South African Statistical Journal
Vol. 58, No. 2, pp. 99–108
https://doi.org/10.37920/sasj.2024.58.2.2

A new wavelet estimator of multivariate copula densities based on Sklar's theorem, with optimal strong uniform convergence rate

Jan Swanepoel[1], *Cheikh Tidiane Seck*[2] *and Salha Mamane*[3]

[1]Unit for Data Science and Computing, North-West University, Potchefstroom, South Africa
[2]Department of Mathematics, Alioune Diop University, Bambey, Senegal
[3]School of Statistics and Actuarial Science, University of the Witwatersrand, Johannesburg, South Africa

In this paper, we propose a natural wavelet estimator of multivariate copula densities as a ratio of the linear wavelet estimator of the underlying joint population density function and the product of the linear wavelet estimators of the corresponding marginal density functions. It is proven that the new estimator attains the optimal almost sure convergence rate over Besov balls for the supremum norm whenever the resolution level is suitably chosen.

Keywords: Convergence rate, Copula, Nonparametric estimation, Wavelet.

1. Introduction

Given a vector $\mathbf{X} = (X_1, \ldots, X_d)$ of continuous random variables with marginal distribution functions $F_1, \ldots, F_d$, the copula function of $\mathbf{X}$ is defined as the joint cumulative distribution function of the random vector $(F_1(X_1), \ldots, F_d(X_d))$. It gives a full characterisation of the dependence between random variables, be it linear or nonlinear. According to Durante and Sempi (2010), the study of copulas dates back to Fréchet (1951) and the term "copula" was first used in this sense by Sklar (1959). Nelsen (2003) dates the study of copulas even further back in time to Hoeffding (1940, 1941). Copulas have applications in various fields such as finance, insurance, environmental science, healthcare, hydrology, economics, marketing, demography, climate science, psychology (see, for instance, Manner and Reznikova, 2012; Patton, 2012; Bhatti and Do, 2019; Jaworski et al., 2010; Größer and Okhrin, 2022).

Nonparametric estimation of copula densities is a vibrant research area, historically dominated by the use of kernel methods. For instance, Gijbels and Mielniczuk (1990) and Fermanian and Scaillet (2003) used convolution kernel methods to construct consistent estimators for the copula density. However, kernel methods are prone to boundary effect issues due to the compact support of the copula function. To address this issue, Gijbels and Mielniczuk (1990) introduced a mirror-reflection technique, and Chen and Huang (2007) used a local linear kernel procedure. Omelka et al. (2009) also introduced improved copula kernel estimators to mitigate boundary bias and Geenens et al. (2017) introduced kernel-type estimators for the copula density based on a probit transformation method, which effectively handle boundary effects.

Corresponding author: Salha Mamane (Salha.Mamane@wits.ac.za)
MSC2020 subject classifications: 62G07, 62G20

Wavelet bases (Meyer, 1992; Daubechies, 1992; Vidakovic, 1999; Mallat, 2009) offer an alternative approach as they automatically adapt to the properties of the curve being estimated, thus handling boundary effects inherently. Genest et al. (2009) studied a rank-based linear wavelet estimator of the bivariate copula density, establishing its optimality under certain conditions in terms of the L_2-norm loss as well as on Hölder balls for the pointwise-norm loss. Autin et al. (2010) extended these results to nonlinear thresholded estimators of multivariate copula densities, demonstrating near-optimal performance for the L_2-norm loss. Similarly, Gannoun and Hosseinioun (2012) established upper bounds on L_p-losses for linear wavelet-based estimators of bivariate copula densities. Seck and Mamane (2024) investigated the almost sure convergence, in supremum norm, of the rank-based linear wavelet estimator of a multivariate copula density. However, the convergence rate achieved by this estimator was found to be suboptimal.

In this paper, we propose a natural wavelet estimator of the multivariate copula density as a ratio of the linear wavelet estimator of the joint density function of $\mathbf{X}$ and the product of linear wavelet estimators of the corresponding marginal density functions. Density estimation using wavelet methods is discussed, for example, in Härdle et al. (1998), Giné and Nickl (2009), and Guo and Kou (2019). We establish the exact almost sure convergence rate in supremum norm loss of this new estimator. Contrary to the linear wavelet estimator studied by Seck and Mamane (2024), this Sklar theorem-based estimator achieves the optimal convergence rate over Besov balls for the supremum norm loss.

The rest of the paper is organised as follows. In Section 2, we recall some facts on wavelet theory and define the proposed wavelet estimator of the multivariate copula density. In Section 3, the main theoretical results are presented.

