DEPENDENCE MODELING FOR STOCHASTIC SIMULATION

Bahar Biller

Tepper School of Business
Carnegie Mellon University
Pittsburgh, PA 15213, U.S.A.

Soumyadip Ghosh

IBM T.J. Watson Research Center
Yorktown Heights, NY 10598, U.S.A.

ABSTRACT

An important step in designing stochastic simulation is modeling the uncertainty in the input environment of the system being studied. Obtaining a reasonable representation of this uncertainty can be challenging in the presence of dependencies in the input process. This tutorial attempts to provide a coherent narrative of the central principles that underlie methods that aim to model and sample a wide variety of dependent input processes.

1 INTRODUCTION

An important step in the design of stochastic simulation is input modeling, i.e., modeling the uncertainty in the input environment of the system being studied. Input modeling is often characterized as selecting appropriate univariate probability distributions to represent the primitive inputs of interest, and it would indeed be this simple if the relevant input processes could be represented as a sequence of independent random variables having identical distributions. When such univariate models do apply, there are a number of software packages that support automated, or nearly automated, input modeling; refer Vincent (1998) or Law and Kelton (2000) for a good review.

However, these simple models fail to capture the stochastic properties of input processes that exhibit marked dependencies that occur naturally in different forms in many service, communications, and manufacturing systems. For example, Melamed, Hill, and Goldsman (1992) observe autocorrelation in sequences of compressed video frame bit-rates, while Ware, Page, and Nelson (1998) report that the times between file accesses on a computer network frequently exhibit burstiness, as characterized by a sequence of short inter-access times followed by one or more long ones. The ability to capture these dependencies is crucial because even detailed logical models combined with sound experimental designs and output analysis cannot compensate for inaccurate input models. This might further lead to performance measures that are seriously in error as illustrated by Livny, Melamed, and Tsiolis (1993), who examine the impact of autocorrelated interarrival times on the mean waiting time of a single-server queue following the FIFO discipline. When the exponential interarrival and service times are independent and the system utilization is 50%, the mean waiting time is 1.0 time unit. Livny et al. take this case as benchmark and observe that introduction of an autocorrelation of -0.55 reduces the mean waiting time by 32%, while an autocorrelation of 0.25 increases the mean waiting time by 80%. The impact gets even more dramatic with the strength of autocorrelation: an autocorrelation of 0.85 results in a mean waiting time that is more than 200 times larger than the benchmark case. It is clear that, even in the case of an exponential single-server queue, the independence assumption leads to very poor estimates of the performance measures.

We focus here on developing input models for two types of multivariate processes. The first type captures dependencies among a finite number of random variables, jointly called a random vector, and consists of independent samples of identically distributed random vectors. The second kind captures temporal dependence that arises over time and is traditionally studied in terms of time series.

A good input modeling methodology should be flexible in modeling varied situations, and should allow for efficient generation of samples of the input model. A message the reader would consistently get from the ideas presented in the tutorial is that multivariate input modeling is far from being a complete discipline. It is, if anything, the next frontier of methodology research in input modeling. Interesting problems which promise both theoretical content as well as practical impact abound.

The rest of this tutorial is structured as follows. Section 2 provides the relevant definitions and some key concepts that are used in modeling dependence for stochastic simulation. The key ideas that underlie modeling both random vectors and time series are often the same, and hence the subsequent sections present methods in a unified manner where possible. The methods are grouped into sections in the descending order of control we have on the construction of joint distributions (defined in Section 2.1) of the

multivariate processes. We describe the methods that work with the fully specified or estimated joint distribution of dependent random variables in Section 3, while we present the practical alternative of using partially specified random processes for input model development in Section 4. A substantially comprehensive version of this tutorial will be part of a forthcoming handbook on simulation (see Biller and Ghosh 2005).

2 BASIC PROPERTIES OF MULTIVARIATE INPUT PROCESSES

2.1 Random Vectors

A *random vector* $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ denotes a finite collection of d random components. Each of these components is a random variable with a distribution function in $\Re$. We associate with the random vector a probability distribution, that is called the *joint* (or *multivariate*) *distribution function*, in the space $\Re^d$. The joint cumulative distribution function (cdf) of a d-dimensional random vector $\mathbf{X}$ is defined as

$$F(\mathbf{x}) = P(\mathbf{X} \le \mathbf{x}) = P(X_1 \le x_1, \ldots, X_d \le x_d)$$

for any fixed d-vector $\mathbf{x} = (x_1, x_2, \ldots, x_d)'$. This cdf satisfies some basic properties: It is non-decreasing and right-continuous in each argument x_i for $i = 1, 2, \ldots, d$. Moreover, $\lim_{\mathbf{x} \to -\infty} F(\mathbf{x}) = 0$ and $\lim_{\mathbf{x} \to +\infty} F(\mathbf{x}) = 1$, where the limits hold jointly for all components.

The joint distribution function completely characterizes the behavior of $\mathbf{X}$: it defines the distribution of each of its component random variables, termed the *marginal* distributions, and determines their stochastic relationships.

The random variables X_i, $i = 1, 2, \ldots, d$, are stochastically independent if and only if their joint cdf can be factored into the product of the d marginal cdfs. The case when X_i, $i = 1, 2, \ldots, d$, are not stochastically independent is the focus of this tutorial.

