The Australian Journal of Mathematical Analysis and Applications

http://ajmaa.org

Volume 11, Issue 1, Article 2, pp. 1-13, 2014

SOME PROPERTIES OF THE MARSHALL-OLKIN AND GENERALIZED CUADRAS-AUGÉ FAMILIES OF COPULAS

EDWARD DOBROWOLSKI AND PRANESH KUMAR*

Received 24 May, 2013; accepted 31 October, 2013; published 19 February, 2014.

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF NORTHERN BRITISH COLUMBIA, PRINCE GEORGE, BC, CANADA, V2N 4Z9
Pranesh.Kumar@unbc.ca

ABSTRACT. We investigate some properties of the families of two parameter Marshall-Olkin and Generalized Cuadras-Augé copulas. Some new results are proved for copula parameters, dependence measure and mutual information. A numerical application is discussed.

Key words and phrases: Copulas, Dependence measures, Entropy measures, Mutual information.

2000 *Mathematics Subject Classification.* Primary 62H20,60E05. Secondary 2H86.

ISSN (electronic): 1449-5910

1. INTRODUCTION

Copulas allow to express multivariate distributions in terms of their marginal distributions and multivariate dependence structure. Copulas are a very popular tool and are widely applied, for example, to perform stress-tests and robustness checks in situations where extreme events may occur, in flexible modeling of the dependence structure for portfolios of large dimensions, in financial risk assessment and actuarial analysis, in database formulation for the reliability analysis, in various multivariate simulation studies in engineering applications, in climate and weather modeling research, in various simulation-based performance studies. In this paper we consider the Marshall-Olkin [12] and Generalized Cuadras-Augé [3] bivariate copulas and investigate some of its properties. Marshall and Olkin [12] described bivariate exponential distribution to study complex systems in which the two components are not independent and are subject to shocks which are fatal to one or both components.

2. MARSHALL-OLKIN AND GENERALIZED CUADRAS-AUGÉ FAMILIES OF COPULAS

Let X and Y be the lifetimes of the two components in a system. Assume that shocks to the two components follow three independent Poisson processes with parameters λ_1, λ_2 and $\lambda_{12} \geq 0$, depending on whether the shock kills only component 1, only component 2 or both components simultaneously. Hence for $x, y \geq 0$, survival function is

$$S(x,y) = P(X > x,\ Y > y) = \exp\left[-\lambda_1 x - \lambda_2 y - \lambda_{12}\ \max(x,y)\right], \tag{2.1}$$

and since $\max(x,y) = x + y - \min(x,y)$,

$$\begin{aligned} S(x,y) &= \exp[(-\lambda_1 + \lambda_{12})\ x - (-\lambda_2 + \lambda_{12}) \quad y + \lambda_{12}\ \min(x,y) \\ &= S(x)\ S(y) \min[\exp(\lambda_{12}\, x), \exp(\lambda_{12}\ y)] \end{aligned} \tag{2.2}$$

The marginal survival functions for X and Y are

$$S(x) = \exp[(-\lambda_1 + \lambda_{12})\ x], \tag{2.3}$$

and

$$S(y) = \exp[(-\lambda_2 + \lambda_{12})\ y]. \tag{2.4}$$

Setting $u = F(x)$, $v = G(y)$, $\alpha = \lambda_{12}/(\lambda_1 + \lambda_{12})$ and $\beta = \lambda_{12}/(\lambda_2 + \lambda_{12})$, survival copulas $\hat{C}$ associated with survival function is

$$\hat{C}(u,v) = uv \min(u^{-\alpha}, v^{-\beta}) = \min(u^{1-\alpha}v, uv^{1-\beta}). \tag{2.5}$$

Since λs are positive and both α and $\beta \in\ (0,1)$, survival copulas for the Marshall-Olkin bivariate exponential distribution leads to a two- parameter family of copulas

$$C(u,v) = \left[\begin{array}{ll} u^{1-\alpha}v, & u^{\alpha} \geq v^{\beta}, \\ uv^{1-\beta}, & u^{\alpha} \leq v^{\beta}. \end{array} \right. \tag{2.6}$$

This copula family is known as both the two-parameter Marshall-Olkin family [12] and the Generalized Cuadras-Augé [3] family when $\alpha = \beta$ corresponding to the case in which $\lambda_1 = \lambda_2$, that is, X and Y are exchangeable. For some interesting relevant readings, refer to [8, 5, 14]. In what follows now, we will refer this family of copulas as copula $C(u,v)$.

3. PROPERTIES OF THE MARSHALL-OLKIN FAMILY AND THE GENERALIZED CUADRAS-AUGÉ FAMILY

We will present and prove some properties of copula $C(u,v)$ in respect of probability distributions, dependence measures and information measures. An algorithm to generate copula $C(u,v)$ is also given.

