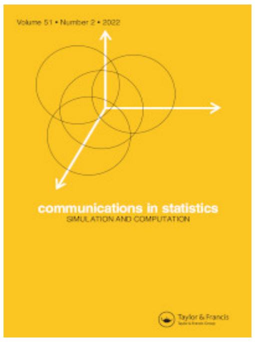

Communications in Statistics - Simulation and Computation

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lssp20

Testing bivariate independence based on α-divergence by improved probit transformation method for copula density estimation

M. Mohammadi, M. Emadi & M. Amini

COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
https://doi.org/10.1080/03610918.2022.2025836

Testing bivariate independence based on α-divergence by improved probit transformation method for copula density estimation

M. Mohammadi[a,b], M. Emadi[a], and M. Amini[c]

[a]Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran; [b]Department of Statistics, University of Zabol, Zabol, Iran; [c]Department of Statistics, Ordered Data, Reliability and Dependency Center of Excellence, Ferdowsi University of Mashhad, Mashhad, Iran

ABSTRACT
Independence test based on empirical copula does not perform well in the presence of weak dependency or when dependency occurs only in the tails. The copula density, which is estimated by a local likelihood probit transformation method, is used to detect the independence. In this article, three nonparametric tests of independence based on α-divergence and copula density are introduced. These tests are capable of considering weak dependency. The asymptotic consistency of the copula-based α-divergence estimator is also derived. In addition, the empirical powers of the proposed tests are computed through extensive simulations. The results show that the new tests outperform in small sample sizes or weak dependencies. Finally, an application in uranium exploration is presented to illustrate the applicability of the proposed tests.

ARTICLE HISTORY
Received 7 April 2021
Accepted 2 January 2022

KEYWORDS
α-divergence; copula density; probit transformation; test of independence

2010 MATHEMATICS SUBJECT CLASSIFICATION
62G10; 62G07; 62H15; 62H20

1. Introduction

Testing independence between two random variables is momentous in statistics and other fields of science. Most classical testing independence was initially based on some measures of dependence such as the Pearson linear correlation, Kendall's τ (τ), and Spearman's ρ. Blum, Kiefer, and Rosenblatt (1961) showed that these tests are usually inconsistent, meaning that their power functions do not tend to one as the sample size under certain alternatives tends to infinity. To solve this issue, they used the Cramér-von Mises (CvM) distance to compare the joint empirical distribution function with the product of its corresponding marginal empirical distributions.

Copulas are a useful tool to model multivariate distributions. Sklar (1959) was the first to introduce the fundamental concept of the copula. The joint distribution H can then be represented by using copula function C as

$$F(x, y) = C(F_X(x), G_Y(y)), \qquad x, y \in \mathbb{R}, \tag{1}$$

where F_X and F_Y are the marginal distributions of X and Y, respectively. A bivariate copula function C is a cumulative distribution function of random vector (U, V), defined on the unit square $[0, 1]^2$, with uniform marginal distributions as $U = F(X)$ and $V = G(Y)$.

If C is an absolutely continuous copula distribution on $[0, 1]^2$, then its density function is $c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}$. The distribution of independence copula is defined as $\Pi(u, v) = uv$ and so, the density of independence copula is $\pi(u, v) = 1$. As a result, the relationship between the copula density c and the joint density function f of (X, Y) according to Eq. (1) can be represented as

CONTACT M. Emadi emadi@um.ac.ir Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran.

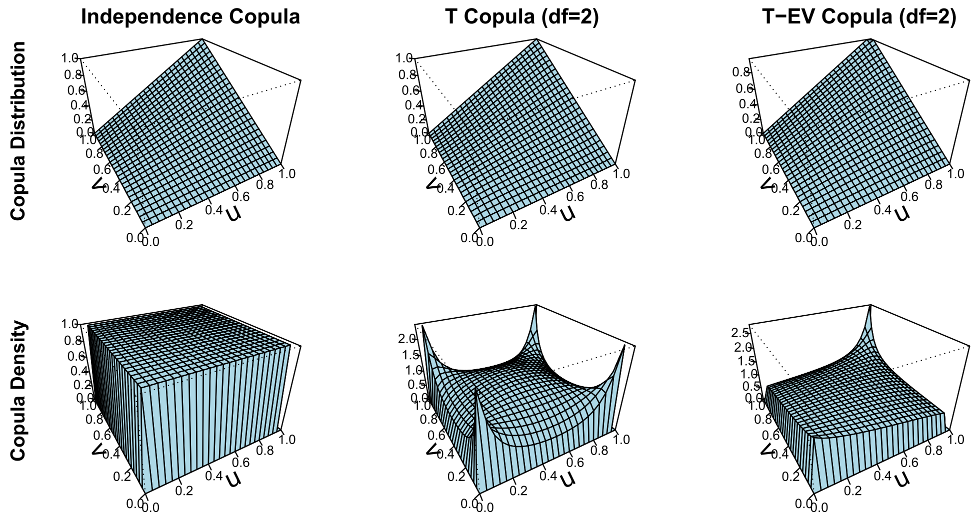

Figure 1. Copula distribution and density of independent, T, and T-EV copulas.

$$f(x, y) = c(F(x), G(y))f_X(x)f_Y(y), \qquad (x, y) \in \mathbb{R}^2, \tag{2}$$

where f_X and f_Y are the marginal density function of X and Y, respectively.

The null hypothesis of copula-based independence test can be expressed as $H_0 : C(u, v) = uv$. Genest and Rémillard (2004) suggested the test statistic

$$S_n = n\int_{[0,1]^2} (C_n(u, v) - uv)^2 dC_n(u, v), \tag{3}$$

for testing independence based on the CvM distance, where C_n is the empirical copula that was initially introduced by Deheuvels (1979). Empirical copula is defined as $C_n(u, v) = \frac{1}{n}\sum_{i=1}^{n} I\{\hat{U}_i \leq u, \hat{V}_i \leq v\}$, where $\hat{U}_i = \frac{n\hat{F}_X(x_i)}{n+1}$ and $\hat{V}_i = \frac{n\hat{F}_Y(y_i)}{n+1}$, $i = 1, \ldots, n$, are the pseudo observations in which $\hat{F}_X$ and $\hat{F}_Y$ are the empirical cumulative distribution function of the observation X_i and Y_i, respectively.

Student-T (T) copula is a useful distribution in the elliptical copula class. Demarta and McNeil (2007) showed that the lower and upper tail dependence coefficients for this copula with parameter ρ and ν degree of freedom are equal to $2t_{\nu+1}\left(-\sqrt{\nu+1}\sqrt{1-\rho}/\sqrt{1+\rho}\right)$, where t_ν^{-1} is the inverse of the univariate Student's-T distribution with ν degrees of freedom and $\tau = \frac{2}{\pi}\arcsin(\rho)$. The lower and upper tail dependence for T copula with zero Kendall's τ and 2 and 5 degrees of freedom are equal to 0.182 and 0.05, respectively. Thus, in these cases, dependency occurs only at the tails. Also, dependency occurs only at the upper tail when the data are produced from Student-T-Extreme Value (T-EV) copula with zero Kendall's τ and low degree of freedom. Therefore, in T and T-EV copulas with a small degree of freedom, zero Kendall's τ does not necessarily imply independence because dependency may occur in the tails.

Belalia et al. (2017) showed that the test of independence based on empirical copula fails in term of power when dependency occurs only in the tails. Figure 1 demonstrates the motivation for using copula density in the construction of new tests, which shows that the copula density is flexible in detecting the independence between the variables of interest. This figure compares the distribution (top row) and density (bottom row) of independent, T($\nu = 2$), and T-EV($\nu = 2$) copulas with zero Kendall's τ, respectively. It can be seen that there is no difference between the independent copula distribution function and the copula distribution functions of T and T-EV copulas. Whereas, the shape of the copula densities changes with respect to the type of dependencies in T and T-EV copulas compared with independence copula density. Thus, the density of copula is appropriate for the detection of independence and so this result inspired us to use copula density instead of the empirical copula to test of independence.

The estimation of the copula density is needed to perform the independence test based on divergence. A specific class of nonparametric copula density estimators is kernel estimators. Charpentier, Fermanian, and Scaillet (2006) and Nagler (2014, 2018) presented different approaches to nonparametric estimation of the copula density such as mirror-reflection method, beta kernel, and transformation technique. The *R* package **kdecopula** (Nagler and Wen (2018)) implements several bivariate kernel copula density estimators that have been proposed in recent years. In this article, the local likelihood probit transformation ($\mathcal{LLPT}$) technique is used to estimate the copula density suggested by Geenens, Charpentier, and Paindaveine (2017). A comprehensive simulation study shown by Nagler (2018) is that the $\mathcal{LLPT}$ method for copula density estimation is very good and easy to implement estimators, fixes boundary issues in a natural way, and is able to cope with unbounded copula densities.

Kullback–Leibler (KL) divergence was introduced by Kullback and Leibler (1951). By combining the KL divergence and copula density in Eq. (2), Blumentritt and Schmid (2012) considered a measure of bivariate association as $\int_{[0,1]^2} c(u,v)\log c(u,v)\,dudv$. Initially, Chernoff (1952) proposed the α-divergence, which is a generalization of the KL divergence. Eguchi and Kato (2010) showed that the α-divergence is a robust divergence with respect to outliers and consequently has a flexible performance. For some α-divergence investigations, see, for example, Cichocki and Amari (2010) and Read and Cressie (2012).