2. Wavelet theory and estimation procedure

In this section we recall some facts about wavelet theory. For a general introduction we refer to Meyer (1992), Daubechies (1992), Vidakovic (1999) and Mallat (2009). A multiresolution analysis of $L_2(\mathbb{R}^d)$ is a sequence of closed sub-spaces $(V_j)_{j\in\mathbb{Z}}$ of $L_2(\mathbb{R}^d)$ satisfying the following properties:

i) **Increasing**: $V_j \subset V_{j+1}$;

ii) **Separability**: $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$;

iii) **Density**: $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L_2(\mathbb{R}^d)$;

iv) **Dilatation**: $f(\mathbf{x}) \in V_j \Leftrightarrow f(2\mathbf{x}) \in V_{j+1}$;

v) **Translation invariance**: $f(\mathbf{x}) \in V_j \Rightarrow f(\mathbf{x}+\mathbf{k}) \in V_j, \forall\, \mathbf{k} \in \mathbb{Z}^d$;

vi) **Existence of scaling function**: $\exists\ \phi \in L_2(\mathbb{R}^d)$, with $\int_{\mathbb{R}^d} \phi(\mathbf{x})d\mathbf{x} = 1$, such that the family $\{\phi(\mathbf{x}-\mathbf{k}), \mathbf{k} \in \mathbb{Z}^d\}$ is an orthonormal basis for V_0.

The scaling function ϕ is called a father wavelet, and the family $\{\phi_{j,\mathbf{k}}(\mathbf{x}) = 2^{jd/2}\phi(2^j\mathbf{x}-\mathbf{k}), \mathbf{k} \in \mathbb{Z}^d\}$ is an orthonormal basis for V_j, for all j. A multiresolution analysis is called r-regular if $\phi \in C^r$ and all its partial derivatives up to total order r are rapidly decreasing, i.e., for every integer $i \geq 0$, there exists a constant A_i such that

$$|(D^\beta\phi)(\mathbf{x})| \leq \frac{A_i}{(1+\|\mathbf{x}\|)^i}, \qquad \text{for all} \quad |\beta| \leq r,$$

where $\beta = (\beta_1, \ldots, \beta_d)$ is a vector of positive integers and $|\beta| = \sum_{i=1}^{d} \beta_i$.

Let W_j denote the orthogonal complement of the subspace V_j in V_{j+1}, i.e., $V_{j+1} = V_j \oplus W_j$, then, for all fixed $j_0 \in \mathbb{Z}$, we have the decomposition

$$L_2(\mathbb{R}^d) = V_{j_0} \oplus \left(\bigoplus_{j \geq j_0} W_j \right), \tag{1}$$

and there exist $M = 2^d - 1$ wavelet functions $\{\psi^{(m)}, m = 1, \ldots, M\}$ called mother wavelets such that

- $\{\psi^{(m)}(\mathbf{x} - \mathbf{k}), \mathbf{k} \in \mathbb{Z}^d, m = 1, \ldots, M\}$ is an orthonormal basis for W_0;
- $\{\psi_{j,\mathbf{k}}^{(m)}(\mathbf{x}) = 2^{\frac{jd}{2}} \psi^{(m)}(2^j \mathbf{x} - \mathbf{k}), \mathbf{k} \in \mathbb{Z}^d, m = 1, \ldots, M\}$ is an orthonormal basis for W_j;
- $\{\phi_{j_0,\mathbf{k}}, \psi_{j,\mathbf{k}}^{(m)}\}_{j \geq j_0, \mathbf{k} \in \mathbb{Z}^d, 1 \leq m \leq M}$ is an orthonormal basis for $L_2(\mathbb{R}^d)$;
- each $\psi^{(m)}$ has the same regularity properties as ϕ, which will be assumed to be compactly supported on $[-L, L]^d$, $L > 0$.

