



HAL
open science

Contributions to the theory of U-statistics and empirical processes

Amel Nezzal

► **To cite this version:**

Amel Nezzal. Contributions to the theory of U-statistics and empirical processes. Mathematics [math]. Université de Technologie de Compiègne, 2022. English. NNT : 2022COMP2704 . tel-04640072

HAL Id: tel-04640072

<https://theses.hal.science/tel-04640072v1>

Submitted on 9 Jul 2024

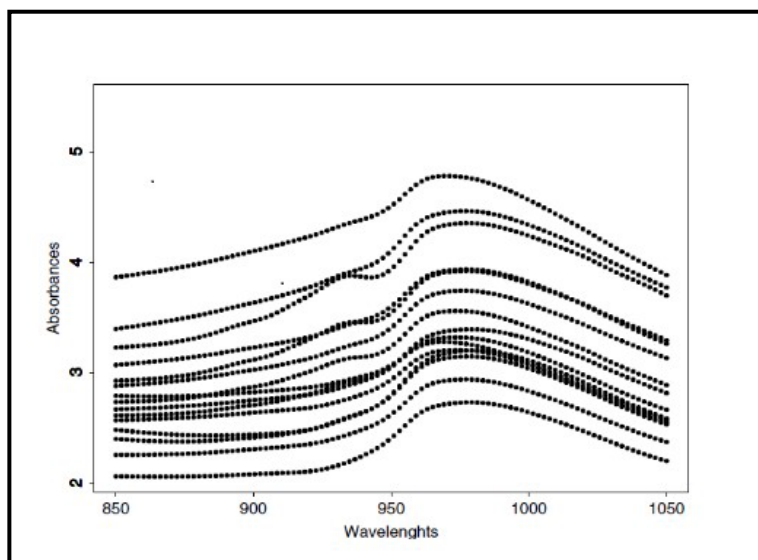
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Par Amel NEZZAL

*Contributions à la théorie des U-statistiques
et processus empiriques*

Thèse présentée
pour l'obtention du grade
de Docteur de l'UTC



Soutenue le 28 octobre 2022

Spécialité : Mathématiques Appliquées : Laboratoire de
Mathématiques Appliquées de Compiègne (Unité de
recherche EA-2222)

D2704

Alliance Sorbonne Université

THÈSE

pour obtenir le grade de docteur délivré par

Université de Technologie de Compiègne

Spécialité : **Mathématiques Appliquées**

présentée et soutenue publiquement par

Amel NEZZAL

le 28/10/2022

Contributions à la théorie des U-statistiques et processus empiriques

Directeur de thèse : **Salim BOUZEBDA**

Jury

M. Mokhtar Z. Alaya,	Maître de Conférences	Examineur
M. Samir Ben Hariz,	Maître de Conférences	Examineur
M. Salim Bouzebda,	Professeur des universités	Directeur
Mme. Sophie Dabo-Niang,	Professeur des universités	Rapporteuse
M. Amor Keziou,	Maître de Conférences	Rapporteur
M. Nikolaos Limnios,	Professeur des universités	Président du jury

UTC

Laboratoire de Mathématiques Appliquées de Compiègne (LMAC)

CS 60319 - 57 avenue de Landshut Compiègne Cedex, 60203 France

Publications

- [P.1] *Uniform consistency and uniform in number of neighbors consistency for nonparametric regression estimates and conditional U-statistics involving functional data.* **Japanese Journal of Statistics and Data Science.** **5(2022)**, no. 2, pp. 431—533. Joint work with S. Bouzebda.
- [P.2] *Asymptotic properties of conditional U-statistics using delta sequences for functional data.* **Mathematics.** **(11)(2023)** pp. 1–40. Joint work with S. Bouzebda and T. Zari.
- [P.1] *Asymptotic properties of conditional U-statistics using delta sequences.* **Communications in Statistics-Theory and Methods (2023)** pp. 1–56. Joint work with S. Bouzebda. Submitted

Conferences

- [C.1] *Uniform consistency and uniform in number of neighbors consistency for nonparametric regression estimates and conditional U-statistics involving functional data.* 14th International Conference of the ERCIM WG on Computational and Methodological Statistics
15th International Conference on Computational and Financial Econometrics King's College was held in London, UK 18-20 December 2021.

Résumé de la thèse

Dans cette thèse, nous nous intéressons principalement à l'estimation des modèles de régression non paramétrique. Plus précisément, nous nous intéressons à une classe d'estimateurs des U-statistiques conditionnelles. Les U-statistiques conditionnelles peuvent être considérées comme une généralisation de l'estimateur de Nadaraya-Watson. Ce dernier utilise les méthodes à noyau pour "moyenner" les valeurs des variables réponses. Stute généralise l'estimateur de Nadaraya-Watson en remplaçant les moyennes pondérées simples au numérateur et au dénominateur par des U-statistiques. Ensuite, en utilisant une collection de variables aléatoires prédictives, il prédit l'espérance conditionnelle de la fonction noyau des U-statistiques. Cette généralisation est prospère et influente dans la statistique mathématique en raison de son utilité scientifique exceptionnelle et de sa fascinante complexité théorique. Cependant, comme toute autre technique d'estimation par noyau, la question du choix de la fenêtre de lissage appropriée pour équilibrer le compromis variance-biais, est un sujet qui reste mal abordé dans la littérature sur les U-statistiques conditionnelles lorsque les variables explicatives sont fonctionnelles. Dans la première partie de la thèse, nous introduisons l'estimateur k des plus proches voisins des U-statistiques conditionnelles dépendant d'une covariable de dimension infinie. La convergence uniforme en termes de nombre de voisinages (UINN) pour l'estimateur proposé est présentée. Un tel résultat permet de varier le nombre de voisinages dans une plage complète pour laquelle l'estimateur reste consistant. Par conséquent, ce résultat représente une ligne directrice intéressante dans la pratique pour sélectionner le nombre de voisinages optimal dans l'analyse des données fonctionnelles non paramétriques. De plus, nos résultats sont uniformes sur une classe de fonctions convenablement choisie \mathcal{F} , dans les deux cas bornée et non bornée, satisfaisant des conditions de moment et certaines conditions générales sur l'entropie. En tant que sous-produit de nos preuves, nous énonçons des résultats de convergence pour le k -NN des U-statistiques conditionnelles, dans le cadre de la censure aléatoire, uniformes en nombre de voisins. La deuxième partie de la thèse traite un cadre général d'estimation non paramétrique incluant l'estimateur de Stute comme cas particulier. La classe des « estimateurs de delta séquence » est introduite et traitée dans ce travail. Cette classe comprend également les séries orthogonales et les méthodes d'histogramme. Nous étendons partiellement ces résultats au cadre des données fonctionnelles. La majeure partie de la thèse est motivée par les problèmes d'apprentissage, y compris parmi beaucoup d'autres, les problèmes de discrimination, l'apprentissage métrique et le "classement multipartite".

Mots-clés: Analyse des données fonctionnelles; Classe de fonctions de type VC; Consistance uniforme; Convergence Uniforme presque complète; Données fonctionnelles; Entropie métrique; Espace fonctionnel; Estimateurs à noyaux; Estimateurs des k plus proche voisins; Estimation par les séquences delta; Paramètre de lissage; Probabilité de petites boules, Processus empiriques ; U-statistiques; U-statistiques conditionnelles; U-processus.

Thesis abstract

In this dissertation, we are interested in nonparametric regression estimation models. More precisely, we are concerned with a class of conditional U-statistics estimators. Conditional U-statistics can be viewed as a generalization of the Nadaraya-Watson estimator. The latter uses a smoothing kernel function to “average” response variable values within a predictor range. Stute generalizes the Nadaraya-Watson estimator first by replacing simple weighted averages in the numerator and denominator with U-statistics. Then, using a collection of predictor random variables, he predicts the conditional expectation of the U-statistic kernel function. This generalization is prosperous and influential in mathematical statistics due to its outstanding scientific utility and fascinating theoretical complexity. However, like any other kernel estimation technique, the question of choosing a suitable bandwidth to balance the variance-bias trade-off is a subject that remains insufficiently addressed in the literature on conditional U-statistics when explanatory variables are functional. In the first part, we introduce the k nearest neighborhoods estimator of the conditional U-statistics depending on an infinite-dimensional covariate. A sharp uniform in the number of neighborhoods (UINN) limit law for the proposed estimator is presented. Such a result allows the NN to vary within a complete range for which the estimator is consistent. Consequently, it represents an interesting guideline in practice to select the optimal NN in nonparametric functional data analysis. In addition, uniform consistency is also established over $\varphi \in \mathcal{F}$ for a suitably restricted class \mathcal{F} , in both cases bounded and unbounded, satisfying some moment conditions and some mild conditions on the model. As a by-product of our proofs, we state consistency results for the k -NN conditional U-statistics, under the random censoring, are uniform in the number of neighbors. The second part of the thesis deals with a general nonparametric statistical curve estimation setting, including the Stute estimator as a particular case. The class of “delta sequence estimators” is defined and treated here. This class also includes the orthogonal series and histogram methods. We partially extend these results to the setting of the functional data. The major part of the thesis is motivated by machine learning problems, including, among many others, the discrimination problems, the metric learning, and the multipartite ranking.

Key words Functional data analysis; VC-Class of functions; Uniform consistency; Almost complete Uniform Convergence; Functional data; Metric entropy; Functional space; Kernel estimators; k nearest neighbor estimators; Data-driven estimator; Delta sequences; Smoothing parameter; Small-ball probability; Empirical processes; U-statistics; Conditional U-statistics; U-processes.

Remerciements

This thesis is a result of research carried out at the Laboratory of Applied Mathematics of Compiègne, LMAC University of Technology of Compiègne.

First and foremost, I would like to express my deepest gratitude to my supervisor **Prof. Salim BOUZEEDA**, for his patience, enthusiastic encouragement, and valuable critique of this research work. I am indebted to him for all his effort, devotion and support; without him, I could not have undertaken this journey.

I would like to thank also the members of my thesis committee, starting with the referees: I would like to thank **Prof. Sophie Dabo-Niang** and **Dr. Amor KEZIOU** not only for the time they devoted to reading my manuscript but for agreeing to evaluate my thesis results and for their insightful comments and encouragement. I am indebted to the rest of the committee **Dr. Mokhtar Z. Alaya**, **Samir BEN HARIZ** and the president **Prof. Nikolaos LIMNIOS** to accept taking part in this committee thesis.

I take this opportunity to express gratitude to all the members of the lab LMAC, especially, **Dr. Faten Jelassi**, whom I have worked with as a temporary teaching assistant, for her encouragement and great help.

I gratefully acknowledge the Algerian government for giving me this excellent opportunity to prepare my thesis in France. Furthermore, I would like to thank them for the funding I received during my thesis. I am particularly grateful to all of my professors at the University of Biskra, Algeria, for their teaching and guidance.

I owe special thanks to my friends **Boutheina**, **Inass**, **Thouria** and my best one **NourElhouda**. I would like to express my thanks for their support, encouragement, and love.

I would like to extend my sincere thanks to my dear parents, **Mohamed NEZZAL** and **Samira HOUHOUNE** for their unconditional love, understanding, prayers, and continuing support during my studies. I would like to say a heartfelt thank you for everything they have done for me. A special thanks to my beloved sisters **Hala** and **Maroua** for their deep love and encouragement.

Amel NEZZAL

Contents

Contents	ix
List of Figures	1
1 Introduction	3
1.1 Motivation	5
1.2 Organization of the dissertation	11
1.3 References	14
2 Mathematical background	21
2.1 Mathematical Background	21
2.2 References	33
3 k-NN functional conditional U-statistics	35
3.1 Introduction	36
3.2 The functional framework	40
3.3 UIB and UINN consistency for functional regression	47
3.4 UIB and UINN consistency of functional conditional U -statistics	49
3.5 Extension to the censored case	55
3.6 Some potential applications	58
3.7 Examples of kernel for the U -statistics	62
3.8 The bandwidth selection criterion	64
3.9 Mathematical developments	66
3.10 References	117
4 Conditional U-statistics using delta sequences	127
4.1 Introduction	128
4.2 Preliminaries and estimation procedure	131
4.3 Main results	134
4.4 Conditional U-statistics for censored data	141
4.5 Applications	144
4.6 The bandwidth selection criterion	154
4.7 Mathematical development	155

4.8	Appendix	176
4.9	References	177
5	Functional conditional U-statistics by delta-sequences	187
5.1	Introduction	188
5.2	Preliminaries and estimation procedure	193
5.3	Some asymptotic results	199
5.4	Conditional U -statistics for censored data	203
5.5	Applications	206
5.6	Mathematical development	211
5.7	References	227
6	Conclusions and perspectives	237
6.1	Concluding remarks : Chapter 3	237
6.2	Concluding remarks : Chapter 4	238
6.3	Concluding remarks : Chapter 5	238
6.4	References	239

List of Figures

1.1	Some Units Among The Original Chemometric Data	11
-----	--	----

Chapter 1

Introduction

Statistics and data analysis are becoming increasingly relevant and essential in this era of big data to evaluate and examine the vast amounts of information now available. Therefore, the tasks of estimation and prediction are of the utmost significance. The estimation theory can be considered one of the fundamental components of mathematical statistics, which can be divided into parametric estimation and nonparametric estimation. A nonparametric procedure is typically valid regardless of the distribution of the sampled observations. Instead, it estimates some unknown function elements belonging to a certain class of functions from the observations. Therefore, estimating functional characteristics associated with the distribution of the observations, such as the distribution function, density function, or regression function, constitutes a significant part of the problems studied in the nonparametric estimation theory. Within the scope of this thesis, our primary focus will be on the investigation of nonparametric regression estimators, more specifically, conditional U-statistics. Our study aims to generalize and improve the performance of the conditional U-statistics (and/or processes) estimates.

Let Z_1, \dots, Z_n be independent identically distributed random variables. Suppose we want to estimate an unknown parameter θ from the observations, using a known function $\varphi(\cdot)$ depending on $m \leq n$ variables. If one assumes that the function $\varphi(\cdot)$ is an unbiased estimator of the parameter θ , then :

$$\theta = \mathbb{E} [\varphi (Z_1, \dots, Z_m)].$$

Assume that $\varphi(\cdot)$ is a symmetric function, we define the U-statistic based on $\varphi(\cdot)$ by :

Definition 1.0.0.1 *Let $\{Z_n\}_{n \in \mathbb{N}^*}$ be a sequence of independent identically distributed random variables, taking values in some measurable space $(\mathcal{Z}, \mathcal{A}(\mathcal{Z}))$. For $m \leq n$, we define a U statistic based on a symmetric kernel function $\varphi : \mathcal{Z}^m \rightarrow \mathbb{R}$, by :*

$$u_n^{(m)}(\varphi) := \frac{(n-m)!}{n!} \sum_{I_m^n} \varphi(Z_{i_1}, \dots, Z_{i_m}), \quad (1.0.1)$$

where, $I_m^n = \{(i_1, \dots, i_m) : 1 \leq i_j \leq n \text{ and } i_j \neq i_r \text{ if } j \neq r\}$, and the integer m is called the degree or the rank of the U-statistic (or also the rank or the degree of the kernel $\varphi(\cdot)$).

These estimators were first considered in [Halmos \[1946\]](#) in connection with unbiased statistics, from which comes the letter U (unbiased), and were formally introduced by [Hoeffding \[1948\]](#). From (1.0.1), we can observe that the U-statistic is an arithmetic average of unbiased estimators for θ and because the observations are independent then $u_n^{(m)}(\cdot)$ is an unbiased estimator of θ too, and moreover it has a smaller variance than $\varphi(\cdot)$. They often have the smallest variance among the class of unbiased estimators in θ . Generally, we use the notation $U_{\mathbb{R}}$ -statistic to design a U-statistic that the kernel takes values in the set of real numbers \mathbb{R} , U_H -statistic to design a U-statistic that the kernel takes values in a Hilbert space. More broadly, if the kernel φ belongs to some separable Banach space B , we note the U-statistic by U_B -statistic. This thesis uses the same notation U-statistic in all settings.

Since its introduction by [Hoeffding \[1948\]](#), the simple U-statistic has been an active research field in statistics. [Hoeffding \[1961\]](#), established some fundamental properties of U-statistic. [Sen \[1974\]](#) added some valuable contributions on this topic. [Gregory \[1977\]](#) got the asymptotic distribution for degenerate U-statistics with rank two. The asymptotic distribution of the U-statistic with an arbitrary rank was extended by [Janson \[1979\]](#) and [Rubin and Vitale \[1980\]](#).

[Serfling \[2009\]](#) has shown that $(nu_n^{(m)})_{n \in \mathbb{N}^*}$ converges in distribution to a random variable which is written, in terms of the eigenvalues of an operator associated with the kernel $\varphi(\cdot)$, as a weighted sum of Chi-squared independent variables. In turn, [Borovskikh \[1986\]](#) established the asymptotic distribution of the U-statistic in a Hilbert space. [de la Peña \[1992\]](#) introduced a general decoupling inequality for U-statistics. We refer to [Serfling \[2009\]](#), [Lee \[1990\]](#) and [Korolyuk and Borovskikh \[2013\]](#) books for a detailed review and major historical developments in this field. The asymptotic theory of a collection of U-statistics is also of considerable importance; that's why [Arcones and Giné \[1993\]](#) studied the convergence of a family of U-statistics built over a sequence of i.i.d r.v.'s. [Arcones and Yu \[1994\]](#) established the functional central limit theorems under absolute regular observations for U-processes. In (1992), [Giné and Zinn \[1992\]](#) obtained Hoffmann-Jorgensen type moment inequalities for completely degenerate U-statistics of an arbitrary order m . [Klass and Nowicki \[1997\]](#) also gave moment inequalities for non-negative generalized U-statistics, for order $m = 2$. In turn, [Ibragimov and Sharakhmetov \[1999\]](#) established analogs of Rosenthal inequality for non-negative and canonical U-statistics. On the other hand, exponential and obtained moment inequalities citeEvarist2000 specifically, a Bernstein-type exponential inequality for (generalized) canonical U-statistics of order 2. Somewhere else, [de la Peña and Giné \[1999\]](#) presented, for U-statistics and U-processes, the decoupling theory and its application to limit theorems. More recent works on both U-statistics and U-processes have also been established, such as [Adamczak \[2006\]](#), who generalized the results of [Giné et al. \[2000\]](#) to canonical U-statistics of arbitrary order. [Giné and Mason \[2007b\]](#) who introduced a local U-statistic process based on an estimator given by [Frees \[1994\]](#), and established its central limit theorems for diverse norms, namely, the sup and the L_p -norms (for $p \geq 1$). In another work [Giné and Mason \[2007a\]](#), they established the laws of the iterated logarithm for the same statistic (the local U-statistic process). Others investigated the fields of application of U-statistics. For example, [Cléménçon et al. \[2008\]](#) used the U-statistics in the ranking prob-

lems and clustering Cl  men  on [2014]. A natural occurrence of U-statistics also took place in learning on graphs Biau and Bleakley [2006] or in expansions of smooth statistics Robins *et al.* [2009] as components of higher-order terms. In 2015, Joly and G  bor published an article Joly and Lugosi [2016] named Robust estimation of U-statistics, in which they investigate the estimation of the mean of multivariate functions in some specific cases.

Up to 1991, only the simple U-statistics (U-processes) were studied. The first paper devoted to conditional U-statistic, in which it was introduced, belongs to Stute [1991]. From there, several authors have been interested in studying asymptotic properties of the conditional U-statistics. More details on the established works on conditional U-statistics are given in the next chapter.

1.1 Motivation

1.1.1 Ranking problems

A considerable number of works concerning conditional U-statistics have emerged following the two decades of their introduction, allowing many applications in diverse fields. For instance, the ranking problem has received particular attention in machine learning. Comparing two observations based on their characteristics is necessary for some specific ranking problems and deciding which is better instead of simply classifying them. Furthermore, the ordering problems have many applications in different areas of banking (Data mining process for direct marketing data extraction), document type classification, and many more. U-statistics are frequently used to solve ordering and ranking problems where the objective is to establish a universal and consistent ordering method. Suppose that we want to establish an order between the first components of the two pairs $(X, Y), (X', Y')$ of independent and identically distributed observations in $\mathcal{X} \times \mathbb{R}$. The variables Y and Y' are respective labels of the variables X and X' that we want to order by observing them (and not their labels). Usually, we decide that X is better than X' if $Y > Y'$. To see things more clearly, we introduce a new variable :

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

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

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

The following ranking risk gives the performance measure of r :

$$L(r) = \mathbb{P}(W \cdot r(X, X') < 0).$$

A natural estimate for L according to Cl emen on *et al.* [2008] is :

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

where $(X_1, Y_1), \dots, (X_n, Y_n)$ are n independent, identically distributed copies of (X, Y) , and $W_{i,j} = \frac{Y_i - Y_j}{2}$. One can easily see that L_n is a U-statistic with $m = 2$. Cl emen on *et al.* [2008], considered a minimizer over a class of ranking rules. So to learn a rigorous ranking rule, one should minimize the ranking risk (empirically) over a class \mathcal{R} of ranking rules, i.e.,

$$r_n = \underset{r \in \mathcal{R}}{\operatorname{argmin}} L_n(r).$$

Then one needs to investigate the performance measure of r_n given by :

$$L(r_n) = \mathbb{P}(W \cdot r_n(X, X') \mid (X_1, Y_1), \dots, (X_n, Y_n))$$

The initial step to studying the empirical risk minimizer performance is based on the inequality:

$$L(r_n) - \inf_{r \in \mathbb{R}} L(r) \leq 2 \sup_{r \in \mathbb{Z}} |L_n(r) - L(r)|.$$

The latter indicates that evaluating the performance of an empirical minimizer of the ranking risk comes up for discussing the properties of conditional U-statistics, that is, the properties of a collection of conditional U-statistics indexed by a class of ranking rules. The ordering problem presented here is a simple ordering problem that is a problem of ranking just two observations X, X' . More general ranking problems those of ordering m independent observations $X_1^{(m)}, \dots, X_n^{(m)}$, involve the conditional U-statistic and the conditional U-process with degree m .

1.1.2 Kendall rank correlation coefficient

To test the independence of one-dimensional random variables X and Y Kendall [1938] proposed a method based on the U-statistic K_n with the kernel function :

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

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

$$\hat{r}_n^{(2)}(\varphi, \mathbf{t}) = \frac{\sum_{i \neq j}^n \varphi(Y_i, Y_j) \mathbb{K}\left(\frac{t_1 - X_i}{h_n}\right) \mathbb{K}\left(\frac{t_2 - X_j}{h_n}\right)}{\sum_{i \neq j}^n \mathbb{K}\left(\frac{t_1 - X_i}{h_n}\right) \mathbb{K}\left(\frac{t_2 - X_j}{h_n}\right)}, \quad (1.1.1)$$

where $\mathbf{t} = (t_1, t_2) \in \mathbb{I} \subset \mathbb{R}^2$ and $\varphi(\cdot)$ is Kendall's kernel (1.1.1). Suppose that $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are d_1 and d_2 -dimensional random vectors respectively and $d_1 + d_2 = d$. Furthermore, suppose that Y_1, \dots, Y_n are observations of $(\boldsymbol{\xi}, \boldsymbol{\eta})$, we are interested in testing :

$$H_0 : \boldsymbol{\xi} \text{ and } \boldsymbol{\eta} \text{ are conditionally independent given } X. \text{ vs } H_a : H_0 \text{ is not true.} \quad (1.1.2)$$

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

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

As in [Zhang \[2001\]](#), for $m = 2$ and the class of functions :

$$\mathcal{F}_{\mathbf{a}} \mathcal{K}^2 = \left\{ \varphi^{\mathbf{a}}(\cdot, \cdot) \mathbb{K}\left(\frac{\cdot - t_1}{h}\right) \mathbb{K}\left(\frac{\cdot - t_2}{h}\right) : \mathbf{a} \in \mathbb{R}^d : \|\mathbf{a}\| = 1 \right\},$$

it is easy to see that $\sup_{\|\mathbf{a}\|=1} \sqrt{nh^2} \hat{r}_2^{(2)}(\varphi^{\mathbf{a}}, \mathbf{t})$ null hypothesis is true, then :

$$D_n = \sup_{\|\mathbf{a}\|=1} \sqrt{nh^2} \left| \hat{r}_2^{(2)}(\varphi^{\mathbf{a}}, \mathbf{t}) \right| = \sup_{\varphi^{\mathbf{a}} \times \mathbb{K} \times \mathbb{K} \in \mathcal{F}_{\mathbf{a}} \mathcal{K}^2} \sqrt{nh^2} \left| \hat{r}_2^{(2)}(\varphi^{\mathbf{a}}, \mathbf{t}) \right|.$$

1.1.3 High-dimensional independence test ([Drton et al. \[2020\]](#))

Let the random vector $\mathbf{X} := (X_1, \dots, X_p)^\top$ taking values in \mathbb{R}^p with existing continuous univariate marginal distributions. Suppose we are interested in testing the dependence between the elements of \mathbf{X} , for any value of p (even for a large p) :

$$H_0 : \text{The components } X_1, \dots, X_p \text{ are mutually independent.} \quad (1.1.3)$$

This kind of test has always been of considerable regard; it has been the focus of several works, e.g., the likelihood ratio test, Roy's largest root test [Roy \[1957\]](#), and Nagao's L_2 -type test [Nagao \[1973\]](#). However, these methods give inferior results when p is large enough, so to deal with this challenge, more recent methods were investigated; for example, [Jiang and Yang \[2013\]](#) derived corrected likelihood ratio tests for Gaussian data. Using covariance/correlation statistics such as Pearson r , Spearman ρ , and Kendall τ , [Bao et al. \[2012\]](#), [Gao et al. \[2017\]](#), [Han et al. \[2018\]](#) and [Bao \[2019\]](#) proposed revised versions of Roy largest root test. [Schott \[2005\]](#) and [Leung and Drton \[2018\]](#) derived corrected Nagao L_2 -type tests. Finally, [Jiang \[2004\]](#), [Zhou \[2007\]](#), and [Han et al. \[2017\]](#) proposed tests using the magnitude of the largest pairwise correlation statistics. Nevertheless, most of the cited tests are largely based on linear and simple rank correlations, which makes them unable to detect non-linear and non-monotone dependencies. Let X_1, X_2 two variables with absolutely continuous joint distribution; most of the dependence measures respect the following criterion:

1. I-consistency. If X_1 and X_2 are independent, their correlation equals zero.
2. D-consistency. If X_1 and X_2 are dependent, their correlation is nonzero.
3. Monotonic-invariance. The correlation measurement does not change by monotonic transformations.

Drton *et al.* [2020] used the U-statistic to introduce a pair-wise rank correlation that was involved in a maximum-type test of the independence hypothesis (1.1.3). Assume that $\mathbf{X}_1, \dots, \mathbf{X}_n$ are independent copies of \mathbf{X} and let the fixed function $\varphi : \mathbb{R}^{2m} \rightarrow \mathbb{R}$. The U-statistic based on the kernel φ is :

$$\hat{u}_{j,k}^{(m)} := \frac{(n-m)!}{n!} \sum_{\epsilon \in l(m,n)} \varphi \left(\begin{pmatrix} X_{i_1,j} \\ X_{i_1,k} \end{pmatrix}, \dots, \begin{pmatrix} X_{i_m,j} \\ X_{i_m,k} \end{pmatrix} \right),$$

for $i \neq j \in \{1, \dots, p\}$ and $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,p})$. One should note that if the kernel φ is rank based then the correlation statistics $\hat{u}_{j,k}^{(m)}$ and the corresponding measures $\mathbb{E}(\hat{u}_{j,k}^{(m)})$ are also rank based. They are both I-consistent and D-consistent. Some notions are required to understand how the U-statistics are involved in the test. For m independent identically distributed random vectors W_1, \dots, W_m with distribution \mathbb{P}_W , and $k \in \{1, \dots, m\}$ let :

$$\begin{aligned} \mathbb{E}[\varphi] &= \mathbb{E}[\varphi(W_1, \dots, W_m)], \\ \varphi_k(w_1, \dots, w_k; \mathbb{P}_W) &:= \mathbb{E}[\varphi(w_1, \dots, w_k, W_{k+1}, \dots, W_m)], \\ \varphi^{(k)}(w_1, \dots, w_k; \mathbb{P}_W) &:= \varphi_k(w_1, \dots, w_k; \mathbb{P}_W) - \mathbb{E}[\varphi] \sum_{j=1}^{k-1} \sum_{1 \leq i_1 \leq \dots \leq i_j \leq k} \varphi^{(j)}(w_{i_1}, \dots, w_{i_j}; \mathbb{P}_W). \end{aligned}$$

We say that the kernel $\varphi(\cdot)$ and the U-statistic based on, are degenerate under \mathbb{P}_W if

$$\text{Var}[\varphi_1] = 0.$$

Let \mathcal{U}_0 be the uniform distribution on $[0; 1]$ and $\mathcal{U}_0 \otimes \mathcal{U}_0$ the uniform distribution on $[0; 1]^2$. The kernel φ is supposed to satisfy :

- φ is symmetric, bounded, mean-zero and rank-based.
- φ satisfies, as $W_1 \mathcal{U}_0 \otimes \mathcal{U}_0$:

$$\mathbb{E}[\varphi_1(W_1; \mathcal{U}_0 \otimes \mathcal{U}_0)]^2 = 0.$$

- $\varphi_2(w_1, w_2; \mathcal{U}_0 \otimes \mathcal{U}_0)$ satisfies :

$$\varphi_2(w_1, w_2; \mathcal{U}_0 \otimes \mathcal{U}_0) = \sum_{n=1}^{\infty} \lambda_n h_n(w_1) h_n(w_2),$$

where $\{\lambda_n\}_{n \in \mathbb{N}^*}$ $\{h_n(\cdot)\}_{n \in \mathbb{N}^*}$ are the eigenvalues and the eigenfunctions respectively, that satisfy for $w_1 \in \mathbb{R}^2$:

$$\mathbb{E}[\varphi_2(w_1, W_2) h_2(z_2)] = \lambda_1 h_1(z_1)$$

and $W_2 \mathcal{U}_0 \otimes \mathcal{U}_0, \{\lambda_n\}_{n \in \mathbb{N}^*}$ is a decreasing sequence with

$$\sum_{n=1}^{\infty} \lambda_n \in (0; \infty).$$

Finally, the eigen functions satisfy:

$$\sup_n \|h_n\|_{\infty} < \infty.$$

The proposed maximum-type test of the mutual independence hypothesis H_0 is given by :

$$M_n := (n - 1) \max_{j < k} \hat{u}_{j,k}^{(m)},$$

then H_0 is rejected if M_n is bigger than some lower limit.

1.1.4 U-statistics in clustering

The clustering method brings together a set of learning algorithms that aim to group unlabeled data with similar properties. Isolating schemas or families in this way also prepares the ground for the subsequent application of supervised learning algorithms (such as KNN). Clustering is used in particular when it is expensive to label data. However, it is a mathematically ill-defined problem: different metrics and/or different data representations will result in different clustering without any of them necessarily being better than the other. Thus, the clustering method must be carefully chosen according to the expected result and the intended use of the data. An approach based on U-statistics was introduced to assure a statistically significant clustering. Indeed, let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be a sample of n random vectors taking values in \mathbb{R}^p forming two groups G_1 and G_2 of sizes n_1, n_2 , respectively, with $n_1 + n_2 = n$. In the i -th group, observations $\mathbf{X}_{j_1}, \dots, \mathbf{X}_{j_{n_j}}$ are assumed to be independent and identically distributed (i.i.d) with a p -variate distribution F_j . Assume that the distribution F_j admits mean vector μ_j and positive-definite dispersion matrix Σ_j . We want to test if the groups G_1 and G_2 really are separated groups or if they have the same distribution, i.e.:

$$H_0: \text{The distributions } F_1 \text{ and } F_2 \text{ are equal, } H_1: \text{The distributions } F_1 \text{ and } F_2 \text{ differ.} \quad (1.1.4)$$

For a 2 order symmetric function $\varphi(\cdot, \cdot)$, the functional distance is defined according to [Sen \[2006\]](#) and [Pinheiro *et al.* \[2009\]](#), by :

$$\theta(F_1, F_2) = \iint \varphi(x_1, x_2) dF_1(x_1) dF_2(x_2).$$

The latter can be used to define both distances within and between groups. It follows from U-statistics theory that an unbiased estimator of this functional for within-group distance $\theta(F_j, F_j)$ is the generalized U-statistic based on the symmetric kernel φ , that is :

$$u_{n_j}^{(j)} = \binom{n_j}{2}^{-1} \sum_{1 \leq i < k \leq n_j} \varphi(\mathbf{X}_{ji}, \mathbf{X}_{jk}).$$

While the unbiased estimator to measure the distance between the groups :

$$u_{n_1, n_2}^{(1,2)} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=k}^{n_2} \varphi(\mathbf{X}_{1i}, \mathbf{X}_{2k}).$$

One should note that the U-statistic based on the whole sample can be written as :

$$\begin{aligned} u_n^{(2)} &= \binom{n}{2}^{-1} \sum_{1 \leq i < k \leq n} \varphi(X_i, X_k) \\ &= \sum_{j=1}^2 \frac{n_j}{n} u_{n_j}^{(j)} + \frac{n_1 n_2}{n(n-1)} (2u_{n_1, n_2}^{(1,2)} - u_{n_1}^{(1)} - u_{n_2}^{(2)}), \\ &= I_n + II_n. \end{aligned}$$

The second term II_n is centered under the null hypothesis, i.e., $\mathbb{E}(II_n) = 0$. Alternatively, $\mathbb{E}(II_n) \geq 0$. When II_n takes large values H_0 is then rejected.

1.1.5 Functional data

Statistical problems involving the analysis of infinite-dimensional variables, such as curves, sets, and images, have garnered considerable attention in the literature for some years. The development of this research theme is promoted by modern technological progress in measuring devices and the abundance of data measured on increasingly fine grids: this is the case; take, for example, in meteorology, medicine, satellite imagery, and many other areas of research. Notably, the statistical analysis of functional data can be traced back to Rao [1958] where a set of growth curves was analyzed by functional principal component analysis in a rudimentary form for the first time and subsequently rigorously investigated by Kleffe [1973] and Dauxois *et al.* [1982]. Numerous recent works, from both a theoretical and a practical standpoint, confirm the relevance of investigating this area, such as the monographs of Horváth and Kokoszka [2012], Shi and Choi [2011], Bosq and Blanke [2007], Geenens [2011], Cuevas [2014], Shang [2014], Horváth and Rice [2015], Müller [2016], Nagy [2017], Vieu [2018], Aneiros *et al.* [2019], and Ling and Vieu [2018], where we have collect these references.

Functional variables can model many phenomena, such as forecasting problems, electricity consumption, speech recognition, and spectrometric data (cf Ferraty and Vieu [2006]). The following example concerns the analysis of chemical components in a food substance (meat) by spectrometry, and as mentioned in Ferraty and Vieu [2006] the original data come from a quality control problem in the food industry; they were recorded on a Tecator Infratec Food and Feed Analyzer working in the wavelength range 850 – 1050 nm by the near-infrared (NIR) transmission principle <http://lib.stat.cmu.edu/datasets/tecator>. We have a sample of $n = 15$ pieces of meat. For each unit i , we observe one spectrometric discredited curve corresponding to the absorbance measured at a grid of 100 wavelengths Figure 1.1. By considering these data, many statistical questions arise, including the fact that the main objective of the spectrometric analysis is to make it possible to discover the proportion of certain chemical contents. In contrast, the treatment would take more time and be much more expensive. If we suppose that Y is the proportion of some components, to predict Y , one would like to use the spectrometric curves: This problem is a typical nonparametric functional regression problem.

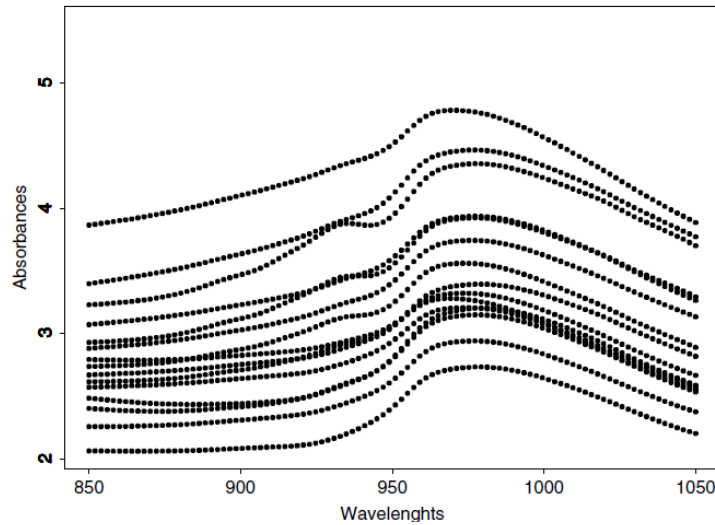


Figure 1.1: Some Units Among The Original Chemometric Data

1.2 Organization of the dissertation

Chapter 2. Mathematical background

This chapter is devoted to the preliminary results for a few specific topics, which we will need to be self-contained and better understood in the forthcoming chapters. We also review some of the standard facts concerning empirical processes and their convergence, with special attention given to the conditional U-statistics (processes) estimation.

Chapter 3. Uniform consistency and uniform in the number of neighbors consistency for functional conditional U-statistics

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be i.i.d random copies of the random vector functions $(X, Y) \in \mathcal{X} \times \mathcal{Y}$, where \mathcal{X} and \mathcal{Y} are some abstract spaces. In this chapter, we consider that $(\mathcal{X}, d_{\mathcal{X}})$ is a semi-metric space endowed with a semi-metric $d(\cdot, \cdot)$ ¹. Let $\varphi : \mathcal{Y}^m = \mathcal{Y} \times \dots \times \mathcal{Y} \rightarrow \mathbb{R}$ be a measurable function belonging to some class of functions \mathcal{F}_m , we are interested in the estimation of the regression function

$$r^{(m)}(\varphi, \mathbf{t}) := \mathbb{E}(\varphi(Y_1, \dots, Y_m) | X_1, \dots, X_m = \mathbf{t}), \quad (1.2.1)$$

whenever it exists. One of the widely studied estimates of (1.2.1) are the conditional U-statistics introduced by [Stute \[1991\]](#) and extended later by [Fu \[2012\]](#) to the functional setting. These estimates and many other kernel-based estimators depend on a real-valued non-random bandwidth sequence h_n . However, the difficulty of these methods is the choice of a smoothing parameter

¹A semi-metric (sometimes called pseudo-metric) $d(\cdot, \cdot)$ is a metric which allows $d(x_1, x_2) = 0$ for some $x_1 \neq x_2$.

that does not lead to either over-smoothness or under-smoothness of the estimate, and also the fact that the bandwidth h_n often does not take into consideration the local features of the data. Hence, we substitute the bandwidth h_n with data-dependent bandwidths generated using the k nearest neighbor (k NN) method. We define the following functional conditional U-statistic in the k NN setting given for all $\mathbf{t} \in \mathcal{X}^m$ by:

$$\hat{r}_n^{*(m)}(\varphi, \mathbf{t}, \mathbf{h}_{n,k}(\mathbf{t})) = \frac{\sum_{(i_1, \dots, i_m) \in \mathcal{I}(m, n)} \varphi(Y_{i_1}, \dots, Y_{i_m}) \mathbb{K}\left(\frac{d(t_1, X_{i_1})}{H_{n,k}(t_1)}\right) \cdots \mathbb{K}\left(\frac{d(t_m, X_{i_m})}{H_{n,k}(t_m)}\right)}{\sum_{(i_1, \dots, i_m) \in \mathcal{I}(m, n)} \mathbb{K}\left(\frac{d(t_1, X_{i_1})}{H_{n,k}(t_1)}\right) \cdots \mathbb{K}\left(\frac{d(t_m, X_{i_m})}{H_{n,k}(t_m)}\right)}, \quad (1.2.2)$$

where :

$$\begin{aligned} \mathcal{I}(m, n) &= \mathcal{I}_n^m := \{(i_1, \dots, i_m) : i_j \in \mathbb{N}^* \text{ and } i_r \neq i_j \text{ if } r \neq j\}, \\ \tilde{\mathbb{K}}(\mathbf{t}) &= \prod_{i=1}^m \mathbb{K}(t_i), \quad \mathbf{t} = (t_1, \dots, t_m) \in \mathcal{X}^m = \mathcal{X} \times \dots \times \mathcal{X}, \end{aligned}$$

and $\varphi : \mathcal{Y}^m = \mathcal{Y} \times \dots \times \mathcal{Y} \rightarrow \mathbb{R}$ is a measurable function belonging to some class of functions \mathcal{F}_m , with envelope function F , and $\mathbf{h}_{n,k}(\mathbf{t}) = (H_{n,k}(t_1), \dots, H_{n,k}(t_m))$ is a vector of positive random variables that depend on (X_1, \dots, X_n) such that, for all $\mathbf{x} = (x_1, \dots, x_m) \in \mathcal{X}^m$ and $j = 1, \dots, m$:

$$H_{n,k}(x_j) = \min \left\{ h \in \mathbb{R}^+ : \sum_{i=1}^n \mathbb{1}_{B(x_j, h)}(X_i) = k \right\}, \quad (1.2.3)$$

where $B(t, r) = \{z \in \mathcal{X} : d(z, t) \leq r\}$ is a ball in \mathcal{X} with the center $t \in \mathcal{X}$ and radius r , and $\mathbb{1}_A$ is the indicator function of the set A . For a kernel function $\mathbb{K}(\cdot)$ and a subset $\mathcal{S}_{\mathcal{X}} \subset \mathcal{X}$, we define the pointwise measurable class of functions, for $1 \leq m \leq n$:

$$\mathcal{K}^m := \left\{ (x_1, \dots, x_m) \mapsto \prod_{i=1}^m \mathbb{K}\left(\frac{d(x_i, t_i)}{h_i}\right), (h_1, \dots, h_m) \in \mathbb{R}_+^m \setminus \{0\} \text{ and } (\mathbf{t}_1, \dots, \mathbf{t}_m) \in \mathcal{S}_{\mathcal{X}}^m \right\}.$$

In this chapter, we are interested in establishing the uniform consistency and the uniform in the number of neighbors k and uniform consistency on a VC-type class of functions of the non-parametric functional regression estimator and the functional conditional U-process (statistic) by giving the consistency rates almost completely (a.co). The results of this chapter combine three different aspects in a doubly infinite context, namely the functional framework and the nonparametric framework. The investigated results in this chapter complement and extend the previous results of [Bouzebda and Nemouchi \[2020\]](#). However, it should be noted that applying the k NN methods on conditional U-statistics has never been developed in the functional framework. We are also based on the works of [Burba et al. \[2009\]](#) and [Kara et al. \[2017\]](#) to give a good rate of consistency almost completely under very mild conditions. i.e., for the metric entropy $\Psi_{\mathcal{S}_{\mathcal{X}}}$ of some subset $\mathcal{S}_{\mathcal{X}} \subset \mathcal{X}$:

$$\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}}\left(\frac{\log n}{n}\right)}{n\phi(a_n)}}$$

This chapter unifies the approaches in some other recent papers. As a by-product of our proofs, we state consistency results for the k -NN conditional U-statistics, under the random censoring, are uniform in the number of neighbors. We provide some applications, including the set indexed conditional U-statistics, the multi-sample U processes, discrimination, and metric learning. Finally, we briefly discuss the bandwidth selection criterion.

Chapter 4. Asymptotic properties of conditional U-statistics using delta sequences

Let $(\mathbf{X}, \mathbf{Y}), (\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_n, \mathbf{Y}_n)$ be independent and identically distributed random vectors with common joint density function $f_{\mathbf{X}, \mathbf{Y}}: \mathbb{R}^d \times \mathbb{R}^{d'} \rightarrow [0, \infty[$. Let $\varphi: \mathbb{R}^{d+k} \rightarrow \mathbb{R}$ be a measurable function. We are interested in the estimation of the conditional expectation, or regression function

$$r^{(k)}(\varphi, \tilde{\mathbf{x}}) = \mathbb{E}(\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \tilde{\mathbf{x}}), \text{ for } \tilde{\mathbf{x}} \in \mathbb{R}^{dk}, \quad (1.2.4)$$

whenever it exists, i.e, $\mathbb{E}(|\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k)|) < \infty$. To be more precise, we want to extend the results of conditional U-statistics presented in [Stute \[1991\]](#), using the delta sequences method. The delta-sequence technique is a device to study the properties of estimates of several usual methods in nonparametric function estimation simultaneously. Introduced by [Watson and Leadbetter \[1964\]](#), the delta-sequence method relative to the density estimation was studied among others by [Földes and Révész \[1974\]](#) and [Walter and Blum \[1979\]](#) whereas it was extended to the multivariate case by [Susarla and Walter \[1981\]](#). To estimate the regression function $r^{(k)}(\varphi, \cdot)$ given in (1.2.4), we propose a new general family of estimates, for each $\tilde{\mathbf{x}} \in \mathbb{R}^{dk}$, defined by

$$\hat{r}_n^{(k)}(\varphi, \tilde{\mathbf{x}}; m_n) = \begin{cases} \frac{\sum_{(i_1, \dots, i_k) \in \mathcal{I}(k, n)} \varphi(Y_{i_1}, \dots, Y_{i_k}) \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k})}{\sum_{(i_1, \dots, i_k) \in \mathcal{I}(k, n)} \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k})}, \\ \text{if } \sum_{(i_1, \dots, i_k) \in \mathcal{I}(k, n)} \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k}) \neq 0, \\ \frac{n!}{(n-k)!k!} \sum_{(i_1, \dots, i_k) \in \mathcal{I}(k, n)} \varphi(Y_{i_1}, \dots, Y_{i_k}) \\ \text{if } \sum_{(i_1, \dots, i_k) \in \mathcal{I}(k, n)} \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k}) = 0, \end{cases} \quad (1.2.5)$$

where $\{\delta_m\}$ are positive delta-sequences which will be defined later in chapter 4. In this chapter, we study the asymptotic properties of the new family of estimators defined in (1.2.5) and establish the exponential inequalities and the limiting law. We provide an important application for the censored data setting that is of its own interest. We give some applications, including the discrimination, the metric learning, the multipartite ranking and the Ranking problems. A short discussion of the bandwidth selection criterion is provided.

Chapter 5. Uniform consistency for functional conditional U-statistics using delta-sequences

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $(\mathcal{X}, d(\cdot, \cdot))$ be an infinite-dimensional separable Banach space equipped with a norm $\|\cdot\|$ with $d(x, y) := \|x - y\|$ and \mathcal{B} be the σ -algebra of Borel subsets of \mathcal{X} . We consider a sequence $\{X_i, Y_i : i \geq 1\}$ of independent identically distributed random copies of the random vector (X, Y) , where X is a random element defined on $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in $(\mathcal{X}, \mathcal{B})$ and Y takes values in some abstract space \mathcal{Y} . We are mainly interested in studying the functional framework's uniform consistency for the conditional U-statistic defined by (1.2.5). This work includes, in particular cases, the works of Prakasa Rao [2010] and Ouassou and Rachdi [2012]. This is a natural and nontrivial higher-order generalization that needs more involved treatment. As an application, we consider the Kendall rank correlation coefficient's uniform consistency, which is one of the main tools in testing conditional independence.

Chapter 6. Conclusions and perspectives

This chapter concludes this thesis by giving some remarks and future developments.

