := 1 - F$ and $\overline{G} := 1 - G$ are regularly varying at infinity, with respective indices $(-1/\gamma_1)$ and $(-1/\gamma_2)$, i.e, $\overline{F} \in RV_{-1/\gamma_1}$ and $\overline{G} \in RV_{-1/\gamma_2}$. That is, for any $t > 0$,

$$\lim_{x\to\infty} \overline{F}(tx)/\overline{F}(x) = t^{-1/\gamma_1} \quad \text{and} \quad \lim_{x\to\infty} \overline{G}(tx)/\overline{G}(x) = t^{-1/\gamma_2} \tag{1}$$

where $\gamma_j > 0$ $(j = 1, 2)$ is the so-called extreme value index (e.v.i) is a well-known parameter to measure the tail heaviness of a distribution. Distributions satisfying (1) play a very crucial role in extreme value analysis. They include many commonly used models such as Pareto, Burr, Fréchet and Lévy-stable distributions, known to be suitable models for adjusting large insurance claims, log-returns, large fluctuations, etc. (see for instance, Resnick, 2006). Recently, Benatmane et al. (2020) have proposed a new so-called composite Rayleigh-Pareto distribution, and they showed that such a model will be a better fit for some heavy tailed insurance claims data (actual data on Algerian fire insurance losses and Danish fire loss data).

In many real applications, in case of presence of random right truncation (RRT), the rv of interest X may not be fully available. This truncation can occur in many areas, for example, it is usual that the insurer's claim data do not correspond to the underlying losses, because they are truncated from above. We refer to Escudero and Ortega (2008) for a recent paper on insurance claims under RRT.

In what follows, the star notation $(*)$ relates to any characteristic of the observed subsequence denoted by (X_i^*, Y_i^*); $1 \leq i \leq n$, $(n \leq N)$ subject to $X_i^* \leq Y_i^*$ from the N-sample. As a consequence of truncation, the size of actually observed sample, n, is a binomial rv with parameters N and $p := P(X \leq Y)$. We shall assume that $p > 0$, otherwise, nothing will be observed. Consequently, the marginal cdf's of X^* and Y^*, respectively denoted by F^* and G^*, becomes

$$F^*(x) := p^{-1} \int_0^x \overline{G}(t) dF(t) \quad \text{and} \quad G^*(y) := p^{-1} \int_0^y F(t) dG(t),$$

the corresponding tails are

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

We can easily show that (see Proposition B.1.10 in de Haan and Ferreira, 2007) $\overline{F^*} \in RV_{-1/\gamma_1^*}$ and $\overline{G^*} \in RV_{-1/\gamma_2^*}$ with respective indices

$$\gamma_1^* = \gamma_1 \gamma_2 / (\gamma_1 + \gamma_2) \quad \text{and} \quad \gamma_2^* = \gamma_2. \tag{2}$$

Since F and G are heavy-tailed. Therefore, the Woodroofe's nonparametric estimator (see, Woodroofe, 1985) of F, is defined by

$$F_n^{(W)}(x) := \prod_{j: X_j^* > x} \exp\left(-1/nC_n(X_j^*)\right), \quad \text{where } C_n(x) := \frac{1}{n} \sum_{j=1}^n 1_{\left(X_j^* \leq x \leq Y_j^*\right)},$$

in which C_n is the empirical estimator of

$$C(z) := P(X \leq z \leq Y | X \leq Y) = p^{-1} \overline{G}(z) F(z).$$

Another most popular estimator for F, is the well known Lynden-Bell nonparametric maximum likelihood estimator, originally proposed in Lynden-Bell (1971), defined by

$$F_n^{(LB)}(x) := \prod_{j: X_j^* > x} \left(1 - 1/nC_n(X_j^*)\right).$$

Recently, Gardes and Stupfler (2015) was briefly exploited the above relations (2) to define an estimator for the parameter of interest γ_1 by considering the classical Hill (see, Hill, 1975) estimators of γ_1^* and γ_2^* as functions of two distinct numbers of upper observations:

$$\hat{\gamma}_1^{(GS)}(k_1, k_2) = \hat{\gamma}_1^*(k_1) \hat{\gamma}_2^*(k_2) / \left(\hat{\gamma}_2^*(k_2) - \hat{\gamma}_1^*(k_1)\right) \tag{3}$$

where

$$\hat{\gamma}_1^*(k_1) := \frac{1}{k_1} \sum_{j=1}^{k_1} \log\left(X_{(n-j+1)}^* / X_{(n-k_1)}^*\right) \quad \text{and} \quad \hat{\gamma}_2^*(k_2) := \frac{1}{k_2} \sum_{j=1}^{k_2} \log\left(Y_{(n-j+1)}^* / Y_{(n-k_2)}^*\right),$$

