In The Name of Allah

Proceedings of the

2nd Workshop on Copula and its Applications

Ordered and Spatial Data Center of Excellence Ferdowsi University of Mashhad

29-30 May, 2012

Faculty of Mathematical Sciences
Ferdowsi University of Mashhad
Mashhad, Iran

Scientific Committee:

Amini, M. (Chairman)	**Ferdowsi University of Mashhad**
Alamatsaz, M.H.	**Isfahan**
Dolati, A.	**Yazd University**
Jabbari Nooghabi, H.	**Ferdowsi University of Mashhad**

Organizing Committee:

Ahmadi, J.	**Ferdowsi University of Mashhad**
Amini, M.	**Ferdowsi University of Mashhad**
Jabbari Nooghabi, H.	**Ferdowsi University of Mashhad**
Mohtashami Borzadaran, G.R.	**Ferdowsi University of Mashhad**

Student Committee:

Ahmadzade, H.	**Ferdowsi University of Mashhad**
Okhli, K.	**Ferdowsi University of Mashhad**
Mohtashami, H.A.	**Ferdowsi University of Mashhad**

Topics of the Workshop:

- History of Copula
- Construction of Copula
- Copulas and Dependence Concepts
- Modelling using Copula
- Applications of Copula
- Copula and Spatial Statistics
- Software Concepts of Copula

Preface

On behalf of the organizing and scientific committees, we would like to extend a very warm welcome to all the participants of the 2nd Workshop on Copula and its Applications.

We hope this seminar provides an environment of useful discussions and would also exchange scientific ideas and opinions.

The program of this seminar is organized in 5 key speakers and 14 oral contributions.

We wish to express our gratitude to the numerous individuals and organizations that have contributed to the success of this workshop, in which more than 50 colleagues, researchers, and postgraduate students have participated.

Finally, we would like to extend our sincere gratitude to the students of the Faculty of Mathematical Sciences at Ferdowsi University of Mashhad for their kind help and cooperation. We wish them all success.

Ordered and Spatial Data
Center of Excellence

Contents

Constructing a generalized family of dependent bivariate distributions

Amini, M. , Dolati, A. and Mirhoseini S.M.
Department of Statistics, Ferdowsi University of Mashhad, Meshhad, Iran,
Department of Statistics, Faculty of Mathematics, Yazd University, Yazd, Iran,

Abstract *A generalized family of bivariate distribution functions is constructed by a stochastic mixture of the order and record statistics of a sample of size two, Generalized Linear means method and The Ruschendorf Method. Various properties of the proposed models are studied. Also, dependence structure and formulas for dependence measures are given via the copula of these distributions.*

Keywords *Copula function, Ordered Statistics, record Statistics, Associated Measures* .

1 Introduction

The various ways to construct a bivariate distributions have been investigated by authors. One of the most popular bivariate distributions is The FGM distribution that has been studied by Farli (1960), Gumbel (1960) and Morgensten (1956). Also, the FGM family generalized by many authors for example,Lai and Xie (2000), Fischer and Klein (2007),Sarmanov (1960), Lallena and Flores (2004),uang and Kotz (1999),Bairamov and Kotz (2002), Baker (2008),Dolati, Mirhoseini and Amini (2011) and Amblard and Girard (2009). Constructing of the bivariate FGM based on stochastic mixture method studied by Baker (2008). Noting that, for continuous marginals distribution functions Pearson correlation in the FGM is restricted to $[-\frac{1}{3}, \frac{1}{3}]$. Many authors, generalized the FGM distribution for increase of the limit of Pearson correlation and have been studied the dependence structure it. There are several of new approach to the construction of bivariate distributions, for instance, Baker, (2008), Lin and Huang, (2000), Dolati, Mirhoseini, and Amini (2011).
This paper is organized as follows: In Section 2, we introduce a generalized family of bivariate distributions. Dependence structure of generalized models is studied, also, some concepts of dependence stochastic orders for these bivariate distribution are derived. We apply the stochastic mixture method based on order and record statistics for constructing some sub-families of proposed model in Section 3. Moreover, in view of the Ruschendorf Method we derive some sub-families, particularly the family studied by Lallena and Flores (2004). The generalized linear means method introduced by Klein and Christa (2012) evaluate finally.

2 Generalized bivariate distribution

The approaches here allows construction of extended bivariate distribution functions as:

$$H(x, y) = F(x).G(y) + W_\theta(F(x), G(y)), \qquad (1)$$

where θ is suitable parameter such that $H(x, y)$ is a distribution function and $W(x, y)$ satisfying the following conditions for all $x, y \in [0, 1]$:

- $W(x, y) \geq 0$.
- $W(x, 0) = W(0, y) = 0$.
- $W^{12}(x, y) \geq -1$, $W^1(x, 1) \geq -1$ and $W^2(1, y) \geq -1$.
- $\int_0^1 \int_0^1 W^{12}(x, y)dxdy = 0$.

Where $W^{12}(x,y)=\frac{\partial^2 W(x,y)}{\partial x \partial y}$, $W^1(x,y)=\frac{\partial W(x,1)}{\partial x}$ and $W^2(x,y)=\frac{\partial W(1,y)}{\partial y}$. The suitable functions that satisfy in the above conditions called kernel functions. In addition, θ is a parameter that shows dependence structure of $H(x,y)$ and $\theta=0$, leads to the independence of X and Y. Also, It is easy to check that the marginal distributions, joint density function and marginal density function of Model (1) are as follows respectively:

$$H_1(x)=F(x)+W_\theta(F(x),1), \qquad H_2(x)=G(y)+W_\theta(1,G(y)).$$

$$h(x,y)=f(x).g(y)[1+W_\theta^{12}(F(x),G(y))],$$

$$h_1(x)=f(x)[1+W_\theta^1(F(x),1)]$$

and

$$h_2(y)=g(y)[1+W_\theta^2(1,G(y))].$$

The conditional density functions $Y|X=x$ and $X|Y=y$ are as follow respectively:

$$h(y|x)=\frac{1+W_\theta^{12}(F(x),G(y))}{1+W_\theta^1(F(x),1)}.g(y)$$

and

$$h(x|y)=\frac{1+W_\theta^{12}(F(x),G(y))}{1+W_\theta^1(1,G(y))}.f(x).$$

Also, the joint survival function obtain as:

$$\begin{aligned}\overline{H}(x,y) &= P[X>x,Y>y]\\ &= \overline{F}(x)\overline{G}(y)-W_\theta(F(x),1)-W_\theta(1,G(y))+W_\theta(F(x),G(y)).\end{aligned}$$

Copula function
assuming $\psi_1(t) = t + W_\theta(t, 1)$ and $\psi_2(t) = t + W_\theta(1, t)$, the copula corresponding to the model given in (1)obtain as:

$$C(\psi_1(u), \psi_2(v)) = uv + W_\theta(u, v), \quad u, v \in [0, 1].$$

2.1 Some Concepts of Dependence

Let X and Y be two random variables with joint distribution (density) function $H(x, y)$ $(h(x, y))$ and marginals $H_1(x)$ and $H_2(y)$ respectively. Several concepts of positive(negative) dependence and dependence stochastic orders have been introduced in literatures. We recall some this concepts and then, we study dependence structure of family given in (1).
The random variables X and Y are said to be:

- Positive likelihood ratio dependent $(PLRD(X, Y))$ if $h(x, y)$ is total positive of order two (TP2) means, for all $x_1 \leq x_2, y_1 \leq y_2$

$$h(x_1, y_1)h(x_2, y_2) - h(x_1, y_2)h(x_2, y_1) \geq 0.$$

- Negative likelihood ratio dependent $(NLRD(X, Y))$ if $h(x, y)$ is Rivers regular of order two (RR2), means, for all $x_1 \leq x_2, y_1 \leq y_2$

$$h(x_1, y_1)h(x_2, y_2) - h(x_1, y_2)h(x_2, y_1) \leq 0.$$

- Stochastically decreasing (increasing) $SD(Y|X)(SI(Y|X))$ if $P[Y > y|X = x]$ is non-increasing(non-decreasing) in x for all y.
- Left corner set decreasing(increasing) $LCSD(X, Y)(LCSI(X, Y))$ if $P[X \leq x, Y \leq y|X \leq t, Y \leq s]$ is non-increasing(non-decreasing) in t and s for all x and y.
- Let (X_1, X_2) and (Y_1, Y_2) be two random vectors with respective distribution functions $F(t, s)$ and $G(t, s)$ belong to Frechet class of all bivariate distributions with the univariate marginals F_1, F_2. We say that (X_1, X_2) smaller than (Y_1, Y_2) in the lower orthant decreasing ratio order (denoted by $(X_1, X_2) \leq_{lodr} (Y_1, Y_2)$ or $F \leq_{lodr} G$) if

$$\frac{G(t, s)}{F(t, s)} \;\; is \;\; decreasing \;\; in \;\; (t, s) \in \{(t, s) : G(t, s) > 0\}.$$

  Nothing that, for a parametric family of bivariate distributions $H_\theta(x, y)$ we have
  $H_{\theta_1} \leq_{lodr} H_{\theta_2}$ for all $\theta_1 \leq \theta_2$ if

$$\frac{H_{\theta_2}(t, s)}{H_{\theta_1}(t, s)} \;\; is \;\; increasing \;\; in \;\; (t, s).$$

i) If $1+W_\theta^{12}(x, y)$ is $TP2(RR2)$, then $h(x, y)$ is $TP2(RR2)$, consequently we have $PLRD(X, Y)(NLRD(X, Y))$ respectively.

ii) Let $x_1 \leq x_2$, $y_1 \leq y_2$ and $x_1W_\theta(x_2, y) = x_2W_\theta(x_1, y)$ or $y_1W_\theta(x, y_2) = y_2W_\theta(x, y_1)$ for all $x, y \in (0, 1)$. Then $W_\theta(x, y)$ is TP2(RR2) if and only if $H_\theta(x, y)$ TP2(RR2) consequently we have $LCSD(X, Y)(LCSI(X, Y))$.

iii) If for all $t \in R$, $\frac{G(t)+W_\theta^1(F(x),G(t))}{1+W_\theta^1(F(x),1))}$, increasing (decreasing) in x, then $NRD(Y|X)(PRD(Y|X))$. Similarly, $\frac{F(t)+W_\theta^2(F(t),G(y))}{1+W_\theta^2(1,G(y))}$, increasing (decreasing) in y, then $SD(X|Y)(SI(X|Y))$.

i) The bivariate distributions family generated by
$W_\theta(x, y) = \theta xy(1 - x)(1 - y)$, having to positive (negative) dependence structure if and only if $0 \leq \theta \leq 1(-1 \leq \theta \leq 0)$.

ii) The bivariate distributions family generated by $W_\theta(x, y) = x.y(\theta(2 - x - y) + (1 - x)(1 - y))$, having to positive dependence structure if and only if $0 \leq \theta \leq 1$.

iii) The bivariate distributions family generated by $W_\theta(x, y) = xy(\theta(y - x) - (1 - x)(1 - y))$, having to negative dependence structure if and only if $0 \leq \theta \leq 1$.

Let $H_\theta(x, y) = F(x)G(y) + \theta W(F(x), G(y))$, and $W(x, y)$ be decreasing in x, y. Then for all $0 \leq \theta_1 < \theta_2 \leq 1$,

$$H_{\theta_1} \preceq^{lodr} H_{\theta_2}.$$

Proof. Let $\theta_1 < \theta_2$ and $(x_1, x_2) \prec (y_1, y_2)$ (i.e. $x_i < y_i, \quad i = 1, 2$), we obtain

$$\begin{aligned} & H_{\theta_1}(x_1, x_2).H_{\theta_2}(y_1, y_2) - H_{\theta_1}(y_1, y_2).H_{\theta_2}(x_1, x_2) \\ = & (\theta_2 - \theta_1)[F(x_1)G(x_2)W(F(y_1), G(y_2)) - F(y_1)G(y_2)W(F(x_1), G(x_2))] \leq 0, \end{aligned}$$

□

In Particulate if $W(x, y) = xy\psi(x, y)$, then for all $\theta_1 < \theta_2$,

$$H_{\theta_1} \preceq^{lodr} H_{\theta_2} \Longleftrightarrow \psi(x, y) \searrow \quad in \quad x, y,$$

2.2 Measures of Association

In this section, we compute four measures of association, Kendall's tau, Spearman's rho, tail dependence coefficients and extremal dependence coefficients for family given (1). Let $H(x, y)$ be a absolutely continuous bivariate distribution whit margins $H_1(x)$ and $H_2(y)$, then

$$\int\int H(x, y)dH(x, y) = \frac{1}{2} - \int\int \frac{\partial H(x, y)}{\partial x}.\frac{\partial H(x, y)}{\partial y}dxdy.$$

Let X and Y be a random variables with joint distribution $H(x, y)$, margins $H_1(x)$ and $H_2(y)$ respectively,then

$$\begin{aligned} \tau &= 4\int\int H(x, y)dH(x, y) - 1 \\ &= 1 - 4\int\int \frac{\partial H(x, y)}{\partial x}.\frac{\partial H(x, y)}{\partial y}dxdy. \end{aligned}$$

and

$$\begin{aligned} \rho_s &= 12\int\int H(x, y)dH^I(x, y) - 3 \\ &= 12\int\int H^I(x, y)dH(x, y) - 3. \end{aligned}$$

Moreover, applying Lemma 1, we can derive another formula for ρ_s as:

$$\rho_s = 3 - 6\int\int [\frac{\partial H(x, y)}{\partial x}.\frac{\partial H^I(x, y)}{\partial y} + \frac{\partial H^I(x, y)}{\partial x}.\frac{\partial H(x, y)}{\partial y}]dxdy.$$

Where, $H^I(x, y) = H_1(x).H_2(y)$. Let $H(x, y)$ be a bivariate distribution given as (1). Then
i)

$$\begin{aligned} \tau(\theta) &= 4\int_0^1\int_0^1 [W_\theta(x, y) + xy.W_\theta^{12}(x, y) + W_\theta(x, y).W_\theta^{12}(x, y)]dxdy \\ &= -4\int_0^1\int_0^1 [x.W_\theta^1(x, y) + y.W_\theta^2(x, y) + W_\theta^1(x, y).W_\theta^2(x, y)]dxdy. \end{aligned}$$

ii)

$$\begin{aligned}\rho_s(\theta) &= 12\int_0^1\int_0^1 [(x+W_\theta(x,1))(y+W_\theta(1,y))(1+W_\theta^{12}(x,y))]dxdy-3\\ &= 12\int_0^1\int_0^1 [(xy+W_\theta(x,y))(1+W_\theta^{2}(1,y))(1+W_\theta^{1}(x,1))]dxdy-3.\end{aligned}$$

Moreover, another formula for ρ_s is as:

$$\begin{aligned}\rho_s(\theta)=3 &- 6\int_0^1\int_0^1 [x+W_\theta^2(x,y)][1+W_\theta^1(x,1)][y+W_\theta(1,y)]dxdy\\ &- 6\int_0^1\int_0^1 [y+W_\theta^1(x,y)][1+W_\theta^2(1,y)][x+W_\theta(x,1)]dxdy.\end{aligned}$$

For many families of bivariate distributions, that exhibit weak dependence, the sample value of Spearman's rho is about 50 percent larger than the sample value of Kendall's tau. Fredricks and Nelsen (2007)showed that for the population versions of these statistics, the ration of $\frac{\rho_s(\theta)}{\tau(\theta)}$ approaches $\frac{3}{2}$ as the joint distribution approaches that of two independent random variables. We also, derive the direction of inequality between 3τ and $2\rho_s$ for the family of bivariate distribution given as (1), under some suitable conditions on $W_\theta(x,y)$. Let $H(x,y)$ be an absolutely continuous bivariate distribution given as (1). Suppose that, for all $x,y\in[0,1]$,

$$A(x,y)=W_\theta(x,y)-y.W_\theta(x,1)-x.W_\theta(1,y)-W_\theta(x,1)W_\theta(1,y)\neq 0.$$

i) If $\frac{\partial^2 \ln|A(x,y)|}{\partial x\partial y}\geq 0$ then $2\rho\leq 3\tau$.
ii) If $\frac{\partial^2 \ln|A(x,y)|}{\partial x\partial y}\leq 0$ then $2\rho\geq 3\tau$.
ii) If $\frac{\partial^2 \ln|A(x,y)|}{\partial x\partial y}= 0$ then $2\rho= 3\tau$.

Proof. Using various forms for τ and ρ given in Proposition 4, we have

$$\int\int_{R^2}[H(x,y)\frac{\partial^2 H(x,y)}{\partial x.\partial y}-\frac{\partial H(x,y)}{\partial x}\frac{\partial H(x,y)}{\partial y}]dxdy=\frac{\tau}{2},$$

and

$$\int\int_{R^2}[H^I(x,y)\frac{\partial^2 H(x,y)}{\partial x.\partial y}+H(x,y)-\frac{\partial H(x,y)}{\partial x}\frac{\partial H^I(x,y)}{\partial y}-\frac{\partial H^I(x,y)}{\partial x}\frac{\partial H(x,y)}{\partial y}]dxdy=\frac{\rho_s}{3}.$$

So, elementary calculus yield,

$$\begin{aligned}\frac{\tau}{2}-\frac{\rho}{3} &= \int\int_{R^2}(H(x,y)-H^I(x,y))^2\frac{\partial^2\ln(H(x,y)-H^I(x,y))}{\partial x\partial y}dxdy\\ &= \int_0^1\int_0^1 A^2(x,y)\frac{\partial^2 \ln A(x,y)}{\partial x\partial y}dxdy.\end{aligned}$$

This complete the proof. □

If $A(x,y)=\theta\phi(x)\psi(y)$, then $\frac{\partial^2 \ln|A(x,y)|}{\partial x\partial y}=0$ so $2\rho=3\tau$. An alternate proof of Daniel's inequality between ρ_s and τ, as $-1\leq 3\tau-2\rho\leq 1$, given in Theorem 5.1.10 of Nelsen (2006). If $-\frac{1}{12}\leq\int_0^1\int_0^1 B(x,y)dxdy\leq\frac{1}{12}$. Where for all $x,y\in[0,1]$,

$$\begin{aligned}B(x,y) &= [W(x,1)+W^2(1,y)+W^2(1,y)-W^2(x,y)]\\ &\times\ [W(1,y)+W^1(x,1)+W^1(x,1)-W^1(x,y)].\end{aligned}$$

Then $-1 \leq 3\tau - 2\rho \leq 1$. Applying formulas of ρ and τ in proposition 4, it is easy to show that

$$\begin{aligned}\int\int_{R^2}[\frac{\partial H^I(x,y)}{\partial x}-\frac{\partial H(x,y)}{\partial x}].[\frac{\partial H^I(x,y)}{\partial y}\frac{\partial H(x,y)}{\partial x}]dxdy &= \int_0^1\int_0^1 B(x,y)dxdy \\ &= \frac{2\rho-3\tau}{12}.\end{aligned}$$

Tail dependence coefficients Another dependence coefficient is tail dependence, which measures the dependence between the variables in the upper-right quadrant and in the lower-left quadrant of $I^2 = [0,1] \times [0,1]$. Let (X, Y) be a random vector with joint distribution function H and marginals H_1 and H_2. The quantity

$$\begin{aligned}\lambda_u &= \lim_{t\to 1^-} P\left(H_2(X) > t | H_1(Y) > t\right) \\ &= \lim_{t\to 1^-}\frac{\hat{C}(t,t)}{1-t} \\ &= 1 - \lim_{t\to 1^-}\frac{t-C(t,t)}{1-t},\end{aligned}$$

is called the upper tail dependence coefficient (UTDC) provided the limit exists. We say that (X, Y) has upper tail dependence if $\lambda_u > 0$ and upper tail independent if $\lambda_u = 0$. Similarly, we define the lower tail dependence coefficient (LTDC) by

$$\lambda_l = \lim_{t\to 0^+} P\left(H_2(X) \leq t | H_1(Y) \leq t\right) = \lim_{t\to 0^+}\frac{C(t,t)}{t}.$$

For more details, see Coles et al. (1999). If $W_\theta(x,y) = \theta xy\psi(x,y)$, Then $\lambda_l = \lambda_u = 0$.

Proof. Set $\phi_1^{-1}(u) = u + \theta u\psi(t,1)$, $\quad \phi_2^{-1}(u) = u + \theta u\psi(1,t)$, we obtain

$$C(u,v) = \phi_1^{-1}(u)\phi_2^{-1}(u)[1+\theta\psi(\phi_1^{-1}(u),\phi_2^{-1}(u))],$$

Now Simple calculation complete the proof. □

3 Sub-families and examples

The corresponding sub-families and examples of bivariate distributions in each case studied in this section. A well-known limitation to the family given in (1) is that it dose not allow the modelling of large dependences since the maximum Pearson correlation is not achieve to 1. Based on this remark, several generalization of $W(x,y)$ have been introduced by researchers in recent years that given some of them in the following.

