



INFORMS Journal on Computing

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

C-NORTA: A Rejection Procedure for Sampling from the Tail of Bivariate NORTA Distributions

Soumyadip Ghosh, Raghu Pasupathy,

To cite this article:

Soumyadip Ghosh, Raghu Pasupathy, (2012) C-NORTA: A Rejection Procedure for Sampling from the Tail of Bivariate NORTA Distributions. INFORMS Journal on Computing 24(2):295-310. <http://dx.doi.org/10.1287/ijoc.1100.0447>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2012, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

C-NORTA: A Rejection Procedure for Sampling from the Tail of Bivariate NORTA Distributions

Soumyadip Ghosh

IBM T. J. Watson Research Center, Yorktown Heights, New York 10598, ghoshs@us.ibm.com

Raghu Pasupathy

Industrial and Systems Engineering, Virginia Tech, Blacksburg, Virginia 24061, pasupath@vt.edu

We propose C-NORTA, an exact algorithm to generate random variates from the tail of a bivariate NORTA random vector. (A NORTA random vector is specified by a pair of marginals and a rank or product-moment correlation, and it is sampled using the popular NORmal-To-Anything procedure.) We first demonstrate that a rejection-based adaptation of NORTA on such constrained random vector generation problems may often be fundamentally intractable. We then develop the C-NORTA algorithm, relying on strategic conditioning of the NORTA vector, followed by efficient approximation and acceptance/rejection steps. We show that, in a certain precise asymptotic sense, the sampling efficiency of C-NORTA is exponentially larger than what is achievable through a naïve application of NORTA. Furthermore, for at least a certain class of problems, we show that the acceptance probability within C-NORTA decays only linearly with respect to a defined rarity parameter. The corresponding decay rate achievable through a naïve adaptation of NORTA is exponential. We provide directives for efficient implementation.

Key words: statistics; simulation; random variable generation; multivariate distribution; correlation

History: Accepted by Marvin Nakayama, Area Editor for Simulation; received December 2009; revised July 2010, October 2010; accepted November 2010. Published online in *Articles in Advance* May 17, 2011.

1. Introduction and Motivation

We consider the question of generating random vectors with specified marginal distributions and a correlation matrix, and constrained to a feasible region characterized by linear constraints. The reader might recognize this as the usual random vector generation problem but with the additional stipulation that the generated random variates fall within a prescribed feasible region. We are particularly interested in contexts where the feasible region is so “rare” as to render naïve adaptations of existing methods for the unconstrained problem (e.g., NORTA of Cario and Nelson 1997, 1998; Avramidis et al. 2009; and copula-based methods of Sklar 1959, Nelsen 1999) unimplementable. This last statement will become much clearer further along in the paper, when we rigorously characterize the performance of a straightforward adaptation of NORTA to the current problem. The solution we present, by contrast, is a certain nontrivial extension of NORTA, accomplished through strategic conditioning and efficient approximation of the probability laws inherent to the problem.

Applications of the ability to generate samples of correlated random vectors from a constrained space seem widespread. Consider, for example, estimating tail measures of multivariate distributions in the context of rare-event simulations (Juneja

and Shahabuddin 2006). This problem is important because the quantity being estimated requires generating random vectors from a low-probability set, and inefficient sampling will have a direct bearing on the quality of the estimation. A specific application of such estimation can be found in Huang and Subramanian (2009), where a stochastic optimization problem is posed to determine the most profitable portfolio, subject to a constraint on the risk of large losses associated with tail events modeled using NORTA vectors. A similar need for such constrained random vector generation problems arises (directly or indirectly) in simulation optimization problems within various settings (e.g., production flow lines, call-center staffing, tandem production lines, and stochastic PERT) and vehicle routing problems.

Another rather novel application of the constrained random vector generation problem comes from the Bayesian inference modeling literature. A common Bayesian learning setting (Gelfand et al. 1992, Damien and Walker 2001) models the distribution of uncertain parameters using joint normal distributions, and a learning step updates to a posterior that limits the parameter values with linear constraints. The next learning step needs to then sample parameter values from this constrained joint normal space. This is typically accomplished by the naïve approach

of sampling and rejecting until a sample that satisfies all constraints is obtained. A more sophisticated, faster approach uses Markov chain Monte Carlo techniques such as the Metropolis-Hastings algorithm (Ross 2006), but these methods can produce samples with only approximately close distributions. An efficient sampler that can directly sample from the rare set would provide a big speed boost to such learning algorithms.

It is worth mentioning here that the methods we present in this paper may be mildly reminiscent of techniques that underlie importance sampling. We hasten to add, however, that the change-of-measure constructions within importance sampling are aimed at a much more specific problem, i.e., that of reducing the variance of a point estimator of a specific quantity that can be expressed as an expectation. (See Asmussen and Glynn 2007 for an introduction, and see Glasserman et al. 2000, 2002; Glasserman and Li 2005 for certain specific contexts.) We ask a different and arguably more fundamental question in this paper: How can correlated random vectors constrained to lie in a feasible region be generated efficiently? Although it is true that the methods we present apply within estimation contexts, the question we tackle often arises well before any estimation happens.

1.1. Notation

Throughout this paper, random vectors are represented in bold capitalized letters \mathbf{X}, \mathbf{N} , etc., whereas plain capitalized letters represent univariate random variables. \mathbb{R} is used to denote the real field, and $\bar{\mathbb{R}}$ is used to represent the set of extended reals, i.e., real numbers, ∞ , and $-\infty$. Call N the standard (i.e., zero mean, unit variance) univariate normal random variable with density ϕ and distribution Φ . The related function $\bar{\Phi}(\cdot) = 1 - \Phi(\cdot)$ will be used heavily. The random vector \mathbf{N} denotes the standard (zero mean, unit variance, independent marginals) bivariate joint normal random vector. (Note that the covariance matrix of \mathbf{N} is the 2×2 identity matrix.) By contrast, the random vector \mathbf{Z} represents a bivariate joint normal with standard normal marginals but a nonidentity covariance matrix. In all sections except §5, linear constraints of the form $\{c^t x = c_1 x_1 + c_2 x_2 \geq v\}$ are assumed to be posed such that the vector c is a unit vector; i.e., $c_1^2 + c_2^2 = 1$. In §5, for ease of exposition, we switch to the equivalent representation $\{c_1 x_1 + x_2 \geq v\}$.

1.2. Problem Statement and Preliminaries

The question of sampling random vectors with specified marginals and correlation matrix is well studied. Among available methods, the NORmal-To-Anything (NORTA) method (Cario and Nelson 1997, 1998) is arguably the most popular. The NORTA method

essentially involves a component-wise transformation of a multivariate normal random vector and exploits the fact that multivariate normals are easily generated (Law and Kelton 2000, p. 480). Specifically, suppose that we wish to generate independent and identically distributed (i.i.d.) replicates of a random vector $\mathbf{X} = (X_1, X_2, \dots, X_q)$ with prescribed marginal univariate distributions

$$F_i(x) = \Pr(X_i \leq x), \quad i = 1, 2, \dots, q, \quad x \in \mathcal{X}_i \subseteq \mathbb{R}$$

and product-moment or rank correlation matrix

$$\rho_{\mathbf{X}} = \rho_{\mathbf{X}}(i, j), \quad 1 \leq i, j \leq q, \quad \text{where } \rho_{\mathbf{X}}(i, j) = \text{Corr}(X_i, X_j).$$

Let $F_i^{-1}(u) = \inf\{x: F_i(x) \geq u\}$ be the inverse of the i th marginal distribution. Assume $\rho_{\mathbf{X}}$ to be feasible for the given marginals. (Not all product-moment correlations are theoretically feasible for a given pair of marginals; see Whitt 1976 for more on this point.) Then the NORTA method generates i.i.d. replicates of \mathbf{X} by the following procedure.

Algorithm 1 (NORTA)

Inputs: Marginal distributions $F_i, i = 1, 2, \dots, q$; correlation matrix $\rho_{\mathbf{X}}$.

Outputs: Random variate \mathbf{X} .

1. Generate an \mathbb{R}^d valued joint normal random vector $\mathbf{Z} = (Z_1, Z_2, \dots, Z_q)$ such that $Z_i, i = 1, 2, \dots, q$ have a standard normal distribution, and \mathbf{Z} has correlation matrix $\rho_{\mathbf{Z}}^*$.
2. Compute the vector $\mathbf{X} = (X_1, X_2, \dots, X_q)$ via $X_i = \Psi_i(Z_i) = F_i^{-1}(\Phi(Z_i))$, for $i = 1, 2, \dots, q$. (1)

The vector function $\Psi: \mathbb{R}^q \rightarrow \mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_q$ is the NORTA transformation from the joint normal \mathbf{Z} to the NORTA vector \mathbf{X} . Random vector \mathbf{X} will have the prescribed marginal distributions. To see this, note that each Z_i has a standard normal distribution so that $\Phi(Z_i)$ is uniformly distributed on $(0, 1)$, and so $F_i^{-1}(\Phi(Z_i))$ will have the required marginal distribution. The correlation matrix $\rho_{\mathbf{Z}}^*$ is chosen in a preprocessing phase so as to ensure that it induces the prescribed correlation matrix $\rho_{\mathbf{X}}$ for \mathbf{X} . This is usually the most difficult step in implementing the NORTA method but has recently been well studied (Cario and Nelson 1998, Chen 2001, Ghosh and Henderson 2003, Avramidis et al. 2009).

Whereas NORTA in its generic form addresses the question of generating \mathbf{X} on its original support \mathcal{X} , this paper concerns the question of generating NORTA random vectors from constrained regions $\mathcal{F} = \{c^t x \geq v\}$ on the support \mathcal{X} of \mathbf{X} . Of course, a naive adaptation of NORTA is certainly possible on the latter problem. Specifically, to generate random vectors $\mathbf{X} | \mathcal{F}$, simply generate unconstrained NORTA random vectors \mathbf{X} and accept only those that lie within \mathcal{F} .

Downloaded from informs.org by [128.173.125.76] on 05 February 2014, at 07:19 . For personal use only, all rights reserved.

Such a naïve rejection procedure is akin to sampling geometrically, with the acceptance probability equal to the probability assigned to \mathcal{F} by the NORTA generated joint distribution of \mathbf{X} .

How efficient is this approach as the set \mathcal{F} becomes increasingly rare? A simple example of sampling the standard bivariate normal \mathbf{N} constrained on a set $\mathcal{F} = \{c^t z \geq v\}$ is illustrative. With the specific intent of studying the behavior of the acceptance probability as the set \mathcal{F} becomes rarer, let us parameterize the constraint set as $\mathcal{F}^\lambda = \{c^t(z - \lambda d) \geq v\} = \{c^t z \geq v + \lambda c^t d\}$, where $\lambda \geq 0$ is the “rarity” or translation parameter, d is a unit vector, and the resulting half-space \mathcal{F}^λ is called the translated set. (Such a parameterization allows us to make the feasible set rarer by increasing a single parameter λ .) For each λ and d , the linear subspace that defines the boundary of the set \mathcal{F} when translated by λd gives the boundary of the translated set \mathcal{F}^λ . The random variable $c^t \mathbf{N}$ is univariate normal with zero mean and variance $c^t c = 1$. The acceptance probability is the mass $P(\lambda) = \Pr(\mathbf{N} \in \mathcal{F}^\lambda)$ assigned to \mathcal{F}^λ by the standard bivariate normal random vector \mathbf{N} . Thus, $P(\lambda)$ is the tail-probability $\bar{\Phi}(v + \lambda c^t d)$, and the set \mathcal{F}^λ becomes rarer as $\lambda \rightarrow \infty$ for $c^t d > 0$. This leads us to the following result.

PROPOSITION 1. *If $c^t d > 0$, the naïve procedure for sampling $\mathbf{N} | \mathcal{F}^\lambda$ obeys*

$$\lim_{\lambda \rightarrow \infty} \lambda e^{(v + \lambda c^t d)^2 / 2} P(\lambda) = \frac{1}{c^t d \sqrt{2\pi}}.$$

PROOF. Equation (2.1) in Lu and Li (2009) gives the following bounds for the univariate Gaussian tail:

$$\frac{z^2}{1 + z^2} \frac{1}{\sqrt{2\pi}} \frac{e^{-z^2/2}}{z} \leq \bar{\Phi}(z) \leq \frac{1}{\sqrt{2\pi}} \frac{e^{-z^2/2}}{z}, \quad z > 0. \quad (2)$$

Using these bounds for $P(\lambda) = \bar{\Phi}(v + \lambda c^t d)$ for all $\lambda \geq \lambda_0$, where λ_0 is sufficiently large such that $v + \lambda_0 c^t d > 0$, we obtain the result by multiplying throughout with $\lambda e^{(v + \lambda c^t d)^2 / 2}$ and taking the appropriate limit on either side. \square

Proposition 1 asserts that the acceptance probability of the naïve adaptation of NORTA decays exponentially as the set \mathcal{F} as defined above becomes rarer. (Although we have stated this proposition for bivariate standard normals for illustrative purposes, a similar exponential decay can be shown to hold for more general vectors under mild conditions.) As we shall see, the C-NORTA method we describe in this paper improves substantially on this by reducing the decay to a linear rate (Theorem 1). Furthermore, such gains are achieved with no corresponding increase in the effort or complexity of the generation steps within the algorithm.

