Univariate and multivariate Pareto models

Barry C Arnold

Correspondence:
barry.arnold@ucr.edu
Department of Statistics, University
of California, Riverside, USA

Abstract

The Pareto distribution has long been recognized as a suitable model for many non-negative socio-economic variables. Univariate and multivariate variations abound. Some unification is possible by representing the Pareto variables in terms of independent gamma distributed components. Further unification is sometimes possible since some of the frequently used multivariate Pareto models share the same copula. In some cases, inference strategies can be developed to take advantage of the stochastic representations in terms of gamma components.

Keywords: Inequality; Heavy tails; Generalized Pareto; Feller-Pareto; Kumaraswamy distribution; Hidden truncation; Conditional specification

1 Introduction

Discussion of Pareto and Pareto-like distributions can be traced back to Vilfredo Pareto's Economics textbook published in Rome in 1897. His observation that the number of persons in a population whose incomes exceed x is often well approximated by $Cx^{-\alpha}$ for some positive C and some positive α, led inexorably to consideration of the following well known model used for fitting univariate income data.

$$P(X > x) = \overline{F}_X(x) = (x/\sigma)^{-\alpha}, \quad x > \sigma. \tag{1}$$

Here σ, the scale parameter, is positive and α (Pareto's index of inequality) is also positive.

This model is typically referred to as the classical Pareto model. Improved fitting of data is encountered when more general Pareto-like distributions are considered. In this survey, the classical model will be embedded in a hierarchy of more complicated Pareto models. In this hierarchy of generalized Pareto distributions, the classical model will be called the Pareto (I) distribution. Multivariate income distributions are also of interest and, in that arena, a hierarchy of multivariate Pareto distributions is available, paralleling and closely related to the univariate hierarchy.

Even more flexible models have been proposed using these univariate and multivariate Pareto models as building blocks. Several of these will be described in this paper.

The end result is an impressively flexible array of income models from which the researcher can select a parsimonious model for the particular data set at hand. The emphasis in this survey will be on distributional properties of the models but some attention will be paid to estimation and inference strategies.

2 A hierarchy of generalized Pareto models

As the basic distribution in our hierarchy of generalized Pareto models we use the classical Pareto distribution, called here the Pareto (I) distribution. Its survival function is of the form (1). In practice, α is frequently assumed to be larger than 1, so that the distribution has a finite mean. If a random variable X has (1) as its survival function, then we write $X \sim P(I)(\sigma, \alpha)$. In this basic model, the parameter σ has a dual role. It is indeed a scale parameter, but also it determines the lower bound of the support of the distribution and, in a sense, plays to some extent the role of a location parameter. A slightly more general model, which separates the roles of location and scale parameters has then been frequently used.

This distribution will be called the Pareto (II) distribution. Its survival function is of the form

$$\overline{F}(x) = \left[1 + \left(\frac{x - \mu}{\sigma} \right) \right]^{-\alpha}, \quad x > \mu \tag{2}$$

where μ, the location parameter, is real valued, σ is positive and α is positive. In most applications μ will be non-negative, but negative values for μ pose no problems. If X has (2) as its survival function, we will write $X \sim P(II)(\mu, \sigma, \alpha)$.

There is an intimate relation between the Pareto (II) distribution and the Pickands generalized Pareto model, which is much used in the study of extreme values and peaks over thresholds. A good general reference for discussion of the role of the Pickands generalized Pareto distribution is the book by Falk et al. (2011). The density of the Pickands generalized Pareto model is

$$f(x; \sigma, k) = \frac{1}{\sigma} \left(1 - \frac{kx}{\sigma} \right)^{(1-k)/k} I(x > 0, (kx)/\sigma < 1). \tag{3}$$

where $\sigma > 0$ and $-\infty < k < \infty$. The density corresponding to $k = 0$ is obtained by taking the limit as $k \uparrow 0$ in (3). This model, (3), includes three sub-models. When $k < 0$, it yields a Pareto (II) density (with $\mu = 0$), when $k = 0$ it yields an exponential density, while for $k > 0$, it corresponds to a scaled Beta distribution (of the first or standard kind). In fact, several results for the Pareto (II) distribution can be proved to remain valid in the more general context of the Pickands generalized Pareto model. In an income modeling setting, as Pareto observed, heavy tailed distributions are typically encountered and, for this reason, we will concentrate on the Pareto (II) sub-model. We remark, in passing, that despite Pareto's insistence on the ubiquity of heavy tails, several authors have utilized scaled Beta distributions as income models, and some have even argued in favor of the exponential distribution as a model. The most general model in our hierarchy, the Feller-Pareto model will be seen to actually include some light-tailed distributions corresponding to scaled Beta models with an additional location parameter. Applications of such light tailed models are more likely to be encountered outside of the income distribution context.

Truncated versions of the Pareto (II) distribution, which clearly are not heavy tailed, are sometimes appropriate models for data sets which, for some reason, exclude large values. Such distributions are of the form

$$
\begin{aligned}
F(x) &= 0, & x \leq \mu, \\
&= \frac{1 - \left(1 + \frac{x - \mu}{\sigma} \right)^{-\alpha}}{1 - \left(1 + \frac{\tau - \mu}{\sigma} \right)^{-\alpha}}, & \mu < x < \tau, \\
&= 1, & x \geq \tau,
\end{aligned}
\tag{4}
$$

where $-\infty < \mu < \tau < \infty$, $\sigma > 0$, and $\alpha > 0$. Some discussion of such models may be found in Aban et al. (2006) and Arnold and Austin (1987).

The Pareto (III) distribution is a variant model with tail behavior comparable to that of the Pareto (II) distribution. Its survival function is of the form

$$\overline{F}(x) = \left[1 + \left(\frac{x - \mu}{\sigma} \right)^{1/\gamma} \right]^{-1}, \quad x > \mu \tag{5}$$

where μ is real, σ is positive and γ is positive. We will call γ the inequality parameter. If $\mu = 0$ and $\gamma \leq 1$, then γ turns out to be precisely the Gini index of inequality for this distribution. If X has (5) as its survival function, we will write $X \sim P(\text{III})(\mu, \sigma, \gamma)$.