$X_{(1)}^* \leq ... \leq X_{(n)}^*$ and $Y_{(1)}^* \leq ... \leq Y_{(n)}^*$ denote the usual order statistics of both observed samples, k_1 and k_2 are the numbers of top statistics (upper observations) which are kept for estimating γ_1^* and γ_2^*. The estimator given by (3) suffer from some kind of calibration problem, because of the difficulty in assessing the correlation between $\hat{\gamma}_1^*$ and $\hat{\gamma}_1^*$, the authors of Gardes and Stupfler (2015) they don't consider the situation where the upper statistics are equal. Benchaira et al. (2015) considered the case where $k := k_1 = k_2$ in the expression (3) of $\hat{\gamma}_1^{(GS)}$. They proved the asymptotic normality of this estimator under the tail dependence and the second-order regular variation conditions. Recently, Worms and Worms (2016) proposed an asymptotically normal estimator for γ_1 by considering a Lynden-Bell integrals with a deterministic threshold $t_n > 0$ given by

$$\hat{\gamma}_1^{(W)}(t_n) := \frac{1}{nF_n^{(LB)}(t_n)} \sum_{j=1}^n \frac{F_n^{(LB)}\left(X_j^*\right)}{C_n\left(X_j^*\right)} \log\left(X_j^* / t_n\right) 1_{\left(X_j^* > t_n\right)}. \tag{4}$$

The case of a random threshold, is addressed by Benchaira et al. (2016) who propose a Hill-type estimator under RRT

based on a Woodroofe integration as follows:

$$\hat{\gamma}_1^{(B)}(k) := \frac{1}{nF_n^{(W)}\left(X^*_{(n-k)}\right)} \sum_{i=1}^{k} \frac{F_n^{(W)}\left(X^*_{(n-i+1)}\right)}{C_n\left(X^*_{(n-i+1)}\right)} \log\left(X^*_{(n-i+1)}/X^*_{(n-k)}\right). \tag{5}$$

All of these e.v.i estimators, as well as the classical Hill-type (in complete data case) are non-robust, in the sense that they are very sensitive to extreme observations, data contamination or model deviation and tend to be highly volatile for small samples (this is illustrated in our simulation study). Also, the rate of convergence of these estimators are based on the optimal value of the numbers of top statistics k or the threshold t_n, but this rate are slower than the parametric rate $\sqrt{n}$. Moreover, estimating the optimal value of k is virtually impossible when the sample size n is small and this leads to unstable estimates for small samples and large confidence intervals (see, Resnick, 1997, for a detailed discussion). The alternative approach is inspired by the theory of robust inference (see for instance, Huber (1981) and Hampel et al. (1986)) instead of exact consistency, this theory aim at stability for small samples, possibly at the cost of a small asymptotic bias. However, as observed by Beran and Schell (2012), in some practical cases, such as natural disasters, operational risk assessment or reinsurance data are sparse (with n often somewhere between 20 and 50) and distributions are expected to be heavy tailed with an unknown e.v.i. Robust estimation of e.v.i. focuses primarily on complete data case, see Brazauskas and Serfling (2000), Beran and Schell (2012) and references therein. The incomplete data case has first been considered by Sayah et al. (2014), who dealt with heavy-tailed and right censored data. The aim of the current paper is to provide a robust e.v.i. estimator for heavy tailed data under RLT.

The paper is organized as follows. In Section 2, we introduce our new e.v.i. estimator under RRT, and establish its consistency and asymptotic normality. The proposed estimator is compared with those already existed and its finite sample behavior is checked by simulation in Section 3. As an application, we introduce, in Section 4, an estimator for very high quantiles, which we apply to a real data-set of lifetimes of automobile brake pads. Section 5 contain some concluding notes.

2. Framework and statement of the results

Recall that the condition (1) can be rephrased as $\bar{F}(x) = x^{-1/\gamma_1} L_F(x)$ and $\overline{G}(x) = x^{-1/\gamma_2} L_G(x)$, where L_F and L_G are slowly varying functions at infinity. Assuming that $\lim_{x\to\infty} L_F(x) = c > 0$ leads to the class of so-called Pareto-like (or heavy-tailed) distributions, i.e. distribution satisfying $1 - F(x) \sim cx^{-1/\gamma_1}$ as x tends to infinity. This, the tail of such distribution behaves asymptotically like the tail of Pareto distribution. Thus, we suggests to robustify the Pareto maximum likelihood estimator of γ_1 in order to obtain sensible estimates for the class of Pareto-type distributions despite possible deviations from the single-parameter Pareto model (see, Beran and Schell, 2012, for a detailed discussion). A natural estimate of γ_1 can therefore be based on a Huberized Pareto score function :

$$\begin{aligned}\psi_{v,u}(x,\gamma) &= [\gamma^{-1}\log(x) - 1]_v^u - \int [\gamma^{-1}\log(z) - 1]_v^u dF_{Par,\gamma}(z) \\ &= [\gamma^{-1}\log(x) - 1]_v^u - (v + \exp(-(v+1)) - \exp(-(u+1))),\end{aligned} \tag{6}$$

where $F_{Par,\gamma}(x) := 1 - x^{-1/\gamma}$, for $x \geq 1$ and $[y]_v^u := \min(\max(y,v),u)$. The reason for huberization is that the Pareto distribution is only an approximation of the true cdf. By huberizing, the estimate becomes robust against a large class of deviations from this approximation. Since deviations are mainly relevant in the center of the distribution, the lower truncation parameter v is more important. As an alternative to Hill-type estimation, Beran (1997) proposed to use all data but huberize the Pareto score function at lower quantiles. This method has been investigated in the complete data case in Beran and Schell (2012). Moreover, $\psi_{v,u}(x,\gamma)$ is defined for any choice of $\gamma > 0$ and $-1 \leq v < u \leq \infty$. Thus, as shown by Beran and Schell (2012), robustness needs to be achieved for lower quantiles whereas extreme observations on the right are those we are interested in. In particular, $\psi_{-1,\infty}(x,\gamma) = \gamma^{-1}\log(x) - 1$ for $x \geq 1$. Consequently, a natural choice is $u = \infty$ and robustification on the left is obtained only if $v > -1$.

