Hacettepe Journal of
Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 52 (5) (2023), 1303 – 1348
DOI : 10.15672/hujms.1134334

Research Article

On the weak convergence and the uniform-in-bandwidth consistency of the general conditional U-processes based on the copula representation: multivariate setting

Salim Bouzebda*

Laboratoire de Mathématiques Appliquées de Compiègne Université de Technologie de Compiégne

Abstract

U-statistics represent a fundamental class of statistics from modeling quantities of interest defined by multi-subject responses. U-statistics generalise the empirical mean of a random variable X to sums over every m-tuple of distinct observations of X. Stute [Conditional U -statistics, Ann. Probab., 1991] introduced a class of estimators called conditional U-statistics. In the present work, we provide a new class of estimators of conditional U-statistics. More precisely, we investigate the conditional U-statistics based on copula representation. We establish the uniform-in-bandwidth consistency for the proposed estimator. In addition, uniform consistency is also established over $\varphi \in \mathscr{F}$ for a suitably restricted class $\mathscr{F}$, in both cases bounded and unbounded, satisfying some moment conditions. Our theorems allow data-driven local bandwidths for these statistics. Moreover, in the same context, we show the uniform bandwidth consistency for the nonparametric Inverse Probability of Censoring Weighted estimators of the regression function under random censorship, which is of its own interest. We also consider the weak convergence of the conditional U-statistics processes. We discuss the wild bootstrap of the conditional U-statistics processes. These results are proved under some standard structural conditions on the Vapnik-Chervonenkis class of functions and some mild conditions on the model.

Mathematics Subject Classification (2020). 62G07, 60F05, 60F15, 62E20, 62G30.

Keywords. Conditional U-statistics, consistency, data-dependent bandwidth selection, empirical process, kernel estimation, Nadaraya-Watson, regression, copula function, uniform in bandwidth, weak convergence, wild bootstrap

1. Introduction

Nonparametric density and regression function estimation has been the subject of intense investigation for many years, leading to the development of many methods. For good sources of references to the research literature in this area, along with statistical applications, consult [41, 55, 60, 78, 120, 128, 144] and the references therein. In the last decades, empirical process theory has provided very useful and powerful tools to analyze the large

*Corresponding Author.
Email addresses: salim.bouzebda@utc.fr (S.Bouzebda)
Received: 22.06.2022; Accepted: 28.02.2023

sample properties of several nonparametric estimators of functionals of the distribution, such as the regression function and the density function, refer to [57, 58, 97, 114, 142]. Nolan and Pollard [110] were the first to introduce the notion of uniform in bandwidth consistency for kernel density estimators and they applied empirical process methods in their study. In the series of papers, [11, 18, 19, 20, 21, 23, 28, 30, 45, 52, 53, 56, 61, 62, 105], the authors established uniform consistency results for such estimators, where h_n varies within suitably chosen intervals indexed by n. U-statistics, first considered by [81] in connection with unbiased statistics, and formally introduced by [84]. The theory of U-statistics and U-processes has received considerable attention in the last decades due to its great number of applications and usefulness for solving complex statistical problems. Examples are density estimation, nonparametric regression tests and goodness-of-fit tests. More precisely, U-processes appear in statistics in many instances, e.g., as the components of higher-order terms in von Mises expansions. In particular, U-statistics play a role in analyzing estimators (including function estimators) with varying degrees of smoothness. For example, Stute [136] applies the a.s. uniform bounds for $\mathbb{P}$-canonical U-processes to analyze the product limit estimator for truncated data. Arcones and Wang [5] present two new tests for normality based on U-processes. Making use of the results of [75, 76], Schick et al. [119] introduced new tests for normality, which are based on the weighted L_1-distances between the standard normal density and local U-statistics based on standardized observations. Joly and Lugosi [91] discussed the estimation of the mean of the multivariate functions in the case of possibly heavy-tailed distributions and introduced the median-of-means, which is based on U-statistics. U-processes are important tools for a broad range of statistical applications such as testing for qualitative features of functions in nonparametric statistics [1, 72, 100], cross-validation for density estimation [110], and establishing limiting distributions of M-estimators (see, e.g., [4, 50, 125, 126]). Infinite-order U-statistics are useful tools for constructing simultaneous prediction intervals that quantify the uncertainty of ensemble methods such as subbagging and random forests. Peng et al. [113] develop in great detail the notion of generalized U-statistics random forest predictions. The MeanNN approach estimation for differential entropy introduced by [63] is a particular of the U- statistic. Using U-statistics, [102] proposed a new test statistic for goodness-of-fit tests. Halmos [81], Hoeffding [84] and vonMises [143] provided (amongst others) the first asymptotic results for the case that the underlying random variables are independent and identically distributed. This paper uses the copula representation to consider the so-called conditional U-statistics introduced by [135]. These statistics may be viewed as generalizations of the Nadaraya-Watson [107, 145] estimates of a regression function. To better understand the problem, we first introduce Stute's estimators.

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent random vectors with common joint density function $f_{X,Y} : \mathbb{R}^p \times \mathbb{R}^q \to [0, \infty[$. Let $\varphi : \mathbb{R}^{mq} \to \mathbb{R}$ be a measurable function. In this paper, we are primarily concerned with the estimation of the conditional expectation or regression function

$$r^{(m)}(\varphi, \mathbf{t}) = \mathbb{E}(\varphi(Y_1, \ldots, Y_m) \mid (X_1, \ldots, X_m) = \mathbf{t}), \text{ for } \mathbf{t} \in \mathbb{R}^{mp}, \tag{1.1}$$

whenever it exists, i.e., $\mathbb{E}(|\varphi(Y_1, \ldots, Y_m)|) < \infty$. We now introduce a kernel function $K : \mathbb{R}^p \to \mathbb{R}$ with support contained in $[-B, B]^p$, $B > 0$ satisfying

$$\sup_{x \in \mathbb{R}^p} |K(x)| =: \kappa < \infty \ \text{ and } \ \int_{\mathbb{R}^p} K(x)dx = 1 \tag{K.i}$$

introduced [135] and a class of estimators for $r^{(m)}(\varphi, \mathbf{t})$, called conditional U-statistics, which is defined for each $\mathbf{t} \in \mathbb{R}^{mp}$ to be

$$\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n) = \frac{\displaystyle\sum_{(i_1,\ldots,i_m)\in I(m,n)} \varphi(Y_{i_1}, \ldots, Y_{i_m}) K\left(\frac{t_1 - X_{i_1}}{h_n}\right) \cdots K\left(\frac{t_m - X_{i_m}}{h_n}\right)}{\displaystyle\sum_{(i_1,\ldots,i_m)\in I(m,n)} K\left(\frac{t_1 - X_{i_1}}{h_n}\right) \cdots K\left(\frac{t_m - X_{i_m}}{h_n}\right)}, \quad (1.2)$$

where

$$I(m,n) = \{\mathbf{i} = (i_1, \ldots, i_m) : 1 \le i_j \le n \quad \text{and} \quad i_j \ne i_r \quad \text{if} \ \ j \ne r\},$$

is the set of all m-tuples of different integers between 1 and n and $0 < h_n < 1$ goes to zero at a certain rate. Notice that $\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$ generalize the Nadaraya-Watson estimate of a regression function, [107, 145]. Indeed, the particular case $m = 1$, the $r^{(m)}(\varphi, t)$ is reduced to $r^{(1)}(\varphi, t) = \mathbb{E}(\varphi(Y)|X = t)$ and Stute's estimator becomes the Nadaraya-Watson estimator of $r^{(1)}(\varphi, t)$ given by :

$$\widehat{r}_n^{(1)}(\varphi, t, h_n) = \sum_{i=1}^{n} \varphi(Y_i) K\left(\frac{X_i - t}{h_n}\right) \Big/ \sum_{i=1}^{n} K\left(\frac{X_i - t}{h_n}\right). \quad (1.3)$$

The work of [122] was devoted to estimating the rate of convergence of $\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$ to $r^{(m)}(\varphi, \mathbf{t})$. In the paper of [115], the limit distributions of $\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$ are discussed and compared with those obtained by Stute. Harel and Puri [82] extended the results of [135], under appropriate mixing conditions, to weakly dependent data. Stute [139] proposed symmetrized nearest neighbour conditional U-statistics as alternatives to the usual kernel-type estimators. Sen [122] obtained results on the uniform in $\mathbf{t}$ consistency of $\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$. An important contribution is given in the paper [56] where a much stronger form of consistency holds, namely, uniform in $\mathbf{t}$ and in bandwidth consistency (i.e., h_n, $h_n \in [a_n, b_n]$ where $a_n < b_n \to 0$ at some specific rate) of $\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$. In addition, uniform consistency is also established over $\varphi \in \mathscr{F}$ for a suitably restricted class $\mathscr{F}$, for recent references see [31, 32, 34, 35, 36, 37, 132]. The main tool in their result is the use of the local conditional U process investigated in [75]. For excellent resource of references on the U-statistics and U-processes the interested reader may refer to [4, 50, 99]. For the U-statistics with random kernels of diverging orders we refer to [67, 83, 131, 133]. Infinite-order U-statistics are useful tools for constructing simultaneous prediction intervals that quantify the uncertainty of ensemble methods such as subbagging and random forests.

Consider a random vector (X, Y) with joint cumulative distribution function [df] $\mathbb{F}(\cdot)$ and the corresponding marginal df's $\mathbb{F}_i(x_i) := \mathbb{P}(X_i \le x_i)$, $i = 1, \ldots, p$ and $\mathbb{F}_{0,j}(y_j) = \mathbb{P}(Y_j \le y_j)$, $j = 1, \ldots, q$ are continuous. The characterization theorem of [129] implies that there exists a copula function $\mathbb{C}(\cdot, \cdot)$, such that,

$$\mathbb{F}(x_1, \ldots, x_p, y_1, \ldots, y_q) = \mathbb{C}(\mathbf{F}(x), \mathbf{F}_0(y)), \quad \text{for all} \quad (x, y) \in \mathbb{R}^{p\times q}, \quad (1.4)$$

where $\mathbf{F}(x) = (F_1(x_1), \ldots, F_p(x_p))$ and $\mathbf{F}_0(y) = (F_{0,1}(y_1), \ldots, F_{0,q}(y_q))$. By definition, the copula function $\mathbb{C}(\cdot, \cdot)$ is a $p+q$-variate cumulative distribution function, on the unit cube $[0, 1]^{p+q}$, the margins of which are standard uniform distributions on the interval $[0, 1]$. If not stated otherwise, we assume that the $F_i(\cdot)$, $i = 1, \ldots, p$ and $F_{0,j}(\cdot)$, $j = 1, \ldots, q$ are continuous functions, in this case, the copula function $\mathbb{C}(\cdot)$ is unique. In the monographs by [89, 90, 108] the reader may find detailed ingredients of the modeling theory as well as surveys of the commonly used copulas; we also refer to [43, 59, 124]. For in-depth and overview historical notes, we refer to [49]. We can refer also to [130], where the author sketches the proof of (1.4), develops some of its consequences, and surveys some of the work on copulas. Copulas are a flexible and versatile tool for analyzing dependency structures. More specifically, copula $\mathbb{C}(\cdot)$ "couples" the joint distribution function $\mathbb{F}(\cdot)$ to its univariate marginals, capturing as such the dependence structure between the components of (X, Y). Indeed, most conventional dependence measures can be explicitly expressed in terms of the copula. This feature has motivated successful applications in actuarial science and survival analysis (see, e.g., [68]). In the literature on risk management and, more generally, in mathematical economics and mathematical finance modeling, a number of illustrations are provided (refer to books of [44] and [106]), in particular, in the context of asset pricing and credit risk management. Notice that we have the following representation

$$\begin{aligned} r^{(m)}(\varphi, \mathbf{t}) &= \mathbb{E}(\varphi(Y_1, \ldots, Y_m) \mid (X_1, \ldots, X_m) = \mathbf{t}) \\ &= \int_{\mathbb{R}^{mq}} \varphi(y_1, \ldots, y_m) \left\{ \frac{\prod_{j=1}^{m} f(t_j, y_j)}{\prod_{j=1}^{m} f_X(t_j)} \right\} dy_1 \ldots dy_m, \quad \text{for } \mathbf{t} \in \mathbb{R}^{mp}. \end{aligned} \tag{1.5}$$

Let

$$\mathbf{c}(u_1, \ldots, u_p, v_1, \ldots, v_q) = \frac{\partial^{p+q}}{\partial u_1 \cdots \partial u_p \partial v_1 \cdots \partial v_q} \mathbb{C}(u_1, \ldots, u_p, v_1, \ldots, v_q)$$

be the copula density, i.e., the density of $W = (\mathbf{F}(X), \mathbf{F}_0(Y))$, that we assume to exist. The copula density of $\mathbf{F}(X)$, is given by

$$\check{\mathbf{c}}(u_1, \ldots, u_p) = \int_{[0,1]^q} \mathbf{c}(u_1, \ldots, u_p, v_1, \ldots, v_q) dv_1 \ldots dv_q.$$

Then, we have the following representations

$$\begin{aligned} f_{X,Y}(x, y) &= f_{X_1}(x_1) \ldots f_{X_p}(x_p) f_{Y_1}(y_1) \ldots f_{Y_q}(y_q) \mathbf{c}(\mathbf{F}(x), \mathbf{F}_0(y)), \\ \mathbf{g}_{Y|X}(y \mid x) &= \frac{f_{X_1}(x_1) \ldots f_{X_p}(x_p) f_{Y_1}(y_1) \ldots f_{Y_q}(y_q) \mathbf{c}(\mathbf{F}(x), \mathbf{F}_0(y))}{g_X(x)} \\ &= f_{Y_1}(y_1) \ldots f_{Y_q}(y_q) \frac{\mathbf{c}(\mathbf{F}(x), \mathbf{F}_0(y))}{\check{\mathbf{c}}(\mathbf{F}(x))}, \end{aligned} \tag{1.6}$$

where $\mathbf{g}_{Y|X}$ denotes the conditional density of Y given $X = x$, $g_X(\cdot)$ and $g_Y(\cdot)$ are the joint densities of X and Y respectively $f_{XY}(\cdot)$ is the joint density of (X, Y). Making use of relation (1.5) and (1.6), we have the following identity, for $\mathbf{t} \in \mathbb{R}^{mp}$,

$$\begin{aligned} & r^{(m)}(\varphi, \mathbf{t}) = \mathbb{E}(\varphi(Y_1, \ldots, Y_m) \mid (X_1, \ldots, X_m) = \mathbf{t}) \\ &= \int_{\mathbb{R}^{mq}} \varphi(y_1, \ldots, y_m) \left\{ \prod_{j=1}^{m} \mathbf{g}_{Y|X}(y_j \mid t_j) \right\} dy_1 \ldots dy_m, \\ &= \int_{\mathbb{R}^{mq}} \varphi(y_1, \ldots, y_m) \left\{ \prod_{j=1}^{m} f_{Y_1}(y_{1,j}) \ldots f_{Y_q}(y_{q,j}) \frac{\mathbf{c}(\mathbf{F}(t_j), \mathbf{F}_0(y_j))}{\check{\mathbf{c}}(\mathbf{F}(t_j))} \right\} dy_1 \ldots dy_m. \end{aligned} \tag{1.7}$$

In the particular cases of a single covariate ($p = 1$) or mutually independent predictors, $\check{\mathbf{c}}(\mathbf{F}(x)) = 1$, we have

$$r^{(m)}(\varphi, \mathbf{t}) = \int_{\mathbb{R}^{mq}} \varphi(y_1, \ldots, y_m) \left\{ \prod_{j=1}^{m} f_{Y_1}(y_{1,j}) \ldots f_{Y_q}(y_{q,j}) \mathbf{c}(\mathbf{F}(t_j), \mathbf{F}_0(y_j)) \right\} dy_1 \ldots dy_m.$$

To illustrate the idea, we briefly cite two examples for $m = p = q = 1$. If the copula density of (Y, X) belongs to the Farlie-Gumbel-Morgenstern (FGM) family with a parameter θ, then we have

$$\begin{aligned} r^{(1)}(Id, t) &= \mathbb{E}(Y \mid X = t) \\ &= \mathbb{E}(Y) + \theta(2\mathbb{G}_X(t) - 1) \int \mathbb{G}_Y(y)(1 - \mathbb{G}_Y(y)) dy, \end{aligned}$$

where Id denotes the identity function. In this example $m = q = 1$, $p \geq 1$. If the copula of $(Y, \mathbf{X}^\top)^\top$ is Gaussian with correlation matrix

$$\Sigma_{Y,\mathbf{X}} = \begin{bmatrix} 1 & \boldsymbol{\rho}^\top \\ \boldsymbol{\rho} & \Sigma_{\mathbf{X}} \end{bmatrix},$$

then we have

$$\begin{aligned} r^{(1)}(Id, \mathbf{t}) &= \mathbb{E}(Y \mid \mathbf{X} = \mathbf{t}) \\ &= \mathbb{E}\left(\mathbb{G}_Y^{-1}\left(\Phi\left(\mathbf{u}^\top \Sigma_{\mathbf{X}}^{-1}\boldsymbol{\rho} + \sqrt{1 - \boldsymbol{\rho}^\top \Sigma_{\mathbf{X}}^{-1}\boldsymbol{\rho}}Z\right)\right)\right), \end{aligned}$$

where $\mathbf{u} = (\Phi^{-1}(F_1(x_1), \ldots, \Phi^{-1}(F_d(x_d))^\top$, $Z \sim N(0, 1)$ and $\Phi(\cdot)$ is the cdf of standard normal distribution, we may refer to [101], we refer also to [38, 54, 71, 95, 109, 146]. In the present work, we consider the regression-based copula representation given in (1.7). We propose a new estimation methodology for the regression function $r^{(m)}(\varphi, \mathbf{t})$. The present work largely extends and completes the work [22] to the multivariate setting in several ways. In addition, we consider new results like the weak convergence and the censored setting. More precisely, we will consider one of the most commonly used classes of estimators formed by the so-called kernel-type estimators based on the copula regression representation. There are no restrictions on the choice of the kernel function in our setup, apart from satisfying some mild conditions that we will give after. The selection of the bandwidth, however, is more problematic. It is worth noting that the choice of bandwidth is crucial to obtain a good rate of consistency; for example, it greatly influences the size of the estimate's bias. In general, we are interested in the selection of bandwidth that produces an estimator which has a good balance between the bias and the variance of the considered estimators. It is more appropriate to consider the bandwidth varying according to the criteria applied and to the available data and location, which cannot be achieved by using classical methods. The interested reader may refer to [104] for more details and discussion on the subject. In the present paper, we develop methods that permit the study of kernel-type copula regression estimators using data or location-dependent bandwidth sequences. To the best of our knowledge, the problems we investigate in the present paper form an unsolved open problem in the literature, and it gives the main motivation for our paper.

The layout of the article is as follows. The focus of Section 2 is on introducing the new methodology of estimating U-statistics based on the copula representation. We establish general uniform in bandwidth consistency results for kernel copula estimators in Theorems 2.2 and 2.3. In Section 3, the weak convergence of the U-processes are studied. Section 4 is devoted to the wild bootstrap of the U-proposed investigated in Section 3. In Section 6, we provide some potential applications: Simultaneous prediction intervals for random forests in 6.1, Discrimination in 6.2, Ranking problems in 6.3. We provide some examples of classes of functions together with conditional U-statistics in Section 7. In Section 8, we present how to select the bandwidth through the cross-validation procedures. This article concludes with a brief discussion in Section 9. All mathematical developments are relegated to Section 10. A few relevant technical results are given in the appendix.

2. Estimation

Recall the idea of [23]. To construct estimate of $r^{(1)}(\varphi, t)$, we estimate copula density $\mathbf{c}(\cdot, \cdot)$ by the usual kernel density estimator and estimate integration with respect to $f_Y(\cdot)$ by integration with respect to the empirical measure, then delete the diagonal terms. Notice a similar idea was introduced in [80] in estimating the integral of a squared probability density (derivative). As mentioned in the last paper, this avoids the addition of a type of bias in the estimator. A similar idea was used in estimating other density functions like the extropy; refer to [98] for definition. Jansen et al. [88] investigate the performance of

Bernstein estimators for the copula partial derivatives; one can also refer to [140]. Using the estimators of the copula first-order partial derivatives, Jansen et al. [88] also construct estimators for the conditional distribution function and its important functional of it, like the mean regression function and the quantile regression function. Let us introduce the estimators

$$\begin{aligned}
F_n(x) &:= \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{X_i \leq x\} \quad \text{for,} \quad x \in \mathbb{R}^p,\\
F_{n,j}(x_j) &:= \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{X_{ij} \leq x_j\} \quad \text{for,} \quad x_j \in \mathbb{R}, j=1,\ldots,p,\\
F_{0;n}(y) &:= \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{Y_i \leq y\} \quad \text{for} \quad y \in \mathbb{R}^q,\\
F_{0;n,j}(y_j) &:= \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{Y_{ij} \leq y_j\} \quad \text{for} \quad y_j \in \mathbb{R}, j=1,\ldots,q,\\
&\widehat{\mathbf{c}}^{j}_{n-1;h_n}\left(\mathbf{F}_n(x), \mathbf{F}_{0,n}(y)\right)\\
&:= \frac{1}{(n-1)h_n^{p+q}}\sum_{\substack{i=1\\ i\neq j}}^{n} K\left(\frac{\mathbf{F}_n(x)-\mathbf{F}_n(X_i)}{h_n}\right) K_0\left(\frac{\mathbf{F}_{0,n}(y)-\mathbf{F}_{0,n}(Y_i)}{h_n}\right), \quad (2.1)
\end{aligned}$$

where

$$\mathbf{F}_n(x) = (F_{n,1}(x_1),\ldots,F_{n,p}(x_p)), \mathbf{F}_{0,n}(y) = (F_{0,n,1}(y_1),\ldots,F_{0,n,q}(y_q)),$$

and the kernel function $K_0 : \mathbb{R}^p \to \mathbb{R}$ with support contained in $[-B_0, B_0]$, $B_0 > 0$ satisfying

$$\sup_{x\in\mathbb{R}^q} |K_0(x)| =: \kappa < \infty \quad \text{and} \quad \int_{\mathbb{R}^q} K(x)dx = 1 \tag{K.i.0}$$

The kernel estimation of copula is a very rich topic of research; we only mention some recent references [11, 32, 42, 111], see their lists of references for related studies.

Remark 2.1. For notational convenience, we have chosen the same bandwidth sequence for each margin. This assumption can be dropped easily. If one wants to use the vector bandwidths (see, in particular, Chapter 12 of [55]). With obvious changes of notation, our results and their proofs remain true when h_n is replaced by a vector bandwidth $\mathbf{h}_n = (h_n^{(1)},\ldots,h_n^{(p)})$, where $\min h_n^{(i)} > 0$. In this situation we set $h_n = \prod_{i=1}^{p} h_n^{(i)}$, and for any vector $\mathbf{v} = (v_1,\ldots,v_p)$ we replace $\mathbf{v}/h_n$ by $(v_1/h_n^{(1)},\ldots,v_p/h_n^{(p)})$.

