



## Construction of Continuous Bivariate Distribution by Transmuting Dependent Distribution

Hüseyin ÜNÖZKAN<sup>1</sup> , Mehmet YILMAZ<sup>2\*</sup> 

<sup>1</sup> Ankara University, Graduate School of Natural and Applied Science, Department of Statistics, Turkey

<sup>2</sup> Ankara University, Faculty of Science, Department of Statistics, 06100, Ankara, Turkey

Received: 10.09.2019; Accepted: 02.10.2019

<http://dx.doi.org/10.17776/csj.618236>

**Abstract.** In this study, a new bivariate distribution family is introduced by adding an appropriate term to independent class. By choosing a base distribution which is negatively dependent from the same marginals we derive a new distribution around the product of marginals, i.e. independent class of distribution. We note that the new distribution has additional parameter which would provide additional flexibility in applications. The joint probability density, joint reliability and reversed hazard rate functions of the new bivariate distribution are obtained. Furthermore, we obtain lower and upper bounds of Spearman's correlation coefficient. Two example are given to illustrate this family. This new bivariate continuous distribution can make more appropriate modeling of some data sets in terms of the Spearman rank coefficient.

**Keywords:** Transmuted bivariate distribution, Dependence, Bivariate distribution, Spearman's Rho correlation coefficient, Fréchet bounds.

## Bağımlı Dağılım Fonksiyonunun Dönüştürülmesi ile İki Boyutlu Sürekli Dağılım Fonksiyonu Oluşturulması

**Özet.** Bu çalışmada, bağımsız dağılım fonksiyonuna uygun bir terim eklenerek yeni bir iki boyutlu dağılım fonksiyonu tanıtılmıştır. Bu yeni dağılım karmaşık bir yapıda değildir. Aynı marjinalere sahip dağılım fonksiyonları sınıfından bir temel dağılım seçilerek, bağımsız dağılım fonksiyonu etrafında yeni dağılım türetilmiştir. Bu yeni dağılımın fazladan bir parametresi olup, uygulama alanlarında modelleme esnasında esneklik sağlayacağı düşünülmektedir. Bu yeni iki boyutlu dağılımın ortak olasılık yoğunluk fonksiyonu, ortak güvenilirlik fonksiyonu ve ters bozulma oranı fonksiyonu elde edilmiştir. Bunun yanı sıra, bu metot ile elde edilen yeni dağılım fonksiyonunun Spearman Sıra Korelasyonu katsayısı bakımından biraz daha esneklik kazandırabileceği söylenebilir.

**Anahtar Kelimeler:** Dönüştürülmüş iki boyutlu dağılım, Bağımlılık, İki boyutlu dağılım, Spearman sıra korelasyonu, Fréchet sınırları.

### 1. INTRODUCTION

In both statistical theory and practice, univariate models are sometimes insufficient to explain random phenomena. Bivariate distributions are very important in modeling dependent random quantities in many different areas. We need to construct a joint distribution with specific marginals and higher or lower correlations. Dolati and Ubeda-Flores [1] introduced a method based on the choice of pairs of order statistics of the marginal distributions. Lai and Xie [2] studied on construction of continuous bivariate distributions that possesses the Positive Quadrant Dependence property. According to similar work of [2], [3] introduces some conditions for negatively dependent families. Inspired by these studies, we desire to propose a simpler but useful model. After giving the necessary conditions to construct a new distribution,

\* Corresponding author. Email address: mehmetyilmaz@ankara.edu.tr  
<http://dergipark.gov.tr/csj> ©2016 Faculty of Science, Sivas Cumhuriyet University

Spearman's rank correlation coefficient is calculated on two illustrative examples and the usefulness of this family is discussed. Furthermore, some reliability properties are studied for this family.

