

Article

Goodness-of-Fit Tests For Elliptical and Independent Copulas through Projection Pursuit

Jacques Touboul

Université Pierre et Marie Curie, Laboratoire de Statistique Théorique et Appliquée,
175 rue du Chevaleret, 75013 Paris, France; E-Mail: jack_touboul@hotmail.com;
Tel.: +33-1-4427-8562; Fax: +33-1-4427-4670

Received: 15 March 2011; in revised form: 5 April 2011 / Accepted: 8 April 2011 /

Published: 26 April 2011

Abstract: Two goodness-of-fit tests for copulas are being investigated. The first one deals with the case of elliptical copulas and the second one deals with independent copulas. These tests result from the expansion of the projection pursuit methodology that we will introduce in the present article. This method enables us to determine on which axis system these copulas lie as well as the exact value of these very copulas in the basis formed by the axes previously determined irrespective of their value in their canonical basis. Simulations are also presented as well as an application to real datasets.

Keywords: copulas; goodness-of-fit; projection pursuit; elliptical distributions

Classification: MSC 62H05 62H15 62H40 62G15

1. Introduction

The need to describe the dependency between two or more random variables triggered the concept of copulas. We consider a joint cumulative distribution function (cdf) F on $\mathbb{R}^d$ and its cdf margins $F_1, F_2, \dots, F_d$. A copula C is a function such that $F = C(F_1, F_2, \dots, F_d)$. Sklar [1] is the first to lay the foundations of this new theory. Several parametric families of copulas have been defined, namely elliptical, archimedean, periodic copulas *etc.*, see Joe [2] and Nelsen [3] as well as Appendix A for an overview of these families. Finding criteria to determine the best copula for a given problem can only be achieved through a goodness-of-fit (GOF) approach. So far several GOF copula approaches have been proposed in the literature, e.g., Carriere [4], Genest and Remillard [5], Fermanian [6], Genest Quessy and

Rémillard [7], Michiels and De Schepper [8], Genest Favre Béliveau and Jacques [9], Mesfioui Quessy and Toupin [10], Genest Rémillard and Beaudoin [11], Berg [12], Bücher and Dette [13], among others. However, the field is still at an embryonic stage which explains the current shortage in recommendations. In univariate distributions, the GOF assessment can be performed using for instance the well-known Kolmogorov test. In the multivariate field, there are fewer alternatives. A simple way to build GOF approaches for multivariate random variables is to consider multi-dimensional chi-square approaches, as in for example Broniatowski [14]. However, these approaches present feasibility issues for high dimensional problems due to the curse of dimensionality. In order to solve this, we recall some facts from the theory of projection pursuit.

The objective of projection pursuit is to generate one or several projections providing as much information as possible about the structure of the dataset regardless of its size. Once a structure has been isolated, the corresponding data are transformed through a Gaussianization. Through a recursive approach, this process is iterated to find another structure in the remaining data, until no further structure can be evidenced in the data left at the end. Friedman [15] and Huber [16] count among the first authors who introduced this type of approaches for evidencing structures. They each describe, with many examples, how to evidence such a structure and consequently how to estimate the density of such data through two different methodologies each. Their work is based on maximizing Kullback-Leibler divergence. In the present article, we will introduce a new projection pursuit methodology based on the minimisation of any ϕ -divergence greater than the L^1 -distance (ϕ -PP). We will show that this algorithm presents the extra advantage of being robust and fast from a numerical standpoint. Its key rationale lies in the fact that it allows not only to carry out GOF tests for elliptical and independent copulas but also to determine the axis system upon which these very copulas are based. The exact expression of these copulas in the basis constituted by these axes can therefore be derived.

This paper is organised as follows: Section 2 contains preliminary definitions and properties. In Section 3, we present in details the ϕ -projection pursuit algorithm. In Section 4, we present our first results. In Section 5, we introduce our tests. In Section 6, we provide three simulations pertaining to the two major situations described herein and we will study a real case.

2. Basic theory

2.1. An Introduction to Copulas

In this section, we recall the concept of copula. We will also define the family of elliptical copulas through a brief reminder of elliptical distributions—see Appendix A for an overview of other families.

Sklar's Theorem

First, let us define a copula in $\mathbb{R}^d$.

Definition 2.1. A d -dimensional copula is a joint cumulative distribution function C defined on $[0, 1]^d$, with uniform margins.

The following theorem explains in what extent a copula does describe the dependency between two or more random variables.

Theorem 2.1 (Sklar [1]). *Let F be a joint multivariate distribution with margins $F_1, \dots, F_d$, then, there exists a copula C such that*

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad (2.1)$$

If marginal cumulative distributions are continuous, then the copula is unique. Otherwise, the copula is unique on the range of values of the marginal cumulative distributions.

Remark 2.1. *First, for any copula C and any u_i in $[0, 1]$, $1 \leq i \leq d$, we have*

$$W(u_1, \dots, u_d) = \max \left\{ 1 - d + \sum_{i=1}^d u_i, 0 \right\} \leq C(u_1, \dots, u_d) \leq \min_{j \in \{1, \dots, d\}} u_j = M(u_1, \dots, u_d)$$

where W and M are called the Frechet-Hoeffding copula boundaries and are also copulas.

We set the independent copula Π as $\Pi(u_1, \dots, u_d) = \prod_{i=1}^d u_i$, for any u_i in $[0, 1]$, $1 \leq i \leq d$.

Moreover, we define the density of a copula as the density associated with the cdf C , which we will name as c :

Definition 2.2. *Whenever there exists, the density of C is defined by $c(u_1, \dots, u_d) = \frac{\partial^d}{\partial u_1 \dots \partial u_d} C(u_1, \dots, u_d)$, for any u_i in $[0, 1]$, $1 \leq i \leq d$.*

Finally, let us present several examples of copulas (see also Appendix A to find an overview).

Example 2.1.

The Gaussian copula C_ρ (in $\mathbb{R}^2$):

Defining Ψ_ρ as the standard bivariate normal cumulative distribution function with ρ correlation, the Gaussian copula function is

$$C_\rho(u, v) = \Psi_\rho(\Psi^{-1}(u), \Psi^{-1}(v))$$

where $u, v \in [0, 1]$ and where Ψ is the standard normal cumulative distribution function.

The Student copula C_ρ (in $\mathbb{R}^2$):

Defining $T_{\rho,k}$ as the standard bivariate student cumulative distribution function with ρ as the correlation coefficient and with k as the degree of freedom of the distribution, the Student copula function is

$$C_\rho(u, v) = T_{\rho,k}(T_k^{-1}(u), T_k^{-1}(v))$$

where $u, v \in [0, 1]$ and where T_k is the standard Student cumulative distribution function.

The Elliptical copula :

Similarly as above, elliptical copulas are the copulas of elliptical distributions (an overview is provided in Appendix A).

2.2. Brief Introduction to the ϕ -Projection Pursuit Methodology (ϕ -PP)

Let us first introduce the concept of ϕ -divergence.

The Concept of ϕ -Divergence

Let φ be a strictly convex function defined by $\varphi : \overline{\mathbb{R}^+} \rightarrow \overline{\mathbb{R}^+}$, and such that $\varphi(1) = 0$. We define a ϕ -divergence of P from Q , where P and Q are two probability distributions over a space Ω such that Q is absolutely continuous with respect to P -by

$$D_\phi(Q, P) = \int \varphi \left(\frac{dQ}{dP} \right) dP$$

or $D_\phi(q, p) = \int \varphi \left(\frac{q(x)}{p(x)} \right) p(x) dx$, if P and Q present p and q as density respectively.

Throughout this article, we will also assume that $\varphi(0) < \infty$, that φ' is continuous and that this divergence is greater than the L^1 distance—see also Appendix B page 109.

Functioning of the Algorithm

Let f be a density on $\mathbb{R}^d$. We consider an instrumental density g with the same mean and variance as f . We start with performing the $D_\phi(g, f) = 0$ test; should this test turn out to be positive, then $f = g$ and the algorithm stops, otherwise, the first step of our algorithm consists in defining a vector a_1 and a density $g^{(1)}$ by

$$a_1 = \arg \inf_{a \in \mathbb{R}_*^d} D_\phi \left(g \frac{f_a}{g_a}, f \right) \text{ and } g^{(1)} = g \frac{f_{a_1}}{g_{a_1}} \quad (2.2)$$

where $\mathbb{R}_*^d$ is the set of non-null vectors of $\mathbb{R}^d$ and f_a (resp. g_a) stands for the density of $a^\top X$ (resp. $a^\top Y$) when f (resp. g) is the density of X (resp. Y).

In our second step, we replace g with $g^{(1)}$ and we repeat the first step, and so on. By iterating this process, we end up obtaining a sequence $(a_1, a_2, \dots)$ of vectors in $\mathbb{R}_*^d$ and a sequence of densities $g^{(i)}$.

Remark 2.2. First, to obtain an approximation of f , we stop our algorithm when the divergence equals zero, i.e., we stop when $D_\phi(g^{(j)}, f) = 0$ since it implies $g^{(j)} = f$ with $j \leq d$, or when our algorithm reaches the d^{th} iteration, i.e., we approximate f with $g^{(d)}$.

Second, we get $D_\phi(g^{(0)}, f) \geq D_\phi(g^{(1)}, f) \geq \dots \geq 0$ with $g^{(0)} = g$.

Finally, the specific form of the relationship (2.2) implies that we deal with M-estimation. We can therefore state that our method is robust—see Sections 6, Yohai [19], Toma [20] as well as Huber [21].

The main steps of the present algorithm have been summarized in Table 1.

Table 1. Proposal.

0.	We define g , a density with same mean and variance as f and we set $g^{(0)} = g$.
$i - 1$.	We perform the goodness-of-fit test $D_\phi(g^{(i-1)}, f) = 0$: • Should this test be passed, we derive f from $f = g \prod_{i=1}^j \frac{f_{a_i}}{g_{a_i}^{(i-1)}}$ And the algorithm stops. • Should this test not be verified, and should we look to approximate f, when we get to the d^{th} iteration of the algorithm, we derive f from $f = g \prod_{i=1}^d \frac{f_{a_i}}{g_{a_i}^{(i-1)}}$ Otherwise, let us define a vector a_i and a density $g^{(i)}$ by $a_i = \arg \inf_{a \in \mathbb{R}_*^d} D_\phi \left(g^{(i-1)} \frac{f_a}{g_a^{(i-1)}}, f \right), \text{ and } g^{(i)} = g^{(i-1)} \frac{f_{a_i}}{g_{a_i}^{(i-1)}}$
i .	Then we replace $g^{(i-1)}$ with $g^{(i)}$ and go back to $i - 1$.

At present, let us study the following example:

Example 2.2. Let f be a density defined on $\mathbb{R}^3$ by $f(x_1, x_2, x_3) = n(x_1, x_2)h(x_3)$, with n being a bi-dimensional Gaussian density, and h being a non-Gaussian density. Let us also consider g , a Gaussian density with the same mean and variance as f .

Since $g(x_1, x_2/x_3) = n(x_1, x_2)$, we have $D_\phi(g_{g_3}^{f_3}, f) = D_\phi(n \cdot f_3, f) = D_\phi(f, f) = 0$ as $f_3 = h$, i.e., the function $a \mapsto D_\phi(g_{g_a}^{f_a}, f)$ reaches zero for $e_3 = (0, 0, 1)'$, where f_3 and g_3 are the third marginal densities of f and g respectively. We therefore obtain $g(x_1, x_2/x_3) = f(x_1, x_2/x_3)$.

To recapitulate our method, if $D_\phi(g, f) = 0$, we derive f from the relationship $f = g$; whenever a sequence $(a_i)_{i=1, \dots, j}$, $j < d$, of vectors in $\mathbb{R}_*^d$ defining $g^{(j)}$ and such that $D_\phi(g^{(j)}, f) = 0$ exists, then $f(\cdot/a_i^\top x, 1 \leq i \leq j) = g(\cdot/a_i^\top x, 1 \leq i \leq j)$, i.e., f coincides with g on the complement of the vector subspace generated by the family $\{a_i\}_{i=1, \dots, j}$ —see also Section 3 for a more detailed explanation.

