

## FITTING A NORMAL COPULA FOR A MULTIVARIATE DISTRIBUTION WITH BOTH DISCRETE AND CONTINUOUS MARGINALS

Nabil Channouf

GERAD, Pavillon Aisenstadt  
2920 Chemin de la Tour, Université de Montréal  
Montréal (Qc), H3T 1J4, CANADA

Pierre L'Ecuyer

DIRO, Université de Montréal  
C.P. 6128, Succ. Centre-Ville  
Montréal (Qc), H3C 3J7, CANADA

### ABSTRACT

We consider a multivariate distribution with both discrete and continuous marginals, for which the dependence is modeled by a normal copula (sometimes called the NORTA method), and provide an algorithm for fitting the copula in that situation. The fitting is done by matching (approximately) either the rank correlations or the product moment correlations for all pairs of marginals. Numerical illustrations are provided.

### 1 INTRODUCTION

Random vectors are essential to capture the dependence between various quantities in stochastic models. Capturing this dependence is often even more important than picking the correct marginal distributions. The usual way of defining a multivariate distribution is to first specify the marginal distributions, and then select a copula that determines the dependence structure between these marginals (Joe 1997, Nelsen 1999, Hörmann, Leydold, and Derflinger 2004).

A popular and convenient type of copula, especially when the dimension is large, is the normal copula, which depends only on the pairwise rank correlations between the marginals when these marginal are continuous (Mardia 1970). Modeling dependence by a normal copula is also known as the NORTA (*normal to anything*) method (Cario and Nelson 1997, Ghosh and Henderson 2003). It consists in specifying a multinormal distribution with given correlation matrix, where each marginal is standard normal (with mean 0 and variance 1). To generate a random vector with arbitrary marginal distributions, one generate a vector from this multinormal distribution, then transform each coordinate by applying first the standard normal distribution function, and then the inverse distribution function of the desired marginal. All the dependence is determined by the correlation matrix of the multinormal.

This correlation matrix is usually specified by specifying a target rank correlation for each pair of marginals (e.g., to match the sample rank correlation observed in some data set), and then finding the correlation between the corresponding pair of standard normals that would produce that target rank correlation for the marginals. We refer to the latter as the *pairwise correlation matching problem*. Solving this problem for each pair of marginals provides a proposal correlation matrix for the standard normal distribution. It often happens, particularly when the dimension is large, that this proposal matrix is not positive semi-definite (Ghosh and Henderson 2003), but in that case one can compute the “closest” positive semidefinite matrix (for a given norm on matrices) by solving an optimization problem that minimizes the distance to the proposal (Ghosh and Henderson 2003, Lurie and Goldberg 1998). Typically, this distance turns out to be small.

Our concern in this paper is to solve the pairwise correlation matching problem. When the two marginals are continuous, the solution is straightforward, because a simple analytic formula gives directly the correlation  $\rho$  between the standard normals as a function of the desired rank correlation  $r$ . We have  $\rho = 2\sin(\pi r/6)$ . But no direct formula is available if one (or both) marginal is discrete. The case where both marginals are discrete was studied by Avramidis, Channouf, and L'Ecuyer (2009), who proposed efficient approximation algorithms to match either the rank or the linear correlations. But their methods do not cover the case where one marginal is discrete and the other one is continuous. The aim of the present paper is to fill this gap.

Our proposed method builds on similar ideas as in Avramidis, Channouf, and L'Ecuyer (2009), with adaptations, and we end up with a different and simpler algorithm. In particular, the method involves the computation of the bivariate normal

distribution, which requires a double integration in Avramidis, Channouf, and L'Ecuyer (2009). But here we transform the two-dimensional integral to a one-dimensional one before solving it numerically. Moreover, whereas Avramidis, Channouf, and L'Ecuyer (2009) have double infinite sums to approximate, here we only have simple sums to handle.

