Modeling and Generating Random Vectors with Arbitrary Marginal Distributions and Correlation Matrix

Marne C. Cario
Delphi Packard Electric Systems
Warren, OH 44486, USA

Barry L. Nelson
Department of Industrial Engineering and Management Sciences
Northwestern University, Evanston, IL 60208, USA

April 9, 1997

Abstract

We describe a model for representing random vectors whose component random variables have arbitrary marginal distributions and correlation matrix, and describe how to generate data based upon this model for use in a stochastic simulation. The central idea is to transform a multivariate normal random vector into the desired random vector, so we refer to these vectors as having a NORTA (NORmal To Anything) distribution. NORTA vectors are most useful when the marginal distributions of the component random variables are neither identical nor from the same family of distributions, and they are particularly valuable when the dimension of the random vector is greater than two. Several numerical examples are provided.

Keywords: simulation, random vector, input modeling, correlation matrix, copulas

1 Introduction

In many stochastic simulations, simple input models—idependent and identically distributed sequences from standard probability distributions—are not faithful representations of the physical input process. For example, the processing times for a single product at a series of k machining stations may be dependent, due to characteristics of that particular product. Similarly, the service times for a single customer at the order desk, cashier and loading dock of a store may be dependent due to characteristics of the order. And the quantities of each of k items that a factory demands from a multi-item inventory system would typically be related. Ignoring such dependence can significantly distort the simulated performance of the system.

There are numerous models available for representing and generating random vectors with dependent components and marginal distributions from a common family. Excellent surveys can be found in Devroye [1986] and Johnson [1987]. However, when the component random variables have different marginal distributions, from different families, then there are few alternatives available.

In this paper we present a model for representing a $k \times 1$ random vector $\mathbf{X} = (X_1, X_2, \ldots, X_k)'$ with arbitrary marginal distributions and any feasible correlation matrix. We use a transformation-oriented approach to represent $\mathbf{X}$. This approach takes a random vector with a known correlation matrix, the base vector $\mathbf{Z}$, and transforms it to achieve the desired marginal distributions for the components of the input vector, $\mathbf{X}$. The target correlation matrix of $\mathbf{X}$ is obtained by adjusting the correlation matrix of the base vector. In our model, the base vector $\mathbf{Z}$ is a standard multivariate normal random vector, so we refer to $\mathbf{X}$ as having a NORTA (NORmal To Anything) distribution.

The idea of transforming multivariate normal vectors into vectors with other marginal distributions has long been folklore in statistics and simulation. The first reference appears to be Mardia [1970] who described transforming bivariate normal random variables. Li and Hammond [1975] discussed the extension to random vectors of any finite dimension having continuous marginal distributions. There are numerous other references that hint at the same idea. Therefore, the primary contribution of the present work is to pull together, extend and (in some cases) simplify previous results. In particular, we extend the idea to discrete or mixed marginal distributions, and we establish properties of the transformation that make fitting a NORTA distribution feasible. The results in this paper also extend the results of Cario and Nelson [1996], who defined ARTA (AutoRegressive To Anything) processes to model a stationary time series with an arbitrary marginal distribution and autocorrelation structure specified through lag p. Their results apply only to a common marginal distribution, while we allow each element of a NORTA vector to have a different marginal distribution.

Our work is closely related to methods that transform a random vector with uniformly distributed marginals; see, for example, Cook and Johnson [1981] and Johnson [1987, Chapter 10]. In fact, the NORTA transformation can be viewed as a two-step process, first transforming a multivariate normal vector $\mathbf{Z}$ into a multivariate uniform vector $\mathbf{U}$, then transforming the multivariate uniform vector into the desired input vector $\mathbf{X}$. The joint distribution of $\mathbf{U}$ is known as a *copula*, and any joint distribution has a representation as a transformation of a copula [Schweizer 1991]. Our approach is quite different from techniques that randomly mix distributions with extremal correlations to obtain intermediate correlations (e.g., Hill and Reilly [1994]), or methods that exploit special properties of a particular family of distributions. What we gain is a general-purpose, easy-to use tool; what we sacrifice is computational efficiency in fitting and random-variate generation.

We present our model in Section 2. In Section 3 we develop some relationships between the multivariate normal base vector $\mathbf{Z}$ and the input vector $\mathbf{X}$. We then discuss how to use these relationships to select the correlation matrix for the base vector that gives the desired correlation matrix for the input vector. In Section 4 we describe how NORTA random vectors are generated for use as simulation inputs and in Section 5 we provide several examples. Our conclusions appear in Section 6.

2 Model

The goal of our model is to define a random vector $\mathbf{X}$ with the following properties:

- $X_i \sim F_{X_i}, i = 1, 2, \ldots, k$, where each F_{X_i} is an arbitrary cumulative distribution function (cdf); and
- $\text{Corr}[\mathbf{X}] = \mathbf{\Sigma}_X$, where $\mathbf{\Sigma}_X$ is given.

We represent $\mathbf{X}$ as a transformation of a k-dimensional, standard multivariate normal (MVN) vector $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_k)'$ with correlation matrix $\mathbf{\Sigma}_Z$. Specifically, the NORTA vector $\mathbf{X}$ is

$$\mathbf{X} = \begin{pmatrix} F_{X_1}^{-1}[\Phi(Z_1)] \\ F_{X_2}^{-1}[\Phi(Z_2)] \\ \vdots \\ F_{X_k}^{-1}[\Phi(Z_k)] \end{pmatrix}$$

where Φ is the univariate standard normal cdf and $F_X^{-1}(u) \equiv \inf\{x : F_X(x) \geq u\}$ denotes the inverse cdf.

