

COMPARING TWO SYSTEMS: BEYOND COMMON RANDOM NUMBERS

Samuel M. T. Ehrlichman
Shane G. Henderson

School of Operations Research and Information Engineering
Cornell University
Ithaca, NY 14850, U.S.A.

ABSTRACT

Suppose one wishes to compare two closely related systems via stochastic simulation. Common random numbers (CRN) involves using the same streams of uniform random variates as inputs for both systems to sharpen the comparison. One can view CRN as a particular choice of copula that gives the joint distribution of the inputs of both systems. We discuss the possibility of using more general copulae, including simple examples that show how this can outperform CRN.

1 INTRODUCTION

Let X and Y be random variables quantifying the performance of two systems. Consider the problem of determining which of these systems has greater mean performance. A typical stochastic simulation approach to this problem is to generate an IID sequence $(X_j, Y_j : j = 1, \dots, n)$ of pairs of random variables where $X_j \stackrel{d}{=} X$ and $Y_j \stackrel{d}{=} Y$, $j = 1, \dots, n$, and estimate

$$E(X - Y) \approx \frac{1}{n} \sum_{j=1}^n (X_j - Y_j).$$

The sign of the resulting estimator indicates which of the two systems is preferable.

It is crucial that the random vectors (X_j, Y_j) be IID in order for the usual limit theorems to hold. However, there is no reason that X_j and Y_j must be independent for fixed j . Indeed, it may be helpful to induce such dependence; if $\text{cov}(X_j, Y_j) > 0$, then

$$\text{var}(X_j - Y_j) < \text{var} X + \text{var} Y.$$

The right-hand side, above, is the variance that would be achieved if X_j and Y_j were sampled independently.

A particularly simple method for inducing positive dependence between X and Y is common random number

(CRN) sampling (e.g., [Kelton 2006](#)), discussed in §2. Our purpose in this paper is to propose a more general technique for introducing such dependence; the technique has a quite similar flavor to CRN sampling, and indeed has CRN sampling as a special case.

The outline of the paper is as follows. Section 2 establishes notation, defines CRN sampling, and introduces our new method. Section 3 discusses earlier work on the subject, especially on known conditions under which CRN is optimal (in a sense to be defined). Section 4 proposes using a particular class of copula, the Gaussian copula, and shows its effectiveness in several toy examples. Section 5 discusses two algorithms for computing an optimal Gaussian copula. In §6 we prove a key property of the set of optimal Gaussian copulae in a particularly simple case. Section 7 offers concluding remarks and directions for further research.

In the sequel, we drop the subscript j and make reference to the joint distribution of the random vector (X, Y) .

2 COMMON RANDOM NUMBERS

Common random number sampling entails using identical sequences $\mathbf{U}_X = \mathbf{U}_Y = \mathbf{U} = (U_1, U_2, \dots)$ of pseudorandom variates to compute both X and Y . This can often be accomplished by resetting a seed for a pseudorandom number generator to a common value s for simulating both X and Y . If the ways in which X and Y are computed in terms of \mathbf{U} are fairly similar, we may hope that this technique induces the positive dependence, and hence variance reduction, discussed in the previous section.

We shall ignore the fact that \mathbf{U} is actually deterministic and treat it as random for the remainder of this paper. Under this convention, \mathbf{U} is treated as a sequence of IID uniform $[0,1]$ random numbers. In fact, we will assume further that \mathbf{U} has finite dimension d , i.e., $\mathbf{U} = (U_1, \dots, U_d)$. Technically this is without loss of generality; in fact, we could even assume $d = 1$ since there exists a bijection between $[0, 1]$ and $[0, 1]^\infty$. But practically speaking, such a bijection is not particularly useful. Hence, the assumption

of finite dimensionality may limit the situations in which the approach we discuss below is applicable.

Let us make the dependence of X and Y on \mathbf{U} explicit by defining functions $f_U, g_U : [0, 1]^d \rightarrow \mathbf{R}$ so that $X = f_U(\mathbf{U}_X)$ and $Y = g_U(\mathbf{U}_Y)$, for $\mathbf{U}_X, \mathbf{U}_Y \sim \mathcal{U}([0, 1]^d)$. Notice that if X or Y depends on $j < d$ uniform random variables, then the function f_U or g_U will simply depend on the first j components of \mathbf{U}_X or \mathbf{U}_Y .

