A TRIANGULAR SYMMETRIC COPULA

Alexander Andronov

Transport and Telecommunication Institute, Riga, Latvia
lora@mailbox.riga.lv

1. Introduction. An important element of a probabilistic model is a description of considered random variable distributions. It is usually supposed that all random variables are independent. But numerous statistical data prove the opposite. For example, it has been experimentally stated that characteristics of Internet flows are the dependent ones [3]. Analogously, flows of insurance claims for damages have a dependent structure [2]. In the first case, a correlation between interarrivals of the claims is described by the so called Markov-Additive Processes of Arrivals [5]. In the second case, copulas are usually used for a description of the dependence [1, 2, 4]. Lately, the copulas have been used in the reliability theory [7].

Joint distribution function $C(u_1,u_2,...,u_n)=P\{U_1\le u_1,U_2\le u_2,...,U_n\le u_n\}$ is called *a copula* if the marginal distributions of all components $U_1,U_2,...,U_n$ are uniform on [0, 1]. The following fact is basic [6]: any multivariate continuous distribution function $G(x_1,x_2,...,x_n)=$ $=P\{X_1\le x_1,X_2\le x_2,...,X_n\le x_n\}$ can be presented uniquely via cumulative distribution functions of its components $F_i(x_i)=P\{X_i\le x_i\}$ by the corresponding copula C: $G(x_1,x_2,...,x_n)=C(F_1(x_1),F_2(x_2),...,F_n(x_n))$.

In our paper we wish to use a natural and direct way to describe the dependence between continue random variables X_0, $X_1,...$. We suppose that the last ones correspond to a Markov chain with continue state space Ω = (0, 1). To describe transit probability per one step we consider a family $\Psi=\{f_\theta(x):x,\theta\in\Omega\}$ of triangular distributions that is determined by a probability density function

$$f_\theta(x)=\begin{cases}2\dfrac{x}{\theta}, & 0<x<\theta,\\ 2\dfrac{1-x}{1-\theta}, & \theta\le x<1.\end{cases}\tag{1}$$

The corresponding cumulative distribution function is calculated as

$$F_\theta(x)=\begin{cases}\dfrac{1}{\theta}x^2, & 0<x<\theta,\\ 1-\dfrac{(1-x)^2}{1-\theta}, & \theta\le x<1.\end{cases}\tag{2}$$

Further, let $X_1,X_2,...,X_n,...$ be a random variable sequence that forms Markov chain: if $X_n=x$, then X_{n+1} has a distribution from family Ψ with parameter $\theta=x$. Our aim is to find for this a Markov chain: marginal stationary distribution $G(x)=P\{X\le x\}$, two-dimensional stationary distribution $F(x,y)=P\{X_n\le x,X_{n+1}\le y\}$ and its moments, a corresponding copula. This copula allows us to form a corresponding Frechet family and to consider some new distributions.

2. Stationary marginal distribution. Let $g(x) = G(x)'$ be the probability density function for the stationary marginal distribution. Then, a consideration of two consecutive terms of a stationary Markov chain gives:

$$G(x)=\int_0^x g(\theta)\left[1-\frac{(1-x)^2}{1-\theta}\right]d\theta+\int_x^1 g(\theta)\frac{x^2}{\theta}d\theta=G(x)-\int_0^x g(\theta)\frac{(1-x)^2}{1-\theta}d\theta+\int_x^1 g(\theta)\frac{x^2}{\theta}d\theta.$$

$$0=-(1-x)^2\int_0^x g(\theta)\frac{1}{1-\theta}d\theta+x^2\int_x^1 g(\theta)\frac{1}{\theta}d\theta. \tag{3}$$

$$\int_0^x g(\theta)\frac{1}{1-\theta}d\theta=\left(\frac{x}{1-x}\right)^2\int_x^1 g(\theta)\frac{1}{\theta}d\theta. \tag{4}$$

Taking a derivative form (3) with respect to x and using (4), we have:

$$0=2(1-x)\int_0^x g(\theta)\frac{1}{1-\theta}d\theta-(1-x)^2 g(x)\frac{1}{1-x}+2x\int_x^1 g(\theta)\frac{1}{\theta}d\theta-x^2 g(x)\frac{1}{x},$$

$$0=2\frac{x^2}{1-x}\int_x^1 g(\theta)\frac{1}{\theta}d\theta-(1-x)g(x)+2x\int_x^1 g(\theta)\frac{1}{\theta}d\theta-xg(x),$$

$$g(x)=\frac{2x}{1-x}\int_x^1 g(\theta)\frac{1}{\theta}d\theta,\quad g(x)\frac{1-x}{2x}=\int_x^1 g(\theta)\frac{1}{\theta}d\theta.$$