It follows that for any function $f \in L_2(\mathbb{R}^d)$,

$$f(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{Z}^d} \alpha_{j_0,\mathbf{k}} \phi_{j_0,\mathbf{k}}(\mathbf{x}) + \sum_{j \geq j_0} \sum_{m=1}^{M} \sum_{\mathbf{k} \in \mathbb{Z}^d} \beta_{j,\mathbf{k}}^{(m)} \psi_{j,\mathbf{k}}^{(m)}(\mathbf{x}), \qquad \mathbf{x} \in \mathbb{R}^d, \tag{2}$$

where $j_0 \in \mathbb{Z}$ is called a resolution level, $\alpha_{j_0,\mathbf{k}}$ are the scaling coefficients and $\beta_{j,\mathbf{k}}^{(m)}$ are the wavelet coefficients. Due to the orthogonality of the basis, the scaling coefficients and the wavelet coefficients are respectively given by

$$\alpha_{j_0,\mathbf{k}} = \int_{\mathbb{R}^d} f(\mathbf{x}) \phi_{j_0,\mathbf{k}}(\mathbf{x}) d\mathbf{x}, \qquad \beta_{j,\mathbf{k}}^{(m)} = \int_{\mathbb{R}^d} f(\mathbf{x}) \psi_{j,\mathbf{k}}^{(m)}(\mathbf{x}) d\mathbf{x}.$$

Besov spaces can be characterised in terms of wavelet coefficients. If $P_{V_j} f$ denotes the orthogonal projection of f onto the subspace V_j, then for all $j \geq j_0$, we have

$$(P_{V_j} f)(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{Z}^d} \alpha_{j,\mathbf{k}} \phi_{j,\mathbf{k}}(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{Z}^d} \alpha_{j_0,\mathbf{k}} \phi_{j_0,\mathbf{k}}(\mathbf{x}) + \sum_{l=j_0}^{j} \sum_{m=1}^{M} \sum_{\mathbf{k} \in \mathbb{Z}^d} \beta_{l,\mathbf{k}}^{(m)} \psi_{l,\mathbf{k}}^{(m)}(\mathbf{x}).$$

Assume that the multiresolution analysis is r-regular. Then a function $f \in L_2(\mathbb{R}^d)$ belongs to the Besov space $B_{p,q}^t(\mathbb{R}^d)$, with $t < r$, if and only if

$$\|P_{V_0}(f)\|_{L_p} + \left(\sum_{j \geq 0} (2^{jt} \|P_{W_j}(f)\|_{L_p})^q \right)^{1/q} < \infty,$$

for $1 \leq p, q \leq \infty$. For more details regarding Besov spaces, we refer to Härdle et al. (1998) and the Appendix of Masry (2000).

Let $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ be a d-dimensional random vector with absolutely continuous (w.r.t. Lebesgue measure) distribution function $F(\mathbf{x})$ and corresponding joint probability density function $f(\mathbf{x})$, where $\mathbf{x} = (x_1, \ldots, x_d)$.

Denote the corresponding marginal distribution functions and probability density functions by $F_1(x_1), \ldots, F_d(x_d)$ and $f_1(x_1), \ldots, f_d(x_d)$ respectively.

Suppose that $(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$ is a random sample from the population $\mathbf{X}$, where $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{id})$, $i = 1, \ldots, n$.

The linear wavelet estimator of $f(\mathbf{x})$ at resolution level $j_n(d) > j_0$ is given by

$$\widehat{f}_{j_n(d)}(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} \widehat{\alpha}_{j_n(d),\mathbf{k}} \phi_{j_n(d),\mathbf{k}}(\mathbf{x}), \quad \text{with } \widehat{\alpha}_{j_n(d),\mathbf{k}} = \frac{1}{n}\sum_{i=1}^{n} \phi_{j_n(d),\mathbf{k}}(\mathbf{X}_i).$$

Henceforth, write $j(d)$ instead of $j_n(d)$ for the resolution level.

The corresponding univariate linear wavelet estimator of $f_\ell(x_\ell)$ is denoted by $\widehat{f}_{\ell,j(1)}(x_\ell)$ for each $\ell = 1, \ldots, d$.

Assuming that the copula density $c(\mathbf{u})$, $\mathbf{u} = (u_1, \ldots, u_d)$, associated with the random vector $\mathbf{X}$ exists and is continuous and bounded on $(0,1)^d$, then it follows from the well-known Sklar's theorem that

$$c(\mathbf{u}) := \frac{f(Q_1(u_1), \ldots, Q_d(u_d))}{\prod_{\ell=1}^d f_\ell(Q_\ell(u_\ell))} =: \frac{A(\mathbf{u})}{B(\mathbf{u})}, \qquad \text{for all } \mathbf{u} \in (0,1)^d, \tag{3}$$

where $Q_\ell(u_\ell) := F_\ell^{-1}(u_\ell)$ is the generalised inverse function of $F_\ell(x_\ell)$, for $\ell = 1, \ldots, d$.

