Statistics of Extremes under Random Censoring

Einmahl, J.H.J.; Fils-Villetard, A.; Guillou, A.

Publication date:
2006

No. 2006–62

STATISTICS OF EXTREMES UNDER RANDOM CENSORING

By John H.J. Einmahl, Amélie Fils-Villetard, Armelle Guillou

June 2006

ISSN 0924-7815

Statistics of Extremes under Random Censoring

John H.J. Einmahl
Tilburg University

Amélie Fils-Villetard
Université Paris VI

Armelle Guillou
Université Paris VI

26th June 2006

Abstract. We investigate the estimation of the extreme value index, when the data are subject to random censorship. We prove in a unified way detailed asymptotic normality results for various estimators of the extreme value index and use these estimators as the main building block for estimators of extreme quantiles. We illustrate the quality of these methods by a small simulation study and apply the estimators to medical data.

Running title: Statistics of censored extremes.
AMS 2000 subject classifications. 62G05, 62G20, 62G32, 62N02.
JEL codes. C13, C14, C41.
Keywords and phrases. Asymptotic normality, extreme value index, extreme quantiles, random censoring.

1 Introduction

Let $X_1, ..., X_n$ be a sample of n independent and identically distributed (i.i.d.) random variables, distributed according to an unknown distribution function (df) F. A question of great interest is how to obtain a good estimator for a quantile

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

where ε is so small that this quantile is situated on the border of or beyond the range of the data. Estimating such extreme quantiles is directly linked to the accurate modeling and estimation of the tail of the distribution

$$\overline{F}(x) := 1 - F(x) = \mathbb{P}(X > x)$$

for large thresholds x. From extreme value theory, the behaviour of such extreme quantile estimators is known to be governed by one crucial parameter of the underlying distribution, the extreme value index. This parameter is important since it measures the tail heaviness of F. This estimation has been widely studied in the literature: we mention for instance Hill (1975), Smith (1987), Dekkers *et al.* (1989), and Drees *et al.* (2004).

However, in classical applications such as the analysis of lifetime data (survival analysis, reliability theory, insurance) a typical feature which appears is censorship. Quite often, X represents the time elapsed from the entry of a patient, say, in a follow-up study until death. If at the time that the data collection is performed, the patient is still alive or has withdrawn from the study for some reason, the variable of interest X will not be available. A convenient way to model this situation is the introduction of a random variable Y, independent of X, such that only

$$Z = X \wedge Y \quad \text{and} \quad \delta = \mathbb{1}_{\{X \leq Y\}} \tag{1}$$

are observed. The indicator variable δ determines whether X has been censored or not. Given a random sample (Z_i, δ_i), $1 \leq i \leq n$, of independent copies of (Z, δ), it is our goal to make inference on the right tail of the unknown lifetime df F, while G, the df of Y, is considered to be a nonparametric nuisance parameter.

Statistics of extremes of randomly censored data is a new research field. The statistical problems in this field are difficult, since typically only a small fraction of the data can be used for inference in the far tail of F and in the case of censoring these data are, moreover, not fully informative. The topic has first been mentioned in Reiss and Thomas (1997, Section 6.1) where an estimator of a positive extreme value index is introduced, but no (asymptotic) results are derived. Recently, Beirlant *et al.* (2006) proposed estimators for the general

extreme value index and for an extreme quantile. They made a start with the analysis of the asymptotic properties of some estimators, but only when using the data above a deterministic threshold. Obviously, in practice, the threshold is random (typically an order statistic), which renders the proof of the asymptotics much more complicated.

For almost all applications of extreme value theory, the estimation of the extreme value index is of primordial importance. Consequently, it is the main aim of this paper to propose a unified method to prove asymptotic normality for various estimators of the extreme value index under random censoring. We apply our estimators to the problem of extreme quantile estimation under censoring. We illustrate our results with simulations and also apply our methods to AIDS survival data.

We consider data on patients diagnosed with AIDS in Australia before 1 July 1991. The source of these data is Dr P. J. Solomon and the Australian National Centre in HIV Epidemiology and Clinical Research; see Venables and Ripley (2002). The information on each patient includes gender, date of diagnosis, date of death or end of observation, and an indicator which of the two is the case. The data set contains 2843 patients, of which 1761 died; the other survival times are right censored. We will apply our methodology to the 2754 male patients (there are only 89 women in the data set). Apart from assessing the heaviness of the right tail of the survival function $1-F$ by means of the estimation of the extreme value index, it is also important to estimate very high quantiles of F, thus getting a good indication of how long very strong men will survive AIDS.

Another application, that we will not pursue in this paper, is to annuity insurance contracts. Life annuities are contractual guarantees, issued by insurance companies, pension plans, and government retirement systems, that offer promises to provide periodic income over the lifetime of individuals. If we monitor the policyholders during a certain period, the data are right censored since many policyholders survive until the end of the observation period. We are interested in the far right tail of the future lifetime distribution of the annuitants, since longevity is an important and difficult risk to evaluate for insurance companies. In the

case of life annuities it needs to be estimated as accurately as possible for setting adequate insurance premiums.

We will study estimators for the extreme value index of F, assuming that F and G are both in the max-domain of attraction of an extreme value distribution. In Section 2, we introduce various estimators of this extreme value index and we establish in a unified way their asymptotic behavior; we also introduce an estimator for very high quantiles. Some examples are given in Section 3 and a small simulation study is performed. Our estimators are applied to the AIDS data in Section 4.

2 Estimators and main results

Let $X_1, ..., X_n$ be a sequence of i.i.d. random variables from a df F. We denote the order statistics by

$$X_{1,n} \leq \ldots \leq X_{n,n}.$$

The weak convergence of the centered and standardized maxima $X_{n,n}$, means the existence of sequences of constants $a_n > 0$ and b_n and a df $\widetilde{G}$ such that

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{X_{n,n} - b_n}{a_n} \leq x\Big) = \widetilde{G}(x), \tag{2}$$

for all x where $\widetilde{G}$ is continuous. The work by Fisher and Tippett (1928), Gnedenko (1943), and de Haan (1970) answered the question on the possible limits and characterized the classes of dfs F having a certain limit in (2).

This convergence result is our main assumption. Up to location and scale, the possible limiting dfs $\widetilde{G}$ in (2) are given by the so-called *extreme value distributions* G_γ defined by

$$G_\gamma(x) = \begin{cases} \exp(-(1+\gamma x)^{-1/\gamma}) & \text{if } \gamma \neq 0, \\ \exp(-\exp(-x)) & \text{if } \gamma = 0. \end{cases} \tag{3}$$

We say that F is in the (max-)domain of attraction of G_γ, notation $F \in D(G_\gamma)$. Here γ is the extreme value index. Knowledge of γ is crucial for estimating the right tail of F.

We review briefly some estimators of γ that have been proposed in the literature. The most famous one is probably the Hill (1975) estimator

$$\widehat{\gamma}^{(H)}_{X,k,n} := M^{(1)}_{X,k,n} := \frac{1}{k}\sum_{i=1}^{k} \log X_{n-i+1,n} - \log X_{n-k,n}, \tag{4}$$

where $k \in \{1, \ldots, n\}$. However, this estimator is only useful when $\gamma > 0$. A generalization which works for any $\gamma \in \mathbb{R}$ is the so-called Moment estimator, introduced in Dekkers *et al.* (1989):

$$\widehat{\gamma}^{(M)}_{X,k,n} := M^{(1)}_{X,k,n} + S_{X,k,n} := M^{(1)}_{X,k,n} + 1 - \frac{1}{2}\left(1 - \frac{(M^{(1)}_{X,k,n})^2}{M^{(2)}_{X,k,n}}\right)^{-1}, \tag{5}$$

with

$$M^{(2)}_{X,k,n} := \frac{1}{k}\sum_{i=1}^{k} (\log X_{n-i+1,n} - \log X_{n-k,n})^2.$$

The Hill estimator can be derived in several ways, one very appealing being the slope of the Pareto quantile plot, which consists of the points

$$\left(\log \frac{n+1}{i}, \log X_{n-i+1,n}\right),\ i = 1, \ldots, k.$$

This plot has been generalized in Beirlant *et al.* (1996) by defining $UH_{i,n} = X_{n-i,n}\widehat{\gamma}^{(H)}_{X,i,n}$ and by considering the points

$$\left(\log \frac{n+1}{i}, \log UH_{i,n}\right),\ i = 1, \ldots, k.$$

This generalized quantile plot becomes almost linear for small enough k, i.e. for extreme values. It follows immediately that the slope of this graph will estimate γ no matter if it is positive, negative or zero. An estimator of this slope is given by

$$\widehat{\gamma}^{(UH)}_{X,k,n} := \frac{1}{k}\sum_{i=1}^{k} \log UH_{i,n} - \log UH_{k+1,n}. \tag{6}$$

A quite different estimator of γ, is the so-called maximum likelihood (ML) estimator. (Note that the classical, parametric ML approach is not applicable, because

F is not in a parametric family.) The approach relies on results in Balkema and de Haan (1974) and Pickands (1975), stating that the limit distribution of the exceedances $E_j = X_j - t$ $(X_j > t)$ over a threshold t, when t tends to the right endpoint of F, is given by a Generalized Pareto distribution depending on two parameters, γ and σ. In practice t is replaced by an order statistic $X_{n-k,n}$, and the resulting ML-estimators are denoted by $\widehat{\gamma}^{(ML)}_{X,k,n}$ and $\widehat{\sigma}^{(ML)}_{X,k,n}$.

