S.G. Henderson and B.L. Nelson (Eds.), *Handbook in OR & MS, Vol. 13*

DOI: 10.1016/S0927-0507(06)13005-4

Chapter 5
Multivariate Input Processes

Bahar Biller
Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213, USA
E-mail: billerb@andrew.cmu.edu

Soumyadip Ghosh
Mathematical Sciences Department, IBM T.J. Watson Research Center,
Yorktown Heights, NY 10598, USA
E-mail: ghoshs@us.ibm.com

Abstract

Representing uncertainty in a simulation study is referred to as input modeling, and is often characterized as selecting probability distributions to represent the input processes. This is a simple task when the input processes can be represented as sequences of independent random variables with identical distributions. However, dependent and multivariate input processes occur naturally in many service, communications, and manufacturing systems. This chapter focuses on the development of multivariate input models which incorporate the interactions and interdependencies among the inputs for the stochastic simulation of such systems.

1 Introduction

An important step in the design of stochastic simulation is input modeling, which refers to modeling the uncertainty in the input environment of the system being studied. Input modeling is often thought of as limited to selecting appropriate univariate probability distributions to represent the primitive inputs of a simulation, and this would indeed be true if the relevant input processes could be represented as independent sequences of identically distributed random variables. When such univariate models are appropriate, a number of generation procedures and tools that support automated input modeling are available; Chapter 4 provides a good overview.

Often, however, univariate models fail to adequately capture the effect of dependencies between input processes that occur naturally in different forms in many service, communications, and manufacturing systems; Melamed et al. (1992) and Ware et al. (1998) give good examples. The ability to capture these

dependencies is crucial because inaccurate input models often lead to performance estimates that are seriously in error. This is illustrated powerfully in a straightforward example provided by Livny et al. (1993), who examine the impact of autocorrelated interarrival times on the mean waiting time of a single-server queue. They simulate an $M/M/1$ queue process with independent sequences of exponential interarrival and service times and observe that the estimated mean waiting time changes dramatically with the introduction of autocorrelation in the interarrival process, showing increases by a factor of two orders of magnitude when the autocorrelation is set high. This shows that the independence assumption could lead to very poor estimates of performance measures in the presence of dependence. It is thus imperative to develop simulation input models that incorporate dependence. Our goal here is to present a coherent narrative of the central principles that underlie many promising methods available for constructing dependent input models.

We consider models that can be broadly classified into two groups. The first group produces a sequence of independent samples of a *random vector* $\mathbf{X} = (X_1, X_2, \ldots, X_d)'$, which is a finite collection of d random components. Each component is a real-valued random variable and is associated with a univariate distribution function called its *marginal* distribution. We associate with the random vector a probability distribution function, called its *joint distribution function*, in the space $\Re^d$. This joint distribution completely defines the stochastic (and hence dependence) properties of $\mathbf{X}$. The second group of input models we consider in this chapter captures dependence that arises between subsequent samples of a sequence. Sequences of samples are frequently indexed over time and stochastic processes that exhibit this *temporal* dependence are called *time series*. In this chapter, we use the term time series to denote a sequence of random variables $\{X_t; t \geqslant 1\}$ indexed by a discrete set $t = 1, 2, \ldots$. For example, the month-to-month order quantities for a product placed by a customer can be considered as a univariate time series. The time-series process is a stochastic process with a probability distribution on the space of all possible path realizations. Such a probability distribution completely describes the dependence properties of the stochastic process $\{X_t; t \geqslant 1\}$. However, almost all statistical and engineering applications specify and analyze time series in terms of the probability distributions of the individual random elements X_t and the *autocorrelation* structure of the sequence, which (refer to Section 4.2) is a measure that tries to capture temporal dependencies in sufficient detail.

Simulation input modeling literature focuses on random vectors and time series as separate cases, although the underlying concepts are often similar. We maintain this distinction while describing the methods, but strive to present them as different avatars of the same central concept. A recent unifying work (Biller and Nelson, 2003) represents random vector and time series as special cases of the more general framework of *multivariate time series*. This denotes a time series where the elements $\mathbf{X}_t$ are finite d-dimensional random vectors with distributions in $\Re^d$, and thus represent input processes where dependence exists both over time and among the components of the input process. For ex-

ample, a distributor has several warehouses and each places monthly orders for a product. The month-to-month dependence still exists, but there may also be dependence between the orders from different warehouses in the same month if they are able to share inventory or supply the same customers.

A multivariate input modeling procedure can be judged *good* or *effective* based on a number of criteria that can be grouped into two main categories, those that judge its modeling capabilities and the effectiveness of its sampling procedures respectively. Any model needs to perform well in both.

From the modeling point of view, a method should firstly be able to represent a broad class of real-life situations. For example, a procedure that models random vectors to have multivariate- or joint-normal distributions may not be appropriate in situations where the true input process is known to be either multi-modal, nonsymmetric, skewed or heavy-tailed. The goal of representing a wide variety of situations is best achieved when the procedure works directly with a complete characterization of the input process, which would be the joint distribution in the case of random vectors and in the more general case of a multivariate time series a probability distribution on the space of all possible $\Re^d$ sample path realizations of the process. Some random vector modeling procedures indeed work on this principle (Section 2) or with a more modest aim of modeling the random vector as a member of an appropriately chosen family of joint distributions (Section 3).

Modeling by matching the joint distribution, though sound from a modeling power point of view, can be quite restrictive in practice. This is primarily due to the quantity and quality of information needed to adequately fit a joint distribution to the situation at hand. The information needed could be data or expert opinion to estimate the true joint distribution function or the parameters of a chosen model distribution function. In light of this difficulty, most of the focus in input-modeling research has been on the development of methods that match only certain key properties of the input process (Section 4), mainly its marginal distributions and an aptly chosen dependence measure. This approach is not completely general, but one can be reasonably certain that the chosen properties capture the effect of dependence to a satisfactory degree in the model. Moreover, these models are presumably easier to construct from data since either fewer parameters need to be estimated or the estimates needed are easily obtained from available data.

From a sampling point of view, the sampling schema associated with a procedure should be easy to implement on a computer and fast in generating samples. Thus, it should avoid complex, computationally intensive function evaluations, which might be needed in many of the approaches in Sections 2 and 3.

One should additionally be able to easily validate the success of a procedure in representing input dependencies once a sample of data is generated. This statistical model validation is an important aspect of the input model development but it will not be covered here; texts such as Law and Kelton (2000)

provide a detailed discussion of issues such as goodness of the fit and validation.

The methods discussed in this chapter are presented in descending order of control we have on the stochastic description of the multivariate input processes. In Sections 2 and 3 this discussion is essentially centered around random vector processes and their joint distributions. Much of the time series modeling literature focuses on processes that are partially specified with distributions for individual elements and autocorrelations. These methods have much in common with random vector methods that work with similar partial specifications and are therefore discussed in a unified manner in Section 4. We conclude with promising areas for input modeling research in Section 5.

2 Constructing full joint distributions

The *joint* (or *multivariate*) *cumulative distribution function* (c.d.f.) of a d-dimensional random vector $\mathbf{X} = (X_1, X_2, \ldots, X_d)'$ is a nondecreasing d-dimensional real-valued function F that takes values in $[0, 1]$ and is right-continuous in each argument. The joint distribution function completely characterizes the stochastic behavior of $\mathbf{X}$. Random vector $\mathbf{X}$ is said to have a *joint* (or *multivariate*) *probability density function* (p.d.f.) f, if a nonnegative, integrable function f exists such that

$$F(\mathbf{x}) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_d} f(u_1, u_2, \ldots, u_d)\, \mathrm{d}u_1\, \mathrm{d}u_2 \cdots \mathrm{d}u_d.$$

In this section, we present different approaches that construct input models with fully specified joint c.d.f.'s or p.d.f.'s. Although these methods are typically used for constructing random vectors and generating data from the pre-specified joint cumulative distribution functions, they can be applied to time-series processes of fixed lengths as well.

Some methods assume that such a joint distribution is explicitly given (Sections 2.1 and 2.3) while some assume that it is available in a special form (Section 2.2). These requirements entail knowledge of a multidimensional function, which can be a major drawback especially if the dimensions involved are large. Other methods reviewed (Sections 2.4 and 2.5) take the alternative approach of obtaining nonparametric estimations of the joint distribution from available historical data. In particular, Bézier curves (Section 2.4) allow the inclusion of various types of expert opinions into the model development using visualization tools, while kernel density estimators (Section 2.5) provide a generalization and improvement over histograms.

2.1 Acceptance/rejection method

The acceptance/rejection method has been used extensively for generating univariate data, but its impact on multivariate data generation has been muted.

This is partly due to the relatively limited attention researchers have paid to multivariate generation, but also because significant practical difficulties exist in the implementation of the method.

The acceptance/rejection principle has a long history; Marsaglia and Bray (1964) is an early reference. To generate random vector $\mathbf{X}$ from a multivariate joint density function $f(\cdot)$, first a joint density $h(\mathbf{x})$ is selected such that $ch(\mathbf{x})$ dominates $f(\mathbf{x})$, i.e., $ch(\mathbf{x}) \geqslant f(\mathbf{x})$, for any $\mathbf{x}$ in the domain D of the function f, where c is some positive constant. A random vector sample generated from $h(\mathbf{x})$ is then accepted as a sample from $f(\mathbf{x})$ with probability $f(\mathbf{x})/[ch(\mathbf{x})]$. It is straightforward to show that this procedure generates vector samples with density f; a proof of this result is available in Appendix 8A of Law and Kelton (2000).

Two important challenges arise in implementing this method: finding a dominating density h which is easy to sample, and computing the constant $c = \sup_{\mathbf{x}\in D} f(\mathbf{x})/h(\mathbf{x})$. If D is bounded, then one can choose the multivariate uniform density function as the dominating h. Otherwise, a reasonable choice is to let h be the joint distribution of independent random variables with the same marginals as those of f. However, for these choices of h, the constant c (which also represents the expected number of h-samples needed to obtain an f-sample, and hence the average time needed to generate one f-sample) increases rapidly as the dimension increases. A more complicated h might be more efficient, but constructing h, generating data from h and computing c may be more difficult and expensive. Gilks and Wild (1992), Hörmann (1995) and Leydold (1998) suggest the transformed-density-rejection method to construct a dominating function, in which one uses a monotonic function to transform f into a concave function, takes the minimum of several tangent hyper-planes, and then transforms it back into the original scale.

The choice of the dominating function h is by no means a trivial one, and one needs to exercise great care. Lack of a codified procedure to obtain efficiently sampled dominating functions makes this method unappealing even in cases of moderate dimensional random vectors.

