Journal of Data Science **11**(2013), 739-766

The Kumaraswamy GP Distribution

Saralees Nadarajah* and Sumaya Eljabri
University of Manchester

Abstract: The generalized Pareto (GP) distribution is the most popular model for extreme values. Recently, Papastathopoulos and Tawn [*Journal of Statistical Planning and Inference* **143** (2013), 131-143] have proposed some generalizations of the GP distribution for improved modeling. Here, we point that Papastathopoulos and Tawn's generalizations are in fact not new and then go on to propose a tractable generalization of the GP distribution. For the latter generalization, we provide a comprehensive treatment of mathematical properties, estimate parameters by the method of maximum likelihood and provide the observed information matrix. The proposed model is shown to give a better fit for the real data set used in Papastathopoulos and Tawn.

Key words: Beta distribution, GP distribution, Kumaraswamy distribution, maximum likelihood, order statistics.

1. Introduction

The generalized Pareto (GP) distribution is the most widely applied model for univariate extreme values. Possible applications cover most areas of science, engineering and medicine. Some published applications are: lifetime data analysis, coupon collector's problem, analysis of radio audience data, analysis of rainfall time series, comparing investment risk between Chinese and American stock markets, regional flood frequency analysis, drought modeling, value at risk, analysis of turbine steady-state, second-order material property closures, wind extremes, analysis of a Spanish motor liability insurance database, analysis of finite buffer queues, river flow modeling, measuring liquidity risk of open-end funds, modeling of extreme earthquake events, estimation of the maximum inclusion size in clean steels, and modeling of high-concentrations in short-range atmospheric dispersion.

For details on the GP distribution, its theory and further applications, we refer the readers to Leadbetter *et al.* (1987), Embrechts *et al.* (1997), Castillo *et al.* (2005), and Resnick (2008).

*Corresponding author.

However, the GP distribution has been misused in too many areas, as can be seen from the list given. It does not give adequate fits in many areas. For example, Madsen and Rosbjerg (1998) find that the GP distribution does not give a good fit to drought deficit volumes due to many small drought events. In an illustrative example of the SAS/ETS SEVERITY procedure, Joshi (2010) finds "Both plots indicate that the Exp (exponential), Pareto, and Gpd (generalized Pareto) distributions are a poor fit".

In this paper, we propose a simple generalization of the GP distribution. We provide two different motivations for this simple generalization. The first is based on the definition of the GP distribution. The GP distribution arises as the conditional distribution of exceedances of a process over a large threshold (Pickands, 1975). If $F(\cdot)$ denotes the cumulative distribution function of the process then we can write

$$1 - F(x) \approx p \left(1 + \xi \frac{x - t}{\sigma} \right)^{-1/\xi}, \tag{1}$$

for $x > t$ and some large t, where $p = 1 - F(t)$, $x > t$ if $\xi \geq 0$, $t < x \leq t - \sigma/\xi$ if $\xi < 0$, $-\infty < \xi < \infty$ is a shape parameter and $\sigma > 0$ is a scale parameter. One way to improve on (1) is to take a mixture of GP cumulative distribution functions. That is, write

$$1 - F(x) \approx \sum_{i=1}^{k} w_i \left(1 + \xi_i \frac{x - t}{\sigma_i} \right)^{-1/\xi_i}, \tag{2}$$

for $x > t$ and some large t. But mixtures of the form (2) are notoriously difficult to handle not just because of the complicated mathematical form. Inferences and fitting of (2) are also difficult. For example, on the subject of estimating a mixture of Pareto distributions, Bee *et al.* (2009) say "Application of standard techniques to a mixture of Pareto is problematic". Indeed, applications of mixtures of Pareto distributions have been very limited.

A way around is to rewrite (2) in a simple mathematical form. There are many choices for the mathematical form. A choice motivated by the works of Kumaraswamy (1980) and Cordeiro and de Castro (2011) is

$$1 - F(x) = \left\{ 1 - G(x)^a \right\}^b, \tag{3}$$

where $G(\cdot)$ denotes a GP cumulative distribution function and $a > 0$, $b > 0$ are two additional parameters whose role is partly to introduce skewness and to vary tail weights. Note that the right hand side of (3) can be expanded as

$$\left\{ 1 - G(x)^a \right\}^b = \sum_{i=0}^{\infty} c_i \left[1 - G(x) \right]^{b+i}, \tag{4}$$

a mixture taking the form of (2). The coefficients c_i are functions of a and b. For instance, $c_0 = a^b$. The parameter b mainly dictates the tail behaviors of the mixture components. The parameter a mainly dictates the mixture coefficients.

Following the terminology used in Cordeiro and de Castro (2011), we shall refer to the distribution given by (3) as the KumGP distribution. The probability density function corresponding to (3) is

$$f(x) = ab\ g(x)\ G(x)^{a-1} \left\{1 - G(x)^a\right\}^{b-1}, \tag{5}$$

where $g(x) = dG(x)/dx$ is a GP probability density function. Because $g(\cdot)$ and $G(\cdot)$ are tractable, the KumGP distribution can be used quite effectively even if the data are censored. Moreover, existing software for the GP distribution (say, to compute probability density function, cumulative distribution function, quantile function, moments, maximum likelihood estimates, random numbers, etc) can be easily adapted for the KumGP distribution. Clearly, the GP distribution is a special case of the KumGP distribution for $a = b = 1$ with a continuous crossover towards cases with different shapes (for example, a particular combination of skewness and kurtosis).

The role of the two additional parameters, $a > 0$ and $b > 0$, is to govern skewness and generate distributions with heavier/ligther tails. If $a < 1$ then the tails of $f(\cdot)$ will be heavier than those of $g(\cdot)$. Similarly, if $b < 1$ then the tails of $f(\cdot)$ will be heavier than those of $g(\cdot)$. On the other hand, if $a > 1$ then the tails of $f(\cdot)$ will be lighter than those of $g(\cdot)$. Similarly, if $b > 1$ then the tails of $f(\cdot)$ will be lighter than those of $g(\cdot)$. Further description of the role of a and b is given in Sections 2 and 3.

Another physical interpretation for the KumGP distribution when a and b are positive integers is as follows. Suppose a system is made of b independent components and that each component is made up of a independent subcomponents. Suppose the system fails if any of the b components fails and that each component fails if all of the a subcomponents fail. Let $X_{j1}, X_{j2}, \cdots, X_{ja}$ denote the lifetimes of the subcomponents within the jth component, $j = 1, 2, \cdots, b$ with a common GP cumulative distribution function. Let X_j denote the lifetime of the jth component, $j = 1, \cdots, b$, and let X denote the lifetime of the entire system. So, the cumulative distribution function of X is

$$\begin{aligned}
\Pr(X \le x) &= 1 - \Pr^b\left(X_1 > x\right) = 1 - \left\{1 - \Pr\left(X_1 \le x\right)\right\}^b \\
&= 1 - \left\{1 - \Pr\left(X_{11} \le x, X_{12} \le x, \cdots, X_{1a} \le x\right)\right\}^b \\
&= 1 - \left\{1 - \Pr^a\left(X_{11} \le x\right)\right\}^b = 1 - \left\{1 - G^a_{\xi,\sigma}(x)\right\}^b,
\end{aligned} \tag{6}$$

where $G_{\xi,\sigma}(\cdot)$ denotes the cumulative distribution function of the GP distribution. So, it follows that the KumGP distribution given by (3) and (5) is precisely the time to failure distribution of the entire system. The GP distribution has been widely used to model lifetimes: see, for example, Mahmoudi (2011). For the particular case, $\xi = 0$, (6) is the cumulative distribution function of the Kumexponential distribution. The Kumexponential distribution has been used to model lifetimes, see Cordeiro *et al.* (2010).

There are other ways to generalize the GP distribution. The most recent generalizations of the GP distribution were proposed by Papastathopoulos and Tawn (2013). They referred to their generalizations as EGP1, EGP2 and EGP3 distributions. The EGP1 distribution is specified by the cumulative distribution function

$$F(x) = \frac{1}{B\left(\kappa, 1/\mid \xi \mid\right)} B_{1-\left(1+\xi\frac{x}{\sigma}\right)^{-|\xi|/\xi}}\left(\kappa, 1/\mid \xi \mid\right), \tag{7}$$

for $x > 0$ (if $\xi \geq 0$), $0 < x \leq -\sigma/\xi$ (if $\xi < 0$), $\sigma > 0$, $\kappa > 0$ and $-\infty < \xi < \infty$, where $B_x(\cdot, \cdot)$ denotes the incomplete beta function defined by

$$B_x(a, b) = \int_0^x t^{a-1}(1-t)^{b-1} dt,$$

and $B(\cdot, \cdot)$ denotes the beta function defined by

$$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt.$$

The EGP2 distribution is specified by the cumulative distribution function

$$F(x) = \frac{1}{\Gamma(\kappa)} \gamma \left[\kappa, \frac{1}{\xi} \ln\left(1 + \xi\frac{x}{\sigma}\right)\right], \tag{8}$$

for $x > 0$ (if $\xi \geq 0$), $0 < x \leq -\sigma/\xi$ (if $\xi < 0$), $\sigma > 0$, $\kappa > 0$ and $-\infty < \xi < \infty$, where $\Gamma(\cdot)$ denotes the gamma function defined by

$$\Gamma(a) = \int_0^\infty t^{a-1} \exp(-t) dt,$$

and $\gamma(\cdot, \cdot)$ denotes the incomplete gamma function defined by

$$\gamma(a, x) = \int_0^x t^{a-1} \exp(-t) dt.$$

The EGP3 distribution is specified by the cumulative distribution function

$$F(x) = \left\{1 - \left(1 + \xi\frac{x}{\sigma}\right)^{-1/\xi}\right\}^\kappa, \tag{9}$$

for $x > 0$ (if $\xi \geq 0$), $0 < x \leq -\sigma/\xi$ (if $\xi < 0$), $\sigma > 0$, $\kappa > 0$ and $-\infty < \xi < \infty$.