If we introduce both a shape and an inequality parameter, we arrive at the Pareto (IV) family:

$$\overline{F}(x) = \left[1 + \left(\frac{x - \mu}{\sigma} \right)^{1/\gamma} \right]^{-\alpha}, \quad x > \mu \tag{6}$$

where μ (location) is real, σ (scale) is positive, γ (inequality) is positive and α (shape) is positive. Although we continue to call γ the inequality parameter it will only be identifiable with the Gini index when $\alpha = 1$ and $\mu = 0$. One might argue instead that in the P(IV) model both γ and α would be best described as shape parameters, since neither of them has a direct inequality interpretation. An anonymous referee points out that the two parameters γ and α govern the behavior of the P(IV) density as x approaches μ from above and as x approaches infinity. Thus, $f(x; \mu, \sigma, \gamma, \alpha) \sim x^{-\alpha/\gamma - 1}$ as $x \to \infty$ and $f(x - \mu; \mu, \sigma, \gamma, \alpha) \sim (x - \mu)^{1/\gamma - 1}$ as $x \to \mu$. He suggests that an argument might be advanced in favor of a reparameterization in which we define $\beta = \alpha/\gamma$, to highlight the roles of α and β in determining the limiting behavior of the density. However, in this paper, to be consistent with the notation in Arnold (1983), we will continue with the $\mu, \sigma, \gamma, \alpha$ parameterization and continue to call γ the inequality parameter. If a random variable X has (6) as its survival function, we will write $X \sim P(\text{IV})(\mu, \sigma, \gamma, \alpha)$. Note that the Pareto (IV) distribution, with $\mu = 0$, is also known as a Burr Type XII distribution.

The three more specialized families. P(I)-P(III), may be identified as special cases of the Pareto (IV) family as follows:

$$\begin{aligned} P(\text{I})(\sigma, \alpha) &= P(\text{IV})(\sigma, \sigma, 1, \alpha), \\ P(\text{II})(\mu, \sigma, \alpha) &= P(\text{IV})(\mu, \sigma, 1, \alpha), \\ P(\text{III})(\mu, \sigma, \gamma) &= P(\text{IV})(\mu, \sigma, \gamma, 1). \end{aligned} \tag{7}$$

Feller (1971), p. 49, suggested a different definition of a Pareto distribution. It can be recognized as the distribution of a ratio of two independent gamma variables (a distribution also known as Beta distribution of the second kind). By considering a linear function of a power of such a random variable, we arrive at a very general family, called the Feller–Pareto family. Thus if $X_i \sim \Gamma(\delta_i, 1)$ $i = 1, 2$, are independent random variables, and if for μ real, $\sigma > 0$ and $\gamma > 0$ we define

$$W = \mu + \sigma (X_2/X_1)^\gamma, \tag{8}$$

then W has a Feller–Pareto distribution, and we write $W \sim FP(\mu, \sigma, \gamma, \delta_1, \delta_2)$.

It may be verified that the Pareto (IV) distributions are identifiable with the Feller–Pareto distributions with $\delta_2 = 1$, i.e.,

$$P(IV)(\mu, \sigma, \gamma, \alpha) = FP(\mu, \sigma, \gamma, \alpha, 1). \tag{9}$$

The density of the general Feller–Pareto distribution defined by (8) is of the form

$$f_W(w) = \left(\frac{w-\mu}{\sigma}\right)^{(\delta_2/\gamma)-1} \left[1 + \left(\frac{w-\mu}{\sigma}\right)^{1/\gamma}\right]^{-\delta_1-\delta_2} \Bigg/ \left[\gamma \sigma B(\delta_1, \delta_2)\right],$$
$$w > \mu. \tag{10}$$

The corresponding survival function is obtainable from tables of the incomplete beta function. For many computations it is simpler to work directly with the representation (8). The Pareto (IV) distributions correspond to the case in which X_2 has a gamma distribution while X_1 has an exponential distribution. The Pareto (III) distributions are encountered when both X_1 and X_2 are exponential variables.

Kalbfleisch and Prentice (1980) call the Feller-Pareto density (with $\mu = 0$) a generalized F density. Instead we might describe a Feller Pareto variable as being a location and scale transform of a generalized beta variable of the second kind. Recall that a beta variable of the second kind is just a ratio of independent gamma variables.

An even more general model might be built using independent variables $X_i \sim \Gamma(\delta_i, 1)$, $i = 1, 2$. One could define

$$W = \mu + \sigma \left(\frac{X_2^{\gamma_2}}{X_1^{\gamma_1}}\right). \tag{11}$$

The additional flexibility provided by the introduction of such a sixth parameter in the model has not been investigated.

The full array of generalized univariate Pareto distributions to be considered in this paper are subsumed in the Feller–Pareto family and a unified derivation of many distributional results is possible. However, in the case of Pareto (I)-(IV) distributions, some alternative representations are also useful.

A random variable X has a $P(I)(\sigma, \alpha)$ distribution if it is of the form

$$X = \sigma e^{V/\alpha} \tag{12}$$

where V is a standard exponential random variable. An analogous representation of a Pareto (II) variable in terms of an exponential random variable is possible, i.e.,

$$X = \mu + \sigma \left(e^{V/\alpha} - 1\right). \tag{13}$$

Likewise Pareto (III) and (IV) variables can be represented as $X = \mu + \sigma \left(e^V - 1\right)^\gamma$ and $X = \mu + \sigma \left(e^{V/\alpha} - 1\right)^\gamma$ respectively. The representation (12) for a classical Pareto variable (i.e., Pareto (I)) highlights the useful observation that the logarithm of such a variable has a shifted exponential distribution. This will permit the recognition of many distributional properties of Pareto (I) variables as reflections of parallel properties of exponential variables.

A second important representation of the Pareto (II) distribution, known to Maguire et al. (1952), is as a mixture of exponentials. We may describe it in terms of the conditional survival function, given an auxiliary gamma distributed random variable Z. Thus, if

$$P(X > x \mid Z = z) = e^{-z(x-\mu)/\sigma}, \quad x > \mu,$$

i.e., a (translated) exponential distribution, and if $Z \sim \Gamma(\alpha, 1)$, then it follows that unconditionally $X \sim P(\text{II})(\mu, \sigma, \alpha)$. Alternatively, this can be viewed as being equivalent to the representation in (8) after setting $\gamma = \delta_2 = 1$.

This representation of the Pareto (II) distribution as a gamma mixture of exponential distributions is often encountered in reliability and survival contexts, see e.g., Keiding et al. (2002). It is also familiar in Bayesian analysis of exponential data, where the gamma density enters as a convenient prior. In this context the Pareto (II) distribution is sometimes called the Lomax distribution.

