Journal of Uncertain Systems
Vol.7, No.2, pp.83-91, 2013
Online at: www.jus.org.uk

Modeling and Generating Multi-Variate-Attribute Random Vectors Using a New Simulation Method Combined with NORTA Algorithm

M.R. Niavarani*, Alan J.R. Smith
Department of Mechanical Engineering, The University of Melbourne, Melbourne, Australia

Received 12 September 2012; Revised 12 November 2012

Abstract

The NORmal-To-Anything (NORTA) algorithm requires a correlation matrix of multivariate normal variables to convert a multivariate normal vector to any other distribution. This paper presents a new simulation method that works in combination with the NORTA algorithm yet avoids having to solve some complicated equations which need to be solved to achieve this matrix. The performance of the proposed method is investigated in three examples and the results indicate that the proposed method works well in terms of closeness of the correlation matrix calculated for generated random vector with the desired input correlation matrix.

Keywords: multi-variate-attribute random vector, NORTA, simulation, correlation matrix, multi-variate-attribute distributions

1 Introduction and Literature Review

There are many areas that researchers, in their simulation studies, need to generate correlated multi-variate-attribute random vectors including both continuous and discrete distributions. For instance, In order to undertake a simulation study on human behavioral and physical characteristics, some of the variables and attribute characteristics are correlated and needs to be generated as a correlated multi-variate-attribute vector. Similarly, in a simulation study on monitoring quality of a product or service, one deals with some correlated continuous and discrete quality characteristics. It this case, generation of correlated random variables including both continuous and discrete distributions seems desirable.

There are some methods in multi-variate-attribute domains to generate random vectors with dependent components and marginal distributions. Niaki and Abbasi [18] placed random vector generation methods that have been proposed for multi-variate-attribute cases into three categories: Analytical approaches that employ conditional distributions, Numerical procedures using acceptance/rejection methods, and Simulation approaches that apply partially specified properties like a set of marginal distributions and a correlation matrix in transformation procedures but not full joint distribution for the random vector.

Amongst the analytical approaches, Johnson's [10] proposed method to generate a p-dimensional random vector $\mathbf{X}=\left[x_1,x_2,...,x_p\right]^T$ by using a marginal distribution and a conditional distribution. He generates x_1 from cumulative marginal distribution F_1 and afterward generates $x_p, p=2,...,p$ from the cumulative conditional distribution $F\left(x_p \mid x_1,x_2,...,x_{p-1}\right)$. This approach may be difficult to apply because cumulative conditional distributions usually are not easy to derive and also the joint distribution of the random vectors is not often known [18]. Johnson [10] also has developed a method for generating multivariate lognormal distributions. As a more general method, Johnson [10] proposed a multivariate Johnson generation procedure but it was limited to a continuous multivariate random vector. Moreover, if some of the empirical marginal distributions of the original data (or the corresponding underlying

* Corresponding author. Email: m.rezazadehniavarani@pgrd.unimelb.edu.au (M.R. Niavarani).

theoretical marginals) possess marked skewness, then the correlation matrix of the fitted multivariate Johnson distribution will not match the sample correlation matrix of the original data set (Stanfield, [26]). Stanfield [26] proposed a modified multivariate input modeling based on the Johnson model but the above-mentioned limitations are also the case for his method. There are some other analytical methods that are limited to generating data for specific bivariate or multivariate distributions. Moonan [17] proposed a method for generating normal random vectors based on the linear transformation of a set of independent standard normal random variables. There have been a couple of studies on multivariate Gamma distribution: Ronning [21], Schmeiser and Lal [22] and Lewis [12] have proposed generation models for the nonnegative correlated multivariate gamma, bivariate gamma, and negatively correlated multivariate gamma distributions, respectively. A multivariate log-normal distribution data generation method has been proposed by Press [20] for when only positive random variables are required. Parrish [19] has proposed another multivariate generation method; this one is restricted to Pearson distributions and is claimed to work approximately with those non-Pearson multivariate distributions which have known joint moments.

Johnson [10] has also reported a numerical procedure that is similar to that of Devroye [3]. In this procedure, a joint distribution is used to generate random vectors by the acceptance/rejection approach. In these methods, one of the joint distributions that dominate the original joint probability density function is selected. Then a random vector is generated based on the selected joint distribution. Finally, this random vector is accepted or rejected based on using a distribution with uniform centre and exponential tails. It has been reported that, acceptance–rejection methods typically become inefficient as the number of variables of the problem increases [26, 7, 8, 13] suggested a transformed density rejection method to construct the dominating density function. Using a joint distribution as an input for this problem is a major constraint for the proposed method in this category because it does not happen often in practice [18].

Other than the above-mentioned limitations and difficulties of the aforementioned analytical and numerical approaches, all of the proposed methods are only utilizable to generate random vectors whose variables have a similar distribution and most of them have constraints on the size of the random vector that can be generated [18].