The analogous estimators based on copula representations of (1.3) is given, for $p = q = 1$, by

$$\begin{aligned}
\widetilde{m}_{n;h_n}(x,\varphi) &:= \frac{1}{n(n-1)h_n^2}\sum_{1\leq i\neq j\leq n}\varphi(Y_i)\\
&\times K\left(\frac{\widehat{\mathbb{G}}_{Y;n}(Y_j)-\widehat{\mathbb{G}}_{Y;n}(Y_i)}{h_n}\right) K\left(\frac{\widehat{G}_{X;n}(x)-\widehat{\mathbb{G}}_{X;n}(X_j)}{h_n}\right).
\end{aligned}\tag{2.2}$$

By setting $\varphi(y) = y$ into (2.2) we get the copula kernel regression function estimator of $m(x) := \mathbb{E}(Y \mid X = x)$ given by

$$\begin{aligned}
\widetilde{m}_{n;h_n}(x) &:= \frac{1}{n(n-1)h_n^2}\sum_{1\leq i\neq j\leq n} Y_i\\
&\times K\left(\frac{\widehat{\mathbb{G}}_{Y;n}(Y_j)-\widehat{\mathbb{G}}_{Y;n}(Y_i)}{h_n}\right) K\left(\frac{\widehat{G}_{X;n}(x)-\widehat{\mathbb{G}}_{X;n}(X_j)}{h_n}\right),
\end{aligned}\tag{2.3}$$

By setting $\varphi_t(y) = \mathbb{1}\{y \le t\}$, $t \in \mathbb{R}$, into (2.2) we obtain the kernel estimator of the conditional distribution function $F(t|x) := \mathbb{P}(Y \le t|X = x)$ given by

$$\widetilde{F}_{n;h_n}(t|x) := \frac{1}{n(n-1)h_n^2} \sum_{1\le i\ne j\le n} \mathbb{1}\{Y_i \le t\} \times K\left(\frac{\widehat{\mathbb{G}}_{Y;n}(Y_j) - \widehat{\mathbb{G}}_{Y;n}(Y_i)}{h_n}\right) K\left(\frac{\widehat{G}_{X;n}(x) - \widehat{\mathbb{G}}_{X;n}(X_j)}{h_n}\right). \tag{2.4}$$

Let us introduce the weight, $(i_1, \ldots, i_m), (\ell_1, \ldots, \ell_m) \in I(m, n)$,

$$\mathbf{W}(\mathbf{i}, \boldsymbol{\ell}, \mathbf{t}; h_n) = \frac{\frac{1}{h_n^{m(p+q)}} \prod_{\nu=1}^{m} K\left(\frac{\mathbf{F}_n(t_\nu) - \mathbf{F}_n(X_{\ell_\nu})}{h_n}\right) K\left(\frac{\mathbf{F}_{0,n}(Y_{i_\nu}) - \mathbf{F}_{0,n}(Y_{\ell_\nu})}{h_n}\right)}{\frac{(n-m)!}{n! h_n^{mp}} \sum_{(i_1,\ldots,i_m)\in I(m,n)} \prod_{\nu=1}^{m} K\left(\frac{\mathbf{F}_n(t_\nu) - \mathbf{F}_n(X_{i_\nu})}{h_n}\right)}, \tag{2.5}$$

where

$$I(m, n) = \{\mathbf{i} = (i_1, \ldots, i_m) : 1 \le i_j \le n \quad \text{and} \quad i_j \ne i_r \quad \text{if} \quad j \ne r\},$$

is the set of all m-tuples of different integers between 1 and n. Remark that we take 'delete one' observation estimators because they are natural in that they have more expressed expected values and lead to U-statistics rather than to V-statistics, but the bias introduced if we did not delete the diagonal observations would make no difference whatsoever. By extending the idea of the construction of the estimator (2.2), we now introduce our estimator of $r^{(m)}(\varphi, \mathbf{t})$

$$\widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h_n) = \frac{(n-m)!}{n!(n-1)^m} \sum_{\substack{(i_1,\ldots,i_m),(\ell_1,\ldots,\ell_m)\in I(m,n)\\ i_k\ne \ell_k, k=1,\ldots,m}} \varphi(Y_{i_1}, \ldots, Y_{i_m}) \mathbf{W}(\mathbf{i}, \boldsymbol{\ell}, \mathbf{t}; h_n).$$

In the particular cases of a single covariate ($p = 1$) or mutually independent predictors, the weight reduces to

$$\mathbf{W}(\mathbf{i}, \boldsymbol{\ell}, \mathbf{t}; h_n) = \frac{1}{h_n^{m(1+q)}} \prod_{\nu=1}^{m} K\left(\frac{\mathbf{F}_n(t_\nu) - \mathbf{F}_n(X_{\ell_\nu})}{h_n}\right) K\left(\frac{\mathbf{F}_{0,n}(Y_{i_\nu}) - \mathbf{F}_{0,n}(Y_{\ell_\nu})}{h_n}\right). \tag{2.6}$$

From our proofs, we have

$$\mathbb{E}\left(\widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)\right) = r^{(m)}(\varphi, \mathbf{t}) + o(1).$$

For $m \le n$, consider a class $\mathscr{F}$ of measurable functions $g : \mathbb{R}^{qm} \to \mathbb{R}$ such that

$$\mathbb{E}g^2(Y_1, \ldots, Y_m) < \infty,$$

which satisfies the following conditions, (F.i)–(F.iii). First, to avoid measurability problems, we assume that

$$\mathscr{F} \text{ is a pointwise measurable class,} \tag{F.i}$$

that is, there exists a countable subclass $\mathscr{F}_0$ of $\mathscr{F}$ such that we can find, for any function $g \in \mathscr{F}$, a sequence of functions $g_m \in \mathscr{F}_0$ for which

$$g_m(y) \to g(y), \quad y \in \mathbb{R}^{qm}.$$

This condition is discussed in [141, Example 2.3.4. p 110] and [97, 8.2. p. 110]. We also assume that $\mathscr{F}$ has a measurable envelope function

$$F(y) \ge \sup_{g\in\mathscr{F}} |g(y)| \quad \text{for} \quad y \in \mathbb{R}^{qm}. \tag{F.ii}$$

Notice that condition (F.i) implies that the supremum in (F.ii) is measurable. Finally, we assume that $\mathscr{F}$ is of VC-type, with characteristics A and v ("VC" for Vapnik and Červonenkis), meaning that for some $A \geq 3$ and $v \geq 1$,

$$\mathcal{N}(\mathscr{F}, L_2(Q), \varepsilon) \leq \left(\frac{A\|F\|_{L_2(Q)}}{\varepsilon}\right)^v, \quad \text{for} \quad 0 < \varepsilon \leq 2\|F\|_{L_2(Q)}, \tag{F.iii}$$

where Q is any probability measure on $(\mathbb{R}^m, \mathcal{B})$, where $\mathcal{B}$ represents the σ-field of Borel sets of $\mathbb{R}^m$, such that

$$\|F\|_{L_2(Q)} < \infty,$$

and where for $\varepsilon > 0$, $\mathcal{N}(\mathscr{F}, L_2(Q), \varepsilon)$ is defined as the smallest number of $L_2(Q)$-open balls of radius ε required to cover $\mathscr{F}$. For instance, see [114, Examples 26 and 38], [110, Lemma 22], [58, Section 4.7.], [141, Theorem 2.6.7], [97, Section 9.1] provide a number of sufficient conditions under which (F.i) holds, we may refer also to [51, Section 3.2] for further discussions. For instance, it is satisfied, for general $p \geq 1$, whenever $g(x) = \phi(\iota(x))$, with $\iota(x)$ is a polynomial in p variables and $\phi(\cdot)$ is a real-valued function of bounded variation, we refer the reader to [62, p. 1381].
Let $\Psi_\phi = \{g(x) : x \in \mathbb{R}^d\}$ be a class of functions such that $g(x) = \psi(\phi(x))$, where $\phi(\cdot)$ is either a real polynomial or the αth power of the absolute value of a real polynomial, for some $\alpha > 0$, and $\psi(r)$, $r \in \mathbb{R}$, is some real-valued function of bounded variation.
To state our results, we need the following conditions for the measurable real-valued function $H(t), t \in \mathbb{R}^p$.

(K.1): $H(t)$ is twice differentiable, and the second partial derivatives are bounded;
(K.2): $H(t)$ is compactly supported;
(K.3): $H(t)$ and its first partial derivatives belong to the class Ψ_ϕ.

If (F.iii) holds for $\mathscr{F}$, then we say that the VC-type class $\mathscr{F}$ admits the characteristics A and v. Consider the class of functions

$$\mathscr{K} := \left\{K\left(\frac{t_1 - \cdot}{h}\right)K\left(\frac{t_2 - \cdot}{h}\right) : h > 0, (t_1, t_2) \in \mathbb{R}^{p+q}\right\}.$$

Introduce the class of functions formed from the Lipschitz continuous copula density $\mathbf{c}(\cdot, \cdot)$

$$\mathscr{C} = \left\{\prod_{k=1}^m \mathbf{c}(t_k, \mathbf{F}_0(\cdot)) : t_k \in [0,1]^p\right\}.$$

We define the class of functions

$$\mathscr{F} \cdot \mathscr{C} = \left\{\varphi(\cdot)\prod_{k=1}^m \mathbf{c}(t_k, \mathbf{F}_0(\cdot)) : \varphi \in \mathscr{F}, t_k \in [0,1]^p\right\}.$$

Making use of condition (F.iii) and the Lipschitz continuity of the copula density readily implies that $\mathscr{F} \cdot \mathscr{C}$ is of VC-type for the envelope

$$\mathfrak{F}(\cdot) = F(\cdot)\prod_{k=1}^m \max_{t_k \in [0,1]^p} \mathbf{c}(t_k, \mathbf{F}_0(\cdot)).$$

To prove the strong consistency of $\widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$, we shall consider another, but more appropriate and more computationally convenient, centering factor than the expectation $\mathbb{E}\widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$, which is delicate to handle. We set

$\mathbb{E}\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$

$$= \frac{\mathbb{E}\left\{\frac{1}{h_n^{m(p+q)}}\varphi(Y_1,\ldots,Y_m)\prod_{j=1}^{m}K\left(\frac{\mathbf{F}_n(t_j)-\mathbf{F}_n(X_j)}{h_n}\right)K\left(\frac{\mathbf{F}_{0,n}(Y_{m+1})-\mathbf{F}_{0,n}(Y_j)}{h_n}\right)\right\}}{\mathbb{E}\left\{\frac{1}{h_n^{mp}}\prod_{j=1}^{m}K\left(\frac{\mathbf{F}_n(t_j)-\mathbf{F}_n(X_{i_j})}{h_n}\right)\right\}}.$$

We denote by $\mathbf{I}$ and $\mathbf{J}$ two fixed subsets of $\mathbb{R}^m$ such that

$$\mathbf{I}=\prod_{j=1}^{m}[a_j,b_j]\subset\mathbf{J}=\prod_{j=1}^{m}[c_j,d_j]\subset\mathbb{R}^m,$$

where

$$-\infty<c_j<a_j<b_j<d_j<\infty,\quad\text{for}\quad j=1,\ldots,m.$$

The main results in this section, concerning the uniform consistency of $\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h_n)$, to be proved here may now be stated precisely as follows.

Theorem 2.2. *Suppose that the copula density $\mathbf{c}(\cdot)$ is Lipschitz continuous on $(0,1)^{p+q}$ and let $a_n=\varrho(\log n/n)$ for $\varrho>0$. If the class of functions $\mathscr{F}\cdot\mathscr{C}$ is bounded, in the sense that for some $0<M<\infty$,*

$$\mathfrak{F}(y)<M.$$

Let $K\times K$ satisfy **(K.1-2-3)**. *Assume the conditions (F.i)-(F.iii) on $\mathscr{F}$ are satisfied. Then, for all $\varrho>0$ and $0<b_0<1$, there exists a constant $0<\Sigma<\infty$ such that*

$$\limsup_{n\to\infty}\sup_{a_n\le h^{p+q}\le b_0}\sup_{\varphi\in\mathscr{F}}\sup_{\mathbf{t}\in\mathbf{I}^p}\frac{\sqrt{nh^{p+q}}|\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h)-\mathbb{E}\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h)|}{\sqrt{|\log h|\vee\log\log n}}\le\Sigma,\quad a.s.$$

The proof of Theorem 2.2 is postponed until Section 10.

Theorem 2.3. *Suppose that the copula density $\mathbf{c}(\cdot)$ is Lipschitz continuous on $(0,1)^{p+q}$ and let $a_n=\varrho(\log n/n)$ for $\varrho>0$. If $\mathscr{F}\cdot\mathscr{C}$ is unbounded, but satisfies*

$$\mathbb{E}(\mathfrak{F}^2(Y))<\infty. \tag{2.7}$$

Let $K\times K$ satisfy **(K.1-2-3)**. *Assume the conditions (F.i)-(F.iii) on $\mathscr{F}$ are satisfied. Then for all $\varrho>0$ and $0<b_0<1$, there exists a constant $0<\breve{\Sigma}<\infty$ such that*

$$\limsup_{n\to\infty}\sup_{a_n\le h^{p+q}\le b_0}\sup_{\varphi\in\mathscr{F}}\sup_{\mathbf{t}\in\mathbf{I}^p}\frac{\sqrt{nh^{p+q}}|\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h)-\mathbb{E}\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h)|}{\sqrt{|\log h|\vee\log\log n}}\le\breve{\Sigma},\quad a.s.$$

The proof of Theorem 2.3 is postponed until Section 10.
To handle the bias term, we shall assume that the copula density $\mathbf{c}(\cdot)$ admits derivatives of order s such that

(C.i) There exists a constant $0<\mathfrak{C}<\infty$ such that

$$\sup_{u,v\in[0,1]^{p+q}}\left|\frac{\partial^s\mathbf{c}(u,v)}{\partial^{j_1}u_1\cdots\partial^{j_p}u_p\partial^{j_{p+1}}v_{p+1}\cdots\partial^{j_{p+q}}v_q}\right|\le\mathfrak{C},\quad\sum_{i=1}^{p+q}j_i=s.$$

(K.v) $\mathbb{K}(u,v)=K(u)K(v)$ is of order s ; i.e.,

$$\int_{\mathbb{R}^{p+q}}u^{j_1}\cdots u^{j_p}v^{j_{p+1}}\cdots v^{j_{p+q}}\mathbb{K}(u,v)dudv=0,\quad j_i\ge0,\quad\sum_{i=1}^{p+q}j_i=1,\ldots,s-1,$$

$$\int_{\mathbb{R}^{p+q}}u^{j_1}\cdots u^{j_p}v^{j_{p+1}}\cdots v^{j_{p+q}}\mathbb{K}(u,v)dudv<\infty,\quad j_i\ge0,\quad\sum_{i=1}^{p+q}j_i=s.$$

Remark 2.4. Notice that the conditions (K.v) and (K.i) are classical in the nonparametric estimation procedures. In particular, by imposing the condition (K.v), the kernel function exploits the smoothness of the copula density function. In the univariate case, it is well known that the best obtainable rate of convergence of the kernel estimator, in the AMISE sense, is of order $n^{-4/5}$. If we lose the condition that the kernel function $K(\cdot)$ must be a density, the convergence rate could be faster. Indeed, the convergence rate can be arbitrarily close to the parametric n^{-1} as the order increases. In fact, Chacón and Duong [40] showed that the parametric rate n^{-1} can be attained by using superkernels, and that superkernel density estimators automatically adapt to the unknown degree of smoothness of the density. In this situation, the main drawback of higher-order kernels is that the negative contributions of the kernel may make the estimated density, not a density itself. The interested reader may refer to, e.g., [92].

Corollary 2.1. Under the assumption of Theorem 2.3, in addition, we assume that (C.i) and (K.v) hold. It follows that for all sequences $0 < a_n \leq \tilde{a}_n \leq b_n$ satisfying $b_n \to 0$ and $n\tilde{a}_n/\log n \to \infty$,

$$\sup_{\widetilde{a}_n \leq h^{p+q} \leq b_n} \sup_{\varphi \in \mathscr{F}} \sup_{\mathbf{t} \in \mathbf{I}^p} |\widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h) - r^{(m)}(\varphi, \mathbf{t})| \to 0, \quad a.s.$$

Remark 2.5. Deheuvels and Mason [53] consider local plug-in type estimators $\widehat{h}_n = \widehat{h}_n(\mathbf{x})$, which satisfy,

$$\mathbb{P}\left(a_n \leq \widehat{h}_n(t) \leq b_n : t \in \mathbb{R}\right) \to 1,$$

with $a_n = c_1 h_n$ and $b_n = c_2 h_n$, where $0 < c_1 \leq c_2 < \infty$, or fulfill, for any $\varepsilon > 0$

$$\mathbb{P}\left(\sup_{x \in I} \left|\frac{\widehat{h}_n(t)}{h_n} - \eta(t)\right| > \varepsilon\right) \to 0, \tag{2.8}$$

where $\eta(\cdot)$ is an appropriate continuous function on $\mathbb{R}$ and $I = [a, b] \subset \mathbb{R}$, for $a < b$. We refer to their Example 2.1 p. 246, where they show subject to smoothness conditions that the optimal $\widehat{h}_n(t)$ satisfies (2.8) with $h_n = n^{-1/5}$, for $d = 1$, in terms of asymptotic mean square error for estimating the density function $f(\cdot)$ or regression function $m(\cdot)$. Following their methods, it will be interesting to derive our results for local plug-in estimators $\widehat{h}_n(t)$, where the convergence is either in probability or with probability 1, depending on conditions on $\widehat{h}_n(t)$. We omit the corresponding details here.

Remark 2.6. We note that the main problem in using an estimator such as in (2.6) is properly choosing the smoothing parameter h. The uniform in bandwidth consistency results given in Theorems 2.2, 2.3 show that any choice of h between h'_n and h''_n ensures the consistency of $\widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h)$. Namely, the fluctuation of the bandwidth in a small interval does not affect the consistency of the nonparametric estimator $\widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h)$ of $r^{(m)}(\varphi, \mathbf{t})$.

Remark 2.7. Note that the condition (2.7) may be replaced by more general hypotheses upon moments of Y as in [51]. That is

(M.1)$''$ We denote by $\{\mathcal{M}(x) : x \geq 0\}$ a nonnegative continuous function, increasing on $[0, \infty)$, and such that, for some $s > 2$, ultimately as $x \uparrow \infty$,

$$(i)\ x^{-s}\mathcal{M}(x) \downarrow; (ii)\ \ x^{-1}\mathcal{M}(x) \uparrow. \tag{2.9}$$

For each $t \geq \mathcal{M}(0)$, we define $\mathcal{M}^{inv}(t) \geq 0$ by $\mathcal{M}(\mathcal{M}^{inv}(t)) = t$. We assume further that:

$$\mathbb{E}\left(\mathcal{M}\left(|\mathfrak{F}(\mathbf{Y})|\right)\right) < \infty.$$

The following choices of $\mathcal{M}(\cdot)$ are of particular interest:

(i) $\mathcal{M}(x) = x^p$ for some $p > 2$;
(ii) $\mathcal{M}(x) = \exp(sx)$ for some $s > 0$.

The introduction of the function $\varphi(\cdot)$ in our setting is motivated by Remark 1.2 of [53] or Remark 1.1 of [51].

Remark 2.8. We start with examples of which θ varies within subsets of $\mathbb{R}$. Such is the case for the extreme value copulas, namely

$$\mathbb{C}_A(u_1, u_2) := \exp\left\{\log u_1 u_2 A\left(\frac{\log u_1}{\log u_1 u_2}\right)\right\}, \tag{2.10}$$

where $A(\cdot)$ is a convex function on $[0, 1]$, satisfying

$$A : [0, 1] \mapsto [1/2, 1] \text{ such that } \max(t, 1-t) \leq A(t) \leq 1 \text{ for all } 0 \leq t \leq 1.$$

For

$$A(t) := A_\theta(t) = (t^\theta + (1-t)^\theta)^{1/\theta}; \quad \theta \in [1, \infty[\tag{2.11}$$

we have [77] family of copulas, which is one of the most popular model used to model bivariate extreme values. For

$$A_\theta(t) = 1 - (t^{-\theta} + (1-t)^{-\theta})^{-1/\theta}; \quad \theta \in [0, \infty[\tag{2.12}$$

we obtain [69] family of copulas. Finally for

$$A_\theta(t) = t\Phi\left(\theta^{-1} + \frac{1}{2}\theta \log\left(\frac{t}{1-t}\right)\right) + (1-t)\Phi\left(\theta^{-1} - \frac{1}{2}\theta \log\left(\frac{t}{1-t}\right)\right), \tag{2.13}$$

where $\theta \in [0, \infty[$ and $\Phi(\cdot)$ denoting the standard normal $N(0, 1)$ distribution function, we obtain the [87] family of copulas. A useful family of copulas, due to, is given, for $0 < u_1, u_2 < 1$, by

$$\mathbb{C}_\theta(u_1, u_2) := 1 - \left[(1-u_1)^\theta + (1-u_2)^\theta - (1-u_1)^\theta(1-u_2)^\theta\right]^{1/\theta}; \quad \theta \in [1, \infty[. \tag{2.14}$$

The Gumbel-Barnett copulas are given, for $0 < u_1, u_2 < 1$, by

$$\mathbb{C}_\theta(u_1, u_2) := u_1 u_2 \exp\{-(1-\theta)(\log u_1)(\log u_2)\}; \quad \theta \in [0, 1]. \tag{2.15}$$

The Clayton copulas of positive dependence are such that, for $0 < u_1, u_2 < 1$,

$$\mathbb{C}_\theta(u_1, u_2) = \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta}; \quad \theta \in]0, \infty[. \tag{2.16}$$

Parametric families of copulas with parameter θ varying in $\mathbb{R}^p$, for some $p \geq 2$, include the following classical examples. Below, we set $\theta = (\theta_1, \theta_2)^\top \in \mathbb{R}^2$.

$$\mathbb{C}_\theta(u_1, u_2) := \left\{1 + \left[(u_1^{-\theta_1} - 1)^{\theta_2} + (u_2^{-\theta_1} - 1)^{\theta_2}\right]^{1/\theta_2}\right\}^{-1/\theta_1}, \ \theta \in]0, \infty[\times[1, \infty[; \tag{2.17}$$

$$\begin{aligned}\mathbb{C}_\theta(u_1, u_2) := \exp\Big\{ -\Big[\theta_2{}^{-1} \log\Big(&\exp\left(-\theta_2(\log u_1)^{\theta_1}\right) \\ &+ \exp\left(-\theta_2(\log u_2)^{\theta_1}\right) - 1\Big)\Big]^{1/\theta_1}\Big\}, \ \theta \in [1, \infty[\times]0, \infty[.\end{aligned} \tag{2.18}$$

For other examples of this kind, we refer to [89]. One can see, for example, that the copula density function given in (2.16) is not bounded on $[0, 1]^2$. It will be of interest to weak conditions used in [121]. Let $V_{d,j} = \{u \in [0, 1]^{p+q} : 0 < u_j < 1\}$ for $j \in \{1, \ldots, p+q\}$. For every $i, j \in \{1, \ldots, p+q\}$, the second-order partial derivative of the copula function $\ddot{C}_{ij}$ is defined and continuous on the set $V_{d,i} \cap V_{d,j}$, and there exists a constant $K > 0$ such that

$$|\ddot{C}_{ij}(u)| \leq K \min\left(\frac{1}{u_i(1-u_i)}, \frac{1}{u_j(1-u_j)}\right), \qquad u \in V_{d,i} \cap V_{d,j}.$$

A similar condition was used in [16] in approximating the Kac empirical copula processes by appropriate Poisson processes. A weaker than imposing an Hölder condition on the derivatives like in [121] was considered in [65], condition C.1. Another way to circumvent the boundedness condition on the copula density function is to consider weighted versions of Theorems 2.2 and 2.3 in a similar way as in Proposition 1.1 of [74]. We point out that we are not mainly concerned with estimating the copula density itself, with a particular focus on estimation near the boundaries of the unit square. It is well known that the estimations based on symmetric kernels are inconsistent on the boundaries. They suffer from the so-called boundary bias. Such bias can be significant in the neighborhood of the boundaries too, depending on the size of the bandwidth. Several solutions can be proposed to cope with such issues: mirror image modification transformed kernels and boundary kernels. In the last one, a smooth distortion is considered near the border so that the bandwidth and the kernel shape can be modified. In this setting, the beta kernels have received particular interest.

