*Probability in the Engineering and Informational Sciences*, **29**, 2015, 277–290.
doi:10.1017/S0269964814000345

# EXTREME VALUE DISTRIBUTIONS FOR BIASED SAMPLES

GEORGE TZAVELAS
*Department of Statistics and Insurance Science,*
*University of Piraeus, 80 Karaoli and Dimitriou str., 185 34 Piraeus, Greece*
*E-mail: tzafor@unipi.gr (Corresponding author)*

POLYCHRONIS ECONOMOU
*Department of Civil Engineering, University of Patras,*
*Rion-Patras, Greece*
*E-mail: peconom@upatras.gr*

In this work, the relationship between the extreme value distributions of the parent and its weighted counterpart distribution is studied. Sufficient conditions are provided in terms of the weight $w(x)$ under which the weighted and parent distributions belong to the same attractor (Fréchet, Weibull, or Gumbel) and the relation of the corresponding shape parameters of the limiting distributions is presented. Additionally, a biased sampling corrected extreme value index is proposed, when the extreme value index is estimated from a biased sample. Finally, some simulation results are presented that suggest the superiority of a biased sample in estimating the extreme value index.

## 1. INTRODUCTION

A biased sample arises when the probability of a unit of the population to be selected in a sample is proportional to some non-negative weight function $w(x)$ of its size $x$. If $f(x)$ is the probability density function (p.d.f.) of $X$, the obtained weighted sample can be interpreted as a random sample from a population with p.d.f. given by

$$f_w(x) = \frac{w(x)}{E\left(w(X)\right)} f(x).$$

The above weighted p.d.f. is well defined when $E\left(w(X)\right) < \infty$.

Any truncated distribution can be viewed and interpreted as a weighted distribution with $w(x) = I(X \in A)$ for some $A \in S_X$ (with $S_X$ the support of $X$) with $P(A) > 0$. In the present paper, only left truncated distributions will be considered since we do not want the truncation to change the right end-point of the support of the rv. Another weight function which is frequently used for non-negative rvs is the size-biased weight $x^r$. In this case, the p.d.f. of the weighted version of the random variable (rv) $X$ (with p.d.f. $f$) is

© Cambridge University Press 2015 0269-9648/15 

https://doi.org/10.1017/S0269964814000345 Published online by Cambridge University Press

expressed as

$$f_r(x) = \frac{x^r}{E\,(X^r)} f(x),$$

when $E(X^r) = \mu_r < \infty$.

The case of biased sampling is a rather common phenomenon not only in lifetime data (Simon [11], Zelen and Feinleib [12]), but also in environmental and meteorological data. Biased sampling cannot be avoided in some fields (tumor detections, chronic diseases identified by early detection screening programs, objects detection using radars, individuals with values below detection limits and many others). In many such studies, we are interested in and we record the maximum or the minimum value of a biased sample. Consider for example the case where the maximum values of some biological parameters such as the cholesterol level or the blood pressure of patients are recorded from a series of measurements taken during their stay at the intensive care unit after an ischemic heart disease problem. This is a typical case of a block maxima study. A similar situation appears in extreme environmental phenomena studies where the block is the year or the season (see for example Finkenstädt and Rootzén [2] and references therein). The aim of this work is to study the effects of the biased sampling scheme at the extreme value distribution of the rv under consideration. Let us consider a continuous type rv $X > 0$ with p.d.f. $f$ and cumulative distribution function (cdf) $F$ as well as its $w$-weighted version with p.d.f. and cdf $f_w$ and $F_w$, respectively. Let also $X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$, where $X_1, X_2, \ldots, X_n$ is a sequence of i.i.d. observations from the rv $X$ with p.d.f. $f$. The fundamental theorem of extreme value theory (Gnedenko [3], Gumbel [4]) claims that if there exist a sequence of numbers $b_n$ and a sequence of positive numbers $a_n$ such that for all real values $x$

$$P\left(a_n X_{(n)} + b_n \leq x\right) \to G(x;\alpha) \text{ as } n \to \infty,$$

where $G$ is a non-degenerate distribution function, then $G$ belongs to one of the following families:

$$\begin{aligned}
G_1(x;\alpha) &= \exp[-x^{-\alpha}] \quad x > 0 \quad \alpha > 0 \quad \text{(Fréchet distr.)},\\
G_2(x;\alpha) &= \exp[-(-x)^{\alpha}] \quad x \leq 0 \quad \alpha > 0 \quad \text{(Weibull distr.)},\\
G_3(x;0) &= \exp[-\exp(-x)] \quad x \in \mathbb{R} \quad \text{(Gumbel distr.)}.
\end{aligned}$$

The parameter $\alpha$ is the shape parameter of the distribution. A useful tool in modeling extreme values is the generalized extreme-value (GEV) or Jenkinson distribution which combines into a single expression the three possible limiting distributions. The distribution function of the GEV distribution is given by

$$G(x;\mu,\sigma,\xi) = \begin{cases} \exp\left[-\left(1+\xi\dfrac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right], & 1+\xi\dfrac{x-\mu}{\sigma} > 0,\ \xi \neq 0,\\ \exp\left[-e^{-(x-\mu)/\sigma}\right], & -\infty < x < \infty,\ \xi = 0, \end{cases}$$

where $\mu \in \mathbb{R}$, $\sigma > 0$, and $\xi \in \mathbb{R}$ are the location, scale and shape parameters, respectively. For $\xi = 0$, $\xi > 0$ and $\xi < 0$ the GEV distribution corresponds to the Gumbel, Fréchet, and Weibull distribution, respectively. It is worth mentioning that the shape parameter $\xi$ of the distribution is connected with the parameter $\alpha$ of the $G_1, G_2$ distributions above through the relationships $\alpha = 1/\xi, \alpha = -1/\xi$, respectively. The $G_3$ distribution corresponds to $\xi = 0$. We write $F \in D(G_i)$ (or $F \in D(G_\xi)$ depending on what we want to emphasize) to denote

https://doi.org/10.1017/S0269964814000345 Published online by Cambridge University Press

that the limiting distribution of $X_n^* = a_n X_{(n)} + b_n$ (with $X_n \sim F$) is $G_i$ for $i = 1, 2, 3$. In this case, we say that $F$ belongs to the domain of attraction $D(G_i)$ of $G_i$. Apparently the fundamental theorem of the extreme value theory holds if the sample is a $w$-weighted sample with distribution $F_w$. The main question under study in this work is whether the $F$ and the $F_w$ share the same attractor $G_i$.

