7659, England, UK
Journal of Information and Computing Science
Vol. 2, No. 3, 2007, *pp*. 191-196

Generating Multivariate Nonnormal Distribution Random Numbers Based on Copula Function

Xiaoping Hu [+] , Jianmin He and Hongsheng Ly

School of Economics and Management, Southeast University, Nanjing 210096

(*Received 30 June 2006, accepted 4 August 4 2006*)

Abstract. Random numbers of multivariate nonnormal distribution are strongly requested by the area of theoretic research and application in practice. A new algorithm of generating multivariate nonnormal distribution random numbers is given based on the Copula function, and theoretic analysis suggests that the algorithm is suitable to be feasible. Furthermore, simulation shows that the empirical distribution which is formed by random numbers generating from the proposed algorithm can well approach the original distribution.

Keywords: Multivariate Nonnormal Distribution,Random Number, Copula,Algorithm

1. Introduction

There are only a few methods of generating random numbers from multivariate nonnormal distribution, and such random numbers are strongly requested by the area of multivariate analysis and statistical modeling.

Nagahara(2004) stated that the Pearson distribution system could represent wide class of distributions with various skewness and kurtosis. The Pearson system included some well-known distributions, for example, gamma, beta, t-distribution, etc. Generating random numbers from the Pearson distribution system was given in that paper.

In this paper, a new method of generating multivariate nonnormal distribution random numbers is proposed based on the Copula function. Firstly, every marginal distribution is obtained from the multivariate distribution. Secondly, the copula function of the multivariate distribution is gained according to the Sklar's theory. Thirdly, uniform distribution random number vector between 0 and 1 are generated using the Bayesian conditional probability formula. lastly, every component of the uniform distribution random number vector obtained in 3ith step is transformed by corresponding to the quasi-inverses function of the marginal distribution function . Cogent theoretical analysis shows that the proposed generator is suitable to be a reliable and efficient multivariate nonnormal distribution random numbers generator which can be used widely in multivariate analysis and statistical modeling.

The outline of the paper is as follows. A brief review of Copula and Notion are introduced in Section 2. A algorithm of generating multivariate nonnormal distributions by using the Copula method is shown in Section 3.The simulation for a concrete example are applied in Section 4. The conclusion is shown in Section 5.

2. A Brief Review of Copulas and Notion

Nelsen defines copulas as "functions that join or couple multivariate distribution functions to their one-dimensional marginal distribution functions" (Nelsen, 1999, page 5). Copulas contain all the information about the dependence structure of a vector of random variables. They can capture nonlinear dependence among random variables, while correlation is only a linear measure of dependence. In particular, copulas contain information about the joint behavior of the random variables in the tails of the distribution, which should be of primary interest in a study of contagion of financial crises. Moreover, copulas are able to capture tail behavior without the need of using discretion to define extreme outcomes.

We now assume that we are using the increasing function definition of a Copula, and the relationship

[+] Corresponding author. Tel.: +86-025-8379 3035; fax: +86-025-8379 3035.
E-mail address:hxpnj@163.com.

between the Copula and joint probability distribution function can be described by theorem 1 and 2.

Theorem 1. (Sklar,1959): Let $F(z_1, z_2, ..., z_n)$ be the joint distribution with margins $F_i(z_i)$, and let $F_i^{-1}(U_i)$ be quasi-inverses[1], then there exists a copula function

$$C(U_1, U_2, ..., U_n) = F(F_1^{-1}(U_1), F_2^{-1}(U_2), ..., F_n^{-1}(U_n)) \tag{1}$$

If the F_i is continuous, then C is unique.

If the F_i is not continuous, there are some technicalities that relate to what are called sub-copulas and the range of the corresponding variables.

If Copula and marginal distribution functions are known, the multivariate joint probability distribution function will be solved by theorem 2.

Theorem 2 (Sklar,1959): Let $C(U_1, U_2, ..., U_n)$ be a Copula ,and assume that $F_i(z_i)$ are distribution functions. Then there exists a joint distribution function $F(z_1, z_2, ..., z_n)$ given by

$$F(z_1, z_2, ..., z_n) = C(F_1(z_1), F_2(z_2), ..., F_n(z_n)) \tag{2}$$

and the $F_i(z_i)$ are the marginal distribution functions.

