Constructing Discrete Unbounded Distributions with Gaussian-Copula Dependence and Given Rank Correlation

Athanassios N. Avramidis
School of Mathematics, University of Southampton
Highfield, Southampton, SO17 1BJ, UNITED KINGDOM
aa1w07@soton.ac.uk

April 25, 2013

Abstract

A random vector $\mathbf{X}$ with given univariate marginals can be obtained by first applying the normal distribution function to each coordinate of a vector $\mathbf{Z}$ of correlated standard normals to produce a vector $\mathbf{U}$ of correlated uniforms over $(0, 1)$ and then transforming each coordinate of $\mathbf{U}$ by the relevant inverse marginal. One approach to fitting requires, separately for each pair of coordinates of $\mathbf{X}$, the rank correlation, $r(\rho)$, or the product-moment correlation, $r_{\mathrm{L}}(\rho)$, where ρ is the correlation of the corresponding coordinates of $\mathbf{Z}$, to equal some target r^*. We prove the existence and uniqueness of a solution for any feasible target, without imposing restrictions on the marginals. For the case where $r(\rho)$ cannot be computed exactly due to an infinite discrete support, the relevant infinite sums are approximated by truncation, and lower and upper bounds on the truncation errors are developed. With a function $\tilde{r}(\rho)$ defined by the truncated sums, a bound on the error $r(\rho^*) - r^*$ is given, where ρ^* is a solution to $\tilde{r}(\rho^*) = r^*$. Based on this bound, an algorithm is proposed that determines truncation points so that the solution has any specified accuracy. The new truncation method has potential for significant work reduction relative to truncating heuristically, largely because as required accuracy decreases, so does the number of terms in the truncated sums. This is quantified with examples. The gain appears to increase with the heaviness of tails.

Keywords: statistics; multivariate distribution; unbounded discrete distribution; correlation; Gaussian copula.

History: Accepted by Marvin Nakayama, Area Editor for Simulation; received September 2011; revised October 2012, April 2013; accepted May 2013.

1 Introduction

A multivariate distribution may be specified via marginal univariate distributions and with dependence between marginals induced via a Gaussian (normal) copula. This is also known as the NORmal To Anything (NORTA) approach (Cario and Nelson, 1996, 1997). More precisely, let F_k, $k = 1, \ldots, d$ be univariate (cumulative) distribution functions, write $\mathcal{N}_{\mathbf{R}}$ for the multivariate normal distribution with mean the zero vector and $d \times d$ correlation matrix $\mathbf{R}$, and construct $\mathbf{X}$ as

$$\begin{aligned} \mathbf{Z} &= (Z_1, \ldots Z_d) \sim \mathcal{N}_{\mathbf{R}} \\ \mathbf{X} &= (X_1, \ldots, X_d) = \left(F_1^{-1}[\Phi(Z_1)], \ldots, F_d^{-1}([\Phi(Z_d)])\right), \end{aligned} \tag{1}$$

where Φ is the standard normal distribution function (with mean 0 and variance 1) and $F_k^{-1}(u) = \inf\{x : F_k(x) \geq u\}$ for $0 < u < 1$ is the inverse of F_k. By construction, the k-th marginal of $\mathbf{X}$ is F_k. Relative to other multivariate approaches, this model may be appealing by its separating the marginals from the dependence, which is contained in $\mathbf{R}$. The choice of Gaussian copula, while restrictive, facilitates fitting the model and sampling from it.

Consider the case $d = 2$. The construction reduces to selecting the scalar correlation $\rho = \mathrm{Corr}(Z_1, Z_2)$. One approach to specifying ρ is to require that the *rank correlation* between X_1 and X_2, $r(\rho) = r(\rho; F_1, F_2) = \mathrm{Corr}(F_1(X_1), F_2(X_2))$, equals (matches) a target value r^*, which may be the sample rank correlation computed from data (observations of $\mathbf{X}$), or determined otherwise. This leads to the *rank-correlation matching* problem of solving

$$r(\rho; F_1, F_2) = r^*. \tag{2}$$

If F_1 and F_2 are both continuous (meaning absolutely continuous with respect to Lebesgue measure), then the rank-correlation matching problem is resolved by inversion of the formula $r(\rho) = \mathrm{Corr}(\Phi(Z_1), \Phi(Z_2)) = (6/\pi) \arcsin(\rho/2)$ (Kruskal, 1958). An alternative approach seeks ρ so that the product-moment correlation matches a target. Avramidis et al. (2009) have studied the *discrete* problem, where each marginal is discrete. Channouf and L'Ecuyer (2009) have studied the *mixed* problem, where one marginal is discrete and the other one is continuous. Correlation-matching is only one possible route to specifying a model with given marginals. Joe (2005) describes an alternative where marginals and the dependence parameter (of a general copula, not necessarily normal) are estimated in two separate phases, based on maximum-likelihood ideas.

The problem in dimension $d = 2$ is central to Gaussian-copula-based constructions of random vectors in dimension $d > 2$ and the VARTA class of stationary multivariate time series (Biller and Nelson, 2003). In these constructions, a correlation-matching problem is solved for certain pairs of coordinates. In the random-vector construction, a positive semi-definite matrix $\mathbf{R}$ is computed from the solutions of all coordinate pairs (Ghosh and Henderson, 2003). Channouf and L'Ecuyer (2012) use this methodology to model arrival counts in call centers over several periods of a day and find it most effective in fitting the full set of correlations.

Our first contribution is a proof of existence and uniqueness of a solution for any feasible target, without imposing restrictions on the marginals. Intermediate results we obtain are expressions for the derivatives of the mean products $\mathbb{E}[X_1X_2]$ and $\mathbb{E}[F_1(X_1)F_2(X_2)]$ with respect to ρ; and that $r(\rho)$ and $r_{\mathrm{L}}(\rho)$ are differentiable and strictly increasing on $(-1, 1)$.

Our second contribution is to approximate the function $r(\rho)$, with bounds on error, when it cannot be computed exactly due to an infinite (discrete) support, and to bound the error in induced correlation when solutions are computed via the approximation. An X_1 with infinite support gives rise to infinite sums in $\mathbb{E}[F_1(X_1)]$, $\mathrm{Var}[F_1(X_1)]$, and $\mathbb{E}[F_1(X_1)F_2(X_2)]$. In the mean product, a doubly infinite sum arises if, additionally, X_2 is infinite. Avramidis et al. (2009) and Channouf and L'Ecuyer (2009) replace $r(\rho)$ by a version in which the relevant infinite sums are truncated. A simple heuristic is used there: truncate each infinite tail to the right at the quantile x_p associated to a tail probability p (quantile of order $1-p$), resulting in x_p^2 terms in the mean product. In this paper, the relevant infinite sums are approximated by truncation, and lower and upper bounds on the truncation errors are developed. With a function $\tilde{r}(\rho)$ defined by the truncated sums, a bound on the error $r(\rho^*) - r^*$ is given, where ρ^* is a solution to $\tilde{r}(\rho^*) = r^*$ (the solution here is assumed to exist); such a problem can be (and is) solved as in Avramidis et al. (2009); Channouf and L'Ecuyer (2009). A simple algorithm is proposed that determines truncation points so that the solution has any required accuracy. Thus, we enable solving to desired accuracy, which is new.

Our focus on rank correlation is motivated by the fact that nonlinear dependence may be "missed" by product-moment correlation: Embrechts et al. (2002, Example 5) present a sequence of random vectors (X, Y) that are comonotonic (or counter-monotonic), i.e., have a perfect positive (negative) dependence, and such that the product-moment correlation tends to zero; rank correlation, in contrast, captures the dependence. Separately, our bounding method does not apply to product-moment correlation.

Our approach may require less work than the heuristic, especially as the relevant quantile(s) become large. To quantify this point, consider the widely-used discrete Pareto family (Parulekar and Makowski, 1997; Suárez-González et al., 2002; Axtell, 2001; Deuchert and Brody, 2007). Defined for $\alpha > 1$, and supported on the positive integers, the probability mass is $f(k) = k^{-\alpha}/\zeta(\alpha)$, where $\zeta(\alpha) = \sum_{k=1}^{\infty} k^{-\alpha}$ is Riemann's *zeta function*. For $X \sim \mathrm{Pareto}(\alpha)$, a simple calculation gives $x_p \sim \big((\alpha-1)\zeta(\alpha)p\big)^{1/(1-\alpha)}$ as $p \to 0$ ($a_x \sim b_x$ means $a_x/b_x \to 1$). For $\alpha = 2.1$ (finite mean, infinite variance) and $p = 10^{-5}$, we have $x_p = 21488$, and x_p^2 is large; in comparison, our method, with error tolerance 10^{-3} on r^*, requires about 35 thousand terms.

For the discrete problem, the truncation algorithm can be summarized as follows. The total error bound is the sum of a *rightward error bound*, which only depends on rightward truncation point(s), and a *leftward error bound* (that depends on leftward and rightward truncation points). We determine truncation points so that the total error bound is small enough as follows. First, we

iteratively increase candidate rightward truncation points, increasing one of them by one at each iteration, until the rightward error bound is small enough. Then, we decrease candidate leftward truncation points, in similar fashion, until the leftward error bound is small enough. In the mixed problem, the algorithm is similar.

