

## ON A CONSTRUCTION OF MULTIVARIATE DISTRIBUTIONS GIVEN SOME MULTIDIMENSIONAL MARGINALS

NABIL KAZI-TANI,\* \*\* AND  
DIDIER RULLIÈRE,\* \*\*\* *Université Lyon 1*

### 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 marginals. In the case where the data does 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 marginal; copula; elliptical distribution

2010 Mathematics Subject Classification: Primary 60E05; 60E10  
Secondary 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 a 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 [36] 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 are of particular interest because, as 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  $X_K$  to denote the random vector with values in  $\mathbb{R}^k$  given by  $(X_{j_1}, \dots, X_{j_k})$  and  $F_K$  the cumulative

Received 10 January 2018; revision received 1 March 2019.

\* Postal address: Laboratoire SAF, ISFA, Université Lyon 1, 50 Avenue Tony Garnier, F-69366 Lyon Cedex 07, France.

\*\* Email address: [nabil.kazi-tani@univ-lyon1.fr](mailto:nabil.kazi-tani@univ-lyon1.fr)

\*\*\* Email address: [didier.rulliere@univ-lyon1.fr](mailto:didier.rulliere@univ-lyon1.fr)

distribution function of  $X_K$ . We will abuse notation 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

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

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. We could also try to extend the notion of Copula functions to the case of nonoverlapping multidimensional marginals. Genest *et al.* [15] showed that this approach is useless, since it only allows modeling of 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  $C(u, v) = uv$ .

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: for instance, 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 Sections 3.4.3–3.7, 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 was proved that if  $\mu$  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  $\mu$ . 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], [7], [10], [18], [19], [20], [25], [26], [27], [34], and [35] for overlapping or nonoverlapping 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 two-dimensional copula is doable, and then construct a valid  $d$ -dimensional vector having the prescribed two-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 (hierarchical, vine copulas, or nested, see [4], [22], [29], 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 multidimensional marginals (Definition 2.1). We will say that a probability distribution satisfying Definition 2.1 is in a *projective class*. The main result of this 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). Note 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 nonlinear transformation to a characterizing function, and then obtain a linear relation in the form of (2.2). Stated otherwise, given a family of probability distributions, if some linearity can be found in the expression of one of its characterizing functions, then our approach allows us to exploit this linearity to construct multidimensional distributions. The case of elliptical random vectors effectively illustrates 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. In Section 4 we also propose some practical implementations.

## 2. Projective class of random vectors

### 2.1. Definitions

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

We assume the existence of a link between the joint distribution of  $X$  and the joint distributions of its projections  $X_K$ . In general, this link could be derived from the characteristic function of  $X$ , 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 depend only 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.** (Projective 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 nonempty  $K \in \mathcal{P}(\mathcal{D})$ ,*

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

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

In the rest of the paper, to simplify notation, we will denote  $m(t, X, K)$  by  $m(t, X_K)$ .

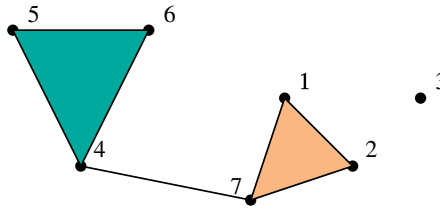


FIGURE 1: Illustration of a multivariate distribution in the dimension 7, knowing the marginal distribution {3}, the bivariate projection {4, 7}, and the trivariate projections {1, 2, 7} and {4, 5, 6}. All subsets of these projections also correspond to known marginal distributions.

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

**Remark 2.2.** Such a function  $m$  always exists. Typically,  $m$  can be given by  $m(t, X_K) = \mathbb{E}[e^{it_K^\top X_K}]$ , with  $a = 0$  and  $m_0 = 1$ . Another example is  $m(t, X_K) = F_K(t_K)$ , the CDF of  $X_K$ , with  $a = +\infty$  and  $m_0 = 1$ . A third example is  $m(t, X_K) = \phi^{-1} \circ F_K(t_K)$ , 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(ii) states 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  $P_K t$ . However, note that not every function characterizing the distribution of random variables satisfies this assumption. Let us give an example of the potential function (see [3]). Define the potential kernel  $v$  on  $\mathbb{R}^d$  as

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

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

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

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

We aim at defining the whole distribution of  $X$  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 allows us to easily know the distribution of every subvector. In the algebraic topology terminology,  $\mathcal{S}$  is a simplicial complex [37]. 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).

**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  $X \in \mathbb{R}^d$  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  $t \in \bar{\mathbb{R}}^d$ ,

$$m(t, X) = \sum_{K \in \mathcal{S}} \alpha_{K, \mathcal{D}} m(t, X_K). \tag{2.2}$$

We note that if a random vector  $X$  belongs to the projective class  $\mathcal{F}_{\mathcal{D}}(\mathcal{S})$  then the set of suitable coefficients  $\{\alpha_{K,\mathcal{D}}, K \subset \mathcal{D}\}$  satisfying (2.2) is not necessarily unique. Note 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, note that *a priori*, the coefficients  $\alpha_{K,\mathcal{D}}$  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 the choice of  $m$  (see Remark 2.6).

It is not trivial to assess the compatibility conditions between arbitrary characterizing functions of the projections; given arbitrary  $m(t, X_K)$  in (2.2), determining if the resulting  $m(t, X)$  indeed corresponds to a characterizing function is difficult. As an example, if  $m$  is a characteristic function, the verification may rely on known criteria such 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 [8], [21], and [24], for supplementary material on the question of projections compatibility.

### 2.2. Properties

In this subsection we discuss several properties of projective distributions regarding uniqueness, stability, and statistical inference. In particular, we provide explicit expressions for the coefficients appearing in Definition 2.1, in the case of fixed projections up to a given dimension.

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

**Proposition 2.1.** (Condition on the projections for uniqueness.) *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  $M_S(\mathcal{T}) := (m(t, X_K))_{t \in \mathcal{T}, K \in S}$ . Assume that the distribution of  $X$  is projective, and denote the vector of coefficients  $\alpha := (\alpha_{K,\mathcal{D}})_{K \in S}$ , then the following results hold.*

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

**Remark 2.4.** The combination of (i) and (ii) above implies the following necessary and sufficient condition for uniqueness:  $\alpha$  is unique if and only if there exists a set  $\mathcal{T}$  with  $|\mathcal{T}| = |S|$  such that the matrix  $M_S(\mathcal{T})$  is invertible.

*Proof of Proposition 2.1.* It is clear that if  $m_0 = 0$  then  $\alpha_{\emptyset, \mathcal{D}} m(t, X_{\emptyset}) = 0$  whatever the value of  $\alpha_{\emptyset, \mathcal{D}}$ . 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 (i). Write the vector  $m_X(\mathcal{T}) = (m(t, X))_{t \in \mathcal{T}}$ . The main linear equality of (2.2) is

$$m_X(\mathcal{T}) = M_S(\mathcal{T})\alpha \tag{2.3}$$

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

To prove (ii), consider a set  $\mathcal{T}$  of distinct points with  $|\mathcal{T}| = |S|$ . Again, (2.2) can be written as (2.3). This linear system of  $|S|$  equations admits either no solution, which is excluded by the assumption that the distribution of  $X$  is projective, an infinite number of solutions, which is excluded by the assumption that  $\alpha$  is unique, or a unique solution, which is the only possible case. This implies, by the Rouché–Capelli theorem, that the rank of the matrix  $M_S(\mathcal{T})$  is equal to  $|S|$ , or, equivalently, that this matrix is invertible.