The transformation $F_{X_i}^{-1}[\Phi(\cdot)]$ ensures that X_i has the desired marginal distribution F_{X_i}. Therefore, the central problem is to select the correlation matrix $\mathbf{\Sigma}_Z$ that gives the desired correlation matrix $\mathbf{\Sigma}_X$.

3 Properties of NORTA Vectors

For $i \neq j$, let $\rho_Z(i,j)$ be the i,jth element of $\mathbf{\Sigma}_Z$, and let $\rho_X(i,j)$ be the i,jth element of $\mathbf{\Sigma}_X$. The correlation matrix of $\mathbf{Z}$ directly determines the correlation matrix of $\mathbf{X}$, since $\rho_X(i,j) = \mathrm{Corr}[X_i, X_j] = \mathrm{Corr}\left\{F_{X_i}^{-1}[\Phi(Z_i)], F_{X_j}^{-1}[\Phi(Z_j)]\right\}$ for all $i \neq j$. To adjust this correlation, we can restrict attention to adjusting $\mathrm{E}[X_i X_j]$, since

$$\mathrm{Corr}[X_i, X_j] = \frac{\mathrm{E}[X_i X_j] - \mathrm{E}[X_i]\mathrm{E}[X_j]}{\sqrt{\mathrm{Var}[X_i]\mathrm{Var}[X_j]}}$$

and $\mathrm{E}[X_i], \mathrm{E}[X_j], \mathrm{Var}[X_i]$ and $\mathrm{Var}[X_j]$ are fixed by F_{X_i} and F_{X_j}. Then, since (Z_i, Z_j) has a standard bivariate normal distribution with correlation $\mathrm{Corr}[Z_i, Z_j] = \rho_Z(i,j)$, we have

$$\begin{aligned} \mathrm{E}[X_i X_j] &= \mathrm{E}\left\{F_{X_i}^{-1}[\Phi(Z_i)] F_{X_j}^{-1}[\Phi(Z_j)]\right\} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_{X_i}^{-1}[\Phi(z_i)] F_{X_j}^{-1}[\Phi(z_j)] \varphi_{\rho_Z(i,j)}(z_i, z_j)\, dz_i\, dz_j, \end{aligned} \tag{1}$$

where $\varphi_{\rho_Z(i,j)}$ is the standard bivariate normal probability density function (pdf) with correlation $\rho_Z(i,j)$. We are only interested in distributions for which this expectation exists.

Observe from Equation (1) that the correlation between X_i and X_j is a function only of the correlation between Z_i and Z_j, which appears in the expression for $\varphi_{\rho_Z(i,j)}$. We denote this function by $c_{ij}[\rho_Z(i,j)]$. Thus, the problem of determining $\mathbf{\Sigma}_Z$ for $\mathbf{Z}$ that gives the desired correlation matrix $\mathbf{\Sigma}_X$ for $\mathbf{X}$ reduces to $k(k-1)/2$ independent problems: For each $i \neq j$, find the value $\rho_Z(i,j)$ for which $c_{ij}[\rho_Z(i,j)] = \rho_X(i,j)$. Unfortunately, it is not possible to express the ρ_Z-values in closed form except in special cases (see §5); however, we establish some properties of the function c_{ij} that enable us to perform an efficient numerical search to find the ρ_Z-values to within any desired precision.

The first two properties concern the sign and the range of $c_{ij}[\rho_Z(i,j)]$ for $-1 \leq \rho_Z(i,j) \leq 1$. The results in this section extend results in Cambanis and Marsy [1978] and Cario and Nelson [1996], which apply to time-series input processes with identical marginal distributions.

Proposition 1. *For any distributions F_{X_i} and F_{X_j}, $c_{ij}(0) = 0$, and $\rho_Z(i,j) \geq 0$ (≤ 0) implies that $c_{ij}[\rho_Z(i,j)] \geq 0$ (≤ 0).*

Proof. If $\rho_Z(i,j) = 0$, then

$$\mathrm{E}[X_i X_j] = \mathrm{E}\left\{F_{X_i}^{-1}[\Phi(Z_i)]F_{X_j}^{-1}[\Phi(Z_j)]\right\} = \mathrm{E}\left\{F_{X_i}^{-1}[\Phi(Z_i)]\right\}\mathrm{E}\left\{F_{X_j}^{-1}[\Phi(Z_j)]\right\} = \mathrm{E}[X_i]\mathrm{E}[X_j]$$

since $\rho_Z(i,j) = 0$ implies that Z_i and Z_j are independent. If $\rho_Z(i,j) \geq 0$ (≤ 0), then $\mathrm{Cov}[g_1(Z_i,Z_j), g_2(Z_i,Z_j)] \geq 0$ (≤ 0) for all nondecreasing functions g_1 and g_2 such that the covariance exists (Tong [1990], p. 20). Taking $g_1(Z_i,Z_j) \equiv F_{X_i}^{-1}[\Phi(Z_i)]$ and $g_2(Z_i,Z_j) \equiv F_{X_j}^{-1}[\Phi(Z_j)]$, the result follows since $F_X^{-1}[\Phi(\cdot)]$ is a nondecreasing function. □

It follows from the proof of Proposition 1 that taking $\rho_Z(i,j) = 0$ results in a vector in which X_i and X_j are not only uncorrelated, but are also independent.

