INAUGURAL - DISSERTATION

zur

Erlangung der Doktorwürde

der

Naturwissenschaftlich-Mathematischen Gesamtfakultät

der

Ruprecht - Karls - Universität

Heidelberg

vorgelegt von

Diplom-Mathematiker Johannes Dueck

aus

Bielefeld / Deutschland

Tag der mündlichen Prüfung:

Dependencies in Complex Systems

Betreuer: **Prof. Dr. Rainer Dahlhaus**
Prof. Dr. Tilmann Gneiting

Acknowledgements

I would like to thank...

... Prof. Dr. Rainer Dahlhaus, my main supervisor, of the Faculty of Mathematics and Computer Science of Heidelberg University, for the great support and very helpful discussions. For giving me the opportunity to work in diverse research fields on the highest academic level. For teaching me everything I know about statistics.

... Prof. Dr. Tilmann Gneiting. He brought my attention to distance correlation and without him a significant part of the thesis would have been impossible.

... Prof. Dr. Donald Richards and Prof. Dr. Mercedes Richards for being great friends. I have learned so much from them in many hours of discussions during their visit in Heidelberg and via Skype. Thank you so much for everything.

... the *Deutsche Forschungsgemeinschaft* (German Research Foundation) that supported me within the program “Spatio/Temporal Graphical Models and Applications in Image Analysis,” grant GRK 1653.

... all of my colleagues, especially Fabian Bachl, Fabian Rathke, Eva Didden, Bernhard Schmitzer, Sebastian Schweer and Dr. Cornelia Wichelhaus.

... Dominic Edelmann for being a truly amazing friend and coauthor. We have been together in one office for over five years. He could bear my whistling and singing. We know each other so well that it is scary and people think, that we are one person.

... Peter Büermann, Maike Frank, Matthias Katzfuss, Kristin Meyer, Sarah Müller and Steffen Müller for being the best friends that I could have ever hoped for. For always being there for me and sharing the happiest moments of my life with me. PL for life. I would do everything for you.

... Mathias Stotz for taking me to the ”Therme Erding” for the first time and being a brilliant brother-in-law.

... Stefan, Helen, Phil, Sasha and Nick for celebrating Thanksgiving with me with a huge turkey. It is just such a pleasure to be a part of your family.

... Gisi and Dietmar Doleschel. You are my second home. I celebrated every Christmas of my life there. I will never forget about all the support and everything you have done for me.

... my older sister Anne Dueck. She is a passionate Bayern München fan like me and she always goes to the stadium with me. She is a role model for me and I can’t express in words how much she means to me. She always has advice and she is the only person I know that instantaneously realizes if I am telling the truth or not.

... my parents, Monika and Gunter Dueck. There is not a single thing that I can complain about concerning my upbringing. I had no sorrow at all. My whole life I could grow up in a well protected, happy and caring environment. They taught me everything I know and never asked for anything back. So this way: You are the perfect parents and I love you from the bottom of my heart! Everything I have achieved is due to your guidance.

Abstract

A task in statistics is to find meaningful associations or dependencies between multivariate random variables or in multivariate, time-dependent stochastic processes. Hawkes (1971) introduced the powerful multivariate point process model of mutually exciting processes (Hawkes model) to explain causal structure in data. Therefore, we discuss several causality concepts and show that causal structure is fully encoded in the corresponding Hawkes kernels. Hence, for causal inference and for establishing graphical models induced by causality it is necessary to estimate the Hawkes kernels. We provide a nonparametric, consistent and asymptotically normal estimator of the Hawkes kernels depending on the increments on a time scale with mesh Δ using methods from infinite order regression and time series analysis. To illustrate our results we apply our method to EEG data from the spinal dorsal horn of a rat.

To tackle the problem for random samples of random vectors we examine a new dependence measure, namely distance correlation (Székely, Rizzo and Bakirov; 2007). Distance correlation provides a strikingly simple sample version in order to test for independence between two random vectors of arbitrary dimensions and finite first moments. However, distance correlation is not well understood on the population side and it fails to be invariant under the group of all invertible affine transformations. Hence, we introduce the affinely invariant distance correlation and compute the analytic usual distance correlation and affinely invariant distance correlation in various settings: for multivariate normal distributions and for Lancaster probabilities (e.g. the bivariate gamma distribution) explicitly. Furthermore, we generalize an integral which is at the core of distance correlation.

Abstract (Deutsch)

Eine Aufgabe der Statistik besteht darin, aussagekräftige Verknüpfungen und Abhängigkeiten zwischen multivariaten Zufallsvariablen oder in multivariaten, zeitabhängigen stochastischen Prozessen zu finden. Hawkes (1971) führte das mächtige Punktprozessmodell der wechselseitig anregenden Prozesse ein (Hawkes-Modell), um kausale Strukturen in Daten zu erklären. Daher diskutieren wir verschiedene Kausalitätskonzepte und zeigen, dass die kausale Struktur vollständig in den Hawkes-Kernen verschlüsselt ist. Deshalb ist es für die kausale Inferenz und für das Betrachten von induzierten Kausalitätsgraphen nötig, die Hawkes-Kerne zu schätzen. Wir leiten einen nichtparametrischen, konsistenten und asymptotisch normalen Schätzer für die Hawkes-Kerne her. Dieser beruht auf einer Diskretisierungsmethode und den Zuwächsen des Prozesses auf einem Gitter der Feinheit Δ. Folglich benutzen wir verallgemeinerte Methoden der Regression unendlicher Ordnung und der Zeitreihenanalyse. Wir veranschaulichen unseren Schätzer am Beispiel von EEG-Daten aus dem Hinterhorn des Rückenmarks einer Ratte.

Um auf das Problem bezüglich zufälliger Stichproben von Zufallsvektoren einzugehen, untersuchen wir ein neues Abhängigkeitsmaß, nämlich die Distanzkorrelation (Székely, Rizzo and Bakirov; 2007). Die Distanzkorrelation bietet eine erstaunlich einfache empirische Version, um die Unabhängigkeit zwischen zwei Zufallsvektoren von beliebiger Dimension und endlichen ersten Momenten zu testen. Jedoch ist die Distanzkorrelation auf der Populationsseite nicht gut verstanden. Außerdem ist sie nicht invariant unter der Gruppe aller invertierbarer affiner Transformationen. Daher führen wir die affin invariante Distanzkorrelation ein und berechnen explizit die gewöhnliche Distanzkorrelation und die affin invariante Distanzkorrelation in diversen Situationen: für multivariate Normalverteilungen und für Lancasterverteilungen (z.B. die bivariate Gammaverteilung). Darüber hinaus verallgemeinern wir ein Integral, welches das Herzstück der Distanzkorrelation bildet.

Contents

1 Notations

Throughout this thesis we use the following notations: We denote by $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ the sets of nonnegative integers, integers, rational numbers, real numbers and complex numbers, respectively. For a complex number $z \in \mathbb{C}$ we let $|z|$ be its modulus and $\Re(z)$ the real part of z. For a matrix $A \in \mathbb{R}^{n\times k}$ we denote by A' its transpose. If we face a real vector $t \in \mathbb{R}^d$, then $|t|_d$ represents the standard Euclidean norm of t. Hence, if $t = (t_1, \dots, t_d)'$ then

$$|t|_d = (t_1^2 + \cdots + t_d^2)^{1/2}.$$

For vectors u and v of the same dimension, we let $\langle u, v\rangle$ be the standard Euclidean scalar product of u and v. If we need to put emphasis on the dimension of the space where the vectors are elements from, we use the notation $\langle u, v\rangle_d$ for $u, v \in \mathbb{R}^d$. We call the identity matrix, in the matrix space $\mathbb{R}^{d\times d}$, I_d. Furthermore, $\mathbb{P}$ always stands for a probability measure on the corresponding probability space and $\mathbb{E}$ for the expected value. For two random variables X and Y we denote by $\mathrm{var}(X)$ the variance of X and by $\mathrm{cov}(X, Y)$ the covariance between X and Y (or the covariance matrix). We also make use of the abbreviations Σ_X and Σ_Y for the variance/covariance matrices of X and Y, respectively, and Σ_{XY} for the cross-covariance matrix between X and Y. The normal distribution is represented by the symbol $\mathcal{N}$. Stochastic convergence is abbreviated $\xrightarrow{p}$ and convergence in distribution $\xrightarrow{\mathcal{D}}$. We use the standard big O and small o notation. For a matrix $A \in \mathbb{R}^{n\times k}$ we denote by $\|A\|$ the Frobenius norm and by $\|A\|_1$ the spectral radius.

For a stationary time series y, f_y denotes its spectral density. Finally, we let Δ be an arbitrary positive real number and k a positive integer.

2 Introduction

In this thesis we examine dependencies in complex systems. In particular, we mainly focus on meaningful associations, such as causality and stochastic independence, for multivariate data, e.g. generated by multivariate random variables or multivariate processes. It is natural to consider causality in time-dependent systems (the cause should be always before the effect). Therefore, we introduce an appropriate point process model, the so-called Hawkes model, which reflects causal structures in point process data. In order to analyze multivariate random variables (in general not time-dependent) we suggest to use the new dependence measure distance correlation. Still, the population quantity of distance correlation is not well understood. In our work we put a lot of effort in clarifying the analytic distance correlation version for certain probability distributions.

Hawkes (1971) first introduced mutually exciting processes (Hawkes model). Historically, it was motivated by modeling aftershocks and seismological phenomena. In several papers over the last years, many authors dealt with modeling earthquake effects using mutually exciting processes, e.g. see Ogata (1999), Vere-Jones (1970) and Vere-Jones and Ozaki (1982) for details. However, the usage of the Hawkes model has been more and more spread out to different research areas: Brantingham, et al. (2011) examine insurgency in Iraq; Mohler, et al. (2011) use it for modeling crime and Reynaud-Bouret and Schbath (2010) apply it to genome analysis. The most recent advance of the Hawkes model is in finance for price fluctuations or transactions, see Bacry, et al. (2011), Bacry, et al. (2012) and Embrechts, Liniger and Lu (2011).

In this thesis we put the emphasis on the causal structure induced by Hawkes model. We connect concepts for causality with mutually exciting processes. Granger (1969) defined the notion of Granger causality. It reflects the belief that a cause should always occur before the effect and that a prediction of a process with the knowledge of a possible cause should improve if there is a causal relation present. Since then, new concepts concerning causality in different settings have been developed, namely local independence, Didelez (2008), and weakly instantaneously causality, Florens and Fougere (1996), below others.

A Hawkes process is a multivariate point process with conditional arrival rate

$$\Lambda(t) = \nu + \int_{-\infty}^{t} \gamma(t-u)\, dN(u),$$

where ν is a vector of positive constants often referred to as background or Poisson rates and $\gamma(\cdot)$ is a matrix of nonnegative functions that vanish on the negative half axis. It may be viewed as a continuous analogon to classical auto-regression in time series analysis. Every time a component process jumps, it can excite the other processes according to the so called decay or Hawkes kernel $\gamma(\cdot)$. Hence, a causal structure is encoded in the Hawkes kernels. The problem is to estimate these decay kernels. The most popular approach is an ML-approach as in Ozaki (1979). Therefore, γ is assumed to be of a parametric form, e.g. exponential functions or Laguerre polynomials. In recent literature other estimation procedures evolved. Bacry, et al. (2012) use a numerical method for the nonparametric estimation based on martingale and Laplace transform techniques and Lewis and Mohler (2011) EM-algorithms. We estimate Hawkes kernels nonparametrically using infinite order regression methods as Lewis and Reinsel (1985) by discretizing the axis corresponding to a mesh Δ.

For the purpose of testing for noncausality the asymptotic distribution of the estimator is needed. We show that the estimator is indeed asymptotically normal. Finally, we apply our estimation procedure to EEG data from the spinal dorsal horn of a rat and to simulated data.

Székely, Rizzo and Bakirov (2007) and Székely and Rizzo (2009), in two seminal papers, introduced the distance covariance and distance correlation as powerful measures of dependence. Contrary to the classical Pearson correlation coefficient, the population distance covariance vanishes only in the case of independence, and it applies to random vectors of arbitrary dimensions, rather than to univariate quantities only.

As noted by Newton (2009), the "distance covariance not only provides a *bona fide* dependence measure, but it does so with a simplicity to satisfy Don Geman's *elevator test* (i.e., a method must be sufficiently simple that it can be explained to a colleague in the time it takes to go between floors on an elevator!)." In the case of the sample distance covariance, find the pairwise distances between the sample values for the first variable, and center the resulting distance matrix; then do the same for the second variable. The square of the sample distance covariance equals the average entry in the componentwise or Schur product of the two centered distance matrices. Given the theoretical appeal of the population quantity, and the striking simplicity of the sample version, it is not surprising that the distance covariance is experiencing a wealth of applications, despite having been introduced merely half a decade ago.

In later papers, Rizzo and Székely (2010, 2011) and Székely and Rizzo (2012, 2013, 2014) gave applications of the distance correlation concept to several problems in mathematical statistics. In recent years, an enormous number of papers have appeared in which the distance correlation coefficient has been applied to many fields. In particular, the concept of distance covariance has been extended to abstract metric spaces (Lyons, 2013) and has been related to machine learning (Sejdinovic, Sriperumbudur, Gretton, and Fukumizu, 2013); to detecting associations in large astrophysical databases (Martínez-Gomez, Richards, and Richards, 2014) and to interpreting those associations (Richards, Richards, and Martínez-Gomez, 2014); to measuring nonlinear dependence in time series data (Zhou, 2012); and to numerous other fields.

To recapitulate the definition of distance correlation we let p and q be positive integers. For column vectors $s \in \mathbb{R}^p$ and $t \in \mathbb{R}^q$, denote by $|s|_p$ and $|t|_q$ the standard Euclidean norms on the corresponding spaces; thus, if $s = (s_1, \ldots, s_p)'$ then $|s|_p = (s_1^2 + \cdots + s_p^2)^{1/2}$, and similarly for $|t|_q$. Given d-dimensional vectors u and v, we let $\langle u, v\rangle_d$ 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

$$f_{X,Y}(s,t) = \mathbb{E} \exp\big[i\, \langle s, X\rangle_p + i\, \langle t, Y\rangle_q\big]$$

be the joint characteristic function of (X, Y), and let $f_X(s) = f_{X,Y}(s, 0)$ and $f_Y(t) =$

$f_{X,Y}(0,t)$ be the marginal characteristic functions of X and Y, respectively. Székely, et al. (2007) defined $\mathcal{V}(X,Y)$, the *distance covariance* between X and Y, as

$$\mathcal{V}(X,Y) = \left[\frac{1}{c_p c_q}\int_{\mathbb{R}^{p+q}} \frac{|f_{X,Y}(s,t) - f_X(s) f_Y(t)|^2}{|s|_p^{p+1}\,|t|_q^{q+1}}\,\mathrm{d}s\;\mathrm{d}t\right]^{1/2}, \tag{2.1}$$

where $|z|^2$ denotes the squared modulus of $z \in \mathbb{C}$ and

$$c_p = \frac{\pi^{(p+1)/2}}{\Gamma\big((p+1)/2\big)}. \tag{2.2}$$

The *distance correlation* between X and Y is the nonnegative number defined by

$$\mathcal{R}(X,Y) = \frac{\mathcal{V}(X,Y)}{\sqrt{\mathcal{V}(X,X)\mathcal{V}(Y,Y)}} \tag{2.3}$$

if both $\mathcal{V}(X,X)$ and $\mathcal{V}(Y,Y)$ are strictly positive, and $\mathcal{R}(X,Y)$ is defined otherwise to be zero. For distributions with finite first moments, the distance correlation coefficient characterizes independence in that $0 \le \mathcal{R}(X,Y) \le 1$, and $\mathcal{R}(X,Y) = 0$ if and only if X and Y are mutually independent.

When using the concept of distance correlation in applications one faces the problem that distance correlation is not invariant under the group of all invertible affine transformations. A main contribution of this thesis consists of the introduction of an affinely invariant distance correlation coefficient. This new measure inherits all basic characteristics from usual distance correlation but is equipped with the additional group invariance. We review the sample version of the affinely invariant distance correlation introduced by Székely, et al. (2007), and we prove that the sample version is strongly consistent. Moreover, we provide exact expressions for the affinely invariant distance correlation in the case of subvectors from a multivariate normal population of arbitrary dimension, thereby generalizing a result of Székely, et al. (2007) in the bivariate case. Our result is non-trivial: It is derived using the theory of zonal polynomials and the hypergeometric functions of matrix argument, and it enables the explicit and efficient calculation of the affinely invariant distance correlation in the multivariate normal case. To get a better understanding of affinely invariant distance correlation in higher dimensions, we outline the behavior of the affinely invariant distance measures for subvectors of multivariate normal populations in limiting cases as the Frobenius norm of the cross-covariance matrix converges to zero, or as the dimensions of the subvectors converge to infinity. We expect that these results will motivate and provide the theoretical basis for many applications of distance correlation measures for high-dimensional data.

After the presentation of our theoretical results we study the example of time series of wind vectors at the Stateline wind energy center in Oregon and Washington; we shall derive the empirical auto and cross distance correlation functions between wind vectors at distinct meteorological stations.

In further sections, we are able to compute regular distance correlation for the normal distribution and the Laplace distribution. Unfortunately, the calculation of population distance correlation coefficients remains an intractable problem. As a consequence, it is not possible to calculate distance correlation coefficients explicitly for given nonnormal distributions in terms of the parameters that parametrize these distributions; nor is it possible to ascertain for nonnormal distribution any analogs of the mentioned limit theorems.

We describe in detail the difficulties in calculating the population distance covariance coefficient. For any pair of random vectors X and Y, the fundamental obstacle in calculating the population distance correlation coefficient is the computation of the singular integral (2.1) which defines these coefficients. We solve this problem as follows: First, we note that the singular nature of the integrand precludes evaluation of the integral by squaring the denominator and then integrating each of the resulting three terms. Then we compute 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 (1958). It is well-known that the distribution functions of the Lancaster family have appealing expansions in terms of certain orthogonal functions (Koudou, 1998; Diaconis, et al., 2008). By applying these orthogonal expansions, we deduce that the corresponding characteristic functions can be expanded as infinite series, and it is those series which lead to explicit expressions for the corresponding distance covariance and distance correlation coefficients (e.g. for a bi-variate gamma distribution).

In the final section we generalize the fundamental singular integral, which allows distance correlation to possess a strikingly simple sample quantity.

3 Nonparametric Estimation of Hawkes Kernels and Graphical Models

In this section we follow Dahlhaus, Dueck and Eichler (2014). We introduce the powerful Hawkes model which yields a framework for examining causal structures in point processes. We make the connection between Hawkes model and Granger causality and conclude that it is necessary for causal inference to estimate the Hawkes kernels. Therefore, we provide a nonparametric estimator, which is shown to be consistent and asymptotically normal.

3.1 Hawkes Process and Graphical Models

We first define a mutually exciting or Hawkes process and then show and discuss connections to Granger causality and Granger causality graphs. In our setting we consider a stationary d-variate point process $N = (N(t), t \in \mathbb{R})$. $N_a(t) = N_a([0,t])$ for $t \geq 0$ and $a \in \{1, \ldots, d\}$ is supposed to represent the cumulative number of events in the ath process from time zero up to time t. We denote the intensity vector of N by

$$\lambda = \mathbb{E}[dN(t)]/dt.$$

A mutually exciting process N does not have multiple jumps at the same time instant, which is regularly in literature referred to as a simple point process. Additionally, a Hawkes process is orderly, which means that the jumps of the process are isolated. Following Hawkes (1971) we establish a linear dependence of the conditional intensity function at a specific time instant t on the full history of the process N by defining

$$\mathbb{P}(N_a(t+h) - N_a(t) > 0 \,|\, N(s), s \leq t) = \Lambda_a(t)\, h + o(h), \tag{3.1}$$

where $a \in \{1, \ldots, d\}$. By specifying the conditional intensity function, the distribution of Hawkes process is fully determined which is a basic result on point processes. Hawkes sets

$$\Lambda(t) = \nu + \int_{-\infty}^{t} \gamma(t-u)\, dN(u)$$

with $\gamma_{ij}(u) = 0$, for $u < 0$, such that

$$\Lambda_a(t) = \nu_a + \sum_{j=1}^{d} \int_0^\infty \gamma_{aj}(u)\, dN_j(t-u), \tag{3.2}$$

where ν is a deterministic vector whose elements are all nonnegative. To guarantee the nonnegativity of the conditional intensity function the Hawkes kernels have to fulfill

$\gamma_{ij}(u) \geq 0$ for all $u \in \mathbb{R}$. The vector ν is the basic Poisson rate (independent of each other) for the component processes to jump. After having jumped each component process excites another process according to the link function γ. That is why Hawkes refers to it as "model of mutually exciting point processes".
If the process is stationary it is straight forward to derive the useful relation

$$\lambda = \Big(I - \int_{-\infty}^{\infty} \gamma(t)\, dt\Big)^{-1} \nu. \tag{3.3}$$

A full explicit analysis of the covariance structure can be found in Bacry, et al. (2012). In his paper Hawkes computes the point spectral density of the mutually exciting process under the condition that the link functions are of exponential decay. In the following we will stick to the notation of Hawkes. To summarize define the covariance density matrix as

$$\mu(\tau) = \mathbb{E}\big[dN(t+\tau)\, dN'(t)\big]/(dt)^2 - \lambda\lambda'.$$

Due to stationarity μ does not depend on t. Because of Bartlett (1963) it is well known that the covariance density matrix for the Hawkes Process is well defined. Furthermore it holds $\mu(-\tau) = \mu'(\tau)$. Since Hawkes Process is a simple point process it is degenerate for $\tau = 0$, i.e. $\mathbb{E}\big[dN_i(t)^2\big] = \mathbb{E}\big[dN_i(t)\big]$. Thus, Hawkes defines the complete covariance density as

$$\mu^{(c)}(\tau) = D\,\delta(\tau) + \mu(\tau),$$

where $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_d)$ and δ is the Dirac delta function. Notice that $\mu(\tau)$ is continuous at the origin. Then, the point spectral density matrix takes the form

$$\begin{aligned} f(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega\tau}\, \mu^{(c)}(\tau)\, d\tau \\ &= M(\omega) + \frac{1}{2\pi} D \end{aligned}$$

where M is the Fourier transform of μ, i.e.

$$M(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega\tau}\, \mu(\tau)\, d\tau.$$

Additionally Hawkes derives an implicit expression for μ,

$$\mu(\tau) = \gamma(\tau)\, D + \int_{-\infty}^{\infty} \gamma(\tau - u)\, \mu(u)\, du, \quad \text{for } \tau > 0,$$

which results in solving a fundamental integral equation subject to the condition $\mu(-\tau) = \mu'(\tau)$. The latter equality is similar to the Wiener-Hopf integral equation. Now defining

$$\beta(\tau) = \gamma(\tau)\, D + \int_{-\infty}^{\infty} \gamma(\tau - u)\, \mu(u)\, du - \mu(\tau), \quad -\infty < \tau < \infty,$$

and assuming

$$\gamma_{ij}(u) < A\, e^{-\eta u},$$
$$|\beta_{ij}(u)| < B\, e^{\eta u}$$

for some constants A and B, Hawkes is able to explicitly compute

$$f(\omega) = \frac{1}{2\pi}\big[I_d - G(\omega)\big]^{-1} D \,\big[I_d - G'(-\omega)\big]^{-1},$$

where $G(\omega) = \int_{-\infty}^{\infty} e^{-i\omega\tau}\, \gamma(\tau)\, d\tau$.

If γ vanishes in any component then it should be intuitive, that there is no respective causal relation. Hence, desiring to establish a graphical model based on noncausality in this sense, we have to find an appropriate definition of causality in order to capture the latter intuition. Also notice, that in the case where the ath row of γ vanishes identically, then N_a is a regular Poisson process with rate ν_a.

Definition 3.1. *Set $\mathcal{F}_t = \sigma(N(s), s \leq t)$, $\mathcal{F}_t^{-a} = \sigma(N_{V\setminus\{a\}}(s), s \leq t)$, ($V = \{1, \ldots, d\}$) and*

$$\Lambda_b^{-a}(t) = \lim_{h\downarrow 0} \frac{\mathbb{P}(N_b(t+h) - N_b(t) > 0 \,|\, \mathcal{F}_t^{-a})}{h}.$$

We say that N_a is globally or Granger noncausal for N_b relative to the process N_V, denoted by $N_a \not\rightarrow_G N_b[N_V]$, if

$$\Lambda_b(t) = \Lambda_b^{-a}(t) \quad a.s.\ \forall\, t \geq 0.$$

Granger causality (Granger 1969) displays, that N_a is noncausal for N_b relative to the process N_V if the prediction of N_b with respect to the full history of the process is as good as the prediction with respect to the full history except for the history of the process N_a. As we will find in the next theorem, Granger noncausality corresponds to vanishing of the associated Hawkes kernels contained in the matrix function γ in the Hawkes model. This immediately implies that we may establish Granger causality graphs based on the estimation of the link functions in the Hawkes model.

Theorem 3.2. *For the Hawkes model N_a is globally noncausal for N_b relative to the process N_V if and only if γ_{ba} vanishes, i.e.*

$$N_a \not\rightarrow_G N_b[N_V] \quad \Leftrightarrow \quad \gamma_{ba} \equiv 0. \tag{3.4}$$

PROOF. "$\Leftarrow$" First suppose $\gamma_{ba} \equiv 0$. Then

$$
\begin{aligned}
\Lambda_b^{-a}(t) &= \lim_{h\downarrow 0} \frac{\mathbb{P}(N_b(t+h)-N_b(t)>0\,|\,\mathcal{F}_t^{-a})}{h} \\
&= \lim_{h\downarrow 0} \frac{\mathbb{E}(\mathbb{1}_{\{N_b(t+h)-N_b(t)>0\}}\,|\,\mathcal{F}_t^{-a})}{h} \\
&= \lim_{h\downarrow 0} \frac{\mathbb{E}(\mathbb{E}(\mathbb{1}_{\{N_b(t+h)-N_b(t)>0\}}\,|\,\mathcal{F}_t)\,|\,\mathcal{F}_t^{-a})}{h} \\
&= \lim_{h\downarrow 0} \frac{\mathbb{E}([\nu_b+\sum_{k\in V\setminus\{a\}}\int_0^\infty \gamma_{bk}(u)\,dN_k(t-u)]\,h+o(h)\,|\,\mathcal{F}_t^{-a})}{h} \\
&= \nu_b + \sum_{k\in V\setminus\{a\}} \int_0^\infty \gamma_{bk}(u)\,dN_k(t-u) \\
&= \Lambda_b(t),
\end{aligned}
$$

where we used (3.1), measurability with respect to $\mathcal{F}_t^{-a}$.

" $\Rightarrow$ " Next assume $\Lambda_b(t) = \Lambda_b^{-a}(t)$ for all t. This identity is equivalent to the equation

$$
\int_0^\infty \gamma_{ba}(u)\,dN_a(t-u) = \mathbb{E}\Big[\int_0^\infty \gamma_{ba}(u)\,dN_a(t-u)\,|\,\mathcal{F}_t^{-a}\Big].
$$

Clearly, this immediately implies that the left hand side is measurable with respect to $\mathcal{F}_t^{-a}$. Since the state space of N is polish, we find by some factorization lemma of conditional expectations, that there must exist an $\mathcal{F}_t^{-a}$-measurable function h with

$$
\int_0^\infty \gamma_{ba}(u)\,dN_a(t-u) = h(\bar{N}_{V\setminus a}(t)),
$$

which must mean $\gamma_{ba} \equiv 0$.

□

The next proposition works out the fact, that the equality of the above conditional rates (noncausality) can be rephrased in terms of conditional independence.

Proposition 3.3. *With the same notations as before*

$$
N_a \not\to_G N_b[N_V] \quad \Leftrightarrow \quad \Lambda_b^{-a}(t) \perp \bar{N}_a(t) \big| \bar{N}_{V\setminus a}(t) \quad \forall t,
$$

where for any $A \subseteq V$, $\bar{N}_A(t) = \{N_A(s), s \leq t\}$.

Proof.
" $\Rightarrow$ " For the first step let $N_a \not\to_G N_b[N_V]$. Then Theorem **3.2** tells us $\gamma_{ba} \equiv 0$ and $\Lambda_b(t) = \Lambda_b^{-a}(t)$. The latter yields

$$\Lambda_b^{-a}(t) = v_b + \sum_{k \in V \setminus \{a\}} \int_0^\infty \gamma_{bk}(u)\, dN_k(t-u),$$

which is clearly independent of $\bar{N}_a(t)$ given $\bar{N}_{V\setminus a}(t)$.

" $\Leftarrow$ " Now assume $\Lambda_b^{-a}(t) \perp \bar{N}_a(t) \big| \bar{N}_{V\setminus a}(t)$ for all t. Parallel computations as above lead to

$$\Lambda_b^{-a}(t) = \mathbb{E}\Big(\Big[v_b + \sum_{k=1}^{d} \int_0^\infty \gamma_{bk}(u)\, dN_k(t-u) \Big] \Big| \mathcal{F}_t^{-a} \Big)$$

If γ_{ba} would not vanish for all t, this would directly lead to contradiction. Hence, Theorem **3.2** again gives us $N_a \not\to_G N_b[N_V]$. □

In the next subsections we state different kinds of causal relation definitions and discuss them. We find that the causality concepts are equivalent for the Hawkes model.

3.1.1 Local Independence

Didelez (2008) defines local independence for finite marked point processes in continuous time. The key idea of local independence is that under rather mild conditions one may decompose the component processes $N_k(t)$ into a compensator and a martingale dependent on the chosen filtration (Doob Meyer decomposition). If we choose the internal filtration of the whole process the compensator concerning $N_k(t)$ takes the form

$$Z_k(t) = \int_0^t \mathbb{E}(N_k(ds) | \mathcal{F}_{t^-}).$$

We now state a specified version for the component processes of local independence by Didelez.

Definition 3.4 (Local Independence). *Let $N = (N_1, \dots, N_d)$ be a multivariate point process. We say that N_a is locally independent of N_b, for $a \neq b$, given $N_{V\setminus\{a,b\}}$ if the $\mathcal{F}_t$ compensator Z_a is measurable with respect to $\mathcal{F}_t^{-b}$ for all t. We denote this by $N_a \not\to_L N_b[N_V]$.*

The key idea of local independence is, that if we reduce the whole filtration $\mathcal{F}_t$ to $\mathcal{F}_t^{-b}$, the compensator Z_a, which is a short-term prediction, remains unchanged.

In view of the linear structure of the conditional intensity function in the Hawkes model, one would expect equivalence of the concepts of local and global noncausality in this situation. This is the subject of the next theorem.

Theorem 3.5. *In the Hawkes model $N_b \not\to_G N_a[N_V]$ if and only if $N_b \not\to_L N_a[N_V]$.*

PROOF. We show $N_b \not\to_L N_a[N_V]$ if and only if $\gamma_{ba} \equiv 0$ and hence by Theorem **3.2** the assertion. We obtain

$$\begin{aligned} Z_a(t) &= \int_0^t \mathbb{E}(N_a(ds)|\mathcal{F}_{t^-}) = \int_0^t \Lambda_a(s)\,ds \\ &= \int_0^t \Big(\nu_a + \sum_{j=1}^d \int_0^\infty \gamma_{aj}(u)\,dN_j(s-u)\Big)ds \\ &= \nu_a t + \sum_{j=1}^d \int_0^t \int_0^\infty \gamma_{aj}(u)\,dN_j(s-u)ds. \end{aligned}$$

Therefore, $Z_a(t)$ is measurable with respect to $\mathcal{F}_t^{-b}$ if and only if $\gamma_{ba} \equiv 0$. $\square$

3.1.2 Weakly Instantaneously Causality

Florens and Fougere (1996) yield a similar definition for a certain type of noncausality as the local independence concept, namely the weakly instantaneously causality. However, this definition is formulated in a rather general context and in a slightly more technical way. In their setting they consider Markov and counting processes as well as a very general class of stochastic processes. The authors put the focus on the close relation between causality and martingale properties of a stochastic process. Let $\mathcal{F} = (\mathcal{F}_t)_{t\in I}$ be a filtration and z_t a real-valued stochastic process adapted to $\mathcal{F}$. Furthermore set $\mathcal{G} = (\mathcal{G}_t)_{t\in I}$ to be a sub-σ field of $\mathcal{F}$, which means $\mathcal{G}_t \subset \mathcal{F}_t$ for all $t \in I$. We are ready to state the definition of weakly instantaneously causality out of Florens and Fougere.

Definition 3.6 (Weakly Instantaneously Causality)**.** *Assume that z_t is a semi-martingale with respect to $\mathcal{G} = (\mathcal{G}_t)_{t\in I}$, such that a decomposition $z_t = z_0 + H_t^* + M_t^*$ exists. Then $\mathcal{F}$ does not weakly instantaneously cause z_t given $\mathcal{G}$, if z_t remains a semi-martingale with respect to $\mathcal{F}$ with the same decomposition.*

Let us recapitulate this definition for our special setting:

Definition 3.7 (Weakly Instantaneously Causality)**.** *Let $N = (N_1, \dots, N_d)$ be a multivariate point process. Let $\mathcal{F}$ be the canonical filtration generated by N. Suppose that N_a is a special semi-martingale with respect to $\mathcal{F}^{-b}$, so that a decomposition $N_a(t) = N_a(0) + H_t^* + M_t^*$ exists. Then N_b does not weakly instantaneously cause N_a given N_V ($N_b \not\to_{WI} N_a[N_V]$), if $N_a(t)$ remains a semi-martingale with respect to $\mathcal{F}$ with the same decomposition (for all $t \in I$).*

The proof of the next theorem is immediate.