Unfortunately, none of the distributions given by (7)-(9) are new. There have been many published papers (possibly in hundreds) proposing distributions same as (7)-(9) or containing (7)-(9) as special cases.

Besides, the distributions given by Papastathopoulos and Tawn (2013) appear complicated: at least (7) and (8) involve the incomplete beta function and the incomplete gamma function, special functions requiring numerical routines. We shall also see later that none of (7)-(9) provide significant improvements over the GP distribution for the data set considered in Papastathopoulos and Tawn (2013).

We now explain why the distributions given by (7)-(9) are not new. Firstly, (7) is a special case of the class of beta-G distributions introduced by Eugene *et al.* (2002) and followed by Jones (2004) and many others. The beta-G distribution is specified by the cumulative distribution function

$$F(x) = \frac{1}{B(a,b)} \int_0^{G(x)} t^{a-1}(1-t)^{b-1} dt, \tag{10}$$

for $a > 0$ and $b > 0$. Note that (7) is a special case of (10) for $G(\cdot)$ specified by

$$G(x) = 1 - \left(1 + \xi \frac{x}{\sigma}\right)^{-|\xi|/\xi}.$$

This special case is considered in detail by Akinsete *et al.* (2008, Section 2.2), Mahmoudi (2011) and many others.

Secondly, (8) is a special case of the class of gamma-G distributions introduced by Zografos and Balakrishnan (2009) and followed by Ristic and Balakrishnan (2012), Nadarajah *et al.* (2012) and many others. The gamma-G distribution is specified by the cumulative distribution function

$$F(x) = \frac{\gamma\left(a, -\log\left[1 - G(x)\right]\right)}{\Gamma(a)}, \tag{11}$$

for $a > 0$. Note that (8) is a special case of (11) for $G(\cdot)$ a GP cumulative distribution function. Furthermore, the formula for the cumulative distribution function of the EGP2 distribution given in Papastathopoulos and Tawn (2013) is not a valid cumulative distribution function!

Finally, (9) is identical to the exponentiated Pareto distribution studied by Adeyemi and Adebanji (2004), Shawky and Abu-Zinadah (2008, 2009), Afify (2010) and many others.

In this paper, we study the mathematical properties of the KumGP distribution. From now on, we write the cumulative distribution function and the probability density function of the GP distribution by

$$G_{\xi,\sigma}(x) = 1 - u, \tag{12}$$

and

$$g_{\xi,\sigma}(x) = \sigma^{-1} u^{1+\xi}, \tag{13}$$

respectively, where $u = \{1+\xi(x-t)/\sigma\}^{-1/\xi}$. The cumulative distribution function and the probability density function of the KumGP distribution can be written as

$$F(x) = 1 - \{1 - (1-u)^a\}^b, \tag{14}$$

and

$$f(x) = \sigma^{-1} abu^{1+\xi}(1-u)^{a-1}\{1-(1-u)^a\}^{b-1}, \tag{15}$$

respectively. The EGP3 distribution given by (9) is a particular case of the KumGP distribution. Unlike the EGP1 and EGP2 distributions, the KumGP distribution does not involve special functions. So, one can expect that the KumGP distribution could attract wider applicability than the EGP1, EGP2 and EGP3 distributions.

The KumGP distribution given by (15) is much more flexible than the GP distribution and can allow for greater flexibility of tails. Plots of the probability density function in (15) for some parameter values are given in Figure 1.

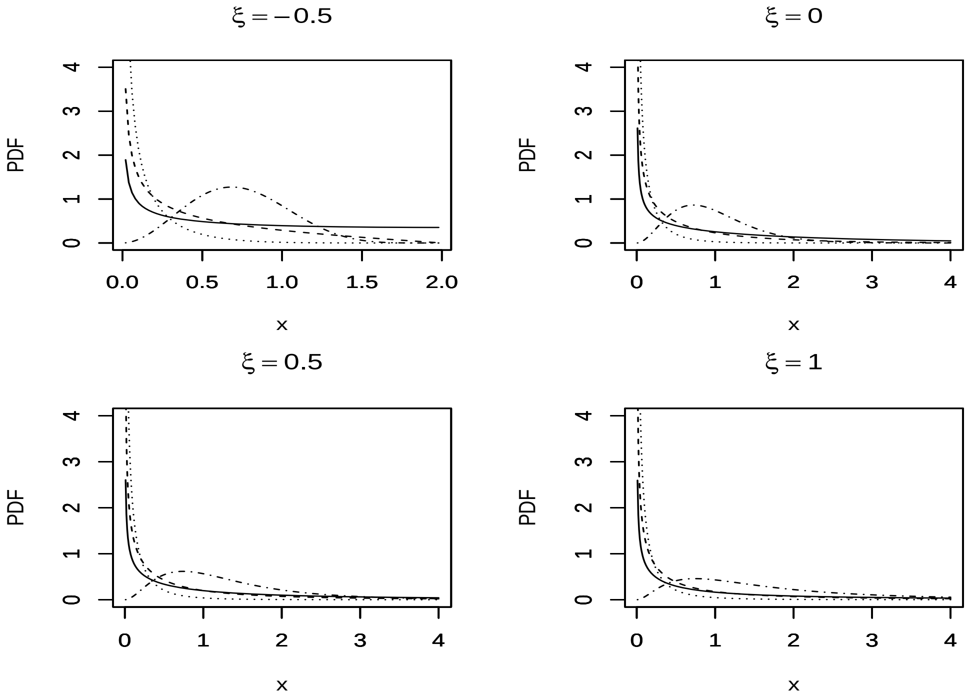

Figure 1: Plots of (15) for $u = 0$, $\sigma = 1$, $(a, b) = (0.5, 0.5)$ (solid curve), $(a, b) = (0.5, 1)$ (curve of dashes), $(a, b) = (0.5, 3)$ (curve of dots) and $(a, b) = (3, 3)$ (curve of dots and dashes)

If X is a random variable with probability density function, (15), we write $X \sim \mathrm{KumGP}(a, b, \sigma, \xi)$. The KumGP quantile function is obtained by inverting (14):

$$x = Q(z) = F^{-1}(z) = t + \frac{\sigma}{\xi}\left\{\left[1 - \left\{1 - (1-z)^{1/b}\right\}^{1/a}\right]^{-\xi} - 1\right\}. \tag{16}$$

So, one can generate KumGP variates from (16) by setting $X = Q(U)$, where U is a uniform variate on the unit interval $(0, 1)$.

In the rest of this paper, we provide a comprehensive description of the mathematical properties of (15). We examine the shape of (15) and its associated hazard rate function in Sections 2 and 3, respectively. We derive expressions for moments in Section 4. Order statistics, their moments and L moments are calculated in Section 5. Asymptotic distributions of the extreme values are provided in Section 6. Estimation by the method of maximum likelihood – including the observed information matrix – is presented in Section 7. A simulation study is presented in Section 8 to assess the performance of the maximum likelihood estimators. An application of the KumGP distribution to the real data set in Papastathopoulos and Tawn (2013) is illustrated in Section 9.

The results in Sections 4 and 5 involve infinite series representations. The terms of these infinite series are elementary, so the infinite series can be computed by truncation using any standard package, perhaps even pocket calculators.

2. Shape of Probability Density Function

The first derivative of $\log\{f(x)\}$ for the KumGP distribution is:

$$\frac{d \log f(x)}{dx} = -\frac{u^{1+\xi}}{\sigma}\left\{\frac{1+\xi}{u} - \frac{a-1}{1-u} + \frac{a(b-1)(1-u)^{a-1}}{1-(1-u)^a}\right\},$$

where $u = \{1 + \xi(x-t)/\sigma\}^{-1/\xi}$. So, the modes of $f(x)$ are the roots of the equation

$$\frac{a(b-1)(1-u)^{a-1}}{1-(1-u)^a} = \frac{a-1}{1-u} - \frac{1+\xi}{u}. \tag{17}$$

There may be more than one root to (17).

Furthermore, the asymptotes of $f(x)$ and $F(x)$ as $u \to 0, 1$ are given by

$$f(x) \sim a^b b \sigma^{-1} u^{b+\xi},$$

as $u \to 0$,

$$f(x) \sim ab\sigma^{-1}(1-u)^{a-1},$$

as $u \to 1$,

$$1 - F(x) \sim (au)^b,$$

as $u \to 0$, and

$$F(x) \sim b(1-u)^a,$$

as $u \to 1$. Note that both the upper and lower tails of $f(x)$ are polynomials with respect to u. Larger values of a correspond to heavier upper tails of f. Larger values of b correspond to lighter upper tails of f.

Plots of the shapes of (15) for $t = 0$, $\sigma = 1$ and selected values of (a, b, ξ) are given in Figure 1. Both unimodal and monotonically decreasing shapes appear possible. Unimodal shapes appear when both a and b are large. Monotonically decreasing shapes appear when either a or b is small.

3. Shape of Hazard Rate Function

The hazard rate function defined by $h(x) = f(x)/\{1 - F(x)\}$ is an important quantity characterizing life phenomena of a system. For the KumGP distribution, $h(x)$ takes the form

$$h(x) = \frac{abu^{1+\xi}(1-u)^{a-1}}{\sigma\left[1-(1-u)^a\right]}, \tag{18}$$

where $u = \{1 + \xi(x-t)/\sigma\}^{-1/\xi}$. The first derivative of $\log h(x)$ is:

$$\frac{d \log h(x)}{dx} = -\frac{u^{1+\xi}}{\sigma}\left[\frac{1+\xi}{u} - \frac{a-1}{1-u} + \frac{a(1-u)^{a-1}}{1-(1-u)^a}\right].$$

So, the modes of $h(x)$ are the roots of the equation

$$\frac{a(1-u)^{a-1}}{1-(1-u)^a} = \frac{a-1}{1-u} - \frac{1+\xi}{u}. \tag{19}$$

There may be more than one root to (19).