In the case of censoring, we would like to adapt all these methods. Actually, we will provide a general adaptation of estimators of the extreme value index and a unified proof of their asymptotic normality; the four estimators above are special cases of this. We assume that both F and G are absolutely continuous and that $F \in D(G_{\gamma_1})$ and $G \in D(G_{\gamma_2})$, for some $\gamma_1, \gamma_2 \in \mathbb{R}$. The extreme value index of H, the df of Z defined in (1), exists and is denoted by γ. Let $\tau_F = \sup\{x : F(x) < 1\}$ (resp. τ_G and τ_H) denote the right endpoint of the support of F (resp. G and H). In the sequel, we assume that the pair (F, G) is in one of the following three cases:

$$\begin{cases} \text{case 1:} & \gamma_1 > 0, \gamma_2 > 0, & \text{in this case} & \gamma = \frac{\gamma_1\gamma_2}{\gamma_1+\gamma_2} \\ \text{case 2:} & \gamma_1 < 0, \gamma_2 < 0,\ \tau_F = \tau_G, & \text{in this case} & \gamma = \frac{\gamma_1\gamma_2}{\gamma_1+\gamma_2}\,. \\ \text{case 3:} & \gamma_1 = \gamma_2 = 0,\ \tau_F = \tau_G = \infty, & \text{in this case} & \gamma = 0 \end{cases} \tag{7}$$

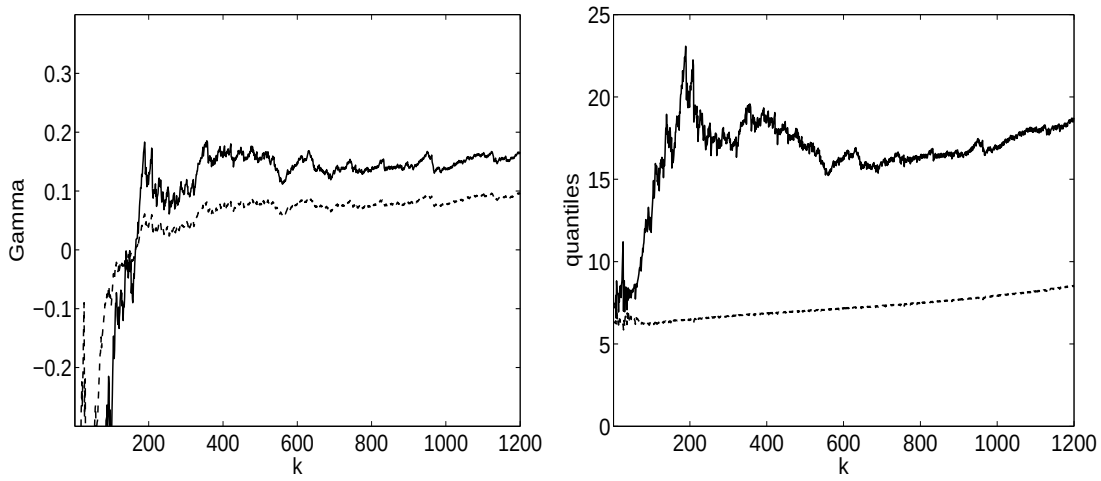
(In case 3 we also define for convenient presentation $\frac{\gamma_1\gamma_2}{\gamma_1+\gamma_2} = \gamma$.) The other possibilities are not very interesting. Typically they are very close to the "uncensored case" (which has been studied in detail in the literature) or the "completely censored situation" (where estimation is impossible).

The first important point that should be mentioned is the fact that all the preceding estimators (Hill, Moment, UH or ML) are obviously not consistent if they are based on the sample $Z_1, ..., Z_n$, that means if the censoring is not taken into account. Indeed, they all converge to γ, the extreme value index of the Z-sample, and not to γ_1, the extreme value index of F. Consequently, we have to adapt all these estimators to censoring. We will divide all these estimators by the proportion of non-censored observations in the k largest Zs:

$$\widehat{\gamma}^{(c,.)}_{Z,k,n} = \frac{\widehat{\gamma}^{(.)}_{Z,k,n}}{\widehat{p}} \quad \text{where} \quad \widehat{p} = \frac{1}{k}\sum_{j=1}^{k} \delta_{[n-j+1,n]},$$

with $\delta_{[1,n]}, ..., \delta_{[n,n]}$ the δs corresponding to $Z_{1,n}, ..., Z_{n,n}$, respectively. $\widehat{\gamma}^{(.)}_{Z,k,n}$ could be any estimator not adapted to censoring, in particular $\widehat{\gamma}^{(H)}_{Z,k,n}, \widehat{\gamma}^{(M)}_{Z,k,n}, \widehat{\gamma}^{(UH)}_{Z,k,n}$ or $\widehat{\gamma}^{(ML)}_{Z,k,n}$. It will be our main aim to study in detail the asymptotic normality of these estimators.

To illustrate the difference between the estimators, adapted and not adapted to censoring, we plot in Figure 1(a), $\widehat{\gamma}^{(UH)}_{Z,k,n}$ (dashed line) and $\widehat{\gamma}^{(c,UH)}_{Z,k,n}$ (full line) as a function of k for the AIDS survival data. Note that the censoring in the tail is higher than 70%, much higher than the censoring in the whole sample. Nevertheless, we see a quite stable plot when k ranges from about 200 (or 350) to 1200 and a substantial difference between both estimators. Similar graphs could be presented for the other estimators.



(a) (b)

Figure 1: UH-estimator adapted (full line) and not adapted (dashed line) to censoring (a) for the extreme value index and (b) for the extreme quantile with $\varepsilon = 0.001$ for the AIDS survival data.

Let us now consider the estimation of an extreme quantile $x_\varepsilon = F^{\leftarrow}(1-\varepsilon)$. Denoting with $\widehat{F}_n$ the Kaplan-Meier (1958) product-limit estimator, we can adapt the classical estimators proposed in the literature as follows:

$$\widehat{x}_{\varepsilon,k}^{(c,.)} = Z_{n-k,n} + \widehat{a}_{Z,k,n}^{(c,.)} \frac{\left(\frac{1-\widehat{F}_n(Z_{n-k,n})}{\varepsilon}\right)^{\widehat{\gamma}_{Z,k,n}^{(c,.)}} - 1}{\widehat{\gamma}_{Z,k,n}^{(c,.)}} , \tag{8}$$

where

$$\widehat{a}_{Z,k,n}^{(c,.)} = \frac{Z_{n-k,n}\, M_{Z,k,n}^{(1)}\,(1 - S_{Z,k,n})}{\widehat{p}} \text{ , for } M \text{ and } UH,$$

and

$$\widehat{a}_{Z,k,n}^{(c,ML)} = \frac{\widehat{\sigma}_{Z,k,n}^{(ML)}}{\widehat{p}} .$$

Note that these estimators are defined under the assumption that the two endpoints τ_F and τ_G are equal, but possibly infinite. This is true for the three cases defined in (7). Note also that we have excluded the Hill estimator since it only works in case 1.

Again, to observe the difference between the adapted and not adapted estimators, we plot in Figure 1(b), $\widehat{x}_{0.001,k}^{(UH)}$ (dashed line) and $\widehat{x}_{0.001,k}^{(c,UH)}$ (full line) for the AIDS data. The difference between both estimators (for k between 250 and 500) is about 10 years.