Under the assumptions above, and denote by F a set of distributions with support in R_+. Then the functional $T(F)$ defined on $\mathscr{F}$ as the solution $t = t_0$ of the equation

$$\beta_F(t) = \int \psi_{v,u}(x,t)\, dF(x) = 0, \quad (F \in \mathscr{F})$$

is called huberized tail index M-functional. Consequently, by using relations (1.9) and (1.10) in Stute and Wang (2008) in the left truncation case, a natural adaptation of this integral $\beta_F(t)$ in the framework of RRT, leads to the corresponding Huberized Lynden-Bell integral estimator of the e.v.i. γ_1 as any solution sequence T_n of the empirical equation

$$\hat{\beta}_{F_n}(T_n) := \sum_{j=1}^{n} \psi_{v,u}\left(X_j^*, T_n\right) F_n^{(LB)}\left(X_j^*\right) / C_n\left(X_j^*\right) = 0. \tag{7}$$

Remark 2.1. It is worth mentioning that for complete data case (no truncation), we have $n = N$, $X^* = X$ and $C_n = F_n = F_n^*$, it follows that $\hat{\beta}_{F_n}(T_n) = \sum_{i=1}^{n} \psi_{v,u}(X_i, T_n)$ and consequently T_n reduce to the Beran and Schell estimator (see e.g. Beran and Schell, 2012).

Next, we investigate the asymptotic properties of the estimator of the tail index γ_1 under the large class of Pareto-type distributions assumptions. To formulate our main result, the following conditions are required:

(A1) Let $\overline{F} \in RV_{-1/\gamma_1}$ and $\overline{G} \in RV_{-1/\gamma_2}$ with $0 < \gamma_1 < \gamma_2$.

(A2) $\int \left(1/\overline{G}(x)\right) \psi_{v,u}^2(x,t)\, dF(x) < \infty$ and $\int \left(1/\overline{G}(x)\right) dF(x) < \infty$.

Remark 2.2. In comparison with the optimal value of the numbers of top statistics k in the Hill-type estimators, the parameter v play a less crucial role, since the rate of convergence does not depend on v. In contrast to Hill-type estimators under truncation (see, equations 3 and 5), all data are used. The role of v is only to determine a threshold below which data have a bounded influence on the estimator. Note also that, the equation (7) defining our estimator has a solution for $n \geq 2$ almost surely.

Theorem 2.1. *Assume that assumptions (A1) and (A2) hold. Moreover, let $F_n := F_n^{(LB)}$ be the Lynden-Bell estimator of the cdf F. Then, provided the existence of γ_1 as a unique solution of $\beta_F(t) = 0$, any solution sequence $\hat{\gamma}_1^{(Z)} := \hat{\gamma}_{1n}^{(Z)}(v,u)$ of*

$$\hat{\beta}_{F_n}(t) = \int \psi_{v,u}(x,t)\, dF_n(x) = 0 \qquad (n \in \mathbb{N})$$

is a consistent estimator of γ_1. Assume further that $\int \frac{\partial}{\partial t} \psi_{v,u}(x,t)\, dF(x) \neq 0$ hold in a neighborhood of γ_1. Then

$$\sqrt{n}\left(\hat{\gamma}_1^{(Z)} - \gamma_1\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_{v,u}^2\right), \quad as\ n \to \infty$$

where $\xrightarrow{d}$ stands for convergence in distribution and

$$\sigma_{v,u}^2 := \sigma^2 \left(\int \frac{\partial}{\partial t} \psi_{v,u}(x,t)\, dF(x) \right)^{-2} \tag{8}$$

in which

$$\sigma^2 = Var\left\{ \frac{\Lambda(X^*)}{C(X^*)} + \int_{X^*}^{Y^*} \frac{\Lambda(z)}{C^2(z)} dF^*(z) \right\},$$

where

$$\Lambda(z) := \int_{z>x} \left[\psi_{v,u}(z,\gamma_1) - \psi_{v,u}(x,\gamma_1)\right] dF(x).$$

Remark 2.3. Condition (A1) is standard in heavy-tailed and RRT context. The condition $\gamma_1 < \gamma_2$ ensures that the tail of the truncated rv of interest X is note too contaminated by the truncation rv Y. In addition, (A1) implies that, the right endpoints of X and Y are infinite and thus they are equal. Assumption (A2) already appeared in Stute and Wang (2008), who showed that, $\sigma^2 < \infty$ under (A2), therefore, $\sigma_{v,u}^2 < \infty$ too. Since $\overline{G} \leq 1$, it implies $\int \psi_{v,u}^2(x,t)\, dF(x) < \infty$, which is the assumption when no truncation occurs (see, Theorem 2 in Beran and Schell, 2012). In our case, (A2) is satisfied when $0 < \gamma_1 < \gamma_2$.