Let  $(X_1, Y_1)$  and  $(X_2, Y_2)$  be two independent vectors of random variables with common distribution function  $H(x, y)$ . Note that,  $H(x, y)$  belongs to the distribution family  $\mathcal{F}(F, G)$  where  $F$  and  $G$  denote respectively marginals of  $X$  and  $Y$ . Let  $X_{(1)}, X_{(2)}$  and  $Y_{(1)}, Y_{(2)}$  be their corresponding order statistics. According to [4], consider the random vector

$$(Z_1, Z_2) = \begin{cases} (X_{(1)}, Y_{(2)}), & \text{with probability } 1/2 \\ (X_{(2)}, Y_{(1)}), & \text{with probability } 1/2. \end{cases}$$

Then the distribution of  $(Z_1, Z_2)$  is given by

$$H_1(x, y) = H(x, y)[1 - \bar{H}(x, y)], \quad (1)$$

where  $\bar{H}(x, y)$  denotes survival function of  $(X, Y)$  i.e.,  $Pr(X > x, Y > y)$ . If we consider the random vector

$$(T_1, T_2) = \begin{cases} (X_{(1)}, Y_{(1)}), & \text{with probability } 1/2 \\ (X_{(2)}, Y_{(2)}), & \text{with probability } 1/2. \end{cases}$$

Then the distribution function of  $(T_1, T_2)$  is given by

$$H_2(x, y) = F(x)G(y) + H(x, y)\bar{H}(x, y). \quad (2)$$

If we look at the equations (1) and (2) immediately, it is possible to say that the model formed by obtaining a mixture of  $H_1$  and  $H_2$  is not a simple structure. With this in mind, our contribution is to propose a simpler model. According to eq. (2.1) and the conditions (2.2)-(2.4) of Han (2011) [3], the following assumption is sufficient for our purpose: Suppose that  $\psi(x, y) = -H(x, y)\bar{H}(x, y)$ . Then we have a function defined as  $H_1^*(x, y) = F(x)G(y) + \psi(x, y)$ . For subsequent discussions, following theorem explains the necessity of having negative dependence (or independence) condition on  $H(x, y)$ .

**Theorem1.** Let  $H(x, y)$  be a distribution function belongs to the distribution family  $\mathcal{F}(F, G)$  which is differentiable on  $\mathbb{R}^2$  and  $h(x, y) = \frac{\partial^2 H(x, y)}{\partial x \partial y}$  denote the joint probability density function. Then  $H_1^*(x, y) = F(x)G(y) - H(x, y)\bar{H}(x, y)$  is a distribution function if  $H(x, y) \leq F(x)G(y)$ , for all  $(x, y) \in \mathbb{R}^2$  (or  $H(x, y) = F(x)G(y)$ , for all  $(x, y) \in \mathbb{R}^2$ ).

**Proof.** Multivariate distribution function must satisfy following properties (see, Barlow and Proschan, 1975, Chapter 5), [4]:

**(P1)**

$$\lim_{x \rightarrow \infty} F(x)G(y) - H(x, y)\bar{H}(x, y) = G(y),$$

$$\lim_{y \rightarrow \infty} F(x)G(y) - H(x, y)\bar{H}(x, y) = F(x),$$

$$\lim_{x \wedge y \rightarrow \infty} F(x)G(y) - H(x, y)\bar{H}(x, y) = 1.$$

**(P2)**  $\frac{\partial H_1^*(x, y)}{\partial x} \geq 0$  and  $\frac{\partial H_1^*(x, y)}{\partial y} \geq 0$ . For the simplicity  $f_x = \frac{dF(x)}{dx}$  and  $g_y = \frac{dG(y)}{dy}$ . Then

$$\begin{aligned} \frac{\partial H_1^*(x, y)}{\partial x} &= f_x G(y) - \frac{\partial H(x, y)}{\partial x} \bar{H}(x, y) - H(x, y) \frac{\partial \bar{H}(x, y)}{\partial x} \\ &= f_x (G(y) - Pr(Y \leq y|X = x)\bar{H}(x, y) + H(x, y)Pr(Y > y|X = x)) \end{aligned}$$

Now, by noting that negatively dependence implies  $Pr(Y \leq y|X = x) \leq G(y)$ , then we have

$$\begin{aligned} \frac{\partial H_1^*(x, y)}{\partial x} &\geq f_x (Pr(Y \leq y|X = x)[1 - \bar{H}(x, y)] + H(x, y)Pr(Y > y|X = x)) \\ &\geq 0. \end{aligned}$$

Obviously,  $\frac{\partial H_1^*(x, y)}{\partial y} \geq 0$ .