2.2 *Conditional distributions*

These methods factorize the joint distribution into a set of conditional and marginal distributions that easily yields samples with the desired joint distributional properties. The key idea is to reduce the problem of generating a d-dimensional random vector into a series of smaller multivariate, and often d univariate, generation problems. Thus, these methods can utilize the vast body of techniques available for univariate generation problems.

There are a large number of ways in which joint distributions can be specified in terms of conditional and marginal distributions, and such a specification can arise naturally in many systems through its dynamics. Gelman and Speed (1993) provide a discussion of the combinations of marginal and conditional distributions that ensure the existence of a (unique) joint distribution. We shall

outline two such approaches that we feel are attractive from a random vector generation point of view.

The first approach (Rosenblatt, 1952) factorizes the joint density function $f(\mathbf{x})$ of the d-dimensional random vector $\mathbf{X} = (X_1, X_2, \ldots, X_d)'$ into a marginal and $(d-1)$ conditional distributions as follows:

$$f(\mathbf{x}) = f_{X_1}(x_1) \prod_{m=2}^{d} f_{X_m|X_1,\ldots,X_{m-1}}(x_m|x_1, \ldots, x_{m-1}), \tag{1}$$

where f_{X_1} is the marginal density of the first component X_1 and $f_{X_m|X_1,\ldots,X_{m-1}}$ is the conditional density of the mth component X_m, given the first $m-1$ components $X_1, \ldots, X_{m-1}$. Any such set of conditional distributions always defines a valid joint distribution via (1). Notice that as long as the structure of the factorization remains the same, the order of the conditioning has no impact on the resulting input model: the first factor on the right-hand side of the factorization might as well correspond to the dth component.

Using the factorization (1), random vector $\mathbf{X}$ can be sampled by first generating $X_1 = x_1$ from the marginal distribution f_{X_1} of X_1, then generating $X_2 = x_2$ from the conditional distribution of X_2 given $X_1 = x_1$, $f_{X_2|X_1}$, and so forth through the d components. Despite its simple logic, this method may be difficult to apply as conditional distributions are not easy to derive except for some special cases. Such easily factored multivariate distributions include the multivariate normal (Tong, 1990), the Cauchy and the Burr (Johnson and Kotz, 1972). In cases where the joint distribution is not known, the conditional distributions have to be obtained either from an estimate of the joint distribution or estimated directly.

Another class of conditional distribution based methods that has drawn some academic interest in recent years can be of potential interest to the simulation practitioner. Arnold et al. (2001) provide a good survey of this approach, where the joint distribution is specified in terms of the conditional distribution of the mth component given all the other components $X_1, \ldots, X_{m-1}, X_{m+1}, \ldots, X_d$, $m = 1, \ldots, d$. We denote the associated conditional distribution by $F_{m|-m}(x_m|x_1, \ldots, x_{m-1}, x_{m+1}, \ldots, x_d)$. The sample generation procedure is simple. The $(n+1)$st sample $\mathbf{x}^{n+1}$ is obtained using $\mathbf{x}^n$ of the previous step as follows:

(i) x_1^{n+1} is sampled from $F_{1|-1}(\cdot|x_2^n, \ldots, x_d^n)$.
(ii) For $m = 2, \ldots, d-1$, x_m^{n+1} is sampled from $F_{m|-m}(\cdot|x_1^{n+1}, \ldots, x_{m-1}^{n+1}, x_{m+1}^n, \ldots, x_d^n)$.
(iii) x_d^{n+1} is sampled from $F_{d|-d}(\cdot|x_1^{n+1}, \ldots, x_{d-1}^{n+1})$.

This procedure is known as the Gibbs sampling technique (Geman and Geman, 1984) and is primarily driven by the theory of Markov chains. Under fairly reasonable conditions, the joint distribution of the sample points $\mathbf{x}^n$, $n \geqslant 1$, can be shown to converge to a limiting stationary joint distribution

geometrically fast (under a certain norm). This limiting distribution is unique if all of the conditional distributions have the same support (i.e., the subset over which the functions are nonzero in $\Re$). However, the convergence to a limiting distribution is not guaranteed for any arbitrary set of conditional distributions, unlike the earlier approach, where the factorization (1) ensures the correct joint distribution produced. Gelman and Speed (1993) discuss sufficient conditions for the convergence in the bivariate case. Although compatible conditional distributions can be chosen with relative ease in lower dimensions, this problem will likely become intractable in higher dimensions.

Independent of the approach used, it gets harder to derive the conditional distributions from a joint distribution as the dimension of the input process increases. Additionally, conditional distributions are hard to understand in higher dimensions; therefore using expert opinion in the context of these methods with increasing dimensions will be difficult. These issues limit the usage of this method in higher dimensional random vector sampling beyond special cases where the conditional distributions arise naturally.

2.3 *Method of copulas*

The distinguishing feature of the method of this section is its use of a family of distributions called *copulas*. A copula may be thought of in two equivalent ways: as a function that maps points in the unit hypercube in $\Re^d$ to values in the unit interval or as a multivariate joint distribution function with standard uniform marginal distributions. Sklar (1959) shows that every joint distribution H with marginals F_i, $i = 1, \ldots, d$, can be written as

$$H(x_1, x_2, \ldots, x_d) = C\big(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)\big), \tag{2}$$

where C is a copula that is uniquely defined if the marginals F_i are continuous, and is unique to an equivalence class in the case of discrete marginals. In either case, C can be interpreted as the dependence structure of H. Thus, we can transform the problem of estimating the joint distribution from a given data set to that of estimating a function in the unit hypercube. The copula can be determined empirically or can belong to an appropriately chosen parametric family. However, the central issue of estimating a function in d variates remains, and hence these methods can very often be haunted by the curse of dimensionality.

The main advantage of a copula is that it remains invariant under strictly increasing transformations of its component variables, simplifying the random vector generation. To sample a random vector $\mathbf{X} = (X_1, X_2, \ldots, X_d)'$ with joint distribution H as defined in (2), we first generate a multivariate uniform vector $\mathbf{U} = (U_1, U_2, \ldots, U_d)'$ from copula C and then set the random vector $\mathbf{X}$ to be $(F_1^{-1}(U_1), \ldots, F_d^{-1}(U_d))'$, where $F_i^{-1}(u) = \inf\{x\colon\ F_i(x) \geqslant u\}$ is the generalized inverse of F_i. Thus, copulas provide an easy method to model and generate random vectors when the modeler believes that the dependence between the components of the random vector can be expressed independently of the marginal distributions. Later in Sections 4.3.2 and 4.3.3, we shall discuss

the methods that use copulas with a special structure as a base for a fast sampling procedure that match partial specifications of random vectors.

Normal copulas are commonly used in the literature for summarizing the dependence structure of a random vector since one can calibrate them to yield pre-specified pair-wise product-moment or rank correlations (Section 4.1); see Clemen and Reilly (1999) for an example. This kind of dependence modeling is also becoming increasingly popular in cases where linear correlation falls short of capturing complex interactions and interdependencies among the model inputs. Further, numerous other parametric families of copulas can be coupled with arbitrary marginal distributions without worries about consistency. Joe (1997) and Nelsen (1999) provide a survey of some extensively studied parametric families of copulas.

2.4 Bézier distributions

This subsection presents a graphical, interactive technique for modeling simulation input processes whose distributions are based on Bézier curves used extensively in computer graphics to approximate smooth functions on bounded intervals. The distinguishing feature of this type of modeling is that since low dimensional Bézier curves can be easily visually modified, various types of expert opinions can be incorporated into the model development using visualization tools. Wagner and Wilson (1995, 1996) develop such tools for the univariate and bivariate cases and show that these are quite effective, especially when few historical data points are available from the existing system. This method is limited to the estimation of continuous joint distributions with bounded support.

We present the key idea behind the construction of Bézier distributions considering the simplest case of representing a univariate, independent process with a marginal distribution denoted by F. To estimate the Bézier approximation $\widehat{F}$ for the true distribution function F, we first define a Bézier curve in two dimensions via the parametric bivariate function

$$\begin{aligned}\mathbf{P}(t) &= \left[P_{\mathbf{x}_1}(t; n, \mathbf{x}_1), P_{\mathbf{x}_2}(t; n, \mathbf{x}_2)\right] \\ &= \sum_{i=0}^{n} B_{n,i}(t)\mathbf{p}_i,\end{aligned}$$

where $t \in [0, 1]$, $\mathbf{x}_1 = (x_{10}, x_{11}, \ldots, x_{1n})'$, $\mathbf{x}_2 = (x_{20}, x_{21}, \ldots, x_{2n})'$ and $\{\mathbf{p}_i = (x_{1i}, x_{2i})'; i = 0, 1, \ldots, n\}$ are $n + 1$ bivariate control points. The control points can be thought of as magnets attracting the curve $\mathbf{P}$ to themselves. The attraction is quantified in terms of the Bernstein polynomials $\{B_{n,i}(t); i = 0, 1, \ldots, n\}$ (Farin, 1990). The attraction from the ith control point is the strongest at i/n, where $B_{n,i}(t)$ attains its maximum value. The curve passes through $\mathbf{p}_i$ if $B_{n,i} = 1$, which is true for the endpoints $\mathbf{p}_0$ and $\mathbf{p}_n$.

We can now define the Bézier approximation $\widehat{F}$ for F via the curve $\mathbf{P}$ as

$$x(t) = P_{\mathbf{x}_1}(t; n, \mathbf{x}_1) = \sum_{i=0}^{n} B_{n,i}(t) x_{1i},$$

$$\widehat{F}\big[x(t)\big] = P_{\mathbf{x}_2}(t; n, \mathbf{x}_2) = \sum_{i=0}^{n} B_{n,i}(t) x_{2i}$$

for all $t \in [0, 1]$. Further, the density function estimate $\hat{f}$ is given parametrically for $x(t)$ as

$$\hat{f}\big[x(t)\big] = \frac{\sum_{i=0}^{n-1} B_{n-1,i}(t) \Delta x_{1i}}{\sum_{i=0}^{n-1} B_{n-1,i}(t) \Delta x_{2i}}$$

for all $t \in [0, 1]$, where $\Delta x_{ji} = x_{j(i+1)} - x_{ji}$ for $i = 0, 1, \ldots, n-1$ and $j = 1, 2$. It is straightforward to generate a Bézier random variate with this density via inversion. After a uniformly distributed random number $u \in [0, 1]$ is obtained, the equality $P_{\mathbf{x}_2}(t_u; n, \mathbf{x}_2) = u$ is solved for $t_u \in [0, 1]$ and the random variate x is set to $P_{\mathbf{x}_1}(t_u; n, \mathbf{x}_1)$.