In the remaining of our study of the algorithm, after having clarified the choice of g , we will consider the statistical solution to the representation problem, assuming that f is unknown and that $X_1, X_2, \dots, X_m$ are i.i.d. with density f . We will provide asymptotic results pertaining to the family of optimizing vectors $a_{k,m}$ —which we will define more precisely below—as m goes to infinity. Our results also prove that the empirical representation scheme converges towards the theoretical one.

3. The Algorithm

3.1. The Model

Let f be a density on $\mathbb{R}^d$. We assume there exists d non-null linearly independent vectors a_j , with $1 \leq j \leq d$, of $\mathbb{R}^d$, such that

$$f(x) = n(a_{j+1}^\top x, \dots, a_d^\top x)h(a_1^\top x, \dots, a_j^\top x) \quad (3.1)$$

with $j < d$, n being an elliptical density on $\mathbb{R}^{d-j}$ and with h being a density on $\mathbb{R}^j$, which does not belong to the same family as n . Let $X = (X_1, \dots, X_d)$ be a vector with f as density.

We define g as an elliptical distribution with the same mean and variance as f .

For simplicity, let us assume that the family $\{a_j\}_{1 \leq j \leq d}$ is the canonical basis of $\mathbb{R}^d$:

The very definition of f implies that $(X_{j+1}, \dots, X_d)$ is independent from $(X_1, \dots, X_j)$. Hence, the density of $(X_{j+1}, \dots, X_d)$ given $(X_1, \dots, X_j)$ is n .

Let us assume that $D_\phi(g^{(j)}, f) = 0$, for some $j \leq d$. We then get $\frac{f(x)}{f_{a_1} f_{a_2} \dots f_{a_j}} = \frac{g(x)}{g_{a_1}^{(1-1)} g_{a_2}^{(2-1)} \dots g_{a_j}^{(j-1)}}$, since, by induction, we have $g^{(j)}(x) = g(x) \frac{f_{a_1}}{g_{a_1}^{(1-1)}} \frac{f_{a_2}}{g_{a_2}^{(2-1)}} \dots \frac{f_{a_j}}{g_{a_j}^{(j-1)}}$.

Consequently, lemma C.1 and the fact that the conditional densities with elliptical distributions are also elliptical, as well as the above relationship, lead us to infer that $n(a_{j+1}^\top x, \dots, a_d^\top x) = f(\cdot/a_i^\top x, 1 \leq i \leq j) = g(\cdot/a_i^\top x, 1 \leq i \leq j)$. In other words, f coincides with g on the complement of the vector subspace generated by the family $\{a_i\}_{i=1, \dots, j}$.

Now, if the family $\{a_j\}_{1 \leq j \leq d}$ is no longer the canonical basis of $\mathbb{R}^d$, then this family is again a basis of $\mathbb{R}^d$. Hence, lemma C.2 implies that

$$g(\cdot/a_1^\top x, \dots, a_j^\top x) = n(a_{j+1}^\top x, \dots, a_d^\top x) = f(\cdot/a_1^\top x, \dots, a_j^\top x) \quad (3.2)$$

which is equivalent to $D_\phi(g^{(j)}, f) = 0$, since by induction $g^{(j)} = g \frac{f_{a_1}}{g_{a_1}^{(1-1)}} \frac{f_{a_2}}{g_{a_2}^{(2-1)}} \dots \frac{f_{a_j}}{g_{a_j}^{(j-1)}}$.

The end of our algorithm implies that f coincides with g on the complement of the vector subspace generated by the family $\{a_i\}_{i=1, \dots, j}$. Therefore, the nullity of the ϕ -divergence provides us with information on the density structure.

In summary, the following proposition clarifies our choice of g which depends on the family of distribution one wants to find in f :

Proposition 3.1. *With the above notations, $D_\phi(g^{(j)}, f) = 0$ is equivalent to*

$$g(\cdot/a_1^\top x, \dots, a_j^\top x) = f(\cdot/a_1^\top x, \dots, a_j^\top x)$$

More generally, the above proposition defines the co-support of f as the vector space generated by the vectors $a_1, \dots, a_j$.

Definition 3.1. *Let f be a density on $\mathbb{R}^d$. We define the co-vectors of f as the sequence of vectors $a_1, \dots, a_j$ which solves the problem $D_\phi(g^{(j)}, f) = 0$ where g is an elliptical distribution with the same mean and variance as f . We define the co-support of f as the vector space generated by the vectors $a_1, \dots, a_j$.*

Remark 3.1. *Any (a_i) family defining f as in (3.1) is an orthogonal basis of $\mathbb{R}^d$ —see lemma C.3*

3.2. Stochastic Outline of Our Algorithm

Let $X_1, X_2, \dots, X_m$ (resp. $Y_1, Y_2, \dots, Y_m$) be a sequence of m independent random vectors with the same density f (resp. g). As customary in nonparametric ϕ -divergence optimizations, all estimates of f and f_a , as well as all uses of Monte Carlo methods are being performed using subsamples $X_1, X_2, \dots, X_n$

and $Y_1, Y_2, \dots, Y_n$ —extracted respectively from $X_1, X_2, \dots, X_m$ and $Y_1, Y_2, \dots, Y_m$ —since the estimates are bounded below by some positive deterministic sequence θ_m —see Appendix D.

Let $\mathbb{P}_n$ be the empirical measure based on the subsample $X_1, X_2, \dots, X_n$. Let f_n (resp. $f_{a,n}$ for any a in $\mathbb{R}_*^d$) be the kernel estimate of f (resp. f_a), which is built from $X_1, X_2, \dots, X_n$ (resp. $a^\top X_1, a^\top X_2, \dots, a^\top X_n$).

As defined in Section 2.2, we consider the following sequences $(a_k)_{k \geq 1}$ and $(g^{(k)})_{k \geq 1}$ such that

$$\begin{aligned} a_k \text{ is a non null vector of } \mathbb{R}^d \text{ defined by } a_k &= \arg \min_{a \in \mathbb{R}_*^d} D_\phi \left(g^{(k-1)} \frac{f_a}{g_a^{(k-1)}}, f \right) \\ g^{(k)} \text{ is the density defined by } g^{(k)} &= g^{(k-1)} \frac{f_{a_k}}{g_{a_k}^{(k-1)}} \text{ with } g^{(0)} = g \end{aligned} \quad (3.3)$$

The stochastic setting up of the algorithm uses f_n and $g_n^{(0)} = g$ instead of f and $g^{(0)} = g$ —since g is known. Thus, at the first step, we build the vector $\tilde{a}_1$ which minimizes the ϕ -divergence between f_n and $g_n^{\frac{f_{a,n}}{g_a}}$ and which estimates a_1 . First, since proposition D.1 and lemma C.4 show how the infimum of the criteria (or index)

$$\check{D}_\phi \left(g \frac{f_{a,n}}{g_a}, f_n \right) = \frac{1}{n} \sum_{i=1}^n \varphi \left(\frac{g(X_i) f_{a,n}(a^\top X_i)}{f_n(X_i) g_a(a^\top X_i)} \right)$$

is reached, we are then able to minimize the ϕ -divergence between f_n and $g_n^{\frac{f_{a,n}}{g_a}}$. Second, defining $\tilde{a}_1$ as the argument of this minimization, proposition 4.3 infers that this vector tends to a_1 . Finally, we define the density $\check{g}_n^{(1)}$ as $\check{g}_n^{(1)} = g_n^{\frac{f_{\tilde{a}_1,n}}{g_{\tilde{a}_1}^{(1)}}}$ which estimates $g^{(1)}$ through theorem 4.1.

Now, from the second step and as defined in Section 2.2, we derive the fact that the density $g^{(2-1)}$ is unknown. Consequently, once again, the samples have to be truncated.

All estimates of f and f_a (resp. $g^{(1)}$ and $g_a^{(1)}$) are being performed using a subsample $X_1, X_2, \dots, X_n$ (resp. $Y_1^{(1)}, Y_2^{(1)}, \dots, Y_n^{(1)}$) extracted from $X_1, X_2, \dots, X_m$ (resp. $Y_1^{(1)}, Y_2^{(1)}, \dots, Y_m^{(1)}$), which is a sequence of m independent random vectors with same density $g^{(1)}$ such that the estimates are bounded below by some positive deterministic sequence θ_m —see Appendix D.

Let $\mathbb{P}_n$ be the empirical measure of the subsample $X_1, X_2, \dots, X_n$. Let f_n (resp. $g_n^{(1)}, f_{a,n}, g_{a,n}^{(1)}$ for any a in $\mathbb{R}_*^d$) be the kernel estimate of f (resp. $g^{(1)}$ and f_a as well as $g_a^{(1)}$) which is built from $X_1, X_2, \dots, X_n$ (resp. $Y_1^{(1)}, Y_2^{(1)}, \dots, Y_n^{(1)}$ and $a^\top X_1, a^\top X_2, \dots, a^\top X_n$ as well as $a^\top Y_1^{(1)}, a^\top Y_2^{(1)}, \dots, a^\top Y_n^{(1)}$).

The stochastic setting up of the algorithm uses f_n and $g_n^{(1)}$ instead of f and $g^{(1)}$. Thus, we build the vector $\tilde{a}_2$, which minimizes the ϕ -divergence between f_n and $g_n^{(1)} \frac{f_{a,n}}{g_{a,n}^{(1)}}$, since $g^{(1)}$ and $g_a^{(1)}$ are unknown—and which estimates a_2 . First, since proposition D.1 and lemma C.4 show how the infimum of the criteria (or index)

$$\check{D}_\phi \left(g_n^{(1)} \frac{f_{a,n}}{g_{a,n}^{(1)}}, f_n \right) = \frac{1}{n} \sum_{i=1}^n \varphi \left(\frac{g_n^{(1)}(X_i) f_{a,n}(a^\top X_i)}{f_n(X_i) g_{a,n}^{(1)}(a^\top X_i)} \right)$$

is reached, we are then able to minimize the ϕ -divergence between f_n and $g_n^{(1)} \frac{f_{a,n}}{g_{a,n}^{(1)}}$. Second, defining $\tilde{a}_2$ as the argument of this minimization, proposition 4.3 infers that this vector tends to a_2 . Finally, we define the density $\check{g}_n^{(2)}$ as $\check{g}_n^{(2)} = g_n^{(1)} \frac{f_{\tilde{a}_2,n}}{g_{\tilde{a}_2}^{(1)}}$ which estimates $g^{(2)}$ through theorem 4.1.

And so on, we end up obtaining a sequence $(\tilde{a}_1, \tilde{a}_2, \dots)$ of vectors in $\mathbb{R}_*^d$ estimating the co-vectors of f and a sequence of densities $(\check{g}_n^{(k)})_k$ such that $\check{g}_n^{(k)}$ estimates $g^{(k)}$ through theorem 4.1.

Let us now summarize the main steps of the stochastic implementation of our algorithm (the dual representation of the estimators will be further detailed in Table 2 below).

Table 2. Stochastic outline of the algorithm.

0.	We define g , a density with same mean and variance as f and we set $g^{(0)} = g$.
$i - 1$.	Given $\check{g}_n^{(i-1)}$, find $\check{a}_i$ such that the index $\check{D}_\phi \left(\check{g}_n^{(i-1)} \frac{f_{a,n}}{\check{g}_{a,n}^{(i-1)}}, f_n \right) = \frac{1}{n} \sum_{i=1}^n \varphi \left(\frac{\check{g}_n^{(i-1)}(X_i) f_{a,n}(a^\top X_i)}{f_n(X_i) \check{g}_{a,n}^{(i-1)}(a^\top X_i)} \right)$ is minimized, where $f_{a,n}$ is a marginal density estimate based on $a^\top X_1, a^\top X_2, \dots, a^\top X_n$, and where $\check{g}_{a,n}^{(i-1)}$ is a density estimate based on the projection to a of a Monte Carlo random sample from $\check{g}_n^{(i)}$. And we set $\check{g}_n^{(i)} = \check{g}_n^{(i-1)} \frac{f_{\check{a}_1,n}}{\check{g}_{\check{a}_1,n}^{(i-1)}}$ i. Then we replace $\check{g}_n^{(i-1)}$ with $\check{g}_n^{(i)}$ and go back to $i - 1$ until the criteria reaches the stopping rule of this procedure (see below).

4. Results

4.1. Hypotheses on f

In this paragraph, we define the set of hypotheses on f which could possibly be of use in our work. Discussion on several of these hypotheses can be found in Appendix E.

