Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE @ DIRECT•

Journal of Multivariate Analysis 86 (2003) 310–329

Journal of Multivariate Analysis

http://www.elsevier.com/locate/jmva

Semi-parametric multivariate modelling when the marginals are the same ☆

J.S. Marron,[a] Miguel Nakamura,[b,*] and Víctor Pérez-Abreu[b]

[a] *Department of Statistics, University of North Carolina, Chapel Hill, NC 27599, USA*
[b] *Department of Probability and Statistics, Centro de Investigación en Matemáticas A. C., Apdo. Postal 402, Guanajuato, Gto. 36000, Mexico*

Received 13 November 2000

Abstract

A model is developed for multivariate distributions which have nearly the same marginals, up to shift and scale. This model, based on "interpolation" of characteristic functions, gives a new notion of "correlation". It allows straightforward nonparametric estimation of the common marginal distribution, which avoids the "curse of dimensionality" present when nonparametically estimating the full multivariate distribution. The method is illustrated with environmental monitoring network data, where multivariate modelling with common marginals is often appropriate.

AMS 2000 subject classifications: 62H05; 60E07; 62H20.

Keywords: Multivariate characteristic function; Measures of correlation; Infinitely divisible random vectors; Environmental data

1. Introduction

When the data are sparse in multivariate statistical analysis, the statistician often has little alternative to normal theory methods, even when the data are clearly not normal, because there is insufficient information in the data for the usual nonparametric alternatives, as for example density estimation. This is a common

☆ Research partially supported by CONACYT Grant No. 0328P-E9506.
* Corresponding author.
E-mail address: nakamura@cimat.mx (M. Nakamura).

0047-259X/03/$ - see front matter
doi:10.1016/S0047-259X(03)00051-4

situation in environmental monitoring data, where, up to shift and scale, it is sensible to view the marginal distributions of observations at different monitoring stations as all the same.

Here we propose a semiparametric multivariate model for distributions where the marginals are the same. The problem of multivariate modelling from given marginals has been extensively studied in the literature; see for example [7,9,13,21,22] for a comprehensive review of this subject, and [10] for an earlier history. Some of the general attempts for this kind of modelling are within the framework of the so-called *frailty* distributions which are generated by mixtures of distribution or survival functions [12]. These methods require knowledge of the explicit form of the one-dimensional distribution function and are particularly useful for reliability and extreme value distributions. Other approaches, like the *random-additive-effects* model [2], are based on moment generating or characteristic functions.

In our approach, very mild nonparametric assumptions are made about the common marginal distribution. The multivariate model is constructed through decomposition of *characteristic functions* by an "interpolation formula" having the totally dependent and the independent models as extreme points, and in such a way that the usual multivariate Gaussian results when the marginal distributions are Gaussian. This last property does not hold for other approaches to this problem. A by-product of this modelling is a new look at measures of dependence. Our model gives a class of such measures which includes the usual Pearson correlation coefficient as a special case. A new view as to why other measures of correlation are probably more useful when the distribution is not multivariate Gaussian is provided.

The problem of multivariate density estimation with common marginals arises naturally in the context of environmental monitoring network design and evaluation; see for example [3,14]. Specifically, a network consists of d possible monitoring sites where one or several environmental variables are monitored. In the case of a single variable, it is assumed that at each site the variable follows a common distribution with density $f(x_j),\ j = 1, \ldots, d$. The corresponding multivariate density $f(x_1, \ldots, x_d)$ for the d stations is needed to compute the Shannon index, which is a quantification of the quality of the performance of the network. Experience shows that the assumption of equal marginals after a location-scale transformation is reasonable here. One then expects an environmental variable to originate from the same family of distributions at each monitoring station.

Our multivariate model is developed in Section 2, where we also propose a method for estimating the dependence parameters. Section 3 presents some theoretical properties of the model, its relationship to cumulants, to the multivariate Gaussian distribution, and a discussion of the connection between the new dependence parameters and the usual correlation coefficients and other concepts of dependence. Results of a computational study to evaluate the performance of the estimation method are presented in Section 4. In Section 5 we fit the model to CO and ozone data from the environmental monitoring network in Mexico City.

2. Dependence model

2.1. Proposed model

A useful tool for understanding the multivariate probability distribution of a random vector $\underset{\rightarrow}{X} = (X_1, \ldots, X_d)' \in \mathbb{R}^d$ is its joint probability density function $f_{\underset{\rightarrow}{X}}(x_1, \ldots, x_d)$ (when it exists). Another representation of the joint distribution is through its characteristic function

$$\phi_{\underset{\rightarrow}{X}}(\underset{\rightarrow}{t}) = Ee^{i\underset{\rightarrow}{t}'\underset{\rightarrow}{X}} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{i\underset{\rightarrow}{t}'\underset{\rightarrow}{x}} f_{\underset{\rightarrow}{X}}(\underset{\rightarrow}{x})\, dx_1 \cdots dx_d,$$

which is the Fourier transform of the density, or in general of a multivariate probability distribution. For a comprehensive review of multivariate characteristic functions we refer to the book by Cuppens [4].

Our multivariate model assumes a common marginal density $f(x)$, i.e. for $j = 1, \ldots, d$

$$f(x_j) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\underset{\rightarrow}{X}}(x_1, \ldots, x_d)\, dx_1 \cdots dx_{j-1}\, dx_{j+1} \cdots dx_d.$$

This entails a common marginal characteristic function

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx.$$

When the components of $\underset{\rightarrow}{X}$ are independent, the joint density factors as

$$f_{\underset{\rightarrow}{X}}(\underset{\rightarrow}{x}) = \prod_{j=1}^{d} f(x_j),$$

and the joint characteristic function also factors as

$$\phi_{\underset{\rightarrow}{X}}(\underset{\rightarrow}{t}) = \prod_{j=1}^{d} \phi(t_j). \tag{1}$$

When $d = 2$, and the components of $\underset{\rightarrow}{X}$ are the *same*, i.e. $X_1 = X_2$, the joint characteristic function has the simple form

$$\phi_{\underset{\rightarrow}{X}}\left(\begin{pmatrix} t_1 \\ t_2 \end{pmatrix}\right) = Ee^{i(t_1 X_1 + t_2 X_1)} = \phi(t_1 + t_2). \tag{2}$$

Our multivariate model is a ‘‘geometric mixture’’ of the characteristic functions (1) and (2). Hence it can be viewed as a combination (in the Fourier domain) of distributions that are independent and have marginal variables that are the same. In particular, in the case $d = 2$, for a given marginal characteristic function $\phi(t)$, and given a parameter $\alpha \in [0, 1]$, our bivariate model is the distribution (when it exists) with *interpolated characteristic function*

$$\phi_{\underset{\rightarrow}{X}}\left(\begin{pmatrix} t_1 \\ t_2 \end{pmatrix}\right) = \phi(t_1 + t_2)^{\alpha}[\phi(t_1)\phi(t_2)]^{1-\alpha}. \tag{3}$$

When $\alpha = 1$, the total dependence model is obtained, while $\alpha = 0$ yields the independent case, in analogy with the usual correlation coefficient. In general, powers of the characteristic function $\phi(t)$ are defined by $[\phi(t)]^{\alpha} = \exp[\alpha \log \phi(t)]$, where we take for $\log \phi(t)$ the principal branch of the complex logarithm, that is, the one for which $\log \phi(0) = 0$. As will be seen in Section 3.1, this leads to a proper multivariate distribution, when ϕ corresponds to an infinitely divisible distribution.

In Section 3.2 it is seen that when the marginal distribution is the standard Gaussian, this reduces to the bivariate Gaussian distribution with correlation α. In Section 3.3 it is seen that the usual Pearson correlation coefficient (defined for any multivariate distribution with second moments) is a special case of α which results from fitting this model (with respect to different norms) to the joint characteristic function.