For (iii), write  $S = \{K_1, \dots, K_s\}$ , set  $t \in \mathbb{R}^d$ , and set  $\mathcal{T} = \{t_K, K \in S\}$ . Write  $P$  the matrix with components  $P_{ij} = 1$  if  $K_i \subset K_j$  or 0 otherwise. We can check that the component  $i, j$  of the matrix  $M_S(\mathcal{T})$  is

$$(M_S(\mathcal{T}))_{ij} = m(t_{K_i}, X_{K_j}) = m(t, X_{K_i \cap K_j}).$$

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

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

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 on the right-hand side,  $K \in S, J \subset K, J \neq K$  implies that  $|J| = 1$  and  $|K| = 2$ , and  $K \subset L$  with  $|K| = |L| = 2$  implies that  $K = L$ , i.e.

$$\sum_{K \in S, K \subsetneq L} m(t, X_K) = \sum_{K \in S, K \subset L} \sum_{J \subset K, J \neq K} m(t, X_J).$$

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

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

so that  $M_S(\mathcal{T}) = P^{\top}DP$ , where  $D$  is the diagonal matrix with diagonal  $\{m_K^*(t), K \in S\}$ . 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 (iii),  $\det(D) > 0$  and the result holds. □

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 belongs to  $\mathcal{F}_{\mathcal{D}}(\mathcal{S})$ , and we compute the corresponding coefficients.

**Proposition 2.2.** (Projection stability.) *Let  $X$  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 nonempty  $L \subset \mathcal{D}$ , the subvector  $X_L$  belongs to  $\mathcal{F}_L(\mathcal{S})$ , where for any nonempty subset  $J$  of  $L$ , a suitable set of associated coefficients is given by*

$$\alpha_{J,L} = \sum_{K \in \mathcal{S}, K \cap L = J} \alpha_{K,\mathcal{D}}, \tag{2.4}$$

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  $\{\alpha_{J,L}, J \subset L\}$  depend only 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

$$\alpha_{j,\ell} = \sum_{k=j}^{\min(d_0, d-\ell+j)} \binom{d-\ell}{k-j} \alpha_{k,d} \quad \text{if } j \leq \min(\ell, d_0), \tag{2.5}$$

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

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

$$\begin{aligned} m(t, X_L) &= m(P_L t, X) \\ &= \sum_{K \in \mathcal{S}} \alpha_{K,\mathcal{D}} m(P_L t, X_K) \\ &= \sum_{K \in \mathcal{S}} \alpha_{K,\mathcal{D}} m(t, X_{K \cap L}) \\ &= \sum_{J \subset L} \sum_{K \in \mathcal{S}, K \cap L = J} \alpha_{K,\mathcal{D}} m(t, X_J). \end{aligned}$$

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 (2.4), we finally obtain

$$m(t, X_L) = \sum_{J \subset L, J \in \mathcal{S}} \alpha_{J,L} m(t, X_J);$$

hence,  $X_L$  belongs to  $\mathcal{F}_L(\mathcal{S})$ .

Finally, note 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 (2.4) is empty and  $\alpha_{J,L} = 0$ .

We now prove the second part of the proposition. When  $\mathcal{S} = \{K \subset \mathcal{D}, |K| \leq d_0\}$ , we have, using (2.4),

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

Now, let  $j = |J|$  and  $\ell = |L|$ . If  $j > \ell$ , it is clear that  $J \not\subset L$  and  $\alpha_{J,L} = 0$ . 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  $J \subset L$ ,

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

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

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

yielding the second result. □

**Remark 2.5.** (*Case of  $\emptyset$ .*) Note that  $\emptyset$  necessarily belongs to  $\mathcal{S}$ . In the case where  $m(t, X_\emptyset) \neq 0$ , (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 nonempty  $L$ , (2.4) becomes

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

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

As previously noted, 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  $\tilde{m}(t, X) = cm(t, X)$  for a positive constant  $c$  leads to  $\tilde{m}(t, X_\emptyset) = cm(t, X_\emptyset)$  and to unchanged coefficients  $\tilde{\alpha}_{K, \mathcal{D}} = \alpha_{K, \mathcal{D}}$ .

**Corollary 2.1.** (Given projections up to dimension  $d_0$ .) *Let  $X$  be a  $d$ -dimensional random vector in  $\mathcal{F}_{\mathcal{D}}(\mathcal{S})$ . Assume that all projections of  $X$  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  $\{\alpha_{J,L}, J \subset L\}$  depend only 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*

$$\alpha_{d_0-z,d} = 1 - \sum_{i=1}^z \binom{d-d_0+z}{i} \alpha_{i+d_0-z,d} \tag{2.6}$$

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

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

*Proof.* Under the assumption that  $\alpha_{k,k} = 1$  for all  $k \leq d_0$ , this follows directly from Proposition 2.2, by writing (2.5) in the case where  $j = \ell \leq d_0$ , and setting  $i = k - j, j = d_0 - z$ . We obtain  $\alpha_{d_0,d} = 1$  when  $z = 0$ , and (2.6) when  $z \geq 1$ . □

**Remark 2.6.** (*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 depend only 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 variables 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 2.7.** (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 nonempty  $L \subset \mathcal{D}$ , we can reformulate (2.2) as

$$m(t, X_L) = \alpha_{0,|L|}m_0 - (|L| - 2) \sum_{J \subset L, |J|=1} m(t, X_J) + \sum_{J \subset L, |J|=2} m(t, X_J), \tag{2.7}$$

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

Let  $X \in \mathbb{R}^d$  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  $\mathcal{D}$ :

$$A = (\alpha_{J,L})_{J \subset \mathcal{D}, L \subset \mathcal{D}}. \tag{2.8}$$

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.*

- (i) *If the set of associated coefficients  $\{\alpha_{J,L}\}$  is unique then the matrix  $A$  defined in (2.8) is idempotent, i.e.  $A^2 = A$ .*
- (ii) *If, furthermore, the coefficients depend only on the subset's cardinal, i.e.  $\alpha_{J,L} = \alpha_{j,\ell}$  with  $j = |J|$  and  $\ell = |L|$ , we obtain, for  $0 \leq j \leq \ell \leq d$ ,*

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

Now define  $s_{j,\ell} := \binom{\ell-2}{j-2} \alpha_{j,\ell}$  when  $j \geq 2$ . When  $j \geq 2$ , the above equation becomes

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

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

$$m(t, X_L) = \sum_{K \subset L, K \in \mathcal{S}} \alpha_{K,L} m(t, X_K).$$

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

$$\begin{aligned} m(t, X_L) &= \sum_{K \subset L, K \in \mathcal{S}} \sum_{J \subset K, J \in \mathcal{S}} \alpha_{J,K} \alpha_{K,L} m(t, X_J) \\ &= \sum_{J \subset L, J \in \mathcal{S}} \sum_{K \subset L, K \in \mathcal{S}} \alpha_{J,K} \alpha_{K,L} m(t, X_J) \end{aligned}$$