Instead of specifying target rank correlations, one can alternatively specify the usual linear (product-moment) correlations between the marginals. These correlations do depend on the marginals and have other drawbacks discussed in Hörmann, Leydold, and Derflinger (2004) and Avramidis, Channouf, and L'Ecuyer (2009). Our recommendation is to use the rank correlations, which are generally preferable and easier to handle. But our methodology nevertheless covers the product-moment correlations as well.

The remainder is organized as follow. In Section 2, we recall basic facts and definitions on normal copulas and the correlation matching problem, for both the rank and linear correlations. In Section 3 we present our method for the mixed case. We give numerical examples in Section 4 and we conclude in Section 5.

## 2 THE NORMAL COPULA

### 2.1 Definition and the Matching Problem

We want to specify the probability distribution of a random vector  $\mathbf{X} = (X_1, \dots, X_d)^t$  (we use column vectors and “t” means “transposed”) by specifying (i) the marginal distribution of each  $X_j$  and (ii) the dependence structure between them. There is an extremely rich variety of copulas that can be used to determine this dependence structure (Joe 1997, Nelsen 1999). In this paper, we restrict ourselves to the normal copula, whose main advantage is its simplicity and its ability to match (exactly or approximately) the pairwise correlations.

Let  $F_j$  be the distribution function of  $X_j$ , defined by  $F_j(x) = \mathbb{P}[X_j \leq x]$ . Recall that  $F_j(X_j)$  has the uniform distribution over  $(0, 1)$  when  $X_j$  has a continuous distribution ( $F_j$  is continuous), but not necessarily when  $X_j$  is discrete. For  $1 \leq i, j \leq d$ , the product moment (linear) *correlation coefficient* between  $X_i$  and  $X_j$  is defined by

$$\rho_{i,j}^X = \text{Corr}(X_i, X_j) = \frac{\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]}{(\text{Var}[X_i]\text{Var}[X_j])^{1/2}}$$

and the *rank correlation* (also called the *Spearman's rho*) is

$$r_{i,j}^X = \text{Corr}(F_i(X_i), F_j(X_j)).$$

A normal copula is defined by specifying a correlation matrix  $\mathbf{R}_\rho^Z$  with elements  $\rho_{i,j}^Z = \text{Corr}(Z_i, Z_j)$ ,  $1 \leq i, j \leq d$ , for a multinormal vector  $\mathbf{Z} = (Z_1, \dots, Z_d)^t$  with standard normal marginals. We then define

$$X_i = F_i^{-1}(\Phi(Z_i)), \quad i = 1, \dots, d, \tag{1}$$

where  $\Phi$  is the standard normal cumulative distribution function and  $F_i^{-1}(u) = \inf\{x : F_i(x) \geq u\}$ . To generate a realization of  $\mathbf{X}$ , it suffices to generate a realization of  $\mathbf{Z}$  and to compute  $\mathbf{X}$  via (1). The multinormal vector  $\mathbf{Z}$  can be generated in the standard way: decompose  $\mathbf{R}_\rho^Z$  as  $\mathbf{L}\mathbf{L}^t$  for some matrix  $\mathbf{L}$ , generate a vector  $\mathbf{W}$  of  $d$  independent standard normal random variables, and put  $\mathbf{Z} = \mathbf{L}\mathbf{W}$ . The correlation matrix  $\mathbf{R}_\rho^Z$  must be symmetric and its diagonal elements must be ones, so there are  $d(d-1)/2$  elements to specify. The pairwise correlation matching problem studied in this paper consists in finding, for each pair  $1 \leq i < j \leq d$ , the value of  $\rho_{i,j}^Z$  that produces a given target rank correlation  $r_{i,j}^X$  or a given product-moment correlation  $\rho_{i,j}^X$ .