Section 3 gives an account of the methods that either require the complete specification of the joint distribution via expert opinion or the estimation of an appropriate distribution via the use of data. These methods highlight the fact that capturing the properties of a multivariate function can often be a tough task. In practice, matching certain key measures of dependence reasonably well is a compromise that is often found satisfactory. The hope is that a properly chosen partial specification can capture the essence of the dependence between the components while sparing the practitioner the arduous task of estimating the complete joint distribution. The next section defines some of the common dependence measures used.

2.2 Measures of Dependence

The most widely used and understood measure of dependence in simulation experiments is the *product-moment correlation*. The (i, j)th product-moment correlation (with sample analog being the Pearson correlation) between components X_i and X_j is defined as

$$\rho(i, j) = E[(X_i - \mu_i)(X_j - \mu_j)]/(\sigma_i \sigma_j),$$

where $E[X_\ell] = \mu_\ell$ and $Var(X_\ell) = \sigma_\ell^2$ for $\ell = i, j$. If X_i and X_j are independent, then $\rho(i, j) = 0$. On the other hand, zero correlation between X_i and X_j does not necessarily imply that X_i and X_j are independent. As an example, consider a two-dimensional random vector $\mathbf{X} = (X_1, X_2)'$ that is uniformly distributed in the unit circle. Although it is true that $\rho(1, 2) = 0$, the random variables X_1 and X_2 are related by $X_1^2 + X_2^2 = 1$.

Despite its wide use, product-moment correlation suffers several limitations. Perhaps, the most critical one is that correlation is simply a scalar measure of dependence and, therefore, it cannot capture nonlinear dependence relationships that may exist between real-world input processes as well as the degree of dependence in the tails of the underlying distributions. Figure 1 shows 10000 bivariate realizations from two different probability models for $\mathbf{X} = (X_1, X_2)'$. In both models, X_1 and X_2 have identical normal marginal distributions and the linear correlation between them is 0.7. However, it is clear that the dependence between X_1 and X_2 in the two models is qualitatively quite different, e.g., extreme positive realizations have a tendency to occur together in the second model. Thus the dependence in the two models can not be distinguished on the grounds of correlations alone.

Correlations are also not invariant under any transformation of the inputs, and the values $\rho(i, j)$ can vary with the marginal distributions of the random variables X_i and X_j. For example, any correlation value from the interval $[-1, 1]$ can be attained for two normal random variables, but the attainable interval for the correlation between two lognormal random variables, which are often generated via a transformation of normals, with zero means and standard deviations of 1 and 2 is $[-0.09, 0.67]$.

Fortunately, there exists measures of dependence that avoid some of these pitfalls. One such measure is the fractile correlation

$$r(i, j) = 12E\left[F_i(X_i) F_j(X_j)\right] - 3, \tag{1}$$

whose sample analog is called Spearman or rank correlation. The two measures coincide if the marginal distributions F_i and F_j are all uniform. Rank correlations can be shown to be invariant to strictly increasing transformations of the components. Thus, it provides a natural way to separate
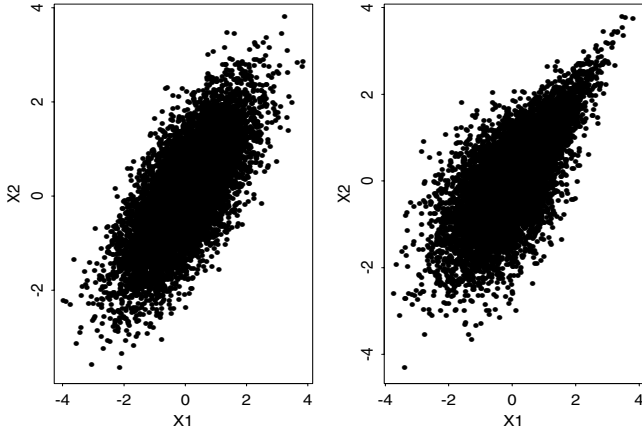
Figure 1: 10000 Simulated Data Points from Two Bivariate Distributions with Identical Normal Marginal Distributions and Identical Correlation of 0.7, but Different Dependence Structures (Embrechts, McNeil, and Straumann 2002)

the estimation of the individual marginal distributions from that of the dependence among the variables. Of course, the uniform random vector can be modeled with any number of joint cdfs with the same rank correlation matrix. The joint cdfs can be uniquely determined only under certain special cases that will be discussed in Section 3.3. We refer the reader to Nelsen (1999) for other measures of dependence that have been suggested for the special case of uniform marginal distributions.

Note that in matching a correlation-based specification, or any partial specification, one has to ensure that the assumed values for the dependence measures are *feasible*, that is, there exists at least one joint distribution that possesses dependence measures with these values. In the lognormal example cited earlier, clearly one can not find a joint cdf that can match a correlation value of 0.7. When feasible, a specification might of course be matched by many joint distributions.

2.3 Multivariate Time Series

A *multivariate time series* $\{\mathbf{X}_t; t = 1, 2, \ldots\}$ denotes a series of finite d-dimensional random vectors with joint distribution functions in $\Re^d$. It is a stochastic process with a probability distribution on the space of all possible path realizations. Such a probability distribution provides a complete description of the stochastic properties of a multivariate time series. However, almost all statistical methods for the analysis of time series at the present time specify the series in terms of the probability distributions of the individual random elements of $\mathbf{X}_t$ and rely on the fundamental concept of *autocorrelation* to capture the dependence. Autocorrelations are the correlations between observations within the series.

