STATISTICA, anno LXXIV, n. 2, 2014

A FAMILY OF BIVARIATE PARETO DISTRIBUTIONS

P. G. Sankaran
Department of Statistics Cochin University of Science and Technology, Cochin
N. Unnikrishnan Nair
Department of Statistics Cochin University of Science and Technology, Cochin
Preethi John
Department of Statistics Cochin University of Science and Technology, Cochin

1. Introduction

The univariate Pareto distribution was first proposed in literature as a model for income analysis. The probability density function $f(x)$ of the Pareto distribution is defined by

$$f(x) = \frac{a\,k^a}{x^{a+1}} \quad ; \ x \geq k,\ a, k > 0. \tag{1}$$

The graph of $f(x)$ shows that the fraction of the population that owns a small amount of wealth per person is rather high and then decreases steadily as wealth increases. Arnold (1985) has studied various properties of (1) and its extensions using transformations of the random variable. The distribution (1) can also be used to model the sizes of human settlement, the values of oil reserves in oil fields, hard disk drive error rates, the standarized price returns on individual stocks, sizes of sand particles and large casualty losses for certain lines of business. For various applications of the Pareto distribution, one could refer to Arnold (1985) and Johnson *et al.* (1994).

As in the case of univariate Pareto distributions, mathematical simplicity and tractability have provided a lot of interest in the theory and applications of multivariate Pareto distributions. Accordingly, Mardia (1962) introduced a bivariate Pareto distribution with joint density function $f(x, y)$ as

$$f(x,y) = \begin{cases} \frac{a(a+1)(pq)^{a+1}}{(px+qy-pq)^{a+2}}; & x \geq p, y \geq q, a > 0 \\ 0 & \text{other wise.} \end{cases} \tag{2}$$

The distribution (2) is referred as bivariate Pareto distribution of first kind, since the marginal distributions have univariate Pareto (1). The bivariate Pareto distribution of the second kind was also introduced by Mardia (1962) using two dependent gamma variables in the sense of Kibble (1941). Later Lindley and Singpurwalla (1986) have introduced a bivariate Pareto II distribution which has simple joint survival function with Pareto II marginals. This distribution was further studied and generalized by Nayak (1987), Barlow and Mendel (1992), Sankaran and Nair (1993), Langseth (2002), Balakrishnan and Lai (2009) and Sankaran and Kundu (2014). For various other bivariate Pareto distributions and its generalizations, one may refer to Arnold (1990), Arnold (1992) and Kotz *et al.* (2002).

When there is very little information about the data generating mechanism, it is desirable in modeling problems to begin with a family of distributions which is quite flexible in the desired characterstics. The distributions discussed earlier are individual in nature and suits only for a particular set of data that meets the specified requirements. Motivated by this, in the present paper, we introduce a class of bivariate Pareto distributions arising from a generalization of the univariate dullness property by which Talwalker (1980) has characterized the Pareto law (1). It is shown that the marginal distributions of the proposed model are univariate Pareto I models. The proposed bivariate family consists of some known and several new models. It also imparts enough flexibility in terms of desirable properties that are generally used in modelling problems.

The rest of the article is organized as follows. In Section 2, we introduce a family of bivariate Pareto distributions. Various members belonging to the family are identified. The distributional properties of the family are discussed in Section 3. In Section 4, we study dependence structure of the family of distributions. In Section 5 we discuss the inference procedure and apply the proposed class of models to two real data sets. Finally, Section 6 summarizes the major conclusions of the study.

2. The Model

Let (X, Y) be a non-negative random vector having absolutely continuous survival function $\bar{F}(x, y) = P(X > x, Y > y)$. In order to construct the proposed family of bivariate Pareto distributions, we assume that Z is a non-negative random variable with continuous and strictly decreasing survival function $\bar{G}(z)$ and cumulative hazard function $H(z)$ defined by $H(z) = -\log \bar{G}(z)$. We require the following theorem to construct the proposed bivariate Pareto family.

Theorem 1. *The random variable Z satisfies the property*

$$P(Z > \log g(x, y) | Z > a \log x) = P(Z > b \log y) \tag{3}$$

for all $a, b > 0, x, y > 1$ and some $g(x, y) > x^a$ if and only if

$$H(\log g(x, y)) = H(a \log x) + H(b \log y) \tag{4}$$

Proof. Since $H^{-1}(t) = \bar{G}^{-1}(e^{-t})$ for all $t > 0$

$$\begin{aligned} H^{-1}(H(a \log x) + H(b \log y)) &= \bar{G}^{-1}(\exp[-H(a \log x) - H(b \log y)]) \\ &= \bar{G}^{-1}(\bar{G}(a \log x).\bar{G}(b \log y)) \end{aligned} \tag{5}$$

or

$$\bar{G}\, H^{-1}(H(a \log x) + H(b \log y)) = \bar{G}(a \log x).\bar{G}(b \log y) \tag{6}$$

Now to prove Theorem 1 first assume (3). It is equivalent to

$$\begin{aligned}\bar{G}(\log g(x,y)) &= \bar{G}(a\log x).\bar{G}(b\log y)\\ &= \bar{G}\,[H^{-1}(H(a\log x)+H(b\log y))]\end{aligned}$$

using (5) and (6) from which we have (4). Conversely assuming (4), we obtain

$$\begin{aligned}P(Z>\log g(x,y)|\,Z>a\log x) &= \frac{\bar{G}(\log g(x,y))}{\bar{G}(a\log x)}\\ &= \frac{\exp[-H(\log g(x,y))]}{\exp[-H(a\log x)]}\\ &= P[Z>b\log y].\end{aligned}$$

□

We notice that $g(x,y)$ is a function of (x,y) in $R_2^+=\{(x,y)|x,y>0\}$ satisfying the property (4). Further

(a) $g(1,y)=y^b, g(x,1)=x^a$,

(b) $g(\infty,y)=\infty, g(x,\infty)=\infty$,

(c) since $H(.)$ is increasing and continuous, $g(x,y)$ is also increasing and continuous in x and y and

(d) it is assumed that $g(x,y)$ satisfies the inequality $\frac{2}{g(x,y)}\,\frac{\partial g}{\partial x}\frac{\partial g}{\partial y}\;-\;\frac{\partial^2 g}{\partial x\partial y}\geq 0$.

