Afrika Statistika

ISSN 2316-090X

Afrika Statistika
Vol. 13 (1), 2018, pages 1551–1568.
DOI: http://dx.doi.org/10.16929/as/1551.120

Uniform in Bandwidth Law of the Iterated Logarithm for a Transformation Kernel Estimator of Copulas

Cheikh Tidiane Seck[1,2,*], **Diam Ba**[2] **and Gane Samb Lo**[2,3,4]

[1] Cheikh Tidiane Seck, Université Alioune Diop de Bambey, Sénégal
[2] LERSTAD, Université Gaston Berger, Sénégal.
[3] LSTA, Université Pierre et Marie Curie, France (Affiliated)
[4] African University of Sciences and Technology, AUST, Abuja, Nigeria (Affiliated)

Selected Papers presented at the East African Conference of Statistical Mathematics and Applications, EACSMA 2017

Received September 26, 2017; Accepted February 27, 2018

Abstract. In this paper, we establish a uniform in bandwidth law of the iterated logarithm for the Transformation kernel estimator of bivariate copulas introduced in Omelka et al. (2009). To this end, we make use of a general empirical process approach inspired by the works in Mason and Swanepoel (2011). We obtain the asymptotic order of the maximal deviation of this estimator from its expectation. Then, we show that the bias converges asymptotically to zero at the same order provided that the second-order partial derivatives of the copula exist and are bounded. We also propose a bandwidth selection method by using a cross-validation approach. Finally, we compare in a simulation study the performances of the Transformation kernel estimator by considering two different methods of selecting the bandwidth.

Key words: Copula function; Kernel estimation; Transformation estimator; Law of the iterated logarithm; Uniform in bandwidth consistency; Cross-validation method.
AMS 2010 Mathematics Subject Classification : 62G05; 62G20; 62G15

---

*Corresponding author Cheikh Tidiane Seck : cheikhtidianeseck@uadb.edu.sn

Résumé (French) Nous établissons une loi du logarithme itéré uniforme en la fenêtre pour l'estimateur à noyau de Transformation de la copule bivariée proposé par Omelka et al. (2009). Notre méthodologie utilise une approche générale de processus empirique développée dans les travaux de Mason and Swanepoel (2011) et nous a permis d'obtenir la vitesse de convergence asymptotique de la deviation maximale de cet estimateur par rapport à son espérance. Nous avons aussi montré, sous des hypothèses assez douces concernant la fonction copule, que le biais de l'estimateur converge asymptotiquement vers 0, avec la même vitesse. Une méthode de sélection du paramètre de lissage h, utilisant l'approche de la validation croisée est proposée. Enfin, nous comparons par une étude de simulation cette approche avec une autre approche de sélection de fenêtre basée sur une distribution de référence et proposée par Omelka et al. (2009).

1. Introduction

Nonparametric kernel estimation of copulas suffers from boundary bias problem, mainly because a copula function has its support in the hypercube $[0,1]^d$. Because of this difficulty, there are a few works in the literature that paid special attention to the correction of this boundary bias problem. For instance, Gijbels and Mielniczuk (1990) proposed an estimator for the copula density by using a technique called *mirror-reflexion*. This technique permits to overcome the boundary bias by reflecting the original data with respect to the four edges and four corners of the unit square $[0,1]^2$. In the same spirit, Chen and Huang (2007) introduced a *local linear* kernel estimator for the copula function and a simple mathematical correction to remove the boundary bias. More recently, improved versions of the two previous estimators have been introduced in Omelka et al. (2009), where the boundary bias is taken into account by using a *shrinkage* principle that reduces the bandwidth near the corners of the unit square. In addition, Omelka et al. (2009) also provided a modification of the smoothed empirical copula estimator proposed in Fermanian (2004). This modification consists of transforming the data which allows to deal with unbounded domains in which the estimation of the copula can be done freely of boundary bias.

There exist in the literature some Transformation kernel estimators for the copula density (see, e.g.Geenens et al. (2014)). But in this paper, we are concerned with kernel estimation of the copula function itself considered as a distribution function. Namely, a copula function C is a multivariate distribution function with uniform margins in $[0,1]$; i.e. for example, for $d=2$,

$$C(u,v)=\mathbb{P}(U\leq u,V\leq v),\quad (u,v)\in[0,1]^2,$$

where U,V are [0,1]-Uniform random variables. It is a hidden function which is not directly observable in empirical studies, but reveals to be very useful for modeling dependence structure between random variables.

Let $(X_1,Y_1),...,(X_n,Y_n)$ be an independent and identically distributed random sample of a random vector (X,Y) with joint cumulative distribution function H and marginal distribution functions F and G, respectively. It is customary in Transformation kernel estimation to work with the pseudo-observations $\hat{U}_i=\frac{n}{n+1}F_n(X_i)$ and $\hat{V}_i=\frac{n}{n+1}G_n(Y_i)$, where F_n and G_n

denote the empirical marginal cumulative distribution functions associated with F and G, respectively. But to simplify calculations, we shall consider the uniform pseudo-observations $\hat{U}_i = F_n(X_i)$ and $\hat{V}_i = G_n(Y_i)$ which are asymptotically equivalent to the formers. For a bivariate multiplicative kernel $K(\cdot,\cdot)$; i.e., $K(x,y) = K(x)K(y)$, the Transformation kernel estimator for copula introduced in Omelka et al. (2009) is defined for all $(u,v) \in [0,1]^2$ as

$$\hat{C}_n^{(T)}(u,v) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{\phi^{-1}(u)-\phi^{-1}(\hat{U}_i)}{h_n}\right) K\left(\frac{\phi^{-1}(v)-\phi^{-1}(\hat{V}_i)}{h_n}\right), \tag{1}$$

where $\phi : \mathbb{R} \mapsto [0,1]$ is an increasing transformation, and $0 < h_n < 1$ denotes a bandwidth sequence. This estimator presents an advantage comparatively to the one proposed in Fermanian (2004), since it does not depend on the marginal distributions F and G. Indeed, the marginals of the transformed pseudo-observations, say $\hat{S}_i = \phi^{-1}(\hat{U}_i)$ and $\hat{T}_i = \phi^{-1}(\hat{V}_i)$, only depend on the transformation ϕ which determines their nature. For example, taking ϕ equal to the standard Gaussian distribution leads to Gaussian margins ; this procedure is known as the Probit transformation (see, e.g. Marron and Ruppert (1994) for more details). Note that the bias problem may be removed under certain regularity conditions on the function ϕ (see, e.g. Omelka et al. (2009) for more details).

A major problem of the estimator defined in (1) is the choice of the bandwidth h_n, as pointed out in Omelka et al. (2009). Indeed, since the asymptotic expressions of the bias and variance are not tractable, one cannot apply here the plug-in method which minimizes the asymptotic mean square error to determine the optimal bandwidth. But, one has to experiment other approaches in order to obtain an optimal bandwidth ensuring the consistency of the estimator. It turns out that deterministic sequences h_n are not suitable for many optimal bandwidth selectors which often produce data-dependent smoothing parameters. This motivated us to consider in this paper the following estimator :

$$\hat{C}_{n,h}^{(T)}(u,v) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{\phi^{-1}(u)-\phi^{-1}(\hat{U}_i)}{h}\right) K\left(\frac{\phi^{-1}(v)-\phi^{-1}(\hat{V}_i)}{h}\right), \tag{2}$$

where the deterministic bandwidth h_n in (1) is replaced with a variable bandwidth h, that may depend either on the data or the location point (u,v) at which the copula is estimated. In the literature, it is found that, for practical use, the most interesting choice for the bandwidth in kernel distribution function estimation is the data-driven method (see, e.g. Altman and Léger (1995)).

