



HAL
open science

On a construction of multivariate distributions given some multidimensional marginals

Nabil Kazi-Tani, Didier Rullière

► **To cite this version:**

Nabil Kazi-Tani, Didier Rullière. On a construction of multivariate distributions given some multidimensional marginals. 2017. hal-01575169v2

HAL Id: hal-01575169

<https://hal.science/hal-01575169v2>

Preprint submitted on 20 Nov 2018 (v2), last revised 1 Mar 2019 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On a construction of multivariate distributions given some multidimensional marginals.*

Nabil Kazi-Tani [†] Didier Rullière [‡]

20/11/2018

Abstract

In this paper, we investigate the link between the joint law of a d -dimensional random vector and the law of some of its multivariate marginals. We introduce and focus on a class of distributions, that we call *projective*, for which we give detailed properties. This allows us to obtain necessary conditions for a given construction to be projective. We illustrate our results by proposing some theoretical projective distributions, as elliptical distributions or a new class of distribution having given bivariate margins. In the case where the data do not necessarily correspond to a projective distribution, we also explain how to build proper distributions while checking that the distance to the prescribed projections is small enough.

Keywords: Multidimensional marginals, Copulas, Elliptical Distributions.

AMS Subject Classification(2010): 60E05; 60E10; 62H05; 62H20.

1 Introduction

Multivariate analysis is now a cornerstone in probability theory. Constructing multivariate distributions from given marginals is mathematically interesting on its own, but also has huge impact in practical problems.

Let $d \geq 1$ be an integer and let X_1, \dots, X_d be continuous random variables with cumulative distribution functions (CDFs) F_1, \dots, F_d , respectively. Sklar's theorem [35] states that the joint distribution H of (X_1, \dots, X_d) can be written, for all $(x_1, \dots, x_d) \in \mathbb{R}^d$ in the form

$$H(x_1, \dots, x_d) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d) = C(F_1(x_1), \dots, F_d(x_d)),$$

where C is a copula function, i.e. the cumulative distribution function of a probability measure on \mathbb{R}^d whose marginals are uniform on $[0, 1]$. Copula models have shown their interest in particular because, as it can be seen from the previous equation, they separate the study of the margins and the study of the dependence structure.

In this paper, we want to investigate the link between the joint law of a d -dimensional random vector and the law of its multivariate marginals. For any subset $K = (j_1, \dots, j_k)$ of $\{1, \dots, d\}$ with cardinal k , and any random vector $X = (X_1, \dots, X_d)$, we will write X_K the random vector with values in \mathbb{R}^k given by $(X_{j_1}, \dots, X_{j_k})$ and F_K the cumulative distribution function of X_K .

*This manuscript version: 20/11/2018 16:13:50.

[†]Laboratoire SAF, Université Lyon I, France, nabil.kazi-tani@univ-lyon1.fr

[‡]Laboratoire SAF, Université Lyon I, France, didier.rulliere@univ-lyon1.fr

We will abuse notations and call F_K a probability distribution on \mathbb{R}^k . Let $n \leq 2^d$ be a positive integer and let K_1, \dots, K_n be n subsets of $\{1, \dots, d\}$ with cardinals k_1, \dots, k_n .

A question that has been extensively studied in the literature is the following: given n probability measures P_1, \dots, P_n such that P_i is a probability measure on \mathbb{R}^{k_i} , is it possible to construct a probability measure F on \mathbb{R}^d such that

$$F_{K_i} = P_i \text{ for each } i = 1, \dots, n? \tag{1.1}$$

The existence of such a measure F is not guaranteed. In the case where the subsets $\{K_i, i = 1, \dots, n\}$ are disjoint, the product distribution guarantees its existence. One could also try to extend the notion of Copula function to the case of non overlapping multidimensional marginals: Genest & al. [15] showed that this approach is useless, since it only allows to model the product distribution. More precisely, they proved that if

$$H(x_1, \dots, x_m, y_1, \dots, y_n) = C(F(x_1, \dots, x_m), G(y_1, \dots, y_n))$$

defines a proper $(m+n)$ -dimensional distribution function (with $m+n \geq 3$) for every F and G with respective dimensions m and n , then $C(u, v) = u \cdot v$.

When the subsets $\{K_i, i = 1, \dots, n\}$ are not disjoint, an obvious necessary condition is that the prescribed measures $\{P_i, i = 1, \dots, n\}$ have the same marginals on common subspaces. But this condition is not sufficient : Kellerer [24] gave a necessary and sufficient condition only involving the structure of the sets $\{K_i, i = 1, \dots, n\}$. We refer to [21] (in particular, chapters 3 and 4, and section 3.4.3 to 3.7, for some compatibility conditions) and [8] and the references therein for further details and related problems, in particular extremal distributions with fixed multidimensional marginals and related optimization problems.

The question of uniqueness is also tricky. In [16], it is proved that if μ is a probability measure on \mathbb{R}^d with a density f with respect to the Lebesgue measure, then there exists a subset U of \mathbb{R}^d such that all the lower dimensional projections of the uniform distribution on U coincide with the lower dimensional projections of μ . This shows, at least in the case where measures have a density, that there is in general no hope for uniqueness of a measure with prescribed projections.

Possible explicit constructions of distribution functions F satisfying (1.1) have been given in [6], [33], [34], [7], [18], [19], [20], [27], [25], [26], [10] for overlapping or non-overlapping marginals.

All these constructions are very useful, in particular in statistics. Indeed, when the dimension d is large, one can first estimate all the bivariate marginals, since fitting a 2-dimensional copula is doable, and then construct a valid d -dimensional vector having the prescribed 2-dimensional margins. One problem of this approach is that it may not provide a unique d -dimensional distribution, but as pointed out in [21], one can then use entropy maximization techniques to choose a distribution among all those that have the prescribed marginals. By comparison, directly fitting the right copula in large dimension is however quite difficult and often makes use of recent research developments (nested, hierarchical or vine copulas, see [29], [4], [22], including pair-copula constructions or copulas with prescribed bivariate projections [1], [2], [9]).

The previously mentioned constructions all make use of the joint CDF, or the joint density. There are however other functions characterizing the joint distribution, for instance the characteristic function. We will call any such function a *characterizing function*. In this paper, we assume that a characterizing function m is given, and that there is a linear explicit decomposition of the characterizing function of the d -dimensional vector with respect to the characterizing functions of certain of its multi-dimensional marginals (Definition 2.1). We will say that a probability distribution satisfying Definition 2.1 is in a *projective class*. The main result of the paper is a complete analysis of the coefficients appearing in a decomposition of a projective class. Indeed, the distributions satisfying our Definition 2.1 are stable by projection, in the sense that they are such that all their multidimensional marginals also satisfy the same Definition 2.1. This allows us to give precise, and simple necessary conditions for a sequence of

coefficients to generate a probability distribution on \mathbb{R}^d having fixed multidimensional marginals and belonging to a projective class. In particular, a necessary condition is that a matrix containing these coefficients is idempotent (Proposition 2.3). Notice that the *linear* form of the decomposition in a projective class is not as restrictive as one would initially imagine, since one could first apply a bijective non linear transformation to a characterizing function, and then obtain a linear relation in the form of (2.2). Said otherwise, given a family of probability distributions, if some linearity can be found in the expression of one of its characterizing function, then our approach allows to exploit this linearity to construct multi-dimensional distributions. The case of Elliptical random vectors illustrates well this last point.

In Section 2 we define the projective class that we are going to work with, and in Section 3 we give and analyse examples of elements of this class.

2 Projective class of random vectors

Let $\mathcal{D} = \{1, \dots, d\}$, $\mathbf{t} \in \bar{\mathbb{R}}^d$, and denote by $\mathcal{P}(\mathcal{D})$ the power set of \mathcal{D} . Consider a random vector $\mathbf{X} = (X_i)_{i \in \mathcal{D}}$; for $K \in \mathcal{P}(\mathcal{D})$ a subset of \mathcal{D} , we denote by $\mathbf{X}_K = (X_i)_{i \in K}$ the subvector of \mathbf{X} , and by $\mathbf{t}_K = (t_i)_{i \in K}$ the subvector of \mathbf{t} .

We assume the existence of a link between the joint distribution of \mathbf{X} and the joint distributions of its projections \mathbf{X}_K . In general this link could be derived from the characteristic function of \mathbf{X} , from its CDF, or from some other quantity. We thus define a function m for which the link will be investigated.

We will denote by \mathcal{X}_d the space of \mathbb{R}^d -valued random variables that we will work with (with $k \leq d$). In the rest of the paper, the quantities involved (and our constructions) will only depend on the distributions of the random variables, and not on the random variables themselves. Because it has no impact on our results, we will nonetheless use random variables.

Assumption 2.1 (Characterizing function) *We assume that there exists a function $m : \bar{\mathbb{R}}^d \times \mathcal{X}_d \times \mathcal{P}(\mathcal{D}) \rightarrow \mathbb{R}$ such that for any non-empty $K \in \mathcal{P}(\mathcal{D})$:*

1. $\{m(\mathbf{t}, \mathbf{X}, K)\}_{\mathbf{t} \in \bar{\mathbb{R}}^d}$ characterizes the joint distribution of \mathbf{X}_K , i.e. $m(\mathbf{t}, \mathbf{X}, K) = m(\mathbf{t}, \mathbf{Y}, K)$ for all $\mathbf{t} \in \bar{\mathbb{R}}^d$ if and only if \mathbf{X}_K and \mathbf{Y}_K have the same distribution.
2. there exists $a \in \bar{\mathbb{R}}$ such that for all $\mathbf{t} \in \bar{\mathbb{R}}^d$, $m(P_K \mathbf{t}, \mathbf{X}, \mathcal{D}) = m(\mathbf{t}, \mathbf{X}, K)$, where

$$(P_K \mathbf{t})_i := \begin{cases} t_i & \text{if } i \in K \\ a & \text{if } i \notin K. \end{cases} \quad (2.1)$$

In the rest of the paper, to simplify notations, we will denote $m(\mathbf{t}, \mathbf{X}, K)$ by $m(\mathbf{t}, \mathbf{X}_K)$.

Remark 2.1 *Assumption 2.1 2. implies that $P_\emptyset \mathbf{t} = (a, \dots, a) =: \mathbf{a} \in \bar{\mathbb{R}}^d$, and that $m(\mathbf{t}, \mathbf{X}, \emptyset) = m(\mathbf{a}, \mathbf{X}, \emptyset)$ for all \mathbf{t} . For simplicity, we will write this quantity $m_0 := m(\mathbf{a}, \mathbf{X}, \emptyset)$, or abuse notation and write $m(\mathbf{t}, \mathbf{X}_\emptyset) = m_0$.*

Remark 2.2 *Such a function m always exists. Typically, m can be given by $m(\mathbf{t}, \mathbf{X}_K) = \mathbb{E} \left[e^{i \mathbf{t}_K^\top \mathbf{X}_K} \right]$, with $a = 0$ and $m_0 = 1$. Another example is $m(\mathbf{t}, \mathbf{X}_K) = F_K(\mathbf{t}_K)$, the CDF of \mathbf{X}_K , with $a = +\infty$ and $m_0 = 1$. A third example is $m(\mathbf{t}, \mathbf{X}_K) = \phi^{-1} \circ F_K(\mathbf{t}_K)$ where $\phi : \mathbb{R}_+ \rightarrow [0, 1]$ is an invertible Archimedean generator, with $a = +\infty$ and $m_0 = 0$. Direct transformations of these functions, as entropy or survival functions, are also suitable characterizing functions.*

Remark 2.3 *Assumption 2.1 2. says that to study the marginal distribution of X on the subset K , it is enough to study the distribution of X , with the characterizing function m restricted to $P_K \mathbf{t}$. However, notice that not every function characterizing the distribution of random variables*

satisfies this Assumption. Let us give the example of the potential function ([3]). Define the potential kernel v on \mathbb{R}^d as follows:

$$v(x) = -|x|, d = 1, \quad v(x) = -\log|x|, d = 2, \quad v(x) = |x|^{-d+2}, d \geq 3.$$

Then the potential U_X of a random vector X on \mathbb{R}^d is defined by

$$U_X(\mathbf{t}) = \mathbb{E}[v(X - \mathbf{t})]$$

when the expectation exists. We have that $U_X = U_Y$ if and only if X and Y have the same distribution, but the potential function does not satisfy Assumption 2.1 2.