Proof. The proof is essentially based on Theorem 4.3 in He and Yang (1998) and Corollary 1.1. in Stute and Wang (2008). Note that $\psi_{v,u}(x,t)$ is monotone and continuous in t and $\beta_F(t)$ possesses an isolated root at $t=\gamma_1$. Let $\varepsilon>0$, then under (A1) and (A2) by strong low of large numbers under truncation (see, Theorem 4.3 in He and Yang, 1998), we have

$$\hat{\beta}_{F_n}(\gamma_1-\varepsilon)=\int \psi_{v,u}(x,\gamma_1-\varepsilon)\,dF_n(x)\to\beta_F(t_0-\varepsilon)>0 \quad \text{almost surely}$$

and

$$\hat{\beta}_{F_n}(\gamma_1+\varepsilon)=\int \psi_{v,u}(x,\gamma_1+\varepsilon)\,dF_n(x)\to\beta_F(t_0+\varepsilon)<0 \quad \text{almost surely.}$$

Hence, there exists some $n\in\mathbb{N}$ such that

$$P\left(\hat{\beta}_{F_m}(\gamma_1+\varepsilon)<0<\hat{\beta}_{F_m}(\gamma_1-\varepsilon),\ \forall m\geq n\right)\to 1 \quad as\ n\to\infty. \tag{9}$$

According to the monotonicity of $\psi_{v,u}(x,t)$ in t, together with the assumption of the existence of a solution sequence $\hat{\gamma}_1^{(Z)}$ of the empirical equation

$$\hat{\beta}_{F_n}(t)=\int \psi_{v,u}(x,t)\,dF_n(x)=0 \qquad (n\in\mathbb{N})$$

we then get

$$P\left(\gamma_1+\varepsilon<\hat{\gamma}_1^{(Z)}<\gamma_1-\varepsilon,\ \forall n\geq m\right)\to 1 \quad as\ n\to\infty.$$

The existence of such a solution sequence for a continuous function in a neighborhood of γ_1 follows from (9) for n large enough. Thus, $\hat{\gamma}_1^{(Z)}$ is a consistent estimator of γ_1.

Let us now focus on the asymptotic normality of $\hat{\gamma}_1^{(Z)}$. Recall that,

$$\begin{aligned}\int \psi_{v,u}\left(x,\hat{\gamma}_1^{(Z)}\right)dF_n(x)-\int \psi_{v,u}(x,\gamma_1)\,dF(x)&=\int\left(\psi_{v,u}\left(x,\hat{\gamma}_1^{(Z)}\right)-\psi_{v,u}(x,\gamma_1)\right)dF_n(x)\\&\quad+\int \psi_{v,u}(x,\gamma_1)\,d\left(F_n(x)-F(x)\right),\end{aligned} \tag{10}$$

The assumed differentiability of $\psi_{v,u}(x,t)$ in t allows a Taylor expansion around γ_1 which yields

$$\sqrt{n}\left(\hat{\gamma}_1^{(Z)}-\gamma_1\right)\int\frac{\partial}{\partial t}\psi_{v,u}(x,t)\,dF_n(x)=\sqrt{n}\int\left(-\psi_{v,u}(x,\gamma_1)\right)d\left(F_n(x)-F(x)\right).$$

Then,

$$\sqrt{n}\left(\hat{\gamma}_1^{(Z)}-\gamma_1\right)=\sqrt{n}\left(\int\frac{\partial}{\partial t}\psi_{v,u}(x,t)\,dF_n(x)\right)^{-1}\int\left(-\psi_{v,u}(x,\gamma_1)\right)d\left(F_n(x)-F(x)\right).$$

It was shown in Theorem 4.3 of He and Yang (1998) that for any non-negative measurable real function $\varphi:=\frac{\partial}{\partial t}\psi$, under the condition that $\int\varphi_{v,u}(x,t)\,dF(x)\neq 0$ hold in a neighborhood of γ_1, we get

$$\int\varphi_{v,u}(x,t)\,dF_n(x)=\int\varphi_{v,u}(x,t)\,dF(x)+op(1). \tag{11}$$

Under assumptions (A1) and (A2), we can apply the central limit theorem under right truncation (see, Corollary 1.1 in Stute and Wang, 2008) for the Lynden-Bell integral, obtaining

$$\sqrt{n}\int\left(-\psi_{v,u}(x,\gamma_1)\right)d\left(F_n(x)-F(x)\right)\xrightarrow{d}\mathcal{N}\left(0,\sigma^2\right),\quad \text{as } n\to\infty \tag{12}$$

where σ^2 is given by (8). Consequently, the limit variance follows from (11) and (12). This concludes the proof of Theorem 2.1. □