Furthermore, the asymptotes of $h(x)$ as $u \to 0, 1$ are given by

$$h(x) \sim b\sigma^{-1}u^{\xi},$$

as $u \to 0$ and

$$h(x) \sim ab\sigma^{-1}(1-u)^{a-1},$$

as $u \to 1$. Note that both the upper and lower tails of $h(x)$ are polynomials with respect to u. Larger values of a correspond to lighter lower tails. Larger values of b correspond to heavier lower tails and heavier upper tails of h.

Figure 2 illustrates some of the possible shapes of $h(x)$ for $t = 0$, $\sigma = 1$ and selected values of (a, b, ξ). Both monotonically increasing, monotonically decreasing and bathtub shapes appear possible. Bathtub shapes appear for negative values of ξ. Monotonically increasing shapes appear when both a and b are large. Monotonically decreasing shapes appear when either a or b is small and ξ is not negative.

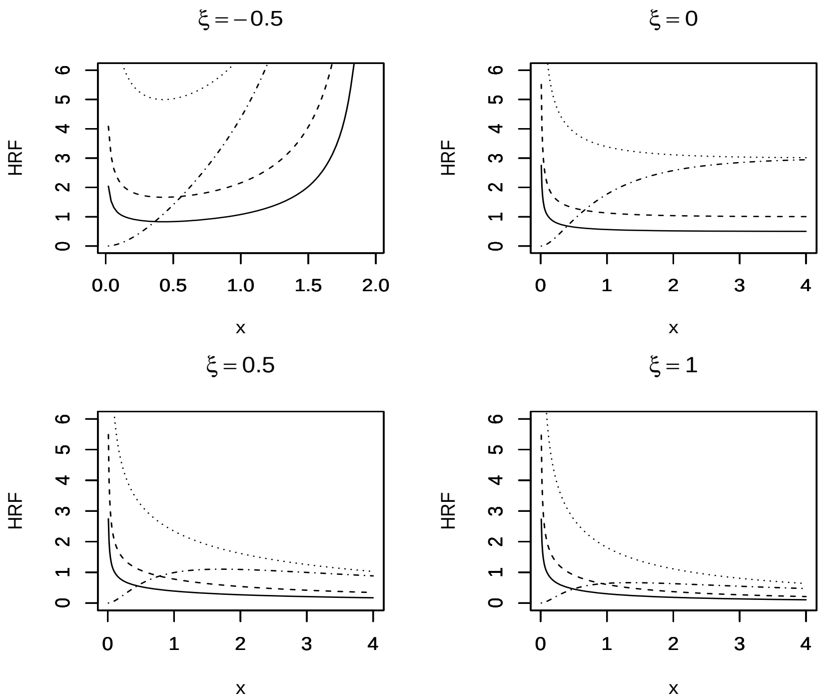

Figure 2: Plots of (18) for $u = 0$, $\sigma = 1$, $(a, b) = (0.5, 0.5)$ (solid curve), $(a, b) = (0.5, 1)$ (curve of dashes), $(a, b) = (0.5, 3)$ (curve of dots) and $(a, b) = (3, 3)$ (curve of dots and dashes)

Bathtub shaped hazard rates are the most realistic ones in practice. It is interesting to note that the KumGP distribution can exhibit this shape. The GP distribution cannot exhibit bathtub shaped hazard rates.

4. Moments

Let $X \sim \text{KumGP}(a, b, \sigma, \xi)$. Using the transformation $u = \{1+\xi(x-t)/\sigma\}^{-1/\xi}$, we can write

$$
\begin{aligned}
E\left(X^n\right) &= ab \int_0^1 \left[\frac{\sigma}{\xi}\left(u^{-\xi}-1\right)+t\right]^n (1-u)^{a-1}\left[1-(1-u)^a\right]^{b-1} du \\
&= ab \sum_{i=0}^{n} \binom{n}{i} \left(\frac{\sigma}{\xi}\right)^i \left(t-\frac{\sigma}{\xi}\right)^{n-i} \int_0^1 u^{-i\xi}(1-u)^{a-1}\left[1-(1-u)^a\right]^{b-1} du \\
&= ab \sum_{i=0}^{n} \binom{n}{i} \left(\frac{\sigma}{\xi}\right)^i \left(t-\frac{\sigma}{\xi}\right)^{n-i} \sum_{j=0}^{\infty} \binom{b-1}{j} (-1)^j \int_0^1 u^{-i\xi}(1-u)^{a+aj-1} du \\
&= ab \sum_{i=0}^{n} \binom{n}{i} \left(\frac{\sigma}{\xi}\right)^i \left(t-\frac{\sigma}{\xi}\right)^{n-i} \sum_{j=0}^{\infty} \binom{b-1}{j} (-1)^j B\left(1-i\xi, a+aj\right) \qquad (20)
\end{aligned}
$$

for $n \geq 1$ provided that $1-i\xi$ is not an integer for all $i=0,1,\cdots,n$. The first four moments are:

$$
\begin{aligned}
E(X) &= ab\Bigg[\left(t-\frac{\sigma}{\xi}\right)\sum_{j=0}^{\infty}\binom{b-1}{j}\frac{(-1)^j}{a+aj} \\
&\quad +\frac{\sigma}{\xi}\sum_{j=0}^{\infty}\binom{b-1}{j}(-1)^j B\left(1-\xi, a+aj\right)\Bigg], \qquad (21)
\end{aligned}
$$

$$
\begin{aligned}
E\left(X^2\right) &= ab\Bigg[\left(t-\frac{\sigma}{\xi}\right)^2\sum_{j=0}^{\infty}\binom{b-1}{j}\frac{(-1)^j}{a+aj} \\
&\quad +2\left(t-\frac{\sigma}{\xi}\right)\frac{\sigma}{\xi}\sum_{j=0}^{\infty}\binom{b-1}{j}(-1)^j B\left(1-\xi, a+aj\right) \\
&\quad +\left(\frac{\sigma}{\xi}\right)^2\sum_{j=0}^{\infty}\binom{b-1}{j}(-1)^j B\left(1-2\xi, a+aj\right)\Bigg], \qquad (22)
\end{aligned}
$$

$$
\begin{aligned}
E\left(X^3\right) &= ab\Bigg[\left(t-\frac{\sigma}{\xi}\right)^3\sum_{j=0}^{\infty}\binom{b-1}{j}\frac{(-1)^j}{a+aj} \\
&\quad +3\left(t-\frac{\sigma}{\xi}\right)^2\frac{\sigma}{\xi}\sum_{j=0}^{\infty}\binom{b-1}{j}(-1)^j B\left(1-\xi, a+aj\right) \\
&\quad +3\left(t-\frac{\sigma}{\xi}\right)\left(\frac{\sigma}{\xi}\right)^2\sum_{j=0}^{\infty}\binom{b-1}{j}(-1)^j B\left(1-2\xi, a+aj\right) \\
&\quad +\left(\frac{\sigma}{\xi}\right)^3\sum_{j=0}^{\infty}\binom{b-1}{j}(-1)^j B\left(1-3\xi, a+aj\right)\Bigg], \qquad (23)
\end{aligned}
$$

and

$$E\left(X^4\right) = ab\left[\left(t-\frac{\sigma}{\xi}\right)^4 \sum_{j=0}^{\infty}\binom{b-1}{j}\frac{(-1)^j}{a+aj}\right.$$
$$+4\left(t-\frac{\sigma}{\xi}\right)^3 \frac{\sigma}{\xi}\sum_{j=0}^{\infty}\binom{b-1}{j}(-1)^j B\left(1-\xi, a+aj\right)$$
$$+6\left(t-\frac{\sigma}{\xi}\right)^2 \left(\frac{\sigma}{\xi}\right)^2\sum_{j=0}^{\infty}\binom{b-1}{j}(-1)^j B\left(1-2\xi, a+aj\right)$$
$$+4\left(t-\frac{\sigma}{\xi}\right)\left(\frac{\sigma}{\xi}\right)^3\sum_{j=0}^{\infty}\binom{b-1}{j}(-1)^j B\left(1-3\xi, a+aj\right)$$
$$\left.+\left(\frac{\sigma}{\xi}\right)^4\sum_{j=0}^{\infty}\binom{b-1}{j}(-1)^j B\left(1-4\xi, a+aj\right)\right], \quad (24)$$

provided that $1-\xi$, $1-2\xi$, $1-3\xi$ and $1-4\xi$ are not integers. The infinite series in (20)-(24) all converge.

The expressions given by (21)-(24) can be used to compute the mean, variance, skewness and kurtosis of X. The values of these four quantities versus ξ are plotted in Figure 3 for $t=0$, $\sigma=1$ and selected values of (a,b). It is evident each of the quantities is an increasing function of ξ for all choices of (a,b).