From properties (a) through (d) it follows that

$$\bar{F}(x,y)=[g(x,y)]^{-1},\;\;x,y\;>1 \tag{7}$$

is the survival function of a random vector (X,Y) with Pareto I marginals $\bar{F}_X(x)=x^{-a}$, $x>1$ and $\bar{F}_Y(y)=y^{-b}$, $y>1$.

This completes the procedure for constructing the family of bivariate Pareto distributions based on $g(x,y)$ arising from a property characterizing a class of univariate distributions. We designate $\bar{G}(z)$ as the baseline distribution that corresponds to $\bar{F}(x,y)$, since the members of the family are generated through the functional equation (4) based on $H(z)$, the cumulative hazard rate of $\bar{G}(z)$.

We derive some members of the family as follows.

1. Let Z be exponential with $\bar{G}_1(z)=\exp(-\lambda\,z), z>0$ so that $H(z)=\lambda z$.Then $g(x,y)=x^a y^b$. Then the bivariate distribution is

$$\bar{F}_1(x,y)=x^{-a}y^{-b};\;x>y>1;a,b>0. \tag{8}$$

2. When Z has Gompertz distribution $\bar{G}_2(z)=\exp[-\theta(e^{\alpha z}-1)]; z\geq 0;\alpha,\theta>0$ $H(z)=\theta(e^{\alpha z}-1)$ and the resulting bivariate distribution is

$$\bar{F}_2(x,y)=(x^{a\alpha}+y^{b\alpha}-1)^{\frac{-1}{\alpha}};x,y>1,\alpha,a>0. \tag{9}$$

Setting $\alpha = \frac{1}{a} = \frac{1}{b}$, we obtain

$$\bar{F}_3(x,y) = (x+y-1)^{-a}; x, y > 1, \tag{10}$$

the well known Mardia's(1962) type I bivariate Pareto model.

3. Take Z to be a Pareto II variable with $\bar{G}_4(z) = (1+\beta z)^{-\alpha}$ to get $H(z) = \alpha \log(1+\beta z)$. Then we have the bivariate law

$$\bar{F}_4(x,y) = x^{-a-c\log y} y^{-b}, x, y > 1, a, b > 0; 0 \le c \le 1. \tag{11}$$

4. If Z has half-logistic distribution specified by the survival function

$$\bar{G}_5(z) = 2(1+e^{\frac{z}{\sigma}})^{-1},\ z > 0, \sigma > 0.$$

The model is

$$\bar{F}_5(x,y) = [\frac{1}{2}(x^{\alpha} + y^{\beta} + x^{\alpha}y^{\beta} - 1)]^{-\sigma}; \alpha = \frac{a}{\sigma} > 0, \sigma > 0, \beta = \frac{b}{\sigma} > 0. \tag{12}$$

5. The Burr XII distribution (Pareto IV) $\bar{G}_6(z) = (1+z^c)^{-k},\ z > 0;\ c,\ k > 0$ with $H(z) = k\log(1+z^c)$ for Z leaves the bivariate model as

$$\bar{F}_6(x,y) = \exp[-(a\log x)^c - (b\log y)^c - (ab\log x \log y)^c]^{\frac{1}{c}}. \tag{13}$$

6. Suppose Z follows the distribution $\bar{G}_7(z) = (2e^z - 1)^{-\sigma}, z > 0; \sigma > 0$, Then $H(z) = \sigma\log(2e^z - 1)$ and

$$\bar{F}_7(x,y) = (1 + 2x^a y^b - x^a - y^b)^{-1}. \tag{14}$$

7. When Z is distributed as Weibull $\bar{G}_8(z) = e^{-(\lambda z)^{\alpha}}\ \ \alpha, \lambda > 0, z > 0$ gives $H(z) = (\lambda z)^{\alpha}$ and

$$\bar{F}_8(x,y) = \exp[\frac{-1}{\lambda}\{(\lambda a\log x)^{\alpha} + (\lambda b\log y)^{\alpha}\}^{\frac{1}{\alpha}}]. \tag{15}$$

8. If Z has generalized exponential distribution $\bar{G}_9(z) = \frac{p}{e^{\lambda z} - q},\ z > 0;$ $\lambda > 0,\ 0 < p < 1,\ q = 1 - p$
   we have $H(z) = \log\frac{e^{\lambda z} - q}{p}$ and

$$\bar{F}_9(x,y) = (q + p^{-1}(x^{a\lambda} - q)(y^{b\lambda} - q))^{\frac{-1}{\lambda}}. \tag{16}$$

9. Taking $\bar{G}_{10}(z) = (1 + \frac{e^{\lambda z} - 1}{\alpha})^{-1},\ \alpha, \lambda > 0$, the cumulative hazard function $H(z) = \log(1 + \alpha^{-1}(e^{\lambda z} - 1))$ provides the bivariate Pareto

$$\bar{F}_{10}(x,y) = (1 + \alpha^{-1}(\alpha + x^{a\lambda} - 1)(\alpha + y^{b\lambda} - 1) - \alpha)^{\frac{-1}{\lambda}}. \tag{17}$$

