Uniform in Bandwidth Consistency of Conditional U-statistics Adaptive to Intrinsic Dimension in Presence of Censored Data

Salim Bouzebda and Anouar Abdeldjaoued Ferfache
Université de Technologie de Compiègne, LMAC (Laboratory of Applied Mathematics of Compiègne), Compiègne Cedex, France
Thouria El-hadjali
Université Frères Mentouri, Constantine 1, Algeria

Abstract

*U-statistics represent a fundamental class of statistics from modelling quantities of interest defined by multi-subject responses. U-statistics generalize the empirical mean of a random variable X to sums over every m-tuple of distinct observations of Stute (Ann. Probab. **19**, 812–825 1991) introduced a class of so-called conditional U-statistics, which may be viewed as a generalization of the Nadaraya-Watson estimates of a regression function. Stute proved their strong pointwise consistency to:*

$$r(\mathbf{t}) := \mathbb{E}[\varphi(Y_1, \ldots, Y_m)|(X_1, \ldots, X_m) = \mathbf{t}], \quad \text{for} \quad \mathbf{t} \in \mathbb{R}^{dm}.$$

*We apply the methods developed in Dony and Mason (Bernoulli **14**(4), 1108–1133 2008) to establish uniform in $\mathbf{t}$ and in bandwidth consistency (i.e., h, $h \in [a_n, b_n]$ where $0 < a_n < b_n \to 0$ at some specific rate) to $r(\mathbf{t})$ of the estimator proposed by Stute when Y, under weaker conditions on the kernel than previously used in the literature. We extend existing uniform bounds on the kernel conditional U-statistic estimator and make it adaptive to the intrinsic dimension of the underlying distribution of $\mathbf{X}$ which the so-called intrinsic dimension will characterize. In addition, uniform consistency is also established over $\varphi \in \mathscr{F}$ for a suitably restricted class $\mathscr{F}$, in both cases bounded and unbounded, satisfying some moment conditions. Our theorems allow data-driven local bandwidths for these statistics. Moreover, in the same context, we show the uniform bandwidth consistency for the nonparametric inverse probability of censoring weighted (I.P.C.W.) estimators of the regression function under random censorship, which is of its own interest. The theoretical uniform consistency results established in this paper are (or will be) key tools for many further developments in regression analysis.*

empirical processes, conditional U-processes, kernel estimation, functional estimation, VC-classes.

1   Introduction

Motivated by numerous applications, the theory of U-statistics (introduced in the seminal work by Hoeffding (1948)) and U-processes have received considerable attention in the past decades. U-processes are useful for solving complex statistical problems. Examples are density estimation, nonparametric regression tests and goodness-of-fit tests. More precisely, U-processes appear in statistics in many instances, e.g., as the components of higher order terms in von Mises expansions. In particular, U-statistics play a role in the analysis of estimators (including function estimators) with varying degrees of smoothness. For example, Stute (1993) applies a.s. uniform bounds for $\mathbb{P}$-canonical U-processes to the analysis of the product limit estimator for truncated data. Arcones and Wang (2006) present two new tests for normality based on U-processes. Making use of the results of Giné and Mason (2007a), Giné and Mason (2007b), Schick et al. (2011) introduced new tests for normality which used as test statistics weighted L_1-distances between the standard normal density and local U-statistics based on standardized observations. Joly and Lugosi (2016) discussed the estimation of the mean of multivariate functions in case of possibly heavy-tailed distributions and introduced the median-of-means, which is based on U-statistics. U-processes are important tools for a broad range of statistical applications such as testing for qualitative features of functions in nonparametric statistics (Lee et al., 2009; Ghosal et al., 2000; Abrevaya and Jiang, 2005) and establishing limiting distributions of M-estimators (see, e.g., Arcones and Giné 1993; Sherman 1993; Sherman 1994; de la Peña and Giné 1999). Halmos (1946), von Mises (1947) and Hoeffding (1948), who provided (amongst others) the first asymptotic results for the case that the underlying random variables are independent and identically distributed. Under weak dependency assumptions asymptotic results are for instance shown in Borovkova et al. (2001), in Denker and Keller (1983) or more recently in Leucht (2012) and in more general setting in Leucht and Neumann (2013), Bouzebda and Nemouchi (2019), Bouzebda and Nemouchi (2022). For excellent resource of references on the U-statistics and U-processes, the interested reader may refer to Borovskikh (1996), Koroljuk and Borovskich (1994), Lee (1990), Arcones and Giné (1995), Arcones et al. (1994) and Arcones and Giné (1993). A profound insight into the theory of U-processes is given by de la Peña and Giné (1999). In this paper, we consider the so-called conditional U-statistics

introduced by Stute ([1991](#)). These statistics may be viewed as generalizations of the Nadaraya-Watson (Nadaraja, [1964](#); Watson, [1964](#)) estimates of a regression function.

To be more precise, let us consider a 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 $\mathbf{Y}_i \in \mathbb{R}^q$, $d, q \geq 1$. Let $\mathbb{P}_{\mathbf{X}} = \mathbb{P}$ be an unknown marginal Borel probability distribution in $\mathbb{R}^d$. Let $\varphi : \mathbb{R}^{qm} \to \mathbb{R}$ be a measurable function. In this paper, we are primarily concerned with the estimation of the conditional expectation, or regression function of $\varphi(\mathbf{Y}_1, \ldots, \mathbf{Y}_m)$ evaluated at $(\mathbf{X}_1, \ldots, \mathbf{X}_m) = \mathbf{t}$, given by

$$r^{(m)}(\varphi, \mathbf{t}) \; = \; \mathbb{E}\left(\varphi(\mathbf{Y}_1, \ldots, \mathbf{Y}_m) \,|\, (\mathbf{X}_1, \ldots, \mathbf{X}_m) = \mathbf{t}\right), \quad \text{for } \mathbf{t} \in \mathbb{R}^{dm}, (1.1)$$

whenever it exists, i.e., $\mathbb{E}\left(|\varphi(\mathbf{Y}_1, \ldots, \mathbf{Y}_m)|\right) < \infty$. We now introduce a kernel function $K : \mathbb{R}^{dm} \to \mathbb{R}$. Stute ([1991](#)) presented a class of estimators for $r^{(m)}(\varphi, \mathbf{t})$, called the conditional U-statistics, which is defined for each $\mathbf{t} \in \mathbb{R}^{dm}$ to be :

$$\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n) = \frac{\displaystyle\sum_{(i_1, \ldots, i_m) \in I(m,n)} \varphi(Y_{i_1}, \ldots, Y_{i_m}) K\left(\frac{\mathbf{t}_1 - \mathbf{X}_{i_1}}{h_n}\right) \cdots K\left(\frac{\mathbf{t}_m - \mathbf{X}_{i_m}}{h_n}\right)}{\displaystyle\sum_{(i_1, \ldots, i_m) \in I(m,n)} K\left(\frac{\mathbf{t}_1 - \mathbf{X}_{i_1}}{h_n}\right) \cdots K\left(\frac{\mathbf{t}_m - \mathbf{X}_{i_m}}{h_n}\right)},$$

$$(1.2)$$

where:

$$I(m, n) = \left\{\mathbf{i} = (i_1, \ldots, i_m) : 1 \leq i_j \leq n \quad \text{and} \quad i_j \neq i_r \quad \text{if} \quad j \neq r\right\}, \quad (1.3)$$

is the set of all m-tuples of different integers between 1 and n and $\{h_n\}_{n \geq 1}$ denotes a sequence of positive constants converging to zero and $nh_n^m \to \infty$. For notational simplicity, we let $h = h_n$. In the particular case $m = 1$, the $r^{(m)}(\varphi, \mathbf{t})$ is reduced to $r^{(1)}(\varphi, \mathbf{t}) = \mathbb{E}(\varphi(\mathbf{Y})|\mathbf{X} = \mathbf{t})$ and Stute's estimator becomes the Nadaraya-Watson estimator of $r^{(1)}(\varphi, \mathbf{t})$ given by:

$$\widehat{r}_n^{(1)}(\varphi, \mathbf{t}, h) = \sum_{i=1}^n \varphi(\mathbf{Y}_i) K\left(\frac{\mathbf{X}_i - \mathbf{t}}{h}\right) \bigg/ \sum_{i=1}^n K\left(\frac{\mathbf{X}_i - \mathbf{t}}{h}\right).$$

The work of Sen ([1994](#)) was devoted to estimate the rate of the uniform convergence in $\mathbf{t}$ of $\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h)$ to $r^{(m)}(\varphi, \mathbf{t})$. In the paper of Prakasa Rao and Sen ([1995](#)), the limit distributions of $\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h)$ are discussed and compared with those obtained by Stute. Harel and Puri ([1996](#)) extend the results

of Stute (1991), under appropriate mixing conditions, to weakly dependent data and have applied their findings to verify the Bayes risk consistency of the corresponding discrimination rules. Stute (1996) proposed symmetrized nearest neighbour conditional U-statistics as alternatives to the usual kernel-type estimators. An important contribution is given in the paper Dony and Mason (2008) where a much stronger form of consistency holds, namely, uniform in $\mathbf{t}$ and in bandwidth consistency (i.e., $h \in [a_n, b_n]$ where $a_n < b_n \to 0$ at some specific rate) of $\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h)$. In addition, uniform consistency is also established over $\varphi \in \mathscr{F}$ for a suitably restricted class $\mathscr{F}$. The main tool in their result is the use of the local conditional U-process investigated in Giné and Mason (2007a). In the last decades, empirical process theory has provided very useful and powerful tools to analyze the large sample properties of several nonparametric estimators of functionals of the distribution, such as the regression function and the density function, refer to, van der Vaart and Wellner (1996) and Kosorok (2008). Nolan and Pollard (1987) were the first to introduce the notion of uniform in bandwidth consistency for kernel density estimators and they applied empirical process methods in their study. In the series of papers, Deheuvels (2000), Deheuvels and Mason (2004), Einmahl and Mason (2005), Dony and Mason (2008), Maillot and Viallon (2009), Mason and Swanepoel (2011), Bouzebda and Elhattab (2009, 2010), Bouzebda (2012), Bouzebda et al. (2018, 2021), Bouzebda and Nemouchi (2020), Bouzebda and El-hadjali (2020), Bouzebda and Nezzal (2022) the authors established uniform consistency results for such kernel estimators, where h varies within suitably chosen intervals indexed by n. More precisely, we will consider one of the most commonly used classes of estimators that is formed by the so-called kernel-type estimators. There are basically no restrictions on the choice of the kernel function in our setup, apart from satisfying some mild conditions that we will give after. The selection of the bandwidth, however, is more problematic. It is worth noticing that the choice of the bandwidth is crucial to obtain a good rate of consistency, for example, it has a big influence on the size of the estimate's bias. In general, we are interested in the selection of bandwidth that produces an estimator which has a good balance between the bias and the variance of the considered estimators. It is then more appropriate to consider the bandwidth varying according to the criteria applied and the available data and location, which cannot be achieved by using classical methods. The interested reader may refer to Mason (2012) for more details and discussion on the subject. In the present paper, we develop methods that permit the study of the kernel-type under nonrestrictive conditions. It is worth noticing that the high-dimensional data sets have several unfortunate properties

that make them hard to analyze. The phenomenon that the computational and statistical efficiency of statistical techniques degrade rapidly with the dimension is often referred to as the "curse of dimensionality". Density and regression estimation on manifolds has received much less attention than the "full-dimensional" counterpart. However, understanding density estimation in situations where the intrinsic dimension can be much lower than the ambient dimension is becoming ever more important: modern systems can capture data at an increasing resolution while the number of degrees of freedom stays relatively constant. One of the limiting aspects of density (regression)-based approaches is their performance in high dimensions.

We know that the notion of the intrinsic dimension, say d_M, has been studied in the statistical machine learning literature so as to establish fast estimation rates in high-dimensional kernel regression settings. There are numerous known techniques for doing so e.g. Kégl (2002), Levina and Bickel (2004), Hein and Audibert (2005), Farahmand et al. (2007). We first introduce a concept proposed by Kim et al. (2018, 2019), the so-called *volume dimension*, to characterize the intrinsic dimension of the underlying distribution. More specifically, the volume dimension d_{vol} is the decay rate of the probability of vanishing Euclidean balls. Let $\|\cdot\|$ be the Euclidean 2-norm. For $\mathbf{x} \in \mathbb{R}^d$ and $r > 0$, we use the notation $\mathbb{B}_{\mathbb{R}^d}(\mathbf{x}, r)$ for the open Euclidean ball centered at $\mathbf{x}$ and radius r, i.e.,

$$\mathbb{B}_{\mathbb{R}^d}(\mathbf{x}, r) = \left\{ \mathbf{y} \in \mathbb{R}^d : \|\mathbf{y} - \mathbf{x}\| < r \right\}.$$

When a probability distribution $\mathbb{P}$ has a bounded density $f_{\mathbf{X}}(\cdot)$ supported on a well-behaved manifold M of dimension d_M, it is known that, for any point $\mathbf{x} \in M$, the measure on the ball $\mathbb{B}_{\mathbb{R}^d}(\mathbf{x}, r)$ centered at $\mathbf{x}$ and radius r decays as

$$\mathbb{P}\left(\mathbb{B}_{\mathbb{R}^d}(\mathbf{x}, r)\right) \sim r^{d_M},$$

when r is small enough. From this, Kim et al. (2018) define the volume dimension of a probability distribution $\mathbb{P}$ to be the maximum possible exponent rate that can dominate the probability volume decay on balls, i.e., fix a subset $\mathbb{X} \subset \mathbb{R}^d$, then

$$d_{\mathrm{vol}}(\mathbb{P}) := \sup \left\{ \nu \geq 0 : \limsup_{r \to 0} \sup_{\mathbf{x} \in \mathbb{X}} \frac{\mathbb{P}(\mathbb{B}_{\mathbb{R}^d}(\mathbf{x}, r))}{r^{\nu}} < \infty \right\}. \qquad (1.4)$$

The primary purpose of the present work is to extend the work of Kim et al. (2018) to the more general estimators, including the kernel density estimator as a particular case (studied in Kim et al. 2018), this generalization is far

from being trivial and harder to control some complex classes of functions, which form a basically unsolved open problem in the literature. We aim to fill this gap in the literature by combining results Kim et al. (2018) with techniques developed in Einmahl and Mason (2005) and Dony and Mason (2008). The present extends our previous work in Bouzebda and El-hadjali (2020) in several directions; the main one is that the present paper considers the conditional U-statistics, including the regression treated in the last mentioned paper. More precisely, a uniform in bandwidth consistency results for some general kernel-type estimators are established. However, as will be seen later, the problem requires much more than "simply" combining ideas from the existing results. Delicate mathematical derivations will be required to cope with the empirical processes that we consider in this extended setting. In addition, we will consider the nonparametric Inverse Probability of Censoring Weighted (I.P.C.W.) estimators of the multivariate regression function under random censorship and obtain uniform in bandwidth consistency results that are of independent interest.

An outline of the remainder of our paper is as follows. In the forthcoming section, we introduce the mathematical framework and provide our main results concerning the uniform in bandwidth consistency of conditional U-statistics adaptive to intrinsic dimension extending the setting of the work of Dony and Mason (2008). In Section 3, we consider the conditional U-statistics in the right censored data framework. Examples of U-statistics kernel are provided in Section 3.1. In Section 4, we present how to select the bandwidth through the cross-validation procedures. An application to the nonparametric discrimination problem is discussed in Section 4.1. Some concluding remarks are given in Section 5. To prevent interrupting the presentation flow, all proofs are gathered in Section 6. A few relevant technical results are given in the Appendix for easy reference.

2   Main Results

For a fixed integer $m \leq n$, consider a class $\mathscr{F}_q$ of measurable functions $g : \mathbb{R}^{qm} \to \mathbb{R}$ defined on $\mathbb{R}^{qm}$, such that $\mathbb{E}g^2(\mathbf{Y}_1, \ldots, \mathbf{Y}_m) < \infty$, which satisfies the conditions, (F.i)–(F.iii) given below. First, to avoid measurability problems, we assume that

$$\mathscr{F}_q \text{ is a pointwise measurable class,} \qquad \text{(F.i)}$$

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

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

This condition is discussed in van der Vaart and Wellner (1996). We also assume that $\mathscr{F}_q$ has a measurable envelope function

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

Notice that condition (F.i) implies that the supremum in Eq. F.ii is measurable. Finally, we assume that $\mathscr{F}_q$ is of VC-type, with characteristics A and v ("VC" for Vapnik and Červonenkis), meaning that for some $A \geq 3$ and $v_1 \geq 1$,

$$\mathcal{N}(\mathscr{F}_q, L_2(Q), \varepsilon) \leq \left(\frac{A\|F\|_{L_2(Q)}}{\varepsilon} \right)^{v_1}, \quad \text{for} \quad 0 < \varepsilon \leq 2\|F\|_{L_2(Q)}, \tag{F.iii}$$

where Q is any probability measure on $(\mathbb{R}^{qm}, \mathcal{B})$, where $\mathcal{B}$ represents the σ-field of Borel sets of $\mathbb{R}^{qm}$, such that $\|F\|_{L_2(Q)} < \infty$, and where for $\varepsilon > 0$, $\mathcal{N}(\mathscr{F}_q, L_2(Q), \varepsilon)$ is defined as the smallest number of $L_2(Q)$-open balls of radius ε required to cover $\mathscr{F}_q$. (If Eq. F.iii holds for $\mathscr{F}_q$, then we say that the VC-type class $\mathscr{F}_q$ admits the characteristics A and v_1). In this section, we follow Kim et al. (2018) for weakening the conditions on the kernel and making it adaptive to the intrinsic dimension of the underlying distribution and without assumptions on the distribution. It is worth noticing that for general distributions such as the one with support is a lower-dimensional manifold, the usual change of variables argument is no longer directly applicable. However, we can provide a bound based on the volume dimension under an integrability condition on the kernel, given below. Let $K(\cdot)$ be a kernel function defined on $\mathbb{R}^{dm}$, that is a measurable function satisfying

$$\int_{\mathbb{R}^d} K(\mathbf{t})d\mathbf{t} = 1.$$

ASSUMPTION 1. *(Integrability condition) Let* $\|K\|_\infty = \sup_{\mathbf{x} \in \mathbb{R}^d} |K(\mathbf{x})| =: \kappa < \infty$, *and fix* $k > 0$. *We have: either* $d_{vol} = 0$ *or*

$$\int_0^\infty x^{d_{vol}-1} \sup_{\|\mathbf{t}\| \geq x} |K|^k(\mathbf{t})dx < \infty. \tag{K.ii}$$

REMARK 2.1. (Kim et al., 2018) It is important to emphasize that Assumption 1 is weak, as it is satisfied by commonly used kernels. For instance, if the kernel function $K(\mathbf{x})$ decays at a polynomial rate strictly faster than d_{vol}/k (which is at most d/k) as $\mathbf{x} \to \infty$, that is, if

$$\limsup_{\mathbf{x} \to \infty} \|\mathbf{x}\|^{d_{\mathrm{vol}}/k+\epsilon} K(x) < \infty$$

for any $\epsilon > 0$, the integrability condition Eq. K.ii is satisfied. Also, if the kernel function $K(\mathbf{x})$ is spherically symmetric, that is, if there exists $\widetilde{K} : [0, \infty) \to \mathbb{R}$ with $K(\mathbf{x}) = \widetilde{K}(\|\mathbf{x}\|_2)$, then the integrability condition Eq. K.ii is satisfied provided $\|K\|_k < \infty$. Kernels with bounded support also satisfy the condition Eq. K.ii. Thus, most of the commonly used kernels including Uniform, Epanechnikov, and Gaussian kernels satisfy the above integrability condition.

Now, we consider the class of functions

$$\mathscr{K} := \left\{ (\mathbf{t}, h) \mapsto K\left(\frac{\mathbf{t} - \cdot}{h}\right) : \mathbf{t} \in \mathbb{X}, h \geq l_n \right\},$$

with $\|K\|_2 < \infty$.

ASSUMPTION 2. *Assume that $\mathscr{K}$ is bounded VC-class with envelope κ and dimension ν_2, i.e., there exists positive number $A_2 \geq 1$ and $\nu_2 \geq 1$ such that, for every probability measure $\mathbb{Q}$ on $\mathbb{R}^d$ and for every $\varepsilon \in (0, \|K\|_\infty)$, the covering numbers $\mathcal{N}(\mathcal{K}, L_2(\mathbb{Q}), \varepsilon)$ satisfies*

$$\mathcal{N}(\mathcal{K}, L_2(\mathbb{Q}), \varepsilon) \leq \left(\frac{A_2 \|K\|_\infty}{\varepsilon}\right)^{\nu_2}. \tag{K.iii}$$

Furthermore, let

$$\widetilde{\mathbf{K}}(\mathbf{t}) := \prod_{j=1}^m K(\mathbf{t}_j), \quad \mathbf{t} = (\mathbf{t}_1, \ldots, \mathbf{t}_m) \tag{K.iv}$$

denote the product kernel. Next, if $(S, \mathcal{S})$ is a measurable space, define the general U-statistic with kernel $H : S^k \to \mathbb{R}$ based on S-valued random variables $Z_1, \cdots, Z_n$ as

$$U_n^{(k)}(H) := \frac{(n-k)!}{n!} \sum_{i \in I(k,n)} H(Z_{i_1}, \ldots, Z_{i_k}), \quad 1 \leq k \leq n, \tag{2.1}$$

where $I(k, n)$ is defined as in Eq. 1.3 with $m = k$. Note that we do not require H to be symmetric here. For a bandwidth $0 < h$ and $g \in \mathscr{F}_q$, consider the U-kernel

$$G_{g,h,\mathbf{t}}(\mathbf{x}, \mathbf{y}) := g(\mathbf{y})\widetilde{\mathbf{K}}_h(\mathbf{t} - \mathbf{x}) \qquad \mathbf{x}, \mathbf{t} \in \mathbb{R}^{dm} \quad \text{and} \quad \mathbf{y} \in \mathbb{R}^{qm},$$

where, as usual, $K_h(z) = h^{-d}K(z/h)$, $z \in \mathbb{R}^d$, and for the sample $(\mathbf{X}_1, \mathbf{Y}_1)$, $\ldots, (\mathbf{X}_n, \mathbf{Y}_n)$, define

$$U_n(g, h, \mathbf{t}) := U_n^{(m)}(G_{g,h,\mathbf{t}}) = \frac{(n-m)!}{n!} \sum_{i \in I_n^m} G_{g,h,\mathbf{t}}(\mathbf{X}_i, \mathbf{Y}_i),$$

where, throughout this paper, we shall use the notation

$$\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_m) \in \mathbb{R}^{dm} \quad \text{and} \quad \mathbf{X}_\mathbf{i} = (\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_k}) \in \mathbb{R}^{dk} \quad \mathbf{i} \in \mathbf{I}_n^k,$$
$$\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_m) \in \mathbb{R}^{qm} \quad \text{and} \quad \mathbf{Y}_\mathbf{i} = (\mathbf{Y}_{i_1}, \ldots, \mathbf{Y}_{i_k}) \in \mathbb{R}^{dk} \quad \mathbf{i} \in \mathbf{I}_n^k.$$

Now, introduce the U-statistic process

$$u_n(g, h, \mathbf{t}) := \sqrt{n}\{U_n(g, h, \mathbf{t}) - \mathbb{E}U_n(g, h, \mathbf{t})\}. \tag{2.2}$$

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

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

where

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

Introduce the class of functions defined on the compact subset $\mathbf{J}^m$ of $\mathbb{R}^{dm}$,

$$\mathcal{M} = \left\{ r^{(m)}(\varphi, \cdot)\widetilde{f}(\cdot) : \varphi \in \mathscr{F}_q \right\},$$

where $r^{(m)}(\varphi, \cdot)$ is defined in Eq. 1.1 and the function $\widetilde{f} : \mathbb{R}^{dm} \to \mathbb{R}$ is defined as

$$\widetilde{f}(\mathbf{t}) := \int_{\mathbb{R}^{qm}} f(\mathbf{t}_1, \mathbf{y}_1) \cdots f(\mathbf{t}_m, \mathbf{y}_m) \, \mathrm{d}\mathbf{y}_1 \ldots \mathrm{d}\mathbf{y}_m = f_\mathbf{X}(\mathbf{t}_1) \ldots f_\mathbf{X}(\mathbf{t}_m),$$

where $f(\cdot, \cdot)$ denote the joint density of $(\mathbf{X}, \mathbf{Y})$. We fix a subset $\mathbb{X} \subset \mathbb{R}^d$ on which we are considering the uniform convergence of the kernel regression estimator. We first characterize the intrinsic dimension of the distribution $\mathbb{P}$, proposed by Kim et al. (2018), by its rate of the probability volume growth on balls. If a probability distribution has a positive measure on a manifold with a positive reach, then the volume dimension is always between 0 and the manifold's dimension. In particular, the volume dimension of any probability distribution is between 0 and the ambient dimension d.