In this article, new nonparametric independence tests based on copula density and α-divergence measures are presented to tackle the issue mentioned above. These tests are simple to implement and also provide bigger power compared to the empirical copula-based test in the presence of weak dependencies. Also, these tests reduce the complexity of computation because those only depend on the copula density and there is no need to estimate the joint and the marginals density functions.

The rest of the article is arranged as follows. In Section 2, the estimation of copula density function using $\mathcal{LLPT}$ method is provided. The α-divergence based on the copula density with their basic properties is introduced in Section 3. In Section 4, the behavior of copula-based α-divergence for some copula functions is interpreted. Thereafter, the tests of independence based on α-divergence and copula density are defined, and the asymptotic distribution of new tests is achieved in Section 5. The simulation results are provided to compare the empirical size and power of independence tests based on α-divergence measure in Section 6. In the end, the performance of the considered tests for real data is evaluated.

2. Local likelihood probit transformation estimation

The transformation method for kernel copula density estimation was introduced by Charpentier, Fermanian, and Scaillet (2006). The simple idea is to transform the data so that it is supported on the full $\mathbb{R}^2$ (instead of the unit cube). On this transformed domain, standard kernel techniques can be used to estimate the density. An adequate back-transformation then yields an estimate of the copula density. The inverse of the standard normal cumulative distribution function is most commonly used for the transformation since it is known that kernel estimators tend to do well for Gaussian random variables.

Let $(U_i, V_i)_{i=1,\dots,n}$ be independent and identically distributed observations from the bivariate copula C and the purpose is to estimate the corresponding copula density function. Denote Φ as the standard Gaussian distribution and ϕ as its first order derivative. Then $(S_i, T_i) = (\Phi^{-1}(U_i), \Phi^{-1}(V_i))$ is a random vector with Gaussian margins and copula C. According to (2), the corresponding density function can be written as $f(s,t) = c(\Phi(s),\Phi(t))\phi(s)\phi(t)$. Thus, an estimation of the copula density function can be given by

$$\hat{c}_n^{(\mathcal{PT})}(u,v) = \frac{\hat{f}_n(\Phi^{-1}(u),\Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}, \qquad (u,v) \in (0,1)^2. \tag{4}$$

However, as (U_i, V_i)'s are unavailable and one has to use $(\hat{S}_i, \hat{T}_i) = (\Phi^{-1}(\hat{U}_i), \Phi^{-1}(\hat{V}_i))$ the pseudo-transformed sample, instead. As a first natural idea, the standard kernel density estimator for $\hat{f}_n$ in (4) can be considered as follows:

$$\hat{f}_n(s,t) = \frac{1}{n|\mathbf{H}_{ST}|^{\frac{1}{2}}} \sum_{i=1}^{n} \mathbf{K}\left(\mathbf{H}_{ST}^{-\frac{1}{2}}\begin{pmatrix} s-\hat{S}_i \\ t-\hat{T}_i \end{pmatrix}\right), \tag{5}$$

where $\mathbf{K}: \mathbb{R}^2 \to \mathbb{R}$ is a kernel function and $\mathbf{H}_{ST} = \begin{bmatrix} b_n & 0 \\ 0 & b_n \end{bmatrix}$ is a bandwidth matrix.

This kernel estimator has asymptotic problems at the edges of the distribution support. To remedy this problem, an $\mathcal{LLPT}$ method was recently suggested by Geenens, Charpentier, and Paindaveine (2017). Instead of applying the standard kernel estimator, they locally fit a polynomial to the log-density of the transformed sample. The motivation and the advantages of estimating $f(s, t)$ by local likelihood methods instead of raw kernel density estimation are related to the detailed discussion in Geenens (2014). The notations are similar to ones used in Geenens, Charpentier, and Paindaveine (2017).

Around $(s,t) \in \mathbb{R}^2$ and (s',t') close to (s,t), the local log-quadratic likelihood estimation of $\log f(s,t)$ from the pseudo-transformed sample is defined as follows:

$$\begin{aligned}\log f(s',t') &= a_{2,0}(s,t) + a_{2,1}(s,t)(s'-s) + a_{2,2}(s,t)(t'-t) \\ &\quad + a_{2,3}(s,t)(s'-s)^2 + a_{2,4}(s,t)(t'-t)^2 + a_{2,5}(s,t)(s'-s)(t'-t) \\ &\equiv P_{a_2}(s'-s, t'-t).\end{aligned}$$

The vector $a_2(s,t) \equiv (a_{2,0}(s,t), \ldots, a_{2,5}(s,t))$ is then estimated by solving a weighted maximum likelihood problem as

$$\begin{aligned}\hat{a}_2(s,t) = \arg\max_{a_2} \Bigg\{ &\sum_{i=1}^{n} \mathbf{K}\left(\mathbf{H}_{ST}^{-\frac{1}{2}}\begin{pmatrix} s-\hat{S}_i \\ t-\hat{T}_i \end{pmatrix}\right) P_{a_2}(\hat{S}_i - s, \hat{T}_i - t) \\ &- n\int_{\mathbb{R}^2} \mathbf{K}\left(\mathbf{H}_{ST}^{-\frac{1}{2}}\begin{pmatrix} s-s' \\ t-t' \end{pmatrix}\right) \exp\left(P_{a_2}(s'-s, t'-t)\right) ds'dt' \Bigg\}.\end{aligned}$$

Therefore, the estimation of $f(s, t)$ is $\tilde{f}^p(s,t) = \exp\{\hat{a}_2(s,t)\}$ and thus the $\mathcal{LLPT}$ estimator of a copula density is

$$\hat{c}_n^{(\mathcal{LLPT})}(u,v) = \frac{\tilde{f}^p(\Phi^{-1}(u),\Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}, \qquad (u,v) \in [0,1]^2. \tag{6}$$

When the underlying density is on $[0,1]^2$, the performance of the kernel estimator depends on the choice of the kernel function and the bandwidth (smoothing parameter). For bandwidth choice, a practical approach is to consider the minimization of the asymptotic mean integrated squared error (AMISE) on the level of the transformed data. In this article, the nearest-neighbor method is used for bandwidth choice such that smoothing parameters are selected based on univariate least-squares cross-validation on the first principal component in the transformed domain (see, Geenens, Charpentier, and Paindaveine (2017), Section 4).

The asymptotic normality of the copula density estimator based on local likelihood transformation was demonstrated by Geenens, Charpentier, and Paindaveine (2017) as

$$\hat{c}_n^{(\mathcal{LLPT})}(u,v) \text{ is } AN\left(\mu(u,v), \frac{\sigma^2(u,v)}{nb_n^2}\right),$$

where

$$\sigma^2(u,v) = \frac{5c(u,v)}{8\pi\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}, \tag{7}$$

$$\begin{aligned}\mu(u,v) = c(u,v) - \frac{b_n^4}{8}\frac{c(u,v)}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}\Bigg\{\frac{\partial^4 g}{\partial x^4} + \frac{\partial^4 g}{\partial y^4} + 2\frac{\partial^4 g}{\partial x^2\partial y^2} \\ + 4\left(\frac{\partial^3 g}{\partial x^3}\frac{\partial g}{\partial x} + \frac{\partial^3 g}{\partial y^3}\frac{\partial g}{\partial y} + \frac{\partial^3 g}{\partial x^2\partial y}\frac{\partial g}{\partial y} + \frac{\partial^3 g}{\partial x\partial y^2}\frac{\partial g}{\partial x}\right)\Bigg\}(x,y),\end{aligned} \tag{8}$$

and $x = \Phi^{-1}(u)$, $y = \Phi^{-1}(v)$, and $g(x,y) = \log c(\Phi(x),\Phi(y)) + \log\phi(x) + \log\phi(y)$.

3. α-Divergence based on copula density

The α-divergence was defined by Rényi (1961) for $a>0$ with $a\neq 1$ and by Liese and Vajda (1987) for $a<0$. This divergence measure can be derived from the Csiszár φ-divergence if $\varphi(u) = \frac{u^\alpha - \alpha(u-1)-1}{\alpha(\alpha-1)}, u\geq 0, \alpha\neq 0,1$. Following Cichocki and Amari (2010), the asymmetric α-divergence between two probability density functions f_1 and f_2 of a continuous random variable can be defined as

$$\begin{aligned}\mathcal{AD}_\alpha(f_1\,\|\,f_2) &= \frac{1}{\alpha(\alpha-1)}\int_{\mathbb{R}}\left(f_1^\alpha(x)f_2^{1-\alpha}(x) - \alpha f_1(x) + (\alpha-1)f_2(x)\right)dx \\ &= \frac{1}{\alpha(\alpha-1)}\left(\int_{\mathbb{R}} f_1^\alpha(x)f_2^{1-\alpha}(x)dx - 1\right), \qquad \alpha\in\mathbb{R}\backslash\{0,1\}.\end{aligned} \tag{9}$$

If $\alpha\to 1$, then the KL divergence and in special cases for $\alpha = 2, 0.5$, the well-known Neyman Chi-square (χ_N^2) divergence and Hellinger (He) distance can be obtained from Eq. (9).