The classical multivariate time series model is the $d \times 1$ vector linear process given by

$$\mathbf{X}_t = \sum_{i=-\infty}^{\infty} \boldsymbol{\alpha}_i \mathbf{Y}_{t-i}, \quad t = 0, \pm 1, \pm 2, \ldots, \quad (2)$$

where $\{\boldsymbol{\alpha}_i; i = 0, \pm 1, \pm 2, \ldots\}$ is a sequence of $d \times d$ matrices such that $\sum_{i=-\infty}^{\infty} \| \boldsymbol{\alpha}_i \| < \infty$ ($\| \cdot \|$ denotes the usual eigenvalue norm) and $\{\mathbf{Y}_t; t = 0, \pm 1, \pm 2, \ldots, \}$ is a sequence of i.i.d. d-dimensional white noise vectors with mean zero and covariance matrix Σ_Y (Wei 1990). If the $\mathbf{Y}_t$'s are Gaussian, then clearly so are the $\mathbf{X}_t$'s. Furthermore, if the $\mathbf{X}_t$'s are Gaussian with mean zero and absolutely continuous spectrum, then there is a sequence of i.i.d. normal mean-zero random vectors $\mathbf{Y}_t$, $t = 0, \pm 1, \pm 2, \ldots$, and a sequence of matrices $\{\boldsymbol{\alpha}_i; i = 0, \pm 1, \pm 2, \ldots\}$, such that the processes $\mathbf{X}_t$ and $\sum_{i=-\infty}^{\infty} \boldsymbol{\alpha}_i \mathbf{Y}_{t-i}$, $t = 0, \pm 1, \pm 2, \ldots$, are stochastically equal.

Mallows (1967) shows that the linearity of the multivariate time series model (2) implies normal marginal distributions, but there are many physical situations in which the marginals of the time-series input processes are non-normal. If the $\mathbf{X}_t$ process is non-Gaussian, then the decomposition (2) may not exist and the statistical inference procedures developed for processes satisfying (2) do not apply. It has been suggested that time series that depart slightly from normality be handled by data transformations. For other cases, where the departure from normality is more substantial, it has been suggested that new time series models be developed. Over the past decade, there has been a considerable amount of research on modeling time series with exponential, gamma, geometric, and general discrete marginal distributions. We refer the reader to Block, Langberg, and Stoffer (1990) that present bivariate linear time-series models with non-normal marginal distributions and the references therein.

3 CONSTRUCTING FULLY SPECIFIED JOINT DISTRIBUTIONS

Although the methods of this section are typically used for constructing random vectors, they can be applied to time-series processes of fixed length. The first three methods work when the joint distribution function is available. The objective of the next two methods is to obtain a non-parametric characterization of the joint distribution function from available data. Finally, we review some of the popular parameterized forms used to obtain approximations to the joint cdf. All of these methods suffer from a common limitation that the amount of information they require to do a reasonable job increases enormously with dimension.

3.1 Acceptance/Rejection Method

The acceptance/rejection principle of generating random vector $\mathbf{X}$ from a multivariate distribution with joint density function f can be explained as follows (Johnson 1987): A joint density function h is selected such that $ch(\mathbf{x}) \geq f(\mathbf{x})$ for any $\mathbf{x}$ in the support D of the function f, where c is some positive constant. The random vector sample $\mathbf{x}$ generated from the joint density function h is accepted as from f with probability $f(\mathbf{x})/[ch(\mathbf{x})]$. This procedure has been shown to generate vector samples with density f (Law and Kelton 2000).

One obvious drawback of this method is that the joint density f needs to be known. Other issues include finding an efficient h from which it is easy to generate observations, and computing the constant $c = \sup_{\mathbf{x} \in D} f(\mathbf{x})/h(\mathbf{x})$. There is a trade-off between choosing a complicated h and computing the constant c. Since samples are chosen on average in every c trials, a large c will also mean a large average generation time. As a rule of thumb, if D is finite, then a quick choice of h is the multivariate uniform; if D is infinite, a reasonable choice is to let h be the joint distribution of independent random variables with the same marginals as those of f. However, choosing an h is by no means trivial, and requires great care.

3.2 Conditional Distribution Based Methods

These methods factor the joint distribution into a set of conditional and marginal distributions that support easier data generation. Gelman and Speed (1993) provide a discussion of the combinations of marginal and conditional distributions that suffice to determine a joint distribution.

There are two such methods that merit attention from the random vector generation point of view. The first one samples a d-dimensional random vector $\mathbf{X} = (X_1, X_2, \ldots, X_d)'$ by generating $X_1 = x_1$ from its marginal distribution, then $X_2 = x_2$ from the conditional distribution of X_2 given $X_1 = x_1$, $X_3 = x_3$ from the conditional distribution of X_3 given X_1 and X_2, and so forth through the d components. This method guarantees that a joint distribution exists that corresponds to the generated random vector. The second method differs from the first one in that it requires conditional distributions of each component X_m given the other components $X_1, \ldots, X_{m-1}, X_{m+1}, \ldots, X_d$ for $m = 1, \ldots, d$. Arnold and Sarabia (2001) give sufficient conditions on the conditional distributions that guarantee that this method works.