Beirlant *et al.* (2006) considered asymptotic properties of some of these estimators, but only for a deterministic threshold, that is when $Z_{n-k,n}$ is replaced by t in the preceding formulas. Note also that the asymptotic bias of these estimators has not been studied. Our aim in this paper is to establish the asymptotic normality (including bias and variance) of all these estimators based on k upper order statistics (or equivalently on a random threshold $Z_{n-k,n}$), using a unified approach.

To specify the asymptotic bias of the different estimators, we use a second order condition phrased in terms of the tail quantile function $U_H(x) = H^{-1}(1 - \frac{1}{x})$. From the theory of generalized regular variation of second order outlined in de Haan and Stadtmüller (1996), we assume the existence of a positive function a and a second eventually positive function a_2 with $\lim_{x\to\infty} a_2(x) = 0$, such that the

limit

$$\lim_{x\to\infty}\frac{1}{a_2(x)}\left\{\frac{U_H(ux)-U_H(x)}{a(x)}-h_\gamma(u)\right\}=k(u) \tag{9}$$

exists for $u\in(0,\infty)$, with $h_\gamma(u)=\int_1^u z^{\gamma-1}dz$. It follows that there exists a $c\in\mathbb{R}$ and a second order parameter $\rho\le 0$, for which the function a satisfies

$$\lim_{x\to\infty}\left\{\frac{a(ux)}{a(x)}-u^\gamma\right\}/a_2(x)=cu^\gamma h_\rho(u). \tag{10}$$

The function a_2 is regularly varying with index ρ. As usual, we will assume that $\rho<0$ and we will also assume that the slowly varying part of a_2 is asymptotically equivalent to a positive constant. For an appropriate choice of the function a, the function k that appears in (9) admits the representation

$$k(u)=Ah_{\gamma+\rho}(u), \tag{11}$$

with $A\neq 0$; now c in (10) is equal to 0. We denote the class of second order regularly varying functions U_H (satisfying (9)-(11) with $c=0$) by $GRV_2(\gamma,\rho;a(x),a_2(x);A)$. In Appendix 1, we give an overview of the possible forms of the GRV_2 functions and the corresponding representations for U_H.

In the statement of our results, we use the following notation (see Appendix 1):

$$b(x)=\begin{cases}\frac{A\rho[\rho+\gamma(1-\rho)]}{(\gamma+\rho)(1-\rho)}a_2(x) & \text{if } 0<-\rho<\gamma \text{ or if } 0<\gamma<-\rho \text{ with } D=0,\\ -\frac{\gamma^3}{(1+\gamma)}x^{-\gamma}L_2(x) & \text{if } \gamma=-\rho,\\ -\frac{\gamma^3 D}{(1+\gamma)}x^{-\gamma} & \text{if } 0<\gamma<-\rho \text{ with } D\neq 0,\\ \frac{1}{\log^2 x} & \text{if } \gamma=0,\\ \frac{A\rho(1-\gamma)}{(1-\gamma-\rho)}a_2(x) & \text{if } \gamma<\rho,\\ -\frac{\gamma}{1-2\gamma}\frac{\ell_+}{\tau_H}x^\gamma & \text{if } \rho<\gamma<0,\\ \frac{\gamma}{1-2\gamma}\left[A(1-\gamma)-\frac{\ell_+}{\tau_H}\right]x^\gamma & \text{if } \gamma=\rho,\end{cases}$$

and

$$\widetilde{\rho}=\begin{cases}-\gamma & \text{if } 0<\gamma<-\rho \text{ with } D\neq 0,\\ \rho & \text{if } -\rho\le\gamma \text{ or if } 0<\gamma<-\rho \text{ with } D=0, \text{ or if } \gamma<\rho,\\ \gamma & \text{if } \rho\le\gamma\le 0.\end{cases}$$

Before stating our main result, define:

$$p(z) = \mathbb{P}\left(\delta = 1 \,|\, Z = z\right).$$

Note that in cases 1 and 2, $\lim_{z\to\tau_H} p(z)$ exists and is equal to $\frac{\gamma_2}{\gamma_1+\gamma_2} =: p \in (0,1)$. Assume that in case 3 this limit also exists and is positive and denote it again by p. By convention we also define $\frac{\gamma_2}{\gamma_1+\gamma_2} = p$ for that case.

In the sequel, $k = k_n$ is an intermediate sequence, i.e. a sequence such that $k \to \infty$ and $\frac{k}{n} \to 0$, as $n \to \infty$. Our main result now reads as follows.

Theorem 1. *Under the assumptions that, for $n \to \infty$,*

$$\begin{cases} \sqrt{k}\, a_2\left(\frac{n}{k}\right) \to \alpha_1 \in \mathbb{R} & \text{for the } ML\text{-estimator} \\ \sqrt{k}\, b\left(\frac{n}{k}\right) \to \alpha_1 \in \mathbb{R} & \text{for the other three estimators,} \end{cases} \tag{12}$$

$$\frac{1}{\sqrt{k}} \sum_{i=1}^{k} \Big[p\Big(H^{-1}\Big(1 - \frac{i}{n}\Big)\Big) - p \Big] \longrightarrow \alpha_2 \in \mathbb{R}, \tag{13}$$

and

$$\sqrt{k} \sup_{\{1-\frac{k}{n} \le t < 1, |t-s| \le C\frac{\sqrt{k}}{n}, s<1\}} \Big| p\Big(H^{-1}(t)\Big) - p\Big(H^{-1}(s)\Big) \Big| \longrightarrow 0 \quad \text{for all} \quad C > 0, \tag{14}$$

we have for the four estimators (for the Hill estimator we assume case 1 holds and for the ML-estimator $\gamma > -\frac{1}{2}$)

$$\sqrt{k}\Big(\widehat{\gamma}^{(c,.)}_{Z,k,n} - \gamma_1\Big) \xrightarrow{d} \mathcal{N}\Big(\frac{1}{p}\Big(\alpha_1 b_0 - \gamma_1 \alpha_2\Big), \frac{\sigma^2 + \gamma_1^2\, p(1-p)}{p^2}\Big),$$

where $\alpha_1 b_0$ (resp. σ^2) denotes the bias (resp. the variance) of $\sqrt{k}\left(\widehat{\gamma}^{(.)}_{Z,k,n} - \gamma\right)$.

This leads to the following corollary, the proof of which is straightforward using the expressions for the asymptotic bias-terms of the four "uncensored" estimators, see Beirlant *et al.* (2005) and Drees *et al.* (2004).

Corollary 1. *Under the assumptions of Theorem 1, we have*

$$\sqrt{k}\Big(\widehat{\gamma}^{(c,H)}_{Z,k,n} - \gamma_1\Big) \xrightarrow{d} \mathcal{N}\Big(\mu^{(c,H)}, \frac{\gamma_1^3}{\gamma}\Big) \qquad \text{in case 1}$$

$$\sqrt{k}\Big(\widehat{\gamma}^{(c,M)}_{Z,k,n}-\gamma_1\Big) \xrightarrow{d} \begin{cases} \mathcal{N}\Big(\mu^{(c,M)},\frac{\gamma_1^2}{\gamma^2}\Big(1+\gamma_1\gamma\Big)\Big) & \textit{in case 1}\\ \mathcal{N}\Big(\mu^{(c,M)},\frac{\gamma_1^2(1-\gamma)^2(1-2\gamma)(1-\gamma+6\gamma^2)}{\gamma^2(1-4\gamma)(1-3\gamma)}+\gamma_1^2\Big(\frac{\gamma_1}{\gamma}-1\Big)\Big) & \textit{in case 2}\\ \mathcal{N}\Big(\mu^{(c,M)},\Big(\frac{\gamma_1+\gamma_2}{\gamma_2}\Big)^2\Big) & \textit{in case 3}\end{cases}$$

$$\sqrt{k}\Big(\widehat{\gamma}^{(c,UH)}_{Z,k,n}-\gamma_1\Big) \xrightarrow{d} \begin{cases} \mathcal{N}\Big(\mu^{(c,UH)},\frac{\gamma_1^2}{\gamma^2}\Big(1+\gamma_1\gamma\Big)\Big) & \textit{in case 1}\\ \mathcal{N}\Big(\mu^{(c,UH)},\frac{\gamma_1^2(1-\gamma)(1+\gamma+2\gamma^2)}{\gamma^2(1-2\gamma)}+\gamma_1^2\Big(\frac{\gamma_1}{\gamma}-1\Big)\Big) & \textit{in case 2}\\ \mathcal{N}\Big(\mu^{(c,UH)},\Big(\frac{\gamma_1+\gamma_2}{\gamma_2}\Big)^2\Big) & \textit{in case 3}\end{cases}$$