3.1. **Copulas $C(u,v)$ and Probability distributions.** The density function associated with copula $C(u,v)$ is given by

$$c(u,v) = \frac{\partial^2 C(u,v)}{\partial u \partial v} = \left[\begin{array}{ll} (1-\alpha)u^{-\alpha}, & u^\alpha > v^\beta, \\ (1-\beta)v^{-\beta}, & u^\alpha < v^\beta, \end{array} \right. , \; 0 < \alpha, \beta < 1. \tag{3.1}$$

Thus, the mass of the singular component must be concentrated on the curve $u^\alpha > v^\beta$ in $[0,1]^2$.

The copulas $C(u,v)$ although have full support for α,$\beta \in (0,1)$, they are neither absolutely continuous nor singular. However they have both, the absolutely continuous component

$$\begin{aligned} A_{\alpha,\beta}(u,v) &= \int_0^u \int_0^v \frac{\partial^2 C(s,t)}{\partial s \partial t} dt ds \\ &= C(u,v) - \frac{\alpha\beta}{\alpha+\beta-\alpha\beta} \left[\min \left(u^\alpha, v^\beta\right)\right]^{(\alpha+\beta-\alpha\beta)/\alpha\beta}, \end{aligned} \tag{3.2}$$

and the singular component [14] given by

$$\begin{aligned} S_{\alpha,\beta}(u,v) = C(u,v) - A_{\alpha,\beta}(u,v) &= \frac{\alpha\beta}{\alpha+\beta-\alpha\beta} \left[\min \left(u^\alpha, v^\beta\right)\right]^{(\alpha+\beta-\alpha\beta)/\alpha\beta} \\ &= \int_0^{\min \left(u^\alpha, v^\beta\right)} t^{\frac{1}{\alpha}+\frac{1}{\beta}-2} dt. \end{aligned} \tag{3.3}$$

For the Marshall-Olkin copula $C(u,v)$ with $\alpha = 0.3$ and $\beta = 0.75$, the singular component is shown in Figure 1.

Two functions closely related to copulas and survival copulas, but which are not the copulas, are dual of a copula and the co-copula [15]. The dual of copula $C(u,v)$ is the function

$$\begin{aligned} \tilde{C}(u,v) &= u + v - C(u,v) \\ &= \left[\begin{array}{ll} u + v(1-u^{1-\alpha}), & u^\alpha \geq v^\beta, \\ v + u(1-v^{1-\beta}), & u^\alpha \leq v^\beta, \end{array} \right. \end{aligned} \tag{3.4}$$

and the co-copula is the function given by

$$\begin{aligned} C^*(u,v) &= 1 - C(1-u, 1-v) \\ &= \left[\begin{array}{ll} 1-(1-u)^{1-\alpha}(1-v), & u^\alpha \geq v^\beta, \\ 1-(1-u)(1-v)^{1-\beta}, & u^\alpha \leq v^\beta, \end{array} \right. \end{aligned} \tag{3.5}$$

The dual of copula and the co-copula both express probability of events X and Y. For $P(X \leq x, Y \leq y) = C(u,v)$ and $P(X > x, Y > y) = \hat{C}(1-u, 1-v)$, it is noted that

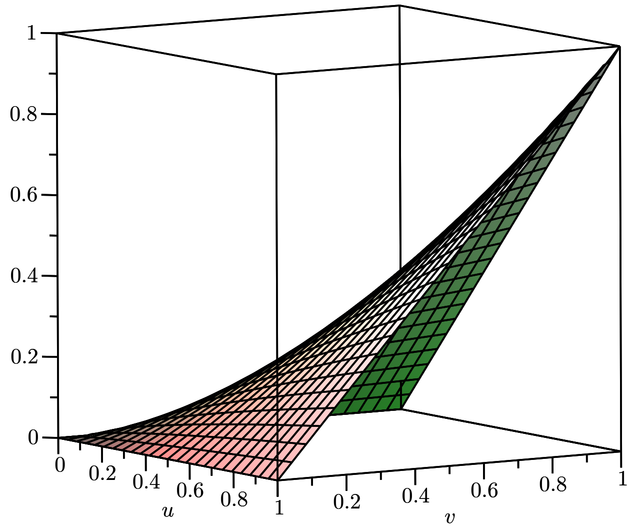
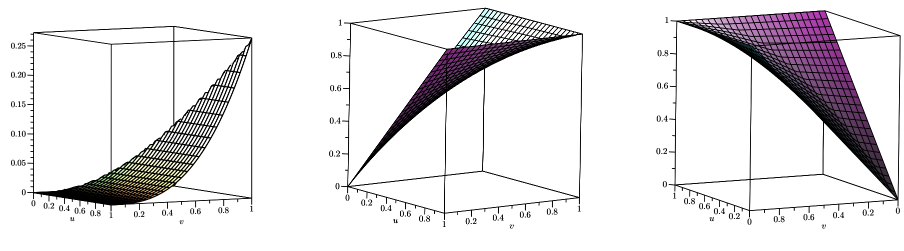