Therefore, if F is a continuous multivariate distribution function, Sklar's theorem suggests that it is possible to separate the univariate margins from the dependence structure. The dependence structure is represented by the copula. This can be seen even more clearly if we assume the F_i 's are differentiable, and C and F are n-times differentiable. Then, deriving both sides to get the density of F, we get:

$$\frac{\partial^n F(z_1, z_2, ..., z_n)}{\partial z_1 \partial z_2 ..., \partial z_n} = \frac{\partial^n C(U_1, U_2, ..., U_n)}{\partial U_1 \partial U_2 ..., \partial U_n} \frac{\partial U_1}{\partial z_1} \cdots \frac{\partial U_n}{\partial z_n} \tag{3}$$

where $U_i = F_i(z_i)(i = 1, 2, ..., n)$, that is, the density of F has been expressed as the product of the copula density and the univariate marginal densities. In this sense, we state that the copula has all the information about the dependence structure. Consequently, a copula is in essence a multivariate distribution whose marginal distributions are $U(0,1)$, which is uniform distribution on the interval (0, 1). Copulas allow one to model the marginal distributions and the dependence structure of a multivariate random variable separately. For more discussions on the theory of copulas and specific examples of copulas, see Nelsen (1998).

3. Algorithm for generating random number

Now, we give the idea, that is, how the random number vector $(x_1, \ldots, x_n)$ is generated by the Cumulated Density Function $F(X_1, \ldots, X_n)$. Firstly, obtain the marginal distribution F_i of a variable X_i from $F(X_1, \ldots, X_n)$. Secondly, get the Copula function $C(U_1, \ldots, U_n)$ according to the theorem 2, and then generate from Copula function $C(U_1, \ldots, U_n)$ random number vector $(u_1, \ldots, u_n)$ whose marginal distribution follows $U(0,1)$. Finally, inversely transform the marginal distribution F_i of a variable X_i, and gain $x_i = F_i^{-1}(u_i)$ .consequently, we have the random number vector $(x_1, \ldots, x_n)$ from the existed $F(X_1, \ldots, X_n)$.

In the above-mentioned analysis, the most significant disposal lies with generating random number vector $(u_1, \ldots, u_n)$ whose the marginal distribution follows $U(0,1)$ from Copula function $C(U_1, \ldots, U_n)$. The following is the algorithm to gain the random number vector $(x_1, \ldots, x_n)$ from the Cumulated Density Function $F(X_1, \ldots, X_n)$.

---

[1] The quasi-inverses definition of a function $f(x)$ is as follow: $x = \inf \{x \mid f(x) = y\}$.

Algorithm:

Step 1. Generate $U(0,1)$ random number U_1, let $i=2$

Step 2. Generate independent $U(0,1)$ random number p, then U_i is given by

$$U_i = F_{ICCDF,i}(p \mid U_1,...,U_{i-1})$$

where

$$F_{ICCDF,i}(p \mid U_1,...,U_{i-1}) = \int_0^{U_i} \frac{\partial^{i-1} C(U_1,\ldots,U_{i-1},U,1,\ldots,1)}{\partial U_1,\ldots,\partial U_{i-1}\partial U} \bullet \frac{1}{f_m(U_1,\ldots,U_{i-1})} dU$$

$$= \frac{\partial^{i-1} C(U_1,\ldots,U_{i-1},U_i,1,\ldots,1)}{\partial U_1,\ldots,\partial U_{i-1}} \bullet \frac{1}{f_m(U_1,\ldots,U_{i-1})}$$

$$f_m(U_1,\ldots,U_{i-1}) = \frac{\partial^{i-1} C(U_1,\ldots,U_{i-1},1,\ldots,1)}{\partial U_1,\ldots,\partial U_{i-1}}$$

Step 3. Let $i=i+1$. if $i <= n$, go to step 2; else ,stop.

Theorem 3. The random number vector $U = \left(U_1,U_2,...,U_n\right)$ given by the aforementioned algorithm follows the joint probability distribution $C(U_1,U_2,...,U_n)$.

Proof: For $n=1$, by all appearance, the theory is valid.