REMARK 2. *The method of construction provide a large class of bivariate Pareto distributions. Any $\bar{G}(z)$ which is strictly increasing and provides a $g(x, y)$ satisfying conditions (a) to (d) can give rise to a bivariate Pareto model. The above cases 1 to 9 comprises only of some simple forms that does not exhaust the members of the family.*

REMARK 3. *When $a = b$ in the above scheme, we have an exchangeable family of Pareto distributions. Such a restriction becomes quite handy in inference problems using Bayesian approach. In that case, $\bar{F}_1(x, y)$ is the only Schur-constant model belonging to the family.*

REMARK 4. *A random variable W (or its probability distribution) satisfies dullness property (Talwalker (1980)) if for all $x, y \geq 1$*

$$P(W \geq x\,y|\, W \geq x) \, = P(W \geq y) \tag{18}$$

It may be easy to observe that the property (3) *reduces to the dullness property* (18) *when $Z = \log W$, $g(x, y) = xy$ and $a = b = 1$.*

REMARK 5. *Although the family* (7) *comprises of a large number of members, every bivariate Pareto distribution does not belong to it. For example, the survival function*

$$S(x, y) = \, x^{\frac{-a}{2}} y^{\frac{-a}{2}} \exp[-\frac{1}{2}((a \log x)^2 + (a \log y)^2)^{\frac{1}{2}}] \;\; x, y > 1, a > 0 \tag{19}$$

represents a bivariate Pareto model with Pareto I marginals. If it belongs to the family one must have

$$g(x, y) = x^{\frac{a}{2}} y^{\frac{a}{2}} \exp[\frac{1}{2}((a \log x)^2 + (a \log y)^2)^{\frac{1}{2}}] \tag{20}$$

that satisfies(4) *for some cumulative hazard function $H(.)$ of a nonnegative random variable Z, for all x, y. If* (20) *is true for all x, y, it should also hold for*

$$H(\log g(x, x)) = 2H(a \log x)$$

or

$$H \log(x^{(\frac{\sqrt{2}+1}{\sqrt{2}} a)}) = 2H(a \log x)$$

or

$$\frac{1}{2} H(\frac{\sqrt{2}+1}{\sqrt{2}} t) = H(t)\,;\, t = a \log x \tag{21}$$

for all $t > 0$. It is known from Kagan et al. *(1973) that the functional equation*

$$A(x) = kA(\theta x),\; \theta > 0\;;\; A(0) = 0 \tag{22}$$

has a solution only if $0 < \theta < 1 < k$. By analogy (21) *is a particular case of* (22) *with $\theta = \frac{\sqrt{2}+1}{\sqrt{2}} > 1$ and hence there is no admissible $H(x)$ that satisfy* (21). *Thus* (19) *does not belong to the proposed family* (7).

3. Properties

3.1. Marginal Distributions

As shown in Section 2, the marginal distributions of the family (7) are univariate Pareto I models. The joint density functions of the various models are presented in Table 1.

3.2. Conditional Distributions

There are two kinds of conditional distributions of interest. One is the usual $f_1(x|y) = \frac{f(x,y)}{a_2(y)}$ and $f_2(y|x) = \frac{f(x,y)}{a_1(x)}$ where $f(x,y)$ is the joint density function and $a_1(x)$ and $a_2(y)$ are respectively the marginal density functions of X and Y. These conditional density functions are given respectively in Table 2. The second type of conditional distributions required in the sequel are conditional distributions of X (Y) given $Y > y$ $(X > x)$ denoted by $f_1(x|Y > y)$ and $f_2(y|X > x)$ or equivalently the conditional survival functions $P(X > x|Y > y)$ and $P(Y > y|X > x)$. These are exhibited in Table 3. Note that these two sets of conditional distributions determine the joint distributions in the family.

3.3. Regression Functions

The bivariate Pareto family (7) is rich enough in the sense that it contains a large number of members that could be candidates for different data situations. The members of the family have basically two shape parameters besides location and scale parameters that can be arbitrarily introduced in the models. Further, the members are highly flexible in various distributional characterstics to represent a wide variety of models. The last aspect needs a detailed consideration with reference to some important model characterstics that are often required in data analysis problems.

We represent the regression functions $A(x) = E(Y|X = x)$ and $B(y) = E(X|Y = y)$ with suffixies corresponding to the member distributions. Accordingly for $\bar{F}_1(x,y)$, the regression functions are constants, being the respective means. In the case of $\bar{F}_3(x,y)$, we obtain

$$A_3(x) = (1 + \frac{x}{a})$$

and

$$B_3(y) = (1 + \frac{y}{b}),$$

both linearly increasing functions. They intersect on the means $(E(X), E(Y))$ of the distributions. However for $\bar{F}_2(x,y)$, the regression functions are

$$A_2(x) = \frac{1+\alpha}{\alpha} \frac{x^{a(1+\alpha)}}{(x^a - 1)^{1+\frac{b-1}{b\alpha}}}$$

and

$$B_2(y) = \frac{1+\alpha}{\alpha} \frac{y^{b(1+\alpha)}}{(y^b - 1)^{1+\frac{a-1}{a\alpha}}}$$

which are non-linear in character. These functions do not intersect at the means. All the remaining distributions also have non-linear regressions, but with different functional forms. For instance, $\bar{F}_4(x, y)$ has

$$B_4(y) = \frac{b(a + c \log y)(a + c \log y - 1) + c}{b(a + c \log y - 1)^2}$$

which is a logarithmic function, where as for $\bar{F}_7(x, y)$

$$B_7(y) = \frac{2(1+t)^2}{t^{\frac{a-1}{a}}(2t+1)^{\frac{1}{a}+1}}[B_u(\frac{1}{a}+1, \frac{a-1}{a}) - \frac{1}{t}B_u(\frac{1}{a}+1, \frac{2a-1}{a})]$$

where $u = \frac{3-2y^b}{y^b-2}$, $t = 1-y^b$ and $B_u(p, q) = \int\limits_u^1 z^{p-1}(1-z)^{q-1}dz$ is the incomplete beta function. The expressions for $A_4(x)$ and $A_7(x)$ are obtained by changing a to b, b to a and y to x.

