Distance Correlation Coefficients for Lancaster Distributions

Johannes Dueck,* Dominic Edelmann,* and Donald Richards†

February 6, 2015

Abstract

We consider the problem of calculating distance correlation coefficients between random vectors whose joint distributions belong to the class of Lancaster distributions. We derive under mild convergence conditions a general series representation for the distance covariance for these distributions. To illustrate the general theory, we apply the series representation to derive explicit expressions for the distance covariance and distance correlation coefficients for the bivariate and multivariate normal distributions, and for the bivariate gamma and Poisson distributions which are of Lancaster type.

Key words and phrases. Affine invariance; bivariate gamma distribution; bivariate Poisson distribution; characteristic function; distance correlation coefficient; Lancaster distributions; multivariate normal distribution.

2010 Mathematics Subject Classification. Primary: 60E05, 62H20; Secondary: 33C05, 42C05, 60E10.

Running head: Distance Correlation and Lancaster Distributions.

1 Introduction

The concepts of distance covariance and distance correlation, introduced by Székely, et al. [31, 27], have been shown to be widely applicable for measuring dependence between collections of random variables. As examples of the ubiquity of distance correlation methods, we note the results on distance correlation given recently by: Székely, et al. [20, 31, 28, 29, 30], on statistical inference; Sejdinovic, et al. [25], on machine learning;

*Institut für Angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany.
†Department of Statistics, Pennsylvania State University, University Park, PA 16802, U.S.A.
Kong, et al. [10], on familial relationships and mortality; Zhou [33], on nonlinear time series; Lyons [16], on abstract metric spaces; Martínez-Gomez, et al. [17] and Richards, et al. [19], on large astrophysical databases; Dueck, et al. [6], on high-dimensional inference and the analysis of wind data; and Dueck, et al. [7], on a connection with singular integrals on Euclidean spaces.

A result which is of fundamental importance in distance correlation theory is the explicit formula for the empirical distance correlation coefficient [31, pp. 2773-2774]. By combining that explicit formula with the fast algorithm of Huo and Székely [9], it becomes straightforward to apply distance correlation methods to real-world data sets.

On the other hand, the calculation of population distance correlation coefficients remains an intractable problem generally. Székely, et al. [31, p. 2785] calculated the distance correlation coefficient for the bivariate normal distribution; Dueck, et al. [5, Appendix] extended that result to the general multivariate normal distribution; and Dueck, et al. [6] calculated the affinely invariant distance correlation coefficient for the multivariate normal distribution. Otherwise, no such results are yet available for any other distribution. Hence, the state of distance correlation theory hitherto is that the empirical coefficients could be calculated readily but their population counterparts were unknown, generally. Consequently, it was not possible to calculate distance correlation coefficients explicitly for given nonnormal distributions in terms of the usual parameters that parametrize these distributions, or to ascertain for nonnormal distributions any analogs of the limit theorems derived by Dueck, et al. [6, Section 4].

We describe in detail the difficulties arising in attempts to calculate the population distance correlation coefficients. Let $p$ and $q$ be positive integers. For column vectors $s \in \mathbb{R}^p$ and $t \in \mathbb{R}^q$, denote by $\|s\|$ and $\|t\|$ the standard Euclidean norms on the corresponding spaces; thus, if $s = (s_1, \ldots, s_p)'$ then $\|s\| = (s_1^2 + \cdots + s_p^2)^{1/2}$, and similarly for $\|t\|$. Given vectors $u$ and $v$ of the same dimension, we let $\langle u, v \rangle$ be the standard Euclidean scalar product of $u$ and $v$. For jointly distributed random vectors $(X,Y) \in \mathbb{R}^p \times \mathbb{R}^q$ and non-random vectors $(s,t) \in \mathbb{R}^p \times \mathbb{R}^q$, let

$$
\psi_{X,Y}(s,t) = \mathbb{E}\exp\left[i\langle s, X \rangle + i\langle t, Y \rangle\right],
$$

be the joint characteristic function of $(X,Y)$, and let $\psi_X(s) = \psi_{X,Y}(s,0)$ and $\psi_Y(t) = \psi_{X,Y}(0,t)$ be the corresponding marginal characteristic functions. For any $z \in \mathbb{C}$, let $|z|^2$ denote the squared modulus of $z$; also, we use the notation

$$
\gamma_p = \frac{\pi^{(p+1)/2}}{\Gamma((p + 1)/2)}.
$$

In the case of distributions with finite first moments, Székely, et al. [31, p. 2772] defined $\mathcal{V}(X,Y)$, the distance covariance between $X$ and $Y$, to be the positive square-root of

$$
\mathcal{V}^2(X,Y) = \frac{1}{\gamma_p \gamma_q} \int_{\mathbb{R}^{p+q}} \left| \psi_{X,Y}(s,t) - \psi_X(s)\psi_Y(t) \right|^2 \frac{ds \, dt}{\|s\|^{p+1} \|t\|^{q+1}}.
$$
and they defined the distance correlation coefficient between $X$ and $Y$ as

$$R(X, Y) = \frac{V(X, Y)}{\sqrt{V(X, X) V(Y, Y)}}$$ (1.3)

if both $V(X, X)$ and $V(Y, Y)$ are strictly positive, and otherwise to be zero [31, p. 2773]. For distributions with finite first moments we have $0 \leq R(X, Y) \leq 1$, and $R(X, Y) = 0$ if and only if $X$ and $Y$ are mutually independent.

For given random vectors $X$ and $Y$, the fundamental obstacle in calculating the population distance correlation coefficient (1.3) is the computation of the singular integral (1.2). In particular, the singular nature of the integrand precludes evaluation of the integral by expanding the numerator, $|\psi_{X,Y}(s, t) - \psi_X(s)\psi_Y(t)|^2$, and subsequent term-by-term integration of each of the resulting three terms.

In this paper, we calculate the distance correlation coefficients for pairs $(X, Y)$ of random vectors whose joint distributions are in the class of Lancaster distributions, a class of probability distributions which was made prominent by Lancaster [14, 15] and by Sarmanov [21]. The distribution functions of the Lancaster family are well-known to have attractive expansions in terms of certain orthogonal functions (Koudou [13]; Diaconis, et al. [4]). By applying those expansions, we deduce series expansions for the corresponding characteristic functions and then we obtain explicit expressions for the distance covariance and distance correlation coefficients.

Consequently we derive under mild convergence conditions a general formula for the distance covariance for the Lancaster distributions. As examples, we apply the general formula to obtain explicit expressions for the distance covariance and distance correlation for the bivariate and multivariate normal distributions, and for bivariate gamma and Poisson distributions. We remark that explicit results can also be obtained for certain negative binomial distributions and for other Lancaster-type expansions obtained by Bar-Lev, et al. [3]; because the formulas derived here are fully representative of the general case then we will omit the details for other cases.

## 2 The Lancaster distributions

To recapitulate the class of Lancaster distributions we generally follow the standard notation in that area, as given by Koudou [12, 13]; cf., Lancaster [15], Pommeret [18], or Diaconis, et al. [4, Section 6].

Let $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$ be locally compact, separable probability spaces, such that $L^2(\mu)$ and $L^2(\nu)$ are separable. Let $\sigma$, a probability measure on $\mathcal{X} \times \mathcal{Y}$, have marginal distributions $\mu$ and $\nu$; then there exist functions $K_\sigma$ and $L_\sigma$ such that

$$\sigma(dx, dy) = K_\sigma(x, dy)\mu(dx) = L_\sigma(dx, y)\nu(dy).$$

We note that $K_\sigma$ and $L_\sigma$ represent the conditional distributions of $Y$ given $X = x$, and $X$ given $Y = y$, respectively.
Let $\mathcal{C}$ denote a countable index set with a zero element, denoted by 0. Let $\{P_n : n \in \mathcal{C}\}$ and $\{Q_n : n \in \mathcal{C}\}$ be sequences of functions on $\mathcal{X}$ and $\mathcal{Y}$ which form orthonormal bases for the separable Hilbert spaces $L^2(\mu)$ and $L^2(\nu)$, respectively. We assume, by convention, that $P_0 \equiv 1$ and $Q_0 \equiv 1$.