**(P3)**  $\frac{\partial^2 H_1^*(x, y)}{\partial x \partial y} \geq 0$ . For the simplicity, let  $h_0 = f(x)g(y)$ ,  $H_0 = F(x)G(y)$  and  $h_{xy} = \frac{\partial^2 H(x, y)}{\partial x \partial y}$ . Then

$$\frac{\partial^2 H_1^*(x, y)}{\partial x \partial y} = h_0 - h_{xy}[H(x, y) + \bar{H}(x, y)] - \frac{\partial H(x, y)}{\partial x} \frac{\partial \bar{H}(x, y)}{\partial y} - \frac{\partial H(x, y)}{\partial y} \frac{\partial \bar{H}(x, y)}{\partial x}.$$

Now, according to Domma (2011) [5], by noting that negative dependence implies both  $h_{xy}H(x, y) \leq \frac{\partial H(x, y)}{\partial x} \frac{\partial H(x, y)}{\partial y}$  and  $h_{xy}\bar{H}(x, y) \leq \frac{\partial \bar{H}(x, y)}{\partial x} \frac{\partial \bar{H}(x, y)}{\partial y}$ . Hence, we have

$$\frac{\partial^2 H_1^*(x, y)}{\partial x \partial y} \geq h_0 - \left[ \frac{\partial H(x, y)}{\partial x} + \frac{\partial \bar{H}(x, y)}{\partial x} \right] \left[ \frac{\partial H(x, y)}{\partial y} + \frac{\partial \bar{H}(x, y)}{\partial y} \right].$$

By considering  $\frac{\partial \bar{H}(x, y)}{\partial x} = -f_x + \frac{\partial H(x, y)}{\partial x}$  and  $\frac{\partial \bar{H}(x, y)}{\partial y} = -f_y + \frac{\partial H(x, y)}{\partial y}$ , the expressions in square brackets are rewritten as  $h_0 - 2f_y \frac{\partial H(x, y)}{\partial x} - 2f_x \frac{\partial H(x, y)}{\partial y} + 4 \frac{\partial H(x, y)}{\partial x} \frac{\partial H(x, y)}{\partial y}$ . After some simplification, the above inequality becomes:

$$\begin{aligned} \frac{\partial^2 H_1^*(x, y)}{\partial x \partial y} &\geq 2 \frac{\partial H(x, y)}{\partial x} \left( f_y - \frac{\partial H(x, y)}{\partial y} \right) + 2 \frac{\partial H(x, y)}{\partial y} \left( f_x - \frac{\partial H(x, y)}{\partial x} \right) \\ &\geq 0. \end{aligned}$$

This completes the proof.

According to Theorem1, by assuming  $H(x, y)$  be negatively dependent or independent, we can define a new pairs of random variables  $U$  and  $V$  as below:

$$(U, V) = \begin{cases} (T_1, T_2), & \text{with probability } \alpha \\ (Z_1^*, Z_2^*), & \text{with probability } 1 - \alpha, \end{cases}$$

where  $(Z_1^*, Z_2^*)$  is distributed as  $H_1^*$ . Hence, the distribution of  $(U, V)$  is given by

$$\begin{aligned} F(x, y) &= Pr(U \leq x, V \leq y) = \alpha H_2(x, y) + (1 - \alpha)H_1^*(x, y) \\ &= F(x)G(y) + (2\alpha - 1)H(x, y)\bar{H}(x, y). \end{aligned} \tag{3}$$

By letting  $\lambda = 2\alpha - 1$ , where  $\lambda \in [-1, 1]$ , eq. (3) can be rewritten as

$$F(x, y) = F(x)G(y) + \lambda H(x, y)\bar{H}(x, y). \tag{4}$$

$\lambda = 0$  indicates  $F = F(x)G(y)$ ,  $\lambda = -1$  indicates that  $F$  is negatively dependent and  $\lambda = 1$  indicates that  $F$  is positively dependent. Note that, if  $H(x, y) = F(x)G(y)$ ,  $F$  indicates well-known bivariate distribution which is called as Farlie-Gumbel-Morgenstern distribution (see, [6] and [7]).