In the general case having $d \geqslant 3$, given parameters $\alpha_{j,k} = \alpha_{k,j} \in [0, 1], j, k = 1, \ldots, d$, $j \neq k$, with $\sum_{k=1,\ k\neq j}^{d} \alpha_{j,k} \leqslant 1$ for each j, our model is the multivariate distribution (when it exists) with characteristic function

$$\phi_{\underset{\rightarrow}{X}}(\underset{\rightarrow}{t}) = \left[\prod_{1\leqslant j<k\leqslant d} \phi(t_j + t_k)^{\alpha_{j,k}}\right]\left[\prod_{j=1}^{d} \phi(t_j)^{1-\sum_{k=1,\ k\neq j}^{d} \alpha_{j,k}}\right]. \tag{4}$$

Again in the standard Gaussian case, $\alpha_{j,k}$ are the usual correlations.

2.2. Estimation

Data $\underset{\rightarrow}{X}^{(1)}, \ldots, \underset{\rightarrow}{X}^{(n)}$ is assumed to be in the form of marginal standardizations, that is, each original sample is modified by subtracting the marginal mean and dividing by the marginal standard deviation. To estimate the marginal characteristic function ϕ and the parameters $\alpha_{j,k}$ in this model recall that an unbiased estimate of the joint characteristic function is the "empirical characteristic function" (see for example [6]):

$$\hat{\phi}_{\underset{\rightarrow}{X}}(\underset{\rightarrow}{t}) = n^{-1} \sum_{\ell=1}^{n} e^{i\,\underset{\rightarrow}{t}'\,\underset{\rightarrow}{X}^{(\ell)}}. \tag{5}$$

"Pooling" the marginal versions of this empirical characteristic function (using the assumption of same marginals) gives the following estimate of ϕ:

$$\hat{\phi}(t) = d^{-1} \sum_{j=1}^{d} \hat{\phi}_{X_j}(t) = (nd)^{-1} \sum_{\ell=1}^{n} \sum_{j=1}^{d} e^{itX_j^{(\ell)}}.$$

$\alpha_{j,k}$ are then taken to "make the model match the joint distribution as well as possible". In particular, given some norm $||\cdot||$ on $\mathbb{R}^d$, take the vector $\underset{\rightarrow}{\hat{\alpha}}$ of estimates

$\hat{\alpha}_{j,k}$ to be

$$\arg\min_{\underset{\rightarrow}{\alpha}} \left\| \hat{\phi}_{\underset{\rightarrow}{X}}(\underset{\rightarrow}{t}) - \left[\prod_{1 \leqslant j < k \leqslant d} \hat{\phi}(t_j + t_k)^{\alpha_{j,k}} \right] \left[\prod_{j=1}^{d} \hat{\phi}(t_j)^{1 - \sum_{k=1,\ k \neq j}^{d} \alpha_{j,k}} \right] \right\|.$$

Insight into the anticipated behavior of these estimates, in particular, how well this model fits various types of multivariate distributions having characteristic function $\psi(\underset{\rightarrow}{t})$, comes from studying the "theoretical version", $\underset{\rightarrow}{\tilde{\alpha}}$ of approximations $\tilde{\alpha}_{j,k}$, defined to be

$$\arg\min_{\underset{\rightarrow}{\alpha}} ||\psi(\underset{\rightarrow}{t}) - \phi_{\underset{\rightarrow}{X}}(\underset{\rightarrow}{t})||.$$

In Section 3.3 we show that under a limiting argument, the minimizing values $\underset{\rightarrow}{\hat{\alpha}}$ and $\underset{\rightarrow}{\tilde{\alpha}}$ are the empirical and theoretical (respectively) Pearson correlation coefficients. Thus we have developed generalizations of the notion of "measure of correlation" which can be viewed as more sensible than the Pearson version, as shown by the computational results of Section 4.

In many applications it is of interest to obtain an estimate of the joint density of the observations. For this we use an inversion formula on the fitted joint characteristic function (see [4, Theorem 2.3.1]). Let $\hat{\phi}_{\underset{\rightarrow}{X}}(\underset{\rightarrow}{t})$ be the fitted characteristic function obtained by substituting $\underset{\rightarrow}{\hat{\alpha}}$ and $\hat{\phi}(t)$ in (4). The estimate of $f_{\underset{\rightarrow}{X}}(\underset{\rightarrow}{x})$ is

$$\hat{f}_{\underset{\rightarrow}{X}}(\underset{\rightarrow}{x}) = \frac{1}{2\pi} \int_{|\underset{\rightarrow}{t}| < B} \exp(-i\, \underset{\rightarrow}{t}'\, \underset{\rightarrow}{x}) \hat{\phi}_{\underset{\rightarrow}{X}}(\underset{\rightarrow}{t})\, d\, \underset{\rightarrow}{t},$$

where B acts as a smoothing parameter, see e.g. [15, Section 2.7]. Our numerical approach is Monte Carlo integration, which involves generating $\underset{\rightarrow}{t_1}, \ldots, \underset{\rightarrow}{t_M}$ independent uniform vectors on $|\underset{\rightarrow}{t}| < B$, for large M and then approximating

$$\hat{f}_{\underset{\rightarrow}{X}}(\underset{\rightarrow}{x}) = \frac{1}{2\pi} \frac{1}{M} \mathrm{real} \left\{ \sum_{i=1}^{M} \exp(-i\, \underset{\rightarrow}{t_i}'\, \underset{\rightarrow}{x}) \hat{\phi}_{\underset{\rightarrow}{X}}(\underset{\rightarrow}{t_i}) \right\}.$$

3. Theoretical properties

3.1. Connection to infinitely divisible laws and cumulants

As pointed out by Olkin [13], any construction of multivariate distributions with given marginals has limitations, since they apply to many different situations. The restriction that $0 \leqslant \alpha \leqslant 1$ ensures that $\phi_{\underset{\rightarrow}{X}}$ given by (3) is indeed the characteristic function of a joint distribution for a large class of marginal distributions which includes the infinitely divisible laws (we do not assume existence of a density here). This class is reasonably rich, and includes many distributions recently used in nonparametric and parametric statistical modelling such as the stable and the self-decomposable distributions; see for example [1,5,18].

In general, our model assumes that the left-hand side of (3) is a bivariate characteristic function having the particular α-decomposition-type (see [4]) given by its right-hand side. We do not know which kind of marginals other than the univariate infinitely divisible give a valid multivariate model (4). However, we conjecture that (4) well approximates several models, as suggested by Theorem 9.2.1 in [4] and the results for uniform marginals obtained in Section 4.

Knowledge of the univariate characteristic function ϕ provides important information about (4) and conversely. Specifically, if the marginal characteristic function ϕ is infinitely divisible, so are the multivariate models (3) and (4). Conversely, if (4) is multivariate infinitely divisible with additional properties (see [4, Theorems 9.3.1 and 9.3.2]), under the assumption of common marginals, ϕ is also (univariate) infinitely divisible. Moreover, it is not difficult to see that if ϕ is a univariate stable (more generally self-decomposable) characteristic function then (4) gives a multivariate stable (self-decomposable) distribution. The specific connection with the multivariate Gaussian distribution is illustrated in the next section.

If ϕ is the (real) characteristic function of a symmetric univariate distribution, then the dependence model (4) is the (real) characteristic function of a symmetric multivariate random vector. On the other hand, if $\int_{-\infty}^{\infty} |\phi(t)|^{\alpha}\, dt$ is finite, then

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\phi_{\underset{\rightarrow}{X}}(t_1, t_2)|\, dt_1\, dt_2 < \infty,$$

and therefore, using Corollary 2.3.1 in [4], the bivariate density of (3) exists, and similarly for the general multivariate situation. From now on we will always assume the existence of the multivariate density, which exists for all nondegenerate multivariate self-decomposable distributions [19].