Figure 1: Singular component of $C(u, v)$ *for* $\alpha = 0.3$ *and* $\beta = 0.75$.

$$P(X \leq x \text{ or } Y \leq y) = \tilde{C}(u, v) \text{ and } P(X > x \text{ or } Y > y) = C^*(1 - u, 1 - v)\ . \tag{3.6}$$

The Marshall-Olkin copula $C(u, v)$, dual of copula $\tilde{C}(u, v)$ and co-copula $C^*(u, v)$ for $\alpha = 0.3$ and $\beta = 0.75$, are shown in Table 1 from the left to the right, respectively.

Table 3.1: Marshal-Olkin copula $(C(u, v)$, *its dual* $\tilde{C}(u, v)$, *an co-copula* $C^*(u, v)$ *for* $\alpha = 0.3$ *and* $\beta = 0.75$.



3.2. **Copulas $C(u, v)$ and Dependence measures.** Dependence or association between variables is one of the most studied concept in probability and statistics. The nature of dependence can take a variety of forms and unless some specific assumptions are made about the dependence, no meaningful statistical model can be contemplated [9, 10]. Several copula dependence properties and measures are scale invariant, that is, they remain unchanged under strictly increasing transformations of the variables [7, 18]. Copulas have been explored extensively in the study of dependence. Schweizer and Wolff [17] noted, "It is precisely the copula which captures those properties of the joint distribution which are invariant under almost surely strictly increasing transformations." Most widely used scale-invariant measures of association are the Kendall's τ and Spearman's ρ both of which measure a form of dependence known as concordance. We now present some results of copulas $C(u,v)$ related to dependence measures.

Theorem 3.1. *Let $C(u,v)$ be a member of two parameter Marshall-Olkin family of copulas with $0<\alpha,\beta<1$. Then Kendall's τ can be expressed as*

$$\begin{aligned}\tau &= 4\iint_{[0,1]^2} C(u,v)\,dC(u,v)-1\\ &= 1-4\iint_{[0,1]^2}\frac{\partial C(u,v)}{\partial u}\,\frac{\partial C(u,v)}{\partial v}\,du\,dv\\ &= \alpha\beta/(\alpha+\beta-\alpha\beta).\end{aligned} \tag{3.7}$$

Theorem 3.2. *Let $C(u,v)$ be a member of two parameter Marshall-Olkin family of copulas with $0<\alpha,\beta<1$. Then Spearman's ρ can be expressed as*

$$\begin{aligned}\rho &= 12\iint_{[0,1]^2} uv\;\, dC(u,v)-3\\ &= 12\iint_{[0,1]^2} C(u,v)\,du\,dv-3\\ &= 3\alpha\beta/(2\alpha+2\beta-\alpha\beta)\;.\end{aligned} \tag{3.8}$$

Blomqvist [1] proposed a measure of association known as the medical correlation coefficient which we denote by B

$$B = P[(X-X_d)(Y-Y_d)>0]-P[(X-X_d)(Y-Y_d)<0]. \tag{3.9}$$

Theorem 3.3. *Let $C(u,v)$ be a member of two parameter Marshall-Olkin family of copulas with $0<\alpha,\beta<1$.Then*

$$B = 4C(\frac{1}{2},\frac{1}{2})-1 = 4(2)^{\min(\alpha,\beta)\,-2}-1. \tag{3.10}$$

Proof. Since $0\le u,\alpha,\beta\le 1$, it may be noted

$$C(u,u)=\min(u^{2-\alpha},u^{2-\beta})=u^{\max(2-\alpha,2-\beta)}=u^{2-\min(\alpha,\beta)},$$

hence the proof. ∎

For the proofs of Theorems 3.1 and 3.2, reference is made to pages 165 and 168 in Nelson [14].

Gini introduced a measure of association γ which he called the *indice di cograduazione seplice*. Now we prove a theorem expressing Marshall-Olkin copula parameters in terms of association parameter γ.