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

$$m(t, X_L) = \sum_{J \subset L, J \in \mathcal{S}} \left\{ \sum_{K \subset L, K \in \mathcal{S}} \alpha_{J,K} \alpha_{K,L} \right\} m(t, X_J)$$

so that, using the uniqueness 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  $\ell$  and  $k \leq \ell$ , define  $[L]^k := \{K \subset L \text{ such that } |K| = k\}$ . Assume that when  $K \subset L$  the coefficients  $\alpha_{K,L}$  depend only on the cardinals  $k = |K|$  and  $\ell = |L|$  of the considered sets, and assume that  $\alpha_{K,L} = 0$  if  $K \notin \mathcal{S}$ , i.e.  $\alpha_{k,\ell} = 0$  if  $k > d_0$ . Then

$$\begin{aligned} m(t, X_L) &= \sum_{K \subset L} \alpha_{K,L} m(t, X_K) \\ &= \sum_{k=0}^{\min(d_0, \ell)} \alpha_{k,\ell} \sum_{K \in [L]^k} m(t, X_K) \\ &= \sum_{k=0}^{\min(d_0, \ell)} \alpha_{k,\ell} \sum_{K \in [L]^k} \left\{ \sum_{j=0}^{\min(d_0, k)} \alpha_{j,k} \sum_{J \in [K]^j} m(t, X_J) \right\}. \end{aligned}$$

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

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

which entails that

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

On the other hand, we have

$$m(t, X_L) = \sum_{j=0}^{\min(d_0, \ell)} \alpha_{j,\ell} \sum_{J \in [L]^j} m(t, X_J)$$

so that, for all  $t$ , for all  $j \leq k$ ,

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

thus, the second result holds. □

**Remark 2.8.** Note that, due to the projection stability property of Proposition 2.2, any column of the matrix  $A$  can be deduced from the last column by multiplication by a matrix with values in  $\{0, 1\}$ , i.e.  $A_{\cdot,L} = P^{(L)} A_{\cdot, \mathcal{D}}$ , 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,

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

Let us finish this section with a useful property of our construction for statistical inference. Assume that we have at our disposal an independent and identically distributed sequence  $(X_1, \dots, X_n)$ , where, for each  $i$ ,  $X_i \in \mathbb{R}^d$ . The following proposition highlights the fact that if we have estimators with good properties for the distribution of subvectors  $X_K$ , then these properties are maintained for the estimator of the distribution of the whole vector  $X$ .

As in Definition 2.1, let  $\mathcal{S} \in \mathcal{P}(\mathcal{D})$  be decreasing. For any  $K$  in  $\mathcal{S}$ , assume that we can construct an estimator  $\widehat{m}_n(t, X_K)$  of  $m(t, X_K)$  using the sample  $(X_1, \dots, X_n)$  for  $t \in \mathbb{R}^d$ . From (2.2), a natural expression for an estimator  $\widehat{m}_n(t, X)$  of the distribution of the whole vector  $X$  is given by

$$\widehat{m}_n(t, X) = \sum_{K \in \mathcal{S}} \alpha_{K,\mathcal{D}} \widehat{m}_n(t, X_K). \tag{2.9}$$

**Proposition 2.4.** (Unbiasedness and consistency.) *The natural estimator given by (2.9) preserves unbiasedness and consistency.*

- (i) *If, for each  $K$ ,  $\widehat{m}_n(t, X_K)$  is unbiased then  $\widehat{m}_n(t, X)$  is also unbiased.*
- (ii) *If, for each  $K$ ,  $\widehat{m}_n(t, X_K)$  is consistent then  $\widehat{m}_n(t, X)$  is also consistent.*

*Proof.* Property (i) is obvious by the linearity in (2.2) and the linearity of expectation.

To prove property (ii), first suppose that, for each  $K$ ,  $\widehat{m}_n(t, X_K)$  converges in probability to  $m(t, X_K)$  as  $n$  goes to  $\infty$ . By Slutsky’s theorem, the vector  $(\widehat{m}_n(t, X_K))_{K \in \mathcal{S}}$  converges in probability to  $(m(t, X_K))_{K \in \mathcal{S}}$ , and by the continuous mapping theorem, any linear combination of elements of  $(\widehat{m}_n(t, X_K))_{K \in \mathcal{S}}$  converges in probability to the same linear combination of elements of  $(m(t, X_K))_{K \in \mathcal{S}}$ , which is the desired result. Notice that we are able to use Slutsky’s theorem, which is a statement about convergence in distribution, because all the limits are towards real constant values. □

**Remark 2.9.** If a central limit theorem (CLT) is available for each  $\widehat{m}_n(t, X_K)$ , then further work is needed to obtain a CLT for  $\widehat{m}_n(t, X)$  defined in (2.9). Indeed for  $K_1, K_2 \in \mathcal{S}$ ,  $K_1 \cap K_2$  is not necessarily empty, which implies that the elements in the linear combination (2.9) are dependent. A CLT for  $\widehat{m}_n(t, X)$  is thus determined by the strength of this dependence.

### 3. Examples

#### 3.1. Elliptical random vectors

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

$$\mathbb{E}[e^{it^\top(X-\mu)}] = \phi(t^\top \Sigma t)$$

for a given function  $\phi: \mathbb{R}_+ \rightarrow \mathbb{R}$  which is called the generator of the distribution, where  $\mu$  is the mean of the vector  $X$  and  $\Sigma$  is a nonnegative definite matrix. We assume here that the generator  $\phi$  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 of [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 (see [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  $\alpha$ -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 of [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, which we explore in this subsection.

Let us first remark that, for a given generator  $\phi$ , when one considers a centered multivariate elliptical distribution, the distribution is fully characterized by all components  $\sigma_{ij}$  of the matrix  $\Sigma$ , that is, 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  $\Sigma$  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  $\Sigma$  rather than its inverse  $\Sigma^{-1}$ . It thus seems easier to work with characteristic functions or entropy (which are expressed using  $\Sigma$ ) rather than densities or cumulative distribution functions (which are expressed using  $\Sigma^{-1}$ ).

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

**Definition 3.1.** ( *$\mathcal{S}$ -admissible sequence.*) Let  $\mathcal{D} \subset \mathbb{N}$ , and let  $\mathcal{S}$  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  $\Sigma$ , for all  $t$ ,

$$t^\top \Sigma t = \sum_{K \subset \mathcal{D}, K \in \mathcal{S}} \alpha_{K, \mathcal{D}} \cdot t_K^\top \Sigma_K t_K, \tag{3.1}$$

where  $\Sigma_K$  is the submatrix of  $\Sigma$  with indices in  $K$ .

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*

$$\alpha_{1,d} = - \sum_{k=2}^{d_0} s_k \frac{d-k}{k-1} \quad \text{and} \quad \alpha_{k,d} = s_k / \binom{d-2}{k-2}, \quad k = 2, \dots, d_0, \tag{3.2}$$

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

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

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

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

*Proof.* Assume that  $\alpha_d$  is  $\mathcal{S}$ -admissible, and that  $\alpha_{K,\mathcal{D}}$  depend only on  $|K|$  and  $|\mathcal{D}|$ . Let  $(i_0, j_0) \in \mathcal{D}^2, i_0 \neq j_0$ . Isolating the coefficient  $t_{i_0} \sum_{i_0, j_0} t_{j_0}$  on both sides of (3.1), we obtain

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

Denoting  $s_k = \alpha_{k,d} \binom{d-2}{k-2}$  for all  $k$ , we obtain  $\sum_{k=2}^{d_0} s_k = 1$ . Now considering the coefficient  $t_{i_0} \sum_{i_0, i_0} t_{i_0}$  on both sides of (3.1), we obtain

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

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  $1 = \sum_{k=2}^{d_0} s_k$ , (3.2) holds. The remainder of the proof follows by a direct application of this last equation.  $\square$

A direct application of such an  $\mathcal{S}$ -admissible sequence is given in the following proposition. As a consequence, the further defined *distance to admissibility* (see subsection 4.1) is always 0 for elliptical distributions.