We aim at defining the whole distribution of \mathbf{X} , using only some of its projections, i.e. using only the laws of X_K for $K \in \mathcal{S}$, where \mathcal{S} is a given subset of $\mathcal{P}(\mathcal{D})$. For example, \mathcal{S} can gather some subsets of cardinal 3, their subsets, and some singletons, or \mathcal{S} can gather only subsets of cardinal 1, as in copula theory. We assume that \mathcal{S} is *decreasing*, in the sense that for all $K \subset J$, $J \in \mathcal{S}$ implies $K \in \mathcal{S}$: knowing the distribution of a projection easily allows to know the distribution of every sub-vector. In the algebraic topology terminology, \mathcal{S} is a *simplicial complex* [36]. Simplicial complexes can be represented using points, line segments, triangles, and simplices in higher dimensions, which may ease the understanding of the projections and the model (see Figure 1 for an illustration).

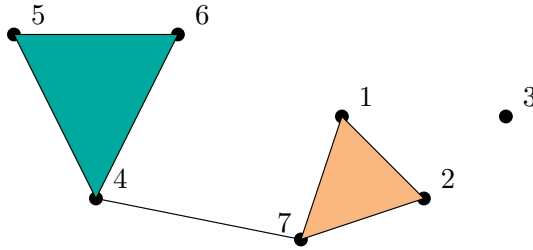


Figure 1: Illustration of a multivariate distribution in the dimension 7, knowing the marginal distribution $\{3\}$, the bivariate projection $\{4, 7\}$, and the trivariate projections $\{1, 2, 7\}$ and $\{4, 5, 6\}$. All subsets of these projections also correspond to known marginal distributions.

Definition 2.1 (Projective class) Let $\mathcal{S} \in \mathcal{P}(\mathcal{D})$ be decreasing. For a given characterizing function m , we say that a random vector $\mathbf{X} \in \mathbb{R}^d$ belongs to the projective class $\mathcal{F}_{\mathcal{D}}(\mathcal{S})$ if there exist some real coefficients $\alpha_{K, \mathcal{D}}$, $K \subset \mathcal{D}$, such that for all $\mathbf{t} \in \mathbb{R}^d$,

$$m(\mathbf{t}, \mathbf{X}) = \sum_{K \in \mathcal{S}} \alpha_{K, \mathcal{D}} m(\mathbf{t}, \mathbf{X}_K), \quad (2.2)$$

Remark that if a random vector \mathbf{X} belongs the projective class $\mathcal{F}_{\mathcal{D}}(\mathcal{S})$, then the set of suitable coefficients $\{\alpha_{K, \mathcal{D}}, K \subset \mathcal{D}\}$ satisfying Equation (2.2) is not necessarily unique. Notice also that if $\mathcal{D} \in \mathcal{S}$, any d -dimensional random vector is in $\mathcal{F}_{\mathcal{D}}(\mathcal{S})$, using for example constant coefficients $\alpha_{K, \mathcal{D}}$ equal to 1 if $K = \mathcal{D}$ and equal to 0 otherwise. This is quite natural, since the class $\mathcal{F}_{\mathcal{D}}(\mathcal{S})$ intends to define multivariate distributions that can be fully determined by some of their projections. This is obviously the case when the initial joint distribution is already in \mathcal{S} .

The projective class explicitly depends on the set \mathcal{S} of given projections. It also depends implicitly on the choice of the characterizing function m , but for the sake of simplicity, it will

not be indexed by m . A distribution (e.g. a centered Gaussian distribution) can be projective for a characterizing function m_1 (e.g. the logarithm of the characteristic function), but not for a characterizing function m_2 (e.g. the cumulative distribution function). At last, notice that *a priori*, the coefficients $\alpha_{K,D}$ may also depend on the choice of the characterizing function m , but for the sake of simplicity they will not be indexed by m . We will show however that under some simple conditions, one can find coefficients that do not depend on of the choice of m (see further Remark 2.5).

It is not trivial to assess the compatibility conditions between arbitrary characterizing functions of the projections: given arbitrary $m(\mathbf{t}, \mathbf{X}_K)$ in Equation (2.2), determining if the resulting $m(\mathbf{t}, \mathbf{X})$ corresponds indeed to a characterizing function is hard. As an example, if m is a characteristic function, the verification may rely on known criteria as Bochner's theorem or multivariate extensions of Pólya's theorem. If m is a transformation of a cumulative distribution function, the verification may rely on differentiation of this function using chain rule differentiation and multivariate extensions of Faa Di Bruno's formula. For known projective families (see Section 3), the compatibility is ensured, but in practice (see Section 4) it may rely on some numerical verifications. As in Section 1, we refer to [24, 21, 8] for supplementary material on the question of projections compatibility.

Assuming that a distribution is projective, we give here a sufficient condition ensuring the unicity of the coefficients $\{\alpha_{K,D}\}_{K \in \mathcal{S}}$, which further implies that the distribution characterized by Equation 2.2 is unique. The condition relies on the given projections $m(\mathbf{t}, \mathbf{X}_K)$, $K \in \mathcal{S}$.

Proposition 2.1 (Condition on the projections for unicity) *Define $S := \mathcal{S}$ if $m_0 \neq 0$, and $S := \mathcal{S} \setminus \{\emptyset\}$ if $m_0 = 0$. Consider a finite set of points $\mathcal{T} \subset \bar{\mathbb{R}}^d$, and denote the matrix $M_S(\mathcal{T}) := (m(\mathbf{t}, \mathbf{X}_K))_{\mathbf{t} \in \mathcal{T}, K \in \mathcal{S}}$. Assume that the distribution of \mathbf{X} is projective, and denote the vector of coefficients $\boldsymbol{\alpha} := (\alpha_{K,D})_{K \in \mathcal{S}}$, then the following results hold:*

- (i) $\boldsymbol{\alpha}$ is unique if there exists a set \mathcal{T} with $|\mathcal{T}| = |\mathcal{S}|$ such that the matrix $M_S(\mathcal{T})$ is invertible.
- (ii) In particular when $\mathcal{S} = \{K \subset \mathcal{D}, |K| \leq 2\}$ and $m_0 = 0$, $\boldsymbol{\alpha}$ is unique if there exists $\mathbf{t} \in \bar{\mathbb{R}}^d$ such that for all $K \in \mathcal{S}$, $m_K^*(\mathbf{t}) = m(\mathbf{t}, \mathbf{X}_K) - \sum_{J \subset K, J \neq K} m(\mathbf{t}, \mathbf{X}_J) > 0$.

Proof. It is clear that if $m_0 = 0$, then $\alpha_{\emptyset,D} m(\mathbf{t}, \mathbf{X}_\emptyset) = 0$ whatever the value of $\alpha_{\emptyset,D}$. Multiplied by zero, this coefficient has no impact, we thus exclude it from the analysis by setting $S = \mathcal{S} \setminus \{\emptyset\}$ in the case when $m_0 = 0$. Let us show the item (i). Write the vector $\mathbf{m}_X(\mathcal{T}) = (m(\mathbf{t}, X))_{\mathbf{t} \in \mathcal{T}}$. The main linear equality of Equation (2.2) writes:

$$\mathbf{m}_X(\mathcal{T}) = M_S(\mathcal{T}) \cdot \boldsymbol{\alpha} \quad (2.3)$$

so that when $M_S(\mathcal{T})$ is invertible, $\boldsymbol{\alpha} = [M_S(\mathcal{T})]^{-1} \mathbf{m}_X(\mathcal{T})$ is uniquely determined.

For the second part of the proposition, item (ii), write $S = \{K_1, \dots, K_s\}$, set $\mathbf{t} \in \bar{\mathbb{R}}^d$ and set $\mathcal{T} = \{\mathbf{t}_K, K \in \mathcal{S}\}$. Write P the matrix with components $P_{ij} = 1$ if $K_i \subset K_j$ or 0 otherwise. One can check that the component i, j of the matrix $M_S(\mathcal{T})$ writes:

$$(M_S(\mathcal{T}))_{ij} = m(\mathbf{t}_{K_i}, \mathbf{X}_{K_j}) = m(\mathbf{t}, \mathbf{X}_{K_i \cap K_j}) \quad (2.4)$$

For any set L , $K = L$ if and only if K is a subset of L but not a strict subset of L , thus

$$(M_S(\mathcal{T}))_{ij} = \sum_{K \in \mathcal{S}, K \subset K_i \cap K_j} m(\mathbf{t}, \mathbf{X}_K) - \sum_{K \in \mathcal{S}, K \subsetneq K_i \cap K_j} m(\mathbf{t}, \mathbf{X}_K). \quad (2.5)$$

By assumption, for all $K \in \mathcal{S}$, $|K| \leq 2$. The following equality holds for $|L| = 1$, since both the left and right hand terms are equal to 0 in this case. The equality also holds for $|L| = 2$,

because in the right hand side, $K \in \mathcal{S}, J \subset K, J \neq K$ implies $|J| = 1$ and $|K| = 2$, and $K \subset L$ with $|K| = |L| = 2$ implies $K = L$:

$$\sum_{K \in \mathcal{S}, K \subsetneq L} m(\mathbf{t}, \mathbf{X}_K) = \sum_{K \in \mathcal{S}, K \subset L} \sum_{J \subset K, J \neq K} m(\mathbf{t}, \mathbf{X}_J). \quad (2.6)$$

Finally, for $h \in \{1, \dots, s\}$, as $K_h \subset K_i \cap K_j$ if and only if $P_{hi}P_{hj} = 1$,

$$(M_S(\mathcal{T}))_{ij} = \sum_{K_h \in \mathcal{S}, K_h \subset K_i \cap K_j} m_{K_h}^*(\mathbf{t}) = \sum_{K_h \in \mathcal{S}} P_{hi} \cdot m_{K_h}^*(\mathbf{t}) \cdot P_{hj} \quad (2.7)$$

so that $M_S(\mathcal{T}) = P^\top DP$, where D is the diagonal matrix with diagonal $\{m_K^*(\mathbf{t}), K \in \mathcal{S}\}$. When $\mathcal{S} = \{K \subset \mathcal{D}, |K| \leq 2\}$, up to a rearrangement, P is an upper triangular matrix with ones on its diagonal. Thus $\det(P) = 1$ and finally $\det(M_S(\mathcal{T})) = \det(D)$. Under the assumption of (ii), $\det(D) > 0$ and the result holds. \square

In the next Proposition, we prove the following projection stability property: if a random vector belongs to the class $\mathcal{F}_{\mathcal{D}}(\mathcal{S})$, then any subvector also belong to $\mathcal{F}_{\mathcal{D}}(\mathcal{S})$, and we compute the corresponding coefficients.

Proposition 2.2 (Projection stability) *Let \mathbf{X} be a d -dimensional random vector in $\mathcal{F}_{\mathcal{D}}(\mathcal{S})$, with associated coefficients $\{\alpha_{K,\mathcal{D}}, K \in \mathcal{S}\}$. Then for any non-empty $L \subset \mathcal{D}$, the subvector \mathbf{X}_L belongs to $\mathcal{F}_L(\mathcal{S})$, where for any non-empty subset J of L , a suitable set of associated coefficients is given by*

$$\alpha_{J,L} = \sum_{K \in \mathcal{S}, K \cap L = J} \alpha_{K,\mathcal{D}}, \quad (2.8)$$

which implies in particular that $\alpha_{J,L} = 0$ if $J \not\subset L$.

Let d_0 be an integer such that $1 \leq d_0 \leq d$. When $\mathcal{S} = \{K \subset \mathcal{D}, |K| \leq d_0\}$, and when the coefficients $\{\alpha_{J,L}, J \subset L\}$ only depend on the subsets cardinals, i.e. $\alpha_{J,L} = \alpha_{j,\ell}$ with $j = |J| \in \{1, \dots, d\}$ and $\ell = |L| \in \{0, \dots, d\}$, then a suitable set of associated coefficients is given by

$$\alpha_{j,\ell} = \sum_{k=j}^{\min(d_0, d-\ell+j)} \binom{d-\ell}{k-j} \alpha_{k,d} \quad \text{if } j \leq \min(\ell, d_0), \quad (2.9)$$

and $\alpha_{j,\ell} = 0$ otherwise.