With the simulation approaches, full joint distribution for the random vector is not required but a set of marginal distributions and a correlation matrix (as the partial specifications) are used to generate a multivariate random vector. The NORmal-To Anything (NORTA) transformation detailed by Cairo and Nelson [1], which is based on work by Marida [16] and Li and Hammond [14], is an example of the research in this category. The NORTA transformation can be found in “partially specified” approaches to random vector generation [26]. Song and Hsiao [23] and Song et al. [24] have applied a similar concept to generate time series random variables [18].

The NORTA approach has a number of desirable attributes; however, there are also some practical limitations to the application of the algorithm. These issues are discussed in some detail in the next section.

In this paper, a new simulation method is presented that works in combination with the NORTA algorithm for generating multi-variate-attribute random vectors, but overcomes the difficulties identified with the present implementations of the algorithm. Section 2 discusses the existing literature concerning the past developments involving the NORTA algorithm and refers all previous researches in this area. In Section 3, the new proposed method is developed and its performance is evaluated in using three examples in the following section.

2 NORmal-To-Anything (NORTA) Literature Review and Algorithm

The NORTA transformation demonstrates that samples can be obtained from the partially specified distribution by transforming the elements of a sample from a multivariate standard normal distribution according to the appropriate desired marginal distribution, where the correlations of the elements of the deriving normally distributed random vector are set to generate the desired correlations in the transformed random vector. Setting the correlation matrix appropriately amounts to solving a number of one-dimensional root finding problems corresponding to each desired pair-wise correlation value, each of which can be solved by bisection. Solving some complex nonlinear system of equations is the most important problem in partially specified approaches.

In the NORTA algorithm, firstly a p-dimensional standard normal random vector is generated and then transformed into a vector of uniform random vectors. After this, by the inverse transformation technique, the uniform random vector is transformed back into a random vector with the target marginals. The initialization step in the NORTA algorithm requires finding the correlation matrix of the normal random vector such that it guarantees the last random vectors generated have a specific correlation matrix [10, 17, 11]. To reach this goal a number of one-dimensional simultaneous equations must be solved, which usually is difficult to do analytically. Moreover, there are two complications in NORTA algorithm. The first is that a desired pair-wise correlation may not be feasible; that is, there may not exist a pair-wise correlation for the driving multivariate normal random vector such that the corresponding transformed elements have the desired correlation value. The second, and more serious restriction, is that even if all desired correlation values are feasible, the full correlation matrix for the driving multivariate normal random vector may not be positive. This becomes an issue where the dimensionality of the random vector increases[18].

In order to determine when the first complication might arise, Ghosh and Henderson [6] developed a computational approach for establishing whether a given covariance matrix is feasible for a given set of marginals in the initialization step in the NORTA algorithm. They mentioned that in cases where NORTA cannot precisely match a feasible covariance matrix, it is still possible to use NORTA to obtain the desired marginals exactly, but that the desired covariance matrix is only approximated. In their proposed method a potentially large linear program (LP) needs to be solved and it increases with the size of the problem. Later, they investigated the behavior of the NORTA method in having more variables and proposed a procedure to modify NORTA that can be used more generally for sampling uniformly from the space of all asymmetric positive definite matrices with the diagonal fixed at some positive value. They proposed an augmented NORTA as a semi-definite program (SDP) to be solved in the generation step of NORTA as a modification. They concluded that despite the feasibility problem, the NORTA method is a viable method even in high-dimensional problems. Furthermore, computational results show that the SDP problem in the SDP-augmented method is solved within a reasonable amount of time for dimensions less than 10.

In order to resolve the second issue, Lurie and Goldberg [15] proposed a method in which the correlation matrix is first checked for mathematical consistency (positive semi-definiteness), and adjusted if necessary to have the closest positive semi-definite correlation matrix. Then the correlated random vectors are generated using a combination of Cholesky decomposition and Gauss-Newton iteration. However, solving this problem is potentially computationally expensive; additionally, random vectors generated with the resulting NORTA transform will not have the desired correlation structure [26].

Stanhope [26] developed a multivariate-to-anything transform in place of NORTA for partially specified random vector generation and proposed several alternatives to the NORTA algorithm in some specific cases. As the first alternative, he proposed bilateral bivariate Pareto-to-anything (BBPTA) transforms as a limited method for modeling bivariate partial specifications. As another alternative, Dirichlet-to-anything (DIRTA) transform were proposed. One unattractive aspect of using either the BBPTA or DIRTA transform for partially specified random vector generation is that only partial specifications with correlation matrices having negative pair-wise correlation values can be modeled. To loosen this restriction, he proposed a multivariate-t-to-anything (TTA) transformation. However, for the stated goals of partially specified random number generation, a low value of n (degree of freedom) is useful in order to maximize the difference between the TTA and NORTA transforms. So, a considerable difference between their outcomes is not expected particularly in high dimension cases.