Instead of using characteristic functions, sometimes it is easier to work with the moment generating functions (when they exist)

$$\mathrm{M}_{\underset{\rightarrow}{X}}(\underset{\rightarrow}{t}) = \left[\prod_{1 \leqslant j < k \leqslant d} \mathrm{M}(t_j + t_k)^{\alpha_{j,k}}\right]\left[\prod_{j=1}^{d} \mathrm{M}(t_j)^{1-\sum_{k=1,\ k \neq j}^{d} \alpha_{j,k}}\right], \tag{6}$$

$$\mathrm{M}_{\underset{\rightarrow}{X}}(\underset{\rightarrow}{t}) = Ee^{\underset{\rightarrow}{t}'\underset{\rightarrow}{X}} = \int_{-\infty}^{\infty} e^{\underset{\rightarrow}{t}'\underset{\rightarrow}{x}} f_{\underset{\rightarrow}{X}}(\underset{\rightarrow}{x})\, dx_1 \cdots dx_d$$

and

$$\mathrm{M}(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx.$$

From [4, Theorem 9.2.1] it follows that if the distribution of the left-hand side of (4) has multivariate moment generating function $\mathrm{M}_{\underset{\rightarrow}{X}}$, then the moment generating function M of the distribution of ϕ also exists.

Norman L. Johnson has pointed out the following fact. Since the log of a characteristic function generates cumulants, our log model becomes

$$\log \phi_{\underset{\to}{X}}\left(\begin{pmatrix} t_1 \\ t_2 \end{pmatrix}\right) = \alpha \log \phi(t_1 + t_2) + (1-\alpha)\log[\phi(t_1)\phi(t_2)] \tag{7}$$

and therefore, the dependence model (3) has the interpretation that its cumulants are interpolated averages of cumulants from the total dependence model and the independent one.

We finally observe that expression (3) has recently been used—in a completely different context—by Houdré et al. [8], as a basic tool for proving correlation inequalities and in studying association of infinitely divisible random vectors.

3.2. Connection to the multivariate normal distribution

When the joint distribution is multivariate normal, with mean vector $\underset{\to}{0}$ and covariance matrix $\boldsymbol{\Sigma}$, the joint density is

$$f_{\underset{\to}{X}}(\underset{\to}{x}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}\underset{\to}{x}'\boldsymbol{\Sigma}^{-1}\underset{\to}{x}}.$$

Much insight about this distribution comes from its characteristic function

$$\phi_{\underset{\to}{X}}(\underset{\to}{t}) = e^{-\frac{1}{2}\underset{\to}{t}'\boldsymbol{\Sigma}\underset{\to}{t}}$$

(see [20, p. 28]).

If we use approach (4) to create a multivariate distribution from univariate standard normals, using the relationship $(t_j + t_k)^2 = t_j^2 + t_k^2 + 2t_jt_k$, we obtain

$$\begin{aligned}
&\left[\prod_{1\leqslant j<k\leqslant d} \phi(t_j + t_k)^{\alpha_{j,k}}\right]\left[\prod_{j=1}^{d} \phi(t_j)^{1-\sum_{k=1,\ k\neq j}^{d}\alpha_{j,k}}\right] \\
&\quad = \exp\left\{-\frac{1}{2}\left[\sum_{1\leqslant j<k\leqslant d} \alpha_{j,k}(t_j+t_k)^2 + \sum_{j=1}^{d}\left(1 - \sum_{k=1,\ k\neq j}^{d}\alpha_{j,k}\right)t_j^2\right]\right\} \\
&\quad = \exp\left\{-\frac{1}{2}\left[\sum_{1\leqslant j<k\leqslant d} 2\ \alpha_{j,k}t_jt_k + \sum_{j=1}^{d} t_j^2\right]\right\} \\
&\quad = e^{-\frac{1}{2}\underset{\to}{t}'\boldsymbol{\Sigma}\underset{\to}{t}},
\end{aligned}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix with entries $\alpha_{j,k}$,

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \alpha_{12} & \cdots & \alpha_{1d} \\ \alpha_{12} & 1 & & \\ \vdots & & \ddots & \vdots \\ \alpha_{1d} & & \cdots & 1 \end{pmatrix}.$$

Hence our model is multivariate Gaussian when we start with Gaussian marginals, and the $\alpha_{j,k}$ are just the usual correlations.

Conversely, if (4) is a multivariate Gaussian model, then using Theorem 9.3.1 in [4] it follows that ϕ is the characteristic function of a univariate Gaussian distribution.

3.3. Connection to Pearson's correlation

For simplicity and clarity only the case $d = 2$ is handled here, but the extension to general d is straightforward. In this section we do not need to assume that the marginal distributions are the same, but instead only need common second moments. In particular, assume $EX_1 = EX_2 = 0$, $EX_1^2 = EX_2^2 = 1$, and all third moments exist. Then, as $t_1, t_2 \to 0$, standard Taylor expansion gives

$$E(e^{it_jX_j}) = 1 - \tfrac{1}{2}t_j^2 + O(|t_j|^3),$$

$$E(e^{i(t_1+t_2)X_1}) = 1 - \tfrac{1}{2}t_1^2 - \tfrac{1}{2}t_2^2 - t_1t_2 + O(|t_1|^3) + O(|t_2|^3),$$

$$E(e^{i(t_1X_1+t_2X_2)}) = 1 - \tfrac{1}{2}t_1^2 - \tfrac{1}{2}t_2^2 - t_1t_2E(X_1X_2) + O(|t_1|^3) + O(|t_2|^3).$$

The relation (as $s \to 0$)

$$(1 + s + o(s))^\beta = 1 + \beta s + o(s),$$

together with straightforward algebra gives

$$|E(e^{i(t_1X_1+t_2X_2)}) - [E(e^{i(t_1+t_2)X_1})]^\alpha [E(e^{it_1X_1})E(e^{it_2X_2})]^{(1-a)}|$$
$$= |t_1t_2(\alpha - E(X_1X_2)) + O(|t_1|^3) + O(|t_2|^3)|.$$

Thus, when t_1, t_2 are near to 0, we obtain that $\tilde{\alpha}$ is near to $E(X_1X_2)$, the Pearson correlation coefficient under these assumptions.

This development has been in terms of the "theoretical" correlation coefficient, but the same calculation also applies to the "empirical" version, by replacing the expectation operator with its sample version, i.e. by replacing the operation

$$Eg(\underset{\rightarrow}{X}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\underset{\rightarrow}{x}) f_{\underset{\rightarrow}{X}}(\underset{\rightarrow}{x})\, dx_1 \cdots dx_d$$

with

$$\hat{E}g(\underset{\rightarrow}{X}) = n^{-1} \sum_{\ell=1}^{n} g(\underset{\rightarrow}{X}^{(\ell)})$$

at all points. The assumption of common marginal mean 0 and variance 1 is achieved by "standardizing", i.e. by assuming that the $X_j^{(\ell)}$ come from data $Y_j^{(\ell)}$ as

$$X_j^{(\ell)} = (Y_j^{(\ell)} - \bar{Y}_j)/\hat{\sigma}_j \tag{8}$$

for $\ell = 1, \ldots, n$ and $j = 1, \ldots, d$, where $\bar{Y}_j = n^{-1} \sum_{\ell=1}^{n} Y_j^{(\ell)}$ and

$$\hat{\sigma}_j = \left[n^{-1} \sum_{\ell=1}^{n} (Y_j^{(\ell)} - \bar{Y}_j)^2 \right]^{1/2}.$$

Thus, in the limit as $t_1, t_2 \to 0$, we get

$$\hat{\alpha} = n^{-1} \sum_{\ell=1}^{n} X_1^{(\ell)} X_2^{(\ell)} = n^{-1} \sum_{\ell=1}^{n} (Y_1^{(\ell)} - \bar{Y}_1)_1^{(\ell)} (Y_2^{(\ell)} - \bar{Y}_2)_2^{(\ell)} / (\hat{\sigma}_1 \hat{\sigma}_2),$$

which is the empirical Pearson correlation coefficient.

