STATISTICA, anno LXXVI, n. 2, 2016

THE CAMBANIS FAMILY OF BIVARIATE DISTRIBUTIONS: PROPERTIES AND APPLICATIONS

N. Unnikrishnan Nair
Department of Statistics Cochin University of Science and Technology, Cochin
Johny Scaria [1]
Department of Statistics Nirmala College, Muvattupuzha
Sithara Mohan
Department of Statistics Nirmala College, Muvattupuzha

1. Introduction

Bivariate distributions specified by their marginals has been a topic of considerable interest in distribution theory. Among these the Farlie-Gumbel-Morgenstern (FGM) family of distributions was studied extensively by many authors. This family is represented by the bivariate distribution function, with marginal distribution functions $F_1(x_1)$ and $F_2(x_2)$, as

$$F(x_1, x_2) = F_1(x_1)F_2(x_2)[1 + \alpha(1 - F_1(x_1))(1 - F_2(x_2))]; |\alpha| < 1. \tag{1}$$

For the properties and applications of (1), we refer to Johnson and Kotz (1975), Schucany et al. (1978), Drouet Mari and Kotz (2001) and various references there in. One important limitation of the family is that its coefficient correlation is restricted to the narrow range of $(-\frac{1}{3}, \frac{1}{3})$, so that its application is confined to data that exhibits low correlation. Accordingly there has been several attempts to modify (1) by several researchers like Kotz and Johnson (1977), Cambanis (1977), Huang and Kotz (1984, 1999), Bairamov et al. (2001), Amblard and Girard (2009), Bekrizadeh et al. (2012) and Carles et al. (2012) to enhance the range of correlation as well as to impart more flexibility aimed at extending the domain of application. The Cambanis system defined by the distribution function

$$\begin{aligned} F(x_1, x_2) =& F_1(x_1)F_2(x_2)[1 + \alpha_1(1 - F_1(x_1)) + \alpha_2(1 - F_2(x_2)) \\ &+ \alpha_3(1 - F_1(x_1))(1 - F_2(x_2))], \end{aligned} \tag{2}$$

obtained by the addition of two linear terms in F_1 and F_2 to (1) differs from the other extensions of the Farlie-Gumbel-Morgenstern family in the method of construction and in the properties. While introducing this family Cambanis (1977) has

[1] Corresponding Author. E-mail: johny.kakkattil@gmail.com

shown that if the random variables corresponding to (2) are uncorrelated, they are independent and also gave an interpretation to the parameters. Other than this, a systematic study of the properties of the distribution and its application does not appear to have been discussed in literature. The present paper is an attempt in this direction. Apart from the generalization of a well discussed family there are several factors that makes our work worthwhile. In the first place under some simple conditions on the parameters it is totally positive (reverse rule) of order two, the strongest condition for positive (negative) association. The coefficients of association like Kendall's tau, Spearman's rho and various time-dependent measures of association have quite simple expressions and mutual implications. We can use the family as a model for lifetimes in a two-component device or system. A peculiar feature of the bivariate reliability function derived of this model is that it can be expressed in terms of the univariate reliability functions of the marginals of (2) and those of the baseline distributions F_1 and F_2. Further it can be used as a model in all situations where the Farlie-Gumbel-Morgenstern family is applied.

The present work is organized into four sections. In section 2, we enumerate the distributional characteristics. This is followed by the reliability properties in section 3. We demonstrate the application of the family in modelling bivariate lifetime data in section 4. The study ends with a brief conclusion in section 5.

2. Distributional properties

The survival function and the probability density function of the vector (X_1, X_2) following Cambanis distribution are, respectively

$$\bar{H}(x_1, x_2) = \bar{F}_1(x_1)\bar{F}_2(x_2)[1 - \alpha_1 F_1(x_1) - \alpha_2 F_2(x_2) + \alpha_3 F_1(x_1)F_2(x_2)], \quad (3)$$

and

$$\begin{aligned} h(x_1, x_2) &= f_1(x_1)f_2(x_2)[1 + \alpha_1(1 - 2F_1(x_1)) + \alpha_2(1 - 2F_2(x_2)) \\ &+ \alpha_3(1 - 2F_1(x_1))(1 - 2F_2(x_2))]. \end{aligned} \quad (4)$$

When (3) is absolutely continuous the parameters satisfy the conditions

$(1 + \alpha_1 + \alpha_2 + \alpha_3) > 0$, $(1 - \alpha_1 - \alpha_2 + \alpha_3) > 0$, $(1 - \alpha_1 + \alpha_2 - \alpha_3) > 0$ and $(1 + \alpha_1 - \alpha_2 - \alpha_3) > 0$, where $\alpha's$ are real constants. The marginal survival functions are

$$\bar{H}_i(x_i) = \bar{F}_i(x_i)[1 - \alpha_i F_i(x_i)], i = 1, 2. \quad (5)$$

Unlike the FGM, the marginals are not F_1 and F_2 from which the bivariate family is constructed, but the marginals H_1 and H_2 are uniquely determined by F_1 and F_2. From (5), the marginal densities work out to be

$$h_i(x_i) = f_i(x_i)[1 + \alpha_i(1 - 2F_i(x_i))]. \quad (6)$$

Let Z_1 and Z_2 be the random variables with distribution functions F_1 and F_2.