In following section, the NORTA algorithm is explained in detail with the same notation as used in Niaki and Abbasi's paper [18]. The goal of the NORTA algorithm is to generate a p-dimensional random vector $\mathbf{X}=\left[x_1,x_2,...,x_p\right]^T$ with the following properties: $x_i \sim F_{x_i}, i=1,2,...,p$, where F_{x_i} is an arbitrary cumulative distribution function (cdf) and $Corr\left[X\right]=\boldsymbol{\Sigma_X}$, where $\boldsymbol{\Sigma_X}$ is given.

The vector X is generated by a transformation of a p-dimensional standard multivariate normal (MVN) vector $\mathbf{Z}=\left[z_1,z_2,...,z_p\right]^T$ with correlation matrix $\boldsymbol{\Sigma_Z}$ by using transformation equation (1).

$$\mathrm{X}=(F_{x_1}^{-1}[\Phi(z_1)],F_{x_2}^{-1}[\Phi(z_2)],\ldots,F_{x_p}^{-1}[\Phi(z_p)])^T \tag{1}$$

where $\boldsymbol{\Phi}$ is the univariate standard normal cdf and $F_x^{-1}(u)\equiv\inf\{x{:}F_x(x)\geq u\}$ denotes the inverse cdf. It is noted that since the exact value of $F_x^{-1}(\Phi(z))$ for each distribution (especially for discrete distributions) may not be known, the infimum value of x that satisfies $F_x(x)\geq\Phi(z)$ is selected for x.

The transformation $F_{x_i}^{-1}(\Phi(\cdot))$ ensures that x_i has the desired cumulative marginal distribution F_{x_i}. To achieve this, the general problem is to select the correlation matrix $\boldsymbol{\Sigma_Z}$ that gives the desired correlation matrix $\boldsymbol{\Sigma_X}$ .Therefore, in the NORTA transformation process, a multivariate normal vector **Z** is transformed into a multivariate uniform vector **U** and then the multivariate uniform vector **U** into the desired vector **X**. So, the joint distribution of **U** is known as a copula and any joint distribution has a representation as a transformation of a copula (Niaki and Abbasi, 2008). The NORTA method is summarized in Figure 1.

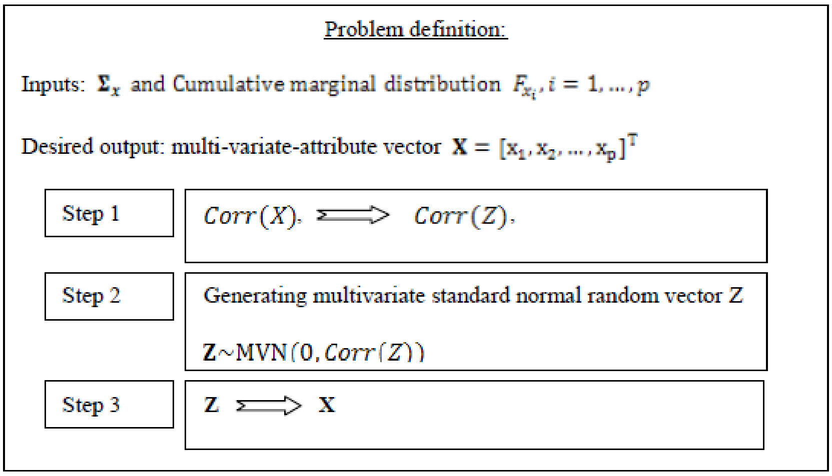

Figure 1: Generating a multi-variate-attribute random vector by using NORTA

Each element of Σ_Z , $\rho_z(i,j),\ i \neq j,$ shows the correlation between z_i and z_j . Similarly, $\rho_x(i,j),\ i \neq j,$ denotes the correlation between x_i and x_j in Σ_X . That is,

$$\rho_x(i,j) = Corr(x_i, x_j) = Corr\{F_{x_i}^{-1}[\Phi(z_i)], F_{x_j}^{-1}[\Phi(z_j)]\} \quad i \neq j\,. \tag{2}$$

Since

$$Corr(x_i, x_j) = \frac{E(x_i, x_j) - E(x_i)E(x_j)}{\sqrt{Var(x_i)Var(x_j)}}, \tag{3}$$

$E(x_i), E(x_j)$, $Var(x_i)$ and $Var(x_j)$ are fixed by F_{x_i} and F_{x_j} ; and (z_i, z_j) follows a standard bivariate normal distribution with correlation $Corr(z_i, z_j) = \rho_z(i,j)$, $E(x_i, x_j)$ in equation (3) can be calculated as,

$$E(x_i, x_j) = E\{F_{x_i}^{-1}[\Phi(z_i)]F_{x_j}^{-1}[\Phi(z_j)]\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F_{x_i}^{-1}[\Phi(z_i)]F_{x_j}^{-1}[\Phi(z_j)]\varphi_{\rho_z(i,j)} d_{z_i} d_{z_j} \tag{4}$$

where $\varphi_{\rho_z(i,j)} d_{z_i} d_{z_j}$ is the standard bivariate normal probability density function (pdf) with correlation $\rho_z(i,j)$. In equation (4), summation is applied rather than integration for the discrete distributions.