- One of the most popular parametric family bivariate distributions is the Farlie-Gumbel-Morgenstern (FGM) family defined when $W_\theta(x,y) = \theta.xy[1-x][1-y]$. This family of bivariate distribution discussed by Morgenstren (1956), Farlie (1960) and Gumbel (1958). In this family of bivariate distribution $\rho = \frac{2}{9}\theta$ and $\rho_s = \frac{\theta}{3}$.

- If $W_\theta(x,y) = \theta\varphi(x)\psi(y)$, where φ and ψ are functions defined on $I = [0,1]$, such that satisfy in the Lipschitz condition and $\varphi(0) = \varphi(1) = \psi(0) = \psi(1) = 0$. The family of bivariate distributions generated by $W_\theta(x,y)$ introduced Sarmanov (1960) and re-discovered by Lallena and Flores (2004).

- If $W_\theta(x,y) = \theta(xy)^p[(1-x)(1-y)]^q$, where $p, q > 1$. The family of distributions generated by this function studied in Lai and Xie (2000).

- The kernel extensions of FGM are studied by Huang and Kotz (1999), if $W_\theta(x,y) = \theta(xy)[(1-x)(1-y)]^q$ for $q \geq 1$ and $W_\theta(x,y) = \theta(xy)[(1-x^q)(1-y^q)]$ for $q \geq \frac{1}{2}$. It is proved that positive correlation can be increased up to approximatively 0.39 while the maximal negative correlation remains $-1/3$.

- Amblard and Girard (2009) considered the family of functions defined on I^2 as follow:

$$C_{\theta,\varphi}(u,v) = uv + \theta[\max\{u,v\}]\varphi(u)\varphi(v),$$

where θ and φ satisfy in suitable conditions, such that this function was a copula.They Studied dependence structure the family of distribution generated by this copula function.

- If $W_\theta(x,y) = (1-\theta)x.y + \frac{\theta}{n}\sum_{k=1}^{n}[\sum_{j=k}^{n} C_n^k x^j(1-x)^{n-j}.\sum_{j=k}^{n} C_n^k y^j(1-y)^{n-j}]$, for all $\theta \in [0,1]$, then we obtain the bivariate distributions family with positive dependent structure introduced by Baker (2008). In particular, if

$$W_\theta(x,y) = (1-\theta)x.y + \frac{\theta}{n}[x^n(1-(1-y)^n) + y^n(1-(1-x)^n)]$$

we get a bivariate distributions family induced by mixture of (X,Y) with $(X_{(1)}, Y_{(n)})$ and $(Y_{(1)}, X_{(n)})$.

- Lin and Huang (2010) investigated Baker's bivariate distributions with fixed margins which are based on order statistics and found conditions under which the correlation converges to the maximum for Frchet-Hoeffding upper bound as the sample size tends to infinity.

- Another extension of FGM family introduced by Lallena (1992) and extensively studied by Amblard and Girard (2002, 2005), if $W_\theta(x,y) = \theta\phi(x)\phi(y)$. They proved that for properly chosen function $\phi(.)$, the range of the correlation coefficient is extended to $[-\frac{3}{4}, \frac{3}{4}]$.

4 Order Statistics Based Method

Baker (2008), proposed a new characterization of the bivariate FGM distribution based on order statistics as follows:

$$(V_1, V_2) = \begin{cases} (X_{(2)}, Y_{(2)}) & \text{with, p. } \frac{1}{2} \\ (X_{(1)}, Y_{(1)}) & \text{with, p. } \frac{1}{2} \end{cases}$$

It is easy to show that $(V_1, V_2) \sim FGM(1)$. Also if we define,

$$(U_1, U_2) = \begin{cases} (X_{(1)}, Y_{(2)}) & \text{with, p. } \frac{1}{2} \\ (X_{(2)}, Y_{(1)}) & \text{with, p. } \frac{1}{2} \end{cases}$$

then, we can check that $(U_1, U_2) \sim FGM(-1)$. Combining the above statements we obtain,

$$H(x,y) = FGM(\theta) = \frac{1+\theta}{2}FGM(1) + \frac{1-\theta}{2}FGM(-1).$$

where $-1 \leq \theta \leq 1$.
Several statistician have tried to extend the FGM distribution to allow large correlation, for instance: Huang and Kotz(1999), Bairamov et al. (2001), Amblard and Girard(2009), Fisher and Klein(2007), Baker (2008), Praha and Kotz (2010), Mirhoseini et al. (2011) and Klein and Cheista (2012). In this section, we derived some new versions of $W_\theta(x,y)$ for all $x, y \in [0,1]$ constructed using mixture of order statistics. Also, we studied distribution functions, copula. Spearman and Kendall correlations and some dependence ordering.

5 Record Statistics Based Method

Let X and Y be independent random variables with distribution functions F and G respectively. It is well-known that, the distribution function for lower record and upper record of a random sample of X and Y is as: Let X_1 and X_2, Y_1 and Y_2 be two sample of X and Y respectively. Define

$$F_{L_2}(x) = F(x) + F(x).H(x), \quad and \quad F_{U_2}(x) = F(x) - \bar{F}(x)R(x),$$

where $H(x) = -\ln(F(x))$ and $R(x) = -\ln(\bar{F}(x))$. The corresponding quantities for random variable Y denoted by $H(y)$ and $R(y)$.
In this section discussed some bivariate distributions constructed using mixture of recorded statistics. Also, we obtained distribution functions, copula, Spearman and Kendall correlations and some dependence ordering.

6 The Ruschendorf Method

Ruschendorf (1985) developed a method as follow: Let $f^1(x, y)$ has its integral to be zero on the unit square,as well as its two marginals integrate to zero, i.e.

$$\int_I f^1(x, y)dxdy = 0, \quad \int_0^1 f^1(x, y)dx = \int_0^1 f^1(x, y)dy = 0.$$

In that case $1 + f^1(x, y)$ is a density of a copula. There is however a constraint, namely that $1 + f^1(x, y)$ must be positive. If it is not the case, but f^1 is bounded, one can then find a constant θ such that $1 + \theta.f^1(x, y)$ is positive.It is easy to construct a function $f^1(x, y)$ as:
Step 1. Select an arbitrary real integrable function f on the unit square with its marginals being uniform, and computes:

$$A = \int_I f(x, y)dxdy, \quad f_1(x) = \int_0^1 f(x, y)dy, \quad f_2(y) = \int_0^1 f(x, y)dx.$$

Step 2- Put $f^1(x, y) = f(x, y) + A - f_1(x) - f_2(y)$.

Let $f(x, y) = \phi_1(x)\phi_2(y)$, such that $\phi_1(x)$ and $\phi_2(y)$ are bounded, we obtain

$$c(x, y) = 1 + \theta[\phi_1(x) - (\Phi_1(1) - \Phi_1(0))][\phi_2(x) - (\Phi_2(1) - \Phi_2(0))],$$

and

$$C(x, y) = x.y + \theta\psi(x, y)$$

where, $\Phi_i(x) = \int_0^x \phi(t)dt, \quad i = 1, 2.$ and

$$\psi(x, y) = [\Phi_1(x) - \Phi_1(0) - x(\Phi_1(1) - \Phi_1(0))][\Phi_2(x) - \Phi_2(0) - y(\Phi_2(1) - \Phi_2(0))]$$

1- If $\Phi_i(x), \quad i = 1, 2$ are decreasing, then the admissible region of θ is:

$$\theta \geq -[\phi_1(1) - (\Phi_1(1) - \Phi_1(0))]^{-1}[\phi_2(1) - (\Phi_2(1) - \Phi_2(0))]^{-1}.$$

2- If least one of $\Phi_1(x)$ or $\Phi_2(x)$ are increasing, then the admissible region of θ is:
$\theta \geq -\max\{\theta_1, \theta_2\}$. where

$$\theta_1 = [\phi_1(0) - (\Phi_1(1) - \Phi_1(0))]^{-1}[\phi_2(1) - (\Phi_2(1) - \Phi_2(0))]^{-1},$$

and

$$\theta_2 = [\phi_1(1) - (\Phi_1(1) - \Phi_1(0))]^{-1}[\phi_2(0) - (\Phi_2(1) - \Phi_2(0))]^{-1}.$$

Noting that this family is a member of the family of Lallena (1992), with

$$\phi(x) = [\Phi_1(x) - \Phi_1(0) - x(\Phi_1(1) - \Phi_1(0))]$$

and

$$\psi(x) = [\Phi_2(x) - \Phi_2(0) - y(\Phi_2(1) - \Phi_2(0))],$$

It is obvious that $\phi(1) = \phi(0) = \psi(1) = \psi(0) = 0$. So, using Proposition 6 we obtain $\frac{\partial^2 \ln(A(x,y))}{\partial x \partial y} = \frac{\partial^2 \ln(\theta\phi(x).\psi(y))}{\partial x \partial y} = 0$, consequently $2\rho_s = 3\tau$.

1- If $f(x, y) = x^i y^j, \quad i, j \geq 1, \quad x, y \in [0, 1]$. Then, we derive

$$c(x, y) = 1 + \theta(x^i - \frac{1}{i+1})(y^j - \frac{1}{j+1}),$$

and

$$C(x, y) = xy + \theta \frac{xy}{(i+1)(j+1)}.(x^i - 1)(y^j - 1),$$

where $0 \leq \theta \leq \min\{\frac{(i+1)(j+1)}{i}, \frac{(i+1)(j+1)}{j}\} \leq 1$. This is known as family of bivariate distributions Hung and Kotz (1999) if $i = j = q$.

Ruschendorf (1985), using the function $f(x, y) = x^i y^j, \quad i, j \geq 1, \quad x, y \in [0, 1]$ generated polynomial copula of power m as follows:

$$C(x, y) = xy[1 + \sum_{i \geq 1, j \geq 1}^{i+j \leq m-2} \frac{\theta_{ij}}{(i+1)(j+1)}(x^i - 1)(y^j - 1)],$$

where

$$0 \leq \theta \leq \min\{\sum_{i \geq 1, j \geq 1} \theta_{ij} \frac{i}{(i+1)(j+1)}, \sum_{i \geq 1, j \geq 1} \theta_{ij} \frac{j}{(i+1)(j+1)}\}.$$

In particular, if $i = j = 1$ we derive the family of bivariate distributions $FGM^+(\theta)$.

2- If $f(x, y) = e^{-ix}e^{-jy}$, Then, we obtain

$$c(x, y) = 1 + \theta[e^{-ix} - \frac{1 - e^{-i}}{i}][e^{-jy} - \frac{1 - e^{-j}}{j}],$$

and

$$C(x, y) = xy + \frac{\theta}{ij}[1 - e^{-ix} - x(1 - e^{-i})][1 - e^{-jy} - y(1 - e^{-j})],$$

where

$$-1 \leq \max\{-A_{ij}, -B_{ij}\} \leq \theta \leq B_{ij} \leq 1,$$

$$A_{ij} = \frac{ij}{[(i+1)e^{-i} - 1][(j+1)e^{-j} - 1]}$$

and

$$B_{ij} = \frac{ij}{[i - 2 + (i+2)e^{-i}][j - 2 + (j+2)e^{-j}]}.$$

3-If $f(x, y) = \phi(x) + \psi(y)$, where $\phi(x)$ and $\psi(y)$ are bounded and satisfy in suitable conditions, then we obtain the independence copula i.e.$C(x, y) = \Pi(x, y) = x.y$.

References

[1] Amblard, C. and Girard, S.(2009) A new extension of bivariate FGM copulas. Metrika:70:1-17.

[2] N. Balakrishnan and C. Lai, Continuous bivariate distributions, Second Edition, Springer New York, 2009.

[3] Bairamov ,I.,Kotz S.and Bekc M. (2001) New generalized Farlie-Gumbel-Morgenstern distributions and concomitants of order statistics. J.Appl.Stat. 28(5);521-536.

[4] Bairamov, I. and Kotz, S. (2002) Dependence structure and symmetry of huang-Kotz FGM distributions and their extension. Metrika, 56:55-72.

[5] Baker, R.(2008) An order-statistics-based method for constructing multivariate distributions with fixed marginals. J. Multiv.Analysis, 99:2312-2327.

[6] R.E. Barlow and F. Proschan, Statistical Theory of Reliability and Life Testing. New York: Holt, Rinehart and Winston, 1975.

[7] P.J Boland and F.J Samaniego, The signature of a coherent system and its applications in reliability. In Mathematical reliability: An expository eerspective, Soyer R, Mazzuchi T, Singpurwalla ND (eds), Kluwer Publishers: Boston, 2004.

[8] Coles, S., Heffernan, J. and Tawn, J. (1999). Dependence Measures for Extreme value Analyses. *Exterems 2(4),339-365.*

[9] Cuadras, C.M.(2009) Constructing copula functionc with weighted geometric means. J.Statist.Plann. Infe. 139, 3766-3772.

[10] Fisher, M. and Klein, I.(2007) Constructing generalized FGM copulas by means of certain univariate distributions. Metrica, 65; 243-260.

[11] fredricks, G.A. and Nelsen, R.B. (2007) On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables. J.Stat.Plann. Inference. 137: 2143-2150.

[12] H. David and H.N. Nagaraja, Order Statistics, Wiley, Hoboken, NJ, 3rd ed, 2003.

[13] Dolati, A. and Ubeda-Flores,M.(2009) Construdting copulas by means of pairs of order statistics. Kybernetika-Manuscript, 45(6), 992-1002.

[14] Huang,J.S. and Kotz, S.(1999) Modifications of the Farlie-Gumbel-Morgenstern distributions. A tough hill to climb. Metrika, 49: 135-145.

[15] Praha,H., D. and Kotz, S. (2010) The generalized FGM distribution and application to stereology of extremes. Applications of Mathematics. 6; 495-512.

[16] M. C. Jones, Families of distributions arising from distributions of order statistics (with discussion), Test 13 (2004) 143.

[17] Klein, I. and Christa, F. (2012) Families of copulas closed under the construction of generalized linears means. IWQW.

[18] D. Drouet-Mari and S. Kotz, Correlation and dependence, Imperial College Press, London, 2001.

[19] S.N.U.A Kirmani and R. Gupta, On the proportional odds model in survival analysis, Ann. Inst. Statist. Math. 53 (2) (2001) 203–216.

[20] S. Kotz, N. Balakrishnan and N.L Johnson, Continuous multivariate distributions, Vol. 1, second ed., Wiley: New York, 2000.

[21] T.J. Kozubowskiand S. Nadarajah, Multitude of Laplace distributions, Statistical Papers, 51, 2010, 127-148.

[22] Lai, C.D. and Xie, M.(2000) A new family of positive quadrant dependent bivariate distributions. Stat. Probab. Lett. 46; 359-364.

[23] Lin, G.D. and Huang, J.S. (2010) A note on the maximum correlation for Baker's bivariate distributions with fixed marginals. J. Multiv. Anal. 101, 2227-2233.

[24] A. W. Marshall and I. Olkin, Adding a parameter to a family of distributions with application to the exponential and Weibull families, Biometrika 84 (1997) 641–652.

[25] Mirhoseini, S.M., Dolati, A. and Amini, M. (2011) On A Class of Distributions Generated by Stochastic Mixture of the Extreme Order Statistics of A Sample of Size Two Journal of Statistical Theory and Applications Volume 10, Number 3, 2011

[26] R.B Nelsen, An introduction to copulas, Second Edition, Springer, New York, 2006.

[27] M. Shaked and G. Shanthikumar, Stochastic orders, Springer, New York, 2007.

A new class of positive dependent bivariate copula and its properties

Bekrizadeh H. ,Parham G. A. ,Zadkarmi M. R.
Department of Statistics Shahid Chamran University, Ahvaz, Iran,

Abstract

We propose to give up the polynomial form to work with a semi parametric family of copulas. The induced parametric families of copulas are generated as simply as Archimedean copulas, that is by a univariate function.This new class covers Farlie-Gumbel-Morgenstern and Gumbell-Barnnet copulas. Also, seeing that, the domain of correlation Farlie-Gumbel-Morgenstern copula has been limited, it attempts to study values of dependence and asserts that these extending families of copula are used in accordance modeling with high postive dependence. Furthermore, dependence properties of copulas with polynomial sections are preserved and the dependence degree can be increased without significantly complexifying the model

Keywords: Copulas, semiparametric family, Farlie–Gumbel–Morgenstern copula, measures of association, positive dependence

1 Introduction

Copulas are functions that join or "couple" joint distribution functions to their one-dimensional marginal distribution functions. Alternatively, copulas are joint distribution functions whose one-dimensional margins are uniform on the interval $(0, 1]$.

Copulas are of interest to statisticians for two main reasons: Firstly, as a way of studying scale-free measures of dependence; and secondly, as a starting point for constructing families of bivariate distributions, sometimes with a view to simulation. The recent book by R.B. Nelsen is an important monograph about copulas. As for the relationship with problems of given marginals, it can be seen [Benes, and Stepan (1997), Cuadras, et. al. (2002), Joe (1997)].

Several families of copulas, such as Archimedean copulas [Coles,Currie,and Tawn (1999)] or copulas with polynomial sections [Nelsen et. al.(1997) and Quesada-Molina et. al. (1995)] have been proposed. In [Nelsen et. al. (1997)], the authors point out that the copulas with quadratic section proposed in [Quesada-Molina et. al. (1995)] are not able to modelize large dependences. Then, they introduce copulas with cubic sections and conclude that copulas with higher order polynomial section would increase the dependence degrees but simultaneously the complexity of the model. We propose to give up the polynomial form to work with a semi parametric family of copulas. The induced parametric families of copulas are generated as simply as Archimedean copulas, that is by an univariate function. Furthermore, dependence properties of copulas with polynomial sections are preserved and the dependence degree can be increased without significantly complexifying the model. Note that, in Ferguson (1995), a class of symmetric bivariate copulas with wide correlation coefficients range is introduced, but these copulas are less convenient to perform classical calculations on a probability law than the copulas studied here.

One of the most popular parametric families of copulas, that studied in Farlie (1960), Gumbel (1960), and Morgenstern (1956), is the Farlie-Gumbel-Morgenstern (FGM) family that defined by

$$C^{FGM}(u, v) = uv(1 + \theta(1 - u)(1 - v)), \theta \in [-1, 1]$$

and the FGM copula density is provided by

$$c^{FGM}(u, v) = (1 + \theta(2u - 1)(2v - 1)), \theta \in [-1, 1]$$

These are the only copulas whose functional form is a polynomial quadratic in u and in v. They are commonly denoted FGM copulas. Members of the FGM family are symmetric, i.e., $C_\theta(u, v) = C_\theta(v, u)$

for all (u, v) in I^2. An alternative approach to generalize the FGM family of copulas is to consider the semi-parametric family of symmetric copulas defined by

$$C^{RL}(u,v) = uv + \theta\Theta(u)\Theta(v), \theta \in [-1,1]$$

It was first introduced in Rodriguez-Lallena (1992), and extensively studied in Amblard and Girard (2002, 2005). Kim and Sungur (2004) utilized the FGM family to models involved censoring. Durante and Jaworski (2009) analyzed the structure of the FGM distribution with gamma marginals and discussed the various parameterizations of the FGM family. Kim et al. (2011) proposed a new class of bivariate copula to quantify dependence and incorporate into various iterated copula families. A pair (X, Y) of random variables is said to be exchangeable if the vectors (X, Y) and (Y, X) are identically distributed. For identically distributed continuous random variables, exchangeability is equivalent to the symmetry of the copula.

In this paper, we propose to give up the polynomial form to work with a semi parametric family of copulas. In Section 2, the semi parametric family is defined and its basic properties are derived. Properties of measures of association are investigated in Section 3. Section 4 is devoted to the dependence structure and the dependence ordering of the family. Finally the results of this research carried out in section 5.

2 new class of bivariate copula

Throughout this paper, we note $I = [0, 1]$. A bivariate copula defined on the unit square I^2 is a bivariate cumulative distribution function with univariate uniform margins. Equivalently, C must satisfy the following properties:

(P1) $C(u, 0) = C(0, v) = 0$, for every u, v in I,

(P2) $C(u, 1) = u$ and $C(1, v) = v$, for every u, v in I,

(P3) For every u_1, v_1, u_2, v_2 in I such that $u_1 \leq u_2$ and $v_1 \leq v_2$,

$$\Delta(u_1, v_1, u_2, v_2) = C(u_2, v_2) - C(u_2, v_1) - C(u_1 v_2) + C(u_1, v_1) \geq 0$$

We consider the semi parametric family of functions defined on I^2 by

$$C_\theta^{\Phi,\Psi}(u,v) = uve^{\theta\Phi(u)\Psi(v)} \tag{1}$$

for some parameter θ, where Φ and Ψ are a differentiable function on I. Let us note first that, the independent copula $C_0^{\Phi,\Psi}(u, v) = uv$ belongs to any parametric family $\{C_\theta^{\Phi,\Psi}\}$ generated by a functions Φ and Ψ. Second, the functions Φ and $-\Phi$ or Ψ and $-\Psi$ clearly define the same functions. The function Φ and Ψ plays a role similar to the generating function in Archimedean copulas [Genest and MacKay (1986)]. The next theorem gives sufficient and necessary conditions on Φ , Ψ, and θ to ensure that $C_\theta^{\Phi,\Psi}(u, v)$ is a copula.