Hence, a natural wavelet estimator for $c(\mathbf{u})$ can now be defined as

$$\widehat{c}_n(\mathbf{u}) := \frac{\widehat{f}_{j(d)}(\widehat{Q}_1(u_1), \ldots, \widehat{Q}_d(u_d))}{\prod_{\ell=1}^d \widehat{f}_{\ell,j(1)}(\widehat{Q}_\ell(u_\ell))} =: \frac{\widehat{A}_n(\mathbf{u})}{\widehat{B}_n(\mathbf{u})}, \tag{4}$$

where $\widehat{Q}_\ell(u_\ell)$, $\ell = 1, \ldots, d$, are strongly consistent estimators of the quantile functions $Q_\ell(u_\ell)$. Examples could be to choose $\widehat{Q}_\ell(u_\ell)$ equal to the generalised inverse of the empirical distribution function or a smooth kernel distribution function estimator based on $X_{i\ell}$, $i = 1, \ldots, n$.

3. Asymptotic behaviour of the estimator

Let us assume that the following conditions hold:

Condition 1. The multiresolution analysis is r-regular and the father wavelet $\phi : \mathbb{R}^d \to \mathbb{R}$ is bounded with a compact support.

Condition 2. $f(\mathbf{x}) \in B^t_{p,q}(\mathbb{R}^d)$ is bounded for some $t > d/p$, and $f_\ell(x_\ell) \in B^s_{p,q}(\mathbb{R})$ is bounded for $\ell = 1, \ldots, d$ and some $s > 1/p$, $1 \le p, q \le \infty$.

Condition 3.

$$\sup_{u_\ell \in (0,1)} |\widehat{Q}_\ell(u_\ell) - Q_\ell(u_\ell)| = O_{a.s.}(n^{-1/2}(\log n)^\delta),$$

for $\ell = 1, \ldots, d$ and some $\delta > 0$.

Condition 4. The resolution level $j(d)$ satisfies

$$2^{j(d)} \approx \left(\frac{n}{\log n}\right)^{\gamma(d)},$$

where, for $d \ge 1$,

$$\gamma(d) = \frac{1}{d + 2(t - d/p)}.$$

Remark 1. All the conditions (except for Condition 3) are quite standard as they are usually imposed in the study of wavelet density estimation. Conditions 1 and 2 are satisfied for a large variety of wavelets. One example is the Daubechies' wavelets (see Daubechies, 1992, Chapter 6, or Härdle et al., 1998). The following proposition proves that Condition 3 is fulfilled if we choose $\widehat{Q}_\ell(u_\ell) := F_{\ell,n}^{-1}(u_\ell)$, $\ell = 1, \ldots, d$, where $F_{\ell,n}(x_\ell)$ is the empirical distribution function of $\mathbf{X}_\ell$. Similar results exist in the statistical literature for other types of quantile estimators. A typical example is the well-known smooth kernel quantile estimator.

Proposition 1. *Assume that* $\inf_{0<u_\ell<1} f_\ell(F_\ell^{-1}(u_\ell)) > 0$, *for all* $\ell = 1, \ldots, d$. *Then we have*

$$\sup_{0<u_\ell<1} |F_{\ell,n}^{-1}(u_\ell) - F_\ell^{-1}(u_\ell)| = O_{a.s.}\left(n^{-1/2}(\log\log n)^{1/2}\right).$$

Proof. For ease of notation, write $F_{\ell,n}^{-1}(u_\ell) =: F_n^{-1}(u)$, $F_\ell^{-1}(u_\ell) =: F^{-1}(u)$, and $f_\ell(u_\ell) =: f(u)$. Define (see Theorem 2.5.1 in Serfling, 1980)

$$R_n = \left(F_n^{-1}(u) - F^{-1}(u)\right) - \frac{u - F_n(F^{-1}(u))}{f(F^{-1}(u))}.$$

We then have, by applying respectively Theorem D and Theorem B appearing on pages 101 and 62 in Serfling (1980),

$$\begin{aligned}
&\sup_{0<u<1} f(F^{-1}(u))|F_n^{-1}(u) - F^{-1}(u)| \\
&\leq \sup_{0<u<1} f(F^{-1}(u))|R_n(u)| + \sup_{0<u<1} |u - F_n(F^{-1}(u))| \\
&= O_{a.s.}\left(n^{-3/4}(\log n)^{1/2}(\log\log n)^{1/4}\right) + \sup_{-\infty<x<\infty} |F_n(x) - F(x)| \\
&= O_{a.s.}\left(n^{-3/4}(\log n)^{1/2}(\log\log n)^{1/4}\right) + O_{a.s.}\left(n^{-1/2}(\log\log n)^{1/2}\right) \\
&= O_{a.s.}\left(n^{-1/2}(\log\log n)^{1/2}\right).
\end{aligned}$$