Proof. First remark that, due to the second assumption on $m(\mathbf{t}, \mathbf{X})$, we have $m(\mathbf{t}, \mathbf{X}_{K \cap L}) = m(P_{K \cap L} \mathbf{t}, \mathbf{X}) = m(P_K P_L \mathbf{t}, \mathbf{X}) = m(P_L \mathbf{t}, \mathbf{X}_K)$. Thus, using (2.2), we have

$$\begin{aligned} m(\mathbf{t}, \mathbf{X}_L) &= m(P_L \mathbf{t}, \mathbf{X}) = \sum_{K \in \mathcal{S}} \alpha_{K,\mathcal{D}} m(P_L \mathbf{t}, \mathbf{X}_K) = \sum_{K \in \mathcal{S}} \alpha_{K,\mathcal{D}} m(\mathbf{t}, \mathbf{X}_{K \cap L}) \\ &= \sum_{J \subset L} \sum_{K \in \mathcal{S}, K \cap L = J} \alpha_{K,\mathcal{D}} m(\mathbf{t}, \mathbf{X}_J). \end{aligned}$$

By assumption, $K \in \mathcal{S}$ implies that $J = K \cap L \in \mathcal{S}$. Setting $\alpha_{J,L} = \sum_{K \in \mathcal{S}, K \cap L = J} \alpha_{K,\mathcal{D}}$ as in Equation (2.8) we finally get

$$m(\mathbf{t}, \mathbf{X}_L) = \sum_{J \subset L, J \in \mathcal{S}} \alpha_{J,L} m(\mathbf{t}, \mathbf{X}_J), \quad (2.10)$$

hence \mathbf{X}_L belongs to $\mathcal{F}_L(\mathcal{S})$.

Eventually remark that $K \cap L = J$ if and only if $J \subset L$ and $J \subset K$ and $K \cap L \subset J$. As a consequence, if $J \not\subset L$, the sum in (2.8) is empty and $\alpha_{J,L} = 0$.

We prove now the second part of the Proposition. When $\mathcal{S} = \{K \subset \mathcal{D}, |K| \leq d_0\}$, we have using (2.8)

$$\alpha_{J,L} = \sum_{K' \subset \mathcal{D} \setminus L, |K'| \leq d_0 - |J|} \alpha_{K' \cup J, \mathcal{D}}.$$

Now, let $j = |J|$ and $\ell = |L|$. If $j > \ell$, it is clear that $J \not\subset L$ and that $\alpha_{J,L} = 0$. As $K' \subset \mathcal{D} \setminus L$ and $|K'| \leq d_0 - |J|$, we get $0 \leq |K'| \leq \min(d_0 - j, d - \ell)$. Thus when $J \subset L$,

$$\alpha_{J,L} = \sum_{k=0}^{\min(d_0-j, d-\ell)} \sum_{K' \subset \mathcal{D} \setminus L, |K'|=k} \alpha_{K' \cup J, \mathcal{D}} \quad (2.11)$$

and if the coefficients $\alpha_{J,L}$ do only depend on the cardinals,

$$\alpha_{j,\ell} = \sum_{k=0}^{\min(d_0-j, d-\ell)} \sum_{K' \subset \mathcal{D} \setminus L, |K'|=k} \alpha_{k+j, d} = \sum_{k=0}^{\min(d_0-j, d-\ell)} \binom{d-\ell}{k} \alpha_{k+j, d}. \quad (2.12)$$

Hence the second result. \square

Remark 2.4 (case of \emptyset) Notice that \emptyset necessarily belongs to \mathcal{S} . In the case where $m(\mathbf{t}, \mathbf{X}_\emptyset) \neq 0$, Equation (2.2) may involve a constant $\alpha_{\emptyset, \mathcal{D}}$, and implies that $\sum_{K \subset \mathcal{D}, K \in \mathcal{S}} \alpha_{K, \mathcal{D}} = 1$. In this case, it becomes useful to determine the coefficients $\alpha_{\emptyset, L}$. For any non-empty L , the Equation (2.8) writes

$$\alpha_{\emptyset, L} = \sum_{K \subset \mathcal{D} \setminus L, K \in \mathcal{S}} \alpha_{K, \mathcal{D}}. \quad (2.13)$$

When $L = \emptyset$, the Equation (2.2) remains valid for $D = \emptyset$ if one defines $\alpha_{\emptyset, \emptyset} = 1$.

As previously noticed, the coefficients $\alpha_{K, \mathcal{D}}$ and $\alpha_{\emptyset, L}$ may depend on the choice of the characterizing function m . However, one can check that using $\tilde{m}(t, \mathbf{X}) = cm(t, \mathbf{X})$ for a positive constant c leads to $\tilde{m}(t, \mathbf{X}_\emptyset) = cm(t, \mathbf{X}_\emptyset)$ and to unchanged coefficients $\tilde{\alpha}_{K, \mathcal{D}} = \alpha_{K, \mathcal{D}}$.

Corollary 2.1 (Given projections up to dimension d_0) Let \mathbf{X} be a d -dimensional random vector in $\mathcal{F}_{\mathcal{D}}(\mathcal{S})$. Assume that all projections of \mathbf{X} are given up to a dimension d_0 , so that $\mathcal{S} = \{K \subset \mathcal{D}, |K| \leq d_0\}$. Assume that the associated coefficients $\{\alpha_{J,L}, J \subset L\}$ only depend on the subsets cardinals, i.e. $\alpha_{J,L} = \alpha_{j,\ell}$ with $j = |J| \in \{1, \dots, d\}$ and $\ell = |L| \in \{0, \dots, d\}$. Assume furthermore that $\alpha_{k,k} = 1$ for all $k \leq d_0$. Then the coefficients $\alpha_{d_0-z, d}$ can be obtained recursively, using

$$\alpha_{d_0-z, d} = 1 - \sum_{i=1}^z \binom{d-d_0+z}{i} \alpha_{i+d_0-z, d}, \quad (2.14)$$

for $z = 1, \dots, d_0$, starting with $\alpha_{d_0, d} = 1$. In particular for $2 \leq d_0 \leq d$, we get $\alpha_{d_0, d} = 1$, $\alpha_{d_0-1, d} = -(d-d_0)$, $\alpha_{d_0-2, d} = 1 + \frac{d-d_0+2}{2}(d-d_0-1)$.

When $d_0 \geq 3$, $\alpha_{d_0-3, d} = 1 - (d-d_0+3)\{1 - \frac{d-d_0+2}{2}(1 - \frac{d-d_0+1}{3})\}$. For higher orders, one can check by induction that these coefficients do only depend on $d-d_0$, but their expression is omitted here.

Proof. Under the assumption $\alpha_{k,k} = 1$ for all $k \leq d_0$, this follows directly from Proposition 2.2, by writing Equation (2.9) in the case where $j = \ell \leq d_0$, and setting $i = k - j$, $j = d_0 - z$: we get $\alpha_{d_0, d} = 1$ when $z = 0$, and Equation (2.14) when $z \geq 1$. \square

Remark 2.5 (Explicit coefficients) The fact that $\alpha_{k,k} = 1$ for all $k \leq d_0$ means that if we are given a k -dimensional marginal, we do not try to retrieve it from the given lower dimensional marginals. Under this assumption, and under the assumption that the coefficients in the projective decomposition only depend on the subsets cardinals, the previous Corollary provides a set of suitable coefficients $\{\alpha_{j,\ell}\}$ which are explicitly given, independently of the choice of m .

The case where all bivariate projections are given is a very natural and interesting case: in practical applications, bivariate projections can be graphically visualized, and the estimation of the dependence structure among each pair of random variable is still tractable. The following Remark shows that in this case, under some simple conditions, the coefficients $\alpha_{J,L}$ can be computed explicitly.

Remark 2.6 (Given bivariate projections) *Consider the same assumptions as in Corollary 2.1 and assume that all bivariate projections of a multivariate distribution are given, so that $d_0 = 2$ and $\mathcal{S} = \{J \subset \mathcal{D}, |J| \leq 2\}$. Then for all non-empty $L \subset \mathcal{D}$, we can reformulate equation (2.2) as follows:*

$$m(\mathbf{t}, \mathbf{X}_L) = \alpha_{0,|L|}m_0 - (|L| - 2) \sum_{J \subset L, |J|=1} m(\mathbf{t}, \mathbf{X}_J) + \sum_{J \subset L, |J|=2} m(\mathbf{t}, \mathbf{X}_J), \quad (2.15)$$

where $\alpha_{0,0} = 1$, $\alpha_{0,\ell} = 1 + \frac{1}{2}\ell(\ell - 3)$, $\ell \geq 1$ and where $m_0 = m(\mathbf{t}, \mathbf{X}_\emptyset)$ is defined in Remark 2.1.

Let $\mathbf{X} \in \mathbb{R}^d$ be a random vector in $\mathcal{F}_{\mathcal{D}}(\mathcal{S})$. Since $\mathcal{D} = \{1, \dots, d\}$ is a finite set, the set of subsets of \mathcal{D} is also finite, and we can define the following matrix, indexed by the subsets of \mathcal{D} :

$$A = (\alpha_{J,L})_{J \subset \mathcal{D}, L \subset \mathcal{D}}. \quad (2.16)$$

We will write $A_{\cdot,L}$ for the column vector relative to the subset L .

Proposition 2.3 *The coefficients in the matrix A satisfy the following constraints:*

1. *If the set of associated coefficients $\{\alpha_{J,L}\}$ is unique, then the matrix A defined in (2.16) is idempotent: $A^2 = A$.*
2. *If furthermore the coefficients do only depend on the subset's cardinal, i.e. $\alpha_{J,L} = \alpha_{j,\ell}$ with $j = |J|$ and $\ell = |L|$, we get for $0 \leq j \leq \ell \leq d$:*

$$\alpha_{j,\ell} = \sum_{k=j}^{\min(d_0,\ell)} \binom{\ell-j}{k-j} \alpha_{j,k} \alpha_{k,\ell}. \quad (2.17)$$

Now define $s_{j,\ell} := \binom{\ell-2}{j-2} \alpha_{j,\ell}$ when $j \geq 2$. The previous equation writes, when $j \geq 2$,

$$s_{j,\ell} = \sum_{k=j}^{\min(d_0,\ell)} s_{j,k} s_{k,\ell}. \quad (2.18)$$

Proof. Let $L \subset \mathcal{D}$, $|L| \geq 1$. From Proposition 2.2,

$$m(\mathbf{t}, \mathbf{X}_L) = \sum_{K \subset L, K \in \mathcal{S}} \alpha_{K,L} m(\mathbf{t}, \mathbf{X}_K)$$

From Proposition 2.2, we also have $m(\mathbf{t}, \mathbf{X}_K) = \sum_{J \subset K, J \in \mathcal{S}} \alpha_{J,K} m(\mathbf{t}, \mathbf{X}_J)$, so that finally

$$\begin{aligned} m(\mathbf{t}, \mathbf{X}_L) &= \sum_{K \subset L, K \in \mathcal{S}} \sum_{J \subset K, J \in \mathcal{S}} \alpha_{J,K} \alpha_{K,L} m(\mathbf{t}, \mathbf{X}_J) \\ &= \sum_{J \subset L, J \in \mathcal{S}} \sum_{K \subset L, K \in \mathcal{S}} \alpha_{J,K} \alpha_{K,L} m(\mathbf{t}, \mathbf{X}_J), \end{aligned}$$

as $\alpha_{J,K} = 0$ if $J \not\subset K$. Then for all \mathbf{t} ,

$$m(\mathbf{t}, \mathbf{X}_L) = \sum_{J \subset L, J \in \mathcal{S}} \left\{ \sum_{K \subset L, K \in \mathcal{S}} \alpha_{J,K} \alpha_{K,L} \right\} m(\mathbf{t}, \mathbf{X}_J)$$

so that, using the unicity of the set of coefficients $\{\alpha_{J,L}\}$,

$$\alpha_{J,L} = \sum_{K \subset L, K \in \mathcal{S}} \alpha_{J,K} \alpha_{K,L}$$

and thus A is idempotent.