In the remaining of this section, for legibility reasons, we replace g with $g^{(k-1)}$. Let

$$\Theta = \mathbb{R}^d, \quad \Theta^{D_\phi} = \left\{ b \in \Theta \mid \int \varphi^* \left(\varphi' \left(\frac{g(x) f_b(b^\top x)}{f(x) g_b(b^\top x)} \right) \right) d\mathbf{P} < \infty \right\}$$

$$M(b, a, x) = \int \varphi' \left(\frac{g(x) f_b(b^\top x)}{f(x) g_b(b^\top x)} \right) g(x) \frac{f_a(a^\top x)}{g_a(a^\top x)} dx - \varphi^* \left(\varphi' \left(\frac{g(x) f_b(b^\top x)}{f(x) g_b(b^\top x)} \right) \right)$$

$$\mathbb{P}_n M(b, a) = \int M(b, a, x) d\mathbb{P}_n, \quad \mathbf{P} M(b, a) = \int M(b, a, x) d\mathbf{P}$$

where $\mathbf{P}$ is the probability measure presenting f as density.

Similarly as in chapter V of Van der Vaart [22], let us define :

(A1) : For all $\varepsilon > 0$, there is $\eta > 0$, such that for all $c \in \Theta^{D_\phi}$ verifying $\|c - a_k\| \geq \varepsilon$,

we have $\mathbf{P} M(c, a) - \eta > \mathbf{P} M(a_k, a)$, with $a \in \Theta$.

(A2) : $\exists Z < 0, n_0 > 0$ such that $(n \geq n_0 \Rightarrow \sup_{a \in \Theta} \sup_{c \in \{\Theta^{D_\phi}\}^c} \mathbb{P}_n M(c, a) < Z)$

(A3) : There exists V , a neighbourhood of a_k , and H , a positive function, such that, for all $c \in V$,

we have $|M(c, a_k, x)| \leq H(x)(\mathbf{P} - a.s.)$ with $\mathbf{P} H < \infty$,

(A4) : There exists V , a neighbourhood of a_k , such that for all ε , there exists a η such that for all $c \in V$

and $a \in \Theta$, verifying $\|a - a_k\| \geq \varepsilon$, we have $\mathbf{P} M(c, a_k) < \mathbf{P} M(c, a) - \eta$.

Putting $I_{a_k} = \frac{\partial^2}{\partial a^2} D_\phi(g \frac{f_{a_k}}{g_{a_k}}, f)$, let us consider now four new hypotheses:

(A5) : $\mathbf{P} \left\| \frac{\partial}{\partial b} M(a_k, a_k) \right\|^2$ and $\mathbf{P} \left\| \frac{\partial}{\partial a} M(a_k, a_k) \right\|^2$ are finite and the expressions $\mathbf{P} \frac{\partial^2}{\partial b_i \partial b_j} M(a_k, a_k)$ and I_{a_k} exist and are invertible.

(A6) : There exists k such that $\mathbf{P} M(a_k, a_k) = 0$.

(A7) : $(\text{Var}_{\mathbf{P}}(M(a_k, a_k)))^{1/2}$ exists and is invertible.

(A0) : f and g are assumed to be positive and bounded and such that $K(g, f) \geq \int |f(x) - g(x)| dx$ where K is the Kullback-Leibler divergence.

Estimation of the First Co-Vector of f

Let $\mathcal{R}$ be the class of all positive functions r defined on $\mathbb{R}$ and such that $g(x)r(a^\top x)$ is a density on $\mathbb{R}^d$ for all a belonging to $\mathbb{R}_*^d$. The following proposition shows that there exists a vector a such that $\frac{f_a}{g_a}$ minimizes $D_\phi(gr, f)$ in r :

Proposition 4.1. *There exists a vector a belonging to $\mathbb{R}_*^d$ such that*

$$\arg \min_{r \in \mathcal{R}} D_\phi(gr, f) = \frac{f_a}{g_a} \text{ and } r(a^\top x) = \frac{f_a(a^\top x)}{g_a(a^\top x)}$$

Following Broniatowski [33], let us introduce the estimate of $D_\phi(g \frac{f_{a,n}}{g_a}, f_n)$, through $\check{D}_\phi(g \frac{f_{a,n}}{g_a}, f_n) = \sup_{b \in \Theta} \int M(b, a, x) d\mathbb{P}_n(x)$.

Proposition 4.2. *Let $\check{a}$ be such that $\check{a} := \arg \inf_{a \in \mathbb{R}_*^d} \check{D}_\phi(g \frac{f_{a,n}}{g_a}, f_n)$.*

Then, $\check{a}$ is a strongly convergent estimate of a , as defined in proposition 4.1.

Let us also introduce the following sequences $(\check{a}_k)_{k \geq 1}$ and $(\check{g}_n^{(k)})_{k \geq 1}$, for any given n —see Section 3.2—such that

$\check{a}_k$ is an estimate of a_k as defined in proposition 4.2 with $\check{g}_n^{(k-1)}$ instead of g ,

$\check{g}_n^{(k)}$ is defined by $\check{g}_n^{(0)} = g$, $\check{g}_n^{(k)}(x) = \check{g}_n^{(k-1)}(x) \frac{f_{\check{a}_k, n}(\check{a}_k^\top x)}{[\check{g}_n^{(k-1)}]_{\check{a}_k, n}(\check{a}_k^\top x)}$, i.e., $\check{g}_n^{(k)}(x) = g(x) \prod_{j=1}^k \frac{f_{\check{a}_j, n}(\check{a}_j^\top x)}{[\check{g}_n^{(j-1)}]_{\check{a}_j, n}(\check{a}_j^\top x)}$.

We also note that $\check{g}_n^{(k)}$ is a density.

Convergence Study at the k^{th} Step of the Algorithm:

In this paragraph, we show that the sequence $(\check{a}_k)_n$ converges towards a_k and that the sequence $(\check{g}_n^{(k)})_n$ converges towards $g^{(k)}$.

Let $\check{c}_n(a) = \arg \sup_{c \in \Theta} \mathbb{P}_n M(c, a)$, with $a \in \Theta$, and $\check{\gamma}_n = \arg \inf_{a \in \Theta} \sup_{c \in \Theta} \mathbb{P}_n M(c, a)$. We state

Proposition 4.3. *Both $\sup_{a \in \Theta} \|\check{c}_n(a) - a_k\|$ and $\check{\gamma}_n$ converge toward a_k a.s.*

Finally, the following theorem shows that $\check{g}_n^{(k)}$ converges almost everywhere towards $g^{(k)}$:

Theorem 4.1. *It holds $\check{g}_n^{(k)} \rightarrow_n g^{(k)}$ a.s.*

Testing of the Criteria

In this paragraph, through a test of our criteria, namely $a \mapsto D_\phi(\check{g}_n^{(k)} \frac{f_{a,n}}{[\check{g}_n^{(k)}]_{a,n}}, f_n)$, we build a stopping rule for this procedure. First, the next theorem enables us to derive the law of our criteria:

Theorem 4.2. *For a fixed k , we have $\sqrt{n}(\text{Var}_{\mathbf{P}}(M(\check{c}_n(\check{\gamma}_n), \check{\gamma}_n)))^{-1/2}(\mathbb{P}_n M(\check{c}_n(\check{\gamma}_n), \check{\gamma}_n) - \mathbb{P}_n M(a_k, a_k)) \xrightarrow{\mathcal{L}aw} \mathcal{N}(0, I)$, where k represents the k^{th} step of our algorithm and where I is the identity matrix in $\mathbb{R}^d$.*

Note that k is fixed in theorem 4.2 since $\tilde{\gamma}_n = \arg \inf_{a \in \Theta} \sup_{c \in \Theta} \mathbb{P}_n M(c, a)$ where M is a known function of k —see Section 4.1. Thus, in the case when $D_\phi(g^{(k-1)} \frac{f_{a_k}}{g_{a_k}^{(k-1)}}, f) = 0$, we obtain

Corollary 4.1. We have $\sqrt{n}(\text{Var}_{\mathbf{P}}(M(\tilde{c}_n(\tilde{\gamma}_n), \tilde{\gamma}_n)))^{-1/2} \mathbb{P}_n M(\tilde{c}_n(\tilde{\gamma}_n), \tilde{\gamma}_n) \xrightarrow{\mathcal{L}^{aw}} \mathcal{N}(0, I)$.

Hence, we propose the test of the null hypothesis

$$(H_0) : D_\phi(g^{(k-1)} \frac{f_{a_k}}{g_{a_k}^{(k-1)}}, f) = 0 \text{ versus the alternative } (H_1) : D_\phi(g^{(k-1)} \frac{f_{a_k}}{g_{a_k}^{(k-1)}}, f) \neq 0.$$

Based on this result, we stop the algorithm, then, defining a_k as the last vector generated, we derive from corollary 4.1 a α -level confidence ellipsoid around a_k , namely $\mathcal{E}_k = \{b \in \mathbb{R}^d; \sqrt{n}(\text{Var}_{\mathbf{P}}(M(b, b)))^{-1/2} \mathbb{P}_n M(b, b) \leq q_\alpha^{N(0,1)}\}$, where $q_\alpha^{N(0,1)}$ is the quantile of a α -level reduced centered normal distribution and where $\mathbb{P}_n$ is the empirical measure arising from a realization of the sequences $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_n)$.

Consequently, the following corollary provides us with a confidence region for the above test:

Corollary 4.2. $\mathcal{E}_k$ is a confidence region for the test of the null hypothesis (H_0) versus (H_1) .

5. Goodness-of-Fit Tests

5.1. The Basic Idea

Let f be a density defined on $\mathbb{R}^2$. Let us also consider g , a known elliptical density with the same mean and variance as f . Let us also assume that the family (a_i) is the canonical basis of $\mathbb{R}^2$ and that $D_\phi(g^{(2)}, f) = 0$.

Hence, since lemma C.1 page 110 implies that $g_{a_j}^{(j-1)} = g_{a_j}$ if $j \leq d$, we then have $g^{(2)}(x) = g(x) \frac{f_1}{g_1} \frac{f_2}{g_2} = g(x) \frac{f_1}{g_1} \frac{f_2}{g_2}$. Moreover, we get f with $g^{(2)} = f$, as derived from property B.1 page 110.

Consequently, $f = g(x) \frac{f_1}{g_1} \frac{f_2}{g_2}$, i.e., $\frac{f}{f_1 f_2} = \frac{g}{g_1 g_2}$, and then $\frac{\partial^2}{\partial x \partial y} C_f = \frac{\partial^2}{\partial x \partial y} C_g$ where C_f (resp. C_g) is the copula of f (resp. g).

More generally, if f is defined on $\mathbb{R}^d$, then the family (a_i) is once again free (see lemma C.5), i.e., the family (a_i) is once again a basis of $\mathbb{R}^d$. The relationship $D_\phi(g^{(d)}, f) = 0$ therefore implies that $g^{(d)} = f$, i.e., for any $x \in \mathbb{R}^d$, $f(x) = g^{(d)}(x) = g(x) \prod_{k=1}^d \frac{f_{a_k}(a_k^\top x)}{[g^{(k-1)}]_{a_k}(a_k^\top x)} = g(x) \prod_{k=1}^d \frac{f_{a_k}(a_k^\top x)}{g_{a_k}(a_k^\top x)}$ since lemma C.1 page 110 implies that $g_{a_k}^{(k-1)} = g_{a_k}$ if $k \leq d$. In other words, for any $x \in \mathbb{R}^d$, it holds

$$\frac{g(x)}{\prod_{k=1}^d g_{a_k}(a_k^\top x)} = \frac{f(x)}{\prod_{k=1}^d f_{a_k}(a_k^\top x)} \quad (5.1)$$

Finally, putting $A = (a_1, \dots, a_d)$ and defining vector y (resp. density $\tilde{f}$, copula $\tilde{C}_f$ of $\tilde{f}$, density $\tilde{g}$, copula $\tilde{C}_g$ of $\tilde{g}$) as the expression of vector x (resp. density f , copula C_f of f , density g , copula C_g of g) in basis A , then, the following proposition provides us with the density associated with the copula of f as being equal to the density associated with the copula of g in basis A :

Proposition 5.1. With the above notations, should a sequence $(a_i)_{i=1, \dots, d}$ of not null vectors in $\mathbb{R}_*^d$ defining $g^{(d)}$ and such that $D_\phi(g^{(d)}, f) = 0$ exist, then $\frac{\partial^d}{\partial y_1 \dots \partial y_d} \tilde{C}_f = \frac{\partial^d}{\partial y_1 \dots \partial y_d} \tilde{C}_g$.