By using Eq. (2), the copula-based α-divergence between joint density function h and marginal density functions f and g for X and Y, respectively, can be represented as

$$\begin{aligned}\mathcal{AD}_\alpha(c) &\equiv \mathcal{AD}_\alpha(h\,\|\,fg) \\ &= \frac{1}{\alpha(\alpha-1)}\left(\int_{\mathbb{R}^2} h(x,y)^\alpha (f(x)g(y))^{1-\alpha}dxdy - 1\right) \\ &= \frac{1}{\alpha(\alpha-1)}\left(\int_{\mathbb{R}^2}\left(\frac{h(x,y)}{f(x)g(y)}\right)^\alpha f(x)g(y)dxdy - 1\right) \\ &= \frac{1}{\alpha(\alpha-1)}\left(\int_{[0,1]^2} c^\alpha(u,v)dudv - 1\right), \qquad \alpha\in\mathbb{R}\backslash\{0,1\}.\end{aligned} \tag{10}$$

This representation of α-divergence reduces the complexity because it depends only on the copula density. In Proposition 1, some theoretical aspects of this measure of the association will be reviewed.

Proposition 1. *Let C denote the copula distribution of (X,Y) and let c be the corresponding copula density. Then the following assertions hold:*

a. *$\mathcal{AD}_\alpha(c)\geq 0$ and $\mathcal{AD}_\alpha(c) = 0$ if and only if X and Y are independent.*
b. *$\mathcal{AD}_\alpha(c) < \infty$ for $\alpha > 1$ and $\mathcal{AD}_\alpha(c)\leq -\frac{1}{\alpha(\alpha-1)}$ for $\alpha < 1$.*

c. *$\mathcal{AD}_\alpha(c)$ is invariant under strictly monotone transformations of one or two components of (X, Y).*

Proof.

a. Using Jensen's inequality, we get

$$\begin{aligned}
\mathcal{AD}_\alpha(c) &= \frac{1}{\alpha(\alpha-1)}\left(\int_{[0,1]^2} c^\alpha(u,v)dudv - 1\right)\\
&= \frac{1}{\alpha(\alpha-1)}\left(E\left(c^{\alpha-1}(u,v)\right) - 1\right)\\
&\geq \frac{1}{\alpha(\alpha-1)}\left(E^{1-\alpha}\left(\frac{1}{c(u,v)}\right) - 1\right)\\
&= 0,
\end{aligned}$$

where $\alpha \in \mathbb{R}\backslash\{0, 1\}$. If X and Y are independent random variables, then $c(u, v) = 1$. Therefore $\mathcal{AD}_\alpha(c) = 0$.

b. It is observed that $\varphi''(u) = u^{\alpha-2}$, $\varphi(0) = \frac{1}{\alpha}$, and $\varphi(1) = 0$. The function φ is convex for $u > 0$, and strictly convex at $u = 1$. Thus, from Micheas and Zografos (2006), the upper bound of α-divergence is

$$\begin{aligned}
\gamma &= \varphi(0) - \varphi(1) + \lim_{u\to+\infty}\frac{\varphi(u)}{u}\\
&= \frac{1}{\alpha} + \lim_{u\to+\infty}\frac{u^\alpha - \alpha(u-1) - 1}{u\alpha(\alpha-1)}\\
&= \frac{1}{\alpha(\alpha-1)}\left(\lim_{u\to+\infty} u^{\alpha-1} - 1\right).
\end{aligned}$$

Hence, $\gamma = +\infty$ for $a > 1$ and also, $\gamma = -\frac{1}{\alpha(\alpha-1)}$ for $a < 1$. Note that the measure satisfies the important axioms *(A3)–(A5)* in Micheas and Zografos (2006), when $a < 1$, and axioms *(A3)* and *(A4)*, when $a > 1$. Also, we note that, from axiom *(A5)*, when $a < 1$, $\mathcal{AD}_\alpha(c) = -\frac{1}{\alpha(\alpha-1)}$ if and only if the random variables X and Y are completely dependent.

c. The basic invariance properties of copulas under strictly monotone transformations of random variables can be found in Nelsen (2007). Under increasing transformations of the random variables, the copula is invariant. Thus the copula density and $\mathcal{AD}^\alpha(c)$ are also invariant. If δ and η are strictly decreasing transformations, then $C_{\delta(X),Y}(u, v) = v - C_{X,Y}(1-u, v)$ and $C_{\delta(X),\eta(Y)}(u, v) = u + v - 1 - C_{X,Y}(1-u, 1-v)$. For the copula density, we can get $c_{\delta(X),Y}(u, v) = c_{X,Y}(1-u, v)$ and $c_{\delta(X),\eta(Y)}(u, v) = c_{X,Y}(1-u, 1-v)$. Thus,

$$\begin{aligned}
\mathcal{AD}_\alpha(c)_{(\delta(X),Y)} &= \frac{1}{\alpha(\alpha-1)}\left(\int_{[0,1]^2} c^\alpha(1-u,v)dudv - 1\right)\\
(\text{substitution of } 1-\text{u} = z) &= \frac{1}{\alpha(\alpha-1)}\left(\int_{[0,1]^2} c^\alpha(z,v)dzdv - 1\right)\\
&= \mathcal{AD}_\alpha(c)_{(X,Y)}.
\end{aligned}$$

In the same way,

$$\mathcal{AD}_\alpha(c)_{(\delta(X),\eta(Y))} = \frac{1}{\alpha(\alpha-1)}\left(\int_{[0,1]^2} c^\alpha(1-u,1-v)dudv - 1\right),$$

and by replacing $1-u=z,\ 1-v=w,$ we get

$$\begin{aligned}\mathcal{AD}_\alpha(c)_{(\delta(X),\eta(Y))} &= \frac{1}{\alpha(\alpha-1)}\left(\int_{[0,1]^2} c^\alpha(z,w)dzdw - 1\right)\\ &= \mathcal{AD}_\alpha(c)_{(X,Y)}.\end{aligned}$$

□

Now, we will focus on representing the special cases of $\mathcal{AD}_\alpha(c)$ that can be considered as a measure of association. If $\alpha \to 1,$ then the KL divergence is obtained of the form

$$\begin{aligned}KL(c) &= \lim_{\alpha\to 1}\mathcal{AD}_\alpha(c)\\ &= \int_{[0,1]^2} c(u,v)\log(c(u,v))\,dudv \\ &= E\big(\log(c(U,V))\big).\end{aligned} \tag{11}$$

In special cases when $\alpha=2$, the copula-based Neyman Chi-square (χ^2_N) divergence by using the Eq. (2) can be represent as

$$\begin{aligned}\chi^2_N(c) &= \mathcal{AD}_2(c)\\ &= \frac{1}{2}\int_{[0,1]^2}(c(u,v)-1)^2\,dudv\\ &= \frac{1}{2}E(c(U,V)) - \frac{1}{2}.\end{aligned} \tag{12}$$

Also, in a special case for $\alpha = 1/2,$ the copula-based Hellinger (He) distance can be rewritten as follows:

$$\begin{aligned}He(c) &= \frac{1}{4}\mathcal{AD}_{1/2}(c)\\ &= \frac{1}{2}\int_{[0,1]^2}(\sqrt{c(u,v)}-1)^2\,dudv\\ &= 1 - E\left(\frac{1}{\sqrt{c(U,V)}}\right).\end{aligned} \tag{13}$$

These three divergences separate the dependence structure from the marginal distributions and are only based on copula density. Now, the properties of copula-based α-divergence for some copulas will be illustrated. The behavior of $\mathcal{AD}_\alpha(c)$ for the Gaussian copula is described in Section 3.1, and the behavior of $\mathcal{AD}_\alpha(c)$ for T, T-EV, Clayton, and Gumbel–Hougaard (Gumbel) copulas is discussed in Section 3.2.

3.1. $\mathcal{AD}_\alpha(c)$ for the Gaussian copula

In practice, the Gaussian (normal) copula is the most important case in quantitative finance applications. The bivariate Gaussian copula with parameter ρ is defined as

$$C_{\mathcal{G}}(u,v,\rho) = \Phi_2(\Phi^{-1}(u),\Phi^{-1}(v);\rho), \qquad (u,v)\in[0,1]^2,\ \rho\in[-1,1], \tag{14}$$

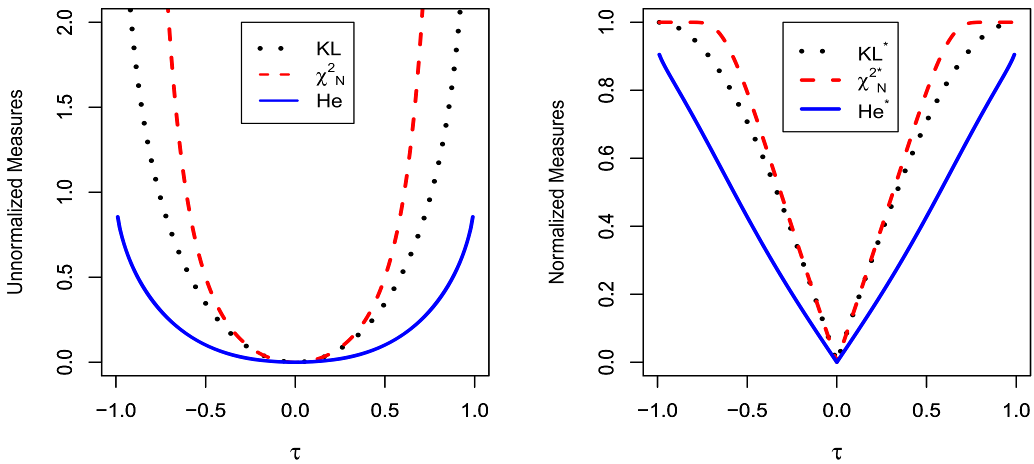

Figure 2. Normalize and unnormalize measure of α-divergence for Gaussian copula.

where Φ_2 is the bivariate normal distribution function with zero means, variances one, and correlation matrix $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, and Φ^{-1} denotes the quantile function of the standard normal distribution. The density of the Gaussian copula is determined by

$$c_{\mathcal{G}}(u,v;\rho)=\frac{1}{\sqrt{1-\rho^2}}\exp\left\{\frac{2\rho\Phi^{-1}(u)\Phi^{-1}(v)-\rho^2(\Phi^{-1}(u)^2+\Phi^{-1}(v)^2)}{2(1-\rho^2)}\right\},\ (u,v)\in[0,1]^2, \tag{15}$$

where $\rho\in[-1,1]$.