The Pareto (III) distribution was apparently first considered by (Fisk 1961a; 1961b) who called it a sech^2 distribution. It is closely related to the logistic distribution. We say that a random variable X has a logistic (μ, σ) distribution, if its distribution function assumes the form

$$F_X(x) = \left[1 + e^{-(x-\mu)/\sigma} \right]^{-1}, \quad -\infty < x < \infty$$

and we write $X \sim L(\mu, \sigma)$. It is not difficult to verify that

$$X \sim L(\mu, \sigma) \Leftrightarrow e^X \sim P(\text{III})(0, e^\mu, \sigma). \tag{14}$$

It is as a consequence of the relation (14) that the Pareto (III) distribution, with $\mu = 0$, is sometimes called the log-logistic distribution.

Remark 1. *Johnson et al. (1994) refer to a Pareto distribution of the third kind that is not to be confused with the Pareto (III) distribution discussed in this paper. The survival function of this "third kind" distribution is of the form*

$$\overline{F}(x) = \left(1 + \frac{x}{\sigma}\right)^{-\alpha} e^{-\beta x}, \ x > 0. \tag{15}$$

This distribution, which was suggested by Pareto (1897), was proposed to accommodate cases in which the basic Pareto model (1) was inadequate for fitting certain data configurations. This model is closely related to the Pareto (II) distribution, but with an additional exponential factor. Note that it could be viewed as the distribution of the minimum of a Pareto (II) variable (with $\mu = 0$) and an independent exponential variable. This model has been used infrequently, but recently it has reappeared, this time called a tapered Pareto distribution (Kagan and Schoenberg 2001).

2.1 Distributional properties

The Feller–Pareto distributions are unimodal. The mode is at μ if $\gamma > \delta_2$, while if $\gamma \leq \delta_2$, we find (here $W \sim FP(\mu, \sigma, \gamma, \delta_1, \delta_2)$)

$$\text{mode}(W) = \mu + \sigma \left[(\delta_2 - \gamma)/(\delta_1 + \gamma)\right]^\gamma \tag{16}$$

In order to compute moments of the Pareto distributions, it is convenient to work with the representation (8). With $W \sim FP(\mu, \sigma, \gamma, \delta_1, \delta_2)$, if we define $W^* = (W - \mu)/\sigma$, then $W^* \sim FP(0, 1, \gamma, \delta_1, \delta_2)$, i.e., $W^* =^d (X_2/X_1)^\gamma$ where $X_i \sim \Gamma(\delta_i, 1)$ $i = 1, 2$, are independent random variables. It then can be readily verified that for a real number τ, the τ'th moment of W^* when it exists is of the form

$$E\left(W^{*\tau}\right) = \Gamma(\delta_1 - \gamma\tau)\Gamma(\delta_2 + \gamma\tau)/\Gamma(\delta_1)\Gamma(\delta_2), \quad -(\delta_2/\gamma) < \tau < (\delta_1/\gamma). \tag{17}$$

From this expression moments for the Feller-Pareto and the P(II)-P(IV) distributions are readily obtained.

Moments of the Pareto (I) distribution cannot be obtained in this way since, for it, $\mu = \sigma \neq 0$. They are obtainable by direct integration:

$$\text{(Pareto (I))} \quad E(X^\tau) = \sigma^\tau \left(1 - \frac{\tau}{\alpha}\right)^{-1}, \quad \tau < \alpha. \tag{18}$$

Sums of independent Pareto variables typically do not have analytically tractable distributions. If we multiply independent Pareto variables rather than adding them, it is sometimes possible to get simple expressions for the density of the resulting product. In the case of the Pareto I distribution the key lies in utilization of representation (12). Thus, if $X_1, X_2, \ldots, X_n$ are independent Pareto I variables with $X_i \sim P(\text{I})(\sigma_i, \alpha_i)$, then their product W has the representation

$$W = \left(\prod_{i=1}^{n} \sigma_i\right) \exp\left(\sum_{i=1}^{n} (V_i/\alpha_i)\right) \tag{19}$$

where the V_i's are independent standard exponential variables. In some cases expressions are available for the distribution of $\sum_{i=1}^{n} V_i/\alpha_i$. In particular, if $\alpha_i = \alpha, (i = 1, 2, \ldots, n)$, then $\sum_{i=1}^{n} V_i/\alpha \sim \Gamma(n, 1/\alpha)$, and we may readily obtain the density of W.

A second case in which simple closed form expressions are available is one in which all the α_i's are distinct. In this situation we can use a result for weighted sums of exponentials given in, for example, Feller (1971), p. 40, and write the survival function of the product in the form:

$$P(W > w) = \sum_{i=1}^{n} \left(\frac{w}{\sigma}\right)^{-\alpha_i} \prod_{\substack{k=1 \\ k \neq i}}^{n} \left(\frac{\alpha_k}{\alpha_i - \alpha_k}\right), \quad w > \sigma \tag{20}$$

where $\sigma = \prod_{i=1}^{n} \sigma_i$ and $\alpha_i \neq \alpha_j$ if $i \neq j$. The distribution of products of independent Pareto (IV) variables with μ_i's equal to 0 can, via the representation (8), be reduced to a problem involving the distribution of products of powers of independent gamma random variables. Unlike the Pareto (I) case, closed form expressions for the resulting density are apparently not obtainable, although moments of such products are readily available.

The Pareto (IV) family is closed under minimization when certain parameters are common to the minimands. Thus, if X_1 and X_2 are independent random variables with $X_i \sim P(\text{IV})(\mu, \sigma, \gamma, \alpha_i), i = 1, 2$, then

$$\min(X_1, X_2) \sim P(\text{IV})(\mu, \sigma, \gamma, \alpha_1 + \alpha_2). \tag{21}$$

Note that in this situation the X_i's share common values for the parameters μ, σ and γ.

Pareto (III) variables exhibit an interesting closure property with respect to geometric minimization and maximization. Indeed, this was used as a justification for use of the Pareto (III) distribution as a suitable model for income distributions based on a scenario involving competitive bidding for employment (Arnold and Laguna 1976). For this, consider a sequence $X_1, X_2, \ldots$ of i.i.d. Pareto (III)(μ, σ, γ) random variables. Suppose that for some $p \in (0, 1)$, N_p is independent of the X_i's and has a geometric(p) distribution, i.e., $P(N = n) = p(1 - p)^{n-1}, n = 1, 2, \ldots$. Define the corresponding random extrema by

$$U_p = \min\{X_1, X_2, \ldots, X_{N_p}\}, \tag{22}$$

and

$$V_p = \max\{X_1, X_2, \ldots, X_{N_p}\}. \tag{23}$$

It is readily verified, by conditioning on N_p, that U_p and V_p each have Pareto (III) distributions. Thus

$$U_p \sim P(III)(\mu, \sigma p^\gamma, \gamma), \tag{24}$$

and

$$V_p \sim P(III)(\mu, \sigma p^{-\gamma}, \gamma). \tag{25}$$

Observe that, if $\mu = 0$, then

$$p^{-\gamma} U_p \overset{d}{=} p^\gamma V_p \overset{d}{=} X_1.$$

Some characterization results based on this observation were discussed in Arnold et al. (1986).