5.2. With the Elliptical Copula

Let f be an unknown density defined on $\mathbb{R}^d$. The objective of the present section is to determine whether the copula of f is elliptical. We thus define an instrumental elliptical density g with the same mean and variance as f , and we follow the procedure of Section 3.2. As explained in Section 5.1, we infer from proposition 5.1 that the copula of f equals the copula of g when $D_\phi(g^{(d)}, f) = 0$, i.e., when a_d is the last vector generated from the algorithm and when (a_i) is the canonical basis of $\mathbb{R}^d$. Thus, in order to verify this assertion, corollary 4.1 page 96 provides us with a α -level confidence ellipsoid around this vector, namely

$$\mathcal{E}_d = \{b \in \mathbb{R}^d; \sqrt{n}(\text{Var}_{\mathbf{P}}(M(b, b)))^{-1/2} \mathbb{P}_n M(b, b) \leq q_\alpha^{\mathcal{N}(0,1)}\}$$

where $q_\alpha^{\mathcal{N}(0,1)}$ is the quantile of a α -level reduced centered normal distribution, where $\mathbb{P}_n$ is the empirical measure arising from a realization of the sequences $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_n)$ —see Appendix D—and where M is a known function of d , f_n and $g_n^{(d-1)}$ —see Section 4.1.

Consequently, keeping the notations introduced in Section 5.1, we perform a statistical test of the null hypothesis

$$(H_0) : \frac{\partial^d}{\partial x_1 \dots \partial x_d} C_f = \frac{\partial^d}{\partial x_1 \dots \partial x_d} C_g \text{ versus } (H_1) : \frac{\partial^d}{\partial x_1 \dots \partial x_d} C_f \neq \frac{\partial^d}{\partial x_1 \dots \partial x_d} C_g$$

Since, under (H_0) , we have $D_\phi(g^{(d)}, f) = 0$, then the following theorem provides us with a confidence region for this test.

Theorem 5.1. *The set $\mathcal{E}_d$ is a confidence region for the test of the null hypothesis (H_0) versus the alternative (H_1) .*

Remark 5.1. 1/ If $D_\phi(g^{(k)}, f) = 0$, for $k < d$, then we reiterate the algorithm until $g^{(d)}$ is created in order to obtain a relationship for the copula of f .

2/ If the a_i do not constitute the canonical basis, then keeping the notations introduced in Section 5.1, our algorithm meets the test:

$$(H_0) : \frac{\partial^d}{\partial y_1 \dots \partial y_d} \tilde{C}_f = \frac{\partial^d}{\partial y_1 \dots \partial y_d} \tilde{C}_g \text{ versus } (H_1) : \frac{\partial^d}{\partial y_1 \dots \partial y_d} \tilde{C}_f \neq \frac{\partial^d}{\partial y_1 \dots \partial y_d} \tilde{C}_g$$

Thus, our method permits to determine whether the copula of f equals the copula of g in the $(a_1, \dots, a_d)$ basis.

5.3. With the Independent Copulas

Let f be a density on $\mathbb{R}^d$ and let X be a random vector with f as density. The objective of this section is to determine whether f is the product of its margins, i.e., whether the copula of f is the independent copula. Let g be an instrumental product of univariate Gaussian density—with $\text{diag}(\text{Var}(X_1), \dots, \text{Var}(X_d))$ as covariance matrix and with the same mean as f . As explained at Section 5.2, we follow the procedure described at Section 3.2, i.e., proposition 5.1 infers that the copula of f is the independent copula when $D_\phi(g^{(d)}, f) = 0$, we then perform a statistical test of the null hypothesis:

$$(H_0) : f = \prod_{i=1}^d f_i \text{ versus the alternative } (H_1) : f \neq \prod_{i=1}^d f_i$$

Since, under (H_0) , we have $D_\phi(g^{(d)}, f) = 0$, the following theorem provides us with a confidence region for our test.

Theorem 5.2. *Keeping the notations of Section 5.2, the set $\mathcal{E}_d$ is a confidence region for the test of the null hypothesis (H_0) versus the alternative (H_1) .*

Remark 5.2. (1) As explained in Section 5.2, if $D_\phi(g^{(k)}, f) = 0$, for $k < d$, we reiterate the algorithm until $g^{(d)}$ is created in order to derive a relationship for the copula of f .

(2) If the a_i do not constitute the canonical basis, then keeping the notations of Section 5.1, our algorithm meets the test:

$$(H_0) : f = \Pi_{i=1}^d f_{a_i} \text{ versus the alternative } (H_1) : f \neq \Pi_{i=1}^d f_{a_i}$$

Thus, our method enables us to determine if the copula of f is the independent copula in the $(a_1, \dots, a_d)$ basis.

5.4. Study of the Subsequence $(g^{(k')})$ Defined by $D_\phi(g^{(k')}, f) = 0$ for Any k'

Let $\mathcal{Q}$ be the set of non-negative integers defined by $\mathcal{Q} = \{k'_i; k'_1 = 1, k'_q = d, k'_i < k'_{i+1}\}$, where q —such that $q \leq d$ —is its cardinal. In the present section, our goal is to study the subsequence $(g^{(k')})$ of the sequence $(g^{(k)})_{k=1..d}$ defined by $D_\phi(g^{(k')}, f) = 0$ for any k' belonging to $\mathcal{Q}$.

First, we have:

$$D_\phi(g^{(d)}, f) = 0 \Leftrightarrow g^{(d)} = f, \text{ through property B.1}$$

$$\Leftrightarrow \frac{g(x)}{\Pi_{k=1}^d g_{a_k}(a_k^\top x)} = \frac{f(x)}{\Pi_{k=1}^d f_{a_k}(a_k^\top x)}, \text{ as explained in Section 5.2}$$

$$\Leftrightarrow \frac{\tilde{g}(y)}{\Pi_{k=1}^d \tilde{g}_k(y_k)} = \frac{\tilde{f}(y)}{\Pi_{k=1}^d \tilde{f}_k(y_k)}, \text{ which amounts to the previous relationship written in the } A = (a_1, \dots, a_d) \text{ basis with the notations introduced in Section 5.2.}$$

Moreover, defining $\tilde{k}'_i$ as the previous integer k'_i , in the space $\{1, \dots, d\}$, with $i > 1$, and as explained in Section 3.1, the relationship $D_\phi(g^{(k')}, f) = 0$ implies that

$$\tilde{f}(y_i, \dots, y_{\tilde{k}'_{i+1}} / y_1, \dots, y_{\tilde{k}'_i}, y_{k'_{i+1}}, \dots, y_d) = \tilde{f}_{i,i+1}(y_i, \dots, y_{\tilde{k}'_{i+1}})$$

where $\tilde{f}_{i,i+1}$ is the density of vector $(a_i^\top X, \dots, a_{\tilde{k}'_{i+1}}^\top X)$ in the $A = (a_1, \dots, a_d)$ basis. Consequently, $\tilde{f}(y) = \tilde{f}_{1,2}(y_1, \dots, y_{\tilde{k}'_2}) \cdot \tilde{f}_{2,3}(y_{k'_2}, \dots, y_{\tilde{k}'_3}) \dots \tilde{f}_{q-1,d}(y_{k'_{q-1}}, \dots, y_{\tilde{k}'_d})$.

Hence, we can infer that

$$\frac{\tilde{f}(y)}{\Pi_{k=1}^d \tilde{f}_k(y_k)} = \frac{\tilde{f}_{1,2}(y_1, \dots, y_{\tilde{k}'_2})}{\Pi_{k=1}^{\tilde{k}'_2} \tilde{f}_k(y_k)} \cdot \frac{\tilde{f}_{2,3}(y_{k'_2}, \dots, y_{\tilde{k}'_3})}{\Pi_{k=k'_2}^{\tilde{k}'_3} \tilde{f}_k(y_k)} \dots \frac{\tilde{f}_{q-1,d}(y_{k'_{q-1}}, \dots, y_{\tilde{k}'_d})}{\Pi_{k=k'_{q-1}}^d \tilde{f}_k(y_k)} \quad (5.2)$$

The following theorem explicitly describes the form of the f copula in the $A = (a_1, \dots, a_d)$ basis:

Theorem 5.3. *Defining $\tilde{C}_{f_{i,j}}$ as the copula of $\tilde{f}_{i,j}$ and keeping the notations introduced in Sections 5.1 and 5.4, it holds*

$$\frac{\partial^d}{\partial y_1 \dots \partial y_d} \tilde{C}_f = \frac{\partial^{\tilde{k}'_2}}{\partial y_1 \dots \partial y_{\tilde{k}'_2}} \tilde{C}_{f_{1,2}} \cdot \frac{\partial^{\tilde{k}'_3 - k'_2 + 1}}{\partial y_{k'_2} \dots \partial y_{\tilde{k}'_3}} \tilde{C}_{f_{2,3}} \dots \frac{\partial^{d - k'_{q-1} + 1}}{\partial y_{k'_{q-1}} \dots \partial y_d} \tilde{C}_{f_{q-1,d}}$$

Remark 5.3. If there exists i such that $i < d$ and $k'_i = \tilde{k}'_{i+1}$, then the notation $\tilde{f}_{i,i+1}(y_{k'_i}, \dots, y_{\tilde{k}'_{i+1}})$ means $\tilde{f}_{k'_i}(y_{k'_i})$. Thus, if, for any k , we have $D_\phi(g^{(k)}, f) = 0$, then, for any $i < d$, we have $k'_i = \tilde{k}'_{i+1}$, i.e., we have $\tilde{f} = \prod_{k=1}^d \tilde{f}_k(y_k)$, where $\tilde{f}_k$ is the k^{th} marginal density of $\tilde{f}$.

At present, using relationship 5.2 and remark 5.3, the following corollary gives us the copula of f as equals to 1 in the $\{a_1, \dots, a_d\}$ basis when, for any k , $D_\phi(g^{(k)}, f) = 0$:

Corollary 5.1. In the case where, for any k , $D_\phi(g^{(k)}, f) = 0$, it holds:

$$\frac{\partial^d}{\partial y_1 \dots \partial y_d} \tilde{C}_f = 1$$

6. Simulations

Let us examine three simulations and an application to real datasets. The first simulation studies the elliptical copula and the second studies the independent copula. In each simulation, our program will aim at creating a sequence of densities $(g^{(j)})$, $j = 1, \dots, d$ such that $g^{(0)} = g$, $g^{(j)} = g^{(j-1)} f_{a_j} / [g^{(j-1)}]_{a_j}$ and $D_\phi(g^{(d)}, f) = 0$, where D_ϕ is a divergence—see Appendix B for its definition—and $a_j = \arg \inf_b D_\phi(g^{(j-1)} f_b / g_b^{(j-1)}, f)$, for all $j = 1, \dots, d$. We will therefore perform the tests introduced at theorems 5.1 and 5.2. Finally, the third simulation compares the optimisations obtained, when we execute the process with, each time, a new ϕ -divergence.

Simulation 6.1. We are in dimension 2(=d), and we use the χ^2 divergence to perform our optimisations. Let us consider a sample of 50(=n) values of a random variable X with a density law f defined by :

$$f(x) = c_\rho(F_{Gumbel}(x_1), F_{Exponential}(x_2)).Gumbel(x_1).Exponential(x_2)$$

where c is the Gaussian copula with correlation coefficient $\rho = 0.5$, and where the Gumbel distribution parameters are -1 and 1 and the exponential density parameter is 2 .

Let us generate then a Gaussian random variable Y with a density—that we will name as g —presenting the same mean and variance as f .

We theoretically obtain $k = 2$ and $(a_1, a_2) = ((1, 0), (0, 1))$.