Because the tensor product Hilbert space $L^2(\mu \otimes \nu) \equiv L^2(\mu) \otimes L^2(\nu)$ is separable there holds, for $\sigma \in L^2(\mu \otimes \nu)$, the expansion
\[
\sigma(dx, dy) = \sum_{m \in \mathcal{C}} \sum_{n \in \mathcal{C}} \rho_{m,n} P_m(x)Q_n(y) \mu(dx) \nu(dy),
\]
($x, y) \in \mathcal{X} \times \mathcal{Y}$). Letting $\delta_{m,n}$ denote Kronecker’s delta, the probability measure $\sigma$ is called a Lancaster distribution if there exists a positive sequence $\{\rho_n : n \in \mathcal{C}\}$ such that
\[
\int P_m(x)Q_n(y) \sigma(dx, dy) = \rho_m \delta_{m,n}
\]
for all $m, n \in \mathcal{C}$; in particular, $\rho_0 = 1$. The sequence $\{\rho_n : n \in \mathcal{C}\}$ is called a Lancaster sequence, and the expansion (2.1) reduces to
\[
\sigma(dx, dy) = \sum_{n \in \mathcal{C}} \rho_n P_n(x)Q_n(y) \mu(dx) \nu(dy).
\]

Koudou [12, pp. 255–256] characterized the Lancaster sequences $\{\rho_n : n \in \mathcal{C}\}$ such that the associated probability distribution $\sigma$ is absolutely continuous with respect to $\mu \otimes \nu$ and has Radon-Nikodym derivative
\[
\frac{\sigma(dx, dy)}{\mu(dx)\nu(dy)} = \sum_{n \in \mathcal{C}} \rho_n P_n(x)Q_n(y) \in L^2(\mu \otimes \nu),
\]
($x, y) \in \mathcal{X} \times \mathcal{Y}$).

In the sequel, we consider the case in which $\mathcal{X} = \mathbb{R}^p$ and $\mathcal{Y} = \mathbb{R}^q$ and the underlying random vectors $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ have joint distribution $\sigma$ and marginal distributions $\mu$ and $\nu$, respectively. We assume that $\mu$, $\nu$, and $\sigma$ are absolutely continuous with respect to Lebesgue measure or counting measure on the respective sample spaces and we denote their corresponding probability density functions by $\phi_X$, $\phi_Y$, and $\phi_{X,Y}$, respectively. This yields the expansion,
\[
\phi_{X,Y}(x, y) = \phi_X(x) \phi_Y(y) \sum_{n \in \mathcal{C}} \rho_n P_n(x)Q_n(y).
\]

We will refer to (2.2) as the Lancaster expansion of the joint density function $\phi_{X,Y}$.

3 Examples of Lancaster expansions

In this section, we provide examples of Lancaster expansions (2.2) for the bivariate and multivariate normal distributions, and for some bivariate gamma and Poisson distributions. In the sequel, we denote by $\mathbb{N}_0$ the set of nonnegative integers.
3.1 The bivariate normal distribution

Let \((X, Y)\) follow a bivariate normal distribution with mean 0 and covariance matrix
\[
\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},
\]
denoted by \((X, Y) \sim \mathcal{N}_2(0, \Sigma)\). The joint probability density function of \((X, Y)\) is
\[
\phi_{X,Y}(x, y) = \frac{1}{2\pi(1 - \rho^2)^{\frac{1}{2}}} \exp \left( -\frac{x^2 + y^2 - 2\rho xy}{2(1 - \rho^2)} \right),
\]
\(x, y \in \mathbb{R}\), and the marginal density functions are given by
\[
\phi_X(x) = \phi_Y(y) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2}x^2 \right).
\]
In this case, the index set \(C\) is \(\mathbb{N}_0\). For \(n \in \mathbb{N}_0\), let
\[
H_n(x) = (-1)^n \exp \left( \frac{1}{2}x^2 \right) \left( \frac{d}{dx} \right)^n \exp \left( -\frac{1}{2}x^2 \right),
\]
x \(\in \mathbb{R}\), denote the \(n\)th Hermite polynomial, \(n = 0, 1, 2, \ldots\). It is well-known that the polynomials \(\{H_n : n \in \mathbb{N}_0\}\) are orthogonal with respect to the standard normal distribution and form a complete orthogonal basis for the Hilbert space \(L^2(X)\). Also, the Lancaster expansion of \(\phi_{X,Y}\) is given by the classical formula of Mehler: For \(x, y \in \mathbb{R}\),
\[
\phi_{X,Y}(x, y) = \phi_X(x) \phi_Y(y) \sum_{n=0}^{\infty} \frac{\rho^n}{n!} H_n(x) H_n(y),
\]
and this series converges absolutely for all \(x \in \mathbb{R}\) and \(y \in \mathbb{R}\).

We remark that there are numerous extensions of Mehler’s formula which represent Lancaster-type expansions for generalizations of the bivariate normal distribution; for such expansions, see Sarmanov and Bratoeva [24] and Srivastava and Singhal [26] and the references given in those articles. The details in those cases are similar to the results which we derive, and we can obtain analogous formulas for the distance correlation coefficients for those distributions.

3.2 The multivariate normal distribution

Let \(X \in \mathbb{R}^p\) and \(Y \in \mathbb{R}^q\) be random vectors such that \((X, Y) \sim \mathcal{N}_{p+q}(0, \Sigma)\), a \((p + q)\)-dimensional multivariate normal distribution with mean vector 0 and positive definite covariance matrix
\[
\Sigma = \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix}
\]
\((3.2)\).
Dueck, Edelmann, and Richards

where \( \Sigma_X, \Sigma_Y, \) and \( \Sigma_{XY} = \Sigma_{YX}' \) are \( p \times p, q \times q \) and \( p \times q \) matrices, respectively. We denote by \( \phi_{X,Y} \) the joint probability density function of \((X,Y)\), and by \( \phi_X \) and \( \phi_Y \) the marginal density functions of \( X \) and \( Y \), respectively.

We now describe the Lancaster expansion of \( \phi_{X,Y} \), a result derived in [32]. In this case, the index set \( C \) is \( \mathbb{N}_{0}^{p \times q} \), the set of \( p \times q \) matrices with nonnegative integer entries.

For a matrix of summation indices \( N = (N_{rc}) \in \mathbb{N}_{0}^{p \times q} \), define \( N! = \prod_{r=1}^{p} \prod_{c=1}^{q} N_{rc}! \). For \( r = 1, \ldots, p \), let

\[
N_{r,*} = \sum_{c=1}^{q} N_{rc}
\]

and set \( N_{*,*} = (N_{1,*}, \ldots, N_{p,*}) \). Similarly, for each \( c = 1, \ldots, q \), define

\[
N_{*,c} = \sum_{r=1}^{p} N_{rc}
\]

and set \( N_{*,*} = (N_{*,1}, \ldots, N_{*,q}) \). Further, we define

\[
N_{**,*} = \sum_{r=1}^{p} \sum_{c=1}^{q} N_{rc},
\]

and note that \( N_{**,*} = \sum_{r=1}^{p} N_{r,*} = \sum_{c=1}^{q} N_{*,c} \).

Denoting by \( (\Sigma_{XY})_{rc} \) the \((r,c)\)th entry of \( \Sigma_{XY} \), we also define

\[
\Sigma_{XY}^N = \prod_{r=1}^{p} \prod_{c=1}^{q} (\Sigma_{XY})_{rc}^{N_{rc}}.
\]

We now introduce the multivariate Hermite polynomials. For any \( p \in \mathbb{N} \), \( k = (k_1, \ldots, k_p) \in \mathbb{N}_0^p \), and \( x = (x_1, \ldots, x_p) \in \mathbb{R}^p \), define \( x^k = x_1^{k_1} \cdots x_p^{k_p} \) and define the differential operator,

\[
\left( -\frac{\partial}{\partial x} \right)^k = \left( -\frac{\partial}{\partial x_1} \right)^{k_1} \cdots \left( -\frac{\partial}{\partial x_p} \right)^{k_p}.
\]

The \( k \)th multivariate Hermite polynomial with respect to the marginal density function \( \phi_X \) is defined as

\[
H_k(x; \Sigma_X) = \frac{1}{\phi_X(x)} \left( -\frac{\partial}{\partial x} \right)^k \phi_X(x).
\]