Standard, or IID, sampling consists of sampling the random vector $(\mathbf{U}_X, \mathbf{U}_Y)$ according to the uniform probability measure on $[0, 1]^{2d}$. We denote this measure by P_{iid} . In contrast, CRN sampling consists of sampling under the probability measure P_{CRN} where

1. \mathbf{U}_X and \mathbf{U}_Y are each uniform on $[0, 1]^d$, and
2. $\mathbf{U}_X = \mathbf{U}_Y$ P_{CRN} -almost surely.

Both probability measures described above are examples of *copulae* on $[0, 1]^{2d}$. That is, they are distributions on this hypercube having uniform marginals. In this paper, we consider the possibility of using other copulae that satisfy Condition 1. We say that a copula on $[0, 1]^{2d}$ satisfying Condition 1 is *admissible*, and we denote by \mathcal{C} the set of all such copulae. Within any given class of copulae, a copula minimizing $\text{var}(X - Y)$ is called *optimal* in that class. If $\text{var}(X - Y) = 0$ is achieved, the copula is called *perfect*. Of course, $\text{var} X$ and $\text{var} Y$ are unaffected by the choice of admissible copula. This implies that we cannot expect to find a perfect copula except possibly in the case where $\text{var} X = \text{var} Y$.

3 PREVIOUS WORK

At its essence, our work involves a computational approach to constructing a coupling between two stochastic systems. For excellent reviews of coupling see Lindvall (1992) and Thorisson (2000). Schmeiser and Kachitvichyanukul (1986) described a number of approaches for coupling two random variables based on generation methods other than inversion. Devroye (1990) developed various couplings between two random vectors that attempts to maximize the number of components that are identical. Glasserman and Yao (1992) consider the question of when common random numbers is optimal for a class of performance measures that includes the variance of the difference between two random variables as considered here. As noted there, this question is poorly defined without further structure, which they impose in various ways. Glasserman and Yao (2004) provide a characterization of optimal couplings using a property they call the “nonintersection” property.

Throughout the remainder of the paper we will make considerable use of standard methods from linear algebra. The necessary background can be found in many books, e.g., Horn and Johnson (1985).

4 GAUSSIAN COPULAE

We now restrict our attention to a smaller class of copulae than \mathcal{C} , the class \mathcal{G} of *Gaussian copulae* that are admissible. In our setting, a Gaussian copula on $[0, 1]^{2d}$ is a probability measure P such that the random vector $(\mathbf{Z}_X, \mathbf{Z}_Y)$, defined componentwise by

$$\begin{aligned} Z_X[i] &= \Phi^{-1}(U_X[i]), \\ Z_Y[i] &= \Phi^{-1}(U_Y[i]), \end{aligned}$$

for $i = 1, \dots, d$, has a multivariate normal distribution with standard marginals under P . Here, Φ denotes the standard normal cdf.

Let

$$\Sigma = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{XY}^T & \Sigma_{YY} \end{bmatrix}$$

be the covariance matrix of $(\mathbf{Z}_X, \mathbf{Z}_Y)$, where the blocks are $d \times d$ matrices. In order for \mathbf{U}_X and \mathbf{U}_Y to be uniform on $[0, 1]^d$ (Condition 1 of §2) we must have $\Sigma_{XX} = \Sigma_{YY} = \mathbf{I}_d$, the $d \times d$ identity matrix. Therefore we consider only covariance matrices of the form

$$\Sigma = \begin{bmatrix} \mathbf{I}_d & \Sigma_{XY} \\ \Sigma_{XY}^T & \mathbf{I}_d \end{bmatrix}. \tag{1}$$

A positive semidefinite matrix of the form (1) will be called *admissible*, and we denote by \mathcal{S}_d the set of all such matrices. We denote by P_Σ the copula on $[0, 1]^{2d}$ associated with the covariance matrix Σ .

It is immediate that both P_{iid} and P_{CRN} are elements of \mathcal{G} . IID sampling corresponds to $\Sigma = \mathbf{I}_{2d}$, or in other words $\Sigma_{XY} = \mathbf{0}_d$, the $d \times d$ zero matrix. On the other hand, Condition 2 of §2 describing CRN sampling corresponds to $\Sigma_{XY} = \Sigma_{YX} = \mathbf{I}_d$.