1.3 References

- Adamczak, R. a. (2006). Moment inequalities for U-statistics. *Ann. Probab.*, **34**(6), 2288–2314. [4](#)
- Aneiros, G., Cao, R., Fraiman, R., Genest, C., and Vieu, P. (2019). Recent advances in functional data analysis and high-dimensional statistics. *J. Multivariate Anal.*, **170**, 3–9. [10](#)
- Arcones, M. A. and Giné, E. (1993). Limit theorems for U-processes. *Ann. Probab.*, **21**(3), 1494–1542. [4](#)
- Arcones, M. A. and Yu, B. (1994). Central limit theorems for empirical and U-processes of stationary mixing sequences. *J. Theoret. Probab.*, **7**(1), 47–71. [4](#)
- Bao, Z. (2019). Tracy-Widom limit for Kendall's tau. *Ann. Statist.*, **47**(6), 3504–3532. [7](#)
- Bao, Z., Pan, G., and Zhou, W. (2012). Tracy-Widom law for the extreme eigenvalues of sample correlation matrices. *Electron. J. Probab.*, **17**, no. 88, 32. [7](#)
- Biau, G. and Bleakley, K. (2006). Statistical inference on graphs. *Statist. Decisions*, **24**(2), 209–232. [5](#)
- Borovskikh, Y. V. (1986). The theory of U-statistics in a Hilbert space. *Akad. Nauk Ukrain. SSR Inst. Mat. Preprint*, (78), 56. [4](#)

- Bosq, D. and Blanke, D. (2007). *Inference and prediction in large dimensions*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester; Dunod, Paris. 10
- 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**(2), 452–509. 12
- Burba, F., Ferraty, F., and Vieu, P. (2009). k -nearest neighbour method in functional nonparametric regression. *J. Nonparametr. Stat.*, **21**(4), 453–469. 12
- Cléménçon, S. (2014). A statistical view of clustering performance through the theory of U-processes. *J. Multivariate Anal.*, **124**, 42–56. 5
- Cléménçon, S., Lugosi, G., and Vayatis, N. (2008). Ranking and empirical minimization of U-statistics. *Ann. Statist.*, **36**(2), 844–874. 4, 6
- Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *J. Statist. Plann. Inference*, **147**, 1–23. 10
- Dauxois, J., Pousse, A., and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *J. Multivariate Anal.*, **12**(1), 136–154. 10
- de la Peña, V. H. and Giné, E. (1999). *Decoupling*. Probability and its Applications (New York). Springer-Verlag, New York. From dependence to independence, Randomly stopped processes. U-statistics and processes. Martingales and beyond. 4
- de la Pena, V. H. (1992). Decoupling and khintchine’s inequalities for u-statistics. *The Annals of Probability*, pages 1877–1892. 4
- Drton, M., Han, F., and Shi, H. (2020). High-dimensional consistent independence testing with maxima of rank correlations. *Ann. Statist.*, **48**(6), 3206–3227. 7, 8
- Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis*. Springer Series in Statistics. Springer, New York. Theory and practice. 10
- Földes, A. and Révész, P. (1974). A general method for density estimation. *Stud. Sci. Math. Hungar.*, **9**, 81–92 (1975). 13
- Frees, E. W. (1994). Estimating densities of functions of observations. *J. Amer. Statist. Assoc.*, **89**(426), 517–525. 4
- Fu, K.-A. (2012). An application of U-statistics to nonparametric functional data analysis. *Comm. Statist. Theory Methods*, **41**(9), 1532–1542. 11

- Gao, J., Han, X., Pan, G., and Yang, Y. (2017). High dimensional correlation matrices: the central limit theorem and its applications. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, **79**(3), 677–693. [7](#)
- Geenens, G. (2011). Curse of dimensionality and related issues in nonparametric functional regression. *Stat. Surv.*, **5**, 30–43. [10](#)
- Giné, E. and Mason, D. M. (2007a). Laws of the iterated logarithm for the local U-statistic process. *J. Theoret. Probab.*, **20**(3), 457–485. [4](#)
- Giné, E. and Mason, D. M. (2007b). On local U-statistic processes and the estimation of densities of functions of several sample variables. *Ann. Statist.*, **35**(3), 1105–1145. [4](#)
- Giné, E. and Zinn, J. (1992). On Hoffmann-Jørgensen’s inequality for U-processes. In *Probability in Banach spaces, 8 (Brunswick, ME, 1991)*, volume 30 of *Progr. Probab.*, pages 80–91. Birkhäuser Boston, Boston, MA. [4](#)
- Giné, E., Latała, R., and Zinn, J. (2000). Exponential and moment inequalities for U-statistics. In *High dimensional probability, II (Seattle, WA, 1999)*, volume 47 of *Progr. Probab.*, pages 13–38. Birkhäuser Boston, Boston, MA. [4](#)
- Gregory, G. G. (1977). Large sample theory for U-statistics and tests of fit. *Ann. Statist.*, **5**(1), 110–123. [4](#)
- Halmos, P. R. (1946). The theory of unbiased estimation. *Ann. Math. Statistics*, **17**, 34–43. [4](#)
- Han, F., Chen, S., and Liu, H. (2017). Distribution-free tests of independence in high dimensions. *Biometrika*, **104**(4), 813–828. [7](#)
- Han, F., Xu, S., and Zhou, W.-X. (2018). On Gaussian comparison inequality and its application to spectral analysis of large random matrices. *Bernoulli*, **24**(3), 1787–1833. [7](#)
- Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics*, **19**, 293–325. [4](#)
- Hoeffding, W. (1961). The strong law of large numbers for u-statistics. Technical report, North Carolina State University. Dept. of Statistics. [4](#)
- Horváth, L. and Kokoszka, P. (2012). *Inference for functional data with applications*. Springer Series in Statistics. Springer, New York. [10](#)
- Horváth, L. and Rice, G. (2015). An introduction to functional data analysis and a principal component approach for testing the equality of mean curves. *Rev. Mat. Complut.*, **28**(3), 505–548. [10](#)

- Ibragimov, R. and Sharakhmetov, S. (1999). Analogues of Khintchine, Marcinkiewicz-Zygmund and Rosenthal inequalities for symmetric statistics. *Scand. J. Statist.*, **26**(4), 621–633. [4](#)
- Janson, S. (1979). *The asymptotic distribution of degenerate U-statistics*. [Uppsala] Department of Mathematics, Uppsala University. [4](#)
- Jiang, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.*, **14**(2), 865–880. [7](#)
- Jiang, T. and Yang, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist.*, **41**(4), 2029–2074. [7](#)
- Joly, E. and Lugosi, G. (2016). Robust estimation of U-statistics. *Stochastic Process. Appl.*, **126**(12), 3760–3773. [5](#)
- Kara, L.-Z., Laksaci, A., Rachdi, M., and Vieu, P. (2017). Data-driven k NN estimation in nonparametric functional data analysis. *J. Multivariate Anal.*, **153**, 176–188. [12](#)
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**(1/2), 81–93. [6](#)
- Klass, M. J. and Nowicki, K. (1997). Order of magnitude bounds for expectations of Δ_2 -functions of nonnegative random bilinear forms and generalized U-statistics. *Ann. Probab.*, **25**(3), 1471–1501. [4](#)
- Kleffe, J. (1973). Principal components of random variables with values in a separable Hilbert space. *Math. Operationsforsch. Statist.*, **4**(5), 391–406. [10](#)
- Korolyuk, V. S. and Borovskich, Y. V. (2013). *Theory of U-statistics*, volume 273. Springer Science & Business Media. [4](#)
- Lee, A. J. (1990). *U-statistics*, volume 110 of *Statistics: Textbooks and Monographs*. Marcel Dekker, Inc., New York. Theory and practice. [4](#)
- Leung, D. and Drton, M. (2018). Testing independence in high dimensions with sums of rank correlations. *Ann. Statist.*, **46**(1), 280–307. [7](#)
- Ling, N. and Vieu, P. (2018). Nonparametric modelling for functional data: selected survey and tracks for future. *Statistics*, **52**(4), 934–949. [10](#)
- Müller, H.-G. (2016). Peter Hall, functional data analysis and random objects. *Ann. Statist.*, **44**(5), 1867–1887. [10](#)
- Nagao, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.*, **1**, 700–709. [7](#)
- Nagy, S. (2017). An overview of consistency results for depth functionals. In *Functional statistics and related fields*, Contrib. Stat., pages 189–196. Springer, Cham. [10](#)

- Ouassou, I. and Rachdi, M. (2012). Regression operator estimation by delta-sequences method for functional data and its applications. *AStA Adv. Stat. Anal.*, **96**(4), 451–465. [14](#)
- Pinheiro, A., Sen, P. K., and Pinheiro, H. P. (2009). Decomposability of high-dimensional diversity measures: quasi-U-statistics, martingales and nonstandard asymptotics. *J. Multivariate Anal.*, **100**(8), 1645–1656. [9](#)
- Prakasa Rao, B. L. S. (2010). Nonparametric density estimation for functional data by delta sequences. *Braz. J. Probab. Stat.*, **24**(3), 468–478. [14](#)
- Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics*, **14**(1), 1–17. [10](#)
- Robins, J., Li, L., Tchetgen, E., and van der Vaart, A. W. (2009). Quadratic semiparametric von mises calculus. *Metrika*, **69**(2), 227–247. [5](#)
- Roy, S. N. (1957). *Some aspects of multivariate analysis*. John Wiley & Sons, Inc., New York; Indian Statistical Institute, Calcutta. [7](#)
- Rubin, H. and Vitale, R. A. (1980). Asymptotic distribution of symmetric statistics. *Ann. Statist.*, **8**(1), 165–170. [4](#)
- Schott, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika*, **92**(4), 951–956. [7](#)
- Sen, P. K. (1974). Almost sure behaviour of u-statistics and von mises’ differentiable statistical functions. *The Annals of Statistics*, pages 387–395. [4](#)
- Sen, P. K. (2006). Robust statistical inference for high-dimensional data models with application to genomics. *Austrian journal of statistics*, **35**(2&3), 197–214. [9](#)
- Serfling, R. J. (2009). *Approximation theorems of mathematical statistics*, volume 162. John Wiley & Sons. [4](#)
- Shang, H. L. (2014). A survey of functional principal component analysis. *AStA Adv. Stat. Anal.*, **98**(2), 121–142. [10](#)
- Shi, J. Q. and Choi, T. (2011). *Gaussian process regression analysis for functional data*. CRC Press, Boca Raton, FL. [10](#)
- Stute, W. (1991). Conditional U-statistics. *Ann. Probab.*, **19**(2), 812–825. [5](#), [11](#), [13](#)
- Susarla, V. and Walter, G. (1981). Estimation of a multivariate density function using delta sequences. *Ann. Statist.*, **9**(2), 347–355. [13](#)
- Vieu, P. (2018). On dimension reduction models for functional data. *Statist. Probab. Lett.*, **136**, 134–138. [10](#)

- Walter, G. and Blum, J. (1979). Probability density estimation using delta sequences. *Ann. Statist.*, **7**(2), 328–340. [13](#)
- Watson, G. S. and Leadbetter, M. R. (1964). Hazard analysis. II. *Sankhyā Ser. A*, **26**, 101–116. [13](#)
- Zhang, D. (2001). Bayesian bootstraps for U-processes, hypothesis tests and convergence of Dirichlet U-processes. *Statist. Sinica*, **11**(2), 463–478. [7](#)
- Zhou, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Trans. Amer. Math. Soc.*, **359**(11), 5345–5363. [7](#)

Chapter 2

Mathematical background

2.1 Mathematical Background

In this chapter, we present some of the essential tools and concepts that will be used in the remainder of this thesis. In particular, we make a presentation of essential notions and tools concerning theory, this being with the most important bibliographical references; for more details, we refer to [Kosorok \[2008\]](#), [van der Vaart and Wellner \[1996\]](#) and [Athreya and Lahiri \[2006\]](#).

2.1.1 Metric and Normed Spaces

We introduce some concepts and results for metric spaces. Before defining metric spaces, we briefly review the σ -fields and the measure spaces.

Definition 2.1.1.1 *Let Ω be a nonempty set, and $\mathcal{P}(\Omega) \equiv \{A : A \subset \Omega\}$ be the power set of Ω , i.e., the class of all subsets of Ω .*

A collection of sets $\mathcal{A} \subset \mathcal{P}(\Omega)$ is called an algebra if

- (a) $\Omega \in \mathcal{A}$;
- (b) $A \in \mathcal{A}$ implies $A^c \in \mathcal{A}$;
- (c) $A, B \in \mathcal{A}$ implies $A \cup B \in \mathcal{A}$ (i.e., closure under pairwise unions).

In other words, an algebra is a class of sets containing Ω that is closed under complementation and pairwise (and hence finite) unions. It is easy to see that one can equivalently define an algebra by requiring that properties (a), (b) hold and that the property

$$(c') \quad A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$$

holds (i.e., closure under finite intersections).

Definition 2.1.1.2 *A class $\mathcal{A} \subset \mathcal{P}(\Omega)$ is called a σ -algebra if it is an algebra and if it satisfies*

(d) $A_n \in \mathcal{A}$ for $n \geq 1 \Rightarrow \bigcup_{n \geq 1} A_n \in \mathcal{A}$.

Thus, a σ -algebra class of subsets of Ω contains Ω and is closed under complementation and countable unions. An instrumental class of σ -algebras is those generated by open sets of a topological space. These are called Borel σ -algebras.

Definition 2.1.1.3 A topological space is a pair $(\mathbb{S}, \mathcal{T})$ where \mathbb{S} is a nonempty set and \mathcal{T} is a collection of subsets of \mathbb{S} such that

(i) $\mathbb{S} \in \mathcal{T}$;

(ii) $\mathcal{O}_1, \mathcal{O}_2 \in \mathcal{T} \Rightarrow \mathcal{O}_1 \cap \mathcal{O}_2 \in \mathcal{T}$;

(iii) $\{\mathcal{O}_\alpha : \alpha \in \mathbb{I}\} \subset \mathcal{T} \Rightarrow \bigcup_{\alpha \in \mathbb{I}} \mathcal{O}_\alpha \in \mathcal{T}$.

Elements of \mathcal{T} are called open sets.

Definition 2.1.1.4 The Borel σ -algebra on a topological space \mathbb{S} (in particular, on a metric space or a Euclidean space) is defined as the σ -algebra generated by the collection of open sets in \mathbb{S} .

Definition 2.1.1.5 Let Ω be a nonempty set and \mathcal{A} be an algebra on Ω . Then, a set function μ on \mathcal{A} is called a measure if

(a) $\mu(A) \in [0, \infty]$ for all $A \in \mathcal{A}$;

(b) $\mu(\emptyset) = 0$;

(c) for any disjoint collection of sets $A_1, A_2, \dots, \in \mathcal{A}$ with $\bigcup_{n \geq 1} A_n \in \mathcal{A}$,

$$\mu\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

Definition 2.1.1.6 A measure μ is called finite or infinite according as $\mu(\Omega) < \infty$ or $\mu(\Omega) = \infty$. A finite measure with $\mu(\Omega) = 1$ is called a probability measure. A measure μ on a σ -algebra \mathcal{A} is called σ -finite if there exist a countable collection of sets $A_1, A_2, \dots, \in \mathcal{A}$, not necessarily disjoint, such that

(a) $\bigcup_{n \geq 1} A_n = \Omega$;

(b) $\mu(A_n) < \infty$ for all $n \geq 1$.

The triple $(\Omega, \mathcal{A}, \mu)$ is called a measure space. For a probability measure P on a set Ω with σ -field \mathcal{A} , the triple (Ω, \mathcal{A}, P) is called a probability space.

Definition 2.1.1.7 Let \mathbb{S} be a nonempty set. Let $d : \mathbb{S} \times \mathbb{S} \rightarrow \mathbb{R}_+ = [0, \infty)$ be such that

(i) $d(x, y) = d(y, x)$ for any x, y in \mathbb{S} ;

(ii) $d(x, z) \leq d(x, y) + d(y, z)$ for any x, y, z in \mathbb{S} (triangle inequality);

(iii) $d(x, y) = 0$ iff $x = y$.

Such a d is called a metric on \mathbb{S} and the pair (\mathbb{S}, d) a metric space. If the map $d(\cdot, \cdot)$ satisfies only (i) and (ii) then it is called a semimetric or pseudometric.

Definition 2.1.1.8 A metric space (\mathbb{S}, d) is separable if there exists a subset $D \subset \mathbb{S}$ that is countable and dense in \mathbb{S} , i.e., for each x in \mathbb{S} and $\varepsilon > 0$, there is a y in D such that $d(x, y) < \varepsilon$.

A very important example of a metric space is a normed space, defined below.

Definition 2.1.1.9 1. A map $\|\cdot\| : \mathbb{V} \rightarrow [0, \infty)$ is a norm if the following axioms are satisfied; for all $x, y \in \mathbb{V}$ and $\alpha \in \mathbb{R}$

(i) $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality);

(ii) $\|\alpha x\| = |\alpha| \times \|x\|$ (scalar homogeneity);

(iii) $\|x\| = 0$ if and only if $x = 0$.

2. A normed space \mathbb{V} is a vector space (also called a linear space) equipped with a norm.

The map $\|\cdot\|$ is a seminorm if it satisfies only (i) and (ii). Note that a normed (respectively seminormed) space is a metric (respectively semimetric) space with $d(x, y) = \|x - y\|$, for all $x, y \in \mathbb{V}$.

Definition 2.1.1.10 A Banach space is a complete normed linear space $(\mathbb{V}, \|\cdot\|)$ (completeness being understood with respect to the metric induced by the norm).

Definition 2.1.1.11 Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. We define $L^p(\Omega, \mathcal{A}, \mu)$, $0 < p \leq \infty$, as the set of all measurable functions f on $(\Omega, \mathcal{A}, \mu)$ such that $\|f\|_p < \infty$, where for $0 < p < \infty$,

$$\|f\|_p = \left(\int |f|^p d\mu \right)^{\min\{\frac{1}{p}, 1\}}$$

and for $p = \infty$,

$$\|f\|_\infty \equiv \sup\{k : \mu(\{|f| > k\}) > 0\}$$

(called the essential supremum of f).

All $L^p(\Omega, \mathcal{A}, \mu)$ spaces are Banach spaces, provided $p \geq 1$ and in particular, all Euclidean spaces are Banach spaces. An example is the set of bounded real functions $f : S \rightarrow \mathbb{R}$, where S is an arbitrary set, this space is denoted $\ell^\infty(S)$. Another example is the space $\mathcal{C}[0, 1]$ of all real valued continuous functions on $[0, 1]$ with the norm (called the supnorm) is defined by $\|f\| = \sup\{|f(x)| : 0 \leq x \leq 1\}$.

2.1.2 Empirical processes and classes of functions

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space on which we define the sequence X_1, \dots, X_n and a collection of random variables $X = \{X(t) = X(t, \omega), \omega \in \Omega, t \in T\}$, T is an arbitrary index set.

Definition 2.1.2.1 *The collection $X = \{X(t) = X(t, \omega), \omega \in \Omega, t \in T\}$, is a stochastic process. For a fixed point $\omega \in \Omega$, the map:*

$$X(\cdot, \omega) : T \mapsto \mathbb{S},$$

is called the sample path of the stochastic process X .

An empirical process is a stochastic process based on random observations X_1, \dots, X_n .

Note that the space $\ell^\infty(T)$ is where most of the action occurs for statistical applications of empirical processes, so in next we will consider $\mathbb{S} = \ell^\infty(T)$, and for $x, y \in \mathbb{S} : d = \sup_{t \in T} |x(t) - y(t)|$ is the uniform distance on \mathbb{S} .

Generally, when dealing with empirical processes or U-processes, the index set $T = \mathcal{F}$ is a class of measurable functions φ . In the first part of the thesis, we mainly deal with classes of functions of VC-type subgraph classes.

Definition 2.1.2.2 *A class of subsets \mathcal{C} on a set C is called a VC-class if there exists a polynomial $P(\cdot)$ such that, for every set of N points in C , the class \mathcal{C} picks out at most $P(N)$ distinct subsets.*

Definition 2.1.2.3 *A class of functions \mathcal{F} is called a VC sub-graph class if the graphs of the functions in \mathcal{F} form a VC class of sets, that is, if we define the sub-graph of a real-valued function f on some space \mathcal{Z} as the following subset \mathcal{G}_φ on $\mathcal{Z} \times \mathbb{R}$:*

$$\mathcal{G}_f = \{(x, t) : 0 \leq t \leq \varphi(x) \quad \text{or} \quad \varphi(x) \leq t \leq 0\},$$

the class $\mathcal{C} = \{\mathcal{G}_\varphi : \varphi \in \mathcal{F}\}$ is a VC class of sets on $\mathcal{Z} \times \mathbb{R}$

Definition 2.1.2.4 *The class \mathcal{F} is pointwise measurable if there exists a countable subset $\mathcal{G} \subset \mathcal{F}$ such that for every $f \in \mathcal{F}$ there exists a sequence $\{g_l\} \in \mathcal{G}$ with $g_l(x) \rightarrow f(x)$ for every x .*

Assume that for $\varphi, \psi \in \mathcal{F}$, we have a metric function $d(\varphi, \psi)$.

Definition 2.1.2.5 *Let $\epsilon > 0$ be a positive real number and $\varphi_1, \dots, \varphi_N \in \mathcal{F}$. The elements $\varphi_1, \dots, \varphi_N$ are said ϵ -separated if:*

$$\text{for any } 1 \leq i \neq j \leq N : d(\varphi_i, \varphi_j) > \epsilon.$$

We call the ϵ -packing number, $N_{\square}(\epsilon, \mathcal{F}, d)$ the maximal cardinality of an ϵ -separated set.

Definition 2.1.2.6 *Let $\epsilon > 0$ be a positive real number and $\varphi_1, \dots, \varphi_N \in \mathcal{F}$. The elements $\varphi_1, \dots, \varphi_N$ are said ϵ -cover of \mathcal{F} if for any $\varphi \in \mathcal{F}$:*

$$\exists i \leq N : d(\varphi_i, \varphi) \leq \epsilon.$$

We call the ϵ -covering number, $N(\epsilon, \mathcal{F}, d)$ the minimal cardinality of an ϵ -cover of \mathcal{F} .

The metric entropy of the class \mathcal{F} is $\log(N(\epsilon, \mathcal{F}, d))$. The next lemma presents a link between the covering and the packing numbers of a function class \mathcal{F} .

Lemma 2.1.2.7 *For a class of functions \mathcal{F} we have :*

$$N_{[]} (2\epsilon, \mathcal{F}, d) \leq N(\epsilon, \mathcal{F}, d) \leq N_{[]} (\epsilon, \mathcal{F}, d).$$

The following lemma concerns the covering numbers of a VC- type class of functions.

Lemma 2.1.2.8 (Arcones and Yu [1994]) *If \mathcal{F} is VC-class of functions with an envelope function F , then for any $1 \leq p < \infty$, there exists a constants a and v such that :*

$$N(\epsilon, \mathcal{F}, \|\cdot\|_{Q,r}) \leq a \left(\frac{(\int F^r dQ)^r}{\epsilon} \right)^v$$

for all $\epsilon > 0$ and probability measure Q with

$$QF^r = \int F^r dQ < \infty.$$

Example 2.1.2.9 *The set \mathcal{F} of all indicator functions $\mathbb{1}_{(-\infty, t]}$ of cells in \mathbb{R} satisfies :*

$$N(\epsilon, \mathcal{F}, L_2(Q)) \leq \frac{2}{\epsilon^2},$$

for any probability measure Q and $\epsilon \leq 1$. Notice that :

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

For more details and discussion on this example refer to Example 2.5.4 of *van der Vaart and Wellner [1996]* and [*Kosorok, 2008, p. 157*]. The covering numbers of the class of cells $(-\infty, t]$ in higher dimension satisfy a similar bound, but with higher power of $(1/\epsilon)$, see Theorem 9.19 of *Kosorok [2008]*.

Example 2.1.2.10 *(Classes of functions that are Lipschitz in a parameter, Section 2.7.4 in van der Vaart and Wellner [1996]). Let \mathcal{F} be the class of functions $x \mapsto \varphi(t, x)$ that are Lipschitz in the index parameter $t \in T$. Suppose that:*

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

for some metric d on the index set T , the function $\kappa(\cdot)$ defined on the sample space \mathcal{X} , and all x . According to Theorem 2.7.11 of *van der Vaart and Wellner [1996]* and Lemma 9.18 of *Kosorok [2008]*, it follows, for any norm $\|\cdot\|_{\mathcal{F}}$ on \mathcal{F} , that :

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

Hence if (T, d) satisfy $J(\infty, T, d) = \int_0^\infty \sqrt{\log N(\epsilon, T, d)} d\epsilon < \infty$, then the conclusions holds for \mathcal{F} .

Example 2.1.2.11 Let us consider, as an example, the classes of functions that are smooth up to order α defined as follows, see Section 2 of [van der Vaart and Wellner \[1996\]](#). For $0 < \alpha < \infty$ let $[\alpha]$ be the greatest integer strictly smaller than α . For any vector $k = (k_1, \dots, k_d)$ of d integers define the differential operator :

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

where :

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

Then, for a function $\varphi : \mathcal{X} \rightarrow \mathbb{R}$, let :

$$\|\varphi\|_\alpha := \max_{k \leq [\alpha]} \sup_x |D^k \varphi(x)| + \max_{k \leq [\alpha]} \sup_{x, y} \frac{D^k \varphi(x) - D^k \varphi(y)}{\|x - y\|^{\alpha - [k]}}$$

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

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

Note that for $\alpha \leq 1$ this class consists of bounded functions $\varphi(\cdot)$ that satisfy a Lipschitz condition. [Kolmogorov and Tihomirov \[1961\]](#) computed the entropy of the classes of $C_M^\alpha(\mathcal{X})$ for the uniform norm. As a consequence of their results [van der Vaart and Wellner \[1996\]](#) shows that there exists a constant K depending only on α, d and the diameter of \mathcal{X} such that for every measure γ and every $\epsilon > 0$:

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

$N_{[\cdot]}$ is the bracketing number; refer to Definition 2.1.6 of [van der Vaart and Wellner \[1996\]](#) and we refer to Theorem 2.7.1 of [van der Vaart and Wellner \[1996\]](#) for a variant of the last inequality. By Lemma 9.18 of [Kosorok \[2008\]](#), we have :

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

2.1.3 Conditional U-statistics

In this section, we present the historical background and the main results concerning the conditional U-statistics.

Let $(\mathbf{X}_i, \mathbf{Y}_i), i \in \mathbb{N}^*$ be a sequence of i.i.d random real vectors defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with $\mathbf{X}_i \in \mathbb{R}^d$ and $\mathbf{Y}_i \in \mathbb{R}^{d'}$, $\varphi : \mathbb{R}^{d+m} \rightarrow \mathbb{R}$ a measurable symmetric function. In order to estimate the regression function

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

Stute [1991] introduced a class of estimators, called conditional U-statistics, which is defined for every $\mathbf{t} = (t_1, \dots, t_m) \in \mathbb{R}^{dm}$ by:

$$\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n) = \frac{\sum_{(i_1, \dots, i_m) \in I(m, n)} \varphi(Y_{i_1}, \dots, Y_{i_m}) \prod_{i=1}^m \mathbb{K}\left(\frac{\mathbf{t}_i - \mathbf{X}_{i_i}}{h_n}\right)}{\sum_{(i_1, \dots, i_m) \in I(m, n)} \prod_{i=1}^m \mathbb{K}\left(\frac{\mathbf{t}_i - \mathbf{X}_{i_i}}{h_n}\right)}, \quad (2.1.1)$$

$\mathbb{K}(\cdot)$ is the so-called smoothing kernel satisfying $\int_{\mathbb{R}^d} \mathbb{K}(t) dt = 1$, $\{h_n\}_{n \geq 1}$ is a sequence of bandwidth tending to zero at appropriate rates. Here summation extends over all permutations of length m , that is, over all pairwise distinct :

$$I(m, n) = \{\mathbf{i} = (i_1, \dots, i_m) : 1 \leq i_j \leq n \text{ and } i_j \neq i_r \text{ if } j \neq r\}.$$

Stute [1991] proved the asymptotic normality and weak and strong point-wise consistency of $\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$ when the observations are real ($d = 1, p = 1$) independent and identically distributed. He affirmed the possible extension of the results to the case when $d \geq 1$ and $p \geq 1$.

2.1.4 Historical background and main results for conditional U-statistics

Let's assume that:

- A.1 $h_n \rightarrow 0, nh_n \rightarrow \infty$ as $n \rightarrow \infty$.
- A.2 $\mathbb{K}(\cdot)$ is bounded with compact support.
- A.3 Let $\mathbf{t} = (t_1, \dots, t_m)$ be a point of continuity for each :

$$r_{jl}(\mathbf{t}) = \begin{cases} 0 & \text{if } t_j \neq t_l \\ \mathbb{E}_{j,l}(\mathbf{t}) & \text{if } t_j = t_l \end{cases}$$

where

$$\mathbb{E}_{j,l}(\mathbf{t}) = \mathbb{E} \left\{ \varphi(Y_1, \dots, Y_{j-1}, Y, Y_{j+1}, \dots, Y_m) \times \varphi(Y_{m+1}, \dots, Y_{m+j-1}, Y, Y_{m+j+1}, \dots, Y_{2m}) \right.$$

$$\left. | X_i = t_i \text{ for } i \neq j, X_{m+k} = t_k \text{ for } k \neq l \text{ and } X = t_j = t_l \right\}$$

- A.4 The density function of X_1 is continuous at each $t_j, 1 \leq j \leq m$, with $f(t_j) > 0$.
- A.5 $r_{jlk}(\cdot, \cdot, \cdot)$ is bounded in a neighborhood of \mathbf{t} , for all $1 \leq j, l, k \leq m$, where :

$$\begin{aligned} r_{jlk}(\mathbf{z}_m, \mathbf{z}_{2m}, \mathbf{z}_{3m}) &= \mathbb{E} \left\{ \varphi(Y_1, \dots, Y_{j-1}, Y, Y_{j+1}, \dots, Y_m) \right. \\ &\quad \times \varphi(Y_{m+1}, \dots, Y_{m+j-1}, Y, Y_{m+j+1}, \dots, Y_{2m}) \\ &\quad \times \varphi(Y_{2m+1}, \dots, Y_{2m+j-1}, Y, Y_{2m+j+1}, \dots, Y_{3m}) \\ &\quad \left. | X_i = z_i \text{ for } 1 \leq i \leq 3m, i \neq j, m+1, 2m+k, X = z \right\} \end{aligned}$$

and for $1 \leq k \leq 3, \mathbf{z}_{km} = (z_{(k-1)m+1}, \dots, z_{j-1}, z_1, z_{j+1}, \dots, z_{km})$.

A.6 $r_{1,2}^{(m)}(\cdot, \cdot)$ is bounded in a neighborhood of (\mathbf{t}, \mathbf{t}) , where

$$r_{1,2}^{(m)}(\mathbf{t}_1, \mathbf{t}_2) = \mathbb{E} \{ \varphi(Y_{i_1}, \dots, Y_{i_m}) \varphi(Y_{j_1}, \dots, Y_{j_m}) \mid (X_{i_1}, \dots, X_{i_m}) = \mathbf{t}_1, (X_{j_1}, \dots, X_{j_m}) = \mathbf{t}_2 \}.$$

A.1' The sequence h_n satisfies

$$\sum_n n^{-3/2} h_n^{-2} < \infty.$$

A.2' For some positive constants $c, c_1, c_2, r > 0$, and some decreasing function $H(\bullet)$ we have

$$K(t) \geq c \mathbb{1}_{\{|x| \leq r\}},$$

and

$$c_1 H(|t|) \leq K(t) \leq c_2 H(|t|).$$

A.3' The 3rd moment satisfies : $\mathbb{E} |\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_m)|^3 < \infty$.

The main first results on conditional U-statistics are the following :

Theorem 2.1.4.1 *Under assumptions (A.1 – A.6) and if $r^{(m)}(\cdot, \cdot)$ is continuous at \mathbf{t} , then :*

$$(nh_n)^{1/2} (\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n) - \mathbb{E} [G_{\varphi, \mathbf{t}}(\mathbf{X}_1, \dots, \mathbf{X}_n; \mathbf{Y}_1, \dots, \mathbf{Y}_n)]) \xrightarrow{w} \mathcal{N}(0, \rho^2),$$

where

$$G_{\varphi, \mathbf{t}}(\mathbf{X}_1, \dots, \mathbf{X}_n; \mathbf{Y}_1, \dots, \mathbf{Y}_n) := \frac{(n-m)!}{n!} \frac{\sum_{\mathbf{I}(m,n)} \varphi(Y_1, \dots, Y_m) \prod_{j=1}^m K\left(\frac{X_{i_j} - t_j}{h_n}\right)}{\sum_{\mathbf{I}(m,n)} \prod_{i=1}^m K\left(\frac{X_i - t_i}{h_n}\right)},$$

$$\rho^2 := \sum_{i=1}^m \sum_{j=1}^m \mathbb{1}_{\{t_i = t_j\}} \left[r_{jl}(\mathbf{t}) - (r^{(m)}(\mathbf{t}))^2 \right] \int K^2(u) du / f(t_i).$$

Theorem 2.1.4.2 *If the assumptions (A.1), (A.1') and (A.3') are satisfied, then for all-most all \mathbf{t}*

$$\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n) \rightarrow r^{(m)}(\varphi, \mathbf{t}) \quad \text{with probability 1.}$$

Much research has been dedicated to investigating the asymptotic properties of conditional U-statistics. [Sen \[1994\]](#) established rates of uniform strong convergence of $\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$ to $r^{(m)}(\varphi, \mathbf{t})$ under appropriate conditions while [Stute \[1994a\]](#) gave the almost sure convergence of $\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$ in p^{th} -norm, $p > 2$ for i.i.d random variables. [Stute \[1994b\]](#) showed that $\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$ is universally consistent. He also proved universal consistencies of the kernel and k_n -nearest neighbor estimators as two special cases of the conditional U-statistics. [Harel and Puri \[1996\]](#) extended the results of [Stute \[1991\]](#), under appropriate mixing conditions, to weakly dependent data and have applied their findings to verify Bayes risk consistency of the corresponding discrimination rules.

The paper of [Dony and Mason \[2008\]](#) is another work investigating the consistency of $\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$. More precisely, they established the uniform in bandwidth and the uniform in \mathbf{t} consistency $\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$. They suppose that φ and the kernel $K(\cdot)$ belong to some VC-type classes of functions. It is worth mentioning that the consistency rates obtained by [Dony and Mason \[2008\]](#) are better than those obtained by [Sen \[1994\]](#). In turn, the study of functional conditional U-statistics (when the explanatory variable takes its values in some functional space), has had its share of attention; [Fu \[2012\]](#) adopted the conditional U-statistic based on functional observations, i.e., He consider a sequence $\{(X_i, Y_i) : i \geq 1\}$ of independent and identically distributed random variables, where X_i take their values in some abstract space \mathcal{X} and Y in \mathbb{R}^p . Suppose that \mathcal{X} is endowed with a semi-metric $d(\cdot, \cdot)$. The proposed estimator is :

$$\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_K) = \frac{\sum_{(i_1, \dots, i_m) \in I(m, n)} \varphi(Y_{i_1}, \dots, Y_{i_m}) K\left(\frac{d(t_1, X_{i_1})}{h_K}\right) \dots K\left(\frac{d(t_m, X_{i_m})}{h_K}\right)}{\sum_{(i_1, \dots, i_m) \in I(m, n)} K\left(\frac{d(t_1, X_{i_1})}{h_K}\right) \dots K\left(\frac{d(t_m, X_{i_m})}{h_K}\right)}. \quad (2.1.2)$$

Under i.i.d [Fu \[2012\]](#) established the asymptotic finite-dimensional normality for (2.1.2). Similar to the contribution of [Dony and Mason \[2008\]](#), [Bouzebda and Nemouchi \[2020\]](#) studied a strong form of consistency of (2.1.2), that is, uniform consistency and uniform in bandwidth consistency, in addition to the consistency over some VC-type class of functions \mathcal{F} . Recently, [Bouzebda and Nemouchi \[2019\]](#) established central limit theorems for conditional U-processes of stationary mixing conditions, and they also investigated the weak convergence of the conditional U-processes with functional mixing data (cf. [Bouzebda and Nemouchi \[2022\]](#)). [Bouzebda et al. \[2021\]](#) also studied the uniform in bandwidth consistency of the general conditional U-statistics based on the copula representation.

Generally speaking, we may take for $\varphi(\cdot)$ any function found interesting in the unconditional setup; cf. [Serfling \[2009\]](#). As mentioned before, the case $m = 1$ leads to the Nadaraya-Watson estimator if we set $\varphi = Id$, the identity function; $\varphi = \mathbb{1}_{\{t, \leq t\}}$ yields the conditional d.f. evaluated at t , refer to [Stute \[1986\]](#). We now present several examples for $m = 2$. We suppose that (X_1, Y_1) and (X_2, Y_2) are independent.

Example 2.1.4.3 Let $\varphi(Y_1, Y_2) = Y_1 Y_2$, then :

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

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

Example 2.1.4.4 For :

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

we obtain:

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

Example 2.1.4.5 For $\varphi(Y_1, Y_2) = \mathbb{1}\{Y_1 + Y_2 > 0\}$ we obtain a conditional U-statistic which may be viewed as a conditional version of the Wilcoxon one-sample statistic. It may be used for testing the hypothesis that the conditional distribution at X_1 is symmetric at zero. Obviously :

$$r^{(2)}(\varphi, t_1) = \mathbb{P}(Y_1 + Y_2 > 0 \mid X_1 = t_1 = X_2)$$

Example 2.1.4.6 For $\varphi(Y_1, Y_2) = \mathbb{1}\{Y_1 \leq Y_2\}$:

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

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

Example 2.1.4.7 Assume $\{Y_i = (Y_{i,1}, Y_{i,2})^\top\}_{i=1,2}$ and define φ by :

$$\varphi(y_1, y_2) := \frac{1}{2}(y_{1,1}y_{1,2} + y_{2,1}y_{2,2} - y_{1,1}y_{2,2} - y_{1,2}y_{2,1}),$$

and

$$r^{(2)}(\varphi, t_1, t_2) = \frac{1}{2} \{ \mathbb{E}(Y_{1,1}Y_{1,2} \mid X_1 = t_1) + \mathbb{E}(Y_{2,1}Y_{2,2} \mid X_2 = t_2) \\ - \mathbb{E}(Y_{1,1}Y_{2,2} \mid X_1 = t_1, X_2 = t_2) - \mathbb{E}(Y_{1,2}Y_{2,1} \mid X_1 = t_1, X_2 = t_2) \}$$

In particular:

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

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

Example 2.1.4.8 For $m = 3$, let :

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

We have

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

Moreover, the corresponding CU – S can be considered a conditional analog of the Hollander-Proschan test-statistic (*Hollander and Proschan [1972]*). It may be used to test the hypothesis that the conditional distribution of Y_1 given $X_1 = t$, is exponential against the alternative that it is of the New-Better than-Used-type.

Example 2.1.4.9 Let :

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

and for $m = 5$ define :

$$\varphi(Y_1, \dots, Y_5) = \frac{1}{4} \psi(Y_1, Y_2, Y_3) \psi(Y_1, Y_4, Y_5) \\ \times \psi(Y_1, Y_2, Y_3) \psi(Y_1, Y_4, Y_5)$$

We have :

$$r^{(5)}(\varphi, t_1 = t_2 = t_3 = t_4 = t_5 = t) = \mathbb{E}(\varphi(Y_1, \dots, Y_5) \mid X_1 = X_2 = X_3 = X_4 = X_5 = t)$$

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

Example 2.1.4.10 Let $\widehat{Y_1 Y_2}$ 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 :

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

Silverman [1978] has used this kernel to propose a U -process to test uniformity on the circle. Let

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

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

Example 2.1.4.11 Let us suppose $\{Z_1 = (X_1, Y_1), Z_2 = (X_2, Y_2)\}$ two random variables, if we want to test the symmetrization about zero of the conditional distribution at $X_1 = X_2 = t_1$, we often use so-called Wilcoxon one sample statistic. In this case, the U -statistic :

$$\widehat{r}_2^{(2)}(\varphi, t_1, t_2) = \frac{1}{2} \frac{\mathbb{1}_{\{Y_1 + Y_2 > 0\}} K\left(\frac{X_1 - t_1}{h}\right) K\left(\frac{X_2 - t_1}{h}\right)}{K\left(\frac{X_1 - t_1}{h}\right) K\left(\frac{X_2 - t_1}{h}\right)},$$

can be viewed as a conditional version of the Wilcoxon one-sample statistic, with the kernel:

$$\mathbb{1}_{\{Y_1 + Y_2 > 0\}} K\left(\frac{X_1 - t_1}{h}\right) K\left(\frac{X_2 - t_1}{h}\right),$$

for

$$\varphi(y_1, y_2) = \mathbb{1}_{\{y_1 + y_2 > 0\}}.$$

2.1.5 Useful tools : Hoeffding's decomposition

Hoeffding's decomposition can be considered the primary tool for understanding the asymptotic behavior of U -statistics. This section is based on Serfling's book. Let's suppose that the symmetric kernel function $\varphi(\cdot)$ satisfies $\mathbb{E}(\varphi(Z_1, \dots, Z_m)) < 0$, and define some associated functions :

$$\varphi_k(z_1, \dots, z_k) = \mathbb{E}(\varphi(z_1, \dots, z_k, Z_{k+1}, \dots, Z_m)),$$

for $1 \leq k \leq m-1$ with $\varphi_m = \varphi$. It is clear that $\varphi_k(\cdot)$ is the conditional expectation of $\varphi(Z_1, \dots, Z_m)$ given Z_1, \dots, Z_k , i.e.,

$$\varphi_k(z_1, \dots, z_k) = \mathbb{E}(\varphi(Z_1, \dots, Z_m) | Z_1 = z_1, \dots, Z_k = z_k).$$

Moreover, for $1 \leq c \leq m-1$:

$$\varphi_k(z_1, \dots, z_k) = \mathbb{E}(\varphi_{k+1}(z_1, \dots, z_k, Z_{k+1})),$$

the functions $\varphi_k(\cdot)$ being introduced, we will need in the following to center them and define new functions based on the centered $\varphi_k(\cdot)$. Recall that we have :

$$\theta = \mathbb{E}[\varphi(Z_1, \dots, Z_m)],$$

then, the centered $\varphi_k(\cdot)$ functions :

$$\begin{cases} \varphi^{(c)} = \varphi - \theta, \\ \varphi_k^{(c)} = \varphi_k - \theta \quad \text{for } 1 \leq k \leq m. \end{cases}$$

Set now :

$$\begin{aligned} \tilde{\varphi}_1(z_1) &:= \varphi_1^{(c)}(z_1), \\ \tilde{\varphi}_2(z_1, z_2) &:= \varphi_2^{(c)}(z_1, z_2) - \tilde{\varphi}_1(z_1) - \tilde{\varphi}_1(z_2), \\ \tilde{\varphi}_3(z_1, z_2, z_3) &:= \varphi_3^{(c)}(z_1, z_2, z_3) - \sum_{i=1}^3 \tilde{\varphi}_1(z_i) - \sum_{1 \leq i_1 \neq i_2 \leq 3} \tilde{\varphi}_2(z_{i_1}, z_{i_2}), \\ &\vdots \\ \tilde{\varphi}_m(z_1, \dots, z_m) &:= \varphi_m^{(c)}(z_1, \dots, z_m) - \sum_{i=1}^m \tilde{\varphi}_1(z_i) - \sum_{1 \leq i_1 \neq i_2 \leq m} \tilde{\varphi}_2(z_{i_1}, z_{i_2}) - \dots - \\ &\quad - \dots - \sum_{I(m,n)} \tilde{\varphi}_{m-1}(z_{i_1}, \dots, z_{i_{m-1}}). \end{aligned}$$

The $\tilde{\varphi}_k$'s functions are symmetric in their arguments, and we can easily check that :

$$\begin{aligned} \mathbb{E}(\tilde{\varphi}_1(z_1)) &= 0, \\ \mathbb{E}(\tilde{\varphi}_2(z_1, z_2)) &= 0, \\ \mathbb{E}(\tilde{\varphi}_3(z_1, z_2, z_3)) &= 0, \\ &\vdots \\ \mathbb{E}(\tilde{\varphi}_m(z_1, \dots, z_m)) &= 0. \end{aligned}$$

Define the sums of centered φ_k 's and $\tilde{\varphi}_k$'s by :

$$\begin{aligned} S_n(\varphi) &:= \sum_{I(m,n)} \varphi^{(c)}(Z_{i_1}, \dots, Z_{i_m}), \\ S_{nk}(\varphi) &:= \sum_{I(k,n)} \tilde{\varphi}_k(Z_{i_1}, \dots, Z_{i_k}), \quad \text{for } 1 \leq k \leq m. \end{aligned}$$

Then we can write :

$$u_n^{(m)}(\varphi) - \theta = \frac{(n-m)!}{n!} S_n(\varphi).$$

We are now equipped to state Hoeffding's theorem.

Theorem 2.1.5.1 (Hoeffding [1948]) *Let the symmetric kernel function $\varphi(\cdot)$ satisfies $\mathbb{E}(\varphi(Z_1, \dots, Z_m)) < \infty$, then*

$$u_n^{(m)}(\varphi) - \theta = \sum_{k=1}^m \binom{m}{k} \binom{n}{k}^{-1} S_{nk}(\varphi).$$

Moreover, for $1 \leq k \leq m$:

$$\mathbb{E}(S_{nk} | Z_1, \dots, Z_c) = S_{kc} \quad k \leq c \leq n.$$

Thus, the sequence $\{S_{nk}, \sigma(Z_1, \dots, Z_n)\}_{n \geq k}$ is a forward martingale.

2.2 References

- Arcones, M. A. and Yu, B. (1994). Central limit theorems for empirical and U -processes of stationary mixing sequences. *Journal of Theoretical Probability*, **7**(1), 47–71. [25](#)
- Athreya, K. B. and Lahiri, S. N. (2006). *Measure theory and probability theory*, volume 19. Springer. [21](#)
- Bouzebda, S. and Nemouchi, B. (2019). Central limit theorems for conditional empirical and conditional U -processes of stationary mixing sequences. *Math. Methods Statist.*, **28**(3), 169–207. [29](#)
- 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**(2), 452–509. [29](#)
- Bouzebda, S. and Nemouchi, B. (2022). Weak-convergence of empirical conditional processes and conditional u -processes involving functional mixing data. *Statistical Inference for Stochastic Processes*, pages 1–56. [29](#)
- 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**(2), 321–358. [29](#)
- Dony, J. and Mason, D. M. (2008). Uniform in bandwidth consistency of conditional U -statistics. *Bernoulli*, **14**(4), 1108–1133. [29](#)
- Fu, K.-A. (2012). An application of U -statistics to nonparametric functional data analysis. *Comm. Statist. Theory Methods*, **41**(9), 1532–1542. [29](#)
- Harel, M. and Puri, M. L. (1996). Conditional U -statistics for dependent random variables. *J. Multivariate Anal.*, **57**(1), 84–100. [28](#)
- Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics*, **19**, 293–325. [32](#)
- Hollander, M. and Proschan, F. (1972). Testing whether new is better than used. *Ann. Math. Statist.*, **43**, 1136–1146. [30](#)
- Kolmogorov, A. N. and Tihomirov, V. M. (1961). ϵ -entropy and ϵ -capacity of sets in functional space. *Amer. Math. Soc. Transl. (2)*, **17**, 277–364. [26](#)
- Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York. [21](#), [25](#), [26](#)
- Sen, A. (1994). Uniform strong consistency rates for conditional U -statistics. *Sankhyā Ser. A*, **56**(2), 179–194. [28](#), [29](#)

- Serfling, R. J. (2009). *Approximation theorems of mathematical statistics*, volume 162. John Wiley & Sons. 29
- Silverman, B. W. (1978). Distances on circles, toruses and spheres. *J. Appl. Probability*, **15**(1), 136–143. 31
- Stute, W. (1986). Conditional empirical processes. *Ann. Statist.*, **14**(2), 638–647. 29
- Stute, W. (1991). Conditional U-statistics. *Ann. Probab.*, **19**(2), 812–825. 27, 28
- Stute, W. (1994a). L^p -convergence of conditional U-statistics. *J. Multivariate Anal.*, **51**(1), 71–82. 28
- Stute, W. (1994b). Universally consistent conditional U-statistics. *Ann. Statist.*, **22**(1), 460–473. 28
- van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics. 21, 25, 26

Chapter 3

Uniform consistency and uniform in number of neighbors consistency for functional conditional U-statistics

Ce chapitre développe le contenu d'un article publié [Japanese Journal of Statistics and Data Science (2022). <https://doi.org/10.1007/s42081-022-00161-3>], mis en forme pour être inséré dans le présent manuscrit de thèse.

Title : *Uniform consistency and uniform in number of neighbors consistency for nonparametric regression estimates and conditional U-statistics involving functional data.*

abstract

U-statistics represent a fundamental class of statistics arising from modeling 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 X . W. Stute [Ann. Probab. 19 (1991) 812–825] 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:

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

In the present paper, we introduce the k nearest neighborhoods estimator of the conditional U-statistics depending on an infinite-dimensional covariate. A sharp uniform in the number of neighborhoods (UINN) limit law for the proposed estimator is presented. Such results allow the NN to vary within a complete range for which the estimator is consistent. Consequently, it represents an interesting guideline in practice to select the optimal NN in nonparametric functional data analysis. In addition, uniform consistency is also established over $\varphi \in \mathcal{F}$ for a suitably restricted class \mathcal{F} , in both cases bounded and unbounded, satisfying some moment conditions and some mild conditions on the model. This paper unifies the approaches in some

other recent papers. As a by-product of our proofs, we state consistency results for the k -NN conditional U-statistics under the random censoring, which are uniform in the number of neighbors. The theoretical uniform consistency results established in this paper are (or will be) key tools for many further developments in functional data analysis.

Keywords : Uniform almost complete convergence ; Conditional U-statistics ; Functional data analysis ; Functional regression; Kolmogorov's entropy ; Small ball probability ; Uniform consistency ; Uniform in number of neighbors consistency ; k -NN estimator ; Data-driven estimator.

Mathematics Subject Classification: 62G05; 62G08; 62G20; 62G35; 62G07; 62G32; 62G30; Secondary: 62E20.