Proposition 2. *Let $\overline{\rho}_{ij}$ and $\underline{\rho}_{ij}$ be the maximum and minimum feasible bivariate correlations, respectively, for random variables having marginal distributions F_{X_i} and F_{X_j}. Then, $c_{ij}(1) = \overline{\rho}_{ij}$ and $c_{ij}(-1) = \underline{\rho}_{ij}$.*

Proof. A correlation of 1 is the maximum possible for bivariate normal random variables. Therefore, taking $\rho_Z(i,j) = 1$ is equivalent (in distribution) to setting $Z_i \leftarrow \Phi^{-1}(U)$ and $Z_j \leftarrow \Phi^{-1}(U)$, where U is a U$(0,1)$ random variable (Whitt [1976]). But this definition of Z_i and Z_j implies that $X_i \leftarrow F_{X_i}^{-1}[U]$ and $X_j \leftarrow F_{X_j}^{-1}[U]$, from which it follows that $c_{ij}(1) = \overline{\rho}_{ij}$ by the same reasoning. Similarly, taking $\rho_Z(i,j) = -1$ is equivalent to setting $X_i \leftarrow F_{X_i}^{-1}[U]$ and $X_j \leftarrow F_{X_j}^{-1}[1-U]$, from which it follows that $c_{ij}(-1) = \underline{\rho}_{ij}$. □

Our next two results shed light on the shape of the function $c_{ij}[\rho_Z(i,j)]$.

Theorem 1. *The function $c_{ij}[\rho_Z(i,j)]$ is nondecreasing for $-1 \leq \rho_Z(i,j) \leq 1$.*

Proof. See the Appendix. □

Theorem 2. *If there exists $\epsilon > 0$ such that $\mathrm{E}[|X_i X_j|^{1+\epsilon}] < \infty$ for all values of $-1 \leq \rho_Z(i,j) \leq 1$, where X_i, X_j are defined by a NORTA transformation, then the function $c_{ij}[\rho_Z(i,j)]$ is continuous for $-1 \leq \rho_Z(i,j) \leq 1$.*

Proof. See the Appendix. □

Since $c_{ij}[\rho_Z(i,j)]$ is a continuous, nondecreasing function under the mild conditions stated in Theorem 2, any reasonable search procedure can be used to find $\rho_Z(i,j)$ such that $c_{ij}[\rho_Z(i,j)] \approx \rho_X(i,j)$. Proposition 1 provides the initial bounds for such a procedure. Proposition 2 shows that the extremal values of $\rho_X(i,j)$ are attainable under our model. Furthermore, from Proposition 2, Theorem 2 and the Intermediate Value Theorem, any feasible bivariate correlation for F_{X_i}, F_{X_j} is attainable under our model. Theorem 1 provides the theoretical basis for adjusting the values of $\rho_Z(i,j)$, and is the key to establishing convergence of a search procedure.

Throughout the previous discussion we assumed that there exists a joint distribution with marginal distributions $F_{X_i}, i = 1, 2, \ldots, k$, and correlation matrix $\mathbf{\Sigma}_X$. However, not all combinations of $F_{X_i}, i = 1, 2, \ldots, k$, and $\mathbf{\Sigma}_X$ are feasible. Clearly, for $\mathbf{\Sigma}_X$ to be feasible we must have $\underline{\rho}_{ij} \leq \rho_X(i,j) \leq \overline{\rho}_{ij}$ for each

$i \neq j$. In addition, $\mathbf{\Sigma}_X$ must be nonnegative definite. Our next result indicates that $\mathbf{\Sigma}_X$ will be nonnegative definite if $\mathbf{\Sigma}_Z$ is.

Proposition 3. If $\mathbf{\Sigma}_Z$ is nonnegative definite, then so is $\mathbf{\Sigma}_X$ implied by the NORTA transformation.

Proof. Provided that $\mathbf{\Sigma}_Z$ is nonnegative definite,

$$\begin{aligned} F_{\mathbf{X}}(x_1, x_2, \ldots, x_k) &= \Pr\{X_1 \leq x_1, X_2 \leq x_2, \ldots, X_k \leq x_k\} \\ &= \Pr\{Z_1 \leq \Phi^{-1}[F_{X_1}(x_1)], Z_2 \leq \Phi^{-1}[F_{X_2}(x_2)], \ldots, Z_k \leq \Phi^{-1}[F_{X_k}(x_k)]\} \end{aligned}$$

is a well-defined joint cdf for $-\infty < x_i < \infty$, $i = 1, 2, \ldots, k$. Therefore, $\mathbf{\Sigma}_X$ must be nonnegative definite. □

Comment: In a sense, the problem of representing multivariate random vectors with given marginals and dependence structure has been made difficult by the popularity of the product-moment correlation as the measure of dependence. Certain other measures, such as Spearman's ρ and Kendall's τ, are invariant under monotone transformations, so that fixing these measures for the base vector $\mathbf{Z}$ guarantees the same measures on the input vector $\mathbf{X}$. More precisely, both Spearman's ρ and Kendall's τ depend only on the copula of a pair of random variables, and $\mathbf{X}$ has the same copula as $\mathbf{Z}$ by construction.

4 Generating Simulation Input

Let $\mathbf{\Sigma}_Z$ be the correlation matrix such that $c_{ij}[\rho_Z(i,j)] \approx \rho_X(i,j)$ for all $i \neq j$. We can check whether $\mathbf{\Sigma}_Z$ is nonnegative definite to determine the existence of the desired joint distribution for $\mathbf{X}$. Random vectors are generated as follows:

NORTA Generation Procedure

1. Set up: Determine a lower-triangular, nonsingular factorization $\mathbf{M}$ of $\mathbf{\Sigma}_Z$ so that $\mathbf{M}\mathbf{M}' = \mathbf{\Sigma}_Z$.
2. Generate $\mathbf{W} = (W_1, W_2, \ldots, W_k)'$, a $k \times 1$ random vector whose elements are i.i.d. standard normal random variables.