To get this result, we perform the following test:

$$(H_0) : (a_1, a_2) = ((1, 0), (0, 1)) \text{ versus } (H_1) : (a_1, a_2) \neq ((1, 0), (0, 1))$$

Then, theorem 5.1 enables us to verify (H_0) by the following $0.9(=\alpha)$ level confidence ellipsoid

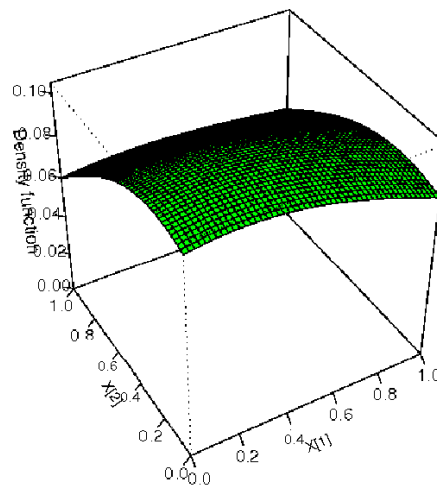
$$\mathcal{E}_2 = \{b \in \mathbb{R}^2; (Var_{\mathbf{P}}(M(b, b)))^{(-1/2)} \mathbb{P}_n M(b, b) \leq q_\alpha^{N(0,1)} / \sqrt{n} \simeq 0, 2533 / 7.0710 = 0.03582\}$$

Results of this optimisation can be found in Table 3 and Figure 1.

Table 3. Simulation 1: Numerical results of the optimisation.

Our Algorithm	
	minimum : 0.445199
Projection Study 0:	at point : (1.0171,0.0055)
	<i>P</i> -Value : 0.94579
Test:	$H_1 : a_1 \notin \mathcal{E}_1 : \text{True}$
	minimum : 0.009628
Projection Study 1:	at point : (0.0048,0.9197)
	<i>P</i> -Value : 0.99801
Test:	$H_0 : a_2 \in \mathcal{E}_2 : \text{True}$
$\chi^2(\text{Kernel Estimation of } g^{(2)}, g^{(2)})$	3.57809

Therefore, we can conclude that H_0 is verified.

Figure 1. Graph of the estimate of $(x_1, x_2) \mapsto c_\rho(F_{\text{Gumbel}}(x_1), F_{\text{Exponential}}(x_2))$.

Simulation 6.2. We are in dimension 2(=d), and we use the χ^2 divergence to perform our optimisations.

Let us consider a sample of 50(=n) values of a random variable X with a density law f defined by

$$f(x) = \text{Gumbel}(x_1) \cdot \text{Exponential}(x_2)$$

where the Gumbel distribution parameters are -1 and 1 and the exponential density parameter is 2 .

Let g be an instrumental product of univariate Gaussian densities with $\text{diag}(\text{Var}(X_1), \dots, \text{Var}(X_d))$ as covariance matrix and with the same mean as f .

We theoretically obtain $k = 2$ and $(a_1, a_2) = ((1, 0), (0, 1))$. To get this result, we perform the following test:

$$(H_0) : (a_1, a_2) = ((1, 0), (0, 1)) \text{ versus } (H_1) : (a_1, a_2) \neq ((1, 0), (0, 1))$$

Then, theorem 5.2 enables us to verify (H_0) by the following $0.9(=\alpha)$ level confidence ellipsoid

$$\mathcal{E}_2 = \{b \in \mathbb{R}^2; (Var_{\mathbf{P}}(M(b, b)))^{(-1/2)} \mathbb{P}_n M(b, b) \leq q_{\alpha}^{N(0,1)} / \sqrt{n} \simeq 0.03582203\}$$

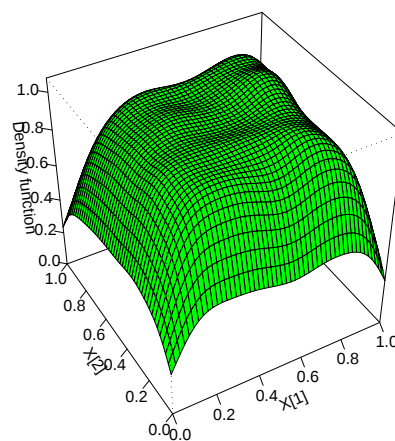
Results of this optimisation can be found in Table 4 and Figure 2.

Table 4. Simulation 2: Numerical results of the optimisation.

Our Algorithm	
	minimum : 0.057833
Projection Study 0 :	at point : (0.9890,0.1009)
	P -Value : 0.955651
Test :	$H_1 : a_1 \notin \mathcal{E}_1$: True
	minimum : 0.02611
Projection Study 1 :	at point : (−0.1105,0.9290)
	P -Value : 0.921101
Test :	$H_0 : a_2 \in \mathcal{E}_2$: True
χ^2 (Kernel Estimation of $g^{(2)}, g^{(2)}$)	1.25945

Therefore, we can conclude that $f = \Pi_{i=1}^d f_i$.

Figure 2. Graph of the independent copula estimate.



Simulation 6.3 (On the choice of a ϕ -divergence). In this paragraph, we perform our algorithm several times. We first use several ϕ -divergences (see Appendix B for their definitions and their notations). We then perform a sensitivity analysis by varying the number n of simulated variables. Finally we introduce outliers.

At present, we consider a sample of n values of a random variable X with a density f defined by $f(x) = \text{Laplace}(x_1) \cdot \text{Gumbel}(x_0)$,

where the Gumbel distribution parameters are $(1, 2)$ and where the Laplace distribution parameters are 4 and 3. In theory, we get $a_1 = (0, 1)$ and $a_2 = (1, 0)$. Then, following the procedure of the first simulation, we get

$n = 50$	Outliers = 0	Time	Outliers = 2	Time
Relative Entropy	(0.10, 0.83) (1.13, 0.11)	30 mn	(0.1, 0.8) (0.80, 0.024)	43 mn
χ^2 -divergence	(0, 0.8) (1.021, 0.09)	22 mn	(0.12, 0.79) (0.867, -0.104)	31 mn
Hellinger distance	(0.1, 0.9) (0.91, 0.15)	35 mn	(0.1, 0.85) (0.81, 0.14)	46 mn
$n = 100$	Outliers = 0	Time	Outliers = 5	Time
Relative Entropy	(0.09, 0.89) (1.102, 0.089)	50 mn	(0.1, 0.88) (1.15, 0.144)	60 mn
χ^2 -divergence	(0, 0.9) (0.97, -0.1)	43 mn	(-0.1, 0.9) (0.87, 0.201)	52 mn
Hellinger distance	(0.1, 0.91) (0.93, -0.11)	57 mn	(-0.05, 1.1) (0.79, 0.122)	62 mn
$n = 500$	Outliers = 0	Time	Outliers = 25	Time
Relative Entropy	(0, 1.07) (1.1, -0.05)	107 mn	(0.13, 0.75) (0.79, 0.122)	121 mn
χ^2 -divergence	(0, 0.95) (1.12, -0.02)	91 mn	(0.15, 0.814) (0.922, 0.147)	103 mn
Hellinger distance	(-0.01, 0.95) (1.01, -0.073)	100 mn	(-0.17, 1.3) (0.973, 0.206)	126 mn

Remark 6.1.

(1) We have worked with a calculator presenting the following characteristics :

- Processor : Mobile AMD 3000+,
- Memory RAM : 512 DDR,
- Windows XP.

(2) Our method, which uses the χ^2 as ϕ -divergence, is faster and its performance is as good if not better than any other divergence method.

This results from the fact that the projection index (or criteria) of χ^2 is a second degree polynomial. It is consequently easier and faster to assess. Moreover, these simulations illustrate the robustness of our method.

6.1. Application to Real Datasets

Let us for instance study the moves in the stock prices of Renault and Peugeot from January 4, 2010 to July 25, 2010. We thus gather 140(=n) data from these stock prices, see Table 7 and Table 8 below.

Let us also consider X_1 (resp. X_2) the random variable defining the stock price of Renault (resp. Peugeot). We will assume—as it is commonly done in mathematical finance—that the stock market abides by the classical hypotheses of the Black-Scholes model—see Black and Scholes [34].

Consequently, X_1 and X_2 each present a log-normal distribution as probability distribution.

Let f be the density of vector $(\ln(X_1), \ln(X_2))$, let us now apply our algorithm to f with the Kullback-Leibler divergence as ϕ -divergence. Let us generate then a Gaussian random variable Y with a density—that we will name as g —presenting the same mean and variance as f .

We first assume that there exists a vector a such that $D_\phi(g_{\frac{f_a}{g_a}}, f) = 0$.

In order to verify this hypothesis, our reasoning will be the same as in Simulation 6.1. Indeed, we assume that this vector is a co-factor of f . Consequently, corollary 4.2 enables us to estimate a by the following $0.9(=\alpha)$ level confidence ellipsoid

$$\mathcal{E}_1 = \{b \in \mathbb{R}^2; (Var_{\mathbf{P}}(M(b, b)))^{(-1/2)} \mathbb{P}_n M(b, b) \leq q_{\alpha}^{N(0,1)} / \sqrt{n} \simeq 0,2533 / \sqrt{140} = 0.02140776\}.$$

Numerical results of the first projection are summarized in Table 5.

Table 5. Numerical results: First projection.

Our Algorithm	
	minimum : 0.02087685
Projection Study 0:	at point : $a_1=(19.1,-12.3)$
	P-Value : 0.748765
Test:	$H_0 : a_1 \in \mathcal{E}_1 : \text{True}$
K(Kernel Estimation of $g^{(1)}, g^{(1)}$)	4.3428735

Therefore, our first hypothesis is confirmed.

However, our goal is to study the copula of $(\ln(X_1), \ln(X_2))$. Then, as explained in Section 5.4, we formulate another hypothesis assuming that there exists a vector a such that $D_{\phi}(g^{(1)} \frac{f_a}{g_a^{(1)}}, f) = 0$.

In order to verify this hypothesis, we use the same reasoning as above. Indeed, we assume that this vector is a co-factor of f . Consequently, corollary 4.2 enables us to estimate a by the following $0.9(=\alpha)$ level confidence ellipsoid $\mathcal{E}_2 = \{b \in \mathbb{R}^2; (Var_{\mathbf{P}}(M(b, b)))^{(-1/2)} \mathbb{P}_n M(b, b) \leq q_{\alpha}^{N(0,1)} / \sqrt{n} \simeq 0,2533 / \sqrt{140} = 0.02140776\}$. Numerical results of the second projection are summarized in Table 6.

Table 6. Numerical results: Second projection.

Our Algorithm	
	minimum : 0.0198753
Projection Study 1:	at point : $a_2=(8.1,3.9)$
	P-Value : 0.8743401
Test:	$H_0 : a_2 \in \mathcal{E}_2 : \text{True}$
K(Kernel Estimation of $g^{(2)}, g^{(2)}$)	4.38475324

