Ann. Inst. Statist. Math.
Vol. 51, No. 2, 331–344 (1999)

BIVARIATE DISTRIBUTIONS WITH PEARSON TYPE VII CONDITIONALS

ATHANASIOS KOTTAS*, KONSTANTINOS ADAMIDIS AND SOTIRIOS LOUKAS

Department of Mathematics, University of Ioannina, 451 10 Ioannina, Greece

(Received July 8, 1997; revised November 26, 1997)

Abstract. In this article the most general class of bivariate distributions such that both conditional densities are Pearson Type VII, with fixed shape parameter, is fully characterized. Some of its properties and relations with other distributions are explored. The estimation of parameters is considered by the methods of maximum likelihood and pseudolikelihood and a method for random variate generation is presented along with a simulation experiment. Bivariate and multivariate extensions of the Pearson Type VII conditionals distribution are also discussed.

Key words and phrases: Characterizations, conditional distributions, functional equations, maximum likelihood estimation, maximum pseudolikelihood estimation, Pearson Type VII distribution.

1. Introduction

The study of bivariate distributions such that both conditional distributions are members of specific parametric families (i.e. conditionally specified distributions) has received a considerable amount of interest in the last decade. The motivation for conditional specification stems from the fact that it is usually easier to determine the forms of the conditionals rather than the joint or marginal distributions. Ground breaking contributions to the theory associated with conditional specifications are the papers by Castillo and Galambos (1987*a*, 1989), which pioneered the utility of functional equation arguments in specifying the complete class of bivariate distributions with normal conditionals. The recent monograph by Arnold *et al.* (1992) is an excellent reference for the work done on conditionally specified distributions and related topics (see also Arnold *et al.* (1993, 1995) and Sarabia (1995)).

This paper is concerned with bivariate distributions possessing conditional

* Now at Department of Statistics, University of Connecticut, U-120, Storrs, CT 06269-3120, U.S.A.

densities of the form

$$f_X(x) = \frac{\sqrt{\sigma}\Gamma(p)}{\sqrt{\pi}\Gamma\left(p-\frac{1}{2}\right)(1+\sigma x^2)^p}, \qquad x \in R, \tag{1.1}$$

which defines the Pearson Type VII distribution (PVII for brevity) with parameters $\sigma \in R_+$ and $p > 2^{-1}$; in the sequel we shall be referring to such distributions as bivariate distributions with PVII conditionals (BPVIIC for brevity). Note that when $p = 1$ the PVII becomes the Cauchy distribution with zero location parameter and by restricting $2p - 1 \in N$, (1.1) becomes the density function of a t distribution with $2p-1$ degrees of freedom and scale parameter σ^{-1}. After deriving the class of BPVIIC distributions with fixed shape parameter we study its properties and explore relations with other distributions. The estimation of parameters is considered by the methods of maximum likelihood and pseudolikelihood and their performances are compared by carrying out a simulation experiment. Finally, we discuss briefly the general case where the shape parameters are allowed to depend on the conditioned variable and sketch an extension to higher dimensions.

2. The BPVIIC distribution

2.1 *Feasible models with PVII conditionals*

Assume a pair of random variables (X, Y) with positive joint density $f_{X,Y}(x, y)$ over R^2 and denote by $f_X(x)$, $f_Y(y)$ and $f_{X|Y}(x \mid y)$, $f_{Y|X}(y \mid x)$ the associated marginal and conditional densities respectively. To characterize the wider class of BPVIIC distributions suppose that $X \mid Y$ and $Y \mid X$ are PVII variables with densities

$$\begin{aligned} f_{X|Y}(x \mid y) &= \frac{\sqrt{\sigma(y)}\Gamma(p)}{\sqrt{\pi}\Gamma\left(p-\frac{1}{2}\right)(1+\sigma(y)x^2)^p}, \qquad x \in R, \\ f_{Y|X}(y \mid x) &= \frac{\sqrt{\tau(x)}\Gamma(p)}{\sqrt{\pi}\Gamma\left(p-\frac{1}{2}\right)(1+\tau(x)y^2)^p}, \qquad y \in R, \end{aligned} \tag{2.1}$$

where $p > 2^{-1}$ is a fixed shape parameter and $\sigma(y)$, $\tau(x)$ are some positive functions with real arguments. Now write the joint density as the product of a marginal and a conditional density in both possible ways to obtain the functional equation

$$\frac{f_Y(y)\sqrt{\sigma(y)}}{(1+\sigma(y)x^2)^p} = \frac{f_X(x)\sqrt{\tau(x)}}{(1+\tau(x)y^2)^p}, \qquad x, y \in R, \tag{2.2}$$

and set

$$g(y) = \{f_Y(y)\sqrt{\sigma(y)}\}^{1/p}, \qquad h(x) = \{f_X(x)\sqrt{\tau(x)}\}^{1/p} \tag{2.3}$$

so that, after rearranging, (2.2) becomes

$$g(y) + y^2 g(y)\tau(x) - h(x) - x^2 h(x)\sigma(y) = 0, \tag{2.4}$$

which must be solved for σ, τ, g and h. It is readily recognized that the latter is a special case of the functional equation $\sum_{k=1}^{n} f_k(x)g_k(y) = 0$, whose most general solution is given in the classical book by Aczel ((1966), p. 161). Thus, with $h(x)$, $x^2h(x)$ and $g(y)$, $y^2g(y)$ being the systems of mutually linearly independent functions, the solution of (2.4) is found to be