1.3. Contributions

The question of generating random vectors having a prescribed correlation and prescribed marginal distributions is a well-studied problem in the stochastic simulation literature. Its extension—that of generating such random vectors but when constrained to a given space \mathcal{F} —seems to have received much less attention. In fact, apart from the naïve extension of algorithms available for the unconstrained context, we know of no other technique that is currently available. To this extent, we see the family of algorithms presented in this paper (C-NORTA) as the first (to our knowledge) available nontrivial method for solving the two-dimensional constrained version of the random vector generation problem.

Our specific contributions through this paper include the following.

1. We present an exact rejection algorithm for generating bivariate normals constrained by a single hyperplane. This algorithm is shown to possess an acceptance probability that decays linearly with respect to a certain rigorously defined rarity parameter.
2. We present C-NORTA—an exact rejection algorithm for generating bivariate random vectors of the NORTA type, constrained to a region bounded by linear constraints.
3. We demonstrate that the acceptance probability within C-NORTA algorithms is exponentially better (in a certain precisely defined asymptotic sense) than what is obtainable through the naïve adaptation of NORTA to constrained problems. Furthermore, on certain classes of problems, we demonstrate that C-NORTA achieves an acceptance probability that decays linearly with respect to the rarity parameter. The corresponding decay rate in NORTA is always exponential.

2. An Overview of C-NORTA

Recall from the previous section that we are concerned with the problem of generating a random vector $\mathbf{X} = (X_1, X_2)$ such that (i) X_1 and X_2 have the prescribed marginals F_1 and F_2 , respectively; (ii) $\text{Corr}(X_1, X_2) = \rho$ for prescribed $\rho \in (-1, 1)$; and (iii) the vector \mathbf{X} is constrained to the prescribed region $\mathcal{F} \subset \mathcal{X}_1 \times \mathcal{X}_2$. In achieving this, and in a manner much more efficient than a naïve adaptation of NORTA, we now present an algorithm called C-NORTA. Algorithm C-NORTA relies on the following two simple facts.

- (a) Under mild conditions (see §5), there exists a one-to-one mapping between the required vector $\mathbf{X} \in \mathcal{F}$ and a corresponding vector $\mathbf{Z} = (Z_1, Z_2) \in \mathcal{F}_N$ in the bivariate normal space.

(b) If (Z_1, Z_2) is a bivariate normal random vector, the distributions of the conditional random variables $Z_2 | \{Z_1 = z_1, \mathbf{Z} \in \mathcal{G}\}$ and $Z_1 | \{\mathbf{Z} \in \mathcal{G}\}$ are fully characterizable for any set $\mathcal{G} \subset \mathbb{R}^2$.

That (a) is true is evident from the basis of NORTA itself. Specifically, we know from the theory underlying NORTA (as discussed in §1.2) that a bivariate normal vector $\mathbf{Z} = \mathbf{Z}(\rho^*)$ can be chosen to have standard normal marginals and correlation ρ^* , such that the two random variables X_1 and X_2 defined via the NORTA transformation (1) have marginals F_1 and F_2 , respectively, and $\text{Corr}(X_1, X_2) = \rho$. Furthermore, the choice of ρ^* that accomplishes this is a well-studied problem; see, for instance, Cario and Nelson (1998), Avramidis et al. (2009), and Chen (2001). The image \mathcal{F}_N of the set \mathcal{F} in the normal space is then

$$\mathcal{F}_N = \{(z_1, z_2) : \psi(z_1, z_2) = (\psi_1(z_1), \psi_2(z_2)) \in \mathcal{F}\},$$

and the measures $\mathbf{Z}(\rho^*) | \mathcal{F}_N$ and $\mathbf{X} | \mathcal{F}$ are equivalent.

For exposition of (b), denote the set $\mathcal{G}(z_1) = \{z_2 \in \mathbb{R} : (z_1, z_2) \in \mathcal{G}\}$. Then, Proposition 2 asserts that, under certain conditions, the random variable $Z_2 | \{Z_1 = z_1, \mathbf{Z} \in \mathcal{G}\}$ has a normal distribution that is truncated and appropriately scaled over the feasible set $\mathcal{G}(z_1)$. Similarly, Proposition 2 states that the random variable $Z_1 | \{\mathbf{Z} \in \mathcal{G}\}$ has a distribution that is a normal mixture of certain normal probabilities.

PROPOSITION 2. *Let (Z_1, Z_2) be a bivariate normal random vector with standard normal marginals and correlation $\text{Corr}(Z_1, Z_2) = \rho^*$. Assume that $\mathcal{G}(z_1) = \{z_2 \in \mathbb{R} : (z_1, z_2) \in \mathcal{G}\}$ is nonempty for all $z_1 \in \mathbb{R}$, and denote $\mathcal{T}(z_1, \rho^*) = \{t : t\sqrt{1 - \rho^{*2}} + \rho^*z_1 \in \mathcal{G}(z_1)\}$. Then*

(i) $Z_1 | \{\mathbf{Z} \in \mathcal{G}\}$ has the density function

$$f_1(z_1) = \frac{1}{\Pr\{\mathbf{Z} \in \mathcal{G}\}} \phi(z_1) \Pr\{Z_2 \in \mathcal{T}(z_1, \rho^*)\} \quad z_1 \in \mathbb{R}.$$

(ii) $Z_2 | \{Z_1 = z_1, \mathbf{Z} \in \mathcal{G}\}$ has the density function

$$f_{z_1}(z_2) = \frac{1}{\Pr\{Z_2 \in \mathcal{G}(z_1) | Z_1 = z_1\}} \cdot \frac{1}{\sqrt{1 - \rho^{*2}}} \phi\left(\frac{z_2 - \rho^*z_1}{\sqrt{1 - \rho^{*2}}}\right) I\{z_2 \in \mathcal{G}(z_1)\} \quad z_2 \in \mathbb{R}.$$

PROOF. To prove (i), we recall that $Z_2 | \{Z_1 = z_1\}$ is normal with mean ρ^*z_1 and variance $1 - \rho^{*2}$ (Johnson et al. 1994), and we write

$$\Pr\{Z_1 \leq z_1 | \mathbf{Z} \in \mathcal{G}\} = \frac{1}{\Pr\{\mathbf{Z} \in \mathcal{G}\}} \int_{-\infty}^{z_1} \int_{y \in \mathcal{G}(x)} \phi(x) \frac{1}{\sqrt{1 - \rho^{*2}}} \phi\left(\frac{y - \rho^*x}{\sqrt{1 - \rho^{*2}}}\right) dy dx. \quad (3)$$

Differentiate the above expression with respect to z_1 using the Leibniz rule (Bartle 1976, p. 245) to conclude that the density is

$$f_1(z_1) = \frac{1}{\Pr\{\mathbf{Z} \in \mathcal{G}\}} \phi(z_1) \Pr\{Z_2 \in \mathcal{T}(z_1, \rho^*)\}.$$

To prove (ii), we note that the density of the conditional random variable $Z_2 | \{Z_1 = z_1, \mathbf{Z} \in \mathcal{G}\}$ is

$$f_{z_1}(z_2) = \frac{f_{Z_2|Z_1=z_1}(z_2)}{\Pr\{\mathbf{Z} \in \mathcal{G} | Z_1 = z_1\}},$$

where $f_{Z_2|Z_1=z_1}(z_2)$ —the density function of the conditional random variable $Z_2 | \{Z_1 = z_1\}$ —is the normal density with mean ρ^*z_1 and variance $1 - \rho^{*2}$. \square

Algorithm C-NORTA exploits fact (a) and Proposition 2 to arrive at a very simple generation technique.

Algorithm 2 (C-NORTA)

Inputs: Marginal distributions F_1, F_2 ; correlation ρ ; feasible region \mathcal{F} .

Outputs: Random variate (x_1, x_2) .

0. Characterize \mathcal{F}_N from \mathcal{F} .
1. Select \mathcal{G} such that $\mathcal{G} \supset \mathcal{F}_N$.
2. Generate $z_1 \sim f_1(\cdot)$.
3. Generate $z_2 \sim f_{z_1}(\cdot)$.
4. Set $x_1 = F_1^{-1}(\Phi(z_1)), x_2 = F_2^{-1}(\Phi(z_2))$.
5. If $(x_1, x_2) \in \mathcal{F}$, then deliver (x_1, x_2) .
 Otherwise, go to Step 2.

The correctness of the C-NORTA algorithm is self-evident. The efficiency, however, depends critically on Steps 1, 2, and 3. For instance, the naïve adaptation of NORTA that was discussed in §1.2 is essentially C-NORTA with the outer set \mathcal{G} chosen to be \mathbb{R}^2 . This turns out to be very inefficient because, when \mathcal{F}_N has very small measure, choosing $\mathcal{G} = \mathbb{R}^2$ ensures that most random variates generated in Step 4 of the C-NORTA algorithm end up getting rejected in Step 5. (Why do we choose $\mathcal{G} \supset \mathcal{F}_N$ in Step 1 of C-NORTA? Choosing $\mathcal{G} \subset \mathcal{F}_N$ in Step 1 would lead to incorrect marginals owing to no generation in certain regions of \mathcal{F} having positive probability under the desired bivariate distribution. Choosing $\mathcal{G} = \mathcal{F}_N$, although ideal from an efficiency standpoint, may not lend itself to easy generation because the structure of the set \mathcal{F}_N is not readily available.)

Throughout the rest of this paper, our interest will be limited to *piecewise-linear* (PL) sets \mathcal{G} , i.e., sets \mathcal{G} that are obtained by outer-bounding \mathcal{F}_N using an appropriately chosen piecewise-linear curve. As we shall see, the reason for such stipulation is that PL sets, in addition to having the ability to outer-approximate \mathcal{F}_N to an arbitrary level of accuracy, afford great tractability in the generation Steps 2 and 3. Also, because the identification of ρ^* and the inversion in Step 4 are well-studied problems (Cario and Nelson 1997, 1998; Avramidis et al. 2009; Ghosh and Henderson 2003; Chen 2001), all our ensuing discussion is limited to Steps 1, 2, and 3. Accordingly, §§3 and 4 simply focus on the question of generating bivariate normal random vectors constrained to a given PL set \mathcal{G} , i.e., on Steps 2 and 3 in the C-NORTA

Downloaded from informs.org by [128.173.125.76] on 05 February 2014, at 07:19. For personal use only, all rights reserved.

algorithm. In §5, we discuss the choice of an appropriate PL set \mathcal{G} , i.e., Step 1 in the C-NORTA algorithm and its implications on Step 5.

3. C-NORTA on Normal Half-Spaces

In this section, we treat the special case where the required random vector \mathbf{X} is itself bivariate normal, and the PL set \mathcal{G} is a linear half-space. In addition to this special case being important in its own right, it serves to ease the exposition of the more general case. In what follows, we thus provide a generation technique for $\mathbf{N} | \mathcal{G}$ when \mathcal{G} is a linear half-space, and we establish a certain asymptotic efficiency property for the technique.

Define $\mathcal{G} = \{c^t z \geq v\}$. Assume the coefficients $c_1, c_2 > 0$; each of the other three cases are treated identically by the symmetry of the standard bivariate joint normal \mathbf{N} . Define $l(z_1) = \inf\{z_2: z_2 \in \mathcal{G}(z_1)\} = (v - c_1 z_1)/c_2$. Then, the set $\mathcal{G}(z_1) = \{z_2 \geq l(z_1)\}$. The marginal f_1 and conditional f_{z_1} densities are given by

$$\begin{aligned} f_1 &= \frac{1}{\bar{\Phi}(v)} \phi(z_1) \bar{\Phi}(l(z_1)), \\ f_{z_1} &= \frac{1}{\bar{\Phi}(l(z_1))} \phi(z_2) I\{z_2 \in \mathcal{G}(z_1)\}. \end{aligned} \quad (4)$$

The first term in each expression represents the normalizing constants for each density. The f_1 marginal is a mixture of normal tail distributions by a normal density.

Let $a = v/c_1$ represent the z_1 value at which the linear function $l(z_1) = 0$. We construct the following majorizing function $t_1(z_1)$ for $f_1(z_1)$:

$$t_1(z_1) = \begin{cases} \frac{1}{2} \frac{1}{\bar{\Phi}(v)} \phi(z_1) e^{-l^2(z_1)/2} & \text{if } z_1 < a, \\ \frac{1}{\bar{\Phi}(v)} \phi(z_1) & \text{if } z_1 \geq a. \end{cases} \quad (5)$$

That the function t_1 dominates the marginal f_1 on the left side $z_1 < a$ is evident since, for $y \in \mathbb{R}$,

$$\begin{aligned} \bar{\Phi}(y) &= 1 - \Phi(y) \leq 1 - \frac{1}{2}(1 + \sqrt{1 - e^{-y^2/2}}) \\ &\leq 1 - \frac{1}{2}(1 + (1 - e^{-y^2/2})) = \frac{1}{2} e^{-y^2/2}. \end{aligned} \quad (6)$$

The first inequality in (6) follows from the left side of the famous bound on the normal cumulative distribution function (cdf) (Johnson et al. 1994, p. 115), and the second inequality holds since $\sqrt{1 - e^{-y^2/2}} \geq 1 - e^{-y^2/2}$ for all $y \in \mathbb{R}$. That the function t_1 dominates the marginal f_1 on the right side $z_1 \geq a$ of (5) follows from using the loose upper bound 1 for the normal tail probability $\bar{\Phi}$.