**Proposition 3.1.** (Elliptical distributions are projective.) Consider a  $d$ -dimensional random vector  $X$  having elliptical distribution with mean  $\mu$ , matrix  $\Sigma$ , and invertible generator  $\phi$ . 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  $\alpha_d = (\alpha_{1,d}, \dots, \alpha_{d_0,d})$ ,

$$\phi(t^\top \Sigma t) = \phi\left(\sum_{k=1}^{d_0} \alpha_{k,d} \sum_{K \in [\mathcal{D}]^k} t_K^\top \Sigma_K t_K\right)$$

holds, where  $\Sigma_K$  is the submatrix of  $\Sigma$  with indices in  $K$ . In other words, setting  $m(t, X) = \phi^{-1}(\mathbb{E}[e^{it^\top(X-\mu)}])$  and  $\mathcal{S} = \{K \subset \mathcal{D}, |K| \leq 2\}$ , we have

$$X \in \mathcal{F}_{\mathcal{D}}(\mathcal{S}).$$

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,  $t^\top \Sigma t = \sum_{k=1}^{d_0} \alpha_{k,d} \sum_{K \in [\mathcal{D}]^k} t_K^\top \Sigma_K t_K$ . 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

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{3.3}$$

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 2.5). One can 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.

Consider a centered elliptical random vector with zero mean and covariance matrix  $\Sigma$ . As seen before, its distribution is thus projective, so that, for all  $t \in \mathbb{R}^d$ ,

$$t^\top \Sigma t = \sum_{K \subset \mathcal{D}, K \in \mathcal{S}} \alpha_{K,\mathcal{D}} \cdot t_K^\top \Sigma_K t_K, \tag{3.4}$$

Let us denote by  $D_K$  the  $d \times d$  diagonal matrix having 1s only at indices in  $K$ . We can write

$$t^\top \Sigma t = t^\top \left( \sum_{K \subset \mathcal{D}, K \in \mathcal{S}} \alpha_{K,\mathcal{D}} (D_K)^\top \Sigma D_K \right) t$$

so that this holds for all  $t$  if and only if

$$\Sigma = \sum_{K \subset \mathcal{D}, K \in \mathcal{S}} \alpha_{K,\mathcal{D}} \text{Ext}_{\mathcal{D}}(\Sigma_K),$$

where  $\text{Ext}_{\mathcal{D}}(\Sigma_K)$  is the extension of the matrix  $\Sigma_K$  to the dimension  $d \times d$ , i.e. the matrix having components  $(\Sigma_K)_{ij}$  for all  $i, j \in K$  and 0 when  $i \notin K$  or  $j \notin K$ . Now assume that, for all  $K \in \mathcal{S}$ , we have a given estimator  $\widehat{\Sigma}_K$  of the covariance matrix of  $X_K$ . From the previous equation, a natural estimator of the full covariance matrix  $\Sigma$  is defined as

$$\widehat{\Sigma} := \sum_{K \subset \mathcal{D}, K \in \mathcal{S}} \alpha_{K,\mathcal{D}} \text{Ext}_{\mathcal{D}}(\widehat{\Sigma}_K).$$

Now assume that classical estimators  $\widehat{\Sigma}_K$  are also available for any  $K \subset \mathcal{D}$ , then we get the following result: if, for all  $K \in \mathcal{J}$ ,  $\widehat{\Sigma}_K$  is the submatrix of  $\widehat{\Sigma}_{\mathcal{D}}$  for indices in  $K$  then

$$\widehat{\Sigma} = \widehat{\Sigma}_{\mathcal{D}}.$$

This can be checked using (3.4), when choosing  $t$  the vector having 0 everywhere except for two given components  $i$  and  $j$ .

In particular, consider for example a centered Gaussian random vector, if  $\widehat{\Sigma}_K$  are maximum likelihood estimators,  $K \subset \mathcal{D}$ , it is well known that they are directly proportional to the sample covariance matrix, so that all  $\widehat{\Sigma}_K$  are submatrices of  $\widehat{\Sigma}_{\mathcal{D}}$ , and, thus,  $\widehat{\Sigma} = \widehat{\Sigma}_{\mathcal{D}}$ . This also holds for many shrinkage estimators. On the contrary, when  $\widehat{\Sigma}_K$  are built by inverting an estimated precision matrix,  $\widehat{\Sigma}_{\mathcal{D}}$  and  $\widehat{\Sigma}$  may differ. The study of all possible estimators of  $\Sigma$  is however outside of the scope of the present paper.

When the generator  $\phi$  is unknown, another interesting perspective is to use the underlying linearity of the projective class in order to build a nonparametric estimator of  $\phi$ .

### 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 nonnegative parameters  $\theta_i, \theta_j$ , and  $\theta_{i,j}$  specific to each couple  $(i, j) \in D^2$ , and that each  $\mu_{i,j}$  satisfies

$$\mu_{i,j}((t, +\infty), (u, +\infty)) := \psi(\theta_i t + \theta_j u + \theta_{i,j} t u), \quad (t, u) \in \mathbb{R}_+^2, \quad (3.5)$$

where  $\psi$  is a given appropriate function.

Thanks to our previous results, we can construct a random variable  $X$ , 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 (3.5).

Let us first analyze the copula function associated to  $(X_i, X_j)$ . To do so, assume that  $\psi$  is a decreasing bijection from  $\mathbb{R}_+$  to  $(0, 1]$  such that  $\psi(0) = 1$  and that derivatives of  $\psi$  exist up to the order  $d$ . Denote by  $\psi^{-1}$  the inverse function of  $\psi$ . 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 obtain  $\theta_i t_i = \psi^{-1}(S_i(t))$ , 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

$$C_{S_{ij}}(u, v) = \psi \left( \psi^{-1}(u) + \psi^{-1}(v) + \frac{\theta_{ij}}{\theta_i \theta_j} \psi^{-1}(u) \psi^{-1}(v) \right), \quad (u, v) \in [0, 1]^2. \quad (3.6)$$

**Example 3.1.** Examples of survival functions in the form of (3.5), or (3.7) in the  $d$ -dimensional case, include the following particular cases.

- (i) If  $\psi(x) = \exp(-x)$  then (3.6) 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 [31, Copula (4.2.9)]).

- (ii) In the case where  $\theta_{i,j} = 0$  for every  $(i, j) \in D^2$  and if the generator  $\psi$  is  $d$ -monotone (see Definition 3.2), we obtain a survival copula which is Archimedean with generator  $\psi$ . It is clear that in this specific case, the multivariate distribution in a higher dimension will still have an Archimedean survival copula with the same generator, as it will appear further in (3.7).
- (iii) In the case where  $\theta_{i,j} = 0$  for every  $(i, j) \in D^2$  and all the coefficients  $\theta_i$  are equal, we obtain the class of Schur constant vectors, studied for example in [32]. In that case, the function  $\psi$  corresponds to the generator of the Schur constant vector, which has an Archimedean survival copula in the bivariate case, whose generator is given by  $\psi^{-1}$ .