LEMMA 2.2. (Kim et al., 2018) *Let $\mathbb{P}$ be a probability distribution on $\mathbb{R}^d$, and d_{vol} be its volume dimension. Then for any $\nu \in [0, d_{vol})$, there exists a constant $C_{\nu,\mathbb{P}}$ depending only on $\mathbb{P}$ and ν such that for all $\mathbf{x} \in \mathbb{X}$ and $r > 0$,*

$$\frac{\mathbb{P}(\mathbb{B}_{\mathbb{R}^d}(\mathbf{x}, r))}{r^\nu} \le C_{\nu,\mathbb{P}}. \tag{2.3}$$

For the exact optimal rate, we impose conditions on how the probability volume decay in Eq. 2.3.

ASSUMPTION 3. *Let $\mathbb{P}$ be a probability distribution on $\mathbb{R}^d$, and d_{vol} be its volume dimension. For $\nu \in [0, d_{vol})$, we assume that*

$$\limsup_{r \to 0} \sup_{\mathbf{x} \in \mathbb{X}} \frac{\mathbb{P}(\mathbb{B}_{\mathbb{R}^d}(\mathbf{x}, r))}{r^\nu} < \infty. \tag{2.4}$$

ASSUMPTION 4. *Let $\mathbb{P}$ be a probability distribution on $\mathbb{R}^d$, and d_{vol} be its volume dimension. For $\nu \in [0, d_{vol})$, we assume that*

$$\sup_{\mathbf{x} \in \mathbb{X}} \liminf_{r \to 0} \frac{\mathbb{P}(\mathbb{B}_{\mathbb{R}^d}(\mathbf{x}, r))}{r^\nu} > 0. \tag{2.5}$$

These assumptions are in fact weak and hold for common probability distributions. In particular, Assumptions 3 and 4 hold when the probability distribution has a bounded density with respect to the d-dimensional Lebesgue measure. By combining Assumption 1 and Lemma (2.0.2) of Kim et al. (2018), we can bound $\mathbb{E}_{\mathbb{P}}\left[K^2\right]$ in terms of the volume dimension d_{vol}.

LEMMA 2.3. *Let $(\mathbb{R}^d, \mathbb{P})$ be a probability space and let $X \sim \mathbb{P}$. For any kernel $K(\cdot)$ satisfying Assumption 1 with $k > 0$, the expectation of the k-moment of the kernel is upper bounded as*

$$\mathbb{E}_{\mathbb{P}}\left[\left|K\left(\frac{\mathbf{t} - \mathbf{X}}{h}\right)\right|^k\right] \leq C_{k,\mathbb{P},K,\varepsilon} h^{d_{vol}-\varepsilon} \tag{2.6}$$

for any $\varepsilon \in (0, d_{vol})$, where $C_{k,\mathbb{P},K,\varepsilon}$ is a constant depending only on $k, \mathbb{P}, K$ and ε. Further, if $d_{vol} = 0$ or under Assumption 1 in Kim et al. (2018), ε can be 0 in Eq. 2.6.

We give an example from Kim et al. (2018) of an unbounded density. In this case, the volume dimension is strictly smaller than the dimension of the support, which illustrates why the dimension of the support is not enough to characterize the dimensionality of a distribution.

EXAMPLE 2.4. (Kim et al., 2018) Let $\mathbb{P}$ be a distribution on $\mathbb{R}^d$ having a density p with respect to the d-dimensional Lebesgue measure. Fix $\beta < d$, and suppose $p : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$p(\mathbf{x}) = \frac{(d - \beta)\Gamma\left(\frac{d}{2}\right)}{2\pi^{\frac{d}{2}}} \|\mathbf{x}\|_2^{-\beta} \mathbb{1}(\|\mathbf{x}\|_2 \leq 1).$$

Then, for each fixed $r > 0$,

$$\sup_{\mathbf{x} \in \mathbb{R}^d} \mathbb{P}(\mathbb{B}_{\mathbb{R}^d}(\mathbf{x}, r)) = \mathbb{P}(\mathbb{B}_{\mathbb{R}^d}(\mathbf{0}, r)) = r^{d-\beta}.$$

Hence from definition in Eq. 1.4, the volume dimension is

$$d_{\mathrm{vol}}(\mathbb{P}) = d - \beta.$$

In our setting, we will use

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

$$= \begin{cases} \dfrac{\displaystyle\sum_{(i_1,\ldots,i_m) \in I(m,n)} \varphi(Y_{i_1}, \ldots, Y_{i_m}) K\left(\dfrac{\mathbf{t}_1 - \mathbf{X}_{i_1}}{h}\right) \cdots K\left(\dfrac{\mathbf{t}_m - \mathbf{X}_{i_m}}{h}\right)}{\displaystyle\sum_{(i_1,\ldots,i_m) \in I(m,n)} K\left(\dfrac{\mathbf{t}_1 - \mathbf{X}_{i_1}}{h}\right) \cdots K\left(\dfrac{\mathbf{t}_m - \mathbf{X}_{i_m}}{h}\right)}, \\[2em] \text{if} \displaystyle\sum_{(i_1,\ldots,i_m) \in I(m,n)} K\left(\dfrac{\mathbf{t}_1 - \mathbf{X}_{i_1}}{h}\right) \cdots K\left(\dfrac{\mathbf{t}_m - \mathbf{X}_{i_m}}{h}\right) \neq 0, \\[2em] \dfrac{(n-m)!}{n!} \displaystyle\sum_{(i_1,\ldots,i_m) \in I(m,n)} \varphi(Y_{i_1}, \ldots, Y_{i_m}) \\[1em] \text{if} \displaystyle\sum_{(i_1,\ldots,i_m) \in I(m,n)} K\left(\dfrac{\mathbf{t}_1 - \mathbf{X}_{i_1}}{h}\right) \cdots K\left(\dfrac{\mathbf{t}_m - \mathbf{X}_{i_m}}{h}\right) = 0. \end{cases}$$

It is clear that $\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h)$ can be rewritten, for all $\varphi \in \mathscr{F}$, as

$$\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h) = \frac{\sum_{\mathbf{i} \in I_n^m} \varphi(\mathbf{Y_i}) \widetilde{K}_h(\mathbf{t} - \mathbf{X_i})}{\sum_{\mathbf{i} \in I_n^m} \widetilde{K}_h(\mathbf{t} - \mathbf{X_i})} = \frac{U_n(\varphi, \mathbf{t}; h)}{U_n(1, \mathbf{t}; h)},$$

where we denote by $U_n(1, \mathbf{t}; h)$ the U-statistic $U_n(\varphi, \mathbf{t}; h)$ with $\varphi \equiv 1$. To prove the uniform consistency of $\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h)$ to $r^{(m)}(\varphi, \mathbf{t})$, we shall consider another, more appropriate, centering factor than the expectation $\mathbb{E}\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h)$, which may not exist or may be difficult to compute. Define the centring

$$\widehat{\mathbb{E}}\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h) := \frac{\mathbb{E}U_n(\varphi, \mathbf{t}; h)}{\mathbb{E}U_n(1, \mathbf{t}; h)}. \tag{2.7}$$

This centering permits us to derive results on the convergence rates of the process

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

to zero and the consistency of $\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h)$ uniform in $\mathbf{t}$ and in bandwidth.

THEOREM 2.5. *Let $l_n = c(\log n/n)^{1/dm}$ for $c > 0$. If the class of functions $\mathscr{F}_q$ is bounded, in the sense that for some $0 < M < \infty$,*

$$F(\mathbf{y}) < M, \quad for \ \ \mathbf{y} \in \mathbb{R}^{qm}. \tag{2.8}$$

Fix $\varepsilon \in (0, d_{vol})$. Further, if $d_{\mathrm{vol}} = 0$ or under Assumption 3, ε can be 0. Suppose

$$\limsup_n \frac{\left(\log\left(\frac{1}{l_n}\right)\right)_+ + \log\left(\frac{2}{\delta}\right)}{n l_n^{d_{\mathrm{vol}}-\varepsilon}} < \infty.$$

Then we can infer, under the above mentioned assumptions on $\mathscr{F}_q$ and Assumptions 1, 2 and 4, that for all $\delta > 0$, there exists a constant $0 < \mathfrak{C}_1 < \infty$ such that we have with probability at least $1 - \delta$

$$\sup_{h \geq l_n} \sup_{g \in \mathscr{F}_q} \sup_{\mathbf{t} \in \mathbb{R}^{dm}} \frac{\sqrt{n l_n^{m(2d-d_{\mathrm{vol}}+\varepsilon)}} |U_n(g, \mathbf{t}; h) - \widehat{\mathbb{E}} U_n(g, \mathbf{t}; h)|}{\sqrt{|\log l_n| \vee \log(2/\delta)}} \leq \mathfrak{C}_1. \tag{2.9}$$

THEOREM 2.6. *Let $a_n = c((\log n/n)^{1-2/p})^{1/dm}$ for $c > 0$. If $\mathscr{F}_q$ is unbounded, but satisfies, for some $p > 2$,*

$$\mu_p := \sup_{\mathbf{x} \in \mathbb{R}^{dm}} \mathbb{E}(F^p(\mathbf{Y}) \mid \mathbf{X} = \mathbf{x}) < \infty, \tag{2.10}$$

then we can infer, under the above mentioned assumptions on $\mathscr{F}_q$ and and Assumptions 1, 2 and 4, that for all $c > 0$ and $0 < b_0 < 1$, there exists a constant $0 < \mathfrak{C}_2 < \infty$ such that

$$\limsup_{n \to \infty} \sup_{a_n \leq h \leq b_0} \sup_{g \in \mathscr{F}_q} \sup_{\mathbf{t} \in \mathbb{R}^{dm}} \frac{\sqrt{n h^{m(2d-d_{\mathrm{vol}}+\varepsilon)}} |U_n(g, \mathbf{t}; h) - \widehat{\mathbb{E}} U_n(g, \mathbf{t}; h)|}{\sqrt{|\log h| \vee \log \log n}} \leq \mathfrak{C}_2,$$

for any $\varepsilon \in (0, d_{vol})$.

We mention that Kim et al. (2018) do not need continuity of the density for their results. (Of course, continuity of the density is crucial for controlling the bias.) Some related results on uniform convergence over compact subsets have been obtained by Bouzebda and El-hadjali (2020) for a much larger class of estimators including kernel estimators for regression functions among others. In this general setting, however, it is often not possible to obtain the convergence uniformly over $\mathbb{R}^d$. Density estimators are in that sense somewhat exceptional.

THEOREM 2.7. *Besides being bounded, suppose that the marginal density function $f_{\mathbf{X}}$ of $\mathbf{X}$ is continuous and strictly positive on the interval $\mathbf{I}$. Assume that the class of functions $\mathcal{M}$ is uniformly equicontinuous. It then follows that for all sequences $0 < b_n < 1$ with $b_n \to 0$,*

$$\sup_{0 < h \leq b_n} \sup_{\varphi \in \mathscr{F}_q} \sup_{\mathbf{t} \in \mathbf{I}^m} \left| \widehat{\mathbb{E}} \widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h) - r^{(m)}(\varphi, \mathbf{t}) \right| = o(1),$$

where $\mathbf{I}^m = \mathbf{I} \times \ldots \times \mathbf{I}$.

COROLLARY 2.8. *Besides being bounded, suppose that the marginal density function $f_{\mathbf{X}}$ of $\mathbf{X}$ is continuous and strictly positive on the interval $\mathbf{I}$. It then follows, under the above mentioned assumptions on $\mathscr{F}_q$ and and Assumptions 1, 2 and 4, that for all $c > 0$ and all sequences $0 < b_n < 1$ with $a_n' \leq b_n \to 0$, there exists a constant $0 < \mathfrak{C}_3 < \infty$ such that*

$$\limsup_{n \to \infty} \sup_{a_n' \leq h \leq b_0} \sup_{\varphi \in \mathscr{F}_q} \sup_{\mathbf{t} \in \mathbf{I}^m} \frac{\sqrt{n l_n^{m(2d - d_{\mathrm{vol}} + \varepsilon)}} |\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h) - \widehat{\mathbb{E}} \widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h)|}{\sqrt{|\log h| \vee \log \log n}} \leq \mathfrak{C}_3,$$

where for any $\varepsilon \in (0, d_{vol})$ and a_n' is either a_n or l_n, depending on whether the class $\mathscr{F}_q$ is bounded or not.

We can now state the main result of this section which follows easily from Theorems 2.5 and 2.6.

COROLLARY 2.9. *Under the conditions of Theorems 2.5 and 2.6 on f_X and the class of functions $\mathscr{F}_q$ and Assumptions 1, 2 and 4, it follows that for all sequences $0 < a_n' \leq \widetilde{a}_n \leq b_n < 1$ satisfying $b_n \to 0$ and $n\widetilde{a}_n / \log n \to \infty$,*

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

REMARK 2.10. Under additional, weak regularity conditions on $\mathbb{P}$, the value of ε can be taken equal to 0 (2.9). Under the assumption that the distribution has a bounded Lebesgue density, $d_{vol} = d$ so our result recovers existing results in literature in terms of rates of convergence, in particular the results presented in Dony and Mason (2008). Our results complement those Dony and Mason (2008) by relaxing the condition on the kernel functions as it was done by Kim et al. (2018). At this point, we mention that our results are stated in the multivariate setting and more importantly are adaptive to the dimension volume. This alleviates the problem of the curse of dimension. To be more precise, it is well known that the estimation problems of a regression function are especially hard in the case when the dimension of the explanatory $\mathbf{X}$ is large. It is worth noticing that one consequence of

this is that the optimal minimax rate of convergence $n^{-2k/(2k+d)}$ for the estimation of a k times differentiable regression function converges to zero rather slowly if the dimension d of $\mathbf{X}$ is large compared to k. To circumvent the so-called curse of dimensionality, the only way is to impose additional assumptions on the regression functions. The simplest way is to consider the linear models but this rather restrictive parametric assumption can be extended in several ways. An idea is to consider the additive models to simplify the problem of regression estimation by fitting only functions to the data which have the same additive structure. In projection pursuit one generalizes this further by assuming that the regression function is a sum of univariate functions applied to projections of $\mathbf{x}$ in various directions, we note that this includes the single index models as particular cases, the interested reader may refer to Györfi et al. (2002, Chapter 22) for more rigorous developments of such techniques. Other ways to be investigated are the semi-parametric models, considered intermediary models between linear and nonparametric ones, aiming to combine the flexibility of nonparametric approaches with the interpretability of the parametric ones, for details on these methods for functional data, one can refer to Ling and Vieu (2018, Section 4.2) and the reference therein.

REMARK 2.11. We note that the main problem in using an estimator such as in Eq. 1.2 is to choose properly the smoothing parameter h. The uniform in bandwidth consistency results given in Corollary 2.9 shows that any choice of h between a_n and b_n ensures the consistency of $\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h)$. Namely, the fluctuation of the bandwidth in a small interval does not affect the consistency of the nonparametric estimator $\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h)$ of $r^{(m)}(\varphi, \mathbf{t})$.

REMARK 2.12. For notational convenience, we have chosen the same bandwidth sequence for each margins. This assumption can be dropped easily. If one wants to make use of the vector bandwidths (see, in particular, Chapter 12 of Devroye and Lugosi (2001)). With obvious changes of notation, our results and their proofs remain true when h is replaced by a vector bandwidth $\mathbf{h}_n = (h_n^{(1)}, \ldots, h_n^{(d)})$, where $\min h_n^{(i)} > 0$. In this situation we set $h = \prod_{i=1}^{d} h^{(i)}$, and for any vector $\mathbf{v} = (v_1, \ldots, v_d)$ we replace $\mathbf{v}/h$ by $(v_1/h^{(1)}, \ldots, v_1/h^{(d)})$. For ease of presentation, we chose to use real-valued bandwidths throughout.

REMARK 2.13. In the sequel, we will need to symmetrize the functions $G_{g,h,\mathbf{t}}(\mathbf{x}, \mathbf{y})$. To do this, we have

$$\bar{G}_{g,h,\mathbf{t}}(\mathbf{x}, \mathbf{y}) := \frac{1}{m!} \sum_{\sigma \in \mathbf{I}_m^m} G_{g,h,\mathbf{t}}(\mathbf{x}_\sigma, \mathbf{y}_\sigma) = \frac{1}{m!} \sum_{\sigma \in \mathbf{I}_m^m} g(\mathbf{y}_\sigma) \widetilde{\mathbf{K}}_h(\mathbf{t} - \mathbf{x}_\sigma),$$

where $\mathbf{x}_\sigma := (x_{\sigma_1}, \ldots, x_{\sigma_m})$ et $\mathbf{y}_\sigma := (y_{\sigma_1}, \ldots, y_{\sigma_m})$. Obviously, after symmetrization we have

$$\mathbb{E}(G_{g,h,\mathbf{t}}(\mathbf{x}, \mathbf{y})) = \mathbb{E}(\bar{G}_{g,h,\mathbf{t}}(\mathbf{x}, \mathbf{y})) \quad \text{and} \quad U_n^{(m)}(\bar{G}_{g,h,\mathbf{t}}(\cdot, \cdot)) = U_n(g, \mathbf{t}; h)$$

So the U-statistic process in Eq. 2.2 may be redefined using the symmetrized kernels, hence we consider

$$u_n^{(m)}(g, h, \mathbf{t}) := \sqrt{n}\{U_n^{(m)}(\bar{G}_{g,h,\mathbf{t}}) - \mathbb{E}U_n^{(m)}(\bar{G}_{g,h,\mathbf{t}})\}. \tag{2.11}$$

For more details, consult for instance the book of de la Peña and Giné (1999).

3 Extension to the Censored Case

Consider a triple $(Y, C, \mathbf{X})$ of random variables defined in $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^d$. Here Y is the variable of interest, C is a censoring variable and $\mathbf{X}$ is a concomitant variable. Throughout, we will use Maillot and Viallon (2009) 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

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

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

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

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

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

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

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

the upper point of the corresponding distribution. Now consider a pointwise measurable class $\mathscr{F}$ 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 have recently gained popularity in the censored data literature (see Kohler et al. (2002), Carbonez et al. (1995), Brunel and Comte (2006)). 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

$$\Phi_\psi(y, c) = \frac{\mathbb{1}\{y \le c\}\psi(y \wedge c)}{1 - G(y \wedge c)}. \tag{3.1}$$

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

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

We have

$$
\begin{aligned}
r^{(1)}(\Phi_\psi, \mathbf{x}) &:= \mathbb{E}(\Phi_\psi(Y, C) \mid \mathbf{X} = \mathbf{x}) \\
&= \mathbb{E}\left\{ \frac{\mathbb{1}\{Y \le C\}\psi(Z)}{1 - G(Z)} \mid \mathbf{X} = \mathbf{x} \right\} \\
&= \mathbb{E}\left\{ \frac{\psi(Y)}{1 - G(Y)} \mathbb{E}(\mathbb{1}\{Y \le C\} \mid \mathbf{X}, Y) \mid \mathbf{X} = \mathbf{x} \right\} \\
&= r^{(1)}(\psi, \mathbf{x}).
\end{aligned} \tag{3.2}
$$

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

$$\overline{\omega}_{n,K,h,i}^{(1)}(\mathbf{x}) := K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h} \right) \Big/ \sum_{j=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{X}_j}{h} \right). \tag{3.3}$$

We assume that h satisfies **(H.1)**. In view of Eqs. 3.1, 3.2, and 3.3, whenever $G(\cdot)$ is known, a kernel estimator of $r^{(1)}(\psi, \cdot)$ is given by

$$\check{r}_n^{(1)}(\psi, \mathbf{x}; h) = \sum_{i=1}^n \overline{\omega}_{n,K,h,i}^{(1)}(\mathbf{x}) \frac{\delta_i \psi(Z_i)}{1 - G(Z_i)}. \tag{3.4}$$

The function $G(\cdot)$ is generally unknown and has to be estimated. We will denote by $G_n^*(\cdot)$ the Kaplan-Meier estimator of the function $G(\cdot)$ (Kaplan and Meier, 1958). Namely, adopting the conventions

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

and $0^0 = 1$ and setting

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

we have

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

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

$$\check{r}_n^{(1)*}(\psi, \mathbf{x}; h) = \sum_{i=1}^n \overline{\omega}_{n,K,h,i}^{(1)}(\mathbf{x}) \frac{\delta_i \psi(Z_i)}{1 - G_n^*(Z_i)}, \tag{3.5}$$

refer to Kohler et al. (2002) and Maillot and Viallon (2009). Adopting the convention $0/0 = 0$, this quantity is well defined, since $G_n^*(Z_i) = 1$ if and only if $Z_i = Z_{(n)}$ and $\delta_{(n)} = 0$, where $Z_{(k)}$ is the kth ordered statistic associated with the sample $(Z_1, \ldots, Z_n)$ for $k = 1, \ldots, n$ and $\delta_{(k)}$ is the δ_j corresponding to $Z_k = Z_j$. When the variable of interest is right-censored, functional of the (conditional) law can generally not be estimated on the complete support (see Brunel and Comte 2006). To obtain our results, we will work under the following assumptions.

