South African Statistical Journal
Vol. 58, No. 1, pp. 35–56
https://doi.org/10.37920/sasj.2024.58.1.3

Strong uniform convergence rates of the linear wavelet estimator of a multivariate copula density

Cheikh Tidiane Seck[1] *and Salha Mamane*[2]

[1]Department of Mathematics, Alioune Diop University, Bambey, Senegal
[2]School of Statistics and Actuarial Science, University of the Witwatersrand, Johannesburg, South Africa

In this paper, we investigate the almost sure convergence, in supremum norm, of the rank-based linear wavelet estimator for the multivariate copula density over Besov classes. Using empirical process tools, we establish a uniform limit law for the deviation of an oracle estimator (which assumes known margins) from its expectation. This enables us to derive strong convergence rates for the rank-based linear estimator.

Keywords: Almost sure uniform convergence rates, Copula density, Nonparametric estimation, Wavelet methods.

1. Introduction

A copula is a multivariate distribution function C defined on $[0,1]^d$, $d \geq 2$, with uniform margins. Unlike the linear correlation coefficient, it gives a full characterisation of the dependence between random variables, be it linear or nonlinear. Given a vector $(X_1, \ldots, X_d)$ of continuous random variables with marginal distribution functions $F_1, \ldots, F_d$, the copula C may be defined as the joint cumulative distribution function of the random vector $(F_1(X_1), \ldots, F_d(X_d))$. If it exists, the copula density is defined as the derivative, c, of the copula distribution function C with respect to the Lebesgue measure:

$$c(u_1 \ldots, u_d) = \frac{\partial^d}{\partial u_1 \cdots \partial u_d} C(u_1, \ldots, u_d), \;\; \forall\, (u_1, \ldots, u_d) \in (0,1)^d.$$

Nonparametric estimation of a copula density is an active reseach domain that has been investigated by many authors. For instance, Gijbels and Mielniczuk (1990) and Fermanian and Scaillet (2003) used convolution kernel methods to construct consistent estimators for the copula density, while Sancetta and Satchell (2004) employed techniques based on Bernstein polynomials. A drawback of kernel methods is the existence of boundary effects due to the compact support of the copula function. To overcome this difficulty, some approaches have been proposed. For example Gijbels and Mielniczuk (1990) used a mirror-reflexion technique, while Chen and Huang (2007) employed a local linear kernel procedure. In the same vein, Omelka et al. (2009) proposed improved copula

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

kernel estimators to mitigate the boundary bias problem. Recently, Geenens et al. (2017) introduced kernel-type estimators for the copula density, based on a probit transformation method that can take care of the boundary effects.

In this paper, we deal more neatly with the boundary bias problem by using wavelet methods, which are very convenient to describe features of functions at the edges and corners of the unit cube, because of their good localisation properties. Indeed, wavelet bases automatically handle the boundary effects by locally adapting to the properties of the curve being estimated. The use of wavelet methods in density and regression estimation problems is surveyed in Härdle et al. (1998), where approximation properties of wavelets are discussed at length. For more details on wavelet theory we refer to Meyer (1992), Daubechies (1992), Mallat (2009), and Vidakovic (1999) and references therein.

Wavelet methods have been already used in nonparametric copula density estimation. For instance, Genest et al. (2009) dealt with a rank-based linear wavelet estimator of the bivariate copula density and established, under certain conditions, its optimality in the minimax sense on Besov-balls for the L_2-norm loss, as well as on Hölder-balls for the pointwise-norm loss. Autin et al. (2010) extended these results to the nonlinear thresholded estimators of multivariate copula densities. These nonlinear estimates are near optimal (up to a logarithmic factor) for the L_2-norm loss, and have the advantage of being adaptive to the regularity of the copula density function. In a similar vein, Gannoun and Hosseinioun (2012) established an upper bound on L_p-losses, $2 \leq p < \infty$, for linear wavelet-based estimators of the bivariate copula density, when this latter is bounded.

Our goal in this paper, is to establish almost sure convergence rates, in supremum norm loss, for the linear wavelet estimator of the multivariate copula density. Our methodology of proof is inspired by Giné and Nickl (2009), who provided almost sure convergence rates, in supremum norm loss, for the linear wavelet estimator of a univariate density function on $\mathbb{R}$. Here, we want to extend this result to a multivariate copula density on $(0,1)^d$. In fact, we prove that under the condition of sufficient regularity of the multivariate copula density c (i.e., c belongs to the Besov space of regularity t, denoted as $B^t_{\infty,\infty}((0,1)^d)$ and corresponding to the Hölder space of order t) and the resolution level, say j_n, satisfies: $2^{j_n} \simeq (n/\log n)^{1/(2t+d)}$, then the rank-based linear wavelet estimator of c converges almost surely, in supremum norm, with a rate of the order $O((\log n/n)^{[2(t-1)-d]/[2(2t+d)]})$. Moreover, we show that, in contrast, the oracle estimator (obtained for known margins) attains the optimal convergence rate which is $O((\log n/n)^{t/(2t+d)})$.

The rest of the paper is organised as follows. In Section 2, we recall some facts on wavelet theory and define the rank-based linear wavelet estimator of the multivariate copula density as in Autin et al. (2010). Section 3 presents the main theoretical results along with some comments. In Appendix A, we recall some useful facts on empirical process theory. Appendix B contains the proof of the uniform limit law given in Proposition 1.

2. Wavelet theory and Estimation procedure

Let ϕ be a father wavelet and ψ its associated mother wavelet, which are both assumed compactly supported. Cohen et al. (1993) proposed orthonormal wavelet bases for $L_2([0,1])$, the space of all measurable and square integrable functions on $[0,1]$. Precisely, for all fixed $j_0 \in \mathbb{N}$, the family $\{\phi_{j_0,l} : l = 1,\ldots,2^{j_0}\} \bigcup \{\psi_{j,l} : j \geq j_0, l = 1,\ldots,2^j\}$ is an orthonormal basis for $L_2([0,1])$, where $\phi_{j,l}(u) = 2^{j/2}\phi(2^j u - l)$ and $\psi_{j,l}(u) = 2^{j/2}\psi(2^j u - l)$, $\forall j,l \in \mathbb{Z}$, $u \in [0,1]$. Using the tensorial product, one can construct a multivariate wavelet basis for $L_2([0,1]^d)$, $d \geq 2$. For $\mathbf{k} = (k_1, \ldots, k_d) \in \mathbb{Z}^d$, define the following functions of $\mathbf{u} = (u_1, \ldots, u_d) \in [0,1]^d$:

$$\phi_{j_0,\mathbf{k}}(u_1, \ldots, u_d) = \prod_{m=1}^{d} \phi_{j_0,k_m}(u_m), \tag{1}$$

$$\psi^{\epsilon}_{j,\mathbf{k}}(u_1, \ldots, u_d) = \prod_{m=1}^{d} \phi^{1-\epsilon_m}_{j,k_m}(u_m)\psi^{\epsilon_m}_{j,k_m}(u_m),$$

where $\epsilon = (\epsilon_1, \ldots, \epsilon_d) \in \mathcal{S}_d = \{0,1\}^d \setminus \{(0, \ldots, 0)\}$. Then the family $\{\phi_{j_0,\mathbf{k}}, \psi^{\epsilon}_{j,\mathbf{h}} : j \geq j_0, \mathbf{k} \in \{1, \ldots, 2^{j_0}\}^d, \mathbf{h} \in \{1, \ldots, 2^j\}^d, \epsilon \in \mathcal{S}_d\}$ is an orthonormal basis for $L_2([0,1]^d)$, for any fixed $j_0 \in \mathbb{N}$. Thus, assuming that the copula density c belongs to $L_2([0,1]^d)$, we have the following representation:

$$c(\mathbf{u}) = \sum_{\mathbf{k} \in \{1,\ldots,2^{j_0}\}^d} \alpha_{j_0,\mathbf{k}} \phi_{j_0,\mathbf{k}}(\mathbf{u}) + \sum_{j \geq j_0} \sum_{\mathbf{k} \in \{1,\ldots,2^j\}^d} \sum_{\epsilon \in \mathcal{S}_d} \beta^{\epsilon}_{j,\mathbf{k}} \psi^{\epsilon}_{j,\mathbf{k}}(\mathbf{u}), \tag{2}$$

for all $\mathbf{u} \in [0,1]^d$, where the scaling coefficients $\alpha_{j_0,\mathbf{k}}$ and wavelet coefficients $\beta^{\epsilon}_{j,\mathbf{k}}$ are respectively defined as

$$\alpha_{j_0,\mathbf{k}} = \int_{[0,1]^d} c(\mathbf{u})\phi_{j_0,\mathbf{k}}(\mathbf{u})d\mathbf{u} \quad \text{and} \quad \beta^{\epsilon}_{j,\mathbf{k}} = \int_{[0,1]^d} c(\mathbf{u})\psi^{\epsilon}_{j,\mathbf{k}}(\mathbf{u})d\mathbf{u}.$$

Now, let $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ be an independent and identically distributed (i.i.d) sample of the random vector $\mathbf{X} = (X_1, \ldots, X_d)$, with continuous marginal distribution functions $F_1, \ldots, F_d$, and where $\mathbf{X}_i = (X_{i1}, \ldots, X_{id})$, $i = 1, \ldots, n$. The distribution function of the random vector $\mathbf{U}_i = (F_1(X_{i1}), \ldots, F_d(X_{id}))$ is the copula C and its density, if it exists, is c. Denoting the expectation operator by $\mathbb{E}$, the coefficients $\alpha_{j_0,\mathbf{k}}$ and $\beta^{\epsilon}_{j,\mathbf{k}}$ can be rewritten as follows:

$$\alpha_{j_0,\mathbf{k}} = \mathbb{E}[\phi_{j_0,\mathbf{k}}(\mathbf{U}_i)], \qquad \beta^{\epsilon}_{j,\mathbf{k}} = \mathbb{E}[\psi^{\epsilon}_{j,\mathbf{k}}(\mathbf{U}_i)].$$

If the margins $F_1, \ldots, F_d$ were known, natural estimators for $\alpha_{j_0,\mathbf{k}}$ and $\beta^{\epsilon}_{j,\mathbf{k}}$ would be given by

$$\tilde{\alpha}_{j_0,\mathbf{k}} = \frac{1}{n}\sum_{i=1}^{n} \phi_{j_0,\mathbf{k}}(\mathbf{U}_i), \qquad \tilde{\beta}^{\epsilon}_{j,\mathbf{k}} = \frac{1}{n}\sum_{i=1}^{n} \psi^{\epsilon}_{j,\mathbf{k}}(\mathbf{U}_i).$$

But, usually the marginal distribution functions $F_1, \ldots, F_d$ are unknown; and it is customary to replace them by their empirical counterparts $F_{1n}, \ldots, F_{dn}$ (or rescaled versions thereof), with

$$F_{jn}(x_j) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(X_{ij} \leq x_j), \qquad j = 1, \ldots, d, \quad x_j \in \mathbb{R},$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. Then, putting $\hat{\mathbf{U}}_i = (\hat{U}_{i1}, \ldots, \hat{U}_{id})$, where $\hat{U}_{ij} = F_{jn}(X_{ij})$, $j = 1, \ldots, d$; $i = 1, \ldots, n$, the modified empirical coefficients are

$$\hat{\alpha}_{j_0,\mathbf{k}} = \frac{1}{n}\sum_{i=1}^{n} \phi_{j_0,\mathbf{k}}(\hat{\mathbf{U}}_i), \qquad \hat{\beta}^{\epsilon}_{j,\mathbf{k}} = \frac{1}{n}\sum_{i=1}^{n} \psi^{\epsilon}_{j,\mathbf{k}}(\hat{\mathbf{U}}_i).$$

Now, choosing a suitable resolution level $j_n \geq j_0$ and considering the orthogonal projection of c onto the sub-space V_{j_n} of the underlying multiresolution analysis on $L_2([0,1]^d)$, we obtain the rank-based linear wavelet estimator of c:

$$\hat{c}_{j_n}(\mathbf{u}) = \sum_{\mathbf{k}\in\{1,\ldots,2^{j_n}\}^d} \hat{\alpha}_{j_n,\mathbf{k}}\phi_{j_n,\mathbf{k}}(\mathbf{u}), \qquad \mathbf{u} \in (0,1)^d. \tag{3}$$

As remarked in Genest et al. (2009), the estimator $\hat{c}_{j_n}$ is not necessarily a density, because it can take negative values on parts of its domain and fails to integrate to 1. In practice, some truncations and normalisations are necessary for its use.