Let us now explicitly construct a distribution on  $\mathbb{R}_+^d$ , with the given bivariate marginals. Given the form (3.5) of these marginals, we choose the following characterizing function:

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

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

Now consider the decreasing set  $\mathcal{J} = \{J \subset \mathcal{D}, |J| \leq 2\}$ , and assume that  $m$  belongs to the class  $\mathcal{F}_{\mathcal{D}}(\mathcal{J})$  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,  $J \subset K$ , depend only on cardinals  $j = |J|$  and  $k = |K|$ , so that  $\alpha_{J,K} = \alpha_{j,k}$ .

Due to Remark 2.7, if a valid multivariate distribution belongs to the class  $\mathcal{F}_{\mathcal{D}}(\mathcal{J})$  then its survival function must take the form

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

Note that  $m(t, X_{\{i\}}) = \theta_i t_i$  and  $m(t, X_{\{i,j\}}) = \theta_i t_i + \theta_j t_j + \theta_{i,j} t_i t_j$ . 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 obtain

$$\bar{F}_K(t) = \psi \left( \sum_{\{i\} \subset K} \theta_i t_i + \sum_{\{i,j\} \subset K} \theta_{i,j} t_i t_j \right). \tag{3.7}$$

Proposition 3.2 below shows that under some sufficient conditions this expression is a proper multivariate survival function. This proposition makes use of the following definition of  $d$ -monotony, as given in [30].

**Definition 3.2.** (*d-monotone function.*) A real function  $f$  is called  $d$ -monotone in  $(a, b)$ , where  $a, b \in \bar{\mathbb{R}}$  and  $d \geq 2$ , if it is differentiable there up to the order  $d - 2$  and the derivatives satisfy

$$(-1)^k f^{(k)}(x) \geq 0, \quad k = 0, 1, \dots, d - 2,$$

for any  $x \in (a, b)$  and further if  $(-1)^{d-2} f^{(d-2)}$  is nonincreasing and convex in  $(a, b)$ . For  $d = 1$ ,  $f$  is called 1-monotone in  $(a, b)$  if it is nonnegative and nonincreasing over  $(a, b)$ .



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

- (i)  $\psi$  and its derivatives goes fast enough to zero: for every  $n \leq k - 1$ ,  $\lim_{x \rightarrow +\infty} x\psi^{(n)}(x) = 0$ ;
- (ii)  $\psi$  is  $k$ -monotone;
- (iii) for all distinct  $i, j$  in  $K$ ,  $\theta_{i,j} \in [0, \theta_i\theta_j\rho_{\psi,k}]$ ,

where

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

For example, if  $|K| = k = 3$  and  $\psi(x) = \exp(-x)$ , then  $\psi$  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 (3.7) 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

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

with

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

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

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

If  $f_K$  is a nonnegative function whose integral is 1, then it will be the density of a random vector, and  $\bar{F}_K(t)$  will be a proper multivariate CDF.

*Positivity.* Let us first establish conditions under which  $f_K$  is a nonnegative function. The multivariate Faà Di Bruno’s formula gives us

$$f_K(t) = (-1)^k \sum_{\pi \in \Pi_K} \psi^{(|\pi|)}(Q(t)) \cdot \prod_{B \in \pi} \frac{\partial^{|B|} Q(t)}{\prod_{j \in B} \partial t_j}, \tag{3.8}$$

where  $\Pi_K$  is the set of all partitions of  $K$ ,  $B \in \pi$  means that  $B$  runs through all nonempty blocks of a considered partition  $\pi$ . In the following, we will write  $\partial^2 Q(t)/\partial t_B = \partial Q(t)/\partial t_i \partial t_j$  and  $\theta_B = \theta_{i,j}$ , where  $B = \{i, j\}$ .

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

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

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

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

We can write  $f_K(t) = \sum_{r \in \mathcal{N}_k} \xi(r)$ . Assume that all  $\theta_i \geq 0$  and  $\theta_{i,j} \geq 0, i, j \in K$ . Under this assumption, we 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  $r \geq 1$ ,

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

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

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 as

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

and, thus,

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

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

$$-\frac{\partial^2 Q(t)}{\partial t_B} |\psi^{(k-r)}(Q(t))| + \gamma_{k,r}^{-1} |\psi^{(k+1-r)}(Q(t))| \frac{\partial Q(t)}{\partial t_i} \frac{\partial Q(t)}{\partial t_j} \geq 0.$$

Thus yielding 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 1. If so,  $\bar{F}_K$  would be a valid absolutely continuous distribution, without a singular component. First assume that, for all integers  $n \leq k - 1$ ,

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

Recall that  $\bar{F}_K(u) = \psi(Q(u_1, \dots, u_k))$ . We now make use of the multivariate Faa Di Bruno's formula as in (3.8). Seen as a function of  $u_n$ , the derivative of  $\bar{F}_K(u)$  with respect to  $u_{n+1}, \dots, u_k$  can be written as a sum of terms  $\psi^{(i)}(au_n + b)P(u_n)$ , where  $P$  is a polynomial of degree at most 1 and  $a, b$  some real values. Thus, under assumption (i), for any  $u \in \mathbb{R}^k$  and all integers  $n \leq k - 1$ ,

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

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

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

Using the fact that  $\bar{F}_K(0, \dots, 0) = 1$ , we conclude that the derivative function  $f_K$  is nonnegative and is integrating to 1 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 (3.7). 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 (2.2), may not be a proper CDF. However, by construction, this function has exactly the prescribed multivariate margins.

It is quite a usual problem that estimators do not always satisfy all the required constraints. Among many examples include an empirical copula is not a copula [11], some Kaplan–Meier estimated distributions are defective [33], some estimated quantiles do not satisfy monotonicity [5], some nested copulas do not satisfy C-measure positivity on all hyperrectangles [17], [29], 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, we 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  $[-\infty, x] = \{(s_1, \dots, s_d) : s_1 \leq x_1, \dots, s_d \leq x_d\}$  for any vector  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ .

**Proposition 4.1.** (Maximal admissibility distance.) *Let  $\mathcal{D} \subset \mathbb{R}^d$  and denote  $\mathcal{D}_x = \mathcal{D} \cap [-\infty, x]$ . 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  $x \in \mathcal{D}$ ,  $F(x) = \int_{\mathcal{D}_x} f(s) ds$ . Assume, furthermore, that  $\int_{\mathcal{D}} f(s) ds = 1$ . Denote  $\mathcal{D}^- = \{s \in \mathcal{D}, f(s) < 0\}$  and  $\Delta = \int_{\mathcal{D}^-} |f(s)| ds$ . 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)  $d_{KS}(F, F^+) \leq \Delta / (1 + \Delta)$ ;
- (iii) for any  $K \subset \{1, \dots, d\}$ ,  $d_{KS}(F_K, F_K^+) \leq \Delta / (1 + \Delta)$ ,

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

$$F^+(x) = \int_{\mathcal{D}_x} f^+(x) dx,$$

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

*Proof.* Let  $F^+(x) = \int_{\mathcal{D}_x} f^+(x) dx$ , where  $f^+$  is defined in the proposition, and let  $\mathcal{D}^+ := \mathcal{D} \setminus \mathcal{D}^-$  for simplicity. First,  $f^+$  is nonnegative and using  $\int_{\mathcal{D}} f(x) dx = 1$ , we obtain  $\int_{\mathcal{D}^+} f^+(x) dx = 1$ , so that  $f^+$  is a proper PDF and (i) holds.