Then

$$\begin{aligned}\mu_{r,0} = E(X_1^r) &= \int_{-\infty}^{\infty} x_1^r h_1(x_1) dx_1 \\ &= \int_{-\infty}^{\infty} x_1^r (1 + \alpha_1 - 2\alpha_1 F_1(x_1)) f_1(x_1) dx_1 \\ &= (1+\alpha_1) M_1(r,0,0) - 2\alpha_1 M_1(r,1,0),\end{aligned}$$

where

$$M_1(r,s,t) = E(Z_1^r F_1^s \bar{F}_1^{\,t}), \tag{7}$$

is the probability weighted moment (PWM) of order (r,s,t) of Z_1. For a detailed study of the properties and application of PWM we refer to Greenwood et al. (1975). For brevity we suppress the suffix *t=0* in the following deliberations. Defining $M_2(r,s,t)$ as the PWM of Z_2,

$$\mu_{0r} = E(X_2^r) = (1+\alpha_2) M_2(r,0) - 2\alpha_2 M_2(r,1). \tag{8}$$

Further from (4),

$$\begin{aligned}E(X_1X_2) = &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 h(x_1,x_2) dx_1 dx_2 \\ &= M_1(1,0)M_2(1,0) + \alpha_1[M_1(1,0) - 2M_1(1,1)]M_2(1,0) + \alpha_2[M_2(1,0) \\ &-2M_2(1,1)]M_1(1,0) + \alpha_3[M_1(1,0) - 2M_1(1,1)][M_2(1,0) - 2M_2(1,1)],\end{aligned}$$

and hence

$$\begin{aligned}Cov(X_1X_2) &= (\alpha_3 - \alpha_1\alpha_2)[M_1(1,0)M_2(1,0) - 2M_1(1,1)M_2(1,0) \\ &- 2M_1(1,0)M_2(1,1) + 4M_1(1,1)M_2(1,1)].\end{aligned} \tag{9}$$

From (7), (8), and (9) the means and variances of X_i and the correlation coefficient between X_1 and X_2 follow.

Two kinds of conditional distributions are of interest in studying the dependence structures of bivariate distributions. Of these the distributions of X_i given $X_j > x_j$, i, j= 1, 2, $i \neq j$ has survival function

$$P[X_i > x_i | X_j > x_j] = \bar{F}_j(x_j)\frac{[1 - \alpha_1 F_1(x_1) - \alpha_2 F_2(x_2) + \alpha_3 F_1(x_1) F_2(x_2)]}{1 - \alpha_j F_j(x_j)},$$

and that of X_i given $X_j{=}x_j$ is given by

$$P[X_i > x_i | X_j = x_j] = \frac{\bar{F}_i(x_i) f_j(x_j)[\alpha_i F_i(x_i) - 1]}{[1 + \alpha_j(1 - 2F_j(x_j))]}.$$

2.1. Dependence

There are different approaches to study the dependence or association between two random variables. The classical method is to construct a coefficient that indicates the extent and nature of dependence. Prominent measures belonging to this family are the Kendall's tau, Spearman's rho, and Blomqvist's beta (Nelson (2006)). Although one of these measures is sufficient to find the nature of dependence for the Cambanis family, we discuss all so that the practitioner can choose one of his choice.

When X_1 and X_2 are continuous random variables Kendall's tau is calculated as

$$\tau = 4\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(x_1,x_2)h(x_1,x_2)dx_1dx_2 - 1,$$

For the Cambanis family

$$\tau = \frac{2}{9}(\alpha_3 - \alpha_1\alpha_2). \tag{10}$$

For the FGM, $\alpha_1 = \alpha_2 = 0$ and $-1 < \alpha_3 < 1$ so that $\tau \in (\frac{-2}{9}, \frac{2}{9})$. The Cambanis family improves upon the dependence coefficient compared to the FGM. For example, assume that α_3=0.9, α_1 = -0.41 and α_2 = -0.5. Then FGM gives $\tau = \frac{2\alpha_3}{9}$ = 0.2 while for the Cambanis family τ=0.24. Notice that the choice of α_1 and α_2 are subject to the constraints on the parameters of (3).

The Spearman's rho derived from

$$\begin{aligned}\rho &= 12\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[H(x_1,x_2) - H_1(x_1)H_2(x_2)]dH_1(x_1)dH_2(x_2), \\ &= \frac{(\alpha_3 - \alpha_1\alpha_2)}{3}.\end{aligned}$$

Finally the Blomqvists measure is the simplest of all expressed as,

$$\beta = 4H(M_{X_1}, M_{X_2}) - 1,$$

where M_{X_1} and M_{X_2} are the medians of X_1 and X_2. From (5), using $H_i(M_{x_i}) = \frac{1}{2}$, i=1,2 we obtain

$$F_i(M_{x_i}) = \frac{1 + \alpha_i - (1+\alpha_i^2)^{1/2}}{2\alpha_i}.$$

Substituting this in $H(.,.)$, we find $H(M_{X_1}, M_{X_2})$ to obtain β. Thus,

$$\begin{aligned}\beta = {} & \alpha_1^{-1}\alpha_2^{-1}(1+\alpha_1-(1+\alpha_1^2)^{\frac{1}{2}})(1+\alpha_2-(1+\alpha_2^2)^{\frac{1}{2}}) \\ & \{1 + \frac{(\alpha_1 - 1 - (1+\alpha_1^2)^{\frac{1}{2}})}{2} + \frac{(\alpha_2 - 1 - (1+\alpha_2^2)^{\frac{1}{2}})}{2} \\ & + \frac{\alpha_3(\alpha_1 - 1 - (1+\alpha_1^2)^{\frac{1}{2}})(\alpha_2 - 1 - (1+\alpha_2^2)^{\frac{1}{2}})}{4\alpha_1\alpha_2}\} - 1.\end{aligned}$$

A second approach to study the dependence structure is to make use of various dependence concepts. The six basic positive dependence concepts used for the purpose in the order of stringency are positive correlation, positive quadrant dependence (PQD), association, right tail increase, stochastic increase, total positivity of order 2. Among these, a function $g(x, y)$ is said to be totally positive (reverse rule) of order two, abbreviated as TP_2 (RR_2) if

$$g(x,y)g(u,v) - g(x,v)g(u,y) \geq (\leq)0,$$

for all $x < u$, $y < v$. When the support of the vector (X_1, X_2) is a cartesian product set, then an equivalent condition for TP_2 (RR_2) is that the density function $h(x_1, x_2)$ satisfies

$$\frac{\partial^2}{\partial x_1 \partial x_2} \log h(x_1, x_2) \geq (\leq)0.$$

For the Cambanis family

$$\frac{\partial^2 \log h}{\partial x_1 \partial x_2} = \frac{(\alpha_3 - \alpha_1\alpha_2)h_1(x_1)h_2(x_2)}{[1 - \alpha_1 F_1(x_1) - \alpha_2 F_2(x_2) + \alpha_3 F_1(x_1)F_2(x_2)]^2},$$

showing that $h(x_1, x_2)$ is TP_2 (RR_2) if $\alpha_3 \geq (\leq)\alpha_1\alpha_2$. The value of τ in (10) further reveals that $\tau > 0 \Longrightarrow (X_1, X_2)$ is TP_2 (RR_2). Thus $\tau > (<)0$ is a sufficient condition for all the six positive (negative) dependent concepts to hold.

