Far East Journal of Theoretical Statistics
Volume 30, Number 1, 2010, Pages 61-70
Published Online: January 5, 2010
This paper is available online at http://www.pphmj.com

A NEW COPULA FUNCTION FOR CONSTRUCTION OF MULTIVARIATE DISTRIBUTIONS WITH SPECIFIED MARGINALS

S. A. SHABAN and AYMAN ORABI

Institute of Statistical Studies and Research
Cairo University, Egypt
e-mail: drshabanshaban@yahoo.com

College of Business Administration
King Saud University
Saudi Arabia
e-mail: orabi_issr@yahoo.com

Abstract

A new copula function for construction of bivariate and multivariate distributions with specified marginals will be introduced. The properties of linear correlation and rank correlation of a new copula function in bivariate case will be studied. The comparison between a new copula and Farlie-Gumbel-Morgenstern bivariate copula will be studied. In section a new copula in multivariate case will be given. The general properties of a new copula function in multivariate case and constrains on its parameters α’s in trivariate case will be studied.

1. Introduction

Since the second half of 20th century, the construction of bivariate and multivariate distributions with specified marginals, which can be normal or nonnormal, has been become a basic problem in statistical science for understanding

2010 Mathematics Subject Classification: Primary 60E05; Secondary 62E15.

Keywords and phrases: copula function, bivariate distributions, multivariate distributions, linear correlation and rank correlation.

Received July 13, 2009

relationships among bivariate or multivariate outcomes, also it is an important for the other sciences. For an example in actuarial science, when two lives are subject to failure, such as under a joint life insurance or annuity policy, we are concerned with joint distribution of lifetimes. As another example, when we simulate the distribution of a scenario that arises out of a financial security system, we need to understand the distribution of several variables interacting simultaneously, not in isolation of one another. So the construction of bivariate and multivariate distributions is discussed by many literatures, Gumbel [3], Johnson and Kotz [5], D'Este [1], Schweizer and Wolff [11], Gupta and Wong [4], Long and Krzysztofowicz [8], Johnson et al. [6], Fang et al. [2], Nelsen [9], Kole et al. [7] and others.

In this paper, we review the Farlie-Gumbel-Morgenstern bivariate copula function in Section 2. In Section 3, we introduce a new copula function for construction of multivariate distributions with specified marginals, study its properties and constrains on its parameters α's in trivariate case. In Section 4, we introduce a new copula function for construction of bivariate distributions with specified marginals and study linear correlation, rank correlation and comparison between a new copula and Farlie-Gumbel-Morgenstern bivariate copula.

2. Farlie-Gumbel-Morgenstern Bivariate Construction

Johnson and Kotz [5] discussed the construction of bivariate distributions using method due to Farlie, Gumbel and Morgenstern system. (FGM system of bivariate distribution). The FGM system provides a general technique by which a bivariate distribution can be constructed directly from its marginal distribution and correlation between the variables. The FGM bivariate cumulative distribution function of the random variables X_1 and X_2 is given by

$$F(X_1, X_2) = F_1(X_1) F_2(X_2)[1 + \alpha(F_1(X_1) - 1)(F_2(X_2) - 1)], \tag{1}$$

where $F(X_j)$ is the marginal cumulative density function of the random variable X_j for $j = 1, 2,$ and the parameter α is directly proportional to the correlation coefficient and the limits of α were obtained by Johnson and Kotz [5] as $|\alpha| < 1.$ Pearson's correlation coefficient ρ between the random variables has the same sign as α, and is proportional to it. An upper bound for ρ in the FGM system is given by Schucany et al. [10] who proved that $|\rho| < 1/3.$ The corresponding joint probability density function for (1) is given by

$$f(X_1,\ X_2) = f_1(X_1)\,f_2(X_2)[1 + \alpha(2F_1(X_1) - 1)(2F_2(X_2) - 1)], \tag{2}$$

where $f(X_j)$ is the marginal probability density distribution, i.e., p.d.f. of the random variable X_j for $j = 1,\ 2.$

Note that, the equation (1) can be express as the copula function as the following:

$$C(U,\ V,\ \alpha) = UV[1 + \alpha(U - 1)(V - 1)] \tag{3}$$

and the corresponding joint probability density function for (3) is given by

$$c(U,\ V,\ \alpha) = 1 + \alpha(2U - 1)(2V - 1), \tag{4}$$

where $0 \le U \le 1,\ 0 \le V \le 1$ are uniform distributions.