Theorem1. $C_\theta^{\Phi,\Psi}(u, v)$ is a copula if and only if Φ and Ψ satisfies the following conditions:

A1. $\Phi(1) = 0$ or $\Psi(1) = 0$,

A2. $[1 + \theta u\Phi'(u)\Psi(v)][1 + \theta v\Phi(u)\Psi'(v)] + \theta uv\Phi'(u)\Psi'(v) \geq 0$,

A3. $\theta \geq -\min\left\{[\Phi'(1)\Psi(0)]^{-1}, [\Phi'(1)\Psi'(1)]^{-1}\right\}$.

Proof: The proof involves three steps.

1. It is clear that (P2)$\Longleftrightarrow$ (A1).

2. We show that (P3)$\Longleftrightarrow$ (B1). In this case, for all u,u_1,v, and v_1 in I such that $u_1 > u$ and $v_1 > v$, and

$\Delta(u, u_1, v, v_1)$ can be written as

$$\Delta(u, u_1, v, v_1) = v_1 \left[u_1 e^{\theta\Phi(u_1)\Psi(v_1)} - ue^{\theta\Phi(u)\Psi(v_1)}\right] - v \left[u_1 e^{\theta\Phi(u_1)\Psi(v)} - ue^{\theta\Phi(u)\Psi(v)}\right]$$

and thus $\Delta(u, u_1, v, v_1) \geq 0$ implies

$$\frac{v\left[u_1 e^{\theta\Phi(u_1)\Psi(v)} - u e^{\theta\Phi(u)\Psi(v)}\right]}{v_1\left[u_1 e^{\theta\Phi(u_1)\Psi(v_1)} - u e^{\theta\Phi(u)\Psi(v_1)}\right]} \leq 1 \tag{2}$$

multiplying and divide (2) by $u_1 - u$ and letting $u_1 \longrightarrow u$, we have

$$v_1\left[\frac{\partial}{\partial u} u e^{\theta\Phi(u)\Psi(v)}\right] - v\left[\frac{\partial}{\partial u} u e^{\theta\Phi(u)\Psi(v)}\right] \geq 0$$

Divided both side by $v_1 - v$ and letting $v_1 \longrightarrow v$ in the previous inequality yields (A2) as follows:

$$\frac{\partial^2}{\partial u \partial v} uv e^{\theta\Phi(u)\Psi(v)} = e^{\theta\Phi(u)\Psi(v)}\left\{[1 + \theta u \Phi'(u)\Psi(v)][1 + \theta v \Phi(u)\Psi'(v)] + \theta uv \Phi'(u)\Psi'(v)\right\} \geq 0 \tag{3}$$

3. To determine the admissible range of θ it is sufficient to determine the range of values of θ for which the density is nonnegative. For the density (3), the overall constraint on θ is given by

$$c_\theta^{\Phi,\Psi}(u,v) = \frac{\partial^2}{\partial u \partial v} C_\theta^{\Phi,\Psi}(u,v) = \frac{\partial^2}{\partial u \partial v} uv e^{\theta\Phi(u)\Psi(v)} \geq 0$$

Thus the range of θ arises from:

I: When $u = v = 1$, since $\Phi(1) = 0$ and $\Psi(1) = 0$, then $1 + \theta\Phi'(1)\Psi'(1) \geq 0$. Thus $\theta \geq -\left[\Phi'(1)\Psi'(1)\right]^{-1}$.
II: When $u = 1$ and $v = 0$ or $u = 0$ and $v = 1$, then $1 + \theta\Phi'(1)\Psi(0) \geq 0$. Thus $\theta \geq -[\Phi'(1)\Psi(0)]^{-1}$.
Therefore, the admissible range of θ is

$$\theta \geq -\min\left\{[\Phi'(1)\Psi(0)]^{-1}, [\Phi'(1)\Psi'(1)]^{-1}\right\}$$

Corollary 2.1 Under assumption of the Theorem 1, if $\Phi(x) = \Psi(x) = \Omega(x)$ then generate parametric families of copulas as

$$C_\theta^{\Omega}(u,v) = uv e^{\theta\Omega(u)\Omega(v)}$$

where $C_\theta^{\Omega}(u,v)$ if and only if Ω satisfies the following conditions:
B1. $\Omega(1) = 0$,
B2. $[1 + \theta u \Omega'(u)\Omega(v)][1 + \theta v \Omega(u)\Omega'(v)] + \theta uv \Omega'(u)\Omega'(v) \geq 0$,
B3. $\theta \geq -\min\left\{[\Omega'(1)\Omega(0)]^{-1}, [\Omega'(1)]^{-2}\right\}$.
Corollary 2.2 Under assumption of the Theorem 1, The novel family (1) has the following extension to a $(2^d - d - 1)$-parameter family of d-copulas, $d \geq 3$:

$$C_\theta^{\Phi,\Psi}(u_1, u_2, ..., u_d) = \left[{}_{i=1}^{d} u_i\right] e^{\sum_{k=2}^{d} \sum_{1 \leq j_1 < j_2 < \ldots < j_k \leq d} \theta_{j_1 j_2 \ldots j_k} \Phi(uj_1)\Psi(vj_1)\Phi(uj_2)\Psi(vj_2)\ldots\Phi(uj_k)\Psi(vj_k)}$$

Corollary 2.3 The log-likelihood function for the parameter given set of observed pairs, is given by

$$l^n(\theta | u, v) = \sum_i \log c_\theta^{\Phi,\Psi}(u_i, v_i)$$

where $c_\theta^{\Phi,\Psi}(u_i, v_i)$ is the density function of the distribution function $C_\theta^{\Phi,\Psi}(u_i, v_i)$. The maximum likelihood estimators are obtained by maximizing. Then, we have,

$$\frac{\partial}{\partial \theta} l^n(\theta | u, v) = 0$$

equivalently

$$A\hat{\theta}^2 + B\hat{\theta} + C = 0 \tag{4}$$

where

$$\begin{aligned} A &= \sum_i u_i v_i \Phi^2(u_i)\Psi^2(v_i)\Phi'(u_i)\Psi'(v_i) \\ B &= \sum_i u_i v_i \Phi(u_i)\Psi(v_i)\left[u_i\Phi'(u_i)\Psi(v_i) + v_i\Phi(u_i)\Psi'(v_i) + 3u_i v_i \Phi'(u_i)\Psi'(v_i)\right] \\ C &= \sum_i \left[\Phi(u_i)\Psi(v_i) + u_i\Phi'(u_i)\Psi(v_i) + v_i\Phi(u_i)\Psi'(v_i) + 3u_i v_i \Phi'(u_i)\Psi'(v_i)\right] \end{aligned}$$

The MLE of θ, then could be canonically represented as

$$\hat{\theta} = \arg\max_{\theta} l^n \left(\theta | \, u, v\right) \tag{5}$$

The Copula can be used to make generalized bivariate distributions when marginal distributions are $F(x)$ and $G(y)$. Some examples are mentioned in the following sections.
Example 2.1 The bivariate Taylor expansion of exponential part (1) at $(u, v) = (1, 1)$ with the function $\Phi(x) = \Psi(x) = 1 - x$ generate parametric families copulas

$$C_\theta(u, v) = uv \left\{ 1 + \sum_{r=1}^{\infty} \frac{[\theta(1-u)(1-v)]^r}{r!} \right\}, \, \theta \in [-1, 1]$$

This is higher order polynomial section.
Example 2.2 The bivariate Taylor expansion exponential part of (1) with order 3 or 4 at $(u, v) = (1, 1)$ and $\Phi(x) = \Psi(x) = 1 - x$, generate parametric families of FGM copulas.
Example 2.3 Relation (1) with the function $\Phi(x) = \ln(x)$ and $\Psi(x) = -\ln(x)$ generate parametric families of Gumbell-Barnnet copulas.
Some properties are mentioned in the following sections.
Theorem2. Suppose $C_\theta^{\Phi,\Psi}(u, v)$ that defined in (1) under conditions in theorem 1:
I) The conditional distribution function for V given $U = u$ is

$$C_u^{\Phi,\Psi}(v) = P\left(V \le v | \, V \le v\right) = v e^{\theta\Phi(u)\Psi(v)} \left[1 + u\Phi'(u)\Psi(v)\right] \tag{6}$$

II) Let (U, V) be a pair of random variable with copula (1) and uniform marginals. Then,

$$P(U < V) = \frac{1}{2} \tag{7}$$

Proof: I) The conditional distribution function $P\left(V \le v | \, V \le v\right)$ in (6) follows from the derivative of $C_\theta^{\Phi,\Psi}(u, v)$ with respect to u , i.e., $\frac{\partial}{\partial u} C_\theta^{\Phi,\Psi}(u, v)$.
II) From (3) we have

$$\begin{aligned} P(U &< V) = \int_0^1 \int_0^v \frac{\partial^2}{\partial u \partial v} C_\theta^{\Phi,\Psi}(u, v) du dv \\ &= \int_0^1 \int_0^v e^{\theta\Phi(u)\Psi(v)} \left\{[1 + \theta u\Phi'(u)\Psi(v)][1 + \theta v\Phi(u)\Psi'(v)] + \theta uv\Phi'(u)\Psi'(v)\right\} du dv \\ &= \int_0^1 e^{\theta\Phi(v)\Psi(v)} v \left[1 + \theta v\Phi(v)\Psi'(v)\right] dv \\ &= \int_0^1 v e^{\theta\Phi(v)\Psi(v)} dv + \left\{ \frac{v^2}{2} e^{\theta\Phi(v)\Psi(v)} \right|_0^1 - \int_0^1 v e^{\theta\Phi(v)\Psi(v)} dv \\ &= \frac{1}{2} \end{aligned}$$

So, $P(U < V) = \frac{1}{2}$. We know, if X and$Y$ be independent and identically distributed continuous random variables, then $P(X < Y) = \frac{1}{2}$.

3 Measures of dependence

In this section we will look at different ways in which copulas can be used in the study of dependence between random variables. For a historical review of measures of association and concepts of independence, see [Joe (1997), Caari'ere (2004), and Nelsen (2006)].

3.1 Spearman's rho

Let X and Y be continuous random variables whose copula is C. then the population version of Spearman's rho for X and Y is given by

$$\rho = 12\int_0^1\int_0^1 C(u,v)dudv - 3 \tag{8}$$

we now present the results on the properties of a novel family of copula. Note that if $\theta = 0$, then U and$V$ are independent and $\rho = 0$.

Since direct computation of the Spearman correlation for new copula is impossible, we compute the Spearman correlation for special case of functions Φ and Ψ.

Example 3.1 From (8) and (1) with the function $\Phi(x) = \Psi(x) = 1 - x$, Spearman's Rho is

$$\begin{aligned} \rho &= 12\int_0^1\int_0^1 uve^{\theta(1-u)(1-v)}dudv - 3 \\ &= 12\int_0^1 \frac{ve^{\theta(1-v)} - 1 - \theta(1-v)}{\theta^2(1-v)^2}dv - 3 \\ &= 12\frac{(1-\theta)\{E_i(1,-\theta) + \gamma + \ln(-\theta)\} - e^{\theta} + 2\theta + 1}{\theta^2} - 3\ . \end{aligned}$$

where $\gamma = -0.5772156649$ and $E_i(t,z)$ is the exponential integrals by $E_i(t,z) = z^{t-1}\Gamma(1-t,z)$. It may be noted that $\rho \in [-0.2916, 0.3806]$.

On the other hand, by using the example 2.1, we can also compute Spearman correlation function. The amplitude correlation is

$$\begin{aligned} \rho &= 12\int_0^1 uv\left\{1 + \sum_{r=1}^{\infty}\frac{[\theta(1-u)(1-v)]^r}{r!}\right\}dudv - 3 \\ &= 12\left\{\sum_{r=1}^{\infty}\frac{\theta^r}{(r+2)!}\frac{1}{(r+1)(r+2)}\right\} \end{aligned}$$

now we approximate the above relation with finite term m, such as

$$\rho \approx 12\left\{\sum_{r=1}^{\infty}\frac{\theta^r}{(r+2)!}\frac{1}{(r+1)(r+2)}\right\}$$

the numerical results of this approximation for the lover and upper bound of ρ with $\theta = 1$ and $\theta = -1$ respectively are shown in Table 3.1.

Table 3.1: Approximate of the lover and upper bound of ρ.

m	Lover bound of ρ	Upper bound of ρ
1	-0.3333333333	0.3333333333
5	-0.2961678005	0.3806122449
10	-0.2961629109	0.3806180585
25	-0.2961629109	0.3806180585
50	-0.2961629109	0.3806180585
100	-0.2961629109	0.3806180585
200	-0.2961629109	0.3806180585

Note that in Table 3.1 when $m = 1$, we have $\rho = \frac{\theta}{3} \in [-0.33, 0.33]$ which is bound Spearman's Rho for FGM copula. Example 3.1 shows that the upper bound Spearman's Rho in new family has improved with respect to spearman correlation amplitude in FGM family, because polynomial section has higher order with respect to FGM copula.

3.2 Tail Dependence

The concept of tail dependence relates to the amount of dependence in the upper-right quadrant tail or lower-left-quadrant tail of a bivariate distribution (Farlie, 1960). It is a concept that is relevant for the study of dependence between extreme values. It turns out that tail dependence between two continuous random variables and is a copula property and hence the amount of tail dependence is invariant under strictly increasing transformations of X and Y.

Definition 3.2.1 If a bivariate copula C is such that,

$$L_U = \lim_{u \longrightarrow 1} \frac{1 - 2u + C(u,u)}{1 - u} \tag{9}$$

exists, then C has upper tail dependence if $L_U \in (0, 1]$, and upper tail independence if $L_U = 0$. The measure is extensively used in extreme value theory. It is the probability that one variable is extreme given that the other is extreme, i.e., $L_U = P\{U > u|\, V > v\}$. Thus L_U could be viewed as a quantile dependent measure of dependence (Coles, Currie and Tawn, 1999).

Definition 3.2.2 The concept of lower tail dependence can be defined in a similar way. If the limit,

$$L_L = \lim_{u \longrightarrow 0} \frac{C(u,u)}{u} \tag{10}$$

exists, then C has lower tail dependence if $L_L \in (0, 1]$, and lower tail independence if $L_L = 0$. Similarly lower tail dependence is defined as $L_U = P\{U < u|\, V < v\}$. For copulas without a simple closed form an alternative formula for L_L is more useful. So, the upper tail dependence of the copula (1) by using (9) is,

$$L_U = \lim_{u \longrightarrow 1} \frac{1 - 2u + u^2 e^{\theta\Phi(u)\Psi(u)}}{1 - u} = 0$$

since $\Phi(1) = 0$ and $\Psi(1) = 0$.

The lower tail dependence of the copula (1) by using (10) is,

$$L_L = \lim_{u \longrightarrow 0} \frac{u^2 e^{\theta\Phi(u)\Psi(u)}}{u} = 0$$

4 Concepts of dependence

In this section, for the sake of simplicity, we assume that X and Y are exchangeable. Several concepts of positive dependence have been introduced and characterized in terms of copulas. The random variables X and Y are

a) Positively Quadrant Dependent (PQD) if for all

$$P(X \leq x, Y \leq y) \geq P(X \leq x)P(Y \leq y)\ ,\ (x, y)\ \in \mathbb{R}^2$$

or equivalently

$$C(u, v) \geq uv\ ,\ \forall (u, v) \in I^2$$

b) Left Tail Decreasing (LTD) if $P\{Y \leq y|\, X \leq x\}$ is non-increasing in x for all y, or equivalently, see Theorem 5.2.5 in Nelsen (2006), $u \longrightarrow \frac{C(u,v)}{u}$ is non-increasing for all $v \in I$.

c) Right Tail Increasing (RTI) if $P\{Y > y|\, X > x\}$ is non-decreasing in x for all y, or equivalently, $u \longrightarrow \frac{v - C(u,v)}{1-u}$ is non-increasing for all $v \in I$.

d) Left Corner Set Decreasing (LCSD) if $P\{X \leq x, Y \leq y | X \leq x_1, Y \leq y_1\}$is non increasing in x_1 and y_1 for all x and y , or equivalently, see Corollary 5.2.17 in Nelsen (2006), C is a totally positive function of order 2, i.e. for all $(u_1, u_2, v_1, v_2) \in I^4$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$, one has

$$\Lambda = C(u_1, v_1)C(u_2, v_2) - C(u_1, v_2)C(u_2, v_1) \geq 0$$

e) Right Corner Set Increasing (RCSI) if $P\{X > x, Y > y | X > x_1, Y > y_1\}$is non-decreasing in x_1 and y_1 for all x and y , or equivalently, the survival copula $\hat{C}$ associated to C is a totally positive function of order 2. Negative dependence can be similarly defined.

In the following theorem establishing the concepts of positive dependencies using the above definitions for the new copula family.

Theorem 3. Let (X, Y) a random pair with copula $C_\theta^{\Phi,\Psi}(u, v)$ in (1). X and Y are
(i) PQD if and only if $\theta\Phi(u)\Psi(v) \geq 0$.
(ii) LTD if and only if $\theta\Phi'(u)\Psi(v) \leq 0$.
(iii) RTI if and only if $e^{\theta\Phi(u)\Psi(v)}\{1 + \theta u(1-u)\Phi'(u)\Psi(v)\} + 1 \geq 0$.
(iv) LCSD if and only if they are LTD.
(v) RCSI if and only if they are RTI.
(i) By using definition (a) can be rewritten as

$$e^{\theta\Phi(v)\Psi(v)} \geq 1, \ \forall (u, v) \in I^2$$

so, $\theta\Phi(u)\Psi(v) \geq 0$.
(ii) By using definition (b), $\frac{C_\theta^{\Phi,\Psi}(u,v)}{u} = ve^{\theta\Phi(u)\Psi(v)}$ is non-increasing with respect to u if and only if,

$$\frac{\partial}{\partial u} ve^{\theta\Phi(u)\Psi(v)} = \theta v\Phi'(u)\Psi(v)e^{\theta\Phi(u)\Psi(v)} \leq 0$$

Therefore, $\theta\Phi'(u)\Psi(v) \leq 0$.
(iii) by using definition (c), $\frac{v - C_\theta^{\Phi,\Psi}(u,v)}{1-u} = \frac{v(1-ue^{\theta\Phi(u)\Psi(v)})}{1-u}$ is non-increasing with respect to u if and only if,

$$\frac{\partial}{\partial u}\left\{\frac{v(1 - ue^{\theta\Phi(u)\Psi(v)})}{1-u}\right\} = \frac{ve^{\theta\Phi(u)\Psi(v)}\left\{\theta u^2\Phi'(u)\Psi(v) - \theta u\Phi'(u)\Psi(v) - 1\right\} + v}{(1-u)^2} \leq 0$$

The above condition, reduce to $e^{\theta\Phi(u)\Psi(v)}\left\{\theta u^2\Phi'(u)\Psi(v) - \theta u\Phi'(u)\Psi(v) - 1\right\} + 1 \leq 0$ or $e^{\theta\Phi(u)\Psi(v)}\{1 + \theta u(1-u)\Phi'(u)\Psi(v)\}$
$1 \geq 0$.
(iv) Suppose $u_1 \leq u_2$ and $v_1 \leq v_2$, and by definition (d), we have

$$\begin{aligned}\Lambda &= u_1v_1e^{\theta\Phi(u_1)\Psi(v_1)}u_2v_2e^{\theta\Phi(u_2)\Psi(v_2)} - u_1v_2e^{\theta\Phi(u_1)\Psi(v_2)}u_2v_1e^{\theta\Phi(u_2)\Psi(v_1)} \\ &= u_1v_1u_2v_2\left\{e^{\theta\Phi(u_1)\Psi(v_1)}e^{\theta\Phi(u_2)\Psi(v_2)} - e^{\theta\Phi(u_1)\Psi(v_2)}e^{\theta\Phi(u_2)\Psi(v_1)}\right\} \geq 0\end{aligned}$$

so, $e^{\theta\Phi(u_1)\Psi(v_1)}e^{\theta\Phi(u_2)\Psi(v_2)} \geq e^{\theta\Phi(u_1)\Psi(v_2)}e^{\theta\Phi(u_2)\Psi(v_1)}$ and $\Phi(u_1)\Psi(v_1) + \Phi(u_2)\Psi(v_2) \geq \Phi(u_1)\Psi(v_2) + \Phi(u_2)\Psi(v_1)$, then

$$\Phi(u_1)[\Psi(v_1) - \Psi(v_2)] \geq \Phi(u_2)[\Psi(v_1) - \Psi(v_2)]$$

thus, $\Phi(u_1) \geq \Phi(u_2)$. Since $u_1 \leq u_2$ and $\Phi(u_1) \geq \Phi(u_2)$, thus $\Phi(u)$ is non-decreasing.
(v) The proof of similar (iv)

The three remaining situations are equivalent to the three previous ones since the considered copulas are symmetric in the arguments.