Our main purpose here is to establish a uniform in bandwidth law of the iterated logarithm for the Transformation kernel estimator $\hat{C}_{n,h}^{(T)}$ defined in (2). We obtain the uniform in bandwidth convergence rate of the maximal deviation of this estimator from its expectation by adapting the methodology developed in Mason and Swanepoel (2011), which utilizes general empirical process tools to establish such results for a wide class of kernel-type function estimators including : kernel density, regression and distribution function estimators. We also obtain the uniform in bandwidth convergence of the bias to zero, under some smooth conditions on the copula function C and the transformation ϕ.
The rest of the paper is organized as follows. In Section 2, we state our main theoretical results and give their proofs. Section 3 presents a practical method of bandwidth selection

which relies on a cross-validation criterion inspired by the works in Sard (1993). In Section 4, we make a simulation study with data generated from Archimedean copula. Finally, we give in Appendix the details of the proof of Proposition 1, which constitutes the fondation of this paper.

2. Main results and Proofs

Let $R_n = \left(\frac{n}{2\log\log n}\right)^{1/2}$. We shall also assume that the function $K(\cdot)$ is the integral of a symmetric bounded kernel $k(\cdot)$ supported on $[-1,1]$ and satisfying the following conditions :

(K.1) $\int_{-1}^{1} k(s)ds = 1$;

(K.2) $k(\cdot)$ is a 2-order kernel ; i.e., $\int_{-1}^{1} sk(s)ds = 0$ and $0 \neq \int_{-1}^{1} s^2k(s)ds < \infty$.

Here are our main results.

Theorem 1. *Suppose that the transformation ϕ admits a bounded derivative ϕ'. Then, for any sequence of positive constants $(b_n)_{n\geq 1}$ satisfying $0 < b_n < 1, b_n \to 0$ and $b_n \geq (\log n)^{-1}$, we have almost surely, for some $c > 0$,*

$$\limsup_{n\to\infty} \left\{ R_n \sup_{\frac{c\log n}{n} \leq h \leq b_n} \sup_{(u,v)\in[0,1]^2} \left| \hat{C}_{n,h}^{(T)}(u,v) - \mathbb{E}\hat{C}_{n,h}^{(T)}(u,v) \right| \right\} \leq 3. \tag{3}$$

Theorem 2. *Suppose that the copula function $C(u,v)$ has bounded second-order partial derivatives on $[0,1]^2$ and the transformation ϕ admits a bounded derivative ϕ'. Then, for any sequence of positive constants $(b_n)_{n\geq 1}$ satisfying $0 < b_n < 1$ and $\sqrt{n}b_n^2/\sqrt{\log\log n} = o(1)$, we have almost surely, as $n \to \infty$,*

$$R_n \sup_{0<h\leq b_n} \sup_{(u,v)\in[h,1-h]^2} \left| \mathbb{E}\hat{C}_{n,h}^{(T)}(u,v) - C(u,v) \right| \longrightarrow 0. \tag{4}$$

Remark.

1) A similar result can be found in Bouzebda (2012) for copula derivatives estimators in the multivariate case. But in this paper we deal with the Transformation Kernel Estimator of bivariate copula for which a law of the iterated logarithm is not available yet.
2) The uniformity in h allows us to apply various bandwidth selection rules such as plug-in, cross-validation or reference distribution methods, provided that the resulting optimal bandwidth belongs to a suitable interval $[a_n, b_n]$.
3) In Deheuvels and Mason (2004) a local plug-in type estimator $\hat{h}_n(x)$ of the bandwidth h is considered, with the following condition

$$\mathbb{P}(a_n \leq \hat{h}_n(x) \leq b_n,\ x \in \mathbb{R}) \to 1,$$

where $a_n = c_1h_n, b_n = c_2h_n,\ 0 < c_1 \leq c_2 < \infty$ and h_n is a sequence of positive constants converging to zero. Under slight conditions on the sequence h_n, Deheuvels and Mason (2004) proved probability versions of such results for the density and regression function estimators which may be applied to construct uniform confidence bands for these functionals. Note also that this methodology has been applied in Bâ et al. (2015) to establish simultaneous confidence bands for the bivariate copula, using the *local linear* kernel estimator proposed in Chen and Huang (2007).

Proof. (**Theorem 1**) We begin by some notation. Recall that H_n, F_n and G_n are the empirical cumulative distribution functions of H, F and G, respectively. Then the copula estimator based directly on Sklar's Theorem (See also **?** for a new and direct proof) can be defined as

$$C_n(u,v) = H_n(F_n^{-1}(u), G_n^{-1}(v)),$$

with $F_n^{-1}(u) = \inf\{x : F_n(x) \geq u\}$ and $G_n^{-1}(v) = \inf\{x : F_n(x) \geq v\}$ the quantile functions corresponding to F_n and G_n. Denote the bivariate empirical copula process as

$$\mathbb{C}_n(u,v) = \sqrt{n}[C_n(u,v) - C(u,v)], \qquad (u,v) \in [0,1]^2$$

and introduce the following quantity

$$\widetilde{C}_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{U_i \leq u, V_i \leq v\}$$

which represents the uniform bivariate empirical distribution function based on a sample $(U_1, V_1), \cdots, (U_n, V_n)$ of independent and identically distributed random variables with marginals uniformly distributed on $[0,1]$. Define the following empirical process :

$$\widetilde{\mathbb{C}}_n(u,v) = \sqrt{n}[\widetilde{C}_n(u,v) - C(u,v)], \quad (u,v) \in [0,1]^2.$$

We have

$$\begin{aligned}
\widetilde{\mathbb{C}}_n(u,v) &= \sqrt{n}[\widetilde{C}_n(u,v) - C_n(u,v)] + \sqrt{n}[C_n(u,v) - C(u,v)] \\
&= \sqrt{n}[\widetilde{C}_n(u,v) - C_n(u,v)] + \mathbb{C}_n(u,v).
\end{aligned}$$

Applying Proposition 1.8 in Deheuvels (2009), we can write

$$\widetilde{C}_n(u,v) - C_n(u,v) = O\left(\frac{1}{n}\right);$$

that is,

$$\widetilde{\mathbb{C}}_n(u,v) = \mathbb{C}_n(u,v) + O\left(\frac{1}{\sqrt{n}}\right). \tag{5}$$

For every $n \geq 1$, $0 < h < 1$, set

$$D_{n,h}(u,v) := \hat{C}_{n,h}^{(T)}(u,v) - \mathbb{E}\hat{C}_{n,h}^{(T)}(u,v)$$

and

$$g_{n,h} := \hat{C}_{n,h}^{(T)}(u,v) - \widetilde{C}_n(u,v).$$

Then, one can write

$$\begin{aligned}
g_{n,h} &= \frac{1}{n}\sum_{i=1}^{n}\left[K\left(\frac{\phi^{-1}(u) - \phi^{-1}(\hat{U}_i)}{h}\right)K\left(\frac{\phi^{-1}(v) - \phi^{-1}(\hat{V}_i)}{h}\right) - \mathbb{I}\{U_i \leq u, V_i \leq v\}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[K\left(\frac{\phi^{-1}(u) - \phi^{-1}(F_n \circ F^{-1}(U_i))}{h}\right)K\left(\frac{\phi^{-1}(v) - \phi^{-1}(G_n \circ G^{-1}(V_i))}{h}\right) - \mathbb{I}\{U_i \leq u, V_i \leq v\}\right] \\
&=: \frac{1}{n}\sum_{i=1}^{n} g(U_i, V_i, h),
\end{aligned}$$

where g belongs to the class of measurable functions $\mathcal{G}_n$ defined as

$$\mathcal{G}_n = \left\{ g : g(s,t,h) = K\left(\frac{\phi^{-1}(u)-\phi^{-1}(\zeta_{1,n}(s))}{h}\right) K\left(\frac{\phi^{-1}(v)-\phi^{-1}(\zeta_{2,n}(t))}{h}\right) - \mathbb{I}\{s \leq u, t \leq v\}, u,v \in [0,1], 0 < h < 1 \right\}, \tag{6}$$

where $\zeta_{1,n}(s) = F_n \circ F^{-1}(s)$ and $\zeta_{2,n}(t) = G_n \circ G^{-1}(t)$.
Since $\mathbb{E}\widetilde{C}_n(u,v) = C(u,v)$, one can observe that

$$\sqrt{n}|g_{n,h} - \mathbb{E}g_{n,h}| = |\sqrt{n}D_{n,h}(u,v) - \widetilde{\mathbb{C}}_n(u,v)|.$$