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

$$\alpha(x) := \frac{\int_{\mathcal{D}_x \cap \mathcal{D}^-} |f(x)| \, dx}{\int_{\mathcal{D}^-} |f(x)| \, dx}.$$

and  $\alpha(x) = 0$  otherwise. Note that  $\alpha(x)$  is such that  $\int_{\mathcal{D}_x \cap \mathcal{D}^-} f(x) \, dx = -\alpha(x)\Delta$ , and

$$F^+(x) = \frac{1}{1 + \Delta} \int_{\mathcal{D}_x \cap \mathcal{D}^+} f(x) \, dx = \frac{1}{1 + \Delta} \left( \int_{\mathcal{D}_x} f(x) \, dx - \int_{\mathcal{D}_x \cap \mathcal{D}^-} f(x) \, dx \right).$$

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

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

Note that  $\Delta = \int_{\mathcal{D}^-} |f(s)| \, ds = -\int_{\mathcal{D}^-} f(s) \, ds = -(1 - \int_{\mathcal{D}^+} f(s) \, ds) = \int_{\mathcal{D}^+} f(s) \, ds - 1$ , so that  $\int_{\mathcal{D}^+} f(s) \, ds = 1 + \Delta$ . As a consequence, we can define the failure ratio

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

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  $m(t, X) = \varphi \circ F(t)$ , where  $\varphi$  is a bijection from  $(0, 1]$  to  $\mathbb{R}^+$ , we obtain

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

In this case if  $\varphi, \varphi^{-1}$  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, we can show that  $F$  satisfies the assumptions of Proposition 4.1.

We note that, in practice, depending on the application, and when  $\Delta$  is small, we may use  $F$  instead of  $F^+$ , as the expression of  $F$  does not require computation of 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 has 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 us to build a valid trivariate cumulative distribution with projections close enough to each of the prescribed ones.

4.2.1. *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 to some usual data. Surely, better fits can be proposed, but are outside 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`.

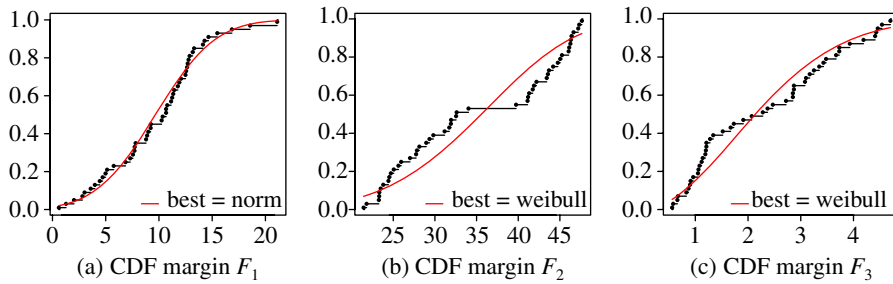


FIGURE 2: Univariate margins fits: empirical CDF (discontinuous line) and fitted CDF (continuous line) obtained by R software package fitdistrplus.

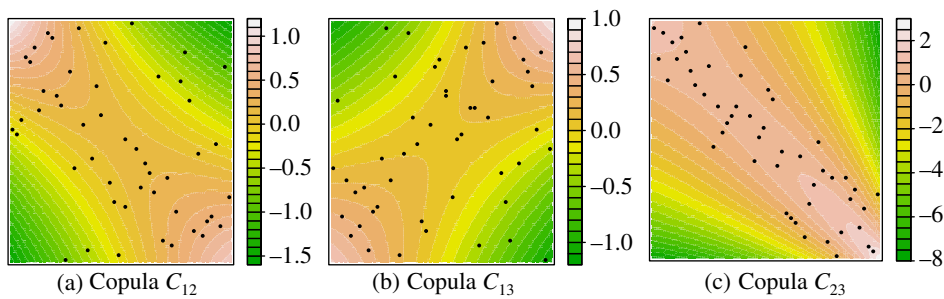


FIGURE 3: Copula fits: pseudo observations scatterplot and heatmap of the copula log-density, obtained by R software package VineCopula.

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 the data here, 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.

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.

Object	Parametric fit
$F_1$	<code>norm(9.671, 4.435 376 534 185 12)</code>
$F_2$	<code>weibull(4.483 565 871 569 89, 38.603 816 962 111 1)</code>
$F_3$	<code>weibull(1.898 639 544 002 89, 2.594 687 972 112 36)</code>
$C_{12}$	<code>Rotated BB8 90 degrees(- 6, -0.396 616 316 431 917)</code>
$C_{13}$	<code>BB8(6, 0.316 859 224 873 494)</code>
$C_{23}$	<code>Rotated BB8 270 degrees(- 6, -0.857 389 852 442 906)</code>

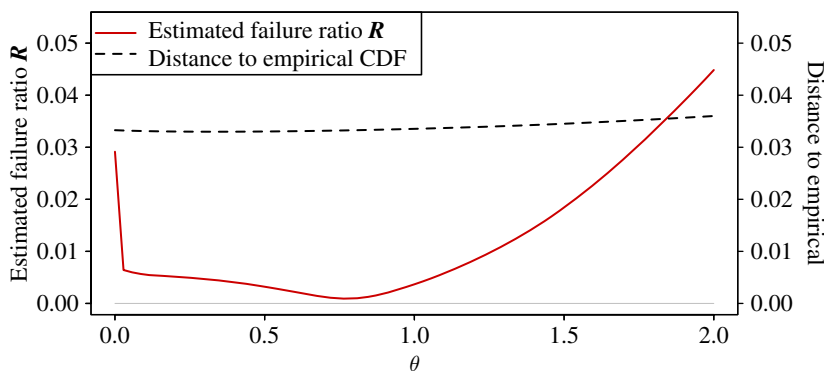


FIGURE 4: Estimated failure ratio  $R$  (solid line) and distance  $\delta$  to empirical CDF (dashed line), as a function of the parameter  $\theta$ .

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 paper.

**4.2.2. 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:

$$m(t, X_L) = \varphi_\theta[F_{X_L}(t_L)].$$

For the link function  $\varphi_\theta$ , we have chosen some parametric monotone functions such that  $\varphi_\theta(1) = 0$ , detailed hereafter. We can check that in this case  $m_0 = m(t, X_\emptyset) = \varphi_\theta(F_X(+\infty, \dots, +\infty)) = 0$  as defined in Remark 2.1. Using the proposed method and the result in Remark 2.7, when  $|L| = 3$  and with  $t_L \in \mathbb{R}^3$ , we obtain a fitted trivariate function, denoted  $F_\theta$ ,

$$F_\theta(t_L) = \varphi_\theta^{-1} \left[ -(|L| - 2) \sum_{J \subset L, |J|=1} \varphi_\theta(F_{X_J}(t_J)) + \sum_{J \subset L, |J|=2} \varphi_\theta(F_{X_J}(t_J)) \right].$$

Note 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  $\varphi_\theta$ . We have considered strictly positive and decreasing functions  $\varphi_\theta: [0, 1] \rightarrow \mathbb{R}^+$  of Table 4.1 of [31], that are also known to be Archimedean generators. We have tried the six first strict generators of [31, 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  $R = \Delta / (1 + \Delta)$  in (4.1). The parameter  $\theta$  is thus used to reduce the maximal distance between the obtained function  $F_\theta$  and a proper multivariate CDF  $F_\theta^+$ , as detailed in Proposition 4.1. We also compared the trivariate function  $F_\theta$  and the empirical trivariate CDF  $F_{\text{emp}}$  of the data, by computing the average absolute error  $\delta = (1/n) \sum_{i=1}^n |F_\theta(x_i) - F_{\text{emp}}(x_i)|$ , where  $x_1, \dots, x_n$  are the points in the dataset. In Figure 4 we show the different values of the estimated failure ratio  $R$  and the distance  $\delta$  with the empirical CDF. The ratio  $R$  was estimated by replacing the