Theorem 3.4. *Let $C(u,v)$ be a member of two parameter Marshall-Olkin family of copulas with $0<\alpha,\beta<1$. Then Gini's measure of association γ can be expressed as*

$$\begin{aligned}\gamma &= 2\iint_{[0,1]^2} (|u+v-1| - |u-v|)\, dC(u,v) \\ &= 4\left[\int_{[0,1]} C(u,1-u)\, du - \int_{[0,1]} \{u - C(u,u)\}\, du\right] \\ &= \frac{4}{3-\min(\alpha,\beta)} - 2 + 4\big[\frac{1}{3-\alpha}u_0^{3-\alpha} + \frac{1}{3-\beta}(1-u_0)^{3-\beta} - \frac{1}{2-\alpha}u_0^{2-\alpha} \\ &\quad - \frac{1}{2-\beta}(1-u_0)^{2-\beta} + \frac{1}{2-\alpha} + \frac{1}{2-\beta} - \frac{1}{3-\alpha} - \frac{1}{3-\beta}\big]. \end{aligned} \tag{3.11}$$

Proof. Consider

$$\gamma = 4[\int_{[0,1]} C(u,1-u)\, du - \int_{[0,1]} \{u - C(u,u)\, du\}] \tag{3.12}$$

expressed as

$$\gamma = 4\int_{[0,1]} C(u,1-u)\, du - 2 + 4\int_{[0,1]} C(u,u)\, du \tag{3.13}$$

Since $u \in [0,1]$ and $\alpha, \beta \in (0,1)$, we have

$$C(u,u) = \min(u^{2-\alpha}, u^{2-\beta}) = u^{\max(2-\alpha, 2-\beta)} = u^{2-\min(\alpha,\beta)}. \tag{3.14}$$

Hence

$$4\int_{[0,1]} C(u,u)\, du = 4\int_{[0,1]} u^{2-\min(\alpha,\beta)}\, du = \frac{4}{3-\min(\alpha,\beta)}. \tag{3.15}$$

Now

$$C(u,1-u) = \min(u^{1-\alpha}(1-u), u(1-u)^{1-\beta}) = u^{1-\alpha}(1-u)^{1-\beta}\min((1-u)^{\beta}, u^{\alpha}). \tag{3.16}$$

Let u_0 be the solution of $(1-u)^{\beta} = u^{\alpha}$. With our conditions on u, α, and β there is unique such solution. Then

$$\min((1-u)^{\beta}, u^{\alpha}) = \begin{cases} u^{\alpha} & u \le u_0, \\ (1-u)^{\beta} & u > u_0. \end{cases} \tag{3.17}$$

Hence

$$C(u,1-u) = \begin{cases} u(1-u)^{1-\beta} & u \le u_0, \\ u^{1-\alpha}(1-u) & u > u_0. \end{cases} \tag{3.18}$$

$$\begin{aligned}\int_{[0,1]} C(u,1-u)\,du &= \int_{[0,u_0]} u(1-u)^{1-\beta}\,du + \int_{[u_0,1]} u^{1-\alpha}(1-u)\,du \\ &= \frac{1}{3-\alpha}u_0^{3-\alpha} + \frac{1}{3-\beta}(1-u_0)^{3-\beta} - \frac{1}{2-\alpha}u_0^{2-\alpha} \\ &\quad - \frac{1}{2-\beta}(1-u_0)^{2-\beta} + \frac{1}{2-\alpha} + \frac{1}{2-\beta} - \frac{1}{3-\alpha} - \frac{1}{3-\beta}. \end{aligned} \tag{3.19}$$

We can replace $1-u_0$ by $u_0^{\alpha/\beta}$ if this seems more elegant. Finally we reach at the result of the theorem.

∎

Remark 3.1. Concerning u_0, it is the solution of $(1-u)^\beta = u^\alpha$, or equivalently $u^{\alpha/\beta}+u-1=0$. Thus u_0 is a function of α/β and can be tabulate via Maple for example. We can get also an idea of $u_0(\alpha/\beta)$ by plotting its inverse $\alpha/\beta = u_0^{-1} = \frac{\log(1-u)}{\log u}$.

Corollary 3.5. *In case of $\alpha=\beta$, for the single parameter (α) Marshall-Olkin copula $C(u,v)$, dependence measures*

$$\tau = \frac{\alpha}{2-\alpha^2}, \tag{3.20}$$

$$\rho = \frac{3\alpha}{4-\alpha}, \tag{3.21}$$

$$\gamma = \frac{2(2+3\alpha-\alpha^2)-(4-\alpha)2^\alpha}{(2-\alpha)(3-\alpha)}. \tag{3.22}$$

The relationship between the Marshall-Olkin copula parameter concordance measures of Kendall, Spearman and Gini are plotted in Figure 2.