$$\begin{aligned} \tau(x) &= \frac{\lambda_3 + \lambda_4 x^2}{\lambda_1 + \lambda_2 x^2}, & \sigma(y) &= \frac{\lambda_2 + \lambda_4 y^2}{\lambda_1 + \lambda_3 y^2}, \\ h(x) &= \frac{1}{\lambda_1 + \lambda_2 x^2}, & g(y) &= \frac{1}{\lambda_1 + \lambda_3 y^2}, \end{aligned} \tag{2.5}$$

for real constants λ_j, $j = 1, \dots, 4$. Finally, substituting (2.3) and (2.5) in (2.2) the joint density function is derived in the form

$$f_{X,Y}(x,y) = \frac{N_p(\lambda_1, \lambda_2, \lambda_3, \lambda_4)}{(\lambda_1 + \lambda_2 x^2 + \lambda_3 y^2 + \lambda_4 x^2 y^2)^p}, \qquad x, y \in R, \tag{2.6}$$

where $N_p(\cdot)$ denotes the normalizing constant and $\lambda_j \in R$, $j = 1, \dots, 4$.

In order to meet the non-negativity condition for $f_{X,Y}$ we must postulate that $\lambda_j \in R_+ \cup \{0\}$. Furthermore, for (2.6) to represent a well defined density function certain compatibility conditions, studied by Arnold and Press (1989), must be satisfied. Specifically, denoting by $S_{\mathcal{X}}$ and $S_{\mathcal{Y}}$ the supports of X and Y respectively, Arnold and Press proved that if (X, Y) is absolutely continuous with respect to some product measure $m_1 \times m_2$ on $S_{\mathcal{X}} \times S_{\mathcal{Y}}$ then a joint density with conditionals $f_{X|Y}$ and $f_{Y|X}$ will exist iff

(i) $\{(x,y) : f_{X|Y}(x \mid y) > 0\} = \{(x,y) : f_{Y|X}(y \mid x) > 0\} = T$, and

(ii) $\forall (x,y) \in T, \ni a(x), b(y) : f_{X|Y}/f_{Y|X} = a(x)b(y)$, where $\int a(x)dm_1(x) < \infty$.

Obviously, in our case, the supports of the conditional densities coincide. The equality in (ii) holds with $a(x) = (\lambda_3 + \lambda_4 x^2)^{-1/2}(\lambda_1 + \lambda_2 x^2)^{1/2-p}$ and finally, to ensure that $a(x)$ is integrable we must have $\lambda_1 \in R_+ \cup \{0\}$ and $\lambda_j \in R_+$, $j = 2, 3, 4$; moreover if $\lambda_1 = 0$ then $p \in (2^{-1}, 1)$.

2.2 *The normalizing constant*

The final step in the characterization of the BPVIIC distribution involves the determination of the constant of proportionality in (2.6). To this end we must evaluate the integral

$$\{N_p(\lambda_1, \lambda_2, \lambda_3, \lambda_4)\}^{-1} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\lambda_1 + \lambda_2 x^2 + \lambda_3 y^2 + \lambda_4 x^2 y^2)^{-p} dx dy. \tag{2.7}$$

We treat the case $\lambda_1 \neq 0$ first. Making the transformation $s = \lambda_2 \lambda_1^{-1} x^2$, $t = \lambda_3 \lambda_1^{-1} y^2$, calling $\varphi = \lambda_1 \lambda_4 (\lambda_2 \lambda_3)^{-1}$ and using the well known integral representation of the Beta function, $B(m,n) = \int_0^\infty x^{m-1}(1+x)^{-m-n} dx$, for $m, n \in R_+$, we obtain

$$\{N_p(\lambda_1,\lambda_2,\lambda_3,\lambda_4)\}^{-1} = \frac{B\left(\frac{1}{2}, p-\frac{1}{2}\right)}{\lambda_1^{p-1}\sqrt{\lambda_2\lambda_3}} \int_0^\infty \frac{dt}{(1+t)^{p-1/2}\sqrt{t(1+\varphi t)}}. \tag{2.8}$$

Letting $\omega = t(1+t)^{-1}$ and manipulating we end up with

$$N_p(\lambda_1,\lambda_2,\lambda_3,\lambda_4) = \frac{\lambda_1^{p-1}\sqrt{\lambda_2\lambda_3}}{B\left(\frac{1}{2}, p-\frac{1}{2}\right) I\left(\frac{1}{2}, \frac{1}{2}, p; 1-\varphi\right)}, \tag{2.9}$$

where $\lambda_j \in R_+$, $j = 1,\dots,4$, $\varphi = \lambda_1\lambda_4(\lambda_2\lambda_3)^{-1}$, $p > 2^{-1}$ and

$$I(a,b,c;z) = \int_0^1 \omega^{b-1}(1-\omega)^{c-b-1}(1-z\omega)^{-a}d\omega,$$

for $c > b \in R_+$. The latter integral converges for $z < 1$ and it is fairly well known from Euler's formula ${}_2F_1(a,b;c;z) = \Gamma(c)\{\Gamma(b)\Gamma(c-b)\}^{-1}I(a,b,c;z)$ which extends the definition of the hypergeometric function ${}_2F_1(a,b;c;z)$ beyond $|z| < 1$ (see Magnus *et al.* (1966), p. 54). Alternatively, the normalizing constant can be expressed in terms of the usual infinite series representation of ${}_2F_1(a,b;c;z)$. Making in (2.8) the changes of variables $\omega = t(1+t)^{-1}$ and $\omega = (1+t)^{-1}$ in turn and placing appropriate restrictions on φ to ensure that the relevant hypergeometric series converge, we find