3. Simulation study

This following section examines the performance of our estimator $\hat{\gamma}_1^{(Z)}$ given by solving the empirical equation (7), in which, the huberizing constants are $v = 0$ and $u = \infty$, and compare it with estimators proposed by Gardes and Stupfler (2015), Worms and Worms (2016) and Benchaira et al. (2016) given by (3), (4) and (5) respectively.
Firstly, we generate 1000 pseudo-random samples X and Y of size $N = 100, 150$ and 200 from Burr's models, $\overline{F}(x) = \left(1+x^{1/\theta}\right)^{-\theta/\gamma_1}$ and $\bar{G}(x) = \left(1+x^{1/\theta}\right)^{-\theta/\gamma_2}$ $x \geq 0$. We fix $\theta = 1/4$ and choose the values 0.6 and 0.8 for γ_1 and $p = 0.7$ (resp. 0.9), that means the percentage of truncation is 30% (resp. 10%). The pertaining γ_2-value is obtained by solving the equation $p = \gamma_2/(\gamma_1+\gamma_2)$, for each couple (γ_1, p). We obtained 1000 pseudo-random samples X^* and Y^* of size $n \simeq pN$.
Next, we calculate the estimators values frame the observed data X^* and Y^*. For choosing the optimal number k_n of upper order statistics used in the computation of $\hat{\gamma}_1^{(GS)}$, $\hat{\gamma}_1^{(W)}$ and $\hat{\gamma}_1^{(B)}$ we adopt the Reiss and Thomas algorithm Reiss and Thomas (2007). In those simulations, we used the random threshold $X^*_{(n-k_n)}$ instead of t_n in the definition of $\hat{\gamma}_1^{(W)}$.
Also note that we only consider $k_n := k_1 = k_2$ in the expression (3), in this case $\hat{\gamma}_1^{(GS)}$ is the one considered in Benchaira et al. (2015).
Finally, we compute the absolute bias (abias) and root mean squared error (rmse) of these estimators, the results are summarized in Table 1 and Table 2. We see that our new estimator shows good performance compared to existing methods for small sample sizes.

Table 1: Bias and rmse of the estimators based on 1000 samples of Burr's models with $\gamma_1 = 0.6$, for $p = 0.7$ (top) and $p = 0.9$ (bottom).

			$\hat{\gamma}_1^{(Z)}$		$\hat{\gamma}_1^{(GS)}$		$\hat{\gamma}_1^{(W)}$		$\hat{\gamma}_1^{(B)}$	
p	N	n	*abias*	*rmse*	*abias*	*rmse*	*abias*	*rmse*	*abias*	*rmse*
	100	70	.008	.028	.422	7.310	.014	.243	.197	.447
0.7	150	104	.006	.013	.225	1.892	.011	.212	.154	.399
	200	139	.003	.010	.227	.993	.009	.187	.148	.363
	100	80	.004	.171	.122	4.751	.007	.178	.050	.556
0.9	150	120	.005	.073	.072	.537	.007	.143	.061	.392
	200	159	.006	.019	.084	.651	.006	.121	.068	.309

Table 2: Bias and rmse of the estimators based on 1000 samples of Burr's models with $\gamma_1 = 0.8$, for $p = 0.7$ (top) and $p = 0.9$ (bottom).

			$\hat{\gamma}_1^{(Z)}$		$\hat{\gamma}_1^{(GS)}$		$\hat{\gamma}_1^{(W)}$		$\hat{\gamma}_1^{(B)}$	
p	N	n	*abias*	*rmse*	*abias*	*rmse*	*abias*	*rmse*	*abias*	*rmse*
	100	70	.006	.019	.315	9.594	.017	.379	.247	.617
0.7	150	104	.009	.011	.308	2.803	.018	.365	.190	.515
	200	139	.008	.012	.256	1.192	.019	.291	.200	.513
	100	80	.023	.027	.093	5.440	.037	.183	.090	.713
0.9	150	120	.018	.020	.138	.786	.036	.161	.137	.467
	200	159	.010	.014	.110	.487	.034	.138	.102	.407

Now, in order to study the sensitivity to outliers of our newly estimator, we consider an ε-contaminated model known by mixture of Pareto distributions

$$F_{\gamma_1,\lambda,\varepsilon}(z) = 1-(1-\varepsilon)z^{-1/\gamma_1}+\varepsilon z^{-1/\lambda}, \qquad \gamma_1,\lambda > 0 \quad \textit{and} \quad 0 < \varepsilon < 0.5 \tag{13}$$

Note that, for $\gamma_1 < \lambda$ and $\varepsilon > 0$, (13) corresponds to a Pareto distribution contaminated by a longer tailed distribution.

In this context, we proceed our study as follows:
We fix $\lambda = 2$ and consider four different contamination levels $\varepsilon = 0.05, 0.15, 0.25, 0.35$, and we vary γ_1 among 0.6 and 0.8. For each value of ε, 1000 samples of size $N = 200$ were generated from the model (13) truncated by a simple Pareto model $\bar{G}(x) = x^{-1/\gamma_2}$, with $p = 0.7$ and 0.9.
Our illustration, made with respect to the biases and rmse's, are summarized in Table 3. The values of the first line are those of the case where $\varepsilon = 0$ (i.e., uncontaminated case). Both the bias and the rmse of our estimator are note sensitive to outliers. Then we can conclude its robustness, giving us, in fact, an excellent level of protection against contamination data.