From equation (4), it is clear that the correlation between x_i and x_j is a function of the correlation between z_i and z_j . This function is denoted by $C_{ij}[\rho_z(i,j)]$. In other words, $\rho_x(i,j) = C_{ij}[\rho_z(i,j)]$.

So, in order to generate a *p*-dimensional random vector using the NORTA algorithm, equation (4) needs to be solved for each pair of the variables. Hence, *p(p −1)/2* complicated equations need to be solved that for many cumulative marginal distributions are usually unsolvable by analytical methods.

Regarding having Σ_X as one of the two inputs (in addition to the marginal distribution), Σ_Z is obtainable. To generate Σ_Z from Σ_X (step 1), Niaki and Abbasi [18] identified four generic approaches that might be followed. Each of these approaches has its own area of application and drawbacks:

1) The analytical approach has been found to work for some special cases such as uniform random vectors. It is difficult to apply because the joint distributions and consequently the conditional distributions are not easy to obtain in most cases. Chen [2] and Hull [9] have applied conditional distribution (assumed known) to solve equation (4) for multivariate normal distributions.

2) In the numerical approach, one employs numerical root finding methods to solve *p(p −1)/2* equations. In this approach, the double integral (summation for discrete distributions) function values of the form (4) are evaluated by numerical integration methods (summation for discrete distributions). Li and Hammond [14] and Cario and Nelson[1] used Newton’s method and Yen [27] applied the efficient Gaussian-Quadrature integration and Newton’s method. In these methods, the computational time increases quadratically with *p*. Furthermore, the equispaced integration methods may be inefficient, and the Gaussian quadrature methods may be inaccurate [2].

3) In the simulation approach, for any set of the root candidates, first the NORTA algorithm is applied to generate *m* random vectors. Then, the correlations of the *m* generated observations are calculated and checked to identify they have reached the required correlation matrix. Chen [2] employed this approach to solve the $p(p-1)/2$ equations in (4) by treating it as a stochastic root-finding problem, solving equations using only the estimates of the function values. Yen [27] mentioned that the disadvantage of this approach is that the computation time is usually longer than the numerical approach.

4) In the most recent approach, Niaki and Abbasi [18], have proposed the concept of function fitting to generate $\boldsymbol{\Sigma}_Z$ using artificial neural networks(ANN). In order to generate $\rho_z(i,j)$ for each pair of random variables in matrix $\boldsymbol{\Sigma}_Z$ they employed an ANN that first must be trained with marginal distributions of the random variables and then be used to generate the $\rho_z(i,j)$ values between each pair of the variables. It needs to employ one network for each pair of the variables that have different marginal distributions. In other words, similar to the number of equations (4) that needs to be solved in the numerical method, in ANN $p(p-1)/2$ networks are required for a p-dimensional problem with different marginal distributions. It is argued that it is not always easy to employ and train $p(p-1)/2$ networks. For example, in a 5 dimensional random vectors with different marginal distributions, 10 separate networks are required. Niaki and Abbasi showed that the ANN method works better than Newton's method as applied by Cario and Nelson[1] in terms of the sum of squared errors (SSE) of the difference between Σ_x, the known input of the problem, and $\overline{\boldsymbol{\Sigma}_x}$, calculated by the random vector generated by NORTA algorithm. However this slightly better result was not considered sufficient to justify the use of this approach given its complexity and the general drawbacks in applying ANN identified by Friedman et al. [4].

Five issues concerning ANN were identified, the first of which is the trial-and-error approach to building models during the training phase. Another concern was that ANN techniques do not provide an understanding of how the system works with only the inputs and outputs being observed. Over-fitting of the iterative model on the training dataset is another issue. As another concern, model performance is not independent of the starting input values and parameters. The fifth issue was that variable scaling can be considered as a limitation where the model fits better if all variables are of a similar scale.

3 Proposed Method

In this method an initial starting matrix $\boldsymbol{\Sigma}^s{}_z$ could be calculated using the desired input correlation matrix $\boldsymbol{\Sigma}^1{}_x$ and starting random matrix $\mathbf{D}^s$ $\mathbf{D}^s$ is a random diagonal matrix with the same dimension as $\boldsymbol{\Sigma}^1{}_x$, in which all the elements on the main diagonal are equal to zero and all other elements are random numbers between $(-1-\rho^i{}_x(i,j), 1-\rho^i{}_x(i,j))$ where $\rho^i{}_x(i,j)$ are elements of $\boldsymbol{\Sigma}^1{}_x$.