3.4. Connection to other dependence concepts

If $0 \leqslant \alpha_j \leqslant 1$, $j = 1, \ldots, d$ and ϕ is infinitely divisible, model (4) gives the characteristic function of an *associated* random vector. Recall that a random vector $\underset{\to}{X}$ is associated if $\mathrm{Cov}(G_1(\underset{\to}{X}), G_2(\underset{\to}{X})) \geqslant 0$, for all componentwise nondecreasing functions $G_1, G_2 : \mathbb{R}^d \to \mathbb{R}$, for which the covariance exists (see [8,17], for the association of infinitely divisible random vectors).

Recently, Rosinski and Zak [16] have studied a useful measure of dependence for a general pair of infinitely divisible random variables. Namely, if (X_1, X_2) is an infinitely divisible random vector with characteristic function $\phi\binom{t_1}{t_2}$, the *codifference* $\tau(X_1, X_2)$ is defined as

$$\tau(X_1, X_2) = \log \phi \begin{pmatrix} 1 \\ -1 \end{pmatrix} - \log \phi \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \log \phi \begin{pmatrix} 0 \\ -1 \end{pmatrix}. \tag{9}$$

For the Gaussian case the codifference is the usual correlation coefficient and in general, it holds that X_1 and X_2 are independent if and only if $\tau(X_1, X_2) = \tau(X_1, -X_2) = 0$.

When (X_1, X_2) follows model (3),

$$\tau(X_1, X_2) = -\alpha \log \phi(1)\phi(-1),$$

that is, the codifference is proportional to α. In general, if $(X_1, \ldots, X_d)$ follows model (4), $\tau(X_j, X_k) = -\alpha_{j,k} \log \phi(1)\phi(-1)$.

4. Computational study

To gain insight about the properties of our common marginal dependence model, and the proposed estimators, we conducted a computational study which include both theoretical computations and simulations. Of particular interest is how well our model (3) approximates joint distributions that are not of exactly that form.

The examples considered were bivariate distributions, normalized so that $EX_1 = EX_2 = 0$ and $\mathrm{var}(X_1) = \mathrm{var}(X_2) = 1$. The dependence structures were:

D1. X_1, X_2, independent.

D2. $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \Sigma^{-1/2} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$, for $EZ_1 = EZ_2 = 0$, $\mathrm{var}(Z_1) = \mathrm{var}(Z_2) = 1$, and $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$. This intended to be a "moderate positive correlation" model.

D3. Same as D2, except $\Sigma = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$. This is intended to be a "high positive correlation" model.

D4. A mixture of the degenerate distribution $X_1 = X_2$ (with probability 1/2) and the X_1, X_2 independent distribution (with probability 1/2). This distribution has positive correlation of a very nonstandard type.

D5. A mixture of the two degenerate distributions $X_1 = X_2$ (with probability 1/2) and $X_1 = -X_2$ (with probability 1/2). This distribution is supported on the 45° lines in the plane and is a very nonstandard distribution. Even the notion of "correlation" could be defined in many quite different ways.

D6. Same as D2, except $\Sigma = \begin{pmatrix} 1 & -0.9 \\ -0.9 & 1 \end{pmatrix}$. This is intended to be a "high negative correlation" model. Note that our model is not expected to work at all here, because it is specifically designed for positive correlation.

The marginal distributions considered were:

M1. Gaussian.

M2. Laplace (also called the "double exponential"). Intended to represent nonGaussian shapes.

M3. Uniform. Even further from the Gaussian in shape.

M4. Exponential. A different type of nonGaussian, and also asymmetric.

Four natural norms were considered for the estimation procedure of Section 2:

N1. Standard L^2: $||f_1 - f_2||_1^2 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |f_1(\underset{\rightarrow}{t}) - f_2(\underset{\rightarrow}{t})|^2 \, d\underset{\rightarrow}{t}$,

N2. Weighted L^2: $||f_1 - f_2||_2^2 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |f_1(\underset{\rightarrow}{t}) - f_2(\underset{\rightarrow}{t})|^2 e^{(\underset{\rightarrow}{t}' \underset{\rightarrow}{t})} \, d\underset{\rightarrow}{t}$, (this puts more weight on the origin)

N3. Weakly weighted L^2:

$$||f_1 - f_2||_3^2 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |f_1(\underset{\rightarrow}{t}) - f_2(\underset{\rightarrow}{t})|^2 e^{(\underset{\rightarrow}{t}' \underset{\rightarrow}{t})/2} \, d\underset{\rightarrow}{t},$$

(this uses weights in between $||\cdot||_1$ and $||\cdot||_2$)

N4. Evaluation at 0: $||f_1 - f_2||_4 = |f_1(\underset{\rightarrow}{0}) - f_2(\underset{\rightarrow}{0})|$ (this is only a seminorm, but is included to study the connection to Pearson's correlation discussed in Section 3.3).

We have done both theoretical and empirical computations for all combinations of the above settings, but just show a few here, chosen to highlight the main ideas, to save space. The case of the mixture dependence (D4), with the Laplace marginal distributions (M2) was fairly representative (our model gave better performance in most cases). Fig. 1 shows the theoretical joint characteristic function, $\phi_{\underset{\rightarrow}{X}}(\underset{\rightarrow}{t})$, together with the characteristic function of our model as defined in (3), fit by the L^2

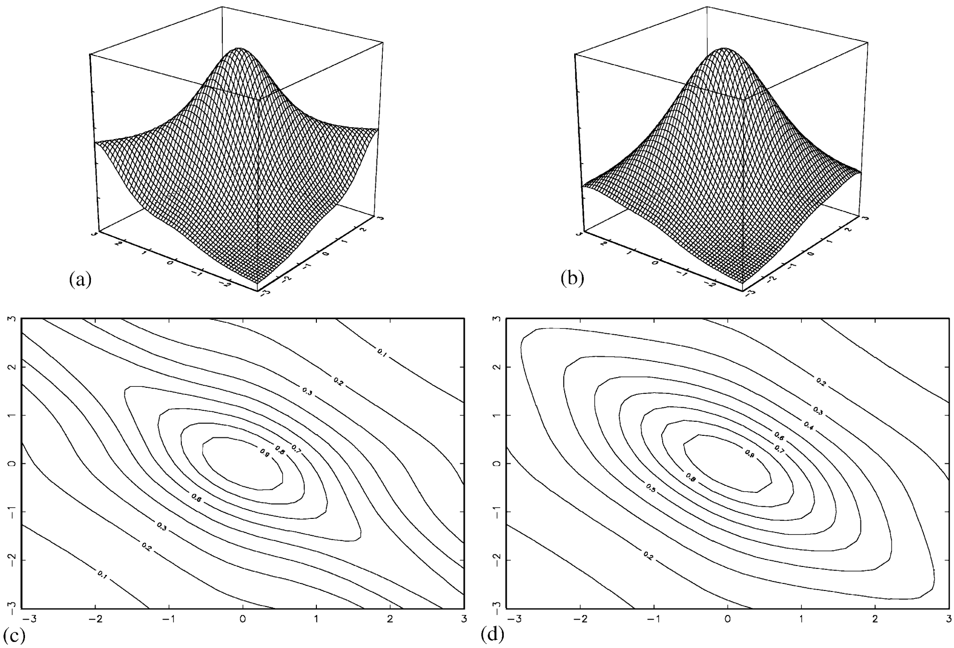


Fig. 1. Characteristic functions, for the underlying joint distribution. Left-hand panels are for the true underlying joint distribution, right-hand panels are for the model of the form (3) fit by norm (N1). Top panels are mesh plots, and bottom panels are contour plots of the same surfaces. For the mixture joint distribution (D4), with common Laplace marginal distributions (M2).

norm (N1). The approximation is reasonable, but not perfect. We observed substantially better approximation for the more standard dependence models (D1), (D2) and (D3), and also for the Gaussian common marginal distributions (M1).