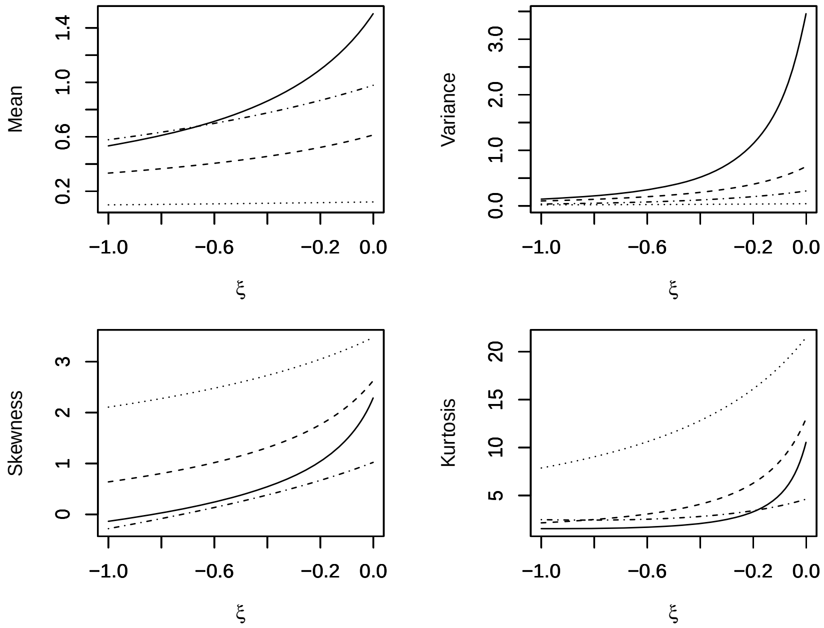

Figure 3: Mean, variance, skewness and kurtosis versus ξ for $t=0$, $\sigma=1$, $(a,b)=(0.5,0.5)$ (solid curve), $(a,b)=(0.5,1)$ (curve of dashes), $(a,b)=(0.5,3)$ (curve of dots) and $(a,b)=(3,3)$ (curve of dots and dashes)

5. Order Statistics

Order statistics make their appearance in many areas of statistical theory and practice. Let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ denote the order statistics for a random sample $X_1, X_2, \cdots, X_n$ from (15). Then the probability density function of the kth order statistic, say $Y = X_{k:n}$, can be expressed as

$$
\begin{aligned}
f_Y(y) &= \frac{abn!}{\sigma(k-1)!(n-k)!} u^{1+\xi}(1-u)^{a-1}\left[1-(1-u)^a\right]^{b(n-k+1)-1} \\
&\quad \times \left\{1-\left[1-(1-u)^a\right]^b\right\}^{b-1} \\
&= \frac{abn!}{\sigma(k-1)!(n-k)!} \\
&\quad \times \sum_{i=0}^{\infty} \binom{k-1}{i} (-1)^i u^{1+\xi}(1-u)^{a-1}\left[1-(1-u)^a\right]^{b(i+n-k+1)-1} \\
&= \frac{n!}{(k-1)!(n-k)!} \sum_{i=0}^{\infty} \binom{k-1}{i} (-1)^i f_{a,b(i+n-k+1),\sigma,\xi}(y),
\end{aligned}
$$

where $u = \{1+\xi(y-t)/\sigma\}^{-1/\xi}$ and $f_{a,b,\sigma,\xi}(\cdot)$ denotes the probability density function of $X_{a,b,\sigma,\xi} \sim \mathrm{KumGP}(a,b,\sigma,\xi)$. So, the probability density function of Y is a linear combination of probability density functions of $\mathrm{KumGP}(a,b,\sigma,\xi)$. Hence, other properties of Y can be easily derived. For instance, the cumulative distribution function of Y can be expressed as

$$
F_Y(y) = \frac{n!}{(k-1)!(n-k)!} \sum_{i=0}^{\infty} \binom{k-1}{i} (-1)^i F_{a,b(i+n-k+1),\sigma,\xi}(y),
$$

where $F_{a,b,\sigma,\xi}(\cdot)$ denotes the cumulative distribution function corresponding to $f_{a,b,\sigma,\xi}(\cdot)$. The qth moment of Y can be expressed as

$$
E\left[Y^q\right] = \frac{n!}{(k-1)!(n-k)!} \sum_{i=0}^{\infty} \binom{k-1}{i} (-1)^i E\left[X^q_{a,b(i+n-k+1),\sigma,\xi}\right], \tag{25}
$$

where $X_{a,b,\sigma,\xi} \sim \mathrm{KumGP}(a,b,\sigma,\xi)$.

L-moments are summary statistics for probability distributions and data samples (Hoskings, 1990). They are analogous to ordinary moments but are computed from linear functions of the ordered data values. The rth L moment is defined by

$$
\lambda_r = \sum_{j=0}^{r-1} (-1)^{r-1-j} \binom{r-1}{j} \binom{r-1+j}{j} \beta_j,
$$

where $\beta_j = E\{XF(X)^j\}$. In particular, $\lambda_1 = \beta_0$, $\lambda_2 = 2\beta_1 - \beta_0$, $\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0$ and $\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0$. In general, $\beta_r = (r+1)^{-1}E(X_{r+1:r+1})$, so it can be computed using (25). The L moments have several advantages over ordinary moments: for example, they apply for any distribution having finite mean; no higher-order moments need be finite.

6. Extreme Values

Suppose $X_1, \cdots, X_n$ is a random sample from (15). If $\overline{X} = (X_1 + \cdots + X_n)/n$ denotes the sample mean, then by the usual central limit theorem, $\sqrt{n}(\overline{X} - \mathrm{E}(X))/\sqrt{\mathrm{Var}(X)}$ approaches the standard normal distribution as $n \to \infty$ provided that $\xi < 1/2$. Sometimes one would be interested in the asymptotes of the extreme order statistics $M_n = \max(X_1, \cdots, X_n)$ and $m_n = \min(X_1, \cdots, X_n)$.

Firstly, suppose that G in (12) belongs to the max domain of attraction of the Gumbel extreme value distribution. Then by Leadbetter *et al.* (1987, Chapter 1), there must exist a strictly positive function, say $h(t)$, such that

$$\lim_{t \uparrow x(G)} \frac{1 - G\left(t + xh(t)\right)}{1 - G(t)} = \exp(-x),$$

for every $x \in (-\infty, \infty)$, where $x(G) = \sup\{x : G(x) < 1\}$. But, using L'Hopital's rule and since $x(F) = x(G)$, we note that

$$\begin{aligned}\lim_{t \uparrow x(F)} \frac{1 - F\left(t + xh(t)\right)}{1 - F(t)} &= \lim_{t \uparrow x(G)} \left\{\frac{1 - G^a\left(t + xh(t)\right)}{1 - G^a(t)}\right\}^b \\ &= \lim_{t \uparrow x(G)} \left\{\frac{1 - G\left(t + xh(t)\right)}{1 - G(t)}\right\}^b = \exp(-bx),\end{aligned}$$

for every $x \in (-\infty, \infty)$. So, it follows that F also belongs to the max domain of attraction of the Gumbel extreme value distribution with

$$\lim_{n \to \infty} \Pr\left\{a_n\left(M_n - b_n\right) \le x\right\} = \exp\left\{-\exp(-bx)\right\},$$

for some suitable norming constants $a_n > 0$ and b_n.

Secondly, suppose that G in (12) belongs to the max domain of attraction of the Fréchet extreme value distribution. Then by Leadbetter *et al.* (1987, Chapter 1), $x(G) = \infty$ and there must exist a $\beta < 0$ such that

$$\lim_{t \uparrow \infty} \frac{1 - G(tx)}{1 - G(t)} = x^{\beta},$$

for every $x > 0$. But, using L'Hopital's rule, we note that

$$\begin{aligned}\lim_{t \uparrow \infty} \frac{1 - F(tx)}{1 - F(t)} &= \lim_{t \uparrow \infty} \left\{\frac{1 - G^a(tx)}{1 - G^a(t)}\right\}^b = \lim_{t \uparrow \infty} \left\{\frac{1 - G(tx)}{1 - G(t)}\right\}^b \\ &= x^{b\beta},\end{aligned}$$

for every $x > 0$. Also $x(F) = x(G) = \infty$. So, it follows that F also belongs to the max domain of attraction of the Fréchet extreme value distribution with

$$\lim_{n\uparrow\infty} \Pr\left\{a_n\left(M_n - b_n\right) \leq x\right\} = \exp\left(-x^{b\beta}\right),$$

for some suitable norming constants $a_n > 0$ and b_n.

Thirdly, suppose that G in (12) belongs to the max domain of attraction of the Weibull extreme value distribution. Then by Leadbetter *et al.* (1987, Chapter 1), $x(G) < \infty$ and there must exist a $\alpha > 0$ such that

$$\lim_{t\downarrow 0} \frac{1 - G\left(x(G) - tx\right)}{1 - G\left(x(G) - t\right)} = x^{\alpha},$$

for every $x > 0$. But, using L'Hopital's rule and since $x(F) = x(G) < \infty$, we note that

$$\begin{aligned}\lim_{t\downarrow 0} \frac{1 - F\left(x(F) - tx\right)}{1 - F\left(x(F) - t\right)} &= \lim_{t\downarrow 0} \left[\frac{1 - G^a\left(x(G) - tx\right)}{1 - G^a\left(x(G) - t\right)}\right]^b \\ &= \lim_{t\downarrow 0} \left[\frac{1 - G\left(x(G) - tx\right)}{1 - G\left(x(G) - t\right)}\right]^b = x^{\alpha b}.\end{aligned}$$

So, it follows that F also belongs to the max domain of attraction of the Weibull extreme value distribution with

$$\lim_{n\to\infty} \Pr\left\{a_n\left(M_n - b_n\right) \leq x\right\} = \exp\left\{-(-x)^{\alpha b}\right\},$$

for some suitable norming constants $a_n > 0$ and b_n.

The same argument applies to min domains of attraction. That is, F belongs to the same min domain of attraction as that of G.