In the next section, the main results of the paper are presented and a biased sampling corrected extreme value index is proposed, when the extreme value index is estimated from a biased sample. In Section 3, some simulation results, based on the block maxima method are presented by generating independent and identically distributed random samples from the Burr ($D(G_1)$), the beta ($D(G_2)$), the log-normal ($D(G_3)$) and the Weibull ($D(G_3)$) distributions. The simulations results, using the bias and the mean square error (MSE) of the maximum likelihood estimator for the extreme value index, suggest the superiority of a biased sample in the estimation procedure. It is worth mentioning that for all the above distributions except the lognormal, the bias and MSE converge fast to zero as n tends to infinity. This relatively bad behavior of the maximum-likelihood estimator for the extreme value index in the log-normal case is an interesting problem and it deserves more study. Finally, some concluding remarks are given in Section 4.

In terms of the notation we write $E_w(X)$ for the mean value of $X$ with respect to $f_w$. The mean excess function (or mean residual life) for $F$ and $F_w$ is denoted by $m(x) = E(X - x|X > x)$ and $m_w(x) = E_w(X - x|X > x)$, respectively. Also $A(x) = E(w(X) - w(x)|X > x)$ is the mean residual weight. The notation $g'(t)$ stands for the derivative of $g$ in terms of $t$. The right end-point of the distribution $F$ is defined as $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$. Also the notation $f \sim g$ as $x \to x_F$ means that $f(x)/g(x) \to d$ for $x \to x_F$, where $0 < d < \infty$. The failure (hazard) rate of an rv $X$ is $\lambda(x) = f(x)/(1 - F(x))$. Throughout this paper we will assume that $E(w(X)) < \infty$.

## 2. MAIN RESULTS

In this section, the main results of the paper are presented. The results are obtained with the help of the next known theorem (Gnedenko [3] and Resnick [10]) presented here only for completeness purposes and to facilitate the reader.

THEOREM 1: *Let $F(x)$ be the cdf of $X$. Then the following hold*

(i) *$F \in D(G_1)$ if and only if $x_F = \infty$ and*

$$\lim_{x\to\infty} \frac{1 - F(x)}{1 - F(kx)} = k^{\alpha}, \text{ for every } k > 0 \text{ and } \alpha > 0. \tag{1}$$

(ii) *$F \in D(G_2)$ if and only if $x_F < \infty$ and*

$$\lim_{x\to 0-} \frac{1 - F(kx + x_F)}{1 - F(x + x_F)} = k^{\alpha}, \quad \text{for every } k > 0 \text{ and } \alpha > 0. \tag{2}$$

(iii) *$F \in D(G_3)$ if and only if there exists some $z < x_F$ such that the survival function $\overline{F}(x) = 1 - F(x)$ has a representation*

$$\overline{F}(x) = c(x) \exp\left\{ - \int_z^x \frac{g(t)}{a(t)} dt \right\}, \quad z < x < x_F, \tag{3}$$

*where $c$ and $g$ are measurable functions satisfying*

$$c(x) \to c > 0, \quad g(x) \to 1 \text{ as } x \to x_F$$

https://doi.org/10.1017/S0269964814000345 Published online by Cambridge University Press

*and $a(x)$ is a positive absolutely continuous functions (with respect to Lebesque measure) with density $a'(x)$ having $\lim a'(x) \to 0$ as $x \to x_F$. One possible choice for $a(x)$ is the mean excess function $m(t)$ for which also holds that $\lim m'(x) \to 0$ as $x \to x_F$. For $x_F = \infty$ it holds that $(m(x))/x) \to 0$ as $x \to \infty$.*

### 2.1. A Closure Property of the Domain of Attractors $D(G_i)$

The next theorem provides a closure property of $D(G_i)$ $i = 1, 2, 3$ under biased sampling for a non-negative rv $X$.

THEOREM 2: *Let $F$ be a distribution function such that the $F_w$ exists.*

(i) *If it holds that*

$$w(x) \sim x^r \text{ for } r \in \mathbb{R} \text{ as } x \to \infty, \tag{4}$$

*then $F \in D(G_1(\alpha))$ if and only if $F_w \in D(G_1(\alpha^*))$ with $\alpha^* = \alpha - r$, provided that $\alpha^* \geq 0$.*

(ii) *If it holds that*

$$w(x) \sim (x_F - x)^r \text{ for } r \in \mathbb{R} \text{ as } x \to x_F < \infty, \tag{5}$$

*then $F \in D(G_2(\alpha))$ if and only if $F_w \in D(G_2(\alpha^*))$ with $\alpha^* = \alpha + r$, provided that $\alpha^* \geq 0$.*

(iii) *If the weight function $w(x)$ satisfies the relationship*

$$\lim_{x \to x_F} [\log(w(x))]' m(x) = 0, \tag{6}$$

*then, $F \in D(G_3)$ if and only if $F_w \in D(G_3)$.*

PROOF:

(i) The existence of $F_w$ implies that $E(w(X)) < \infty$. Then under (4) we can easily conclude that $E(X^r) < \infty$. Indeed for sufficiently large $x_0$ we have that $x^r/w(x) < 3/(2d)$ for $x > x_0$. Thus

$$\int_{x_0}^{\infty} x^r f(x) dx \leq \frac{3}{2b} \int_{x_0}^{\infty} w(x) f(x) dx < \infty.$$

Since $f \in D(G_1(\alpha))$ we have that $r < \alpha$ (Embrechts, Klüppelberg, and Mikosch [1]). Furthermore, (4) implies that

$$\lim_{x \to \infty} \frac{w(x)}{w(kx)} = k^{-r}. \tag{7}$$

From Theorem 1(i) we have that if $F \in D(G_1)$ then

$$\lim_{x \to \infty} \frac{1 - F(x)}{1 - F(kx)} = k^{\alpha}, \quad \text{for every } k > 0 \text{ and } \alpha > 0. \tag{8}$$