Insight into the behavior of the different norms (N1)–(N4) is given in Fig. 2, which shows these as a function of the parameter α. In what follows $\tilde{\alpha}_i$ denotes the minimizing value under norm Ni. The theoretical value of Pearson's correlation for this joint distribution is 0.5, so the norm (N4) works as expected, with $\tilde{\alpha}_4 = 0.499$. When "correlation" is instead taken to be the α that minimizes other norms, the value is somewhat different. Most different is the standard L^2 norm (N1), with $\tilde{\alpha}_1 = 0.714$. Putting more weight on the origin gives something in between (N1) and (N4), so it is not surprising that the minimizers are $\tilde{\alpha}_2 = 0.569$ and $\tilde{\alpha}_3 = 0.609$ (note (N3) is "between" (N1) and (N2)).

An interesting variation on Fig. 2 showed up for the unusual dependency model (D5), with Gaussian marginal distributions (M1), as shown in Fig. 3. For this model the notion of "correlation" that follows from the standard L^2 norm has two solutions $\tilde{\alpha}_1 = \pm 0.790$. This is consistent with the fact that "positive" and "negative" correlations are not simply defined notions for this model (where the bivariate probability puts mass symmetrically on the two 45° lines in the plane).

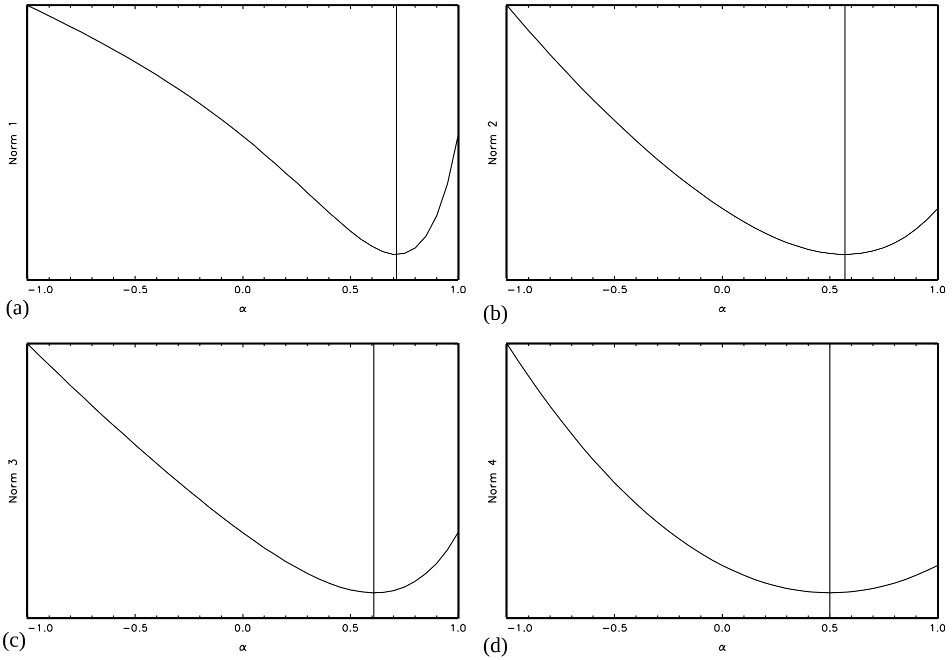


Fig. 2. Theoretical norms between joint characteristic functions and our common marginal dependency model, as a function of the dependency parameter α. Norms are (N1) for (a), (N2) for (b), (N3) for (c), and (N4) for (d). For the mixture joint distribution (D4), with common Laplace marginal distributions (M2).

However, the other three norms all result in $\tilde{\alpha}_i = 0$ (only (N3) is shown but the others are similar) which is also a sensible definition of "correlation". Insight as to why these answers are different comes from studying the mesh plots of the characteristic functions, as shown in the lower row. Note that the joint characteristic function is approximately radially symmetric near the origin, but has distinct "shoulders" away from the origin. The standard L^2 norm is more strongly influenced by points away from the center, so these "shoulders" give the multiple minima apparent in Fig. 3a (which reflect fitting quite elliptical Gaussian distributions, of the type shown in Fig. 3d). But the other norms are more strongly influenced by points near the origin, where the radially symmetric part is dominant, so the best Gaussian fit is spherical.

Next we studied the performance of the empirical version of our model, in these various contexts. For comparison with the theoretical case, we again focus on the nonstandard mixture distribution (D4) and common Laplace marginals (M2). This case was again fairly representative. Using one pseudo-data set of size $n = 100$, gave the empirical version of Fig. 1 that is shown in Fig. 4. This is the empirical joint characteristic function, $\hat{\phi}_{\vec{X}}(\vec{t})$, together with the empirical fit of our model, fit by the L^2 norm (N1). Again as in Fig. 1, the approximation is reasonable, but not perfect. Again we observed substantially better approximation for the more standard

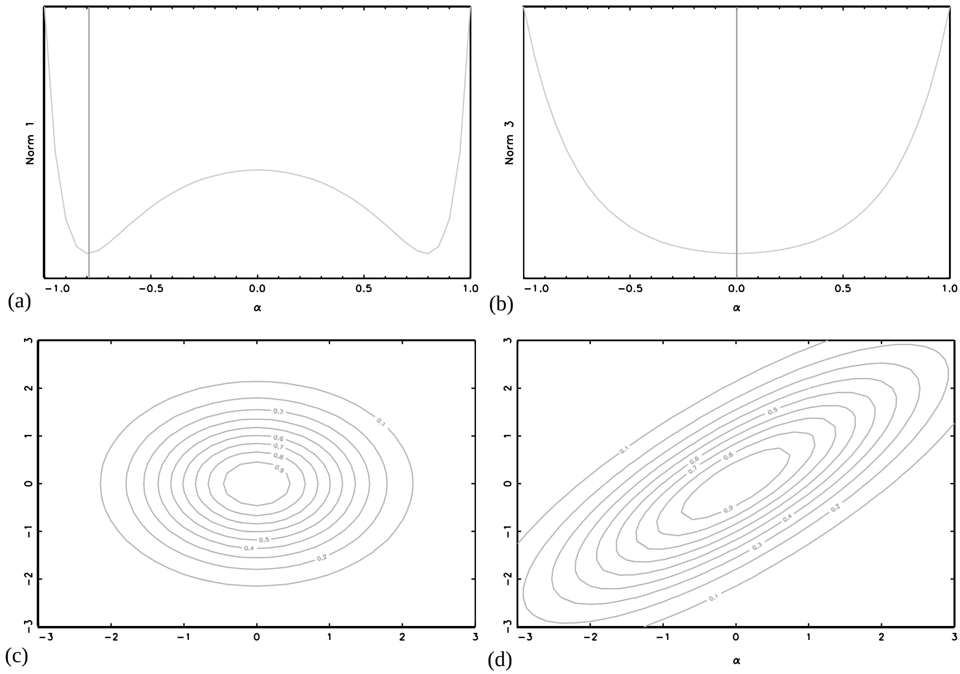


Fig. 3. Top row shows theoretical norms between joint characteristic functions and our common marginal dependency model, as a function of the dependency parameter α. Norms are the standard L^2 norm (N1) for (a), the weighted L^2 norm (N3) for (b). Large differences are explained by characteristic functions shown in the bottom row, joint for (c), and our model, fit by (N1) for (d).

dependence models (D1), (D2) and (D3), and when the common marginals were Gaussian (M1).

The empirical versions of Fig. 2 are not worth the space, because the norms were quite similar to each other in this case (and had a shape very similar to what is seen in Fig. 2). The analogs of this for most other settings were similar. Even the analog of Fig. 3 had a similar "single minimum" shape (because the special symmetry that created the two minima disappeared in the empirical characteristic function).