Wagner and Wilson (1995) provide two approaches for obtaining the Bézier density estimate $\hat{f}$ when the functional form of the true distribution is not known. The first approach finds a Bézier approximation whose distributional properties match expert opinion using a visualization software. This software allows the user to choose appropriate control points, and then makes sure that $\hat{f}$ satisfies the density validity conditions. The second approach is used when historical data are available. An empirical distribution F_n can be constructed from the data sample and one can then minimize the distance between F_n and $\widehat{F}$ in any functional distance norm subject to the constraints that $\widehat{F}$ has the extreme values 0 and 1, and that $\hat{f}$, the density function defined above, is nonnegative.

The extension of the key idea to the characterization of the multivariate case is straightforward. The parametric function $\mathbf{P}$ in $d + 1$ dimensions,

$$\mathbf{P}(t) = \big[P_{\mathbf{x}_1}(t; n, \mathbf{x}_1), P_{\mathbf{x}_2}(t; n, \mathbf{x}_2), \ldots, P_{\mathbf{x}_d}(t; n, \mathbf{x}_d), P_{\mathbf{x}_{d+1}}(t; n, \mathbf{x}_{d+1})\big],$$

is used to derive a d-dimensional joint distribution function and the density function estimate can again be obtained in several ways. However, parameter estimation would become increasingly difficult. Using visualization tools becomes noticeably harder even in the bivariate case (Wagner and Wilson, 1996) and can be ruled out in higher dimensions. Wagner and Wilson (1996) suggest some simple easily solvable optimization formulations to obtain a proper fit in the bivariate case. These can be extended to higher dimensions, but care must be given to the formulation to ensure feasibility and computational ease. Nevertheless, to the best of our knowledge, the extension of these methods to higher-dimensional generation has not been tried and it deserves attention.

2.5 Kernel density estimation

These methods estimate joint densities using kernel density estimates. They are also well suited for efficient multivariate generation in cases where the kernel estimators are carefully chosen. We note that generating random variates from a density estimate by using kernel estimates is the same as the smoothed bootstrap, the subject of Chapter 14.

The literature credits Rosenblatt (1956) for proposing this method and Parzen (1962) for providing an early framework. We refer the reader to Devroye and Györfi (1985) and Silverman (1982) for a good introduction. These methods do not seem to be very popular in the simulation input modeling literature, and this is perhaps due to the fact that selecting the right parameters to obtain good estimates of joint distributions with these methods can increasingly be an art. Some recent attention to these methods, however, suggest that they have the potential to perform very well for problems with moderate dimensions.

The kernel density estimate of an unknown d-dimensional joint density function f is obtained from an observed d-dimensional data sample of size n by averaging an appropriately scaled *noise* density centered around each of the data points. (Although we refer here to density estimates, this procedure can be applied to estimate c.d.f.'s using kernel distributions.) The noise is chosen to be a multivariate random vector whose d-variate probability density function is called the *kernel*, denoted by $K_{\mathbf{H}}$ and defined as

$$K_{\mathbf{H}}(\mathbf{x}) = |\mathbf{H}|^{-1/2} K\big(\mathbf{H}^{-1/2}\mathbf{x}\big),$$

where $\mathbf{H}$ is a symmetric positive definite $d \times d$ matrix called the bandwidth matrix. The density function K is assumed to be symmetric around the origin, and we shall henceforth consider this condition to be a part of the definition of the kernel density function. In its most general form, the d-dimensional kernel density estimator is (Deheuvels, 1977)

$$\hat{f}_n(\mathbf{x}; \mathbf{H}) = \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_i).$$

This probability density function is the equiprobable mixture of n noise distributions, each centered around one of the data points. Sampling from this density estimate is achieved by resampling from the original n data points, each being equally likely, and then adding some noise to the selected data point in the form of a sample from the scaled kernel $K_{\mathbf{H}}$ associated with it. Notice that the generation procedure does not need an explicit calculation of the kernel density estimate.

Under certain conditions on the bandwidth parameter $\mathbf{H}$ and the target p.d.f. f, the kernel density estimate $\hat{f}_n$ converges asymptotically to the true density in the integrated mean squared error sense as n grows. Thus, the generation procedure produces samples asymptotically consistent with f, and the

kernel density and the bandwidth parameters that minimize the integrated mean squared error is the best choice for the given n-sized sample. However, the selection of such a pair has its challenges in practice. In the univariate case, where the minimization problem has been analytically solved (Wand and Jones, 1995), it is often better from a generation-effort perspective to choose a near-optimal kernel that is straightforward to sample. Although the literature sees a lot of attention paid to this issue there is no generally accepted procedure for specifying the bandwith and kernel parameters. We refer the reader to Wand and Jones (1995) for a good overview and Silverman (1982) for quick-and-dirty estimates of these parameters. These estimates can be far from optimal for many distributions. In such a case, Hörmann et al. (2001) suggest the use of the normal kernel together with a variance-correction algorithm. It is based on the fact that fitting a multivariate normal distribution is optimal if the unknown distribution is normal (with $|\mathbf{H}| \to \infty$), and naive resampling is optimal if the unknown distribution is not continuous (with $|\mathbf{H}| = 0$). In other cases, it would intuitively seem that intermediate values of $|\mathbf{H}|$ lead to a better approximation of the unknown distribution. So, even if the guess of $|\mathbf{H}|$ is far from optimal, it is still very likely that it is better than using naive resampling or fitting the normal distribution for most continuous distributions.

Although the asymptotic approximation properties of the kernel density estimation procedure are theoretically well understood as compared to many other density estimation procedures, the problem of selecting an efficient pair of kernel density and bandwidth parameters in the d-dimensional case remains to be addressed. Preliminary evidence indicates that the method performs satisfactorily in moderate dimensional input processes.

3 Parametric families of joint distributions

As underscored in the previous section, the quantity and the quality of expert opinion or data needed to satisfactorily model a multivariate input process with a fully specified joint distribution might turn the input model development into a formidable task even in moderate dimensions. A classical approach around this problem is to assume a parameterized form for the unknown joint distribution and then determine the parameter values based on any available expert opinion or historical data. However, many of the standard distributions available in literature are insufficient for Monte Carlo applications. Johnson (1987) provides a comprehensive list of these insufficiencies, where he notes that often these families are insufficiently flexible and require formidable computational power to fit and generate samples. These limitations should not be viewed as grounds to abandon these specific distributions entirely, and clearly ample opportunity exists to make significant contributions to this approach.

Many of the parametric families of the multivariate distributions are tied directly to the multivariate normal distribution, which has played a dominant role in both theoretical and applied statistics since the time of Laplace. The

normal curve cannot provide adequate representation for many of the input processes encountered in practice, and there exists a significant body of work exploring reasonable departures from the multivariate normal distribution. For example, *(elliptically) symmetric distributions* generalize the symmetric form of the multivariate normal distribution to distributions that possess (elliptically) symmetric contours, while maintaining the advantage of being easy to sample. Fang et al. (1990), Joe (1997) and Johnson (1987) provide a good introduction to the properties of symmetric distributions.

The constraint that the underlying joint distributions possess symmetric contours is also found to be quite restrictive in practice. A class of distributions called the systems of skew frequency curves has been developed to overcome the limitation of the multivariate normal distribution in representing skewed processes. Recall that the probability density function of a multivariate normal random variable is completely determined by the first two moments; therefore, it is not possible to have a skewness factor that could have an impact on the shape of the distributions. The systems of skew frequency curves are constructed from a multivariate normal random vector by applying component-wise transformations and include the t-distribution, the exponential power distribution, and the log-normal distribution. Similar distributions in common use are the Pearson-type distributions proposed by Pearson (1895) and Charlier (1906), and those proposed by Edgeworth (1898), who uses transformations which can be represented by polynomials.

In the last few years, a translation system developed by Johnson (1949a) has been one of the most popular flexible system of distributions used in simulation applications (e.g., see Stanfield et al., 1996; Mirka et al., 2000; Biller and Nelson, 2003). Johnson (1987) is a good source of information on many members of the so-called Johnson translation system. We start with the univariate system, for which a random variable X is defined by a cumulative distribution function of the form

$$F(X) = \Phi\left\{\gamma + \delta g\left[\frac{X-\xi}{\lambda}\right]\right\}, \tag{3}$$

where γ and δ are shape parameters, ξ is a location parameter, λ is a scale parameter, Φ is the standard univariate normal distribution function and $g(\cdot)$ is one of the following transformations:

$$g(y) = \begin{cases} \log(y) & \text{for the } S_{\mathrm{L}} \text{ (log-normal) family,} \\ \log\left(y + \sqrt{y^2+1}\right) & \text{for the } S_{\mathrm{U}} \text{ (unbounded) family,} \\ \log\left(\frac{y}{1-y}\right) & \text{for the } S_{\mathrm{B}} \text{ (bounded) family,} \\ y & \text{for the } S_{\mathrm{N}} \text{ (normal) family.} \end{cases} \tag{4}$$

There is a unique family (choice of g) for each feasible combination of the skewness and the kurtosis that determine the parameters γ and δ. Any mean and (positive) variance can be attained by a g in (4) by manipulating the parameters λ and ξ. Within each family, a distribution is completely specified by

the values of the parameters $[\gamma, \delta, \lambda, \xi]$ and the range of X depends on the family of interest. A detailed illustration for the shapes of the Johnson-type probability density functions can be found in Johnson (1987).

The multivariate version of this system is obtained by using the following d-dimensional normalizing translation (Johnson, 1949b):

$$\mathbf{Z} = \boldsymbol{\gamma} + \boldsymbol{\delta}\mathbf{g}\big[\boldsymbol{\lambda}^{-1}(\mathbf{X} - \boldsymbol{\xi})\big] \sim \mathrm{N}_d(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{Z}}), \tag{5}$$

where $\mathbf{Z}$ is a d-dimensional normal random variable with mean $\mathbf{0}$ and a $d \times d$ covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Z}}$, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_d)'$, $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_d)'$ are the d-dimensional vectors of shape and location parameters, $\boldsymbol{\delta} = \mathrm{diag}(\delta_1, \delta_2, \ldots, \delta_d)$ and $\boldsymbol{\lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_d)$ are the diagonal matrices whose entries are the shape and scale parameters and the transformation $\mathbf{g}(\cdot)$ is defined as

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

where $g_i(y_i)$, $i = 1, 2, \ldots, d$, are defined as in (4). This characterization ensures that the marginal distribution of each X_i is a univariate Johnson distribution.

Modeling using the Johnson family of distributions can be advantageous since they have the ability to represent any set of (finite) first four moments. Moreover, simulation output performance measures are other insensitive to the specific input distribution chosen when enough moments of the distribution are correctly captured. Examples include diffusion approximations and many queuing applications (Burman, 1981). Moreover, the Johnson translation system represents a wide variety of unimodal and bimodal distributional shapes, though shapes with three or more nodes cannot be represented. Further, it allows for the marginal distributions to be different, while most parametric families require that they be from the same family.