We can consider X and Y to depend on the Gaussian random vector $(\mathbf{Z}_X, \mathbf{Z}_Y)$ directly. Thus, in order to simplify notation, we introduce functions $f, g : \mathbb{R}^{2d} \rightarrow \mathbb{R}$ given by

$$\begin{aligned} f(\mathbf{z}) &= f_U(\Phi^{-1}(z[1]), \dots, \Phi^{-1}(z[d])), \\ g(\mathbf{z}) &= g_U(\Phi^{-1}(z[1]), \dots, \Phi^{-1}(z[d])). \end{aligned}$$

Examples 1 and 2, below, are simple cases where a Gaussian copula outperforms both CRN and independent sampling.

Example 1. Take $d = 2$, $f(\mathbf{Z}_X) = (Z_X[1] + Z_X[2])/\sqrt{2}$, and $g(\mathbf{Z}_Y) = Z_Y[1]$. Here X and Y have linear relationships to \mathbf{Z}_X and \mathbf{Z}_Y , respectively. The random variable X depends on \mathbf{Z}_X through both its components equally, whereas Y depends only upon the first component of \mathbf{Z}_Y . The admissible covariance matrix with

$\Sigma_{XY} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ defines an optimal, and in fact a perfect, Gaussian copula for this problem; this covariance matrix corresponds to setting $\mathbf{Z}_Y[1] = (Z_X[1] + Z_X[2])/\sqrt{2}$ and $\mathbf{Z}_Y[2] = (Z_X[1] - Z_X[2])/\sqrt{2}$, so that $X = Y$.

Observe that the columns of Σ_{XY} have L_2 norm 1. We may interpret this fact to mean that the optimal copula results in perfect correlation between the appropriate linear functions of \mathbf{Z}_X and \mathbf{Z}_Y . This copula is strongly related to CRN sampling in that \mathbf{U}_Y is a deterministic transformation of \mathbf{U}_X . We prove that this happens whenever f and g are linear in §6.

Example 2. Take $d = 1$, $f_U(U_X) = \chi_{[.5, .6]}(U_X)$, and $g_U(U_Y) = \chi_{[.7, .8]}(U_Y)$. Here, χ_A denotes the indicator function (characteristic function) of the set A . In order to maximize the covariance of these indicator random variables, we would like to have the events $U_X \in [.5, .6]$ and $U_Y \in [.7, .8]$ tend to occur at the same time. In fact, a copula satisfying $U_X \in [.5, .6] \iff U_Y \in [.7, .8]$ would be perfect.

Unfortunately, there is no Gaussian copula satisfying this condition. The optimal Gaussian copula, in contrast, is given by the covariance matrix $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, where ρ is chosen to maximize the probability of $(Z_X, Z_Y) \in [\Phi^{-1}(.5), \Phi^{-1}(.6)] \times [\Phi^{-1}(.7), \Phi^{-1}(.8)]$. It is easy to see that $\rho = 1$ sets this probability to zero, so clearly $\rho = 1$ is not optimal. On the other hand, it is intuitive that $\rho > 0$ is desirable since we want Z_X and Z_Y to tend to have the same sign. Figure 1 shows that the optimal value of ρ is about .97. This value yields $\text{var}(X - Y) \approx .15$. Contrast this result with Example 1, where the optimal Gaussian copula had every column of Σ_{XY} having norm 1.

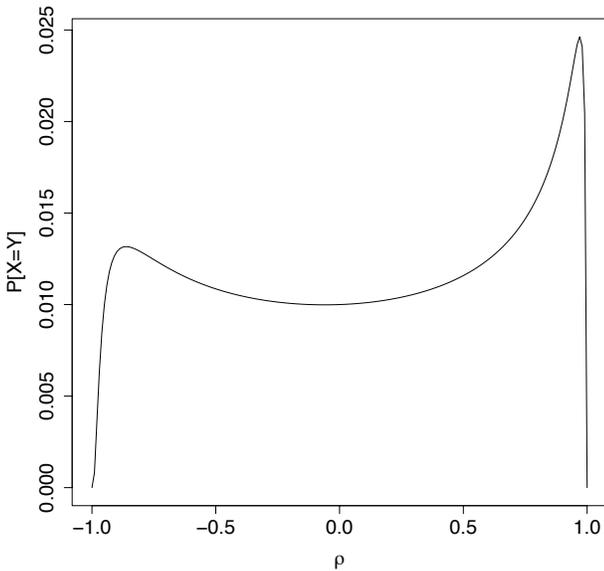


Figure 1: $P[X = Y]$ as a function of ρ .