Although there is no simple way to specify equivalent tolerances between the heuristic and our approach, we nevertheless report numerical comparisons to help assess the potential for higher efficiency (work reduction). Comparing the heuristic with $p = 10^{-6}$ to our method with tolerance 10^{-3} for Poisson, negative binomial, and Pareto marginals, we observe a consistent efficiency improvement, and far more pronounced with the Pareto heavier tails.

The remainder is organized as follows. Section 2 develops the existence and uniqueness results. The discrete problem is studied in Section 3; preliminary results have appeared in Avramidis (2009). The algorithm for determining truncation points is detailed in Section 3.5. The mixed problem is studied in Section 4. Numerical results appear in Section 5.

2 General Marginals

We assume throughout the paper that the marginals are non-degenerate. The rank correlation between X_1 and X_2 as in (1) is

$$r(\rho) = \mathrm{Corr}(F_1(X_1), F_2(X_2)) = \frac{g(\rho) - \mu_1\mu_2}{\sigma_1\sigma_2}, \tag{3}$$

where $\mu_k = \mathbb{E}[F_k(X_k)]$; $\sigma_k^2 = \mathrm{Var}[F_k(X_k)]$; and

$$g(\rho) = \mathbb{E}[F_1(X_1)F_2(X_2)] = \int_0^1 \int_0^1 \mathbb{P}(h_1(Z_1) > x, h_2(Z_2) > y)\mathrm{d}x\mathrm{d}y \tag{4}$$

where $h_k = F_k \circ F_k^{-1} \circ \Phi$ (the composite function). The last equality is based on the fact that for any random variables X and Y,

$$\mathbb{E}[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{P}(X > x, Y > y)\mathrm{d}x\mathrm{d}y, \tag{5}$$

provided the expectation is finite (e.g. Lehmann, 1966, Lemma 2).

We now represent g and its derivative with respect to ρ as integrals involving the bivariate normal density. We need a (generalized) inverse of the functions h_k. To this end, let F be a (cumulative) distribution function (c.d.f. in short), let D_F be the set of discontinuity points of F, and put $G(F) = \cup_{x \in D_F}(F(x-), F(x))$; this is the set of u for which there exists no v such that $F(v) = u$, due to discontinuity of F. The inverse of F is $F^{-1}(u) = \inf\{x : F(x) \geq u\}$. We now define the inverse of $F \circ F^{-1}$ as

$$(F \circ F^{-1})^{-1}(u) = \begin{cases} F(F^{-1}(u)-) & u \in G(F) \\ v & \text{otherwise} \end{cases} \tag{6}$$

for $u \in (0,1)$, where $F(F^{-1}(u)-)$ is the left limit of F at $F^{-1}(u)$. A special case of (6) that we need later has a discrete F with cumulative probabilities $0 = f_0 < f_1 < \ldots$; then $(F \circ F^{-1})^{-1}(u) = f_{i-1}$ whenever $u \in (f_{i-1}, f_i]$. Now define $h_k^{-1} = \Phi^{-1} \circ (F_k \circ F_k^{-1})^{-1}$, where Φ^{-1} is the inverse of Φ. For any F, one may verify that

$$F \circ F^{-1}(v) > u \Longleftrightarrow v > (F \circ F^{-1})^{-1}(u) \tag{7}$$

and thus (4) gives

$$g(\rho) = \int_0^1 \int_0^1 \mathbb{P}(Z_1 > h_1^{-1}(x), Z_2 > h_2^{-1}(y))\mathrm{d}x\mathrm{d}y = \int_0^1 \int_0^1 \bar{\Phi}_\rho(h_1^{-1}(x), h_2^{-1}(y))\mathrm{d}x\mathrm{d}y, \tag{8}$$

where $\bar{\Phi}_\rho(x,y) = \int_x^\infty \int_y^\infty \phi_\rho(z,w)\mathrm{d}z\mathrm{d}w$, where $\phi_\rho(x,y)$ is the density at (x,y) of the bivariate standard normal distribution with correlation ρ. Certain invariance properties of rank correlation can now be seen. First, if F_k is continuous, then $h_k^{-1}() = \Phi^{-1}()$, and then (8) shows that the dependence on F_k disappears. In particular, for F_1 discrete and F_2 continuous, $r(\rho; F_1, F_2)$ is a function of ρ and F_1 only. Second, the locations of any discontinuity points do not matter (to g and r)—only the *values* attained by the c.d.f. do. This property is inherited from the inverse $(F \circ F^{-1})^{-1}$ defined above. Differentiation of (8) gives

$$\frac{\mathrm{d}}{\mathrm{d}\rho} g(\rho) = \int_0^1 \int_0^1 \frac{\mathrm{d}}{\mathrm{d}\rho} \bar{\Phi}_\rho(h_1^{-1}(x), h_2^{-1}(y))\mathrm{d}x\mathrm{d}y = \int_0^1 \int_0^1 \phi_\rho(h_1^{-1}(x), h_2^{-1}(y))\mathrm{d}x\mathrm{d}y, \quad \rho \in (-1,1). \tag{9}$$

The derivative can pass inside the integral by an argument as in the proof of Theorem 9.42 in Rudin (1976) on noting that $\phi_\rho(h_1^{-1}(x), h_2^{-1}(y))$ has a bounded gradient with respect to (ρ, x, y) almost everywhere on $(-1,1) \times \mathbb{R}^2$. We then use that $(\mathrm{d}/\mathrm{d}\rho)\bar{\Phi}_\rho(x,y) = \phi_\rho(x,y)$ (e.g. Avramidis et al., 2009, eq. (13)).

An analogous development for the product-moment correlation follows: $r_{\mathrm{L}}(\rho) = \mathrm{Corr}(X_1, X_2) = (g_{\mathrm{L}}(\rho) - \mathbb{E}[X_1]\mathbb{E}[X_2])/\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}$, where, using (5),

$$\begin{aligned} g_{\mathrm{L}}(\rho) = \mathbb{E}[X_1 X_2] &= \int_{-\infty}^\infty \int_{-\infty}^\infty \mathbb{P}(F_1^{-1}(\Phi(Z_1)) > x, F_2^{-1}(\Phi(Z_2)) > y)\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^\infty \int_{-\infty}^\infty \mathbb{P}(Z_1 > \Phi^{-1}(F_1(x)), Z_2 > \Phi^{-1}(F_2(y)))\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^\infty \int_{-\infty}^\infty \bar{\Phi}_\rho(\Phi^{-1}(F_1(x)), \Phi^{-1}(F_2(y)))\mathrm{d}x\mathrm{d}y. \end{aligned} \tag{10}$$

We used above the equivalence $F^{-1}(u) > x \Longleftrightarrow u > F(x)$, valid for any c.d.f. F and $u \in (0,1)$ (Asmussen and Glynn, 2007, Proposition 2.2(a), page 38). Differentiation of (10) gives

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\rho} g_{\mathrm{L}}(\rho) &= \int_{-\infty}^\infty \int_{-\infty}^\infty \frac{\mathrm{d}}{\mathrm{d}\rho} \bar{\Phi}_\rho((\Phi^{-1}(F_1(x)), \Phi^{-1}(F_2(y)))\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^\infty \int_{-\infty}^\infty \phi_\rho(\Phi^{-1}(F_1(x)), \Phi^{-1}(F_2(y)))\mathrm{d}x\mathrm{d}y, \quad \rho \in (-1,1). \end{aligned} \tag{11}$$

The derivative can pass inside the integral because $\phi_\rho(\Phi^{-1}(F_1(x)), \Phi^{-1}(F_2(y)))$ has a bounded gradient with respect to (ρ, x, y) almost everywhere on $(-1, 1) \times \mathbb{R}^2$. Thus:

Proposition 1 *Let F_1 and F_2 be c.d.f.'s of non-degenerate distributions. Put $X_k = F_k^{-1}(\Phi(Z_k))$ for $k = 1, 2$, where (Z_1, Z_2) is bivariate normal with standard-normal marginals and correlation ρ. The function $g(\rho)$ in (4) has derivative (9), and the function $g_{\mathrm{L}}(\rho)$ in (10) has derivative (11).*

The differentiability of the function g implies its continuity, and the same properties hold for function r. Moreover, continuity at the endpoints, -1 and 1, follows by Cario and Nelson (1996, Theorem 2). Then, the Intermediate Value Theorem implies that for any $r^* \in [r(-1), r(1)]$, the equation $r(\rho) = r^*$ has a solution. Likewise, the differentiability of g_{L} on $(-1, 1)$ implies that for any $\delta > 0$, g_{L} and r_{L} are continuous on $[-1 + \delta, 1 - \delta]$ (continuity at -1 and 1 may require extra conditions), proving the solution's existence for $r^* \in [r_{\mathrm{L}}(-1 + \delta), r_{\mathrm{L}}(1 - \delta)]$. Moreover, the solution in each case is unique, as each of r and r_{L} is strictly increasing in ρ (the integrand in each of (9) and (11) is positive on a set of positive Lebesgue measure and non-negative everywhere). Thus, the existence and uniqueness results in Cario and Nelson (1996, Theorem 1) and Avramidis et al. (2009) extend, without any restrictions on the marginals:

Corollary 1 *The functions g and g_{L} are differentiable and strictly increasing everywhere. For any $r^* \in [r(-1), r(1)]$, the equation $r(\rho) = r^*$ has a unique solution. For any $\delta > 0$ and for $r^* \in [r_{\mathrm{L}}(-1 + \delta), r_{\mathrm{L}}(1 - \delta)]$, the equation $r_{\mathrm{L}}(\rho) = r^*$ has a unique solution.*

3 The Discrete Problem

3.1 Preliminaries

The *discrete rank-correlation-matching problem* refers to solving $r(\rho) = r^*$ where the marginals F_1 and F_2 are both discrete. For simplicity, we assume the marginals have an infinite tail to the right only. We enumerate the support points after putting them in increasing order as $\{0, 1, 2, \ldots\}$. For the k-th marginal, $p_{k,i}$ denotes the probability mass at i; we put $f_{k,i} = \sum_{j=0}^{i} p_{k,j}$ and $f_{k,-1} = 0$. As stated following (6), $(F_k \circ F_k^{-1})^{-1}(v) = f_{k,i-1}$ for all $v \in (f_{k,i-1}, f_{k,i}]$. Then (8) gives

$$g(\rho) = \sum_{i=0}^{\infty} p_{1,i} \sum_{j=0}^{\infty} p_{2,j} \bar{\Phi}_\rho(z_{1,i-1}, z_{2,j-1}), \tag{12}$$

where $z_{k,i} = \Phi^{-1}(f_{k,i})$ and $z_{k,-1} = -\infty$. This is equation (10) of Avramidis et al. (2009), seen here to be a special case of (8).

3.2 Approximation of the Mean and the Variance

The task is to approximate the means and variances in (3). To lighten notation, we work with a single marginal and later apply the forthcoming results to each marginal. Denote p_i the probability

mass at i and $f_i = \sum_{j=0}^{i} p_j$ the cumulative probability at i. We will approximate the mean $\mu = \mathbb{E}[F(X)] = \sum_{i=0}^{\infty} f_i p_i$ and the variance $\sigma^2 = \sum_{i=0}^{\infty} f_i^2 p_i - \mu^2$. Note that $\mu < 1$ and $\sigma^2 > 0$, by non-degeneracy. The approximation is via the corresponding exact moments of the finite-support random variable, X_n, obtained by shifting to the point $n+1$ the probability mass of all the points to its right, so that the resulting mass at $n+1$ is the tail probability $t_n = 1 - f_n = \sum_{i>n} p_i$. With F_n denoting the c.d.f. of X_n, the approximate mean is

$$\tilde{\mu}_n = \mathbb{E}[F_n(X_n)] = \sum_{i=0}^{n} f_i p_i + 1 - f_n$$

and the approximate variance is

$$\tilde{\sigma}_n^2 = \mathrm{Var}[F_n(X_n)] = \tilde{\mu}_n^{(2)} - \tilde{\mu}_n^2, \tag{13}$$

where $\tilde{\mu}_n^{(2)} = \mathbb{E}[F_n^2(X_n)] = \sum_{i=0}^{n} f_i^2 p_i + 1 - f_n$.

We now derive sequences that bound μ and σ^2 from below and above and that converge to these targets in each case. We define $x^+ = \max(x, 0)$.

Lemma 1 *(i) (Sequences bounding μ below and above and converging to it.) Define $\underline{\mu}_n = (\tilde{\mu}_n - t_n t_{n+1})^+$. We have*

$$\underline{\mu}_n \le \mu \le \tilde{\mu}_n \quad \textit{for all } n, \tag{14}$$

and $\tilde{\mu}_n \downarrow \mu$ and $\underline{\mu}_n \to \mu$ as $n \to \infty$.

(ii) (Sequences bounding σ^2 below and above and converging to it.) Define

$$\underline{\sigma}_n^2 = \tilde{\sigma}_n^2 - 2(1 - \underline{\mu}_n) t_n t_{n+1} \quad \textit{and} \quad \bar{\sigma}_n^2 = \begin{cases} \tilde{\sigma}_n^2 - l_n & n < n^* \\ \tilde{\sigma}_n^2 & n \ge n^* \end{cases}$$

where $l_n = (1 + f_n - \tilde{\mu}_{n-1} - \tilde{\mu}_n) t_n t_{n+1}$ and $n^* = \min\{n : 1 + f_n - \tilde{\mu}_{n-1} - \tilde{\mu}_n > 0\} < \infty$. We have

$$\underline{\sigma}_n^2 \le \sigma^2 \le \bar{\sigma}_n^2 \quad \textit{for all } n, \tag{15}$$

and $\{\tilde{\sigma}_n^2\}_{n=n^*}^{\infty} \downarrow \sigma^2$ and $\underline{\sigma}_n^2 \to \sigma^2$ as $n \to \infty$.

Proof. Part (i). Write

$$\tilde{\mu}_{i-1} - \tilde{\mu}_i = \sum_{k \le i-1} f_k p_k + 1 - f_{i-1} - \sum_{k \le i} f_k p_k - (1 - f_i) = p_i(1 - f_i) > 0 \tag{16}$$

and $\tilde{\mu}_n - \mu = \sum_{i>n} (\tilde{\mu}_{i-1} - \tilde{\mu}_i)$, by the nonnegativity of the summands. Thus

$$0 < \tilde{\mu}_n - \mu = \sum_{i>n} (\tilde{\mu}_{i-1} - \tilde{\mu}_i) = \sum_{i>n} p_i(1 - f_i) \le (1 - f_{n+1})(1 - f_n) = t_{n+1} t_n. \tag{17}$$

The assertion $\lim \underline{\mu}_n = \mu$ follows from $\lim_{n\to\infty} t_n t_{n+1} = 0$.

Part (ii). We have

$$\begin{aligned}\tilde{\sigma}_{i-1}^2 - \tilde{\sigma}_i^2 &= \tilde{\mu}_{i-1}^{(2)} - \tilde{\mu}_i^{(2)} - (\tilde{\mu}_{i-1}^2 - \tilde{\mu}_i^2) \\ &= \sum_{k\le i-1} f_k^2 p_k + 1 - f_{i-1} - \sum_{k\le i} f_k^2 p_k - (1-f_i) - (\tilde{\mu}_{i-1} - \tilde{\mu}_i)(\tilde{\mu}_{i-1} + \tilde{\mu}_i) \\ &= p_i(1-f_i^2) - p_i(1-f_i)(\tilde{\mu}_{i-1} + \tilde{\mu}_i) \\ &= p_i(1-f_i)(1 + f_i - \tilde{\mu}_{i-1} - \tilde{\mu}_i) \qquad (18)\end{aligned}$$

by using (16) in the third step. The above is nonnegative for all i large enough, so

$$\tilde{\sigma}_n^2 - \sigma^2 = \sum_{i>n} (\tilde{\sigma}_{i-1}^2 - \tilde{\sigma}_i^2). \qquad (19)$$

We claim that

$$2(1-\underline{\mu}_n)t_n t_{n+1} \ge \sum_{i>n} p_i(1-f_i)(1+f_i-\tilde{\mu}_{i-1}-\tilde{\mu}_i) \ge \begin{cases} (1+f_n-\tilde{\mu}_{n-1}-\tilde{\mu}_n)t_n t_{n+1}, & n < n^* \\ 0, & n \ge n^* \end{cases} \qquad (20)$$

and observe that the quantity in the middle is $\tilde{\sigma}_n^2 - \sigma^2$; then a simple rearrangement will prove (15). It remains to prove (20). Note that the sequence $\{1 + f_i - \tilde{\mu}_{i-1} - \tilde{\mu}_i\}_{i=1}^{\infty}$ is monotonically increasing to $2(1-\mu)$ (since $\{f_i\}_{i=0}^{\infty} \uparrow 1$ and $\{\tilde{\mu}_i\}_{i=0}^{\infty} \downarrow \mu$), so

$$2(1-\mu)\sum_{i>n} p_i(1-f_i) \ge \sum_{i>n}(1+f_i-\tilde{\mu}_{i-1}-\tilde{\mu}_i)p_i(1-f_i) \ge (1+f_n-\tilde{\mu}_{n-1}-\tilde{\mu}_n)\sum_{i>n} p_i(1-f_i). \qquad (21)$$

In the above, we may substitute looser bounds, as follows. The upper bound (left side) is positive, so we may substitute for $\sum_{i>n} p_i(1-f_i)$ and $1-\mu$ the respective upper bounds $t_n t_{n+1}$ and $1-\underline{\mu}_n$. The lower bound (right side) is negative (positive) when $n < n^*$ ($n \ge n^*$) respectively; in the negative case, we may substitute for $\sum_{i>n} p_i(1-f_i)$ the upper bound $t_n t_{n+1}$; in the positive case, we may substitute zero. These substitutions give (20), and this completes the proof of (15). The assertion $\{\tilde{\sigma}_n^2\}_{n=n^*}^{\infty} \downarrow \sigma^2$ holds on noting that the sequence $\{1 + f_i - \tilde{\mu}_{i-1} - \tilde{\mu}_i\}_{i=0}^{\infty}$ is monotonically increasing and its n^*-th term is positive, so each summand in (19) is positive for $n \ge n^*$. The assertion $\lim \underline{\sigma}_n^2 = \sigma^2$ follows from $\lim_{n\to\infty} t_n t_{n+1} = 0$. □

Results (14) and (15) hold at any n, so truncation of a finite support (a special case of an infinite one) is also covered. In view of $\lim \underline{\sigma}_n^2 = \sigma^2 > 0$, we may define for n large enough the real number $\underline{\sigma}_n = \sqrt{\underline{\sigma}_n^2}$.