3. A New Copula Function in Multivariate Case

The new copula function which hence after denoted by GFGM copula. It can be considered as generalization of FGM copula. The GFGM copula function of p random variables can be given by

$$\begin{aligned} &C(U_1,\ U_2,\ \ldots,\ U_p) \\ &= \prod_{j=1}^{p}(U_j) + A^2 \sum_{j_1<}^{p-1}\sum_{j_2}^{p} \alpha_{j_1 j_2} G_{j_1}(U_{j_1}) G_{j_2}(U_{j_2}) \prod_{i=1}^{p} H_1(U_i) \\ &\quad + A^3 \sum_{j_1<}^{p-2}\sum_{j_2<}^{p-1}\sum_{j_3}^{p} \alpha_{j_1 j_2 j_3} G_{j_1}(U_{j_1}) G_{j_2}(x_{j_2}) G_{j_3}(U_{j_3}) \prod_{i=1}^{p} H_2(U_i) \\ &\quad + \cdots + A^p \alpha_{1,2,3,\ldots,p} \prod_{j=1}^{p} G_j(U_j), \end{aligned} \tag{5}$$

where

$$U_j = F_j(x_j), \quad G_j(U_j) = \left[(2U_j - 1)^{\frac{2(\beta+1)}{2\beta+1}} - 1\right],$$

$$H_1(U_i) = \begin{cases} U_i, & \forall\ i \neq j_1,\ j_2, \\ 1, & \forall\ i = j_1,\ j_2, \end{cases}$$

$$H_2(U_i) = \begin{cases} U_i, & \forall\ i \neq j_1,\ j_2,\ j_3, \\ 1, & \forall\ i = j_1,\ j_2,\ j_3, \end{cases}$$

and so on

$$A = \frac{2\beta + 1}{4(\beta + 1)} \quad \text{and } \beta = 0, 1, 2, \ldots.$$

Properties.

(i) $\lim_{U_j \to 0} C(U_1, U_2, \ldots, U_p) = 0$, for all $j = 1, 2, \ldots, p$, since

$$\lim_{U_j \to 0} \prod_{j=1}^{p} (U_j) = 0 \text{ and } \lim_{U_j \to 0} G_j(U_j) = \lim_{U_j \to 0} \left[(2U_j - 1)^{\frac{2(\beta+1)}{2\beta+1}} - 1 \right] = 0.$$

(ii) $C(1, 1, \ldots, 1) = 1$, since $\prod_{j=1}^{p} (1) = 1$ and $G_j(1) = \left[(2-1)^{\frac{2(\beta+1)}{2\beta+1}} - 1 \right] = 0.$

(iii) $C(U_1, U_2, \ldots, U_{P-1}, 1) = C(U_1, U_2, \ldots, U_{P-1})$, since

$$G_j(U_p = 1) = \left[(2-1)^{\frac{2(\beta+1)}{2\beta+1}} - 1 \right] = 0.$$

The corresponding joint probability density to equation (5) is given by

$$f_{1,2,\ldots,p}(X_1, X_2, \ldots, X_p) = \prod_{j=1}^{p} (f_j(X_j)) \Bigg[1 + \sum_{j_1<}^{p-1} \sum_{j_2}^{p} \alpha_{j_1 j_2} g_{j_1}(X_{j_1}) g_{j_2}(X_{j_2})$$

$$+ \sum_{j_3}^{p} \alpha_{j_1 j_2 j_3} g_{j_1}(X_{j_1}) g_{j_2}(X_{j_2}) g_{j_3}(X_{j_3})$$

$$+ \cdots + \alpha_{1,2,3,\ldots,p} \prod_{j=1}^{p} g_j(X_j) \Bigg], \tag{6}$$

where

$$g_j(X_j) = (2F_j(X_j) - 1)^{\frac{1}{2\beta+1}}.$$

Note that, when $\beta = 0$ in equation (6) is multivariate distribution for FGM construction.