(A.1) $\mathscr{F} = \{\psi := \psi_1 \mathbb{1}\{(-\infty, \tau)^m\}, \psi_1 \in \mathscr{F}_1\}$, where $\tau < T_H$ and $\mathscr{F}_1$ is a pointwise measurable class of real measurable functions defined on $\mathbb{R}$ and of type VC.

(A.2) The class of functions $\mathscr{F}$ has a measurable and uniformly bounded envelope function Υ with,

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

(A.3) The class of functions $\mathcal{M}$ is relatively compact concerning the sup-norm topology on $\mathcal{I}_\alpha^m$.

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

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

This choice is justified by the fact that, under hypothesis $(\mathscr{I})$ we have

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

$$(3.6)$$

Let us assume the following conditions.

(H.1) $h \downarrow 0, 0 < h < 1$, and $nh^d \uparrow \infty$;

(H.2) $nh^d/\log n \to \infty$ as $n \to \infty$;

(H.3) $\log(1/h)/\log\log n \to \infty$ as $n \to \infty$;

We now have all the ingredients to state the result corresponding to the censored case. Let $\{h_n'\}_{n\geq 1}$ and $\{h_n''\}_{n\geq 1}$ be two sequences of positive constants fulfilling Assumptions (**H.1-H.3**) with

$$0 < h_n' \leq h_n'' < 1.$$

Bouzebda and El-hadjali (2020) showed under assumptions **(A.1–3)**, for $m = 1$, **(I)**, Assumption 3, for any kernel $K(\cdot)$ satisfying Assumptions 1 and 2, with probability at least $1 - \delta$,

$$\sup_{h'_n \leq h \leq h''_n} \sup_{\mathbf{x} \in \mathcal{I}} \left| \breve{r}_n^{(1)*}(\psi, \mathbf{x}; h) - \widehat{\mathbb{E}}(\breve{r}_n^{(1)*}(\psi, \mathbf{x}; h)) \right| \leq \mathfrak{C}_4 \sqrt{\frac{\log(1/l_n)_+ + \log(2/\delta)}{n l_n^{2d - d_{\text{vol}} + \varepsilon}}},$$
(3.7)

for some positive constant $\mathfrak{C}_4$. A right-censored version of an unconditional U-statistic with a kernel of degree $m \geq 1$ is introduced by the principle of a mean preserving reweighting scheme in Datta et al. (2010). Stute and Wang (1993) have proved almost sure convergence of multi-sample U-statistics under random censorship and provided application by considering the consistency of a new class of tests designed for testing equality in distribution. To overcome potential biases arising from right-censoring of the outcomes and the presence of confounding covariates, Chen and Datta (2019) proposed adjustments to the classical U-statistics. Yuan et al. (2017) proposed a different way in the estimation procedure of the U-statistic by using a substitution estimator of the conditional kernel given the observed data. To our best knowledge, the problem of the estimation of the conditional U-statistics was opened up to the present, and it gives and main motivation to the study of this section. A natural extension of the function defined in Eq. 3.1 is given by

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

From this, we have an analogous relation to Eq. 3.2 given by

$$\mathbb{E}(\mathbf{\Phi}_\psi(Y_1, \ldots, Y_m, C_1, \ldots, C_m) \mid (\mathbf{X}_1, \ldots, \mathbf{X}_m) = \mathbf{t})$$

$$= \mathbb{E}\left(\frac{\prod_{i=1}^m \{\mathbb{1}\{Y_i \leq C_i\} \psi(Y_1 \wedge C_1, \ldots, Y_m \wedge C_m)}{\prod_{i=1}^m \{1 - G(Y_i \wedge C_i)\}} \mid (\mathbf{X}_1, \ldots, \mathbf{X}_m) = \mathbf{t} \right)$$

$$= \mathbb{E}\left(\frac{\psi(Y_1, \ldots, Y_m)}{\prod_{i=1}^m \{1 - G(Y_i)\}} \mathbb{E}\left(\prod_{i=1}^m \{\mathbb{1}\{Y_i \leq C_i\} \mid (Y_1, X_1), \ldots (Y_m, X_m) \right) \mid$$

$$(\mathbf{X}_1, \ldots, \mathbf{X}_m) = \mathbf{t} \right)$$

$$= \mathbb{E}(\psi(Y_1, \ldots, Y_m) \mid (\mathbf{X}_1, \ldots, \mathbf{X}_m) = \mathbf{t}) = m_\psi(\mathbf{t}).$$

An analogue estimator to Eq. 1.2 in the censored case is given by

$$\breve{r}_n^{(m)}(\psi, \mathbf{t}; h) = \sum_{(i_1, \ldots, i_m) \in I(m,n)} \frac{\delta_{i_1} \cdots \delta_{i_m} \psi(Z_{i_1}, \ldots, Z_{i_m})}{(1 - G(Z_{i_1})) \cdots (1 - G(Z_{i_m}))} \overline{\omega}_{n,K,h,\mathbf{i}}^{(m)}(\mathbf{t}), \quad (3.8)$$

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

$$\overline{\omega}_{n,K,h,\mathbf{i}}^{(m)}(\mathbf{t}) := \frac{K\left(\dfrac{\mathbf{t}_1 - \mathbf{X}_{i_1}}{h}\right) \cdots K\left(\dfrac{\mathbf{t}_m - \mathbf{X}_{i_m}}{h}\right)}{\displaystyle\sum_{(i_1,\ldots,i_m)\in I(m,n)} K\left(\dfrac{\mathbf{t}_1 - \mathbf{X}_{i_1}}{h}\right) \cdots K\left(\dfrac{\mathbf{t}_m - \mathbf{X}_{i_m}}{h}\right)}. \quad (3.9)$$

The estimator that we will investigate is given by

$$\breve{r}_n^{(m)*}(\psi, \mathbf{t}; h) = \sum_{(i_1,\ldots,i_m)\in I(m,n)} \frac{\delta_{i_1} \cdots \delta_{i_m} \psi(Z_{i_1}, \ldots, Z_{i_m})}{(1 - G_n^*(Z_{i_1})) \cdots (1 - G_n^*(Z_{i_m}))} \overline{\omega}_{n,K,h,\mathbf{i}}^{(m)}(\mathbf{t}). \tag{3.10}$$

THEOREM 3.1. *Let $a_n = c((\log n / n)^{1 - 2/p})^{1/dm}$ for $c > 0$. If the class of functions $\mathscr{F}$ is bounded, in the sense (**A.2**). Fix $\varepsilon \in (0, d_{vol})$. Further, if $d_{\mathrm{vol}} = 0$ or under Assumption 3, ε can be 0. Suppose that*

$$\limsup_n \frac{\left(\log\left(\dfrac{1}{l_n}\right)\right)_+ + \log\left(\dfrac{2}{\delta}\right)}{n l_n^{d_{\mathrm{vol}} - \varepsilon}} < \infty.$$

*Then we can infer, under the above mentioned assumptions on $\mathscr{K}$, (**A.1-2**) (**$\mathscr{I}$**), that for all $c > 0$ and $0 < b_0 < 1$, there exists a constant $0 < \mathfrak{C}_5 < \infty$ such that*

$$\sup_{h \geq l_n} \sup_{\varphi \in \mathscr{F}} \sup_{\mathbf{t} \in \mathcal{I}^m} \frac{\sqrt{n l_n^{(2d - d_{\mathrm{vol}} + \varepsilon)m}} |\breve{r}_n^{(m)*}(\psi, \mathbf{t}; h) - \widehat{\mathbb{E}}\breve{r}_n^{(m)*}(\psi, \mathbf{t}; h)|}{\sqrt{|\log l_n| \vee \log(2/\delta)}} \leq \mathfrak{C}_5.$$

PROPOSITION 3.2. *Under assumptions (**A.1-3**), (**$\mathscr{I}$**) and for any kernel $K(\cdot)$ satisfying Assumptions 1, 2 and 3. Let $\{h'_n\}_{n \geq 1}$ and $\{h''_n\}_{n \geq 1}$ be two sequences of positive constants fulfilling Assumptions (**H.1-H.3**) with*

$$0 < h'_n \leq h''_n < 1.$$

With probability at least $1 - \delta$, there exists a constant $0 < \mathfrak{C}_6 < \infty$ such that

$$\sup_{h'_n \leq h \leq h''_n} \sup_{\mathbf{x} \in \mathcal{I}^m} \sup_{\psi \in \mathscr{F}} \left|\breve{r}_n^{(m)*}(\psi, \mathbf{t}; h) - \widehat{\mathbb{E}}\breve{r}_n^{(m)*}(\psi, \mathbf{t}; h)\right|$$

$$\leq \mathfrak{C}_6 \sqrt{\frac{\log(1/l_n)_+ + \log(2/\delta)}{n l_n^{(2d - d_{\mathrm{vol}} + \varepsilon)m}}}. \tag{3.11}$$

3.1. Examples of U-statistics

EXAMPLE 3.3. Let $\widehat{Y_1Y_2}$ denote the oriented angle between $Y_1, Y_2 \in T$, T is the circle of radius 1 and center 0 in $\mathbb{R}^2$. Let:

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

Silverman (1978) has used this kernel to propose a U-process to test uniformity on the circle.

EXAMPLE 3.4. Hoeffding (1948) introduced the parameter

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

where $D(y_1, y_2) = F(y_1, y_2) - F(y_1, \infty)F(\infty, y_2)$ and $F(\cdot, \cdot)$ is the distribution function of Y_1 and Y_2. The parameter $\triangle$ has the property that $\triangle = 0$ if and only if Y_1 and Y_2 are independent. From Lee (1990), an alternative expression for $\triangle$ can be developed by introducing the functions

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

and

$$\varphi(y_{1,1}, y_{1,2}, \ldots, y_{5,1}, y_{5,2}) = \frac{1}{4}\psi(y_{1,1}, y_{1,2}, y_{1,3})\,\psi(y_{1,1}, y_{1,4}, y_{1,5})\\ \psi(y_{1,2}, y_{2,2}, y_{3,2})\,\psi(y_{1,2}, y_{4,2}, y_{5,2}).$$

We have

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

We have

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

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

EXAMPLE 3.5. For $m = 3$, let $\varphi(Y_1, Y_2, Y_3) = \mathbb{1}\{Y_1 - Y_2 - Y_3 > 0\}$, the corresponding U-Statistic corresponds to the Hollander-Proschan test-statistic (Hollander and Proschan, 1972).

EXAMPLE 3.6. For:

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

we obtain the variance of Y.

EXAMPLE 3.7. For:

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

we obtain:

$$r^{(2)}(\varphi(Y_1, Y_2) \mid t_1, t_2) = \text{Var}(Y_1 \mid X_1 = t_1).$$

EXAMPLE 3.8. Let $\mathbf{Y} = (Y_1, Y_2)$ such that Y_2 is a smooth curve, $Y_2 \in L_2([0,1])$ and $Y_1 \in \mathbb{R}$ has a continuous distribution. For

$$\varphi(\mathbf{Y}_1, \mathbf{Y}_2) = \mathbb{1}\{(Y_{1,1} - Y_{2,1})(Y_{1,2}(t) - Y_{2,2}(t)) > 0\} - 0.5,$$

which can be used to treat the problem of testing for conditional association between a functional variable belonging to Hilbert space and a scalar variable. More precisely, this gives the conditional Kendall's Tau type statistics.

4   The Bandwidth Selection Criterion

Many methods have been established and developed to construct, in asymptotically optimal ways, bandwidth selection rules for nonparametric kernel estimators especially for Nadaraya-Watson regression estimator we quote among them Hall (1984), Härdle and Marron (1985). This parameter has to be selected suitably, either in the standard finite-dimensional case, or in the infinite dimensional framework for ensuring good practical performances. Following Dony and Mason (2008), the leave-one-out cross-validation procedure allows to define, for any fixed $\mathbf{i} = (i_1, \dots, i_m) \in I(m, n)$:

$$\widehat{r}_{n,\mathbf{i}}^{(m)}(\varphi, \mathbf{t}; h)$$

$$= \frac{\displaystyle\sum_{(j_1,\dots,j_m)\in I(m,n)(\mathbf{i})} \varphi(Y_{j_1}, \dots, Y_{j_m}) K\left(\frac{\mathbf{t}_1 - \mathbf{X}_{j_1}}{h}\right) \cdots K\left(\frac{\mathbf{t}_m - \mathbf{X}_{j_m}}{h}\right)}{\displaystyle\sum_{(i_1,\dots,i_m)\in I(m,n)} K\left(\frac{\mathbf{x}_1 - \mathbf{X}_{j_1}}{h}\right) \cdots K\left(\frac{\mathbf{x}_m - \mathbf{X}_{j_m}}{h}\right)}, \quad (4.1)$$

where

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

In order to minimize the quadratic loss function, we introduce the following criterion, we have for some (known) non-negative weight function $\mathcal{W}(\cdot)$:

$$CV\left(\varphi, h\right) := \frac{(n-k)!}{n!} \sum_{\mathbf{i} \in I(m,n)} \left(\varphi\left(\mathbf{Y_i}\right) - \widehat{r}_{n,\mathbf{i}}^{(m)}(\varphi, \mathbf{X_i}; h)\right)^2 \widetilde{\mathcal{W}}\left(\mathbf{X_i}\right), \quad (4.2)$$

where

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

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

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

we can conclude, by Corollary 2.9, that :

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

The main interest of our results is the possibility to derive the asymptotic properties of our estimate even if the bandwidth parameter is a random variable, like in the last equation. One can replace (4.2) by

$$CV\left(\varphi, h\right) := \frac{(n-k)!}{n!} \sum_{\mathbf{i} \in I(m,n)} \left(\varphi\left(\mathbf{Y_i}\right) - \widehat{r}_{n,\mathbf{i}}^{(m)}(\varphi, \mathbf{X_i}; h)\right)^2 \widehat{\mathcal{W}}\left(\mathbf{X_i}, \mathbf{t}\right), \quad (4.3)$$

where

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

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

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

For sake of brevity, we have just considered the most popular method, that is, the cross-validated selected bandwidth. This may be extended to any other bandwidth selector such as the bandwidth based on Bayesian ideas (Shang, 2014).

4.1. Discrimination Now, we apply the results to the problem of discrimination described in Section 3 of Stute (1994b), refer to also to Stute (1994a). We will use a similar notation and setting. Let $\varphi(\cdot)$ be any function taking at most finitely many values, say $1, \ldots, M$. The sets

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

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

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

where

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

The above inequality becomes equality if

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

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

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

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

$$m_n^j(\mathbf{x}) = \frac{\displaystyle\sum_{(i_1,\ldots,i_m) \in I(m,n)} \mathbb{1}\{\varphi(Y_{i_1}, \ldots, Y_{i_m})=j\} K\left(\frac{\mathbf{t}_1-\mathbf{X}_{i_1}}{h}\right) \cdots K\left(\frac{\mathbf{t}_m-\mathbf{X}_{i_m}}{h}\right)}{\displaystyle\sum_{(i_1,\ldots,i_m) \in I(m,n)} K\left(\frac{\mathbf{t}_1-\mathbf{X}_{i_1}}{h}\right) \cdots K\left(\frac{\mathbf{t}_m-\mathbf{X}_{i_m}}{h}\right)}.$$

$$\tag{4.4}$$

Set

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

Let us introduce

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

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

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

5 Concluding Remarks and Future Works

In the present work, we have used general methods based upon empirical process techniques to prove uniform in bandwidth consistency for kernel-type estimators of the conditional U-statistics. We have considered an extended setting when the dimension can be lower than the ambient dimension. In addition, our work complements the paper Kim et al. (2018) by considering other examples of kernel estimates. Our proof relies on the work of Dony and Mason (2008). Our results extend and complement the last cited reference by establishing the convergence rate adaptive to the volume dimension. Our results are especially useful to establish uniform consistency of data-driven bandwidth kernel-type function estimators. The interest in doing so would be to extend our work to k-nearest neighbours estimators. Presently it is beyond reasonable hope to achieve this program without new technical arguments. We will not treat the uniform consistency of such estimators in the present paper, and leave this for future investigation.

6 Mathematical Developments

This section is devoted to the proof of our results. The previously defined notation continues to be used in what follows.

Our main tool to analyze $u_n^{(m)}(g, h, \mathbf{t})$ will be the Hoeffding decomposition, which we recall here for the reader's convenience.

6.1. Hoeffding Decomposition The Hoeffding decomposition, Hoeffding (1948), states the following, which is easy to check,

$$u_n^{(m)}(g, h, \mathbf{t}) = \sqrt{n} \sum_{k=1}^{m} \binom{m}{k} U_n^{(k)}(\pi_k \overline{G}_{g,h,\mathbf{t}}(\cdot, \cdot)), \qquad (6.1)$$

where the kth Hoeffding projection for a (symmetric) function $L : S^m \times S^m \to \mathbb{R}$ with respect to $\mathbb{P}$ is defined for $\mathbf{x}_k = (x_1, \ldots, x_k) \in S^k$ and $\mathbf{y}_k = (y_1, \ldots, y_k) \in S^k$ as

$$\pi_k L(\mathbf{x}_k, \mathbf{y}_k) := (\delta_{(x_1, y_1)} - \mathbb{P}) \times \ldots \times (\delta_{(x_k, y_k)} - \mathbb{P}) \times \mathbb{P}^{m-k}(L),$$

where $\mathbb{P}$ is any probability measure on $(S, \mathcal{S})$ and for measures $\mathbb{Q}_i$ on S we have

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

Considering $(X_i, Y_i), i \geq 1$, i.i.d.-$\mathbb{P}$ and assuming $L \in L_2(\mathbb{P}^m)$, this is an orthogonal decomposition and

$$\mathbb{E}[\pi_k L(\mathbf{X}_k, \mathbf{Y}_k) \mid (X_2, Y_2)), \ldots, (X_k, Y_k)] = 0, k \geq 1,$$

where we denote $\mathbf{X}_k$ and $\mathbf{Y}_k$ for $(X_1, \ldots, X_k)$ and $(Y_1, \ldots, Y_k)$, respectively. Thus the kernels $\pi_k L$ are canonical for $\mathbb{P}$. Also, $\pi_k, k \geq 1$, are nested projections, that is, $\pi_k \circ \pi_l = \pi_k$ if $k \leq l$, and

$$\mathbb{E}[(\pi_k L)^2(\mathbf{X}_k, \mathbf{Y}_k)] \leq \mathbb{E}[(L - \mathbb{E}L)^2(\mathbf{X}, \mathbf{Y})] \leq \mathbb{E}L^2(\mathbf{X}, \mathbf{Y}). \quad (6.2)$$

For more details, consult de la Peña and Giné (1999). The proofs of our results are largely inspired from Dony and Mason (2008), Kim et al. (2018) and Bouzebda and El-hadjali (2020).

6.2. *Proof of Theorem 2.5: the Bounded Case*

6.3. *Linear Term* To establish the relation (2.9), we need to study the linear term (the first term) of Eq. 6.1, given by

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

Keeping in mind the fact that the class $\mathscr{F}_q$ is a VC-type class of functions with an envelope function F and the class $\mathscr{K}$ is a VC-type with envelope κ, which implies that the class of functions on $\mathbb{R}^{qm} \times \mathbb{R}^{dm}$ given by $\{h^{dm} G_{g,h,\mathbf{t}}(\cdot, \cdot) : g \in \mathscr{F}_q, \mathbf{t} \in \mathbb{R}^{dm}\}$ is of VC-type (via Lemma A.1 in Einmahl and Mason (2000)), as well as the class

$$\mathcal{G} = \{h^{dm} \overline{G}_{g,h,\mathbf{t}}(\cdot, \cdot) : g \in \mathscr{F}_q, h \geq l_n, \mathbf{t} \in \mathbb{R}^{dm}\}, \quad (6.3)$$

for which we denote the VC-type characteristics by A and v, and the envelope function by

$$\widetilde{F}(\mathbf{y}) \equiv \widetilde{F}(\mathbf{x}, \mathbf{y}) = \kappa^m \sum_{\sigma \in I_m^m} F(\mathbf{y}_\sigma), \quad \mathbf{y} \in \mathbb{R}^{qm}. \quad (6.4)$$

By considering the following class of functions on $\mathbb{R}^{dk} \times \mathbb{R}^{qk}$, for $k = 1, \ldots, m$,

$$\mathcal{G}^{(k)} = \{h^{dm} \pi_k \overline{G}_{g,h,\mathbf{t}}(\cdot, \cdot) : g \in \mathscr{F}_q, h \geq l_n, \mathbf{t} \in \mathbb{R}^{dm}\}, \quad (6.5)$$

and following Giné and Mason (2007a) one can show that each class $\mathcal{G}^{(k)}$ is of VC-type with characteristics A and v and envelope function

$$\mathbf{F}_k \leq 2^k \|\widetilde{\mathbf{F}}\|_\infty. \quad (6.6)$$

Recall that the sample $(X_i, Y_i), 1 \leq i \leq n$ is i.i.d. and from the definition of the Hoeffding projections, for all $(x, y) \in \mathbb{R}^d \times \mathbb{R}^q$, we get

$$\pi_1 \overline{G}_{g,h,\mathbf{t}}(x, y) = \mathbb{E}\left(\overline{G}_{g,h,\mathbf{t}}((x, X_2, \ldots, X_m), (y, Y_2, \ldots, Y_m))\right)$$

$$-\mathbb{E}\left(\overline{G}_{g,h,\mathbf{t}}(\mathbf{X},\mathbf{Y})\right)$$
$$= \mathbb{E}\left(\overline{G}_{g,h,\mathbf{t}}(\mathbf{X},\mathbf{Y}) \mid (X_1,Y_1)=(x,y)\right) - \mathbb{E}\left(\overline{G}_{g,h,\mathbf{t}}(\mathbf{X},\mathbf{Y})\right).$$

Introduce the following function on $\mathbb{R}^d \times \mathbb{R}^q$:

$$
\begin{aligned}
S_{g,h,\mathbf{t}} : \mathbb{R}^d \times \mathbb{R}^q &\rightarrow \mathbb{R} \\
(x,y) &\mapsto mh^{dm}\mathbb{E}\left(\overline{G}_{g,h,\mathbf{t}}(\mathbf{X},\mathbf{Y}) \mid (X_1,Y_1)=(x,y)\right).
\end{aligned}
$$

Making use of this notation, we can write

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

For all $g \in \mathscr{F}_q$, $h \geq l_n$ and $\mathbf{t} \in \mathbb{R}^{dm}$, the linear term of the decomposition in Eq. 6.1 times h^{dm} is given by

$$
\begin{aligned}
m\sqrt{n}h^{dm}U_n^{(1)}(\pi_1\overline{G}_{g,h,\mathbf{t}}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{S_{g,h,\mathbf{t}}(X_i,Y_i) - \mathbb{E}(S_{g,h,\mathbf{t}}(X_i,Y_i))\} \\
&=: \alpha_n(S_{g,h,\mathbf{t}}),
\end{aligned}
$$

where we recall that the last expression is the empirical process $\alpha_n(\cdot)$ based on the sample $(X_1,Y_1),\ldots,(X_n,Y_n)$ and we set for $\mathbf{t} \in \mathbf{I} \subset \mathbb{X}$, $g \in \mathscr{F}_q$ and $h \geq l_n$ the class of normalised functions on $\mathbb{R}^{dm} \times \mathbb{R}^{qm}$,

$$\mathcal{S}_n = \{S_{g,h,\mathbf{t}}(\cdot,\cdot) : g \in \mathscr{F}_q, h \geq l_n, \mathbf{t} \in \mathbf{I} \subset \mathbb{X}\}. \qquad (6.7)$$

