mathematics

Article

Uniform Consistency for Functional Conditional *U*-Statistics Using Delta-Sequences

Salim Bouzebda [1,*,†], Amel Nezzal [1,†] and Tarek Zari [2]

[1] LMAC (Laboratory of Applied Mathematics of Compiègne), Université de Technologie de Compiègne, 57 Avenue de Landshut CS 60319, CEDEX, 60203 Compiegne, France
[2] Laboratory MAEGE, FSJES Ain-Sebaa, University Hassan II, Casablanca 20000-20200, Morocco
* Correspondence: salim.bouzebda@utc.fr
† These authors contributed equally to this work.

Abstract: *U*-statistics are a fundamental class of statistics derived from modeling quantities of interest characterized by responses from multiple subjects. *U*-statistics make generalizations the empirical mean of a random variable X to the sum of all k-tuples of X observations. This paper examines a setting for nonparametric statistical curve estimation based on an infinite-dimensional covariate, including Stute's estimator as a special case. In this functional context, the class of "delta sequence estimators" is defined and discussed. The orthogonal series method and the histogram method are both included in this class. We achieve almost complete uniform convergence with the rates of these estimators under certain broad conditions. Moreover, in the same context, we show the uniform almost-complete convergence for the nonparametric inverse probability of censoring weighted (I.P.C.W.) estimators of the regression function under random censorship, which is of its own interest. Among the potential applications are discrimination problems, metric learning and the time series prediction from the continuous set of past values.

Keywords: nonparametric estimation; regression-type models; *U*-statistics; conditional distribution; functional estimation; delta sequences; kernel estimation; machine learning problems

MSC: 60F05; 60G15; 60G10; 62G05; 62G07; 62H12

Citation: Bouzebda, S.; Nezzal, A.; Zari, A. Uniform Consistency for Functional Conditional *U*-Statistics Using Delta-Sequences. *Mathematics* **2023**, *11*, 161. https://doi.org/10.3390/math11010161

Academic Editor: Manuel Alberto M. Ferreira

Received: 28 November 2022
Revised: 16 December 2022
Accepted: 21 December 2022
Published: 28 December 2022

1. Introduction

The regression problem has been studied by statisticians and probability theorists for many years to keep up with the various problems and topics brought up by technological and computational advances, resulting in the creation of many advanced and complex techniques. Among the problems addressed are modeling, estimation method applications, and tests. In this paper, we are interested in nonparametric regression estimation. Unlike the parametric framework, where one must estimate a finite number of parameters based on a specified structural model a priori, nonparametric estimation does not require any specific structure; instead, it allows the data to speak for themselves. However, as natural drawbacks, nonparametric procedures are more susceptible to estimation biases and losses in convergence rates than parametric methods. Since their introduction, kernel nonparametric function estimation approaches have garnered a significant amount of attention; for references to research literature and statistical applications in this area, consult [1–7] and the references therein. Popular as they may be, they represent only one of the numerous methods for developing accurate function estimators. These include nearest-neighbor, spline, neural network, and wavelet approaches. In addition, these techniques have been applied to a vast array of data types. This article will focus on constructing consistent estimators for the conditional *U*-statistics for functional data based on the delta sequence. The theory of *U*-statistics and *U*-processes, which was initially introduced in the seminal work of [8], has garnered a significant amount of interest over the course of the last few decades as a result of the diverse applications to which it has been applied. U-statistics can be utilized to solve complex statistical problems. Among the examples are nonparametric regression, density estimation, and goodness-of-fit tests. Furthermore, U-statistics contribute to the study of estimators with various degrees of smoothness (including function estimators). [9], for instance, analyzes the product limit estimator for truncated data applying a.s. uniform bounds for $\mathbb{P}$-canonical U-processes. [10] introduces two new normality tests based on U-processes. Using the findings of [11–13] provided new normality tests that used as test statistics weighted L_1-distances between the standard normal density and local U-statistics that were based on standardized observations. These tests were used to determine whether or not the data was normally distributed. Ref. [14] worked on the estimate of the mean of multivariate functions under the premise of possibly heavy-tailed distributions and presented the U-based median-of-means. U-processes are also necessary for a wide variety of statistical applications, such as the examination of qualitative aspects of functions in nonparametric statistics [15,16] as well as establishing limiting distributions of M-estimators (see, for example, [17–19]). In [20] the authors consider the problem of detecting distributional changes in a sequence of high dimensional data by using the weighted cumulative sums of U-statistics stemming from L_p norms. In [21], the authors proposed tests based on U-statistics for testing the equality of marginal density functions. In the paper [22], the following problem is considered: Is it possible, given a sample of random variables that are independent, identically distributed, and have a finite variance, to build an estimator of the unknown mean that performs almost as well as if the data were normally distributed? The argument that was presented in the previous work is based on a new deviation inequality for the U-statistics of order that is permitted to rise with sample size. This inequality is the most important part of the argument. The first asymptotic results for the scenario in which the underlying random variables are assumed to be independent and distributed in an identical fashion were presented by [23,24] and [8] (amongst others), respectively. In contrast, the asymptotic results under weak dependency assumption were demonstrated in [25], in [26] or more recently in [27] and in more general settings in [28–31]. The interested reader may refer to [17,32] for an excellent collection of references on U-statistics and U-processes. We also refer to [19] for a profound understanding of the theory of U-processes.

In the present work, we consider the conditional U-statistics introduced by [33], that can be considered as generalizations of the Nadaraya-Watson ([34,35]) regression function estimates. To be more precise, let us consider the sequence of independent and identically distributed random vectors $\{(\mathbf{X}_i,\mathbf{Y}_i), i\in\mathbb{N}^*\}$ with $\mathbf{X}_i\in\mathbb{R}^d$ and $\mathbf{Y}_i\in\mathbb{R}^{d'}$, $d,d'\geq 1$. Let $\varphi:\mathbb{R}^{d'k}\to\mathbb{R}$ denote a measurable function. Within the scope of this work, our primary focus is on the estimation of the conditional expectation or regression function, as follows:

$$r^{(k)}(\varphi,\mathbf{t}) \quad = \quad \mathbb{E}(\varphi(\mathbf{Y}_1,\ldots,\mathbf{Y}_k)\mid(\mathbf{X}_1,\ldots,\mathbf{X}_k)=\mathbf{t}),\ \text{ for }\ \mathbf{t}\in\mathbb{R}^{dk},$$

whenever it exists, i.e., $\mathbb{E}(|\varphi(\mathbf{Y}_1,\ldots,\mathbf{Y}_k)|)<\infty$. Now, we are going to present a kernel function $K:\mathbb{R}^d\to\mathbb{R}$ with support contained in $[-B,B]^d$, $B>0$, fulfilling:

$$\sup_{\mathbf{x}\in\mathbb{R}^d}|K(\mathbf{x})|=:\kappa<\infty \ \text{ and } \ \int K(\mathbf{x})d\mathbf{x}=1. \tag{1}$$

Ref. [33] established a new category of estimators for $r^{(k)}(\varphi,\mathbf{t})$, called conditional U-statistics, that is defined for each $\mathbf{t}\in\mathbb{R}^{dk}$ to be:

$$\widehat{r}_n^{(k)}(\varphi,\mathbf{t};h_n)=\frac{\displaystyle\sum_{(i_1,\ldots,i_k)\in I(k,n)}\varphi(Y_{i_1},\ldots,Y_{i_k})K\left(\frac{\mathbf{t}_1-\mathbf{X}_{i_1}}{h_n}\right)\cdots K\left(\frac{\mathbf{t}_k-\mathbf{X}_{i_k}}{h_n}\right)}{\displaystyle\sum_{(i_1,\ldots,i_k)\in I(k,n)}K\left(\frac{\mathbf{x}_1-\mathbf{X}_{i_1}}{h_n}\right)\cdots K\left(\frac{\mathbf{x}_k-\mathbf{X}_{i_k}}{h_n}\right)}, \tag{2}$$

where:
$$I(k,n) = \left\{\mathbf{i} = (i_1, \ldots, i_k) : 0 \le i_j \le n \quad \text{and} \quad i_j \ne i_r \quad \text{if} \quad j \ne r\right\},$$
is the set of all k-tuples of different integers between 1 and n and $\{h_n\}_{n\ge1}$ is a sequence of positive constants that, at a certain rate, converge to the value zero, $nh_n^{dk} \to \infty$. In the particular case $k = 1$, the $r^{(k)}(\varphi, \mathbf{t})$ reduces to $r^{(1)}(\varphi, \mathbf{t}) = \mathbb{E}(\varphi(\mathbf{Y})|\mathbf{X} = \mathbf{t})$ and the estimator developed by Stute is now known as the Nadaraya-Watson estimator $r^{(1)}(\varphi, \mathbf{t})$, given by:
$$\widehat{r}_n^{(1)}(\varphi, \mathbf{t}, h_n) = \sum_{i=1}^{n} \varphi(\mathbf{Y}_i) K\left(\frac{\mathbf{X}_i - \mathbf{t}}{h_n}\right) \Big/ \sum_{i=1}^{n} K\left(\frac{\mathbf{X}_i - \mathbf{t}}{h_n}\right).$$
The work of [36] focused on the estimation of the rate of the uniform convergence in $\mathbf{t}$ of $\widehat{r}_n^{(k)}(\varphi, \mathbf{t}; h_n)$ to $r^{(k)}(\varphi, \mathbf{t})$. In [37], the limit distributions of $\widehat{r}_n^{(k)}(\varphi, \mathbf{t}; h_n)$ are analyzed and compared to those produced by Stute. Under suitable mixing conditions, ref. [38] extend the results of [33] to weakly dependent data and uses their findings to evaluate the Bayes risk consistency of the corresponding discriminating rules.

As alternatives to the standard kernel-type estimators, [39] presented symmetrized nearest neighbor conditional U-statistics. This work has observed a major advancement because of the contributions of [40], where a far more strong version of consistency can be found; to be specific, 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 $\widehat{r}_n^{(k)}(\varphi, \mathbf{t}; h_n)$. Additionally, uniform consistency is achieved across $\varphi \in \mathscr{F}$ for a suitably restricted class of functions $\mathscr{F}$, extended in [41–44] and [45]. The key component of their findings is the local conditional U-process studied in [11].

The case of functional data is the primary focus of this research. We present an excerpt from [46]: "Functional data analysis (FDA) is a branch of statistics concerned with analyzing infinite-dimensional variables such as curves, sets, and images. It has undergone phenomenal growth over the past 20 years, stimulated partly by significant data collection technology advances that have brought about the "Big Data" revolution. Often perceived as a somewhat arcane area of research at the turn of the century, FDA is now one of the most active and relevant fields of investigation in data science". For an introduction to the subject of FDA, we refer to the books of [47,48], which contain different case studies in economics, archaeology, criminology, and neurophysiology, as well as fundamental analysis techniques. It is important to note that the extension of probability theory to random variables with values in normed vector spaces (such as Banach and Hilbert spaces), in conjunction with extensions of certain classical asymptotic limit theorems, predates the recent literature on functional data. This fact can be demonstrated by tracing back the history of the subject (see for instance, [49]). Ref. [50] investigated density and mode estimation for data with values in a normed vector space. At the same time, he brought attention to the issue of the curse of dimensionality, which affects functional data, and he suggested potential remedies to the problem. In the context of regression estimation, ref. [48] considered the nonparametric models. We may refer also to [51–53]. Recently, Modern theory has been applied to the treatment of functional data. For instance, ref. [54] provided the consistency rates of several functionals of the conditional distribution, such as the regression function, the conditional cumulative distribution, and the conditional density, amongst others, uniformly over a subset of the explanatory variable. Other examples include conditional density, which is a measure of the density of the conditional distribution, and conditional cumulative distribution, which is a measure of the conditional cumulative distribution. [55] also investigated the consistency rates for some functionals nonparametric models, such as the regression function, the conditional distribution, the conditional density, and the conditional hazard function, uniformly in bandwidth (UIB consistency) extended to the ergodic setting by [56]. In the paper [57], the topic of local linear estimation of the regression function in the case when the regressor is functional was investigated, and the results indicated robust convergence (with rates) consistently across bandwidth parameters. In the work of [58], the k-nearest neighbors (kNN) estimate of the nonparametric regression model for heavy mixing of functional time series data was explored. Under some mild conditions, a uniform and practically perfect convergence rate of the k-nearest neighbors estimator was established. In the work [59], the authors offer a variety of solutions for limiting laws for the conditional mode in the functional setting for ergodic data; for some current references, see the following: [45,60–65].

We will consider a general method for functional estimation by using the delta sequences. Delta sequences (also called "approximate identities" or "summability kernels") arise in a wide variety of subfields within mathematics. Still, the applications that pertain to the theory of generalized functions are likely the most significant ones. The regularization of generalized functions is the major application for delta sequences. The proposed methods generalize several nonparametric estimation methods, including the kernel estimators given in (2) of [33]. To be more precise, the broad class of delta-sequence estimators includes the histogram estimators, Chentsov's projection estimators [66], and nearest-neighbor estimators, among others. Certain types of these sequences were already studied by [67], who called them "δ-function sequences". They established, among other things, the asymptotic unbiasedness and the asymptotic variance of estimators based on them but did not consider convergence rates. Ref. [68] obtained the rate of strong consistency and the rate of asymptotic bias for estimators associated with delta sequences arising from the Fejér kernel of the Fourier series. The delta sequence method of density estimation of [69] is extended to certain non-i.i.d. cases in [70], where it is assumed that the observations are taken from a stationary Markov process. Ref. [71] considered the delta-sequence estimator for the marginal distribution of a strictly stationary stochastic process satisfying some mixing conditions. In [72], the author investigated the local and global convergence rates of delta-sequence type estimators of the density function, its derivative, and its mode. Ref. [73] proved the uniform strong consistency of delta-sequence estimators. Ref. [74] partially generalized the usual nonparametric estimators of a regression function by using an estimator based on quasi-positive delta sequences. Ref. [75] considered a general nonparametric statistical curve estimation setting called the class of "fractional delta sequence estimators". Ref. [76] used the delta method to investigate the correlation model. [77] looked at the problem of estimating the density function of functional data with values in an infinite-dimensional separable Banach space using the method of delta sequences; for further information, we can also look into [78]. The copula estimation using the delta sequences methods is considered in [79]. The problem of the nonparametric minimax estimation of a multivariate density at a given point by the delta sequences was investigated in [80]. Ref. [81] used the delta sequence to propose an essential application to the classification problem of the value of the discrete random variable.

The goal of the current study is to present and investigate the delta sequences estimators for the conditional U-statistics for functional data, more specifically for random elements taking values in an infinite-dimensional separable Banach space, such as the space of continuous functions on the interval $[0, 1]$ endowed with the supremum norm. This will allow the delta sequences estimators for the conditional U-statistics for functional data to be utilized for functional data analysis. Examples of functional data that can appear in these spaces include stochastic processes with continuous sample paths on a finite interval associated with the supremum norm and stochastic processes whose sample paths are square-integrable on the real line. Both of these types of stochastic processes can occur on the real line. The dimensionality problem must be addressed in a nonparametric functional data analysis in two ways: first, by working with data that have an infinite number of dimensions, and second, by making universal assumptions about the infinite number of dimensions for the probability distribution of variables in nonparametric modeling. This structure's twofold infinity of dimensions is the basis for all subsequent developments in the discipline. Our previous work, delivered in the multivariate setting and cited as [82], is extended here in the present study. Although the concept behind our estimation approach is similar to that presented in [82] (containing the Stute estimator), we make allowances for the infinite dimensionality of the covariate. have determined the asymptotic characteristics of the multivariate delta sequence estimators, ref. [82]. Their findings do not directly apply to the current situation since we are working with a covariate with an unlimited number of dimensions. As a result, we must utilize other reasoning in our proofs to deal with the broader framework. These findings are beneficial in their own right, but they are also necessary for the inquiry being conducted in this work. To "simply" combine ideas from other publications would not be sufficient to solve the issue, as will be demonstrated in the following paragraphs. To be able to deal with delta sequence U-statistic estimators for functional data, you will need to resort to intricate mathematical derivations. Compared to the previous studies written on delta sequence estimators, the current paper considers the setting of an unbounded function φ, which adds a significant amount of complexity to the proof. The general assumptions that are required for the derivations of the asymptotic results for the conditional U-statistics delta sequence estimators are presented in this study.

The format of this article is structured as follows. Section 2 is devoted to introducing the delta sequences and the definitions we need in our work, where we introduce the new family of estimators. Section 3 gives the paper's main results concerning the uniform convergences. In Section 4, we present a significant application for the censored data context of its interest. In Section 5, we provide some applications, including the Kendall rank correlation coefficient in Section 5.1, the discrimination in Section 5.2, the the metric learning in Section 5.3 and the time series prediction from a continuous set of past values in Section 5.4. Some final observations and possible developments in the future are moved to Section 6. To maintain a smooth flow throughout the presentation, all proofs have been compiled in Section 7. A selection of significant technical findings is presented in Appendix A.

2. Preliminaries and Estimation Procedure

Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a probability space, $(\mathcal{X}, d(\cdot,\cdot))$ denote an infinite-dimensional separable Banach space equipped with a norm $\|.\|$ such that $d(u,v) = \|u-v\|$ and $\mathcal{B}$ be the σ-algebra of Borel subsets of $\mathcal{X}$. Let us consider a sequence $\{X_i, Y_i : i \geq 1\}$ of independent identically distributed random copies of the random element (X, Y), where X is a random element defined on $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in $(\mathcal{X}, \mathcal{B})$ and Y takes values in some abstract space $(\mathcal{Y}, \mathcal{B}')$. Ref. [83] introduced the functional conditional U-statistics when $\mathbf{x} \in \mathcal{X}^m$ some semi-metric space as a generalization of Stute's estimator by:

$$\widehat{r}_n^{(m)}(\varphi, \mathbf{x}; h_K) = \frac{\sum\limits_{(i_1,\ldots,i_m)\in I(m,n)} \varphi(Y_{i_1}, \ldots, Y_{i_m}) K\left(\frac{d(x_1, X_{i_1})}{h_K}\right) \cdots K\left(\frac{d(x_m, X_{i_m})}{h_K}\right)}{\sum\limits_{(i_1,\ldots,i_m)\in I(m,n)} K\left(\frac{d(x_1, X_{i_1})}{h_K}\right) \cdots K\left(\frac{d(x_m, X_{i_m})}{h_K}\right)}. \quad (3)$$

As we mentioned, the delta-sequences procedures can be considered a more general class, including kernel estimation techniques. Therefore, we can naturally obtain a more general class of functional conditional U-statistics by replacing the kernel $K(\cdot)$ in Equation (3) with positive delta sequences $\delta_m(\cdot,\cdot)$ (see Definition 1), which allows us to introduce the following conditional U-statistic for each $\mathbf{x} = (x_1, \ldots, x_k) \in \mathcal{X}^k$ and $\varphi : \mathcal{Y}^k \to \mathbb{R}$ a measurable function, by

$$\widehat{r}_n^{(k)}(\varphi,\mathbf{x};m_n) = \begin{cases} \dfrac{\sum_{(i_1,\ldots,i_k)\in I(k,n)} \varphi(Y_{i_1},\ldots,Y_{i_k})\delta_{m_n}(x_1,X_{i_1})\cdots\delta_{m_n}(x_k,X_{i_k})}{\sum_{(i_1,\ldots,i_k)\in I(k,n)} \delta_{m_n}(x_1,X_{i_1})\cdots\delta_{m_n}(x_k,X_{i_k})}, \\ \text{if} \quad \sum_{(i_1,\ldots,i_k)\in I(k,n)} \delta_{m_n}(x_1,X_{i_1})\cdots\delta_{m_n}(x_k,X_{i_k}) \neq 0, \\ \dfrac{n!}{(n-k)!k!} \sum_{(i_1,\ldots,i_k)\in I(k,n)} \varphi(Y_{i_1},\ldots,Y_{i_k}) \\ \text{if} \quad \sum_{(i_1,\ldots,i_k)\in I(k,n)} \delta_{m_n}(x_1,X_{i_1})\cdots\delta_{m_n}(x_k,X_{i_k}) = 0, \end{cases} \tag{4}$$

which we consider estimating the regression function

$$r^{(k)}(\varphi,\mathbf{x}) = \mathbb{E}(\varphi(Y_1,\ldots,Y_k) \mid (X_1,\ldots,X_k) = \mathbf{x}), \;\text{ for }\; \mathbf{x} \in \mathcal{X}^k, \tag{5}$$

whenever it exists, i.e, $\mathbb{E}(|\varphi(Y_1,\ldots,Y_k)|) < \infty$.