To obtain the strong convergence rate of the linear estimator $\hat{c}_{j_n}$, our methodology of proof follows the empirical process approach developed in Einmahl and Mason (2000); see also Giné and Nickl (2009), Giné and Guillou (2002). In fact, we can rewrite $\hat{c}_{j_n}$ in terms of the empirical measure. Let us define the following kernel functions: for all $(x,y) \in \mathbb{R}^2$,

$$\begin{aligned} \widetilde{K}(x,y) &= \sum_{l=1}^{2^{j_n}} \phi(x-l)\phi(y-l), \qquad (4)\\ \widetilde{K}_{j_n}(x,y) &= 2^{j_n}\widetilde{K}(2^{j_n}x, 2^{j_n}y). \end{aligned}$$

For $\mathbf{x} = (x_1,\ldots,x_d) \in \mathbb{R}^d,\ \mathbf{y} = (y_1,\ldots,y_d) \in \mathbb{R}^d$,

$$\mathbf{K}(\mathbf{x},\mathbf{y}) = \prod_{m=1}^{d} \widetilde{K}(x_m, y_m),$$

$$\mathbf{K}_{j_n}(\mathbf{x},\mathbf{y}) = \prod_{m=1}^{d} \widetilde{K}_{j_n}(x_m, y_m).$$

Then, the linear wavelet estimator $\hat{c}_{j_n}$ can be rewritten as

$$\hat{c}_{j_n}(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{K}_{j_n}(\hat{\mathbf{U}}_i, \mathbf{u}) = \frac{2^{dj_n}}{n}\sum_{i=1}^{n} \mathbf{K}(2^{j_n}\hat{\mathbf{U}}_i, 2^{j_n}\mathbf{u}). \tag{5}$$

3. Asymptotic behaviour of the estimator

Let us introduce an auxillary estimator $\tilde{c}_{j_n}$ corresponding to the case where the marginal distribution functions $F_1,\ldots,F_d$ are known. In this situation, $(U_{i1},\ldots,U_{id}) = (F_1(X_{i1}),\ldots,F_d(X_{id}))$, $i = 1,\ldots,n$, are direct observations of the copula C, and $\tilde{c}_{j_n}$ may be defined as

$$\tilde{c}_{j_n}(\mathbf{u}) = \sum_{\mathbf{k}\in\{1,\ldots,2^{j_n}\}^d} \tilde{\alpha}_{j_n,\mathbf{k}}\phi_{j_n,\mathbf{k}}(\mathbf{u}), \tag{6}$$

where

$$\tilde{\alpha}_{j_n,\mathbf{k}} = \frac{1}{n}\sum_{i=1}^{n} \phi_{j_n,\mathbf{k}}(F_1(X_{i1}),\ldots,F_d(X_{id}))$$

is an unbiased estimator of $\alpha_{j_n,\mathbf{k}}$.

For all $\mathbf{u} \in (0,1)^d$, we can decompose the estimation error $\hat{c}_{j_n} - c$ as

$$\begin{aligned}\hat{c}_{j_n}(\mathbf{u}) - c(\mathbf{u}) &= [\hat{c}_{j_n}(\mathbf{u}) - \tilde{c}_{j_n}(\mathbf{u})] + [\tilde{c}_{j_n}(\mathbf{u}) - \mathbb{E}\tilde{c}_{j_n}(\mathbf{u})] + [\mathbb{E}\tilde{c}_{j_n}(\mathbf{u}) - c(\mathbf{u})] \\ &=: R_n(\mathbf{u}) + D_n(\mathbf{u}) + B_n(\mathbf{u}). \end{aligned} \tag{7}$$

To obtain the almost sure convergence rate of $\hat{c}_{j_n}$ uniformly in $\mathbf{u} \in (0,1)^d$, we have to investigate the limiting behaviour of each of the three above terms. We need the following assumptions in the sequel :

(H.1) The father wavelet $\phi \in L^2(\mathbb{R})$ is bounded, compactly supported and admits a bounded derivative ϕ'.

(H.2) There exists a bounded and compactly supported function $\Phi : \mathbb{R} \longrightarrow \mathbb{R}_+$ such that $|\widetilde{K}(x,y)| \leq \Phi(x-y), \forall x, y \in \mathbb{R}$ and the function $\theta_\phi(\cdot) = \sum_{k=1}^{2^{j_n}} |\phi(\cdot - k)|$ is bounded.

(H.3) The kernel $\widetilde{K}$ satisfies, for all $y \in \mathbb{R}$, $\int_{-\infty}^{\infty} \widetilde{K}(x,y)dx = 1$.

(H.4) As $n \longrightarrow \infty$, the sequence $(j_n)_{n\geq 0}$ satisfies

$$j_n \nearrow \infty, \qquad \frac{n}{j_n 2^{2(d+1)j_n}} \longrightarrow \infty, \qquad \frac{j_n}{\log\log n} \longrightarrow \infty.$$

Remark 1. Assumptions (H.1), (H.2) and (H.3) are usual conditions that are satisfied by many wavelets bases, for example the Daubechies wavelets and the Haar wavelet $\phi(u) = 1_{[0,1]}(u)$. The conditions in Assumption (H.4) are analogous to some conditions imposed on the bandwidth parameter in convolution-kernel estimation methods.

The following proposition gives the asymptotic behaviour of the second term $D_n(\mathbf{u})$ in (7), corresponding to the deviation of the auxillary estimator $\tilde{c}_{j_n}$ from its expectation. In the sequel, we denote $I = (0,1)$, $\|c\|_\infty = \sup_{\mathbf{u}\in I^d} |c(\mathbf{u})|$ and for any bounded real function φ defined on $\mathbb{R}^d$, $d \geq 1$, $\|\varphi\|_\infty = \sup_{\mathbf{x}\in\mathbb{R}^d} |\varphi(\mathbf{x})|$.

Proposition 1. *Suppose that Assumptions (H.1–4) hold and that the father wavelet ϕ is uniformly continuous with support $[0, B]$, B being a positive integer. If, moreover, the copula density c is continuous and bounded on I^d, then we have almost surely (a.s.),*

$$\lim_{n\to\infty} r_n \sup_{\mathbf{u}\in I^d} \frac{|\tilde{c}_{j_n}(\mathbf{u}) - \mathbb{E}\tilde{c}_{j_n}(\mathbf{u})|}{\sqrt{\int_{\mathbb{R}^d} \mathbf{K}^2(\mathbf{x}, 2^{j_n}\mathbf{u})d\mathbf{x}}} = \sqrt{\|c\|_\infty}, \tag{8}$$

with

$$r_n = \sqrt{\frac{n}{(2d\log 2)j_n 2^{dj_n}}}. \tag{9}$$

Proof. It is largely inspired by Giné and Nickl (2009) and is postponed to Appendix B. It will consist of establishing a lower bound and an upper bound for the limit in (8), a methodology borrowed from Einmahl and Mason (2000); see also Giné and Guillou (2002). ∎

Remark 2. Proposition 1 gives the exact almost sure convergence rate, in supremum norm, of the deviation D_n to zero, which is of order $O(\sqrt{j_n 2^{dj_n}/n})$. In fact, by Assumptions (H.1), (H.2) and

(H.3), the quantity $\int_{\mathbb{R}^d} \mathbf{K}^2(\mathbf{x}, 2^{j_n}\mathbf{u})d\mathbf{x}$ can be bounded: there exist two positive constants D_1 and D_2 independent of $\mathbf{u}$ and n such that

$$D_1 \leq \int_{\mathbb{R}^d} \mathbf{K}^2(\mathbf{x}, 2^{j_n}\mathbf{u})d\mathbf{x} \leq D_2. \tag{10}$$

This readily implies

$$\sup_{\mathbf{u}\in I^d} |D_n(\mathbf{u})| = O_{a.s.}\left(\sqrt{\frac{j_n 2^{d j_n}}{n}}\right). \tag{11}$$

The following theorem constitutes our principal result. We need some notation before stating it. Let N be a positive integer and $t = N + \alpha, 0 < \alpha \leq 1$. For any bounded real function f defined on I^d and possessing derivatives up to order N, set

$$\|f\|_{t,\infty,\infty} = \|f\|_\infty + \sum_{k=0}^{N} \sup_{u\neq v, u,v\in I^d} \frac{|f^{(k)}(u) - f^{(k)}(v)|}{|u-v|^\alpha}. \tag{12}$$

We say that f belongs to the Besov space of regularity t, $B^t_{\infty,\infty}(I^d)$, if and only if $\|f\|_{t,\infty,\infty} < \infty$.

The following condition is also needed for the proof:
Condition $1(N)$: the father wavelet ϕ admits weak derivatives up to order $N \in \mathbb{N}$, that are all in $\mathcal{L}^p(\mathbb{R}^d)$ for some $1 \leq p \leq \infty$.

Theorem 1. *Suppose that the assumptions of Proposition 1 are fulfilled. If, moreover, c belongs to* $B^t_{\infty,\infty}(I^d)$ *and* ϕ *satisfies Condition* $1(N)$, *with* $(d+2)/2 < t < N+1$, *then, if* $2^{j_n} \simeq (n/\log n)^{\frac{1}{2t+d}}$, *we have as* $n \longrightarrow \infty$,

$$\sup_{\mathbf{u}\in I^d} |\hat{c}_{j_n}(\mathbf{u}) - c(\mathbf{u})| = O_{a.s.}(\mathrm{R_n}), \tag{13}$$

where $\mathrm{R_n} = \sqrt{2^{2(1+d)j_n}(\log\log n)/n}$.