3. Weak convergence

Let us introduce the U-statistic process

$$u_n(\varphi, h_n, \mathbf{t}) := \sqrt{nh_n^{p+q}}(\widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h_n) - \mathbb{E}\widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)).$$

As in [4], we say that the CLT (Central Limit Theorem) holds for the process

$$\{n^{1/2}(U_m^n(f, \mathbb{P}) - \mathbb{E}f) : f \in \mathscr{F}\}$$

if there is a Gaussian process $\{G(f) : f \in \mathscr{F}\}$ which has a version with bounded and d-uniformly continuous paths, d being the pseudo distance defined by

$$d^2(f, g) := \mathbb{V}ar(\mathbb{P}^{m-1}(f - g)),$$

and if

$$n^{1/2}(U_m^n(f, \mathbb{P}) - \mathbb{E}f) \stackrel{\mathscr{L}}{\to} G(f) \quad \text{uniformly in} \quad \ell^\infty(\mathscr{F}) \tag{3.1}$$

where convergence is in the sense of Hoffmann-Jørgensen, see, e.g., [50, 97, 142] for definition and further details. Then $\{G(f) : f \in \mathscr{F}\}$ is a centered Gaussian process indexed by the class $\mathscr{F}$, with covariance, for $f, g \in \mathscr{F}$,

$$\mathbb{E}G(f)G(g) = m^2\mathbb{P}[(\mathbb{P}^{m-1}f)(\mathbb{P}^{m-1}g)] - m^2(\mathbb{P}^m f)(\mathbb{P}^m g).$$

It is well-known that (3.1) is equivalent to both $(\mathscr{F}, d)$ being totally bounded and

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}^* \left\{ \sup_{d(f,g)\leq\delta} n^{1/2} \left|U_m^n(f, \mathbb{P}) - \mathbb{E}f - U_m^n(g, \mathbb{P}) + \mathbb{E}g\right| > \epsilon \right\} = 0 \tag{3.2}$$

for all $\epsilon > 0$, where $\mathbb{P}^*$ is the outer measure if the variable within the bracket is not measurable, refer to [97, 142] for definition. If $\{G(f) : f \in \mathscr{F}\}$ has a version with bounded, ρ-uniformly continuous paths for some pseudodistance ρ on $\mathscr{F}$, then (3.2) with d replaced by ρ is sufficient for the CLT (3.1). We introduce now a few more notation to state the asymptotic normality of $\widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$. For $1 \leq j, l, k \leq m$ and $t_1, \ldots, t_{3m} \in \mathbb{R}^{3pm}$, define

$$\begin{aligned} & r_{j,l}(t_1, \ldots, t_m) \\ & \quad := \mathbb{E}\big[\varphi(Y_1, \ldots, Y_{j-1}, Y, Y_{j+1}, \ldots, Y_m)\varphi(Y_{m+1}, \ldots, Y_{m+l-1}, Y, Y_{m+l+1}, \ldots, Y_{2m}) \\ & \qquad \big|X = t_j\ ;\ X_i = t_i, \forall i = 1, \ldots, m, i \neq j\ ;\ X_{m+i} = t_i, \forall i = 1, \ldots, m, i \neq l\big], \end{aligned} \tag{3.3}$$

$$\begin{aligned} & \tilde{r}_{j,l}(t_1, \ldots, t_{2k}) \\ & \quad := \mathbb{E}\big[\varphi(Y_1, \ldots, Y_{j-1}, Y, Y_{j+1}, \ldots, Y_m) \\ & \qquad \varphi(Y_{m+1}, \ldots, Y_{m+l-1}, Y, Y_{m+l+1}, \ldots, Y_{2m}) \\ & \qquad \big|X_j = t_j\ ;\ X_i = t_i, \forall i = 1, \ldots, 2m, i \notin \{j, m+i\}\big]. \end{aligned} \tag{3.4}$$

$$\begin{aligned} & r_{j,l,m}(t_1, \ldots, t_{3k}) \\ & \quad := \mathbb{E}\big[\varphi(Y_1, \ldots, Y_{j-1}, Y, Y_{j+1}, \ldots, Y_m) \\ & \qquad \varphi(Y_{m+1}, \ldots, Y_{m+l-1}, Y, Y_{m+l+1}, \ldots, Y_{2m}) \\ & \qquad \varphi(Y_{2m+1}, \ldots, Y_{2m+k-1}, X, Y_{2m+k+1}, \ldots, Y_{3m}) \end{aligned}$$

$$|X_i = t_i, \forall i = 1, \ldots, 3m, X = t_j, i \notin \{j, m+l, 2m+k\}]. \tag{3.5}$$

Assumption 3.1. [(i)]

(1) $h_n \to 0$ and $nh_n^p \to \infty$;
(2) $\mathbb{K} = K \times K$ is symmetric at 0, bounded and compactly supported ;
(3) $r_{j,l}$ is continuous at $(t_1, \ldots, t_m)$ for all $1 \le j, l \le m$;
(4) $r^{(m)}$ is two times continuously differentiable in a neighborhood of $(t_1, \ldots, t_m)$;
(5) $r_{j,l,k}$ is bounded in a neighborhood of $(t_1, \ldots, t_m, t_1, \ldots, t_m, t_1, \ldots, t_m) \in \mathbb{R}^{3pm}$, for all $1 \le j, l, k \le m$;
(6) $\breve{\mathbf{c}}$ is twice differentiable in neighborhoods of $t_i, 1 \le i \le m$.

Proposition 3.2. *[Asymptotic normality of $\widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$, Corollary 2.4 in [135]] Under Assumption 3.1, we have*

$$\sqrt{nh_n^{p+q}}(\widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h_n) - r^{(m)}(\varphi, \mathbf{t})) \to N(0, \rho^2),$$

where

$$\rho^2 := \sum_{j,i=1}^{m} \mathbb{1}_{\{t_j = t_l\}}(r_{j,l}(t_1, \ldots, t_m) - r^{(m)^2}(\varphi, \mathbf{t})\|\mathbb{K}\|_2^2/\breve{\mathbf{c}}(\mathbf{F}(t_j)).$$

Moreover, let $\mathfrak{B}$ be a positive integer, and $(t_1^{(1)}, \ldots, t_m^{(1)}, \ldots, t_1^{(\mathfrak{B})}, \ldots, t_k^{(\mathfrak{B})}) \in \mathbb{R}^{pm \times \mathfrak{B}}$. Then under similar regularity conditions,

$$\sqrt{nh_n^{p+q}}(\widetilde{r}_n^{(m)}(\varphi, \mathbf{t}^{(i)}; h_n) - r^{(m)}(\varphi, \mathbf{t}^{(i)}))_{i=1,\ldots,\mathfrak{B}} \to N(0, \Sigma),$$

where, for $1 \le \tilde{j}, \tilde{i} \le \mathfrak{B}$,

$$(\Sigma)_{\tilde{j},\tilde{i}} := \sum_{j,i=1}^{k} \mathbb{1}_{\left\{t_j^{(\tilde{j})} = t_i^{(\tilde{i})}\right\}} \left(r_{j,i} \left(t_1^{(\tilde{j})}, \ldots, t_m^{(\tilde{j})}, t_1^{(\tilde{i})}, \ldots, t_m^{(\tilde{i})} \right) - r^{(m)}(\varphi, \mathbf{t}^{(\tilde{i})}) r^{(m)}(\varphi, \mathbf{t}^{(\tilde{j})}) \right) \frac{\|\mathbb{K}\|_2^2}{\breve{\mathbf{c}}(\mathbf{F}(t_j^{\tilde{j}}))}.$$

Note that the second part of Proposition 3.2 above is a consequence of the first one. Indeed, for every $(c_1, \ldots, c_N) \in \mathbb{R}^N$, we can define

$$r(x_1^{(1)}, \ldots, x_k^{(1)}, \ldots, x_1^{(N)}, \ldots, x_k^{(N)}) := \sum_{\tilde{i}=1}^{N} c_{\tilde{i}} r^{(m)}(\varphi, \mathbf{t}^{(\tilde{i})})$$

and corresponding versions of φ, $\widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$ and ρ^2. Finally, the conclusion follows from the Cramér-Wold device.

Theorem 3.3. *Let $\mathscr{F}$ be a measurable class of symmetric functions on S^m such that $t^2\mathbb{P}\{\mathbf{F} > t\} = 0$. Then the following statement hold, in law,*

$$\{u_n(\varphi, h_n, \mathbf{t}) : \varphi \in \mathscr{F}, K \in \mathscr{K}\} \to \{mG_{\mathbb{P}} \circ \mathbb{P}^{m-1}\psi : \psi \in \mathscr{F} \cdot \mathscr{K}\}.$$

The proof of Theorem 3.3 is postponed until Section 10.

Remark 3.4. If φ is not symmetric we will need to symmetrize it. To do this we define the following function

$$\overline{\varphi}(\mathbf{y}) := \frac{1}{m!} \sum_{\sigma \in I_m^m} \varphi(\mathbf{y}_\sigma),$$

where $\mathbf{y}_\sigma = (y_{\sigma_1}, \ldots, y_{\sigma_m})$. After symmetrization the expectation remain unchanged and the U-statistic $u_n(\varphi, h_n, \mathbf{t}) = u_n(\overline{\varphi}, h_n, \mathbf{t})$ do not change.

4. Wild bootstrap

Define a sequence $(Z_n)_{n\geq 1}$ of i.i.d. replicæ of a *strictly positive* random variable Z with distribution function $G(\cdot)$, independent of the (X_n, Y_n)'s. In the sequel, the following assumptions on the Z_n's will prevail:

(A1) $\mathbb{E}(Z) = 1$; $\quad\mathbb{E}(Z^2) = 2$ (or, equivalently, $\mathbf{Var}(Z) = 1$).

We now introduce the wild bootstrap of the estimator of $r^{(m)}(\varphi, \mathbf{t})$

$$\widetilde{r}_n^{(m)^*}(\varphi, \mathbf{t}; h_n) = \frac{(n-m)!}{n!(n-1)^m} \sum_{\substack{(i_1,\ldots,i_m),(\ell_1,\ldots,\ell_m)\in I(m,n)\\ i_k\neq \ell_k, k=1,\ldots,m}} Z_{i_1}\cdots Z_{i_m}\varphi(Y_{i_1},\ldots,Y_{i_m})\mathbf{W}(\mathbf{i},\boldsymbol{\ell},\mathbf{t};h_n).$$

Let us introduce the bootstrapped U-statistic process

$$u_n^*(\varphi, h_n, \mathbf{t}) := \sqrt{n}(\widetilde{r}_n^{(m)^*}(\varphi, \mathbf{t}; h_n) - \widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)).$$

Notice that

$$\begin{aligned}
&\widetilde{r}_n^{(m)^*}(\varphi, \mathbf{t}; h_n) - \widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)\\
&\quad= \frac{(n-m)!}{n!(n-1)^m} \sum_{\substack{(i_1,\ldots,i_m),(\ell_1,\ldots,\ell_m)\in I(m,n)\\ i_k\neq \ell_k, k=1,\ldots,m}} (Z_{i_1}\cdots Z_{i_m}-1)\varphi(Y_{i_1},\ldots,Y_{i_m})\mathbf{W}(\mathbf{i},\boldsymbol{\ell},\mathbf{t};h_n).
\end{aligned}$$

We have

$$\mathbb{E}(\widetilde{r}_n^{(m)^*}(\varphi, \mathbf{t}; h_n) - \widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)) = 0.$$

Let us introduce

$$\Psi(z_1,\ldots,z_m,y_1,\ldots,y_m) = \left(\prod_{i=1}^m z_i - 1\right)\varphi(y_1,\ldots,y_m).$$

Remark that we have

$$\begin{aligned}
u_n^*(\varphi, h_n, \mathbf{t}) &= \sqrt{nh_n^{p+q}}(\widetilde{r}_n^{(m)^*}(\varphi, \mathbf{t}; h_n) - \widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h_n))\\
&= \sqrt{nh_n^{p+q}}(\widetilde{r}_n^{(m)^*}(\varphi, \mathbf{t}; h_n) - \widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)\\
&\quad -\mathbb{E}\widetilde{r}_n^{(m)^*}(\varphi, \mathbf{t}; h_n) - \widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h_n))\\
&= u_n(\Psi, h_n, \mathbf{t}).
\end{aligned}$$

In this setting, the envelope to be considered is given by

$$\widetilde{\mathbf{F}}(\cdot) = F(\cdot)\left(\prod_{i=1}^m z_i - 1\right)\prod_{k=1}^m \max_{0\leq t_k\leq 1}\mathbf{c}(t_k, \mathbb{G}_Y(\cdot)).$$

The following theorem can be shown by following similar arguments as those used in Theorem 2.3. More precisely, it suffices to replace φ by Ψ.

Theorem 4.1. *Suppose that the copula density $\mathbf{c}(\cdot,\cdot)$ is continuous and let $a_n = \varrho(\log n/n)$ for $\varrho > 0$. If $\mathscr{F}\cdot\mathscr{C}$ is unbounded, but satisfies*

$$\mathbb{E}(\widetilde{\mathbf{F}}^2(Y)) < \infty, \tag{4.1}$$

then we can infer, under the above-mentioned assumptions on $\mathscr{F}$ and $\mathscr{K}$, that for all $\varrho' > 0$ and $0 < b_0 < 1$, there exists a constant $0 < \Sigma' < \infty$ such that

$$\limsup_{n\to\infty}\sup_{a_n\leq h^{p+q}\leq b_0}\sup_{\varphi\in\mathscr{F}}\sup_{\mathbf{t}\in\mathbf{I}^p}\frac{\sqrt{nh^{p+q}}|\widetilde{r}_n^{(m)^*}(\varphi, \mathbf{t}; h) - \widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h))|}{\sqrt{|\log h|\vee\log\log n}} \leq \Sigma', \quad a.s.$$

The main application of this result is the following

$$\limsup_{n\to\infty}\sup_{\varphi\in\mathscr{F}}\sup_{\mathbf{t}\in\mathbf{I}^p}\frac{|\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h)-\mathbb{E}\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h)|}{|\widetilde{r}_n^{(m)^*}(\varphi,\mathbf{t};h)-\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h))|}\leq\Gamma,$$

where Γ is a positive constant. It will be of interest to refine this result by showing

$$\limsup_{n\to\infty}\sup_{\varphi\in\mathscr{F}}\sup_{\mathbf{t}\in\mathbf{I}^p}\frac{|\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h)-\mathbb{E}\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h)|}{|\widetilde{r}_n^{(m)^*}(\varphi,\mathbf{t};h)-\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h))|}=1+o(1).$$

The last relation may be used to construct a uniform asymptotic certainty band similar to Remark 2.3 of [51].

Theorem 4.2. *Let $\mathscr{F}$ be a measurable class of symmetric functions on S^m such that $v^2\mathbb{P}\{\mathfrak{F}>v\}=0$. Then the following statement hold, in law,*

$$\{u_n^*(\varphi,h_n,\mathbf{t}):\varphi\in\mathscr{F},K\in\mathscr{K}\}\to\{mG_{\mathbb{P}}^*\circ\mathbb{P}^{m-1}\psi:\psi\in\mathscr{F}\cdot\mathscr{K}\},$$

where $G_{\mathbb{P}}^*$ is an independent copy of the Gaussian process $G_{\mathbb{P}}$ defined in Theorem 3.3.

Application of Theorem 4.2, is well documented, we may refer for example to [11, 13, 14, 17, 26, 27, 33, 35, 132, 133].

Remark 4.3. It is well known that Theorem 4.2 can be used easily through routine bootstrap sampling, which we describe briefly as follows. Let N be a large integer. Let $Z_1^{(k)},\ldots,Z_n^{(k)}$, for $k=1,\ldots,N$, be a sample of weights satisfying the preceding conditions and independent of (X_n,Y_n)'s. Moreover, for any $k=1,\ldots,N$, let

$$u_n^{*k}(\varphi,h_n,\mathbf{t})=\sqrt{nh_n^{p+q}}(\widetilde{r}_n^{(m)^{*k}}(\varphi,\mathbf{t};h_n)-\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h_n)),$$

where, $\widetilde{r}_n^{(m)^{*k}}(\varphi,\mathbf{t};h_n)$ is defined by

$$\begin{aligned}&\widetilde{r}_n^{(m)^{*k}}(\varphi,\mathbf{t};h_n)\\&\quad=\frac{(n-m)!}{n!(n-1)^m}\sum_{\substack{(i_1,\ldots,i_m),(\ell_1,\ldots,\ell_m)\in I(m,n)\\ i_k\neq\ell_k,k=1,\ldots,m}}Z_{i_1}^{(k)}\cdots Z_{i_m}^{(k)}\varphi(Y_{i_1},\ldots,Y_{i_m})\mathbf{W}(\mathbf{i},\boldsymbol{\ell},\mathbf{t};h_n).\end{aligned}$$

Now, according to Theorem 4.2, we readily obtain that,

$$\{u_n^{*k}(\varphi,h_n,\mathbf{t}):\varphi\in\mathscr{F},K\in\mathscr{K}\}\to\{mG_{\mathbb{P}}^k\circ\mathbb{P}^{m-1}\psi\in\mathscr{F}\cdot\mathscr{K}\}.$$

where $G_{\mathbb{P}}^1,\ldots,G_{\mathbb{P}}^N$ are independent copies of $G_{\mathbb{P}}$. In order to approximate the limiting distribution of $\{u_n^*(\varphi,h_n,\mathbf{t}):\varphi\in\mathscr{F},K\in\mathscr{K}\}$, one can use the empirical distribution of $\{u_n^{*k}(\varphi,h_n,\mathbf{t}):\varphi\in\mathscr{F},K\in\mathscr{K}\}$, $k=1,\ldots,N$, for N large enough. To be more precise, if we are interested in performing a statistical test based on a *smooth* functional

$$S_n:=\varphi(u_n(\varphi,h_n,\mathbf{t})),$$

with the convention that large values of S_n lead to the rejection of the null hypothesis, $\mathcal{H}_0$ say, under some regularity conditions, a valid approximation to the p-value for the test based on S_n, for N large enough, is given by

$$\frac{1}{N}\sum_{k=1}^{N}\mathrm{1I}\{S_n^{(k)}\geq S_n\},$$

where

$$S_n^{(k)}:=\varphi(u_n^{*k}(\varphi,h_n,\mathbf{t})).$$

5. The censored case

Consider a triple $(Y, C, \mathbf{X})$ of random variables defined in $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^d$. Here Y is the variable of interest, C a censoring variable and $\mathbf{X}$ a concomitant variable. Throughout, we will use [103] notation and we work with a sample $\{(Y_i, C_i, \mathbf{X}_i)_{1\leq i\leq n}\}$ of independent and identically distributed replication of $(Y, C, \mathbf{X})$, $n \geq 1$. Actually, in the right censorship model, the pairs (Y_i, C_i), $1 \leq i \leq n$, are not directly observed and the corresponding information is given by $Z_i := \min\{Y_i, C_i\}$ and $\delta_i := \mathbb{1}\{Y_i \leq C_i\}$, $1 \leq i \leq n$. Accordingly, the observed sample is

$$\mathcal{D}_n = \{(Z_i, \delta_i, \mathbf{X}_i), i = 1, \ldots, n\}.$$

For example, survival data in clinical trials or failure time data in reliability studies are often subject to censoring. More specifically, many statistical experiments result in incomplete samples, even under well-controlled conditions. For example, clinical data for surviving most types of disease are usually censored by other competing risks to life, which result in death. In the sequel, we impose the following assumptions upon the distribution of $(\mathbf{X}, Y)$. Denote by $\mathbf{I}$ a given compact set in $\mathbb{R}^d$ with nonempty interior and set, for any $\alpha > 0$,

$$I_\alpha = \{\mathbf{x} : \inf_{\mathbf{u}\in\mathbf{I}} \|\mathbf{x} - \mathbf{u}\| \leq \alpha\}.$$

We will assume that, for a given $\alpha > 0$, $(\mathbf{X}, Y)$ [resp. $\mathbf{X}$] has a density function $f_{\mathbf{X},Y}$ [resp. $f_{\mathbf{X}}$] with respect to the Lebesgue measure on $I_\alpha \times \mathbb{R}$ [resp. I_α]. For $-\infty < t < \infty$, set

$$F_Y(t) = \mathbb{P}(Y \leq t), \quad G(t) = \mathbb{P}(C \leq t), \quad \text{and} \quad H(t) = \mathbb{P}(Z \leq t),$$

the right-continuous distribution functions of Y, C and Z respectively. For any right-continuous distribution function L defined on $\mathbb{R}$, denote by

$$T_L = \sup\{t \in \mathbb{R} : L(t) < 1\}$$

the upper point of the corresponding distribution. Now consider a pointwise measurable class $\mathcal{F}$ of real, measurable functions defined on $\mathbb{R}$, and assume that $\mathcal{F}$ is of VC-type. In this section, we will mostly focus on the regression function of $\psi(Y)$ evaluated at $\mathbf{X} = \mathbf{x}$, for $\psi \in \mathcal{F}$ and $\mathbf{x} \in I_\alpha$, given by

$$r^{(1)}(\psi, \mathbf{x}) = \mathbb{E}(\psi(Y) \mid \mathbf{X} = \mathbf{x}),$$

when Y is right-censored. To estimate $r^{(1)}(\psi, \cdot)$, we make use of the Inverse Probability of Censoring Weighted (I.P.C.W.) estimators wish have recently gained popularity in the censored data literature (see [39, 94]). The key idea of I.P.C.W. estimators is as follows. Introduce the real-valued function $\Phi_\psi(\cdot, \cdot)$ defined on $\mathbb{R}^2$ by

$$\Phi_\psi(y, c) = \frac{\mathbb{1}\{y \leq c\}\psi(y \wedge c)}{1 - G(y \wedge c)}. \tag{5.1}$$

Assuming the function $G(\cdot)$ to be known, first note that $\Phi_\psi(Y_i, C_i) = \delta_i\psi(Z_i)/(1 - G(Z_i))$ is observed for every $1 \leq i \leq n$. Moreover, under the Assumption **(I)** below,

(I): C and $(Y, \mathbf{X})$ are independent.

We have

$$\begin{aligned}
r^{(1)}(\Phi_\psi, \mathbf{x}) &:= \mathbb{E}(\Phi_\psi(Y, C) \mid \mathbf{X} = \mathbf{x}) \\
&= \mathbb{E}\left\{\frac{\mathbb{1}\{Y \leq C\}\psi(Z)}{1 - G(Z)} \mid \mathbf{X} = \mathbf{x}\right\} \\
&= \mathbb{E}\left\{\frac{\psi(Y)}{1 - G(Y)}\mathbb{E}(\mathbb{1}\{Y \leq C\} \mid \mathbf{X}, Y) \mid \mathbf{X} = \mathbf{x}\right\} \\
&= r^{(1)}(\psi, \mathbf{x}).
\end{aligned} \tag{5.2}$$

Therefore, any estimate of $r^{(1)}(\Phi_\psi,\cdot)$, which can be built on fully observed data, turns out to be an estimate for $r^{(1)}(\psi,\cdot)$ too. Thanks to this property, most statistical procedures that provide estimates of the regression function in the uncensored case can be naturally extended to the censored case. For instance, kernel-type estimates are particularly easy to construct. Set, for $\mathbf{x}\in\mathbf{I}$, $h\geq 0$, $1\leq i\leq n$,

$$\overline{\omega}^{(1)}_{n,K,h,i}(\mathbf{x}) := K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h}\right) \Big/ \sum_{j=1}^{n} K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h}\right). \tag{5.3}$$

In view of (5.1), (5.2), and (5.3), whenever $G(\cdot)$ is known, a kernel estimator of $r^{(1)}(\psi,\cdot)$ is given by

$$\breve{r}^{(1)}_n(\psi,\mathbf{x};h_n) = \sum_{i=1}^{n} \overline{\omega}^{(1)}_{n,K,h,i}(\mathbf{x})\frac{\delta_i\psi(Z_i)}{1-G(Z_i)}. \tag{5.4}$$

The function $G(\cdot)$ is generally unknown and has to be estimated. We will denote by $G_n^*(\cdot)$ the Kaplan-Meier estimator of the function $G(\cdot)$ [93]. Namely, adopting the conventions

$$\prod_{\emptyset} = 1$$

and $0^0=1$ and setting

$$N_n(u) = \sum_{i=1}^{n} \mathbb{1}\{Z_i\geq u\},$$

we have

$$G_n^*(u) = 1 - \prod_{i:Z_i\leq u}\left\{\frac{N_n(Z_i)-1}{N_n(Z_i)}\right\}^{(1-\delta_i)}, \quad \text{for} \quad u\in\mathbb{R}.$$

Given this notation, we will investigate the following estimator of $r^{(1)}(\psi,\cdot)$

$$\breve{r}^{(1)*}_n(\psi,\mathbf{x};h_n) = \sum_{i=1}^{n} \overline{\omega}^{(1)}_{n,K,h,i}(\mathbf{x})\frac{\delta_i\psi(Z_i)}{1-G_n^*(Z_i)}, \tag{5.5}$$

refer to [94] and [103]. Adopting the convention $0/0=0$, this quantity is well defined, since $G_n^*(Z_i)=1$ if and only if $Z_i=Z_{(n)}$ and $\delta_{(n)}=0$, where $Z_{(k)}$ is the kth ordered statistic associated with the sample $(Z_1,\ldots,Z_n)$ for $k=1,\ldots,n$ and $\delta_{(k)}$ is the δ_j corresponding to $Z_k=Z_j$. When the variable of interest is right-censored, functionals of the (conditional) law can generally not be estimated on the complete support. To obtain our results, we will work under the following assumptions.

- **(A.1):** $\mathscr{F}=\{\psi:=\psi\mathbb{1}\{(-\infty,\tau)^m\},\psi\in\mathscr{F}\}$, where $\tau<T_H$ and $\mathscr{F}_1$ is a pointwise measurable class of real measurable functions defined on $\mathbb{R}$ and of type VC.
- **(A.2):** The class of functions $\mathscr{F}$ has a measurable and uniformly bounded envelope function Υ with,

$$\Upsilon(y_1,\ldots,y_m)\geq \sup_{\psi\in\mathscr{F}} \mid \psi(y_1,\ldots,y_m)\mid, \quad y_i\leq T_H.$$

- **(A.3):** The class of functions $\mathcal{M}$ is relatively compact with respect to the sup- norm topology on I_α^m.