Remark 1. *It is worth noting that X may admit a probability density function $f(\cdot)$ in relation to a σ-finite measure μ on $(\mathcal{X},\mathcal{B})$ in such a way that (for instance, refer [77,78,84]):*

$$\mathbb{P}(X \in A) = \int_A f(x)\mu(dx), \;\textit{ for every }\; A \in \mathcal{B} \;\textit{ such that }\; 0 < \mu(A) < \infty. \tag{6}$$

The concept of this remark is elaborated upon in [77] and its references. We denote by $(\Omega,\mathcal{F},\mathbb{P})$ a probability space and a nondecreasing family of sub- σ-algebras of $\mathcal{F}$ is denoted by $\{\mathcal{F}_t\}_{t\geq 0}$. Let $\{W(t)\}_{t\geq 0}$ denote a standard Wiener process defined on $(\Omega,\mathcal{F},\mathbb{P})$ in such a way that W_t is $\mathcal{F}_t$-measurable. We highlight that the probability measure μ_W on the space $\mathcal{C}^0(0,T)$ is connected with a Borel σ-algebra generated by the supremum norm topology is induced by the standard Wiener process. Let $\{X(t)\}_{0\leq t\leq T}$ be a diffusion process defined the stochastic differential equation:

$$dX(t) = a(t,X(t))dt + b(t,X(t))dW(t),$$

where $X(0) = x_0$ for $0 \leq t \leq T$. By imposing some assumptions on the functions $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$, we can establish that the probability measure μ_X on the space $\mathcal{C}^0(0,T)$ induced by the process X is absolutely continuous with respect to the probability measure μ_W. In addition, applying Girsanov's Theorem permits the computation of the Radon-Nikodym derivative of μ_X with respect to μ_W. The probability density of X on the space $\mathcal{C}^0(0,T)$ is the μ_W derivative, for instance, see [85]. From this point of view, the main motivation leading to the analysis of functional data is the inference of stochastic processes; the reader is referred to [85,86]. For the purposes of drawing conclusions, we make the assumption that the entirety of the process can be observed. However, if the process can only be observed at discrete times, either on a tiny grid or when the data are sparse, then other approaches, such as parametric inference for discrete data, need to be devised. For instance, for the diffusion processes, these new methods are necessary (cf. [85,87,88]). Observe that if $A = B(x,\kappa)$ for $(x,\kappa) \in \mathcal{X} \times \mathbb{R}_+^*$, then (6) allows for the small ball probability to be considered.

Assume also that $S_{\mathcal{X}}$ is a pseudo-compact subset of $\mathcal{X}$ satisfying the following property: for any $\epsilon > 0$, there exists $t_\ell \in \mathcal{X}, 1 \leq \ell \leq d_n$ such that

$$S_{\mathcal{X}} \subset \mathcal{S}_n := \bigcup_{\ell=1}^{d_n} B(t_\ell,\epsilon), \tag{7}$$

and there exists $\kappa > 0$ such that $d_n \epsilon^\kappa$ is a constant $C > 0$. Here, the open ball with center t_ℓ and radius ϵ is denoted by $B(t_\ell, \epsilon)$.

It is worth mentioning that the hypothesis (7) is essential for assuming a geometrical link between the number d_n of balls and their radius ϵ. In addition, this condition is fulfilled in usual nonparametric problems when $\mathcal{X} = \mathbb{R}^p$ is endowed with the Euclidean metric on $\mathbb{R}^p$ (because $\kappa = p$ suffices). However, this topological characteristic does not hold for any abstract semi-metric space, as [89] explains. Before we can use the delta-sequences approach to estimate the value of the regression operator $r^{(k)}(\cdot)$ in the model (5), we must first have the following definition.

Definition 1. *A sequence of non-negative functions $\{\delta_{m_n}(x, y) : m_n \geq 1\} = \{\delta_m(x, y), m \geq 1\}$ defined on $\mathcal{X}^k \times \mathcal{X}^k$ is called a delta-sequence with respect to the measure μ if the following properties are satisfied:*

(C.1) *For each γ in such a way that $0 < \gamma \leq \infty$:*

$$\lim_{m \to +\infty} \sup_{\mathbf{x} \in S^k_{\mathcal{X}}} \left| \int_{\mathbb{B}(\mathbf{x}, \gamma)} \delta_m(\mathbf{x}, \mathbf{y}) \mu(d\mathbf{y}) - 1 \right| = 0, \tag{8}$$

where $\mathbb{B}(\mathbf{x}, \gamma) := \prod_{j=1}^{k} B(x_j, \gamma)$, for all $\mathbf{x} = (x_1, \ldots, x_k)$.

(C.2) *There exists a positive constant C_1, in such a way that*

$$\sup_{(\mathbf{x},\mathbf{y}) \in S^k_{\mathcal{X}} \times \mathcal{X}^k} \delta_m(\mathbf{x}, \mathbf{y}) \leq C_1 s_m < \infty, \tag{9}$$

where $0 < s_m \to \infty$ as $m \to \infty$ and $\lim_{m \to \infty} \dfrac{m}{s_m \log(m)} = \infty$.

(C.3) *There exist $C_2 > 0, \beta_1 > 0$ and $\beta_2 > 0$, in such a way that*

$$|\delta_m(\mathbf{x}_1, \mathbf{y}) - \delta_m(\mathbf{x}_2, \mathbf{y})| \leq C s_m^{\beta_2} d(\mathbf{x}_1, \mathbf{x}_2)^{\beta_1} \quad \text{for all } \mathbf{x}_1, \mathbf{x}_2, \mathbf{y} \in \mathcal{X}^k, \tag{10}$$

where

$$d(\mathbf{x}, \mathbf{y}) := \frac{1}{k} d(x_1, y_1) + \cdots + \frac{1}{k} d(x_k, y_k),$$

for all $\mathbf{x} = (x_1, \ldots, x_k)$ and $\mathbf{y} = (y_1, \ldots, y_k) \in \mathcal{X}^k$.

(C.4) *For any $\gamma > 0$:*

$$\lim_{m \to \infty} \sup_{\substack{\mathbf{x} \in S^k_{\mathcal{X}} \\ \mathbf{y} \in \bar{\mathbb{B}}(x, \gamma)}} \delta_m(\mathbf{x}, \mathbf{y}) d(\mathbf{x}, \mathbf{y}) = 0, \tag{11}$$

where the complement set of the open ball $\mathbb{B}(\mathbf{x}, \gamma)$ is denoted by $\bar{\mathbb{B}}(\mathbf{x}, \gamma)$.

Notice that the conditions **(C.1)–(C.4)** of Definition 1 are modelled after a similar set of conditions for kernel-type estimators. The condition **(C.2)** corresponds to the bound of δ_m over $S^k_{\mathcal{X}} \times \mathcal{X}^k$ whereas the condition **(C.3)** is pertains to the uniform Lipschitz property of $\delta_m(\mathbf{x}, \mathbf{y})$. Contrarily, the condition **(C.4)** is and it is not an assumption on the bound of $d(\mathbf{x}, \mathbf{y})$ over $S^k_{\mathcal{X}} \times S^k_{\mathcal{X}}$ but an assumption on the limiting behaviour of $\delta_m(\mathbf{x}, \mathbf{y}) d(\mathbf{x}, \mathbf{y})$ as $m \to +\infty$.

Proposition 1. *Let $\{\delta_{m,1}(x_1, y_1)\}, \ldots, \{\delta_{m,k}(x_k, y_k)\}$ each be non-negative delta-sequence with respect to the measure μ, then*

$$\delta_m(\mathbf{x}, \mathbf{y}) := \prod_{j=1}^{k} \delta_{m,j}(x_j, y_j), \tag{12}$$

is also a non-negative delta sequence.

This proposition is similar to Proposition 2.2 [69] when $\mathcal{X} = \mathbb{R}^d$, which means that the product of non-negative delta-sequences is also a positive delta sequence. The proposition provides a flexible way to construct delta sequences in high dimensions in a similar way to kernel type estimation. Unless otherwise specified, we will set

$$\delta_m(\mathbf{x},\mathbf{y}) := \prod_{j=1}^{k} \delta_{m_n}(x_j, y_j)$$

for all $\mathbf{x}$ and $\mathbf{y} \in \mathcal{X}^k$. This notation will unburden our results in the forthcoming theorems.

2.1. Examples of Delta Sequence

In this section, following the notation of [78], we provide guidelines for constructing and recovering some well-known estimators in literature.

Example 1. *Kernel estimator*
Let $\mathcal{X} = \mathcal{C}^0(0,1)$ denote the space of the real-valued continuous functions that vanishes at 0. Suppose that $\mathcal{X}$ is equipped with the uniform topology that is induced by the supremum norm, i.e., if $x \in \mathcal{C}^0(0,1)$ then x is continuous on $(0,1)$ with $x(0) = 0$ and that

$$\|x\| = \sup_{t\in(0,1)} |x(t)|.$$

The Wiener measure on the space $\mathcal{X}$ induced by the standard Wiener process is denoted by μ. Let us define

$$\delta_m(x,y) = \frac{1}{\mu(B(x,1/m))} \mathbb{1}_{B(x,1/m)}(y),$$

where as usual $\mathbb{1}_A$ denotes the indicator function of the set A. Set for all $\mathbf{x} = (x_1,\ldots,x_k)$ and $\mathbf{y} = \mathbf{x} = (y_1,\ldots,y_k) \in \mathcal{X}^k$

$$\delta_m(\mathbf{x},\mathbf{y}) = \prod_{j=1}^{k} \delta_m(x_j, y_j),$$

then, by Proposition 1, $\delta_m(\mathbf{x},\mathbf{y})$ is a non-negative delta sequence, and the conditional U-statistic is defined in this case by

$$\begin{aligned}
\widehat{r}_n^{(k)}(\varphi,\mathbf{x};m_n) &= \frac{\sum\limits_{\mathbf{i}\in I(k,n)} \varphi(Y_{i_1},\ldots,Y_{i_k}) \prod\limits_{j=1}^{k} \mathbb{1}_{B(x_j,1/m)}(X_{i_j})}{\sum\limits_{\mathbf{i}\in I(k,n)} \prod\limits_{j=1}^{k} \mathbb{1}_{B(x_j,1/m)}(X_{i_j})} \\
&= \frac{\sum\limits_{\mathbf{i}\in I(k,n)} \varphi(Y_{i_1},\ldots,Y_{i_k}) \prod\limits_{j=1}^{k} \mathbb{1}_{B(x_j,1)}\left(\dfrac{d(x_j,X_{i_j})}{1/m}\right)}{\sum\limits_{\mathbf{i}\in I(k,n)} \prod\limits_{j=1}^{k} \mathbb{1}_{B(x_j,1)}\left(\dfrac{d(x_j,X_{i_j})}{1/m}\right)},
\end{aligned}$$

*which can be considered as the naive kernel estimator of $r^{(k)}(\cdot)$. We can observe clearly that $\delta_m(\cdot,\cdot)$ in this example satisfies the condition **(C.1)**. In fact*

$$\begin{aligned}
&\lim_{m\to+\infty} \sup_{\mathbf{x}\in S_{\mathcal{X}}^k} \left| \int_{\mathbb{B}(\mathbf{x},\gamma)} \prod_{j=1}^{k} \frac{1}{\mu(B(x_j,1/m))} \mathbb{1}_{B(x_j,1)}\left(\frac{d(x_j,y_j)}{1/m}\right) \mu(dy_1)\cdots\mu(dy_k) - 1 \right| \\
&= \lim_{m\to+\infty} \sup_{\mathbf{x}\in S_{\mathcal{X}}^k} \left| \prod_{j=1}^{k} \frac{1}{\mu(B(x_j,1/m))} \mu\big(B(x_j,\gamma)\cap B(x_j,1/m)\big) - 1 \right|
\end{aligned}$$

$$= \lim_{m\to+\infty} \sup_{\mathbf{x}\in S^k_{\mathcal{X}}} \left| \prod_{j=1}^{k} \frac{1}{\mu(B(x_j, 1/m))} \mu\big(B\big(x_j, \min(\gamma, 1/m)\big)\big) - 1 \right|,$$

this quantity tends to zero when m is sufficiently large.

For a bandwidth h_n^x, that is a sequence of positive numbers, define

$$\delta_m(x, y) = \frac{1}{h_n^x} K_n(d(x, y)),$$

*where $K_n(\cdot)$ is a sequence of functions fullfilling **(C.1)–(C.4)**.*

Example 2. *Histogram estimator*
Let $\mathcal{P}_n = \{A_{n,j}, j \in J_n\}$ be a partition of the set F (cf. [90]), such that

$$|J_n| = m_n, \quad \max_{j\in J_n} \mu\big(A_{n,j}\big) \to 0 \quad \textit{and} \quad n \min_{j\in J_n} \mu\big(A_{n,j}\big) \to \infty \quad \textit{as } n \to \infty.$$

Denote

$$\delta_m(x, y) = \sum_{j\in J_n} \frac{1}{\mu\big(A_{n,j}\big)} \mathbb{1}_{A_{n,j}}(x) \mathbb{1}_{A_{n,j}}(y).$$

We can now construct the histogram and regressogram estimators in the conditional U-statistics framework by taking $\delta_m(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{k} \delta_m(x_i, y_i)$.

Example 3. *Orthogonal series estimator*
Let $\{e_p\}_{p\geq 1}$ be a complete orthonormal system of the space $\mathcal{X}$, comprising eigenfunctions of a compact operator in the square integral functions space ($L_2(\mathcal{X})$, say). Define

$$\delta_m(x, y) = \sum_{p=1}^{m} e_p(x) e_p(y) \quad \textit{for } x, y \in F.$$

As stated in [91], $\delta_m(\cdot, \cdot)$ in this case are delta sequences. Now using Proposition 1, we can observe that

$$\delta_m(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{k} \delta_m(x_i, y_i)$$

are also a positive delta sequences.

For more examples of delta sequences, we refer to [69,92].

2.2. Conditions and Comments

In order to study the consistency of the proposed estimator, let us first state the following conditions:

(C.5) We assume that $d_n = n^{\zeta}$ for $\zeta > 0$ and

$$\epsilon^{\beta_1} s_m^{\beta_2} < \sqrt{\frac{s_m \log(m)}{m}}.$$

(C.6) Suppose that $m \to \infty$ and that

$$\exists 0 < \tau < 1 \quad \text{in such a way that } n^{\tau} \leq m \leq n, \quad \text{for large } n.$$

(C.7) We assume the following usual boundedness condition:

$$\sup_{\mathbf{y}\in\mathcal{Y}^k} |\varphi(\mathbf{y})| = M < \infty.$$

(C.7′) The function φ is unbounded and fullfils for some $q > 2$:

$$\mu_q := \sup_{\mathbf{t}\in S^k_{\mathcal{X}}} \mathbb{E}(\varphi^q(\mathbf{Y})|\mathbf{X} = \mathbf{t}) < \infty.$$

(C.8) For every $\gamma \geq 0$:

$$\sup_{\mathbf{x}\in S^k_{\mathcal{X}}} \left| \int_{\bar{\mathbb{B}}(\mathbf{x},\gamma)} \delta_m(\mathbf{x},\mathbf{y})\mu(d\mathbf{y}) - 1 \right| = O(D_m),$$

where $D_m = \left\{ d(\mathbf{x},\mathbf{y}), \mathbf{x} \in S^k_{\mathcal{X}} \text{ and } \mathbf{y} \in \mathcal{X}^k, \text{ such that } \delta_m(\mathbf{x},\mathbf{y}) > 0 \right\} = o(1)$ as $m \to +\infty$.

(C.9) The regression operator $r^{(k)}(\varphi,\cdot)$ is Lipschitzian in the following sense: $\exists C_3 > 0$ in such a way that, for any $\mathbf{x}_1 \in S^k_{\mathcal{X}}$ and $\mathbf{x}_2 \in \mathcal{X}^k$, we have

$$|r^{(k)}(\varphi,\mathbf{x}_1) - r^{(k)}(\varphi,\mathbf{x}_2)| \leq C_3 d(\mathbf{x}_1,\mathbf{x}_2).$$

2.3. Comments on the Assumptions

Similar to conditions **(C.1)–(C.4)**, assumption **(C.5)** is also modelled after some kernel-type conditions, and it allows us to select β_1 and β_2 in condition **(C.3)**. Due to the infinite nature of the problem, additional constraints are required to achieve uniform consistency across the pseudo-compact set. Ref. [89], discussed the assumption (7). This condition holds trivially for any finite-dimensional Euclidean space and remains valid for projection-based metric spaces with infinite dimensions. Condition **(C.7)** concerning the boundedness of the function $\varphi(\cdot)$ is essential to establish exponential bounds, this, coupled with the technical condition **(C.6)**, allows us to obtain the almost complete convergence later in the proofs. Note that we can replace condition **(C.7)** with a more general one, that is, condition **(C.7′)**, to obtain the results when the function $\varphi(\cdot)$ is unbounded. Finally, to establish precise rates of almost complete convergence in the functional context, additional conditions related to the topological nature of the problem are required. Mainly the assumption **(C.8)** and **(C.9)**, where the latter condition concerning the Lipschitz property of the operator $r^{(k)}(\cdot)$ is standard when studying with uniform consistency.

Remark 2. *Note that the condition* ***(C.7)*** *can be replaced by more broad hypotheses at specific times of* $\mathbf{Y}$*, as shown in [93]. That is*

(C.7′)*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* $x \uparrow \infty$*,*

$$(i)\ x^{-s}\mathcal{M}(x) \downarrow; (ii)\ \ x^{-1}\mathcal{M}(x) \uparrow. \tag{13}$$

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:*

$$\mathbb{E}(\mathcal{M}(|\varphi(\mathbf{Y})|)) < \infty.$$

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 boundedness assumption on $\varphi(\cdot)$ can be substituted by a finite moment assumption **(C.7′)**, but doing so will add a significant amount of additional complexity to the proofs; for further information, check the most recent reference [42,45,62,94,95] for more details.*

3. Some Asymptotic Results

In this subsection, we will discuss the uniform consistency of the functional conditional U-statistic, which is defined by (4). First, let us provide basic notation

$$\begin{aligned} \mathbf{X} &:= (X_1, \ldots, X_k) \in \mathcal{X}^k, \quad \mathbf{Y} := (Y_1, \ldots, Y_k) \in \mathcal{Y}^k, \\ \mathbf{X}_i &:= (X_{i_1}, \ldots, X_{i_k}), \quad \mathbf{Y}_i := (Y_{i_1}, \ldots, Y_{i_k}), \\ G_{\varphi,\mathbf{x}}(\mathbf{X}, \mathbf{Y}) &:= \varphi(\mathbf{Y})\delta_m(\mathbf{x}, \mathbf{X}) \quad \text{for} \quad \mathbf{x} \in S_{\mathcal{X}}^k, \\ u_n(\varphi, \mathbf{x}, m_n) &= u_n^{(k)}(G_{\varphi,\mathbf{x}}) := \frac{(n-k)!}{n!} \sum_{i \in I(k,n)} G_{\varphi,\mathbf{x}}(\mathbf{X}_i, \mathbf{Y}_i). \end{aligned}$$

It is clear that, for all $\mathbf{x} \in \mathcal{X}^k$:

$$\widehat{r}_n^{(k)}(\varphi, \mathbf{x}; m_n) = \frac{u_n(\varphi, \mathbf{x}, m_n)}{u_n(1, \mathbf{x}, m_n)},$$

and $u_n(\varphi, \mathbf{x}, m_n)$ is a classical U-statistic with the U-kernel $G_{\varphi,\mathbf{x},m_n}(\mathbf{x}, \mathbf{y})$. Therefore, to establish the uniform consistency of $\widehat{r}_n^{(k)}(\varphi, \mathbf{x}; m_n)$ to $r^{(k)}(\varphi, \mathbf{x})$ we need to study the uniform consistency of $u_n(\varphi, \mathbf{x}, m_n)$ to $\mathbb{E}(u_n(\varphi, \mathbf{x}, m_n))$. In this case, we will be considering a suitable centering parameter different from the expectation $\mathbb{E}\left(\widehat{r}_n^{(k)}(\varphi, \mathbf{x}; m_n)\right)$; hence, we define:

$$\widehat{\mathbb{E}}\left(\widehat{r}_n^{(k)}(\varphi, \mathbf{x}; m_n)\right) = \frac{\mathbb{E}(u_n(\varphi, \mathbf{x}, m_n))}{\mathbb{E}(u_n(1, \mathbf{x}, m_n))}. \tag{14}$$

The notation and facts that are presented below should be included in the continuation of this discussion. For a kernel L of $k \geq 1$ variables, we define

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

Suppose that L is a function of $\ell \geq 1$ variables, symmetric in its entries. Then, the Hoeffding projections (see [8,19]) with respect to $\mathbb{P}$, for $1 \leq k \leq \ell$, are defined as

$$\pi_{k,\ell}L(x_1, \ldots, x_k) = (\Delta_{x_1} - \mathbb{P}) \times \cdots \times (\Delta_{x_k} - \mathbb{P}) \times \mathbb{P}^{\ell-k}(L),$$

and

$$\pi_{0,\ell}L = \mathbb{E}L(X_1, \ldots, X_\ell),$$

for some measures Q_i on S, we denote

$$Q_1 \cdots Q_k L = \int_{S^k} L(x_1, \ldots, x_k) dQ_1(x_1) \cdots dQ_k(x_k),$$

and Δ_x denote Dirac measure at point $x \in \mathcal{X}$. Then, the Hoeffding decomposition give the following

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

which is easy to check. For $L \in L_2\left(\mathbb{P}^\ell\right)$ this denotes an orthogonal decomposition and $\mathbb{E}(\pi_k L \mid X_2, \ldots, X_k) = 0$ for $k \geq 1$; which is, the kernels $\pi_{k,\ell}L$ are canonical for $\mathbb{P}$. Moreover, $\pi_{k,\ell}, k \geq 1$, are nested projections, that is, $\pi_{k,\ell} \circ \pi_{k',\ell} = \pi_{k,\ell}$ if $k \leq k'$, and

$$\mathbb{E}(\pi_{k,\ell}L)^2 \leq \mathbb{E}(L - \mathbb{E}L)^2 \leq \mathbb{E}L^2.$$

For example,

$$\pi_{1,\ell}h(x) = \mathbb{E}(h(X_1, \ldots, X_\ell) \mid X_1 = x) - \mathbb{E}h(X_1, \ldots, X_\ell).$$

Remark 3. *The function $G_{\varphi,\mathbf{x},m_n}$ is not necessarily symmetric; when we need to symmetrize them, we have:*

$$\overline{G}_{\varphi,\mathbf{x}}(\boldsymbol{x}, \boldsymbol{y}) := \frac{1}{k!}\sum_{\sigma \in I_k^k} G_{\varphi,\mathbf{x},m_n}(\boldsymbol{x}_\sigma, \boldsymbol{y}_\sigma) = \frac{1}{k!}\sum_{\sigma \in I_k^k} \varphi(\boldsymbol{y}_\sigma)\delta_{m_n}(\boldsymbol{x}_\sigma, \mathbf{y}_\sigma),$$

where $\boldsymbol{x}_\sigma = (x_{\sigma_1}, \ldots, x_{\sigma_k})$ and $\boldsymbol{y}_\sigma = (y_{\sigma_1}, \ldots, y_{\sigma_k})$. After symmetrization, the expectation

$$\mathbb{E}\big(\overline{G}_{\varphi,\mathbf{x},m_n}(\boldsymbol{x}, \boldsymbol{y})\big) = \mathbb{E}\big(G_{\varphi,\mathbf{x},m_n}(\boldsymbol{x}, \boldsymbol{y})\big),$$

and the U-statistic

$$u_n^{(k)}(G_{\varphi,\mathbf{x},m_n}) = u_n^{(k)}(\overline{G}_{\varphi,\mathbf{x},m_n}) := u_n(\varphi, \mathbf{x}, m_n)$$

do not change.