Proof. In view of decomposition (7), it suffices to handle the first and the last term. The behaviour of the second term $D_n(\mathbf{u})$ is given by the previous Proposition 1. Let us begin with the first term $R_n(\mathbf{u})$. We have, for $\mathbf{k} = (k_1, \ldots, k_d) \in \mathbb{Z}^d$,

$$\begin{aligned}\hat{\alpha}_{j_n,\mathbf{k}} - \tilde{\alpha}_{j_n,\mathbf{k}} &= \frac{1}{n}\sum_{i=1}^{n}\left[\phi_{j_n,\mathbf{k}}(F_{1n}(X_{i1}), \ldots, F_{dn}(X_{id})) - \phi_{j_n,\mathbf{k}}(F_1(X_{i1}), \ldots, F_d(X_{id}))\right] \\ &=: \frac{1}{n}\sum_{i=1}^{n} \xi_{\mathbf{k}}(X_{i1}, \ldots, X_{id}),\end{aligned}$$

where we set

$$\xi_{\mathbf{k}}(X_{i1}, \ldots, X_{id}) = \phi_{j_n,\mathbf{k}}(F_{1n}(X_{i1}), \ldots, F_{dn}(X_{id})) - \phi_{j_n,\mathbf{k}}(F_1(X_{i1}), \ldots, F_d(X_{id})).$$

For $d = 2$, by using the multiplicativity of $\phi_{j_n,\mathbf{k}}$ (see (1)), one can prove that (see also Genest et al., 2009) that, with $k = (k_1, k_2)$:

$$\xi_k(X_{i1}, X_{i2}) = \xi_{k_1}(X_{i1})\xi_{k_2}(X_{i2}) + \xi_{k_1}(X_{i1})\phi_{j_n k_2}(F_2(X_{i2})) + \xi_{k_2}(X_{i2})\phi_{j_n k_1}(F_1(X_{i1})), \tag{14}$$

where $\xi_{k_m}(X_{im}) = \phi_{j_n k_m}(F_{mn}(X_{im})) - \phi_{j_n k_m}(F_m(X_{im}))$, for $m = 1, 2$.

By induction of (14), we obtain for all fixed $d \geq 2$ that

$$\xi_{\mathbf{k}}(X_{i1}, \ldots, X_{id}) = \sum_{q=0}^{d-1} \sum_{\epsilon_1+\ldots+\epsilon_d=q} \prod_{m=1}^{d} \left[\xi_{k_m}(X_{im})\right]^{1-\epsilon_m} \left[\phi_{j_n k_m}(F_m(X_{im}))\right]^{\epsilon_m}, \tag{15}$$

where $(\epsilon_1, \ldots \epsilon_d) \in \{0, 1\}^d$. Recall that $\phi_{jl}(u) = 2^{j/2}\phi(2^j u - l), \forall j, l \in \mathbb{Z}$. By using the derivability of ϕ by hypothesis, we can write, for all $m = 1, \ldots, d$,

$$\begin{aligned} \xi_{k_m}(X_{im}) &= 2^{\frac{j_n}{2}}\phi(2^{j_n}F_{mn}(X_{im}) - k_m) - 2^{\frac{j_n}{2}}\phi(2^{j_n}F_m(X_{im}) - k_m) \\ &= 2^{\frac{3}{2}j_n}\left[F_{mn}(X_{im}) - F_m(X_{im})\right]\phi'(\zeta_{im}), \end{aligned}$$

where ζ_{im} lies between $F_{mn}(X_{im})$ and $F_m(X_{im})$. Now, combining the Chung's law of the iterated logarithm (Chung, 1949) with the boundedness of ϕ and ϕ', we obtain, for all $m = 1, \ldots, d$,

$$|\xi_{k_m}(X_{im})| \leq 2^{\frac{3}{2}j_n}\sqrt{\frac{\log\log n}{2n}}\|\phi'\|_\infty, \quad a.s.$$

Thus, for $d = 2$, the expression in (15) can be bounded above; that is

$$|\xi_k(X_{i1}, X_{i2})| \leq \left(2^{\frac{3}{2}j_n}\sqrt{\frac{\log\log n}{2n}}\|\phi'\|_\infty\right)^2 + 2\times 2^{\frac{3}{2}j_n}\sqrt{\frac{\log\log n}{2n}}\|\phi'\|_\infty \left(2^{\frac{j_n}{2}}\|\phi\|_\infty\right)^{2-1}, \quad a.s. \tag{16}$$

Since

$$\frac{2^{3j_n}\left(\frac{\log\log n}{2n}\right)}{2^{2j_n}\sqrt{\frac{\log\log n}{2n}}} = \frac{1}{\sqrt{2}}\left(\frac{j_n 2^{2j_n}}{n}\right)^{1/2}\left(\frac{\log\log n}{j_n}\right)^{1/2},$$

which, by (H.4), converges to 0 as $n \to \infty$, then $2^{3j_n}(\log\log n)/(2n) = o(2^{2j_n}\sqrt{(\log\log n)/(2n)})$. That is, for $d = 2$,

$$|\xi_k(X_{i1}, X_{i2})| = O_{a.s.}\left(2^{2j_n}\sqrt{\frac{\log\log n}{n}}\right).$$

By induction of formula (16), we get for all $d \geq 2$, with C_d^r the binomial coefficients,

$$\begin{aligned} |\xi_{\mathbf{k}}(X_{i1}, \ldots, X_{id})| \leq\ & C_d^0\left(2^{\frac{3}{2}j_n}\sqrt{\frac{\log\log n}{2n}}\|\phi'\|_\infty\right)^d \\ & + C_d^1\left(2^{\frac{3}{2}j_n}\sqrt{\frac{\log\log n}{2n}}\|\phi'\|_\infty\right)^{d-1}\left(2^{\frac{j_n}{2}}\|\phi\|_\infty\right) \\ & \ldots \\ & + C_d^{d-1}\left(2^{\frac{3}{2}j_n}\sqrt{\frac{\log\log n}{2n}}\|\phi'\|_\infty\right)\left(2^{\frac{j_n}{2}}\|\phi\|_\infty\right)^{d-1}. \end{aligned} \tag{17}$$

Note align the number of terms in the summation on the right-hand side of inequality (17) is equal to d. Moreover, as we observe in the case $d = 2$, all these terms are dominated (*small-o's*) by the last one, which is of order $O_{a.s.}(2^{\frac{3}{2}j_n}\sqrt{(\log\log n)/n}\,2^{\frac{d-1}{2}j_n})$. Then

$$|\xi_{\mathbf{k}}(X_{i1},\ldots,X_{id})| = O_{a.s.}\left(2^{\frac{2+d}{2}j_n}\sqrt{\frac{\log\log n}{n}}\right).$$

and

$$|\hat{\alpha}_{j_n,\mathbf{k}} - \tilde{\alpha}_{j_n,\mathbf{k}}| = \frac{1}{n}\sum_{i=1}^{n}|\xi_{\mathbf{k}}(X_{i1},\ldots,X_{id})| = O_{a.s.}\left(2^{\frac{2+d}{2}j_n}\sqrt{\frac{\log\log n}{n}}\right).$$

Finally, by using the boundedness of the function $\theta_\phi(x) = \sum_{l=1}^{2^{j_n}}|\phi(x-l)|$, we obtain

$$\begin{aligned}
|\hat{c}_{j_n,\mathbf{k}}(\mathbf{u}) - \tilde{c}_{j_n,\mathbf{k}}(\mathbf{u})| &\leq \sum_{\mathbf{k}\in\{1,\ldots,2^{j_n}\}^d}|\hat{\alpha}_{j_n,\mathbf{k}} - \tilde{\alpha}_{j_n,\mathbf{k}}|2^{\frac{d}{2}j_n}\prod_{m=1}^{d}\phi(2^{j_n}u_m - k_m)\\
&= O_{a.s.}\left[\|\theta_\phi\|_\infty^d 2^{\frac{2+2d}{2}j_n}\sqrt{\frac{\log\log n}{n}}\right]\\
&= O_{a.s.}\left[\left(\frac{j_n 2^{2(1+d)j_n}}{n}\right)^{1/2}\left(\frac{\log\log n}{j_n}\right)^{1/2}\right] =: O_{a.s.}(\mathrm{R_n}).
\end{aligned}$$

Hence,

$$\sup_{\mathbf{u}\in I^d}|R_n(\mathbf{u})| = O_{a.s.}(\mathrm{R_n}). \tag{18}$$

To handle the last term $B_n(\mathbf{u})$ corresponding to the bias of $\tilde{c}_{j_n}$, we make use of approximation properties in Besov spaces. Let K_{j_n} denote the orthogonal projection kernel onto the sub-space V_{j_n}. That is

$$K_{j_n}(c)(\mathbf{u}) = \int_{I^d}K_{j_n}(\mathbf{u},\mathbf{v})c(\mathbf{v})d\mathbf{v}, \quad \mathbf{u}\in I^d.$$

Then, we can write

$$B_n(\mathbf{u}) = \mathbb{E}\tilde{c}_{j_n}(\mathbf{u}) - c(\mathbf{u}) = K_{j_n}(c)(\mathbf{u}) - c(\mathbf{u}).$$

Since ϕ satisfies *Condition* $1(N)$ and $c \in B^t_{\infty,\infty}(I^d)$, $(d+2)/2 < t < N+1$, then Theorem 9.4 in Härdle et al. (1998) gives:

$$\|K_{j_n}(c) - c\|_\infty \leq A2^{-j_n t},$$

where A is a positive constant depending on the Besov norm of c. Hence

$$\sup_{\mathbf{u}\in I^d}|B_n(\mathbf{u})| = O(2^{-j_n t}). \tag{19}$$

In view of (11), (18) and (19), we can write

$$\sup_{\mathbf{u}\in I^d}|\hat{c}_{j_n}(\mathbf{u}) - c(\mathbf{u})| = O_{a.s.}\left(\sqrt{\frac{j_n 2^{dj_n}}{n}}\right) + O(2^{-j_n t}) + O_{a.s.}(\mathrm{R_n}).$$

Now, if $2^{j_n} \simeq (n/\log n)^{\frac{1}{2t+d}}$, the terms $\sqrt{j_n 2^{dj_n}/n}$ and $2^{-j_n t}$ are equivalent and are both less than $\mathrm{R_n}$, because $\sqrt{j_n 2^{dj_n}/n}/\mathrm{R_n} \to 0$, as $n \to \infty$. Thus,

$$\sup_{\mathbf{u}\in I^d} |\hat{c}_{j_n}(\mathbf{u}) - c(\mathbf{u})| = O_{a.s.}(\mathrm{R_n}),$$

which completes the proof of Theorem 1. ■

Remark 3. Note that $R_n = \sqrt{2^{2(1+d)j_n}(\log\log n)/n} \leq \sqrt{2^{2(1+d)j_n}(\log n)/n}$. Therefore, we can write $O_{a.s.}(R_n) = O_{a.s.}\left(\sqrt{2^{2(1+d)j_n}(\log n)/n}\right)$. Thus, Theorem 1 implies that if $2^{j_n} \simeq (n/\log n)^{1/(2t+d)}$ then the rank-based linear wavelet estimator $\hat{c}_{j_n}$ converges almost surely to c, in supremum norm, with a convergence rate of the order of $(\log n/n)^{(2(t-1)-d)/(2(2t+d))}$. One can note that this rate is weaker than $(\log n/n)^{t/(2t+d)}$, which is the best attainable rate for this norm; see, e.g., Juditsky and Lambert-Lacroix (2004). However, it is obtained for very standard conditions and covers a large class of wavelet bases, such as Haar, Daubechies, and Meyer. In contrast, the oracle estimator $\tilde{c}_{j_n}$ attains the optimal rate of convergence for the supremum norm. In fact, for all $\mathbf{u} \in I^d$, we have

$$\tilde{c}_{j_n}(\mathbf{u}) - c(\mathbf{u}) = \tilde{c}_{j_n}(\mathbf{u}) - \mathbb{E}\tilde{c}_{j_n}(\mathbf{u}) + \mathbb{E}\tilde{c}_{j_n}(\mathbf{u}) - c(\mathbf{u}) = D_n(\mathbf{u}) + B_n(\mathbf{u})$$

which implies

$$\sup_{\mathbf{u}\in I^d} |\tilde{c}_{j_n}(\mathbf{u}) - c(\mathbf{u})| = O_{a.s.}\left(\sqrt{\frac{j_n 2^{dj_n}}{n}} + 2^{-j_n t}\right).$$

Thus, if $2^{j_n} \simeq (n/\log n)^{1(2t+d)}$, the terms $\sqrt{j_n 2^{dj_n}/n}$ and $2^{-j_n t}$ are equivalent and equal to $(\log n/n)^{\frac{t}{2t+d}}$ which is the optimal rate for the supremum norm over the Besov class $B^t_{\infty,\infty}(I^d)$.