Now, we have to apply the main Theorem of Mason and Swanepoel (2011), which gives the order of convergence of the deviation of kernel-type function estimators from their expectation. To this end, the above class of functions $\mathcal{G}_n$ must satisfy the following four conditions :

(G.i) There exists a finite constant $\kappa > 0$ such that

$$\sup_{0\leq h\leq 1} \sup_{g\in\mathcal{G}_n} \|g(\cdot,\cdot,h)\|_\infty = \kappa < \infty.$$

(G.ii) There exists a constant $C' > 0$ such that for all $h \in [0,1]$,

$$\sup_{g\in\mathcal{G}_n} \mathbb{E}\left[g^2(U,V,h)\right] \leq C'h.$$

(F.i) $\mathcal{G}_n$ satisfies the uniform entropy condition, i.e.,

$$\exists\, C_0 > 0, \nu_0 > 0 \;:\; N(\epsilon, \mathcal{G}_n) \leq C_0\epsilon^{-\nu_0}.$$

(F.ii) $\mathcal{G}_n$ is a pointwise measurable class, i.e there exists a countable sub-class $\mathcal{G}_0$ of $\mathcal{G}_n$ such that for all $g \in \mathcal{G}$, there exits $(g_m)_m \subset \mathcal{G}_0$ such that $g_m \longrightarrow g$.

Checking of these conditions will be done in Appendix and constitutes the proof of the following proposition.

Proposition 1. *Suppose that the transformation ϕ admits a bounded derivative ϕ'. Then assuming (G.i), (G.ii), (F.i) and (F.ii), we have for some $c > 0,\ 0 < h_0 < 1$, with probability one,*

$$\limsup_{n\to\infty} \sup_{\frac{c\log n}{n}\leq h\leq h_0} \sup_{(u,v)\in(0,1)^2} \frac{|\sqrt{n}D_{n,h}(u,v) - \tilde{\mathbb{C}}_n(u,v)|}{\sqrt{h(|\log h| \vee \log\log n)}} = A(c),$$

where $A(c)$ is a positive constant.

Corollary 1. *Under the assumptions of Proposition 1, one has for any sequence of constants $0 < b_n < 1$, satisfying $b_n \to 0,\ b_n \geq (\log n)^{-1}$, with probability one, as $n \to \infty$,*

$$\sup_{\frac{c\log n}{n}\leq h\leq b_n} \sup_{(u,v)\in(0,1)^2} \frac{|\sqrt{n}D_{n,h}(u,v) - \tilde{\mathbb{C}}_n(u,v)|}{\sqrt{\log\log n}} = O(\sqrt{b_n}).$$

Proof. (**Corollary 1**)
First, observe that the condition $b_n \geq (\log n)^{-1}$ implies

$$\frac{|\log b_n|}{\log\log n} \leq 1. \tag{7}$$

Next, by the monotonicity of the function $x \mapsto x|\log x|$ on $[0, 1/e]$, one can write for n large enough, $h|\log h| \leq b_n|\log b_n|$ and hence,

$$h(|\log h| \vee \log\log n) \leq b_n(|\log b_n| \vee \log\log n). \tag{8}$$

Combining this with Proposition 1, we obtain

$$\sup_{\frac{c\log n}{n} \leq h \leq b_n} \sup_{(u,v)\in(0,1)^2} \frac{|\sqrt{n}D_{n,h}(u,v) - \tilde{\mathbb{C}}_n(u,v)|}{\sqrt{b_n \log\log n \left(\frac{|\log b_n|}{\log\log n} \vee 1\right)}} = O(1).$$

Thus the Corollary 1 follows from (7).

Now, coming back to the proof of Theorem 1, we have to show that the deviation $D_{n,h}(u,v)$, suitably normalized, is almost surely uniformly bounded, as $n \to \infty$. To this end, it suffices to prove that

$$\limsup_{n\to\infty} \sup_{\frac{c\log n}{n} \leq h \leq b_n} \sup_{(u,v)\in[0,1]^2} \frac{|\sqrt{n}D_{n,h}(u,v)|}{\sqrt{2\log\log n}} \leq 3. \tag{9}$$

We will make use of an approximation of the empirical copula process $\mathbb{C}_n$ by a Kiefer process (see e.g., Zari (2010), page 100). Let $\mathbb{W}(u,v,t)$ denote a 3-parameter Wiener process defined on $[0,1]^2 \times [0,\infty)$. Then the Gaussian process $\mathbb{K}(u,v,t) = \mathbb{W}(u,v,t) - \mathbb{W}(1,1,t).uv$ is called a 3-parameter Kiefer process defined on $[0,1]^2 \times [0,\infty)$.

By Theorem 3.2 in Zari (2010), there exists a sequence of Gaussian processes $\{\mathbb{K}_C(u,v,n), u,v \in [0,1], n > 0\}$ such that

$$\sup_{(u,v)\in[0,1]^2} \left|\sqrt{n}\mathbb{C}_n(u,v) - \mathbb{K}_C^*(u,v,n)\right| = O\left(n^{3/8}(\log n)^{3/2}\right),$$

where

$$\mathbb{K}_C^*(u,v,n) = \mathbb{K}_C(u,v,n) - \mathbb{K}_C(u,1,n)\frac{\partial C(u,v)}{\partial u} - \mathbb{K}_{\mathbb{C}}(1,v,n)\frac{\partial C(u,v)}{\partial v}.$$

This yields

$$\limsup_{n\to\infty} \sup_{(u,v)\in[0,1]^2} \frac{|\mathbb{C}_n(u,v)|}{\sqrt{2\log\log n}} = \limsup_{n\to\infty} \sup_{(u,v)\in[0,1]^2} \frac{|\mathbb{K}_C^*(u,v,n)|}{\sqrt{2n\log\log n}}. \tag{10}$$

By the works on the law of the iterated logarithm in Wichura (1973), one has almost surely, for $d = 2$,

$$\limsup_{n\to\infty} \sup_{(u,v)\in[0,1]^2} \frac{|\mathbb{K}_{\mathbb{C}}^*(u,v,n)|}{\sqrt{2n\log\log n}} \leq 3, \tag{11}$$

which entails

$$\limsup_{n\to\infty} \sup_{(u,v)\in[0,1]^2} \frac{|\mathbb{C}_n(u,v)|}{\sqrt{2\log\log n}} \leq 3.$$

Since $\mathbb{C}_n(u,v)$ and $\tilde{\mathbb{C}}_n(u,v)$ are asymptotically equivalent in view of (5), one obtains

$$\limsup_{n\to\infty} \sup_{(u,v)\in[0,1]^2} \frac{\left|\tilde{\mathbb{C}}_n(u,v)\right|}{\sqrt{2\log\log n}} \leq 3.$$

Applying Corollary 1 and recalling the fact that $b_n \to 0$, one obtains (9) which completes the proof of Theorem 1.