Now, we have to bound $S_{g,h,\mathbf{t}}$. From Eq. 2.8, we get

$$|S_{g,h,\mathbf{t}}(x,y)| \leq mh^{dm}M\left\|K\right\|_\infty^m.$$

In order to bound the VC dimension of $\mathcal{S}_n$, we remark that $\mathcal{S}_n = m\mathcal{G}^{(1)}$ is VC-type with characteristics A and v as defined in Eq. 6.5 for $k = 1$. For reader convenience, we give more details. Let us give the bound for the VC dimension of $\mathcal{S}_n$. Fix $\eta < l_n^{-dm}\left\|S_{g,h,\mathbf{t}}\right\|_\infty$ and a probability measure $\mathbb{Q}$ on $\mathbb{R}^d$. Suppose

$$\left[l_n, \left(\frac{\eta}{2\left\|S_{g,h,\mathbf{t}}\right\|_\infty}\right)^{-1/dm}\right]$$

is covered by balls of the form

$$\left\{\left(h_i - \frac{\sqrt{n}\eta l_n^{dm+1}}{3dm\left\|S_{g,h,\mathbf{t}}\right\|_\infty}, h_i + \frac{\sqrt{n}\eta l_n^{dm+1}}{3dm\left\|S_{g,h,\mathbf{t}}\right\|_\infty}\right), 1 \leq i \leq N_1\right\},$$

and $(\mathcal{S}_n, \mathbb{L}_2(\mathbb{Q}))$ is covered by

$$\left\{ \mathbb{B}_{\mathbb{L}_2(\mathbb{Q})}\left(K_j, \frac{l_n^{dm}\eta}{3mM}\right) \cup \mathbb{B}_{\mathbb{L}_2(\mathbb{Q})}\left(g_k, \widetilde{\varepsilon}\right), 1 \le j \le N_2, 1 \le k \le N_3 \right\},$$

where

$$\widetilde{\varepsilon} \le \frac{l_n^{dm}\eta}{3mM \left\|K\right\|_\infty^m}.$$

For $1 \le i \le N_1, 1 \le j \le N_2$ and $1 \le k \le N_3$, we let

$$S_{i,j,k} = \frac{1}{\sqrt{n}h_i^{dm}} S_{j,k} = \frac{m}{h_i^{dm}} \mathbb{E}\left(g_k(\mathbf{Y})\widetilde{K}_j\left(\frac{\mathbf{t} - \mathbf{X}}{h}\right) \mid (X_1, Y_1) = (x, y)\right).$$

Also, choose $b_0 > \left(\dfrac{\sqrt{n}\eta}{2\left\|S_{g,h,\mathbf{t}}\right\|_\infty}\right)^{-1/dm}$, $\mathbf{t}_0 \in \mathbf{I}, g_0 \in \mathscr{F}_q$ and let

$$S_0 = \frac{1}{\sqrt{n}b_0} S_{g_0, b_0, \mathbf{t}_0}.$$

We will show that

$$\begin{aligned} &\left\{\mathbb{B}_{\mathbb{L}_2(\mathbb{Q})}\left(S_{i,j,k}, \eta\right) : 1 \le i \le N_1, 1 \le j \le N_2 \quad \text{and} \right.\\ &\left. 1 \le k \le N_3\right\} \cup \left\{\mathbb{B}_{\mathbb{L}_2(\mathbb{Q})}\left(S_0, \eta\right)\right\} \quad \text{covers } \mathcal{S}_n. \end{aligned} \tag{6.8}$$

For the first case when $h \le \left(\dfrac{\sqrt{n}\eta}{2\left\|S_{g,h,\mathbf{t}}\right\|_\infty}\right)^{-1/dm}$, find h_i, K_j and g_k with

$$\begin{aligned} h &\in \left(h_i - \frac{\sqrt{n}\eta l_n^{dm+1}}{3dm\left\|S_{g,h,\mathbf{t}}\right\|_\infty}, h_i + \frac{\sqrt{n}\eta l_n^{dm+1}}{3dm\left\|S_{g,h,\mathbf{t}}\right\|_\infty}\right),\\ g &\in \mathbb{B}_{\mathbb{L}_2(\mathbb{Q})}\left(g_k, \widetilde{\varepsilon}\right),\\ K &\in \mathbb{B}_{\mathbb{L}_2(\mathbb{Q})}\left(K_j, \frac{l_n^{dm}\eta}{3mM}\right). \end{aligned}$$

Then the distance between $\frac{1}{\sqrt{n}h^{dm}}S_{g,h,\mathbf{t}}$ and $\frac{1}{\sqrt{n}h_i^{dm}}S_{j,k}$ is upper bounded as follows

$$\left\|\frac{1}{\sqrt{n}h^{dm}}S_{g,h,\mathbf{t}} - \frac{1}{\sqrt{n}h_i^{dm}}S_{j,k}\right\|_{\mathbb{L}_2(\mathbb{Q})}$$

$$= \left\|\frac{1}{h^{dm}}m\mathbb{E}(\overline{G}_{g,h,\mathbf{t}}(\mathbf{X}, \mathbf{Y}) \mid (X_1, Y_1) = (x, y)) - \frac{1}{\sqrt{n}h_i^{dm}}S_{j,k}(X_1, Y_1)\right\|_{\mathbb{L}_2(\mathbb{Q})}$$

$$\leq \left\| \frac{m}{h^{dm}} \mathbb{E}\left(g(\mathbf{Y})\widetilde{K}\left(\frac{\mathbf{t}-\mathbf{X}}{h} \right) \mid (X_1, Y_1) = (x, y) \right) \right.$$

$$\left. - \frac{m}{h_i^{dm}} \mathbb{E}\left(g(\mathbf{Y})\widetilde{K}\left(\frac{\mathbf{t}-\mathbf{X}}{h} \right) \mid (X_1, Y_1) = (x, y) \right) \right\|_{\mathbb{L}_2(\mathbb{Q})}$$

$$+ \left\| \frac{m}{h_i^{dm}} \mathbb{E}\left(g(\mathbf{Y})\widetilde{K}\left(\frac{\mathbf{t}-\mathbf{X}}{h} \right) \mid (X_1, Y_1) = (x, y) \right) - \frac{1}{\sqrt{n}h_i^{dm}} S_{i,k} \right\|_{\mathbb{L}_2(\mathbb{Q})}$$

$$\leq \left\| \frac{1}{\sqrt{n}h^{dm}} S_{g,h,\mathbf{t}} - \frac{1}{\sqrt{n}h_i^{dm}} S_{g,h,\mathbf{t}} \right\|_{\mathbb{L}_2(\mathbb{Q})}$$

$$+ \left\| \frac{1}{\sqrt{n}h_i^{dm}} S_{g,h,\mathbf{t}} - \frac{m}{h_i^{dm}} \mathbb{E}\left(g(\mathbf{Y})\widetilde{K}_j\left(\frac{\mathbf{t}-\mathbf{X}}{h} \right) \mid (X_1, Y_1) = (x, y) \right) \right\|_{\mathbb{L}_2(\mathbb{Q})}$$

$$+ \left\| \frac{m}{h_i^{dm}} \mathbb{E}\left(g(\mathbf{Y})\widetilde{K}_j\left(\frac{\mathbf{t}-\mathbf{X}}{h} \right) \mid (X_1, Y_1) = (x, y) \right) - \right.$$

$$\left. \frac{m}{h_i^{dm}} \mathbb{E}\left(g_k(\mathbf{Y})\widetilde{K}_j\left(\frac{\mathbf{t}-\mathbf{X}}{h} \right) \mid (X_1, Y_1) = (x, y) \right) \right\|_{\mathbb{L}_2(\mathbb{Q})}. \tag{6.9}$$

Now the first term of Eq. 6.9 is upper bounded as

$$\left\| \frac{1}{\sqrt{n}h^{dm}} S_{g,h,\mathbf{t}} - \frac{1}{\sqrt{n}h_i^{dm}} S_{g,h,\mathbf{t}} \right\|_{\mathbb{L}_2(\mathbb{Q})}$$

$$= \left| \frac{1}{h^{dm}} - \frac{1}{h_i^{dm}} \right| \frac{\|S_{g,h,\mathbf{t}}\|_{\mathbb{L}_2(\mathbb{Q})}}{\sqrt{n}}$$

$$= |h_i - h| \sum_{k=0}^{dm-1} h_i^{k-dm} h^{-1-k} \frac{\|S_{g,h,\mathbf{t}}\|_{\mathbb{L}_2(\mathbb{Q})}}{\sqrt{n}}$$

$$\leq |h_i - h| dm l_n^{-dm-1} \frac{1}{\sqrt{n}} \|S_{g,h,\mathbf{t}}\|_\infty < \frac{\eta}{3}. \tag{6.10}$$

Also, the second term of Eq. 6.9 is upper bounded as

$$\left\| \frac{1}{\sqrt{n}h_i^{dm}} S_{g,h,\mathbf{t}} - \frac{m}{h_i^{dm}} \mathbb{E}\left(g(\mathbf{Y})\widetilde{K}_j\left(\frac{\mathbf{t}-\mathbf{X}}{h} \right) \mid (X_1, Y_1) = (x, y) \right) \right\|_{\mathbb{L}_2(\mathbb{Q})}$$

$$= \frac{m}{h_i^{dm}} \left\| \mathbb{E}\left(g(\mathbf{Y})\widetilde{K}\left(\frac{\mathbf{t}-\mathbf{X}}{h} \right) \mid (X_1, Y_1) = (x, y) \right) \right.$$

$$\left. - \mathbb{E}\left(g(\mathbf{Y})\widetilde{K}_j\left(\frac{\mathbf{t}-\mathbf{X}}{h} \right) \mid (X_1, Y_1) = (x, y) \right) \right\|_{\mathbb{L}_2(\mathbb{Q})}$$

$$\leq \frac{mM}{h_i^{dm}} \left\| \mathbb{E}\left(\widetilde{K}\left(\frac{\mathbf{t}-\mathbf{X}}{h} \right) \mid (X_1, Y_1) = (x, y) \right) \right.$$

$$-\mathbb{E}\left(\widetilde{K}_j\left(\frac{\mathbf{t}-\mathbf{X}}{h}\right)\mid (X_1,Y_1)=(x,y)\right)\Bigg\|_{\mathbb{L}_2(\mathbb{Q})}$$

$$\leq \frac{\eta}{3}. \tag{6.11}$$

The last term of Eq. 6.9 is upper bounded as

$$\left\|\frac{m}{h_i^{dm}}\mathbb{E}\left(g(\mathbf{Y})\widetilde{K}_j\left(\frac{\mathbf{t}-\mathbf{X}}{h}\right)\mid (X_1,Y_1)=(x,y)\right)\right.$$

$$\left.-\frac{m}{h_i^{dm}}\mathbb{E}\left(g_k(\mathbf{Y})\widetilde{K}_j\left(\frac{\mathbf{t}-\mathbf{X}}{h}\right)\mid (X_1,Y_1)=(x,y)\right)\right\|_{\mathbb{L}_2(\mathbb{Q})}$$

$$\leq \quad ml_n^{-dm}\|K\|_\infty^m\|g-g_k\|_{\mathbb{L}_2(\mathbb{Q})}$$

$$< \quad ml_n^{-dm}\|K\|_\infty^m\widetilde{\varepsilon}\leq \frac{\eta}{3}. \tag{6.12}$$

By combining the Eqs. 6.10, 6.11 and 6.12 to 6.9, we readily obtain the following bound

$$\left\|\frac{1}{\sqrt{n}h^{dm}}S_{g,h,\mathbf{t}}-\frac{1}{\sqrt{n}h_i^{dm}}S_{j,k}\right\|_{\mathbb{L}_2(\mathbb{Q})}<\eta.$$

For the second case when $h>\left(\frac{\sqrt{n}\eta}{2\|S_{g,h,\mathbf{t}}\|_\infty}\right)^{-1/dm}$, we have

$$\left\|\frac{1}{\sqrt{n}h^{dm}}S_{g,h,\mathbf{t}}\right\|_{\mathbb{L}_2(\mathbb{Q})}\leq\left\|\frac{1}{\sqrt{n}h^{dm}}S_{g,h,\mathbf{t}}\right\|_\infty<\frac{\eta}{2},$$

holds, and hence

$$\left\|\frac{1}{\sqrt{n}h^{dm}}S_{g,h,\mathbf{t}}-S_0\right\|_{\mathbb{L}_2(\mathbb{Q})}\leq\left\|\frac{1}{\sqrt{n}h^{dm}}S_{g,h,\mathbf{t}}\right\|_{\mathbb{L}_2(\mathbb{Q})}+\|S_0\|_{\mathbb{L}_2(\mathbb{Q})}<\eta.$$

Therefore Eq. 6.8 is shown. Hence by combining Eqs. F.ii, F.iii and 2.8 with Lemma 9.9, p.160 of Kosorok (2008), gives that every probability measure $\mathbb{Q}$ on $\mathbb{R}^d$ and for every $\eta\in\left(0,\sqrt{n}h^{-dm}\|S_{g,h,\mathbf{t}}\|_\infty\right)$, the covering number $\mathcal{N}(\mathcal{S}_n,\eta)$ is upper bounded as

$$\sup_{\mathbb{Q}}\mathcal{N}(\mathcal{S}_n,L_2(\mathbb{Q}),\eta)$$

$$\leq \quad \mathcal{N}\left(\left[l_n,\left(\frac{\eta}{2\|S_{g,h,\mathbf{t}}\|_\infty}\right)^{-1/dm}\right],|\cdot|,\frac{\sqrt{n}\eta l_n^{dm+1}}{3dm\|S_{g,h,\mathbf{t}}\|_\infty}\right)$$

$$\sup_{\mathbb{Q}} \mathcal{N}\left(m\mathscr{F}_q, L_2(\mathbb{Q}), \widetilde{\varepsilon}\right)$$

$$\sup_{\mathbb{Q}} \mathcal{N}\left(\mathscr{K}^m, L_2(\mathbb{Q}), \frac{l_n^{dm}\eta}{3mM}\right) + 1$$

$$\leq \frac{3dm \left\|S_{g,h,\mathbf{t}}\right\|_\infty}{\eta l_n^{dm+1}} \left(\frac{2\left\|S_{g,h,\mathbf{t}}\right\|_\infty}{\eta}\right)^{1/dm} \left(\frac{3A_1 mM \|K\|_\infty^m}{\eta l_n^{dm}}\right)^{2\nu_1 - 1}$$

$$\left(\frac{3A_2 mM \|K\|_\infty}{\eta l_n^{dm}}\right)^{m\nu_2} + 1$$

$$\leq \left(\frac{3Adm \left\|S_{g,h,\mathbf{t}}\right\|_\infty}{\eta l_n^{dm}}\right)^{2\nu_1 + m\nu_2 - 1} \left[\left(\frac{3d^{2-2\nu_1 - m\nu_2} \left\|S_{g,h,\mathbf{t}}\right\|_\infty}{\eta l_n^{dm+1}}\right)\right.$$

$$\left.\left(\frac{2\left\|S_{g,h,\mathbf{t}}\right\|_\infty}{\eta}\right)^{1/dm} + \left(\frac{3Adm \left\|S_{g,h,\mathbf{t}}\right\|_\infty}{\eta l_n^{dm}}\right)^{-2\nu_1 - m\nu_2 + 1}\right]$$

$$\leq \left(\frac{3Adm \left\|S_{g,h,\mathbf{t}}\right\|_\infty}{\eta l_n^{dm}}\right)^{2\nu_1 + m\nu_2 - 1}, \tag{6.13}$$

for some finite constant $0 < A < \infty$. For $p > 2$, note that assumption (2.8) implies that

$$\sup_{\mathbf{x} \in \mathbf{I}} \mathbb{E}(F^p(\mathbf{Y}) \mid \mathbf{X} = \mathbf{x}) < M^p < \infty.$$

From Lemma 2.3 and using Jensen's inequality, we observe that, for $p \geq k$

$$\mathbb{E}\left[S_{g,h,\mathbf{t}}(X,Y)\right]^k \leq m^k n^{\frac{k}{2}} h^{kdm} \mathbb{E}\left(\overline{G}_{g,h,\mathbf{t}}^k(\mathbf{X},\mathbf{Y})\right)$$

$$\leq m^k n^{\frac{k}{2}} \mathbb{E}\left(\widetilde{K}^k\left(\frac{\mathbf{t}-\mathbf{X}}{h}\right) \mathbb{E}\left(g^k(\mathbf{Y}) \mid \mathbf{X} = \mathbf{x}\right)\right),$$

where

$$\mathbb{E}(g^k(\mathbf{Y})) \mid \mathbf{X} = \mathbf{x}) \leq \sup_{\mathbf{x} \in \mathbf{J}} \mathbb{E}(F^k(\mathbf{Y}) \mid \mathbf{X} = \mathbf{x}).$$

Then, we have

$$\mathbb{E}\left[S_{g,h,\mathbf{t}}(X,Y)\right]^2 \leq M^p m^2 \mathbb{E}\left[\prod_{j=1}^m K^2\left(\frac{\mathbf{t}_j - \mathbf{X}}{h}\right)\right].$$

By Hölder inequality and using once more Lemma 2.3 we obtain

$$\mathbb{E}\left[\prod_{j=1}^m K^2\left(\frac{\mathbf{t}_j - \mathbf{X}}{h}\right)\right] \leq \prod_{j=1}^m \left(\mathbb{E}\left[K^{2p_j}\left(\frac{\mathbf{t}_j - \mathbf{X}}{h}\right)\right]\right)^{1/p_j},$$

where

$$\sum_{j=1}^{m} \frac{1}{p_j} = 1, \quad 0 < p_j \le \infty, j = 1, \ldots, m.$$

Hence, from Lemma 2.3, we have

$$\mathbb{E}\left[\prod_{j=1}^{m} K^2\left(\frac{\mathbf{t}_j - \mathbf{X}}{h}\right)\right] \le \widetilde{C}_{\mathbb{P},K,\varepsilon}^m h^{m(d_{\mathrm{vol}}-\varepsilon)}, \tag{6.14}$$

where

$$\widetilde{C}_{\mathbb{P},K,\varepsilon} = \max_{1 \le j \le m} C_{2p_j,\mathbb{P},K,\varepsilon}.$$

Then, we readily obtain

$$\mathbb{E}\left[S_{g,h,\mathbf{t}}(X,Y)\right]^2 \le \widetilde{C}_{\mathbb{P},K,\varepsilon}^m M^p m^2 h^{m(d_{\mathrm{vol}}-\varepsilon)}. \tag{6.15}$$

Now from Eq. 6.15, applying Theorem 7.1 to the class $\mathcal{S}_n$ defined in Eq. 6.7 gives that

$$\sup_{S_{g,h,\mathbf{t}} \in \mathcal{S}_n} \left| \frac{1}{n} \sum_{i=1}^{n} S_{g,h,\mathbf{t}}(X_i, Y_i) - \mathbb{E}(S_{g,h,\mathbf{t}}(X_1, Y_1)) \right|$$

is upper bounded with probability at least $1 - \delta$ as

$$\sup_{S_{g,h,\mathbf{t}} \in \mathcal{S}_n} \left| \frac{1}{n} \sum_{i=1}^{n} S_{g,h,\mathbf{t}}(X_i, Y_i) - \mathbb{E}(S_{g,h,\mathbf{t}}(X_1, Y_1)) \right| \le C_{A,\|\widetilde{\mathbf{F}}\|_\infty,\nu,d_{\mathrm{vol}},\widetilde{C}_{\mathbb{P},K,\varepsilon},m,\varepsilon}$$

$$\times \left(\frac{\left(\log\left(\frac{1}{h}\right)\right)_+}{n} + \sqrt{\frac{h^{m(d_{\mathrm{vol}}-\varepsilon)}\left(\log\left(\frac{1}{h}\right)\right)_+}{n}} \right.$$

$$\left. + \sqrt{\frac{h^{m(d_{\mathrm{vol}}-\varepsilon)}\log\left(\frac{2}{\delta}\right)}{n}} + \frac{\log\left(\frac{2}{\delta}\right)}{n} \right).$$

Using the condition

$$\limsup_{n} \frac{\left(\log\left(\frac{1}{l_n}\right)\right)_+ + \log\left(\frac{2}{\delta}\right)}{n l_n^{d_{\mathrm{vol}}-\varepsilon}} < \infty,$$

we conclude that

$$\sup_{h \geq l_n} \sup_{g \in \mathscr{F}_q} \sup_{\mathbf{t} \in \mathbf{I}} \frac{m \sqrt{n l_n^{m(2d - d_{\mathrm{vol}} + \varepsilon)}} |U_n^{(1)}(\pi_1 \overline{G}_{g,h,\mathbf{t}})|}{\sqrt{|\log l_n| \vee \log(2/\delta)}} \leq C_1.$$

6.4.   The Other Terms of Eq. 6.1 We will follow the steps of the proof of Dony and Mason (2008). Now we consider the other terms of the Hoeffding decomposition (6.1) and show that is almost surely upper bounded, that is, for each $k = 2, \ldots, m$,

$$\sup_{h \geq l_n} \sup_{g \in \mathscr{F}_q} \sup_{\mathbf{t} \in \mathbb{R}^{dm}} \frac{\binom{m}{k} \sqrt{n l_n^{m(2d - d_{\mathrm{vol}} + \varepsilon)}} |U_n^{(k)}(\pi_k \overline{G}_{g,h,\mathbf{t}})|}{\sqrt{|\log l_n| \vee \log(2/\delta)}} \leq C_2. \tag{6.16}$$

By the fact that $n l_n^{dm} = c^{dm} \log n$, this will be established if we can obtain that for each $k = 2, \ldots, m$,

$$\sup_{h \geq l_n} \sup_{g \in \mathscr{F}_q} \sup_{\mathbf{t} \in \mathbb{R}^{dm}} \frac{\sqrt{n l_n^{m(2d - d_{\mathrm{vol}} + \varepsilon)}} |U_n^{(k)}(\pi_k \overline{G}_{g,h,\mathbf{t}})|}{(\sqrt{|\log l_n| \vee \log(2/\delta)})^k} = O\left(\frac{1}{\sqrt{l_n^{dm} n^{k-1}}}\right). \tag{6.17}$$