The conditional distributions used are univariate, and hence the methods reduce the problem of generating a d-dimensional random vector into a series of d univariate generation problems. Thus, if appropriate sets of conditional distributions are easily specified, then variate generation is easy. However, despite their simple logic, both methods may

be difficult to implement because conditional distributions are not easy to derive except for special cases. Additionally, the joint distribution is assumed to be known so that the conditional distributions can be derived. However, if the objective is to fit the joint distribution to data, then the conditional distributions have to be also estimated from data, which can be difficult in practice.

3.3 The Method of Copulas

This method characterizes the joint distribution of a random vector $\mathbf{X}$ in terms of the marginal distributions F_i, $i = 1, 2, \ldots, d$, and the dependence structure defined by a *copula*. Copulas are joint distribution functions with uniform marginal distributions. Sklar (1959) shows that every joint distribution H with marginals F_i, $i = 1, \ldots, d$, can be written as

$$H(\mathbf{x}) = C[F_1(x_1), F_2(x_2), \ldots, F_d(x_d)],$$

where C is a copula. In general, the copula C is unique to an equivalence class and can be interpreted as the dependence structure of the joint distribution H. Copulas thus allows the dependence between the components of the random vector to be expressed in a manner unrelated to the marginal distributions.

The data generation scheme is simple: a sample $\mathbf{U} = (U_1, \ldots, U_d)$ is first generated from copula C, and then the random vector $\mathbf{X}$ is set to be $(F_1^{-1}(U_1), \ldots, F_d^{-1}(U_d))'$, where F_i^{-1} refers to the generalized inverse of the ith marginal distribution. The data in the second plot of Figure 1 are generated similarly by applying the so-called Gumbel copula

$$C_\beta(u, v) = \exp\left[-\{(-\log u)^{1/\beta} + (-\log v)^{1/\beta}\}^\beta\right]$$

to standard normal marginals F_1 and F_2. We choose the parameter $0 < \beta \leq 1$ so that the linear correlation of X_1 and X_2 is 0.7. We refer the reader to Nelsen (1999) for numerous parametric families of copulas that have been studied in the literature.

3.4 Bézier Distributions

Bézier curves are used extensively in modeling multidimensional functions and surfaces. Estimating joint distributions with Bézier curves allows various expert opinions to be incorporated into the model using, for example, visualization tools. Wagner and Wilson (1996) develop these modeling tools for the univariate and bivariate case and have shown them to be quite effective. The method can be extended to higher dimensions as well, but the parameter estimation will become increasingly harder: visualization tools can be

ruled out for instance. Wagner and Wilson (1996) use an alternate approach of formulating feasible, easily solvable mathematical programs to obtain a good fit in the bivariate case. A similar approach can possibly be employed in higher dimensions. This has not, to the best of our knowledge, been tried before and deserves attention.

3.5 Kernel Density Estimation

This method resamples from an observed sample of size n and then adds noise to these resampled data points. For the one-dimensional case, the noise is chosen to be a continuous random variable with zero mean and a small variance. Its probability density function is called the *kernel* and denoted by $k(x)$. This method is in essence a joint density estimation based procedure. The kernel estimate of the unknown probability density function of the observed sample, denoted by $\hat{f}$, is

$$\hat{f}(x) = \frac{1}{nb} \sum_{i=1}^{n} k\left(\frac{x - X_i}{b}\right), \qquad (3)$$

where b is a scale parameter called the *bandwidth*. The density estimate $\hat{f}$ approaches the true density function asymptotically as $n \to \infty$ in an appropriate norm. Notice that the density (3) is the equiprobable mixture of n noise distributions, each centered around one of the data points X_i, $i = 1, 2, \ldots, n$. The generation procedure is easy to implement: we first choose an observed sample point from an equally-likely draw of the n points and then sample from the kernel density associated with the chosen point. In this respect, this procedure is similar to using the smoothed bootstrap method. The kernel method enjoys the advantage that it is not even necessary to compute the density estimate explicitly once an appropriate kernel density k and bandwidth b are chosen.

This procedure extends transparently to d-variate random vectors. The central idea remains same: one resamples from available data and adds some noise to it. The joint density estimate is constructed as in (3) and it approaches the true joint density function asymptotically as $n \to \infty$. However, complications arise when implementing this method in the higher dimensional case as the choice of kernel k and bandwidth b is not as straightforward as it is in the one-dimensional case. Hörmann, Leydold, and Deflinger (2001) provide an algorithm that finds the best pair of kernel k and bandwidth b by minimizing a specific density function distance measure. This algorithm performs better than both naive resampling and fitting the normal distribution to data. Kernel density estimation is not widely utilized in the simulation input modeling literature, and despite the challenges it faces regarding parameter selection, it deserves

attention for problems with moderate dimensions where it has the potential to perform very well.

3.6 Special Families of Joint Distributions

A classical statistical approach for developing multivariate input models assumes a parameterized form for the unknown distribution and then determines the parameter values of the chosen form based on any available expert opinion or data. The distribution family is chosen based on factors such as its flexibility in modeling and ease in variate generation. The most popular such family in input modeling research of recent years is the Johnson's translation system (see Stanfield et al. 1996, Biller and Nelson 2003). It is defined by a d-dimensional normalizing translation

$$\mathbf{Z} = \boldsymbol{\gamma} + \boldsymbol{\delta} \, \mathbf{g}\left[\boldsymbol{\lambda}^{-1}(\mathbf{X} - \boldsymbol{\xi})\right] \sim N_d\left(\mathbf{0}, \Sigma\right), \qquad (4)$$

