Les Cahiers du GERAD ISSN: 0711–2440

Efficient Correlation Matching for Normal-Copula Dependence when Univariate Marginals Are Discrete

A.N. Avramidis, N. Channouf, P. L'Ecuyer

G–2006–70

November 2006

Efficient Correlation Matching for Normal-Copula Dependence when Univariate Marginals Are Discrete

Athanassios N. Avramidis
Nabil Channouf
Pierre L'Ecuyer*

Département d'informatique et de recherche opérationnelle
Université de Montréal
C.P. 6128, Succ. Centre-ville
Montréal (Québec) Canada H3C 3J7
{avramidi, channoun, lecuyer}@iro.umontreal.ca

* *and GERAD*

November 2006

Les Cahiers du GERAD
G–2006–70

Abstract

A popular approach for modeling dependence in a finite-dimensional random vector $\mathbf{X}$ with given univariate marginals is via a normal copula that fits the rank or linear correlations for the bivariate marginals of $\mathbf{X}$. In this approach, known as the NORTA method, the normal distribution function is applied to each coordinate of a vector $\mathbf{Z}$ of correlated standard normals to produce a vector $\mathbf{U}$ of correlated uniforms random variables over $(0, 1)$; then $\mathbf{X}$ is obtained by applying the inverse of the target marginal distribution function for each coordinate of $\mathbf{U}$. The fitting requires finding the appropriate correlation ρ between any two given coordinates of $\mathbf{Z}$ that would yield the target rank or linear correlation r between the corresponding coordinates of $\mathbf{X}$. This root-finding problem is easy to solve when the marginals are continuous, but not when they are discrete. In this paper, we provide a detailed analysis of the NORTA method for discrete marginals. We prove key properties of r and of its derivative as a function of ρ. It turns out that the derivative is easier to evaluate than the function itself. Based on that, we propose and compare alternative methods for finding or approximating the appropriate ρ. The case of discrete distributions with unbounded support is covered as well. In our numerical experiments, a derivative-supported method is faster and more accurate than a state-of-the-art, non-derivative-based method. We also characterize the asymptotic convergence rate of the function r (as a function of ρ) to the continuous-marginals limiting function, when the discrete marginals converge to continuous distributions.

Key Words: Statistics; distribution; estimation; correlation; mathematics; simulation.

Acknowledgments: This work has been supported by a grant from Bell Canada via the Bell University Laboratories, Grants No. CRDPJ-251320 and ODGP0110050 from NSERC-Canada, and a Canada Research Chair to the third author. We thank Richard Simard for his assistance in programming aspects and some numerical experiments.

Résumé

Une approche populaire pour modéliser la dépendance dans un vecteur aléatoire de dimension finie $\mathbf{X}$, dont on connaît ses distributions marginales, consiste à utiliser une copule normale qui ajuste la corrélation de rang ou la corrélation linéaire. Dans cette approche, connue sous le nom NORTA, la fonction de distribution de la loi normale standard est appliquée à chaque composante du vecteur $\mathbf{Z}$ de variables aléatoires normales standards corrélées pour obtenir un vecteur $\mathbf{U}$ de variables aléatoires uniformes corrélées sur $(0, 1)$; par la suite $\mathbf{X}$ est obtenu en appliquant l'inverse de chaque distribution marginale associée à chaque composante de $\mathbf{U}$. Pour faire l'ajustement, il est nécessaire de trouver chaque corrélation ρ appropriée à chaque couple de composantes de $\mathbf{Z}$ qui donne la corrélation de rang ou la corrélation linéaire désirée dans le couple correspondant de $\mathbf{X}$. Le problème de recherche de la racine est très facile dans le cas où les distributions marginales sont continues, contrairement au cas où elles sont discrètes. Dans cet article, nous fournissons une analyse détaillée de la méthode NORTA pour des distributions discrètes. Nous démontrons les propriétés clés de r en tant que fonction de ρ et de sa fonction dérivée. Il s'avère que la fonction dérivée est plus facile à évaluer que la fonction elle-même. Sur cette base, nous proposons et nous comparons différentes méthodes pour trouver ou approximer le ρ approprié. Le cas des distributions marginales avec support infini est aussi étudié. Dans nos exemples numériques, la méthode basée sur la fonction dérivée est plus rapide et plus précise que la méthode standard, qui fait l'état-de-l'art, basée sur la fonction et non sa dérivée. Nous spécifions aussi le taux de convergence asymptotique de r (comme function de ρ) vers la distribution marginale continue limite, dans le cas où les distributions marginales discrètes convergent vers des distributions continues.

1 Introduction

This paper develops methods that support the estimation (fitting) of *discrete* multivariate distributions. A powerful scheme for modeling multivariate distributions in general is based on the concept of *copula*; it permits one to specify separately the marginal distributions and the stochastic dependence. To put our work in the proper perspective, we start by recalling basic facts from copula theory. For a concise introduction to copulas, see Embrechts et al. (2002) or Joe (1997); for a more complete treatment, see Nelsen (1999).

A function $C : [0,1]^d \to [0,1]$ is called a *copula* if it is the distribution function of a random vector in $\mathbb{R}^d$ with $U(0,1)$ marginals (uniform over the interval (0,1)). Consider a random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with joint distribution F and write F_j for the marginal distribution of X_j. A *copula associated with F (equivalently, $\mathbf{X}$)* is a copula C that satisfies

$$F(\mathbf{x}) = C(F_1(x_1), \ldots, F_d(x_d)), \quad \mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d. \tag{1}$$

Given an arbitrary F, a copula C satisfying (1) always exists. If each X_j is a continuous random variable, then C is unique, and this uniqueness means that we have separated the marginals from the dependence structure, which is captured by C. (Otherwise, there may be more than one C satisfying (1), so the dependence cannot be uniquely characterized.) We will shortly specify a class of distributions F via (1) by specifying the dependence via a d-variate copula C that is selected *after* the marginals have been selected. For given marginals, the choice of copula can have a dramatic impact; see Embrechts et al. (2003, Sec. 7.1) for an example.

In this paper, we nevertheless restrict our attention to *normal copulas*; these are the copulas defined by taking F as a multivariate distribution in (1). This family of copulas has been suggested by several authors, dating back to Mardia (1970). Attractive features of normal copulas are that they facilitate estimation (as will be explained) and simulation. They are sufficient and very convenient for a wide range of applications where fitting only the marginals and the correlations is a reasonable compromise. In more than two or three dimensions, estimating the entire copula in a complicated real-life situation is often an insurmountable challenge.

Other models of discrete multivariate distributions can be found, e.g., in Joe (1997, sec. 7.2). A limitation of several of these models is that the same parameters affect the marginal distributions and the dependence. For example, in Model (7.27) of Joe (1997) the X_i's are conditionally independent Poisson with mean A_i, where the A_i, $i = 1, \ldots, d$, obey some multivariate continuous distribution; but the upper limit $\mathrm{Corr}(X_i, X_j) = 1$ is only possible in the limit where X_i and X_j have identical marginals and $\mathrm{Var}(X_i)/\mathbb{E}(X_i) \to \infty$; a further limitation is that if one wanted negative binomial marginals for the X_i, then one would need the A_i to obey a multivariate distribution with gamma marginals, which is not convenient to use (Joe, 1997, p. 236).

Returning to the normal copula, if we write $\mathcal{N}_{\mathbf{R}}$ for the normal distribution with mean the zero vector and $d \times d$ correlation matrix $\mathbf{R}$, and $C_{\mathbf{R}}$ for the associated copula defined via (1) with $F = \mathcal{N}_{\mathbf{R}}$, we have the representation

$$\begin{aligned} &\mathbf{Z} = (Z_1, \dots Z_d) \sim \mathcal{N}_{\mathbf{R}} \\ &\mathbf{X} = (X_1, \dots, X_d) = \left(F_1^{-1}[\Phi(Z_1)], \dots, F_d^{-1}([\Phi(Z_d)])\right), \end{aligned} \tag{2}$$

where Φ is the standard normal distribution function (with mean 0 and variance 1) and F_i^{-1}, defined by $F_i^{-1}(u) = \inf\{x : F_i(x) \geq u\}$ for $0 \leq u \leq 1$, is the *quantile function* of the marginal distribution F_i. It is easily seen that $C_{\mathbf{R}}$ is a copula associated with $\mathbf{X}$ in (2). This $C_{\mathbf{R}}$ is a *normal copula*. Model (2) is also known under the name *NORTA* (Cario and Nelson, 1996, 1998; Chen, 2001), an acronym for *NORmal To Anything*, since normal variates are transformed to variates with general nonuniform marginals.

The main issue here is how to find a matrix $\mathbf{R}$ such that the vector $\mathbf{X}$ has the desired rank or linear correlation matrix, either exactly or approximately. The natural way of doing this is element-wise, so we start by discussing the bivariate case ($d = 2$). Later, we will discuss the extension to $d > 2$.

Suppose that $d = 2$ and that the marginals F_1 and F_2 have been specified. Selecting $\mathbf{R}$ in (2) reduces to selecting the scalar correlation $\rho = \mathrm{Corr}(Z_1, Z_2)$. The *rank correlation* between X_1 and X_2 is

$$\begin{aligned} r_X(\rho) = r_X(\rho; F_1, F_2) &= \mathrm{Corr}(F_1(X_1), F_2(X_2)) \\ &= \mathrm{Corr}\left(F_1 \circ F_1^{-1} \circ \Phi(Z_1), F_2 \circ F_2^{-1} \circ \Phi(Z_2)\right) \end{aligned}$$

where $\rho = \mathrm{Corr}(Z_1, Z_2)$ and "$\circ$" denotes function composition. We will explain shortly that r_X may depend on the marginals only if at least one of them is not continuous. One approach to specifying ρ is to require that $r_X(\rho; F_1, F_2)$ equals a given target value $\tilde{r}$, which may be the sample rank correlation computed from data, (observations of $\mathbf{X}$), or determined otherwise. This leads to the *NORTA rank-correlation matching* problem of solving

$$r_X(\rho; F_1, F_2) = \tilde{r}. \tag{3}$$

The dependence of r_X on the marginals disappears when F_1 and F_2 are both continuous: $F_l \circ F_l^{-1}$, $l = 1, 2$ are the identity map, and thus

$$r_X(\rho; F_1, F_2) = \mathrm{Corr}(\Phi(Z_1), \Phi(Z_2)) = (6/\pi) \arcsin(\rho/2),$$

where the second equality is a well-known property of the bivariate normal distribution. Thus, solving (3) poses a problem only when at least one of the marginals is not continuous.

Another possibility would be to work analogously with the *linear correlation*

$$\rho_X(\rho; F_1, F_2) = \mathrm{Corr}(X_1, X_2) = \mathrm{Corr}(F_1^{-1} \circ \Phi(Z_1), F_2^{-1} \circ \Phi(Z_2)),$$

which leads to the *NORTA linear-correlation matching* equality:

$$\rho_X(\rho; F_1, F_2) = \tilde{\rho}, \tag{4}$$

where $\tilde{\rho}$ is the sample linear correlation computed from data. Embrechts et al. (2002) give a detailed account of measures of dependence and strong arguments that rank correlation is a more appropriate measure than linear correlation. We review their Example 5, which illuminates this issue. Consider the marginals $X_1 \sim \text{Lognormal}(0, 1)$ and $X_2 \sim \text{Lognormal}(0, \sigma^2)$ for $\sigma > 0$. Under several measures of dependence discussed there, extreme positive and negative dependence occur when X_2 is an increasing (decreasing) function of X_1, i.e., in the stochastic representations $(X_1, X_2) = (e^Z, e^{\sigma Z})$ and $(X_1, X_2) = (e^Z, e^{-\sigma Z})$, respectively, where $Z \sim \text{Normal}(0, 1)$. Then the rank correlation of the pair (X_1, X_2) equals 1 and -1, respectively. On the other hand, we have: $\text{Corr}(e^Z, e^{\sigma Z}) = (e^\sigma - 1)/\sqrt{(e-1)(e^{\sigma^2}-1)}$ and $\text{Corr}(e^Z, e^{-\sigma Z}) = (e^{-\sigma} - 1)/\sqrt{(e-1)(e^{\sigma^2}-1)}$; these continuous functions of σ are far from 1 and -1 over most of their domain, and they converge to zero as $\sigma \to \infty$. Here, linear correlation fails to capture well the dependence, and the failure is dramatic in the limit. Hörmann et al. (2004, Section 12.5) give additional examples of this phenomenon and strongly recommend matching the rank correlations instead of the linear correlations.

When $d > 2$, (2) is specified by constructing $\mathbf{R}$ elementwise; that is, for each pair (i, j), one has a target value $\tilde{r}_{i,j}$ (or $\tilde{\rho}_{i,j}$) and one sets the (i, j)-th element of $\mathbf{R}$ to the solution of (3) with $\tilde{r} = \tilde{r}_{i,j}$ (or the solution of (4) with $\tilde{\rho} = \tilde{\rho}_{i,j}$). Thus, one needs to solve $d(d-1)/2$ such independent equations. In case the resulting matrix $\mathbf{R}$ is not positive semidefinite, various authors suggest replacing it by another matrix that is positive semidefinite and minimizes some measure of distance from $\mathbf{R}$ (Mardia, 1970; Lurie and Goldberg, 1998; Ghosh and Henderson, 2003). According to Ghosh and Henderson (2003), this appears to work well, in the sense that the minimized distance was very small in their tests.

Another related setting is the VARTA class of multivariate stationary time series (Biller and Nelson, 2003), $\{\mathbf{X}_t = (X_{1,t}, \ldots, X_{k,t}), t = 1, 2, \ldots\}$, where one specifies the marginals F_l for $l = 1, \ldots, k$ and dependence via the normal copula, i.e., via correlations between $X_{i,t}$ and $X_{j,t-h}$ for $h = 0, 1, \ldots, p$ and $i, j \in \{1, 2, \ldots, k\}$; the univariate case $k = 1$ is known as ARTA (Cario and Nelson, 1996). That is, the i-th component time series is obtained by the transformation $X_{i,t} = F_i^{-1}(\Phi(Z_{i,t}))$, where $\{\mathbf{Z}_t\} = (Z_{1,t}, \ldots, Z_{k,t})$ is a k-variate vector autoregressive process of order p and whose noise vectors are Gaussian; see Biller and Nelson (2003, Section 3.1.1). Here, the number of equations that must be solved is $pk^2 + k(k-1)/2$. (The complications and remedies mentioned earlier have analogs in the time-series setting). Because the number of equations to be solved can be considerable, efficient methods for solving equations of the form (3) and (4) are of interest.