On the other hand $F_w \in D(G_1)$ if and only if

$$\lim_{x \to \infty} \frac{1 - F_w(x)}{1 - F_w(kx)} = k^{\alpha^*}, \quad \text{for every } k > 0 \text{ and } \alpha^* > 0. \tag{9}$$

https://doi.org/10.1017/S0269964814000345 Published online by Cambridge University Press

Applying the L'Hospital rule in the last relationship we have that

$$\lim_{x\to\infty}\frac{1-F_w(x)}{1-F_w(kx)}=\lim_{x\to\infty}\frac{f_w(x)}{kf_w(kx)}$$

or equivalently

$$\lim_{x\to\infty}\frac{1-F_w(x)}{1-F_w(kx)}=\lim_{x\to\infty}\frac{w(x)f(x)}{w(kx)kf(kx)}=\lim_{x\to\infty}\frac{w(x)}{w(kx)}\lim_{x\to\infty}\frac{1-F(x)}{1-F(kx)}.$$

So, from the last equation and (7) it holds that

$$\lim_{x\to\infty}\frac{1-F_w(x)}{1-F_w(kx)}=k^{-r}\lim_{x\to\infty}\frac{1-F(x)}{1-F(kx)}, \tag{10}$$

which implies that $F\in D(G_1(\alpha))$ if and only if $F_w\in D(G_1(\alpha^*))$ for $\alpha^*=\alpha-r$.

(ii) From Theorem 1(ii) we have that $F$ belongs to $D(G_2)$ if and only if $x_F<\infty$ and

$$\lim_{x\to 0-}\frac{1-F(kx+x_F)}{1-F(x+x_F)}=k^{\alpha},\ \text{ for every } k>0.$$

With the use of L'Hospital rule the last relation implies

$$\lim_{x\to 0-}\frac{kf(kx+x_F)}{f(x+x_F)}=k^{\alpha},\ \text{ for every } k>0.$$

Concerning the $F_w$ distribution we have firstly that $x_{F_w}=x_F$ and secondly that

$$\lim_{x\to 0-}\frac{F_w(kx+x_{F_w})}{F_w(x+x_{F_w})}=\lim_{x\to 0-}\frac{w(kx+x_F)}{w(x+x_F)}\frac{kf(kx+x_F)}{f(x+x_F)}=k^rk^{\alpha}=k^{\alpha+r}$$

for $k>0$. Thus we conclude that $F\in D(G_2(\alpha))$ iff $F_w\in D(G_2(\alpha^*))$, where $\alpha^*=\alpha+r$ provided that $\alpha^*>0$.

(iii) First note that with a simple integration we obtain

$$\overline{F}_w(x)=\frac{w(x)+A(x)}{\mu_w}\overline{F}(x). \tag{11}$$

An application of L'Hospital rule yields that

$$\lim_{x\to x_F}\frac{A(x)}{w(x)}=\lim_{x\to x_F}\frac{1}{1-[\log(w(x))]'\lambda^{-1}(x)}-1. \tag{12}$$

On the other hand, it is easy to verify the relation

$$\lambda(x)=\frac{1+m'(x)}{m(x)}$$

(see also Gupta and Keating [5]) which implies that

$$\lambda(x)m(x)=1+m'(x). \tag{13}$$

For a cdf $F$ which belongs to the domain of attraction of $D(G_3)$ holds that $\lim_{x\to x_F}m'(x)=0$ (see Embrechts et al. [1]). Thus from (13) we have

$$m(x)\sim\lambda^{-1}(x)\text{ as } x\to x_F. \tag{14}$$

https://doi.org/10.1017/S0269964814000345 Published online by Cambridge University Press

From assumption (6) and relations (12) and (14) we have that

$$\lim_{x\to x_F}\frac{A(x)}{w(x)}\to 0 \; as \; x\to x_F. \tag{15}$$

From (11) and Theorem 1(iii) we have that

$$\overline{F}_w(x)=\frac{w(x)+A(x)}{\mu_w}c(x)\exp\left\{-\int_z^x\frac{g(t)}{a(t)}dt\right\}\quad z<x<x_F$$

from which we obtain

$$\overline{F}_w(x)=\frac{1+\frac{A(x)}{w(x)}}{\mu_w w(z)}c(x)\exp\left\{-\int_z^x\left(\frac{g(t)}{m(t)}-[\log(w(x))]'\right)dt\right\}\quad z<x<x_F.$$

The last equation is equivalent to

$$\overline{F}_w(x)=\frac{1+\frac{A(x)}{w(x)}}{w(z)\mu_w}c(x)\exp\left\{-\int_z^x\frac{1}{m(t)}[g(t)-[\log(w(x))]'m(x)]dt\right\},\quad z<x<x_F. \tag{16}$$

So Eq. (16) is of the form

$$\overline{F}_w(x)=\tilde{c}(x)\exp\left\{-\int_z^x\frac{\tilde{g}(t)}{m(t)}dt\right\},\quad z<x<x_F, \tag{17}$$

where

$$\tilde{c}(x)=\frac{1+\frac{A(x)}{w(x)}}{w(z)\mu_w}\;\text{ and }\;\tilde{g}(t)=g(t)-[\log(w(x))]'m(x). \tag{18}$$

From (6) and (15) all the assumptions of Theorem 1(iii) are met and $F_w\in D(G_3)$. For the inverse direction, that is if the $F_w\in D(G_3)$ then $F\in D(G_3)$ it is enough to consider the $w^*(x)=1/w(x)$ weight. Then if relation (6) holds for $w(x)$ it holds also for $w^*(x)$.