Theorem 3.8. *In the Hawkes model $N_a \not\to_{WI} N_b[N_V]$ if and only if $N_a \not\to_L N_b[N_V]$.*

$\square$

In all of the latter concepts causality is explicitly expressed via the link functions γ_{ab} in the Hawkes model. One can therefore induce a causality graph structure. In this graph there is no directed edge from a to b, if N_a is noncausal for N_b, i.e. if $\gamma_{ba} \equiv 0$. For the inference of the causal structure two steps are necessary. We have to estimate the Hawkes kernels nonparametrically and we have to derive the asymptotic distribution of our estimator. The next section is devoted to the latter problem. The asymptotic distribution can be used in future work to test if the Hawkes kernels are identically equal to zero or not.

3.2 Nonparametric Estimation and Identification

Our approach for the estimation of the link function γ is via discretization and consequently using methods from time series analysis. Again, as in section 3.1 we observe a multi-dimensional point process $N = (N(t) : t \in \mathbb{R})$ with component processes N_i, $1 \leq i \leq d$, where d is the dimension of the process N. The conditional intensity function is once more given by (3.1) and (3.2), where the component functions of γ belong to a nicely behaving but general class, that is specified later on in the section. Our desire is to estimate the link functions of the Hawkes process nonparametrically, i.e. we do not assume any parametric form of the link functions, e.g. an exponential form. For the purpose of discretization define

$$\begin{aligned} y^i_{\Delta,t} &:= N_i((t+1)\cdot\Delta) - N_i(t\cdot\Delta), \\ \mathcal{F}^{\Delta,t-1} &:= \mathcal{F}_{t\cdot\Delta}, \end{aligned}$$

for all $t \in \mathbb{Z}$, $1 \leq i \leq d$ and for any positive constant Δ. This is equivalent to dividing the real line into intervals of width Δ. For every fixed Δ, y_Δ represents a d-dimensional time series displaying the number of jumps in time intervals of the form $[t\cdot\Delta, (t+1)\cdot\Delta]$. Thus, for Δ small enough the random variables $y_{\Delta,t}$ are approximately binary. Notice that the considered point processes are simple so that the probability that more than one jump takes place in the interval $[t, t+\Delta]$ is of order $o(\Delta)$. Additionally Hawkes process is stochastically continuous. With these observations we may calculate

$$\begin{aligned} \mathbb{E}[y^i_{\Delta,t}|\mathcal{F}^{\Delta,t-1}] &= \mathbb{P}(y^i_{\Delta,t} = 1 \,|\, \mathcal{F}^{\Delta,t-1}) + o(\Delta) \\ &= \mathbb{P}(N_i((t+1)\cdot\Delta) - N_i(t\cdot\Delta) = 1 \,|\, \mathcal{F}^{\Delta,t-1}) + o(\Delta) \\ &= \Big[\nu_i + \sum_{j=1}^{d} \int_0^\infty \gamma_{ij}(s)\, dN_j(t\cdot\Delta - s)\Big]\cdot\Delta + o(\Delta) \\ &= \Big[\nu_i + \sum_{j=1}^{d} \sum_{u\in\mathbb{N}_0} \int_{u\Delta}^{(u+1)\Delta} \gamma_{ij}(s)\, dN_j(t\cdot\Delta - s)\Big]\cdot\Delta + o(\Delta). \end{aligned}$$

The latter equation motivates a least squares approach for the nonparametric estimation. In order to see this make the following observations or approximations. For Δ small enough we can think of $\mathbb{E}[y^i_{\Delta,t}|\mathcal{F}^{\Delta,t}] \approx y^i_{\Delta,t}$ and the component function of γ being constant on the intervals $[t\Delta,(t+1)\Delta]$ for $t \in \mathbb{Z}$ as well as $o(\Delta)$ displaying some white noise process. Then the latter computation gives us approximately the structure

$$y_{\Delta,t} \approx \Delta\,\nu + \Delta \sum_{u\in\mathbb{N}_0} \gamma(\Delta u)\, y_{\Delta,t-u-1} + \varepsilon_t,$$

which is a regular infinite order regression problem and is known to be solvable by a least squares approach (at least for a fixed Δ).

However, in our case it is more complicated so that we need to be exact and have a closer look at the possible class of functions that we may handle properly with such an approach. From an applicational point of view connected to e.g. EEG data as well as from a mathematical point of view it is reasonable to assume that γ_{ij} is directly Riemann integrable and integrable with respect to the point process N for every $1 \le i,j \le d$. Furthermore, this implies

$$\begin{aligned} &\gamma_{ij}(t) \to 0 \quad \text{as } t \to \infty, \\ &\gamma_{ij}(t) < C \quad \text{uniformly for all } 1 \le i,j \le d \text{ and all } t \in \mathbb{R} \end{aligned}$$

for some positive constant $C > 0$. Our strategy is to discretize the real line to approximate the Hawkes process by some kind of infinite order regression. Then to obtain a consistent estimator we use methods derived by Lewis and Reinsel (1985). For simplicity we stick to the same notation as in Lewis and Reinsel (1985).

For this purpose define $\Gamma_\Delta(j) := \mathbb{E}(y_{\Delta,t}y'_{\Delta,t+j}) = \Gamma_\Delta(-j)'$ to emphasize that the $y_{\Delta,t}$ strongly depend on Δ and therefore as well does Γ.

We set $\Gamma'_{1,k,\Delta} := (\Gamma_\Delta(1)',\dots,\Gamma_\Delta(k)')$. Note that this is a $(d \times dk)$ matrix. Also set $\Gamma_{k,\Delta}$ to be the $(dk \times dk)$ matrix whose (m,n)th $(d \times d)$ block of elements is $\Gamma_\Delta(m-n)$, m,$n = 1,\dots,k$.
Suppose now that we observe the realizations $y_{\Delta,1},\dots,y_{\Delta,T}$. Without loss of generality we assume that $y_{\Delta,1},\dots,y_{\Delta,T}$ are centered. If this is not the case, we initially center the data by subtracting $\Delta\bar{y} := \Delta T^{-1}\sum_{t=1}^{T} y_{1,t}$. It is well known that the sample mean converges in probability to the population mean. We notice carefully that T is not the observation time but the number of observed time slots. The observation time itself equals $T \cdot \Delta$. We define our estimator as

$$\widehat{\boldsymbol{\gamma}}_k = \big(\widehat{\gamma}^k(0),\widehat{\gamma}^k(\Delta),\dots,\widehat{\gamma}^k((k-1)\Delta)\big) := \widehat{\Gamma}'_{1,k,\Delta}\,\widehat{\Gamma}^{-1}_{k,\Delta}\,\frac{1}{\Delta},$$

where $\widehat{\Gamma}_{1,k,\Delta} := (T-k)^{-1}\sum_{t=k}^{T-1} Y_{\Delta,t,k}\, y'_{\Delta,t+1}$, $\widehat{\Gamma}_{k,\Delta} := (T-k)^{-1}\sum_{t=k}^{T-1} Y_{\Delta,t,k}\, Y'_{\Delta,t,k}$ and $Y_{\Delta,t,k} = (y'_{\Delta,t},\dots,y'_{\Delta,t-k+1})'$ which is a column vector of length dk.
With these definitions we are able to derive the desired asymptotic results for $\widehat{\boldsymbol{\gamma}}_k$. Therefor we denote by $\|B\|^2 := tr(B'B)$ the Frobenius norm and by $\|B\|_1^2 := \lambda_{\max}(B'B)$

the spectral or operator norm. For the proof of the following main theorem we remind ourselves of the inequalities

$$\|AB\|^2 \leq \|A\|_1^2 \|B\|^2 \quad \text{and} \quad \|AB\|^2 \leq \|A\|^2 \|B\|_1^2,$$

as well as $\|A\|_1 \leq \|A\| \leq \sqrt{r}\|A\|_1$, where r is the rank of A. Furthermore the Frobenius norm is sub-multiplicative.

Theorem 3.9. *For $k \geq 1$ define $\boldsymbol{\gamma}_k := (\gamma(0), \gamma(\Delta), \ldots, \gamma((k-1)\Delta))$ and make the following assumptions:*

1. *Let k and Δ be functions of T such that*

$$k\Delta \to \infty \quad \text{and} \quad T^{-1}\frac{k^2}{\Delta} \to 0$$

as $T \to \infty$.

2. *We demand the stability condition $\left\| \int_0^\infty \gamma(u)du \right\|_1 < 1$.*

3. *The component functions of γ decrease in a way such that*

$$k^{1/2} \sum_{j=k}^{\infty} \|\gamma(\Delta j)\| \to 0$$

as $T \to \infty$.

4. *$\Delta^2 k \to 0$ as $T \to \infty$.*

Then our estimator is consistent in the sense

$$\Delta \, \|\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k\| \xrightarrow{p} 0 \quad \text{as } T \to \infty.$$

Proof. First of all we realize, that assumption 2 assures the stationarity of the mutually exciting process. Moreover we observe as in Lewis and Reinsel (1985) the useful relation

$$\begin{aligned}
\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k &= \widehat{\Gamma}'_{1,k,\Delta} \, \widehat{\Gamma}^{-1}_{k,\Delta} \, \frac{1}{\Delta} - \boldsymbol{\gamma}_k \, \widehat{\Gamma}_{k,\Delta} \, \widehat{\Gamma}^{-1}_{k,\Delta} \\
&= \left(\widehat{\Gamma}'_{1,k,\Delta} - \Delta \, \boldsymbol{\gamma}_k \, \widehat{\Gamma}_{k,\Delta}\right) \widehat{\Gamma}^{-1}_{k,\Delta} \frac{1}{\Delta} \\
&= \left((T-k)^{-1} \sum_{t=k}^{T-1} y_{\Delta,t+1} Y'_{\Delta,t,k} - \Delta \boldsymbol{\gamma}_k (T-k)^{-1} \sum_{t=k}^{T-1} Y_{\Delta,t,k} Y'_{\Delta,t,k}\right) \widehat{\Gamma}^{-1}_{k,\Delta} \frac{1}{\Delta} \\
&= \left((T-k)^{-1} \sum_{t=k}^{T-1} \left(y_{\Delta,t+1} - \Delta \boldsymbol{\gamma}_k Y_{\Delta,t,k}\right) Y'_{\Delta,t,k}\right) \widehat{\Gamma}^{-1}_{k,\Delta} \frac{1}{\Delta} \\
&= \left((T-k)^{-1} \sum_{t=k}^{T-1} \varepsilon_{\Delta,t+1,k} Y'_{\Delta,t,k}\right) \widehat{\Gamma}^{-1}_{k,\Delta} \frac{1}{\Delta},
\end{aligned}$$

where $\varepsilon_{\Delta,t,k} := y_{\Delta,t} - \sum_{j=0}^{k-1} \Delta\gamma(\Delta j)\, y_{\Delta,t-j-1}$.

Thus, we find

$$\Delta\|\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k\| \leq \|\Delta\widehat{\Gamma}_{k,\Delta}^{-1}\|_1 \Big(\|U_{1T}\| + \|U_{2T}\| + \|U_{3T}\| + \|U_{4T}\|\Big),$$

with

$$\begin{aligned}
U_{1T} &:= (T-k)^{-1}\sum_{t=k}^{T-1}\Big(\sum_{u=k}^{\infty}\gamma(\Delta u) y_{\Delta,t-u}\Big) Y'_{\Delta,t,k}\\
U_{2T} &:= (T-k)^{-1}\sum_{t=k}^{T-1}\Big(\frac{y_{\Delta,t+1} - \mathbb{E}[y_{\Delta,t+1}|\mathcal{F}^{\Delta,t}]}{\Delta}\Big) Y'_{\Delta,t,k}\\
U_{3T} &:= (T-k)^{-1}\sum_{t=k}^{T-1}\omega\, Y'_{\Delta,t,k} \text{ where } \omega = \nu - \lambda\\
U_{4T} &:= (T-k)^{-1}\sum_{t=k}^{T-1}\Big(\int_0^{\infty}\gamma(u)dN(t\Delta - u) - \sum_{j=0}^{\infty}\gamma(\Delta j)y_{\Delta,t-j}\Big) Y'_{\Delta,t,k},
\end{aligned}$$

where N is here meant to be the centered point process.

First we find

$$\Delta\|\widehat{\Gamma}_{k,\Delta}^{-1}\|_1 \leq \Delta\|\Gamma_{k,\Delta}^{-1}\|_1 + \Delta\|\widehat{\Gamma}_{k,\Delta}^{-1} - \Gamma_{k,\Delta}^{-1}\|_1.$$

We now show that the first summand in the latter inequality is an $O(1/\Delta)$. In order to prove that fact we realize that $\Gamma_{k,\Delta}$ is a block Toeplitz matrix for all fixed Δ. Upper bounding the spectral radius of $\Delta\Gamma_{k,\Delta}^{-1}$ for all k,Δ is equivalent to bounding the smallest eigenvalue of $1/\Delta\Gamma_{k,\Delta}$ away from zero for all k,Δ. It is well known that this can be translated to the world of the respective spectral density generating the block Toeplitz matrix. To recapitulate this denote for any fixed Δ the spectral density of y_Δ by $f_{y,\Delta}$. Let ρ be the function that maps a matrix onto its smallest eigenvalue. If

$$\inf_{\omega}\rho\big(f_{y,\Delta}(\omega)\big) > C > 0,$$

then the smallest eigenvalue of $1/\Delta\Gamma_{k,\Delta}$ is bounded away from zero for all k. This yields for our situation that we need to show

$$\inf_{\Delta}\inf_{\omega}\rho\big(f_{y,\Delta}(\omega)\big) > C > 0$$

in order to obtain $\Delta\|\Gamma_{k,\Delta}^{-1}\|_1 < \kappa$ for all k,Δ for some constant $\kappa > 0$. Since the point process is stationary we have with the notation of Hawkes

$$\operatorname{cov}\big(dN(s), dN(t)\big) = \mu(s-t)\,dsdt + D\,\delta_{\{s=t\}}ds := \mu^{(c)}(s-t)\,dsdt$$

and for the spectral density matrix

$$f(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega\tau}\mu^{(c)}(\tau)\,d\tau = M(\omega) + \frac{1}{2\pi}D.$$

Under suitable regularity assumptions we have the inverse Fourier-transform

$$\mu(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau}\, M(\omega)\, d\omega.$$

Hawkes proves

$$f(\omega) = \frac{1}{2\pi}\{I - G(\omega)\}^{-1} D\,\{I - G'(-\omega)\}^{-1}.$$

Since $y_{\Delta,t}$ is stationary we have

$$\begin{aligned} c_\Delta(j) := \operatorname{cov}\big(y_{\Delta,j}\,,y_{\Delta,0}\big) &= \int_0^\Delta \int_{j\Delta}^{(j+1)\Delta} \operatorname{cov}\big(dN(s), dN'(t)\big)\, ds\, dt \\ &= \int_0^\Delta \int_{j\Delta}^{(j+1)\Delta} \int_{-\infty}^{\infty} e^{i\omega(s-t)}\, M(\omega)\, d\omega\, ds\, dt + \Delta\, D\, \delta_j. \end{aligned}$$

Since $\int_0^\Delta e^{-i\omega t} dt = \Delta + O(\Delta^2)$ we obtain

$$\begin{aligned} c_\Delta(j) &= \Delta^2 \int_{-\infty}^{\infty} e^{i\omega j\Delta}\, M(\omega)\, d\omega + \Delta\, D\, \delta_j \\ &= \Delta \int_{-\infty}^{\infty} e^{i\lambda j}\, M\Big(\frac{\lambda}{\Delta}\Big)\, d\lambda + \Delta\, D\, \delta_j \\ &= \Delta \int_{-\pi}^{\pi} e^{i\lambda j} \sum_{\ell=-\infty}^{\infty} M\Big(\frac{\lambda + 2\pi\ell}{\Delta}\Big)\, d\lambda + \Delta\, D\, \delta_j. \end{aligned}$$

This implies that the spectral density matrix of the process $y_{\Delta,t}$ is

$$f_{y,\Delta}(\lambda) = \Delta \sum_{\ell=-\infty}^{\infty} M\Big(\frac{\lambda + 2\pi\ell}{\Delta}\Big) + \frac{1}{2\pi}\, \Delta D$$

Since $M(\lambda)$ is positive definite for all λ we have that the minimal eigenvalue of $f_{y,\Delta}(\lambda)$ is larger than the minimal eigenvalue of $\frac{1}{2\pi}\,\Delta D$ which is $\frac{1}{2\pi}\,\Delta \min_{i\in\{1,\ldots,d\}} \lambda_i > 0$. Hence, it holds that $\Delta\|\Gamma_{k,\Delta}^{-1}\|_1 < \kappa$ for all k and all Δ for some $\kappa > 0$.

Remark 3.10. *Under suitable regularity conditions (e.g. exponential decay of M) we have*

$$\lim_{\Delta\to 0} \frac{1}{\Delta} f_{y,\Delta}(\lambda) = M(0)\, \delta_{\{\lambda=0\}} + \frac{1}{2\pi} D.$$

This means that the above lower bound is 'optimal' for $\lambda \neq 0$ and 'rate optimal' for $\lambda = 0$.

We claim, that the second summand of the inequality, $\Delta\|\widehat{\Gamma}_{k,\Delta}^{-1} - \Gamma_{k,\Delta}^{-1}\|_1$, converges to zero in probability as $T \to \infty$. For that purpose note that $\|\widehat{\Gamma}_{k,\Delta} - \Gamma_{k,\Delta}\|_1 \xrightarrow{p} 0$ since

$$\begin{aligned} \mathbb{E}\big(\|\widehat{\Gamma}_{k,\Delta} - \Gamma_{k,\Delta}\|_1^2\big) &\le \mathbb{E}\big(\|\widehat{\Gamma}_{k,\Delta} - \Gamma_{k,\Delta}\|^2\big) \\ &= \mathbb{E}\Big[tr\Big(\big(\widehat{\Gamma}_{k,\Delta} - \Gamma_{k,\Delta}\big)'\big(\widehat{\Gamma}_{k,\Delta} - \Gamma_{k,\Delta}\big)\Big)\Big] \\ &\le C_2 \frac{k^2}{T-k} \mathbb{E}\big(\big(y_{\Delta,1}' y_{\Delta,1}\big)^2\big), \end{aligned}$$

which converges to zero as $T \to \infty$ according to assumption 1. Do note that $\mathbb{E}\big((y'_{\Delta,1}y_{\Delta,1})^2\big) = o(1)$ as $\Delta \to 0$. Therefore we find

$$\begin{aligned}
\Delta\|\widehat{\Gamma}_{k,\Delta}^{-1} - \Gamma_{k,\Delta}^{-1}\|_1 &= \Delta\,\|\widehat{\Gamma}_{k,\Delta}^{-1}\,(\widehat{\Gamma}_{k,\Delta} - \Gamma_{k,\Delta})\,\Gamma_{k,\Delta}^{-1}\|_1 \\
&\le \Delta\,\|\widehat{\Gamma}_{k,\Delta}^{-1}\|_1\,\|\widehat{\Gamma}_{k,\Delta} - \Gamma_{k,\Delta}\|_1\,\|\Gamma_{k,\Delta}^{-1}\|_1 \\
&\le \Delta\,\big(\|\widehat{\Gamma}_{k,\Delta}^{-1} - \Gamma_{k,\Delta}^{-1}\|_1 + \|\Gamma_{k,\Delta}^{-1}\|_1\big)\,\|\Gamma_{k,\Delta}^{-1}\|_1\,\|\widehat{\Gamma}_{k,\Delta} - \Gamma_{k,\Delta}\|_1 \\
&\le \big(\Delta\,\|\widehat{\Gamma}_{k,\Delta}^{-1} - \Gamma_{k,\Delta}^{-1}\|_1 + \kappa\big)\,\kappa\,\|\widehat{\Gamma}_{k,\Delta} - \Gamma_{k,\Delta}\|_1/\Delta.
\end{aligned}$$

Now we have

$$\begin{aligned}
0 \le Z_{k,\Delta,T} &= \Delta\|\widehat{\Gamma}_{k,\Delta}^{-1} - \Gamma_{k,\Delta}^{-1}\|_1/\big(\Delta\|\widehat{\Gamma}_{k,\Delta}^{-1} - \Gamma_{k,\Delta}^{-1}\|_1 + \kappa\big)\kappa \\
&\le \|\widehat{\Gamma}_{k,\Delta} - \Gamma_{k,\Delta}\|_1/\Delta \xrightarrow{p} 0 \quad \text{as } T \to \infty,
\end{aligned}$$

due to assumption 4.55. Additionally, we obtain

$$\begin{aligned}
\Delta\|\widehat{\Gamma}_{k,\Delta}^{-1} - \Gamma_{k,\Delta}^{-1}\|_1 &= \frac{\kappa^2\,\Delta\,\|\widehat{\Gamma}_{k,\Delta}^{-1} - \Gamma_{k,\Delta}^{-1}\|_1}{\big(\Delta\,\|\widehat{\Gamma}_{k,\Delta}^{-1} - \Gamma_{k,\Delta}^{-1}\|_1 + \kappa\big)\,\kappa - \kappa\,\Delta\,\|\widehat{\Gamma}_{k,\Delta}^{-1} - \Gamma_{k,\Delta}^{-1}\|_1} \\
&= \frac{\kappa^2\,Z_{k,\Delta,T}}{1 - \kappa\,Z_{k,\Delta,T}}.
\end{aligned}$$

Thus, this yields that also $\Delta\|\widehat{\Gamma}_{k,\Delta}^{-1} - \Gamma_{k,\Delta}^{-1}\|_1 \xrightarrow{p} 0$ as $T \to \infty$.

We proceed in showing that each U_{iT}, $1 \le i \le 4$, converges to zero in probability. For this purpose it suffices to show that the expected values of the corresponding norms vanish (Markov inequality). Examining U_{1T} we find

$$\begin{aligned}
\mathbb{E}(\|U_{1T}\|) &\le (T-k)^{-1}\sum_{t=k}^{T-1}\mathbb{E}\Big(\Big\|\sum_{u=k}^{\infty}\gamma(\Delta u)y_{\Delta,t-u}Y'_{\Delta,t,k}\Big\|\Big) \\
&\le \Big\{\mathbb{E}\Big(\Big\|\sum_{u=k}^{\infty}\gamma(\Delta u)y_{\Delta,t-u}\Big\|^2\Big)\Big\}^{1/2}\{\mathbb{E}(\|Y_{\Delta,t,k}\|^2)\}^{1/2} \\
&\le \{k \cdot tr(\Gamma_\Delta(0))\}^{1/2}\Big\{\sum_{i=k}^{\infty}\sum_{j=k}^{\infty}\|\Gamma_\Delta(i-j)\|\|\ \gamma(\Delta i)\|\|\ \gamma(\Delta j)\|\Big\}^{1/2} \\
&\le Ck^{1/2}\sum_{j=k}^{\infty}\|\gamma(\Delta j)\|,
\end{aligned}$$

which converges to zero in probability according to assumption 3. For the second summand we obtain

$$
\begin{aligned}
\mathbb{E}\Big(\|U_{2T}\|^2\Big) &= (T-k)^{-2}\mathbb{E}\Big(\Big\|\sum_{t=k}^{T-1}\Big(\frac{y_{\Delta,t+1}-\mathbb{E}[y_{\Delta,t+1}|\mathcal{F}^{\Delta,t}]}{\Delta}\Big)Y'_{\Delta,t,k}\Big\|^2\Big)\\
&= (T-k)^{-2}\frac{1}{\Delta^2}\sum_{t=k}^{T-1}\sum_{s=k}^{T-1}\mathbb{E}\Big[\Big(y_{\Delta,t+1}-\mathbb{E}[y_{\Delta,t+1}|\mathcal{F}^{\Delta,t}]\Big)'\\
&\qquad\times\Big(y_{\Delta,s+1}-\mathbb{E}[y_{\Delta,s+1}|\mathcal{F}^{\Delta,s}]\Big)Y'_{\Delta,t,k}Y_{\Delta,s,k}\Big]\\
&= (T-k)^{-2}\frac{1}{\Delta^2}\sum_{t=s}\mathbb{E}\Big[\Big(y_{\Delta,t+1}-\mathbb{E}[y_{\Delta,t+1}|\mathcal{F}^{\Delta,t}]\Big)'\\
&\qquad\times\Big(y_{\Delta,s+1}-\mathbb{E}[y_{\Delta,s+1}|\mathcal{F}^{\Delta,s}]\Big)Y'_{\Delta,t,k}Y_{\Delta,s,k}\Big]\\
&\le C(T-k)^{-1}\frac{k}{\Delta^2}.
\end{aligned}
$$

Therefore, U_{2T} converges to zero in probability according to assumption 1.

For U_{3T} observe

$$
\begin{aligned}
\mathbb{E}\Big(\Big\|U_{3T}\Big\|^2\Big) &= \mathbb{E}\Bigg(tr\Bigg(\Big[(T-k)^{-1}\sum_{t=k}^{T-1}\omega Y'_{\Delta,t,k}\Big]'\Big[(T-k)^{-1}\sum_{t=k}^{T-1}\omega Y'_{t,k}\Big]\Bigg)\Bigg)\\
&= (T-k)^{-2}\mathbb{E}\Bigg(tr\Bigg(\Big[\sum_{t=k}^{T-1}\omega Y'_{\Delta,t,k}\Big]'\Big[\sum_{t=k}^{T-1}\omega Y'_{\Delta,t,k}\Big]\Bigg)\Bigg)\\
&= \omega'\omega(T-k)^{-2}\sum_{t=k}^{T-1}\sum_{s=k}^{T-1}\mathbb{E}\Bigg(Y'_{\Delta,t,k}Y_{\Delta,s,k}\Bigg)\\
&= k\cdot\omega'\omega\cdot(T-k)^{-2}\sum_{t=k}^{T-1}\sum_{s=k}^{T-1}tr\Big(\Gamma_\Delta(t-s)\Big)\\
&\le k\cdot\omega'\omega\cdot(T-k)^{-2}\sum_{t=k}^{T-1}\sum_{s=-\infty}^{\infty}tr\Big(\Gamma_\Delta(t-s)\Big)\\
&\le C\cdot k\cdot(T-k)^{-1}
\end{aligned}
$$

which again converges to zero in probability by assumption 1. For the last inequality we notice, that for all $\Delta < 1$, $\sum_{s=-\infty}^{\infty} tr\big(\Gamma_\Delta(t-s)\big)$ is uniformly bounded by $\sum_{s=-\infty}^{\infty} tr\big(\Gamma_1(t-s)\big)$.

Finally we find for U_{4T}

$$
\begin{aligned}
&\Big(\mathbb{E}\Big(\|U_{4T}\|^2\Big)\Big)^2 \le C_1\, k\, \Delta\, \mathbb{E}\Big\| \sum_{j=0}^{\infty} \int_{\Delta j}^{\Delta(j+1)} (\gamma(u) - \gamma(\Delta j))\, dN(t\,\Delta - u) \Big\|^2 \\
&= C_1\, k\, \Delta \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \int_{\Delta j}^{\Delta(j+1)} \int_{\Delta k}^{\Delta(k+1)} tr\big[(\gamma(u) - \gamma(\Delta j))'(\gamma(v) - \gamma(\Delta k))\big] \\
&\qquad\qquad \times \mathbb{E}\big[dN(t\,\Delta - v)'\, dN(t\,\Delta - u)\big] \\
&= C_1\, k\, \Delta \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \int_{\Delta j}^{\Delta(j+1)} \int_{\Delta k}^{\Delta(k+1)} tr\big[(\gamma(u) - \gamma(\Delta j))'(\gamma(v) - \gamma(\Delta k))\big] \\
&\qquad\qquad \times tr\big[\mu^{(c)}(u-v)\big]\, d(t\,\Delta - v)\, d(t\,\Delta - u) \\
&= C_1\, k\, \Delta \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \int_{0}^{\Delta} \int_{0}^{\Delta} tr\big[(\gamma(u + \Delta\, j) - \gamma(\Delta j))'(\gamma(v + \Delta\, k) - \gamma(\Delta k))\big] \\
&\qquad\qquad \times tr\big[\mu^{(c)}(u - v + \Delta(j-k))\big]\, d(t\,\Delta - v - k\,\Delta)\, d(t\,\Delta - u - j\,\Delta) \\
&\le C_2\, k\, \Delta^3,
\end{aligned}
$$

which converges to zero in probability by assumption 4.

Putting all factors and summands as well as the inequalities together it falls out nicely that

$$
\|\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k\| \xrightarrow{p} 0 \quad \text{as } T \to \infty.
$$

□

After having proved our main Theorem we discuss the made assumptions. Assumption 1, $T^{-1}\frac{k^2}{\Delta} \to 0$ as $T \to \infty$, tells us that the number of observed slots T has to be much larger than the width of the slots and the finite regression order k, i.e. $T >> 1/\Delta$ and $T >> k$. We may also reformulate assumption 1 by $\frac{k^2}{\widetilde{T}} \to 0$ as $\widetilde{T} \to \infty$, where $\widetilde{T} = \Delta\, T$ is the observation time. Furthermore $k\Delta \to \infty$ assures that the infinite integrals and sums converge. The second assumption $\| \int_0^{\infty} \gamma(u) du\| < 1$ is a standard stability condition for any infinite order regression to make the process stationary and to establish that the autocovariances are absolutely summable with respect to the corresponding norm. Assumption 3 is some kind of lower bound for the convergence rate of the norm of γ to zero. It tells us that the link functions have to decrease quickly.

The next theorem generalizes Theorem **3.9** to a functional convergence. Therefor we set $\widetilde{\gamma}^T$ to be the step function defined by

$$
\widetilde{\gamma}^T(s) = \sum_{u=1}^{k} I_{[\Delta(u-1), \Delta u)}(s) \widehat{\gamma}^k(\Delta(u-1)).
$$

Theorem 3.11. *Under the assumptions of Theorem **3.9** it holds*

$$
\int_{\mathbb{R}} \|\gamma(s) - \widetilde{\gamma}^T(s)\| ds \xrightarrow{p} 0 \qquad \text{as} \quad T \to \infty.
$$

PROOF. Decomposing the integral into the approximation error and the estimation error we find

$$\int_{\mathbb{R}} \|\gamma(s) - \widetilde{\gamma}^T(s)\| ds = \int_0^{\Delta k} \|\gamma(s) - \widetilde{\gamma}^T(s)\| ds + \int_{\Delta k}^{\infty} \|\gamma(s) - \widetilde{\gamma}^T(s)\| ds$$
$$= \sum_{u=1}^{k} \int_{\Delta(u-1)}^{\Delta u} \|\gamma(s) - \widetilde{\gamma}^T(s)\| ds + \int_{\Delta k}^{\infty} \|\gamma(s)\| ds$$
$$= \sum_{u=1}^{k} \int_{\Delta(u-1)}^{\Delta u} \|\gamma(s) - \gamma(\Delta(u-1))\| ds + \Delta \|\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k\| + \int_{\Delta k}^{\infty} \|\gamma(s)\| ds,$$

where the first and third summand converge to zero according to integrability conditions and the second summand converges to zero in probability by Theorem **3.9**.

□

Figure **1** illustrates our estimation procedure based on simulated data. To simulate the three dimensional Hawkes process we used a Lewis simulation method. The true link functions, given by the dashed lines, were taken to have the form $\gamma_{ij}(s) = \alpha_{ij} \exp(-\beta_{ij}\, s)$ for $1 \leq i, j \leq 3$ with different coefficients. In total we recorded approximately five thousand events. We chose $\Delta = 0.05$ and $k = 20$ as our discretization parameters. The solid lines are the outcome of our estimation technique.

Notice that we have estimated the link functions separately from the constant vector ν. This is intuitively possible due to the fact, that the least squares estimator estimates the function or parameters that concern the covariance structure of the process. However, ν displays the Poisson amount of the component processes. Hence, this part is "independent" of the covariance structure. Having estimated the link functions we may also estimate ν in the stationary setting via the well known relation (see (3.3))

$$\nu = (I_d - G)\lambda,$$

where $G := \int_{-\infty}^{\infty} \gamma(t) dt$. This is a nice side result, however, since ν is a component that has no causal influence, it is not of importance for establishing causality graphs.

Corollary 3.12. *Define $\widehat{\nu} = (I_d - \widehat{G}^T)\,\widehat{\lambda}$, where $\widehat{G}^T := \int_0^{k\Delta} \widetilde{\gamma}^T(s) ds$ and $\widehat{\lambda} = T^{-1} \sum_{t=1}^{T} y_{1,t}$. Then with the assumptions of Theorem **3.9** it holds*

$$\widehat{\nu} \xrightarrow{p} \nu \quad \text{as } T \to \infty.$$

PROOF. Follows directly by Theorem **3.9** and Slutsky's theorem.

□

We now turn to the derivation of asymptotic normality of our estimator. In the following $vec(\cdot)$ denotes the vec-operator and $A \otimes B$ the Kronecker product of matrices A and B with suitable dimensions.