$$N_p(\lambda_1,\lambda_2,\lambda_3,\lambda_4) = \begin{cases} \dfrac{\lambda_1^{p-1}\sqrt{\lambda_2\lambda_3}}{{}_2F_1\left(\frac{1}{2},\frac{1}{2};p;1-\varphi\right)B^2\left(\frac{1}{2},p-\frac{1}{2}\right)}, \\ \qquad 0 < \varphi \le \dfrac{1}{2}, \\ \dfrac{\lambda_1^{p-1/2}\sqrt{\pi\lambda_4}}{{}_2F_1\left(\frac{1}{2},p-\frac{1}{2};p;1-\frac{1}{\varphi}\right)\Gamma\left(p-\frac{1}{2}\right)B^2\left(\frac{1}{2},p-\frac{1}{2}\right)}, \\ \qquad \varphi > \dfrac{1}{2}. \end{cases} \tag{2.10}$$

The advantage of (2.9) over the latter expression becomes evident when the process of fitting the model to data is encountered. It is useful to note that for two values of p the formula for the normalizing constant in (2.9) may be restated in terms of elliptic integrals using the expressions

$$\begin{aligned} R_F(a,b,c) &= \frac{1}{2}\int_0^\infty \{(t+a)(t+b)(t+c)\}^{-1/2}dt, \\ R_D(a,b,c) &= \frac{3}{2}\int_0^\infty \{(t+a)(t+b)(t+c)^3\}^{-1/2}dt, \end{aligned} \tag{2.11}$$

obtained from Carlson ((1977), Chapter 9) for the elliptic integrals of the first and second kind respectively. For the former integral a, b, c must be non negative and at most one may be equal to zero, while for the latter, c must be positive, a, b non negative and only one of a and b may be allowed to take the value of zero. The expressions in (2.11) are symmetrised variants of the classic elliptic integrals with special advantages in computations and they are available in standard software such as NAG (Numerical Algorithms Group, 1984). For $p = 1$ and 2 the integrands in (2.8) (or 2.9) can be easily rearranged to accomplish the forms required by (2.11) giving the normalizing constants as

$$N_p(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = \begin{cases} \sqrt{\lambda_1\lambda_4}\{2\pi R_F(0, \varphi^{-1}, 1)\}^{-1}, & p = 1, \\ 3\sqrt{\lambda_1^3\lambda_4}\{\pi R_D(0, \varphi^{-1}, 1)\}^{-1}, & p = 2. \end{cases} \tag{2.12}$$

The evaluation of the integral in (2.7) when $\lambda_1 = 0$ is dealt with arguments analogous to those leading to (2.9) and subsequently yields

$$N_p(\lambda_2, \lambda_3, \lambda_4) = \frac{(\lambda_2\lambda_3)^{p-1/2}\lambda_4^{1-p}}{B\left(\frac{1}{2}, p - \frac{1}{2}\right) B\left(1 - p, p - \frac{1}{2}\right)}, \tag{2.13}$$

with $\lambda_j \in R_+$ for $j = 2, 3, 4$ and $p \in (2^{-1}, 1)$.

Summarizing, we have obtained that for fixed p the most general BPVIIC distribution has density function given by (2.6) with normalizing constant given by (2.9) or (2.13) for $p > 2^{-1}$, $\lambda_1 \in R_+ \cup \{0\}$, $\lambda_j \in R_+$, $j = 2, 3, 4$ and the further restriction $p < 1$ when $\lambda_1 = 0$.

3. Properties of the distribution

3.1 *Conditionals, marginals and moments*

Hereinafter we shall concentrate on the case where $p \geq 1$ (which implies that $\lambda_1 > 0$) and for simplicity move to the parameterization $\mu_j = \lambda_{j+1}\lambda_1^{-1}$, $j = 1, 2$. Consequently (2.6) assumes the form

$$f_{X,Y}(x, y) = N_p(\mu_1, \mu_2, \varphi)(1 + \mu_1 x^2 + \mu_2 y^2 + \varphi\mu_1\mu_2 x^2 y^2)^{-p}, \qquad x, y \in R, \tag{3.1}$$

where

$$N_p(\mu_1, \mu_2, \varphi) = \frac{\sqrt{\mu_1\mu_2}}{B\left(\frac{1}{2}, p - \frac{1}{2}\right) I\left(\frac{1}{2}, \frac{1}{2}, p; 1 - \varphi\right)},$$

with $p \geq 1$ and $\mu_j, \varphi \in R_+$ for $j = 1, 2$; this is denoted by $(X, Y) \sim$ BPVIIC$(\mu_1, \mu_2, \varphi; p)$. From (2.1) and (2.5) we immediately have that

$$\begin{aligned} X \mid Y &\sim \text{PVII}\{\mu_1(1 + \varphi\mu_2 y^2)(1 + \mu_2 y^2)^{-1}, p\}, \\ Y \mid X &\sim \text{PVII}\{\mu_2(1 + \varphi\mu_1 x^2)(1 + \mu_1 x^2)^{-1}, p\}, \end{aligned} \tag{3.2}$$

with densities given using (1.1) and subsequently the marginal densities are

$$f_X(x) = \sqrt{\mu_1}\left\{I\left(\frac{1}{2},\frac{1}{2},p;1-\varphi\right)(1+\varphi\mu_1 x^2)^{1/2}(1+\mu_1 x^2)^{p-1/2}\right\}^{-1}, \quad x \in R,$$

$$f_Y(y) = \sqrt{\mu_2}\left\{I\left(\frac{1}{2},\frac{1}{2},p;1-\varphi\right)(1+\varphi\mu_2 y^2)^{1/2}(1+\mu_2 y^2)^{p-1/2}\right\}^{-1}, \quad y \in R.$$

In this formulation μ_1 and μ_2 are intensity parameters for X and Y respectively (μ_1^{-1} and μ_2^{-1} are scale parameters) and φ, p are dependence and shape parameters respectively. It is easy to prove that X and Y are independent if and only if $\varphi = 1$ and furthermore this is also the unique case for encountering PVII marginals. The graph of the joint density is always a symmetrical bell-shaped curve and for $\mu_1 = \mu_2 = 1$ we get the standard form of the distribution.