In the case where  $X_i$  has a continuous distribution,  $F_i(X_i) = \Phi(Z_i)$  has the uniform distribution over  $(0, 1)$ . If both  $X_i$  and  $X_j$  have continuous distributions, then  $r_{i,j}^X$  is the same as the rank correlation between  $Z_i$  and  $Z_j$ , and the matching problem is solved by the simple formula  $\rho_{i,j}^Z = 2 \sin(\pi r_{i,j}^X/6)$ . In the remainder, we focus on the case where some  $X_i$  has a discrete distribution, in which case  $F_i(F_i^{-1})$  is not the identity function. We consider a single pair  $(i, j)$  at a time. When  $d > 2$ , this pairwise correlation matching problem must be solved for all  $d(d-1)/2$  pairs  $(i, j)$  with  $i < j$ .

## 2.2 Matching the Rank or the Product-Moment Correlations

The rank correlation  $r_{i,j}^X$  satisfies

$$\begin{aligned} r_{i,j}^X &= \text{Corr}(F_i(X_i), F_j(X_j)) = \text{Corr}(F_i(F_i^{-1}(\Phi(Z_i))), F_j(F_j^{-1}(\Phi(Z_j)))) \\ &= \frac{\mathbb{E}\left[F_i(F_i^{-1}(\Phi(Z_i)))F_j(F_j^{-1}(\Phi(Z_j)))\right] - \mu_{F_i}\mu_{F_j}}{\sigma_{F_i}\sigma_{F_j}} = \frac{g_r(\rho_{i,j}^Z) - \mu_{F_i}\mu_{F_j}}{\sigma_{F_i}\sigma_{F_j}}, \end{aligned} \quad (2)$$

where

$$g_r(\rho_{i,j}^Z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_i(F_i^{-1}(\Phi(z_i)))F_j(F_j^{-1}(\Phi(z_j)))\phi(z_i, z_j, \rho_{i,j}^Z)dz_idz_j, \quad (3)$$

for  $1 \leq i < j \leq d$ ,  $\mu_{F_i}$  and  $\sigma_{F_i}^2$  are the mean and variance of  $F_i(X_i)$ , and  $\phi(\cdot, \cdot, \rho)$  is the density of the bivariate standard normal distribution with correlation  $\rho$ . Rearranging (2), we see that the correlation matching problem amounts to finding a root  $\rho = \rho_{i,j}^Z$  of the equation:

$$g_r(\rho) - \mu_{F_i}\mu_{F_j} - r_{i,j}^X\sigma_{F_i}\sigma_{F_j} = 0, \quad (4)$$

for  $1 \leq i < j \leq d$ .

For the product-moment correlation  $\rho_{i,j}^X$ , we have

$$\begin{aligned} \rho_{i,j}^X &= \text{Corr}(X_i, X_j) = \text{Corr}(F_i^{-1}(\Phi(Z_i)), F_j^{-1}(\Phi(Z_j))) \\ &= \frac{\mathbb{E}\left[F_i^{-1}(\Phi(Z_i))F_j^{-1}(\Phi(Z_j))\right] - \mu_i\mu_j}{\sigma_i\sigma_j} = \frac{g_l(\rho_{i,j}^Z) - \mu_i\mu_j}{\sigma_i\sigma_j}, \end{aligned} \quad (5)$$

where

$$g_l(\rho_{i,j}^Z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_i^{-1}(\Phi(z_i))F_j^{-1}(\Phi(z_j))\phi(z_i, z_j, \rho_{i,j}^Z)dz_idz_j, \quad (6)$$

where  $\mu_i$  and  $\sigma_i^2$  are the mean and variance of  $X_i$ . The corresponding correlation matching problem amounts to finding a root  $\rho = \rho_{i,j}^Z$  of

$$g_l(\rho) - \mu_i\mu_j - \rho_{i,j}^X\sigma_i\sigma_j = 0, \quad (7)$$

for  $1 \leq i < j \leq d$ .

Avramidis, Channouf, and L'Ecuyer (2009) provide efficient algorithms for the root-finding problems (4) and (7) when both marginal are discrete. In the next section, we provide an algorithm for the situation where one marginal is discrete and the other is continuous.

## 3 FITTING IN THE CASE OF DISCRETE AND CONTINUOUS MARGINALS

Here we assume that  $X_i$  has a continuous distribution over the real line and that  $X_j$  has a discrete distribution over the integers. We develop algorithms for solving (4) and (7) for this case. The method differs from that of Avramidis, Channouf, and L'Ecuyer (2009), although it is based on similar ideas.