$$\sqrt{k}\Big(\widehat{\gamma}^{(c,ML)}_{Z,k,n}-\gamma_1\Big) \xrightarrow{d} \mathcal{N}\Big(\mu^{(c,ML)},\frac{\gamma_1^2}{\gamma^2}\Big[1+\gamma(2+\gamma_1)\Big]\Big) \qquad \textit{in cases 1, 3, and 2 with } \gamma>-\tfrac{1}{2}$$

where

$$\mu^{(c,H)} := -\frac{\gamma_1\alpha_2}{p}+\frac{\alpha_1}{p}\frac{\gamma}{\widetilde{\rho}+\gamma(1-\widetilde{\rho})}$$

$$\mu^{(c,M)} := -\frac{\gamma_1\alpha_2}{p}+\frac{\alpha_1}{p}\cdot\begin{cases} \frac{1}{1-\widetilde{\rho}} & \textit{in case 1}\\ \frac{2\gamma-1}{\widetilde{\rho}(1-\widetilde{\rho})} & \textit{in case 2, if } \rho<\gamma\\ \frac{1-2\gamma}{(1-\gamma)(1-3\gamma)}\frac{A(1-\gamma)^2-(\gamma+1)\frac{\ell_+}{\tau_H}}{A(1-\gamma)-\frac{\ell_+}{\tau_H}} & \textit{in case 2, if } \rho=\gamma\\ \frac{1-2\gamma}{1-2\gamma-\widetilde{\rho}} & \textit{in case 2, if } \gamma<\rho\\ 1 & \textit{in case 3}\end{cases}$$

$$\mu^{(c,UH)} := -\frac{\gamma_1\alpha_2}{p}+\frac{\alpha_1}{p(1-\widetilde{\rho})}$$

$$\mu^{(c,ML)} := -\frac{\gamma_1\alpha_2}{p}+\frac{\alpha_1}{p}\frac{\rho(\gamma+1)A}{(1-\rho)(1-\rho+\gamma)}\,.$$

Proof of Theorem 1. We consider the following decomposition

$$\begin{aligned}\sqrt{k}\Big(\widehat{\gamma}^{(c,.)}_{Z,k,n}-\gamma_1\Big) &= \frac{1}{\widehat{p}}\sqrt{k}\Big(\widehat{\gamma}^{(.)}_{Z,k,n}-\gamma\Big)+\frac{1}{\widehat{p}}\sqrt{k}\Big(\gamma-\gamma_1\widehat{p}\Big)\\ &= \frac{1}{\widehat{p}}\sqrt{k}\Big(\widehat{\gamma}^{(.)}_{Z,k,n}-\gamma\Big)+\frac{\gamma_1}{\widehat{p}}\sqrt{k}\Big(\frac{\gamma_2}{\gamma_1+\gamma_2}-\widehat{p}\Big). \qquad (15)\end{aligned}$$

The asymptotic behavior of $\sqrt{k}\Big(\widehat{\gamma}^{(.)}_{Z,k,n}-\gamma\Big)$ is well-known since this estimator is

based on the Z-sample, i.e. on the uncensored situation; see Beirlant *et al.* (2005) and Drees *et al.* (2004).

First note that in case 3, $\gamma_1 = \gamma = 0$. Therefore, the second term in the decomposition (15) is exactly 0 as long as $\widehat{p} > 0$. That means that this case follows, since $\widehat{p} \xrightarrow{\mathbb{P}} p > 0$. Now we focus in detail on the second term of the decomposition in (15) for the cases 1 and 2.

To this aim, consider the following construction: Let Z be a random variable with df H. Let U have a uniform-$(0,1)$ distribution and be independent of Z. Define now

$$\delta = \begin{cases} 1 & \text{if } U \leq p(Z) \\ 0 & \text{if } U > p(Z) \end{cases}$$

and

$$\widetilde{\delta} = \begin{cases} 1 & \text{if } U \leq p \\ 0 & \text{if } U > p. \end{cases}$$

We repeat this construction independently n times. It is easy to show that the resulting pairs (Z_i, δ_i), $i = 1, \ldots, n$, have the same distribution as the initial pairs (Z_i, δ_i), $i = 1, \ldots, n$, for all $n \in \mathbb{N}$, so we continue with the new pairs (Z_i, δ_i).

Moreover, clearly Z and $\widetilde{\delta}$ are independent and

$$\mathbb{P}\left(|\delta - \widetilde{\delta}| = 1 \,|\, Z = z\right) = |p - p(z)|.$$

Consider the order statistics $Z_{1,n} \leq \ldots \leq Z_{n,n}$ and denote the induced order statistics of the Us by $U_{[1,n]}, \ldots, U_{[n,n]}$. We can write $\widehat{p}$ as follows:

$$\widehat{p} = \frac{1}{k} \sum_{j=1}^{k} \mathbb{1}_{\{U_{[n-j+1,n]} \leq p(Z_{n-j+1,n})\}},$$

and similarly

$$\widetilde{p} := \frac{1}{k} \sum_{j=1}^{k} \widetilde{\delta}_{[n-j+1,n]} = \frac{1}{k} \sum_{j=1}^{k} \mathbb{1}_{\{U_{[n-j+1,n]} \leq p\}}.$$

Clearly $U_{[1,n]}, ..., U_{[n,n]}$ are i.i.d. and independent of the Z-sample.

We use the following decomposition:

$$\sqrt{k}\Big(\widehat{p}-p\Big)=\sqrt{k}\Big(\widehat{p}-\widetilde{p}\Big)+\sqrt{k}\Big(\widetilde{p}-p\Big). \tag{16}$$

Since $\widetilde{p}\overset{d}{=}\frac{1}{k}\sum_{j=1}^{k}\mathbb{1}_{\{U_j\leq p\}}$, we have

$$\sqrt{k}\Big(\widetilde{p}-p\Big)\overset{d}{\longrightarrow}\mathcal{N}\Big(0,p(1-p)\Big).$$

Now, we are interested in $\sqrt{k}\Big(\widehat{p}-\widetilde{p}\Big)$, which turns out to be a bias-term. It can be rewritten as follows

$$\begin{aligned}\sqrt{k}\Big(\widehat{p}-\widetilde{p}\Big) &\overset{d}{=} \frac{1}{\sqrt{k}}\sum_{j=1}^{k}\Big[\mathbb{1}_{\{U_j\leq p(Z_{n-j+1,n})\}}-\mathbb{1}_{\{U_j\leq p\}}\Big]\\ &= \frac{1}{\sqrt{k}}\sum_{j=1}^{k}\Big[\mathbb{1}_{\{U_j\leq p(Z_{n-j+1,n})\}}-\mathbb{1}_{\{U_j\leq p(H^{-1}(1-\frac{j}{n}))\}}\Big]\\ &\quad+\frac{1}{\sqrt{k}}\sum_{j=1}^{k}\Big[\mathbb{1}_{\{U_j\leq p(H^{-1}(1-\frac{j}{n}))\}}-\mathbb{1}_{\{U_j\leq p\}}\Big]\\ &=: T_{1,k}+T_{2,k}.\end{aligned}$$

Then, under the assumptions (13) and (14), the convergence in probability of $T_{2,k}$ to α_2 follows from a result in Chow and Teicher (1997), p.356.

So we need now to show that $T_{1,k}\overset{\mathbb{P}}{\longrightarrow}0$. To this aim, write $V_i=H(Z_i)$, so that $Z_i=H^{-1}(V_i)$. The V_i are i.i.d. uniform-$(0,1)$. Write also $r(t)=p(H^{-1}(t))$. Then

$$T_{1,k}=\frac{1}{\sqrt{k}}\sum_{j=1}^{k}\Big[\mathbb{1}_{\{U_j\leq r(V_{n-j+1,n})\}}-\mathbb{1}_{\{U_j\leq r(1-\frac{j}{n})\}}\Big].$$

By the weak convergence of the uniform tail quantile process we have uniformly in $1\leq j\leq k$,

$$V_{n-j+1,n}-\Big(1-\frac{j}{n}\Big)=O_{\mathbb{P}}\Big(\frac{\sqrt{k}}{n}\Big).$$

Let $\eta > 0$. Using (14), we have with arbitrarily high probability, for large n,

$$\begin{aligned} |T_{1,k}| &\leq \frac{1}{\sqrt{k}} \sum_{j=1}^{k} \left| \mathbb{1}_{\{U_j \leq r(V_{n-j+1,n})\}} - \mathbb{1}_{\{U_j \leq r(1-\frac{j}{n})\}} \right| \\ &\stackrel{d}{=} \frac{1}{\sqrt{k}} \sum_{j=1}^{k} \mathbb{1}_{\{U_j \leq |r(V_{n-j+1,n}) - r(1-\frac{j}{n})|\}} \leq \frac{1}{\sqrt{k}} \sum_{j=1}^{k} \mathbb{1}_{\{U_j \leq \frac{\eta}{\sqrt{k}}\}}. \end{aligned}$$

Using the aforementioned result in Chow and Teicher (1997), p.356, and the fact that $\eta > 0$ can be chosen arbitrarily small, $T_{1,k} \xrightarrow{\mathbb{P}} 0$ follows.