Comment. We are currently working on a different estimator proposed by a reviewer. Under different assumptions which are potentially satisfied by different classes of wavelets, this estimator achieves an optimal uniform convergence rate.

Appendix A: Useful results on empirical process

Bernstein's inequality (maximal version):

Let $Z_1, \ldots, Z_n$ be independent random variables with $\mathbb{E}(Z_i) = 0, i = 1, \ldots, n$ and $\mathrm{Var}(\sum_{i=1}^n Z_i) \leq v$. Assume further that for some constant $M > 0$, $|Z_i| < M, i = 1, ..., n$. Then for all $t > 0$

$$\mathbb{P}\left(\max_{q\leq n}\left|\sum_{i=1}^q Z_i\right| > t\right) \leq 2\exp\left\{\frac{-t^2}{2v + (2/3)Mt}\right\}. \tag{20}$$

Lemma 1 (Lemma A.1, Einmahl and Mason, 2000)**.** *Let $\mathcal{F}$ and $\mathcal{G}$ be two classes of real-valued measurable functions on $\mathcal{X}$ satisfying*

$$|f(x)| \leq F(x), \qquad f \in \mathcal{F}, \quad x \in \mathcal{X},$$

where F is a finite valued measurable envelope function on $\mathcal{X}$;

$$\|g\| \leq M, \qquad g \in \mathcal{G},$$

where $M > 0$ is a finite constant. Assume that for all probability measures Q with $0 < Q(F^2) < \infty$,

$$N(\varepsilon(Q(F^2))^{1/2}, \mathcal{F}, d_Q) \leq C_1 \varepsilon^{-\nu_1}, \qquad 0 < \varepsilon < 1,$$

and

$$N(\varepsilon M, \mathcal{G}, d_Q) \leq C_2 \varepsilon^{-\nu_2}, \qquad 0 < \varepsilon < 1$$

where $\nu_1, \nu_1, C_1, C_2 \geq 1$ are suitable constants. Then we have for all probability measure Q with $0 < Q(F^2) < \infty$,

$$N(\varepsilon M(Q(F^2))^{1/2}, \mathcal{F}\mathcal{G}, d_Q) \leq C_3 \varepsilon^{-\nu_1-\nu_2}, \qquad 0 < \varepsilon < 1$$

for some finite constant $0 < C_3 < \infty$.

Proposition 2 (Einmahl and Mason, 2000)**.** *Let $Z, Z_1, Z_2, \ldots,$ be a sequence of i.i.d. random vectors taking values in $\mathbb{R}^m$, $m \geq 1$. For each $n \geq 1$, consider the empirical distribution function based on the first n of these random vectors, defined by*

$$G_n(s) = \frac{1}{n} \sum_{i=1}^{n} 1_{Z_i \leq s}, \quad s \in \mathbb{R}^m,$$

where as usual $z \leq s$ means that each component of z is less than or equal to the corresponding component of s. For any measurable real valued function g defined on $\mathbb{R}^m$, set

$$G_n(g) = \int_{\mathbb{R}^m} g(s) dG_n(s), \qquad \mu(g) = \mathbb{E}g(Z) \qquad \text{and} \quad \sigma(g) = Var(g(Z)).$$

Let $a_n : n \geq 1$ denote a sequence of positive constants converging to zero. Consider a sequence $\mathcal{G}_n = \{g_i^{(n)} : i = 1, ..., k_n\}$ of sets of real-valued measurable functions on $\mathbb{R}^2$, satisfying, whenever $g_i^{(n)} \in \mathcal{G}_n$:

$$\mathbb{P}(g_i^{(n)}(Z) = 0,\ g_j^{(n)}(Z)) = 0, \quad i \neq j \quad \text{and} \quad \sum_{i=1}^{k_n} \mathbb{P}(g_i^{(n)}(Z) \neq 0) \leq 1/2.$$

Further assume the following:

- *For some $0 < r < \infty$, $a_n k_n \to r$, as $n \to \infty$.*
- *For some $-\infty < \mu_1, \mu_2 < \infty$, uniformly in $i = 1, \ldots, k_n$, for all large n, $a_n \mu_1 \leq \mu(g_i^{(n)}) \leq a_n \mu_2$.*
- *For some $0 < \sigma_1 < \sigma_2 < \infty$, uniformly in $i = 1, \ldots k_n$, for all large n, $\sigma_1 \sqrt{n} a_n \leq \bar{\sigma}(g_i^{(n)}) \leq \sigma_2 \sqrt{n} a_n$.*
- *For some $0 < B < \infty$, uniformly in $i = 1, \ldots, k_n$, for all large n, $|g_i^{(n)}| \leq B$.*

Proposition 3. *Under these assumptions, with probability one, for each $0 < \varepsilon < 1$, there exists an N_ε such that for $n \geq N_\varepsilon$,*

$$\max_{1 \leq i \leq k_n} \frac{\sqrt{n}\{G_n(g_i^{(n)}) - \mu(g_i^{(n)})\}}{\bar{\sigma}(g_i^{(n)})\sqrt{2|\log a_n|}} \geq 1 - \varepsilon.$$

Talagrand's inequality:
Let X_i, $i = 1, \ldots, n$, be an independent and identically distributed random sample of X with probability law P on $\mathbb{R}$, and $\mathcal{G}$ a P-centered (i.e., $\int g dP = 0$ for all $g \in \mathcal{G}$) countable class of real-valued functions on $\mathbb{R}$, uniformly bounded by the constant U. Let σ be any positive number such that $\sigma^2 \geq \sup_{g\in\mathcal{G}} \mathbb{E}(g^2(X))$. Then, Talagrand's inequality (Talagrand, 1996) implies that there exists a universal constant L such that for all $t > 0$,

$$\mathbb{P}\left(\max_{q\leq n}\left\|\sum_{i=1}^{q} g(X_i)\right\|_{\mathcal{G}} > E + t\right) \leq L\exp\left\{\frac{-t}{LU}\log\left(1+\frac{tU}{V}\right)\right\}, \tag{21}$$

with

$$E = \mathbb{E}\left\|\sum_{i=1}^{n} g(X_i)\right\|_{\mathcal{G}} \quad \text{and} \quad V = \mathbb{E}\left\|\sum_{i=1}^{n}(g(X_i))^2\right\|_{\mathcal{G}}.$$

Further, if $\mathcal{G}$ is a VC(Vapnik–Červonenkis)-type class of functions, with characteristics A and v, then there exist a universal constant B such that [see, e.g., Giné and Guillou, 2002]

$$E \leq B\left[vU\log\frac{AU}{\sigma} + \sqrt{v}\sqrt{n\sigma^2}\frac{AU}{\sigma}\right] \tag{22}$$

Next, if $\sigma < U/2$, the constant A may be replaced by 1 at the price of changing the constant B, and then if, moreover, $n\sigma^2 > C_0\log(U/\sigma)$, we have

$$E \leq C_1\sqrt{n\sigma^2\log\left(\frac{U}{\sigma}\right)} \quad \text{and} \quad V \leq L' n\sigma^2, \tag{23}$$

where C_1, L' are constants depending only on A ,v and C_0. Finally, it follows from (36) and (23) that, for all $t > 0$ satisfying: $C_1\sqrt{n\sigma^2\log(U/\sigma)} \leq t \leq C_2 n\sigma^2/U$ for all constants $C_2 \geq C_1$,

$$\mathbb{P}\left(\max_{n_{k-1}\leq n\leq n_k}\left\|\sum_{i=1}^{n} g(X_i)\right\|_{\mathcal{G}} > t\right) \leq R\exp\left\{\frac{-1}{C_3}\frac{t^2}{n\sigma^2}\right\}, \tag{24}$$

where $C_3 = \log(1 + C_2/L')/RC_2$ and R a constant depending only on A and v.

Appendix B: Proof of Proposition 1

Upper bound

Lemma 2. *Under the assumptions of Proposition 1, one has almost surely*

$$\limsup_{n\to\infty} r_n \sup_{\mathbf{u}\in I^d}\frac{|D_n(\mathbf{u})|}{\sqrt{\int_{\mathbb{R}^d} K^2(\mathbf{x}, 2^{j_n}\mathbf{u})d\mathbf{x}}} \leq \sqrt{\|c\|_\infty}. \tag{25}$$

Proof. Given $\lambda > 1$, define $n_k = [\lambda^k]$, $k \in \mathbb{N}$, where $[a]$ denotes the integer part of a real a. Let $\delta_m = 1/m$, $m \geq 1$ integer, then we can cover the set I^d by a number l_k of small cubes $S_{k,r}$, each of side length $\delta_m 2^{-j_{n_k}}$, with

$$l_k \leq \left(\frac{1}{\delta_m 2^{-j_{n_k}}} + 1\right)^d \leq \left(\frac{2}{\delta_m 2^{-j_{n_k}}}\right)^d, \tag{26}$$

for k large enough. Let us choose points $\mathbf{u}_{k,r} \in S_{k,r} \cap I^d$, $r = 1, \dots, l_k$. We want to prove Lemma 2 over the discrete grid of points $\{\mathbf{u}_{k,r} : r = 1, \dots, l_k\}$. For all $\eta \in (0,1)$ we claim that

$$\limsup_{k\to\infty} \sqrt{\frac{n_k}{(2d\log 2) j_{n_k} 2^{dj_{n_k}}}} \max_{1\le r\le l_k} \max_{n_{k-1}\le n\le n_k} |D_n(\mathbf{u}_{k,r})| \le (1+\eta)\sqrt{\|c\|_\infty [K^2]}, \tag{27}$$

where we note

$$[K^2] = \int_{\mathbb{R}^d} \mathbf{K}^2(\mathbf{x}, 2^{j_n}\mathbf{u}) d\mathbf{x}.$$

To prove (27), we apply the maximal version of Bernstein inequality (see, Appendix A above). Given $\mathbf{u} \in I^d$ and $k \in \mathbb{N}$, for all n satisfying : $n_{k-1} \le n \le n_k$ let

$$Z_i(\mathbf{u}) = \mathbf{K}(2^{j_n}\mathbf{U}_i, 2^{j_n}\mathbf{u}) - \mathbb{E}\mathbf{K}(2^{j_n}\mathbf{U}_i, 2^{j_n}\mathbf{u}), \qquad i = 1, \dots, n.$$

Observe that for each n, the $Z_i(\mathbf{u})$ are independent and identically distributed zero-mean random variables, and for all $\mathbf{u} \in I^d$,

$$D_n(\mathbf{u}) = \frac{2^{dj_n}}{n} \sum_{i=1}^{n} Z_i(\mathbf{u}). \tag{28}$$