Using the fact that

$$\phi(z_i, z_j, \rho_{i,j}^Z) = \phi(z_i)\phi\left(\frac{z_j - \rho_{i,j}^Z z_i}{1 - (\rho_{i,j}^Z)^2}\right)(1 - (\rho_{i,j}^Z)^2)^{-1/2}$$

where  $\phi$  is the density of the standard normal, we can write (3) as

$$\begin{aligned}
 g_r(\rho_{i,j}^Z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_i(F_i^{-1}(\Phi(z_i))) F_j(F_j^{-1}(\Phi(z_j))) \phi(z_i, z_j, \rho_{i,j}^Z) dz_i dz_j \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi(z_i) F_j(F_j^{-1}(\Phi(z_j))) \phi(z_i) \phi\left(\frac{z_j - \rho_{i,j}^Z z_i}{\sqrt{1 - (\rho_{i,j}^Z)^2}}\right) (1 - (\rho_{i,j}^Z)^2)^{-1/2} dz_i dz_j \\
 &= \int_{-\infty}^{\infty} \frac{\Phi(z_i) \phi(z_i)}{\sqrt{1 - (\rho_{i,j}^Z)^2}} \left[ \int_{-\infty}^{\infty} F_j(F_j^{-1}(\Phi(z_j))) \phi\left(\frac{z_j - \rho_{i,j}^Z z_i}{\sqrt{1 - (\rho_{i,j}^Z)^2}}\right) dz_j \right] dz_i \\
 &= \int_{-\infty}^{\infty} \frac{\Phi(z_i) \phi(z_i)}{\sqrt{1 - (\rho_{i,j}^Z)^2}} \left[ \sum_{l=-\infty}^{\infty} f_{j,l} \int_{z_{j,l-1}}^{z_{j,l}} \phi\left(\frac{z_j - \rho_{i,j}^Z z_i}{\sqrt{1 - (\rho_{i,j}^Z)^2}}\right) dz_j \right] dz_i \\
 &= \sum_{l=-\infty}^{\infty} f_{j,l} \left[ \int_{-\infty}^{\infty} \Phi(z_i) \phi(z_i) \left( \Phi\left(\frac{z_{j,l} - \rho_{i,j}^Z z_i}{\sqrt{1 - (\rho_{i,j}^Z)^2}}\right) - \Phi\left(\frac{z_{j,l-1} - \rho_{i,j}^Z z_i}{\sqrt{1 - (\rho_{i,j}^Z)^2}}\right) \right) dz_i \right], \tag{8}
 \end{aligned}$$

where  $f_{j,l} = F_j(l)$  and  $z_{j,l} = \Phi^{-1}(f_{j,l})$ . Here we have transformed the integral involving the discrete distribution of  $X_j$  to a sum of terms where each term is a constant multiplied by an area under the normal density. Note that in Avramidis, Channouf, and L'Ecuyer (2009), we had to apply the same transformation with a bivariate normal density, because both marginals were discrete. We now apply the change of variable  $u_i = \Phi(z_i)$  to obtain an integral over the interval (0, 1):

$$g_r(\rho_{i,j}^Z) = \sum_{l=-\infty}^{\infty} f_{j,l} \left[ \int_0^1 u_i \left( \Phi\left(\frac{z_{j,l} - \rho_{i,j}^Z \Phi^{-1}(u_i)}{\sqrt{1 - (\rho_{i,j}^Z)^2}}\right) - \Phi\left(\frac{z_{j,l-1} - \rho_{i,j}^Z \Phi^{-1}(u_i)}{\sqrt{1 - (\rho_{i,j}^Z)^2}}\right) \right) du_i \right]. \tag{9}$$