The Lancaster expansion of the multivariate normal density function \( \phi_{X,Y} \) is given by the generalized Mehler formula [32]:

\[
\phi_{X,Y}(x,y) = \phi_X(x) \phi_Y(y) \sum_{N \in \mathbb{N}_{0}^{p \times q}} \frac{\Sigma_{XY}^N}{N!} H_{N,**}(x; \Sigma_X) H_{N,**}(y; \Sigma_Y),
\]

with absolute convergence for all \( x \in \mathbb{R}^p, y \in \mathbb{R}^q \).
To calculate the affinely invariant distance correlation coefficient between $X$ and $Y$, as defined by Dueck, et al. (2014), we need the Lancaster expansion of the joint density function of the standardized random vectors $\tilde{X} = \Sigma_X^{-1/2} X$ and $\tilde{Y} = \Sigma_Y^{-1/2} Y$. It is straightforward to verify that $(\tilde{X}, \tilde{Y}) \sim N_{p+q}(0, \Lambda)$ where

$$\Lambda = \begin{pmatrix} I_p & \Lambda_{XY} \\ \Lambda_{XY}' & I_q \end{pmatrix}$$

with $\Lambda_{XY} = \Sigma_X^{-1/2} \Sigma_{XY} \Sigma_Y^{-1/2}$, and then we deduce from (3.4) that the Lancaster expansion for $(\tilde{X}, \tilde{Y})$ is

$$\phi_{\tilde{X}, \tilde{Y}}(x, y) = \phi_{\tilde{X}}(x) \phi_{\tilde{Y}}(y) \sum_{N \in \mathbb{N}^{p \times q}} \frac{\Lambda_{XY}}{N!} H_{N..}(x; I_p) H_{N..}(y; I_q).$$

(3.5)

### 3.3 The bivariate gamma distribution

The Lancaster expansion for a bivariate gamma distribution, which was derived by Sarmanov [22, 23], can be stated as follows (cf., Kotz, et al. [11, pp. 437–438]).

For $\alpha > -1$ and $n \in \mathbb{N}_0$, the classical *Laguerre polynomial* is defined by

$$L_n^{(\alpha)}(x) = \frac{1}{n!} x^{-\alpha} \exp(x) \left( \frac{d}{dx} \right)^n x^{n+\alpha} \exp(-x)$$

$$= \frac{(\alpha + 1)_n}{n!} \sum_{j=0}^{n} \frac{(-n)_j}{(\alpha + 1)_j} \frac{x^j}{j!},$$

(3.6)

$x > 0$, where $(\alpha)_n = \Gamma(\alpha + n)/\Gamma(\alpha)$ denotes the rising factorial.

Let $\lambda \in (0, 1)$, and let $\alpha$ and $\beta$ satisfy $\alpha \geq \beta > 0$. Sarmanov [22, 23] derived for certain bivariate gamma random variables $(X, Y)$ the joint probability density function,

$$\phi_{X,Y}(x, y) = \phi_X(x) \phi_Y(y) \sum_{n=0}^{\infty} a_n L_n^{(\alpha-1)}(x) L_n^{(\beta-1)}(y),$$

(3.7)

$x, y > 0$, where

$$a_n = \lambda^n \left[ \frac{(\beta)_n}{(\alpha)_n} \right]^{1/2},$$

(3.8)

$n = 0, 1, 2, \ldots$. The corresponding marginal density functions are

$$\phi_X(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \exp(-x)$$

and

$$\phi_Y(y) = \frac{1}{\Gamma(\beta)} y^{\beta-1} \exp(-y),$$

which we recognize as the density functions of one-dimensional gamma random variables with index parameters $\alpha$ and $\beta$, respectively.
We remark that if $\alpha = \beta$ then the density function (3.7) reduces to the Kibble-Moran bivariate gamma density function and $\text{Corr}(X, Y) = \lambda$ (Kotz, et al. [11, pp. 436–437]). Also, (3.7) represents the Lancaster expansion for $(X, Y)$.

3.4 The bivariate Poisson distribution

For $a > 0$ and $x, n \in \mathbb{N}_0$, let

$$C_n(x; a) = \left(\frac{a^n}{n!}\right)^{1/2} \sum_{k=0}^{n} (-1)^k \binom{n}{k} \binom{x}{k} k! a^k$$

(3.9)

denote the Poisson-Charlier polynomial of degree $n$. For $\lambda \in [0, 1]$, Koudou [13, Section 5] (cf., Bar-Lev, et al. [3], Pommeret [18]) shows that there exists a bivariate random vector $(X, Y)$ with probability density function

$$\phi_{X,Y}(x, y) = \phi_X(x) \phi_Y(y) \sum_{n=0}^{\infty} \lambda^n C_n(x; a) C_n(y; a),$$

(3.10)

$x, y \in \mathbb{N}_0$. The corresponding marginal density functions $\phi_X$ and $\phi_Y$ are given by

$$\phi_X(k) = \phi_Y(k) = \frac{a^k \exp(-a)}{k!},$$

$k \in \mathbb{N}_0$, so that $X$ and $Y$ are distributed marginally according to a Poisson distribution with parameter $a$. The series (3.10) is an expansion of Lancaster type, a special case of (2.2), and the resulting distribution is called a bivariate Poisson distribution.

4 Distance correlation coefficients for Lancaster distributions

In this section, we derive a general series expression for the distance correlation coefficients for Lancaster distributions with density functions of the form (2.2). For a joint density function $\phi_{X,Y}$ given by (2.2) and $n \in \mathcal{C}$, we introduce the notation

$$\mathcal{P}_n(s) = \mathbb{E} \exp(i \langle s, X \rangle) P_n(X),$$

(4.1)

$s \in \mathbb{R}^p$, and

$$\mathcal{Q}_n(t) = \mathbb{E} \exp(i \langle t, Y \rangle) Q_n(Y),$$

(4.2)

t $\in \mathbb{R}^q$. To verify that each expectation $\mathcal{P}_n(s)$ converges absolutely for all $s \in \mathbb{R}^p$, we apply the Cauchy-Schwarz inequality to obtain

$$|\mathcal{P}_n(s)|^2 = |\mathbb{E} \exp(i \langle s, X \rangle) P_n(X)|^2$$

$$\leq \left(\mathbb{E} |\exp(i \langle s, X \rangle)|^2 \right) \cdot \left(\mathbb{E} |P_n(X)|^2 \right) = 1,$$
because \( \{P_n : n \in \mathcal{C}\} \) is an orthonormal basis for the Hilbert space \( L^2(\mu) \). Similarly, \( |Q_n(t)| \leq 1 \) for all \( t \in \mathbb{R}^q \).

In the following result, we will use the notation

\[
A_{j,k} = \int_{\mathbb{R}^p} P_j(s) P_k(-s) \frac{ds}{\|s\|^{p+1}}
\]

and

\[
B_{j,k} = \int_{\mathbb{R}^q} Q_j(t) Q_k(-t) \frac{dt}{\|t\|^{q+1}},
\]

\( j, k \in \mathcal{C} \), whenever these integrals converge absolutely.

We now state the main result.

**Theorem 4.1.** Suppose that the random vectors \( X \in \mathbb{R}^p \) and \( Y \in \mathbb{R}^q \) have the joint probability density function (2.2). Then,

\[
\gamma^2(X, Y) = \frac{1}{\gamma^p \gamma^q} \sum_{j \in \mathcal{C}, j \neq 0} \sum_{k \in \mathcal{C}, k \neq 0} \rho_j \rho_k A_{j,k} B_{j,k},
\]

whenever the sum converges absolutely.

**Proof.** Rewriting the Lancaster expansion (2.2) in the form,

\[
\phi_{X,Y}(x, y) - \phi_X(x) \phi_Y(y) = \phi_X(x) \phi_Y(y) \sum_{n \in \mathcal{C}, n \neq 0} \rho_n P_n(x) Q_n(y),
\]

and taking Fourier transforms on both sides of this identity, we obtain for all \( s \in \mathbb{R}^p \) and \( t \in \mathbb{R}^q \) the expansion

\[
\psi_{X,Y}(s, t) - \psi_X(s) \psi_Y(t) = \sum_{n \in \mathcal{C}, n \neq 0} \rho_n P_n(s) Q_n(t).
\]

This identity is valid subject to the requirement that we may interchange summation and integration, which is justified by the assumption that the sum in the final result converges absolutely. Using (4.4) we deduce that

\[
|\psi_{X,Y}(s, t) - \psi_X(s) \psi_Y(t)|^2 = (\psi_{X,Y}(s, t) - \psi_X(s) \psi_Y(t))(\psi_{X,Y}(s, t) - \psi_X(s) \psi_Y(t))
\]

\[
= \sum_{j \in \mathcal{C}, j \neq 0} \sum_{k \in \mathcal{C}, k \neq 0} \rho_j \rho_k P_j(s) P_k(-s) Q_j(t) Q_k(-t).
\]

Next, we integrate this expansion with respect to the measures \( ds/\|s\|^{p+1} \) and \( dt/\|t\|^{q+1} \); this requires that we again interchange summation and integration which, by assumption, we are able to do. On carrying through these procedures, we obtain (4.3). \( \square \)
5 Examples

In this section, we demonstrate the versatility of Theorem 4.1 by applying it to compute the distance correlation coefficients for the bivariate normal, multivariate normal, and bivariate gamma and Poisson distributions. We verify for each example the absolute convergence of the series resulting from Theorem 4.1, for that convergence property cannot in general be obtained from abstract Lancaster expansions. In developing each example, we retain the corresponding notation in Section 3.