5 FINDING AN OPTIMAL GAUSSIAN COPULA

The primary reason we have restricted attention to Gaussian copulae is that they are easily parameterized by a covariance matrix. This allows us to perform a numerical search for a locally optimal copula. We formulate the optimization problem in two distinct ways. Both involve maximizing the covariance between X and Y , but the underlying space over which the optimization is performed differs.

The first formulation is a nonlinear semidefinite program (NLP-SDP) (e.g., Kočvara and Stingl 2003). The decision variable in the optimization problem is the covariance matrix Σ of the Gaussian copula, which varies over the feasible region of admissible covariance matrices \mathcal{S}_d as defined in (1). Let \mathbf{Z} be a $2d$ -dimensional standard multivariate normal random vector. We then wish to maximize $Eh(\Sigma, \mathbf{Z})$, where

$$h(\Sigma, \mathbf{Z}) = f\left(\left(\Sigma^{1/2}\mathbf{Z}\right)[1, \dots, d]\right) g\left(\left(\Sigma^{1/2}\mathbf{Z}\right)[d+1, \dots, 2d]\right). \tag{2}$$

Here, $\Sigma^{1/2}$ denotes the Cholesky factor of Σ .

One can view $\Sigma^{1/2}$ as a differentiable function (admittedly complicated) of Σ , so that (2) is differentiable in Σ for each fixed \mathbf{Z} . One might then apply gradient-based methods for performing the optimization. Unfortunately, actually computing the gradient is difficult. One might resort to some rather complicated approach based on infinitesimal perturbation, or a more straightforward approach based on finite differences. In either case there are computational disadvantages, so we turn to a different formulation that seems more readily adapted to computation.

In our second formulation, rather than treating the covariance matrix Σ as the decision variable, we optimize over the space of all appropriate linear transformations of $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{2d})$. The key to this formulation is the following proposition. A matrix M with at least as many rows as columns is called orthogonal if $M^T M$ gives the identity matrix.

Proposition 1. Let Σ_{XY} and M_2 be $d \times d$ matrices such that

$$M := \begin{bmatrix} \Sigma_{XY} \\ M_2 \end{bmatrix}$$

is orthogonal. Then the covariance matrix of

$$\begin{bmatrix} \mathbf{Z}_X \\ \mathbf{Z}_Y \end{bmatrix} := \begin{bmatrix} \mathbf{Z}[1, \dots, d] \\ M^T \mathbf{Z} \end{bmatrix}$$

is given by

$$\Sigma = \begin{bmatrix} \mathbf{I}_d & \Sigma_{XY} \\ \Sigma_{XY}^T & \mathbf{I}_d \end{bmatrix}.$$

Conversely, if Σ is admissible then there exists \mathbf{M}_2 such that

$$\begin{bmatrix} \Sigma_{XY} \\ \mathbf{M}_2 \end{bmatrix}$$

is orthogonal.

Proof. For the first statement, we have

$$\begin{aligned} \text{cov} \begin{bmatrix} \mathbf{Z}_X & \mathbf{Z}_Y \end{bmatrix} &= E \begin{bmatrix} \mathbf{Z}_X \mathbf{Z}_X^T & \mathbf{Z}_X \mathbf{Z}^T \mathbf{M} \\ \mathbf{M}^T \mathbf{Z} \mathbf{Z}^T & \mathbf{M}^T \mathbf{Z} \mathbf{Z}^T \mathbf{M} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I}_d & [\mathbf{I}_d \ \mathbf{0}_d] \mathbf{M} \\ \mathbf{M}^T \begin{bmatrix} \mathbf{I}_d \\ \mathbf{0}_d \end{bmatrix} & \mathbf{M}^T \mathbf{M} \end{bmatrix} \\ &= \Sigma. \end{aligned}$$