The third class of dependence measures are time bound, they being dependent on x_1 and x_2. In reliability and survival analysis the lifetime remaining to an individual or device after it has attained a specific age is of substantial interest. For example, in medical studies such time-dependent measures are useful in determining the time of maximum association between interval from remission to relapse and the next interval from relapse to death. Many time-dependent measures are proposed to analyse the nature of association.

Clayton (1978) proposed

$$\theta(x_1, x_2) = \frac{\bar{H}(x_1, x_2)\frac{\partial^2 \bar{H}(x_1, x_2)}{\partial x_1 \partial x_2}}{\frac{\partial}{\partial x_1}\bar{H}(x_1, x_2)\frac{\partial}{\partial x_2}\bar{H}(x_1, x_2)},$$

where $\bar{H}$ is an arbitrary absolutely continuous bivariate survival function, as a measure of association. The variables X_1 and X_2 are positively or negatively associated according as $\theta > 1$ or < 1 and X_1 and X_2 are independent of $\theta = 1$. In terms of logarithms of the survival functions it can be seen that

$$\theta(x_1, x_2) - 1 = \frac{\frac{\partial^2}{\partial x_1 \partial x_2} \log \bar{H}(x_1, x_2)}{\frac{\partial \log \bar{H}}{\partial x_1}\frac{\partial \log \bar{H}}{\partial x_2}}.$$

Taking $\bar{H}$ to be (3) after some algebra we find

$$\begin{aligned} \theta(x_1, x_2) - 1 &= \frac{\bar{F}_1(x_1)\bar{F}_2(x_2)}{[1 - 2\alpha_1 F_1(x_1) - \alpha_2 F_2(x_2) + 2\alpha_3 F_1(x_1)F_3(x_3)]} \\ &\times \frac{(\alpha_3 - \alpha_1\alpha_2)}{[1 - \alpha_1 F_1(x_1) - 2\alpha_2 F_2(x_2) + 2\alpha_3 F_1(x_1)F_2(x_2)]}. \end{aligned}$$

Since the denominator is positive, it can be seen that the association is positive (negative) iff $\alpha_3 > (<)\alpha_1\alpha_2$ and the variables are independent if $\alpha_3 = \alpha_1\alpha_2$. We say that two continuous random variables X_1 and X_2 are right corner set increasing (decreasing)- RCSI (RCSD), if

$$P(X_1 > x_1, X_2 > x_2 | X_1 > x_1', X_2 > x_2'),$$

is non-decreasing (non-increasing) in x_1' and x_2' for all x_1 and x_2. Further Gupta (2003) has shown that X_1 and X_2 are RCSI (RCSD) if and only if $\theta > (<) 1$. Accordingly $\alpha_3 > \alpha_1\alpha_2$ is a necessary and sufficient condition that (X_1, X_2) is RCSI (RCSD) which is equivalent to $\tau > (<) 0$.

A second time-dependent measure discussed in Anderson et al. (1992) is

$$\psi(x_1, x_2) = \frac{\bar{H}(x_1, x_2)}{\bar{H}_1(x_1)\bar{H}_2(x_2)},$$

which for the Cambanis family simplifies to

$$\psi(x_1, x_2) = 1 + \frac{(\alpha_3 - \alpha_1\alpha_2)F_1(x_1)F_2(x_2)}{(1 - \alpha_1 F_1(x_1))(1 - \alpha_2 F_2(x_2))}.$$

The interpretation of ψ is that when ψ=1, X_1 and X_2 are independent and larger (smaller) values of ψ greater (lesser) than unity implies positive (negative) association. Thus $\alpha_3 > \alpha_1\alpha_2$ ($\alpha_3 < \alpha_1\alpha_2$) indicate positive (negative) association. By definition (X_1, X_2) is positive (negative) quadrant dependent - PQD (NQD), if $\bar{H}(x_1, x_2) \geq (\leq) \bar{H}_1(x_1)\bar{H}_2(x_2)$, so that $\psi(x_1, x_2) > (<) 1$ is equivalent to PQD (NQD) and this happens only if $\alpha_3 > \alpha_1\alpha_2$. Notice that whenever $\tau > (<) 0$ both $\theta(x_1, x_2)$ and $\psi(x_1, x_2)$ exhibit positive (negative) association, but the former cannot account for the extent of association as it does not depend on the terms containing x_1 and x_2 in the latter measures. The advantage of time-dependent measures is that a further analysis of its behaviour is possible. For example in the case of the uniform distribution with

$$F_i(x_i) = x_i, 0 \leq x_i \leq 1, i = 1, 2,$$

$$\bar{H}(x_1, x_2) = (1 - x_1)(1 - x_2)[1 - \alpha_1 x_1 - \alpha_2 x_2 + \alpha_3 x_1 x_2],$$

$$\psi(x_1, x_2) = 1 + \frac{(\alpha_3 - \alpha_1\alpha_2)x_1 x_2}{(1 - \alpha_1 x_1)(1 - \alpha_2 x_2)},$$

$$\frac{\partial\psi}{\partial x_1} = \frac{(\alpha_3 - \alpha_1\alpha_2)x_2}{(1 - \alpha_1 x_1)^2(1 - \alpha_2 x_2)}.$$

Suppose $\alpha_3 > \alpha_1\alpha_2$ so that X_1 and X_2 are positively associated. This association will show a decreasing trend in x_1 whenever $\alpha_1 < 0$ and increasing trend for $\alpha_1 > 0$ and similarly in x_2.