The distributions of the Johnson translation system are easy to sample using the normalizing translation defined in (5) once the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Z}}$ of the multivariate normal random variable $\mathbf{Z}$, and the marginal distribution parameters $\mathbf{g}$, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, $\boldsymbol{\lambda}$ and $\boldsymbol{\xi}$ are given. Estimating these distribution parameters from available data can however pose significant challenges. The marginal parameters can be estimated from the corresponding univariate marginal information (DeBrota et al. (1988) provide an efficient numerical procedure), but estimating the normal covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Z}}$, for instance from the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X}}$ of the Johnson random variable $\mathbf{X}$, can be tricky. If $\boldsymbol{\Sigma}_{\mathbf{Z}}$ were being estimated from $\boldsymbol{\Sigma}_{\mathbf{X}}$, then Theorems 1 and 2 in Cario and Nelson (1997) show that under mild conditions on the marginals the input correlation $\boldsymbol{\Sigma}_{\mathbf{X}}(i, j)$ is a continuous nondecreasing function of the base correlation $\boldsymbol{\Sigma}_{\mathbf{Z}}(i, j)$ and does not depend on the other base correlations. This suggests that a straightforward numerical search procedure should yield the correct $\boldsymbol{\Sigma}_{\mathbf{Z}}$ for the given $\boldsymbol{\Sigma}_{\mathbf{X}}$. Such a procedure might however not always produce positive semidefinite estimates for $\boldsymbol{\Sigma}_{\mathbf{Z}}$ (Ghosh and Henderson, 2002a), a necessary (and sufficient) condition for a matrix to be a multivariate normal covariance matrix.

Stanfield et al. (1996) suggest a Johnson-type marginal-based data-generation procedure that does not need to determine the covariance matrix of the multivariate normal random vector. To generate data for a random vector to match mean $\boldsymbol{\mu}_{\mathbf{X}}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X}}$, the authors first define a random vector $\mathbf{Y}$ whose components are independent standardized Johnson variates, i.e., $\mathrm{E}(\mathbf{Y}_i) = 0$ and $\mathrm{Var}(\mathbf{Y}_i) = 1$ and $\mathrm{Cov}(\mathbf{Y}_i, \mathbf{Y}_j) = 0$ for $i, j = 1, 2, \ldots, d$. Let $\mathbf{L}_{\mathbf{X}} = \boldsymbol{\Sigma}_{\mathbf{X}}^{1/2}$ correspond to the lower triangular Cholesky-decomposition factor of the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X}}$ and $\boldsymbol{\sigma}_{\mathbf{X}}$ be a diagonal matrix whose entries are the standard deviations of the components of $\mathbf{X}$. Then, random vector $\mathbf{X} = \boldsymbol{\mu}_{\mathbf{X}} + \boldsymbol{\sigma}_{\mathbf{X}}\mathbf{L}_{\mathbf{X}}\mathbf{Y}$ has desired mean $\boldsymbol{\mu}_{\mathbf{X}}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X}}$. Further, Stanfield et al. (1996) give expressions that relate the skewness and kurtosis values of $\mathbf{X}$ to those of $\mathbf{Y}$. Thus, one can start with the right parameter values for $\mathbf{Y}$ to obtain the desired first four moments in $\mathbf{X}$. The computational effort needed in finding the right parameters is relatively little. Hence, it is an attractive generation method for obtaining samples with given mean and covariance matrix. However, their approach has some limitations. First, since the marginals do not in general survive the transformation from $\mathbf{Y}$ to $\mathbf{X}$, the exact expressions and properties of the form of the joint c.d.f. and marginal distributions of $\mathbf{X}$ are not clear. Second, they note that although any given skewness factor can be achieved for the standardized Johnson components Y_i, $i = 1, 2, \ldots, d$, not all possible values of the kurtosis can be achieved. Moreover, the order in which the component parameters are determined can affect the kurtosis determination problem.

Flexible families of distributions are, of course, not limited to the Johnson translation system. Johnson et al. (1997) and Kotz et al. (2000) are good sources for multivariate discrete distributions and multivariate continuous distributions, respectively. Other distributions that have been given particular attention in the literature include Tukey's g and h transformations (Johnson, 1987), the multivariate Pearson (Parrish, 1990), and the four-parameter families designed explicitly for use in simulation experiments by Ramberg and Schmeiser (1974) and Schmeiser and Deutsch (1977). The last two families of distributions are particularly easy to use, but the one-to-one relationship between the distribution parameters and the moments is lost. They also fall short in capturing distributional characteristics such as bimodality and heavy tails.

4 Constructing partially specified joint distributions

So far, we have discussed the methods that work either with completely specified c.d.f.'s or with estimates picked from special parametric families of distributions. We have seen that both approaches can face severe practical limitations. In this section, we discuss the construction of input models that match partial specifications of the input process. Here, the input process is commonly specified in terms of marginal distributions of its components and a measure

of dependence. An argument in support of modeling using only marginals and a measure of dependence relates to the use of diffusion approximations for modeling stochastic systems. In many cases the limiting diffusions depend only on the first two moments of the input distributions. Performance measures in many queuing applications (Burman, 1981) can be insensitive to moments higher than the second. Therefore, there is some insensitivity in performance measures computed from these models to the exact form of the input distributions. In general then, if this form of insensitivity is present in a model, then the approach discussed here is quite reasonable.

The methods of this section aim, rather successfully, for efficient input model development (model fitting and variate generation) in higher dimensions in return for a compromise on the modeling front. The hope is that a proper specification can capture the essence of the dependence between the components while sparing the practitioner the often arduous task of trying to estimate the full joint distribution. However, partial characterization may not necessarily uniquely or even correctly specify a joint distribution, and more than one, or sometimes no, joint distribution can be defined that satisfy the desired characterization. We consider a partial specification as *feasible* if a multivariate process (or equivalently a probability distribution in the appropriate space), which has the specified properties, exists. The methods presented in this section should ideally be guaranteed to work with any feasible partial specification.

The most commonly specified dependence measure quantifies linear correlation between the components, as correlation matrices for random vectors and a series of correlation matrices called the correlogram for time series. Section 4.1 reviews some relevant properties of correlation and other measures of dependence used in the simulation input model development. Time series modeling literature focuses primarily on processes that are partially specified with distributions for individual elements and autocorrelations. Therefore, we provide a brief overview of basic time series notations and some important models in Section 4.2. In the remainder of the section, we present key approaches that model and generate input processes (random vectors and time series) specified with almost any set of arbitrary marginal distributions and correlation matrices.

4.1 Measures of dependence

Most of the dependence measures we discuss in this section are *pair-wise* measures, in that they are used to quantify the dependence between pairs of components. We refer the reader to Nelsen (1999) for a complete treatment of other dependence measures.

The most widely used and understood measure in engineering applications is the *product-moment covariance*. The (i, j)th product-moment covariance between random variables X_i and X_j (whose variances are necessarily finite) is

defined as

$$\mathrm{Cov}(X_i, X_j) = \mathrm{E}\big[(X_i - \mu_i)(X_j - \mu_j)\big] \quad \text{for } 1 \leqslant i \leqslant j \leqslant d,$$

where $\mathrm{E}(X_i) = \mu_i$ and $\mathrm{E}(X_j) = \mu_j$. A related measure is the product-moment correlation (the sample analog is called the Pearson correlation) of X_i and X_j, defined as

$$\rho(i, j) = \mathrm{Cor}(X_i, X_j) = \frac{\mathrm{Cov}(X_i, X_j)}{\sigma_i \sigma_j}, \tag{6}$$

where $\mathrm{Var}(X_i) = \sigma_i^2$ and $\mathrm{Var}(X_j) = \sigma_j^2$. Correlation values, limited to be in a subset of $[-1, 1]$, are preferred over covariances because of their independence from scale parameters of the marginal distributions of X_i and X_j. We shall discuss correlations in the sequel, but each measure can be recovered from the other given the marginal distributions and hence can be treated equivalently.

Product-moment correlation is a measure of *linear* dependence between two components, in that it takes on the maximum magnitude of 1 if and only if a linear relationship exists between the components. This can be a modeling limitation, and can be especially disadvantageous in cases where strong nonlinear dependencies exist since the product-moment correlation value fails to register this.

Some efficient procedures try to incorporate additional dependence information beyond correlations. The algorithm developed by Schmeiser and Lal (1982) is worth mentioning here: it models and constructs bivariate gamma processes using specified correlations, and can also incorporate regression curves of the form $\mathrm{E}[X_1|X_2 = x_2]$ and $\mathrm{E}[X_2|X_1 = x_1]$. Unfortunately, this algorithm relies heavily on the properties of the gamma random variables, and a generalization to other marginal distributions is not obvious.

The correlations (covariances) between the components of a random vector $\mathbf{X}$ are collectively represented by the *correlation (covariance) matrix*. The (i, j)th element of this matrix represents the correlation between components X_i and X_j. By definition, the correlation matrix of any random vector is symmetric, positive semidefinite, and has unit diagonal elements. We refer the reader to Ghosh and Henderson (2002a) for a discussion on the implications of the positive semidefiniteness of a correlation matrix. Biller and Nelson (2005) establish similar requirements for correlograms of time-series processes. However, positive semidefiniteness is not sufficient to guarantee feasibility of a marginal and covariance (or correlation) matrix specification. The marginal distributions also play a crucial role in determining the feasibility of correlation matrices for random vectors and correlograms for multivariate time series. As an example, consider a bivariate input process with a standard exponential and a standard uniform as marginals. The magnitude of correlation between them cannot exceed 0.866. Indeed, a correlation value of 1 would imply a linear relationship between the two, which clearly cannot be the case. In general then,

the set of feasible correlation matrices is a strict subset of the set of correlation matrices.

The definition of the product-moment correlation in (6) implies that it is defined only when the variances of the components of the input process are finite. Thus, it is not an appropriate measure of dependence for very heavy-tailed inputs. A dependence measure that avoids this pitfall is the *Spearman rank correlation* (see Jong-Dev, 1984), defined as

$$r(i,j) = \frac{\mathrm{E}[F_i(X_i)F_j(X_j)] - \mathrm{E}[F_i(X_i)]\,\mathrm{E}[F_j(X_j)]}{\sqrt{\mathrm{Var}(F_i(X_i))\,\mathrm{Var}(F_j(X_j))}}. \tag{7}$$

(The transform $F_i(X_i)$ is called the *probability integral* transform of X_i.) The random variables $F_i(X_i)$ and $F_j(X_j)$ are bounded within [0, 1] and hence rank correlations always exist.