3.1. Uniform Consistency of Functional Conditional U-Statistics

Let (z_n) for $n \in \mathbb{N}$, be a sequence of real r.v.'s. We say that (z_n) converges almost-completely (a.co.) toward zero if, and only if, for all

$$\epsilon > 0, \sum_{n=1}^{\infty} \mathbb{P}(|z_n| > \epsilon) < \infty.$$

Moreover, we say that the rate of the almost-complete convergence of (z_n) toward zero is of order u_n (with $u_n \to 0$) and we write $z_n = O_{a.co.}(u_n)$ if, and only if, there exists $\epsilon > 0$ such that

$$\sum_{n=1}^{\infty} \mathbb{P}(|z_n| > \epsilon u_n) < \infty.$$

This kind of convergence implies both the almost-sure convergence and the convergence in probability. The following result concerns the uniform deviation of the estimate $u_n(\varphi, \mathbf{x}, m_n)$ with respect to $\mathbb{E}(u_n(\varphi, \mathbf{x}, m_n))$ when the function φ is bounded.

Theorem 1. *Under the conditions **(C.1)–(C.4)**, and if conditions **(C.5)**, **(C.7)** are satisfied, then we have:*

$$\sup_{\mathbf{x} \in S_{\mathcal{X}}^k} |u_n(\varphi, \mathbf{x}, m_n) - \mathbb{E}(u_n(\varphi, \mathbf{x}, m_n))| = O_{a.co}\left(\sqrt{\frac{s_m \log(m)}{m}}\right).$$

We present a more general result concerning the case when the function φ is unbounded in the sense of the condition **(C.7′)**. That being said, the preceding theorem constitutes an important step in the truncation method used in the proof of the following theorem.

Theorem 2. *Under the conditions **(C.1)–(C.4)**, and if conditions **(C.5)**, **(C.6)** and **(C.7′)** are satisfied, then we have:*

$$\sup_{\mathbf{x} \in S_{\mathcal{X}}^k} |u_n(\varphi, \mathbf{x}, m_n) - \mathbb{E}(u_n(\varphi, \mathbf{x}, m_n))| = O_{a.co}\left(\sqrt{\frac{s_m \log(m)}{m}}\right).$$

The following result handles the uniform deviation of the estimator $\widehat{r}_n^{(k)}(\varphi, \mathbf{x}; m_n)$ with respect to $\widehat{\mathbb{E}}\left[\widehat{r}_n^{(k)}(\varphi, \mathbf{x}; m_n)\right]$ in the case of the function φ is bounded or unbounded.

Theorem 3. *Under the conditions **(C.1)–(C.4)**, and if conditions **(C.5)**, **(C.6)** and condition **(C.7)** (or **(C.7′)**) are satisfied, then we have:*

$$\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\left|\widehat{r}_n^{(k)}(\varphi, \mathbf{x}; m_n) - \widehat{\mathbb{E}}\left[\widehat{r}_n^{(k)}(\varphi, \mathbf{x}; m_n)\right]\right| = O_{a.co}\left(\sqrt{\frac{s_m \log(m)}{m}}\right),$$

where $(s_m)_{m\in N^*}$ is a sequence of positive real numbers, in such a way that $m(s_m \log(m))^{-1} \to \infty$ as $n \to \infty$.

Theorem 4. *Under the conditions **(C.1)–(C.4)** and **(C.9)** , we have:*

$$\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\left|\widehat{\mathbb{E}}\left[\widehat{r}_n^{(k)}(\varphi, \mathbf{x}; m_n)\right] - r^{(k)}(\varphi, \mathbf{x})\right| \to 0.$$

The following corollary is more or less straightforward, given Theorems 3 and 4.

Corollary 1. *Under the conditions of Theorems 3 and 4 it follows that, as m tends to infinity:*

$$\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\left|\widehat{r}_n^{(k)}(\varphi, \mathbf{x}; m_n) - r^{(k)}(\varphi, \mathbf{x})\right| \to 0, \quad a.co.,$$

where $(s_m)_{m\in \mathbb{N}^*}$ is a sequence of positive real numbers, in such a way that $m(s_m \log(m))^{-1} \to \infty$ as $n \to \infty$.

3.2. Uniform Strong Consistency with Rates

This section is devoted to the uniform version with the rate of Theorem 1's result. More specifically, our objective is to obtain the uniform almost-complete convergence of $\widehat{r}_n^{(k)}(\cdot)$ on some subset $S_{\mathcal{X}}^k$ of $\mathcal{X}^k$ satisfying condition (7). In the following theorem, we establish the bias order.

Theorem 5. *Under the conditions **(C.1)–(C.4)**, and if conditions **(C.8)** and **(C.9)** are satisfied, then we have:*

$$\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\left|\widehat{\mathbb{E}}\left(\widehat{r}_n^{(k)}(\varphi, \mathbf{x}; m_n)\right) - r^{(k)}(\varphi, \mathbf{x})\right| = O(D_m). \tag{15}$$

The almost-complete convergence is then given by the corollary that follows, which uses a rate of $\widehat{r}_n^{(k)}(\cdot)$.

Corollary 2. *Under the conditions of Theorems 3 and 5 it follows that:*

$$\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\left|\widehat{r}_n^{(k)}(\varphi, \mathbf{x}; m_n) - r^{(k)}(\varphi, \mathbf{x})\right| = O(D_m) + O_{a.co}\left(\sqrt{\frac{s_m \log(m)}{m}}\right). \tag{16}$$

4. Conditional *U*-Statistics for Censored Data

Consider a triple $(Y, C, \mathbf{X})$ of random variables defined in $\mathbb{R} \times \mathbb{R} \times \mathscr{X}$. Here Y is the variable of interest, C is a censoring variable, and $\mathbf{X}$ is a concomitant variable. Throughout, we will use [96] 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

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

This type of censoring is commonly applied to the survival data collected during clinical trials as well as the failure time data collected during reliability studies, for example. To be more specific, the majority of statistical experiments end up producing incomplete samples, even when the conditions are carefully monitored. For instance, clinical data for surviving the majority of diseases are typically censored due to the presence of other competing risks to life that ultimately result in death. In the sequel, we impose the following assumptions upon the distribution of $(\mathbf{X}, Y)$. Denote by $\mathcal{I}$ a given compact set in $\mathscr{X}$ with nonempty interior and set, for any $\alpha > 0$,

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

For $-\infty < t < \infty$, set

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

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 $\mathscr{F}$ of real measurable functions defined on $\mathbb{R}$, and assume that $\mathscr{F}$ is of VC-type. We recall the regression function of $\psi(Y)$ evaluated at $\mathbf{X} = \mathbf{x}$, for $\psi \in \mathscr{F}$ and $\mathbf{x} \in \mathcal{I}_\alpha$, 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 that have recently gained popularity in the censored data literature (see [97,98]). 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

$$\Phi_\psi(y, c) = \frac{\mathbb{1}\{y \leq c\}\psi(y \wedge c)}{1 - G(y \wedge c)}. \tag{17}$$

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 $1 \leq i \leq n$. In addition, under Assumption **(I)** below

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

We have

$$\begin{aligned} r^{(1)}(\Phi_\psi, \mathbf{x}) &:= \mathbb{E}(\Phi_\psi(Y, C) \mid \mathbf{X} = \mathbf{x}) \\ &= \mathbb{E}\left\{\frac{\mathbb{1}\{Y \leq C\}\psi(Z)}{1 - G(Z)} \mid \mathbf{X} = \mathbf{x}\right\} \\ &= \mathbb{E}\left\{\frac{\psi(Y)}{1 - G(Y)} \mathbb{E}(\mathbb{1}\{Y \leq C\} \mid \mathbf{X}, Y) \mid \mathbf{X} = \mathbf{x}\right\} \\ &= r^{(1)}(\psi, \mathbf{x}). \end{aligned} \tag{18}$$

Therefore, every estimate of $r^{(1)}(\Phi\psi, \cdot)$ that can be constructed using completely observed data is also an estimate of $r^{(1)}(\psi, \cdot)$. This characteristic permits the natural application of the majority of statistical procedures known to produce estimates of the regression function in the uncensored case to the censored case. Estimates of the kernel type, for instance, are exceptionally straightforward to build. Set, for $\mathbf{x} \in \mathcal{I}$, $h \geq 0$, $1 \leq i \leq n$,

$$\overline{\omega}^{(1)}_{n,K,h,i}(\mathbf{x}) := \delta_{m_n}(\mathbf{x}, \mathbf{X}_i) \Big/ \sum_{j=1}^{n} \delta_{m_n}(\mathbf{x}, \mathbf{X}_j). \tag{19}$$

Making use of the Equations (17)–(19), whenever $G(\cdot)$ is known, we define the kernel estimator of $r^{(1)}(\psi, \cdot)$ by

$$\breve{r}_n^{(1)}(\psi, \mathbf{x}; h_n) = \sum_{i=1}^{n} \overline{\omega}^{(1)}_{n,K,h,i}(\mathbf{x}) \frac{\Delta_i \psi(Z_i)}{1 - G(Z_i)}. \tag{20}$$

Since the function $G(\cdot)$ is unknown, it is to be estimated. Let $G_n^*(\cdot)$ denote the Kaplan-Meier estimator of the function $G(\cdot)$ [99]. To be precise, 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

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

Given this notation, we will examine the next estimate of $r^{(1)}(\psi, \cdot)$

$$\breve{r}_n^{(1)*}(\psi, \mathbf{x}; h_n) = \sum_{i=1}^{n} \overline{\omega}^{(1)}_{n,K,h,i}(\mathbf{x}) \frac{\Delta_i \psi(Z_i)}{1 - G_n^*(Z_i)}, \tag{21}$$

the reader is invited to see the papers of [96,97]. The convention $0/0 = 0$ is used, 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 related 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, it is often impossible to estimate the function of the (conditional) law on the whole support (see [100]). Ref. [101] introduces a right-censored version of an unconditional U-statistic with a kernel of degree $m \geq 1$ based on the notion of a mean-preserving reweighting technique. Ref. [102] have demonstrated the almost sure convergence of multi-sample U-statistics under random censorship and presented an application by analyzing the consistency of a novel class of tests meant to evaluate distribution equality. Ref. [103] presented improvements to the traditional U-statistics to counteract potential biases caused by right-censoring of the outcomes and the existence of confounding factors. Ref. [104] suggested an alternative method for estimating the U-statistic by employing a substitution estimator of the conditional kernel given observed data. We also refer to [44,45,105]. To our best knowledge, estimating the conditional U-statistics in the censored data setting is a current open problem, and it gives the main motivation for the study of this section.

The function described by (17) has a natural expansion given by

$$\Phi_\psi(y_1, \ldots, y_k, c_1, \ldots, c_k) = \frac{\prod_{i=1}^{k} \{\mathbb{1}\{y_i \leq c_i\} \psi(y_1 \wedge c_1, \ldots, y_k \wedge c_m)}{\prod_{i=1}^{k} \{1 - G(y_i \wedge c_i)\}}.$$

We have an analogous relationship to (18) based on the formula:

$$\mathbb{E}(\Phi_\psi(Y_1, \ldots, Y_k, C_1, \ldots, C_k) \mid (\mathbf{X}_1, \ldots, \mathbf{X}_k) = \mathbf{x})$$

$$
\begin{aligned}
&= \mathbb{E}\left(\frac{\prod_{i=1}^{k}\{\mathbb{1}\{Y_i \leq C_i\}\psi(Y_1 \wedge C_1, \ldots, Y_k \wedge C_k)}{\prod_{i=1}^{k}\{1 - G(Y_i \wedge C_i)\}} \mid (\mathbf{X}_1, \ldots, \mathbf{X}_k) = \mathbf{x}\right) \\
&= \mathbb{E}\left(\frac{\psi(Y_1, \ldots, Y_k)}{\prod_{i=1}^{k}\{1 - G(Y_i)\}}\mathbb{E}\left(\prod_{i=1}^{k}\{\mathbb{1}\{Y_i \leq C_i\} \mid (Y_1, X_1), \ldots (Y_k, X_k)\right) \mid (\mathbf{X}_1, \ldots, \mathbf{X}_k) = \mathbf{x}\right) \\
&= \mathbb{E}(\psi(Y_1, \ldots, Y_k) \mid (\mathbf{X}_1, \ldots, \mathbf{X}_k) = \mathbf{x}).
\end{aligned}
$$

An analogue estimator to (4) in the censored situation is given by

$$
\breve{r}_n^{(k)}(\psi, \mathbf{x}; m_n) = \sum_{(i_1, \ldots, i_k) \in I(k,n)} \frac{\Delta_{i_1} \cdots \Delta_{i_k} \psi(Z_{i_1}, \ldots, Z_{i_k})}{(1 - G(Z_{i_1}) \cdots (1 - G(Z_{i_k}))} \overline{\omega}_{n,\delta,m_n,\mathbf{i}}^{(k)}(\mathbf{x}), \tag{22}
$$

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

$$
\overline{\omega}_{n,\delta,m_n,\mathbf{i}}^{(k)}(\mathbf{x}) = \frac{\delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k})}{\sum_{(i_1, \ldots, i_k) \in I(k,n)} \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k})}. \tag{23}
$$

The estimator we shall examine is provided by

$$
\breve{r}_n^{(k)*}(\psi, \mathbf{x}; m_n) = \sum_{(i_1, \ldots, i_k) \in I(k,n)} \frac{\Delta_{i_1} \cdots \Delta_{i_k} \psi(Z_{i_1}, \ldots, Z_{i_k})}{(1 - G_n^*(Z_{i_1}) \cdots (1 - G_n^*(Z_{i_k}))} \overline{\omega}_{n,\delta,m_n,\mathbf{i}}^{(k)}(\mathbf{x}). \tag{24}
$$

In a similar way as in [44], we arrive to the following conclusion.

Corollary 3. *Assume that the condition* (**I**) *and the assumptions of Theorems 3 and 5 are satisfied. Then, we have*

$$
\left|\breve{r}_n^{(k)*}(\psi, \mathbf{x}; m_n) - r^{(k)}(\varphi, \tilde{\mathbf{x}})\right| = O(D_m) + O_{a.co}\left(\sqrt{\frac{s_m \log(m)}{m}}\right), \quad a.s.
$$

This last result is a direct consequence of Corollary (2) and the law of iterated logarithm for $G_n^*(\cdot)$ obtained in [106] gives

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

At this point, we may refer to [44,45,105].

5. Applications

5.1. Kendall Rank Correlation Coefficient

To test the independence of one-dimensional random variables Y_1 and Y_2 [107] proposed a method based on the U-statistic K_n with the kernel function:

$$
\varphi((s_1, t_1), (s_2, t_2)) = \mathbb{1}_{\{(s_2 - s_1)(t_2 - t_1) > 0\}} - \mathbb{1}_{\{(s_2 - s_1)(t_2 - t_1) \leqslant 0\}}. \tag{25}
$$

Its rejection on the region is of the form $\{\sqrt{n}K_n > \gamma\}$. In this example, we consider a multivariate case. To test the conditional independence of $\boldsymbol{\xi}, \boldsymbol{\eta} : Y = (\boldsymbol{\xi}, \boldsymbol{\eta})$ given X, we propose a method based on the conditional U-statistic:

$$
\widehat{r}_n^{(2)}(\varphi, \mathbf{t}) = \frac{\sum_{i \neq j}^{n} \varphi(Y_i, Y_j) \delta_m(t_1, X_i) \delta_m(t_2, X_j)}{\sum_{i \neq j}^{n} \delta_m(t_1, X_i) \delta_m(t_2, X_j)},
$$

where $\mathbf{t} = (t_1, t_2) \in \mathbb{I} \subset \mathbb{R}^2$ and $\varphi(\cdot)$ is Kendall's kernel (25). Suppose that $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are d_1 and d_2-dimensional random vectors, respectively, and $d_1 + d_2 = d$. Furthermore, suppose that $\mathrm{Y}_1, \ldots, \mathrm{Y}_n$ are observations of $(\boldsymbol{\xi}, \boldsymbol{\eta})$, we are interested in testing:

$$\mathrm{H}_0 : \boldsymbol{\xi} \text{ and } \boldsymbol{\eta} \text{ are conditionally independent given } X. \text{ vs } \mathrm{H}_a : \mathrm{H}_0 \text{ is not true.} \tag{26}$$

Let $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2) \in \mathbb{R}^d$ such as $\|\mathbf{a}\| = 1$ and $\mathbf{a}_1 \in \mathbb{R}^{d_1}, \mathbf{a}_2 \in \mathbb{R}^{d_2}$, and $\mathrm{F}(\cdot), \mathrm{G}(\cdot)$ be the distribution functions of $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$, respectively. Suppose $F^{a_1}(\cdot)$ and $G^{a_2}(\cdot)$ to be continuous for any unit vector $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2)$ where $\mathrm{F}^{\mathbf{a}_1}(t) = \mathbb{P}\left(\mathbf{a}_1^\top \boldsymbol{\xi} < t\right)$ and $\mathrm{G}^{\mathbf{a}_2}(t) = \mathbb{P}\left(\mathbf{a}_2^\top \boldsymbol{\eta} < t\right)$ and $\mathbf{a}_1^{\mathrm{T}}$ means the transpose of the vector $\mathbf{a}_i, 1 \leqslant i \leqslant 2$. For $n = 2$, let $Y^{(1)} = \left(\boldsymbol{\xi}^{(1)}, \boldsymbol{\eta}^{(1)}\right)$ and $Y^{(2)} = \left(\boldsymbol{\xi}^{(2)}, \boldsymbol{\eta}^{(2)}\right)$ such as $\boldsymbol{\xi}^{(i)} \in \mathbb{R}^{d_1}$ and $\boldsymbol{\eta}^{(i)} \in \mathbb{R}^{d_2}$ for $i = 1, 2$, and :

$$\varphi^a\left(\mathrm{Y}^{(1)}, \mathrm{Y}^{(2)}\right) = \varphi\left(\left(\mathbf{a}_1^\top \boldsymbol{\xi}^{(1)}, \mathbf{a}_2^\top \boldsymbol{\eta}^{(1)}\right), \left(\mathbf{a}_1^\top \boldsymbol{\xi}^{(2)}, \mathbf{a}_2^\top \boldsymbol{\eta}^{(2)}\right)\right).$$