By hypothesis (H.2), we have

$$\left|\mathbf{K}(2^{j_n}\mathbf{U}_i, 2^{j_n}\mathbf{u})\right| = \prod_{m=1}^{d} \left|\tilde{K}(2^{j_n}U_{im}, 2^{j_n}u_m)\right| \le \prod_{m=1}^{d} \Phi(2^{j_n}(U_{im} - u_m)) \le \|\Phi\|_\infty^d,$$

where $\|\Phi\|_\infty = \sup_{x\in\mathbb{R}} |\Phi(x)|$. This implies

$$\left|\mathbb{E}\mathbf{K}[(2^{j_n}\mathbf{U}_i, 2^{j_n}\mathbf{u})]\right| \le \mathbb{E}\|\Phi\|_\infty^2 = \|\Phi\|_\infty^d.$$

Thus, for all $\mathbf{u} \in I^d$,

$$|Z_i(\mathbf{u})| \le 2\|\Phi\|_\infty^d := M.$$

Since the $Z_i(\mathbf{u})$, $i = 1, \dots, n$ are independent and centered, we can write for $n = n_k$

$$Var\left(\sum_{i=1}^{n_k} Z_i(\mathbf{u})\right) = n_k Var(Z_1(\mathbf{u})) = n_k \mathbb{E}(Z_1^2(\mathbf{u})).$$

Then using the change of variables $\mathbf{s} = 2^{-j_{n_k}}\mathbf{x}$, $\mathbf{s} = (s_1, \dots, s_d)$, $\mathbf{x} = (x_1, \dots, x_d)$, we obtain

$$\begin{aligned}
\mathbb{E}(Z_1^2(\mathbf{u})) &\le \mathbb{E}\mathbf{K}^2[(2^{j_{n_k}}\mathbf{U}_1, 2^{j_{n_k}}\mathbf{u}) \\
&\le \int_{I^d} \mathbf{K}^2(2^{j_{n_k}}\mathbf{s}, 2^{j_{n_k}}\mathbf{u}) c(\mathbf{s}) d\mathbf{s} \\
&\le 2^{-dj_{n_k}} \|c\|_\infty \int_{[0,2^{j_{n_k}}]^d} \mathbf{K}^2(\mathbf{x}, 2^{j_{n_k}}\mathbf{u}) d\mathbf{x},
\end{aligned}$$

which yields

$$Var\left(\sum_{i=1}^{n_k} Z_i(\mathbf{u})\right) \le n_k 2^{-dj_{n_k}} \|c\|_\infty \int_{\mathbb{R}^d} \mathbf{K}^2(\mathbf{x}, 2^{j_n}\mathbf{u}) d\mathbf{x} := \sigma_k^2.$$

Now, applying the maximal version Bernstein's inequality, for each point $\mathbf{u}_{k,r}$, we obtain for all $t > 0$,

$$\mathbb{P}\left(\max_{n_{k-1}\le n\le n_k}\left|\sum_{i=1}^{n} Z_i(\mathbf{u}_{k,r})\right| > t\right) \le 2\exp\left\{\frac{-t^2}{2\sigma_k^2 + (2/3)Mt}\right\}, \tag{29}$$

which yields

$$\begin{aligned}\mathbb{P}\left(\max_{1\le r\le l_k}\max_{n_{k-1}\le n\le n_k}\left|\sum_{i=1}^{n} Z_i(\mathbf{u}_{k,r})\right| > t\right) &= \mathbb{P}\left(\bigcup_{r=1}^{l_k}\left\{\max_{n_{k-1}\le n\le n_k}\left|\sum_{i=1}^{n} Z_i(\mathbf{u}_{k,r})\right| > t\right\}\right)\\ &\le \sum_{r=1}^{l_k}\mathbb{P}\left(\max_{n_{k-1}\le n\le n_k}\left|\sum_{i=1}^{n} Z_i(\mathbf{u}_{k,r})\right| > t\right)\\ &\le l_k 2\exp\left\{\frac{-t^2}{2\sigma_k^2 + (2/3)Mt}\right\}.\end{aligned}$$

Let $t = \sqrt{2(1+\eta)n_k 2^{-dj_{n_k}}\log 2^{dj_{n_k}}\|c\|_\infty[K^2]}$. Then, for k large enough, $t \to \infty$. Combining this with (26), we obtain after some little algebra,

$$\begin{aligned}\mathbb{P}\left(\max_{1\le r\le l_k}\frac{\max_{n_{k-1}\le n\le n_k}\left|\sum_{i=1}^{n} Z_i(\mathbf{u}_{k,r})\right|}{\sqrt{2n_k 2^{-dj_{n_k}}\log 2^{dj_{n_k}}\|c\|_\infty[K^2]}} > \sqrt{1+\eta}\right) &\le 2l_k\exp\left\{\frac{-t^2}{\frac{t^2}{(1+\eta)\log 2^{dj_n}} + \frac{4}{3}\|\Phi\|^2 t}\right\}\\ &\le 2l_k\exp\left\{-(1+\eta)\log 2^{dj_{n_k}}\right\}\\ &\le 2^d\delta_m^{-d}2^{-d\eta j_{n_k}}.\end{aligned}$$

Since the series $\sum_{k\ge 0}2^{-d\eta j_{n_k}} < \infty$, the Borel–Cantelli lemma yields

$$\mathbb{P}\left(\max_{1\le r\le l_k}\frac{\max_{n_{k-1}\le n\le n_k}\left|\sum_{i=1}^{n} Z_i(\mathbf{u}_{k,r})\right|}{\sqrt{(2d\log 2)n_k j_{n_k}2^{-dj_{n_k}}\|c\|_\infty[K^2]}} > \sqrt{1+\eta}\right) = o(1). \tag{30}$$

That is

$$\limsup_{k\to\infty}\max_{1\le r\le l_k}\frac{\max_{n_{k-1}\le n\le n_k}\left|\sum_{i=1}^{n} Z_i(\mathbf{u}_{k,r})\right|}{\sqrt{(2d\log 2)n_k j_{n_k}2^{-dj_{n_k}}\|c\|_\infty[K^2]}} \le \sqrt{1+\eta},\quad a.s. \tag{31}$$

Since the function $x \mapsto x2^{-2x}$ is decreasing for $x > 2\log 2$, we have for $n_{k-1} \le n \le n_k$, and k large enough,

$$\sqrt{\frac{n_k j_{n_k}2^{-dj_{n_k}}}{n j_n 2^{-dj_n}}} \le \sqrt{\frac{n_k}{n}} \le \sqrt{\frac{n_k}{n_{k-1}}} \le \sqrt{\lambda}. \tag{32}$$

In view of inequality (32), Statement (31) yields

$$\limsup_{k\to\infty}\max_{1\le r\le l_k}\frac{\max_{n_{k-1}\le n\le n_k}\left|\sum_{i=1}^{n} Z_i(\mathbf{u}_{k,r})\right|}{\sqrt{(2d\log 2)n j_n 2^{-dj_n}\|c\|_\infty[K^2]}} \le \sqrt{\lambda(1+\eta)},\quad a.s. \tag{33}$$

Now, multiplying the numerator and the denominator of the fraction in (33) by the factor $2^{dj_n}/n$, and recalling the expression of $D_n(\mathbf{u})$ in (28), we finally get for all $\eta \in (0,1)$,

$$\limsup_{k\to\infty} \max_{1\le r\le l_k} \frac{\max_{n_{k-1}\le n\le n_k} \sqrt{n}\left|D_n(\mathbf{u}_{k,r})\right|}{\sqrt{(2d\log 2) j_n 2^{dj_n}}} \le \sqrt{\lambda(1+\eta)\|c\|_\infty [K^2]}, \tag{34}$$

which proves Lemma 2 over the discrete grid.

Next, to prove Lemma 2 between the grid points, we shall make use of Talagrand's (1996) inequality in (21). Let us introduce the sequence of functions defined as follows: for all $n \ge 1$, $k \ge 1$, $1 \le r \le l_k$ and any fixed $\mathbf{u} \in S_{k,r}$, define

$$g_{k,r}^{(n)}(\mathbf{s},\mathbf{u}) = \mathbf{K}(2^{j_{n_k}}\mathbf{s}, 2^{j_{n_k}}\mathbf{u}_{k,r}) - \mathbf{K}(2^{j_n}\mathbf{s}, 2^{j_n}\mathbf{u}), \qquad \mathbf{s} \in I^d. \tag{35}$$

and set, for all $\lambda > 1$,

$$\mathcal{G}_{k,r}(\lambda) = \left\{ g : \mathbf{s} \mapsto g_{k,r}^{(n)}(\mathbf{s},\mathbf{u}) : \mathbf{u} \in S_{k,r} \cap I^d,\ n_{k-1} \le n \le n_k \right\}.$$

Let $\mathbf{S} = (S_1,\dots,S_d)$ be a vector of $[0,1]$ uniform random variables, now we have to check the following conditions in order to apply Talagrand's inequality:

i) The classes $\mathcal{G}_{k,r}(\lambda)$, $1 \le r \le l_k$, are of VC-type with characteristics A and v;

ii) $\forall g \in \mathcal{G}_{k,r}(\lambda)$, $\|g\|_\infty \le \mathrm{U}$;

iii) $\forall g \in \mathcal{G}_{k,r}(\lambda)$, $\mathrm{Var}[g(\mathbf{S})] \le \sigma_k^2$;

iv) $\sigma_k < U/2$ and $n_k\sigma_k^2 > C_0 \log(U/\sigma_k)$, $C_0 > 0$.

These conditions will be checked below.

Recall that $\mathbf{U}_i = (F_{1i}(X_{i1}),\dots,F_{di}(X_{id}))$, $i = 1,\dots,n$, is a sequence of independent and identically distributed vectors of $[0,1]$ uniform components. We have shown (see below) that each class $\mathcal{G}_{k,r}(\lambda)$ satisfies all the conditions i), ii), iii) and iv) for $\mathrm{U} = 2\|\Phi\|_\infty^d$ and $\sigma_k^2 = D_0 2^{-dj_{n_k}}\|c\|_\infty \omega_\phi^2(\delta_m)$, where ω_ϕ is the modulus of continuity of ϕ defined below in (43) and D_0 is a positive constant depending on $\|\Phi\|_\infty$ and d. Then, Talagrand's inequality gives, for all $t > 0$,

$$\mathbb{P}\left(\max_{n_{k-1}\le n\le n_k} \left\| \sum_{i=1}^n (g(\mathbf{U}_i) - \mathbb{E}g(\mathbf{U}_i)) \right\|_{\mathcal{G}_{k,r}(\lambda)} > t \right) \le R\exp\left\{ \frac{-1}{C_3}\frac{t^2}{n_k\sigma_k^2} \right\}, \tag{36}$$

which yields, by taking the maximum over r and $t = C_1\sqrt{n_k\sigma_k^2 \log\left(\frac{\mathrm{U}}{\sigma_k}\right)}$,

$$\mathbb{P}\left(\max_{1\le r\le l_k} \max_{n_{k-1}\le n\le n_k} \left\| \sum_{i=1}^n (g(\mathbf{U}_i) - \mathbb{E}g(\mathbf{U}_i)) \right\|_{\mathcal{G}_{k,r}(\lambda)} > t \right) \le Rl_k \exp\left\{ \frac{-C_1^2}{C_3}\log\left(\frac{\mathrm{U}}{\sigma_k}\right) \right\}. \tag{37}$$