Let us now focus on the second part of the proposition. For a subset $L \subset \mathcal{D}$ with cardinal ℓ and $k \leq \ell$, define $[L]^k := \{K \subset L \text{ such that } |K| = k\}$. Assume that when $K \subset L$ the coefficients $\alpha_{K,L}$ do only depend on the cardinals $k = |K|$ and $\ell = |L|$ of the considered sets, and assume $\alpha_{K,L} = 0$ if $K \notin \mathcal{S}$, i.e. $\alpha_{k,\ell} = 0$ if $k > d_0$. Then

$$\begin{aligned} m(\mathbf{t}, \mathbf{X}_L) &= \sum_{K \subset L} \alpha_{K,L} m(\mathbf{t}, \mathbf{X}_K) \\ &= \sum_{k=0}^{\min(d_0, \ell)} \alpha_{k,\ell} \sum_{K \in [L]^k} m(\mathbf{t}, \mathbf{X}_K) \\ &= \sum_{k=0}^{\min(d_0, \ell)} \alpha_{k,\ell} \sum_{K \in [L]^k} \left\{ \sum_{j=0}^{\min(d_0, k)} \alpha_{j,k} \sum_{J \in [K]^j} m(\mathbf{t}, \mathbf{X}_J) \right\} \end{aligned}$$

Notice that, by a simple combinatorial argument, for $j \leq k$,

$$\sum_{K \in [L]^k} \sum_{J \in [K]^j} m(\mathbf{t}, \mathbf{X}_J) = \binom{\ell - j}{k - j} \sum_{J \in [L]^j} m(\mathbf{t}, \mathbf{X}_J),$$

which entails that

$$m(\mathbf{t}, \mathbf{X}_L) = \sum_{k=0}^{\min(d_0, \ell)} \sum_{j=0}^{\min(d_0, k)} \alpha_{k,\ell} \alpha_{j,k} \binom{\ell - j}{k - j} \sum_{J \in [L]^j} m(\mathbf{t}, \mathbf{X}_J).$$

On the other hand we have

$$m(\mathbf{t}, \mathbf{X}_L) = \sum_{j=0}^{\min(d_0, \ell)} \alpha_{j,\ell} \sum_{J \in [L]^j} m(\mathbf{t}, \mathbf{X}_J)$$

so that for all \mathbf{t} , for all $j \leq k$,

$$\alpha_{j,\ell} = \sum_{k=j}^{\min(d_0, \ell)} \binom{\ell - j}{k - j} \alpha_{j,k} \alpha_{k,\ell}$$

and the second result holds. \square

Remark 2.7 Notice that due to the projection stability property of Proposition 2.2, any column of the matrix A can be deduced from the last one by multiplication by a matrix with values in $\{0, 1\}$: $A_{\cdot, L} = P^{(L)} A_{\cdot, \mathcal{D}}$, where the $2^d \times 2^d$ matrix $P^{(L)}$ is defined by its components

$$P_{J,K}^{(L)} = \begin{cases} 1 & \text{if } J = K \cap L \text{ and } K \in \mathcal{S} \\ 0 & \text{otherwise.} \end{cases},$$

for J, K, L subsets of \mathcal{D} . Indeed, we have from Proposition 2.2:

$$\alpha_{J,L} = \sum_{K \in \mathcal{S}, K \cap L = J} \alpha_{K,\mathcal{D}} = \sum_{K \in \mathcal{S}} P_{J,K}^{(L)} \alpha_{K,\mathcal{D}}. \quad (2.19)$$

3 Examples

3.1 Elliptical random vectors

Recall [21] that d -dimensional elliptical distributions are characterized by the fact that their characteristic function can be written in the following form: for any $\mathbf{t} \in \mathbb{R}^d$,

$$\mathbb{E} \left[e^{i\mathbf{t}^\top (\mathbf{X} - \boldsymbol{\mu})} \right] = \phi(\mathbf{t}^\top \Sigma \mathbf{t}) \quad (3.1)$$

for a given function $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$ which is called the generator of the distribution, and where $\boldsymbol{\mu}$ is the mean of the vector \mathbf{X} and Σ is a non-negative definite matrix. We assume here that the generator ϕ does not depend on the dimension d of the random vector, i.e. that the elliptical distribution is consistent in the sense of [23].

One interesting feature of families of elliptical distributions is that they allow heavy tails, while preserving some advantageous properties of multivariate Gaussian distributions. Indeed, an elliptical distribution has a stochastic representation as a product of two independent random elements: a univariate radial distribution and a uniform distribution on an ellipsoid (see Chapter 2 in [12]). This representation makes possible the analysis of the density functions, the moments, the conditional distributions, the symmetries, and the infinite-divisibility properties of elliptical distributions ([13]). Other than multivariate Gaussians, elliptical distributions include the Student t -distributions, the symmetric generalized hyperbolic distribution, the power exponential distributions and the sub-Gaussian α -stable distributions among others. Families of elliptical distributions are widely applied in statistics, in particular in the area of robust statistics, as a starting point for the development of the M -estimates of multivariate location and scale (see Chapter 2 in [28]). It is also a rather standard family of distributions in financial modelling; see [14] and the references therein. Thus, constructing a distribution with given multivariate elliptical marginals is a useful and interesting problem, this is what we do in this subsection.

Let us first remark that, for a given generator ϕ , when one considers a centered multivariate elliptical distribution, the distribution is fully characterized by all components σ_{ij} of the matrix Σ , that is to say by all bivariate elliptical projections of the distribution (it does not mean that the multivariate elliptical distribution is the only one having those projections).

It is thus quite natural to analyse, in the case of elliptical distributions, the links between the matrix Σ and a given set of submatrices $\Sigma_{K_1}, \dots, \Sigma_{K_n}$, for K_1, \dots, K_n subsets of $\mathcal{D} = \{1, \dots, d\}$. This is easier to do using the matrix Σ rather than its inverse Σ^{-1} . It thus seems easier to work with characteristic functions or entropy (which are expressed using Σ) rather than densities of cumulative distribution functions (which are expressed using Σ^{-1}).

We will try to express the quantity $\mathbf{t}^\top \Sigma \mathbf{t}$ as a linear combination of products $\mathbf{t}_K^\top \Sigma_K \mathbf{t}_K$, where K belongs to known projections indexes in \mathcal{S} .

Definition 3.1 (\mathcal{S} -admissible sequence) Let $\mathcal{D} \subset \mathbb{N}$, and \mathcal{S} be a decreasing subset of \mathcal{D} . A sequence of coefficients $\alpha_{K,\mathcal{D}}$, $K \in \mathcal{S}$, is said to be \mathcal{S} -admissible if for all matrix Σ , for all \mathbf{t} ,

$$\mathbf{t}^\top \Sigma \mathbf{t} = \sum_{K \subset \mathcal{D}, K \in \mathcal{S}} \alpha_{K,\mathcal{D}} \cdot \mathbf{t}_K^\top \Sigma_K \mathbf{t}_K, \quad (3.2)$$

The following Lemma provides a characterization of such coefficients:

Lemma 3.1 (Characterization of \mathcal{S} -admissible sequence) *Let $d_0 \in \mathbb{N}$ be such that $2 \leq d_0 \leq d$, and assume that $\mathcal{S} = \{K \subset \mathcal{D}, |K| \leq d_0\}$. Assume furthermore that for any sets K, \mathcal{D} , $\alpha_{K, \mathcal{D}} = \alpha_{|K|, |\mathcal{D}|}$. A sequence $\alpha_d = (\alpha_{k,d})_{k \leq d}$ of coefficients is \mathcal{S} -admissible if and only if it can be written*

$$\alpha_{1,d} = -\sum_{k=2}^{d_0} s_k \frac{d-k}{k-1} \quad \text{and} \quad \alpha_{k,d} = s_k / \binom{d-2}{k-2}, \quad k = 2, \dots, d_0 \quad (3.3)$$

for some real values s_2, \dots, s_{d_0} such that $s_2 + \dots + s_{d_0} = 1$.

In the particular case where the coefficients are deduced from only two given dimensions, i.e. if there exists $k_0 \geq 2$ such that $s_i = 0$ whenever $i \notin \{1, k_0\}$, we get a particular \mathcal{S} -admissible sequence

$$\alpha_{1,d} = -\frac{d-k_0}{k_0-1} \quad \text{and} \quad (3.4)$$

$$\alpha_{k,d} = \frac{1}{\binom{d-2}{k_0-2}}, \text{ if } k = k_0 \text{ and } \alpha_{k,d} = 0 \text{ otherwise, } k = 2, \dots, d_0. \quad (3.5)$$

Furthermore, when $d_0 = 2$, the only \mathcal{S} -admissible sequence is

$$\alpha_{1,d} = -(d-2) \quad \text{and} \quad \alpha_{2,d} = 1 \quad (3.6)$$

Proof. Assume that α_d is \mathcal{S} -admissible, and that $\alpha_{K, \mathcal{D}}$ does only depend on $|K|$ and $|\mathcal{D}|$. Let $(i_0, j_0) \in \mathcal{D}^2$, $i_0 \neq j_0$. Isolating the coefficient $\mathbf{t}_{i_0} \sum_{i_0, j_0} \mathbf{t}_{j_0}$ on both sides of Equation (3.2), we get

$$1 = \sum_{k=2}^{d_0} |\{K : K \subset \mathcal{D}, |K| = k, \{i_0, j_0\} \subset K\}| \alpha_{k,d} = \sum_{k=2}^{d_0} \binom{d-2}{k-2} \alpha_{k,d}$$

Denoting $s_k = \alpha_{k,d} \binom{d-2}{k-2}$ for all k , we get $\sum_{k=2}^{d_0} s_k = 1$. Now considering the coefficient $\mathbf{t}_{i_0} \sum_{i_0, i_0} \mathbf{t}_{i_0}$ on both sides of Equation (3.2), we get

$$1 = \sum_{k=1}^{d_0} \alpha_{k,d} |\{K : K \subset \mathcal{D}, |K| = k, i_0 \in K\}| = \alpha_{1,d} + \sum_{k=2}^{d_0} \binom{d-1}{k-1} \alpha_{k,d}$$

Now as $s_k = \alpha_{k,d} \binom{d-2}{k-2}$ for all k ,

$$1 = \alpha_{1,d} + \sum_{k=2}^{d_0} s_k \frac{\binom{d-1}{k-1}}{\binom{d-2}{k-2}} = \alpha_{1,d} + \sum_{k=2}^{d_0} s_k \frac{d-1}{k-1}$$

Finally, using $1 = \sum_{k=2}^{d_0} s_k$, Equation (3.3) holds. The rest of the proposition are direct applications of this last equation. \square

A direct application of such \mathcal{S} -admissible sequence is the following one:

Proposition 3.1 (Elliptical distributions are projective) *Consider a d -dimensional random vector \mathbf{X} having elliptical distribution with mean $\boldsymbol{\mu}$, matrix Σ and invertible generator ϕ . Let $\mathcal{D} = \{1, \dots, d\}$ and $[\mathcal{D}]^k = \{K \subset \mathcal{D}, |K| = k\}$. Consider that all projections are given up to a dimension d_0 , $2 \leq d_0 \leq d$, so that $\mathcal{S} = \{K \subset \mathcal{D}, |K| \leq d_0\}$. Then for any \mathcal{S} -admissible sequence $\alpha_d = (\alpha_{1,d}, \dots, \alpha_{d_0,d})$, the following equality holds*

$$\phi(\mathbf{t}^\top \Sigma \mathbf{t}) = \phi \left(\sum_{k=1}^{d_0} \alpha_{k,d} \sum_{K \in [\mathcal{D}]^k} \mathbf{t}_K^\top \Sigma_{KK} \mathbf{t}_K \right). \quad (3.7)$$

In other words, setting $m(\mathbf{t}, \mathbf{X}) = \phi^{-1} \left(\mathbb{E} \left[e^{i\mathbf{t}^\top (\mathbf{X} - \boldsymbol{\mu})} \right] \right)$, and $\mathcal{S} = \{K \subset \mathcal{D}, |K| \leq 2\}$, we have

$$\mathbf{X} \in \mathcal{F}_{\mathcal{D}}(\mathcal{S}) \quad (3.8)$$

In particular, when $d_0 = 2$ (i.e. starting from all bivariate projections), the admissible sequence is $\alpha_{1,d} = -(d-2)$ and $\alpha_{2,d} = 1$.

Proof. By definition, for any \mathcal{S} -admissible sequence, $\mathbf{t}^\top \Sigma \mathbf{t} = \sum_{k=1}^{d_0} \alpha_{k,d} \sum_{K \in [\mathcal{D}]^k} \mathbf{t}_K^\top \Sigma_{KK} \mathbf{t}_K$. One can also check that the functions m are suitable characterizing functions satisfying Assumption 2.1 with $a = 0$ and $m_0 = \phi^{-1}(1) = 0$. Hence the result. \square

Remark 3.1 (Matrix A in the elliptical case, $d_0 = 2$) Consider the elliptical case with $d_0 = 2$. Then we get by Lemma 3.1,

$$\alpha_{J,L} = \begin{cases} 0 & \text{if } J \not\subset L, \\ -(|L| - 2) & \text{if } |J| = 1 \text{ and } J \subset L, \\ 1 & \text{if } |J| = 2 \text{ and } J \subset L. \end{cases}$$

In particular, if $d = 3$ and $\mathcal{D} = \{1, 2, 3\}$, the matrix $A = (\alpha_{J,L})_{J \subset \mathcal{D}, L \subset \mathcal{D}}$ is

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad (3.9)$$

where the seven rows and columns correspond to successive subsets of \mathcal{D} :

$$\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}$$

As $m_0 = 0$ in the elliptical case, it is not necessary to compute the coefficients $\alpha_{J,L}$ for $J = \emptyset$ or $L = \emptyset$ (see Remark 2.4). One easily check that we can apply Proposition 2.3 to deduce that A is idempotent, which can also be verified by hand in this example.