An application of Corollary 2 gives

$$\sup_{\mathbf{x} \in S_{\mathcal{X}}^2} \left| \widehat{r}_n^{(2)}(\varphi^a, \mathbf{x}; m_n) - r^{(2)}(\varphi^a, \mathbf{x}) \right| = O(D_m) + O_{a.co}\left(\sqrt{\frac{s_m \log(m)}{m}}\right). \tag{27}$$

5.2. Discrimination Problems

Now, we apply these findings to the discrimination problem outlined in Section 3 of [108], refer to also to [109]. We will employ a similar setup and notation. Let $\varphi(\cdot)$ be any function taking at most finitely many values, say $1, \ldots, M$. The sets

$$A_j = \{(\mathbf{y}_1, \ldots, \mathbf{y}_k) : \varphi(\mathbf{y}_1, \ldots, \mathbf{y}_k) = j\}, \;\; 1 \leq j \leq M$$

subsequently, produce a partition of the feature space. Predicting the value of $\varphi(\mathbf{Y}_1, \ldots, \mathbf{Y}_k)$ is equivalent to making a guess about which set will be in the partition to which $(\mathbf{Y}_1, \ldots, \mathbf{Y}_k)$ belongs. For any discrimination rule g, we have

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

where

$$\mathfrak{M}^j(\tilde{\mathbf{x}}) = \mathbb{P}(\varphi(\mathbf{Y}) = j \mid \mathbf{X} = \tilde{\mathbf{x}}), \;\; \tilde{\mathbf{x}} \in \mathbb{R}^d.$$

The inequality described above becomes equality if

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

$g_0(\cdot)$ is known as the Bayes rule, and the associated error probability

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

is called the Bayes risk. Each of the unknown $\mathfrak{M}^j$ functions can be reliably estimated using one of the techniques described in the prior sections. Let, for $1 \leq j \leq M$,

$$\mathfrak{M}_n^j(\tilde{\mathbf{x}}) = \frac{\sum_{(i_1, \ldots, i_k) \in I(k,n)} \mathbb{1}\{\varphi(Y_{i_1}, \ldots, Y_{i_k}) = j\} \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k})}{\sum_{(i_1, \ldots, i_k) \in I(k,n)} \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k})}, \tag{28}$$

Set
$$g_{0,n}(\tilde{\mathbf{x}}) = \arg\max_{1\le j\le M} \mathfrak{M}_n^j(\tilde{\mathbf{x}}).$$
Let us define
$$\mathbf{L}_n^* = \mathbb{P}(g_{0,n}(\mathbf{X}) \neq \varphi(\mathbf{Y})).$$
The discrimination rule $g_{0,n}(\cdot)$ is asymptotically Bayes' risk consistent
$$\mathbf{L}_n^* \to \mathbf{L}^*.$$
This is a consequence of the relation
$$|\mathbf{L}^* - \mathbf{L}_n^*| \le 2\mathbb{E}\left[\max_{1\le j\le M}\left|\mathfrak{M}_n^j(\mathbf{X}) - \mathfrak{M}^j(\mathbf{X})\right|\right].$$

5.3. Metric Learning

Metric learning seeks to adapt the metric to the data and has garnered a great deal of attention in recent years; for instance, see [110,111] for an account of metric learning and its applications. This is driven by a wide variety of applications, spanning from information retrieval via bioinformatics to computer vision as the primary source of inspiration. For the purpose of demonstrating the applicability of this idea, we will now discuss the metric learning problem for supervised classification as shown in [111]. Let us consider independent copies $(X_1, Y_1), \ldots, (X_n, Y_n)$ of a $\mathcal{X} \times \mathcal{Y}$ valued random couple (X, Y), where $\mathcal{X}$ is some feature space and $\mathcal{Y} = \{1, \ldots, C\}$, with $C \ge 2$ say, a finite set of labels. Let $\mathcal{D}$ be a set distance measures $D : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$. The purpose of metric learning in this context is, intuitively speaking, to identify a metric under which pairs of points with the same label are close to each other, while those with different labels are far away from each other. A natural way to characterize the risk associated with a metric D is as follows
$$R(D) = \mathbb{E}[\phi((1 - D(X, X') \cdot (2\mathbb{1}\{Y = Y'\} - 1))], \tag{29}$$
where $\phi(u)$ is a convex loss function upper bounding the indicator function $\mathbb{1}\{u \ge 0\}$, for instance, the hinge loss $\phi(u) = \max(0, 1 - u)$. To estimate $R(D)$, we consider the usual empirical estimator
$$R_n(D) = \frac{2}{n(n-1)} \sum_{1\le i<j\le n} \phi((D(X_i, X_j) - 1) \cdot (2\mathbb{1}\{Y_i = Y_j\} - 1)), \tag{30}$$
which is one sample U-statistic of degree two with kernel given by:
$$\varphi_{\mathrm{D}}((x, y), (x', y')) = \phi((D(x, x') - 1) \cdot (2\mathbb{1}\{y = y'\} - 1)).$$
The convergence to (29) of a minimizer of (30) has been studied in the frameworks of algorithmic stability ([112]), algorithmic robustness ([113]) and based on the theory of U-processes under appropriate regularization ([114]).

5.4. Time Series Prediction from Continuous Set of Past Values

Let $\{Z_n(t),\ t \in \mathbb{R}\}_{n\ge1}$ denote a sequence of processes with value in $\mathbb{R}$. Let s denote a fixed positive real number. In this model, we suppose that the process is observed from $t = 0$ until $t = t_{\max}$, and assume without loss of generality that $t_{\max} = nT + s < \tau$. The method ensures splitting the observed process into n fixed-length segments. Let us denote each piece of the process by
$$X_i = \{Z(t), (i-1)T \le t < iT\}.$$

The response value is therefore $Y_i = Z(iT+s)$, and this can be formulated as a regression problem:

$$\varphi(Z_1(\tau+s),\ldots,Z_k(\tau+s)) = r^{(k)}(Z_1(t),\ldots,Z_k(t)), \text{ for } \tau - T \le t < \tau. \quad (31)$$

provided that we make the assumption that a function of this kind, r, does not depend on i (which is satisfied if the process is stationary, for example). Because of this, when we get to time τ, we can use the following predictor, which is directly derived from our estimator, to make a prediction about the value that will be at time $\tau+s$

$$\widehat{r}_n^{(k)}(\varphi,\mathbf{z};m_n) = \frac{\sum_{(i_1,\ldots,i_k)\in I(k,n)} \varphi(Z_{i_1}(\tau+s),\ldots,Z_{i_k}(\tau+s))\delta_{m_n}(z_1,X_{i_1})\cdots\delta_{m_n}(z_k,X_{i_k})}{\sum_{(i_1,\ldots,i_k)\in I(k,n)} \delta_{m_n}(z_1,X_{i_1})\cdots\delta_{m_n}(z_k,X_{i_k})}$$

where $\mathbf{z} = (z_1,\ldots,z_k) = \{(Z_1(t),\ldots,Z_k(t)), \text{ for } \tau - T \le t < \tau\}$. Corollary 2 provides mathematical support for this nonparametric functional predictor and extends previous results in numerous ways in [48,78]. Notice that this modelization encompasses a wide variety of practical applications, as this procedure allows for the consideration of a large number of past process values without being affected by the curse of dimensionality.

5.5. Example of U-Kernels

Example 4. *Hoeffding's D From the symmetric kernel,*

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

We obtain Hoeffding's D statistic, which is a rank-based U-statistic of order 5.

Example 5 (Blum-Kiefer-Rosenblatt's R)**.** *The symmetric kernel*

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

gives Blum-Kiefer-Rosenblatt's R statistic (see [115]), which is a rank-based U-statistic of order 6*, refer also to [116–120].*

Example 6. *Bergsma-Dassios-Yanagimoto's* τ^* *[121] introduced a rank correlation statistic as a U-statistic of order* 4 *with the symmetric kernel*

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

Here

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

Example 7. *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*

$$\tau_{mk} = \frac{1}{n(n-1)} \sum_{1 \le i \ne j \le n} \mathbf{1}\Big\{ (X_{im} - X_{jm}) \Big(X_{ik} - X_{jk} \Big) > 0 \Big\}.$$

Then, Kendall's tau rank correlation coefficient matrix $T = \{\tau_{mk}\}_{m,k=1}^{p}$ denotes a matrix-valued U-statistic, for wich the kernel is bounded. It is obvious that τ_{mk} quantifies the monotonic dependency between (X_{1m}, X_{1k}) and (X_{2m}, X_{2k}) and it is an unbiased estimator of

$$\mathbb{P}((X_{1m} - X_{2m})(X_{1k} - X_{2k}) > 0),$$

that is, the probability that (X_{1m}, X_{1k}) and (X_{2m}, X_{2k}) are concordant.

Example 8. *The Gini mean difference. The Gini index provides another usual measure of dispersion. It corresponds to the case where $E \subset \mathbb{R}$ and $h(x, y) = |x - y|$:*

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

6. Concluding Remarks

In this paper, the conditional U-statistics regression operator estimation methods for random elements taking values in an infinite-dimensional separable Banach space are generalized to the delta-sequences techniques. The space of continuous functions on the interval $(0, 1)$ with the supremum norm illustrates a separable Banach space. Notably, the method of delta-sequences unifies the kernel method of the probability density function estimation, the histogram method, and a few other methods, including the method of orthogonal series for appropriate choices of orthonormal bases in the one-dimensional and finite-dimensional cases. We have obtained strong uniform consistency results in abstract settings under some conditions on the model. The general framework that we consider extends the existing methods to higher-order statistics; this has a significant impact both from a theoretical and practical point of view. In a future investigation, considering the limiting law of the conditional U-statistics regression estimators based on the delta sequence will be of interest. A natural extension of the present investigation is to consider the serial-dependent setting such as the mixing (see [61,62,122]) or the ergodic processes (see [56,123]). In a future investigation of the functional delta sequence local linear approach estimators, it will be natural to think about the possibility of obtaining an alternative estimator that benefits from the advantages of both methods, the local linear method and the delta sequence approach. This is because both methods have their own distinct advantages. Many methods have been developed and established to construct, in asymptotically optimal ways, bandwidth selection rules for nonparametric kernel estimators, particularly for the Nadaraya-Watson regression estimator. We quote several of these methods, including [44,45,124]. This parameter needs to have an appropriate value chosen for it to ensure that satisfactory practical performances are achieved, either in the typical situation of finite dimensions or in the framework of infinite dimensions. On the other hand, to the best of our knowledge, no such studies are currently conducted to treat generic functional conditional U-statistics. This exemplifies a potential new avenue for research in the future.

7. Mathematical Development

This section contains the proof of our results. The preceding notation is also used in the subsequent text. Keeping in mind the relation (7), we can conclude that, for each $\mathbf{x} =$

$(x_1, \ldots, x_k) \in S_{\mathcal{X}}^k$, there exists $\boldsymbol{\ell}(\mathbf{x}) = (\ell(x_1), \ldots, \ell(x_k))$ where $\forall 1 \leq i \leq k$, $1 \leq \ell(x_i) \leq d_n$ and such that

$$\mathbf{x} \in \prod_{i=1}^{k} B(x_{\ell(x_i)}, \varepsilon_n) \text{ and } d(x_i, x_{\ell(x_i)}) = \underset{1 \leq \ell \leq d_n}{\arg\min} d(x_i, x_\ell).$$

We denote for each $\mathbf{x} = (x_1, \ldots, x_k) \in S_{\mathcal{X}}^k$ and $\mathbf{x}_{\boldsymbol{\ell}(\mathbf{x})} = (x_{\ell(x_1)}, \ldots, x_{\ell(x_k)})$:

$$\prod_{i=1}^{k} B(x_{\ell(x_i)}, \varepsilon_n) =: B(\mathbf{x}_{\boldsymbol{\ell}(\mathbf{x})}, \varepsilon_n).$$

Hence, for each $\mathbf{x} \in S_{\mathcal{X}}^k$, we can reformulated the U-statistic as

$$\begin{aligned}
&|u_n(\varphi, \mathbf{x}; m_n) - \mathbb{E}[u_n(\varphi, \mathbf{x}; m_n)]| \\
&\quad \leq \quad \left| u_n(\varphi, \mathbf{x}; m_n) - u_n(\varphi, \mathbf{x}_{\boldsymbol{\ell}(\mathbf{x})}; m_n) \right| \\
&\qquad\quad + \left| \mathbb{E}[u_n(\varphi, \mathbf{x}_{\boldsymbol{\ell}(\mathbf{x})}; m_n)] - \mathbb{E}[u_n(\varphi, \mathbf{x}; m_n)] \right| \\
&\qquad\quad + \left| u_n(\varphi, \mathbf{x}_{\boldsymbol{\ell}(\mathbf{x})}; m_n) - \mathbb{E}[u_n(\varphi, \mathbf{x}_{\boldsymbol{\ell}(\mathbf{x})}; m_n)] \right|.
\end{aligned}$$

Proof of Theorem 1. We need to establish that there exists some $\eta > 0$, in such a way that

$$\sum_{n \geq 1} \mathbb{P}\left\{ \sup_{\mathbf{x} \in S_{\mathcal{X}}^k} \sqrt{\frac{m}{s_m \log(m)}} |u_n(\varphi, \mathbf{x}, m_n) - \mathbb{E}(u_n(\varphi, \mathbf{x}, m_n))| \geq \eta \right\} < \infty. \tag{32}$$

To do that, we need to obtain an exponential bound for

$$\mathbb{P}\left(\sup_{\mathbf{x} \in S_{\mathcal{X}}^k} |u_n(\varphi, \mathbf{x}, m_n) - \mathbb{E}[u_n(\varphi, \mathbf{x}, m_n)]| > \eta \sqrt{\frac{s_m \log(m)}{m}} \right).$$

We first remark that we have

$$\begin{aligned}
&|u_n(\varphi, \mathbf{x}, m_n) - \mathbb{E}[u_n(\varphi, \mathbf{x}, m_n)]| \\
&= \frac{(n-k)!}{n!} \left| \sum_{i \in I(k,n)} \left\{ \varphi(Y_{i_1}, \ldots, Y_{i_k}) \prod_{j=1}^{k} \delta_{m_n}(x_j, X_{i_j}) - \mathbb{E}\left[\varphi(Y_{i_1}, \ldots, Y_{i_k}) \prod_{j=1}^{k} \delta_{m_n}(x_j, X_{i_j}) \right] \right\} \right| \\
&= \frac{(n-k)!}{n!} \left| \sum_{i \in I(k,n)} \left\{ G_{\varphi,\mathbf{x}}(\mathbf{X}_i, \mathbf{Y}_i) - \mathbb{E}\left[G_{\varphi,\mathbf{x}}(\mathbf{X}_i, \mathbf{Y}_i) \right] \right\} \right| \\
&= \frac{(n-k)!}{n!} \left| \sum_{i \in I(k,n)} H(\mathbf{X}_i, \mathbf{Y}_i) \right|,
\end{aligned}$$

where

$$H(\mathbf{X}, \mathbf{Y}) = G_{\varphi,\mathbf{x}}(\mathbf{X}, \mathbf{Y}) - \mathbb{E}\left[G_{\varphi,\mathbf{x}}(\mathbf{X}, \mathbf{Y}) \right].$$

In order to get the desired result, we apply Lemma A1 on the function $H(\cdot, \cdot)$. Throughout the rest of the proof, we suppose the function $G_{\varphi,\mathbf{x}}$ is symmetric. Moreover, it is clear that the function $H(\cdot, \cdot)$ is bounded by $2MC_1 s_m$ by condition **(C.2)** and the fact that the function $\varphi(\cdot)$ is bounded by the condition **(C.7)**. We obviously remark that,

$$\theta = \mathbb{E}[H(\mathbf{X}, \mathbf{Y})] = 0$$

by design, and

$$\sigma^2 = \mathrm{Var}(H(\mathbf{X},\mathbf{Y})) \le 2(MC_1 s_m)^2.$$

For any $\eta > 0$ and m large enough, we obtain that

$$\begin{aligned}
&\mathbb{P}\left(|u_n(\varphi,\mathbf{x},m_n) - \mathbb{E}[u_n(\varphi,\mathbf{x},m_n)]| > \eta\sqrt{\frac{s_m\log(m)}{m}}\right) \\
&\le 2\exp\left[-\frac{n((s_m\log(m))/m)\eta^2}{4(MC_1s_m)^2 + \frac{4}{3}MC_1s_m\eta\sqrt{(s_m\log(m))/m}}\right]. \qquad (33)
\end{aligned}$$

We can write

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}|u_n(\varphi,\mathbf{x};m_n) - \mathbb{E}[u_n(\varphi,\mathbf{x};m_n)]| > 2\eta\sqrt{\frac{s_m\log(m)}{m}}\right) \\
\le \quad & \mathbb{P}\left(\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\Big|u_n(\varphi,\mathbf{x};m_n) - u_n(\varphi,\mathbf{x}_{\ell(\mathbf{x})};m_n)\right. \\
&\left.+\mathbb{E}[u_n(\varphi,\mathbf{x}_{\ell(\mathbf{x})};m_n)] - \mathbb{E}[u_n(\varphi,\mathbf{x};m_n)]\Big| > \eta\sqrt{\frac{s_m\log(m)}{m}}\right) \qquad (34) \\
&+\mathbb{P}\left(\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\Big|u_n(\varphi,\mathbf{x}_{\ell(\mathbf{x})};m_n) - \mathbb{E}[u_n(\varphi,\mathbf{x}_{\ell(\mathbf{x})};m_n)]\Big| > \eta\sqrt{\frac{s_m\log(m)}{m}}\right). \qquad (35)
\end{aligned}$$

Taking into account the condition **(C.3)**, we have

$$\begin{aligned}
&\Big|u_n(\varphi,\mathbf{x};m_n) - u_n(\varphi,\mathbf{x}_{\ell(\mathbf{x})};m_n)\Big| \\
&\le \frac{(n-k)!}{n!}\sum_{i\in I(k,n)}\left|\varphi(Y_{i_1},\ldots,Y_{i_k})\left\{\prod_{j=1}^k \delta_{m_n}(x_j,X_{i_j}) - \prod_{j=1}^k \delta_{m_n}(x_{\ell(x_j)},X_{i_j})\right\}\right| \\
&\le M\frac{(n-k)!}{n!}\sum_{i\in I(k,n)}\Big|\delta_{m_n}(\mathbf{x},\mathbf{X}_i) - \delta_{m_n}(\mathbf{x}_{\ell(\mathbf{x})},\mathbf{X}_i)\Big| \\
&\le M\frac{(n-k)!}{n!}\sum_{i\in I(k,n)} C_2 s_m^{\beta_2} d(\mathbf{x},\mathbf{x}_{\ell(\mathbf{x})})^{\beta_1} \\
&\le MC_2 s_m^{\beta_2} d(\mathbf{x},\mathbf{x}_{\ell(\mathbf{x})})^{\beta_1} \\
&\le MC_2 s_m^{\beta_2}\epsilon_n^{\beta_1}.
\end{aligned}$$

Consequently, we obtain uniformly on $\mathbf{x} \in S_{\mathcal{X}}^k$:

$$\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\Big|u_n(\varphi,\mathbf{x};m_n) - u_n(\varphi,\mathbf{x}_{\ell(\mathbf{x})};m_n)\Big| \le O(s_m^{\beta_2}\epsilon_n^{\beta_1}) = O\left(\sqrt{\frac{s_m\log(m)}{m}}\right), \qquad (36)$$

by condition **(C.5)**. We deduce from (36) that:

$$\begin{aligned}
&\Big|\mathbb{E}[u_n(\varphi,\mathbf{x}_{\ell(\mathbf{x})};m_n)] - \mathbb{E}[u_n(\varphi,\mathbf{x};m_n)]\Big| \\
&\quad = \Big|\mathbb{E}\Big[u_n(\varphi,\mathbf{x}_{\ell(\mathbf{x})};m_n) - u_n(\varphi,\mathbf{x};m_n)\Big]\Big| \\
&\quad \le \mathbb{E}\Big|\Big[u_n(\varphi,\mathbf{x}_{\ell(\mathbf{x})};m_n) - u_n(\varphi,\mathbf{x};m_n)\Big]\Big|. \qquad (37)
\end{aligned}$$