Conversely, suppose $\begin{bmatrix} \mathbf{Z}_X & \mathbf{Z}_Y \end{bmatrix}$ has admissible covariance matrix Σ . Then

$$\begin{aligned} \text{cov} \left(\mathbf{Z}_Y - \Sigma_{XY}^T \mathbf{Z}_X \right) &= \begin{bmatrix} -\Sigma_{XY}^T & \mathbf{I}_d \end{bmatrix} \Sigma \begin{bmatrix} -\Sigma_{XY} \\ \mathbf{I}_d \end{bmatrix} \\ &= \begin{bmatrix} -\Sigma_{XY}^T & \mathbf{I}_d \end{bmatrix} \begin{bmatrix} \mathbf{0}_d \\ \mathbf{I}_d - \Sigma_{XY}^T \Sigma_{XY} \end{bmatrix} \\ &= \mathbf{I}_d - \Sigma_{XY}^T \Sigma_{XY}, \end{aligned}$$

implying that the matrix on the right is positive semidefinite. Therefore, there exists \mathbf{M}_2 such that $\mathbf{M}_2^T \mathbf{M}_2 = \mathbf{I}_d - \Sigma_{XY}^T \Sigma_{XY}$ and hence $\mathbf{M}^T \mathbf{M} = \Sigma_{XY}^T \Sigma_{XY} + \mathbf{M}_2^T \mathbf{M}_2 = \mathbf{I}_d$. \square

Proposition 1 demonstrates that we can compute an optimal Gaussian copula by solving an optimization problem on the space $\mathcal{V}_{2d,d}$ of $2d \times d$ orthogonal matrices rather than on \mathcal{S}_d . The space of such matrices is an example of a *Stiefel manifold*; see [Edelman, Arias, and Smith \(1999\)](#) for a discussion of the geometry of these and related manifolds, and of optimization algorithms thereon. In this formulation, our objective function (2) is replaced by

$$h(\mathbf{M}, \mathbf{Z}) = f(\mathbf{Z}_X) g(\mathbf{M}^T \mathbf{Z}). \quad (3)$$

Again we wish to maximize $Eh(\mathbf{M}, \mathbf{Z})$, but this time the feasible region is $\mathbf{M} \in \mathcal{V}_{2d,d}$.

It is now straightforward to compute derivatives of $h(\cdot, \mathbf{Z})$ with respect to \mathbf{M} for a fixed \mathbf{Z} . We have that

$$\begin{aligned} h'_{ij}(\mathbf{M}, \mathbf{Z}) &:= \frac{\partial h(\mathbf{M}, \mathbf{Z})}{\partial \mathbf{M}[i, j]} \\ &= f(\mathbf{Z}_X) g_j(\mathbf{M}^T \mathbf{Z}) \mathbf{Z}[i], \end{aligned}$$

where $g_j(x)$ is the partial derivative of $g(\cdot)$ with respect to the j th component, evaluated at x .

One can use these derivatives in various gradient-based optimization approaches such as stochastic approximation. Our optimization approach is based upon Sample Average Approximation (SAA) (e.g., [Shapiro 2004](#)). SAA is a general method for solving optimization problems of the form

$$\max_{x \in \mathcal{X}} Eh(x, \xi)$$

where ξ is a random object. Initially a small ‘‘pilot’’ sample ξ_1, \dots, ξ_m is generated. These values are then treated as fixed, and the optimization problem is replaced by

$$\max_{x \in \mathcal{X}} \frac{1}{m} \sum_{i=1}^m h(x, \xi_i).$$

Since the sample is fixed, the problem can be viewed as a deterministic optimization problem, and one can then employ specialized deterministic optimization algorithms to solve the problem. We use exactly this approach using optimization algorithms designed for differentiable functions over a Stiefel manifold. The solution to the optimization problem using a sample of size m , say, $\mathbf{Z}_1, \dots, \mathbf{Z}_m$ yields a matrix \mathbf{M}_m^* that defines a copula, which can then be used in a ‘‘production’’ run to actually compare the systems in question. Under mild regularity conditions, it is known that \mathbf{M}_m^* will not only be a locally optimal solution for the sample-average problem, but will also be a nearly locally optimal solution for the true problem. See [Shapiro \(2004\)](#), Proposition 7, p. 363 and [Bastin, Cirillo, and Toint \(2006\)](#).