Table 3: Bais and rmse of the estimators based on 1000 samples of a contaminated Pareto distribution, with tail index $\gamma_1 = 0.6$ (left) and $\gamma_1 = 0.8$ (right), $N = 200$.

p	0.7		0.9		0.7		0.9	
$\varepsilon\%$	*bias*	*rmse*	*bias*	*rmse*	*bias*	*rmse*	*bias*	*rmse*
0	.0088	.0137	.0052	.0998	.0265	.0180	.0189	.0558
5	.0104	.0558	.0644	.1112	.0562	.0591	.0698	.0954
15	.0153	.0921	.0905	.1568	.0872	.0938	.0954	.1589
25	.0256	.3336	.1256	.4451	.1010	.7470	.1615	.4785
35	.1414	.5330	.2115	.6121	.1726	.9221	.2121	.7787

We conclude from tables 1, 2 and 3 that our newly estimator perform better (with the smallest bias and root mean squared error), compared to existing methods, for small sample sizes and for both uncontaminated and contaminated cases.

4. Application

4.1. Estimation of an upper quantile

Estimation of e.v.i. is very important in the determination of high quantiles, upper tail probabilities, mean excess functions, and excess-of-loss and stop-loss reinsurance premiums. Consequently, small errors in estimation of this quantity can produce substantial impact in applications. Thus, for robust estimation of quantities based on γ_1 robust estimation of γ_1 itself is crucial. In other words, for a heavy tailed distributions, robust estimation of the high quantile Q_ε corresponding to upper tail probability ε, becomes of interest, and this may be carried out by robust estimation of γ_1. We refer to Brazauskas and Serfling (2000) for a detailed account of this issue.
Let (α_n) be some sequence of quantiles orders tending to 0, such that $\alpha_n = o\left(\overline{F}(s_n)\right)$, where (s_n) is a sequence of positive deterministic thresholds growing to infinity with n. Consequently, the quantile of F of order $(1-\alpha_n)$ is such that $\overline{F}(Q_{\alpha_n}) = \alpha_n$. We can then define an estimator $\hat{Q}_{\alpha_n,s_n}$ of Q_{α_n} :

$$\hat{Q}_{\alpha_n,s_n} = s_n \left(\alpha_n^{-1}\left(1 - F_n^{(LB)}(s_n)\right)\right)^{\hat{\gamma}_1^{(Z)}}.$$

A similar estimator is proposed in Worms and Worms (2016), but instead of $\hat{\gamma}_1^{(Z)}$ they consider the estimator $\hat{\gamma}_1^{(W)}(t_n)$ given by (4). Before we state the asymptotic normality of $\hat{Q}_{\alpha_n,s_n}$, we set $d_n := \overline{F}(s_n)/\alpha_n$ and assume that

$$d_n \to \infty \quad \text{and} \quad \sqrt{n}/\log(d_n) \to \infty, \quad \text{as } n \to \infty. \tag{14}$$

Moreover, from the classical second order condition (see, Bingham et al. 1989) for L_F, it follows that

$$\forall x > 0, \ \frac{L_F(tx)}{L_F(t)} - 1 \overset{t\to\infty}{\sim} \frac{x^\rho - 1}{\rho} h(t) \quad (\forall x > 1)$$

where L_F is slowly varying function at infinity and h is a positive measurable function, slowly varying with index $\rho < 0$. Set $\bar{H} := \bar{F}\bar{G}$, where H is the distribution function of $min(X,Y)$. The asymptotic normality result will then

require the following conditions on s_n :

$$n\bar{H}(s_n) \to \infty, \quad \text{as } n \to \infty \tag{15}$$

and

$$\sqrt{n\bar{H}(s_n)}h(s_n) \to \lambda, \quad \text{as } n \to \infty \quad (\text{for some } \lambda > 0). \tag{16}$$

Theorem 4.1. *Under (14), (15), (16) and the assumptions of Theorem 2.1, we have*

$$\frac{\sqrt{n}}{\log(d_n)}\left(\frac{\hat{Q}_{\alpha_n,s_n}}{Q_{\alpha_n}} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{v,u}\right), \quad as\ n \to \infty.$$

Proof. The result follows by analogous arguments as in the proof of Theorem 2 in Worms and Worms (2016). Recall that the high quantile Q_{α_n} corresponding to order $(1-\alpha_n)$ is such that $\overline{F}(Q_{\alpha_n}) = \alpha_n$, and its estimator is defined by

$$\hat{Q}_{\alpha_n,s_n} = s_n\left(\frac{\bar{F}_n(s_n)}{\alpha_n}\right)^{\hat{\gamma}_1^{(Z)}}.$$

For convenience, we set $\Lambda_n := \bar{F}_n(s_n)/\bar{F}(s_n)$. Indeed, we have

$$\begin{aligned}
\frac{\hat{Q}_{\alpha_n,s_n}}{Q_{\alpha_n}} - 1 &= \frac{s_n}{Q_{\alpha_n}}\left(\frac{\bar{F}_n(s_n)}{\alpha_n}\Lambda_n\right)^{\hat{\gamma}_1^{(Z)}} - 1 \\
&= \Lambda_n^{\hat{\gamma}_1^{(Z)}}\left\{\left(\frac{s_n}{Q_{\alpha_n}}d_n^{\gamma_1}d_n^{\left(\hat{\gamma}_1^{(Z)}-\gamma_1\right)} - 1\right) + \left(1 - \Lambda_n^{-\hat{\gamma}_1^{(Z)}}\right) + \left(\frac{s_n}{Q_{\alpha_n}}d_n^{\gamma_1} - 1\right)\right\} \\
&=: \Lambda_n^{\hat{\gamma}_1^{(Z)}}\left\{I_{n1} + I_{n2} + I_{n3}\right\}.
\end{aligned}$$