5 Conclusion

Copulas are useful devices to explain the dependence structure among variables by eliminating the influence of marginal. We propose a new family of copulas to quantify and incorporate it into various iterated copula families. We investigate properties of the new class of bivariate copula and derive the measure of association, such as spearman and tail dependence for the new class. The main feature of this family is to permit the modeling of high positive dependence. The new copula permit to us that calculate Spearman correlation for high order polynomial section. Necessary and sufficient conditions are given on the generating functions in order to obtain various dependence properties. Some examples of parametric subfamilies are provided. We also provide the concept directional dependence in bivariate regression setting by using copula. In addition, it is possible to build a new class of multivariate copulas.

References

[1] Amblard, C. and Girard, S., Symmetry and dependence properties within a semi parametric family of bivariate copulas. Nonparametric Stat. 2002,14(6):715–727.

[2] Amblard, C. and Girard, S., Estimation procedures for a semiparametric family of bivariate copulas. J. Comput. Graph Stat. 2005, 14(2):1–15.

[3] Benes, V. and Stepan, J., Distributions with Given Marginals and Moment Problems. Kluwer Academic Publishers, Dordrecht, 1997.

[4] Caari'ere, J. F., Copulas. Encyclopedia of Actuarial Science, Wiley, 2004,Vol. 1-3.

[5] Coles, S., Currie, J. and Tawn, J., Dependence measures for extreme value analyses.Working paper, Department of Mathematics and Statistics, Lancaster University,1999.

[6] Cuadras, C., Fortiana, J. and Rodr´ıguez Lallena, J. A., Distributions with Given Marginals and Statistical Modeling. Kluwer Academic Publishers, Dordrecht,2002.

[7] Dall'Aglio, G., Fr´echet classes: the beginnings in Advances in Probability Distributions with Given Marginals. Kluwer Academic Publishers, Dordrecht, 1991, pp. 1-12.

[8] Durante, F. and Jaworski, P., A new characterization of bivariate copulas. AMS Subject Classification,2009, 60E05, 62H20, 26A24.

[9] Farlie, D. G. J. , The performance of some correlation coefficients for a general bivariate distribution. Biometrika,1960, 47:307–323.

[10] Ferguson, T. S., A class of symmetric bivariate uniform copulas. Statistical paper, 1995, 36:31–40.

[11] Genest, C. and MacKay, J. , Copules archim´ediennes et familles de lois bidimensionnelles dont les marges sont donn´ees. Canad. J. Statist.,1986, 14:145–159.

[12] Gumbel, E. J., Bivariate exponential distributions. J. Am. Stat. Assoc., 1960,55:698–707.

[13] Joe, H., Multivariate Models and Dependence Concepts. Chapman Hall London,1997.

[14] Kim, J. et. al., Generalized bivariate copulas and their properties. IOS Press, Model Assisted Statistics and Application, 2011, 6:127-136.

[15] Kim, J. M., Sungur, E. A., New class of bivariate copulas. In: Proceedings for the Spring Conference 2004, Korean Statistical Society, 2004, pp. 207–212.

[16] Morgenstern, D., Einfache beispiele zweidimensionaler verteilungen. Mitteilungsblatt fürMathematische. Statistik,1956, 8:234–235.

[17] Nelsen, R. B., An Introduction to Copulas. Springer, New York,2006.

[18] Nelsen, R. B., Quesada-Molina, J. J., and Rodr´ıguez-Lallena, J. A., Bivariate copulas with cubic sections. Nonparametric Statistics,1997, 7:205–220.

[19] Quesada-Molina, J. J. and Rodr´ıguez-Lallena, J. A., Bivariate copulas with quadratic sections. Nonparametric Statistics, 1995, 5:323–337.

[20] Rodriguez-Lallena, J. A., Estudio de la compabilidad y dise64co de nuevas familias en la teoria de c64epulas.Aplicaciones. Tesis doctoral, Universidad de Granada, 1992.

Some Families of Multivariate Copulas

Dolati, A.

Yazd University, Department of Statistics,Yazd, Iran.

e-mail: adolati@yazdunu.ac.ir

Abstract

In this talk we review some recent results on constructing multivariate copulas.

Keywords and phrases: Multivariate copula, Dependence, Measure of association, Multivariate distribution.

AMS Subject Classification 2000: 62E15; Secondary 62H10.

1 Introduction

The construction of multivariate distributions with given margins has been a problem of interest to statisticians for many years. The past decade has seen an incredible evolution in copulas (distributions with uniform univariate margins) and their applications. Sklar's Theorem (Sklar, 1959) establishes the grounds for separate investigations of the marginal distribution and the dependence structure that empowers copula as a more flexible modelling technique than the conventional multivariate approach. Nelsen (2006) summarizes different methods of constructing copulas. Any joint distribution can be uniquely expressed as a copula function of individual marginal distributions provided they are continuous. Conversely, copulas can be constructed from multivariate distributions without any constraints on marginal distributions. A natural framework for the construction of multivariate non-normal distributions is the method of copulas, justified by the Sklar's theorem. According to Joe [(1997), Section 4.1], a parametric family of distributions should satisfy four desirable properties:

(a) There should exist an interpretation like a mixture or other stochastic representation.

(b) The margins, at least the univariate and bivariate ones, should belong to the same parametric family and numerical evaluation should be possible.

(c) The bivariate dependence between the margins should be described by a parameter and cover a wide range of dependence.

(d) The multivariate distribution and density should preferably have a closed-form representation; at least numerical evaluation should be possible.

In general, these desirable properties cannot be fulfilled simultaneously. For example, multivariate normal distributions satisfy properties (a), (b), and (c) but not (d). The method of copulas satisfies property (c) but implies only partial closedness under the taking of margins, and can lead to computational complexity as the dimension increases. It is yet an open problem to find parametric families of copulas that satisfy all of the desirable properties. In this talk we review some recent results on constructing multivariate copulas.

2 Preliminaries

Let $n \geq 2$ be a natural number. An n-copula is the restriction to $[0,1]^n$ of a continuous multivariate distribution function whose univariate margins are uniform on $[0,1]$. More precisely:

Definition 1. An n-copula is a function $C\colon [0,1]^n \longrightarrow [0,1]$ which satisfies:

(C1) For every $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ in $[0,1]^n$, $C(\mathbf{u}) = 0$ if at least one coordinate of $\mathbf{u}$ is 0, and $C(\mathbf{u}) = u_k$ whenever all coordinates of $\mathbf{u}$ are 1 except u_k;

(C2) C is n increasing; i.e., for every $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_n)$ in $[0,1]^n$ such that $a_k \leq b_k$ for all $k = 1, 2, \ldots, n$,

$$V_C([\mathbf{a}, \mathbf{b}]) = \sum \operatorname{sgn}(\mathbf{c}) C(\mathbf{c}) \geq 0$$

where $[\mathbf{a}, \mathbf{b}]$ denotes the *n-box*$[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$, the sum is taken over all the *vertices* $\mathbf{c} = (c_1, c_2, \ldots, c_n)$ of $[\mathbf{a}, \mathbf{b}]$ such that each c_k is equal to either a_k or b_k, and $\operatorname{sgn}(\mathbf{c})$ is 1 if $c_k = a_k$ for an even number of k's, and -1 if $c_k = a_k$ for an odd number of k's.

The importance of copulas is described in the following result: Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vector with joint distribution function H and respective margins $F_1, \ldots, F_n$. Then there exists an n-copula C (which is uniquely determined on $\text{Range}F_1 \times \cdots \times \text{Range}F_n$) such that

$$H(x_1, \ldots, x_n) = C\{F_1(x_1), \ldots, F_n(x_n)\},$$

for all $\mathbf{x} = (x_1, \ldots, x_n)$ in $[-\infty, \infty]^n$. If $F_1, \ldots, F_n$ are continuous, then C is unique. In the sequel, we shall suppose that $\mathbf{X}$ is continuous. For a complete review, see Nelsen (2006).

If $\mathbf{u} = (u_1, \ldots, u_n) \in [0,1]^n$, $\Pi^n(\mathbf{u}) = \prod_{i=1}^{n} u_i$ denotes the n-copula of independent continuous random variables. For every $\mathbf{u}$ in $\mathbb{I}^n$, any n-copula C satisfies that

$$W^n(\mathbf{u}) = \max(\sum_{i=1}^{n} u_i - n + 1, 0) \leq C(\mathbf{u}) \leq \min(u_1, u_2, \ldots, u_n) = M^n(\mathbf{u}).$$

For every $n \geq 2$, M^n is an n-copula; however W^n is an n-copula if and only if $n = 2$. Let $\mathbf{X} > \mathbf{x}$ denote the component-wise inequality, and let $\mathbf{1}$ be an n-dimensional vector of $1's$. If $\mathbf{U}$ is a vector of uniform $[0,1]$ random variables with n-copula C, then $\hat{C}(\mathbf{u}) = P[\mathbf{1} - \mathbf{U} < \mathbf{u}]$ and $\overline{C}(\mathbf{u}) = P[\mathbf{U} > \mathbf{u}]$ are the *survival copula* and the *survival function*, respectively, associated to C. Given two n-copulas C_1 and C_2, let $C_1 \leq C_2$ denote the inequality $C_1(\mathbf{u}) \leq C_2(\mathbf{u})$ for all $\mathbf{u}$. C is *radially symmetric* if $C = \hat{C}$. $C(\pi\mathbf{u})$ denotes the n-copula given by any permutation π of $\mathbf{u}$.

The population versions of two non-parametric measures of multivariate association between components of a continuous random vector $\mathbf{U}$ with associated n-copula C are two multivariate generalizations of Kendall's tau and the Spearman's rho coefficient, which are given by

$$\tau_n(C) = \frac{1}{2^{n-1} - 1}\left(2^n \int_{[0,1]^n} C(\mathbf{u}) \mathrm{d}C(\mathbf{u}) - 1\right) \quad \text{and}$$

$$\rho_n(C) = \frac{n+1}{2^n - (n+1)}\left[2^{n-1}\left(\int_{[0,1]^n} C(\mathbf{u}) \mathrm{d}\Pi^n(\mathbf{u}) + \int_{[0,1]^n} \Pi^n(\mathbf{u}) \mathrm{d}C(\mathbf{u})\right) - 1\right],$$

respectively ; see Nelsen (2002).

3 Multivariate FGM type distributions

The well-known Farle-Gumbel-Morgrnstern (FGM) family of copulas is a copula of the form

$$C(u, v) = uv[1 + \theta(1 - u)(1 - v)], \quad \theta \in [-1, 1]$$

for all $u, v \in [0,1]$. The natural multivariate generalization of this copula could be

$$C(u_1, ..., u_n) = \prod_{i=1}^{n} u_i[1 + \theta \prod_{i=1}^{n}(1 - u_i)], \quad \theta \in [-1, 1].$$

Rodríguez-Lallena and Úbeda-Flores (2004) proposed and studied a family of bivariate copulas of the form

$$C_\lambda(u, v) = uv + \lambda f(u) g(v), \quad 0 \leq u, v \leq 1, \tag{1}$$

where f and g are two absolutely continuous functions such that $f(0) = f(1) = g(0) = g(1) = 0$, and

$$\lambda \in [-1/\max(\alpha\gamma, \beta\delta), -1/\min(\alpha\delta, \beta\gamma)],$$

where

$$\alpha = \inf\{f'(u) : u \in A\}, \quad \beta = \sup\{f'(u) : u \in A\},$$

$$\gamma = \inf\{g'(v) : v \in B\}, \quad \delta = \sup\{g'(v) : v \in B\},$$

with

$$A = \{u \in [0,1] : f'(u)\text{exists}\}, \quad B = \{v \in [0,1] : g'(v)\text{exists}\}.$$

Some multivariate extensions for this model are introduced by Dolati and Úbeda-Flores (2006).

3.1 First multivariate extension

A natural extension of (1) is a multivariate copula of the form

$$C_\theta(u_1, ..., u_n) = \prod_{i=1}^{n} u_i + \theta \prod_{i=1}^{n} f_i(u_i), \tag{2}$$

where f_i, $1 \le i \le n$, are n non-zero absolutely continuous functions such that $f_i(0) = f_i(1) = 0$ and

$$\frac{-1}{\sup\limits_{(u_1,...,u_n)\in D^+}\left(\prod\limits_{i=1}^{n} f_i'(u_i)\right)} \le \theta \le \frac{-1}{\inf\limits_{(u_1,...,u_n)\in D^-}\left(\prod\limits_{i=1}^{n} f_i'(u_i)\right)},$$

$D^- = \{(u_1, ..., u_n) \in [0,1] : \prod_{i=1}^{n} f_i'(u_i) < 0\}$ and $D^+ = \{(u_1, ..., u_n) \in [0,1]^n : \prod_{i=1}^{n} f_i'(u_i) > 0\}$.

Different properties of this family are studied in Dolati and Úbeda-Flores (2006). The survival function and the survival n-copula associated to this family are given by

$$\begin{aligned}\overline{C}_\theta(u_1, ..., u_n) &= P(U_1 > u_1, ..., U_n > u_n)\\ &= \prod_{i=1}^{n}(1-u_i) + (-1)^n\theta\prod_{i=1}^{n} f_i(u_i)\end{aligned}$$

and

$$\begin{aligned}\hat{C}_\theta(u_1, ..., u_n) &= P(1-U_1 \le u_1, ..., 1-U_n \le u_n)\\ &= \prod_{i=1}^{n} u_i + (-1)^n\theta\prod_{i=1}^{n} f_i(1-u_i).\end{aligned}$$

Observe that for this class of n-copulas we have the relationship

$$\rho(C_\theta) = \frac{(n+1)(2^{n-1}-1)\tau(C_\theta)}{2(2^n-n-1)}.$$

For $n = 2$, it reduces to $\frac{\rho(C)}{\tau(C)} = \frac{3}{2}$.

Example. Let $f_i(u) = u^b(1-u)^a$, $1 \le i \le 3$, with $a, b \ge 1$. Then

$$C_\theta(u_1, u_2, u_3) = u_1u_2u_3 + \theta u_1^b(1-u_1)^a u_2^b(1-u_2)^a u_3^b(1-u_3)^a,$$

is a 3-copula if and only if

$$-\min\left\{\frac{1}{(K_{ab}^+)^2K_{ab}^-}, \frac{1}{(K_{ab}^-)^3}\right\} \le \theta \le \min\left\{\frac{1}{(K_{ab}^-)^2K_{ab}^+}, \frac{1}{(K_{ab}^+)^3}\right\}.$$

with $K_{ab}^{+} = (R(a,b))^{b-1}(1-R(a,b))^{a-1}(R(a,b)-b)$ and $K_{ab}^{-} = (S(a,b))^{b-1}(1-S(a,b))^{a-1}(S(a,b)-b)$, where

$$R(a,b) = \frac{b(a+b-1)+\sqrt{ab(a+b-1)}}{(a+b)(a+b-1)}, \quad \text{and}$$

$$S(a,b) = \frac{b(a+b-1)-\sqrt{ab(a+b-1)}}{(a+b)(a+b-1)}.$$

3.2 Second multivariate extension

Let f and g be non-zero absolutely continuous functions such that $f(0) = f(1) = 0$ and $g(0) = g(1) = 0$. Let α be a real number, and consider the function C_α^* given by

$$C_\alpha^*(\mathbf{u}) = \prod_{i=1}^{n} u_i + \alpha \sum_{1 \le i < j \le n} f(u_i)g(u_j) \prod_{\substack{k=1 \\ k \ne i,j}}^{n} u_k, \quad \mathbf{u} \in [0,1]^n. \tag{3}$$

Then each k-margin, $2 \le k < n$, of C_α^* is again in this parametric family. C_α^* is an absolutely continuous function with density $c_\alpha^*(\mathbf{u}) = 1 + \alpha \sum\limits_{1 \le i < j \le n} f'(u_i)g'(u_j)$. Thus, the above function is an n-copula if and only if

$$\frac{-1}{\max\{\sum\limits_{1 \le i < j \le n} f'(u_i)g'(u_j)\}} \le \alpha \le \frac{-1}{\min\{\sum\limits_{1 \le i < j \le n} f'(u_i)g'(u_j)\}}.$$

Different properties of this family are studied in Dolati and Úbeda-Flores (2006). The measures Kendall's tau and the Spearman's rho associated with this family are given by

$$\tau(C_\alpha^*) = \frac{4\alpha n(n-1)}{2^{n-1}-1} \int_0^1 f(u)\mathrm{d}u \int_0^1 g(u)\mathrm{d}u$$

and

$$\rho(C_\alpha^*) = \frac{2\alpha n(n^2-1)}{2^n-(n+1)} \int_0^1 f(u)\mathrm{d}u \int_0^1 g(u)\mathrm{d}u.$$

Observe that for this class of n-copulas we have, again, the relationship

$$\rho(C_\alpha^*) = \frac{(n+1)(2^{n-1}-1)\tau(C_\alpha^*)}{2(2^n-n-1)}.$$

3.3 Third multivariate extension

Let $\Theta = \{\theta_{i_1 \cdots i_k} : 1 \le i_1 < \cdots < i_k \le n,\ 2 \le k \le n\}$ be a set of $2^n - n - 1$ parameters. A third class of distributions which generalizes (1) can be the following (2^n-n-1)-parameter family:

$$C_\Theta^{**}(u_1,...,u_n) = \prod_{i=1}^{n} u_i + \sum_{k=2}^{n} \sum_{1 \le i_1 < \cdots < i_k \le n} \theta_{i_1 \cdots i_k} \prod_{j=i_1}^{i_k} f_j(u_j) \prod_{\substack{l=1 \\ l \ne i_1,...,i_k}}^{n} u_l, \tag{4}$$

with $f_i(0) = f_i(1) = 0$ and

$$1 + \sum_{k=2}^{n} \sum_{1 \le i_1 < \cdots < i_k \le n} \theta_{i_1 \cdots i_k} \prod_{j=i_1}^{i_k} f_j'(u_j) \ge 0$$

Each k-margin, $2 \le k < n$, of a member this family is again an k-copula of the same type.

The survival copula associated to C_Θ^{**} is given by

$$\hat{C}_\Theta^{**}(u_1,...,u_n) = \prod_{i=1}^{n} u_i + \sum_{k=2}^{n} (-1)^k \sum_{1 \le i_1 < \cdots < i_k \le n} \theta_{i_1 \cdots i_k} \prod_{j=i_1}^{i_k} f_j(1-u_j) \prod_{\substack{l=1 \\ l \ne i_1,...,i_k}}^{n} u_l.$$

The Kendall's tau and Spearman'd rho for this family are given by

$$\tau(C^{**}_\Theta) = \frac{2}{2^{n-1}-1}\sum_{k=2}^{n} 2^{k-1}(1+(-1)^k)\sum_{1\le i_1<\cdots<i_k\le n}\theta_{i_1\cdots i_k}\prod_{j=i_1}^{i_k}\int_0^1 f_j(u_j)\,\mathrm{d}u_j,$$

$$\rho(C^{**}_\Theta) = \frac{n+1}{2^{n}-n-1}\sum_{k=2}^{n} 2^{k-1}(1+(-1)^k)\sum_{1\le i_1<\cdots<i_k\le n}\theta_{i_1\cdots i_k}\prod_{j=i_1}^{i_k}\int_0^1 f_j(u_j)\,\mathrm{d}u_j.$$

Again, we have the relationship

$$\rho_n(C^{**}_\Theta) = \frac{(n+1)(2^{n-1}-1)\tau_n(C^{**}_\Theta)}{2(2^n-n-1)}.$$

3.4 Forth multivariate extension

For a given bivariate copula C^*, Kim and Sugur (2004), considered a copula of the form

$$C(u,v) = C^*(u,v) + \theta f(u)g(v), \tag{5}$$

where f and g are absolutely continuous functions with $f(1) = g(1) = f0) = g(0) = 0$ and

$$\theta \in [-1/\max(\alpha\gamma,\beta\delta), -1/\min(\alpha\delta,\beta\gamma)], \quad \min(\alpha\delta,\beta\gamma) \ge -1,$$

$$\alpha = \inf\{f'(u) : u \in A\}, \quad \beta = \sup\{f'(u) : u \in A\},$$

$$\gamma = \inf\{g'(v) : v \in B\}, \quad \delta = \sup\{g'(v) : v \in B\},$$

with

$$A = \{u \in [0,1] : f'(u)\text{exists}\}, \quad B = \{v \in [0,1] : g'(v)\text{exists}\}.$$

A simple subclass of this model is

$$C(u,v) = C^*(u,v) + \theta[\psi_1(u)-\psi_2(u)][\psi_1(v)-\psi_2(v)],$$

where ψ_1 and ψ_2 are two absolutely continuous functions such that $\psi_1(0) = \psi_1(1) = \psi_2(0) = \psi_2(1) = 0$. This function is also a copula under the condition

$$[\psi_1'(u)-\psi_2'(u)][\psi_1'(v)-\psi_2'(v)] \ge 0,$$

$$\frac{-\frac{\partial C^*(u,v)}{\partial u\partial v}}{[\psi_1'(u)-\psi_2'(u)][\psi_1'(v)-\psi_2'(v)]} \le \theta,$$

for all $u, v \in [0,1]$, see Kim and Sugur (2004).