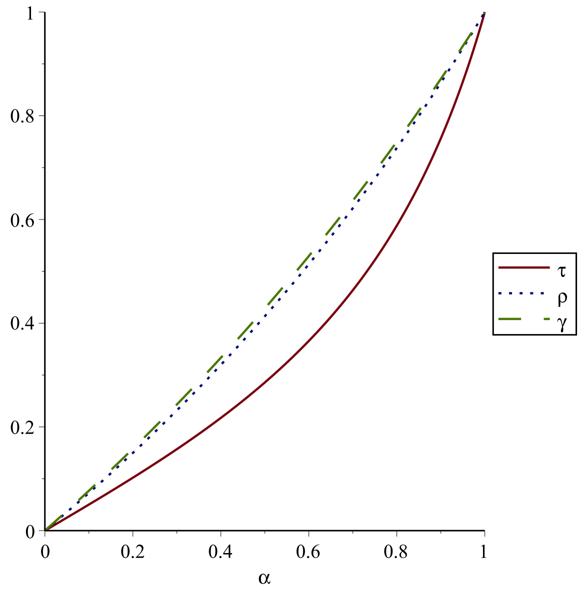


Figure 2: Copula $C(u,v)$ parameter α and concordant measures.

3.3. **Copulas $C(u,v)$ and Tail dependence.** Dependence measure in general are used in describing how large (small) values of one variable appear with large(small) values of the other variable. Concept of tail dependence in a copula $C(u,v)$ measures the dependence between the variables u and v in the upper-right quadrant and lower-left quadrant of $[0,1]^2$. Tail dependence can be used to analyze the dependence among extreme random events. Joe [9] defined the upper and lower tail dependence as

$$\lambda_U = \lim_{t\to 1^-} P\left[Y > G^{-1}(t) | X > F^{-1}(t)\right] = \lim_{t\to 1^-} \frac{1-2t+C(t,t)}{1-t}$$
$$= 2 - \lim_{t\to 1^-} \frac{1-C(t,t)}{1-t},$$
$$\lambda_L = \lim_{t\to 0^+} P\left[Y \le G^{-1}(t) | X \le F^{-1}(t)\right] = \lim_{t\to 1^+} \frac{C(t,t)}{t}. \tag{3.23}$$

These parameters are nonparametric and depend only on copula $C(u,v)$. Copula has lower (upper) tail dependence for $\lambda_U\ (\lambda_L) \in (0,1]$ and no lower (upper) tail dependence for $\lambda_U(\lambda_L) = 0$. This tail dependence measure is the probability that one variable is extreme given that other is extreme. Tail dependence measures are copula-based and copula is related to the full distribution via quantile transformations, i.e.,

$$C(u,v) = F\left(F^{-1}\ (u),\ G^{-1}(v)\right). \tag{3.24}$$

The tail dependence for Marshall-Olkin copulas are given in the following theorem:

Theorem 3.6. *Let $C(u,v)$ be a member of two parameter Marshall-Olkin family of copulas with $0 < \alpha, \beta < 1$. Then the copula $C(u,v)$ has lower and upper tail dependence parameters $\lambda_L = 0$ and $\lambda_U = \min(\alpha,\beta)$.*

Proof. From the definition of tail dependence parameters,

$$\lambda_L = \lim_{t\to 0^+} \frac{C(t,t)}{t} = \lim_{t\to 0^+} \frac{t^2 - \min(\alpha,\beta)}{t} = 0, \tag{3.25}$$

This is because $1-\min(\alpha,\beta) > 0$. Further,

$$\lambda_U = 2 - \lim_{t\to 1^-} \frac{1-C(t,t)}{1-t} = \lim_{t\to 1^-} \frac{1-t^{2-\min(\alpha,\beta))}}{1-t}, \tag{3.26}$$

and by the l'Hospital's rule,

$$\lambda_U = 2 - \lim_{u\to 1^-} \frac{-(2-\min(\alpha,\beta))u^{1-\min(\alpha,\beta)}}{-1} = \min(\alpha,\beta). \tag{3.27}$$

Thus, Marshall-Olkin copulas have no lower tail dependence but have the upper tail dependence which is qual to $\min(\alpha,\beta)$. ∎

3.4. **Copulas $C(u, v)$ and the Mutual information.** Uncertainty prevails in several forms and various kinds of uncertainties may arise from random fluctuations, incomplete information, imprecise perception, vagueness etc. Probabilistic uncertainty may be viewed as one associated with a random outcome of an experiment or which is associated with the manner in which data are collected and analyzed following statistical designs. Often such type of uncertainty is summarized in terms of bias, standard error, and measures based on the statistical probability distributions. Shannon [16] laid the mathematical foundation of information theory in the context of communication theory and defined a probabilistic measure of uncertainty referred to as entropy. However earlier contributions in this direction have been noted due to Nyquist [15] and Hartley [6]. The entropy measures the expected uncertainty in a variable X and is $H(X) = -\int f(x) \log f(x)\, dx$. For two continuous random variables X and Y having the marginal density functions $f(x)$ and $f(y)$ and jointly distributed according to the joint density function $f(x, y)$, then

Definition 3.1. The joint entropy of two random variables X and Y having the joint density function $f(x, y)$ is

$$H(X, Y) = -\iint f(x, y) \log f(x, y)\, dx\, dy, \tag{3.28}$$

and the conditional entropy is

$$H(X|Y) = -\iint f(x, y) \log f(x|y)\, dx\, dy, \tag{3.29}$$

where $f(x|y)$ is the conditional density function of X given Y.