where $\mathbf{X}$ is the random vector we are interested in generating, $\mathbf{Z}$ is a d-dimensional multivariate normal random variable with mean $\mathbf{0}$ and a $d \times d$ covariance matrix Σ, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_d)'$ and $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_d)'$ are vectors of shape parameters, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_d)'$ is a vector of scale parameters, $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_d)'$ is a vector of location parameters, and the transformation $\mathbf{g}$ is given by

$$\mathbf{g}\left[(y_1, y_2, \ldots, y_d)\right] = \left[g_1(y_1), g_2(y_2), \ldots, g_d(y_d)\right],$$

where $g_i(y_i)$, $i = 1, 2, \ldots, d$, are defined by

$$g_i(y_i) = \begin{cases} \log(y_i) & \text{for the } S_L \text{ (lognormal) family,} \\ \sinh^{-1}(y_i) & \text{for the } S_U \text{ (unbounded) family,} \\ \log\left(\frac{y_i}{1-y_i}\right) & \text{for the } S_B \text{ (bounded) family,} \\ y_i & \text{for the } S_N \text{ (normal) family.} \end{cases}$$

Thus, the marginal distribution of each X_i is a univariate Johnson distribution (Johnson 1949). Notice that it is easy to sample from a d-dimensional Johnson-type distribution by using the normalizing translation defined in (4) when we are given the characterization of $\mathbf{Z}$ and the vectors of the Johnson-type parameters $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, $\boldsymbol{\xi}$, $\boldsymbol{\lambda}$, and $\mathbf{g}$.

These distributions have been found to work wonderfully well for applications reported in Stanfield et al. (1996). However, the control we have on the final form of the distribution used for sampling is limited to parameters $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, $\boldsymbol{\xi}$, and $\boldsymbol{\lambda}$, and that might be insufficient. For example, the Johnson's translation system provides good representations for unimodal distributions and can represent certain bimodal shapes, but not three or more modes.

Flexible families of distributions are, of course, not limited to the Johnson's translation system, e.g., Tukey's g and h transformations (Johnson 1987) and multivariate Pearson

distribution (Parrish 1990). We refer the reader to Johnson (1987) for more on specialized families of distributions.

4 CONSTRUCTING PARTIALLY SPECIFIED JOINT DISTRIBUTIONS

We now describe how multivariate input processes can be modeled using partial specifications, usually by specifying only the marginal distributions and some dependence measures. The goal is to construct input processes in higher dimensions efficiently in return for a compromise on the modeling front.

4.1 Transformation Based Methods

The methods we cover in this section work in essence on the same principle as the transformation-based univariate generation procedures, where a random variable with an arbitrary cdf F is sampled as the F^{-1} transformation of a uniform $(0, 1]$ random variable U. In the multivariate setting, we apply component-wise inverse marginal transformations to a multivariate base process with $U(0, 1]$ marginals. This procedure ensures that the marginal distributions of the resulting multivariate process are as specified. The challenge is to find a way to construct the multivariate base process so that uniform random variates are efficiently sampled and the process defined by the transformation has the desired dependence structure, which is mostly represented by either product-moment correlation or rank correlations.

4.1.1 ARTA, NORTA, and VARTA Processes

Autoregressive-To-Anything (ARTA) processes of Cario and Nelson (1996) define a time series with uniform marginals on $(0, 1]$ via the transformation $U_t = \Phi(Z_t)$, where the base process $\{Z_t; t = 0, 1, 2, \ldots\}$ is a stationary, standard, Gaussian autoregressive process of order p with the representation

$$Z_t = \sum_{h=1}^{p} \alpha_h Z_{t-h} + Y_t, \quad t = p+1, p+2, \ldots, n.$$

This autoregressive process is a special case of the vector linear process defined in (2). The α_h, $h = 1, 2, \ldots, p$, are fixed autoregressive coefficients that uniquely determine the autocorrelation structure of the base process, $\rho_Z(h)$, $h = 1, 2, \ldots, p$, and Y_t, $t = p+1, p+2, \ldots, n$, are i.i.d. Gaussian random variables with mean zero and variance σ_Y^2. The time series $\{X_t; t = 0, 1, 2, \ldots\}$ is obtained via the transformation $X_t = F^{-1}(U_t) = F^{-1}[\Phi(Z_t)]$ which ensures that X_t has distribution F by well-known properties of the inverse cumulative distribution function. Therefore, the central problem is to select the autocorrelation structure, $\rho_Z(h)$, $h = 1, \ldots, p$, for the base process Z_t that gives the desired

autocorrelation structure, $\rho_X(h)$, $h = 1, \ldots, p$, for the input process X_t. It is easily shown that the correlation $\rho_Z(h)$ depends only on the correlation $\rho_X(h)$. The determination of the dependence structure for the base process is thus equivalent to solving p independent correlation-matching problems.

A general, and related, method for obtaining random vectors with arbitrary marginal distributions and correlation matrix is described by Cario and Nelson (1997). The central idea is to transform a standard multivariate normal vector into the desired random vector. These vectors are referred as having a Normal-To-Anything (NORTA) distribution. Specifically, we let

$$\mathbf{X} = \left[F_1^{-1}[\Phi(Z_1)], F_2^{-1}[\Phi(Z_2)], \ldots, F_d^{-1}[\Phi(Z_d)] \right]',$$

be the input random vector with marginal distributions F_1, F_2, ..., F_d and correlation matrix $\Sigma_{\mathbf{X}}$, where the base vector $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_d)'$ is a standard multivariate normal vector with correlation matrix $\Sigma_{\mathbf{Z}}$. The problem then becomes finding a base correlation matrix $\Sigma_{\mathbf{Z}}$ for $\mathbf{Z}$ that implies the desired correlation matrix $\Sigma_{\mathbf{X}}$ for $\mathbf{X}$, a total of $d(d-1)/2$ independent correlation-matching problems.