A multivariate extension of this family could be constructed as follows. Given a n-copula C^*, consider the function

$$C(u_1,...,u_n) = C^*(u_1,...,u_n) + \theta\prod_{i=1}^{n}[\psi_1(u_i)-\psi_2(u_i)].$$

It is of interest to find conditions for which this function a n-copula.

3.5 Fifth multivariate extension

Let D be a $(n-1)$-copula and $A:[0,1]^{n-1}\to[0,1]$ and $f:[0,1]\to[0,1]$ be two functions with $f(0)=f(1)=0$, $A(1,...,1)=0$ and $A(u_1,...,u_{n-1})=0$, if at least one component is zero. A new extension for FGM type copulas proposed in Durante et al. (2010) by

$$C(u_1,...,u_n)=D(u_1,...,u_{n-1})u_n+A(u_1,...,u_{n-1})f(u_n).$$

They provided conditions for which C is a copula.
Example. With $D(u_1,...,u_{n-1})=\prod_{i=1}^{n-1}u_i$ and $A=\prod_{i=1}^{n-1}u_i(1-u_i)$, they showed that

$$C(u_1,...,u_n)=\prod_{i=1}^{n}u_i+\prod_{i=1}^{n-1}u_i(1-u_i)f(u_n),$$

is an n-copula if and only if f is 1-Lipschitz function. For instance with $f(t)=\lambda t$, $\lambda\in[-1,1]$, one get an n- copula of the form

$$C(u_1,...,u_n)=\prod_{i=1}^{n}u_i[1+\lambda\prod_{i=1}^{n-1}(1-u_i)].$$

With $f(t)=t(1+t)$, $D(u_1,...,u_{n-1})=u_1u_2H(u_3,...,u_{n-1})$, and $A(u_1,...,u_{n-1})=\lambda u_1(1-u_1)u_2(1-u_2)H(u_3,...,u_{n-1})$, where H is a $(n-3)$-copula and $\lambda\in[-1/3,1/3]$, we have

$$C(u_1,...,u_n)=u_1u_2u_nH(u_3,...,u_{n-1})[1+\lambda(1-u_1)(1-u_2)(1+u_n)].$$

4 Compatibility Problem

A difficult problem related to the theory of multivariate distributions is to construct a multivariate distribution with prescribed multivariate margins. Some aspects of this problem including existence of such distributions, compatibility and methods of constructing are discussed, for instance, by Dall'Aglio (1972), Cohen (1984), Rüschendorf (1985), Cuadras (1992), Marco and Ruiz-Rivas (1992), Li et al. (1996, 1999) and Joe (1997). The problem can be formulated as follows: For $1<m\le n$, n-copula C has $\binom{n}{m}$ m-marginals. If an n-copula exists for which the given copulas are the m-marginals, then these are said to be compatible.
Example. Let X,Y and Z be three random variables with

$$C_{X,Y}=W,C_{X,Z}=W,C_{Y,Z}=W.$$

Is there a 3-copula C such that

$$C(x,y,1)=W(x,y),C(x,1,z)=W(x,z),C(1,y,z)=W(y,z)?$$

The answer is negative. Since we must have $X=g(Y)$, $X=h(Z)$ and $Y=h(Z)$, with f,g,h decreasing, which is impossible.

4.1 Some partial answers

Let A,B be two 2-copulas. The $*$-product operation is defined in Darsow et al. (1992) by

$$A*B(x,y)=\int_0^1 D_2A(x,t)D_1B(t,y)dt$$

where

$$D_1C(x,y)=\frac{\partial}{\partial x}C(x,y),\quad D_2C(x,y)=\frac{\partial}{\partial y}C(x,y).$$

They showed that $A*B$ is also a copula. Kolesárová et al. (2006) defined a function C by

$$C(x,y,z)=\int_0^y D_2A(x,t)D_1B(t,z)dt.$$

They showed that it is a 3-copula and

$$C(x,y,1)=A(x,y),C(1,y,z)=B(y,z),C(x,1,z)=A*B(x,z),$$

i.e., the copulas A, B and $A*B$ are always compatible.

Note that

$$A*B(x,y)=\int_0^1 \Pi\{D_2A(x,t),D_1B(t,y)\}dt,$$

where $\Pi(x,y)=xy$. For given 2-copula A,B and C, Durante et al. (2008) introduced the generalized $*$-product operation by

$$A*_C B(x,y)=\int_0^1 C\{D_2A(x,t),D_1B(t,y)\}dt.$$

They showed that it is also a copula. They showed that the function

$$E(x,y,z)=\int_0^y C\{D_2A(x,t),D_1B(t,z)\}dt.$$

is also a 3-copula and

$$E(x,y,1)=A(x,y),E(1,y,z)=B(y,z),E(x,1,z)=A*_C B(x,z),$$

i.e., the copulas A, B and $A*_C B$ are always compatible.

4.2 A multivariate generalization

Let $A_1,...,A_n$ be a 2-copula. Consider the function $C:[0,1]^n\to[0,1]$ by

$$C(u_1,...,u_n)=\int_0^1 \frac{\partial A_1(u_1,t)}{\partial t}\frac{\partial A_2(u_2,t)}{\partial t}....\frac{\partial A_n(u_n,t)}{\partial t}dt.$$

Then C is an n-copula. Note that

$$\begin{aligned}C_{12}(u_1,u_2) &= C(u_1,u_2,1,...,1)=\int_0^1 \frac{\partial A_1(u_1,t)}{\partial t}\frac{\partial A_2(u_2,t)}{\partial t}dt\\ &= \int_0^1 \frac{\partial A_1(u_1,t)}{\partial t}\frac{\partial A_2^T(t,u_2)}{\partial t}dt\\ &= A_1*A_2^T(u_1,u_2),\end{aligned}$$

where $A^T(u,v)=A(v,u)$.

Let $U_1,...,U_n$ and Z be uniform $[0,1]$ random variables such that $(U_1,Z),...,(U_n,Z)$ are independent random vectors with the 2-copulas $A_1,...,A_n$. Then

$$\begin{aligned}P(U_1\le u_1,...,U_n\le u_n|Z=t) &= \prod_{j=1}^n P(U_j\le u_j|Z=t)\\ &= \prod_{j=1}^n \frac{\partial A_j(u_j,t)}{\partial t},\end{aligned}$$

and thus

$$\begin{aligned}C(u_1,...,u_n) &= P(U_1\le u_1,...,U_n\le u_n)\\ &= \int_0^1 P(U_1\le u_1,...,U_n\le u_n|Z=t)dt\\ &= \int_0^1 \prod_{j=1}^n \frac{\partial A_j(u_j,t)}{\partial t}dt.\end{aligned}$$

Example. For $j = 1, ..., n$, let $A_j(u,t) = \min\{ut^\alpha, u^\alpha t\}$, with $\alpha \in [0,1]$, be members of the Cuadras-Auge (1981) family of copulas. Direct calculations shows that

$$\int_0^1 \prod_{j=1}^{n} \frac{\partial A_j(u_j,t)}{\partial t} dt = u_{(1)} \prod_{j=2}^{n} u_{(j)}^{1-\alpha}, \tag{6}$$

where $u_{(1)} \leq u_{(2)} \leq ... \leq u_{(n)}$ denote the components of $(u_1, ..., u_n) \in \mathbb{I}^n$ rearranged in increasing order.

5 Multivariate copulas with given bivariate marginals

Dolati and Ubeda-Flores (2005) proposed a method for construction multivariate copulas with given bivariate margins.

Let $\{C_{ij} : 1 \leq i < j \leq n\}$ be a set of $\binom{n}{2}$ 2-copulas, and let

$$C(u_1, ..., u_n) = \sum_{1\leq i<j\leq n} C_{ij}(u_i, u_j) \prod_{\substack{k=1 \\ k\neq i,j}}^{n} u_k - \frac{(n-2)(n+1)}{2} \prod_{i=1}^{n} u_i.$$

Then C is an n-copula whose bivariate margins are C_{ij} if and only if

$$\sum_{1\leq i<j\leq n} \frac{C_{ij}(v_i, v_j) - C_{ij}(v_i, u_j) - C_{ij}(u_i, v_j) + C_{ij}(u_i.u_j)}{(v_i - u_i)(v_j - u_j)} \geq \frac{(n-2)(n+1)}{2},$$

for every u_k, v_k in $[0,1]$, $k = 1, 2, \ldots, n$, such that $u_k < v_k$.

The expression for the Spearman's rho for this family is

$$\rho(C) = \frac{n+1}{3[2^n - (n+1)]} \cdot \sum_{1\leq i<j\leq n} \rho(C_{ij}),$$

where $\rho(C_{ij})$ denotes the Spearman's rho coefficient associated to C_{ij}. For $n = 3$, we have

$$\rho(C) = \frac{\rho(C_{12}) + \rho(C_{13}) + \rho(C_{23})}{3}.$$

Example. Let

$$\begin{aligned} C_{12}(u_1, u_2) &= u_1 u_2, \\ C_{13}(u_1, u_3) &= \alpha u_1 u_3 + (1-\alpha) M(u_1, u_3), \\ C_{23}(u_2, u_3) &= \beta u_2 u_3 + (1-\beta) W(u_2, u_3) \end{aligned}$$

where $\alpha, \beta \in [0,1]$, and

$$M(u_1, u_3) = \min(u_1, u_3), \quad W(u_2, u_3) = \max(u_2 + u_3 - 1, 0).$$

Then, it could be proved that the function

$$C(u_1, u_2, u_3) = (\alpha + \beta - 1) u_1 u_2 u_3 + (1-\alpha) u_2 M(u_1, u_3) + (1-\beta) u_1 W(u_2, u_3)$$

is a 3-copula if and only if

$$\alpha + \beta \geq 1.$$

Example. Let $\{C_{\lambda_{ij}} : 1 \leq i < j \leq n\}$ be a set of $\binom{n}{2}$ FGM 2-copulas

$$C_{ij}(u, v) = uv[1 + \lambda_{ij} uv(1-u)(1-v)], \quad \lambda_{ij} \in [-1, 1].$$

Then the function C defined by

$$C(u_1, ..., u_n) = \sum_{1\le i<j\le n} C_{ij}(u_i, u_j) \prod_{\substack{k=1\\k\neq i,j}}^{n} u_k - \frac{(n-2)(n+1)}{2}\prod_{i=1}^{n} u_i,$$

is an n-copula if and only if

$$\sum_{1\le i<j\le n} \lambda_{ij}(1-2u_i)(1-2u_j) \ge -1.$$

Thus, a sufficient condition is that $|\sum_{1\le i<j\le n} \lambda_{ij}| \le 1$.

As a particular case, if all the bivariate margins are

$$C_{ij}(u,v) = uv[1+\lambda(1-u)(1-v)],$$

for all (i,j), then the function C is an copula if and only if

$$-\frac{2}{n(n-1)} \le \lambda \le \frac{1}{\lfloor \frac{n}{2} \rfloor}.$$

where $\lfloor x \rfloor$ denotes the integer part of the real number x.In such a case,

$$\rho(C) = \frac{n(n-1)(n+1)\lambda}{18[2^n-(n+1)]}.$$

5.1 An application

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables such that X_n has the type I Extreme Value distribution

$$F_n(x) = e^{-e^{\frac{-(x-a_n)}{b}}}, \quad x \in \mathbb{R},$$

where $\{a_n\}_{n\in\mathbb{N}}$ is a sequence of real numbers and $b > 0$. For any $n > 2$, let C_n consider the copula

$$C_n(u_1, ..., u_n) = \sum_{1\le i<j\le n} C_{ij}(u_i, u_j) \prod_{\substack{k=1\\k\neq i,j}}^{n} u_k - \frac{(n-2)(n+1)}{2}\prod_{i=1}^{n} u_i.$$

with

$$C_{ij}(u,v) = uv[1+\theta_{|i-j|}(1-u)(1-v)],$$

with the parameterization $\theta_{|i-j|} = \frac{1}{\kappa_0|i-j|^\alpha}$, for $\alpha > 1$, where $\kappa_0 > \zeta(\alpha)$, and $\zeta(\alpha) = \sum_{j=1}^{\infty} \frac{1}{j^\alpha}$ is the so-called Riemman zeta function.

Let C_n, be the copula associated to the vector $(X_0, ..., X_{n-1})$, which implies that the copula related to (X_r, X_s) is C_{rs}, $r, s \in \{0, 1, ..., n-1\}$ and $r \neq s$. Then

$$Cov(X_t, X_{t+h}) = \frac{\ln(2)^2}{b^2\kappa_0} h^{-\alpha}.$$

Since the above construction is an n-copula for any $n > 2$, Sklar's theorem guarantees the existence of all finite dimensional distribution functions with the marginals, bivariate copulas and n-dimensional copulas as specified in the construction above. Therefore, by the Kolmogorov's existence theorem, we have just constructed a weakly stationary process $\{X_n\}$ by reparameterizing a certain family of parametric copulas and its marginals.

6 Multivariate extensions of Cuadras-Augé family of copulas

For $\alpha \in [0,1]$, Cuadras and Augé (1981) proposed a family of copulas of the form

$$\begin{aligned} C_\alpha(u,v) &= \min\{uv^\alpha, u^\alpha v\} \\ &= \min(u,v)\left(\max(u,v)\right)^\alpha, \quad u,v \in [0,1]. \end{aligned}$$

For $f : [0,1] \to [0,1]$, Durante (2007) considered a function

$$C(u,v) = \min(u,v) f\left(\max(u,v)\right).$$

The function C is a copula if and only if $f(1) = 1$ and f is increasing, and $\frac{f(t)}{t}$ is decreasing on (0,1].

When $f(t) = t^\alpha$, it reduces to Cuadras-Aug'e family of copulas.

Durante, Quesada-Molina and Úbeda-Flores (2007) considered a multivariate extension of the above model of the form

$$C(u_1, ..., u_n) = u_{[1]} \prod_{i=2}^{n} f(u_{[i]}),$$

where $u_{[1]}, ..., u_{[n]}$ denote the components of $(u_1, ..., u_n)$ arranged in increasing order; i.e., $u_{[1]} = \min(u_1, ..., u_n)$ and $u_{[n]} = \max(u_1, ..., u_n)$. They proved that for every $n \geq 2$, the function C is a n-copula if and only if $f(1) = 1$ and f is increasing, and $\frac{f(t)}{t}$ is decreasing on (0,1].

6.1 A probabilistic interpretation

Let $W_1, ..., W_n$ and Z be $n+1$ independent random variables. Let $W_i \sim f(t)$ and $Z \sim g(t) = \frac{t}{f(t)}$.

Consider $U_i = \max(W_i, Z)$, for $i = 1, ..., n$. Then

$$\begin{aligned} P(U_1 \leq u_1, ..., U_n \leq u_n) &= P(W_1 \leq u_1, ..., W_n \leq u_n, Z \leq u_{[1]}) \\ &= u_{[1]} \prod_{i=2}^{n} f(u_{[i]}) \\ &= C(u_1, ..., u_n). \end{aligned}$$

6.2 Another possible generalization

Consider a continuous and increasing function $\phi : [0,1] \to [0,1]$ and suppose that ϕ^{-1} is absolutely monotonic of order n; i.e.,

$$\frac{d^i(\phi^{-1}(t))}{dt^i} \geq 0, \quad \text{for} \quad i = 1, 2, ..., n.$$

From Morillas (2005), Theorem 4.7, for every n- copula C we have that

$$C_\phi(u_1, ..., u_n) = \phi^{-1}\{C(\phi(u_1), ..., \phi(u_n))\}$$

is also a copula.

In particular, if

$$C(u_1, ..., u_n) = u_{[1]} \prod_{i=2}^{n} f(u_{[i]}),$$

then the function

$$C_{\phi,f}(u_1, ..., u_n) = \phi^{-1}\left[\phi(u_{[1]}) \prod_{i=2}^{n} f(\phi(u_{[i]}))\right]$$

is also a copula.

For $g_1(t) = -\ln\phi(t)$ and $g_2(t) = -\ln(f(\phi(t))$, the above family can be written as

$$C_{g_1,g_2}(u_1, ..., u_n) = g_1^{-1}\left[g_1(u_{[1]}) + \sum_{i=2}^{n} g_2(u_{[i]})\right].$$

For the case $g_1 = g_2$, we get a direct generalization of the Archimedean family with the generator g

$$C_g(u_1, ..., u_n) = g^{-1}\left[g(u_1) + ... + g(u_n)\right].$$

References

[1] Cohen, L. (1984). Probability distributions with given multivariate marginals. *J. Math. Phys.* **25**, 2402–2403.

[2] Cuadras, C. (1992). Probability distributions with given multivariate marginals and given dependence structure. *J. Multivariate Anal.* **42**, 51–66.

[3] Cuadras, C.M., Aug; J. (1981). A continuous general multivariate distribution and its properties, *Comm. Statist. A Theory Methods*, **10**, 339–353.

[4] Darsow, W., Nguyen, B.A. and Olsen, E.T. (1992). Copulas and Markov processes, *Illinois J. Math*, **36**, 600–642.

[5] Dall'Aglio, G. (1972). Fréchet classes and compatibility of distribution functions. *Symp. Math.* **9**, 131–150.

[6] Dolati, A. and Úbeda-Flores, M. (2006). Some new parametric families of multivariate copulas. *Int. Math. Forum*, **1**, 17–25.

[7] Dolati, A. and Úbeda-Flores, M. (2005). A method for constructing multivariate distributions with given bivariate margins. *Brazilian Journal of Probability and Statistics*, **19**, 85–92.

[8] Durante, F. (2007). A new family of symmetric bivariate copulas, *C.R. Math. Acad. Sci. Paris*, **344**, 195–198.

[9] Durante, F., Quesada-Molina, J.J., Úbeda-Flores, M. (2007). On a family of multivariate copulas for aggregation processes. *Inf Sci*, **177**, 5715–5724.

[10] Durante, F., Klement, P.E. and Quesada-Molina, J.J. (2008) Bounds for Trivariate Copulas with Given Bivariate Marginals. *Journal of Inequalities and Applications doi:10.1155/2008/161537.*

[11] Durante, F., Foscolo, E., Rodrýguez-Lallena, J. and Úbeda-Flores, M. (2010). A method for constructing higher-dimensional copulas.

[12] Joe, H. (1997). *Multivariate Models and Dependence Concepts*, Monographs on Statistics and Applied Probability, vol. 73, Chapman & Hall, London.

[13] Kolesárová, A., Mesiar, R., Sempi, C. (2006). Three copulas and compatibility. In: Bouchon-Meunier, B., Yager, R., Eds., Proceedings of IPMU 2006, Vol. 1. Paris: Éditions E.D.K.

[14] Li, X. and Fang, R. (2012). A New Family of Bivariate Copulas Generated by Univariate Distributions, *Journal of Data Science*, **10**, 661–677.

[15] Li, H., Scarsini, M. and Shaked, M. (1996). Linkages: a tool for the construction of multivariate distributions with given nonoverlapping multivariate marginals. *J. Multivariate Anal.* **56**, 20–41.

[16] Li, H., Scarsini, M. and Shaked, M. (1999). Dynamic linkages for multivariate distributions with given nonoverlapping multivariate marginals. *J. Multivariate Anal.* **68**, 54–77.

[17] Marco, J.M. and Ruiz-Rivas, C. (1992). On the construction of multivariate distributions with given nonoverlapping multivariate marginals. *Statist. Probab. Lett.* **15**, 259–265.

[18] Morillas, P.M. (2005). A method to obtain new copulas from a given one, *Metrika*, **61**, 169–184.