It is possible to write down expressions for the densities of order statistics from a Pareto (IV) sample. The corresponding distribution functions will involve incomplete beta functions. Simulation of such order statistics may be accomplished by utilizing the relatively simple form of the Pareto (IV) quantile function, i.e.,

$$F^{-1}(u) = \mu + \sigma \left[(1-u)^{-1/\alpha} - 1 \right]^\gamma . \tag{26}$$

From this we have that if $X_{i:n}$ is the ith order statistic from a sample of size n from a Pareto (IV) distribution, then

$$X_{i:n} \overset{d}{=} F^{-1}(U_{i:n}) \tag{27}$$

where $\overset{d}{=}$ means that the two random variables are identically distributed, where F^{-1} is as given in (26), and where $U_{i:n}$ is the ith order statistic of a sample of size n from a uniform (0,1) distribution. It is well known (see e.g. David and Nagaraja 2003) that $U_{i:n} \sim$ Beta$(i, n-i+1)$.

In some special cases the density of the ith order statistic (27) assumes a known form. For example:

$$X_i's \sim P(III)(\mu, \sigma, \gamma) \Rightarrow X_{i:n} \sim FP(\mu, \sigma, \gamma, n-i+1, i). \tag{28}$$

Another case involves minima:

$$X_i's \sim P(IV)(\mu, \sigma, \gamma, \alpha) \Rightarrow X_{1:n} \sim P(IV)(\mu, \sigma, \gamma, n\alpha). \tag{29}$$

3 Some related extensions

A variety of models have been proposed to add more flexibility to the generalized Pareto models discussed in Section 2. Most of them include Pareto models as special cases. In this Section we will make note of a selection of these models.

Many early researchers modeled the logarithm of income (called income power by Champernowne (1937)). Thus, instead of postulating a simple distribution for income, a relatively simple distribution was assumed for some function of income. More flexibility may be introduced by considering a parametric family of monotonic transformations of the income data whose parameters must be estimated from the data. For example, we might begin with a parametric family of increasing functions $\psi(x; \underline{\tau})$ with corresponding inverse functions $\psi^{-1}(x; \underline{\tau})$ and assume that $\psi(X; \underline{\tau})$ has a Pareto (IV)$(0, \sigma, \gamma, \alpha)$

distribution. If we denote the corresponding $P(IV)(0, \sigma, \gamma, \alpha)$ distribution by $F_{\sigma, \gamma, \alpha}(x)$ then the distribution of X will be

$$F_X(x; \sigma, \gamma, \alpha, \underline{\tau}) = F_{\sigma, \gamma, \alpha}\left(\psi^{-1}(x; \underline{\tau})\right). \tag{30}$$

A parallel extension involves quantile functions instead of distribution functions. For this the quantile function of X is assumed be of the form

$$F_X^{-1}(u; \sigma, \gamma, \alpha, \underline{\tau}) = F_{\sigma, \gamma, \alpha}^{-1}\left(\widetilde{\psi}(u; \underline{\tau})\right). \tag{31}$$

where $\widetilde{\psi}(u; \underline{\tau})$ is a parametric family of monotone functions mapping $(0, 1)$ onto $(0, 1)$. A popular model of this genre, introduced by Jones (2004), makes use of the family of quantile functions of Beta distributions. The density function of this Beta-generalized Pareto distribution is given by

$$f_X(x; \sigma, \gamma, \alpha, \lambda_1, \lambda_2) = \frac{\alpha\left[1 - \left(1 + \left(\frac{x}{\sigma}\right)^{1/\gamma}\right)^{-\alpha}\right]^{\lambda_1 - 1}\left(1 + \left(\frac{x}{\sigma}\right)^{1/\gamma}\right)^{-\alpha\lambda_2 - 1}\left(\frac{x}{\sigma}\right)^{(1/\gamma) - 1}}{\sigma\gamma B(\lambda_1, \lambda_2)},$$
$$\tag{32}$$

where $x \in (0, \infty)$. Of course, if $\lambda_1 = \lambda_2 = 1$, the Beta-generalized distribution simplifies to become a Pareto distribution.

Another popular model of the form (31) involves the simple choice $\widetilde{\psi}(u) = u^{1/\theta}$ where $\theta > 0$. In such a case we have

$$F_X(x; \sigma, \gamma, \alpha, \theta) = [F_{\sigma, \gamma, \alpha}(x)]^\theta, \quad x > 0, \tag{33}$$

and the distribution is usually called the exponentiated generalized Pareto distribution. It can be recognized as a special case of the Beta-generalized Pareto model with the parameters chosen to be $\lambda_1 = \theta$ and $\lambda_2 = 1$.

Instead of the Beta distribution, one might use the Kumaraswamy distribution to obtain an alternative generalized Pareto distribution. First, we must recall the definition of the Kumaraswamy distribution. We say that X has a Kumaraswamy (λ_1, λ_2) distribution if its density and distribution functions are :

$$f_K(x) = \lambda_1 \lambda_2 x^{\lambda_1 - 1}\left(1 - x^{\lambda_1}\right)^{\lambda_2 - 1}, \quad 0 < x < 1, \tag{34}$$

and

$$F_K(x) = 1 - \left(1 - x^{\lambda_1}\right)^{\lambda_2}, \quad 0 < x < 1. \tag{35}$$

See Jones (2009) for a comprehensive introduction to the Kumaraswamy distribution. Let $F_P(x)$ denote the Pareto (IV) distribution function and suppose that K has a Kumaraswamy (λ_1, λ_2) distribution. Define $Y = F_P^{-1}(K)$, then Y has a Kumaraswamy-Pareto (IV) distribution with corresponding density

$$f_Y(y) = f_K(F_P(y))f_P(y).$$

Akinsete et al. (2008) consider some special subcases of the Beta-generalized Pareto distribution, while Paranaiba et al. (2013) discuss the Kumaraswamy-generalized Pareto distribution. Submodels of the Kumaraswamy-generalized Pareto model are often of interest. For example the exponentiated generalized Pareto distribution (33) is such a submodel.

And, of course, one can concatenate these constructions and consider a Beta-Kumaraswamy-Pareto distribution. Going one step further we would arrive at a generalized-Beta-Kumaraswamy-Pareto model. Each generalization adds flexibility at the cost of introducing more parameters. Some degree of parsimony is evidently called for here.