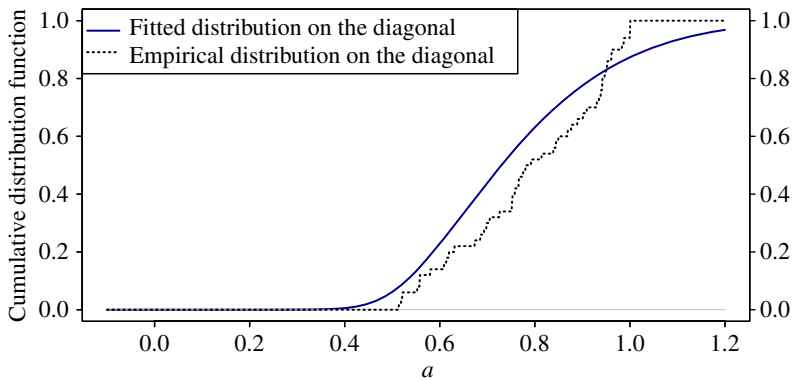


FIGURE 5: Values of the fitted function  $F_\theta$  (solid line) and the empirical CDF  $F_{\text{emp}}$  (dotted line), on several points of a diagonal line  $m + a(M - m)$  in  $\mathbb{R}^3$ , as a function of the abscissa  $a$ .

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  $\delta$  is quite stable on this data, the failure ratio  $R$  is sensitive to the choice of the parameter  $\theta$ . 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  $\varphi_\theta$  given by (4.2.9) in Table 4.1 of [31], are gathered in the following table:

Dataset (columns)	Link function $\varphi_\theta$ (parameter $\theta$ )	Estimation of $R$	Distance $\delta$
LifeCycleSavings(1,2,3)	(4.2.9) Hougaard ( $\theta = 0.768$ )	0.000 872	0.0332

For the estimation of  $R$ , using a more precise grid of 64000 points, in the extended domain  $[0.9m, 1.1M]$ , where  $m$  and  $M$  are the componentwise minimum and maximum of the points in the data, we obtained quite similar results, i.e. an estimated  $R$  equal to 0.000 984 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, p. 75].

We can check that the estimated admissibility of the fitted function  $F_\theta$  is very good so that, in practice, as we can see in Figures 5 and 6, it may be unnecessary to compute the proper CDF  $F_\theta^+$  of Proposition 4.1. Indeed, the numerical approximations involved by numerical differentiation or integration may be greater to the maximal distance  $\Delta$  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. We can instead compute the values of the fitted function for a set of points belonging to a one-dimensional or a two-dimensional 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  $m = (m_1, m_2, m_3)$  and  $M = (M_1, M_2, M_3)$  (so that the cube  $[m, M]$  contains all observations of the dataset). Let us also define the two points  $A = (M_1, m_2, M_3)$  and  $B = (m_1, M_2, M_3)$ .

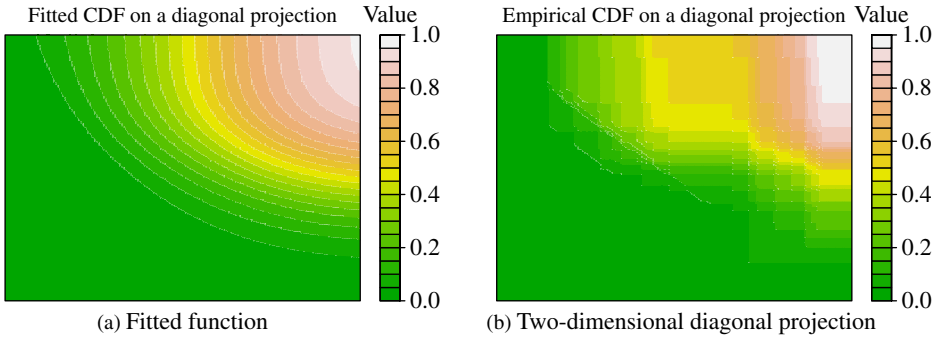


FIGURE 6: Level curves of the fitted function  $F_\theta$  (left panel) and the empirical CDF  $F_{\text{emp}}$  (right panel), on several points of the diagonal plane  $m + a(A - m) + b(B - m)$ , as a function of the abscissa  $a$  and the ordinate  $b$  (right panel).

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

$$Z := \max\left(\frac{X_1 - m_1}{M_1 - m_1}, \frac{X_2 - m_2}{M_2 - m_2}, \frac{X_3 - m_3}{M_3 - m_3}\right).$$

In Figure 5 we have drawn the estimated CDF of  $Z$  coming from our construction and 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  $F_\theta$  that we constructed, evaluated at several points of a diagonal  $m + a(M - m)$ , 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  $F_\theta$  for points belonging to a plane  $m + a(A - m) + b(B - m)$ , for different values of  $a$  and  $b$  (left panel), the empirical counterpart is also drawn (right panel).

We 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 Table 3. For some datasets (stackloss, rock), the estimated values of  $R$  are not negligible and more investigations are needed, i.e. change in the bivariate projections, in the one-parameter link function  $\varphi_\theta$ , 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 the MathAchieve dataset 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  $R$ .



TABLE 3: Final estimated failure ratio  $R$  and distance  $\delta$  to empirical for different datasets or for different considered columns. Results are sorted by estimated ratio  $R$ . Details of each bivariate fit are omitted here. The horizontal dashed lines highlight the previously detailed case study. The horizontal plain line separates the cases with estimated failure ratio greater than 0.05.

Dataset (columns)	Link function $\varphi_\theta$ (parameter $\theta$ )	Estimation of $R$	Distance $\delta$
MathAchieve(4,5,6)	(4.2.1) Clayton ( $\theta = 1.10$ )	0.000 003 92	0.006 324 558
LifeCycleSavings(1,3,5)	(4.2.5) Frank ( $\theta = 2.15$ )	0.000 136 078	0.075 148 513
LifeCycleSavings(2,3,5)	(4.2.9) Hougaard ( $\theta = 0.66$ )	0.000 243 695	0.071 234 939
LifeCycleSavings(1,2,5)	(4.2.9) Hougaard ( $\theta = 0.96$ )	0.000 365 772	0.068 639 313
LifeCycleSavings(1,2,3)	(4.2.9) Hougaard ( $\theta = 0.768$ )	0.000 872	0.0332
airquality(1,2,4)	(4.2.1) Clayton ( $\theta = 0.61$ )	0.003 274 217	0.026 502 691
airquality(1,3,4)	(4.2.5) Frank ( $\theta = -3.57$ )	0.022 312 892	0.010 637 131
mtcars(1,5,6)	(4.2.5) Frank ( $\theta = -5.55$ )	0.032 730 639	0.026 896 243
mtcars(3,4,6)	(4.2.4) Gumbel ( $\theta = 5.95$ )	0.033 680 629	0.034 147 215
swiss(1,2,3)	(4.2.9) Hougaard ( $\theta = 1.12$ )	0.056 176 129	0.013 166 825
trees(1,2,3)	(4.2.1) Clayton ( $\theta = 1.24$ )	0.058 231 572	0.021 837 067
stackloss(2,3,4)	(4.2.5) Frank ( $\theta = 3.14$ )	0.076 542 359	0.037 345 426
rock(1,2,3)	(4.2.6) Joe ( $\theta = 1.00$ )	0.099 751 509	0.028 614 845

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.