We now review past work on NORTA correlation matching. This literature has emphasized linear-correlation matching (Cario and Nelson, 1998; Chen, 2001; Biller and Nelson,

2003), despite the existing arguments in favor of rank correlation, and in principle applies to both continuous and discrete marginals, unless otherwise said. Cario and Nelson (1998) employ root bracketing combined with approximating $\rho_X(\rho; F_1, F_2)$ (a function of ρ) via two-dimensional numerical integration (Gauss and Kronrod quadrature rules). With discrete marginals, the integrand has a discontinuity at every support point, so these general-purpose quadrature rules are not well-suited. Chen (2001) proposed a simulation-based approach. Biller and Nelson (2003) restricted the marginals to the Johnson family and thus simplified the solution. For the case of discrete marginals, we were unable to find a published or unpublished example of NORTA rank- or linear-correlation matching.

The main contributions of this paper are a detailed study of the NORTA correlation matching problems (3) and (4) and the development of efficient methods for solving these problems *when the marginal distributions are discrete*. We express $r_X(\rho; F_1, F_2)$ as an infinite series, where each term involves a bivariate normal integral to the northeast of a bivariate support point. We obtain the derivative of r_X with respect to ρ as a series of terms that only involve the exponential function. For finite support, it turns out that the derivative is considerably faster to evaluate than r_X, even if one uses state-of-the-art methods to compute the bivariate normal integrals. We then develop solution methods that exploit the derivative. In particular, we propose a simple Newton-type method, which in numerical experiments is faster and more accurate than a state-of-the-art, non-derivative-based method. For unbounded marginals, we propose a method that does not require evaluating r_X and that substitutes an approximation of the derivative (obtained by truncating the series); and we provide bounds on the resulting error.

Another contribution is a convergence result on the L_∞ distance (i.e., the supremum over $\rho \in [-1, 1]$ of the absolute difference) between the rank-correlation function $r_X(\rho; F_1, F_2)$ for given discrete marginals F_1 and F_2 and the explicitly known analog for continuous marginals, in terms of the maximum probability masses of F_1 and F_2, as these masses go to zero.

Our results and methods for the rank-correlation problem extend immediately to the linear-correlation problem, under mild uniform convergence conditions. For reasons given earlier, we emphasize the rank-correlation problem and discuss only briefly the extension to the linear-correlation problem.

The remainder is organized as follows. Section 2.1 summarizes relevant background. In Section 2.2, we prove key properties of the rank and linear correlations as a function of ρ, we obtain expressions for their derivatives, and we discuss implications. Section 2.3 proposes an approximation to the derivative, with error bounds, for the infinite-support case. The convergence rate result to the continuous case is proved in Section 2.4. Section 3 specifies the benchmark and the new methods for bivariate NORTA correlation matching, for either finite or infinite support. In Section 4 we give numerical examples.

2 Mathematical properties

2.1 Background

Theorem 1 below summarizes useful known results that hold for arbitrary marginals. Let

$$\phi_\rho(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-(x^2 - 2\rho xy + y^2)/[2(1-\rho^2)]\right\}, \tag{5}$$

the bivariate standard normal density function with correlation ρ.

Theorem 1 *Assume F_1 and F_2 are arbitrary c.d.f.'s and define $r_X(\rho) = \mathrm{Corr}(F_1(X_1), F_2(X_2))$ and $\rho_X(\rho) = \mathrm{Corr}(X_1, X_2)$ with (X_1, X_2) defined as in (2) with $\rho = \mathrm{Corr}(Z_1, Z_2)$.*

1 *The functions r_X and ρ_X are nondecreasing on $[-1,1]$. We have $r_X(0) = 0$ and $\rho_X(0) = 0$.*

2 *Assume there exists $\delta > 0$ such that $\mathbb{E}[|X_1X_2|^{1+\delta}] < \infty$ for all $\rho \in [-1,1]$. Then r_X and ρ_X are continuous on $[-1,1]$.*

3 *If the marginals F_l are continuous, then*

$$\mathrm{Corr}(F_1(X_1), F_2(X_2)) = 12 g_{\mathrm{C}}(\rho) - 3 = \frac{6}{\pi}\arcsin(\rho/2) =: r_{\mathrm{C}}(\rho), \tag{6}$$

where

$$g_{\mathrm{C}}(\rho) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Phi(x_1)\Phi(x_2)\phi_\rho(x_1,x_2)dx_1dx_2.$$

Proof. For the linear correlation ρ_X, parts 1 and 2 are Theorems 1 and 2 of Cario and Nelson (1996), respectively. To prove the analogous results for r_X, it suffices to replace the nondecreasing functions $F_l^{-1} \circ \Phi$ in the proofs of Theorems 1 and 2 of Cario and Nelson (1996), respectively, by the nondecreasing functions $F_l \circ F_l^{-1} \circ \Phi$ for $l = 1, 2$. According to Kurowicka and Cooke (2001), part 3 was obtained by Karl Pearson in 1907. A more recent reference is Kruskal (1958). □

Parts 1 and 2 provide the basis for solving (3) and (4) via root-bracketing; see method NI1 in Section 3. In Section 2.4 we provide a theoretical result that establishes $r_{\mathrm{C}}(\rho)$ as a natural approximation of $r_X(\rho; F_1, F_2)$. The derivative-based solution methods of Section 3 can work without this approximation, but the approximation usually helps increase their speed.

2.2 Derivatives and further properties

This section develops the basis for the proposed solution methods. We assume that marginals are discrete and satisfy weak conditions and we develop explicit formulæ for the derivatives of the functions r_X and ρ_X.

For $l = 1, 2$, we assume that the positive support can be (and is) enumerated in increasing order as $0 \leq x_{l,0} < x_{l,1} < x_{l,2} < ...$ and that the negative support is enumerated as $0 > x_{l,-1} > x_{l,-2} > ....$ An example where this condition fails is a support having a finite $x_0 > 0$ as an accumulation point from the left, combined with $F_l(x_0) < 1$. The condition holds for most discrete distributions usually encountered in applications, e.g., discrete uniform, binomial, geometric, Poisson, negative binomial, and certainly for many more, e.g., any finite mixture of any of these. From this practical standpoint, the assumption does not appear restrictive.

Denote the probability mass of $x_{l,j}$ as $p_{l,j}$. For any integer k, the cumulative probability mass is $f_{l,k} = \sum_{j=-\infty}^{k} p_{l,j}$. For $l = 1, 2$, $\lim_{k\to\infty} p_{l,k} = \lim_{k\to\infty} p_{l,-k} = 0$. Write $z_{l,k} = \Phi^{-1}(f_{l,k})$, and note that $\lim_{k\to\infty} z_{l,k} = -\lim_{k\to\infty} z_{l,-k} = \infty$. If the probability mass above zero is concentrated on a finite number of points, then an increasing sequence of artificial points $x_{l,j}$ with probability $p_{l,j} = 0$ can be added as needed, and similarly for the probability mass below zero.

Derivative of the rank correlation. The rank correlation between X_1 and X_2 is

$$r_X(\rho) = \mathrm{Corr}(F_1(X_1), F_2(X_2)) = \frac{g(\rho) - \mu_1\mu_2}{\sigma_1\sigma_2}, \tag{7}$$

where:

$$\begin{aligned} g(\rho) &= \mathbb{E}\left[F_1(X_1)F_2(X_2)\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_1\{F_1^{-1}[\Phi(x_1)]\}F_2\{F_2^{-1}[\Phi(x_2)]\}\phi_\rho(x_1, x_2)dx_1dx_2, \end{aligned} \tag{8}$$

where μ_k and σ_k are the known mean and standard deviation of $F_k(X_k)$, respectively. Note that r_X involves only shifting and scaling of g by known constants. We rewrite the double integral in (8) as

$$g(\rho) = \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} f_{1,i}f_{2,j}\left(\int_{z_{1,i-1}}^{z_{1,i}}\int_{z_{2,j-1}}^{z_{2,j}} \phi_\rho(x_1, x_2)dx_1dx_2\right) \tag{9}$$

$$\begin{aligned} &= \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} f_{1,i}f_{2,j}\Big[\bar{\Phi}_\rho(z_{1,i-1}, z_{2,j-1}) - \bar{\Phi}_\rho(z_{1,i-1}, z_{2,j}) \\ &\qquad\qquad -\bar{\Phi}_\rho(z_{1,i}, z_{2,j-1}) + \bar{\Phi}_\rho(z_{1,i}, z_{2,j})\Big] \\ &= \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} (f_{1,i+1} - f_{1,i})(f_{2,j+1} - f_{2,j})\bar{\Phi}_\rho(z_{1,i}, z_{2,j}) \\ &= \sum_{i=-\infty}^{\infty} p_{1,i+1}\sum_{j=-\infty}^{\infty} p_{2,j+1}\bar{\Phi}_\rho(z_{1,i}, z_{2,j}) \end{aligned} \tag{10}$$

which involves the *bivariate normal integral* $\bar{\Phi}_\rho(x,y) = \int_x^\infty \int_y^\infty \phi_\rho(z_1,z_2)dz_1dz_2$. In the derivation above, (9) follows directly from the definition (2); the second step rewrites each double integral over a square as the signed summation of four terms involving four related integrals at the square's corners; the third step is a simple rearrangement of the summation. Observe that in (10), the weight $p_{1,i+1}\,p_{2,j+1}$ multiplies the value of $\bar{\Phi}_\rho$ at $(z_{1,i},\, z_{2,j})$, not at $(z_{1,i+1},\, z_{2,j+1})$. If $x_{1,i+1}$ and $x_{2,j+1}$ are the smallest values with positive probabilities for X_1 and X_2, respectively, then $z_{1,i} = z_{2,j} = -\infty$, so $\bar{\Phi}_\rho(z_{1,i},\, z_{2,j}) = 1$ and the corresponding term in (10) is $p_{1,i+1}\,p_{2,j+1}$. As a special case, suppose X_1 is degenerate to a single value, say $p_{1,i+1} = 1$. Then, (10) yields

$$g(\rho) = \sum_{j=-\infty}^{\infty} p_{2,j+1}\bar{\Phi}_\rho(-\infty, z_{2,j}) = \sum_{j=-\infty}^{\infty} p_{2,j+1}\bar{\Phi}(\Phi^{-1}(f_{2,j}))$$
$$= \sum_{j=-\infty}^{\infty} p_{2,j+1}(1 - f_{2,j}) = \mathbb{E}[\bar{F}_2(X_2)]$$

(a constant), where $\bar{F}_2(x) := P[X_2 \geq x]$. If both X_1 and X_2 are degenerate, this gives $g(\rho) \equiv 1$.

Proposition 1 *The function $g(\rho)$ is infinitely differentiable on the interval $(-1,1)$, with first derivative*

$$g'(\rho) = \sum_{i=-\infty}^{\infty} p_{1,i+1} \sum_{j=-\infty}^{\infty} p_{2,j+1}\phi_\rho(z_{1,i}, z_{2,j}). \tag{11}$$

Proof. We start with the first derivative. We will exploit the property of the bivariate standard normal density that for $-1 < \rho < 1$,

$$\frac{d}{d\rho}\phi_\rho(x,y) = \frac{\partial^2}{\partial x \partial y}\phi_\rho(x,y) \quad \text{for any } x, y \tag{12}$$

(Kendall and Stuart, 1977, p. 393, exercise 15.4). We have

$$\begin{aligned}
\frac{d}{d\rho}\bar{\Phi}_\rho(x,y) &= \int_x^\infty \int_y^\infty \frac{d}{d\rho}\phi_\rho(z_1,z_2)dz_2dz_1 \\
&= \int_x^\infty \frac{d}{dz_1}\left[\int_y^\infty \frac{\partial}{\partial z_2}\phi_\rho(z_1,z_2)dz_2\right]dz_1 \\
&= \int_x^\infty \frac{d}{dz_1}\left[-\phi_\rho(z_1,y)\right]dz_1 \\
&= \phi_\rho(x,y).
\end{aligned} \tag{13}$$

In steps one and two, the interchange of differentiation and integration is valid because of the existence and boundedness of the derivatives over the integration domain; in step two, we used (12); steps three and four use the fundamental theorem of calculus.

Equation (13) shows that the derivative of each term in the series (10) is the corresponding term in the series (11). It remains to show the validity of interchanging the order of differentiation and summation. A sufficient condition for this is that for each $\rho_0 \in (-1,1)$, there is a neighborhood of ρ_0, $N_\epsilon(\rho_0) = (\rho_0 - \epsilon,\, \rho_0 + \epsilon) \subset (-1,1)$, such that the series on the right side of (11) converges uniformly for $\rho \in N_\epsilon(\rho_0)$ (Rudin, 1976, Theorem 7.17). This uniform convergence holds in particular if there is an increasing sequence of finite sets $S_k \subset \mathbb{Z}^2$, $k \geq 0$, such that

$$\lim_{k\to\infty} \sup_{\rho\in N_\epsilon(\rho_0)} \sum_{(i,j)\in\mathbb{Z}^2\setminus S_k} p_{1,j+1}\, p_{2,j+1} \phi_\rho(z_{1,i}, z_{2,j}) = 0. \tag{14}$$

(Since all the terms in (11) are non-negative, this condition is actually a special case of the well-known Cauchy criterion for uniform convergence (Rudin, 1976, Theorem 7.8).) The latter condition is easily verified if we take S_k as the bounded rectangle $\{(i,j) : \max(|i|,|j|) \leq k\}$:

$$\begin{aligned} & \sup_{\rho\in N_\epsilon(\rho_0)} \sum_{(i,j):\max(|i|,|j|)>k} p_{1,i+1}\, p_{2,j+1} \phi_\rho(z_{1,i}, z_{2,j}) \\ \leq\ & \frac{1}{2\pi\sqrt{1-(\rho_0+\epsilon)^2}} \left[\sum_{i:|i|>k} p_{1,i+1} + \sum_{j:|j|>k} p_{2,j+1} \right] \to 0 \quad \text{as } k \to \infty. \end{aligned} \tag{15}$$