Proof. (**Theorem 2**) Let

$$B_{n,h}(u,v) = \mathbb{E}\hat{C}^{(T)}_{n,h}(u,v) - C(u,v).$$

By hypothesis (K.1) on the kernel $k(\cdot)$, one can write for all $(u,v) \in [0,1]^2$,

$$C(u,v) = \int_{-1}^{1}\int_{-1}^{1} C(u,v)k(s)k(t)dsdt.$$

Recall that $\zeta_{1,n}(U_i) = \hat{U}_i = F_n \circ F^{-1}(U_i)$ and $\zeta_{2,n}(V_i) = \hat{V}_i = G_n \circ G^{-1}(V_i)$. Then we have

$$\begin{aligned}\mathbb{E}\hat{C}^{(T)}_{n,h}(u,v) &= \mathbb{E}\left[K\left(\frac{\phi^{-1}(u)-\phi^{-1}(\hat{U}_i)}{h}\right)K\left(\frac{\phi^{-1}(v)-\phi^{-1}(\hat{V}_i)}{h}\right)\right]\\ &= \int_{-1}^{1}\int_{-1}^{1} \mathbb{E}\mathbb{I}\{U_i \leq \zeta^{-1}_{1,n}[\phi(\phi^{-1}(u)-sh)], V_i \leq \zeta^{-1}_{2,n}[\phi(\phi^{-1}(v)-th)]\}k(s)k(t)dsdt\\ &= \int_{-1}^{1}\int_{-1}^{1} C\left(\zeta^{-1}_{1,n}[\phi(\phi^{-1}(u)-sh)], \zeta^{-1}_{2,n}[\phi(\phi^{-1}(v)-th)]\right)k(s)k(t)dsdt\end{aligned}$$

and

$$B_{n,h}(u,v) = \int_{-1}^{1}\int_{-1}^{1}\left[C\left(\zeta^{-1}_{1,n}[\phi(\phi^{-1}(u)-sh)], \zeta^{-1}_{2,n}[\phi(\phi^{-1}(v)-th)]\right) - C(u,v)\right]k(s)k(t)dsdt. \tag{12}$$

Making use of Chung's (1949) law of the iterated logarithm, we can infer that, whenever F is continuous and admits a bounded density f, for all $u \in [0,1]$, as $n \to \infty$,

$$\begin{aligned}\zeta^{-1}_{1,n}(u) - u = F\circ F_n^{-1}(u) - F\circ F^{-1}(u) &= f(c_n)[F_n^{-1}(u) - F^{-1}(u)],\ c_n \in [F_n^{-1}(u)\wedge F^{-1}(u), F_n^{-1}(u)\vee F^{-1}(u)]\\ &= O\left(\sqrt{\log\log n/2n}\right).\end{aligned}$$

This implies that $\zeta^{-1}_{1,n}(u)$ is asymptotically equivalent to u. As well, we have $\zeta^{-1}_{2,n}(v) = G\circ G_n^{-1}(v)$ is asymptotically equivalent to v, for all $v \in [0,1]$. Thus, for all large n, one can write

$$B_{n,h}(u,v) = \int_{-1}^{1}\int_{-1}^{1}\left[C(\phi(\phi^{-1}(u)-sh), \phi(\phi^{-1}(v)-th)) - C(u,v)\right]k(s)k(t)dsdt + o(1).$$

By applying a 2-order Taylor expansion for the copula function C, we obtain

$$\begin{gathered} C(\phi(\phi^{-1}(u)-sh),\phi(\phi^{-1}(v)-th))-C(u,v)= \\ [\phi(\phi^{-1}(u)-sh)-u]C_u(u,v)+[\phi(\phi^{-1}(v)-th)-v]C_v(u,v)+[\phi(\phi^{-1}(u)-sh)-u]^2\tfrac{C_{uu}(u,v)}{2} \\ +[\phi(\phi^{-1}(v)-th)-v]^2\tfrac{C_{vv}(u,v)}{2}+[\phi(\phi^{-1}(u)-sh)-u][\phi(\phi^{-1}(v)-th)-v]C_{uv}(u,v)+o(h^2), \end{gathered}$$

where

$$C_u(u,v)=\frac{\partial C}{\partial u}(u,v) \;\; ; \;\; C_v(u,v)=\frac{\partial C}{\partial v}(u,v)$$

and where

$$C_{uu}(u,v)=\frac{\partial^2 C}{\partial u^2}(u,v)\;;\;C_{vv}(u,v)=\frac{\partial^2 C}{\partial v^2}(u,v)\;;\;C_{uv}(u,v)=\frac{\partial^2 C}{\partial u\partial v}(u,v).$$

Applying again a 1-order Taylor expansion for the function ϕ, we get

$$\phi(\phi^{-1}(u)-sh)-u=\phi(\phi^{-1}(u)-sh)-\phi(\phi^{-1}(u))=-\phi'(\phi^{-1}(u))sh+o(h)$$

and

$$\phi(\phi^{-1}(v)-sh)-v=\phi(\phi^{-1}(v)-th)-\phi(\phi^{-1}(v))=-\phi'(\phi^{-1}(v))th+o(h).$$

Thus

$$\begin{aligned} C(\phi(\phi^{-1}(u)-sh),\phi(\phi^{-1}(v)-th))-C(u,v) \;=\; & -\phi'(\phi^{-1}(u))shC_u(u,v)-\phi'(\phi^{-1}(v))thC_v(u,v)+o(h) \\ & +[\phi'(\phi^{-1}(u))sh]^2\frac{C_{uu}(u,v)}{2}+[\phi'(\phi^{-1}(v))th]^2\frac{C_{vv}(u,v)}{2} \\ & +[\phi'(\phi^{-1}(u))][\phi'(\phi^{-1}(u))]sth^2C_{uv}(u,v)+o(h^2). \end{aligned}$$

Now, by using the fact that $k(\cdot)$ is a 2-order kernel function ; i.e., $\int_{-1}^{1} sk(s)ds=0$ and $\int_{-1}^{1} s^2k(s)ds\neq 0$, we obtain, by Fubini's Theorem, that for all $(u,v)\in[h,1-h]^2$,

$$B_{n,h}(u,v)=\frac{h^2}{2}\left[\phi'(\phi^{-1}(u))^2C_{uu}(u,v)\int_{-1}^{1}s^2k(s)ds+\phi'(\phi^{-1}(v))^2C_{vv}(u,v)\int_{-1}^{1}t^2k(t)dt\right]+o(1). \tag{13}$$

Since the second-order partial derivatives C_{uu}, C_{vv} and ϕ' are assumed to be bounded, we obtain

$$\sup_{0<h\leq b_n}\sup_{(u,v)\in[h,1-h]^2}B_{n,h}(u,v)=O(b_n^2).$$

Thus

$$\left(\frac{n}{2\log\log n}\right)^{1/2}\sup_{0<h\leq b_n}\sup_{(u,v)\in[h,1-h]^2}B_{n,h}(u,v)=O\left(\frac{\sqrt{n}b_n^2}{\sqrt{2\log\log n}}\right)=o(1), \tag{14}$$

which completes the proof of Theorem 2.

3. Bandwidth choice

As pointed out in Omelka et al. (2009), the choice of the bandwidth for the Transformation estimator $\hat{C}_{n,h}^{(T)}(u,v)$ is problematic. Indeed, since the asymptotic expressions of the bias and variance of this estimator are not tractable, one cannot apply the plug-in method which relies on the minimization of the asymptotic *mean integrated square error*. Instead, one may use a cross-validation method to select an optimal bandwidth for the kernel copula estimator $\hat{C}_{n,h}^{(T)}(u,v)$.
Let

$$C_n(u,v) = H_n(F_n^{-1}(u), G_n^{-1}(v))$$

be the empirical copula estimator directly based on Sklar's Theorem, where H_n, F_n and G_n represent the empirical cumulative distribution functions of H, F and G, respectively. F_n^{-1} and G_n^{-1} are the empirical quantile distributions of F and G. For any couple of pseudo-observations $(\hat{U}_i,\ \hat{V}_i\ )$, let

$$\hat{C}_{n,h,-i}^{(T)}(u,v) = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} K\left(\frac{\phi^{-1}(u)-\phi^{-1}(\hat{U}_j)}{h}\right) K\left(\frac{\phi^{-1}(v)-\phi^{-1}(\hat{V}_j)}{h}\right)$$

denote the leave-$(\hat{U}_i,\ \hat{V}_i)$-out version of $\hat{C}_{n,h}^{(T)}(u,v)$, computed with all pseudo-observations except the couple $(\hat{U}_i,\ \hat{V}_i)$. By taking a same weight equal to the unity for all pseudo-observations, we may here define Sarda's criterion (see, Sard (1993)) as follows :

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{C}_{n,h,-i}^{(T)}(\hat{U}_i,\hat{V}_i) - C_n(\hat{U}_i,\hat{V}_i)\right]^2. \tag{15}$$

Let $a_n = \frac{c\log n}{n}$ and (b_n) a sequence satisfying conditions of Theorems 1 and 2. The optimal bandwidth, say $\hat{h}_{opt}$, is then solution to the following minimization problem :

$$\min_{h\in[a_n,b_n]} CV(h).$$

Since $\hat{h}_{opt}$ must be in the interval $[a_n, b_n]$, then the strong consistency of the estimator $\hat{C}_{n,\hat{h}_{opt}}^{(T)}$ is guarranted by applying Theorems 1 and 2.