Example 3. We conclude this section with an example of a stochastic activity network (e.g., [Avramidis and Wilson 1993](#)). The network in Figure 2 is an abstraction of a set of jobs which must be completed. Each arc corresponds to a job. Nodes represent constraints on the order in which the jobs must be performed. All the jobs whose arcs enter a given node must be completed before any job whose arc leaves that node commences. The arcs are labelled by random variables corresponding to the length of time required by each task. Two nodes are distinguished as the source and the sink, respectively representing the state in which no tasks have begun and the state in which all tasks are completed. The total completion time for the set of all tasks is equal to the maximum length of all paths from source to sink.

Let us compare two possible configurations of the stochastic activity network in Figure 2, where the different configurations correspond to different joint distributions on the activity times V_1, \dots, V_4 . In Configuration 1, the activity times are IID exponential with rate 1. In Configuration

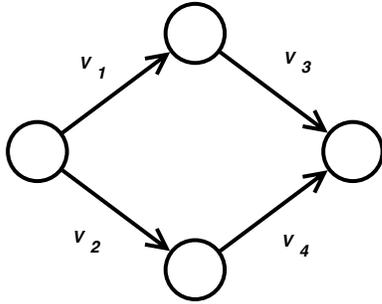


Figure 2: Stochastic Activity Network Example.

2, V_1 and V_2 are as in Configuration 1, but V_3 and V_4 are exponential conditional on V_1 and V_2 with respective rates $\frac{1}{2}(1 + V_2)$ and $\frac{1}{2}(1 + V_1)$. Let X and Y respectively be the completion times of Configurations 1 and 2.

The functions connecting the underlying normal random variates to the service times are given by

$$f(\mathbf{z}) = \max(-\log \Phi(z_1) - \log \Phi(z_3), -\log \Phi(z_2) - \log \Phi(z_4)),$$

$$g(\mathbf{z}) = \max\left(-\log \Phi(z_1) - \frac{2}{1 - \log \Phi(z_2)} \log \Phi(z_3), -\log \Phi(z_2) - \frac{2}{1 - \log \Phi(z_1)} \log \Phi(z_4)\right).$$

We solved this problem using the Stiefel manifold formulation with the freely available MATLAB procedure `sgmin` (Lippert and Edelman 1999), using the SAA framework sketched above. We solved it over both the spaces $\mathcal{V}_{2d,d}$ and $\mathcal{V}_{d,d}$ and achieved quite similar resulting covariance matrices. This strongly suggests that the optimal Gaussian copula is in fact a change of variables applied to CRN sampling.

The Gaussian copula returned by the optimization procedure is defined by (4). We performed longer runs, independent from the pilot, under IID sampling, CRN sampling, and the optimal Gaussian copula. The resulting variance of $(X - Y)$ in each case is given in Table 1.

$$\Sigma_{XY} = \begin{bmatrix} .958 & -.038 & .160 & .237 \\ -.037 & .960 & .239 & .141 \\ -.158 & -.238 & .957 & -.048 \\ -.239 & -.143 & -.026 & .960 \end{bmatrix}. \quad (4)$$

We can see that the optimal covariance matrix defined (4) returned by the optimization algorithm is quite close to that of CRN sampling. Although the random variables $Z_X[i]$ and $Z_Y[i]$, $i = 1, \dots, 4$, are not identical, they are very highly correlated. However, the difference in performance between the two copulae is great, with the optimal Gaussian copula resulting in more than a 50% reduction in variance.

Table 1: Example 3 results.

Sampling Strategy	Variance
IID	5.257
CRN	0.565
OPT	0.280

6 ANALYSIS OF THE LINEAR CASE

The optimal copula given in Example 1 has the property that Σ_{XY} is itself an orthogonal matrix; equivalently, the lower d rows of this solution \mathbf{M} are all zero. A natural question to ask is under what conditions we may assume that an optimal solution of this type exists. Knowledge of such conditions would allow the optimization problem to be solved on the smaller space $\mathcal{V}_{d,d}$. Moreover, the resulting copula can be sampled from using only d independent normal variates per sample, as opposed to the $2d$ normal variates required in the general case. We do not have a complete answer to this question at present, although we are able to show that a sufficient condition is for X and Y to be linear in \mathbf{Z}_X and \mathbf{Z}_Y , respectively.