For $n=2$, variables U_1 and p follow $U(0,1)$,

$$p = F_{ICCDF,2} = \frac{\partial C(U_1,U_2)}{\partial U_1} \bullet \frac{1}{f_m(U_1)} = \frac{\partial C(U_1,U_2)}{\partial U_1}$$

So get the conditional density function of U_2: $f\left(U_2 \mid U_1\right) = \frac{\partial^2 C(U_1,U_2)}{\partial U_1 \partial U_2}$

The two dimensions joint density function of U_1 and U_2 is

$$f(U_1,U_2) = f\left(U_2 \mid U_1\right) f(U_1) = \frac{\partial^2 C(U_1,U_2)}{\partial U_1 \partial U_2} \bullet 1 = \frac{\partial^2 C(U_1,U_2)}{\partial U_1 \partial U_2}$$

Consequently, (U_1,U_2) follows $C(U_1,U_2)$.

For $n=k$, suppose $(U_1,\ldots,U_k)$ follows $C(U_1,\ldots,U_k)$

For $n=k+1$,

$(U_1,\ldots,U_k)$ follows $C(U_1,\ldots,U_k)$ and p follows $U(0,1)$

$$p = F_{ICCDF,k+1} = \frac{\partial^k C(U_1,\ldots,U_k,U_{k+1})}{\partial U_1,\ldots,\partial U_k} \bullet \frac{1}{f_m(U_1,\ldots,U_k)}$$

where

$$f_m(U_1,\ldots,U_k) = \frac{\partial^k C(U_1,\ldots,U_k,1)}{\partial U_1,\ldots,\partial U_k}$$

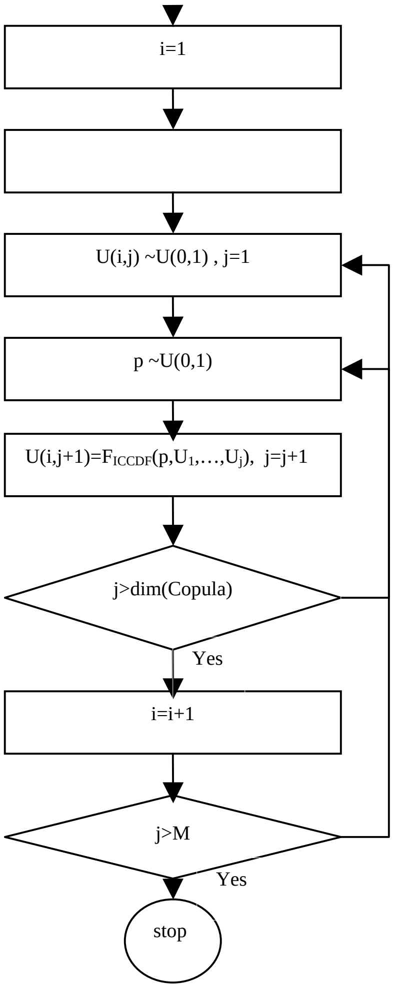
$$f(U_{k+1} \mid U_1,\ldots,U_k) = \frac{\partial^{k+1} C(U_1,\ldots,U_k,U_{k+1})}{\partial U_1,\ldots,\partial U_{k+1}} \bullet \frac{1}{f_m(U_1,\ldots,U_k)}$$

$$f(U_1,\ldots,U_{k+1}) = f(U_{k+1}\,|\,U_1,\ldots,U_k)\bullet f_m(U_1,\ldots,U_k)$$
$$= \frac{\partial^{k+1}C(U_1,\ldots,U_k,U_{k+1})}{\partial U_1,\ldots,\partial U_{k+1}}\bullet\frac{1}{f_m(U_1,\ldots,U_k)}\bullet f_m(U_1,\ldots,U_k)$$
$$= \frac{\partial^{k}C(U_1,\ldots,U_k,U_{k+1})}{\partial U_1,\ldots,\partial U_{k+1}}$$

Consequently, random number vector $(U_1,\ldots,U_{k+1})$ follows $C(U_1,\ldots,U_{k+1})$ .According to the principle of induction, the theorem is valid. □

If Copula belongs to the Elliptical Copula family, for example, normal (or t) Copula, multivariate random numbers may be generated by other simpler method as well. Random vectors from these copulas can be generated by creating random vectors from the multivariate Elliptical distribution, then transforming them to uniform marginal using the Elliptical CDF.