To establish the uniform in bandwidth convergence rates, we have to use a blocking argument and a decomposition of the interval $[l_n, b_0]$, for b_0 large enough, into smaller intervals. For this, set $n_\ell = 2^\ell, \ell \geq 0$ and consider the intervals $\mathcal{H}_{\ell,j} := [h_{\ell,j-1}, h_{\ell,j}]$ where the boundaries are given by $h_{\ell,j}^m := 2^j l_{n_\ell}^m$. By setting

$$L(\ell) = \max\{j : h_{\ell,j} \leq 2b_0\},$$

remark that

$$[l_{n_\ell}, b_0] \subseteq \bigcup_{\ell=1}^{L(\ell)} \mathcal{H}_{\ell,j} \quad \text{and} \quad L(\ell) \sim \log\left(\frac{n_\ell b_0}{c \log n_\ell}\right) / \log 2, \tag{6.18}$$

implying, in particular, that $L(\ell) \leq 2 \log n_\ell$. This fact will be used tacitly to conclude some crucial steps of the proofs. Next, for $1 \leq j \leq L(\ell)$, consider the class of functions on $\mathbb{R}^{dm} \times \mathbb{R}^{qm}$,

$$\mathcal{G}_{\ell,j} := \{h^{dm} \bar{G}_{g,h,\mathbf{t}}(\cdot, \cdot) : g \in \mathcal{F}_q, h \in \mathcal{H}_{\ell,j}, \mathbf{t} \in \mathbb{R}^{dm}\},$$

as well the class on $\mathbb{R}^{dm} \times \mathbb{R}^{qm}$,

$$\mathcal{G}_{\ell,j}^{(k)} := \left\{ \frac{h^{dm} \pi_k \bar{G}_{g,h,\mathbf{t}}(\cdot, \cdot)}{M_k} : g \in \mathcal{F}_q, h \in \mathcal{H}_{\ell,j}, \mathbf{t} \in \mathbb{R}^{dm} \right\},$$

where $M_k = 2^k \kappa^m M$. Clearly, each class $\mathcal{G}_{\ell,j}$ is of VC-type with the same characteristics as $\mathcal{G}^k$ (and thus as $\mathcal{G}$) with envelope function $M_k^{-1} F_k$, where F_k is the envelope function of $\mathcal{G}^k$. Notice that from Eqs. 2.8 and 6.6,

$$M_k \geq \sup_{\mathbf{x}, \mathbf{y} \in \mathbb{R}^k} \left\{ |\pi_k \bar{G}_{g,h,\mathbf{t}(\mathbf{x},\mathbf{y})}| : g \in \mathcal{F}_q, 0 < h < 1, \mathbf{t} \in \mathbb{R}^{dm} \right\}$$

and hence each function in $\mathcal{G}_{\ell,j}^{(k)}$ is bounded by 1. Define now for $n_{\ell-1} \leq n \leq n_\ell, \ell = 1, 2, \dots,$

$$\mathcal{U}_n(j, k, \ell) = n_\ell^{-k/2} \sup_{H \in \mathcal{G}_{\ell,j}^{(k)}} \left| \sum_{\mathbf{i} \in I_n^k} H(\mathbf{X_i}, \mathbf{Y_i}) \right|. \tag{6.19}$$

From Theorem 4 of Giné and Mason (2007b) as in Dony and Mason (2008), we get for $c = 1/2, r = 2$ and all $x > 0$ that for any $\ell \geq 1$,

$$\mathbb{P} \left\{ \max_{n_{\ell-1} < n \leq n_\ell} \mathcal{U}_n(j, k, \ell) > x \right\} \leq \frac{2}{x} \mathbb{P} \left\{ \mathcal{U}_{n_\ell}(j, k, \ell) > x/2 \right\}^{1/2} \mathbb{E}[\mathcal{U}_{n_\ell}^2(j, k, \ell)]^{1/2}. \tag{6.20}$$

We shall apply an exponential inequality and a moment bound for U-statistics, due to, respectively, de la Peña and Giné (1999) and Giné and Mason (2007b), on the class $\mathcal{G}_{\ell,j}(k)$ to bound (6.20). To use these results, we must first derive some bounds. First, it is readily checked that

$$\begin{aligned}
\mathcal{U}_n(j, k, \ell) &= n_\ell^{-k/2} \sup_{H \in \mathcal{G}_{\ell,j}^{(k)}} \left| \sum_{\mathbf{i} \in I_n^k} H(\mathbf{X_i}, \mathbf{Y_i}) \right| \\
&= n_\ell^{-k/2} \binom{n}{k} \binom{n}{k}^{-1} \sup_{H \in \mathcal{G}_{\ell,j}^{(k)}} \left| \sum_{\mathbf{i} \in I_n^k} H(\mathbf{X_i}, \mathbf{Y_i}) \right| \\
&= n_\ell^{-k/2} \binom{n}{k} \| U_n^{(k)}(\pi_k G) \|_{\mathcal{G}_{\ell,j}^{(k)}} \\
&\leq n_\ell^{k/2} \| U_n^{(k)}(\pi_k G) \|_{\mathcal{G}_{\ell,j}^{(k)}}, \tag{6.21}
\end{aligned}$$

for all $n_{\ell-1} < n \leq n_\ell$. Second, notice that in Assumption 2, the kernel $K(\cdot)$ is assumed to be bounded by κ and, for notational convenience in the proofs, to have support in $[-1/2, 1/2]$, so that by assumption (2.8) and $M_k = 2^k \kappa^m M$, for $H \in \mathcal{G}_{\ell,j}^{(k)}$, we have by Eqs. 6.2 and 6.14,

$$\mathbb{E} H^2(\mathbf{X}, \mathbf{Y}) \leq M_k^{-2} h^{2dm} \mathbb{E} \bar{G}_{g,h,\mathbf{t}}^2(\mathbf{X}, \mathbf{Y})$$

$$
= M_k^{-2} \mathbb{E}\left[g^2(\mathbf{Y}) \tilde{K}^2 \left(\frac{\mathbf{t} - \mathbf{X}}{h} \right) \right]
$$

$$
\leq \widetilde{C}_{k,\mathbb{P},K,\varepsilon}^m 4^{-k} \kappa^{-2m} h^{m(d_{\mathrm{vol}}-\varepsilon)}.
$$

For $D_{m,k} = \widetilde{C}_{k,\mathbb{P},K,\varepsilon}^m 4^{-k}\kappa^{-2m}$, this gives us that

$$
\sup_{H \in \mathcal{G}_{\ell,j}^{(k)}} \mathbb{E} H^2(\mathbf{X},\mathbf{Y}) \leq D_{m,k} h_{\ell,j}^{m(d_{\mathrm{vol}}-\varepsilon)} =: \sigma_{\ell,j}^2. \tag{6.22}
$$

Since $\pi_k \pi_k L = \pi_k L$ for all $k \leq 1$, we can now apply Corollary 1 of Giné and Mason ([2007b](#)) to the class $\mathcal{G}_{\ell,j}^{(k)}$ with σ^2 as in Eq. [6.22](#) and easily obtain that for some constant A_k,

$$
\mathbb{E}\mathcal{U}_{n_\ell}^2(j,k,\ell) \leq n_\ell^k \| U_{n_\ell}^{(k)}(\pi_k G) \|_{\mathcal{G}_{\ell,j}^{(k)}}^2 \leq 2^k A_k h_{\ell,j}^{m(d_{\mathrm{vol}}-\varepsilon)} |\log(h_{\ell,j})|^k
$$

$$
\leq 2^k A_k h_{\ell,j}^{m(d_{\mathrm{vol}}-\varepsilon)} |\log(l_n)|^k. \tag{6.23}
$$

To control the probability term in Eq. [6.20](#), we shall apply an exponential inequality to the same class $\mathcal{G}_{\ell,j}^{(k)}$ recall that each $H \in \mathcal{G}_{\ell,j}^{(k)}$ is bounded by 1. Setting

$$
y^* = C_{1,k}(|\log l_n| \vee \log\log(2/\delta))^{k/2} =: C_{1,k}\lambda_{n,k}, \tag{6.24}
$$

where $C_{1,k} < \infty$, Theorem 5.3.14 of de la Peña and Giné ([1999](#)) gives us constants $C_{2,k}, C_{3,k}$ and $C_{4,k}$ such that for $j = 1, \ldots, L(\ell)$ and for any $\rho > 1$,

$$
\mathbb{P}\left\{ \mathcal{U}_{n_\ell}(j,k,\ell) > \rho^{k/2} y^* \right\} \leq C_{2,k} \exp\{-C_{3.k}\rho y^{*2/k}\}
$$

$$
\leq \exp\{-C_{4.k}\rho \log\log(2/\delta)\}, \tag{6.25}
$$

plugging the bounds Eqs. [6.23](#) and [6.25](#) into Eq. [6.20](#), we then get for some $C_{5,k} > 0$, and $\rho \geq 2$ and ℓ large enough,

$$
\mathbb{P}\left\{ \max_{n_{\ell-1} < n \leq n_\ell} \mathcal{U}_n(j,k,\ell) > 2\rho^{k/2} y^* \right\}
$$

$$
\leq \frac{(\log 2/\delta)^{-\rho C_{4,k}/2} \sqrt{2^k A_k h_{\ell,j}^{m(d_{\mathrm{vol}}-\varepsilon)} |\log(l_n)|^k}}{C_{1,k}\sqrt{\rho^k(|\log l_n| \vee \log(2/\delta))^k}}
$$

$$
\leq \sqrt{h_{\ell,j}^{m(d_{\mathrm{vol}}-\varepsilon)}} \log(2/\delta)^{-\rho C_{5,k}}. \tag{6.26}
$$

Finally, note also that

$$
n_\ell^{k/2} \| U_n^{(k)}(\pi_k G) \|_{\mathcal{G}_{\ell,j}(k)} = n_\ell^k \binom{n}{k}^{-1} n_\ell^{-k/2} \sup_{H \in \mathcal{G}_{\ell,j}^{(k)}} \left| \sum_{\mathbf{i} \in I_n^k} H(\mathbf{X_i}, \mathbf{Y_i}) \right|
$$

$$= n_\ell^k \binom{n}{k}^{-1} \mathcal{U}_n(j,k,\ell)$$

$$\leq k! \prod_{j=0}^{k-1} \left(\frac{n_\ell}{n-j}\right) \mathcal{U}_n(j,k,\ell)$$

$$\leq C_k M_k \mathcal{U}_n(j,k,\ell), \tag{6.27}$$

for some $C_k > 0$. Therefore, by Eq. 6.18, for each $k = 2, \ldots, m$ and ℓ large enough,

$$\max_{n_{\ell-1} < n \leq n_\ell} A_{n,k}$$

$$:= \max_{n_{\ell-1} < n \leq n_\ell} \sup_{l_n \leq h \leq b_0} \sup_{g \in \mathscr{F}_q} \sup_{\mathbf{t} \in \mathbf{I}} \frac{\sqrt{n l_n^{m(2d-d_{\mathrm{vol}}+\varepsilon)}} |U_n^{(k)}(\pi_k \overline{G}_{g,h,\mathbf{t}}))|}{\sqrt{(|\log l_n| \vee \log(2/\delta))^k}}$$

$$\leq \max_{n_{\ell-1} < n \leq n_\ell} \max_{1 \leq j \leq L(\ell)} \sup_{h \in \mathcal{H}_{\ell,j}} \sup_{g \in \mathscr{F}_q} \sup_{\mathbf{t} \in \mathbf{I}} \frac{\sqrt{n_\ell l_n^{m(2d-d_{\mathrm{vol}}+\varepsilon)}} |U_n^{(k)}(\pi_k \overline{G}_{g,h,\mathbf{t}}))|}{\sqrt{(|\log l_n| \vee \log(2/\delta))^k}}$$

$$\leq \frac{C_k M_k}{\sqrt{n_\ell^{k-1}}} \max_{n_{\ell-1} < n \leq n_\ell} \max_{1 \leq j \leq L(\ell)} \frac{\mathcal{U}_{n_\ell}(j,k,\ell)}{\lambda_{n,k}}$$

$$\leq \frac{C_k M_k}{\sqrt{l_{n_\ell}^{dm} n_\ell^{k-1}}} \max_{n_{\ell-1} < n \leq n_\ell} \max_{1 \leq j \leq L(\ell)} \frac{\mathcal{U}_{n_\ell}(j,k,\ell)}{\lambda_{n,k}},$$

where $\lambda_{n,k}$ was defined as in Eq. 6.24. Now, recall that $L(\ell) \leq 2\log(n_\ell)$. Then Eq. 6.26 applied with $\rho \geq (2+\gamma)/C_{5,k}$, $\gamma > 0$ and in combination with the above inequality and the obvious bound

$$\sqrt{l_n^{dm} n^{k-1}} A_{n,k} \leq \sqrt{l_{n_\ell}^{dm} n_\ell^{k-1}} A_{n,k},$$

valid for all $n_{\ell-1} < n \leq n_\ell$, implies for $C_{6,k} \geq 2\rho^{k/2} C_k M_k C_{1,k}$ and for the choice $\delta = 2^{-\ell+1}$ that for $k = 2, \ldots, m$,

$$\mathbb{P}\left\{ \max_{n_{\ell-1} < n \leq n_\ell} \sqrt{l_n^{dm} n^{k-1}} A_{n,k} > C_{6,k} \right\} \leq \sum_{j=1}^{L(\ell)} \sqrt{h_{\ell,j}^{m(d_{\mathrm{vol}}-\varepsilon)}} \log(2/\delta)^{-\rho C_{5,k}}$$

$$\leq L(\ell) \log(2/\delta)^{-\rho C_{5,k}}$$

$$\leq 2(\ell \log 2)^{-(1+\gamma)}. \tag{6.28}$$

This proves, via Borel-Cantelli, that Eq. 6.17 holds, which obviously implies Eq. 6.16 and hence complete the proof of Theorem 2.5.

6.5. *Proof of Theorem 2.6: the Unbounded Case* To prove this theorem, we will need to truncate the conditional U-statistic $u_n(g, h, \mathbf{t})$. If condition (2.8) is not satisfied, we consider bandwidths lying in the smaller interval $\mathcal{H}'_{n_\ell} = [a'_{n_\ell}, b_0]$, that may be divided into subintervals as follows

$$\mathcal{H}'_{\ell,j} := [h'_{\ell,j-1}, h'_{\ell,j}] \tag{6.29}$$

where the boundaries are given by $h'^{dm}_{\ell,j} := 2^j a'^{dm}_{n_\ell}$. Note that it is straightforward to show that Eq. 6.18 remains valid if we replace $h_{\ell,j}$ by $h'_{\ell,j}$. In particular, we still have $L(\ell) \leq 2 \log n_\ell$, where $L(\ell)$ is now defined as

$$L(\ell) := \max\{j : h'_{\ell,j} \leq 2b_0\}.$$

Recall that $n_\ell = 2^\ell, \ell \geq 0$, and set for, $\ell \geq 1$,

$$\gamma_\ell = n_\ell / \log n_\ell. \tag{6.30}$$

For an arbitrary $\epsilon > 0$, we truncate each function $\mathcal{G}$, either the envelope function as follows

$$
\begin{aligned}
\overline{G}_{g,h,\mathbf{t}}(\mathbf{x}, \mathbf{y}) &= \overline{G}_{g,h,\mathbf{t}}(\mathbf{x}, \mathbf{y}) \mathbb{1}\{\tilde{F}(\mathbf{y}) \leq \epsilon \gamma_\ell^{1/p}\} + \overline{G}_{g,h,\mathbf{t}}(\mathbf{x}, \mathbf{y}) \mathbb{1}\{\tilde{F}(\mathbf{y}) > \epsilon \gamma_\ell^{1/p}\} \\
&= \overline{G}^{(\ell)}_{g,h,\mathbf{t}}(\mathbf{x}, \mathbf{y}) + \tilde{G}^{(\ell)}_{g,h,\mathbf{t}}(\mathbf{x}, \mathbf{y}),
\end{aligned}
$$

where $\tilde{F}$ is the symmetric envelope function of the class $\mathcal{G}$ as defined in Eq. 6.4. $u_n(g, h, \mathbf{t})$ can then also be decomposed for any $n_{\ell-1} < n \leq n_\ell$ since, from Eq. 2.11,

$$
\begin{aligned}
u_n(g, h, \mathbf{t}) &= \sqrt{n}\{U_n^{(m)}(\bar{G}^{(\ell)}_{g,h,\mathbf{t}}) - \mathbb{E}U_n^{(m)}(\bar{G}^{(\ell)}_{g,h,\mathbf{t}})\} + \sqrt{n}\{U_n^{(m)}(\tilde{G}^{(\ell)}_{g,h,\mathbf{t}}) \\
&\quad - \mathbb{E}U_n^{(m)}(\tilde{G}^{(\ell)}_{g,h,\mathbf{t}})\} \\
&=: u_n^{(\ell)}(g, h, \mathbf{t}) + \tilde{u}_n^{(\ell)}(g, h, \mathbf{t}).
\end{aligned}
$$

The term $u_n^{(\ell)}(g, h, \mathbf{t})$ will be called the truncated part and $\tilde{u}_n^{(\ell)}(g, h, \mathbf{t})$ the remainder part. To prove Theorem (2.6), we shall apply the Hoeffding decomposition to the truncated part and analyze each of the terms separately, while the remaining part can be treated directly using simple arguments based on standard inequalities. Note, for further use, that

$$a'^{dm}_{n_\ell} = c^{dm} \gamma_\ell^{2/p-1}, \qquad \ell \geq 1. \tag{6.31}$$

6.5.1. *Truncated Part.* First, note that by Hoeffding decomposition (6.1), we need to consider the terms of

$$\sum_{k=1}^{m} \binom{m}{k} U_n^{(k)} (\pi_k \overline{G}_{g,h,\mathbf{t}}^{(\ell)}).$$

We shall start with the linear term in this decomposition. Following the same reasoning as in the previous section, we can show that $\pi_1 \overline{G}_{g,h,\mathbf{t}}^{(\ell)}$ is a centered conditional expectation and that the first term of Eq. 6.1 can be written as an empirical process based on the sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ and indexed by the class of functions

$$\mathcal{S}'_\ell := \left\{ S_{g,h,\mathbf{t}}^{(\ell)}(\cdot, \cdot) : g \in \mathscr{F}_q, h \in \mathcal{H}'_{n_\ell}, \mathbf{t} \in \mathbf{I} \subset \mathbb{X} \right\},$$

where $h \in \mathcal{H}'_{n_\ell}$ was defined at the beginning of this section and where

$$S_{g,h,\mathbf{t}}^{(\ell)}(x, y) = mh^{dm} \mathbb{E} \left(\overline{G}_{g,h,\mathbf{t}}^{(\ell)}(\mathbf{X}, \mathbf{Y}) \mid (X_1, Y_1) = (x, y) \right).$$

To show that $\mathcal{S}'_\ell$ is a VC-class, introduce the class of functions of $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{dm} \times \mathbb{R}^{qm}$,

$$\mathcal{C} := \left\{ h^{dm} \overline{G}_{g,h,\mathbf{t}}(\mathbf{x}, \mathbf{y}) \mathbb{1}\{\tilde{F}(\mathbf{y}) \leq c\} : g \in \mathscr{F}_q, 0 < h < 1, \mathbf{t} \in \mathbf{I} \subset \mathbb{X} \right\}.$$

Since both $\mathcal{G}'$ as defined below

$$\mathcal{G}' = \{ h^{dm} \overline{G}_{g,h,\mathbf{t}}(\cdot, \cdot) : g \in \mathscr{F}_q, 0 < h < 1, \mathbf{t} \in \mathbb{R}^{dm} \}, \tag{6.32}$$

and the class of functions of $\mathbf{y} \in \mathbb{R}^{qm}$ given by $\mathcal{I} = \{\mathbb{1}\{\tilde{F}(\mathbf{y}) \leq c\} : c > 0\}$ are of VC-type (and note that $\mathcal{I}$ has a bounded envelope function), we can apply Lemma A.1 in Einmahl and Mason (2000) to conclude that $\mathcal{C}$ is also of VC-type. Therefore, so is the class of functions $m\mathcal{C}^{(1)}$ on $\mathbb{R}^{d+q}$, where $\mathcal{C}^{(1)}$ consists of the π_1 -projections of the functions in the class $\mathcal{C}$. Thus, we see that $\mathcal{S}'_\ell \subset m\mathcal{C}^{(1)}$ and hence $\mathcal{S}'_\ell$ is of VC -type with the same characteristics as $m\mathcal{C}^{(1)}$. Now, to find an envelope function for $\mathcal{S}'_\ell$, set $\mathbf{t}_j := (t_1, \ldots, t_{j-1}, t_{j+1}, \ldots, t_m) \in \mathbb{R}^{d(m-1)}$ and $\mathbf{Z}_j(u) := (Z_1, \ldots, Z_{j-1}, u, Z_{j+1}, \ldots, Z_m) \in \mathbb{R}^{qm}$ for $u \in \mathbb{R}^q$ and $\mathbf{Z} \in \mathbb{R}^{qm}$. We can then rewrite the function

$S_{g,h,\mathbf{t}}^{(\ell)}(x,y) \in \mathcal{S}_\ell'$ as

$$
\begin{aligned}
S_{g,h,\mathbf{t}}^{(\ell)}(x,y) &= K\left(\frac{t_1 - x}{h}\right) \mathbb{E}\left[g\left(\mathbf{Y}_1(y)\right) \tilde{K}\left(\frac{\mathbf{t}_1 - \mathbf{X}^*}{h}\right)\right. \\
&\qquad \left. \times \mathbb{1}\left\{\widetilde{F}\left(\mathbf{Y}_1(y)\right) \le \varepsilon \gamma_\ell^{1/p}\right\}\right] \\
&\quad + K\left(\frac{t_2 - x}{h}\right) \mathbb{E}\left[g\left(\mathbf{Y}_2(y)\right) \widetilde{K}\left(\frac{\mathbf{t}_2 - \mathbf{X}^*}{h}\right)\right. \\
&\qquad \left. \times \mathbb{1}\left\{\widetilde{F}\left(\mathbf{Y}_2(y)\right) \le \varepsilon \gamma_\ell^{1/p}\right\}\right] \\
&\quad + \cdots + K\left(\frac{t_m - x}{h}\right) \mathbb{E}\left[g\left(\mathbf{Y}_m(y)\right) \widetilde{K}\left(\frac{\mathbf{t}_m - \mathbf{X}^*}{h}\right)\right. \\
&\qquad \left. \times \mathbb{1}\left\{\widetilde{F}\left(\mathbf{Y}_m(y)\right) \le \varepsilon \gamma_\ell^{1/p}\right\}\right],
\end{aligned}
$$

where $\mathbf{X}^* = (X_2, \ldots, X_m) \in \mathbb{R}^{d(m-1)}$ and where (with a little abuse of notation here) the product kernel in (K.iii) is now defined for $d(m-1)$-dimensional vectors, that is, $\widetilde{K}(\mathbf{u}) = \prod_{i=1}^{m-1} K(u_i)$, $\mathbf{u} \in \mathbb{R}^{d(m-1)}$. Hence, we can bound $S_{g,h,\mathbf{t}}^{(\ell)}(x,y) \in \mathcal{S}_\ell'$ simply as

$$
\begin{aligned}
\left|S_{g,h,\mathbf{t}}^{(\ell)}(x,y)\right| &\le K\left(\frac{t_1 - x}{h}\right) \mathbb{E}\left[F\left(y, Y_2, \ldots, Y_m\right) \tilde{K}\left(\frac{\mathbf{t}_1 - \mathbf{X}^*}{h}\right)\right] \\
&\quad + K\left(\frac{t_2 - x}{h}\right) \mathbb{E}\left[F\left(Y_2, y, Y_3, \ldots, Y_m\right) \widetilde{K}\left(\frac{\mathbf{t}_2 - \mathbf{X}^*}{h}\right)\right] \\
&\quad + \cdots + K\left(\frac{t_m - x}{h}\right) \mathbb{E}\left[F\left(Y_2, \ldots, Y_m, y\right) \widetilde{K}\left(\frac{\mathbf{t}_m - \mathbf{X}^*}{h}\right)\right] \\
&=: G_m(x,y).
\end{aligned}
$$