3.1 Introduction

U-statistics, first considered by [Halmos \[1946\]](#) in connection with unbiased statistics and formally introduced by [Hoeffding \[1948\]](#). The theory of U-statistics and U-processes has received considerable attention in the last decades due to its great number of applications and usefulness for solving complex statistical problems. Examples are density estimation, nonparametric regression tests, and goodness-of-fit tests. More precisely, U-processes appear in statistics in many instances, e.g., as the components of higher-order terms in von Mises expansions. In particular, U-statistics play a role in the analysis of estimators (including function estimators) with varying degrees of smoothness. For example, [Stute \[1993\]](#) applies the a.s. uniform bounds for \mathbb{P} -canonical U-processes to analyze 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 \[2007a,b\]](#), [Schick et al. \[2011\]](#) introduced new tests for normality which are based on the weighted L_1 -distances between the standard normal density and local U-statistics based on standardized observations. [Joly and Lugosi \[2016\]](#) discussed the estimation of the mean of the multivariate functions in the case of possibly heavy-tailed distributions and introduced the median-of-means, which is based on U-statistics. U-processes are important tools for a broad range of statistical applications such as testing for qualitative features of functions in nonparametric statistics [[Lee et al. \[2009\]](#), [Ghosal et al. \[2000\]](#), [Abrevaya and Jiang \[2005\]](#)], cross-validation for density estimation [Nolan and Pollard \[1987\]](#), and establishing limiting distributions of M-estimators [see, e.g., [Arcones and Giné \[1993\]](#), [Sherman \[1993\]](#), [Sherman \[1994\]](#), [de la Peña and Giné \[1999\]](#)]. Infinite-order U-statistics are useful tools for constructing simultaneous prediction intervals that quantify the uncertainty of ensemble methods such as sub-bagging and random forests. [Peng et al. \[2019\]](#) develop in great detail the notion of generalized U-statistics random forest predictions. The Mean NN approach estimation for differential entropy introduced by [Faivishevsky and Goldberger \[2008\]](#) is a particular of the U-statistic. Using U-statistics, [Liu et al. \[2016\]](#) proposed a new test statistic for goodness-of-fit tests. [Cléménçon \[2011\]](#) defined a measure by U-statistics to quantify the clustering quality of

a partition. **Halmos [1946]**, **v. Mises [1947]**, and **Hoeffding [1948]** provided (amongst others) the first asymptotic results for the case that the underlying random variables are independent and identically distributed. In this paper, we consider the so-called conditional U-statistics introduced by **Stute [1991]** when the covariates are functional. These statistics may be viewed as generalizations of the Nadaraya-Watson (**Nadaraja [1964]**, and **Watson [1964]**) estimates of a regression function. To better understand the problem, we first introduce Stute's estimators. Let us consider independent and identically distributed sequence of random elements $\{(\mathbf{X}_i, Y_i), i \in \mathbb{N}^*\}$ with $\mathbf{X}_i \in \mathbb{R}^d$ and $Y_i \in \mathcal{Y}$ some polish space and $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$. Let $\varphi : \mathcal{Y}^m \rightarrow \mathbb{R}$ be a measurable function. In this paper, we are primarily concerned with the estimation of the conditional expectation or regression function :

$$r^{(m)}(\varphi, \mathbf{t}) = \mathbb{E}(\varphi(Y_1, \dots, Y_m) | (\mathbf{X}_1, \dots, \mathbf{X}_m) = \mathbf{t}), \text{ for } \mathbf{t} \in \mathbb{R}^{dm}, \quad (3.1.1)$$

whenever it exists, i.e.,

$$\mathbb{E}(|\varphi(Y_1, \dots, Y_m)|) < \infty.$$

We now introduce a kernel function $K : \mathbb{R}^d \rightarrow \mathbb{R}$ with support contained in $[-B, B]^d$, $B > 0$, satisfying :

$$\sup_{\mathbf{x} \in \mathbb{R}^d} |K(\mathbf{x})| =: \kappa < \infty \text{ and } \int K(\mathbf{x}) d\mathbf{x} = 1. \quad (3.1.2)$$

Stute [1991] introduced a class of estimators for $r^{(m)}(\varphi, \mathbf{t})$, called conditional U-statistics, which is defined for each $\mathbf{t} \in \mathbb{R}^{dm}$ to be :

$$\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n) = \frac{\sum_{(i_1, \dots, i_m) \in I(m, n)} \varphi(Y_{i_1}, \dots, Y_{i_m}) K\left(\frac{\mathbf{t}_1 - \mathbf{X}_{i_1}}{h_n}\right) \dots K\left(\frac{\mathbf{t}_m - \mathbf{X}_{i_m}}{h_n}\right)}{\sum_{(i_1, \dots, i_m) \in I(m, n)} K\left(\frac{\mathbf{t}_1 - \mathbf{X}_{i_1}}{h_n}\right) \dots K\left(\frac{\mathbf{t}_m - \mathbf{X}_{i_m}}{h_n}\right)}, \quad (3.1.3)$$

where $\{h_n\}_{n \geq 1}$ is a sequence of positive constants converging to zero at the rate $nh_n^{dm} \rightarrow \infty$. 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 :

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

The work of **Sen [1994]** was devoted to estimate the rate of the uniform convergence in \mathbf{t} of $\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$ to $r^{(m)}(\varphi, \mathbf{t})$. In the paper of **Prakasa Rao and Sen [1995]**, the limit distributions of $\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$ 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 neighbor 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., $h_n, h_n \in [a_n, b_n]$ where $0 < a_n < b_n \rightarrow 0$ at some specific rate) of $\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$. In addition, uniform consistency is also established over $\varphi \in \mathcal{F}$ for a suitably restricted class \mathcal{F} , refer to [Bouzebda et al. \[2021\]](#). The main tool in their result is the use of the local conditional U-process investigated in [Giné and Mason \[2007a\]](#). Our main goal is to extend the work [Dony and Mason \[2008\]](#) to the functional setting; this generalization is far from being trivial, in the same spirit of [Bouzebda and Nemouchi \[2020\]](#).

Statistical problems concerned with the analysis of infinite-dimensional variables, such as curves, sets, and images, have known phenomenal growth and interest in literature for some years. The development of this research theme is indeed motivated by modern technological progress in measuring devices and the abundance of data measured on increasingly fine grids: this is the case, for example, in meteorology, medicine, satellite imagery, and many other areas of research. Thus, the statistical modeling of these data, comparable to random functions, has recently undergone intense development from theoretical aspects (related to the study of random variables with values in an infinite dimensional space) and applied point of view (implementation of estimators). Notice that the statistical analysis of functional data can date back to [Rao \[1958\]](#) where a set of growth curves was analyzed by functional principal component analysis in a rudimentary form for the first time, and subsequently rigorously investigated by [Kleffe \[1973\]](#) and [Dauxois et al. \[1982\]](#). Numerous recent work, from both theoretical and practical point of view, confirm the pertinence of investigating this area, for instance, refer to the monographs of [Bosq \[2000\]](#), [Ramsay and Silverman \[2005\]](#), [Ferraty and Vieu \[2006\]](#), [Horváth and Kokoszka \[2012\]](#), [Shi and Choi \[2011\]](#), [Bosq and Blanke \[2007\]](#), [Geenens \[2011\]](#), [Cuevas \[2014\]](#), [Shang \[2014\]](#), [Horváth and Rice \[2015\]](#), [Müller \[2016\]](#), [Nagy \[2017\]](#), [Vieu \[2018\]](#), [Aneiros et al. \[2019\]](#), and [Ling and Vieu \[2018\]](#), where we have collect these references. A recent review of functional data analysis is provided by [Goia and Vieu \[2016\]](#). Survey on recent advances in nonparametric FDA and also highlight the need for adaptative bandwidth in FDA. The general mathematical background for analyzing functional variables is developed in [Hsing and Eubank \[2015\]](#). In this article we focus on the almost complete¹ (a.co) uniform rates of consistency of the nonparametric functional regression model and functional conditional U-processes (statistics). More precisely, in this paper, we are interested in establishing the a.co uniform consistency and the a.co uniform in the number of neighborhoods (UINN) consistency of the nonparametric functional regression estimator and also the functional con-

¹Let $(z_n)_{n \in \mathbb{N}}$ be a sequence of real random variables. We say that (z_n) converges almost completely (a.co.) towards zero if for all $\epsilon > 0$,

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

The rate of convergence is of order u_n (with $u_n \rightarrow 0$) and we write $z_n = O_{\text{a.co.}}(u_n)$ if there exists $\epsilon > 0$ such that

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

This kind of convergence implies both almost-sure convergence and convergence in probability.

ditional U -processes (statistic). This is motivated by a series of papers, among many others, Einmahl and Mason [2000], Deheuvels and Mason [2004], Einmahl and Mason [2005], Dony and Einmahl [2009], Bouzebda and Elhattab [2011], Bouzebda [2012], Bouzebda *et al.* [2018], Bouzebda and El-hadjali [2020], Bouzebda and Nemouchi [2020], Bouzebda *et al.* [2021] the authors established uniform in bandwidth (UIB) consistency results for such estimators in the finite dimensional setting, where h_n varies within suitably chosen intervals indexed by n . In the FDA, several authors have been interested in studying nonparametric functional estimators; for example, Ferraty *et al.* [2010a] provided the consistency rates of some functionals of the conditional distribution, including the regression function, the conditional cumulative distribution, the conditional density, and some other ones, uniformly over a certain subset of the explicative variable. Kara-Zaitri *et al.* [2017] established the uniform consistency rate for some conditional models, including the regression function, the conditional distribution, the conditional density, and the conditional hazard function. The last mentioned paper is extended by Bouzebda and Nemouchi [2020]. Kara *et al.* [2017] and Almanjahie *et al.* [2022] established the almost complete convergence of the k -nearest neighbors (k -NN) estimators, which are uniform in the number of neighbors, under some classical assumptions on the kernel and on the small ball probabilities of the functional variable in connection with the entropy condition controlling the space complexity. Attouch *et al.* [2019] considered the problem of local linear estimation of the regression function when the covariate is functional and proved the strong uniform-in-bandwidth (UIB) convergence. Ling *et al.* [2019] investigated the k -NN estimation of the nonparametric regression model for strong mixing functional time series data and established the uniform a.co convergence rate of the k -NN estimator under some mild conditions. Novo *et al.* [2019] stated some new uniform asymptotic results for kernel estimates in the functional single-index model. Fu [2012] investigated the functional conditional U -statistic and established its finite-dimensional limiting law. Most of this literature focuses on UIB or UINN consistency or uniform consistency on some functional subset but never the both together. We aim to fill this gap in the literature by combining results from the FDA and the empirical processes theory. However, as will be seen later, the problem requires much more than ‘simply’ combining ideas from these two domains. In fact, delicate manipulation of several topological notions and establishing relations between them is needed. To our best knowledge, these problems form a basically unsolved open problem in the literature, giving the primary motivation for our paper. The present paper extends and complements our previous paper Bouzebda and Nemouchi [2020] in several, and the proofs for UINN are more involved; we will give more details in the forthcoming sections. Moreover, we have established new results and tools in the proof that are of independent interest.

The layout of present article is structured as follows. Section 3.2 is devoted to introducing the functional framework and the definitions we need in our work; we give the assumptions used in our asymptotic analysis with a short discussion. Section 3.3 provides the UINN consistency in the functional regression framework. Section 3.4 gives the paper’s main results concerning the UINN consistency with rates for the conditional U -statistics. In Section 3.5, we provide

an important application for the censored data setting that is of its own interest. In Section 3.6, we provide some applications including, the set indexed conditional U -statistics in §3.6.1, the multi-sample U processes in §3.6.2, the discrimination in §3.6.3, the metric learning in §3.6.4. In Section 3.7, we collect some examples of classes of functions and examples of conditional U -statistics. In Section 3.8, we briefly discussed the bandwidth selection criterion. Some concluding remarks and possible future developments are relegated to 6.1. To prevent interrupting the flow of the presentation, all proofs are gathered in Section 3.9. Finally, a few relevant technical results are given in the Appendix.

3.2 The functional framework

3.2.1 Generality on the model

We consider a sequence $\{X_i, Y_i : i \geq 1\}$ of independent and identically distributed random copies of the random vector functions (X, Y) , where X takes its values in some abstract space \mathcal{X} and Y in the abstract space \mathcal{Y} . Suppose that \mathcal{X} is endowed with a semi-metric $d(\cdot, \cdot)$ ² defining a topology to measure the proximity between two elements of \mathcal{X} and which is disconnected from the definition of X in order to avoid measurability problems. We are mainly interested in studying the uniform in the number of neighbors consistency of the following functional conditional U -statistic in the k -NN setting given for all $\mathbf{t} \in \mathcal{X}^m$ by:

$$\hat{r}_n^{*(m)}(\varphi, \mathbf{t}, \mathbf{h}_{n,k}(\mathbf{t})) = \frac{\sum_{(i_1, \dots, i_m) \in I(m, n)} \varphi(Y_{i_1}, \dots, Y_{i_m}) \mathbb{K}\left(\frac{d(t_1, X_{i_1})}{H_{n,k}(t_1)}\right) \dots \mathbb{K}\left(\frac{d(t_m, X_{i_m})}{H_{n,k}(t_m)}\right)}{\sum_{(i_1, \dots, i_m) \in I(m, n)} \mathbb{K}\left(\frac{d(t_1, X_{i_1})}{H_{n,k}(t_1)}\right) \dots \mathbb{K}\left(\frac{d(t_m, X_{i_m})}{H_{n,k}(t_m)}\right)}, \quad (3.2.1)$$

to the regression function $r^{(m)}(\varphi, \mathbf{t}) := \mathbb{E}(\varphi(Y_1, \dots, Y_m) \mid X_1, \dots, X_m = \mathbf{t})$, where :

$$\tilde{\mathbb{K}}(\mathbf{t}) = \prod_{i=1}^m \mathbb{K}(t_i), \quad \mathbf{t} = (t_1, \dots, t_m) \in \mathcal{X}^m = \mathcal{X} \times \dots \times \mathcal{X},$$

and $\varphi : \mathcal{Y}^m = \mathcal{Y} \times \dots \times \mathcal{Y} \rightarrow \mathbb{R}$ is a measurable function belonging to some class of functions \mathcal{F}_m , and $\mathbf{h}_{n,k}(\mathbf{t}) = (H_{n,k}(t_1), \dots, H_{n,k}(t_m))$ is a vector of positive random variables that depend on (X_1, \dots, X_n) such that, for all $\mathbf{x} = (x_1, \dots, x_m) \in \mathcal{X}^m$ and $j = 1, \dots, m$:

$$H_{n,k}(x_j) = \min \left\{ h \in \mathbb{R}^+ : \sum_{i=1}^n \mathbb{1}_{B(x_j, h)}(X_i) = k \right\}, \quad (3.2.2)$$

where $B(t, r) = \{z \in \mathcal{X} : d(z, t) \leq r\}$ is a ball in \mathcal{X} with the center $t \in \mathcal{X}$ and radius r , and $\mathbb{1}_A$ is the indicator function of the set A . Notice that the estimator (3.2.1) is the first time introduced and investigated in the present paper. In fact this k -NN estimate can be considered as an extension to random and locally adaptive neighborhood of the functional conditional U -statistics

²A semi-metric (sometimes called pseudo-metric) $d(\cdot, \cdot)$ is a metric which allows $d(x_1, x_2) = 0$ for some $x_1 \neq x_2$.

estimate of $r^{(m)}(\varphi, \mathbf{t})$ defined for all $t \in \mathcal{X}^m$ as :

$$\hat{r}_n^{(m)}(\varphi, \mathbf{t}, h_{\mathbb{K}}(\mathbf{t})) = \frac{\sum_{(i_1, \dots, i_m) \in \mathcal{I}(m, n)} \varphi(Y_{i_1}, \dots, Y_{i_m}) \mathbb{K}\left(\frac{d(t_1, X_{i_1})}{h_{\mathbb{K}}(t_1)}\right) \cdots \mathbb{K}\left(\frac{d(t_m, X_{i_m})}{h_{\mathbb{K}}(t_m)}\right)}{\sum_{(i_1, \dots, i_m) \in \mathcal{I}(m, n)} \mathbb{K}\left(\frac{d(t_1, X_{i_1})}{h_{\mathbb{K}}(t_1)}\right) \cdots \mathbb{K}\left(\frac{d(t_m, X_{i_m})}{h_{\mathbb{K}}(t_m)}\right)}, \quad (3.2.3)$$

where $h_{\mathbb{K}}(\mathbf{t}) = (h_{\mathbb{K}, n}(t_1), \dots, h_{\mathbb{K}, n}(t_m)) =: (h_{\mathbb{K}}(t_1), \dots, h_{\mathbb{K}}(t_m))$ are positive real numbers decreasing to zero as n goes to infinity. The k -NN method is a fundamental statistical method presenting several advantages. In general, the method is computationally fast without the requirement of too much parameter tuning. One of the main features is that the k -NN methods are purely nonparametric, meaning they can automatically adapt to any continuous underlying distributions without relying on specific models. Finally, for important statistical problems, including, among many others, density estimation, classification, and regression, k -NN methods are all proven to be consistent as long as a proper k is selected. Historically, the k -NN was first introduced by [Fix and Hodges \[1951\]](#)—see also [Fix and Hodges \[1989\]](#)—in the context of nonparametric discrimination, and further investigated by [Loftsgaarden and Quesenberry \[1965\]](#), for more details, we refer to [Biau and Devroye \[2015\]](#). [Bouzebda and Nemouchi \[2020\]](#) studied the uniform in \mathbf{t} and the uniform in bandwidth consistency of the estimator (3.2.3) where $h_{\mathbb{K}}(\mathbf{t}) \in [a_n, b_n]$ where $a_n \rightarrow 0$ and $b_n \rightarrow 0$. Our main interest is to study the uniform consistency when the bandwidth $h_{\mathbb{K}}(\mathbf{t})$ is adaptive to the data X_1, \dots, X_n and thus depending on k and \mathbf{t} . The randomness of the bandwidths $H_{n,k}(t)$ is its major advantage and disadvantage since it creates some technical difficulties in proofs, see Remark 3.4.2.7. To solve these difficulties, we will generalize, in several directions, the lemma of [Burba et al. \[2009\]](#), which was studied in many articles in the case of nonparametric functional regression. More precisely, we will extend it in a non-trivial way to the functional conditional U -statistics setting by the uniformity in φ , \mathbf{t} , and k simultaneously. This situation is not investigated before.

The main purpose is to prove the uniform consistency, the uniform in number of neighbors consistency and in the class of functions, for certain classes of functions of the estimator $\hat{r}_n^{*(m)}(\varphi, \mathbf{t}, \mathbf{h}_{n,k}(\mathbf{t}))$ to $r^{(m)}(\varphi, \mathbf{t})$. It should be noted that since our objective is to generalize the results obtained for the estimator defined in (3.2.3), and given the fact that one of the main differences is that the smoothing parameter, $\mathbf{h}_{n,k}(\mathbf{t})$ is a vector of random variables instead of a univariate parameter $h_{\mathbb{K}}$ our first course of action would be to extend the results of [Bouzebda and Nemouchi \[2020\]](#) the UIB results to the multivariate setting. First, we need to introduce some notation. Let $\mathcal{F}_m = \{\varphi : \mathcal{Y}^m \rightarrow \mathbb{R}\}$ denote a pointwise measurable class of real valued symmetric functions on \mathcal{Y}^m with a measurable envelope function :

$$F(\mathbf{y}) \geq \sup_{\varphi \in \mathcal{F}_m} |\varphi(\mathbf{y})|, \text{ for } \mathbf{y} \in \mathcal{Y}^m. \quad (3.2.4)$$

For a kernel function $\mathbb{K}(\cdot)$ and a subset $\mathcal{S}_{\mathcal{X}} \subset \mathcal{X}$, we define the pointwise measurable class of functions, for $1 \leq m \leq n$:

$$\mathcal{K}^m := \left\{ (x_1, \dots, x_m) \mapsto \prod_{i=1}^m \mathbb{K}\left(\frac{d(x_i, t_i)}{h_i}\right), (h_1, \dots, h_m) \in \mathbb{R}_+^m \setminus \{0\} \text{ and } (t_1, \dots, t_m) \in \mathcal{S}_{\mathcal{X}}^m \right\}.$$

For a certain interval $\mathcal{H}_n^{(m)} \subset \mathbb{R}_+^m \setminus \{0\}$, we denote

$$\mathcal{H}_n^{(m)} := \prod_{j=1}^m (a_{n,j}, b_{n,j}),$$

where

$$0 < a_{n,j} < b_{n,j} \text{ and } \lim_{n \rightarrow \infty} a_{n,j} = \lim_{n \rightarrow \infty} b_{n,j} = 0, \forall j = 1, \dots, m.$$

In the sequel, we denote (unless stated otherwise) :

$$\tilde{a}_n = \min_{1 \leq j \leq m} a_{n,j} \text{ and } \tilde{b}_n = \max_{1 \leq j \leq m} b_{n,j}.$$

We also note for all $\mathbf{b} = (b_1, \dots, b_m) \in (0, 1)^m$:

$$\mathcal{H}_0^{(m)} := \prod_{j=1}^m (a_{n,j}, b_j),$$

and

$$\tilde{b}_0 := \max_{1 \leq j \leq m} b_j.$$

For notational convenience, in the case of $m = 1$, we will be noting $\mathcal{H}_n^{(1)}$ and $\mathcal{H}_0^{(1)}$ simply \mathcal{H}_n and \mathcal{H}_0 . The same goes for other similar notations unless stated otherwise. We need to introduce some concepts that are related to the topological structure of functional spaces. First, we define the small-ball probability for a fixed $t \in \mathcal{X}$ and for all $r > 0$ by:

$$\mathbb{P}(X \in B(t, r)) =: \phi_t(r), \quad (3.2.5)$$

this notion is widely used in nonparametric functional data analysis to avoid introducing density assumptions on the functional variable X and address the issues associated with the infinite-dimensional nature of the functional spaces. At this point, we can refer to [Gasser *et al.* \[1998\]](#), [Masry \[2005\]](#) and [Ferraty and Vieu \[2006\]](#). Furthermore, taking into account the uniformity aspect of our study, we consider for a fixed subset $\mathcal{S}_{\mathcal{X}}$ of \mathcal{X} the following assumptions.

(A.1) The small ball probability ϕ_t is such that

(A.1.a) There exist constants $0 < C_1 \leq C_2 < \infty$, an invertible function $\phi : \mathbb{R}^+ \rightarrow (0, \infty)$ such that, for all $\mathbf{t} = (t_1, \dots, t_m) \in \mathcal{S}_{\mathcal{X}}^m$ and $\mathbf{h}(\mathbf{t}) = (h_1(t_1), \dots, h_m(t_m))$, we have

$$0 < C_1 \tilde{\phi}(\mathbf{h}(\mathbf{t})) \leq \phi_{\mathbf{t}}(\mathbf{h}(\mathbf{t})) = \phi_{t_1}(h_1(t_1)) \times \dots \times \phi_{t_m}(h_m(t_m)) \leq C_2 \tilde{\phi}(\mathbf{h}(\mathbf{t})) < \infty,$$

where $\tilde{\phi}(\mathbf{h}(\mathbf{t})) := \prod_{j=1}^m \phi(h_j(t_j))$, $\phi(0) = 0$ and $\phi(u)$ is absolutely continuous in a neighborhood of the origin. Moreover, the function $\phi(\cdot)$ satisfies, for a sequence of positive real numbers h_K belonging to $(\tilde{a}_n, \tilde{b}_n)$, we have: $n\phi(h_K) \rightarrow \infty$ as $n \rightarrow \infty$ and

$$\exists A > 0, \exists r_0 > 0 \text{ such that } \forall r < r_0 : \phi'(r) < A.$$

(A.1.b) For $h_K \in (\tilde{a}_n, \tilde{b}_n)$, there exist constants $0 < C_3 \leq C_4 < \infty$ such that, with n large enough,

$$C_3 \leq \frac{\phi(h_K/2)}{\phi(h_K)} \leq C_4.$$

Notice that the function $\tilde{\phi}(\cdot)$ allows us to control the concentration of the probability measure of the functional variable X on small balls of different radius and different centers. This allows us to get more information on the local structure of the data. Therefore, the study of the uniform consistency is done using the function $\tilde{\phi}(\cdot)$ (or simply $\phi(\cdot)$, in the case of $m = 1$) instead of the concentration function $\phi_t(\cdot)$ that only controls the measure of probability on a small ball.

We also need to deal with the VC-subgraph classes (“VC” for Vapnik and Červonenkis).

Definition 3.2.1.1 A class of subsets \mathcal{C} on a set C is called a VC-class if there exists a polynomial $P(\cdot)$ such that, for every set of N points in C , the class \mathcal{C} picks out at most $P(N)$ distinct subsets.

Definition 3.2.1.2 A class of functions \mathcal{F} is called a VC-subgraph class if the graphs of the functions in \mathcal{F} form a VC class of sets, that is, if we define the subgraph of a real-valued function f on S as the following subset \mathcal{G}_f on $S \times \mathbb{R}$:

$$\mathcal{G}_f = \{(s, t) : 0 \leq t \leq f(s) \text{ or } f(s) \leq t \leq 0\}$$

the class $\{\mathcal{G}_f : f \in \mathcal{F}\}$ is a VC-class of sets on $S \times \mathbb{R}$. Informally, a VC class of functions is characterized by having a polynomial covering number (the minimal number of required functions to make a covering on the entire class of functions).

3.2.2 Conditions and comments

To prove the uniform consistency and the UINN consistency of the functional regression and conditional U -statistic in the k -NN setting, some conditions and results of the UIB ideas developed in Bouzebda and Nemouchi [2020] are needed. Therefore we will be stating them in the following subsection, together with the additional conditions needed for UINN consistency.

(A.2.) For n large enough and for some $\beta > 1$, the Kolomogorov’s entropy satisfies :

$$\frac{(\log n)^2}{n\phi(h_K)} \leq m\psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right) \leq \frac{n}{\log n} \phi(h_K), \quad (3.2.6)$$

$$\sum_{n=1}^{\infty} \exp \left\{ m(1 - \beta)\psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right) \right\} < \infty, \quad (3.2.7)$$

where $\psi_{\mathcal{S}_{\mathcal{X}}}(\varepsilon) := \log N_{\varepsilon}(\mathcal{S}_{\mathcal{X}})$, and $N_{\varepsilon}(\mathcal{S}_{\mathcal{X}})$ is the minimal number of open balls of radius ε in \mathcal{X} , needed to cover $\mathcal{S}_{\mathcal{X}}$.

(A.3.) Suppose that the regression function satisfies for the elements $\varphi(\cdot)$ of the class \mathcal{F} , and for $1 \leq m \leq n$:

$$\exists \gamma > 0, \forall \mathbf{t}_1, \mathbf{t}_2 \in \mathcal{V}_t: |r^{(m)}(\varphi, \mathbf{t}_1) - r^{(m)}(\varphi, \mathbf{t}_2)| \leq C_5 d_{\mathcal{X}^m}^\gamma(\mathbf{t}_1, \mathbf{t}_2),$$

where for $1 \leq i \leq 2$: $\mathbf{t}_i := (t_{i1}, \dots, t_{im})$, \mathcal{V}_t denotes a fixed neighborhood of \mathbf{t} in $\mathcal{S}_{\mathcal{X}}^m$ and for $\mathbf{x} = (x_1, \dots, x_m), \mathbf{y} = (y_1, \dots, y_m) \in \mathcal{X}^m$:

$$d_{\mathcal{X}^m}(\mathbf{x}, \mathbf{y}) := \frac{1}{m} d(x_1, y_1) + \dots + \frac{1}{m} d(x_m, y_m).$$

(A.4.) The classes of functions \mathcal{F}_m and \mathcal{K}^m are such that:

(A.4.a) The class of functions \mathcal{F}_m is bounded and its envelope function satisfy for some $0 < M < \infty$:

$$F(\mathbf{y}) \leq M, \quad \mathbf{y} \in \mathcal{Y}^m$$

(A.4.b) The kernel functions $K(\cdot)$ is Lipschitz, supported within $(0, 1/2)$ and there exists some constants $0 < \kappa_1 \leq \kappa_2 < \infty$, such that:

$$\int K(x) dx = 1,$$

and:

$$0 < \kappa_1 \mathbb{1}_{(0,1/2)}(\cdot) \leq K(\cdot) \leq \kappa_2 \mathbb{1}_{(0,1/2)}(\cdot).$$

(A.5.) The classes of functions \mathcal{F}_m and \mathcal{K}^m are such that:

$$\sup_Q \int_0^1 \sqrt{1 + \log N(\epsilon, \mathcal{F}_m, \|\cdot\|_{L_2(Q)})} d\epsilon < \infty, \quad (3.2.8)$$

$$\sup_Q \int_0^1 \sqrt{1 + \log N(\epsilon, \mathcal{K}^m, \|\cdot\|_{L_2(Q)})} d\epsilon < \infty, \quad (3.2.9)$$

where the supremum is taken over all probability measures Q on the probability space (Ω, \mathcal{A}) , with $Q(F^2) < \infty$, and for each class $N(\epsilon, \cdot, \|\cdot\|_{L_2(Q)})$ is the covering number respect to $L_2(Q)$ -norm.

(A.5'.) The class of functions $\mathcal{F}_m \mathcal{K}^m$ is supposed to be of VC-type with envelope function previously defined. Hence, there are two finite constants b and α such that:

$$N(\epsilon, \mathcal{F}_m \mathcal{K}^m, \|\cdot\|_{L_2(Q)}) \leq \left(\frac{b \|F \mathcal{K}^m\|_{L_2(Q)}}{\epsilon} \right)^\alpha \quad (3.2.10)$$

for any $\epsilon > 0$ and each probability measure such that $Q(F)^2 < \infty$.

(A.6) The sequences $\{\tilde{a}_n\}$ and $\{\tilde{b}_n\}$ (resp. $\{a_n\}$ and $\{b_n\}$) verify

$$\tilde{b}_n \rightarrow 0 \quad \text{and} \quad \frac{(\log n/n)}{\min\{\tilde{a}_n^2, \phi^2(\tilde{a}_n)\}} \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (3.2.11)$$

(A.7.) There exist sequences $\{\rho_{n,1}, \dots, \rho_{n,m}\} \subset (0, 1)^m$, $\{k_{1,n}\} \subset \mathbb{Z}^+$ and $\{k_{2,n}\} \subset \mathbb{Z}^+$ ($k_{1,n} \leq k \leq k_{2,n}$) and constants $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)$ and $\mathbf{v} = (v_1, \dots, v_m)$, such that

$$0 < \mu_j \leq v_j < \infty, \text{ for all } j = 1, \dots, m \text{ (and we note } \boldsymbol{\mu} \leq \mathbf{v}\text{),}$$

and

$$\mu_j \phi^{-1} \left(\frac{\rho_{n,j} k_{1,n}}{n} \right) \leq \phi_{t_j}^{-1} \left(\frac{\rho_{n,j} k_{1,n}}{n} \right) \text{ and } \phi_{t_j}^{-1} \left(\frac{k_{2,n}}{\rho_{n,j} n} \right) \leq v_j \phi^{-1} \left(\frac{k_{2,n}}{\rho_{n,j} n} \right), \quad (3.2.12)$$

$$\phi^{-1} \left(\frac{k_{2,n}}{\rho_{n,j} n} \right) \rightarrow 0, \quad (3.2.13)$$

$$\min \left\{ \frac{1 - \rho_{n,j}}{4} \frac{k_{1,n}}{\ln n}, \frac{(1 - \rho_{n,j})^2}{4\rho_{n,j}} \frac{k_{1,n}}{\ln n} \right\} > 2, \quad (3.2.14)$$

$$\frac{(\log n/n)}{\min \left\{ \mu_j \phi^{-1} \left(\frac{\rho_{n,j} k_{1,n}}{n} \right), \phi \left(\mu_j \phi^{-1} \left(\frac{\rho_{n,j} k_{1,n}}{n} \right) \right) \right\}} \rightarrow 0. \quad (3.2.15)$$

3.2.3 Comments on the conditions

Functional nonparametric models pose significant theoretical issues that can not be easily solved using traditional statistical approaches, particularly in a double infinite-dimensional context. To deal with this challenge, some conditions utilizing the different characteristics of the infinite-dimensional space are imposed. The condition (A.1.a) on small-ball probabilities is standard; it can be considered as an extension to the multivariate context, where we assume that the density function of the variable X is strictly positive. However, a more general condition can be used:

(A.1'.a) For all $\mathbf{t} = (t_1, \dots, t_m) \in \mathcal{S}_{\mathcal{X}}^m$, and $\mathbf{h} = (h_1, \dots, h_m) \in \mathbb{R}_+^m$ we have :

$$0 < C_1 \prod_{i=1}^m \phi(h_i) g(t_i) \leq \phi_{\mathbf{t}}(\mathbf{h}(\mathbf{t})) = \phi_{t_1}(h_1(t_1)) \times \dots \times \phi_{t_m}(h_m(t_m)) \leq C_2 \prod_{i=1}^m \phi(h_i) g(t_i) < \infty,$$

where $g(\cdot)$ is a non-negative function on $\mathcal{S}_{\mathcal{X}}$.

Nevertheless, using (A.1'.a) results in a complicated computation, which is why we retain the condition (A.1.a) to get the required results. Also, it is worth mentioning that, if in particular, we denote $\mathbf{h}_{\mathcal{K}} := (h_{\mathcal{K}}, \dots, h_{\mathcal{K}}) \in (\tilde{a}_n, \tilde{b}_n)^m$, we can find two positive constants C'_1, C'_2 , such that

$$0 < C'_1 \phi(h_{\mathcal{K}}) \leq \phi_{\mathbf{t}}(\mathbf{h}_{\mathcal{K}}) \leq C'_2 \phi(h_{\mathcal{K}}) < \infty, \quad (3.2.16)$$

which is similar to condition (C.1) used in Bouzebda and Nemouchi [2020], so we will deal with $\phi(h_{\mathcal{K}})$ instead of $\phi^m(h_{\mathcal{K}})$, (whenever we encounter a similar situation in the proofs). This approach is not only for notational purposes but also to make it easier to bridge the UIB and the UINN results.

It is worth noting that, in general, when we deal with functional data, we need some information about the variability of the small-ball probability to adapt to the bias of nonparametric estimators; this information is usually obtained by supposing that:

(A.1'.b)

$$\forall u \in [0, 1]: \quad \lim_{r \rightarrow \infty} \frac{\Phi_t(ur)}{\Phi_t(r)} = \lim_{r \rightarrow \infty} \mathbb{P}(d(\mathbf{X}, t) \leq ur | d(\mathbf{X}, t) \leq r) =: \tau_t(u) < \infty,$$

assumption (A.1.b) is a weaker version of (A.1'.b) and will be mainly used in the proofs. On the other hand, condition (A.2.) takes into account the topological considerations by controlling the Kolmogorov entropy³ of the set $\mathcal{S}_{\mathcal{X}}^m$, which is standard in nonparametric models when we study the uniform consistency and the uniform in bandwidth consistency, we refer to Ferraty *et al.* [2010a] and Kudraszow and Vieu [2013] for discussions. As mentioned in Kudraszow and Vieu [2013], there are special cases of functional spaces \mathcal{X} and subsets $S_{\mathcal{X}}$ where $\psi_{S_{\mathcal{X}}}(\log(n)/n) \gg \log(n)$. Some examples are the closed ball in a Sobolev space, the unit ball of the Cameron-Martin space, and a compact subset in a Hilbert space with a projection semi-metric (see Kolmogorov and Tihomirov [1959], van der Vaart and van Zanten [2007] and Ferraty *et al.* [2010b], respectively, for further details). In all these cases, it is easy to see that (A.2.) is verified as soon as $\beta > 2$. (A.3.) is another classical condition that is imposed on the model and that enables us to establish precise rates of consistency. The uniform integral entropy criteria described in (A.5.) are less restrictive than the VC-type classes of functions requirement (see Einmahl and Mason [2005]) and allow us to characterize the empirical process moments. Even so, to investigate the uniform consistency, the uniform in bandwidth consistency, and later the uniform in the number of neighbors consistency for the functional U-statistic/process, we have to employ specific procedures and inequalities that need the class of functions to be of VC-type, which is why we add the condition (A.5'.). However, we should keep in mind that every VC-type class of functions satisfies the condition (A.5.), for instance, see [Nolan and Pollard, 1987, Lemma 22], [Dudley, 2014, §4.7.], [van der Vaart and Wellner, 1996, Theorem 2.6.7], [Kosorok, 2008, §9.1] provide a number of sufficient conditions under which (A.5'.) holds, we may refer also to Bouzebda [2012] and [Deheuvels, 2011, §3.2] for further discussions. The condition (A.4.a) is used when dealing with bounded functions, which means that we will be replacing it when dealing with an unbounded class of functions by:

(A.4.a') The class of functions \mathcal{F}_m is unbounded and its envelope function satisfy for some $p > 2$:

$$\theta_p := \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \mathbb{E}(F^p(\mathbf{Y}) | \mathbf{X} = \mathbf{t}) < \infty.$$

Assumption (A.6) is essential to establish the rates of convergence (consistency) of the estimator defined in (3.2.3), while assumption (A.7) adapts condition (A.6) to the the case of the functional conditional U-statistics in the k -NN setting.

³The concept of metric entropy was introduced by Kolmogorov (*cf.* Kolmogorov and Tihomirov [1959]) and studied subsequently for numerous metric spaces. This concept was used by Dudley [1967] to give sufficient conditions for the continuity of Gaussian processes and was the basis for striking generalizations of Donsker's theorem on the weak convergence of the empirical process, refer to van der Vaart [2002] for the connection of this notion with Le Cam's work.

Remark 3.2.3.1 The condition (A.4.b) regarding the kernel function $K(\cdot)$ could be replaced by the following one for treating the regression function estimator when $m = 1$:

(A.4.b') The kernel functions $K(\cdot)$ is Lipschitz supported within $(0, 1/2)$ with existing first derivative on $(0, 1/2)$ satisfying for two real constants $-\infty < \kappa'_3 \leq \kappa'_4 < 0$:

$$-\infty < \kappa'_3 \leq K'(\cdot) \leq \kappa'_4 < 0.$$

But it should be noted that if we consider the previous assumptions instead of (A.4.b) in order to bound the expectation of the kernel, we should suppose that the small ball probability $\phi(\cdot)$ satisfies

$$\exists C > 0, \exists \eta_0 > 0, \forall \eta < \eta_0 : \int_0^\eta \phi(u) du > C\eta\phi(\eta),$$

for instance, refer to Lemma 4.3 and Lemma 4.4 of [Ferraty and Vieu \[2006\]](#).

Remark 3.2.3.2 Note that the condition (A.4.a') may be replaced by more general hypotheses upon moments of \mathbf{Y} as in [Deheuvels \[2011\]](#). That is

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

$$(i) \ x^{-s}\mathcal{M}(x) \downarrow; (ii) \ x^{-1}\mathcal{M}(x) \uparrow. \quad (3.2.17)$$

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

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

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

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

3.3 UIB and UINN consistency for functional regression

In this section, we consider the uniform consistency and the uniform in number of neighbors consistency of the functional regression operator in its general form which is given for all $t \in \mathcal{X}$, by:

$$\hat{r}_n^{*(1)}(\varphi, \mathbf{t}, H_{n,k}(t)) = \frac{\sum_{i=1}^n \varphi(Y_i) K\left(\frac{d(t, X_i)}{H_{n,k}(t)}\right)}{\sum_{i=1}^n K\left(\frac{d(t, X_i)}{H_{n,k}(t)}\right)}, \quad (3.3.1)$$

where $k_{1,n} \leq k \leq k_{2,n}$, ($k = k_n$ depends on n) and

$$H_{n,k}(t) = \min \left\{ h \in \mathbb{R}^+ \text{ such that } \sum_{i=1}^n \mathbb{1}_{B(t,h)}(X_i) = k \right\}.$$

In fact the k -NN operator presented in (3.3.1) can be considered as a generalization of the usual kernel regression

$$\widehat{r}_n^{(1)}(\varphi, \mathbf{t}, h_K) = \frac{\sum_{i=1}^n \varphi(Y_i) K\left(\frac{d(t, X_i)}{h_K}\right)}{\sum_{i=1}^n K\left(\frac{d(t, X_i)}{h_K}\right)}, \quad \forall t \in \mathcal{X}, \quad (3.3.2)$$

where the bandwidth $h_K \in [a_n, b_n] \subset \mathbb{R}_+^*$ depends on n (but does not depend on t). We start by reminding the uniform asymptotic properties of $\widehat{r}_n^{(1)}$ defined in (3.3.2).

3.3.1 UIB consistency for functional regression

We give in the following results the uniform convergence in the a. co sense.

Theorem 3.3.1.1 *Under the assumptions (A.1.)-(A.6.), (for $m = 1$) we have,*

$$\sup_{\varphi \in \mathcal{F}} \sup_{\mathcal{H}} \sup_{a_n \leq h_K \leq b_n} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |\widehat{r}_n^{(1)}(\varphi, t; h_K) - r^{(1)}(\varphi, t)| = O(b_n^Y) + O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}}\left(\frac{\log n}{n}\right)}{n\phi(a_n)}} \right). \quad (3.3.3)$$

The following result gives the uniform consistency when the class of functions is unbounded.

Corollary 3.3.1.2 *Under the assumptions (A.1.)- (A.3.), and (A.4.a'), (A.4.b), (A.5.), (A.6.) (for $m = 1$), we have*

$$\sup_{\varphi \in \mathcal{F}} \sup_{\mathcal{H}} \sup_{a_n \leq h_K \leq b_n} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |\widehat{r}_n^{(1)}(\varphi, t; h_K) - r^{(1)}(\varphi, t)| = O(b_n^Y) + O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}}\left(\frac{\log n}{n}\right)}{n\phi(a_n)}} \right). \quad (3.3.4)$$

Now, we can state our results for the k -NN functional regression.

3.3.2 UINN consistency for functional regression

Theorem 3.3.2.1 *Under the assumptions (A.2.)-(A.5.) and (A.7.), we have*

$$\begin{aligned} \sup_{\varphi \in \mathcal{F}} \sup_{\mathcal{H}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |\widehat{r}_n^{*(1)}(\varphi, t; h_{n,k}(t)) - r^{(1)}(\varphi, t)| &= O\left(\phi^{-1}\left(\frac{k_{2,n}}{\rho_n n}\right)^Y\right) \\ &+ O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}}\left(\frac{\log n}{n}\right)}{n\phi\left(\mu\phi^{-1}\left(\frac{\rho_n k_{1,n}}{n}\right)\right)}} \right). \end{aligned}$$

The following result gives the uniform consistency when the class of functions is unbounded.

Corollary 3.3.2.2 *Under the assumptions (A.1.)- (A.3.), (A.4.a'), (A.4.b), (A.5.), (A.6.) and (A.7.) (for $m = 1$), we have*

$$\begin{aligned} \sup_{\varphi \in \mathcal{F}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |\hat{r}_n^{*(1)}(\varphi, t; h_{n,k}(t)) - r^{(1)}(\varphi, t)| &= O\left(\phi^{-1}\left(\frac{k_{2,n}}{\rho_n n}\right)^Y\right) \\ &+ O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}}\left(\frac{\log n}{n}\right)}{n\phi\left(\mu\phi^{-1}\left(\frac{\rho_n k_{1,n}}{n}\right)\right)}} \right) \end{aligned} \quad (3.35)$$

3.4 UIB and UINN consistency of functional conditional U-statistics

In this section, we consider the uniform consistency, the UIB and the UINN consistency of the functional conditional U-statistic given by (3.2.1). First, let's introduce some notation :

$$\begin{aligned} \mathbf{X} &:= (X_1, \dots, X_m) \in \mathcal{X}^m, \quad \mathbf{Y} := (Y_1, \dots, Y_m) \in \mathcal{Y}^m, \\ \mathbf{X}_i &:= (X_{i_1}, \dots, X_{i_m}), \quad \mathbf{Y}_i := (Y_{i_1}, \dots, Y_{i_m}), \\ \mathbf{h}_{n,k}(\mathbf{t}) &:= (H_{n,k}(t_1), \dots, H_{n,k}(t_m)) \quad \text{for } \mathbf{t} = (t_1, \dots, t_m) \in \mathcal{S}_{\mathcal{X}}^m, \\ G_{\varphi, \mathbf{t}, \mathbf{h}_{n,k}}(\mathbf{X}, \mathbf{Y}) &:= \frac{1}{\phi_{\mathbf{t}}(\mathbf{h}_{n,k}(\mathbf{t}))} \varphi(Y_1, \dots, Y_m) \prod_{i=1}^m K\left(\frac{d(X_i, t_i)}{H_{n,k}(t_i)}\right) \quad \text{for } \mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m, \\ u_n(\varphi, \mathbf{t}, \mathbf{h}_{n,k}(\mathbf{t})) &= u_n^{(m)}(G_{\varphi, \mathbf{t}, \mathbf{h}_{n,k}}) := \frac{(n-m)!}{n!} \sum_{i \in \mathbb{I}(m,n)} G_{\varphi, \mathbf{t}, \mathbf{h}_{n,k}}(\mathbf{X}_i, \mathbf{Y}_i), \end{aligned}$$

and for some symmetric measurable function $f(\cdot)$ define the \mathbb{P} -canonical function $\pi_{p,m}$, $p = 1, \dots, m$, see Arcones and Giné [1993] (we replace the index k with p to avoid confusing it with the smoothing parameter k), by :

$$\pi_{p,m} f(t_1, \dots, t_p) := (\delta_{t_1} - \mathbb{P}) \cdots (\delta_{t_p} - \mathbb{P}) \mathbb{P}^{m-p} f.$$

where for measures Q_i on S we let

$$Q_1 \cdots Q_m h = \int_{S^m} h(x_1, \dots, x_m) dQ_1(x_1) \cdots dQ_m(x_m),$$

and δ_x denote Dirac measure at point $x \in \mathcal{X}$. For example,

$$\pi_{1,m} h(x) = \mathbb{E}(h(X_1, \dots, X_m) | X_1 = x) - \mathbb{E}h(X_1, \dots, X_m).$$

It's clear that, for all $\varphi \in \mathcal{F}_m$:

$$\hat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h}_{n,k}(\mathbf{t})) = \frac{u_n(\varphi, \mathbf{t}, \mathbf{h}_{n,k}(\mathbf{t}))}{u_n(1, \mathbf{t}, \mathbf{h}_{n,k}(\mathbf{t}))},$$

and $u_n(\varphi, \mathbf{t}, \mathbf{h}_{n,k}(\mathbf{t}))$ is a classical U-statistic with the U-kernel $G_{\varphi, \mathbf{t}, \mathbf{h}_{n,k}}(\mathbf{x}, \mathbf{y})$. However, the study of the uniform consistency and the UIB consistency of $\hat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h}_{n,k}(\mathbf{t}))$ to $r^{(m)}(\varphi, \mathbf{t})$ can not

be done with a straightforward approach due to the randomness of the bandwidth vector $\mathbf{h}_{n,k}(\mathbf{t})$ which poses some technical problems. To circumvent this, our strategy is to first study the uniform consistency and the UIB consistency of $\widehat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h})$, where $\mathbf{h} = (h_1, \dots, h_m) \in \mathcal{H}_n^{(m)}$ is a multivariate bandwidth that does not depend on \mathbf{t} and k . Hence, we study the uniform consistency and the UIB consistency of $u_n(\varphi, \mathbf{t}; \mathbf{h})$ to $\mathbb{E}(u_n(\varphi, \mathbf{t}; \mathbf{h}))$ when $\varphi \in \mathcal{F}_m$ and when $\varphi \equiv 1$, and we shall consider an appropriate centering factor than the expectation $\mathbb{E}(\widehat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h}))$, hence we define :

$$\widehat{\mathbb{E}}(\widehat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h})) = \frac{\mathbb{E}(u_n(\varphi, \mathbf{t}; \mathbf{h}))}{\mathbb{E}(u_n(1, \mathbf{t}; \mathbf{h}))}. \quad (3.4.1)$$

The second step will be the use a general lemma, adapted to our setting, similar to that of [Burba et al. \[2009\]](#) and [Kudraszow and Vieu \[2013\]](#) (see Subsection 3.9.1) to derive the results for the bandwidth $\mathbf{h}_{n,k}(\mathbf{t})$.

3.4.1 Uniform consistency and UIB of functional conditional U-statistics

Before we start studying the uniform consistency and UIB of functional conditional U-statistics in the multivariate setting, for reader convenience, let us recall the results of [Bouzebda and Nemouchi \[2020\]](#) for a univariate bandwidth $h_K \in [a_n, b_n]$.

Uniform consistency and UIB consistency for a univariate bandwidth

The following result handles the uniform deviation of the estimate $u_n(\varphi, \mathbf{t}, h_K)$ with respect to $\mathbb{E}(u_n(\varphi, \mathbf{t}, h_K))$ when the class of functions is bounded.

Theorem 3.4.1.1 *Suppose that the conditions (A.1.), (A.2.), (A.4.), (A.6.) and (A.5'.) are fulfilled, we infer that for*

$$\sup_{\varphi \in \mathcal{F}_m} \sup_{\mathbf{K} \in \mathcal{K}^m} \sup_{a_n \leq h_K \leq b_n} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} |u_n(\varphi, \mathbf{t}, h_K) - \mathbb{E}(u_n(\varphi, \mathbf{t}, h_K))| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}{n\phi(a_n)}} \right), \quad (3.4.2)$$

where, for $\mathbf{t} = (t_1, \dots, t_m) \in \mathcal{S}_{\mathcal{X}}^m$,

$$\widetilde{\mathbf{K}} \left(\frac{d(\cdot, \mathbf{t})}{h_K} \right) := \prod_{i=1}^m \mathbf{K} \left(\frac{d(\cdot, t_i)}{h_K} \right).$$

The following result handles the uniform deviation of the estimate $u_n(\varphi, \mathbf{t}, h_K)$ with respect to $\mathbb{E}(u_n(\varphi, \mathbf{t}, h_K))$ when the class of functions is unbounded satisfying general moment condition.

Theorem 3.4.1.2 *Suppose that the conditions (A.1.), (A.2.), (A.4.a'), (A.4.b), (A.5'.) and (A.6.) are fulfilled, we infer that for all $0 < b_0 < 1$:*

$$\sup_{\varphi \in \mathcal{F}_m} \sup_{\mathbf{K} \in \mathcal{K}^m} \sup_{a_n \leq h_K \leq b_0} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} |u_n(\varphi, \mathbf{t}, h_K) - \mathbb{E}(u_n(\varphi, \mathbf{t}, h_K))| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}{n\phi(a_n)}} \right). \quad (3.4.3)$$

The following result handles the uniform deviation of the estimate $\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_K)$ with respect to $\hat{\mathbb{E}}(\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_K))$ in both cases, when the class of functions is bounded or unbounded satisfying general moment condition.

Theorem 3.4.1.3 *Suppose that the conditions (A.1.), (A.2.), (A.4.), (A.5'.) and (A.6.) (or the following (A.1.)-(A.4.a'), (A.4.b), (A.5'.) and (A.6.)) are fulfilled, we infer that for all $0 < b_0 < 1$:*

$$\sup_{\varphi \tilde{K} \in \mathcal{F}_m \mathcal{K}^m} \sup_{a_n \leq h_K \leq b_0} \sup_{\mathbf{t} \in \mathcal{S}_x^m} |\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_K) - \hat{\mathbb{E}}(\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_K))| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_x} \left(\frac{\log n}{n} \right)}{n\phi(a_n)}} \right). \quad (3.4.4)$$

Theorem 3.4.1.4 *Suppose that the conditions (A.1.)-(A.3.) and (A.4.b) are fulfilled, we infer that for all $0 < b_n < 1, b_n \rightarrow 0$:*

$$\sup_{\varphi \tilde{K} \in \mathcal{F}_m \mathcal{K}^m} \sup_{a_n \leq h_K \leq b_n} \sup_{\mathbf{t} \in \mathcal{S}_x^m} |\hat{\mathbb{E}}(\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_K)) - r^{(m)}(\varphi, \mathbf{t})| = O(b_n^Y). \quad (3.4.5)$$

Corollary 3.4.1.5 *Under the assumptions of Theorems 3.4.1.3 and 3.4.1.4 it follows that :*

$$\sup_{\varphi \tilde{K} \in \mathcal{F}_m \mathcal{K}^m} \sup_{a_n \leq h_K \leq b_n} \sup_{\mathbf{t} \in \mathcal{S}_x^m} |\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_K) - r^{(m)}(\varphi, \mathbf{t})| = O(b_n^Y) + O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_x} \left(\frac{\log n}{n} \right)}{n\phi(a_n)}} \right). \quad (3.4.6)$$

Remark 3.4.1.6 *If the elements φ of the class of functions \mathcal{F}_m are not symmetric, which implies that $G_{\varphi, \mathbf{t}, h_K}$ are not symmetric, we will need to symmetrize them. To do this, we have :*

$$\bar{G}_{\varphi, \mathbf{t}, h_K}(\mathbf{x}, \mathbf{y}) := \frac{1}{m!} \sum_{\sigma \in I_m^m} G_{\varphi, \mathbf{t}, h_K}(\mathbf{x}_\sigma, \mathbf{y}_\sigma) = \frac{1}{m!} \sum_{\sigma \in I_m^m} \varphi(\mathbf{y}_\sigma) \tilde{K}(\mathbf{x}_\sigma - \mathbf{t}),$$

where $\mathbf{x}_\sigma = (x_{\sigma_1}, \dots, x_{\sigma_m})$ and $\mathbf{y}_\sigma = (y_{\sigma_1}, \dots, y_{\sigma_m})$. After symmetrization the expectation

$$\mathbb{E}(\bar{G}_{\varphi, \mathbf{t}, h_K}(\mathbf{x}, \mathbf{y})) = \mathbb{E}(G_{\varphi, \mathbf{t}, h_K}(\mathbf{x}, \mathbf{y})),$$

and the U-statistic

$$u_n^{(m)}(G_{\varphi, \mathbf{t}, h_K}) = u_n^{(m)}(\bar{G}_{\varphi, \mathbf{t}, h_K}) := u_n(\varphi, \mathbf{t}, h_K)$$

do not change. Note that if the class of functions $\mathcal{F}_m \mathcal{K}^m$ satisfies the entropy condition, the class $\bar{\mathcal{F}}_m \mathcal{K}^m$ of symmetrized functions, satisfies it too and its envelope function is :

$$\bar{F}(\mathbf{y}) \equiv \bar{F}(\mathbf{x}, \mathbf{y}) = k^m \sum_{\sigma \in I_m^m} F(\mathbf{y}_\sigma).$$

This remark remains true also for the functions $G_{\varphi, \mathbf{t}, h_{n,k}}$.