4. Simulation study

Here, we give numerical experiments to compare the performances of our Transformation kernel estimator $\hat{C}_{n,h}^{(T)}$ constructed separately with two optimal bandwidths, say $\hat{h}_{CV}$ and $\hat{h}_{Om}$, selected *via* the above cross-validation method and the reference distribution method (see, e.g.Omelka et al. (2009)), respectively. Towards this end, we first determine graphically the optimal bandwidth for each method by visualizing the curve of its criterion over $h \in [a_n, b_n]$. In this study, we take $a_n = \log n/n$ (with c=1) and $b_n = \left(\frac{\log\log n}{n^2}\right)^{1/4}$ in such a way that Theorems 1 and 2 hold. Thus, for a given sample size n, we may compute the interval $[a_n, b_n]$, which is equal to $[0.08, 0.15]$ for $n = 50$ and $[0.04, 0.11]$ for $n = 100$.

The reference distribution rule proposed in Omelka et al. (2009) is based on the minimization of the following criterion:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\hat{H}_n(x,y)-H(x,y)\right)^2 h(x,y)dxdy, \tag{16}$$

where $H(x,y)$ is taken equal to a bivariate normal distribution with density $h(x,y)$, and $\hat{H}_n(x,y)$ is a bivariate kernel distribution function estimator defined as

$$\hat{H}_n(x,y)=\frac{1}{n}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)K\left(\frac{y-Y_i}{h}\right), \tag{17}$$

where the (X_i,Y_i)'s are independent replicae of the random couple (X,Y). Notice that the expression in (16) can be seen as the expectation of the quantity $[\hat{H}_n(X,Y)-H(X,Y)]^2$, and hence may be approximated by the criterion

$$Om(h)=\frac{1}{n}\sum_{i=1}^{n}\left[\hat{H}_n(X_i,Y_i)-H(X_i,Y_i)\right]^2. \tag{18}$$

Let us now consider a 0.001-step grid of points in the interval $[a_n,b_n]$ and represent the curves of the two criteria $CV(h)$ and $Om(h)$ in Figure 1 below. To compute these criteria,

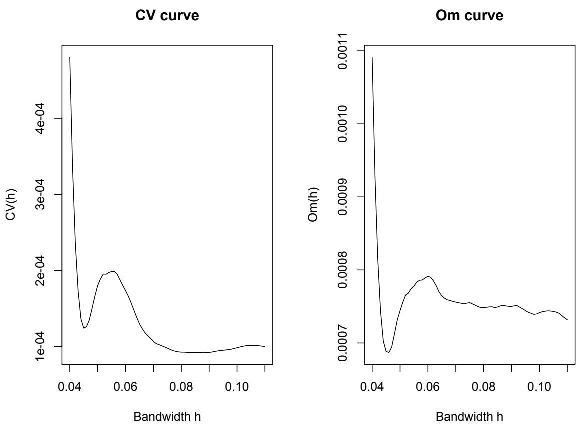

Fig. 1. Curves of criteria $CV(h)$ and $Om(h)$, over $h\in[a_n,b_n]$ with $n=100$

we generate data from Frank's copula by using the conditional sampling method for fixed parameters. Frank's copula has bounded second-order partial derivatives and is defined for any given parameter $\theta\in\mathbb{R}^*$, as

$$C_\theta(u,v)=-\frac{1}{\theta}\log\left[1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{(e^{-\theta}-1)}\right]. \tag{19}$$

We also choose the Epanechnikov kernel density $k(t) = 0.75(1-t^2)\mathbb{I}(|t| \leq 1)$ to compute the integral function $K(\cdot)$, whereas the transformation $\phi(\cdot)$ is taken to be the standard Gaussian distribution.

From Figure 1, we can observe that the cross-validation criterion $CV(h)$ is minimal at value $h = 0.085$, while the reference distribution criterion $Om(h)$ reaches its minimum at value $h = 0.045$. So, we may take $\hat{h}_{CV} = 0.085$ and $\hat{h}_{Om} = 0.045$ as the optimal bandwidths for $CV(h)$ and $Om(h)$ criteria, respectively.

Now, we estimate the mean square error (mse), with respect to Frank's copula C_θ, of our Transformation kernel estimator $\hat{C}^{(T)}_{n,h}(u,v)$, where h is replaced by $\hat{h}_{CV}$ or $\hat{h}_{Om}$. To this end, we replicate $B = 1000$ samples of size n from C_θ and apply the formula

$$mse\left(\hat{C}^{(T)}_{n,h}(u,v)\right) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{C}^{(T)}_{n,h,b}(u,v) - C_\theta(u,v)\right)^2,$$

where $\hat{C}^{(T)}_{n,h,b}(u,v)$ is the value of $\hat{C}^{(T)}_{n,h}(u,v)$ obtained with the b^{th} sample. For arbitrary values of $\theta = -2, 1, 5$ and randomly chosen couples $(u,v) \in [0,1]^2$; we obtain the results in Table 1 which contains the values of the ratio r defined as

$$r(u,v) = \frac{mse\left(\hat{C}^{(T)}_{n,\hat{h}_{CV}}(u,v)\right)}{mse\left(\hat{C}^{(T)}_{n,\hat{h}_{Om}}(u,v)\right)}.$$

This ratio fluctuates around the unity and suggests that the proposed cross-validation method is as performant as the distribution reference method in selecting the optimal bandwidth for the Transformation kernel estimator of bivariate copula. We also observe that for negative values of θ, our cross-validation method seems to be better than the reference distribution rule, because the ratio r is most of the time less than the unity.

	$\theta = -2$		$\theta = 1$		$\theta = 5$	
(u,v)	$n=50$	$n=100$	$n=50$	$n=100$	$n=50$	$n=100$
$(0.48, 0.25)$	0.59	0.93	1.03	0.97	0.83	1.16
$(0.12, 0.80)$	1.73	0.95	1.02	0.80	1.03	1.10
$(0.35, 0.68)$	0.85	0.45	0.95	0.81	1.27	2.05
$(0.73, 0.21)$	1.26	1.09	1.48	1.00	0.79	0.89
$(0.84, 0.44)$	0.38	0.63	1.12	0.69	1.06	1.86
$(0.14, 0.67)$	0.89	1.02	1.16	0.07	0.82	1.00
$(0.29, 0.25)$	0.53	0.81	0.91	0.98	0.71	1.70
$(0.57, 0.50)$	0.70	0.62	1.48	1.03	1.31	0.62

Table 1. Values of the ratio r for some couples (u,v).

Appendix

Proof. (**Proposition 1**)
Recall that $K(\cdot,\cdot)$ is the integral of a symmetric bounded multiplicative kernel $k(s,t) = k(s)k(t)$, supported on $[-1,1]^2$; i.e., $K(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} k(s,t)dsdt$. We have to check (G.i), (G.ii), (F.i) and (F.ii).