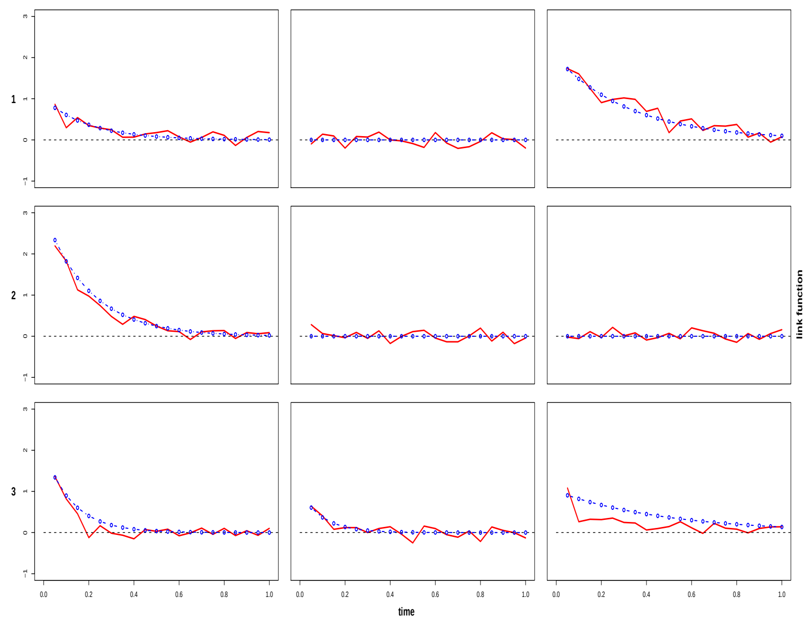

Figure 1: Estimation of the Hawkes link functions based on simulated data. The solid line displays the estimated values and the dashed lines the real link functions.

Theorem 3.13. *Make the following assumptions:*

1. *Let k and Δ be functions of T such that*
$$k\Delta \to \infty \qquad \text{and} \qquad T^{-1}\frac{k^3}{\Delta} \to 0$$
as $T \to \infty$.

2. *We demand the stability condition $\left\| \int_0^\infty \gamma(u)du \right\|_1 < 1$.*

3. *The component functions of γ decrease in a way such that*
$$T^{1/2} \sum_{j=k}^{\infty} \|\gamma(\Delta j)\| \to 0$$
as $T \to \infty$.

4. $T^{1/2}\left(\mathbb{E}\left\| \int_0^\infty \gamma(u)dN(t\Delta - u) - \sum_{j=0}^\infty \gamma(\Delta j) y_{\Delta,t-j} \right\|^2\right)^{1/2} \to 0$ *as* $T \to \infty$.

5. *A sequence of $(kd^2 \times 1)$ vectors, $\{l(k)\}$ is given satisfying*
$$0 < C_1 \leq \|l(k)\|^2 \leq C_2 < \infty \quad \text{for} \quad k = 1, 2, \ldots$$

Then

$$
\begin{aligned}
&(T-k)^{1/2}\Delta l(k)' vec(\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k) \\
&\quad - (T-k)^{1/2}\Delta l(k)' vec\Big(\Big\{(T-k)^{-1} \\
&\qquad \times \sum_{t=k}^{T-1}\Big(\frac{y_{\Delta,t+1} - \mathbb{E}[y_{\Delta,t+1}|\mathcal{F}^{\Delta,t}]}{\Delta} + \omega\Big) Y'_{\Delta,t,k}\Big\}\Gamma_{k,\Delta}^{-1}\Big) \xrightarrow{p} 0
\end{aligned}
$$

as $T \to \infty$.

Proof.

Observe the following computation:

$$
\begin{aligned}
&(T-k)^{1/2}\Delta l(k)' vec(\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k) \\
&\quad - (T-k)^{1/2}\Delta l(k)' vec\Big(\Big\{(T-k)^{-1} \\
&\qquad \sum_{t=k}^{T-1}\Big(\frac{y_{\Delta,t+1} - \mathbb{E}[y_{\Delta,t+1}|\mathcal{F}^{\Delta,t}]}{\Delta} + \omega\Big) Y'_{\Delta,t,k}\Big\}\Gamma_{k,\Delta}^{-1}\Big) \\
&= (T-k)^{1/2}\Delta l(k)'\Big\{vec\Big[\sum_{i=1}^{4} U_{iT}\widehat{\Gamma}_{k,\Delta}^{-1}\Big] - vec\Big[U_{2T}\Gamma_{k,\Delta}^{-1}\Big] - vec\Big[U_{3T}\Gamma_{k,\Delta}^{-1}\Big]\Big\} \\
&= (T-k)^{1/2}\Delta l(k)'\Big\{\sum_{i=1}^{4} I_d \otimes (\widehat{\Gamma}_{k,\Delta}^{-1} - \Gamma_{k,\Delta}^{-1}) vec[U_{iT}] + \sum_{i=1,4} I_d \otimes \Gamma_{k,\Delta}^{-1} vec[U_{iT}]\Big\} \\
&= \sum_{i=1}^{4} w_{iT} + \sum_{i=1,4} v_{iT},
\end{aligned}
$$

where w_{iT} and v_{iT} are defined in the obvious way. First we notice for $i = 1, \dots, 4$

$$
|w_{iT}| \leq \|l(k)\| k^{1/2}\Delta \|\widehat{\Gamma}_{k,\Delta}^{-1} - \Gamma_{k,\Delta}^{-1}\|_1 \|k^{-1/2}(T-k)^{1/2} U_{iT}\|.
$$

Following the arguments of the proof of Theorem **3.9** it holds that $k^{1/2}\Delta\|\widehat{\Gamma}_{k,\Delta}^{-1} - \Gamma_{k,\Delta}^{-1}\|_1 \xrightarrow{p} 0$. Hence, again using the same methods as in the proof of Theorem **3.9** we obtain

$$
\begin{aligned}
|w_{1T}| &\le \|l(k)\|\, k^{1/2}\, \Delta\, \|\widehat{\Gamma}_{k,\Delta}^{-1} - \Gamma_{k,\Delta}^{-1}\|_1\, C\, (T-k)^{1/2} \sum_{j=k}^{\infty} \|\gamma(\Delta j)\|, \\
|w_{2T}| &\le \|l(k)\|\, k^{1/2}\, \Delta\, \|\widehat{\Gamma}_{k,\Delta}^{-1} - \Gamma_{k,\Delta}^{-1}\|_1\, \frac{1}{\Delta} \sum_{s=k}^{\infty} \mathbb{E}\Big((y'_{\Delta,t+1}\, y_{\Delta,s+1})^2 \Big), \\
|w_{3T}| &\le \|l(k)\|\, k^{1/2}\, \Delta \|\widehat{\Gamma}_{k,\Delta}^{-1} - \Gamma_{k,\Delta}^{-1}\|_1\, C, \\
|w_{4T}| &\le \|l(k)\|\, k^{1/2}\ \Delta\, \|\widehat{\Gamma}_{k,\Delta}^{-1} - \Gamma_{k,\Delta}^{-1}\|_1\, C\, (T-k)^{1/2} \\
&\qquad \times \Big[\mathbb{E}\Big(\Big\| \int_0^{\infty} \gamma(u)\, dN(t\Delta - u) - \sum_{j=0}^{\infty} \gamma(\Delta j)\, y_{\Delta,t-j} \Big\|^2 \Big) \Big]^{1/2}.
\end{aligned}
$$

Therefore, $|w_{iT}| \xrightarrow{p} 0$ for all $1, \ldots, 4$ ($|w_{1T}|$, $|w_{2T}|$ by assumption 3 and $|w_{4T}|$ by assumption 4).

Furthermore, for the remaining summands we find

$$
\begin{aligned}
|v_{1T}| &\le (T-k)^{-1/2}\, d^{1/2}\, \Delta\, \Big\| \sum_{t=k}^{T-1} \Big(\sum_{u=k}^{\infty} \gamma(\Delta u) y_{\Delta,t-u} \Big) l(k)' (\Gamma_{k,\Delta}^{-1} Y_{\Delta,t,k}) \otimes I_d \Big\| \\
&\le C\, (T-k)^{1/2}\, M_2\, \kappa \sum_{j=k}^{\infty} \|\gamma(\Delta j)\| \xrightarrow{p} 0 \quad \text{as} \quad T \to \infty,
\end{aligned}
$$

due to assumption 3, where κ is the uniform bound of $\|\Delta \Gamma_{k,\Delta}^{-1}\|_1$ and analogously

$$
\begin{aligned}
|v_{4T}| &\le (T-k)^{-1/2}\, d^{1/2}\, \Delta \\
&\qquad \times \Big\| \sum_{t=k}^{T-1} \Big(\int_0^{\infty} \gamma(u) dN(t\Delta - u) - \sum_{j=0}^{\infty} \gamma(\Delta j) y_{\Delta,t-j} \Big) l(k)' (\Gamma_{k,\Delta}^{-1} Y_{\Delta,t,k}) \otimes I_d \Big\| \\
&\le C (T-k)^{1/2} M_2 \kappa \Big(\mathbb{E} \Big\| \int_0^{\infty} \gamma(u) dN(t\Delta - u) - \sum_{j=0}^{\infty} \gamma(\Delta j) y_{\Delta,t-j} \Big\|^2 \Big)^{1/2} \xrightarrow{p} 0,
\end{aligned}
$$

as $T \to \infty$ due to assumption 4. Thus, the proof is complete. □

The next theorem establishes asymptotic normality of the modified score function of the last theorem.

Theorem 3.14. *Define*

$$
s_T = (T-k)^{1/2} \Delta l(k)' vec\Big(\Big\{ (T-k)^{-1} \sum_{t=k}^{T-1} \Big(\frac{y_{\Delta,t+1} - \mathbb{E}[y_{\Delta,t+1} | \mathcal{F}^{\Delta,t}]}{\Delta} + \omega \Big) Y'_{\Delta,t,k} \Big\} \Gamma_{k,\Delta}^{-1} \Big)
$$

and

$$
v_T^2 = Var(s_T).
$$

Then under the assumptions of Theorem ***3.13*** *it holds*

$$\frac{s_T}{\upsilon_T} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad \textit{as} \quad T \to \infty.$$

Proof. First we notice that

$$\begin{aligned} Var(s_T) =&(T-k)^{-1}\,\Delta^2 \sum_{t=k}^{T-1}\sum_{s=k}^{T-1} l(k)'\,(I_d \otimes \Gamma_{k,\Delta}^{-1}) \\ &\times \mathbb{E}\Big[vec\Big(\Big(\frac{y_{\Delta,t+1} - \mathbb{E}[y_{\Delta,t+1}|\mathcal{F}^{\Delta,t}]}{\Delta} + \omega\Big)\,Y'_{\Delta,t,k}\Big) \\ &\times vec\Big(\Big(\frac{y_{\Delta,s+1} - \mathbb{E}[y_{\Delta,s+1}\,|\,\mathcal{F}^{\Delta,s}]}{\Delta} + \omega\Big)\,Y'_{\Delta,s,k}\Big)'\Big](I_d \otimes \Gamma_{k,\Delta}^{-1})l(k). \end{aligned}$$

For explicit expressions for the covariance structure see Bacry, et al. (2012). Define for $k+1 \le t \le T$

$$X_t(T) = (T-k)^{-1/2}\,\Delta\,l(k)'\,\Big(\Gamma_{k,\Delta}^{-1}Y'_{\Delta,t-1,k} \otimes I_d\Big)\,\Big(\frac{y_{\Delta,t} - \mathbb{E}[y_{\Delta,t}|\mathcal{F}^{\Delta,t-1}]}{\Delta} + \omega\Big)/\upsilon_T$$

and set $X_t(T) = 0$ for $0 \le t \le k$. Then we can rewrite $s_T/\upsilon_T = \sum_{t=1}^{T} X_t(T)$. By the structure of the latter term (e.g. by Doob-Meyer decomposition) we immediately find that $(S_n(T) = \sum_{t=1}^{n} X_t(T), 0 \le n \le T$ a.e. is a martingale sequence for every fixed $T \ge 1$. Thus, we have to prove asymptotic normality for a triangular array. Theorem 2 of Scot (1973) gives sufficient conditions for the convergence of the triangular array in the sense $s_t/\upsilon_T \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$. Here we want to derive conditions (C) of Theorem 2 in Scot (1973), namely

$$\begin{aligned} &\text{(a)} \quad \sup_{t\le T} X_t^2(T) \xrightarrow{p} 0 \,\text{as}\, T \to \infty, \\ &\text{(b)} \quad \sum_{t=1}^{n_T(\tau)} X_t^2(T) \xrightarrow{p} \tau,\ 0 \le \tau \le 1,\ \text{as}\, T \to \infty, \end{aligned}$$

where $n_T(\tau) = \max_{n\le T}\{n : \mathbb{E}[(\sum_{t=1}^{n} X_t(T))^2] \le \tau\}$. For (a) notice that for any $\delta > 0$

$$\begin{aligned} &\mathbb{P}\Big(\sup_{t\le T} X_t^2(T) \ge \delta\Big) \le \sum_{t=k+1}^{T} \mathbb{P}(X_t^2(T) \ge \delta) \le \delta^{-2}(T-k)\mathbb{E}(X_t^4(T)) \\ &\le (T-k)^{-1}\,\|l(k)\|^4\,\big(\Delta\,\|\Gamma_{k,\Delta}^{-1}\|\big)^4 \upsilon_t^{-4} \times \\ &\qquad \Big\{\mathbb{E}\Big(\Big\|\frac{y_{\Delta,t+1} - \mathbb{E}[y_{\Delta,t+1}|\mathcal{F}^{\Delta,t}]}{\Delta} + \omega\Big\|^8\Big)\Big\}^{1/2}\,\Big\{\|Y_{\Delta,t-1,k}\|^8\Big\}^{1/2} \\ &\le C\,\Delta^{-1}\,(T-k)^{-1}\,k^2\,\mathbb{E}\Big(\big(y'_{\Delta,t}\,y_{\Delta,t}\big)^4\Big), \end{aligned}$$

which converges to zero as $T \to \infty$ due to Assumption 1. Condition (b) can be verified in analogous fashion to the proof of Theorem 3 in Lewis (1985) by Assumption 1.

$\square$

Finally asymptotic normality of our estimator is an immediate consequence of the latter theorem.

Corollary 3.15. *With the assumptions of Theorem* ***3.14*** *it holds*

$$(T-k)^{1/2}\Delta l(k)' vec\big(\widehat{\gamma}_k - \gamma_k\big) \,/\, v_T \to \mathcal{N}(0,1) \quad as \quad T \to \infty.$$

In future work it would be desirable to develop a test based on the latter asymptotic behavior to decide whether an estimated link function is believed to be identically equal to zero or not.

3.3 Application

As an application we analyzed spike train data from the lumbar spinal dorsal horn of a pentobarbital-anaesthetised rat during noxious stimulation. The firing times of ten neurons were recorded simultaneously by a single electrode with an observation time of 100s. The data have been measured and analyzed by Sandkühler and Eblen-Zajjur (1994) who studied discharge patterns of spinal dorsal horn neurons under various conditions.

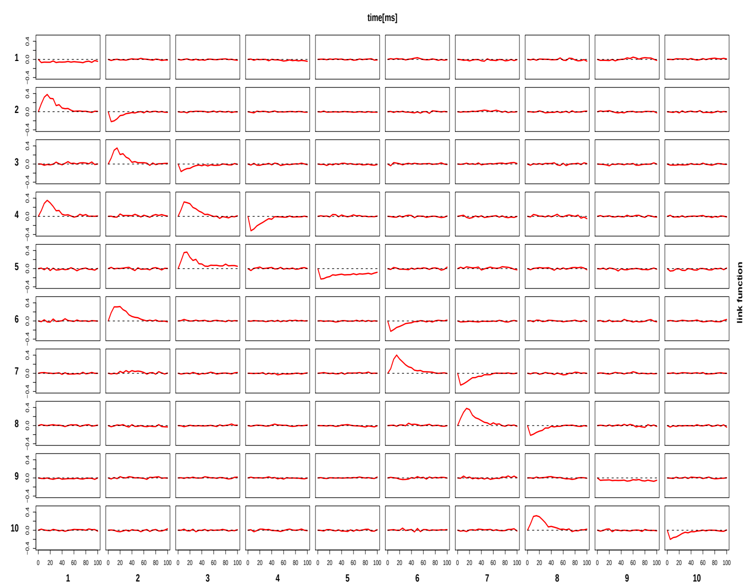

Figure 2: Estimated link functions of Hawkes model based on spike train data.

As an underlying model it is reasonable to choose Hawkes model, since it reflects the mutually exciting structure of neurons dependent on time. Hence, we estimated the Hawkes kernels with the presented estimation procedure (Figure **2**). As discretization parameters we set $\Delta = 0.5$ and $k = 20$. We can clearly see that most of the 100 link functions vanish and only few show significant peaks. To statistically significantly decide whether the link functions vanish or not a test has to be constructed in future work. As we have seen a causality graph is induced by the estimated link functions, which is displayed in Figure **3**. It is remarkable, that neuron nine is completely isolated from the other neurons.

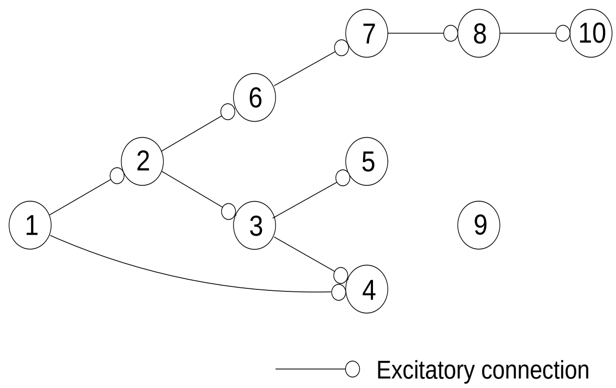

Figure 3: Causality Graph induced by estimated link functions of Hawkes model based on spike train data.

Furthermore Figure **2** shows, that if there is a peak, then shape, time and intensity behave very similarly. The time instants of the peaks happen roughly around 17 ms with an intensity of approximately 0.38.

Hawkes model has two drawbacks in this application. First it is only capable of modeling excitement, not inhibition, which is well known to play a major role in neuronal data. Secondly neurons possess a refractory period. Thus, a neuron inhibits itself after having fired. As a consequence negative correlations appear and Hawkes model fails to catch this phenomenon. Hence, we observe in Figure **2** that link functions on the diagonal become negative in contradiction to Hawkes model. On the other hand incorporating refractory period into the model would destroy its linear structure.

However, in many different applications such as modeling aftershock effects (Ogata, 1999; Vere-Jones, 1970 and Vere-Jones and Ozaki, 1982), insurgency in Iraq (Brantingham, et al. 2011), crime (Mohler et al. 2011) or genome analysis (Reynaud-Bouret and Schbath 2010), the Hawkes model does not encounter these problems.

4 Distance Correlation

In this section we introduce the new dependence measure of distance correlation. We state its most important properties and explain why it makes sense in certain situations to make the transition from usual distance correlation to affinely invariant distance correlation. Therefore, we define affinely invariant distance correlation and show that in the Gaussian case it is a function of the canonical correlation coefficients. Moreover, we derive explicit formulas for distance correlation for Lancaster distributions.

4.1 Definition and Properties

In this section we follow the revolutionizing papers by Székely, Rizzo and Bakirov (2007) and Székely and Rizzo (2009). We state the definition of distance correlation and recapitulate their most important results. Distance correlation is a new measure for classical stochastic independence between random variables of arbitrary dimensions with finite first moments in the usual sense, that the characteristic function of the joint distribution splits up into the product of the marginal characteristic functions. Specifically, let p and q be positive integers that will represent the dimensions of the corresponding random variables. For jointly distributed random vectors $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$, let

$$f_{X,Y}(s,t) = \mathbb{E}\exp\big[i\,\langle s, X\rangle_p + i\,\langle t, Y\rangle_q\big]$$

be the joint characteristic function of (X, Y), and let $f_X(s) = f_{X,Y}(s, 0)$ and $f_Y(t) = f_{X,Y}(0, t)$ be the marginal characteristic functions of X and Y, where $s \in \mathbb{R}^p$ and $t \in \mathbb{R}^q$. Then an obvious natural choice for a dependence measure would be of the form

$$\mathcal{V}^2(X, Y, \omega) = \int_{\mathbb{R}^{p+q}} |f_{X,Y}(s,t) - f_X(s)f_Y(t)|^2\,\omega(s,t)\,\mathrm{d}s\,\mathrm{d}t, \tag{4.1}$$

where $|z|$ denotes the modulus of $z \in \mathbb{C}$. This definition leads to a family of distance covariances and then to a definition of distance correlations analogously to classical product moment correlations, that heavily depend on the choice of the weight function ω. Székely, Rizzo and Bakirov (2007) point out that they are only interested in scale invariant nonnegative distance correlations. Furthermore, the authors show that the weight function ω should not be integrable, since otherwise

$$\lim_{\varepsilon\to 0} \frac{\mathcal{V}^2(\varepsilon X, \varepsilon Y; \omega)}{\sqrt{\mathcal{V}^2(\varepsilon X, \varepsilon X; \omega)\mathcal{V}^2(\varepsilon Y, \varepsilon Y; \omega)}} = \rho^2(X, Y).$$

This would imply that distance correlation for dependent but uncorrelated random variables could be arbitrarily close to zero. On the other hand a weight function ω is needed that allows for a simple consistent estimator. Otherwise one would have to deal with estimating high-dimensional integrals using e.g. numerical methods which seems unsatisfying. The striking result of Székely, Rizzo and Bakirov (2007) is that they find a weight function ω such that the sample distance correlation is of a very simple form. The following lemma is the crucial integral observation and the basis of distance covariance and the whole distance correlation calculus.

Lemma 4.1. (Székely, Rizzo and Bakirov, p. 2771) *Suppose that* $0 < \alpha < 2$. *Then for all* $x \in \mathbb{R}^d$

$$\int_{\mathbb{R}^d} \frac{1 - \cos\langle t, x\rangle_d}{|t|_d^{d+\alpha}} = C(d, \alpha)|x|_d^\alpha, \tag{4.2}$$

where

$$C(d, \alpha) = \frac{2\pi^{d/2}\,\Gamma(1 - \alpha/2)}{\alpha\, 2^\alpha\, \Gamma\big((d + \alpha)/2\big)}.$$

The integrals in the neighborhood of 0 *and* ∞ *are meant in the principal value sense:* $\lim_{\tau\to 0+} \int_{\mathbb{R}^d \setminus \{\epsilon B + \epsilon^{-1} B^c\}}$, *where* B *is the unit ball centered at* 0 *in* $\mathbb{R}^k$ *and* $B^c = \mathbb{R}^d \setminus B$ *is the complement of* B.

In section **5** we will generalize this integral. Indeed, it is not necessary to use any regularization method for the integral to converge. The term $1 - \cos\langle t, x\rangle_d$ acts as a natural regularizer. Therefore, the authors choose as a canonical weight function

$$\omega(s, t) = (c_p c_q |s|_p^{p+1} |t|_q^{q+1})^{-1},$$

where $c_p = C(p, 1)$. Then, they show, that the integral in (4.1) is well defined as long as the first moments of X and Y, respectively, exist, which finally leads to the definition of distance covariance.

Definition 4.1. (Székely, Rizzo and Bakirov, p. 2772) *The distance covariance between random vectors* X *and* Y *with finite first moments is the nonnegative number* $\mathcal{V}(X, Y)$ *defined by*

$$\mathcal{V}^2(X, Y) = \frac{1}{c_p c_q} \int_{\mathbb{R}^{p+q}} \frac{|f_{X,Y}(s, t) - f_X(s) f_Y(t)|^2}{|s|_p^{1+p} |t|_q^{1+q}} \,\mathrm{d}s\,\mathrm{d}t$$

and distance variance is then just defined as $\mathcal{V}^2(X, X)$.

Now, distance correlation is defined analogously to classical product moment correlation.

Definition 4.2. (Székely, Rizzo and Bakirov, p. 2773) *The distance correlation between random vectors* X *and* Y *with finite first moments is the nonnegative number* $\mathcal{R}(X, Y)$ *defined by*

$$\mathcal{R}(X, Y) = \frac{\mathcal{V}(X, Y)}{\sqrt{\mathcal{V}(X, X)\mathcal{V}(Y, Y)}} \tag{4.3}$$

if both $\mathcal{V}(X, X)$ *and* $\mathcal{V}(Y, Y)$ *are strictly positive, and defined to be zero otherwise.*

For distributions with finite first moments, the distance correlation characterizes independence in that $0 \le \mathcal{R}(X, Y) \le 1$ with $\mathcal{R}(X, Y) = 0$ if and only if X and Y are independent. We now review the sample versions of the standard distance covariance and distance correlation. Given a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ from jointly distributed random vectors $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$, we set

$$\boldsymbol{X} = [X_1, \ldots, X_n] \in \mathbb{R}^{p\times n} \qquad \text{and} \qquad \boldsymbol{Y} = [Y_1, \ldots, Y_n] \in \mathbb{R}^{q\times n}.$$

An ad hoc way of introducing a sample version of distance covariance is to let

$$f^n_{\boldsymbol{X},\boldsymbol{Y}}(s,t) = \frac{1}{n}\sum_{j=1}^{n} \exp\bigl[i\langle s, X_j\rangle_p + i\langle t, Y_j\rangle_q\bigr]$$

be the corresponding empirical characteristic function, and to write $f^n_{\boldsymbol{X}}(s) = f^n_{\boldsymbol{X},\boldsymbol{Y}}(s,0)$ and $f^n_{\boldsymbol{Y}}(t) = f^n_{\boldsymbol{X},\boldsymbol{Y}}(0,t)$ for the respective marginal empirical characteristic functions. The *sample distance covariance* then is the nonnegative number $\mathcal{V}_n(\boldsymbol{X},\boldsymbol{Y})$ defined by

$$\mathcal{V}_n^2(\boldsymbol{X},\boldsymbol{Y}) = \frac{1}{c_p c_q}\int_{\mathbb{R}^{p+q}} \frac{|f^n_{\boldsymbol{X},\boldsymbol{Y}}(s,t) - f^n_{\boldsymbol{X}}(s) f^n_{\boldsymbol{Y}}(t)|^2}{|s|_p^{1+p}\,|t|_q^{1+q}}\,\mathrm{d}s\,\mathrm{d}t.$$

Székely, et al. (2007), in a *tour de force*, showed that

$$\mathcal{V}_n^2(\boldsymbol{X},\boldsymbol{Y}) = \frac{1}{n^2}\sum_{k,l=1}^{n} A_{kl}B_{kl}, \tag{4.4}$$

where

$$a_{kl} = |X_k - X_l|_p, \quad \bar{a}_{k\cdot} = \frac{1}{n}\sum_{l=1}^{n} a_{kl}, \quad \bar{a}_{\cdot l} = \frac{1}{n}\sum_{k=1}^{n} a_{kl}, \quad \bar{a}_{\cdot\cdot} = \frac{1}{n^2}\sum_{k,l=1}^{n} a_{kl},$$

and

$$A_{kl} = a_{kl} - \bar{a}_{k\cdot} - \bar{a}_{\cdot l} + \bar{a}_{\cdot\cdot},$$

and similarly for $b_{kl} = |Y_k - Y_l|_q$, $\bar{b}_{k\cdot}$, $\bar{b}_{\cdot l}$, $\bar{b}_{\cdot\cdot}$, and B_{kl}, where $k,l = 1,\ldots,n$. Thus, the squared sample distance covariance equals the average entry in the component wise or Schur product of the centered distance matrices for the two variables. The *sample distance correlation* then is defined by

$$\mathcal{R}_n(\boldsymbol{X},\boldsymbol{Y}) = \frac{\mathcal{V}_n(\boldsymbol{X},\boldsymbol{Y})}{\sqrt{\mathcal{V}_n(\boldsymbol{X},\boldsymbol{X})\mathcal{V}_n(\boldsymbol{Y},\boldsymbol{Y})}}, \tag{4.5}$$

if both $\mathcal{V}_n(\boldsymbol{X},\boldsymbol{X})$ and $\mathcal{V}_n(\boldsymbol{Y},\boldsymbol{Y})$ are strictly positive, and defined to be zero otherwise. Székely, Rizzo and Bakirov (2007) show that the sample distance covariance and the sample distance correlation are strongly consistent estimators with $0 \le \mathcal{R}_n \le 1$ and if $\mathcal{R}_n(\boldsymbol{X},\boldsymbol{Y}) = 1$, then there exists a vector a, a nonzero real number b and an orthogonal matrix C such that $\boldsymbol{Y} = a + b\boldsymbol{X}C$. Another crucial property of the distance correlation is that it is invariant under transformations of the form

$$(X,Y) \longmapsto (a_1 + b_1C_1X, a_2 + b_2C_2Y), \tag{4.6}$$

where $a_1 \in \mathbb{R}^p$ and $a_2 \in \mathbb{R}^q$, b_1 and b_2 are nonzero real numbers, and the matrices $C_1 \in \mathbb{R}^{p\times p}$ and $C_2 \in \mathbb{R}^{q\times q}$ are orthogonal. One further major achievement of the paper is a hypothesis test for independence based on a test statistic consisting of sample distance covariance components. Computer code for calculating these sample versions and for applying the test for independence is available in an R package (energy package) by Rizzo and Székely (2011).

One problem concerning distance correlation is that it lacks physical interpretation. The sample distance correlation is very simple but distance correlation on the population side is complicated. One does not know exactly what the sample distance correlation even estimates. Therefore, it is of great value to understand distance correlation for standard distributions as functions of the parameters of the distributions. Székely, Rizzo and Bakirov (2007, p. 2785) take the first step into that direction, that we pursue in the next sections.

Theorem 4.2. *If X and Y are standard normal with correlation ρ, then*

1. $\mathcal{R}(X,Y) \le |\rho|$,
2. $\mathcal{R}^2(X,Y) = \frac{\rho \arcsin \rho + \sqrt{1-\rho^2} - \rho \arcsin \rho/2 - \sqrt{4-\rho^2} + 1}{1+\pi/3-\sqrt{3}}$,
3. $\lim_{\rho \to 0} \frac{\mathcal{R}(X,Y)}{|\rho|} = \frac{1}{2(1+\pi/3-\sqrt{3})^{1/2}} \approx 0.89066.$

Now we discuss an alternative distance correlation, the so called *affinely invariant distance correlation*, Dueck, et al. (2014), and compute affinely invariant distance correlation for the multivariate Gaussian distribution. Furthermore, we derive explicit formulas for distance correlation for Lancaster distributions and generalize the integral in (4.2).

4.2 Affinely Invariant Distance Correlation for the Multivariate Normal Distribution

This subsection concerning the affinely invariant distance correlation is due to Dueck, Edelmann, Gneiting and Richards (2014). We tackle the problem that distance correlation fails to be invariant under the group of all invertible affine transformations.

4.2.1 Affinely Invariant Distance Correlation

Despite of the very nice properties reviewed in the last section, distance correlation fails to be invariant under the group of all invertible affine transformations of (X,Y), which led Székely, et al. (2007, pp. 2784–2785) and Székely and Rizzo (2009, pp. 1252–1253) to propose an affinely invariant sample version of the distance correlation.

Adapting this proposal to the population setting, the *affinely invariant distance covariance* between distributions X and Y with finite second moments can be introduced as the nonnegative number $\widetilde{\mathcal{V}}(X,Y)$ defined by

$$\widetilde{\mathcal{V}}^2(X,Y) = \mathcal{V}^2(\Sigma_X^{-1/2}X, \Sigma_Y^{-1/2}Y), \tag{4.7}$$

where Σ_X and Σ_Y are the respective population covariance matrices. The *affinely invariant distance correlation* between X and Y is the nonnegative number defined by

$$\widetilde{\mathcal{R}}(X,Y) = \frac{\widetilde{\mathcal{V}}(X,Y)}{\sqrt{\widetilde{\mathcal{V}}(X,X)\widetilde{\mathcal{V}}(Y,Y)}} \tag{4.8}$$

if both $\widetilde{\mathcal{V}}(X,X)$ and $\widetilde{\mathcal{V}}(Y,Y)$ are strictly positive, and defined to be zero otherwise. In the sample versions proposed by Székely, et al. (2007), the population quantities are replaced by their natural estimators. Clearly, the population affinely invariant distance correlation and its sample version are invariant under the group of invertible affine transformations, and in addition to satisfying this often-desirable group invariance property (Eaton, 1989), they inherit the desirable properties of the standard distance dependence measures. In particular, $0 \le \widetilde{\mathcal{R}}(X,Y) \le 1$ and, for populations with finite second moments and positive definite covariance matrices, $\widetilde{\mathcal{R}}(X,Y) = 0$ if and only if X and Y are independent.

4.2.2 The Sample Version of the Affinely Invariant Distance Correlation

In this section we describe sample versions of the affinely invariant distance covariance and distance correlation as introduced by Székely, et al. (2007, pp. 2784–2785) and Székely and Rizzo (2009, pp. 1252–1253).