To find a root of (4) with  $g_r$  given in (9), as in the algorithm NII of Avramidis, Channouf, and L'Ecuyer (2009), we use the procedure `zero` of Brent (1971), also described in Press, Teukolsky, Vetterling, and Flannery (1992), and called Brent's method. It combines root bracketing, bisection, and inverse quadratic interpolation. To find the root of a function  $f$ , at each step of the algorithm, we have an interval  $[a, b]$  that contains the root, so  $f(a)$  and  $f(b)$  have opposite signs. The algorithm then selects another point  $c$  in the interval, either by bisection, or by inverse quadratic interpolation at steps where this is safe. In our case, we are looking for a root of the function

$$f(\rho) = g_r(\rho) - \mu_{F_i} \mu_{F_j} - r_{i,j}^X \sigma_{F_i} \sigma_{F_j}, \tag{10}$$

with  $g_r$  given in (9). We know that the solution is in the interval  $[-1, 0]$  when the target correlation is negative, and in  $[0, 1]$  when it is positive, so we start with one of those intervals. The algorithm stops when it finds a number that is guaranteed to be at a distance smaller than  $\varepsilon$  from the true root, for a given (selected) tolerance  $\varepsilon > 0$ . We compute the integral in (9) numerically, which is not problematic because the integrand is fairly smooth as a function of  $u_i$ . In our implementation, we took the Gauss-Legendre quadrature, using the numerical integration classes from the Java library of Flanagan (2007). The sum in (9) is truncated to some index  $\ell_0$  such that  $\mathbb{P}[|X_j| > \ell_0] < \varepsilon_1$ . In our implementation, we took  $\varepsilon_1 = 10^{-6}$ . The usual discrete distributions are over the non-negative integers; then we only need to truncate on the right. The correlation-matching procedure for an arbitrary pair  $(i, j)$  is summarized in Algorithm 1.

---

**Algorithm 1** Pairwise correlation-matching algorithm.

// This algorithm takes as input the marginal distributions  $F_i$  and  $F_j$  of  $X_i$  and  $X_j$  where  $F_i$  is continuous and  $F_j$  is discrete, their rank correlation  $r_{i,j}^X$ , the maximum number of iterations  $N_{\max}$ , and the desired tolerance  $\varepsilon$ .

It returns the product moment correlation  $\rho = \rho_{i,j}^Z$  that solves (approximately) (10).

1. **Initialize** the search interval:  $[a, b] = [0, 1]$  if  $r_{i,j}^X > 0$ , and  $[a, b] = [-1, 0]$  otherwise;
  2. Apply Brent's method to find a root of (10), as described in Press, Teukolsky, Vetterling, and Flannery (1992);
  3. **Return**  $\rho_{i,j}^Z$ .
-