Finally, combining (15) and (16) yields

$$\sqrt{k}\Big(\widehat{\gamma}^{(c,.)}_{Z,k,n} - \gamma_1\Big) = \frac{1}{\widehat{p}}\Big(\sqrt{k}\Big(\widehat{\gamma}^{(.)}_{Z,k,n} - \gamma\Big) - \gamma_1\sqrt{k}\Big(\widetilde{p} - p\Big)\Big) - \frac{\gamma_1\alpha_2}{\widehat{p}} + o_{\mathbb{P}}(1), \quad (17)$$

with the two terms within the brackets independent, since the first one is based on the Z-sample and the second one on the U-sample. Therefore, under the assumptions (12)–(14), we have

$$\sqrt{k}\Big(\widehat{\gamma}^{(c,.)}_{Z,k,n} - \gamma_1\Big) \xrightarrow{d} \mathcal{N}\Big(\frac{1}{p}\Big(\alpha_1 b_0 - \gamma_1\alpha_2\Big), \frac{\sigma^2 + \gamma_1^2\, p(1-p)}{p^2}\Big). \qquad \square$$

3 Examples and small simulation study

In this section we consider two examples: first a Burr distribution censored by another Burr distribution (so an example of case 1), and second a ReverseBurr distribution censored by another ReverseBurr distribution (an example of case 2). We show that these distributions satisfy all the assumptions and calculate the bias-terms explicitly. In particular, we will see how assumptions (12) and (13) compare. We also provide simulations to illustrate the behavior of our estimators for these distributions.

Example 1: $X \sim Burr(\beta_1, \tau_1, \lambda_1)$ and $Y \sim Burr(\beta_2, \tau_2, \lambda_2)$, $\beta_1, \tau_1, \lambda_1, \beta_2, \tau_2, \lambda_2 > 0$.

In that case

$$1-F(x) = \Big(\frac{\beta_1}{\beta_1+x^{\tau_1}}\Big)^{\lambda_1} = x^{-\tau_1\lambda_1}\beta_1^{\lambda_1}\Big(1+\beta_1 x^{-\tau_1}\Big)^{-\lambda_1}, x>0;$$
$$1-G(x) = \Big(\frac{\beta_2}{\beta_2+x^{\tau_2}}\Big)^{\lambda_2} = x^{-\tau_2\lambda_2}\beta_2^{\lambda_2}\Big(1+\beta_2 x^{-\tau_2}\Big)^{-\lambda_2}, x>0.$$

We can infer that

$$U_H(x) = H^{-1}\Big(1-\frac{1}{x}\Big) = \Big(\beta_1^{\lambda_1}\,\beta_2^{\lambda_2}x\Big)^{\frac{1}{\tau_1\lambda_1+\tau_2\lambda_2}}\left[1-\gamma\eta\Big(\beta_1^{\lambda_1}\,\beta_2^{\lambda_2}\,x\Big)^{\rho}(1+o(1))\right]$$

with $\tau=\min(\tau_1,\tau_2)$, $\rho=-\gamma\tau$, and $\eta=\begin{cases}\lambda_1\beta_1 & \text{if } \tau_1<\tau_2\\ \lambda_2\beta_2 & \text{if } \tau_1>\tau_2\\ \lambda_1\beta_1+\lambda_2\beta_2 & \text{if } \tau_1=\tau_2\end{cases}$.

The different parameters of interest are the following

$$\gamma_1=\frac{1}{\lambda_1\tau_1}\ ,\ \gamma_2=\frac{1}{\lambda_2\tau_2} \qquad \text{and} \qquad \gamma=\frac{1}{\lambda_1\tau_1+\lambda_2\tau_2}.$$

First, we check assumption (14). Using the above approximation of H^{-1}, it follows for $s\le t<1$ and s large enough, that

$$|p(H^{-1}(t))-p(H^{-1}(s))| \le \widetilde{C}\left((1-s)^{\gamma\tau}-(1-t)^{\gamma\tau}\right),$$

for some $\widetilde{C}>0$. It now easily follows that in case $\gamma\tau\ge 1$, the left hand side of (14) tends to 0. In case $\gamma\tau<1$ the left hand side of (14) is of order $\sqrt{k}\left(\frac{\sqrt{k}}{n}\right)^{\gamma\tau} = \sqrt{k}\left(\frac{n}{k}\right)^{\rho}k^{\rho/2}$, which tends to 0 when (13) holds (see below).

The asymptotic bias of $\sqrt{k}(\widehat{\gamma}^{(.)}_{Z,k,n}-\gamma)$ can be explicitly computed (from Corollary 1) and is asymptotically equivalent to:

$$-\eta\Big(\beta_1^{\lambda_1}\beta_2^{\lambda_2}\Big)^{\rho}\sqrt{k}\left(\frac{n}{k}\right)^{\rho}\cdot\begin{cases}\frac{\gamma\rho}{1-\rho} & \text{for the Hill estimator}\\ \frac{\rho(1+\gamma)(\gamma+\rho)}{(1-\rho)(1-\rho+\gamma)} & \text{for the } ML\text{-estimator}\\ \frac{\rho[\rho+\gamma(1-\rho)]}{(1-\rho)^2} & \text{for the Moment and } UH\text{-estimators}\end{cases}.$$

They are all of the same order.

We obtain another bias-term from assumption (13). Direct computations lead to

$$p(z)-p=\frac{\gamma^2}{\gamma_1\gamma_2}\Big[-\beta_1 z^{-\tau_1}(1+o(1))+\beta_2 z^{-\tau_2}(1+o(1))\Big],$$

when $\tau_1 \neq \tau_2$, or $\tau_1 = \tau_2$ and $\beta_1 \neq \beta_2$. Consequently, assumption (13) is equivalent to

$$\beta \frac{\gamma^2}{\gamma_1 \gamma_2} \Big(\beta_1^{\lambda_1} \beta_2^{\lambda_2}\Big)^{\rho} \frac{1}{1-\rho} \sqrt{k} \Big(\frac{n}{k}\Big)^{\rho} \longrightarrow \alpha_2,$$

$$\text{with } \beta = \begin{cases} -\beta_1 & \text{if } \tau_1 < \tau_2 \\ \beta_2 & \text{if } \tau_1 > \tau_2 \\ \beta_2 - \beta_1 & \text{if } \tau_1 = \tau_2 \end{cases}.$$

So both bias-terms are of the same order. Only, when $\tau_1 = \tau_2$ and $\beta_1 = \beta_2$ (in particular when $F \equiv G$) the biases of the estimators of γ dominate.

Example 2: $X \sim ReverseBurr(\beta_1, \tau_1, \lambda_1, x_+)$ and $Y \sim ReverseBurr(\beta_2, \tau_2, \lambda_2, x_+)$, $\beta_1, \tau_1, \lambda_1, \beta_2, \tau_2, \lambda_2, x_+ > 0$.