Constraints on α's

The α's must be such that the joint probability density function is nonnegative for all *X*. From (18), we see that we must have

$$\Bigg[1+\sum_{j_1<}^{p-1}\sum_{j_2}^{p}\alpha_{j_1j_2}g_{j_1}(X_{j_1})g_{j_2}(X_{j_2})$$
$$+\sum_{j_1<}^{p-2}\sum_{j_2<}^{p-1}\sum_{j_3}^{p}\alpha_{j_1j_2j_3}g_{j_1}(X_{j_1})g_{j_2}(X_{j_2})g_{j_3}(X_{j_3})$$
$$+\cdots+\alpha_{1,2,3,\dots,p}\prod_{j=1}^{p}g_j(X_j)\Bigg]\geq 0, \tag{7}$$

hence the substitution of $F_j(X_j) = 0, 1\ \ (j = 1, \dots,\ p)$ in every one of the possible combinations will yield necessary and sufficient condition on the values of the α's. These conditions are therefore

$$1+\sum_{j_1<}^{p-1}\sum_{j_2}^{p}c_{j_1}c_{j_2}\alpha_{j_1j_2}+\sum_{j_1<}^{p-2}\sum_{j_2<}^{p-1}\sum_{j_3}^{p}c_{j_1}c_{j_2}c_{j_3}\alpha_{j_1j_2j_3}$$
$$+\cdots+\alpha_{1,2,3,\dots,p}\prod_{j=1}^{p}c_j\geq 0, \tag{8}$$

where $c_j = \pm 1$.

For $p = 3$, we have

$$1+\alpha_{12}+\alpha_{23}+\alpha_{13}+\alpha_{123}\geq 0 \text{ at } (c_1 = c_2 = c_3 = 1), \tag{9}$$

$$1+\alpha_{12}+\alpha_{23}+\alpha_{13}-\alpha_{123}\geq 0 \text{ at } (c_1 = c_2 = c_3 = -1), \tag{10}$$

$$1-\alpha_{12}+\alpha_{23}-\alpha_{13}+\alpha_{123}\geq 0 \text{ at } (c_1 = 1,\ c_2 = c_3 = -1), \tag{11}$$

$$1-\alpha_{12}-\alpha_{23}+\alpha_{13}+\alpha_{123}\geq 0 \text{ at } (c_2 = 1,\ c_1 = c_3 = -1), \tag{12}$$

$$1+\alpha_{12}-\alpha_{23}-\alpha_{13}+\alpha_{123}\geq 0 \text{ at } (c_3 = 1,\ c_1 = c_2 = -1), \tag{13}$$

$$1+\alpha_{12}-\alpha_{23}-\alpha_{13}-\alpha_{123}\geq 0 \text{ at } (c_3 = -1,\ c_1 = c_2 = 1), \tag{14}$$

$$1-\alpha_{12}-\alpha_{23}+\alpha_{13}-\alpha_{123}\geq 0 \text{ at } (c_2 = -1,\ c_1 = c_3 = 1), \tag{15}$$

$$1-\alpha_{12}+\alpha_{23}-\alpha_{13}-\alpha_{123}\geq 0 \text{ at } (c_1 = -1,\ c_2 = c_3 = 1). \tag{16}$$

We can summarize the conditions on the α's, for $p = 3$, as follows:

From inequalities (13) and (14)

$$|\alpha_{13} + \alpha_{23} \pm \alpha_{123}| \leq 1 + \alpha_{12}.$$

From inequalities (12) and (15)

$$|\alpha_{12} + \alpha_{23} \pm \alpha_{123}| \leq 1 + \alpha_{13}.$$

From inequalities (11) and (16)

$$|\alpha_{12} + \alpha_{13} \pm \alpha_{123}| \leq 1 + \alpha_{23},$$

and from inequalities (9) and (10)

$$|\alpha_{123}| \leq 1 + \alpha_{12} + \alpha_{13} + \alpha_{23}.$$

4. A New Copula Function in Bivariate Case

The GFGM copula in bivariate case is given by

$$C(U, V, \alpha, \beta) = UV + \alpha\left[\frac{2\beta + 1}{4(\beta + 1)}\right]^2 \left[(2U - 1)^{\frac{2(\beta+1)}{2\beta+1}} - 1\right]\left[(2V - 1)^{\frac{2(\beta+1)}{2\beta+1}} - 1\right], \quad (17)$$

where $0 \leq U \leq 1$, $0 \leq V \leq 1$ are uniform distributions, $|\alpha| \leq 1$ and $\beta = 0, 1, 2, \ldots$. The corresponding joint density function for $C(U, V, \alpha, \beta)$ of U and V is given by