When $p \in (2^{-1}, 1]$ the BPVIIC distribution does not possess finite moments. If on the contrary $p > 1$, the raw moments of the pair (X,Y) may be determined from (3.1) by direct integration in a manner similar to that used for obtaining the normalizing constant. For $p > \max\{\frac{k+1}{2}, \frac{r+1}{2}\}$ and $k, r \in N \cup \{0\}$ we find that

$$E(X^kY^r) = \begin{cases} \dfrac{\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(p-\frac{k+1}{2}\right)I\left(\frac{k+1}{2},\frac{r+1}{2},p;1-\varphi\right)}{\sqrt{\pi\mu_1^k\mu_2^r}\,\Gamma\left(p-\frac{1}{2}\right)I\left(\frac{1}{2},\frac{1}{2},p;1-\varphi\right)}, & \\ \quad k,\ r \text{ both even or zero}, & \\ 0, \quad \text{at least one of } k,\ r \text{ is odd}. & \end{cases} \tag{3.3}$$

Accordingly, the BPVIIC model may be an appropriate candidate for uncorrelated but non independent bivariate data. A set of real data arising from such situation is analyzed in Arnold and Strauss (1991*a*) and also discussed in Sarabia (1995).

The conditional moments are easily derived using (3.2) and (1.1). For even $k < 2p-1$ we obtain

$$E(X^k \mid y) = \frac{\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(p-\frac{k+1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(p-\frac{1}{2}\right)}\left(\frac{1+\mu_2 y^2}{\mu_1(1+\varphi\mu_2 y^2)}\right)^{k/2},$$

$$E(Y^k \mid x) = \frac{\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(p-\frac{k+1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(p-\frac{1}{2}\right)}\left(\frac{1+\mu_1 x^2}{\mu_2(1+\varphi\mu_1 x^2)}\right)^{k/2},$$

while for k odd the above moments equal zero. We note that the conditional moments are rational functions of the conditioned variable.

3.2 *Relations to other distributions*

From (3.1) with $p = 1$ we get the centered Cauchy conditionals model of Anderson and Arnold (1991). The equivalence of their expression (2.21), which gives the normalizing constant in terms of complete elliptic integrals of the first kind $K(m)$, with (2.10) can be seen through the formula ${}_2F_1(\frac{1}{2}, \frac{1}{2}; 1; m) = \frac{2}{\pi}K(m)$, $m \in R_+ \cup \{0\}$ (see Abramowitz and Stegun (1968), p. 591). Of course (3.1), with the normalizing constant given alternatively by the first expression in (2.12), gives the joint density in a compact form without having to distinguish two cases for φ.

An interesting limiting case of the BPVIIC$(\mu_1, \mu_2, \varphi; p)$ occurs when $\varphi \to 0$. Let $h(x, y) = \lim_{\varphi \to 0} f_{X,Y}(x, y)$. Then it is straightforward to verify that

$$h(x, y) = \pi^{-1}(p-1)\sqrt{\mu_1\mu_2}(1 + \mu_1 x^2 + \mu_2 y^2)^{-p}, \qquad x, y \in R, \tag{3.4}$$

with $p > 1$ and $\mu_1, \mu_2 \in R_+$. This is a special case of the bivariate Pearson Type VII distribution (Johnson (1987), p. 117) with location parameters equal to zero and uncorrelated components. If we restrict $2p - 2$ to take positive integer values $h(x, y)$ becomes a special case of the general bivariate t distribution (Johnson and Kotz (1972), p. 134, relation 1) with uncorrelated components and $2p - 2$ degrees of freedom. For $\mu_1 = \mu_2 = c^{-2}$ and $p = 3/2$ in (3.4) we get the bivariate Cauchy distribution (Mardia (1970), p. 86), while for $\mu_1 = \mu_2 = \nu^{-1}$ and $p = (\nu + 2)/2$ we obtain the bivariate Student distribution (Johnson and Kotz (1972), p. 134, relation 2) with ν degrees of freedom. In the latter case we get the standard bivariate normal distribution with independent components as a further limiting case,

$$\lim_{\nu \to \infty} \lim_{\varphi \to 0} f_{X,Y}(x, y) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(x^2 + y^2)\right\}, \qquad x, y \in R.$$

Another limiting case of the BPVIIC$(\mu_1, \mu_2, \varphi; p)$ is the centered normal conditionals distribution, studied in detail by Sarabia (1995). Letting $\alpha = 2p\mu_1$, $\beta = 2p\mu_2$, $\gamma = \varphi(2p)^{-1}$ in (3.1) and using that $\lim_{z\to\infty} e^{-a \log z}\Gamma(z + a)\{\Gamma(z)\}^{-1} = 1$, $\lim_{c\to\infty} {}_2F_1(a, b; c; 1 - cz^{-1}) = z^a U(a, a - b + 1, z)$ (see Magnus *et al.* (1966), pp. 12 and 263 respectively) where $U(\cdot)$ is a confluent hypergeometric function, we obtain

$$\begin{aligned}&\lim_{p\to\infty} f_{X,Y}(x, y)\\ &\quad = \sqrt{2\alpha\beta\gamma}\left(2\pi U\left(\frac{1}{2}, 1, \frac{1}{2\gamma}\right)\right)^{-1} \exp\left\{-\frac{1}{2}(\alpha x^2 + \beta y^2 + \alpha\beta\gamma x^2 y^2)\right\},\end{aligned}$$

which is the density function of the centered normal conditionals model with $\alpha, \beta, \gamma \in R_+$.