3. Set $\mathbf{Z} \leftarrow \mathbf{MW}$.
4. Return $\mathbf{X}$ where $X_i \leftarrow F_{X_i}^{-1}[\Phi(Z_i)]$, $i = 1, 2, \ldots, k$.
5. Go to step 2.

Steps 1–3 are a standard method for generating a MVN vector; see Johnson [1987] for a detailed justification.

5 Examples

Two special cases are easier than the general problem. If X_i and X_j have continuous uniformly distributed marginals, then

$$\rho_X(i,j) = \frac{6}{\pi} \sin^{-1}\left(\frac{\rho_Z(i,j)}{2}\right)$$

from Li and Hammond [1975, equation (7)]. When X_i and X_j are exponentially distributed, then the transformation $\rho_X(i,j) = c_{ij}[\rho_Z(i,j)]$ is independent of the parameters of the exponential distributions, meaning that a fine grid of $[\rho_X(i,j), \rho_Z(i,j)]$ pairs can be stored and used as starting values in a search. Figure 1 shows a plot of the function $c_{ij}[\rho_Z(i,j)]$ for the uniformly distributed case, where the relationship is nearly the identity mapping, while Figure 2 shows a similar plot for the exponentially distributed case, where there is significant curvature near the boundary $c_{ij}[-1] \approx -0.645$.

In general, a numerical search is required to find the $\rho_Z(i,j)$ such that $c_{i,j}[\rho_Z(i,j)] \approx \rho_X(i,j)$ for all $i \neq j$. In the special case when the marginals of $X_1, X_2, \ldots, X_k$ are all the same, the `ARTAFACTS` software described in Cario and Nelson [1997] does this automatically.[1] Two examples are given below.

Suppose that we require a trivariate random variable with marginals that are all Gamma($\beta = 0.03424, \alpha =$

[1] `ARTAFACTS` is designed to fit a stationary time series with arbitrary marginal distribution and autocorrelations specified through lag p. More information, and the software itself, can be obtained from `http://www.iems.nwu.edu/~ nelsonb/ARTA/`.

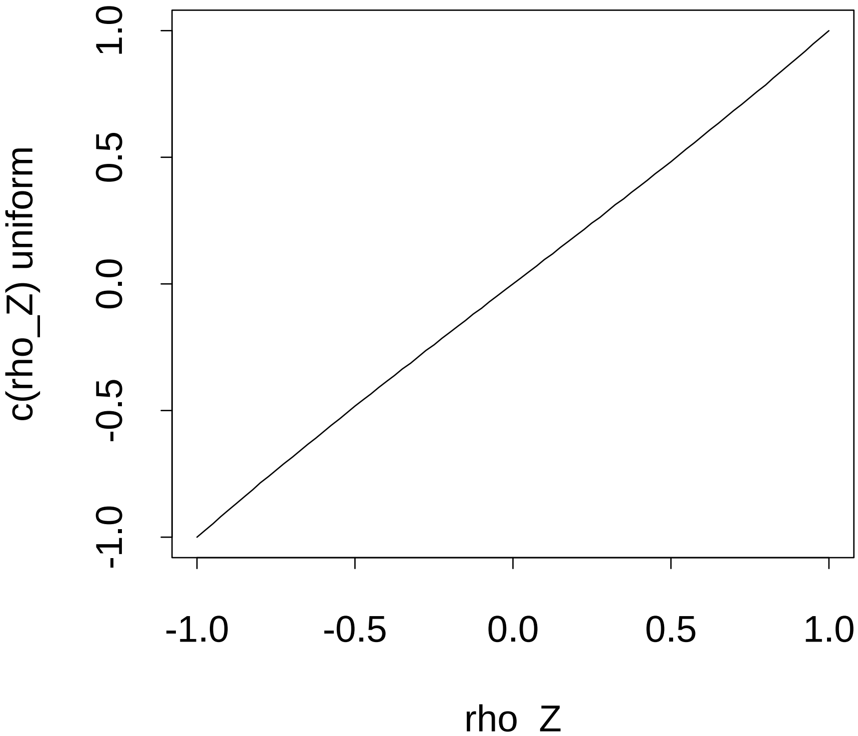

Figure 1: The function $c_{ij}[\rho_Z(i,j)]$ when X_i and X_j have uniformly distributed marginals.