We need the survival and probability density function for subsequent discussions. These functions are respectively given by

$$\bar{F}(x, y) = \bar{F}(x)\bar{G}(y) + \lambda H(x, y)\bar{H}(x, y)$$

and

$$f(x, y) = f(x)g(y) + \lambda h(x, y)[\bar{H}(x, y) + H(x, y)] - \lambda k(x, y),$$

where  $k(x, y) = Pr(X \leq x|Y = y)Pr(Y > y|X = x) + Pr(X > x|Y = y)Pr(Y \leq y|X = x)$ .

## 2. REVERSED HAZARD RATE OF THE NEW FAMILY OF BIVARIATE DISTRIBUTION

The bivariate reversed hazard is defined by Bismi [8] as  $rv(x, y) = f(x, y)/F(x, y)$ . Furthermore, the bivariate hazard rate function defined by Basu [9] as  $r(x, y) = f(x, y)/\bar{F}(x, y)$ . Analogously to the hazard gradient by Johnson and Kotz [10], Roy [11] defined the bivariate reversed hazard rate and the bivariate hazard rate as follows:  $rv_{1,2}(x, y) = (rv_1(x, y), rv_2(x, y))$ , where

$$rv_1(x, y) = \frac{\partial \log F(x, y)}{\partial x}, rv_2(x, y) = \frac{\partial \log F(x, y)}{\partial y},$$

and  $r_{1,2}(x, y) = (r_1(x, y), r_2(x, y))$ , where

$$r_1(x, y) = \frac{-\partial \log \bar{F}(x, y)}{\partial x}, r_2(x, y) = \frac{-\partial \log \bar{F}(x, y)}{\partial y}.$$

Reversed hazard rate gradients of  $F(x, y)$  given by eq. (4) are as follows:

$$\begin{aligned} rv_1(x, y) &= w(x, y, \lambda)rv_1(x, \infty) + (1 - w(x, y, \lambda))(rv_{1H}(x, y) - r_{1H}(x, y)), \\ rv_2(x, y) &= w(x, y, \lambda)rv_2(\infty, y) + (1 - w(x, y, \lambda))(rv_{2H}(x, y) - r_{2H}(x, y)). \end{aligned}$$

Accordingly, after some simplifications, bivariate reversed hazard rate can be given by

$$\begin{aligned} rv(x, y) &= w(x, y, \lambda)rv_1(x, \infty)rv_2(\infty, y) + (1 - w(x, y, \lambda))(rv_H(x, y) + r_H(x, y)) \\ &\quad - (1 - w(x, y, \lambda))[rv_{1H}(x, y)r_{2H}(x, y) + rv_{2H}(x, y)r_{1H}(x, y)], \end{aligned}$$

where  $w(x, y, \lambda) = \frac{F(x)G(y)}{F(x)G(y) + \lambda H(x, y)\bar{H}(x, y)}$ .

## 3. LOWER AND UPPER BOUNDS ON SPEARMAN'S RHO MEASURE FOR THE NEW FAMILY OF BIVARIATE DISTRIBUTION

This section deals with obtaining bounds for the bivariate distribution family given by the eq. (4). According to Hoeffding [12] and Fréchet [13], for any bivariate distribution belonging to  $\mathcal{F}(F, G)$  contains Fréchet a lower bound and an upper bound. These bounds are respectively defined as

$$F^-(x, y) = \max\{F(x) + G(y) - 1, 0\} \quad (5)$$

$$F^+(x, y) = \min\{F(x), G(y)\}. \quad (6)$$

For  $F \in \mathcal{F}(F, G)$ , Spearman's rho can be expressed as

$$\rho_s(X, Y) = 12 \int_{\mathbb{R}} \int_{\mathbb{R}} \{F(x, y) - F(x)G(y)\} dG(y) dF(x) \quad (7)$$

(see, Schweizer and Wolff [14]). The coefficient of Spearman's rho for the new family can be obtained by

$$\rho_s = 12\lambda \int_{\mathbb{R}} \int_{\mathbb{R}} H(x, y)\bar{H}(x, y)dG(y)dF(x). \tag{8}$$

Hence, by using the fact that  $H(x, y) \leq F(x)G(y)$ , for  $\lambda > 0$ , we have the upper bound as  $\rho_s \leq \frac{\lambda}{3}$ . To obtain the lower bound for  $\lambda > 0$ , we use the eq. (5), then the lower bound is  $\rho_s \geq 0$ . According to sign of  $\lambda$ , we achieve the bounds as below:

$$\rho_s \in \begin{cases} \left[ \frac{-\lambda}{3}, 0 \right], & \text{for } \lambda < 0 \\ 0 & \text{for } \lambda = 0 \\ \left[ 0, \frac{\lambda}{3} \right] & \text{for } \lambda > 0. \end{cases}$$

We have two example to illustrate this family.