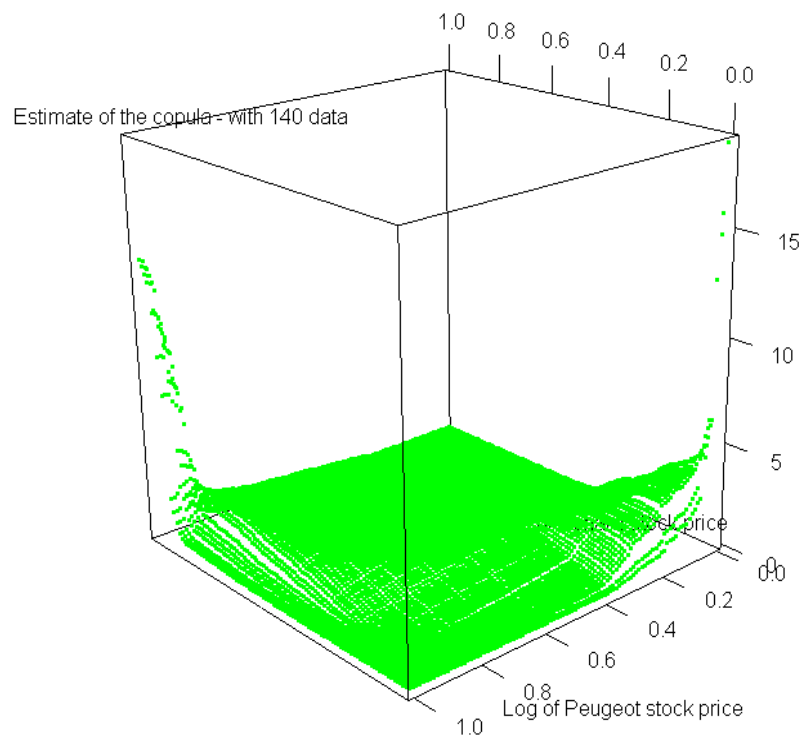
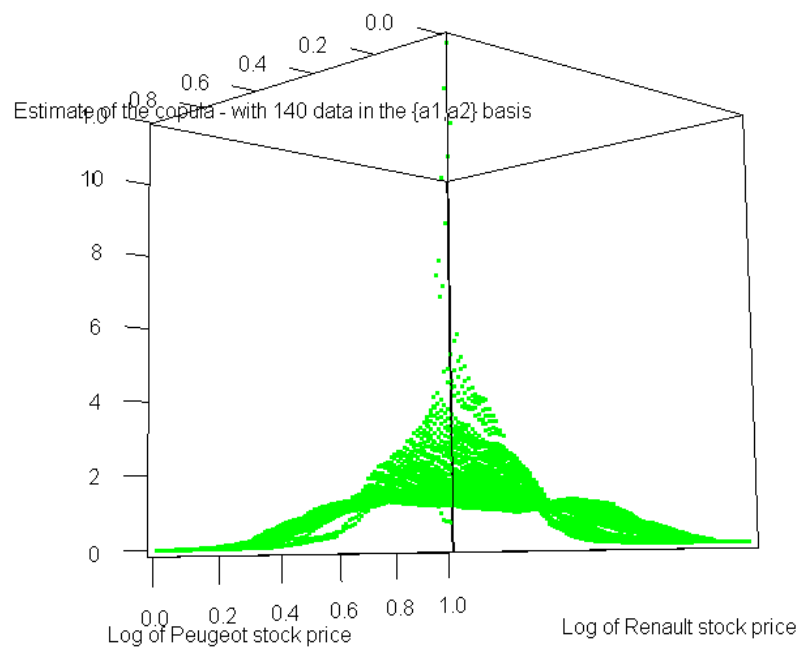
Therefore, our second hypothesis is confirmed.

In conclusion, as explained in corollary 5.1, the copula of f is equal to 1 in the $\{a_1, a_2\}$ basis.

This result has been illustrated at Figures 3,4 and 5.

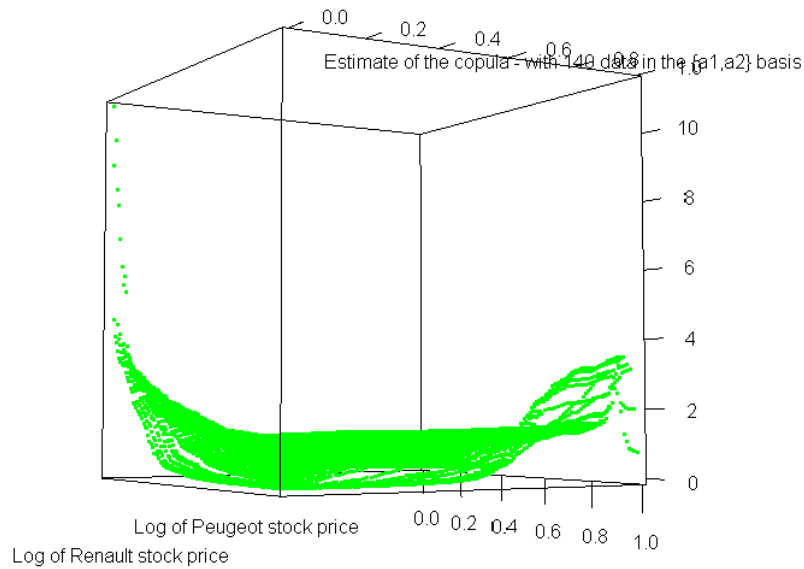
6.2. Critics of the Simulations

In the case where f is unknown, we will never be sure to have reached the minimum of the ϕ -divergence: the simulated annealing method has been used to solve our optimisation problem, and therefore it is only when the number of random jumps tends in theory towards infinity that the probability to get the minimum tends to 1. We also note that no theory on the optimal number of jumps to implement does exist, as this number depends on the specificities of each particular problem.

Figure 3. Graph of the copula of $(\ln(X_1), \ln(X_2))$ in the canonical basis.**Figure 4.** Graph of the copula of $(\ln(X_1), \ln(X_2))$ in the $\{a_1, a_2\}$ basis.

Moreover, we choose the $50^{-\frac{4}{4+d}}$ for the AMISE of the two simulations. This choice leads us to simulate 50 random variables—see Scott [23] page 151, none of which have been discarded to obtain the truncated sample.

This has also been the case in our application to real datasets.

Figure 5. Graph of the copula of $(\ln(X_1), \ln(X_2))$ in the $\{a_1, a_2\}$ basis—other view.**Table 7.** Stock prices of Renault and Peugeot.

Date	Renault	Peugeot	Date	Renault	Peugeot	Date	Renault	Peugeot
23/07/10	34.9	24.2	22/07/10	34.26	24.01	21/07/10	33.15	23.3
20/07/10	32.69	22.78	19/07/10	33.24	23.36	16/07/10	33.92	23.77
15/07/10	34.44	23.71	14/07/10	35.08	24.36	13/07/10	35.28	24.37
12/07/10	33.84	23.16	09/07/10	33.46	23.13	08/07/10	33.08	22.65
07/07/10	32.15	22.19	06/07/10	31.12	21.56	05/07/10	30.02	20.81
02/07/10	30.17	20.85	01/07/10	29.56	20.05	30/06/10	30.78	21.07
29/06/10	30.55	20.97	28/06/10	32.34	22.3	25/06/10	31.35	21.68
24/06/10	32.29	22.25	23/06/10	33.58	22.47	22/06/10	33.84	22.77
21/06/10	34.06	23.25	18/06/10	32.89	22.7	17/06/10	32.08	22.31
16/06/10	31.87	21.92	15/06/10	32.03	22.12	14/06/10	31.45	22.2
11/06/10	30.62	21.42	10/06/10	30.42	20.93	09/06/10	29.27	20.34
08/06/10	28.48	19.73	07/06/10	28.92	20.15	04/06/10	29.19	20.27
03/06/10	30.35	20.46	02/06/10	29.33	19.53	01/06/10	28.87	19.45
31/05/10	29.39	19.54	28/05/10	29.16	19.55	27/05/10	29.18	19.81
26/05/10	27.5	18.5	25/05/10	26.76	18.08	24/05/10	28.75	18.81
21/05/10	28.78	18.82	20/05/10	28.53	18.84	19/05/10	29.49	19.25
18/05/10	30.95	19.76	17/05/10	30.92	19.35	14/05/10	31.35	19.34
13/05/10	33.65	20.76	12/05/10	33.63	20.52	11/05/10	33.38	20.34
10/05/10	33.28	20.3	07/05/10	31	19.24	06/05/10	32.4	20.22
05/05/10	32.95	20.45	04/05/10	33.3	21.03	03/05/10	35.58	22.63
30/04/10	35.41	22.45	29/04/10	35.53	22.36	28/04/10	34.75	22.33

Table 8. Stock prices of Renault and Peugeot.

Date	Renault	Peugeot	Date	Renault	Peugeot	Date	Renault	Peugeot
27/04/10	36.2	22.9	26/04/10	37.65	23.73	23/04/10	36.72	23.5
22/04/10	34.36	22.72	21/04/10	35.01	22.86	20/04/10	35.62	22.88
19/04/10	34.08	21.77	16/04/10	34.46	21.71	15/04/10	35.16	22.22
14/04/10	35.1	22.22	13/04/10	35.28	22.45	12/04/10	35.17	21.85
09/04/10	35.76	21.9	08/04/10	35.67	21.67	07/04/10	36.5	21.89
06/04/10	36.87	22	01/04/10	35.5	21.97	31/03/10	34.7	21.8
30/03/10	34.8	22.24	29/03/10	35.7	22.73	26/03/10	35.54	22.58
25/03/10	35.53	22.73	24/03/10	33.8	21.82	23/03/10	34.1	21.58
22/03/10	33.73	21.64	19/03/10	34.12	21.68	18/03/10	34.44	21.75
17/03/10	34.68	21.98	16/03/10	34.33	21.88	15/03/10	33.57	21.53
12/03/10	33.9	21.86	11/03/10	33.27	21.58	10/03/10	33.12	21.47
09/03/10	32.69	21.54	08/03/10	32.99	21.66	05/03/10	32.89	21.85
04/03/10	31.64	21.26	03/03/10	31.65	20.7	02/03/10	31.05	20.2
01/03/10	30.26	19.54	26/02/10	30.2	19.39	25/02/10	29.42	18.98
24/02/10	30.9	19.49	23/02/10	30.54	19.74	22/02/10	31.89	20.06
19/02/10	32.29	20.67	18/02/10	32.26	20.41	17/02/10	31.69	20.31
16/02/10	31.08	19.8	15/02/10	30.25	19.66	12/02/10	29.56	19.57
11/02/10	31	20.4	10/02/10	32.78	21.21	09/02/10	33.31	22.31
08/02/10	32.63	21.95	05/02/10	32.15	22.33	04/02/10	33.72	22.86
03/02/10	35.32	23.93	02/02/10	35.29	23.8	01/02/10	35.31	24.05
29/01/10	34.26	23.64	28/01/10	33.94	23.31	27/01/10	33.85	23.88
26/01/10	34.97	24.86	25/01/10	35.06	24.35	22/01/10	35.7	24.95
21/01/10	36.1	25	20/01/10	36.92	25.35	19/01/10	38.4	25.81
18/01/10	39.28	25.95	15/01/10	38.6	25.7	14/01/10	39.56	26.67
13/01/10	39.49	26.13	12/01/10	38.36	25.98	11/01/10	39.21	26.65
08/01/10	39.38	26.5	07/01/10	39.69	26.7	06/01/10	39.25	26.32
05/01/10	38.31	24.74	04/01/10	38.2	24.52			

Finally, the shape of the copula in the case of real datasets in the $\{a_1, a_2\}$ basis is also noteworthy.

Figure 4 shows that the curve reaches a quite wide plateau around 1, whereas Figure 5 shows that this plateau prevails on almost the entire $[0, 1]^2$ set. We can therefore conclude that the theoretical analysis is indeed confirmed by the above simulation.

6.3. Conclusions

Projection pursuit is useful in evidencing characteristic structures as well as one-dimensional projections and their associated distribution in multivariate data. This article clearly demonstrates the efficiency of the φ -projection pursuit methodology for goodness-of-fit tests for copulas. Indeed, the

robustness as well as the convergence results that we achieved convincingly fulfilled our expectations regarding the methodology used.

Appendix

All the demonstrations of this article have been gathered in the Technical Report [24].

A. On the Different Families of Copula

There exists many copula families. Let us here present the most important amongst them.

A.1. Elliptical Copulas

The Gaussian Copula

The Gaussian copula can be used in several fields. For example, many credit models are built from this copula, which also presents the property to make extreme values (minimal or maximal) independent in the limit; see Joe [2] for more details. For example, in $\mathbb{R}^2$, it is derived from the bivariate normal distribution and from Sklar's theorem. Defining Ψ_ρ as the standard bivariate normal cumulative distribution function with ρ correlation, the Gaussian copula function is $C_\rho(u, v) = \Psi_\rho(\Psi^{-1}(u), \Psi^{-1}(v))$ where $u, v \in [0, 1]$ and where Ψ is the standard normal cumulative distribution function. Then, the copula density function is :

$$c_\rho(u, v) = \frac{\psi_{X,Y,\rho}(\Psi^{-1}(u), \Psi^{-1}(v))}{\psi(\Psi^{-1}(u))\psi(\Psi^{-1}(v))}$$

where $\psi_{X,Y,\rho}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}[x^2 + y^2 - 2\rho xy]\right)$ is the density function for the standard bivariate Gaussian with Pearson product-moment correlation coefficient ρ and where ψ is the standard normal density. This definition can obviously be extended to $\mathbb{R}^d$.