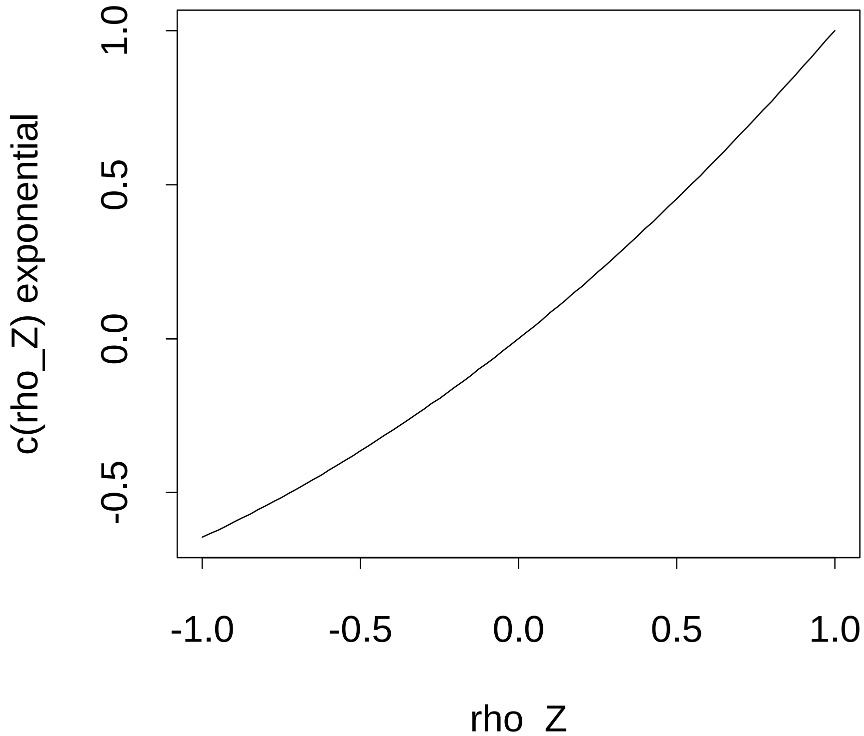

Figure 2: The function $c_{ij}[\rho_Z(i,j)]$ when X_i and X_j have exponentially distributed marginals.

14.4) and correlation matrix

$$\mathbf{\Sigma}_X = \begin{pmatrix} 1.0 & 0.7 & 0.5 & -0.9 \\ & 1.0 & 0.7 & -0.6 \\ & & 1.0 & -0.3 \\ & & & 1.0 \end{pmatrix}.$$

These characteristics are attained by a NORTA vector whose underlying trivariate normal random vector has correlation matrix

$$\mathbf{\Sigma}_Z = \begin{pmatrix} 1.0 & 0.704 & 0.504 & -0.920 \\ & 1.0 & 0.704 & -0.616 \\ & & 1.0 & -0.304 \\ & & & 1.0 \end{pmatrix}.$$

Notice that the correlation matrix of $\mathbf{Z}$ differs only slightly from the desired correlation matrix for $\mathbf{X}$, as is often the case when F_X is continuous and relatively symmetric.

One of the advantages of the NORTA transformation is that discrete marginals are no more difficult than continuous marginals (in fact, the numerical work required for fitting discrete marginals is somewhat less). For instance, suppose we require a trivariate random variable with all marginals Binomial($n = 3, p = 0.5$) and correlation matrix

$$\mathbf{\Sigma}_X = \begin{pmatrix} 1.0 & 0.2 & -0.8 \\ & 1.0 & 0.2 \\ & & 1.0 \end{pmatrix}.$$

These characteristics are attained by a NORTA vector whose underlying trivariate normal random vector has correlation matrix

$$\mathbf{\Sigma}_Z = \begin{pmatrix} 1.0 & 0.2288 & -0.8960 \\ & 1.0 & 0.2288 \\ & & 1.0 \end{pmatrix}.$$

Notice that in this case $\mathbf{\Sigma}_X$ and $\mathbf{\Sigma}_Z$ differ significantly.

Of course, the most important feature of the NORTA transformation is that random vectors that include both continuous and discrete component random variables are handled within the same framework. Although we have not yet modified the `ARTAFACTS` code to fit general NORTA vectors, a crude numerical search will suffice in many cases. For example, if we need a bivariate random vector (X_1, X_2) with X_1 having a discrete uniform distribution on $\{1, 2, \ldots, 10\}$, X_2 having an exponential distribution with mean 10, and (X_1, X_2) having correlation matrix

$$\mathbf{\Sigma}_X = \begin{pmatrix} 1.0 & -0.5 \\ & 1.0 \end{pmatrix}$$

then these characteristics will be attained by a NORTA vector whose underlying bivariate normal random vector has correlation matrix

$$\mathbf{\Sigma}_Z = \begin{pmatrix} 1.0 & -0.576 \\ & 1.0 \end{pmatrix}.$$

In this case we matched the desired correlation by using a bisection search on $\rho_Z(1,2)$, estimating the implied correlation $\rho_X(1,2)$ by generating 200000 random vectors (the standard error of the correlation estimate is approximately 0.0016). For dimension $k > 2$ the same procedure would be followed for each of the $k(k-1)/2$ pairs of correlations. Figure 3 shows a scatterplot of 200 observations from this NORTA vector, where it is clear that small values of the discrete uniform tend to be paired with large values of the exponential, and vice versa.

6 Conclusions

The NORTA method, and the related ARTA method for time-series input processes, provide a general-purpose tool for modeling and generating dependent input processes. This generality comes at the cost of computational efficiency. The fitting process is time consuming, although this expense is incurred only once for each input model. More importantly, the marginal time for generating each NORTA variate can be longer than the fastest available method for a particular distribution, due to the need to evaluate the composite function $F_X^{-1}[\Phi(\cdot)]$. However, in system simulation applications where input/output processing, event-list management, animation, etc. account for the bulk of the execution time, the additional time required to generate NORTA variates will be acceptable, and the generality of the method welcome.