On the left interval $z_1 < a$, the majorizing function $t_1(z_1)$ is proportional to a univariate normal density with mean vc_1 and variance c_2^2 . To see this, write

$$\begin{aligned} \phi(z_1) e^{-l^2(z_1)/2} &= \frac{1}{\sqrt{2\pi}} e^{-(1/(2c_2^2))(z_1^2 - 2vc_1 z_1 + v^2)} \\ &= \frac{1}{\sqrt{2\pi}} e^{-(1/(2c_2^2))(z_1 - vc_1)^2} e^{-\frac{1}{2}v^2}, \end{aligned} \quad (7)$$

where the first equality in (7) follows upon using $l(z_1) = (v - c_1 z_1)/c_2$ and $c_1^2 + c_2^2 = 1$. (Note that this density is different from the importance sampling change-of-measure optimal for estimating the probability of this tail set, which is the *standard* univariate normal density shifted to have mean vc_1 .) On the right interval $z_1 \geq a$, it can be seen that the majorizing function $t_1(z_1)$ is proportional to a standard normal density.

The cdf corresponding to the majorizing function $t_1(z_1)$ is then a mixture of two normal densities defined over disjoint half-intervals of the real line, and it is given by

$$\begin{aligned} R_1(z_1) &= \frac{w_1}{w_1 + w_2} R_{1,1}(z_1) \\ &\quad + \frac{w_2}{w_1 + w_2} R_{1,2}(z_1) \quad z_1 \in \mathbb{R}, \end{aligned} \quad (8)$$

where $R_{1,1} \sim w_1^{-1}(vc_1 + c_2 N)I\{z_1 < a\}$ and $R_{1,2} \sim w_2^{-1}NI\{z_1 \geq a\}$, and the constants w_1 and w_2 simplify to

$$\begin{aligned} w_1 &= \frac{1}{\bar{\Phi}(v)} \frac{1}{2\sqrt{2\pi}} \int_{-\infty}^a e^{-z_1^2/2} e^{-l^2(z_1)/2} dz_1 \\ &= \frac{c_2}{2\bar{\Phi}(v)} e^{-v^2/2} \Phi((v/c_1 - vc_1)/c_2); \\ w_2 &= \frac{1}{\bar{\Phi}(v)} \int_a^\infty \phi(z_1) dz_1 = \frac{1}{\bar{\Phi}(v)} \bar{\Phi}(a). \end{aligned} \quad (9)$$

The composition method (Law and Kelton 2000, p. 433) can be utilized to generate samples from R_1 . Generation from the densities f_1 and f_{z_1} thus follows the steps laid down in Algorithm 2a. (We note that Algorithm 2a is presented for the more general context of generating random vector \mathbf{Z} on the half-space $\mathcal{G} = \{c^t z \geq v\}$.)

Algorithm 2a (Generating from $\mathbf{Z} | \mathcal{G}$, where $\mathbf{Z} = \mathbf{Z}(\rho^*)$ and $\mathcal{G} = \{c^t z \geq v\}$)

Inputs: correlation ρ^* ; constant vector $c = (c_1, c_2)$; constant scalar v .

Outputs: Random variate (z_1, z_2) .

- (i) Initialize $C := \begin{bmatrix} 1 & \rho^* \\ 0 & \sqrt{1 - \rho^{*2}} \end{bmatrix}$, and redefine $c = (C^t)^{-1}c$.
- (ii) Generate $u_1, u_2 \sim U(0, 1)$ independently.
- (iii) If $u_1 < w_1/(w_1 + w_2)$, sample $z_1 \sim R_{1,1}$. Else, sample $z_1 \sim R_{1,2}$.

- (iv) If $u_2 t_1(z_1) > f_1(z_1)$, then go to Step (ii).
- (v) Let $l(z_1) = \frac{v - c_1 z_1}{c_2}$.
- (vi) Generate $u_3 \sim U(0, 1)$ independently.
- (vii) Set $z_2 = \Phi^{-1}(u_3(1 - \Phi(l(z_1))) + \Phi(l(z_1)))$.
- (viii) Return $z = C^t(z_1, z_2)^t$.

The initialization step calculates the linear transform to the standard bivariate normal $\mathbf{N} | \mathcal{G}$, and the final Step (viii) applies the reverse transformation. Step (iii) generates z_1 from the majorizer t_1 , and the variate is accepted as being from f_1 in (4) if the check of Step (iv) is satisfied. Step (vii) generates z_2 given z_1 via inversion of the conditional density in (4).

How efficient is C-NORTA in generating normal random vectors in a specified half-space \mathcal{G} ? Toward understanding this, Theorem 1 provides the asymptotic acceptance probability of C-NORTA on translated half-spaces \mathcal{G}^λ of the sort introduced for Proposition 1.

THEOREM 1. *Let \mathcal{G}^λ denote the translated half-space*

$$\mathcal{G}^\lambda = \{z \in \mathbb{R}^2: c^t z \geq v + \lambda c^t d\},$$

where $c = (c_1, c_2)$ is a constant vector, v is a constant scalar, $\lambda > 0$ is the rarity parameter, and d is a unit vector. If $c^t d > 0$, Algorithm 2 (C-NORTA) to sample from $\mathbf{Z} | \mathcal{G}$ obeys

$$\lim_{\lambda \rightarrow \infty} \lambda P(\lambda) = \frac{2}{c_2} \frac{1}{c^t d \sqrt{2\pi}}.$$

PROOF. It is sufficient if we establish the stated result for the standard bivariate normal \mathbf{N} on the set \mathcal{G}^λ (as opposed to \mathbf{Z} on \mathcal{G}^λ). To see this, we note first that \mathbf{N} and $A^{-1}\mathbf{Z}$ have the same distribution for $A = \begin{bmatrix} 1 & 0 \\ \rho^* & \sqrt{1-\rho^{*2}} \end{bmatrix}$. Consequently, the conditional random variables $\mathbf{Z} | \mathcal{G}$ and $A \times \mathbf{N} | A^{-1}\mathcal{G}$ have the same distribution, and the random vector $\mathbf{Z} | \mathcal{G}$ can be generated by first generating $\mathbf{N} | A^{-1}\mathcal{G}$ and then multiplying it by the matrix A . The probability of accepting a sample $P(\lambda)$ with such generation is then exactly the same as the probability of accepting a sample when generating $\mathbf{N} | A^{-1}\mathcal{G}$. This and the fact that $A^{-1}\mathcal{G}$ is a half-space if \mathcal{G} is a half-space assures us that it is sufficient to prove the stated result for standard bivariate normal vectors. (Also, note that the linear transformation A^{-1} ensures that this correspondence continues to hold for PL sets \mathcal{G} .)

We now prove the stated result for standard bivariate normal random vectors. The probability of accepting a sample in Step (iv) of Algorithm 2a is given by the inverse of the normalizing constant of the majorizer t_1 (Asmussen and Glynn 2007, Remark 2.5). Thus, we have

$$P(\lambda) = \frac{1}{(w_1 + w_2)}$$

$$= \frac{\bar{\Phi}(v_\lambda)}{(c_2/2)e^{-0.5v_\lambda^2}\Phi(\mu) + \bar{\Phi}(v_\lambda/c_1)},$$

$$v_\lambda = v + \lambda c^t d, \mu = (v_\lambda/c_1 - v_\lambda c_1)/c_2,$$

after substituting v_λ for v , μ for $l(z_1)$, and v_λ/c_1 for a , in (9). Using the bounds on $\bar{\Phi}(z)$, $z > 0$ introduced in (2), we can then write

$$\frac{(v_\lambda^2/(v_\lambda^2 + 1))(e^{-0.5v_\lambda^2}/v_\lambda)}{(e^{-0.5(v_\lambda/c_1)^2}/(v_\lambda/c_1)) + (c_2/2)\sqrt{2\pi}e^{-0.5v_\lambda^2}\Phi(\mu)} \leq P(\lambda)$$

$$\leq \frac{(e^{-0.5v_\lambda^2}/v_\lambda)}{(v_\lambda^2/(v_\lambda^2 + c_1^2))(e^{-0.5(v_\lambda/c_1)^2}/(v_\lambda/c_1)) + (c_2/2)\sqrt{2\pi}e^{-0.5v_\lambda^2}\Phi(\mu)},$$

$$\frac{(v_\lambda^2/(v_\lambda^2 + 1))}{c_1 e^{-0.5(v_\lambda c_2/c_1)^2} + v_\lambda (c_2/2)\sqrt{2\pi}\Phi(\mu)} \leq P(\lambda)$$

$$\leq \frac{1}{(v_\lambda^2 c_1/(v_\lambda^2 + c_1^2))e^{-0.5(v_\lambda c_2/c_1)^2} + v_\lambda (c_2/2)\sqrt{2\pi}\Phi(\mu)}.$$

Recalling that $v_\lambda = v + \lambda c^t d$, the result is now obtained by multiplying throughout with λ and taking the limit. \square

Theorem 1 asserts that C-NORTA's efficiency falls linearly with respect to λ . This compares very favorably with the exponential drop in efficiency of the naive sampling technique, rigorized through Proposition 1. Additionally, rejection in C-NORTA is limited to a univariate marginal.

4. C-NORTA on General Piecewise-Linear Sets

In this section, we detail Steps 2 and 3 in Algorithm 2 (C-NORTA) for PL sets \mathcal{G} . Recall that, for the purposes of this paper, PL sets are sets that can be characterized as lying above a piecewise-linear function. Specifically, a set $\mathcal{B} \subset \mathbb{R}^2$ is a PL set if there exists a piecewise-linear function $l: \mathbb{R} \rightarrow \mathbb{R}$ such that $\mathcal{B} = \{(z_1, z_2): z_2 \geq l(z_1)\}$.

We remind the reader that Step 1 of C-NORTA involves generating from the density $f_1(\cdot)$, which is a certain normal mixture of normal probabilities. Step 2 involves generating from $f_{z_1}(\cdot)$, which is a certain normal density that is truncated to the set $\mathcal{G}(z_1)$. As we shall see, both of these steps are accomplished through a rejection technique that remains efficient (in a certain precise sense) even as the set enclosed by \mathcal{G} becomes increasingly rare under the normal measure. Mainly for ease of exposition, we note two properties of the PL set \mathcal{G} without proof.

PROPERTY 1. *The set $\mathcal{G}(z_1) = \{z_2 \in \mathbb{R}: (z_1, z_2) \in \mathcal{G}\}$ is nonempty for all $z_1 \in \mathbb{R}$.*

PROPERTY 2. *The set $\mathcal{G}(z_1)$ is an interval of the form $[l(z_1), \infty)$.*

Our assumption of \mathcal{G} being PL precludes \mathcal{G} being bounded—something that we think is not a major stipulation. Problems with bounded \mathcal{G} sets, although easily addressed through a simple extension of the methods we present, are considered generally uninteresting from the standpoint of this paper. Property 2 essentially states that the set $\mathcal{G}(z_1)$ is an interval that lies “above” the chosen piecewise-linear boundary. This again turns out not to be a serious stipulation. For instance, if \mathcal{G} is chosen so that $\mathcal{G}(z_1)$ is a finite union of intervals for every z_1 (as opposed to an interval), much of what we say in this section can be easily adapted. Furthermore, for many problems where the set \mathcal{G} is unbounded but Property 2 is violated, a simple rotation of the coordinate axes will ensure that the resulting set \mathcal{G} becomes PL. Figure 1 illustrates typical \mathcal{G} and $\mathcal{G}(z_1)$ sets that are of interest.

4.1. Generating from $f_1(\cdot)$ and f_{z_1} in C-NORTA

As in §3, toward constructing a rejection technique for generation, choose the majorizing function of $f_1(\cdot)$ to be

$$t_1(z_1) = \begin{cases} 0.5(\Pr\{\mathbf{Z} \in \mathcal{G}\})^{-1} \phi(z_1) \exp\{-0.5(1 - \rho^*)^{-1} \cdot (l(z_1) - \rho^*z_1)^2\} & \text{if } l(z_1) > \rho^*z_1, \\ (\Pr\{\mathbf{Z} \in \mathcal{G}\})^{-1} \phi(z_1) & \text{if } l(z_1) \leq \rho^*z_1. \end{cases} \quad (10)$$

That the function $t_1(z_1)$ majorizes the function $f_1(z_1)$ follows from the derivation of the density function of the conditional random variable $Z_2 | Z_1$ and a certain bound on the normal tail probability. Proposition 3 asserts this formally.