Table 1: Results for the rank correlations

	$\varepsilon$	$r_{1,2}^X$	$\rho_{1,2}^Z$	CPU (s)	$N_1$	$r_{1,2}^X(\rho_{1,2}^Z)$	Rel. error (%)
$X_1$ <b>Exponential</b> $\lambda_1 = 25$ $X_2$ <b>Binomial</b> $n_2 = 100$ $p_2 = 0.65$	$10^{-2}$	-0.50	-0.5189	0.0033	4	-0.4995	0.106
		0.05	0.0503	0.0023	3	0.0479	4.293
		0.20	0.2099	0.0032	4	0.1999	0.025
		0.90	0.9110	0.0023	3	0.9010	0.109
		0.98	0.9866	0.0023	3	0.9827	0.275
	$10^{-4}$	-0.50	-0.5194	0.0042	5	-0.5000	< 0.001
		0.05	0.0526	0.0032	4	0.0500	< 0.001
		0.20	0.2099	0.0041	5	0.2000	< 0.001
		0.90	0.9101	0.0041	5	0.9000	< 0.001
		0.98	0.9841	0.0041	5	0.9800	< 0.001
$X_1$ <b>Gamma</b> $\alpha_1 = 20$ $\beta_1 = 30$ $X_2$ <b>Negative bin.</b> $s_2 = 60.21$ $p_2 = 0.611$	$10^{-2}$	-0.50	-0.5183	0.0036	4	-0.4997	0.053
		0.05	0.0513	0.0026	3	0.0489	2.300
		0.20	0.2102	0.0026	3	0.2006	0.280
		0.90	0.9088	0.0047	5	0.9000	< 0.001
		0.96	0.9633	0.0047	5	0.9587	0.135
	$10^{-4}$	-0.50	-0.5186	0.0046	5	-0.5000	< 0.001
		0.05	0.0525	0.0036	4	0.0500	< 0.001
		0.20	0.2096	0.0036	4	0.2000	< 0.001
		0.90	0.9088	0.0047	5	0.9000	< 0.001
		0.96	0.9644	0.0058	6	0.9600	< 0.001
$X_1$ <b>Normal</b> $\mu_1 = 25$ $\sigma_1 = 10$ $X_2$ <b>Poisson</b> $\lambda = 25$	$10^{-2}$	-0.50	-0.5185	0.0024	4	-0.4992	0.163
		0.05	0.0507	0.0017	3	0.0482	3.657
		0.20	0.2098	0.0024	4	0.2000	< 0.001
		0.90	0.9118	0.0018	3	0.9018	0.201
		0.98	0.9836	0.0024	4	0.9794	0.058
	$10^{-4}$	-0.50	-0.5193	0.0030	5	-0.5000	< 0.001
		0.05	0.0526	0.0024	4	0.0500	< 0.001
		0.20	0.2099	0.0024	4	0.2000	< 0.001
		0.90	0.9101	0.0031	5	0.9000	< 0.001
		0.98	0.9841	0.0039	6	0.9800	< 0.001

For the product moment correlations  $\rho_{i,j}^X$ ,  $1 \leq i < j \leq d$ , we obtain a similar formula by transforming Equation (6):

$$\begin{aligned}
 g_l(\rho_{i,j}^Z) &= \sum_{l=-\infty}^{\infty} y_{j,l} \left[ \int_{-\infty}^{\infty} F_i^{-1}(\Phi(z_i)) \phi(z_i) \left( \Phi \left( \frac{z_{j,l} - \rho_{i,j}^Z z_i}{\sqrt{1 - (\rho_{i,j}^Z)^2}} \right) - \Phi \left( \frac{z_{j,l-1} - \rho_{i,j}^Z z_i}{\sqrt{1 - (\rho_{i,j}^Z)^2}} \right) \right) dz_i \right] \\
 &= \sum_{l=-\infty}^{\infty} y_{j,l} \left[ \int_0^1 F_i^{-1}(u_i) \left( \Phi \left( \frac{z_{j,l} - \rho_{i,j}^Z \Phi^{-1}(u_i)}{\sqrt{1 - (\rho_{i,j}^Z)^2}} \right) - \Phi \left( \frac{z_{j,l-1} - \rho_{i,j}^Z \Phi^{-1}(u_i)}{\sqrt{1 - (\rho_{i,j}^Z)^2}} \right) \right) du_i \right], \tag{11}
 \end{aligned}$$

where  $y_{j,l} = \mathbb{P}[X_j = l]$ . To find a root of (7), with  $g_l$  given in (11), we use a similar algorithm as for (4).

#### 4 NUMERICAL EXAMPLES

We report the performance of our algorithm for some numerical examples. We consider three pairs of distributions (three cases), as follows:

1.  $X_1$  is exponential with mean  $\lambda_1 = 25$  and  $X_2$  is binomial with parameters  $n_2 = 100$  and  $p_2 = 0.5$ .