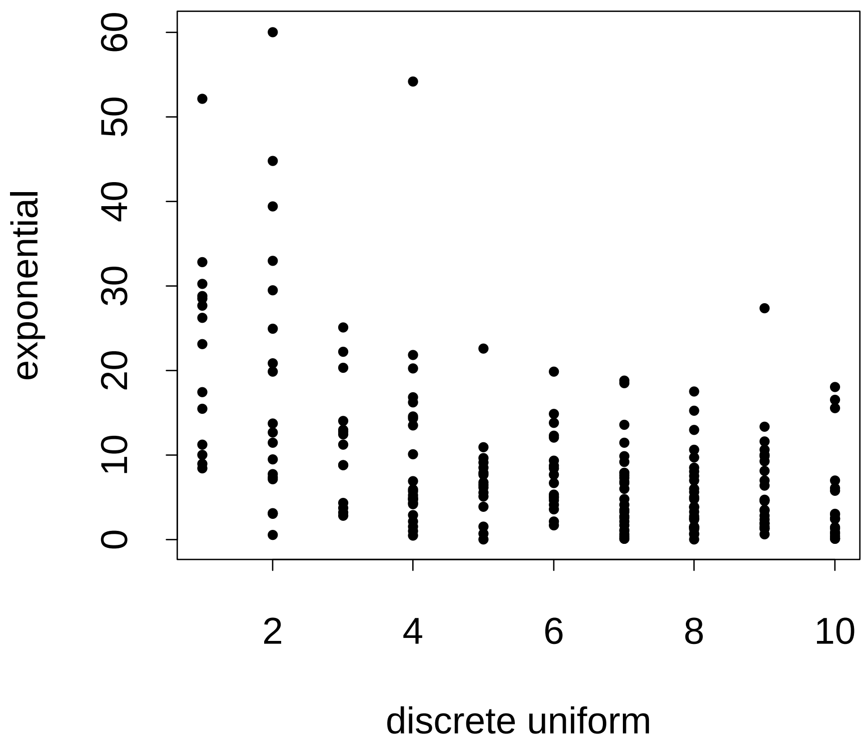

Figure 3: Scatterplot of 200 observations from a bivariate exponential-discrete uniform NORTA random vector with correlation -0.5.

It is also important to note that there is no statistical theory to back up the use of a NORTA distribution; in fact, there seems to be little reason to believe that the "true" underlying distribution is ever NORTA. The philosophy of the NORTA approach is to match distributional properties—marginals and correlation matrix—that are considered important to having a good model, rather than giving any consideration to finding the true model, if such a thing ever exists.

Appendix

Lemma 1. *Let* $\vec{\boldsymbol{\mu}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ *and* $\boldsymbol{\Sigma}_{\rho_i} = \begin{pmatrix} 1 & \rho_i \\ \rho_i & 1 \end{pmatrix}$*, and let* $(N_1, N_2)'$ *and* $(Z_1, Z_2)'$ *be bivariate normal random variables with common mean* $\vec{\boldsymbol{\mu}}$ *and variance-covariance matrices* $\boldsymbol{\Sigma}_{\rho_2}$ *and* $\boldsymbol{\Sigma}_{\rho_1}$*, respectively, where* $0 \leq \rho_1 < \rho_2 < 1$*. Let* $g_1(x)$ *and* $g_2(x)$ *be nondecreasing functions of* x *for* $-\infty < x < \infty$*. Then for any* g_i *for which* $\mathrm{E}[g_i^2(N)]$ *exists,* $i = 1, 2$*, the* $\mathrm{E}[g_1(N_1)g_2(-N_2)] \leq \mathrm{E}[g_1(Z_1)g_2(-Z_2)]$.

Proof. The proof extends Cario and Nelson [1996], Lemma 1.

Let T_1, T_2, V_1, V_2, and W be i.i.d. N(0,1) random variables. Then,

$$(N_1, -N_2) \stackrel{d}{=} \left(\sqrt{1-\rho_2}T_1 + \sqrt{\rho_2 - \rho_1}V_1 + \sqrt{\rho_1}W, -\sqrt{1-\rho_2}T_2 - \sqrt{\rho_2 - \rho_1}V_1 - \sqrt{\rho_1}W\right)$$

and

$$(Z_1, -Z_2) \stackrel{d}{=} \left(\sqrt{1-\rho_2}T_1 + \sqrt{\rho_2 - \rho_1}V_1 + \sqrt{\rho_1}W, -\sqrt{1-\rho_2}T_2 - \sqrt{\rho_2 - \rho_1}V_2 - \sqrt{\rho_1}W\right)$$

where $\stackrel{d}{=}$ denotes equality in distribution. Therefore,

$$\begin{aligned} \mathrm{E}[g_1(N_1)g_2(-N_2)] &= \mathrm{E}\left[\mathrm{E}\left\{\mathrm{E}\left[g_1(\sqrt{1-\rho_2}T_1 + \sqrt{\rho_2-\rho_1}V_1 + \sqrt{\rho_1}W) \times \right.\right.\right. \\ &\qquad \left.\left.\left. g_2(-\sqrt{1-\rho_2}T_2 - \sqrt{\rho_2-\rho_1}V_1 - \sqrt{\rho_1}W) \mid V_1, W\right] \mid W\right\}\right] \\ &= \mathrm{E}\left[\mathrm{E}\left\{\Psi_W^{(1)}(V_1)\Psi_{-W}^{(2)}(-V_1) \mid W\right\}\right] \end{aligned}$$

where

$$\Psi_w^{(i)}(v_1) = \mathrm{E}\left[g_i(\sqrt{1-\rho_2}T + \sqrt{\rho_2 - \rho_1}V_1 + \sqrt{\rho_1}W) \mid V_1 = v_1, W = w\right]$$

and the expectation is with respect to T, an independent $\mathrm{N}(0,1)$ random variable.[2]