Taking a derivative form of the last equation with respect to x, we get

$$g'(x)\frac{1-x}{2x}+g(x)\left(-\frac{1}{2x^2}\right)=-g(x)\frac{1}{x},$$

$$g'(x)\frac{1-x}{2x}=\frac{1}{2}g(x)\left[\frac{1}{x^2}-\frac{2}{x}\right],\quad \frac{g'(x)}{g(x)}=\frac{1-2x}{x(1-x)},$$

$$\ln(g(x))=c+\int\frac{1-2x}{x(1-x)}dx=c+\left(\int\frac{1}{x}dx\right)-\int\frac{1}{1-x}dx=c+\ln(x)+\ln(1-x).$$

$$g(x)=cx(1-x),\ \ 0<x<1.$$

An unknown constant c is determined by the normalization property:

$$1=\int_0^1 g(x)dx=c\int_0^1 x(1-x)dx=c\left[\frac{1}{2}-\frac{1}{3}\right]=c\frac{1}{6}.$$

Therefore, $c = 6$ and finally

$$g(x)=6x(1-x),\ \ 0<x<1. \tag{5}$$

Stationary cumulative function is calculated as

$$G(x)=x^2(3-2x). \tag{6}$$

3. Two-dimensional stationary marginal distribution. Further the two-dimensional case will be considered when $(X,Y)=(X_1,X_2)$. Taking into account formulas (5) and (2), we get for a cumulative distribution function of (X, Y):

$$F(x,y)=\begin{cases}\int_y^x 6z(1-z)\left[\frac{1}{z}y^2\right]dz\ +\int_0^y 6z(1-z)\left(1-\frac{(1-y)^2}{1-z}\right)dz & \textit{for}\quad y\le x,\\ \int_y^x 6z(1-z)\left(1-\frac{(1-y)^2}{1-z}\right)dz & \textit{for}\quad y\le x.\end{cases}$$

After some simplifications, we get a final formula

$$F(x,y)=\begin{cases}3y^2x(2-x)-2y^3, & y\le x,\\ 3x^2y(2-y)-2x^3, & y>x.\end{cases} \tag{7}$$

It allows us to write the two-dimensional probability density function:

$$f(x,y)=\begin{cases}12y(1-x), & y\le x,\\ 12x(1-y), & y>x.\end{cases} \tag{8}$$

4. Moments and Kendall's τ. From (5) we get the marginal expectation, second moment, variance and standard deviation: $E(X) = 1/2, E(X^2) = 3/10, D(X) = 1/20, \sigma = \sqrt{1/20}$. The second mixed moment is calculated from (8):

$$E(XY) = \int_0^1 \left(\int_0^x xy12y(1-x)dy + \int_x^1 xy12x(1-y)dy \right) dx = \frac{4}{15},$$

$$Cov = E(XY) - E(X)^2 = \frac{4}{15} - \frac{1}{4} = \frac{1}{60}.$$

It allows us calculating the correlation coefficient:

$$\rho = \frac{Cov}{D(X)} = \frac{20}{60} = \frac{1}{3}.$$

Now we wish to calculate Kendall's τ that is defined for two independent copies of (*X*, *Y*) and (*X'*, *Y'*) as [4]:

$$\tau = 4P\{X \le X', Y \le Y'\} - 1.$$

From (7) and (8) we have

$$P\{X \le X', Y \le Y'\} = \int_0^1 \int_0^x 12y(1-x)\left[3y^2 x(2-x) - 2y^3\right] dydx + \int_0^1 \int_x^1 12x(1-y)\left[3x^2 y(2-y) - 2x^3\right] dydx =$$

$$= 36\int_0^1 (1-x)x(2-x)\int_0^x y^3 dydx - 24\int_0^1 (1-x)\int_0^x y^4 dydx + \int_0^1 \int_0^y 12x(1-y)\left[3x^2 y(2-y) - 2x^3\right] dxdy =$$

$$= 2\left[36\int_0^1 x(2-3x+x^2)\frac{1}{4}x^4 dx - 24\int_0^1 (1-x)\frac{1}{5}x^5 dx\right] = \frac{903}{2940}.$$

Therefore,

$$\tau = 4P\{X \le X', Y \le Y'\} - 1 = 4\frac{903}{2940} - 1 = \frac{3612}{2940} - 1 = \frac{672}{2940} = \frac{56}{245}.$$

5. Copula and Frechet family. To find a copula for our distribution, we must express cumulative distribution function (7) through its marginal distribution function (6). For that purpose, it is necessary to know quantile $u = G^{-1}(p)$ of probability *p*, $0 < p < 1$. Therefore, we have a cubic equation

$$x^2(3-2x) - p = 0, \quad 0 \le x \le 1. \tag{9}$$

A standard analysis shows that for each $p \in (0, 1)$ this equation has a unique root in interval (0, 1), namely,

$$u = 0.5 - \cos\left(\frac{1}{3}\left(Arc\cos(1-2p) - 2\pi\right)\right), \quad 0 < p < 1. \tag{10}$$