In that case

$$1 - F(x) = \Big(\frac{\beta_1}{\beta_1 + (x_+ - x)^{-\tau_1}}\Big)^{\lambda_1} = (x_+ - x)^{\tau_1 \lambda_1} \beta_1^{\lambda_1} \Big(1 + \beta_1 (x_+ - x)^{\tau_1}\Big)^{-\lambda_1},\ x < x_+ ;$$

$$1 - G(x) = \Big(\frac{\beta_2}{\beta_2 + (x_+ - x)^{-\tau_2}}\Big)^{\lambda_2} = (x_+ - x)^{\tau_2 \lambda_2} \beta_2^{\lambda_2} \Big(1 + \beta_2 (x_+ - x)^{\tau_2}\Big)^{-\lambda_2},\ x < x_+ .$$

Define τ and η as in Example 1, but set $\rho = \gamma\tau$ now. We can infer that

$$U_H(x) = H^{-1}\Big(1 - \frac{1}{x}\Big) = x_+ - \Big(\beta_1^{\lambda_1}\, \beta_2^{\lambda_2} x\Big)^{-\frac{1}{\tau_1 \lambda_1 + \tau_2 \lambda_2}} \Big[1 - \gamma\eta \Big(\beta_1^{\lambda_1}\, \beta_2^{\lambda_2}\, x\Big)^{\rho} (1 + o(1))\Big].$$

The different parameters of interest are the following

$$\gamma_1 = -\frac{1}{\lambda_1 \tau_1}\ ;\ \gamma_2 = -\frac{1}{\lambda_2 \tau_2}\ ;\ \gamma = -\frac{1}{\lambda_1 \tau_1 + \lambda_2 \tau_2} \qquad \text{and} \qquad \tau_F = \tau_G = \tau_H = x_+.$$

Note that we can easily prove (as in Example 1) that assumption (14) is satisfied if we assume (13).

The asymptotic bias of $\sqrt{k}(\widehat{\gamma}^{(.)}_{Z,k,n} - \gamma)$ can be explicitly computed (again from Corollary 1) and is asymptotically equivalent with:

• For the UH-estimator:

$$\begin{cases} -\frac{\gamma^2 \tau (1-\gamma)(1+\tau)}{(1-\gamma-\gamma\tau)(1-\gamma\tau)}\, \eta \left(\beta_1^{\lambda_1} \beta_2^{\lambda_2}\right)^{\rho} \sqrt{k} \left(\frac{n}{k}\right)^{\rho} & \text{if } \tau < 1 \\ \frac{\gamma^2}{(1-\gamma)(1-2\gamma)} \Big(\beta_1^{\lambda_1} \beta_2^{\lambda_2}\Big)^{\rho} \Big[-2\eta(1-\gamma) + \frac{1}{x_+} \Big] \sqrt{k} \left(\frac{n}{k}\right)^{\rho} & \text{if } \tau = 1 \\ \frac{\gamma^2}{(1-\gamma)(1-2\gamma) x_+} \Big(\beta_1^{\lambda_1} \beta_2^{\lambda_2}\Big)^{\gamma} \sqrt{k} \left(\frac{n}{k}\right)^{\gamma} & \text{if } \tau > 1 \end{cases}$$

• For the Moment estimator:

$$\begin{cases} -\frac{\gamma^2\tau(1-\gamma)(1+\tau)(1-2\gamma)}{(1-\gamma-\gamma\tau)(1-2\gamma-\gamma\tau)}\,\eta\Big(\beta_1^{\lambda_1}\beta_2^{\lambda_2}\Big)^{\rho}\sqrt{k}\left(\frac{n}{k}\right)^{\rho} & \text{if } \tau<1 \\ -\frac{\gamma^2}{(1-\gamma)(1-3\gamma)}\Big(\beta_1^{\lambda_1}\beta_2^{\lambda_2}\Big)^{\rho}\Big[2\,\eta(1-\gamma)^2-\frac{\gamma+1}{x_+}\Big]\sqrt{k}\left(\frac{n}{k}\right)^{\rho} & \text{if } \tau=1 \\ -\frac{\gamma}{(1-\gamma)x_+}\Big(\beta_1^{\lambda_1}\beta_2^{\lambda_2}\Big)^{\gamma}\sqrt{k}\left(\frac{n}{k}\right)^{\gamma} & \text{if } \tau>1 \end{cases}$$

• For the ML-estimator, if $\gamma > -\frac{1}{2}$:

$$-\frac{\gamma^2\tau(1+\gamma)(1+\tau)}{(1-\gamma\tau)(1+\gamma-\gamma\tau)}\eta\Big(\beta_1^{\lambda_1}\beta_2^{\lambda_2}\Big)^{\rho}\sqrt{k}\Big(\frac{n}{k}\Big)^{\rho}.$$

They are all of the same order if $\tau \leq 1$, otherwise the biases of the Moment and UH-estimators dominate the one of the ML-estimator.

Similarly to Example 1: if $\tau_1 \neq \tau_2$, or $\tau_1 = \tau_2$ and $\beta_1 \neq \beta_2$, direct computations lead to

$$p(z) - p = \frac{\gamma^2}{\gamma_1\gamma_2}\Big[-\beta_1(x_+ - z)^{\tau_1}(1+o(1)) + \beta_2(x_+ - z)^{\tau_2}(1+o(1))\Big].$$

Consequently, assumption (13) is equivalent in that case to

$$\beta\frac{\gamma^2}{\gamma_1\gamma_2}\Big(\beta_1^{\lambda_1}\beta_2^{\lambda_2}\Big)^{\rho}\frac{1}{1-\rho}\sqrt{k}\Big(\frac{n}{k}\Big)^{\rho} \longrightarrow \alpha_2.$$

Again, this order is the same as the order of the asymptotic bias-terms of all the estimators, in case $\tau \leq 1$ and dominated by the one of the Moment and UH-estimators, otherwise. When $\tau_1 = \tau_2$ and $\beta_1 = \beta_2$ the biases of the estimators of γ dominate.

In order to illustrate these two examples, we simulate 100 samples of size 500 from the following distributions:

- a Burr$(10, 4, 1)$ censored by a Burr$(10, 1, 0.5)$,
- a ReverseBurr$(1, 8, 0.5, 10)$ censored by a ReverseBurr$(10, 1, 0.5, 10)$.

For both examples $p = \frac{8}{9}$, meaning that the percentage of censoring in the right tail is close to 11%. In the first case we have $\gamma_1 = \frac{1}{4}, \gamma = \frac{2}{9}$, and $\rho = -\frac{2}{9}$, in

the second case $\gamma_1 = -\frac{1}{4}, \gamma = -\frac{2}{9}$, and again $\rho = -\frac{2}{9}$. In both examples, the panels (a) and (c) (in Figures 2 and 3) represent the median for the index and the extreme quantile respectively, whereas the panels (b) and (d) represent the empirical mean squared errors (MSE) based on the 100 samples. The (very small) value of ε is $\frac{1}{5000}$; observe that $n\varepsilon = \frac{1}{10}$. All these estimators plotted are adapted to censoring. The horizontal line represents the true value of the parameter.

In the first example, we can observe, in the case of the estimation of the index, the superiority of the Hill estimator adapted to censoring in terms of MSE, the three others being quite similar. For the extreme quantile estimators, however, there is much less to decide between all the estimators: they are very stable and close to the true value of the parameter. A similar observation can be done for the second example, with a slight advantage for the UH-estimator only, in the case of the estimation of the index. Note that the medians of the extreme quantile estimates differ less than 0.1% from the true value.

