IRAQI Academic Scientific Journals

University of Kirkuk Journal For Administrative and Economic Science (2023) 13 (2): 295-300

UKJAES

University of Kirkuk Journal For Administrative and Economic Science

ISSN 2222-2995
University of Kirkuk Journal For Administrative and Economic Science

**Ahmed Nawzad M. & Hamdin Ayad Othman & Mohamad Soran Husen. Copula Probability Model for Determining Dependency and Generations.** ***University of Kirkuk Journal For Administrative and Economic Science*** **(2023) 13 (2):295-300.**

# Copula Probability Model for Determining Dependency and Generations

Nawzad M. Ahmed [1], Ayad Othman Hamdin [2], Soran Husen Mohamad [3]

[(1,2,3)] *Sulaimaniyah University: College Administration & Economics, Statistics & Informatics Department*

[1] nawzad.mahmud@univsul.edu.iq
[2] ayad.hamdin@univsul.edu.iq
[3] soran.abdulrahman@univsul.edu.iq

**Abstract:** The statistical idea of a copula was developed in 1959 under the base of sklar's theorem; it was not applied to financial markets and finance until the late 1990s. In this study, copula was used as a probability model that represents a multivariate uniform distribution (MUD), first to examines the association or dependence between many variables as a vector of random variables (Vn) having different probability densities, that helps to isolate the joint or marginal probabilities of a pair of variables that are enmeshed in a more complex multivariate system, secondly it was used for generating the probability values of dependencies for a vector of random variables together in MVCDF rather than their partial dependencies from their marginals.

**Keywords:** multivariate uniform distribution (MUD), vector of random variables (Vn), multivariate cumulative distribution function (MVCDF), copula distribution function( C), copula probability density function(c).

# نموذج احتمالية الكوبولا لتحديد التبعية والأجيال

د. نوزاد محمد احمد [١]، م .اياد عثمان حمدين [٢]، م.م. سوران حسين محمد [٣]

[(١،٢،٣)] *جامعة السليمانية- كلية الإدارة والاقتصاد، قسم الإحصاء والمعلوماتية، السليمانية، العراق*

**المستخلص:** تم تطوير الفكرة الإحصائية للكوبولا في عام ١٩٥٩ تحت قاعدة نظرية سكلار ؛ لم يتم تطبيقه على الأسواق المالية والتمويل حتى أواخر التسعينيات. في هذه الدراسة ، تم استخدام الكوبولا كنموذج احتمالي يمثل توزيعًا موحدًا متعدد المتغيرات (MUD) ، أولاً لفحص الارتباط أو الاعتماد بين العديد من المتغيرات كمتجه لمتغيرات عشوائية (Vn) لها كثافات احتمالية مختلفة ، مما يساعد على عزل الاحتمالات المشتركة أو الهامشية لزوج من المتغيرات المتضمنة في نظام متعدد المتغيرات أكثر تعقيدًا ، وثانيًا تم استخدامه لتوليد القيم الاحتمالية للتبعيات لمتجه من المتغيرات العشوائية معًا في MVCDF بدلاً من تبعياتها الجزئية من الهامش.

**الكلمات المفتاحية:** التوزيع الموحد متعدد المتغيرات (MUD)، ناقل المتغيرات العشوائية (Vn)، دالة التوزيع التراكمي متعدد المتغيرات (MVCDF)، دالة توزيع الكوبولا (C)، دالة كثافة احتمال الكوبولا (ج).