Pillai (1991) suggested an extension of the Pareto (III) distribution, motivated by its closure under geometric minimization. A random variable is said to have a semi-Pareto (III) distribution if its survival function is of the form

$$\overline{F}(x; \mu, \sigma, \gamma, p) = \left[1 + \left(\frac{x - \mu}{\sigma} \right)^{1/\gamma} h \left(\frac{x - \mu}{\sigma} \right) \right]^{-1}, \quad x > \mu, \tag{36}$$

where $\mu \in (-\infty, \infty)$, $\sigma, \gamma \in (0, \infty)$, $p \in (0, 1)$ and $h(x)$ is a periodic function of $\ln x$ with period $-2\pi/[\gamma \ln p]$, and with $h(0) = 1$. The case in which $h(x) \equiv 1$ for every x corresponds to the usual Pareto (III) model. More generality can be arrived at if $h(x)$ is replaced by a suitable parametric family of periodic functions. Note that, in order for (36) to be a valid survival function, it must be the case that $x^{1/\gamma} h(x)$ is a non-decreasing function of x.

Hidden truncation or selection models may sometimes provide alternative models that are more suitable than basic Pareto models. The corresponding scenario is one in which the variable X is observed only if a covariable Y takes on a value less than some threshold value. Thus the distribution of the observed X's is of the form $P(X \leq x | Y \leq y_0)$. With this in mind, consider the case in which (X, Y) has a bivariate Pareto (IV) distribution with the following joint survival function.

$$P(X > x, Y > y) = \left[1 + \left(\frac{x - \mu}{\sigma} \right)^{1/\gamma} + \left(\frac{y - \nu}{\tau} \right)^{1/\delta} \right]^{-\alpha}, \quad x > \mu, y > \nu. \tag{37}$$

(Such distributions will be discussed in more detail in Section 5). This distribution has Pareto (IV) marginals and has Pareto (IV) conditionals. After suitable reparameterization the corresponding hidden truncation density for X, given that it can only be observed if Y is not too large, is

$$f_{HT}(x; \mu, \sigma, \gamma, \alpha, \theta) = \frac{\alpha \left(\frac{x-\mu}{\sigma} \right)^{(1/\gamma)-1}}{\gamma \sigma [1 - (1 + \theta)^{-\alpha}]} \tag{38}$$

$$\times \left[\left(1 + \left(\frac{x - \mu}{\sigma} \right)^{1/\gamma} \right)^{-(\alpha+1)} \right.$$

$$\left. - \left(1 + \left(\frac{x - \mu}{\sigma} \right)^{1/\gamma} + \theta \right)^{-(\alpha+1)} \right], \quad x > \mu,$$

where a new parameter $\theta = [(y_0 - \nu)/\tau]^{1/\delta}$ has been introduced. In this model, μ is a real valued parameter, often positive, while all of the other parameters, σ, γ, α

and θ are positive valued. An alternative representation of this density is possible as follows.

$$f_{HT}(x; \mu, \sigma, \gamma, \alpha, \theta)$$

$$= \frac{1}{1 - (1+\theta)^{-\alpha}} \left[\frac{\alpha}{\gamma\sigma} \left(\frac{x-\mu}{\sigma} \right)^{(1/\gamma)-1} \left[1 + \left(\frac{x-\mu}{\sigma} \right)^{1/\gamma} \right]^{-(\alpha+1)} \right] \tag{39}$$

$$- \frac{(1+\theta)^{-\alpha}}{1 - (1+\theta)^{-\alpha}} \left[\frac{\alpha}{\gamma\sigma_1} \left(\frac{x-\mu}{\sigma_1} \right)^{(1/\gamma)-1} \left[1 + \left(\frac{x-\mu}{\sigma_1} \right)^{1/\gamma} \right]^{-(\alpha+1)} \right]$$

where $\sigma_1 = \sigma(1+\theta)^\gamma$. This is recognizable as a linear combination of two Pareto (IV) densities. Note that the density is a linear combination of two Pareto (IV) densities, but it is not a convex combination since, although the coefficients add up to 1, the second coefficient is negative. Motivated by this example, one might also consider k-component linear combinations of Pareto (IV) densities as income models, allowing k to be greater than 2. Such models with positive coefficients are natural candidates for fitting multimodal income data sets which may well have a mixture genesis. Note that, by testing the hypothesis that $\theta = 0$ one can decide whether or not the data set at hand has been subject to hidden truncation. More detailed discussion of these hidden truncation Pareto models, in the Pareto (II) case, may be found in Arnold and Ghosh (2011).

4 Inference, briefly

Suppose that $X_1, X_2, \ldots, X_n$ are independent identically distributed random variables with a common Pareto (IV) distribution. The sample size should be reasonably large, since we have four parameters to estimate. The sample minimum, or some minor corrected version of it, will be a suitable estimate of the location parameter μ. After subtracting it from each of the observations, the remaining three parameters may be estimated using maximum likelihood. The corresponding Fisher information matrix is available (as indeed is the Fisher information matrix for the Feller-Pareto model with $\mu = 0$).

Either a global search or numerical solution of the likelihood equations will be required to identify the location of the maximum of the likelihood function. In the Pareto (I) case, a variety of alternative estimates are available including best unbiased estimates. Alternatively, in the Pareto (I) case one can take logarithms of the data and arrive at a shifted exponential model, for which many estimation strategies have been developed.

A diffuse prior Bayesian analysis can be used for Pareto (IV) data. It will, predictably, yield results similar to those obtained via maximum likelihood. In the Pareto (I) case, Lwin (1972) introduced a conjugate family of priors for (σ, α) which can be used to incorporate some degree of prior knowledge of the parameters. Arnold et al. (1998) suggest use of a more flexible family of what they call conditionally conjugate priors in this setting. These priors are tailor-made for subsequent use of Gibbs sampling algorithms to generate realizations from the corresponding posterior distribution.

More details on parametric inference for Pareto models may be found in Arnold (1983) and Arnold (2008).