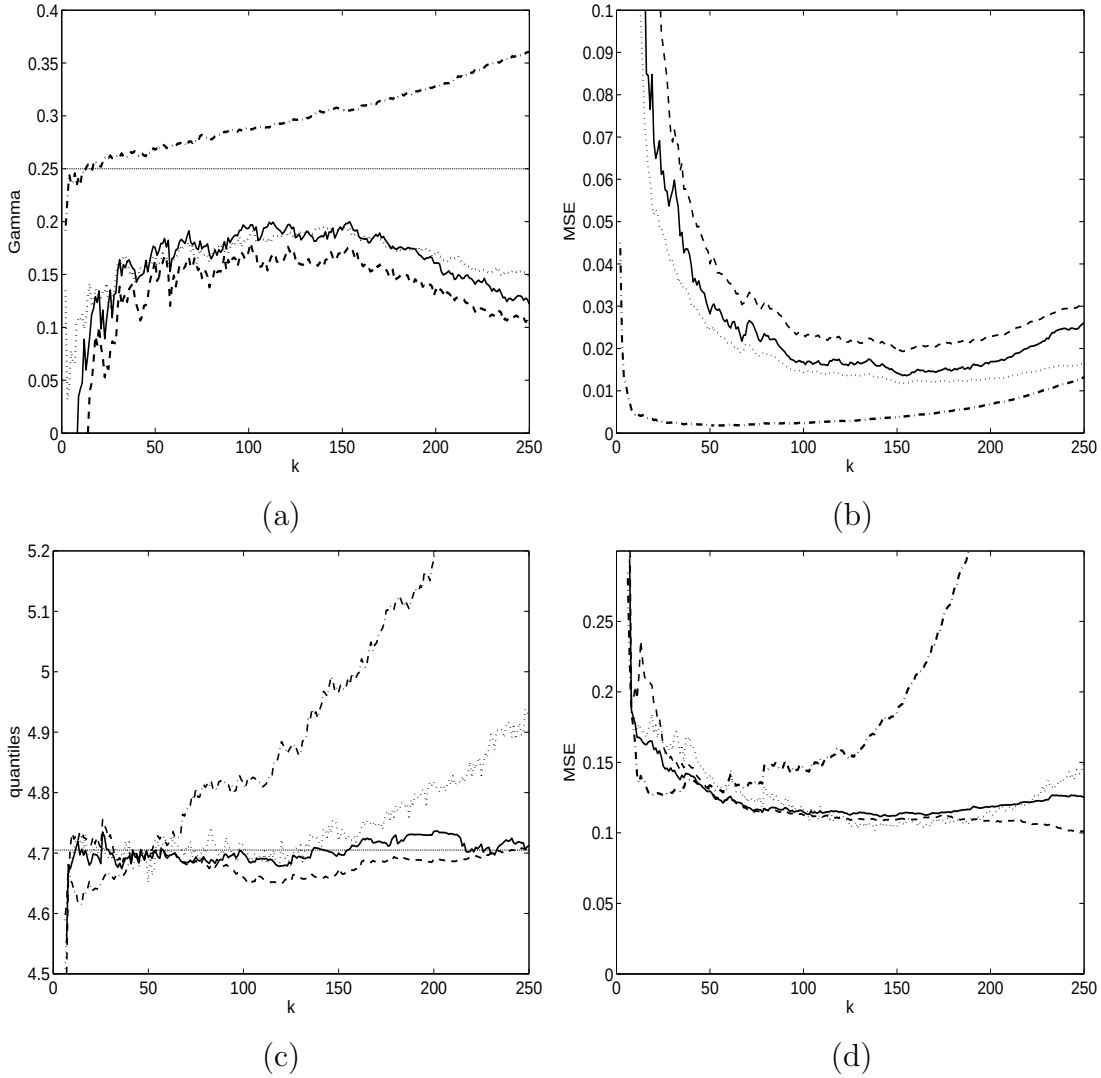


Figure 2: A Burr$(10, 4, 1)$ distribution censored by a Burr$(10, 1, 0.5)$ distribution: UH-estimator (dotted line), Moment estimator (full line), ML-estimator (dashed line) and Hill estimator (dashed-dotted line); (a) Median and (b) MSE for the extreme value index; (c) Median and (d) MSE for the extreme quantile with $\varepsilon = \frac{1}{5000}$.

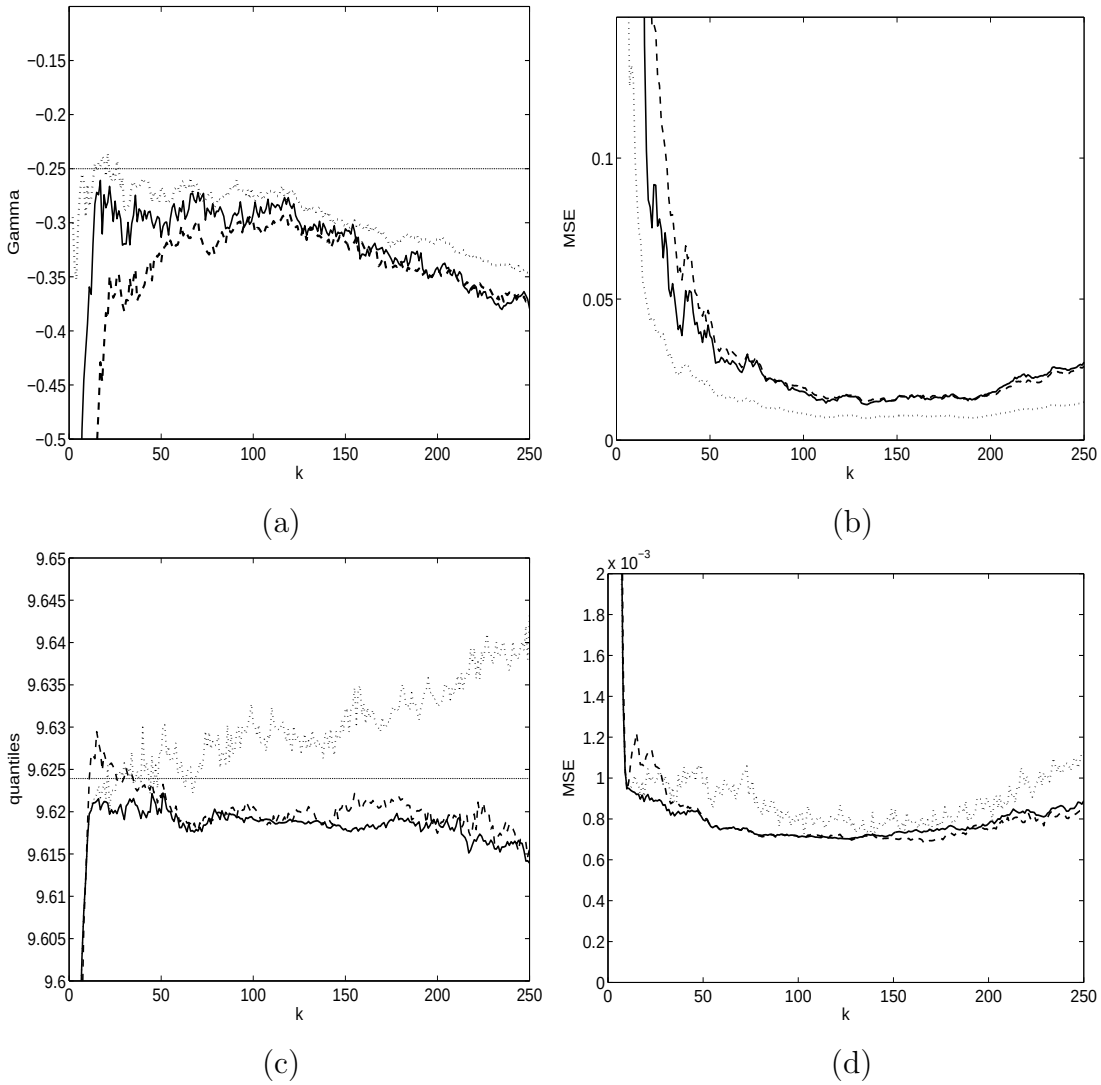


Figure 3: A ReverseBurr$(1, 8, 0.5, 10)$ distribution censored by a ReverseBurr$(10, 1, 0.5, 10)$ distribution: UH-estimator (dotted line), Moment estimator (full line), ML-estimator (dashed line); (a) Median and (b) MSE for the extreme value index; (c) Median and (d) MSE for the extreme quantile with $\varepsilon = \frac{1}{5000}$.

4 Application to AIDS survival data

We return to our real data set presented in Section 1 and used in Section 2, i.e. the Australian AIDS survival data for the male patients diagnosed before 1 July 1991. The sample size is 2754.

First we estimate $p = \lim_{z \to \tau_H} p(z)$. In Figure 4, we see $\widehat{p}$ as a function of k. Clearly there is a stable part in the plot when k ranges from about 75 until 175; for higher k the bias sets in. Note that $\widehat{p}$ is the mean of 0-1 variables, so for a sample of this size, the estimator is already very accurate. Therefore we estimate p with the corresponding vertical level in the plot, which is 0.28.