[19] Kim, J.M., Sungur, E.A. (2004). New class of bivariate copulas. *Proceeding of the Spring Conference, Korean Statistical Society.*

[20] Nelsen, R.B. (1996). Nonparametric measures of multivariate association, In: Rüschendorf, L., Schweizer, B., Taylor, M.D. (Eds) Distributions with Fixed Marginals and Related Topics, Institute of Mathematical Statistics, Hayward, CA, pp. 223–232.

[21] Nelsen, R.B. (2002). Concordance and copulas: A survey, In: Cuadras, C., Fortiana, J., Rodríguez-Lallena, J.A. (Eds.) Distributions with Given Marginals and Statistical Modelling, Kluwer Academic Publishers, Dordrecht, pp. 169–178.

[22] Nelsen, R.B. (2006). *An Introduction to Copulas. Second Edition.* Springer, New York.

[23] Rodríguez-Lallena, J.R. and Úbeda-Flores, M. (2004). A new class of bivariate copulas, *Statistics & Probability Letters,* **66**, 315 - 325.

[24] Rüschendorf, L. (1985). Construction of multivariate distributions with given marginals. *Ann. Inst. Statist. Math.* **37**, 225–233.

GLM-method to estimate a copula's parameter(s)

Payandeh Najafabadi A. ,Farid-Rohani M. ,Qazvini M.

a Faculty member of Mathematical Sciences Department, Shahid Beheshti University,G.C. Evin, 1983963113, Tehran, Iran.

b E.C.O College of Insurance, Allameh Tabatabái University, Tehran, Iran.

Abstract

This study introduces a new approach to problem of estimating parameter(s) of a given copula. More precisely, using the concept of the generalized linear models (GLM) accompanied with least square method, we introduce an estimation method, say GLM-method. A simulation study has been conducted to provide a comparison among the inversion of Kendal's tau, the inversion of Spearman's rho, the pml, the Copula-quantile regression with ($q = 0.25, 0.50, 0.75$), and the GLM-method. Such simulation study shows that the GLM-method is an appropriate method whenever the data distributed according to an elliptical distribution.

Keywords: Parameter estimation; Copula; GLM; Copula-quantile regression

1 Introduction

Copula is used to model the relationship between random variables. It can capture the interdependency that cannot be exhibited by other association measures such as the well-known correlation coefficient. One step in copula modeling is the estimation of parameters. The most efficient method is the maximum likelihood estimator (mle), which is used to evaluate the parameter of any kind of models. It can also be applied to copula, but the problem becomes complicated as the number of parameters and dimension of copula increases, because the parameters of the margins and copula are estimated simultaneously. Therefore, MLE is highly affected by misspecification of marginal distributions. A rather straightforward way at the cost of lack of efficiency is IFM, which is put forward by Joe (2005). The idea of this method came from psychometrics literature for latent models based on the multivariate normal distributions. Similar to MLE in this method the margins of the copula are important, because the parameter estimation is dependent on the choice of the marginal distributions. First the margins' parameters are estimated and then the parameters of copula will be evaluated given the values from the first step.The efficiency of this method is 1 for product copula under some conditions. Efficiency decreases with strong dependence. IFM is not a good estimation technique due to its efficiency for extreme dependence near Fréchet bounds. Both MLE and IFM are placed in the category of parametric methods. Genest et. al. (1995) introduce a semiparametric method, known as maximum pseudolikelihood estimation (mpl), similar to MLE. The only difference between this method and MLE is that the data must be converted to pseudo observations. The consistency, asymptotic normality of this method is established in their paper. They established that this method is efficient for independent copula. Two nonparametric methods based on the rank of observations are inversion of Kendall's tau (itau) and Spearman's rho (irho). These moment estimations are applied to one parameter copula when it is exchangeable (Kojadinovic & Yan, 2010). Kojadinovic & Yan (2010) found that when $\tau \leq 0.4$, the inversion of Spearman's rho is a good approach for estimating the parameter of Gumbel-Hougaard copula. They compared mpl, irho and itau by looking at their mean square error, for different sample sizes and dependency level. It turns out that for $n = 50$ and $\tau \leq 0.2$ i.e. weak dependency, two methods of moment estimation seem to be better than the estimation based on pseudo likelihood. However, as dependency increases and sample size gets larger, mean square error for mpl will reduce. The estimation based on mpl is more biased than the methods of moment estimators, but the biasedness will decrease as n gets larger. The estimation based on the inversion of Spearman's rho performs well for the Gumbel-Hougaard copula. In general, the estimation based on Kendall's tau is better than the Spearman's rho. Tsukahara (2005) introduced a semiparametric estimator, known as "rank approximate Z-estimator". He also proved the asymptotic normality of this estimator. Through a Monte Carlo simulation, he compared τ-inversion, ρ-inversion, pml, minimum Cramér-von Mises distance, minimum Kolmogorov-Smirnov distance and rank approximate Z-estimators. He concluded that pml has the lowest MSE and Z-estimator has the lowest bias. Vandenhende & Lambert

(2005), showed that it was possible to form a univariate distribution from any Archimedean copula (see Theorem 1) by writing the generating function in linear form. They used least square estimation to evaluate the copula parameters. Brahimi & Necir (2012), evaluated the parameters of copula using method of moments. They particularly focused on Archimedean copulas. Consistency and asymptotic normality of their procedures have been verified. They concluded that their method is practically faster and easier. Qu et al. (2009) used the fact that a multivariate Archimedean copula is the same as survival copula of multivariate L_1-norm symmetric distributions and proposed a method for parameter estimation and model selection concentrating on Archimedean families. They established the consistency of the estimator and applied Radia Information Criteria (RIC) for selecting the well-fitted Archimedean copula. Kim et al. (2007) compared two parametric estimation methods i.e. MLE and IFM with pml. They showed that misspecification of margins had significant impact on parameter estimation when the methods used were IFM and MLE. But pml is robust against margins misspecification, therefore it is preferred to other two methods. However, MLE is the most efficient method and at the same time the most efficient one when dealing with multivariate multi parameter copulas. They advised on using pml method. Haan et al. (2008) calculated the parameter of extreme value copula for censored data using pseudo-sample based on MLE. The asymptotic distribution of this method was established in their paper. They also proposed a test statistic for selection of the appropriate copula and found its critical values using bootstrap methods.

This study uses the concept of the generalized linear model (GLM) along with the least square method and introduces a new estimation method, namely GLM-method, for estimating parameter(s) of a given copula. Performance of such GLM-method is compared with performance of other estimation methods through a simulation study. Cramér-von Mises distance is employed as a criteria for such comparison. This article is organized as follows: Section 2 provides a brief review on parameter estimation methods and considers copula quantile regression. GLM-method introduces in Section 3. Section 4 compares different methods of calculating copulas' parameters through a simulation study, and Section 5 concludes.

Methods of parameter estimation

This section reviews some most applicable methods used for parameter estimation of a given copula. As explained above although MLE is misspecified by the margins, it is the most efficient method, which can be implemented by

$$L(\mathbf{\Theta}) = \sum_{t=1}^{T} \ln \mathbf{c}(F_1(x_{1t}), F_2(x_{2t}), \ldots, F_d(x_{dt})) + \sum_{t=1}^{T}\sum_{j=1}^{d} \ln f_j(x_{jt}),$$

where $\mathbf{\Theta}$ is the vector of all parameters of both the marginals and the copula, and c is $\frac{\partial C_\theta(F_1(t_1),\ldots,F_d(t_d))}{\partial F_1(t_1)\ldots\partial F_d(t_d)}$. Another parametric method is IFM, where first the parameter of the margins are estimated and then the parameters of the copula, namely

$$L(\alpha, \theta_1, \ldots, \theta_n) = \sum_{i=1}^{m} \log f(x_i; \theta_1, \ldots, \theta_n, \alpha).$$

The point estimation methods i.e. itau and irho are two nonparametric methods, which are applied given the relationship between the parameter of the copula and the Kendall's tau and Spearman's rho association measures viz. $\theta = g^{-1}(\tau)$. Such relationship is presented in Table 1 .

Table 1: Two dimensional Copulas, the relations between parameters and Kendall's tau, τ.

Copula	Functional form	Parameter
Gaussian	$C_R^{Ga}(u_1, u_2) = \Phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$	$R = \sin(\frac{\pi(\tau)}{2})$
t (df= ν)	$C_R^{t}(u_1, u_2) = t_{v,R}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2))$	$R = \sin(\frac{\pi(\tau)}{2})$
Gumbel	$C_\theta^{Gumbel}(u_1, u_2) = \exp(-[(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{\frac{1}{\theta}})$	$\theta = \frac{1}{1-\tau}$
Clayton	$C_\theta^{Clayton}(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{\frac{-1}{\theta}}$	$\theta = \frac{2\tau}{1-\tau}$

The nonparametric version of MLE, which seems to perform better than other methods when the sample size and level of dependence increases is pml, in which the data must be converted to pseudo observations. Suppose $X_i = (X_{i,1}, \dots, X_{i,p})$ are random vectors and R_{ij} the rank of X_{ij}. Then by applying $\hat{U}_{ij} = \frac{R_{ij}}{n+1}$ the pseudo observations can be calculated. Therefore, the pseudo log-likelihood is given by

$$\log L(\theta) = \sum_{i=1}^{n} \log c_\theta(\hat{\mathbf{u}}_\mathbf{i}),$$

where $\hat{\mathbf{u}}_\mathbf{i} = (\hat{U_{i,1}}, \dots, \hat{U_{i,p}})$. For a bivariate copula this simplifies to

$$L(\theta) = \sum_{i=1}^{n} \log\{c_\theta(\frac{R_i}{n+1}, \frac{S_i}{n+1})\}.$$

Copula-quantile regression

When the distribution function of the variables is not normal, the conditional expectation $E(Y|X)$ and conditional variation $Var(Y|X)$ does not suffice to give full information on the conditional distribution function. In such cases, the quantile regression is used. To estimate the parameters of the quantile regression LSE can be applied, however as opposed to standard mean regression, the loss function is not square error, instead it is the absolute error loss function, hence the sign of the error terms is important (Alexander, § 7, 2008). Copula-quantile regression, which is introduced by Bouyé & Salmon (2009) is a nonlinear form of quantile regression. To apply quantile regression, one needs to know the conditional copula distribution which is given by $C_{U|V}(u|v) = \frac{\partial}{\partial v}C(u,v)$, $C_{V|U}(v|u) = \frac{\partial}{\partial u}C(u,v)$. The copula-quantile regression is defined as the following.

Definition 1. *If $C(.,.,\theta)$ is a parametric copula with parameter θ, the p^{th} quantile curve of v conditional on u is defined by*

$$p = \frac{\partial C(u,v;\theta)}{\partial u},$$

and rearranging with respect to v the copula-quantile regression is given by

$$v = r(u,p;\theta).$$

They studied properties of copula-quantile regression and showed its application on the interdependency between foreign exchange markets, which are skewed. They also compared their findings with tail dependency measure at different α. They noted that tail dependency at different α is not corresponding to copula regression at different quantiles and the results are different. They expressed that their results were more reliable. The following lemma utilized from Bouyé & Salmon (2009)'s finding and established the expression of quantile regression for 5 most applicable copulas, i.e., Clayton, Frank, Gumbel, Normal, and t copulas. (Bouyé & Salmon, 2009) The copula-quantile regression for different class of copulas is given by

(i) Clayton copula $v = ((p^{-\theta/(1+\theta)} - 1)u^{-\theta} + 1)^{-1/\theta}$;

(ii) Frank copula $v = \frac{-1}{\theta}\ln(1 - (1 - e^{-\theta})(1 + e^{-\theta u}(p^{-1} - 1))^{-1})$;

(iii) Normal copula $v = \Phi(\rho\Phi^{-1}(u) + \sqrt{1-\rho^2}\Phi^{-1}(p))$;

(iv) t-copula $v = t_\nu(\rho t_\nu^{-1}(u) + \sqrt{(1-\rho^2)(\nu+1)^{-1}(\nu + t_\nu^{-1}(u)^2)}t_{\nu+1}^{-1}(p))$.

To employ the above lemma, one has to set $u := F_X(x)$ and $v := F_Y(y)$. It should be worthwhile to mention that the Gumbel copula does not have a closed form for the copula-quantile regression method. Therefore, its copula-quantile regression has to be found numerically.

GLM approach to parameter estimation

The idea of using the GLM method for estimating parameter(s) of a given copula has been suggested roughly by several authors, see Genest (1987) and Frees & Valdez (1998) for Frank copula; Parsa & Klugman (2011) for Gaussian copula.

The following theorem uses the GLM method along with the least square method and provides a particle algorithm to estimate copula's parameter(s). Suppose $C(.,.,\theta)$ is a bivariate copula function with parameter θ. Then, based upon continuous random sample $(X_1,Y_1),(X_2,Y_2),\cdots,(X_n,Y_n)$, copula parameter θ can be estimated by minimizing the following least square function in θ.

$$Q(\theta)=\sum_{i=1}^{n}\left(V_i-1+\int_0^1\int_0^k c_\theta(U_i,t)dtdk\right)^2, \tag{1}$$

where $V_i:=F_Y(Y_i)$ and $U_i:=F_X(X_i)$. **Proof.** The GLM expression on two uniform random variables V and U can be stated as

$$\begin{aligned} E(V|U=u) &= \int_0^\infty(1-F_{V|U}(v|u))dv+\int_{-\infty}^0 F_{V|U}(k|u)dk \\ &= 1-\int_0^1 F_{V|U}(k|u)dk \\ &= 1-\int_0^1\int_0^k \frac{f_{U,V}(u,t)}{f_U(u)}dtdk \\ &= 1-\frac{1}{c_{\theta_0}(u)}\int_0^1\int_0^k c_\theta(u,t)f_V(t)dtdk \\ &= 1-\int_0^1\int_0^k c_\theta(u,t)dtdk. \end{aligned}$$

The second and last equations arrived from the fact that V and U are two uniform random variables and the forth equation arrives from identity $f(t_1,\ldots,t_d)=c_\theta(F_1(t_1),\ldots,F_d(t_d))\prod_{i=1}^d f_i(t_i)$.

Now the least square expression can be written as Eqution 1. □

In most cases the above theorem has to be employed numerically. The following corollary explores a situation that the GLM-method has explicit solution. Suppose $c_\theta(\cdot,\cdot)$ is a Farlie-Gumbel-Morgenstern copula function. Then, using random sample $(X_1,Y_1),(X_2,Y_2),\cdots,(X_n,Y_n)$, copula parameter θ is

$$\theta = \frac{3\sum_{i=1}^n(y_i-\frac{1}{2})(x_i-\frac{1}{2})}{\sum_{i=1}^n(x_i-\frac{1}{2})^2}.$$

Proof. The PDF of Farlie-Gumbel-Morgenstern copula is

$$c_\theta(u,v) = 1+\theta(1-2u)(1-2v).$$

Substituting $c_\theta(u,v)$ is least square function given by Theorem 1 leads to

$$Q(\theta)=\sum_{i=1}^{n}(y_i-1+\frac{1}{2}+\frac{1}{6}\theta(1-2x_i))^2.$$

The desire proof arrives after solving $\frac{\partial}{\partial\theta}Q(\theta)=0$ in θ. □

The copula's PDF, $c_\theta(\cdot,\cdot)$, plays a crucial role in calculation of

$$E(V|U=u)=1-\int_0^1\int_0^k c_\theta(u,t)dtdk,$$

in Theorem 1. Table 2 provides the copula's PDF, $c_\theta(\cdot,\cdot)$ for some well known class of copulas.

Table 2: CDF and PDF of two dimensional Copulas.

Copula	CDF	PDF
Gaussian	$\Phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$	$\frac{1}{\sqrt{1-\theta^2}} \exp\left\{-\frac{\theta^2\xi_1^2 - 2\theta\xi_1\xi_2 + \theta^2\xi_2^2}{2(1-\theta^2)}\right\}$
t (df= ν)	$t_{v,R}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2))$	$\frac{1}{\sqrt{\theta}} \frac{\Gamma(\nu/2+1)\Gamma(\nu/2)}{\Gamma(\nu/2+1/2)^2} \left[\frac{\nu(1-\theta^2)(1+\varsigma_1^2/\nu)(1+\varsigma_2^2/\nu)}{\nu(1-\theta^2)+\varsigma_1^2+\varsigma_2^2-2\theta\varsigma_1\varsigma_2}\right]^{(\nu/2+1)}$
Gumbel	$\exp\left\{-[(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{\frac{1}{\theta}}\right\}$	$(A+\theta-1)A^{1-2\theta}\exp(-A)(xt)^{-1}$
Clayton	$(u_1^{-\theta} + u_2^{-\theta} - 1)^{\frac{-1}{\theta}}$	$\frac{1+\theta}{(xt)^{\theta+1}}(x^{-\theta} + t^{-\theta} - 1)^{-2-1/\theta}$
Frank	$\frac{-1}{\theta}\ln(1 + \frac{(\exp(-\theta u_1)-1)(\exp(-\theta u_2)-1)}{\exp(-\theta)-1})$	$\frac{\exp\{-\theta(x+t)\}(1-\exp\{-\theta\})}{(\exp\{\theta(x+t)\}-\exp\{-\theta x\}-\exp\{-\theta t\}+\exp\{-\theta\})^2}$

where $A = ((-\ln x)^\theta + (-\ln t)^\theta)^{1/\theta}$, $\xi_1 = \Phi^{-1}(x)$ and $\xi_2 = \Phi^{-1}(t)$ are the quantiles of standard normal distribution, and $\varsigma_1 = t_\nu^{-1}(x)$ and $\varsigma_2 = t_\nu^{-1}(t)$ are the quantiles of student distribution with ν degree of freedom.

2 Simulation study

Now a simulation study compares these five estimation methods numerically. For this propose five different distributions normal, t, Cauchy, and two extreme value distributions, logistic and Hüsler-Reiss distributions are selected. The first three distributions are symmetric and appropriate for financial data and the last two are appropriate distributions for insurance and reinsurance portfolios, which involve losses with high severity and low frequency. The distribution of insurance claims data are normally skewed and heavy-tailed (Embrechts et. el., 2002, and Kotz & Nadarajah, 2000). The data are simulated from these distributions with high dependence level at 0.9. All data are transformed into pseudo-observations. First, they are ranked and then multiplied by $\frac{1}{n+1}$ to avoid problem that may arise at boundary $[0,1]^d$. Five copulas, which are mostly used, viz, Gumbel, Frank, Clayton, normal and t copulas are selected. The parameters are estimated with 5 approaches: (**i**) Inversion of Kendal's tau; (**ii**) Inversion of Spearman's rho; (**iii**) pml; (**iv**) Copula-quantile regression with ($q = 0.25, 0.50, 0.75$), and (**v**) GLM-method. It is expected that each method is suitable for a particular distribution. To compare the results, the cramér-von Mises distance is used. It is defined by

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(F_{XY}(x,y) - C_\theta(F_X(x), F_Y(y)))^2 f_{XY}(x,y)dxdy.$$

This criteria measures how far is the distance of the real data from the data generated by copula; in other words, this is the weighted average of the errors. Normally the lower this distance, the better would be the results. The simulation is carried out on sample size 1000 with 100 iterations using the packages *copula*, *fCopulae*, *quantreg*, and *evd* in R. The calculation for elliptical copulas and distributions take longer than others. The results are presented in the tables below.