The general form of the α-divergence for multivariate normal distribution was obtained by Micheas and Zografos (2006). Thus, by using part (c) of Proposition 1 and Eq. (10), the α-divergence for bivariate Gaussian copula can be shown as follows:

$$\mathcal{AD}_\alpha(c_{\mathcal{G}})=\frac{1}{\alpha(\alpha-1)}\left(\frac{|1-\rho^2|^{\frac{1-\alpha}{2}}}{(1-(1-\alpha)^2\rho^2)^{\frac{1}{2}}}-1\right),\qquad \alpha\in\mathbb{R}\backslash\{0,1\},\ \rho\in[-1,1]. \tag{16}$$

So, it is easy to see that

$$KL(c_{\mathcal{G}})=-\log(1-\rho^2)/2,$$
$$\chi^2_N(c_{\mathcal{G}})=\frac{\rho^2}{2(1-\rho^2)},$$
$$He(c_{\mathcal{G}})=1-2(1-\rho^2)^{\frac{1}{4}}/(4-\rho^2)^{\frac{1}{2}}.$$

Remark (Normalization). As shown in Eq. (16), the range of $\mathcal{AD}_\alpha(c)$ is $[0,\infty]$, and therefore, normalization to $[0,1]$ is desirable. Thus, a suitable transformation $T:[0,\infty]\to[0,1]$ is required, which is continuous and strictly increasing and satisfies $T(0)=0$ and $T(\infty)=1$. The normalization transformation $T(x)=\sqrt{1-e^{-2x}}$, denoted by $\mathcal{AD}^*_\alpha(c)$, is used for this purpose, which was suggested by Joe (1989).

The behavior of the normalize and unnormalize form of the special cases of the α-divergence for bivariate Gaussian copula versus different values of Kendall's τ is shown in Figure 2, where $\tau=\frac{2}{\pi}\arcsin(\rho)$ in this copula. Note that $\mathcal{AD}_\alpha(c_{\mathcal{G}})$ and $\mathcal{AD}^*_\alpha(c_{\mathcal{G}})$ take their minimum values (zero) in $\tau=0$.

Table 1. Some bivariate Archimedean copulas.

Copula	$C(u, v)$	Parameter space	Kendall's τ
Clayton	$(u^{-\theta}+v^{-\theta}-1)^{-1/\theta}$	$(0, +\infty)$	$\frac{\theta}{\theta+2}$
Gumbel	$\exp\{-[(-\ln u)^{\theta}+(-\ln v)^{\theta}]^{1/\theta}\}$	$[1, +\infty)$	$\frac{\theta-1}{\theta}$

3.2. $\mathcal{AD}_\alpha(c)$ for T, T-EV, Clayton and Gumbel copulas

In this Subsection, the copula-based α-divergence measure for T, T-EV, Clayton, and Gumbel copulas is considered. Moreover, $\mathcal{AD}_\alpha(c)$ for these copulas does not have a closed form and should be computed numerically. For more details on T, T-EV, Clayton, and Gumbel copulas, see Joe (2014).

The bivariate T copula with parameter ρ and ν degrees of freedom takes on the form

$$C(u, v; \rho, \nu) = t_{2,\nu}(t_\nu^{-1}(u), t_\nu^{-1}(v); \rho),\ (u, v) \in [0, 1]^2,\ \rho \in [-1, 1],\ \nu \geq 1,$$

where t_ν^{-1} is the inverse of the univariate Student's-T distribution with ν degrees of freedom.

The bivariate T-EV copula with parameter ρ and ν degrees of freedom is

$$C(u, ; \rho, \nu) = \exp\left\{-(x+y)B\left(\frac{x}{x+y}; \rho, \nu\right)\right\};\ (u, v) \in [0, 1]^2,\ \rho \in [-1, 1],\ \nu > 0,$$

with $x = \log(u)$, $y = \log(v)$, and

$$B(w; \rho, \nu) = w\, t_{\nu+1}\left(\frac{\sqrt{\nu+1}}{\sqrt{1-\rho^2}}\left[\left(\frac{w}{1-w}\right)^{\frac{1}{\nu}} - \rho\right]\right) + (1-w)\, t_{\nu+1}\left(\frac{\sqrt{\nu+1}}{\sqrt{1-\rho^2}}\left[\left(\frac{1-w}{w}\right)^{\frac{1}{\nu}} - \rho\right]\right),$$

where t_ν is the univariate distribution function of the Student-T distribution with ν degrees of freedom.

The Clayton and Gumbel copulas belong to the class of Archimedean copulas. Table 1 presents a summary information for Clayton and Gumbel copulas such as the parameter space and Kendall's τ of them.

The special case of normalized copula-based α-divergence values $\mathcal{AD}^*_\alpha(c)$ for T($\nu = 2$), T-EV($\nu = 2$), Clayton, and Gumbel copulas versus different values of τ and sample size $n = 200$ is obtained numerically and is reported in Table 2. In this table, it is visible that when the Kendall's τ increases, the normalized α-divergence values also increase. As a result, $KL^*(c)$, $\chi^{2*}_N(c)$, and $He^*(c)$ can be considered as new measures of association.

4. Estimation of α-divergence

A nonparametric estimation of $\mathcal{AD}_\alpha(c)$ based on the $\mathcal{LLPT}$ method to estimate the copula density can be written as

$$\begin{aligned}\widehat{\mathcal{AD}}_\alpha(c) &= \frac{1}{\alpha(\alpha-1)}\left(\int_{[0,1]^2} \hat{c}_n^{(\mathcal{LLPT})}(u, v)^{\alpha-1} dC_n(u, v) - 1\right)\\ &= \frac{1}{\alpha(\alpha-1)}\left(\frac{1}{n}\sum_{i=1}^{n} \hat{c}_n^{(\mathcal{LLPT})}(U_i, V_i)^{\alpha-1} - 1\right), \qquad \alpha \in \mathbb{R}\backslash\{0, 1\}.\end{aligned} \tag{17}$$

In Proposition 2, we show the asymptotic first-order consistency of α-divergences estimator in Eq. (17) under conditions. The proof method is similar to the work of Ahmad and Lin (1976) in which the consistency of entropy estimation by the kernel method has been proved.

Proposition 2. *Let $(U_i, V_i)_{i=1,\ldots,n}$ be an i.i.d. sample from the copula C. Assume the following assumptions:*

Table 2. Special case of $\mathcal{AD}^*_\alpha(c)$ for some copulas.

		τ								
Copula	Measure	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
T($\nu = 2$)	$KL^*(c)$	0.373	0.456	0.558	0.662	0.759	0.843	0.911	0.960	0.990
	$\chi^{2*}_N(c)$	0.497	0.602	0.722	0.832	0.917	0.970	0.994	1.000	1.000
	$He^*(c)$	0.171	0.220	0.286	0.362	0.444	0.528	0.616	0.705	0.801
T-EV($\nu = 2$)	$KL^*(c)$	0.346	0.481	0.640	0.747	0.827	0.891	0.938	0.974	0.993
	$\chi^{2*}_N(c)$	0.418	0.593	0.854	0.907	0.974	1.000	1.000	1.000	1.000
	$He^*(c)$	0.176	0.241	0.348	0.450	0.511	0.576	0.638	0.749	0.815
Clayton	$KL^*(c)$	0.179	0.355	0.506	0.637	0.748	0.838	0.909	0.960	0.990
	$\chi^{2*}_N(c)$	0.215	0.477	0.726	0.884	0.962	0.992	0.999	1.000	1.000
	$He^*(c)$	0.080	0.168	0.256	0.348	0.440	0.529	0.622	0.718	0.811
Gumbel	$KL^*(c)$	0.201	0.362	0.502	0.627	0.735	0.827	0.901	0.956	0.989
	$\chi^{2*}_N(c)$	0.222	0.439	0.624	0.772	0.882	0.953	0.989	0.999	1.000
	$He^*(c)$	0.106	0.189	0.269	0.352	0.438	0.530	0.624	0.720	0.813

(i) $b_n \to 0$, *as* $n \to \infty$;
(ii) $nb_n^2 \to \infty$, *as* $n \to \infty$;
(iii) $\mu(u, v) < \infty$;
(iv) $\int_{[0,1]^2} c^{2\alpha-1}(u, v)dudv < \infty$.