■

Theorem 1. *Let Conditions 1–4 hold and suppose that* $\operatorname{supp} f_\ell(x_\ell) = I = [a, b]$, *for some* $-\infty < a < b < \infty$, *and* $\inf_{x_\ell \in I} f_\ell(x_\ell) > 0$ *for* $\ell = 1, \ldots, d$. *Then, as* $n \to \infty$, *we have*

$$\sup_{\mathbf{u}\in(0,1)^d} |\widehat{c}_n(\mathbf{u}) - c(\mathbf{u})| = O_{a.s.}\left(\left(\frac{\log n}{n}\right)^{(t-d/p)\gamma(d)}\right).$$

Proof. We begin with the following decomposition (see (3) and (4)):

$$\begin{aligned}
\widehat{c}_n(\mathbf{u}) - c(\mathbf{u}) &= \frac{\widehat{A}_n(\mathbf{u})}{\widehat{B}_n(\mathbf{u})} - \frac{A(\mathbf{u})}{B(\mathbf{u})} \\
&= \frac{(\widehat{A}_n(\mathbf{u}) - A(\mathbf{u}))}{B(\mathbf{u})} - \frac{(\widehat{B}_n(\mathbf{u}) - B(\mathbf{u}))A(\mathbf{u})}{B(\mathbf{u})^2} \\
&\quad - \frac{(\widehat{A}_n(\mathbf{u}) - A(\mathbf{u}))(\widehat{B}_n(\mathbf{u}) - B(\mathbf{u}))}{\widehat{B}_n(\mathbf{u})B(\mathbf{u})} + \frac{(\widehat{B}_n(\mathbf{u}) - B(\mathbf{u}))^2 A(\mathbf{u})}{\widehat{B}_n(\mathbf{u})B(\mathbf{u})^2} \\
&=: T_{1n}(\mathbf{u}) + T_{2n}(\mathbf{u}) + T_{3n}(\mathbf{u}) + T_{4n}(\mathbf{u}).
\end{aligned} \tag{5}$$

Applying Taylor's expansion (mean-value theorem), the fact that $f(\mathbf{x})$ has uniformly bounded first-order partial derivatives, Theorem 2 of Guo and Kou (2019), the boundedness of $f(\mathbf{x})$, and Condition 3, it follows that for some finite constant $C > 0$,

$$\begin{aligned}
&\sup_{\mathbf{u}\in(0,1)^d}\left|\widehat{A}_n(\mathbf{u}) - A(\mathbf{u})\right| \\
&\leq \sup_{\mathbf{u}\in(0,1)^d} |\widehat{f}_{j(d)}(\widehat{Q}_1(u_1),\ldots,\widehat{Q}_d(u_d)) - f(\widehat{Q}_1(u_1),\ldots,\widehat{Q}_d(u_d))| \\
&\quad + \sup_{\mathbf{u}\in(0,1)^d} |f(\widehat{Q}_1(u_1),\ldots,\widehat{Q}_d(u_d)) - f(Q_1(u_1),\ldots,Q_d(u_d))| \\
&\leq \sup_{\mathbf{y}\in I^d} |\widehat{f}_{j(d)}(\mathbf{y}) - f(\mathbf{y})| + C\sum_{\ell=1}^{d} \sup_{u_\ell\in(0,1)} |\widehat{Q}_\ell(u_\ell) - Q_\ell(u_\ell)| \\
&= O_{a.s.}\left(\left(\frac{\log n}{n}\right)^{(t-d/p)\gamma(d)}\right) + O_{a.s.}\left(\frac{(\log n)^\delta}{n^{1/2}}\right) \\
&= O_{a.s.}\left(\left(\frac{\log n}{n}\right)^{(t-d/p)\gamma(d)}\right),
\end{aligned} \tag{6}$$

for all $\delta > 0$, since $(t - d/p)\gamma(d) < 1/2$. (Note that Guo and Kou (2019) considered the case $a = 0$ and $b = 1$. However, their proof can immediately be generalised to have $\operatorname{supp} f(\mathbf{x}) = [a, b]^d$ by simply applying the transformation $Y = (X - a)/(b - a)$).