For g_1 nondecreasing and fixed $W = w$, the function $\Psi_w^{(1)}(v_1)$ is nondecreasing in v_1. Similarly, $-\Psi_{-w}^{(2)}(v)$ is nonincreasing in v (where v is a dummy variable used only for clarity). Therefore,

$$\mathrm{Var}\left[\Psi_w^{(1)}(V_1) - \left\{-\Psi_{-w}^{(2)}(V)\right\}\right] = \mathrm{Var}\left[\Psi_w^{(1)}(V_1)\right] + \mathrm{Var}\left[-\Psi_{-w}^{(2)}(V)\right] - 2\mathrm{Cov}\left[\Psi_w^{(1)}(V_1), -\Psi_{-w}^{(2)}(V)\right]$$

is minimized with respect to all joint distributions of (V_1, V) with $\mathrm{N}(0,1)$ marginals when $V_1 = \Phi^{-1}(U)$ and $V = \Phi^{-1}(1-U)$, where Φ is the standard normal cdf and U is a $\mathrm{U}(0,1)$ random variable (Rubinstein et al. [1985], Proposition 1). For $\mathrm{N}(0,1)$ random variables this implies that $V = -V_1$. Therefore, $\mathrm{Cov}[\Psi_w^{(1)}(V_1), -\Psi_{-w}^{(2)}(V)]$ is maximized (equivalently, $\mathrm{Cov}[\Psi_w^{(1)}(V_1), \Psi_{-w}^{(2)}(V)]$ is minimized) by letting $V = -V_1$. Thus,

$$\begin{aligned} \mathrm{E}\{\Psi_w^{(1)}(V_1)\Psi_{-w}^{(2)}(-V_1) \mid W = w\} &\leq \mathrm{E}\{\Psi_w^{(1)}(V_1) \mid W = w\}\mathrm{E}\{\Psi_{-w}^{(2)}(-V_1) \mid W = w\} \qquad (2) \\ &= \mathrm{E}\{\Psi_w^{(1)}(V_1) \mid W = w\}\mathrm{E}\{\Psi_{-w}^{(2)}(-V_2) \mid W = w\} \qquad (3) \end{aligned}$$

where (2) holds because the minimum expected value must be smaller than the expected value under independence, and (3) holds because V_1 and V_2 are identically distributed. Since (2) and (3) hold for any value of W, it follows that

$$\mathrm{E}\left[\mathrm{E}\{\Psi_W^{(1)}(V_1)\Psi_{-W}^{(2)}(-V_1) \mid W\}\right] \leq \mathrm{E}\left[\mathrm{E}\{\Psi_W^{(1)}(V_1) \mid W\}\mathrm{E}\{\Psi_{-W}^{(2)}(-V_2) \mid W\}\right].$$

But notice that

$$\mathrm{E}\left[g_1(Z_1)g_2(-Z_2)\right] = \mathrm{E}\left[\mathrm{E}\{\Psi_W^{(1)}(V_1)\Psi_{-W}^{(2)}(-V_2)\} \mid W\right]$$

[2] Notice that $T_1 \stackrel{d}{=} -T_2$, and they are independent.

$$= \mathrm{E}\left[\mathrm{E}\{\Psi_W^{(1)}(V_1) \mid W\}\mathrm{E}\{\Psi_{-W}^{(2)}(-V_2) \mid W\}\right]$$

since V_1 and V_2 are independent. □

Corollary. *Let $(N_1, N_2)'$ and $(Z_1, Z_2)'$ have bivariate normal distributions with common mean $\vec{\boldsymbol{\mu}}$ and variance-covariance matrices $\boldsymbol{\Sigma}_{\rho_2}$ and $\boldsymbol{\Sigma}_{\rho_1}$, respectively, where $-1 < \rho_2 < \rho_1 \leq 0$. Let $g_1(x)$ and $g_2(x)$ be nondecreasing functions of x for $-\infty < x < \infty$. Then,* $\mathrm{E}[g_1(N_1)g_2(N_2)] \leq \mathrm{E}[g_1(Z_1)g_2(Z_2)]$.

Proof. This follows from Lemma 1 since $(N_1, -N_2)'$ and $(Z_1, -Z_2)'$ have bivariate normal distributions with mean $\vec{\boldsymbol{\mu}}$ and covariance matrices $\boldsymbol{\Sigma}_{-\rho_2}$ and $\boldsymbol{\Sigma}_{-\rho_1}$, respectively. □

Lemma 2. Under the same conditions as Lemma 1, $\mathrm{E}[g_1(N_1)g_2(N_2)] \geq \mathrm{E}[g_1(Z_1)g_2(Z_2)]$.

Proof. The proof extends Tong [1990], Theorem 5.3.10, to the case of nonidentical cdfs. It is analogous to the proof of Lemma 1, but makes use of the fact that

$$(N_1, N_2) \stackrel{d}{=} \left(\sqrt{1-\rho_2}T_1 + \sqrt{\rho_2 - \rho_1}V_1 + \sqrt{\rho_1}W, \sqrt{1-\rho_2}T_2 + \sqrt{\rho_2 - \rho_1}V_1 + \sqrt{\rho_1}W\right)$$

$$(Z_1, -Z_2) \stackrel{d}{=} \left(\sqrt{1-\rho_2}T_1 + \sqrt{\rho_2 - \rho_1}V_1 + \sqrt{\rho_1}W, \sqrt{1-\rho_2}T_2 + \sqrt{\rho_2 - \rho_1}V_2 + \sqrt{\rho_1}W\right)$$

and that the covariance between any nondecreasing functions of V_1 and V is maximized with respect to all joint distributions of (V_1, V) with $\mathrm{N}(0,1)$ marginals when $V = V_1 = \Phi(U)$ (Rubinstein et al. [1985], Proposition 1). □

Proof of Theorem 1. By taking $g_i \equiv F_{X_i}^{-1}[\Phi(\cdot)]$ in the Corollary (if $\rho_Z(i,j) < 0$) or in Lemma 2 (if $\rho_Z(i,j) \geq 0$), it follows that $c_{ij}[\rho_Z(i,j)]$ is a nondecreasing function for $-1 \leq \rho_Z(i,j) \leq 1$. □

Lemma 3. *For given cdfs F_{X_i}, $i = 1, 2$, if there exists $\epsilon > 0$ such that*

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sup_{\rho \in [-1,1]} \left\{ \left| F_{X_1}^{-1}[\Phi(z_1)] F_{X_2}^{-1}[\Phi(z_2)] \right|^{1+\epsilon} \varphi_\rho(z_1, z_2) \right\} dz_1 dz_2 \quad < \quad \infty \tag{4}$$

then $c_{12}(\rho)$ is a continuous function for $-1 \leq \rho \leq 1$.