In what follows, we will study the uniform convergence of $\widetilde{m}^*_{\psi,n,h}(\mathbf{x})$ centered by the following centering factor

$$\widehat{\mathbb{E}}\breve{r}^{(1)*}_n(\psi,\mathbf{x};h_n) = \frac{\mathbb{E}\left(\psi(Y)K\left(\dfrac{\mathbf{x}-\mathbf{X}}{h}\right)\right)}{\mathbb{E}\left(K\left(\dfrac{\mathbf{x}-\mathbf{X}}{h}\right)\right)}.$$

This choice is justified by the fact that under hypothesis **(I)**, we have

$$\begin{aligned}\mathbb{E}\left\{\Phi_\psi(Y,C)K\left(\frac{\mathbf{x}-\mathbf{X}}{h}\right)\right\} &= \mathbb{E}\left\{\frac{\mathbb{1}\{Y\le C\}\psi(Z)}{1-G(Z)}K\left(\frac{\mathbf{x}-\mathbf{X}}{h}\right)\right\} \\ &= \mathbb{E}\left\{\frac{\psi(Y)K\left(\frac{\mathbf{x}-\mathbf{X}}{h}\right)}{1-G(Y)}\mathbb{E}[\mathbb{1}\{Y\le C\}\mid \mathbf{X},\mathbf{Y}]\right\} \\ &= \mathbb{E}\left\{\psi(Y)K\left(\frac{\mathbf{x}-\mathbf{X}}{h}\right)\right\}.\end{aligned} \tag{5.6}$$

A natural extension of the function defined in (5.1) is given by

$$\mathbf{\Phi}_\psi(y_1,\ldots,y_m,c_1,\ldots,c_m)=\frac{\prod_{i=1}^m\{\mathbb{1}\{y_i\le c_i\}\psi(y_1\wedge c_1,\ldots,y_m\wedge c_m)}{\prod_{i=1}^m\{1-G(y_i\wedge c_i)\}}.$$

From this, we have an analogous relation to (5.2) given by

$$\begin{aligned}&\mathbb{E}(\mathbf{\Phi}_\psi(Y_1,\ldots,Y_m,C_1,\ldots,C_m)\mid(\mathbf{X}_1,\ldots,\mathbf{X}_m)=\mathbf{t}) \\ &= \mathbb{E}\left(\frac{\prod_{i=1}^m\{\mathbb{1}\{Y_i\le C_i\}\psi(Y_1\wedge C_1,\ldots,Y_m\wedge C_m)}{\prod_{i=1}^m\{1-G(Y_i\wedge C_i)\}}\mid(\mathbf{X}_1,\ldots,\mathbf{X}_m)=\mathbf{t}\right) \\ &= \mathbb{E}\left(\frac{\psi(Y_1,\ldots,Y_m)}{\prod_{i=1}^m\{1-G(Y_i)\}}\mathbb{E}\left(\prod_{i=1}^m\{\mathbb{1}\{Y_i\le C_i\}\mid(Y_i,X_i)_{1\le i\le m}\right)\mid(\mathbf{X}_1,\ldots,\mathbf{X}_m)=\mathbf{t}\right) \\ &= \mathbb{E}\left(\psi(Y_1,\ldots,Y_m)\mid(\mathbf{X}_1,\ldots,\mathbf{X}_m)=\mathbf{t}\right)\end{aligned}$$

An analog estimator to (1.2) in the censored case is given by

$$\breve{\mathbf{r}}_n^{(m)}(\psi,\mathbf{t};h_n)=\sum_{(i_1,\ldots,i_m)\in I(m,n)}\frac{\delta_{i_1}\cdots\delta_{i_m}\psi(Z_{i_1},\ldots,Z_{i_m})}{(1-G(Z_{i_1})\cdots(1-G(Z_{i_m}))}\overline{\omega}_{n,K,h,\mathbf{i}}^{(m)}(\mathbf{t}), \tag{5.7}$$

where, for $\mathbf{i}=(i_1,\ldots,i_m)\in I(m,n)$,

$$\overline{\omega}_{n,K,h,\mathbf{i}}^{(m)}(\mathbf{t}):=\frac{K\left(\frac{\mathbf{t}_1-\mathbf{X}_{i_1}}{h_n}\right)\cdots K\left(\frac{\mathbf{t}_m-\mathbf{X}_{i_m}}{h_n}\right)}{\sum_{(i_1,\ldots,i_m)\in I(m,n)}K\left(\frac{\mathbf{t}_1-\mathbf{X}_{i_1}}{h_n}\right)\cdots K\left(\frac{\mathbf{t}_m-\mathbf{X}_{i_m}}{h_n}\right)}. \tag{5.8}$$

The estimator that we will investigate is given by

$$\breve{\mathbf{r}}_n^{(m)*}(\psi,\mathbf{t};h_n)=\sum_{(i_1,\ldots,i_m)\in I(m,n)}\frac{\delta_{i_1}\cdots\delta_{i_m}\psi(Z_{i_1},\ldots,Z_{i_m})}{(1-G_n^*(Z_{i_1})\cdots(1-G_n^*(Z_{i_m}))}\overline{\omega}_{n,K,h,\mathbf{i}}^{(m)}(\mathbf{t}). \tag{5.9}$$

In the copula representation, this is given by

$$\begin{aligned}&\breve{r}_n^{(m)}(\psi,\mathbf{t};h_n) \\ &= \frac{(n-m)!}{n!(n-1)^m}\sum_{\substack{(i_1,\ldots,i_m),(\ell_1,\ldots,\ell_m)\in I(m,n)\\ i_k\neq\ell_k,k=1,\ldots,m}}\frac{\delta_{i_1}\cdots\delta_{i_m}\psi(Z_{i_1},\ldots,Z_{i_m})}{(1-G(Z_{i_1})\cdots(1-G(Z_{i_m}))}\mathbf{W}(\mathbf{i},\boldsymbol{\ell},\mathbf{t};h_n),\end{aligned}$$

and

$$\begin{aligned}&\breve{r}_n^{(m)*}(\psi,\mathbf{t};h_n) \\ &= \frac{(n-m)!}{n!(n-1)^m}\sum_{\substack{(i_1,\ldots,i_m),(\ell_1,\ldots,\ell_m)\in I(m,n)\\ i_k\neq\ell_k,k=1,\ldots,m}}\frac{\delta_{i_1}\cdots\delta_{i_m}\psi(Z_{i_1},\ldots,Z_{i_m})}{(1-G_n^*(Z_{i_1})\cdots(1-G_n^*(Z_{i_m}))}\mathbf{W}(\mathbf{i},\boldsymbol{\ell},\mathbf{t};h_n).\end{aligned}$$

Let

$$\begin{aligned}
&\widetilde{\mathbb{E}}\breve{r}_n^{(m)*}(\varphi, \mathbf{t}; h_n) \\
&= \mathbb{E}\left\{\frac{1}{h_n^{m(p+q)}} \frac{\psi(Z_{i_1}, \ldots, Z_{i_m})}{(1 - G_n^*(Z_{i_1}) \cdots (1 - G_n^*(Z_{i_m}))} \right. \\
&\left. \prod_{j=1}^{m} K\left(\frac{\mathbf{F}_n(t_j) - \mathbf{F}_n(X_j)}{h_n}\right) K\left(\frac{\mathbf{F}_{0,n}(Y_{m+1}) - \mathbf{F}_{0,n}(Y_j)}{h_n}\right)\right\} \\
&\times \mathbb{E}\left\{\frac{1}{h_n^{mp}} \prod_{j=1}^{m} K\left(\frac{\mathbf{F}_n(t_j) - \mathbf{F}_n(X_{i_j})}{h_n}\right)\right\}^{-1}.
\end{aligned}$$

we will let $h > 0$ vary in such a way that $h_n' \leq h \leq h_n''$, where $\{h_n'\}_{n\geq 1}$ and $\{h_n''\}_{n\geq 1}$ are two sequences of positive constants such that $0 < h_n' \leq h_n'' < \infty$ and, for either choice of $h_n = h_n'$ or $h_n = h_n''$, conditions (H .1-2-3) below are fulfilled by $\{h_n\}_{n\geq 1}$

(H.1) $h_n \downarrow 0, 0 < h_n < 1$, and $nh_n^d \uparrow \infty$;
(H.2) $nh_n^d / \log n \to \infty$ as $n \to \infty$
(H.3) $\log(1/h_n) / \log\log n \to \infty$ as $n \to \infty$.

Assumptions (H.1)-(H.3) are classical in the empirical process theory and are often referred to as the Csörgő-Révész-Stute (CRS) conditions [48, 134]. They primarily allow the controlling of variance-type terms. The condition $h_n \downarrow 0$ is used to obtain the asymptotic unbiasedness of the kernel (density or regression) type estimators. We need a more restrictive assumption on h_n for the consistency, this is given by the condition $nh_n^d \uparrow \infty$; one can refer to [2, 112, 118].

Theorem 5.1. *Suppose that the copula density $\mathbf{c}(\cdot)$ is Lipschitz continuous on $[0,1]^{p+1}$ and let $a_n = \varrho(\log n/n)$ for $\varrho > 0$. If the class of functions $\mathscr{F} \cdot \mathscr{C}$ is bounded, in the sense that for some $0 < M < \infty$,*

$$\mathfrak{F}(y) < M.$$

We infer, under (A.1)-(A.3), (H.1)-(H.3), (I) and the above-mentioned assumptions on $\mathscr{F}$ and $\mathscr{K}$, that for all $\varrho > 0$ and $0 < b_0 < 1$, there exists a constant $0 < \Sigma'' < \infty$ such that

$$\limsup_{n\to\infty} \sup_{h_n' \leq h \leq h_n''} \sup_{\psi\in\mathscr{F}} \sup_{\mathbf{t}\in\mathbf{I}^p} \frac{\sqrt{nh^{p+1}}|\breve{r}_n^{(m)*}(\psi, \mathbf{t}; h_n) - \widetilde{\mathbb{E}}\breve{r}_n^{(m)*}(\psi, \mathbf{t}; h)|}{\sqrt{|\log h| \vee \log\log n}} \leq \Sigma'', \quad a.s.$$

The proof of Theorem 5.1 is postponed until Section 10.
The proof of this theorem, when combined with [56] results, gives the following

$$\limsup_{n\to\infty} \sup_{h_n' \leq h \leq h_n''} \sup_{\psi\in\mathscr{F}} \sup_{\mathbf{t}\in\mathbf{I}^p} \frac{\sqrt{nh^{mp}}|\breve{\mathbf{r}}_n^{(m)*}(\psi, \mathbf{t}; h_n) - \widetilde{\mathbb{E}}\breve{\mathbf{r}}_n^{(m)*}(\psi, \mathbf{t}; h)|}{\sqrt{|\log h| \vee \log\log n}} \leq \breve{\breve{\Sigma}}, \quad a.s.$$

for some constant $0 < \breve{\breve{\Sigma}} < \infty$. Refer to [30] for recent references.

6. Potential applications

6.1. Simultaneous prediction intervals for random forests

This example is given in [131]. Consider a training dataset of size n,

$$\{(Y_1, Z_1), \ldots, (Y_n, Z_n)\} = \{X_1, \ldots, X_n\} = X_1^n,$$

where $Y_i \in \mathcal{Y}$ is a vector of features and $Z_i \in \mathbb{R}$ is a response. Let h be a deterministic prediction rule that takes a sub-sample $\{X_{i_1}, \ldots, X_{i_m}\}$ with $1 \leq m \leq n$ as input and outputs predictions on d testing points $(y_1^*, \ldots, y_d^*)$ in the feature space $\mathcal{Y}$. The tree-based prediction rule is constructed on each sub-sample with additional randomness for random forests. Specifically, let $\{W_\iota : \iota \in I(m,n)\}$ be a collection of i.i.d. random variables taking value in a measurable space $(S', \mathcal{S}')$ that are independent of the data X_1^n. Let $H : S^m \times S' \to \mathbb{R}^d$ be an $\mathcal{S}^m \otimes \mathcal{S}'$-measurable function such that $\mathbb{E}[H(x_1, \ldots, x_m, W)] = h(x_1, \ldots, x_m)$. Then predictions of random forests are given by a d-dimensional U-statistic with random kernel H:

$$\widehat{U}_n := \frac{(n-m)!}{n!} \sum_{\mathbf{i} \in I(m,n)} H(X_{i_1}, \ldots, X_{i_m}, W_{\mathbf{i}}). \tag{6.1}$$

where the random kernel H varies with m.

6.2. Discrimination

Now, we apply the results of the problem of discrimination described in Section 3 of [138], refer to also to [137]. We will use a similar notation and setting. Let $\varphi(\cdot)$ be any function taking at most finitely many values, say $1, \ldots, M$. The sets

$$A_j = \{(y_1, \ldots, y_m) : \varphi(y_1, \ldots, y_k) = j\}, \quad 1 \leq j \leq M$$

then yield a partition of the feature space. Predicting the value of $\varphi(Y_1, \ldots, Y_m)$ is tantamount to predicting the set in the partition to which $(Y_1, \ldots, Y_m)$ belongs. For any discrimination rule g, we have

$$\mathbb{P}(g(\mathbf{X}) = \varphi(\mathbf{Y})) \leq \sum_{j=1}^{M} \int_{\{\mathbf{x}: g(\mathbf{x}) = j\}} \max m^j(\mathbf{x}) d\mathbf{x},$$

where

$$m^j(\mathbf{x}) = \mathbb{P}(\varphi(\mathbf{Y}) = j \mid \mathbf{X} = \mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d.$$

The above inequality becomes equality if

$$g_0(\mathbf{x}) = \arg \max_{1 \leq j \leq M} m^j(\mathbf{x}).$$

$g_0(\cdot)$ is called the Bayes rule, and the pertaining probability of error

$$\mathbf{L}^* = 1 - \mathbb{P}(g_0(\mathbf{X}) = \varphi(\mathbf{Y})) = 1 - \mathbb{E}\left\{\max_{1 \leq j \leq M} m^j(\mathbf{x})\right\}$$

is called the Bayes risk. Each of the above unknown function m^j's can be consistently estimated by one of the methods discussed in Section 2. Let, for $1 \leq j \leq M$,

$$m_n^j(\mathbf{x}) = \frac{(n-m)!}{n!(n-1)^m} \sum_{\substack{(i_1, \ldots, i_m), (\ell_1, \ldots, \ell_m) \in I(m,n) \\ i_k \neq \ell_k, k = 1, \ldots, m}} \mathbb{1}\{\varphi(Y_{i_1}, \ldots, Y_{i_m}) = j\} \mathbf{W}(\mathbf{i}, \boldsymbol{\ell}, \mathbf{t}; h_n).$$

Set

$$g_{0,n}(\mathbf{x}) = \arg \max_{1 \leq j \leq M} m_n^j(\mathbf{x}).$$

Let us introduce

$$\mathbf{L}_n^* = \mathbb{P}(g_{0,n}(\mathbf{X}) \neq \varphi(\mathbf{Y})).$$

Then, one can show that the discrimination rule $g_{0,n}(\cdot)$ is asymptotically Bayes' risk consistent

$$\mathbf{L}_n^* \to \mathbf{L}^*.$$

This follows from the obvious relation

$$|\mathbf{L}^* - \mathbf{L}_n^*| \leq 2\mathbb{E}\left[\max_{1 \leq j \leq M} \left| m_n^j(\boldsymbol{X}) - m^j(\boldsymbol{X}) \right| \right].$$

6.3. Ranking problems

For its great importance, the problem of ranking instances has received special attention in machine learning. In some specific ranking problems, it is necessary to compare two different observations based on their observed characteristics and decide which one is better instead of simply classifying them. The ordering problems have many applications in different areas of banking (Data mining process for direct marketing data extraction), document type classification and so on. The problems of ordering/ranking are frequent problems in which U-statistics come into play. In this challenge, the aim is to establish a universal and consistent ordering method. Suppose that we want to establish an order between the first components of the two pairs (X, Y), (X', Y') of independent and identically distributed observations in $\mathcal{X} \times \mathbb{R}$. The variables Y and Y' are respective labels of the variables X and X' that we want to order by observing them (and not their labels). Usually, we decide that X is better than X' if $Y > Y'$. To see things more clearly, we introduce the new variable:

$$Z = \frac{Y - Y'}{2},$$

then $Y > Y'$ is equivalent to $Z > 0$. As mentioned, the goal is to establish a classification rule between X and X' with minimal risk, i.e., the probability that the label of the highest ranked variable is the smallest, is small. Mathematically speaking, the decision rule is given by the function:

$$r(x, x') = \begin{cases} 1 & \text{if} \quad x > x', \\ -1 & \qquad\quad \text{else.} \end{cases}$$

The following ranking risk gives the performance measure of r:

$$L(r) = \mathbb{P}\left(Z.r\left(X, X'\right)\right).$$

A natural estimate for $L(\cdot)$ according to [47] is:

$$L_n(r) := \frac{1}{n(n-1)} \sum_{i \neq j} \mathbb{1}_{\{Z_{i,j}.r(X_i, X_j) < 0\}},$$

where $(X_1, Y_1), \ldots, (X_n, Y_n)$ are n independent, identically distributed copies of (X, Y), and $Z_{i,j} = \frac{Y_i - Y_j}{2}$. One can easily see that L_n is a U-statistic with $m = 2$. For more details the reader is invited to consult [47] and [117].

7. Examples

7.1. Examples of classes of functions

Example 7.1. The set $\mathscr{F}$ of all indicator functions $\mathbb{I}_{\{(-\infty, t]\}}$ of cells in $\mathbb{R}$ satisfies :

$$N\left(\epsilon, \mathscr{F}, d_{\mathbb{P}}^{(2)}\right) \leq \frac{2}{\epsilon^2},$$

for any probability measure $\mathbb{P}$ and $\epsilon \leq 1$. Notice that :

$$\int_0^1 \sqrt{\log\left(\frac{1}{\epsilon}\right)} d\epsilon \leq \int_0^\infty u^{1/2} \exp(-u) du \leq 1.$$

For more details and discussion on this example, refer to Example 2.5.4 of [142] and [97, p. 157]. The covering numbers of the class of cells $(-\infty, t]$ in higher dimension satisfy a similar bound, but with higher power of $(1/\epsilon)$, see Theorem 9.19 of [97].

Example 7.2. (*Classes of functions that are Lipschitz in a parameter*, Section 2.7.4 in [142]). Let $\mathscr{F}$ be the class of functions $x \mapsto \varphi(t, x)$ that are Lipschitz in the index parameter $t \in T$. Suppose that:

$$|\varphi(t_1, x) - \varphi(t_2, x)| \leq d(t_1, t_2)\kappa(x)$$

for some metric d on the index set T, the function $\kappa(\cdot)$ defined on the sample space $\mathcal{X}$, and all x. According to Theorem 2.7.11 of [142] and Lemma 9.18 of [97], it follows, for any norm $\|\cdot\|_{\mathscr{F}}$ on $\mathscr{F}$, that :

$$N(\epsilon\|F\|_{\mathscr{F}}, \mathscr{F}, \|\cdot\|_{\mathscr{F}}) \leq N(\epsilon/2, T, d).$$

Hence if (T, d) satisfy

$$J(\infty, T, d) = \int_0^\infty \sqrt{\log N(\epsilon, T, d)} d\epsilon < \infty,$$

then the conclusions holds for $\mathscr{F}$.

Example 7.3. Let us consider as an example the classes of functions that are smooth up to order α defined as follows, see Section 2.7.1 of [142] and Section 2 of [141]. For $0 < \alpha < \infty$ let $\lfloor\alpha\rfloor$ be the greatest integer strictly smaller than α. For any vector $k = (k_1, \ldots, k_d)$ of d integers define the differential operator

$$D^{k.} := \frac{\partial^{k.}}{\partial^{k_1} \ldots \partial^{k_d}},$$

where

$$k_. := \sum_{i=1}^d k_i.$$

Then, for a function $f : \mathcal{X} \to \mathbb{R}$, let

$$\|f\|_\alpha := \max_{k. \leq \lfloor\alpha\rfloor} \sup_x |D^k f(x)| + \max_{k. = \lfloor\alpha\rfloor} \sup_x \frac{D^k f(x) - D^k f(y)}{\|x - y\|^{\alpha - \lfloor\alpha\rfloor}},$$

where the suprema are taken over all x, y in the interior of $\mathcal{X}$ with $x \neq y$. Let $C_M^\alpha(\mathcal{X})$ be the set of all continuous functions $f : \mathcal{X} \to \mathbb{R}$ with

$$\|f\|_\alpha \leq M.$$

Note that for $\alpha \leq 1$ this class consists of bounded functions f that satisfy a Lipschitz condition. Kolmogorov and Tihomirov [96] computed the entropy of the classes of $C_M^\alpha(\mathcal{X})$ for the uniform norm. As a consequence of their results van der Vaart [141] shows that there exists a constant K depending only on α, d and the diameter of $\mathcal{X}$ such that for every measure γ and every $\epsilon > 0$,

$$\log \mathcal{N}_{[\,]}(\epsilon M \gamma(\mathcal{X}), C_M^\alpha(\mathcal{X}), L_2(\gamma)) \leq K \left(\frac{1}{\epsilon}\right)^{d/\alpha},$$

$\mathcal{N}_{[\,]}$ is the bracketing number, refer to Definition 2.1.6 of [142] and we refer to Theorem 2.7.1 of [142] for a variant of the last inequality. By Lemma 9.18 of [97], we have

$$\log \mathcal{N}(\epsilon M \gamma(\mathcal{X}), C_M^\alpha(\mathcal{X}), L_2(\gamma)) \leq K \left(\frac{1}{2\epsilon}\right)^{d/\alpha}.$$

7.2. Examples of U-kernels

Example 7.4. For :

$$\varphi(Y_1, Y_2) = \frac{1}{2}(Y_1 - Y_2)^2,$$

we obtain :

$$r^{(2)}(\varphi, t_1) = \operatorname{Var}(Y_1 \mid X_1 = t_1).$$

Example 7.5. Let $\varphi(Y_1, Y_2) = Y_1 Y_2$, then :

$$\begin{aligned} r^{(2)}(\varphi, t_1, t_2) &= \mathbb{E}(Y_1 Y_2 \mid X_1 = t_1, X_2 = t_2) \\ &= \mathbb{E}(Y_1 \mid X_1 = t_1)\mathbb{E}(Y_2 \mid X_2 = t_2) \\ &= \overline{r}^{(1)}(t_1)\overline{r}^{(1)}(t_2), \end{aligned}$$

where $\overline{r}^{(2)}$ denoting the regression of $\mathbf{Y}$ on $\mathbf{X} = \mathbf{t}$. The above $\varphi(\cdot)$ is a simple example of the kernel for a conditional U-statistic where one is interested in functions of $\overline{r}^{(2)}$.

Example 7.6. Let :

$$\psi(Y_1, Y_2, Y_3) = \mathbb{1}\{Y_2 \le Y_1\} - \mathbb{1}\{Y_3 \le Y_1\}$$

and for $m = 5$ define :

$$\varphi(Y_1, \ldots, Y_5) = \frac{1}{4}\psi(Y_1, Y_2, Y_3)^2\psi(Y_1, Y_4, Y_5)^2.$$

We have :

$$\begin{aligned} & r^{(5)}\left(\varphi, t_1, t_2, t_3, t_4, t_5\right) \\ &= \mathbb{E}\left(\varphi(Y_1, \ldots, Y_5) \mid X_1 = X_2 = X_3 = X_4 = X_5 = t\right). \end{aligned}$$

The corresponding U-statistics may be used to test the conditional independence.

Example 7.7. For $\varphi(Y_1, Y_2) = \mathbb{1}\{Y_1 \le Y_2\}$:

$$r^{(2)}(\varphi, t_1, t_2) = \mathbb{P}(Y_1 \le Y_2 \mid X_1 = t_1, X_2 = t_2), \quad \text{for} \quad t_1 \ne t_2$$

equals the probability that the output pertaining to t_1 is less than or equal to the one pertaining to t_2.

Example 7.8. Assume $\left\{\mathbf{Y}_i = (Y_{i,1}, Y_{i,2})^\top\right\}_{i=1,2}$ and define φ by :

$$\varphi(\mathbf{y}_1, \mathbf{y}_2) := \frac{1}{2}(y_{1,1}y_{1,2} + y_{2,1}y_{2,2} - y_{1,1}y_{2,2} - y_{1,2}y_{2,1}),$$

and :

$$\begin{aligned} r^{(2)}(\varphi, t_1, t_2) &= \frac{1}{2}\left\{\mathbb{E}(Y_{1,1}Y_{1,2} \mid X_1 = t_1) + \mathbb{E}(Y_{2,1}Y_{2,2} \mid X_2 = t_2)\right. \\ &\quad \left. - \mathbb{E}(Y_{1,1}Y_{2,2} \mid X_1 = t_1, X_2 = t_2) - \mathbb{E}(Y_{1,2}Y_{2,1} \mid X_1 = t_1, X_2 = t_2)\right\}. \end{aligned}$$

In particular :

$$r^{(2)}(\varphi, t_1) = \mathbb{E}(Y_{1,1}Y_{1,2} \mid X_1 = t_1) - \mathbb{E}(Y_{1,1} \mid X_1 = t_1)\mathbb{E}(Y_{1,2} \mid X_1 = t_1)$$

is the conditional covariance of Y_1 given $X_1 = t_1$.

Example 7.9. For $m = 3$, let :

$$\varphi(Y_1, Y_2, Y_3) = \mathbb{1}\{Y_1 - Y_2 - Y_3 > 0\},$$

We have

$$r^{(3)}(\varphi, t_1, t_2, t_3) = \mathbb{P}(Y_1 > Y_2 + Y_3 \mid X_1 = X_2 = X_3 = t)$$

and the corresponding conditional U-Statistic can be looked upon as a conditional analog of the Hollander-Proschan test-statistic [85]. It may be used to test the hypothesis that the conditional distribution of Y_1 given $X_1 = t$, is exponential, against the alternative that it is of the New-Better than-Used-type.