5   Multivariate Pareto models

The first author to systematically study k-dimensional Pareto distributions was Mardia (1962). Mardia's type I multivariate Pareto distribution has the attractive feature that both marginals and conditional distributions are Paretian in nature. We will say that a k-dimensional random vector $\underline{X}$ has a type I multivariate Pareto distribution, if the joint survival function is of the form

$$\bar{F}_{\underline{X}}(\underline{x}) = \left[\sum_{i=1}^{k} (x_i/\sigma_i) - k + 1 \right]^{-\alpha}, \quad x_i > \sigma_i \qquad (40)$$

and we write $\underline{X} \sim MP^{(k)}(\mathrm{I})(\underline{\sigma}, \alpha)$. The σ_i's are non-negative marginal scale parameters. The non-negative parameter α is an inequality parameter (common to all marginals). It follows from (40) that the one-dimensional marginals are classical Pareto distributions. Thus $X_i \sim P(\mathrm{I})(\sigma_i, \alpha)$, $i = 1, 2, \ldots, k$. By setting selected x_i's equal to σ_i in (40), it is apparent that, for any $k_1 < k$, all k_1 dimensional marginals are again multivariate Pareto. If we use the notational device $\underline{X} = (\dot{\underline{X}}, \ddot{\underline{X}})$ where $\dot{\underline{X}}$ is k_1 dimensional, with an analogous partition of the vector $\underline{\sigma} = (\dot{\underline{\sigma}}, \ddot{\underline{\sigma}})$, we may write

$$\dot{\underline{X}} \sim MP^{(k_1)}(\mathrm{I})(\dot{\underline{\sigma}}, \alpha). \qquad (41)$$

Conditional distributions are also of the form (40), but with a change of location.

It is natural to extend this basic multivariate Pareto model by the introduction of location, scale, inequality and shape parameters in a manner parallel to that used to develop the univariate Pareto (II)-(IV) distributions, as follows:

$(MP^{(k)}(II))$ We will say that $\underline{X}$ has a k-dimensional Pareto distribution of type II, if its joint survival function is of the form

$$\bar{F}_{\underline{X}}(\underline{x}) = \left[1 + \sum_{i=1}^{k} \left(\frac{x_i - \mu_i}{\sigma_i} \right) \right]^{-\alpha}, \qquad x_i > \mu_i, \qquad (42)$$
$$i = 1, 2, \ldots, k$$

and we write $\underline{X} \sim MP^{(k)}(\mathrm{II})(\underline{\mu}, \underline{\sigma}, \alpha)$.

$(MP^{(k)}(III))$ $\underline{X}$ has a k-dimensional Pareto distribution of type III, if its joint survival function is of the form

$$\bar{F}_{\underline{X}}(\underline{x}) = \left[1 + \sum_{i=1}^{k} \left(\frac{x_i - \mu_i}{\sigma_i} \right)^{1/\gamma_i} \right]^{-1}, \qquad x_i > \mu_i, \qquad (43)$$
$$i = 1, 2, \ldots, k$$

and we write $\underline{X} \sim MP^{(k)}(\mathrm{III})(\underline{\mu}, \underline{\sigma}, \underline{\gamma})$.

$(MP^{(k)}(IV))$ $\underline{X}$ has a k-dimensional Pareto distribution of type IV, if its joint survival function is of the form

$$\bar{F}_{\underline{X}}(\underline{x}) = \left[1 + \sum_{i=1}^{k} \left(\frac{x_i - \mu_i}{\sigma_i} \right)^{1/\gamma_i} \right]^{-\alpha}, \qquad x_i > \mu_i, \qquad (44)$$
$$i = 1, 2, \ldots, k$$

and we write $\underline{X} \sim MP^{(k)}(\mathrm{IV})(\underline{\mu}, \underline{\sigma}, \underline{\gamma}, \alpha)$.

The marginals and conditionals of an $MP^{(k)}(II)$ distribution are again of the $MP^{(k)}(II)$ form. An $MP^{(k)}(III)$ distribution has $MP^{(k)}(III)$ marginals, but not conditionals. However

an $MP^{(k)}(IV)$ distribution does have both its marginals and conditionals of the $MP^{(k)}(IV)$ form. Specifically, in the $MP^{(k)}(IV)$ case, using the dot – double dot notation, we have

$$\dot{\underline{X}} \sim MP^{(k_1)}(\text{IV})(\dot{\underline{\mu}}, \dot{\underline{\sigma}}, \dot{\underline{\gamma}}, \alpha) \tag{45}$$

and

$$\dot{\underline{X}} | \ddot{\underline{X}} = \ddot{\underline{x}} \sim MP^{(k_1)}(\text{IV})(\dot{\underline{\mu}}, \dot{\underline{\tau}}, \dot{\underline{\gamma}}, \alpha + k - k_1), \tag{46}$$

where

$$\tau_i = \sigma_i \left[1 + \sum_{j=k_1+1}^{k} \left(\frac{x_j - \mu_j}{\sigma_j} \right)^{1/\gamma_j} \right]^{\gamma_i}, \quad i = 1, 2, \ldots, k_1. \tag{47}$$

Takahasi (1965) discussed the $MP^{(k)}(IV)$ distribution with $\underline{\mu} = \underline{0}$ and $\underline{\sigma} = \underline{1}$. He called it a multivariate Burr's distribution, and noted that the marginal and conditional distributions were of the same form.

As in the univariate case, distributional properties of these multivariate Pareto distributions and possible further extensions are more transparent if one uses a representation of the variables as functions of certain independent gamma variables. Thus if $\underline{X}$ has a k-dimensional Pareto distribution of type IV, we may act as if the X_i's have the representation

$$X_i = \mu_i + \sigma_i (W_i/Z)^{\gamma_i}, \quad i = 1, 2, \ldots, k \tag{48}$$

where the W_i's are independent identically distributed $\Gamma(1, 1)$ variables (i.e., standard exponential variables) and Z, independent of the W_i's, has a $\Gamma(\alpha, 1)$ distribution. This representation, for example, makes it easy to compute the means, variances and covariances of $\underline{X}$.

A generalization of the representation (48) is one in which the W_i's are gamma rather than exponential variables. The resulting distribution will be called k-dimensional Feller-Pareto, since its marginals are of the Feller-Pareto form. Thus $\underline{X} \sim FP^{(k)}(\underline{\mu}, \underline{\sigma}, \underline{\gamma}, \alpha, \underline{\beta})$ if

$$X_i = \mu_i + \sigma_i (W_i/Z)^{\gamma_i}, \quad i = 1, 2, \ldots, k \tag{49}$$

where the W_i's and Z are independent random variables with $W_i \sim \Gamma(\beta_i, 1)$, ($i = 1, 2, \ldots, k$), and $Z \sim \Gamma(\alpha, 1)$. The marginal and conditional distributions of this multivariate Feller-Pareto distribution are again multivariate Feller-Pareto. The covariance structure can be readily obtained from the representation (49). Parallel to the situation in one dimension, there exist alternative names that could be applied to multivariate Feller-Pareto variables. They could be called multivariate generalized F variables or multivariate generalized beta of the second kind variables. An evident drawback of the multivariate Feller-Pareto model (and its various submodels) is the presence of a common value of α which appears in each marginal density. The consequences of this homogeneity are not easy to pin down. Certainly a model with Feller Pareto marginals with different α's for each of the marginals would be desirable, if one can be developed with attractive distributional properties (e,g., "nice" conditional distributions).