5.1 The bivariate normal distribution

In the sequel, we use the standard double-factorial notation,

\[ n!! = n(n-2)(n-4)\cdots = \begin{cases} n(n-2)(n-4)\cdots 2, & \text{if } n \text{ is even} \\ n(n-2)(n-4)\cdots 1, & \text{if } n \text{ is odd} \end{cases} \]

Proposition 5.1. Let \((X,Y) \sim \mathcal{N}_2(0, \Sigma)\), a bivariate normal distribution with correlation coefficient \(\rho\). Then,

\[ V^2(X,Y) = 4\pi^{-1} \sum_{l=1}^{\infty} \frac{((2l-3)!!)^2}{(2l)!} (1 - 2^{-(2l-1)}) \rho^{2l}, \]

and this series converges absolutely for all \(\rho \in (-1, 1)\).

Proof. Starting with the Lancaster expansion of the bivariate normal density function, as given in (3.1), and using the definitions of \(P_n\) and \(Q_n\) in (4.1) and (4.2), respectively, we obtain by substitution and integration-by-parts,

\[ P_n(s) = Q_n(s) = \int_{-\infty}^{\infty} \exp(isx) \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2) H_n(x) \, dx \]

\[ = (is)^n \exp\left(-\frac{1}{2}s^2\right), \]

\(s \in \mathbb{R}\). Therefore,

\[ A_{j,k} = B_{j,k} = (-1)^k i^{j+k} \int_{-\infty}^{\infty} s^{j+k-2} \exp(-s^2) \, ds, \]

\[ = \begin{cases} (-1)^k i^{j+k} \pi^{1/2} \left(\frac{1}{2}\right)^{(j+k-2)/2} (j+k-3)!!, & \text{if } j+k \text{ is even} \\ 0, & \text{otherwise} \end{cases} \]

since the latter integral is a moment of the \(\mathcal{N}(0, \frac{1}{2})\) distribution. By Theorem 4.1, we obtain

\[ V^2(X,Y) = \frac{4}{\pi} \sum_{j+k > 0, j+k \text{ even}} \frac{\rho^{j+k}}{j!k!} \left(\frac{1}{2}\right)^{j+k} ((j+k-3)!!)^2. \]
Setting $j + k = 2l$ with $l \geq 1$, the double series reduces to
\[
V^2(X,Y) = 4 \sum_{l=1}^{\infty} \rho^{2l} \frac{((2l - 3)!!)^2}{2^{2l} ((2l - 3)!!)^2 (2l)!} \sum_{j+l=k} \frac{1}{j! k!}
\]
\[
= 4 \sum_{l=1}^{\infty} \rho^{2l} \frac{((2l - 3)!!)^2}{2^{2l} ((2l - 3)!!)^2 (2l)!} \sum_{j+l=k} \frac{1}{j! (2l - j)!}
\]
\[
= 4 \sum_{l=1}^{\infty} \rho^{2l} \frac{((2l - 3)!!)^2}{2^{2l} (2l)! (2l - 2)},
\]
which is the same as (5.1).

The absolute convergence of (5.1) can be verified by comparison with a geometric series. Moreover, it is straightforward to verify that the series is identical with the result obtained by Székely, et al. \cite{31}, p. 2786. □

Having obtained $V(X,Y)$, we let $\rho \to 1-$ to obtain the distance variances $V(X,X)$ and $V(Y,Y)$; here, we are applying a well-known result that if $(X,Y) \sim \mathcal{N}_2(0, \Sigma)$ where $\text{Var}(X) = \text{Var}(Y)$ and $\rho = 1$ then $X = Y$, almost surely. By applying properties of Gauss’ hypergeometric series, as was done by Dueck, et al. \cite{6}, p. 2318, we obtain
\[
V^2(X,X) = V^2(Y,Y) = \frac{4}{3} - \frac{4(\sqrt{3} - 1)}{\pi}.
\]

### 5.2 The multivariate normal distribution

In this subsection, we will make extensive use of the notation $\mathbf{N}_{r, \cdot}, \mathbf{N}_{\cdot, c}, \mathbf{N}_{* \cdot}, \mathbf{N}_{\cdot *},$ and $\mathbf{N}_{..}$ from Subsection 3.2 for the multi-index matrix $\mathbf{N} \in \mathbb{N}_0^{p \times q}$. We now establish the following result.

**Proposition 5.2.** Let $(X,Y) \sim \mathcal{N}_{p+q}(0, \Sigma)$, where $\Sigma$ is given in (3.2). Then the affinely invariant distance covariance, $\tilde{V}^2(X,Y)$, is given by
\[
\tilde{V}^2(X,Y) = \frac{1}{\gamma_p \gamma_q} \sum_{J \neq 0, K \neq 0} A_{J,K} B_{J,K} \frac{\Lambda_{XY}^{J,K}}{J! K!},
\]
where the sums are taken over all non-zero $J, K \in \mathbb{N}_0^{p \times q}$ such that all components of $J, K$ and $J, K$ are even,
\[
A_{J,K} = \frac{\Gamma \left( \frac{1}{2} (J, .. + K, .. - 1) \right)}{\Gamma \left( \frac{1}{2} (J, .. + K, ..) + \frac{1}{2} p \right)} \prod_{r=1}^{p} \Gamma \left( \frac{1}{2} (J, .. + K, ..) + 1 \right) \tag{5.3}
\]
and
\[
B_{J,K} = \frac{\Gamma \left( \frac{1}{2} (J, .. + K, .. - 1) \right)}{\Gamma \left( \frac{1}{2} (J, .. + K, ..) + \frac{1}{2} q \right)} \prod_{c=1}^{q} \Gamma \left( \frac{1}{2} (J, .. + K, ..) + 1 \right). \tag{5.4}
\]
PROOF. In this case, the index set $\mathcal{C}$ is $\mathbb{N}_0^{p \times q}$, and we write the Lancaster expansion (3.5) of $(\tilde{X}, \tilde{Y})$ in the form

$$
\phi_{\tilde{X},\tilde{Y}}(x, y) - \phi_{\tilde{X}}(x) \phi_{\tilde{Y}}(y) = \phi_{\tilde{X}}(x) \phi_{\tilde{Y}}(y) \sum_{N \neq 0} \frac{\Lambda_{XY}}{N!} H_{N_{++}}(x; I_p) H_{N_{--}}(y; I_q).
$$

To calculate the Fourier transform $\mathcal{P}_N$ corresponding to $\tilde{X}$, we apply the definition (3.3) of the multivariate Hermite polynomials and integration-by-parts to deduce that for $s \in \mathbb{R}^p$,

$$
\mathcal{P}_N(s) = \int_{\mathbb{R}^p} \exp(i \langle s, x \rangle) \phi_{\tilde{X}}(x) H_{N_{++}}(x; I_p) \, dx
$$

$$
= (-1)^{N_{++}} \int_{\mathbb{R}^p} \exp(i \langle s, x \rangle) \left( \frac{\partial}{\partial x} \right)^{N_{++}} \phi_{\tilde{X}}(x) \, dx
$$

$$
= \int_{\mathbb{R}^p} \phi_{\tilde{X}}(x) \left( \frac{\partial}{\partial x} \right)^{N_{++}} \exp(i \langle s, x \rangle) \, dx
$$

$$
= (is)^{N_{++}} \int_{\mathbb{R}^p} \phi_{\tilde{X}}(x) \exp(i \langle s, x \rangle) \, dx
$$

$$
= i^{N_{++}} s^{N_{++}} \exp(-\frac{1}{2} \langle s, s \rangle).
$$

Similarly,

$$
\mathcal{Q}_N(t) = i^{N_{++}} t^{N_{++}} \exp(-\frac{1}{2} \langle t, t \rangle),
$$