Then, for $\alpha \in \mathbb{R}\backslash\{0, 1\}$,

$$E\left|\widehat{\mathcal{AD}}_\alpha(c) - \mathcal{AD}_\alpha(c)\right| \to 0, \qquad n \to \infty.$$

Proof. We will write $\hat{c}(u, v)$ for $\hat{c}_n^{(\mathcal{LLPT})}(u, v)$, $\mathcal{AD}_\alpha$ for $\alpha(\alpha-1)\mathcal{AD}_\alpha(c) + 1$, and $\widehat{\mathcal{AD}}_\alpha$ for $\alpha(\alpha - 1)\widehat{\mathcal{AD}}_\alpha(c) + 1$ on $I = [0, 1]$. Define

$$\mathcal{AD}_\alpha = \int_{I^2} c^{\alpha-1}(u, v)dC(u, v) = E(c^{\alpha-1}(U, V)), \tag{18}$$

$$\widehat{\mathcal{AD}}_\alpha = \int_{I^2} \hat{c}^{\alpha-1}(u, v)dC_n(u, v) = \frac{1}{n}\sum_{i=1}^{n} \hat{c}^{\alpha-1}(\hat{U}_i, \hat{V}_i), \tag{19}$$

$$L_\alpha = \int_{I^2} (E(\hat{c}(u, v)))^{\alpha-1}dC_n(u, v) = \frac{1}{n}\sum_{i=1}^{n} \left(E(\hat{c}(\hat{U}_i, \hat{V}_i))\right)^{\alpha-1}, \tag{20}$$

$$M_\alpha = \int_{I^2} c^{\alpha-1}(u, v)dC_n(u, v) = \frac{1}{n}\sum_{i=1}^{n} c^{\alpha-1}(\hat{U}_i, \hat{V}_i). \tag{21}$$

Then by using the Minkowski inequality, we have

$$\begin{aligned} E\left|\widehat{\mathcal{AD}}_\alpha(c) - \mathcal{AD}_\alpha(c)\right| &= \frac{1}{|\alpha(\alpha-1)|} E\left|\widehat{\mathcal{AD}}_\alpha - \mathcal{AD}_\alpha\right| \\ &\le \frac{1}{|\alpha(\alpha-1)|}\left(E\left|\widehat{\mathcal{AD}}_\alpha - L_\alpha\right| + E|L_\alpha - M_\alpha| + E|M_\alpha - \mathcal{AD}_\alpha|\right) \\ &= \frac{1}{|\alpha(\alpha-1)|}(I_1 + I_2 + I_3). \end{aligned}$$

To prove the theorem, it suffices to show that $I_j \to 0$ as $n \to \infty$ for $j = 1, 2, 3$. Substituting (19) and (20) into I_1, we have

$$
\begin{aligned}
I_1 &= E\left|\int_{I^2}\left[\hat{c}^{\alpha-1}(u,v)-(E(\hat{c}(u,v)))^{\alpha-1}\right]dC_n(u,v)\right| \\
&\leq E\int_{I^2}\left|\hat{c}^{\alpha-1}(u,v)-(E(\hat{c}(u,v)))^{\alpha-1}\right|dC_n(u,v).
\end{aligned}
$$

Since a Taylor expansion of the differentiable function $g(z)$ about a real number b may be expressed as $g(z)=g(b)+(z-b)\,g'[\theta b+(1-\theta)z]$. for some θ, $0<\theta<1$, we have

$$
|g(z)-g(c)|\leq|z-c|\,|g'[\theta c+(1-\theta)z]|, \qquad 0<\theta<1. \tag{22}
$$

Now taking $g(z)=z^{\alpha-1}$ with $z=\hat{c}(u,v)$ and $b=E(\hat{c}(u,v))$, we have

$$
\begin{aligned}
I_1 &\leq |\alpha-1|E\int_{I^2}|\hat{c}(u,v)-E(\hat{c}(u,v))|(\theta E(\hat{c}(u,v))+(1-\theta)\hat{c}(u,v))^{\alpha-2}dC_n(u,v) \quad (0<\theta<1) \\
&\leq |\alpha-1|\theta^{\alpha-2}E\int_{I^2}|\hat{c}(u,v)-E(\hat{c}(u,v))|(E(\hat{c}(u,v)))^{\alpha-2}dC_n(u,v) \qquad (\hat{c}(u,v)\geq 0) \\
&= |\alpha-1|\theta^{\alpha-2}E\left(\frac{1}{n}\sum_{i=1}^{n}|\hat{c}(\hat{U}_i,\hat{V}_i)-E(\hat{c}(\hat{U}_i,\hat{V}_i))|\left(E(\hat{c}(\hat{U}_i,\hat{V}_i))\right)^{\alpha-2}\right) \\
&= |\alpha-1|\theta^{\alpha-2}E\left(|\hat{c}(\hat{U},\hat{V})-E(\hat{c}(\hat{U},\hat{V}))|\left(E(\hat{c}(\hat{U},\hat{V}))\right)^{\alpha-2}\right) \\
&= |\alpha-1|\theta^{\alpha-2}\int E_{\hat{c}}(|\hat{c}(u,v)-E(\hat{c}(u,v))|(E(\hat{c}(u,v)))^{\alpha-2}\ |\hat{U}=u,\hat{V}=v)c(u,v)dudv.
\end{aligned} \tag{23}
$$

The second equality in (23) follows from the fact that $\hat{c}(\hat{U}_i,\hat{V}_i)$ for $i=1,2,\ldots,n$ are identically distributed. From Eqs. (7) and (8), and it can be shown that, for all n sufficiently large,

$$
\begin{aligned}
E(\hat{c}(u,v)) &= c(u,v)+O(b_n^4), \\
Var(\hat{c}(u,v)) &= O\left(\frac{1}{nb_n^2}\right).
\end{aligned}
$$

Therefore, with the aid of assumptions *i* and *ii* we have, for all $(u,v)\in(0,1)^2$,

$$
E(\hat{c}(u,v))\longrightarrow c(u,v),
$$

and

$$
|\hat{c}(u,v)-E(\hat{c}(u,v))|\leq(Var(\hat{c}(u,v)))^{1/2}\longrightarrow 0,
$$

as $n\longrightarrow\infty$. Hence the integrand in the last expression of Eq. (23) converges to zero as $n\longrightarrow\infty$. Furthermore, with the assumption *iii*, note that

$$
\begin{aligned}
E_{\hat{c}}\,(|\hat{c}(u,v)-E(\hat{c}(u,v))|(E(\hat{c}(u,v)))^{\alpha-2})c(u,v) &\leq (E(\hat{c}(u,v)+E(\hat{c}(u,v))(E(\hat{c}(u,v)))^{\alpha-2}c(u,v) \\
&= 2(E(\hat{c}(u,v)))^{\alpha-1}c(u,v)<\infty,
\end{aligned}
$$

which is integrable. By an application of the Lebesgue dominated convergence theorem, the last integral in Eq. (23) converges to zero as $n\longrightarrow\infty$. Therefore, $I_1\longrightarrow 0$ as $n\longrightarrow\infty$.

To show that $I_2\longrightarrow 0$, let $g(z)=z^{\alpha-1}$ with $z=E(\hat{c}(u,v))$ and let $b=c(u,v)$ in Eq. (22). Then, similar to Eq. (23), we have

$$
\begin{aligned}
I_2 &= E\left|\int_{I^2}\left[(E(\hat{c}(u,v)))^{\alpha-1} - c^{\alpha-1}(u,v)\right]dC_n(u,v)\right| \\
&\le E\int_{I^2}\left|(E(\hat{c}(u,v)))^{\alpha-1} - c^{\alpha-1}(u,v)\right|dC_n(u,v) \\
&\le |\alpha-1|E\int_{I^2}|E(\hat{c}(u,v)) - c(u,v)|(\theta c(u,v) + (1-\theta)E(\hat{c}(u,v)))^{\alpha-2}dC_n(u,v) \quad (0<\theta<1) \\
&\le |\alpha-1|\theta^{\alpha-2}E\int_{I^2}|E(\hat{c}(u,v)) - c(u,v)|c(u,v)^{\alpha-2}dC_n(u,v) \quad (E(\hat{c}(u,v)) \ge 0) \\
&= |\alpha-1|\theta^{\alpha-2}E\left(\frac{1}{n}\sum_{i=1}^{n}\left|E(\hat{c}(\hat{U}_i,\hat{V}_i)) - c(\hat{U}_i,\hat{V}_i)\right|c^{\alpha-2}(\hat{U}_i,\hat{V}_i))\right) \\
&= |\alpha-1|\theta^{\alpha-2}E\left(\left|E(\hat{c}(\hat{U},\hat{V})) - c(\hat{U},\hat{V})\right|c^{\alpha-2}(\hat{U},\hat{V}))\right) \\
&= |\alpha-1|\theta^{\alpha-2}\int E_{\hat{c}}\left(|E(\hat{c}(u,v)) - c(u,v)|c^{\alpha-2}(u,v)|\hat{U}=u,\hat{V}=v\right)c(u,v)dudv.
\end{aligned} \tag{24}
$$

Since $|E(\hat{c}(u,v)) - c(u,v)| \to 0$ as $n\to\infty$, for all $(u,v)\in(0,1)^2$ and $|E(\hat{c}(u,v)) - c(u,v)| \le E(\hat{c}(u,v)) + c(u,v)$, which is integrable and converges to an integrable limit $2c(u,v)$ as $n\to\infty$, we conclude that $I_2 \to 0$ as $n\to\infty$ by an application of the Lebesgue dominated convergence theorem.

Finally, for I_3, note that

$$
\begin{aligned}
I_3 &= E\left|\frac{1}{n}\sum_{i=1}^{n}c^{\alpha-1}(\hat{U}_i,\hat{V}_i) - E(c^{\alpha-1}(U,V))\right| \\
&\le \frac{1}{\sqrt{n}}\left(Var(c^{\alpha-1}(U,V))\right)^{1/2} \\
&= \frac{1}{\sqrt{n}}\left(\int_{I^2}c^{2\alpha-1}(u,v)dudv - \mathcal{AD}_\alpha^2\right)^{1/2}.
\end{aligned} \tag{25}
$$

Thus, with the aid of assumptions iv, $I_3 \to 0$ as $n\to\infty$. □

Now, a simple example for Proposition 2 is presented.