**Corresponding Author: E-mail:** nawzad.mahmud@univsul.edu.iq

**1- Introduction**:

Statistically in any random process, the independence among random variables in a vector V_n^'=(X_1,X_2,…,X_d) is hypothetical, and rarely occurs. Then it can be said that they are not significantly depend on each other, which is depends on the multiple correlation or rate of dependency after testing this hypothesis statistically, then their dependencies (direct or indirect), (positive or negative) logically can't be ignored regardless of its size specially in those studied that characterized by extremely sensitive in their data, moreover, its applicably seldom to collect a vector of random variables having independence property, so it's too difficult to evaluate a multivariate probability distribution comes from a vector of random variables each have different population or probability distribution. Copula is a function to express joint distributions for a vector of random variables where they have dependence or not, comes from several populations. With a copula you can separate the joint distribution into two contributions: the marginal distributions of each variable by itself, and the copula that combines these into a joint distribution. One basic result is that any joint distribution can be expressed in this manner. Another convenience is that the conditional distributions can be readily expressed using the copula. What Is a Copula? Copula is a probability model that represents a multivariate uniform distribution, which examines the association or dependence between many variables. Put differently, a copula helps isolate the joint or marginal probabilities of a pair of variables that are collected in a more complex multivariate system. The copula is then the unique index or set of instructions for describing how those pairs fit together in the more complex system. This method is useful as it can help identify spurious correlations observed in the data. It is also useful in fine-tuning derivatives pricing models where the price of one depends on the price of some underlying thing. Today, copulas are employed in advanced financial analysis to better understand outcomes that involve fat tails and skewness.

**2-Methodology:**

This section in the study is to explained the method and criterion of copula and how it can be used for releasing a strongly correlated vector of continuous random variables to another uniform combined uncorrelated marginal vector.

**3- Description of Copula Technique:**

If F(X) and F(Y) are continuous, then C is unique; otherwise C is uniquely determined on Range F(X) × Range F(Y) only. Conversely, for any univariate distribution functions F(X) and F(Y) and any copula C, the function F(XY) is a valid bivariate distribution function with marginal F(X) and F(Y) . The popularity of copulas for dependence modelling largely follows from quotes like 'Copulas allow us to separate the dependence from multivariate to the product of marginal distributions. Clearly, if( C ) is unique, then it unequivocally characterizes how the two marginal F(X) and F(Y) interlock for producing the joint behavior of (X, Y), while staying ignorant of what those marginal are. For instance, if X and Y are independent, and if C is unique, then from the definition of copula ( C) must be the 'independence copula or so called 'product copula'). In order to compute a copula value that corresponding the probability of dependency for a given interval inside the range of (X, and Y), then the values of each corresponding range for (U, and V) can be evaluated from the eq.(1), means: for a given interval( $x_1 < x \le x_2,\ y_1 < y \le y_2$ ) , using $C_{U,V}(u,v)$ produces a corresponding interval for bivariate uniform ( $u_1 < u \le u_2,\ u_1 < y \le u_2$), and then calculating bivariate copula from the property below that's the probability of dependency for random vector (X, And Y) together at a specific value .

$$C_{U,V}(u,v) = \Pr[(u_1 \le U \le u_2),(v_1 \le V \le v_2)$$

$$C_{U,V}(u,v) = C(u_2,v_2) + C(u_1,v_1) - C(u_1,v_2) - C(u_2,v_1) \ \ge 0 \qquad \text{------ (1)}$$

$$\text{For all } 0 \le u_1 \le u \le u_2 \le 1 \text{ , and} 0 \le v_1 \le v \le v_2 \le 1$$

**4-Description of Sklar's Theorem:**

Sklar's theorem provides the theoretical foundation for the application of copulas. This theorem states that every multivariate cumulative distribution function with the form:

$G(x_1, x_2, \dots, x_d) = \Pr\left(X_1 \le x_1, X_2 \le x_2, \dots, X_d \le x_d\right)$ ----------- (2)

of a random vector $(X_1, X_2, \dots, X_d)$ {\displaystyle (X_{1},X_{2},\dots ,X_{d})} Can be expressed in terms of its marginal

$F_i(x_i) = \Pr(X_i \le x_i)$ {\displaystyle F_{i}(x_{i})=\Pr[X_{i}\leq x_{i}]} and a copula (C)is{\displaystyle C}is $G(x_1, x_2, \dots, x_d)$

{\displaystyle H(x_{1},\dots ,x_{d})=C\left(F_{1}(x_{1}),\dots ,F_{d}(x_{d})\right).}In case that the multivariate distribution has a density{\displaystyle h} (g), and if eq. (2) is differentiable then, it holds that:

$g(x_1, x_2, \dots, x_d) = c\big(F_1(x_1), F_2(x_2), \dots, F_d(x_d)\big).f_1(x_1).f_2(x_2) \dots f_d(x_d)$ ----- (3)

Where (c) is the probability density function of copula:

3- Example:

Let $V_n' = (X_1, X_2)$ is a bivariate vector of random variables; denoted two columns functioned with a joint probability density function given by:

$$f(x, y) = \left(\frac{1}{320}\right)(x_1 + 2x_2) \quad , \text{ where } 0 < x_1 \le 4 \text{ , and } \quad 0 < x_2 \le 8$$
$$= 0 \qquad \text{elsewhere} \qquad \text{--------------- (4)}$$

In the first step required to derive the inverse cumulative distribution function formula for each $(X_1, \text{and } X_2)$ as $F_{X_1}^{-1}(u), \text{and } F_{X_2}^{-1}(v)$ , Where$(u, \text{and } v)$ , are two uniformly distributed [0, 1] random variables one to one correspondent respectively to$(X, \text{and } Y)$ . and secondly deriving bivariate copula cumulative distribution function formula as:

$C_{U,V}(u, v) = F_{X_1,X_2}(x_1, x_2)$ ------------------ (5)

The marginals rv's (X, and Y) each alone comes from integrating the joint p.d.f from eq.(4) we can calculate respectively: $f_{X_1}(x_1), \text{and } f_{X_2}(x_2)$ given by:

$f_{X_1}(x_1) = (1/40)(8 + x_1)$ , where $0 < x_1 \le 4$

---------------- (6)

$f_{X_2}(x_2) = ({}^{1}/_{40})(1 + x_2)$ , where $0 < x_2 \le 8$

Using sklar's theorem criteria:

Let $u = F_{X_1}(x_1), \text{ and } v = F_{X_2}(x_2) \rightarrow x_1 = F_{X_1}^{-1}(u), \text{and } x_2 = F_{X_2}^{-1}(v)$ -------- (7)

To get each $u = F_{X_1}(x_1), \text{and } v = F_{X_2}(x_2)$ , it can be calculated by integrating the marginal $f_{X_1}(x_1), \text{and } f_{X_2}(x_2)$ given by:

$$u = F_{X_1}(x_1) = \int_0^{x_1} f_X(x_1)\, dx_1 = (x_1/5)[1 + (x_1/16)] \qquad 0 < x_1 \le 4$$

------------- (8)

And $v = F_{X_2}(x_2) = \int_0^{x_2} f_{X_2}(x_2)\, dt = (x_2/40)[1+(x_2/2)]$ $0 < x_2 \le 8$

From eq(4) each of $x_1$ , and $x_2$ can be calculated

$x_1 = F_{X_1}^{-1}(u) = (-16 \pm \sqrt{256 + 320u})/2$ . Because of $0 < x_1 \le 4$ hence the negative root ignored then

$x_1 = F_{X_1}^{-1}(u) = (-16 + \sqrt{256 + 320u})/2$ , with range $0 \le u \le 1$

--------------- (9)

$x_2 = F_{X_2}^{-1}(u) = (-2 + \sqrt{4 + 320v})/2$ , with range $0 \le v \le 1$

In order to derive copula distribution function given by eq. (2), then

$$C_{U,V}(u, v) = F_{X_1,X_2}(x_1, x_2) = \int_0^{x_2}\int_0^{x_1} f_{X_1X_2}(x_1, x_2)dx_1dx_2$$

$C_{U,V}(u, v) = F_{X_1,X_2}(x_1, x_2) = \left(\frac{1}{640}\right)(x_1^2x_2 + 2x_1x_2^2)$ --------- (10)

Substituting the values of $x_1$, and $x_2$ from eq. (6) in eq. (7) produces bivariate copula cumulative distribution function given by:

$$C_{U,V}(u,v) = (1/640)[\{\frac{-16+\sqrt{256+320u}}{2}\}^2(-2+\frac{\sqrt{4+320v}}{2}) + 2*\left(\frac{-16+\sqrt{256+320u}}{2}\right)\left(\frac{-2+\sqrt{4+320v}}{2}\right)^2\}] \quad \text{--------------- (11)}$$