Uniform consistency and UIB consistency for a multivariate bandwidth

Next, we will give the UIB results for all $\mathbf{t} \in \mathcal{S}_x^m$ and $\mathbf{h} \in \mathcal{H}_n^{(m)}$. We first start with announcing the result concerning the uniform derivation of the estimate $u_n(\varphi, \mathbf{t}, \mathbf{h})$ with respect to $\mathbb{E}(u_n(\varphi, \mathbf{t}, \mathbf{h}))$ when the class of functions is bounded.

Theorem 3.4.1.7 *Suppose that the conditions (A.1.), (A.2.), (A.4.), (A.5'.) and (A.6.) are fulfilled, we infer that*

$$\sup_{\varphi \in \mathcal{F}_m} \sup_{\mathbf{h} \in \mathcal{H}_n^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} |u_n(\varphi, \mathbf{t}, \mathbf{h}) - \mathbb{E}(u_n(\varphi, \mathbf{t}, \mathbf{h}))| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_x} \left(\frac{\log n}{n} \right)}{n\phi(\tilde{a}_n)}} \right). \quad (3.4.7)$$

The following result covers the uniform derivation of the estimate $u_n(\varphi, \mathbf{t}, \mathbf{h})$ with respect to $\mathbb{E}(u_n(\varphi, \mathbf{t}, \mathbf{h}))$ when the class of functions is unbounded satisfying general moments condition.

Theorem 3.4.1.8 *Suppose that the conditions (A.1.), (A.2.), (A.4.a'), (A.4.b), (A.5'.) and (A.6.) are fulfilled, we infer that for all $0 < \tilde{b}_0 < 1$:*

$$\sup_{\varphi \in \mathcal{F}_m} \sup_{\mathbf{h} \in \mathcal{H}_0^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} |u_n(\varphi, \mathbf{t}, \mathbf{h}) - \mathbb{E}(u_n(\varphi, \mathbf{t}, \mathbf{h}))| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_x} \left(\frac{\log n}{n} \right)}{n\phi(\tilde{a}_n)}} \right). \quad (3.4.8)$$

The following result handles the uniform deviation of the estimate $\hat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h})$ with respect to $\hat{\mathbb{E}}(\hat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h}))$ in both situations, where the class of functions is bounded or unbounded satisfying a general moment condition.

Theorem 3.4.1.9 *Suppose that the conditions (A.1.), (A.2.), (A.4.), (A.5'.) and (A.6.) (or the following (A.1.)-(A.4.a'), (A.4.b), (A.6.) and (A.5'.)) are fulfilled, we infer that for all $0 < \tilde{b}_0 < 1$:*

$$\sup_{\varphi \in \mathcal{F}_m} \sup_{\mathbf{h} \in \mathcal{H}_0^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} |\hat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h}) - \hat{\mathbb{E}}(\hat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h}))| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_x} \left(\frac{\log n}{n} \right)}{n\phi(\tilde{a}_n)}} \right). \quad (3.4.9)$$

Theorem 3.4.1.10 *Suppose that the conditions (A.1.)-(A.3.) and (A.4.b) are fulfilled, we infer that or all $0 < \tilde{b}_n < 1, \tilde{b}_n \rightarrow 0$:*

$$\sup_{\varphi \in \mathcal{F}_m} \sup_{\mathbf{h} \in \mathcal{H}_n^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} |\hat{\mathbb{E}}(\hat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h})) - r^{(m)}(\varphi, \mathbf{t})| = O(\tilde{b}_n^Y). \quad (3.4.10)$$

Corollary 3.4.1.11 *Under the assumptions of Theorems 3.4.1.9 and 3.4.1.10 it follows that :*

$$\sup_{\varphi \in \mathcal{F}_m} \sup_{\mathbf{h} \in \mathcal{H}_n^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} |\hat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h}) - r^{(m)}(\varphi, \mathbf{t})| = O(\tilde{b}_n^Y) + O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_x} \left(\frac{\log n}{n} \right)}{n\phi(\tilde{a}_n)}} \right). \quad (3.4.11)$$

3.4.2 Uniform consistency and UINN consistency of functional conditional U -statistics

Let $0 < \mu^* \leq \nu^* < \infty$ be some constants and $\rho_n^* \in (0, 1)$ is a sequence chosen in such a way that

$$\min_{1 \leq j \leq m} \mu_j \phi^{-1} \left(\frac{\rho_{n,j} k_{1,n}}{n} \right) = \mu^* \phi^{-1} \left(\frac{\rho_n^* k_{1,n}}{n} \right)$$

and

$$\max_{1 \leq j \leq m} \nu_j \phi^{-1} \left(\frac{k_{2,n}}{\rho_{n,j} n} \right) = \nu^* \phi^{-1} \left(\frac{k_{2,n}}{\rho_n^* n} \right).$$

The following result deals with the uniform deviation of the estimate $u_n(\varphi, \mathbf{t}, \mathbf{h}_{n,k}(\mathbf{t}))$ with respect to $\mathbb{E}(u_n(\varphi, \mathbf{t}, \mathbf{h}_{n,k}(\mathbf{t})))$, when the class of functions is bounded.

Theorem 3.4.2.1 *Suppose that the conditions (A.1.), (A.2.), (A.4.), (A.5'.) and (A.6.) are fulfilled, if in addition assumption (A.7.) holds, we infer that :*

$$\begin{aligned} & \sup_{\varphi \in \mathcal{F}_m} \sup_{\mathcal{K} \in \mathcal{K}^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} |u_n(\varphi, \mathbf{t}, \mathbf{h}_{n,k}(\mathbf{t})) - \mathbb{E}(u_n(\varphi, \mathbf{t}, \mathbf{h}_{n,k}(\mathbf{t})))| \\ & = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_x} \left(\frac{\log n}{n} \right)}{n \phi \left(\mu^* \phi^{-1} \left(\frac{\rho_n^* k_{1,n}}{n} \right) \right)}} \right) \end{aligned} \quad (3.4.12)$$

The following result deals with the uniform deviation of the estimate $u_n(\varphi, \mathbf{t}, \mathbf{h}_{n,k}(\mathbf{t}))$ with respect to $\mathbb{E}(u_n(\varphi, \mathbf{t}, \mathbf{h}_{n,k}(\mathbf{t})))$ when the class of functions is unbounded satisfying general moments condition.

Theorem 3.4.2.2 *Suppose that the conditions (A.1.), (A.2.), (A.4.a'), (A.4.b), (A.5'.) and (A.6.) are fulfilled, if in addition assumption (A.7.) holds, we infer that :*

$$\begin{aligned} & \sup_{\varphi \in \mathcal{F}_m} \sup_{\mathcal{K} \in \mathcal{K}^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} |u_n(\varphi, \mathbf{t}, \mathbf{h}_{n,k}(\mathbf{t})) - \mathbb{E}(u_n(\varphi, \mathbf{t}, \mathbf{h}_{n,k}(\mathbf{t})))| \\ & = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_x} \left(\frac{\log n}{n} \right)}{n \phi \left(\mu^* \phi^{-1} \left(\frac{\rho_n^* k_{1,n}}{n} \right) \right)}} \right) \end{aligned} \quad (3.4.13)$$

The next results give uniform consistency when the class of functions is bounded or unbounded.

Theorem 3.4.2.3 *Suppose that the conditions (A.1.), (A.2.), (A.4.), (A.5'.) and (A.6.) (or the following (A.1.)-(A.4.a'), (A.4.b), (A.5'.) and (A.6.)) are fulfilled, if in addition assumption (A.7.) holds, we infer that :*

$$\begin{aligned} & \sup_{\varphi \in \mathcal{F}_m} \sup_{\mathcal{K} \in \mathcal{K}^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} |\hat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h}_{n,k}(\mathbf{t})) - \widehat{\mathbb{E}}(\hat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h}_{n,k}(\mathbf{t})))| \\ & = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_x} \left(\frac{\log n}{n} \right)}{n \phi \left(\mu^* \phi^{-1} \left(\frac{\rho_n^* k_{1,n}}{n} \right) \right)}} \right) \end{aligned} \quad (3.4.14)$$

Theorem 3.4.2.4 Suppose that the conditions (A.1.)-(A.3.) and (A.4.b) are fulfilled, if in addition assumption (A.7.) holds, we infer that :

$$\sup_{\varphi \in \mathcal{F}_m} \sup_{\tilde{\mathcal{K}} \in \mathcal{K}^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} |\widehat{\mathbb{E}}(\widehat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h}_{n,k}(\mathbf{t}))) - r^{(m)}(\varphi, \mathbf{t})| = O\left(\phi^{-1}\left(\frac{k_{2,n}}{\rho_n^* n}\right)^\gamma\right). \quad (3.4.15)$$

Corollary 3.4.2.5 Under the assumptions of Theorems 3.4.2.3 and 3.4.2.4 it follows that :

$$\begin{aligned} & \sup_{\varphi \in \mathcal{F}_m} \sup_{\tilde{\mathcal{K}} \in \mathcal{K}^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} |\widehat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h}_{n,k}(\mathbf{t})) - r^{(m)}(\varphi, \mathbf{t})| \\ &= O\left(\phi^{-1}\left(\frac{k_{2,n}}{\rho_n^* n}\right)^\gamma\right) + O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}}\left(\frac{\log n}{n}\right)}{n\phi\left(\mu^* \phi^{-1}\left(\frac{\rho_n^* k_{1,n}}{n}\right)\right)}} \right). \end{aligned} \quad (3.4.16)$$

Remark 3.4.2.6 The choice of the parameters μ^* and ρ_n^* defined in a similar way as in condition (A.7.) affects the rate of convergence of the k -NN estimator. We can choose these parameters depending on the small ball probability function $\phi_t(h)$; it is worth noticing that, in many examples, the latter can be written approximately as a product of two independent functions of t and another of h , (see, for instance, [Mayer-Wolf and Zeitouni \[1993\]](#) for the diffusion process, [Bogachev \[1998\]](#) for a Gaussian measure, and [Li and Shao \[2001\]](#) for a general Gaussian process). The most frequent result available in the literature is of the form $\phi_t(\varepsilon) \sim g(t)\phi(\varepsilon)$ where $\phi(\varepsilon) = \varepsilon^\gamma \exp(-C/\varepsilon^p)$ with $\gamma \geq 0$ and $p \geq 0$. It corresponds to the Ornstein-Uhlenbeck and general diffusion processes (for such processes, $p = 2$ and $\gamma = 0$) and the fractal processes (for such processes, $\gamma > 0$ and $p = 0$). For more examples [Ferraty et al. \[2007\]](#):

1. $\phi_t(h) = \ell(t)h^\nu$, for some $\nu > 0$ where $\tau_t(s) = s^\nu$;
2. $\phi_t(h) = \ell(t)h^\nu \exp(-Ch^{-p})$, for some $\nu > 0$ and $p > 0$ where $\tau_t(s)$ is the Dirac's function;
3. $\phi_t(h) = \ell(t)|\ln(h)|^{-1}$, where $\tau_t(s) = \mathbb{1}_{]0,1]}(s)$ the indicator function in $]0,1]$.

Remark 3.4.2.7 It is worth being noted what is the price to pay by the nice features of the k -NN-based estimators: remembering that $H_{n,k}(X)$ is a random variable (which depends on (X_1, \dots, X_n)), one should expect that additional technical difficulties will appear along the proofs of asymptotic properties. To fix the idea on this point, note that the random elements involved in (3.2.1), $m = 1$, can not be decomposed as sums of independent variables (as it is the case, for instance, with kernel-based estimators), and hence its treatment will need more sophisticated probabilistic developments than standard limit theorems for sums of iid variables. Also, the Hoeffding decomposition can not be applied directly to (3.2.1), which is the main tool for the study U -statistics.

3.5 Extension to the censored case

Consider a triple (Y, C, X) of random variables defined in $\mathbb{R} \times \mathbb{R} \times \mathcal{X}$. Here Y is the variable of interest, C is a censoring variable, and X is a concomitant variable. Throughout, we will use [Maillot and Viallon \[2009\]](#) and [Bouzebda and Chaouch \[2022\]](#) notation and we work with a sample $\{(Y_i, C_i, X_i)_{1 \leq i \leq n}\}$ of independent and identically distributed replication of (Y, C, 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

$$\mathcal{D}_n = \{(Z_i, \Delta_i, X_i), i = 1, \dots, n\}.$$

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 that result in death. For $-\infty < t < \infty$, set

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

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

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

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

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

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.* \[1995\]](#)). The key idea of I.P.C.W. estimators is as follows. Introduce the real-valued function $\Phi_\psi(\cdot, \cdot)$ defined on \mathbb{R}^2 by

$$\Phi_\psi(y, c) = \frac{\mathbb{1}\{y \leq c\}\psi(y \wedge c)}{1 - G(y \wedge c)}. \quad (3.5.1)$$

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

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

We have

$$\begin{aligned}
 r^{(1)}(\Phi_\psi, \mathbf{x}) &:= \mathbb{E}(\Phi_\psi(Y, C) \mid X = x) \\
 &= \mathbb{E}\left\{ \frac{\mathbb{1}\{Y \leq C\} \psi(Z)}{1 - G(Z)} \mid X = x \right\} \\
 &= \mathbb{E}\left\{ \frac{\psi(Y)}{1 - G(Y)} \mathbb{E}(\mathbb{1}\{Y \leq C\} \mid X, Y) \mid X = x \right\} \\
 &= r^{(1)}(\psi, x).
 \end{aligned} \tag{3.5.2}$$

Therefore, any estimate of $r^{(1)}(\Phi_\psi, \cdot)$, which can be built on fully observed data, turns out to be an estimate for $r^{(1)}(\psi, \cdot)$ too. Thanks to this property, most statistical procedures 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 $x \in \mathcal{X}$, $h \geq 0$, $1 \leq i \leq n$,

$$\bar{\omega}_{n,K,h,i}^{(1)}(x) := K\left(\frac{d(x, X_i)}{H_{n,k}(x)}\right) / \sum_{j=1}^n K\left(\frac{d(x, X_j)}{H_{n,k}(x)}\right). \tag{3.5.3}$$

In view of (3.5.1), (3.5.2), and (3.5.3), whenever $G(\cdot)$ is known, a kernel estimator of $r^{(1)}(\psi, \cdot)$ is given by

$$\tilde{r}_n^{(1)}(\psi, x; h_n) = \sum_{i=1}^n \bar{\omega}_{n,K,h,i}^{(1)}(x) \frac{\Delta_i \psi(Z_i)}{1 - G(Z_i)}. \tag{3.5.4}$$

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

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

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

$$\tilde{r}_n^{(1)*}(\psi, x; h_n) = \sum_{i=1}^n \bar{\omega}_{n,K,h,i}^{(1)}(x) \frac{\Delta_i \psi(Z_i)}{1 - G_n^*(Z_i)}, \tag{3.5.5}$$

refer to [Kohler et al. \[2002\]](#) and [Mailhot 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 k th ordered statistic associated with the sample (Z_1, \dots, Z_n) for $k = 1, \dots, n$ 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]). 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. We refer also to Bouzebda and El-hadjali [2020]. To our best knowledge, the problem of the estimation of the conditional U-statistics was open up to the present, and it gives the main motivation for the study of this section. A natural extension of the function defined in (3.5.1) is given by

$$\Phi_{\Psi}(y_1, \dots, y_m, c_1, \dots, c_m) = \frac{\prod_{i=1}^m \{\mathbb{1}\{y_i \leq c_i\} \Psi(y_1 \wedge c_1, \dots, y_m \wedge c_m)\}}{\prod_{i=1}^m \{1 - G(y_i \wedge c_i)\}}.$$

From this, we have an analogous relation to (3.5.2) given by

$$\begin{aligned} & \mathbb{E}(\Phi_{\Psi}(Y_1, \dots, Y_m, C_1, \dots, C_m) \mid (X_1, \dots, X_m) = \mathbf{t}) \\ &= \mathbb{E} \left(\frac{\prod_{i=1}^m \{\mathbb{1}\{Y_i \leq C_i\} \Psi(Y_1 \wedge C_1, \dots, Y_m \wedge C_m)\}}{\prod_{i=1}^m \{1 - G(Y_i \wedge C_i)\}} \mid (X_1, \dots, X_m) = \mathbf{t} \right) \\ &= \mathbb{E} \left(\frac{\Psi(Y_1, \dots, Y_m)}{\prod_{i=1}^m \{1 - G(Y_i)\}} \mathbb{E} \left(\prod_{i=1}^m \{\mathbb{1}\{Y_i \leq C_i\} \mid (Y_1, X_1), \dots, (Y_m, X_m)\} \mid (X_1, \dots, X_m) = \mathbf{t} \right) \right) \\ &= \mathbb{E}(\Psi(Y_1, \dots, Y_m) \mid (X_1, \dots, X_m) = \mathbf{t}). \end{aligned}$$

An analogue estimator to (3.2.1) in the censored case is given by

$$\check{r}_n^{*(m)}(\boldsymbol{\varphi}, \mathbf{t}, \mathbf{h}_{n,k}(\mathbf{t})) = \sum_{(i_1, \dots, i_m) \in \mathcal{I}(m, n)} \frac{\delta_{i_1} \cdots \delta_{i_m} \Psi(Z_{i_1}, \dots, Z_{i_m})}{(1 - G(Z_{i_1})) \cdots (1 - G(Z_{i_m}))} \bar{\omega}_{n,K,k,\mathbf{i}}^{(m)}(\mathbf{t}), \quad (3.5.6)$$

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

$$\bar{\omega}_{n,K,k,\mathbf{i}}^{(m)}(\mathbf{t}) = \frac{K\left(\frac{d(t_1, X_{i_1})}{H_{n,k}(t_1)}\right) \cdots K\left(\frac{d(t_m, X_{i_m})}{H_{n,k}(t_m)}\right)}{\sum_{(i_1, \dots, i_m) \in \mathcal{I}(m, n)} K\left(\frac{d(t_1, X_{i_1})}{H_{n,k}(t_1)}\right) \cdots K\left(\frac{d(t_m, X_{i_m})}{H_{n,k}(t_m)}\right)}. \quad (3.5.7)$$

The estimator that we will investigate is given by

$$\check{r}_n^{(k)*}(\boldsymbol{\psi}, \mathbf{t}, \mathbf{h}_{n,k}(\mathbf{t})) = \sum_{(i_1, \dots, i_k) \in \mathcal{I}(k, n)} \frac{\delta_{i_1} \cdots \delta_{i_k} \Psi(Z_{i_1}, \dots, Z_{i_k})}{(1 - G_n^*(Z_{i_1})) \cdots (1 - G_n^*(Z_{i_k}))} \bar{\omega}_{n,\delta,m_n,\mathbf{i}}^{(k)}(\mathbf{t}). \quad (3.5.8)$$

In order to obtain our results, we will work under the following assumptions.

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

(A.2) The class of functions \mathcal{F}' has a measurable and uniformly bounded envelope function Y with,

$$Y(y_1, \dots, y_m) \geq \sup_{\psi \in \mathcal{F}'} |\psi(y_1, \dots, y_m)|, \quad y_i \leq T_H.$$

(A.3) The class of functions \mathcal{M} is relatively compact with respect to the sup-norm topology on \mathcal{V}_x^m (compact set of \mathcal{X} containing x), where

$$\mathcal{M} = \{r^{(m)}(\varphi, \cdot) \tilde{K}(\cdot) : \varphi \in \mathcal{F}_q\}.$$

Corollary 3.5.0.1 *Under (A.1)-(A.3) and the assumptions of Theorems 3.4.2.3 and 3.4.2.4 it follows that :*

$$\begin{aligned} & \sup_{\varphi \tilde{K} \in \mathcal{F}_m} \sup_{\mathcal{K}^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{V}_x^m} \left| \tilde{r}_n^{(k)*}(\psi, \mathbf{t}; \mathbf{h}_{n,k}(\mathbf{t})) - r^{(m)}(\varphi, \mathbf{t}) \right| \\ &= O\left(\phi^{-1}\left(\frac{k_{2,n}}{\rho_n^* n}\right)^Y\right) + O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{L}_x}\left(\frac{\log n}{n}\right)}{n\phi\left(\mu^* \phi^{-1}\left(\frac{\rho_n^* k_{1,n}}{n}\right)\right)}} \right). \end{aligned} \quad (3.5.9)$$

This last result is a direct consequence of Corollary (3.4.2.5) and the law of iterated logarithm for $G_n^*(\cdot)$ established in [Földes and Rejtő \[1981\]](#) ensures that

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

3.6 Some potential applications

3.6.1 Set indexed conditional U-statistics

We aim to study the links between X and Y by estimating functional operators associated with the conditional distribution of Y given X , such as the regression operator, for $C_1 \times \dots \times C_m := \tilde{\mathcal{C}}$ in \mathcal{a} is a class of sets \mathcal{C}^m ,

$$\mathbb{G}^{(m)}(C_1 \times \dots \times C_m | \mathbf{t}) = \mathbb{E} \left(\prod_{i=1}^m \mathbb{1}_{\{Y_i \in C_i\}} | (X_1, \dots, X_m) = \mathbf{t} \right) \text{ for } \mathbf{t} \in \mathcal{X}^m.$$

We define metric entropy with the inclusion of the class of sets \mathcal{C} . For each $\varepsilon > 0$, the covering number is defined as :

$$\begin{aligned} \mathcal{N}(\varepsilon, \mathcal{C}, \mathbb{G}^{(1)}(\cdot | x)) &= \inf\{n \in \mathbb{N} : \exists C_1, \dots, C_n \in \mathcal{C} \text{ such that } \forall C \in \mathcal{C} \exists 1 \leq i, j \leq n \\ &\quad \text{with } C_i \subset C \subset C_j \text{ and } \mathbb{G}^{(1)}(C_j \setminus C_i | x) < \varepsilon\}, \end{aligned}$$

the quantity $\log(\mathcal{N}(\varepsilon, \mathcal{C}, \mathbb{G}^{(1)}(\cdot | x)))$ is called metric entropy with inclusion of \mathcal{C} with respect to $\mathbb{G}^{(1)}(\cdot | x)$. The quantity $\log \mathcal{N}(\varepsilon, \mathcal{C}, \mathbb{G}^{(1)}(\cdot | x))$ is called metric entropy with inclusion of \mathcal{C} with respect to $\mathbb{G}(\cdot | x)$. Estimates for such covering numbers are known for many classes (see, e.g., [Dudley \[1984\]](#)). We will often assume below that either $\log \mathcal{N}(\varepsilon, \mathcal{C}, \mathbb{G}^{(1)}(\cdot | x))$ or $\mathcal{N}(\varepsilon, \mathcal{C}, \mathbb{G}^{(1)}(\cdot | x))$ behave like powers of ε^{-1} : we say that the condition (R_γ) holds if

$$\log \mathcal{N}(\varepsilon, \mathcal{C}, \mathbb{G}^{(1)}(\cdot | x)) \leq H_\gamma(\varepsilon), \text{ for all } \varepsilon > 0, \quad (3.6.1)$$

where

$$H_\gamma(\varepsilon) = \begin{cases} \log(A\varepsilon) & \text{if } \gamma = 0, \\ A\varepsilon^{-\gamma} & \text{if } \gamma > 0, \end{cases}$$

for some constants $A, r > 0$. As in [Polonik and Yao \[2002\]](#), it is worth noticing that the condition (3.6.1), $\gamma = 0$, holds for intervals, rectangles, balls, ellipsoids, and for classes which are constructed from the above by performing set operations union, intersection and complement finitely many times. The classes of convex sets in \mathbb{R}^d ($d \geq 2$) fulfill the condition (3.6.1), $\gamma = (d-1)/2$. This and other classes of sets satisfying (3.6.1) with $\gamma > 0$ can be found in [Dudley \[1984\]](#). As a particular case of (3.2.3), we estimate $\mathbb{G}^{(m)}(C_1 \times \dots \times C_m | \mathbf{t})$

$$\widehat{\mathbb{G}}_n^{(m)}(\widetilde{\mathcal{C}}, \mathbf{t}) = \frac{\sum_{(i_1, \dots, i_m) \in I(m, n)} \prod_{j=1}^m \mathbb{1}_{\{Y_{i_j} \in C_j\}} \mathbb{K}\left(\frac{d(t_1, X_{i_1})}{H_{n,k}(t_1)}\right) \dots \mathbb{K}\left(\frac{d(t_m, X_{i_m})}{H_{n,k}(t_m)}\right)}{\sum_{(i_1, \dots, i_m) \in I(m, n)} \mathbb{K}\left(\frac{d(t_1, X_{i_1})}{H_{n,k}(t_1)}\right) \dots \mathbb{K}\left(\frac{d(t_m, X_{i_m})}{H_{n,k}(t_m)}\right)}. \quad (3.6.2)$$

One can apply Corollary 3.4.1.5 to infer that

$$\sup_{\widetilde{\mathcal{C}} \times \widetilde{\mathcal{K}} \in \mathcal{C}^m \mathcal{X}^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{\mathbf{t} \in \mathcal{X}^m} |\widehat{\mathbb{G}}_n^{(m)}(\widetilde{\mathcal{C}}, \mathbf{t}) - \mathbb{G}^{(m)}(\widetilde{\mathcal{C}} | \mathbf{t})| \longrightarrow 0 \quad \text{a.co.} \quad (3.6.3)$$

Remark 3.6.1.1 Another point of view is to consider the following situation, for a compact $\mathbf{J} \subset \mathbb{R}^{dm}$,

$$\mathbf{G}^{(m)}(y_1, \dots, y_m | \mathbf{t}) = \mathbb{E} \left(\prod_{i=1}^m \mathbb{1}_{\{Y_i \leq y_i\}} | (X_1, \dots, X_m) = \mathbf{t} \right) \text{ for } \mathbf{t} \in \mathcal{X}^m, (y_1, \dots, y_m) \in \mathbf{J}.$$

Let $L(\cdot)$ be a distribution in \mathbb{R}^d and $\widetilde{H}_{n,k}(t_i)$ is the number of neighborhoods associated with Y_i s. One can estimate $\mathbf{G}^{(m)}(y_1, \dots, y_m | \mathbf{t}) = \mathbf{G}^{(m)}(\mathbf{y} | \mathbf{t})$ by

$$\widehat{\mathbf{G}}_n^{(m)}(\mathbf{y}, \mathbf{t}) := \frac{\sum_{(i_1, \dots, i_m) \in I(m, n)} L\left(\frac{t_1 - Y_{i_1}}{\widetilde{H}_{n,\widetilde{k}}(t_1)}\right) \dots L\left(\frac{t_m - Y_{i_m}}{\widetilde{H}_{n,\widetilde{k}}(t_m)}\right) \mathbb{K}\left(\frac{d(t_1, X_{i_1})}{H_{n,k}(t_1)}\right) \dots \mathbb{K}\left(\frac{d(t_m, X_{i_m})}{H_{n,k}(t_m)}\right)}{\sum_{(i_1, \dots, i_m) \in I(m, n)} \mathbb{K}\left(\frac{d(t_1, X_{i_1})}{H_{n,k}(t_1)}\right) \dots \mathbb{K}\left(\frac{d(t_m, X_{i_m})}{H_{n,k}(t_m)}\right)}.$$

One can use Corollary 3.4.1.5 to infer that

$$\sup_{\widetilde{\mathcal{K}} \in \mathcal{X}^m} \sup_{k_{1,n} \leq k, \widetilde{k} \leq k_{2,n}} \sup_{\mathbf{t} \in \mathcal{X}^m} \sup_{\mathbf{y} \in \mathbf{J}} |\widehat{\mathbf{G}}_n^{(m)}(\mathbf{y}, \mathbf{t}) - \mathbf{G}^{(m)}(\mathbf{y} | \mathbf{t})| \longrightarrow 0 \quad \text{a.co.} \quad (3.6.4)$$

3.6.2 Generalized *U*-Statistics

The extension to the case of several samples is straightforward. Consider k independent collections, of independent observations

$$\left\{ \left(X_1^{(1)}, Y_1^{(1)} \right), \left(X_2^{(1)}, Y_2^{(1)} \right), \dots \right\}, \dots, \left\{ \left(X_1^{(k)}, Y_1^{(k)} \right), \left(X_2^{(k)}, Y_2^{(k)} \right), \dots \right\}.$$

Let, for $\mathbf{t} \in \mathcal{X}^{(m_1 + \dots + m_k)}$,

$$\begin{aligned} r^{(m,k)}(\varphi, \mathbf{t}) &= r^{(m,k)}(\varphi, \mathbf{t}_1, \dots, \mathbf{t}_k) \\ &= \mathbb{E} \left(\varphi \left(Y_1^{(1)}, \dots, Y_{m_1}^{(1)}; \dots; Y_1^{(k)}, \dots, Y_{m_k}^{(k)} \right) \mid \left(X_1^{(j)}, \dots, X_{m_j}^{(j)} \right) = \mathbf{t}_j, j = 1, \dots, k \right), \end{aligned}$$

where φ is assumed, without loss of generality, to be symmetric within each of its k blocks of arguments. Corresponding to the "kernel" φ and assuming $n_1 \geq m_1, \dots, n_k \geq m_k$, the conditional *U*-statistic for estimation of $r^{(m,k)}(\varphi, \mathbf{t})$ is defined as

$$\begin{aligned} \hat{r}_n^{(m,k)}(\varphi, \mathbf{t}, \tilde{\mathbf{h}}_{n,k}(\mathbf{t})) \\ = \frac{\sum_c \varphi \left(Y_{i_{11}}^{(1)}, \dots, Y_{i_{1m_1}}^{(1)}; \dots; Y_{i_{k1}}^{(k)}, \dots, Y_{i_{km_k}}^{(k)} \right) \mathbf{K} \left(X_{i_{11}}^{(1)}, \dots, X_{i_{1m_1}}^{(1)}; \dots; X_{i_{k1}}^{(k)}, \dots, X_{i_{km_k}}^{(k)} \right)}{\sum_c \mathbf{K} \left(X_{i_{11}}^{(1)}, \dots, X_{i_{1m_1}}^{(1)}; \dots; X_{i_{k1}}^{(k)}, \dots, X_{i_{km_k}}^{(k)} \right)}, \end{aligned}$$

where

$$\mathbf{K} \left(X_{i_{11}}^{(1)}, \dots, X_{i_{1m_1}}^{(1)}; \dots; X_{i_{k1}}^{(k)}, \dots, X_{i_{km_k}}^{(k)} \right) = \prod_{j=1}^k \mathbf{K} \left(\frac{d(t_1^{(j)}, X_{i_{j1}}^{(j)})}{H_{n,k}(t_1)} \right) \dots \mathbf{K} \left(\frac{d(t_{m_j}^{(j)}, X_{i_{jm_j}}^{(j)})}{H_{n,k}(t_m)} \right),$$

and $\tilde{\mathbf{h}}_{n,k}(\mathbf{t}) = \left(H_{n,k}(t_1^{(1)}), \dots, H_{n,k}(t_{m_1}^{(1)}), \dots, H_{n,k}(t_1^{(k)}), \dots, H_{n,k}(t_{m_k}^{(k)}) \right) \in \mathcal{H}_n^{m_1 + \dots + m_k} \subset \mathbb{R}_+^{m_1 + \dots + m_k}$

is a vector of positive random variables which depend on $\{X_1^{(1)}, \dots, X_{n_1}^{(1)}\}, \dots, \{X_1^{(k)}, \dots, X_{n_k}^{(k)}\}$. Here $\{i_{j1}, \dots, i_{jm_j}\}$ denotes a set of m_j distinct elements of the set $\{1, 2, \dots, n_j\}$ $1 \leq j \leq k$, and \sum_c denotes summation over all such combinations. The extension of [Hoeffding \[1948\]](#) treatment of one-sample *U*-statistics to the k sample case is due to [Lehmann \[1951\]](#) and [Dwass \[1956\]](#). One can use [Corollary 3.4.1.5](#) to infer that

$$\sup_{\varphi \in \mathcal{F}_{m_1 + \dots + m_k}} \max_{\mathcal{X}^{m_1 + \dots + m_k}} \sup_{1 \leq j \leq k} \sup_{k_{1,n}^j \leq k_j \leq k_{2,n}^j} \sup_{t \in \mathcal{X}^{m_1 + \dots + m_k}} \left| \hat{r}_n^{(m,k)}(\varphi, \mathbf{t}, \tilde{\mathbf{h}}_{n,k}(\mathbf{t})) - r^{(m,k)}(\varphi, \mathbf{t}) \right| \longrightarrow 0 \quad \text{a.co.} \quad (3.6.5)$$

3.6.3 Discrimination

Now, we apply the results on the problem of discrimination described in Section 3 of [Stute \[1994b\]](#), refer to also to [Stute \[1994a\]](#). We will use similar notations and settings. Let $\varphi(\cdot)$ be any function taking at most finitely many values, say $1, \dots, M$. The sets

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

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

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

where

$$m^j(\mathbf{x}) = \mathbb{P}(\varphi(\mathbf{Y}) = j \mid \mathbf{X} = \mathbf{t}), \quad \mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m.$$

The above inequality becomes equality if

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

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

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

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

$$m_n^j(\mathbf{t}) = \frac{\sum_{(i_1, \dots, i_m) \in \mathcal{I}(m, n)} \mathbb{1}\{\varphi(Y_{i_1}, \dots, Y_{i_m}) = j\} \mathbb{K} \left(\frac{d(t_1, X_{i_1})}{H_{n,k}(t_1)} \right) \dots \mathbb{K} \left(\frac{d(t_m, X_{i_m})}{H_{n,k}(t_m)} \right)}{\sum_{(i_1, \dots, i_m) \in \mathcal{I}(m, n)} \mathbb{K} \left(\frac{d(t_1, X_{i_1})}{H_{n,k}(t_1)} \right) \dots \mathbb{K} \left(\frac{d(t_m, X_{i_m})}{H_{n,k}(t_m)} \right)}, \quad (3.6.6)$$

Set

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

Let us introduce

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

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

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

3.6.4 Metric learning

Metric learning aims at adapting the metric to the data and has attracted a lot of interest in recent years; for instance, see [Bellet *et al.* \[2013\]](#) and [Cl  men  on *et al.* \[2016\]](#) for an account of metric learning and its applications. This is motivated by a variety of applications ranging from computer vision to information retrieval through bioinformatics. To illustrate the usefulness of this concept, we present the metric learning problem for supervised classification as in [Cl  men  on *et al.* \[2016\]](#). Let us consider independent copies $(X_1, Y_1), \dots, (X_n, Y_n)$ of a $\mathcal{X} \times \mathcal{Y}$ valued random couple (X, Y) , where \mathcal{X} is some feature space and $\mathcal{Y} = \{1, \dots, C\}$, with $C \geq 2$, say, a finite set of labels. Let \mathcal{D} be a set of distance measures $D : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_+$. Intuitively, the goal

of metric learning in this setting is to find a metric under which pairs of points with the same label are close to each other, and those with different labels are far away. The natural way to define the risk of a metric D is

$$R(D) = \mathbb{E} [\phi((1 - D(X, X')) \cdot (2\mathbb{1}\{Y = Y'\} - 1))], \quad (3.6.7)$$

where $\phi(u)$ is a convex loss function upper bounding the indicator function $\mathbb{1}\{u \geq 0\}$, for instance, the hinge loss $\phi(u) = \max(0, 1 - u)$. To estimate $R(D)$, we consider the natural empirical estimator

$$R_n(D) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \phi((D(X_i, X_j) - 1) \cdot (2\mathbb{1}\{Y_i = Y_j\} - 1)), \quad (3.6.8)$$

which is a one-sample U-statistic of degree two with kernel given by:

$$H_D((x, y), (x', y)) = \phi((D(x, x') - 1) \cdot (2\mathbb{1}\{y = y'\} - 1)).$$

The convergence to (3.6.7) of a minimizer of (3.6.8) has been studied in the frameworks of algorithmic stability (Jin *et al.* [2009]), algorithmic robustness (Bellet and Habrard [2015]) and based on the theory of U-processes under appropriate regularization (Cao *et al.* [2016]).

3.7 Examples of kernel for the U-statistics

Generally speaking, we may take for $\varphi(\cdot)$ any function which has been found interesting in the unconditional setup; cf. Serfling [1980]. As mentioned before, the case $m = 1$ leads to the Nadaraya-Watson estimator if we set $\varphi = Id$ (the identity function); $\varphi = \mathbb{1}\{\cdot \leq t\}$ yields the conditional d.f. evaluated at t , for instance, see Stute [1986].

Example 3.7.0.1 For

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

we obtain a conditional U-statistic which may be viewed as a conditional version of the Wilcoxon one-sample statistic. It may be used for testing the hypothesis that the conditional distribution at X_1 is symmetric at zero. Obviously :

$$r^{(2)}(\varphi(Y_1, Y_2) \mid t_1 = t_2) = \mathbb{P}(Y_1 + Y_2 > 0 \mid X_1 = t_1 = X_2).$$

Example 3.7.0.2 For

$$\varphi(Y_1, Y_2) = \mathbb{1}\{Y_1 \leq Y_2\},$$

we have

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

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

Example 3.7.0.3 For $m=3$, let :

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

we have

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

and the corresponding CU-S can be looked upon as a conditional analogue of the Hollander-Proschan test statistic ([Hollander and Proschan \[1972\]](#)). It may be used to test the hypothesis that the conditional distribution of Y_1 given $X_1 = t$ is exponential against the alternative that it is of the New-Better than-Used-type.

Example 3.7.0.4 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 3.7.0.5 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; for more details, refer to [Jadhav and Ma \[2019\]](#) for the unconditional setting.

Example 3.7.0.6 [Hoeffding \[1948\]](#) introduced the parameter

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

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 Δ has the property that $\Delta = 0$ if and only if Y_1 and Y_2 are independent. From [Lee \[1990\]](#), an alternative expression for Δ can be developed by introducing the functions

$$\Psi(y_1, y_2, y_3) = \begin{cases} 1 & \text{if } y_2 \leq y_1 < y_3 \\ 0 & \text{if } y_1 < y_2, y_3 \text{ or } y_1 \geq y_2, y_3 \\ -1 & \text{if } y_3 \leq y_1 < y_2 \end{cases}$$

and

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

We have

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

We have

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

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

Example 3.7.0.7 (Blum-Kiefer-Rosenblatt's R). Write \mathcal{P}_k for the set of all $k!$ permutations of $\{1, \dots, k\}$. The symmetric kernel

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

yields Blum-Kiefer-Rosenblatt's R statistic (Blum et al. [1961]), which is a rank-based U-statistic of order 6.

Remark 3.7.0.8 It is important to mention that the FDA is a very active field in the current statistical community because of its wide range of applications. The literature on the subject has been impressively large but mostly restricted to the standard multivariate situation where both X and Y are real or multivariate. However, when the response is also of functional feature, although functional linear models have been well developed, see, for instance, Ramsay and Dalzell [1991], Müller and Yao [2008] and the references therein, the functional nonparametric model when both X and Y are functional has not been well developed and we can cite only a few reports in this direction, Ling et al. [2017, 2020] and the references therein.

3.8 The bandwidth selection criterion

Many methods have been established and developed to construct, in asymptotically optimal ways, bandwidth selection rules for nonparametric kernel estimators, especially for Nadaraya-Watson regression estimator we quote among them Hall [1984], Härdle and Marron [1985] and Rachdi and Vieu [2007]. This parameter has to be selected suitably, either in the standard finite-dimensional case or in the infinite-dimensional framework, to ensure good practical performance. However, according to our knowledge, such studies do not presently exist for treating such a general functional conditional U-statistic (unless the real case we could find in the paper of Dony and Mason [2008] a paragraph devoted to the selection of the number k). Nevertheless an extension of the leave-one-out cross-validation procedure allows to define, for any fixed $\mathbf{i} = (i_1, \dots, i_m) \in \mathcal{I}(m, n)$:

$$\hat{r}_{n,\mathbf{i}}^{*(m)}(\varphi, \mathbf{t}, \mathbf{h}_{n,k}(\mathbf{t})) = \frac{\sum_{\mathbf{j} \in \mathcal{I}_n^m(\mathbf{i})} \varphi(Y_{j_1}, \dots, Y_{j_m}) \mathbb{K}\left(\frac{d(t_1, X_{j_1})}{H_{n,k}(t_1)}\right) \dots \mathbb{K}\left(\frac{d(t_m, X_{j_m})}{H_{n,k}(t_m)}\right)}{\sum_{\mathbf{j} \in \mathcal{I}_n^m(\mathbf{i})} \mathbb{K}\left(\frac{d(t_1, X_{j_1})}{H_{n,k}(t_1)}\right) \dots \mathbb{K}\left(\frac{d(t_m, X_{j_m})}{H_{n,k}(t_m)}\right)}, \quad (3.8.1)$$

where

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

The equation (3.8.1) represents the leave-out- $(\mathbf{X}_i, \mathbf{Y}_i)$ estimator of the functional regression and also could be considered as a predictor of $\varphi(\mathbf{Y}_i)$. In order to minimize the quadratic loss function, we introduce the following criterion; we have for some (known) non-negative weight function $\mathcal{W}(\cdot)$:

$$CV(\varphi, k) := \sum_{\mathbf{i} \in I(m, n)} \left(\varphi(\mathbf{Y}_i) - \hat{r}_{n, \mathbf{i}}^{*(m)}(\varphi, \mathbf{X}_i, \mathbf{h}_{n, k}(\mathbf{X}_i)) \right)^2 \widetilde{\mathcal{W}}(\mathbf{X}_i), \quad (3.8.2)$$

where

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

Following the ideas developed by [Rachdi and Vieu \[2007\]](#), a natural way for choosing the bandwidth is to minimize the precedent criterion, so let's choose $\hat{k} \in [k_{1, n}, k_{2, n}]$ minimizing among $k \in [k_{1, n}, k_{2, n}]$:

$$\sup_{\varphi \in \mathcal{F}_m, \mathcal{K}^m} CV(\varphi, k),$$

we can conclude, by Corollary 3.4.1.5, that :

$$\sup_{\varphi \in \mathcal{F}_m, \mathcal{K}^m} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \left| \hat{r}_{n, \hat{k}}^{*(m)}(\varphi, \mathbf{t}, \mathbf{h}_{n, \hat{k}}(\mathbf{t})) - r^{(m)}(\varphi, \mathbf{t}) \right| \longrightarrow 0, \quad \text{a.co.}$$

The main interest of our results is the possibility of deriving asymptotics for any automatic data-driven parameters. Let $K'(\cdot)$ be a density function in \mathbb{R}^d and $H'_{n, k'}(t_i)$ is the number of neighborhoods associated with Y_i s. One can estimate the conditional density $\mathbf{f}^{(m)}(y_1, \dots, y_m | \mathbf{t}) = \mathbf{f}^{(m)}(\mathbf{y} | \mathbf{t})$ by

$$\begin{aligned} & \hat{\mathbf{f}}_n^{(m)}(\mathbf{y}, \mathbf{t}) \\ := & \frac{\sum_{(i_1, \dots, i_m) \in I(m, n)} K' \left(\frac{t_1 - Y_{i_1}}{H'_{n, k'}(t_1)} \right) \dots K' \left(\frac{t_m - Y_{i_m}}{H_{n, k'}(t_m)} \right) K \left(\frac{d(t_1, X_{i_1})}{H_{n, k}(t_1)} \right) \dots K \left(\frac{d(t_m, X_{i_m})}{H_{n, k}(t_m)} \right)}{\sum_{(i_1, \dots, i_m) \in I(m, n)} K \left(\frac{d(t_1, X_{i_1})}{H_{n, k}(t_1)} \right) \dots K \left(\frac{d(t_m, X_{i_m})}{H_{n, k}(t_m)} \right)}. \end{aligned}$$

Hence the leave-one-out estimator is given by

$$\begin{aligned} & \hat{\mathbf{f}}_{n, \mathbf{i}}^{(m)}(\mathbf{y}, \mathbf{t}) \\ := & \frac{\sum_{(j_1, \dots, j_m) \in I_n^m(\mathbf{i})} K' \left(\frac{t_1 - Y_{j_1}}{H'_{n, k'}(t_1)} \right) \dots K' \left(\frac{t_m - Y_{j_m}}{H_{n, k'}(t_m)} \right) K \left(\frac{d(t_1, X_{j_1})}{H_{n, k}(t_1)} \right) \dots K \left(\frac{d(t_m, X_{j_m})}{H_{n, k}(t_m)} \right)}{\sum_{(i_1, \dots, i_m) \in I(m, n)} K \left(\frac{d(t_1, X_{i_1})}{H_{n, k}(t_1)} \right) \dots K \left(\frac{d(t_m, X_{i_m})}{H_{n, k}(t_m)} \right)}. \end{aligned}$$

While the cross-validation procedures described above aim to approximate quadratic errors of estimation, alternative ways for choosing smoothing parameters could be introduced, aiming instead to optimize the predictive power of the method. The criterion is given by

$$(\check{k}, \check{k}') = \arg \min_{k_{1, n} \leq k \leq k_{2, n}, k'_{1, n} \leq k' \leq k'_{2, n}} \sum_{\mathbf{i} \in I(m, n)} \left(\varphi(\mathbf{Y}_i) - \arg \max_{\mathbf{y} \in \mathcal{J}} \hat{\mathbf{f}}_{n, \mathbf{i}}^{(m)}(\mathbf{y}, \mathbf{t}) \right)^2,$$

3.9 Mathematical developments

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

3.9.1 General Lemma

This lemma presents an important tool that will allow us to derive asymptotic results for k -NN estimators by using the results from functional conditional U -statistics of [Bouzebda and Nemouchi \[2020\]](#). We also present Lemma 3.9.1.1 in a general setting which could be useful in many other situations than ours; this is a generalization of a result obtained in [Burba *et al.* \[2009\]](#). More generally, this technical tool could be useful as long as one has to deal with random bandwidths.

Let $(\mathbf{A}_i, \mathbf{B}_i)_{n \geq 1}$ be n random vectors valued in $(\Omega^m \times \Omega^{lm}, \mathcal{A} \times \mathcal{B})$, a general space. Let S_Ω be a fixed subset of Ω and we note that $G : \mathbb{R} \times (S_\Omega \times \Omega) \rightarrow \mathbb{R}^+$ a function such that, $\forall t \in S_\Omega$, $G(\cdot, (t, \cdot))$ is measurable and $\forall x, x' \in \mathbb{R}$:

(L₀)

$$x \leq x' \Rightarrow G(x, \mathbf{z}) \leq G(x', \mathbf{z}), \forall \mathbf{z} \in S_\Omega \times \Omega.$$

We define the pointwise measurable class of functions for $1 \leq m \leq n$:

$$\mathcal{G}^m := \left\{ (x_1, \dots, x_m) \mapsto \prod_{i=1}^m G(h_i, (x_i, t_i)), (h_1, \dots, h_m) \in \mathbb{R}_+^m \setminus \{0\} \text{ and } (t_1, \dots, t_m) \in S_\Omega^m \right\}.$$

Let $(\mathbf{D}_{n,k}(\mathbf{t}))_{n \in \mathbb{N}}$ be a sequence of random real vectors (r.r.v.) such that for all $\mathbf{t} = (t_1, \dots, t_m) \in S_\Omega^m$, $\mathbf{D}_{n,k}(\mathbf{t}) = (D_{n,k}(t_1), \dots, D_{n,k}(t_m))$, and $\varphi : \Omega^m \rightarrow \mathbb{R}$ be a measurable function belonging to some class of functions \mathcal{F}_m and let $\mathfrak{M}^{(m)} : \mathcal{F}_m \times S_\Omega^m \rightarrow \mathbb{R}$ be a nonrandom function such that,

$$\sup_{\varphi \in \mathcal{F}_m} \sup_{\mathbf{t} \in S_\Omega^m} |\mathfrak{M}^{(m)}(\varphi, \mathbf{t})| < \infty.$$

Now, for all $\mathbf{t} \in S_\Omega^m$, $\varphi \in \mathcal{F}_m$ and $n \in \mathbb{N} \setminus \{0\}$ we define

$$\mathfrak{M}_n^{(m)}(\varphi, \mathbf{t}; \mathbf{h}) = \frac{\sum_{i \in I(m,n)} \varphi(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \prod_{j=1}^m G(H_j, (t_j, A_{i,j}))}{\sum_{i \in I(m,n)} \prod_{j=1}^m G(H_j, (t_j, A_{i,j}))},$$

where $\mathbf{h} = (H_1, \dots, H_m) \in \mathbb{R}_+^m$, and $\tilde{G} = \prod_{i=1}^m G(h_i, (x_i, t_i))$, $\mathbf{t} = (t_1, \dots, t_m) \in S_\Omega^m$.