Whenever $m \to \infty$, $\omega_\phi(\delta_m) \to 0$. Hence, for any $\varepsilon > 0$, there exists $m_0 \in \mathbb{N}$ such that $\omega_\phi(\delta_m) < \varepsilon$ for $m \ge m_0$. Using this fact, we can replace σ_k^2 by $4D2^{-dj_{n_k}}\varepsilon\|c\|_\infty$, for m large enough. We also have, for k large enough,

$$\log\left(\frac{\mathrm{U}}{\sigma_k}\right) = \log\left(\frac{\mathrm{U}}{4D\varepsilon\|c\|_\infty}\right) + j_{n_k}\log 2 \sim j_{n_k}\log 2,$$

and thus, for k, m large enough,

$$t = C_1\sqrt{n_k\sigma_k^2 \log\left(\frac{\mathrm{U}}{\sigma_k}\right)} \sim \sqrt{4DC_1^2 n_k 2^{-dj_{n_k}} d\, j_{n_k} \log 2\, \varepsilon\|c\|_\infty}.$$

By combining these facts with (26) we obtain, with $A_0 = \sqrt{2DC_1}$,

$$\mathbb{P}\left(\max_{1\le r\le l_k} \frac{\max_{n_{k-1}\le n\le n_k} \left\|\sum_{i=1}^n (g(\mathbf{U}_i) - \mathbb{E}g(\mathbf{U}_i))\right\|_{\mathcal{G}_{k,r}(\lambda)}}{\sqrt{(2d\log 2)n_k j_{n_k} 2^{-dj_{n_k}}}} > A_0\sqrt{\varepsilon\|c\|_\infty}\right) \le 4R\delta^{-2}2^{-[C_1^2/2C_3-1]2j_{n_k}}. \tag{38}$$

Now, we can choose the constant C_1 in such a way that $C_1^2/2C_3 - 1 > 0$; in which case the series $\sum_{k\ge 0} 2^{-[C_1^2/2C_3-1]2j_{n_k}}$ converges. Thus, the Borel-Cantelli's lemma implies

$$\mathbb{P}\left(\max_{1\le r\le l_k} \frac{\max_{n_{k-1}\le n\le n_k} \left\|\sum_{i=1}^n (g(\mathbf{U}_i) - \mathbb{E}g(\mathbf{U}_i))\right\|_{\mathcal{G}_{k,r}(\lambda)}}{\sqrt{(2d\log 2)n_k j_{n_k} 2^{-dj_{n_k}}}} > A_0\sqrt{\varepsilon\|c\|_\infty}\right) = o(1), \tag{39}$$

that is

$$\limsup_{k\to\infty} \max_{1\le r\le l_k} \frac{\max_{n_{k-1}\le n\le n_k} \left\|\sum_{i=1}^n (g(\mathbf{U}_i) - \mathbb{E}g(\mathbf{U}_i))\right\|_{\mathcal{G}_{k,r}(\lambda)}}{\sqrt{(2d\log 2)n_k j_{n_k} 2^{-dj_{n_k}}}} \le A_0\sqrt{\varepsilon\|c\|_\infty}, \quad a.s. \tag{40}$$

Arguing as in the discrete case, with Statement (32) in view, we conclude that

$$\limsup_{k\to\infty} \max_{1\le r\le l_k} \frac{\max_{n_{k-1}\le n\le n_k} \sqrt{n}\left\|D_n(\mathbf{u}_{k,r}) - D_n(\mathbf{u})\right\|_{\mathcal{G}_{k,r}(\lambda)}}{\sqrt{(2d\log 2)j_n 2^{dj_n}}} \le A_0\sqrt{\lambda\varepsilon\|c\|_\infty}, \quad a.s., \tag{41}$$

which completes the proof of Lemma 2 between the grid-points.

Now recapitulating, we can infer from (34) and (41) that

$$\limsup_{k\to\infty} \max_{1\le r\le l_k} \frac{\max_{n_{k-1}\le n\le n_k} \sqrt{n}\,|D_n(\mathbf{u})|}{\sqrt{(4\log 2)j_n 2^{-2j_n}}} \le \sqrt{\lambda(1+\eta)\|c\|_\infty[K^2]} + A_0\sqrt{\lambda\varepsilon\|c\|_\infty}, \quad a.s. \tag{42}$$

Since η and ε are arbitrary, letting $\lambda \to 1$ completes the proof of Lemma 2. ∎

Checking conditions i), ii), iii), iv)

Checking i): Observe that the elements of the class $\mathcal{G}_{k,r}(\lambda)$ may be rewritten as

$$g_{k,r}^{(n)}(\mathbf{s},\mathbf{u}) = \prod_{m=1}^d \widetilde{K}(2^{j_{n_k}}s_m, 2^{j_{n_k}}u_{k,r,m}) - \prod_{m=1}^d \widetilde{K}(2^{j_n}s_m, 2^{j_n}u_m),$$

where $\widetilde{K}(x,y) = \sum_{l=1}^{2^{j_n}} \phi(x-l)\phi(y-l)$, with ϕ compactly supported and of bounded variation. For $m = 1,\ldots,d$, define the classes of functions: $\mathcal{F}_m = \{v \mapsto \sum_{l\in\mathbb{Z}} \phi(2^j w - l)\phi(2^j v - l)$:

$w \in [0,1], j \in \mathbb{N}\}$. By Lemma 2 in Giné and Nickl (2009), $\mathcal{F}_1, \ldots, \mathcal{F}_m$ are VC-type classes of functions. Moreover, $\mathcal{F}_1, \ldots, \mathcal{F}_m$ are uniformly bounded. Indeed, for all $w \in [0,1],\ j \in \mathbb{N}$, we have $\left|\sum_{l=1}^{2^{jn}} \phi(2^j w - l)\phi(2^j \cdot -l)\right| \leq \|\phi\|_\infty \|\theta_\phi\|_\infty$, as the function $\theta_\phi(x) = \sum_{l=1}^{2^{jn}} |\phi(x-l)|$ is bounded. By Lemma A.1 in Einmahl and Mason (2000), this implies that the product $\mathcal{F}_1 \cdots \mathcal{F}_m$ is also a VC-type class of functions. Now, using properties (iv) and (v) of Lemma 2.6.18 in van der Vaart and Wellner (1996), we can infer that the classes of functions $\mathcal{G}_{k,r}(\lambda)$ are of VC-type for all k, r fixed.

Checking ii): For all $k \geq 1,\ 0 \leq r \leq l_k,\ n_{k-1} \leq n \leq n_k$, using hypothesis (H.2), we can write

$$\begin{aligned} \left|g_{k,r}^{(n)}(\cdot,\mathbf{u})\right| &\leq \left|\mathbf{K}(2^{j_{n_k}}\cdot, 2^{j_{n_k}}\mathbf{u}_{k,r})\right| + \left|\mathbf{K}(2^{j_n}\cdot, 2^{j_n}\mathbf{u})\right| \\ &\leq \prod_{m=1}^{d} \widetilde{K}(2^{j_{n_k}}\cdot, 2^{j_{n_k}}u_{k,r,m}) + \prod_{m=1}^{d} \widetilde{K}(2^{j_n}\cdot, 2^{j_n}u_m) \leq 2\|\Phi\|_\infty^d \end{aligned}$$

and ii) holds with $\mathrm{U} = 2\|\Phi\|_\infty^d$.

Checking iii): For all $k \geq 1,\ 0 \leq r \leq l_k,\ n_{k-1} \leq n \leq n_k$. As in Giné and Nickl (2009) we choose $\lambda \in (0,1)$, such that $j_{n_k} = j_n$. By a change of variable $\mathbf{s} = \mathbf{u} + 2^{-j_{n_k}}\mathbf{x},\ \mathbf{s} = (s_1, \ldots, s_d),\ \mathbf{x} = (x_1, \ldots, x_d),\ \mathbf{u} = (u_1, \ldots, u_d)$, we have

$$\begin{aligned} \mathbb{E}\left[\left(g_{k,r}^{(n)}(\mathbf{S},\mathbf{u})\right)^2\right] &= \mathbb{E}\left[\left(\mathbf{K}(2^{j_{n_k}}\mathbf{S}, 2^{j_{n_k}}\mathbf{u}_{k,r}) - \mathbf{K}(2^{j_{n_k}}\mathbf{S}, 2^{j_{n_k}}\mathbf{u})\right)^2\right] \\ &= \int_{I^d} \left(\mathbf{K}(2^{j_{n_k}}\mathbf{s}, 2^{j_{n_k}}\mathbf{u}_{k,r}) - \mathbf{K}(2^{j_{n_k}}\mathbf{s}, 2^{j_{n_k}}\mathbf{u})\right)^2 c(\mathbf{s})d\mathbf{s} \\ &\leq \frac{\|c\|_\infty}{2^{-dj_{n_k}}} \int_{[-2^{j_{n_k}}, 2^{j_{n_k}}]^d} \left(\mathbf{K}(2^{j_{n_k}}\mathbf{u}+\mathbf{x}, 2^{j_{n_k}}\mathbf{u}_{k,r}) - \mathbf{K}(2^{j_{n_k}}\mathbf{u}+\mathbf{x}, 2^{j_{n_k}}\mathbf{u})\right)^2 d\mathbf{x} \\ &\leq 2^{-dj_{n_k}}\|c\|_\infty \int_{\mathbb{R}^d} \left(\mathbf{K}(2^{j_{n_k}}\mathbf{u}+\mathbf{x}, 2^{j_{n_k}}\mathbf{u}_{k,r}) - \mathbf{K}(2^{j_{n_k}}\mathbf{u}+\mathbf{x}, 2^{j_{n_k}}\mathbf{u})\right)^2 d\mathbf{x}. \end{aligned}$$

To simplify, let us take $\mathbf{w} = 2^{j_{n_k}}\mathbf{u} + \mathbf{x}$, then $d\mathbf{x} = d\mathbf{w}$, $w = (w_1, \ldots, w_d) \in \mathbb{R}^d$.
Put $A(\mathbf{w}) = \mathbf{K}(\mathbf{w}, 2^{j_{n_k}}\mathbf{u}_{k,r}) - \mathbf{K}(\mathbf{w}, 2^{j_{n_k}}\mathbf{u})$; using the multiplicativity of the kernel $\mathbf{K}$, we can rewrite $A(\mathbf{w})$ as

$$\begin{aligned} A(\mathbf{w}) &= \prod_{l=1}^{d} \widetilde{K}(w_l, 2^{j_{n_k}}u_{k,r,l}) - \prod_{l=1}^{d} \widetilde{K}(w_l, 2^{j_{n_k}}u_{k,r,l}) \\ &= \sum_{l=1}^{d} \left[\widetilde{K}(w_l, 2^{j_{n_k}}u_{k,r,l}) - \widetilde{K}(w_l, 2^{j_{n_k}}u_l)\right] \prod_{p=1, p\neq l}^{d} \widetilde{K}(w_p, 2^{j_{n_k}}u_{k,r,p}). \end{aligned}$$

For any $\delta > 0$, the modulus of continuity of ϕ is defined as

$$\omega_\phi(\delta) = \{\sup|\phi(x) - \phi(y)| : |x-y| \leq \delta\}. \tag{43}$$