We also studied some sampling properties of our model via a simulation study. The results are reported in [11]. The main insight gained is that norm (N3) has marginal smaller variability, so we prefer it. Furthermore, we conclude that our dependence model is not robust against violation of the assumption of positive correlation, and recommend that adjusting for this by changing appropriate signs of variables is worthwhile. We also studied the performance of our final marginal density estimates $\hat{f}_{\underset{\rightarrow}{X}}(\underset{\rightarrow}{x})$, defined in Section 2.2. The estimation was very good for the case of Gaussian (M1) marginals, but not so good for the Laplace (M3) marginals.

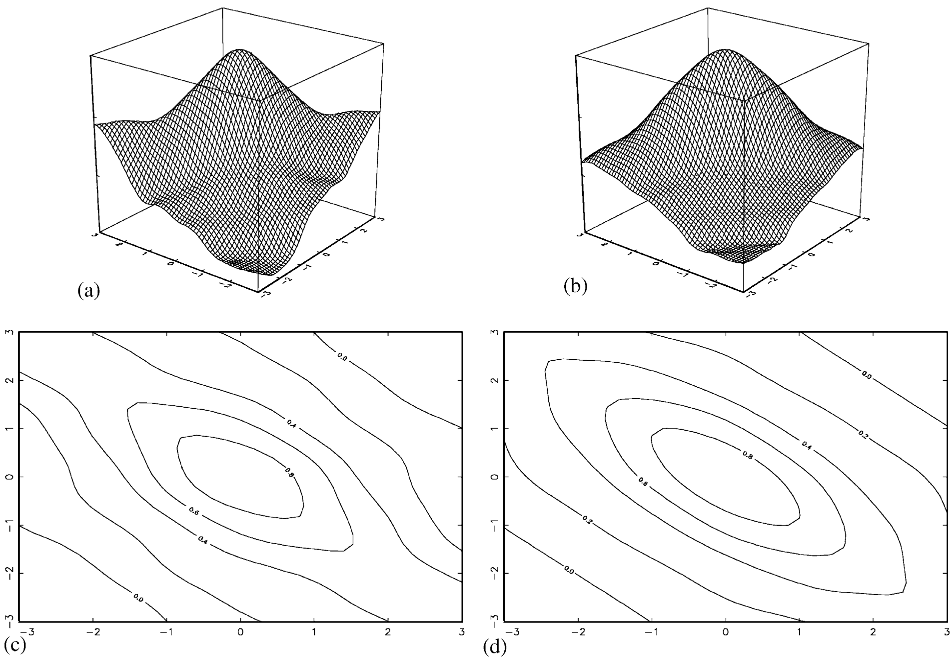


Fig. 4. Empirical characteristic functions, based on a single pseudo data set of size $n = 100$. Left-hand panels are for the empirical joint distribution, right-hand panels are for the model of the form (3) fit by norm (N3). Top panels are mesh plots, and bottom panels are contour plots of the same surfaces. For the mixture joint distribution (D4), with common Laplace marginal distributions (M2).

5. Data application

5.1. CO and ozone data from a monitoring network

When modelling a single variable in environmental monitoring, it is often appropriate to assume common marginals; moreover, the Gaussian assumption is not often met so one must build a different dependence model in order to estimate the associated multivariate density. In this section we illustrate our multivariate modelling by considering data at four stations of the automatic air monitoring network in Mexico City (known as RAMA), examining two pollution variables, one at a time: CO and ozone. Additional RAMA stations were not considered here because they did not provide an adequate amount of data which was complete in these variables. We presently focus only on fitting the model; this is a first step towards further and more general analysis in this setting, for example, in finding the least informative station using the Shannon index as a measure of performance (see [3,14]).

The RAMA stations considered are: Merced (MER), Pedregal (PED), Cerro de la Estrella (EST) and Plateros (PLA). PED and PLA are located SW, EST is NE, and MER is near the center of the city. The data consist of vectors of dimension $d = 4$, with each entry corresponding to the weekly maxima of CO and ozone at each station for the years 1988–1993. We consider only those weeks for which complete observations (simultaneously in all stations) were available; 122 observations for CO, and 128 for ozone were obtained. Stationwise scatterplots show distinguishing features for which our model seems especially appropriate: positive correlations, and common marginals (of an unspecified nature). In both examples below, the parameters in the characteristic function (4) are estimated using norm (N3).

When considering CO, preliminary inspection showed us clearly that common marginals are plausible in three of the stations considered, so our first example concerns CO observations disregarding PED. The exclusion of PED may be debatable, because the degree to which its distribution does not conform to the other three is not serious; but we prefer to be cautious in this illustrative example. Upon standardizing each data entry by substracting marginal means and dividing by marginal standard deviations, we observe that the marginals are approximately the same, and heavily right-skewed (i.e. far from Gaussian). The estimated α values are

	EST	PLA
MER	0.20	0.12
EST		0.44

Estimates of α are nonnegative, as expected in an air pollution monitoring network, and correlation is stronger between EST and PLA despite the fact that there is a large distance between these two stations. A possible explanation for this is that station PLA is aligned with respect to EST in a north-easterly direction, so that correlation could be induced by transport due to dominant trade winds.

Our second example concerns Ozone observations at the four stations. In this case we observe that the marginal distributions approximately follow the same distribution in all four stations after appropriate standardizing. Results of estimates of α are

	PED	EST	PLA
MER	−0.01	0.57	0.03
PED		0.06	0.51
EST			0.27

Note, in contrast to CO, that zero correlation between some pairs of stations (MER–PED, MER–PLA, PED–EST) is suggested, which correspond to pairs of stations which are geographically far apart from each other, and in different types of location (residential vs. industrial).

5.2. Graphical results

A specific goodness of fit tool has not yet been developed especially for this model, so for this purpose we use the following ad hoc graphical device. Let $\underset{\rightarrow}{v}$ represent a direction in $\mathbb{R}^d$ with $|\underset{\rightarrow}{v}| = 1$. We compare real and imaginary parts of the empirical characteristic function (5) and the fitted characteristic function (4), by plotting these at $t\underset{\to}{v}$, for $-2 \leqslant t \leqslant 2$ for different choices of $\underset{\to}{v}$. When $\underset{\to}{v}$ is made equal to each of d orthogonal directions—principal components based on the sample covariance matrix of data, for example—this amounts to comparing two characteristic functions in d variables along orthogonal slices in $\mathbb{R}^d$. If $\underset{\to}{v}$ was instead taken to be standard directions, the d plots obtained would compare each of the marginal distributions; but note that at least one additional plot in a nonstandard direction is required in order to better resolve dependence structure, because plots for a distribution with common marginals and *any* correlation structure would always show agreement in the standard directions. The alternative graphical display used in Fig. 4 also