Example 1. Let U and V be uniform (0, 1) random variables whose joint distribution function is an arbitrary copula function. In Proposition 2, consider $b_n = n^{-1/4}$ and $\alpha = 1/2$. It can be shown that assumptions $(i)-(iv)$ based on Gaussian copula density in Eq. (15), are established as

$$
\int_{[0,1]^2}c^{2\alpha-1}(u,v)dudv = \int_{[0,1]^2}dudv < \infty.
$$

According to the proof of Proposition 2, we have

$$
E\left|\widehat{\mathcal{AD}_\alpha}(c) - \mathcal{AD}_\alpha(c)\right| \le \frac{1}{|\alpha(\alpha-1)|}(I_1+I_2+I_3) \to 0 \qquad as \qquad n\to\infty.
$$

Under these assumptions, we have

$$
\begin{aligned}
E(\hat{c}(u,v)) &= c(u,v) + O\left(\frac{1}{n}\right), \\
Var(\hat{c}(u,v)) &= O\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

and so, $I_j \to 0$ as $n\to\infty$ for $j=1,2,3$.

4.1. High-dimensional copula-based α-divergence

A multivariate d-dimensional of copula-based α-divergence according to Eq. (10) can be presented as

$$\mathcal{AD}_\alpha(c) = \frac{1}{\alpha(\alpha-1)}\left(\int_{[0,1]^d} c^\alpha(u_1,u_2,\dots,u_d)du_1du_2\dots du_d - 1\right), \qquad \alpha \in \mathbb{R}\backslash\{0,1\}. \tag{26}$$

In practice, to use Eq. (26), the copula density must be estimated. Nonparametric density estimators in high-dimensional ($d > 3$) suffer a great deal from the well-known curse of dimensionality. In this situation, nonparametric density estimators converge more slowly to the true density as dimension increases (see Scott 2008). Sparseness of the data is a probable reason of the curse of dimensionality.

Nagler and Czado (2016) showed that one can evade the curse of dimensionality by assuming a simplified vine copula model for the dependence between variables, which use marginal densities and bivariate copulas as building blocks. Vine copula models follow the idea of Joe (1996) that any d-dimensional copula can be expressed in terms of $d(d-1)/2$ bivariate (conditional) copulas. Nagler and Czado (2016) formulated a general nonparametric estimator for such a model and showed under high-level assumptions that the speed of convergence is independent of dimension. Also, they discussed the asymptotic normality of their estimator under these assumptions and their simulation experiments illustrated a large gain in finite sample performance when the assumptions are at least approximately true.

4.2. Test of independence

In practical applications, a first natural question is whether the copula function in Eq. (1) is actually different from the independence copula. If variables are independent, then only marginal modeling is necessary, which can be carried out by using classical statistical approaches for univariate *i.i.d.* observations. If independence is rejected, then a typical next step is to fit an appropriate parametric copula family to the available data. So, according to Genest and Rémillard (2004), this amounts to testing

$$H_0 : C(u,v) = uv \qquad \textit{versus} \qquad H_1 : C(u,v) \neq uv. \tag{27}$$

The classical alternative consists of testing

$$H_0 : \tau = 0 \qquad \textit{versus} \qquad H_1 : \tau \neq 0.$$

From Hofert et al. (2018), we know that $C(u,v) = uv$ implies $\tau = 0$, but that the converse is false in general. This lack of equivalence, however, is usually not an issue in practice as copulas $C(u,v) \neq uv$ such that $\tau = 0$ do not seem to arise often in applications. In any case, the latter configuration can be easily discarded by a scatter plot of the observations.

Thus, the null hypothesis of independence test in Eq. (27) based on copula density is equivalent to testing

$$H_0 : c(u,v) = 1 \qquad \textit{versus} \qquad H_1 : c(u,v) \neq 1. \tag{28}$$

In order to test the null hypothesis (28), nonparametric estimators of the special cases α-divergence as $KL(c)$, $\chi^2_N(c)$, and $He(c)$ are considered as test statistics. Therefore, by replacing the unknown copula density in Eqs. (11)–(13) with its $\mathcal{LLPT}$ estimation in Eq. (6), the plug-in estimators of the copula-based α-divergence measures can be obtained as

$$\widehat{KL}(c) = \frac{1}{n}\sum_{i=1}^{n} \log\left(\hat{c}_n^{(\mathcal{LLPT})}(\hat{U}_i, \hat{V}_i)\right), \tag{29}$$

Table 3. Empirical size and power of the test statistics for Gaussian, Clayton, and Gumbel copulas and different values of Kendall's τ coefficient and sample size.

		$n=50$			$n=100$			$n=200$		
Copula	Statistic	$\tau=0$	$\tau=0.1$	$\tau=0.25$	$\tau=0$	$\tau=0.1$	$\tau=0.25$	$\tau=0$	$\tau=0.1$	$\tau=0.25$
Gaussian	S_n	0.045	0.180	0.695	0.060	0.261	0.960	0.050	0.463	1.000
	$\widehat{KL}(c)$	0.047	0.185	0.711	0.052	0.269	0.972	0.053	0.456	0.998
	$\widehat{\chi^2_N}(c)$	0.053	0.181	0.698	0.056	0.255	0.941	0.054	0.448	0.996
	$\widehat{He}(c)$	0.049	0.191	0.723	0.051	0.274	0.978	0.050	0.471	0.999
Clayton	S_n	0.051	0.195	0.625	0.062	0.285	0.935	0.072	0.491	1.000
	$\widehat{KL}(c)$	0.049	0.206	0.712	0.058	0.292	0.941	0.053	0.476	0.996
	$\widehat{\chi^2_N}(c)$	0.048	0.197	0.665	0.061	0.281	0.932	0.065	0.469	0.994
	$\widehat{He}(c)$	0.049	0.216	0.715	0.052	0.298	0.945	0.051	0.493	0.999
Gumbel	S_n	0.031	0.165	0.692	0.041	0.291	0.932	0.065	0.502	1.000
	$\widehat{KL}(c)$	0.047	0.188	0.705	0.053	0.309	0.941	0.054	0.491	0.996
	$\widehat{\chi^2_N}(c)$	0.046	0.179	0.696	0.055	0.286	0.928	0.055	0.476	0.991
	$\widehat{He}(c)$	0.048	0.192	0.719	0.049	0.315	0.947	0.051	0.511	1.000

$$\widehat{\chi^2_N}(c) = \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n} \hat{c}_n^{(\mathcal{LLPT})}(\hat{U}_i, \hat{V}_i) - 1\right), \tag{30}$$

$$\widehat{He}(c) = 1 - \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\sqrt{\hat{c}_n^{(\mathcal{LLPT})}(\hat{U}_i, \hat{V}_i)}}, \tag{31}$$

where $(\hat{U}_i, \hat{V}_i)_{i=1,\dots,n}$ are pseudo observations. It is notable that the null hypothesis of independence is equivalent to the nullity of the measures $\widehat{KL}(c)$, $\widehat{\chi^2_N}(c)$, and $\widehat{He}(c)$ and the null hypothesis Eq. (28) will be rejected for large value of these measures.

5. Simulation study

Simulation studies are performed under the null hypothesis to evaluate the finite-sample properties of the nonparametric tests of independence proposed in the previous sections. Under the alternative hypothesis, the Gaussian, T, T-EV, Clayton and Gumbel copulas are considered because they cover every degree of dependence, as measured by Kendall's τ. The procedure provided in Section 5.1 is used to calculate the critical value ($C.V.$), $P-value$, and empirical power ($E.P.$) of the independence tests based on S_n, $\widehat{KL}(c)$, $\widehat{\chi^2_N}(c)$, and $\widehat{He}(c)$ at 5% significance level.

5.1. Bootstrap procedure for calculating the empirical power

A procedure to compute the empirical power was proposed by Genest and Rémillard (2008) and Genest, Rémillard, and Beaudoin (2009) and it is a pattern for authors. The following procedure leads to the $C.V.$, approximate $P-value$, and $E.P.$ for the test of independence based on S_n, $\widehat{KL}(c)$, $\widehat{\chi^2_N}(c)$, and $\widehat{He}(c)$. For example, $C.V.$, approximate $P-value$, and $E.P.$ for $\widehat{KL}(c)$ can be obtained by means of the following procedure:

1. Generate B (a large integer) random samples of size n from independence copula (null hypothesis) and for each of these samples, determine the values of the test statistics; $\widehat{KL}_j(c); j = 1, 2, \dots, B$.
2. Compute $\widehat{KL}^{\dagger}(c)$ by using pseudo-observations from the alternative hypothesis (or real data).
3. If $\widehat{KL}_{1:B}(c), \widehat{KL}_{2:B}(c), \dots, \widehat{KL}_{B:B}(c)$ denote the ordered test statistical values calculated in step 3, then an estimate of the $C.V.$ at 0.05 level of significance ($C.V.\,5\%$) is $\widehat{KL}_{\lfloor(1-\alpha)B\rfloor:B}(c)$, where $\lfloor a \rfloor$ denotes the integer part of a. Thus, the approximate $P-value$ is given by

Table 4. Empirical size and power of the test statistics for T copula with 2, 5, and 10 degree of freedom and different values of Kendall's τ coefficient and sample size.