Table (1): the univariate probability of dependency for each $(x_{1,}$ and$x_2)$, separately with corresponding uniform random variable (U, and V) respectively using eq. (11):

| Range of r.v, X | Range of r.v, U | Prob. of Dependency Between (x,y) | Range of r.v, Y | Range of r.v, V | Prob. of Dependency Between (u,v) |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 1.000 | Completely dependent | 8 | 1.000 | Completely dependent |
| 1 | 0.2125 | weak | 1 | 0.03750 | weak |
| 1.5 | 0.3281 | weak | 1.5 | 0.06563 | weak |
| 2.2 | 0.5001 | strong | 2 | 0.10000 | weak |
| 3 | 0.7125 | strong | 3 | 0.18750 | weak |
| 3.5 | 0.8531 | strong | 3.5 | 0.24060 | weak |
| 2.5 | 0.578 | strong | 4 | 0.30000 | weak |
| 2.7 | 0.6311 | strong | 2.5 | 0.13200 | weak |
| 2.9 | 0.6851 | strong | 5.5 | 0.51560 | strong |
| 3.8 | 0.9405 | strong | 6 | 0.60000 | strong |
| 3.9 | 0.97013 | strong | 7.5 | 0.89100 | strong |

This table explains dependency values for each $(X_{1,}$ and$X_2)$ versus uniform random Variables (U, and V) .

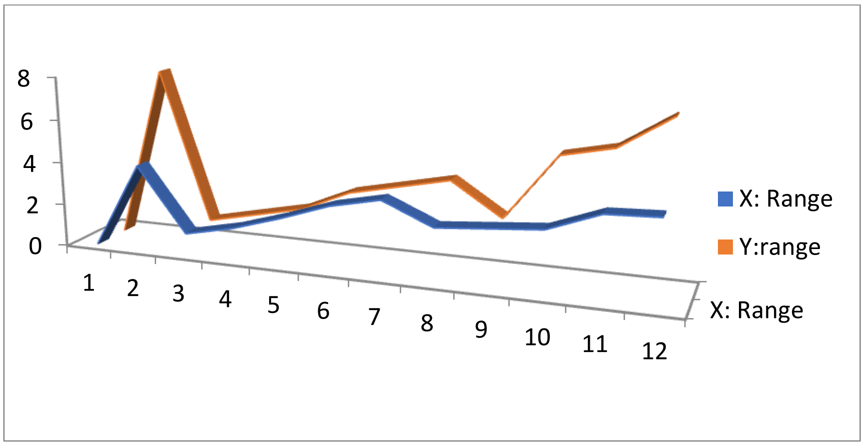


**Figure 1**: Dependency among (x,and y)

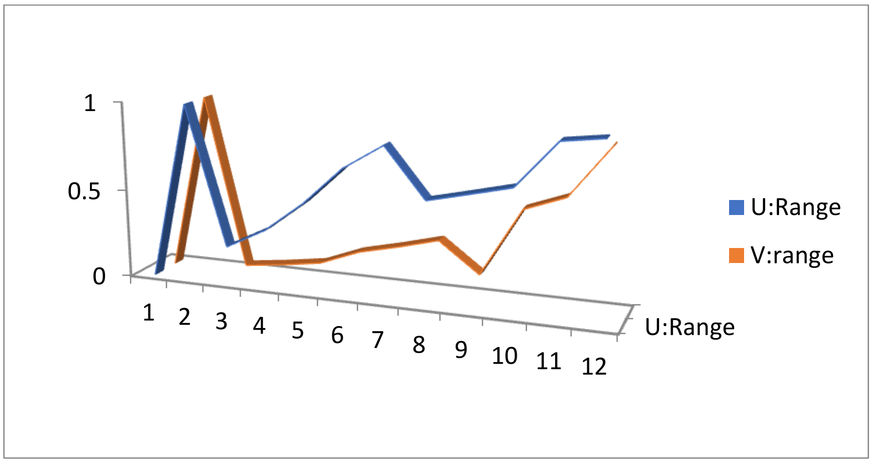


**Figure 2**: Dependency among (U, and V)

**Note:** figures (1,2) are from data in table(1) above by executing (Microsoft Excel).

## 5- Analysis and Results:

Copulas are popular in high-dimensional statistical applications as they allow one to easily stochastic modeling and estimating the distribution of random vectors by estimating marginal and copula separately. There are many parametric copula families available, which usually have parameters that control the strength of dependence.