■

*Remark 1*: Many weight functions $w(x)$ which appear in literature can be studied with the help of conditions (4)–(6). Some of them as listed in Liang [8] are

$$w_1(x)=x^r,\qquad w_2(x)=x(x-1)(x-2)\cdots(x-r),$$
$$w_3(x)=c+bx,\qquad w_4(x)=\exp(c+bx).$$

With the help of (4)–(6) we can verify that the $D(G_i)$ $i=1,2,3$ are closed for the weights $w_j$, $j=1,2,3$ (provided that $\alpha>r$ for $D(G_1)$ and $\alpha>-r$ for $D(G_2)$). Concerning the $w_4(x)$ the $D(G_1(\alpha))$ and $D(G_3)$ are not necessarily closed, while the $D(G_2(\alpha))$ it does since $w_4(x_F)<\infty$. An interesting example of this case is the Pareto type I distribution with p.d.f.

$$f(x)=\frac{\theta}{x^{\theta+1}}\text{ for }x\geq 1.$$

The Pareto distribution is a typical example of heavy tail distribution and belongs to the attractor of the Fréchet distribution. With the use of $w(x)=e^{-x}$ ($w(x)=w_4(x)$ for $c=0$

https://doi.org/10.1017/S0269964814000345 Published online by Cambridge University Press

and $b = -1$) as a weight function we obtain

$$f_w(x) = \frac{e^{-x}}{E(w(X))} \frac{\theta}{x^{\theta+1}}, \text{ for } x \geq 1$$

with $E(w(X)) = \theta\Gamma(-\theta, 1)$, where $\Gamma(\cdot, \cdot)$ is the incomplete gamma function. But the $f_w$ also can be viewed as the weighted version of the truncated at one exponential (1) distribution with weight

$$w^*(x) = \frac{1}{x^{\theta+1}} I_{[1,\,\infty)}(x).$$

The exponential distribution, and so the truncated at one exponential (1) distribution, has constant mean excess function and belongs to the attractor of the Gumbel distribution. Since condition (6) is satisfied for the weight $w^*$, we conclude that $f_w$ belongs to the attractor of the Gumbel distribution too.

*Remark 2*: Special attention must be given in the choice of $r$ for a given weighted function for the cases (i) and (ii) of Theorem 2 above since it may correspond to a totally different value. For example, the weight function $w(x) = x^2$ corresponds to $r = 2$ for the first case and to $r = 0$ for the second case since $w(x)/w(x_F) \to 1$ as $x \to x_F$ and thus $w(x) \sim (x_F - x)^0$ for $d = w(x_F)$.

*Remark 3*: Theorem 2 provides conditions related to the weight $w(x)$ under which the distribution functions $F$ and $F_w$ share the same attractor. An example for which the $F$ and $F_w$ share the same attractor but the corresponding condition (6) is not satisfied, is the gamma distribution ($D(G_3)$) with shape parameter $m$ and scale parameter $\lambda$. In this case, the mean excess function is $m(x) = (1/\lambda) + ((m-1)/x\lambda^2)(1 + o(1))$. For weight function $w(x) = e^{-x}$ the $f_w$ is again gamma distribution with shape parameter $m + 1$ and scale parameter $\lambda$ and so shares the same attractor with its parent distribution (more specifically $F$ and $F_w \in D(G_3)$), but the condition (6) is not satisfied since $\lim_{x \to x_F} [\log(w(x))]' m(x) = -(1/\lambda) \neq 0$. We must also mention that for $w(x) = x$ and $w(x) = x_F - x$ for $x_F = \infty$ and $x_F < \infty$, respectively, condition (6) results to the relations $\lim_{x\to\infty} x^{-1} m(x) = 0$ and $\lim_{x \to x_F} (x_F - x)^{-1} m(x) = 0$ as they appear in Embrechts et al. [1].

*Remark 4*: On noticing that $\min\{X_1, \ldots, X_n\} = -\max\{-X_1, \ldots, -X_n\}$ we can obtain similar results for the $X_{(1)}$ rv.

### 2.2. Truncated Distributions

As already mentioned, any truncated distribution can be viewed as a weighted distribution with $w(x) = I(X \in A)$ for some $A \in S_X$. Truncated distributions play an important role in modeling physical phenomena. A characteristic example is daily rainfall, where the accumulated daily precipitation can be modeled as a mixture of a degenerate distribution on zero and a continuous truncated at zero distribution. In such cases, extreme events are of interest and the study of the distribution of the maximum (or the minimum) value is crucial. Hernández, Guenni, and Sanso [7] provided sufficient conditions for weak convergence of the distribution of maxima which are used to determine the domain of attraction of the truncated normal and the truncated $t$-student distribution in the context of (left) truncated and power transformed models. The results of Hernández et al. [7] concerning not only the left normal and the Student's $t$-distribution but also any left truncated distribution can be obtained by the following corollary.

https://doi.org/10.1017/S0269964814000345 Published online by Cambridge University Press

COROLLARY 1: *Any left truncated distribution belongs to the same domain of attraction with its no truncated parent distribution.*

PROOF: The proof is easily obtained by noticing that a left truncated distribution (say at $x_0$) can be viewed as a weighted distribution with $w(x) = I(X \in (x_0, x_F))$, which satisfies all the necessary conditions of Theorem 2 related to the weight $w(x)$ under which the distribution functions $F$ and $F_w$ share the same attractor $D(G_i)$. In this paper, we consider only the case of left truncated distribution and not the case of right truncated. This is because we focus mainly on the extreme maximum value distribution, whose form depends on the behavior of the right tail of the distribution function of the sample. By right truncating a distribution we actually restrict the domain of the rv and this may result to a change of the domain of attraction for a distribution. For example if $F \in D(G_1)$ then, the right truncated distribution does not belong to $D(G_1)$ and as a consequence in this case we cannot draw any conclusion for the true domain attraction for a distribution based on a sample from the right truncated distribution. ■