The conditional entropy is a measure of how much uncertainty remains about X when we know the value of Y.

The mutual information or information transmission (distance from statistical independence) between X and Y in terms of the Kullback-Liebler divergence [11] is:

Definition 3.2. The mutual information between two random variables X and Y with joint density function $f(x, y)$ and marginal density functions $f(x)$ and $f(y)$ is

$$I(X, Y) = -\iint f(x, y) \log \frac{f(x, y)}{f(x)f(y)}\, dx\, dy. \tag{3.30}$$

Definition 3.3. The joint probability density function $f(x, y)$ can be expressed in terms of copula density function $c(u, v)$ as

$$f(x, y) = c(u, v)f(x)f(y), \tag{3.31}$$

where copula density function $c(u, v)$ is

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}. \tag{3.32}$$

We thus have the following theorem about the Marshall-Olkin copula $C(u, v)$:

Theorem 3.7. *Let $C(u, v)$ be a member of two parameter Marshall-Olkin family of copulas with $0 < \alpha, \beta < 1$. Then the mutual information between two random variables X and Y is*

$$I(X,Y) = -\iint c(u,v)\log c(u,v)\,du\,dv$$

(3.33) $$= -\left[\frac{(1-\alpha)\beta\log(1-\alpha)+(1-\beta)\alpha\log(1-\beta)}{\alpha-\alpha\beta+\beta}+\frac{\alpha\beta(\alpha-2\alpha\beta+\beta)}{(\alpha-\alpha\beta+\beta)^2}\right].$$

Proof. For the Marshall-Olkin copula $C(u,v)$

(3.34) $$C(u,v) = \left[\begin{array}{ll} u^{1-\alpha}v, & u^\alpha \geq v^\beta, \\ uv^{1-\beta}, & u^\alpha \leq v^\beta, \end{array}\right.$$

we get

(3.35) $$c(u,v) = \begin{cases} u^{-\alpha}(1-\alpha), & u^\alpha \geq v^\beta, \\ v^{-\beta}(1-\beta), & u^\alpha \leq v^\beta. \end{cases}$$

Let $[0,1]^2 = E \cup F$, where

(3.36) $$E = \{(u,v) \mid 0 \leq u \leq 1, 0 \leq v \leq u^{\frac{\alpha}{\beta}}\},$$

(3.37) $$F = \{(u,v) \mid 0 \leq u \leq 1, u^{\frac{\alpha}{\beta}} \leq v \leq 1\} = \{(u,v) \mid 0 \leq v \leq 1, 0 \leq u \leq v^{\frac{\beta}{\alpha}}\}.$$

Thus,

(3.38) $$c(u,v) = u^{-\alpha}(1-\alpha) \text{ on } E \text{ and } v^{-\beta}(1-\beta) \text{ on } F.$$

E and F overlap only on the boundary, but it is clear from the calculations that we can ignore that fact. Set

(3.39) $$I(X,Y) = -\iint_{[0,1]^2} c(u,v)\log c(u,v)\,du\,dv = -\left(I_1+I_2\right),$$

where

$$I_1 = \iint_E c(u,v)\log c(u,v)\,du\,dv = \int_0^1\int_0^{u^{\frac{\alpha}{\beta}}} u^{-\alpha}(1-\alpha)\log(u^{-\alpha}(1-\alpha))\,dv\,du,$$

$$I_2 = \iint_F c(u,v)\log c(u,v)\,du\,dv = \int_0^1\int_0^{v^{\frac{\beta}{\alpha}}v^{-\beta}} (1-\beta)\log(v^{-\beta}(1-\beta))\,du\,dv.$$

Due to symmetry we can calculate one of these integrals, then interchange α with β to obtain the second.

$$I_1 = \int_0^1 u^{\frac{\alpha}{\beta}-\alpha}(1-\alpha)\log((1-\alpha)u^{-\alpha})\,du$$

$$= (1-\alpha)\log(1-\alpha)\int_0^1 u^{\frac{\alpha}{\beta}-\alpha}\,du - \alpha(1-\alpha)\int_0^1 u^{\frac{\alpha}{\beta}-\alpha}\log u\,du$$

(3.40) $$= \frac{(1-\alpha)\beta\log(1-\alpha)}{\alpha-\alpha\beta+\alpha} + \frac{\alpha(1-\alpha)\beta^2}{(\alpha-\alpha\beta+\beta)^2}.$$