7. Maximum Likelihood Estimation

Suppose $x_1, x_2, \cdots, x_n$ is a random sample of size n from (15). Let $u_i = \{1 + \xi(x_i - t)/\sigma\}^{-1/\xi}$ for $i = 1, 2, \cdots, n$. Then the log-likelihood function for the vector of parameters (a, b, σ, ξ) can be written as

$$\begin{aligned}\log L(a, b, \sigma, \xi) = n\log(ab) - n\log\sigma + (1+\xi)\sum_{i=1}^{n}\log u_i + (a-1)\sum_{i=1}^{n}\log\left(1 - u_i\right) \\ +(b-1)\sum_{i=1}^{n}\log\left[1 - \left(1 - u_i\right)^a\right]. \qquad (26)\end{aligned}$$

The first-order partial derivatives of (26) with respect to the four parameters are:

$$\frac{\partial \log L}{\partial a} = \frac{n}{a} + \sum_{i=1}^{n} \log(1-u_i) - (b-1)\sum_{i=1}^{n} \frac{(1-u_i)^a \log(1-u_i)}{1-(1-u_i)^a}, \tag{27}$$

$$\frac{\partial \log L}{\partial b} = \frac{n}{b} + \sum_{i=1}^{n} \log\left[1-(1-u_i)^a\right], \tag{28}$$

$$\frac{\partial \log L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1+\xi}{\sigma^2}\sum_{i=1}^{n} u_i^{\xi}(x_i-t) - \frac{a-1}{\sigma^2}\sum_{i=1}^{n}\frac{u_i^{1+\xi}(x_i-t)}{1-u_i} + \frac{a(b-1)}{\sigma^2} \times \sum_{i=1}^{n} \frac{u_i^{1+\xi}(1-u_i)^{a-1}(x_i-t)}{1-(1-u_i)^a}, \tag{29}$$

and

$$\frac{\partial \log L}{\partial \xi} = \sum_{i=1}^{n}\log u_i + \frac{1+\xi}{\xi^2}\sum_{i=1}^{n}\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right] - \frac{\xi(x_i-t)}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\}$$
$$-\frac{a-1}{\xi^2}\sum_{i=1}^{n}\frac{u_i}{1-u_i}\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right] - \frac{\xi(x_i-t)}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\}$$
$$+\frac{a(b-1)}{\xi^2}\sum_{i=1}^{n}\frac{u_i(1-u_i)^{a-1}}{1-(1-u_i)^a}\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right] - \frac{\xi(x_i-t)}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\}. \tag{30}$$

The maximum likelihood estimates of (a, b, σ, ξ), say $(\widehat{a}, \widehat{b}, \widehat{\sigma}, \widehat{\xi})$, are the simultaneous solutions of the equations $\partial \log L/\partial a = 0$, $\partial \log L/\partial b = 0$, $\partial \log L/\partial \sigma = 0$ and $\partial \log L/\partial \xi = 0$. As $n \to \infty$, $\sqrt{n}(\widehat{a}-a, \widehat{b}-b, \widehat{\sigma}-\sigma, \widehat{\xi}-\xi)$ approaches a multivariate normal vector with zero means and variance-covariance matrix, $-(E\boldsymbol{J})^{-1}$, where

$$\boldsymbol{J} = \begin{pmatrix} \frac{\partial^2 \log L}{\partial a^2} & \frac{\partial^2 \log L}{\partial a \partial b} & \frac{\partial^2 \log L}{\partial a \partial \sigma} & \frac{\partial^2 \log L}{\partial a \partial \xi} \\ \frac{\partial^2 \log L}{\partial b \partial a} & \frac{\partial^2 \log L}{\partial b^2} & \frac{\partial^2 \log L}{\partial b \partial \sigma} & \frac{\partial^2 \log L}{\partial b \partial \xi} \\ \frac{\partial^2 \log L}{\partial \sigma \partial a} & \frac{\partial^2 \log L}{\partial \sigma \partial b} & \frac{\partial^2 \log L}{\partial \sigma^2} & \frac{\partial^2 \log L}{\partial \sigma \partial \xi} \\ \frac{\partial^2 \log L}{\partial \xi \partial a} & \frac{\partial^2 \log L}{\partial \xi \partial b} & \frac{\partial^2 \log L}{\partial \xi \partial \sigma} & \frac{\partial^2 \log L}{\partial \xi^2} \end{pmatrix}.$$

The matrix, $-E\boldsymbol{J}$, is known as the expected information matrix. The matrix, $-\boldsymbol{J}$, is known as the observed information matrix.

In simulations and real data applications described later on, we maximized the log-likelihood function using the nlm function in the R statistical package (R Development Core Team, 2012). For each maximization, the nlm function was executed for a wide range of initial values. This sometimes resulted in more than one maximum, but at least one maximum was identified each time. In cases of more than one maximum, we took the maximum likelihood estimates to correspond to the largest of the maxima.

In practice, n is finite. The literature (see, for example, Efron and Hinkley (1978)) suggests that it is best to approximate the distribution of $\sqrt{n}(\widehat{a}-a, \widehat{b}-b, \widehat{\sigma}-\sigma, \widehat{\xi}-\xi)$ by a multivariate normal distribution with zero means and variance-covariance matrix given by $-\boldsymbol{J}^{-1}$, inverse of the observed information matrix, with (a, b, σ, ξ) replaced $(\widehat{a}, \widehat{b}, \widehat{\sigma}, \widehat{\xi})$. So, it is useful to have explicit expressions for the elements of $\boldsymbol{J}$. They are given in Appendix A.

The multivariate normal approximation can be used to construct approximate confidence intervals and confidence regions for the individual parameters and for the hazard and survival functions.

8. Simulation Study

Here, we assess the performance of the maximum likelihood estimates given by (27)-(30) with respect to sample size n. The assessment is based on a simulation study:

1. generate ten thousand samples of size n from (15). The inversion method is used to generate samples, i.e variates of the KumGP distribution are generated using (16).
2. compute the maximum likelihood estimates for the ten thousand samples, say $(\widehat{a}_i, \widehat{b}_i, \widehat{\sigma}_i, \widehat{\xi}_i)$ for $i = 1, 2, \cdots, 10000$.
3. compute the biases and mean squared errors given by
$$\mathrm{bias}_h(n) = \frac{1}{10000} \sum_{i=1}^{10000} \left(\widehat{h}_i - h\right),$$
and
$$\mathrm{MSE}_h(n) = \frac{1}{10000} \sum_{i=1}^{10000} \left(\widehat{h}_i - h\right)^2,$$
for $h = a, b, \sigma, \xi$.

We repeat these steps for $n = 10, 20, \cdots, 1000$ with $a = 3$, $b = 3$, $t = 0$, $\sigma = 1$ and $\xi = 0.5$, so computing $\mathrm{bias}_a(n)$, $\mathrm{bias}_b(n)$, $\mathrm{bias}_\sigma(n)$, $\mathrm{bias}_\xi(n)$ and $\mathrm{MSE}_a(n)$, $\mathrm{MSE}_b(n)$, $\mathrm{MSE}_\sigma(n)$, $\mathrm{MSE}_\xi(n)$ for $n = 10, 20, \cdots, 1000$.

We know from theory that maximum likelihood estimates have biases of the order $O(1/n)$ and mean squared errors of the order $O(1/n)$. With this in mind, we have shown in Figures 4 and 5 how n times the four biases and n times the four mean squared errors vary with respect to n. The following observations can be made:

1. the biases for a and b appear generally positive;
2. n times the biases for a appear to level out for all n greater than 200;
3. n times the biases for b appear to level out for all n greater than 200;
4. n times the biases for σ appear to level out for all n greater than 400;
5. n times the biases for ξ appear to level out for all n greater than 200;
6. n times the mean squared errors for a appear to level out for all n greater than 200;
7. n times the mean squared errors for b appear to level out for all n greater than 200;
8. n times the mean squared errors for σ appear to level out for all n greater than 200;
9. n times the mean squared errors for ξ appear to level out for all n greater than 400.

We have presented results for only one choice for (a, b, σ, ξ), namely that (a, b, σ, ξ) $= (3, 3, 1, 0.5)$. But the results were similar for other choices.

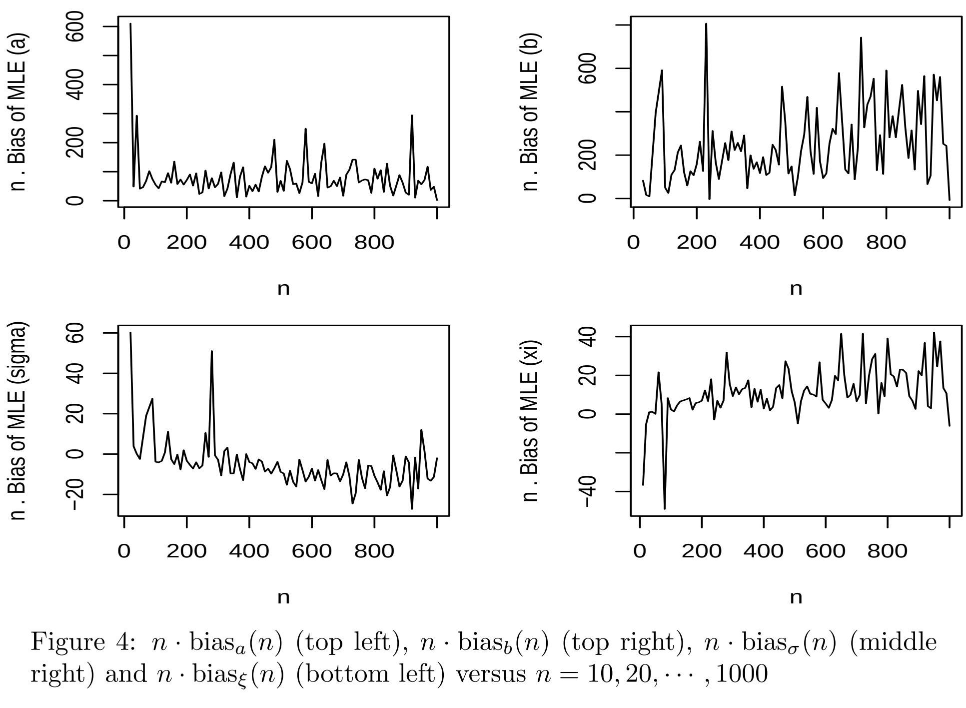

Figure 4: $n \cdot \text{bias}_a(n)$ (top left), $n \cdot \text{bias}_b(n)$ (top right), $n \cdot \text{bias}_\sigma(n)$ (middle right) and $n \cdot \text{bias}_\xi(n)$ (bottom left) versus $n = 10, 20, \cdots, 1000$