3.3 Approximation of the Mean Product

For a vector $\mathbf{n} = (l_1, r_1, l_2, r_2)$, define the approximation $g_{\mathbf{n}}(\rho)$ of $g(\rho)$ as the right side of (12) truncated so the range of i is restricted to $l_1 \le i \le r_1$ and the range of j is restricted to $l_2 \le j \le r_2$.

Lemma 2 *We have*

$$0 \le g(\rho) - g_{\mathbf{n}}(\rho) \le \sum_{k=1}^{2} (f_{k,l_k-1} + t_{k,r_k}^2) \quad \textit{for all } \rho. \tag{22}$$

Proof. By the non-negativity of each summand in (12), we have, for any ρ,

$$\begin{aligned} 0 \le g(\rho) - g_{\mathbf{n}}(\rho) \quad \le \quad & \sum_{i<l_1} p_{1,i} \sum_{j=0}^{\infty} p_{2,j} \bar{\Phi}_\rho(z_{1,i-1}, z_{2,j-1}) + \sum_{i>r_1} p_{1,i} \sum_{j=0}^{\infty} p_{2,j} \bar{\Phi}_\rho(z_{1,i-1}, z_{2,j-1}) \\ & + \sum_{j<l_2} p_{2,j} \sum_{i=0}^{\infty} p_{1,i} \bar{\Phi}_\rho(z_{1,i-1}, z_{2,j-1}) + \sum_{j>r_2} p_{2,j} \sum_{i=0}^{\infty} p_{1,i} \bar{\Phi}_\rho(z_{1,i-1}, z_{2,j-1}). \end{aligned} \tag{23}$$

Since $\bar{\Phi}_\rho(x, y)$ is non-decreasing in ρ, we have

$$\bar{\Phi}_\rho(x, y) \le \bar{\Phi}_1(x, y) = \bar{\Phi}(\max(x, y)) = \min(\bar{\Phi}(x), \bar{\Phi}(y)) \quad \text{for all } \rho, \tag{24}$$

where $\bar{\Phi} = 1 - \Phi$ is the standard univariate normal complementary c.d.f.. Using this, an upper bound for the first of the four terms on the right in (23) is

$$\sum_{i<l_1} p_{1,i} \bar{\Phi}(z_{1,i-1}) \sum_{j=0}^{\infty} p_{2,j} = \sum_{i<l_1} p_{1,i} t_{1,i-1} \le \sum_{i<l_1} p_{1,i} = f_{1,l_1-1} \tag{25}$$

upon noting that $\bar{\Phi}(z_{1,i-1}) = t_{1,i-1}$ and $\sum_{j=0}^{\infty} p_{2,j} = 1$; and an upper bound for the second term on the right of (23) is

$$\sum_{i>r_1} p_{1,i} \bar{\Phi}(z_{1,i-1}) \sum_{j=0}^{\infty} p_{2,j} = \sum_{i>r_1} p_{1,i} t_{1,i-1} \le t_{1,r_1} t_{1,r_1+1}. \tag{26}$$

The bounds (25) and (26) and their analogs for the third and fourth term in (23) give (22). □

3.4 Approximation of the Rank Correlation

For $k \in \{1, 2\}$, and for the purpose of approximating μ_k and σ_k, we truncate marginal k to the right of r_k, as described in Section 3.2. We will approximate the function $r(\rho)$ as $\tilde{r}_{\mathbf{n}}(\rho) = (g_{\mathbf{n}}(\rho) - \tilde{\mu}_{1,r_1}\tilde{\mu}_{2,r_2})/(\tilde{\sigma}_{1,r_1}\tilde{\sigma}_{2,r_2})$, where $\mathbf{n} = (l_1, r_1, l_2, r_2)$ gives the truncation of the sum about $g_{\mathbf{n}}$, as in Section 3.3, and the sums μ_k and σ_k are truncated to the right only. Left-truncation of these sums would complicate the error analysis while having little impact on computing cost.

Observe that $\tilde{r}_{\mathbf{n}}$ is a continuous strictly increasing function on $[-1, 1]$, and thus has an inverse; that is, for $r^* \in [\tilde{r}_{\mathbf{n}}(-1), \tilde{r}_{\mathbf{n}}(1)]$, there exists a unique ρ such that $\tilde{r}_{\mathbf{n}}(\rho) = r^*$, which we denote $\tilde{r}_{\mathbf{n}}^{-1}(r^*)$. This follows immediately from Corollary 1 by observing that $g_{\mathbf{n}}$ is the g in (12) corresponding to the finite support that results when for each $k \in \{1, 2\}$ we shift to the point r_k the probability mass of the points to its right and we shift to the point l_k the probability mass of the points to its left. Our main result is as follows.

Proposition 2 *Let $\rho^* = \tilde{r}_{\mathbf{n}}^{-1}(r^*)$, where $r^* \in [\tilde{r}_{\mathbf{n}}(-1), \tilde{r}_{\mathbf{n}}(1)]$. Provided that $\underline{\sigma}_1^2$ and $\underline{\sigma}_2^2$ are positive, we have*

$$\zeta_{\mathbf{n}} \le r(\rho^*) - r^* \le \eta_{\mathbf{n}} + \theta_{\mathbf{n}} \quad \textit{for all } n, \tag{27}$$

where

$$\zeta_{\mathbf{n}} = \begin{cases} r^* \left(\frac{\tilde{\sigma}_{1,r_1}\tilde{\sigma}_{2,r_2}}{\bar{\sigma}_{1,r_1}\bar{\sigma}_{2,r_2}} - 1 \right), & r^* > 0 \\ r^* \left(\frac{\tilde{\sigma}_{1,r_1}\tilde{\sigma}_{2,r_2}}{\underline{\sigma}_{1,r_1}\underline{\sigma}_{2,r_2}} - 1 \right), & r^* < 0, \end{cases} \qquad \eta_{\mathbf{n}} = \frac{f_{1,l_1-1} + f_{2,l_2-1}}{\underline{\sigma}_{1,r_1}\underline{\sigma}_{2,r_2}},$$

and

$$\theta_{\mathbf{n}} = \begin{cases} \frac{t_{1,r_1}^2 + t_{2,r_2}^2 + \tilde{\mu}_{1,r_1}\tilde{\mu}_{2,r_2} - \underline{\mu}_{1,r_1}\underline{\mu}_{2,r_2}}{\underline{\sigma}_{1,r_1}\underline{\sigma}_{2,r_2}} + r^* \left(\frac{\tilde{\sigma}_{1,r_1}\tilde{\sigma}_{2,r_2}}{\underline{\sigma}_{1,r_1}\underline{\sigma}_{2,r_2}} - 1 \right), & r^* > 0 \\ \frac{t_{1,r_1}^2 + t_{2,r_2}^2 + \tilde{\mu}_{1,r_1}\tilde{\mu}_{2,r_2} - \underline{\mu}_{1,r_1}\underline{\mu}_{2,r_2}}{\underline{\sigma}_{1,r_1}\underline{\sigma}_{2,r_2}} + r^* \left(\frac{\tilde{\sigma}_{1,r_1}\tilde{\sigma}_{2,r_2}}{\bar{\sigma}_{1,r_1}\bar{\sigma}_{2,r_2}} - 1 \right), & r^* < 0. \end{cases}$$

Proof. Putting $\tilde{h}_{\mathbf{n}}(\rho) = g_{\mathbf{n}}(\rho) - \tilde{\mu}_{1,r_1}\tilde{\mu}_{2,r_2} - r^*\tilde{\sigma}_{1,r_1}\tilde{\sigma}_{2,r_2}$, we have $\tilde{h}_{\mathbf{n}}(\rho^*) = 0$ and

$$r(\rho^*) - r^* = \frac{g(\rho^*) - g_{\mathbf{n}}(\rho^*) + \tilde{h}_{\mathbf{n}}(\rho^*) + \tilde{\mu}_{1,r_1}\tilde{\mu}_{2,r_2} - \mu_1\mu_2 + r^*(\tilde{\sigma}_{1,r_1}\tilde{\sigma}_{2,r_2} - \sigma_1\sigma_2)}{\sigma_1\sigma_2}. \tag{28}$$

Now (27) follows from the bounds on $g(\rho) - g_{\mathbf{n}}(\rho)$ in (22); the bounds on μ_k as in (14); and the bounds on σ_k^2 as in (15). □

We observe, even though we never use, that the set $\{\rho : \mathrm{sign}(\rho) \neq \mathrm{sign}(\tilde{r}_{\mathbf{n}}(\rho))\}$ is the interval between zero and $\tilde{r}_{\mathbf{n}}^{-1}(0)$, and that this interval can be made arbitrarily small: for $l_1 = l_2 = 0$, we can show that $\lim_{r_1,r_2\to\infty} \tilde{r}_{\mathbf{n}}^{-1}(0) = 0$.