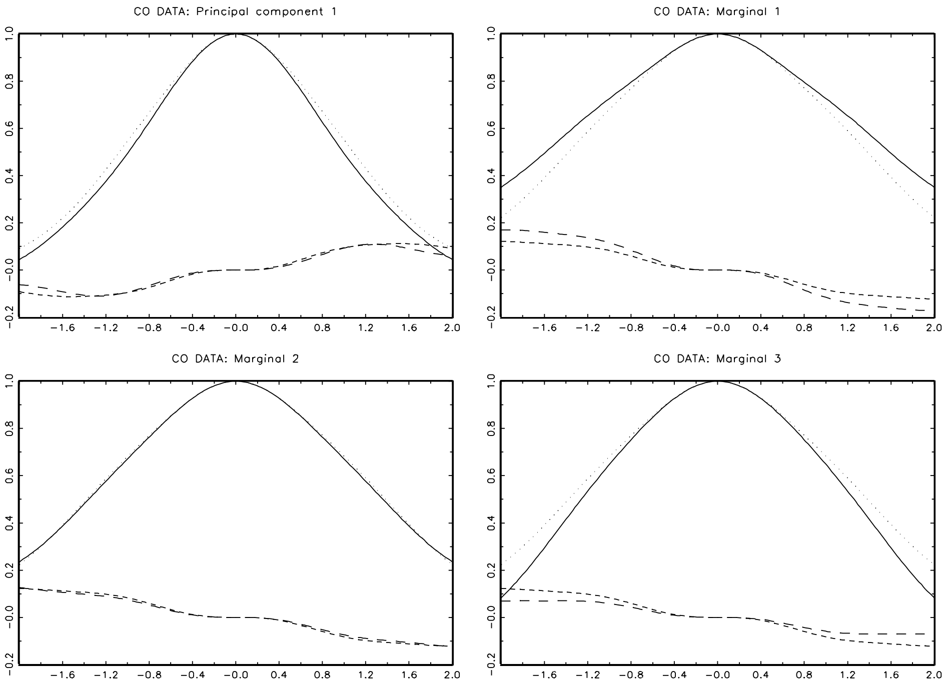
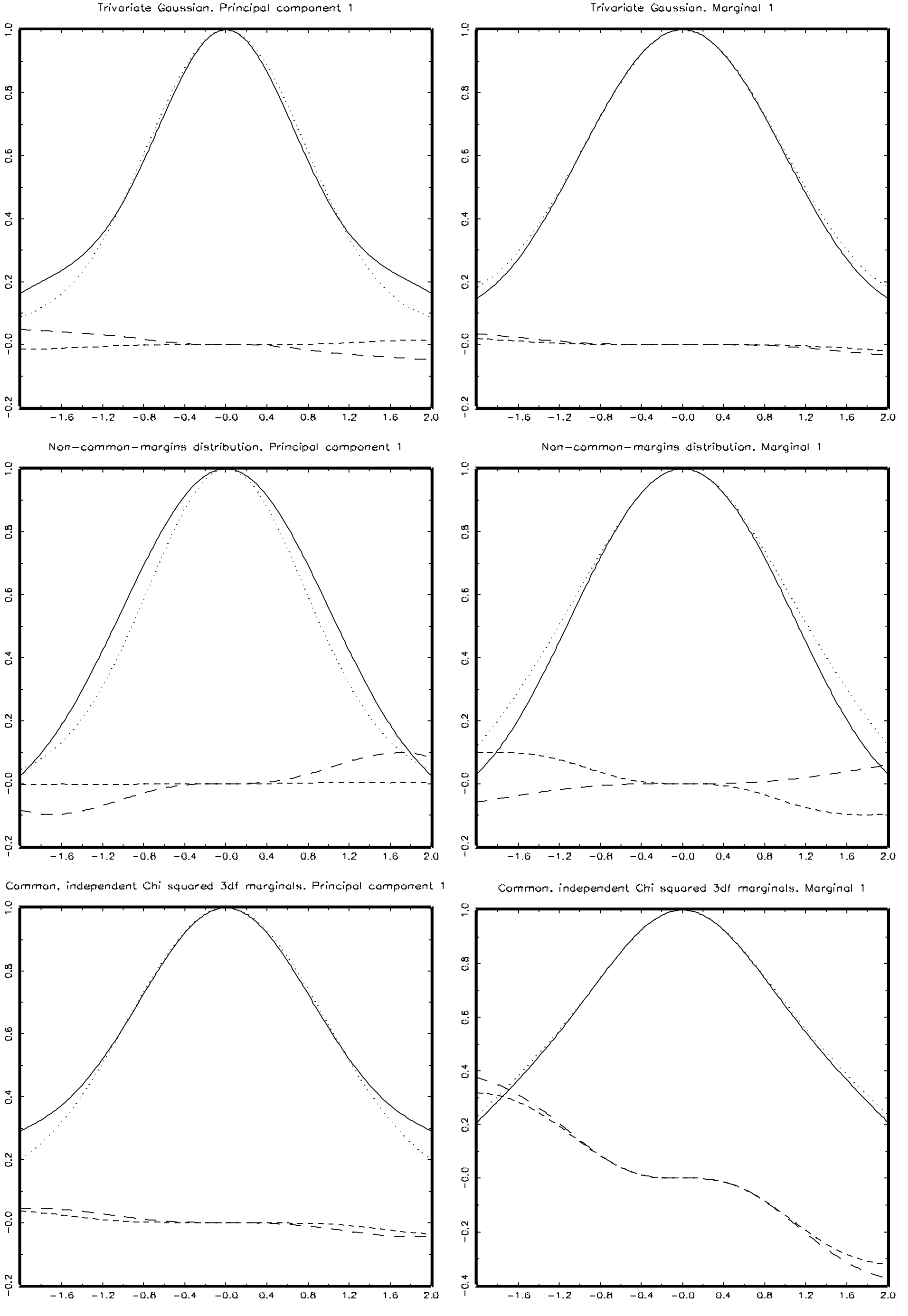


Fig. 5. CO Data. Comparison of real (solid) and imaginary (long dashed) parts of empirical and real (dotted) and imaginary (short dashed) parts of fitted characteristic functions, corresponding to the first principal component and the three standard directions.

Fig. 6. Comparison of real (solid) and imaginary (long dashed) parts of empirical and real (dotted) and imaginary (short dashed) parts of fitted characteristic functions, corresponding to the first principal component and the three standard directions, for three sets of standardized three-dimensional data of size $n = 125$. The first row originated from a trivariate Gaussian distribution with correlation among entries; second row from i.i.d. realizations of a $(N(0,1), U(0,1), \chi_1^2)$ vector with independent entries; third row from i.i.d. realizations of independent common χ_3^2 random variables.



Trivariate Gaussian. Principal component 1
Trivariate Gaussian. Marginal 1
Non-common-margins distribution. Principal component 1
Non-common-margins distribution. Marginal 1
Common, independent Chi squared 3df marginals. Principal component 1
Common, independent Chi squared 3df marginals. Marginal 1