The BPVIIC distribution can be related to a special case of the bivariate second kind Beta conditionals of Castillo and Sarabia (1990), by means of the change of variables $(U, V) = (X^2, Y^2)$. Specifically if $(X, Y) \sim$ BPVIIC$(\lambda_j, j = 1, \ldots, 4; p)$ with density given by (2.6) and normalizing constant as in (2.9) or (2.13) then using their notation $(X^2, Y^2) \sim$ BE$(\lambda_1, \lambda_2, \lambda_3, \lambda_4; \frac{1}{2}, p - \frac{1}{2})$.

4. Estimation of parameters

4.1 *Maximum likelihood and pseudolikelihood estimation*

We assume that data $Y_{obs} = (x_i, y_i; i = 1, \ldots, n)$ are observed and are known to be i.i.d. from the BPVIIC$(\mu_1, \mu_2, \varphi; p)$ distribution with density function (3.1). From (3.3) it is evident that moment estimates for the parameters cannot be obtained in closed forms and therefore there is little point in considering the method any further. Proceeding with the method of maximum likelihood (ml), the log likelihood function from the n pairs of observations is given by

$$\begin{aligned} l(\mu_1, \mu_2, \varphi, p; Y_{obs}) = & -n\left\{\log\sqrt{\pi} + \log\Gamma\left(p - \frac{1}{2}\right) - \log\Gamma(p)\right\} \\ & + \frac{n}{2}(\log\mu_1 + \log\mu_2) - n\log I\left(\frac{1}{2}, \frac{1}{2}, p; 1 - \varphi\right) \\ & - p\sum_{i=1}^{n}\log(1 + \mu_1 x_i^2 + \mu_2 y_i^2 + \varphi\mu_1\mu_2 x_i^2 y_i^2). \end{aligned}$$

To avoid additional complexities in an already burdensome estimation problem we assume that p is known and consequently the ml equations are found to be

$$\frac{n}{\mu_1} = 2p\sum_{i=1}^{n}\frac{x_i^2 + \varphi\mu_2 x_i^2 y_i^2}{1 + \mu_1 x_i^2 + \mu_2 y_i^2 + \varphi\mu_1\mu_2 x_i^2 y_i^2},$$

$$\frac{n}{\mu_2} = 2p\sum_{i=1}^{n}\frac{y_i^2 + \varphi\mu_1 x_i^2 y_i^2}{1 + \mu_1 x_i^2 + \mu_2 y_i^2 + \varphi\mu_1\mu_2 x_i^2 y_i^2},$$

$$\frac{nI\left(\frac{3}{2}, \frac{3}{2}, p+1; 1 - \varphi\right)}{I\left(\frac{1}{2}, \frac{1}{2}, p; 1 - \varphi\right)} = 2\mu_1\mu_2 p\sum_{i=1}^{n}\frac{x_i^2 y_i^2}{1 + \mu_1 x_i^2 + \mu_2 y_i^2 + \varphi\mu_1\mu_2 x_i^2 y_i^2},$$

where we have used that $\frac{d}{dz}I(a, b, c; z) = aI(a+1, b+1, c+1; z)$ for the integral in (2.9). The ml estimates of μ_1, μ_2 and φ must be derived numerically and the evaluation of the integral $I(\cdot)$ in each iteration imposes a further drawback. However, the presence of an awkward normalizing constant is a common problem in conditionally specified models.

One alternative estimation scheme is provided by maximum pseudo likelihood (mpsl) estimation, which was introduced by Besag (1975, 1977). Arnold and Strauss (1991*b*) demonstrated that, under standard regularity conditions, the mpsl estimators are consistent and asymptotically normal but are not, in general, efficient. They also gave some interesting examples of the use of mpsl estimators and derived expressions for their asymptotic variances and covariances; see also Arnold *et al.* ((1992), Chapter 9). The estimation scheme for conditionally specified models is based on the maximization of the product of conditional densities and hence the problem is tackled without call to the normalizing constant. When the sampling distribution is the BVPVIIC$(\mu_1, \mu_2, \varphi; p)$, the required conditional densities are given using (3.2) and (1.1) and therefore we have the pseudo log likelihood function in the form

$$\begin{aligned}
l_{ps}(\mu_1, \mu_2, \varphi, p; Y_{obs}) &= \prod_{i=1}^{n} f_{X_i|Y_i}(x_i \mid y_i) f_{Y_i|X_i}(y_i \mid x_i) \\
&= -2n\left\{\log\sqrt{\pi} + \log\Gamma\left(p - \frac{1}{2}\right) - \log\Gamma(p)\right\} \\
&\quad + \frac{n}{2}(\log\mu_1 + \log\mu_2) \\
&\quad + \frac{1}{2}\sum_{i=1}^{n}\log(1 + \varphi\mu_2 y_i^2) + \frac{1}{2}\sum_{i=1}^{n}\log(1 + \varphi\mu_1 x_i^2) \\
&\quad + \left(p - \frac{1}{2}\right)\left\{\sum_{i=1}^{n}\log(1 + \mu_2 y_i^2) + \sum_{i=1}^{n}\log(1 + \mu_1 x_i^2)\right\} \\
&\quad - 2p\sum_{i=1}^{n}\log(1 + \mu_1 x_i^2 + \mu_2 y_i^2 + \varphi\mu_1\mu_2 x_i^2 y_i^2).
\end{aligned}$$