4. Dependence Structure

Since the bivariate distributions in the proposed family have identical marginal distributions, in modelling and analysis of data, a crucial aspect that differentiate them in a practical situation is the differences in the dependence or association between the constituent random variables. Thus a study of various dependence concepts and measures become crucial when discussing family properties, as they tell us the extent to which the variables are associated and also the nature of their relationships. There are three distinct approaches in the study of association. The first one is through numerical measures like the Pearson's correlation coefficient, the Kendall's tau, Spearman's rho, Gini's measure and Blomqvist's β. Presently we discuss the correlation coefficient and postpone the study of the other measures in a seperate work when the copulas of the member distributions are taken up. A second approach is to study the dependence properties. The six basic properties of positive dependence are (1) total positivity of order 2 (2) stochastic increase (3) right tail increase (4) positive association (5) positive quadrant dependence and (6) positive correlation or $Cov(X, Y) \geq 0$. Negative dependence properties are defined as the dual's of these. Among the six properties, the relative stringency is expressed as follows

(1) $\Rightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (4) $\Rightarrow$ (5) $\Rightarrow$ (6).

Finally, we have local measures of dependence, which measures the dependence structure at specific values of x and y. These become important in survival studies where the duration spent in a specific state of a disease is crucial and also in economics where income of individuals below the poverty line or above the affluence level is of importance. The Holland and Wang dependence function, Clayton-Oakes measure, Bjerve and Doksum's correlation curve, Anderson measure, Nair and Sankaran function etc belong to this category (See Nair and Sankaran (2010) for references). We will now discuss each of these approaches in some detail, with illustrative examples from the members of the family.

TABLE 1

Joint density functions for various types of Pareto models

Model	Joint density function
$f_1(x,y)$	$a\,b\,x^{-a-1}\,y^{-b-1}$
$f_2(x,y)$	$a\,b\,(1+\alpha)x^{\alpha a-1}\,y^{\alpha b-1}(x^{\alpha a}+y^{\alpha b}-1)^{\frac{-1}{\alpha}-2}$
$f_3(x,y)$	$a(a+1)(x+y-1)^{-a-2}$
$f_4(x,y)$	$x^{-a-c\log y-1}y^{-b-1}[(a+c\log y)(b+c\log x)-c],\ 0\leq c\leq a\,b$
$f_5(x,y)$	$\frac{\alpha\,\beta\,\sigma}{4}\ [\frac{1}{2}(1+x^{\alpha})(1+y^{b})-1]^{-\sigma-2}(1+\sigma(1+x^{\alpha})(1+y^{b})x^{\alpha-1}y^{\beta-1}$
$f_7(x,y)$	$2\,ab\frac{(2x^a y^b-x^a-y^b)}{(1+2x^a y^b-x^a-y^b)^3}x^{a-1}y^{b-1}$
$f_8(x,y)$	$a\,b\ \exp[-\frac{1}{\lambda}\{(\lambda\,a\ \log x)^{\alpha}+(\lambda\,b\ \log y)^{\alpha}\}^{\frac{1}{\alpha}}][(\lambda\,a\ \log x)^{\alpha}+(\lambda\,b\ \log y)^{\alpha}]^{\frac{1}{\alpha}-2}\frac{(\lambda\,a\ \log x)^{\alpha-1}(\lambda\,b\ \log y)^{\alpha-1}}{x\,y}$
$f_9(x,y)$	$\frac{\lambda\,ab}{p}[q+\frac{(x^{\lambda\,a}-q)(y^{\lambda\,b}-q)}{p}]^{-\frac{1}{\lambda}-2}[\frac{(x^{\lambda\,a}-q)(y^{\lambda\,b}-q)}{\lambda p}-q]x^{\lambda\,a-1}y^{\lambda\,b-1}$