The behaviour of the local dependence function of Holland and Wang (1987) viz,

$$\delta(x, y) = \frac{\partial^2}{\partial x_1 \partial x_2} \log h(x_1, x_2),$$

is already clear from our discussions on TP_2. There are other measures like local correlation coefficient of Bjerve and Doksum (1993) and linear dependence function of Bairamov et al. (2003) discussed in literature. The expressions for these functions for the general family are intractable, but there is closed form expressions for some of the members

Example 2.1. *The Bjerve and Doksum measure is*

$$\rho(x_1) = \frac{\sigma_{X_1}\beta(x_1)}{(\sigma_{X_1}\beta(x_1))^2 + \sigma^2(x_1)},$$

where $\sigma^2_X = V(X_1)$, $\beta(x_1) = \frac{\partial}{\partial x_1}[E(X_2|X_1 = x_1)]$, $\sigma^2(x_1) = V(X_2|X_1 = x_1)$. *For the bivariate exponential law*

$$h(x_1, x_2) = e^{-x_1}e^{-x_2}[1 + \alpha_1(2e^{-x_1} - 1) + \alpha_2(2e^{-x_2} - 1) + \alpha_3(2e^{-x_1} - 1)(2e^{-x_2} - 1)], \qquad x_1, x_2 > 0,$$

$$\rho(x_1) = \frac{\sigma_{X_1}(\alpha_3 - 2\alpha_1)e^{-x_1}}{\sigma^2_{X_1}(\alpha_3 - 2\alpha_1)^2 e^{-2x_1} + \sigma^2(x_1)},$$

where

$$\sigma^2_{X_1} = 1 - \frac{5\alpha_1}{2} - \alpha_1^2,$$

and

$$\begin{aligned}\sigma^2(x_1) &= (1 + \frac{\alpha_2}{2} - \frac{\alpha_2^2}{4}) + (\alpha_2(\alpha_1 - \frac{\alpha_3}{2}))(2e^{-x_1} - 1) \\ &+ (\alpha_1 - \frac{\alpha_3}{2})^2(2e^{-x_1} - 1)^2.\end{aligned}$$

Generally, the coefficient $\rho(x_1)$ *lies between -1 and +1. But we can see from the above expression that the association is positive when* $\alpha_3 > 2\alpha_1$ *and negative when* $\alpha_3 < 2\alpha_1$.

3. Application to reliability modelling

3.1. Hazard rates

Gupta (2016) has analysed the FGM from a reliability point of view. We extend some of these results to the Cambanis family and provide some new applications. Two basic concepts required for our discussions are the bivariate hazard rate and the mean residual life. There are several definitions for the hazard rate in the multivariate case, of which we first consider the bivariate scalar hazard rate of Basu (1971), defined as

$$a(x_1, x_2) = \frac{h(x_1, x_2)}{\bar{H}(x_1, x_2)}.$$

For the Cambanis family

$$a(x_1, x_2) = \frac{f_1(x_1)f_2(x_2)[1 + \alpha_1(1 - 2F_1(x_1)) + \alpha_2(1 - 2F_2(x_2)) + \alpha_3(1 - 2F_1(x_1))(1 - 2F_2(x_2))]}{\bar{F}_1(x_1)\bar{F}_2(x_2)[1 - \alpha_1F_1(x_1) - \alpha_2F_2(x_2) + \alpha_3F_1(x_1)F_2(x_2)]}. \qquad (11)$$

Denoting by $r_i(x_i) = f_i(x_i)/\bar{F}_i(x_i)$ and $s_i(x_i) = h_i(x_i)/\bar{H}_i(x_i)$, the marginal hazard rates of the distributions F_i and H_i, i=1, 2, we can obtain a relationship between $s_i(x_i)$ and $r_i(x_i)$ in the form

$$s_i(x_i) = r_i(x_i) + \frac{\alpha_i r_i(x_i)\bar{F}_i(x_i)}{1 - \alpha_i\bar{F}_i(x_i)}.$$

Solving

$$F_i(x_i) = \frac{1}{\alpha_i}[(\alpha_i - 1) + \frac{r_i(x_i)}{s_i(x_i)}], i = 1, 2. \qquad (12)$$

Substituting (12) into (11), we get an expression for the bivariate hazard rate in terms of the univariate hazard rates r_i and s_i , after noting that the right side of (11), $f_1 f_2 / \bar{F}_1 \bar{F}_2 = r_1 r_2$. Thus an important property of the Cambanis family is that its scalar hazard rate is determined from the marginal hazard rates. Further note that such a result is not true for the FGM, as the expression (12) holds for $\alpha_i > 0$ only and to get FGM we need $\alpha_i = 0$.

The hazard gradient of (X_1, X_2) is

$$(b_1(x_1, x_2), b_2(x_1, x_2)) = (\frac{-\partial \log \bar{H}}{\partial x_1}, \frac{-\partial \log \bar{H}}{\partial x_2}),$$

which is a second approach to define bivariate hazard rate. Now

$$\begin{aligned} b_i(x_1, x_2) &= r_i - \frac{(\alpha_3 f_{3-i} - \alpha_i) f_i}{(1 - \alpha_1 F_1 - \alpha_2 F_2 + \alpha_3 F_1 F_2)} \\ &= \frac{r_i[1 - \frac{s_i - r_i}{\alpha_i s_i}\{\alpha_3 \frac{\alpha_{3-i} - 1}{\alpha_{3-i}} + \frac{r_i}{s_i}\} - \alpha_1]}{1 - ((\alpha_1 - 1) + \frac{r_1}{s_1}) - ((\alpha_2 - 1) + \frac{r_2}{s_2}) + \frac{\alpha_3}{\alpha_1 \alpha_2}((\alpha_1 - 1) + \frac{r_1}{s_1})(\alpha_2 - 2)(\frac{r_2}{s_2})}, \\ & \qquad i = 1, 2. \end{aligned}$$

A third definition hazard rate is the bivariate conditional hazard rate defined as

$$(c_1(x_1, x_2), c_2(x_1, x_2)) = (\frac{h(x_1|x_2)}{P(X_1 > x_1 | X_2 = x_2)}, \frac{h(x_2|x_1)}{P(X_2 > x_2 | X_1 = x_1)})$$

where $h(x_1|x_2), h(x_2|x_1)$ are the conditional density functions of

$$(X_1 | X_2 = x_2), (X_2 | X_1 = x_1)$$

. These rates are calculated as

$$c_i(x_1, x_2) = \frac{f_i(x_i)[1 + \alpha_1(1 - 2F_1(x_1)) + \alpha_2(1 - 2F_2(x_2)) + \alpha_3(1 - 2F_1(x_1))(1 - 2F_2(x_2))]}{\bar{F}_i(x_i)(1 - \alpha_i F_i(x_i))}, \quad i = 1, 2.$$

Thus the conditional failure rates can also be expressed in terms of $r_i(x_i)$ and $s_i(x_i)$.