		$n=50$			$n=100$			$n=200$		
Copula	Statistic	$\tau=0$	$\tau=0.1$	$\tau=0.25$	$\tau=0$	$\tau=0.1$	$\tau=0.25$	$\tau=0$	$\tau=0.1$	$\tau=0.25$
T($\nu=2$)	S_n	0.120	0.215	0.730	0.112	0.410	0.945	0.165	0.680	1.000
	$\widehat{\ }$	0.291	0.338	0.782	0.551	0.791	0.975	0.891	0.976	1.000
	$\widehat{\chi^2_N}(c)$	0.288	0.331	0.773	0.532	0.746	0.966	0.875	0.966	1.000
	$\widehat{He}(c)$	0.296	0.345	0.797	0.598	0.813	0.979	0.909	0.982	1.000
T($\nu=5$)	S_n	0.070	0.205	0.680	0.075	0.380	0.935	0.085	0.520	1.000
	$\widehat{KL}(c)$	0.112	0.256	0.741	0.143	0.402	0.946	0.301	0.536	1.000
	$\widehat{\chi^2_N}(c)$	0.102	0.225	0.733	0.131	0.397	0.939	0.297	0.524	1.000
	$\widehat{He}(c)$	0.125	0.282	0.752	0.153	0.415	0.961	0.316	0.549	1.000
T($\nu=10$)	S_n	0.061	0.171	0.686	0.055	0.365	0.960	0.055	0.521	1.000
	$\widehat{KL}(c)$	0.052	0.212	0.706	0.056	0.383	0.975	0.053	0.509	1.000
	$\widehat{\chi^2_N}(c)$	0.055	0.209	0.698	0.059	0.354	0.943	0.059	0.497	0.998
	$\widehat{He}(c)$	0.052	0.221	0.713	0.051	0.395	0.979	0.050	0.526	1.000

Table 5. Empirical size and power of the test statistics for T-EV copula with 2, 5, and 10 degree of freedom and different values of Kendall's τ coefficient and sample size.

		n = 50			n = 100			n = 200		
Copula	Statistic	$\tau=0$	$\tau=0.1$	$\tau=0.25$	$\tau=0$	$\tau=0.1$	$\tau=0.25$	$\tau=0$	$\tau=0.1$	$\tau=0.25$
T-EV($\nu=2$)	S_n	0.216	0.482	0.971	0.348	0.583	1.000	0.581	0.918	1.000
	$\widehat{KL}(c)$	0.286	0.523	0.981	0.488	0.706	1.000	0.726	0.943	1.000
	$\widehat{\chi^2_N}(c)$	0.273	0.498	0.978	0.452	0.681	1.000	0.717	0.933	1.000
	$\widehat{He}(c)$	0.293	0.541	0.985	0.513	0.726	1.000	0.749	0.968	1.000
T-EV($\nu=5$)	S_n	0.091	0.221	0.521	0.095	0.415	0.890	0.095	0.932	1.000
	$\widehat{KL}(c)$	0.112	0.229	0.543	0.099	0.466	0.912	0.128	0.937	0.999
	$\widehat{\chi^2_N}(c)$	0.096	0.226	0.536	0.098	0.431	0.902	0.115	0.934	0.999
	$\widehat{He}(c)$	0.118	0.232	0.557	0.102	0.476	0.926	0.131	0.946	1.000
T-EV($\nu=10$)	S_n	0.075	0.134	0.552	0.061	0.245	0.795	0.031	0.473	1.000
	$\widehat{KL}(c)$	0.054	0.142	0.568	0.053	0.265	0.821	0.055	0.449	0.989
	$\widehat{\chi^2_N}(c)$	0.056	0.139	0.558	0.058	0.236	0.788	0.061	0.443	0.971
	$\widehat{He}(c)$	0.051	0.148	0.572	0.051	0.273	0.837	0.050	0.461	0.996

$$P - value = \frac{1}{B}\sum_{j=1}^{B} I_{\left\{\widehat{KL}_j(c) \geq \widehat{KL}^{\dagger}(c)\right\}}.$$

5) Finally, to evaluate the *E.P.* at the α level for some large integer M, repeat the calculating approximate P – *value* and thus *E.P.* is equal to the number of approximate P – *values* less than α.

The **kdecopula** package (Nagler and Wen (2018)) in R software is used for estimation of the copula density. A Monte Carlo experiment based on 1000 replications is performed according to the bootstrap procedure. The power function of the new tests in Eqs. (29)–(31) is compared with the power function of the classical test, which is based on the empirical copula for various sample sizes $n = 50, 100, 200$ and different degrees of dependency based on different Kendall's τ coefficient, $\tau = 0, 0.1, 0.25$. For Kendall's τ coefficient greater than 0.5, all the considered tests provide good and comparable results.

The empirical size and power of all the considered tests for the Gaussian, Clayton, and Gumbel copulas are reported in Table 3, and for T and T-EV, copulas with 2, 5, and 10 degrees of freedom are reported in Tables 4 and 5. In Tables 3–5, the power of the test increases as the sample size increases and also, the power of the test increases as the degree of dependency measured by Kendall's τ increases.

In Table 3, the empirical size of all the tests is obtained when Kendall's τ is equal to zero. For T and T-EV copulas, when Kendall's τ is equal to zero and the degree of freedom is small ($\nu=2$ and $\nu=5$), dependency occurs in the tails and in this case there is no independence. Thus, in

Table 6. Independence tests for some pairs of uranium exploration dataset.

		Pairs					
		(U,Co) ($\tau = 0.0596$)	(U,Sc) ($\tau = 0.0923$)	(Li,Co) ($\tau = 0.0061$)	(Li,Ti) ($\tau = 0.0028$)	(K,Ti) ($\tau = 0.0406$)	$C.V.\,5\%$
S_n	Statistic	0.2256	0.4480	0.0995	0.0959	0.1832	0.1124
	P-value	0.004	< 0.001	0.082	0.094	0.007	
$\widehat{KL}(c)$	Statistic	0.0498	0.0656	0.0142	0.0251	0.0352	0.0189
	P-value	< 0.001	< 0.001	0.441	0.003	< 0.001	
$\widehat{\chi^2_N}(c)$	Statistic	0.0381	0.0473	0.0088	0.0169	0.0251	0.0121
	P-value	< 0.001	< 0.001	0.43	0.005	< 0.001	
$\widehat{He}(c)$	Statistic	0.0211	0.0268	0.0052	0.0139	0.0153	0.0072
	P-value	< 0.001	< 0.001	0.462	0.002	< 0.001	

Tables 4 and 5, the empirical size of the tests for T and T-EV copulas is obtained when Kendall's τ is equal to zero and the degree of freedom is equal to 10. In Tables 3–5, it can be observed that all the tests generally control the size.

A comprehensive survey of the results leads to general ideas as follows. In Tables 3–5, $\widehat{He}(c)$ has the best performance among the new α-divergence tests. Based on Proposition 1, part (b), it is observed that the maximum value $\widehat{He}(c)$ is finite, and according to Micheas and Zografos (2006), we suggest that this measure of dependence will be preferred from another. Also, the results show that the performance of α-divergence independence test decreases with increasing alpha value in Eq. (17). Overall, the proposed tests based on α-divergence have a better performance than S_n for a small sample size or weak dependency.

According to the results of Table 3, the empirical power functions of all the tests for Gaussian, Clayton, and Gumble copulas are comparable. In this table, for all degrees of dependence, the outcomes can be stated as follows:

a. When the sample size is equal to 50, the new independence tests based on α-divergence perform better than S_n.
b. When the sample size is equal to 100, the independence tests based on $\widehat{He}(c)$ and $\widehat{KL}(c)$ have outperform than S_n.
c. When the sample size is greater than 100 and the degree of dependency is very low ($\tau \leq 0.1$), the independence tests based on $\widehat{He}(c)$ perform better than S_n.

In Tables 4 and 5, when the degree of freedom is equal to 10, the same result as in Table 3 is obtained for $n = 50, 100, 200$. In these two tables, when the degree of freedom is equal to 2 or 5, the new independence tests based on α-divergence outperform than S_n and $\widehat{He}(c)$ has the best performance among the tests based on α-divergence. In these cases, the difference becomes even more important when the sample size increases.

For example, in T copula with 2 degree of freedom, $n = 200$, and $\tau = 0$, the empirical powers of the tests $\widehat{KL}(c)$, $\widehat{\chi^2_N}(c)$, and $\widehat{He}(c)$ are equal to 0.891, 0.875, and 0.909, respectively, where as the power of test S_n is equal to 0.165. The same remark applies for T-EV copula when the degree of dependencies is small. For example, in T-EV copula with 2 degree of freedom, $n = 200$, and $\tau = 0$, the empirical powers of the tests $\widehat{KL}(c)$, $\widehat{\chi^2_N}(c)$, and $\widehat{He}(c)$ are equal to 0.726, 0.717, and 0.749, respectively, where as the power of test S_n is equal to 0.581. These results for T and T-EV copulas were to be expected according to Figure 1.