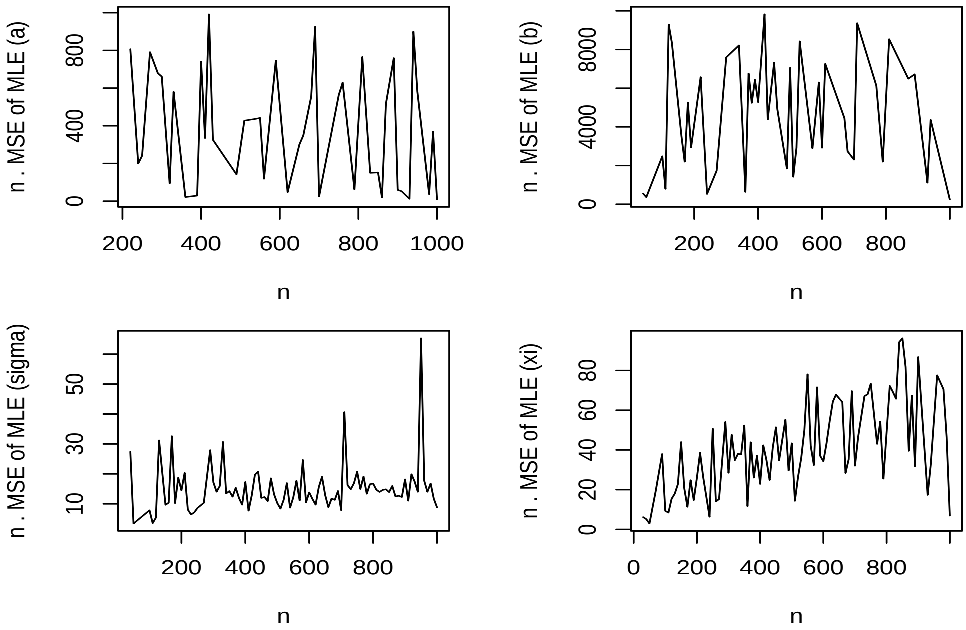

Figure 5: $n \cdot \text{MSE}_a(n)$ (top left), $n \cdot \text{MSE}_b(n)$ (top right), $n \cdot \text{MSE}_\sigma(n)$ (middle right) and $n \cdot \text{MSE}_\xi(n)$ (bottom left) versus $n = 10, 20, \cdots, 1000$

In addition to computing the biases and mean squared errors, we also computed p values to check for multivariate normality and validity of likelihood ratio tests. The p values for multivariate normality were based on the Shapiro-Wilk test (Royston, 1982). The p values for the validity of likelihood ratio tests were based on the chi-square goodness of fit test. Plots of the p values versus n showed that they remained above 0.05 for all values of n greater than 200. The plots are not shown here for reasons of space.

9. An Application

Here, we illustrate the flexibility of the KumGP distribution using a real data set analyzed in Papastathopoulos and Tawn (2013). The data set consists of one hundred and fifty four exceedances of the threshold $65m^3s^{-1}$ by the River Nidd at Hunsingore Weir from 1934 to 1969. The data is taken from NERC (1975).

A mean residual life plot is a tool used to select the threshold t for the GP distribution. The same tool can be used to select t for the KumGP distribution because of (3) and (4). The mean residual life plot of the data is shown in Figure 6. From this plot we choose $t = 65.3m^2s^{-1}$. This threshold shown in red seems appropriate.

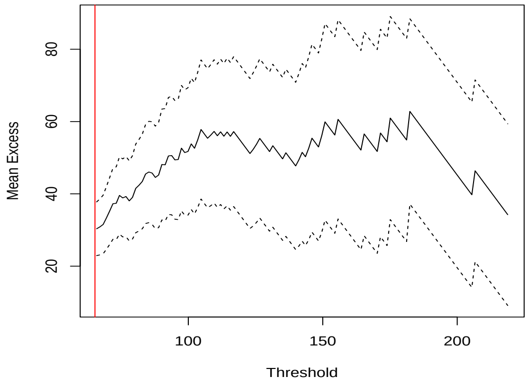

Figure 6: Mean residual life plot for exceedances of the levels of River Nidd over the threshold $65m^3s^{-1}$

We fitted the distributions (13), (7), (8), (9) and (15) to the data. The middle three distributions are those considered by Papastathopoulos and Tawn (2013). The maximum likelihood procedure described in Section 7 was used for fitting (15). The parameter estimates, log-likelihood values, AIC values and BIC values are shown in Table 1. The numbers within brackets are standard errors computed by inverting the observed information matrices.

Table 1: Parameter estimates, log-likelihood, AIC and BIC

Model	Parameter estimate (s.e)	$-\log L$	AIC	BIC
(13)	$\widehat{\sigma} = 19.034$ (2.970), $\widehat{\xi} = 0.418$ (0.134)	554.2696	1112.539	1120.647
(7)	$\widehat{\sigma} = 9.766$ (4.722), $\widehat{\xi} = 0.625$ (0.190), $\widehat{\kappa} = 1.488$ (0.423)	553.0091	1112.018	1120.599
(8)	$\widehat{\sigma} = 7.789$ (5.537), $\widehat{\xi} = 0.587$ (0.139), $\widehat{\kappa} = 1.554$ (0.508)	552.9237	1111.847	1120.38
(9)	$\widehat{\sigma} = 10.613$ (4.879), $\widehat{\xi} = 0.629$ (0.197), $\widehat{\kappa} = 1.468$ (0.409)	553.057	1112.114	1120.551
(15)	$\widehat{\sigma} = 72.876$ (20.654), $\widehat{\xi} = 0.742$ (0.264), $\widehat{a} = 3.594$ (0.651), $\widehat{b} = 1.124$ (0.451)	550.611	1109.222	1118.228

None of the three-parameter distributions (EGP1, EGP2 and EGP3) provide significant improvements over the GP distribution. Among these three distributions, the EGP2 distribution has the largest likelihood value, the smallest AIC value and the smallest BIC value. But the fit of the EGP2 distribution is not significantly better than that of the GP distribution.

The proposed four-parameter distribution provides a significant improvement over the GP distribution and the three three-parameter distributions (EGP1, EGP2 and EGP3). It has the largest likelihood value, the smallest AIC value and the smallest BIC values among all fitted distributions. Furthermore, chi-square goodness of fit tests give the p-values of 0.0373, 0.0461, 0.048, 0.041 and

0.068 for (13), (7), (8), (9) and (15), respectively, suggesting that (15) provides the only adequate fit.

The conclusion based on the likelihood values, AIC values, BIC values and the chi-square goodness of fit tests can be verified by means of probability-probability plots, quantile-quantile plots and density plots. A probability-probability plot consists of plots of the observed probabilities against probabilities predicted by the fitted model. For example, for the model given by (13), $1 - [1 + \widehat{\xi}(x_{(j)} - t)/\widehat{\sigma}]^{-1/\widehat{\xi}}$ are plotted versus $(j - 0.375)/(n + 0.25)$, $j = 1, 2, \cdots, n$, as recommended by Blom (1958) and Chambers *et al.* (1983), where $x_{(j)}$ are the sorted values of the data in ascending order and n is the number of observations. A quantile-quantile plot consists of plots of the observed quantiles against quantiles predicted by the fitted model. For example, for the model given by (13), $t+(\widehat{\sigma}/\widehat{\xi})\{(1-(j-0.375)/(n+0.25))^{-\widehat{\xi}}-1\}$ are plotted versus $x_{(j)}$, $j = 1, 2, \cdots, n$, as recommended by Blom (1958) and Chambers *et al.* (1983).

The probability-probability plots and quantile-quantile plots for the five fitted models are shown in Figures 7 and 8. We can see that the model given by (15) has points closest to the diagonal line especially in the upper tail. This is evident from the sum of the absolute differences in probabilities and quantiles shown in Table 2.

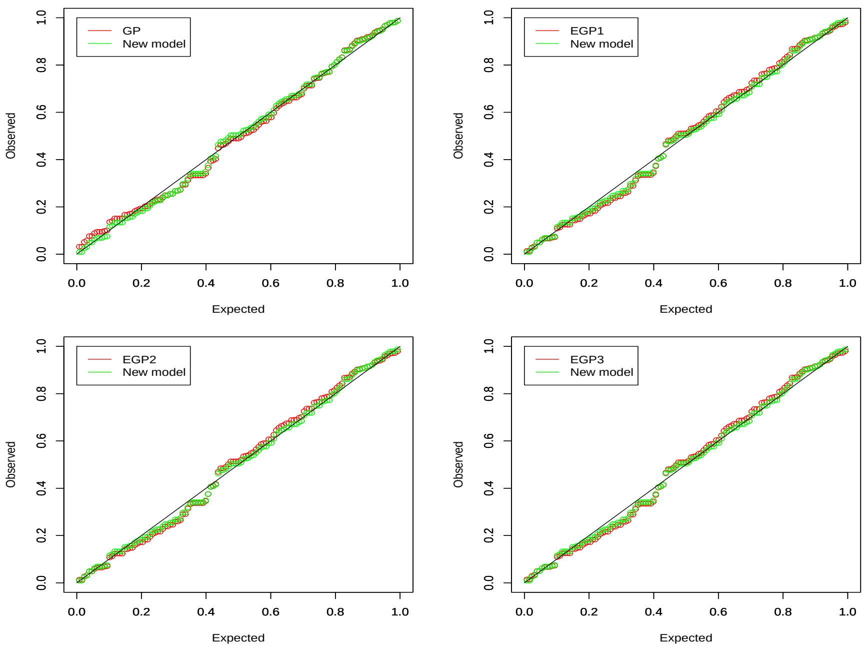

Figure 7: Probability plots for the fits of (13), (7), (8), (9) and (15) for exceedances of the levels of River Nidd over the threshold $t = 65.3m^3s^{-1}$