3.2. Mean residual life

The mean residual life of X_i is given by

$$\begin{aligned} m_1(x_1) &= E(X_1 - x_1 | X_1 > x_1) = \frac{1}{H_1(x_1)} \int_{x_1}^{\infty} H_1(t) dt \\ &= \frac{1}{\bar{F}_1(x_1)(1 - \alpha_1 F_1(x_1))}[\int_{x_1}^{\infty} \bar{F}_1(t) dt - \alpha_1 \int_{x_1}^{\infty} \bar{F}_1(t) F_1(t) dt]. \quad (13) \end{aligned}$$

Likewise for the distribution F_1, the mean residual life is

$$n_1(x_1) = \frac{1}{\bar{F}_1(x_1)} \int_{x_1}^{\infty} \bar{F}_1(t)dt. \tag{14}$$

From (13) and (14)

$$\int_{x_1}^{\infty} \bar{F}_1(t)dt = \bar{F}_1(x_1)n_1(x_1), \tag{15}$$

and

$$\int_{x_1}^{\infty} \bar{F}_1(t)F_1(t)dt = \alpha_1^{-1}[n_1(x_1)\bar{F}_1(x_1) - \bar{F}_1(x_1)(1-\alpha_1 F_1(x_1))m_1(x_1)]. \tag{16}$$

Now the bivariate mean residual life is the vector $(\mu_1(x_1,x_2), \mu_2(x_1,x_2))$ where

$$\begin{aligned} \mu_1(x_1,x_2) &= E(X_1 - x_1 | X_1 > x_1, X_2 > x_2) \\ &= \frac{1}{\bar{H}(x_1,x_2)} \int_{x_1}^{\infty} \bar{H}(t,x_2)dt, \end{aligned} \tag{17}$$

and

$$\mu_2(x_1,x_2) = \frac{1}{\bar{H}(x_1,x_2)} \int_{x_2}^{\infty} \bar{H}(x_1,t)dt. \tag{18}$$

Using (15) and (16) in the expression for (17) simplifies to a closed form expression for μ_i as

$$\mu_1(x_1,x_2) = \frac{(\frac{\alpha_3}{\alpha_1} - \alpha_2)F_2(x_2)n_1(x_1) - (\frac{\alpha_3}{\alpha_1}F_2(x_2) - 1)(1-\alpha_1 F_1(x_1))m_1(x_1)}{1 - \alpha_1 F_1(x_1) - \alpha_2 F_2(x_2) + \alpha_3 F_1(x_1)F_2(x_2)},$$

Similarly from (18)),

$$\mu_2(x_1,x_2) = \frac{(\frac{\alpha_3}{\alpha_2} - \alpha_1)F_1(x_1)n_2(x_2) - (\frac{\alpha_3}{\alpha_2}F_1(x_1) - 1)(1-\alpha_2 F_2(x_2))m_2(x_2)}{1 - \alpha_1 F_1(x_1) - \alpha_2 F_2(x_2) + \alpha_3 F_1(x_1)F_2(x_2)}.$$

where n_2 and m_2 are the mean residual life functions corresponding to F_2 and H_2. An aspect of special interest in reliability analysis is the monotonicity of the reliability functions. From the expressions of the various functions, it is easy to recognize that the regions of the parameter space for which the hazard rates of mean residual life are increasing or decreasing are not easily determined analytically. However we note from Shaked (1977) that since $h(x_1,x_2)$ is TP_2 (RR_2) if $\alpha_3 \geq (\leq)\alpha_1\alpha_2$.

(a) $b_1(x_1,x_2)$ is decreasing when $\alpha_3 > \alpha_1\alpha_2$

(b) $c_1(x_1,x_2)$ is decreasing when $\alpha_3 > \alpha_1\alpha_2$

(c) $\mu_1(x_1,x_2)$ is increasing when $\alpha_3 > \alpha_1\alpha_2$.

3.3. The exponential case

Since exponential distribution play an important role in reliability analysis. In this section, we discuss the reliability properties of the Cambanis model when $F_i(x_i) = 1 - \exp[-\lambda_i x_i]$, $\lambda_i > 0$. In this case, the marginal distributions of X_i, i=1, 2 have survival functions

$$\begin{aligned}\bar{H}_i(x_i) &= e^{-\lambda_i x_i}[1 - \alpha_i(1 - e^{-\lambda_i x_i})] \\ &= (1 - \alpha_i)e^{-\lambda_i x_i} + \alpha_i e^{-2\lambda_i x_i}, \qquad (19)\end{aligned}$$

which is a generalized mixture of two exponential laws with parameters λ_i and $2\lambda_i$, respectively. Although the sum of the mixing constants is unity, α_i can be negative. The hazard rate of H_i is

$$s_i(x_i) = \frac{(1 - \alpha_i)\lambda_i e^{-\lambda_i x_i} + 2\lambda_i \alpha_i e^{-2\lambda_i x_i}}{(1 - \alpha_i)e^{-\lambda_i x_i} + 2\alpha_i e^{-2\lambda_i x_i}}.$$

Differentiating $s_i(x_i)$ with respect to x_i, we see that the sign of $\frac{ds_i}{dx_i}$ depends on

$$A_i(x_i) = -3\alpha_i(1 - \alpha_i)e^{-3\lambda_i x_i}.$$

Thus the marginal hazard rate is decreasing (DHR) for $0 < \alpha_i < 1$ and increasing (IHR) if α_i lies outside (0,1). On account of this the marginals can accommodate IFR and DFR data. The mean residual life function of X_i is

$$m_i(x_i) = \frac{\lambda_i^{-1}(1 - \alpha_i) + (2\lambda_i)^{-1}\alpha_i e^{-\lambda_i x_i}}{(1 - \alpha_i) + \alpha_i e^{-\lambda_i x_i}}.$$

Since X_i is DHR in $0 < \alpha_i < 1$ and IHR otherwise, it follows that X_i has increasing mean residual life (IMRL) in $0 < \alpha_i < 1$ and decreasing mean residual life for other values of α_i. From theorem 2 in Nair and Preeth (2009), the distributions (19) is characterized by the relationship

$$m_i(x_i) = \frac{3}{2\lambda_i} - \frac{1}{2\lambda_i^2} s_i(x_i),$$

with $\beta_i = 1 - \alpha_i$, $\beta_i \geq 0$, so that seperate calculation of $m_i(x_i)$ from $\bar{H}_i(x_i)$ is not neccesary.