Differentiating with respect to μ_1, μ_2, φ and p we get the mpsl equations in the form,

$$\begin{aligned}
\frac{n}{\mu_1} &= -\sum_{i=1}^{n}\frac{\varphi x_i^2}{1 + \varphi\mu_1 x_i^2} - (2p - 1)\sum_{i=1}^{n}\frac{x_i^2}{1 + \mu_1 x_i^2} \\
&\quad + 4p\sum_{i=1}^{n}\frac{x_i^2 + \varphi\mu_2 x_i^2 y_i^2}{1 + \mu_1 x_i^2 + \mu_2 y_i^2 + \varphi\mu_1\mu_2 x_i^2 y_i^2}, \\
\frac{n}{\mu_2} &= -\sum_{i=1}^{n}\frac{\varphi y_i^2}{1 + \varphi\mu_2 y_i^2} - (2p - 1)\sum_{i=1}^{n}\frac{y_i^2}{1 + \mu_2 y_i^2} \\
&\quad + 4p\sum_{i=1}^{n}\frac{y_i^2 + \varphi\mu_1 x_i^2 y_i^2}{1 + \mu_1 x_i^2 + \mu_2 y_i^2 + \varphi\mu_1\mu_2 x_i^2 y_i^2},
\end{aligned}$$

$$0 = \sum_{i=1}^{n}\frac{\mu_1 x_i^2}{1 + \varphi\mu_1 x_i^2} + \sum_{i=1}^{n}\frac{\mu_2 y_i^2}{1 + \varphi\mu_2 y_i^2} - 4\mu_1\mu_2 p\sum_{i=1}^{n}\frac{x_i^2 y_i^2}{1 + \mu_1 x_i^2 + \mu_2 y_i^2 + \varphi\mu_1\mu_2 x_i^2 y_i^2},$$

$$\begin{aligned}
0 &= 2n\psi(p) - 2n\psi\left(p - \frac{1}{2}\right) + \sum_{i=1}^{n}\log(1 + \mu_2 y_i^2) + \sum_{i=1}^{n}\log(1 + \mu_1 x_i^2) \\
&\quad - 2\sum_{i=1}^{n}\log(1 + \mu_1 x_i^2 + \mu_2 y_i^2 + \varphi\mu_1\mu_2 x_i^2 y_i^2),
\end{aligned}$$

where $\psi(\cdot)$ is the digamma function.

Table 1. A simulation experiment: the upper and lower figures in each pair are the maximum likelihood and pseudolikelihood estimators respectively. Each is the average of 100 replications and the figures in parentheses are the root-mean-squared deviations of the 100 estimates from the true value.

True values	$n = 50$	$n = 100$	$n = 500$
$\mu_1 = 1$	1.138(.437)	1.055(.320)	1.009(.115)
	1.153(.462)	1.063(.325)	1.010(.115)
$\mu_2 = 1$	1.121(.340)	1.042(.291)	1.017(.133)
	1.135(.417)	1.050(.297)	1.018(.134)
$\varphi = 0.5$	.738(1.079)	.689(.554)	.551(.224)
	.730(1.113)	.682(.563)	.549(.223)
$\mu_1 = 1$	1.113(.373)	1.051(.276)	1.001(.123)
	1.112(.378)	1.052(.279)	1.001(.123)
$\mu_2 = 1$	1.128(.413)	1.075(.313)	1.019(.131)
	1.127(.416)	1.075(.316)	1.019(.131)
$\varphi = 1$	1.429(2.032)	1.201(.838)	1.103(.430)
	1.574(2.663)	1.208(.855)	1.104(.433)
$\mu_1 = 1$	1.127(.401)	1.056(.298)	1.001(.126)
	1.123(.407)	1.052(.301)	1.000(.127)
$\mu_2 = 1$	1.127(.446)	1.075(.315)	1.023(.137)
	1.120(.453)	1.072(.320)	1.022(.137)
$\varphi = 1.5$	2.037(3.064)	1.846(1.585)	1.594(.581)
	2.284(3.421)	1.928(1.904)	1.601(.592)
$\mu_1 = 1$	1.147(.502)	1.057(.299)	1.000(.131)
	1.146(.531)	1.050(.298)	1.000(.131)
$\mu_2 = 1$	1.128(.500)	1.090(.320)	1.022(.129)
	1.125(.516)	1.084(.317)	1.020(.129)
$\varphi = 2$	2.888(3.840)	2.268(1.518)	2.102(.712)
	3.357(4.010)	2.341(1.695)	2.114(.721)

4.2 *Simulation*

In order to allow a comparison of the ml and mpsl estimators a simulation experiment was conducted as follows. We generated 100 samples of sizes $n = 50$, 100 and 500, each randomly sampled from a BPVIIC distribution with density given by (3.1) for $p = 2$, $\mu_1 = \mu_2 = 1$ and each of four values of $\varphi = 0.5(0.5)2$.

The generations were accomplished using a rejection scheme which required only the kernel of the BPVIIC density and therefore circumvented the lack of analytic expression for the constant of proportionality. Briefly, denote by $g_1(x, y)$ the kernel of the density in (3.1) and by $g_2(x, y) = (1 + \mu_1 x^2 + \mu_2 y^2)^{-p}$ the kernel of the density $h(x, y)$ of the uncorrelated t distribution given by (3.4). Then, to generate a BPVIIC pair (x, y) we sample from $h(x, y)$ and accept it with probability $g_1(x, y)/g_2(x, y)$.