A block diagram of the proposed procedure generating M groups of random number vectors is shown in Figure 1.



procedure block diagram

4. Simulation

In this section, we use Clayton Copula given by Nelsen (1998) to check computations as a concrete application example. The expression of Clayton Copula is

$$\begin{cases} C(U,V) = (U^{-\alpha} + V^{-\alpha} - 1)^{-\frac{1}{\alpha}} \\ \alpha \in [-1,\infty]\setminus\{0\} \end{cases}$$

In this simulation, we endue $\alpha = 2$, and define a new residua function:

$$Z(U,V) = \left| C(U,V) - F_{emp}(U,V) \right|$$

where $C(U,V)$ is theoretic distribution function, and $F_{emp}(U,V)$ is empirical distribution function. According to theorem (Glivenko-Cantenlli), when the quantity of observed samples is of greatness, the empirical distribution function $F_{emp}(U,V)$ can well approximate the real distribution function. Now, we set the parameters. Simulation numbers are 10000, and the square $[0,1]\times[0,1]$ is partitioned to the grids by step 0.05. The figure 2 is the residual figure.

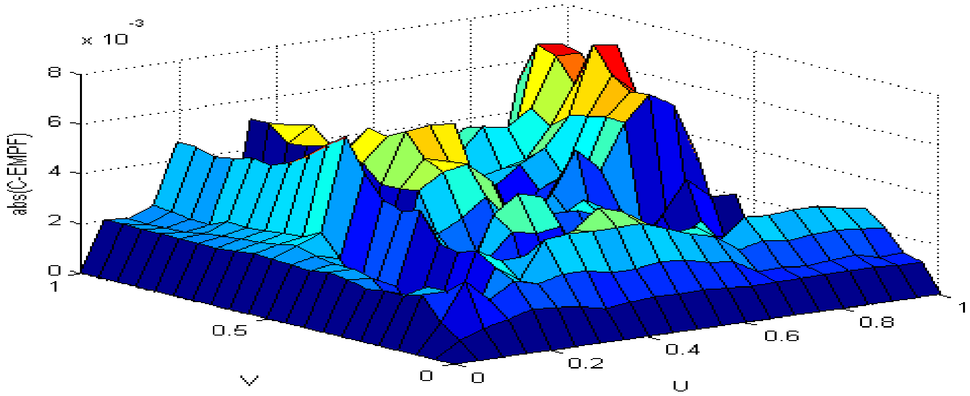


Figure 2 residual figure

In figure 2, the mean and maximal residual value is respectively the lesser value: 0.0022, 0.074. It is shown that the theoretic distribution function $C(U,V)$ can be well approached by the empirical distribution function $F_{emp}(U,V)$. For the scenarios of different parameter value of α and some of other Copulas, we achieve the similar result. So, random numbers generated by the proposed algorithm are assuredly sampled from the theoretic distribution function $C(U,V)$.

If marginal distributions don't follow $U(0,1)$, for example, follow gamma distribution, it is necessary to inversely transform random numbers generated by the Copula in the Algorithm. Transform form is given by

$$x_i = inv\ GammaCDF(u_i)$$

5. Conclusion

In this paper we have proposed a Copula-based algorithm to generate random numbers from multivariate nonnormal distribution. The Copula approach allowed us to construct algorithm by two stages. At one stage, generate random numbers with marginal distribution $U(0,1)$ from Copula function corresponding to cumulative distribution function. At the other stage, transform the random numbers from the first stage by implementing the quasi-inverse function of marginal distribution. Theoretic proof suggests that the proposed algorithm is suitable to be reliable. Furthermore, simulation shows that the empirical distribution which is formed by random numbers generating from the proposed algorithm can well approach the original distribution. In conclusion, the copula-based algorithm is found to perform well in generating random numbers from multivariate nonnormal distribution.

6. Acknowledgements

Hu Xiaoping acknowledges the support of The National Natural Science Foundation of China, grant N70371035. The author would also like to thank Academician Wun Quan and colleagues at the school of economic &management ,southeast university for their interest to the problems studied here and for their helpful comments. The author thanks also Professor Wu Guangmo for interesting discussions and the reviewers for their suggestions that allow to improve the paper.

7. References

[1] R. B. Nelsen. *An Introduction to Copulas*. NewYork: Springer-Verlag. 1998.

[2] E. W. Frees, & E. Valdez. Understanding relationships using copulas. *North American Actuarial Journal*. 2002, **2**(1): 1-25.

[3] Y. Nagahara. A method of simulating multivariate nonnormal distributions by the Pearson distribution system and estimation. *Computational Statistics & Data Analysis*.