Example 7.10. Let $\widehat{Y_1Y_2}$ denote the oriented angle between $Y_1, Y_2 \in T$, T is the circle of radius 1 and center 0 in $\mathbb{R}^2$. Let :

$$\varphi_t(Y_1, Y_2) = \mathbb{1}\{\widehat{Y_1Y_2} \le t\} - t/\pi, \quad \text{for} \quad t \in [0, \pi).$$

Silverman [127] has used this kernel in order to propose U-process to test uniformity on the circle. Let

$$r^{(2)}(\varphi_t, t_1, t_2) = \mathbb{E}(\varphi_t(Y_1, Y_2) \mid X_1 = X_2 = t).$$

In this setting, one can propose a conditional U-process to test conditional uniformity on the circle.

Example 7.11. Hoeffding [84] introduced the parameter

$$\triangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} D^2(y_1, y_2)dF(y_1, y_2),$$

where $D(y_1, y_2) = F(y_1, y_2) - F(y_1, \infty)F(\infty, y_2)$ and $F(\cdot, \cdot)$ is the distribution function of Y_1 and Y_2. The parameter $\triangle$ has the property that $\triangle = 0$ if and only if Y_1 and Y_2 are independent. From [99], an alternative expression for $\triangle$ can be developed by introducing the functions

$$\psi(y_1, y_2, y_3) = \begin{cases} 1 & \text{if } y_2 \le y_1 < y_3 \\ 0 & \text{if } y_1 < y_2, y_3 \text{ or } y_1 \ge y_2, y_3 \\ -1 & \text{if } y_3 \le y_1 < y_2 \end{cases}$$

and

$$\begin{aligned}\varphi(y_{1,1}, y_{1,2}, \ldots, y_{5,1}, y_{5,2}) &= \frac{1}{4}\psi(y_{1,1}, y_{1,2}, y_{1,3})\,\psi(y_{1,1}, y_{1,4}, y_{1,5}) \\ &\quad \times\psi(y_{1,2}, y_{2,2}, y_{3,2})\,\psi(y_{1,2}, y_{4,2}, y_{5,2}).\end{aligned}$$

We have

$$\triangle = \int\ldots\int \varphi(y_{1,1}, y_{1,2}, \ldots, y_{5,1}, y_{5,2})\,dF(y_{1,1}, y_{1,2})\ldots dF(y_{1,5}, y_{2,5}).$$

We have

$$\begin{aligned}& r^{(5)}(\varphi, t_1, t_2, t_3, t_4, t_5) \\ &= \mathbb{E}(\varphi((Y_{1,1}, Y_{1,2}), \ldots, (Y_{5,1}, Y_{5,2})) \mid X_1 = X_2 = X_3 = X_4 = X_5 = t).\end{aligned}$$

The corresponding U-statistics may be used to test the conditional independence.

Example 7.12. (Hoeffding's D). From the symmetric kernel,

$$\begin{aligned}& h_D(z_1, \ldots, z_5) \\ &:= \frac{1}{16}\sum_{(i_1,\ldots,i_5)\in\mathcal{P}_5} [\{\mathbb{1}(z_{i_1,1} \le z_{i_5,1}) - \mathbb{1}(z_{i_2,1} \le z_{i_5,1})\}\{\mathbb{1}(z_{i_3,1} \le z_{i_5,1}) - \mathbb{1}(z_{i_4,1} \le z_{i_5,1})\}] \\ &\quad \times [\{\mathbb{1}(z_{i_1,2} \le z_{i_5,2}) - \mathbb{1}(z_{i_2,2} \le z_{i_5,2})\}\{\mathbb{1}(z_{i_3,2} \le z_{i_5,2}) - \mathbb{1}(z_{i_4,2} \le z_{i_5,2})\}],\end{aligned}$$

we recover Hoeffding's D statistic, which is a rank-based U-statistic of order 5 and gives rise to Hoeffding's D correlation measure $\mathbb{E}h_D$.

Example 7.13. (Blum-Kiefer-Rosenblatt's R). The symmetric kernel

$$\begin{aligned}& h_R(z_1, \ldots, z_6) := \frac{1}{32}\sum_{(i_1,\ldots,i_6)\in\mathcal{P}_6} \\ &\times [\{\mathbb{1}(z_{i_1,1} \le z_{i_5,1}) - \mathbb{1}(z_{i_2,1} \le z_{i_5,1})\}\{\mathbb{1}(z_{i_3,1} \le z_{i_5,1}) - \mathbb{1}(z_{i_4,1} \le z_{i_5,1})\}] \\ &\times [\{\mathbb{1}(z_{i_1,2} \le z_{i_6,2}) - \mathbb{1}(z_{i_2,2} \le z_{i_6,2})\}\{\mathbb{1}(z_{i_3,2} \le z_{i_6,2}) - \mathbb{1}(z_{i_4,2} \le z_{i_6,2})\}]\end{aligned}$$

yields Blum-Kiefer-Rosenblatt's R statistic [8], which is a rank-based U-statistic of order 6. At this point, we refer to [10, 12, 15, 24, 25].

Example 7.14. (Bergsma-Dassios-Yanagimoto's τ^*). Bergsma and Dassios [7] introduced a rank correlation statistic as a U-statistic of order 4 with the symmetric kernel

$$\begin{aligned}
&h_{\tau^*}(z_1, \ldots, z_4)\\
&:=\frac{1}{16} \sum_{(i_1,\ldots,i_4)\in\mathcal{P}_4} \{\mathbb{1}(z_{i_1,1}, z_{i_3,1} < z_{i_2,1}, z_{i_4,1}) + \mathbb{1}(z_{i_2,1}, z_{i_4,1} < z_{i_1,1}, z_{i_3,1})\\
&\quad -\mathbb{1}(z_{i_1,1}, z_{i_4,1} < z_{i_2,1}, z_{i_3,1}) - \mathbb{1}(z_{i_2,1}, z_{i_3,1} < z_{i_1,1}, z_{i_4,1})\}\\
&\quad \times\{\mathbb{1}(z_{i_1,2}, z_{i_3,2} < z_{i_2,2}, z_{i_4,2}) + \mathbb{1}(z_{i_2,2}, z_{i_4,2} < z_{i_1,2}, z_{i_3,2})\\
&\quad -\mathbb{1}(z_{i_1,2}, z_{i_4,2} < z_{i_2,2}, z_{i_3,2}) - \mathbb{1}(z_{i_2,2}, z_{i_3,2} < z_{i_1,2}, z_{i_4,2})\}
\end{aligned}$$

Here, $\mathbb{1}(y_1, y_2 < y_3, y_4) := \mathbb{1}(y_1 < y_3)\,\mathbb{1}(y_1 < y_4)\,\mathbb{1}(y_2 < y_3)\,\mathbb{1}(y_2 < y_4)$.

Example 7.15. The sample covariance matrix

$$\hat{S}_n = (n-1)^{-1} \sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)\left(X_i - \bar{X}_n\right)^\top,$$

is an unbiased estimator of the covariance matrix $\Sigma = \operatorname{Cov}(X_1)$. Here, $\hat{S}_n$ is a matrix-valued U-statistic with the quadratic kernel $h(x_1, x_2) = (x_1 - x_2)(x_1 - x_2)^\top/2$ for $x_1, x_2 \in \mathbb{R}^p$.

Example 7.16. Two generic vectors $y = (y_1, y_2)$ and $z = (z_1, z_2)$ in $\mathbb{R}^2$ are said to be concordant if $(y_1 - z_1)(y_2 - z_2) > 0$. For $m, k = 1, \ldots, p$, define

$$\tau_{\ell k} = \frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} \mathbb{1}\{(X_{i\ell} - X_{j\ell})(X_{ik} - X_{jk}) > 0\}.$$

Then Kendall's tau rank correlation coefficient matrix $T = \{\tau_{\ell k}\}_{\ell,k=1}^{p}$ is a matrix-valued U-statistic with a bounded kernel. It is clear that $\tau_{\ell k}$ quantifies the monotonic dependency between $(X_{1\ell}, X_{1k})$ and $(X_{2\ell}, X_{2k})$ and it is an unbiased estimator of $\mathbb{P}((X_{1\ell} - X_{2\ell})(X_{1k} - X_{2k}) > 0)$, that is, the probability that $(X_{1\ell}, X_{1k})$ and $(X_{2\ell}, X_{2k})$ are concordant.

Example 7.17. The Gini mean difference. The Gini index provides another popular measure of dispersion. It corresponds to the case where $E \subset \mathbb{R}$ and $h(x, y) = |x - y|$:

$$G_n = \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} |X_i - X_j|$$

Example 7.18. The Wilcoxon Statistic. Suppose that $E \subset \mathbb{R}$ is symmetric around zero. As an estimate of the quantity

$$\int_{(x,y)\in E^2} \left\{2\mathbb{1}_{\{x+y>0\}} - 1\right\} dF(x)dF(y),$$

it is pertinent to consider the statistic

$$W_n = \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} \left\{2\cdot\mathbb{1}_{\{X_i+X_j>0\}} - 1\right\},$$

which is relevant for testing whether or not μ is located at zero.

Example 7.19. The Takens estimator. Suppose that $E \subset \mathbb{R}^d, d \geq 1$. Denote by $\|\cdot\|$ the usual Euclidean norm on $\mathbb{R}^d$. In [9], the following estimate of the correlation integral,

$$C_F(r) = \int \mathbb{I}_{\{\|x-x'\|\leq r\}} dF(x)dF(x'), \quad r > 0,$$

is considered:

$$C_n(r) = \frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} \mathbb{I}_{\{\|X_i-X_j\|\leq r\}}$$

In the case where a scaling law holds for the correlation integral, i.e., when there exists $(\alpha, r_0, c) \in \mathbb{R}_+^{*3}$ such that $C_F(r) = c \cdot r^{-\alpha}$ for $0 < r \leq r_0$, the U-statistic

$$T_n = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \log\left(\frac{\|X_i - X_j\|}{r_0}\right),$$

is used in order to build the Takens estimator $\hat{\alpha}_n = -T_n^{-1}$ of the correlation dimension α.

8. The bandwidth selection criterion

Many methods have been established and developed to construct, in asymptotically optimal ways, bandwidth selection rules for nonparametric kernel estimators, especially for Nadaraya-Watson regression estimator we quote among them [29, 56, 79, 116]. This parameter has to be selected suitably, either in the standard finite-dimensional case or in the infinite-dimensional framework to insuring good practical performances. The leave-one-out cross-validation procedure allows to define, for any fixed $\mathbf{i} = (i_1, \ldots, i_m) \in I(m, n)$:

$$\widetilde{r}_{n,\mathbf{i}}^{(m)}(\varphi, \mathbf{t}; h_n) = \frac{(n-m)!}{n!(n-1)^m} \sum_{\substack{(i_1,\ldots,i_m),(\ell_1,\ldots,\ell_m) \in I_n^m(\mathbf{i}) \\ i_k \neq \ell_k, k=1,\ldots,m}} \varphi(Y_{i_1}, \ldots, Y_{i_m}) \mathbf{W}^{(\mathbf{i})}(\mathbf{j}, \boldsymbol{\ell}, \mathbf{t}; h_n) \quad (8.1)$$

where

$$I_n^m(\mathbf{i}) := \{\mathbf{j} \in I(m,n) \text{ and } \mathbf{j} \neq \mathbf{i}\} = I(m,n) \backslash \{\mathbf{i}\},$$

and the weight, $(i_1, \ldots, i_m), (\ell_1, \ldots, \ell_m) \in I(m,n) \backslash \{\mathbf{i}\}$,

$$\mathbf{W}^{(\mathbf{i})}(\mathbf{j}, \boldsymbol{\ell}, \mathbf{t}; h_n) = \frac{\frac{1}{h_n^{m(p+q)}} \prod_{\nu=1}^{m} K\left(\frac{\mathbf{F}_n(t_\nu) - \mathbf{F}_n(X_{\ell_\nu})}{h_n}\right) K\left(\frac{\mathbf{F}_{0,n}(Y_{i_\nu}) - \mathbf{F}_{0,n}(Y_{\ell_\nu})}{h_n}\right)}{\frac{(n-m)!}{n! h_n^{mp}} \sum_{(i_1,\ldots,i_m) \in I(m,n)} \prod_{\nu=1}^{m} K\left(\frac{\mathbf{F}_n(t_\nu) - \mathbf{F}_n(X_{i_\nu})}{h_n}\right)}. \quad (8.2)$$

The equation (8.1) represents the leave-out-$(\mathbf{X_i}, \mathbf{Y_i})$ estimator of the functional regression and also could be considered as a predictor of $\varphi(\mathbf{Y_i})$. In order to minimize the quadratic loss function, we introduce the following criterion, we have for some (known) non-negative weight function $\mathcal{W}(\cdot)$:

$$CV(\varphi, h) := \frac{(n-m)!}{n!(n-1)^m} \sum_{\mathbf{i} \in I(m,n)} \left(\varphi(\mathbf{Y_i}) - \widetilde{r}_{n,\mathbf{i}}^{(m)}(\varphi, \mathbf{X_i}; h)\right)^2 \widetilde{\mathcal{W}}(\mathbf{X_i}), \quad (8.3)$$

where

$$\widetilde{\mathcal{W}}(\mathbf{t}) := \prod_{i=1}^{m} \mathcal{W}(t_i).$$

Following the ideas developed by [116], a natural way for choosing the bandwidth is to minimize the precedent criterion, so let's choose $\widehat{h}_n \in [a_n, b_n]$ minimizing among $h \in [a_n, b_n]$:

$$\sup_{\varphi \in \mathscr{F}} CV(\varphi, h),$$

we can conclude, by Corollary 2.1, that :

$$\sup_{\varphi \in \mathscr{F}} \sup_{\mathbf{t} \in \mathbf{I}^p} \left| \widehat{r}_n^{(m)}(\varphi, \mathbf{t}; \widehat{h}_n) - r^{(m)}(\varphi, \mathbf{t}) \right| \longrightarrow 0, \qquad \text{p.s.}$$

The main interest of our results is the possibility to derive the asymptotic properties of our estimate even if the bandwidth parameter is a random variable, like in the last equation. Following [6] where the bandwidths are locally chosen by a data-driven method based on the minimization of a functional version of a cross-validated criterion, one can replace (8.3) by

$$CV\left(\varphi, h\right) := \frac{(n-m)!}{n!(n-1)^m} \sum_{\mathbf{i} \in I(m,n)} \left(\varphi\left(\mathbf{Y_i}\right) - \widehat{r}_{n,\mathbf{i}}^{(m)}(\varphi, \mathbf{X_i}; h)\right)^2 \widehat{\mathcal{W}}\left(\mathbf{X_i}, \mathbf{t}\right), \tag{8.4}$$

where

$$\widehat{\mathcal{W}}\left(\mathbf{s}, \mathbf{t}\right) := \prod_{i=1}^{m} \widehat{W}(s_i, t_i).$$

In practice, one takes for $\mathbf{i} \in I(m,n)$, the uniform global weights $\widetilde{\mathcal{W}}\left(\mathbf{X_i}\right) = 1$, and the local weights

$$\widehat{W}(X_i, t) = \begin{cases} 1 & \text{if} \quad \|X_i - t\| \leq h, \\ 0 & \quad\text{otherwise.} \end{cases}$$

For the sake of brevity, we have just considered the most popular method, that is, the cross-validated selected bandwidth. This may be extended to any other bandwidth selector, such as the bandwidth based on Bayesian ideas [123].

Remark 8.1. We can use a different bandwidth criterion suggested by [128], the rule of thumb. Strictly speaking, since the cross-validated bandwidth m_n is random, the asymptotic theory can only be justified using a specific stochastic equicontinuity argument. For testing a parametric model for conditional mean function against a nonparametric alternative, Horowitz and Spokoiny [86] proposed an adaptive-rate-optimal rule. Gao and Gijbels [70] present the other method for selecting a proper bandwidth. Gao and Gijbels [70] propose, utilizing the Edgeworth expansion of the asymptotic distribution of the test, to select the bandwidth such that the power function of the test problem is maximized while the size function is controlled. Although any choice of bandwidth h_n that satisfies the assumption will produce the result in Corollary 2.1, we need guidance on choosing h_n in practice. Idealistically, we should choose a m_n that provides the greatest power (e.g., test based on the Kendall tau) or small MSE for a given sample size, but deriving this procedure is complicated enough to warrant a separate study.

9. Concluding remarks

In this work, we have considered the nonparametric conditional U-statistics estimation. Using the copula representation, we have proposed an alternative estimator of Stute's conditional U-statistics estimator, which is a generalization of the Nadaraya - Watson estimator. Similarly, our estimator is a direct extension of the copula regression estimator proposed in [23] and [22]. We have obtained the uniform-in-bandwidth consistency of the proposed estimator, which may be interesting in some applications. Copula regression models provide flexible tools, e.g., elliptical copulas have been used to combine generalized linear models for the components of auto or multi-peril homeowners insurance claims, Gaussian copula regression with Gamma margins to model whole-life and term insurance demand jointly. The copula models have to be used in cases when the errors are not jointly normally distributed; for example, in risk management, the marginal distributions of the error terms are far from normal. The conditions used in our analysis are close to those imposed in some papers dealing with the uniformity in bandwidth; for instance, refer [56]. As mentioned in the introduction, there are no restrictions on the choice of the kernel function, but copula density estimation has a crucial role. There are at least two reasons for this. First, copula densities are defined on bounded supports, but the standard kernel estimators are known to suffer from boundary biases. Second, the consistency of kernel density estimators requires that the underlying densities are bounded on their supports. However, many copula densities are unbounded at the boundaries. This unboundedness violates a key assumption of the kernel density estimation and renders it inconsistent. Since we need the copula density estimation as an intermediate step, it is important to consider the boundaries-related problems by using adapted kernels like beta kernels or wavelets. Another problem to be studied in the future is the characterization of the asymptotic properties of our estimator in the serially dependent setting. We would be interested in extending our work to k-nearest neighbours estimators. Presently it is beyond reasonable hope to achieve this program without new technical arguments. Another direction of research is to consider the projection pursuit regression and projection pursuit conditional distribution, which need an extension and generalization of the methods used in the present work. If we assume that the regression function $r^{(m)}(\varphi, \cdot)$ is smooth enough, that is differentiable at a fixed $\mathbf{t}_0$, it will be better to use the local polynomial regression techniques, refer to [64], to obtain a more appropriate estimate at $\mathbf{t}_0$ than that given by the Nadaraya-Watson type estimator. We will not treat the uniform consistency of such estimators in the present paper and leave for future investigation.

10. Mathematical developments

This section is devoted to the proof of our results. The previously presented notation continues to be used in the following.

Proof of Theorem 2.3

Let L be a kernel function of m variables, symmetric in its entries. Then, for $1 \leq k \leq m$, the Hoeffding projections with respect to $\mathbb{P}$ are defined as

$$\pi_k L(x_1, \ldots, x_k) = (\delta_{x_1} - \mathbb{P}) \times \cdots \times (\delta_{x_k} - \mathbb{P})\mathbb{P}^{m-k}(L)$$

with $\pi_0 L = \mathbb{E}L(X_1, \ldots, X_m)$, where

$$\mathbb{Q}_1 \cdots \mathbb{Q}_m L = \int \cdots \int L(x_1, \ldots, x_m) d\mathbb{Q}_1(x_1) \cdots d\mathbb{Q}_m(x_m).$$

For more details refer to [50]. If a function L is not necessarily symmetric, we will write $\mathbb{S}_m L$ for its symmetrization, that is,

$$\mathbb{S}_m L(x_1, \ldots, x_m) = \frac{1}{m!} \sum L(x_{\sigma(1)}, \ldots, x_{\sigma(m)}), \tag{10.1}$$

where the summation is over all permutations σ of $\{1, \ldots, m\}$. The [84] decomposition states the following, which is easy to check:

$$U_n^{(m)}(L) - \mathbb{E}L = \sum_{k=1}^{m} \binom{m}{k} U_n^{(k)}(\pi_k L),$$

where for a kernel L of k variables, $1 \leq k \leq m$, and we set

$$U_n^{(k)} = \frac{(n-k)!}{n!} \sum_{\mathbf{i} \in I(k,n)} L(X_{i_1}, \ldots, X_{i_k}).$$

Assuming L is in $L_2(\mathbb{P}^m)$, this is an orthogonal decomposition and

$$\mathbb{E}(\pi_k L \mid X_2, \ldots, X_k) = 0, \text{ for } k \geq 1,$$

that is, the kernels $\pi_k L$ are canonical for $\mathbb{P}$ (or completely degenerate, or completely centered). Also, $\pi_k, k \geq 1$, are nested projections, that is

$$\pi_k \circ \pi_\ell, \text{ if } k \leq \ell,$$

and

$$\mathbb{E}((\pi_k L)^2(X_1, \ldots, X_k)) \leq \mathbb{E}((L - \mathbb{E}L)^2(X_1, \ldots, X_m)) \leq \mathbb{E}L^2(X_1, \ldots, X_m).$$

Let us now recall the following definitions and notation from [3]. The reproducing kernel Hilbert space H_p associated with random variable $\delta_x - \mathbb{P}$ is the subspace $\ell_\infty(\mathscr{F})$ consisting of all functions of the form

$$u_h(f) = \mathbb{E}f(X)h(X), \text{ for } f \in \mathscr{F},$$

with h in the closed linear span in $L_2(\mathbb{P})$ of

$$\pi_{1,1}\mathscr{F} = \{f - \mathbb{P}f : f \in \mathscr{F}\}.$$

The unit ball of H_p, under the inner product $\langle u_{h_1}, u_{h_2}\mathbb{P}h_1h_2\rangle$ is

$$K_{\mathscr{F}} = \{u_h \in H_p : \|h\|_2 \leq 1\}.$$

Let $\mathbb{P}_n$ be the empirical measure associated with these random variables is defined as placing mass $1/n$ on each of the observations X_i, $i = 1, \ldots, n$, i.e.,

$$\mathbb{P}_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}.$$

Let $f : \mathcal{X} \mapsto \mathbb{R}$ be a measurable function. In the modern theory of the empirical, it is customary to identify $\mathbb{P}$ and $\mathbb{P}_n$ with the mappings given by

$$f \to \mathbb{P}f = \int_{\mathcal{X}} f d\mathbb{P}, \quad \text{and} \quad f \to \mathbb{P}_n f = \int_{\mathcal{X}} f d\mathbb{P}_n = \frac{1}{n}\sum_{k=1}^{n} f(\mathbf{X}_i).$$

For any class $\mathscr{F}$ of measurable functions $f : \mathcal{X} \mapsto \mathbb{R}$, an empirical process

$$\{\mathbb{G}_n f = \sqrt{n}\,(\mathbb{P}_n f - \mathbb{P}f) : f \in \mathscr{F}\}$$

can be defined. Then the law of iterated logarithm is said to hold for $\mathscr{F}$ if

$$\{(n/2\log\log n)^{1/2}(\mathbb{P}_n - \mathbb{P})f : f \in \mathscr{F}\}$$

is almost surely compact and its limit set is $K_{\mathscr{F}}$. In the U-process case, we say that $\mathscr{F}$ satisfies the law of iterated logarithm (LIL) if

$$\pi_1\mathscr{F} \text{ satisfies the LIL} \tag{10.2}$$

and

$$\lim_{n\to\infty}\left\{\frac{n}{\log\log n}\right\}^{1/2} \sup_{f\in\mathscr{F}} |U_n(\pi_k f)| = 0, \text{ a.s.} \quad \text{for } k = 2, \ldots, m. \tag{10.3}$$

Then

$$\{(n/2\log\log n)^{1/2}(U_n(f) - \mathbb{E}f) : f \in \mathscr{F}\}$$

is almost surely compact and its limit set is $K_{\pi_1\mathscr{F}}$. Let us introduce the following process

$$\mathbb{U}_n(\psi, \mathbf{t}) = \frac{(n-m)!}{n!}\sum_{(i_1,\ldots,i_m)\in I(m,n)} \varphi(Y_{i_1}, \ldots, Y_{i_m})\left\{\prod_{k=1}^{m} \mathbf{c}(\mathbf{F}(t_k), \mathbf{F}_0(Y_{i_k}))\right\}, \text{ for } \psi \in \mathscr{F}\cdot\mathscr{C}.$$

Under the condition of Theorem 2.2, making use of Theorem 2.2. of [3], we infer that

$$\{(n/2\log\log n)^{1/2}(\mathbb{U}_n(\psi, \mathbf{t}) - \mathbb{E}\psi) : \psi \in \mathscr{F}\cdot\mathscr{C}\}$$

is almost surely compact and its limit set is $K_{\pi_1\mathscr{F}\cdot\mathscr{C}}$. We have also

$$\sup_{\psi\in\mathscr{F}\cdot\mathscr{C}}\left\{\frac{n}{\log\log n}\right\}^{1/2}\left|\mathbb{U}_n(\psi, \mathbf{t}) - \mathbb{E}\psi - \frac{1}{n}\sum_{i=1}^{n}\pi_1\psi(Y_i)\right| = 0. \text{ a.s.}$$

For positive constant ϱ, we have

$$\sup_{\psi\in\mathscr{F}\cdot\mathscr{C}}\left\{\frac{n}{\log\log n}\right\}^{1/2} |\mathbb{U}_n(\psi, \mathbf{t}) - \mathbb{E}\psi| \leq \varrho. \tag{10.4}$$