### 2.3. The Distribution of Exceedances

When the distribution $F$ is unbounded then there is an interesting connection between the limit distribution of $X_{(n)}$ and the distribution of the excess life $X - t$ conditional on $\{X > t\}$. More specifically Pickands [9] proved that for sufficiently large $t$

$$F \in D(G_\xi) \text{ iff } P\left(X - t > x | X > t\right) = \begin{cases} \left(1 + \xi \frac{x}{\sigma}\right)^{-\frac{1}{\xi}} & \xi \neq 0, \\ e^{-\frac{x}{\sigma}} & \xi = 0, \end{cases} \tag{19}$$

where $\sigma > 0$ (depending on threshold $t$, and denoted sometimes as $\sigma_t$ for that reason), and $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -(\sigma/\xi)$ when $\xi < 0$. The distribution appearing in relationship (19) is the Generalized Pareto distribution. For $\xi > 0$, it is the ordinary Pareto distribution, while for $\xi = 0$ collapses to the exponential distribution. For $\xi < 1$, the Generalized Pareto Distribution has a finite mean and the mean excess function of a unit that have exceeded a threshold $u > t$ is given by

$$m(u; \xi) = \frac{\sigma + \xi u}{1 - \xi} = \frac{\sigma}{1 - \xi} + \frac{\xi}{1 - \xi} u, \quad \sigma + \xi u > 0. \tag{20}$$

Thus we conclude that $F \in D(G_1(\alpha))$ implies that

$$m(u; \alpha) = C_\alpha + \frac{1}{\alpha - 1} u, \quad \text{provided } \alpha > 1,$$

while $F \in D(G_3)$ implies that

$$m_0(t) = C,$$

where $C_\alpha$ and $C$ are constants independent of the threshold $u$. The following theorem provides the relationship of the mean residual life of an rv $X$ with distribution function $F$ and its corresponding weighted distribution $F_w$.

https://doi.org/10.1017/S0269964814000345 Published online by Cambridge University Press

THEOREM 3: *Under the assumptions of Theorem 2 the following statements hold.*

(i) *If at least one of* $F$ *and* $F_w \in D(G_1)$ *and* $\alpha - r > 1$, *then the rate of change with respect to the threshold* $u$ *of the mean residual life of* $F$ *and* $F_w$ *depends only on the* $\alpha$ *and* $r$ *and are given by*

$$m'(u;\alpha) = \frac{1}{\alpha - 1} \quad \text{and} \quad m'(u;\alpha) = \frac{1}{\alpha - r - 1},$$

*respectively.*

(ii) *If at least one of* $F$ *and* $F_w \in D(G_3)$ *and* $\alpha = 0$, *then* $F$ *and* $F_w$ *have constant but different mean residual life.*

PROOF: The proof can be easily obtained by the closure property of Theorem 2 and the Pickands' theorem. ■

## 2.4. Biased Sampling Corrected (BSC) Extreme Value Index (EVI)

It is well known that the shape parameter $\xi$ of the GEV distribution, which is also referred as the EVI, determines the domain of attraction of a distribution or equivalently its tail behavior. More specifically, a positive $\xi$ ($\xi = 1/\alpha$) indicates an underline heavy tail distribution which implies more severe and more frequently observed extreme values, whereas $\xi < 0$ ($\xi = -1/\alpha$) suggests a light tail distribution with a finite right endpoint $x_F$ and as a result rare (or none) observed extreme values. For $\xi = 0$, the underline distribution belongs to the domain of attraction of the Gumbel distribution and its tail distribution decays exponentially. Theorem 2 suggests the need of a BSC EVI for the parent distribution when the inference is made using the extremes of biased samples (with weight function $w(x)$). More specifically, if $\xi_w$ is the EVI of the weighted version of a distribution then a BSC EVI for the parent distribution could be defined as

$$\xi_p = \frac{\xi_w}{1 + \xi_w r}, \tag{21}$$

provided that $1 + \xi_w r > 0$, where $r$ is determined so that $w(x) \sim x^r$ as $x \to \infty$, when $x_F = \infty$ and $w(x) \sim (x_F - x)^r$ as $x \to x_F$ when $x_F < \infty$.

*2.4.1. The Bias and the MSE of the BSC EVI.* In most of the applications, the $\xi_p$ is unknown and needs to be estimated. In this section, the bias and the MSE of the MLE of the BSC EVI based on a biased sample as defined in relationship (21) are studied. From the invariant property of the MLE and (21), the MLE of the $\xi_p$ is given by

$$\hat{\xi}_p = \frac{\hat{\xi}_w}{1 + \hat{\xi}_w r},$$

where $\hat{\xi}_w$ is the MLE of the EVI based on a biased sample.

By expanding $\hat{\xi}_p$ in Taylor series around the true value $\xi_w$ we obtain

$$\text{Bias}(\hat{\xi}_p) = \frac{1}{(1 + \xi_w r)^2}\text{Bias}(\hat{\xi}_w) - \frac{r}{(1 + \xi_w r)^3}E(\hat{\xi}_w - \xi_w)^2 + R_1(r,n) \tag{22}$$

https://doi.org/10.1017/S0269964814000345 Published online by Cambridge University Press

and

$$\mathrm{MSE}(\hat{\xi}_p) = \frac{1}{(1+\xi_w r)^4}\mathrm{MSE}(\hat{\xi}_w) - \frac{2r}{(1+\xi_w r)^5}E(\hat{\xi}_w - \xi_w)^3 + R_2(r,n) \tag{23}$$

where $E(\hat{\xi}_w - \xi_w)^2 = O(n^{-1})$, $R_1(r,n) = o(n^{-1})$, $E(\hat{\xi}_w - \xi_w)^3 = O(n^{-3/2})$, and $R_2(r,n) = o(n^{-3/2})$ (Gusev [6]). From the above expansions, for large $n$, we obtain the following approximations:

$$\mathrm{Bias}(\hat{\xi}_p) \approx \frac{1}{(1+\xi_w r)^2}\mathrm{Bias}(\hat{\xi}_w) \tag{24}$$

and

$$\mathrm{MSE}(\hat{\xi}_p) \approx \frac{1}{(1+\xi_w r)^4}\mathrm{MSE}(\hat{\xi}_w). \tag{25}$$

for the bias and the MSE of $\hat{\xi}_p$, respectively.

If $\xi_w < 0$ (light tail-$D(G_2)$) the coefficients of $\mathrm{Bias}(\hat{\xi}_w)$ and $\mathrm{MSE}(\hat{\xi}_w)$ at (24) and (25), respectively, are increasing with respect to $r$. This implies that the bias and the MSE of $\hat{\xi}_p$ will increase as $r$ increases. On the contrary, if $\xi_w > 0$ (heavy tail – $D(G_1)$) the coefficients of $\mathrm{Bias}(\hat{\xi}_w)$ and $\mathrm{MSE}(\hat{\xi}_w)$ are decreasing for $r$ and so the bias and the MSE of $\hat{\xi}_p$ is expected to decrease as $r$ increases.