$t \in \mathbb{R}^q$. Therefore,

$$
\int_{\mathbb{R}^p} \mathcal{P}_J(s) \mathcal{P}_K(-s) \frac{ds}{\|s\|^{p+1}} = (-1)^{K_{++}} i^{J_{++}+K_{--}} \int_{\mathbb{R}^p} s^{J_{++}+K_{--}} \exp(-\langle s, s \rangle) \frac{ds}{\|s\|^{p+1}}.
$$

We now change variables to hyperspherical coordinates: $s = r \omega$, where $r > 0$ and $\omega = (\omega_1, \ldots, \omega_p) \in S^{p-1}$, the unit sphere in $\mathbb{R}^p$. Then the latter integral reduces to

$$
\int_{\mathbb{R}_+} r^{J_{++}+K_{--}-2} \exp(-r^2) \, dr \int_{S^{p-1}} \omega^{J_{++}+K_{--}} \, d\omega.
$$

The integral over $\mathbb{R}_+$ is evaluated by replacing $r$ by $r^{1/2}$, and we obtain its value as

$$
\frac{1}{2} \Gamma \left( \frac{1}{2} (J_{++} + K_{--} - 1) \right).
$$

It is easy to see that the integral over $S^{p-1}$ equals zero if any component of $J_{++} + K_{--}$ is odd. For the case in which each component of $J_{++} + K_{--}$ is even, we obtain

$$
\int_{S^{p-1}} \omega^{J_{++}+K_{--}} \, d\omega = A(S^{p-1}) \mathbb{E}(\omega^{J_{++}+K_{--}}),
$$

where $A(S^{p-1}) = 2\pi^{p/2} / \Gamma(\frac{1}{2} p)$ is the surface area of $S^{p-1}$ and $\omega$ now is a uniformly distributed random vector on $S^{p-1}$. It is well-known that the random vector $(\omega_1^2, \ldots, \omega_p^2) \sim$
$D(\frac{1}{2}, \ldots, \frac{1}{2})$, a Dirichlet distribution with parameters $(\frac{1}{2}, \ldots, \frac{1}{2})$; so, by a classical formula for the moments of the Dirichlet distribution [11, p. 488],

$$
\mathbb{E}(\omega^{J\cdot K\cdot}) = \frac{\Gamma(\frac{1}{2}p) \prod_{r=1}^{p} \Gamma(\frac{1}{2}J_{r\cdot} + K_{r\cdot} + 1)}{\Gamma(\frac{1}{2}p) \prod_{r=1}^{p} \Gamma(\frac{1}{2}J_{r\cdot} + K_{r\cdot} + \frac{1}{2}p)}.
$$

Collecting together these results, we obtain

$$
\int_{\mathbb{R}^p} \mathcal{P}_J(s) \mathcal{P}_K(-s) \frac{ds}{\|s\|^{p+1}} = (-1)^{K\cdot} (-1)^{(J\cdot + K\cdot)/2} A_{J,K},
$$

where $A_{J,K}$ is given in (5.3). A similar expression can be obtained for

$$
\int_{\mathbb{R}^q} \mathcal{Q}_J(t) \mathcal{Q}_K(-t) \frac{dt}{\|t\|^{q+1}},
$$

from which the final result (5.2) follows.

Similar to the bivariate normal case, the affinely invariant distance variance $\widetilde{V}^2(X, X)$ in the multivariate case can be calculated by taking $p = q$ and $\Lambda_{XY} = \rho I_p$, where $-1 < \rho < 1$, and then letting $\rho \to 1$ in the expression for $\widetilde{V}^2(X, Y)$.

We remark also that the distance covariance and distance correlation for non-standardized jointly normal random vectors can be calculated using the arguments used earlier, and we refer to Dueck, et al. [6] for the explicit formula.

5.3 The bivariate gamma distribution

**Proposition 5.3.** Suppose that the random vector $(X, Y)$ is distributed according to a Sarmanov bivariate gamma distribution, as given by (3.7). Then,

$$
\mathcal{V}^2(X, Y) = 2^{(1-\alpha-\beta)} \frac{\Gamma(2\alpha + 1) \Gamma(2\beta + 1)}{\Gamma(\alpha) \Gamma(\beta)} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} a_j a_k A_{j,k}(\alpha) A_{j,k}(\beta),
$$

where

$$
A_{j,k}(\alpha) = \frac{(\alpha)_j (\alpha)_k (1 - \alpha - j)_{j+k-2}}{j! k! (\alpha - j + 2)_{j+k-2} \Gamma(\alpha - j + 2)} \, _2F_1(-j - k + 2, 2\alpha; \alpha - k + 2; \frac{1}{2}).
$$

**Proof.** By (3.7), there holds the expansion,

$$
\phi_{X,Y}(x, y) - \phi_X(x) \phi_Y(y) = \phi_X(x) \phi_Y(y) \sum_{n=1}^{\infty} a_n L_n^{(\alpha-1)}(x) L_n^{(\beta-1)}(y),
$$

$x, y > 0$. Then, it follows from (4.1) that for $s, t \in \mathbb{R},$

$$
\mathcal{P}_n(s) = \int_{0}^{\infty} \exp(isx) L_n^{(\alpha-1)}(x) \phi_X(x) \, dx
$$

$$
= \frac{1}{\Gamma(\alpha)} \int_{0}^{\infty} \exp\left(- (1-is)x\right) x^{\alpha-1} L_n^{(\alpha-1)}(x) \, dx.
$$
By a direct calculation using (3.6), we obtain

\[ P_n(s) = \frac{(\alpha)_n}{n!} (1 - is)^{-\alpha} (1 - (1 - is)^{-1})^n \]

\[ = \frac{(\alpha)_n}{n!} (1 - is)^{-(\alpha+n)} (-is)^n \]

and, analogously,

\[ Q_n(t) = \frac{(\beta)_n}{n!} (1 - it)^{-(\beta+n)} (-it)^n. \]

We now calculate the integral

\[ \int_{\mathbb{R}} P_j(s) P_k(-s) \frac{ds}{s^2} = \frac{(\alpha)_j}{j!} \frac{(\alpha)_k}{k!} \int_{\mathbb{R}} g(s) ds, \quad (5.6) \]

where

\[ g(s) = s^{j+k-2} (1 - is)^{-(\alpha+j)} (1 + is)^{-(\alpha+k)}, \quad (5.7) \]

\( s \in \mathbb{R}. \) To calculate the integral on the right-hand side of (5.6), we utilize Cauchy’s beta integral [2, p. 48]: For \( a, u, v \in \mathbb{C} \) such that \( \text{Re}(a) > 0 \) and \( \text{Re}(u + v) > 1, \)

\[ \int_{\mathbb{R}} (1 - is)^{-u} (1 + ias)^{-v} \, ds = 2\pi \frac{\Gamma(u+v-1)}{\Gamma(u)\Gamma(v)} a^{u-1} (a+1)^{2-u-v}. \quad (5.8) \]

To differentiate the left-hand side of (5.8) \( m \) times with respect to \( a, \) we apply the formula,

\[ \left( \frac{\partial}{\partial a} \right)^m (1 + ias)^{-v} = (-i)^m s^m (v)_m (1 + ias)^{-v-m}; \]

by differentiating under the integral we obtain

\[ \left( \frac{\partial}{\partial a} \right)^m \int_{\mathbb{R}} (1 - is)^{-u} (1 + ias)^{-v} \, ds = (-i)^m (v)_m \int_{\mathbb{R}} s^m (1 - is)^{-u} (1 + ias)^{-v-m} \, ds. \]

To differentiate the right-hand side of (5.8) \( m \) times with respect to \( a, \) we apply Leibniz’s formula:

\[ \left( \frac{\partial}{\partial a} \right)^m [a^{u-1} (a+1)^{2-u-v}] = \sum_{l=0}^{m} \binom{m}{l} \left( \frac{\partial}{\partial a} \right)^{m-l} a^{u-1} \cdot \left( \frac{\partial}{\partial a} \right)^{l} (a+1)^{2-u-v}. \]

Noting that

\[ \binom{m}{l} = \frac{(-1)^l (-m)_l}{l!}, \]

\[ \left( \frac{\partial}{\partial a} \right)^{m-l} a^{u-1} = (-1)^l a^{u-1-m+l} \frac{(1 - u)_m}{(u - m)_l}, \]
and

\[ \left( \frac{\partial}{\partial a} \right)^l (a + 1)^{2-u-v} = (-1)^l (a + 1)^{2-u-v} (-2 + u + v)_l, \]

we obtain

\[ \left( \frac{\partial}{\partial a} \right)^m \left[ a^{u-1} (a + 1)^{2-u-v} \right] \\
= (-1)^m a^{u-1-m} (a + 1)^{2-u-v} (1-u)_m \sum_{l=0}^{m} \frac{(-m)_l (-2 + u + v)_l}{l! (u-m)_l} a^l (a + 1)^{-l} \\
= (-1)^m a^{u-1-m} (a + 1)^{2-u-v} (1-u)_m \, _2F_1 \left( -m, -2 + u + v; u - m; \frac{a}{a+1} \right), \]

where \( _2F_1 \) denotes Gauss’ hypergeometric function.