Table 2: Results for the product moment correlations

	$\epsilon$	$\rho_{1,2}^X$	$\rho_{1,2}^Z$	CPU (s)	$N_1$	$r_{1,2}^X(\rho_{1,2}^Z)$	Rel. error (%)
<b><math>X_1</math> Exponential</b> $\lambda_1 = 25$ <b><math>X_2</math> Binomial</b> $n_2 = 100$ $p_2 = 0.5$	$10^{-2}$	-0.50	-0.5365	0.0038	4	-0.5002	0.044
		0.05	0.0824	0.0049	5	0.0500	< 0.001
		0.20	0.2531	0.0038	4	0.2018	0.886
		0.50	0.5884	0.0038	4	0.5000	< 0.001
		0.80	0.9277	0.0028	3	0.8017	0.211
	$10^{-4}$	-0.50	-0.5362	0.0049	5	-0.5000	< 0.001
		0.05	0.0824	0.0049	5	0.0500	< 0.001
		0.20	0.2511	0.0049	5	0.2000	< 0.001
		0.50	0.5885	0.0038	4	0.5000	< 0.001
		0.80	0.9258	0.0039	4	0.8000	< 0.001
<b><math>X_1</math> Gamma</b> $\alpha_1 = 20$ $\beta_1 = 30$ <b><math>X_2</math> Negative bin.</b> $s_2 = 60.21$ $p_2 = 0.611$	$10^{-2}$	-0.50	-0.5067	0.0343	4	-0.4998	0.044
		0.05	0.0528	0.0253	3	0.0489	2.264
		0.20	0.2062	0.0255	3	0.1998	0.101
		0.50	0.5098	0.0341	4	0.4992	0.167
		0.90	0.9149	0.0430	5	0.9000	< 0.001
	$10^{-4}$	-0.50	-0.5070	0.0424	5	-0.5000	< 0.001
		0.05	0.0540	0.0342	4	0.0500	< 0.001
		0.20	0.2064	0.0428	5	0.2000	< 0.001
		0.50	0.5106	0.0428	5	0.5000	< 0.001
		0.90	0.9149	0.0428	5	0.9000	< 0.001
<b><math>X_1</math> Normal</b> $\mu_1 = 25$ $\sigma_1 = 10$ <b><math>X_2</math> Poisson</b> $\lambda = 25$	$10^{-2}$	-0.50	-0.5080	0.0019	3	-0.5024	0.487
		0.05	0.0502	0.0019	3	0.0496	0.724
		0.20	0.2009	0.0019	3	0.1986	0.704
		0.90	0.9104	0.0027	4	0.8995	0.053
		0.98	0.9896	0.0027	4	0.9777	0.236
	$10^{-4}$	-0.50	-0.5055	0.0026	4	-0.5000	< 0.001
		0.05	0.0506	0.0027	4	0.0500	< 0.001
		0.20	0.2024	0.0026	4	0.2000	< 0.001
		0.90	0.9109	0.0034	5	0.9000	< 0.001
		0.98	0.9919	0.0041	6	0.9800	< 0.001

- $X_1$  is gamma with shape parameter  $\alpha_1 = 20$  and scale parameter  $\beta_1 = 30$  (so the mean is  $\alpha_1\beta_1$ ), and  $X_2$  is negative binomial with parameters  $s_2 = 60.21$  and  $p_2 = 0.611$ .
- $X_1$  is normal with mean  $\mu_1 = 25$  and variance  $\sigma_1^2 = 100$ , and  $X_2$  is Poisson with parameter (mean)  $\lambda_2 = 25$ .

We tested the algorithm for both the rank and the product moment correlations, with several target values of  $r_{1,2}^X$  and  $\rho_{1,2}^X$ , in the interval  $(-1, 1)$ , that provide a reasonable coverage of the different situations that might occur. We considered two choices of tolerances for the root-finding algorithm based on Brent's method:  $\epsilon = 10^{-2}$  and  $\epsilon = 10^{-4}$ . In each case, we also computed the average CPU time over  $10^4$  replications, the number  $N_1$  of iterations (that is, the number of evaluations of the function  $g_r$  or  $g_l$ ), and the relative error. The results are in Tables 1 and 2.