3.2 Vectors built from bivariate distributions

Assume that a family $(\mu_{i,j})_{(i,j) \in D^2}$ of probability measures on \mathbb{R}^2 is given, for some non-negative parameters θ_i, θ_j and $\theta_{i,j}$ specific to each couple $(i, j) \in D^2$, and that each $\mu_{i,j}$ satisfies:

$$\mu_{i,j}((t, +\infty), (u, +\infty)) := \psi(\theta_i t + \theta_j u + \theta_{i,j} t u), \quad (t, u) \in \mathbb{R}_+^2, \quad (3.10)$$

where ψ is a given appropriate function.

Thanks to our previous results, we can construct a random variable \mathbf{X} , with values in \mathbb{R}_+^d , such that the survival function of each subvector (X_i, X_j) is given by the right hand side of equation (3.10).

Let us first analyze the copula function associated to (X_i, X_j) . To do so, assume that ψ is a decreasing bijection from \mathbb{R}_+ to $(0, 1]$, such that $\psi(0) = 1$ and that derivatives of ψ exist up to the order d . Denote by ψ^{-1} the inverse function of ψ . Let $t = (t_i, t_j) \in \mathbb{R}_+^2$ and $S_{ij}(t) = \psi(\theta_i t_i + \theta_j t_j + \theta_{i,j} t_i t_j)$. The one-dimensional survival functions are then given by $S_i(t) = \psi(\theta_i t_i)$, from which we get $\theta_i t_i = \psi^{-1}(S_i(t))$, so that finally

$$S_{ij}(t) = \psi \left(\psi^{-1}(S_i(t)) + \psi^{-1}(S_j(t)) + \frac{\theta_{ij}}{\theta_i \theta_j} \psi^{-1}(S_i(t)) \psi^{-1}(S_j(t)) \right).$$

Therefore we can write

$$S_{ij}(t) = C_{S_{ij}}(S_i(t), S_j(t)),$$

where the survival copula $C_{S_{ij}}$ is given by

$$C_{S_{ij}}(u, v) = \psi \left(\psi^{-1}(u) + \psi^{-1}(v) + \frac{\theta_{ij}}{\theta_i \theta_j} \psi^{-1}(u) \psi^{-1}(v) \right), \quad (u, v) \in [0, 1]^2. \quad (3.11)$$

Example 3.1 *Examples of survival functions in the form of Equation(3.10), or Equation (3.13) in the d -dimensional case, include the following particular cases:*

1. If $\psi(x) = \exp(-x)$, then (3.11) reduces to

$$C_{S_{ij}}(u_i, u_j) = u_i u_j \exp \left(-\frac{\theta_{ij}}{\theta_i \theta_j} \ln u_i \ln u_j \right),$$

which is the survival copula associated with Gumbel's bivariate exponential distribution (see [30], copula (4.2.9)).

2. In the case where $\theta_{i,j} = 0$ for every $(i, j) \in D^2$ and if the generator ψ is d -monotone, we obtain a survival copula which is Archimedean with generator ψ . It is clear that in this specific case, the multivariate distribution in higher dimension will still have an Archimedean survival copula with the same generator, as it will appear further in Equation (3.13).
3. In the case where $\theta_{i,j} = 0$ for every $(i, j) \in D^2$ and all the coefficients θ_i are equal, we obtain the class of Schur constant vectors, studied for example in [31]. In that case the function ψ correspond to the generator of the Schur constant vector, which has an Archimedean survival copula in the bivariate case, whose generator is given by ψ^{-1} .

Let us now explicitly construct a distribution on \mathbb{R}_+^d , with the given bivariate marginals. Given the form (3.10) of these marginals, we choose the following characterizing function:

$$m(\mathbf{t}, \mathbf{X}) = \psi^{-1} (\mathbb{P}(X_i > t_i, i \in \mathcal{D})) . \quad (3.12)$$

With the assumptions we made for ψ , it is clear that the function m satisfies Assumption 2.1, with constant $a = +\infty$ and $m_0 = m(\mathbf{t}, \emptyset) = 0$. The function m is thus a suitable characterizing function.

Now consider the decreasing set $\mathcal{S} = \{J \subset \mathcal{D}, |J| \leq 2\}$, and assume that m belongs to the class $\mathcal{F}_{\mathcal{D}}(\mathcal{S})$ in Definition 2.1: each multivariate distribution is assumed here to depend only on its bivariate projections. Assume furthermore that the associated coefficients $\alpha_{J,K}$ in Definition 2.1, $J \subset K$, do only depend on cardinals $j = |J|$ and $k = |K|$, so that $\alpha_{J,K} = \alpha_{j,k}$.

Due to Remark 2.6, if a valid multivariate distribution belongs to the class $\mathcal{F}_{\mathcal{D}}(\mathcal{S})$, then its survival function must take the form

$$\bar{F}_K(\mathbf{t}) = \psi \left(-(k-2) \sum_{J \subset K, |J|=1} m(\mathbf{t}, \mathbf{X}_J) + \sum_{J \subset K, |J|=2} m(\mathbf{t}, \mathbf{X}_J) \right),$$

Notice that $m(t, \mathbf{X}_{\{i\}}) = \theta_i t_i$ and $m(t, \mathbf{X}_{\{i,j\}}) = \theta_i t_i + \theta_j t_j + \theta_{i,j} t_i t_j$. Now using $\sum_{\{i,j\} \subset K} (\theta_i t_i + \theta_j t_j) = (k-1) \sum_{\{i\} \subset K} \theta_i t_i$ we get

$$\bar{F}_K(\mathbf{t}) = \psi \left(\sum_{\{i\} \subset K} \theta_i t_i + \sum_{\{i,j\} \subset K} \theta_{i,j} t_i t_j \right). \quad (3.13)$$

The following Proposition shows that under some sufficient conditions this expression is a proper multivariate survival function.

Proposition 3.2 *The following three conditions ensure that, for any fixed subset K , $\bar{F}_K(\mathbf{t})$ is a proper multivariate survival function:*

(i) ψ and its derivatives goes fast enough to zero, i.e. $\lim_{x \rightarrow +\infty} x\psi^{(n)}(x) = 0$.

(ii) ψ is k -monotone, where $k := |K|$,

(iii) for all distinct i, j in K , $\theta_{i,j} \in [0, \theta_i\theta_j\rho_{\psi,k}]$,

where $\rho_{\psi,k} = \inf_{x \in \mathbb{R}^+, r \leq k/2, r \text{ odd}} \gamma_{k,r}^{-1} \left| \frac{\psi^{(k+1-r)}(x)}{\psi^{(k-r)}(x)} \right|$ and $\gamma_{k,r} = \frac{1}{r} \binom{k-2r+2}{2}$.

For example, if $|K| = k = 3$ and $\psi(x) = \exp(-x)$, then ψ is a k -monotone function satisfying condition (i) and (ii). It also satisfies (iii) with coefficient $\rho_{\psi,k} = \frac{1}{3}$, and the function \bar{F}_K defined in Equation (3.13) is a valid multivariate survival function if $\theta_{i,j} \leq \theta_i\theta_j/3$ for all $i, j \in K$.

Proof. We have

$$\bar{F}_K(\mathbf{t}) = \psi \left(\sum_{\{i\} \subset K} \theta_i t_i + \sum_{\{i,j\} \subset K} \theta_{i,j} t_i t_j \right) = \psi(Q(\mathbf{t})),$$

with

$$Q(\mathbf{t}) = \sum_{\{i\} \subset K} \theta_i t_i + \sum_{\{i,j\} \subset K} \theta_{i,j} t_i t_j.$$

Let us consider without loss of generality $K = \{1, \dots, k\}$ and $\mathbf{t} = (t_1, \dots, t_k)$. Let us define

$$f_K(\mathbf{t}) = (-1)^k \frac{\partial^k}{\partial t_1 \dots \partial t_k} \bar{F}_K(\mathbf{t}).$$

If f_K is a non-negative function whose integral is one, then it will be the density of a random vector, and $\bar{F}_K(\mathbf{t})$ will be a proper multivariate cdf.

Positivity Let us first establish conditions under which f_K is a non-negative function. The multivariate Faa Di Bruno's formula gives

$$f_K(\mathbf{t}) = (-1)^k \sum_{\pi \in \Pi_K} \psi^{(|\pi|)}(Q(\mathbf{t})) \cdot \prod_{B \in \pi} \frac{\partial^{|B|} Q(\mathbf{t})}{\prod_{j \in B} \partial t_j}. \quad (3.14)$$

where Π_K is the set of all partitions of K . $B \in \pi$ means that B runs through all non-empty blocks of a considered partition π . In the following, we will write $\frac{\partial^2}{\partial \mathbf{t}_B} Q(\mathbf{t}) = \frac{\partial}{\partial t_i \partial t_j} Q(\mathbf{t})$, and $\theta_B = \theta_{i,j}$, where $B = \{i, j\}$.

Notice that $\partial^{|B|} Q(\mathbf{t})$ is 0 when $|B| \geq 3$. Thus the only partitions π involved in the calculation contain blocks of 1 or 2 elements only. Hereafter, we denote by Π_K^r the partitions in Π_K that contains exactly r distinct blocs of size 2. For a partition $\pi \in \Pi_K^r$, these r blocs will be denoted B_1^π, \dots, B_r^π . Such a partition $\pi \in \Pi_K^r$ contains r blocs of size 2 and $k - 2r$ blocs of size 1, so that $|\pi| = k - r$. Thus we get (with the convention $\prod_{i=1}^0 = 1$)

$$f_K(\mathbf{t}) = (-1)^k \sum_{r=0}^{\lfloor k/2 \rfloor} \sum_{\pi \in \Pi_K^r} \psi^{(k-r)}(Q(\mathbf{t})) \cdot \prod_{i=1}^r \frac{\partial^2 Q(\mathbf{t})}{\partial \mathbf{t}_{B_i^\pi}} \prod_{j \in K \setminus \cup_i B_i^\pi} \frac{\partial Q(\mathbf{t})}{\partial t_j}.$$

If ψ is k -monotone, $\psi^{(k-r)} = (-1)^{k-r} |\psi^{(n-r)}|$, and setting $\mathcal{N}_k = \{0, \dots, \lfloor k/2 \rfloor\}$,

$$f_K(\mathbf{t}) = \sum_{r \in \mathcal{N}_k} \sum_{\pi \in \Pi_K^r} \left| \psi^{(k-r)}(Q(\mathbf{t})) \right| \cdot \prod_{i=1}^r \left(-\frac{\partial^2 Q(\mathbf{t})}{\partial \mathbf{t}_{B_i^\pi}} \right) \prod_{j \in K \setminus \cup_i B_i^\pi} \frac{\partial Q(\mathbf{t})}{\partial t_j}.$$

One can write $f_K(\mathbf{t}) = \sum_{r \in \mathcal{N}_k} \xi(r)$. Assume that all $\theta_i \geq 0$ and $\theta_{i,j} \geq 0$, $i, j \in K$. Under this assumption, one can check that, when r is even, $\xi(r) \geq 0$. As a consequence,

$$\sum_{r \in \mathcal{N}_k} \xi(r) \geq \sum_{r \in \mathcal{N}_k, r \text{ odd}} [\xi(r) + \xi(r-1)].$$

Let us try to simplify $\xi(r) + \xi(r-1)$. First remark that for $r \geq 1$,

$$|\Pi_K^r| = \binom{k}{2} \cdots \binom{k-2(r-1)}{2} / r! = 2^{-r} k! / (k-2r)! r!$$

and $|\Pi_K^r| = \gamma_{k,r} |\Pi_K^{r-1}|$, with $\gamma_{k,r} = \frac{1}{r} \binom{k-2r+2}{2}$.

Let us write $\xi(r) = \sum_{\pi \in \Pi_K^r} z(B_1^\pi, \dots, B_r^\pi)$. The term $\xi(r-1)$ can be written

$$\sum_{\pi \in \Pi_K^{r-1}} z(B_1^\pi, \dots, B_{r-1}^\pi) = \gamma_{k,r}^{-1} \sum_{\pi \in \Pi_K^r} z(B_1^\pi, \dots, B_{r-1}^\pi).$$

and thus

$$\xi(r) + \xi(r-1) = \sum_{\pi \in \Pi_K^r} \left[z(B_1^\pi, \dots, B_r^\pi) - \gamma_{k,r}^{-1} z(B_1^\pi, \dots, B_{r-1}^\pi) \right].$$

As a consequence, a sufficient condition to ensure that $f_K(\mathbf{t}) \geq 0$ is that: for any $B = \{i, j\}$, for any odd r , for any \mathbf{t} ,

$$-\frac{\partial^2 Q(\mathbf{t})}{\partial \mathbf{t}_{B_i}} \left| \psi^{(k-r)}(Q(\mathbf{t})) \right| + \gamma_{k,r}^{-1} \left| \psi^{(k+1-r)}(Q(\mathbf{t})) \right| \frac{\partial Q(\mathbf{t})}{\partial t_i} \frac{\partial Q(\mathbf{t})}{\partial t_j} \geq 0$$

Thus the sufficient condition to ensure the positivity of $f_K(x)$.