Figure 6

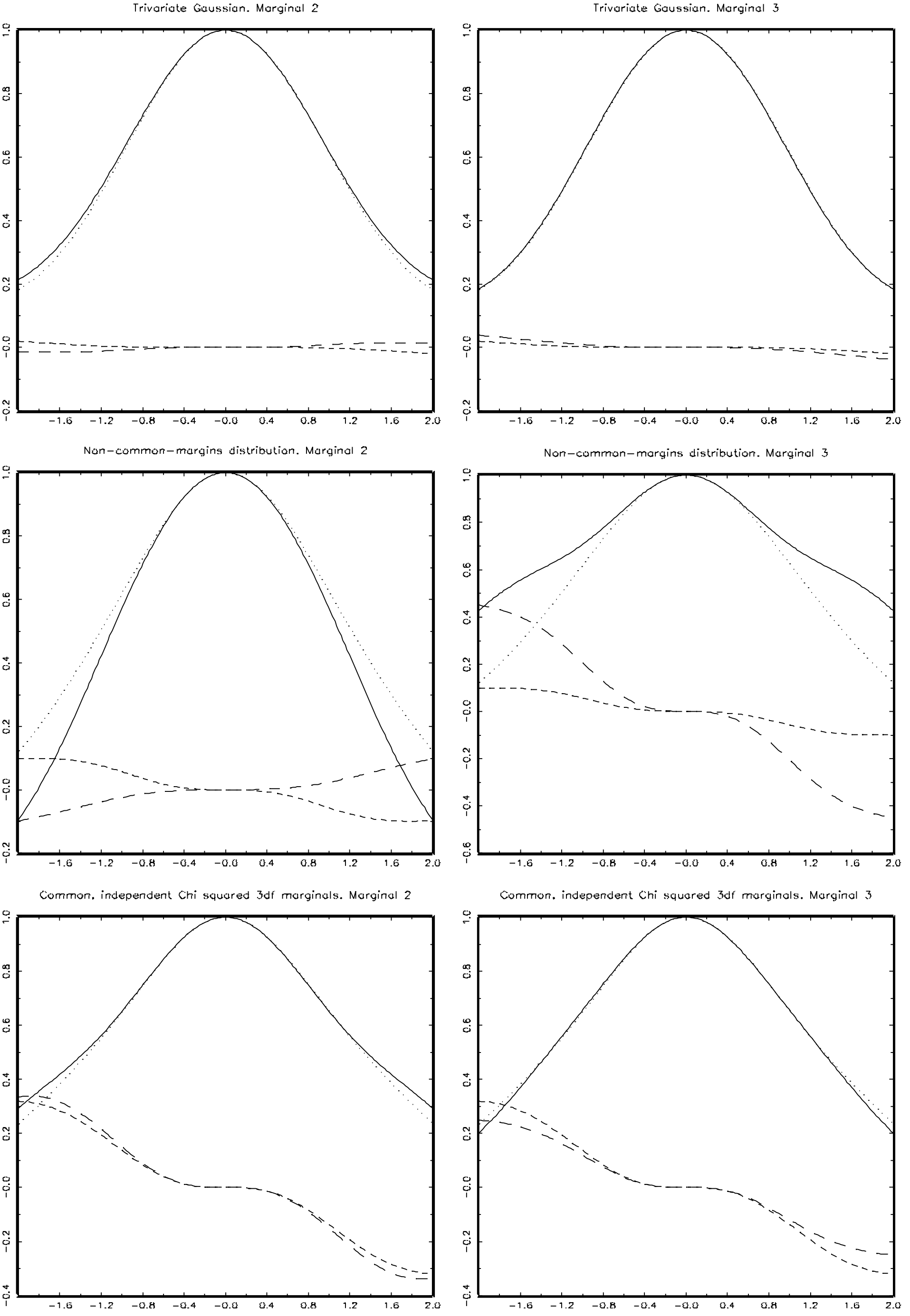


Fig. 6 (*continued*).

compares an empirical and a fitted joint characteristic function, but it is well suited for two dimensions only.

Fig. 5 shows the plots which result from the three-dimensional fit to the CO data, letting $\underset{\rightarrow}{v}$ be the first principal component and the three standard directions. Notice that there is general agreement in these plots. The nonzero imaginary parts in the marginal plots show that this distribution is asymmetrical.

Similar plots constructed for ozone as a result of the four-dimensional fit in the directions of all four principal components and standard directions give even better agreement, and are not shown here. In these, imaginary parts of characteristic functions are practically zero, indicating that the marginal distributions involved are more symmetric; furthermore, the shape of the real part of the marginal characteristic functions suggests that ozone may be described by a distribution which is nearly multivariate Gaussian.

In order to get a feel for what this graphical method is doing, consider three standardized three-dimensional data sets of size 125 (this value was chosen because it is central to our real data examples): the first originated from a Gaussian distribution with correlation among entries; the second from i.i.d. realizations of a $(N(0,1), U(0,1), \chi_1^2)$ vector with independent entries (that is, a joint distribution which does not even share common marginals); and the third from i.i.d. realizations of independent common χ_3^2 random variables. Assuming model (4), estimating α parameters, and constructing the described plots in each case produces Fig. 6. Note that there is a general agreement in characteristic functions in the first and last data sets, whereas in the second there clearly is not. An important aspect of Fig. 6 is that it illustrates by how much these plots (for the given sample size) can differ for data whose model is (4). Also, it is here evident that inspection of various directions is necessary, because there is not a single direction which tells the whole story. Discrepancies for the first and third simulated data sets are qualitatively similar to the ones obtained for CO in Fig. 5 and for ozone (not shown); we interpret this to mean that there are not severe objections to the validity of model (4) in these cases.

6. Conclusions

We constructed a semiparametric model for multivariate observations when the marginals are the same. The model incorporates parameters which give a new notion of dependence for a wide family of distributions.

The model (4) has nice properties and is useful when the marginals are infinitely divisible. It enables easy multivariate modelling with common marginals taking into account dependence parameters between all pairs of marginals, and reduces to the multivariate Gaussian in the case of Gaussian marginals.

To estimate the dependence parameters we considered the empirical characteristic functions in such a way that the assumption of same marginals is involved. A computational study was conducted to evaluate several norms used in the estimation of the dependence parameters. Our recommendation is that norm (N3) should be used for this purpose. The proposed method was applied to two sets of pollution data from an environmental monitoring network, showing that the proposed distribution has potential for modelling this type of environmental data.

References

[1] O.E. Barndorff-Nielsen, Probability and statistics: selfdecomposability, finance and turbulence, in: L. Accardi, C.C. Heyde (Eds.), Probability Towards 2000, Lecture Notes in Statistics, 128, Springer, New York (1998) 47–57.

[2] O.E. Barndorff-Nielsen, P. Blaesild, V. Seshadri, Multivariate distributions with generalized inverse Gaussian margins, and associated Poisson mixtures, Canad. J. Statist. 20 (1992) 109–120.

[3] W.F. Caselton, J.V. Zidek, Optimal monitoring network design, Statist. Probab. Lett. 2 (1984) 223–227.

[4] R. Cuppens, Decomposition of Multivariate Distributions, Academic Press, New York, 1975.

[5] P. Damien, P.W. Laud, A.F.M. Smith, Approximate random variable generation from infinitely distributions with applications to Bayesian inference, J. Roy. Statist. Soc. B 57 (1995) 547–563.

[6] T.W. Epps, Characteristic functions and their empirical counterparts: geometrical interpretations and applications to statistical inference, Amer. Statist. 47 (1993) 33–38.

[7] C. Genest, R.J. Mackay, The joy of copulas: bivariate distributions with uniform marginals, Amer. Statist. 40 (1986) 280–283.

[8] C. Houdré, V. Pérez-Abreu, D. Surgailis, Interpolation, correlation identities and inequalities for infinitely divisible variables, J. Fourier Anal. Appl. 4 (1998) 651–668.

[9] T.P. Huchinson, C.D. Lai, Continuous Bivariate Distributions, Emphasizing Applications, Rumsby Scientific Publishing, Adelaide, 1990.

[10] K.V. Mardia, Families of Bivariate Distributions, Griffin, London, 1970.

[11] J.S. Marron, M. Nakamura, V. Pérez-Abreu, Semi-parametric multivariate modelling when the marginals are the same, Comunicaciones Técnicas del CIMAT, I–02–33 (PE), 2002.

[12] A.W. Marshall, I. Olkin, Families of multivariate distributions, J. Amer. Statist. Assoc. 83 (1988) 834–841.

[13] I. Olkin, Multivariate nonnormal distributions and models of dependency, Multivariate Anal. Appl. IMS Lecture Notes 24 (1994) 37–53.

[14] V. Pérez-Abreu, J.E. Rodríguez, Index of effectiveness of a multivariate environmental monitoring network, Environmetrics 7 (1996) 489–501.

[15] B.L.S. Prakasa Rao, Nonparametric Functional Estimation, Academic Press, New York, 1983.

[16] J. Rosiński, T. Zak, The equivalence of ergodicity and weak mixing for infinitely divisible processes, J. Theor. Probab. 10 (1997) 73–86.

[17] G. Samorodnitsky, Association of infinitely divisible random vectors, Stochastic Processes Appl. 55 (1995) 45–55.

[18] G. Samorodnitsky, M. Taqqu, Stable Non-Gaussian Random Processes, Chapman & Hall, New York, 1994.

[19] K. Sato, Absolute continuity of multivariate distributions of class *L*, J. Multivariate Anal. 12 (1982) 89–94.

[20] Y.L. Tong, The Multivariate Normal Distribution, Springer, New York, 1990.

[21] A.I. Yashin, I.A. Iachine, Survival of related individuals: an extension of some fundamental results of heterogeneity analysis, Math. Popul. Studies 5 (1995) 321–339.

[22] A.I. Yashin, J.W. Vaupel, I.A. Iachine, Correlated frailty models: an advantageous approach to the analysis of bivariate data, Math. Popul. Studies 5 (1996) 1–15.