Note that $\zeta_{\mathbf{n}} \le 0$ and $\eta_{\mathbf{n}}, \theta_{\mathbf{n}} \ge 0$. We now derive asymptotic relations about the error bounds under the assumption $l_1 = l_2 = 0$ and $r_1, r_2 \to \infty$. To lighten the notation here, put $\tilde{\mu}_k = \tilde{\mu}_{k,r_k}$, $\underline{\mu}_k = \underline{\mu}_{k,r_k}$, $\tilde{\sigma}_k = \tilde{\sigma}_{k,r_k}$, $\underline{\sigma}_k = \underline{\sigma}_{k,r_k}$, $\bar{\sigma}_k = \bar{\sigma}_{k,r_k}$. Now observe: (a) for all r_k large enough, we have $\underline{\mu}_k = \tilde{\mu}_k - t_{k,r_k}t_{k,r_k+1}$, so $\tilde{\mu}_1\tilde{\mu}_2 - \underline{\mu}_1\underline{\mu}_2 = \mu_2 t_{1,r_1}^2 + \mu_1 t_{2,r_2}^2 + o(t_{1,r_1}^2 + t_{2,r_2}^2)$, where $o(\cdot)$ has the usual meaning; (b) from $\underline{\sigma}_k = \tilde{\sigma}_k\sqrt{1 - 2(1-\underline{\mu}_k)t_{k,r_k}t_{k,r_k+1}/\tilde{\sigma}_k^2}$ and the Taylor expansion $\sqrt{1-x} = 1 - x/2 + o(x)$ as $x \downarrow 0$, we obtain $\underline{\sigma}_k = \tilde{\sigma}_k[1 - (1-\mu_k)t_{k,r_k}^2/\sigma_k^2] + o(t_{k,r_k}^2)$; then a simple calculation gives $\tilde{\sigma}_1\tilde{\sigma}_2/\underline{\sigma}_1\underline{\sigma}_2 - 1 = a_1 t_{1,r_1}^2 + a_2 t_{2,r_2}^2 + o(t_{1,r_1}^2 + t_{2,r_2}^2)$, where $a_k = (1-\mu_k)/\sigma_k^2$; and (c) $\tilde{\sigma}_1\tilde{\sigma}_2/(\bar{\sigma}_1\bar{\sigma}_2) = 1$, provided $r_k \ge n_k^*$, the n_k^* being as in Lemma 1. From these observations, we obtain: (i) when $r^* < 0$, we have $\zeta_{\mathbf{n}} = r^*(a_1 t_{1,r_1}^2 + a_2 t_{2,r_2}^2) + o(t_{1,r_1}^2 + t_{2,r_2}^2)$; when $r^* > 0$, we have $\zeta_{\mathbf{n}} = 0$, provided $r_k \ge n_k^*$; and (ii) $\theta_{\mathbf{n}} = b_1 t_{1,r_1}^2 + b_2 t_{2,r_2}^2 + o(t_{1,r_1}^2 + t_{2,r_2}^2)$, where we define $b_k = (1 + \mu_{3-k})/(\sigma_1\sigma_2) + r^* a_k$ for $r^* > 0$ and $b_k = (1 + \mu_{3-k})/(\sigma_1\sigma_2)$ for $r^* < 0$.

Remark 1 Write c for the right side of (22). It is not difficult to see that

$$\left.\begin{aligned} \frac{g_{\mathbf{n}}(\rho) - \tilde{\mu}_{1,r_1}\tilde{\mu}_{2,r_2}}{\bar{\sigma}_{1,r_1}\bar{\sigma}_{2,r_2}} \le \frac{g(\rho) - \mu_1\mu_2}{\sigma_1\sigma_2} \le \frac{g_{\mathbf{n}}(\rho) + c - \underline{\mu}_{1,r_1}\underline{\mu}_{2,r_2}}{\underline{\sigma}_{1,r_1}\underline{\sigma}_{2,r_2}}, \quad \rho > 0 \\ \frac{g_{\mathbf{n}}(\rho) - \tilde{\mu}_{1,r_1}\tilde{\mu}_{2,r_2}}{\underline{\sigma}_{1,r_1}\underline{\sigma}_{2,r_2}} \le \frac{g(\rho) - \mu_1\mu_2}{\sigma_1\sigma_2} \le \frac{g_{\mathbf{n}}(\rho) + c - \underline{\mu}_{1,r_1}\underline{\mu}_{2,r_2}}{\bar{\sigma}_{1,r_1}\bar{\sigma}_{2,r_2}}, \quad \rho < 0 \end{aligned}\right\} \tag{29}$$

The distance between the lower and upper bounds above converges to zero when $l_1 = l_2 = 0$ and $r_1, r_2 \to \infty$ (by results (a) to (c) after Proposition 2). Thus, (29) may be used to compute $r(\rho)$ for any ρ, including the extreme correlations $r(-1)$ and $r(1)$, to any desired accuracy.

3.5 Truncation Algorithm

The work to compute the root of $\tilde{r}_{\mathbf{n}}(\rho) = r^*$ can be expected to be *roughly* linear in $w = (r_1 - l_1 + 1)(r_2 - l_2 + 1)$. This is because w bivariate normal integrals are involved in evaluating $g_{\mathbf{n}}(\rho)$ at any candidate; if derivatives are to be used (Avramidis et al., 2009), then w derivatives, one for each term of (12), are involved at any candidate; and empirical results in Avramidis et al. (2009) are consistent with our claim. Then, accuracy and efficiency considerations suggest that $\mathbf{n}$ be chosen to minimize w subject to the error bounds in (27) being within given limits.

Rather than solving such a minimization problem exactly, we propose Algorithm 1 below. This is an approximate algorithm, and it is designed for simplicity and not efficiency because the ensuing root-finding work is far more demanding, as numerical results will show. First we reduce the quantity $\max(-\zeta_{\mathbf{n}}, \theta_{\mathbf{n}})$ —called the *rightward* error bound, as it only depends on the r_k—as follows: we initialize r_1 and r_2 as the smallest support point, 0 and iteratively increase r_1 or r_2 by one, choosing for simplicity the one that corresponds to the larger tail probability to the right, t_{k,r_k}, until both lower bounds on variance ($\underline{\sigma}^2_{k,r_k}$) are positive and the rightward error bound is no larger than δ_{r}, where $\delta_{\mathrm{r}} > 0$ is a specified tolerance. Having determined r_1 and r_2, we then reduce the quantity $\eta_{\mathbf{n}}$ —called the *leftward* error bound because the r_k have been fixed—as follows: we initialize l_k as the r_k ($k = 1, 2$) determined in phase one, and iteratively decrease l_1 or l_2 by one, choosing the one that corresponds to the larger probability to the left, f_{k,l_k-1}, until the leftward error bound is no larger than δ_{l}, where $\delta_{\mathrm{l}} \geq 0$ is a specified tolerance. The output is a truncation $\mathbf{n} = (l_1, r_1, l_2, r_2)$ and the numbers $\zeta_{\mathbf{n}}$, $\eta_{\mathbf{n}}$, and $\theta_{\mathbf{n}}$. For any $\delta_{\mathrm{r}} > 0$ and $\delta_{\mathrm{l}} > 0$, there exist finite r_k as required; then $\underline{\sigma}_{1,r_1}\underline{\sigma}_{2,r_2} > 0$, and there exist finite l_k as required. The solution to $r_{\mathbf{n}}(\rho) = r^*$ (to be computed elsewhere) satisfies (27) and in particular $-\delta_{\mathrm{r}} \leq r(\rho) - r^* \leq \delta_{\mathrm{r}} + \delta_{\mathrm{l}}$. By $\delta_{\mathrm{l}} = 0$ we will mean no leftward truncation, i.e., $l_1 = l_2 = 0$.

To see where Algorithm 1 truncates, consider the case $l_1 = l_2 = 0$ and some $\delta_{\mathrm{r}} > 0$. Suppose $r^* > 0$. In the limit as $\delta_{\mathrm{r}} \to 0$, the only requirement is $\theta_{\mathbf{n}} \leq \delta_{\mathrm{r}}$, and we have seen that $\theta_{\mathbf{n}} \sim b_1 t^2_{1,r_1} + b_2 t^2_{2,r_2}$. with b_k defined in point (ii) following Proposition 2. It is easy to see that $t_{1,r_1} \sim t_{2,r_2} \sim \sqrt{\delta_{\mathrm{r}}/(b_1 + b_2)}$, so r_k is simply the quantile of F_k corresponding to this tail probability. The case $r^* < 0$ gives similar behavior.

4 The Mixed Problem

The *mixed correlation-matching problem* refers to solving $r(\rho) = r^*$ where F_1 is discrete and F_2 is continuous. This indexing involves no loss of generality. The discrete support points are $0, 1, 2, \ldots$;

Algorithm 1: Truncate

Input: Probability masses $\{p_{k,i}\}_{i=0}^{\infty}$ for $k = 1, 2$; target r^*; tolerances $\delta_{\mathrm{l}} \geq 0$ and $\delta_{\mathrm{r}} > 0$

. **Output**: Vector $\mathbf{n} = (l_1, r_1, l_2, r_2)$; error-bound components $\zeta_{\mathbf{n}}$, $\eta_{\mathbf{n}}$, and $\theta_{\mathbf{n}}$.