Absolute continuity Let us now check if the integral of f_K is summing to one. If so, \bar{F}_K would be a valid absolutely continuous distribution, without singular component. First assume that for all integers $n \leq k-1$,

$$\lim_{x \rightarrow +\infty} x \psi^{(n)}(x) = 0. \quad (3.15)$$

Recall that $\bar{F}_K(\mathbf{u}) = \psi(Q(u_1, \dots, u_k))$. We now make use of the multivariate Faa Di Bruno's formula as is Equation (3.14). Seen as function of u_n , the derivative of $\bar{F}_K(\mathbf{u})$ with respect to u_{k+1}, \dots, u_d writes as a sum of terms $\psi^{(i)}(au_k + b) \cdot P(u_k)$, where P is a polynomial of degree at most 1 and a, b some real values. Thus under chosen assumption,

$$\lim_{u_n \rightarrow +\infty} \frac{\partial^n}{\partial u_{n+1} \cdots \partial u_k} \bar{F}_K(\mathbf{u}) = 0. \quad (3.16)$$

first show that for any $\mathbf{u} \in \mathbb{R}^k$, for all integer $n \leq k-1$,

$$\lim_{u_n \rightarrow +\infty} \frac{\partial^n}{\partial u_{n+1} \cdots \partial u_k} \bar{F}_K(\mathbf{u}) = 0. \quad (3.17)$$

As a consequence, one can show by recursion that in this case

$$\bar{F}_K(t_1, \dots, t_k) = \int_{t_1}^{+\infty} \cdots \int_{t_k}^{+\infty} \frac{\partial^k}{\partial u_1 \cdots \partial u_k} \bar{F}_K(u_1, \dots, u_k) du_k \cdots du_1. \quad (3.18)$$

Using the fact that $\bar{F}_K(0, \dots, 0) = 1$, we conclude that the derivative function f_K is non-negative and is integrating to one on the whole domain \mathbb{R}_+^k . Under chosen assumptions, it thus defines a proper probability measure and \bar{F}_K is a valid multivariate survival function. \square

4 Practical implementation

4.1 Distance to admissibility

We have seen several examples of projective distributions that can be used in practice, as Elliptically contoured distributions or distributions satisfying Equation (3.13). In the following, we show that the method is also applicable in situations where the data does not necessarily come from a projective distribution.

In practice, it can naturally happen that the multivariate distribution of a considered data is not projective. In such a case, the resulting multivariate function F , obtained from the characterizing function and from multivariate projections by Equation (2.2), may not be a proper cdf. However, by construction, this function have exactly the prescribed multivariate margins.

It is quite a usual problem that estimators do not always satisfy all the required constraints: among many examples, an empirical copula is not a copula ([11]), some Kaplan-Meier estimated distributions are defective ([32]), some estimated quantiles do not satisfy monotonicity ([5]), some nested copulas do not satisfy C-measure positivity on all hyperrectangles ([29, 17]), etc. Here the obtained F , even when it is increasing on each component, may not satisfy the positivity of its cross-derivatives. In this Section, we approximate the function F that we obtain by an admissible cdf.

We show hereafter that, even in the case where the resulting function F is not a cdf, one can find a proper cdf F^+ that is close to F in some sense, and such that its projections are close to the prescribed ones. Furthermore, we will see in numerical illustrations that the maximal distance between F and F^+ can be easily estimated, and that it is very small in some considered applications.

In the following we denote $[-\infty, \mathbf{x}] = \{(s_1, \dots, s_d) : s_1 \leq x_1, \dots, s_d \leq x_d\}$ for any vector $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$.

Proposition 4.1 (Maximal admissibility distance) *Let $\mathcal{D} \subset \mathbb{R}^d$ and denote $\mathcal{D}_{\mathbf{x}} = \mathcal{D} \cap [-\infty, \mathbf{x}]$. Consider a function $F : \mathcal{D} \rightarrow [0, 1]$, and assume that there exists a function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ such that for any $\mathbf{x} \in \mathcal{D}$, $F(\mathbf{x}) = \int_{\mathcal{D}_{\mathbf{x}}} f(\mathbf{s}) d\mathbf{s}$. Assume furthermore that $\int_{\mathcal{D}} f(\mathbf{s}) d\mathbf{s} = 1$. Denote $\mathcal{D}^- = \{\mathbf{s} \in \mathcal{D}, f(\mathbf{s}) < 0\}$ and $\Delta = \int_{\mathcal{D}^-} |f(\mathbf{s})| d\mathbf{s}$. Then there exists a function $F^+ : \mathcal{D} \rightarrow [0, 1]$ such that*

(i) $F^+ : \mathcal{D} \rightarrow [0, 1]$ is a proper multivariate cdf,

(ii) $d_{\text{KS}}(F, F^+) \leq \frac{\Delta}{1+\Delta}$,

(iii) for any $K \subset \{1, \dots, d\}$, $d_{\text{KS}}(F_K, F_K^+) \leq \frac{\Delta}{1+\Delta}$

where $F_K(\mathbf{x}) = F(P_K \mathbf{x})$, $F_K^+(\mathbf{x}) = F^+(P_K \mathbf{x})$ and $P_K \mathbf{x} = (p_1, \dots, p_d)$ with $p_i = x_i$ if $i \in K$, $p_i = +\infty$ otherwise, as defined in (2.1). Here $d_{\text{KS}}(F, G) = \sup_{\mathbf{x} \in \mathcal{D}} |F(\mathbf{x}) - G(\mathbf{x})|$ denotes the Kolmogorov-Smirnov distance between two functions F and G .

Such a function F^+ is given by

$$F^+(\mathbf{x}) = \int_{\mathcal{D}_{\mathbf{x}}} f^+(\mathbf{x}) d\mathbf{x},$$

where $f^+(\mathbf{x}) = \frac{1}{1+\Delta} f(\mathbf{x})$ if $\mathbf{x} \in \mathcal{D} \setminus \mathcal{D}^-$ and $f^+(\mathbf{x}) = 0$ otherwise.

Proof. Let $F^+(\mathbf{x}) = \int_{\mathcal{D}_{\mathbf{x}}} f^+(\mathbf{x}) d\mathbf{x}$, where f^+ is defined in the Proposition, and let $\mathcal{D}^+ := \mathcal{D} \setminus \mathcal{D}^-$ for simplicity. First, f^+ is nonnegative and using $\int_{\mathcal{D}} f(\mathbf{x}) d\mathbf{x} = 1$, we get $\int_{\mathcal{D}^+} f^+(\mathbf{x}) d\mathbf{x} = 1$, so that f^+ a proper pdf and item (i) holds.

When $\Delta > 0$, define $\alpha(\mathbf{x}) \in [0, 1]$ by

$$\alpha(\mathbf{x}) := \frac{\int_{\mathcal{D}_{\mathbf{x}} \cap \mathcal{D}^-} |f(\mathbf{x})| d\mathbf{x}}{\int_{\mathcal{D}^-} |f(\mathbf{x})| d\mathbf{x}}.$$

and $\alpha(\mathbf{x}) = 0$ otherwise. Notice that $\alpha(\mathbf{x})$ is such that $\int_{\mathcal{D}_{\mathbf{x}} \cap \mathcal{D}^-} f(\mathbf{x}) d\mathbf{x} = -\alpha(\mathbf{x})\Delta$, and

$$F^+(\mathbf{x}) = \frac{1}{1 + \Delta} \int_{\mathcal{D}_{\mathbf{x}} \cap \mathcal{D}^+} f(\mathbf{x}) d\mathbf{x} = \frac{1}{1 + \Delta} \left(\int_{\mathcal{D}_{\mathbf{x}}} f(\mathbf{x}) d\mathbf{x} - \int_{\mathcal{D}_{\mathbf{x}} \cap \mathcal{D}^-} f(\mathbf{x}) d\mathbf{x} \right).$$

From this, we get $F(\mathbf{x}) - F^+(\mathbf{x}) = \frac{\Delta}{1 + \Delta} (F(\mathbf{x}) - \alpha(\mathbf{x}))$. As $\sup_{\mathbf{x} \in \mathcal{D}} |F(\mathbf{x}) - \alpha(\mathbf{x})| \leq 1$ the result (ii) holds for this specific function F^+ .

As (ii) holds, $|F(\mathbf{x}) - F^+(\mathbf{x})| \leq \frac{\Delta}{1 + \Delta}$ for any $\mathbf{x} \in D$, and in particular for any $\mathbf{x}' = P_K \mathbf{x}$, so that (iii) holds. \square

Notice that $\Delta = \int_{\mathcal{D}^-} |f(\mathbf{s})| d\mathbf{s} = -\int_{\mathcal{D}^-} f(\mathbf{s}) d\mathbf{s} = -(1 - \int_{\mathcal{D}^+} f(\mathbf{s}) d\mathbf{s}) = \int_{\mathcal{D}^+} f(\mathbf{s}) d\mathbf{s} - 1$, so that $\int_{\mathcal{D}^+} f(\mathbf{s}) d\mathbf{s} = 1 + \Delta$. As a consequence, one can define the following failure ratio

$$R = \frac{\int_{\mathcal{D}^-} |f(\mathbf{s})| d\mathbf{s}}{\int_{\mathcal{D}^+} |f(\mathbf{s})| d\mathbf{s}} = \frac{\Delta}{1 + \Delta}, \quad (4.1)$$

which bounds the errors in Proposition 4.1. This ratio can easily be estimated by a discrete approximation of each integral, using classical techniques which avoid any normalization of the integrals.

In the case where $m(\mathbf{t}, X) = \varphi \circ F(\mathbf{t})$ where φ is a bijection from $(0, 1]$ to \mathbb{R}^+ , one gets

$$F_L(\mathbf{t}_L) = \varphi^{-1} \left(\sum_{J \subset L, J \in \mathcal{S}} \alpha_{J,L} \cdot \varphi \circ F_J(\mathbf{t}_J) \right).$$

In this case, if φ , φ^{-1} and all given projections F_K are differentiable up to a sufficient order, then applying chain differentiation rules and multivariate Faa Di Bruno's formula, one can show that F satisfies the assumptions of Proposition 4.1.

One can notice that in practice, depending on the application, and when Δ is small, one may use F instead of F^+ , as the expression of F do not require to compute an integral. In particular, some sampling procedures like MCMC may be easily adapted to the function F instead of F^+ . In the latter case, the produced sample have by definition a valid multivariate cdf.

4.2 Numerical illustration on a real data set

In the following, we give an example on a real data set. First, we illustrate a natural fit procedure on a trivariate data, by adjusting each marginal distribution, and then each copula. Second, we show that the proposed construction allows to build a valid trivariate cumulative distribution with projections close enough to each of the prescribed ones.

Marginals and copulas fit The purpose here is to provide reasonable univariate and bivariate fits using standard tools, without distorting the data, in order to illustrate the flexibility of our result and its applicability on some usual data. Surely, better fits can be proposed, but are out of the scope of the present paper.

We have obtained best univariate fits using classical existing tools in R software, and in particular the package `fitdistrplus`. Best copulas were obtained by the package `VineCopula`.

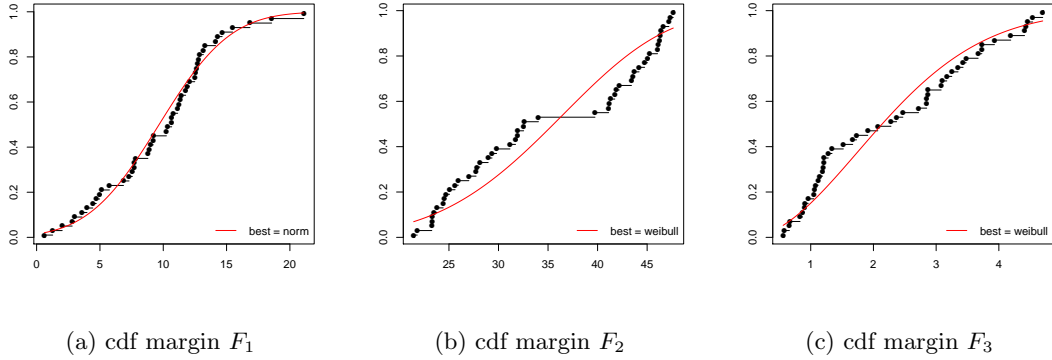


Figure 2: Univariate margins fits: empirical cdf (discontinuous line, black) and fitted cdf (continuous line, red) obtained by R software package `fitdistrplus`.