Proposition 2. *Let \mathbf{Z} be a $2d$ -dimensional random vector with mean zero and covariance matrix \mathbf{I}_{2d} . Suppose f and g are functions on \mathbb{R}^d given by $f(\mathbf{z}) = \mathbf{a}^T \mathbf{z}$, $g(\mathbf{z}) = \mathbf{b}^T \mathbf{z}$, for some $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$. Then*

$$\begin{aligned} & \max_{\mathbf{M} \in \mathcal{V}_{2d,d}} Ef([\mathbf{I}_d \ \mathbf{0}_d]^T \mathbf{Z})g(\mathbf{M}^T \mathbf{Z}) \\ &= \max_{\Sigma_{XY} \in \mathcal{V}_{d,d}} Ef([\mathbf{I}_d \ \mathbf{0}_d]^T \mathbf{Z})g([\Sigma_{XY} \ \mathbf{0}_d]^T \mathbf{Z}). \end{aligned}$$

Proof. Let LHS and RHS respectively denote the left- and right-hand sides of the desired equality. The inequality $\text{LHS} \geq \text{RHS}$ follows immediately from the fact that $[\Sigma_{XY} \ \mathbf{0}_d] \in \mathcal{V}_{2d,d}$, so we need only prove the converse.

Let $\mathbf{M} \in \mathcal{V}_{2d,d}$ be arbitrary. Let us write

$$\begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{bmatrix} \text{ and } \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \end{bmatrix}$$

for \mathbf{Z} and \mathbf{M} respectively. Then

$$\begin{aligned} Ef([\mathbf{I}_d \ \mathbf{0}_d]^T \mathbf{Z})g(\mathbf{M}^T \mathbf{Z}) &= E\mathbf{a}^T \mathbf{Z}_1 \mathbf{b}^T \mathbf{M}^T \mathbf{Z} \\ &= E\mathbf{a}^T \mathbf{Z}_1 \mathbf{Z}_1^T \mathbf{M}_1 \mathbf{b} + E\mathbf{a}^T \mathbf{Z}_1 \mathbf{Z}_2^T \mathbf{M}_2 \mathbf{b} = \langle \mathbf{a}, \mathbf{M}_1 \mathbf{b} \rangle. \end{aligned} \quad (5)$$

Now,

$$\begin{aligned} \langle \mathbf{a}, \mathbf{M}_1 \mathbf{b} \rangle^2 &\leq \|\mathbf{a}\|^2 \|\mathbf{M}_1 \mathbf{b}\|^2 \\ &\leq \|\mathbf{a}\|^2 \|\mathbf{M}_1\|_2^2 \|\mathbf{b}\|^2 \\ &\leq \|\mathbf{a}\|^2 \|\mathbf{b}\|^2. \end{aligned} \quad (6)$$

Here, $\|\mathbf{M}_1\|_2$ denotes the spectral norm (greatest absolute eigenvalue) of \mathbf{M}_1 . The first inequality is Cauchy-Schwartz. The second inequality is a property of the spectral norm. The third inequality is proven as follows: if λ is an eigenvalue of \mathbf{M}_1 with corresponding eigenvector \mathbf{w} , then

$$\begin{aligned} 1 &= \|\mathbf{w}\|^2 = \mathbf{w}^T \mathbf{M}^T \mathbf{M} \mathbf{w} \\ &= \mathbf{w}^T (\mathbf{M}_1^T \mathbf{M}_1 + \mathbf{M}_2^T \mathbf{M}_2) \mathbf{w} \\ &= \lambda^2 \|\mathbf{w}\|^2 + \mathbf{w}^T \mathbf{M}_2^T \mathbf{M}_2 \mathbf{w} \\ &\geq \lambda^2, \end{aligned}$$

since $\mathbf{M}_2^T \mathbf{M}_2$ is positive semidefinite.