Now let $S_{\boldsymbol{X}}$ and $S_{\boldsymbol{Y}}$ denote the usual sample covariance matrices of the data $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. Following Székely, et al. (2007, p. 2785) and Székely and Rizzo (2009, p. 1253), the *sample affinely invariant distance covariance* is the nonnegative number $\widetilde{\mathcal{V}}_n(\boldsymbol{X},\boldsymbol{Y})$ defined by

$$\widetilde{\mathcal{V}}_n^2(\boldsymbol{X},\boldsymbol{Y}) = \mathcal{V}_n^2(S_{\boldsymbol{X}}^{-1/2}\boldsymbol{X}, S_{\boldsymbol{Y}}^{-1/2}\boldsymbol{Y}) \tag{4.9}$$

if $S_{\boldsymbol{X}}$ and $S_{\boldsymbol{Y}}$ are positive definite, and defined to be zero otherwise. The *sample affinely invariant distance correlation* is defined by

$$\widetilde{\mathcal{R}}_n(\boldsymbol{X},\boldsymbol{Y}) = \frac{\widetilde{\mathcal{V}}_n(\boldsymbol{X},\boldsymbol{Y})}{\sqrt{\widetilde{\mathcal{V}}_n(\boldsymbol{X},\boldsymbol{X})\widetilde{\mathcal{V}}_n(\boldsymbol{Y},\boldsymbol{Y})}} \tag{4.10}$$

if the quantities in the denominator are strictly positive, and defined to be zero otherwise. The sample affinely invariant distance correlation inherits the properties of the sample distance correlation; in particular

$$0 \le \widetilde{\mathcal{R}}_n(\boldsymbol{X},\boldsymbol{Y}) \le 1,$$

and $\widetilde{\mathcal{R}}_n(\boldsymbol{X},\boldsymbol{Y}) = 1$ implies that $p = q$, that the linear spaces spanned by $\boldsymbol{X}$ and $\boldsymbol{Y}$ have full rank, and that there exist a vector $a \in \mathbb{R}^p$, a nonzero number $b \in \mathbb{R}$, and an orthogonal matrix $C \in \mathbb{R}^{p\times p}$ such that $S_{\boldsymbol{Y}}^{-1/2}\boldsymbol{Y} = a + b\,CS_{\boldsymbol{X}}^{-1/2}\boldsymbol{X}$.

Our next result shows that the sample affinely invariant distance correlation is a consistent estimator of the respective population quantity.

Theorem 4.3. *Let $(X,Y) \in \mathbb{R}^{p+q}$ be jointly distributed random vectors with positive definite marginal covariance matrices $\Sigma_X \in \mathbb{R}^{p\times p}$ and $\Sigma_Y \in \mathbb{R}^{q\times q}$, respectively. Suppose that $(X_1,Y_1),\ldots,(X_n,Y_n)$ is a random sample from (X,Y), and let $\boldsymbol{X} = [X_1,\ldots,X_n] \in \mathbb{R}^{p\times n}$ and $\boldsymbol{Y} = [Y_1,\ldots,Y_n] \in \mathbb{R}^{q\times n}$. Also, let $\widehat{\Sigma}_{\boldsymbol{X}}$ and $\widehat{\Sigma}_{\boldsymbol{Y}}$ be strongly consistent estimators for Σ_X and Σ_Y, respectively. Then*

$$\mathcal{V}_n^2(\widehat{\Sigma}_{\boldsymbol{X}}^{-1/2}\boldsymbol{X}, \widehat{\Sigma}_{\boldsymbol{Y}}^{-1/2}\boldsymbol{Y}) \to \widetilde{\mathcal{V}}^2(X,Y),$$

almost surely, as $n \to \infty$. In particular, the sample affinely invariant distance correlation satisfies

$$\widetilde{\mathcal{R}}_n(\boldsymbol{X}, \boldsymbol{Y}) \to \widetilde{\mathcal{R}}(X, Y), \tag{4.11}$$

almost surely.

PROOF. As the covariance matrices Σ_X and Σ_Y are positive definite, we may assume that the strongly consistent estimators $\widehat{\Sigma}_{\boldsymbol{X}}$ and $\widehat{\Sigma}_{\boldsymbol{Y}}$ also are positive definite. Therefore, in order to prove the first statement it suffices to show that

$$\mathcal{V}_n^2(\widehat{\Sigma}_{\boldsymbol{X}}^{-1/2}\boldsymbol{X}, \widehat{\Sigma}_{\boldsymbol{Y}}^{-1/2}\boldsymbol{Y}) - \mathcal{V}_n^2(\Sigma_X^{-1/2}\boldsymbol{X}, \Sigma_Y^{-1/2}\boldsymbol{Y}) \to 0, \tag{4.12}$$

almost surely. By the decomposition of Székely, et al. (2007, p. 2776, Equation (2.18)), the left-hand side of (4.12) can be written as an average of terms of the form

$$\big|\widehat{\Sigma}_{\boldsymbol{X}}^{-1/2}(X_k - X_l)\big|_p \, \big|\widehat{\Sigma}_{\boldsymbol{Y}}^{-1/2}(Y_k - Y_m)\big|_q - \big|\Sigma_X^{-1/2}(X_k - X_l)\big|_p \, \big|\Sigma_Y^{-1/2}(Y_k - Y_m)\big|_q.$$

Using the identity

$$\begin{aligned}
&\big|\widehat{\Sigma}_{\boldsymbol{X}}^{-1/2}(X_k - X_l)\big|_p \, \big|\widehat{\Sigma}_{\boldsymbol{Y}}^{-1/2}(Y_k - Y_m)\big|_q \\
&= \big|(\widehat{\Sigma}_{\boldsymbol{X}}^{-1/2} - \Sigma_X^{-1/2} + \Sigma_X^{-1/2})(X_k - X_l)\big|_p \, \big|(\widehat{\Sigma}_{\boldsymbol{Y}}^{-1/2} - \Sigma_Y^{-1/2} + \Sigma_Y^{-1/2})(Y_k - Y_m)\big|_q,
\end{aligned}$$

we obtain

$$\begin{aligned}
\big|\widehat{\Sigma}_{\boldsymbol{X}}^{-1/2}(X_k - X_l)\big|_p \, \big|\widehat{\Sigma}_{\boldsymbol{Y}}^{-1/2}(Y_k - Y_m)\big|_q &- \big|\Sigma_X^{-1/2}(X_k - X_l)\big|_p \, \big|\Sigma_Y^{-1/2}(Y_k - Y_m)\big|_q \\
&\le \|\widehat{\Sigma}_{\boldsymbol{X}}^{-1/2} - \Sigma_X^{-1/2}\|_1 \, \|\widehat{\Sigma}_{\boldsymbol{Y}}^{-1/2} - \Sigma_Y^{-1/2}\|_1 \, |X_k - X_l|_p \, |Y_k - Y_m|_q \\
&\quad + \|\widehat{\Sigma}_{\boldsymbol{X}}^{-1/2} - \Sigma_X^{-1/2}\|_1 \, |X_k - X_l|_p \, \big|\Sigma_Y^{-1/2}(Y_k - Y_m)\big|_q \\
&\quad + \|\widehat{\Sigma}_{\boldsymbol{Y}}^{-1/2} - \Sigma_Y^{-1/2}\|_1 \, \big|\Sigma_X^{-1/2}(X_k - X_l)\big|_p \, |Y_k - Y_m|_q,
\end{aligned}$$

where the matrix norm $\|\Lambda\|_1$ is the largest eigenvalue of Λ in absolute value. Now we can separate the three sums in the decomposition of Székely, et al. (2007, p. 2776, Equation (2.18)) and place the factors like $\|\widehat{\Sigma}_{\boldsymbol{X}}^{-1/2} - \Sigma_X^{-1/2}\|_1$ in front of the sums, since they appear in every summand. Then, $\|\widehat{\Sigma}_{\boldsymbol{X}}^{-1/2} - \Sigma_X^{-1/2}\|_1$ and $\|\widehat{\Sigma}_{\boldsymbol{Y}}^{-1/2} - \Sigma_Y^{-1/2}\|_1$ tend to zero and the remaining averages converge to constants (representing some distance correlation components) almost surely as $n \to \infty$, and this completes the proof of the first statement. Finally, the property (4.11) of strong consistency of $\widetilde{\mathcal{R}}_n(\boldsymbol{X}, \boldsymbol{Y})$ is obtained immediately upon setting $\widehat{\Sigma}_{\boldsymbol{X}} = S_{\boldsymbol{X}}$ and $\widehat{\Sigma}_{\boldsymbol{Y}} = S_{\boldsymbol{Y}}$. □

Székely, et al. (2007, p. 2783) proposed a test for independence that is based on the sample distance correlation. From their results, we see that the asymptotic properties of the test statistic are not affected by the transition from the standard distance correlation to the affinely invariant distance correlation. Hence, a completely analogous but different test can be stated in terms of the affinely invariant distance correlation. Noting the results of Kosorok (2009, Section 4), we raise the possibility that the specific details can be devised in a judicious, data-dependent way so that the power of the test for independence increases when the transition is made to the affinely invariant distance correlation. An alternative multivariate test for independence based on ranks of distances can be found in Heller, et al. (2012).

4.2.3 The Affinely Invariant Distance Correlation for Multivariate Normal Populations

We now consider the problem of calculating the affinely invariant distance correlation between the random vectors X and Y where $(X, Y) \sim \mathcal{N}_{p+q}(\mu, \Sigma)$, a multivariate normal distribution with mean vector $\mu \in \mathbb{R}^{p+q}$ and covariance matrix $\Sigma \in \mathbb{R}^{(p+q)\times(p+q)}$. We assume, without loss of generality that Σ_X and Σ_Y are nonsingular; otherwise, the problem reduces to a calculation on a lower-dimensional space.

For the case in which $p = q = 1$, i.e., the bivariate normal distribution, the problem was solved by Székely, et al. (2007), see Theorem **4.2**. In that case, the formula for the affinely invariant distance correlation depends only on ρ, the correlation coefficient, and appears in terms of the functions $\sin^{-1}\rho$ and $(1-\rho^2)^{1/2}$, both of which are well-known to be special cases of Gauss' hypergeometric series. Therefore, it is natural to expect that the general case will involve generalizations of Gauss' hypergeometric series, and Theorem 4.4 below demonstrates that such is indeed the case. To formulate this result, we need to recall the rudiments of the theory of zonal polynomials (Muirhead 1982, Chapter 7).

A *partition* κ is a vector of nonnegative integers $(k_1, \ldots, k_q)$ such that $k_1 \geq \cdots \geq k_q$. The integer $|\kappa| = k_1 + \cdots + k_q$ is called the *weight* of κ; and $\ell(\kappa)$, the *length* of κ, is the largest integer j such that $k_j > 0$. The *zonal polynomial* $C_\kappa(\Lambda)$ is a mapping from the class of symmetric matrices $\Lambda \in \mathbb{R}^{q\times q}$ to the real line which satisfies several properties, the following of which are crucial for our results:

(a) Let $O(q)$ denote the group of orthogonal matrices in $\mathbb{R}^{q\times q}$. Then

$$C_\kappa(K'\Lambda K) = C_\kappa(\Lambda) \tag{4.13}$$

for all $K \in O(q)$; thus, $C_\kappa(\Lambda)$ is a symmetric function of the eigenvalues of Λ.

(b) $C_\kappa(\Lambda)$ is homogeneous of degree $|\kappa|$ in Λ: For any $\delta \in \mathbb{R}$,

$$C_\kappa(\delta\Lambda) = \delta^{|\kappa|} C_\kappa(\Lambda). \tag{4.14}$$

(c) If Λ is of rank r then $C_\kappa(\Lambda) = 0$ whenever $\ell(\kappa) > r$.

(d) For any nonnegative integer k,

$$\sum_{|\kappa|=k} C_\kappa(\Lambda) = (\operatorname{tr}\Lambda)^k. \tag{4.15}$$

(e) For any symmetric matrices $\Lambda_1, \Lambda_2 \in \mathbb{R}^{q\times q}$,

$$\int_{O(q)} C_\kappa(K'\Lambda_1 K \Lambda_2)\, \mathrm{d}K = \frac{C_\kappa(\Lambda_1)\, C_\kappa(\Lambda_2)}{C_\kappa(I_q)}, \tag{4.16}$$

where $I_q = \operatorname{diag}(1, \ldots, 1) \in \mathbb{R}^{q\times q}$ denotes the identity matrix and the integral is with respect to the Haar measure on $O(q)$, normalized to have total volume 1.

(f) Let $\lambda_1, \ldots, \lambda_q$ be the eigenvalues of Λ. Then, for a partition (k) with one part,

$$C_{(k)}(\Lambda) = \frac{k!}{(\frac{1}{2})_k} \sum_{i_1+\cdots+i_q=k} \prod_{j=1}^{q} \frac{(\frac{1}{2})_{i_j}\, \lambda_j^{i_j}}{i_j!}\,, \tag{4.17}$$

where the sum is over all nonnegative integers $i_1, \ldots, i_q$ such that $i_1+\cdots+i_q = k$, and

$$(\alpha)_k = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)} = \alpha(\alpha+1)(\alpha+2)\cdots(\alpha+k-1),$$

$\alpha \in \mathbb{C}$, is standard notation for the rising factorial. In particular, on setting $\lambda_j = 1$, $j = 1, \ldots, q$, we obtain from (4.17)

$$C_{(k)}(I_q) = \frac{(\frac{1}{2}q)_k}{(\frac{1}{2})_k}, \tag{4.18}$$

(Muirhead, 1982, p. 237, equation (18); Gross and Richards, 1987, p. 807, Lemma 6.8).

With these properties of the zonal polynomials, we are ready to state our key result which obtains an explicit formula for the affinely invariant distance covariance in the case of a Gaussian population of arbitrary dimension and arbitrary covariance matrix with positive definite marginal covariance matrices. This formula turns out to be a function depending only on the dimensions p and q and the eigenvalues of the matrix $\Lambda = \Sigma_Y^{-1/2}\, \Sigma_{YX}\, \Sigma_X^{-1}\, \Sigma_{XY}\, \Sigma_Y^{-1/2}$, i.e. the squared canonical correlation coefficients of the subvectors X and Y. For fixed dimensions this implies $\widetilde{\mathcal{R}}(X,Y) = g(\lambda_1, \ldots, \lambda_r)$, where $r = \min(p,q)$ and $\lambda_1, \ldots, \lambda_r$ are the canonical correlation coefficients of X and Y. Due to the functional invariance the maximum likelihood estimator (MLE) for affinely invariant distance correlation in the Gaussian setting is hence defined by $g(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_r)$, where $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_r$ are the MLEs of the canonical correlation coefficients.

Theorem 4.4. *Suppose that* $(X,Y) \sim \mathcal{N}_{p+q}(\mu, \Sigma)$, *where*

$$\Sigma = \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix}$$

with $\Sigma_X \in \mathbb{R}^{p\times p}$, $\Sigma_Y \in \mathbb{R}^{q\times q}$, *and* $\Sigma_{XY} \in \mathbb{R}^{p\times q}$. *Then*

$$\widetilde{\mathcal{V}}^2(X,Y) = 4\pi\, \frac{c_{p-1}}{c_p}\frac{c_{q-1}}{c_q} \sum_{k=1}^{\infty} \frac{2^{2k}-2}{k!\, 2^{2k}}\, \frac{(\frac{1}{2})_k\, (-\frac{1}{2})_k\, (-\frac{1}{2})_k}{(\frac{1}{2}p)_k\, (\frac{1}{2}q)_k}\, C_{(k)}(\Lambda), \tag{4.19}$$

where

$$\Lambda = \Sigma_Y^{-1/2}\, \Sigma_{YX}\, \Sigma_X^{-1}\, \Sigma_{XY}\, \Sigma_Y^{-1/2} \in \mathbb{R}^{q\times q}. \tag{4.20}$$

Proof. We may assume, with no loss of generality, that μ is the zero vector. Since Σ_X and Σ_Y both are positive definite the inverse square-roots, $\Sigma_X^{-1/2}$ and $\Sigma_Y^{-1/2}$, exist.

By considering the standardized variables $\widetilde{X} = \Sigma_X^{-1/2} X$ and $\widetilde{Y} = \Sigma_Y^{-1/2} Y$, we may replace the covariance matrix Σ by

$$\widetilde{\Sigma} = \begin{pmatrix} I_p & \Lambda_{XY} \\ \Lambda_{XY}' & I_q \end{pmatrix},$$

where

$$\Lambda_{XY} = \Sigma_X^{-1/2} \, \Sigma_{XY} \, \Sigma_Y^{-1/2}. \tag{4.21}$$

Once we have made these reductions, it follows that the matrix Λ in (4.20) can be written as $\Lambda = \Lambda_{XY}'\Lambda_{XY}$ and that it has norm less than 1. Indeed, by the *partial Iwasawa decomposition* of $\widetilde{\Sigma}$, viz., the identity,

$$\widetilde{\Sigma} = \begin{pmatrix} I_p & 0 \\ \Lambda_{XY}' & I_q \end{pmatrix} \begin{pmatrix} I_p & 0 \\ 0 & I_q - \Lambda_{XY}'\Lambda_{XY} \end{pmatrix} \begin{pmatrix} I_p & \Lambda_{XY} \\ 0 & I_q \end{pmatrix},$$

where the zero matrix of any dimension is denoted by 0, we see that the matrix $\widetilde{\Sigma}$ is positive semidefinite if and only if $I_q - \Lambda$ is positive semidefinite. Hence, $\Lambda \le I_q$ in the Loewner ordering and therefore $\|\Lambda\|_1 \le 1$.

We proceed to calculate the distance covariance $\widetilde{\mathcal{V}}(X,Y) = \mathcal{V}(\widetilde{X},\widetilde{Y})$. It is well-known that the characteristic function of $(\widetilde{X},\widetilde{Y})$ is

$$f_{\widetilde{X},\widetilde{Y}}(s,t) = \exp\Big[-\tfrac{1}{2}\binom{s}{t}'\widetilde{\Sigma}\binom{s}{t}\Big] = \exp\big[-\tfrac{1}{2}(|s|_p^2 + |t|_q^2 + 2s'\Lambda_{XY}t)\big],$$

where $s \in \mathbb{R}^p$ and $t \in \mathbb{R}^q$. Therefore,

$$\big|f_{\widetilde{X},\widetilde{Y}}(s,t) - f_{\widetilde{X}}(s) f_{\widetilde{Y}}(t)\big|^2 = \big(1 - \exp(-s'\Lambda_{XY}t)\big)^2 \exp(-|s|_p^2 - |t|_q^2),$$

and hence

$$\begin{aligned} c_p c_q \, \mathcal{V}^2(\widetilde{X},\widetilde{Y}) &= \int_{\mathbb{R}^{p+q}} \big(1 - \exp(-s'\Lambda_{XY}t)\big)^2 \exp(-|s|_p^2 - |t|_q^2) \, \frac{\mathrm{d}s}{|s|_p^{p+1}} \, \frac{\mathrm{d}t}{|t|_q^{q+1}} \\ &= \int_{\mathbb{R}^{p+q}} \big(1 - \exp(s'\Lambda_{XY}t)\big)^2 \exp(-|s|_p^2 - |t|_q^2) \, \frac{\mathrm{d}s}{|s|_p^{p+1}} \, \frac{\mathrm{d}t}{|t|_q^{q+1}}, \end{aligned} \tag{4.22}$$

where the latter integral is obtained by making the change of variables $s \mapsto -s$ within the former integral.

By a Taylor series expansion, we obtain

$$\begin{aligned} \big(1 - \exp(s'\Lambda_{XY}t)\big)^2 &= 1 - 2\exp(s'\Lambda_{XY}t) + \exp(2s'\Lambda_{XY}t) \\ &= \sum_{k=2}^{\infty} \frac{2^k - 2}{k!} (s'\Lambda_{XY}t)^k. \end{aligned}$$

Substituting this series into (4.22) and interchanging summation and integration, a procedure which is straightforward to verify by means of Fubini's theorem, and noting that the odd-order terms integrate to zero, we obtain

$$c_p c_q \, \mathcal{V}^2(\widetilde{X},\widetilde{Y}) = \sum_{k=1}^{\infty} \frac{2^{2k} - 2}{(2k)!} \int_{\mathbb{R}^{p+q}} (s'\Lambda_{XY}t)^{2k} \exp(-|s|_p^2 - |t|_q^2) \, \frac{\mathrm{d}s}{|s|_p^{p+1}} \, \frac{\mathrm{d}t}{|t|_q^{q+1}}. \tag{4.23}$$

To calculate, for $k \ge 1$, the integral

$$\int_{\mathbb{R}^{p+q}} (s'\Lambda_{XY}t)^{2k} \exp(-|s|_p^2 - |t|_q^2) \frac{\mathrm{d}s}{|s|_p^{p+1}} \frac{\mathrm{d}t}{|t|_q^{q+1}}, \tag{4.24}$$

we change variables to polar coordinates, putting $s = r_x\theta$ and $t = r_y\phi$ where $r_x, r_y > 0$, $\theta = (\theta_1, \ldots, \theta_p)' \in S^{p-1}$, and $\phi = (\phi_1, \ldots, \phi_q)' \in S^{q-1}$. Then the integral (4.24) separates into a product of multiple integrals over (r_x, r_y), and over (θ, ϕ), respectively. The integrals over r_x and r_y are standard gamma integrals,

$$\int_0^\infty \int_0^\infty r_x^{2k-2} r_y^{2k-2} \exp(-r_x^2 - r_y^2)\, \mathrm{d}r_x \mathrm{d}r_y = \tfrac{1}{4} [\Gamma(k - \tfrac{1}{2})]^2 = \left[(-\tfrac{1}{2})_k\right]^2 \pi, \tag{4.25}$$

and the remaining factor is the integral

$$\int_{S^{q-1}} \int_{S^{p-1}} (\theta'\Lambda_{XY}\phi)^{2k}\, \mathrm{d}\theta\, \mathrm{d}\phi, \tag{4.26}$$

where $\mathrm{d}\theta$ and $\mathrm{d}\phi$ are unnormalized surface measures on S^{p-1} and S^{q-1}, respectively. By a standard invariance argument,

$$\int_{S^{p-1}} (\theta' v)^{2k}\, \mathrm{d}\theta = |v|_p^{2k} \int_{S^{p-1}} \theta_1^{2k}\, \mathrm{d}\theta,$$

$v \in \mathbb{R}^p$. Setting $v = \Lambda_{XY}\phi$ and applying some well-known properties of the surface measure $\mathrm{d}\theta$, we obtain

$$\begin{aligned} \int_{S^{p-1}} (\theta'\Lambda_{XY}\phi)^{2k}\, \mathrm{d}\theta &= |\Lambda_{XY}\phi|_p^{2k} \int_{S^{p-1}} \theta_1^{2k}\, \mathrm{d}\theta \\ &= 2c_{p-1} \frac{\Gamma(k + \frac{1}{2})\Gamma(\frac{1}{2}p)}{\Gamma(k + \frac{1}{2}p)\Gamma(\frac{1}{2})} (\phi'\Lambda\phi)^k. \end{aligned}$$

Therefore, in order to evaluate (4.26), it remains to evaluate

$$J_k(\Lambda) = \int_{S^{q-1}} (\phi'\Lambda\phi)^k\, \mathrm{d}\phi.$$

Since the surface measure is invariant under transformation $\phi \mapsto K\phi$, $K \in O(q)$, it follows that $J_k(\Lambda) = J_k(K'\Lambda K)$ for all $K \in O(q)$. Integrating with respect to the normalized Haar measure on the orthogonal group, we conclude that

$$J_k(\Lambda) = \int_{O(q)} J_k(K'\Lambda K)\, \mathrm{d}K = \int_{S^{q-1}} \int_{O(q)} (\phi' K'\Lambda K\phi)^k\, \mathrm{d}K\, \mathrm{d}\phi. \tag{4.27}$$

We now use the properties of the zonal polynomials. By (4.15),

$$(\phi' K'\Lambda K\phi)^k = (\operatorname{tr} K'\Lambda K\phi\phi')^k = \sum_{|\kappa|=k} C_\kappa(K'\Lambda K\phi\phi');$$

therefore, by (4.16),

$$\int_{O(q)} (\phi' K' \Lambda K \phi)^k \, \mathrm{d}K = \sum_{|\kappa|=k} \int_{O(q)} C_\kappa(K' \Lambda K \phi\phi') \, \mathrm{d}K = \sum_{|\kappa|=k} \frac{C_\kappa(\Lambda) \, C_\kappa(\phi\phi')}{C_\kappa(I_q)}.$$

Since $\phi\phi'$ is of rank 1 then, by property (c), $C_\kappa(\phi\phi') = 0$ if $\ell(\kappa) > 1$; it now follows, by (4.15) and the fact that $\phi \in S^{q-1}$, that

$$C_{(k)}(\phi\phi') = \sum_{|\kappa|=k} C_\kappa(\phi\phi') = (\operatorname{tr} \phi\phi')^k = (\phi'\phi)^k = |\phi|_q^{2k} = 1.$$

Therefore,

$$\int_{O(q)} (\phi' K' \Lambda K \phi)^k \, \mathrm{d}K = \frac{C_{(k)}(\Lambda)}{C_{(k)}(I_q)} = \frac{(\frac{1}{2})_k}{(\frac{1}{2}q)_k} \, C_{(k)}(\Lambda),$$

where the last equality follows by (4.18). Substituting this result at (4.27), we obtain

$$J_k(\Lambda) = 2c_{q-1} \frac{(\frac{1}{2})_k}{(\frac{1}{2}q)_k} \, C_{(k)}(\Lambda). \tag{4.28}$$

Collecting together these results, and using the well-known identity $(2k)! = k! \, 2^{2k} \, (\frac{1}{2})_k$, we obtain the representation (4.19), as desired. $\square$

We remark that by interchanging the roles of X and Y in Theorem 4.4, we would obtain (4.19) with Λ in (4.20) replaced by

$$\Lambda_0 = \Sigma_X^{-1/2} \, \Sigma_{XY} \, \Sigma_Y^{-1} \, \Sigma_{YX} \, \Sigma_X^{-1/2} \in \mathbb{R}^{p \times p}.$$

Since Λ and Λ_0 have the same characteristic polynomial and hence the same set of nonzero eigenvalues, and noting that $C_\kappa(\Lambda)$ depends only on the eigenvalues of Λ, it follows that $C_{(k)}(\Lambda) = C_{(k)}(\Lambda_0)$. Therefore, the series representation (4.19) for $\widetilde{\mathcal{V}}^2(X, Y)$ remains unchanged if the roles of X and Y are interchanged.

The series appearing in Theorem 4.4 can be expressed in terms of the generalized hypergeometric functions of matrix argument (James, 1964; Muirhead, 1982; Gross and Richards, 1987). For this purpose, we introduce the *partitional rising factorial* for any $\alpha \in \mathbb{C}$ and any partition $\kappa = (k_1, \ldots, k_q)$ as

$$(\alpha)_\kappa = \prod_{j=1}^{q} \left(\alpha - \tfrac{1}{2}(j-1)\right)_{k_j}.$$

Let $\alpha_1, \ldots, \alpha_l, \beta_1, \ldots, \beta_m \in \mathbb{C}$ where $-\beta_i + \frac{1}{2}(j-1)$ is not a nonnegative integer, for all $i = 1, \ldots, m$ and $j = 1, \ldots, q$. Then the ${}_lF_m$ generalized hypergeometric function of matrix argument is defined as

$${}_lF_m(\alpha_1, \ldots, \alpha_l; \beta_1, \ldots, \beta_m; S) = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{|\kappa|=k} \frac{(\alpha_1)_\kappa \cdots (\alpha_l)_\kappa}{(\beta_1)_\kappa \cdots (\beta_m)_\kappa} C_\kappa(S),$$

where S is a symmetric matrix. A complete analysis of the convergence properties of this series was derived by Gross and Richards (1987, p. 804, Theorem 6.3), and we refer the reader to that paper for the details.

Corollary 4.5. *In the setting of Theorem 4.4, we have*

$$\widetilde{\mathcal{V}}^2(X,Y) = 4\pi \frac{c_{p-1}}{c_p}\frac{c_{q-1}}{c_q} \times \Big({}_3F_2\big(\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{1}{2};\tfrac{1}{2}p,\tfrac{1}{2}q;\Lambda\big) - 2\,{}_3F_2\big(\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{1}{2};\tfrac{1}{2}p,\tfrac{1}{2}q;\tfrac{1}{4}\Lambda\big) + 1\Big). \quad (4.29)$$

Proof. It is evident that

$$(\tfrac{1}{2})_\kappa = \begin{cases} (\tfrac{1}{2})_{k_1}, & \text{if } \ell(\kappa) \le 1, \\ 0, & \text{if } \ell(\kappa) > 1. \end{cases}$$

Therefore, we now can write the series in (4.19), up to a multiplicative constant, in terms of a generalized hypergeometric function of matrix argument, in that

$$\begin{aligned} &\sum_{k=1}^{\infty} \frac{2^{2k}-2}{k!\,2^{2k}} \frac{(\frac{1}{2})_k(-\frac{1}{2})_k(-\frac{1}{2})_k}{(\frac{1}{2}p)_k(\frac{1}{2}q)_k} C_{(k)}(\Lambda) \\ &= \sum_{k=1}^{\infty} \frac{2^{2k}-2}{k!\,2^{2k}} \sum_{|\kappa|=k} \frac{(\frac{1}{2})_\kappa(-\frac{1}{2})_\kappa(-\frac{1}{2})_\kappa}{(\frac{1}{2}p)_\kappa(\frac{1}{2}q)_\kappa} C_\kappa(\Lambda) \\ &= \sum_{k=1}^{\infty} \frac{1}{k!} \sum_{|\kappa|=k} \frac{(\frac{1}{2})_\kappa(-\frac{1}{2})_\kappa(-\frac{1}{2})_\kappa}{(\frac{1}{2}p)_\kappa(\frac{1}{2}q)_\kappa} C_\kappa(\Lambda) - 2\sum_{k=1}^{\infty} \frac{1}{k!\,2^{2k}} \sum_{|\kappa|=k} \frac{(\frac{1}{2})_\kappa(-\frac{1}{2})_\kappa(-\frac{1}{2})_\kappa}{(\frac{1}{2}p)_\kappa(\frac{1}{2}q)_\kappa} C_\kappa(\Lambda) \\ &= \left[{}_3F_2\left(\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{1}{2};\tfrac{1}{2}p,\tfrac{1}{2}q;\Lambda\right) - 1\right] - 2\left[{}_3F_2\left(\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{1}{2};\tfrac{1}{2}p,\tfrac{1}{2}q;\tfrac{1}{4}\Lambda\right) - 1\right]. \end{aligned}$$

Due to property (4.14) it remains to show that the zonal polynomial series expansion for the ${}_3F_2\left(\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{1}{2};\tfrac{1}{2}p,\tfrac{1}{2}q;\Lambda\right)$ generalized hypergeometric function of matrix argument converges absolutely for all Λ with $\Lambda \le I_q$ in the Loewner ordering. By (4.18)

$$\begin{aligned} {}_3F_2\left(\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{1}{2};\tfrac{1}{2}p,\tfrac{1}{2}q;\Lambda\right) &\le \sum_{k=0}^{\infty} \frac{2^{2k}}{k!\,2^{2k}} \frac{(-\frac{1}{2})_k(-\frac{1}{2})_k}{(\frac{1}{2}p)_k} \\ &= {}_2F_1\left(-\tfrac{1}{2},-\tfrac{1}{2};\tfrac{1}{2}p;1\right). \end{aligned}$$

The latter series converges due to Gauss' Theorem for hypergeometric functions and so we have absolute convergence at (4.29) for all Σ with positive definite marginal covariance matrices. □

Consider the case in which $q = 1$ and p is arbitrary. Then Λ is a scalar; say, $\Lambda = \rho^2$ for some $\rho \in [-1,1]$. Then the ${}_3F_2$ generalized hypergeometric functions in (4.29) each reduce to a Gaussian hypergeometric function, denoted by ${}_2F_1$, and (4.29) becomes

$$\widetilde{\mathcal{V}}^2(X,Y) = 4\frac{c_{p-1}}{c_p}\Big({}_2F_1\left(-\tfrac{1}{2},-\tfrac{1}{2};\tfrac{1}{2}p;\rho^2\right) - 2\,{}_2F_1\left(-\tfrac{1}{2},-\tfrac{1}{2};\tfrac{1}{2}p;\tfrac{1}{4}\rho^2\right) + 1\Big).$$

For the case in which $p = q = 1$, we may identify ρ with the Pearson correlation coefficient and the hypergeometric series can be expressed in terms of elementary functions. By well-known results (Andrews, Askey, and Roy (2000), pp. 64 and 94),

$${}_2F_1(-\tfrac{1}{2},-\tfrac{1}{2};\tfrac{1}{2};\rho^2) = \rho\sin^{-1}\rho + (1-\rho^2)^{1/2}, \quad (4.30)$$

and thus we derive the same result for $p = q = 1$ as in Székely, et al. (2007, p. 2786).