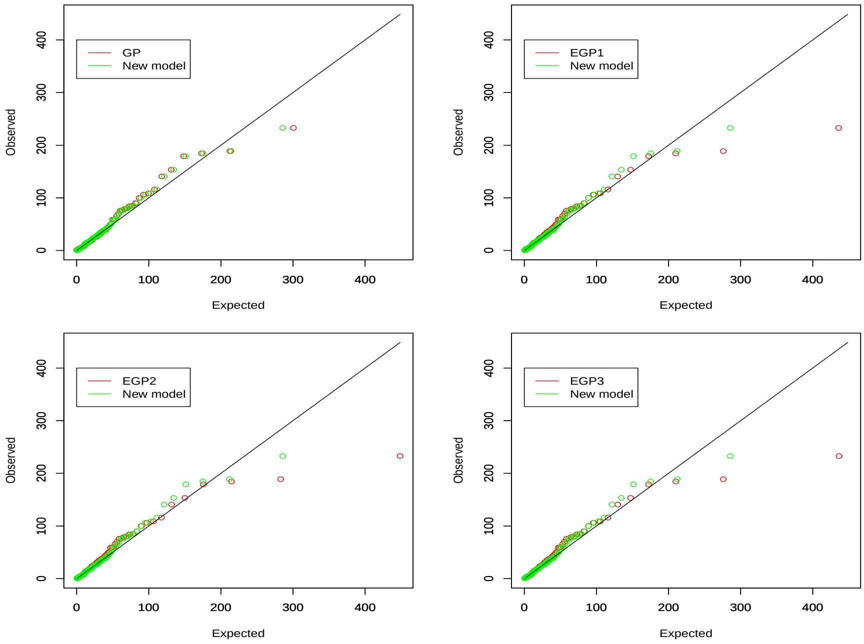

Figure 8: Quantile plots for the fits of (13), (7), (8), (9) and (15) for exceedances of the levels of River Nidd over the threshold $t = 65.3m^3s^{-1}$

Table 2: Sum of the absolute differences in probabilities and quantiles

Model	Probabilities	Quantiles
(13)	2.28	430.6654
(7)	2.64	639.2003
(8)	2.61	623.6487
(9)	2.59	623.0249
(15)	1.83	361.5818

A density plot compares the fitted probability density functions of the models with the empirical histogram of the observed data. The density plots are shown in Figure 9. Again the fitted probability density function for (15) appears to capture the general pattern of the empirical histogram best.

Quantities of interest for practitioners of extreme value models are the return levels. A T year return level, say x_T, is defined as the level that is exceeded on average every T years. For the GP model given by (13),

$$x_T = t + \frac{\sigma}{\xi}\left\{(mT)^{\xi} - 1\right\}, \tag{31}$$

where m is the average number of exceedances per year. For the KumGP model given by (15),

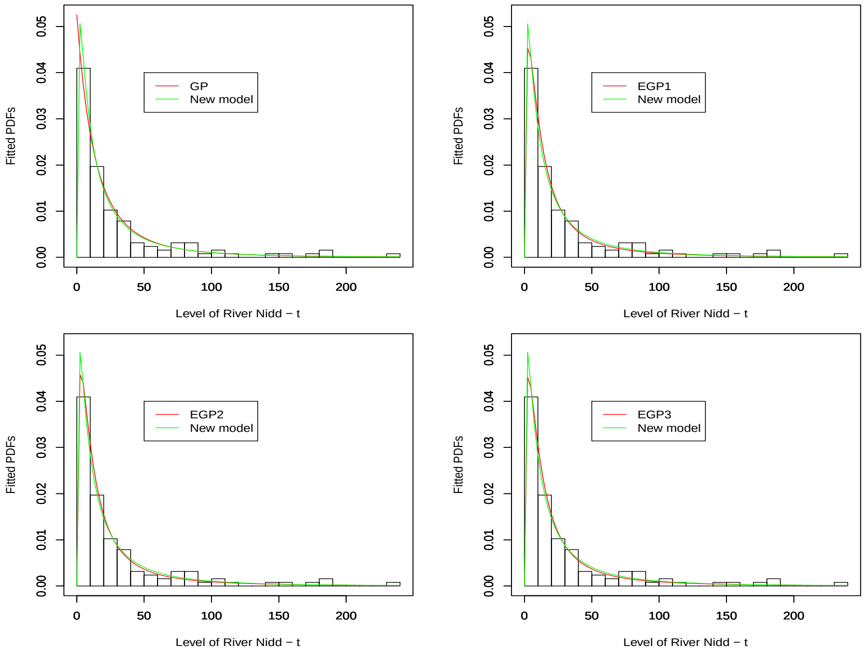

Figure 9: Fitted probability density functions of (13), (7), (8), (9) and (15) for exceedances of the levels of River Nidd over the threshold $t = 65.3 m^3 s^{-1}$

$$x_T = t + \frac{\sigma}{\xi} \left\{ \left[1 - \left\{ 1 - (mT)^{-1/b} \right\}^{1/a} \right]^{-\xi} - 1 \right\}, \tag{32}$$

where m is again the average number of exceedances per year. Plots of (31) and (32) for $T = 2, 3, \cdots, 50$ along with 95 confidence intervals computed by the delta method (Rao, 1973, pp. 387-389) are shown in Figure 10.

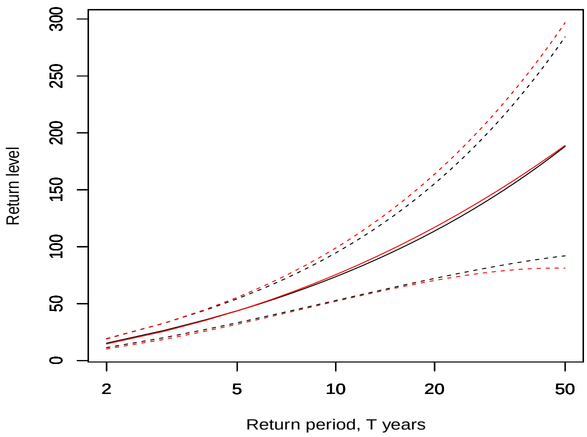

Figure 10: Return levels for exceedances of the levels of River Nidd and their 95 percent confidence intervals for the fits of (15) (in red) and (13) (in black)

Return levels are important quantities. They are used to determine, for example, dimensions of sea walls, water dams, flood defences, etc. Figure 10 suggests that the return levels given by (31) and (32) do not differ so much. The confidence bands for (32) appear only slightly wider than those for (31). One would expect the former to be wider because the KumGP model has more parameters than the GP model.

As a final remark, we like to mention that the results reported here must be treated conservatively because of the sample size. For $n = 154$, some of the biases and mean squared errors reported in Figures 4 and 5 appear large. Furthermore, asymptotic normality does not appear to have been reached. Better estimation methods (for example, bias-corrected estimation methods or bootstrapping based methods) will be needed to draw more sensible results.

Appendix A

Here, we give explicit expressions for the elements of $\boldsymbol{J}$ defined in Section 7:

$$J_{11} = -\frac{n}{a^2} + (1-b)\sum_{i=1}^{n} \frac{(1-u_i)^a \log^2(1-u_i)}{1-(1-u_i)^a} + (1-b)\sum_{i=1}^{n} \frac{(1-u_i)^{2a} \log^2(1-u_i)}{\left[1-(1-u_i)^a\right]^2},$$

$$J_{12} = -\sum_{i=1}^{n} \frac{(1-u_i)^a \log(1-u_i)}{1-(1-u_i)^a},$$

$$J_{13} = -\frac{1}{\sigma^2}\sum_{i=1}^{n} \frac{u_i^{1+\xi}(x_i-t)}{1-u_i} + \frac{b-1}{\sigma^2}\sum_{i=1}^{n} \frac{u_i^{1+\xi}(x_i-t)(1-u_i)^{a-1}\left[a\log(1-u_i)+1\right]}{1-(1-u_i)^a}$$
$$+\frac{a(b-1)}{\sigma^2}\sum_{i=1}^{n} \frac{u_i^{1+\xi}(x_i-t)(1-u_i)^{2a-1}\log(1-u_i)}{\left[1-(1-u_i)^a\right]^2},$$

$$J_{14} = -\frac{1}{\xi^2}\sum_{i=1}^{n} \frac{u_i}{1-u_i}\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right] - \frac{\xi(x_i-t)}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\}$$
$$+\frac{b-1}{\xi^2}\sum_{i=1}^{n} \frac{u_i(1-u_i)^{a-1}}{1-(1-u_i)^a}\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right] - \frac{\xi(x_i-t)}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\}$$
$$+\frac{a(b-1)}{\xi^2}\sum_{i=1}^{n} \frac{u_i(1-u_i)^{a-1}\log(1-u_i)}{1-(1-u_i)^a}\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right]\right.$$
$$\left.-\frac{\xi(x_i-t)}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\}$$
$$+\frac{a(b-1)}{\xi^2}\sum_{i=1}^{n} \frac{u_i(1-u_i)^{2a-1}\log(1-u_i)}{\left[1-(1-u_i)^a\right]^2}\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right]\right.$$

$$-\frac{\xi\left(x_i-t\right)}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\Bigg\},$$

$$J_{22}=-\frac{n}{b^2},$$

$$J_{23}=\frac{a}{\sigma^2}\sum_{i=1}^n\frac{u_i^{1+\xi}\left(1-u_i\right)^{a-1}\left(x_i-t\right)}{1-\left(1-u_i\right)^a},$$

$$J_{24}=\frac{a}{\xi^2}\sum_{i=1}^n\frac{u_i\left(1-u_i\right)^{a-1}}{1-\left(1-u_i\right)^a}\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right]-\frac{\xi\left(x_i-t\right)}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\},$$