We shall now apply the moment bound in Theorem 7.3 to the subclasses

$$
\mathcal{S}_{\ell,j}' := \left\{S_{g,h,\mathbf{t}}^{(\ell)}(\cdot,\cdot) : g \in \mathcal{F}_q, h \in \mathcal{H}_{\ell,j}', \mathbf{t} \in \mathbb{R}^{dm}\right\}, \quad 1 \le j \le L(\ell),
$$

where $\mathcal{H}_{\ell,j}'$ was defined in Eq. 6.29. Since $\mathcal{S}_{\ell,j}' \subset \mathcal{S}_\ell'$ for $j = 1, \ldots, L(\ell)$, all of these subclasses are of VC-type, with the same envelope function and characteristics as the class $m\mathcal{C}^{(1)}$ (which is independent of ℓ), verifying (ii) in Theorem 7.3. For (i), recall that although all of the terms of the envelope function $G_m(x,y)$ are different, their expectations are the same. Therefore, writing $\mathbf{Y}^*$ for $(Y_2, \ldots, Y_m)$ and applying Minkowski's inequality followed by Jensen's inequality, we obtain from assumption (2.10) the following upper bound for the second moment of the envelope function:

$$
\mathbb{E}G_m^2(X,Y) = \kappa^{2m} \mathbb{E}_Y\left\{\mathbb{E}_{\mathbf{Y}^*}\left[F\left(Y, Y_2, \ldots, Y_m\right)\right]\right.
$$

$$+\mathbb{E}_{\mathbf{Y}^*}\left[F\left(Y_2, Y, Y_3, \ldots, Y_m\right)\right] + \cdots + \mathbb{E}_{\mathbf{Y}^*}\left[F\left(Y_2, \ldots, Y_m, Y\right)\right]\}^2$$
$$\leq m^2 \kappa^{2m} \mathbb{E} F^2\left(Y_1, \ldots, Y_m\right)$$
$$\leq m^2 \kappa^{2m} \mu_p^{2/p}.$$

Note, further, that by the symmetry of $\widetilde{F}$,

$$\mathbb{E}\bar{G}_{g,h,\mathbf{t}}^{(\ell)}(\mathbf{X}, \mathbf{Y}) = h^{-dm}\mathbb{E}\left[g(\mathbf{Y})\widetilde{K}\left(\frac{\mathbf{t}-\mathbf{X}}{h}\right)\mathbb{1}\left\{\widetilde{F}(\mathbf{Y}) \leq \varepsilon\gamma_\ell^{1/p}\right\}\right],$$

so that Jensen's inequality, the change of variable $\mathbf{u} = (\mathbf{t}-\mathbf{x})/h$ and the assumption in Eq. 2.10 give the following upper bound for the second moment of any function in $\mathcal{S}_\ell'$:

$$\begin{aligned}
\mathbb{E}\left(S_{g,h,\mathbf{t}}^{(\ell)}(X, Y)\right)^2 &\leq m^2 \mathbb{E}\left[g^2(\mathbf{Y})\widetilde{K}^2\left(\frac{\mathbf{t}-\mathbf{X}}{h}\right)\mathbb{1}\left\{\widetilde{F}(\mathbf{Y}) \leq \varepsilon\gamma_\ell^{1/p}\right\}\right] \\
&\leq m^2 \mathbb{E}\left[\widetilde{K}^2\left(\frac{\mathbf{t}-\mathbf{X}}{h}\right)\mathbb{E}\left[F^2(\mathbf{Y}) \mid \mathbf{X} = \mathbf{x}\right]\right] \\
&\leq m^2 \mu_p^{2/p}\mathbb{E}\left[\widetilde{K}^2\left(\frac{\mathbf{t}-\mathbf{X}}{h}\right)\right] = m^2 \mu_p^{2/p} C_{2,\mathbb{P},K,\varepsilon} h^{m(d_{\mathrm{vol}}-\varepsilon)}.
\end{aligned}$$
$$(6.33)$$

Therefore, with $\beta \equiv m\mu_p^{1/p}\left(\kappa^m \vee C_{2,\mathbb{P},K,\varepsilon}^{1/2}\right)$, our previous calculations give us that

$$\mathbb{E}G_m^2(X, Y) \leq \beta^2 \quad \text{and} \quad \sup_{S \in \mathcal{S}_{\ell,j}'} \mathbb{E}S^2(X, Y) \leq \beta^2 h_{\ell,j}'^{m(d_{\mathrm{vol}}-\varepsilon)} =: \sigma_{\ell,j}^2,$$

verifying condition (iii) as well. Finally, recall from Eq. 6.4 that since $\mathcal{G}$ has envelope function $\widetilde{F}(\mathbf{y})$, it holds for all $x, y \in \mathbb{R}^{d+q}$ that

$$\left|S_{g,h,\mathbf{t}}^{(\ell)}(x, y)\right| \leq m\mathbb{E}\left[\widetilde{F}(\mathbf{Y})\mathbb{1}\left\{\widetilde{F}(\mathbf{Y}) \leq \varepsilon\gamma_\ell^{1/p}\right\} \mid (X_1, Y_1) = (x, y)\right] \leq m\varepsilon\gamma_\ell^{1/p},$$

so that by taking $\varepsilon > 0$ small enough, Theorem 7.3 is now applicable. Thus, for an absolute constant $A_1 < \infty$, we have

$$\begin{aligned}
\mathbb{E}\left\|\sum_{i=1}^{n_\ell} \epsilon_i S\left(X_i, Y_i\right)\right\|_{\mathcal{S}_{\ell,j}'} &\leq A_1 \sqrt{n_\ell h_{\ell,j}'^{m(d_{\mathrm{vol}}-\varepsilon)}\left|\log h_{\ell,j}'\right|} \\
&\leq A_1 \sqrt{n_\ell h_{\ell,j}'^{m(d_{\mathrm{vol}}-\varepsilon)}}\left(\left|\log h_{\ell,j}'\right| \vee \log\log n_\ell\right)
\end{aligned}$$

$$=: \quad A_1 \lambda_j'(\ell), \tag{6.34}$$

where $\epsilon_1, \ldots, \epsilon_{n_\ell}$ are independent Rademacher variables, independent of $(X_i, Y_i), 1 \leq i \leq n_\ell$. Consequently, applying the exponential inequality of Talagrand (1994) to the class $\mathcal{S}_{\ell,j}'$ (see Theorem 7.5 in the Appendix) with $M = m\varepsilon\gamma_\ell^{1/p}, \sigma_{\mathcal{S}_{l,i}'}^2 = \beta^2 h_{\ell,j}'^{m(d_{\mathrm{vol}}-\varepsilon)}$ and the moment bound in Eq. 6.34, we get, for an absolute constant $A_2 < \infty$ and all $t > 0$, that

$$\mathbb{P}\left\{ \max_{n_{\ell-1} < n \leq n_\ell} \left\| \sqrt{n}\alpha_n \right\|_{\mathcal{S}_{\ell,j}'} \geq C_1 \left(A_1 \lambda_j'(\ell) + t \right) \right\}$$

$$\leq \quad 2 \left[\exp\left(-\frac{A_2 t^2}{n_\ell \beta^2 h_{\ell,j}'^{m(d_{\mathrm{vol}}-\varepsilon)}} \right) + \exp\left(-\frac{A_2 t}{m\varepsilon\gamma_\ell^{1/p}} \right) \right]. \tag{6.35}$$

Regarding the application of this inequality with $t = \rho\lambda_j'(\ell), \rho > 1$, note that it clearly follows from Eq. 6.31 and the definitions of $h_{\ell,j}'$ as in Eq. 6.29, γ_ℓ as in Eq. 6.31 and $\lambda_j'(\ell)$ as in Eq. 6.34 that for all $j \geq 0$,

$$\frac{\lambda_j'^2(\ell)}{n_\ell h_{\ell,j}'^{m(d_{\mathrm{vol}}-\varepsilon)}} = \left| \log h_{\ell,j}' \right| \vee \log\log n_\ell \geq \log\log n_\ell,$$

$$\frac{\lambda_j'^2(\ell)}{\gamma_\ell^{2/p}} = 2^j c^{dm} h_{\ell,j}'^{m(d_{\mathrm{vol}}-d-\varepsilon)} \log n_\ell \left(\left| \log h_{\ell,j}' \right| \vee \log\log n_\ell \right) \geq c^{dm} (\log\log n_\ell)^2.$$

Consequently, Eq. 6.35, when applied with $t = \rho\lambda_j'(\ell)$ and any $\rho > 1$ with ℓ large enough, yields, for suitable constants A_2', A_2'' and A_3, the inequality

$$\mathbb{P}\left\{ \max_{n_{\ell-1} < n \leq n_\ell} \left\| \sqrt{n}\alpha_n \right\|_{\mathcal{S}_{\ell,j}'} \geq C_1 \left(A_1 + \rho \right) \lambda_j'(\ell) \right\}$$

$$\leq \quad 2 \left[\exp\left(-A_2'\rho^2 \log\log n_\ell \right) + \exp\left(-A_2''\rho \log\log n_\ell \right) \right]$$

$$\leq \quad 4 \left(\log n_\ell \right)^{-A_3\rho}. \tag{6.36}$$

Keeping in mind that $mh^{dm}\sqrt{n}U_n^{(1)}\left(\pi_1\bar{G}_{g,h,\mathbf{t}}^{(\ell)} \right)$ is the empirical process α_n $\left(S_{g,h,\mathbf{t}}^{(\ell)} \right)$ indexed by the class $\mathcal{S}_\ell'$ and recalling Eq. 6.18, since $d_{\mathrm{vol}} - \varepsilon < d$, we obtain, for $\ell \geq 1$, that

$$\max_{n_{\ell-1} < n \leq n_\ell} A_{n,\ell}' \quad := \quad \max_{n_{\ell-1} < n \leq n_\ell} \sup_{a_n' \leq h \leq b_0} \sup_{g \in \mathcal{F}} \sup_{\mathbf{t} \in \mathbb{R}^{dm}}$$

$$\frac{m\sqrt{n}h^{m(2d-d_{\mathrm{vol}}+\varepsilon)} \left| U_n^{(1)} \left(\pi_1\bar{G}_{g,h,\mathbf{t}}^{(\ell)} \right) \right|}{\sqrt{|\log h| \vee \log\log n}}$$

$$
\leq \max_{n_{\ell-1} < n \leq n_\ell} \max_{\leq j \leq L(\ell)} \sup_{h \in \mathcal{H}'_{\ell,j}} \sup_{g \in \mathcal{F}} \sup_{\mathbf{t} \in \mathbb{R}^{dm}}
$$

$$
\frac{2\sqrt{2^{\frac{d_{\mathrm{vol}} - \varepsilon}{d}}} \left| \sqrt{n} \alpha_n \left(S^{(\ell)}_{g,h,\mathbf{t}} \right) \right|}{\sqrt{n_\ell h'^{m(d_{\mathrm{vol}} - \varepsilon)}_{\ell,j} \left(\left| \log h'_{\ell,j} \right| \vee \log \log n_\ell \right)}}
$$

$$
\leq \max_{n_{\ell-1} < n \leq n_\ell} \max_{\leq j \leq L(\ell)} \sup_{H \in \mathcal{S}'_{\ell,j}} \frac{3 \left| \sqrt{n} \alpha_n(H) \right|}{\lambda'_j(\ell)}.
$$

Consequently, recalling once again that $L(\ell) \leq 2 \log n_\ell$, we can infer from Eq. 6.36 that for some constant $C_5(\rho) \geq 3C_1 (A_1 + \rho)$,

$$
\mathbb{P} \left\{ \max_{n_{\ell-1} < n \leq n_\ell} A'_{n,\ell} > C_5(\rho) \right\}
$$

$$
\leq \sum_{j=1}^{L(\ell)} \mathbb{P} \left\{ \max_{n_{\ell-1} < n \leq n_\ell} \left\| \sqrt{n} \alpha_n \right\|_{\mathcal{S}'_{\ell,j}} > C_1 (A_1 + \rho) \lambda'_j(\ell) \right\}
$$

$$
\leq 8 \left(\log n_\ell \right)^{1 - A_3 \rho}.
$$

The Borel-Cantelli lemma, when combined with this inequality for $\rho \geq (2 + \delta)/A_3, \delta > 0$ and with the choice $n_\ell = 2^\ell$, establishes, for some $C' < \infty$ and with probability 1, that

$$
\limsup_{\ell \to \infty} \max_{n_{\ell-1} < n \leq n_\ell} \sup_{a'_n \leq h \leq b_0} \sup_{g \in \mathcal{F}} \sup_{\mathbf{t} \in \mathbb{R}^{dm}} \frac{m \sqrt{n h^{m(2d - d_{\mathrm{vol}} + \varepsilon)}} \left| U^{(1)}_n \left(\pi_1 \bar{G}^{(\ell)}_{g,h,\mathbf{t}} \right) \right|}{\sqrt{|\log h| \vee \log \log n}} \leq C',
$$
(6.37)

this achieves the control of the first term in Eq. 6.1. We now treat the nonlinear terms. The purpose is to prove that, for $k = 2, \ldots, m$ and with probability 1, all of the other terms of Eq. 6.1 are asymptotically bounded or go to zero at the proper rate, that is

$$
\max_{n_{\ell-1} < n \leq n_\ell} \sup_{a'_n \leq h \leq b_0} \sup_{g \in \mathcal{F}} \sup_{\mathbf{t} \in \mathbb{R}^{dm}} \frac{\sqrt{n h^{m(2d - d_{\mathrm{vol}} + \varepsilon)}} \left| U^{(k)}_n \left(\pi_k \bar{G}^{(\ell)}_{g,h,\mathbf{t}} \right) \right|}{\sqrt{|\log h| \vee \log \log n}} = O \left(\gamma_\ell^{1 - k/2} \right).
$$
(6.38)

By following the same reasoning as in the bounded case, we define some classes of functions on $\mathbb{R}^{dm} \times \mathbb{R}^{dm}$ and $\mathbb{R}^{dk} \times \mathbb{R}^{dk}$,

$$
\mathcal{G}'_{\ell,j} := \left\{ h^m \bar{G}^{(\ell)}_{g,h,\mathbf{t}}(\cdot, \cdot) : g \in \mathcal{F}, h \in \mathcal{H}'_{\ell,j}, \mathbf{t} \in \mathbb{R}^{dm} \right\},
$$

$$\mathcal{G}_{\ell,j}^{\prime(k)} := \left\{ h^m \left(\pi_k \bar{G}_{g,h,\mathbf{t}}^{(\ell)} \right) (\cdot, \cdot) / \left(2^k \varepsilon \gamma_\ell^{1/p} \right) : g \in \mathcal{F}, h \in \mathcal{H}_{\ell,j}', \mathbf{t} \in \mathbb{R}^{dm} \right\}.$$

It is then easily verified that these classes are of VC-type with characteristics that are independent of ℓ and with envelope functions $\widetilde{F}$ and $\left(2^k \varepsilon \gamma_\ell^{1/p} \right)^{-1} F_k$, respectively. The function $\widetilde{F}$ is defined as in Eq. 6.4 and F_k is determined just as in the proof of Theorem 1 of Giné and Mason (2007a). Note that just as in Eqs. 6.19 and 6.21, by setting

$$\mathcal{U}_n'(j,k,\ell) := \sup_{H \in \mathcal{G}_{\ell,j}^{\prime(k)}} \left| \frac{1}{n_\ell^{k/2}} \sum_{\mathbf{i} \in I_n^k} H\left(\mathbf{X_i}, \mathbf{Y_i} \right) \right|, \quad n_{\ell-1} < n \le n_\ell,$$

we see that for all $k = 2, \dots, m$ and $n_{\ell-1} < n \le n_\ell$,

$$\mathcal{U}_n'(j,k,\ell) \le n_\ell^{k/2} \left\| U_n^{(k)} \left(\pi_k G \right) \right\|_{\mathcal{G}_{\varepsilon,j}^{(k)}}.$$

Consequently, applying Theorem 7.2 with $c = 1/2$ and $r = 2$ gives us precisely (6.20) with $\mathcal{U}_n(j,k,\ell)$ and $\mathcal{U}_{n_\ell}(j,k,\ell)$ replaced by $\mathcal{U}_n'(j,k,\ell)$ and $\mathcal{U}_{n_\ell}'(j,k,\ell)$, respectively. Therefore, the same methodology as in the bounded case will be applied. Note also that, as held for all the functions in $\mathcal{G}_{\ell,j}^{(k)}$, the functions in $\mathcal{G}_{\ell,j}^{\prime(k)}$ are bounded by 1 and have second moments that can be bounded by $h^{m(d_{\mathrm{vol}}-\varepsilon)} D_{m,k}$ for a suitable $D_{m,k}$ (by arguing as in Eqs. 6.33 and 6.22). Hence, the expression in Eq. 6.22 is also satisfied for functions in $\mathcal{G}_{\ell,j}^{\prime(k)}$, that is,

$$\sup_{H \in \mathcal{G}_{\ell,j}^{\prime(k)}} \mathbb{E} H^2(\mathbf{X}, \mathbf{Y}) \le D_{m,k} h_{\ell,j}^{\prime m(d_{\mathrm{vol}}-\varepsilon)} =: \sigma_{\ell,j}'^2.$$

Thus, all the conditions for Theorems 7.4 and 7.6 are satisfied so that, after some obvious identifications and modifications, the second part of the proof of Theorem 2.5 (and Eq. 6.26 in particular) gives us, for some $C_{7,k} > 0$, all $j = 1, \dots, L(\ell)$ and any $\rho > 2$,

$$\mathbb{P} \left\{ \max_{n_{\ell-1} < n \le n_\ell} \mathcal{U}_n'(j,k,\ell) > 2\rho^{k/2} y'^* \right\} \le \sqrt{h_{\ell,j}^{\prime m(d_{\mathrm{vol}}-\varepsilon)}} \left(\log n_\ell \right)^{-\rho C_{7,k}}, \quad (6.39)$$

with $y'^* = C_{1,k}' \lambda_{j,k}'(\ell)$ for some $C_{1,k}' > 0$ and where $\lambda_{j,k}'(\ell)$ is defined as in Eq. 6.24 with $h_{\ell,j}$ replaced by $h_{\ell,j}'$, that is,

$$\lambda_{j,k}'(\ell) = \left(\left| \log h_{\ell,j}' \right| \vee \log \log n_\ell \right)^{k/2}.$$

Now, to finish the proof of Eq. 6.38, note that, similarly to Eq. 6.27, for some $C_k > 0$, for $n_{\ell-1} < n \leq n_\ell$

$$n_\ell^{k/2} \left\| U_n^{(k)} (\pi_k G) \right\|_{\mathcal{G}'_{\ell,j}} \leq 2^k C_k \varepsilon \gamma_\ell^{1/p} \mathcal{U}'_n(j, k, \ell).$$

This gives that for some $c_k > 0$,

$$\max_{n_{\ell-1} < n \leq n_\ell} A'_{n,\ell,k}$$

$$:= \max_{n_{\ell-1} < n \leq n_\ell} \sup_{a'_n \leq h \leq b_0} \sup_{g \in \mathcal{F}} \sup_{\mathbf{t} \in \mathbb{R}^{dm}} \frac{\sqrt{n h^{m(2d - d_{\mathrm{vol}} + \varepsilon)}} \left| U_n^{(k)} \left(\pi_k \bar{G}_{g,h,\mathbf{t}}^{(\ell)} \right) \right|}{\sqrt{(|\log h| \vee \log\log n)^k}}$$

$$\leq \frac{2^k c_k \varepsilon \gamma_\ell^{1/p}}{\sqrt{n_\ell^{k-1}}} \max_{n_{\ell-1} < n \leq n_\ell} \max_{j \leq L(\ell)} \frac{\mathcal{U}'_n(j, k, \ell)}{\lambda'_{j,k}(\ell)}$$

$$\leq \frac{2^k c_k \varepsilon \gamma_\ell^{1/p}}{\sqrt{a'^{md}_{n_\ell} n_\ell^{k-1}}} \max_{n_{\ell-1} < n \leq n_\ell} \max_{j \leq L(\ell)} \frac{\mathcal{U}'_n(j, k, \ell)}{\lambda'_{j,k}(\ell)}.$$

From Eq. 6.31, we now see that

$$\gamma_\ell^{2/p} / a'^{md}_{n_\ell} n_\ell^{k-1} = c^{-m} n_\ell^{2-k} / \log n_\ell.$$

By the fact that $\log n / n^{2-k}$ is monotone increasing in $n \geq 2$ whenever $k \geq 2$, so that for some constant $C_{8,k} > 0$, we infer that

$$\mathbb{P} \left\{ \max_{n_{\ell-1} < n \leq n_\ell} \sqrt{\frac{\log n}{n^{2-k}}} A'_{n,\ell,k} > C_{8,k} \right\}$$

$$\leq \mathbb{P} \left\{ \max_{n_{\ell-1} < n \leq n_\ell} \max_{1 \leq j \leq L(\ell)} \frac{U'_n(j, k, \ell)}{\lambda'_{j,k}(\ell)} > \frac{C_{8,k}}{2^k c_k \varepsilon \gamma_\ell^{1/p}} \sqrt{\frac{n_\ell^{2-k} a'^m_{n_\ell} n_\ell^{k-1}}{\log n_\ell}} \right\}$$

$$\leq \sum_{j=1}^{L(\ell)} \mathbb{P} \left\{ \max_{n_{\ell-1} < n \leq n_\ell} \mathcal{U}'_n(j, k, \ell) > \frac{C_{8,k} c^{m/2}}{2^k c_k \varepsilon} \lambda'_{j,k}(\ell) \right\}.$$