In an early work on rank correlations, Hotelling and Pabst (1936), citing previous work by Karl Pearson, show that for jointly normal random variables, the product-moment correlation ρ is related to the rank correlation r by

$$\rho = 2\sin\Big(\frac{\pi}{6}r\Big). \tag{8}$$

The main advantage of rank correlations over product-moment correlations is that they are invariant under transformations of the input random variables. As an example, consider two log-normal random variables denoted by X_1 and X_2. The rank correlation between $\log(X_1)$ and $\log(X_2)$ is the same as the rank correlation between X_1 and X_2. Thus, the use of the rank correlation provides a natural way to separate the characterization of the individual marginal distributions from that of the dependence among the variables, simplifying the sample generation. A random vector $\mathbf{X} = (X_1, X_2, \ldots, X_d)'$ with marginals F_i, $i = 1, 2, \ldots, d$, and rank correlation matrix $\boldsymbol{\Sigma}_{\mathbf{X}}$ can be sampled by first generating a random vector $\mathbf{U} = (U_1, U_2, \ldots, U_d)'$ with uniform marginals and the same rank correlation matrix $\boldsymbol{\Sigma}_{\mathbf{X}}$, and then applying the inverse probability transform to each variate individually: $X_i = F_i^{-1}(U_i)$ for $i = 1, 2, \ldots, d$.

Nelsen (1999) describes other monotonic-transformation-invariant measures of dependence that have been used in various applications. For example, the *orthant probability* measures the probability that both variables are above or both below their medians. It is alternatively called the *median deviation concordance* probability. It is invariant to monotonic transformations of the variates. For the bivariate normal distribution with correlation ρ, this probability is $1/2 + \sin^{-1}(\rho)/\pi$ (Stuart and Ord, 1987). Another example of a widely known monotonic-transformation-invariant measure is Kendall's τ and is typically defined using independent and identically distributed random vectors, say (X_1, Y_1) and (X_2, Y_2). It is defined as the probability of concordance minus the probability of discordance

$$\begin{aligned}\tau(X,Y) = {} & \mathrm{P}\big[(X_1 - X_2)(Y_1 - Y_2) > 0\big] \\ & - \mathrm{P}\big[(X_1 - X_2)(Y_1 - Y_2) < 0\big].\end{aligned}$$

Clearly, concordance between two random variables arises if large values of one variable occur with the large values of the other and small values occur with small values of the other. Both the orthant probability and Kendall's τ have been studied for the special case of uniform marginals and their extension to more general cases remains an open area of interest.

4.2 Multivariate time series

Statistical time series analysis literature focuses attention chiefly on studying models that specify marginal distributions for individual elements and model the dependence within the series in terms of the correlogram, which is a system of lag-indexed product-moment correlations of the sequence. Assuming the stochastic process is stationary, the correlogram can be defined to be the set of correlations of the form $\boldsymbol{\Sigma}_X(h) = \mathrm{Cor}(\mathbf{X}_t, \mathbf{X}_{t+h})$, $h \geqslant 1$. Here, $\boldsymbol{\Sigma}_X(h)$ might correspond to a scalar or a matrix depending on whether the $\{\mathbf{X}_t; t \geqslant 1\}$ is a univariate or a multivariate time-series process.

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

$$\mathbf{X}_t = \sum_{i=0}^{\infty} \boldsymbol{\psi}_i \mathbf{Y}_{t-i}, \quad t \geqslant 1. \tag{9}$$

The series $\{\boldsymbol{\psi}_i; i \geqslant 1\}$ is a sequence of $d \times d$ matrices such that $\sum_{i=0}^{\infty} \| \boldsymbol{\psi}_i \| < \infty$, where $\| \cdot \|$ denotes the usual eigenvalue norm, and $\{\mathbf{Y}_t; t = 0, \pm 1, \pm 2, \ldots\}$ is a sequence of independent and identically distributed d-dimensional white noise vectors with mean zero and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Y}}$ such that

$$\mathrm{E}[\mathbf{Y}_t] = \mathbf{0}_{(d\times 1)} \quad \text{and} \quad \mathrm{E}\big[\mathbf{Y}_t \mathbf{Y}'_{t+h}\big] = \begin{cases} \boldsymbol{\Sigma}_{\mathbf{Y}} & \text{if } h = 0, \\ \mathbf{0}_{(d\times d)} & \text{otherwise.} \end{cases}$$

Hannan (1970) shows that if the $\mathbf{X}_t$'s are Gaussian with mean zero and have an absolutely continuous spectrum, then there exists a sequence of independent and identically distributed normal mean-zero random vectors $\mathbf{Y}_t$, $t = 0, \pm 1, \pm 2, \ldots$, and a sequence of matrices $\{\boldsymbol{\psi}_i; i \geqslant 0\}$, such that the processes $\mathbf{X}_t$ and $\sum_{i=0}^{\infty} \boldsymbol{\psi}_i \mathbf{Y}_{t-i}$ are stochastically equal. Furthermore, if the $\mathbf{Y}_t$'s are Gaussian, then so are the $\mathbf{X}_t$'s in (9). If we choose a $d \times d$ matrix $\boldsymbol{\psi}$ for which $\sum_{i=0}^{\infty} \| \boldsymbol{\psi}^i \| < \infty$ holds and set $\boldsymbol{\psi}_i = \boldsymbol{\psi}^i$ for $i \geqslant 0$, then $\mathbf{X}_t$ simplifies to a stationary first-order vector autoregressive process that is written as $\mathbf{X}_t = \boldsymbol{\psi}\mathbf{X}_{t-1} + \mathbf{Y}_t$. This process is very well suited to modeling multivariate time series with normal marginal distributions. We refer the reader to Lütkepohl (1993) for a comprehensive review of the distributional properties of the multivariate autoregressive time-series processes.

However, there are many physical situations in which the marginals of the time-series input processes are nonnormal. If the input process is non-Gaussian, then the decomposition in (9) may not exist and the statistical inference procedures developed for processes satisfying (9) do not apply. Over

the past decade, a considerable amount of research effort has been devoted to develop models of time series with exponential, gamma, geometric, or general discrete marginal distributions. We refer the reader to Lewis et al. (1989) and Block et al. (1990) for example work on univariate and bivariate autoregressive moving-average models with gamma marginals and exponential and geometric marginals, respectively. The primary shortcoming of the approach taken to construct these input processes is that a different model is required for each marginal distribution of interest and the sample paths of these processes, while adhering to the desired marginal distribution and autocorrelation structure, sometimes have unexpected features. In the succeeding subsection, we discuss a more general approach for modeling nonnormal time series.

4.3 Transformation-based methods

The methods covered in this section are in essence an extension of the transformation-based univariate generation procedure of Chapter 4, where any arbitrary distribution F is sampled using the inverse transformation of the uniform random variable $U(0, 1]$, $F^{-1}(U)$. This procedure works because the random variable $F^{-1}(U)$ has the cumulative distribution function F, and is extensively used since univariate uniform variables are very easily generated. The key idea remains the same in the multivariate setting, where we apply component-wise inverse marginal transformations to a base multivariate process with $U(0, 1]$ marginal distributions. Consider the generic case of modeling a d-variate stationary time series $\{\mathbf{X}_t; t \geqslant 1\}$ whose component time series $\{X_{i,t}; i = 1, 2, \ldots, d, t \geqslant 1\}$ has a marginal distribution denoted by F_i. For $i = 1, 2, \ldots, d$, we let $\{U_{i,t}; t \geqslant 1\}$ be the ith component of the d-variate time-series process $\{\mathbf{U}_t; t \geqslant 1\}$ with uniform marginals. We obtain the ith time series via the transformation

$$X_{i,t} = F_i^{-1}(U_{i,t}). \tag{10}$$

In the remainder of the section, we shall refer to the time series $\{\mathbf{U}_t; t \geqslant 1\}$ as the base process.

Two important issues need to be addressed: the construction of the base process $\mathbf{U}_t$ has to enable fast and efficient sampling and the dependence structure of the base process has to achieve the desired dependence structure of the input process. Both of these issues are tackled in different ways by the methods we shall discuss below. In Section 4.3.1, we present methods that construct the uniform base process by using transformations of a multivariate process with normal marginals. They can be used in the context of random vectors as well as times series, and can match both product-moment and rank correlation specifications. Section 4.3.2 presents a method for modeling uniform random vectors using chessboard copulas, while the method in Section 4.3.3 uses vine copulas. These copula construction procedures are typically used to match rank correlation specifications. Finally, Section 4.3.4 presents transform-expand-sample

univariate time series processes whose base process is a sequence of autocorrelated uniforms generated in an autoregressive manner.

4.3.1 ARTA, NORTA and VARTA processes

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

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

The α_h, $h = 1, \ldots, p$, are fixed autoregressive coefficients that uniquely determine the autocorrelation structure of the base process, $\rho_Z(h)$, $h = 1, \ldots, p$, and Y_t, $t = p+1, p+2, \ldots$, are independent and identically distributed Gaussian random variables with mean zero and variance σ_Y^2. The univariate time series $\{X_t; t \geqslant 1\}$ is obtained via the transformation

$$X_t = F^{-1}(U_t) = F^{-1}\big[\Phi(Z_t)\big] \tag{11}$$

which ensures that X_t has distribution F. 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 value of the lag-h base correlation $\rho_Z(h)$ depends only on the corresponding input correlation $\rho_X(h)$. The correlation structure determination step 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). It can be considered as broadening the ARTA process beyond a common marginal distribution. The central idea is to transform a standard multivariate normal vector into the desired random vector, which is referred as having a Normal-To-Anything (NORTA) distribution. Specifically, we let

$$\mathbf{X} = \big[F_1^{-1}\big(\Phi(Z_1)\big), F_2^{-1}\big(\Phi(Z_2)\big), \ldots, F_d^{-1}\big(\Phi(Z_d)\big)\big]', \tag{12}$$

where the base vector $\mathbf{Z} = (Z_1, \ldots, Z_d)'$ is a standard multivariate normal vector with correlation matrix $\boldsymbol{\Sigma}_\mathbf{Z}$ and $F_1, F_2, \ldots, F_d$ are the desired marginal distributions. The base process setup problem then boils down to finding the base correlation matrix $\boldsymbol{\Sigma}_\mathbf{Z}$ for $\mathbf{Z}$ that transforms to the desired correlation matrix $\boldsymbol{\Sigma}_\mathbf{X}$ for $\mathbf{X}$. As in the earlier case, each $\boldsymbol{\Sigma}_\mathbf{Z}(i, j)$ depends only on the corresponding component $\boldsymbol{\Sigma}_\mathbf{X}(i, j)$, and hence a total of $d(d-1)/2$ independent correlation-matching problems have to be solved.