To study the higher-order derivatives, we note that $\phi_\rho(x,y) = (1-\rho^2)^{-1/2}\phi(x)\phi[(y - \rho x)(1-\rho^2)^{-1/2}]/(2\pi)$ and we change from coordinates (x,y) to polar coordinates (r,θ), i.e., set $x = r\cos\theta$, $y = r\sin\theta$, where $r \geq 0$ and $\theta \in [0, 2\pi]$. Let $\delta > 0$ and write $\phi_\rho^{(d)}$ for the d-th derivative of ϕ_ρ with respect to ρ for $|\rho| \leq 1 - \delta$. Differentiation gives

$$\begin{aligned} \phi_\rho^{(1)}(r,\theta) &= \phi(r\cos\theta)\phi(ra(\theta,\rho)/\sqrt{1-\rho^2}) \frac{(1-\rho^2)[2r^2a(\theta,\rho)-1]\cos\theta + r^2a^2(\theta,\rho)}{4\pi(1-\rho^2)^{5/2}} \\ &\leq K_1 r^2 \exp\left(-r^2 b(\theta,\rho)/2\right) \quad \text{for all } r,\theta, \text{ and } |\rho| \leq 1-\delta, \end{aligned} \tag{16}$$

where $a(\theta,\rho) = \sin\theta - \rho\cos\theta$, $b(\theta,\rho) = (1 - 2\rho\sin\theta\cos\theta)/(1-\rho^2)$, and K_1 is a positive constant. First, observe that for any $\alpha > 0$ and positive integer d, $r^d \exp(-\alpha r^2)$ is a bounded function of r for $r \geq 0$. Second, for any θ, simple calculus shows that $\inf_{\rho\in[-1,1]} b(\theta,\rho) \geq 1/2$. This shows that

$$\sup_{|\rho|\leq 1-\delta, r\geq 0, \theta\in[0,2\pi]} \phi_\rho^{(d)}(r,\theta) < \infty \tag{17}$$

for $d = 1$. Thus, the analog of (15) holds when we substitute $\phi_\rho^{(1)}$ for ϕ_ρ; this proves that g has a second derivative on $(-1,1)$ and that this derivative is an infinite series analogous to (11) (in each term, one replaces ϕ_ρ by the intermediate quantity in (16), in the preferred

coordinate system). The existence of higher-order derivatives of g follows along similar lines, which we only sketch: $\phi_\rho^{(d)}$ obeys a generalized expression as in (16), where the ϕ terms remain intact (the multiplying fraction becomes more complicated); a bound as in the right of (16) applies with the exponential term intact, a power no larger than r^{2d} outside the exponential, and a different constant K_1; Thus, (17) holds for any integer $d > 1$, and the remaining argument is as before. □

Proposition 1, combined with the strict positivity of $\phi_\rho(z_{1,i}, z_{2,i})$ when $z_{1,i} > -\infty$ and $z_{2,i} > -\infty$, and part 2 of Theorem 1, yield:

Corollary 1 *If both F_1 and F_2 are non-degenerate distributions, then the function r_X is strictly increasing on $[-1, 1]$, and has therefore an inverse, i.e., there exists a mapping $r_X^{-1} : [r_X(-1),\ r_X(1)] \to [-1, 1]$ such that $r_X \circ r_X^{-1}$ is the identity map.*

Corollary 1 guarantees the existence and *uniqueness* of a solution to equation (3), under the condition that $\tilde{r} \in [r_X(-1),\ r_X(1)]$.

Derivative of the linear correlation. Analogous properties can be derived for the linear correlation between X_1 and X_2, defined as

$$\rho_X(\rho) = \mathrm{Corr}(X_1, X_2) = \frac{g_{\mathrm{L}}(\rho) - \lambda_1\lambda_2}{\tau_1\tau_2},$$

where

$$g_{\mathrm{L}}(\rho) = \mathbb{E}\left[X_1X_2\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_1^{-1}[\Phi(x_1)]F_2^{-1}[\Phi(x_2)]\phi_\rho(x_1, x_2)dx_1dx_2, \tag{18}$$

and λ_i and $\tau_i^2 < \infty$ are the known mean and variance of F_i, respectively. Paralleling the development that led to (10), we obtain the analogous series representation

$$g_{\mathrm{L}}(\rho) = \sum_{i=-\infty}^{\infty}(x_{1,i+1} - x_{1,i})\sum_{j=-\infty}^{\infty}(x_{2,j+1} - x_{2,j})\bar{\Phi}_\rho(z_{1,i}, z_{2,j}). \tag{19}$$

Cario and Nelson (1998, eq. (5)) have stated an expression analogous to (9) for the function g_{L}, where they heuristically truncate both summations to a finite number of terms; they do not provide an estimate of the truncation error.

To obtain an analogue of Proposition 1, we must justify the interchange of derivative with summation when we differentiate (19) with respect to ρ. A sufficient uniform convergence condition in this case is

Condition 1 For each $\rho_0 \in (0, 1)$, there is a neighborhood $N_\epsilon(\rho_0) = (\rho_0-\epsilon,\ \rho_0+\epsilon) \subset (0, 1)$ such that

$$\lim_{k\to\infty}\sup_{\rho\in N_\epsilon(\rho_0)}\sum_{(i,j):\max(|i|,|j|)>k}(x_{1,i+1} - x_{1,i})(x_{2,j+1} - x_{2,j})\phi_\rho(z_{1,i}, z_{2,j}) = 0. \tag{20}$$

Proposition 2 *If Condition 1 holds, then the function $g_{\mathrm{L}}(\rho)$ is differentiable on $(-1,1)$ with first derivative*

$$g'_{\mathrm{L}}(\rho) = \sum_{i=-\infty}^{\infty} (x_{1,i+1} - x_{1,i}) \sum_{j=-\infty}^{\infty} (x_{2,j+1} - x_{2,j}) \phi_\rho(z_{1,i}, z_{2,j}). \tag{21}$$

Moreover, if Condition 1 holds with $\phi_\rho(z_{1,i}, z_{2,j})$ replaced by its nth derivative with respect to ρ for $n = 1, \ldots, d$, then $g_{\mathrm{L}}(\rho)$ is d times continuously differentiable over $(-1,1)$.

Proof. The proof parallels that of Proposition 1 and we omit the details. □

Condition 1 is clearly verified if both F_1 and F_2 have finite support. A bounded support (i.e., if all the probability mass of the joint distribution is contained in a bounded rectangle) is also a (weaker) sufficient condition. For discrete distributions with unbounded support, the condition will hold if the tail probabilities $1 - F_l(x)$ converge to zero at a fast enough rate when $x \to \infty$. If the support is the set of non-negative integers (this is the case for most popular discrete distributions with infinite support), it is natural to take $x_{l,i} = i$ for all i. We then have $(x_{1,i+1} - x_{1,i})(x_{2,j+1} - x_{2,j}) = 1$ so all we need is that $|z_{l,i}| = |\Phi^{-1}(F_l(x_{l,i}))|$ increases quickly enough with i, for $l = 1, 2$.

Suppose for example that $x_{l,i} = i$ and that the tail of F_l decreases at an exponential rate: $1 - F_l(x) \le \exp[-\gamma x^\alpha]$ for $l = 1, 2$ when x is large enough, for some positive constants α and γ. Several common distributions such as the geometric, negative binomial, Poisson, etc., satisfy this condition. Using the fact that $\Phi^{-1}(y) \sim \sqrt{-2\ln(1-y)}$ when $y \to 1$, we have that for large i,

$$z_{l,i} = \Phi^{-1}(F_l(i)) \ge \Phi^{-1}\left(1 - \exp[-\gamma i^\alpha]\right) \ge (1-\delta)\sqrt{2\gamma i^\alpha}$$

for some small constant $\delta > 0$. Putting this in (5) yields (for i and j large enough)

$$\begin{aligned} \phi_\rho(z_{1,i}, z_{2,j}) &\le \phi_\rho((1-\delta)\sqrt{2\gamma i^\alpha}, (1-\delta)\sqrt{2\gamma j^\alpha}) \\ &\le \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[- \frac{2(1-\delta)^2\gamma(i^\alpha + j^\alpha - 2\rho(ij)^{\alpha/2})}{2(1-\rho^2)} \right]. \end{aligned}$$

But observe that $i^\alpha + j^\alpha - 2\rho(ij)^{\alpha/2} = (i^{\alpha/2} - \rho j^{\alpha/2})^2 + (1-\rho^2) j^\alpha = (j^{\alpha/2} - \rho i^{\alpha/2})^2 + (1-\rho^2) i^\alpha$. Using this, we can easily show that for j large enough,

$$\sum_{i=0}^{\infty} \sup_{\rho \in N_\epsilon(\rho_0)} \phi_\rho(z_{1,i}, z_{2,j}) \le K_0 \exp[-K_1 j^\alpha]$$

for some positive constants K_0 and K_1 that may depend on ρ_0 but not on j. Summing this over $j > k$, for k large enough, we obtain that

$$\sum_{(i,j):j>k} \sup_{\rho \in N_\epsilon(\rho_0)} \phi_\rho(z_{1,i}, z_{2,j}) \le K_0 \sum_{j>k} \exp[-K_1 j^\alpha] \;\to\; 0 \quad \text{when } k \to \infty.$$

The same property obviously holds if we permute i and j, which means that the sum over $\{(i,j) : i > k\}$ also vanishes when $k \to \infty$. This implies (20).

As another example, let $x_{l,i} = i$ and $p_{l,i} = O(i^{-3/2})$. Then each X_l has infinite variance and Condition 1 does not hold.

Corollary 2 *If both F_1 and F_2 are non-degenerate distributions and Condition 1 holds, then ρ_X is strictly increasing on $[-1,1]$, so it has an inverse $\rho_X^{-1} : [\rho_X(-1),\ \rho_X(1)] \to [-1,1]$, and (4) possesses a unique solution in $[-1,1]$ if $\tilde{\rho} \in [\rho_X(-1),\ \rho_X(1)]$.*

We conclude this section by studying the limit when $|\rho| \to 1$. The behavior of $g'(\rho)$ as $\rho \to 1$ depends on whether

$$\text{there exist } i \text{ and } j \text{ such that } 0 < f_{1,i} = f_{2,j} < 1; \tag{22}$$

the behavior as $\rho \to -1$ depends on whether

$$\text{there exist } i \text{ and } j \text{ such that } 0 < f_{1,i} = 1 - f_{2,j} < 1. \tag{23}$$

In words, (22) says that F_1 and F_2 are non-degenerate discrete distributions whose c.d.f. *values* meet at least once at a value that is strictly between 0 and 1. The interpretation of (23) is analogous.

Proposition 3

(a) (22) implies $\lim_{\rho\to 1} g'(\rho) = \infty$. (23) implies $\lim_{\rho\to -1} g'(\rho) = \infty$.

(b) Assume F_1 and F_2 have finite support. If (22) fails, then $\lim_{\rho\to 1} g'(\rho) = 0$. If (23) fails, then $\lim_{\rho\to -1} g'(\rho) = 0$.

(c) Analogs of (a) and (b), obtained by replacing g' by g'_{L}, hold.

Proof. We use well-known properties of ϕ_ρ as $|\rho| \to 1$. If $y = x$, then $\lim_{\rho\to 1} \phi_\rho(x,y) = \infty$. Analogously, if $y = -x$, then $\lim_{\rho\to -1} \phi_\rho(x,y) = \infty$. For all (x,y) that lie outside the lines $y = x$ and $y = -x$, we have $\lim_{\rho\to \pm 1} \phi_\rho(x,y) = 0$. Condition (22) implies that there exist i and j with finite $z_{1,i} = z_{2,j}$ and with $p_{1,i+1}\ p_{2,j+1} > 0$. Then $g'(\rho) \geq p_{1,i+1}\ p_{2,j+1}\ \phi_\rho(z_{1,i},\ z_{2,j}) \to \infty$ as $\rho \to 1$. Similarly, (23) implies that there exist i and j with finite $z_{1,i} = -z_{2,j}$ and with $p_{1,i+1}\ p_{2,j+1} > 0$, which gives $g'(\rho) \to \infty$ as $\rho \to -1$. This completes the proof of part (a). For part (b), there are only finitely many terms, so the failure of (22) implies that all finite pairs $(z_{1,i},\ z_{2,j})$ lie outside the line $y = x$; as $\rho \to 1$, each of the finitely many terms in (11) converges to zero, yielding $g'(\rho) \to 0$. The result as $\rho \to -1$ follows analogously. The above arguments remain intact if we replace g' by g'_{L}; this proves part (c). □

2.3 Approximating g' when the support is unbounded

For the case where one or both marginals have unbounded support, we propose approximate computation of the derivative g' via truncation of (11), provide a bound on the truncation error, and outline the computation. This supports the approximate method detailed in Section 3.2. We discuss the case where both marginals have unbounded support; straightforward modifications apply otherwise.

We rewrite (11) as

$$g'(\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \sum_{i=-\infty}^{\infty} p_{1,i+1}\phi(z_{1,i})S_i, \tag{24}$$

where

$$S_i = \sum_{j=-\infty}^{\infty} p_{2,j+1}\phi\left(\frac{z_{2,j}-\rho z_{1,i}}{\sqrt{1-\rho^2}}\right). \tag{25}$$

Our bound of the upper tail of S_i is based on the observation that $\phi\left((z_{2,j}-\rho z_{1,i})/\sqrt{1-\rho^2}\right)$ is decreasing as j increases beyond $j^*(i)$, where $j^*(i) = \min\{j : z_{2,j} \geq \rho z_{1,i}\}$. This yields

$$\sum_{j=k+1}^{\infty} p_{2,j+1}\phi\left(\frac{z_{2,j}-\rho z_{1,i}}{\sqrt{1-\rho^2}}\right) \leq (1-f_{2,k})\phi\left(\frac{z_{2,k}-\rho z_{1,i}}{\sqrt{1-\rho^2}}\right) \quad \text{for any } k \geq j^*(i). \tag{26}$$

The lower tail is bounded similarly:

$$\sum_{j=-\infty}^{k-1} p_{2,j+1}\phi\left(\frac{z_{2,j}-\rho z_{1,i}}{\sqrt{1-\rho^2}}\right) \leq f_{2,k-1}\phi\left(\frac{z_{2,k}-\rho z_{1,i}}{\sqrt{1-\rho^2}}\right) \quad \text{for any } k \leq j^*(i), \tag{27}$$

because $\phi\left((z_{2,j}-\rho z_{1,i})/\sqrt{1-\rho^2}\right)$ is decreasing as j decreases beyond $j^*(i)$. A similar approach allows bounding the tails of the summation in (24). Observe that $S_i \leq \phi(0)$ for all i and $\phi(z_{1,i})$ is decreasing as i increases beyond i^*, where $i^* = \min\{i : z_{1,i} \geq 0\}$. This yields

$$\sum_{i=k+1}^{\infty} p_{1,i+1}\phi(z_{1,i})S_i \leq \phi(0)\phi(z_{1,k})(1-f_{1,k}) \quad \text{for any } k \geq i^*. \tag{28}$$

Similarly,

$$\sum_{i=-\infty}^{k-1} p_{1,i+1}\phi(z_{1,i})S_i \leq \phi(0)\phi(z_{1,k})f_{1,k-1} \quad \text{for any } k \leq i^*. \tag{29}$$

Select small real numbers $\epsilon_1 > 0$ and $\epsilon_2 > 0$. We truncate the summation in (24), keeping terms between the indices

$$\begin{aligned} i^- := i^-(\epsilon_1) &:= \max\{k : k \le i^*,\ \phi(0)\phi(z_{1,k})f_{1,k-1} \le \epsilon_1 2\pi\sqrt{1-\rho^2}\}, \\ i^+ := i^+(\epsilon_1) &:= \min\{k : k \ge i^*,\ \phi(0)\phi(z_{1,k})(1-f_{1,k}) \le \epsilon_1 2\pi\sqrt{1-\rho^2}\}. \end{aligned} \tag{30}$$

For i in this finite range, we truncate the summation in (25), keeping terms between the indices

$$\begin{aligned} j^-(i) &= \max\{k : k \le j^*(i),\ p_{1,i+1}\phi(z_{1,i})f_{2,k-1}\phi((z_{2,k}-\rho z_{1,i})/\sqrt{1-\rho^2}) \le \epsilon_2\}, \\ j^+(i) &= \min\{k : k \ge j^*(i),\ p_{1,i+1}\phi(z_{1,i})(1-f_{2,k})\phi((z_{2,k}-\rho z_{1,i})/\sqrt{1-\rho^2}) \le \epsilon_2\}. \end{aligned} \tag{31}$$

(Note the truncation indices depend on ρ; our notation does not emphasize this). Define the finite-term approximation of g',

$$\tilde{g}'(\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \sum_{i=i^-}^{i^+} p_{1,i+1}\phi(z_{1,i}) \sum_{j=j^-(i)}^{j^+(i)} p_{2,j+1}\phi\left(\frac{z_{2,j}-\rho z_{1,i}}{\sqrt{1-\rho^2}}\right). \tag{32}$$

The bounds stated in (26), (27), (28) and (29) easily imply the following result.