- The obtained fitted distribution has exactly the prescribed projections, by construction.
- 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.
- 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

In this paper we have considered 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 distributions to some natural survival models, as presented in Section 3.

In theory, for these 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, (2.6), and (2.7).

In practice, when dealing with a given dataset, one possibility is to use 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 this 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 Section 3.2 (but it involves many chain rule 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 theoretically build new classes of projective distributions, or try to satisfy, at least numerically, the validity of the fitted functions on some data.

### Acknowledgements

This paper was written when the authors visited the Vietnam Institute for Advanced Study in Mathematics (VIASM). The authors warmly thank the VIASM institute for 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. (2009). Pair-copula constructions of multiple dependence. *Insurance Math. Econom.* **44**, 182–198.
- [2] ACAR, E. F., GENEST, C. AND NEŠLEHOVÁ, J. (2012). Beyond simplified pair-copula constructions. *J. Multivariate Anal.* **110**, 74–90.
- [3] BAXTER, J. R. AND CHACON, R. V. (1974). Potentials of stopped distributions. *Illinois J. Math.* **18**, 649–656.
- [4] BRECHMANN, E. C. (2014). Hierarchical kendall copulas: Properties and inference. *Canad. J. Statist.* **42**, 78–108.
- [5] CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. AND GALICHON, A. (2010). Quantile and probability curves without crossing. *Econometrica* **78**, 1093–1125.
- [6] COHEN, L. (1984). Probability distributions with given multivariate marginals. *J. Math. Phys.* **25**, 2402–2403.
- [7] CUADRAS, C. M. (1992). Probability distributions with given multivariate marginals and given dependence structure. *J. Multivariate Anal.* **42**, 51–66.
- [8] DALL'AGLIO, G., KOTZ, S. AND SALINETTI, G. (eds) (1991). *Advances in Probability Distributions with Given Marginals: Beyond the Copulas* (Mathematics and Its Applications **67**), Springer, Netherlands.
- [9] DI BERNARDINO, E. AND RULLIÈRE, D. (2016). On an asymmetric extension of multivariate archimedean copulas based on quadratic form. *Depend. Model.* **4**, 328–347.
- [10] DOLATI, A. AND ÚBEDA-FLORES, M. (2005). A method for constructing multivariate distributions with given bivariate margins. *Braz. J. Prob. Statist.* **19**, 85–92.
- [11] ERDELY, A. AND GONZÁLEZ-BARRIOS, J. M. (2008). Exact distribution under independence of the diagonal section of the empirical copula. *Kybernetika (Prague)* **44**, 826–845.
- [12] FANG, K., KOTZ, S. AND NG, K. W. (1990). *Symmetric Multivariate and Related Distributions* (Monogr. Statist. Appl. Prob. **36**). Chapman and Hall, London.
- [13] FRAHM, G. (2004). Generalized elliptical distributions: theory and applications. Doctoral Thesis. Universität zu Köln.
- [14] FRAHM, G., JUNKER, M. AND SZIMAYER, A. (2003). Elliptical copulas: applicability and limitations. *Statist. Prob. Lett.* **63**, 275–286.
- [15] GENEST, C., QUESEDA MOLINA, J. J. AND RODRIGUEZ LALLENA, J. A. (1995). De l'impossibilité de construire des lois à marges multidimensionnelles données à partir de copules. *C. R. Acad. Sci. Ser. I Math.* **320**, 723–726.
- [16] GUTMANN, S., KEMPERMAN, J. H. B., REEDS, J. A. AND SHEPP, L. A. (1991). Existence of probability measures with given marginals. *Ann. Prob.* **19**, 1781–1797.
- [17] HOFERT, M. AND PHAM, D. (2013). Densities of nested archimedean copulas. *J. Multivariate Anal.* **118**, 37–52.

- [18] JOE, H. (1993). Parametric families of multivariate distributions with given margins. *J. Multivariate Anal.* **46**, 262–282.
- [19] JOE, H. (1994). Multivariate extreme-value distributions with applications to environmental data. *Canad. J. Statist.* **22**, 47–64.
- [20] JOE, H. (1996). *Families of  $m$ -variate distributions with given margins and  $m(m - 1)/2$  bivariate dependence parameters* (IMS Lecture Notes Monogr. Ser. **28**), Hayward, CA, pp. 120–141.
- [21] JOE, H. (1997). *Multivariate Models and Multivariate Dependence Concepts* (Monogr. Stat. Appl. Prob. **73**). Chapman and Hall, London.
- [22] JOE, H., LI, H. AND NIKOLOULOPOULOS, A. K. (2010). Tail dependence functions and vine copulas. *J. Multivariate Anal.* **101**, 252–270.
- [23] KANO, Y. (1994). Consistency property of elliptic probability density functions. *J. Multivariate Anal.* **51**, 139–147.
- [24] KELLERER, H. G. (1964). Verteilungsfunktionen mit gegebenen marginalverteilungen. *Z. Wahrscheinlichkeitsth.* **3**, 247–270.
- [25] LI, H., SCARSINI, M. AND SHAKED, M. (1996). Linkages: A tool for the construction of multivariate distributions with given nonoverlapping multivariate marginals. *J. Multivariate Anal.* **56**, 20–41.
- [26] LI, H., SCARSINI, M. AND SHAKED, M. (1999). Dynamic linkages for multivariate distributions with given nonoverlapping multivariate marginals. *J. Multivariate Anal.* **68**, 54–77.
- [27] MARCO, J. M. AND RUIZ-RIVAS, C. (1992). On the construction of multivariate distributions with given nonoverlapping multivariate marginals. *Statist. Prob. Lett.* **15**, 259–265.
- [28] MARONNA, R., MARTIN, D. AND YOHAI, V. (2006). *Robust Statistics: Theory and Methods*. John Wiley, Chichester.
- [29] MCNEIL, A. J. (2008). Sampling nested archimedean copulas. *J. Statist. Comput. Simul.* **78**, 567–581.
- [30] MCNEIL, A. J. AND NEŠLEHOVÁ, J. (2009). Multivariate Archimedean copulas,  $d$ -monotone functions and  $\ell_1$ -norm symmetric distributions. *Ann. Statist.* **37**, 3059–3097.
- [31] NELSEN, R. B. (1999). *An Introduction to Copulas* (Lecture Notes Statist. **139**). Springer, New York.
- [32] NELSEN, R. B. (2005). Some properties of schur-constant survival models and their copulas. *Braz. J. Prob. Statist.* **19**, 179–190.
- [33] PORTNOY, S. (2003). Censored regression quantiles. *J. Amer. Statist. Assoc.* **98**, 1001–1012.
- [34] RÜSCHENDORF, L. (1985). Construction of multivariate distributions with given marginals. *Ann. Inst. Statist. Math.* **37**, 225–233.
- [35] SÁNCHEZ ALGARRA, P. (1986). Construcción de distribuciones con marginales multivariantes dadas. *Qüestió* **10**, 133–141.
- [36] SKLAR, M. (1959). Fonctions de répartition à  $n$  dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8**, 229–231.
- [37] SPANIER, E. H. (1994). *Algebraic Topology*. Springer, New York.