Lemma 3.9.1.1 *Let $(\mathcal{U}_n)_{n \in \mathbb{N}}$ be a decreasing positive sequence such that $\lim_{n \rightarrow \infty} \mathcal{U}_n = 0$. If, for all increasing sequence $\beta_n \in (0, 1)$ with $\beta_n - 1 = O(\mathcal{U}_n)$, there exist two sequences of r.r.v. $(\mathbf{D}_{n,k}^-(\beta_n, \mathbf{t}))_{n \in \mathbb{N}}$ and $(\mathbf{D}_{n,k}^+(\beta_n, \mathbf{t}))_{n \in \mathbb{N}}$ such that*

(L₁)

$$\forall n \in \mathbb{N} \text{ and } \mathbf{t} \in \mathcal{S}_\Omega^m, D_{n,k}^-(\beta_n, t_j) \leq D_{n,k}^+(\beta_n, t_j), \forall j = 1, \dots, m,$$

 (L₂)

$$\prod_{j=1}^m \mathbb{1}_{\{D_{n,k}^-(\beta_n, t_j) \leq D_{n,k}(t_j) \leq D_{n,k}^+(\beta_n, t_j)\}} \rightarrow 1, \text{ a.co.}, \forall \mathbf{t} \in \mathcal{S}_\Omega^m.$$

 (L₃)

$$\sup_{\tilde{G} \in \mathcal{G}^m} \sup_{\mathbf{t} \in \mathcal{S}_\Omega^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} \left| \frac{\sum_{i \in I(m,n)} \prod_{j=1}^m G(D_{n,k}^-(\beta_n, t_j), (t_j, A_{i,j}))}{\sum_{i \in I(m,n)} \prod_{j=1}^m G(D_{n,k}^+(\beta_n, t_j), (t_j, A_{i,j}))} - \beta_n \right| = O_{a.co}(\mathcal{U}_n),$$

 (L₄)

$$\sup_{\varphi \in \mathcal{F}_m} \sup_{\tilde{G} \in \mathcal{G}^m} \sup_{\mathbf{t} \in \mathcal{S}_\Omega^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} \left| \mathfrak{M}_n^{(m)}(\varphi, \mathbf{t}; D_{n,k}^-(\beta_n, \mathbf{t})) - \mathfrak{M}^{(m)}(\varphi, \mathbf{t}) \right| = O_{a.co}(\mathcal{U}_n),$$

 (L₅)

$$\sup_{\varphi \in \mathcal{F}_m} \sup_{\tilde{G} \in \mathcal{G}^m} \sup_{\mathbf{t} \in \mathcal{S}_\Omega^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} \left| \mathfrak{M}_n^{(m)}(\varphi, \mathbf{t}; D_{n,k}^+(\beta_n, \mathbf{t})) - \mathfrak{M}^{(m)}(\varphi, \mathbf{t}) \right| = O_{a.co}(\mathcal{U}_n).$$

Then

$$\sup_{\varphi \in \mathcal{F}_m} \sup_{\tilde{G} \in \mathcal{G}^m} \sup_{\mathbf{t} \in \mathcal{S}_\Omega^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} \left| \mathfrak{M}_n^{(m)}(\varphi, \mathbf{t}; D_{n,k}(\mathbf{t})) - \mathfrak{M}^{(m)}(\varphi, \mathbf{t}) \right| = O_{a.co}(\mathcal{U}_n). \quad (3.9.1)$$

Proof of Lemma 3.9.1.1

The proof of this Lemma extends the scheme of [Burba et al. \[2009\]](#) and [Kudraszow and Vieu \[2013\]](#). For technical reasons, we assume that the function $\varphi(\cdot)$ is a non-negative function. To obtain the results for any real valued function $\varphi(\cdot)$ we can write $\varphi(\cdot) = \varphi^+(\cdot) - \varphi^-(\cdot)$, thus, for all $\mathbf{y} \in \Omega^m$ we have:

$$\varphi^+(\mathbf{y}) = \max(\varphi(\mathbf{y}), 0) \text{ and } \varphi^-(\mathbf{y}) = -\min(\varphi(\mathbf{y}), 0).$$

Let us denote

$$\begin{aligned} \mathfrak{M}_{n-}^{(m)}(\varphi, \mathbf{t}; \beta_n) &= \frac{\sum_{i \in I(m,n)} \varphi(B_{i,1}, \dots, B_{i,m}) \prod_{j=1}^m G(D_{n,k}^-(\beta_n, t_j), (t_j, A_{i,j}))}{\sum_{i \in I(m,n)} \prod_{j=1}^m G(D_{n,k}^+(\beta_n, t_j), (t_j, A_{i,j}))}, \\ \mathfrak{M}_{n+}^{(m)}(\varphi, \mathbf{t}; \beta_n) &= \frac{\sum_{i \in I(m,n)} \varphi(B_{i,1}, \dots, B_{i,m}) \prod_{j=1}^m G(D_{n,k}^+(\beta_n, t_j), (t_j, A_{i,j}))}{\sum_{i \in I(m,n)} \prod_{j=1}^m G(D_{n,k}^-(\beta_n, t_j), (t_j, A_{i,j}))}. \end{aligned}$$

For all sequence $\beta_n \in (0, 1)$ with $\beta_n - 1 = O(\mathcal{U}_n)$, we deduce from (L₃) and (L₄) that

$$\begin{aligned} & \sup_{\varphi \tilde{G} \in \mathcal{F}_m \mathcal{G}^m} \sup_{\mathbf{t} \in S_\Omega^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} |\mathfrak{M}_{n-}^{(m)}(\varphi, \mathbf{t}; \beta_n) - \mathfrak{M}^{(m)}(\varphi, \mathbf{t})| \\ & \leq \sup_{\varphi \tilde{G} \in \mathcal{F}_m \mathcal{G}^m} \sup_{\mathbf{t} \in S_\Omega^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} |\mathfrak{M}_{n-}^{(m)}(\varphi, \mathbf{t}; \beta_n) - \beta_n \mathfrak{M}^{(m)}(\varphi, \mathbf{t})| \\ & \quad + |\mathfrak{M}^{(m)}(\varphi, \mathbf{t})| |\beta_n - 1| \end{aligned} \quad (3.9.2)$$

$$= O_{a.co}(\mathcal{U}_n). \quad (3.9.3)$$

In the same way, (L₃) and (L₅) give us

$$\sup_{\varphi \tilde{G} \in \mathcal{F}_m \mathcal{G}^m} \sup_{\mathbf{t} \in S_\Omega^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} |\mathfrak{M}_{n+}^{(m)}(\varphi, \mathbf{t}; \beta_n) - \mathfrak{M}^{(m)}(\varphi, \mathbf{t})| = O_{a.co}(\mathcal{U}_n). \quad (3.9.4)$$

For all $\varepsilon > 0$, we note

$$T_n(\varepsilon_n) = \left\{ \sup_{\varphi \tilde{G} \in \mathcal{F}_m \mathcal{G}^m} \sup_{\mathbf{t} \in S_\Omega^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} |\mathfrak{M}_n^{(m)}(\varphi, \mathbf{t}; \mathbf{D}_{n,k}(\mathbf{t})) - \mathfrak{M}^{(m)}(\varphi, \mathbf{t})| \leq \varepsilon \mathcal{U}_n \right\},$$

and for all sequence $\beta_n \in (0, 1)$ with $\beta_n - 1 = O(\mathcal{U}_n)$

$$\begin{aligned} S_n^-(\varepsilon_n, \beta_n) &= \left\{ \sup_{\varphi \tilde{G} \in \mathcal{F}_m \mathcal{G}^m} \sup_{\mathbf{t} \in S_\Omega^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} |\mathfrak{M}_{n-}^{(m)}(\varphi, \mathbf{t}; \beta_n) - \mathfrak{M}^{(m)}(\varphi, \mathbf{t})| \leq \varepsilon \mathcal{U}_n \right\}, \\ S_n^+(\varepsilon_n, \beta_n) &= \left\{ \sup_{\varphi \tilde{G} \in \mathcal{F}_m \mathcal{G}^m} \sup_{\mathbf{t} \in S_\Omega^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} |\mathfrak{M}_{n+}^{(m)}(\varphi, \mathbf{t}; \beta_n) - \mathfrak{M}^{(m)}(\varphi, \mathbf{t})| \leq \varepsilon \mathcal{U}_n \right\}, \\ S_n(\beta_n) &= \left\{ \mathfrak{M}_{n-}^{(m)}(\varphi, \mathbf{t}; \beta_n) \leq \mathfrak{M}^{(m)}(\varphi, \mathbf{t}) \leq \mathfrak{M}_{n+}^{(m)}(\varphi, \mathbf{t}; \beta_n), \forall \mathbf{t} \in S_\Omega^m \right\}. \end{aligned}$$

Then for all $\beta_n \in (0, 1)$ with $\beta_n - 1 = O(\mathcal{U}_n)$,

$$\forall \varepsilon_n > 0, S_n^-(\varepsilon_n, \beta_n) \cap S_n^+(\varepsilon_n, \beta_n) \cap S_n(\beta_n) \subset T_n(\varepsilon_n). \quad (3.9.5)$$

For all $j = 1, \dots, m$, set

$$G_{n,j}(\beta_n) = \left\{ D_{n,k}^-(\beta_n, t_j) \leq D_{n,k}(\beta_n, t_j) \leq D_{n,k}^+(\beta_n, t_j), \forall \mathbf{t} \in S_\Omega^m \right\},$$

and

$$\mathbf{G}_n(\beta_n) = \bigcap_{j=1}^m G_{n,j}(\beta_n).$$

Then (L₀) implies that $G_n(\beta_n) \subset S_n(\beta_n)$ and, from (3.9.5), we have

$$\forall \varepsilon_n > 0, T_n(\varepsilon_n)^c \subset S_n^-(\varepsilon_n, \beta_n)^c \cup S_n^+(\varepsilon_n, \beta_n)^c \cup G_n(\beta_n)^c.$$

Consequently,

$$\begin{aligned}
 & \mathbb{P} \left\{ \sup_{\varphi \in \tilde{\mathcal{G}}} \sup_{\mathcal{F}_m} \sup_{\mathcal{G}^m} \sup_{\mathbf{t} \in \mathcal{S}_\Omega^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} \left| \mathfrak{M}_n^{(m)}(\varphi, \mathbf{t}; \mathbf{D}_{n,k}(\mathbf{t})) - \mathfrak{M}^{(m)}(\varphi, \mathbf{t}) \right| > \varepsilon \mathcal{U}_n \right\} \\
 & \leq \mathbb{P} \left\{ \sup_{\varphi \in \tilde{\mathcal{G}}} \sup_{\mathcal{F}_m} \sup_{\mathcal{G}^m} \sup_{\mathbf{t} \in \mathcal{S}_\Omega^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} \left| \mathfrak{M}_{n^-}^{(m)}(\varphi, \mathbf{t}; \beta_{n,\varepsilon_n}) - \mathfrak{M}^{(m)}(\varphi, \mathbf{t}) \right| > \varepsilon \mathcal{U}_n \right\} \\
 & \quad + \mathbb{P} \left\{ \sup_{\varphi \in \tilde{\mathcal{G}}} \sup_{\mathcal{F}_m} \sup_{\mathcal{G}^m} \sup_{\mathbf{t} \in \mathcal{S}_\Omega^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} \left| \mathfrak{M}_{n^+}^{(m)}(\varphi, \mathbf{t}; \beta_{n,\varepsilon_n}) - \mathfrak{M}^{(m)}(\varphi, \mathbf{t}) \right| > \varepsilon \mathcal{U}_n \right\} \\
 & \quad + \sum_{j=1}^m \mathbb{P} \left\{ \mathbb{1}_{\left\{ D_{n,k}^-(\beta_n, t_j) \leq D_{n,k}(\beta_n, t_j) \leq D_{n,k}^+(\beta_n, t_j) \right\}} = 0 \right\}.
 \end{aligned}$$

Taking into account (L₂), (3.9.2) and (3.9.4), we get that, for $\varepsilon_0 > 0$, we have

$$\sum_{n=1}^{\infty} \mathbb{P} \left\{ \sup_{\varphi \in \tilde{\mathcal{G}}} \sup_{\mathcal{F}_m} \sup_{\mathcal{G}^m} \sup_{\mathbf{t} \in \mathcal{S}_\Omega^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} \left| \mathfrak{M}_n^{(m)}(\varphi, \mathbf{t}; \mathbf{D}_{n,k}(\mathbf{t})) - \mathfrak{M}^{(m)}(\varphi, \mathbf{t}) \right| > \varepsilon_0 \mathcal{U}_n \right\} < \infty. \quad (3.9.6)$$

Hence the proof is complete. \square

Proofs of the intermediate results

Using the fact that

$$\hat{r}_n^{(1)}(\varphi, t; h_K) = \frac{W_n(\varphi K, t, h_K)}{Q_n(K, t, h_K)},$$

where

$$W_n(\varphi K, t; h_K) := \frac{1}{n\phi_t(h_K)} \sum_{i=1}^n \varphi(Y_i) K\left(\frac{d(t, X_i)}{h_K}\right),$$

and

$$Q_n(K, t; h_K) := \frac{1}{n\phi_t(h_K)} \sum_{i=1}^n K\left(\frac{d(t, X_i)}{h_K}\right),$$

we can proceed as follows:

$$\begin{aligned}
 & \hat{r}_n^{(1)}(\varphi, t; h_K) - r^{(1)}(\varphi, t) \\
 & = \frac{1}{Q_n(K, t; h_K)} (W_n(\varphi K, t; h_K) - \mathbb{E}[W_n(\varphi K, t; h_K)]) \\
 & \quad - \frac{\mathbb{E}[W_n(\varphi K, t; h_K)]}{Q_n(K, t; h_K) \mathbb{E}[Q_n(K, t; h_K)]} (Q_n(K, t; h_K) - \mathbb{E}[Q_n(K, t; h_K)]) \\
 & \quad - \left(\mathbb{E}(\varphi(Y)|X=t) - \frac{\mathbb{E}[W_n(\varphi K, t; h_K)]}{\mathbb{E}[Q_n(K, t; h_K)]} \right).
 \end{aligned}$$

The following lemmas are instrumental in the proof of Theorem 3.3.2.1.

Lemma 3.9.1.2 (a) *Under the assumptions (A.1)-(A.3), (A.4.b), (A.6) and if we suppose that (3.2.9) holds, then we get:*

$$\sup_{K \in \mathcal{K}} \sup_{a_n \leq h_K \leq b_n} \sup_{t \in \mathcal{S}_X} |Q_n(K, t; h_K) - \mathbb{E}[Q_n(K, t; h_K)]| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_X} \left(\frac{\log n}{n} \right)}{n\phi(a_n)}} \right). \quad (3.9.7)$$

(b) If in addition assumption (A.7.) holds, then we have

$$\sup_{K \in \mathcal{K}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |Q_n(K, t; h_{n,k}(t)) - \mathbb{E}(Q_n(K, t; h_{n,k}(t)))| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}}\left(\frac{\log n}{n}\right)}{n\Phi\left(\mu\Phi^{-1}\left(\frac{\rho_n k_{1,n}}{n}\right)\right)}} \right). \quad (3.9.8)$$

Moreover, there exists $C > 0$ such that

$$\sum_{n=1}^{\infty} \mathbb{P} \left(\inf_{K \in \mathcal{K}} \inf_{k_{1,n} \leq k \leq k_{2,n}} \inf_{t \in \mathcal{S}_{\mathcal{X}}} Q_n(K, t, h_{n,k}(t)) < C \right) < \infty. \quad (3.9.9)$$

Lemma 3.9.1.3 (a) If we suppose that the assumptions (A.1.)-(A.5.) and (A.6.) hold, then we get

$$\sup_{\varphi K \in \mathcal{F}} \sup_{a_n \leq h_K \leq b_n} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |W_n(\varphi K, t; h_K) - \mathbb{E}(W_n(\varphi K, t; h_K))| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}}\left(\frac{\log n}{n}\right)}{n\Phi(a_n)}} \right). \quad (3.9.10)$$

(b) If in addition assumption (A.7.) holds, then we have

$$\sup_{\varphi K \in \mathcal{F}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |W_n(\varphi K, t; h_{n,k}(t)) - \mathbb{E}(W_n(\varphi K, t; h_{n,k}(t)))| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}}\left(\frac{\log n}{n}\right)}{n\Phi\left(\mu\Phi^{-1}\left(\frac{\rho_n k_{1,n}}{n}\right)\right)}} \right). \quad (3.9.11)$$

In order to complete this proof, we need to restate Lemma 3.9.1.3 for the unbounded class of functions:

Lemma 3.9.1.4 (a) If we suppose that the assumptions (A.1.)-(A.3.) and (A.4.a'), (A.4.b), (A.5.), (A.6.) hold then we get :

$$\sup_{\varphi K \in \mathcal{F}} \sup_{a_n \leq h_K \leq b_n} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |W_n(\varphi K, t; h_K) - \mathbb{E}(W_n(\varphi K, t; h_K))| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}}\left(\frac{\log n}{n}\right)}{n\Phi(a_n)}} \right). \quad (3.9.12)$$

(b) If in addition assumption (A.7.) holds, then we have

$$\sup_{\varphi K \in \mathcal{F}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |W_n(\varphi K, t; h_{n,k}(t)) - \mathbb{E}(W_n(\varphi K, t; h_{n,k}(t)))| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}}\left(\frac{\log n}{n}\right)}{n\Phi\left(\mu\Phi^{-1}\left(\frac{\rho_n k_{1,n}}{n}\right)\right)}} \right). \quad (3.9.13)$$

Remark 3.9.1.5 The statements (a) in each one of the above lemmas is already proved in Bouzebda and Nemouchi [2020], but since some of the elements of these proofs are needed in the following sections, we will present a brief version of those proofs and adapt them to our case.

Proof of Lemma 3.9.1.2

Proof of Lemma 3.9.1.2 (a)

Taking condition (A.1.a) in consideration, it suffices to prove that there exists $\epsilon_0 > 0$ and $b_0 > 0$ such that

$$\sum_{n \geq 1} \mathbb{P} \left\{ \sup_{K \in \mathcal{K}} \sup_{a_n \leq h_K \leq b_0} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sqrt{\frac{n\phi(a_n)}{\Psi_{\mathcal{S}_{\mathcal{X}}}\left(\frac{\log n}{n}\right)}} |Q_n(K, t; h_K) - \mathbb{E}(Q_n(K, t; h_K))| \geq \epsilon_0 \right\} < \infty.$$

Let $\{t_1, \dots, t_{N_\epsilon(\mathcal{S}_{\mathcal{X}})}\}$ be an ϵ -net for $\mathcal{S}_{\mathcal{X}}$, we know that:

$$\mathcal{S}_{\mathcal{X}} \subseteq \bigcup_{\ell=1}^{N_\epsilon(\mathcal{S}_{\mathcal{X}})} B(t_\ell, \epsilon),$$

where $N_\epsilon(\mathcal{S}_{\mathcal{X}})$ is the minimal number of balls in \mathcal{X} of radius ϵ ($\epsilon_n = (\log n)/n$) needed to cover $\mathcal{S}_{\mathcal{X}}$. Therefore, for all $t \in \mathcal{S}_{\mathcal{X}}$, there exists $\ell(t)$, $1 \leq \ell(t) \leq N_\epsilon(\mathcal{S}_{\mathcal{X}})$ such that

$$t \in B(t_\ell, \epsilon_n) \text{ and } d(t, t_{\ell(t)}) = \operatorname{argmin}_{1 \leq \ell \leq N_{\epsilon_n}(\mathcal{S}_{\mathcal{X}})} d(t, t_\ell).$$

Thus, the problem can be written as

$$\begin{aligned} & |Q_n(K, t; h_K) - \mathbb{E}(Q_n(K, t; h_K))| \\ & \leq |Q_n(K, t; h_K) - Q_n(K, t_{\ell(t)}; h_K)| + |Q_n(K, t_{\ell(t)}; h_K) - \mathbb{E}(Q_n(K, t_{\ell(t)}; h_K))| \\ & + |\mathbb{E}(Q_n(K, t_{\ell(t)}; h_K)) - \mathbb{E}(Q_n(K, t; h_K))| \end{aligned} \quad (3.9.14)$$

$$= |Q_{n,1}(K, t; h_K)| + |Q_{n,2}(K, t; h_K)| + |Q_{n,3}(K, t; h_K)|. \quad (3.9.15)$$

By (3.9.15), we can see that, in order to prove (3.9.7), we need to study

$$\sup_{K \in \mathcal{K}} \sup_{a_n \leq h_K \leq b_0} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \{|Q_{n,1}(K, t; h_K)| + |Q_{n,2}(K, t; h_K)| + |Q_{n,3}(K, t; h_K)|\}.$$

Let us start with $Q_{n,2}(K, t; h_K)$. We define

$$h_{K,j} = 2^j a_n, \quad L(n) = \max\{j : h_{K,j} \leq 2b_0\}.$$

Notice that, if we denote $\mathcal{H}_{K,j} := [h_{K,j-1}, h_{K,j}]$, we can clearly see that $[a_n, b_0] \subseteq \bigcup_{j=1}^{L(n)} \mathcal{H}_{K,j}$, which means in order to study the uniformity of $Q_{n,2}(K, t; h_K)$, we need to study

$$\max_{1 \leq j \leq L(n)} \sup_{K \in \mathcal{K}} \sup_{h_K \in \mathcal{H}_{K,j}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |Q_{n,2}(K, t; h_K)|.$$

Now, for $1 \leq j \leq L(n)$ and $1 \leq \ell \leq N_{\epsilon_n}(\mathcal{S}_{\mathcal{X}})$, we consider the following class of functions

$$\mathcal{G}_{K,j}^{(\ell)} = \left\{ x \mapsto K \left(\frac{d(x, t_\ell)}{h_K} \right) \text{ where } K \in \mathcal{K} \text{ and } h_K \in \mathcal{H}_{K,j} \right\}. \quad (3.9.16)$$

In addition, we denote

$$\alpha_n(g) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (g(X_i) - \mathbb{E}(g(X_i))),$$

the empirical process based on the functional variables X_1, \dots, X_n . Hence, using the fact that, for $1 \leq \ell(t) \leq N_\varepsilon(\mathcal{S}_{\mathcal{X}})$

$$Q_{n,2} = \frac{1}{\phi_{t_\ell}(h_K) \sqrt{n}} \alpha_n(K),$$

and, by the condition **(A.1.a)**, we have

$$\frac{\phi(a_n)}{\phi_{t_\ell}^2(h_K)} \leq \frac{\phi(a_n)}{C_1^2 \phi^2(h_K)},$$

and taking into account that $h_{K,j} = 2^j a_n$, which means that $h_{K,j}/2 = 2^{j-1} a_n =: h_{K,j-1}$, we get

$$\frac{\phi(a_n)}{C_1^2 \phi^2(h_K)} \leq \frac{\phi(a_n)}{C_1^2 \phi^2(h_{K,j-1})} = \frac{\phi(a_n)}{C_1^2 \phi^2(h_{K,j}/2)}.$$

In view of condition **(A.1.b)**, we have

$$C_3 \leq \frac{\phi(h_{K,j}/2)}{\phi(h_{K,j})} \leq \frac{\phi(h_{K,j}/2)}{\phi(a_n)},$$

this implies

$$\frac{\phi(a_n)}{\phi(h_{K,j}/2)} \leq \frac{1}{C_3}.$$

Hence, we conclude that

$$\begin{aligned} \frac{\phi(a_n)}{C_1^2 \phi^2(h_{K,j}/2)} &= \frac{1}{C_1^2 \phi(h_{K,j}/2)} \left(\frac{\phi(a_n)}{\phi(h_{K,j}/2)} \right) \\ &\leq \frac{1}{C_1^2 C_3 \phi(h_{K,j}/2)}. \end{aligned}$$

For some $\varrho > 0$, we have

$$\frac{\phi(a_n)}{\phi_{t_\ell}^2(h_K)} \leq \frac{\varrho^2}{\phi(h_{K,j}/2)}.$$

It follows that

$$\begin{aligned} &\mathbb{P} \left\{ \sup_{K \in \mathcal{K}} \sup_{a_n \leq h_K \leq b_0} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sqrt{\frac{n\phi(a_n)}{\Psi_{\mathcal{S}_{\mathcal{X}}}(\varepsilon_n)}} |Q_{n,2}(K, t_{\ell(t)}; h_K)| \geq \varepsilon_0 \right\} \\ &\leq \sum_{\ell=1}^{N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}})} \sum_{j=1}^{L(n)} \mathbb{P} \left\{ \frac{1}{\sqrt{n\phi(h_{K,j}/2)\Psi_{\mathcal{S}_{\mathcal{X}}}(\varepsilon_n)}} \max_{1 \leq \ell \leq N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}})} \max_{1 \leq j \leq L(n)} \|\sqrt{n}\alpha_n(K)\|_{\mathcal{G}_{K,j}^{(\ell)}} \geq \varepsilon_0/\varrho \right\} \\ &\leq L(n)N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}}) \max_{1 \leq j \leq L(n)} \max_{1 \leq \ell \leq N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}})} \mathbb{P} \left\{ \max_{1 \leq p \leq n} \|\sqrt{p}\alpha_p(K)\|_{\mathcal{G}_{K,j}^{(\ell)}} \geq \sqrt{n\phi(h_{K,j}/2)\Psi_{\mathcal{S}_{\mathcal{X}}}(\varepsilon_n)\varrho} \right\} \end{aligned}$$

We apply Bernstein's inequality in order to bound the probability in (3.9.17). To do that, we first study the asymptotic behavior of

$$\mu'_n = \mathbb{E} \|\sqrt{n}\alpha_n(K)\|_{\mathcal{G}_{K,j}^{(\ell)}},$$

and

$$\sigma^2 = \mathbb{E} \left(G_{K,j}^{(\ell)}(\mathbf{X}) \right)^2,$$

where $G_{K,j}^{(\ell)}$ denotes the envelope function of the class $\mathcal{G}_{K,j}^{(\ell)}$. Taking into consideration the condition (A.4.b), it follows that

$$G_{K,j}^{(\ell)}(\mathbf{X}) \leq \kappa_2 \mathbb{1}_{B(t_\ell, h_{K,j}/2)}(\mathbf{X}),$$

and, from condition (A.1.), we get that, for all $2 \leq q \leq p$,

$$\begin{aligned} \mathbb{E} \left(G_{K,j}^{(\ell)}(\mathbf{X}) \right)^q &\leq \mathbb{E} \left(\kappa_2 \mathbb{1}_{B(t_\ell, h_{K,j}/2)}(\mathbf{X}) \right)^q \\ &\leq \kappa_2^q \mathbb{P} \left(\mathbf{X} \in B(t_\ell, h_{K,j}/2) \right)^q \\ &\leq \kappa_2^q \phi_{t_\ell}(h_{K,j}/2)^q, \end{aligned}$$

and since the probabilities $0 \leq \phi_{t_\ell}(h_{K,j}/2) \leq 1$, then it is clear that $\phi_{t_\ell}(h_{K,j}/2)^q \leq \phi_{t_\ell}(h_{K,j}/2)$. Now using condition (A.1.a), there exists a constant $C_2 > 0$ such that $\phi_{t_\ell}(h_{K,j}/2) \leq C_2 \phi(h_{K,j}/2)$. we get that, for all $2 \leq q \leq p$,

$$\mathbb{E} \left(G_{K,j}^{(\ell)}(\mathbf{X}) \right)^q \leq C_2 \kappa_2^q \phi_{t_\ell}(h_{K,j}/2),$$

which implies that

$$\sigma^2 = C_2 \kappa_2^2 \phi(h_{K,j}/2).$$

Moreover, combining Lemma 3.9.1.6 and 3.9.1.7 in appendix together with condition (A.5.), we could observe that

$$\mu'_n = \mathbb{E} \left\| \sqrt{n} \alpha_n(\mathbf{K}) \right\|_{\mathcal{G}_{K,j}^{(\ell)}} \leq C \sqrt{n \phi(h_{K,j}/2)}.$$

Hence we have

$$\mu'_n = O \left(\sqrt{n \phi(h_{K,j}/2)} \right).$$

Now, we can apply Bernstein's inequality for the empirical processes, therefore for

$$z = (\epsilon_0/2) \sqrt{n \phi(h_{K,j}/2) \psi_{\mathcal{S}_{\mathcal{X}}}(\epsilon_n)} \quad \text{and} \quad H = \kappa_2 \left(\frac{2}{q!} \right)^{1/(q-2)}, \quad \text{or simply } H = \kappa_2 \text{ if } q = 2.$$

Making use of Lemma 3.9.1.8, we get

$$\begin{aligned} &\mathbb{P} \left\{ \max_{1 \leq p \leq n} \left\| \sqrt{p} \alpha_p(\mathbf{K}) \right\|_{\mathcal{G}_{K,j}^{(\ell)}} \geq \sqrt{n \phi(h_{K,j}/2) \psi_{\mathcal{S}_{\mathcal{X}}}(\epsilon_n)} \epsilon_0 / \varrho \right\} \\ &\leq \mathbb{P} \left\{ \max_{1 \leq p \leq n} \left\| \sqrt{p} \alpha_p(\mathbf{K}) \right\|_{\mathcal{G}_{K,j}^{(\ell)}} \geq \mu'_n + z \right\} \end{aligned} \quad (3.9.18)$$

$$\begin{aligned} &\leq \exp \left(\frac{-(\epsilon_0/\varrho)^2 \sqrt{n} \psi_{\mathcal{S}_{\mathcal{X}}}(\epsilon_n)}{8 \kappa_2^2 \sqrt{n} + 4 \epsilon_0 \kappa_2 \sqrt{\frac{\psi_{\mathcal{S}_{\mathcal{X}}}(\epsilon_n)}{\phi(h_{K,j}/2)}}} \right) \\ &\leq n^{-C \epsilon_0}, \end{aligned} \quad (3.9.19)$$

where $\mathbf{C} > 0$. Moreover, from equations (3.2.7), and the fact that $L(n) \sim \frac{\log(b_0/a_n)}{\log(2)}$ and by choosing

$$\epsilon_0^2 \mathbf{C} > 1, \epsilon_n = \log n/n,$$

we get from (3.9.19) and (3.9.17),

$$\begin{aligned} \sum_{n \geq 1} \mathbb{P} \left\{ \sup_{K \in \mathcal{K}} \sup_{a_n \leq h_K \leq b_0} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sqrt{\frac{n\phi(a_n)}{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}} |Q_{n,2}(K, t_{\ell(t)}; h_K)| \geq \epsilon_0 \right\} \\ \leq \sum_{n \geq 1} \text{CL}(n) N_{\epsilon_n}(\mathcal{S}_{\mathcal{X}}) n^{-\mathbf{C}\epsilon_0} < \infty. \end{aligned} \quad (3.9.20)$$

Let's us now study the first term, $|Q_{n,1}(K, t; h_K)|$ in (3.9.15). By condition (A.1.a), for $\mathbf{C} > 0$, we have

$$\begin{aligned} \sup_{K \in \mathcal{K}} \sup_{a_n \leq h_K \leq b_0} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |Q_{n,1}(K, t; h_K)| \\ \leq \frac{1}{n} \sup_{K \in \mathcal{K}} \sup_{a_n \leq h_K \leq b_0} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \left\{ \sum_{i=1}^n \left| \frac{1}{\phi_t(h_K)} \mathbb{K} \left(\frac{d(X_i, t)}{h_K} \right) - \frac{1}{\phi_{t_{\ell(t)}}(h_K)} \mathbb{K} \left(\frac{d(X_i, t_{\ell(t)})}{h_K} \right) \right| \right\} \quad (3.9.21) \\ \leq \sup_{K \in \mathcal{K}} \sup_{a_n \leq h \leq b_0} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \frac{\mathbf{C}}{n\phi(h_K)} \sum_{i=1}^n \left| \mathbb{K} \left(\frac{d(X_i, t)}{h_K} \right) - \mathbb{K} \left(\frac{d(X_i, t_{\ell(t)})}{h_K} \right) \right| \mathbb{1}_{B(t, h_K) \cup B(t_{\ell(t)}, h_K)}(X_i) \quad (3.9.22) \end{aligned}$$

We denote:

$$\sup_{K \in \mathcal{K}} \sup_{a_n \leq h_K \leq b_0} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \frac{1}{n\phi(h_K)} \sum_{i=1}^n \left| \mathbb{K} \left(\frac{d(X_i, t)}{h_K} \right) - \mathbb{K} \left(\frac{d(X_i, t_{\ell(t)})}{h_K} \right) \right| \mathbb{1}_{B(t, h_K) \cup B(t_{\ell(t)}, h_K)}(X_i) =: Q_{n,11}.$$

Since the kernel $\mathbb{K}(\cdot)$ is supposed to be Lipschitz, we have

$$\begin{aligned} Q_{n,11} &\leq \sup_{a_n \leq h \leq b_0} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \frac{1}{n} \sum_{i=1}^n \frac{1}{h_K \phi(h_K)} d(t, t_{\ell(t)}) \mathbb{1}_{B(t, h_K) \cup B(t_{\ell(t)}, h_K)}(X_i) \\ &\leq \sup_{a_n \leq h \leq b_0} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \frac{1}{n} \sum_{i=1}^n \frac{\epsilon_n}{h_K \phi(h_K)} \mathbb{1}_{B(t, h_K) \cup B(t_{\ell(t)}, h_K)}(X_i) \\ &\leq \sup_{a_n \leq h \leq b_0} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \frac{1}{n} \sum_{i=1}^n Z_i^{(11)}, \end{aligned}$$

where, for $1 \leq i \leq n$:

$$Z_i^{(11)} = \frac{\epsilon_n}{h_K \phi(h_K)} \mathbb{1}_{B(t, h_K) \cup B(t_{\ell(t)}, h_K)}(X_i).$$

Using condition (A.6.) with the fact that the bandwidths h_K belongs to the interval $[a_n, b_n]$, we get for $q \geq 2$:

$$\mathbb{E} \left(Z_1^{(11)} \right)^q \leq C_1^q \left(\frac{1}{\sqrt{\phi(a_n)}} \right)^{2(q-1)},$$

where we assume that C_1 is the bound of the sequence $\log n/na_n$ which converges to 0. Hence, by applying a standard inequality (see Corollary A.8 [Ferraty and Vieu \[2006\]](#)) that the conditions are satisfied here, uniformly on $t \in \mathcal{S}_{\mathcal{X}}$ and on h_K , on :

$$\frac{1}{n} \sum_{i=1}^n Z_i^{(11)},$$

in combination with condition (A.2.) and (A.6.), we readily obtain

$$Q_{n,11} = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_X} \left(\frac{\log n}{n} \right)}{n\phi(a_n)}} \right).$$

Therefore

$$\sup_{K \in \mathcal{K}} \sup_{a_n \leq h \leq b_0} \sup_{t \in \mathcal{S}_X} |Q_{n,1}(K, t; h_K)| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_X} \left(\frac{\log n}{n} \right)}{n\phi(a_n)}} \right). \quad (3.9.23)$$

Let's consider now $|Q_{n,3}(K, t; h_K)|$ in (3.9.15). It is clear that

$$|Q_{n,3}(K, t; h_K)| \leq \mathbb{E} |Q_n(K, t_{\ell(t)}; h_K) - Q_n(K, t; h_K)|,$$

hence, by the fact that the kernel $K(\cdot)$ is Lipschitz and for any constant a , $\mathbb{E}(a) = a$, then we deduce directly by proceeding in the same way as in treating $|Q_n^{(1)}(K, t; h_K)|$, that :

$$\sup_{K \in \mathcal{K}} \sup_{a_n \leq h \leq b_0} \sup_{t \in \mathcal{S}_X} |Q_{n,3}(K, t; h_K)| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_X} \left(\frac{\log n}{n} \right)}{n\phi(a_n)}} \right). \quad (3.9.24)$$

Consequently, from (3.9.20), (3.9.23) and (3.9.24), we get :

$$\sup_{K \in \mathcal{K}} \sup_{a_n \leq h_K \leq b_0} \sup_{t \in \mathcal{S}_X} |Q_n(K, t; h_K) - \mathbb{E}(Q_n(K, t; h_K))| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_X} \left(\frac{\log n}{n} \right)}{n\phi(a_n)}} \right).$$

Proof of Lemma 3.9.1.2 (b)

The proof of Lemma 3.9.1.2 (b) is a direct consequence of Lemma 3.9.1.2 (a). Specifically, taking assumptions (3.2.13), (3.2.14) and (3.2.15) into account, it suffices to consider

$$a_n = \mu\phi^{-1} \left(\frac{\rho_n k_{1,n}}{n} \right) \quad \text{and} \quad b_n = \nu\phi^{-1} \left(\frac{k_{2,n}}{\rho_n n} \right),$$

$$\mathbb{E}(Q_n(K, t, h_K)) \geq \kappa_1 \frac{\phi_t(h_K/2)}{\phi_t(h_K)} \geq \frac{\kappa_1 C_1}{C_2} \frac{\phi(h_K/2)}{\phi(h_K)} \geq \frac{\kappa_1 C_1 C_4}{C_2} = C' > 0. \quad (3.9.25)$$

Thus, denoting $C = C'/2$ verifies that

$$\begin{aligned}
 & \mathbb{P} \left(\inf_{K \in \mathcal{K}} \inf_{k_{1,n} \leq k \leq k_{1,n}} \inf_{t \in \mathcal{S}_{\mathcal{X}}} Q_n(K, t, h_K) < C \right) \\
 & \leq \mathbb{P} \left(\sup_{K \in \mathcal{K}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |Q_n(K, t, h_K) - \mathbb{E}(Q_n(K, t, h_K))| \geq C \right) \\
 & \leq \mathbb{P} \left(\sup_{K \in \mathcal{K}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |Q_n(K, t, h_K) - \mathbb{E}(Q_n(K, t, h_K))| \mathbb{1}_{\left\{ \mu\phi^{-1} \left(\frac{\rho_n k_{1,n}}{n} \right) \leq H_{n,k}(t) \leq \nu\phi^{-1} \left(\frac{k_{2,n}}{\rho_n n} \right) \right\}} \geq C/2 \right) \\
 & \quad + \mathbb{P} \left(H_{n,k}(t) \notin \left(\mu\phi^{-1} \left(\frac{\rho_n k_{1,n}}{n} \right), \nu\phi^{-1} \left(\frac{k_{2,n}}{\rho_n n} \right) \right) \right) \tag{3.9.26}
 \end{aligned}$$

It is shown, in [Kara *et al.* \[2017\]](#), see (3.9.42) and (3.9.43), that

$$\sum_n \sum_{k=k_{1,n}}^{k_{2,n}} \mathbb{P} \left(H_{n,k}(t) \leq \mu\phi^{-1} \left(\frac{\rho_n k_{1,n}}{n} \right) \right) < \infty \text{ and } \sum_n \sum_{k=k_{1,n}}^{k_{2,n}} \mathbb{P} \left(H_{n,k}(t) \geq \nu\phi^{-1} \left(\frac{k_{2,n}}{\rho_n n} \right) \right) < \infty.$$

So, all that is left to be investigated is

$$\sup_{K \in \mathcal{K}} \sup_{a_n \leq h_K \leq b_0} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |Q_n(K, t; h_K) - \mathbb{E}(Q_n(K, t; h_K))|.$$

That is considered in the previous lemma. Hence the proof is complete. \square

Proof of Lemma 3.9.1.3

Similar to the proof of Lemma 3.9.1.2, to prove (3.9.10), we can write

$$\begin{aligned}
 & |W_n(\varphi K, t; h_K) - \mathbb{E}(W_n(\varphi K, t; h_K))| \\
 & \leq |W_n(\varphi K, t; h_K) - W_n(\varphi K, t_{\ell(t)}; h_K)| \\
 & \quad + |W_n(\varphi K, t_{\ell(t)}; h_K) - \mathbb{E}(W_n(\varphi K, t_{\ell(t)}; h_K))| \\
 & \quad + |\mathbb{E}(W_n(\varphi K, t_{\ell(t)}; h_K)) - \mathbb{E}(W_n(\varphi K, t; h_K))|.
 \end{aligned}$$

In the sequel, we use the following notation

$$\begin{aligned}
 |W_{n,1}(\varphi K, t; h_K)| & := |W_n(\varphi K, t; h_K) - W_n(\varphi K, t_{\ell(t)}; h_K)|, \\
 |W_{n,2}(\varphi K, t_{\ell(t)}; h_K)| & := |W_n(\varphi K, t_{\ell(t)}; h_K) - \mathbb{E}(W_n(\varphi K, t_{\ell(t)}; h_K))|, \\
 |W_{n,3}(\varphi K, t; h_K)| & := |\mathbb{E}(W_n(\varphi K, t_{\ell(t)}; h_K)) - \mathbb{E}(W_n(\varphi K, t; h_K))|.
 \end{aligned}$$

Proceeding in the same way as in the proof of Lemma 3.9.1.2, we observe that for $1 \leq \ell \leq N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}})$:

$$\begin{aligned}
 W_{n,2}(\varphi K, t_{\ell}; h_K) & = \frac{1}{n\phi_{t_{\ell}}(h_K)} \sum_{i=1}^n \left\{ \varphi(Y_i) K \left(\frac{d(X_i, t_{\ell})}{h_K} \right) - \mathbb{E} \left(\varphi(Y_i) K \left(\frac{d(X_i, t_{\ell})}{h_K} \right) \right) \right\} \\
 & =: \frac{1}{\phi_{t_{\ell}}(h_K) \sqrt{n}} \alpha'_n(\varphi K),
 \end{aligned}$$

where $\alpha'_n(\varphi K)$ is an empirical process indexed by the class of functions $\mathcal{F}\mathcal{K}$. We consider also the following class of functions for $1 \leq j \leq L(n)$ and $1 \leq \ell \leq N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}})$:

$$\mathcal{G}_{\varphi K, j}^{(\ell)} = \left\{ (x, y) \mapsto \varphi(y) K \left(\frac{d(x, t_\ell)}{h_K} \right) \text{ where } K \in \mathcal{K}, \varphi \in \mathcal{F} \text{ and } h_K \in \mathcal{H}_{K, j} \right\},$$

with the envelope function $G_{\varphi K, j}^{(\ell)}$, and proceed as follows :

$$\begin{aligned} & \mathbb{P} \left\{ \max_{1 \leq j \leq L(n)} \sup_{\varphi K \in \mathcal{F}\mathcal{K}} \sup_{h_K \in \mathcal{H}_{K, j}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sqrt{\frac{n\phi(a_n)}{\Psi_{\mathcal{S}_{\mathcal{X}}}(\varepsilon_n)}} |W_{n,2}(\varphi K, t_{\ell(t)}; h_K)| \geq \varepsilon_0 \right\} \\ & \leq L(n) N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}}) \max_{1 \leq j \leq L(n)} \max_{1 \leq \ell \leq N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}})} \mathbb{P} \left\{ \max_{1 \leq p \leq n} \left\| \sqrt{p} \alpha'_p(\varphi K) \right\|_{\mathcal{G}_{\varphi K, j}^{(\ell)}} \geq \varepsilon_0 \sqrt{n\phi(h_{K, j}) \Psi_{\mathcal{S}_{\mathcal{X}}}(\varepsilon_n)} \right\}. \end{aligned}$$

Since the envelope function $G_{\varphi K, j}^{(\ell)}(\cdot, \cdot)$, satisfy

$$G_{\varphi K, j}^{(\ell)}(X, Y) \leq M\kappa_2 \mathbb{1}_{B(t_\ell, h_{K, j}/2)}(X),$$

and from (A.1.), (A.4.b) and (A.5.), with the use of the Lemma 3.9.1.6 in combination with Lemma 3.9.1.7 in appendix, we observe that

$$\mu_n'' := \mathbb{E} \left\| \sqrt{n} \alpha'_n(\varphi K) \right\|_{\mathcal{G}_{\varphi K, j}^{(\ell)}} = O \left(\sqrt{n\phi(h_{K, j}/2)} \right),$$

and

$$\sigma_1^2 := \mathbb{E} \left(G_{\varphi K, j}^{(\ell)}(X, Y) \right)^2 = O \left(\phi(h_{K, j}/2) \right).$$

By applying Bernstein's inequality for the empirical processes $\{\alpha'_n(\varphi K)\}_{\mathcal{F}\mathcal{K}}$ with the same choice of parameters as in the proof of Lemma 3.9.1.2, and from (3.2.6) we obtain :

$$\begin{aligned} \mathbb{P} \left\{ \max_{1 \leq p \leq n} \left\| \sqrt{p} \alpha'_p(\varphi K) \right\|_{\mathcal{G}_{\varphi K, j}^{(\ell)}} \geq \varepsilon_0 \sqrt{n\phi(h_{K, j}/2) \Psi_{\mathcal{S}_{\mathcal{X}}}(\varepsilon_n)} \right\} & \leq \exp \left(\frac{-\varepsilon_0^2 \Psi_{\mathcal{S}_{\mathcal{X}}}(\varepsilon_n)}{16\kappa_2^2} \right) \\ & \leq n^{-\varepsilon_0^2 C''}. \end{aligned} \quad (3.9.27)$$

Therefore, from condition (A.2.) and if we choose $\varepsilon_0^2 C'' > 1$ and $\varepsilon_n = \frac{\log n}{n}$, we get

$$\begin{aligned} & \sum_{n \geq 1} \mathbb{P} \left\{ \sup_{\varphi K \in \mathcal{F}\mathcal{K}} \sup_{a_n \leq h_K \leq b_0} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sqrt{\frac{n\phi(a_n)}{\Psi_{\mathcal{S}_{\mathcal{X}}}\left(\frac{\log n}{n}\right)}} |W_{n,2}(\varphi K, t_{\ell(t)}; h_K)| \geq \varepsilon_0 \right\} \\ & \leq \sum_{n \geq 1} CL(n) n^{-\varepsilon_0^2 C''} < \infty. \end{aligned} \quad (3.9.28)$$

Remark, from condition (A.4.a), that $W_{n,1}(\varphi K, t; h_K)$ satisfies :

$$\begin{aligned} & |W_{n,1}(\varphi K, t; h_K)| \\ & = \left| \frac{1}{n\phi_t(h_K)} \sum_{i=1}^n \varphi(Y_i) K \left(\frac{d(X_i, t)}{h_K} \right) - \frac{1}{n\phi_{t_{\ell(t)}}(h_K)} \sum_{i=1}^n \varphi(Y_i) K \left(\frac{d(X_i, t_{\ell(t)})}{h_K} \right) \right| \\ & \leq M \left| \frac{1}{n\phi_t(h_K)} \sum_{i=1}^n K \left(\frac{d(X_i, t)}{h_K} \right) - \frac{1}{n\phi_{t_{\ell(t)}}(h_K)} \sum_{i=1}^n K \left(\frac{d(X_i, t_{\ell(t)})}{h_K} \right) \right| \\ & = M |Q_n(K, t; h_K) - Q_n(K, t_{\ell(t)}; h_K)| \\ & = MQ_{n,1}(K, t; h_K). \end{aligned}$$

Hence, the study of the term $|W_{n,1}(\varphi\mathbf{K}, t; h_{\mathbf{K}})|$ is similar to that of $Q_{n,1}(\mathbf{K}, t; h_{\mathbf{K}})$. Therefore, we obtain

$$\sup_{\varphi\mathbf{K} \in \mathcal{F} \cdot \mathcal{K}} \sup_{a_n \leq h_{\mathbf{K}} \leq b_0} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |W_{n,1}(\varphi\mathbf{K}, t; h_{\mathbf{K}})| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}}\left(\frac{\log n}{n}\right)}{n\phi(a_n)}} \right). \quad (3.9.29)$$

Using similar arguments as those before, we remark that $|W_{n,3}(\varphi\mathbf{K}, t; h_{\mathbf{K}})|$ can be handled as $Q_{n,3}(\mathbf{K}, t; h_{\mathbf{K}})$. Hence

$$\sup_{\varphi\mathbf{K} \in \mathcal{F} \cdot \mathcal{K}} \sup_{a_n \leq h_{\mathbf{K}} \leq b_0} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |W_{n,3}(\varphi\mathbf{K}, t; h_{\mathbf{K}})| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}}\left(\frac{\log n}{n}\right)}{n\phi(a_n)}} \right). \quad (3.9.30)$$

Then (3.9.28), (3.9.29) and (3.9.30) yield :

$$\sup_{\varphi\mathbf{K} \in \mathcal{F} \cdot \mathcal{K}} \sup_{a_n \leq h_{\mathbf{K}} \leq b_0} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |W_n(\varphi\mathbf{K}, t; h_{\mathbf{K}}) - \mathbb{E}(W_n(\varphi\mathbf{K}, t; h_{\mathbf{K}}))| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}}\left(\frac{\log n}{n}\right)}{n\phi(a_n)}} \right).$$

Therefore the proof is complete. \square

Proof of Lemma 3.9.1.4

In this lemma, we suppose that the class of functions \mathcal{F} is not necessarily bounded but satisfying condition (A.4.a'). In order to prove Lemma 3.9.1.4, we proceed as follows, for an arbitrary $\lambda > 0$ and each function $\varphi \in \mathcal{F}$:

$$\begin{aligned} \varphi(y) &= \varphi(y) \mathbb{1}_{\{F(y) \leq \lambda \eta_n^{1/p}\}} + \varphi(y) \mathbb{1}_{\{F(y) > \lambda \eta_n^{1/p}\}} \\ &= \varphi^{(T)}(y) + \varphi^{(R)}(y), \end{aligned}$$

where we take $\eta_n = n / (\log n)^2$. So we write

$$\begin{aligned} W_n(\varphi\mathbf{K}, t; h_{\mathbf{K}}) &:= \frac{1}{n\phi_t(h_{\mathbf{K}})} \sum_{i=1}^n \varphi^{(T)}(Y_i) \mathbf{K}\left(\frac{d(X_i, t)}{h_{\mathbf{K}}}\right) \\ &\quad + \frac{1}{n\phi_t(h_{\mathbf{K}})} \sum_{i=1}^n \varphi^{(R)}(Y_i) \mathbf{K}\left(\frac{d(X_i, t)}{h_{\mathbf{K}}}\right) \\ &= W_{n,(T)}(\varphi\mathbf{K}, t; h_{\mathbf{K}}) + W_{n,(R)}(\varphi\mathbf{K}, t; h_{\mathbf{K}}). \end{aligned}$$