For the case $\xi_w = 0$, we need one more terms at the Taylor expansions. In this case, relation (24) can be written as

$$\mathrm{Bias}(\hat{\xi}_p) \approx \mathrm{Bias}(\hat{\xi}_w) - rE(\hat{\xi}_w)^2.$$

From the last expression we conclude that as $r$ increases the $\mathrm{Bias}(\hat{\xi}_p)$ decreases. For the MSE we have

$$\mathrm{MSE}(\hat{\xi}_p) \approx \mathrm{MSE}(\hat{\xi}_w) - 2rE(\hat{\xi}_w)^3.$$

We were not able to prove that $E(\hat{\xi}_w)^3 > 0$. However, we performed numerous simulations that support this claim. Under this assumption, the MSE of $\hat{\xi}_p$ is a decreasing function for $r$.

## 3. SIMULATIONS

In this section, the MLE estimator of the EVI $\xi_p$ of the parent distribution is obtained with the help of simulated random samples and it is compared with the MLE obtained with the use of simulated biased samples. The comparison is in terms of the bias and the MSE and covers all the three attractors. For the attractor $D(G_1)$ the Burr distribution

$$F(x) = 1 - \left(1 + x^{-\frac{\rho}{\gamma}}\right)^{1/\rho}, \quad x > 0, \quad \gamma > 0, \quad \rho < 0$$

is selected. It holds that EVI $\xi = \gamma$. For the simulation we take $\gamma = 0.25$, $\rho = -1$. For the attractor $D(G_2)$ the beta distribution with p.d.f.

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, \quad 0 < x < 1, \quad \alpha, \beta > 0$$

is selected. The EVI equals to $\xi = -(1/\beta)$ and for the simulation the values $\alpha = 2$ and $\beta = 4$ are used. The attractor $D(G_3)$ is much wider compared to the other two attractors

https://doi.org/10.1017/S0269964814000345 Published online by Cambridge University Press

and therefore two distributions are selected. The first is the log-normal distribution

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\log x-\mu)^2}{2\sigma^2}}, \quad 0 < x$$

with EVI $\xi = 0$. For the simulation, the values $\mu = 0$ and $\sigma = 1$ are used. While all the moments of the log-normal distribution exist, the moment generating function is defined only on the negative half-axis. The second distribution is the Weibull distribution with p.d.f.

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-(x/\beta)^{\alpha}} \quad x > 0, \quad \beta > 0, \quad \alpha > 0.$$

It holds again that EVI $\xi = 0$ and for the simulation we use $\beta = 1$ and $\alpha = 0.5$.

Additionally, for comparison purposes, four different weight functions were used, namely $w(x) = 1$ (which produces random samples), $w(x) = x^r, r = 1, 2$ (which produce size-biased samples) and $w(x) = I(X > x_0)$ (which produces samples from the truncated distribution). The $x_0$ is selected to be the 10% percentile, namely the point such that $P(X < x_0) = 0.1$.

For each combination of distribution and weight function the following procedure was followed 10,000 times. First, $n = 50, 100$ random samples $Y_{i1}, \ldots, Y_{ik}$ $(k = 10, 30, 100, 250, 500$ – the block size) was generated by the $f_w$ distribution (note that these samples are $w$-weighted samples from the initial distribution $f$) and the $X_i = \max(Y_{i1}, \ldots, Y_{ik}), 1 \le i \le n$ were obtained. Then the GEV distribution was fitted to the $X_1, X_2, \ldots, X_n$ sample and the MLE $\hat{\xi}_w$ of the shape parameter $\xi_w$ is obtained.

For each case the mean and the MSE of $\hat{\xi}_p$ were calculated in order to examine the influence of the different weights at the estimation of $\xi$. The results are presented in Table 1.

Comparing the columns $w(x) = 1$ and $w(x) = I(X > x_0)$ of Table 1 we conclude that for all cases the use of a truncated sample does not have serious effects at the estimator of $\xi$. In order to examine the influence of the $w(x) = x$ and $w(x) = x^2$ weights at the estimation of $\xi$ we compare the bias and the MSE of the estimators of $\xi$ for each case $(k, n)$ as $r$ increases. More specifically, by comparing the results for the $w(x) = 1$, $w(x) = x$ and $w(x) = x^2$ weights in the case of Burr distribution $(D(G_1))$ we can conclude that for the estimation of the EVI $\xi$ it is better to use a biased sample than a random sample. Indeed the convergence to the true value it seems to be achieved already for $r = 1$ while at the same time the MSE decreases as $r$ increases, providing more accurate estimates for the biased samples. It is worth mentioning that the above findings are in agreement with the comments in Section 2.4.1.

For the beta distribution, which belongs to $D(G_2)$, we observe that for the estimation of $\xi$ it is better to use a random sample than a biased one with weight of the form $w(x) = x^r$, $r > 0$. This can be explained by the fact that if we multiply the p.d.f. $f$ by $x^r$ then it becomes more skewed to the left and thus the probability of obtaining values close to $x_F < \infty$ increases. As a consequence, the maximum value of a sample becomes less informative and thus less accurate estimators of the EVI are obtained as it was actually expected according to the results regarding the bias and the MSE of the BSC EVI when $\xi < 0$.

However, if we repeat the simulation process for the beta distribution by using this time the weight functions $w(x) = (1 - x)^{-1}$ or $w(x) = (1 - x)^{-2}$ we result to more accurate estimators compared to those obtained when we use a random sample (Table 2). This happens because multiplying the $f$ by $w(x) = (1 - x)^{-1}$ or $w(x) = (1 - x)^{-2}$, the beta distribution becomes right skewed with increasing (with respect to $r$) weight to the right tail of the distribution.