The Elliptical Copula

Let us begin with defining the class of elliptical distributions and its properties—see also Cambanis [17], Landsman [18] :

Definition A.1. *X is said to abide by a multivariate elliptical distribution, denoted $X \sim E_d(\mu, \Sigma, \xi_d)$, if X has the following density, for any x in $\mathbb{R}^d$:*

$$f_X(x) = \frac{\alpha_d}{|\Sigma|^{1/2}} \xi_d\left(\frac{1}{2}(x - \mu)' \Sigma^{-1}(x - \mu)\right)$$

where Σ is a $d \times d$ positive-definite matrix and where μ is a d -column vector,

where ξ_d is referred as the “density generator”,

where α_d is a normalisation constant, such that $\alpha_d = \frac{\Gamma(d/2)}{(2\pi)^{d/2}} \left(\int_0^\infty x^{d/2-1} \xi_d(x) dx \right)^{-1}$,

with $\int_0^\infty x^{d/2-1} \xi_d(x) dx < \infty$.

Property A.1. (1) For any $X \sim E_d(\mu, \Sigma, \xi_d)$, for any $m \times d$ matrix with rank $m \leq d$, A , and for any m -dimensional vector b , we have $AX + b \sim E_m(A\mu + b, A\Sigma A', \xi_m)$.

Therefore, any marginal density of multivariate elliptical distribution is elliptical, i.e.,

$X = (X_1, X_2, \dots, X_d) \sim E_d(\mu, \Sigma, \xi_d) \Rightarrow X_i \sim E_1(\mu_i, \sigma_i^2, \xi_1)$, $1 \leq i \leq d$, with $f_{X_i}(x) = \frac{\alpha_1}{\sigma_i} \xi_1\left(\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2\right)$. (2) Corollary 5 of Cambanis [17] states that conditional densities with elliptical distributions are also elliptical. Indeed, if $X = (X_1, X_2)' \sim E_d(\mu, \Sigma, \xi_d)$, with X_1 (resp. X_2) of size $d_1 < d$ (resp. $d_2 < d$), then $X_1/(X_2 = a) \sim E_{d_1}(\mu', \Sigma', \xi_{d_1})$ with $\mu' = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(a - \mu_2)$ and $\Sigma' = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, with $\mu = (\mu_1, \mu_2)$ and $\Sigma = (\Sigma_{ij})_{1 \leq i, j \leq 2}$.

Remark A.1. Landsman [18] shows that multivariate Gaussian distributions derive from $\xi_d(x) = e^{-x}$ and that if $X = (X_1, \dots, X_d)$ has an elliptical density such that its marginals verify $E(X_i) < \infty$ and $E(X_i^2) < \infty$ for $1 \leq i \leq d$, then μ is the mean of X and Σ is a multiple of the covariance matrix of X . Consequently, from now on, we will assume this is indeed the case.

Definition A.2. Let t be an elliptical density on $\mathbb{R}^k$ and let q be an elliptical density on $\mathbb{R}^{k'}$. The elliptical densities t and q are said to belong to the same family of elliptical densities, if their generating densities are ξ_k and $\xi_{k'}$ respectively, which belong to a common given family of densities.

Example A.1. Consider two Gaussian densities $\mathcal{N}(0, 1)$ and $\mathcal{N}((0, 0), Id_2)$. They are said to belong to the same elliptical family as they both present $x \mapsto e^{-x}$ as generating density.

Finally, let us introduce the definition of an elliptical copula which generalizes the above overview of the Gaussian copula:

Definition A.3. Elliptical copulas are the copulas of elliptical distributions.

A.2. Archimedean Copulas

These copulas exhibit a simple form as well as properties such as associativity. They also present a variety of dependent structures. They can generally be defined under the following form

$$\mathcal{A}(u_1, u_2, \dots, u_n) = \xi^{-1}\left(\sum_{i=1}^n \xi(F_i(u_i))\right)$$

where $(u_1, u_2, \dots, u_n) \in [0, 1]^n$ and where ξ is known as a “generator function”. This ξ function must be at least $d - 2$ times continuously differentiable, must have a decreasing and convex $d - 2$ derivative, and must be such that $\xi(1) = 0$.

Let us now present several examples :

(1) Clayton copula:

The Clayton copula is an asymmetric Archimedean copula, displaying greater dependency in the negative tail than in the positive tail. Let us define X (resp. Y) as the random vector having F (resp. G) as cumulative distribution function (CDF). Assuming that the vector (X, Y) has a Clayton copula, then this copula is given by:

$$\mathcal{A}(x, y) = (F(x)^\theta + G(y)^\theta - 1)^{1/\theta}$$

And its generator is:

$$\xi(x) = x^\theta - 1$$

For $\theta = 0$, the random variables are independent.

(2) *Gumbel copula*:

The Gumbel copula (Gumbel-Hougaard copula) is an asymmetric Archimedean copula, presenting greater dependency in the positive tail than in the negative tail. This copula is given by:

$$\xi(x) = (-\ln(x))^\alpha$$

(3) *Frank copula*:

The Frank copula is a symmetric Archimedean copula given by:

$$\xi(x) = \ln \left(\frac{e^{\alpha x} - 1}{e^\alpha - 1} \right)$$

A.3. Periodic Copula

In 2005, Alfonsi and Brigo [25] derived a new way of generating copulas based on periodic functions. Defining h (resp. $\mathcal{A}$) as a 1-periodic non-negative function that integrates to 1 over $[0, 1]$ (resp. as a double primitive of h), then both

$$\mathcal{A}(u+v) - \mathcal{A}(u) - \mathcal{A}(v) \text{ and } -\mathcal{A}(u-v) + \mathcal{A}(u) + \mathcal{A}(-v)$$

are copula functions, the second one not necessarily being exchangeable.

B. ϕ -Divergence

Let us call h_a the density of $a^\top Z$ if h is the density of Z . Let φ be a strictly convex function defined by $\varphi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, and such that $\varphi(1) = 0$.

Definition B.1. We define a ϕ -divergence of P from Q , where P and Q are two probability distributions over a space Ω such that Q is absolutely continuous with respect to P , by

$$D_\phi(Q, P) = \int \varphi \left(\frac{dQ}{dP} \right) dP \quad (\text{B.1})$$

The above expression (B.1) is also valid if P and Q are both dominated by the same probability.

The most used distances (Kullback, Hellinger or χ^2) belong to the Cressie-Read family (see Cressie-Read [26], Csiszár I. [27] and the books of Friedrich and Igor [28], Pardo Leandro [29] and Zografos K. [30]). They are defined by a specific φ . Indeed,

- with the Kullback-Leibler divergence, we associate $\varphi(x) = K(x) = x \ln(x) - x + 1$
- with the Hellinger distance, we associate $\varphi(x) = H(x) = 2(\sqrt{x} - 1)^2$
- with the χ^2 distance, we associate $\varphi(x) = \chi^2(x) = \frac{1}{2}(x - 1)^2$
- more generally, with power divergences, we associate $\varphi(x) = \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma - 1)}$, where $\gamma \in \mathbb{R} \setminus (0, 1)$
- and, finally, with the L^1 norm, which is also a divergence, we associate $\varphi(x) = |x - 1|$.

Let us now expose some well-known properties of divergences.

Property B.1. We have $D_\phi(P, Q) = 0 \Leftrightarrow P = Q$.

Property B.2. The divergence function $Q \mapsto D_\phi(Q, P)$ is convex and lower semi-continuous for the topology that makes all the applications of the form $Q \mapsto \int f dQ$ continuous (where f is bounded and continuous), and lower semi-continuous for the topology of the uniform convergence.

Finally, we will also use the following property derived from the first part of corollary (1.29) page 19 of Friedrich and Igor [28],

Property B.3.

If $T : (X, A) \rightarrow (Y, B)$ is measurable and if $D_\phi(P, Q) < \infty$, then $D_\phi(P, Q) \geq D_\phi(PT^{-1}, QT^{-1})$ with equality being reached when T is surjective for (P, Q) .

C. Miscellaneous

Lemma C.1. For any $p \leq d$, we have $g_{a_p}^{(p-1)} = g_{a_p}$.

Lemma C.2. We have $g(\cdot/a_1^\top x, \dots, a_j^\top x) = n(a_{j+1}^\top x, \dots, a_d^\top x) = f(\cdot/a_1^\top x, \dots, a_j^\top x)$.

Lemma C.3. Should there exist a family $(a_i)_{i=1\dots d}$ such that $f(x) = n(a_{j+1}^\top x, \dots, a_d^\top x)h(a_1^\top x, \dots, a_j^\top x)$, with $j < d$, with f , n and h being densities, then this family is an orthogonal basis of $\mathbb{R}^d$.

Lemma C.4. $\inf_{a \in \mathbb{R}_*^d} D_\phi(g_{a_a}^{f_a}, f)$ is reached when the ϕ -divergence is greater than the L^1 distance as well as the L^2 distance.

Lemma C.5. Whenever there exists p , $p \leq d$, such that $D_\phi(g^{(p)}, f) = 0$, then the family of $(a_i)_{i=1,\dots,p}$ is free and is orthogonal.

Lemma C.6. For any continuous density f , we have $y_m = |f_m(x) - f(x)| = O_{\mathbf{P}}(m^{-\frac{2}{4+d}})$.

D. Study of the Sample

Let $X_1, X_2, \dots, X_m$ be a sequence of independent random vectors with the same density f . Let $Y_1, Y_2, \dots, Y_m$ be a sequence of independent random vectors with the same density g . Then, the kernel estimators $f_m, g_m, f_{a,m}$ and $g_{a,m}$ of f, g, f_a and g_a , for all $a \in \mathbb{R}_*^d$, almost surely and uniformly converge since we assume that the bandwidth h_m of these estimators meets the following conditions (see Bosq [32]):

(Hyp): $h_m \searrow_m 0$, $mh_m \nearrow_m \infty$, $mh_m/L(h_m^{-1}) \rightarrow_m \infty$ and $L(h_m^{-1})/LLm \rightarrow_m \infty$,

with $L(u) = \ln(u \vee e)$.

Let us consider

$$B_1(n, a) = \frac{1}{n} \sum_{i=1}^n \varphi' \left\{ \frac{f_{a,n}(a^\top Y_i)}{g_{a,n}(a^\top Y_i)} \frac{g_n(Y_i)}{f_n(Y_i)} \right\} \frac{f_{a,n}(a^\top Y_i)}{g_{a,n}(a^\top Y_i)} \text{ and } B_2(n, a) = \frac{1}{n} \sum_{i=1}^n \varphi^* \left\{ \frac{f_{a,n}(a^\top X_i)}{g_{a,n}(a^\top X_i)} \frac{g_n(X_i)}{f_n(X_i)} \right\}$$

Our objective is to estimate the minimum of $D_\phi(g_{a_a}^{f_a}, f)$. To achieve this, samples have to be truncated:

Let us consider now a positive sequence θ_m such that $\theta_m \rightarrow 0$, $y_m/\theta_m^2 \rightarrow 0$, where y_m is the almost sure convergence rate of the kernel density estimator— $y_m = O_{\mathbf{P}}(m^{-\frac{2}{4+d}})$, see lemma C.6— $y_m^{(1)}/\theta_m^2 \rightarrow 0$, where $y_m^{(1)}$ is defined by

$$\left| \varphi \left(\frac{g_m(x) f_{b,m}(b^\top x)}{f_m(x) g_{b,m}(b^\top x)} \right) - \varphi \left(\frac{g(x) f_b(b^\top x)}{f(x) g_b(b^\top x)} \right) \right| \leq y_m^{(1)}$$

for all b in $\mathbb{R}_*^d$ and all x in $\mathbb{R}^d$, and finally $\frac{y_m^{(2)}}{\theta_m^2} \rightarrow 0$, where $y_n^{(2)}$ is defined by

$$|\varphi' \left(\frac{g_m(x) f_{b,m}(b^\top x)}{f_m(x) g_{b,m}(b^\top x)} \right) - \varphi' \left(\frac{g(x) f_b(b^\top x)}{f(x) g_b(b^\top x)} \right)| \leq y_m^{(2)}$$

for all b in $\mathbb{R}_*^d$ and all x in $\mathbb{R}^d$.

We then generate f_m , g_m and $g_{b,m}$ from the starting sample and we select the X_i and Y_i vectors such that $f_m(X_i) \geq \theta_m$ and $g_{b,m}(b^\top Y_i) \geq \theta_m$, for all i and for all $b \in \mathbb{R}_*^d$.