TABLE 2
Conditional densities $f_1(x|y)$ and $f_2(y|x)$

Joint density function	$f_1(x\|y)$	$f_2(y\|x)$
$f_1(x,y)$	$a\,x^{-a-1}$	$b\,y^{-b-1}$
$f_2(x,y)$	$(1+\alpha)a(x^{a\alpha}+y^{b\alpha}-1)^{-\frac{1}{\alpha}-2}x^{a\alpha-1}y^{(1+\alpha)b}$	$(1+\alpha)b(x^{a\alpha}+y^{b\alpha}-1)^{-\frac{1}{\alpha}-2}y^{b\alpha-1}x^{(1+\alpha)a}$
$f_3(x,y)$	$(a+1)\,(x+y-1)^{-a-2}y^{a+1}$	$(a+1)\,(x+y-1)^{-a-2}x^{a+1}$
$f_4(x,y)$	$a^{-1}\,[(a+c\log y)(b+c\log x)-c]\,x^{-c\log y}y^{-b-1}$	$b^{-1}[(a+c\log y)(b+c\log x)-c]x^{-a-c\log y-1}$
$f_5(x,y)$	$\frac{1}{4}\beta[\frac{1}{2}(1+x^{\alpha})(1+y^{\beta})-1]^{-\sigma-2}$ $(1+\sigma(1+x^{\alpha})(1+y^{\beta}))x^{\alpha-1}y^{\beta(\sigma+1)}$	$\frac{1}{4}\beta[\frac{1}{2}(1+x^{\alpha})(1+y^{\beta})-1]^{-\sigma-2}(1+\sigma(1+x^{\alpha})(1+y^{\beta}))$ $x^{(\sigma+1)\alpha}y^{\beta-1)}$
$f_7(x,y)$	$2\,b\frac{(2x^ay^b-x^a-y^b)\,x^{a-1}y^{2b}}{1+2x^ay^b-x^a-y^b}$	$2\,a\frac{(2x^ay^b-x^a-y^b)\,x^{2a}y^{b-1}}{1+2x^ay^b-x^a-y^b}$
$f_8(x,y)$	$\exp[-\frac{1}{\lambda}\{(\lambda\,a\log x)^{\alpha}+(\lambda\,b\log y)^{\alpha}\}^{\frac{1}{\alpha}}[(\lambda\,a\log x)^{\alpha}$ $+(\lambda\,b\log y)^{\alpha}]^{\frac{1}{\alpha}-2}\{(\lambda\,a\log x)^{\alpha}+(\lambda\,b\log y)^{\alpha}\}^{\frac{2}{\alpha}}$ $-(1-\alpha)\lambda](\lambda\,a\log x)^{\alpha-1}(\lambda\,b\log y)^{\alpha-1}x^{-1}y^{b}$	$\exp[-\frac{1}{\lambda}\{(\lambda\,a\log x)^{\alpha}+(\lambda\,b\log y)^{\alpha}\}^{\frac{1}{\alpha}}$ $[(\lambda\,a\log x)^{\alpha}+(\lambda\,b\log y)^{\alpha}]^{\frac{1}{\alpha}-2}\{[(\lambda\,a\log x)^{\alpha}+(\lambda\,b\log y)^{\alpha}]^{\frac{2}{\alpha}}$ $-(1-\alpha)\lambda\}x^{a}y^{-1}$
$f_9(x,y)$	$[q+\frac{(x^{\lambda\,a}-q)(y^{\lambda\,b}-q)}{p}]^{-\frac{1}{\lambda}-2}[\frac{(x^{\lambda\,a}-q)(y^{\lambda\,b}-q)}{\lambda p}-q]$ $p^{-1}\lambda\,a\,x^{\lambda a-1}y^{(1+\lambda)b}$	$[q+\frac{(x^{\lambda\,a}-q)(y^{\lambda\,b}-q)}{p}]^{-\frac{1}{\lambda}-2}[\frac{(x^{\lambda\,a}-q)(y^{\lambda\,b}-q)}{\lambda p}-q]p^{-1}\lambda\,b\,x^{(1+\lambda)a}y^{\lambda b-1}$

TABLE 3
Conditional survival functions

Distribution	$P(X > x \mid Y > y)$	$P(Y > y \mid X > x)$
$F_1(x,y)$	x^{-a}	y^{-b}
$\bar{F}_2(x,y)$	$(x^{a\alpha} + y^{b\alpha} - 1)^{-\frac{1}{\alpha}} y^b$	$(x^{a\alpha} + y^{b\alpha} - 1)^{-\frac{1}{\alpha}} x^a$
$\bar{F}_3(x,y)$	$(\frac{x+y-1}{y})^{-a}$	$(\frac{x+y-1}{x})^{-a}$
$F_4(x,y)$	$x^{-a-c\log y}$	$x^{-c\log y} y^{-b}$
$\bar{F}_5(x,y)$	$[\frac{1}{2}(\frac{x^\alpha+y^\beta+x^\alpha y^\beta-1}{y^\beta})]^{-\sigma}$	$[\frac{1}{2}(\frac{x^\alpha+y^\beta+x^\alpha y^\beta-1}{x^\alpha})]^{-\sigma}$
$\bar{F}_7(x,y)$	$\frac{y^b}{1+2x^a y^b - x^a - y^b}$	$\frac{x^a}{1+2x^a y^b - x^a - y^b}$
$\bar{F}_8(x,y)$	$\exp[-\frac{1}{\lambda}\{(\lambda\, a \log x)^\alpha + (\lambda\, b \log x)^\alpha\}^{\frac{1}{\alpha}}] y^b$	$\exp[-\frac{1}{\lambda}\{(\lambda\, a \log x)^\alpha + (\lambda\, b \log x)^\alpha\}^{\frac{1}{\alpha}}] x^a$
$\bar{F}_9(x,y)$	$[q + \frac{(x^{\lambda\, a}-q)(y^{\lambda\, b}-q)}{p}]^{-\frac{1}{\lambda}} y^b$	$[q + \frac{(x^{\lambda\, a}-q)(y^{\lambda\, b}-q)}{p}]^{-\frac{1}{\lambda}} x^a$

4.1. Correlation Coefficient

Obviously $\bar{F}_1(x,y)$, in which the variables are independent, has zero correlation coefficient. The Mardia form $\bar{F}_3(x,y)$, has coefficient of correlation $R_3 = \frac{1}{a}$ (Kotz *et al.* (2002)). Since the variances of X and Y exist only when $a > 2$, we see that the model exhibits a low correlation lying in $(0, \frac{1}{2})$. As regards $\bar{F}_4(x,y)$,the correlation coefficient (R) has the form

$$R_4 = [\frac{(a-2)(b-2)}{a\,b}]^{\frac{1}{2}}[\frac{(a-1)(b-1)}{c}\,e^{\frac{(a-1)(b-1)}{c}}E_1(\frac{(a-1)(b-1)}{c}) - 1]$$

where $E_1(z) = \int\limits_1^\infty \frac{e^{-z\,t}}{t}\,dt$, $Rez > 0$ is the exponential integral discussed in Abramowitz and Stegun (1966). When $c = 0$, the distribution $\bar{F}_4(x,y)$ is the product of the marginal distributions of X and Y which means that X and Y are independent and hence $R_4 = 0$. For any fixed values of $a, b > 2$, R_4 is a decreasing function of c. Thus as c runs through $[0, ab]$, the correlation coefficient becomes increasingly negative. When $c = ab$

$$R_4 = (\frac{(a-2)(b-2)}{a\,b})^{\frac{1}{2}}[p\,e^p E_1(p) - 1]$$

where $p = \frac{(a-1)(b-1)}{ab}$. As a, b tends to infinity, $\lim_{a,b\to\infty} R_4 = [eE_1(1) - 1]$ which is always negative. Thus (X,Y) is always negatively correlated.