From (7) we have for $0 < v \le u < 1$:

$$F(x,y) = 3y^2 x(2-x) - 2y^3 = 3y^2 x(2-x) + \left[y^2(3-2y)\right] - 3y^2 =$$

$$= 3y^2\left[x(2-x) - 1\right] + G(y) = G(y) - 3y^2(1-x)^2.$$

Now from (10) we have the following copula for $0 < v \le u < 1$:

$$C(u,v) = v - 3\left[\frac{1}{2} - \cos\left(\frac{1}{3}\left(Arc\cos(1-2v) - 2\pi\right)\right)\right]^2 \left(\frac{1}{2} + \cos\left(\frac{1}{3}\left(Arc\cos(1-2u) - 2\pi\right)\right)\right)^2. \tag{11}$$

An expression for $0 < u \le v < 1$ has the same view with the replacement of *u* by *v* and on the contrary. Now, our two-dimensional cumulative distribution function can be presented as

$$F(x,y) = C(G(x), G(y)) \quad \forall x, y. \tag{12}$$

But we can use arbitrary marginal distribution $G_1(x), G_2(y)$ as an argument of the copula, receiving Frechet family of distributions $\{F(x, y)\}$:

$$F(x,y) = C(G_1(x), G_2(y)) \quad \forall x, y. \tag{13}$$

6. Example. Let us find a distribution for a sum of two consecutive terms of our stationary Markov chain: $H(x) = P\{X + Y \le x\}, \quad x > 0$. Obviously

$$H(x) = \begin{cases} \int_0^x g(\theta) F_\theta(x-\theta) d\theta, & 0 \le x \le 1, \\ G(x-1) + \int_{x-1}^1 g(\theta) F_\theta(x-\theta) d\theta, & 1 \le x \le 2. \end{cases} \tag{14}$$

We wish to compare it with a case when random variables X and Y are independent and have distribution (5), (6). Here,

$$H^*(x) = \int_0^x g(\theta) G(x-\theta) d\theta = \int_0^x 6\theta(1-\theta)\left[(x-\theta)^2 (3 - 2(x-\theta)\right] d\theta, \quad x \ge 0.$$

Table 1 contains the corresponding values. A comparison of the presented results shows that the dependence between terms of the sum influences remarkably on the sum's distribution function.

Table 1

Cumulative distribution functions $H(x)$ and $H^*(x)$ of the sum $X + Y$

x	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
$H(x)$	0.005	0.029	0.092	0.196	0.417	0.709	0.916	0.993	1.000
$H^*(x)$	0.027	0.110	0.274	0.500	0.726	0.890	0.973	0.998	1.000

Now we consider a distribution from Frechet family (13) when marginal cumulative distribution function $U(x)$ is uniformly distributed on the interval $\left(0.5 - \frac{1}{2}\sqrt{3/5}, 0.5 + \frac{1}{2}\sqrt{3/5}\right)$. Such a distribution has the same values of the expectation and the variance. In this case a two-dimensional cumulative distribution function is

$$\widetilde{H}(x,y) = C(U(x), U(y)).$$

Some function values are presented in Table 2. We can see again a difference between the compared cases.

Table 2

Cumulative distribution functions $F(x, y)$ and $\widetilde{H}(x,y)$

x	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$F(x, 0.5)$	0.093	0.192	0.309	0.432	0.551	0.652	0.725	0.769	0.784
$\widetilde{H}(x, 0.5)$	0.099	0.210	0.318	0.423	0.523	0.616	0.698	0.758	0.758
$F(x, 0.7)$	0.074	0.149	0.232	0.313	0.380	0.432	0.470	0.492	0.500
$\widetilde{H}(x, 0.7)$	0.080	0.165	0.243	0.313	0.372	0.423	0.467	0.500	0.500

References

1. F. Abegaz, U.V. Naik-Nimbalkar. Modelling statistical dependence of Markov chains via copula models. *Journal of statistical planning and inference*, 138 (4) (2008), 1131-1146 pp.
2. P.Embrechts, F. Lindskog, A. McNeil. Modelling Dependence with Copulas and Applications to Risk Management (Chapter 8). *Handbook of Heavy Tailed Distributions in Finance, Amsterdam*, Elsevier (2003), 329 – 384 pp.
3. F. Hernandez-Campos, K.F. Jeffay, C. Park, J.S. Marron, S.I. Resnick. Extremal dependence: Internet traffic applications. *Stochastic Models*, 22(1) (2005), 1-35 pp.
4. B. R. Nelsen. An Introduction to Copulas. Second Edition. New York, Springer, 2006. 270 p.
5. Pacheco, L.C. Tang, U N. Prabhu. Markov-Modulated processes and Semiregenerative Phenomena. New Jersey - London – Singapore, World Scientific (2009), 224 p.
6. Sklar, A. Fonctions de repartition a *n* dimensions et leurs marges. *Publ. Inst Statist. Univ Paris 8,* (1959). 229 – 231 pp.
7. L. Spizzichino. Ageing and positive dependence. *Encyclopedia of Statistics for Quality and Reliability.* F. Ruggeri, R. Kennett, and F.W. Faltin (Eds). Chichester, Wiley & Son Limited (2007), 82 – 95 pp.