This, together with the fact that $\inf_{x_\ell\in I} f_\ell(x_\ell) > 0$ for $\ell = 1,\ldots,d$ (implying that $\inf_{\mathbf{u}\in(0,1)^d} B(\mathbf{u}) > 0$), allows us to conclude that

$$\sup_{\mathbf{u}\in(0,1)^d} |T_{1n}(\mathbf{u})| = O_{a.s.}\left(\left(\frac{\log n}{n}\right)^{(t-d/p)\gamma(d)}\right). \tag{7}$$

To take care of the second term $T_{2n}(\mathbf{u})$, we first recall the fact that for any sequences $\{a_\ell : 1 \leq \ell \leq d\}$ and $\{b_\ell : 1 \leq \ell \leq d\}$ of real numbers, we have

$$\prod_{\ell=1}^{d} a_\ell - \prod_{\ell=1}^{d} b_\ell = \sum_{\ell=1}^{d}(a_\ell - b_\ell)\prod_{i=1}^{\ell-1} b_i \prod_{h=\ell+1}^{d} a_h, \tag{8}$$

with the product on the empty set being equal to one. See, for instance, Lemma 5.1 in Bouzebda et al. (2011). Now, since $\widehat{f}_{h,j(1)}(\widehat{Q}_h(u_h))$ converges almost surely to $f_h(Q_h(u_h))$ for each $h = 1,\ldots,d$, we can write for n large enough

$$\widehat{f}_{h,j(1)}(\widehat{Q}_h(u_h)) = f_h(Q_h(u_h)) + o_{a.s.}(1).$$

It follows from this and (8) that

$$\begin{aligned}
\widehat{B}_n(\mathbf{u}) - B(\mathbf{u}) &= \prod_{\ell=1}^{d} \widehat{f}_{\ell,j(1)}(\widehat{Q}_\ell(u_\ell)) - \prod_{\ell=1}^{d} f_\ell(Q_\ell(u_\ell)) \\
&= \sum_{\ell=1}^{d} \left\{ \widehat{f}_{\ell,j(1)}(\widehat{Q}_\ell(u_\ell)) - f_\ell(Q_\ell(u_\ell)) \right\} \prod_{i=1}^{\ell-1} f_i(Q_i(u_i)) \prod_{h=\ell+1}^{d} \widehat{f}_{h,j(1)}(\widehat{Q}_h(u_h)) \\
&= \sum_{\ell=1}^{d} \left\{ \widehat{f}_{\ell,j(1)}(\widehat{Q}_\ell(u_\ell)) - f_\ell(Q_\ell(u_\ell)) \right\} \left\{ \prod_{i\neq\ell}^{d} f_i(Q_i(u_i)) + o_{a.s.}(1) \right\}.
\end{aligned}$$

Therefore, since f_ℓ is bounded for $\ell = 1, \dots, d$, we have for some finite constant $C > 0$ that

$$\sup_{\mathbf{u}\in(0,1)^d} |\widehat{B}_n(\mathbf{u}) - B(\mathbf{u})| \le C \sum_{\ell=1}^{d} \sup_{y\in I} |\widehat{f}_{\ell,j(1)}(y) - f_\ell(y)| + o_{a.s.}\left(\sum_{\ell=1}^{d} \sup_{y\in I} |\widehat{f}_{\ell,j(1)}(y) - f_\ell(y)| \right).$$

That is,

$$\sup_{\mathbf{u}\in(0,1)^d} |\widehat{B}_n(\mathbf{u}) - B(\mathbf{u})| = O_{a.s.}\left(\left(\frac{\log n}{n} \right)^{(t-1/p)\gamma(1)} \right), \tag{9}$$

by applying Theorem 2 of Guo and Kou (2019) with $d = 1$. Since $\inf_{\mathbf{u}\in(0,1)^d} B^2(\mathbf{u}) > 0$, this readily implies that

$$\sup_{\mathbf{u}\in(0,1)^d} |T_{2n}(\mathbf{u})| = O_{a.s.}\left(\left(\frac{\log n}{n} \right)^{(t-1/p)\gamma(1)} \right). \tag{10}$$