6. Application

An application of the new tests is presented to a given dataset for the uranium exploration. This dataset is considered by Cook and Johnson (1986), and Genest, Quessy, and Rémillard (2006)

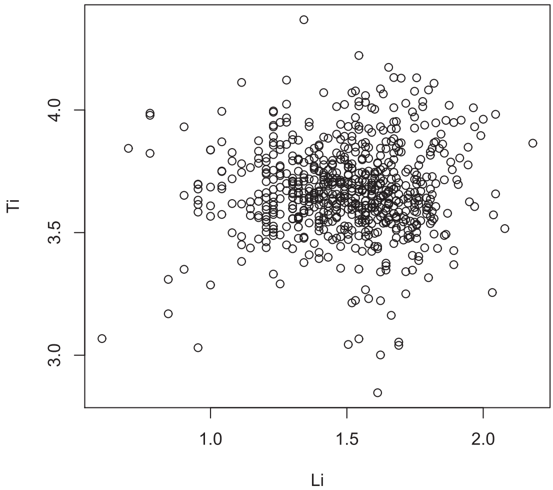

Figure 3. Scatter plot for the pair (Li, Ti) of uranium exploration dataset.

and is given in the package **copula** by Hofert et al. (2018). The uranium exploration dataset is consist of 655 chemical analyses of water samples collected from the Montrose quadrangle of western Colorado (USA). Concentrations for the following elements were measured: uranium (U), lithium (Li), cobalt (Co), potassium (K), cesium (Cs), scandium (Sc), and titanium (Ti). In Table 6, the values of the test statistics S_n, $\widehat{KL}(c)$, $\widehat{\chi^2_N}(c)$, and $\widehat{He}(c)$ for selected pairs of variables are shown. For these pairs, the corresponding $P-value$ and $C.V.$ are calculated at 5% level of significance with 1000 iterations according to the bootstrap procedure that is illustrated in the Section 5.1.

Based on the results in Table 6, the copula-based α-divergence tests are generally in agreement with the empirical copula-based test for the pairs (U, Co), (U, Sc), (Li, Co), and (K, Ti). These tests lead to a different result in rejecting the independence hypothesis as for the pair (Li, Ti). According to the test S_n, the independence test is not rejected with a $P-value$ of 0.094, whereas the independence tests based on α-divergence are rejected and their $P-values$ are less than 0.05. Also, the value of test statistics S_n is less than $C.V\,5\%$ that it can be observed in the last column of Table 6 and thus the independence hypothesis is not rejected.

The scatter plot of the pair (Li, Ti) can be seen in Figure 3. The presence of tail dependency or outlying observations in the pair (Li, Ti) may be reasons to get different results. According to the results of Section 5, when Kendall's τ is close to zero ($\tau \leq 0.1$), the independence test based on $\widehat{He}(c)$ is more reliable than the independence test based on empirical copula, because dependency may occur at the tails and these tests can measure this kind of dependency. Thus, based on the results of $\widehat{He}(c)$, the independence hypothesis for the pair (Li, Ti) is rejected.

7. Conclusion

This study introduced three nonparametric independence tests based on the α-divergence in the class of continuous random variables. The special cases of α-divergence were calculated in terms of copula density, and their features were investigated. The properties of copula-based α-divergence for some copulas were described. The asymptotic first-order consistency of α-divergence estimators was demonstrated. The test statistics $\widehat{KL}(c)$, $\widehat{\chi^2_N}(c)$, and $\widehat{He}(c)$ were estimated by the $\mathcal{LLPT}$ estimator of passive copula density. A Monte Carlo experiment was run to investigate the performance of the tests in comparison with the existing test based on empirical copula. The proposed tests based on the α-divergence have a better performance than S_n for a small sample size or weak dependency. Moreover, $\widehat{He}(c)$ has the best performance among the new α-divergence tests. The simulation results showed that the suggested tests outperform for T and T-EV copulas with small degrees of freedom.

Acknowledgments

We are very grateful to the Associate Editor and the reviewers for their appropriate and constructive suggestions.

ORCID

M. Mohammadi http://orcid.org/0000-0001-9850-0078
M. Emadi http://orcid.org/0000-0003-1663-9119
M. Amini http://orcid.org/0000-0002-8336-201X

References

Ahmad, I., and P. E. Lin. 1976. A nonparametric estimation of the entropy for absolutely continuous distributions. *IEEE Transactions on Information Theory* 22 (3):372–5. doi:10.1109/TIT.1976.1055550.

Belalia, M., T. Bouezmarni, F. C. Lemyre, and A. Taamouti. 2017. Testing independence based on Bernstein empirical copula and copula density. *Journal of Nonparametric Statistics* 29 (2):346–80. doi:10.1080/10485252.2017.1303063.

Blum, J. R., J. Kiefer, and M. Rosenblatt. 1961. Distribution free tests of independence based on the sample distribution function. *The Annals of Mathematical Statistics* 32 (2):485–98. doi:10.1214/aoms/1177705055.

Blumentritt, T., and F. Schmid. 2012. Mutual information as a measure of multivariate association: Analytical properties and statistical estimation. *Journal of Statistical Computation and Simulation* 82 (9):1257–74. doi:10.1080/00949655.2011.575782.

Charpentier, A., J. Fermanian, and O. Scaillet. 2006. *Copulas: From theory to application in finance, chapter the estimation of copulas: Theory and Practice*, vol. 1, 35–62. Torquay, UK: Risk Books.

Chernoff, H. 1952. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics* 23 (4):493–507. doi:10.1214/aoms/1177729330.

Cichocki, A., and S. I. Amari. 2010. Families of alpha-beta-and gamma-divergences: Flexible and robust measures of similarities. *Entropy* 12 (6):1532–68. doi:10.3390/e12061532.

Cook, R. D., and M. E. Johnson. 1986. Generalized Burr–Pareto–logistic distributions with applications to a uranium exploration data set. *Technometrics* 28 (2):123–31. doi:10.1080/00401706.1986.10488113.

Deheuvels, P. 1979. La fonction de dependence empirique et ses proprietes, Un test non parametrique d'independance. *Bulletin de la Classe Des Sciences* 65 (1):274–92. doi:10.3406/barb.1979.58521.

Demarta, S., and A. J. McNeil. 2007. The t copula and related copulas. *International Statistical Review* 73 (1):111–29. doi:10.1111/j.1751-5823.2005.tb00254.x.

Eguchi, S., and S. Kato. 2010. Entropy and divergence associated with power function and the statistical application. *Entropy* 12 (2):262–74. doi:10.3390/e12020262.

Geenens, G. 2014. Probit transformation for kernel density estimation on the unit interval. *Journal of the American Statistical Association* 109 (505):346–58. doi:10.1080/01621459.2013.842173.

Geenens, G., A. Charpentier, and D. Paindaveine. 2017. Probit transformation for nonparametric kernel estimation of the copula density. *Bernoulli* 23 (3):1848–73. doi:10.3150/15-BEJ798.

Genest, C., J. F. Quessy, and B. Rémillard. 2006. Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics* 33 (2):337–66. doi:10.1111/j.1467-9469.2006.00470.x.

Genest, C., and B. Rémillard. 2004. Test of independence and randomness based on the empirical copula process. *Test* 13 (2):335–69. doi:10.1007/BF02595777.

Genest, C., and B. Rémillard. 2008. Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'IHP Probabilités et Statistiques* 44 (6):1096–127.

Genest, C., B. Rémillard, and D. Beaudoin. 2009. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics* 44 (2):199–213. doi:10.1016/j.insmatheco.2007.10.005.

Hofert, M., I. Kojadinovic, M. Maechler, and J. Yan. 2018. copula: Multivariate dependence with copulas. R package version 0.999-19. https://CRAN.R-project.org/package=copula.

Joe, H. 1989. Relative entropy measures of multivariate dependence. *Journal of the American Statistical Association* 84 (405):157–64. doi:10.1080/01621459.1989.10478751.

Joe, H. 1996. Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. *Lecture Notes-Monograph Series* 28:120–41.

Joe, H. 2014. *Dependence modeling with copulas*. New York: Chapman and Hall/CRC.

Kullback, S., and R. A. Leibler. 1951. On information and sufficiency. *The Annals of Mathematical Statistics* 22 (1): 79–86. doi:10.1214/aoms/1177729694.

Liese, F., and I. Vajda. 1987. *Convex statistical distances*, vol. 95. Leipzig: BG Teubner Verlagsgesellschaft.

Micheas, A. C., and K. Zografos. 2006. Measuring stochastic dependence using Φ-divergence. *Journal of Multivariate Analysis* 97 (3):765–84.

Nagler, T. 2018. kdecopula: An R package for the kernel estimation of bivariate copula densities. *Journal of Statistical Software* 84 (7):1–22. doi:10.18637/jss.v084.i07.

Nagler, T., and C. Czado. 2016. Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. *Journal of Multivariate Analysis* 151:69–89. doi:10.1016/j.jmva.2016.07.003.

Nagler, T. 2014. Kernel methods for vine copula estimation. Master's thesis, Technische Universität München.

Nagler, T., and K. Wen. 2018. kdecopula: Kernel smoothing for bivariate copula densities. R package version 0.9.2. https://CRAN.R-project.org/package=kdecopula.

Nelsen, R. B. 2007. *An introduction to copulas*. New York: Springer Science and Business Media.

Read, T. R., and N. A. Cressie. 2012. *Goodness-of-fit statistics for discrete multivariate data*. New York: Springer Science and Business Media.

Rényi, A. 1961. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, 547–61. California: University of California Press.

Scott, D. W. 2008. The curse of dimensionality and dimension reduction. *In Multivariate Density Estimation: Theory, Practice, and Visualization*, vol. 1, 2nd ed., 217–40. New York: John Wiley & Sons.

Sklar, M. 1959. Fonctions de repartition an dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris* 8:229–31.