If the $\boldsymbol{\Sigma}^s{}_z$ matrix is not positive-definite, a new matrix $\mathbf{D}$ is generated and used, otherwise the $\boldsymbol{\Sigma}^s{}_z$ matrix is used to generate the random vector $\mathbf{X}^s = [x^s{}_1, x^s{}_2, \ldots, x^s{}_p]^T$. $\overline{\boldsymbol{\Sigma}_x}$ is the correlation matrix calculated based on the generated random vector $\mathbf{X}^s$. In the other words, based on an iterative procedure, the generated $\mathbf{X}^s = [x^s{}_1, x^s{}_2, \ldots, x^s{}_p]^T$ and calculated $\overline{\boldsymbol{\Sigma}_x}$ are the starting random vector and correlation matrix for **X**, respectively.

Based on the value of the $\mathrm{SSE}_{\mathrm{target}}$, which is set by the user, if the generated $\overline{\boldsymbol{\Sigma}_x}$ is close enough to the given $\boldsymbol{\Sigma}^1{}_x$ (SSE value less than $\mathrm{SSE}_{\mathrm{target}}$), the generated vector $\mathbf{X}^s$ could be used as $\mathbf{X}$, otherwise another Σ_Z is calculated by using a different randomly generated **D**.

In order to have a clear view of the proposed method, a detailed step by step algorithm is shown in Figure 2.

There are three advantages in using this method over previous methods; namely,

The two general complications in NORTA algorithm have been resolved in this method. The first one is concerned with there being sometimes infeasible solutions to equation (4). This is overcome by setting a Σ_Z as an input by the user. It means that equation (4) does not need to be solved and consequently there is no possibility of not having feasible pair-wise correlation matrixes. The second issue is in regards to having a non-positive definite Σ_Z. The algorithm checks for this condition and if found.

Programming is considerably easier and computation time is pleasingly less than other methods.

The outcome is technically more appropriate than other methods in terms of closeness of generated Σ_x and input Σ_x and also smaller SSE (as shown and discussed in the case studies below.