Recall that $\widetilde{K}(x,y) = \sum_{h=1}^{2^{jn}} \phi(x-h)\phi(y-h)$. Combining these facts with the inequality $(a_1 + \cdots + a_d)^2 \leq d(a_1^2 + \cdots + a_d^2)$, and Fubini's Theorem, we get

$$\int_{\mathbb{R}^d} |A(\mathbf{w})|^2 d\mathbf{w} \leq d\int_{\mathbb{R}^d} \sum_{l=1}^{d} \left[\widetilde{K}(w_l, 2^{j_{n_k}}u_{k,r,l}) - \widetilde{K}(w_l, 2^{j_{n_k}}u_l)\right]^2 \prod_{p=1, p\neq l}^{d} \widetilde{K}^2(w_l, 2^{j_{n_k}}u_{k,r,p})d\mathbf{w}.$$

Then

$$\begin{aligned}
\int_{\mathbb{R}^d} |A(\mathbf{w})|^2 d\mathbf{w} &\leq d\sum_{l=1}^{d}\int_{\mathbb{R}^d}\left[\sum_{h=1}^{2^{jn}}\phi(2^{j_{n_k}}w_l-h)[\phi(2^{j_{n_k}}u_{k,r,l}-h)-\phi(2^{j_{n_k}}u_l-h)]\right]^2 \\
&\quad\times\prod_{p=1,p\neq l}^{d}\widetilde{K}^2(w_p,2^{j_{n_k}}u_{k,r,p})d\mathbf{w} \\
&\leq d\omega_\phi^2(\delta_m)\sum_{l=1}^{d}\int_{\mathbb{R}^d}\left[\sum_{h}^{2^{jn}}\phi(2^{j_{n_k}}w_l-h)\right]^2\prod_{p=1,p\neq l}^{d}\widetilde{K}^2(w_p,2^{j_{n_k}}u_{k,r,p})d\mathbf{w} \\
&\leq d\omega_\phi^2(\delta_m)\sum_{l=1}^{d}\int_{\mathbb{R}}\left[\sum_{h=1}^{2^{jn}}\phi(2^{j_{n_k}}w_l-h)\right]^2 dw_l\prod_{p=1,p\neq l}^{d}\int_{\mathbb{R}}\widetilde{K}^2(w_p,2^{j_{n_k}}u_{k,r,p})dw_p.
\end{aligned}$$

Now, since the family $\{\phi(\cdot-h) : h=1,\ldots,2^{j_n}\}$ is an orthonormal basis, the quantity $\int_{\mathbb{R}}\left(\sum_{h=1}^{2^{jn}}\phi(w_l-h)\right)^2 dw_l$ can be bounded by a constant M_0; thus

$$\begin{aligned}
\int_{\mathbb{R}^d} |A(\mathbf{w})|^2 d\mathbf{w} &\leq M_0 d\omega_\phi^2(\delta_m)\sum_{l=1}^{d}\prod_{p=1,p\neq l}^{d}\int_{\mathbb{R}}\widetilde{K}^2(w_p,2^{j_{n_k}}u_{k,r,p})dw_p \\
&\leq M_0 d^2\omega_\phi^2(\delta_m)D,
\end{aligned}$$

where we use Hypothesis (H.2) for the last inequality, with D a positive constant depending on $\|\Phi\|_\infty$. Finally, we obtain

$$\mathbb{E}\left[\left(g_{k,r}^{(n)}(\mathbf{S},\mathbf{u})\right)^2\right] \leq M_0 d^2 D 2^{-dj_{n_k}}\|c\|_\infty\omega_\phi^2(\delta_m), \tag{44}$$

and iii) holds with

$$\sigma_k^2 = D_0 2^{-dj_{n_k}}\|c\|_\infty\omega_\phi^2(\delta_m), \qquad D_0 = M_0 d^2 D. \tag{45}$$

Checking iv): For $m>0$ fixed, we have

$$\frac{\sigma_k}{\mathrm{U}} = \frac{D_0^{1/2}2^{-\frac{d}{2}j_{n_k}}\|c\|_\infty^{1/2}\omega_\phi(\delta_m)}{2\|\Phi\|_\infty^d} \to 0, k\to\infty,$$

which implies that $\sigma_k/U < \varepsilon$, for all $\varepsilon>0$ and k large enough. Hence, for $\varepsilon = 1/2$, we have $\sigma_k < U/2$. We also have, for all large k,

$$\frac{n_k\sigma_k^2}{\log\left(\frac{U}{\sigma_k}\right)} = \frac{D_0 n_k\|c\|_\infty\omega_\phi^2(\delta_m)}{j_{n_k}2^{dj_{n_k}}\ \log 2} \longrightarrow \infty,$$

by Hypothesis (H.4). This readily implies that, for any constant $C_0>0$, $n_k\sigma_k^2 > C_0\log(U/\sigma_k)$ for all large k, and iv) holds.

Lower bound

Lemma 3. *Under the assumptions of Proposition 1, one has almost surely*

$$\liminf_{n\to\infty} r_n \sup_{\mathbf{u}\in I^d} \frac{|D_n(\mathbf{u})|}{\sqrt{\int_{\mathbb{R}^d} K^2[(\mathbf{x}, 2^{j_n}\mathbf{u})]d\mathbf{x}}} \geq \sqrt{\|c\|_\infty}. \tag{46}$$

Proof. It is an adaptation of the proof of Proposition 2 in Giné and Nickl (2009), which is, itself, inspired by Proposition 2 in Einmahl and Mason (2000). According to this latter proposition, (46) holds if and only if for all $\tau > 0$, and all large n, there exists $k_n =: k_n(\tau)$ points $\mathbf{z}_{i,n} = (z_{1,i,n}, \dots, z_{d,i,n}) \in I^d$, $i = 1, \dots, k_n$ such that, for functions $g_i^{(n)}(\mathbf{s}) = \mathbf{K}(2^{j_n}\mathbf{s}, 2^{j_n}\mathbf{z}_{i,n})$, $\mathbf{s} \in I^d$, and for $\mathbf{U} = (U_1, \dots, U_d)$ a random vector with joint density c, the following conditions hold :

C.1) $\mathbb{P}(g_i^{(n)}(\mathbf{U}) \neq 0,\ g_{i'}^{(n)}(\mathbf{U}) \neq 0) = 0, \quad \forall i \neq i'$;

C.2) $\sum_{i=1}^{k_n} \mathbb{P}(g_i^{(n)}(\mathbf{U}) \neq 0) \leq 1/2$;

C.3) $2^{-j_n}k_n \longrightarrow r \in]0, \infty[$;

C.4) $\exists\, \mu_1, \mu_2 \in \mathbb{R} : 2^{-dj_n}\mu_1 \leq \mathbb{E}g_i^{(n)}(\mathbf{U}) \leq 2^{-dj_n}\mu_2, \quad \forall i = 1, \dots, k_n$;

C.5) $\exists\, \sigma_1, \sigma_2 > 0 : 2^{-dj_n}\sigma_1^2 \leq \mathrm{Var}[g_i^{(n)}(\mathbf{U})] \leq 2^{-dj_n}\sigma_2^2, \quad \forall i = 1, \dots, k_n$;

C.6) $\|g_i^{(n)}\|_\infty < \infty, \forall i = 1, \dots, k_n, \forall n \geq 1$;

Now, we have to check these conditions. By hypothesis the copula density c is continuous and bounded on I^d, then there exists some orthrotope $D \subset I^d$ such that $\max_{\mathbf{s}\in D} c(\mathbf{s}) = \|c\|_\infty$. Thus, for all $\tau > 0$ there exists $\mathbf{s}_0 \in D$ such that $c(\mathbf{s}_0) \geq (1-\tau)\|c\|_\infty$. Let

$$D_\tau = \{\mathbf{s} \in D : c(\mathbf{s}) \geq (1-\tau)\|c\|_\infty\}, \tag{47}$$

and choose a subset $D_0 \subset D_\tau$ such that $\mathbb{P}(\mathbf{U} \in D_0) \leq \frac{1}{2}$. Suppose that $D_0 = \prod_{j=1}^d [a_j, b_j]$, with $0 \leq a_j < b_j \leq 1$ and $b_j - a_j = \ell$, $\forall j = 1, \dots, d$.

Set $\delta = 3B$ and define

$$z_{j,i,n} = a + i\delta 2^{-j_n}, \quad i = 1, \dots, \left[\frac{b-a}{\delta 2^{-j_n}}\right] - 1 := k_n, \quad j = 1, \dots, d,$$

where $[x]$ denotes the integer part of a real x.

Checking C.1): Recall that ϕ is supported on $[0, B]$, then

$$g_i^{(n)}(\mathbf{U}) \neq 0 \iff \forall k, l \in \mathbb{Z} \quad \begin{cases} 0 \leq \quad 2^{j_n}U_j - l \quad \leq B,\ j = 1, \dots, d, & (1) \\ 0 \leq \quad 2^{j_n}z_{j,i,n} - l \quad \leq B,\ j = 1, \dots, d, & (2) \end{cases}$$

and

$$g_{i'}^{(n)}(\mathbf{U}) \neq 0 \iff \forall k, l \in \mathbb{Z} \quad \begin{cases} 0 \leq \quad 2^{j_n}U_j - l \quad \leq B,\ j = 1, \dots, d, & (1)' \\ 0 \leq \quad 2^{j_n}z_{j,i',n} - l \quad \leq B,\ j = 1, \dots, d. & (2)' \end{cases}$$

Combining (2) and (2)′ gives, for every $j = 1, \ldots, d$,

$$|z_{j,i,n} - z_{j,i',n}| \leq 2^{-j_n} B. \quad (3)$$

But, by definition, for all $i \neq i'$, $|z_{j,i,n} - z_{j,i',n}| > \delta 2^{-j_n} = 3B2^{-j_n}$, which contradicts (3). Hence, the event $\{g_i^{(n)}(\mathbf{U}) \neq 0, g_{i'}^{(n)}(\mathbf{U}) \neq 0\}$ is empty for $i \neq i'$ and condition C.1) holds.

Checking C.2): For all $n \geq 1$, the sets $\{g_i^{(n)}(\mathbf{U}) \neq 0\}$, $i = 1, \ldots, k_n$ are disjoint in view of Condition C.1). Then, we have

$$\sum_{i=1}^{k_n} \mathbb{P}(\{g_i^{(n)}(\mathbf{U}) \neq 0\}) = \mathbb{P}\left(\bigcup_{i=1}^{k_n} \{g_i^{(n)}(\mathbf{U}) \neq 0\}\right).$$

Now, it suffices to show that $\bigcup_{i=1}^{k_n} \{g_i^{(n)}(\mathbf{U}) \neq 0\} \subset \{\mathbf{U} \in D_0\}$. From statements (1) and (3) above, we can write, for all $j = 1, \ldots, d$,

$$\begin{gathered} -B \leq 2^{j_n}(U_j - u_{j,i,n}) \leq B \\ u_{j,i,n} - 2^{-j_n} B \leq U_j \leq u_{j,i,n} + 2^{-j_n} B \\ a_j \leq a_j + (3i-1)2^{-j_n} B \leq U_j \leq a_j + (3i+1)2^{-j_n} B \leq b_j. \end{gathered}$$

That is $U_j \in [a_j, b_j]$, and hence $\mathbf{U} = (U_1, \ldots, U_d) \in \prod_{j=1}^d [a_j, b_j] = D_0$. It follows that,

$$\begin{gathered} \forall\ i = 1, \ldots, k_n, \quad \{g_i^{(n)}(\mathbf{U}) \neq 0\} \subset \{\mathbf{U} \in D_0\} \\ \bigcup_{i=1}^{k_n} \{g_i^{(n)}(\mathbf{U}) \neq 0\} \subset \{\mathbf{U} \in D_0\} \\ \mathbb{P}\left(\bigcup_{i=1}^{k_n} \{g_i^{(n)}(\mathbf{U}) \neq 0\}\right) \leq \mathbb{P}(\{\mathbf{U} \in D_0\}) \leq \tfrac{1}{2}. \end{gathered}$$

Hence, C.2) is fulfilled.