We will show that $\frac{\sqrt{n}}{\log(d_n)}I_{n1}$ is asymptotically centred Gaussian rv with variance $\sigma^2_{v,u}$ and $\frac{\sqrt{n}}{\log(d_n)}I_{nj} \xrightarrow{P} 0,\ j = 2,3$. Concerning the term I_{n1}, by using the mean value theorem, it follows that

$$\frac{\sqrt{n}}{\log(d_n)}I_{n1} = \sqrt{n}\left(\hat{\gamma}_1^{(Z)} - \gamma_1\right)\exp(\delta_n),$$

where $\delta_n \leq \left|\hat{\gamma}_1^{(Z)} - \gamma_1\right|\log(d_n)$. Assumption (14) and Theorem 2.1, allows us to conclude that δ_n tends to 0. We use then Theorem 2.1 to get.

$$\frac{\sqrt{n}}{\log(d_n)}I_{n1} \to N\left(0, \sigma^2_{v,u}\right), \quad \text{as } n \to \infty.$$

Let us now focus on the negligible terms I_{n2} and I_{n3}. By using the mean value theorem, we get

$$I_{n2} = \hat{\gamma}_1^{(Z)}M_n^{-\hat{\gamma}_1^{(Z)}-1}(\Lambda_n - 1),$$

with M_n tending to 1. In view of assumptions (A1) and (15), the sequence (Λ_n) converge to 1 in probability (see, Lemma 2 in Worms and Worms, 2016), we have then

$$\frac{\sqrt{n}}{\log(d_n)}(\Lambda_n - 1) = o_p(1).$$

Hence

$$\frac{\sqrt{n}}{\log(d_n)}I_{n2} = o_p(1).$$

For I_{n3}, in view of the regular variation of $\bar{F}$, (1) can be rephrased as $\bar{F}(x) = x^{-1/\gamma_1}L_F(x)$, where L_F is slowly varying

function at infinity and by definition of Q_{α_n}, we get

$$I_{n3} = \frac{s_n}{Q_{\alpha_n}} \left(\frac{\bar{F}(s_n)}{\bar{F}(Q_{\alpha_n})} \right)^{-\gamma_1} - 1 = \left(\frac{L_F(Q_{\alpha_n})}{L_F(s_n)} \right)^{-\gamma_1} - 1.$$

Therefore, we use the following representation of L_F (see, Smith 1987, page 1195)

$$L_F(x) = c\left(1 + \rho^{-1} h(x) + o(h(x))\right), \quad \text{for } x \to \infty$$

where h is a positive measurable function, slowly varying with index $\rho < 0$. We have, Q_{α_n}/s_n tends to infinity, then $h(Q_{\alpha_n})/h(s_n)$ tends to 0 and

$$\left| \frac{h(Q_{\alpha_n})}{h(s_n)} - \left(\frac{Q_{\alpha_n}}{s_n} \right)^{\rho} \right| \leq \sup_{w \geq 1} \left| \frac{h(ws_n)}{h(s_n)} - w^{\rho^{-1}} \right| \to 0.$$

It follows that

$$\begin{aligned} \frac{L_F(Q_{\alpha_n})}{L_F(s_n)} &= 1 - \rho^{-1} h(s_n) \left(1 - \frac{h(Q_{\alpha_n})}{h(s_n)} + o\left(\frac{h(Q_{\alpha_n})}{h(s_n)} \right) + o_p(1) \right) \\ &= 1 - \rho^{-1} h(s_n) (1 + o_p(1)). \end{aligned}$$

Therefore $|I_{n3}| \leq C |L_F(Q_{\alpha_n})/L_F(s_n) - 1|$, then

$$\frac{\sqrt{n}}{\log(d_n)} |I_{n3}| \leq C \frac{\sqrt{n}}{\log(d_n)} \rho^{-1} h(s_n) (1 + o_p(1))$$

and then the desired negligibility of I_{n3} follows from assumption (16), which ends the proof of the Theorem. □

4.2. Real data example : automobile brake pad lifetime

In reliability, a real data-set of lifetimes of automobile brake pads already considered in Lawless (2002), was recently analyzed in Gardes and Stupfler (2015) and Benchaira et al. (2016) as an application of heavy-tailed and RRT data. We follow the same steps as those in Gardes and Stupfler (2015) who transformed this sample into a right-truncated data, which originally is left-truncated. We are dealing with a data-set of small size ($n = 98$), consequently, robust estimation of γ_1 can produce substantial robust estimation of the high quantile. Then, our procedure should be preferred to that based on no robust estimation of γ_1. In these situation, we used the random threshold $X^*_{(n-k_n)}$ instead of s_n in the definition of $\hat{Q}_{\alpha_n, s_n}$. We select the optimal number of top statistics, via the numerical procedure of (Reiss and Thomas, 2007, page 137) and we get $k = 10$ and we estimate the tail index γ_1 given in (5) and (7) we get $\hat{\gamma}_1^{(B)} = 0.4701$ and $\hat{\gamma}_1^{(Z)} = 0.4925$ respectively.