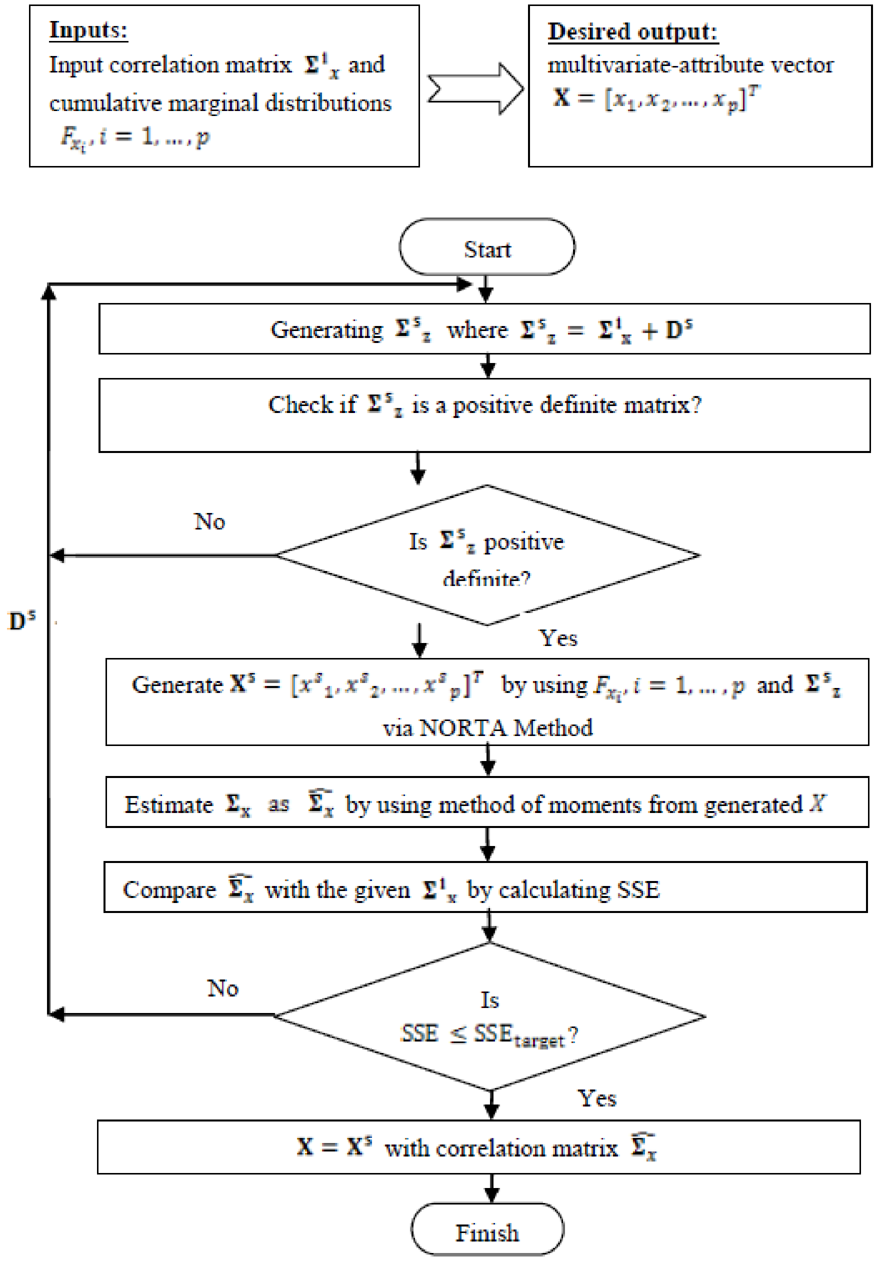

Figure 2: Combined simulation method in using NORTA algorithm for generating multi-variate-attribute random vector

4 Case Studies

Cario and Nelson [1] presented three numerical examples in which they obtained the correlation matrices needed for the NORTA algorithm by numerically solving equation (4) using Newton's Method. Niaki and Abbasi [18] compared the estimated correlation matrices from their ANN method to the ones generated in those examples.

In these case studies, MATLAB software is applied to generate random vectors using the new proposed method discussed above.

In order to do this, the correlation matrix of the normal random vector $\mathbf{D}^s{}_z$ is calculated using the generated matrix $\mathbf{D}^s$. Then, if it passes the positive-definite condition, to compare with the Newton and ANN methods in a consistent manner, it is used to generate 5000 random vectors with the NORTA algorithm. Finally, the correlation matrices are calculated and compared to results with the two other methods in terms of sum of squared error (SSE) to evaluate closeness of the generated correlation matrix $\overline{\boldsymbol{\Sigma}_x}$ and the desired input correlation matrix $\boldsymbol{\Sigma}^1{}_x$.

Case 1: As a multivariate case, all of the random variables of a four-dimensional random vector come from a Gamma distribution with parameters of $\alpha = 14.4$ and $= 0.03424$, in which the input correlation matrix is

$$\begin{bmatrix} 1 & 0.7 & 0.5 & -0.9 \\ 0.7 & 1 & 0.7 & -0.6 \\ 0.5 & 0.7 & 1 & -0.3 \\ -0.9 & -0.6 & -0.3 & 1 \end{bmatrix}.$$

The value of $\mathrm{SSE}_{\mathrm{target}}$ is set as 0.00035.

The generated correlation matrix based on the generated random vector **X** and corresponding SSE using the proposed method are

$$\begin{bmatrix} 1 & 0.7020 & 0.4984 & -0.8838 \\ 0.7020 & 1 & 0.7009 & -0.5979 \\ 0.4984 & 0.7009 & 1 & -0.3076 \\ -0.8838 & -0.5979 & -0.3076 & 1 \end{bmatrix}$$

and 0.000332, respectively.

Case 2: As a multi-attribute case, three-dimensional random vectors are generated in which the variables all share the same Binomial distribution with parameters $n = 3$ and $p = 0.5$ and correlation matrix

$$\begin{bmatrix} 1 & 0.2 & -0.8 \\ 0.2 & 1 & 0.2 \\ -0.8 & 0.2 & 1 \end{bmatrix}.$$

The value of $\mathrm{SSE}_{\mathrm{target}}$ is set as 0.00006.

The generated correlation matrix based on the generated random vector **X** and corresponding SSE using the proposed method are

$$\begin{bmatrix} 1 & 0.1980 & -0.7941 \\ 0.1980 & 1 & 0.2022 \\ -0.7941 & 0.2022 & 1 \end{bmatrix}$$

and 0.000043, respectively.

Case 3: As a multivariate-attribute case, a mixed random vector is considered containing the discrete random variable $X_1 \sim U(1,10)$ and continuous random vector $X_2 \sim Exponential(10)$ with the correlation matrix

$$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

The value of $\mathrm{SSE}_{\mathrm{target}}$ is set as 0.000003.

The generated correlation matrix based on the generated random vector X and corresponding SSE using the proposed method are

$$\begin{bmatrix} 1 & -0.5001 \\ -0.5001 & 1 \end{bmatrix}$$

and 0.000001, respectively.

Comparison of these results with those reported by Cario and Nelson [1] and Niaki and Abbasi [18] are shown in Tables 1 and 2. In Table 1, the second and the third columns show the marginal probability distributions of the variables in the random vectors and the input correlation matrices of the random vectors as desired correlation matrix, respectively. The fourth, fifth and sixth columns show the generated correlation matrices using the Newton, ANN and the proposed methods for implementing NORTA algorithm respectively. In Table 2, the second, third and fourth columns show the estimated correlation matrix of 5000 random vectors generated by the NORTA algorithm using the three methods. Finally, columns five to seven show the SSE value between the desired input correlation matrices and the ones generated by the three methods.