Hence the required result. Note that in these computations we use the fact that $\lim_{u\to 0^+} u^\delta \log u = 0$ for $\delta > 0$. ∎

Theorem3.7 results in the following corollary in case of $\alpha = \beta$:

Corollary 3.8. *Let $C(u,v)$ be a single parameter (α) Marshall-Olkin copula. Then the mutual information between two random variables X and Y is*

(3.41) $$I(X,Y) = -2\frac{1-\alpha}{2-\alpha}\log(1-\alpha) + \frac{2\alpha(1-\alpha)}{(2-\alpha)^2}.$$

Mercier [13] has also arrived at this corollary result.

3.5. **Algorithm to simulate Marshall-Olkin copula $C(u,v)$.** For the Marshall-Olkin bivariate exponential distribution with parameters λ_1, λ_2 and λ_{12}, the following algorithm due to Devroye [4] simulates uniform random variables U and V whose joint distribution function is a Marshall-Olkin copula $C(u,v)$:

Step 1. Generate three independent uniform $(0,1)$ random variables p, q and r.

Step 2. Let $x = \min(\frac{-\ln p}{\lambda_1}, \frac{-\ln r}{\lambda_{12}})$ and $y = \min(\frac{-\ln q}{\lambda_2}, \frac{-\ln r}{\lambda_{12}})$.

Step 3. The desired pair of variables (u,v) is obtained from $u = \exp[-(\lambda_1 + \lambda_{12})x]$ and $v = \exp[-(\lambda_2 + \lambda_{12})x]$.

In Figure 3, we present the scatter plot of 500 pairs of (u,v) simulated using the Marshall-Olkin copula for $\alpha = 0.3$ and $\beta = 0.75$. The singular component is clearly seen in this figure.

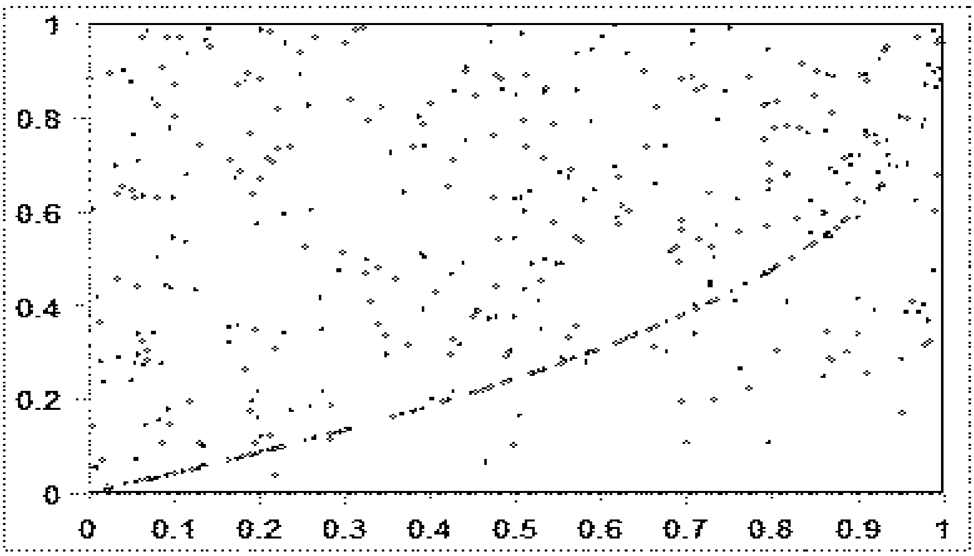


Figure 3: Scatterplot of 500 pairs of (u,v) simulated usuing the marshal-Olkin copula for $\alpha = 0.3,\ \beta = 0.75$.

4. APPLICATION OF THE MARSHALL-OLKIN COPULA BASED MUTUAL INFORMATION

We consider the data arising in the production of nitric acid in the process of oxidizing ammonia [2]. The response variable is stack loss which is expressed as the percentage of the ingoing ammonia that escapes unabsorbed and key process variables are the airflow, the cooling water inlet temperature in ^0C and the acid concentration is percent.

Table 2. Mutual information using response variable stack loss and process variables air flow, cooling temperature and acid concentration in the process of oxidizing ammonia.