PROPOSITION 3. *The function $t_1(z_1) \geq f_1(z_1)$ for all $z_1 \in \mathbb{R}$.*

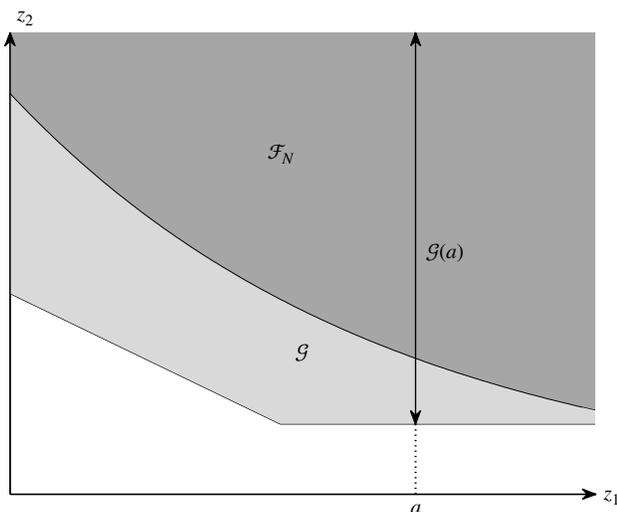


Figure 1 Illustration of Typical \mathcal{G} and $\mathcal{G}(z_1)$ Sets in C-NORTA
Notes. The region above the piecewise-linear boundary is the \mathcal{G} set. The set $\mathcal{G}(z_1)$ is the set of ordinates of the points in \mathcal{G} whose abscissa is z_1 . The set \mathcal{F}_N is the image in the normal space of the feasible region \mathcal{F} .

PROOF. Recall from Proposition 2 that

$$f_1(z_1) = \frac{1}{\Pr\{\mathbf{Z} \in \mathcal{G}\}} \phi(z_1) \Pr\{Z_2 \in \mathcal{T}(z_1, \rho^*)\}, \quad z_1 \in \mathbb{R},$$

where $\mathcal{T}(z_1, \rho^*) = \{t: t\sqrt{1 - \rho^*^2} + \rho^*z_1 \in \mathcal{G}(z_1)\}$. Since $\Pr\{Z_2 \in \mathcal{T}(z_1, \rho^*)\} \leq 1$, $t_1(z_1)$ clearly majorizes $f_1(z_1)$ when $l(z_1) \leq \rho^*z_1$. Next, note that since $l(z_1) = \inf\{z_2: (z_2, z_1) \in \mathcal{G}\}$, the set $\mathcal{T}(z_1, \rho^*)$ is the interval $[(l(z_1) - \rho^*z_1)/\sqrt{1 - \rho^*^2}, \infty)$. Therefore, the probability $\Pr\{Z_2 \in \mathcal{T}(z_1, \rho^*)\} = 1 - \Phi((l(z_1) - \rho^*z_1)/\sqrt{1 - \rho^*^2})$. Now use the inequality $1 - \Phi(z) \leq 0.5 \exp(-0.5z^2)$ for all $z > 0$ (as demonstrated through (6)) to conclude that $t_1(z_1)$ serves as a majorizer when $l(z_1) > \rho^*z_1$. \square

Again because of the equivalence (shown in the beginning of the proof of Theorem 1) of the general bivariate constrained normal $\mathbf{Z} | \mathcal{G}_Z$ and the constrained standard bivariate normal $\mathbf{N} | \mathcal{G}_N$ in the present context, we limit the rest of this section to the bivariate standard normal \mathbf{N} constrained by \mathcal{G} (i.e., simply assuming $\rho^* = 0$). Denote $\mathcal{A} = \{a_0 = -\infty, a_1, a_2, \dots, a_n, a_{n+1} = \infty\}$ as the ordered set that includes the locations where the function $l(\cdot)$ either changes slope or $l(z_1) = \rho^*z_1 = 0$ (for \mathbf{N} , $\rho^* = 0$). See Figure 2 to get a sense of the set \mathcal{A} . Because the piecewise-linear boundary defining \mathcal{G} has only a finite number of “pieces,” it is clear that each a_i , $i = 1, 2, \dots, n$ is finite. (If an entire segment of the piecewise-linear boundary of \mathcal{G} aligns with the z_1 axis, only the end points of the segment are included in \mathcal{A} .) Also denote the masses

$$w_i = \int_{a_{i-1}}^{a_i} t_1(z_1) dz_1, \quad \forall i = 1, 2, \dots, n + 1.$$

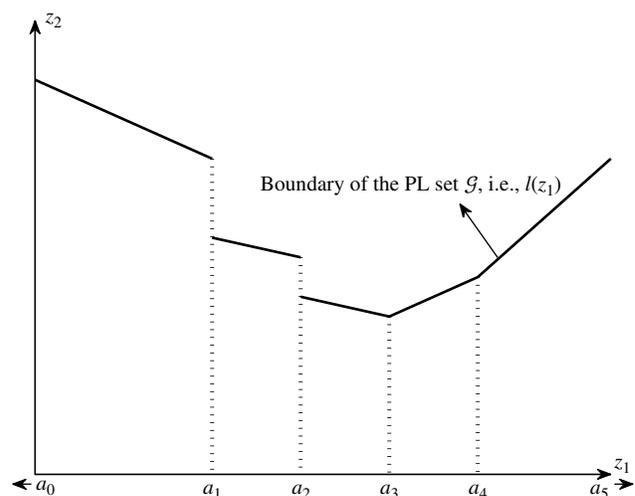


Figure 2 Illustration of the Piecewise-Linear Boundary and the $l(z_1)$ Function in C-NORTA

Notes. The region above the piecewise-linear boundary is the \mathcal{G} set. Note that the function $l(z_1)$ is also piecewise linear but can have jump discontinuities as shown.

The calculation of these areas w_1, w_2, \dots, w_{n+1} is made easy by the fact that the function $l(z_1)$ neither changes slope nor its sign in each of the above integrals. More specifically, if the line segment of the boundary of \mathcal{G} in (a_i, a_{i+1}) is defined by $c_{1i}z_1 + c_{2i}z_2 = v_i$, then $l(z_1)$ can be written as $l(z_1) = \sum_{i=0}^n (v_i/c_{2i} - c_{1i}z_1/c_{2i})I\{z_1 \in (a_i, a_{i+1})\}$ for $z_1 \in \mathbb{R} \setminus \mathcal{A}$. This implies, after some algebra, that in intervals (a_i, a_{i+1}) where $l(z_1) > 0$, the integrand $t_1(z_1)$ is proportional to a normal density with mean $c_{1i}v_i$ and variance c_{2i}^2 , and is proportional to a standard normal density otherwise. The computation of the areas $w_i, i = 1, 2, \dots, n+1$ hence amounts to the calculation of a known normal probability on a finite segment.

The above characterization allows us to write the cdf corresponding to the majorizing function $t_1(z_1)$ as a mixture of known normal cdfs, i.e.,

$$R_1(z_1) = \sum_{i=1}^{n+1} \frac{w_i}{w} R_{1,i}(z_1) \quad z_1 \in \mathbb{R}, \quad (11)$$

where $w = \sum_{i=1}^{n+1} w_i$, $R_{1,i}(z_1)$ is the cdf corresponding to a normal density with mean $c_{1i}v_i$ and variance c_{2i}^2 when $-c_{1i}z_1 + v_i > 0$ and restricted to $z_1 \in (a_{i-1}, a_i)$; otherwise, it is a standard normal cdf. This provides for easy generation using the composition method (Law and Kelton 2000, p. 433). The corresponding ideas for generating from the density $f_{z_1}(\cdot)$ in Step 3 of Algorithm 2 are similar but far easier. It can be seen from the expression in Property 2 that $f_{z_1}(\cdot)$ is a standard normal density truncated to the set $\mathcal{G}(z_1)$. Generation from the densities f_1 and f_{z_1} thus follow the steps laid down in Algorithm 3.

Algorithm 3 (Generating from $\mathbf{Z} | \mathcal{G}$, where $\mathbf{Z} = \mathbf{Z}(\rho^*)$ and \mathcal{G} is a PL set)

Inputs: correlation ρ^* ; PL set \mathcal{G} .

Outputs: Random variate (z_1, z_2) .

- (i) Initialize $C := \begin{bmatrix} 1 & \rho^* \\ 0 & \sqrt{1-\rho^{*2}} \end{bmatrix}$. Use C to define the set of points \mathcal{A} , piecewise-linear curve $l(z_1)$, and composite-distribution R_1 with parameters $\{w_i, i = 1, 2, \dots, n\}$ as described in §4.2. Set $w = \sum_{i=1}^{n+1} w_i$.
- (ii) Generate $u_1, u_2, u_3 \sim U(0, 1)$ independently.
- (iii) Use u_1 to select an index $i \in \{1, 2, \dots, n\}$ from the discrete mass function $\{w_i/w, i = 1, 2, \dots, n+1\}$.
- (iv) Set $z_1 = R_{1,i}^{-1}(R_{1,i}(a_i) + u_2(R_{1,i}(a_{i+1}) - R_{1,i}(a_i)))$.
- (v) If $u_3 t_1(z_1) > f_1(z_1)$, then go to Step (i).
- (vi) Let $l(z_1) = (v_i - c_{1i}z_1)/c_{2i}$.
- (vii) Sample $z_2 \sim \{N | z_2 \geq l(z_1)\}$.
- (viii) Return $z = C^t(z_1, z_2)^t$.

Step (iii) requires generating from a discrete distribution, and with some preprocessing, it is possible to do this in constant time (see, for example,

Walker 1977, Law and Kelton 2000, p. 472). Steps (iv) and (vii) generate from a univariate normal conditioned on a half-interval. Observe that the probability $\Pr\{\mathbf{N} \in \mathcal{G}\}$ need not be calculated explicitly for any of the steps in Algorithm 3, including Step (iii)'s discrete-mass sampling and rejection Step (v).

4.2. Preprocessing in C-NORTA

Recall that Steps 2 and 3 of C-NORTA (Algorithm 2) involve generating from the density functions $f_1(\cdot)$ and $f_{z_1}(\cdot)$, respectively. As we saw in §4.1, the density f_1 can be generated through a rejection technique with a majorizing function that is a mixture of two or more normal densities. Then, we can generate directly from the density $f_{z_1}(\cdot)$ after noting that it is a normal density that is truncated to an interval. The generation Steps 2 and 3 of C-NORTA are thus fairly straightforward and can use fast, well-established techniques for generating from normal densities.

The preprocessing steps in C-NORTA, however, are slightly more involved, and it is not immediately evident as to whether they will be efficient during implementation. Accordingly, we now detail all the operations that are required for preprocessing within C-NORTA.

Preprocessing Steps in C-NORTA

Step P.1. Identify the PL set \mathcal{G} such that $\mathcal{G} \supset \mathcal{F}_N = \{(z_1, z_2) : (\Psi_1(z_1), \Psi_2(z_2)) \in \mathcal{F}\}$, where \mathcal{F} is the feasible set, and Ψ is defined as in (1). Define the boundary of \mathcal{G} as the set of line segments $\{h_i^t z = v_i, i = 1, 2, \dots, m\}$ with intersection points $\{(\bar{z}_1^0, \bar{z}_2^0) = (-\infty, -\infty), (\bar{z}_1^1, \bar{z}_2^1), (\bar{z}_1^2, \bar{z}_2^2), \dots, (\bar{z}_1^m, \bar{z}_2^m), (\bar{z}_1^{m+1}, \bar{z}_2^{m+1}) = (\infty, \infty)\}$.

Step P.2. Identify corresponding PL set \mathcal{G} under \mathbf{N} . To do this, set $C := \begin{bmatrix} 1 & \rho^* \\ 0 & \sqrt{1-\rho^{*2}} \end{bmatrix}$, and to define line segments and intersection points, set $\{(C^t)^{-1}\bar{z}^i, (C^t)^{-1}h_i, v_i, i = 1, 2, \dots, m\}$.

Step P.3. Identify the function $l(z_1) = \inf\{z_2 : z_2 \in \mathcal{G}(z_1)\}$, where $\mathcal{G}(z_1) = \{z_2 : (z_1, z_2) \in \mathcal{G}\}$.

Step P.4. Identify set $\mathcal{A} = \{a_0 = -\infty, a_1, a_2, \dots, a_n, a_{n+1} = \infty\}$ —the ordered set where $l(z_1)$ either changes slope or $l(z_1) = 0$. These include the z_1 intercepts of intersection points.

Step P.5. Compute $w_i = \int_{a_{i-1}}^{a_i} t_1(z_1) dz_1$ for $i = 1, 2, \dots, n+1$, where $t_1(z_1)$ is the majorizing function given in (10).

Among preprocessing Steps P.1–P.5, Step P.1 is the least clear. Accordingly, we devote the ensuing §5 in its entirety to this step.

Suppose for now that the piecewise-linear set \mathcal{G} has been constructed. This implies that we have “constructed” the piecewise-linear function $l(z_1)$ as well. Upon such construction of the function $l(z_1)$, the set $\mathcal{A} = \{a_0 = -\infty, a_1, a_2, \dots, a_n, a_{n+1} = \infty\}$ (Step P.4) is simply the set of locations $z_1 \in \mathbb{R}$ where $l(z_1)$ attains zero and the set of points $z_1 \in \mathbb{R}$ where $l(z_1)$

changes slope. Both of these sets can be identified in a straightforward and efficient manner. Because the majorizing function $t_1(\cdot)$ involves only normal densities, Step P.5 is easily accomplished using readily available numerical quadrature. Note that for discrete mass sampling Step (ii) of Algorithm 3, it is sufficient to express all the w_i as multiples of the quantity $(\Pr\{\mathbf{N} \in \mathcal{G}\})^{-1}$.