Comparing the derivatives of the left- and right-hand sides of (5.8), we obtain

\[ \int_R s^m (1 - is)^{-u} (1 + ias)^{-v-m} \, ds = 2\pi (-i)^m a^{u-1-m} (a + 1)^{2-u-v} \frac{\Gamma(u+v-1)}{\Gamma(u)\Gamma(v)} \]

\[ \times \frac{(1-u)^m}{(v)_m} \, _2F_1 \left( -m, -2 + u + v; u - m; \frac{a}{a+1} \right). \]

Substituting \( a = 1, m = j + k - 2, u = \alpha + j, \) and \( v = \alpha + k - m \equiv \alpha - j + 2, \) the latter equation reduces to

\[ \int_R g(s) \, ds = 2^{-2\alpha+1} \pi (-1)^{j+k-2} \frac{\Gamma(2\alpha+1)}{\Gamma(\alpha+j)\Gamma(\alpha-j+2)} \]

\[ \times \frac{(1-\alpha-j)_{j+k-2}}{(\alpha-j+2)_{j+k-2}} \, _2F_1 \left( -j - k + 2, 2\alpha; \alpha - k + 2; \frac{1}{2} \right). \]

Therefore,

\[ \int_R \mathcal{P}_j(s) \mathcal{P}_k(-s) \, \frac{ds}{s^2} = 2^{-2\alpha+1} \pi (-1)^{j-1} \frac{(\alpha)_j (\alpha)_k}{j! k!} \frac{\Gamma(2\alpha+1)}{\Gamma(\alpha+j)\Gamma(\alpha-j+2)} \]

\[ \times \frac{(1-\alpha-j)_{j+k-2}}{(\alpha-j+2)_{j+k-2}} \, _2F_1 \left( -j - k + 2, 2\alpha; \alpha - k + 2; \frac{1}{2} \right), \]

and similarly for \( Y. \) Substituting these expressions into Theorem 4.1 and simplifying the outcome, we obtain the series (5.5) as a formal expression for \( \mathcal{V}^2(X, Y). \)

Finally, we verify that (5.5) converges absolutely. By (5.7),

\[ \int_R |g(s)| \, ds = \int_R |s|^{j+k-2} (1 + s^2)^{-(2\alpha+j+k)/2} \, ds. \]

Making the change-of-variables \( s^2 = t/(1-t), \) the latter integral is transformed to

\[ \int_0^1 t^{\frac{1}{2}(j+k-3)} (1-t)^{\alpha-\frac{1}{2}} \, dt = B \left( \frac{1}{2}(j + k - 1), \alpha + \frac{1}{2} \right), \quad (5.9) \]
where $B(\cdot, \cdot)$ is the classical beta function, and this integral converges absolutely because $j + k - 1 > 0$ and $\alpha + 1/2 > 0$ for all $j, k \in \mathbb{N}$ and $\alpha > 0$. Hence, to establish that (5.5) converges absolutely, we need only show that the series

$$
\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} a_j a_k \frac{(\alpha)_j (\beta)_j}{(j!)^2} \frac{(\alpha)_k (\beta)_k}{(k!)^2} \times B\left(\frac{j}{2}(j + k - 1), \alpha + \frac{1}{2}\right) B\left(\frac{k}{2}(j + k - 1), \beta + \frac{1}{2}\right)
$$

(5.10)

converges absolutely.

By (3.8), $0 \leq a_j \leq \lambda^j \leq 1$ for all $j$. Also, for $j + k \geq 3$, it follows from (5.9) that

$$
B\left(\frac{1}{2}(j + k - 1), \alpha + \frac{1}{2}\right) \leq \int_0^1 (1 - t)^{\alpha - \frac{1}{2}} dt = \frac{1}{\alpha + \frac{1}{2}}.
$$

Therefore, (5.10) is bounded above by

$$
\alpha^2 \beta^2 \lambda^2 \left[ B\left(\frac{1}{2}, \alpha + \frac{1}{2}\right)^2 + \frac{1}{(\alpha + \frac{1}{2})^2} \sum_{j, k \geq 1}^{\infty} \frac{(\alpha)_j (\beta)_j (\alpha)_k (\beta)_k}{(j!)^2 (k!)^2} \lambda^{j+k} \right] \leq \alpha^2 \beta^2 \lambda^2 \left[ B\left(\frac{1}{2}, \alpha + \frac{1}{2}\right)^2 + \frac{1}{(\alpha + \frac{1}{2})^2} \left( \sum_{j=0}^{\infty} \frac{(\alpha)_j (\beta)_j}{(j!)^2} \lambda^j \right)^2 \right] \equiv \alpha^2 \beta^2 \lambda^2 \left[ B\left(\frac{1}{2}, \alpha + \frac{1}{2}\right)^2 + \frac{1}{(\alpha + \frac{1}{2})^2} \left[ \beta \alpha \right] \right],
$$

and it is well-known that this Gaussian hypergeometric series converges absolutely for all $\alpha, \beta \in \mathbb{C}$ and all $\lambda \in [0, 1]$. \hfill \Box

In calculating the distance variances $\mathcal{V}(X, X)$ and $\mathcal{V}(Y, Y)$, only the marginal distributions are relevant. Therefore, we may assume that $X$ and $Y$ have any joint distribution for which the marginal distributions are gamma with parameters $\alpha$ and $\beta$, respectively. Letting $\beta \to \alpha$, the Sarmanov bivariate gamma distribution reduces to the Kibble-Moran distribution, and then the joint characteristic function of $(X, Y)$ is

$$
\left((1 - it_1)(1 - it_2) + \lambda t_1 t_2\right)^{-\alpha};
$$
cf. [11, p. 436]. Next, we let $\lambda \to 1-$; then this characteristic function converges to

$$
(1 - i(t_1 + t_2))^{-\alpha} \equiv \mathbb{E} \exp \left( i(t_1 + t_2)X \right),
$$

proving that, for $\lambda = 1$, $X = Y$, almost surely. Therefore, the distance variance $\mathcal{V}(X, X)$ is a limiting case of $\mathcal{V}(X, Y)$, viz.,

$$
\mathcal{V}^2(X, Y) = \frac{1}{\gamma_1^2} \int_{\mathbb{R}^2} |\psi_X(s + t) - \psi_X(s)\psi_X(t)|^2 \frac{ds}{s^2} \frac{dt}{t^2} = \lim_{\lambda \to 1-} \lim_{\beta \to \alpha} \frac{1}{\gamma_1^2} \int_{\mathbb{R}^2} |\psi_{X,Y}(s, t) - \psi_X(s)\psi_Y(t)|^2 \frac{ds}{s^2} \frac{dt}{t^2} = \lim_{\lambda \to 1-} \lim_{\beta \to \alpha} \mathcal{V}^2(X, Y).
$$
Similarly, 
\[ V^2(Y, Y) = \lim_{\lambda \to 1^-} \lim_{\alpha \to \beta^+} V^2(X, Y). \]

### 5.4 The bivariate Poisson distribution

**Proposition 5.4.** Suppose that the random vector \((X, Y)\) is distributed according to a bivariate Poisson distribution, as given by (3.10). Then

\[ V^2(X, Y) = \frac{1}{\pi} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \lambda^{j+k} \left( \frac{(4a)^{j+k}}{j!k!} \right) A_{jk}^2, \quad (5.11) \]

where

\[ A_{jk} = \sum_{l=0}^{j-k} \binom{j-k}{l} (-1)^{l/2} \sum_{n=0}^{l/2} (-1)^n \frac{\Gamma(n+j - \frac{1}{2}l - \frac{1}{2})}{\Gamma(n+j - \frac{1}{2}l)} \times {}_1F_1(n+j - \frac{1}{2}l - \frac{1}{2}; n+j - \frac{1}{2}l; -4a) = \frac{(a^{n+1})^{1/2}}{n!} (1-e^{i\pi})^n \exp (-a(1-e^{i\pi})). \]

Therefore, for \(j \geq k\), and \(A_{jk} = A_{kj}\) for \(j < k\).