For cases in which $q = 1$ and p is odd, we can again obtain explicit expressions for $\widetilde{\mathcal{V}}^2(X,Y)$. In such cases, the ${}_3F_2$ generalized hypergeometric functions in (4.29) reduce to Gaussian hypergeometric functions of the form ${}_2F_1(-\frac{1}{2},-\frac{1}{2};k+\frac{1}{2};\rho^2)$, $k \in \mathbb{N}$, and it can be shown that these latter functions are expressible in closed form in terms of elementary functions and the $\sin^{-1}(\cdot)$ function. For instance, for $p = 3$, the contiguous relations for the ${}_2F_1$ functions can be used to show that

$$ {}_2F_1(-\tfrac{1}{2},-\tfrac{1}{2};\tfrac{3}{2};\rho^2) = \frac{3(1-\rho^2)^{1/2}}{4} + \frac{(1+2\rho^2)\sin^{-1}\rho}{4\rho}. \tag{4.31} $$

Further, by repeated application of the same contiguous relations, it can be shown that for $k = 2,3,4,\ldots$,

$$ {}_2F_1(-\tfrac{1}{2},-\tfrac{1}{2};k+\tfrac{1}{2};\rho^2) = \rho^{-2(k-1)}(1-\rho^2)^{1/2}P_{k-1}(\rho^2) + \rho^{-(2k-1)}Q_k(\rho^2)\sin^{-1}\rho, $$

where P_k and Q_k are polynomials of degree k. Therefore, for $q = 1$ and p odd, the distance covariance $\widetilde{\mathcal{V}}^2(X,Y)$ can be expressed in closed form in terms of elementary functions and the $\sin^{-1}(\cdot)$ function.

The appearance of the generalized hypergeometric functions of matrix argument also yields a useful expression for the affinely invariant distance variance. In order to state this result, we shall define for each positive integer p the quantity

$$ A(p) = \frac{\Gamma(\frac{1}{2}p)\,\Gamma(\frac{1}{2}p+1)}{\big[\Gamma\big(\frac{1}{2}(p+1)\big)\big]^2} - 2\,{}_2F_1\big(-\tfrac{1}{2},-\tfrac{1}{2};\tfrac{1}{2}p;\tfrac{1}{4}\big) + 1. \tag{4.32} $$

Corollary 4.6. *In the setting of Theorem 4.4, we have*

$$ \widetilde{\mathcal{V}}^2(X,X) = 4\pi\frac{c_{p-1}^2}{c_p^2}\,A(p). \tag{4.33} $$

Proof. We are in the special case of Theorem 4.4 for which $X = Y$, so that $p = q$ and $\Lambda = I_p$. By applying (4.18) we can write the series in (4.19) as

$$ \begin{aligned} & 4\pi\frac{c_{p-1}^2}{c_p^2}\sum_{k=1}^{\infty}\frac{2^{2k}-2}{k!\,2^{2k}}\frac{(\frac{1}{2})_k(-\frac{1}{2})_k(-\frac{1}{2})_k}{(\frac{1}{2}p)_k(\frac{1}{2}p)_k}\,C_{(k)}(I_p) \\ &\qquad = 4\pi\frac{c_{p-1}^2}{c_p^2}\sum_{k=1}^{\infty}\frac{2^{2k}-2}{k!\,2^{2k}}\frac{(-\frac{1}{2})_k(-\frac{1}{2})_k}{(\frac{1}{2}p)_k} \\ &\qquad = 4\pi\frac{c_{p-1}^2}{c_p^2}\Big(\big[{}_2F_1\big(-\tfrac{1}{2},-\tfrac{1}{2};\tfrac{1}{2}p;1\big)-1\big] - 2\big[{}_2F_1\big(-\tfrac{1}{2},-\tfrac{1}{2};\tfrac{1}{2}p;\tfrac{1}{4}\big)-1\big]\Big). \end{aligned} $$

By Gauss' Theorem for hypergeometric functions the series ${}_2F_1(-\frac{1}{2},-\frac{1}{2};\frac{1}{2}p;z)$ also converges for the special value $z = 1$, and then

$$ {}_2F_1(-\tfrac{1}{2},-\tfrac{1}{2};\tfrac{1}{2}p;1) = \frac{\Gamma(\frac{1}{2}p)\,\Gamma(\frac{1}{2}p+1)}{\big[\Gamma\big(\frac{1}{2}(p+1)\big)\big]^2}, $$

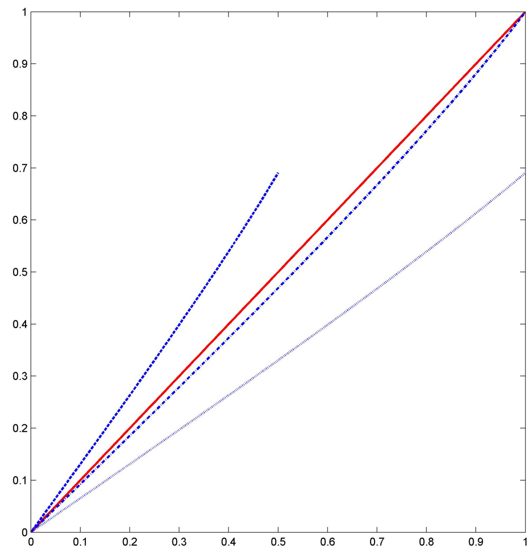

Figure 4: The affinely invariant distance correlation for subvectors of a multivariate normal population, where $p = q = 2$, as a function of the parameter r in three distinct settings. The solid diagonal line is the identity function and is provided to serve as a reference for the three distance correlation functions. See the text for details.

thereby completing the proof. $\square$

For cases in which p is odd, we can proceed as explained at (4.31) to obtain explicit values for the Gaussian hypergeometric function remaining in (4.33). This leads in such cases to explicit expressions for the exact value of $\widetilde{\mathcal{V}}^2(X, X)$. In particular, if $p = 1$ then it follows (4.30) that

$$\widetilde{\mathcal{V}}^2(X, X) = \frac{4}{3} - \frac{4(\sqrt{3} - 1)}{\pi};$$

and for $p = 3$, we deduce from (4.31) that

$$\widetilde{\mathcal{V}}^2(X, X) = 2 - \frac{4(3\sqrt{3} - 4)}{\pi}.$$

Corollaries 4.5 and 4.6 enable the explicit and efficient calculation of the affinely invariant distance correlation (4.8) in the case of subvectors of a multivariate normal population. In doing so, we use the algorithm of Koev and Edelman (2006) to evaluate the generalized hypergeometric function of matrix argument, with C and Matlab code being available at these authors' websites.

Figure 4 concerns the case $p = q = 2$ in various settings, in which the matrix Λ_{XY} depends on a single parameter r only. The dotted line shows the affinely invariant distance correlation when

$$\Lambda_{XY} = \begin{pmatrix} 0 & 0 \\ 0 & r \end{pmatrix};$$

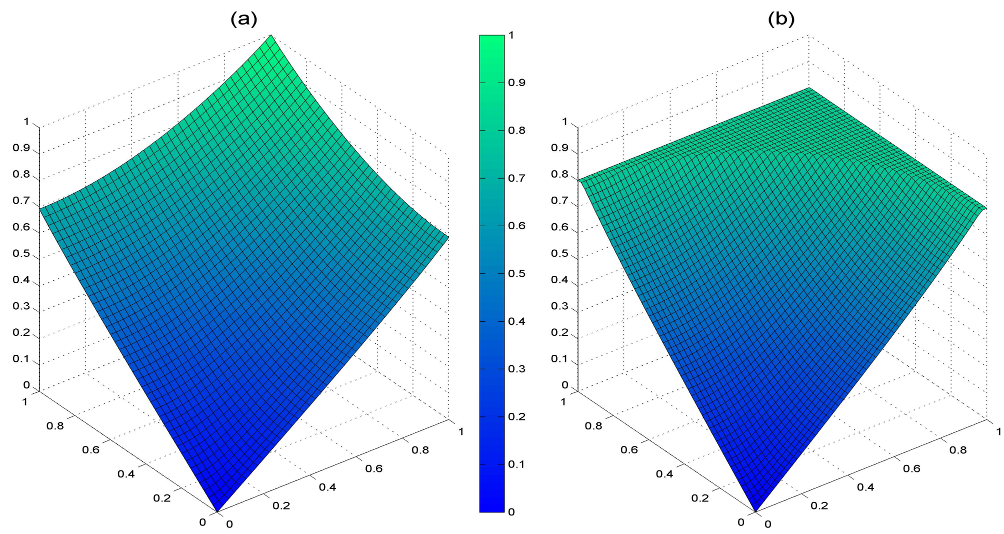

Figure 5: The affinely invariant distance correlation between the p- and q-dimensional subvectors of a $(p+q)$-dimensional multivariate normal population, where (a) $p = q = 2$ and $\Lambda_{XY} = \mathrm{diag}(r, s)$, and (b) $p = 2$, $q = 1$ and $\Lambda_{XY} = (r, s)'$.

this is the case with the weakest dependence considered here. The dash-dotted line applies when

$$\Lambda_{XY} = \begin{pmatrix} r & 0 \\ 0 & r \end{pmatrix}.$$

The strongest dependence corresponds to the dashed line, which shows the affinely invariant distance correlation when

$$\Lambda_{XY} = \begin{pmatrix} r & r \\ r & r \end{pmatrix};$$

in this case we need to assume that $0 \le r \le \frac{1}{2}$ in order to retain positive definiteness.

In Figure 5, panel (a) shows the affinely invariant distance correlation when $p = q = 2$ and

$$\Lambda_{XY} = \begin{pmatrix} r & 0 \\ 0 & s \end{pmatrix},$$

where $0 \le r, s \le 1$. With reference to Figure 4, the margins correspond to the dotted line and the diagonal corresponds to the dash-dotted line.

Panel (b) of Figure 5 concerns the case in which $p = 2$, $q = 1$ and $\Lambda_{XY} = (r, s)'$, where $r^2 + s^2 \le 1$. Here, the affinely invariant distance correlation attains an upper limit as $r^2 + s^2 \uparrow 1$, and we have evaluated that limit numerically as 0.8252.

4.2.4 Limit Theorems

We now study the limiting behavior of the affinely invariant distance correlation measures for subvectors of multivariate normal populations.

Our first result quantifies the asymptotic decay of the affinely invariant distance correlation in the case in which the cross-covariance matrix converges to the zero matrix, in that

$$\operatorname{tr}(\Lambda) = \|\Lambda_{XY}\|^2 \longrightarrow 0,$$

where $\|\cdot\|$ denotes the Frobenius norm, and the matrices $\Lambda = \Lambda_{XY}{}'\Lambda_{XY}$ and Λ_{XY} are defined in (4.20) and (4.21), respectively.

Theorem 4.7. *Suppose that* $(X, Y) \sim \mathcal{N}_{p+q}(\mu, \Sigma)$, *where*

$$\Sigma = \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix}$$

with $\Sigma_X \in \mathbb{R}^{p\times p}$ *and* $\Sigma_Y \in \mathbb{R}^{q\times q}$ *being positive definite, and suppose that the matrix* Λ *in (4.20) has positive trace. Then,*

$$\lim_{\operatorname{tr}(\Lambda)\to 0} \frac{\widetilde{\mathcal{R}}^2(X,Y)}{\operatorname{tr}(\Lambda)} = \frac{1}{4\,pq\sqrt{A(p)A(q)}}, \tag{4.34}$$

where $A(p)$ *is defined in (4.32).*

Proof. We first note that $\widetilde{\mathcal{V}}^2(X,X)$ and $\widetilde{\mathcal{V}}^2(Y,Y)$ do not depend on Σ_{XY}, as can be seen from their explicit representations in terms of $A(p)$ and $A(q)$ given in (4.33).

In studying the asymptotic behavior of $\widetilde{\mathcal{V}}^2(X,Y)$, we may interchange the limit and the summation in the series representation (4.19). Hence, it suffices to find the limit term-by-term. Since $C_{(1)}(\Lambda) = \operatorname{tr}(\Lambda)$ then the ratio of the term for $k=1$ and $\operatorname{tr}(\Lambda)$ equals

$$\frac{c_{p-1}}{c_p}\frac{c_{q-1}}{c_q}\frac{\pi}{pq}.$$

For $k \ge 2$, it follows from (4.17) that $C_{(k)}(\Lambda)$ is a sum of monomials in the eigenvalues of Λ, with each monomial being of degree k, which is greater than the degree, viz. 1, of $\operatorname{tr}(\Lambda)$; therefore,

$$\lim_{\operatorname{tr}(\Lambda)\to 0} \frac{C_{(k)}(\Lambda)}{\operatorname{tr}(\Lambda)} = \lim_{\Lambda\to 0} \frac{C_{(k)}(\Lambda)}{\operatorname{tr}(\Lambda)} = 0.$$

Collecting these facts together, we obtain (4.34). □

If $p = q = 1$ we are in the situation of Theorem 7(iii) in Székely, et al. (2007). Applying the identity (4.30), we obtain

$${}_2F_1(-\tfrac{1}{2}, -\tfrac{1}{2}; \tfrac{1}{2}; \tfrac{1}{4}) = \frac{\pi}{12} + \frac{\sqrt{3}}{2},$$

and $(\operatorname{tr}(\Lambda))^{1/2} = |\rho|$. Thus we obtain

$$\lim_{\rho\to 0} \frac{\widetilde{\mathcal{R}}(X,Y)}{|\rho|} = \frac{1}{2\left(1 + \frac{1}{3}\pi - \sqrt{3}\right)^{1/2}},$$

as shown by Székely, et al. (2007, p. 2785).

In the remainder of this section we consider situations in which one or both of the dimensions p and q grow without bound. We will repeatedly make use of the fact that

$$\frac{c_{p-1}}{\sqrt{p}\, c_p} \longrightarrow \frac{1}{\sqrt{2\pi}} \tag{4.35}$$

as $p \to \infty$, which follows easily from the functional equation for the gamma function along with Stirling's formula.

Theorem 4.8. *For each positive integer p, suppose that $(X_p, Y_p) \sim \mathcal{N}_{2p}(\mu_p, \Sigma_p)$, where*

$$\Sigma_p = \begin{pmatrix} \Sigma_{X,p} & \Sigma_{XY,p} \\ \Sigma_{YX,p} & \Sigma_{Y,p} \end{pmatrix}$$

with $\Sigma_{X,p} \in \mathbb{R}^{p\times p}$ and $\Sigma_{Y,p} \in \mathbb{R}^{p\times p}$ being positive definite and such that

$$\Lambda_p = \Sigma_{Y,p}^{-1/2}\, \Sigma_{YX,p}\, \Sigma_{X,p}^{-1}\, \Sigma_{XY,p}\, \Sigma_{Y,p}^{-1/2} \neq 0.$$

Then

$$\lim_{p\to\infty} \frac{p}{\operatorname{tr}(\Lambda_p)}\, \widetilde{\mathcal{V}}^2(X_p, Y_p) = \frac{1}{2} \tag{4.36}$$

and

$$\lim_{p\to\infty} \frac{p}{\operatorname{tr}(\Lambda_p)}\, \widetilde{\mathcal{R}}^2(X_p, Y_p) = 1. \tag{4.37}$$

In particular, if $\Lambda_p = r^2 I_p$ for some $r \in [0,1]$, then $\operatorname{tr}(\Lambda_p) = r^2 p$, and so (4.36) and (4.37) reduce to

$$\lim_{p\to\infty} \widetilde{\mathcal{V}}^2(X_p, Y_p) = \frac{1}{2}\, r^2 \qquad \text{and} \qquad \lim_{p\to\infty} \widetilde{\mathcal{R}}(X_p, Y_p) = r,$$

respectively. The following corollary concerns the special case in which $r = 1$; we state it separately for emphasis.

Corollary 4.9. *For each positive integer p, suppose that $X_p \sim \mathcal{N}_p(\mu_p, \Sigma_p)$, with Σ_p being positive definite. Then*

$$\lim_{p\to\infty} \widetilde{\mathcal{V}}^2(X_p, X_p) = \frac{1}{2}. \tag{4.38}$$

Proof of Theorem 4.8 and Corollary 4.9. In order to prove (4.36) we study the limit for the terms corresponding separately to $k = 1$, $k = 2$, and $k \geq 3$ in (4.19).

For $k = 1$, on recalling that $C_{(1)}(\Lambda_p) = \operatorname{tr}(\Lambda_p)$, it follows from (4.35) that the ratio of that term to $\operatorname{tr}(\Lambda_p)/p$ tends to $1/2$.

For $k = 2$, we first deduce from (4.15) that $C_{(2)}(\Lambda_p) \leq (\operatorname{tr}\Lambda_p)^2$. Moreover, $\operatorname{tr}(\Lambda_p) \leq p$ because $\Lambda_p \leq I_p$ in the Loewner ordering. Thus, the ratio of the second term in (4.19) to $\operatorname{tr}(\Lambda_p)/p$ is a constant multiple of

$$\frac{p}{\operatorname{tr}(\Lambda_p)}\, \frac{c_{p-1}^2}{c_p^2}\, \frac{C_{(2)}(\Lambda_p)}{(\frac{1}{2}p)_2\, (\frac{1}{2}p)_2} \leq \frac{c_{p-1}^2}{c_p^2}\, \frac{p^2}{(\frac{1}{2}p)_2\, (\frac{1}{2}p)_2} = 4\, \frac{p}{(p+1)^2}\, \frac{c_{p-1}^2}{p\, c_p^2}$$

which, by (4.35), converges to zero as $p \to \infty$.

Finally, suppose that $k \geq 3$. Obviously, $\Lambda_p \leq \|\Lambda_p\|_1 I_p$ in the Loewner ordering inequality, and so it follows from (4.17) that $C_{(k)}(\Lambda_p) \leq \|\Lambda_p\|_1^k \, C_{(k)}(I_p)$. Also, since $\mathrm{tr}\,(\Lambda_p) \geq \|\Lambda_p\|$ then by again applying the Loewner ordering inequality and (4.18) we obtain

$$\frac{C_{(k)}(\Lambda_p)}{\mathrm{tr}\,(\Lambda_p)} \leq \frac{\|\Lambda_p\|_1^k \, C_{(k)}(I_p)}{\|\Lambda_p\|_1} = \|\Lambda_p\|_1^{k-1} C_{(k)}(I_p) \leq C_{(k)}(I_p) = \frac{(\frac{1}{2}p)_k}{(\frac{1}{2})_k}. \tag{4.39}$$

Therefore,

$$\begin{aligned} 4\pi \, \frac{p}{\mathrm{tr}\,(\Lambda_p)} \, \frac{c_{p-1}^2}{c_p^2} \sum_{k=3}^{\infty} \frac{2^{2k}-2}{k!\,2^{2k}} \, \frac{(\frac{1}{2})_k \, (-\frac{1}{2})_k \, (-\frac{1}{2})_k}{(\frac{1}{2}p)_k \, (\frac{1}{2}p)_k} \, C_{(k)}(\Lambda_p) & \\ \leq 4\pi \, p \, \frac{c_{p-1}^2}{c_p^2} \sum_{k=3}^{\infty} \frac{2^{2k}-2}{k!\,2^{2k}} \frac{(-\frac{1}{2})_k \, (-\frac{1}{2})_k}{(\frac{1}{2}p)_k}. & \end{aligned}$$

By (4.35), each term $pc_{p-1}^2/(\frac{1}{2}p)_k c_p^2$ converges to zero as $p \to \infty$, and this proves both (4.36) and its special case, (4.38). Then, (4.37) follows immediately. $\square$

Finally, we consider the situation in which q, the dimension of Y, is fixed while p, the dimension of X, grows without bound.

Theorem 4.10. *For each positive integer p, suppose that $(X_p, Y) \sim \mathcal{N}_{p+q}(\mu_p, \Sigma_p)$, where*

$$\Sigma_p = \begin{pmatrix} \Sigma_{X,\,p} & \Sigma_{XY,\,p} \\ \Sigma_{YX,\,p} & \Sigma_Y \end{pmatrix}$$

with $\Sigma_{X,\,p} \in \mathbb{R}^{p \times p}$ and $\Sigma_Y \in \mathbb{R}^{q \times q}$ being positive definite and such that

$$\Lambda_p = \Sigma_Y^{-1/2} \, \Sigma_{YX,\,p} \, \Sigma_{X,\,p}^{-1} \, \Sigma_{XY,\,p} \, \Sigma_Y^{-1/2} \neq 0.$$

Then

$$\lim_{p \to \infty} \frac{\sqrt{p}}{\mathrm{tr}\,(\Lambda_p)} \, \widetilde{\mathcal{V}}^2(X_p, Y) = \sqrt{\frac{\pi}{2}} \frac{c_{q-1}}{q \, c_q} \tag{4.40}$$

and

$$\lim_{p \to \infty} \frac{\sqrt{p}}{\mathrm{tr}\,(\Lambda_p)} \, \widetilde{\mathcal{R}}^2(X_p, Y) = \frac{1}{2q\sqrt{A(q)}}. \tag{4.41}$$

Proof. By (4.19),

$$\widetilde{\mathcal{V}}^2(X_p, Y) = 4\pi \, \frac{c_{p-1}}{c_p} \frac{c_{q-1}}{c_q} \sum_{k=1}^{\infty} \frac{2^{2k}-2}{k!\,2^{2k}} \frac{(\frac{1}{2})_k \, (-\frac{1}{2})_k \, (-\frac{1}{2})_k}{(\frac{1}{2}p)_k \, (\frac{1}{2}q)_k} \, C_{(k)}(\Lambda_p).$$

We now examine the limiting behavior, as $p \to \infty$, of the terms in this sum for $k = 1$ and, separately, for $k \geq 2$.

For $k = 1$, the limiting value of the ratio of the corresponding term to $\operatorname{tr}(\Lambda_p)/\sqrt{p}$ equals

$$\pi \, \frac{c_{q-1}}{q \, c_q} \lim_{p \to \infty} \frac{\sqrt{p}}{\operatorname{tr}(\Lambda_p)} \frac{c_{p-1}}{p \, c_p} \, C_{(1)}(\Lambda_p) = \sqrt{\frac{\pi}{2}} \frac{c_{q-1}}{q \, c_q}$$

by (4.35) and the fact that $C_{(1)}(\Lambda_p) = \operatorname{tr}(\Lambda_p)$.

For $k \geq 2$, the ratio of the sum to $\operatorname{tr}(\Lambda_p)/\sqrt{p}$ equals

$$\begin{aligned} 4\pi \, \frac{\sqrt{p}}{\operatorname{tr}(\Lambda_p)} \frac{c_{p-1}}{c_p} \frac{c_{q-1}}{c_q} \sum_{k=2}^{\infty} \frac{2^{2k} - 2}{k! \, 2^{2k}} \frac{(\frac{1}{2})_k \, (-\frac{1}{2})_k \, (-\frac{1}{2})_k}{(\frac{1}{2}p)_k \, (\frac{1}{2}q)_k} \, C_{(k)}(\Lambda_p) & \\ \leq 4\pi \, \frac{\sqrt{p}}{\|\Lambda_p\|_1} \frac{c_{p-1}}{c_p} \frac{c_{q-1}}{c_q} \sum_{k=2}^{\infty} \frac{2^{2k} - 2}{k! \, 2^{2k}} \frac{(-\frac{1}{2})_k \, (-\frac{1}{2})_k}{(\frac{1}{2}p)_k} \, \|\Lambda_p\|_1^k & \\ \leq 4\pi \, \sqrt{p} \, \frac{c_{p-1}}{c_p} \frac{c_{q-1}}{c_q} \sum_{k=2}^{\infty} \frac{2^{2k} - 2}{k! \, 2^{2k}} \frac{(-\frac{1}{2})_k \, (-\frac{1}{2})_k}{(\frac{1}{2}p)_k}, & \end{aligned}$$

where we have used (4.39) to obtain the last two inequalities. By applying (4.35), we see that the latter upper bound converges to 0 as $p \to \infty$, which proves (4.40), and then (4.41) follows immediately. $\square$

The results in this section have practical implications for affine distance correlation analysis of large-sample, high-dimensional Gaussian data. In the setting of Theorem 4.10, $\operatorname{tr}(\Lambda_p) \leq q$ is bounded, and so

$$\lim_{p \to \infty} \widetilde{\mathcal{R}}(X_p, Y) = 0.$$

As a consequence of Theorem 4.3 on the consistency of sample measures, it follows that the direct calculation of affine distance correlation measures for such data will return values which are virtually zero. In practice, in order to obtain values of the sample affine distance correlation measures which permit statistical inference, it will be necessary to calculate $\widehat{\Lambda}_p$, the maximum likelihood estimator of Λ_p, and then to rescale the distance correlation measures with the factor $\sqrt{p}/\operatorname{tr}(\widehat{\Lambda}_p)$. In the scenario of Theorem 4.8 the asymptotic behavior of the affine distance correlation measures depends on the ratio $p/\operatorname{tr}(\Lambda_p)$; and as $\operatorname{tr}(\Lambda_p)$ can attain any value in the interval $[0, p]$, a wide range of asymptotic rates of convergence is conceivable.

In all these settings, the series representation (4.19) can be used to obtain complete asymptotic expansions in powers of p^{-1} or q^{-1}, of the affine distance covariance or correlation measures, as p or q tend to infinity.

4.2.5 Time Series of Wind Vectors at the Stateline Wind Energy Center

Rémillard (2009) proposed the use of the distance correlation to explore nonlinear dependencies in time series data. Zhou (2012) pursued this approach recently and defined the auto distance covariance function and the auto distance correlation function, along with natural sample versions, for a strongly stationary vector-valued time series, say $(X_j)_{j=-\infty}^{\infty}$.

It is straightforward to extend these notions to the affinely invariant distance correlation. Thus, for an integer k, we refer to

$$\widetilde{\mathcal{R}}_X(k) = \frac{\widetilde{\mathcal{V}}(X_j, X_{j+k})}{\widetilde{\mathcal{V}}(X_j, X_j)} \tag{4.42}$$

as the *affinely invariant auto distance correlation* at the lag k. Similarly, given jointly strongly stationary, vector-valued time series $(X_j)_{j=-\infty}^{\infty}$ and $(Y_j)_{j=-\infty}^{\infty}$, we refer to

$$\widetilde{\mathcal{R}}_{X,Y}(k) = \frac{\widetilde{\mathcal{V}}(X_j, Y_{j+k})}{\sqrt{\widetilde{\mathcal{V}}(X_j, X_j)\widetilde{\mathcal{V}}(Y_j, Y_j)}} \tag{4.43}$$

as the *affinely invariant cross distance correlation* at the lag k. The corresponding sample versions can be defined in the natural way, as in the case of the non-affine distance correlation (Zhou, 2012).

We illustrate these concepts on time series data of wind observations at and near the Stateline wind energy center in the Pacific Northwest of the United States. Specifically, we consider time series of bivariate wind vectors at the meteorological towers at Vansycle, right at the Stateline wind farm at the border of the states of Washington and Oregon, and at Goodnoe Hills, 146 km west of Vansycle along the Columbia River Gorge. Further information can be found in the paper by Gneiting, et al. (2006), who developed a regime-switching space-time (RST) technique for 2-hour-ahead forecasts of hourly average wind speed at the Stateline wind energy center, which was then the largest wind farm globally. For our purposes, we follow Hering and Genton (2010) in studying the time series at the original 10-minute resolution, and we restrict our analysis to the longest continuous record, the 75-day interval from August 14 to October 28, 2002.

Thus, we consider time series of bivariate wind vectors over $10,800$ consecutive 10-minute intervals. We write V_j^{NS} and V_j^{EW} to denote the north-south and the east-west component of the wind vector at Vansycle at time j, with positive values corresponding to northerly and easterly winds. Similarly, we write G_j^{NS} and G_j^{EW} for the north-south and the east-west component of the wind vector at Goodnoe Hills at time j, respectively.

Figure 6 shows the classical (Pearson) sample auto and cross correlation functions for the four univariate time series. The auto correlation functions generally decay with the temporal, but do so non-monotonously, due to the presence of a diurnal component. The cross correlation functions between the wind vector components at Vansycle and Goodnoe Hills show remarkable asymmetries and peak at positive lags, due to the prevailing westerly and southwesterly wind (Gneiting, et al. 2006). In another interesting feature, the cross correlations between the north-south and east-west components at lag zero are strongly positive, documenting the dominance of southwesterly winds.

Figure 7 shows the sample auto and cross distance correlation functions for the four time series; as these variables are univariate, there is no distinction between the standard and the affinely invariant version of the distance correlation. The patterns seen resemble those in the case of the Pearson correlation. For comparison, we also

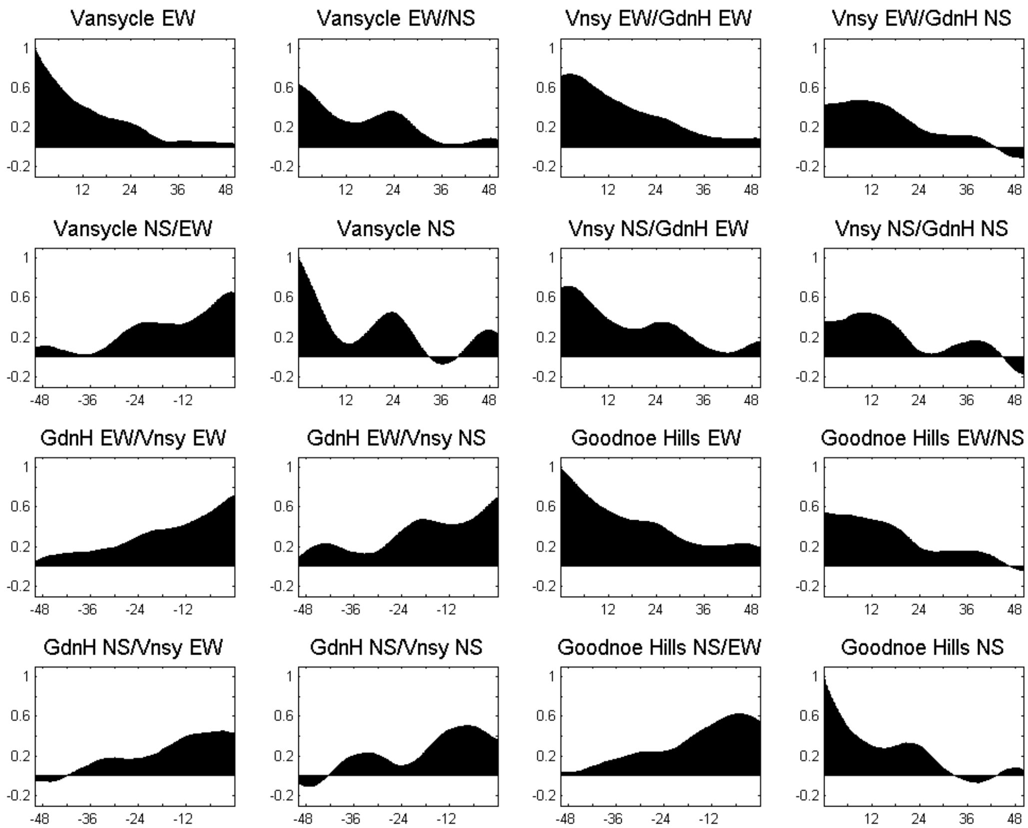

Figure 6: Sample auto and cross Pearson correlation functions for the univariate time series V_j^{EW}, V_j^{NS}, G_j^{EW}, and G_j^{NS}, respectively. Positive lags indicate observations at the westerly site (Goodnoe Hills) leading those at the easterly site (Vansycle), or observations of the north-south component leading those of the east-west component, in units of hours.

display values of the distance correlation based on the sample Pearson correlations shown in Figure 6, and converted to distance correlation under the assumption of bivariate Gaussianity, using the results of Székely, et al. (2007, p. 2786) and Section 4.2.3; in every single case, these values are smaller than the original ones.

Having considered the univariate time series setting, it is natural and complementary to look at the wind vector time series $(V_j^{\mathrm{EW}}, V_j^{\mathrm{NS}})$ at Vansycle and $(G_j^{\mathrm{EW}}, G_j^{\mathrm{NS}})$ at Goodnoe Hills from a genuinely multivariate perspective. To this end, Figure 8 shows the sample affinely invariant auto and cross distance correlation functions for the bivariate wind vector series at the two sites. Again, a diurnal component is visible, and there is a remarkable asymmetry in the cross-correlation functions, which peak at lags of about two to three hours.