Therefore, by choosing $C_{8,k} > 2^{k+1} c^{-m/2} \varepsilon c_k C'_{1,k} ((2+\delta)/C_{7,k})^{k/2}$ and noting that by definition $L(\ell) \leq 2 \log n_\ell$ and $h'_{\ell,j} < 2$ for all $j = 1, \ldots, L(\ell)$, we can infer from Eq. 6.39 with

$$\rho = (2 + \delta)/C_{7,k},$$

that

$$\mathbb{P}\left\{\max_{n_{\ell-1}<n\leq n_\ell}\sqrt{\frac{\log n}{n^{2-k}}}A'_{n,\ell,k}>C_{8,k}\right\}$$

$$\leq\sum_{j=1}^{L(\ell)}\mathbb{P}\left\{\max_{n_{\ell-1}<n\leq n_\ell}\mathcal{U}'_n(j,k,\ell)>2\left(\frac{2+\delta}{C_{7,k}}\right)^{k/2}C'_{1,k}\lambda'_{j,k}(\ell)\right\}$$

$$=\sum_{j=1}^{L(\ell)}\mathbb{P}\left\{\max_{n_{\ell-1}<n\leq n_\ell}\mathcal{U}'_n(j,k,\ell)>2\left(\frac{2+\delta}{C_{7,k}}\right)^{k/2}y'^*\right\}$$

$$\leq L(\ell)\sqrt{h_{\ell,j}'^m}\left(\log n_\ell\right)^{-\rho C_{7,k}}$$

$$\leq 2\sqrt{2^m}\left(\log n_\ell\right)^{-(1+\delta)}.$$

This immediately implies, via Borel-Cantelli, that for all $k = 2, \ldots, m$ and $\ell \geq 1$,

$$\max_{n_{\ell-1}<n\leq n_\ell}\sup_{a'_n\leq h\leq b_0}\sup_{g\in\mathcal{F}}\sup_{\mathbf{t}\in\mathbb{R}^{dm}}\frac{\sqrt{nh^{m(2d-d_{\mathrm{vol}}+\varepsilon)}}\left|U_n^{(k)}\left(\pi_k\bar{G}_{g,h,\mathbf{t}}^{(\ell)}\right)\right|}{\sqrt{(|\log h|\vee\log\log n)^k}}$$

$$=O\left(\sqrt{\frac{n_\ell^{2-k}}{\log n_\ell}}\right)$$

a.s., which obviously implies Eq. 6.38. Finally, recalling the Hoeffding decomposition (6.1), this implies, together with Eq. 6.37, that for some $C'' > 0$ with probability 1,

$$\limsup_{\ell\to\infty}\max_{n_{\ell-1}<n\leq n_\ell}\sup_{a'_n\leq h\leq b_0}\sup_{g\in\mathcal{F}}\sup_{\mathbf{t}\in\mathbb{R}^{dm}}$$

$$\frac{\sqrt{nh^{m(2d-d_{\mathrm{vol}}+\varepsilon)}}\left|U_n^{(m)}\left(\bar{G}_{g,h,\mathbf{t}}^{(\ell)}\right)-\mathbb{E}U_n^{(m)}\left(\bar{G}_{g,h,\mathbf{t}}^{(\ell)}\right)\right|}{\sqrt{|\log h|\vee\log\log n}}\leq C''. \qquad (6.40)$$

6.6. Remainder Part Consider now the remainder process $\widetilde{u}_n^{(\ell)}(g,h,\mathbf{t})$ based on the unbounded (symmetric) U-kernel given by

$$\widetilde{G}_{g,h,\mathbf{t}}^{(\ell)}(\mathbf{x},\mathbf{y}):=\bar{G}_{g,h,\mathbf{t}}(\mathbf{x},\mathbf{y})\mathbb{1}\left\{\widetilde{F}(\mathbf{y})>\varepsilon\gamma_\ell^{1/p}\right\},$$

where we defined γ_ℓ as in Eq. 6.30. We shall show that this U-process is asymptotically negligible at the rate given in Theorem 2.6. More precisely, we shall prove that as $\ell \to \infty$,

$$\max_{n_{\ell-1} < n \le n_\ell, \alpha' < h \le b_0} \sup_{g \in \mathcal{F}} \sup_{\mathbf{t} \in \mathbb{R}^{dm}} \frac{\sqrt{nh^{m(2d-d_{\mathrm{vol}}+\varepsilon)}} \left| U_n^{(m)} \left(\widetilde{G}_{g,h,\mathbf{t}}^{(\ell)} \right) - \mathbb{E} U_n^{(m)} \left(\widetilde{G}_{g,h,\mathbf{t}}^{(\ell)} \right) \right|}{\sqrt{|\log h| \vee \log \log n}} = o(1) \quad a.s. \tag{6.41}$$

Recall that for all $g \in \mathcal{F}, h \in [a'_n, b_0]$ and $\mathbf{t}, \mathbf{x} \in \mathbb{R}^m$, $\widetilde{F}(\mathbf{y}) \ge h^m \left| \bar{G}_{g,h,\mathbf{t}}(\mathbf{x}, \mathbf{y}) \right|$, so from the symmetry of $\widetilde{F}$, it holds that

$$\left| U_n^{(m)} \left(\widetilde{G}_{g,h,\mathbf{t}}^{(\ell)} \right) \right| \le h^{-dm} U_n^{(m)} \left(\widetilde{F} \cdot \mathbb{1} \left\{ \widetilde{F} > \varepsilon \gamma_\ell^{1/p} \right\} \right),$$

where $U_n^{(m)} \left(\widetilde{F} \cdot \mathbb{1} \left\{ \widetilde{F} > \varepsilon \gamma_\ell^{1/p} \right\} \right)$ is a U-statistic based on the positive and symmetric kernel $\mathbf{y} \to \widetilde{F}(\mathbf{y}) \mathbb{1} \left\{ \widetilde{F}(\mathbf{y}) > \varepsilon \gamma_\ell^{1/p} \right\}$. Recalling that $a'^{md}_n = c^{md} (\log n/n)^{1-2/p}$, we obtain easily that for all $g \in \mathcal{F}, h \in [a'_n, b_0], \mathbf{t} \in \mathbb{R}^m$ and some $C > 0$

$$\max_{n_{\ell-1} < n \le n_\ell} \frac{\sqrt{nh^{m(2d-d_{\mathrm{vol}}+\varepsilon)}} \left| U_n^{(m)} \left(\widetilde{G}_{g,h,\mathbf{t}}^{(\ell)} \right) \right|}{\sqrt{|\log h| \vee \log \log n}} \le \frac{\sqrt{n_\ell a_{n_\ell}^{m(-d_{\mathrm{vol}}+\varepsilon)}} U_{n_\ell}^{(m)} \left(\widetilde{F} \cdot \mathbb{1} \left\{ \widetilde{F} > \varepsilon \gamma_\ell^{1/p} \right\} \right)}{\sqrt{\left(\left| \log a'_{n_\ell} \right| \vee \log \log n_\ell \right)}}$$

$$\le C \gamma_\ell^{1-1/p} a_{n_\ell}^{\frac{m}{2}(-d_{\mathrm{vol}}+\varepsilon)} U_{n_\ell}^{(m)} \left(\widetilde{F} \cdot \mathbb{1} \left\{ \widetilde{F} > \varepsilon \gamma_\ell^{1/p} \right\} \right).$$

Arguing in the same way, since a U-statistic is an unbiased estimator of its kernel, we get that, uniformly in $g \in \mathcal{F}, h \in [a'_n, b_0]$ and $\mathbf{t} \in \mathbb{R}^m$,

$$\max_{n_{\ell-1} < n \le n_\ell} \frac{\sqrt{nh^{m(2d-d_{\mathrm{vol}}+\varepsilon)}} \left| \mathbb{E} U_n^{(m)} \left(\widetilde{G}_{g,h,t}^{(\ell)} \right) \right|}{\sqrt{|\log h| \vee \log \log n}}$$

$$\le C \gamma_\ell^{1-1/p} a_{n_\ell}^{\frac{m}{2}(-d_{\mathrm{vol}}+\varepsilon)} \mathbb{E} U_{n_\ell}^{(m)} \left(\widetilde{F} \cdot \mathbb{1} \left\{ \widetilde{F} > \varepsilon \gamma_\ell^{1/p} \right\} \right)$$

$$\le C' \mathbb{E} \left[\widetilde{F}^p(\mathbf{Y}) \mathbb{1} \left\{ \widetilde{F}(\mathbf{Y}) > \varepsilon \gamma_\ell^{1/p} \right\} \right]. \tag{6.42}$$

From Eq. 6.42, we see that as $\ell \to \infty$,

$$\max_{1 < n \le n_\ell} \sup_{a'_n \le h \le b_0} \sup_{g \in \mathcal{F}} \sup_{\mathbf{t} \in \mathbb{R}^{dm}} \frac{\sqrt{nh^{m(2d-d_{\mathrm{vol}}+\varepsilon)}} \left| \mathbb{E} U_n^{(m)} \left(\widetilde{G}_{g,h,\mathbf{t}}^{(\ell)} \right) \right|}{\sqrt{|\log h| \vee \log \log n}} = o(1). \tag{6.43}$$

Thus, to finish the proof of Eq. 6.41, it suffices to show that

$$U_{n_\ell}^{(m)}\left(\widetilde{F}\cdot\mathbb{1}\left\{\widetilde{F}>\varepsilon\gamma_\ell^{1/p}\right\}\right)=o\left(\gamma_\ell^{1/p-1}\right)\qquad\text{a.s.}\tag{6.44}$$

First, note that from Chebyshev's inequality and a well-known inequality for the variance of a U-statistic (see Theorem 5.2 of Hoeffding (1948)), we get, for any $\delta>0$,

$$\begin{aligned}\mathbb{P}&\left\{\left|U_{n_\ell}^{(m)}\left(\widetilde{F}\cdot\mathbb{1}\left\{\widetilde{F}>\varepsilon\gamma_\ell^{1/p}\right\}\right)-\mathbb{E}U_{n_\ell}^{(m)}\left(\widetilde{F}\cdot\mathbb{1}\left\{\widetilde{F}>\varepsilon\gamma_\ell^{1/p}\right\}\right)\right|\right.\\&\left.\qquad>\delta\gamma_\ell^{-(1-1/p)}\right\}\\&\le\delta^{-2}\gamma_\ell^{2-2/p}\,\mathrm{Var}\left(U_{n\ell}^{(m)}\left(\widetilde{F}\cdot\mathbb{1}\left\{\widetilde{F}>\varepsilon\gamma_\ell^{1/p}\right\}\right)\right)\\&\le m\delta^{-2}\frac{n_\ell^{1-2/p}}{(\log n_\ell)^{2-2/p}}\mathbb{E}\left[\widetilde{F}^2(\mathbf{Y})\mathbb{1}\left\{\widetilde{F}(\mathbf{Y})>\varepsilon\gamma_\ell^{1/p}\right\}\right].\end{aligned}\tag{6.45}$$

Next, in order to establish the finite convergence of the series of the above probabilities, we split the indicator function $\mathbb{1}\left\{\widetilde{F}(\mathbf{Y})>\varepsilon\gamma_\ell^{1/p}\right\}$ into two distinct parts determined by whether $\widetilde{F}(\mathbf{Y})>n_\ell^{1/p}$ or $\varepsilon\gamma_\ell^{1/p}<\widetilde{F}(\mathbf{Y})\le n_\ell^{1/p}$, and consider the corresponding second moments in Eq. 6.45 separately. In the first case, note that, from Eqs. 2.10 and 6.4, $\mathbb{E}\widetilde{F}^p(\mathbf{Y})\le\mu_p\kappa^{pm}(m!)^p$ and observe that since $p>2$ and $n_\ell=2^\ell$,

$$\begin{aligned}\sum_{\ell=1}^\infty&\frac{n_\ell^{1-2/p}}{(\log n_\ell)^{2-2/p}}\mathbb{E}\left[\widetilde{F}^2(\mathbf{Y})\mathbb{1}\left\{\widetilde{F}(\mathbf{Y})>n_\ell^{1/p}\right\}\right]\le\mathbb{E}\left[\widetilde{F}^p(\mathbf{Y})\right]\\&\times\sum_{\ell=1}^\infty(\log n_\ell)^{-(2-2/p)}<\infty.\end{aligned}$$

To handle the second case, we shall need the following fact from Einmahl and Mason (2000).

FACT 6.1. *Let $(c_n)_{n\ge1}$ be a sequence of positive constants such that $c_n/n^{1/s}\nearrow\infty$ for some $s>0$ and let Z be a random variable satisfying*

$$\sum_{n=1}^\infty\mathbb{P}\left\{|Z|>c_n\right\}<\infty.$$

We then have, for any $q>s$,

$$\sum_{k=1}^\infty2^k\mathbb{E}\left[|Z|^q\mathbb{1}\left\{|Z|\le c_{2^k}\right\}\right]/(c_{2^k})^q<\infty.$$

Notice that for any $p < r \leq 2p$,

$$\sum_{\ell=1}^{\infty} \frac{n_\ell^{1-2/p}}{(\log n_\ell)^{2-2/p}} \mathbb{E}\left[\widetilde{F}^2(\mathbf{Y})\mathbb{1}\left\{\varepsilon\gamma_\ell^{1/p} < \widetilde{F}(\mathbf{Y}) \leq n_\ell^{1/p}\right\}\right]$$

$$\leq \quad \varepsilon^{r-2} \sum_{\ell=1}^{\infty} \frac{n_\ell \mathbb{E}\left[\widetilde{F}^r(\mathbf{Y})\mathbb{1}\left\{\widetilde{F}(\mathbf{Y}) \leq n_\ell^{1/p}\right\}\right]}{(\log n_\ell)^{2-r/p}\, n_\ell^{r/p}}$$

$$\leq \quad \varepsilon^{r-2} \sum_{\ell=1}^{\infty} \frac{n_\ell \mathbb{E}\left[\widetilde{F}^r(\mathbf{Y})\mathbb{1}\left\{\widetilde{F}(\mathbf{Y}) \leq n_\ell^{1/p}\right\}\right]}{n_\ell^{r/p}}.$$

Now, set $Z = \widetilde{F}(\mathbf{Y})$, $c_n = n^{1/p}$ and $q = r$ in Fact 6.1 and note that $c_n/n^{1/s} \nearrow \infty$ for any s such that $q = r > s > p$. Since $q = r > s$, we can conclude from Fact 6.1 that this last bound is finite. Finally, note that the bound leading to Eq. 6.45 implies that

$$\gamma_\ell^{1-1/p}\mathbb{E}U_{n\ell}^{(m)}\left(\widetilde{F}\cdot\mathbb{1}\left\{\widetilde{F} > \varepsilon\gamma_\ell^{1/p}\right\}\right) = o(1).$$

Consequently, the above results, together with Eq. 6.45, imply via Borel-Cantelli and the arbitrary choice of $\delta > 0$ that Eq. 6.44 holds, which, when combined with Eqs. 6.43 and 6.45, completes the proof of Eq. 6.41. This also completes the proof of Theorem 2.6 since we have already established the result in Eq. 6.40.

  6.7. *Proof of Theorem 2.7* Theorem 2.7 is essentially a consequence of Theorem 7.7, details are similar to the proof of Dony and Mason (2008) and therefore omitted.

  6.8. *Proof of Corollary 2.8* We now turn to the proof of Corollary 2.8. We observe the following standard inequalities

$$|\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h) - \widehat{\mathbb{E}}\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h)| = \left|\frac{U_n(\varphi, h, \mathbf{t})}{U_n(1, h, \mathbf{t})} - \frac{\mathbb{E}U_n(\varphi, h, \mathbf{t})}{\mathbb{E}U_n(1, h, \mathbf{t})}\right|$$

$$\leq \frac{|U_n(\varphi, h, \mathbf{t}) - \mathbb{E}U_n(\varphi, h, \mathbf{t})|}{|U_n(1, h, \mathbf{t})|}$$

$$+ \frac{|\mathbb{E}U_n(\varphi, h, \mathbf{t})| \cdot |U_n(1, h, \mathbf{t}) - \mathbb{E}U_n(1, h, \mathbf{t})|}{|U_n(1, h, \mathbf{t})| \cdot |\mathbb{E}U_n(1, h, \mathbf{t})|}$$

$$=: (\mathrm{I}) + (\mathbb{I}).$$

We can infer from Theorem 7.7 that

$$\sup_{a_n \leq h < b_n} \sup_{\mathbf{t} \in \mathbf{I}^m} \left|\mathbb{E}U_n(1, h, \mathbf{t}) - \widetilde{f}(\mathbf{t})\right| \to 0. \tag{6.46}$$

Then, from Theorem 2.5, Eq. 6.46 and $f_{\mathbf{X}}(\cdot)$ bounded away from zero on $\mathbf{J}$, we get, for some $\mathbf{x}i_1, \mathbf{x}i_2 > 0$ and c large enough in $a_n = c(\log n/n)^{1/dm}$,

$$\liminf_{n\to\infty} \sup_{a_n \leq h < b_n} \sup_{\mathbf{t}\in\mathbf{I}^m} |U_n(1, h, \mathbf{t})| = \mathbf{x}i_1 > 0 \quad \text{a.s.},$$

and, for n large enough,

$$\sup_{a_n \leq h < b_n} \sup_{\mathbf{t}\in\mathbf{I}^m} |\mathbb{E}U_n(1, h, \mathbf{t})| = \mathbf{x}i_2 > 0.$$

Further, for a_n' equalling either ℓ_n or a_n, we readily obtain from the assumptions (2.8) or (2.10) on the envelope function that

$$\sup_{a_n' \leq h < b_n} \sup_{\varphi\in\mathscr{F}_q} \sup_{\mathbf{t}\in\mathbf{I}^m} |\mathbb{E}U_n(\varphi, h, \mathbf{t})| = O(1).$$

Hence, we can now use Theorem 2.5 to handle ($\mathbb{I}$), while for (I), depending on whether the class $\mathscr{F}_q$ satisfies Eq. 2.8 or 2.10, we apply Theorem 2.5 or Theorem 2.6, respectively. Taking everything together, we conclude that for c large enough and some $\mathfrak{C}_3 > 0$, with probability 1,

$$\limsup_{n\to\infty} \sup_{a_n' \leq h < b_n} \sup_{\varphi\in\mathscr{F}_q} \sup_{\mathbf{t}\in\mathbf{I}^m} \frac{\sqrt{nl_n^{m(2d-d_{\mathrm{vol}}+\varepsilon)}} \left|\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h) - \widehat{\mathbb{E}r}_n^{(m)}(\varphi, \mathbf{t}; h)\right|}{\sqrt{|\log h| \vee (\log(2/\delta) \vee \log\log n)}}$$

$$\leq \limsup_{n\to\infty} \sup_{a_n' \leq h < b_n} \sup_{\varphi\in\mathscr{F}_q} \sup_{\mathbf{t}\in\mathbf{I}^m} \frac{\sqrt{nh^{m(2d-d_{\mathrm{vol}}+\varepsilon)}}(\mathrm{I})}{\sqrt{|\log h| \vee \log\log n}}$$

$$+ \limsup_{n\to\infty} \sup_{a_n' \leq h < b_n} \sup_{\varphi\in\mathscr{F}_q} \sup_{\mathbf{t}\in\mathbf{I}^m} \frac{\sqrt{nl_n^{m(2d-d_{\mathrm{vol}}+\varepsilon)}}(\mathbb{I})}{\sqrt{|\log l_n| \vee \log(2/\delta)}}$$

$$\leq \mathfrak{C}_3.$$

We readily obtain the assertion of the theorem by choosing appropriately δ. $\square$

6.9.   Proof of Proposition 3.2 Recalling the Definition 3.1 of

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

it is obvious that $\boldsymbol{\Phi}_\psi$ is uniformly bounded, in $(y_1, \ldots, y_m, c_1, \ldots, c_m) \in \mathbb{R}^{2m}$ and $\psi \in \mathscr{F}$, since $\mathscr{F}$ is uniformly bounded, $\psi(t) = 0$ for all $t > \tau$ and

$G(\tau) < 1$. This property, when combined with the VC property of $\mathscr{F}$, ensures that the class of function

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

verifies Eqs. F.ii, F.iii. Similarly, it can be shown that $\mathscr{F}_{\mathbf{\Phi}}$ is a pointwise measurable class of functions (F.i). Moreover, by **(A.3)** and Eq. 3.2, the class

$$\mathcal{M}_{\mathbf{\Phi}} := \{m_{\mathbf{\Phi}_{\psi}} f_{\mathbf{X}} : \psi \in \mathscr{F}\}$$

is almost surely relatively compact concerning the sup- norm topology on $\mathcal{I}_{\alpha}$. So we can apply Theorem 2.8 with $\mathbf{Y} = (Y, C)$ and $\mathbf{\Psi} = \mathbf{\Phi}_{\psi}$. The result of Proposition 3.2 is straightforward. □

LEMMA 6.2. *Under assumptions of Theorem 3.1, we have with probability one,*

$$\sup_{h \geq l_n} \sup_{\mathbf{t} \in \mathcal{I}^m} \sup_{\psi \in \mathscr{F}} \left| \breve{r}_n^{(m)*}(\psi, \mathbf{t}; h) - \breve{r}_n^{(m)}(\psi, \mathbf{t}; h) \right| = o\left(\sqrt{\frac{\log(1/h)}{nh^d}} \right) \quad as \quad n \to \infty.$$

(6.47)

6.10.   Proof of Lemma 6.2 Recall the following useful lemma.

LEMMA 6.3. *Let a_i, $i = 1, \ldots, k$, b_i $i = 1, \ldots, k$ be real number*

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

An application of the preceding lemma gives

$$\sup_{h \geq l_n} \sup_{\mathbf{t} \in \mathcal{I}^m} \sup_{\psi \in \mathscr{F}} \left| \breve{r}_n^{(m)*}(\psi, \mathbf{t}; h) - \breve{r}_n^{(m)}(\psi, \mathbf{t}; h) \right|$$

$$= \sup_{h \geq l_n} \sup_{\mathbf{t} \in \mathcal{I}^m} \sup_{\psi \in \mathscr{F}} \left| \sum_{(i_1, \ldots, i_m) \in I(m,n)} \delta_{i_1} \cdots \delta_{i_m} \psi(Z_{i_1}, \ldots, Z_{i_m}) \overline{\omega}_{n,K,h,\mathbf{i}}^{(m)}(\mathbf{t}) \right.$$

$$\left. \left(\frac{1}{(1 - G_n^*(Z_{i_1})) \cdots (1 - G_n^*(Z_{i_m}))} - \frac{1}{(1 - G^*(Z_{i_1})) \cdots (1 - G^*(Z_{i_m}))} \right) \right|$$

$$\leq \sup_{h \geq l_n} \sup_{\mathbf{t} \in \mathcal{I}^m} \sum_{i=1}^{n} |\overline{\omega}_{n,K,h,\mathbf{i}}^{(m)}(\mathbf{t})| \sup_{\mathbf{t} \in [0,\tau)^m} \sup_{\psi \in \mathscr{F}_q} |\psi(\mathbf{t})|$$

$$\times \sum_{\eta=1}^{m} \left(\frac{1}{1 - G_n^*(Z_{i_\eta})} - \frac{1}{1 - G^*(Z_{i_\eta})} \right) \prod_{j=1}^{\eta-1} \left(\frac{1}{1 - G_n^*(Z_{i_j})} \right)$$

$$\prod_{k=\eta+1}^{m} \left(\frac{1}{1 - G(Z_{i_k})} \right)$$

$$\leq \sup_{h \geq l_n} \sup_{\mathbf{t} \in \mathcal{I}^m} \sum_{i=1}^{n} |\overline{\omega}_{n,K,h,\mathbf{i}}^{(m)}(\mathbf{t})| \sup_{\mathbf{t} \in [0,\tau)^m} \sup_{\psi \in \mathscr{F}_q} |\psi(\mathbf{t})|$$

$$\times \sum_{\eta=1}^{m} \frac{\sup_{x \leq \tau} |G_n^*(x) - G(x)|}{[1 - G_n^*(\tau)][1 - G(\tau)]} \prod_{j=1}^{\eta-1} \left(\frac{1}{1 - G_n^*(Z_{i_j})} \right) \prod_{k=\eta+1}^{m} \left(\frac{1}{1 - G(Z_{i_k})} \right)$$

$$\leq C \sup_{h \geq l_n} \sup_{\mathbf{t} \in \mathcal{I}^m} \sum_{i=1}^{n} |\overline{\omega}_{n,K,h,\mathbf{i}}^{(m)}(\mathbf{t})| \sup_{\mathbf{t} \in [0,\tau)^m} \sup_{\psi \in \mathscr{F}} |\psi(\mathbf{t})| \frac{\sup_{x \leq \tau} |G_n^*(x) - G(x)|}{[1 - G_n^*(\tau)][1 - G(\tau)]}.$$

$$(6.48)$$

Since

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

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

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

the law of iterated logarithm for $G_n^*(\cdot)$ established in Földes and Rejtő (1981) ensures that

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

By combining the results of Proposition 3.2 and Lemma 6.2, the result of the Theorem 3.1 is immediate by noting that, under the conditions **(H.1-3)**, we have, for n sufficiently large,

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

Hence the proof is complete.