The computing time is less than 5 milliseconds in almost all cases, except for the combination of gamma and negative binomial distributions when we specify the product moment correlations, which is about 10 times slower with the smallest tolerance. These times are faster than those reported by Avramidis, Channouf, and L'Ecuyer (2009) for the NI1 algorithm, by a factor of 10 or more for the similar examples.. When the tolerance is reduced from  $\epsilon = 10^{-2}$  to  $\epsilon = 10^{-4}$ ,  $N_1$  increases by only 1 or 2, and the CPU times increase only marginally. The largest relative error is observed with a target rank correlation of 0.05 and  $\epsilon = 10^{-2}$ ; in this case the target rank correlation is close to zero, so the "large" relative error corresponds to a small absolute error.

## 5 CONCLUSION

We proposed an efficient correlation matching algorithm for a normal copula in the case of a mixture of discrete and continuous marginal distributions. The method is faster than the algorithms proposed in Avramidis, Channouf, and L'Ecuyer (2009) for the case where both marginals are discrete.

## ACKNOWLEDGMENTS

This work has been supported by Grants OGP-0110050 and CRDPJ-320308 from NSERC-Canada, a grant from Bell Canada via the Bell University Laboratories, and a Canada Research Chair to the second author.

## REFERENCES

- Avramidis, A. N., N. Channouf, and P. L'Ecuyer. 2009. Efficient correlation matching for fitting discrete multivariate distributions with arbitrary marginals and normal copula dependence. *INFORMS Journal of Computing* 21:88–106.
- Brent, R. P. 1971. An algorithm with guaranteed convergence for finding a zero of a function. *Computer Journal* 14:422–425.
- Cario, M. C., and B. L. Nelson. 1997. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. working paper.
- Flanagan, M. T. 2007. *Michael Thomas Flanagan's scientific Java library*. Available at <<http://www.ee.ucl.ac.uk/~mflanaga>>.
- Ghosh, S., and S. Henderson. 2003. Behaviour of the NORTA method for correlated random vector generation as the dimension increases. *ACM Transactions on Modeling and Computer Simulation* 13:276–294.
- Hörmann, W., J. Leydold, and G. Derflinger. 2004. *Automatic nonuniform random variate generation*. Berlin: Springer-Verlag.
- Joe, H. 1997. *Multivariate models and dependence concepts*. London: Chapman and Hall.
- Lurie, P. M., and M. S. Goldberg. 1998. An approximate method for sampling correlated random variables from partially-specified distributions. *Management Science* 44:203–218.
- Mardia, K. V. 1970. A translation family of bivariate distributions and Fréchet's bounds. *Sankhya* A32:119–122.
- Nelsen, R. B. 1999. *An introduction to copulas*, Volume 139 of *Lecture Notes in Statistics*. New York, NY: Springer-Verlag.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. *Numerical recipes in C: The art of scientific computing*. second ed. New York, NY: Cambridge University Press.

## AUTHOR BIOGRAPHIES

**NABIL CHANNOUF** received a M.Sc. in statistics in 2001 and a PhD in Operations Research in 2008, both from the Université de Montréal. He is currently postdoctoral fellow in the GERAD research center. His main research interests are in applied statistics, with a focus on the modeling of non-stationary and doubly stochastic arrival processes in call centers and other types of service systems.

**PIERRE L'ECUYER** is Professor in the Département d'Informatique et de Recherche Opérationnelle, at the Université de Montréal, Canada. He holds the Canada Research Chair in Stochastic Simulation and Optimization. He is a member of the CIR-RELT and GERAD research centers. His main research interests are random number generation, quasi-Monte Carlo methods, efficiency improvement via variance reduction, sensitivity analysis and optimization of discrete-event stochastic systems, and discrete-event simulation in general. He is currently Associate/Area Editor for *ACM Transactions on Modeling and Computer Simulation*, *ACM Transactions on Mathematical Software*, *Statistics and Computing*, *Management Science*, *International Transactions in Operational Research*, *The Open Applied Mathematics Journal*, and *Cryptography and Communications*. He obtained the *E. W. R. Steacie* fellowship in 1995-97, a *Killam* fellowship in 2001-03, and became an INFORMS Fellow in 2006. His recent research articles are available on-line from his web page: <<http://www.iro.umontreal.ca/~lecuyer>>.