$$c(U, V, \alpha, \beta) = 1 + \alpha\left[(2U - 1)^{\frac{1}{2\beta+1}}\right]\left[(2V - 1)^{\frac{1}{2\beta+1}}\right] \quad (18)$$

and in terms of $f_1(x_1)$, $f_2(x_2)$ is given by

$$f(X_1, X_2) = f_1(X_1) f_2(X_2)\left[1 + \alpha(2F_1(X_1) - 1)^{\frac{1}{2\beta+1}}(2F_2(X_2) - 1)^{\frac{1}{2\beta+1}}\right], \quad (19)$$

where $f_j(X_j)$ is the marginal density function of X_j and $F_j(X_j)$ is the marginal cumulative distribution function of x_j for $j = 1, 2$. Note that when $\beta = 0$ the equation $f(X_1, X_2)$ becomes joint density of Farlie-Gumbel-Morgenstern.

4.1. Measures of correlation

(a) Linear correlation

Theorem 1. *If* x_1 *and* x_2 *are bivariate GFGM with finite nonzero variances, then*

$$|\rho| \leq \alpha\left(\frac{2\beta+1}{2\beta+3}\right), \tag{20}$$

where ρ *is Pearson's correlation coefficient.*

Proof. Let

$$\delta_j = \int X_j(2F_j(X_j)-1)^{\frac{1}{2\beta+1}} dF_j(X_j). \tag{21}$$

From equations (19) and (21), we have

$$\operatorname{cov}(X_1, X_2) = \alpha\delta_1\delta_2. \tag{22}$$

Let σ_j^2 be the variance of x_j, $j = 1, 2$, the

$$\rho = \frac{\alpha\delta_1\delta_2}{\sigma_1\sigma_2}, \tag{23}$$

from equation (21)

$$\delta^2 = \left[\int X(2F(X)-1)^{\frac{1}{2\beta+1}} dF(X)\right]^2$$

since

$$\int \mu(2F(X)-1)^{\frac{1}{2\beta+1}} dF(X) = 0,$$

we have

$$\delta^2 = \left[\int X(2F(X)-1)^{\frac{1}{2\beta+1}} dF(X)\right]^2 = \left[\int (X-\mu)(2F(X)-1)^{\frac{1}{2\beta+1}} dF(X)\right]^2$$

from the Cauchy-Schwarz inequality

$$\delta^2 = \left[\int X(2F(X)-1)^{\frac{1}{2\beta+1}} dF(X)\right]^2 = \left[\int (X-\mu)(2F(X)-1)^{\frac{1}{2\beta+1}} dF(X)\right]^2$$

$$\leq \int (X-\mu)^2 dF(X) \int (2F(X)-1)^{\frac{2}{2\beta+1}} dF(X),$$

but

$$\int (X - \mu)^2 dF(X) = \sigma^2 \text{ and } \int (2F(X) - 1)^{\frac{2}{2\beta+1}} dF(X) = \frac{2\beta + 1}{2\beta + 3},$$

then

$$\delta^2 = \left[\int X(2F(X) - 1)^{\frac{1}{2\beta+1}} dF(X)\right]^2 = \left[\int (X - \mu)(2F(X) - 1)^{\frac{1}{2\beta+1}} dF(X)\right]^2$$

$$\leq \int (X - \mu)^2 dF(X) \int (2F(X) - 1)^{\frac{2}{2\beta+1}} dF(X) = \sigma^2 \left(\frac{2\beta + 1}{2\beta + 3}\right), \qquad (24)$$

from equation (23)

$$|\rho| \leq \alpha\left(\frac{2\beta + 1}{2\beta + 3}\right).$$

Special cases

(i) From Theorem 1, if $\alpha = 1$ and $\beta = 0$, then $|\rho| \leq \frac{1}{3}$, that is to say FGM is a special case of our GFGM copula.

(ii) From Theorem 1, if $\alpha = 1$ and $\beta \to \infty$, then $|\rho| \leq 1$.