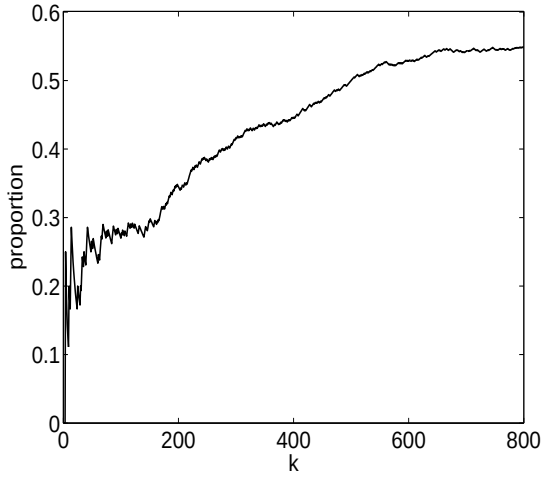
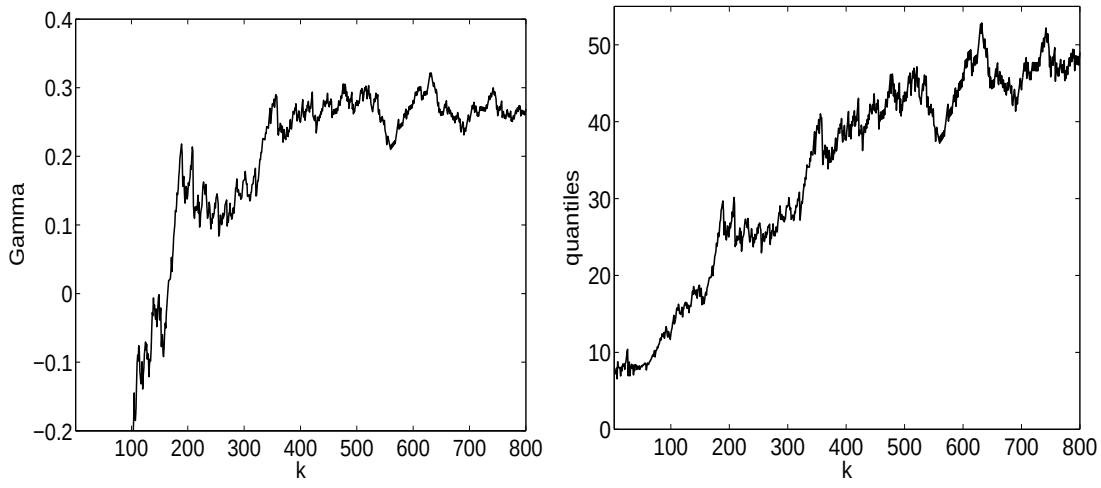


Figure 4: Estimator of p for the Australian AIDS survival data for the male patients.

Now we continue with the estimation of the extreme value index γ_1 and an extreme quantile $F^{\leftarrow}(1-\varepsilon)$, using the UH-method (as in Section 2). We will plot these estimators again as a function of k, but replace $\widehat{p} = \widehat{p}(k)$ already with its estimate 0.28, in order to prevent that the bias plays a dominant role for values of k larger than 200, say.

In Figure 5(a), the estimator of the extreme value index is presented, whereas Figure 5(b) shows the extreme quantile estimator for $\varepsilon = 0.001$. The estimator of



(a) (b)

Figure 5: UH-estimator (a) for the extreme value index and (b) for the extreme quantile with $\varepsilon = 0.001$, for the Australian AIDS survival data for the male patients.

γ_1 is quite stable for values of k between 200 and 300; we estimate it with 0.14. This indicates that the survival times are heavy tailed. We estimate the extreme quantile with k-values in the same range, because that range gives again a stable part in the plot. The corresponding estimated survival time is as high as about 25 years. So, although the estimated median survival time has the low value 1.3 years, due to the somewhat heavy tailed nature of the survival distribution, we find that exceptionally strong males can survive AIDS for 25 years.

References

[1] Balkema, A. and de Haan, L. (1974). Residual life at great age, *Ann. Probab.*, **2**, 792-804.

[2] Beirlant, J., Dierckx, G and Guillou, A. (2005). Estimation of the extreme-value index and generalized quantile plots, *Bernoulli*, **11**, 949-970.

[3] Beirlant, J., Guillou, A., Delafosse, E. and Fils-Villetard, A. (2006). Estimation of the extreme value index and extreme quantiles under random censoring, *Extremes*,

under revision.

[4] Beirlant, J., Vynckier, P. and Teugels, J.L. (1996). Tail index estimation, Pareto quantile plots, and regression diagnostics, *J. Amer. Statist. Assoc.*, **91**, 1659-1667.

[5] Chow and Teicher (1997). *Probability theory. Independence, interchangeability, martingales*, Third edition, Springer-Verlag, New York.

[6] Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution, *Ann. Statist.*, **17**, 1833-1855.

[7] Draisma, G., de Haan, L., Peng, L. and Pereira, T.T. (1999). A bootstrap based method to achieve optimality in estimating the extreme value index, *Extremes*, **2**, 367-404.

[8] Drees, H., Ferreira, A. and de Haan, L. (2004). On maximum likelihood estimation of the extreme value index, *Ann. Appl. Probab.*, **14**, 1179-1201.

[9] Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution in the largest particle size and smallest member of a sample, *Proc. Camb. Phil. Soc.*, **24**, 180-190.

[10] Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire, *Ann. Math.*, **44**, 423-453.

[11] de Haan, L. (1970). *On regular variation and its application to the weak convergence of sample extremes*, Mathematical Centre Tracts 32, Amsterdam.

[12] de Haan, L. and Stadtmüller, U. (1996). Generalized regular variation of second order, *J. Austral. Math. Soc. Ser. A*, **61**, 381-395.

[13] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Ann. Statist.*, **3**, 1163-1174.

[14] Kaplan, E.L. and Meier, P. (1958). Non-parametric estimation from incomplete observations, *J. Amer. Statist. Assoc.*, **53**, 457–481.

[15] Pickands III, J. (1975). Statistical inference using extreme order statistics, *Ann. Statist.*, **3**, 119-131.

[16] Reiss, R.D. and Thomas, M. (1997). *Statistical analysis of extreme values with applications to insurance, finance, hydrology and other fields*, Birkhäuser Verlag, Basel.

[17] Smith, R.L. (1987). Estimating tails of probability distributions, *Ann. Statist.*, **15**, 1174-1207.

[18] Vanroelen, G. (2003). *Generalized regular variation, order statistics and real inversion formulas*, Ph.D. thesis, Katholieke Universiteit Leuven, Leuven, Belgium.

[19] Venables, W.N. and Ripley, B.D. (2002). *Modern Applied Statistics with S.* Fourth edition, Springer-Verlag, New York.

Appendix 1: Overview of GRV_2 functions with $\rho < 0$

From Vanroelen (2003), we obtain the following representations of U_H; see also the appendix in Draisma *et al.* (1999).

• $0 < -\rho < \gamma$: for $U_H \in GRV_2(\gamma, \rho; \ell_+ x^\gamma, a_2(x); A)$:

$$U_H(x) = \ell_+ x^\gamma \Big\{ \frac{1}{\gamma} + \frac{A}{\gamma+\rho} a_2(x)(1+o(1)) \Big\},$$

• $\gamma = -\rho$: for $U_H \in GRV_2(\gamma, -\gamma; \ell_+ x^\gamma, x^{-\gamma}\ell_2(x); A)$:

$$U_H(x) = \ell_+ x^\gamma \Big\{ \frac{1}{\gamma} + x^{-\gamma} L_2(x) \Big\}$$

with $L_2(x) = B + \int_1^x (A + o(1)) \frac{\ell_2(t)}{t} dt + o(\ell_2(x))$ for some constant B and some slowly varying function ℓ_2,

• $0 < \gamma < -\rho$: for $U_H \in GRV_2(\gamma, \rho; \ell_+ x^\gamma, a_2(x); A)$:

$$U_H(x) = \ell_+ x^\gamma \Big\{ \frac{1}{\gamma} + D x^{-\gamma} + \frac{A}{\gamma+\rho} a_2(x)(1+o(1)) \Big\},$$

• $\gamma = 0$: for $U_H \in GRV_2(0, \rho; \ell_+, a_2(x); A)$:

$$U_H(x) = \ell_+ \log x + D + \frac{A\ell_+}{\rho} a_2(x)(1+o(1)),$$

• $\gamma < 0$: for $U_H \in GRV_2(\gamma, \rho; \ell_+ x^\gamma, a_2(x); A)$:

$$U_H(x) = \tau_H - \ell_+ x^\gamma \Big\{ \frac{1}{-\gamma} - \frac{A}{\gamma+\rho} a_2(x)(1+o(1)) \Big\},$$

where $\ell_+ > 0, A \neq 0, D \in \mathbb{R}$.

JHJE
Dept. of Econometrics & OR and CentER
Tilburg University
P.O. Box 90153
5000 LE Tilburg
The Netherlands
j.h.j.einmahl@uvt.nl

AF-V & AG
Laboratoire de Statistique Théorique et Appliquée
Université Paris VI
Boîte 158
4 Place Jussieu
75252 Paris Cedex 05
France
fils@ccr.jussieu.fr & guillou@ccr.jussieu.fr