The estimation results of our based (right-panel) and that of Benchaira et al. (2016) based (left-panel) extreme quantiles estimators with three different quantile levels corresponding to $\alpha_n = 0.001, 0.005, 0.010$ are summarized in Table 4. For instance, we conclude that the brake pad lifetime is estimated to be less than 17063 km for 1 % of the cars while only be out of a thousand brake pads lasts less than 10200 km.

Table 4: Extreme quantiles for automobile brake pad lifetimes.

Quantile level	$\hat{Q}_{\alpha_n}$ via $\hat{\gamma}_1^{(B)}$	$\hat{Q}_{\alpha_n}$ via $\hat{\gamma}_1^{(Z)}$
0.990	17604	17063
0.995	14641	14138
0.999	10559	10203

5. Concluding notes

The main objective of this paper was to propose a robust estimator for the extreme value index of Pareto-type distributions under randomly right-truncated data by using a Lynden-Bell integral and a useful huberized M-functional and M-estimators of the tail index. It has been shown that our newly estimator is more robust and perform better than the estimators proposed in Gardes and Stupfler (2015), Worms and Worms (2016), Benchaira et al. (2016), for small sample sizes and for both uncontaminated and contaminated cases. In our further research we will study this robust estimator in more detail. We will investigate its influence function and its relative asymptotic efficiency. Note also that, The degree of robustness is determined by the tuning parameters v and u. This paper does not treat the choice of these parameters, it remains a likely topic for future investigations. Finally, we emphasize that our approach may also be employed to derive several robust estimators of upper tail probabilities, mean excess functions, and excess-of-loss and stop-loss reinsurance premiums, in case of presence of random right truncation.

Acknowledgment: The authors wish to thank the editor and reviewers for their helpful comments in the earlier version of this paper.

References

1. Benatmane, C., Zeghdoudi, H., Shanker, R., and Lazri, N. (2020). Composite rayleigh-pareto distribution: Application to real fire insurance losses data set. *Journal of Statistics and Management Systems*, pages 1–13.
2. Benchaira, S., Meraghni, D., and Necir, A. (2015). On the asymptotic normality of the extreme value index for right-truncated data. *Statistics & Probability Letters*, 107:378–384.
3. Benchaira, S., Meraghni, D., and Necir, A. (2016). Tail product-limit process for truncated data with application to extreme value index estimation. *Extremes*, 19(2):219–251.
4. Beran, J. (1997). On heavy tail modeling and teletraffic data. *The Annals of statistics*, 25(5):1852–1856.
5. Beran, J. and Schell, D. (2012). On robust tail index estimation. *Computational Statistics & Data Analysis*, 56(11):3430–3443.
6. Bingham, N., Goldie, C., and Teugels, J. (1989). *Regular variation*, volume 27. Cambridge university press.
7. Brazauskas, V. and Serfling, R. (2000). Robust and efficient estimation of the tail index of a single-parameter pareto distribution. *North American Actuarial Journal*, 4(4):12–27.
8. de Haan, L. and Ferreira, A. (2007). *Extreme value theory: an introduction*. Springer.
9. Escudero, L. F. and Ortega, E. (2008). Actuarial comparisons for aggregate claims with randomly right-truncated claims. *Insurance: Mathematics and Economics*, 43(2):255–262.
10. Gardes, L. and Stupfler, G. (2015). Estimating extreme quantiles under random truncation. *Test*, 24(2):207–227.
11. Hampel, F., Ronchetti, E., Rousseeuw, P., and Stahel, W. (1986). *Robust statistics: the approach based on influence functions*. John Wiley & Sons, Inc.,New York.
12. He, S. and Yang, G. (1998). The strong law under random truncation. *Annals of statistics*, pages 992–1010.
13. Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *The annals of statistics*, pages 1163–1174.
14. Huber, P. (1981). *Robust statistics*. John Wiley & Sons, New York.
15. Lawless, J. (2002). *Statistical models and methods for lifetime data*. Second Edition. Wiley Series in Probability and Statistics.
16. Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3cr quasars. *Monthly Notices of the Royal Astronomical Society*, 155(1):95–118.
17. Reiss, R. and Thomas, M. (2007). *Statistical analysis of extreme values with Applications to Insurance, Finance, Hydrology and Other Fields*. 3rd ed. Birkhäuser Verlag, Basel, Boston, Berlin.
18. Resnick, S. (1997). Heavy tail modeling and teletraffic data: special invited paper. *The Annals of Statistics*, 25(5):1805–1869.
19. Resnick, S. (2006). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer.
20. Sayah, A., Yahia, D., and Brahimi, B. (2014). On robust tail index estimation under random censorship. *Afrika Statistika*, 9(1):671–683.
21. Smith, R. (1987). Estimating tails of probability distributions. *The annals of Statistics*, pages 1174–1207.
22. Stute, W. and Wang, J. (2008). The central limit theorem under random truncation. *Bernoulli*, 14(3):604–622.

23. Woodroofe, M. (1985). Estimating a distribution function with truncated data. *The Annals of Statistics*, 13(1):163–177.
24. Worms, J. and Worms, R. (2016). A lynden-bell integral estimator for extremes of randomly truncated data. *Statistics & Probability Letters*, 109:106–117.