For the log-normal and the Weibull distribution $(D(G_3))$ we observe that the use of the biased samples results to the reduction of the bias of the estimator as $r$ increases as it

https://doi.org/10.1017/S0269964814000345 Published online by Cambridge University Press

https://doi.org/10.1017/S0269964814000345 Published online by Cambridge University Press



**TABLE 1.** The mean and the MSE in parentheses of the $\hat{\xi}_p$ based on 10,000 simulated samples.

| | | $w(x)=1$ | | $w(x)=x$ | | $w(x)=x^2$ | | $w(x)=I(X>x_0)$ | |
|---|---|---|---|---|---|---|---|---|---|
| | | Sample size | | Sample size | | Sample size | | Sample size | |
| | $k$ | 50 | 100 | 50 | 100 | 50 | 100 | 50 | 100 |
| Burr | 10 | 0.2275 (0.0319) | 0.2243 (0.0175) | 0.2472 (0.0136) | 0.2465 (0.0062) | 0.2515 (0.0033) | 0.2536 (0.0022) | 0.2367 (0.0407) | 0.2469 (0.0298) |
| ($\xi=0.25$) | 30 | 0.2468 (0.0306) | 0.2454 (0.0130) | 0.2503 (0.0109) | 0.2511 (0.0050) | 0.2514 (0.0046) | 0.2525 (0.0019) | 0.2452 (0.0248) | 0.2463 (0.0117) |
| | 100 | 0.2540 (0.0317) | 0.2504 (0.0088) | 0.2499 (0.0125) | 0.2502 (0.0043) | 0.2505 (0.0034) | 0.2506 (0.0016) | 0.2495 (0.0245) | 0.2517 (0.0101) |
| | 250 | 0.2554 (0.0309) | 0.2520 (0.0089) | 0.2505 (0.0131) | 0.2498 (0.0066) | 0.2494 (0.0038) | 0.2499 (0.0014) | 0.2513 (0.0277) | 0.2498 (0.0083) |
| | 500 | 0.2488 (0.0236) | 0.2493 (0.0080) | 0.2508 (0.0130) | 0.2501 (0.0044) | 0.2499 (0.0033) | 0.2497 (0.0010) | 0.2507 (0.0236) | 0.2505 (0.0082) |
| Beta | 10 | −0.2974 (0.0156) | −0.2846 (0.0077) | −0.3284 (0.0215) | −0.3169 (0.0107) | −0.3465 (0.0224) | −0.3344 (0.0125) | −0.2922 (0.0187) | −0.2831 (0.0067) |
| ($\xi=-0.25$) | 30 | −0.2975 (0.0175) | −0.2895 (0.0061) | −0.3161 (0.0187) | −0.3043 (0.0095) | −0.3277 (0.0190) | −0.3180 (0.0103) | −0.2930 (0.0183) | −0.2835 (0.0057) |
| | 100 | −0.2921 (0.0184) | −0.2824 (0.0062) | −0.3018 (0.0193) | −0.2934 (0.0078) | −0.3130 (0.0195) | −0.3031 (0.0080) | −0.2878 (0.0180) | −0.2796 (0.0054) |
| | 250 | −0.2920 (0.0154) | −0.2831 (0.0058) | −0.2955 (0.0160) | −0.2846 (0.0076) | −0.3027 (0.0149) | −0.2922 (0.0068) | −0.2838 (0.0177) | −0.2766 (0.0053) |
| | 500 | −0.2889 (0.0181) | −0.2830 (0.0057) | −0.2880 (0.0158) | −0.2824 (0.0064) | −0.2959 (0.0153) | −0.2868 (0.0062) | −0.2829 (0.0156) | −0.2742 (0.0051) |
| Lognormal | 10 | 0.3132 (0.1238) | 0.3137 (0.1069) | 0.2342 (0.0676) | 0.2347 (0.0595) | 0.1853 (0.0400) | 0.1889 (0.0373) | 0.3132 (0.1512) | 0.3102 (0.1059) |
| ($\xi=0$) | 30 | 0.2703 (0.1000) | 0.2696 (0.0826) | 0.2119 (0.0589) | 0.2115 (0.0501) | 0.1668 (0.0346) | 0.1715 (0.0323) | 0.2719 (0.1160) | 0.2703 (0.0831) |
| | 100 | 0.2425 (0.0805) | 0.2415 (0.0666) | 0.1895 (0.0502) | 0.1923 (0.0430) | 0.1532 (0.0328) | 0.1585 (0.0283) | 0.2416 (0.0803) | 0.2423 (0.0680) |
| | 250 | 0.2238 (0.0681) | 0.2253 (0.0588) | 0.1790 (0.0449) | 0.1794 (0.0369) | 0.1449 (0.0270) | 0.1467 (0.0248) | 0.2224 (0.0704) | 0.2247 (0.0586) |
| | 500 | 0.2143 (0.0640) | 0.2137 (0.0537) | 0.1707 (0.0418) | 0.1725 (0.0345) | 0.1408 (0.0260) | 0.1452 (0.0232) | 0.2131 (0.0680) | 0.2134 (0.0531) |
| Weibull | 10 | 0.3931 (0.1802) | 0.3952 (0.1654) | 0.1477 (0.0341) | 0.1527 (0.0280) | 0.0770 (0.0162) | 0.0857 (0.0113) | 0.3805 (0.1689) | 0.3796 (0.1533) |
| ($\xi=0$) | 30 | 0.2678 (0.0911) | 0.2665 (0.0791) | 0.1108 (0.0281) | 0.1162 (0.0187) | 0.0584 (0.0140) | 0.0672 (0.0086) | 0.2565 (0.0873) | 0.2593 (0.0759) |
| | 100 | 0.1949 (0.0576) | 0.1957 (0.0460) | 0.0890 (0.0246) | 0.0934 (0.0146) | 0.0478 (0.0125) | 0.0572 (0.0064) | 0.1934 (0.0596) | 0.1923 (0.0447) |
| | 250 | 0.1619 (0.0426) | 0.1624 (0.0332) | 0.0762 (0.0192) | 0.0832 (0.0121) | 0.0410 (0.0119) | 0.0504 (0.0067) | 0.1589 (0.0526) | 0.1609 (0.0334) |
| | 500 | 0.1438 (0.0369) | 0.1440 (0.0280) | 0.0687 (0.0184) | 0.0743 (0.0116) | 0.0379 (0.0118) | 0.0452 (0.0063) | 0.1390 (0.0364) | 0.1441 (0.0277) |