The passage from (37) to (37) follows by applying Jensen's inequality further to some properties of the absolute value function. Now using the fact that the function $\varphi(\cdot)$ is bounded and that the function δ_m is Lipschitz in addition for any constant a, $\mathbb{E}[a]=a$, we can directly conclude that

$$\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\left|\mathbb{E}[u_n(\varphi,\mathbf{x}_{\boldsymbol{\ell}(\mathbf{x})};m_n)]-\mathbb{E}[u_n(\varphi,\mathbf{x};m_n)]\right|\leq O\left(s_m^{\beta_2}\epsilon_n^{\beta_1}\right)=O\left(\sqrt{\frac{s_m\log(m)}{m}}\right).$$

For some $\eta>0$ and for sufficiently large n and large m, we have

$$\begin{aligned}\mathbb{P}\Bigg(&\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\Big|u_n(\varphi,\mathbf{x};m_n)-u_n(\varphi,\mathbf{x}_{\boldsymbol{\ell}(\mathbf{x})};m_n)\\&+\mathbb{E}[u_n(\varphi,\mathbf{x}_{\boldsymbol{\ell}(\mathbf{x})};m_n)]-\mathbb{E}[u_n(\varphi,\mathbf{x};m_n)]\Big|>\eta\sqrt{\frac{s_m\log(m)}{m}}\Bigg)=0.\end{aligned}$$

Continue, now, with (35), supposing that the kernel function $G_{\varphi,\mathbf{x}_\ell}(\cdot)$ is symmetric, we have to decompose the U-statistic by making use of the [8] decomposition, we infer that

$$\begin{aligned}&u_n(\varphi,\mathbf{x}_\ell;m_n)-\mathbb{E}[u_n(\varphi,\mathbf{x}_\ell;m_n)]\\&=\sum_{p=1}^{k}\frac{k!}{(k-p)!}u_n^{(p)}\left(\pi_{p,k}(G_{\varphi,\mathbf{x}_\ell,m_n})\right)\\&=ku_n^{(1)}(\pi_{1,k}(G_{\varphi,\mathbf{x}_\ell}))+\sum_{p=2}^{k}\frac{k!}{(k-p)!}u_n^{(p)}\left(\pi_{p,k}(G_{\varphi,\mathbf{x}_\ell})\right).\end{aligned}\tag{38}$$

Let us first start with the linear term. We have

$$ku_n^{(1)}(\pi_{1,k}(G_{\varphi,\mathbf{x}_\ell}))=\frac{k}{n}\sum_{j=1}^{n}\pi_{1,k}(G_{\varphi,\mathbf{x}_\ell})(\mathbf{X}_i,\mathbf{Y}_i).$$

From Hoeffding's projection, we have

$$\begin{aligned}&\pi_{1,k}(G_{\varphi,\mathbf{x}_\ell})(x,y)\\&=\left\{\mathbb{E}\left[G_{\varphi,\mathbf{x}_\ell}((x,X_2,\ldots,X_k),(y,Y_2,\ldots,Y_k))\right]-\mathbb{E}[G_{\varphi,\mathbf{x}_\ell}(\mathbf{X},\mathbf{Y})]\right\}\\&=\left\{\mathbb{E}[G_{\varphi,\mathbf{x}_\ell}(\mathbf{X},\mathbf{Y})|(X_1,Y_1)=(x,y)]-\mathbb{E}[G_{\varphi,\mathbf{x}_\ell}(\mathbf{X},\mathbf{Y})]\right\}.\end{aligned}$$

Set

$$Z_i=\pi_{1,k}(G_{\varphi,\mathbf{x}_\ell})(X_i,Y_i).$$

We can see that Z_i are independent and identically distributed random variables bounded by $2kMC_1s_m$ with zero mean and

$$\sigma^2\leq(MC_1s_m)^2.$$

An application of Bernstein's inequality yields

$$\begin{aligned}&\mathbb{P}\left(\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\left|u_n^{(1)}(\pi_{1,k}(G_{\varphi,\mathbf{x}_\ell}))\right|>\eta\sqrt{\frac{s_m\log(m)}{m}}\right)\\&\leq\sum_{i=1}^{d_n}\mathbb{P}\left(\max_{1\leq\ell_i\leq d_n}\left|u_n^{(1)}(\pi_{1,k}(G_{\varphi,\mathbf{x}_\ell}))\right|>\eta\sqrt{\frac{s_m\log(m)}{m}}\right)\\&\leq 2d_n\exp\left[-\frac{n((s_m\log(m))/m)\eta^2}{4(MC_1s_m)^2+\frac{4}{3}MC_1s_m\eta\sqrt{(s_m\log(m))/m}}\right]\end{aligned}$$

$$\leq \quad n^{\alpha-\tau\eta^2/C_4},$$

resulting from the fact $m \leq n$ and $\log(m) \geq \tau \log(n)$. This implies that

$$\sum_{n\geq 1} \mathbb{P}\left(\sup_{\mathbf{x}\in S_{\mathcal{X}}^k} \left| u_n^{(1)}\left(\pi_{1,k}(G_{\varphi,\mathbf{x}_\ell})\right) \right| > \eta \sqrt{\frac{s_m \log(m)}{m}} \right)$$
$$\leq \sum_{n\geq 1} n^{\alpha-\tau\eta^2/C_4} < \infty.$$

Consequently, we obtain the following:

$$\sup_{\mathbf{x}\in S_{\mathcal{X}}^k} \left| u_n^{(1)}\left(\pi_{1,k}(G_{\varphi,\mathbf{x}_\ell})\right) \right| = O_{a.co}\left(\sqrt{\frac{s_m \log(m)}{m}} \right).$$

Moving to the nonlinear term, we want to prove that for $2 \leq p \leq k$:

$$\sup_{\mathbf{x}\in S_{\mathcal{X}}^k} \frac{\binom{k}{p} \sqrt{m} \left| u_n^{(p)}\left(\pi_{p,k} G_{\varphi,\mathbf{x}_{\ell(\mathbf{x})}}\right) \right|}{\sqrt{s_m \log(m)}} = \mathrm{O}_{a.co}(1),$$

which implies that, for $1 \leq i \leq k$ and $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_k)$:

$$\max_{1\leq \ell_i \leq d_n} \frac{\binom{k}{p} \sqrt{m} \left| u_n^{(p)}\left(\pi_{p,k} G_{\varphi,\mathbf{x}_{\ell(\mathbf{x})}}\right) \right|}{\sqrt{s_m \log(m)}} = \mathrm{O}_{a.co}(1).$$

To prove the above-mentioned equation, we need to apply Proposition 1 of [125] (see Lemma A2). We can see that $G_{\varphi,\mathbf{x}_\ell}$ is bounded by $MC_1 s_m$, hence for $\eta > 0$ we have

$$\begin{aligned}
&\mathbb{P}\left(n^{1/2} \left| \sum_{p=2}^{k} \frac{k!}{(k-p)!} u_n^{(p)}\left(\pi_{p,k}(G_{\varphi,\mathbf{x}_\ell})\right) \right| > \eta \sqrt{\frac{s_m \log(m)}{m}} \right) \\
&= \mathbb{P}\left(\left| \sum_{p=2}^{k} \frac{k!}{(k-p)!} u_n^{(p)}\left(\pi_{p,k}(G_{\varphi,\mathbf{x}_\ell})\right) \right| > n^{-1/2}\eta \sqrt{\frac{s_m \log(m)}{m}} \right) \\
&= \mathbb{P}\left(\left| \sum_{p=2}^{k} \frac{k!}{(k-p)!} u_n^{(p)}\left(\pi_{p,k}(G_{\varphi,\mathbf{x}_\ell})\right) \right| > \varepsilon_0 \sqrt{\frac{s_m \log(m)}{m}} \right),
\end{aligned}$$

where $\varepsilon_0 = \dfrac{\eta}{\sqrt{n}}$. Now for $t = \eta \sqrt{\dfrac{s_m \log(m)}{m}}$, Lemma A2 gives us:

$$\begin{aligned}
&\mathbb{P}\left(\left| \sum_{p=2}^{k} \frac{k!}{(k-p)!} u_n^{(p)}\left(\pi_{p,k}(G_{\varphi,\mathbf{x}_\ell})\right) \right| > \varepsilon_0 \sqrt{\frac{s_m \log(m)}{m}} \right) \\
&\leq \quad 2\exp\left(-\frac{t(n-1)^{1/2}}{2^{k+2} k^{k+1} M C_1 s_m} \right) \\
&\leq \quad 2\exp\left(-\frac{\eta\sqrt{s_m \log(m)/m}(n-1)^{1/2}}{2^{k+2} k^{k+1} M C_1 s_m} \right) \\
&\leq \quad 2\exp\left(-\frac{\eta\sqrt{\log(m)/m}(n-1)^{1/2}}{2^{k+2} k^{k+1} M C_1 \sqrt{s_m}} \right).
\end{aligned}$$

By the fact that $m \leq n$ and $\log(m) \geq \tau \log(n)$, it follows that there exists $\eta > 0$ in such a way that

$$\mathbb{P}\left(\left|\sum_{p=2}^{k} \binom{k}{p} u_n^{(p)}\left(\pi_{p,k}(G_{\varphi,\mathbf{x}_\ell})\right)\right| > \varepsilon_0 \sqrt{\frac{s_m \log(m)}{m}}\right) \leq n^{-\tau/2C_5},$$

where $C_5 = C''2^{k+2}k^{k+1}MC_1\sqrt{s_m}$ with $C'' > 0$. Therefore, for each $\varepsilon_0 > 0$, $1 \leq i \leq k$ and $\ell = (\ell_1, \ldots, \ell_k)$:

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\left|\sum_{p=2}^{k} \binom{k}{p} u_n^{(p)}\left(\pi_{p,k}(G_{\varphi,\mathbf{x}_\ell})\right)\right| > \varepsilon_0 \sqrt{\frac{s_m \log(m)}{m}}\right) \\
&\leq \; d_n \max_{1\leq \ell_i \leq d_n} \mathbb{P}\left(\left|\sum_{p=2}^{k} \binom{k}{p} u_n^{(p)}\left(\pi_{p,k}(G_{\varphi,\mathbf{x}_\ell})\right)\right| > \varepsilon_0 \sqrt{\frac{s_m \log(m)}{m}}\right) \\
&\leq \; n^{-k(\tau/2C_5)}.
\end{aligned}$$

Consequently, we have

$$\begin{aligned}
\sum_{n\geq 1}\mathbb{P}\left(\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\left|\sum_{p=2}^{k} \binom{k}{p} u_n^{(p)}\left(\pi_{p,k}(G_{\varphi,\mathbf{x}_\ell})\right)\right| > \varepsilon_0 \sqrt{\frac{s_m \log(m)}{m}}\right)& \\
\leq \sum_{n\geq 1} n^{\alpha-\tau/2C_5} \to 0 \text{ as } n \to 0.&
\end{aligned}$$

Hence, the proof is achieved. $\square$

Proof of Theorem 2. We will need to truncate the conditional U-statistic to prove this theorem. Taking the condition **(C.7′)** into account, for each $\lambda > 0$ and

$$\xi_n := \xi_{m_n} = \frac{m_n}{\log m_n} =: \frac{m}{\log m},$$

we have

$$\begin{aligned}
G_{\varphi,\mathbf{x}}(\mathbf{x},\mathbf{y}) &= G_{\varphi,\mathbf{x}}(\mathbf{x},\mathbf{y})\mathbb{1}_{\{\varphi(\mathbf{y})\leq\lambda\xi_n^{1/q}\}} + G_{\varphi,\mathbf{x}}(\mathbf{x},\mathbf{y})\mathbb{1}_{\{\varphi(\mathbf{y})>\lambda\xi_n^{1/q}\}} \\
&=: G_{\varphi,\mathbf{x}}^{(T)}(\mathbf{x},\mathbf{y}) + G_{\varphi,\mathbf{x}}^{(R)}(\mathbf{x},\mathbf{y}),
\end{aligned}$$

which means that each function $\varphi(\cdot)$ is truncated as follows:

$$\begin{aligned}
\varphi(\mathbf{y}) &= \varphi(\mathbf{y})\mathbb{1}_{\left\{\varphi(\mathbf{y})\leq\lambda\xi_n^{1/q}\right\}} + \varphi(\mathbf{y})\mathbb{1}_{\left\{\varphi(\mathbf{y})>\lambda\xi_n^{1/q}\right\}} \\
&= \varphi^{(T)}(\mathbf{y}) + \varphi^{(R)}(\mathbf{y}).
\end{aligned}$$

Notice that the $G_{\varphi,\mathbf{x}}^{(T)}(\mathbf{x},\mathbf{y})$ denotes the truncated part and $G_{\varphi,\mathbf{x}}^{(R)}(\mathbf{x},\mathbf{y})$ refers to the reminder part. It is possible to write the U-statistic in the following way

$$\begin{aligned}
u_n(\varphi,\mathbf{x},m_n) &= u_n^{(k)}\left(G_{\varphi,\mathbf{x}}^{(T)}\right) + u_n^{(k)}\left(G_{\varphi,\mathbf{x}}^{(R)}\right) \\
&=: u_n^{(T)}(\varphi,\mathbf{x},m_n) + u_n^{(R)}(\varphi,\mathbf{x},m_n).
\end{aligned}$$

The first term of the right side $u_n^{(T)}(\varphi,\mathbf{x},m_n)$, as usual, is called the truncated part and the second one $u_n^{(R)}(\varphi,\mathbf{x},m_n)$ is the remainder part. Let us investigate the term $u_n^{(T)}(\varphi,\mathbf{x},m_n)$.

7.1. Truncated Part

In a similar way as in the preceding proof, we infer

$$\begin{aligned}
&\left|u_n^{(T)}(\varphi,\mathbf{x},m_n)-\mathbb{E}\left(u_n^{(T)}(\varphi,\mathbf{x},m_n)\right)\right|\\
&= \frac{(n-k)!}{n!}\left|\sum_{i\in I(k,n)}\left\{\varphi^{(T)}(Y_{i_1},\ldots,Y_{i_k})\prod_{j=1}^{k}\delta_{m_n}(x_j,X_{i_j})\right.\right.\\
&\qquad \left.\left.-\mathbb{E}\left[\varphi^{(T)}(Y_{i_1},\ldots,Y_{i_k})\prod_{j=1}^{k}\delta_{m_n}(x_j,X_{i_j})\right]\right\}\right|\\
&= \frac{(n-k)!}{n!}\left|\sum_{i\in I(k,n)}\left\{G_{\varphi,\mathbf{x}}^{(T)}(\mathbf{X}_i,\mathbf{Y}_i)-\mathbb{E}\left[G_{\varphi,\mathbf{x}}^{(T)}(\mathbf{X}_i,\mathbf{Y}_i)\right]\right\}\right|\\
&= \frac{(n-k)!}{n!}\left|\sum_{i\in I(k,n)}H^{(T)}(\mathbf{X}_i,\mathbf{Y}_i)\right|,
\end{aligned}$$

where

$$H^{(T)}(\mathbf{X},\mathbf{Y})=G_{\varphi,\mathbf{x}}^{(T)}(\mathbf{X},\mathbf{Y})-\mathbb{E}\left[G_{\varphi,\mathbf{x}}^{(T)}(\mathbf{X},\mathbf{Y})\right].$$

Similar to the proof of Theorem 1, we apply Lemma A1 on the function $H^{(T)}(\cdot,\cdot)$. Throughout the rest of the proof, we suppose that the function $G_{\varphi,\mathbf{x}}^{(T)}$ is symmetric. Moreover, it is clear that the function $H^{(T)}(\cdot,\cdot)$ is bounded by $2\lambda\xi_n^{1/q}C_1 s_m$ by condition **(C.2)**. We obviously remark that,

$$\theta=\mathbb{E}[H^{(T)}(\mathbf{X},\mathbf{Y})]=0$$

by design, and

$$\sigma^2=\operatorname{Var}(H^{(T)}(\mathbf{X},\mathbf{Y}))\leq 2(\lambda\xi_n^{1/q}C_1 s_m)^2.$$

For any $\eta>0$ and m large enough, we obtain that

$$\begin{aligned}
&\mathbb{P}\left(\left|u_n^{(T)}(\varphi,\mathbf{x},m_n)-\mathbb{E}\left(u_n^{(T)}(\varphi,\mathbf{x},m_n)\right)\right|>\eta\sqrt{\frac{s_m\log(m)}{m}}\right)\\
&\leq 2\exp\left[-\frac{n((s_m\log(m))/m)\eta^2}{4(\lambda\xi_n^{1/q}C_1 s_m)^2+\frac{4}{3}\lambda\xi_n^{1/q}C_1 s_m\eta\sqrt{(s_m\log(m))/m}}\right].
\end{aligned}$$

We can write

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\left|u_n^{(T)}(\varphi,\mathbf{x},m_n)-\mathbb{E}\left(u_n^{(T)}(\varphi,\mathbf{x},m_n)\right)\right|>2\eta\sqrt{\frac{s_m\log(m)}{m}}\right)\\
&\leq \mathbb{P}\left(\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\left|u_n^{(T)}(\varphi,\mathbf{x};m_n)-u_n^{(T)}(\varphi,\mathbf{x}_{\boldsymbol{\ell}(\mathbf{x})};m_n)\right.\right.\\
&\qquad \left.\left.+\mathbb{E}[u_n^{(T)}(\varphi,\mathbf{x}_{\boldsymbol{\ell}(\mathbf{x})};m_n)]-\mathbb{E}[u_n^{(T)}(\varphi,\mathbf{x};m_n)]\right|>\eta\sqrt{\frac{s_m\log(m)}{m}}\right)\\
&\qquad +\mathbb{P}\left(\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\left|u_n^{(T)}(\varphi,\mathbf{x}_{\boldsymbol{\ell}(\mathbf{x})};m_n)-\mathbb{E}[u_n^{(T)}(\varphi,\mathbf{x}_{\boldsymbol{\ell}(\mathbf{x})};m_n)]\right|>\eta\sqrt{\frac{s_m\log(m)}{m}}\right). \qquad (39)
\end{aligned}$$

Notice that

$$
\begin{aligned}
&\left|u_n^{(T)}(\varphi,\mathbf{x};m_n)-u_n^{(T)}(\varphi,\mathbf{x}_{\ell(\mathbf{x})};m_n)\right|\\
&\leq \frac{(n-k)!}{n!}\sum_{i\in I(k,n)}\left|\varphi^{(T)}(Y_{i_1},\ldots,Y_{i_k})\left\{\prod_{j=1}^{k}\delta_{m_n}(x_j,X_{i_j})-\prod_{j=1}^{k}\delta_{m_n}(x_{\ell(x_j)},X_{i_j})\right\}\right|\\
&\leq \frac{(n-k)!}{n!}\sum_{i\in I(k,n)}\left|\varphi^{(T)}(Y_{i_1},\ldots,Y_{i_k})\right|\left|\delta_{m_n}(\mathbf{x},\mathbf{X}_i)-\delta_{m_n}(\mathbf{x}_{\ell(\mathbf{x})},\mathbf{X}_i)\right|\\
&\leq \frac{(n-k)!}{n!}\sum_{i\in I(k,n)}C_2 s_m^{\beta_2} d(\mathbf{x},\mathbf{x}_{\ell(\mathbf{x})})^{\beta_1}\left|\varphi^{(T)}(Y_{i_1},\ldots,Y_{i_k})\right|\\
&\leq \frac{(n-k)!}{n!}\sum_{i\in I(k,n)}C_2 s_m^{\beta_2}\epsilon_n^{\beta_1}\left|\varphi^{(T)}(Y_{i_1},\ldots,Y_{i_k})\right|\\
&\leq \frac{(n-k)!}{n!}\sum_{i\in I(k,n)}\frac{1}{n}\sum_{j=1}^{n}C_2 s_m^{\beta_2}\epsilon_n^{\beta_1}\left|\varphi^{(T)}(Y_{i_1},\ldots,Y_{i_k})\right|\\
&\leq \frac{(n-k)!}{n!}\sum_{i\in I(k,n)}\frac{1}{n}\sum_{j=1}^{n}W_{j,T},
\end{aligned}
$$

where for $1\leq j\leq n$,

$$W_{j,T}:=C_2 s_m^{\beta_2}\epsilon_n^{\beta_1}\left|\varphi^{(T)}(Y_1,\ldots,Y_k)\right|,$$

and we can write

$$
\begin{aligned}
\mathbb{E}\left[W_{j,T}\right] &= C_2 s_m^{\beta_2}\epsilon_n^{\beta_1}\mathbb{E}\left[\left|\varphi^{(T)}(Y_1,\ldots,Y_k)\right|\right]\\
&= C_2 s_m^{\beta_2}\epsilon_n^{\beta_1}\mathbb{E}\left[\mathbb{E}\left[\varphi^{(T)}(Y_1,\ldots,Y_k)\mid \mathbf{X}=\mathbf{x}\right]\right],
\end{aligned}
$$

which means that for $2\leq\nu\leq q$:

$$
\begin{aligned}
\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\mathbb{E}\left[W_{j,T}\right]^{\nu} &= \sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\left(C_2 s_m^{\beta_2}\epsilon_n^{\beta_1}\right)^{\nu}\mathbb{E}\left[\mathbb{E}\left[\varphi^{(T)}(Y_{i_1},\ldots,Y_{i_k})\mid \mathbf{X}=\mathbf{x}\right]\right]^{\nu} && (40)\\
&\leq \left(C_2 s_m^{\beta_2}\epsilon_n^{\beta_1}\right)^{\nu}(\lambda\xi_n)^{\nu/q}\mu_q^{\nu/q} && (41)\\
&\leq C_2^{\nu}\left(\frac{s_m\log m}{m}\right)^{\nu/2}(\lambda\xi_n)^{\nu/q}\mu_q^{\nu/q} && (42)\\
&\leq C_2^{\nu}\lambda^{\nu/q}(\xi_n)^{\nu/q-\nu/2}(s_m)^{1-\nu/2}\mu_q^{\nu/q}(s_m)^{\nu-1}\\
&\leq \mathfrak{C}s_m^{(\nu-1)}.
\end{aligned}
$$