In light of our analytical results in Section 4.2.3, we can compute the affinely invariant distance correlation between subvectors of a multivariate normally distributed random vector. In particular, we can compute the affinely invariant auto and cross

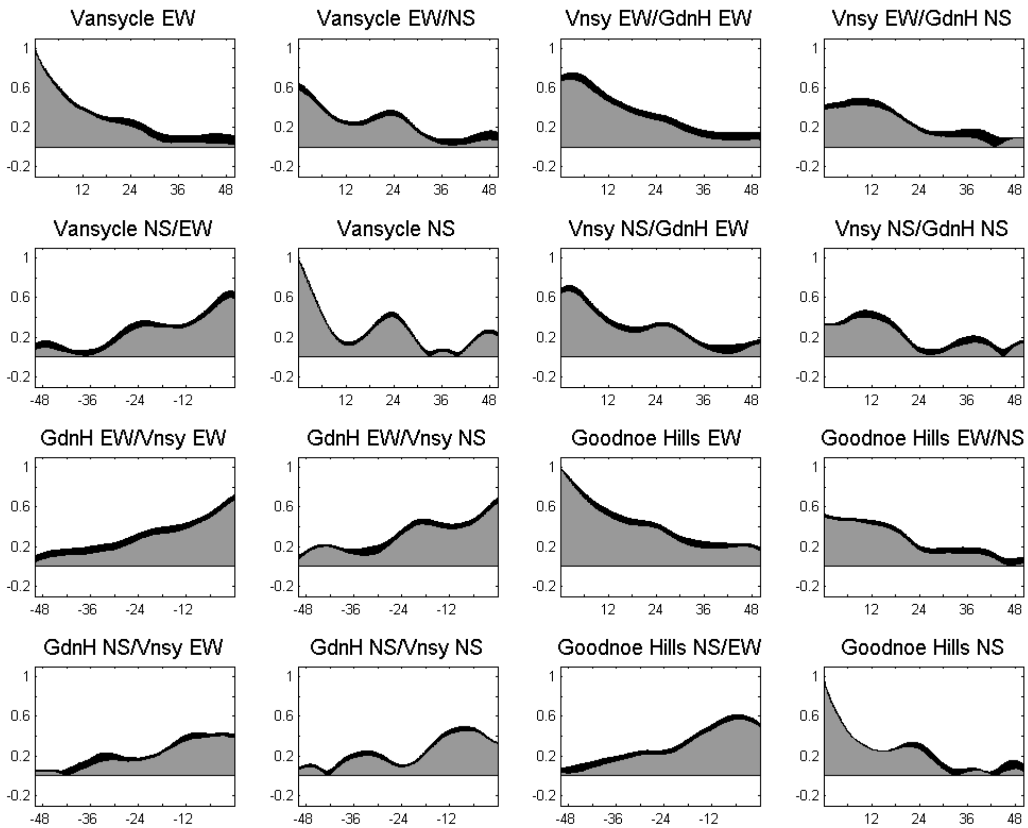

Figure 7: Sample auto and cross distance correlation functions for the univariate time series V_j^{EW}, V_j^{NS}, G_j^{EW}, and G_j^{NS}, respectively. For comparison, we also display, in grey, the values that arise when the sample Pearson correlations in Figure 6 are converted to distance correlation under the assumption of Gaussianity; these values generally are smaller than the original ones. Positive lags indicate observations at Goodnoe Hills leading those at Vansycle, or observations of the north-south component leading those of the east-west component, in units of hours.

distance correlation between bivariate subvectors of a 4-variate Gaussian process with Pearson auto and cross correlations as shown in Figure 6. In Figure 8, values of the affinely invariant distance correlation that have been derived from Pearson correlations in these ways are shown in grey; the differences from those values that are computed directly from the data are substantial, with the converted values being smaller, possibly suggesting that assumptions of Gaussianity may not be appropriate for this particular data set.

We wish to emphasize that our study is purely exploratory: it is provided for illustrative purposes and to serve as a basic example. In future work, the approach hinted at here may have the potential to be developed into parametric or nonparametric bootstrap tests for Gaussianity. For this purpose recall that, in the Gaussian setting, the affinely invariant distance correlation is a function of the canonical correlation

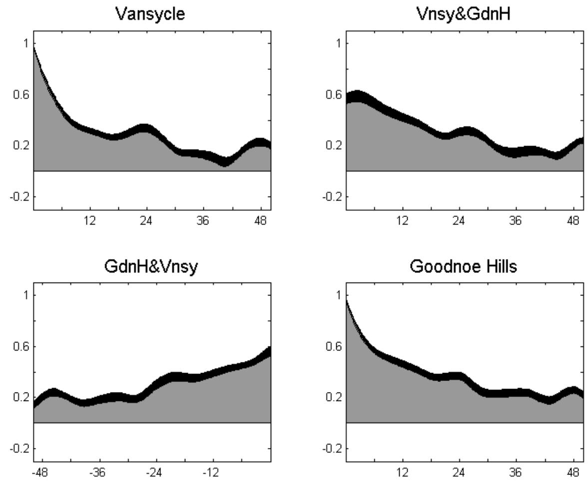

Figure 8: Sample auto and cross affinely invariant distance correlation functions for the bivariate time series $(V_j^{\mathrm{EW}}, V_j^{\mathrm{NS}})'$ and $(G_j^{\mathrm{EW}}, G_j^{\mathrm{NS}})'$ at Vansycle and Goodnoe Hills. For comparison, we also display, in grey, the values that are generated when the Pearson correlation in Figure 6 is converted to the affinely invariant distance correlation under the assumption of Gaussianity; these converted values generally are smaller than the original ones. Positive lags indicate observations at Goodnoe Hills leading those at Vansycle, in units of hours.

coefficients, i.e. $\widetilde{\mathcal{R}} = g(\lambda_1, \ldots, \lambda_r)$. For a parametric bootstrap test, one could generate B replicates of $g(\lambda_1^\star, \ldots, \lambda_r^\star)$, leading to a pointwise $(1-\alpha)$-confidence band. The test would now reject Gaussianity if the sample affinely invariant distance correlation function does not lie within this band. For the nonparametric bootstrap test, one could obtain ensembles $\widetilde{\mathcal{R}}_n^\star$ by resampling methods, again defining a pointwise $(1-\alpha)$-confidence band and checking if $g(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_r)$ is located within this band.

Following the pioneering work of Zhou (2012), the distance correlation may indeed find a wealth of applications in exploratory and inferential problems for time series data.

We proceed with the calculation of regular distance correlation in the Gaussian case.

4.3 Distance Correlation for Multivariate Normal Populations

In Theorem 4.4 and Corollary 4.5 we calculated the affinely invariant distance covariance for multivariate normal populations. Here, we consider the problem of deriving a formula for the standard distance covariance and distance correlation.

We first consider the case in which Σ_X and Σ_Y are scalar matrices, say, $\Sigma_X = \sigma_x^2 I_p$ and $\Sigma_Y = \sigma_y^2 I_q$ with $\sigma_x, \sigma_y > 0$. Thus, suppose that $(X, Y) \sim \mathcal{N}_{p+q}(\mu, \Sigma)$, where

$$\Sigma = \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix} = \begin{pmatrix} \sigma_x^2 I_p & \Sigma_{XY} \\ \Sigma_{YX} & \sigma_y^2 I_q \end{pmatrix}.$$

Putting $\Lambda = \Sigma_{YX}\Sigma_{XY}$, we follow the proofs of Theorem 4.4 and Corollary 4.5 to obtain

$$\begin{aligned}
\mathcal{V}^2(X,Y) &= 4\pi \frac{c_{p-1}}{c_p}\frac{c_{q-1}}{c_q}\sum_{k=1}^{\infty} \frac{2^{2k}-2}{k!\,2^{2k}} \frac{(\frac{1}{2})_k\,(-\frac{1}{2})_k\,(-\frac{1}{2})_k}{(\frac{1}{2}p)_k(\frac{1}{2}q)_k} \frac{1}{(\sigma_x\sigma_y)^{2k-1}}\,C_{(k)}(\Lambda) \\
&= 4\pi\sigma_x\sigma_y \frac{c_{p-1}}{c_p}\frac{c_{q-1}}{c_q}\bigg(\Big[{}_3F_2\big(\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{1}{2};\tfrac{1}{2}p,\tfrac{1}{2}q;\Lambda/\sigma_x^2\sigma_y^2\big)-1\Big] \\
&\qquad\qquad -2\Big[{}_3F_2\big(\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{1}{2};\tfrac{1}{2}p,\tfrac{1}{2}q;\Lambda/4\sigma_x^2\sigma_y^2\big)-1\Big]\bigg).
\end{aligned}$$

Next we reduce the general case to the scalar case above. By making a diagonal transformation of the form (4.6) we see that we may assume, without loss of generality, that Σ_X and Σ_Y are diagonal matrices. Now denote by σ_x^2 and σ_y^2 the smallest eigenvalues of Σ_X and Σ_Y, respectively. Also, let $\Lambda_X = \Sigma_X - \sigma_x^2 I_p$ and $\Lambda_Y = \Sigma_Y - \sigma_y^2 I_q$; then, $\Sigma_X = \Lambda_X + \sigma_x^2 I_p$ and $\Sigma_Y = \Lambda_Y + \sigma_y^2 I_q$. Substituting these decompositions into the integral which defines $\mathcal{V}^2(X,Y)$, we obtain

$$\begin{aligned}
&\int_{\mathbb{R}^{p+q}} \big(1-\exp(s'\Sigma_{XY}t)\big)^2 \exp(-s'\Sigma_X s - t'\Sigma_Y t)\,\frac{\mathrm{d}s}{|s|_p^{p+1}}\,\frac{\mathrm{d}t}{|t|_q^{q+1}} \\
&= \int_{\mathbb{R}^{p+q}} \big(1-\exp(s'\Sigma_{XY}t)\big)^2 \exp(-s'\Lambda_X s - t'\Lambda_Y t)\exp(-\sigma_x^2|s|_p^2 - \sigma_y^2|t|_q^2)\,\frac{\mathrm{d}s}{|s|_p^{p+1}}\,\frac{\mathrm{d}t}{|t|_q^{q+1}}.
\end{aligned}$$

Next, we apply a Taylor expansion,

$$\big(1-\exp(s'\Sigma_{XY}t)\big)^2 = \sum_{k=2}^{\infty} \frac{2^k-2}{k!}(s'\Sigma_{XY}t)^k$$

and, writing $\Lambda_X = \operatorname{diag}(\lambda_{x1},\ldots,\lambda_{xp})$, we have

$$\begin{aligned}
\exp(-s'\Lambda_X s) &= \sum_{l=0}^{\infty} \frac{(-1)^l}{l!}(s'\Lambda_X s)^l \\
&= \sum_{l=0}^{\infty} \frac{(-1)^l}{l!}(\lambda_{x1}s_1^2+\cdots+\lambda_{xp}s_p^2)^l \\
&= \sum_{l=0}^{\infty} \frac{(-1)^l}{l!} \sum_{l_1+\cdots+l_p=l} \binom{l}{l_1,\ldots,l_p} \prod_{i=1}^{p} \lambda_{xi}^{l_i} s_i^{2l_i}.
\end{aligned}$$

Similarly, on writing $\Lambda_Y = \operatorname{diag}(\lambda_{y1},\ldots,\lambda_{yq})$, we obtain

$$\exp(-t'\Lambda_Y t) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!} \sum_{m_1+\cdots+m_q=m} \binom{m}{m_1,\ldots,m_q} \prod_{j=1}^{q} \lambda_{yj}^{m_j} t_j^{2m_j}.$$

Integrating these series term-by-term, we find that the typical integral to be evaluated is

$$\int_{\mathbb{R}^{p+q}} (s'\Sigma_{XY}t)^k \prod_{i=1}^{p} s_i^{2l_i} \prod_{j=1}^{q} t_j^{2m_j} \exp(-\sigma_x^2|s|_p^2 - \sigma_y^2|t|_q^2)\,\frac{\mathrm{d}s}{|s|_p^{p+1}}\,\frac{\mathrm{d}t}{|t|_q^{q+1}}.$$

By the substitution $t \mapsto -t$, we find that this integral vanishes if k is odd, and so we need to calculate

$$\int_{\mathbb{R}^{p+q}} (s'\Sigma_{XY}t)^{2k} \prod_{i=1}^{p} s_i^{2l_i} \prod_{j=1}^{q} t_j^{2m_j} \exp(-\sigma_x^2|s|_p^2 - \sigma_y^2|t|_q^2) \, \frac{\mathrm{d}s}{|s|_p^{p+1}} \, \frac{\mathrm{d}t}{|t|_q^{q+1}}.$$

We transform to polar coordinates $s = r_x\theta$ and $t = r_y\phi$, where $r_x, r_y > 0$, $\theta \in S^{p-1}$, and $\phi \in S^{q-1}$. Then the integrals over r_x and r_y are standard gamma integrals:

$$\int_0^\infty \int_0^\infty r_x^{2k+2l_{\bullet}-2} r_y^{2k+2m_{\bullet}-2} \exp(-\sigma_x^2 r_x^2 - \sigma_y^2 r_y^2) \, \mathrm{d}r_x \mathrm{d}r_y = \frac{\Gamma(k+l_{\bullet}-\frac{1}{2})\,\Gamma(k+m_{\bullet}-\frac{1}{2})}{4\,\sigma_x^{2k+2l_{\bullet}-1}\,\sigma_y^{2k+2m_{\bullet}-1}},$$

where $l_{\bullet} = l_1 + \cdots + l_p$ and $m_{\bullet} = m_1 + \cdots + m_q$. As for the integrals over θ and ϕ, they are

$$\int_{S^{q-1}} \int_{S^{p-1}} (\theta'\Sigma_{XY}\phi)^{2k} \prod_{i=1}^{p} \theta_i^{2l_i} \prod_{j=1}^{q} \phi_j^{2m_j} \, \mathrm{d}\theta \, \mathrm{d}\phi.$$

To evaluate these integrals, we expand $(\theta'\Sigma_{XY}\phi)^{2k}$ using the multinomial theorem, obtaining a sum of terms, each of which is homogeneous in θ and ϕ. Then we integrate term-by-term by transforming the surface measures $\mathrm{d}\theta$ and $\mathrm{d}\phi$ to Euler angles (Anderson, 2003, pp. 285–286). The outcome is a multiple series expansion for the distance covariance. It does not appear to be a series that can be made simple in the general case, but it does provide an explicit expression in terms of Σ, p, and q.

Although we chose σ_x^2 and σ_y^2 to be the smallest eigenvalues of Σ_X and Σ_Y, respectively, we could have chosen them to be any positive numbers. This is reminiscent of the comprehensive work of Kotz, Johnson, and Boyd (1967a, 1967b) on the distribution of positive definite quadratic forms in normal variables. Bearing in mind those results, it seems likely that an optimal choice for σ_x^2 and σ_y^2 will be close to the arithmetic, geometric, or harmonic mean of the eigenvalues of Σ_X and Σ_Y, respectively. At least, the issue of optimal choices for σ_x^2 and σ_y^2 that will accelerate the convergence of the above series is worthy of further investigation.

Finally, we note that our techniques allow for similar explicit expressions in the case of the α-distance dependence measures described by Székely, et al. (2007, p. 2784) and Székely and Rizzo (2009, pp. 1251–1252; 2012, p. 2282).

Computing the analytic distance correlation for general distributions is a nontrivial task. However, we demonstrate a direct computation via the definition of affinely invariant distance correlation once more: for the multivariate Laplace distribution.

4.4 Affinely Invariant Distance Correlation for the Multivariate Laplace Distribution

Now let $(X,Y) \sim L_{p+q}(\Sigma)$, i.e.

$$f_{X,Y}(s,t) = \left(1 + \tfrac{1}{2}\begin{pmatrix} s \\ t \end{pmatrix}' \Sigma \begin{pmatrix} s \\ t \end{pmatrix}\right)^{-1},$$

where $f_{X,Y}$ is the characteristic function of (X,Y). Hence, the characteristic functions of the marginals are

$$f_X(s) = \Big(1+\tfrac{1}{2}s'\Sigma_X s\Big)^{-1} \quad \text{and} \quad f_Y(t) = \Big(1+\tfrac{1}{2}t'\Sigma_Y t\Big)^{-1},$$

respectively. Therefore, the affinely invariant distance covariance between X and Y can be computed as

$$c_p c_q \widetilde{\mathcal{V}}(X,Y) = \int_{\mathbb{R}^{p+q}} \left| \left(1+\tfrac{1}{2}\begin{pmatrix} s \\ t \end{pmatrix}' \Sigma \begin{pmatrix} s \\ t \end{pmatrix}\right)^{-1} - \left(1+\tfrac{1}{2}s'\Sigma_X s\right)^{-1}\left(1+\tfrac{1}{2}t'\Sigma_Y t\right)^{-1} \right|^2 \\ \times \frac{\sqrt{|\Sigma_X|}\,ds\,\sqrt{|\Sigma_Y|}\,dt}{(s'\Sigma_X s)^{(p+1)/2}(t'\Sigma_Y t)^{(q+1)/2}}.$$

By substituting $u = \sqrt{1/2}\,\Sigma_X^{1/2}s$ and $v = \sqrt{1/2}\,\Sigma_Y^{1/2}t$ we obtain for the latter integral

$$2\int_{\mathbb{R}^{p+q}} \Big| \big(1+u'u+v'v+2u'\,\Sigma_X^{-\frac{1}{2}}\,\Sigma_{XY}\,\Sigma_Y^{-\frac{1}{2}}\,v\big)^{-1} \\ - \big(1+u'u\big)^{-1}\big(1+v'v\big)^{-1}\Big|^2 \frac{du\,dv}{(u'u)^{(p+1)/2}(v'v)^{(q+1)/2}}.$$

Now we change variables to polar coordinates, putting $u = r_1\theta$ and $v = r_2\phi$ where $r_1, r_2 > 0$, $\theta = (\theta_1,\ldots,\theta_p)' \in S^{p-1}$, and $\phi = (\phi_1,\ldots,\phi_q)' \in S^{q-1}$. With $\Lambda := \Sigma_X^{-\frac{1}{2}}\,\Sigma_{XY}\,\Sigma_Y^{-\frac{1}{2}}$ the integral is equal to

$$2\int_{S^{p-1}\times S^{q-1}}\int_{\mathbb{R}_+\times\mathbb{R}_+} \Big| \big(1+r_1^2+r_2^2+2r_1r_2\,\theta'\,\Lambda\,\phi\big)^{-1} \\ - \big(1+r_1^2\big)^{-1}\big(1+r_2^2\big)^{-1}\Big|^2 \frac{dr_1\,dr_2\,d\theta\,d\phi}{r_1^2 r_2^2}.$$

Again substituting $u = r_1^2$ and $v = r_2^2$ the latter integral equals

$$\frac{1}{2}\int_{S^{p-1}\times S^{q-1}}\int_{\mathbb{R}_+\times\mathbb{R}_+} \Big| \big(1+u+v+2\sqrt{uv}\,\theta'\,\Lambda\,\phi\big)^{-1} \\ - \big(1+u\big)^{-1}\big(1+v\big)^{-1}\Big|^2 \frac{du\,dv\,d\theta\,d\phi}{u^{3/2}\,v^{3/2}}.$$

Furthermore, we change coordinates to $s = \frac{u}{1+u}$ and $t = \frac{v}{1+v}$. Observing that $1+u = \frac{1}{1-s}$, $1+v = \frac{1}{1-t}$ and

$$1+u+v+2\sqrt{uv}\,\theta'\,\Lambda\,\phi = \frac{1-st+2\,\theta'\Lambda\phi\sqrt{st}\,\sqrt{(1-s)(1-t)}}{(1-s)(1-t)}$$

the inner integral transforms to

$$\int_{[0,1]\times[0,1]} \Big| \Big(1-st+2\theta'\,\Lambda\,\phi\,\sqrt{st}\;\sqrt{(1-s)(1-t)}\Big)^{-1} - 1\Big|^2 \left(\frac{(1-s)\,(1-t)}{s\,t}\right)^{3/2} ds\,dt.$$

By expanding into negative binomial series, we obtain

$$\begin{aligned}
&\left|\left(1 - st + 2\,\theta'\,\Lambda\,\phi\,\sqrt{st}\;\sqrt{(1-s)(1-t)}\right)^{-1} - 1\right|^2 \\
&\qquad= \left(1 - st + 2\,\theta'\,\Lambda\,\phi\,\sqrt{st}\;\sqrt{(1-s)(1-t)}\right)^{-2} \\
&\qquad\quad - 2\left(1 - st + 2\,\theta'\,\Lambda\,\phi\,\sqrt{st}\;\sqrt{(1-s)(1-t)}\right)^{-1} + 1 \\
&\qquad= \sum_{k=2}^{\infty}(k-1)\left(st - 2\,\theta'\Lambda\phi\sqrt{st}\sqrt{(1-s)(1-t)}\right)^k.
\end{aligned}$$

Moreover, by expanding into binomial series, the latter term reads

$$\sum_{k=2}^{\infty}(k-1)\sum_{i=0}^{k}\binom{k}{i}(st)^{k-i}(-1)^i\,(2\,\theta'\Lambda\phi\sqrt{st}\sqrt{(1-s)(1-t)})^i.$$

Hence,

$$\begin{aligned}
\widetilde{\mathcal{V}}(X,Y) =& \frac{1}{2c_pc_q}\sum_{k=2}^{\infty}(k-1)\sum_{i=0}^{k}\binom{k}{i}(-1)^i\left(\int_0^1 s^{k-i-3/2}(1-s)^{(i+3)/2}ds\right)^2 \\
&\times\int_{S^{p-1}\times S^{q-1}}(2\,\theta'\Lambda\phi)^i d\theta\,d\phi.
\end{aligned}$$

Since $\int_{S^{p-1}\times S^{q-1}}(2\,\theta'\Lambda\phi)^i d\theta\,d\phi$ vanishes for i odd, this can be written as

$$\begin{aligned}
\widetilde{\mathcal{V}}(X,Y) =& \frac{1}{2c_pc_q}\sum_{k=2}^{\infty}(k-1)\sum_{j=0}^{\lfloor\frac{k}{2}\rfloor}\binom{k}{2j}\left(\int_0^1 s^{k-j-3/2}(1-s)^{j+3/2}ds\right)^2 \\
&\times\int_{S^{p-1}\times S^{q-1}}(2\,\theta'\Lambda\phi)^{2j} d\theta\,d\phi.
\end{aligned}$$

The integral with respect to s is a standard beta integral

$$\int_0^1 s^{k-j-3/2}(1-s)^{j+3/2}ds = B\left(k-j-\frac{1}{2}, j+\frac{5}{2}\right),$$

where B is the beta function. Moreover the integral with respect to the spheres is readily evaluated in (4.28) as

$$4c_{p-1}c_{q-1}\frac{(\frac{1}{2})_j\,(\frac{1}{2})_j}{(\frac{1}{2}p)_j(\frac{1}{2}q)_j}C_{(j)}(\Lambda),$$

where $(\alpha)_j$ denotes the rising factorial and $C_{(j)}(\cdot)$ is the top order zonal polynomial with weight j. As a result, we finally find

$$\widetilde{\mathcal{V}}(X,Y) = 2\,\frac{c_{p-1}\,c_{q-1}}{c_p\,c_q}\sum_{k=2}^{\infty}(k-1)\sum_{j=0}^{\lfloor\frac{k}{2}\rfloor}2^{2j}\binom{k}{2j}B\left(k-j-\frac{1}{2}, j+\frac{5}{2}\right)^2\frac{(\frac{1}{2})_j\,(\frac{1}{2})_j}{(\frac{1}{2}p)_j(\frac{1}{2}q)_j}C_{(j)}(\Lambda).$$

In the special case $\Sigma = I_{p+q}$ affinely invariant distance correlation between X and Y reduces to

$$2\,\frac{c_{p-1}\,c_{q-1}}{c_p\,c_q}\sum_{k=2}^{\infty}(k-1)\,B\left(k-\frac{1}{2},\frac{5}{2}\right)^2 > 0,$$

which is a strictly positive constant.

4.5 Distance Correlation for Lancaster Distributions

In this section (see Dueck, Edelmann and Richards; 2014) we introduce a large class of distributions, the so-called Lancaster distributions. We provide a formula to compute distance correlation for distributions within this class. In order to prove its relevance we apply the formula to several exemplary distributions. As in the previous sections we once more compute distance correlation and affinely invariant distance correlation for the normal distribution in order to confirm our results and to show how the method simplifies computations, e.g. no elegant tricks are needed. Moreover, we are able to calculate distance correlation for non-Gaussian random variables: for the bivariate gamma distribution.

4.5.1 The Lancaster distributions

To recapitulate the class of Lancaster distributions we will use the standard notation in that area, as given by Koudou (1996 and 1998); cf., Lancaster (1969), Pommeret (2004), or Diaconis, et al. (2008).

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 σ be a probability measure on $\mathcal{X} \times \mathcal{Y}$ such that σ has marginal distributions μ and ν, respectively; then there exist functions K_σ and L_σ such that

$$\sigma(\,\mathrm{d}x,\,\mathrm{d}y) = K_\sigma(x,\,\mathrm{d}y)\mu(\,\mathrm{d}x) = L_\sigma(\,\mathrm{d}x, y)\nu(\,\mathrm{d}y).$$

We note that K_σ and L_σ represent the conditional distributions of Y given $X = x$, and X given $Y = y$, respectively.

Let $\mathbb{N}_0$ denote the set of nonnegative integers. Let $\{P_n : n \in \mathbb{N}_0\}$ be a sequence of functions on $\mathcal{X}$ which forms an orthonormal basis for the Hilbert space $L^2(\mu)$; and similarly, let $\{Q_n : n \in \mathbb{N}_0\}$ be a sequence of functions on $\mathcal{Y}$ which forms an orthonormal basis for $L^2(\nu)$. We suppose that $P_0(x) \equiv 1$ and $Q_0(y) \equiv 1$.

Suppose that $\sigma \in L^2(\mu \otimes \nu)$. Then there holds the expansion

$$\sigma(\,\mathrm{d}x,\,\mathrm{d}y) = \sum_{m\in\mathbb{N}_0}\sum_{n\in\mathbb{N}_0}\rho_{m,n}P_m(x)Q_n(y)\,\mu(\,\mathrm{d}x)\,\nu(\,\mathrm{d}y), \tag{4.44}$$

$(x, y) \in \mathcal{X} \times \mathcal{Y}$. The probability measure σ is called a *Lancaster distribution* if there exists a positive sequence $\{\rho_n : n \in \mathbb{N}_0\}$ such that

$$\int P_m(x)\,Q_n(y)\,\sigma(\,\mathrm{d}x,\,\mathrm{d}y) = \rho_m\,\delta_{m,n}$$

for all n and m, where $\delta_{m,n}$ denotes Kronecker's delta. In this case, the sequence $\{\rho_n : n \in \mathbb{N}_0\}$ is called a *Lancaster sequence*, and the expansion (4.44) reduces to

$$\sigma(\,\mathrm{d}x,\,\mathrm{d}y) = \sum_{n\in\mathbb{N}_0} \rho_n P_n(x) Q_n(y) \mu(\,\mathrm{d}x) \nu(\,\mathrm{d}y).$$

Koudou (1996, pp. 255–256) characterized the Lancaster sequences $\{\rho_n : n \in \mathbb{N}_0\}$ such that the associated probability distribution σ is absolutely continuous with respect to $\mu \otimes \nu$ and has Radon-Nikodym derivative

$$\frac{\sigma(\,\mathrm{d}x,\,\mathrm{d}y)}{\mu(\,\mathrm{d}x)\,\nu(\,\mathrm{d}y)} = \sum_{n\in\mathbb{N}_0} \rho_n\, P_n(x)\, Q_n(y) \in L^2(\mu \otimes \nu),$$

$(x, y) \in \mathcal{X} \times \mathcal{Y}$.

In the sequel, we are particularly interested in the case in which $\mathcal{X} = \mathbb{R}^p$ and $\mathcal{Y} = \mathbb{R}^q$. Then, the underlying random vectors $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ have joint distribution σ and marginal distributions μ and ν, respectively. We assume that μ, ν, and σ 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 ϕ_X, ϕ_Y, and $\phi_{X,Y}$, respectively. This yields the simplified expansion,

$$\phi_{X,Y}(x, y) = \phi_X(x)\, \phi_Y(y) \sum_{n\in\mathbb{N}_0} \rho_n\, P_n(x)\, Q_n(y),$$

equivalently,

$$\phi_{X,Y}(x, y) - \phi_X(x)\, \phi_Y(y) = \phi_X(x)\, \phi_Y(y) \sum_{n\neq 0} \rho_n\, P_n(x)\, Q_n(y). \tag{4.45}$$

We will refer to the latter expansion as the *Lancaster expansion* of the joint probability density function $\phi_{X,Y}$.

4.5.2 Examples of Lancaster expansions

Here, we provide examples of Lancaster expansions (4.45) for the bivariate and multivariate normal distributions, and for the bivariate gamma, Poisson, and negative binomial distributions.

4.5.2.1 The bivariate normal distribution

Let (X, Y) follow a bivariate normal distribution, denoted $(X, Y) \sim \mathcal{N}_2(0, \Sigma)$, where

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

with joint probability density function

$$\phi_{X,Y}(x, y) = \frac{1}{2\pi}(1 - \rho^2)^{-\frac{1}{2}} \exp\left(-\frac{x^2 + y^2 - 2\rho\, x\, y}{2(1 - \rho^2)}\right),$$

and with standard normal marginal distributions

$$\phi_X(x) \equiv \phi_Y(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}x^2\right),$$

$x, y \in \mathbb{R}$. Let

$$H_n(x) = (-1)^n \exp\left(\tfrac{1}{2}x^2\right) \frac{\mathrm{d}^n}{\mathrm{d}x^n} \exp(-x^2/2),$$

$x \in \mathbb{R}$, denote the nth Hermite polynomial, $n = 0, 1, 2, \ldots$. It is well-known that the Hermite polynomials are orthogonal with respect to the standard normal distribution.

The Lancaster expansion of $\phi_{X,Y}$ is given by the classical formula of Mehler,

$$\phi_{X,Y}(x, y) = \phi_X(x)\, \phi_Y(y) \left[1 + \sum_{n=1}^{\infty} \frac{\rho^n}{n!} H_n(x)\ H_n(y)\right], \tag{4.46}$$

$x, y \in \mathbb{R}$.

We remark that there are several extensions of Mehler's formula which can be interpreted as Lancaster expansions for generalizations of the bivariate normal distribution; cf. Srivastava and Singhal (1972) for one such expansion. However, we will not study these generalizations here because the developments are similar to the results which we obtain.

4.5.2.2 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 multivariate normal distribution with mean vector 0 and positive definite covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix}$$

where Σ_X, Σ_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 ϕ_X and ϕ_Y the marginal density functions of X and Y, respectively. Withers and Nadarajah (2010) derived Lancaster-type expansions for $\phi_{X,Y}$ under various assumptions on Σ, e.g., for partitioned covariance matrices or integrated versions.

We now describe the Lancaster expansion of $\phi_{X,Y}$, a result which is provided in Theorem 3.1, p. 1314 (Withers and Nadarajah; 2010). In order to state this result, we introduce additional notation drawn from Withers and Nadarajah (2010).

For each $j = 1, \ldots, p$ and $k = 1, \ldots, q$, let $N_{jk} \in \mathbb{N}_0$ be an index of summation. Let $\boldsymbol{N} = (N_{jk} : 1 \le j \le p,\ 1 \le k \le q)$ be the $p \times q$ matrix of summation indices and let $\boldsymbol{N}! = \prod_{j=1}^{p} \prod_{k=1}^{q} N_{jk}!$. Define, for $j = 1, \ldots, p$,

$$\boldsymbol{A}_{\boldsymbol{N},j} = \sum_{k=1}^{q} N_{jk},$$

and set $\boldsymbol{A}_{\boldsymbol{N}} = (\boldsymbol{A}_{\boldsymbol{N},1}, \ldots, \boldsymbol{A}_{\boldsymbol{N},p})$. Similarly, define, for $k = 1, \ldots, q$,

$$\boldsymbol{B}_{\boldsymbol{N},k} = \sum_{j=1}^{p} N_{jk},$$

and then set $\boldsymbol{B}_{\boldsymbol{N}} = (\boldsymbol{B}_{\boldsymbol{N},1}, \ldots, \boldsymbol{B}_{\boldsymbol{N},q})$. Denoting by $(\Sigma_{XY})_{jk}$ the (j,k)th entry of Σ_{XY}, we also define

$$\Sigma_{XY}^{\boldsymbol{N}} = \prod_{j=1}^{p} \prod_{k=1}^{q} [(\Sigma_{XY})_{jk}]^{N_{jk}}.$$

We also need the multivariate Hermite polynomials. For $\boldsymbol{k} = (k_1, \ldots, k_p)$, a p-dimensional vector of nonnegative integers, and $x \in \mathbb{R}^p$, define the differential operator,

$$\Big(-\frac{\partial}{\partial x}\Big)^{\boldsymbol{k}} = \Big(-\frac{\partial}{\partial x_1}\Big)^{k_1} \cdots \Big(-\frac{\partial}{\partial x_p}\Big)^{k_p}.$$

Then the *multivariate Hermite polynomial* with respect to the marginal density function ϕ_X is defined as

$$H_{\boldsymbol{k}}(x; \Sigma_X) = \frac{1}{\phi_X(x)} \Big(-\frac{\partial}{\partial x}\Big)^{\boldsymbol{k}} \phi_X(x).$$

Then, the Lancaster expansion of $\phi_{X,Y}$ is given by the generalized Mehler formula:

$$\phi_{X,Y}(x,y) = \phi_X(x)\, \phi_Y(y) \Big[1 + \sum_{\boldsymbol{N} \neq \boldsymbol{0}} \frac{\Sigma_{XY}^{\boldsymbol{N}}}{\boldsymbol{N}!} H_{\boldsymbol{A}_{\boldsymbol{N}}}(x; \Sigma_X)\, H_{\boldsymbol{B}_{\boldsymbol{N}}}(y; \Sigma_Y)\Big], \tag{4.47}$$

with absolute convergence for all $x \in \mathbb{R}^p$, $y \in \mathbb{R}^q$.