The NORTA method can be considered as broadening the ARTA process beyond a common marginal distribution. Also, notice that the NORTA method is the same as the copula method of Section 3.3 when the marginal distributions are uniformly distributed on $(0, 1]$ and the base correlation matrix $\Sigma_{\mathbf{Z}}$ is positive definite.

The most recent addition to these transformation-based family of processes is the VARTA framework (Biller and Nelson 2003) that simply pulls together the theory behind ARTA and NORTA processes, and extends it to the multivariate time series by choosing the base process as the standardized Gaussian vector autoregressive process of order p (Wei 1990). For a fully specified VARTA process, there are $pd^2 + d(d-1)/2$ individual correlation-matching problems to solve. Notice that if $d = 1$, then the VARTA process reduces to an ARTA process; and if $d > 1$, but $p = 0$, the VARTA process corresponds to a NORTA vector.

The central problem to all of these three processes is the selection of the base dependence structure that is equivalent to solving the correlation-matching problems of the form $c(\rho_Z) = \rho_X$ for ρ_Z. The function c is defined as

$$\frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_1^{-1}[\Phi(z_1)] F_2^{-1}[\Phi(z_2)] \vartheta_{\rho_Z}(z_1, z_2) dz_1 dz_2 - \mu_1 \mu_2}{\sigma_1 \sigma_2}.$$

The function $c(\rho_Z)$ is nondecreasing for $-1 \leq \rho_Z \leq 1$ and continuous under fairly general conditions (Biller and Nelson 2003). And also, $|c(\rho_Z)| \leq \rho_Z$ (Lancaster 1957). Although it is not possible to solve the correlation-matching problems analytically except in special cases (Li and Ham-

mond 1975), these properties enable the development of efficient numerical procedures to solve them. We refer the reader to Cario and Nelson (1998), Chen (2001), and Biller and Nelson (2003) for such procedures.

The major drawback of using the methods of this section is that there exist sets of marginals with a feasible dependence structure that cannot be represented by any of the ARTA, NORTA, and VARTA transformations as the estimated base correlation matrix $\Sigma_{\mathbf{Z}}$ is not feasible for the base multivariate process. In this case, the procedures proposed by Lurie and Goldberg (1998) and Ghosh and Henderson (2002) can be applied to find a feasible approximation of the base correlation matrix.

4.1.2 Chessboard Distributions

The next two methods construct copulas which match a set of correlation values. These can also be used to model random vectors with arbitrary continuous marginals and given rank correlation values via component-wise transformations of the uniform random vector. This works because the inverse of continuous marginals are strictly increasing and hence preserve rank correlation values when applied component-wise to the uniform random vector.

Ghosh and Henderson (2002) propose a class of copulas that they call the chessboard distributions. The bivariate chessboard copula is illustrative of the general structure of chessboards. The support of the bivariate copula, the unit square, is divided into a grid of equally sized squares with sides of length $1/n$, where parameter n is said to index the level of discretization performed. The joint density of this copula is held constant over each square. Thus, the density over cell (i, j), $i, j = 1, 2, \ldots, n$, is defined as $f(x) = n^2 q(i, j)$, where $q(i, j)$ represents the probability that the random vector $\mathbf{X}$ takes values in the cell (i, j). Ghosh and Henderson (2002) show that the constraints on the probabilities $q(i, j)$ that ensure that f is a bivariate density function and match a given rank correlation are linear in the $q(i, j)$s. Thus, any given correlation matrices can be matched via the use of a linear program. A possible drawback of this method is that the number of linear programs that need to be solved can be very large. Nevertheless, these distributions can be important in modeling random vectors of moderate dimensions, and has the attractive property that it allows more control on the exact form of the joint distribution of interest.

4.1.3 Vine Copula Method

The vine copula method matches a given $d \times d$ rank correlation matrix to a copula. The first step is to construct a vine on the d components. A vine is a nested set of spanning trees, where the edges of tree j are the nodes of tree $j + 1$. The second step is to associate each edge on this vine with an appropriately chosen bivariate distribution. Kurowicka

and Cooke (2001) give a procedure to determine the correct bivariate distributions given the desired rank correlations. They also show that this procedure can match all feasible rank correlation matrices in the trivariate uniform vector case and can come very close to achieving any correlation matrix in higher dimensions. The vine copula method can be used to generate copula samples on the fly, and hence can be used as a base for a fast sampling procedure that matches a set of arbitrary continuous marginals and rank correlation matrix.

4.1.4 TES Processes

Melamed (1991) describe the Transfer-Expand-Sample (TES) process as a sequence of autocorrelated uniforms generated in an autoregressive manner to be used as a base process. In our opinion, TES is the only method that approaches the generality of the ARTA process for the case of generating stationary time series processes. Although the theory behind this method has not been extended to random vector and multivariate time series processes, its development has been a significant contribution to the time series input modeling literature.