For the third term $T_{3n}(\mathbf{u})$, we can write, in view of (9) and for sufficiently large n, that

$$\begin{aligned}
|T_{3n}(\mathbf{u})| = \left| \frac{(\widehat{A}_n(\mathbf{u}) - A(\mathbf{u}))(\widehat{B}_n(\mathbf{u}) - B(\mathbf{u}))}{\widehat{B}_n(\mathbf{u})B(\mathbf{u})} \right| &= \left| \frac{(\widehat{A}_n(\mathbf{u}) - A(\mathbf{u}))}{B(\mathbf{u})} \left(1 - \frac{B(\mathbf{u})}{\widehat{B}_n(\mathbf{u})} \right) \right| \\
&\le \left| \frac{(\widehat{A}_n(\mathbf{u}) - A(\mathbf{u}))}{B(\mathbf{u})} \right|.
\end{aligned}$$

Since $\inf_{\mathbf{u}\in(0,1)^d} B(\mathbf{u}) > 0$, it follows from (6) that

$$\sup_{\mathbf{u}\in(0,1)^d} |T_{3n}(\mathbf{u})| = O_{a.s.}\left(\left(\frac{\log n}{n} \right)^{(t-d/p)\gamma(d)} \right). \tag{11}$$

The last term $T_{4n}(\mathbf{u})$ is handled by using the same arguments. We have for sufficiently large n that

$$\begin{aligned}
|T_{4n}(\mathbf{u})| = \left| \frac{(\widehat{B}_n(\mathbf{u}) - B(\mathbf{u}))^2 A(\mathbf{u})}{\widehat{B}_n(\mathbf{u})B(\mathbf{u})^2} \right| &= \left| \left(1 - \frac{B(\mathbf{u})}{\widehat{B}_n(\mathbf{u})} \right) \frac{(\widehat{B}_n(\mathbf{u}) - B(\mathbf{u}))A(\mathbf{u})}{B(\mathbf{u})^2} \right| \\
&\le \left| (\widehat{B}_n(\mathbf{u}) - B(\mathbf{u})) \frac{A(\mathbf{u})}{B(\mathbf{u})^2} \right|.
\end{aligned}$$

The ratio $A(\mathbf{u})/B(\mathbf{u})^2$ is bounded due to the fact that the density $f(\mathbf{x})$ is bounded and $\inf_{\mathbf{u}\in(0,1)^d} B(\mathbf{u})^2 > 0$. Then it follows from (9) that

$$\sup_{\mathbf{u}\in(0,1)^d} |T_{4n}(\mathbf{u})| = O_{a.s.}\left(\left(\frac{\log n}{n} \right)^{(t-1/p)\gamma(1)} \right). \tag{12}$$

Finally, we conclude from (7), (10), (11) and (12) that

$$\begin{aligned}\sup_{\mathbf{u}\in(0,1)^d} |\widehat{c}_n(\mathbf{u}) - c(\mathbf{u})| &= O_{a.s.}\left(\left(\frac{\log n}{n}\right)^{(t-d/p)\gamma(d)}\right) + O_{a.s.}\left(\left(\frac{\log n}{n}\right)^{(t-1/p)\gamma(1)}\right)\\ &= O_{a.s.}\left(\left(\frac{\log n}{n}\right)^{(t-d/p)\gamma(d)}\right),\end{aligned}$$

since $(t-1/p)\gamma(1) \geq (t-d/p)\gamma(d)$ for $d \geq 1$. ■

Remark 2. Our result is the same as the convergence rate derived by Masry (1997) and Guo and Kou (2019). It also coincides with the convergence rate in Theorem 3 of Giné and Nickl (2009) when $d = 1$ and $p = q = \infty$.

This result also improves the convergence rate achieved by Seck and Mamane (2024) under different assumptions which are potentially satisfied by different classes of wavelet bases.

References

Autin, F., Le Pennec, E., and Tribouley, K. (2010). Thresholding methods to estimate copula density. *Journal of Multivariate Analysis*, **101**, 200–222.

Bhatti, M. I. and Do, H. Q. (2019). Recent development in copula and its applications to the energy, forestry and environmental sciences. *International Journal of Hydrogen Energy*, **44**, 19453–19473.

Bouzebda, S., El Faouzi, N. E., and Zari, T. (2011). On the multivariate two-sample problem using strong approximations of empirical copula processes. *Communications in Statistics – Theory and Methods*, **40**, 1490–1509.

Chen, S. and Huang, T.-M. (2007). Nonparametric estimation of copula functions for dependence modeling. *Canadian Journal of Statistics*, **35**, 265–282.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. Number 61 in CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM).

Durante, F. and Sempi, C. (2010). Copula theory: An introduction. *In* Jaworski, P., Durante, F., Härdle, W. K., and Rychlik, T. (Editors) *Copula Theory and Its Applications*, volume 198 of *Lecture Notes in Statistics*. Springer, New York, NY, 3–31.