Combining (5) with (6) yields $\text{LHS} \leq \|\mathbf{a}\| \|\mathbf{b}\|$. Now let $\mathbf{v} = \mathbf{b} - \frac{\|\mathbf{b}\|}{\|\mathbf{a}\|} \mathbf{a}$. Let Σ_{XY} be the Householder reflection induced by \mathbf{v} ,

$$\Sigma_{XY} = \mathbf{I}_d - \frac{2}{\|\mathbf{v}\|^2} \mathbf{v} \mathbf{v}^T.$$

It is easy to check that $\Sigma_{XY} \in \mathcal{V}_{d,d}$ and that $\langle \mathbf{a}, \Sigma_{XY} \mathbf{b} \rangle^2 = \|\mathbf{a}\|^2 \|\mathbf{b}\|^2$. This implies $\|\mathbf{a}\| \|\mathbf{b}\| \leq \text{RHS}$, completing the proof. \square

7 CONCLUSION

We have shown that it is possible to compute couplings of two random vectors that have IID components with the goal of minimizing the variance of the difference between real-valued functions of the random vectors. We use an underlying Gaussian copula because it is amenable to computation, although one could certainly consider other copula families as well. We have given simple examples where the gains beyond common random numbers are significant.

ACKNOWLEDGMENTS

This research was partially supported by National Science Foundation Grants DMI 0400287 and CMMI 0800688.

REFERENCES

Avramidis, A. N., and J. R. Wilson. 1993. Integrated variance reduction strategies. In *Proceedings of the 1993 Winter Simulation Conference*, ed. G. W. Evans, M. Molgahasemi, E. C. Russell, and W. E. Biles, 445–454. Institute of Electrical and Electronics Engineers, Inc.

Bastin, F., C. Cirillo, and P. L. Toint. 2006. Convergence theory for nonconvex stochastic programming with an application to mixed logit. *Mathematical Programming B* 108:207–234.

Devroye, L. 1990. Coupled samples in simulation. *Operations Research* 38:115–126.

Edelman, A., T. A. Arias, and S. T. Smith. 1999. The geometry of algorithms with orthogonality constraints. *SIAM journal on matrix analysis and applications* 20 (2): 303–353.

Glasserman, P., and D. D. Yao. 1992. Some guidelines and guarantees for common random numbers. *Management Science* 38:884–908.

Glasserman, P., and D. D. Yao. 2004. Optimal couplings are totally positive and more. *Journal of Applied Probability* 41:321–332.

Horn, R. A., and C. R. Johnson. 1985. *Matrix analysis*. Cambridge, UK: Cambridge University Press.

Kelton, W. D. 2006. Implementing representations of uncertainty. In *Simulation*, ed. S. G. Henderson and B. L. Nelson, Volume 13 of *Handbooks in Operations Research and Management Science*. Amsterdam: Elsevier.

Kočvara, M., and M. Stingl. 2003, June. PENNON: a code for convex nonlinear and semidefinite programming. *Optimization Methods and Software* 18 (3): 317–333.

Lindvall, T. 1992. *Lectures on the coupling method*. New York: Wiley.

Lippert, R. A., and A. Edelman. 1999. sgmin [computer software]. Available at <http://www-math.mit.edu/lippert/sgmin.html>.

Schmeiser, B., and V. Kachitvichyanukul. 1986. Correlation induction without the inverse transformation. In *Proceedings of the 1986 Winter Simulation Conference*, ed. J. Wilson, J. Henriksen, and S. Roberts, 266–274. New York, NY: Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Shapiro, A. 2004. Monte Carlo sampling methods. In *Stochastic Programming*, ed. A. Ruszczyński and A. Shapiro, Volume 10 of *Handbooks in Operations Research and Management Science*. Amsterdam: Elsevier.

Thorisson, H. 2000. *Coupling, stationarity, and regeneration*. New York, NY: Springer.

AUTHOR BIOGRAPHIES

SAMUEL M. T. EHRLICHMAN is a Ph.D. student in the School of Operations Research and Information Engineering at Cornell University. His interests include simulation-optimization, with emphasis on applications in financial engineering. He received a John McMullen Graduate Fellowship from Cornell and an NDSEG Fellowship. He expects to graduate and join Jane Street Capital, LLC as a quantitative researcher in August, 2008.

SHANE G. HENDERSON is an associate professor in the School of Operations Research and Information Engineering at Cornell University. He is the simulation area editor at *Operations Research*, and an associate editor for the *ACM Transactions on Modeling and Computer Simulation* and *Operations Research Letters*. He co-edited the handbook *Simulation* as part of Elsevier's series of Handbooks in Operations Research and Management Science, and also co-edited the Proceedings of the 2007 Winter Simulation Conference. He likes cats but is allergic to them. His research interests include discrete-event simulation and simulation optimization, and he has worked for some time with emergency services. His web page can be found via <http://www.orie.cornell.edu>.