The TES process can attain the full range of feasible lag-1 autocorrelations for a given marginal distribution and can frequently match autocorrelations at higher lags. TES^+ covers the positive lag-1 range $[0, 1]$ and TES^- covers the negative lag-1 range $[-1, 0]$. TES^+ is defined as

$$
U_t^+ = \begin{cases} U_0^+ & t = 0, \\ < U_{t-1}^+ + V_t > & t = 1, 2, \ldots, \end{cases}
$$

where the notation $< x >$ denotes modulo-1 arithmetic. TES^- is defined as

$$
U_t^- = \begin{cases} U_t^+ & t \text{ is even}, \\ 1 - U_t^+ & t \text{ is odd}, \end{cases}
$$

where U_0^+ is a $(0, 1]$ uniform random variable. Thus, the recursions written for U_t^+ and U_t^- define random variables with $(0, 1]$ uniform marginals whose autocorrelation structure depends only on the distribution of V_t, that is assumed to be independent of $\{U_0^+, U_1^+, \ldots, U_{t-1}^+\}$. Notice that extreme jumps may appear in the sample path due to the modula-1 arithmetic, i.e., U_{t-1} can be very close to 1 while U_t is very close to 0, or vice versa. The stitching transformation S_ξ, which is parameterized by ξ with $0 \le \xi \le 1$ as

$$
S_\xi(U_t) = \begin{cases} U_t/\xi & 0 \le U_t < \xi \\ (1 - U_t)/(1 - \xi) & \xi \le U_t < 1, \end{cases}
$$

can be used to mitigate this effect. The random variables $S_\xi(U_t)$, $t = 1, 2, \ldots$, still have $(0, 1]$ uniform marginals,

but no longer have extreme jumps. Unfortunately, stitching changes the autocorrelation structure of $S_\xi(U_t)$, and the change is not a simple function of ξ. TEStool allows the user to interactively change ξ and the distribution of V_t, and then displays the implied autocorrelation structure. The user tweaks the distribution until the autocorrelations of the input process match the desired autocorrelations. Experience is required to adjust the distribution in a systematic way. Regrettably, TEStool is no longer available.

4.2 Mixture Methods

In the bivariate case, the problem of interest here is to obtain a distribution with given marginals and correlation $\underline{\rho}_{ij} \le \rho \le \bar{\rho}_{ij}$, where $\underline{\rho}_{ij}$ and $\bar{\rho}_{ij}$ represent, respectively, the minimum and maximum correlations that can be induced between random variables X_i and X_j. The key idea is to mix the distributions corresponding to 0 correlation (independent) and an extremal correlation ($\underline{\rho}_{ij}$ or $\bar{\rho}_{ij}$). After the generation of the independent pair and the pair with the extreme correlation, we choose the independent pair with probability $1 - p$ and the highly correlated one with probability p. This results in a bivariate distribution with the same marginals and correlation $\rho = p\underline{\rho}_{ij}$ if $\rho < 0$ and correlation $\rho = p\bar{\rho}_{ij}$ if $\rho \ge 0$.

This method has been generalized to higher dimensions by Hill and Reilly (1994). To construct a d-variate random vector $\mathbf{X}$ with marginals F_i, $i = 1, 2, \ldots, d$, and a $d \times d$ correlation matrix $\Sigma = [\rho(i, j)]_{d \times d}$, Hill and Reilly define extreme correlation probability density functions (pdfs) by using probability mixtures or convex combinations of extremal distributions. An extreme correlation cdf is simply the joint distribution of X_i, $i = 1, 2, \ldots, d$, for which $\rho(i, j) = \bar{\rho}_{ij}$ or $\rho(i, j) = \underline{\rho}_{ij}$ for every $i < j$. They identify 2^{d-1} such extreme-correlation pdfs for $\mathbf{X}$. We let f_0 correspond to the joint pdf under the assumption of independence, i.e., $f_0(x_1, x_2, \ldots, x_d) = \prod_{i=1}^{d} f_i(x_i)$, and we let f_ℓ correspond to the extreme correlation pdf for $\ell = 1, 2, \ldots, d$. Now, we can define the joint composite probability mass function for $\mathbf{X}$ as

$$ f(\mathbf{x}) \;\; = \;\; \sum_{\ell=0}^{2^{d-1}} \lambda_\ell \, f_\ell(\mathbf{x}), $$

where $\sum_{\ell=0}^{2^{d-1}} \lambda_\ell = 1$ and $\lambda_\ell \ge 0$ for $\ell = 0, 1, \ldots, 2^{d-1}$. Here, the challenge is to find the composite probabilities λ_ℓ, $\ell = 1, 2, \ldots, 2^{d-1}$. Unfortunately, in most cases, there is an infinite number of solutions to this problem. Hence, an appropriate criterion is to be used to select a composite pdf with the desired characteristics. Exactly which criterion should be used is still an open question.

REFERENCES

Arnold, B. C. and J. M. Sarabia. 2001. Conditionally specified distributions. *Statistical Science* 16: 249–274.

Biller, B. and S. Ghosh. 2005. Multivariate Input Processes. In *Handbooks in Operations Research and Management Science: Simulation*, ed. B. L. Nelson and S. G. Henderson. Elsevier Science, Amsterdam. *Forthcoming*.

Biller, B. and B. L. Nelson. 2003. Modeling and generating multivariate time-series input processes using a vector autoregressive technique. *ACM Transactions on Modeling and Computer Simulation* 13: 211–237.