To analyse the bivariate reliabiliy functions, we note that

$$\begin{aligned}\bar{H}(x_1, x_2) = e^{-\lambda_1 x_1 - \lambda_2 x_2}[&1 - \alpha_1(1 - e^{-\lambda_1 x_1}) - \alpha_2(1 - e^{-\lambda_2 x_2}) \\ &+ \alpha_3(1 - e^{-\lambda_1 x_1})(1 - e^{-\lambda_2 x_2})],\end{aligned}$$

and hence the hazard gradient has components

$$b_1(x_1, x_2) = \lambda_1 + \frac{(\alpha_2 - \alpha_3(1 - e^{-\lambda_2 x_2})\lambda_1 e^{-\lambda_1 x_1})}{1 - \alpha_1(1 - e^{-\lambda_1 x_1}) - \alpha_2(1 - e^{-\lambda_2 x_2}) + \alpha_3(1 - e^{-\lambda_1 x_1})(1 - e^{-\lambda_2 x_2})},$$

and

$$b_2(x_1,x_2) = \lambda_2 + \frac{(\alpha_1 - \alpha_3(1-e^{-\lambda_1 x_1})\lambda_2 e^{-\lambda_2 x_2})}{1-\alpha_1(1-e^{-\lambda_1 x_1}) - \alpha_2(1-e^{-\lambda_2 x_2}) + \alpha_3(1-e^{-\lambda_1 x_1})(1-e^{-\lambda_2 x_2})}.$$

Using (19), the mean residual life function reduces to

$$\mu_1(x_1,x_2) = \frac{1-(2-e^{-\lambda_1 x_1})(\frac{\alpha_1}{2} - \frac{\alpha_3}{2}(1-e^{-\lambda_2 x_2})) - \alpha_2(1-e^{-\lambda_2 x_2})}{\lambda_1[1-\alpha_1(1-e^{-\lambda_1 x_1}) - \alpha_2(1-e^{-\lambda_2 x_2}) + \alpha_3(1-e^{-\lambda_1 x_1})(1-e^{-\lambda_2 x_2})]},$$

Similarly

$$\mu_2(x_1,x_2) = \frac{1-(2-e^{-\lambda_2 x_2})(\frac{\alpha_2}{2} - \frac{\alpha_3}{2}(1-e^{-\lambda_1 x_1})) - \alpha_1(1-e^{-\lambda_1 x_1})}{\lambda_2[1-\alpha_1(1-e^{-\lambda_1 x_1}) - \alpha_2(1-e^{-\lambda_2 x_2}) + \alpha_3(1-e^{-\lambda_1 x_1})(1-e^{-\lambda_2 x_2})]}.$$

3.4. Series and parallel systems

Recall that when a two component system with (X_1, X_2) as the lifetimes of the components is connected in parallel, the lifetimes of the system is W=*max* (X_1, X_2) . If the joint lifetimes of the components is of Cambanis form, the survival function of W is

$$\begin{aligned} \bar{F}_W(x) &= 1 - H(x,x) \\ &= 1 - F_1(x)F_2(x)[1-\alpha_1\bar{F}_1(x) - \alpha_2\bar{F}_2(x) + \alpha_3\bar{F}_1(x)\bar{F}_2(x)]. \end{aligned}$$

The hazard rate function of the system becomes

$$\begin{aligned} F_W(x) = &- \frac{\partial \log \bar{F}_W}{\partial x} \\ =&[\{1-\alpha_1\bar{F}_1(x) - \alpha_2\bar{F}_2(x) + \alpha_3\bar{F}_1(x)\bar{F}_2(x)\}\{F_1(x)f_2(x) \\ &+ f_1(x)F_2(x) + F_1F_2\{\alpha_1 f_1(x) + \alpha_2 f_2(x) - \alpha_3\bar{F}_1(x)f_2(x) \\ &- \alpha_3\bar{F}_2(x)f_1(x)\}][1-H(x,x)]^{-1}. \end{aligned} \tag{20}$$

This can be expressed in terms of r_1, r_2, s_1, and s_2 and was done in the previous section. We have the mean residual life of W as

$$\begin{aligned} m_W(x) &= \frac{1}{\bar{F}_W(x)}\int_x^\infty \bar{F}_W(t)dt \\ &= \frac{\int_x^\infty [1-H(t,t)]dt}{1-H(x,x)} \\ &= \frac{\int_0^\infty [1-H(x+t,x+t)]dt}{1-H(x,x)}. \end{aligned}$$

Sometimes a maintenance strategy requires that maintenance be undertaken when the two components are in working condition. In that case, the residual life of interest is

$$\begin{aligned} w(x) &= E(W-x|U>x), U = min(X_1, X_2), \\ &= \int_0^\infty \frac{2\bar{H}(x+t,x) - \bar{H}(x+t,x+t)}{\bar{H}(x,x)} dt. \end{aligned}$$

When the system is connected in series, $U = min(X_1, X_2)$ is the system life length with survival function

$$\begin{aligned} \bar{F}_U(x) &= \bar{H}(x,x) \\ &= \bar{F}_1(x)\bar{F}_2 x[1 - \alpha_1 F_1(x) - \alpha_2 F_2(x) + \alpha_3 F_1(x) F_2(x)] \end{aligned}$$

and hazard rate

$$\begin{aligned} r_U(x) &= -\frac{\partial \log \bar{F}_U(x)}{\partial x} \\ &= r_1(x) + r_2(x) + \frac{\alpha_1 f_1(x) + \alpha_2 f_2(x) - \alpha_3(F_1(x) f_2(x) + f_1(x) F_2(x))}{1 - \alpha_1 F_1(x_1) - \alpha_2 F_2(x_2) + \alpha_3 F_1(x_1) F_2(x_2)} \\ &= r_1(x) + r_2(x) + A(x_1, x_2). \end{aligned} \tag{21}$$