The vectors meeting these conditions will be called $X_1, X_2, \dots, X_n$ and $Y_1, Y_2, \dots, Y_n$.

Consequently, the next proposition provides us with the condition required to derive our estimates:

Proposition D.1. *Using the notations introduced in Broniatowski [33] and in Section 4.1, it holds*

$$\lim_{n \rightarrow \infty} \sup_{a \in \mathbb{R}_*^d} |(B_1(n, a) - B_2(n, a)) - D_\phi(g_{\frac{f_a}{g_a}}, f)| = 0.$$

Remark D.1. *With the Kullback-Leibler divergence, we can take for θ_m the expression $m^{-\nu}$, with $0 < \nu < \frac{1}{4+d}$.*

E. Hypotheses' Discussion

Not all hypotheses will be used simultaneously.

Hypotheses (A1) and (A4) lead us to assume we deal with a saddle point: being used to demonstrate the convergence of $\check{c}_n(a)$ and γ_k towards a_k , they make it easier to use the dual form of the divergence. Moreover, since our criteria $a \mapsto D_\phi(g_{\frac{f_a}{g_a}}, f)$ is differentiable on $\mathbb{R}_*^d$ and continuously differentiable on $\mathbb{R}^d$, these hypotheses can be easily obtained. However, if other discontinuities, for which the criteria can not be extended by continuity, do exist, then the above hypotheses would be very difficult to verify even in very favorable cases.

As shown by the below subsection for relative entropy, hypothesis (A2) generally holds.

Hypotheses (A5) and (A7) are classical hypotheses from which a limit distribution for the criteria can be derived. Yet these hypotheses are difficult to obtain when the criteria $a \mapsto D_\phi(g_{\frac{f_a}{g_a}}, f)$ admits discontinuities—close to the co-vectors of f —for which it can not be continuously differentiable.

Hypothesis (A6) thus enables to create a stopping rule for the process since this hypothesis is equivalent to the nullity of the application $D_\phi(g_{\frac{f_a}{g_a}}, f)$ in a_k .

Hypothesis (A0) constitutes an alternative to the starting hypothesis according to which the divergence should be greater than the L^1 distance. Although weaker, this hypothesis also requires that for all i , we have $K(g^{(i)}, f) \geq \int |f(x) - g^{(i)}(x)| dx$ at each iteration of the algorithm.

E.1. Discussion of (A2).

Let us work with the Kullback-Leibler divergence and with g and a_1 .

For all $b \in \mathbb{R}_*^d$, we have $\int \varphi^*(\varphi'(\frac{g(x)f_b(b^\top x)}{f(x)g_b(b^\top x)}))f(x)dx = \int (\frac{g(x)f_b(b^\top x)}{f(x)g_b(b^\top x)} - 1)f(x)dx = 0$, since, for any b in $\mathbb{R}_*^d$, the function $x \mapsto \frac{g(x)f_b(b^\top x)}{g_b(b^\top x)}$ is a density. The complement of Θ^{D_ϕ} in $\mathbb{R}_*^d$ is $\emptyset$ and then the supremum looked for in $\mathbb{R}$ is $-\infty$. We can therefore conclude. It is interesting to note that we obtain the same verification with f , $g^{(k-1)}$ and a_k .

E.2. Discussion of (A3).

This hypothesis consists in the following assumptions:

(0) We work with the Kullback-Leibler divergence,

(1) We have $f(\cdot/a_1^\top x) = g(\cdot/a_1^\top x)$, i.e., $K(g_{\frac{f_1}{g_1}}, f) = 0$ —we could also derive the same proof with f , $g^{(k-1)}$ and a_k

Preliminary (A): Shows that $A = \{(c, x) \in \mathbb{R}_^d \setminus \{a_1\} \times \mathbb{R}^d; \frac{f_{a_1}(a_1^\top x)}{g_{a_1}(a_1^\top x)} > \frac{f_c(c^\top x)}{g_c(c^\top x)}, g(x) \frac{f_c(c^\top x)}{g_c(c^\top x)} > f(x)\} = \emptyset$ through a reductio ad absurdum, i.e., if we assume $A \neq \emptyset$.*

Thus, our hypothesis enables us to derive

$f(x) = f(\cdot/a_1^\top x)f_{a_1}(a_1^\top x) = g(\cdot/a_1^\top x)f_{a_1}(a_1^\top x) > g(\cdot/c^\top x)f_c(c^\top x) > f$
 since $\frac{f_{a_1}(a_1^\top x)}{g_{a_1}(a_1^\top x)} \geq \frac{f_c(c^\top x)}{g_c(c^\top x)}$ implies $g(\cdot/a_1^\top x)f_{a_1}(a_1^\top x) = g(x)\frac{f_{a_1}(a_1^\top x)}{g_{a_1}(a_1^\top x)} \geq g(x)\frac{f_c(c^\top x)}{g_c(c^\top x)} = g(\cdot/c^\top x)f_c(c^\top x)$,
 i.e., $f > f$. We can thus conclude.

Preliminary (B): Shows that $B = \{(c, x) \in \mathbb{R}_^d \setminus \{a_1\} \times \mathbb{R}^d; \frac{f_{a_1}(a_1^\top x)}{g_{a_1}(a_1^\top x)} < \frac{f_c(c^\top x)}{g_c(c^\top x)}, g(x) \frac{f_c(c^\top x)}{g_c(c^\top x)} < f(x)\} = \emptyset$ through a reductio ad absurdum, i.e., if we assume $B \neq \emptyset$.*

Thus, our hypothesis enables us to derive

$$f(x) = f(\cdot/a_1^\top x)f_{a_1}(a_1^\top x) = g(\cdot/a_1^\top x)f_{a_1}(a_1^\top x) < g(\cdot/c^\top x)f_c(c^\top x) < f$$

We can consequently conclude as above.

Let us now verify (A3):

We have $PM(c, a_1) - PM(c, a) = \int \ln\left(\frac{g(x)f_c(c^\top x)}{g_c(c^\top x)f(x)}\right) \left\{\frac{f_{a_1}(a_1^\top x)}{g_{a_1}(a_1^\top x)} - \frac{f_c(c^\top x)}{g_c(c^\top x)}\right\} g(x) dx$. Moreover, the logarithm $\ln$ is negative on $\{x \in \mathbb{R}_*^d; \frac{g(x)f_c(c^\top x)}{g_c(c^\top x)f(x)} < 1\}$ and is positive on $\{x \in \mathbb{R}_*^d; \frac{g(x)f_c(c^\top x)}{g_c(c^\top x)f(x)} \geq 1\}$.

Thus, the preliminary studies (A) and (B) show that $\ln\left(\frac{g(x)f_c(c^\top x)}{g_c(c^\top x)f(x)}\right)$ and $\left\{\frac{f_{a_1}(a_1^\top x)}{g_{a_1}(a_1^\top x)} - \frac{f_c(c^\top x)}{g_c(c^\top x)}\right\}$ always present a negative product. We can therefore conclude, since $(c, a) \mapsto PM(c, a_1) - PM(c, a)$ is not null for all c and for all a , with $a \neq a_1$.

References

1. Sklar, M. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Stat. Univ.* **1959**, *8*, 229–231.
2. Joe, H. *Multivariate Models and Dependence Concepts*. *Monographs on Statistics and Applied Probability*, 1st ed.; Chapman and Hall/CRC: London, UK, 1997.
3. Nelsen, R.B. *An introduction to Copulas*. *Springer Series in Statistics*, 2nd ed.; Springer: New York, NY, USA, 2006.
4. Carriere, J.F. A large sample test for one-parameter families of Copulas. *Comm. Stat. Theor. Meth.* **1994**, *23*, 1311–1317.
5. Genest, C.; Rémillard, B. Tests of independence and randomness based on the empirical Copula process. *Test* **2004**, *13*, 335–370.
6. Fermanian, J.D. Goodness of fit tests for copulas. *J. Multivariate Anal.* **2005**, *95*, 119–152.
7. Genest C.; Quessy, J.F.; Rémillard, B. Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scand. J. Stat.* **2006**, *33*, 337–366.
8. Michiels, F.; De Schepper, A. A Copula Test Space Model—How to Avoid the Wrong Copula Choice. *Kybernetika* **2008**, *44*, 864–878.

9. Genest, C. Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resour. Res.* **2009**, *43*, W09401:1–W09401:12.
10. Mesfioui, M.; Quessy, J.F.; Toupin, M.H. On a new goodness-of-fit process for families of copulas. *La Revue Canadienne de Statistique* **2009**, *37*, 80–101.
11. Genest, C.; Rémillard, B. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Math. Econ.* **2009**, *44*, 199–213.
12. Berg, D. Copula goodness-of-fit testing: An overview and power comparison. *Eur. J. Finance* **2009**, *15*, 675–701.
13. Bücher, A.; Dette, H. Some comments on goodness-of-fit tests for the parametric form of the copula based on L^2 -distances. *J. Multivar. Anal.* **2010**, *101*, 749–763.
14. Broniatowski, M.; Leorato, S. An estimation method for the Neyman chi-square divergence with application to test of hypotheses. *J. Multivar. Anal.* **2006**, *97*, 1409–1436.
15. Friedman, J.H.; Stuetzle, W.; Schroeder, A. Projection pursuit density estimation. *J. Am. Statist. Assoc.* **1984**, *79*, 599–608.
16. Huber, P.J. Projection pursuit. *Ann. Stat.* **1985**, *13*, 435–525.
17. Cambanis, S.; Huang, S.; Simons, G. On the theory of elliptically contoured distributions. *J. Multivar. Anal.* **1981**, *11*, 368–385.
18. Landsman, Z.M.; Valdez, E.A. Tail conditional expectations for elliptical distributions. *N. Am. Actuar. J.* **2003**, *7*, 55–71.
19. Yohai, V.J. Optimal robust estimates using the Kullback-Leibler divergence. *Stat. Probab. Lett.* **2008**, *78*, 1811–1816.
20. Toma, A. Optimal robust M-estimators using divergences. *Stat. Probab. Lett.* **2009**, *79*, pp. 1–5.
21. Huber, P.J. *Robust Statistics*; Wiley: New York, NY, USA, 1981 (republished in paperback, 2004).
22. van der Vaart, A.W. *Asymptotic statistics. Cambridge Series in Statistical and Probabilistic Mathematics*; Cambridge University Press: Cambridge, UK, 1998.
23. Scott, D.W. *Multivariate Density Estimation. Theory, Practice, and Visualization. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*; A Wiley-Interscience Publication. John Wiley and Sons, Inc.: New York, NY, USA, 1992.
24. Touboul, J. Goodness-of-fit Tests For Elliptical And Independent Copulas Through Projection Pursuit. *arXiv. Statistics Theory*, 2011, arXiv: 1103.0498.
25. Aurélien, A.; Damiano, B. New families of Copulas based on periodic functions. *Commun. Stat. Theor. Meth.* **2005**, *34*, 1437–1447.
26. Cressie, N.; Read, T.R.C. Multinomial goodness-of-fit tests. *J. R. Stat. Soc. Series B* **1984**, *46*, 440–464.
27. Csizsár, I. On topology properties of f -divergences. *Studia Sci. Math. Hungar.* **1967**, *2*, 329–339.
28. Liese, F.; Vajda, I. *Convex Statistical Distances*; BSB B. G. Teubner Verlagsgesellschaft: Leipzig, Germany, 1987.
29. Pardo, L. *Statistical Inference Based on Divergence Measures. Statistics: Textbooks and Monographs*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2006.

30. Zografos, K.; Ferentinos, K.; Papaioannou, T. φ -divergence statistics: sampling properties and multinomial goodness of fit and divergence tests. *Commun. Stat. Theor. Meth.* **1990**, *19*, 1785–1802.
31. Azé, D. *Eléments D'analyse Convexe et Variationnelle, Ellipses*; Dunod: Paris, French, 1997.
32. Bosq, D.; Lecoutre, J.P. *Livre-Theorie De L'Estimation Fonctionnelle*; Economica; 1999.
33. Broniatowski, M.; Keziou, A. Parametric estimation and tests through divergences and the duality technique. *J. Multivar. Anal.* **2009**, *100*, 16–36.
34. Black and Scholes. The pricing of options and corporate liabilities. *J. Polit. Econ.* **1973**, *81*, 635–654.

© 2011 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>.)