Checking for (G.i): Recall that $(U_i,V_i), i \geq 1$ are iid random variables uniformly distributed on $[0,1]^2$, $\zeta_{1,n}(U_i) = F_n o F^{-1}(U_i)$ and $\zeta_{2,n}(V_i) = G_n o G^{-1}(V_i)$. For any function $g \in \mathcal{G}_n$ and $0 < h < 1$, we can write

$$\begin{aligned} g\left(U_i,V_i,h\right) &= K\left(\frac{\phi^{-1}(u)-\phi^{-1}(\zeta_{1,n}(U_i))}{h}\right)K\left(\frac{\phi^{-1}(v)-\phi^{-1}(\zeta_{2,n}(V_i))}{h}\right) - \mathbb{I}\{U_i \leq u, V_i \leq v\} \\ &= \int_{-\infty}^{\frac{\phi^{-1}(u)-\phi^{-1}(\zeta_{1,n}(U_i))}{h}} \int_{-\infty}^{\frac{\phi^{-1}(v)-\phi^{-1}(\zeta_{2,n}(V_i))}{h}} k(s,t)dsdt - \mathbb{I}\{U_i \leq u, V_i \leq v\} \\ &= \int_{-1}^{1}\int_{-1}^{1} \mathbb{I}\left\{U_i \leq \zeta_{1,n}^{-1}\circ\phi(\phi^{-1}(u)-th), V_i \leq \zeta_{2,n}^{-1}\circ\phi(\phi^{-1}(v)-sh)\right\}k(s,t)dsdt \\ &\quad -\mathbb{I}\{U_i \leq u, V_i \leq v\} \\ &\leq \int_{-1}^{1}\int_{-1}^{1} k(s,t)dsdt - \mathbb{I}\{U_i \leq u, V_i \leq v\} \ \leq 4\|k\|^2+1, \end{aligned}$$

where $\|k\| = \sup_{(s,t)\in[-1,1]^2} |k(s,t)|$ represents the supremum norm on $[-1,1]^2$. Thus (G.i) holds by taking $\kappa := 4\|k\|^2 + 1$.

Checking for (G.ii): We have to show that $\sup_{g\in\mathcal{G}_n} \mathbb{E}g^2(U,V,h) \leq C'h$, where C' is a positive constant. One can write

$$\begin{aligned} \mathbb{E}g^2(U,V,h) &= \mathbb{E}\left[K\left(\tfrac{\phi^{-1}(u)-\phi^{-1}(\zeta_{1,n}(U))}{h}\right)K\left(\tfrac{\phi^{-1}(v)-\phi^{-1}(\zeta_{2,n}(V))}{h}\right) - \mathbb{I}\{U\leq u, V \leq v\}\right]^2 \\ &= \mathbb{E}\left[K^2\left(\tfrac{\phi^{-1}(u)-\phi^{-1}(\zeta_{1,n}(U))}{h}\right)K^2\left(\tfrac{\phi^{-1}(v)-\phi^{-1}(\zeta_{2,n}(V))}{h}\right)\right] \\ &\quad -2\mathbb{E}\left[K\left(\tfrac{\phi^{-1}(u)-\phi^{-1}(\zeta_{1,n}(U))}{h}\right)K\left(\tfrac{\phi^{-1}(v)-\phi^{-1}(\zeta_{2,n}(V))}{h}\right)\mathbb{I}\{U\leq u, V \leq v\}\right] + C(u,v) \\ &=: A - 2B + C(u,v) \end{aligned}.$$

Since the function $K(\cdot,\cdot)$ is a kernel of a distribution function, we may assume that it takes its values in $[0,1]$. Then, we can use the inequality $K^2(x,y) \leq K(x,y)$ to bound up the term A in the right hand side of the previous equality.

$$\begin{aligned} A &= \mathbb{E}\left[K^2\left(\tfrac{\phi^{-1}(u)-\phi^{-1}(\zeta_{1,n}(U))}{h}\right)K^2\left(\tfrac{\phi^{-1}(v)-\phi^{-1}(\zeta_{2,n}(V))}{h}\right)\right] \\ &\leq \mathbb{E}\left[K\left(\tfrac{\phi^{-1}(u)-\phi^{-1}(\zeta_{1,n}(U))}{h}\right)K\left(\tfrac{\phi^{-1}(v)-\phi^{-1}(\zeta_{2,n}(V))}{h}\right)\right] \\ &\leq \mathbb{E}\left[\int_{-1}^{1}\int_{-1}^{1}\mathbb{I}\left\{U \leq \zeta_{1,n}^{-1}\circ\phi(\phi^{-1}(u)-sh), V \leq \zeta_{2,n}^{-1}\circ\phi(\phi^{-1}(v)-th)\right\}k(s,t)dsdt\right]. \end{aligned}$$

The other term B can be written into

$$B = \mathbb{E}\left[K\left(\tfrac{\phi^{-1}(u)-\phi^{-1}(\zeta_{1,n}(U))}{h}\right)K\left(\tfrac{\phi^{-1}(v)-\phi^{-1}(\zeta_{2,n}(V))}{h}\right)\mathbb{I}\{U\le u, V\le v\}\right]$$
$$= \mathbb{E}\left[\int_{-1}^{1}\int_{-1}^{1}\mathbb{I}\left\{s\le \tfrac{\phi^{-1}(u)-\phi^{-1}(\zeta_{1,n}(U))}{h}, t\le \tfrac{\phi^{-1}(v)-\phi^{-1}(\zeta_{2,n}(V))}{h}\right\}\mathbb{I}\{U\le u, V\le v\}k(s,t)dsdt\right]$$
$$= \mathbb{E}\left[\int_{-1}^{1}\int_{-1}^{1}\mathbb{I}\left\{U\le u\wedge \zeta_{1,n}^{-1}\circ\phi(\phi^{-1}(u)-sh), V\le v\wedge \zeta_{2,n}^{-1}\circ\phi(\phi^{-1}(v)-th)\right\}k(s,t)dsdt\right],$$

where $x\wedge y=\min(x,y)$. Note that

$$C(u,v)=\int_{-1}^{1}\int_{-1}^{1}C(u,v)k(s,t)dsdt,$$

as the kernel $k(\cdot,\cdot)$ satisfies $\int_{-1}^{1}\int_{-1}^{1}k(s,t)dsdt=1$. Thus

$$\mathbb{E}g^2(U,V,h)\le \mathbb{E}\left[\int_{-1}^{1}\int_{-1}^{1}\mathbb{I}\left\{U\le \zeta_{1,n}^{-1}\circ\phi(\phi^{-1}(u)-sh), V\le \zeta_{2,n}^{-1}\circ\phi(\phi^{-1}(v)-th)\right\}k(s,t)dsdt\right]$$
$$-2\mathbb{E}\left[\int_{-1}^{1}\int_{-1}^{1}\mathbb{I}\left\{U\le u\wedge \zeta_{1,n}^{-1}\circ\phi(\phi^{-1}(u)-sh), V\le v\wedge \zeta_{2,n}^{-1}\circ\phi(\phi^{-1}(v)-th)\right\}k(s,t)dsdt\right]$$
$$+\int_{-1}^{1}\int_{-1}^{1}C(u,v)k(s,t)dsdt.$$

By using the same arguments (Chung's (1949) LIL) as in the proof of Theorem 2, we can write for all large n,

$$\mathbb{E}g^2(U,V,h)\le \mathbb{E}\left[\int_{-1}^{1}\int_{-1}^{1}\mathbb{I}\left\{U\le \phi(\phi^{-1}(u)-sh), V\le \phi(\phi^{-1}(v)-th)\right\}k(s,t)dsdt\right]$$
$$-2\mathbb{E}\left[\int_{-1}^{1}\int_{-1}^{1}\mathbb{I}\left\{U\le u\wedge \phi(\phi^{-1}(u)-sh), V\le v\wedge \phi(\phi^{-1}(v)-th)\right\}k(s,t)dsdt\right]$$
$$+\int_{-1}^{1}\int_{-1}^{1}C(u,v)k(s,t)dsdt+o(1).$$

That is,

$$\begin{aligned}\mathbb{E}g^2(U,V,h) \;\le\; & \int_{-1}^{1}\int_{-1}^{1}C\left(\phi(\phi^{-1}(u)-sh),\phi(\phi^{-1}(v)-th)\right)k(s,t)dsdt\\ & -2\int_{-1}^{1}\int_{-1}^{1}C\left(u\wedge\phi(\phi^{-1}(u)-sh), v\wedge\phi(\phi^{-1}(v)-th)\right)k(s,t)dsdt \\ & +\int_{-1}^{1}\int_{-1}^{1}C(u,v)k(s,t)dsdt+o(1).\end{aligned} \tag{20}$$

Now, we have to discuss condition (G.ii) in the four following cases:

Case 1. $u\wedge\phi(\phi^{-1}(u)-sh)=\phi(\phi^{-1}(u)-sh)$ and $v\wedge\phi(\phi^{-1}(v)-th)=\phi(\phi^{-1}(v)-th)$. In this case the second member of inequality (20) is reduced, and we have

$$\begin{aligned}\mathbb{E}g^2(U,V,h) \;\le\; & -\int_{-1}^{1}\int_{-1}^{1}C\left(\phi(\phi^{-1}(u)-sh),\phi(\phi^{-1}(v)-th)\right)k(s,t)dsdt\\ & +\int_{-1}^{1}\int_{-1}^{1}C(u,v)k(s,t)dsdt+o(1)\\ \le\; & \int_{-1}^{1}\int_{-1}^{1}\left[C(u,v)-C\left(\phi(\phi^{-1}(u)-sh),\phi(\phi^{-1}(v)-th)\right)\right]k(s,t)dsdt+o(1).\end{aligned}$$

By a Taylor expansion for the copula function C, we have

$$\begin{aligned} C(u,v) - C(\phi(\phi^{-1}(u)-sh), \phi(\phi^{-1}(v)-th)) &= [u-\phi(\phi^{-1}(u)-sh)]C_u(u,v) \\ &+ [v-\phi(\phi^{-1}(v)-th)]C_v(u,v) + o(h). \end{aligned}$$

Applying again a Taylor-Young expansion for the function ϕ, we obtain

$$u-\phi(\phi^{-1}(u)-sh) = \phi(\phi^{-1}(u)) - \phi(\phi^{-1}(u)-sh) = \phi'(u)sh + o(h)$$

and

$$v-\phi(\phi^{-1}(v)-sh) = \phi(\phi^{-1}(v)) - \phi(\phi^{-1}(v)-th) = \phi'(v)th + o(h).$$

Thus

$$\begin{aligned} \mathbb{E}g^2(U,V,h) &\leq \int_{-1}^{1}\int_{-1}^{1} [\phi'(u)C_u(u,v)h + \phi'(v)C_v(u,v)h]\, k(s,t)dsdt \\ &\leq 4h\, [\|C_u\| + \|C_v\|] \sup_{x\in\mathbb{R}} |\phi'(x)| \|k\|. \end{aligned}$$

Taking $C' = 4\,[\|C_u\| + \|C_v\|]\,\|\phi'\|\,\|k\|$ gives condition (G.ii).

Case 2. $u \wedge \phi(\phi^{-1}(u)-sh) = u$ and $v \wedge \phi(\phi^{-1}(v)-th) = v$.
Here, inequality (20) is reduced to

$$\mathbb{E}g^2(U,V,h) \leq \int_{-1}^{1}\int_{-1}^{1} \left[C\left(\phi(\phi^{-1}(u)-sh), \phi(\phi^{-1}(v)-th)\right) - C(u,v)\right] k(s,t)dsdt + o(1).$$

Using the same arguments as in Case 1, we obtain condition (G.ii),with $C' = 4\,[\|C_u\| + \|C_v\|]\,\|\phi'\|\,\|k\|$.

Case 3. $u \wedge \phi(\phi^{-1}(u)-sh) = \phi(\phi^{-1}(u)-sh)$ and $v \wedge \phi(\phi^{-1}(v)-th) = v$.
Here, inequality (20) is rewritten into

$$\begin{aligned} &\mathbb{E}g^2(U,V,h) \leq \textstyle\int_{-1}^{1}\int_{-1}^{1} C\left(\phi(\phi^{-1}(u)-sh), \phi(\phi^{-1}(v)-th)\right) k(s,t)dsdt \\ &-2\textstyle\int_{-1}^{1}\int_{-1}^{1} C\left(\phi(\phi^{-1}(u)-sh), v\right) k(s,t)dsdt + \int_{-1}^{1}\int_{-1}^{1} C(u,v)k(s,t)dsdt + o(1) \\ &\leq \textstyle\int_{-1}^{1}\int_{-1}^{1} \left[C\left(\phi(\phi^{-1}(u)-sh), \phi(\phi^{-1}(v)-th)\right) - C\left(\phi(\phi^{-1}(u)-sh), v\right)\right] k(s,t)dsdt \\ &- \textstyle\int_{-1}^{1}\int_{-1}^{1} \left[C\left(\phi(\phi^{-1}(u)-sh), v\right) - C(u,v)\right] k(s,t)dsdt + o(1). \end{aligned}$$

By applying successively a Taylor expansion for C and for ϕ, we get

$$\begin{aligned} \mathbb{E}g^2(U,V,h) &\leq \int_{-1}^{1}\int_{-1}^{1} C_v\left(\phi(\phi^{-1}(u)-sh), \theta_1\right) \left[\phi(\phi^{-1}(v)-th) - \phi(\phi^{-1}(v))\right] k(s,t)dsdt \\ &\quad - \int_{-1}^{1}\int_{-1}^{1} C_u\left(\theta_2, v\right) \left[\phi(\phi^{-1}(u)-sh) - \phi(\phi^{-1}(u))\right] k(s,t)dsdt + o(1) \\ &\leq \int_{-1}^{1}\int_{-1}^{1} C_v\left(\phi(\phi^{-1}(u)-sh), \theta_1\right) \phi'(\gamma_1).(-th)k(s,t)dsdt \\ &\quad - \int_{-1}^{1}\int_{-1}^{1} C_u\left(\theta_2, v\right) \phi'(\gamma_2).(-sh)k(s,t)dsdt + o(1), \end{aligned}$$

where $\theta_1 \in \left(\phi(\phi^{-1}(v)-th),v\right)$; $\theta_2 \in \left(\phi(\phi^{-1}(u)-sh),u\right)$; $\gamma_1 \in \left(\phi^{-1}(v)-th,\phi^{-1}(v)\right)$; $\gamma_2 \in \left(\phi^{-1}(u)-sh,\phi^{-1}(u)\right)$.
This implies

$$\begin{aligned}\mathbb{E}g^2(U,V,h) &\leq 4h\left\|C_v\right\|\left\|\phi'\right\|\left\|k\right\|^2|t|+4h\left\|C_u\right\|\left\|\phi'\right\|\left\|k\right\||s|\\ &\leq 4h\left\|\phi'\right\|\left\|k\right\|\left(\left\|C_v\right\|+\left\|C_u\right\|\right).\end{aligned}$$

Thus condition (G.ii) holds, with $C'=4\left\|\phi'\right\|\left\|k\right\|\left(\left\|C_v\right\|+\left\|C_u\right\|\right)$.

Case 4. $u\wedge\phi(\phi^{-1}(u)-sh)=u$ and $v\wedge\phi(\phi^{-1}(v)-th)=\phi(\phi^{-1}(v)-th)$.
This case is analogous to Case 3, where the roles of u and v are interchanged. Hence, condition (G.ii) is fulfilled, with the same constant $C'=4\left\|\phi'\right\|\left\|k\right\|\left(\left\|C_v\right\|+\left\|C_u\right\|\right)$.

Checking for (F.i): We have to check the uniform entropy condition for the class $\mathcal{G}_n$ defined in (6).
To this end, we introduce the following classes of functions, where φ is a fixed non-decreasing function :

$$\mathbb{F}=\left\{x\mapsto\frac{m-\varphi(x)}{h},0<h<1,\ m\in\mathbb{R}\right\}$$
$$\mathbb{K}_0=\left\{x\mapsto K\left(\frac{m-\varphi(x)}{h}\right),\ 0<h<1,\ m\in\mathbb{R}\right\}$$
$$\mathbb{K}=\left\{(x,y)\mapsto K\left(\frac{m-\varphi(x)}{h}\right)K\left(\frac{m-\varphi(y)}{h}\right),\ 0<h<1,\ m\in\mathbb{R}\right\}$$
$$\mathbb{H}=\left\{(x,y)\mapsto K\left(\frac{m-\varphi(x)}{h}\right)K\left(\frac{m-\varphi(y)}{h}\right)-\mathbb{I}\left\{x\leq u,y\leq v\right\}\ ;0<h<1,\ m\in\mathbb{R},(u,v)\in[0,1]^2\right\}.$$

It is clear that by applying successively Lemmas 2.6.15 and 2.6.18 in van der Vaart and Wellner (1996), pages 146-147, the sets of functions $\mathbb{F}$ and $\mathbb{K}_0$ are VC-subgraph classes. Theorem 2.6.7 of the same reference implies that the class $\mathbb{K}_0$ admits a uniform polynomial covering number. Then applying Lemma A.1 in Einmahl and Mason (2000), we can infer that $\mathbb{K}$ is a class of functions with a uniform polynomial covering number too. It follows from this that the class $\mathbb{H}$ admits a uniform polynomial covering number. Since $\mathbb{H}$ and $\mathcal{G}_n$ have the same structure, we conclude that $\mathcal{G}_n$ satisfies this property too ; i.e. $\mathcal{G}_n$ fulfills condition (F.i).