Recently, Biller and Nelson (2003) have pulled together the theory behind ARTA and NORTA processes and extended it to multivariate time series. They

construct the Vector-Autoregressive-To-Anything (VARTA) process, which is a stationary, d-variate, vector time series $\{\mathbf{X}_t; t \geqslant 1\}$, by taking the inverse transformation of a standardized Gaussian vector autoregressive process of order p

$$\mathbf{Z}_t = \sum_{h=1}^{p} \boldsymbol{\alpha}_h \mathbf{Z}_{t-h} + \mathbf{Y}_t, \quad t = p+1, p+2, \ldots,$$

where the $\boldsymbol{\alpha}_i$, $i = 1, 2, \ldots, p$, are fixed $d \times d$ autoregressive coefficient matrices and $\{\mathbf{Y}_t; t \geqslant p+1\}$ is a sequence of independent and identically distributed d-variate multivariate normal random vectors with mean zero and a $d \times d$ covariance matrix $\boldsymbol{\Sigma}_Y$ chosen as $\boldsymbol{\Sigma}_Z(0) - \sum_{h=1}^{p} \boldsymbol{\alpha}_h \boldsymbol{\Sigma}'_Z(h)$. This selection of $\mathbf{Y}_t$ ensures that each $Z_{i,t}$ is marginally standard normal. For a fully specified VARTA process, there are $pd^2 + d(d-1)/2$ individual correlation matching problems to solve. Note 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 setup problem central to the ARTA, NORTA and VARTA processes is to solve correlation-matching problems of the form $c(\rho_Z) = \rho_X$ for ρ_Z, where c is a complex nonlinear function that requires the evaluation of double integrals. For the VARTA process, we can express input product-moment correlation $\rho_{\mathbf{X}}(i, j, h) = \mathrm{Cor}[X_{i,t}, X_{j,t+h}]$ as

$$\begin{aligned}
&\mathrm{Cor}\big[F_i^{-1}\big[\Phi(Z_{i,t})\big], F_j^{-1}\big[\Phi(Z_{j,t+h})\big]\big] \\
&\quad = \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_i^{-1}\big[\Phi(z_{i,t})\big] F_j^{-1}\big[\Phi(z_{j,t+h})\big] \right. \\
&\qquad\qquad \left. \times\, \vartheta_{\rho_{\mathbf{Z}}(i,j,h)}\big(z_{i,t}, z_{j,t+h}\big)\, \mathrm{d}z_{i,t}\, \mathrm{d}z_{j,t+h} - \mu_i \mu_j \right) (\sigma_i \sigma_j)^{-1} \\
&\quad = c_{ijh}\big[\rho_{\mathbf{Z}}(i, j, h)\big], \qquad (13)
\end{aligned}$$

where μ_i and σ_i^2 are the mean and the variance of the ith component of the input process and $\vartheta_{\rho_{\mathbf{Z}}(i,j,h)}$ is the standard bivariate normal probability density function with correlation $\rho_{\mathbf{Z}}(i, j, h)$. Clearly, $\rho_{\mathbf{X}}(i, j, h)$ depends only on $\rho_{\mathbf{Z}}(i, j, h)$. The function $c_{ijh}(\rho)$ is nondecreasing, lies on the origin for $\rho = 0$ and satisfies $|c_{ijh}(\rho)| \leqslant |\rho|$ for any $\rho \in [-1, 1]$ (Lancaster, 1957). Moreover, under mild conditions on the marginal distributions F_i and F_j, the function c_{ijh} is continuous (see Theorems 3.4 and 3.5 in Cario and Nelson, 1996; Biller and Nelson, 2003). Similar relations hold between the rank correlation $r_{\mathbf{X}}$ of the input process and the product-moment correlation $\rho_{\mathbf{Z}}$ of the base normal process.

The correlation-matching problems (13) can be solved analytically, numerically or using a simulation-based approach. It is possible to find the base dependence $\rho_{\mathbf{Z}}(i, j, h)$ values analytically only in special cases, e.g.,

if the marginal distributions F_i are uniform on $(0, 1]$, then the product-moment (and rank) correlation matching problem (13) reduces to $\rho_{\mathbf{X}}(i, j, h) = (6/\pi)\sin^{-1}[\rho_{\mathbf{Z}}(i, j, h)/2]$ (Kruskal, 1958). The astute reader will notice that this is the inverse of the relation in (8).

The idea behind the numerical approach is to take some initial base correlations, transform them into the implied correlations for the specified pair of marginals (using a numerical integration technique), and then employ a numerical root-finding method until we find a base correlation that approximates the desired correlation within a pre-specified level of accuracy. The properties of the function c_{ijh} ensure that a numerical search finds a value for the base correlation $\rho_{\mathbf{Z}}(i, j, h)$ efficiently. Both Cario and Nelson (1998) and Biller and Nelson (2003) provide such a numerical search method. The efficiency of the numerical search methods varies with the marginals and the $\rho_{\mathbf{X}}$ values: when the integrands are smooth, the numerical approach is very efficient, and may be inaccurate otherwise. Specialized numerical integration tricks might be required to overcome problematic situations.

The simulation-based approach can be used in cases where the form of the marginal distributions makes numerical evaluations of (13) prohibitively expensive. Equation (13) can be evaluated using sampling techniques as follows: A sample of pairs of bivariate standard normal random variates with correlation $\varphi_{\mathbf{Z}}(i, j, h)$ are generated. Each pair is then transformed via (11) into samples with the desired marginals, and the sample correlation estimate $\varphi_{\mathbf{X}}(i, j, h)$ is obtained from these transformed pairs. The key idea is that if $\varphi_{\mathbf{X}}(i, j, h)$ is sufficiently close to the desired input correlation $\rho_{\mathbf{X}}(i, j, h)$, then the base correlation $\varphi_{\mathbf{Z}}(i, j, h)$ is a good approximation to the base correlation $\rho_{\mathbf{Z}}(i, j, h)$ that solves the correlation-matching problem $c_{ijh}(\rho_{\mathbf{Z}}(i, j, h)) = \rho_{\mathbf{X}}(i, j, h)$. After a number of such sampling evaluations of (13), stochastic root finding algorithms are applied to search for the correlation of interest within a predetermined precision. This approach is very general and is a sensible alternative when the dimension of the problem is moderate and a diverse collection of marginal distributions have to be considered.

Throughout the previous discussion, we have assumed that a multivariate process exists with marginal distributions F_i, for $i = 1, 2, \ldots, d$, and an autocorrelation structure characterized by $\boldsymbol{\Sigma}_{\mathbf{X}}(h)$, $h = 0, 1, \ldots, p$. We need to exercise some care here because not all combinations of F_i, $i = 1, 2, \ldots, d$, and $\boldsymbol{\Sigma}_{\mathbf{X}}(h)$, $h = 0, 1, \ldots, p$, are feasible, as discussed in Section 4.1. Any multivariate process constructed by the transformation (11) or (12) from a base normal process with an appropriately chosen positive semidefinite base autocorrelation structure $\boldsymbol{\Sigma}_{\mathbf{Z}}(h)$ is automatically assured to be feasible. Unfortunately, the converse does not necessarily hold; that is, there exist sets of marginals with a feasible autocorrelation structure that cannot be represented by the VARTA transformation. This problem occurs because $\boldsymbol{\Sigma}_{\mathbf{Z}}(h)$ corresponding to a given feasible $\boldsymbol{\Sigma}_{\mathbf{X}}(h)$ structure is not positive semidefinite. This problem was postulated for the random vector case in Li and Hammond (1975) and was proved to exist in Ghosh and Henderson (2002a). Ghosh and Henderson (2002b) report

that, in the case of random vectors, the failure of the transformation-based methods is relatively rare in moderate dimensions and that the method fails when the correlations lie either on the boundary or in close proximity to the set of correlations achievable for the specified marginals of the process. They note, however, that the problem does become increasingly evident as the dimension increases.

If, after solving the correlation-matching problems, the estimate for the base correlations $\mathbf{\Sigma}_Z$ is not positive definite, then the procedures proposed by Lurie and Goldberg (1998) and by Ghosh and Henderson (2002a) can be applied to find a symmetric, positive definite approximation of the base correlation matrix. The hope is that a multivariate process constructed in this manner will have a correlation structure close to the desired one, since the function c_{ijh} defined by (13) is known to be continuous under fairly general conditions.

4.3.2 Chessboard distributions

Ghosh and Henderson (2002a) study a class of copulas that they call chessboard distributions to model a random vector with uniform marginals given the correlation matrix. These distributions are closest in nature to the piecewise-uniform copulas developed by Mackenzie (1994). The construction of a chessboard distribution starts with the division of the d-dimensional unit hypercube into a grid of equally sized cells with sides of length $1/n$, where parameter n is said to index the level of discretization. The joint density of this copula is defined to be constant over each cell. Thus, the density over cell $C(j_1, \ldots, j_d)$, $j_i = 1, 2, \ldots, n$ for each $i = 1, \ldots, d$, is defined as $f(x) = n^d q(j)$, where $j = (j_1, \ldots, j_d)$ and $q(j)$ is the probability that the random vector takes values in the cell $C(j)$. Ghosh and Henderson (2002a) find the values of the $q(j)$s that match a given correlation matrix by solving a linear program. They additionally impose conditions on the values of the $q(j)$s to ensure that f is a d-variate density function.

After appropriate $q(j)$ values are determined, the data-generation procedure works as follows: an index vector $j = (j_1, \ldots, j_d)$ is generated from the discrete distribution represented by the $q(j)$s, and then the random vector sample is chosen uniformly from the chosen cell $C(j)$. Efficient algorithms are available to sample from the discrete distribution q, e.g., the alias method that is developed by Walker (1977) and discussed in detail in Law and Kelton (2000).

This correlation-matching procedure was first introduced in Ghosh and Henderson (2002a) to study feasibility of correlation matrices for any arbitrary set of marginals. They show that this procedure matches almost any (in a precise sense) feasible correlation matrix. The framework is advantageous because one can impose additional constraints on the joint distribution (Ghosh and Henderson, 2001). Thus, it gives the user a degree of control over the joint distribution.

Chessboard copulas can be used to sample random vectors with specified arbitrary continuous marginals and rank correlation matrices via the transfor-

mation (10). This approach can be extended to match arbitrary marginals with a specified product-moment correlation matrix (Ghosh, 2004).

However, this modeling technique suffers significant limitations. The chessboard distributions tend to be sparse in nature as the fraction of cells with positive valued $q(j)$s gets smaller with increasing dimensions. Moreover, the size of the linear programs (indexed by the discretization parameter n) that are solved to match a given correlation matrix can be very large. Ghosh (2004) develops bounds on the required discretization level n. Nevertheless, the chessboard distributions can be of significant help in modeling random vectors of low to moderate dimensions, and are especially attractive since the constructed joint distributions can be controlled.