5.1 Other multivariate Pareto distributions

Although the title of this section promises discussion of multivariate models, only the bivariate case will be treated. It will be left to the reader to visualize the, usually straightforward, extension to the multivariate case. Notational complexity is avoided to a great extent by focusing on the case $k = 2$.

It is not difficult to verify that a $P(\text{IV})(\mu, \sigma, \gamma, \alpha)$ distribution can be represented as a scale mixture of Weibull distributions. Equivalently, as remarked earlier, that a $P(IV)$ random variable admits a representation as

$$X = \mu + \sigma (U/Z)^{\gamma}$$

where $U \sim exp(1)$ and $Z \sim \Gamma(\alpha, 1)$ are independent variables. A natural bivariate version of this construction begins with (U_1, U_2) having a bivariate exponential distribution with standard exponential marginals, perhaps one of the Marshall-Olkin type with parameters 1, 1 and λ. Then, with $Z \sim \Gamma(\alpha, 1)$ independent of (U_1, U_2), we define (X_1, X_2) by

$$X_i = \mu_i + \sigma_i (U_i/Z)^{\gamma_i}, \quad i = 1, 2. \tag{50}$$

Observe that in any bivariate Pareto (IV) distribution generated by this method, the marginals share a common value of α.

A second approach to generating bivariate $P(IV)$ distributions makes use of the following representation of a $P(IV)$ variable. Suppose that $U \sim exp(1)$, then

$$\mu + \sigma \left(e^{U/\alpha} - 1\right)^{\gamma} \sim P(\text{IV})(\mu, \sigma, \gamma, \alpha) \tag{51}$$

Here too then, we can begin with (U_1, U_2) having an arbitrary bivariate distribution with standard exponential marginals and construct a variable (X_1, X_2) with a bivariate Pareto (IV) distribution by defining

$$X_i = \mu_i + \sigma_i \left(e^{U_i/\alpha_i} - 1\right)^{\gamma_i}, \quad i = 1, 2. \tag{52}$$

A third approach makes use of the fact that minima of independent Pareto (IV) random variables themselves have Pareto (IV) distributions. Thus if X_i, $i = 1, 2$, are independent with $X_i \sim P(\text{IV})(\mu, \sigma, \gamma, \alpha_i)$, then $\min(X_1, X_2) \sim P(\text{IV})(\mu, \sigma, \gamma, \alpha_1 + \alpha_2)$. We then begin with three independent random variables Y_1, Y_2, Y_3 with $Y_i \sim P(\text{IV})(\mu, \sigma, \gamma, \alpha_i)$ and define

$$\begin{aligned} X_1 &= \min(Y_1, Y_3), \\ X_2 &= \min(Y_2, Y_3) \end{aligned} \tag{53}$$

(this approach is often called the method of trivariate reduction). In addition to having Pareto IV marginals, it is clear that the distribution described by (54) has the property that $\min(X_1, X_2) \sim P(\text{IV})(\mu, \sigma, \gamma, \alpha_1 + \alpha_2 + \alpha_3)$. This distribution has the perhaps undesirable property that $P(X_1 = X_2) > 0$, and has another unfortunate property in that the marginals share common values of μ, σ and γ. This latter problem can be avoided to some extent by assuming that the Y_i's have $P(\text{IV})(0, 1, \gamma, \alpha_i)$ distributions and then defining

$$\begin{aligned} X_1 &= \mu_1 + \sigma_1 \min(Y_1, Y_3), \\ X_2 &= \mu_2 + \sigma_2 \min(Y_2, Y_3). \end{aligned} \tag{54}$$

In this case the X_i's share only a common value of γ.

Finally we mention the popular Copula based approach to constructing bivariate distributions with given marginals. For this, we begin with an analytically tractable bivariate

distribution for (Z_1, Z_2) and apply marginal transformations to produce a bivariate distribution with Pareto (IV) marginals. A popular choice for the distribution of (Z_1, Z_2) is a bivariate normal with standard normal marginals and correlation ρ, but of course any other bivariate distribution can be used in its place. Now using $F_{\mu,\sigma,\gamma,\alpha}$ to denote the distribution function of a $P(\mathrm{IV})(\mu, \sigma, \gamma, \alpha)$ random variable and Φ to denote a standard normal distribution function, we define

$$
\begin{aligned}
X_1 &= F_{\mu_1,\sigma_1,\gamma_1,\alpha_1}^{-1}(\Phi(Z_1)), \\
X_2 &= F_{\mu_2,\sigma_2,\gamma_2,\alpha_2}^{-1}(\Phi(Z_2)).
\end{aligned}
\tag{55}
$$

The correlation structure of the X_i's is inherited from the correlation structure of the Z_i's. In this case the extension to k dimensions is particularly transparent. Note also that the model has one dependence parameter which, if set equal to 0, yields a model with independent $P(IV)$ marginals. It will be noted that this feature of having a single dependence parameter is shared by the other bivariate models introduced in this Section.

6 Multivariate extensions

Several of the univariate extensions, discussed in Section 3, can be readily modified to yield k-dimensional versions. For example a random variable with the univariate Beta-generalized-Pareto (IV) distribution can be viewed as being defined by

$$
X = F_{\mu,\sigma,\gamma,\alpha}^{-1}(V),
\tag{56}
$$

where $V \sim Beta(\lambda_1, \lambda_2)$. For a bivariate version of this construction, we begin with (V_1, V_2) having a bivariate Beta distribution, perhaps of the the type introduced by Arnold and Ng (2011), and make suitable marginal transformations. Thus we define

$$
\begin{aligned}
X_1 &= F_{\mu_1,\sigma_1,\gamma_1,\alpha_1}^{-1}(V_1), \\
X_2 &= F_{\mu_2,\sigma_2,\gamma_2,\alpha_2}^{-1}(V_2).
\end{aligned}
\tag{57}
$$

Higher dimensional versions of this construction require only the identification of a suitable k-dimensional Beta distribution. A Dirichlet distribution might be used here. Some other alternatives are described in Arnold and Ng (2011).

To identify a suitable bivariate analog of the Kumaraswamy-Pareto (IV) distribution, all that is required is a bivariate-Kumaraswamy distribution. One possible such distribution was suggested by Nadarajah et al. (2011).