Checking for (F.ii):
Define the class of functions

$$\mathcal{G}_0=\left\{\ (s,t)\mapsto K\left(\tfrac{\phi^{-1}(u)-\phi^{-1}(\zeta_{1,n}(s))}{h}\right)K\left(\tfrac{\phi^{-1}(v)-\phi^{-1}(\zeta_{2,n}(t))}{h}\right)-\mathbb{I}\{s\leq u,t\leq v\},\ u,v\in[0,1]\cap\mathbb{Q},0<h<1\ \right\}$$

It's clear that $\mathcal{G}_0$ is countable and $\mathcal{G}_0\subset\mathcal{G}_n$. Let $g\in\mathcal{G}_n$,

$$g(s,t)=K\left(\frac{\phi^{-1}(u)-\phi^{-1}(\zeta_{1,n}(s))}{h}\right)K\left(\frac{\phi^{-1}(v)-\phi^{-1}(\zeta_{2,n}(t))}{h}\right)-\mathbb{I}\{s\leq u,s\leq v\},\qquad\forall(s,t)\in[0,1]^2$$

and for $m\geq 1$,

$$g_m(s,t)=K\left(\frac{\phi^{-1}(u_m)-\phi^{-1}(\zeta_{1,n}(s))}{h}\right)K\left(\frac{\phi^{-1}(v_m)-\phi^{-1}(\zeta_{2,n}(y))}{h}\right)-\mathbb{I}\{s\leq u_m,t\leq v_m\},$$

where $u_m = \frac{1}{m^2}[m^2u] + \frac{1}{m^2}$ and $v_m = \frac{1}{m^2}[m^2v] + \frac{1}{m^2}$.

Let $\alpha_m = u_m - u, \quad \beta_m = v_m - v$. Then, we have $0 < \alpha_m \leq \dfrac{1}{m^2}$ and $0 < \beta_m \leq \dfrac{1}{m^2}$. Hence $u_m \searrow u$ and $v_m \searrow v$. By continuity, $\phi^{-1}(u_m) \searrow \phi^{-1}(u)$ and $\phi^{-1}(v_m) \searrow \phi^{-1}(v)$. Define

$$\delta_{m,u} = \left(\frac{\phi^{-1}(u_m) - \phi^{-1}(\zeta_{1,n}(s))}{h}\right) - \left(\frac{\phi^{-1}(u) - \phi^{-1}(\zeta_{1,n}(s))}{h}\right) = \frac{\phi^{-1}(u_m) - \phi^{-1}(u)}{h}$$

and

$$\delta_{m,v} = \left(\frac{\phi^{-1}(v_m) - \phi^{-1}(\zeta_{2,n}(t))}{h}\right) - \left(\frac{\phi^{-1}(v) - \phi^{-1}(\zeta_{2,n}(t))}{h}\right) = \frac{\phi^{-1}(v_m) - \phi^{-1}(v)}{h}.$$

Then $\delta_{m,u} \searrow 0$ and $\delta_{m,v} \searrow 0$, which are equivalent to

$$\left(\frac{\phi^{-1}(u_m) - \phi^{-1}(\zeta_{1,n}(s))}{h}\right) \searrow \left(\frac{\phi^{-1}(u) - \phi^{-1}(\zeta_{1,n}(s))}{h}\right)$$

and

$$\left(\frac{\phi^{-1}(v_m) - \phi^{-1}(\zeta_{2,n}(t))}{h}\right) \searrow \left(\frac{\phi^{-1}(v) - \phi^{-1}(\zeta_{2,n}(t))}{h}\right).$$

By right-continuity of the kernel $K(\cdot)$, we obtain for all $(s,t) \in [0,1]^2$,

$$g_m(s,t) \longrightarrow g(s,t), \quad m \to \infty$$

and conclude that $\mathcal{G}_n$ is pointwise measurable class.

References

Altman, N. and Léger, C. (1995). Bandwidth selection for Kernel distribution function estimation. *Journal of Statistical Planning and Inference*, 46 (2), 195-214.

Bâ, D. Seck, C.T. and Lô, G.S. (2015). Asymptotic Confidence Bands for Copulas Based on the Local Linear Kernel Estimator. *Applied Mathematics*, 6, 2077-2095.

Bouzebda, S. (2012). On the Strong Approximation of Bootstrapped Empirical Copula Processes with Applications. *Mathematical Methods of Statistics*, Vol. 21, No. 3, 153-188.

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

Chung, K-L. (1949). An estimate concerning the Kolmogoroff limit distribution. *Trans Am Math Soc*, 67, 36-50.

Deheuvels, P. (2009). A Multivariate Bahadur-Kiefer Representation for the Empirical Copula Process. *Journal of Mathematical Sciences*, Vol.163, No. 4, 382-398.

Deheuvels, P. and Mason, D. M. (2004). General Asymptotic Confidence Bands Based on Kernel-type Function Estimators. *Statistical Inference stochastic process*, 7, 225-277.

Einmahl, U. and Mason, D. M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators, *J. Theoret. Prob*, 13, 1-37.

Fermanian, J.D. and Scaillet, O. (2003). Nonparametric Estimation of Copulas for Time Series. *Journal of risk*, 5, 25-54.

Fermanian, J. Radulovic, D. and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli* 10(5), 847-860.

Geenens, G. Charpentier, A. and Paindaveine, D. (2014). Probit transformation for non-parametric kernel estimation of the copula density. *ECARES working paper 2014-23*.

Genest, C. Ghoudi, K. and Rivest, L. P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82, 543-552.

Gijbels, J. and Mielniczuk, J. (1990). Estimation of the density of a copula function. *Communications in Statistics*, Ser. A , Vol.19, 445-464.

Lo, G.S. (2016). *Weak Convergence (IA). Sequences of random vectors.* SPAS Books Series. Saint-Louis, Senegal - Calgary, Canada. doi: 10.16929/sbs/2016.0001. Arxiv: 1610.05415. ISBN : 978-2-9559183-1-9.

Lo, G.S.(2018). A simple proof of the theorem of Sklar and its extension to distribution functions. Arxiv:1803.00409

Marron, J. S. and Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(4), 653-671.

Mason, D. M. and Swanepoel, J. W. (2011). A general result on the uniform in bandwidth consistency of kernel-type function estimators. *Test,* 20(1), 72-94.

Nelsen, R. B. (1999). *An Introduction to Copulas.* New York, Springer-Verlag.

Omelka, M. Gijbels, I. and Veraverbeke, N. (2009). Improved kernel estimators of copulas : Weak convergence and goodness-of-fit testing. *The Annals of Statistics*, Vol. 37, 5B, 3023-3058.

Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society*, Series B (44), 414-422.

Sarda, P. (1993). Smoothing parameter selection for smooth distribution functions. *Journal of Statistical Planning and Inference*, 35, 65-75.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statistic. Univ. Paris*, 8, 229-231.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* New-York, Springer.

Wichura, M. J. (1973). Some Strassen-type laws of the iterated logarithm for multi-parameter stochastic processes with independent increments. *The Annals of Probability*, 1, 272-296.

Zari, T. (2010). *Contribution à l'étude du processus empirique de copule.* Thèse de doctorat, Université de Paris 6, France.