4.3. Asymptotic Sampling Efficiency

Sections 1.2 and 3 established the asymptotic sampling efficiency of the naïve adaptation of NORTA and the C-NORTA methods for a region \mathcal{G} bounded by a single linear constraint. Specifically, we demonstrated that when normal random vectors are generated from the set $\mathcal{G}^\lambda = \{z: c^t z \geq v + \lambda c^t d\}$ ($\lambda > 0$, $c^t d > 0$), C-NORTA guarantees an acceptance probability that decays linearly. This is remarkably more efficient compared with the naïve adaptation of NORTA, where the corresponding acceptance probability decays exponentially. In this section, we prove a similar but slightly weaker efficiency guarantee for the C-NORTA method when generating from arbitrary PL sets \mathcal{G} . To ease exposition, we first establish this result for the specific case of generating normal random vectors on “acute” and “obtuse” cones (Proposition 4), and we then extend this to general random vectors on general PL sets (Theorem 2).

PROPOSITION 4. Let $\mathcal{H}_1 = \{r^t z \geq u + \lambda r^t d = u_\lambda\}$ and $\mathcal{H}_2 = \{s^t z \geq v + \lambda s^t d = v_\lambda\}$, where $d, r = (r_1, r_2), s = (s_1, s_2)$ are vectors satisfying $r^t d > 0, s^t d > 0$. Denote the intersection of the lines $r^t z = u_\lambda$ and $s^t z = v_\lambda$ as θ_λ .

(i) (Acute Cone). Let the translated cone $\mathcal{C}^\lambda = \mathcal{H}_1 \cap \mathcal{H}_2$; $r_1, r_2, s_2 > 0$, and $s_1 < 0$. Then the C-NORTA procedure to sample from $\mathbf{Z} | \mathcal{C}^\lambda$ has acceptance probability $P(\lambda)$ that obeys

$$\liminf_{\lambda \rightarrow \infty} \lambda^{-1} \exp\left\{-\frac{1}{2} \|\theta_\lambda\|^2\right\} \frac{P(\lambda)}{\Pr\{\mathbf{Z} \in \mathcal{C}^\lambda\}} = \kappa_1, \quad 0 < \kappa_1 < \infty.$$

(ii) (Obtuse Cone). Let the translated cone \mathcal{C}^λ be of the form $\mathcal{C}^\lambda = \mathcal{H}_1 \cap \mathcal{H}_2$ or $\mathcal{C}^\lambda = \mathcal{H}_1 \cup \mathcal{H}_2$. Also, assume $s_1, s_2, r_2 > 0$ and $r_1 < 0$. Then the C-NORTA procedure to sample from $\mathbf{Z} | \mathcal{C}^\lambda$ has acceptance probability $P(\lambda)$ that obeys

$$\liminf_{\lambda \rightarrow \infty} \exp\{-\eta \lambda^2\} \frac{P(\lambda)}{\Pr\{\mathbf{Z} \in \mathcal{C}^\lambda\}} = \kappa_2 > 0$$

for some $\eta > 0$.

PROOF. We again appeal to the equivalence of general bivariate and standard bivariate normal random vectors (shown in the beginning of the proof to Theorem 1) and establish this result only for the standard bivariate normal. See Figure 3 for a depiction of acute and obtuse cones.

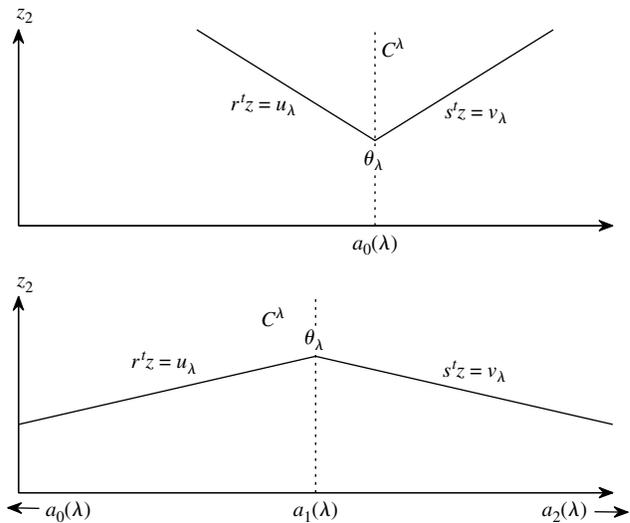


Figure 3 Illustration of Acute (Top) and Obtuse (Bottom) Cones Considered in Proposition 4

Notes. The shaded regions represent the translated cones \mathcal{C}^λ , where λ is the rarity or translation parameter. The direction of translation d is such that the probability measure assigned to the cones tends to zero as $\lambda \rightarrow \infty$; i.e., d satisfies $r^t d > 0, s^t d > 0$.

For proving (i), we first note that, for large-enough λ , the “breakpoint” set \mathcal{A} (the set where $l(z_1)$ either changes slope or is equal to zero, as illustrated in Figure 3) corresponding to the feasible set \mathcal{G}^λ consists of exactly one point θ_λ —the intersection of the lines $r^t z = u_\lambda$ and $s^t z = v_\lambda$. Denote the abscissa of θ_λ as $a_0(\lambda)$. Then, following the proof of Theorem 1, we see that the acceptance probability $P(\lambda) = 1/(w_1 + w_2)$, where

$$w_1 = \int_{-\infty}^{a_0(\lambda)} t_1(z_1) dz, \quad w_2 = \int_{a_0(\lambda)}^{\infty} t_1(z_1) dz,$$

and the majorizing function $t_1(z_1)$ is as given by expression (10). Specifically, because the cone remains above the z_1 axis for large-enough λ , and because we have argued that it is sufficient to consider the standard bivariate case, we see that

$$\begin{aligned} w_1 &= \frac{1}{2 \Pr\{\mathbf{N} \in \mathcal{C}^\lambda\}} \int_{-\infty}^{a_0(\lambda)} \phi(z_1) \exp\{-0.5l^2(z_1)\} dz_1 \\ &= \frac{r_2}{2 \Pr\{\mathbf{N} \in \mathcal{C}^\lambda\}} \exp\left\{-\frac{1}{2} u_\lambda^2\right\} \Phi\left(\frac{a_0(\lambda) - u_\lambda r_1}{r_2}\right), \end{aligned} \quad (12)$$

where $r = (r_1, r_2)$. Similarly, we have

$$w_2 = \frac{s_2}{2 \Pr\{\mathbf{N} \in \mathcal{C}^\lambda\}} \exp\left\{-\frac{1}{2} v_\lambda^2\right\} \bar{\Phi}\left(\frac{a_0(\lambda) - v_\lambda s_1}{s_2}\right). \quad (13)$$

The statement of Proposition 4 assumes (without loss of generality) that $s_2, r_2, r_1 > 0$ and $s_1 < 0$. Then, noting that $a_0(\lambda) = (s_2 u_\lambda - r_2 v_\lambda) / (s_2 r_1 - r_2 s_1)$, it is seen with some algebra that $(a_0(\lambda) - u_\lambda r_1) / r_2 \leq 0$. This implies that $\Phi((a_0(\lambda) - u_\lambda r_1) / r_2) \rightarrow 0$ as $\lambda \rightarrow \infty$. Furthermore, using the Gaussian tail bound given in (2),

we can write

$$\begin{aligned} &\Phi((a_0(\lambda) - u_\lambda r_1)/r_2) \\ &= \bar{\Phi}((-a_0(\lambda) + u_\lambda r_1)/r_2) \\ &= O(\lambda^{-1} \exp\{-\frac{1}{2}(a_0(\lambda) - u_\lambda r_1)^2/r_2^2\}). \end{aligned} \quad (14)$$

Using (12) and (14), and noticing that

$$u_\lambda^2 + \frac{(a_0(\lambda) - u_\lambda r_1)^2}{r_2^2} = a_0^2(\lambda) + \frac{(u_\lambda - a_0(\lambda)r_1)^2}{r_2^2} = \|\theta_\lambda\|^2,$$

we see that

$$w_1 \Pr\{\mathbf{N} \in \mathcal{C}^\lambda\} = O(\lambda^{-1} \exp\{-\frac{1}{2}\|\theta_\lambda\|^2\}). \quad (15)$$

Using analogous arguments, we also arrive at

$$w_2 \Pr\{\mathbf{N} \in \mathcal{C}^\lambda\} = O(\lambda^{-1} \exp\{-\frac{1}{2}\|\theta_\lambda\|^2\}). \quad (16)$$

Use (15) and (16) to conclude that the assertion in (i) holds.

In proving assertion (ii), if the translated cone is of the form $\mathcal{C}^\lambda = \mathcal{H}_1 \cap \mathcal{H}_2$, then the corresponding breakpoint set \mathcal{A} will again contain only the vertex θ_λ of the cone \mathcal{C}^λ . If the translated cone is of the form $\mathcal{C}^\lambda = \mathcal{H}_1 \cup \mathcal{H}_2$, then the corresponding breakpoint set $\mathcal{A} = \{a_0(\lambda), a_1(\lambda), a_2(\lambda)\}$, where $a_0(\lambda)$ and $a_2(\lambda)$ are the locations where the lines $r^t z = u_\lambda$ and $s^t z = v_\lambda$ cross the z_1 axis, respectively, and $a_1(\lambda)$ is the abscissa of their intersection (see Figure 3). Specifically, $a_0(\lambda) = u_\lambda/r_1$, $a_2(\lambda) = v_\lambda/s_1$, and $a_1(\lambda) = (s_2 u_\lambda - r_2 v_\lambda)/(s_2 r_1 - r_2 s_1)$. In what follows, we will prove assertion (ii) only for the latter context. The proof for the former context follows in a very similar fashion, and we omit it.

Suppose that the translated cone is of the form $\mathcal{C}^\lambda = \mathcal{H}_1 \cup \mathcal{H}_2$. Then corresponding to the breakpoint set $\mathcal{A} = \{a_0(\lambda), a_1(\lambda), a_2(\lambda)\}$, we have the four integrals $w_1 = \int_{-\infty}^{a_0(\lambda)} t_1(z_1) dz$, $w_2 = \int_{a_0(\lambda)}^{a_1(\lambda)} t_1(z_1) dz$, $w_3 = \int_{a_1(\lambda)}^{a_2(\lambda)} t_1(z_1) dz$, and $w_4 = \int_{a_2(\lambda)}^{\infty} t_1(z_1) dz$. Noting that $P(\lambda) = 1/(w_1 + w_2 + w_3 + w_4)$, and following the approach leading to expressions (12) and (13), we have

$$\begin{aligned} w_1 &= \frac{r_2}{2 \Pr\{\mathbf{N} \in \mathcal{C}^\lambda\}} \Phi\left(\frac{a_0(\lambda) - u_\lambda r_1}{r_2}\right); \\ w_2 &= \frac{r_2}{2 \Pr\{\mathbf{N} \in \mathcal{C}^\lambda\}} \exp\left\{-\frac{1}{2}u_\lambda^2\right\} \\ &\quad \cdot \left(\Phi\left(\frac{a_1(\lambda) - u_\lambda r_1}{r_2}\right) - \Phi\left(\frac{a_0(\lambda) - u_\lambda r_1}{r_2}\right)\right); \\ w_3 &= \frac{s_2}{2 \Pr\{\mathbf{N} \in \mathcal{C}^\lambda\}} \exp\left\{-\frac{1}{2}v_\lambda^2\right\} \\ &\quad \cdot \left(\Phi\left(\frac{a_2(\lambda) - v_\lambda s_1}{s_2}\right) - \Phi\left(\frac{a_1(\lambda) - v_\lambda s_1}{s_2}\right)\right); \\ w_4 &= \frac{s_2}{2 \Pr\{\mathbf{N} \in \mathcal{C}^\lambda\}} \bar{\Phi}\left(\frac{a_2(\lambda) - v_\lambda s_1}{s_2}\right). \end{aligned} \quad (17)$$

Since $r_2, s_1, s_2 > 0$ and $r_1 < 0$, we know that $a_0(\lambda) \rightarrow -\infty$ as $\lambda \rightarrow \infty$. Again appealing to the Gaussian tail bound given in (2), we see that $w_1 = O(\lambda^{-1} \exp\{-\frac{1}{2}((a_0(\lambda) - u_\lambda r_1)^2/r_2^2)\})$. Similarly, we see that $a_2(\lambda) \rightarrow \infty$ as $\lambda \rightarrow \infty$ and $w_4 = O(\lambda^{-1} \exp\{-\frac{1}{2}(a_2(\lambda) - v_\lambda s_1)^2/s_2^2\})$. Using these, and the expressions for w_3 and w_4 appearing in (17), we conclude that the assertion in (ii) holds. \square

Three observations about Proposition 4 are noteworthy. First, the term $\Pr\{\mathbf{N} \in \mathcal{C}^\lambda\}$ appearing in the assertions of Proposition 4 has the interpretation of the acceptance probability of the naïve adaptation of NORTA method. Hence, the assertions of Proposition 4 imply that, under all reasonable conditions, i.e., the translated feasible region is an obtuse or an acute cone that progressively becomes “rarer,” the acceptance probability in C-NORTA can be expected to be exponentially larger (asymptotically) than that of naïve NORTA. Second, when the translated feasible region is an acute cone, the rate at which the limiting ratio of the acceptance probabilities tends to infinity can be explicitly calculated. Interestingly, this rate is dependent only on the distance of the vertex of the acute cone from the origin. The corresponding rate for the obtuse cone does not seem to be explicitly calculable. This is why we allow the constant κ_2 that appears in assertion (ii) to attain ∞ . Third, unlike Theorem 1, the assertions in Proposition 4 say nothing about the absolute performance of C-NORTA. They are weaker in the sense that they only say how much better C-NORTA can be expected to perform in comparison with a naïve adaptation of NORTA.