Hidden truncation models, likewise, can be considered in higher dimensions. For example we may begin with $\underline{X} \sim MP^{(k)}(\mathrm{IV})(\underline{\mu}, \underline{\sigma}, \underline{\gamma}, \alpha)$. Then, using our dot $-$ double dot notation, we have

$$
f_{HT}(\dot{\underline{x}}) = f_{\dot{\underline{X}}|\ddot{\underline{X}} \leq \ddot{\underline{x}}}(\dot{\underline{x}}) = f_{\dot{\underline{X}}}(\dot{\underline{x}}) \frac{P(\ddot{\underline{X}} \leq \ddot{\underline{x}}|\dot{\underline{X}} = \dot{\underline{x}})}{P(\ddot{\underline{X}} \leq \ddot{\underline{x}})}
\tag{58}
$$

which is not difficult to evaluate, since the conditional distribution of $\ddot{\underline{X}}$ given that $\dot{\underline{X}} = \dot{\underline{x}}$ is of the $MP^{(k-k_1)}(IV)$ form (refer to equation (46), being careful to switch the roles of $\dot{\underline{X}}$ and $\ddot{\underline{X}}$).

We conclude this section by noting the availability of multivariate distributions with Pareto conditionals rather than Pareto marginals. Detailed discussion of such models may be found in Arnold et al. (1999), Chapter 5.

7 Envoi

The survey presented in this paper is far from complete. A more detailed and extensive survey (though somewhat out of date) can be found in Arnold (1983). A revision of that book is, however, currently in preparation. In the interim, see Arnold (2008) for a more up-to-date presentation and, as mentioned in Section 4, for more details on inferential strategies. More work is still needed on the development of estimation and hypothesis testing strategies, especially for multivariate Pareto data. Creative Bayesian analyses involving informative priors, in multivariate settings and in cases involving covariates, are also notable for their absence. Finally, I apologize to those readers whose important contributions have been overlooked in this survey. I excuse myself by repeating that the survey is necessarily incomplete. However, please do advise me of any glaring omissions that you might note.

References
Aban, IB, Meerschaert, MM, Panorska, AK: Parameter estimation for the truncated Pareto distribution. J. Amer. Statist. Assoc. **101**, 270–277 (2006)
Akinsete, A, Famoye, F, Lee, C: The beta-Pareto distribution. Statistics. **42**, 547–563 (2008)
Arnold, BC: Pareto Distributions. International Cooperative Publishing House, Burtonsville, MD (1983)
Arnold, BC: Pareto and generalized Pareto distributions. In: Chotikapanich, D (ed.) Modeling Income Distributions and Lorenz Curves, pp. 119–145. Springer, New York, (2008)
Arnold, BC, Austin, K: Truncated Pareto distributions: flexible, tractable and familiar. Technical Report #150, Department of Statistics, University of California, Riverside, CA (1987)
Arnold, BC, Castillo, E, Sarabia, JM: Bayesian analysis for classical distributions using conditionally specified priors. Sankhya: Ind. J. Stat. Series B. **60**, 228–245 (1998)
Arnold, BC, Castillo, E, Sarabia, JM: Conditional Specification of Statistical Models. Springer, New York (1999)
Arnold, BC, Ghosh, I: Inference for Pareto data subject to hidden truncation. J. Ind. Soc. Probability Stat. **13**, 1–16 (2011)
Arnold, BC, Laguna, L: A stochastic mechanism leading to asymptotically Paretian distributions. In: Business and Economic Statistics Section, Proceedings of the American Statistical Association, pp. 208–210, (1976)
Arnold, BC, Ng, HKT: Flexible bivariate beta distributions. J. Multivariate Anal. **102**, 1194–1202 (2011)
Arnold, BC, Robertson, CA, Yeh, HC: Some properties of a Pareto-type distribution. Sankhya: Ind. J. Stat. Series A. **48**, 404–408 (1986)
Champernowne, DG: The theory of income distribution. Econometrica. **5**, 379–381 (1937)
David, HA, Nagaraja, HN: Order Statistics. Third edition. Wiley, Hoboken, NJ (2003)
Falk, M, Hüsler, J, Reiss, R: Laws of Small Numbers, Extremes and Rare Events. Third edition. Birkhäuser/Springer, Basel (2011)
Feller, W: An Introduction to Probability Theory and its Applications, Vol. 2. Second edition. Wiley, New York (1971)
Fisk, PR: The graduation of income distributions. Econometrica. **29**, 171–185 (1961a)
Fisk, PR: Estimation of location and scale parameters in a truncated grouped sech-square distribution. J. Am. Stat. Assoc. **56**, 692–702 (1961b)
Johnson, NL, Kotz, S, Balakrishnan, N: Continuous Univariate Distributions, Vol. 1. Second edition. Wiley, New York (1994)
Jones, MC: Families of distributions arising from distributions of order statistics. TEST. **13**, 1–43 (2004)
Jones, MC, Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. Stat. Methodol. **6**, 70–81 (2009)
Kagan, YY, Schoenberg, F: Estimation of the upper cutoff parameter for the tapered Pareto distribution. J. Appl. Probability. **38A**, 158–175 (2001)
Kalbfleisch, JD, Prentice, RL: The Statistical Analysis of Failure Time Data. Wiley, New York (1980)
Keiding, N, Kvist, K, Hartvig, H, Tvede, M, Juul, S: Estimating time to pregnancy from current durations in a cross-sectional sample. Biostatistics. **3**, 565–578 (2002)
Lwin, T: Estimation of the tail of the Paretian law. Skand. Aktuarietidskr. **55**, 170–178 (1972)
Maguire, BA, Pearson, ES, Wynn, AHA: The time intervals between industrial accidents. Biometrika. **39**, 168–180 (1952)

Mardia, KV: Multivariate Pareto distributions. Ann. Math. Stat. **33**, 1008–1015 (1962)

Nadarajah, S, Cordeiro, GM, Ortega, EMM: General results for the Kumaraswamy-G distribution. J. Stat. Comput. Simul. **82**, 951–979 (2011)

Pareto, V: Cours d'economie Politique, Vol. II, F. Rouge, Lausanne (1897)

Paranaiba, PF, Ortega, EMM, Cordeiro, GM, de Pascoa, MAR: The Kumaraswamy Burr XII distribution: theory and practice. J. Stat. Comput. Simul. **83**, 2117–2143 (2013)

Pillai, RN: Semi-Pareto processes. J. Appl. Probab. **28**, 461–465 (1991)

Takahasi, K: Note on the multivariate Burr's distribution. Ann. Inst. Statist. Math. **17**, 257–260 (1965)