Let us introduce the copula density estimator

$$\breve{\mathbf{c}}_{n;h_n}\left(\mathbf{F}_n(x)\right) := \frac{1}{nh_n^p}\sum_{i=1}^{n} K\left(\frac{\mathbf{F}_n(x)-\mathbf{F}_n(X_i)}{h_n}\right). \tag{10.5}$$

We define the following kernel estimate when the margins are known

$$\widetilde{\mathbf{c}}^{j}_{n;h_n}\left(\mathbf{F}(x),\mathbf{F}_0(y)\right)) := \frac{1}{nh_n^{p+q}}\sum_{i=1}^{n} K\left(\frac{\mathbf{F}(x)-\mathbf{F}(X_i)}{h_n}\right) K_0\left(\frac{\mathbf{F}_0(y)-\mathbf{F}_0(Y_i)}{h_n}\right), \tag{10.6}$$

$$\breve{\breve{\mathbf{c}}}_{n;h_n}\left(\mathbf{F}(x)\right) := \frac{1}{nh_n^p}\sum_{i=1}^{n} K\left(\frac{\mathbf{F}(x)-\mathbf{F}(X_i)}{h_n}\right). \tag{10.7}$$

The next step in our proof is to give bounds to the following terms

$$\breve{\Delta}^{(m)}_{1,n}(\mathbf{x}) = \left\{\prod_{k=1}^{m}\breve{\mathbf{c}}_{n;h_n}\left(\mathbf{F}_n(x_k)\right)\right\} - \left\{\prod_{k=1}^{m}\breve{\breve{\mathbf{c}}}_{n;h_n}\left(\mathbf{F}(x_k)\right)\right\}, \tag{10.8}$$

$$\breve{\Delta}^{(m)}_{2,n}(\mathbf{x}) = \left\{\prod_{k=1}^{m}\breve{\breve{\mathbf{c}}}_{n;h_n}\left(\mathbf{F}(x_k)\right)\right\} - \left\{\prod_{k=1}^{m}\mathbb{E}\breve{\breve{\mathbf{c}}}_{n;h_n}\left(\mathbf{F}(x_k)\right)\right\}, \tag{10.9}$$

$$\breve{\Delta}^{(m)}_{3,n}(\mathbf{x}) = \left\{\prod_{k=1}^{m}\mathbb{E}\breve{\breve{\mathbf{c}}}_{n;h_n}\left(\mathbf{F}(x_k)\right)\right\} - \left\{\prod_{k=1}^{m}\mathbf{c}(\mathbf{F}(x_k))\right\}. \tag{10.10}$$

Similarly, we will treat the following terms

$$\Delta^{(m)}_{1,n}(\mathbf{x},\mathbf{y}) = \left\{\prod_{k=1}^{m}\widehat{\mathbf{c}}_{n;h_n}\left(\mathbf{F}_n(x_k),\mathbf{F}_{n,0}(y_k)\right)\right\} - \left\{\prod_{k=1}^{m}\widetilde{\mathbf{c}}_{n;h_n}\left(\mathbf{F}(x_k),\mathbf{F}_0(y_k)\right)\right\}, \tag{10.11}$$

$$\Delta^{(m)}_{2,n}(\mathbf{x},\mathbf{y}) = \left\{\prod_{k=1}^{m}\widetilde{\mathbf{c}}_{n;h_n}\left(\mathbf{F}(x_k),\mathbf{F}_0(y_k)\right)\right\} - \left\{\prod_{k=1}^{m}\mathbb{E}\widetilde{\mathbf{c}}_{n;h_n}\left(\mathbf{F}(x_k),\mathbf{F}_0(y_k)\right)\right\}, \tag{10.12}$$

$$\Delta^{(m)}_{3,n}(\mathbf{x},\mathbf{y}) = \left\{\prod_{k=1}^{m}\mathbb{E}\widetilde{\mathbf{c}}_{n;h_n}\left(\mathbf{F}(x_k),\mathbf{F}_0(y_k)\right)\right\} - \left\{\prod_{k=1}^{m}\mathbf{c}(\mathbf{F}(x_k),\mathbf{F}_0(y_k))\right\}. \tag{10.13}$$

Notice that the terms $\breve{\Delta}^{(m)}_{3,n}(\mathbf{x})$ and $\boldsymbol{\Delta}^{(m)}_{3,n}(\mathbf{t},\mathbf{y})$ will be used in the evaluation of the bias term. Let us give the following that we need in the sequel.

Lemma 10.1. *[Telescoping] Let a_i, $i=1,\ldots,k$, b_i $i=1,\ldots,k$ be real number*

$$\prod_{i=1}^{k} a_i - \prod_{i=1}^{k} b_i = \sum_{i=1}^{k}(a_i-b_i)\prod_{j=1}^{i-1} b_j \prod_{h=1+i}^{k} a_h. \tag{10.14}$$

This can be easily verified (see, e.g., [73]). Firstly, according to [105, eq. (4.10)], we have, for any $j=1,\ldots,p$,

$$\limsup_{n\to\infty}\ \sup_{\{u,v:|F_j(u)-F_j(v)|\geq n^{-1}\log n\}}\left\{\sqrt{\frac{n}{2\log n}}\frac{|(F_{n,j}(u)-F_{n,j}(u))-(F_j(u)-F_j(v))|}{|F_j(u)-F_j(v)|^{1/2}}\right\} < \infty. \tag{10.15}$$

This, in turn, implies, for all large n, whenever

$$|F_j(u)-F_j(v)| \geq \frac{\log n}{n},$$

and, for some constant $D>0$, that

$$|(F_{n,j}(u)-F_{n,j}(u))-(F_j(u)-F_j(v))| \leq 2D|F_j(u)-F_j(v)|^{1/2}\sqrt{\frac{\log n}{n}}.$$

Notice that we can restrict ourselves to the set

$$\left\{u,v\in[0,1]:|F_j(u)-F_j(v)|\leq D'h\right\},\quad \text{for}\quad j=1,\ldots,p,\quad \text{and}\quad D'>1.$$

An application of the Chung law of the iterated logarithm (see, e.g. [46]), shows that, for each $j = 1, \ldots, d$, almost surely,

$$\limsup_{n\to\infty} \left\{ (\log\log n)^{-1/2} \sup_{u\in[0,1]} |\sqrt{n}(F_{n,j}(u) - F_j(u))| \right\} = 2^{-1/2}. \tag{10.16}$$

Let

$$\begin{aligned}
\mathfrak{W}_n(\mathbf{F}(u), \mathbf{F}(v)) &= \begin{pmatrix} (F_{n,1}(u_1) - F(u_1)) - (F_{n,1}(v_1) - F(v_1)) \\ \vdots \\ \vdots \\ (F_{n,p}(u_p) - F(u_p)) - (F_{n,p}(v_p) - F(v_p)) \end{pmatrix}, \\
\nabla K(u) &= \begin{pmatrix} \frac{\partial}{\partial u_1} K(u) \\ \vdots \\ \vdots \\ \frac{\partial}{\partial u_d} K(u) \end{pmatrix}, \\
\nabla^2 K(u) &= \begin{pmatrix} \frac{\partial^2}{\partial u_1^2} K(u) & \ldots & \frac{\partial^2}{\partial u_1 \partial u_d} K(u) \\ \vdots & \ldots & \vdots \\ \vdots & \ldots & \vdots \\ \frac{\partial^2}{\partial u_d \partial u_1} K(u) & \ldots & \frac{\partial^2}{\partial u_d^2} K(u) \end{pmatrix}.
\end{aligned}$$

By successive Taylor expansions, we readily obtain

$$\begin{aligned}
&\breve{\mathbf{c}}_{n;h_n}(\mathbf{F}_n(x)) - \breve{\breve{\mathbf{c}}}_{n;h_n}(\mathbf{F}(x)) \\
&= \frac{1}{nh_n^{p+1}} \sum_{i=1}^{n} \mathfrak{W}_n(\mathbf{F}(x), \mathbf{F}(X_i))^\top \nabla K\left(\frac{\mathbf{F}(x) - \mathbf{F}(X_i)}{h_n} \right) \\
&\quad + \frac{1}{nh_n^{p+2}} \sum_{i=1}^{n} \mathfrak{W}_n(\mathbf{F}(x), \mathbf{F}(X_i))^\top \nabla^2 K\left(\frac{\mathfrak{X}_i}{h_n} \right) \mathfrak{W}_n(\mathbf{F}(x), \mathbf{F}(X_i)),
\end{aligned} \tag{10.17}$$

where $\mathfrak{X}_i = (\mathfrak{X}_{i,1}, \ldots, \mathfrak{X}_{i,p})$, is a point between $\mathbf{F}(x) - \mathbf{F}(X_i)$ and $\mathbf{F}_n(x) - \mathbf{F}_n(X_i)$. Making use of (10.15) with (10.17), we obtain

$$\sup_{x\in\mathbb{R}^p} |\breve{\mathbf{c}}_{n;h_n}(\mathbf{F}_n(x)) - \breve{\breve{\mathbf{c}}}_{n;h_n}(\mathbf{F}(x))| = O\left(\sqrt{\frac{\log n}{n}} \frac{1}{h_n^{p+1/2}} \right). \tag{10.18}$$

By Lemma 10.1 and condition (K.i), we infer that we have

$$\begin{aligned}
\breve{\Delta}_{1,n}^{(m)}(\mathbf{x}) &= \left\{ \prod_{k=1}^{m} \breve{\mathbf{c}}_{n;h_n}(\mathbf{F}_n(x_k)) \right\} - \left\{ \prod_{k=1}^{m} \breve{\breve{\mathbf{c}}}_{n;h_n}(\mathbf{F}(x_k)) \right\} \\
&= \sum_{j=1}^{m} \left\{ \breve{\mathbf{c}}_{n;h_n}(\mathbf{F}_n(x_j)) - \breve{\breve{\mathbf{c}}}_{n;h_n}(\mathbf{F}(x_j)) \right\} \\
&\quad \times \left\{ \prod_{\ell=1}^{j-1} \breve{\breve{\mathbf{c}}}_{n;h_n}(\mathbf{F}(x_\ell)) \right\} \left\{ \prod_{s=j+1}^{m} \breve{\mathbf{c}}_{n;h_n}(\mathbf{F}(x_s)) \right\} \\
&\leq \kappa^{2m-2} \sum_{j=1}^{m} \left\{ \breve{\mathbf{c}}_{n;h_n}(\mathbf{F}_n(x_j)) - \breve{\breve{\mathbf{c}}}_{n;h_n}(\mathbf{F}(x_j)) \right\} \\
&\leq m\kappa^{2m-2} \max_{1\leq j\leq m} \left\{ \breve{\mathbf{c}}_{n;h_n}(\mathbf{F}_n(x_j)) - \breve{\breve{\mathbf{c}}}_{n;h_n}(\mathbf{F}(x_j)) \right\}.
\end{aligned} \tag{10.19}$$

By using (10.18), we infer that

$$\sup_{x\in\mathbb{R}^{mp}}\left|\breve{\Delta}_{1,n}^{(m)}(\mathbf{x})\right| = O\left(\sqrt{\frac{\log n}{n}}\frac{1}{h_n^{p+1/2}}\right). \tag{10.20}$$

By using similar arguments, we infer that we have

$$\begin{aligned}
\breve{\Delta}_{2,n}^{(m)}(\mathbf{x}) &= \left\{\prod_{k=1}^{m}\breve{\mathbf{c}}_{n;h_n}\left(\mathbf{F}(x_k)\right)\right\} - \left\{\prod_{k=1}^{m}\mathbb{E}\breve{\mathbf{c}}_{n;h_n}\left(\mathbf{F}(x_k)\right)\right\}\\
&= \sum_{j=1}^{m}\left\{\breve{\mathbf{c}}_{n;h_n}\left(\mathbf{F}(x_j)\right) - \mathbb{E}\breve{\mathbf{c}}_{n;h_n}\left(\mathbf{F}(x_j)\right)\right\}\\
&\quad\times\left\{\prod_{\ell=1}^{j-1}\breve{\mathbf{c}}_{n;h_n}\left(\mathbf{F}(x_\ell)\right)\right\}\left\{\prod_{s=j+1}^{m}\mathbb{E}\breve{\mathbf{c}}_{n;h_n}\left(\mathbf{F}(x_s)\right)\right\}\\
&\le \kappa^{2m-2}\sum_{j=1}^{m}\left\{\breve{\mathbf{c}}_{n;h_n}\left(\mathbf{F}(x_j)\right) - \mathbb{E}\breve{\mathbf{c}}_{n;h_n}\left(\mathbf{F}(x_j)\right)\right\}\\
&\le m\kappa^{2m-2}\max_{1\le j\le m}\left\{\breve{\mathbf{c}}_{n;h_n}\left(\mathbf{F}(x_j)\right) - \mathbb{E}\breve{\mathbf{c}}_{n;h_n}\left(\mathbf{F}(x_j)\right)\right\}.
\end{aligned} \tag{10.21}$$

By Theorem 1 of [62] or Theorem A in the Appendix, when $\breve{\mathbf{c}}(\cdot)$ is bounded, we have for each $\varrho > 0$, and for a suitable function $\Sigma(\varrho)$, with probability 1,

$$\limsup_{n\to\infty}\sup_{\varrho n^{-1}\log n\le h^p\le 1}\frac{\sqrt{nh^p}\|\breve{\mathbf{c}}_{n;h} - \mathbb{E}\breve{\mathbf{c}}_{n;h}\|_\infty}{\sqrt{\log(1/h^p)\vee\log\log n}} \le \Sigma(\varrho) < \infty. \tag{10.22}$$

This implies that

$$\limsup_{n\to\infty}\sup_{\varrho n^{-1}\log n\le h^p\le 1}\frac{\sqrt{nh^p}\|\breve{\Delta}_{2,n}^{(m)}\|_\infty}{\sqrt{\log(1/h^p)\vee\log\log n}} \le m\kappa^{2m-2}\Sigma(\varrho) < \infty. \tag{10.23}$$

We first bound the bias. Let $M > 1$, such that $\breve{\mathbf{c}}(u) < M$,

$$\begin{aligned}
\breve{\Delta}_{3,n}^{(m)}(\mathbf{x}) &= \left\{\prod_{k=1}^{m}\mathbb{E}\breve{\mathbf{c}}_{n;h_n}\left(\mathbf{F}(x_k)\right)\right\} - \left\{\prod_{k=1}^{m}\mathbf{c}(\mathbf{F}(x_k))\right\}\\
&= \sum_{j=1}^{m}\left\{\mathbb{E}\breve{\mathbf{c}}_{n;h_n}\left(\mathbf{F}(x_j) - \mathbf{c}(\mathbf{F}(x_j))\right)\right\}\\
&\quad\times\left\{\prod_{\ell=1}^{j-1}\mathbb{E}\breve{\mathbf{c}}_{n;h_n}\left(\mathbf{F}(x_\ell)\right)\right\}\left\{\prod_{s=j+1}^{m}\mathbf{c}(\mathbf{F}(x_s))\right\}\\
&\le \kappa^m M^m\sum_{j=1}^{m}\left\{\mathbb{E}\breve{\mathbf{c}}_{n;h_n}\left(\mathbf{F}(x_j)\right) - \mathbf{c}(\mathbf{F}(x_j))\right\}\\
&\le m\kappa^m M^m\max_{1\le j\le m}\left\{\mathbb{E}\breve{\mathbf{c}}_{n;h_n}\left(\mathbf{F}(x_j)\right) - \mathbf{c}(\mathbf{F}(x_j))\right\}.
\end{aligned}$$

We have as usual

$$\begin{aligned}
&\mathbb{E}\breve{\mathbf{c}}_{n;h_n}(u) - \mathbf{c}(u)\\
&= \frac{1}{h^p}\int_{[0,1]^p}K\left(\frac{u-v}{h_n}\right)\mathbf{c}(v)dv - \mathbf{c}(u)\\
&= \int_{\prod_{i=1}^{p}\left[\frac{u_i-1}{h_n},\frac{u_i}{h_n}\right]}\mathbf{c}(u-h_nv)K(v)dv - \mathbf{c}(u)
\end{aligned}$$

$$\begin{aligned}
&= \int_{\prod_{i=1}^{p}\left[\frac{u_i-1}{h_n},\frac{u_i}{h_n}\right]}(\mathbf{c}(u-h_nv)-\mathbf{c}(u))K(v)dv\\
&\quad +\mathbf{c}(u)\left\{\int_{\prod_{i=1}^{p}\left[\frac{u_i-1}{h_n},\frac{u_i}{h_n}\right]}K(v)dv-1\right\}\\
&:= \boldsymbol{\nabla}_{1;n}(u)+\boldsymbol{\nabla}_{2;n}(u).
\end{aligned}$$

Making use of condition **(K.4)** in combination with Taylor expansion of order s, we readily obtain that,

$$\begin{aligned}
&\sup_{u\in[0,1]^p}|\boldsymbol{\nabla}_{1;n}(u)|\\
&\quad = \sup_{\mathbf{u}\in[0,1]^d}\left|\int_{\prod_{i=1}^{p}\left[\frac{u_i-1}{h_n},\frac{u_i}{h_n}\right]}(\mathbf{c}(u-h_nv)-\mathbf{c}(u))K(v)dv\right|\\
&\quad = \frac{h_n^s}{s!}\sup_{u\in[0,1]^p}\left|\int_{\prod_{i=1}^{p}\left[\frac{u_i-1}{h_n},\frac{u_i}{h_n}\right]}\sum_{j_1+\cdots+j_p=s}v_1^{j_1}\ldots v_p^{j_p}\frac{\partial^s\mathbf{c}(u-h_n\theta v)}{\partial u_1^{j_1}\ldots\partial u_p^{j_p}}K(v)dv\right|,
\end{aligned}$$

where

$$\theta=(\theta_1,\ldots,\theta_p)\ \text{ and }\ 0<\theta_i<1.$$

Thus, a routine application of Lebesgue dominated convergence theorem, in turn, implies that

$$\begin{aligned}
&h_n^{-s}\sup_{u\in[0,1]^p}|\boldsymbol{\nabla}_{1;n}(u)|\\
&\quad = \frac{1}{s!}\sup_{u\in[0,1]^p}\left|\sum_{j_1+\cdots+j_p=s}\frac{\partial^s\mathbf{c}(u)}{\partial u_1^{j_1}\ldots\partial u_p^{j_p}}\int_{\mathbb{R}^p}v_1^{j_1}\ldots v_d^{j_p}K(v)dv\right|.
\end{aligned}\tag{10.24}$$

Condition **(F.iii)** allows us to infer from the last statement (10.24) that,

$$\sup_{u\in[0,1]^p}|\boldsymbol{\nabla}_{1;n}(u)|=O(h_n^s).\tag{10.25}$$

Making use of the condition **(K.2)**, we readily obtain that, for n enough large,

$$\sup_{u\in[0,1]^p}|\boldsymbol{\nabla}_{2;n}(u)|=o(1).\tag{10.26}$$

From equations (10.25) and (10.26), we infer, in turn, that

$$\sup_{u\in[0,1]^p}\left|\mathbb{E}\breve{\mathbf{c}}_{n;h_n}(u)-\mathbf{c}(u)\right|=O(h_n^s).$$

Using the last equation gives

$$\sup_{\mathbf{x}\in\mathbb{R}^{mp}}\left|\breve{\Delta}_{3,n}^{(m)}(\mathbf{x})\right|=O(h_n^s).\tag{10.27}$$

In a similar way as in (10.20), (10.22) and (10.27), we have

$$\sup_{(\mathbf{x},\mathbf{y})\in\mathbb{R}^{m(p+q)}}\left|\Delta_{1,n}^{(m)}(\mathbf{x},\mathbf{y})\right| = O\left(\sqrt{\frac{\log n}{n}}\frac{1}{h_n^{p+q+1/2}}\right),\tag{10.28}$$

$$\limsup_{n\to\infty}\sup_{\varrho n^{-1}\log n\le h^{p+q}\le 1}\frac{\sqrt{nh^{p+q}}\|\Delta_{2,n}^{(m)}\|_\infty}{\sqrt{\log(1/h^{p+q})\vee\log\log n}} \le m\kappa^{2m-2}\Sigma'(\varrho)<\infty,\tag{10.29}$$

$$\sup_{(\mathbf{x},\mathbf{y})\in\mathbb{R}^{m(p+q)}}\left|\Delta_{3,n}^{(m)}(\mathbf{x},\mathbf{y})\right|=O(h_n^\gamma).\tag{10.30}$$

Let us introduce the weight when the margins are known

$$\mathcal{W}(\mathbf{i},\boldsymbol{\ell},\mathbf{t};h_n)=\frac{\frac{1}{h_n^{m(p+q)}}\prod_{\nu=1}^{m}K\left(\frac{\mathbf{F}(t_\nu)-\mathbf{F}(X_{\ell_\nu})}{h_n}\right)K\left(\frac{\mathbf{F}_0(Y_{i_\nu})-\mathbf{F}_0(Y_{\ell_\nu})}{h_n}\right)}{\frac{(n-m)!}{n!h_n^{mp}}\sum_{(i_1,\ldots,i_m)\in I(m,n)}\prod_{\nu=1}^{m}K\left(\frac{\mathbf{F}(t_\nu)-\mathbf{F}(X_{i_\nu})}{h_n}\right)}. \tag{10.31}$$

By combining (10.20) and (10.28), we obtain

$$\sup_{\mathbf{t}\in\mathbb{R}^{mp}}|\mathbf{W}(\mathbf{i},\boldsymbol{\ell},\mathbf{t};h_n)-\mathcal{W}(\mathbf{i},\boldsymbol{\ell},\mathbf{t};h_n)|=O\left(\sqrt{\frac{\log n}{n}}\frac{1}{h_n^{p+q+1/2}}\right).$$

This gives

$$\begin{aligned}\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h_n) &= \frac{(n-m)!}{n!(n-1)^m}\sum_{\substack{(i_1,\ldots,i_m),(\ell_1,\ldots,\ell_m)\in I(m,n)\\ i_k\neq\ell_k,k=1,\ldots,m}}\varphi(Y_{i_1},\ldots,Y_{i_m})\mathcal{W}(\mathbf{i},\boldsymbol{\ell},\mathbf{t};h_n)\\ &\quad +O\left(\sqrt{\frac{\log n}{n}}\frac{1}{h_n^{p+q+1/2}}\right)\\ &= \breve{r}_n^{(m)}(\varphi,\mathbf{t};h_n)+O\left(\sqrt{\frac{\log n}{n}}\frac{1}{h_n^{p+q+1/2}}\right).\end{aligned} \tag{10.32}$$

By using (10.22), (10.26), (10.27) and (10.29), we have

$$\begin{aligned}\breve{r}_n^{(m)}(\varphi,\mathbf{t};h_n) &= \frac{(n-m)!}{n!}\sum_{(i_1,\ldots,i_m)\in I(m,n)}\varphi(Y_{i_1},\ldots,Y_{i_m})\left\{\prod_{j=1}^{m}\frac{\mathbf{c}(\mathbf{F}(t_j),\mathbf{F}_0(Y_{i_j}))}{\breve{\mathbf{c}}(\mathbf{F}(t_j))}\right\}\\ &\quad +O\left(\sqrt{\frac{\log(1/h^{p+q})\vee\log\log n}{nh_n^{p+q}}}\right)+O(h_n^{\gamma}).\end{aligned} \tag{10.33}$$

By combining the results (10.4), (10.32) and (10.33), we obtain the desired result

$$\limsup_{n\to\infty}\sup_{a_n\leq h^{p+q}\leq b_0}\sup_{\varphi\in\mathscr{F}}\sup_{\mathbf{t}\in\mathbb{R}^{pm}}\frac{\sqrt{nh^{p+q}}|\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h)-\mathbb{E}\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h)|}{\sqrt{|\log h|\vee\log\log n}}\leq\Sigma(\varrho).$$

Hence the proof of Theorem 2.3 is complete. □

Proof of Theorem 2.2

The proof of this theorem is very similar to the preceding one. By the preceding steps, using the fact that the class of functions $\mathscr{F}\cdot\mathscr{C}$ is bounded, we infer that, as in the preceding proof,

$$\begin{aligned}&\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h_n)\\ &= \frac{(n-m)!}{n!}\sum_{(i_1,\ldots,i_m)\in I(m,n)}\varphi(Y_{i_1},\ldots,Y_{i_m})\left\{\prod_{j=1}^{m}\frac{\mathbf{c}(\mathbf{F}(t_j),\mathbf{F}_0(Y_{i_j}))}{\breve{\mathbf{c}}(\mathbf{F}(t_j))}\right\}\\ &\quad +O\left(\sqrt{\frac{\log n}{n}}\frac{1}{h_n^{p+q+1/2}}\right)+O\left(\sqrt{\frac{\log(1/h^{p+q})\vee\log\log n}{nh_n^{p+q}}}\right)+O(h_n^{\gamma}).\end{aligned}$$