**Proof.** By (3.10) and (4.1), we have

\[ P_n(s) = Q_n(s) = \mathbb{E} \exp(isX) C_n(X; a), \]

for \(s \in \mathbb{R}\). Substituting the definition (3.9) of the Poisson-Charlier polynomials \(C_n\) into the expectation and reversing the order of summation, we obtain

\[ P_n(s) = Q_n(s) = \sum_{x=0}^{\infty} \exp(isx) C_n(x; a) \frac{e^{-ax}}{x!} = \frac{(a^n)^{1/2}}{n!} (1-e^{is})^n \exp (-a(1-e^{is})). \]

Therefore, for \(j, k \geq 1\),

\[ \int_{\mathbb{R}} P_j(s) P_k(-s) \frac{ds}{s^2} = \frac{(a^{j+k})^{1/2}}{j! k!} \int_{\mathbb{R}} (1-e^{i\pi})^j (1-e^{-i\pi})^k \exp (-a(1-e^{i\pi} + 1-e^{-i\pi})) \frac{ds}{s^2} = \frac{(a^{j+k})^{1/2}}{j! k!} \int_{\mathbb{R}} (1-e^{i\pi})^j (1-e^{-i\pi})^k \exp (-2a(1-\cos s)) \frac{ds}{s^2}. \quad (5.13) \]

Because this integral is symmetric in \(j\) and \(k\) then we can assume, with no loss of generality, that \(j \geq k\). We now write

\[ (1-e^{is})^j (1-e^{-is})^k = (1-e^{is})^{j-k}(1-e^{is})^k(1-e^{-is})^k = (1-e^{is})^{j-k}(2(1-\cos s))^k, \]
and apply the binomial theorem in the form,
\[(1 - e^{is})^{j-k} = (1 - \cos s - i \sin s)^{j-k}\]
\[= \sum_{l=0}^{j-k} \binom{j-k}{l} (-i \sin s)^l (1 - \cos s)^{j-k-l}.\]

Then, it follows that the integral in (5.13) equals
\[2^k \sum_{l=0}^{j-k} \binom{j-k}{l} (-i)^l \int_{\mathbb{R}} (\sin s)^l (1 - \cos s)^{j-l} \exp \left( -2a(1 - \cos s) \right) \frac{ds}{s^2} = (5.14)\]

Expanding the exponential term,
\[\exp \left( -2a(1 - \cos s) \right) = \sum_{m=0}^{\infty} \frac{(-2a)^m}{m!} (1 - \cos s)^m,\]

applying the half-angle identities, \(\sin s = 2 \sin \frac{1}{2}s \cos \frac{1}{2}s\) and \(1 - \cos s = 2(\sin \frac{1}{2}s)^2\), and integrating term-by-term, we deduce that (5.14) equals
\[2^k \sum_{l=0}^{j-k} \binom{j-k}{l} (-i)^l \int_{\mathbb{R}} (\sin s)^l (1 - \cos s)^{j-l} \exp \left( -2a(1 - \cos s) \right) \frac{ds}{s^2} = (5.15)\]

If \(l\) is odd then the latter integral is an odd function of \(s\), so the integral equals 0. For the case in which \(l\) is even, we apply the identity \(\sin^2 s = 1 - \cos^2 s\) to write the integral in (5.15) as
\[\int_{\mathbb{R}} (\cos^2 \frac{1}{2}s)^{l/2} (\sin^2 \frac{1}{2}s)^{2(j+m)-l} \frac{ds}{s^2} = \int_{\mathbb{R}} (1 - \sin^2 \frac{1}{2}s)^{l/2} (\sin^2 \frac{1}{2}s)^{2(j+m)-l} \frac{ds}{s^2} = (5.16)\]

To calculate the latter integral, we will expand the first term in the integrand by the binomial theorem and then integrate termwise. Applying the formula (Gradshteyn and Ryzhik [8, p. 483, 3.821(10)]),
\[\int_{\mathbb{R}} (\sin^2 \frac{1}{2}s)^{2k} \frac{ds}{s^2} = \begin{cases} \pi, & k = 1 \\ (2k - 3)!!, & k = 2, 3, 4, \ldots \end{cases} = (5.17)\]

we find that (5.16) equals
\[\sum_{n=0}^{l/2} (-1)^n \int_{\mathbb{R}} (\sin^2 \frac{1}{2}s)^{2(n+j+m)-l} \frac{ds}{s^2} = \pi \sum_{n=0}^{l/2} (-1)^n \frac{(2(n+j+m)-l-3)!!}{(2(n+j+m)-l-2)!!}.\]
Substituting this result into (5.15), and interchanging the order of summation over \( m \) and \( n \), we obtain

\[
\int_{\mathbb{R}} P_j(s)P_k(-s)\frac{ds}{s^2} = \pi \left( \frac{(4a)^{j+k}}{j!k!} \right)^{1/2} \sum_{l=0}^{j-k} \binom{j-k}{l} (-1)^{l/2} \times \frac{1}{n/2} \sum_{l=0}^{n/2} \binom{n}{l} \frac{(-4a)^{n/2}}{m!} \frac{(2(n+j+m) - l - 3)!!}{(2(n+j+m) - l - 2)!!}.
\]

(5.18)

Writing each double factorial in terms of rising factorials, and simplifying the resulting expressions, we find that (5.18) equals

\[
\pi^{1/2} \left( \frac{(4a)^{j+k}}{j!k!} \right)^{1/2} \sum_{l=0}^{j-k} \binom{j-k}{l} (-1)^{l/2} \times \frac{1}{n/2} \sum_{l=0}^{n/2} \binom{n}{l} \frac{\Gamma(n+j-l-\frac{1}{2})}{\Gamma(n+j-l-\frac{1}{2})} \, _1F_1(n+j-l-\frac{1}{2}; n+j-l-\frac{1}{2}; -4a),
\]

(5.19)

where \(_1F_1\) denotes the confluent hypergeometric function.

We remark that the individual terms in this series can be calculated in a straightforward way by differentiating a simpler hypergeometric series. Note that each confluent hypergeometric function in (5.19) is of the form \(_1F_1(r-\frac{1}{2}; r; -4a)\) for \( r \in \mathbb{N} \); for \( r = 1 \), this latter function satisfies the well-known Kummer transformation [2, p. 191],

\[ _1F_1(\frac{1}{2}; 1; -4a) \equiv e^{-2a} \, _0F_1(1; a^2); \]

and for \( r \geq 1 \) we may differentiate this identity with respect to \( a \), using the well-known formula [2, p. 94],

\[ _1F_1(r-\frac{1}{2}; r; a) = \left( \frac{1}{2} \right)^{r-1} \frac{\partial}{\partial a} \right)^{r-1} _1F_1(\frac{1}{2}; 1; a). \]

Finally, we establish the absolute convergence of the resulting series for \( \mathcal{L}^2(X,Y) \). On applying to (5.13) the identity

\[ |1 - e^{is}| = |1 - e^{-is}| = (2(1 - \cos s))^{1/2} = 2(\sin^2 \frac{1}{2}s)^{1/2} \]

and the inequality

\[ \exp \left( -2(1 - \cos s) \right) \leq 1, \]