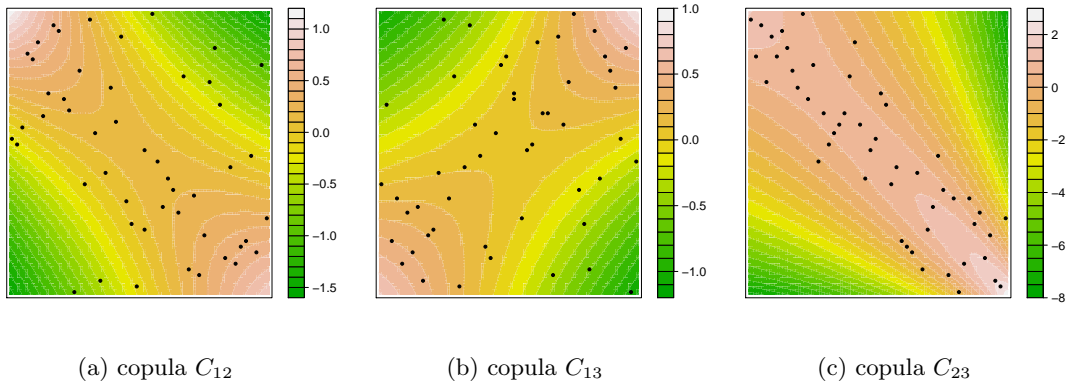


Figure 3: Copula fits: pseudo observations scatterplot and heatmap of the copula log-density, obtained by R software package `VineCopula`.

We present here the results obtained for the data `LifeCycleSavings` from the standard R Software library `datasets`. We have used the first three columns of this data. We do not detail here the data, as the purpose here is just to build a parametric fit on a multivariate data, with given projections.

The marginal fits obtained for the considered data are gathered in Figure 2. One can see that, independently of our method, the fits with usual tools may be sensitive to multimodality of the data. The copula fit illustrations are presented in Figure 3. All univariate and bivariate fits are summarized in Table 1.

Concerning both the cdf and the copulas, they exhibit quite different shapes, as it usually happens on many datasets. It thus seems particularly challenging to propose a trivariate function having exactly those fitted univariate and bivariate margins. In the next paragraph we show how to build such a function, using the method proposed in this article.

object	parametric fit
F_1	norm(9.671,4.43537653418512)
F_2	weibull(4.48356587156989,38.6038169621111)
F_3	weibull(1.89863954400289,2.59468797211236)
C_{12}	Rotated BB8 90 degrees(-6,-0.396616316431917)
C_{13}	BB8(6,0.316859224873494)
C_{23}	Rotated BB8 270 degrees(-6,-0.857389852442906)

Table 1: Fits obtained for the considered dataset, obtained by R software packages `fitdistrplus` and `VineCopula`. Parametric expressions of the fits are those indicated in these packages.

Trivariate fit with prescribed bivariate projections For more flexibility, we consider here a one-parameter class of characterizing functions, in the sense of Definition 2.1. As the given bivariate fits rely here on copulas, it is easier to deal with multivariate cdf or survival functions. The considered class is a parametric transformation of multivariate cdf:

$$m(\mathbf{t}, \mathbf{X}_L) = \varphi_\theta [F_{\mathbf{X}_L}(\mathbf{t}_L)] . \quad (4.2)$$

For the link function φ_θ , we have chosen some parametric monotone functions such that $\varphi_\theta(1) = 0$, detailed hereafter. One can check that in this case, $m_0 = m(\mathbf{t}, \mathbf{X}_\emptyset) = \varphi_\theta(F_{\mathbf{X}}(+\infty, \dots, +\infty)) = 0$ as defined in Remark 2.1. Using the proposed method and the result in Remark 2.6, when $|L| = 3$ and with $\mathbf{t}_L \in \mathbb{R}^3$, we obtain a fitted trivariate function, denoted F_θ :

$$F_\theta(\mathbf{t}_L) = \varphi_\theta^{-1} \left[-(|L| - 2) \sum_{J \subset L, |J|=1} \varphi_\theta(F_{\mathbf{X}_J}(\mathbf{t}_J)) + \sum_{J \subset L, |J|=2} \varphi_\theta(F_{\mathbf{X}_J}(\mathbf{t}_J)) \right] . \quad (4.3)$$

Notice that for some specific link functions and margins, theoretically valid distributions with this shape are given in Proposition 3.2.

As it was easier to use classical expressions with known parametric inverse functions, we have used specific link functions φ_θ . We have considered strictly positive and decreasing functions $\varphi_\theta : [0, 1] \rightarrow \mathbb{R}^+$ of Table 4.1. in [30], that are also known to be Archimedean generators. We have tried the six first strict generators of this table 4.1. (respectively Clayton, AMH, Gumbel, Frank, Joe and Hougaard generators). Then, we have selected the generator and its parameter that was minimizing the estimation of the failure ratio $R = \frac{\Delta}{1+\Delta}$ in Equation (4.1). The parameter θ is thus used to reduce the maximal distance between the obtained function F_θ and a proper multivariate cdf F_θ^+ , as detailed in Proposition 4.1. We also compared the trivariate function F_θ and the empirical trivariate cdf F_{emp} of the data, by computing the average absolute error $\delta = \frac{1}{n} \sum_{i=1}^n |F_\theta(\mathbf{x}_i) - F_{\text{emp}}(\mathbf{x}_i)|$, where $\mathbf{x}_1, \dots, \mathbf{x}_n$ are the points in the dataset. The Figure 4 shows the different values of the estimated failure ratio R and the distance δ with the empirical cdf. The ratio R was estimated by replacing the integrals by sums over a regular grid of 1000 points between minimal and maximal values of each component of the dataset. Whereas the average absolute error δ is quite stable on this data, the failure ratio R is sensitive to the choice of the parameter θ . We verify hereafter that the results are not too sensitive to the grid used for the estimation.

Finally, the results for this dataset, using the Hougaard generator φ_θ given by (4.2.9) in Table 4.1 of [30], are gathered in the following table:

dataset (columns)	link function φ_θ (parameter θ)	estimation of R	distance δ
LifeCycleSavings(1,2,3)	(4.2.9) Hougaard ($\theta = 0.768$)	0.000872	0.0332

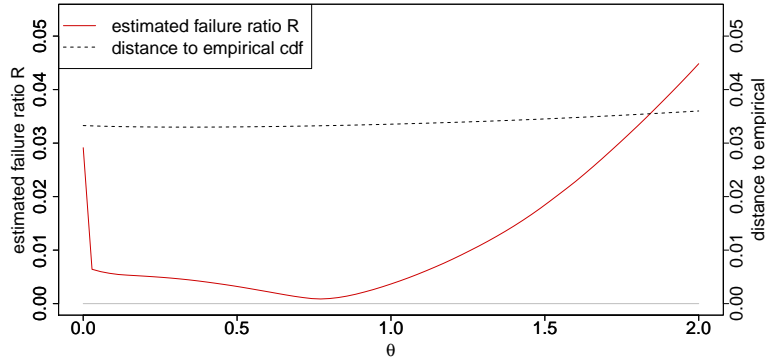


Figure 4: Estimated failure ratio R (red solid line) and distance δ to empirical cdf (black dashed line), as a function of the parameter θ .

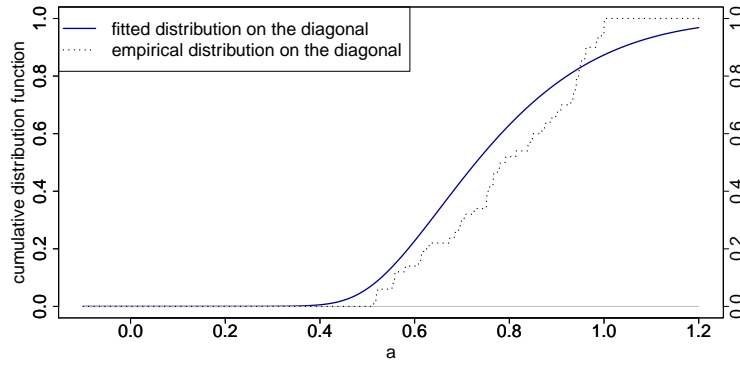


Figure 5: Values of the fitted function F_θ (blue solid line) and the empirical cdf F_{emp} (black dotted line), on several points of a diagonal line $\mathbf{m} + a(\mathbf{M} - \mathbf{m})$ in \mathbb{R}^3 , as a function of the abscissa a .

For the estimation of R , using a more precise grid of 64000 points, in the extended domain $[0.9\mathbf{m}, 1.1\mathbf{M}]$ where \mathbf{m} and \mathbf{M} are the componentwise minimum and maximum of the points in the data, we obtained quite similar results: an estimated R equal to 0.000984 at $\theta = 0.766$. Notice that it is usual that some compatibility conditions can only be verified numerically over a grid of points, see [21], page 75.

One can check that the estimated admissibility of the fitted function F_θ is very good, so that in practice, as one can see on Figures 5 and 6, it may be unnecessary to compute the proper cdf F_θ^+ of Proposition 4.1. Indeed, the numerical approximations involved by numerical differentiation or integration may be greater to the maximal distance Δ and the failure ratio $R \leq \Delta$.

Values of the fitted function on given bivariate projections correspond exactly, by construction, to the prescribed ones so that it is useless to draw these values. One can instead compute the values of the fitted function for a set of points belonging to a 1D or a 2D diagonal hyperplane. The considered data has 3 columns. For each column i in $\{1, 2, 3\}$, denote m_i the minimal observed value in this column, over all observations, and M_i the maximal observed value. Let $\mathbf{m} = (m_1, m_2, m_3)$ and $\mathbf{M} = (M_1, M_2, M_3)$ (so that the cube $[\mathbf{m}, \mathbf{M}]$ contains all observations of the dataset). Let us also define the two points $A = (M_1, m_2, M_3)$ and $B = (m_1, M_2, M_3)$.

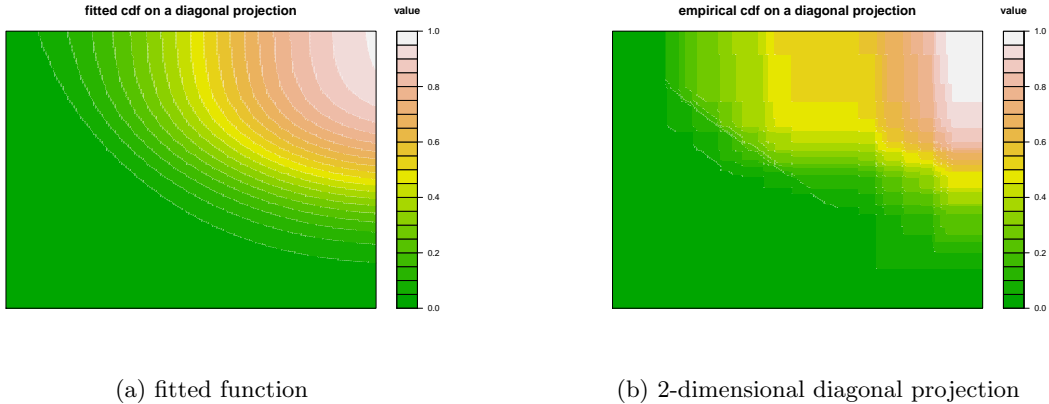


Figure 6: Level curves of the fitted function F_θ (left panel) and the empirical cdf F_{emp} (right panel), on several points of the diagonal plane $\mathbf{m} + a(\mathbf{A} - \mathbf{m}) + b(\mathbf{B} - \mathbf{m})$, as a function of the abscissa a and the ordinate b (right panel).

Suppose that the true distribution of the data set is given by the law of a vector (X_1, X_2, X_3) and suppose that we are interested in estimating the cdf of the univariate random variable $Z := \max(\frac{X_1 - m_1}{M_1 - m_1}, \frac{X_2 - m_2}{M_2 - m_2}, \frac{X_3 - m_3}{M_3 - m_3})$. In Figure 5, we have drawn the estimated cdf of Z coming from our construction, using $\mathbb{P}[Z \leq a] = \mathbb{P}[X_1 \leq m_1 + a(M_1 - m_1), \dots, X_3 \leq m_3 + a(M_3 - m_3)]$. More precisely, the figure shows the values of the function F_θ that we constructed, evaluated at several points of a diagonal $\mathbf{m} + a(\mathbf{M} - \mathbf{m})$, for different values of a . Despite that the fit is not perfect compared to the empirical values (the fits of the bivariate projections in Figure 2 were not perfect either), we observe that the estimated univariate cdf behaves normally (starting from 0 up to 1, and increasing).