In all other cases R involves integrals of incomplete beta function to enable an algebraic analysis of R difficult. However, the nature of the correlation will be deduced below using other dependence concepts.

4.2. Dependence Concepts

While studying the dependence concepts in relation to the members of the bivariate Pareto family, we begin with the strongest concepts in view of the implications to others already considered. We say that a bivariate probability density function $f(x,y)$ is totally positive of order 2- TP_2 (reverse regular of order 2-RR_2) if and only if for all $x_1 < x_2, y_1 < y_2$

$$f(x_1,y_1)f(x_2,y_2) \geq (\leq) f(x_1,y_2)f(x_2,y_1) \tag{23}$$

(Barlow and Proschan (1975)).
In the case of the Mardia form $f_3(x,y)$ from Table 1,we consider the difference

$$f_3(x_1,y_1)f_3(x_2,y_2) - f_3(x_1,y_2)f_3(x_2,y_1)$$

$$= \frac{a(a+1)}{(x_1+y_1-1)^{a+2}}\,\frac{a(a+1)}{(x_2+y_2-1)^{a+2}} - \frac{a(a+1)}{(x_1+y_2-1)^{a+2}}\,\frac{a(a+1)}{(x_2+y_1-1)^{a+2}}$$

The sign of the above expression depends on $(x_1 - x_2)(y_1 - y_2)$ which is non-negative for all $x_1 < x_2$ and $y_1 < y_2$. Hence $f_3(x,y)$ is TP_2. For the more general $f_2(x,y)$, the difference leads to the determination of the sign from $(x_1^{a\alpha} - x_2^{a\alpha})(y_1^{b\alpha} - y_2^{b\alpha})$ which is positive. Thus $f_2(x,y)$ is TP_2. Since $TP_2 \Rightarrow Cov(X,Y) \geq 0$, we deduce that in the case of $f_2(x,y)$, X and Y are positively correlated.

The density function $f_4(x,y)$ is neither TP_2 nor RR_2. However $\bar{F}_4(x,y)$ is RR_2 as evidenced from

$$\bar{F}(x_1,y_1)\bar{F}(x_2,y_2) - \bar{F}(x_1,y_2)\bar{F}(x_2,y_1)$$
$$= x_1^{-a}x_2^{-a}y_1^{-b}y_2^{-b}(x_1^{-c\log y_1}x_2^{-c\log y_2} - x_1^{-c\log y_2}x_2^{-c\log y_1}) \leq 0 \tag{24}$$

for all $x_1 < x_2$ and $y_1 < y_2$. Recall that X and Y are positive(negative) quadrant dependent- *PQD (NQD)* if and only if

$$\bar{F}(x,y) \geq (\leq)\bar{F}(x,0)\bar{F}(0,y)$$

and that $\bar{F}(x,y)$ is RR_2 implies *NQD*. Thus $\bar{F}_4(x,y)$ possess negative dependence.

In the case of $\bar{F}_5(x,y)$, it is TP_2 since the sign of the expressions on the left of (24) with respect to $\bar{F}_5(x,y)$ depends on $(2x_2^\alpha + x_1^\alpha)(y_2^\beta - y_1^\beta)$ which is positive for $y_1 < y_2$. We conclude that $\bar{F}_5(x,y)$ has positive dependence through PQD and further that this implies positive correlation.

While considering the nature of dependence in $\bar{F}_7(x,y)$ we note that

$$\bar{F}_7(x_1,y_1)\bar{F}_7(x_2,y_2) - \bar{F}_7(x_1,y_2)\bar{F}_7(x_2,y_1)$$
$$= (x_1^a - x_2^a)(y_2^b - y_1^b)\bar{F}_7(x_1,y_1)\bar{F}_7(x_2,y_2)\bar{F}_7(x_1,y_2)\bar{F}_7(x_2,y_1)$$

which is non-positive for $x_1 < x_2$. Accordingly we see that $\bar{F}_7(x,y)$ is RR_2 with the consequent implication that the distribution is NQD and the associated random variables are negatively correlated. Similar calculations show that $\bar{F}_9(x,y)$ is PQD and hence the corresponding variables are positively correlated.

A more interesting result emerges for the Weibull based bivariate distribution $\bar{F}_8(x,y)$. The TP_2 nature of the survival function depends on α. For example $\alpha = \frac{1}{2}$, $\bar{F}_8(x,y)$ is RR_2 and for $\alpha = 2$, $\bar{F}_8(x,y)$ is TP_2. Accordingly the distribution can be NQD or PQD depending on α. This means that the random variables X and Y can have negative as well as positive correlation depending on α.

4.3. Dependence Functions

Among the various dependence functions available in literature we choose the Clayton function (Clayton (1978)), which seems to be more popular. It is defined as

$$\theta(x,y) = \frac{\bar{F}(x,y)\frac{\partial^2 \bar{F}}{\partial x \partial y}}{\frac{\partial \bar{F}}{\partial x}\frac{\partial \bar{F}}{\partial y}}$$

The interpretation of $\theta(x,y)$ is that when X and Y are positively (negatively) associated $\theta(x,y) > (<)1$ and $\theta(x,y) = 1$ implies independence of X and Y. For detailed study of interpretation, properties and applications of the measure we refer to Oakes (1989), Anderson *et al.* (1992), Gupta (2003) and Nair and Sankaran (2014).