Table 1: Comparison of correlation matrices (based on Niaki and Abbasi [18])

case	Marginal Probability Distribution	Desired input correlation Matrix Σ^1_x	Standard multivariate normal correlation matrix Σ_z by ...		
			Newton method	ANN method	Proposed method
1	$X_i \sim \Gamma(14.4, 0.034424)$; $i = 1,2,3,4$	$\begin{bmatrix} 1 & 0.7 & 0.5 & -0.9 \\ 0.7 & 1 & 0.7 & -0.6 \\ 0.5 & 0.7 & 1 & -0.3 \\ -0.9 & -0.6 & -0.3 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.7040 & 0.5040 & -0.9200 \\ 0.7040 & 1 & 0.7040 & -0.6160 \\ 0.5040 & 0.7040 & 1 & -0.3040 \\ -0.9200 & -0.6160 & -0.3040 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.7008 & 0.5052 & -0.9289 \\ 0.7008 & 1 & 0.7008 & -0.6133 \\ 0.5052 & 0.7008 & 1 & -0.3016 \\ -0.9289 & -0.6133 & -0.3016 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.6934 & 0.4977 & -0.9054 \\ 0.6934 & 1 & 0.6946 & -0.6054 \\ 0.4977 & 0.6946 & 1 & -0.3054 \\ -0.9054 & -0.6054 & -0.3054 & 1 \end{bmatrix}$
2	$X_i \sim Binomial(3, 0.5)$; $i = 1,2,3$	$\begin{bmatrix} 1 & 0.2 & -0.8 \\ 0.2 & 1 & 0.2 \\ -0.8 & 0.2 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.2288 & -0.8960 \\ 0.2288 & 1 & 0.22288 \\ -0.8960 & 0.2288 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.2304 & -0.8981 \\ 0.2304 & 1 & 0.2304 \\ -0.8981 & 0.2304 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.2278 & -0.8880 \\ 0.2278 & 1 & 0.2226 \\ -0.8880 & 0.22226 & 1 \end{bmatrix}$
3	$X_i \sim Discrete\ U(1,10)$ $X_j \sim Exp.(10)$	$\begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -0.5760 \\ -0.5760 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -0.5719 \\ -0.5719 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -0.5710 \\ -0.5710 & 1 \end{bmatrix}$

Table 2: Comparison of results (based on Niaki and Abbasi [18])

$\widehat{\Sigma_x}$ of 5000 data sets generated using Σ_z in				SSE		
case	Newton method	ANN method	Proposed method	Newton method	ANN method	Proposed method
1	$\begin{bmatrix} 1 & 0.7087 & 0.5175 & -08902 \\ 0.7087 & 1 & 0.7072 & -0.6070 \\ 0.5175 & 0.7072 & 1 & -0.3151 \\ -0.8902 & -0.6070 & -0.3151 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.7005 & 0.5160 & -0.9008 \\ 0.7005 & 1 & 0.7013 & -0.6015 \\ 0.5160 & 0.7013 & 1 & -0.3118 \\ -0.9008 & -0.6015 & -0.3118 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.7020 & 0.4984 & -0.8838 \\ 0.7020 & 1 & 0.7009 & -0.5979 \\ 0.4984 & 0.7009 & 1 & -0.3076 \\ -0.8838 & -0.5979 & -0.3076 & 1 \end{bmatrix}$	0.000807	0.000400	0.000332
2	$\begin{bmatrix} 1 & 0.2116 & -0.7959 \\ 0.2116 & 1 & 0.1964 \\ -0.7959 & 0.1964 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.2061 & -0.8017 \\ 0.2061 & 1 & 0.1930 \\ -0.8017 & 0.1930 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.1980 & -0.7941 \\ 0.1980 & 1 & 0.2022 \\ -0.7941 & 0.2022 & 1 \end{bmatrix}$	0.000164	0.000080	0.000043
3	$\begin{bmatrix} 1 & -0.5037 \\ -0.5037 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -0.5024 \\ -0.5024 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -0.5001 \\ -0.5001 & 1 \end{bmatrix}$	0.000013	0.000005	0.000001

5 Conclusion

As can be seen in Tables 1 and 2, the proposed method works better than the methods proposed by both Cario and Nelson's [1] and Niaki and Abbasi [18] for generating the standard multivariate normal correlation matrix Σ_X and consequently using the NORTA algorithm for generating a multivariate-attribute random vector. In the proposed method, the correlation matrix, which is calculated based on a generated random vector, is closer to the desired input correlation matrix, as shown by the SSE value being less than for the other two methods.

In addition to the power of the proposed method in generating random vectors with the closest correlation matrix to the desired input one, the proposed method, unlike the Newton method, does not need to solve *p(p-1)/2* complicated double integral (summation) equations for each coefficient of the correlation matrix.

Furthermore, over and above avoiding the general disadvantages of ANN mentioned in section 3, the requirement of designing *p(p-1)/2* networks for each pair of variables that have different marginal distributions is not required in the proposed method.