Fermanian, J.-D. and Scaillet, O. (2003). Nonparametric estimation of copulas for time series. *Journal of Risk*, **5**, 25–54.

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Université de Lyon. Série A*, **14**, 53–77.

Gannoun, A. and Hosseinioun, N. (2012). On wavelet-based methods for estimating the copula function. *Rev. Roumaine Math. Pures Appl.*, **57**, 205–213.

Geenens, G., Charpentier, A., and Paindaveine, D. (2017). Probit transformation for nonparametric kernel estimation of the copula density. *Bernoulli*, **23**, 1848–1873.

Genest, C., Masiello, E., and Tribouley, K. (2009). Estimating copula densities through wavelets. *Insurance: Mathematics and Economics*, **44**, 170–181.

Gijbels, I. and Mielniczuk, J. (1990). Estimating the density of a copula function. *Communications in Statistics – Theory and Methods*, **19**, 445–464.

GINÉ, E. AND NICKL, R. (2009). Uniform limit theorems for wavelet density estimators. *The Annals of Probability*, **37**, 1605–1646.

GRÖẞER, J. AND OKHRIN, O. (2022). Copulae: An overview and recent developments. *Wiley Interdisciplinary Reviews: Computational Statistics*, **14**, e1557.

GUO, H. AND KOU, J. (2019). Strong uniform convergence rates of wavelet density estimators with size-biased data. *Journal of Function Spaces*.

HÄRDLE, W., KERKYACHARIAN, G., PICARD, D., AND TSYBAKOV, A. (1998). *Wavelets, Approximation, and Statistical Applications*, volume 129 of *Lecture Notes in Statistics*. Springer, New York, NY.

HOEFFDING, W. (1940). Masstabinvariante Korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, **5**, 179–233. Reprinted as: Scale-invariant correlation theory. In: Fisher, N. I. and Sen, P. K., editors, (1994). The Collected Works of Wassily Hoeffding, 57–107. Springer, New York, NY.

HOEFFDING, W. (1941). Masstabinvariante Korrelationsmasse für diskontinuierliche Verteilungen. *Arkiv für Matematischen Wirtschaften und Sozialforschung*, **7**, 49–70. Reprinted as: Scale-invariant correlation measures for discontinuous distributions. In: Fisher, N. I. and Sen, P. K., editors, (1994). The Collected Works of Wassily Hoeffding, 109–133. Springer, New York, NY.

JAWORSKI, P., DURANTE, F., HÄRDLE, W. K., AND RYCHLIK, T. (2010). *Copula Theory and Its Applications*, volume 198. Springer, New York, NY.

MALLAT, S. (2009). *A Wavelet Tour of Signal Processing*. 3rd edition. Elsevier, Amsterdam.

MANNER, H. AND REZNIKOVA, O. (2012). A survey on time-varying copulas: Specification, simulations, and application. *Econometric Reviews*, **31**, 654–687.

MASRY, E. (1997). Multivariate probability density estimation by wavelet methods: strong consistency and rates for stationary time series. *Stochastic Processes and their Applications*, **67**, 177–193.

MASRY, E. (2000). Wavelet-based estimation of multivariate regression functions in Besov spaces. *Journal of Nonparametric Statistics*, **12**, 283–308.

MEYER, Y. (1992). *Wavelets and Operators*, volume 1. Cambridge University Press, New York, NY.

NELSEN, R. B. (2003). Properties and applications of copulas: A brief survey. *In Proceedings of the First Brazilian Conference on Statistical Modeling in Insurance and Finance*. University Press USP Sao Paulo, 10–28.

OMELKA, M., GIJBELS, I., AND VERAVERBEKE, N. (2009). Improved kernel estimators of copulas: weak convergence and goodness-of-fit testing. *The Annals of Statistics*, **37**, 3023–3058.

PATTON, A. J. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis*, **110**, 4–18. Special Issue on Copula Modeling and Dependence.

SECK, C. T. AND MAMANE, S. (2024). Strong uniform convergence rates of the linear wavelet estimator of a multivariate copula density. *South African Statistical Journal*, **58**, 35–56.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley and Sons, New York, NY.

SKLAR, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, **8**. https://hal.science/hal-04094463.

VIDAKOVIC, B. (1999). *Statistical Modeling by Wavelets*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley and Sons, New York, NY.

Manuscript received 2024-03-29, revised 2024-08-16, accepted 2024-08-16.