The values of $\theta(x,y)$ and the nature of dependence for various models are presented in Table 4. Other dependence functions mentioned in Nair and Sankaran (2010) can be obtained in closed forms for certain bivariate Pareto models. As the nature of dependence is similar to the one based on $\theta(x,y)$, we do not present details on the dependence using other functions.

TABLE 4
Clayton measure for bivariate Pareto models

Distribution	$\theta(x,y)$	dependence
$\bar{F}_1(x,y)$	1	independent
$\bar{F}_2(x,y)$	$1+\alpha$	positive
$\bar{F}_3(x,y)$	$1+\frac{1}{\theta}$	positive
$\bar{F}_4(x,y)$	$1-\frac{c}{(a+c\log y)(b+c\log x)}$	negative
$\bar{F}_5(x,y)$	$1+\frac{1}{\sigma(1+x)^{\alpha}(1+y)^{\beta}}$	positive
$\bar{F}_7(x,y)$	$\frac{4x^a y^b-2x^a-2y^b}{1+4x^a y^b-2x^a-2y^b}$	negative

TABLE 5
American football league data

Sl.No.	X	Y	Sl.No.	X	Y
1	2.05	3.98	22	10.85	38.07
2	7.78	7.78	23	0.85	0.85
3	7.23	9.68	24	7.05	7.05
4	31.13	49.88	25	32.45	42.35
5	7.25	7.25	26	5.78	25.98
6	4.22	9.48	27	1.65	1.65
7	6.42	6.42	28	2.90	2.90
8	10.40	14.25	29	10.15	10.15
9	11.63	17.37	30	3.88	6.43
10	14.58	14.58	31	10.35	10.35
11	17.83	17.83	32	5.52	11.27
12	9.05	9.05	33	3.43	3.43
13	10.57	14.28	34	2.58	2.58
14	6.85	34.58	35	8.53	14.57
15	14.58	20.57	36	13.80	49.75
16	4.25	4.25	37	6.42	15.08
17	15.53	15.53	38	7.02	7.02
18	8.98	8.98	39	8.87	8.87
19	2.98	2.98	40	0.75	0.75
20	1.38	1.38	41	12.13	12.13
21	11.82	11.82	42	19.65	10.70

TABLE 6
Failure time data (transformed)

Sl.No.	X	Y	Sl.No.	X	Y
1	6.96	1.45	11	16.12	1366.49
2	1702.75	1.06	12	2.34	16.61
3	1.15	1.22	13	4914.77	62.18
4	8.50	5.05	14	412503.51	290.03
5	6.75	298.87	15	86.49	2.61
6	3751.83	9.49	16	33.12	1286.91
7	4.06	12.18	17	321980003.30	1.38
8	2.20	11.47	18	145.47	1510.20
9	2.51	1.13	19	29436.77	13.20
10	2.08	2.20	20	9.21	5.64

5. Inference and Data Analysis

The estimators of parameters of the models belonging to the family (7) can be generally derived using the method of maximum likelihood. When the number of parameters is not large, one can easily get estimates by solving likelihood equation. If the model involves more than three parameters, as in the case of (17), we need to solve a four dimensional optimization problem which may not give unique solutions. Alternatively one can use a computationally efficient two stage estimation procedure as suggested by Xu (1996), see also Joe (1997), Joe (2005), in this respect. In the two stage estimation procedure, the first stage involves the maximum likelihood estimation from univariate marginals and the second stage involves the maximum likelihood estimation of the dependent parameters keeping the univariate parameters held fixed obtained from the first stage. It is proved that the estimators so obtained satisfy large sample properties of the maximum likelihood estimators (MLE).

We now apply the proposed family of distributions to two real life data sets. We first apply the model (16) to the American football league data obtained from the matches played on three consecutive week ends in 1986. The data were first published in 'Washington Post'and they are also available in Csörgă and Welsh (1989).

It is a bivariate data set and the variables X_1 and X_2 are as follows; X_1 represents the game time to the first points scored by kicking the ball between goal posts and X_2 represents the game time to the first points scored by moving the ball into the end zone. These times are of interest to a casual spectator who wants to know how long one has to wait to watch a touchdown or to a spectator who is interested only at the beginning stages of a game. The data was first analyzed by Csörgă and Welsh (1989), by converting the seconds to decimal minutes i.e 2:03 has been converted to 2.05. We have adopted the same procedure. The data are presented in Table 5. We use exponential transformation to the data points to make observations larger than one. The maximum likelihood estimates of the parameters of the model (16) are obtained as $\hat{a} = 0.1128$, $\hat{b} = 0.0750$, $\hat{p} = 0.991$

and $\hat{\lambda} = 67.913$, The variables X and Y are positively correlated.

To test the goodness of fit, we use the bivariate version of Kolmogrov-Smirnov(K-S) test given in Justel *et al.* (1997). The K-S statistic values are $D_1 = 0.1976, D_2 = 0.2085, D_3 = 0.0474, D_4 = 0.0237$ and $D_5 = 0.0183$. Thus $D^* = Max(D_1, D_2, D_3, D_4, D_5) = 0.2085$. The above value is less than the value 0.2517 at 20^{th} percentile so that the model (16) is an appropriate fit for the given data.

The second data set is taken from Kim and Kvam (2004), which consists of the failure times of 20 sample units from a system consisting of three components. We use failure times of first two components. Since some data values are smaller than one, we make exponential transformation, so that all the observations have values larger than one. The transformed data are given in Table 6. The model (9) is applied to the data. The method of maximum likelihood provides the estimates of the parameters as $\hat{a} = 0.2088, \hat{b} = 0.3573$ and $\hat{\alpha} = 0.0505$.

The goodness of fit of Justel *et al.* (1997) is applied and the test statistic value $D^* = 0.20939$. This value is less than the value 0.2922 at 25^{th} percentile, we conclude the model (9) is a goodfit for the given data set.