From the two figures above for measures (x, and y), and (u, and v) its clear that the dependency among (u,and v) is explained more precisely than ( x,and y)that's because of the regular joint d.f is not assigned for (x, and y) to be identical, but copula required for (u,and v) to be identically independent distributed. Also one can see easily that the correlation for any two order pair (u, v) in copula is smaller than the joint d.f for (x, y) is one of the important property for copula function, see the dependency probability for (x,y) and compare with this probability for (u, and v).

In other hand copula function makes generation of random numbers and simulating for any two independent random variables is much easier after transforming the joint distribution function for (x, y) to the uniform one having ranges[0,1], this because the copula breaking down the correlation among variables if its exist.

## 6- conclusions:

-It's concluded that about the tails of boundaries, always there are completely independence in the lower bound of the range for the variables (x=0 versus to u=0), and (y=0 versus to v=0), and completely dependency in the upper bound(x=4 versus to u= 1), and (y=8 versus to v=1).

-from the table(1) and figures(1,2) above, the effect of copula appears clearly that it succeeded to decrease the dependency among origin variables(x,y) after transforming them to copula function by more than three times.

## 7- recommendations:

-Its important to the researchers whom works in applied statistics, to use copula technique especially for (Regression or Classification) data modelling to get certain independence as it's required among factors or that commonly named explanatory.

-it's recommended as a future work, to expand copula for discrete type of data for joint probability density function p(x1, x2,….,xn ).

**8-References:**

1- Alidoost, F., Su, Z., & Stein, A. (2019). Evaluating the effects of climate extremes on crop yield, production and price using multivariate distributions: A new copula application. Weather and climate extremes, 26, 100227.
2- Chesneau, C. (2023). Theoretical Validation of New Two-Dimensional One-Variable-Power Copulas. Axioms, 12(4), 392.France; chesneau.christophe@gmail.com.
3- Czado, C. (2010). Pair-copula constructions. In: F. Durante, W. Hardle, P. Jaworski, T. Rychlik ¨ (eds.) Workshop on Copula Theory and its Applications. Springer, Dortrecht (NL)
4- Durante, F., Fernandez-Sanchez, J., & Sempi, C. (2013). A topological proof of Sklar's theorem. Applied Mathematics Letters, 26(9), 945-948., doi:10.1016/j.aml.2013.04.005
5- Eban, E., Rothschild, G., Mizrahi, A., Nelken, I., & Elidan, G. (2013, April). Dynamic copula networks for modeling real-valued time series. In Artificial Intelligence and Statistics (pp. 247-255). PMLR. (PDF), Journal of Machine Learning Research, 31
6- Giacomini, E., Härdle, W., & Spokoiny, V. (2009). Inhomogeneous dependence modeling with time-varying copulae. Journal of Business & Economic Statistics, 27(2), 224-234.
7- Jan,-F., Mai, (March 8, 2022)," About the exact simulation of bivariate (reciprocal) Archimax copulas" https://doi.org/10.1515/demo-2022-0102 .
8- Lazoglou, G., & Anagnostopoulou, C. (2019). Joint distribution of temperature and precipitation in the Mediterranean, using the Copula method. Theoretical and Applied Climatology, 135, 1399-1411.
9- Nelsen, R. B. (1999). An introduction to copulas. New York, USA. Lecture Notes in Statistics.
10- Schoelzel, C., & Friederichs, P. (2008). Multivariate non-normally distributed random variables in climate research–introduction to the copula approach. Nonlinear Processes in Geophysics, 15(5), 761-772.. doi:10.5194/npg-15-761-2008