$$\begin{aligned}J_{33}=&\frac{n}{\sigma^2}-\frac{2(1+\xi)}{\sigma^3}\sum_{i=1}^n u_i^\xi\left(x_i-t\right)+\frac{\xi(1+\xi)}{\sigma^4}\sum_{i=1}^n u_i^{2\xi}\left(x_i-t\right)^2\\&-\frac{a-1}{\sigma^4}\sum_{i=1}^n\frac{u_i^{2\xi+2}\left(x_i-t\right)^2}{\left(1-u_i\right)^2}-\frac{(a-1)(1+\xi)}{\sigma^4}\sum_{i=1}^n\frac{u_i^{2\xi+1}\left(x_i-t\right)^2}{1-u_i}\\&+\frac{2(a-1)}{\sigma^3}\sum_{i=1}^n\frac{u_i^{\xi+1}\left(x_i-t\right)}{1-u_i}-\frac{a(a-1)(b-1)}{\sigma^4}\sum_{i=1}^n\frac{u_i^{2\xi+2}\left(1-u_i\right)^{a-2}\left(x_i-t\right)^2}{1-\left(1-u_i\right)^a}\\&-\frac{a^2(b-1)}{\sigma^4}\sum_{i=1}^n\frac{u_i^{2\xi+2}\left(1-u_i\right)^{2a-2}\left(x_i-t\right)^2}{\left[1-\left(1-u_i\right)^a\right]^2}\\&+\frac{a(b-1)(1+\xi)}{\sigma^4}\sum_{i=1}^n\frac{u_i^{2\xi+1}\left(1-u_i\right)^{a-1}\left(x_i-t\right)^2}{1-\left(1-u_i\right)^a}\\&-\frac{2a(b-1)}{\sigma^3}\sum_{i=1}^n\frac{u_i^{\xi+1}\left(1-u_i\right)^{a-1}\left(x_i-t\right)}{1-\left(1-u_i\right)^a},\end{aligned}$$

$$\begin{aligned}J_{34}=&\frac{1}{\sigma^2}\sum_{i=1}^n u_i^\xi\left(x_i-t\right)-\frac{1+\xi}{\sigma^3}\sum_{i=1}^n u_i^\xi\left(x_i-t\right)^2\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\\&-\frac{a-1}{\xi^2\sigma^2}\sum_{i=1}^n\frac{u_i^{1+\xi}\left(x_i-t\right)}{1-u_i}\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right]-\frac{\xi(1+\xi)\left(x_i-t\right)}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\}\\&-\frac{a-1}{\xi^2\sigma^2}\sum_{i=1}^n\frac{u_i^{2+\xi}\left(x_i-t\right)}{\left(1-u_i\right)^2}\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right]-\frac{\xi\left(x_i-t\right)}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\}\\&+\frac{a(b-1)}{\xi^2\sigma^2}\sum_{i=1}^n\frac{u_i^{1+\xi}\left(x_i-t\right)\left(1-u_i\right)^{a-1}}{1-\left(1-u_i\right)^a}\Bigg\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right]\\&-\frac{\xi(1+\xi)\left(x_i-t\right)}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\Bigg\}-\frac{a(a-1)(b-1)}{\xi^2\sigma^2}\\&\times\sum_{i=1}^n\frac{u_i^{2+\xi}\left(x_i-t\right)\left(1-u_i\right)^{a-2}}{1-\left(1-u_i\right)^a}\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right]-\frac{\xi\left(x_i-t\right)}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\}\end{aligned}$$

$$-\frac{a^2(b-1)}{\xi^2\sigma^2}\sum_{i=1}^{n}\frac{u_i^{2+\xi}\,(x_i-t)\,(1-u_i)^{2a-2}}{[1-(1-u_i)^a]^2}$$
$$\times\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right]-\frac{\xi\,(x_i-t)}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\},$$

and

$$J_{44}=\frac{1}{\xi^2}\sum_{i=1}^{n}\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right]-\frac{\xi\,(x_i-t)}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\}$$
$$-\frac{2+\xi}{\xi^5}\sum_{i=1}^{n}\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right]-\frac{\xi\,(x_i-t)}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\}$$
$$+\frac{1+\xi}{\sigma^2\xi}\sum_{i=1}^{n}(x_i-t)^2\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-2}$$
$$+\frac{2(a-1)}{\xi^3}\sum_{i=1}^{n}\frac{u_i}{1-u_i}\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right]-\frac{\xi\,(x_i-t)}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\}$$
$$-\frac{a-1}{\xi^4}\sum_{i=1}^{n}\frac{u_i}{1-u_i}\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right]-\frac{\xi\,(x_i-t)}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\}^2$$
$$-\frac{a-1}{\xi^4}\sum_{i=1}^{n}\frac{u_i^2}{(1-u_i)^2}\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right]-\frac{\xi\,(x_i-t)}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\}^2$$
$$-\frac{a-1}{\xi\sigma^2}\sum_{i=1}^{n}\frac{u_i}{1-u_i}\,(x_i-t)^2\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-2}$$
$$+\frac{a(b-1)}{\xi^4}\sum_{i=1}^{n}\frac{u_i\,(1-u_i)^{a-2}\,(1-au_i)}{1-(1-u_i)^a}\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right]\right.$$
$$\left.-\frac{\xi\,(x_i-t)}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\}^2$$
$$-\frac{a^2(b-1)}{\xi^4}\sum_{i=1}^{n}\frac{u_i^2\,(1-u_i)^{2a-2}}{[1-(1-u_i)^a]^2}\left\{\log\left[1+\xi\frac{x_i-t}{\sigma}\right]-\frac{\xi\,(x_i-t)}{\sigma}\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-1}\right\}^2$$
$$+\frac{a(b-1)}{\xi\sigma^2}\sum_{i=1}^{n}\frac{u_i\,(1-u_i)^{a-1}}{1-(1-u_i)^a}\,(x_i-t)^2\left[1+\xi\frac{x_i-t}{\sigma}\right]^{-2}.$$

Explicit expressions for the remaining elements of $\boldsymbol{J}$ follow by symmetry.

Acknowledgements

The authors would like to thank the Editor and the referee for careful reading and for comments which greatly improved the paper.

References

Adeyemi, S. and Adebanji, T. (2004). The exponentiated generalized Pareto distribution. *Ife Journal of Science* **6**, 127-133.

Afify, W. M. (2010). On estimation of the exponentiated Pareto distribution under different sample schemes. *Statistical Methodology* **7**, 77-83.

Akinsete, A., Famoye, F. and Lee, C. (2008). The beta-Pareto distribution. *Statistics* **42**, 547-563.

Bee, M., Benedetti, R. and Espa, G. (2009). A note on maximum likelihood estimation of a Pareto mixture. No 903, Department of Economics Working Papers from Department of Economics, University of Trento, Italia.

Blom, G. (1958). *Statistical Estimates and Transformed Beta-Variables.* Wiley, New York.

Castillo, E., Hadi, A. S., Balakrishnan, N. and Sarabia, J. M. (2005). *Extreme Value and Related Models with Applications in Engineering and Science.* Wiley, Hoboken, New Jersey.

Chambers, J., Cleveland, W., Kleiner, B. and Tukey, P. (1983). *Graphical Methods for Data Analysis.* Chapman & Hall/CRC, London.

Cordeiro, G. M. and de Castro, M. (2011). A new family of generalized distributions. *Journal of Statistical Computation and Simulation* **81**, 883-898.

Cordeiro, G. M., Ortega, E. M. M. and Nadarajah, S. (2010). The Kumaraswamy Weibull distribution with application to failure data. *Journal of the Franklin Institute* **347**, 1399-1429.

Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of maximum likelihood estimation: observed versus expected information. *Biometrika* **65**, 457-482.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events: For Insurance and Finance.* Springer, Berlin.

Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics - Theory and Methods* **31**, 497-512.

Hosking, J. R. M. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B* **52**, 105-124.

Jones, M. C. (2004). Families of distributions arising from distributions of order statistics (with discussion). *Test* **13**, 1-43.

Joshi, M. V. (2010). Modeling the severity of random events with the SAS/ETS SEVERITY procedure. Paper 325-2010. SAS Institute Inc., Cary, North Carolina.

Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random-processes. *Journal of Hydrology* **46**, 79-88.

Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1987). *Extremes and Related Properties of Random Sequences and Processes.* Springer, New York.

Madsen, H. and Rosbjerg, D. (1998). A regional Bayesian method for estimation of extreme streamflow droughts. In *Statistical and Bayesian Methods in Hydrological Sciences* (Edited by E. Parent, B. Bobeé, P. Hubert and J. Miquel), 327-340. Studies and Reports in Hydrology, UNESCO, Paris.

Mahmoudi, E. (2011). The beta generalized Pareto distribution with application to lifetime data. *Mathematics and Computers in Simulation* **81**, 2414-2430.

Nadarajah, S., Cordeiro, G. M. and Ortega, E. M. M. (2012). The Zografos-Balakrishnan-G family of distributions: mathematical properties and applications. To appear in *Communications in Statistics - Theory and Methods.*

NERC (1975). *Flood Studies Report.* Natural Environment Research Council, London.

Papastathopoulos, I. and Tawn, J. A. (2013). Extended generalised Pareto models for tail estimation. *Journal of Statistical Planning and Inference* **143**, 131-143.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics* **3**, 119-131.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd edition. Wiley, New York.

Resnick, S. I. (2008). *Extreme Values, Regular Variation, and Point Processes.* Springer, New York.

Ristić, M. M. and Balakrishnan, N. (2012). The gamma exponentiated exponential distribution. *Journal of Statistical Computation and Simulation* **82**, 1191-1206.

Royston, J. P. (1982). Algorithm AS 181: the W test for normality. *Applied Statistics* **31**, 176-180.

Shawky, A. I. and Abu-Zinadah, H. H. (2008). Characterizations of the exponentiated Pareto distribution based on record values. *Applied Mathematical Sciences* **2**, 1283-1290.

Shawky, A. I. and Abu-Zinadah, H. H. (2009). Exponentiated Pareto distribution: different method of estimations. *International Journal of Contemporary Mathematical Sciences* **4**, 677-693.

Zografos, K. and Balakrishnan, N. (2009). On families of beta- and generalized gamma-generated distributions and associated inference. *Statistical Methodology* **6**, 344-362.

Received January 10, 2013; accepted April 16, 2013.

Saralees Nadarajah
School of Mathematics
University of Manchester
Manchester M13 9PL, UK
Saralees.Nadarajah@manchester.ac.uk

Sumaya Eljabri
School of Mathematics
University of Manchester
Manchester M13 9PL, UK