For each sample we found the ml and mpsl estimates of μ_1, μ_2 and φ using the elliptic integral representation for the normalizing constant, given by the second expression in (2.11) and appropriate routines from the NAG Library. The results are reported in Table 1 which gives the averages of the 100 ml and mpsl estimators together with their root-mean squared deviations from the actual parameter value. It appears that both estimators behaved very much the same. The methods tended to overestimate the parameters and the estimators improved as the sample size increased, in the sense that the estimates were generally closer to actual values and the spread decreased. The estimators of μ_1, μ_2 seemed tolerable even for $n = 50$ and steadily improving, as n increased, to rather satisfactory values. Both bias and spread were unpleasantly serious for φ when $n = 50$; although such sample size is rather unrealistically small for bivariate data. However, they drastically decreased as n increased to 100 and seemed to mitigate, reaching acceptable values for $n = 500$. We also increased the number of replications to 200 and 500, and repeated the experiment to investigate for possible significant changes in the behaviours of the two methods. However, apart from the expected reductions in bias and spread, the results of both methods persisted to remain close and therefore the contents of Table 1 seem to be convincing. For example, with 500 replications, $n = 50$ and $\mu_1 = \mu_2 = \varphi = 1$, the respective average ml estimates were 1.028, 1.048 and 1.414, with root-mean-squared deviations 0.281, 0.308 and 1.781 respectively; the corresponding results for the mpsl estimation were 1.028, 1.048, 1.490 and 0.285, 0.314, 2.217. Overall, the results in Table 1 provide with experimental evidence that for samples with moderate-to-large sizes ($n = 100$ to 500) the ml are close to the mpsl estimates and therefore the computationally simpler method of mpsl seems to be an attractive alternative to the, by far, more difficult method of ml.

5. Generalizations of the BPVIIC distribution

5.1 *Bivariate case*

In the previous sections the most general class of bivariate distributions which have PTVII conditionals, with a common constant shape parameter, was examined in detail. Here we treat in passing the general case where both the scale and shape parameters depend on the conditioned variable. Therefore, under the settings of Section 2, we wish to identify the most general class of bivariate distributions with conditionals that satisfy $X \mid Y \sim \mathrm{PVII}\{\sigma(y), p(y)\}$, $\forall y \in R$ and $Y \mid X \sim \mathrm{PVII}\{\tau(x), q(x)\}$, $\forall x \in R$, where $\tau(x), \sigma(y) \in R_+$ and $q(x), p(y) \in (2^{-1}, \infty)$. In this case the functional equation to be solved assumes the form

$$a(x)\{1 + \tau(x)y^2\}^{c(x)} = b(y)\{1 + \sigma(y)x^2\}^{d(y)}, \qquad x, y \in R, \tag{5.1}$$

where

$$a(x) = f_X(x)\sqrt{\tau(x)}\Gamma\{q(x)\}\left[\Gamma\left\{q(x) - \frac{1}{2}\right\}\right]^{-1}, \qquad c(x) = -q(x),$$

$$b(y) = f_Y(y)\sqrt{\sigma(y)}\Gamma\{p(y)\}\left[\Gamma\left\{p(y) - \frac{1}{2}\right\}\right]^{-1}, \qquad d(y) = -p(y).$$

Setting $u = x^2$ and $v = y^2$ in (5.1) we get four functional equations with each one assuming the form

$$\tilde{a}(u)\{1 + \tilde{\tau}(u)v\}^{\tilde{c}(u)} = \tilde{b}(v)\{1 + \tilde{\sigma}(v)u\}^{\tilde{d}(v)}, \qquad u, v \in R_+, \tag{5.2}$$

where $\tilde{a}(u)$ is either $a(\sqrt{u})$ or $a(-\sqrt{u})$ and similarly for the others. The functions $\tilde{a}(u)$, $\tilde{b}(v)$, $\tilde{\tau}(u)$ and $\tilde{\sigma}(v)$ are positive while $\tilde{c}(u)$ and $\tilde{d}(v)$ assume values in $(-\infty, -2^{-1})$. The solution to the functional equation (5.2) was given by Castillo and Galambos (1987*b*); Arnold *et al.* (1993) used it in their characterization of multivariate distributions with generalized Pareto conditionals. Two families of solutions of (5.2) exist. The first is such that

$$\tilde{a}(u)\{1+\tilde{\tau}(u)v\}^{\tilde{c}(u)} = (\lambda_1 + \lambda_2 u + \lambda_3 v + \lambda_4 uv)^{\lambda_5},$$

and the second

$$\tilde{a}(u)\{1+\tilde{\tau}(u)v\}^{\tilde{c}(u)} = \exp\{\theta_1 + \theta_2 \log(1+\theta_5 u) + \theta_3 \log(1+\theta_6 v) \\ +\theta_4 \log(1+\theta_5 u)\log(1+\theta_6 v)\},$$

where λ_i, $i = 1,\ldots,5$ and θ_j, $j = 1,\ldots,6$ are constants. Substituting back for all the functions we get the two possible models with PVII conditionals in the general case,

$$f_{X,Y}(x,y) \propto (\lambda_1 + \lambda_2 x^2 + \lambda_3 y^2 + \lambda_4 x^2 y^2)^{\lambda_5}, \qquad x, y \in R \tag{5.3}$$

and

$$f_{X,Y}(x,y) \propto \exp\{\theta_1 + \theta_2 \log(1+\theta_5 x^2) + \theta_3 \log(1+\theta_6 y^2) \\ +\theta_4 \log(1+\theta_5 x^2)\log(1+\theta_6 y^2)\}, \qquad x, y \in R. \tag{5.4}$$