In order to calculate the affinely invariant distance correlation between X and Y we also need the Lancaster expansion of the joint density function of the standardized variables $\widetilde{X} = \Sigma_X^{-1/2} X$ and $\widetilde{Y} = \Sigma_Y^{-1/2} Y$. We deduce from (4.47) that the Lancaster expansion for $(\widetilde{X}, \widetilde{Y})$ is given by

$$\phi_{\widetilde{X},\widetilde{Y}}(x,y) = \phi_{\widetilde{X}}(x)\, \phi_{\widetilde{Y}}(y) \Big[1 + \sum_{\boldsymbol{N} \neq \boldsymbol{0}} \frac{\Lambda_{XY}^{\boldsymbol{N}}}{\boldsymbol{N}!} H_{\boldsymbol{A}_{\boldsymbol{N}}}(x; I_p)\, H_{\boldsymbol{B}_{\boldsymbol{N}}}(y; I_q)\Big], \tag{4.48}$$

where

$$\begin{aligned}\Lambda &= \begin{pmatrix} \Sigma_X^{-1/2} & 0 \\ 0 & \Sigma_Y^{-1/2} \end{pmatrix} \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix} \begin{pmatrix} \Sigma_X^{-1/2} & 0 \\ 0 & \Sigma_Y^{-1/2} \end{pmatrix} \\ &\equiv \begin{pmatrix} I_p & \Lambda_{XY} \\ \Lambda_{XY}' & I_q \end{pmatrix},\end{aligned}$$

with

$$\Lambda_{XY} = \Sigma_X^{-1/2}\, \Sigma_{XY}\, \Sigma_Y^{-1/2}.$$

4.5.2.3 The bivariate gamma distribution

The Lancaster expansion for a bivariate gamma distribution, which was derived by Sarmanov (1970), can be stated as follows (cf., Kotz, et al.; 2000).

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

$$\begin{aligned} L_n^{(\alpha)}(x) &= \frac{1}{n!} x^{-\alpha} \exp(x) \Big(\frac{\mathrm{d}}{\mathrm{d}x}\Big)^n x^{n+\alpha} \exp(-x) \\ &= \frac{(\alpha+1)_n}{n!} \sum_{j=0}^{n} \frac{(-n)_j}{(\alpha+1)_j} \frac{x^j}{j!}, \end{aligned} \tag{4.49}$$

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

Let $\lambda \in (0,1)$, and let α and β satisfy $\alpha \geq \beta > 0$. Sarmanov (1970) derived for certain bivariate gamma random variables (X,Y) the joint probability density function,

$$\phi_{X,Y}(x,y) = \phi_X(x)\, \phi_Y(y) \left[1 + \sum_{n=1}^{\infty} a_n L_n^{(\alpha-1)}(x)\, L_n^{(\beta-1)}(y)\right], \tag{4.50}$$

$x, y > 0$, where

$$a_n = \lambda^n \left[\frac{(\beta)_n}{(\alpha)_n}\right]^{1/2} \tag{4.51}$$

$n = 0, 1, 2, \ldots$. 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 α and β, respectively.

We remark that if $\alpha = \beta$ then the density function (4.50) reduces to the Kibble-Moran bivariate gamma density function and $\operatorname{Corr}(X,Y) = \lambda$ (Kotz, et al., pp. 436–437; 2000).

4.5.3 Distance correlation coefficients for Lancaster distributions

We derive under mild conditions a general result which enables the calculation of distance correlation coefficients for general Lancaster distributions with density functions of the form (4.45). For $\phi_{X,Y}$ given by (4.45) and $n \neq 0$, we introduce the notation

$$\mathcal{L}_n^X(s) = \int_{\mathbb{R}^p} \exp(\mathrm{i}\, \langle s, x \rangle)\, \phi_X(x)\, P_n(x)\, \mathrm{d}x, \tag{4.52}$$

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

$$\mathcal{L}_n^Y(t) = \int_{\mathbb{R}^q} \exp(\mathrm{i}\, \langle t, y \rangle)\, \phi_Y(y)\, Q_n(y)\, \mathrm{d}y, \tag{4.53}$$

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

$$\begin{aligned}
|\mathcal{L}_n^X(s)|^2 &\equiv \left| \int_{\mathbb{R}^p} \exp(\mathrm{i}\langle s, x\rangle)\, [\phi_X(x)]^{1/2}\, [\phi_X(x)]^{1/2}\, P_n(x)\ \mathrm{d}x \right|^2 \\
&\le \Big(\int_{\mathbb{R}^p} |\exp(\mathrm{i}\langle s, x\rangle)\, [\phi_X(x)]^{1/2}|^2\ \mathrm{d}x \Big) \cdot \Big(\int_{\mathbb{R}^p} \big|[\phi_X(x)]^{1/2}\, P_n(x)\big|^2\ \mathrm{d}x \Big) \\
&= \Big(\int_{\mathbb{R}^p} \phi_X(x)\ \mathrm{d}x \Big) \cdot \Big(\int_{\mathbb{R}^p} \phi_X(x)\, |P_n(x)|^2\ \mathrm{d}x \Big) \\
&= 1,
\end{aligned}$$

because ϕ_X is a density function and because $\{P_n : n \in \mathbb{N}_0\}$ is an orthonormal basis for the Hilbert space $L^2(\mu)$. Similarly, $|\mathcal{L}_n^Y(t)| \le 1$ for all $t \in \mathbb{R}^q$.

We now state the main result.

Theorem 4.11. *Let X and Y be random vectors with values in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively, and with joint probability density function given by (4.45). Then,*

$$\mathcal{V}^2(X,Y) = \frac{1}{c_p c_q} \sum_{j\neq 0} \sum_{k\neq 0} \rho_j\, \bar{\rho}_k \int_{\mathbb{R}^p} \mathcal{L}_j^X(s)\, \mathcal{L}_k^X(-s)\, \frac{\mathrm{d}s}{|s|_p^{p+1}} \int_{\mathbb{R}^q} \mathcal{L}_j^Y(t)\, \mathcal{L}_k^Y(-t)\, \frac{\mathrm{d}t}{|t|_q^{q+1}}, \tag{4.54}$$

whenever the latter double sum converges absolutely.

Proof. By taking Fourier transforms on both sides of (4.45), we obtain for all $s \in \mathbb{R}^p$ and $t \in \mathbb{R}^q$ the identity

$$\psi_{X,Y}(s,t) - \psi_X(s)\, \psi_Y(t) = \sum_{n\neq 0} \rho_n \mathcal{L}_n^X(s)\, \mathcal{L}_n^Y(t), \tag{4.55}$$

subject to the requirement that we may interchange summation and integration; however, that interchange is justified by the assumption that the sum in the final result converges absolutely. Using the identity (4.55) we deduce, moreover, that

$$|\psi_{X,Y}(s,t) - \psi_X(s)\psi_Y(t)|^2 = \sum_{j\neq 0} \sum_{k\neq 0} \rho_j\, \bar{\rho}_k\, \mathcal{L}_j^X(s)\, \mathcal{L}_k^X(-s)\, \mathcal{L}_j^Y(t)\, \mathcal{L}_k^Y(-t).$$

Finally, to obtain a formula for $\mathcal{V}^2(X,Y)$, we need only interchange summation and integration once more, which we are allowed to do so by assumption. □

4.5.4 Examples

To display the versatility of Theorem 4.11, we apply that result to compute the distance correlation coefficients for several distributions, namely the bivariate normal, the multivariate normal, the bivariate gamma and certain bivariate discrete distributions. In each case, we retain the same notation as in Section 4.5.2.

4.5.4.1 The bivariate normal distribution

Theorem 4.12. *Let (X, Y) follow a bivariate normal distribution with correlation coefficient ρ. Then*

$$\mathcal{V}^2(X, Y) = \frac{4}{\pi} \sum_{r=1}^{\infty} \frac{\rho^{2r} (\frac{1}{2})^{2r} ((2r-3)!!)^2}{(2r)!} (2^{2r} - 2). \tag{4.56}$$

PROOF. Starting with the Lancaster expansion of the bivariate normal density function, as given in (4.46), and using the definitions of $\mathcal{L}_n^X$ and $\mathcal{L}_n^Y$ in (4.52) and (4.53), respectively, we obtain by substitution and integration-by-parts,

$$\begin{aligned} \mathcal{L}_n^X(s) = \mathcal{L}_n^Y(s) &= \int_{-\infty}^{\infty} \exp(\mathrm{i} sx) \frac{1}{\sqrt{2\pi}} \exp\big(-\tfrac{1}{2}x^2\big) H_n(x) \, \mathrm{d}x \\ &= (\mathrm{i}\, s)^n \exp\big(-\tfrac{1}{2}s^2\big), \end{aligned}$$

$s \in \mathbb{R}$. Using the notation

$$(2n-1)!! = \prod_{j=1}^{n} (2j-1) \equiv \frac{(2n)!}{2^n \, n!}$$

for the double factorial, we obtain

$$\begin{aligned} \int_{\mathbb{R}} \mathcal{L}_j^X(s) \, \mathcal{L}_k^X(-s) \, \frac{\mathrm{d}s}{s^2} &= (-1)^k \, \mathrm{i}^{j+k} \int_{-\infty}^{\infty} s^{j+k-2} \exp(-s^2) \, \mathrm{d}s, \\ &= \begin{cases} (-1)^k \, \mathrm{i}^{j+k} \, \pi^{1/2} \left(\frac{1}{2}\right)^{(j+k-2)/2} (k+j-3)!!, & \text{if } j+k \text{ is even} \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

since the latter integral is a moment of the $\mathcal{N}(0, \frac{1}{2})$ distribution. By Theorem 4.11, we obtain

$$\mathcal{V}^2(X, Y) = \frac{4}{\pi} \sum_{\substack{j,\, k > 0 \\ j+k \text{ even}}} \frac{\rho^{j+k}}{j! \, k!} \left(-\tfrac{1}{2}\right)^{j+k} \big((j+k-3)!!\big)^2.$$

To reduce the double-series (4.56) to a single series, let $j+k = 2r$ with $r \geq 1$. Then, (4.56) reduces to

$$\begin{aligned} \mathcal{V}^2(X, Y) &= \frac{4}{\pi} \sum_{r=1}^{\infty} \rho^{2r} (-\tfrac{1}{2})^{2r} ((2r-3)!!)^2 \sum_{\substack{j,k \geq 1 \\ j+k=2r}} \frac{1}{j! \, k!} \\ &= \frac{4}{\pi} \sum_{r=1}^{\infty} \rho^{2r} (\tfrac{1}{2})^{2r} \frac{((2r-3)!!)^2}{(2r)!} \sum_{j=1}^{2r-1} \frac{(2r)!}{j! \, (2r-j)!} \\ &= \frac{4}{\pi} \sum_{r=1}^{\infty} \rho^{2r} (\tfrac{1}{2})^{2r} \frac{((2r-3)!!)^2}{(2r)!} \left(\sum_{j=0}^{2r} \binom{2r}{j} - 2 \right) \\ &= \frac{4}{\pi} \sum_{r=1}^{\infty} \rho^{2r} (\tfrac{1}{2})^{2r} \frac{((2r-3)!!)^2}{(2r)!} (2^{2r} - 2). \end{aligned}$$

The absolute convergence of this series can be verified by comparison with a geometric series. Moreover, it is straightforward to verify that this series is identical with the result obtained by Székely, et al. (p. 2786; 2007). □

Once the distance covariance, $\mathcal{V}(X,Y)$ is obtained, we let $\rho \to 1-$ to obtain the distance variances $\mathcal{V}(X,X)$ and $\mathcal{V}(Y,Y)$; here, we are using the well-known result that if (X,Y) is bivariate normally distributed and $\rho = 1$ then $X = Y$, almost surely.

4.5.4.2 The multivariate normal distribution

Theorem 4.13. *Let (X,Y) be multivariate normally distributed, i.e. $(X,Y) \sim \mathcal{N}_{p+q}(0,\Sigma)$, where*

$$\Sigma = \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix}$$

with positive definite Σ_X and Σ_Y. Then

$$\begin{aligned}\widetilde{\mathcal{V}}^2(X,Y) = \frac{1}{4\,c_p\,c_q} \sum_{\boldsymbol{K}\neq\boldsymbol{0}} \sum_{\boldsymbol{J}\neq\boldsymbol{0}} \widetilde{S}(\boldsymbol{K},\boldsymbol{J})\, \widetilde{T}(\boldsymbol{K},\boldsymbol{J})\, (-1)^{|\boldsymbol{A_J}|+|\boldsymbol{B_J}|}\, \mathrm{i}^{|\boldsymbol{A_K}|+|\boldsymbol{A_J}|+|\boldsymbol{B_K}|+|\boldsymbol{B_J}|} \\ \times \Gamma\left((|\boldsymbol{A_K}|+|\boldsymbol{A_J}|-1)/2\right)\, \Gamma\left((|\boldsymbol{B_K}|+|\boldsymbol{B_J}|-1)/2\right)\, \frac{\Lambda_{XY}^{\boldsymbol{K}}}{\boldsymbol{K}!} \frac{\Lambda_{XY}^{\boldsymbol{J}}}{\boldsymbol{J}!},\end{aligned}$$

where

$$\begin{aligned}\widetilde{S}(\boldsymbol{K},\boldsymbol{J}) = \frac{1}{\sqrt{\pi}} \frac{\left(\sum_{\nu=1}^{p}(\boldsymbol{A_K}+\boldsymbol{A_J})_\nu\right)!}{\left(\prod_{\nu=1}^{p}(\boldsymbol{A_K}+\boldsymbol{A_J})_\nu!\right)} \frac{\left(\prod_{\nu=1}^{p}[2(\boldsymbol{A_K}+\boldsymbol{A_J})_\nu]!\right)}{\left(\sum_{\nu=1}^{p}2(\boldsymbol{A_K}+\boldsymbol{A_J})_\nu\right)!} \\ \times \frac{\Gamma\left(|\boldsymbol{A_K}|+|\boldsymbol{A_J}|+\frac{1}{2}\right)\,\Gamma\left(\frac{1}{2}p\right)}{\Gamma\left(|\boldsymbol{A_K}|+|\boldsymbol{A_J}|+\frac{1}{2}\,p\right)}\end{aligned}$$

and

$$\begin{aligned}\widetilde{T}(\boldsymbol{K},\boldsymbol{J}) = \frac{1}{\sqrt{\pi}} \frac{\left(\sum_{\nu=1}^{q}(\boldsymbol{B_K}+\boldsymbol{B_J})_\nu\right)!}{\left(\prod_{\nu=1}^{q}(\boldsymbol{B_K}+\boldsymbol{B_J})_\nu!\right)} \frac{\left(\prod_{\nu=1}^{q}[2(\boldsymbol{B_K}+\boldsymbol{B_J})_\nu]!\right)}{\left(\sum_{\nu=1}^{q}2(\boldsymbol{B_K}+\boldsymbol{B_J})_\nu\right)!} \\ \times \frac{\Gamma\left(|\boldsymbol{B_K}|+|\boldsymbol{B_J}|+\frac{1}{2}\right)\,\Gamma\left(\frac{1}{2}q\right)}{\Gamma\left(|\boldsymbol{B_K}|+|\boldsymbol{B_J}|+\frac{1}{2}\,q\right)}.\end{aligned}$$

Proof. The Lancaster expansion of the standardized jointly normal random vectors is given in (4.48) to be

$$\phi_{\widetilde{X},\widetilde{Y}}(x,y) - \phi_{\widetilde{X}}(x)\,\phi_{\widetilde{Y}}(y) = \sum_{\boldsymbol{N}\neq\boldsymbol{0}} \frac{\Lambda_{XY}^{\boldsymbol{N}}}{\boldsymbol{N}!}\, H_{\boldsymbol{A_N}}(x,I_p)\, H_{\boldsymbol{B_N}}(y,I_q)\,.$$

For the Fourier transforms, we find by integration-by-parts

$$\begin{aligned}\mathcal{L}_{\boldsymbol{N}}^{\widetilde{X}}(s) &= \int_{\mathbb{R}^p} e^{\mathrm{i}\langle s,x\rangle}\, \phi_{\widetilde{X}}(x)\, H_{\boldsymbol{A_N}}(x,I_p)\, \mathrm{d}x \\ &= (-1)^{|\boldsymbol{A_N}|} \int_{\mathbb{R}^p} e^{\mathrm{i}\langle s,x\rangle} \left(\frac{\partial}{\partial x}\right)^{\boldsymbol{A_N}} \phi_{\widetilde{X}}(x)\, \mathrm{d}x \\ &= \mathrm{i}^{|\boldsymbol{A_N}|}\, s^{\boldsymbol{A_N}}\, e^{-s's/2};\end{aligned}$$

and, similarly,

$$\mathcal{L}_{\boldsymbol{N}}^{\widetilde{Y}}(t) = \mathrm{i}^{|\boldsymbol{B_N}|}\, s^{\boldsymbol{B_N}}\, e^{-t't/2}.$$

Furthermore, we have

$$\int_{\mathbb{R}^p} \frac{\mathcal{L}_{\boldsymbol{K}}^{\widetilde{X}}(s)\, \mathcal{L}_{\boldsymbol{J}}^{\widetilde{X}}(-s)}{|s|_p^{p+1}}\, \mathrm{d}s = \int_{\mathbb{R}^p} (-1)^{|\boldsymbol{A_J}|}\, \mathrm{i}^{|\boldsymbol{A_K}|+|\boldsymbol{A_J}|}\, s^{\boldsymbol{A_K}+\boldsymbol{A_J}}\, e^{-s's} \frac{\mathrm{d}s}{|s|_p^{p+1}}.$$

To calculate the affinely invariant distance correlation we reduced the problem of solving the integral to integrals of type

$$\int_{\mathbb{R}^p} s^{\boldsymbol{A_K}+\boldsymbol{A_J}}\, e^{-s's}\, \frac{\mathrm{d}s}{|s|_p^{p+1}}.$$

To derive the latter integral we change variables to hyperspherical coordinates, say $s = r\omega$ where $r > 0$, $\omega = (\omega_1, \ldots, \omega_p)' \in S^{p-1}$:

$$\int_{S^{p-1}} \omega^{\boldsymbol{A_K}+\boldsymbol{A_J}} \int_{\mathbb{R}_+} r^{|\boldsymbol{A_K}|+|\boldsymbol{A_J}|-2}\, e^{-r^2}\, \mathrm{d}r\, \mathrm{d}\omega.$$

The integral with respect to r is a standard gamma integral such that we obtain

$$\tfrac{1}{2}\Gamma\left((|\boldsymbol{A_K}| + |\boldsymbol{A_J}| - 1)/2\right) \int_{S^{p-1}} \omega^{\boldsymbol{A_K}+\boldsymbol{A_J}}\, \mathrm{d}\omega.$$

Furthermore we evaluate the remaining monomial over the sphere as in Aubert and Lam (2003) to

$$\int_{S^{p-1}} \omega^{\boldsymbol{A_K}+\boldsymbol{A_J}}\, \mathrm{d}\omega = \widetilde{S}(\boldsymbol{K}, \boldsymbol{J}),$$

such that

$$\int_{\mathbb{R}^p} \frac{\mathcal{L}_{\boldsymbol{K}}^{\widetilde{Y}}(s)\, \mathcal{L}_{\boldsymbol{J}}^{\widetilde{Y}}(-s)}{|s|_p^{p+1}}\, \mathrm{d}s = \tfrac{1}{2}\,(-1)^{|\boldsymbol{A_J}|}\, \mathrm{i}^{|\boldsymbol{A_K}|+|\boldsymbol{A_J}|}\, \Gamma\left((|\boldsymbol{A_K}| + |\boldsymbol{A_J}| - 1)/2\right)\, \widetilde{S}(\boldsymbol{K}, \boldsymbol{J}).$$

Similarly, we derive the integral with respect to t, such that by putting together all factors we get the final result by Theorem 4.11. We may apply Theorem 4.11 by similar arguments as in the bi-variate case.

□

For the usual distance covariance $\mathcal{V}(X, Y)$ we can use exactly the same arguments as before. However, obtaining the result is slightly more complicated and the final outcome is a little lengthy. Therefore, we explain the derivation of distance correlation for the multivariate normal distribution step by step. The Lancaster expansion of $\phi_{X,Y}$ is

$$\phi_{X,Y}(x, y) - \phi_X(x)\, \phi_Y(y) = \sum_{\boldsymbol{N} \neq \boldsymbol{0}} \frac{\Sigma_{XY}^{\boldsymbol{N}}}{\boldsymbol{N}!}\, H_{\boldsymbol{A_N}}\left(x, I_p\right)\, H_{\boldsymbol{B_N}}\left(y, I_q\right).$$

In the same fashion as above we compute

$$\begin{aligned}\mathcal{L}_{\boldsymbol{N}}^{X}(s) &= \int_{\mathbb{R}^p} e^{\mathrm{i}\, s'x}\, \phi_X(x)\, H_{\boldsymbol{A}_{\boldsymbol{N}}}(x, \Sigma_X)\, \mathrm{d}x \\ &= \mathrm{i}^{|\boldsymbol{A}_{\boldsymbol{N}}|}\, s^{\boldsymbol{A}_{\boldsymbol{N}}}\, e^{-s'\Sigma_X s/2},\end{aligned}$$

such that

$$\int_{\mathbb{R}^p} \frac{\mathcal{L}_{\boldsymbol{K}}^{X}(s)\, \mathcal{L}_{\boldsymbol{J}}^{X}(-s)}{|s|_p^{p+1}}\, \mathrm{d}s = \int_{\mathbb{R}^p} \mathrm{i}^{|\boldsymbol{A}_{\boldsymbol{K}}|+|\boldsymbol{A}_{\boldsymbol{J}}|}\, (-1)^{|\boldsymbol{A}_{\boldsymbol{J}}|}\, s^{\boldsymbol{A}_{\boldsymbol{K}}+\boldsymbol{A}_{\boldsymbol{J}}}\, e^{-s'\Sigma_X s} \frac{\mathrm{d}s}{|s|_p^{p+1}}.$$

To compute the integral

$$\int_{\mathbb{R}^p} s^{\boldsymbol{A}_{\boldsymbol{K}}+\boldsymbol{A}_{\boldsymbol{J}}}\, e^{-s'\Sigma_X s}\, \frac{\mathrm{d}s}{|s|_p^{p+1}}$$

we again change variables to hyperspherical coordinates, say $s = r\omega$ where $r > 0$, $\omega = (\omega_1, \ldots, \omega_p)' \in S^{p-1}$. The integral with respect to r is a standard gamma integral such that the above integral simplifies to

$$\left(\tfrac{1}{2}\right)^{3/2-1/2(|\boldsymbol{A}_{\boldsymbol{J}}|+|\boldsymbol{A}_{\boldsymbol{I}}|)}\, \Gamma\left((|\boldsymbol{A}_{\boldsymbol{J}}| + |\boldsymbol{A}_{\boldsymbol{I}}| - 1)/2\right) \int_{S^{p-1}} \left(\phi'\Sigma_X\phi\right)^{-1}\, \phi^{\boldsymbol{A}_{\boldsymbol{J}}+\boldsymbol{A}_{\boldsymbol{I}}}.$$

If $\Sigma_X \equiv CI_p$ for some $C > 0$, the proof is as above. If $\Sigma_{11} \neq CI_p$ we can divide by a scalar, e.g. $\|\Sigma_X\|_1$ such that $\Delta_X := \Sigma_X/\|\Sigma_X\|_1 \prec I_p$ in the positive definite sense. Hence, we may expand with the help of negativ binomial expansion

$$\left(\phi'\Delta_X\phi\right)^{-1} = \sum_{k=0}^{\infty} \binom{-1}{k} \left(\phi'\Delta_X\phi - 1\right)^k.$$

Finally, we may expand $\left(\phi'\Delta_X\phi - 1\right)^k$ with the help of the regular binomial Theorem. Hence, we see that we are left to evaluate an integral over a sum of monomials with respect to the sphere. This can be evluated with exactly the same formula as above, see Aubert and Lam (2003). If we denote all the sums of these integrals with respect to the spheres again by $S(\boldsymbol{J}, \boldsymbol{I})$ and $T(\boldsymbol{J}, \boldsymbol{I})$, respectively, the outcome is

$$\begin{aligned}\mathcal{V}^2(X,Y) = \frac{1}{c_p\, c_q} \sum_{\boldsymbol{J}\neq\boldsymbol{0}} \sum_{\boldsymbol{I}\neq\boldsymbol{0}} S(\boldsymbol{J}, \boldsymbol{I})\, T(\boldsymbol{J}, \boldsymbol{I})\, \left(\tfrac{1}{2}\right)^{3-1/2(|\boldsymbol{A}_{\boldsymbol{J}}|+|\boldsymbol{A}_{\boldsymbol{I}}|+|\boldsymbol{B}_{\boldsymbol{J}}|+|\boldsymbol{B}_{\boldsymbol{I}}|)} \\ \times \Gamma\left((|\boldsymbol{A}_{\boldsymbol{J}}| + |\boldsymbol{A}_{\boldsymbol{I}}| - 1)/2\right)\, \Gamma\left((|\boldsymbol{B}_{\boldsymbol{J}}| + |\boldsymbol{B}_{\boldsymbol{I}}| - 1)/2\right)\, \frac{\Sigma_{XY}^{\boldsymbol{J}}}{\boldsymbol{J}!} \frac{\Sigma_{XY}^{\boldsymbol{I}}}{\boldsymbol{I}!}.\end{aligned}$$

S and T do now of course depend on Σ_X and Σ_Y, respectively.

4.5.4.3 The bivariate gamma distribution

Theorem 4.14. *Suppose that random vector (X,Y) is distributed according to a Sarmanov bivariate gamma distribution, as given by (4.50). Then the distance covariance between X and Y is*

$$
\begin{aligned}
\mathcal{V}^2(X,Y) = \frac{4}{2^{\alpha+\beta}} \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} a_j\, a_k\, \frac{(\alpha)_j}{j!}\, \frac{(\alpha)_k}{k!}\, \frac{(\beta)_j}{j!}\, \frac{(\beta)_k}{k!}\, & \frac{1}{(\alpha-1+j)!(\beta-1+k)!} \\
\times \frac{(-1)^{2\alpha+2\beta+j+k}}{2^{2j}}\, & \frac{\Gamma(j+k-1)}{\Gamma(j-\alpha)}\frac{\Gamma(j+k-1)}{\Gamma(j-\beta)} \\
\times\, & {}_2F_1\left(-\alpha+1-k, \alpha+j; -\alpha+j; \tfrac{1}{2}\right) \\
\times\, & {}_2F_1\left(-\beta+1-k, \beta+j; -\beta+j; \tfrac{1}{2}\right). \qquad (4.57)
\end{aligned}
$$

Proof. By (4.50), we have 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 (4.52) that for $s,t \in \mathbb{R}$,

$$
\begin{aligned}
\mathcal{L}_n^X(s) &= \int_0^\infty \exp(\mathrm{i}sx)\, L_n^{(\alpha-1)}(x)\, \phi_X(x)\, \mathrm{d}x \\
&= \frac{1}{\Gamma(\alpha)} \int_0^\infty \exp\big(-(1-\mathrm{i}s)x\big)\, x^{\alpha-1}\, L_n^{(\alpha-1)}(x)\, \mathrm{d}x.
\end{aligned}
$$

By applying (4.49), we deduce that

$$
\begin{aligned}
\mathcal{L}_n^X(s) &= \frac{(\alpha)_n}{n!}\, (1-\mathrm{i}s)^{-\alpha}\, \left(1-(1-\mathrm{i}s)^{-1}\right)^n \\
&= \frac{(\alpha)_n}{n!}\, (1-\mathrm{i}s)^{-(\alpha+n)}\, (-\mathrm{i}s)^n,
\end{aligned}
$$

and, analogously,

$$\mathcal{L}_n^Y(t) = \frac{(\beta)_n}{n!}\, (1-\mathrm{i}t)^{-(\beta+n)}\, (-\mathrm{i}t)^n.$$

We calculate, next, the integral

$$
\begin{aligned}
\int_{\mathbb{R}} \mathcal{L}_j^X(s)\, \mathcal{L}_k^X(-s)\, \frac{\mathrm{d}s}{s^2} &= \frac{(\alpha)_j}{j!}\, \frac{(\alpha)_k}{k!}\, \mathrm{i}^{-j+k} \int_{\mathbb{R}} s^{j+k-2}\, (1-\mathrm{i}s)^{-(\alpha+j)}\, (1+\mathrm{i}s)^{-(\alpha+k)}\, \mathrm{d}s \\
&\equiv \frac{(\alpha)_k}{k!}\, \frac{(\alpha)_j}{j!}\, \mathrm{i}^{-j+k} \int_{\mathbb{R}} g(s)\, \mathrm{d}s, \qquad (4.58)
\end{aligned}
$$

where

$$g(s) = s^{j+k-2}\, (1-\mathrm{i}s)^{-(\alpha+j)}\, (1+\mathrm{i}s)^{-(\alpha+k)}, \qquad (4.59)$$

$s \in \mathbb{R}$. We provide two approaches to calculating this integral, one by means of Cauchy's residue theorem and another by applying a beta integral which is also due to Cauchy.

For the first approach we assume, to begin with, that $\alpha, \beta \in \mathbb{N}$. In that case, we have $g(s) = \mathcal{P}(s)/\mathcal{Q}(s)$, where $\mathcal{P}$ and $\mathcal{Q}$ are polynomials with $\deg(\mathcal{P}) + 2 \leq \deg(\mathcal{Q})$ and $\mathcal{Q}(s) \neq 0$ for all $s \in \mathbb{R}$. Furthermore, the rational function g, when extended to the complex plane, has one pole in the upper half-plane, at $z = \mathrm{i}$, and this pole is of order $\alpha + k + 1$. Therefore, by Cauchy's residue theorem,

$$\int_{\mathbb{R}} g(s)\,\mathrm{d}s = 2\pi\mathrm{i}\,\mathrm{Res}(g;\mathrm{i}).$$

An explicit formula for this residue can be found *via*

$$\begin{aligned}
\mathrm{Res}(g;\mathrm{i}) =& \frac{1}{(\alpha-1+k)!} \lim_{z\to\mathrm{i}} \frac{\mathrm{d}^{\alpha-1+k}}{\mathrm{d}z^{\alpha-1+k}} \left[(z-\mathrm{i})^{\alpha+k} g(z)\right] \\
=& \frac{1}{(\alpha-1+k)!} \lim_{z\to\mathrm{i}} \frac{\mathrm{d}^{\alpha-1+k}}{\mathrm{d}z^{\alpha-1+k}} \Big[(z-\mathrm{i})^{\alpha+k} (-1)^{-(\alpha+j)}\, \mathrm{i}^{-(2\alpha+j+k)} \\
& \qquad\qquad \times z^{j+k-2} (z+\mathrm{i})^{-(\alpha+j)} (z-\mathrm{i})^{-(\alpha+k)}\Big] \\
=& \frac{1}{(\alpha-1+k)!} (-1)^{-(\alpha+j)}\, \mathrm{i}^{-2\alpha-j-k-1} \left(\frac{1}{2}\right)^{\alpha+j} \\
& \times \sum_{\nu=0}^{\alpha-1+k} \binom{\alpha-1+k}{\nu} \left(-\frac{1}{2}\right)^{\alpha-1+k-\nu} (j+k-\nu-1)_\nu\, (\alpha+j)_{\alpha-1+k-\nu}.
\end{aligned}$$

By reversing the order of summation and simplifying the terms in the latter sum, we find that the sum equals

$$\begin{aligned}
&\sum_{\nu=0}^{\alpha-1+k} \binom{\alpha-1+k}{\nu} \left(-\frac{1}{2}\right)^{\alpha-1+k-\nu} (j+k-\nu-1)_\nu\, (\alpha+j)_{\alpha-1+k-\nu} \\
&\qquad = \sum_{\nu=0}^{\infty} \binom{\alpha-1+k}{\nu} \frac{(j-\alpha+\nu)_{\alpha-1+k-\nu}}{2^\nu} (\alpha+j)_\nu\, (-1)^\nu \\
&\qquad = \sum_{\nu=0}^{\infty} \frac{(-\alpha+1-k)_\nu}{\nu!\, 2^\nu} \frac{(j-\alpha)_{\alpha-1+k}\, (\alpha+j)_\nu}{(j-\alpha)_\nu} \\
&\qquad = \frac{\Gamma(j+k-1)}{\Gamma(j-\alpha)}\, {}_2F_1\left(-\alpha+1-k, \alpha+j; -\alpha+j; \tfrac{1}{2}\right),
\end{aligned}$$

where ${}_2F_1$ denotes the Gaussian hypergeometric series. For general $\alpha \geq \beta > -1$, this result remains valid because of Carlson's Theorem (Andrews, p. 110; 1999).