Proposition 4 *We have*

$$\tilde{g}'(\rho) \le g'(\rho) \le \tilde{g}'(\rho) + \epsilon(\rho) \tag{33}$$

where $\epsilon(\rho) = 2\epsilon_1 + 2(i^+(\epsilon_1) - i^-(\epsilon_1) + 1)\epsilon_2$.

Remark 1 We outline an implementation for computing $\tilde{g}'(\rho)$ and $\epsilon(\rho)$. In a first outer `until` block, i increases from i^* until i^+ is found; for each fixed i in this range, j first increases from $j^*(i)$ until $j^+(i)$ is found (an `until` block nested inside the outer block); then, similarly, j decreases from $j^*(i)$ until $j^-(i)$ is found. A second outer `until` block is analogous to the first outer block: i decreases from i^* until i^- is found. The work of this algorithm is $O\left(\sum_{i=i^-}^{i^+}(j^+(i) - j^-(i))\right)$. This work and the size of the error bound $\epsilon(\rho)$ are unknown a priori in terms of ϵ_1 and ϵ_2; they are both determined during the process of approximating $g'(\rho)$.

2.4 Uniform convergence to the continuous-marginals rank correlation

This section establishes a convergence result relating the rank-correlation function under discrete marginals to the rank-correlation function for continuous marginals, i.e., r_{C} in (6), in a limit we will make precise. Let $(X_{1,n}, X_{2,n})$, $n = 1, 2, \ldots$ be a sequence of pairs of discrete random variables; write $p_{l,j,n}$ for the probability mass corresponding to the j-th mass point of the l-th marginal ($l = 1, 2$) in the n-th pair, and denote by $F_{1,n}$ and $F_{2,n}$ the associated c.d.f.'s in the n-th pair. Write $r_n(\rho) = \mathrm{Corr}(F_{1,n}(X_{1,n}), F_{2,n}(X_{2,n}))$,

where $(X_{1,n}, X_{2,n})$ has marginals $F_{1,n}$ and $F_{2,n}$ and bivariate dependence as in (2) with $\rho = \mathrm{Corr}(Z_1, Z_2)$. To capture the idea that discreteness vanishes in the limit, let $m_{l,n} = \max_j p_{l,j,n}$ and assume

$$\lim_{n\to\infty} m_{l,n} = 0 \quad \text{for } l = 1, 2. \tag{34}$$

We now state an asymptotic upper bound on the L_∞-distance between r_n and r_C that vanishes in the limit as $n \to \infty$.

Proposition 5 *If (34) holds, then*

$$\limsup_{n\to\infty} \sup_{\rho\in[-1,1]} \frac{|r_n(\rho) - r_\mathrm{C}(\rho)|}{m_{1,n} + m_{2,n}} \le 42, \tag{35}$$

and thus $\sup_{\rho\in[-1,1]} |r_n(\rho) - r_\mathrm{C}(\rho)|$ *converges to 0 as* $n \to \infty$.

Proof. For $l = 1, 2$, define the composite functions $h_{l,n} = F_{l,n} \circ F_{l,n}^{-1}$. Each $F_{l,n}(X_{l,n})$ has distribution equal to that of $h_{i,n}(U)$, where U is uniformly distributed on (0,1). The key behind the proof is that $|h_{l,n}(u) - u| \le m_{l,n}$ for all $0 \le u \le 1$. Write $\mu_{l,n} = \mathbb{E}[F_{l,n}(X_{l,n})]$, $\sigma^2_{l,n} = \mathrm{Var}[F_{l,n}(X_{l,n})]$, and $g_n(\rho) = \mathrm{Cov}[F_{1,n}(X_{1,n}), F_{2,n}(X_{2,n})]$. We will use repeatedly below the inequality $|x_1y_1 - x_2y_2| \le |y_1 - y_2| + |x_1 - x_2|$ for any $0 \le x_1, x_2, y_1, y_2 \le 1$. Using (6) and this inequality, we have

$$\begin{aligned}
|r_n(\rho) - r_\mathrm{C}(\rho)| &= \left| \frac{g_n(\rho) - \mu_{1,n}\mu_{2,n}}{\sigma_{1,n}\sigma_{2,n}} - \frac{g_\mathrm{C}(\rho) - 1/4}{1/12} \right| \\
&= \Bigg| [g_n(\rho) - \mu_{1,n}\mu_{2,n}] \left(\frac{1}{\sigma_{1,n}\sigma_{2,n}} - 12 \right) \\
&\qquad\qquad +12\,[g_n(\rho) - \mu_{1,n}\mu_{2,n} - g_\mathrm{C}(\rho) + 1/4] \Bigg| \\
&\le (|g_n(\rho)| + \mu_{1,n}\mu_{2,n}) \left| \frac{12\sigma_{1,n}\sigma_{2,n} - 1}{\sigma_{1,n}\sigma_{2,n}} \right| \\
&\qquad\qquad +12\,(|g_n(\rho) - g_\mathrm{C}(\rho)| + |\mu_{1,n}\mu_{2,n} - 1/4|)\,. \qquad (36)
\end{aligned}$$

We now find asymptotic upper bounds for each of the terms in (36). We have

$$|\mu_{l,n} - 1/2| = \left| \int_0^1 (h_{l,n}(u) - u)du \right| \le \int_0^1 |h_{l,n}(u) - u|du \le m_{l,n},$$

so $\lim_{n\to\infty} \mu_{l,n} = 1/2$ for $l = 1, 2$ and $\limsup_{n\to\infty} |\mu_{1,n}\mu_{2,n} - 1/4|/(m_{1,n} + m_{2,n}) \le 1/2$. Writing $\sigma^2_{l,n} = \int_0^1 \left[(h_{l,n}(u) - u) + (u - \frac{1}{2}) + (\frac{1}{2} - \mu_{l,n})\right]^2 du$ and integrating the expanded square, it is easy to see that

$$\left| \sigma^2_{l,n} - \frac{1}{12} \right| \le m^2_{l,n} + m^2_{l,n} + 4m_{l,n} \int_0^1 \left| u - \frac{1}{2} \right| du + 2m^2_{l,n} = m_{l,n} + 4m^2_{l,n}, \tag{37}$$

proving that $\lim_{n\to\infty} \sigma_{l,n}^2 = 1/12$ for $l = 1, 2$. The Cauchy-Schwartz inequality yields $\sup_\rho |g_n(\rho)| \leq \sigma_{1,n}\sigma_{2,n}$, so $\limsup_{n\to\infty} \sup_\rho |g_n(\rho)| \leq 1/12$. Furthermore,

$$\limsup_{n\to\infty} \frac{\left|\sqrt{\sigma_{1,n}^2\sigma_{2,n}^2} - \frac{1}{12}\right|}{m_{1,n} + m_{2,n}} \leq \limsup_{n\to\infty} \frac{6\left|\sigma_{1,n}^2\sigma_{2,n}^2 - \left(\frac{1}{12}\right)^2\right|}{m_{1,n} + m_{2,n}} \leq \frac{1}{2};$$

in the above, the first inequality follows from a Taylor expansion of $\sqrt{x}$ about $1/12$ with remainder term involving the first derivative, and the second inequality follows from (37). Finally,

$$\begin{aligned}
&\sup_\rho |g_n(\rho) - g_{\mathrm{C}}(\rho)| \\
&\quad = \sup_\rho \left| \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [h_{1,n}(\Phi(x_1))h_{2,n}(\Phi(x_2)) - \Phi(x_1)\Phi(x_2)]\, \phi_\rho(x_1, x_2)dx_1dx_2 \right| \\
&\quad \leq \sup_\rho \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\sup_{(x_1,x_2)\in\mathbb{R}^2} |h_{1,n}(\Phi(x_1))h_{2,n}(\Phi(x_2)) - \Phi(x_1)\Phi(x_2)| \right) \phi_\rho(x_1, x_2)dx_1dx_2 \\
&\quad \leq \sup_\rho \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (m_{1,n} + m_{2,n})\, \phi_\rho(x_1, x_2)dx_1dx_2 \\
&\quad = m_{1,n} + m_{2,n}. \qquad (38)
\end{aligned}$$

The result (35) follows from the asymptotic bounds established for each of the terms in (36). □

For n large, (35) and (6) imply the approximate bound $\sup_{\rho\in[-1,1]} |r_n(\rho) - (6/\pi) \arcsin(\rho/2)| \leq 42(m_{1,n} + m_{2,n})$. In our examples in Section 4, this bound was too large to ensure that $r_X(2\sin(\pi\tilde{r}/6))$ is sufficiently close (for our purposes) to $\tilde{r} = r_{\mathrm{C}}(2\sin(\pi\tilde{r}/6))$. Had the bound been small enough, that would have made our nearly-exact solution methods less intereresting, because the bound by itself would have ensured that $2\sin(\pi\tilde{r}/6)$ is a sufficiently accurate answer. Of course, better bounds than ours may still act in the same way, i.e., as guarantors of the accuracy of $2\sin(\pi\tilde{r}/6)$ as an approximation to the exact solution. Regardless of the bound's effectiveness in our examples, the proof adds to our intuition; it suggests, for example, that the approximation's effectiveness hinges on *both* marginals (as opposed to only one) being nearly continuous.

3 Solution methods

We detail methods for solving either of the two versions of the correlation-matching problem. Our discussion focuses on the rank-correlation variant for reasons given earlier. Assume that we are given a target $\tilde{r} \in (r_X(-1),\ r_X(1))$ and want to compute the value $r_X^{-1}(\tilde{r})$, i.e., the unique solution of (3). A *zero* of a function f is a value ρ such that $f(\rho) = 0$.

To conform with standard algorithms for solving a single equation, which typically seek a zero of an appropriate function, define $f(\rho) = g(\rho) - \mu_1\mu_2 - \tilde{r}\sigma_1\sigma_2$ and note that f has derivatives identical to those of g and that $f(\rho) < (>)\ 0$ if and only if $r_X(\rho) < (>)\ \tilde{r}$. Thus, finding the solution of (3) is equivalent to finding the unique zero of f.

Section 3.1 treats the case where both marginals have finite support. Infinite supports are addressed in Section 3.2, which offers an approximate solution method and a bound on its error.

3.1 Discrete marginals with finite support

If n_i is the number of support points of marginal i, then (10) and (11) imply that the computational work for each evaluation of g (equivalently, f) or of its derivative $f' = g'$ is proportional to $n = n_1 n_2$, the number of terms in the double sums. The proportionality constants may differ substantially between g and g'.

In what follows, we first explain how we compute g and g', then we define three algorithms to find a root of f. The first algorithm uses only evaluations of g and not its derivative, the second integrates f' until the integral reaches zero, and the third is a variant of the Newton-Raphson iterative method to find a root of f.

Evaluation of g and g'. For the evaluation of g, we employ (10) instead of (9), because the literature emphasizes the computation of the bivariate normal integral in the former expression. We considered several methods for evaluating $\bar{\Phi}_\rho(x, y)$, a function of ρ, x, and y, for which no analytic expression is available. Algorithm 462 in Donnelly (1973) implements the method developed in Owen (1956), which expresses $\bar{\Phi}_\rho$ in terms of the functions Φ and $T(h, a)$, where the latter is the area (integral) of an uncorrelated bivariate standard normal distribution (zero means, unit variances) over the subset of the (x, y)-plane contained between $y = ax$ and $y = 0$ and to the right of $x = h$. The function $T(h, a)$ is expressed (and computed efficiently) as a series. A second class of methods exploits property (13) and computes $\bar{\Phi}_\rho(x, y)$ by numerical integration with respect to the correlation. More precisely, $\bar{\Phi}_\rho(x, y)$ is computed as $\bar{\Phi}_s(x, y) + Q$, where: $s = 0$ or $\text{sign}(\rho)$ (when $|\rho|$ is under and above a certain threshold, respectively); $\bar{\Phi}_0(x, y) = \Phi(-x)\Phi(-y)$; $\bar{\Phi}_1(x, y) = \Phi(-\max(x, y))$; $\bar{\Phi}_{-1}(x, y) = \max(0, \Phi(-x) - \Phi(y))$; and $Q = \int_s^\rho \phi_t(x, y)dt$ is computed by numerical integration. This approach is detailed in Drezner and Wesolowsky (1989) and Genz (2004), which focus on moderate accuracy (6-7 decimals) and high accuracy (15 decimals), respectively. For 15-decimal precision, we compared Algorithm 462 to the method of Genz (2004). For $\rho = -0.92, -0.54, -0.16, 0.22, 0.60, 0.98$, we sampled one million pairs (x, y) uniformly in the square $[-3,\ 3]^2$; the observed ratios of CPU times (Algo. 462 to Genz) were about 0.4, 0.6, 1.1, 1.1, 0.6, and 0.7, respectively. In 7-decimal precision, and for the same set of ρ values, the CPU time ratios of Algorithm 462 to

the method of Drezner and Wesolowsky (1989) were about 0.7, 1.3, 1.3, 1.3, 1.3, 0.7. Comparing the 7- to 15-decimal accuracy versions of Algorithm 462, we observed a ratio of CPU total times (sums over 6 evaluations for the values of ρ above) of about 0.67. For all subsequent work, we chose to evaluate f via Algorithm 462 of Donnelly (1973) with 15 decimal digits of accuracy.