Table 2: The mean and (sd) for parameter (1st row) and Cramér distance (2nd row) based on different methods and different copulas

Logistic	Clayton	Frank	Gumbel	Normal	t (df=4
itau	0.2357896 (0.04758373)	0.9549693 (0.1757497)	1.107067 (0.0292168)	0.1642097 (0.02938949)	0.1642097 (0.0
	0.0003710279 (0.0003495556)	0.0002174217 (0.0002381969)	76638e-05 (0.0001018399)	0.579039 (0.02567358)	0.6560563 (0.0
irho	0.2338561 (0.04714266)	0.9471112 (0.1742059)	1.107806 (0.02868843)	0.1629714 (0.02918711)	0.1629714 (0.0
	0.0003569152 (0.0003406244)	0.0002070756 (0.0002300933)	7.428632e-05 (0.0001018399)	0.5783464 (0.02557629)	0.6553711 (0.0
mpl	0.1379168 (0.03819629)	0.96518 (0.1761349)	1.111834 (0.02539398)	0.170902 (0.02850896)	0.1510195 (0.0
	8.43965e-05 (0.0001237738)	0.0002299379 (0.0002474306)	5.685229e-05 (8.676336e-05)	0.5827182 (0.02535498)	0.64889 (0.02
q(0.25)	3.26151 (0.1628287)	0.9533234 (0.2298271)	1.048115 (0.01800756)	0.152806 (0.04361077)	0.1826987 (0.0
	0.07143315 (0.002748488)	0.0002694946 (0.0002294357)	0.0004259352 (0.0002452429)	0.5737317 (0.02902004)	0.6684247 (0.0
q(0.50)	5.496151 (0.4657097)	0.8908815 (0.1989322)	1.050184 (0.01465274)	0.1688641 (0.03974894)	0.1366404 (0.0
	0.09836826 (0.003908395)	0.0001733166 (0.0002426908)	7.76638e-05 (0.0001830703)	0.5822645 (0.02910009)	0.6419208 (0.0
q(0.75)	3.848151 (0.2425259)	1.053014 (0.1889862)	1.052381 (0.01401862)	0.1777977 (0.03607364)	0.1978619 (0.0
	0.08043742 (0.003343596)	0.0003670748 (0.0003141369)	0.000358005 (0.0001654764)	0.5872107 (0.02965161)	0.6761877 (0.0
GLM	0.3045018 (0.06306191)	1.072397 (0.1985979)	0.3806179 (0.3222072)	0.2045057 (0.03607838)	0.2167632 (0.00
	0.001039189 (0.0007416712)	0.0004081506 (0.0003793207)	0.08743481 (0.05128618)	0.603464 (0.03105044)	0.6866389 (0.0

Table 3: The mean and (sd) for parameter(1st row) and Cramér distance (2nd row) based on different methods and different copulas

Hüsler-Reiss	Clayton	Frank	Gumbel	Normal	t
itau	0.5361688 (0.06593831)	1.971938 (0.2056333)	1.261172 (0.03080665)	0.3250657 (0.03023427)	0.3250657
	0.001006468 (0.0006577236)	0.0005220021 (0.0004315309)	6.052187e-05 (7.3808e-05)	0.6646245 (0.03571358)	0.7410599
irho	0.5341152 (0.0658127)	1.955466 (0.2013865)	1.260534 (0.0310642)	0.3231768 (0.02983328)	0.3231768
	0.0009869564 (0.0006519294)	0.0004873584 (0.0004087076)	6.145837e-05 (7.342551e-05)	0.6627463 (0.03520841)	0.7392119
mpl	0.3179717 (0.04836689)	1.988077 (0.2064492)	1.263038 (0.02907468)	0.3409704 (0.02758792)	0.3005748
	0.0001601931 (0.000182298)	0.0005542253 (0.000445662)	5.429113e-05 (7.029836e-05)	0.6803328 (0.03574465)	0.7184925
q(0.25)	3.915307 (0.2113648)	1.853813 (0.2246314)	1.111556 (0.01409183)	0.34047 (0.0410683)	0.3738092
	0.06286956 (0.002568584)	0.0003443664 (0.0003186477)	0.001743771 (0.000365299)	0.681521 (0.045995)	0.7928539
q(0.50)	7.882373 (1.071549)	1.876001 (0.2573724)	1.109802 (0.0117537)	0.3328085 (0.0391115)	0.2827441
	0.09182991 (0.004510189)	0.0004097881 (0.0004342463)	0.001782287 (0.0003082017)	0.6734424 (0.04400504)	0.7040377
q(0.75)	5.14974 (0.423687)	1.907388 (0.2081813)	1.108183 (0.01247685)	0.3232863 (0.03520504)	0.3217019
	0.07545479 (0.003635486)	0.0004090452 (0.000394444)	0.001826502 (0.0003263631)	0.6635333 (0.04039317)	0.7380088
GLM	0.5873128 (0.06814438)	2.220233 (0.2283099)	0.2457962 (0.2544769)	0.3981225 (0.03141991)	0.2252249
	0.001544791 (0.0008159843)	0.001144286 (0.0007005859)	0.154355 (0.05165686)	0.7472655 (0.04667404)	0.6595661

Table 4: The mean and (sd) for parameter(1st row) and Cramér distance (2nd row) based on different methods and different copulas

Cauchy	Clayton	Frank	Gumbel	Normal	t (df=
itau	5.020746 (0.2754016)	12.13805 (0.5648465)	3.485702 (0.1372955)	0.9011049 (0.007734868)	0.9011049 (0.0
	0.0008905475 (0.0003974337)	0.002003717 (0.0006216668)	0.0001228918 (8.643768e-05)	3.04873 (0.0004111389)	3.046683 (0.0
irho	4.96374 (0.3149087)	11.19294 (0.5681407)	3.419493 (0.1543944)	0.8920558 (0.009278915)	0.8920558 (0.0
	0.0008305017 (0.0004305138)	0.001076423 (0.0004769195)	0.0001604845 (0.0001279575)	3.0489 (0.0004261115)	3.046747 (0.00
mpl	3.256898 (0.1943582)	12.03472 (0.5399194)	3.32288 (0.1304416)	0.897043 (0.008644459)	0.9010679 (0.0
	0.003289592 (0.001265076)	0.001887763 (0.0005835376)	0.0002242443 (0.0001916005)	3.048807 (0.0004254471)	3.046684 (0.00
q(0.25)	18.73929 (1.299287)	12.97825 (0.867226)	1.717903 (0.03354662)	0.9201327 (0.008369305)	0.9016264 (0.0
	0.02472225 (0.001173588)	0.00301631 (0.001054648)	0.04919603 (0.003807738)	3.048274 (0.000420709)	3.046669 (0.00
q(0.50)	109.9589 (34.32165)	23.89952 (1.615271)	1.689724 (0.03003661)	0.9166894 (0.01178864)	0.900092 (0.0
	0.04046081 (0.001047012)	0.01560137 (0.001461331)	0.05247478 (0.003606004)	3.048358 (0.0004698031)	3.046649 (0.00
q(0.75)	15.21881 (1.243137)	12.92436 (0.8186823)	1.661964 (0.03243775)	0.9189895 (0.008637368)	0.9003902 (0.0
	0.021036 (0.001544408)	0.002943198 (0.001005811)	0.05597575 (0.004149634)	3.048308 (0.0004267481)	3.046681 (0.00
GLM	2.864082 (0.1523376)	10.38139 (0.3865085)	0.02927618 (0.003790814)	0.8162787 (0.006318219)	0.3148228 (0.0
	0.006577563 (0.001695343)	0.000494686 (0.0001873805)	1.767603 (0)	3.049181 (0.0005945993)	0.5490803 (0.0

Table 5: The mean and (sd) for parameter(1st row) and Cramér distance (2nd row) based on different methods and different copulas

Normal	Clayton	Frank	Gumbel	Normal	t (df
itau	4.974676 (0.2418697)	12.04375 (0.4959966)	3.488032 (0.1048032)	0.8999215 (0.006750667)	0.8999215 (0
	0.0006288804 (0.0003868215)	0.001848868 (0.0005852485)	5.066503e-05 (6.522701e-05)	2.903314 (0)	2.9033
irho	5.232967 (0.2797341)	11.67597 (0.4966714)	3.576344 (0.1171607)	0.8996185 (0.007201842)	0.8996185 (0
	0.001083041 (0.0003868215)	0.001437082 (0.0005418999)	0.000116134 (0.0001220085)	2.903314 (0)	2.9033
mpl	2.990768 (0.1524266)	11.70848 (0.4915929)	3.182947 (0.09453462)	0.8998355 (0.006135488)	0.8838732 (0
	0.004750293 (0.001340524)	0.001471376 (0.0005422346)	0.0003522941 (0.0002747533)	2.903314 (0)	2.9033
q(0.25)	16.6853 (1.209538)	11.36454 (0.7780717)	1.666286 (0.02518844)	0.8999119 (0.01026184)	0.8795 (0.0
	0.02311366 (0.001278929)	0.001165907 (0.0008030268)	0.05413449 (0.003234027)	2.903314 (0)	2.9033
q(0.50)	90.16834 (24.8816)	21.43149 (2.06973)	1.636694 (0.0232912)	0.8995011 (0.01416515)	0.8829991 (0
	0.04011715 (0.00137752)	0.01351139 (0.002229634)	0.05799676 (0.003202957)	2.903314 (0)	2.9033
q(0.75)	13.77557 (0.901369)	11.31259 (0.6306088)	1.607161 (0.0243272)	0.9003276 (0.008839137)	0.8802001 (0
	0.01950232 (0.001298968)	0.001092559 (0.0005934812)	0.06215515 (0.003579388)	2.903314 (0)	2.9033
GLM	2.986528 (0.1503335)	10.69878 (0.3468778)	0.02785433 (0.001735336)	0.821169 (0.005246515)	0.317729 (0.
	0.004786553 (0.001359574)	0.0005129631 (0.0002542829)	1.645124 (4.176584e-09)	2.903314 (0)	0.49552 (0.0

Table 6: The mean and (sd) for parameter(1st row) and Cramér distance (2nd row) based on different methods and different copulas

t	Clayton	Frank	Gumbel	Normal	t (df=
itau	5.020746 (0.2754016)	12.13805 (0.5648465)	3.485702 (0.1372955)	0.9011049 (0.007734868)	0.9011049 (0.0
	0.0007321414 (0.0004329405)	0.001988042 (0.0006621676)	8.070655e-05 (9.936043e-05)	2.985174 (1.287815e-05)	2.98512 (1.51
irho	4.96374 (0.3149087)	11.19294 (0.5681407)	3.419493 (0.1543944)	0.8920558 (0.009278915)	0.8920558 (0.0
	0.0006641577 (0.0004709457)	0.000987452 (0.0005239334)	0.0001006437 (0.0001125403)	2.985177 (1.261057e-05)	2.985122 (1.46
mpl	3.256898 (0.1943582)	12.03472 (0.5399194)	3.32288 (0.1304416)	0.897043 (0.008644459)	0.9010679 (0.0
	0.002828245 (0.001221514)	0.0018648 (0.0006228675)	0.0001384507 (0.0001611368)	2.985175 (1.282694e-05)	2.98512 (1.51
q(0.25)	18.73929 (1.299287)	12.97825 (0.867226)	1.717903 (0.03354662)	0.9201327 (0.008369305)	0.9016264 (0.0
	0.02514491 (0.001188977)	0.003054616 (0.001108994)	0.04806146 (0.003760178)	2.985167 (1.470355e-05)	2.98512 (1.49
q(0.50)	109.9589 (34.32165)	23.89952 (1.615271)	1.689724 (0.03003661)	0.9166894 (0.01178864)	0.900092 (0.0
	0.04099707 (0.001048899)	0.01598359 (0.001485625)	0.05129966 (0.003561991)	2.985168 (1.501052e-05)	2.985119 (1.59
q(0.75)	15.21881 (1.243137)	12.92436 (0.8186823)	1.661964 (0.03243775)	0.9189895 (0.008637368)	0.9003902 (0.0
	0.02140661 (0.00156805)	0.002978578 (0.001056577)	0.05475874 (0.004100294)	2.985167 (1.494597e-05)	2.98512 (1.53
GLM	2.864082 (0.1523376)	10.38139 (0.3865085)	0.02927618 (0.003790814)	0.8162787 (0.006318219)	0.3148228 (0.0
	0.006020398 (0.00165502)	0.0003340173 (0.000222922)	1.72388 (9.7817e-09)	2.985172 (1.154915e-05)	0.5272869 (0.0

The tables present the mean and standard deviation for parameters of different copulas and their corresponding Cramér distance. The first table is based on the data generated from Logistic distribution, which is an extreme value distribution function. For this distribution, when the copulas are Clayton and Gumbel, pml gives better result. For other copulas, quantile regression method dominates other approaches. Copula-quantile regression at $50\%, 25\%$ and 50% performs better than other methods when the underlying copulas are Frank, Normal and t (df=4), respectively. However, the results obtained by different methods is not significantly different when copulas are elliptical. The second table shows the results for Hüsler-Reiss distribution, which is another extreme value distribution function under study. Similar to Logistic distribution, the performance of pml is better than other approaches when the copulas are Clayton and Gumbel. Moreover, when the copula is Frank, quantile regression at 25% gives better result. Although not significantly different, the Cramér distance given by irho is lower than copula-quantile regression at 75% when the copula is normal. GLM performs well when copula is t. Three other distributions from which the data are simulated are elliptical. The third table belongs to the data generated from Cauchy distribution, which is also a heavy-tailed distribution function. Inverse of Spearman's rho with Clayton copula does better than other estimation methods. Inverse of Kendall's tau performs well when the copula is Gumbel. GLM supersedes other approaches when the copulas are Frank and t. The Cramér distance obtained by different methods is roughly the same when the copula is normal. For two other elliptical distributions i.e. normal and t distributions, the performance of the estimation methods with the corresponding copulas is the same. The only difference is that for normally distributed data, itau performs better than irho when the copula is Clayton. In general, the copula-quantile regression is a better approach when the data are heavy-tailed and the GLM gives rise to better result when the data are elliptically distributed. However, with the latter the performance is the worst when copula is Gumbel regardless of the type of data.Finally, one can conclude that the estimation methods are more dependent on the type of data than the chosen copula. Therefore, it is suggested to check the data and their distributions and then pick a method that is supposed to perform better than other methods.

3 Conclusion

Copulas' parameter estimation is the first step in copula modeling. In this study, for the first time, to the best of the authors' knowledge, the copula-quantile regression is included in a simulation study that compares different approaches for parameter evaluation. As is expected such method is only a good approach when data are from extreme value distributions. Also a new method which is based on GLM is proposed for estimating the parameters of copulas. The results indicate that GLM performs better than other methods when data have an elliptical distribution. Therefore, it is suggested to check the distribution of data before selecting a particular estimation method. An extensive simulation study that contains more distributions , say, heavy-tailed, mixed distributions and more copulas must be carried out to assert this suggestion more firmly. Moreover, based on the findings of the simulation study, in spite of the performance of copula quantile regression and the GLM method which is relatively better than others for different distributions, the pml performs relatively well in all situations. This approach turn out to be the preferred method according to Kojadinovic & Yan (2010), Tsukahara (2005) and Kim et. el. (2007).

References

[1] Alexander, C. (2008). *Market risk anlysis* Volume II Practical Financial Econometrics. John Wiley & Sons, England.

[2] Bouyé E. & Salmon M. (2009). Dynamic copula quantile regressions and tail area dynamic dependence in Forex markets. *European Journal of Finance.* **15**, 721–750.

[3] Brahimi B. & Necir A. (2012). A semiparametric estimation of copula models based on the method of moments, *Statistical Methodology.* **9**, 467–477.

[4] Cherubini U., Luciano E., Vecchiato W. (2004).*Copula methods in finance* . Wily, England

[5] Embrechts P., McNeil A., & Straumann D. (2002). Correlation and dependency in risk management: Properties and pitfalls, in: M. Dempster, ed., Risk Management: Value at Risk and Beyond (Cambridge: Cambridge University Press), 176-223.

[6] Frees E. W. & Valdez E. A. (1998). Understanding relationships using copulas, *North American Actuarial Journal*, **1**, 1–25.

[7] Genest C. (1987). Frank's family of bivariate distributions, *Biometrika*, **74**, 549–555.

[8] Genest C. & Favre A.C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, **12**, 347-368.

[9] Genest C., Ghoudi K.& River L.P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, **82**, 543–552.

[10] Haan L., Neves C. & Peng L. (2008). Parametric tail copula estimation and model testing, *Journal of Multivariate Analysis*, **99**, 1260–1275.

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

[12] Kim G., Silvapulle M. & Silvapulle P. (2007). Comparison of semiparametric and parametric methods for estimating copulas, *Computational Statistics & Data Analysis*, **51**, 2836–2850.

[13] Kojadinovic I. & Yan J. (2010). Comparison of three semiparametric methods for estimating dependence parameters in copula models. *Insurance: Mathematics and Economics*, **47**,52–63.

[14] Kotz S. & Nadarajah S. (2000). *Extreme value distributions: theory and applications*. Imperial College press.

[15] Parsa R. & Klugman S.A. (2011). Copula regression, *Variance Advancing and Science of Risk*, **5**, 45–54.

[16] Qu X., Zhou J. & Shen X. (2009). Archimedean copula estimation and model selection via L1-norm symetric distribution, *Insurance: Mathematics and Economics*, **46**, 406–414.

[17] Tsukahara H.(2005). Semiparametric estimation in copula models, *The Canadian Journal of Statistics*, **33**, 357–375.

[18] Vandenhende F. & Lambert P. (2005). Local dependence estimation using semiparametric Archimedean copulas, *The Canadian Journal of Statistics*, **33**, 377–388.

[19] R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

[20] Diethelm Wuertz, many others and see the SOURCE file (2009). fCopulae: Rmetrics - Dependence Structures with Copulas. R package version 2110.78. http://CRAN.R-project.org/package=fCopulae

[21] Roger Koenker ¡rkoenker@uiuc.edu¿ (2011). quantreg: Quantile Regression. R package version 4.71. http://CRAN.R-project.org/package=quantreg

Pair-Copula constructions for a Multivariate Distribution:Parametric and Non-parametric methods

Poorahmadi, M. and Daneshkhah, A.

of Statistics, Faculty of Mathematical Sciences Shahid Chamran University, Ahvaz,

Abstract

This paper focuses on the two methodologies, one of them is a new method proposed by Bedford and Daneshkhah [3] using the Vines as a way of constructing higher dimensional distributions. Using this method, one can approximate any multivariate density as closely as one likes based on the approximated vines. Each pair-copula which is used in constructing the vine is replaced by the minimum information copula ([4], [3]) in conjunction with the information provided from the constraints (on moments, rank correlation, expert elicitation of observables) to define a copula that represents the joint distribution of two random variables of interests.

In the second method, the parametric copula is fitted to each pair-copula considered in the vine construction. The main difficulty in this method is to select the suitable pair-copula families. We use the different tools including contour plots, λ- function and Goodness-of-Fit (GOF) tests, to find the appropriate distributions in the D-vine decomposition.

We eventually apply these methods to approximate a joint distribution of the three variables associated with the Iran's financial data: gross domestic production, oil income index, inflation index.

Keywords: entropy, Lambda plot, minimum information copula, pair-copula, vine.

1 Introduction

There is a growing literature on the use of copulas to model dependencies (see e.g. Kurowicka 2006; Bedford et al, 2012 and reference therein). There are vast applications of Copulas in the different areas of operations research including combining expert opinion and stochastic simulation. A copula is a joint distribution on the unit square (or more generally on the unit n-cube) with uniform marginal distributions. Under reasonable conditions, a joint distribution for n random variables can be uniquely specify by determining the univariate distribution for each variable, and in addition, specifying the copula. The reason is that each variable can be simply transformed by its own distribution function to make sure that the derived variable is uniformly distributed.

Bedford et al (2012) illustrated that the use of a copula to model dependency is simply and clearly a translation of one difficult problem into another: instead of the difficulty of specifying the full joint distribution we have the difficulty of specifying the copula. As many authors restrict the copulas to a particular parametric class (Gaussian, multivariate t, etc) the potential flexibility of the copula approach is not realized in practice. In addition, it is clear that specifying a multivariate copulas for the high-dimensional (larger than 2) data-set in comparison to the class of bivariate copulas is much harder. Bedford and Cooke (2002) developed a graphical model, called *regular vines* or **pair-copula**, to decompose high-dimensional copula into a group of bivariate copulas. As a results, a multivariate density can be decomposed into a cascade of bivariate copulae. We briefly discuss the pair-copula construction in the next section.

We follow Bedford et al [3], using any vine structure, to approximate any given multivariate copula to any required degree of approximation. The technical assumption is required here is that the multivariate density f under of interests with the uniform marginals should be continuous and non-zero.

This approximation approach is based on the *minimum information copulas* ([4]) that can be specified to any required degree of precision based on the data available. It is proved that good approximation 'locally' guarantees good approximation globally. The advantage of this method is that a vine structure imposes no restrictions on the underlying joint probability distribution it represents (as opposed to the situation for Bayesian networks, see Bedford et al, 2012).

Our main contribution to this paper is applying this method to approximate a joint distribution of Iran financial variables including the Investment index, the petroleum export and the Gross Domestic Production index by using some constraints on the moments and rank correlation which created a new base to approximate the joint density.

The rest of the paper is organized as follows. In Section 2, we first briefly introduce the pair-copulae models. In Section 3, we first explain the D_1AD_2 algorithm and the minimum information copulae, we then extend this idea to build a minimally informative pair-copular given the constraints provided from either the expert or data, and from there we can approximate the corresponding multivariate distribution associated with the aforementioned variables of interest to any required degree of precision. In Section 4, we first fit a parametric distribution to each 2-dimensional copula constructing the vine using some well-known tools and then obtain the multivariate distribution associated with these variables, we then using the non-parametric approach mentioned in Section 3, based on some constraints calculated from the Iran financial data, a minimally informative copula for each pair-copula and the corresponding joint density function D-vine are approximated.

2 Pair-copulae constructions of multiple dependencies

Consider n random variables $\mathbf{X} = (X_1, \ldots, X_n)$ with a joint density function $f(x_1, \ldots, x_n)$. This density can be factorised as

$$f(x_1, \ldots, x_n) = f(x_n)f(x_{n-1}|x_n)f(x_{n-2}|x_{n-1}, x_n) \ldots f(x_1|x_2, \ldots, x_n)$$

and this decomposition is unique up to a relabelling of the variables.