The passage from (40) to (41) is possible by the use of the Jensen's inequality for the concave function z^a, for $0<a\leq 1$, while (42) is by condition **(C.5)**. Then, for $\nu\geq 2$

$$\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\mathbb{E}\left[W_{j,T}\right]^{\nu}\quad\leq\quad \mathfrak{C}s_m^{(\nu-1)},$$

where $\mathfrak{C}>0$. Then, an application of classical inequality (see Corollary A.8-(ii) [48]) with $Z_i:=W_{j,T}$ and $a_n^2=s_m$, which gives us

$$u_m=a_n^2\ln(m)m^{-1}=s_m\ln(m)m^{-1},$$

and it is clear that $u_m \to 0$ with m by condition **(C.2)**. Consequently, we obtain uniformly on $\mathbf{x} \in S_{\mathcal{X}}^k$:

$$\sup_{\mathbf{x} \in S_{\mathcal{X}}^k} \left| u_n^{(T)}(\varphi, \mathbf{x}; m_n) - u_n^{(T)}(\varphi, \mathbf{x}_{\boldsymbol{\ell}(\mathbf{x})}; m_n) \right| = O_{a.co}\left(\sqrt{\frac{s_m \log(m)}{m}} \right). \tag{43}$$

Now, we obtain from (43) that:

$$\begin{aligned} & \left| \mathbb{E}[u_n^{(T)}(\varphi, \mathbf{x}_{\boldsymbol{\ell}(\mathbf{x})}; m_n)] - \mathbb{E}[u_n^{(T)}(\varphi, \mathbf{x}; m_n)] \right| \\ = & \left| \mathbb{E}\left[u_n^{(T)}(\varphi, \mathbf{x}_{\boldsymbol{\ell}(\mathbf{x})}; m_n) - u_n^{(T)}(\varphi, \mathbf{x}; m_n) \right] \right| \qquad (44) \\ \leq & \ \mathbb{E}\left| \left[u_n^{(T)}(\varphi, \mathbf{x}_{\boldsymbol{\ell}(\mathbf{x})}; m_n) - u_n^{(T)}(\varphi, \mathbf{x}; m_n) \right] \right|. \qquad (45) \end{aligned}$$

Similar to the bounded case, the transition from (44) to (45) is due to Jensen's inequality and some properties of the absolute value function. Furthermore, using the fact that for any constant a, $\mathbb{E}[a] = a$, we can directly conclude that

$$\sup_{\mathbf{x} \in S_{\mathcal{X}}^k} \left| \mathbb{E}[u_n(\varphi, \mathbf{x}_{\boldsymbol{\ell}(\mathbf{x})}; m_n)] - \mathbb{E}[u_n(\varphi, \mathbf{x}; m_n)] \right| = O_{a.co}\left(\sqrt{\frac{s_m \log(m)}{m}} \right).$$

For sufficiently large n and large m, we infer that, for some $\eta > 0$,

$$\begin{aligned} \mathbb{P}\Bigg(& \sup_{\mathbf{x} \in S_{\mathcal{X}}^k} \Big| u_n(\varphi, \mathbf{x}; m_n) - u_n(\varphi, \mathbf{x}_{\boldsymbol{\ell}(\mathbf{x})}; m_n) \\ & + \mathbb{E}[u_n(\varphi, \mathbf{x}_{\boldsymbol{\ell}(\mathbf{x})}; m_n)] - \mathbb{E}[u_n(\varphi, \mathbf{x}; m_n)] \Big| > \eta \sqrt{\frac{s_m \log(m)}{m}} \Bigg) = 0. \end{aligned}$$

Continue, now, with (39), by imposing that the kernel function $G_{\varphi, \mathbf{x}_\ell}^{(T)}(\cdot)$ is symmetric, the U-statistic is decomposed according to [8] decomposition, that is

$$\begin{aligned} & u_n^{(T)}(\varphi, \mathbf{x}_\ell; m_n) - \mathbb{E}[u_n^{(T)}(\varphi, \mathbf{x}_\ell; m_n)] \\ = & \sum_{p=1}^{k} \frac{k!}{(k-p)!} u_n^{(p)} \left(\pi_{p,k}(G_{\varphi, \mathbf{x}_\ell, m_n}^{(T)}) \right) \\ = & \ k u_n^{(1)} \left(\pi_{1,k}(G_{\varphi, \mathbf{x}_\ell}^{(T)}) \right) + \sum_{p=2}^{k} \frac{k!}{(k-p)!} u_n^{(p)} \left(\pi_{p,k}(G_{\varphi, \mathbf{x}_\ell}^{(T)}) \right). \qquad (46) \end{aligned}$$

Let us first start with the linear term. We have

$$k u_n^{(1)} \left(\pi_{1,k}(G_{\varphi, \mathbf{x}_\ell}^{(T)}) \right) = \frac{k}{n} \sum_{j=1}^{n} \pi_{1,k}(G_{\varphi, \mathbf{x}_\ell}^{(T)})(\mathbf{X}_i, \mathbf{Y}_i).$$

From Hoeffding's projection, we have

$$\begin{aligned} \pi_{1,k}(G_{\varphi, \mathbf{x}_\ell}^{(T)})(x, y) &= \left\{ \mathbb{E}\left[G_{\varphi, \mathbf{x}_\ell}^{(T)}((x, X_2, \ldots, X_k), (y, Y_2, \ldots, Y_k)) \right] - \mathbb{E}[G_{\varphi, \mathbf{x}_\ell}^{(T)}(\mathbf{X}, \mathbf{Y})] \right\} \\ &= \left\{ \mathbb{E}[G_{\varphi, \mathbf{x}_\ell}^{(T)}(\mathbf{X}, \mathbf{Y}) | (X_1, Y_1) = (x, y)] - \mathbb{E}[G_{\varphi, \mathbf{x}_\ell}^{(T)}(\mathbf{X}, \mathbf{Y})] \right\}. \end{aligned}$$

Set

$$Z_i^{(T)} = \pi_{1,k}(G_{\varphi, \mathbf{x}_\ell}^{(T)})(X_i, Y_i).$$

We can clearly observe that $Z_i^{(T)}$ are independent and identically distributed random variables bounded by $2k\lambda\xi_n^{1/q}C_1 s_m$ with zero mean and

$$\sigma^2 \leq (\lambda\xi_n^{1/q}C_1 s_m)^2.$$

An application of Bernstein's inequality yields

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\left|u_n^{(1)}\left(\pi_{1,k}(G_{\varphi,\mathbf{x}_\ell}^{(T)})\right)\right| > \eta\sqrt{\frac{s_m\log(m)}{m}}\right)\\
\leq\ & \sum_{i=1}^{d_n}\mathbb{P}\left(\max_{1\leq\ell_i\leq d_n}\left|u_n^{(1)}\left(\pi_{1,k}(G_{\varphi,\mathbf{x}_\ell}^{(T)})\right)\right| > \eta\sqrt{\frac{s_m\log(m)}{m}}\right)\\
\leq\ & 2d_n\exp\left[-\frac{n((s_m\log(m))/m)\eta^2}{4(\lambda\xi_n^{1/q}C_1s_m)^2+\frac{4}{3}\lambda\xi_n^{1/q}C_1s_m\eta\sqrt{(s_m\log(m))/m}}\right]\\
\leq\ & n^{\alpha-\tau\eta^2/C_4'},
\end{aligned}$$

for some positive constant C_4', resulting from the fact $m \leq n$ and $\log(m) \geq \tau\log(n)$. This implies that

$$\begin{aligned}
&\sum_{n\geq 1}\mathbb{P}\left(\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\left|u_n^{(1)}\left(\pi_{1,k}(G_{\varphi,\mathbf{x}_\ell}^{(T)})\right)\right| > \eta\sqrt{\frac{s_m\log(m)}{m}}\right)\\
&\leq \sum_{n\geq 1} n^{\alpha-\tau\eta^2/C_4'} < \infty.
\end{aligned}$$

Consequently, we obtain:

$$\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\left|u_n^{(1)}\left(\pi_{1,k}(G_{\varphi,\mathbf{x}_\ell}^{(T)})\right)\right| = O_{a.co}\left(\sqrt{\frac{s_m\log(m)}{m}}\right).$$

Moving to the nonlinear term, we want to prove that for $2 \leq p \leq k$:

$$\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\frac{\binom{k}{p}\sqrt{m}\left|u_n^{(p)}\left(\pi_{p,k}G_{\varphi,\mathbf{x}_{\ell(\mathbf{x})}}^{(T)}\right)\right|}{\sqrt{s_m\log(m)}} = \mathrm{O}_{a.co}(1),$$

which implies that, for $1 \leq i \leq k$ and $\boldsymbol{\ell} = (\ell_1,\ldots,\ell_k)$:

$$\max_{1\leq\ell_i\leq d_n}\frac{\binom{k}{p}\sqrt{m}\left|u_n^{(p)}\left(\pi_{p,k}G_{\varphi,\mathbf{x}_{\ell(\mathbf{x})}}^{(T)}\right)\right|}{\sqrt{s_m\log(m)}} = \mathrm{O}_{a.co}(1).$$

To prove the above-mentioned equation, we need to apply Proposition 1 of [125] (see Lemma A2). We can observe that $G_{\varphi,\mathbf{x}_\ell}^{(T)}$ is bounded by $C_1s_m\lambda\xi_n^{1/q}$, hence for $\eta > 0$ we have

$$\begin{aligned}
&\mathbb{P}\left(n^{1/2}\left|\sum_{p=2}^{k}\frac{k!}{(k-p)!}u_n^{(p)}\left(\pi_{p,k}(G_{\varphi,\mathbf{x}_\ell}^{(T)})\right)\right| > \eta\sqrt{\frac{s_m\log(m)}{m}}\right)\\
=\ & \mathbb{P}\left(\left|\sum_{p=2}^{k}\frac{k!}{(k-p)!}u_n^{(p)}\left(\pi_{p,k}(G_{\varphi,\mathbf{x}_\ell}^{(T)})\right)\right| > n^{-1/2}\eta\sqrt{\frac{s_m\log(m)}{m}}\right)\\
=\ & \mathbb{P}\left(\left|\sum_{p=2}^{k}\frac{k!}{(k-p)!}u_n^{(p)}\left(\pi_{p,k}(G_{\varphi,\mathbf{x}_\ell}^{(T)})\right)\right| > \varepsilon_0\sqrt{\frac{s_m\log(m)}{m}}\right),
\end{aligned}$$

where $\varepsilon_0 = \frac{\eta}{\sqrt{n}}$. Now for $t = \eta\sqrt{\frac{s_m \log(m)}{m}}$, Lemma A2 gives us:

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_{p=2}^{k} \frac{k!}{(k-p)!} u_n^{(p)}\left(\pi_{p,k}(G_{\varphi,\mathbf{x}_\ell}^{(T)})\right)\right| > \varepsilon_0\sqrt{\frac{s_m \log(m)}{m}}\right) \\
&\leq \quad 2\exp\left(-\frac{t(n-1)^{1/2}}{2^{k+2}k^{k+1}\lambda\xi_n^{1/q}C_1 s_m}\right) \\
&\leq \quad 2\exp\left(-\frac{\eta\sqrt{s_m \log(m)/m}(n-1)^{1/2}}{2^{k+2}k^{k+1}\lambda\xi_n^{1/q}C_1 s_m}\right) \\
&\leq \quad 2\exp\left(-\frac{\eta\sqrt{\log(m)/m}(n-1)^{1/2}}{2^{k+2}k^{k+1}\lambda\xi_n^{1/q}C_1 \sqrt{s_m}}\right).
\end{aligned}$$

By the fact that $m \leq n$ and $\log(m) \geq \tau\log(n)$, it follows that there exists $\eta > 0$ in such a way that

$$\mathbb{P}\left(\left|\sum_{p=2}^{k} \binom{k}{p} u_n^{(p)}\left(\pi_{p,k}(G_{\varphi,\mathbf{x}_\ell}^{(T)})\right)\right| > \varepsilon_0\sqrt{\frac{s_m \log(m)}{m}}\right) \leq n^{-\tau/2C_6},$$

where

$$C_6 = C'' 2^{k+2}k^{k+1}\lambda\xi_n^{1/q}C_1\sqrt{s_m},$$

with $C'' > 0$. Therefore, for each $\varepsilon_0 > 0$, $1 \leq i \leq k$ and $\ell = (\ell_1, \ldots, \ell_k)$:

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\left|\sum_{p=2}^{k} \binom{k}{p} u_n^{(p)}\left(\pi_{p,k}(G_{\varphi,\mathbf{x}_\ell}^{(T)})\right)\right| > \varepsilon_0\sqrt{\frac{s_m \log(m)}{m}}\right) \\
&\leq \quad d_n \max_{1\leq \ell_i\leq d_n} \mathbb{P}\left(\left|\sum_{p=2}^{k} \binom{k}{p} u_n^{(p)}\left(\pi_{p,k}(G_{\varphi,\mathbf{x}_\ell}^{(T)})\right)\right| > \varepsilon_0\sqrt{\frac{s_m \log(m)}{m}}\right) \\
&\leq \quad n^{-k(\tau/2C_6)}.
\end{aligned}$$

Consequently, we have

$$\begin{aligned}
&\sum_{n\geq 1}\mathbb{P}\left(\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\left|\sum_{p=2}^{k} \binom{k}{p} u_n^{(p)}\left(\pi_{p,k}(G_{\varphi,\mathbf{x}_\ell}^{(T)})\right)\right| > \varepsilon_0\sqrt{\frac{s_m \log(m)}{m}}\right) \\
&\qquad\qquad \leq \sum_{n\geq 1} n^{\alpha-\tau/2C_6} \to 0 \text{ as } n \to 0.
\end{aligned}$$

7.2. Remainder Part

We now consider the remainder part, which is the U-process $u_n^{(R)}(\varphi, \mathbf{x}, m_n)$ related on the unbounded kernel given by:

$$G_{\varphi,\mathbf{x}}^{(R)}(\mathbf{x},\mathbf{y}) = G_{\varphi,\mathbf{x}}(\mathbf{x},\mathbf{y})\mathbb{1}_{\{\varphi(\mathbf{y})>\lambda\xi_n^{1/q}\}}$$

We have establish that the process is negligible, meaning that

$$\sup_{\mathbf{x}\in S_{\mathcal{X}}^k} \frac{\sqrt{m}\left|u_n^{(k)}(G_{\varphi,\mathbf{x}}^{(R)}) - \mathbb{E}\left(u_n^{(k)}\left(G_{\varphi,\mathbf{x}}^{(R)}\right)\right)\right|}{\sqrt{s_m \log(m)}} = o_{a.co}(1). \tag{47}$$

Observe that for $\mathbf{x}, \mathbf{y} \in \mathcal{X}^k$,

$$\begin{aligned}
|G_{\varphi,\mathbf{x}}| &= |\varphi(\mathbf{y})\delta_m(\mathbf{x}, \mathbf{y})| \\
&\leq C_1 s_m |\varphi(\mathbf{y})| =: \widetilde{F}(\mathbf{y}).
\end{aligned}$$

Taking into account that $\widetilde{F}$ is symmetric, we have:

$$\left| u_n^{(k)}\left(G_{\varphi,\mathbf{x}}^{(R)}\right)\right| \leq u_n^{(k)}\left(\widetilde{F}\mathbb{1}_{\{\widetilde{F}>\lambda\xi_n^{1/q}\}}\right),$$

where $u_n^{(k)}\left(\widetilde{F}(\mathbf{y})\mathbb{1}_{\{\varphi(\mathbf{y})>\lambda\xi_n^{1/q}\}}\right)$ is a U-statistic based on the U-kernel $\widetilde{F}\mathbb{1}_{\{\varphi>\lambda\xi_n^{1/q}\}}$:

$$\begin{aligned}
\sup_{\mathbf{x}\in S_{\mathcal{X}}^k} \frac{\sqrt{m}\left|u_n^{(k)}(G_{\varphi,\mathbf{x}}^{(R)})\right|}{\sqrt{s_m\log(m)}} &\leq (s_m^{-1}\xi_n)^{1/2} u_n^{(k)}\left(\widetilde{F}\mathbb{1}_{\{\widetilde{F}>\lambda\xi_n^{1/q}\}}\right) && (48)\\
&\leq C_7\xi_n u_n^{(k)}\left(\widetilde{F}\mathbb{1}_{\{\widetilde{F}>\lambda\xi_n^{1/q}\}}\right), && (49)
\end{aligned}$$

and

$$\begin{aligned}
\sup_{\mathbf{x}\in S_{\mathcal{X}}^k} \frac{\sqrt{m}\left|\mathbb{E}\left(u_n^{(k)}\left(G_{\varphi,\mathbf{x}}^{(R)}\right)\right)\right|}{\sqrt{s_m\log(m)}} &\leq C_7\xi_n\mathbb{E}\left(u_n^{(k)}\left(\widetilde{F}\mathbb{1}_{\{\varphi(\mathbf{Y})>\lambda\xi_n^{1/q}\}}\right)\right)\\
&\leq C_7\mathbb{E}\left(\widetilde{F}^{1+q}\mathbb{1}_{\{\varphi(\mathbf{Y})>\lambda\xi_n^{1/q}\}}\right).
\end{aligned}$$

Therefore, as $m \longrightarrow \infty$ when $n \longrightarrow \infty$, we have

$$\sup_{\mathbf{x}\in S_{\mathcal{X}}^k} \frac{\sqrt{m}\left|\mathbb{E}\left(u_n^{(k)}\left(G_{\varphi,\mathbf{x}}^{(R)}\right)\right)\right|}{\sqrt{s_m\log(m)}} = o(1). \tag{50}$$

Hence to achieve the proof, it remains to establish that:

$$u_n^{(k)}\left(\widetilde{F}\mathbb{1}_{\{\varphi(\mathbf{y})>\lambda\xi_n^{1/q}\}}\right) = o_{a.co}\left(\left(s_m^{-1}\xi_n\right)^{-1/2}\right). \tag{51}$$

An application of the Chebyshev's inequality, for any $\eta > 0$, gives

$$\begin{aligned}
&\mathbb{P}\left\{\left|u_n^{(k)}\left(\widetilde{F}\mathbb{1}_{\{\varphi(\mathbf{Y})>\lambda\xi_n^{1/q}\}}\right) - \mathbb{E}\left(u_n^{(k)}\left(\widetilde{F}\mathbb{1}_{\{\varphi(\mathbf{Y})>\lambda\xi_n^{1/q}\}}\right)\right)\right| \geq \eta(s_m^{-1}\xi_n)^{-1/2}\right\}\\
&\quad\leq \quad \eta^{-2}(s_m^{-1}\xi_n)Var\left(u_n^{(k)}\left(\widetilde{F}\mathbb{1}_{\{\varphi(\mathbf{Y})>\lambda\xi_n^{1/q}\}}\right)\right) \leq k\eta^{-2}\xi_n\mathbb{E}\left(\widetilde{F}^2\mathbb{1}_{\{\varphi(\mathbf{Y})>\lambda\xi_n^{1/q}\}}\right)\\
&\quad\leq \quad \frac{k}{n^2}\eta^{-2}(\xi_n)^q\mathbb{E}\left(\widetilde{F}^2\mathbb{1}_{\{\varphi(\mathbf{Y})>\lambda\xi_n^{1/q}\}}\right) \leq \eta'\mathbb{E}\left(\widetilde{F}^3\mathbb{1}_{\{\varphi(\mathbf{Y})>\lambda\xi_n^{1/q}\}}\right)\frac{1}{n^2},
\end{aligned}$$

so by using the fact that

$$\eta'\mathbb{E}\left(\widetilde{F}^3\mathbb{1}_{\{\varphi(\mathbf{y})>\lambda\xi_n^{1/q}\}}\right)\sum_{n\geq 1}\frac{1}{n^2} < \infty,$$

we deduce that

$$\sum_{n\geq 1}\mathbb{P}\left\{\left|u_n^{(k)}\left(\widetilde{F}\mathbb{1}_{\{\varphi(\mathbf{y})>\lambda\xi_n^{1/q}\}}\right) - \mathbb{E}\left(u_n^{(k)}\left(\widetilde{F}\mathbb{1}_{\{\varphi(\mathbf{y})>\lambda\xi_n^{1/q}\}}\right)\right)\right| \geq \eta(m\xi_n)^{-1/2}\right\} < \infty.$$