Computing the derivative $g'(\rho)$ is easier, because there is an analytic expression for $\phi_\rho(x, y)$. We just use it and sum up the terms. In a preliminary test, we estimated the ratio of work (CPU time) needed to compute $g(\rho)$ over the work needed to compute the derivative $g'(\rho)$ at about 12. This was based on all calls made to these functions when solving the problem $\tilde{r} = 0.90$ in the nearly-continuous negative binomial case shown at the bottom panel of Table 1, Section 4. We feel that this number is fairly representative because the points $z_{i,k} = \Phi^{-1}(f_{i,k})$, $k = 1, 2, \ldots$, provide a good coverage of the normal density for each i.

Method NI1: Root bracketing without derivatives. This first method assumes no knowledge of derivatives of f and serves as the benchmark against which we compare the speed and accuracy of other methods. We know that the zero of f is contained in $[-1, 0]$ if $\tilde{r} < 0$, and is $[0, 1]$ if $\tilde{r} > 0$; this follows from parts 1 and 2 of Theorem 1 and the Intermediate Value Theorem. *Root-bracketing* methods maintain a *bracket*; this is an interval with endpoints b and c such that $f(b)$ and $f(c)$ are of opposite sign, so the interval must contain the root. One such method is *bisection*, which is iterative and halfs the bracket length at each iteration. Root accuracy is usually controlled by a *tolerance* $\epsilon > 0$: if b is the better root estimate among the bracket endpoints, (i.e., $|f(b)| < |f(c)|$), then it is returned as the root on the first iteration such that either $f(b) = 0$ (in the floating-point representation) or $|b-c| \leq \epsilon$. By the definition of bracket, this guarantees that b is within ϵ of the root. According to Press et al. (1992), Procedure *zero* in Brent (1971) (called *Brent's method* for short), is "the method of choice for general one-dimensional root finding where a function's values only (and not its derivative) are available." This method combines root bracketing, bisection, and *inverse quadratic interpolation*, which uses three prior root estimates to fit an inverse quadratic function (ρ as a quadratic function of $f(\rho)$) whose value at $f(\rho) = 0$ is taken as the next estimate of the root. This is what we have used in our experiments.

Method NI2: Finding a root of f by numerically integrating its derivative. This method is summarized as follows.

1. Start at some initial value ρ_0 and evaluate $f(\rho_0)$, as described in the previous subsection.
2. Select an *integration grid* $S = \{\rho_0, \rho_1, \rho_2, \ldots\}$, which is a sequence of increasing (decreasing) values depending on whether $f(\rho_0) < (>)\ 0$, and such that: if $\tilde{r} > 0$ and

$f(\rho_0) < (>)\ 0$, then 1 (0) is an accumulation point of S; if $\tilde{r} < 0$ and $f(\rho_0) < (>)\ 0$, then 0 (-1) is an accumulation point of S.

3. Compute estimates $\hat{f}(\rho_k)$ of $f(\rho_k)$ for $k = 1, 2, \ldots$ by numerically integrating its derivative g'. Stop at the smallest k, say K, such that $\hat{f}(\rho_k) > (<)\ 0$, respectively. By construction, the interval $[\rho_{K-1}, \rho_K]$ contains a zero of $\hat{f}$.
4. Compute the approximation $\bar{\rho}$ of the zero via polynomial interpolation of $\hat{f}$ over $[\rho_{K-\ell}, \rho_K]$, where ℓ is a small positive integer. For example, for linear interpolation, take $\ell = 1$ and output the unique $\bar{\rho}$ satisfying $(\bar{\rho} - \rho_{K-1})/(\rho_K - \rho_{K-1}) = -\hat{f}(\rho_{K-1})/\left[\hat{f}(\rho_K) - \hat{f}(\rho_{K-1})\right]$.

We now discuss the selection of integration rule, the choice of sequence S, and the method's accuracy. We discuss the case $\tilde{r} > 0$ and $\hat{f}(\rho_0) < 0$; the other three cases are similar.

Two effective classes of integration rules over a finite interval $[a, b]$ are the Gaussian and Newton-Cotes quadrature rules (Stoer and Bulirsch, 1980). These rules evaluate the integrand at a finite set of points in $[a, b]$ and compute a weighted sum of these evaluations. In theory, the *Gaussian rules* (Stoer and Bulirsch, 1980, sec. 3.6) give better accuracy than the Newton-Cotes rules for a given number n of evaluation points: they integrate exactly all polynomials of degree less than $2n$. However, if we change a or b slightly, for fixed n, all the evaluation points must change. In our context, since the integration interval changes at each step of the root-finding process, the Gaussian rule on $[0, \rho_k]$ cannot reuse any of the evaluation points of the rule on the previous interval $[0, \rho_{k-1}]$. With *Newton-Cotes rules* (Stoer and Bulirsch, 1980, sec. 3.1), the integral over $[a, b]$ is approximated as a sum of approximations of the integral over the pieces of a partition of $[a, b]$ (see below), and it possible to select the integration grid in our procedure in a way that the evaluation points for $[0, \rho_{k-1}]$ are reused for $[0, \rho_k]$. Thus, from an efficiency standpoint, Newton-Cotes rules are more suitable in our root-finding context.

A well-known special case of a Newton-Cotes rule is Simpson's rule (Stoer and Bulirsch, 1980, pp. 119-120). For this rule, we select a finite sequence S consisting of $\rho_k = \rho_0 + 2kh$ for $k = 0, 1, 2, \ldots, m$, where $h > 0$ is a *step size* and m is such that $1 - 2h < \rho_m < 1$. In our implementation, we first select ρ_m close to 1 ($\rho_m = 1 - \delta$ for some small $\delta \geq 0$) and then select h and m (a positive integer) such that $|1 - \delta - \rho_0| = 2hm$. Proposition 3 implies $\lim_{\rho \to 1} g'(\rho)$ is either 0 or infinity. In the latter case, one can expect numerical integration rules as ours to loose accuracy as the points of interest approach 1. One may often find via (22) what the limit is. Lacking this knowledge, a defensive choice is to avoid setting δ unnecessarily small, depending on the user's root-accuracy need. An example illustrating this difficulty near 1 is given in Section 4. The Simpson estimate of the definite integral

$\int_{\rho_0}^{\rho_0+2kh} g'(t)dt$ is computed recursively by setting $I_0 = 0$ and

$$I_k = I_{k-1} + \frac{h}{3}\left(g'(\rho_0 + 2kh - 2h) + 4g'(\rho_0 + 2kh - h) + g'(\rho_0 + 2kh)\right).$$

This gives the estimate $\hat{f}(\rho_k) = f(\rho_0) + I_k$, whose error will be discussed later.

If the stopping condition in step 3 is not met after m steps for the m selected at the outset (that is, $\hat{f}(\rho_m)$ has the same sign as $\hat{f}(\rho_0)$), then we continue integrating over a new grid defined to the right of the last point of the previous grid, recursively, if necessary, until a stopping condition as in step 3 is met. That is, the mention in step 2 of an *infinite* sequence S only serves to allow an input $\tilde{r}$ that is arbitrarily close to $r_X(1)$ or $r_X(-1)$.

We consider two variants of algorithm NI2, defined according to how ρ_0 is selected: Variant NI2A sets $\rho_0 = 2\sin(\pi\tilde{r}/6)$, which is a natural estimate of the root because it becomes exact in the limit where discreteness disappears (see Proposition 5 and part 3 of Theorem 1). Variant NI2B sets $\rho_0 = 0$. The motivation for NI2A is to try to minimize the length of the integration interval $[\rho_0, \rho_K]$, and thus the number $N_{g'}$ of evaluations of the function g'. On the other hand, it requires one (costly) evaluation of $f(\rho_0)$ in Step 1. Variant NI2B eliminates the cost of this evaluation, because we know $f(0) = -\tilde{r}\sigma_1\sigma_2$, but $N_{g'}$ is typically larger because we must integrate over a longer interval. If the root does not exceed the value ρ_m selected at the outset, then NI2 requires $N_{g'} = 1 + 2\lceil|r_X^{-1}(\tilde{r}) - \rho_0|/2h\rceil$ evaluations of the function g', where h is the value selected at the outset. Which variant will be faster depends on: (i) the ratio of work needed to compute g relative to g'; (ii) the distance $|r_X^{-1}(\tilde{r}) - \rho_0|$; and (iii) the desired accuracy; lower accuracy allows larger h and thus smaller $N_{g'}$.

Method NI3: Hybrid of Newton-Raphson and bisection. Our third algorithm is a modified version of the Newton-Raphson method. This method would produce a sequence of root estimates $\rho_{k+1} = \rho_k - f(\rho_k)/f'(\rho_k)$ for $k = 0, 1, 2, \ldots$, where $-f(\rho_k)/f'(\rho_k)$ is a correction term such that the new root estimate is the zero of the linear function with value $f(\rho_k)$ and slope $f'(\rho_k)$ at abscissa ρ_k. We need to protect against the possibility that at two subsequent iterations k and $k+1$, the correction terms cancel each other and neither ρ_k nor ρ_{k+1} is a root; that is, $f(\rho_k)/f(\rho_k) + f(\rho_{k+1})/f(\rho_{k+1}) = 0$, $f(\rho_k) \neq 0$, and $f(\rho_{k+1}) \neq 0$; in this case, the recursion enters an infinite cycle without ever finding the root ($\rho_{k+2j} = \rho_k$ for all positive j); this is illustrated in Press et al. (1992, Figure 9.4.3). We protect as proposed in Press et al. (1992, routine `rtsafe`, pp. 366–367); this algorithm maintains a root estimate and a bracket formed by the last two root estimates; if the Newton step starting from the current root estimate would fall outside the current bracket or if the current bracket length is more than half the previous bracket length, then the next root estimate is the bracket's midpoint; otherwise, the next root estimate is found by the Newton step. Root accuracy is controlled by a tolerance ϵ as in NI1. This method has good

convergence properties near the root (Press et al., 1992, pages 364-365), so it is particularly attractive when high accuracy is sought. The initial bracket is $[-1, 0]$ if $\tilde{r} < 0$, and $[0, 1]$ if $\tilde{r} > 0$. Our initial root estimate is $\rho_0 = 2\sin(\pi\tilde{r}/6)$; this value is likely to be closer to the root than other uninformative values, e.g., the midpoint of the initial bracket. It is easy to show that the bracket is at least halved over any *two* successive iterations (Press et al. (1992) do not state this); thus, the number of iterations never exceeds $2\lceil\log_2(1/\epsilon)\rceil$, and it is potentially smaller, depending on the Newton steps' effectiveness.

Controlling the accuracy. Efficient algorithms are known for computing the bivariate normal integral $\bar{\Phi}_\rho$ to negligible error (this was discussed earlier); this allows efficiently computing g to negligible error. In view of this, the methods we discussed fall into two classes that should be contrasted: classical root finding (NI1, NI3) versus approximate root finding via integration and interpolation (NI2). In general, none of these methods can provide a guarantee on rank-correlation error (a known multiple of $|f(\bar{\rho})|$, where $\bar{\rho}$ is the estimated root) unless a global bound on the slope of f is known. Classical root-finding methods, however, do deliver a value to within a specified distance from the true root. For the approximate root-finding methods, we do not have integration-error bounds and consequently we offer no guarantee either on root error or on rank-correlation error, regardless of how much work one does. (Note, however, that global bounds on higher-order derivatives of g can be obtained by straightforward derivations and arguments paralleling (16); this would yield such integration-error bounds.) Thus, the approximate root finding approach—as developed here—can be attractive only in special settings, namely: (1) solution speed is more important than a root-accuracy guarantee; or (2) classical root finding is too complicated to implement, e.g., because a good code for computing Φ_ρ is unavailable.

Worst-case work comparison as required accuracy increases. We focus on the rank-correlation error at the estimated root, $|r_X(\bar{\rho}) - \tilde{r}|$, and assume a requirement that it should not exceed $\epsilon > 0$. We explain that if one views the error in evaluating g as negligible, then one should expect NI2 to require more work than NI3 or the bisection method in the limit as $\epsilon \to 0$. In standard polynomial interpolation, function values are known *exactly* at the interpolation points; in this case, a bound on the error (at any point inside the interpolation interval) is given in Stoer and Bulirsch (1980, Theorem 2.1.4.1). *If the integration error was zero at all interpolation points*, this result would imply that the error is of order $O(h^{\ell+1})$ when an order-ℓ interpolating polynomial is used (the error may of course be zero, but that would seem to be a fortunate coincidence). Thus, we can expect the error to decrease at the rate m^{-k} for some positive integer k that depends on the particular Newton-Cotes rule and ℓ. The *worst-case* number of evaluations of g' for NI2 is $pm+1$, where p is a positive integer that depends on the Newton-Cotes rule; for Simpson's rule, we have $p = 2$. To keep the error at most ϵ, this number must grow as $O(\epsilon^{-1/k})$. To allow comparison to NI3 and bisection, we consider a user of these methods that selects a

tolerance ϵ/M, where $M := \sup_{\rho\in\mathcal{I}} |g'(\rho)| < \infty$, where $\mathcal{I}$ is the initial bracket; this ensures that the error is at most ϵ. The bisection method requires $\lceil \log_2(M/\epsilon) \rceil$ evaluations of g. NI3 requires $2\lceil \log_2(M/\epsilon) \rceil$ iterations in the worst case. In conclusion, *if high accuracy is required*, then NI3 (or bisection) are preferred to NI2, because they are likely to require less work.

Linear correlations. For the linear correlation matching problem, all three methods extend immediately. The initial bracketing intervals are identical; we simply replace the functions g and g' by their counterparts g_{L} and g'_{L} stated in Section 2.2. To get a nonzero starting point for NI2 or NI3, we can invert (6), despite the fact that this has no theoretical basis and that it may be a poor choice relative to crude estimates such as the midpoint of the initial bracket, as suggested by the discussion following (4).