\( s \in \mathbb{R} \), we obtain

\[
\left| \int_{\mathbb{R}} P_j(s)P_k(-s)\frac{ds}{s^2} \right| \leq \left( \frac{(4a)^{j+k}}{j!k!} \right)^{1/2} \int_{\mathbb{R}} |1 - e^{is}| |1 - e^{is}| \exp \left( -2a(1 - \cos s) \right) \frac{ds}{s^2}
\]

\[
\leq \left( \frac{(4a)^{j+k}}{j!k!} \right)^{1/2} \int_{\mathbb{R}} \left( \sin^2 \frac{1}{2}s \right)^{(j+k)/2} \frac{ds}{s^2}.
\]
By the Cauchy-Schwarz inequality,
\[
\int_{\mathbb{R}} \left( \sin^2 \frac{1}{2} s \right)^{(j+k)/2} \frac{ds}{s^2} \equiv \int_{\mathbb{R}} \left( \sin^2 \frac{1}{2} s \right)^{j/2} \left( \sin^2 \frac{1}{2} s \right)^{k/2} \frac{ds}{s^2} \leq \left( \int_{\mathbb{R}} \left( \sin^2 \frac{1}{2} s \right)^j \frac{ds}{s^2} \right)^{1/2} \left( \int_{\mathbb{R}} \left( \sin^2 \frac{1}{2} s \right)^k \frac{ds}{s^2} \right)^{1/2}.
\]
Because \((2k-3)!!/(2k-2)!! \leq 1\) for all \(k \in \mathbb{N}\) then it follows from (5.17) that
\[
\int_{\mathbb{R}} \left( \sin^2 \frac{1}{2} s \right)^j \frac{ds}{s^2} \leq \pi;
\]
therefore,
\[
\left| \int_{\mathbb{R}} \mathcal{P}_j(s) \mathcal{P}_k(-s) \frac{ds}{s^2} \right| \leq \left( \frac{(4a)^{j+k}}{j! k!} \right)^{1/2} \pi,
\]
and the same holds for the functions \(Q_j\). Substituting these bounds into the general series expansion (4.3), we obtain the upper bound
\[
V^2(X, Y) \leq \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \frac{(4\lambda a)^{j+k}}{j! k!} = \left( \exp(4\lambda a) - 1 \right)^2 < \infty,
\]
for all \(\lambda \in [0, 1]\) and \(a > 0\). Therefore, the series (5.12) converges absolutely. \(\square\)

To calculate the distance variance, the argument given in the bivariate gamma case remains valid here. By Koudou [13, p. 103], the characteristic function of \((X, Y)\) is
\[
\psi_{X,Y}(s, t) = \exp \left[ a(1 - \lambda)(e^{is} - 1) + a(1 - \lambda)(e^{it} - 1) + a\lambda(e^{i(s+t)} - 1) \right].
\]
Therefore,
\[
\lim_{\lambda \to 1^-} \psi_{X,Y}(s, t) = \exp \left[ a(e^{i(s+t)} - 1) \right] \equiv \psi_X(s + t),
\]
so we obtain
\[
V^2(X, X) = V^2(Y, Y) = \lim_{\lambda \to 1^-} V^2(X, Y).
\]

Acknowledgments

The research of Johannes Dueck and Dominic Edelmann was supported by the Deutsche Forschungsgemeinschaft (German Research Foundation) within the programme “Spatio/Temporal Graphical Models and Applications in Image Analysis,” grant GRK 1653. The research of Donald Richards was supported by the U.S. National Science Foundation, grant DMS-1309808; and by a Romberg Guest Professorship at the Heidelberg University Graduate School for Mathematical and Computational Methods in the Sciences, funded by the German Universities Excellence Initiative grant GSC 220/2.
References

[1] Anderson, T. W. (2003). An Introduction to Multivariate Statistical Analysis, third edition. Wiley, New York.

[2] Andrews, G. E., Askey, R., and Roy, R. (1999). Special Functions. Cambridge University Press, New York.

[3] Bar-Lev, S. K., Bshouty, D., Letac, G., Lu, I. and Richards, D. St. P. (1993). The diagonal multivariate natural exponential families and their classification. J. Theoret. Probab., 7, 883–929.

[4] Diaconis, P., Khare, K. and Saloff-Coste, L. (2008). Gibbs sampling, exponential families and orthogonal polynomials. Statist. Sci., 23, 151–178.

[5] Dueck, J., Edelmann, D., Gneiting, T. and Richards, D. St. P. (2012). The affinely invariant distance correlation. Preprint, http://arxiv.org/abs/1210.2482v1.

[6] Dueck, J., Edelmann, D., Gneiting, T. and Richards, D. St. P. (2014). The affinely invariant distance correlation. Bernoulli, 20, 2305–2330.

[7] Dueck, J., Edelmann, D., and Richards, D. (2015). A generalization of an integral arising in the theory of distance correlation. Statist. & Probab. Lett., 97, 116–119.

[8] Gradshteyn, I. S., and Ryzhik, I. M. (1994). Tables of Integrals, Series, and Products, fifth edition (A. Jeffrey, Editor). Academic Press, New York.

[9] Huo, X., and Székely, G. J. (2014). Fast computing for distance covariance. Preprint, http://arxiv.org/abs/1410.1503.

[10] Kong, J., Klein, B. E. K., Klein, R., and Wahba, G. (2012). Using distance correlation and SS-ANOVA to access associations of familial relationships, lifestyle factors, diseases, and mortality. Proc. Natl. Acad. Sci. U. S. A., 109, 20352–20357.

[11] Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000). Continuous Multivariate Distributions: Models and Applications, second edition. Wiley, New York.

[12] Koudou, A. E. (1996). Probabilités de Lancaster. Expo. Math., 14, 247–275.

[13] Koudou, A. E. (1998). Lancaster bivariate probability distributions with Poisson, negative binomial and gamma margins. Test, 7, 95–110.

[14] Lancaster, H. O. (1958). The structure of bivariate distributions. Ann. Math. Statist., 29, 719–736.

[15] Lancaster, H. O. (1969). The Chi-Squared Distribution. Wiley, New York.
[16] Lyons, R. (2013). Distance covariance in metric spaces. *Ann. Probab.*, **41**, 3284–3305.

[17] Martínez-Gómez, E., Richards, M. T., and Richards, D. St. P. (2014). Distance correlation methods for discovering associations in large astrophysical databases. *Astrophys. J.*, **781**, 39 (11 pp.).

[18] Pommeret, D. (2004). A characterization of Lancaster probabilities with margins in a multivariate additive class. *Sankhyā*, **66**, 1–19.

[19] Richards, M. T., Richards, D. St. P., and Martínez-Gómez, E. (2014). Interpreting the distance correlation results for the COMBO-17 survey. *Astrophys. J. Lett.*, **784**, L34 (5 pp.).

[20] Rizzo, M. L. and Székely, G. J. (2010). DISCO analysis: A nonparametric extension of analysis of variance. *Ann. Appl. Statist.*, **4**, 1034–1055.

[21] Sarmanov, O. V. (1966). Generalized normal correlation and two-dimensional Fréchet classes. *Soviet Math. Dokl.*, **7**, 596–599.

[22] Sarmanov, O. V. (1970a). Gamma correlation process and its properties. *Dokl. Akad. Nauk, SSSR*, **191**, 30–32.

[23] Sarmanov, O. V. (1970b). An approximate calculation of correlation coefficients between functions of dependent random variables. *Math. Notes Acad. Sciences USSR*, **7**, 373–377.

[24] Sarmanov, O. V. and Bratoeva, Z. N. (1967). Probabilistic properties of bilinear expansions in Hermite polynomials. *Theor. Probab. Appl.*, **12**, 470–481.

[25] Sejdinovic, D., Sriperumbudur, B., Gretton, A., and Fukumizu, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.*, **41**, 2263–2291.

[26] Srivastava, H. M., and Singhal, J. P. (1972). Some extensions of the Mehler formula. *Proc. Amer. Math. Soc.*, **31**, 135–141.

[27] Székely, G. J. and Rizzo, M. (2009). Brownian distance covariance. *Ann. Appl. Statist.*, **3**, 1236–1265.

[28] Székely, G. J. and Rizzo, M. (2012). On the uniqueness of distance correlation. *Statist. & Probab. Lett.*, **82**, 2278–2282.

[29] Székely, G. J. and Rizzo, M. (2013). The distance correlation $t$-test of independence in high dimension. *J. Multivariate Anal.*, **117**, 193–213.
[30] Székely, G. J. and Rizzo, M. (2014). Partial distance correlation with methods for dissimilarities. *Ann. Statist.*, 42, 2382–2412.

[31] Székely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007). Measuring and testing independence by correlation of distances. *Ann. Statist.*, 35, 2769–2794.

[32] Withers, C. S. and Nadarajah, S. (2010). Expansions for the multivariate normal. *J. Multivariate Anal.*, 101, 1311–1316.

[33] Zhou, Z. (2012). Measuring nonlinear dependence in time-series, a distance correlation approach. *J. Time Series Anal.*, 33, 438–457.