Recall that Proposition 4 is for the specific case of feasible sets that are cones and where the underlying random variables are Gaussian. We will see, however, that Proposition 4 extends somewhat seamlessly to general PL sets \mathcal{G} and to general NORTA vectors \mathbf{X} . Toward this, define the *characteristic cone* \mathcal{C} of the PL set \mathcal{G} as the cone formed by the two end segments of the boundary of the set \mathcal{G} . Specifically, if the two semi-infinite end segments of the PL set \mathcal{G} lie on the lines $r^t z = u$ and $s^t z = v$, define the characteristic cone \mathcal{C} of \mathcal{G} as $\mathcal{C} = \{z: r^t z \geq u, s^t z \geq v\}$. As usual, the translated characteristic cone $\mathcal{C}^\lambda = \{z: r^t z \geq u + \lambda r^t d, s^t z \geq v + \lambda s^t d\}$, where the unit vector d satisfies $r^t d > 0$ and $s^t d > 0$. Accordingly, the translated PL set \mathcal{G}^λ is obtained by simply moving each of the segments comprising the boundary of \mathcal{G} in the direction d and by an amount λ . It is also clear that the characteristic cone of the translated set \mathcal{G}^λ is simply the translated cone \mathcal{C}^λ .

With the above notation in place, we are now ready to generalize Proposition 4. Part (i) of Theorem 2 establishes the efficiency result for normal random vectors on general PL sets \mathcal{G} . This is extended to more

general random vectors in part (ii) under a further mild condition.

THEOREM 2. (i) Let \mathcal{G}^λ be a translated PL set with translated characteristic cone

$$\mathcal{C}^\lambda = \{z: r^t z \geq u + \lambda r^t d, s^t z \geq v + \lambda s^t d\},$$

$$r^t d > 0, s^t d > 0. \quad (18)$$

Then, the C-NORTA procedure to sample from $\mathbf{Z} | \mathcal{G}^\lambda$ has acceptance probability $P(\lambda)$ that obeys

$$\liminf_{\lambda \rightarrow \infty} \exp\{-\eta \lambda^2\} \frac{P(\lambda)}{\Pr\{\mathbf{Z} \in \mathcal{G}^\lambda\}} = \kappa > 0$$

for some $\eta > 0$.

(ii) Let $\mathbf{X} = (X_1, X_2)$ be the NORTA vector, i.e., $X_i = F_i^{-1}(\Phi(Z_i(\rho^*)))$, $i = 1, 2$, where F_i , $i = 1, 2$ are continuous strictly increasing distributions, and $\mathbf{Z}(\rho^*) = (Z_1(\rho^*), Z_2(\rho^*))$ is a bivariate normal random vector with standard normal marginals and correlation $\text{Corr}(Z_1(\rho^*), Z_2(\rho^*)) = \rho^*$. Denote $\mathbf{X} | \mathcal{F}^\lambda$ as the vector \mathbf{X} constrained to the translated half-space $\mathcal{F}^\lambda = \{z: c^t z \geq v + \lambda c^t d\}$, where d satisfies $c^t d > 0$. Denote the image of \mathcal{F}^λ under the bivariate normal $\mathbf{Z}(\rho^*)$ as \mathcal{F}_N^λ . If the chosen PL set \mathcal{G}^λ with translated characteristic cone \mathcal{C}^λ (defined in (18)) satisfies

$$\liminf_{\lambda \rightarrow \infty} \frac{\Pr\{\mathbf{Z} \in \mathcal{F}_N^\lambda\}}{\Pr\{\mathbf{Z} \in \mathcal{G}^\lambda\}} > 0, \quad (19)$$

then the C-NORTA procedure to sample from $\mathbf{X} | \mathcal{F}^\lambda$ has acceptance probability $P(\lambda)$ that obeys

$$\liminf_{\lambda \rightarrow \infty} \exp\{-\eta \lambda^2\} \frac{P(\lambda)}{\Pr\{\mathbf{X} \in \mathcal{F}^\lambda\}} = \kappa > 0$$

for some $\eta > 0$.

PROOF. For proving (i), note that for any direction d , the PL set \mathcal{G}^λ can be inner- and outer-bounded by appropriate translations of the cone \mathcal{C}^λ along d . Specifically, if $\Delta\lambda_i$ and $\Delta\lambda_o$ are the respective translations, then we have $\mathcal{C}^{\lambda+\Delta\lambda_i} \subseteq \mathcal{G}^\lambda \subseteq \mathcal{C}^{\lambda+\Delta\lambda_o}$. Thus, $P(\lambda) = \Pr\{\mathbf{N} \in \mathcal{G}^\lambda\}$ is bounded on either side by quantities that, as per Proposition 4, possess the desired property.

Part (ii) follows upon noticing that

$$\begin{aligned} \frac{P(\lambda)}{\Pr\{\mathbf{X} \in \mathcal{F}^\lambda\}} &= \frac{P(\lambda)}{\Pr\{\mathbf{Z}(\rho^*) \in \mathcal{F}_N^\lambda\}} \\ &= \frac{P(\lambda)}{\Pr\{\mathbf{Z}(\rho^*) \in \mathcal{G}^\lambda\}} \frac{\Pr\{\mathbf{Z}(\rho^*) \in \mathcal{G}^\lambda\}}{\Pr\{\mathbf{Z}(\rho^*) \in \mathcal{F}_N^\lambda\}} \end{aligned} \quad (20)$$

and then using the assertion in (i) and the assumption in (19) to the two fractions appearing on the right-hand side of (20). \square

For convenience, we have chosen to state only a weaker and more general assertion within Theorem 2. For the specific case of PL sets with an acute characteristic cone, a more specific result of the form appearing as assertion (i) in Proposition 4 is clearly possible. Also, for the context of general random vectors, Theorem 2 assumes that the assumption in (19) holds; i.e., the chosen PL set \mathcal{G}^λ approximates the set \mathcal{F}^λ well. It so happens that making such a choice is reasonably easy. We say more on this in the subsequent section, where we discuss the choice of PL sets \mathcal{G} within C-NORTA algorithms.

5. Implementation

In this section, we discuss C-NORTA's implementation. Specifically, we provide further insight leading to detail on the mechanics of construction of a PL set \mathcal{G} during implementation. We also demonstrate the process on three illustrative examples and present some numerical evidence. We remind the reader that throughout this section, for ease of exposition, we switch to the notation $\{c_1 x_1 + x_2 \geq v\}$ when representing linear constraints.

5.1. Constructing the PL Set

Consider constraint sets of form $\mathcal{F} = \{x | c_1 x_1 + x_2 \geq v\}$ or $\mathcal{F} = \{x | c_1 x_1 + x_2 \leq v\}$. The NORTA image of this set in the bivariate normal space is $\mathcal{F}_N = \{(z_1, z_2) \in \mathbb{R}^2 | (\Psi_1(z_1), \Psi_2(z_2)) \in \mathcal{F}\}$, where the vector transform $\Psi = (\Psi_1, \Psi_2)$ is as defined in (1). Assume that the marginal distributions F_i are continuous and strictly increasing. This condition is mild; for instance, it covers any distribution that has a density f_i with respect to the Lebesgue measure over the support \mathcal{X}_i that is nonzero over the entire set. The assumption implies the continuity and strict monotonicity of the inverses F_i^{-1} and, consequently, that each component of the map Ψ is one-to-one and maps \mathcal{X}_i onto \mathbb{R}^2 . The set \mathcal{F} then has a one-to-one onto analogue $\mathcal{F}_N \in \mathbb{R}^2$ under the normal measure, and samples of NORTA vector \mathbf{X} from within \mathcal{F} are obtained by transforming (via Ψ) samples of \mathbf{Z} from \mathcal{F}_N . The set $\mathcal{F}_N = \{c_1 \Psi_1(z_1) + \Psi_2(z_2) \geq v\}$, and the boundary of this set is determined by the doubles (z_1, z_2) that satisfy the equality. Denote by $\partial\mathcal{F}_N(z_1)$ the set of z_2 that satisfies the boundary condition

$$\partial\mathcal{F}_N(z_1) = \{z_2 | c_1 \Psi_1(z_1) + \Psi_2(z_2) = v\}. \quad (21)$$

Assume $c_1 \neq 0$ to avoid trivialities. Then, we have the following result that is crucial to constructing PL sets.

PROPOSITION 5. *If each marginal distribution F_i has a nonzero density f_i with respect to (w.r.t.) the Lebesgue measure over its support, then the boundary set $\partial\mathcal{F}_N(z_1)$ defines a monotone function in z_1 , where the direction of monotonicity is determined by the sign of c_1 .*

PROOF. For a fixed x_1 , the set $\partial\mathcal{F}(x_1) = \{x_2 \mid c_1x_1 + x_2 = v\}$ is a singleton as per our assumptions. The preceding discussion establishes that Ψ is one-to-one onto. Thus, for a fixed z_1 , the set $\partial\mathcal{F}_N(z_1)$ is a singleton, and $\partial\mathcal{F}_N(z_1)$ defines a function in z_1 . Consider the \mathbb{R}^2 curve $\{z_1, \mathcal{G}(z_1)\}$ that is defined by the relation $\Psi_2(z_2) = v - c_1\Psi_1(z_1)$. Taking derivatives w.r.t. z_1 on either side, we get

$$\begin{aligned}\dot{\Psi}_2(z_2)\dot{z}_2 &= -c_1\dot{\Psi}_1(z_1), \\ \dot{z}_2 &= -c_1\frac{\phi(z_1)}{\phi(z_2)}\frac{f_2(F_2^{-1}(\Phi(z_2)))}{f_1(F_1^{-1}(\Phi(z_1)))}. \quad (22)\end{aligned}$$

The latter is obtained by applying the chain rule over the composition map $\Psi_i = F_i^{-1}(\Phi(\cdot))$ and using $\dot{F}_i^{-1}(y) = \dot{y}/f_i(F_i^{-1}(y))$. Equation (22) gives the derivative of z_2 w.r.t. z_1 , where all the z_1 and z_2 terms are positive. Thus, the curve is strictly monotone, the direction being determined by c_1 . \square

Proposition 5 ensures that, for any linear-bounded \mathcal{F} set in \mathcal{X} , the set \mathcal{F}_N is bounded by a monotone function. This allows the construction of a PL set \mathcal{G} that outer-bounds the set \mathcal{F}_N using only a finite number of points on the boundary $\partial\mathcal{F}_N(z_1)$ of the set \mathcal{F}_N . To see this, we note from Proposition 5 that the boundary $\partial\mathcal{F}_N(z_1)$ can be expressed as the monotone function

$$\begin{aligned}\partial\mathcal{F}_N(z_1) &= \Psi_2^{-1}(v - c_1\Psi_1(z_1)) \\ &= \Phi^{-1}(E_2(v - c_1F_1^{-1}(\Phi_1(z_1)))).\end{aligned}$$

Suppose we observe this boundary at k points $z_1 = z_1^1, z_1 = z_1^2, \dots, z_1 = z_1^k$; i.e., we calculate $\partial\mathcal{F}_N(z_1^1), \partial\mathcal{F}_N(z_1^2), \dots, \partial\mathcal{F}_N(z_1^k)$. Because the function $\partial\mathcal{F}_N(z_1)$ is known to be monotone increasing ($c_1 > 0$) or monotone decreasing ($c_1 < 0$), an appropriate step function $l(z_1)$ can be constructed using the observed doubles $(z_1^1, \partial\mathcal{F}_N(z_1^1)), (z_1^2, \partial\mathcal{F}_N(z_1^2)), \dots, (z_1^k, \partial\mathcal{F}_N(z_1^k))$. The construction ensures that the PL set formed with $l(z_1)$ as the boundary outer-approximates the set \mathcal{F}_N , as shown in Figure 4. We now formally list this construction in the form of an algorithm. (The listing is provided only for the case $c_1 > 0$. The corresponding construction for $c_1 < 0$ is analogous and straightforward.)

Algorithm 4 (Constructing the PL boundary)

Inputs: constants $c_1 > 0, v$.