Checking C.3): It is immediate, since

$$2^{-j_n} k_n = 2^{-j_n}\left(\left[\frac{b-a}{\delta 2^{-j_n}}\right] - 1\right) = \left[\frac{b-a}{\delta}\right] - 2^{-j_n} \to \left[\frac{b-a}{\delta}\right] =: r > 0,\ n \to \infty.$$

Checking C.4) : Using a change of variables $\mathbf{s} = 2^{-j_n}\mathbf{x}$, $\mathbf{s} = (s_1, \ldots, s_d)$, $\mathbf{x} = (x_1, \ldots, x_d)$, we have

$$\begin{aligned} |\mathbb{E} g_i^{(n)}(\mathbf{U})| &\leq \int_{I_\epsilon^d} \left|\mathbf{K}(2^{j_n}\mathbf{s}, 2^{j_n}\mathbf{z}_{i,n})\right| c(\mathbf{s}) d\mathbf{s} \\ &\leq 2^{-dj_n} \|c\|_\infty \int_{\mathbb{R}^d} \left|\mathbf{K}(\mathbf{x}, 2^{j_n}\mathbf{z}_{i,n})\right| d\mathbf{x} \\ &\leq 2^{-dj_n} \|c\|_\infty \int_{\mathbb{R}^d} |\prod_{j=1}^d \widetilde{K}(x_j, 2^{j_n} u_{j,i,n})| dx_j \\ &\leq 2^{-dj_n} \|c\|_\infty \int_{\mathbb{R}^d} \prod_{j=1}^d \Phi(x_j - 2^{j_n} u_{j,i,n}) dx_j \\ &\leq 2^{-2j_n} \mu, \end{aligned}$$

where $\mu = \|c\|_\infty \int_{\mathbb{R}^d} \prod_{j=1}^d \Phi(x_j - 2^{j_n} u_{j,i,n}) dx_j$ exists, because the function Φ is integrable by hypothesis (H.2). The last inequality is equivalent to

$$-2^{-2j_n}\mu \le \mathbb{E}g_i^{(n)}(U,V) \le 2^{-2j_n}\mu, \ \forall i = 1, \cdots, k_n.$$

That is C.4) holds.

Checking C.5) : For $n \ge 1$, $i = 1, \ldots, k_n$, using a change of variables $\mathbf{s} = 2^{-j_n}\mathbf{x} + \mathbf{z}_{i,n}$, $\mathbf{s} = (s_1, \ldots, s_d)$, $\mathbf{x} = (x_1, \ldots, x_d)$, $\mathbf{z}_{i,n} = (z_{1,i,n}, \ldots, z_{d,i,n})$, we can write

$$\begin{aligned} \mathrm{Var}[g_i^{(n)}(\mathbf{U})] &\le \mathbb{E}\left[\left(g_i^{(n)}(\mathbf{U})\right)^2\right] \\ &\le \int_{I^d} \mathbf{K}^2(2^{j_n}\mathbf{s}, 2^{j_n}\mathbf{z}_{i,n}) c(\mathbf{s}) d\mathbf{s} \\ &\le 2^{-dj_n}\|c\|_\infty \int_{\mathbb{R}^d} \mathbf{K}^2(\mathbf{x} + 2^{j_n}\mathbf{z}_{i,n}, 2^{j_n}\mathbf{z}_{i,n}) d\mathbf{x}. \end{aligned}$$

Putting $\sigma_2^2 := \|c\|_\epsilon \int_{\mathbb{R}^d} \mathbf{K}^2(\mathbf{x} + 2^{j_n}\mathbf{z}_{i,n}, 2^{j_n}\mathbf{z}_{i,n}) d\mathbf{x}$ yields

$$\mathrm{Var}[g_i^{(n)}(\mathbf{U})] \le 2^{-dj_n}\sigma_2^2,$$

which is the upper bound in condition C.5). For the lower bound, we have

$$\begin{aligned} \mathrm{Var}[g_i^{(n)}(\mathbf{U})] &= \mathbb{E}\left[\left(g_i^{(n)}(\mathbf{U})\right)^2\right] - \left[\mathbb{E}g_i^{(n)}(\mathbf{U})\right]^2 \\ &= \int_{I^d} \mathbf{K}^2(2^{j_n}\mathbf{s}, 2^{j_n}\mathbf{z}_{i,n}) c(\mathbf{s}) d\mathbf{s} - \left(\int_{I^d} \mathbf{K}(2^{j_n}\mathbf{s}, 2^{j_n}\mathbf{z}_{i,n}) c(\mathbf{s}) d\mathbf{s}\right)^2. \end{aligned}$$

Put $\mu_n^2 = \left(\int_{I^d} \mathbf{K}(2^{j_n}\mathbf{s}, 2^{j_n}\mathbf{z}_{i,n}) c(\mathbf{s}) d\mathbf{s}\right)^2$. Noting that $D_\tau \subset I^d$, by a change of variables $\mathbf{x} = 2^{j_n}\mathbf{s}$, we obtain

$$\begin{aligned} \mathrm{Var}[g_i^{(n)}(\mathbf{U})] &\ge \int_{D_\tau} \mathbf{K}^2(2^{j_n}\mathbf{s}, 2^{j_n}\mathbf{z}_{i,n}) c(\mathbf{s}) d\mathbf{s} - \mu_n^2 \\ &\ge (1-\tau)\|c\|_\infty \int_{D_\tau} \mathbf{K}^2(2^{j_n}\mathbf{s}, 2^{j_n}\mathbf{z}_{i,n}) d\mathbf{s} - \mu_n^2 \\ &\ge (1-\tau)\|c\|_\infty 2^{-dj_n} \int_{\mathbb{R}^d} \mathbf{K}^2(\mathbf{x}, 2^{j_n}\mathbf{z}_{i,n}) d\mathbf{x} - \mu_n^2. \end{aligned}$$

Proceeding again to the same change of variables, and observing from hypothesis (H.3) that $\int_{\mathbb{R}^d} \mathbf{K}(\mathbf{x}, 2^{j_n}\mathbf{z}_{i,n}) d\mathbf{x} = 1$, we can write

$$\mu_n^2 \le \left(\|c\|_\infty 2^{-dj_n} \int_{\mathbb{R}^d} \mathbf{K}(\mathbf{x}, 2^{j_n}\mathbf{z}_{i,n}) d\mathbf{x}\right)^2 \le \|c\|_\infty^2 2^{-2dj_n},$$

which implies $-\mu_n^2 \ge -\|c\|_\epsilon^2 2^{-4j_n}$. Thus, for n large enough, we obtain the lower bound in condition C.5), i.e.,

$$\begin{aligned} \mathrm{Var}[g_i^{(n)}(\mathbf{U})] &\ge 2^{-dj_n}(1-\tau)\|c\|_\infty \int_{\mathbb{R}^d} \mathbf{K}^2(\mathbf{x}, 2^{j_n}\mathbf{z}_{i,n}) d\mathbf{x} - \|c\|_\infty^2 2^{-2dj_n} \\ &\ge 2^{-dj_n}\sigma_1^2 + o(1), \end{aligned}$$

with $\sigma_1^2 := (1-\tau)\|c\|_\infty \int_{\mathbb{R}^d} K^2(\mathbf{x}, 2^{j_n}\mathbf{z}_{i,n}]d\mathbf{x}$. Finally, C.5) holds. Moreover, letting $\tau \to 0$, we get $\sigma_2^2 = \sigma_1^2 = \|c\|_\infty \int_{\mathbb{R}^d} \mathbf{K}^2(\mathbf{x}, 2^{j_n}\mathbf{u}d\mathbf{x}$.

Checking C.6): For all $\mathbf{s} \in I^d$, $n \geq 1$, $i = 1, \ldots, k_n$, by using Assumptions (H.1) and (H.2) and the multiplicativity of kernel $\mathbf{K}$, we have

$$\begin{aligned} |g_i^{(n)}(\mathbf{s})| &= |\mathbf{K}(2^{j_n}\mathbf{s}, 2^{j_n}\mathbf{z}_{i,n})| = \prod_{m=1}^{d} |\widetilde{K}(2^{j_n}s_m, 2^{j_n}z_{i,n,m})| \\ &\leq \prod_{m=1}^{d} \sum_{l=1}^{2^{j_n}} |\phi(2^{j_n}s_m - l)\phi(2^{j_n}z_{i,n,m} - l)| \\ &\leq \|\phi\|_\infty^d \prod_{m=1}^{d} \sum_{l=1}^{2^{j_n}} |\phi(2^{j_n}s_m - l)| \\ &\leq \|\phi\|_\infty^d \|\theta_\phi\|_\infty^d. \end{aligned}$$

Hence, $\sup_{n\geq 1, 1\leq i\leq k_n} \|g_i^{(n)}\| \leq \|\phi\|_\infty^d \|\theta_\phi\|_\infty^d$, and C.6) holds.

Since Conditions C.1) to C.6) are fulfilled, we can now apply Proposition 2 in Einmahl and Mason (2000) to complete the proof of Lemma 3. ■

Finally, Lemma 2 and Lemma 3 give the proof of Proposition 1.

References

Autin, F., Le Pennec, E., and Tribouley, K. (2010). Thresholding methods to estimate copula density. *Journal of Multivariate Analysis*, **101** (1), 200–222.

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

Cohen, A., Daubechies, I., and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, **1**, 54–81.

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).

Einmahl, U. and Mason, D. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *Journal of Theoretical Probability*, **13** (1), 1–37.

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

Gannoun, A. and Hosseinioun, N. (2012). On wavelet-based methods for estimating the copula function. *Rev. Roumaine Math. Pures Appl.*, **57** (3), 205–213.

Geenens, G., Charpentier, A., and Paindaveine, D. (2017). Probit transformation for nonparametric kernel estimation of the copula density. *Bernoulli*, **23** (3), 1848–1873.

Genest, C., Masiello, E., and Tribouley, K. (2009). Estimating copula densities through wavelets. *Insurance: Mathematics and Economics*, **44** (2), 170–181.

Gijbels, I. and Mielniczuk, J. (1990). Estimating the density of a copula function. *Communications in Statistics-Theory and Methods*, **19** (2), 445–464.

GINÉ, E. AND GUILLOU, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Annals I. Poincaré-PR*, **38** (6), 907–921.

GINÉ, E. AND NICKL, R. (2009). Uniform limit theorems for wavelet density estimators. *The Annals of Probability*, **37** (4), 1605–1646.

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

JUDITSKY, A. AND LAMBERT-LACROIX, S. (2004). On minimax density estimation on $\mathbb{R}$. *Bernoulli*, **10**, 187–220.

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

MEYER, Y. (1992). *Wavelets and Operators: Volume 1*. 37. Cambridge university press.

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** (5B), 3023–3058.

SANCETTA, A. AND SATCHELL, S. (2004). The bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory*, **20** (3), 535–562.

VAN DER VAART, A. AND WELLNER, J. (1996). *Weak convergence and empirical processes*. Springer, New-York.

VIDAKOVIC, B. (1999). *Statistical Modelling by Wavelets*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. John Wiley and Sons Inc.

Manuscript received 2023-07-10, revised 2024-01-23, accepted 2024-02-08.