As mentioned above, a copula is multivariate distribution, C, with uniformly distributed marginals $U(0, 1)$ on $[0, 1]$. Sklars theorem states that every multivariate distribution F with marginals $F_1, F_2, \ldots, F_n$ can be written as

$$F(x_1, \ldots, x_n) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)) \tag{1}$$

for some appropriate n-dimensional copula C.

Passing to the joint density function f, for an absolutely continuous F with strictly increasing, continuous marginal densities $F_1, \ldots, F_n$, we have

$$f(x_1, \ldots, x_n) = c_{12\ldots n}(F_1(x_1), \ldots F_n(x_n))f_1(x_1) \ldots f_n(x_n)$$

for some (uniquely identified) n-variate copula density $c_{12\ldots n}(\cdot)$.

Bedford and Cooke [2] proposed a more flexible structure, called *vine* (or *pair-copula* by Aas et al, 2009) which allows for the free specification of (at least) $n(n-1)/2$ copulas between n variables. The modelling scheme is based on a decomposition of a multivariate density into a set of bivariate copulae. Due to lack of space, we explain a vine structure further using the following example.

The density decomposition associated with 3 random variables $\mathbf{X} = (X_1, X_2, X_3)$ with a joint density function $f(x_1, x_2, x_3)$ satisfying a copula-vine structure (called D-vine, see [6], pp. 93) with the marginal densities f_1, f_2, f_3 is

$$f_{123}(x_1, x_2, x_3) = \prod_{i=1}^{3} f(x_i)c_{12}\left(F(x_1), F(x_2)\right)c_{23}\left(F(x_2), F(x_3)\right)c_{13|2}\left(F(x_1 \mid x_2), F(x_3 \mid x_2)\right) \tag{2}$$

The above factorisation gives us a constructive approach to build a multivariate distribution given a vine structure: If we make choices of marginal densities and copulae then the above formula will give us a multivariate density. Hence vines can be used to model general multivariate densities. However, in practice we have to use copulas from a convenient class, and this class should ideally be one that allows us to approximate any given copula to an arbitrary degree. By having this class of copulas, we then can approximate any multivariate distribution using any vine structure.

3 Constructing bivariate minimum information copulas

Bedford et al [3] present a way to approximate a copula using minimum information methods which demonstrate uniform approximation in the class of copulae used.

Bedford and Meeuwissen ([4]) applied a so-called DAD algorithm to produce discretized minimally informative copula with given rank correlation. This approach can be used whenever we wish to specify the expectation of any symmetric function of $U = F(x)$ and $V = F(y)$. In order to have asymmetric specifications we need to use the D_1AD_2 algorithm where A is a positive square matrix, and we can find diagonal matrices D_1 and D_2 such that the product of D_1AD_2 is doubly stochastic. For the variables of interest X and Y, we would like to correlate them by introducing constraints based on some knowledge about functions of these variables. Suppose there are k of these functions, namely $h'_1(X,Y), h'_2(X,Y), \ldots, h'_k(X,Y)$, and we specify mean values $\alpha_1, \ldots, \alpha_k$ for all these functions respectively from the data or the expert judgment. We can find corresponding functions of the copula variables U and V, defined by $h_1(U,V) = h'_1(F_1^{-1}(U); F_2^{-1}(V))$, etc., and clearly these should also have the specified expectation α_1. We form the kernel

$$A(u,v) = \exp(\lambda_1 h_1(u,v) + \ldots + \lambda_k h_k(u,v)) \tag{3}$$

where u denote the realization of U and v the realization of V .

For practical implementations we have to discretize the set of (u,v) values such that the whole domain of the copula is covered. This means that the kernel A described above becomes a 2-dimensional matrix A and that we seek the matrices D_1 and D_2 to create a discretized copula density $P = D_1AD_2$.

Suppose that both U and V are discretized into n points, respectively u_i, and v_j, $i,j = 1,\ldots,n$. Then we write $A = (a_{ij}), D_1 = diag(d_1^{(1)},\ldots,d_n^{(1)}), D_2 = diag(d_1^{(2)},\ldots,d_n^{(2)})$, where $a_{ij} = A(u_i,v_j)$, $d_i^{(1)} = D_1(u_i)$, $d_j^{(2)} = D_2(v_j)$. The double stochastically of D_1AD_2 with the extra assumption of uniform marginals means that

$$\forall i = 1,\ldots n \quad \sum_j d_i^{(1)} d_j^{(2)} a_{ij} = 1/n, \ and \quad \forall j = 1,\ldots n \quad \sum_1 d_i^{(1)} d_j^{(2)} a_{ij} = 1/n,$$

since for any given i and j the selected cell size in the unit square is $1/n^2$. Hence

$$d_i^{(1)} = \frac{n}{\sum_j d_j^{(2)} a_{ij}} \quad and \quad d_j^{(2)} = \frac{n}{\sum_i d_i^{(1)} a_{ij}}$$

This iteration converges geometrically to give us the vectors required. The D_1AD_2 algorithm works by fixed point iteration and is closely related to iterative proportional fitting algorithms.

Bedford et al (2012) discussed that, for a given set of functions $(h_1,\ldots,h_k)$, the mapping from the set of vectors of λs parameterizing the kernel A onto the expectations of the function $(\alpha_1,\ldots,\alpha_k)$ has to be found numerically, and optimization techniques are used to achieve this. We wish to determine the appropriate set of λs for given expectations α_i, where the expectations have been calculated using the discrete copula density D_1AD_2. We define

$$L_l(\lambda_1,\ldots,\lambda_k) := \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d^{(1)}(u_i)d^{(2)}(v_j)A(u_i,v_j)h_l(u_i,v_j) - \alpha_l, \quad l = 1,2,\ldots,k. \tag{4}$$

We seek the roots of these functions. One of the possible solvers for this task would be FSOLVE - MATLAB's optimization routine.

It can be shown that a multivariate distribution can be arbitrarily well approximated by using a fixed family of bivariate copulae . A key step to demonstrating this is to show that the family of bivariate (conditional) copula densities contained in a given multivariate distribution forms a compact set in the space of continuous functions on $[0,1]^2$. Based on this we can then show that the same finite parameter family of copulae can be used to give a given level of approximation to all conditional copulae simultaneously (see Bedford et al. (2012) for the proof and further discussion)

We now able to approximate any copula using a linear functions, called bases. We denote $C([0,1]^2)$ as the space of continuous real valued functions on a $[0,1]^2$, this space can be considered as a vector space, and in this context a basis is simply sequence of functions $h_1, h_2, \cdots \in C([0,1]^2)$ for which any function $g \in C([0,1]^2)$ can be written as $g = \sum_{i=1}^{\infty} \lambda_i h_i$. There are lots of possible bases, for example, $u, v, uv, u^2, v^2, u^2vuv^2, \ldots$.

Given an ordered basis $h_1, h_2, \cdots \in C([0,1]^2)$ and a required degree of approximation $\epsilon > 0$ in the sup metric, any density which is positive and uniformly bounded away from 0, can be approximated to within error $\epsilon > 0$ by a linear combination of $h_1, h_2, \ldots, h_k$ (see Bedford et al. (2012) for further details).

4 Application: Iran's Financial data

In this section, we study three time series of seasonal data in Iran for the period from 20.03.1990 to 20.03.2007. For further notations, we denote the variable with observations converted from the three indices as follows: A: Gross Domestic Production; B: oil income index; C: inflation Index. We first make the data independent. Therefore, we can use the transformed standardized residuals ARMA(1,1)-GARCH(1,1) model instead of the original data, to capture the dependency between the variables and approximate their density function. We first fit the parametric distribution taken from the different families for each copula in the vine construction. Note that, the first tree in a D-vine structure is uniquely determined as $C - B - A$. The whole D-vine structure will then be simply determined after the first tree is specified.

4.1 Pair-copula constructions based on the Parametric bivariate copula

After finding the tree for the first level we have to fit a copula density to each bivariate copula in the first level to be able to fit a density to the other levels of the D-vine structure. Therefore, we use three different methods to select an appropriate copula for the underlying data. Two of them are graphical methods, the contour plots and the λ-function. The third one will be a statistic test. But notice, all these methods give only a hint of the true copula family and are not methods to identify the copula family uniquely. With these three methods together we try to reduce the set of possible copula families and end up with a good guess.

The first method is to look at the behavior of the empirical data with respect to their Kendalls τ and density function. The density function can be illustrated in the so called *contour plot*. We derive the empirical density function of the data and plot it in a contour plot. We then compare this density with the theoretical density of possible copula families using their contour plots. We choose the Normal(N), t-student (t), Clayton (C), Gumbel (G), Frank (F), BB1- and BB7-copula to compare with the empirical contour plot. Figure 1 displays the empirical contour plot. Since the empirical contour plot had a

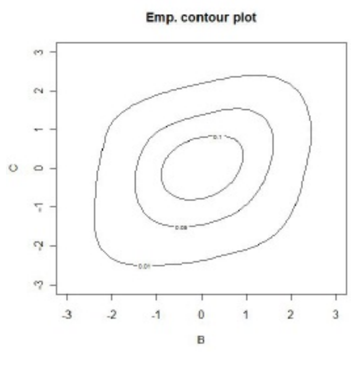

Figure 1: Empirical contour plot

tendency in the lower left and the upper right corner, we can reject the Clayton, Gumbel and Frank copula. We decide that t- , BB1- and BB7- may fit the data well. The Gaussian copula does not have the tail dependency and can not be detected from the empirical contour plot. But from the results of the several tests, it can be concluded that the Gaussian copula may also be a good fit (see Karimi et al, 2012).

As one can see the contour plots of BB1- and BB7- copulas and their flouting intersections cannot be helpful to decide which of these should be accepted, so we need some other methods. Another useful graphical method is the λ-function at which their plots should be a help to identify the underlying copula. Similar to the contour plots, one derives the empirical Kendalls τ and the empirical λ-function for the data. Then, the λ-functions assassinated with the possible copulas are calculated and the resulting plots are compared to each other. Figure 2 shows the plots of empirical λ and 7 corresponding theoretical λ for pair (C,B). In each plot we draw the empirical version (black line), the theoretical version (grey line) and the extreme values of the λ-function (dashed lines).

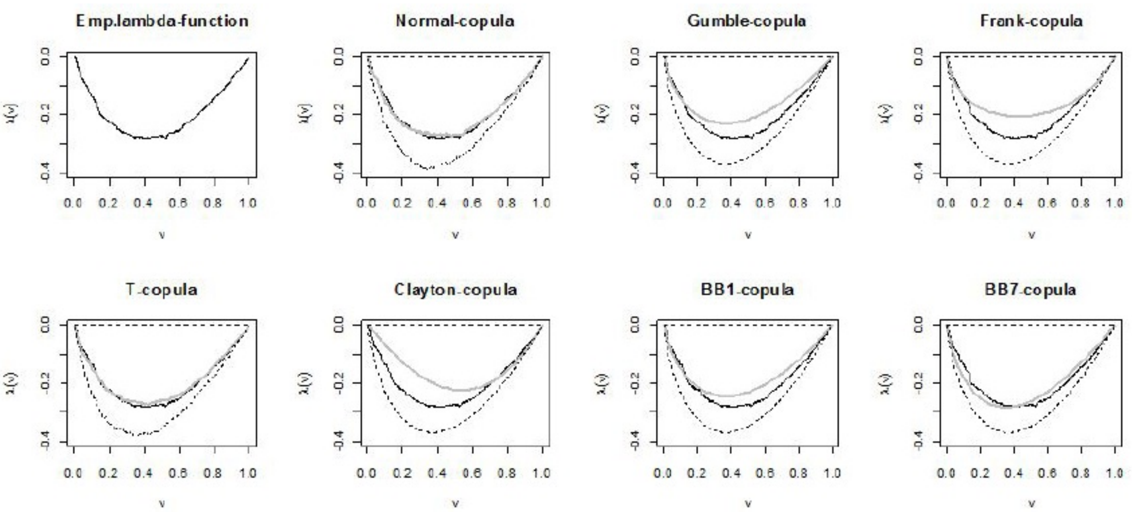

Figure 2: The empirical λ-function of the pair-copula (C,B)(first plot) and the theoretical copula.

We can reject the Clayton, the Gumbel, the Frank and the BB1-copula. The BB7-copula can not be rejected, because it is near the empirical λ function. The λ-function of the Normal and t-copula fit the empirical λ-function well as well, i.e. the grey (theoretical) and black (empirical) lines in the figure overlap.

Beside the two graphical methods explained above, there are some goodness-of-fit tests. One of them is based on the Kendall's transform. These tests are often used in statistical hypothesis testing, in our case we test the hypothesis that if a chosen copula fits the underlying copula of the data well. We calculate the p-values based on the Kendall's transform corresponding to the Carmer von Mises S_n and Kolmogorov Smirnov T_n statistics. We calculate the p-values of the candidate copula for the pair (C, B) at which we can conclude that t-copula has the highest p-value (0.6 for S_n and 0.52 for T_n)s. Thus, from the goodness of fit tests and the graphical tools, we choose the t-copula for the pair (C, B) in the first tree. The same routine will be done for the next copulas. Note that for the nodes C, B and A in the first tree, we used their observations as the underlying data directly, but for the second tree (conditional copula), we first need to compute the conditional distribution function and then use the derived data as the underlying data in each node. Eventually, we complete the D-vine decomposition and determine the most appropriate model for the pair-copula in each step. We conclude that a t- copula is the best choice for pair the (B,A) while a Normal-copula is the best good choice for the copula between $(C \mid B, A \mid B)$ in the second tree. Therefore, the multivariate–three-dimensional given in (2) can be derived by replacing the selected copula in the given decomposition.

4.2 Constructing approximations using minimally informative distributions

We first present the practical guide (given in Bedford et al (2012) in details) to build a minimally informative copulas structure briefly discussed above to approximate any multivariate distribution. A multivariate distribution can be approximated as follows:

- Specify a basic family

Interval	Bases	Lagrange multipliers	Log-like
$0 < B < 0.25$	$(C^3A^2, CA^2, A^2C^4, A^5C, C^2A^5)$	(14.4,35.9,16.6,-26.9,53.6)	9.93
$0.25 < B < 0.5$	$(CA^2, C^3A^2, A^4C^4, A^5C^4, C^3A)$	(0.65,0.8,-3.5,1.4,-0.33)	11.67
$0.5 < B < 0.75$	$(A^3C, C^3A^3, A^2C^2, AC^2, C^2A^4)$	(-12.4,-23.4,98.4,77.7,66.4)	12.43
$0.75 < B < 1$	$(CA^4, C^4A, AC^3, A^5C^2, C^3A^4)$	(5.2,4.3,-4.4,1.5,-6.9)	6.45

Table 1: Bases, parameter values and log-likelihoods for $C_{TB|M}$

- Specify a pair-copula structure
- For each part of pair-copula specify either, (1) mean $\alpha_1, \ldots, \alpha_k$ for $h_1, \ldots, h_k$ on each pairwise copula; (2) functions $\alpha_m(ji \mid D_e)$ for the mean values as functions of the conditioning variables, for $m = 1, \ldots, k$, where D_e is the conditioning set for the edge e.

As an example, we use the Iran's financial data described above to show the construction of a minimally informative copula between two variables A and B denoted by c_{AB} under the following simple polynomials and moving to more complex constraints: $h'_1(A,B) = AB$, $h'_2(A,B) = A^2B$, $h'_3(A,B) = B^2A^2, h'_4(A,B) = A^3B$, $h'_5(A,B) = A^2B^2$, $h'_6(A,B) = B^3A, \ldots$

We propose to assess the log-likelihood of adding each additional basis function to improve the approximation precision. We include the function which produces the largest increase in the log-likelihood (further details can be found in Bedford eta l, 2012). Thus our method is similar to a stepwise regression. We first yield the initial copula in terms of the following basis functions: $A^4B, B^3A^2, A^2B^4, A^5B^2, B^2A^2$.

The number of discretization points (or grid size) is another factor that would effect the approximation. A larger grid size will provide a better approximation to the continuous copula. In order to be consistent throughout the rest of the example, we choose a grid size of 200 by 200 points.

The constraints placed on (A,B) and proposed as the initial bases are $A^4B = 0.2231, B^3A^2 = 0.3586, A^2B^4 = 0.5902, A^5B^2 = 0.7590, B^2A^2 = 0.4113$. We confirm that the D_1AD_2 have converged given the mentioned constraints. The lambda values associated with the fitted minimum information copula to these constraints are $\lambda_1 = .583, \lambda_2.274, \lambda_3 = 26.307, \lambda_4 = 128.045, \lambda_5 = 6.569, \lambda_5 = .3.910$, and the corresponding log-likelihood is $l_{AB} = 35.32$. Similarly, we construct the second copula in the first tree which is denoted by c_{BC}. The calculated log-likelihood for c_{BC} is $l_{BC} = 21.17$.

The conditional copulas in the second tree can similarly be approximated using the minimum information approach. Initially we construct the conditional minimum information copula between $C|B$ and $A|B$. In order to calculate this copula we divide the support of B into some arbitrary sub-intervals or bins and then construct the conditional copula within each bin. To do so we find bases in the same way as for the marginal copulas and fit the copulas to the expectations calculated for these. We use five bins so that the first copula is for $A, C|B \in (0, 0.25)$. Let us consider the first bin as $I_1 = (0.0.025)$. The bases for this copula are

$$[h'_1(A,C) \mid B \in I_1] = A^2C^3, \ [h'_2(A,C) \mid B \in I_1] = CA^2, [h'_3(A,C) \mid B \in I_1] = A^2C^4$$

$$[h'_4(A,C) \mid B \in I_1] = A^5C, [h'_5(A,C) \mid B \in I_1] = A^5C^2$$

We do the same for the remaining bins. Table 1 shows the constraints and corresponding Lagrange multipliers required to build the conditional minimum information copula between $A, C \mid B \in (0,1)$. It also gives the log-likelihood in each bin. The density approximation presented above using the minimum information copula illustrates a very good performance and precision on in comparison to the parametric approaches explain in this paper. In order to make a comparison between these two methods, we have computed the log-likelihoods of the fitted copulas (the parametric copula and the minimum information based copula) to the data for the same vine structure. The results show that the log-likelihood corresponding the parametric approach by choosing the t-copula for the bivariate copulas in the first tree and the Gaussian copula on the conditional copula on the second tree is 80.28, while the log-likelihood corresponding to the minimum information based copula fitted to each bivariate copula in the vine structure is 97.42. Bedford et al (2012) reported the similar results when they study the Norwegian financial data with four variables.

5 Conclusion

In this paper, we briefly review the method of Bedford et al. (2012) [3] to approximate a multivariate distribution by any vine structure to any degree of approximation. We have operationalized the theoretical approximation results by using minimum information copulas that can be specified to any required degree of precision based on the data available. It can be shown that that good approximation 'locally' guarantees good approximation globally (Bedford et al., 2012). We can use the same bases to approximate each copula in each tree of the corresponding vine, with the promise of a uniform level of approximation. Another objective of this paper was to find appropriate data analysis tools to determine the D-vine structure and a suitable parametric copula family for each bivariate copula in the vine structure. To sum it up, the methods discussed in this paper allow to efficiently construct D-vine models even in higher dimensions. The results derived in this paper show that the new method proposed by Bedford et al. (2012) is outperformed the parametric method.

We have extended this method by using the Fourier series and some other orthonormal series as the basis functions to approximate the positive and continuous density function by truncating the series at an appropriate point. The derived results shows a considerable improvement in the density approximation and faster computation. This is can be seen in a working paper by the authors in [7].

References

[1] Aas, K., Czado, K. C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance, Mathematics and Economics*, **44**, 182–198.

[2] Bedford, T., and Cooke. R. M. (2002). Vines - a new graphical model for dependent random variables. *Annals of Statistics*, **30(4)**: 1031–1068.

[3] Bedford, T. Daneshkhah, A. and Wilkinson, K. (2012) Approximate Uncertainty Modeling with Vine copulas. *Submitted to Journal European Journal of Operational Research*

[4] Bedford, T., and Meeuwissen, A. (1997). Minimally informative distributions with given rank correlation for use in uncertainty analysis. *Journal of Statistical Computation and Simulation*, **57(1 - 4)**: 143 - 174.

[5] Karimi, M., Poorahmadi, M. R. and Daneshkhah, A. (2012). Pair-copula constructions based on bivariate copulas from different families. Presented in *Conference on Computer Science and Computational Mathematics.*

[6] Kurowicka, D., and Cooke. R. (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling.* John Wiley.

[7] O. Chatrabgoun, A. Daneshkhah, G. A. Parham and M. Poorahmadi. (2012). Approximation Multivariate Distribution with pair copula Based on the Orthonormal Polynomial Function. *A working paper.*