Outputs: PL boundary function $l(z_1)$.

- (i) Select k points $z_1^1, z_1^2, \dots, z_1^k$ (in ascending order) and compute $\partial\mathcal{F}_N(z_1^1), \partial\mathcal{F}_N(z_1^2), \dots, \partial\mathcal{F}_N(z_1^k)$.
- (ii) If $\mathcal{F} = \{x \mid c_1x_1 + x_2 \geq v\}$ go to Step (iii). Otherwise, go to Step (iv).
- (iii) Set

$$l(z_1) = \begin{cases} -\infty, & z_1 < z_1^1, \\ \partial\mathcal{F}_N(z_1^i), & z_1^i \geq z_1 < z_1^{i+1}, \quad i=1, 2, \dots, k-1, \\ \partial\mathcal{F}_N(z_1^k), & z_1 \geq z_1^k. \end{cases}$$

(iv) Set

$$(z_1) = \begin{cases} \partial\mathcal{F}_N(z_1^1), & z_1 < z_1^1, \\ \partial\mathcal{F}_N(z_1^{i+1}), & z_1^i \geq z_1 < z_1^{i+1}, \quad i=1, 2, \dots, k-1, \\ \infty, & z_1 \geq z_1^k. \end{cases}$$

Algorithm 4 explicitly writes the form of the step function $l(z_1)$ depending on the sign of c_1 (which decides the direction of monotonicity of the function $\partial\mathcal{F}_N(z_1)$) and the side of the constraint constituting the feasible region. Accordingly, the resulting PL set \mathcal{G} will be either $\mathcal{G} = \{(z_1, z_2) \in \mathbb{R}^2 \mid z_2 \geq l(z_1)\}$ or $\mathcal{G} = \{(z_1, z_2) \in \mathbb{R}^2 \mid z_2 \leq l(z_1)\}$.

Algorithm 4 is tractable and affords an arbitrary level of accuracy (in principle) depending on the number of “steps” used. As stated, however, it does not answer the question of how many and which points $z_1^1, z_1^2, \dots, z_1^k$ should be chosen to best approximate $\partial\mathcal{F}_N(z_1)$. This is an important question that lies outside the current scope of this paper. One redeeming factor is that the PL set \mathcal{G} formed by even the simplest of such step functions—a step function with a single step (see Figure 4)—satisfies the conditions stipulated by Theorem 2 and thereby ensures the slow linear drop-off of the acceptance probability $P(\lambda)$.

To further illustrate the construction proposed in Algorithm 4, in Figure 4 we demonstrate its use through three prototypical examples, involving various distributions and types of constraints. For instance, the first row of plots in Figure 4 involves the constraint $2x_1 + x_2 \geq 10$, with the two marginal random variables being exponentially distributed with parameter 1. The second row in Figure 4 involves the constraint $-2x_1 + x_2 \geq 10$, with one of the two marginal random variables having a gamma distribution with parameters (2, 2) and the other a Burr Type XII distribution with parameters (1, 1), respectively. Likewise, the third row in Figure 4 involves the constraint $-x_1 + x_2 \leq -4$, with one of the two marginal random variables having a gamma distribution with parameters (2, 2) and the other a beta distribution with parameters (2, 2), respectively. In each case, the step function constitutes the function constructed using Algorithm 4.

If additional information is available on the structure of the function $\partial\mathcal{F}_N(z_1)$, we could potentially construct a PL boundary $l(z_1)$ that approximates $\partial\mathcal{F}_N(z_1)$ even better than the step function proposed in Algorithm 4. For instance, if it is known that the function $\partial\mathcal{F}_N(z_1)$ is concave or convex (in addition to being monotone), a piecewise-linear approximation obtained by simply connecting the observed

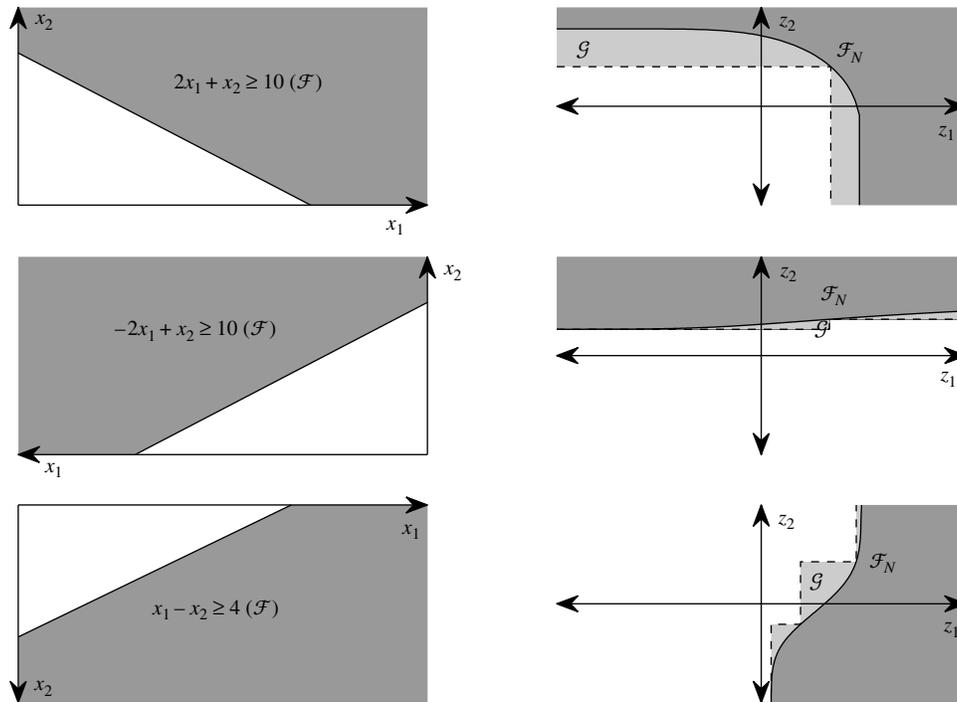


Figure 4 Illustration of the PL Set \mathcal{G} Constructed Using Step Functions (Algorithm 4) for Three Different Distribution Pairs and Feasible Regions

Notes. The plots in the left column depict various feasible regions \mathcal{F} . The plots in the right column depict the corresponding images \mathcal{F}_N in the normal space, with the (dashed) step function being the constructed PL set boundary. The first row corresponds to two exponentials each with parameter 1. The second row corresponds to gamma and Burr Type XII marginal distributions with the respective parameters (2, 2) and (1, 1). The third row corresponds to gamma and beta marginal distributions with the respective parameters (2, 2) and (2, 2).

doubles $(z_1^1, \partial \mathcal{F}_N(z_1^1)), (z_1^2, \partial \mathcal{F}_N(z_1^2)), \dots, (z_1^k, \partial \mathcal{F}_N(z_1^k))$ would better approximate $\partial \mathcal{F}_N(z_1)$. Such convexity or concavity of $\partial \mathcal{F}_N(z_1)$ can be shown for a variety of distribution pairs. Proposition 6, for instance, demonstrates this for the case where both marginal distributions are exponential.

PROPOSITION 6. Let F_1 and F_2 be exponential cdfs with means $\mu_1 = \mu$ and $\mu_2 = 1$, respectively. The boundary set $\partial \mathcal{F}_N(z_1)$ in (23) defines a concave function in z_1 when $c_1, v > 0$.

PROOF. Since F_1 and F_2 are exponentials, $X_i = \Psi_i(Z_i) = -\mu_i \ln(\bar{\Phi}(Z_i))$ with $\mu_1 = \mu$ and $\mu_2 = 1$. This leads to the curve

$$\begin{aligned} \partial \mathcal{G}(z_1) &\triangleq \{z_2 \mid c_1 \Psi_1(z_1) + \Psi_2(z_2) = v\} \\ &= \{z_2 \mid -c_1 \mu \ln(\bar{\Phi}(z_1)) - \ln(\bar{\Phi}(z_2)) = v\} \\ &= \{z_2 \mid \bar{\Phi}(z_2) = e^{-v} (\bar{\Phi}(z_1))^{-c_1 \mu}\}. \end{aligned} \quad (23)$$

The support set of the NORTA vector is the positive quadrant in \mathbb{R}^2 ; so to avoid trivial \mathcal{F} sets, allow the z_2 axis intercept b of the line $c_1 z_1 + z_2 = v$ to take only positive values. Proceeding as in the proof of Proposition 5, differentiate w.r.t. z_1 on either side of (23) to get

$$\begin{aligned} \phi(z_2) \dot{z}_2 &= -c_1 \mu e^{-v} \phi(z_1) (\bar{\Phi}(z_1))^{-c_1 \mu - 1}, \quad \text{or} \\ \dot{z}_2 &= -c_1 \mu \frac{\phi(z_1) \bar{\Phi}(z_2)}{\bar{\Phi}(z_1) \phi(z_2)}. \end{aligned}$$

The last equation uses the definition of the boundary (23). Thus, given the assumption $c_1 > 0$, the derivative \dot{z}_2 is a nonpositive function, and the curve $\partial \mathcal{F}_N(z_1)$ is decreasing.

Call $H(z)$ the hazard-rate function $\phi(z)/\bar{\Phi}(z)$ of the standard univariate normal. The function $H(z)$ is non-decreasing in z . To see why, differentiate to get

$$\dot{H}(z) = \frac{\phi(z)}{\bar{\Phi}(z)} \left(\frac{\phi(z)}{\bar{\Phi}(z)} - z \right),$$

where the term in the parentheses is nonnegative because of the right inequality in (2). Now, $\dot{z}_2 = (-c/\mu)(H(z_1)/H(z_2))$. Consider $z'_1 > z_1$. Then, $z'_2 \leq z_2$ and $H(z_1)/H(z_2) \leq H(z'_1)/H(z'_2)$. We then have that $\dot{z}_2(z_1) \geq \dot{z}_2(z'_1)$, and thus \dot{z}_2 is a nonincreasing function of z_1 , or that the second derivative is nonpositive. This gives us the result. \square

5.2. Numerical Illustration

We now illustrate the efficiency of the approach presented in this paper through a simple numerical example. Suppose we wish to generate a bivariate random vector (X_1, X_2) such that $X_1 \sim \text{Exponential}(1)$, $X_2 \sim \text{Exponential}(1)$, $\text{Corr}(X_1, X_2) = 0$, and $X_1 + X_2 \geq \lambda$ for some known $\lambda > 0$.

The acceptance probability when performing such generation using a naive adaptation of NORTA can

be calculated rather simply. Because the stipulated correlation between X_1 and X_2 is zero, NORTA will generate X_1 and X_2 independently and then accept the resulting vector (X_1, X_2) if $X_1 + X_2 \geq \lambda$. Thus, the acceptance probability in such a case turns out to be

$$P(\lambda) = 1 - \int_0^\lambda \int_0^{\lambda-x_2} \exp(-x_1) \exp(-x_2) dx_1 dx_2 = (1 + \lambda) \exp(-\lambda).$$

As can be seen from Figure 5, it is clear that such generation becomes inefficient very rapidly. (For example, if $\lambda = 7$, $P(\lambda) \approx 0.007$.)

5.2.1. C-NORTA vs. Naïve NORTA. Now consider using C-NORTA for the example. Following the recommendation given after Proposition 5, we choose a PL set with a step function boundary as depicted in Figure 4. For convenience, we choose the vertex of the boundary to be the intersection of the 45° line passing through the origin and the step function boundary. In other words, we choose the vertex (z_1^*, z_2^*) of the boundary to be the point on the boundary satisfying $z_1^* = z_2^*$. After some simple algebra, this gives $z_1^* = z_2^* = z^* = \Phi^{-1}(1 - \exp(-\lambda/2))$. Also, following §4.1, the majorizing function $t_1(z_1)$ of the marginal Z_1 turns out to be

$$t_1(z_1) = \begin{cases} 0.5(1 - \Phi^2(z^*))^{-1} \phi(z_1) \exp\{-0.5z^{*2}\} & \text{if } z_1 \leq z^*, \\ (1 - \Phi^2(z^*))^{-1} \phi(z_1) & \text{if } z_1 > z^*. \end{cases} \quad (24)$$

It then follows that

$$w_1 = \int_{-\infty}^{z^*} t_1(z_1) dz_1 = 0.5(1 - \Phi^2(z^*))^{-1} \Phi(z^*) \exp(-0.5z^{*2})$$

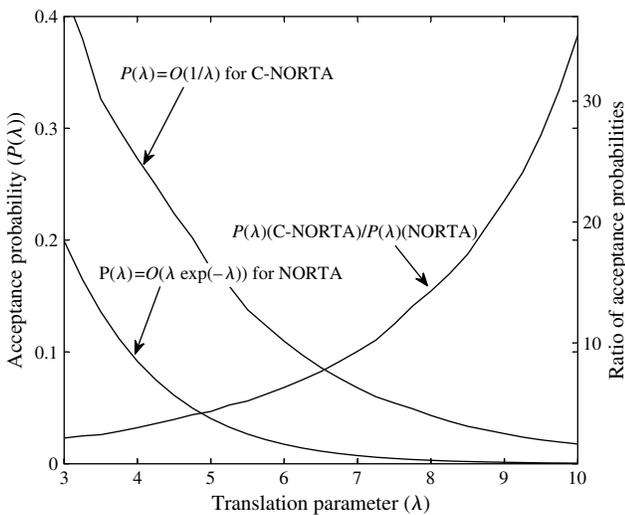


Figure 5 Acceptance Probabilities for Naïve NORTA and C-NORTA on the Simple Problem of Generating Uncorrelated Exponentials X_1 and X_2 (Each with Mean 1) Such That $X_1 + X_2 \geq \lambda$

Notes. The acceptance probability for NORTA, as expected, falls off near exponentially. The acceptance probability for C-NORTA is $O(1/\lambda)$.

and

$$w_2 = \int_{z^*}^\infty t_1(z_1) dz_1 = (1 - \Phi^2(z^*))^{-1} \bar{\Phi}(z^*).$$

This results in an easy-to-generate majorizing function composed of two normals. Again following §4.1, the density of the conditional random variable $Z_2 | Z_1$ is a standard normal if $z_1 > z^*$ and the tail of a standard normal if $z_1 \leq z^*$.