(b) Rank correlation

Schweizer and Wolff [11] showed that two standard nonparametric correlation measures could be expressed solely in terms of the copula function. These are Spearman correlation coefficient, defined by

$$\rho_s(X_1, X_2) = 12 \int\int [C(U, V) - UV] dUdV, \qquad (25)$$

and Kendall's correlation coefficient, defined by

$$\tau(x_1, x_2) = 4 \int\int C(U, V) dC(U, V) - 1. \qquad (26)$$

Theorem 2. *Spearman's correlation coefficient* $\rho_s(\alpha, \beta)$ *between u and v, for the GFGM copula is given by*

$$\rho_s(\alpha, \beta) = 3\alpha\left(\frac{2\beta + 1}{4\beta + 3}\right)^2. \qquad (27)$$

Proof. From equations (5) and (13), we have

$$\rho_S(\alpha,\ \beta) = 12\alpha\left[\frac{2\beta+1}{4(\beta+1)}\right]^2 \int_0^1 \left[(2U-1)^{\frac{2(\beta+1)}{2\beta+1}} - 1\right] dU \int_0^1 \left[(2V-1)^{\frac{2(\beta+1)}{2\beta+1}} - 1\right] dV$$

$$= 12\alpha\left[\frac{2\beta+1}{4(\beta+1)}\right]^2 \left[\frac{2\beta+1}{4\beta+3} - 1\right]^2$$

$$= 3\alpha\left(\frac{2\beta+1}{4\beta+3}\right)^2.$$

Theorem 3. *Kendall's correlation coefficient* $\tau(\alpha,\ \beta)$ *between u and v, for the GFGM copula is given by*

$$\tau(\alpha,\ \beta) = 2\alpha\left(\frac{2\beta+1}{4\beta+3}\right)^2. \tag{28}$$

Proof. From equations (5) and (14), we have

$$\tau(\alpha,\ \beta) = 4\int_0^1\int_0^1 \left\{UV + \alpha\left[\frac{2\beta+1}{4(\beta+1)}\right]^2 \left[(2U-1)^{\frac{2(\beta+1)}{2\beta+1}} - 1\right]\left[(2V-1)^{\frac{2(\beta+1)}{2\beta+1}} - 1\right]\right\}$$

$$= 2\alpha\left(\frac{2\beta+1}{4\beta+3}\right)^2.$$

4.2. Comparison between GFGM copula and FGM copula

In fact the linear or rank correlation coefficient is important factor for comparison among methods of constructing bivariate distribution so we can compare the maximum possible value of ratio ρ/α between GFGM copula and FGM copula for some distributions as the following Table 1:

Table 1. The maximum possible value of ratio ρ/α of FGM and GFGM for some distributions

Family of distribution	FGM construction	GFGM copula
Uniform	$1/3$	$3/4$
Exponential	$1/4$	0.48
Gamma	0.315	0.629
Beta	0.297	0.801
Normal	$1/\pi$	$2/\pi$

This table indicates that the GFGM copula has a wider range than the FGM copula.

References

[1] G. M. D'Este, A Morgenstern-type bivariate gamma distribution, Biometrika 68(1) (1981), 339-340.

[2] Kai-Tai Fang, H.-B. Fang and D. von Rosen, A family of bivariate distributions with non-elliptical contours, Comm. Statist. Theory Methods 29(8) (2000), 1885-1898.

[3] E. J. Gumbel, Bivariate exponential distributions, J. Amer. Statist. Assoc. 55 (1960), 698-707.

[4] A. K. Gupta and C. F. Wong, On three and five-parameter bivariate beta distributions, Metrika 32(2) (1985), 85-91.

[5] N. L. Johnson and S. Kotz, On some generalized Farlie-Gumbel-Morgenstern distributions-II, regression, correlation and further generalizations, Comm. Statist. Theory Methods A6 (1977), 485-496.

[6] N. L Johnson, S. Kotz, and N. Balakrishnan, Continuous Multivariate Distributions, Vol. 2, John Wiley and Sons, New York, 2000.

[7] E. Kole, K. Koedijk and M. Verbeek, Selecting copulas for risk management, Journal of Banking and Finance 31(8) (2007), 2405-2423.

[8] D. Long and R. Krzysztofowicz, A family of bivariate densities constructed from marginals, J. Amer. Statist. Assoc. 90(430) (1995), 739-746.

[9] R. B. Nelsen, An Introduction to Copulas, 2nd ed., New York, Springer Verlag, 2006.

[10] W. R. Schucany, E. W. Parr and J. E. Boyer, Correlation structure in Farlie-Gumbel-Morgenstern distributions, Biometrika 65(3) (1978), 650-653.

[11] B. Schweizer and E. F. Wolff, On nonparametric measures of dependence for random variables, Ann. Statist. 9(4) (1981), 879-885.