In the case of independence of X_1 and X_2, r_U is the sum of the hazard rates $s_1(x)$ and $s_2(x)$. Since $s_i(x)[1 - \alpha_i \bar{F}_i(x)] = r_i(x)$, i=2, we can write

$$r_U(x) = s_1(x) + s_2(x) + [A(x_1, x_2) - \alpha_1 \bar{F}_1(x) s_1(x) - \alpha_2 \bar{F}_2(x) s_2(x)]$$

The term in the square braces on the right gives the effect of dependence between X_1 and X_2 on the hazard rate of the system life. By definition the mean residual life of the parallel system is

$$m_U(x) = \frac{1}{\bar{H}(x,x)} \int_x^\infty \bar{H}(t,t) dt.$$

The condition for the monotonicity of the hazard and mean residual life functions were discussed in the previous section. In the case of series and parallel systems, the hazard rate functions are obtained by (20) and (21) as,

$$r_W(x) = \frac{e^{-(\lambda_1+\lambda_2)x}(\lambda_1 + \lambda_2 - \alpha_1 \lambda_2 e^{-\lambda_1 x} - \alpha_2 \lambda_1 e^{-\lambda_2 x})}{1 - e^{-(\lambda_1+\lambda_2)x}[1 - \alpha_1 e^{-\lambda_1 x} - \alpha_2 e^{-\lambda_2 x} + \alpha_3 e^{-(\lambda_1+\lambda_2)x}]}$$

and

$$r_U(x) = \lambda_1 + \lambda_2 + \frac{[\alpha_1 - \alpha_3(1 - e^{-\lambda_2 x})]\lambda_1 e^{-\lambda_1 x} + [\alpha_2 - \alpha_3(1 - e^{-\lambda_1 x})] e^{-\lambda_2 x}}{1 - \alpha_1(1 - e^{-\lambda_1 x}) - \alpha_2(1 - e^{-\lambda_2 x}) + \alpha_3(1 - e^{-\lambda_1 x})(1 - e^{-\lambda_2 x})}.$$

All the reliability functions have closed form expressions and are easily calculated.

4. Modelling survival data

In this section, we illustrate the application of Cambanis family in modelling real life data. The data is on the survival distribution of incubation times of individuals known to have sexually transmitted disease who were later determined to have had sex with an individual who had the disease verified in the clinic after the time of their *encounter*, reported in Klein and Moeschberger (1997, p. 146). For 25 individuals the time in months from the first encounter and the time in months from the first encounter to the clinical confirmation of disease were recorded for 42 months. We denote by X_1 and X_2, the time to surviving the first encounter and the diagnosis of the disease respectively.

The main problem with associating Cambanis family to real data is that the marginals of X_1 and X_2 are not F_1 and F_2 and these are to be chosen from a list of candidate distributions that satisfy some $H_i(x_i)$, i=1,2 that are relevant to the observations on X_1 and X_2. In the absence of any physical considerations suggesting the form of H_i, the black box modelling approach was pursued. It was found that for a choice of Weibull distributions,

$$\bar{F}_i(x_i) = \exp[-\{\frac{x_i}{a_i}\}^{b_i}], x_i > 0; a_i, b_i > 0, i = 1, 2,$$

and $\bar{H}_i(x_i) = \bar{F}_i(x_i)[1 - \alpha_i F_i(x_i)]$, the observations on X_1 and X_2 gave the maximum likelihood estimates as $\hat{a_1}$=17.5476, $\hat{a_2}$=1.2832, $\hat{b_1}$=1.6826 and $\hat{b_2}$=2.1515, $\hat{\alpha_1}$=0.0148 and $\hat{\alpha_2}$=0.0597. With these estimates the maximum likelihood estimate of α_3 based on (X_1, X_2) values was obtained as $\hat{\alpha_3}$=-0.7972. Accordingly the fitted distributions are

$$\bar{H}(x_1, x_2) = \bar{F}_1\bar{F}_2[1 - 0.0148F_1(x_1) - 0.0597F_2(x_2) - 0.7972F_1(x_1)F_2(x_2)]. \quad (22)$$

$$\bar{H}_1(x_1) = \bar{F}_1(x_1)[1 - 0.0148F_1(x_1)], \quad (23)$$

and

$$\bar{H}_2(x_2) = \bar{F}_2(x_2)[1 - 0.0597F_2(x_2)], \quad (24)$$

where

$$\bar{F}_1(x_1) = \exp[-(\frac{x_1}{17.5476})]^{1.6826}; \bar{F}_2(x_2) = \exp[-(\frac{x_2}{1.2832})]^{2.1515}. \quad (25)$$

The Kolmogorov-Smirnov test statistic for $\bar{H}_1$ is 0.0832, for $\bar{H}_2$ is 0.0934 and the Kolmogorov-Smirnov test for bivariate distributions prescribed in Justel et al. (1997) provided the value for the test statistic as 0.1338. Thus the hypothesis of Cambanis distribution with F_1 and F_2 as Weibull is not rejected.

Regarding the dependence coefficients between X_1 and X_2 we have τ=-0.1773, ρ=-0.2659 as estimated values against the respective sample values -0.1559 and -0.2267 exhibiting the proximity between the sample values and the estimates of the population values. Obviously there is negative dependence between X_1 and X_2 implying that as the time to first encounter increases, the time to be affected by the disease decreases. This fact is also substantiated by the analysis of the reliability functions of the fitted models (22) through (25). In the first place we observe that

F_1 and F_2 being Weibull distributions with $b_1 > 1$ and $b_2 > 1$ respectively, their hazard rate functions are increasing. From Section 3.1, the marginal hazard rates are related through

$$s_i(x_i) = r_i(x_i) + \frac{\alpha_i r_i(x_i)\bar{F}_i(x_i)}{1-\alpha_i \bar{F}_i(x_i)}.$$