To prove (3.9.12) we will first deal with the truncated part $W_{n,(T)}(\varphi\mathbf{K}, t; h_{\mathbf{K}})$, and we have

$$\begin{aligned} &|W_{n,(T)}(\varphi\mathbf{K}, t; h_{\mathbf{K}}) - \mathbb{E}(W_{n,(T)}(\varphi\mathbf{K}, t; h_{\mathbf{K}}))| \\ &\leq |W_{n,(T)}(\varphi\mathbf{K}, t; h_{\mathbf{K}}) - W_{n,(T)}(\varphi\mathbf{K}, t_{\ell(t)}; h_{\mathbf{K}})| \\ &\quad + |W_{n,(T)}(\varphi\mathbf{K}, t_{\ell(t)}; h_{\mathbf{K}}) - \mathbb{E}(W_{n,(T)}(\varphi\mathbf{K}, t_{\ell(t)}; h_{\mathbf{K}}))| \\ &\quad + |\mathbb{E}(W_{n,(T)}(\varphi\mathbf{K}, t_{\ell(t)}; h_{\mathbf{K}})) - \mathbb{E}(W_{n,(T)}(\varphi\mathbf{K}, t; h_{\mathbf{K}}))|. \end{aligned}$$

In the sequel, we will use the following notation

$$\begin{aligned} \left| W_{n,(T)}^{(1)}(\varphi K, t; h_K) \right| &:= \left| W_{n,(T)}(\varphi K, t; h_K) - W_{n,(T)}(\varphi K, t_{\ell(t)}; h_K) \right|, \\ \left| W_{n,(T)}^{(2)}(\varphi K, t_{\ell(t)}; h_K) \right| &:= \left| W_{n,(T)}(\varphi K, t_{\ell(t)}; h_K) - \mathbb{E} \left(W_{n,(T)}(\varphi K, t_{\ell(t)}; h_K) \right) \right|, \\ \left| W_{n,(T)}^{(3)}(\varphi K, t; h_K) \right| &:= \left| \mathbb{E} \left(W_{n,(T)}(\varphi K, t_{\ell(t)}; h_K) \right) - \mathbb{E} \left(W_{n,(T)}(\varphi K, t; h_K) \right) \right|. \end{aligned}$$

In similar way of the proof of Lemma 3.9.1.3, we get that for $1 \leq \ell \leq N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}})$

$$\begin{aligned} &W_{n,(T)}^{(2)}(\varphi K, t_{\ell(t)}; h_K) \\ &= \frac{1}{n\phi_{t_{\ell}}(h_K)} \sum_{i=1}^n \left\{ \varphi^{(T)}(Y_i) K \left(\frac{d(X_i, t_{\ell})}{h_K} \right) - \mathbb{E} \left(\varphi^{(T)}(Y_i) K \left(\frac{d(X_i, t_{\ell})}{h_K} \right) \right) \right\} \\ &=: \frac{1}{\phi_{t_{\ell}}(h_K) \sqrt{n}} \alpha_n(\varphi^{(T)} K), \end{aligned}$$

where $\alpha_n(\varphi^{(T)} K)$ is an empirical process indexed by the class of functions $\mathcal{F} \mathcal{H} \mathcal{I}_{\lambda \eta_n}$, where we define the class of functions of $y \in \mathcal{Y}$, $\mathcal{I}_{\lambda \eta_n}$ by

$$\mathcal{I}_{\lambda \eta_n} := \left\{ \mathbb{1}_{\{F(y) \leq \lambda \eta_n^{1/p}\}} : \lambda > 0 \right\},$$

where p being the moment order given in (A.4.a'). We consider also the following class of functions, for $1 \leq j \leq L(n)$ and $1 \leq \ell \leq N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}})$,

$$\mathcal{G}_{\varphi K, j}^{(\ell, T)} = \left\{ (x, y) \mapsto \varphi(y) K \left(\frac{d(x, t_{\ell})}{h_K} \right) \mathbb{1}_{\{F(y) \leq \lambda \eta_n^{1/p}\}} : K \in \mathcal{K}, \varphi \in \mathcal{F} \text{ and } h_K \in \mathcal{H}_{K, j}, \lambda > 0 \right\},$$

that the envelope function is denoted $G_{\varphi K, j}^{(\ell, T)}$, then :

$$\begin{aligned} &\mathbb{P} \left\{ \max_{1 \leq j \leq L(n)} \sup_{\varphi K \in \mathcal{F} \mathcal{H} \mathcal{K}} \sup_{h_{K, j-1} \leq h_K \leq h_{K, j}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sqrt{\frac{n\phi(a_n)}{\Psi_{\mathcal{S}_{\mathcal{X}}}(\varepsilon_n)}} \left| W_{n,(T)}^{(2)}(\varphi K, t_{\ell(t)}; h_K) \right| \geq \varepsilon_0 \right\} \\ &\leq L(n) N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}}) \max_{1 \leq j \leq L(n)} \max_{1 \leq \ell \leq N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}})} \mathbb{P} \left\{ \max_{1 \leq p \leq n} \left\| \sqrt{p} \alpha_p(\varphi^{(T)} K) \right\|_{\mathcal{G}_{\varphi K, j}^{(\ell, T)}} \geq \varepsilon_0 \sqrt{n\phi(h_{K, j}) \Psi_{\mathcal{S}_{\mathcal{X}}}(\varepsilon_n)} \right\}. \end{aligned} \quad (3.9.31)$$

As it was done previously, we apply Bernstein's inequality to investigate (3.9.31), so we have to study the asymptotic behavior of the quantities

$$\mu_{n, T}^{(3)} = \mathbb{E} \left\| \sqrt{n} \alpha_n(\varphi^{(T)} K) \right\|_{\mathcal{G}_{\varphi K, j}^{(\ell, T)}},$$

and

$$\mathbb{E} \left(G_{\varphi K, j}^{(\ell, T)}(X, Y) \right)^2.$$

From condition (A.4.b), we see that

$$G_{\varphi K, j}^{(\ell, T)}(X, Y) \leq F(Y) \kappa_2 \mathbb{1}_{B(t_{\ell}, h_{K, j/2})}(X) \mathbb{1}_{\{F(y) \leq \lambda \eta_n^{1/p}\}},$$

so thanks to the condition **(A.4.a')** we get

$$\mathbb{E} \left(G_{\varphi K, j}^{(\ell, T)}(X, Y) \right)^2 \leq \lambda^2 \kappa_2^2 \phi(h_{K, j}/2) \eta_n^{2/p} = \sigma_{(T)}^2,$$

indeed

$$\begin{aligned} \mathbb{E} \left(G_{\varphi K, j}^{(\ell, T)}(X, Y) \right)^2 &\leq \mathbb{E} \left(F(Y) \kappa_2 \mathbb{1}_{B(t_\ell, h_{K, j}/2)}(X) \mathbb{1}_{\{F(Y) \leq \lambda \eta_n^{1/p}\}} \right)^2 \\ &\leq \kappa_2^2 \mathbb{E} \left(\lambda \eta_n^{1/p} \mathbb{1}_{B(t_\ell, h_{K, j}/2)}(X) \right)^2 \\ &\leq \kappa_2^2 \lambda^2 \eta_n^{2/p} \mathbb{P}(X \in B(t_\ell, h_{K, j}/2))^2 \\ &\leq \kappa_2^2 \lambda^2 \eta_n^{2/p} \phi_{t_\ell}(h_{K, j}/2) \\ &\leq \kappa_2^2 \lambda^2 \eta_n^{2/p} \phi(h_{K, j}/2). \end{aligned}$$

For all $2 \leq q \leq p$, in a similar way, we have

$$\mathbb{E} \left(G_{\varphi K, j}^{(\ell, T)}(X, Y) \right)^q \leq \kappa_2^q \lambda^q \phi(h_{K, j}/2) \eta_n^{q/p} = \kappa_2^{q-2} \lambda^{q-2} \sigma_{(T)}^2,$$

which gives that

$$\sigma_{(T)}^2 = \kappa_2^2 \lambda^2 \eta_n^{2/p} \phi(h_{K, j}/2).$$

Further, according to Lemma 3.9.1.6 in combination with Lemma 3.9.1.7 in appendix, and since condition **(A.5.)** is satisfied, we infer that

$$\mu_{n, T}^{(3)} = \mathbb{E} \left\| \sqrt{n} \alpha_n(\varphi^{(T)} K) \right\|_{\mathcal{G}_{\varphi K, j}^{(\ell, T)}} \leq C' \kappa_2 \sqrt{n \phi(h_{K, j}/2) \eta_n^{2/p}}.$$

Implying that,

$$\mu_{n, T}^{(3)} = O \left(\sqrt{n \phi(h_{K, j}/2) \eta_n^{2/p}} \right).$$

Now, we can apply Bernstein's inequality for empirical processes, therefore for:

$$z = \epsilon_0 \sqrt{n \phi(h_{K, j}/2) \psi_{\mathcal{S}_X}(\epsilon_n)},$$

in Lemma 3.9.1.8, we get, for $\mathfrak{C}' > 0$,

$$\begin{aligned} &\mathbb{P} \left\{ \max_{1 \leq k \leq n} \left\| \sqrt{k} \alpha_k(\varphi^{(T)} K) \right\|_{\mathcal{G}_{\varphi K, j}^{(\ell, T)}} \geq \epsilon_0 \sqrt{n \phi(h_{K, j}/2) \psi_{\mathcal{S}_X}(\epsilon_n)} \right\} \\ &\leq \mathbb{P} \left\{ \max_{1 \leq k \leq n} \left\| \sqrt{k} \alpha_k(\varphi^{(T)} K) \right\|_{\mathcal{G}_{\varphi K, j}^{(\ell, T)}} \geq \mu_{n, T}^{(3)} + z \right\} \\ &\leq \exp \left(-\epsilon_0^2 \frac{n \phi(h_{K, j}/2) \psi_{\mathcal{S}_X}(\epsilon_n)}{2n \phi(h_{K, j}/2) \eta_n^{2/p} + \mathfrak{C}' \sqrt{n \phi(h_{K, j}/2) \psi_{\mathcal{S}_X}(\epsilon_n)}} \right) \\ &\leq n^{-\epsilon_0^2 \mathfrak{C}'}, \end{aligned} \tag{3.9.32}$$

where $\mathfrak{C}' > 0$. From equations (3.2.7), and fact that $L(n) \sim \frac{\log(b_0/a_n)}{\log(2)}$ and by choosing

$$\epsilon_n = \frac{\log n}{n},$$

we get

$$\sum_{n \geq 1} \mathbb{P} \left\{ \max_{1 \leq j \leq L(n)} \sup_{\varphi \mathbf{K} \in \mathcal{F} \mathcal{K}} \sup_{h_{\mathbf{K}} \in \mathcal{H}_{\mathbf{K},j}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sqrt{\frac{n\phi(a_n)}{\Psi_{\mathcal{S}_{\mathcal{X}}}(\varepsilon_n)}} \left| W_{n,(T)}^{(2)}(\varphi \mathbf{K}, t_{\ell(t)}; h_{\mathbf{K}}) \right| \geq \varepsilon_0 \right\} \leq \sum_{n \geq 1} L(n) n \varepsilon_0^2 C' \quad (\mathfrak{A}633)$$

We have now to prove that the desired result is also satisfied for

$$\left| W_{n,(T)}^{(1)}(\varphi \mathbf{K}, t; h_{\mathbf{K}}) \right| = \left| W_{n,(T)}(\varphi \mathbf{K}, t; h_{\mathbf{K}}) - W_{n,(T)}(\varphi \mathbf{K}, t_{\ell(t)}; h_{\mathbf{K}}) \right|.$$

We proceed as in the bounded case, for $C' > 0$,

$$\begin{aligned} & \left| W_{n,(T)}^{(1)}(\varphi \mathbf{K}, t; h_{\mathbf{K}}) \right| \\ &= \left| \frac{1}{n\phi_t(h_{\mathbf{K}})} \sum_{i=1}^n \varphi^{(T)}(Y_i) \mathbf{K} \left(\frac{d(X_i, t)}{h_{\mathbf{K}}} \right) - \frac{1}{n\phi_{t_{\ell(t)}}(h_{\mathbf{K}})} \sum_{i=1}^n \varphi^{(T)}(Y_i) \mathbf{K} \left(\frac{d(X_i, t_{\ell(t)})}{h_{\mathbf{K}}} \right) \right| \\ &\leq \frac{C'}{n\phi(h_{\mathbf{K}})} \left| \sum_{i=1}^n F^{(T)}(Y_i) \mathbf{K} \left(\frac{d(X_i, t)}{h_{\mathbf{K}}} \right) - \sum_{i=1}^n F^{(T)}(Y_i) \mathbf{K} \left(\frac{d(X_i, t_{\ell(t)})}{h_{\mathbf{K}}} \right) \right|, \end{aligned}$$

where

$$F^{(T)}(y) := F(y) \mathbb{1}_{\{F(y) \leq \lambda \eta_n^{1/p}\}}.$$

Continuing the proof and remark that

$$\begin{aligned} & \sup_{\varphi^{(T)} \mathbf{K} \in \mathcal{F} \mathcal{K}} \sup_{\mathcal{S}_{\lambda \eta_n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sup_{a_n \leq h_{\mathbf{K}} \leq b_0} \left| W_{n,(T)}^{(1)}(\varphi \mathbf{K}, t; h_{\mathbf{K}}) \right| \\ &\leq \sup_{\varphi^{(T)} \mathbf{K} \in \mathcal{F} \mathcal{K}} \sup_{\mathcal{S}_{\lambda \eta_n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sup_{a_n \leq h_{\mathbf{K}} \leq b_0} \frac{1}{n\phi(h_{\mathbf{K}})} \sum_{i=1}^n |F^{(T)}(Y_i)| \left| \mathbf{K} \left(\frac{d(X_i, t)}{h_{\mathbf{K}}} \right) - \mathbf{K} \left(\frac{d(X_i, t_{\ell(t)})}{h_{\mathbf{K}}} \right) \right| \\ &\quad \mathbb{1}_{B(t, h_{\mathbf{K}}) \cup B(t_{\ell(t)}, h_{\mathbf{K}})}(X_i) \\ &=: \left| W_{n,(T)}^{(11)}(\varphi \mathbf{K}, t; h_{\mathbf{K}}) \right|. \end{aligned}$$

Let's go back to $Q_n^{(1)}(\mathbf{K}, t; h_{\mathbf{K}})$ and following the same steps until we get

$$\begin{aligned} \left| W_{n,(T)}^{(11)}(\varphi \mathbf{K}, t; h_{\mathbf{K}}) \right| &\leq \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sup_{a_n \leq h_{\mathbf{K}} \leq b_0} \frac{1}{n} \sum_{i=1}^n \frac{1}{h_{\mathbf{K}} \phi(h_{\mathbf{K}})} |F^{(T)}(Y_i)| d(t, t_{\ell(t)}) \mathbb{1}_{B(t, h_{\mathbf{K}}) \cup B(t_{\ell(t)}, h_{\mathbf{K}})}(X_i) \\ &\leq \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sup_{a_n \leq h_{\mathbf{K}} \leq b_0} \frac{1}{n} \sum_{i=1}^n \frac{\varepsilon_n}{h_{\mathbf{K}} \phi(h_{\mathbf{K}})} |F^{(T)}(Y_i)| \mathbb{1}_{B(t, h_{\mathbf{K}}) \cup B(t_{\ell(t)}, h_{\mathbf{K}})}(X_i) \\ &= \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sup_{a_n \leq h_{\mathbf{K}} \leq b_0} \frac{1}{n} \sum_{i=1}^n Z_{i,T}^{(11)}, \end{aligned}$$

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

$$Z_{i,T}^{(11)} = \frac{\varepsilon_n}{h_{\mathbf{K}} \phi(h_{\mathbf{K}})} |F^{(T)}(Y_i)| \mathbb{1}_{B(t, h_{\mathbf{K}}) \cup B(t_{\ell(t)}, h_{\mathbf{K}})}(X_i)$$

We get uniformly on $t \in \mathcal{S}_{\mathcal{X}}$ and on $h_{\mathbf{K}}$:

$$\sup_{t \in \mathcal{S}_{\mathcal{X}}} \sup_{a_n \leq h_{\mathbf{K}} \leq b_0} Z_{1,T}^{(11)} \leq \sup_{t \in \mathcal{S}_{\mathcal{X}}} \frac{\varepsilon_n}{a_n \phi(a_n)} |F^{(T)}(Y_1)|.$$

So, for $2 \leq q \leq p$, we obtain

$$\begin{aligned} \mathbb{E} \left(Z_{1,T}^{(11)} \right)^q &\leq \left(\frac{\varepsilon_n}{a_n \phi(a_n)} \right)^q \mathbb{E} \left(|F^{(T)}(Y_1)|^q \right) \\ &\leq \sup_{t \in \mathcal{S}_{\mathcal{X}}} \left(\frac{\varepsilon_n}{a_n \phi(a_n)} \right)^q \mathbb{E} \left(|F^{(T)}(Y_1)|^q | X = t \right) \end{aligned} \quad (3.9.34)$$

$$\leq \left(\frac{\varepsilon_n}{a_n \phi(a_n)} \right)^q \theta_p^{q/p} = \frac{\theta_p^{q/p} \varepsilon_n^q}{a_n^q \phi(a_n)} \left(\frac{1}{\phi(a_n)} \right)^{q-1}. \quad (3.9.35)$$

The transition from (3.9.34) to (3.9.35) is done by using Jensen's inequality used for the concave function z^a , for $0 < a \leq 1$. Moreover from condition **(A.6)** we deduce that the quantity $\varepsilon_n/a_n\phi(a_n)$ is bounded, then for $q \geq 2$

$$\sup_{t \in \mathcal{S}_{\mathcal{X}}} \sup_{a_n \leq h_K \leq b_0} \mathbb{E} \left(Z_{1,T}^{(11)} \right)^q \leq \mathfrak{C} \theta_p^{q/p} \left(\frac{1}{\sqrt{\phi(a_n)}} \right)^{2(q-1)},$$

where $\mathfrak{C} > 0$. Hence, by applying a standard inequality (see Corollary A.8 [Ferraty and Vieu \[2006\]](#)) that the conditions are satisfied here, uniformly on $t \in \mathcal{S}_{\mathcal{X}}$, we get

$$\left| W_{n,(T)}^{(11)}(\varphi K, t; h_K) \right| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}{n \phi(a_n)}} \right).$$

Then we obtain

$$\sup_{\varphi^{(T)} K \in \mathcal{F} \mathcal{H} \mathcal{I}_{\lambda \eta n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sup_{a_n \leq h_K \leq b_0} \left| W_{n,(T)}^{(1)}(\varphi K, t; h_K) \right| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}{n \phi(a_n)}} \right).$$

Since the first two terms have been treated, we now evaluate

$$\left| W_{n,(T)}^{(3)}(\varphi K, t; h_K) \right| = \left| \mathbb{E} \left(W_{n,(T)}(\varphi K, t_{\ell(t)}; h_K) \right) - \mathbb{E} \left(W_{n,(T)}(\varphi K, t; h_K) \right) \right|.$$

We have

$$\begin{aligned} \left| W_{n,(T)}^{(3)}(\varphi K, t; h_K) \right| &= \left| \mathbb{E} \left(W_{n,(T)}(\varphi K, t_{\ell(t)}; h_K) \right) - \mathbb{E} \left(W_{n,(T)}(\varphi K, t; h_K) \right) \right|, \\ &= \left| \mathbb{E} \left(W_{n,(T)}(\varphi K, t; h_K) - W_{n,(T)}(\varphi K, t_{\ell(t)}; h_K) \right) \right|, \\ &= \left| \mathbb{E} \left(W_{n,(T)}^{(1)}(\varphi K, t; h_K) \right) \right| \\ &\leq \frac{1}{n \phi(h_K)} \left| \sum_{i=1}^n \mathbb{E} \left(F^{(T)}(Y_i) K \left(\frac{d(X_i, t)}{h_K} \right) \right. \right. \\ &\quad \left. \left. - F^{(T)}(Y_i) K \left(\frac{d(X_i, t_{\ell(t)})}{h_K} \right) \right) \right|, \end{aligned}$$

which means that

$$\begin{aligned}
 & \sup_{\varphi^{(T)} \mathbf{K} \in \mathcal{F} \mathcal{H} \mathcal{I}_{\lambda \eta n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sup_{a_n \leq h_{\mathbf{K}} \leq b_0} \left| W_{n,(T)}^{(3)}(\varphi \mathbf{K}, t; h_{\mathbf{K}}) \right| \\
 & \leq \sup_{\varphi^{(T)} \mathbf{K} \in \mathcal{F} \mathcal{H} \mathcal{I}_{\lambda \eta n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sup_{a_n \leq h_{\mathbf{K}} \leq b_0} \frac{1}{n \phi(h_{\mathbf{K}})} \sum_{i=1}^n \mathbb{E} \left(\left| F^{(T)}(Y_i) \right| \right. \\
 & \quad \left. \left| \mathbf{K} \left(\frac{d(X_i, t)}{h_{\mathbf{K}}} \right) - \mathbf{K} \left(\frac{d(X_i, t_{\ell(t)})}{h_{\mathbf{K}}} \right) \right| \mathbb{1}_{B(t, h_{\mathbf{K}}) \cup B(t_{\ell(t)}, h_{\mathbf{K}})}(X_i) \right) \\
 & =: \left| W_{n,(T)}^{(31)}(\varphi \mathbf{K}, t; h_{\mathbf{K}}) \right|.
 \end{aligned}$$

Let's take a look at $\left| W_{n,(T)}^{(11)}(\varphi \mathbf{K}, t; h_{\mathbf{K}}) \right|$ and following the same steps, by the fact that $\mathbf{K}(\cdot)$ is Lipschitz, we get

$$\begin{aligned}
 \left| W_{n,(T)}^{(31)}(\varphi \mathbf{K}, t; h_{\mathbf{K}}) \right| & \leq \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sup_{a_n \leq h_{\mathbf{K}} \leq b_0} \frac{1}{n} \sum_{i=1}^n \frac{1}{h_{\mathbf{K}} \phi(h_{\mathbf{K}})} \mathbb{E} \left(\left| F^{(T)}(Y_i) \right| d(t, t_{\ell(t)}) \mathbb{1}_{B(t, h_{\mathbf{K}}) \cup B(t_{\ell(t)}, h_{\mathbf{K}})}(X_i) \right) \\
 & \leq \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sup_{a_n \leq h_{\mathbf{K}} \leq b_0} \frac{1}{n} \sum_{i=1}^n \frac{\varepsilon_n}{h_{\mathbf{K}} \phi(h_{\mathbf{K}})} \mathbb{E} \left(\left| F^{(T)}(Y_i) \right| \mathbb{1}_{B(t, h_{\mathbf{K}}) \cup B(t_{\ell(t)}, h_{\mathbf{K}})}(X_i) \right) \\
 & \leq \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sup_{a_n \leq h_{\mathbf{K}} \leq b_0} \frac{1}{n} \sum_{i=1}^n \frac{\varepsilon_n}{h_{\mathbf{K}} \phi(h_{\mathbf{K}})} \mathbb{E} \left(\left| F^{(T)}(Y_i) \right| \right) \\
 & \leq \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sup_{a_n \leq h_{\mathbf{K}} \leq b_0} \frac{1}{n} \sum_{i=1}^n \frac{\varepsilon_n}{h_{\mathbf{K}} \phi(h_{\mathbf{K}})} \mathbb{E} \left(\mathbb{E} \left(F^{(T)}(Y_i) \mid X = t \right) \right) \\
 & = \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sup_{a_n \leq h_{\mathbf{K}} \leq b_0} \frac{1}{n} \sum_{i=1}^n Z_{i,T}^{(31)}, \tag{3.9.36}
 \end{aligned}$$

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

$$Z_{i,T}^{(31)} = \frac{\varepsilon_n}{h_{\mathbf{K}} \phi(h_{\mathbf{K}})} \mathbb{E} \left(\left| F^{(T)}(Y_i) \right| \right),$$

we get for $q \geq 2$ and uniformly on $t \in \mathcal{S}_{\mathcal{X}}$ and on $h_{\mathbf{K}}$:

$$\sup_{t \in \mathcal{S}_{\mathcal{X}}} \sup_{a_n \leq h_{\mathbf{K}} \leq b_0} \mathbb{E} \left(Z_{1,T}^{(31)} \right)^q \leq \sup_{t \in \mathcal{S}_{\mathcal{X}}} \left(\frac{\varepsilon_n}{a_n \phi(a_n)} \right)^q \mathbb{E} \left(\mathbb{E} \left(\left| F^{(T)}(Y_1) \right|^q \mid X = t \right) \right) \tag{3.9.37}$$

$$\leq \left(\frac{\varepsilon_n}{a_n \phi(a_n)} \right)^q \theta_p^{q/p} = \frac{\theta_p^{q/p} \varepsilon_n^q}{a_n^q \phi(a_n)} \left(\frac{1}{\phi(a_n)} \right)^{q-1}. \tag{3.9.38}$$

The transition from (3.9.37) to (3.9.38) is done by using Minkowski's and Jensen's inequalities.

Then following the same way as in treating $\left| W_{n,(T)}^{(11)}(\varphi \mathbf{K}, t; h_{\mathbf{K}}) \right|$ we get

$$\sup_{\varphi^{(T)} \mathbf{K} \in \mathcal{F} \mathcal{H} \mathcal{I}_{\lambda \eta n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sup_{a_n \leq h_{\mathbf{K}} \leq b_0} \left| W_{n,(T)}^{(3)}(\varphi \mathbf{K}, t; h_{\mathbf{K}}) \right| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}{n \phi(a_n)}} \right).$$

Therefore the proof is done for the truncated part. Regarding the remainder term, the idea consists on proving its asymptotic negligibility, that is

$$\sup_{a_n \leq h_{\mathbf{K}} \leq b_n} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \sup_{\varphi \mathbf{K} \in \mathcal{F} \mathcal{H}} \left| W_{n,(R)}(\varphi \mathbf{K}, t; h_{\mathbf{K}}) - \mathbb{E} \left(W_{n,(R)}(\varphi \mathbf{K}, t; h(t)) \right) \right| = o_{a.co}(1),$$

which could be derived directly from the proof of the remainder part of the U-statistics developed in the sequel. \square

Proof of Theorem 3.3.2.1

To prove Theorem 3.3.2.1 we need to check the conditions of Lemma 3.9.1.1 in the case of $m = 1$. For that, we first identify the variables as follows: $S_\Omega = \mathcal{S}_X$, $A_i = X_i$, $\varphi(B_i) = \varphi(Y_i)$,

$$\begin{aligned} G(H, (t, A_i)) &= K(H^{-1}d(t, X_i)), \\ D_{n,k}(t) &= H_{n,k}(t), \\ \mathfrak{M}_n^{(1)}(\varphi, t; H_{n,k}(t)) &= \hat{r}_n^{*(1)}(\varphi, t; H_{n,k}(t)), \\ \mathfrak{M}(\varphi, t) &= r^{(1)}(\varphi, t). \end{aligned}$$

Choosing $D_{n,k}^-(\beta_n, t)$ and $D_{n,k}^+(\beta_n, t)$ such that

$$\begin{aligned} \phi_t(D_{n,k}^-(\beta_n, t)) &= \frac{\sqrt{\beta_n}k}{n}, \\ \phi_t(D_{n,k}^+(\beta_n, t)) &= \frac{k}{n\sqrt{\beta_n}}. \end{aligned}$$

We denote $h^-(t) = D_{n,k}^-(\beta_n, t)$, $h^+(t) = D_{n,k}^+(\beta_n, t)$ and

$$\mathcal{U}_n = \phi^{-1}\left(\frac{k_{2,n}}{\sqrt{\beta_n}n}\right)^Y + \sqrt{\frac{\Psi_{\mathcal{S}_X}\left(\frac{\log n}{n}\right)}{n\phi\left(\mu\phi^{-1}\left(\frac{\rho_n k_{1,n}}{n}\right)\right)}},$$

for all increasing sequence $\beta_n \in (0, 1)$ such that $\beta_n - 1 = O(\mathcal{U}_n)$. Note that for all $\beta_n \in (0, 1)$, $t \in \mathcal{S}_X$ and $k_{1,n} \leq k \leq k_{2,n}$ we have

$$\begin{aligned} \phi_t^{-1}\left(\frac{\sqrt{\beta_n}k_{1,n}}{n}\right) &\leq h^-(t) \leq \phi_t^{-1}\left(\frac{\sqrt{\beta_n}k_{2,n}}{n}\right), \\ \phi_t^{-1}\left(\frac{k_{1,n}}{n\sqrt{\beta_n}}\right) &\leq h^+(t) \leq \phi_t^{-1}\left(\frac{k_{2,n}}{n\sqrt{\beta_n}}\right), \end{aligned}$$

using the condition (3.2.12) we can easily deduce that the bandwidths $h^-(t)$ and $h^+(t)$ both belong to the interval

$$[a_n, b_n] = \left[\mu\phi^{-1}\left(\frac{\rho_n k_{1,n}}{n}\right), \nu\phi^{-1}\left(\frac{k_{2,n}}{\rho_n n}\right) \right].$$

Checking the conditions (L₄) and (L₅)

Let us start with checking (L₄). The fact that β_n is bounded by 1, and the local bandwidth $h^-(t)$ satisfies the conditions of Theorem 3.3.1.1 gives

$$\begin{aligned}
 & \sup_{\varphi \in \mathcal{F}} \sup_{\mathcal{K}} \sup_{a_n \leq h^-(t) \leq b_n} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \left| \mathfrak{M}_n^{(1)}(\varphi, t; D_{n,k}^-(\beta_n, t)) - \mathfrak{M}^{(1)}(\varphi, t) \right| \\
 &= \sup_{\varphi \in \mathcal{F}} \sup_{\mathcal{K}} \sup_{a_n \leq h^-(t) \leq b_n} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \left| \frac{\sum_{i=1}^n \varphi(Y_i) \mathbb{K}\left(\frac{d(t, X_i)}{h^-(t)}\right)}{\sum_{i=1}^n \mathbb{K}\left(\frac{d(t, X_i)}{h^-(t)}\right)} - r^{(1)}(\varphi, t) \right| \\
 &= \sup_{\varphi \in \mathcal{F}} \sup_{\mathcal{K}} \sup_{a_n \leq h^-(t) \leq b_n} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \left| \hat{r}_n^{(1)}(\varphi, t; h^-(t)) - r^{(1)}(\varphi, t) \right| \\
 &= O_{a.co} \left(b_n^\gamma + O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}}\left(\frac{\log n}{n}\right)}{n\phi(a_n)}} \right) \right),
 \end{aligned}$$

which is equivalent to

$$\begin{aligned}
 & \sup_{\varphi \in \mathcal{F}} \sup_{\mathcal{K}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \left| \hat{r}_n^{*(1)}(\varphi, t; h^-(t)) - r^{(1)}(\varphi, t) \right| \\
 &= O_{a.co} \left(\phi^{-1}\left(\frac{k_{2,n}}{\sqrt{\beta_n n}}\right)^\gamma + \sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}}\left(\frac{\log n}{n}\right)}{n\phi\left(\mu\phi^{-1}\left(\frac{\rho_n k_{1,n}}{n}\right)\right)}} \right) \\
 &= O_{a.co}(\mathcal{U}_n).
 \end{aligned}$$

We use the same reasoning to check (L₅) and we readily obtain

$$\sup_{\varphi \in \mathcal{F}} \sup_{\mathcal{K}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \left| \mathfrak{M}_n(\varphi, t; D_{n,k}^+(\beta_n, t)) - \mathfrak{M}(\varphi, t) \right| = O_{a.co}(\mathcal{U}_n).$$

Hence (L₄) and (L₅) are checked.

Checking the condition (L₂)

To check (L₂) we show that for all $t \in \mathcal{S}_{\mathcal{X}}$ and $\varepsilon_0 > 0$,

$$\sum_{n \geq 1} \mathbb{P} \left\{ \left| \mathbb{1}_{\left\{ D_{n,k}^-(\beta_n, t) \leq H_{n,k}(t) \leq D_{n,k}^+(\beta_n, t) \right\}} - 1 \right| > \varepsilon_0 \right\} < \infty. \quad (3.9.39)$$

Let $\varepsilon_0 > 0$ be fixed. Let $\{t_1, \dots, t_{N_{\varepsilon}(\mathcal{S}_{\mathcal{X}})}\}$ be an ε -net for $\mathcal{S}_{\mathcal{X}}$, for all $t \in \mathcal{S}_{\mathcal{X}}$, we have

$$\begin{aligned}
 & \mathbb{P} \left\{ \left| \mathbb{1}_{\{D_{n,k}^-(\beta_n, t) \leq H_{n,k}(t) \leq D_{n,k}^+(\beta_n, t)\}} - 1 \right| > \varepsilon_0 \right\} \\
 & \leq \mathbb{P} \left(H_{n,k}(t) \leq \Phi_t^{-1} \left(\frac{\sqrt{\beta_n} k_{1,n}}{n} \right) \right) + \mathbb{P} \left(H_{n,k}(t) \geq \Phi_t^{-1} \left(\frac{k_{2,n}}{\sqrt{\beta_n n}} \right) \right) \\
 & \leq \sum_{\ell=1}^{N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}})} \sum_{k=k_{1,n}}^{k_{2,n}} \mathbb{P} \left(H_{n,k}(t_\ell) \leq \Phi_t^{-1} \left(\frac{\sqrt{\beta_n} k_{1,n}}{n} \right) \right) \\
 & \quad + \sum_{\ell=1}^{N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}})} \sum_{k=k_{1,n}}^{k_{2,n}} \mathbb{P} \left(H_{n,k}(t_\ell) \geq \Phi_t^{-1} \left(\frac{k_{2,n}}{\sqrt{\beta_n n}} \right) \right) \\
 & \leq N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}}) \sum_{k=k_{1,n}}^{k_{2,n}} \mathbb{P} \left(H_{n,k}(t_\ell) \leq \Phi_t^{-1} \left(\frac{\sqrt{\beta_n} k_{1,n}}{n} \right) \right) \\
 & \quad + N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}}) \sum_{k=k_{1,n}}^{k_{2,n}} \mathbb{P} \left(H_{n,k}(t_\ell) \geq \Phi_t^{-1} \left(\frac{k_{2,n}}{\sqrt{\beta_n n}} \right) \right).
 \end{aligned}$$

Now we use a lemma similar to [Kara et al. \[2017\]](#) (see Lemma 3.9.1.11). For completeness, we give their proof. Making use of Lemma 3.9.1.11, we infer that

$$\begin{aligned}
 \mathbb{P} \left\{ H_{n,k}(t) \leq \Phi_t^{-1} \left(\frac{\alpha k_{1,n}}{n} \right) \right\} &= \mathbb{P} \left\{ \sum_{i=1}^n \mathbb{1}_{B(t, \Phi_t^{-1}(\alpha \frac{k_{1,n}}{n}))} (X_i) > k \right\} \\
 &= \mathbb{P} \left\{ \sum_{i=1}^n \mathbb{1}_{B(t, \Phi_t^{-1}(\alpha \frac{k_{1,n}}{n}))} (X_i) > \frac{k}{\alpha k_{1,n}} \alpha k_{1,n} \right\} \\
 &\leq \exp \{ -(k - \alpha k_{1,n}) / 4 \}.
 \end{aligned} \tag{3.9.40}$$

This implies that

$$\begin{aligned}
 N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}}) \sum_{k=k_{1,n}}^{k_{2,n}} \mathbb{P} \left\{ H_{n,k}(t) \leq \Phi_t^{-1} \left(\frac{\alpha k_{1,n}}{n} \right) \right\} &\leq N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}}) k_{2,n} \exp \{ -(1 - \alpha) k_{1,n} / 4 \} \\
 &\leq N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}}) n^{1 - \{(1 - \alpha) / 4\} \frac{k_{1,n}}{\ln n}}.
 \end{aligned}$$

In a similar way, we obtain

$$\mathbb{P} \left\{ H_{n,k}(t) \geq \Phi_t^{-1} \left(\frac{k_{2,n}}{\alpha n} \right) \right\} \leq \exp \left\{ -\frac{(k_{2,n} - \alpha k)^2}{2\alpha k_{2,n}} \right\}. \tag{3.9.41}$$

It follows that

$$\begin{aligned}
 N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}}) \sum_{k=k_{1,n}}^{k_{2,n}} \mathbb{P} \left\{ H_{n,k}(t) \geq \Phi_t^{-1} \left(\frac{k_{2,n}}{\alpha n} \right) \right\} &\leq N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}}) k_{2,n} \exp \{ -(1 - \alpha) k_{1,n} / 2\alpha \} \\
 &\leq N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}}) n^{1 - \{(1 - \alpha) / 2\alpha\} \frac{k_{2,n}}{\ln n}}.
 \end{aligned}$$

Therefore, by that fact that $k_{i,n}/\ln n \rightarrow \infty$, $i = 1, 2$, we obtain

$$N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}}) \sum_{k=k_{1,n}}^{k_{2,n}} \mathbb{P} \left\{ H_{n,k}(t) \leq \phi_t^{-1} \left(\frac{\alpha k_{1,n}}{n} \right) \right\} \leq N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}}) n^{1-\{(1-\alpha)/4\} \frac{k_{1,n}}{\ln n}} < \infty, \quad (3.9.42)$$

$$N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}}) \sum_{k=k_{1,n}}^{k_{2,n}} \mathbb{P} \left\{ H_{n,k}(t) \geq \phi_t^{-1} \left(\frac{k_{2,n}}{\alpha n} \right) \right\} \leq N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}}) n^{1-\{(1-\alpha)/2\alpha\} \frac{k_{2,n}}{\ln n}} < \infty. \quad (3.9.43)$$

Checking the condition (L₃)

We consider the following quantities:

$$\begin{aligned} f(\mathbb{K}, t, h(t)) &:= \mathbb{E} \left(\mathbb{K} \left(\frac{d(\mathbb{X}_1, t)}{h(t)} \right) \right), \\ Q_{n1} &:= \frac{f(\mathbb{K}, t, h^-(t))}{f(\mathbb{K}, t, h^+(t))}, \\ Q_{n2} &:= \frac{Q_n(\mathbb{K}, t, h^-(t))}{Q_n(\mathbb{K}, t, h^+(t))} - 1, \\ Q_{n3} &:= \frac{f(\mathbb{K}, t, h^+(t))}{f(\mathbb{K}, t, h^-(t))} \beta_n - 1. \end{aligned}$$

The condition (L₃) can be written as

$$\left| \frac{\sum_{i=1}^n \mathbb{K} \left(\frac{d(t, X_i)}{h^-(t)} \right)}{\sum_{i=1}^n \mathbb{K} \left(\frac{d(t, X_i)}{h^+(t)} \right)} - \beta_n \right| \leq |Q_{n1}| |Q_{n2}| + |Q_{n1}| |Q_{n3}|.$$

Hence, by the fact that $\beta_n \rightarrow 1$, our claimed result is

$$\sup_{\mathbb{K} \in \mathcal{K}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \left| \frac{\sum_{i=1}^n \mathbb{K} \left(\frac{d(t, X_i)}{h^-(t)} \right)}{\sum_{i=1}^n \mathbb{K} \left(\frac{d(t, X_i)}{h^+(t)} \right)} - \beta_n \right| = O_{a.co}(\mathcal{U}_n). \quad (3.9.44)$$

The proof of (3.9.44) is based on the following results

$$\sup_{\mathbb{K} \in \mathcal{K}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |Q_{n1}| \leq C, \quad (3.9.45)$$

$$\sup_{\mathbb{K} \in \mathcal{K}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |Q_{n2}| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}{n \phi \left(\mu \phi^{-1} \left(\frac{\rho_n k_{1,n}}{n} \right) \right)}} \right), \quad (3.9.46)$$

$$\sup_{\mathbb{K} \in \mathcal{K}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |Q_{n3}| = O \left(\phi^{-1} \left(\frac{k_{2,n}}{\rho_n n} \right)^Y \right). \quad (3.9.47)$$

Proof of (3.9.45)

Using the condition (A.4.b) one has

$$\mathbb{E} \left(\mathbb{K} \left(\frac{d(X_1, t)}{h(t)} \right) \right) \leq \kappa_2 \phi_t(h(t)/2).$$

Now using the condition (A.1.a) we directly obtain

$$\sup_{\mathbb{K} \in \mathcal{K}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |Q_{n1}| \leq C. \quad (3.9.48)$$

Proof of (3.9.17)

We have

$$\begin{aligned} \sup_{\mathbb{K} \in \mathcal{K}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |Q_{n2}| &= \sup_{\mathbb{K} \in \mathcal{K}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \left| \frac{Q_n(\mathbb{K}, t, h^-(t))}{Q_n(\mathbb{K}, t, h^+(t))} - 1 \right| \\ &\leq \frac{1}{\inf_{\mathbb{K} \in \mathcal{K}} \inf_{k_{1,n} \leq k \leq k_{2,n}} \inf_{t \in \mathcal{S}_{\mathcal{X}}} |Q_n(\mathbb{K}, t, h^+(t))|} \left(\sup_{\mathbb{K} \in \mathcal{K}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |Q_n(\mathbb{K}, t, h^-(t)) - 1| \right. \\ &\quad \left. + \sup_{\mathbb{K} \in \mathcal{K}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |Q_n(\mathbb{K}, t, h^+(t)) - 1| \right). \end{aligned} \quad (3.9.49)$$

To prove this, we use Lemma 3.9.1.2(b), which gives us

$$\sup_{\mathbb{K} \in \mathcal{K}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |Q_n(\mathbb{K}, t, h^-(t)) - 1| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}{n\phi \left(\mu\phi^{-1} \left(\frac{\rho_n k_{1,n}}{n} \right) \right)}} \right), \quad (3.9.50)$$

and

$$\sup_{\mathbb{K} \in \mathcal{K}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |Q_n(\mathbb{K}, t, h^+(t)) - 1| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}{n\phi \left(\mu\phi^{-1} \left(\frac{\rho_n k_{1,n}}{n} \right) \right)}} \right). \quad (3.9.51)$$

Moreover, combining (3.9.50), (3.9.51) with the fact that

$$\mathbb{P} \left(\inf_{\mathbb{K} \in \mathcal{K}} \inf_{k_{1,n} \leq k \leq k_{1,n}} \inf_{t \in \mathcal{S}_{\mathcal{X}}} Q_n(\mathbb{K}, t, h^+(t)) < C \right) < \infty, \quad (3.9.52)$$

it follows that

$$\sup_{\mathbb{K} \in \mathcal{K}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |Q_{n2}| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}{n\phi \left(\mu\phi^{-1} \left(\frac{\rho_n k_{1,n}}{n} \right) \right)}} \right). \quad (3.9.53)$$

Proof of (3.9.47)

Kudraszow and Vieu [2013] use in their proof of part the proof of Lemma 1 in Ezzahrioui and Ould-Saïd [2008]; on the other hand, we will use some computations similar to the steps of the proof of Lemma 5.7.0.3 in Bouzebda and Nemouchi [2020]. Let us consider the following quantity:

$$\begin{aligned}\hat{\mathbf{B}}(t, \varphi\mathbf{K}) &= \frac{\mathbb{E}[W_n(\varphi\mathbf{K}, t, h^+(t))]}{\mathbb{E}[Q_n(\mathbf{K}, t, h^-(t))]} - r^{(1)}(\varphi, t), \\ |\hat{\mathbf{B}}(t, \varphi\mathbf{K})| &= \left| \frac{\phi_t(h^-(t))}{\phi_t(h^+(t))} \frac{\mathbb{E} \left[\mathbf{K} \left(\frac{d(t, X_1)}{h^+(t)} \right) \mathbb{E}[\varphi(Y_1)|X_1] \right]}{\mathbb{E} \left[\mathbf{K} \left(\frac{d(t, X_1)}{h^-(t)} \right) \right]} - r^{(1)}(\varphi, t) \right| \\ &\leq \frac{1}{\mathbb{E} \left[\mathbf{K} \left(\frac{d(t, X_1)}{h^-(t)} \right) \right]} \left| \frac{\phi_t(h^-(t))}{\phi_t(h^+(t))} \mathbb{E} \left[\mathbf{K} \left(\frac{d(t, X_1)}{h^+(t)} \right) (\mathbb{E}[\varphi(Y)_1|X_1] - r^{(1)}(\varphi, t)) \right] \right|,\end{aligned}$$

using the fact that $\frac{\phi_t(h^-(t))}{\phi_t(h^+(t))} = \beta_n$ and supposing that the condition (A.3.) holds which means

$$|r^{(1)}(X, t) - r^{(1)}(\varphi, t)| \leq C_5 d^Y(X, t),$$

and assuming that the conditions (A.1.) and (A.4.b) to be satisfied, then for all $t \in \mathcal{S}_{\mathcal{X}}$, and $h^-(t)$ and $h^+(t)$ in $[a_n, b_n]$, one gets

$$\begin{aligned}\left| \frac{\mathbb{E}[W_n(\varphi\mathbf{K}, t, h^+(t))]}{\mathbb{E}[Q_n(\mathbf{K}, t, h^-(t))]} - r^{(1)}(\varphi, t) \right| &\leq \frac{C_5 \kappa_2 \beta_n}{\kappa_1 \phi_t \left(\frac{h^-(t)}{2} \right)} [\mathbb{E} \mathbb{1}_{B(t, h^+(t)/2)}(X) d^Y(X, t)] \\ &\leq \frac{C_5 \kappa_2 \beta_n}{\kappa_1} \frac{\phi_t \left(\frac{h^+(t)}{2} \right)}{\phi_t \left(\frac{h^-(t)}{2} \right)} (h^+(t))^Y \leq \frac{C_2 C_5 \kappa_2 \beta_n}{C_1 \kappa_1} (h^+(t))^Y \\ &\leq C (h^+(t))^Y.\end{aligned}$$

Keeping in mind the condition (A.4.b), we obtain

$$\sup_{\varphi\mathbf{K} \in \mathcal{F}\mathcal{K}} \sup_{(h^-(t), h^+(t)) \in [a_n, b_n]^2} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \left| \frac{\mathbb{E}[W_n(\varphi\mathbf{K}, t, h^+(t))]}{\mathbb{E}[Q_n(\mathbf{K}, t, h^-(t))]} - r^{(1)}(\varphi, t) \right| \leq C' b_n^Y. \quad (3.9.54)$$

Finally, rewriting (3.9.54) with $\varphi \equiv 1$ gives us

$$\sup_{\mathbf{K} \in \mathcal{K}} \sup_{(h^-(t), h^+(t)) \in [a_n, b_n]^2} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \left| \frac{\mathbb{E} \left[\mathbf{K} \left(\frac{d(t, X_1)}{h^+(t)} \right) \right]}{\mathbb{E} \left[\mathbf{K} \left(\frac{d(t, X_1)}{h^-(t)} \right) \right]} \beta_n - 1 \right| \leq C' b_n^Y,$$

which is equivalent to

$$\sup_{\mathbf{K} \in \mathcal{K}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |Q_{n3}| = O \left(\phi^{-1} \left(\frac{k_{2,n}}{\rho_n n} \right)^Y \right). \quad (3.9.55)$$

Combining the results of (3.9.48), (3.9.53) and (3.9.55) and the fact that $\beta_n \rightarrow 1$, implies that

$$\sup_{K \in \mathcal{K}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} \left| \frac{\sum_{i=1}^n K\left(\frac{d(t, X_i)}{h^-(t)}\right)}{\sum_{i=1}^n K\left(\frac{d(t, X_i)}{h^+(t)}\right)} - \beta_n \right| = O_{a.co}(\mathcal{U}_n).$$

Hence (L₃) is checked. Note that (L₀) is obviously satisfied by (A.4.b), and that (L₁) is also trivially satisfied by construction of $D_n^-(\beta_n, t)$ and $D_n^+(\beta_n, t)$. So one can apply Lemma 3.9.1.1, and (3.9.1) with $m = 1$ is exactly the result of Theorem 3.3.2.1. \square

Proof of corollary 3.3.1.2

To prove corollary 3.3.1.2, we use the same reasoning as the proof of Theorem 3.3.2.1 but instead of taking advantage of the results of Theorem 3.3.1.1 to check the conditions of Lemma 3.9.1.1, we use the results of Corollary 3.3.1.2 and then write the corresponding rates of convergence to obtain

$$\sup_{\varphi \in \mathcal{F}} \sup_{K \in \mathcal{K}} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{t \in \mathcal{S}_{\mathcal{X}}} |\hat{r}_n^{*(1)}(\varphi, t; h_{n,k}(t)) - r^{(1)}(\varphi, t)| = O_{a.co}(\mathcal{U}_n).$$

\square

Preliminaries of the proofs

This part is mainly dedicated to the study of the functional conditional U -statistics. Just like in the case of $m = 1$ where $\mathcal{S}_{\mathcal{X}}$ is covered by

$$\bigcup_{\ell=1}^{N_\varepsilon(\mathcal{S}_{\mathcal{X}})} B(t_\ell, \varepsilon),$$

for some radius ε . Hence, for each $\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m$, there exists $\boldsymbol{\ell}(\mathbf{t}) = (\ell(t_1), \dots, \ell(t_m))$ where $\forall 1 \leq i \leq m, 1 \leq \ell(t_i) \leq N_\varepsilon(\mathcal{S}_{\mathcal{X}})$ such that

$$\mathbf{t} \in \prod_{i=1}^m B(t_{\ell(t_i)}, \varepsilon) \text{ and } d(t_i, t_{\ell(t_i)}) = \operatorname{argmin}_{1 \leq \ell(t_i) \leq N_\varepsilon(\mathcal{S}_{\mathcal{X}})} d(t_i, t_\ell).$$

So for each $\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m$, the closest center is $\mathbf{t}_{\boldsymbol{\ell}(\mathbf{t})}$ and the ball with the closest center

$$\prod_{i=1}^m B(t_{\ell(t_i)}, \varepsilon) := B(\mathbf{t}_{\boldsymbol{\ell}(\mathbf{t})}, \varepsilon).$$

The proofs of the UIB consistency for the multivariate bandwidth will follow the same lines as the proofs of the UIB consistency for the univariate smoothing parameter in (Bouzebda and Nemouchi [2020]) while making the necessary changes to adapt to the k -NN procedures.