**Example1.** The Farlie-Gumbel –Morgenstern (FGM) family of bivariate distributions are given by  $H(x, y) = F(x)G(y)[1 + \theta \bar{F}(x)\bar{G}(y)]$ , for  $\theta \in [-1,1]$ . By taking  $\theta \in [-1,0]$ , the distribution  $F(x, y)$  is given by

$$F(x, y) = F(x)G(y) + \lambda F(x)G(y)\bar{F}(x)\bar{G}(y)[1 + \theta \bar{F}(x)\bar{G}(y)][1 + \theta F(x)G(y)],$$

where  $\lambda \in [-1, 1]$  and  $\theta \in [-1,0]$ . Hence,  $\rho_s = \lambda \left( \frac{1}{3} + \frac{\theta}{6} + \frac{\theta^2}{75} \right)$ . Since  $\theta \in [-1,0]$ ,  $\frac{9\lambda}{50} \leq \rho_s \leq \frac{\lambda}{3}$  for  $\lambda > 0$ , and  $\frac{\lambda}{3} \leq \rho_s \leq \frac{9\lambda}{50}$  for  $\lambda < 0$ .

We conclude that this family model weak dependence as FGM does.

**Example2.** The bivariate Gumbel- Exponential (BGE) distribution is given by  $H(x, y) = 1 - e^{-x} - e^{-y} + e^{-x-y-\theta xy}$ , for  $\theta \in [0,1]$ . The distribution  $F(x, y)$  is given by

$$F(x, y) = 1 - e^{-x} - e^{-y} + e^{-x-y} + \lambda \left( \frac{e^{-x-y-\theta xy} - e^{-2x-y-\theta xy} - e^{-x-2y-\theta xy}}{+e^{-2x-2y-2\theta xy}} \right),$$

where  $\lambda \in [-1, 1]$  and  $\theta \in [0,1]$ . According to Yela and Cuevas [15], the Spearman's rho coefficient of BGE distribution is  $\rho_s^{BGE} = 12 \left[ -\frac{e^{\frac{4}{\theta}} Ei \left( -\frac{4}{\theta} \right) - \frac{1}{4}}{\theta} \right]$ , where  $Ei(\cdot)$  is the exponential integral function.

Then,  $\rho_s$  can be obtained as

$$\rho_s = 12\lambda \frac{e^{\frac{4}{\theta}}}{\theta} \left[ -Ei \left( -\frac{4}{\theta} \right) + 2e^{\frac{2}{\theta}} Ei \left( -\frac{6}{\theta} \right) - \frac{e^{\frac{1}{2\theta}} Ei \left( -\frac{9}{2\theta} \right)}{2} \right].$$

The following plot shows the shape of  $\rho_s^{BGE}$  for different values of  $\theta$ . Here, the approximate values obtained by Matlab package program are used in  $Ei$  evaluations.

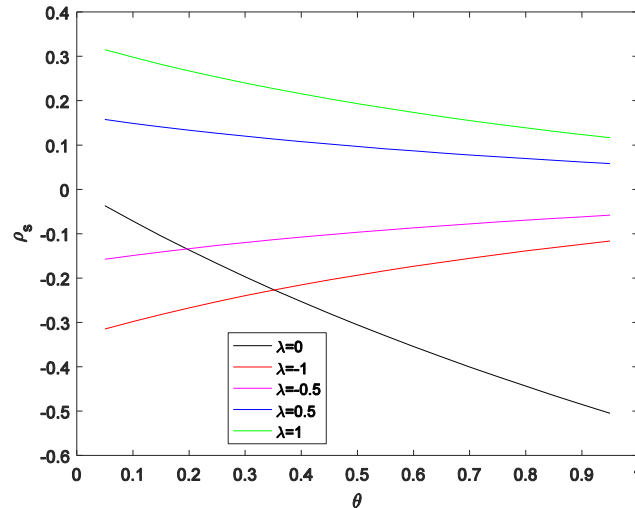


Figure 1. Plots of the  $\rho_s$  for some parameter values.