Along the same lines, on the right panel of Figure 6, we have drawn values of the fitted function F_θ for points belonging to a plane $\mathbf{m} + a(\mathbf{A} - \mathbf{m}) + b(\mathbf{B} - \mathbf{m})$, for different values of a and b (left panel), the empirical counterpart is also drawn (right panel).

One can check that, again, the fitted function behaves normally and is increasing over each component: the admissibility problem can arise even for functions increasing on each components, with values in $[0, 1]$, as this requirement is not sufficient to define a cdf having positive cross derivatives.

At last, we have conducted the same analysis on other datasets of the R Software library **datasets**. Results are gathered in the Table 2. For some data (**stackloss**, **rock**), the estimated values of R are not negligible and more investigations are needed: change in the bivariate projections, in the one-parameter link function φ_θ or in the characterizing function m . For other datasets, we obtain very good numerical results, with sometimes estimated ratios R of order 10^{-6} , and a data as **MathAchieve** also leads to a very good global fit of the trivariate distribution. In our experiments, we did not use the (tedious) parametric expressions of the bivariate projections density, but instead some numerical differentiation. As a result, the distance between the proper cdf F^+ in Proposition 4.1 and the obtained fitted function would probably be comparable to the numerical errors induced by the numerical differentiation, so that we did not build this proper cdf. We have presented here a detailed procedure for a specific one-parameter class of characterizing functions. Introducing more parameters would logically result in smaller ratios R .

The proposed method cannot guarantee that the final fitted multivariate function F does not correspond to a signed measure. However, as a conclusion of this numerical investigation, the following advantages are of practical utility:

dataset (columns)	link function φ_θ (parameter θ)	estimation of R	distance δ
MathAchieve(4,5,6)	(4.2.1) Clayton ($\theta = 1.10$)	0.00000392	0.006324558
LifeCycleSavings(1,3,5)	(4.2.5) Frank ($\theta = 2.15$)	0.000136078	0.075148513
LifeCycleSavings(2,3,5)	(4.2.9) Hougaard ($\theta = 0.66$)	0.000243695	0.071234939
LifeCycleSavings(1,2,5)	(4.2.9) Hougaard ($\theta = 0.96$)	0.000365772	0.068639313
LifeCycleSavings(1,2,3)	(4.2.9) Hougaard ($\theta = 0.768$)	0.000872	0.0332
airquality(1,2,4)	(4.2.1) Clayton ($\theta = 0.61$)	0.003274217	0.026502691
airquality(1,3,4)	(4.2.5) Frank ($\theta = -3.57$)	0.022312892	0.010637131
mtcars(1,5,6)	(4.2.5) Frank ($\theta = -5.55$)	0.032730639	0.026896243
mtcars(3,4,6)	(4.2.4) Gumbel ($\theta = 5.95$)	0.033680629	0.034147215
swiss(1,2,3)	(4.2.9) Hougaard ($\theta = 1.12$)	0.056176129	0.013166825
trees(1,2,3)	(4.2.1) Clayton ($\theta = 1.24$)	0.058231572	0.021837067
stackloss(2,3,4)	(4.2.5) Frank ($\theta = 3.14$)	0.076542359	0.037345426
rock(1,2,3)	(4.2.6) Joe ($\theta = 1.00$)	0.099751509	0.028614845

Table 2: Final estimated failure ratio R and distance δ to empirical, for different datasets or for different considered columns Results are sorted by estimated ratio R . Detail of each bivariate fits are omitted here. The horizontal dashed lines highlights the previously detailed case study. The horizontal plain line separate the cases with estimated failure ratio greater than 0.05.

- (i) The obtained fitted distribution has exactly the prescribed projections, by construction.
- (ii) The Kolmogorov-Smirnov distance between the final fitted function and a proper cdf is bounded by a quantity R that can be estimated, and is small in our experiments. Thus, the fitted function can be directly used for many practical applications.
- (iii) One can build a proper cdf from the final fitted function. In such a case, the projections of this proper cdf are at maximal distance R from the prescribed ones.

5 Conclusion

We have considered in this paper specific multivariate distributions, belonging to a class which was called projective. They rely on a linear link between some functional of the considered distribution and their multivariate margins. The choice of a linear link is not as restrictive as one would initially imagine, since it covers a variety of classical distributions, from Elliptical ones to some natural survival models, as presented in Section 3.

In theory, for those distributions, the compatibility between given multivariate margins and multivariate distribution is automatically ensured, by definition, and the coefficients linking multivariate margins with the whole distribution are easily obtained using for example Proposition 2.2, Equation (2.14) and Equation (2.15).

In practice, when dealing with a given data, one possibility is to use a class projective distributions, as those described in Section 3 (Elliptical distributions, specific survival model, etc.), and estimate its parameters. This way, fitted projections are necessarily compatible with each other, and the admissibility is ensured for the resulting multivariate construction having prescribed projections. However, the fitted projections are then modeled by the same parametric family of functions.

Another possibility is to assume the validity of the linearity assumption for some characterizing functions belonging to a set of functions, as described in Section 4. In that case, given

a data, fitted multivariate marginals, and a chosen characterizing function m , the coefficients and the expression of the candidate function for the whole multivariate distribution are easily obtained. This allows a huge variety of fitted projections. It remains to verify whether the fitted function with prescribed projections is a valid distribution, as done theoretically in the Section 3.2 (but it involves many chain rules differentiations), or numerically in Section 4.2. In the latter case, it is always possible to build a proper multivariate cdf while controlling the distance to the prescribed projections, as detailed in Section 4.1.

At last, other characterizing functions can be tried, eventually relying on several parameters. This way, one can build theoretically new classes of projective distributions, or try to satisfy, at least numerically, the validity of the fitted functions on some data.

Acknowledgment

This paper was written during an invited stay of the authors at the Vietnam Institute for Advanced Study in Mathematics (VIASM). The authors warmly thank the VIASM institute for his support. The authors are very grateful to the anonymous reviewers, the associate editor and the editor, for their careful reading and their many insightful comments and suggestions.

References

- [1] AAS, K., CZADO, C., FRIGESSI, A., AND BAKKEN, H. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics* 44, 2 (2009), 182–198.
- [2] ACAR, E. F., GENEST, C., AND NELEHOV, J. Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis* 110 (2012), 74 – 90. Special Issue on Copula Modeling and Dependence.
- [3] BAXTER, J. R., AND CHACON, R. V. Potentials of stopped distributions. *Illinois J. Math.* 18, 4 (12 1974), 649–656.
- [4] BRECHMANN, E. C. Hierarchical kendall copulas: Properties and inference. *Canadian Journal of Statistics* 42, 1 (2014), 78–108.
- [5] CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I., AND GALICHON, A. Quantile and probability curves without crossing. *Econometrica* 78, 3, 1093–1125.
- [6] COHEN, L. Probability distributions with given multivariate marginals. *Journal of Mathematical Physics* 25, 8 (1984), 2402–2403.
- [7] CUADRAS, C. Probability distributions with given multivariate marginals and given dependence structure. *Journal of Multivariate Analysis* 42, 1 (1992), 51 – 66.
- [8] DALL’AGLIO, G., KOTZ, S., SALINETTI, G., AND UNIVERSITÀ DEGLI STUDI DI ROMA ”LA SAPIENZA.” DIPARTIMENTO DI STATISTICA, P. E. S. A. *Advances in Probability Distributions with Given Marginals: Beyond the Copulas [; Lectures Presented at a ”Symposium on Distributions with Given Marginals” Organized by the Dept. of Statistics of the University La Sapienza, Rome, Italy, Held in Rome in April 1990]*. Mathematics and Its Applications. Springer Netherlands, 1991.
- [9] DI BERNARDINO, E., AND RULLIÈRE, D. On an asymmetric extension of multivariate archimedean copulas based on quadratic form. *Dependence Modeling* 4, 1 (2016), 328–347.

- [10] DOLATI, A., AND ÚBEDA FLORES, M. A method for constructing multivariate distributions with given bivariate margins. *Brazilian Journal of Probability and Statistics* 19 (2005), 85–92.
- [11] ERDELY, A., AND GONZÁLEZ-BARRIOS, J. M. Exact distribution under independence of the diagonal section of the empirical copula. *Kybernetika* 44, 6 (2008), 826–845.
- [12] FANG, K., KOTZ, S., AND NG, K. *Symmetric multivariate and related distributions*. Monographs on statistics and applied probability. Chapman and Hall, 1990.
- [13] FRAHM, G. *Generalized Elliptical Distributions: Theory and Applications*. PhD. thesis from the University of Köln, Germany. 2004.
- [14] FRAHM, G., JUNKER, M., AND SZIMAYER, A. Elliptical copulas: applicability and limitations. *Statistics and Probability Letters* 63, 3 (2003), 275 – 286.
- [15] GENEST, C., QUESEDA MOLINA, J. J., AND RODRIGUEZ LALLENA, J. A. De l'impossibilité de construire des lois marges multidimensionnelles données partir de copules. *Comptes rendus de l'Académie des sciences. Série 1, Mathématique* 320, 6 (1995), 723–726. fre.
- [16] GUTMANN, S., KEMPERMAN, J. H. B., REEDS, J. A., AND SHEPP, L. A. Existence of probability measures with given marginals. *Ann. Probab.* 19, 4 (10 1991), 1781–1797.
- [17] HOFERT, M., AND PHAM, D. Densities of nested archimedean copulas. *Journal of Multivariate Analysis* 118 (2013), 37 – 52.
- [18] JOE, H. Parametric families of multivariate distributions with given margins. *Journal of Multivariate Analysis* 46, 2 (1993), 262 – 282.
- [19] JOE, H. Multivariate extreme-value distributions with applications to environmental data. *Canadian Journal of Statistics* 22, 1 (1994), 47–64.
- [20] JOE, H. *Families of m-variate distributions with given margins and $m(m - 1)/2$ bivariate dependence parameters*, vol. Volume 28 of *Lecture Notes–Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1996, pp. 120–141.
- [21] JOE, H. *Multivariate Models and Multivariate Dependence Concepts*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1997.
- [22] JOE, H., LI, H., AND NIKOLOULOPOULOS, A. K. Tail dependence functions and vine copulas. *Journal of Multivariate Analysis* 101, 1 (2010), 252–270.
- [23] KANO, Y. Consistency property of elliptic probability density functions. *Journal of Multivariate Analysis* 51, 1 (1994), 139 – 147.
- [24] KELLERER, H. G. Distribution functions with given marginal distributions (german: Verteilungsfunktionen mit gegebenen marginalverteilungen). *Zeitschrift fr Wahrscheinlichkeitstheorie und verwandte Gebiete* 3 (1964), 247–270.
- [25] LI, H., SCARSINI, M., AND SHAKED, M. Linkages: A tool for the construction of multivariate distributions with given nonoverlapping multivariate marginals. *Journal of Multivariate Analysis* 56, 1 (1996), 20 – 41.
- [26] LI, H., SCARSINI, M., AND SHAKED, M. Dynamic linkages for multivariate distributions with given nonoverlapping multivariate marginals. *Journal of Multivariate Analysis* 68, 1 (1999), 54 – 77.

- [27] MARCO, J., AND RUIZ-RIVAS, C. On the construction of multivariate distributions with given nonoverlapping multivariate marginals. *Statistics & Probability Letters* 15, 4 (1992), 259 – 265.
- [28] MARONNA, R., MARTIN, D., AND YOHAI, V. *Robust Statistics: Theory and Methods*. Wiley Series in Probability and Statistics. Wiley, 2006.
- [29] MCNEIL, A. J. Sampling nested archimedean copulas. *Journal of Statistical Computation and Simulation* 78, 6 (2008), 567–581.
- [30] NELSEN, R. *An Introduction to Copulas*. Lecture notes in statistics. Springer, 1999.
- [31] NELSEN, R. B. Some properties of schur-constant survival models and their copulas. *Brazilian Journal of Probability and Statistics* 19, 2 (2005), 179–190.
- [32] PORTNOY, S. Censored regression quantiles. *Journal of the American Statistical Association* 98, 464 (2003), 1001–1012.
- [33] RÜSCHENDORF, L. Construction of multivariate distributions with given marginals. *Annals of the Institute of Statistical Mathematics* 37, 1 (1985), 225–233.
- [34] SÁNCHEZ ALGARRA, P. Construcción de distribuciones con marginales multivariantes dadas. *Qüestió* 10, 3 (1986), 133–141.
- [35] SKLAR, M. *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8, 1959.
- [36] SPANIER, E. H. *Algebraic topology*, vol. 55. Springer Science & Business Media, 1994.