This suffices to complete the proof of the theorem. □

Proof of Corollary 2.1

We have to evaluate

$$\mathbb{E}\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h_n)-r^{(m)}(\varphi,\mathbf{t}).$$

From (10.20), (10.22) and (10.27), we readily infer that

$$\limsup_{n\to\infty}\sup_{a_n\le h^{p+q}\le b_0}\sup_{\varphi\in\mathscr{F}}\sup_{\mathbf{t}\in\mathbb{R}^{pm}}|\mathbb{E}\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h_n)-\widetilde{\mathbb{E}}\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h_n)|=o(1),$$

where

$$\mathbb{E}\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h_n) = \frac{\mathbb{E}\left\{\frac{1}{h_n^{m(p+q)}}\varphi(Y_1,\ldots,Y_m)\prod_{j=1}^{m}K\left(\frac{\mathbf{F}_n(t_j)-\mathbf{F}_n(X_j)}{h_n}\right)K\left(\frac{\mathbf{F}_{0,n}(Y_{m+1})-\mathbf{F}_{0,n}(Y_j)}{h_n}\right)\right\}}{\left\{\prod_{j=1}^{m}\breve{\mathbf{c}}(\mathbf{F}(t_j))\right\}}.$$

In a similar way, by combining (10.28), (10.29) and (10.30), we obtain that

$$\limsup_{n\to\infty}\sup_{a_n\le h^{p+q}\le b_0}\sup_{\varphi\in\mathscr{F}}\sup_{\mathbf{t}\in\mathbb{R}^{pm}}|\widetilde{\mathbb{E}}\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h_n)-r^{(m)}(\varphi,\mathbf{t})|=o(1).$$

Hence we have

$$\limsup_{n\to\infty}\sup_{a_n\le h^{p+q}\le b_0}\sup_{\varphi\in\mathscr{F}}\sup_{\mathbf{t}\in\mathbb{R}^{pm}}|\mathbb{E}\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h_n)-r^{(m)}(\varphi,\mathbf{t})|=o(1).$$

This when combined with Theorem 2.3, implies that

$$\begin{aligned}
&\limsup_{n\to\infty}\sup_{a_n\le h^{p+q}\le b_0}\sup_{\varphi\in\mathscr{F}}\sup_{\mathbf{t}\in\mathbb{R}^{pm}}|\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h)-r^{(m)}(\varphi,\mathbf{t})|\\
&\quad\le \limsup_{n\to\infty}\sup_{a_n\le h^{p+q}\le b_0}\sup_{\varphi\in\mathscr{F}}\sup_{\mathbf{t}\in\mathbb{R}^{pm}}|\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h)-\mathbb{E}\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h_n)|\\
&\qquad+\limsup_{n\to\infty}\sup_{a_n\le h^{p+q}\le b_0}\sup_{\varphi\in\mathscr{F}}\sup_{\mathbf{t}\in\mathbb{R}^{pm}}|\mathbb{E}\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h_n)-r^{(m)}(\varphi,\mathbf{t})|\\
&\quad= o(1).
\end{aligned}$$

Hence the proof is complete. □

Proof of Theorem 3.3

Recall that

$$\breve{r}_n^{(m)}(\varphi,\mathbf{t};h_n)=\frac{(n-m)!}{n!(n-1)^m}\sum_{\substack{(i_1,\ldots,i_m),(\ell_1,\ldots,\ell_m)\in I(m,n)\\ i_k\neq\ell_k,k=1,\ldots,m}}\varphi(Y_{i_1},\ldots,Y_{i_m})\mathcal{W}(\mathbf{i},\boldsymbol{\ell},\mathbf{t};h_n).$$

Let us center $\widetilde{r}_n^{(m)}(\varphi,\mathbf{t};h_n)$ and $\breve{r}_n^{(m)}(\varphi,\mathbf{t};h_n)$ as follows

$$\begin{aligned}
&\widetilde{r}_n^{(m,0)}(\varphi,\mathbf{t};h_n)\\
&\quad=\frac{(n-m)!}{n!(n-1)^m}\sum_{\substack{(i_1,\ldots,i_m),(\ell_1,\ldots,\ell_m)\in I(m,n)\\ i_k\neq\ell_k,k=1,\ldots,m}}\left\{\varphi(Y_{i_1},\ldots,Y_{i_m})-r^{(m)}(\varphi,\mathbf{t})\right\}\\
&\qquad\times\mathbf{W}(\mathbf{i},\boldsymbol{\ell},\mathbf{t};h_n),
\end{aligned}$$

$$\begin{aligned}
&\breve{r}_n^{(m,0)}(\varphi,\mathbf{t};h_n)\\
&\quad=\frac{(n-m)!}{n!(n-1)^m}\sum_{\substack{(i_1,\ldots,i_m),(\ell_1,\ldots,\ell_m)\in I(m,n)\\ i_k\neq\ell_k,k=1,\ldots,m}}\left\{\varphi(Y_{i_1},\ldots,Y_{i_m})-r^{(m)}(\varphi,\mathbf{t})\right\}
\end{aligned}$$

$$\times \mathcal{W}(\mathbf{i}, \boldsymbol{\ell}, \mathbf{t}; h_n).$$

Following [139], we have the following decomposition

$$\begin{aligned}
&\widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h_n) - r^{(m)}(\varphi, \mathbf{t}) \\
&= \breve{r}_n^{(m,0)}(\varphi, \mathbf{t}; h_n) + \left\{\widetilde{r}_n^{(m,0)}(\varphi, \mathbf{t}; h_n) - \breve{r}_n^{(m,0)}(\varphi, \mathbf{t}; h_n)\right\} \\
&\quad + \frac{(n-m)!}{n!(n-1)^m} \sum_{\substack{(i_1,\ldots,i_m),(\ell_1,\ldots,\ell_m)\in I(m,n) \\ i_k \neq \ell_k, k=1,\ldots,m}} \{\mathbf{W}(\mathbf{i}, \boldsymbol{\ell}, \mathbf{t}; h_n) - \mathcal{W}(\mathbf{i}, \boldsymbol{\ell}, \mathbf{t}; h_n)\} \\
&\quad \times \left\{r^{(m)}(\varphi, X_{i_1}, \ldots, X_{i_m}) - r^{(m)}(\varphi, \mathbf{t})\right\} \\
&\quad + \frac{(n-m)!}{n!(n-1)^m} \sum_{\substack{(i_1,\ldots,i_m),(\ell_1,\ldots,\ell_m)\in I(m,n) \\ i_k \neq \ell_k, k=1,\ldots,m}} \mathcal{W}(\mathbf{i}, \boldsymbol{\ell}, \mathbf{t}; h_n) \\
&\quad \times \left\{r^{(m)}(\varphi, X_{i_1}, \ldots, X_{i_m}) - r^{(m)}(\varphi, \mathbf{t})\right\} + o_{\mathbb{P}}((nh^{p+q})^{-1/2}) \\
&:= I + II + III + IV.
\end{aligned}$$

An application of Proposition 3.2, implies that

$$\sqrt{nh_n^{p+q}}\, I \to N(0, \rho^2),$$

where

$$\rho^2 := \sum_{j,i=1}^{m} \mathbb{1}_{\{t_j = t_l\}} (\theta_{j,l}(t_1, \ldots, t_m) - r^{(m)}{}^2(\varphi, \mathbf{t}) \|\mathbb{K}\|_2^2 / \breve{\mathbf{c}}(\mathbf{F}(t_j)).$$

An application of Lemmas 3, 4, 5 of [139], permits us to conclude that

$$\begin{aligned}
\sqrt{nh_n^{p+q}}\, II &\to 0, \\
\sqrt{nh_n^{p+q}}\, III &\to 0, \\
\sqrt{nh_n^{p+q}}\, IV &\to 0.
\end{aligned}$$

Since $\breve{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$ is very close to $\widetilde{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$, we will obtain our results for $\breve{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$. Let us introduce some notation from [56] and follow the steps of their proofs. Let us denote $K_h(z) = K_h(z/h)$ and the product kernel

$$\begin{aligned}
\widetilde{K}(\mathbf{t}_1, \mathbf{t}_2) &:= \prod_{j=1}^{m} K(t_{j,1}) K_0(t_{j,2}), \quad (\mathbf{t}_1, \mathbf{t}_2) \\
&= \prod_{j=1}^{m} (t_{j,1,1}, \ldots, t_{j,1,p}, t_{j,2,1}, \ldots, t_{j,2,q}) \in \mathbb{R}^{m(p+q)}.
\end{aligned}$$

For a function $g \in \mathscr{F}$, consider the U-kernel

$$G_{g,h_n,\mathbf{t}}(\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2) := g(\mathbf{y}) \widetilde{K}_{h_n}(\mathbf{t} - \mathbf{x}, \mathbf{y}_1 - \mathbf{y}_2), \quad \mathbf{x}, \mathbf{t} \in \mathbb{R}^{pm}, \quad \mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^{qm},$$

and for the sample $(X_1, Y_1), \ldots, (X_n, Y_n)$, define

$$U_n(g, h_n, \mathbf{t}) := U_n^{(m)}(G_{g,h_n,\mathbf{t}}) = \frac{(n-m)!}{n!} \sum_{\mathbf{i}, \boldsymbol{\ell} \in I_n^m, \mathbf{i} \neq \boldsymbol{\ell}} G_{g,h_n,\mathbf{t}}(\mathbf{X}_{\boldsymbol{\ell}}, \mathbf{Y}_{\mathbf{i}}, \mathbf{Y}_{\boldsymbol{\ell}}),$$

where, throughout this paper, we shall use the notation

$$\begin{aligned}
\mathbf{X} &= (X_1, \ldots, X_m) \in \mathbb{R}^{mp} \ \text{ and } \ \mathbf{X}_{\mathbf{i}} = (X_{i_1}, \ldots, X_{i_k}) \in \mathbb{R}^{kp}, \mathbf{i} \in I_n^k, \\
\mathbf{Y} &= (Y_1, \ldots, Y_m) \in \mathbb{R}^{mq} \ \text{ and } \ \mathbf{Y}_{\mathbf{i}} = (Y_{i_1}, \ldots, Y_{i_k}) \in \mathbb{R}^{kq}, \mathbf{i} \in I_n^k,
\end{aligned}$$

Now, introduce the U-process

$$\eta_n^{(m)}(g, h_n, \mathbf{t}) := \sqrt{nh_n^m}\{U_n(g, h_n, \mathbf{t}) - \mathbb{E}U_n(g, h_n, \mathbf{t})\}. \tag{10.34}$$

In the sequel, we will need to symmetrize the functions $G_{g,h_n,\mathbf{t}}(\cdot,\cdot)$, see Remark 3.4. To do this, we set

$$\begin{aligned}\overline{G}_{g,h_n,\mathbf{t}}(\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2) &= (m!)^{-1} \sum_{\sigma_1,\sigma_2 \in I_m^m} G_{g,h_n,\mathbf{t}}(\mathbf{x}_{\sigma_2}, \mathbf{y}_{\sigma_1}, \mathbf{y}_{\sigma_2}) \\ &= (m!)^{-1} \sum_{\sigma_1,\sigma_2 \in I_m^m} g(\mathbf{y}_{\sigma_1}) \widetilde{K}_{h_n}(\mathbf{t} - \mathbf{x}_{\sigma_2}, \mathbf{y}_{\sigma_1} - \mathbf{y}_{\sigma_2}),\end{aligned}$$

where $\mathbf{z}_\sigma = (z_{\sigma_1}, \ldots, z_{\sigma_m})$. Obviously, the expectation of $G_{g,h_n,\mathbf{t}}(\cdot,\cdot)$ remains unchanged after symmetrization and

$$U_n^{(m)}(\overline{G}_{g,h_n,\mathbf{t}}(\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2)) = U_n(g, h_n, \mathbf{t}),$$

so the U-statistic process in (10.34) may be redefined using the symmetrized kernels, that is, we consider

$$\eta_n^{(m)}(g, h_n, \mathbf{t}) := \sqrt{nh_n^{m(p+q)}}\{U_n^{(m)}(\overline{G}_{g,h_n,\mathbf{t}}) - \mathbb{E}U_n^{(m)}(\overline{G}_{g,h_n,\mathbf{t}})\}. \tag{10.35}$$

Since we assume $\mathscr{F}$ to be of VC-type with an envelope function F and $\mathscr{K}$ to be of VC-type with envelope κ, it is readily checked (via Lemma A.1 in [61]) that the class of functions on $\mathbb{R}^{pm} \times \mathbb{R}^{qm}$ given by $\{h^{m(p+q)} G_{g,h_n,\mathbf{t}}(\cdot,\cdot) : g \in \mathscr{F}, \mathbf{t} \in \mathbb{R}^{pm}\}$ is of VC-type, as well as the class

$$\mathscr{G} = \{h^{m(p+q)} \overline{G}_{g,h_n,\mathbf{t}}(\cdot,\cdot) : g \in \mathscr{F}, h, \mathbf{t} \in \mathbb{R}^{pm}\}, \tag{10.36}$$

for which we denote the VC-type characteristics by A_1 and v_1, and the envelope function by

$$\widetilde{F}(\mathbf{y}_1) \equiv \widetilde{F}(\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2) = \kappa^{m(p+q)} \sum_{\sigma \in I_m^m} F(\mathbf{y}_\sigma), \quad \mathbf{y} \in \mathbb{R}^{qm}. \tag{10.37}$$

Next, for $k = 1, \ldots, m$, introduce the following classes of function on $\mathbb{R}^{kp} \times \mathbb{R}^{2kq}$

$$\mathscr{G}^{(k)} = \{h^{m(p+q)} \pi_k \overline{G}_{g,h_n,\mathbf{t}}(\cdot,\cdot,\cdot) : g \in \mathscr{F}, h, \mathbf{t} \in \mathbb{R}^{pm}\}, \tag{10.38}$$

An argument in [75] then shows that each class $\mathscr{G}^{(k)}$ is of VC-type with characteristics A_1 and v_1 and envelope function

$$\mathbf{F}_k \leq 2^k \|\widetilde{\mathbf{F}}\|_\infty. \tag{10.39}$$

The linear term

$$m\sqrt{n}U_n^{(1)}(\pi_1 \overline{G}_{g,h_n,\mathbf{t}}(\cdot,\cdot,\cdot)) = \frac{m}{\sqrt{n}} \sum_{i=1}^n \pi_1 \overline{G}_{g,h_n,\mathbf{t}}(X_i, Y_i, Y_i),$$

from the definition of the Hoeffding projections and recalling that the sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d., we can say, for all $(x, y) \in \mathbb{R}$, that

$$\begin{aligned}&\pi_1 \overline{G}_{g,h_n,\mathbf{t}}(x, y, y) \\ &\quad= \mathbb{E}(\overline{G}_{g,h_n,\mathbf{t}}((x, X_2, \ldots, X_m), (y, Y_2, \ldots, Y_m), (y, Y_{m+1}, \ldots, Y_{2m}))) \\ &\qquad - \mathbb{E}(\overline{G}_{g,h_n,\mathbf{t}}(\mathbf{X}, \mathbf{Y}, \mathbf{Y}) \\ &\quad= \mathbb{E}(\overline{G}_{g,h_n,\mathbf{t}}(\mathbf{X}, \mathbf{Y}, \mathbf{Y}) \mid (X_1, Y_1) = (x, y)) - \mathbb{E}(\overline{G}_{g,h_n,\mathbf{t}}(\mathbf{X}, \mathbf{Y}, \mathbf{Y})).\end{aligned}$$

Introduce the following function on $\mathbb{R}^{pm} \times \mathbb{R}^{qm}$:

$$\begin{aligned}S_{g,h_n,\mathbf{t}} : \mathbb{R}^{pm} \times \mathbb{R}^{qm} &\to \mathbb{R} \\ (x, y) &\mapsto m h_n^{m(p+q)} \mathbb{E}(\overline{G}_{g,h_n,\mathbf{t}}(\mathbf{X}, \mathbf{Y}, \mathbf{Y}) \mid (X_1, Y_1) = (x, y)).\end{aligned}$$

Using this notation, we write

$$mh_n^m\pi_1\overline{G}_{g,h_n,\mathbf{t}}(x,y,y)=S_{g,h_n,\mathbf{t}}(x,y,y)-\mathbb{E}(S_{g,h_n,\mathbf{t}}(X_1,Y_1,Y_1)).$$

and hence for all $g\in\mathscr{F}$, and $\mathbf{t}\in\mathbb{R}^{pm}$, the linear term of the Hoeffding decomposition times $h_n^{m(p+q)}$ is given by

$$\begin{aligned}
&mU_n^{(1)}(\pi_1\overline{G}_{g,h_n,\mathbf{t}})\\
&= |\widetilde{\eta}_{g,h,\mathbf{t}}-\mathbb{E}\widetilde{\eta}_{g,h,\mathbf{t}}|\\
&= \frac{1}{nh_n^{m(p+q)}}\sum_{i=1}^n S_{g,h_n,\mathbf{t}}(X_i,Y_i,Y_i)-\mathbb{E}(S_{g,h_n,\mathbf{t}}(\mathbf{X},\mathbf{Y},\mathbf{Y}))\\
&= \alpha_n(S_{g,h_n,\mathbf{t}})
\end{aligned}$$

where this last expression is an empirical process α_n based on the sample $(X_1,Y_1),\ldots,(X_n,Y_n)$ and we set for $\mathbf{t}\in\mathbf{I}$, $g\in\mathcal{F}$ and $h\geq 0$ the class of functions on $\mathbb{R}^{pm}\times\mathbb{R}^{qm}$,

$$\mathscr{S}_n=\{S_{g,h_n,\mathbf{t}}:g\in\mathscr{F},h\geq l_n,\mathbf{t}\in\mathbf{I}\}.$$

Clearly, $S_n\subset m\mathscr{G}^{(1)}$ and the class $m\mathscr{G}^{(1)}$ has envelope function mF_1, where F_1 is the envelope function of the class $\mathscr{G}^{(1)}$ defined in (10.38). From the above discussion, this class is of VC-type with the same characteristics as $\mathscr{G}$ and, the conclusion of the theorem follows from the classical theory of the empirical processes. □.

Proof of Theorem 5.1

In the following proposition, we show that Theorem 5.1 naturally follows from Theorem 2.2. We first establish the version of Theorem 5.1 corresponding to the case where $G(\cdot)$ is known (i.e., with $\breve{r}_n^{(m)*}(\psi,\mathbf{t};h_n)$ replaced by $\breve{r}_n^{(m)}(\psi,\mathbf{t};h_n)$). To complete the proof of Theorem 5.1, the consistency of the Kaplan-Meier estimator will be helpful (see Lemma 10.3 below)

Proposition 10.2. *Under assumptions* ***(A.1- 3)****,* ***(I)****, and conditions of Theorem 2.2, assume that h satisfies* ***(H.1-3)****, with probability 1; then*

$$\limsup_{n\to\infty}\sup_{a_n\leq h^{p+1}\leq b_0}\sup_{\psi\in\mathscr{F}}\sup_{\mathbf{t}\in\mathbf{I}^p}\frac{\sqrt{nh^{p+1}}|\breve{r}_n^{(m)}(\psi,\mathbf{t};h_n)-\mathbb{E}\widetilde{r}_n^{(m)}(\psi,\mathbf{t};h)|}{\sqrt{|\log h|\vee\log\log n}}\leq\Sigma''(\varrho).$$

Proof of Proposition 10.2

Recalling the definition 5.1 of Φ_ψ

$$\mathbf{\Phi}_\psi(y_1,\ldots,y_m,c_1,\ldots,c_m)=\frac{\prod_{i=1}^m\{\mathbb{1}\{y_i\leq c_i\}\psi(y_1\wedge c_1,\ldots,y_m\wedge c_m)}{\prod_{i=1}^m\{1-G(y_i\wedge c_i)\}}.$$

$$\Phi_\psi(y,c)=\frac{\mathbb{1}\{y\leq c\}\psi(y\wedge c)}{1-G(y\wedge c)}.\tag{10.40}$$

it is obvious that $\mathbf{\Phi}_\psi$ is uniformly bounded on $\mathbb{R}^{2m}$ and $\psi\in\mathcal{F}$, since $\mathscr{F}$ is uniformly bounded, $\psi(t)=0$ for all $t>\tau$ and $G(\tau)<1$. This property, when combined with the VC property of $\mathscr{F}$, ensures that the class of function

$$\mathcal{F}_\Phi:=\{\Phi_\psi:\psi\in\mathscr{F}\}$$

verifies (F.ii), (F.iii). Similarly, it can be shown that $\mathscr{F}_\Phi$ is a pointwise measurable class of functions (F.i). Moreover, by **(A.3)** and (5.2), the class

$$\mathcal{M}_\Phi:=\{r^{(m)}(\mathbf{\Phi}_\psi,\mathbf{t})\mid f_{\mathbf{X}},\psi\in\mathcal{F}\}$$

is almost surely relatively compact with respect to the sup- norm topology on I_α. So we can apply Theorem 2.2 with $\mathbf{Y} = (Y, C)$ and $\Psi = \Phi_\psi$. The result of Proposition 10.2 is straightforward. □

To complete the demonstration of Theorem 5.1, we will use the result of the next approximation Lemma 10.3 as in [18].

Lemma 10.3. *Under assumptions of Theorem 5.1, we have with probability one,*

$$\sup_{a_n \le h^{p+1} \le b_0} \sup_{\mathbf{t}\in\mathbf{I}} \sup_{\psi\in\mathcal{F}} \left| \breve{r}_n^{(m)*}(\psi, \mathbf{t}; h_n) - \breve{r}_n^{(m)}(\psi, \mathbf{t}; h_n) \right| = o\left(\sqrt{\frac{\log(1/h)}{nh^{p+1}}}\right) \quad \textit{as} \quad n \to \infty. \tag{10.41}$$

Proof of Lemma 10.3

Since

$$\sup_{\psi\in\mathcal{F}} |\psi(t)| < \infty,$$

the kernel $K(\cdot)$ is uniformly bounded and

$$\tau < T_H = T_F \le T_G,$$

the law of iterated logarithm for $G_n^*(\cdot)$ established in [66] ensures that

$$\sup_{t\le\tau} |G_n^* - G(t)| = O\left(\sqrt{\frac{\log\log n}{n}}\right) \quad \text{almost surely as} \quad n \to \infty.$$

By combining the results of Proposition 10.2 and Lemma 10.3, the result of the Theorem 5.1 is immediate by noting that, under the conditions **(H.1-3)**, we have, for n sufficiently large,

$$\sup_{t\le\tau} |G_n^* - G(t)| = o\left(\sqrt{\frac{\log(1/h)}{nh^{p+1}}}\right) \quad \text{almost surely as} \quad n \to \infty.$$

Hence this permits to conclude the proof. □

Appendix

Let $\mathfrak{G}$ denote a class of measurable real valued functions $g(\cdot)$ of $(\mathbf{u}, h) \in [0,1]^d \times (0,1]$. We shall assume that $\mathfrak{G}$ satisfies:

(G.i):

$$\sup_{0<h\le1} \sup_{g\in\mathfrak{G}} \|g(\cdot; h)\|_\infty =: \kappa < \infty.$$

Assume that there exists a constant $\mathfrak{C} > 0$, such that, for all $h \in (0,1]$,

(G.ii):

$$\sup_{g\in\mathfrak{G}} \mathbb{E}[g^2(\boldsymbol{\xi}; h)] \le \mathfrak{C}h.$$

For $\varepsilon > 0$, set

$$N(\varepsilon, \mathfrak{G}) = \sup_Q N(\kappa\varepsilon, \mathfrak{G}, d_Q),$$

where the supremum is taken over all probability measures Q on $(\mathbb{R}^d, \mathcal{B})$, where $\mathcal{B}$ represents the σ-field of Borel sets of $\mathbb{R}^d$. We shall also assume that the class of functions $\mathfrak{G}$ satisfies the following uniform entropy condition.

(F.i): For some $\mathfrak{C}_0 > 0$ and $\nu_0 > 0$,

$$N(\varepsilon, \mathfrak{G}) \le \mathfrak{C}_0 \varepsilon^{-\nu_0}, \quad \text{for} \quad 0 < \varepsilon < 1.$$

Finally, to avoid using outer probability measures in all of the statements, we impose the following measurability assumption.

(F.ii): $\mathfrak{G}$ is a pointwise measurable class, that is, there exists a countable subclass $\mathfrak{G}_0$ of $\mathfrak{G}$ such that we can find for any function $g \in \mathfrak{G}$ a sequence of functions $\{g_m : m \geq 1\}$ in $\mathfrak{G}_0$ for which

$$g_m(\mathbf{z}) \longrightarrow g(\mathbf{z}), \quad \text{for} \quad \mathbf{z} \in \mathbb{R}^d.$$

For any $n \geq 1$, $g \in \mathfrak{G}$ and $0 < h < 1$, let us define

$$g_{n,h} := \frac{1}{n}\sum_{k=1}^{n} g(\boldsymbol{\xi}_k; h).$$

Mason and Swanepoel [105] have proved the following general result.