Proof. The proof extend Cario and Nelson [1996], Lemma 2.

Let Z_1 and Z_3 be i.i.d. N(0,1) random variables. Let $\rho \in [-1, 1]$ be fixed, and $\{\rho_n\}_{n=1}^{\infty}$ be any sequence such that $\rho_n \in [-1, 1]$, for $n = 1, 2, \ldots$, and $\rho_n \to \rho$ as $n \to \infty$. For $n = 1, 2, \ldots$, define

$$Z_{1n} \equiv Z_1, \qquad Z_{2n} \equiv \rho_n Z_1 + \sqrt{1 - \rho_n^2}\, Z_3, \qquad Z_2 \equiv \rho Z_1 + \sqrt{1 - \rho^2}\, Z_3.$$

Further, let $X_{in} \equiv F_{X_i}^{-1}[\Phi(Z_{in})]$, for $i = 1, 2$, and $h\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \equiv F_{X_1}^{-1}[\Phi(z_1)] F_{X_2}^{-1}[\Phi(z_2)]$. Since h is monotone in z_1 and z_2 individually, it has a countable number of discontinuities. Therefore, by the Continuous Mapping Theorem (Billingsley [1995], Theorem 29.2)

$$h\begin{pmatrix} Z_{1n} \\ Z_{2n} \end{pmatrix} \stackrel{d}{\Rightarrow} h\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \text{ as } n \to \infty,$$

since

$$\begin{pmatrix} Z_{1n} \\ Z_{2n} \end{pmatrix} \stackrel{d}{\Rightarrow} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \text{ as } n \to \infty,$$

where $\stackrel{d}{\Rightarrow}$ denotes convergence in distribution. Equivalently,

$$X_{1n} X_{2n} \quad \stackrel{d}{\Rightarrow} \quad X_1 X_2 \text{ as } n \to \infty, \tag{5}$$

where $X_i \equiv F_{X_i}^{-1}[\Phi(Z_i)]$, for $i = 1, 2$. It follows from (4), (5), and Theorem 25.12 of Billingsley [1995], that $\mathrm{E}[X_{1n} X_{2n}] \to \mathrm{E}[X_1 X_2]$ as $n \to \infty$; equivalently, $c_{12}[\rho_n] \to c_{12}[\rho]$ as $n \to \infty$. □

Notice that condition (4) of Lemma 3 is equivalent to stating that $E[\mid X_i X_j \mid^{1+\epsilon}] < \infty$ for all values of $-1 \leq \rho_Z(i,j) \leq 1$, where X_i, X_j are defined by our transformation, which is the condition given in the statement of Theorem 2.

Proof of Theorem 2. Theorem 2 follows immediately from Lemma 3 with $Z_1 \equiv Z_i, Z_2 \equiv Z_j, X_1 \equiv X_i, X_2 \equiv X_j$, and $\rho \equiv \rho_Z(i,j)$. □

References

1. Billingsley, P. 1995. *Probability and Measure.* Third Edition. New York: John Wiley.
2. Cambanis, S. and E. Marsy. 1978. On the reconstruction of the covariance of stationary Gaussian processes observed through zero-memory nonlinearities. *IEEE Transactions on Information Theory* **24**, 485–494.
3. Cario, M. C. and B. L. Nelson. 1996. Autoregressive to Anything: Time Series Input Processes for Simulation. *Operations Research Letters* **19**, 51–58.
4. Cario, M. C. and B. L. Nelson. 1997. Numerical Methods for Fitting and Simulating Autoregressive-to-Anything Processes. *INFORMS Journal on Computing*, forthcoming.
5. R. D. Cook and M. E. Johnson. 1981. A family of distributions for modeling non-elliptically symmetric multivariate data. *Journal of the Royal Statistical Society B* **43**, 210–218.
6. Devroye, L. 1986. *Non-Uniform Random Variate Generation.* New York: Springer-Verlag.
7. Hill, R. R. and C. H. Reilly. 1994. Composition for multivariate random vectors. In *Proceedings of the 1994 Winter Simulation Conference*, 332–339.
8. Johnson, M. E. 1987. *Multivariate Statistical Simulation.* New York: John Wiley.
9. Li, S. T. and J. L. Hammond. 1975. Generation of pseudorandom numbers with specified univariate distributions and correlation coefficients. *IEEE Transactions on Systems, Man and Cybernetics* **5**, 557–561.
10. Mardia, K. V. 1970. A translation family of bivariate distributions and Fréchet's bounds. *Sankhya* **A32**, 119–122.
11. Rubinstein, R. Y., Samorodnitsky, G. and M. Shaked. 1985. Antithetic Variates, Multivariate Dependence and Simulation of Stochastic Systems. *Management Science* **31**, 66–77.
12. Schweizer, B. 1991. Thirty years of copulas. In *Advances in Probabiity Distributions with Given Marginals: Beyond the Copulas* (G. Dall'Aglio, S. Kotz and G. Salinetti, eds.), 13–50. Boston: Kluwer.

13. Tong, Y. L. 1990. *The Multivariate Normal Distribution.* New York: Springer-Verlag.

14. Whitt W. 1976. Bivariate Distributions with Given Marginals. *The Annals of Statistics* **4**, 1280–1289.