4.3.3 Vine copula method

An alternative copula-based input model has recently been suggested by Cooke and his co-authors (Kurowicka and Cooke, 2002; Bedford and Cooke, 2002). This method is founded on Bayesian statistics literature and utilizes graph theoretic concepts. It generates copula samples quickly and efficiently, and can be used as a base for a fast sampling procedure that matches a set of arbitrary marginals and a rank correlation matrix.

To define a d-variate copula, the first step is to construct a graph structure called 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 $j+1$, starting with a tree on a graph with the d components as nodes. The number of nested trees is then $d-2$. A *regular* vine on the d components is a vine in which two edges in tree j are joined by an edge in $j+1$ only if they share a common node in the jth tree for $j = 1, 2, \ldots, d-2$. There are thus a total of $d(d-1)/2$ edges in a regular vine.

Each edge is associated with a *conditional rank correlation* that takes values from the $[-1, 1]$ interval. A conditional rank correlation $r[X_i, X_j|S]$ is defined to be the rank correlation between the conditioned random variables $X_i|S$ and $X_j|S$, where S is a collection of other components. The set S of conditioning components is determined in a systematic manner (Cooke, 1997) and grows as we go deeper into the vine, with the edge of the $(d-2)$nd single-edge tree associated with a conditional rank correlation $r[X_u X_v|\{X_i, i \neq u \text{ or } v\}]$, for some u and v.

Cooke (1997) provides a method to sample from the vine. A bivariate copula is associated with each edge of the vine, and has a correlation equal to the conditional correlation value associated with the edge. Sampling starts with d independent uniform samples and proceeds by traversing the vine in a specific order and applying successive inversions of the conditional distributions derived from the bivariate copulas of each edge. Cooke (1997) associates minimum information copulas (see Meeuwissen and Bedford (1997) for a definition) with the vine edges and shows that the joint distribution of the random vector generated by this procedure is unique and possesses certain favorable properties.

Bedford and Cooke (2002) associate another class of bivariate copulas, the elliptical copulas, with the edges of the vine and their conditional correlations. This case has the advantage that the conditional correlations of the edges can now be related by a set of recursive equations to the set of correlations induced between the components of the random vector generated by the vine-traversing procedure. This implies that a set of conditional rank correlations of a regular vine can be derived for any desired correlation matrix, and thus the method can be used to match a given $d \times d$ rank correlation matrix to uniform marginals. Additionally, conditional distributions of elliptical copulas can be derived in explicit form and are very easy to invert. Thus vine-based sampling, which involves applying $d(d-1)/2$ of these inverses to d independent uniform variates, is fast and efficient. Kurowicka and Cooke (2002) show that this procedure can match any feasible correlation matrix with uniform marginals in the trivariate case. The method can come very close to achieving any feasible correlation matrix in higher dimensions (Cooke, 2003). However, it is not known whether they can achieve all feasible correlation matrices in dimensions greater than three.

4.3.4 TES processes

The Transform-Expand-Sample (TES) process, developed by Melamed (1991), generates a univariate time series with general marginals using as base process a sequence of autocorrelated uniforms generated in an autoregressive manner. Arbitrary marginals are obtained via a transformation similar to (10). TES processes play a significant role in univariate time series modeling and approach the generality of the ARTA process in generating stationary univariate time-series processes. Unfortunately, extensions to random vector and multivariate time-series processes are not available.

The TES process U_t can attain the full range of feasible lag-one autocorrelations for a given marginal distribution, and is defined via processes U_t^+ and U_t^- that are defined as follows:

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

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

where the random variables U_0^+ and V_t, $t \geqslant 1$, have $(0, 1]$-uniform distributions, V_t is independent of $\{U_0^+, V_1, \ldots, V_{t-1}\}$, and the notation $\langle x \rangle$ denotes modulo-1 arithmetic. The process U_t^+ covers the positive lag-one range, while the process U_t^- covers the negative lag-one range and is defined in terms of the U_t^+ process. The autocorrelation structure depends only on the distribution of V_t and can be manipulated by modifying the distribution of V_t without changing the marginal distribution of U_t. However, altering the distribution of V_t typically changes the autocorrelations of the process at all lags, but the user has no control over this behavior.

A disadvantage of the TES process is that extreme jumps may appear in the sample path due to the modulo-1 arithmetic. For example, U_{t-1} can be very close to 1 when U_t is very close to 0. A stitching transformation, $S_\xi(U_t)$ parameterized by $\xi \in [0, 1]$, mitigates this effect. The process $\{S_\xi(U_t); t \geqslant 1\}$ still has $(0, 1]$-uniform marginals, but no longer has extreme jumps. Unfortunately, stitching does not preserve the autocorrelation structure and the change cannot be expressed as a simple function of ξ. The distribution of V_t has to be modified until the autocorrelations of the constructed process match the pre-specified values. This forms one of the main challenges of using the TES process, and the ARTA process is perhaps a better option when the objective is to construct a time-series input process with pre-specified marginal distribution and autocorrelations through lag p. An ARTA process can be fitted to the desired set of $p \geqslant 1$ autocorrelations without user intervention, for any marginal distribution, and has smoother sample paths.

4.4 Mixture methods

Mixture methods sample random vectors with arbitrary marginals and a desired correlation matrix by probabilistically mixing samples from joint distributions with pre-determined correlation matrices. While this approach has been utilized with various configurations of pre-determined matrices, Hill and Reilly (1994) present what is perhaps the most elegant version.

Let $\boldsymbol{\Sigma}_{\mathbf{X}} = [\rho_{ij}]_{(d\times d)}$ be the correlation matrix desired for a d-variate random vector $\mathbf{X}$ with marginals F_i, $i = 1, 2, \ldots, d$. Let $\underline{\rho}_{ij}$ and $\bar{\rho}_{ij}$ be the minimum and maximum correlations attainable between X_i and X_j. Whitt (1976) shows that these values are obtained when the bivariate vector (X_i, X_j) is generated as $(F_i(U), F_j(1 - U))$ and $(F_i(U), F_j(U))$, respectively, where U is a uniform random variable between 0 and 1. Hill and Reilly (1994) define an *extremal distribution* as a joint distribution that has $\rho_{ij} = \bar{\rho}_{ij}$ or $\rho_{ij} = \underline{\rho}_{ij}$ for every $i < j \leqslant d$. Associate a $(d(d-1)/2)$-vector $\boldsymbol{\delta}$ with each extremal distribution, where δ_{ij} takes value 1 if $\rho_{ij} = \bar{\rho}_{ij}$ and 0 otherwise. Hill and Reilly (1994) note that knowledge about any $d - 1$ components of $\boldsymbol{\delta}$ determines the rest of the vector. This comes from the result quoted earlier from Whitt (1976). Hence, there are potentially 2^{d-1} extremal correlation matrices for a d-dimensional $\mathbf{X}$.

Define $\boldsymbol{\Sigma}_0$ to be the zero-correlation matrix that corresponds to a random vector with uncorrelated components, and let $\boldsymbol{\Sigma}_\ell$, $\ell = 1, 2, \ldots, 2^{d-1}$, be the extremal correlation matrices. Hill and Reilly's method tries to match a desired correlation matrix $\boldsymbol{\Sigma}_{\mathbf{X}}$ to a point in the convex hull of the set $\{\boldsymbol{\Sigma}_\ell; \ell = 0, 1, \ldots, 2^{d-1}\}$. The problem, formally, is to determine a set of composition probabilities λ_ℓ, $\ell = 0, 1, \ldots, 2^{d-1}$, that satisfy

$$\sum_{\ell=0}^{2^{d-1}} \lambda_\ell \left[\delta_{ij}^\ell \bar{\rho}_{ij} + \left(1 - \delta_{ij}^\ell\right)\underline{\rho}_{ij}\right] = \rho_{ij}, \quad \forall i < j,\ i, j = 1, \ldots, d, \tag{14}$$

$$\sum_{\ell=0}^{2^{d-1}} \lambda_\ell = 1, \tag{15}$$

$$\lambda_\ell \geqslant 0, \quad \ell = 0, \ldots, 2^{d-1}, \tag{16}$$

where δ_{ij} is as defined earlier. Once the λ_ℓ values are determined, the constructed joint distribution (a probabilistic mixture of the extremal distributions and the independent distribution) can be sampled by first generating an index ℓ from the discrete distribution $\{\lambda_\ell; \ell = 0, \ldots, 2^{d-1}\}$ and then producing a sample from the distribution that corresponds to the chosen index.

In many cases, there are an infinite number of solutions to the Equations (14)–(16). Hill and Reilly (1994) suggest formulating and solving a linear program with an appropriate objective function with Equations (14)–(16) as constraints to select a composite probability density function with desirable characteristics. Additional constraints might also be included; for instance, the frequency of sampling independent random vectors can be minimized by setting $\lambda_0 = 0$.

This approach suffers some major drawbacks. Extremal distributions have a distinctive structure; for instance, in the d-uniform random vector case, each extremal distribution places the entire mass uniformly on a chosen diagonal of the $\Re^d$-unit hypercube. Thus, this mixing procedure constructs distributions that can be quite unnatural to the environment being modeled. The correlation-matching problem defined via (14)–(16) grows exponentially with dimension, which makes the linear programs harder to solve in higher dimensions. The composition finding problem might also have no solution for some feasible correlation matrices. This happens when the set of all feasible correlation matrices for a given set of marginals does not lie within the convex polytope defined by (14)–(16). Ghosh and Henderson (2002a) give an example where this occurs. In light of these difficulties, this method is better suited for low dimensional random vectors.

5 Conclusion

A message the reader would consistently get from this chapter is that multivariate input process modeling is far from being a complete discipline. It is, if anything, a frontier of research in input modeling methodology that awaits conquering. Interesting problems which promise both methodological content as well as practical impact abound.

In terms of developing methods that purport to handle the true joint distribution of random vector input processes or approximations to it, much of the focus has been on developing parametric approximation approaches or on methods that exploit specific properties of these distribution functions. However, nonparametric approximation methods can have a significant impact and

need to be explored more. For instance, a broad and deeply researched literature on procedures like kernel density estimation await our attention and adaptation.

We see that much of the focus has been on constructing methods that match partially specified multivariate processes. The majority of these methods handle dependence in terms of linear correlations in conjunction with marginal distributions. This approach has several limitations. First, most of the available methods are not guaranteed to work for all feasible correlation and marginals specifications. This effect is muted in lower dimensions, but can pose significant challenges as the dimension of the input process grows. Thus, there is a need for methods that can demonstrably work for any given feasible combination. Second, since correlations are single-valued linear dependence measures, they can miss much of the more complex interactions between various components of a multivariate process. Therefore, there is a clear need for models that capture more dependence information than do linear correlations. The ability to control the distribution of the generated multivariate process, for instance in the form of simple constraints on the joint distribution function, is desirable. We believe this objective of capturing more complex interactions and interdependencies will get an increasing impetus once more guidance along these lines become available from practitioners.