Block, H. W., N. A. Langberg, and D. S. Stoffer. 1990. Time series models for non-Gaussian processes. In *Topics in Statistical Dependence*, ed. W. H. Block, A. R. Sampson, and T. H. Savits, 69–83. Heyward, California.

Cario, M. C. and B. L. Nelson. 1996. Autoregressive to anything: Time-series input processes for simulation. *Operations Research Letters* 19: 51–58.

Cario, M. C. and B. L. Nelson. 1997. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL.

Cario, M. C. and B. L. Nelson. 1998. Numerical methods for fitting and simulating autoregressive-to-anything processes. *INFORMS Journal on Computing* 10: 72–81.

Chen, H. 2001. Initialization for NORTA: Generation of random vectors with specified marginals and correlations. *INFORMS Journal on Computing* 13: 312–331.

Embrechts, P., A. McNeil, and D. Straumann. 2002. Correlation and dependence in risk management: properties and pitfalls. In *Risk Management: Value at Risk and Beyond*, ed. M. A. H. Dempster, 176–223. Cambridge: Cambridge University Press.

Gelman, A. and T. P. Speed. 1993. Characterizing a joint probability distribution by conditionals. *Journal of the Royal Statistical Society* 55: 185–188.

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

Hill, R. R. and Reilly, C. H. 1994. Composition for multivariate random vectors. In *Proceedings of the 1994 Winter Simulation Conference*, ed. D. A. Sadowski, A. F. Seila, J. D. Tew, and S. Manivannan, 332–339. Piscataway, New Jersey: IEEE Press.

Hörmann, W., J. Leydold, and G. Deflinger. 2001. Universal methods of non-uniform random variate generation. Technical Report, University of Economics and

Business Administration, Dept for Applied Statistics, Vienna, Austria.

Johnson, N. L. 1949. Systems of frequency curves generated by methods of translation. *Biometrika* 36: 297–304.

Johnson, M. E. 1987. *Multivariate Statistical Simulation*. New York: John Wiley and Sons.

Kurowicka, D. and R. M. Cooke. 2001. Conditional, partial, and rank correlation for the elliptical copula; dependence modeling in uncertainty analysis. Technical Report, Delft University, Delft, Netherlands.

Lancaster, H. O. 1957. Some properties of the bivariate normal distribution considered in the form of a contingency table. *Biometrika* 44: 289–292.

Law, A. M. and W. D. Kelton. 2000. *Simulation Modeling and Analysis*. 3d ed. New York: McGraw-Hill.

Li, S. T. and J. L. Hammond. 1975. Generation of pseudo-random numbers with specified univariate distributions and correlation coefficients. *IEEE Transactions on Systems, Man, and Cybernetics* 5:557–561.

Livny, M., B. Melamed, and A. K. Tsiolis. 1993. The impact of autocorrelation on queueing systems. *Management Science* 39: 322–339.

Lurie, P. M. and M. S. Goldberg. 1998. An approximate method for sampling correlated random variables from partially-specified distributions. *Management Science* 44: 203–218.

Mallows, C. L. 1967. Linear processes are nearly Gaussian. *Journal of Applied Probability* 4: 313–329.

Melamed, B. 1991. TES: A class of methods for generating autocorrelated uniform variates. *ORSA Journal on Computing* 3: 317–329.

Melamed, B., J. R. Hill, and D. Goldsman. 1992. The TES methodology: Modeling empirical stationary time series. In *Proceedings of the 1992 Winter Simulation Conference*, ed. J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson, 135–144. Piscataway, New Jersey: IEEE Press.

Nelsen, R. B. 1999. *An Introduction to Copulas*. New York: Springer-Verlag.

Parrish, R. S. 1990. Generating random deviates from multivariate Pearson distributions. *Computational Statistics and Data Analysis* 9: 283–295.

Sklar, A. 1959. Fonctions de répartition à *n* dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris* 8: 229–231.

Stanfield, P. M., J. R. Wilson, G. A. Mirka, N. F. Glasscock, J. P. Psihogios, and J. R. Davis. 1996. Multivariate input modeling. In *Proceedings of the 1996 Winter Simulation Conference*, ed. J. M. Charnes, D. J. Morrice, D. T. Brunner and J. J. Swain, 1457–1464. Piscataway, New Jersey: IEEE Press.

Vincent, S. 1998. Input data analysis. In *Handbook of Simulation*, ed. J. Banks, 55–91. New York: John Wiley & Sons.

Wagner, M. A. F. and J. R. Wilson. 1996. Recent developments in input modeling with Bézier distributions. In *Proceedings of the 1996 Winter Simulation Conference*, ed. J. M. Charnes, D. J. Morrice, D. T. Brunner and J. J. Swain, 1448–1456. Piscataway, New Jersey: IEEE Press.

Ware, P. P., T. W. Page, and B. L. Nelson. 1998. Automatic modeling of file system workloads using two-level arrival processes. *ACM Transactions on Modeling and Computer Simulation* 8: 305–330.

Wei, W. W. S. 1990. *Time Series Analysis: Univariate and Multivariate Methods*. New York: Addison Wesley.

AUTHOR BIOGRAPHIES

BAHAR BILLER is an assistant professor of Operations Management in Tepper School of Business at Carnegie Mellon University. Her research interests include computer simulation of stochastic systems, queueing processes, and global supply chain design.

SOUMYADIP GHOSH is a Research Staff Member at the IBM T. J. Watson Research Center. He completed his doctoral research in multivariate input modeling at Cornell University.