Model (5.3) corresponds to the case $p(y) = q(x) \equiv$ constant, for every $x, y \in R$ and is the one examined previously. The value of the common constant shape parameter is $-\lambda_5$. The second model, (5.4) is distinct from the one examined so far. It corresponds to the case where $\tau(x)$ and $\sigma(y)$ are constants $\forall x, y \in R$; specifically, $\tau(x) = \theta_5$ and $\sigma(y) = \theta_6$. The PVII distribution with known scale parameter is a member of the one-parameter exponential family of distributions and therefore model (5.4) can be obtained from the results of Arnold and Strauss (1991*a*).

To summarize, we can say that there are two classes of bivariate distributions with PVII conditionals, one with a common constant shape parameter and one with constant values for the scale parameters. Therefore it is not possible to have a model in which both conditionals are PVII with both shape and scale parameters dependent on the conditioned variable.

5.2 *Multivariate case*

Extending the form of the model (2.6) we can identify the form of the pdf of multivariate distributions with PVII conditionals and fixed shape parameter. Let $X = (X_1,\ldots,X_k)$ be a k-dimensional random variable whose joint density $f_X(x_1,\ldots,x_k)$ exists and is positive over R^k. Furthermore, let $X^{(i)}$, $i = 1,\ldots,k$ be the vector X with the i-th coordinate X_i deleted. If we assume that for each

$i = 1, \ldots, k$ the conditional distribution of $X_i \mid X^{(i)}$ is PVII, then the joint density is given by

$$f_X(x_1, \ldots, x_k) = N_p(\lambda_i) \left\{ \sum_{s \in \xi_k} \lambda_s \left(\prod_{j=1}^{k} x_j^{2s_j} \right) \right\}^{-p}, \tag{5.5}$$

where $x_j \in R$, $j = 1, \ldots, k$ and $p > 2^{-1}$, ξ_k denotes the set of all vectors of dimension k, with coordinates 0 and 1 only and $N_p(\lambda_i)$ is the normalizing constant. Restrictions have to be placed on the parameters λ_s so that compatibility conditions are satisfied. All of them must be non-negative, while some are permitted to be zero, depending on the values of k and p.

Acknowledgements

We wish to thank two referees for their useful comments and suggestions.

References

Abramowitz, M. and Stegun, I. A. (1968). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, Dover, New York.

Aczel, J. (1966). *Lectures on Functional Equations and Their Applications*, Academic Press, New York.

Anderson, D. N. and Arnold, B. C. (1991). Centered distributions with Cauchy conditionals, *Comm. Statist. Theory Methods*, **20**(9), 2881–2889.

Arnold, B. C. and Press, S. J. (1989). Compatible conditional distributions, *J. Amer. Statist. Assoc.*, **84**, 152–156.

Arnold, B. C. and Strauss, D. (1991*a*). Bivariate distributions with conditionals in prescribed exponential families, *J. Roy. Statist. Soc. Ser. B*, **53**, 365–375.

Arnold, B. C. and Strauss, D. (1991*b*). Pseudolikelihood estimation: some examples, *Sankhyā Ser. B*, **53**, 233–243.

Arnold, B. C., Castillo, E. and Sarabia, J. M. (1992). Conditionally specified distributions, *Lecture Notes in Statist.*, **73**, Springer, Heidelberg.

Arnold, B. C., Castillo, E. and Sarabia, J. M. (1993). Multivariate distributions with generalized Pareto conditionals, *Statist. Probab. Lett.*, **17**, 361–368.

Arnold, B. C., Castillo, E. and Sarabia, J. M. (1995). General conditional specification models, *Comm. Statist. Theory Methods*, **24**(1), 1–11.

Besag, J. E. (1975). Statistical analysis of non-lattice data, *The Statistician*, **24**, 179–195.

Besag, J. E. (1977). Efficiency of pseudolikelihood estimators for simple Gaussian fields, *Biometrika*, **64**, 616–618.

Carlson, B. C. (1977). *Special Functions of Applied Mathematics*, Academic Press, New York.

Castillo, E. and Galambos, J. (1987*a*). Bivariate distributions with normal conditionals, *Proc. Int. Assoc. Sci. Tech. for Develop., Int. Symp. Simul., Modelling Develop., Cairo* (ed. M. H. Hamza), 59–62, Acta Press, Anaheim, California.

Castillo, E. and Galambos, J. (1987*b*). Lifetime regression models based on a functional equation of physical nature, *J. Appl. Probab.*, **24**, 160–169.

Castillo, E. and Galambos, J. (1989). Conditional distributions and the bivariate normal distribution, *Metrika*, **36**, 209–214.

Castillo, E. and Sarabia, J. M. (1990). Bivariate distributions with second kind Beta conditionals, *Comm. Statist. Theory Methods*, **19**(9), 3433–3445.

Johnson, M. (1987). *Multivariate Statistical Simulation*, Wiley, New York.

Johnson, N. L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*, Wiley, New York.

Magnus, W., Oberhettinger, F. and Soni, R. P. (1966). *Formulas and Theorems for the Special Functions of Mathematical Physics*, Springer, New York.

Mardia, K. V. (1970). *Families of Bivariate Distributions*, Griffin, London.

Sarabia, J. M. (1995). The centered normal conditionals distribution, *Comm. Statist. Theory Methods*, **24**(11), 2889–2900.