	Air flow,%	Cooling temperature, ^0C	Acid concentration,%
Kendall τ	0.802	0.734	0.379
$C(u,v)$ parameter α	0.92209	0.88852	0.61476
Mutual information $I(X,Y)$	0.7774	0.60045	0.4926

In Table 2, concordance measure Kendall's τ between stack loss and process variables air flow, cooling temperature and acid concentration respectively are 0.802, 0.734 and 0.379 while the Marshall-Olkin copula with single parameter α values estimated using the Kendall's τ for these process variables are 0.92209, 0.88852 and 0.61476. The values of mutual information measure $I(X,Y)$ in Table 2 in respect of these process variables are 0.7774, 0.60045 and 0.4926 respectively. Thus, mutual information measure indicates that the information about the amount of decrease in stack loss caused by the knowledge of airflow is maximum followed by the knowledge of the cooling temperature and acid concentration. Higher values of the copula parameter α, i.e., 0.92209 and 0.88852 for air flow and cooling temperature are the indication of the higher degree of association between stack loss and air flow and cooling temperature. Thus, in this application, air flow and cooling temperature can serve as the best predictors in predicting the stack loss. Based on these two predictors, the least squares prediction line for stack loss is estimated as

$$\text{Stack loss} = -5.036 + 0.0674 \text{ Air flow} + 0.130 \text{ Cooling temperature};$$
$$Adj.\ R^2 = 89.9\%,$$

where both predictor variables are highly significant at 1% significance level. For comparison, the least squares prediction line for stack loss using all three predictors is estimated as

$$\begin{aligned}\text{Stack loss} &= 3.614 + 0.072 \text{ Air flow}\\ &\quad + 0.130 \text{ Cooling temperature } - 0.152 \text{ Acid concentration };\\ &\quad Adj.\ R^2 = 89.8\%,\end{aligned}$$

where predictor variables air flow and cooling temperature are highly significant at 1% significance level however acid concentration is not significant because of its P-value being 0.344. It may however be noted that the Marshall-Olkin copulas model only the positive dependence because its parameter lies on the interval $(0,1)$.

REFERENCES

[1] N. BLOMQVIST, On a measure of dependence between two random variables, *Ann. Math. Statist.*, 21 (1950), pp. 593-600.

[2] K. A. BROWNLEE, *Statistical Theory and Methodology in Science and Engineering*, second ed., Wiley, London, 1965.

[3] C. M. CUADRAS, J. AUGÉ, A continuous general multivariate distribution and its properties, *Comm. Statist. A - Theory Methods*, 10 (1981), pp. 339-353.

[4] L. DEVROYE, *Nonuniform Random Variate Generation*, Springer, New York, 1986.

[5] P. EMBRECHTS, F. LINDSKOG, A. MCNEIL, *Modeling dependence with copulas and applications to risk management*, in: S. Rachev (Ed.), *Handbook of Heavy Tailed Distributions in Finance*, Elsevier, New York, 2003, pp. 329-384.

[6] R. V. L. HARTLEY, Transformation of information, *Bell Systems Tech. J.*, 7 (1928), pp. 535-563.

[7] W. HOEFFDING, Masstabinvariance korrelationsmasse, *Schriften des Mathematischen Instituts für Angewandte Mathematik der Universität Berlin*, 5, 3 (1940), pp. 179-233.

[8] T. P. HUTCHINSON, C. D. LAI, *Continuous Bivariate Distributions Emphasizing Applications*, Rumsby Scientific Publishing, Adelaide, 1990.

[9] H. JOE, *Multivariate Models and Dependence Concepts*, Chapman & Hall, NY 1997.

[10] K. JOGDEO, *Concepts of dependence*, in S. Kotz, N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences*, Vol. 1, Wiley, New York, 1982, pp. 324-334.

[11] S. KULLBACK, R. A. LEIBLER, On information and sufficiency, *Ann. Math. Statist.*, 22 (1951), pp. 79-86.

[12] A. W. MARSHALL, I. OLKIN, Families of multivariate distributions, *J. Amer. Statist. Assoc.*, 83 (1988), pp. 834-841.

[13] G. MERCIER, Measures de Dépendance entre Images RSO, GET/ENST Bretagne, *Tech. Rep. RR-2005003-ITI*, 2005, `http//perso.enst-bretagne.fr/126mercierg`.

[14] R. B. NELSON, *An Introduction to Copulas*, second ed., Springer, New York, 2006.

[15] H. NYQUIST, Certain topics in telegraph transmission theory, *Trans. AIEE*, 47 (1928), pp. 617-644. Reprint as classic paper in: *Proc. IEEE* **90**, 2 (2002).

[16] C. E. SHANNON, *A Mathematical Theory of Communication-An Integrated Approach*, Cambridge University Press, 1948.

[17] B. SCHWEIZER, E. F. WOLFF, On nonparametric measures of dependence for random variables, *Ann. Statist.*, 9 (1981), pp. 879-885.

[18] A. SKLAR, Fonctions de répartition à n dimensional et leurs marges, *Publ. Inst. Stat. Univ. Paris*, 8 (1959), pp. 229-231.