$r_1 \leftarrow 0$; $r_2 \leftarrow 0$; $\theta_{\mathbf{n}} \leftarrow \infty$; $\zeta_{\mathbf{n}} \leftarrow -\infty$ `/* Phase 1, rightward truncation */`

while $\max(-\zeta_{\mathbf{n}}, \theta_{\mathbf{n}}) > \delta_{\mathrm{r}}$ **do**

  if $t_{1,r_1} > t_{2,r_2}$ **then**

    $r_1 \leftarrow r_1 + 1$

    Update f_{1,r_1}, t_{1,r_1}, $\tilde{\mu}_{1,r_1}$, $\underline{\mu}_{1,r_1}$, $\tilde{\sigma}^2_{1,r_1}$, $\underline{\sigma}^2_{1,r_1}$ and $\bar{\sigma}^2_{1,r_1}$

    if $\underline{\sigma}^2_{1,r_1} \leq 0$ **then**

      continue while

    end

  else

    $r_2 \leftarrow r_2 + 1$

    Update f_{2,r_2}, t_{2,r_2}, $\tilde{\mu}_{2,r_2}$, $\underline{\mu}_{2,r_2}$, $\tilde{\sigma}^2_{2,r_2}$, $\underline{\sigma}^2_{2,r_2}$ and $\bar{\sigma}^2_{2,r_2}$

    if $\underline{\sigma}^2_{2,r_2} \leq 0$ **then**

      continue while

    end

  end

  Update $\zeta_{\mathbf{n}}$ and $\theta_{\mathbf{n}}$

end

$l_1 \leftarrow r_1$; $l_2 \leftarrow r_2$; $\epsilon \leftarrow f_{1,l_1} - p_{1,l_1}$; $\epsilon' \leftarrow f_{2,l_2} - p_{2,l_2}$ `/* Phase 2, leftward truncation */`

while $\epsilon + \epsilon' > \underline{\sigma}_{1,r_1}\underline{\sigma}_{2,r_2}\delta_{\mathrm{l}}$ **do**

  if $\epsilon > \epsilon'$ **then**

    $l_1 \leftarrow l_1 - 1$; $\epsilon \leftarrow \epsilon - p_{1,l_1}$

  else

    $l_2 \leftarrow l_2 - 1$; $\epsilon' \leftarrow \epsilon' - p_{2,l_2}$

  end

end

$\eta_{\mathbf{n}} \leftarrow (\epsilon + \epsilon')/(\underline{\sigma}_{1,r_1}\underline{\sigma}_{2,r_2})$

p_i is the probability mass at i; and $f_i = \sum_{j=0}^{i} p_j$. The continuity of F_2 means that $F_2(X_2)$ is uniformly distributed on (0,1), so its mean is $\mu_2 = 1/2$ and its variance is $\sigma_2^2 = 1/12$.

The general expression (8) of g would lead to bivariate normal integrals. A more convenient expression is

$$g(\rho) = \sum_{i=0}^{\infty} f_i \int_0^1 u[\Phi(i,u) - \Phi(i-1,u)]\mathrm{d}u, \tag{30}$$

where $\Phi(i,u) = \Phi(i,u,\rho) = \Phi\left((z_i - \rho\Phi^{-1}(u))/\sqrt{1-\rho^2}\right)$, with $z_i = \Phi^{-1}(f_i)$ and $z_{-1} = -\infty$. This is not difficult to see, and agrees with equation (9) in Channouf and L'Ecuyer (2009), except that the support there is unbounded in both directions.

We will develop an approximation of (30) and associated error bounds. We will then develop an approximation of $r(\rho)$ and error bounds in analogy to the discrete problem. Put $\bar{\Phi}(i,u) = 1-\Phi(i,u)$ and $f_{-1} = 0$. Rewrite (30) as $g(\rho) = \int_0^1 I(u,\rho)\mathrm{d}u$, where

$$\begin{aligned} I(u,\rho) &= u\sum_{i=0}^{\infty} f_i[\Phi(i,u) - \Phi(i-1,u)] = u\sum_{i=0}^{\infty} f_i[\bar{\Phi}(i-1,u) - \bar{\Phi}(i,u)] \\ &= u\sum_{i=0}^{\infty} \bar{\Phi}(i-1,u)(f_i - f_{i-1}) = u\sum_{i=0}^{\infty} \bar{\Phi}(i-1,u)p_i. \end{aligned} \tag{31}$$

For an integer n, we truncate the sum expression of the integrand I to obtain

$$I_n(u,\rho) = u\sum_{i=0}^{n} \bar{\Phi}(i-1,u,\rho)p_i, \tag{32}$$

and we approximate $g(\rho)$ by

$$\tilde{g}_n(\rho) = \int_0^1 I_n(u,\rho)\mathrm{d}u. \tag{33}$$

We do not consider truncation to the left for simplicity and because our numerical evidence suggests that the mixed problem is less demanding computationally than the discrete one. To bound the error, observe that

$$I(u,\rho) - I_n(u,\rho) = u\sum_{i>n} \bar{\Phi}(i-1,u,\rho)p_i \geq 0.$$

Integrating this over u, we obtain lower and upper bounds on the error:

$$0 \leq g(\rho) - \tilde{g}_n(\rho) = \int_0^1 u\sum_{i>n} \bar{\Phi}(i-1,u,\rho)p_i\mathrm{d}u \leq t_n\int_0^1 u\bar{\Phi}(n,u,\rho)\mathrm{d}u \leq \frac{t_n}{2}. \tag{34}$$

Computing the integral upper bound above (second from the right) would require numerical integration. For simplicity, we will forego this and use instead the looser upper bound on the right.

Let $t_{1,n}$, $\tilde{\mu}_{1,n}$, $\underline{\mu}_{1,n}$, $\tilde{\sigma}_{1,n}$, $\underline{\sigma}_{1,n}$, and $\bar{\sigma}_{1,n}$ be as in Section 3.2 with truncation point n, and referring to the discrete marginal. Put $\tilde{r}_n(\rho) = (\tilde{g}_n(\rho) - \tilde{\mu}_{1,n}/2)/(\tilde{\sigma}_{1,n}/\sqrt{12})$ as an approximation of $r(\rho)$. Since $\tilde{g}_n$ is the g in (30) that results when we shift to the point n the probability mass

of the points to its right, it follows immediately from Corollary 1 that $\tilde{r}_n$ is a continuous strictly increasing function on $[-1,1]$, and thus has an inverse; that is, for $r^* \in [\tilde{r}_n(-1), \tilde{r}_n(1)]$, there exists a unique ρ such that $\tilde{r}_n(\rho) = r^*$, which we denote $\tilde{r}_n^{-1}(r^*)$. Our main result is as follows.

Proposition 3 *Let $\rho^* = \tilde{r}_n^{-1}(r^*)$, where $r^* \in [\tilde{r}_n(-1), \tilde{r}_n(1)]$. Provided that $\underline{\sigma}_{1,n}^2$ is positive, we have*

$$\zeta_n \leq r(\rho^*) - r^* \leq \theta_n \quad \textit{for all } n, \tag{35}$$

where

$$\zeta_n = \begin{cases} r^* \left(\frac{\tilde{\sigma}_{1,n}}{\bar{\sigma}_{1,n}} - 1 \right), & r^* > 0 \\ r^* \left(\frac{\tilde{\sigma}_{1,n}}{\underline{\sigma}_{1,n}} - 1 \right), & r^* < 0 \end{cases}$$

and

$$\theta_n = \begin{cases} \frac{\sqrt{12}(t_{1,n} + \tilde{\mu}_{1,n} - \underline{\mu}_{1,n})}{2\underline{\sigma}_{1,n}} + r^* \left(\frac{\tilde{\sigma}_{1,n}}{\underline{\sigma}_{1,n}} - 1 \right), & r^* > 0 \\ \frac{\sqrt{12}(t_{1,n} + \tilde{\mu}_{1,n} - \underline{\mu}_{1,n})}{2\underline{\sigma}_{1,n}} + r^* \left(\frac{\tilde{\sigma}_{1,n}}{\bar{\sigma}_{1,n}} - 1 \right), & r^* < 0. \end{cases}$$

Proof. Putting $\tilde{h}_n(y) = \tilde{g}_n(y) - \tilde{\mu}_{1,n}/2 - r^* \tilde{\sigma}_{1,n}/\sqrt{12}$, we have $\tilde{h}_n(\rho^*) = 0$. Equation (28) holds, where $\tilde{\mu}_{2,\cdot} = \mu_2 = 1/2$ and $\tilde{\sigma}_{2,\cdot} = \sigma_2 = 1/\sqrt{12}$ refer to the continuous marginal. The result now follows from (34), (14), and (15). □

Note that $\zeta_n \leq 0$ and $\theta_n > 0$. We can see the asymptotics of the error bounds in (35) as $n \to \infty$, which will show that the error converges to zero. The quantity ζ_n behaves according to point (i) following Proposition 2, modified to eliminate the tail corresponding to the continuous marginal, so $\zeta_n = O(t_{1,n}^2)$. A simple calculation gives $\theta_n = [\sqrt{12}/(2\sigma^2)]t_{1,n} + o(t_{1,n})$. The asymptotic differs from that in the discrete case because the bounding method here is different.