Finally, note that (48) implies

$$\mathbb{E}\left(u_n^{(k)}\left(\widetilde{F}\mathbb{1}_{\{\varphi(\mathbf{y})>\lambda\xi_n^{1/q}\}}\right)\right)=o\left(\left(s_m^{-1}\xi_n\right)^{-1/2}\right).$$

The preceding results of the arbitrary choice of $\lambda>0$ gives that (51) holds, which, by combining with (50) and (48), completes the proof of (47). We finally obtain

$$\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}|u_n(\varphi,\mathbf{x},m_n)-\mathbb{E}(u_n(\varphi,\mathbf{x},m_n))|=O_{a.co}\left(\sqrt{\frac{s_m\log(m)}{m}}\right).$$

Hence, the proof is complete. □

Proof of Theorem 3. The conclusion of Theorem 3 can be obtained from the results of Theorems 1 and 2. We have

$$\begin{aligned}
&\left|\widehat{r}_n^{(k)}(\varphi,\mathbf{x};m_n)-\widehat{\mathbb{E}}\left(\widehat{r}_n^{(k)}(\varphi,\mathbf{x};m_n)\right)\right|\\
&\quad=\left|\frac{u_n(\varphi,\mathbf{x};m_n)}{u_n(1,\mathbf{x};m_n)}-\frac{\mathbb{E}(u_n(\varphi,\mathbf{x};m_n))}{\mathbb{E}(u_n(1,\mathbf{x};m_n))}\right|\\
&\quad\leq\frac{|u_n(\varphi,\mathbf{x};m_n)-\mathbb{E}(u_n(\varphi,\mathbf{x};m_n))|}{|u_n(1,\mathbf{x};m_n)|}\\
&\qquad+\frac{|\mathbb{E}(u_n(\varphi,\mathbf{x};m_n))|\cdot|u_n(1,\mathbf{x};m_n)-\mathbb{E}(u_n(1,\mathbf{x};m_n))|}{|u_n(1,\mathbf{x};m_n)|\cdot|\mathbb{E}(u_n(1,\mathbf{x};m_n))|}\\
&\quad=:\mathscr{I}_1+\mathscr{I}_2.
\end{aligned}$$

Notice that, given the imposed hypothesis and previously obtained results, and for some c_1,c_2, we obtain:

$$\begin{aligned}
\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}|u_n(1,\mathbf{x},m_n)| &= c_1 \qquad \text{a.co},\\
\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}|\mathbb{E}(u_n(1,\mathbf{x},m_n))| &= c_2,\\
\sup_{\mathbf{x}\in S_{\mathcal{X}}^m}|\mathbb{E}(u_n(\varphi,\mathbf{x},m_n))| &= O(1).
\end{aligned}$$

Hence now, depending on whether the function $\varphi(\cdot)$ is bounded or unbounded, we can apply Theorem 1 or Theorem 2 (respectively) to handle both $\mathscr{I}_1$ and $\mathscr{I}_2$, and get for some $c''>0$ with probability 1:

$$\begin{aligned}
&\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\frac{\sqrt{m}\left|\left(\widehat{r}_n^{(k)}(\varphi,\mathbf{x};m_n)\right)-\widehat{\mathbb{E}}\left(\widehat{r}_n^{(k)}(\varphi,\mathbf{x};m_n)\right)\right|}{\sqrt{s_m\log(m)}}\\
&\quad\leq\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\frac{\sqrt{m}(\mathscr{I}_1)}{\sqrt{s_m\log(m)}}+\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\frac{\sqrt{m}(\mathscr{I}_2)}{\sqrt{s_m\log(m)}}\\
&\quad\leq c''.
\end{aligned}$$

Hence, the proof is complete. □

Proof of Theorem 4. Let $\gamma>0$ and $\mathbf{x}\in S_{\mathcal{X}}^k$. We have

$$\widehat{\mathbb{E}}\left(\widehat{r}_n^{(k)}(\varphi,\mathbf{x};m_n)\right)-r^{(k)}(\varphi,\mathbf{x})=\frac{\mathbb{E}[u_n(\varphi,\mathbf{x},m_n)]}{\mathbb{E}[u_n(1,\mathbf{x},m_n)]}-r^{(k)}(\varphi,\mathbf{x}).$$

Notice that

$$\begin{aligned}\widehat{\mathbb{E}}\left(\widehat{r}_n^{(k)}(\varphi,\mathbf{x};m_n)\right) &= \frac{1}{\mathbb{E}[\delta_m(\mathbf{x},\mathbf{X})]}\mathbb{E}\left[\varphi(Y_1,\ldots,Y_k)\prod_{j=1}^{k}\delta_{m_n}(x_j,X_j)\right]\\ &= \frac{1}{\mathbb{E}[\delta_m(\mathbf{x},\mathbf{X})]}\int_{\mathcal{X}^k} r^{(k)}(\varphi,\mathbf{t})\delta_m(\mathbf{x},\mathbf{t})\tilde{f}(\mathbf{t})\mu(\mathbf{dt}),\end{aligned}$$

where for $\mathbf{t}=(t_1,\ldots,t_k)$ we denote $\mu(\mathbf{dt}):=\mu(dt_1)\cdots\mu(dt_k)$, and

$$\tilde{f}(\mathbf{t}):=\prod_{j=1}^{k}f(t_j).$$

Which means that

$$\begin{aligned}&\widehat{\mathbb{E}}\left(\widehat{r}_n^{(k)}(\varphi,\mathbf{x};m_n)\right)-r^{(k)}(\varphi,\mathbf{x})\\ &= \frac{1}{\mathbb{E}[\delta_m(\mathbf{x},\mathbf{X})]}\left(\int_{\mathcal{X}^k} r^{(k)}(\varphi,\mathbf{t})\delta_m(\mathbf{x},\mathbf{t})\tilde{f}(\mathbf{t})\mu(\mathbf{dt})-r^{(k)}(\varphi,\mathbf{x})\int_{\mathcal{X}^k}\delta_m(\mathbf{x},\mathbf{t})\tilde{f}(\mathbf{t})\mu(\mathbf{dt})\right)\\ &= \frac{1}{\mathbb{E}[\delta_m(\mathbf{x},\mathbf{X})]}\left(\int_{\mathcal{X}^k}\left(r^{(k)}(\varphi,\mathbf{t})-r^{(k)}(\varphi,\mathbf{x})\right)\delta_m(\mathbf{x},\mathbf{t})\tilde{f}(\mathbf{t})\mu(\mathbf{dt})\right)\\ &:= I_1(\mathbf{x})+I_2(\mathbf{x}),\end{aligned}$$

where

$$I_1(\mathbf{x}):=\frac{1}{\mathbb{E}[\delta_m(\mathbf{x},\mathbf{X})]}\int_{\mathbb{B}(\mathbf{x},\gamma)}\left(r^{(k)}(\varphi,\mathbf{t})-r^{(k)}(\varphi,\mathbf{x})\right)\delta_m(\mathbf{x},\mathbf{t})\tilde{f}(\mathbf{t})\mu(\mathbf{dt}), \tag{52}$$

and

$$I_2(\mathbf{x}):=\frac{1}{\mathbb{E}[\delta_m(\mathbf{x},\mathbf{X})]}\int_{\bar{\mathbb{B}}(\mathbf{x},\gamma)}\left(r^{(k)}(\varphi,\mathbf{t})-r^{(k)}(\varphi,\mathbf{x})\right)\delta_m(\mathbf{x},\mathbf{t})\tilde{f}(\mathbf{t})\mu(\mathbf{dt}). \tag{53}$$

Therefore, we need to study the asymptotic behavior of both $\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}(I_1(\mathbf{x}))$ and $\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}(I_2(\mathbf{x}))$ to obtain the desired result.
Let us start with the term $\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}(I_1(\mathbf{x}))$, we have

$$\begin{aligned}&\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}|I_1(\mathbf{x})|\\ &= \sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\left|\frac{1}{\mathbb{E}[\delta_m(\mathbf{x},\mathbf{X})]}\int_{\mathbb{B}(\mathbf{x},\gamma)}\left(r^{(k)}(\varphi,\mathbf{t})-r^{(k)}(\varphi,\mathbf{x})\right)\delta_m(\mathbf{x},\mathbf{t})\tilde{f}(\mathbf{t})\mu(\mathbf{dt})\right|\\ &\leq \frac{1}{\mathbb{E}[\delta_m(\mathbf{x},\mathbf{X})]}\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\int_{\mathbb{B}(\mathbf{x},\gamma)}\left|\left(r^{(k)}(\varphi,\mathbf{t})-r^{(k)}(\varphi,\mathbf{x})\right)\delta_m(\mathbf{x},\mathbf{t})\tilde{f}(\mathbf{t})\right|\mu(\mathbf{dt}),\end{aligned}$$

taking into account the fact that the density function $f(\cdot)$ is bounded, and by condition **(C.9)**, we get:

$$\begin{aligned}&\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}|I_1(\mathbf{x})|\\ &\leq \frac{C_f}{\mathbb{E}[\delta_m(\mathbf{x},\mathbf{X})]}\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\int_{\mathbb{B}(\mathbf{x},\gamma)}C_3 d(\mathbf{x},\mathbf{t})\delta_m(\mathbf{x},\mathbf{t})\mu(\mathbf{dt})\\ &\leq \frac{C_f C_3\gamma}{\mathbb{E}[\delta_m(\mathbf{x},\mathbf{X})]}\sup_{\mathbf{x}\in S_{\mathcal{X}}^k}\int_{\mathbb{B}(\mathbf{x},\gamma)}\delta_m(\mathbf{x},\mathbf{t})\mu(\mathbf{dt}),\end{aligned} \tag{54}$$

hence, the term in (54) can be chosen smaller than 2ε as $m \to \infty$ by using the condition (8) . To investigate the term $\sup_{\mathbf{x}\in S^k_{\mathcal{X}}} (I_2(\mathbf{x}))$, notice that

$$
\begin{aligned}
&\sup_{\mathbf{x}\in S^k_{\mathcal{X}}} |I_2(\mathbf{x})| \\
&\quad = \sup_{\mathbf{x}\in S^k_{\mathcal{X}}} \left| \frac{1}{\mathbb{E}[\delta_m(\mathbf{x},\mathbf{X})]} \int_{\mathbb{B}(\mathbf{x},\gamma)} \left(r^{(k)}(\varphi,\mathbf{t}) - r^{(k)}(\varphi,\mathbf{x}) \right) \delta_m(\mathbf{x},\mathbf{t}) \tilde{f}(\mathbf{t}) \mu(\mathbf{dt}) \right| \\
&\quad \leq \frac{C_f C_3}{\mathbb{E}[\delta_m(\mathbf{x},\mathbf{X})]} \sup_{\mathbf{x}\in S^k_{\mathcal{X}}} \int_{\mathbb{B}(\mathbf{x},\gamma)} d(\mathbf{x},\mathbf{t}) \delta_m(\mathbf{x},\mathbf{t}) \mu(\mathbf{dt}).
\end{aligned} \tag{55}
$$

By condition (11), we conclude that

$$
\sup_{\mathbf{x}\in S^k_{\mathcal{X}}} |I_2(\mathbf{x})| \to 0 \ \text{ as } \ m \to \infty.
$$

This concludes the proof of the Theorem. □

Proof of Theorem 5. Following the same steps as the proof of Theorem 4, we can write directly:

$$
\begin{aligned}
&\widehat{\mathbb{E}}\left(\widehat{r}_n^{(k)}(\varphi,\mathbf{x};m_n) \right) - r^{(k)}(\varphi,\mathbf{x}) \\
&\quad = \frac{1}{\mathbb{E}[\delta_m(\mathbf{x},\mathbf{X})]} \left(\int_{\mathcal{X}^k} \left(r^{(k)}(\varphi,\mathbf{t}) - r^{(k)}(\varphi,\mathbf{x}) \right) \delta_m(\mathbf{x},\mathbf{t}) \tilde{f}(\mathbf{t}) \mu(\mathbf{dt}) \right).
\end{aligned}
$$

Taking into account conditions **(C.8)**, we can easily deduce that

$$
\begin{aligned}
&\sup_{\mathbf{x}\in S^k_{\mathcal{X}}} \left| \frac{1}{\mathbb{E}[\delta_m(\mathbf{x},\mathbf{X})]} \left(\int_{\mathcal{X}^k} \left(r^{(k)}(\varphi,\mathbf{t}) - r^{(k)}(\varphi,\mathbf{x}) \right) \delta_m(\mathbf{x},\mathbf{t}) \tilde{f}(\mathbf{t}) \mu(\mathbf{dt}) \right) \right| \\
&\qquad\qquad \leq \sup_{\mathbf{x}\in S^k_{\mathcal{X}}} |I_1(\mathbf{x})| + |I_2(\mathbf{x})|,
\end{aligned}
$$

where $I_1(\mathbf{x})$ and $I_2(\mathbf{x})$ are defined in (52) and (53), respectively. Presently, Equation (54) gives us

$$
\begin{aligned}
&\sup_{\mathbf{x}\in S^k_{\mathcal{X}}} |I_1(\mathbf{x})| \\
&\quad \leq \frac{C_f}{\mathbb{E}[\delta_m(\mathbf{x},\mathbf{X})]} \sup_{\mathbf{x}\in S^k_{\mathcal{X}}} \int_{\mathbb{B}(\mathbf{x},\gamma)} C_3 d(\mathbf{x},\mathbf{t}) \delta_m(\mathbf{x},\mathbf{t}) \mu(\mathbf{dt}) \\
&\quad \leq \frac{C_f C_3 D_m}{\mathbb{E}[\delta_m(\mathbf{x},\mathbf{X})]} \sup_{\mathbf{x}\in S^k_{\mathcal{X}}} \int_{\mathbb{B}(\mathbf{x},\gamma)} \delta_m(\mathbf{x},\mathbf{t}) \mu(\mathbf{dt}) \\
&\quad \leq O(D_m),
\end{aligned}
$$

by conditions **(C.8)** and **(C.1)**. On the other hand, Equation (55) gives us:

$$
\begin{aligned}
&\sup_{\mathbf{x}\in S^k_{\mathcal{X}}} |I_2(\mathbf{x})| \\
&\quad \leq \frac{C_f C_3}{\mathbb{E}[\delta_m(\mathbf{x},\mathbf{X})]} \sup_{\mathbf{x}\in S^k_{\mathcal{X}}} \int_{\mathbb{B}(\mathbf{x},\gamma)} d(\mathbf{x},\mathbf{t}) \delta_m(\mathbf{x},\mathbf{t}) \mu(\mathbf{dt}) \\
&\quad \leq \frac{C_f C_3 D_m}{\mathbb{E}[\delta_m(\mathbf{x},\mathbf{X})]} \sup_{\mathbf{x}\in S^k_{\mathcal{X}}} \int_{\mathbb{B}(\mathbf{x},\gamma)} \delta_m(\mathbf{x},\mathbf{t}) \mu(\mathbf{dt})
\end{aligned}
$$

$$\leq \quad O(D_m),$$

by condition **(C.8)**. This completes the proof of the theorem. □

Author Contributions: S.B., A.N. and T.Z., conceptualization, methodology, investigation, writing—original draft, writing—review and editing. All authors contributed equally to the writing of this paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the three referees for a carefully and thoroughly compiled list of places where the presentation could be improved. The paper has benefited from those points.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Lemma A1 (Theorem A. page 201 [126])**.** *Let f denote a symmetric $\mathcal{X}^k$-valued function fullfilling $\|f\|_\infty \leq c$,*

$$\mathbb{E}f(X_1,\ldots,X_k) = \theta,$$

and

$$\sigma^2 = Var(f(X_1,\ldots,X_k)),$$

then for $t > 0$ and $n \geq k$, we infer:

$$\mathbb{P}\left\{|u_n^{(k)}(f) - \theta| \geq t\right\} \leq \exp\left\{-\frac{[n/k]t^2}{2\sigma^2 - \frac{2}{3}ct}\right\}.$$

Lemma A2 (Proposition 1 [125])**.** *If $G : S^k \to \mathbb{R}$ is a measurable symmetric function with $\|G\|_\infty = b$ then*

$$\mathbb{P}\left\{n^{1/2}\left|\sum_{j=2}^{k}\binom{k}{j}u_n^{(j)}\left(\pi_{j,k}G\right)\right| \geqslant t\right\} \leqslant 2\exp\left(-\frac{t(n-1)^{1/2}}{2^{k+2}k^{k+1}b}\right).$$

Definition A1. *A symmetric and $\mathbb{P}^m$-integrable kernel $f : \mathcal{X}^k \to \mathbb{R}$ is $\mathbb{P}$-degenerate of order $r-1$, notationally $f \in L_2^r\left(\mathbb{P}^k\right)$, if and only if*

$$\int f(x_1,\ldots,x_k)\mathrm{d}\mathbb{P}^{k-r+1}(x_r,\ldots,x_m) = \int f\,\mathrm{d}\mathbb{P}^k$$

holds for any $x_1,\ldots,x_{r-1} \in \mathcal{X}$, and

$$(x_1,\ldots,x_r) \mapsto \int f(x_1,\ldots,x_k)\mathrm{d}\mathbb{P}^{k-r}(x_{r+1},\ldots,x_m)$$

is not a constant function.