Proof of Theorem 3.4.1.7

In this section, we consider a bandwidths $\mathbf{h} = (h_1, \dots, h_m) \in \mathcal{H}_n^{(m)}$. In order to prove Theorem 3.4.1.7 we can write the U-statistic for each $\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m$ as follows

$$\begin{aligned}
 & |u_n(\varphi, \mathbf{t}; \mathbf{h}) - \mathbb{E}[u_n(\varphi, \mathbf{t}; \mathbf{h})]| \\
 & \leq |u_n(\varphi, \mathbf{t}; \mathbf{h}) - u_n(\varphi, \mathbf{t}_{\ell(\mathbf{t})}; \mathbf{h})| \\
 & \quad + |\mathbb{E}[u_n(\varphi, \mathbf{t}_{\ell(\mathbf{t})}; \mathbf{h})] - \mathbb{E}[u_n(\varphi, \mathbf{t}; \mathbf{h})]| \\
 & \quad + |u_n(\varphi, \mathbf{t}_{\ell(\mathbf{t})}; \mathbf{h}) - \mathbb{E}[u_n(\varphi, \mathbf{t}_{\ell(\mathbf{t})}; \mathbf{h})]| \\
 & \leq |u_{n,11}(\varphi, \mathbf{t}; \mathbf{h})| + |u_{n,12}(\varphi, \mathbf{t}; \mathbf{h})| + |u_{n,13}(\varphi, \mathbf{t}; \mathbf{h})|.
 \end{aligned}$$

Note that

$$\begin{aligned}
 & |u_{n,11}(\varphi, \mathbf{t}; \mathbf{h})| \\
 & = \frac{(n-m)!}{n!} \left| \sum_{i \in I(m,n)} \varphi(Y_{i_1}, \dots, Y_{i_m}) \left\{ \frac{1}{\phi_{\mathbf{t}}(\mathbf{h})} \prod_{j=1}^m \mathbb{K}\left(\frac{d(X_{ij}, t_j)}{h_j}\right) - \frac{1}{\phi_{\mathbf{t}_{\ell(\mathbf{t})}}(\mathbf{h})} \prod_{j=1}^m \mathbb{K}\left(\frac{d(X_{ij}, t_{\ell(t_j)})}{h_j}\right) \right\} \right| \\
 & \leq \frac{\mathbb{C}(n-m)!}{n! \tilde{\phi}(\mathbf{h})} \sum_{i \in I(m,n)} \left| \varphi(Y_{i_1}, \dots, Y_{i_m}) \left\{ \prod_{j=1}^m \mathbb{K}\left(\frac{d(X_{ij}, t_j)}{h_j}\right) - \prod_{j=1}^m \mathbb{K}\left(\frac{d(X_{ij}, t_{\ell(t_j)})}{h_j}\right) \right\} \right|.
 \end{aligned}$$

By applying the Telescoping binomial, we get

$$\prod_{j=1}^m \mathbb{K}\left(\frac{d(X_{ij}, t_j)}{h_j}\right) - \prod_{j=1}^m \mathbb{K}\left(\frac{d(X_{ij}, t_{\ell(t_j)})}{h_j}\right) \tag{3.9.56}$$

$$\begin{aligned}
 & = \sum_{j=1}^m \left[\left\{ \mathbb{K}\left(\frac{d(X_{ij}, t_j)}{h_j}\right) - \mathbb{K}\left(\frac{d(X_{ij}, t_{\ell(t_j)})}{h_j}\right) \right\} \right] \\
 & \quad \times \prod_{q=1}^{j-1} \mathbb{K}\left(\frac{d(X_{iq}, t_q)}{h_q}\right) \prod_{p=j+1}^m \mathbb{K}\left(\frac{d(X_{ip}, t_{\ell(t_p)})}{h_p}\right). \tag{3.9.57}
 \end{aligned}$$

From condition (A.4.b) we could claim that

$$\prod_{q=1}^{j-1} \mathbb{K}\left(\frac{d(X_{iq}, t_q)}{h_q}\right) \leq \kappa_2^{j-1} \prod_{q=1}^{j-1} \mathbb{1}_{B(t_q, h_q/2)}(X_{iq}),$$

similarly, we have

$$\prod_{p=j+1}^m \mathbb{K}\left(\frac{d(X_{ip}, t_{\ell(t_p)})}{h_p}\right) \leq \kappa_2^{m-j} \prod_{p=j+1}^m \mathbb{1}_{B(t_{\ell(t_p)}, h_p/2)}(X_{ip}).$$

So (3.9.56) satisfies :

$$\begin{aligned}
 & \prod_{j=1}^m \mathbb{K} \left(\frac{d(X_{ij}, t_j)}{h_j} \right) - \prod_{j=1}^m \mathbb{K} \left(\frac{d(X_{ij}, t_{\ell(t_j)})}{h_j} \right) \\
 & \leq \kappa_2^{m-1} \sum_{j=1}^m \left[\left\{ \mathbb{K} \left(\frac{d(X_{ij}, t_j)}{h_j} \right) - \mathbb{K} \left(\frac{d(X_{ij}, t_{\ell(t_j)})}{h_j} \right) \right\} \right. \\
 & \quad \left. \prod_{q=1}^{j-1} \mathbb{1}_{B(t_q, h_q/2)}(X_{iq}) \prod_{p=j+1}^m \mathbb{1}_{B(t_{\ell(t_p)}, h_p/2)}(X_{ip}) \right] \\
 & \leq \kappa_2^{m-1} \sum_{j=1}^m \left[\left\{ \mathbb{K} \left(\frac{d(X_{ij}, t_j)}{h_j} \right) - \mathbb{K} \left(\frac{d(X_{ij}, t_{\ell(t_j)})}{h_j} \right) \right\} \mathbb{1}_{B(t_j, h_j) \cup B(t_{\ell(t_j)}, h_j)}(X_{ij}) \right. \\
 & \quad \left. \prod_{q=1}^{j-1} \mathbb{1}_{B(t_q, h_q/2)}(X_{iq}) \prod_{p=j+1}^m \mathbb{1}_{B(t_{\ell(t_p)}, h_p/2)}(X_{ip}) \right] \\
 & =: \sum_{j=1}^m \mathbf{K}_{i_j, h_j}(t_j, t_{\ell(t_j)}),
 \end{aligned}$$

where

$$\begin{aligned}
 & \mathbf{K}_{i_j, h_j}^{(\ell)}(t_j, t_{\ell(t_j)}) \\
 & = \kappa_2^{m-1} \left\{ \mathbb{K} \left(\frac{d(X_{ij}, t_j)}{h_j} \right) - \mathbb{K} \left(\frac{d(X_{ij}, t_{\ell(t_j)})}{h_j} \right) \right\} \mathbb{1}_{B(t_j, h_j) \cup B(t_{\ell(t_j)}, h_j)}(X_{ij}) \\
 & \quad \times \prod_{q=1}^{j-1} \mathbb{1}_{B(t_q, h_q/2)}(X_{iq}) \prod_{p=j+1}^m \mathbb{1}_{B(t_{\ell(t_p)}, h_p/2)}(X_{ip}).
 \end{aligned}$$

Therefore

$$\begin{aligned}
 |u_{n,11}(\varphi, \mathbf{t}; \mathbf{h})| & \leq \frac{\mathbb{C}(n-m)!}{n! \tilde{\phi}(\mathbf{h})} \kappa_2^{m-1} \sum_{i \in I(m, n)} \left| \varphi(Y_{i_1}, \dots, Y_{i_m}) \sum_{j=1}^m \mathbf{K}_{i_j, h_j}^{(\ell)}(t_j, t_{\ell(t_j)}) \right| \\
 & \leq \frac{\mathbb{C}(n-m)!}{n! \tilde{\phi}(\mathbf{h})} M \kappa_2^{m-1} \sum_{i \in I(m, n)} \sum_{j=1}^m \left| \left\{ \mathbb{K} \left(\frac{d(X_{ij}, t_j)}{h_j} \right) - \mathbb{K} \left(\frac{d(X_{ij}, t_{\ell(t_j)})}{h_j} \right) \right\} \right. \\
 & \quad \left. \times \mathbb{1}_{B(t_j, h_j) \cup B(t_{\ell(t_j)}, h_j)}(X_{ij}) \prod_{q=1}^{j-1} \mathbb{1}_{B(t_q, h_q/2)}(X_{iq}) \prod_{p=j+1}^m \mathbb{1}_{B(t_{\ell(t_p)}, h_p/2)}(X_{ip}) \right| \quad (3.9.58)
 \end{aligned}$$

$$\begin{aligned}
 & \leq \frac{(n-m)!}{n!} \mathbb{C} m M \kappa_2^{m-1} \sum_{i \in I(m, n)} \frac{1}{m} \sum_{j=1}^m \left[\frac{\varepsilon_n}{\tilde{\phi}(\mathbf{h}) a_{n,j}} \mathbb{1}_{B(t_j, h_j) \cup B(t_{\ell(t_j)}, h_j)}(X_{ij}) \right. \\
 & \quad \left. \times \prod_{q=1}^{j-1} \mathbb{1}_{B(t_q, h_q/2)}(X_{iq}) \prod_{p=j+1}^m \mathbb{1}_{B(t_{\ell(t_p)}, h_p/2)}(X_{ip}) \right]. \quad (3.9.59)
 \end{aligned}$$

The transition from equation (3.9.58) to (3.9.59) is done thanks to the fact that the kernel function $K(\cdot)$ is Lipschitz. Uniformly on $\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m$ and $\mathbf{h} \in \mathcal{H}_n^{(m)}$, we get

$$\begin{aligned}
 & \sup_{\varphi \in \mathcal{F}_m} \sup_{\mathcal{K}^m} \sup_{\mathbf{h} \in \mathcal{H}_n^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} |u_{n,11}(\varphi, \mathbf{t}; \mathbf{h})| \\
 & \leq \sup_{\varphi \in \mathcal{F}_m} \sup_{\mathcal{K}^m} \sup_{\mathbf{h} \in \mathcal{H}_n^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \frac{(n-m)!}{n!} \sum_{i \in I(m,n)} \frac{1}{m} \sum_{j=1}^m \left[\frac{\mathbb{C}m\mathbb{M}\kappa_2^{m-1}\varepsilon_n}{\tilde{\Phi}(\mathbf{A}_n)a_{n,j}} \mathbb{1}_{B(t_j, h_j) \cup B(t_{\ell(t_j), h_j})}(\mathbf{X}_{ij}) \right. \\
 & \quad \left. \times \prod_{q=1}^{j-1} \mathbb{1}_{B(t_q, h_q/2)}(\mathbf{X}_{iq}) \prod_{p=j+1}^m \mathbb{1}_{B(t_{\ell(t_p)}, h_p/2)}(\mathbf{X}_{ip}) \right] \\
 & \leq \sup_{\varphi \in \mathcal{F}_m} \sup_{\mathcal{K}^m} \sup_{\mathbf{h} \in \mathcal{H}_n^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \frac{(n-m)!}{n!} \sum_{i \in I(m,n)} \frac{1}{m} \sum_{j=1}^m \left[\frac{\mathbb{C}m\mathbb{M}\kappa_2^{m-1}\varepsilon_n}{\tilde{\Phi}(\tilde{\mathbf{A}}_n)\tilde{a}_n} \mathbb{1}_{B(t_j, h_j) \cup B(t_{\ell(t_j), h_j})}(\mathbf{X}_{ij}) \right. \\
 & \quad \left. \times \prod_{q=1}^{j-1} \mathbb{1}_{B(t_q, h_q/2)}(\mathbf{X}_{iq}) \prod_{p=j+1}^m \mathbb{1}_{B(t_{\ell(t_p)}, h_p/2)}(\mathbf{X}_{ip}) \right] \\
 & \leq \sup_{\varphi \in \mathcal{F}_m} \sup_{\mathcal{K}^m} \sup_{\mathbf{h} \in \mathcal{H}_n^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \frac{(n-m)!}{n!} \sum_{i \in I(m,n)} \frac{1}{m} \sum_{j=1}^m \left[\frac{\mathbb{C}m\mathbb{M}\kappa_2^{m-1}\varepsilon_n}{C'_1\phi(\tilde{a}_n)\tilde{a}_n} \mathbb{1}_{B(t_j, h_j) \cup B(t_{\ell(t_j), h_j})}(\mathbf{X}_{ij}) \right. \\
 & \quad \left. \times \prod_{q=1}^{j-1} \mathbb{1}_{B(t_q, h_q/2)}(\mathbf{X}_{iq}) \prod_{p=j+1}^m \mathbb{1}_{B(t_{\ell(t_p)}, h_p/2)}(\mathbf{X}_{ip}) \right],
 \end{aligned}$$

by (3.2.16), where $\mathbf{A}_n = (a_{n,1}, \dots, a_{n,m})$ and $\tilde{\mathbf{A}}_n = (\tilde{a}_n, \dots, \tilde{a}_n)$, with $\tilde{\mathbf{A}}_n \leq \mathbf{A}_n$ component by component. The idea is to apply a Bernstein's inequality type to get the desired result by applying Lemma 3.9.1.9 with

$$\begin{aligned}
 \theta &= O\left(\frac{\varepsilon_n}{\tilde{a}_n\phi(\tilde{a}_n)}\right), \\
 \sigma^2 &= O\left(\frac{\varepsilon_n^2}{\tilde{a}_n^2\phi(\tilde{a}_n)}\right), \\
 s &= \sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}}(\varepsilon_n)}{n\phi(\tilde{a}_n)}} - \frac{\varepsilon_n}{\tilde{a}_n\phi(\tilde{a}_n)},
 \end{aligned}$$

on the function

$$f_{\mathbf{t}, \mathbf{h}}(\mathbf{X}_i) = \frac{1}{m} \sum_{j=1}^m \left[\frac{\mathbb{C}m\mathbb{M}\kappa_2^{m-1}\varepsilon_n}{C'_1\phi(\tilde{a}_n)a_{n,j}} \mathbb{1}_{B(t_j, h_j) \cup B(t_{\ell(t_j), h_j})}(\mathbf{X}_{ij}) \prod_{q=1}^{j-1} \mathbb{1}_{B(t_q, h_q/2)}(\mathbf{X}_{iq}) \prod_{p=j+1}^m \mathbb{1}_{B(t_{\ell(t_p)}, h_p/2)}(\mathbf{X}_{ip}) \right],$$

which satisfies for all $\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m$:

$$0 \leq \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \sup_{\mathbf{h} \in \mathcal{H}_n^{(m)}} |f_{\mathbf{t}, \mathbf{h}}(\mathbf{X})| \leq \frac{\mathbb{C}m\mathbb{M}\kappa_2^{m-1}\varepsilon_n}{C'_1\phi(\tilde{a}_n)\tilde{a}_n} \leq Cte = C_7.$$

Notice that the existence of the constant C_7 on the last right side of the preceding inequality is deduced from the condition (3.2.11), which implies that :

$$\mathbb{P} \left\{ \sup_{\varphi \in \mathcal{F}_m} \sup_{\mathcal{K}^m} \sup_{\mathbf{h} \in \mathcal{H}_n^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} |u_{n,11}(\varphi, \mathbf{t}; \mathbf{h})| \geq \sqrt{\frac{\Psi_{\mathcal{S}_x}(\varepsilon_n)}{n\phi(\tilde{a}_n)}} \right\} \quad (3.9.60)$$

$$\begin{aligned} &\leq \mathbb{P} \left\{ |u_n^{(m)}(f) - \theta| \geq \sqrt{\frac{\Psi_{\mathcal{S}_x}(\varepsilon_n)}{n\phi(\tilde{a}_n)} - \frac{\varepsilon_n}{\tilde{a}_n}} \right\} \\ &\leq \exp \left\{ -\frac{[n/m]s^2}{2\sigma^2 - \frac{2}{3}C_7s} \right\}. \end{aligned} \quad (3.9.61)$$

By developing the computation while respecting the imposed conditions, mainly (A.2.) and (A.6.), we get

$$\sum_{n \geq 1} \mathbb{P} \left\{ \sup_{\varphi \in \mathcal{F}_m} \sup_{\mathcal{K}^m} \sup_{\mathbf{h} \in \mathcal{H}_n^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} |u_{n,11}(\varphi, \mathbf{t}; \mathbf{h})| \geq \sqrt{\frac{\Psi_{\mathcal{S}_x}(\varepsilon_n)}{n\phi(\tilde{a}_n)}} \right\} < \infty. \quad (3.9.62)$$

The study of the term $u_{n,12}$ is deduced from the previous one. In fact :

$$\begin{aligned} &|u_{n,12}(\varphi, \mathbf{t}; \mathbf{h})| \\ &= |\mathbb{E}[u_n(\varphi, \mathbf{t}; \mathbf{h}) - u_n(\varphi, \mathbf{t}; \mathbf{h})]| \\ &\leq \frac{\mathbb{C}(n-m)!}{n! \tilde{\phi}(\mathbf{h})} \left| \sum_{i \in \mathbb{I}(m,n)} \mathbb{E} \left(\varphi(Y_{i_1}, \dots, Y_{i_m}) \left\{ \prod_{j=1}^m \mathbf{K} \left(\frac{d(X_{ij}, t_j)}{h_j} \right) - \prod_{j=1}^m \mathbf{K} \left(\frac{d(X_{ij}, t_{\ell(t_j)})}{h_j} \right) \right\} \right) \right| \end{aligned} \quad (3.9.63)$$

$$\leq \frac{\mathbb{C}(n-m)!}{n! \tilde{\phi}(\mathbf{h})} \sum_{i \in \mathbb{I}(m,n)} \mathbb{E} \left| \varphi(Y_{i_1}, \dots, Y_{i_m}) \left\{ \prod_{j=1}^m \mathbf{K} \left(\frac{d(X_{ij}, t_j)}{h_j} \right) - \prod_{j=1}^m \mathbf{K} \left(\frac{d(X_{ij}, t_{\ell(t_j)})}{h_j} \right) \right\} \right|. \quad (3.9.64)$$

To pass from (3.9.63) to (3.9.64) we apply Jensen's inequality in connection with some properties of the absolute value function. Then following the same way already taken, we get

$$|u_{n,12}(\varphi, \mathbf{t}; \mathbf{h})| \leq \mathbb{C} |u'_{n,12}(\varphi, \mathbf{t}; \mathbf{h})|,$$

where

$$|u'_{n,12}(\varphi, \mathbf{t}; \mathbf{h})| := \frac{(n-m)!}{n! \tilde{\phi}(\mathbf{h})} \sum_{i \in \mathbb{I}(m,n)} \mathbb{E} \left| \varphi(Y_{i_1}, \dots, Y_{i_m}) \sum_{j=1}^m \mathbf{K}_{i_j, h_j}^{(\ell)}(t_j, t_{\ell(t_j)}) \right|.$$

Notice that we have

$$\begin{aligned}
 & |u'_{n,12}(\varphi, \mathbf{t}; \mathbf{h})| \\
 & \leq \frac{(n-m)!}{n! \tilde{\phi}(\mathbf{h})} \mathbf{M} \mathbf{K}_2^{m-1} \sum_{i \in \mathcal{I}(m,n)} \sum_{j=1}^m \mathbb{E} \left| \left\{ \mathbf{K} \left(\frac{d(\mathbf{X}_{ij}, t_j)}{h_j} \right) - \mathbf{K} \left(\frac{d(\mathbf{X}_{ij}, t_{\ell(t_j)})}{h_j} \right) \right\} \right. \\
 & \quad \left. \times \mathbb{1}_{\mathcal{B}(t_j, h_j) \cup \mathcal{B}(t_{\ell(t_j)}, h_j)}(\mathbf{X}_{ij}) \prod_{q=1}^{j-1} \mathbb{1}_{\mathcal{B}(t_q, h_q/2)}(\mathbf{X}_{iq}) \prod_{p=j+1}^m \mathbb{1}_{\mathcal{B}(t_{\ell(t_p)}, h_p/2)}(\mathbf{X}_{ip}) \right| \\
 & \leq \frac{(n-m)!}{n!} m \mathbf{M} \mathbf{K}_2^{m-1} \sum_{i \in \mathcal{I}(m,n)} \frac{1}{m} \sum_{j=1}^m \left[\frac{\varepsilon_n}{\tilde{\phi}(\mathbf{A}_n) a_{n,j}} \mathbb{E} \left(\mathbb{1}_{\mathcal{B}(t_j, h_j) \cup \mathcal{B}(t_{\ell(t_j)}, h_j)}(\mathbf{X}_{ij}) \right. \right. \\
 & \quad \left. \left. \times \prod_{q=1}^{j-1} \mathbb{1}_{\mathcal{B}(t_q, h_q/2)}(\mathbf{X}_{iq}) \prod_{p=j+1}^m \mathbb{1}_{\mathcal{B}(t_{\ell(t_p)}, h_p/2)}(\mathbf{X}_{ip}) \right) \right] \\
 & \leq \frac{(n-m)!}{n!} m \mathbf{M} \mathbf{K}_2^{m-1} \sum_{i \in \mathcal{I}(m,n)} \frac{1}{m} \sum_{j=1}^m \left[\frac{\varepsilon_n}{C'_1 \tilde{\phi}(\tilde{a}_n) a_{n,j}} \mathbb{E} \left(\mathbb{1}_{\mathcal{B}(t_j, h_j) \cup \mathcal{B}(t_{\ell(t_j)}, h_j)}(\mathbf{X}_{ij}) \right. \right. \\
 & \quad \left. \left. \times \prod_{q=1}^{j-1} \mathbb{1}_{\mathcal{B}(t_q, h_q/2)}(\mathbf{X}_{iq}) \prod_{p=j+1}^m \mathbb{1}_{\mathcal{B}(t_{\ell(t_p)}, h_p/2)}(\mathbf{X}_{ip}) \right) \right].
 \end{aligned}$$

That implies

$$\begin{aligned}
 \sup_{\varphi \in \mathcal{F}_m} \sup_{\mathcal{K}^m} \sup_{\mathbf{h} \in \mathcal{H}_n^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} |u_{n,12}(\varphi, \mathbf{t}; \mathbf{h})| & \leq m \mathbf{M} \mathbf{K}_2^{m-1} \frac{\log(n)}{n \tilde{a}_n} \leq C'_7 \frac{\log(n)}{n \phi(\tilde{a}_n)} \\
 & = O \left(\sqrt{\frac{\Psi_{\mathcal{S}_x}(\varepsilon_n)}{n \phi(\tilde{a}_n)}} \right).
 \end{aligned}$$

This gives that

$$\sup_{\varphi \in \mathcal{F}_m} \sup_{\mathcal{K}^m} \sup_{\mathbf{h} \in \mathcal{H}_n^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} |u_{n,12}(\varphi, \mathbf{t}; \mathbf{h})| = O \left(\sqrt{\frac{\Psi_{\mathcal{S}_x}(\varepsilon_n)}{n \phi(\tilde{a}_n)}} \right). \quad (3.9.65)$$

Continue, now with $u_{n,13}$,

$$|u_{n,13}(\varphi, \mathbf{t}; \mathbf{h})| = |u_n(\varphi, \mathbf{t}_\ell; \mathbf{h}) - \mathbb{E}[u_n(\varphi, \mathbf{t}_\ell; \mathbf{h})]|.$$

Supposing that the kernel function $G_{\varphi, \mathbf{t}_\ell, \mathbf{h}}(\cdot)$ is symmetric, we need to decompose our U-statistic according to [Hoeffding \[1948\]](#) decomposition, we have

$$\begin{aligned}
 & u_{n,13}(\varphi, \mathbf{t}; \mathbf{h}) \\
 & := u_n(\varphi, \mathbf{t}_\ell; \mathbf{h}) - \mathbb{E}[u_n(\varphi, \mathbf{t}_\ell; \mathbf{h})] \\
 & = \sum_{p=1}^m \frac{m!}{(m-p)!} u_n^{(p)}(\pi_{p,m}(G_{\varphi, \mathbf{t}_\ell, \mathbf{h}})) \\
 & = m u_n^{(1)}(\pi_{1,m}(G_{\varphi, \mathbf{t}_\ell, \mathbf{h}})) + \sum_{p=2}^m \frac{m!}{(m-p)!} u_n^{(p)}(\pi_{p,m}(G_{\varphi, \mathbf{t}_\ell, \mathbf{h}})). \quad (3.9.66)
 \end{aligned}$$

Define new classes of functions, for $\mathbf{h} = (h_1, \dots, h_m) \in \mathcal{H}_n^{(m)}$, $\ell \in \{1, \dots, N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}})\}^m$, and $1 \leq p \leq m$:

$$(\mathcal{F}_m \mathcal{H}^m)^{(p)} := \left\{ \phi_{\mathbf{t}_\ell}(\mathbf{h}) \pi_{p,m}(\mathbf{G}_{\varphi, \mathbf{t}_\ell, \mathbf{h}})(\cdot, \cdot) \text{ for } \mathbf{h} \in \mathcal{H}_0^{(m)} \text{ and } \varphi \in \mathcal{F}_m \right\}.$$

These classes are VC-type classes of functions with the same characteristics, and the envelope function F_p satisfying

$$F_p \leq 2^p \kappa_2^m \|F\|_\infty.$$

Let us start with the linear term of (3.9.66), which is

$$mu_n^{(1)}(\pi_{1,m}(\mathbf{G}_{\varphi, \mathbf{t}_\ell, \mathbf{h}})) = \frac{m}{n} \sum_{j=1}^m \pi_{1,m}(\mathbf{G}_{\varphi, \mathbf{t}_\ell, \mathbf{h}})(X_j, Y_j).$$

From Hoeffding's projection, we have

$$\begin{aligned} \pi_{1,m}(\mathbf{G}_{\varphi, \mathbf{t}_\ell, \mathbf{h}})(x, y) &= \left\{ \mathbb{E} \left[\mathbf{G}_{\varphi, \mathbf{t}_\ell, \mathbf{h}} \left((x, X_2, \dots, X_m), (y, Y_2, \dots, Y_m) \right) \right] - \mathbb{E} \left[\mathbf{G}_{\varphi, \mathbf{t}_\ell, \mathbf{h}}(\mathbf{X}, \mathbf{Y}) \right] \right\} \\ &= \left\{ \mathbb{E} \left[\mathbf{G}_{\varphi, \mathbf{t}_\ell, \mathbf{h}}(\mathbf{X}, \mathbf{Y}) \mid (X_1, Y_1) = (x, y) \right] - \mathbb{E} \left[\mathbf{G}_{\varphi, \mathbf{t}_\ell, \mathbf{h}}(\mathbf{X}, \mathbf{Y}) \right] \right\}. \end{aligned}$$

One can see that

$$mu_n^{(1)}(\pi_{1,m}(\mathbf{G}_{\varphi, \mathbf{t}_\ell, \mathbf{h}})) =: \frac{1}{\sqrt{n}} \alpha_n(S_{1, \mathbf{t}_\ell, \mathbf{h}}),$$

is an empirical process based on a VC-type class of functions contained in $m(\mathcal{F}_m \mathcal{H}^m)^{(1)}$ with the same characteristics and the elements are defined by:

$$S_{1, \mathbf{t}_\ell, \mathbf{h}}(x, y) = m \phi_{\mathbf{t}_\ell}(\mathbf{h}) \mathbb{E} \left[\mathbf{G}_{\varphi, \mathbf{t}_\ell, \mathbf{h}}(\mathbf{X}, \mathbf{Y}) \mid (X_1, Y_1) = (x, y) \right].$$

Hence, the proof of this part is similar to that of the Lemma 3.9.1.4 and then :

$$\sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \sup_{\tilde{a}_n \leq h_\ell \leq b_0} \sup_{\varphi \in \mathcal{F}_m} \left| u_n^{(1)}(\pi_{1,m}(\mathbf{G}_{\varphi, \mathbf{t}_\ell, \mathbf{h}})) \right| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}{n \phi \left(\mu \phi^{-1} \left(\frac{\rho_n k_{1,n}}{n} \right) \right)}} \right).$$

Pass now to the nonlinear terms. The purpose is to prove that, for $2 \leq p \leq m$:

$$\sup_{\varphi \in \mathcal{F}_m} \sup_{\mathbf{h} \in \mathcal{H}_n^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \frac{\binom{m}{p} \sqrt{n \phi(\tilde{a}_n)} \left| u_n^{(p)}(\pi_{p,m}(\mathbf{G}_{\varphi, \mathbf{t}_\ell, \mathbf{h}})) \right|}{\sqrt{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}} = O_{a.co}(1), \quad (3.9.67)$$

to do that, we need to decompose the interval $\prod_{j=1}^m (a_{n,j}, b_{n,j})$ into smaller intervals. First, let us consider the intervals $(a_{n,j}, b_j)$ for all $j = 1, \dots, m$ and $b_j \in (0, 1)$ we note

$$\mathcal{H}_{i_j} = [a_{i_j}, b_{i_j}],$$

where $a_{i_j} = 2^{i-1} a_{n,j}$ and $b_{i_j} = 2^i a_{n,j}$ and we set $L_j(n) = \max\{i : b_{i_j} \leq 2b_j\}$, and $\mathcal{I}_j = \{i_j : 1 \leq i_j \leq L_j(n)\}$.

We can observe that

$$\prod_{j=1}^m (a_{n,j}, b_j) \subseteq \bigcup_{(i_1, \dots, i_m) \in \mathcal{I}_1 \times \dots \times \mathcal{I}_m} \prod_{j=1}^m \mathcal{H}_{i_j},$$

and

$$L_j(n) \sim \frac{\log(b_j/a_{n,j})}{\log(2)} \leq L(n) =: \max L_j(n) \text{ for } 1 \leq j \leq m. \quad (3.9.68)$$

Now, we set the following new classes, for $\mathbf{h} = (h_1, \dots, h_m) \in \mathcal{H}_n^{(m)}$ and $\ell \in \{1, \dots, N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}})\}^m$, $1 \leq j \leq m$, $1 \leq i \leq L(n)$ and $2 \leq p \leq m$:

$$\begin{aligned} (\mathcal{F}_m \mathcal{K}^m)_{i_j, \ell} &:= \left\{ \phi_{\mathbf{t}_\ell}(\mathbf{h}) G_{\varphi, \mathbf{t}_\ell, \mathbf{h}}(\cdot, \cdot) \text{ where } \varphi \in \mathcal{F}_m \text{ and } \mathbf{h} \in \prod_{j=1}^m \mathcal{H}_{i_j} \right\}, \\ (\mathcal{F}_m \mathcal{K}^m)_{i_j, \ell}^{(p)} &:= \left\{ \phi_{\mathbf{t}_\ell}(\mathbf{h}) \pi_{p,m}(G_{\varphi, \mathbf{t}_\ell, \mathbf{h}})(\cdot, \cdot) \text{ where } \varphi \in \mathcal{F}_m \text{ and } \mathbf{h} \in \prod_{j=1}^m \mathcal{H}_{i_j} \right\}. \end{aligned}$$

Thus, to prove (3.9.67), we need to prove that for $1 \leq j \leq m$ and $\ell = (\ell_1, \dots, \ell_m)$:

$$\max_{1 \leq \ell_j \leq N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}})} \max_{1 \leq i \leq L(n)} \sup_{\mathbf{h} \in \prod_{j=1}^m \mathcal{H}_{i_j}} \sup_{\varphi \in \mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \frac{\binom{m}{p} \sqrt{n \phi(\tilde{a}_n)} |u_n^{(p)}(\pi_{p,m}(G_{\varphi, \mathbf{t}_\ell, \mathbf{h}}))|}{\sqrt{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}} = O_{a.co}(1),$$

the key idea is to apply Lemma 3.9.1.10. We see clearly that the class of functions $(\mathcal{F}_m \mathcal{K}^m)_{i_j, \ell}$ is uniformly bounded, i.e.,

$$\sup_{\varphi \tilde{\mathbf{K}} \in (\mathcal{F}_m \mathcal{K}^m)_{i_j, \ell}} |\varphi(\cdot) \tilde{\mathbf{K}}(\cdot)| \leq M \kappa_2^m.$$

Moreover, by applying Proposition 2.6 of Arcones and Giné [1993] we have for each $(x_i, y_i) \in \mathcal{S}_{\mathcal{X}} \times \mathcal{Y}$ and a Rademacher variables ζ_1, \dots, ζ_n :

$$\begin{aligned} &\mathbb{E} \left\{ \left\| n^{1-m} \sum_{I_n^m} \zeta_{i_1} \zeta_{i_2} \varphi(y_{i_1}, \dots, y_{i_m}) \prod_{j=1}^m \mathbf{K} \left(\frac{d(x_{i_j}, t_j)}{h_j} \right) \right\|_{(\mathcal{F}_m \mathcal{K}^m)_{i_j, \ell}} \right\} \\ &\leq c_m \mathbb{E} \int_0^{D_{2,n}} \left[\log N(\epsilon, (\mathcal{F}_m \mathcal{K}^m)_{i_j, \ell}, d_{2,n}) \right]^{m/2} d\epsilon, \end{aligned}$$

where

$$D_{2,n} := \left\| \mathbb{E}_{\zeta}^{1/2} \left\{ n^{2(1-m)} \sum_{I_n^m} \left(\zeta_{i_1} \zeta_{i_2} \varphi(y_{i_1}, \dots, y_{i_m}) \prod_{j=1}^m \mathbf{K} \left(\frac{d(X_{i_j}, t_j)}{h_j} \right) \right)^2 \right\} \right\|_{(\mathcal{F}_m \mathcal{K}^m)_{i_j, \ell}}$$

and

$$d_{2,n} := \mathbb{E}_{\zeta}^{1/2} \left\{ n^{2(1-m)} \sum_{I_n^m} \left(\zeta_{i_1} \zeta_{i_2} \varphi(y_{i_1}, \dots, y_{i_m}) \prod_{j=1}^m \mathbf{K} \left(\frac{d(X_{i_j}, t_j)}{h_j} \right) \right)^2 \right\}.$$

We see that

$$D_{2,n} \leq n^{1-m} \frac{n!}{(n-m)!} M\kappa_2^m \leq nM\kappa_2^m.$$

So using the fact that $\mathcal{F}_m \mathcal{K}^m$ is a VC-type class of functions satisfying (A.5') which implies that the class $(\mathcal{F}_m \mathcal{K}^m)_{i_j, \ell}$ is also a VC-type class of functions with the same characteristics as $\mathcal{F}_m \mathcal{K}^m$, then :

$$\begin{aligned} & \mathbb{E} \left\{ \left\| n^{1-m} \sum_{I_n^m} \zeta_{i_1} \zeta_{i_2} \varphi(y_{i_1}, \dots, y_{i_m}) \prod_{j=1}^m K\left(\frac{d(x_{i_j}, t_j)}{h_j}\right) \right\|_{(\mathcal{F}_m \mathcal{K}^m)_{i_j, \ell}} \right\} \\ & \leq c_m \mathbb{E} \int_0^{D_{2,n}} \left[\log N(\epsilon, (\mathcal{F}_m \mathcal{K}^m)_{i_j, \ell}, d_{2,n}) \right]^{m/2} d\epsilon \\ & \leq c_m n M C^m. \end{aligned} \quad (3.9.69)$$

On the other hand, we have for each ϵ_0 , $1 \leq j \leq m$ and $\ell = (\ell_1, \dots, \ell_m)$:

$$\begin{aligned} & \mathbb{P} \left\{ \sup_{\mathbf{h} \in \mathcal{H}_0^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} \sup_{\varphi \tilde{\mathbf{K}} \in \mathcal{F}_m \mathcal{K}^m} \frac{\binom{m}{p} \sqrt{n\phi(\tilde{a}_n)} |u_n^{(p)}(\pi_{p,m} G_{\varphi, \mathbf{t}_{\ell(\mathbf{v})}, \mathbf{h})|}{\sqrt{\Psi_{\mathcal{S}_x}\left(\frac{\log n}{n}\right)}} \geq \epsilon_0 \right\} \\ & \leq \sum_{1 \leq \ell_j \leq N_{\epsilon_n}(\mathcal{S}_x)} \sum_{1 \leq i \leq L(n)} \mathbb{P} \left\{ \sup_{\mathbf{h} \in \prod_{j=1}^m \mathcal{H}_{i_j}} \sup_{\varphi \tilde{\mathbf{K}} \in \mathcal{F}_m \mathcal{K}^m} \frac{\binom{m}{p} \sqrt{n\phi(\tilde{a}_n)} |u_n^{(p)}(\pi_{p,m} G_{\varphi, \mathbf{t}_{\ell}, \mathbf{h})|}{\sqrt{\Psi_{\mathcal{S}_x}\left(\frac{\log n}{n}\right)}} \geq \epsilon_0 \right\} \\ & \leq L(n) N_{\epsilon_n}^m(\mathcal{S}_x) \max_{1 \leq \ell_j \leq N_{\epsilon_n}(\mathcal{S}_x)} \max_{1 \leq i \leq L(n)} \\ & \quad \times \mathbb{P} \left\{ \left\| \sum_{p=2}^m \binom{m}{p} u_n^{(p)}(\pi_{p,m} G_{\varphi, \mathbf{t}_{\ell(\mathbf{v})}, \mathbf{h}) \right\|_{(\mathcal{F}_m \mathcal{K}^m)_{i_j, \ell}} \geq \epsilon_0 \sqrt{\frac{\Psi_{\mathcal{S}_x}\left(\frac{\log n}{n}\right)}{n\phi(\tilde{a}_n)}} \right\}. \end{aligned}$$

All the conditions of Lemma 3.9.1.10 are fulfilled, so a direct application gives for each ϵ_0 , $1 \leq j \leq m$ and $\ell = (\ell_1, \dots, \ell_m)$:

$$\begin{aligned} & \mathbb{P} \left\{ \left\| \sum_{p=2}^m \binom{m}{p} u_n^{(p)}(\pi_{p,m} G_{\varphi, \mathbf{t}_{\ell(\mathbf{v})}, \mathbf{h}) \right\|_{(\mathcal{F}_m \mathcal{K}^m)_{i_j, \ell}} \geq \epsilon_0 \sqrt{\frac{m \Psi_{\mathcal{S}_x}\left(\frac{\log n}{n}\right)}{n\phi(\tilde{a}_n)}} \right\} \\ & \leq 2 \exp \left(- \frac{\epsilon_0 \sqrt{\frac{m \Psi_{\mathcal{S}_x}\left(\frac{\log n}{n}\right)}{n\phi(\tilde{a}_n)}}}{2^{m+5} m^{m+1} c_m M^2 \kappa_2^{2m}} \right) \\ & \leq n^{-\epsilon_0 C'_m}, \end{aligned} \quad (3.9.70)$$

where

$$\varepsilon_n = \frac{\log n}{n}, \quad C'_m > 0.$$

Furthermore, from equations (3.2.7), (3.9.68) and by choosing $\varepsilon_0 > 2Cte_m$ we get :

$$\begin{aligned} \sum_{n \geq 1} \mathbb{P} \left\{ \sup_{\mathbf{h} \in \mathcal{H}_0^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} \sup_{\varphi \tilde{\mathbf{K}} \in \mathcal{F}_m \mathcal{K}^m} \frac{\binom{m}{p} \sqrt{n\phi(\tilde{a}_n)} |u_n^{(p)}(\pi_{p,m} G_{\varphi, \mathbf{t}(\mathbf{t}), \mathbf{h}})|}{\sqrt{\Psi_{\mathcal{S}_x} \left(\frac{\log n}{n} \right)}} \geq \varepsilon_0 \right\} \\ \leq \sum_{n \geq 1} L(n) n^{-\varepsilon_0 C'_m} \\ \leq \infty. \end{aligned} \quad (3.9.71)$$

By combining (3.9.62), (3.9.65) and (3.9.71), we get

$$\sup_{\varphi \tilde{\mathbf{K}} \in \mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{h} \in \mathcal{H}_n^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} |u_n(\varphi, \mathbf{t}, \mathbf{h}) - \mathbb{E}(u_n(\varphi, \mathbf{t}, \mathbf{h}))| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_x} \left(\frac{\log n}{n} \right)}{n\phi(\tilde{a}_n)}} \right). \quad (3.9.72)$$

□

Proof of Theorem 3.4.1.8

To prove this theorem, we will need to truncate the conditional U-statistic. Taking the condition (A.4.a') into account, we have, for each $\lambda > 0$, $\mathbf{h} = (h_1, \dots, h_m)$ and for $\eta_n = n / (\log n)^2$:

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

which means that we truncate each function $\varphi \in \mathcal{F}_m$, either the envelope function as follows :

$$\begin{aligned} \varphi(\mathbf{y}) &= \varphi(\mathbf{y}) \mathbb{1}_{\{F(\mathbf{y}) \leq \lambda \eta_n^{1/p}\}} + \varphi(\mathbf{y}) \mathbb{1}_{\{F(\mathbf{y}) > \lambda \eta_n^{1/p}\}} \\ &= \varphi^{(T)}(\mathbf{y}) + \varphi^{(R)}(\mathbf{y}). \end{aligned}$$

We can write the U-statistic as follows

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

The first term of the right side $u_n^{(T)}(\varphi, \mathbf{t}, \mathbf{h})$ is called the truncated part and the second one $u_n^{(R)}(\varphi, \mathbf{t}, \mathbf{h})$ is the remainder part. Let's begin with $u_n^{(T)}(\varphi, \mathbf{t}, \mathbf{h})$.

Truncated Part :

Similar to previous procedures, we write

$$\begin{aligned}
 & \left| u_n^{(\text{T})}(\varphi, \mathbf{t}, \mathbf{h}) - \mathbb{E}(u_n^{(\text{T})}(\varphi, \mathbf{t}, \mathbf{h})) \right| \\
 & \leq \left| u_n^{(\text{T})}(\varphi, \mathbf{t}, \mathbf{h}) - u_n^{(\text{T})}(\varphi, \mathbf{t}_{\ell(\mathbf{t})}, \mathbf{h}) \right| \\
 & \quad + \left| \mathbb{E}(u_n^{(\text{T})}(\varphi, \mathbf{t}_{\ell(\mathbf{t})}, \mathbf{h})) - \mathbb{E}(u_n^{(\text{T})}(\varphi, \mathbf{t}, \mathbf{h})) \right| \\
 & \quad + \left| u_n^{(\text{T})}(\varphi, \mathbf{t}_{\ell(\mathbf{t})}, \mathbf{h}) - \mathbb{E}(u_n^{(\text{T})}(\varphi, \mathbf{t}_{\ell(\mathbf{t})}, \mathbf{h})) \right| \\
 & = \left| u_{n,1}^{(\text{T})}(\varphi, \mathbf{t}, \mathbf{h}) \right| + \left| u_{n,2}^{(\text{T})}(\varphi, \mathbf{t}, \mathbf{h}) \right| + \left| u_{n,3}^{(\text{T})}(\varphi, \mathbf{t}_{\ell(\mathbf{t})}, \mathbf{h}) \right|.
 \end{aligned}$$

Clearly

$$\begin{aligned}
 u_{n,1}^{(\text{T})}(\varphi, \mathbf{t}, \mathbf{h}) & := u_n^{(\text{T})}(\varphi, \mathbf{t}, \mathbf{h}) - u_n^{(\text{T})}(\varphi, \mathbf{t}_{\ell(\mathbf{t})}, \mathbf{h}), \\
 u_{n,2}^{(\text{T})}(\varphi, \mathbf{t}, \mathbf{h}) & := \mathbb{E}(u_n^{(\text{T})}(\varphi, \mathbf{t}_{\ell(\mathbf{t})}, \mathbf{h})) - \mathbb{E}(u_n^{(\text{T})}(\varphi, \mathbf{t}, \mathbf{h})), \\
 u_{n,3}^{(\text{T})}(\varphi, \mathbf{t}_{\ell(\mathbf{t})}, \mathbf{h}) & := u_n^{(\text{T})}(\varphi, \mathbf{t}_{\ell(\mathbf{t})}, \mathbf{h}) - \mathbb{E}(u_n^{(\text{T})}(\varphi, \mathbf{t}_{\ell(\mathbf{t})}, \mathbf{h})).
 \end{aligned}$$

Starting with the first term $\left| u_{n,1}^{(\text{T})}(\varphi, \mathbf{t}, \mathbf{h}) \right|$, by following the same arguments as it was done to treat $\left| u_{n,1}(\varphi, \mathbf{t}, \mathbf{h}) \right|$, we get :

$$\left| u_{n,1}^{(\text{T})}(\varphi, \mathbf{t}, \mathbf{h}) \right| := \frac{(n-m)!}{n! \phi_{\mathbf{t}}(\mathbf{h})} \sum_{i \in I(m,n)} \left| \varphi^{(\text{T})}(Y_{i_1}, \dots, Y_{i_m}) \sum_{j=1}^m \mathbb{K}_{i_j, h_j}^{(\ell)}(t_j, t_{\ell_j}) \right|.$$

It follows that

$$\begin{aligned}
 & \left| u_{n,1}^{(\text{T})}(\varphi, \mathbf{t}, \mathbf{h}) \right| \\
 & \leq \frac{\mathbb{C}(n-m)!}{n! \tilde{\phi}(\mathbf{h})} \kappa_2^{m-1} \sum_{i \in I(m,n)} \left| \mathbb{F}^{(\text{T})}(Y_{i_1}, \dots, Y_{i_m}) \right| \sum_{j=1}^m \left[\left| \mathbb{K} \left(\frac{d(X_{i_j}, t_j)}{h_j} \right) - \mathbb{K} \left(\frac{d(X_{i_j}, t_{\ell_j})}{h_j} \right) \right| \right. \\
 & \quad \left. \mathbb{1}_{\mathbb{B}(t_j, h_j) \cup \mathbb{B}(t_{\ell_j}, h_j)}(X_{i_j}) \prod_{q=1}^{j-1} \mathbb{1}_{\mathbb{B}(t_q, h_q/2)}(X_{i_q}) \prod_{p=j+1}^m \mathbb{1}_{\mathbb{B}(t_{\ell(p)}, h_p/2)}(X_{i_p}) \right] \quad (3.9.73)
 \end{aligned}$$

$$\begin{aligned}
 & \leq \frac{(n-m)!}{n!} \sum_{i \in I(m,n)} \frac{1}{m} \sum_{j=1}^m \left[\frac{m \kappa_2^{m-1} \varepsilon_n}{\Phi(\tilde{a}_n) a_{n,j}} \left| \mathbb{F}^{(\text{T})}(Y_{i_1}, \dots, Y_{i_m}) \right| \mathbb{1}_{\mathbb{B}(t_j, h_j) \cup \mathbb{B}(t_{\ell_j}, h_j)}(X_{i_j}) \right. \\
 & \quad \left. \times \prod_{q=1}^{j-1} \mathbb{1}_{\mathbb{B}(t_q, h_q/2)}(X_{i_q}) \prod_{p=j+1}^m \mathbb{1}_{\mathbb{B}(t_{\ell(p)}, h_p/2)}(X_{i_p}) \right]. \quad (3.9.74)
 \end{aligned}$$

Taking into account that

$$\mathbb{F}^{(\text{T})}(Y_1, \dots, Y_m) = \mathbb{F}(Y_1, \dots, Y_m) \mathbb{1}_{\{\mathbb{F}(Y_1, \dots, Y_m) \leq \lambda \eta_n^{1/p}\}}.$$

Then we apply Lemma 3.9.1.9 again on the function

$$\begin{aligned}
 f_{\mathbf{t}, \mathbf{h}}^{(\text{T})}(\mathbf{X}_i, \mathbf{Y}_i) & = \frac{1}{m} \sum_{j=1}^m \left[\frac{m \kappa_2^{m-1} \varepsilon_n}{\Phi(\tilde{a}_n) a_{n,j}} \left| \mathbb{F}^{(\text{T})}(Y_{i_1}, \dots, Y_{i_m}) \right| \mathbb{1}_{\mathbb{B}(t_j, h_j) \cup \mathbb{B}(t_{\ell_j}, h_j)}(X_{i_j}) \right. \\
 & \quad \left. \times \prod_{q=1}^{j-1} \mathbb{1}_{\mathbb{B}(t_q, h_q/2)}(X_{i_q}) \prod_{p=j+1}^m \mathbb{1}_{\mathbb{B}(t_{\ell(p)}, h_p/2)}(X_{i_p}) \right],
 \end{aligned}$$

with

$$\begin{aligned}\theta &= O(1/\phi(\tilde{a}_n)), \\ \sigma^2 &= O(1/(\phi^2(\tilde{a}_n))), \\ t &= \sqrt{\frac{\Psi_{\mathcal{X}}(\varepsilon_n)}{n\phi(\tilde{a}_n)} - \frac{1}{\phi(\tilde{a}_n)}},\end{aligned}$$

the concerned function satisfies for all $\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m$

$$0 \leq \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \sup_{\mathbf{h} \in \mathcal{H}_0^{(m)}} |f_{\mathbf{t}, \mathbf{h}}^{(\mathbb{T})}(\mathbf{X}, \mathbf{Y})| \leq \frac{m\lambda\kappa_2^{m-1}}{\phi(\tilde{a}_n)\tilde{a}_n \log(n)} \leq Cte = C_8.$$

Observe that the presence of the constant κ_2 on the last right side of the precedent inequality is derived from the condition (3.2.11), which indicates that

$$\mathbb{P} \left\{ \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \sup_{\mathbf{h} \in \mathcal{H}_n^{(m)}} \sup_{\varphi \in \mathcal{F}_m \mathcal{K}^m} |u_{n,11}^{(\mathbb{T})}(\varphi, \mathbf{t}, \mathbf{h})| \geq \sqrt{\frac{\Psi_{\mathcal{X}}(\varepsilon_n)}{n\phi(\tilde{a}_n)}} \right\} \quad (3.9.75)$$

$$\begin{aligned}&\leq \mathbb{P} \left\{ |u_n^{(m)}(f) - \theta| \geq \sqrt{\frac{\Psi_{\mathcal{X}}(\varepsilon_n)}{n\phi(\tilde{a}_n)} - \frac{\varepsilon_n}{\tilde{a}_n}} \right\} \\ &\leq \exp \left\{ -\frac{[n/m]t^2}{2\sigma^2 - \frac{2}{3}C_8 t} \right\}.\end{aligned} \quad (3.9.76)$$