**TABLE 2.** The mean and the MSE in parentheses of the $\hat{\xi}_p$ based on 10,000 simulated samples for the beta distribution with weights $w(x) = (1-x)^{-1}$ and $w(x) = (1-x)^{-2}$.

| | | $w(x) = (1-x)^{-1}$ | | $w(x) = (1-x)^{-2}$ | |
|---|---|---|---|---|---|
| | | Sample size | | Sample size | |
| | $k$ | 50 | 100 | 50 | 100 |
| Beta | 10 | −0.2741 (0.0060) | −0.2694 (0.0036) | −0.2494 (0.0030) | −0.2515 (0.0009) |
| ($\xi = -0.25$) | 30 | −0.2757 (0.0069) | −0.2688 (0.0015) | −0.2488 (0.0043) | −0.2542 (0.0009) |
| | 100 | −0.2741 (0.0062) | −0.2660 (0.0034) | −0.2491 (0.0031) | −0.2538 (0.0007) |
| | 250 | −0.2703 (0.0060) | −0.2640 (0.0025) | −0.2496 (0.0025) | −0.2531 (0.0007) |
| | 500 | −0.2683 (0.0061) | −0.2625 (0.0023) | −0.2497 (0.0024) | −0.2525 (0.0006) |

**TABLE 3.** The mean and the MSE in parentheses of the $\hat{\xi}_p$ based on 10,000 simulated samples for $n = 500$.

| | $k$ | $w(x) = 1$ | $w(x) = x$ | $w(x) = x^2$ | $w(x) = I(X > x_0)$ |
|---|---|---|---|---|---|
| Lognormal | 2000 | 0.2002 (0.0414) | 0.1656 (0.0280) | 0.1419 (0.0205) | 0.1996 (0.0412) |
| ($\xi = 0$) | 5000 | 0.1916 (0.0380) | 0.1599 (0.0262) | 0.1376 (0.0193) | 0.1902 (0.0374) |
| Weibull | 2000 | 0.1197 (0.0155) | 0.0698 (0.0058) | 0.0487 (0.0031) | 0.1184 (0.0152) |
| ($\xi = 0$) | 5000 | 0.1073 (0.0127) | 0.0642 (0.0050) | 0.0458 (0.0028) | 0.1065 (0.0125) |

was actually expected. The use of weights makes the estimates for $\xi$ to converge faster to the true value $\xi = 0$ compared to a random sample ($r = 0$). Concerning the behavior of the estimators we observe that the improvement seems to be more significant as we increase the block size $k$ and not the sample size $n$. This can explained from the fact that a large block size $k$ offers a sample of size $n$ of maximum values from the distribution $F_{(k)}$ which is close to the GEV distribution. The larger the $k$, the better the approximation of the GEV distribution $F_{(k)}$ from the $F_{(k)}$. Nevertheless the improvement is important when $k$ and $n$ increase simultaneously. It deserves noticing the large bias of the estimator for the log-normal distribution.

For a better understanding of this phenomenon we run more simulations for the case $\xi = 0$ for large $k$ and $n$. From Table 3, we can see that the performance of estimator is improving for large $k$ and $n$. But even for such large $n$ the bias is still large compared to the other distributions presented here. To our knowledge, such a study of the behavior of the MLE estimator for $\xi$ in terms of the parent distribution of the sample is not available. The various comparison studies of the MLE of EVI in terms of alternative estimating methods are based on simulated samples directly from the extreme value distribution and not from the parent distribution. In any case, it is clear that the behavior of the MLE of the EVI when the sample is drawn from the log-normal distribution deserves more study.

## 4. CONCLUSION AND COMMENTS

Theorem 2 provides sufficient conditions under which the asymptotic distribution from a random sample and from a weighted sample share the same attractor. Based on these results the relationship between the rate of change of the mean residual life of $F$ and $F_w$

https://doi.org/10.1017/S0269964814000345 Published online by Cambridge University Press

were obtained. Additionally, a readjustment of the resulting estimation of the EVI with the use of the relation (21) was introduced when a biased sample is available. The main conclusion of the simulation study is that the use of a carefully chosen weight can result to a significantly better estimation of the EVI. The choice of the weighted function depends only on the available knowledge or on the belief for the value of the right end-point $x_F$ of the distribution.

## References


1. Embrechts, P., Klüppelberg, C. & Mikosch, Th. (2001). *Modelling Extremal Events for Insurance and Finance*. New York: Springer-Verlag.
2. Finkenstädt, B. & Rootzén, H. (2004). *Extreme Values in Finance, Telecommunications, and the Environment*. Chapman & Hall/CRC.
3. Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *The Annals of Mathematics* 44(3): 423–453.
4. Gumbel, E.J. (1958). *Statistics of Extremes*. New York: Columbia University Press.
5. Gupta, R.C. & Keating, J.P. (1986). Relations for reliability measures under length biased sampling. *Scandinavian Journal of Statistics* 13(1): 49–56.
6. Gusev, S.I. (1976). Asymptotic expansions associated with some statistical estimators in the smooth case II. Expansions of moments and distributions. *Theory of Probability and its Applications* **XXI**(1): 14–33.
7. Hernández, A., Guenni, L. & Sanso, B. (2009). Extreme limit distribution of truncated models for daily rainfall. *Environmetrics* 20: 962–980.
8. Liang, S. (2005). *Generalized linear and additive models with weighted distribution*. Ph.D. thesis, National University of Singapore.
9. Pickands III, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics* 3(1): 119–131.
10. Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes*. New York: Springer-Verlag.
11. Simon, R. (1980). Length-biased sampling in etiological studies. *Am. J. Epidemiol.* 111: 444–452.
12. Zelen, M. & Feinleib, M., (1969). On the theory of screening for chronic diseases. *Biometrika* 56: 601–614.


https://doi.org/10.1017/S0269964814000345 Published online by Cambridge University Press