6. Conclusion

In this paper we have introduced a class of bivariate Pareto distributions and studied the distributional properties of the class. The class includes several well known bivariate Pareto distributions. It also contains distributions having positive as well as negative correlations among variables. The dependence structure of the class of distributions were discussed. The proposed class of distributions was applied to real life data set.

The referee points out that the density functions and therefore their likelihoods cannot be referred to a unique functional for the whole family. So it does not make immediate the search for the distribution law that adjusts in the best way the available empirical data. To solve this problem we have developed in a companion paper characterstic properties of the class of distributions using verifiable functional forms of different concepts such as bivariate versions of dullness property, mean residual income, income gap ratio etc. The analysis of dependence structure using copula is another area which is being worked. The work in this direction will be reported in a seperate paper.

ACKNOWLEDGEMENTS

We thank the referee and the editor for the constructive comments. The third author thanks Department of Science and Technology, Government of India for providing financial support.

REFERENCES

M. Abramowitz, I. A. Stegun (1966). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* US Government Printing Office.

J. E. Anderson, T. A. Louis, N. V. Holm, B. Harvald (1992). *Time-dependent association measures for bivariate survival distributions.* Journal of the American Statistical Association, 87, pp. 641–650.

B. C. Arnold (1985). *Pareto Distribution.* Wiley, New York.

B. C. Arnold (1990). *A flexible family of multivariate Pareto distributions.* Journal of Statistical Planning and Inference, 24, pp. 249–258.

B. C. Arnold (1992). *Conditionally Specified Distributions.* Springer, New York.

N. Balakrishnan, C. Lai (2009). *Continuous Bivariate Distributions.* Springer, New York.

R. E. Barlow, M. B. Mendel (1992). *De finetti-type representations for life distributions.* Journal of the American Statistical Association, 87, pp. 1116–1122.

R. E. Barlow, F. Proschan (1975). *Statistical Theory of Reliability and Life Testing: Probability Models.* Holt, Rinehart and Winston,New York.

D. G. Clayton (1978). *A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence.* Biometrika, 65, pp. 141–151.

S. Csörgă, A. Welsh (1989). *Testing for exponential and Marshall–Olkin distributions.* Journal of Statistical Planning and Inference, 23, pp. 287–300.

R. C. Gupta (2003). *On some association measures in bivariate distributions and their relationships.* Journal of Statistical Planning and Inference, 117, pp. 83–98.

H. Joe (1997). *Multivariate Models and Dependence Concepts.* CRC Press.

H. Joe (2005). *Asymptotic efficiency of the two-stage estimation method for copula-based models.* Journal of Multivariate Analysis, 94, pp. 401–419.

N. L. Johnson, S. Kotz, N. Balakrishnan (1994). *Continuous Univariate Distributions.* John Wiley and Sons, New York.

A. Justel, D. Peña, R. Zamar (1997). *A multivariate Kolmogorov-Smirnov test of goodness of fit.* Statistics and Probability Letters, 35, pp. 251–259.

Y. Kagan, Linnik, C. R. Rao, B. Ramachandran (1973). *Characterization Problems in Mathematical Statistics.* Wiley,New York.

W. Kibble (1941). *A two-variate gamma type distribution.* Sankhyā, Series A, 5, pp. 137–150.

H. Kim, P. H. Kvam (2004). *Reliability estimation based on system data with an unknown load share rule.* Lifetime Data Analysis, 10, pp. 83–94.

S. Kotz, N. Balakrishnan, N. L. Johnson (2002). *Continuous Multivariate Distributions*. John Wiley and Sons, New York.

H. Langseth (2002). *Bayesian networks with applications in reliability analysis*. Ph.D. thesis, Norwegian University of Science and Technology, Guntur, India.

D. V. Lindley, N. D. Singpurwalla (1986). *Multivariate distributions for the life lengths of components of a system sharing a common environment*. Journal of Applied Probability, pp. 418–431.

K. V. Mardia (1962). *Multivariate Pareto distributions*. The Annals of Mathematical Statistics, 33, pp. 1008–1015.

N. U. Nair, P. G. Sankaran (2010). *A new measure of association for bivariate survival data*. Journal of Statistical Planning and Inference, 140, pp. 2569–2581.

N. U. Nair, P. G. Sankaran (2014). *Characterizations and time-dependent association measures for bivariate schur-constant distributions*. Brazilian Journal of Probability and Statistics(to appear).

T. K. Nayak (1987). *Multivariate Lomax distribution: properties and usefulness in reliability theory*. Journal of Applied Probability, 24, pp. 170–177.

D. Oakes (1989). *Bivariate survival models induced by frailties*. Journal of the American Statistical Association, 84, pp. 487–493.

P. G. Sankaran, D. Kundu (2014). *A bivariate Pareto model, Statistics*. p. DOI:10.1080/02331888.2012.719521.

P. G. Sankaran, N. U. Nair (1993). *A bivariate Pareto model and its applications to reliability*. Naval Research Logistics, 40, pp. 1013–1020.

S. Talwalker (1980). *On the property of dullness of Pareto distribution*. Metrika, 27, pp. 115–119.

J. J. Xu (1996). *Statistical modelling and Inference for multivariate and longitudinal discrete response data*. Ph.D. thesis, Department of Statistics University of British Columbia, Canada.

SUMMARY

A family of bivariate Pareto distributions

Pareto distributions have been extensively used in literature for modelling and analysis of income and lifetime data. In the present paper, we introduce a family of bivariate Pareto distributions using a generalized version of dullness property. Some important bivariate Pareto distributions are derived as special cases. Distributional properties of the family are studied. The dependency structure of the family is investigated. Finally, the family of distributions is applied to two real life data situations.

Keywords: bivariate Pareto distribution; correlation coefficient; association measures; dullness property