Remark 2 The bounds in (34), (14), and (15) imply lower and upper bounds on $r(\rho)$ analogous to (29). The distance between these bounds converges to zero as $n \to \infty$. This enables the computation of $r(\rho)$, for any $\rho \in [-1,1]$, to any desired accuracy.

The work to solve $\tilde{r}_n(\rho) = r^*$ tends to be *roughly* linear in n as a consequence of the work to compute $I_n(u,\rho)$ being (roughly) linear. Our approach parallels that for the discrete problem: initialize n as the smallest support point and iteratively increase it by one until $\max(-\zeta_n, \theta_n)$ is at most a specified tolerance δ; for any $\delta > 0$, clearly there exists a finite n satisfying this.

5 Numerical Results

We solved test problems with marginals in one of three families: discrete Pareto, Poisson, and negative binomial. For these problems, solutions to equations $\tilde{r}_{\mathbf{n}}(\rho) = r^*$ associated to two different truncations, that is, different $\mathbf{n}$, are computed, as detailed later. The root may be computed via standard derivative-free methods (we use MATLAB's `fzero`) or via derivative-based ones as in

Avramidis et al. (2009). On average, the derivative-based methods were slightly faster. For the mixed problem, we report on the derivative-free method, simply to avoid having to give (integral) expressions for the derivatives. Computations were done in MATLAB, and CPU times measured via `tic/toc` commands. We do not claim these times are competitive; for example, in solving a few problems from Avramidis et al. (2009) with identical truncation and root-finder, our CPU times are larger by a factor of about one thousand. The large timing gap seems to be primarily due to the computer language (these authors use Java). We compute $\bar{\Phi}_\rho(x, y)$, the standard bivariate-normal c.d.f. at $(-x, -y)$, via MATLAB's function `mvncdf` to tolerance 10^{-9}; this method cites Drezner and Wesolowsky (1989), so we think it is reasonably efficient.

Discrete and mixed problems appear in Sections 5.1 and 5.2, respectively.

5.1 Discrete Problems

For the Pareto and Poisson families, four values r^* are chosen between $0.999\tilde{r}_{\mathbf{n_0}}(-0.9999)$ and $0.999\tilde{r}_{\mathbf{n_0}}(0.9999)$, that is, close to the minimal and maximal rank correlation, respectively, inclusively of these and in equal distance. The *benchmark* uses the truncation vector $\mathbf{n_0} = (l_{1,0}, r_{1,0}, l_{2,0}, r_{2,0})$, where $l_{k,0}$ is the leftmost support point and $r_{k,0}$ is the quantile of order $1 - p$, where $p = 10^{-6}$. This p value is also the choice of Avramidis et al. (2009) and Channouf and L'Ecuyer (2009). We compare this against truncation via Algorithm 1 with tolerances specified shortly. The root-finding problem is solved by a hybrid of the Newton-Raphson method and bisection, identical to Press et al. (1992, routine `rtsafe`, pp. 366–367) and to method NI3 in Avramidis et al. (2009, Section 3.1.4), to which we refer for analytical derivatives of $g_{\mathbf{n}}$.

A user of Algorithm 1 with (absolute error) tolerance δ will choose a number $0 < \beta \leq 1$ and set $\delta_{\mathrm{r}} = \delta\beta$ and $\delta_{\mathrm{l}} = \delta(1 - \beta)$. We considered $\beta \in \{1/10,\ 1/2,\ 1\}$ to assess potential sensitivity. In the Pareto case, with $\delta = 10^{-3}$, $\beta = 1$ was optimum (minimized both number of terms and CPU time). This case is shown in Table 1 in detail, but efficiency (the ratio of the benchmark's CPU time to our CPU time) is high for all β (the average efficiency of the best β to the worst one is roughly 3). Each of the six panels in Table 1 specifies a pair of marginals and the benchmark number $w_0 = (r_{1,0} - l_{1,0} + 1)(r_{2,0} - l_{2,0} + 1)$. Each row within a panel corresponds to the problem instance with target r^*; we report the (approximate) solution ρ; our method's number $w = (r_1 - l_1 + 1)(r_2 - l_2 + 1)$; our CPU time; the error estimate $\tilde{r}_{\mathbf{n_0}}(\rho) - r^*$ (where "3e-04" means 3×10^{-4}); and the efficiency. Heavier tails (smaller α) are associated with more work for our method (larger w and CPU) and larger efficiency. Efficiency is also reflected well by the ratio w_0/w. The work reduction is a result of adhering to the accuracy requirement via the bounds, together with the accuracy being modest. With increased accuracy, we can expect work to increase; for example, for $\delta_{\mathrm{r}} = 10^{-6}$ and $\delta_{\mathrm{l}} = 0$, average efficiency in the six panels drops to 1.9, 2.8, 5.2, 4.3, 9.4, and 23.5, respectively. Incidentally, in the first row of each panel, the target is $r^* = 0.999 r_{\mathbf{n_0}}(-0.9999)$ and

the (approximate) solution ρ is far from -0.9999; this happens because the function $r_{\mathbf{n_0}}$ increases very slowly between -0.9999 and that ρ. We also solved the same problems with a second root finder, MATLAB's `fzero`; efficiencies were (again) roughly linear in w_0/w.

Table 1: Discrete problem with Pareto(α_1) and Pareto(α_2) marginals. $\delta_{\mathrm{r}} = 10^{-3}$, $\delta_{\mathrm{l}} = 0$.

	r^*	ρ	w	CPU (sec)	$\tilde{r}_{\mathbf{n_0}}(\rho) - r^*$	efficiency
$\alpha_1 = 5$, $\alpha_2 = 5$	-0.0368	-0.5160	49	0.20	3e-04	16.5
$w_0 = 484$	0.3044	0.6541	49	0.19	3e-04	10.9
	0.6455	0.9157	49	0.32	3e-04	11.2
	0.9867	0.9999	49	0.56	3e-04	11.2
$\alpha_1 = 5$, $\alpha_2 = 4$	-0.0547	-0.5677	72	0.33	4e-04	29.9
$w_0 = 1496$	0.2001	0.4849	72	0.24	4e-04	23.4
	0.4550	0.7892	72	0.32	5e-04	24.8
	0.7099	0.9923	72	0.51	5e-04	22.1
$\alpha_1 = 5$, $\alpha_2 = 3$	-0.0846	-0.6420	190	0.95	5e-04	112.6
$w_0 = 14190$	0.1311	0.3341	190	0.74	5e-04	85.5
	0.3468	0.6875	190	0.74	5e-04	85.7
	0.5625	0.9999	190	1.18	6e-04	80.7
$\alpha_1 = 4$, $\alpha_2 = 4$	-0.0815	-0.6277	100	0.53	5e-04	65.4
$w_0 = 4624$	0.2752	0.5436	100	0.39	5e-04	53.6
	0.6319	0.8777	100	0.60	5e-04	52.8
	0.9887	0.9999	100	1.18	5e-04	53.5
$\alpha_1 = 4$, $\alpha_2 = 3$	-0.1259	-0.7134	261	1.63	4e-04	205.1
$w_0 = 43860$	0.1659	0.3426	261	1.11	4e-04	183.5
	0.4576	0.7269	261	1.10	5e-04	183.7
	0.7494	0.9987	261	1.97	5e-04	170.6
$\alpha_1 = 3$, $\alpha_2 = 3$	-0.1945	-0.7882	529	3.36	5e-04	952.8
$w_0 = 416025$	0.2008	0.3475	529	2.35	5e-04	828.7
	0.5960	0.7933	552	2.82	5e-04	803.8
	0.9913	0.9999	552	7.15	5e-04	809.8

The second set of examples has Poisson marginals. We keep $\delta = 10^{-3}$, and show in Table 2 the (preferred) case $\beta = 1/2$. For $\beta = 1/10$, efficiency is between 1.9 and 5.1, and averages 4.9 in the last panel. For $\beta = 1$, efficiency is between 1.4 and 6.5, except for the last row, where it is 0.83 despite the fact that $w < w_0$.

For the negative-binomial marginals and targets in Avramidis et al. (2009), performance is comparable to the Poisson case. In the largest problems (largest means), w_0 is about 190 thousand, w is 28, 27, and 57 thousand for $\beta = 1/10$, $1/2$, and 1, respectively, and efficiency is about w_0/w.

The work of Algorithm 1 was not significant as a fraction of the overall work. In Table 1, this fraction averaged 0.7%, and the maximum was 4.5%; in Table 2, the respective figures were 1.3% and 9.6%. The larger fractions occurred consistently in the problems requiring less work.