This paper has presented an accurate and robust means of generating multi-variate-attribute vectors for use in subsequent analyses. It is believed that this proposed method is applicable to all forms of continuous and discrete distributions.

Acknowledgement

The authors gratefully acknowledge the helpful comments of the anonymous reviewers, which improved the content of the article.

References

[1] Cairo, M.C., and B.L. Nelson, Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix, *Technique Report*, Department of Industrial Engineering and Management Science, Northwestern University, Evanston, IL, 1997.

[2] Chen, H., Initialization for NORTA: generation of random vectors with specified marginal and correlations, *INFORMS Journal on Computing*, vol.13, pp.312–331, 2001.

[3] Devroye, L., *Non-uniform Random Variate Generation*, Springer-Verlag, New York, 1986.

[4] Friedman, J., T. Hastie, and R. Tibshirani, The element of statistical learning, *Data Mining, Inference and Prediction*, 2001.

[5] Ghosh, S., and S.G. Henderson, Chessboard distributions and random vectors with specified marginals and covariance matrix, *Operations Research*, vol.50, pp.820–834, 2002.

[6] Ghosh, S., and S.G. Henderson, Behavior of the NORTA method for correlated random vector generation as the dimension increases, *ACM Transactions on Modeling and Computer Simulation*, vol.13, pp.1–19, 2003.

[7] Gilks, W.R., and P. Wild, Adaptive rejection sampling for Gibbs sampling, *Applied Statistics*, vol.41, pp.337–348, 1992.

[8] Hörmann, W., A technique for sampling from T-concave distributions, *ACM Transactions on Mathematical Software*, pp.182–193, 1995.

[9] Hull, J.C., Dealing with dependence in risk simulations, *Operational Research Quarterly*, vol.28, pp.201–213, 1987.

[10] Johnson, M.E., *Multivariate Statistical Simulation*, Wiley, New York, 1987.

[11] Law, A.M., and W.D. Kelton, *Simulation Modeling and Analysis*, McGraw-Hill, New York, 2000.

[12] Lewis, P.A.W., Generating negatively correlated gamma variables using the beta-gamma transformation, *Proceedings of the Winter Simulation Conference*, Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp.175–176, 1983.

[13] Leydold, J., A rejection technique for sampling from log concave multivariate distributions, *ACM Transactions on Modeling and Computer Simulation*, vol.8, pp.254–280, 1998.

[14] Li, S.T., and J.L. Hammond, Generation of pseudorandom numbers with specified univariate distributions and correlation coefficients, *IEEE Transactions on Systems, Man and Cybernetics*, vol.5, pp.557–561, 1975.

[15] Lurie, P.M., and M.S. Goldberg, An approximate method for sampling correlated random variables from partially specified distributions, *Management Science*, vol.44, no.2, pp.203–218, 1998.

[16] Marida, K.V., A translation family of bivariate distributions and Fréchet's bounds, *Sankhya*, vol.32, pp.119–122, 1970.

[17] Moonan, W.J., Linear transformation to a set of stochastically dependent normal variables, *Journal of American Statistical Association*, vol.52, pp.247–252, 1957.

[18] Niaki, S.T.A., and B. Abbasi, Generating correlation matrices for normal random vectors in NORTA algorithm using artificial neural networks, *Journal of Uncertain Systems*, vol.2, no.3, pp.192–201, 2008.

[19] Parrish, R.S., Generating random deviates from multivariate Pearson distributions, *Computational Statistics & Data Analysis*, vol.9, pp.283–295, 1990.

[20] Press, S.J., *Applied Multivariate Analysis*, Holt, Rinehart & Winston, New York, 1972.

[21] Ronning, G., A simple scheme for generating multivariate gamma distributions with nonnegative covariance matrix, *Technometrics*, vol.19, pp.179–183, 1997.

[22] Schmeiser, B., and R. Lal, Bivariate gamma random vectors, *Operations Research*, vol.30, pp.355–374, 1982.

[23] Song, W.T., and L. Hsiao, Generating of auto correlated random variables with a specified marginal distributions, *Proceedings of the Winter Simulation Conference*, Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, 1993.

[24] Song, W.T., L. Hsiao, and Y. Chen, Generating pseudo-random time series with specified marginal distributions, *European Journal of Operational Research*, vol.94, pp.194–202, 1996.

[25] Stanfield, P.M., J.R. Wilson, G.A. Mirka, N.F. Glasscock, J.P. Psihogios, and J.R. Davis, Multivariate input modeling with Johnson distributions, *Proceedings of the Winter Simulation Conference*, Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp.1457–1464, 1996.

[26] Stanhope, S., Case studies in multivariate-to-anything transforms for partially specified random vector generation, *Insurance: Mathematics and Economics*, vol.37, pp.68–79, 2005.

[27] Yen, J.C., NORTA initialization for random vector generation by numerical methods, Masters of Industrial Engineering Thesis, Chung Yaun Chiristian University, Taiwan, 2001.