The presentation in this chapter has mostly focused on the modeling and variate-generation aspect of the input model development for stochastic simulation. Another important aspect of the simulation input model development is fitting models to historical data. Although it is common practice to assume that the desired marginal distributions and dependence structure are given in the current input modeling research, it is critically important to develop methods for fitting input models when only raw data generated by an unknown process are available. Recently, Biller and Nelson (2005) propose an automatic and statistically valid algorithm to fit stochastic models to dependent univariate time series for simulation. Development of similar methods for multivariate time-series input processes promises to be a very interesting research area.

References

Arnold, B.C., Castillo, E., Sarabia, J.M. (2001). Conditionally specified distributions. *Statistical Science* 16, 249–274.

Bedford, T., Cooke, R.M. (2002). Vines: A new graphical model for dependent random variables. *The Annals of Statistics* 30 (4), 1031–1068.

Biller, B., Nelson, B.L. (2003). Modeling and generating multivariate time-series input processes using a vector autoregressive technique. *ACM Transactions on Modeling and Computer Simulation* 13 (3), 211–237.

Biller, B., Nelson, B.L. (2005). Fitting arta processes for simulation. *Operations Research* 53 (3), 549–559.

Block, H.W., Langberg, N.A., Stoffer, D.S. (1990). Time series models for non-Gaussian processes. In: Block, W.H., Sampson, A.R., Savits, T.H. (Eds.), *Topics in Statistical Dependence*. Institute of Mathematical Statistics, Hayward, CA, pp. 69–83.

Burman, D. (1981). Insensitivity in queueing systems. *Advances in Applied Probability* 13, 846–859.
Cario, M.C., Nelson, B.L. (1996). Autoregressive to anything: Time-series input processes for simulation. *Operations Research Letters* 19, 51–58.
Cario, M.C., Nelson, B.L. (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., Nelson, B.L. (1998). Numerical methods for fitting and simulating autoregressive-to-anything processes. *INFORMS Journal on Computing* 10, 72–81.
Charlier, C.V.L. (1906). Uber die Darstellung willkurlicher Funktionen. *Arkiv for Matematik, Astronomi och Fysik* 2 (20), 8–20.
Clemen, R.T., Reilly, T. (1999). Correlations and copulas for decision and risk analysis. *Management Science* 45, 208–224.
Cooke, R.M. (1997). Markov and entropy properties of tree- and vine-dependent variables. In: *Proceedings of the ASA Section on Bayesian Statistical Science*. American Statistical Association, Alexandria, VA, pp. 166–175.
Cooke, R.M. (2003). The vine copula method in higher dimensions. Personal communication.
DeBrota, D.J., Roberts, S.D., Swain, J.J., Dittus, R.S., Wilson, J.R., Venkatraman, S. (1988). Input modeling with the Johnson system of distributions. In: Haigh, P.L., Comfort, J.C., Abrams, M.A. (Eds.), *Proceedings of the 1988 Winter Simulation Conference*. IEEE Press, Piscataway, NJ, pp. 165–179.
Deheuvels, P. (1977). Estimation non parametrique de la densité par histogrammes generalisés (ii). *Publications de l'Institut de Statistique de l'Université de Paris* 22, 1–23.
Devroye, L., Györfi, L. (1985). *Nonparametric Density Estimation: The L_1 View*. Wiley, New York.
Edgeworth, F.Y. (1898). Miscellaneous applications of the calculus of probabilities, contd. *Journal of the Royal Statistical Society* 61, 119–131.
Fang, K.-T., Kotz, S., Ng, K.W. (1990). *Symmetric Multivariate and Related Distributions*. *Monographs on Statistics and Applied Probability*, vol. 36. Chapman and Hall, New York.
Farin, G. (1990). *Curves and Surfaces for Computer Aided Geometric Design: A Practical Guide*, 2nd edition. Academic Press, New York.
Gelman, A., Speed, T.P. (1993). Characterizing a joint probability distribution by conditionals. *Journal of the Royal Statistical Society* 55 (1), 185–188.
Geman, S., Geman, D. (1984). Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 721–741.
Ghosh, S. (2004). Dependence in stochastic simulation models. PhD thesis, Department of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY.
Ghosh, S., Henderson, S.G. (2001). Chessboard distributions. In: Peters, B.A., Smith, J.S., Medeiros, D.J., Rohrer, M.W. (Eds.), *Proceedings of the 2001 Winter Simulation Conference*. IEEE Press, Piscataway, NJ, pp. 385–393.
Ghosh, S., Henderson, S.G. (2002a). Chessboard distributions and random vectors with specified marginals and covariance matrix. *Operations Research* 50 (5), 820–834.
Ghosh, S., Henderson, S.G. (2002b). Properties of the NORTA method in higher dimensions. In: Yücesan, E., Chen, C.H., Snowdon, J.L., Charnes, J.M. (Eds.), *Proceedings of the 2002 Winter Simulation Conference*. IEEE Press, Piscataway, NJ, pp. 263–269.
Gilks, W.R., Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* 41, 337–348.
Hannan, E.J. (1970). *Multiple Time Series*. Wiley, New York, NY.
Hill, R.R., Reilly, C.H. (1994). Composition for multivariate random variables. In: Sadowski, D.A., Seila, A.F., Tew, J.D., Manivannan, S. (Eds.), *Proceedings of the 1994 Winter Simulation Conference*. IEEE Press, Piscataway, NJ, pp. 332–339.
Hörmann, W. (1995). A rejection technique for sampling from t-concave distributions. *ACM Transactions on Mathematical Software* 21, 182–193.
Hörmann, W., Leydold, J., Derflinger, G. (2001). Universal methods of non-uniform random variate generation. Technical report, Department for Applied Statistics and Data Processing, University of Economics and Business Administration, Vienna, Austria.

Hotelling, H., Pabst, M.R. (1936). Rank correlation and tests of significance involving no assumption of normality. *Annals of Mathematical Statistics* 7, 29–43.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall, London.

Johnson, M.E. (1987). *Multivariate Statistical Simulation*. Wiley, New York.

Johnson, N.L. (1949a). Systems of frequency curves generated by methods of translation. *Biometrika* 36, 149–176.

Johnson, N.L. (1949b). Bivariate distributions based on simple translation systems. *Biometrika* 36, 297–304.

Johnson, N.L., Kotz, S. (1972). *Distributions in Statistics: Continuous Distributions*. Wiley, New York.

Johnson, N.L., Kotz, S., Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. Wiley, New York.

Jong-Dev, K. (1984). Measures of dependence. In: Krishnaiah, P.R., Sen, P.K. (Eds.), *Handbook of Statistics*, vol. 4. Elsevier (North-Holland), Amsterdam.

Kotz, S., Balakrishnan, N., Johnson, N.L. (2000). *Continuous Multivariate Distributions, volume 1: Models and Applications*. Wiley, New York.

Kruskal, W. (1958). Ordinal measures of association. *Journal of the American Statistical Association* 53, 814–861.

Kurowicka, D., Cooke, R.M. (2002). The vine copula method for representing high dimensional dependent distributions: Application to continuous belief nets. In: Yücesan, E., Chen, C.H., Snowdon, J.L., Charnes, J.M. (Eds.), *Proceedings of the 2002 Winter Simulation Conference*. IEEE Press, Piscataway, NJ, pp. 270–278.

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., Kelton, W.D. (2000). *Simulation Modeling and Analysis*, 3rd edition. McGraw-Hill, New York.

Lewis, P.A.W., McKenzie, E., Hugus, D.K. (1989). Gamma processes. *Communications in Statistics – Stochastic Models* 5, 1–30.

Leydold, J. (1998). A rejection technique for sampling from log-concave multivariate distributions. *ACM Transactions on Modeling and Computer Simulation* 8, 254–280.

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

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

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

Lütkepohl, H.L. (1993). *Introduction to Multiple Time Series Analysis*. Springer-Verlag, New York.

Mackenzie, G.R. (1994). Approximately maximum-entropy multivariate distributions with specified marginals and pairwise correlations. PhD thesis, Department of Decision Sciences, University of Oregon, Eugene, OR.

Marsaglia, G., Bray, T.A. (1964). A convenient method for generating normal variates. *SIAM Review* 6 (3), 260–264.

Meeuwissen, A.M.H., Bedford, T.J. (1997). Minimal informative distributions with given rank correlation for use in uncertainty analysis. *Journal of Statistical Computation and Simulation* 57 (4), 143–175.

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

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

Mirka, G.A., Glasscock, N.F., Stanfield, P.M., Wilson, J.R. (2000). An empirical approach to characterizing trunk muscle coactivation using simulation input modeling techniques. *Journal of Biomechanics* 33, 1701–1704.

Nelsen, R.B. (1999). *An Introduction to Copulas*. *Lecture Notes in Statistics*, vol. 139. Springer-Verlag, New York.

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

Parzen, E. (1962). On the estimation of a probability density function and mode. *Annals of Mathematical Statistics* 33, 1065–1076. Available at *http://citeseer.ist.psu.edu/parzen62estimation.html*.

Pearson, K. (1895). Contributions to the mathematical theory of evolution II: Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society A* 186, 343–414.

Ramberg, J.S., Schmeiser, B.W. (1974). An approximate method for generating asymmetric random variables. *Communications of the ACM* 17, 78–82.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics* 23, 470–472.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics* 27, 832–837.

Schmeiser, B., Deutsch, S. (1977). A versatile fourparameter family of probability distributions, suitable for simulation. *AIIE Transactions* 9, 176–182.

Schmeiser, B., Lal, R. (1982). Bivariate gamma random vectors. *Operations Research* 30, 355–374.

Silverman, B. (1982). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

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., Wilson, J.R., Mirka, G.A., Glasscock, N.F., Psihogios, J.P., Davis, J.R. (1996). Multivariate input modeling with Johnson distributions. In: Charnes, J.M., Morrice, D.J., Brunner, D.T., Swain, J.J. (Eds.), *Proceedings of the 1996 Winter Simulation Conference*. IEEE Press, Piscataway, NJ, pp. 1457–1464.

Stuart, A., Ord, J.K. (1987). *Kendall's Advanced Theory of Statistics*, 5th edition. Oxford University Press, New York, NY.

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

Wagner, M.A.F., Wilson, J.R. (1995). Graphical interactive simulation input modeling with bivariate Bézier distributions. *ACM Transactions on Modeling and Computer Simulation* 5, 163–189.

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

Walker, A.J. (1977). An efficient method for generating discrete random variables with general distributions. *ACM Transactions on Mathematical Software* 3, 253–256.

Wand, M.P., Jones, M.C. (1995). *Kernel Smoothing*. Chapman and Hall, London.

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

Whitt, W. (1976). Bivariate distributions with given marginals. *The Annals of Statistics* 4, 1280–1289.