Figure 5 depicts the performance of C-NORTA on this problem. As the theory predicts, the probability of acceptance in C-NORTA is $O(1/\lambda)$ and decays much slower than that obtainable through naïve NORTA. The difference between these methods becomes particularly prominent as the measure of the feasible set gets rarer. For example, when $\lambda = 10$, the acceptance probability in C-NORTA is almost 40 times that of naïve NORTA.

5.2.2. Effect of Correlation. Recall that the example considered thus far requests two exponential random variables X_1, X_2 having a correlation $\rho = 0$ that are constrained to the region $X_1 + X_2 \geq \lambda$. We now relax the assumption $\rho = 0$ toward gaining some insight on the effect of the correlation ρ on the performance of C-NORTA. Specifically, we would like to characterize the range of acceptance probabilities within C-NORTA because of changing ρ and as a function of the translation parameter λ .

We first note that the feasible regions in the exponential and normal spaces (depicted in the first row of Figure 4) are independent of correlation ρ and thus remain unchanged. The majorizing function, however, depends on the correlation ρ and, following §4.1, becomes

$$t_1(z_1) = \begin{cases} 0.5(\Pr\{Z \in \mathcal{G}\})^{-1} \phi(z_1) \exp\{-0.5(1 - \rho^{*2})^{-1} \cdot (z^* - \rho^* z_1)^2\} & \text{if } a(\rho^*) < z_1 \leq z^*, \\ (\Pr\{Z \in \mathcal{G}\})^{-1} \phi(z_1) & \text{otherwise,} \end{cases} \quad (25)$$

where $a(\rho^*) = -\infty$ if $\rho^* \geq 0$, and z^*/ρ^* if $\rho^* < 0$. Recalling that the acceptance probability $P(\lambda) = (\int_{-\infty}^\infty t_1(z_1))^{-1}$ and integrating the expressions in (25) yields

$$\frac{\Pr\{Z \in \mathcal{G}\}}{P(\lambda)} = \begin{cases} \bar{\Phi}(z^*) + 0.5e^{-0.5z^{*2}} \sqrt{1 - \rho^{*2}} \Phi\left(z^* \sqrt{\frac{1 - \rho^*}{1 + \rho^*}}\right) & \text{if } \rho^* \geq 0, \\ \bar{\Phi}(z^*) + \Phi\left(\frac{z^*}{\rho^*}\right) + 0.5e^{-0.5z^{*2}} \sqrt{1 - \rho^{*2}} \cdot \left(\Phi\left(z^* \sqrt{\frac{1 - \rho^*}{1 + \rho^*}}\right) - \Phi\left(z^* \frac{\sqrt{1 - \rho^*}}{\rho^*}\right)\right) & \text{if } \rho^* < 0. \end{cases} \quad (26)$$

Downloaded from informs.org by [128.173.125.76] on 05 February 2014, at 07:19. For personal use only, all rights reserved.

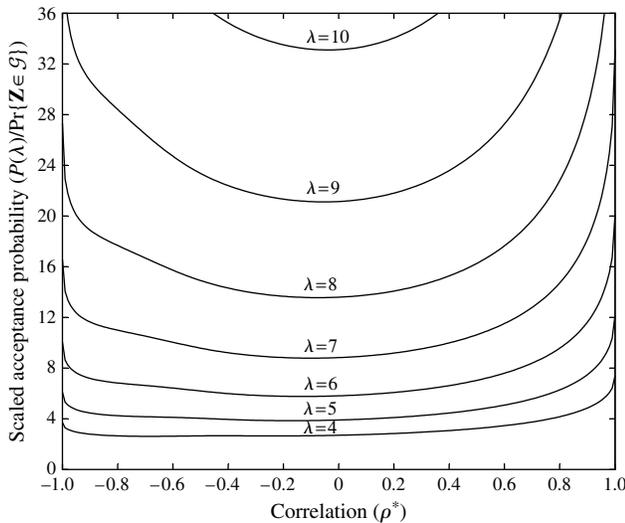


Figure 6 Effect of Correlation on the Acceptance Probabilities Within C-NORTA for the Problem of Generating Correlated Exponentials X_1 and X_2 (Each with Mean 1) Such That $X_1 + X_2 \geq \lambda$

Note. As can be seen, the dependence on correlation increases as a function of the translation parameter, with the smallest acceptance probabilities occurring for the uncorrelated context.

Note that the ratio $P_\lambda(\rho^*) = P(\lambda)/\Pr\{Z \in \mathcal{G}\}$ (inverse of the expression in (26)) has the interpretation of the acceptance probability realized in C-NORTA, scaled by the factor $\Pr\{Z \in \mathcal{G}\}$. Because the factor $\Pr\{Z \in \mathcal{G}\}$ is independent of the correlation ρ^* , plotting $P_\lambda(\rho^*)$ as a function of ρ^* provides insight on how acceptance probabilities within C-NORTA are affected by correlation. Figure 6 does exactly this by depicting the ratio $P_{\rho^*}(\lambda) = P(\lambda)/\Pr\{Z \in \mathcal{G}\}$ as a function of ρ^* for different values of the translation parameter λ .

Three observations about Figure 6, and relating to the example problem, are noteworthy. First, the dependence on ρ^* increases as a function of the translation parameter λ . Furthermore, it can be deduced from the expression in (26) that the limiting values of the ratio $P_\lambda(\rho^*)$ as $\rho^* \rightarrow -1, 1$ are $0.5/\bar{\Phi}(z^*)$ and $1/\bar{\Phi}(z^*)$, respectively. Second, the acceptance probability for C-NORTA seems to have a pronounced minimum for the uncorrelated context but only for large values of the translation parameter λ . Third, for low to moderate values of the translation parameter λ , the acceptance probability within C-NORTA shows a weak dependence on correlation ρ^* .

6. Concluding Remarks

The question of generating random vectors from the tails of bivariate distributions seems to arise in a wide variety of contexts. Despite such presence, little seems to be known on this topic. Specifically, NORTA—arguably the most popular existing method of generating correlated random vectors—is inadequate for

this purpose because it is designed for unconstrained spaces and becomes severely inefficient if modified to work on constrained spaces through simple rejection.

C-NORTA, the algorithm presented in this paper, is an efficient alternative in such contexts. It generalizes NORTA to work on constrained spaces through a strategic conditioning of the NORTA vector, followed by an efficient approximation of the conditional and marginal densities that result from the conditioning. C-NORTA tends to be far more efficient than NORTA in the sense that the acceptance probability of the generated random variates in C-NORTA is exponentially larger, asymptotically. Furthermore, for certain classes of problems, the acceptance probability in C-NORTA is $O(1/\lambda)$, where λ is a certain precisely defined rarity parameter. (A naïve adaptation of NORTA, by contrast, has an acceptance probability that decays exponentially.) C-NORTA differs very little from NORTA from the standpoint of generation. Both methods generate only from appropriate normal densities and use inversion to convert the generated normal random variates to those having the stipulated marginal distributions. C-NORTA’s preprocessing step involves, among other things, constructing a set that encloses the image of the feasible region in the normal space. Owing to the immense structure of the feasible region, such enclosure turns out to be tractable.

Two other remarks relating to future research are now in order.

(i) It is important to note that C-NORTA’s acceptance probability is not bounded away from zero; i.e., although it easily outperforms naïve applications of NORTA for the problem of generating correlated random vectors on constrained spaces, the acceptance probabilities in C-NORTA still decay to zero and can be unacceptably low for certain applications. Given the obvious importance of the problem considered in this paper, further improvements to C-NORTA should be investigated.

(ii) C-NORTA, as presented in this paper, works only in two dimensions. Although we have found numerous applications in the two-dimensional context, extending C-NORTA into higher dimensions would prove to be tremendously useful, particularly if similar gains in efficiency are realizable. As in the bivariate context, the main challenge for such extension is the construction of a majorizing function for an appropriately defined marginal distribution. One possible factorization of the n -dimensional joint normal $\mathbf{X}(n > 2)$ leads to an $(n-1)$ -dimensional joint marginal for $\mathbf{X}_{-1} = \{X_2, X_3, \dots, X_n\}$ and a single-dimensional conditional distribution for $X_1 | \mathbf{X}_{-1}$. This form allows us to again express the joint marginal of \mathbf{X}_{-1} as a normal mixture of normal tail probabilities, which can further be majorized by a mixture

of $(n-1)$ -dimensional joint normal densities. Construction of a generation scheme to sample from such a density, especially with a multidimensional piecewise-linear constraining boundary, is a nontrivial extension and part of ongoing research.

Acknowledgments

The authors thank the anonymous referees and the associate editor for their thorough reviews and suggestions. The second author was supported in part by the Office of Naval Research Grants N000140810066, N000140910997, and N000141110065.

References

- Asmussen, S., P. W. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis. Stochastic Modelling and Applied Probability*, Vol. 57. Springer, New York.
- Avramidis, A. N., N. Channouf, P. L'Ecuyer. 2009. Efficient correlation matching for fitting discrete multivariate distributions with arbitrary marginals and normal-copula dependence. *INFORMS J. Comput.* **21**(1) 88–106.
- Bartle, R. G. 1976. *The Elements of Real Analysis*. John Wiley & Sons, New York.
- Cario, M. C., B. L. Nelson. 1997. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Technical report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL.
- Cario, M. C., B. L. Nelson. 1998. Numerical methods for fitting and simulating autoregressive-to-anything processes. *INFORMS J. Comput.* **10**(1) 72–81.
- Chen, H. 2001. Initialization for NORTA: Generation of random vectors with specified marginals and correlations. *INFORMS J. Comput.* **13**(4) 312–331.
- Damien, P., S. G. Walker. 2001. Sampling truncated normal, beta, and gamma densities. *J. Comput. Graphical Statist.* **10**(2) 206–215.
- Gelfand, A. E., A. F. M. Smith, T.-M. Lee. 1992. Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *J. Amer. Statist. Assoc.* **87**(418) 523–532.
- Ghosh, S., S. G. Henderson. 2003. Behavior of the NORTA method for correlated random vector generation as the dimension increases. *ACM Trans. Model. Comput. Simul.* **13**(3) 276–294.
- Glasserman, P., J. Li. 2005. Importance sampling for portfolio credit risk. *Management Sci.* **51**(11) 1643–1656.
- Glasserman, P., P. Heidelberger, P. Shahabuddin. 2000. Variance reduction techniques for estimating value-at-risk. *Management Sci.* **46**(10) 1349–1364.
- Glasserman, P., P. Heidelberger, P. Shahabuddin. 2002. Portfolio value-at-risk with heavy-tailed risk factors. *Math. Finance* **12**(3) 239–269.
- Huang, P., D. Subramanian. 2009. Iterative estimation maximization for stochastic linear programs with conditional-value-at-risk constraints. Technical report, IBM T. J. Watson Research Center, Yorktown Heights, NY.
- Johnson, N. L., S. Kotz, N. Balakrishnan. 1994. *Continuous Univariate Distributions*, Vol. 1. John Wiley & Sons, New York.
- Juneja, S., P. Shahabuddin. 2006. Rare-event simulation techniques: An introduction and recent advances. S. G. Henderson, B. L. Nelson, eds. *Handbooks in Operations Research and Management Science: Simulation*, Vol. 13. Elsevier Science, Amsterdam, 291–350.
- Law, A. M., W. D. Kelton. 2000. *Simulation Modeling and Analysis*, 3rd ed. McGraw-Hill, New York.
- Lu, D., W. V. Li. 2009. A note on multivariate Gaussian estimates. *J. Math. Anal. Appl.* **354**(2) 704–707.
- Nelsen, R. B. 1999. *An Introduction to Copulas. Lecture Notes in Statistics*, Vol. 139. Springer-Verlag, New York.
- Ross, S. 2006. *Simulation*. Elsevier, Burlington, MA.
- 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.
- Walker, A. J. 1977. An efficient method for generating discrete random variables with general distributions. *ACM Trans. Math. Software* **3**(3) 253–256.
- Whitt, W. 1976. Bivariate distributions with given marginals. *Ann. Statist.* **4**(6) 1280–1289.