Theorem A. *Assuming* **(G.i)**, **(G.ii)**, **(F.i)** *and* **(F.ii)**, *we have for $\varrho > 0$ and $0 < h_0 < 1$, with probability one,*

$$\limsup_{n\to\infty} \sup_{\frac{\varrho \log n}{n} \leq h \leq h_0} \sup_{g\in\mathfrak{G}} \frac{\sqrt{n}|g_{n,h} - \mathbb{E}g_{n,h}|}{\sqrt{h(|\log h| \vee \log\log n)}} =: A(\varrho) < \infty. \tag{10.42}$$

Acknowledgments

The author would like to thank Professor Stute for providing his paper [139] and Professor Veraverbeke for sending his paper [88]. The author is indebted to the Editor-in-Chief, Associate Editor and referees for their helpful comments and suggestions on the first version of our article, which helped us improve the manuscript's content, presentation, and layout. This paper is dedicated to my dearest Badreddine Bouzebda.

References

[1] J. Abrevaya and W. Jiang, *A nonparametric approach to measuring and testing curvature*, J. Bus. Econom. Statist. **23** (1), 1-19, 2005.

[2] H. Akaike, *An approximation to the density function*, Ann. Inst. Statist. Math., Tokyo **6**, 127–132, 1954.

[3] M.A. Arcones, *The law of the iterated logarithm for U-processes,* J. Multivariate Anal. **47** (1), 139–151, 1993.

[4] M.A. Arcones and E. Giné, *Limit theorems for U-processes,* Ann. Probab. **21** (3), 1494–1542, 1993.

[5] M.A. Arcones and Y. Wang, *Some new tests for normality based on U-processes*, Statist. Probab. Lett. **76** (1), 69–82, 2006.

[6] K. Benhenni, F. Ferraty, M. Rachdi, and P. Vieu. *Local smoothing regression with functional data*, Comput. Statist. **22** (3), 353–369, 2007.

[7] W. Bergsma and A. Dassios, *A consistent test of independence based on a sign covariance related to Kendall's tau*, Bernoulli **20** (2), 1006–1028, 2014.

[8] J. R. Blum, J. Kiefer, and M. Rosenblatt, *Distribution free tests of independence based on the sample distribution function*, Ann. Math. Statist. 32, 485–498, 1961.

[9] S. Borovkova, R. Burton, and H. Dehling, *Consistency of the Takens estimator for the correlation dimension*, Ann. Appl. Probab. **9** (2), 376–390, 1999.

[10] S. Bouzebda, *Some new multivariate tests of independence*, Math. Methods Statist. **20** (3), 192–205, 2011.

[11] S. Bouzebda, *On the strong approximation of bootstrapped empirical copula processes with applications*, Math. Methods Statist. **21** (3), 153–188, 2012.

[12] S. Bouzebda and T. Zari, *Asymptotic behavior of weighted multivariate Cramér-von Mises-type statistics under contiguous alternatives*, Math. Methods Statist. **22** (3), 226–252, 2013.

[13] S. Bouzebda, *Bootstrap de l'estimateur de Hill: théorèmes limites*, Ann. I.S.U.P. **54** (1-2), 61–72, 2010.

[14] S. Bouzebda, *Strong approximation of the smoothed Q-Q processes*, East J. Theor. Stat. **31** (2), 169–191, 2010.

[15] S. Bouzebda, *General tests of independence based on empirical processes indexed by functions*, Stat. Methodol. **21**, 59–87, 2014.

[16] S. Bouzebda, *Kac's representation for empirical copula process from an asymptotic viewpoint*, Statist. Probab. Lett. **123**, 107–113, 2017.

[17] S. Bouzebda and M. Cherfi, *Test of symmetry based on copula function*, J. Statist. Plann. Inference **142** (5), 1262–1271, 2012.

[18] S. Bouzebda and T. El-hadjali, *Uniform convergence rate of the kernel regression estimator adaptive to intrinsic dimension in presence of censored data*, J. Nonparametr. Stat. **32** (4), 864–914, 2020.

[19] S. Bouzebda and I. Elhattab, *A strong consistency of a nonparametric estimate of entropy under random censorship*, C. R. Math. Acad. Sci. Paris **347** (13-14), 821–826, 2009.

[20] S. Bouzebda and I. Elhattab, *Uniform in bandwidth consistency of the kernel-type estimator of the Shannon's entropy*, C. R. Math. Acad. Sci. Paris **348** (5-6), 317–321, 2010.

[21] S. Bouzebda and I. Elhattab, *Uniform-in-bandwidth consistency for kernel-type estimators of Shannon's entropy*, Electron. J. Stat. **5**, 440–459, 2011.

[22] S. Bouzebda, I. Elhattab and B. Nemouchi, *On the uniform-in-bandwidth consistency of the general conditional U-statistics based on the copula representation*, J. Nonparametr. Stat. **33** (2), 321–358, 2021.

[23] S. Bouzebda, I. Elhattab and C.T. Seck, *Uniform in bandwidth consistency of nonparametric regression based on copula representation*, Statist. Probab. Lett. **137**, 173–182, 2018.

[24] S. Bouzebda and A. Keziou, *New estimates and tests of independence in semiparametric copula models*, Kybernetika (Prague) **46** (1), 178–201, 2010.

[25] S. Bouzebda and A. Keziou, *A new test procedure of independence in copula models via χ^2-divergence*, Comm. Statist. Theory Methods **39** (1-2), 1–20, 2010.

[26] S. Bouzebda and N. Limnios, *Exchangeably weighted bootstraps of empirical estimators of a semi-Markov kernel*, C. R. Math. Acad. Sci. Paris **351** (13-14), 569–573, 2013.

[27] S. Bouzebda and N. Limnios, *Exchangeably weighted bootstraps of martingale difference arrays under the uniformly integrable entropy,* J. Stoch. Anal. **1** (3), Art. 6, 13, 2020.

[28] S. Bouzebda and B. Nemouchi, *Uniform consistency and uniform in bandwidth consistency for nonparametric regression estimates and conditional U-statistics involving functional data,* J. Nonparametr. Stat. **32** (2), 452–509, 2020.

[29] S. Bouzebda and B. Nemouchi, *Weak-convergence of empirical conditional processes and conditional U-processes involving functional mixing data*, Stat. Inference Stoch. Process. **26** (1), 33–88, 2023.

[30] S. Bouzebda and A. Nezzal, *Uniform consistency and uniform in number of neighbors consistency for nonparametric regression estimates and conditional U-statistics involving functional data*, Jpn. J. Stat. Data Sci. **5** (2), 431–533, 2022.

[31] S. Bouzebda and A. Nezzal, *Asymptotic properties of conditional U-statistics using delta sequences.* Comm. Statist. Theory Methods, pages 1–56, 2023.

[32] S. Bouzebda, A. Nezzal and T. Zari, *Uniform consistency for functional conditional u-statistics using delta-sequences*, Mathematics **11** (1), 1–39, 2023.

[33] S. Bouzebda, C. Papamichail and N. Limnios, *On a multidimensional general bootstrap for empirical estimator of continuous-time semi-Markov kernels with applications*, J. Nonparametr. Stat. **30** (1), 49–86, 2018.

[34] S. Bouzebda and I. Soukarieh, *Non-parametric conditional U-processes for locally stationary functional random fields under stochastic sampling design*, Mathematics **11**

(1), 1–70, 2023.

[35] S. Bouzebda and I. Soukarieh, *Renewal type bootstrap for U-process Markov chains,* Markov Process. Related Fields, 1–52, 2023.

[36] S. Bouzebda and N. Taachouche, *On the variable bandwidth kernel estimation of conditional U-statistics at optimal rates in sup-norm*, Phys. A **625**, Paper No. 129000, 2023.

[37] S. Bouzebda and N. Taachouche, *Rates of the strong uniform consistency for the kernel-type regression function estimators with general kernels on manifolds*, Math. Methods Statist. **32** (1), 27–80, 2023.

[38] T. T. Cai and L. Zhang, *High-dimensional Gaussian copula regression: adaptive estimation and statistical inference,* Statist. Sinica **28** (2), 963–993, 2018.

[39] A. Carbonez, L. Györfi and E.C. van der Meulen, P*artitioning-estimates of a regression function under random censoring*, Statist. Decisions **13** (1), 21–37, 1995.

[40] J. E. Chacón, J. Montanero and A.G. Nogales, *A note on kernel density estimation at a parametric rate*, J. Nonparametr. Stat. **19** (1), 13–21, 2007.

[41] J.E. Chacón and T. Duong, *Multivariate kernel smoothing and its applications*, volume 160 of Monographs on Statistics and Applied Probability, CRC Press, Boca Raton, FL, 2018.

[42] S.X. Chen and T-M. Huang, *Nonparametric estimation of copula functions for dependence modelling*, Canad. J. Statist. **35** (2), 265–282, 2007.

[43] U. Cherubini, F. Gobbi and S. Mulinacci, *Convolution Copula Econometrics*, SpringerBriefs in Statistics. Springer, Cham, 2016.

[44] U. Cherubini, E. Luciano and W. Vecchiato, *Copula Methods in Finance*, Wiley Finance Series. John Wiley & Sons, Ltd., Chichester, 2004.

[45] K. Chokri and S. Bouzebda, *Uniform-in-bandwidth consistency results in the partially linear additive model components estimation*, Comm. Statist. Theory Methods, 1–42, 2023.

[46] K,-L. Chung, *An estimate concerning the Kolmogoroff limit distribution,* Trans. Amer. Math. Soc. **67**, 36–50, 1949.

[47] S. Clémençon, G. Lugosi and N. Vayatis, *Ranking and empirical minimization of U-statistics*, Ann. Statist. **36** (2), 844–874, 2008.

[48] M. Csörgő and P.Révész, *Strong Approximations in Probability and Statistics*, Probability and Mathematical Statistics. Academic Press, Inc, [Harcourt Brace Jovanovich, Publishers], New York-London, 1981.

[49] G. Dall'Aglio, S. Kotz and G. Salinetti, editors, *Advances in probability distributions with given marginals*, volume 67 of *Mathematics and its Applications*, Kluwer Academic Publishers Group, Dordrecht, 1991. Beyond the copulas, Papers from the Symposium on Distributions with Given Marginals held in Rome, April 1990.

[50] V.H. de la Peña and E. Giné, *Decoupling, From dependence to independence, Randomly stopped processes. U-statistics and processes. Martingales and beyond*, Probability and its Applications (New York). Springer-Verlag, New York, 1999.

[51] P. Deheuvels, *One bootstrap suffices to generate sharp uniform bounds in functional estimation*, Kybernetika (Prague) **47** (6), 855–865, 2011.

[52] P. Deheuvels and J. H.J. Einmahl, *Functional limit laws for the increments of Kaplan-Meier product-limit processes and applications*, Ann. Probab. **28** (3), 1301–1335, 2000.

[53] P. Deheuvels and D.M. Mason, *General asymptotic confidence bands based on kernel-type function estimators*, Stat. Inference Stoch. Process. **7** (3), 225–277, 2004.

[54] H. Dette, R. Van Hecke and S. Volgushev, *Some comments on copula-based regression*, J. Amer. Statist. Assoc. **109** (507), 1319–1324, 2014.

[55] L. Devroye and G. Lugosi, *Combinatorial Methods in Density Estimation*, Springer Series in Statistics. Springer-Verlag, New York, 2001.

[56] J. Dony and D.M. Mason, *Uniform in bandwidth consistency of conditional U-statistics*, Bernoulli **14** (4), 1108–1133, 2008.

[57] R.M. Dudley, A course on empirical processes. In *École d'été de probabilités de Saint-Flour, XII—1982*, volume 1097 of *Lecture Notes in Math.*, pages 1–142. Springer, Berlin, 1984.

[58] R. M. Dudley, *Uniform Central Limit Theorems*, volume 142 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, New York, second edition, 2014.

[59] F. Durante and C. Sempi, *Principles of copula theory.* CRC Press, Boca Raton, FL, 2016.

[60] P.P.B. Eggermont and V.N. LaRiccia, *Maximum penalized likelihood estimation., Vol. I.* Springer Series in Statistics. Springer-Verlag, New York, 2001. Density estimation.

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

[62] U. Einmahl and D.M. Mason, *Uniform in bandwidth consistency of kernel-type function estimator*s, Ann. Statist. **33** (3), 1380–1403, 2005.

[63] L. Faivishevsky and J. Goldberger, Ica based on a smooth estimation of the differential entropy. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems* **21**., Curran Associates, Inc., 2008.

[64] J. Fan and I. Gijbels, *Local polynomial modelling and its applications*, volume 66 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London, 1996.

[65] J.D. Fermanian, D. Radulović and M. Wegkamp, *Asymptotic total variation tests for copulas,* Bernoulli **21** (3), 1911–1945, 2015.

[66] A. Földes and L. Rejtő, *A LIL type result for the product limit estimator*, Z. Wahrsch. Verw. Gebiete **56** (1), 75–86, 1981.

[67] E.W. Frees, *Infinite order U-statistics*, Scand. J. Statist. **16** (1), 29–45, 1989.

[68] E.W. Frees and E.A. Valdez, *Understanding relationships using copulas*, N. Am. Actuar. J. **2** (1), 1–25, 1998.

[69] J. Galambos, *Order statistics of samples from multivariate distributions*, J. Amer. Statist. Assoc. **70** (351, part 1), 674–680, 1975.

[70] J. Gao and I. Gijbels, *Bandwidth selection in nonparametric kernel testing*, J. Amer. Statist. Assoc. **103** (484), 1584–1594, 2008.

[71] C. Genest, A.K. Nikoloulopoulos, L.-P. Rivest and M. Fortin, *Predicting dependent binary outcomes through logistic regressions and meta-elliptical copulas*, Braz. J. Probab. Stat. **27** (3), 265–284, 2013.

[72] S. Ghosal, A. Sen and A.W. van der Vaart, *Testing monotonicity of regression*, Ann. Statist. **28** (4), 1054–1082, 2000.

[73] R.D. Gill and S. Johansen, *A survey of product-integration with a view toward application in survival analysis*, Ann. Statist. **18** (4), 1501–1555, 1990.

[74] E. Giné, V. Koltchinskii and J. Zinn, *Weighted uniform consistency of kernel density estimators*, Ann. Probab. **32** (3B):2570–2605, 2004.

[75] E. Giné and D. M. Mason, *Laws of the iterated logarithm for the local U-statistic process*, J. Theoret. Probab. **20** (3), 457–485, 2007.

[76] E. Giné and D.M. Mason, *On local U-statistic processes and the estimation of densities of functions of several sample variables*, Ann. Statist. **35** (3), 1105–1145, 2007.

[77] E.J. Gumbel, *Bivariate exponential distributions*, Journal of the American Statistical Association **55** (292), 698–707, 1960.

[78] L. Györfi, M. Kohler, A. Krzyżak and H. Walk, *A distribution-free theory of nonparametric regression*, Springer Series in Statistics, Springer-Verlag, New York, 2002.

[79] P. Hall, *Asymptotic properties of integrated square error and cross-validation for kernel estimation of a regression function*, Z. Wahrsch. Verw. Gebiete **67** (2), 175–196, 1984.

[80] P. Hall and J.S. Marron, *Estimation of integrated squared density derivatives*, Statist. Probab. Lett. **6** (2), 109–115, 1987.
[81] P.R. Halmos. The theory of unbiased estimation, Ann. Math. Statistics **17**, 34–43, 1946.
[82] M. Harel and M.L. Puri, *Conditional U-statistics for dependent random variables,* J. Multivariate Anal. **57** (1), 84–100, 1996.
[83] C. Heilig and D. Nolan, *Limit theorems for the infinite-degree U-process,* Statist. Sinica **11** (1), 289–302, 2001.
[84] W. Hoeffding, *A class of statistics with asymptotically normal distribution,* Ann. Math. Statistics **19**, 293–325, 1948.
[85] M. Hollander, D. Park and F. Proschan, *Testing whether new is better than used of a specified age, with randomly censored data,* Canad. J. Statist. **13** (1), 45–52, 1985.
[86] J.L. Horowitz and V.G. Spokoiny, *An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative,* Econometrica **69** (3), 599–631, 2001.
[87] J.Hüsler and R.D. Reiss, *Maxima of normal random vectors: between independence and complete dependence,* Statist. Probab. Lett. **7** (4), 283–286, 1989.
[88] P. Janssen, J. Swanepoel and N. Veraverbeke, *Bernstein estimation for a copula derivative with application to conditional distribution and regression functionals*, TEST **25** (2), 351–374, 2016.
[89] H. Joe, *Multivariate models and dependence concepts*, volume 73 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London, 1997.
[90] H. Joe, *Dependence modeling with copulas*, volume 134 of *Monographs on Statistics and Applied Probability*, CRC Press, Boca Raton, FL, 2015.
[91] E. Joly and G. Lugosi, *Robust estimation of U-statistics,* Stochastic Process. Appl. **126** (12), 3760–3773, 2016.
[92] M.C. Jones and D.F. Signorini, *A comparison of higher-order bias kernel density estimators.* J. Amer. Statist. Assoc. **92** (439), 1063–1073, 1997.
[93] E.L. Kaplan and P. Meier, *Nonparametric estimation from incomplete observations,* J. Amer. Statist. Assoc. **53**, 457–481, 1958.
[94] M. Kohler, K. Máthé and M. Pintér, *Prediction from randomly right censored data*, J. Multivariate Anal. **80** (1), 73–100, 2002.
[95] N. Kolev and D. Paiva, *Copula-based regression models: a survey*, J. Statist. Plann. Inference **139** (11), 3847–3856, 2009.
[96] A.N. Kolmogorov and V.M. Tihomirov, *ε-entropy and ε-capacity of sets in functional space*, Amer. Math. Soc. Transl. (2) **17**, 277–364, 1961.
[97] M.R. Kosorok, *Introduction to empirical processes and semiparametric inference*, Springer Series in Statistics, Springer, New York, 2008.
[98] F. Lad, G. Sanfilippo and G. Agrò, *Extropy: complementary dual of entropy*, Statist. Sci. **30** (1), 40–58, 2015.
[99] A.J. Lee, *U-statistics*, volume 110 of *Statistics: Textbooks and Monographs*, Marcel Dekker, Inc., New York, 1990, Theory and practice.
[100] S. Lee, O. Linton and Y.J. Whang, *Testing for stochastic monotonicity*, Econometrica **77** (2), 585–602, 2009.
[101] Y.K. Leong and E.A. Valdez, *Claims prediction with dependence using copula models*, North American Actuarial Journal, 2005.
[102] Q. Liu, J. Lee and M. Jordan, *A kernelized stein discrepancy for goodness-of-fit tests*, In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of Proceedings of Machine Learning Research, pages 276–284, New York, New York, USA, 20–22 Jun 2016. PMLR.
[103] B. Maillot and V. Viallon, *Uniform limit laws of the logarithm for nonparametric estimators of the regression function in presence of censored data*, Math. Methods Statist.

18 (2), 159–184, 2009.

[104] D.M. Mason, *Proving consistency of non-standard kernel estimators*, Stat. Inference Stoch. Process. **15** (2), 151–176, 2012.

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

[106] A.J. McNeil, R. Frey and P. Embrechts, *Quantitative risk management*, Princeton Series in Finance. Princeton University Press, Princeton, NJ, revised edition, 2015. Concepts, techniques and tools.

[107] È.A. Nadaraja, *On a regression estimate*, Teor. Verojatnost. i Primenen. **9** 157–159, 1964.

[108] R.B. Nelsen, *An introduction to copulas*, Springer Series in Statistics, Springer, New York, second edition, 2006.

[109] H. Noh, A. El Ghouch and T. Bouezmarni, *Copula-based regression estimation and inference*, J. Amer. Statist. Assoc. **108** (502), 676–688, 2013.

[110] D. Nolan and D. Pollard, *U-processes: rates of convergence*, Ann. Statist. **15** (2), 780–799, 1987.

[111] M. Omelka, I. Gijbels and N. Veraverbeke, I*mproved kernel estimation of copulas: weak convergence and goodness-of-fit testing*, Ann. Statist. **37** (5B), 3023–3058, 2009.

[112] E. Parzen, *On estimation of a probability density function and mode*, Ann. Math. Statist. **33**, 1065–1076, 1962.

[113] W. Peng, T. Coleman and L. Mentch, *Rates of convergence for random forests via generalized U-statistics*, Electron. J. Stat. **16** (1), 232–292, 2022.

[114] D. Pollard, *Convergence of stochastic processes*, Springer Series in Statistics, Springer-Verlag, New York, 1984.

[115] B.L.S. Prakasa Rao and A. Sen, *Limit distributions of conditional U-statistics*, J. Theoret. Probab. **8** (2), 261–301, 1995.

[116] M. Rachdi and P. Vieu, *Nonparametric regression for functional data: automatic smoothing parameter selection*, J. Statist. Plann. Inference **137** (9), 2784–2801, 2007.

[117] W. Rejchel, *On ranking and generalization bounds*, J. Mach. Learn. Res. **13**, 1373–1392, 2012.

[118] M. Rosenblatt, *Remarks on some nonparametric estimates of a density function*, Ann. Math. Statist. **27**, 832–837, 1956.

[119] A. Schick, Y. Wang and W. Wefelmeyer, *Tests for normality based on density estimators of convolutions*, Statist. Probab. Lett. **81** (2), 337–343, 2011.

[120] D. W. Scott, *Multivariate density estimation*, Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, second edition, 2015. Theory, practice, and visualization.

[121] J. Segers, *Asymptotics of empirical copula processes under non-restrictive smoothness assumptions*, Bernoulli **18** (3), 764–782, 2012.

[122] A. Sen, U*niform strong consistency rates for conditional U-statistics*, Sankhyā Ser. A **56** (2), 179–194, 1994.

[123] H. L. Shang, *Bayesian bandwidth estimation for a functional nonparametric regression model with mixed types of regressors and unknown error density*, J. Nonparametr. Stat. **26** (3), 599–615, 2014.

[124] A. Shemyakin and A. Kniazev, *Introduction to Bayesian estimation and copula models of dependence.* John Wiley & Sons, Inc., Hoboken, NJ, 2017.

[125] R.P. Sherman, The limiting distribution of the maximum rank correlation estimator, Econometrica **61** (1), 123–137, 1993.

[126] R.P. Sherman, *Maximal inequalities for degenerate U-processes with applications to optimization estimators*, Ann. Statist. **22** (1), 439–459, 1994.

[127] B.W. Silverman, *Distances on circles, toruses and spheres.* J. Appl. Probability **15** (1), 136–143, 1978.

[128] B.W. Silverman, *Density estimation for statistics and data analysis.* Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1986.

[129] A. Sklar, *Fonctions de répartition à n dimensions et leurs marges*, Publ. Inst. Statist. Univ. Paris **8**, 229–231, 1959.

[130] A. Sklar, *Random variables, joint distribution functions, and copulas*, Kybernetika (Prague) **9**, 449–460, 1973.

[131] Y. Song, X. Chen and K. Kato, A*pproximating high-dimensional infinite-order U-statistics: statistical and computational guarantees*, Electron. J. Stat. **13** (2), 4794–4848, 2019.

[132] I. Soukarieh and S. Bouzebda, *Exchangeably weighted bootstraps of general markov U-process*, Mathematics **10** (20), 1–42, 2022.

[133] I. Soukarieh and S. Bouzebda, *Renewal type bootstrap for increasing degree U-process of a Markov chain*, J. Multivariate Anal. **195**, Paper No. 105143, 2023.

[134] W. Stute, *Conditional empirical processes*, Ann. Statist. **14** (2), 638–647, 1986.

[135] W. Stute, *Conditional U-statistics*, Ann. Probab. **19** (2), 812–825, 1991.

[136] W. Stute, *Almost sure representations of the product-limit estimator for truncated data*, Ann. Statist. **21** (1), 146–156, 1993.

[137] W. Stute, *L^p-convergence of conditional U-statistics*, J. Multivariate Anal. **51** (1), 71–82, 1994.

[138] W. Stute, *Universally consistent conditional U-statistics*, Ann. Statist. **22** (1), 460–473, 1994.

[139] W. Stute, Symmetrized NN-conditional U-statistics, In *Research developments in probability and statistics*, pages 231–237. VSP, Utrecht, 1996.

[140] A. Tenbusch, *Nonparametric curve estimation with Bernstein estimates*, Metrika **45** (1), 1–30, 1997.

[141] A. van der Vaart, *New Donsker classes*, Ann. Probab. **24** (4), 2128–2140, 1996.

[142] A. van der Vaart and J.A. Wellner, *Weak convergence and empirical processes.* Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics.

[143] R. von Mises, *On the asymptotic distribution of differentiable statistical functions*, Ann. Math. Statistics **18**, 309–348, 1947.

[144] M.P. Wand and M.C. Jones, *Kernel smoothing*, volume 60 of *Monographs on Statistics and Applied Probability*, Chapman and Hall, Ltd., London, 1995.

[145] G.S. Watson, *Smooth regression analysis*, Sankhyā Ser. A **26**, 359–372, 1964.

[146] Z. Wei and D. Kim, *On multivariate asymmetric dependence using multivariate skew-normal copula-based regression*, Internat. J. Approx. Reason. **92**, 376–391, 2018.