3.2 Discrete marginals with infinite (or large) support

If one of the marginals has infinite support, all quantities involved in the definition of $f(\rho)$, namely, μ_l and σ_l for $l = 1, 2$, and $g(\rho)$, involve infinite series; in general, exact computations appear to be impossible—we are not aware of exact formulae, even if the marginals belong to the well-known classes. Approximating $g(\rho)$ (for arbitrary ρ) is the main difficulty, because if one were to truncate the series (10) to a finite number of terms, it would be difficult to bound the error. Approximating the constants μ_l and σ_l is easier, as we will explain. In view of this, method NI2B stands out, because it is the only one among those in Section 3.1 that does not require evaluating $g(\rho)$. Thus, we adapt method NI2B as follows: (i) in the integration (step 3 of method NI2), we replace g' by its approximation $\tilde{g}'$ established in Section 2.3; and (ii) we replace μ_l and σ_l by approximations defined below (the μ_l are involved indirectly via σ_l).

It is straightforward to approximate μ_l and σ_l by truncating the associated series; error bounds are easily obtained and stated in the proof of Proposition 6 below. Select small real numbers $\eta_l > 0$. For $l = 1, 2$, define $k_l^+ = \min\{k : \sum_{j=k+1}^{\infty} p_{l,j} \le \eta_l\}$ and $k_l^- = \max\{k : \sum_{j=-\infty}^{k-1} p_{i,j} \le \eta_l\}$. Define $\tilde{\mu}_l = \sum_{j=k_l^-}^{k_l^+} p_{i,j} f_{i,j}$ and $\tilde{\sigma}_l^2 = \sum_{j=k_l^-}^{k_l^+} p_{i,j} f_{i,j}^2 - \tilde{\mu}_l^2$ as approximations of μ_l and σ_l^2, respectively.

We now define the adaptation of NI2B. We assume that $\rho_0 = 0$ and that we use the sequence S with the Newton-Cotes integration rule. The estimates of $f(\rho_k)$ are $\tilde{f}(0) = -\tilde{r}\tilde{\sigma}_1\tilde{\sigma}_2$ (since $r_X(0) = 0$) and $\tilde{f}(\rho_k) = \tilde{f}(0) + I(\rho_k; \tilde{g}')$ for $k = 1, 2, ...$, where $I(\rho_k; \tilde{g}')$ is the estimate of $\int_0^{\rho_k} g'(t)dt$ via a Newton-Cotes formula applied to $\tilde{g}'$ in (32).

To bound the error in rank correlation at the estimated root, $|r_X(\bar{\rho}) - \tilde{r}|$, define: $I(\rho_k; \epsilon)$ is the Newton-Cotes estimate of $\int_0^{\rho_k} \epsilon(t)dt$, where $\epsilon(\rho)$ is defined following (33); $I(\rho_k; g')$ is the Newton-Cotes estimate of $\int_0^{\rho_k} g'(t)dt$, which will not be explicitly computed, but is

involved in the bound; and write $\Delta_k = |\tilde{f}(\rho_k) - f(\rho_k)|$ for all k. Write K for the index in step 3 of NI2; note that $M_K := \sup_{\rho\in[\rho_{K-1},\rho_K)]} |\tilde{g}'(\rho)| \leq \sup_{|\rho|\leq 1-\delta} |g'(\rho)| < \infty$. The next result bounds the error, and finite support is a special case. The Remarks below discuss how one may reduce this bound.

Proposition 6

(a) Assume all integral estimates are based on Simpson's rule with $h = (1-\delta)/(2m)$, $\rho_0 = 0$, and $|\rho_m| = 1-\delta$ for some $\delta > 0$. Then

$$\Delta_k \leq \zeta(\eta_1, \eta_2) + |I(\rho_k; \epsilon)| + O(m^{-4}) \quad \textit{for any } \rho_k \in S, \tag{39}$$

where

$$\zeta(\eta_1, \eta_2) = |\tilde{r}| \left(\bar{\sigma}_1 \frac{2\eta_2[1 + 2(\tilde{\mu}_2 + \eta_2)]}{\tilde{\sigma}_2 + \underline{\sigma}_2} + \tilde{\sigma}_2 \frac{2\eta_1[1 + 2(\tilde{\mu}_1 + \eta_1)]}{\tilde{\sigma}_1 + \underline{\sigma}_1} \right),$$

$\underline{\sigma}_l := \sqrt{\tilde{\sigma}_l^2 - 2\eta_l[1 + 2(\tilde{\mu}_l + \eta_l)]}$ *and* $\bar{\sigma}_l := \sqrt{\tilde{\sigma}_l^2 + 2\eta_l[1 + 2(\tilde{\mu}_l + \eta_l)]}$ *for* $l = 1, 2$.

(b) For any $\bar{\rho} \in [\rho_{K-1}, \rho_K]$, we have

$$\begin{aligned} |r_X(\bar{\rho}) - \tilde{r}| &\leq \frac{|\tilde{f}(\rho_{K-1}) - \tilde{f}(\rho_K)| + \max(\Delta_{K-1}, \Delta_K)}{\underline{\sigma}_1 \underline{\sigma}_2} \\ &\leq \frac{M_K/m + \zeta(\eta_1, \eta_2) + |I(\rho_K; \epsilon)|}{\underline{\sigma}_1 \underline{\sigma}_2} + O(m^{-4}). \end{aligned} \tag{40}$$

Proof. We have

$$\begin{aligned} |\tilde{f}(\rho_k) - f(\rho_k)| &= \left| \tilde{f}(0) + I(\rho_k; \tilde{g}') - \left(f(0) + \int_0^{\rho_k} g'(s)ds \right) - I(\rho_k; g') + I(\rho_k; g') \right| \\ &\leq |\tilde{f}(0) - f(0)| + |I(\rho_k; \tilde{g}') - I(\rho_k; g')| + \left| I(\rho_k; g') - \int_0^{\rho_k} g'(s)ds \right| \\ &= |\tilde{r}||\tilde{\sigma}_1\tilde{\sigma}_2 - \sigma_1\sigma_2| + |I(\rho_k; \epsilon)| + O(m^{-4}) \\ &\leq |\tilde{r}| \left(\sigma_1|\sigma_2 - \tilde{\sigma}_2| + \tilde{\sigma}_2|\sigma_1 - \tilde{\sigma}_1|\right) + |I(\rho_k; \epsilon)| + O(m^{-4}); \end{aligned} \tag{41}$$

step 2 is the triangle inequality; in step 3, we observe that $I(\rho_k; g') - I(\rho_k; \tilde{g}') = I(\rho_k; \epsilon)$ and that $\left|I(\rho_k; g') - \int_0^{\rho_k} g'(s)ds\right| \leq h^4|\rho_k g^{(5)}(\xi)|/180$ for some ξ with $|\xi| \leq \rho_k$, where $g^{(5)}$ is the fourth derivative of g' (Stoer and Bulirsch, 1980, p. 122), and finally note that $|g^{(5)}(\xi)| < \infty$, since $g^{(5)}$ is continuous on the closed interval $[-1+\delta, 1-\delta]$; step 4 is another application of the triangle inequality. It remains to bound σ_1 and $|\tilde{\sigma}_l - \sigma_l|$ for $l = 1, 2$. We have $|\tilde{\mu}_l - \mu_l| \leq 2\eta_l$ and $|\tilde{\sigma}_l^2 - \sigma_l^2| \leq 2\eta_l[1 + 2(\tilde{\mu}_l + \eta_l)]$ (proofs are easy and omitted), and thus

$$\underline{\sigma}_l \leq \sigma_l \leq \bar{\sigma}_l. \tag{42}$$

Thus

$$|\tilde{\sigma}_l - \sigma_l| = \frac{|\tilde{\sigma}_l^2 - \sigma_l^2|}{\tilde{\sigma}_l + \sigma_l} \leq \frac{2\eta_l[1 + 2(\tilde{\mu}_l + \eta_l)]}{\tilde{\sigma}_l + \underline{\sigma}_l}. \tag{43}$$

Combining (41), (42), and (43), we obtain (39). To prove (40), we note that $|r_X(\bar{\rho}) - \tilde{r}| = |f(\bar{\rho})|/(\sigma_1\sigma_2)$ and

$$\begin{aligned} |f(\bar{\rho})| &\leq \max(|f(\rho_{K-1})|, |f(\rho_K)|) \\ &\leq \max(|\tilde{f}(\rho_{K-1})| + \Delta_{K-1}, |\tilde{f}(\rho_K)| + \Delta_K) \\ &\leq |\tilde{f}(\rho_{K-1})| + |\tilde{f}(\rho_K)| + \max(\Delta_{K-1}, \Delta_K), \\ &= |\tilde{f}(\rho_{K-1}) - \tilde{f}(\rho_K)| + \max(\Delta_{K-1}, \Delta_K). \end{aligned} \tag{44}$$

Step 1 uses the monotonicity of f; step 2 uses the definition of Δ_k; the equality in step 4 holds because $\tilde{f}(\rho_{K-1})$ and $\tilde{f}(\rho_K)$ bracket zero, by construction. This proves the first inequality in (40). To get the second inequality in (40), we use the bound in (39), note that $|I(\rho_k;\epsilon)|$ are nondecreasing in k, and note that $|\tilde{f}(\rho_{K-1}) - \tilde{f}(\rho_K)| = \frac{h}{3}|\tilde{g}'[\rho_K - 2h] + 4\tilde{g}'[\rho_K - h] + \tilde{g}'[\rho_K]| \leq 2hM_K \leq M_K/m$. □

Remark 2 In the special case of finite support, (41) states that $\Delta_k = O(m^{-4})$ for all k. We obtain the rudimentary bound $|r_X(\bar{\rho}) - \tilde{r}| \leq M_K/(m\sigma_1\sigma_2) + O(m^{-4})$, which goes to zero as $m \to \infty$.

Remark 3 In the infinite support case, the first inequality in (40) combined with (39) yields the value $(|\tilde{f}(\rho_{K-1}) - \tilde{f}(\rho_K)| + \zeta(\eta_1, \eta_2) + |I(\rho_K;\epsilon)|)/(\underline{\sigma}_1\underline{\sigma}_2)$ as a computable *approximate* (heuristic) bound on the absolute error in the output correlation, because we dropped the $O(m^{-4})$ integration-error term. Contrary to the finite-support case, it is not enough to let $m \to \infty$ to guarantee that the rank correlation error goes to zero. One must additionally keep small the two new error terms, which may be done as follows. Controlling $\zeta(\eta_1, \eta_2)$ is straightforward by decreasing the η_i, $i = 1, 2$. Controlling $|I(\rho_k;\epsilon)|$ is somewhat complicated; recall the expression for the function $\epsilon(\rho)$ following (33) and note that $2(i^+(\epsilon_1) - i^-(\epsilon_1) + 1)\epsilon_2$ may increase as ϵ_1 decreases. In general, we may expect to reduce $\epsilon(\rho)$ (for any ρ) by appropriately decreasing ϵ_1 and/or ϵ_2 (at the expense of increased work). Also note that fixed ϵ_1 and decreasing ϵ_2 result in decreasing $\epsilon(\rho)$.

4 Numerical examples

We tried our solution methods on two sets of examples, in which the marginal have finite and infinite support, respectively. In our first set of examples, the two marginals are identical binomial distributions, denoted $\text{Bin}(n, p)$, with success probability $p = 1/2$ and varying number of trials n.

Our second set of examples is inspired from modeling the joint distribution of arrival counts to a call center over successive time periods in a day and is based on the case study in Avramidis et al. (2004). We are focusing on bivariate rank-correlation matching for (X_1, X_2), where X_1 and X_2 are the counts on the time periods (8:00am, 8:30am) and (8:30am, 9:00am), respectively. The negative binomial distribution provides a good fit to each marginal. Denote by NegBin(s, p) the negative binomial distribution with mean sp and variance $sp(1 + p)$. The parameters (s, p) of the two marginals estimated from the call center data set in that paper are $s_1 = 15.68$, $s_2 = 60.21$, $p_1 = 0.3861$, $p_2 = 0.6211$. The sample rank correlation between X_1 and X_2 is 0.43. For the correlation matching, we work with bounded (and finite) supports: we upper-bound the support of each marginal at the quantile of order $1 - 10^{-6}$, i.e, $x_l^* = F_l^{-1}(1 - 10^{-6})$, and reset the probability mass of x_l^* accordingly, for $l = 1, 2$. This may significantly impact the correlation relative to the unbounded marginals, but we did not attempt to bound this error. We create additional test problems as follows. In our experiments, we vary s to study the effect of "discreteness strength" on the NORTA correlation matching problem. We also vary the target correlation $\tilde{r}$.

In applications we have in mind, $\tilde{r}$ will be estimated from data; this means high accuracy (either in the root or in the rank correlation) is unlikely to be necessary. With this in mind, we employed NI1 and NI3 with tolerance 10^{-2} and 10^{-4}. Preliminary computations showed that in one of our examples the root is very close to 1; to avoid cumbersome implementations of NI2 that must refine the integration rule to the right of $1 - 2h$ (for the h of interest here), we set $\rho_m = 1 - \delta$ with $\delta = 10^{-4}$. To select the integration-grid spacing $2h$, let d denote the worst-case integration distance, so $d = |1 - \delta - \rho_0|$ if $\tilde{r} > 0$ and $\hat{f}(\rho_0) < 0$ or if $\tilde{r} < 0$ and $\hat{f}(\rho_0) > 0$; and $d = |\rho_0|$ otherwise. For NI2A, we set $2h$ to be as close as possible to 10^{-2}, i.e., $h = d/(2m)$, where $m = \max(1, [100d])$ and $[x]$ is the integer closest to x; this aims to make the accuracy (very roughly) comparable to that of NI1. For a sufficiently small m, NI2B will be faster than NI2A because it does not require the evaluation of $f(\rho_0)$, so with this in mind, we used NI2B with $m = 5$ (so $h = (1 - \delta)/10 \approx 0.2$). This aims towards fast execution achieved at the risk of loss of accuracy. We employ quadratic interpolation in step 4 unless $m = 2$, in which case linear interpolation applies.

Tables 1, 2, 3, and 4 summarize the results for methods NI1, NI2A, NI2B and NI3, respectively. Each of the six panels corresponds to a different pair of marginals; in each case, we give the defining parameters, the extreme correlations $r_X(-1)$ and $r_X(1)$ for these marginals, and the number of bivariate support points $n = n_1 n_2$, where n_i is the number of support points of marginal i. Each row corresponds to a problem instance created by additionally specifying the target $\tilde{r}$. For each problem instance, we report system-independent (method-dependent) measures of work: for NI1, the number N_1 of iterations of the root-bracketing algorithm and thus evaluations of g; for NI2, the number

$N_{g'}$ of evaluations of g'; for NI3, the number N_3 of iterations, and thus evaluations of each of g and g'. Additionally, we report: the computed root $\bar{\rho}$; the CPU time measured in seconds; the correlation $r_X(\bar{\rho})$; and the (absolute) relative error (error, for short) in induced correlation, $|r_X(\bar{\rho}) - \tilde{r}|/\tilde{r}$, shown as a percentage. When the target correlation is small, the reader may prefer to focus on absolute errors. All experiments were done on a 2.4 GHz AMD 64 bit-processor running Linux.