References

ABREVAYA, J. and JIANG, W. (2005). A nonparametric approach to measuring and testing curvature. *J. Bus. Econ. Stat.* **23**, 1–19.

ARCONES, M. A. and GINÉ, E. (1993). Limit theorems for U-processes. *Ann. Probab.* **21**, 1494–1542.

ARCONES, M. A. and GINÉ, E. (1995). On the law of the iterated logarithm for canonical U-statistics and processes. *Stoch. Process. Appl.* **58**, 217–245.

ARCONES, M. A. and WANG, Y. (2006). Some new tests for normality based on U-processes. *Stat. Probab. Lett.* **76**, 69–82.

ARCONES, M. A., CHEN, Z. and GINÉ, E. (1994). Estimators related to U-processes with applications to multivariate medians: asymptotic normality. *Ann. Stat.* **22**, 1460–1477.

BOROVKOVA, S., BURTON, R. and DEHLING, H. (2001). Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. *Trans. Am. Math. Soc.* **353**, 4261–4318.

BOROVSKIKH, Y. V. (1996). *U-statistics in Banach spaces*. VSP, Utrecht.

BOUZEBDA, S. (2012). On the strong approximation of bootstrapped empirical copula processes with applications. *Math. Methods Stat.* **21**, 153–188.

BOUZEBDA, S. and EL-HADJALI, T. (2020). Uniform convergence rate of the kernel regression estimator adaptive to intrinsic dimension in presence of censored data. *J. Nonparametr. Stat.* **32**, 864–914.

BOUZEBDA, S. and ELHATTAB, I. (2009). A strong consistency of a nonparametric estimate of entropy under random censorship. *C. R. Math. Acad. Sci. Paris* **347**, 821–826.

BOUZEBDA, S. and ELHATTAB, I. (2010). Uniform in bandwidth consistency of the kernel-type estimator of the Shannon's entropy. *C. R. Math. Acad. Sci. Paris* **348**, 317–321.

BOUZEBDA, S. and ELHATTAB, I. (2011). Uniform-in-bandwidth consistency for kernel-type estimators of Shannon's entropy. *Electron. J. Stat.* **5**, 440–459.

BOUZEBDA, S. and NEMOUCHI, B. (2019). Central limit theorems for conditional empirical and conditional U-processes of stationary mixing sequences. *Math. Methods Stat.* **28**, 169–207.

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

BOUZEBDA, S. and NEMOUCHI, B. (2022). Weak-convergence of empirical conditional processes and conditional U-processes involving functional mixing data. Stat. Inference Stoch. Process. To appear, pp 1–56.

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

BOUZEBDA, S., ELHATTAB, I. and SECK, C. T. (2018). Uniform in bandwidth consistency of nonparametric regression based on copula representation. *Statist. Probab. Lett.* **137**, 173–182.

BOUZEBDA, S., ELHATTAB, I. and NEMOUCHI, B. (2021). On the uniform-in-bandwidth consistency of the general conditional U-statistics based on the copula representation. *J. Nonparametr. Stat.* **33**, 321–358.

BRUNEL, E. and COMTE, F. (2006). Adaptive nonparametric regression estimation in presence of right censoring. *Math. Methods Stat.* **15**, 233–255.

CARBONEZ, A., GYÖRFI, L. and VAN DER MEULEN, E. C. (1995). Partitioning-estimates of a regression function under random censoring. *Stat. Decis.* **13**, 21–37.

CHEN, Y. and DATTA, S. (2019). Adjustments of multi-sample U-statistics to right censored data and confounding covariates. *Comput. Stat. Data Anal.* **135**, 1–14.

DATTA, S., BANDYOPADHYAY, D. and SATTEN, G. A. (2010). Inverse probability of censoring weighted U-statistics for right-censored data with an application to testing hypotheses. *Scand. J. Stat.* **37**, 680–700.

DE LA PEÑA, V. H. and GINÉ, E. (1999). *Decoupling. Probability and its Applications (New York).* Springer, New York. From dependence to independence, Randomly stopped processes. U-statistics and processes. Martingales and beyond.

DEHEUVELS, P. (2000). Uniform limit laws for kernel density estimators on possibly unbounded intervals. In *Recent advances in reliability theory (Bordeaux, 2000), Stat. Ind. Technol.*, pp 477–492. Birkhäuser, Boston.

DEHEUVELS, P. and MASON, D. M. (2004). General asymptotic confidence bands based on kernel-type function estimators. *Stat. Inference Stoch. Process.* **7**, 225–277.

DENKER, M. and KELLER, G. (1983). On U-statistics and v. Mises' statistics for weakly dependent processes. *Z. Wahrsch. Verw. Gebiete* **64**, 505–522.

DEVROYE, L. and LUGOSI, G. (2001). *Combinatorial methods in density estimation Springer Series in Statistics.* Springer, New York.

DONY, J. and MASON, D. M. (2008). Uniform in bandwidth consistency of conditional U-statistics. *Bernoulli* **14**, 1108–1133.

EINMAHL, U. and MASON, D. M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theor. Probab.* **13**, 1–37.

EINMAHL, U. and MASON, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Stat.* **33**, 1380–1403.

FARAHMAND, A. M., SZEPESVÁRI, C. and AUDIBERT, J. -Y. (2007). Manifold-adaptive dimension estimation. In *Proceedings of the 24th International Conference on Machine Learning, ICML '07*, pp 265–272. Association for Computing Machinery, New York.

FÖLDES, A. and REJTŐ, L. (1981). A LIL type result for the product limit estimator. *Z. Wahrsch. Verw. Gebiete* **56**, 75–86.

GHOSAL, S., SEN, A. and VAN DER VAART, A. W. (2000). Testing monotonicity of regression. *Ann. Stat.* **28**, 1054–1082.

GINÉ, E. and MASON, D. M. (2007a). Laws of the iterated logarithm for the local U-statistic process. *J. Theoret. Probab.* **20**, 457–485.

GINÉ, E. and MASON, D. M. (2007b). On local U-statistic processes and the estimation of densities of functions of several sample variables. *Ann. Stat.* **35**, 1105–1145.

GYÖRFI, L., KOHLER, M., KRZYŻAK, A. and WALK, H. (2002). *A distribution-free theory of nonparametric regression. Springer Series in Statistics.* Springer, New York.

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

HALMOS, P. R. (1946). The theory of unbiased estimation. *Ann. Math. Stat.* **17**, 34–43.

HÄRDLE, W. and MARRON, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Stat.* **13**, 1465–1481.

HAREL, M. and PURI, M. L. (1996). Conditional U-statistics for dependent random variables. *J. Multivariate Anal.* **57**, 84–100.

HEIN, M. and AUDIBERT, J. -Y. (2005). Intrinsic dimensionality estimation of submanifolds in rd. In *Proceedings of the 22nd International Conference on Machine Learning, ICML '05*, pp 289–296. Association for Computing Machinery, New York.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19**, 293–325.

HOLLANDER, M. and PROSCHAN, F. (1972). Testing whether new is better than used. *Ann. Math. Stat.* **43**, 1136–1146.

JOLY, E. and LUGOSI, G. (2016). Robust estimation of U-statistics. *Stoch. Process. Appl.* **126**, 3760–3773.

KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **53**, 457–481.

KÉGL, B. (2002). Intrinsic dimension estimation using packing numbers. In *Proceedings of the 15th International Conference on Neural Information Processing Systems, NIPS'02*, pp. 697–704. MIT Press, Cambridge.

KIM, J., SHIN, J., RINALDO, A. and WASSERMAN, L. (2018). Uniform convergence rate of the kernel density estimator adaptive to intrinsic volume dimension.

KIM, J., SHIN, J., RINALDO, A. and WASSERMAN, L. (2019). Uniform convergence rate of the kernel density estimator adaptive to intrinsic volume dimension. In *Proceedings of the 36th International Conference on Machine Learning, volume 97 of Proceedings of Machine Learning Research*, (K. Chaudhuri and R. Salakhutdinov, eds.), pp. 3398–3407. PMLR, Long Beach.

KOHLER, M., MÁTHÉ, K. and PINTÉR, M. (2002). Prediction from randomly right censored data. *J. Multivar. Anal.* **80**, 73–100.

KOROLJUK, V.S. and BOROVSKICH, Y.V. (1994). *Theory of U-statistics, volume 273 of Mathematics and Its Applications.* Kluwer Academic Publishers Group, Dordrecht. Translated from the 1989 Russian original by P. V. Malyshev and D. V. Malyshev and revised by the authors.

KOSOROK, M. R. (2008). *Introduction to empirical processes and semiparametric inference. Springer Series in Statistics.* Springer, New York.

LEE, A. J. (1990). *U-statistics, volume 110 of Statistics: Textbooks and Monographs.* Marcel Dekker, Inc., New York. Theory and practice.

LEE, S., LINTON, O. and WHANG, Y. -J. (2009). Testing for stochastic monotonicity. *Econometrica* **77**, 585–602.

LEUCHT, A. (2012). Degenerate U- and V-statistics under weak dependence: asymptotic theory and bootstrap consistency. *Bernoulli* **18**, 552–585.

LEUCHT, A. and NEUMANN, M. H. (2013). Degenerate U- and V-statistics under ergodicity: asymptotics, bootstrap and applications in statistics. *Ann. Inst. Stat. Math.* **65**, 349–386.

LEVINA, E. and BICKEL, P. J. (2004). Maximum likelihood estimation of intrinsic dimension. In *Proceedings of the 17th International Conference on Neural Information Processing Systems, NIPS'04*, pp. 777–784. MIT Press, Cambridge.

LING, N. and VIEU, P. (2018). Nonparametric modelling for functional data: selected survey and tracks for future. *Statistics* **52**, 934–949.

MAILLOT, B. and VIALLON, V. (2009). Uniform limit laws of the logarithm for nonparametric estimators of the regression function in presence of censored data. *Math. Methods Stat.* **18**, 159–184.

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

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

NADARAJA, E. A. (1964). On a regression estimate. *Teor. Verojatnost. i Primenen.* **9**, 157–159.

NOLAN, D. and POLLARD, D. (1987). *U*-processes: rates of convergence. *Ann. Stat.* **15**, 780–799.

PRAKASA RAO, B. L. S. and SEN, A. (1995). Limit distributions of conditional *U*-statistics. *J. Theor. Probab.* **8**, 261–301.

SCHICK, A., WANG, Y. and WEFELMEYER, W. (2011). Tests for normality based on density estimators of convolutions. *Stat. Probab. Lett.* **81**, 337–343.

SEN, A. (1994). Uniform strong consistency rates for conditional *U*-statistics. *Sankhyā Ser. A* **56**, 179–194.

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

SHERMAN, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica* **61**, 123–137.

SHERMAN, R. P. (1994). Maximal inequalities for degenerate *U*-processes with applications to optimization estimators. *Ann. Stat.* **22**, 439–459.

SILVERMAN, B. W. (1978). Distances on circles, toruses and spheres. *J. Appl. Probab.* **15**, 136–143.

STUTE, W. (1991). Conditional *U*-statistics. *Ann. Probab.* **19**, 812–825.

STUTE, W. (1993). Almost sure representations of the product-limit estimator for truncated data. *Ann. Stat.* **21**, 146–156.

STUTE, W. (1994a). L^p-convergence of conditional *U*-statistics. *J. Multivar. Anal.* **51**, 71–82.

STUTE, W. (1994b). Universally consistent conditional *U*-statistics. *Ann. Stat.* **22**, 460–473.

STUTE, W. (1996). Symmetrized NN-conditional *U*-statistics. In *Research developments in probability and statistics*, pp. 231–237. VSP, Utrecht.

STUTE, W. and WANG, J. -L. (1993). Multi-sample *U*-statistics for censored data. *Scand. J. Stat.* **20**, 369–374.

TALAGRAND, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.* **22**, 28–76.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak convergence and empirical processes. Springer Series in Statistics.* Springer, New York. With applications to statistics.

VON MISES, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Stat.* **18**, 309–348.

WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26**, 359–372.

YUAN, A., GIURCANU, M., LUTA, G. and TAN, M. T. (2017). U-statistics with conditional kernels for incomplete data models. *Ann. Inst. Stat. Math.* **69**, 271–302.

Appendix

THEOREM 7.1. (Kim et al., 2018) *Let $(\mathbb{R}^d, \mathbb{P})$ be a probability space and let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be i.i.d. from $\mathbb{P}$. Let $\mathcal{F}$ be a class of functions from $\mathbb{R}^d$ to $\mathbb{R}$ that is uniformly bounded VC-class with dimension ν, i.e., there exists positive numbers A, B such that, for all $f \in \mathcal{F}$, $\|f\|_\infty \leq B$, and for every probability measure $\mathbb{Q}$ on $\mathbb{R}^d$ and for every $\epsilon \in (0, B)$, the covering number $\mathcal{N}(\mathcal{F}, L_2(\mathbb{Q}), \epsilon)$ satisfies*

$$\mathcal{N}(\mathcal{F}, L_2(\mathbb{Q}), \epsilon) \leq \left(\frac{AB}{\epsilon} \right)^\nu.$$

Let $\sigma > 0$ with $\mathbb{E}_\mathbb{P} f^2 \leq \sigma^2$ for all $f \in \mathcal{F}$. Then there exists a universal constant C not depending on any parameters such that

$$\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(\mathbf{X}_i) - \mathbb{E}[f(\mathbf{X})] \right|$$

is upper bounded with probability at least $1 - \delta$,

$$\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(\mathbf{X}_i) - \mathbb{E}[f(\mathbf{X})] \right|$$
$$\leq C \left(\frac{\nu B}{n} \log \left(\frac{2AB}{\sigma} \right) + \sqrt{\frac{\nu \sigma^2}{n} \log \left(\frac{2AB}{\sigma} \right)} + \sqrt{\frac{\sigma^2 \log \left(\frac{1}{\delta} \right)}{n}} + \frac{B \log \left(\frac{1}{\delta} \right)}{n} \right).$$

THEOREM 7.2 (Theorem 4 of Giné and Mason (2007a)). *Let $X_1, X_2, \ldots$ be i.i.d. S-valued with probability law $\mathbb{P}$. Let $\mathcal{H}$ be a P-separable collection of measurable functions $f : S^k \to \mathbb{R}$ and assume that $\mathcal{H}$ is $\mathbb{P}$-canonical (which means that every f in $\mathcal{H}$ is $\mathbb{P}$-canonical). Further, assume that*

$$\mathbb{E} \left\| f(X_1, \ldots, X_k) \right\|_\mathcal{H}^r < \infty$$

for some $r > 1$ and let s be the conjugate of r. Then, with S_n Uniform in bandwidth consistency of conditional U-statistics 1131 defined as

$$S_n = \sup_{f \in \mathcal{H}} \left| \sum_{\mathbf{i} \in I_n^k} f(X_{i_1}, \ldots, X_{i_k}) \right|, \quad n \geq k,$$

we have, for all $x > 0$ and $0 < c < 1$,

$$\mathbb{P} \left\{ \max_{k \leq m \leq n} S_m > x \right\} \leq \frac{\mathbb{P} \{S_n > cx\}^{1/s} (\mathbb{E} S_n^r)^{1/r}}{x(1 - c)}.$$

THEOREM 7.3 (Proposition 1 of Einmahl and Mason (2005)). *Let $\mathcal{G}$ be a pointwise measurable class of bounded functions with envelope function G such that for some constants $C, v \geq 1$ and $0 < \sigma \leq \beta$, the following conditions hold:*

(i) $\mathbb{E} G^2(X) \leq \beta^2$;

(ii) $\mathcal{N}(\epsilon, \mathcal{G}) \leq C \epsilon^{-v}, \quad 0 < \epsilon < 1$;

(iii) $\sigma_0^2 := \sup_{g \in \mathcal{G}} \mathbb{E} g^2(X) \leq \sigma^2$

(iv) $\sup_{g \in \mathcal{G}} \|g\|_\infty \leq \frac{1}{4\sqrt{v}} \sqrt{n\sigma^2 / \log(C_1 \beta / \sigma)}$, where $C_1 = C^{1/\nu} \vee e$.

We then have, for some absolute constant A,

$$\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i g(X_i) \right\|_{\mathcal{G}} \leq A \sqrt{vn\,\sigma^2 \log(C_1 \beta / \sigma)},$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. Rademacher variables, independent of $X_1, \ldots, X_n$.

THEOREM 7.4 (Corollary 1 of Giné and Mason (2007a)). *Let $\mathcal{F}$ be a collection of measurable functions $f : S^m \to \mathbb{R}$, symmetric in their entries, with absolute values bounded by $M > 0$, and let P be any probability measure on $(S, \mathcal{S})$ (with X_i i.i.d.- P). Assume that $\mathcal{F}$ is of VC-type with envelope function $F \equiv M$ and with characteristics A and v. Then, for every $m \in \mathbb{N}, A \geq e^m\ v \geq 1$, there exist constants $C_1 := C_1(m, A, v, M)$ and $C_2 = C_2(m, A, v, M)$ such that for $k = 1, \ldots, m$*

$$n^k \mathbb{E} \left\| U_n^{(k)} (\pi_k f) \right\|_{\mathcal{F}}^2 \leq C_1^2 2^k \sigma^2 \left(\log \frac{A}{\sigma} \right)^k,$$

assuming

$$n\sigma^2 \geq C_2 \log(A/\sigma),$$

where σ^2 is any number satisfying

$$\left\| P^m f^2 \right\|_{\mathcal{F}} \leq \sigma^2 \leq M^2.$$

THEOREM 7.5. (Talagrand, 1994) *Let $\mathcal{G}$ be a pointwise measurable class of functions satisfying*

$$\|g\|_\infty \leq M < \infty, \quad g \in \mathcal{G}.$$

We then have, for all $t > 0$,

$$\mathbb{P}\left\{\max_{1 \leq m \leq n}\left\|\sqrt{m}\alpha_m\right\|_{\mathcal{G}} \geq A_1\left(\mathbb{E}\left\|\sum_{i=1}^{n}\varepsilon_i g\left(X_i\right)\right\|_{\mathcal{G}} + t\right)\right\}$$
$$\leq 2\left\{\exp\left(-\frac{A_2 t^2}{n\sigma_{\mathcal{G}}^2}\right) + \exp\left(-\frac{A_2 t}{M}\right)\right\},$$

where

$$\sigma_{\mathcal{G}}^2 = \sup_{g \in \mathcal{G}} \operatorname{Var}(g(X))$$

and A_1, A_2 are universal constants.

We now state the exponential inequality that will permit us to control the probability term in (4.6) and which is stated as Theorem 5.3.14 in de la Peña and Giné (1999).

THEOREM 7.6 (Theorem 5.3.14 of de la Peña and Giné (1999))). *Let $\mathcal{H}$ be a VC-subgraph class of uniformly bounded measurable real-valued kernels H on $(S^m, \mathcal{S}^m)$, symmetric in their entries. Then, for each $1 \leq k \leq m$, there exist constants $c_k, d_k \in]0, \infty[$ such that, for all $n \geq m$ and $t > 0$,,*

$$\left\{\left\|n^{k/2}U_n^{(k)}\left(\pi_k H\right)\right\|_{\mathcal{H}} > t\right\} \leq c_k \exp\left\{-d_k t^{2/k}\right\}.$$

THEOREM 7.7. (Dony and Mason, 2008) *Let $I = [a, b]$ be a compact interval. Suppose that $\mathcal{H}$ is a uniformly equicontinuous family of real-valued functions φ on $J = [a - \eta, b + \eta]^d$ for some $d \geq 1$ and $\eta > 0$. Further assume that K is an L_1-kernel with support in $[-B, B]^d$, with $B > 0$ satisfying $\int_{\mathbb{R}^d} K(\mathbf{u})\mathrm{d}\mathbf{u} = 1$. Then, uniformly in $\varphi \in \mathcal{H}$ and for any sequence of positive constants $b_n \to 0$,*

$$\sup_{0 < h < b_n}\sup_{\mathbf{z} \in I^d}\left|\varphi * K_h(\mathbf{z}) - \varphi(\mathbf{z})\right| \longrightarrow 0 \quad as\ n \to \infty$$

where $K_h(\mathbf{z}) = h^{-d}K(\mathbf{z}/h)$ and

$$\varphi * K_h(\mathbf{z}) := h^{-d}\int_{\mathbb{R}^d}\varphi(\mathbf{x})K\left(\frac{\mathbf{z} - \mathbf{x}}{h}\right)\mathrm{d}\mathbf{x}.$$

Salim Bouzebda and Anouar
Abdeldjaoued Ferfache
Universite de Technologie de
Compiegne, LMAC (Laboratory of
Applied Mathematics of Compiegne),
CS 60 319 - 60 203 Compiegne Cedex,
France
E-mail: salim.bouzebda@utc.fr
E-mail: ferfache.abdeldjaoued@gmail.com

Thouria El-hadjali
Departement de Mathematiques,
Universite Freres Mentouri,
Constantine 1, Algeria
E-mail: t.elhadjali@centre-univ-mila.dz