Table 2: Discrete problem with Poisson(λ_1) and Poisson(λ_2) marginals. $\delta_r = \delta_l = 0.5 \times 10^{-3}$.

	r^*	ρ	w	CPU (sec)	$\tilde{r}_{\mathbf{n_0}}(\rho) - r^*$	efficiency
$\lambda_1 = 1$	-0.8501	-0.9898	30	0.18	9e-05	3.8
$\lambda_2 = 1$	-0.2359	-0.2922	30	0.11	1e-04	3.5
$w_0 = 100$	0.3783	0.4635	30	0.09	2e-04	3.5
	0.9925	0.9999	30	0.39	3e-04	3.0
$\lambda_1 = 1$	-0.9248	-0.9963	100	0.69	7e-05	3.2
$\lambda_2 = 10$	-0.3075	-0.3505	100	0.32	1e-04	3.1
$w_0 = 290$	0.3099	0.3539	100	0.32	3e-04	3.1
	0.9272	0.9981	100	0.63	4e-04	3.3
$\lambda_1 = 1$	-0.9352	-0.9987	330	2.07	3e-04	5.6
$\lambda_2 = 100$	-0.3116	-0.3532	330	1.08	4e-04	5.2
$w_0 = 1520$	0.3121	0.3550	330	1.05	6e-04	5.0
	0.9358	0.9997	330	2.69	6e-04	4.3
$\lambda_1 = 10$	-0.9818	-0.9985	400	2.85	1e-05	2.1
$\lambda_2 = 10$	-0.3222	-0.3394	400	1.44	9e-05	2.2
$w_0 = 841$	0.3374	0.3549	400	1.42	2e-04	2.1
	0.9970	0.9998	400	4.71	3e-04	2.2
$\lambda_1 = 10$	-0.9906	-0.9987	1300	9.65	2e-04	3.3
$\lambda_2 = 100$	-0.3294	-0.3450	1300	4.93	3e-04	3.4
$w_0 = 4408$	0.3317	0.3478	1300	4.89	5e-04	3.4
	0.9928	0.9993	1320	10.82	5e-04	3.0
$\lambda_1 = 100$	-0.9972	-0.9988	4422	34.70	2e-04	5.1
$\lambda_2 = 100$	-0.3320	-0.3460	4422	17.47	3e-04	5.1
$w_0 = 23104$	0.3332	0.3479	4422	17.23	4e-04	5.1
	0.9984	0.9996	4489	41.71	4e-04	5.1

5.2 Mixed Problems

We compare two alternative truncation points: (i) a *benchmark* n_0, set as the quantile of order $1-10^{-6}$; and (ii) the smallest n such that the error bound $\max(-\zeta_n, \theta_n)$ is no larger than $\delta = 10^{-3}$. The values r^* are chosen via near-extremes $0.999\tilde{r}_{n_0}(\pm 0.9999)$, as before. The respective equations, $\tilde{r}_{n_0}(\rho) = r^*$ and $r_n(\rho) = r^*$, are solved with MATLAB's `fzero`, described as "a combination of bisection, secant, and inverse quadratic interpolation methods". The integral in (33) is evaluated via MATLAB's `quadgk` function, described as "adaptive quadrature based on a Gauss-Kronrod pair (15th- and 7th-order formulas)", with error tolerance 10^{-12}.

Mixed problems whose discrete marginal is Pareto are seen in Table 3. Heavier tails are associated with more work for our method (larger n and CPU), and larger efficiency, which is also reflected well by n_0/n. The mixed problem is less demanding than the discrete one with same marginals, in agreement with Channouf and L'Ecuyer (2009).

Table 3: Mixed problem with a Pareto(α) discrete marginal. $\delta = 10^{-3}$.

	r^*	ρ	n	CPU (sec)	$\tilde{r}_{n_0}(\rho) - r^*$	efficiency
$\alpha = 5$	-0.3204	-0.9970	16	0.02	-4e-10	1.3
$n_0 = 22$	-0.1068	-0.2606	16	0.02	-3e-10	1.2
	0.1068	0.2606	16	0.03	6e-09	1.1
	0.3205	0.9971	16	0.03	-1e-08	1.2
$\alpha = 4$	-0.4580	-0.9981	32	0.03	-1e-09	1.5
$n_0 = 68$	-0.1527	-0.2884	32	0.03	-3e-07	1.6
	0.1527	0.2884	32	0.04	3e-07	1.6
	0.4580	0.9981	32	0.05	3e-08	1.6
$\alpha = 3$	-0.6465	-0.9987	126	0.11	-3e-08	4.5
$n_0 = 645$	-0.2155	-0.3194	126	0.11	2e-08	4.8
	0.2155	0.3194	126	0.18	3e-08	5.1
	0.6465	0.9987	126	0.18	-1e-07	5.0
$\alpha = 2.2$	-0.8254	-0.9990	2055	2.34	1e-07	32.7
$n_0 = 61597$	-0.2751	-0.3416	2055	2.39	3e-08	32.7
	0.2751	0.3416	2055	4.76	9e-08	33.2
	0.8254	0.9990	2055	4.70	1e-07	33.9

6 Conclusion

We contributed to the mathematics of constructing a random vector $\mathbf{X}$ of the form (1) by controlling, separately for each pair of coordinates of $\mathbf{X}$, the rank correlations or product-moment correlations. For arbitrary univariate distribution functions F_1 and F_2, we gave expressions for $\mathbb{E}[F_1(X_1)F_2(X_2)]$ and $\mathbb{E}[X_1X_2]$ and their derivatives with respect to ρ and showed that both the rank correlation $r(\rho)$ and the product-moment correlation are differentiable strictly increasing functions on $(-1, 1)$, thus proving existence and uniqueness of the solution for any feasible target. For the case where $r(\rho)$ cannot be computed exactly due to an infinite discrete support, we showed how to construct an approximation $\tilde{r}$ of r such that equations of form $r(\rho) = r^*$ can be solved to any desired accuracy.

In addition to ensuring accuracy, our method may require less work than truncating at quantiles x_p associated to a small tail probability p, because the work decreases as the (absolute error) tolerance increases. With the tolerance fixed, higher efficiency seems to result by setting β (Section 5.1) depending on the probability mass functions (p.m.f.'s): if both p.m.f.'s are nonincreasing in the direction of the infinite right tail (example: Pareto), truncate only the tail: $\beta = 1$; if both p.m.f.'s are (approximately) symmetric, for example normal-like, due to a central-limit effect (examples: large-mean Poisson; negative binomial with large "number of failures" parameter), set $\beta = 1/2$. In other cases, $\beta = 1/2$ seems reasonable, though not necessarily most efficient. Heavier tails, that is, higher sensitivity of x_p to p, seem to translate to higher potential for work reduction.

Some ideas for future inquiry are now proposed. Marginals with large mean(s) tend to result

in large w and according work, even with our approach. More efficient solution of such problems is an open research problem. Another line of inquiry could be to see if our approach can be extended to the product-moment correlation for general discrete and unbounded marginals. A difficulty in this program is that the summands in the corresponding infinite sums do not seem to permit the convenient bound "1" that we used for the cumulative probabilities.

Acknowledgments

The author thanks Professor Pierre L'Ecuyer for his comments on an early draft of the paper.

References

Asmussen, S., P.W. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis*. Springer, New York.

Avramidis, A. N. 2009. Fitting discrete multivariate distributions with unbounded marginals and normal-copula dependence. M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, R. G. Ingalls, eds., *Proceedings of the 2009 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers, Inc., Piscataway, New Jersey, 452–458.

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 Journal on Computing* **21** 88–106.

Axtell, R. L. 2001. Zipf distribution of U.S. firm sizes. *Science* **293** 1818–1820.

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

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

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 Science, Northwestern University.

Channouf, N., P. L'Ecuyer. 2009. Fitting a normal copula for a multivariate distribution with both discrete and continuous marginals. *Proceedings of the 2009 Winter Simulation Conference*. IEEE Press, 352–358.

Channouf, N., P. L'Ecuyer. 2012. A normal copula model for the arrival process in a call center. *International Transactions in Operational Research* **19** 771–787.

Deuchert, E., S. Brody. 2007. Plausible and implausible parameters for mathematical modeling of nominal heterosexual HIV transmission. *Annals of Epidemiology* **17** 237–244.

Drezner, Z., G. O. Wesolowsky. 1989. On the computation of the bivariate normal integral. *Journal of Statistical Computation and Simulation* **35** 101–107.

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

Ghosh, S., S.G. Henderson. 2003. Behavior of the NORTA method for correlated random vector generation as the dimension increases. *ACM Transactions on Modeling and Computer Simulation* **13** 276–294.

Joe, H. 2005. Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis* **94** 401 – 419.

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

Lehmann, E. L. 1966. Some concepts of dependence. *Annals of Mathematical Statistics* **37** 1137–1153.

Parulekar, Minothi, Armand M. Makowski. 1997. Tail probabilities for M/G/∞ input processes (I): Preliminary asymptotics. *Queueing Systems* **27** 271–296.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, B. P. Flannery. 1992. *Numerical Recipes in C: The Art of Scientific Computing*. 2nd ed. Combridge University Press, New York.

Rudin, W. 1976. *Principles of Mathematical Analysis*. 3rd ed. McGraw-Hill, New York.

Suárez-González, A., J. C. López-Ardao, C. López-García, M. Fernández-Veiga, R. F. R. Rubio, M. E. S. Vieira. 2002. A new heavy-tailed discrete distribution for LRD M/G/∞ sample generation. *Performance Evaluation* **47** 197–219.