Differentiating with respect to x_i and simplifying we find

$$\begin{aligned}\frac{ds_i(x_i)}{dx_i} &= \frac{dr_i(x_i)}{dx_i} + \bar{F}_i(x_i)\frac{[(1-\alpha_i\bar{F}_i(x_i))\alpha_i\frac{dr_i(x_i)}{dx_i} - \alpha_i r_i^2(x_i)]}{[1-\alpha_i\bar{F}_i(x_i)]^2} \\ &= \frac{\frac{dr_i(x_i)}{dx_i}[1-\alpha_i\bar{F}_i(x_i)] - \alpha_i r_i^2(x_i)}{[1-\alpha_i\bar{F}_i(x_i)]^2}.\end{aligned}$$

The sign of $s_i(x_i)$ depends on the sign of $[1-\alpha_i\bar{F}_i(x_i)]\frac{dr_i(x_i)}{dx_i} - \alpha_i r_i^2(x_i)$, which is positive at the estimated values indicating increasing nature of the marginal hazard rates. Since $\alpha_3 < \alpha_1\alpha_2$, in the present problem among the components of the bivariate hazard rates $b_i(x_1, x_2)$ and $c_i(x_1, x_2)$, $b_1(x_1, x_2)$ and $c_1(x_1, x_2)$ are increasing in x_1 and $b_2(x_1, x_2)$ and $c_2(x_1, x_2)$ are increasing in x_2. Further conclusions about the various characteristics can be derived from the reliability functions.

5. Conclusion

In the present work, we have studied the Cambanis family of bivariate distributions. Various distributional properties such as moments, covariance etc were derived. An important aspect to be investigated in bivariate models is its dependence structure. The expressions of dependence coefficients like Kendall's tau, Spearmans rho and Blomquist's β were derived along with the conditions for positive and negative dependence. Time dependent measures of Clayton, Anderson, Bjerve and Doksum etc and dependence concepts like, TP_2, RCSI and PQD were examined to find the regions of the parameter space that provide positive and negative dependence. Necessary theoretical results were proposed to use the distribution in the context of modelling lifetime data by finding expressions for reliability functions. It was shown that the bivariate hazard and mean residual life functions can be written in terms of functions of the marginal distributions. Finally the scope of modelling survival data with Cambanis bivariate distribution was illustrated for real data.

Acknowledgements

We wish to thank the referee for his invaluable suggestions for improving the presentation of this work.

References

C. Amblard, S. Girard (2009). *A new extension of FGM copulas.* Metrika, **70**, 1–17.

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, 641–650.

I. Bairamov, S. Kotz, M. Bekci (2001). *New generalized Farlie-Gumbel-Morgenstern distributions and concomitants of order statistics.* Journal of Applied Statistics, 28, 5, 521–536.

I. Bairamov, S. Kotz, T. J. Kozubowski (2003). *A new measure of linear local dependence.* Statistics, 37, 243–258.

A. P. Basu (1971). *Bivariate failure rate.* Jornal of American Statistical Association, 66, 103–104.

H. Bekrizadeh, G. A. Parhan, M. R. Zadkarmi(2012). *A new generalization of Farlie-Gumbel-Morgenstern copulas.* Applied Mathematical Sciences, 6, 71, 3527–3533.

S. Bjerve, K. Doksum (1993). *Correlation curves: Measures of association as function of covariate values.* The Annals of Statistics, 21, 890–902.

S. Cambanis (1977). *Some properties and generalizations of Multivariate Eyraud-Gumbel-Morgenstern Distributions.* Journal of Multivariate Analysis, 7, 551–559.

M. Carles, C. M. Cudras, W. Daz (2012). *Another generaization of the bivariate FGM distribution with two- dimensional extensions.* Acta Et Commentationes Universitatis Tartuinsis De Mathematica, 16, 1.

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, 141–151.

D. Drouet Mari, S. Kotz (2001). *Correlation and Dependence.* Imperial College Press, London.

J. A. Greenwood, J. M. Landwehr, N. C. Matalas, J. R. Wallis (1979). *Probability weighted moments.* Water Resources Research, 15, 1049–1054.

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

R. C. Gupta (2016). *Reliability characteristics of Farlie-Gumbel-Morgenstern family of bivariate distributions.* Communications in Statistics-Theory and Methods (to appear).

P. W. Holland, Y. J. Wang (1987b). *Dependence function for continuous bivariate densities*. Communications in Statistics-Theory and Methods, 16, 863–876.

J. S. Huang, S. Kotz (1984). *Correlation structure in iterated Farlie-Gumbel-Morgenstern distributions*. Biometrika, 71, 633–636.

J. S. Huang, S. Kotz (1999). *Modifications of the Farlie-Gumbel-Morgenstern distributions, A tough hill to climb*. Metrika, 49, 135–145.

N. L. Johnson, S. Kotz (1972). *Distributions in statistics,Continuous Multivariate Distributions*, John Wiley, Newyork.

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

J. P. Klein, M. L. Moeschberger (1997). *Survival Analysis Technique for censored and truncated data*, Springer, New York.

S. Kotz, N. L. Johnson (1977). *Dependence properties of iterated generalized bivariate Farlie-Gumbel-Morgenstern distributions*. Comptes, Rendus Academic Science, Paris, 285 A, 277–280.

N. U. Nair, M. Preeth (2009). *On some properties of equilibrium distributions of order* n. Statistical methods Appl, 18, 453–464.

R. B. Nelson (2006). *An introduction to copulas*. (Second edition), Springer, New York.

W. Schuany, W. C. Parr, J. E. Boyer (1978). *Correlation structure in Farlie-Gumbel-Morgenstern distributions*. Biometrika, 65, 650–653.

M. Shaked (1977). *A family of concepts of dependence for bivariate distributions*. Journal of the American Statistical Association, 72, 642–650.

Summary

The Cambanis family of bivariate distributions was introduced as a generalization of the Farlie-Gumbel-Morgenstern system. The present work is an attempt to investigate the distributional characteristics and applications of the family. We derive various coefficients of association, dependence concepts and time-dependent measures. Bivariate reliability functions such as hazard rates and mean residual life functions are analysed. The application of the family as a model for bivariate lifetime data is also demonstrated.

Keywords: Bivariate Cambanis family; association measures; total positivity; bivariate hazard rates; bivariate mean residual life; series and parallel systems.