Our second approach to calculating the integral (4.58) makes use of Cauchy's beta integral formula:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\mathrm{d}t}{(1+\mathrm{i}at)^x\, (1-\mathrm{i}bt)^y} = \frac{\Gamma(x+y-1)}{\Gamma(x)\Gamma(y)} a^{y-1}\, b^{x-1}\, (a+b)^{1-x-y},$$

where $\mathrm{Re}(x+y) > 1$, $\mathrm{Re}(a) > 0$, $\mathrm{Re}(b) > 0$; see (Andrews, p. 48; 1999). Differentiating both sides of this identity $j+k-2$ times with respect to a and choosing the suiting parameters we find

$$\int_{\mathbb{R}} g(s)\,\mathrm{d}s = \frac{2\pi\,\mathrm{i}^{j+k-2}}{(\alpha+2-k)_{j+k-2}}\left(\frac{1}{2}\right)^{2\alpha+1}\frac{\Gamma(2\alpha+1)}{\Gamma(\alpha+2-j)\,\Gamma(\alpha+2-k)}$$
$$\times\ {}_2F_1\left(-j-k+2, 2\alpha+1; \alpha+2-k; \tfrac{1}{2}\right).$$

Thus, we obtain

$$\int_{\mathbb{R}} \mathcal{L}_j^X(s)\,\mathcal{L}_k^X(-s)\frac{\mathrm{d}s}{s^2} = \frac{(\alpha)_j}{j!}\,\frac{(\alpha)_k}{k!}\,\mathrm{i}^{j-k}\,\frac{1}{(\alpha-1+k)!}\,(-1)^{-(\alpha+j)}\,\mathrm{i}^{-2\alpha-j-k-1}$$
$$\times\left(\tfrac{1}{2}\right)^{\alpha+j}\frac{\Gamma(j+k-1)}{\Gamma(j-\alpha)}\,{}_2F_1\left(-\alpha+1-k, \alpha+j; -\alpha+j; \tfrac{1}{2}\right),$$

and analogously for Y. By Theorem 4.11, we obtain the series (4.57) as a formal expression for $\mathcal{V}^2(X,Y)$.

Finally, we verify that (4.57) converges absolutely. We observe, by (4.59), that

$$\int_{\mathbb{R}} |g(s)|\,\mathrm{d}s = \int_{\mathbb{R}} |s|^{j+k-2}\,(1+s^2)^{-(2\alpha+j+k)/2}\,\mathrm{d}s.$$

Making the change of variables $s^2 = u/(1-u)$, the latter is transformed to

$$\tfrac{1}{2}\int_0^1 u^{(j+k-3)/2}\,(1-u)^{\alpha+\frac{1}{2}}\,\mathrm{d}u = B\left(\tfrac{1}{2}(j+k-1), \alpha+\tfrac{1}{2}\right),$$

where $B(\cdot,\cdot)$ is the standard beta function, and this integral converges absolutely because $j+k-1>0$ for all $j,k\in\mathbb{N}$ and $\alpha+1/2>0$ for all $\alpha>0$. Hence, to establish absolute convergence of the series (4.57), we need to examine the sum

$$\sum_{j=1}^{\infty} a_j\,\frac{(\alpha)_j}{j!}\,\frac{(\beta)_j}{j!}\sum_{k=1}^{\infty} a_k\,\frac{(\alpha)_k}{k!}\,\frac{(\beta)_k}{k!}$$
$$\times\, B\left(\tfrac{1}{2}(j+k-1), \alpha+\tfrac{1}{2}\right)\,B\left(\tfrac{1}{2}(j+k-1), \beta+\tfrac{1}{2}\right). \tag{4.60}$$

By (4.51), we have $a_j \le \lambda^j$ for all j. Also, for fixed j and large k, Stirling's formula for the gamma function yields the asymptotic behavior,

$$B\left(\tfrac{1}{2}(j+k-1), \alpha+\tfrac{1}{2}\right)\,B\left(\tfrac{1}{2}(j+k-1), \beta+\tfrac{1}{2}\right) \sim C_1\,k^{-(\alpha+\beta+1)},$$

and

$$\frac{(\alpha)_j}{j!} \sim C_2\,j^{\alpha-1},$$

where C_1 and C_2 are constants which do not depend on j, and similarly for β. As a result, we find that the inner sum satisfies

$$\sum_{k=1}^{\infty} a_k\,\frac{(\alpha)_k}{k!}\,\frac{(\beta)_k}{k!}\,B\left(\tfrac{1}{2}(j+k-1), \alpha+\tfrac{1}{2}\right)B\left(\tfrac{1}{2}(j+k-1), \beta+\tfrac{1}{2}\right)$$
$$\le C_1 + C_2\sum_{j=1}^{\infty}\frac{\lambda^k}{k^3} = C_3 < \infty,$$

where C_1, C_2, and C_3 are constants which are independent of j due to a standard geometric series argument. Therefore the double sum in (4.60) is bounded above by

$$C_3 \sum_{k=1}^{\infty} \lambda^k \, k^{\alpha+\beta-2} < \infty.$$

Therefore, the series (4.60) converges absolutely and, by Fubini's theorem, the interchange of integral and summation in (4.54) is justified. □

To calculate the distance variances $\mathcal{V}(X, X)$ and $V(Y, Y)$, we note that only the marginal distributions are relevant. Therefore, we may assume that X and Y have any joint distribution for which their marginal distributions are gamma with parameters α and β, respectively. To that end, we first set $\beta = \alpha$; then, the Sarmanov bivariate gamma distribution reduces to the Kibble-Moran distribution, and it is well-known (Kotz, et al.; 2000) that the joint characteristic function of (X, Y) is

$$\big((1 - \mathrm{i}t_1)(1 - \mathrm{i}t_2) + \lambda t_1 t_2\big)^{-\alpha}.$$

Next, we let $\lambda \to 1$; then this characteristic function converges to

$$\big(1 - \mathrm{i}(t_1 + t_2)\big)^{-\alpha} \equiv E \exp\big(\mathrm{i}(t_1 + t_2)X\big),$$

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.,

$$\begin{aligned}
\mathcal{V}^2(X, X) &= \frac{1}{c_1^2} \int_{\mathbb{R}^2} |\psi_X(s + t) - \psi_X(s)\psi_X(t)|^2 \frac{\mathrm{d}s}{s^2} \frac{\mathrm{d}t}{t^2} \\
&= \lim_{\lambda \to 1-} \lim_{\beta \to \alpha} \frac{1}{c_1^2} \int_{\mathbb{R}^2} |\psi_{X,Y}(s, t) - \psi_X(s)\psi_Y(t)|^2 \frac{\mathrm{d}s}{s^2} \frac{\mathrm{d}t}{t^2} \\
&= \lim_{\lambda \to 1-} \lim_{\beta \to \alpha} \mathcal{V}^2(X, Y).
\end{aligned}$$

Similarly,

$$\mathcal{V}(Y, Y) = \lim_{\lambda \to 1-} \lim_{\alpha \to \beta} \mathcal{V}^2(X, Y).$$

5 Extensions of Distance Correlation

Székely and Rizzo (2005), in developing the foundations of distance correlation, derived an intriguing multidimensional singular integral, see (4.2). It is this integral which is the subject of the present section, see Dueck, Edelmann and Richards (2015).

To recapitulate, suppose that $\alpha \in \mathbb{C}$ satisfies $0 < \Re(\alpha) < 2$. Székely and Rizzo (2005) proved that, for all $x \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} \frac{1 - \cos(\langle t, x \rangle)}{|t|_d^{d+\alpha}} \, \mathrm{d}t = C(d, \alpha) \, |x|_d^{\alpha}, \tag{5.1}$$

where

$$C(d, \alpha) = \frac{2\pi^{d/2} \, \Gamma(1 - \alpha/2)}{\alpha \, 2^{\alpha} \, \Gamma\big((d + \alpha)/2\big)}. \tag{5.2}$$

Székely and Rizzo defined the integral (5.1) by means of a regularization procedure, where the integrals at 0 and at ∞ are in a principal value sense: $\lim_{\epsilon \to 0} \int_{\mathbb{R}^d \setminus \{\epsilon B + \epsilon^{-1} B^c\}}$, where B is the unit ball centered at the origin in $\mathbb{R}^d$ and B^c is the complement of B.

In this section, we generalize the integral (5.1) by inserting into the integrand a truncated Maclaurin expansion of the function $\cos(\langle t, x \rangle)$. We show that the generalization is valid for all $\alpha \in \mathbb{C}$ such that $2(m-1) < \Re(\alpha) < 2m$, where m is any positive integer. Moreover, we prove that the generalization converges absolutely under the stated condition on α; as a consequence, we deduce that (5.1) converges without the need for regularization.

We note that the integral (5.1) arises in other areas of probability and statistics. Indeed, in the area of generalized random fields, (5.1) provides the spectral measure of a power law generalized covariance function, which corresponds to fractional Brownian motion; see Reed, Lee and Truong (1995) or Chilès and Delfiner (2012, p. 266, Section 4.5.6). In mathematical analysis, a related integral is treated by Gelfand and Shilov (1964, pp. 192–195), and a similar singular integral arises in Fourier analysis in the derivation of the norms of integral operators between certain Sobolev spaces of functions (Stein, 1970, pp. 140 and 263).

We remark that the extension of (5.1) to more general values of α raises the intriguing possibility that a general theory of distance correlation can be developed for values of α outside the range $(0, 2)$.

Now let $m \in \mathbb{N}$, the set of positive integers. Also, for $v \in \mathbb{R}$, define

$$\cos_m(v) := \sum_{j=0}^{m-1} (-1)^j \frac{v^{2j}}{(2j)!} \tag{5.3}$$

to be the truncated Maclaurin expansion of the cosine function, where the expansion is halted at the mth summand.

The following result generalizes (5.1) to arbitrary $m \in \mathbb{N}$.

Theorem 5.1. *Let $m \in \mathbb{N}$ and $x \in \mathbb{R}^d$. For $\alpha \in \mathbb{C}$,*

$$\int_{\mathbb{R}^d} \frac{\cos_m(\langle t, x \rangle) - \cos(\langle t, x \rangle)}{|t|_d^{d+\alpha}} \, \mathrm{d}t = C(d, \alpha) \, |x|_d^{\alpha}, \tag{5.4}$$

with absolute convergence if and only if $2(m-1) < \Re(\alpha) < 2m$, *where* $C(d,\alpha)$ *is given in (5.2).*

Proof. We shall establish the proof by induction on m.

Throughout the proof, we let $B_a = \{x \in \mathbb{R}^d : |x|_d < a\}$ denote the ball which is centered at the origin and which is of radius a.

Consider the case in which $m = 1$. In this case, observe that for $t \in B_a$ where a is sufficiently small, the function

$$t \mapsto \cos_1(\langle t, x\rangle) - \cos(\langle t, x\rangle) \equiv 1 - \cos(\langle t, x\rangle)$$

is asymptotic to $|t|_d^2$. Then the integrand in (5.4), when restricted to B_a, is asymptotic to $|t|_d^{-d-\alpha+2}$. By a transformation to spherical coordinates to compute the integral over the unit ball B we deduce that the integrand is integrable over B_a, and hence integrable over any compact neighborhood of the origin, if and only if $\Re(\alpha) < 2$.

For $|t|_d \to \infty$, we apply the bound $|1 - \cos(\langle t, x\rangle)| \le 2$ to deduce that the integrand in (5.4) (with $m = 1$) is integrable over $\mathbb{R} \setminus B_a$ if and only if $\Re(\alpha) > 0$. Consequently, for $m = 1$, the integral converges for all $x \in \mathbb{R}^d$ if and only if $0 < \Re(\alpha) < 2$.

To conclude the proof for the case in which $m = 1$, we proceed precisely as did Székely, et al. (2007, p. 2771) to obtain the right-hand side of (5.4).

Next, we assume by inductive hypothesis that the assertion holds for a given positive integer m. Note that the right-hand side of (5.4), as a function of $\alpha \in \mathbb{C}$, is meromorphic with a pole at each nonnegative integral α.

By (5.3),

$$\cos_{m+1}(v) = \cos_m(v) + (-1)^m \frac{v^{2m}}{(2m)!}.$$

For fixed $a > 0$, we decompose the integral (5.4) into a sum of three terms:

$$\int_{\mathbb{R}^d} \frac{\cos_m(\langle t, x\rangle) - \cos(\langle t, x\rangle)}{|t|_d^{d+\alpha}} \, \mathrm{d}t = T_1 + T_2 + T_3, \tag{5.5}$$

where

$$T_1 = \int_{B_a} \frac{\cos_{m+1}(\langle t, x\rangle) - \cos(\langle t, x\rangle)}{|t|_d^{d+\alpha}} \, \mathrm{d}t,$$
$$T_2 = \int_{\mathbb{R}^d \setminus B_a} \frac{\cos_m(\langle t, x\rangle) - \cos(\langle t, x\rangle)}{|t|_d^{d+\alpha}} \, \mathrm{d}t,$$

and

$$T_3 = \frac{(-1)^{m-1}}{(2m)!} \int_{B_a} \frac{\langle t, x\rangle^{2m}}{|t|_d^{d+\alpha}} \, \mathrm{d}t.$$

We now determine the necessary and sufficient condition on the range of α for which the decomposition (5.5) entails absolute convergence of the integral. In so doing, we examine each term individually.

In the case of T_1, we apply (5.3) to write

$$\cos_{m+1}(\langle t,x\rangle) - \cos(\langle t,x\rangle) = \sum_{j=m+1}^{\infty} (-1)^{j+1} \frac{\langle t,x\rangle^{2j}}{(2j)!}. \tag{5.6}$$

Proceeding formally to interchange the integral and summation, we obtain

$$T_1 = \sum_{j=m+1}^{\infty} \frac{(-1)^{j+1}}{(2j)!} \int_{B_a} \frac{\langle t,x\rangle^{2j}}{|t|_d^{d+\alpha}}\, \mathrm{d}t. \tag{5.7}$$

To verify that this series converges absolutely, note that

$$\int_{B_a} \frac{\langle t,x\rangle^{2j}}{|t|_d^{d+\alpha}}\, \mathrm{d}t \tag{5.8}$$

converges absolutely for all $x \in \mathbb{R}^d$ if and only if $\Re(\alpha) < 2j$. Moreover, this integral clearly is a radial function of x, and it can be calculated exactly by a transformation to spherical coordinates. After evaluating the integral and inserting it in the series (5.7), we find that the series converges absolutely for all $x \in \mathbb{R}^d$ if and only if $\Re(\alpha) < 2(m+1)$.

As regards the term T_2 we know, by inductive hypothesis, that it converges absolutely if and only if $\Re(\alpha) > 2(m-1)$.

To analyze the term T_3, we note that T_3 is similar to (5.8); hence we find that T_3 converges absolutely if and only if $\Re(\alpha) < 2m$.

To complete the proof, we need to evaluate T_3. Let S^{d-1} be the unit sphere in $\mathbb{R}^d$ and, for $\omega = (\omega_1, \ldots, \omega_d) \in S^{d-1}$, let $\mathrm{d}\omega$ denote the corresponding surface measure. We define

$$A_{d-1} = \int_{S^{d-1}} \omega_1^{2m}\, \mathrm{d}\omega,$$

a constant which can be calculated exactly but whose exact value is not needed in this context.

Similar to (5.8), T_3 is a radial function of x. Thus, by a standard invariance argument and by a transformation to spherical coordinates, $t = r\omega$, where $\omega \in S^{d-1}$ and $0 \le r \le a$, we obtain

$$\begin{aligned} T_3 &= \frac{(-1)^{m-1}}{(2m)!} A_{d-1} |x|_d^{2m} \int_0^a r^{2m-1-\alpha}\, \mathrm{d}r \\ &= \frac{(-1)^{m-1}}{(2m)!} A_{d-1} |x|_d^{2m} \frac{a^{2m-\alpha}}{2m-\alpha}. \end{aligned} \tag{5.9}$$

Moreover, the last term in (5.9) exists for all $\alpha \in \mathbb{C}$ such that $\Re(\alpha) \neq 2m$ and it is a meromorphic function of α.

To summarize, T_1 converges absolutely for $\Re(\alpha) < 2(m+1)$; T_2 converges absolutely for $\Re(\alpha) > 2(m-1)$; and T_3 converges absolutely for $\Re(\alpha) < 2m$. Therefore, the decomposition (5.5) is valid for $2(m-1) < \Re(\alpha) < 2m$, and it represents an analytic function which equals $C(d,\alpha)\,|x|_d^\alpha$ on the strip $\{\alpha \in \mathbb{C} : 2(m-1) < \Re(\alpha) < 2m\}$. Hence, by analytic continuation, we obtain for $2(m-1) < \Re(\alpha) < 2(m+1)$, $\Re(\alpha) \neq 2m$,

$$C(d,\alpha)\,|x|_d^\alpha = T_1 + T_2 + \frac{(-1)^{m-1}}{(2m)!}\,A_{d-1}\,|x|_d^{2m}\,\frac{a^{2m-\alpha}}{2m-\alpha}. \tag{5.10}$$

Now fix $2m < \Re(\alpha) < 2(m+1)$ and let $a \to \infty$ in (5.10). It is apparent that $T_2 \to 0$ and $a^{2m-\alpha} \to 0$; therefore, for $2m < \Re(\alpha) < 2(m+1)$, we obtain

$$C(d,\alpha)\,|x|_d^\alpha = \lim_{a\to\infty} T_1 = \int_{\mathbb{R}^d} \frac{\cos_{m+1}(\langle t,x\rangle) - \cos(\langle t,x\rangle)}{|t|_d^{d+\alpha}}\,\mathrm{d}t,$$

which concludes the proof. $\square$

In conclusion, we are intrigued by the possibility of applying (5.4) to develop a general theory of distance correlation for values of $\Re(\alpha) > 2$. We expect, *inter alia*, that such a theory will lead for sufficiently large $\Re(\alpha)$ to distance correlation analyses of data modeled by random vectors which do not have finite first moments, e.g., the multivariate stable distributions of index less than 2. Moreover, although the integral (5.4) diverges for $\Re(\alpha) = 2m$, our results raise the possibility of developing a theory of distance correlation at the poles by modifying (5.4) to attain convergence as $\Re(\alpha)$ converges to the poles.

Finally, we remark that our decomposition (5.5) was motivated by the ideas of Gelfand and Shilov (1964, p. 10).

6 Discussion

In this thesis we discussed the concepts of Hawkes processes and distance correlation as powerful tools of recognizing meaningful structures in point process data and in multivariate random samples.

We connected several causality concepts to Hawkes model and showed their equivalence within this model. Causality is fully encoded in the Hawkes kernels. Therefore, we provided a nonparametric, consistent and asymptotic normal estimator based on time series techniques. Furthermore, we illustrated our results by applying our methods to real world data from the spinal dorsal horn of a rat.

In future work, a test for noncausality needs to be established. However, this is a multiple hypothesis testing problem and it might not be tractable in higher dimensions. We also notice that our estimation procedure suffers from high computational costs in high dimensions.

We have studied an affinely invariant version of the distance correlation measure introduced by Székely, et al. (2007) and Székely and Rizzo (2009) in both population and sample settings (see Székely and Rizzo (2012) for further aspects of the role of invariance in properties of distance correlation measures). The affinely invariant distance correlation shares the desirable properties of the standard version of the distance correlation and equals the latter in the univariate case. In the multivariate case, the affinely invariant distance correlation remains unchanged under invertible affine transformations, unlike the standard version, which is preserved under orthogonal transformations only. Furthermore, the affinely invariant distance correlation admits an exact and readily computable expression in the case of subvectors from a multivariate normal population. The standard distance correlation version in the Gaussian case allows for a series expansion, too, but this does not appear to be a series that generally can be made simple, and further research will be necessary to make it accessible to efficient numerical computation. Related results and further asymptotics can be found in Gretton, et al. (2012) and Székely and Rizzo (2013).

Moreover, we computed distance covariance for several distributions, such as a multivariate Laplace distribution or for Lancaster distributions (a mulivariate gamma distribution). The form of the Lancaster expansions simplifies the computation of distance correlation severely. A drawback of the presented method consists of finding those appropriate expansions, even though the existence is immediate.

Finally, we generalize an integral which is at the core of distance correlation by analytic continuation. This raises the possibility that one may introduce distance correlation measures with different parameters in the weight functions.

Competing measures of dependence also have featured prominently recently (Reshef, et al. 2011; Speed, 2011). However, those measures are restricted to univariate settings, and claims of superior performance in exploratory data analysis have been disputed (Gorfine, Heller and Heller, 2012; Simon and Tibshirani, 2012). We therefore opine with Newton (2009) that the distance correlation and the affinely invariant distance correlation might become uniquely useful, and potentially the predominant, measures of dependence and associations for the 21st century.

References

[1] G.E. Andrews, R.A. Askey, and R. Roy. *Special Functions*. Cambridge University Press, Cambridge, 2000.

[2] S. Aubert and Lam C.S. Invariant integration over the unitary group. *Journal of Mathematical Physics*, 44:6112–6131, 2003.

[3] E. Bacry, K. Dayri, and J.F. Muzy. Non-parametric kernel estimation for symmetric hawkes processes. application to high frequency financial data. *The European Physical Journal B*, 85, 2012.

[4] E. Bacry, S. Delattre, M. Hoffmann, and J.F. Muzy. Modeling microstructure noise with mutually exciting point processes. *Accepted for publication in Quantitative Finance*, 2011.

[5] S.K. Bar-Lev, D. Bshouty, G. Letac, I. Lu, and D.St.P. Richards. The diagonal multivariate natural exponential families and their classification. *Journal of Theoretical Probability*, 7:883–929, 1993.

[6] J.F. Barrett and D.G. Lampard. An expansion for some second-order probability distributions and its application to noise problems. *IRE Transactions - Information Theory*, 1:10–15, 1955.

[7] M.S. Bartlett. The spectral analysis of point processes. *J. R. Statist. Soc. B*, 25:264–296, 1963.

[8] K.N. Berk. Consistent autoregressive spectral estimates. *The Annals of Statistics*, 2:489–502, 1974.

[9] P.J. Brantingham, E. Lewis, G. Mohler, and A. Bertozzi. Self-exciting point process models of the insurgency in iraq. 2011.

[10] J.P. Chilès and Delfiner P. *Geostatistics: Modeling Spatial Uncertainty, second edtion*, volume 1. Wiley, New York, 2012.

[11] F. Comte and E. Renault. Noncausality in continuous time models. *Cambridge University Press*, 12:215–256, 1996.

[12] R. Dahlhaus, J. Dueck, and M. Eichler. Nonparametric estimation of hawkes kernels and graphical models. *In preparation*, 2014.

[13] R. Dahlhaus and M. Eichler. Causality and graphical models for time series. *University Press, Oxford*, pages 115–137, 2003.

[14] P. Diaconis, K. Khare, and L. Saloff-Coste. Gibbs sampling, exponential families and orthogonal polynomials. *Statistical Science*, 23:151–178, 2008.

[15] V. Didelez. Graphical models for marked point processes based on local independence. *Journal of the Royal Statistical Society*, 70:245–264, 2008.

[16] J. Dueck, D. Edelmann, T. Gneiting, and D. Richards. The affinely invariant distance correlation. *Bernoulli*, 20:2305–2330, 2014.

[17] J. Dueck, D. Edelmann, and D. Richards. Distance correlation coefficient for lancaster distributions. *In preparation*, 2014.

[18] J. Dueck, D. Edelmann, and D. Richards. A generalization of an integral arising in the theory of distance correlation. *Statistics and Probability Letters: to appear*, 2015.

[19] M.L. Eaton. *Group Invariance Applications in Statistics*. Institute of Mathematical Statistics, Hayward, California, 1989.

[20] P. Embrechts, J.T. Liniger, and L. Lu. Multivariate hawkes processes: an application to financial data. *Journal of Applied Probability*, 48:367–378, 2011.

[21] J. Florens and D. Fougere. Noncausality in continuous time. *Econometrica*, 64:1195–1212, 1996.

[22] I.M. Gelfand and G.E. Shilov. *Generalized Functions*, volume 1. Academic Press, New York, 1964.

[23] T. Gneiting, K. Larson, K. Westrick, M. Genton, and E. Aldrich. Calibrated probabilistic forecasting at the stateline wind energy center: The regime-switching space-time (rst) technique. *Journal of the American Statistical Association*, 101:968–979, 2006.

[24] M. Gorfine, R. Heller, and Y. Heller. Comment on "detecting novel associations in large data sets". *Unpublished manuscript*, 2012. `http://iew3.technion.ac.il/~gorfinm/files/science6.pdf`.

[25] C.W.J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438, 1969.

[26] A. Gretton, K.M. Borgwardt, M.J. Rasch, B. Schoelkopf, and A. Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13:723–773, 2012.

[27] K.I. Gross and D.St.P. Richards. Special functions of matrix argument. i: Algebraic induction, zonal polynomials, and hypergeometric functions. *Transactions of the American Mathematical Society*, 301:781–811, 1987.

[28] A.G. Hawkes. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society*, 33:438–443, 1971.

[29] A.G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58:83–90, 1971.

[30] R. Heller, Y. Heller, and M. Gorfine. A consistent multivariate test of association based on ranks of distances. *Biometrika*, 100:503–510, 2013.

[31] A.S. Hering and M.G. Genton. Powering up with space-time wind forecasting. *Journal of the American Statistical Association*, 105:92–104, 2010.

[32] X. Huo and G.J. Székely. Fast computing for distance covariance. *Preprint: arXiv:1401.1503*, 23, 2014.

[33] A.T. James. Distributions of matrix variates and latent roots derived from normal samples. *Annals of Mathematical Statistics*, 35:475–501, 1964.

[34] P. Koev and A. Edelman. The efficient evaluation of the hypergeometric function of a matrix argument. *Mathematics of Computation*, 75:833–846, 2006.

[35] J. Kong, B.E.K. Klein, R. Klein, and G. Wahba. Using distance correlation and ss-anova to access associations of familial relationships, lifestyle factors, diseases, and mortality. *Proceedings of the National Academy of Sciences*, 109:20352–20357, 2012.

[36] M.R. Kosorok. Discussion of: Brownian distance covariance. *Annals of Applied Statistics*, 3:1270–1278, 2009.

[37] M.R. Kosorok. Correction on discussion of: Brownian distance covariance. *Annals of Applied Statistics*, 7:1247, 2013.

[38] S. Kotz, N. Balakrishnan, and N.L. Johnson. *Continuous Multivariate Distributions: Models and Applications, Second Edition*. Wiley, New York, 2000.

[39] A.E. Koudou. Probabilités de lancaster. *Expositiones Mathematicae*, 14:247–275, 1996.

[40] A.E. Koudou. Lancaster bivariate probability distributions with poisson, negative binomial and gamma margins. *Test*, 7:95–110, 1998.

[41] H.O. Lancaster. The structure of bivariate distributions. *Annals of Mathematical Statistics*, 29:719–736, 1958.

[42] H.O. Lancaster. *The Chi-Squared Distribution*. Wiley, New York, 1969.

[43] E. Lewis and G. Mohler. A nonparametric em algorithm for a multiscale hawkes process. *Joint Satistical Meetings 2011*, 2011.

[44] R. Lewis and G.C. Reinsel. Prediction of multivariate time series by autoregressive model fitting. *Journal of Multivariate Analysis*, 16(3):393–411, 1985.

[45] R. Lyons. Distance covariance in metric spaces. *Annals of Probability*, 41:3284–3305, 2013.

[46] H.B. Mann and A. Wald. On stochastic limit and order relationships. *The Annals of Mathematical Statistics*, 14:217–226, 1974.

[47] E. Martínez-Gómez, M. T. Richards, and D.St.P. Richards. Distance correlation methods for discovering associations in large astrophysical databases. *Astrophysical Journal*, 781:39–50, 2014.

[48] G. Mohler, M. Short, P. Brantingham, F. Schoenberg, and G. Tita. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106:100–108, 2011.

[49] R.J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley, New York, 1982.

[50] M.A. Newton. Introducing the discussion paper by székely and rizzo. *Annals of Applied Statistics*, 3:1233–1235, 2009.

[51] Y. Ogata. The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Ann Inst Stat Math*, 30:243–261, 1978.

[52] Y. Ogata. On lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, 27:23–31, 1981.

[53] Y. Ogata. Seismicity analysis through point-process modeling: A review. *Pure and Applied Geophysics*, 155:471–507, 1999.

[54] Y. Ogata and H. Akaike. On linear intensity models for mixed doubly stochastic poisson and self- exciting point processes. *Journal of the Royal Statistical Society B*, 44:102–107, 1982.

[55] Y. Ogata, H. Akaike, and K. Katsura. On lewis' simulation method for point processes. *Ann Inst Stat Math*, 34:373–387, 1981.

[56] T. Ozaki. Maximum likelihood estimation of hawkes' self-exciting point processes. *Ann Inst Stat Math*, 31:145–155, 1979.

[57] D. Pommeret. A characterization of lancaster probabilities with margins in a multivariate additive class. *The Indian Journal of Statistics*, 66:1–19, 2004.

[58] H. Pruscha. Semiparametric point process and time series models for series of events. 114, 1998. `http://epub.ub.uni-muenchen.de/1503/`.

[59] I.S. Reed, P.C. Lee, and T.K. Truong. Spectral representation of fractional brownian motion in n dimensions and its properties. *IEEE Transactions on Information Theory*, 41:1439–1451, 1995.

[60] B. Rémillard. Discussion of: Brownian distance covariance. *Annals of Applied Statistics*, 3:1295–1298, 2009.

[61] D.N. Reshef, J.A. Reshef, H.K. Finucane, S.R. Grossman, G. McVean, P.J. Turnbaugh, E.S. Lander, M. Mitzenmacher, and P.C. Sabeti. Detecting novel associations in large data sets. *Science*, 334:1518–1524, 2011.

[62] P. Reynaud-Bouret and S. Schbath. Adaptive estimation for hawkes processes; application to genome analysis. *The Annals of Statistics*, 38:2781–2822, 2010.

[63] M.T. Richards, D.St.P. Richards, and E. Martínez-Gómez. Interpreting the distance correlation results for the combo-17 survey. *Astrophysical Journal Letters*, 784:34–39, 2014.

[64] M.L. Rizzo and G.J. Székely. Disco analysis: A nonparametric extension of analysis of variance. *Annals of Applied Statistics*, 4:1034–1055, 2010.

[65] M.L. Rizzo and G.J. Székely. Energy: E-statistics (energystatistics). r package, version 1.4-0. 2011. available at `http://cran.us.r-project.org/web/packages/energy/index.html`.

[66] J. Sandkuehler and A.A. Eblen-Zajjur. Identification and characterization of rhythmic nociceptive and non-nociceptive spinal dorsal horn neurons in the rat. *Neuroscience*, 61:991–1006, 1994.

[67] O.V. Sarmanov. Generalized normal correlation and two-dimensional fréchet classes. *Soviet Mathematics - Doklady*, 7:596–599, 1966.

[68] O.V. Sarmanov. An approximate calculation of correlation coefficients between functions of dependent random variables. *Mathematical Notes, Academy of Sciences of the USSR*, 7:373–377, 1970.

[69] O.V. Sarmanov. Gamma correlation process and its properties. *Doklady Akademii Nauk, SSSR*, 191:30–32, 1970.

[70] O.V. Sarmanov and Z.N. Bratoeva. The probability properties of bilinear expansions in hermite polynomials. *Teoriya Veroyatnostei i ee Primeneniya*, 12.3:520–531, 1967.

[71] D.J. Scot. Central limit theorems for martingales using a skorokhod representation approach. *Adv. Appl. Probab.*, 5:119–137, 1973.

[72] D. Sejdinovic, B. Sriperumbudur, A. Gretton, and K. Fukumizu. Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *Annals of Statistics*, 41:2263–2291, 2013.

[73] N. Simon and R. Tibshirani. Comment on "detecting novel associations in large data sets," by reshef, et al.". Unpublished manuscript, available at `http://www-stat.stanford.edu/~tibs/reshef/comment.pdf`.

[74] T. Speed. A correlation for the 21st century. *Science*, 334:1502–1503, 2011.

[75] H.M. Srivastava and J.P. Singhal. Some extensions of the mehler formula. *Proceedings of the American Mathematical Society*, 31:135–141, 1972.

[76] E.M. Stein. *Singular Integrals and Differentiability Properties of Functions*. Princeton University Press, Princeton, 1970.

[77] G.J. Székely and M. Rizzo. Brownian distance covariance. *Annals of Applied Statistics*, 3:1236–1265, 2012.

[78] G.J. Székely and M. Rizzo. On the uniqueness of distance correlation. *Statistics & Probability Letters*, 82:2278–2282, 2012.

[79] G.J. Székely and M. Rizzo. The distance correlation -test of independence in high dimension. *Journal of Multivariate Analysis*, 2013. In press.

[80] G.J. Székely and M.L. Rizzo. Hierarchical clustering via joint between-within distances: Extending ward's minimum variance method. *Journal of Classification*, 22:151–183, 2005.

[81] G.J. Székely and M.L. Rizzo. Partial distance correlation with methods for dissimilarities. *Annals of Statistics*, 42:2382–2412, 2014.

[82] G.J. Székely, M.L. Rizzo, and N.K. Bakirov. Measuring and testing independence by correlation of distances. *Annals of Statistics*, 35:2769–2794, 2007.

[83] P.D. Thuan. Estimation of the spectral parameters of a stationary point process. *The Annals of Statistics*, 9:615–627, 1981.

[84] D. Vere-Jones. Stochastic models for earthquake occurrence. *Journal of the Royal Statistical Societys B*, 32:1–62, 1970.

[85] D. Vere-Jones and T. Ozaki. Some examples of statistical estimation applied to earthquake data. *Annals of the Institute of Statistical Mathematicss*, 34:189–207, 1982.

[86] P. Whittle. On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika Trust*, 50:129–134, 1963.

[87] C.S. Withers and S. Nadarajah. Expansions for the multivariate normal. *Journal of Multivariate Analysis*, 101:1311–1316, 2010.

[88] Z. Zhou. Measuring nonlinear dependence in time-series, a distance correlation approach. *Journal of Time Series Analysis*, 33:438–457, 2012.

[89] J. Zhuang, Y. Ogata, and D. Vere-Jones. Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97:369–380, 2002.