References

1. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Monographs on Statistics and Applied Probability; Chapman & Hall: London, UK, 1986; p. x+175. [CrossRef]
2. Nadaraya, E.A. *Nonparametric Estimation of Probability Densities and Regression Curves*; Mathematics and Its Applications (Soviet Series); Kluwer Academic Publishers Group: Dordrecht, The Netherland, 1989; Volume 20, p. x+213; Translated from the Russian by Samuel Kotz. [CrossRef]
3. Härdle, W. *Applied Nonparametric Regression*; Econometric Society Monographs; Cambridge University Press: Cambridge, UK, 1990; Volume 19, p. xvi+333. [CrossRef]

4. Wand, M.P.; Jones, M.C. *Kernel Smoothing*; Monographs on Statistics and Applied Probability; Chapman and Hall, Ltd.: London, UK, 1995; Volume 60, p. xii+212. [CrossRef]
5. Eggermont, P.P.B.; LaRiccia, V.N. *Maximum Penalized Likelihood Estimation. Density Estimation. Vol. I*; Springer Series in Statistics; Springer: New York, NY, USA, 2001; p. xviii+510.
6. Devroye, L.; Lugosi, G. *Combinatorial Methods in Density Estimation*; Springer Series in Statistics; Springer: New York, NY, USA, 2001; p. xii+208. [CrossRef]
7. Jiang, J. Nonparametric Statistics. In *Large Sample Techniques for Statistics*; Springer International Publishing: Cham, Switzerland, 2022; pp. 379–415. [CrossRef]
8. Hoeffding, W. A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **1948**, *19*, 293–325. [CrossRef]
9. Stute, W. Almost sure representations of the product-limit estimator for truncated data. *Ann. Statist.* **1993**, *21*, 146–156. [CrossRef]
10. Arcones, M.A.; Wang, Y. Some new tests for normality based on U-processes. *Statist. Probab. Lett.* **2006**, *76*, 69–82. [CrossRef]
11. Giné, E.; Mason, D.M. Laws of the iterated logarithm for the local U-statistic process. *J. Theoret. Probab.* **2007**, *20*, 457–485. [CrossRef]
12. Giné, E.; Mason, D.M. On local U-statistic processes and the estimation of densities of functions of several sample variables. *Ann. Statist.* **2007**, *35*, 1105–1145. [CrossRef]
13. Schick, A.; Wang, Y.; Wefelmeyer, W. Tests for normality based on density estimators of convolutions. *Statist. Probab. Lett.* **2011**, *81*, 337–343. [CrossRef]
14. Joly, E.; Lugosi, G. Robust estimation of U-statistics. *Stoch. Process. Appl.* **2016**, *126*, 3760–3773. [CrossRef]
15. Lee, S.; Linton, O.; Whang, Y.J. Testing for stochastic monotonicity. *Econometrica* **2009**, *77*, 585–602. [CrossRef]
16. Ghosal, S.; Sen, A.; van der Vaart, A.W. Testing monotonicity of regression. *Ann. Statist.* **2000**, *28*, 1054–1082. [CrossRef]
17. Arcones, M.A.; Giné, E. Limit theorems for U-processes. *Ann. Probab.* **1993**, *21*, 1494–1542. [CrossRef]
18. Sherman, R.P. Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Ann. Statist.* **1994**, *22*, 439–459. [CrossRef]
19. de la Peña, V.H.; Giné, E. *Decoupling. From Dependence to Independence, Randomly Stopped Processes. U-Statistics and Processes. Martingales and beyond*; Probability and its Applications (New York); Springer: New York, NY, USA 1999; p. xvi+392. [CrossRef]
20. Boniece, B.C.; Horváth, L.; Jacobs, P. Change point detection in high dimensional data with U-statistics. *arXiv* **2022**, arXiv:2207.08933.
21. Kulathinal, S.; Dewan, I. Weighted U-statistics for likelihood-ratio ordering of bivariate data. *Statist. Papers* **2022** . [CrossRef]
22. Minsker, S. U-statistics of growing order and sub-Gaussian mean estimators with sharp constants. *arXiv* **2022**, arXiv:2202.11842.
23. Halmos, P.R. The theory of unbiased estimation. *Ann. Math. Stat.* **1946**, *17*, 34–43. [CrossRef]
24. von Mises, R. On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Stat.* **1947**, *18*, 309–348. [CrossRef]
25. Borovkova, S.; Burton, R.; Dehling, H. Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. *Trans. Amer. Math. Soc.* **2001**, *353*, 4261–4318. [CrossRef]
26. Denker, M.; Keller, G. On U-statistics and v. Mises' statistics for weakly dependent processes. *Z. Wahrsch. Verw. Geb.* **1983**, *64*, 505–522. [CrossRef]
27. Leucht, A. Degenerate U- and V-statistics under weak dependence: asymptotic theory and bootstrap consistency. *Bernoulli* **2012**, *18*, 552–585. [CrossRef]
28. Leucht, A.; Neumann, M.H. Degenerate U- and V-statistics under ergodicity: Asymptotics, bootstrap and applications in statistics. *Ann. Inst. Statist. Math.* **2013**, *65*, 349–386. [CrossRef]
29. Bouzebda, S.; Soukarieh, I. Renewal type bootstrap for U-process Markov chains. *Markov Process. Relat. Fields* **2023** , 1–52.
30. Soukarieh, I.; Bouzebda, S. Exchangeably Weighted Bootstraps of General Markov U-Process. *Mathematics* **2022**, *10*, 3745. . [CrossRef]
31. Soukarieh, I.; Bouzebda, S. Renewal type bootstrap for increasing degree U-Process A Markov Chain. *J. Multivar. Anal.* **2023**, *195*, 105143. [CrossRef]
32. Lee, A.J. Vol. 110, Statistics: Textbooks and Monographs. In *U-statistics. Theory and Practice*; Marcel Dekker Inc.: New York, NY, USA, 1990; p. xii+302.
33. Stute, W. Conditional U-statistics. *Ann. Probab.* **1991**, *19*, 812–825. [CrossRef]
34. Nadaraja, E.A. On a regression estimate. *Teor. Verojatnost. Primenen.* **1964**, *9*, 157–159.
35. Watson, G.S. Smooth regression analysis. *Sankhyā Indian J. Stat. Ser.* **1964**, *26*, 359–372.
36. Sen, A. Uniform strong consistency rates for conditional U-statistics. *Sankhyā Indian J. Stat. Ser.* **1994**, *56*, 179–194.
37. Prakasa Rao, B.L.S.; Sen, A. Limit distributions of conditional U-statistics. *J. Theoret. Probab.* **1995**, *8*, 261–301. [CrossRef]
38. Harel, M.; Puri, M.L. Conditional U-statistics for dependent random variables. *J. Multivar. Anal.* **1996**, *57*, 84–100. [CrossRef]
39. Stute, W. Symmetrized NN-conditional U-statistics. In *Research Developments in Probability and Statistics*; VSP: Utrecht, The Netherlands, 1996; pp. 231–237.
40. Dony, J.; Mason, D.M. Uniform in bandwidth consistency of conditional U-statistics. *Bernoulli* **2008**, *14*, 1108–1133. [CrossRef]
41. Bouzebda, S.; Nemouchi, B. Central limit theorems for conditional empirical and conditional U-processes of stationary mixing sequences. *Math. Methods Statist.* **2019**, *28*, 169–207. [CrossRef]
42. Bouzebda, S.; Nemouchi, B. Uniform consistency and uniform in bandwidth consistency for nonparametric regression estimates and conditional U-statistics involving functional data. *J. Nonparametr. Stat.* **2020**, *32*, 452–509. [CrossRef]

43. Bouzebda, S.; Elhattab, I.; Nemouchi, B. On the uniform-in-bandwidth consistency of the general conditional U-statistics based on the copula representation. *J. Nonparametr. Stat.* **2021**, *33*, 321–358. [CrossRef]
44. Bouzebda, S.; El-hadjali, T.; Ferfache, A.A. Uniform in bandwidth consistency of conditional U-statistics adaptive to intrinsic dimension in presence of censored data. *Sankhya A* **2022**, 1–59. [CrossRef]
45. Bouzebda, S.; Nezzal, A. 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.* **2022**, *5*, 431–533. [CrossRef]
46. Aneiros, G.; Cao, R.; Fraiman, R.; Genest, C.; Vieu, P. Recent advances in functional data analysis and high-dimensional statistics. *J. Multivar. Anal.* **2019**, *170*, 3–9. [CrossRef]
47. Ramsay, J.O.; Silverman, B.W. *Applied Functional Data Analysis*; Springer Series in Statistics; Springer: New York, NY, USA, 2002; p. x+190; Methods and case studies. [CrossRef]
48. Ferraty, F.; Vieu, P. *Nonparametric Functional Data Analysis. Theory and Practice*; Springer Series in Statistics; Springer: New York, NY, USA, 2006; p. xx+258.
49. Araujo, A.; Giné, E. *The Central Limit Theorem for Real and Banach Valued Random Variables*; Wiley Series in Probability and Mathematical Statistics; John Wiley & Sons: New York, NY, USA; Chichester, UK; Brisbane, Australia, 1980; p. xiv+233.
50. Gasser, T.; Hall, P.; Presnell, B. Nonparametric estimation of the mode of a distribution of random curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1998**, *60*, 681–691. [CrossRef]
51. Bosq, D. *Linear Processes in Function Spaces. Theory and Applications*; Lecture Notes in Statistics; Springer: New York, NY, USA, 2000; Volume 149, p. xiv+283. [CrossRef]
52. Horváth, L.; Kokoszka, P. *Inference for Functional Data with Applications*; Springer Series in Statistics; Springer: New York, NY, USA, 2012; p. xiv+422. [CrossRef]
53. Ling, N.; Vieu, P. Nonparametric modelling for functional data: Selected survey and tracks for future. *Statistics* **2018**, *52*, 934–949. [CrossRef]
54. Ferraty, F.; Laksaci, A.; Tadj, A.; Vieu, P. Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference* **2010**, *140*, 335–352. [CrossRef]
55. Kara-Zaitri, L.; Laksaci, A.; Rachdi, M.; Vieu, P. Uniform in bandwidth consistency for various kernel estimators involving functional data. *J. Nonparametr. Stat.* **2017**, *29*, 85–107. [CrossRef]
56. Bouzebda, S.; Chaouch, M. Uniform limit theorems for a class of conditional Z-estimators when covariates are functions. *J. Multivar. Anal.* **2022**, *189*, 21. [CrossRef]
57. Attouch, M.; Laksaci, A.; Rafaa, F. On the local linear estimate for functional regression: Uniform in bandwidth consistency. *Comm. Statist. Theory Methods* **2019**, *48*, 1836–1853. [CrossRef]
58. Ling, N.; Meng, S.; Vieu, P. Uniform consistency rate of kNN regression estimation for functional time series data. *J. Nonparametr. Stat.* **2019**, *31*, 451–468. [CrossRef]
59. Bouzebda, S.; Chaouch, M.; Laïb, N. Limiting law results for a class of conditional mode estimates for functional stationary ergodic data. *Math. Methods Statist.* **2016**, *25*, 168–195. [CrossRef]
60. Mohammedi, M.; Bouzebda, S.; Laksaci, A. The consistency and asymptotic normality of the kernel type expectile regression estimator for functional data. *J. Multivar. Anal.* **2021**, *181*, 24. [CrossRef]
61. Bouzebda, S.; Mohammedi, M.; Laksaci, A. The k-Nearest Neighbors method in single index regression model for functional quasi-associated time series data. *Rev. Mat. Complut.* **2022** , 1–30. [CrossRef]
62. Bouzebda, S.; Nemouchi, B. Weak-convergence of empirical conditional processes and conditional U-processes involving functional mixing data. *Stat. Inference Stoch. Process.* **2022** , 1–56. [CrossRef]
63. Almanjahie, I.M.; Bouzebda, S.; Kaid, Z.; Laksaci, A. Nonparametric estimation of expectile regression in functional dependent data. *J. Nonparametr. Stat.* **2022**, *34*, 250–281. [CrossRef]
64. Almanjahie, I.M.; Bouzebda, S.; Chikr Elmezouar, Z.; Laksaci, A. The functional kNN estimator of the conditional expectile: uniform consistency in number of neighbors. *Stat. Risk Model.* **2022**, *38*, 47–63. [CrossRef]
65. Litimein, O.; Laksaci, A.; Mechab, B.; Bouzebda, S. Local linear estimate of the functional expectile regression. *Statist. Probab. Lett.* **2023**, *192*, 109682. [CrossRef]
66. Chentsov, N.N. Evaluation of an unknown distribution density from observations. *Sov. Math., Dokl.* **1962**, *3*, 1559–1562.
67. Watson, G.S.; Leadbetter, M.R. Hazard analysis. II. *Sankhyā Indian J. Stat. Ser. A* **1964**, *26*, 101–116.
68. Winter, B.B. Rate of strong consistency of two nonparametric density estimators. *Ann. Statist.* **1975**, *3*, 759–766. [CrossRef]
69. Susarla, V.; Walter, G. Estimation of a multivariate density function using delta sequences. *Ann. Statist.* **1981**, *9*, 347–355. [CrossRef]
70. Prakasa Rao, B.L.S. Density estimation for Markov processes using delta-sequences. *Ann. Inst. Statist. Math.* **1978**, *30*, 321–328. [CrossRef]
71. Ahmad, I.A. A note on nonparametric density estimation for dependent variables using a delta sequence. *Ann. Inst. Statist. Math.* **1981**, *33*, 247–254. [CrossRef]
72. Basu, A.K.; Sahoo, D.K. Speed of convergence in density estimation by delta sequence and estimation of mode. *Calcutta Statist. Assoc. Bull.* **1985**, *34*, 1–13. [CrossRef]
73. Nolan, D.; Marron, J.S. Uniform consistency of automatic and location-adaptive delta-sequence estimators. *Probab. Theory Relat. Fields* **1989**, *80*, 619–632. [CrossRef]

74. Isogai, E. Nonparametric estimation of a regression function by delta sequences. *Ann. Inst. Statist. Math.* **1990**, *42*, 699–708. [CrossRef]
75. Marron, J.S.; Härdle, W. Random approximations to some measures of accuracy in nonparametric curve estimation. *J. Multivar. Anal.* **1986**, *20*, 91–113. [CrossRef]
76. Cristóbal Cristóbal, J.A.; Faraldo Roca, P.; González Manteiga, W. A class of linear regression parameter estimators constructed by nonparametric estimation. *Ann. Statist.* **1987**, *15*, 603–609. [CrossRef]
77. Prakasa Rao, B.L.S. Nonparametric density estimation for functional data by delta sequences. *Braz. J. Probab. Stat.* **2010**, *24*, 468–478. [CrossRef]
78. Ouassou, I.; Rachdi, M. Regression operator estimation by delta-sequences method for functional data and its applications. *AStA Adv. Stat. Anal.* **2012**, *96*, 451–465. [CrossRef]
79. Bouzebda, S. On the strong approximation of bootstrapped empirical copula processes with applications. *Math. Methods Statist.* **2012**, *21*, 153–188. [CrossRef]
80. Belitser, E. Local minimax pointwise estimation of a multivariate density. *Statist. Neerl.* **2000**, *54*, 351–365. [CrossRef]
81. Walter, G.G. A general approach to classification problems. *Inform. Sci.* **1983**, *30*, 67–77. [CrossRef]
82. Bouzebda, S.; Nezzal, A. Asymptotic properties of conditional U-statistics using delta sequences. *Comm. Statist. Theory Methods* **2023**, 1–45.
83. Fu, K.A. An application of U-statistics to nonparametric functional data analysis. *Comm. Statist. Theory Methods* **2012**, *41*, 1532–1542. [CrossRef]
84. Dabo-Niang, S. Kernel density estimator in an infinite-dimensional space with a rate of convergence in the case of diffusion process. *Appl. Math. Lett.* **2004**, *17*, 381–386. [CrossRef]
85. Prakasa Rao, B.L.S. *Statistical Inference for Diffusion Type Processes*; Kendall's Library of Statistics; Arnold: London, UK, 1999; Volume 8.
86. Prakasa Rao, B.L.S. *Semimartingales and Their Statistical Inference*; Monographs on Statistics and Applied Probability; Chapman & Hall/ CRC: Boca Raton, FL, USA, 1999; Volume 83.
87. Rachdi, M.; Monsan, V. Asymptotic properties of p-adic spectral estimates of second order. *J. Combin. Inform. Syst. Sci.* **1999**, *24*, 113–142.
88. Prakasa Rao, B.L.S. Nonparametric functional estimation. In *Probability and Mathematical Statistics*; Academic Press (Harcourt Brace Jovanovich, Publishers: Orlando, FL, USA, 1983; Volume XIV, 522p.
89. Ferraty, F.; Vieu, P. Erratum of:Non-parametric models for functional data, with application in regression, time-series prediction and curve discrimination. *J. Nonparametric Stat.* **2008**, *20*, 187–189. [CrossRef]
90. Maouloud, S.M.O. Quelques Aspects Fonctionnels et Non Fonctionnels des Grandes Déviations et des Déviations Modérées en Estimation Non-Paramétrique. Ph.D. Thesis, Université de Reims-Champagne Ardenne, Reims, France, 2007.
91. Walter, G.G. Expansions of distributions. *Trans. Am. Math. Soc.* **1965**, *116*, 492–510. [CrossRef]
92. Walter, G.; Blum, J. Probability density estimation using delta sequences. *Ann. Stat.* **1979**, 328–340. [CrossRef]
93. Deheuvels, P. One bootstrap suffices to generate sharp uniform bounds in functional estimation. *Kybernetika* **2011**, *47*, 855–865.
94. Bouzebda, S.; Taachouche, N. Rates of the strong uniform consistency for the kernel-type regression function estimators with general kernels on manifolds. *Math. Methods Statist.* **2022**, 1–62.
95. Chokri, K.; Bouzebda, S. Uniform-in-bandwidth consistency results in the partially linear additive model components estimation. *Comm. Statist. Theory Methods* **2022**, *2022*, 2153605. [CrossRef]
96. Maillot, B.; Viallon, V. Uniform limit laws of the logarithm for nonparametric estimators of the regression function in presence of censored data. *Math. Methods Statist.* **2009**, *18*, 159–184. [CrossRef]
97. Kohler, M.; Máthé, K.; Pintér, M. Prediction from randomly right censored data. *J. Multivar. Anal.* **2002**, *80*, 73–100. [CrossRef]
98. Carbonez, A.; Györfi, L.; van der Meulen, E.C. Partitioning-estimates of a regression function under random censoring. *Statist. Decis.* **1995**, *13*, 21–37. [CrossRef]
99. Kaplan, E.L.; Meier, P. Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **1958**, *53*, 457–481. [CrossRef]
100. Brunel, E.; Comte, F. Adaptive nonparametric regression estimation in presence of right censoring. *Math. Methods Statist.* **2006**, *15*, 233–255.
101. Datta, S.; Bandyopadhyay, D.; Satten, G.A. Inverse probability of censoring weighted U-statistics for right-censored data with an application to testing hypotheses. *Scand. J. Stat.* **2010**, *37*, 680–700. [CrossRef]
102. Stute, W.; Wang, J.L. Multi-sample U-statistics for censored data. *Scand. J. Statist.* **1993**, *20*, 369–374.
103. Chen, Y.; Datta, S. Adjustments of multi-sample U-statistics to right censored data and confounding covariates. *Comput. Statist. Data Anal.* **2019**, *135*, 1–14. [CrossRef]
104. Yuan, A.; Giurcanu, M.; Luta, G.; Tan, M.T. U-statistics with conditional kernels for incomplete data models. *Ann. Inst. Statist. Math.* **2017**, *69*, 271–302. [CrossRef]
105. Bouzebda, S.; El-hadjali, T. Uniform convergence rate of the kernel regression estimator adaptive to intrinsic dimension in presence of censored data. *J. Nonparametr. Stat.* **2020**, *32*, 864–914. [CrossRef]
106. Földes, A.; Rejtő, L. A LIL type result for the product limit estimator. *Z. Wahrsch. Verw. Geb.* **1981**, *56*, 75–86. [CrossRef]
107. Kendall, M.G. A New Measure of Rank Correlation. *Biometrika* **1938**, *30*, 81–93. [CrossRef]

108. Stute, W. Universally consistent conditional U-statistics. *Ann. Statist.* **1994**, *22*, 460–473. [CrossRef]
109. Stute, W. L^p-convergence of conditional U-statistics. *J. Multivar. Anal.* **1994**, *51*, 71–82. [CrossRef]
110. Bellet, A.; Habrard, A.; Sebban, M. A Survey on Metric Learning for Feature Vectors and Structured Data. *arXiv* **2013**, arXiv:1306.6709.
111. Clémençon, S.; Colin, I.; Bellet, A. Scaling-up empirical risk minimization: Optimization of incomplete U-statistics. *J. Mach. Learn. Res.* **2016**, *17*, 2682–2717.
112. Jin, R.; Wang, S.; Zhou, Y. Regularized Distance Metric Learning:Theory and Algorithm. In *Advances in Neural Information Processing Systems*; Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C., Culotta, A., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2009; Volume 22.
113. Bellet, A.; Habrard, A. Robustness and generalization for metric learning. *Neurocomputing* **2015**, *151*, 259–267. [CrossRef]
114. Cao, Q.; Guo, Z.C.; Ying, Y. Generalization bounds for metric and similarity learning. *Mach. Learn.* **2016**, *102*, 115–132. [CrossRef]
115. Blum, J.R.; Kiefer, J.; Rosenblatt, M. Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.* **1961**, *32*, 485–498. [CrossRef]
116. Bouzebda, S. General tests of independence based on empirical processes indexed by functions. *Stat. Methodol.* **2014**, *21*, 59–87. [CrossRef]
117. Bouzebda, S. Some new multivariate tests of independence. *Math. Methods Statist.* **2011**, *20*, 192–205. [CrossRef]
118. Bouzebda, S.; Zari, T. Asymptotic behavior of weighted multivariate Cramér-von Mises-type statistics under contiguous alternatives. *Math. Methods Statist.* **2013**, *22*, 226–252. [CrossRef]
119. Bouzebda, S.; El Faouzi, N.E.; Zari, T. On the multivariate two-sample problem using strong approximations of empirical copula processes. *Comm. Statist. Theory Methods* **2011**, *40*, 1490–1509. [CrossRef]
120. Bouzebda, S. General Tests of Conditional Independence Based on Empirical Processes Indexed by Functions. *Jpn. J. Stat. Data Sci.* **2023** , 1–60.
121. Bergsma, W.; Dassios, A. A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli* **2014**, *20*, 1006–1028. [CrossRef]
122. Bouzebda, S.; Soukarieh, I. Non-Parametric Conditional U-Processes for Locally Stationary Functional Random Fields under Stochastic Sampling Design. *Mathematics* **2023**, *11*, 16. [CrossRef]
123. Didi, S.; Bouzebda, S. Wavelet Density and Regression Estimators for Continuous Time Functional Stationary and Ergodic Processes. *Mathematics* **2022**, *10*, 4356. [CrossRef]
124. Rachdi, M.; Vieu, P. Nonparametric regression for functional data: Automatic smoothing parameter selection. *J. Statist. Plann. Inference* **2007**, *137*, 2784–2801. [CrossRef]
125. Arcones, M.A. A Bernstein-type inequality for U-statistics and U-processes. *Stat. Probab. Lett.* **1995**, *22*, 239–247. [CrossRef]
126. Serfling, R.J. *Approximation Theorems of Mathematical Statistics*; Wiley Series in Probability and Mathematical Statistics; John Wiley & Sons, Inc.: New York, NY, USA, 1980; p. xiv+371.