In all cases, NI1 and NI3 with $\epsilon = 10^{-4}$ have good accuracy and require only a modest number of iterations. As the tolerance decreases from $\epsilon = 10^{-2}$ to $\epsilon = 10^{-4}$, the number of iterations of NI3 grows by a factor much smaller than the worst-case number $2\log_2(100) \approx 13$. This suggests that high accuracy would require a small additional computing cost. For all methods, the largest errors occur in the binomial example with $n_1 = 3$, which we examine in more detail later. Except for this example with $\tilde{r} = 0.98$, NI3 always requires less work than NI1, about 30% on average and usually between 20% and 45%. Moreover, with two exceptions in the same example, NI3 is more accurate than NI1. The high-tolerance NI1 ($\epsilon = 10^{-2}$) usually has relative error about 4%-5% when $\tilde{r} = 0.05$, but the absolute error is perhaps more relevant, and this error is small (a simple rough remedy against large *relative* errors would be to set ϵ in proportion to $\tilde{r}$). NI2A is generally fast; it is also accurate, with one exception. This method benefits when the distance $|r_X^{-1}(\tilde{r}) - 2\sin(\pi\tilde{r}/6)|$ is small; in the minimal-discreteness cases (when $n_1 = n_2 = 1000$ for the binomial and for the largest values of r_1 and r_2 for the negative binomial case), this distance is very small, and NI2A is as accurate as NI1 or NI3 and usually faster. The largest observed value of this distance was about 0.09 (binomial marginals, $n_1 = 3$, $\tilde{r} = -0.5$). NI2B does not benefit from such a small integration distance, unless the root is close to zero; it frequently exhibits large errors that tend to increase as the discreteness increases and as the root (or $\tilde{r}$) moves farther from zero; the large errors are not surprising, since a very sparse integration grid was used.

We discuss the binomial problem with $n_1 = n_2 = 3$ and $\tilde{r} = 0.98$. The root is $r_X^{-1}(0.98) \approx 0.999041$ and its approximation is $2\sin(0.98\pi/6) \approx 0.981808$. Figure 1 shows $r_X(\rho)$ for $0.98 \leq \rho \leq 1$. NI3 behaves as pure bisection, because the attempted Newton steps fall outside the bracket at all iterations. NI1 requires fewer iterations than NI3. The low-order polynomial approximations of g supporting NI2 are poor in this areaa, so NI2 suffers from relatively large integration error. (Condition (22) is easily seen to hold in *all* binomial examples, and Proposition 3 gives $\lim_{\rho\to 1} g'(\rho) = \infty$.) We examined NI2 with m varying widely over powers of 2. The inaccuracy of NI2B persists until m quite large enough to make the method slow: at $m = 128$, we obtain $\rho_K = 1 - \delta = 0.9999$, $\hat{f}(\rho_K)$ has relative error about 2.2%, and the final error (the measure in the rightmost column in the tables) is about 1.7%; comparing these two errors suggests that the large final error is due to integration error; it is not due to interpolation error. NI2A fares much better; for example, at $m = 16$, the final error is 0.06%. As mentioned earlier, setting δ too small

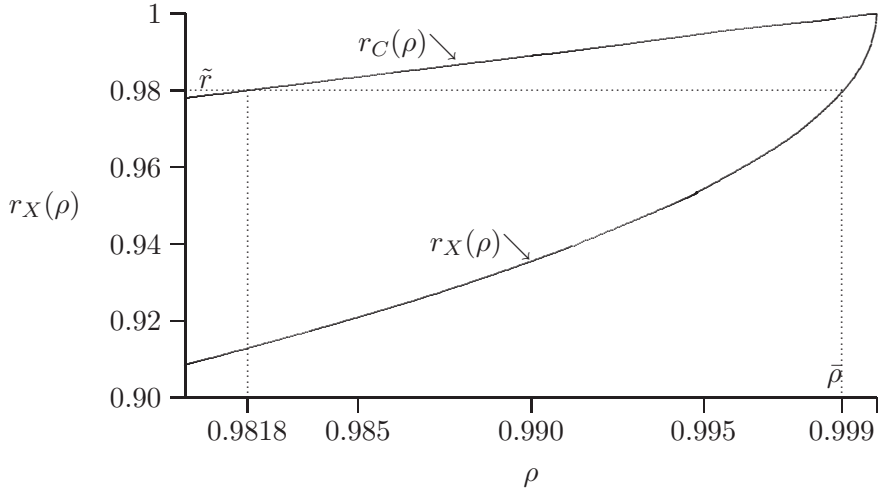

Figure 1: The function $r_X(\rho)$ on $[0.98, 1]$ in the example with binomial marginals with $n_1 = 3$; compared to the function $r_C(\rho)$ given by (6).

is detrimental: for NI2A, changing to $\delta = 10^{-12}$ and maintaining the value $m = 2$ that applies in Table 2, the final error increases to 4.2%.

In summary, if a good code is available for computing the bivariate normal distribution (and thus f), then we recommend NI3; both NI2 variants provide no accuracy guarantee and therefore they should be viewed as cheap, fast alternatives to NI3. If such good code is not available, then NI2B is an easier solution, because it requires only f' and not f.

5 Conclusion

We studied the NORTA correlation-matching problem for the case where the marginals are discrete. We proved some key properties of both the rank and linear correlations and their derivatives as functions of the correlation parameter ρ of the normal copula. We obtained a formula for the derivative f' of the function f whose root is sought. The derivative involves only the exponential function and can be evaluated significantly faster than f. We developed and analyzed algorithms that exploit the derivative. We emphasized rank-correlation matching, but our methods apply immediately to linear-correlation matching. For unbounded univariate marginals and rank-correlation matching, we adapted one of our methods that only requires evaluating f' (and not f) by substituting a finite-term approximation of f', and we provided bounds on the resulting error.

Table 1: Results for method NI1.

	ϵ	$\tilde{r}$	$\bar{\rho}$	CPU (s)	N_1	$r_X(\bar{\rho})$	Rel. error (%)
Binomial	10^{-2}	-0.50	-0.6076	0.084×10^{-3}	5	-0.4999	0.049
$n_1 = n_2 = 3$		0.05	0.0603	0.063×10^{-3}	5	0.0499	0.115
$p_1 = p_2 = 0.5$		0.20	0.2387	0.063×10^{-3}	5	0.1991	0.475
$n = 16$		0.90	0.9760	0.102×10^{-3}	6	0.8999	0.009
$r_X(-1) = -0.9241$		0.98	0.9999	0.050×10^{-3}	3	0.9935	1.382
$r_X(1) = 1$	10^{-4}	-0.50	-0.6079	0.103×10^{-3}	6	-0.5000	< 0.001
		0.05	0.0604	0.082×10^{-3}	5	0.0500	< 0.001
		0.20	0.2399	0.084×10^{-3}	5	0.2000	< 0.001
		0.90	0.9760	0.118×10^{-3}	6	0.9000	< 0.001
		0.98	0.9990	0.130×10^{-3}	7	0.9800	< 0.001
Binomial	10^{-2}	-0.50	-0.5194	0.093	4	-0.4991	0.175
$n_1 = n_2 = 100$		0.05	0.0551	0.069	3	0.0524	4.703
$p_1 = p_2 = 0.5$		0.20	0.2099	0.092	4	0.1999	0.036
$n = 10201$		0.90	0.9107	0.085	4	0.8996	0.043
$r_X(-1) = -0.9971$		0.98	0.9861	0.061	3	0.9811	0.114
$r_X(1) = 1$	10^{-4}	-0.50	-0.5203	0.119	5	-0.5000	< 0.001
		0.05	0.0526	0.094	4	0.0500	< 0.001
		0.20	0.2099	0.117	5	0.2000	< 0.001
		0.90	0.9111	0.113	5	0.9000	< 0.001
		0.98	0.9851	0.098	5	0.9800	< 0.001
Binomial	10^{-2}	-0.50	-0.5171	5.503	4	-0.4992	0.159
$n_1 = n_2 = 1000$		0.05	0.0500	4.410	3	0.0477	4.511
$p_1 = p_2 = 0.5$		0.20	0.2091	5.858	4	0.1999	0.031
$n = 1002001$		0.90	0.9082	4.703	4	0.8999	0.014
$r_X(-1) = -0.9997$		0.98	0.9803	3.153	3	0.9780	0.207
$r_X(1) = 1$	10^{-4}	-0.50	-0.5179	6.995	5	-0.5000	< 0.001
		0.05	0.0524	6.193	4	0.0500	< 0.001
		0.20	0.2091	7.509	5	0.2000	< 0.001
		0.90	0.9083	5.848	5	0.9000	< 0.001
		0.98	0.9821	4.289	4	0.9800	< 0.001
Negative Binomial	10^{-2}	-0.50	-0.5330	6.57×10^{-3}	4	-0.4960	0.209
$r_1 = 1.568$		0.05	0.0518	4.33×10^{-3}	3	0.0478	4.412
$p_1 = 0.3861$		0.43	0.4614	5.49×10^{-3}	4	0.4299	0.030
$r_2 = 6.021$		0.90	0.9323	4.46×10^{-3}	3	0.8986	0.151
$p_2 = 0.6211$		0.96	0.9895	5.88×10^{-3}	4	0.9593	0.076
$n = 768$	10^{-4}	-0.50	-0.5341	8.31×10^{-3}	5	-0.5000	< 0.001
$r_X(-1) = -0.9738$		0.05	0.0542	5.78×10^{-3}	4	0.0500	< 0.001
$r_X(1) = 0.9652$		0.43	0.4616	7.14×10^{-3}	5	0.4300	< 0.001
		0.90	0.9336	5.97×10^{-3}	4	0.9000	< 0.001
		0.96	0.9903	9.59×10^{-3}	6	0.9600	< 0.001
Negative Binomial	10^{-2}	-0.50	-0.5177	0.053	4	-0.4993	0.148
$r_1 = 15.68$		0.05	0.0501	0.041	3	0.0478	4.481
$p_1 = 0.3861$		0.43	0.4467	0.056	4	0.4298	0.042
$r_2 = 60.21$		0.90	0.9090	0.053	4	0.8999	0.017
$p_2 = 0.6211$		0.98	0.9811	0.037	3	0.9778	0.229
$n = 6560$	10^{-4}	-0.50	-0.5184	0.070	5	-0.5000	< 0.001
$r_X(-1) = -0.9971$		0.05	0.0524	0.052	4	0.0500	< 0.001
$r_X(1) = 0.9989$		0.43	0.4469	0.066	5	0.4300	< 0.001
		0.90	0.9092	0.067	5	0.9000	< 0.001
		0.98	0.9832	0.054	4	0.9800	< 0.001
Negative Binomial	10^{-2}	-0.50	-0.5169	1.301	4	-0.4992	0.159
$r_1 = 156.7$		0.05	0.0500	0.871	3	0.0478	4.491
$p_1 = 0.3861$		0.43	0.4464	1.273	4	0.4298	0.042
$r_2 = 602.1$		0.90	0.9080	1.255	4	0.8999	0.013
$p_2 = 0.6211$		0.98	0.9802	0.877	3	0.9780	0.199
$n = 189912$	10^{-4}	-0.50	-0.5177	1.639	5	-0.5000	< 0.001
$r_X(-1) = -0.9997$		0.05	0.0524	1.236	4	0.0500	< 0.001
$r_X(1) = 0.9999$		0.43	0.4465	1.616	5	0.4300	< 0.001
		0.90	0.9081	1.611	5	0.9000	< 0.001
		0.98	0.9819	1.190	4	0.9800	< 0.001

Table 2: Results for method NI2A with $\delta = 10^{-4}$ and $2h$ set as close as possible to 10^{-2}.

	$\tilde{r}$	$\bar{\rho}$	CPU (s)	$N_{g'}$	$r_X(\bar{\rho})$	Rel. error (%)
Binomial	-0.50	-0.6079	0.062×10^{-3}	21	-0.5000	< 0.001
$n_1 = n_2 = 3$	0.05	0.0604	0.032×10^{-3}	5	0.0500	< 0.001
$p_1 = p_2 = 0.5$	0.20	0.2399	0.045×10^{-3}	9	0.2000	< 0.001
$n = 16$	0.90	0.9760	0.067×10^{-3}	17	0.8999	0.011
$r_X(-1) = -0.9241$	0.98	0.9962	0.031×10^{-3}	5	0.9602	2.024
$r_X(1) = 1$						
Binomial	-0.50	-0.5203	0.032	5	-0.5000	< 0.001
$n_1 = n_2 = 100$	0.05	0.0526	0.032	5	0.0500	< 0.001
$p_1 = p_2 = 0.5$	0.20	0.2099	0.032	5	0.2000	< 0.001
$n = 10201$	0.90	0.9111	0.032	5	0.9000	< 0.001
$r_X(-1) = -0.9971$	0.98	0.9851	0.028	5	0.9810	0.099
$r_X(1) = 1$						
Binomial	-0.50	-0.5179	2.17	5	-0.5000	< 0.001
$n_1 = n_2 = 1000$	0.05	0.0524	2.48	5	0.0500	< 0.001
$p_1 = p_2 = 0.5$	0.20	0.2091	2.40	5	0.2000	< 0.001
$n = 1002001$	0.90	0.9083	2.01	5	0.9000	< 0.001
$r_X(-1) = -0.9997$	0.98	0.9821	1.89	5	0.9800	< 0.001
$r_X(1) = 1$						
Negative Binomial	-0.50	-0.5341	2.27×10^{-3}	5	-0.5000	< 0.001
$r_1 = 1.568$	0.05	0.0542	2.09×10^{-3}	5	0.0500	< 0.001
$p_1 = 0.3861$	0.43	0.4616	1.95×10^{-3}	5	0.4300	< 0.001
$r_2 = 6.021$	0.90	0.9336	2.29×10^{-3}	7	0.9000	< 0.001
$p_2 = 0.6211$	0.96	0.9903	2.33×10^{-3}	7	0.9600	< 0.001
$n = 768$						
$r_X(-1) = -0.9738$						
$r_X(1) = 0.9652$						
Negative Binomial	-0.50	-0.5184	0.019	5	-0.5000	< 0.001
$r_1 = 15.68$	0.05	0.0524	0.019	5	0.0500	< 0.001
$p_1 = 0.3861$	0.43	0.4469	0.018	5	0.4300	< 0.001
$r_2 = 60.21$	0.90	0.9092	0.019	5	0.9000	< 0.001
$p_2 = 0.6211$	0.98	0.9832	0.018	5	0.9800	< 0.001
$n = 6560$						
$r_X(-1) = -0.9971$						
$r_X(1) = 0.9989$						
Negative Binomial	-0.50	-0.5177	0.50	5	-0.5000	< 0.001
$r_1 = 156.7$	0.05	0.0524	0.46	5	0.0500	< 0.001
$p_1 = 0.3861$	0.43	0.4465	0.47	5	0.4300	< 0.001
$r_2 = 602.1$	0.90	0.9081	0.47	5	0.9000	< 0.001
$p_2 = 0.6211$	0.98	0.9819	0.45	5	0.9800	< 0.001
$n = 189912$						
$r_X(-1) = -0.9997$						
$r_X(1) = 0.9999$						

Table 3: Results for method NI2B with $m = 5$, $\delta = 10^{-4}$ (so $h = 0.09999$).