As can be seen from the Figure1, for the negative values of  $\lambda$ , it can be said that  $F$  can model negative dependence slightly better than BGE in small theta values. Moreover, for the positive values of  $\lambda$ ,  $F$  can be used to model weak positive dependence.

#### 4. CONCLUSION

In this study, we proposed a new bivariate distribution using a base distribution from the negative dependency class which is in  $\mathcal{F}(F, G)$ . Thus, this new distribution can reveal both negative dependence, positive dependence and independence between the random variables  $X$  and  $Y$ . As a result of illustrative examples, it can be said that distributions can be derived for pairs of random variables with higher correlations considering some base distributions.

If we pay attention to Example 2 again, we can explain a structure by illustrating a random phenomenon which looks essentially negatively dependent but may also be positively dependent as follows:

As the temperatures usually begin to increase in spring, snow starts to melt into the stream. This leads to a rapid increase in the water level of the river. Thereby, it can be thought that there is a positive dependence between temperature and the amount of water level of the river. On the other hand, during the summer season, as the temperatures rise, the water level in the river will decrease due to evaporation. In this case, there is a negative dependence between these two random variables.

The proposed distribution may be useful to model the relationship between river water level and temperature all year round.

#### REFERENCES

- [1] Dolati A. and Ubeda-Flores M., Constructing Copulas by Means of Pairs of Order Statistics, *Kybernetika*, 45-6 (2009) 992-1002.
- [2] Lai C. D. and Xie M., A New Family of Positive Quadrant Dependent Bivariate Distributions, *Statistics and Probability Letters*, 46-4 (2000) 359-364.
- [3] Han Kwang-Hee., A New Family of Negative Quadrant Dependent Bivariate Distributions with Continuous Marginal, *Journal of the Chungcheong Mathematical Society*, 24-4 (2011) 795-805.
- [4] Technical Report, Holt, Rinehart and Winston, New York, 1975.

- [5] Domma F., Bivariate Reversed Hazard Rate, Notions, and Measures of Dependence and their Relationships, *Communications in Statistics - Theory and Methods*, 40-6 (2011) 989-999, DOI: 10.1080/03610920903511777.
- [6] Farlie D., The Performance of Some Correlation Coefficients for a General Bivariate Distribution, *Biometrika*, 47-3/4 (1960) 307-323.
- [7] Gumbel E. J., Bivariate Exponential Distributions, *Journal of the American Statistical Association*, 55-292 (1960) 698-707.
- [8] Bismi G., Bivariate Burr Distributions, unpublished PhD Thesis, Cochin University of Science and Technology, 2005.
- [9] Basu A.P., Bivariate Failure Rate, *Journal of the American Statistical Association*, 66 (1971) 103–104.
- [10] Johnson N. L. and Kotz S., A Vector Multivariate Hazard Rate, *Journal of Multivariate Analysis*, 5-1 (1975) 53-66.
- [11] Roy D., A Characterization of Model Approach for Generating Bivariate Life Distributions Using Reversed Hazard Rates, *Journal of Japan Statistical Society*, 32-2 (2002) 239–245.
- [12] Hoeffding W., Masstabinvariante Korrelationstheorie, *Schriften des Mathematischen Instituts und Instituts für Angewandte Mathematik der Universität Berlin*, 5 (1940) 181-233.
- [13] Fréchet M., Sur Les Tableaux de Corrélation Dont Les marges Sont Donnees, *Annales de l'Université de Lyon, Sciences*, 4 (1951) 13–84.
- [14] Schweizer B. and Wolff E., On Nonparametric Measures of Dependence for Random Variables, *The Annals of Statistics*, 9-4 (1981) 879-885.
- [15] Yela P.J. and Cuevas T.J.R., Estimating the Gumbel-Barnett Copula Parameter of Dependence, *Revista Colombiana de Estadística*, 41-1 (2018) 53-73.