	$\tilde{r}$	$\bar{\rho}$	CPU (s)	$N_{q'}$	$r_X(\bar{\rho})$	Rel. error (%)
Binomial	-0.50	-0.6078	0.024×10^{-3}	9	-0.5000	< 0.001
$n_1 = n_2 = 3$	0.05	0.0601	0.010×10^{-3}	3	0.0497	0.514
$p_1 = p_2 = 0.5$	0.20	0.2389	0.015×10^{-3}	5	0.1999	0.026
$n = 16$	0.90	0.8485	0.029×10^{-3}	11	0.7409	17.676
$r_X(-1) = -0.9241$	0.98	0.8642	0.029×10^{-3}	11	0.7565	22.805
$r_X(1) = 1$						
Binomial	-0.50	-0.5202	9.65×10^{-3}	7	-0.4999	0.029
$n_1 = n_2 = 100$	0.05	0.0525	4.20×10^{-3}	3	0.0499	0.172
$p_1 = p_2 = 0.5$	0.20	0.2099	6.95×10^{-3}	5	0.2000	< 0.001
$n = 10201$	0.90	0.8718	15.60×10^{-3}	11	0.8582	4.641
$r_X(-1) = -0.9971$	0.98	0.9142	15.80×10^{-3}	11	0.9034	7.821
$r_X(1) = 1$						
Binomial	-0.50	-0.5178	1.023	7	-0.4998	0.030
$n_1 = n_2 = 1000$	0.05	0.0523	0.44	3	0.0499	0.157
$p_1 = p_2 = 0.5$	0.20	0.2091	0.73	5	0.2000	< 0.001
$n = 1002001$	0.90	0.8982	1.64	11	0.8892	1.202
$r_X(-1) = -0.9997$	0.98	0.9625	1.62	11	0.9586	2.187
$r_X(1) = 1$						
Negative Binomial	-0.50	-0.5340	0.76×10^{-3}	7	-0.4998	0.031
$r_1 = 1.568$	0.05	0.0541	0.34×10^{-3}	3	0.0499	0.157
$p_1 = 0.3861$	0.43	0.4614	0.75×10^{-3}	7	0.4299	0.028
$r_2 = 6.021$	0.90	0.9567	1.19×10^{-3}	11	0.9246	2.736
$p_2 = 0.6211$	0.96	1.00	1.13×10^{-3}	11	0.9652	0.545
$n = 768$						
$r_X(-1) = -0.9738$						
$r_X(1) = 0.9652$						
Negative Binomial	-0.50	-0.5183	6.38×10^{-3}	7	-0.4999	0.029
$r_1 = 15.68$	0.05	0.0523	2.72×10^{-3}	3	0.0499	0.160
$p_1 = 0.3861$	0.43	0.4468	6.44×10^{-3}	7	0.4299	0.019
$r_2 = 60.21$	0.90	0.8995	9.77×10^{-3}	11	0.8896	1.150
$p_2 = 0.6211$	0.98	0.9644	9.74×10^{-3}	11	0.9595	2.091
$n = 6560$						
$r_X(-1) = -0.9971$						
$r_X(1) = 0.9989$						
Negative Binomial	-0.50	-0.5176	0.18	7	-0.4998	0.030
$r_1 = 156.7$	0.05	0.0523	0.078	3	0.0499	0.155
$p_1 = 0.3861$	0.43	0.4465	0.18	7	0.4296	0.020
$r_2 = 602.1$	0.90	0.9077	0.29	11	0.8996	0.041
$p_2 = 0.6211$	0.98	0.9816	0.28	11	0.9797	0.035
$n = 189912$						
$r_X(-1) = -0.9997$						
$r_X(1) = 0.9999$						

Table 4: Results for method NI3.

	ϵ	$\tilde{r}$	$\bar{\rho}$	CPU (s)	N_1	$r_X(\bar{\rho})$	Rel. error (%)
Binomial	10^{-2}	-0.50	-0.6079	0.038×10^{-3}	2	-0.5000	< 0.001
$n_1 = n_2 = 3$		0.05	0.0604	0.020×10^{-3}	1	0.0500	0.004
$p_1 = p_2 = 0.5$		0.20	0.2399	0.040×10^{-3}	2	0.2000	< 0.001
$n = 16$		0.90	0.9767	0.060×10^{-3}	3	0.9013	0.142
$r_X(-1) = -0.9241$		0.98	0.9922	0.137×10^{-3}	7	0.9429	3.783
$r_X(1) = 1$	10^{-4}	-0.50	-0.6079	0.064×10^{-3}	3	-0.5000	< 0.001
		0.05	0.0604	0.053×10^{-3}	2	0.0500	< 0.001
		0.20	0.2399	0.054×10^{-3}	2	0.2000	< 0.001
		0.90	0.9760	0.106×10^{-3}	5	0.9000	< 0.001
		0.98	0.9990	0.222×10^{-3}	12	0.9800	< 0.001
Binomial	10^{-2}	-0.50	-0.5203	0.028	1	-0.5000	< 0.001
$n_1 = n_2 = 100$		0.05	0.0526	0.029	1	0.0500	< 0.001
$p_1 = p_2 = 0.5$		0.20	0.2099	0.028	1	0.2000	< 0.001
$n = 10201$		0.90	0.9111	0.025	1	0.9000	< 0.001
$r_X(-1) = -0.9971$		0.98	0.9851	0.022	1	0.9800	< 0.001
$r_X(1) = 1$	10^{-4}	-0.50	-0.5203	0.054	2	-0.5000	< 0.001
		0.05	0.0526	0.056	2	0.0500	< 0.001
		0.20	0.2099	0.053	2	0.2000	< 0.001
		0.90	0.9111	0.047	2	0.9000	< 0.001
		0.98	0.9851	0.044	2	0.9800	< 0.001
Binomial	10^{-2}	-0.50	-0.5179	1.642	1	-0.5000	< 0.001
$n_1 = n_2 = 1000$		0.05	0.0524	1.924	1	0.0500	< 0.001
$p_1 = p_2 = 0.5$		0.20	0.2091	1.831	1	0.2000	< 0.001
$n = 1002001$		0.90	0.9083	1.387	1	0.9000	< 0.001
$r_X(-1) = -0.9997$		0.98	0.9821	1.214	1	0.9800	< 0.001
$r_X(1) = 1$	10^{-4}	-0.50	-0.5179	3.360	2	-0.5000	< 0.001
		0.05	0.0524	1.918	1	0.0500	< 0.001
		0.20	0.2091	1.771	1	0.2000	< 0.001
		0.90	0.9083	2.796	2	0.9000	< 0.001
		0.98	0.9821	2.550	2	0.9800	< 0.001
Negative Binomial	10^{-2}	-0.50	-0.5341	3.79×10^{-3}	2	-0.5000	< 0.001
$r_1 = 1.568$		0.05	0.0542	1.78×10^{-3}	1	0.0500	< 0.001
$p_1 = 0.3861$		0.43	0.4616	3.15×10^{-3}	2	0.4300	< 0.001
$r_2 = 6.021$		0.90	0.9336	3.45×10^{-3}	2	0.9000	< 0.001
$p_2 = 0.6211$		0.96	0.9902	3.46×10^{-3}	2	0.9616	< 0.001
$n = 768$	10^{-4}	-0.50	-0.5341	3.77×10^{-3}	2	-0.5000	< 0.001
$r_X(-1) = -0.9738$		0.05	0.0542	3.42×10^{-3}	2	0.0500	< 0.001
$r_X(1) = 0.9652$		0.43	0.4616	3.16×10^{-3}	2	0.4300	< 0.001
		0.90	0.9336	5.13×10^{-3}	3	0.9000	< 0.001
		0.96	0.9903	5.10×10^{-3}	3	0.9600	< 0.001
Negative Binomial	10^{-2}	-0.50	-0.5184	0.017	1	-0.5000	< 0.001
$r_1 = 15.68$		0.05	0.0524	0.016	1	0.0500	< 0.001
$p_1 = 0.3861$		0.43	0.4469	0.016	1	0.4300	< 0.001
$r_2 = 60.21$		0.90	0.9092	0.016	1	0.9000	< 0.001
$p_2 = 0.6211$		0.98	0.9832	0.015	1	0.9800	< 0.001
$n = 6560$	10^{-4}	-0.50	-0.5184	0.031	2	-0.5000	< 0.001
$r_X(-1) = -0.9971$		0.05	0.0524	0.016	1	0.0500	< 0.001
$r_X(1) = 0.9989$		0.43	0.4469	0.031	2	0.4300	< 0.001
		0.90	0.9092	0.031	2	0.9000	< 0.001
		0.98	0.9832	0.028	2	0.9800	< 0.001
Negative Binomial	10^{-2}	-0.50	-0.5177	0.393	1	-0.5000	< 0.001
$r_1 = 156.7$		0.05	0.0524	0.370	1	0.0500	< 0.001
$p_1 = 0.3861$		0.43	0.4465	0.378	1	0.4300	< 0.001
$r_2 = 602.1$		0.90	0.9081	0.383	1	0.9000	< 0.001
$p_2 = 0.6211$		0.98	0.9819	0.332	1	0.9800	< 0.001
$n = 189912$	10^{-4}	-0.50	-0.5177	0.394	1	-0.5000	< 0.001
$r_X(-1) = -0.9997$		0.05	0.0524	0.369	1	0.0500	< 0.001
$r_X(1) = 0.9999$		0.43	0.4465	0.366	1	0.4300	< 0.001
		0.90	0.9081	0.740	2	0.9000	< 0.001
		0.98	0.9819	0.689	2	0.9800	< 0.001

Our numerical experience and findings can be summarized as follows. We initially expected that the ratio of work per evaluation of f compared to work per evaluation of f' would be large, making NI2 competitive. To our surprise, there exist algorithms that compute the bivariate normal integral (and thus f) to negligible error at small computing cost. In our implementation, this ratio was about 12, a value smaller than we expected. (Other users may observe a different value, depending on the method for computing bivariate normal integrals and the implementation quality.) Moreover, NI2 lacks a solution-error guarantee, so it should be viewed as a cheap and approximate alternative to exact methods. Implementing the derivative f' is very simple, requiring just a few lines of simple code. In summary, if a good code is available for computing the bivariate normal integral, then our recommendation is the Newton-type method NI3. Otherwise, NI2B is an easy (approximate) solution, because it requires only f' and not f; but some care is needed to keep the integration errors small enough.

We also contributed a convergence result on the L_∞ distance (i.e., the supremum over $\rho \in [-1, 1]$ of the absolute difference) between the rank-correlation function $r_X(\rho; F_1, F_2)$ for given discrete marginals F_1 and F_2 and the explicitly known analog for continuous marginals, $(6/\pi) \arcsin(\rho/2)$, in terms of the maximum probability masses of F_1 and F_2, as these masses go to zero. In particular, this result justifies the value $2\sin(\pi\tilde{r}/6)$ as an approximation to the solution to (3) and points to it as a starting point for exact solution methods.

Interesting future work is to analyze further the properties of normal-copula dependence for discrete marginals with unbounded support. Problems and approaches of interest are: (1) study the correlation error that results from truncating to finite support for a single given ρ; (2) if this error can be made small uniformly across ρ by an appropriate truncation, then finite-support correlation-matching methods could be proved to be effective; (3) propose and analyze alternatives to our approximate correlation-matching method, perhaps via steps (1) and (2); and (4) evaluate correlation-matching methods experimentally.

References

Avramidis, A. N., A. Deslauriers, P. L'Ecuyer. 2004. Modeling daily arrivals to a telephone call center. *Management Science* **50** 896–908.

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.

Brent, R. P. 1971. An algorithm with guaranteed convergence for finding a zero of a function. *Computer Journal* **14** 422–425.

Bus, J. C. P., T. J. Dekker. 1975. Two efficient algorithms with guaranteed convergence for finding a zero of a function. *ACM Transactions of Mathematical Software* **1** 330–345.

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. 1998. Numerical methods for fitting and simulating autoregressive-to-anything processes. *INFORMS Journal on Computing* **10** 72–81.

Chen, H. 2001. Initialization for NORTA: Generation of random vectors with specified marginals and correlations. *INFORMS Journal on Computing* **13** 312–331.

Donnelly, T. G. 1973. Algorithm 462: Bivariate normal distribution. *Communications of the ACM* **16** 638.

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., F. Lindskog, A. McNeil. 2003. Modelling dependence with copulas and applications to risk management. S. Rachev, ed., *Handbook of Heavy Tailed Distributions in Finance*. Elsevier, 329–384. Chapter 8.

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.

Genz, A. 2004. Numerical computation of rectangular bivariate and trivariate normal and t probabilities. *Statistics and Computing* **14** 151–160. See `http://www.math.wsu.edu/faculty/genz/homepage`.

Ghosh, S., S.G. 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, G. Derflinger. 2004. *Automatic Nonuniform Random Variate Generation*. Springer-Verlag, Berlin.

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

Kendall, M., A. Stuart. 1977. *The advanced theory of statistics*, vol. 1: Distribution theory. 4th ed. Macmillan, New York.

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

Kurowicka, D., R. M. Cooke. 2001. Conditional, partial and rank correlation for the elliptical copula; dependence modelling in uncertainty analysis. E. Zio, M. Demichela, N. Piccinini, eds., *Safety and Reliability, 3rd volume*. 1795–1802.

Lurie, P. M., 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*, *Lecture Notes in Statistics*, vol. 139. Springer-Verlag, New York.

Owen, D. B. 1956. Tables for computing bivariate normal probability. *Annals of Mathematical Statistics* **27** 1075–1090.

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.

Stoer, J., R. Bulirsch. 1980. *Introduction to Numerical Analysis*. Springer-Verlag, New York.