After some algebra, while respecting the conditions (A.2.) and (A.6.), we obtain :

$$\sum_{n \geq 1} \mathbb{P} \left\{ \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \sup_{\mathbf{h} \in \mathcal{H}_n^{(m)}} \sup_{\varphi \in \mathcal{F}_m \mathcal{K}^m} |u_{n,11}^{(\mathbb{T})}(\varphi, \mathbf{t}, \mathbf{h})| \geq \sqrt{\frac{\Psi_{\mathcal{X}}(\varepsilon_n)}{n\phi(\tilde{a}_n)}} \right\} < \infty, \quad (3.9.77)$$

which implies

$$\sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \sup_{\mathbf{h} \in \mathcal{H}_n^{(m)}} \sup_{\varphi \in \mathcal{F}_m \mathcal{K}^m} |u_{n,1}(\varphi, \mathbf{t}, \mathbf{h})| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{X}}\left(\frac{\log n}{n}\right)}{n\phi(\tilde{a}_n)}} \right). \quad (3.9.78)$$

The study of the term $u_{n,2}^{(\mathbb{T})}$ is deduced from the previous one. In fact :

$$\begin{aligned}&|u_{n,2}^{(\mathbb{T})}(\varphi, \mathbf{t}, \mathbf{h})| \\ &= |\mathbb{E}(u_n^{(\mathbb{T})}(\varphi, \mathbf{t}, \mathbf{h}) - u_n^{(\mathbb{T})}(\varphi, \mathbf{t}_{\ell(\mathbf{t})}, \mathbf{h}))| \\ &\leq \frac{\mathbb{C}(n-m)!}{n! \tilde{\phi}(\mathbf{h})} \left| \sum_{i \in I(m,n)} \mathbb{E} \left(\varphi^{(\mathbb{T})}(Y_{i_1}, \dots, Y_{i_m}) \left\{ \prod_{j=1}^m \mathbb{K} \left(\frac{d(X_{i_j}, t_j)}{h_j} \right) - \prod_{j=1}^m \mathbb{K} \left(\frac{d(X_{i_j}, t_{\ell_j})}{h_j} \right) \right\} \right) \right| \quad (3.9.79)\end{aligned}$$

$$\leq \frac{\mathbb{C}(n-m)!}{n! \tilde{\phi}(\mathbf{h})} \sum_{i \in I(m,n)} \mathbb{E} \left| \varphi^{(\mathbb{T})}(Y_{i_1}, \dots, Y_{i_m}) \left\{ \prod_{j=1}^m \mathbb{K} \left(\frac{d(X_{i_j}, t_j)}{h_j} \right) - \prod_{j=1}^m \mathbb{K} \left(\frac{d(X_{i_j}, t_{\ell_j})}{h_j} \right) \right\} \right|. \quad (3.9.80)$$

Similar to the bounded case, we apply Jensen's inequality and some properties of the absolute value function to pass from (3.9.79) to (3.9.80), and by following the same way already taken, we get

$$\left| u_{n,2}^{(T)}(\varphi, \mathbf{t}, \mathbf{h}) \right| \leq \mathbb{C} \left| u_{n,21}^{(T)}(\varphi, \mathbf{t}, \mathbf{h}) \right|,$$

where

$$\left| u_{n,21}^{(T)}(\varphi, \mathbf{t}, \mathbf{h}) \right| := \frac{(n-m)!}{n! \tilde{\phi}(\mathbf{h})} \sum_{i \in I(m,n)} \mathbb{E} \left| \varphi^{(T)}(Y_{i_1}, \dots, Y_{i_m}) \sum_{j=1}^m K_{i_j, h_j}^{(\ell)}(t_j, t_{\ell_j}) \right|.$$

Note that

$$\begin{aligned} & \left| u_{n,21}^{(T)}(\varphi, \mathbf{t}, \mathbf{h}) \right| \\ & \leq \frac{(n-m)!}{n! \tilde{\phi}(\mathbf{h})} \kappa_2^{m-1} \sum_{i \in I(m,n)} \sum_{j=1}^m \mathbb{E} \left| F^{(T)}(Y_{i_1}, \dots, Y_{i_m}) \left\{ K\left(\frac{d(X_{i_j}, t_j)}{h_j}\right) - K\left(\frac{d(X_{i_j}, t_{\ell_j})}{h_j}\right) \right\} \right. \\ & \quad \left. \times \mathbb{1}_{B(t_j, h_j) \cup B(t_{\ell_j}, h_j)}(X_{i_j}) \prod_{q=1}^{j-1} \mathbb{1}_{B(t_q, h_q/2)}(X_{i_q}) \prod_{p=j+1}^m \mathbb{1}_{B(t_{\ell_p}, h_p/2)}(X_{i_p}) \right| \\ & \leq \frac{(n-m)!}{n!} m \lambda \kappa_2^{m-1} \sum_{i \in I(m,n)} \frac{\varepsilon_n}{C'_1 \phi(\tilde{a}_n) a_{n,j}} \mathbb{E}(F^{(T)}(Y_{i_1}, \dots, Y_{i_m})), \\ & \leq m \lambda \kappa_2^{m-1} \frac{\varepsilon_n}{C'_1 \phi(\tilde{a}_n) \tilde{a}_n} \mathbb{E}(F^{(T)}(Y_1, \dots, Y_m)), \\ & = m \lambda \kappa_2^{m-1} \frac{\varepsilon_n}{C'_1 \phi(\tilde{a}_n) \tilde{a}_n} \mathbb{E}(\mathbb{E}(F^{(T)}(Y_1, \dots, Y_m) | \mathbf{X} = \mathbf{t})) \end{aligned} \tag{3.9.81}$$

$$\begin{aligned} & \leq m \lambda \kappa_2^{m-1} \frac{\varepsilon_n}{C'_1 \phi(\tilde{a}_n) \tilde{a}_n} \theta_p^{1/p} \\ & = \frac{1}{n} \sum_{i=1}^n m \lambda \kappa_2^{m-1} \frac{\varepsilon_n}{C'_1 \phi(\tilde{a}_n) \tilde{a}_n} \theta_p^{1/p}. \end{aligned} \tag{3.9.82}$$

The passage from (3.9.81) to (3.9.82) is due to an application of Minkowski's inequality. Consequently, we get

$$\sup_{\mathbf{t} \in \mathcal{S}_x^m} \sup_{\mathbf{h} \in \mathcal{H}_n^{(m)}} \sup_{\varphi \in \mathcal{F}_m \mathcal{K}^m} \left| u_{n,21}^{(T)}(\varphi, \mathbf{t}, \mathbf{h}) \right| \leq \frac{1}{n} \sum_{i=1}^n m \lambda \kappa_2^{m-1} \frac{\varepsilon_n}{C'_1 \phi(\tilde{a}_n) \tilde{a}_n} \theta_p^{1/p}.$$

To finish the proof, we return to (3.9.36) and, accordingly :

$$\sup_{\mathbf{t} \in \mathcal{S}_x^m} \sup_{\mathbf{h} \in \mathcal{H}_n^{(m)}} \sup_{\varphi \in \mathcal{F}_m \mathcal{K}^m} \left| u_{n,2}^{(T)}(\varphi, \mathbf{t}, \mathbf{h}) \right| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_x} \left(\frac{\log n}{n} \right)}{n \phi(\tilde{a}_n)}} \right). \tag{3.9.83}$$

Now we pass to the study of $u_{n,3}^{(T)}$. Using the fact that the functions $\varphi^{(T)}(\cdot)$ are supposed to be symmetric, Hoeffding's decomposition of the considered U-statistic is possible, and get :

$$\begin{aligned} u_{n,3}^{(T)}(\varphi, \mathbf{t}_{\ell(t)}, \mathbf{h}) &= u_n(\varphi, \mathbf{t}_{\ell}; \mathbf{h}) - \mathbb{E}[u_n(\varphi, \mathbf{t}_{\ell}; \mathbf{h})] \\ &= \sum_{p=1}^m \frac{m!}{(m-p)!} u_n^{(p)}(\pi_{p,m}(G_{\varphi, \mathbf{t}_{\ell}, \mathbf{h}})) \end{aligned} \quad (3.9.84)$$

$$= m u_n^{(1)}(\pi_{p,m}(G_{\varphi, \mathbf{t}_{\ell}, \mathbf{h}})) + \sum_{p=2}^m \frac{m!}{(m-p)!} u_n^{(p)}(\pi_{p,m}(G_{\varphi, \mathbf{t}_{\ell}, \mathbf{h}})), \quad (3.9.85)$$

define new classes of functions, for $\mathbf{h} = (h_1, \dots, h_m) \in \mathcal{H}_n^{(m)}$, $\ell \in \{1, \dots, N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}})\}^m$, $1 \leq p \leq m$ and any $\mathbf{y} \in \mathcal{Y}^m$, \mathcal{I}_{λ} :

$$\begin{aligned} \mathcal{I}_{\lambda} &:= \left\{ \mathbb{1}_{\{F(\mathbf{y}) \leq \lambda^{1/p}\}} : \lambda > 0 \right\}. \\ (\mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n})^{(p)} &:= \left\{ \phi_{\mathbf{t}_{\ell}}(\mathbf{h}) \pi_{p,m}(G_{\varphi, \mathbf{t}_{\ell}, \mathbf{h}})(\cdot, \cdot) \text{ for } \mathbf{h} \in \mathcal{H}_0^{(m)} \text{ and } \varphi \in \mathcal{F}_m \mathcal{I}_{\lambda \eta_n} \right\}. \end{aligned}$$

These classes are a VC-type classes of functions (since the classes \mathcal{F}_m , \mathcal{K}^m and $\mathcal{I}_{\lambda \eta_n}$ are too) and the envelope function $F_k^{(T)}(\cdot)$ satisfying :

$$F_p^{(T)} \leq 2^p \kappa_2^m \|F^{(T)}\|_{\infty}.$$

To prove (3.4.2), we begin by studying the linear term of (3.9.84), which means

$$m u_n^{(1)}\left(\pi_{1,m}(G_{\varphi, \mathbf{t}_{\ell(t)}, \mathbf{h}}^{(T)})\right) = \frac{m}{n} \sum_{i=1}^n \pi_{1,m} G_{\varphi, \mathbf{t}_{\ell(t)}, \mathbf{h}}^{(T)}(X_i, Y_i),$$

then we have :

$$\begin{aligned} \pi_{1,m} G_{\varphi, \mathbf{t}_{\ell(t)}, \mathbf{h}}^{(T)}(x, y) &= \mathbb{E} \left[G_{\varphi, \mathbf{t}_{\ell(t)}, \mathbf{h}}^{(T)}((x, X_2, \dots, X_m), (y, Y_2, \dots, Y_m)) \right] - \mathbb{E} \left[G_{\varphi, \mathbf{t}_{\ell(t)}, \mathbf{h}}^{(T)}(\mathbf{X}, \mathbf{Y}) \right] \\ &= \mathbb{E} \left[G_{\varphi, \mathbf{t}_{\ell(t)}, \mathbf{h}}^{(T)}(\mathbf{X}, \mathbf{Y}) \mid (X_1, Y_1) = (x, y) \right] - \mathbb{E} \left[G_{\varphi, \mathbf{t}_{\ell(t)}, \mathbf{h}}^{(T)}(\mathbf{X}, \mathbf{Y}) \right]. \end{aligned}$$

It is easily seen that

$$m u_n^{(1)}\left(\pi_{1,m}(G_{\varphi, \mathbf{t}_{\ell(t)}, \mathbf{h}}^{(T)})\right) =: \frac{1}{\sqrt{n}} \alpha_n \left(S_{\varphi, \mathbf{t}_{\ell(t)}, \mathbf{h}}^{(T)} \right)$$

is an empirical process based on a VC-type class of functions contained in $m(\mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n})^{(1)}$ with the same characteristics and that the elements are defined by :

$$S_{\varphi, \mathbf{t}_{\ell(t)}, \mathbf{h}}^{(T)}(x, y) := m \mathbb{E} \left[G_{\varphi, \mathbf{t}_{\ell(t)}, \mathbf{h}}^{(T)}(\mathbf{X}, \mathbf{Y}) \mid (X_1, Y_1) = (x, y) \right].$$

Thus, the proof of this part is similar to that of the Lemma 3.9.1.4 and then :

$$\sup_{\mathbf{h} \in \mathcal{H}_0^{(1)}} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \sup_{\varphi \tilde{\kappa} \in \mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n}} \left| u_n^{(1)}\left(\pi_{1,m} G_{\varphi, \mathbf{t}_{\ell(t)}, \mathbf{h}}^{(T)}\right) \right| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}{n \phi(\tilde{a}_n)}} \right).$$

We continue, now, with the nonlinear terms; our aim is to prove that, for $2 \leq p \leq m$:

$$\sup_{\mathbf{h} \in \mathcal{H}_0^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \sup_{\varphi \tilde{\mathbf{K}} \in \mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n}} \frac{\binom{m}{p} \sqrt{n \phi(\tilde{a}_n)} \left| u_n^{(p)} \left(\pi_{p,m} G_{\varphi, \mathbf{t}(\ell), \mathbf{h}}^{(\text{T})} \right) \right|}{\sqrt{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}} = O_{a.co}(1),$$

which, for $\ell = (\ell_1, \dots, \ell_m)$ and $1 \leq j \leq m$, is equivalent to :

$$\max_{1 \leq \ell_j \leq N_{\epsilon_n}(\mathcal{S}_{\mathcal{X}})} \max_{1 \leq i \leq L(n)} \sup_{\mathbf{h} \in \prod_{j=1}^m \mathcal{H}_{i_j}} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \sup_{\varphi \tilde{\mathbf{K}} \in \mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n}} \frac{\binom{m}{p} \sqrt{n \phi(\tilde{a}_n)} \left| u_n^{(p)} \left(\pi_{p,m} G_{\varphi, \mathbf{t}(\ell), \mathbf{h}}^{(\text{T})} \right) \right|}{\sqrt{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}} = O_{a.co}(1),$$

where \mathcal{H}_{i_j} is defined in the same fashion as previously, and we consider new classes :

$$\begin{aligned} (\mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n})_{i_j, \ell} &:= \left\{ \phi_{\mathbf{t}(\ell)}(\mathbf{h}) G_{\varphi, \mathbf{t}(\ell), \mathbf{h}}^{(\text{T})}(\cdot, \cdot) \text{ where } \varphi^{(\text{T})} \in \mathcal{F}_m \mathcal{I}_{\lambda \eta_n} \text{ and } \mathbf{h} \in \prod_{j=1}^m \mathcal{H}_{i_j} \right\}, \\ (\mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n})_{i_j, \ell}^{(p)} &:= \left\{ \phi_{\mathbf{t}(\ell)}(\mathbf{h}) \pi_{p,m}(G_{\varphi, \mathbf{t}(\ell), \mathbf{h}}^{(\text{T})})(\cdot, \cdot) \text{ where } \varphi^{(\text{T})} \in \mathcal{F}_m \mathcal{I}_{\lambda \eta_n} \text{ and } \mathbf{h} \in \prod_{j=1}^m \mathcal{H}_{i_j} \right\}, \end{aligned}$$

the key idea is to apply Lemma 3.9.1.10. We see clearly that the class of functions $(\mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n})_{i_j, \ell}$ is uniformly bounded, i.e.,

$$\sup_{\varphi^{(\text{T})} \tilde{\mathbf{K}} \in (\mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n})_{i_j, \ell}} |\varphi^{(\text{T})}(\cdot) \tilde{\mathbf{K}}(\cdot)| \leq \lambda \eta_n \kappa_2^m.$$

Furthermore, by applying Proposition 2.6 of Arcones and Giné [1993], we have for each $(x_i, y_i) \in \mathcal{S}_{\mathcal{X}} \times \mathcal{Y}$ and a Rademacher variables ζ_1, \dots, ζ_n :

$$\begin{aligned} &\mathbb{E} \left\{ \left\| n^{1-m} \sum_{I_n^m} \zeta_{i_1} \zeta_{i_2} \varphi^{(\text{T})}(y_{i_1}, \dots, y_{i_m}) \prod_{j=1}^m \mathbf{K} \left(\frac{d(x_{i_j}, t_j)}{h_j} \right) \right\|_{(\mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n})_{i_j, \ell}} \right\} \\ &\leq c_m \mathbb{E} \int_0^{D_{2,n}} \left[\log N(\epsilon, (\mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n})_{i_j, \ell}, d_{2,n}) \right]^{m/2} d\epsilon, \end{aligned}$$

where

$$D_{2,n} := \left\| \mathbb{E}_{\zeta}^{1/2} \left\{ n^{2(1-m)} \sum_{I_n^m} \left(\zeta_{i_1} \zeta_{i_2} \varphi^{(\text{T})}(Y_{i_1}, \dots, Y_{i_m}) \prod_{j=1}^m \mathbf{K} \left(\frac{d(X_{i_j}, t_j)}{h_j} \right) \right)^2 \right\} \right\|_{(\mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n})_{i_j, \ell}},$$

and

$$d_{2,n} := \mathbb{E}_{\zeta}^{1/2} \left\{ n^{2(1-m)} \sum_{I_n^m} \left(\zeta_{i_1} \zeta_{i_2} \varphi^{(\text{T})}(Y_{i_1}, \dots, Y_{i_m}) \prod_{j=1}^m \mathbf{K} \left(\frac{d(X_{i_j}, t_j)}{h_j} \right) \right)^2 \right\}.$$

Observe that :

$$D_{2,n} \leq n^{1-m} \frac{n!}{(n-m)!} \lambda \kappa_2^m \eta_n \leq n \eta_n \kappa_2^m.$$

Making use of the fact that $\mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n}$ is a VC-type class of functions satisfying **(A.5')** which implies that the class $(\mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n})_{i,j,\ell}$ is also a VC-type class of functions with the same characteristics as $\mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n}$, it follows that :

$$\begin{aligned} & \mathbb{E} \left\{ \left\| n^{1-m} \sum_{I_n^m} \zeta_{i_1} \zeta_{i_2} \varphi^{(\text{T})}(y_{i_1}, \dots, y_{i_m}) \prod_{j=1}^m \mathbb{K} \left(\frac{d(x_{i_j}, t_j)}{h_j} \right) \right\|_{(\mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n})_{j,\ell}} \right\} \\ & \leq c_m \mathbb{E} \int_0^{\text{D}_{2,n}} \left[\log N(\epsilon, (\mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n})_{i,j,\ell}, d_{2,n}) \right]^{m/2} d\epsilon \\ & \leq c_m n \lambda C^m \eta_n. \end{aligned} \quad (3.9.86)$$

On the other hand, we have for each ϵ_0 , $\ell = (\ell_1, \dots, \ell_m)$ and $1 \leq j \leq m$:

$$\begin{aligned} & \mathbb{P} \left\{ \sup_{\mathbf{h} \in \mathcal{H}_0^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \sup_{\varphi^{(\text{T})} \tilde{\mathbf{K}} \in \mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n}} \frac{\binom{m}{p} \sqrt{n \phi(\tilde{a}_n)} |u_n^{(p)}(\pi_{p,m} \mathbf{G}_{\varphi, \mathbf{t}(\ell), \mathbf{h}}^{(\text{T})})|}{\sqrt{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}} \geq \epsilon_0 \right\} \\ & \leq \sum_{1 \leq \ell_j \leq N_{\epsilon_n}(\mathcal{S}_{\mathcal{X}})} \sum_{1 \leq i \leq L(n)} \mathbb{P} \left\{ \sup_{\mathbf{h} \in \prod_{j=1}^m \mathcal{H}_{i,j}} \sup_{\varphi^{(\text{T})} \tilde{\mathbf{K}} \in \mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n}} \frac{\binom{m}{p} \sqrt{n \phi(\tilde{a}_n)} |u_n^{(p)}(\pi_{p,m} \mathbf{G}_{\varphi, \mathbf{t}(\ell), \mathbf{h}}^{(\text{T})})|}{\sqrt{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}} \geq \epsilon_0 \right\} \\ & \leq L(n) N_{\epsilon_n}^m(\mathcal{S}_{\mathcal{X}}) \max_{1 \leq \ell_i \leq N_{\epsilon_n}(\mathcal{S}_{\mathcal{X}})} \max_{1 \leq j \leq L(n)} \\ & \quad \times \mathbb{P} \left\{ \left\| \sum_{p=2}^m \binom{m}{p} u_n^{(p)}(\pi_{p,m} \mathbf{G}_{\varphi, \mathbf{t}(\ell), \mathbf{h}}^{(\text{T})}) \right\|_{(\mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n})_{j,\ell}} \geq \epsilon_0 \sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}{n \phi(\tilde{a}_n)}} \right\}. \end{aligned}$$

We are now in position to apply Lemma 3.9.1.10 , for each ϵ_0 , $1 \leq j \leq m$ and $\ell = (\ell_1, \dots, \ell_m)$, one gets :

$$\begin{aligned} & \mathbb{P} \left\{ \left\| \sum_{p=2}^m \binom{m}{p} u_n^{(p)}(\pi_{p,m} \mathbf{G}_{\varphi, \mathbf{t}(\ell), \mathbf{h}}^{(\text{T})}) \right\|_{(\mathcal{F}_m \mathcal{K}^m \mathcal{I}_{\lambda \eta_n})_{j,\ell}} \geq \epsilon_0 \sqrt{\frac{m \Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}{n \phi(\tilde{a}_n)}} \right\} \\ & \leq 2 \exp \left(- \frac{\epsilon_0^2 \sqrt{\frac{m \Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}{n \phi(\tilde{a}_n)}}}{2^{m+5} m^{m+1} c_m M^2 \kappa_2^{2m}} \right) \\ & \leq n^{-\epsilon_0^2 C_m} \end{aligned} \quad (3.9.87)$$

where

$$\epsilon_n = \frac{\log n}{n}, \quad C_m > 0.$$

Moreover, from equations (3.2.7), (3.9.68) and by choosing $\epsilon_0 > 2Cte_m$ we get :

$$\sum_{n \geq 1} \mathbb{P} \left\{ \sup_{\mathbf{h} \in \mathcal{H}_0^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \sup_{\varphi \in \mathcal{F}_m} \sup_{\mathbf{K} \in \mathcal{K}^m} \sup_{\mathcal{I}_{\lambda \eta_n}} \frac{\binom{m}{p} \sqrt{n\phi(\tilde{a}_n)} \left| u_n^{(p)} \left(\pi_{p,m} G_{\varphi, \mathbf{t}(\mathbf{t}), \mathbf{h}}^{(T)} \right) \right|}{\sqrt{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}} \geq \epsilon_0 \right\} \leq \sum_{n \geq 1} L(n) n^{-\epsilon_0^2 C_m} \leq \infty, \quad (3.9.88)$$

for an appropriate chosen C_m .

Remainder Part

Consider now the remainder part, that is the U-process $u_n^{(R)}(\varphi, \mathbf{t}, \mathbf{h})$ based on the unbounded kernel given by :

$$G_{\varphi, \mathbf{t}, \mathbf{h}}^{(R)}(\mathbf{x}, \mathbf{y}) = G_{\varphi, \mathbf{t}, \mathbf{h}}(\mathbf{x}, \mathbf{y}) \mathbb{1}_{\{F(\mathbf{y}) > \lambda \eta_n^{1/p}\}}.$$

We shall prove that the process is negligible, more precisely

$$\sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \sup_{\mathbf{h} \in \mathcal{H}_0^{(m)}} \sup_{\varphi \in \mathcal{F}_m} \sup_{\mathbf{K} \in \mathcal{K}^m} \frac{\sqrt{n\phi(\tilde{a}_n)} \left| u_n^{(m)}(G_{\varphi, \mathbf{t}, \mathbf{h}}^{(R)}) - \mathbb{E} \left(u_n^{(m)}(G_{\varphi, \mathbf{t}, \mathbf{h}}^{(R)}) \right) \right|}{\sqrt{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}} = o_{a.co}(1). \quad (3.9.89)$$

Denote for each \mathbf{y} ,

$$\tilde{F}(\mathbf{y}) := \kappa_2^m F(\mathbf{y}).$$

It is clearly seen that for all $\varphi \in \mathcal{F}_m, \mathbf{x}, \mathbf{y} \in \mathcal{X}^m$,

$$\tilde{F}(\mathbf{y}) \geq \phi(\tilde{a}_n) |G_{\varphi, \mathbf{t}, \mathbf{h}}(\mathbf{x}, \mathbf{y})|,$$

since \tilde{F} is symmetric it holds that :

$$\left| u_n^{(m)}(G_{\varphi, \mathbf{t}, \mathbf{h}}^{(R)}) \right| \leq \phi^{-1}(\tilde{a}_n) u_n^{(m)}(\tilde{F} \mathbb{1}_{\{F(\mathbf{y}) > \lambda \eta_n^{1/p}\}}),$$

where $u_n^{(m)}(\tilde{F} \mathbb{1}_{\{F(\mathbf{y}) > \lambda \eta_n^{1/p}\}})$ is a U-statistic based on the U- kernel $\tilde{F} \mathbb{1}_{\{F(\mathbf{y}) > \lambda \eta_n^{1/p}\}}$:

$$\begin{aligned} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \sup_{\mathbf{h} \in \mathcal{H}_0^{(m)}} \sup_{\varphi \in \mathcal{F}_m} \sup_{\mathbf{K} \in \mathcal{K}^m} \frac{\sqrt{n\phi(\tilde{a}_n)} \left| u_n^{(m)}(G_{\varphi, \mathbf{t}, \mathbf{h}}^{(R)}) \right|}{\sqrt{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}} &\leq (n\eta_n)^{1/2} u_n^{(m)}(\tilde{F} \mathbb{1}_{\{F(\mathbf{y}) > \lambda \eta_n^{1/p}\}}) \\ &\leq C_9 \eta_n u_n^{(m)}(\tilde{F} \mathbb{1}_{\{F(\mathbf{y}) > \lambda \eta_n^{1/p}\}}), \end{aligned} \quad (3.9.90)$$

and

$$\begin{aligned} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \sup_{\mathbf{h} \in \mathcal{H}_0^{(m)}} \sup_{\varphi \mathbf{K} \in \mathcal{F}_m \mathcal{K}^m} \frac{\sqrt{n\phi(\tilde{a}_n)} \left| \mathbb{E} \left(u_n^{(m)} \left(G_{\varphi, \mathbf{t}, \mathbf{h}}^{(R)} \right) \right) \right|}{\sqrt{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}} &\leq C_9 \eta_n \mathbb{E} \left(u_n^{(m)} \left(\tilde{\mathbb{F}} \mathbb{1}_{\{\mathbb{F}(\mathbf{Y}) > \lambda \eta_n^{1/p}\}} \right) \right) \\ &\leq C_9 \mathbb{E} \left(\tilde{\mathbb{F}}^{1+p} \mathbb{1}_{\{\mathbb{F}(\mathbf{Y}) > \lambda \eta_n^{1/p}\}} \right). \end{aligned}$$

Therefore, as $n \rightarrow \infty$, we have

$$\sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \sup_{\mathbf{h} \in \mathcal{H}_0^{(m)}} \sup_{\varphi \mathbf{K} \in \mathcal{F}_m \mathcal{K}^m} \frac{\sqrt{n\phi(\tilde{a}_n)} \left| \mathbb{E} \left(u_n^{(m)} \left(G_{\varphi, \mathbf{t}, \mathbf{h}}^{(R)} \right) \right) \right|}{\sqrt{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}} = o(1). \quad (3.9.91)$$

Thus to finish the proof, it suffices to show that :

$$u_n^{(m)} \left(\tilde{\mathbb{F}} \mathbb{1}_{\{\mathbb{F}(\mathbf{Y}) > \lambda \eta_n^{1/p}\}} \right) = o_{a.co} \left((n\eta_n)^{-1/2} \right). \quad (3.9.92)$$

By applying Chebyshev's inequality, we get, for any $\delta > 0$

$$\begin{aligned} \mathbb{P} \left\{ \left| u_n^{(m)} \left(\tilde{\mathbb{F}} \mathbb{1}_{\{\mathbb{F}(\mathbf{Y}) > \lambda \eta_n^{1/p}\}} \right) - \mathbb{E} \left(u_n^{(m)} \left(\tilde{\mathbb{F}} \mathbb{1}_{\{\mathbb{F}(\mathbf{Y}) > \lambda \eta_n^{1/p}\}} \right) \right) \right| \geq \delta (n\eta_n)^{-1/2} \right\} \\ \leq \delta^{-2} (n\eta_n) \text{Var} \left(u_n^{(m)} \left(\tilde{\mathbb{F}} \mathbb{1}_{\{\mathbb{F}(\mathbf{Y}) > \lambda \eta_n^{1/p}\}} \right) \right) \leq m \delta^{-2} \eta_n \mathbb{E} \left(\tilde{\mathbb{F}}^2 \mathbb{1}_{\{\mathbb{F}(\mathbf{Y}) > \lambda \eta_n^{1/p}\}} \right) \\ \leq \frac{m}{n^2} \delta^{-2} (\eta_n)^p \mathbb{E} \left(\tilde{\mathbb{F}}^2 \mathbb{1}_{\{\mathbb{F}(\mathbf{Y}) > \lambda \eta_n^{1/p}\}} \right) \leq \delta' \mathbb{E} \left(\tilde{\mathbb{F}}^3 \mathbb{1}_{\{\mathbb{F}(\mathbf{Y}) > \lambda \eta_n^{1/p}\}} \right) \frac{1}{n^2}, \end{aligned}$$

so by using the fact that

$$\delta' \mathbb{E} \left(\tilde{\mathbb{F}}^3 \mathbb{1}_{\{\mathbb{F}(\mathbf{Y}) > \lambda \eta_n^{1/p}\}} \right) \sum_{n \geq 1} \frac{1}{n^2} < \infty,$$

we deduce that

$$\sum_{n \geq 1} \mathbb{P} \left\{ \left| u_n^{(m)} \left(\tilde{\mathbb{F}} \mathbb{1}_{\{\mathbb{F}(\mathbf{Y}) > \lambda \eta_n^{1/p}\}} \right) - \mathbb{E} \left(u_n^{(m)} \left(\tilde{\mathbb{F}} \mathbb{1}_{\{\mathbb{F}(\mathbf{Y}) > \lambda \eta_n^{1/p}\}} \right) \right) \right| \geq \delta (n\eta_n)^{-1/2} \right\} < \infty.$$

Finally, note that (3.9.90) implies

$$\mathbb{E} \left(u_n^{(m)} \left(\tilde{\mathbb{F}} \mathbb{1}_{\{\mathbb{F}(\mathbf{Y}) > \lambda \eta_n^{1/p}\}} \right) \right) = o \left((n\eta_n)^{-1/2} \right).$$

The above results of the arbitrary choice of $\lambda > 0$ imply that (3.9.92) holds, which, when combined with (3.9.91) and (3.9.90), achieves the proof of (3.9.89). Consequently, Lemma 3.4.1.8 is proved since the desired result for the truncated part was also established, hence

$$\sup_{\varphi \tilde{\mathbf{K}} \in \mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{h} \in \mathcal{H}_n^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} |u_n(\varphi, \mathbf{t}, \mathbf{h}) - \mathbb{E}(u_n(\varphi, \mathbf{t}, \mathbf{h}))| = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}{n\phi(\tilde{a}_n)}} \right). \quad (3.9.93)$$

Hence the proof is complete. \square

Proof of Theorem 3.4.1.9

Notice that

$$\begin{aligned}
 & \left| \widehat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h}) - \widehat{\mathbb{E}}\left(\widehat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h})\right) \right| \\
 &= \left| \frac{u_n(\varphi, \mathbf{t}, \mathbf{h})}{u_n(1, \mathbf{t}, \mathbf{h})} - \frac{\mathbb{E}(u_n(\varphi, \mathbf{t}, \mathbf{h}))}{\mathbb{E}(u_n(1, \mathbf{t}, \mathbf{h}))} \right| \\
 &\leq \frac{|u_n(\varphi, \mathbf{t}, \mathbf{h}) - \mathbb{E}(u_n(\varphi, \mathbf{t}, \mathbf{h}))|}{|u_n(1, \mathbf{t}, \mathbf{h})|} \\
 &\quad + \frac{|\mathbb{E}(u_n(\varphi, \mathbf{t}, \mathbf{h}))| \cdot |u_n(1, \mathbf{t}, \mathbf{h}) - \mathbb{E}(u_n(1, \mathbf{t}, \mathbf{h}))|}{|u_n(1, \mathbf{t}, \mathbf{h})| \cdot |\mathbb{E}(u_n(1, \mathbf{t}, \mathbf{h}))|} \\
 &=: \text{II} + \text{III}.
 \end{aligned}$$

Under the imposed hypothesis and the previously obtained results, and for some c'_1, c'_2 see that :

$$\begin{aligned}
 \sup_{\mathbf{h} \in \mathcal{H}_0^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} |u_n(1, \mathbf{t}, \mathbf{h})| &= c'_1 \quad \text{a.co} \\
 \sup_{\mathbf{h} \in \mathcal{H}_0^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} |\mathbb{E}(u_n(1, \mathbf{t}, \mathbf{h}))| &= c'_2 \\
 \sup_{\varphi \widetilde{\mathbf{K}} \in \mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{h} \in \mathcal{H}_0^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} |\mathbb{E}(u_n(\varphi, \mathbf{t}, \mathbf{h}))| &= O(1).
 \end{aligned}$$

Therefore, we can now apply Theorem 3.4.1.7 to handle III and Theorems 3.4.1.7 and 3.4.1.8 to handle II, depending on whether the class \mathcal{F}_m satisfies (A.4.a') or (A.4.a), and get for some $c'' > 0$ with probability 1 :

$$\begin{aligned}
 & \sup_{\varphi \widetilde{\mathbf{K}} \in \mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{h} \in \mathcal{H}_0^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} \frac{\sqrt{n\phi(\widetilde{a}_n)} \left| \widehat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h}) - \widehat{\mathbb{E}}\left(\widehat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h})\right) \right|}{\sqrt{\Psi_{\mathcal{S}_x}(\varepsilon_n)}} \\
 &\leq \sup_{\varphi \widetilde{\mathbf{K}} \in \mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{h} \in \mathcal{H}_0^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} \frac{\sqrt{n\phi(\widetilde{a}_n)} \text{(II)}}{\sqrt{\Psi_{\mathcal{S}_x}(\varepsilon_n)}} \\
 &\quad + \sup_{\varphi \widetilde{\mathbf{K}} \in \mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{h}(\mathbf{t}) \in \mathcal{H}_0^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_x^m} \frac{\sqrt{n\phi(\widetilde{a}_n)} \text{(III)}}{\sqrt{\Psi_{\mathcal{S}_x}(\varepsilon_n)}} \\
 &\leq c''.
 \end{aligned}$$

Hence the proof is complete. □

Proof of Theorem 3.4.1.10

Under (A.4.b) and (H.2.) one has

$$\mathbb{E}[u_n(1, \mathbf{t}; \mathbf{h})] \geq \kappa_1^m \frac{\phi_{\mathbf{t}}(\mathbf{h}/2)}{\phi_{\mathbf{t}}(\mathbf{h})} \geq \kappa_1^m C_3,$$

which implies that

$$\begin{aligned}
 & \left| \frac{\mathbb{E}[u_n(\varphi, \mathbf{t}; \mathbf{h})]}{\mathbb{E}[u_n(1, \mathbf{t}; \mathbf{h})]} - r^{(m)}(\varphi, \mathbf{t}) \right| \\
 & \leq |\kappa_1^{-m} C_3^{-1} \mathbb{E}[u_n(\varphi, \mathbf{t}; \mathbf{h})] - r^{(m)}(\varphi, \mathbf{t})| \\
 & \leq \left| \kappa_1^{-m} C_3^{-1} \mathbb{E} \left(\frac{(n-m)!}{n!} \sum_{i \in I(m,n)} \frac{1}{\phi_{\mathbf{t}}(\mathbf{h})} \varphi(Y_{i_1}, \dots, Y_{i_m}) \prod_{j=1}^m \mathbb{K} \left(\frac{d(X_{i_j}, t_j)}{h_j} \right) \right) \right. \\
 & \quad \left. - r^{(m)}(\varphi, \mathbf{t}) \right| \\
 & \leq \left| \kappa_1^{-m} C_3^{-1} \frac{1}{\phi_{\mathbf{t}}(\mathbf{h})} \mathbb{E} \left(\varphi(Y_1, \dots, Y_m) \prod_{i=1}^m \mathbb{K} \left(\frac{d(X_i, t_i)}{h_i} \right) \right) - r^{(m)}(\varphi, \mathbf{t}) \right| \\
 & \leq \frac{1}{\phi_{\mathbf{t}}(\mathbf{h}) \kappa_1^m C_3} \mathbb{E} \left(\prod_{i=1}^m \mathbb{K} \left(\frac{d(X_i, t_i)}{h_i} \right) \left| r^{(m)}(\varphi, \mathbf{X}) - r^{(m)}(\varphi, \mathbf{t}) \right| \right).
 \end{aligned}$$

Taking in consideration the hypotheses (A.1.)-(A.3.) and (A.4.b), we get $\forall \mathbf{h} \in \prod_{j=1}^m (a_{n,j}, b_j)$ and $\forall \mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m$:

$$\begin{aligned}
 |\kappa_1^{-m} C_3^{-1} \mathbb{E}(u_n(\varphi, \mathbf{t}, \mathbf{h}) - r^{(m)}(\varphi, \mathbf{t}))| & \leq \frac{C_5}{\phi_{\mathbf{t}}(\mathbf{h}) \kappa_1^m C_3} \mathbb{E} \left(\prod_{i=1}^m \mathbb{K} \left(\frac{d(X_i, t_i)}{h_i} \right) d_{\mathcal{X}^m}^Y(\mathbf{X}, \mathbf{t}) \right) \\
 & \leq \frac{C_5}{m} (d(X_1, t_1) + \dots + d(X_m, t_m))^Y \\
 & \leq \frac{C_5}{m} (h_1 + \dots + h_m)^Y \\
 & = \frac{C_5}{m} (b_1 + \dots + b_m)^Y \leq C_5' \tilde{b}_0^Y,
 \end{aligned}$$

where

$$\tilde{b}_0 := \max_{1 \leq j \leq m} b_j.$$

This completes the proof of the theorem. \square

Proof of Corollary 3.4.2.5

In this section we will prove Corollary 3.4.2.5 using Lemma 3.9.1.1. Following the same reasoning as the case of the functional regression, we use the notations: $S_{\Omega} = \mathcal{S}_{\mathcal{X}}$, $\mathbf{A}_i = \mathbf{X}_i$, $\varphi(\mathbf{B}_i) = \varphi(\mathbf{Y}_i)$,

$$\begin{aligned}
 \prod_{i=1}^m G(H_i, (t, \mathbf{A}_i)) & = \prod_{i=1}^m \mathbb{K}(H_i^{-1} d(t, \mathbf{A}_i)), \\
 D_{n,k}(t_j) & = H_{n,k}(t_j), \quad \forall \mathbf{t} = (t_1, \dots, t_m) \in \mathcal{S}_{\mathcal{X}}^m, \text{ and } j = 1, \dots, m, \\
 \mathfrak{M}_n^{(m)}(\varphi, t; \mathbf{h}_{n,k}(\mathbf{t})) & = \hat{r}_n^{*(m)}(\varphi, t; \mathbf{h}_{n,k}(\mathbf{t})), \\
 \mathfrak{M}(\varphi, t) & = r^{(m)}(\varphi, t).
 \end{aligned}$$

Choosing $D_{n,k}^-(\beta_n, t)$ and $D_{n,k}^+(\beta_n, t)$ such that for all $j = 1, \dots, m$,

$$\begin{aligned}\Phi_{t_j}(D_{n,k}^-(\beta_n, t_j)) &= \frac{\sqrt{\beta_{n,j}}k}{n}, \\ \Phi_{t_j}(D_{n,k}^+(\beta_n, t_j)) &= \frac{k}{n\sqrt{\beta_{n,j}}},\end{aligned}$$

where $\beta_{n,j}$ are increasing sequences that belong to $(0, 1)$ and

$$\beta_n = \prod_{j=1}^m \beta_{n,j}.$$

We denote $\mathbf{h}^-(\mathbf{t}) = (h_1^-(t_1), \dots, h_m^-(t_m))$ and $\mathbf{h}^+(\mathbf{t}) = (h_1^+(t_1), \dots, h_m^+(t_m))$, where

$$h_j^-(t_j) = D_{n,k}^-(\beta_n, t_j) \text{ and } h_j^+(t_j) = D_{n,k}^+(\beta_n, t_j), \text{ for all } j = 1, \dots, m.$$

We can easily see that, for all $j = 1, \dots, m$:

$$\Phi_{t_j}^{-1}\left(\frac{\sqrt{\beta_{n,j}}k_{1,n}}{n}\right) \leq h^-(t_j) \leq \Phi_{t_j}^{-1}\left(\frac{\sqrt{\beta_{n,j}}k_{2,n}}{n}\right), \quad (3.9.94)$$

$$\Phi_{t_j}^{-1}\left(\frac{k_{1,n}}{n\sqrt{\beta_{n,j}}}\right) \leq h^+(t_j) \leq \Phi_{t_j}^{-1}\left(\frac{k_{2,n}}{n\sqrt{\beta_{n,j}}}\right). \quad (3.9.95)$$

Using the condition (3.2.12) one gets, for all $j = 1, \dots, m$, there exist constants $0 < \mu_j \leq \nu_j < \infty$ such that

$$\mu_j \Phi^{-1}\left(\frac{\rho_{n,j}k_{1,n}}{n}\right) \leq \Phi_{t_j}^{-1}\left(\frac{\rho_{n,j}k_{1,n}}{n}\right) \text{ and } \Phi_{t_j}^{-1}\left(\frac{k_{2,n}}{\rho_{n,j}n}\right) \leq \nu_j \Phi^{-1}\left(\frac{k_{2,n}}{\rho_{n,j}n}\right),$$

we put $\rho_{n,j} = \sqrt{\beta_{n,j}}$, $a_{n,j} = \mu_j \Phi^{-1}\left(\frac{\rho_{n,j}k_{1,n}}{n}\right)$ and $b_{n,j} = \nu_j \Phi^{-1}\left(\frac{k_{2,n}}{\rho_{n,j}n}\right)$, thus $\mathbf{h}^-(\mathbf{t})$ and $\mathbf{h}^+(\mathbf{t})$ belong to the interval

$$\mathcal{H}_n^{(m)} := \prod_{j=1}^m (a_{n,j}, b_{n,j}).$$

We denote $\tilde{a}_n = \min_{1 \leq j \leq m} a_{n,j}$ and $\tilde{b}_n = \max_{1 \leq j \leq m} b_{n,j}$, therefore

$$h_j(t_j) \in (\tilde{a}_n, \tilde{b}_n), \forall j = 1, \dots, m.$$

We also note for all $\mathbf{b} = (b_1, \dots, b_m) \in (0, 1)^m$:

$$\mathcal{H}_0^{(m)} := \prod_{j=1}^m (a_{n,j}, b_j),$$

and

$$\tilde{b}_0 := \max_{1 \leq j \leq m} b_j.$$

Finally, we can choose constants $0 < \mu^* < \nu^* < \infty$ and a sequence $\{\rho_n^*\} \in (0, 1)$, while respecting the condition (A.7.), in a way that makes $\tilde{a}_n = \mu^* \phi^{-1}\left(\frac{\rho_n^* k_{1,n}}{n}\right)$, $\tilde{b}_n = \nu^* \phi^{-1}\left(\frac{k_{2,n}}{\rho_n^* n}\right)$ and

$$\mathcal{U}_n = \phi^{-1}\left(\frac{k_{2,n}}{\rho_n^* n}\right)^Y + \sqrt{\frac{\Psi_{\mathcal{S}_X}\left(\frac{\log n}{n}\right)}{n\phi\left(\mu^* \phi^{-1}\left(\frac{\rho_n^* k_{1,n}}{n}\right)\right)}}.$$

It is clear that (L₀) is satisfied due to condition (A.4.b), and from (3.9.94) and (3.9.95), we can easily verify that the construction of $\mathbf{h}^-(\mathbf{t})$ and $\mathbf{h}^+(\mathbf{t})$ satisfies the condition (L₁).

Checking the conditions (L₄) and (L₅)

A direct application of Corollary 3.4.1.11 gives

$$\begin{aligned} & \sup_{\varphi \tilde{\mathbf{K}} \in \mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{h}^-(\mathbf{t}) \in \mathcal{H}_n^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_X^m} \left| \mathfrak{M}_n(\varphi, \mathbf{t}; \mathbf{D}_{n,k}^-(\beta_n, \mathbf{t})) - \mathfrak{M}(\varphi, \mathbf{t}) \right| \\ & = O(\tilde{b}_n^Y) + O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_X}\left(\frac{\log n}{n}\right)}{n\phi(\tilde{a}_n)}} \right), \end{aligned}$$

which is equivalent to

$$\begin{aligned} & \sup_{\varphi \tilde{\mathbf{K}} \in \mathcal{F}_m \mathcal{K}^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{\mathbf{t} \in \mathcal{S}_X^m} \left| \mathfrak{M}_n(\varphi, \mathbf{t}; \mathbf{D}_{n,k}^-(\beta_n, \mathbf{t})) - \mathfrak{M}(\varphi, \mathbf{t}) \right| \\ & = O_{a.co} \left\{ \phi^{-1}\left(\frac{k_{2,n}}{\rho_n^* n}\right)^Y + \sqrt{\frac{\Psi_{\mathcal{S}_X}\left(\frac{\log n}{n}\right)}{n\phi\left(\mu^* \phi_t^{-1}\left(\frac{\rho_n^* k_{1,n}}{n}\right)\right)}} \right\}. \end{aligned} \quad (3.9.96)$$

Applying the same reasoning with $\mathbf{h}^+(\mathbf{t})$ we obtain

$$\begin{aligned} & \sup_{\varphi \tilde{\mathbf{K}} \in \mathcal{F}_m \mathcal{K}^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{\mathbf{t} \in \mathcal{S}_X^m} \left| \mathfrak{M}_n(\varphi, \mathbf{t}; \mathbf{D}_{n,k}^+(\beta_n, \mathbf{t})) - \mathfrak{M}(\varphi, \mathbf{t}) \right| \\ & = O_{a.co} \left\{ \phi^{-1}\left(\frac{k_{2,n}}{\rho_n^* n}\right)^Y + \sqrt{\frac{\Psi_{\mathcal{S}_X}\left(\frac{\log n}{n}\right)}{n\phi\left(\mu^* \phi_t^{-1}\left(\frac{\rho_n^* k_{1,n}}{n}\right)\right)}} \right\}. \end{aligned} \quad (3.9.97)$$

Thus the conditions (L₄) and (L₅) are checked.

Checking the condition (L₂)

To check (L₂) we show that for all $\varepsilon_0 > 0$ and $\mathbf{t} = (t_1, \dots, t_m) \in \mathcal{S}_{\mathcal{X}}^m$:

$$\sum_{n \geq 1} \mathbb{P} \left\{ \left| \prod_{j=1}^m \mathbb{1} \left\{ D_{n,k}^-(\beta_n, t_j) \leq H_{n,k}(t_j) \leq D_{n,k}^+(\beta_n, t_j) \right\} - 1 \right| > \varepsilon_0 \right\} < \infty. \quad (3.9.98)$$

We have

$$\begin{aligned} & \mathbb{P} \left\{ \left| \prod_{j=1}^m \mathbb{1} \left\{ D_{n,k}^-(\beta_n, t_j) \leq H_{n,k}(t_j) \leq D_{n,k}^+(\beta_n, t_j) \right\} - 1 \right| > \varepsilon_0 \right\} \\ &= \sum_{j=1}^m \mathbb{P} \left\{ H_{n,k}(t_j) \notin \left(D_{n,k}^-(\beta_n, t_j), D_{n,k}^+(\beta_n, t_j) \right) \right\} \\ &\leq \sum_{j=1}^m \mathbb{P} \left(H_{n,k}(t_j) \leq \Phi_{t_j}^{-1} \left(\frac{\sqrt{\beta_{n,j}} k_{1,n}}{n} \right) \right) + \sum_{j=1}^m \mathbb{P} \left(H_{n,k}(t_j) \geq \Phi_{t_j}^{-1} \left(\frac{k_{2,n}}{\sqrt{\beta_{n,j}} n} \right) \right) \\ &\leq \sum_{\ell=1}^{N_{\varepsilon}(\mathcal{S}_{\mathcal{X}})} \sum_{k=k_{1,n}}^{k_{2,n}} \sum_{j=1}^m \mathbb{P} \left(H_{n,k}(t_{\ell}(t_j)) \leq \Phi_{t_{\ell}(t_j)}^{-1} \left(\frac{\sqrt{\beta_{n,j}} k_{1,n}}{n} \right) \right) \\ &\quad + \sum_{\ell=1}^{N_{\varepsilon}(\mathcal{S}_{\mathcal{X}})} \sum_{k=k_{1,n}}^{k_{2,n}} \sum_{j=1}^m \mathbb{P} \left(H_{n,k}(t_{\ell}(t_j)) \geq \Phi_{t_{\ell}(t_j)}^{-1} \left(\frac{k_{2,n}}{\sqrt{\beta_{n,j}} n} \right) \right), \end{aligned}$$

now, using (3.9.40) and (3.9.41), we get

$$\mathbb{P} \left\{ H_{n,k}(t_{\ell}(t_j)) \leq \Phi_{t_{\ell}(t_j)}^{-1} \left(\frac{\alpha k_{1,n}}{n} \right) \right\} \leq \exp \left\{ -\frac{k - \alpha k_{1,n}}{4} \right\}, \quad (3.9.99)$$

and

$$\mathbb{P} \left\{ H_{n,k}(t_{\ell}(t_j)) \geq \Phi_{t_{\ell}(t_j)}^{-1} \left(\frac{k_{2,n}}{\alpha n} \right) \right\} \leq \exp \left\{ -\frac{(k_{2,n} - \alpha k)^2}{2\alpha k_{2,n}} \right\}. \quad (3.9.100)$$

Consequently, we obtain

$$\begin{aligned} \sum_{\ell=1}^{N_{\varepsilon}(\mathcal{S}_{\mathcal{X}})} \sum_{k=k_{1,n}}^{k_{2,n}} \sum_{j=1}^m \mathbb{P} \left\{ H_{n,k}(t_{\ell}(t_j)) \leq \Phi_{t_{\ell}(t_j)}^{-1} \left(\frac{\alpha k_{1,n}}{n} \right) \right\} &\leq m N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}}) k_{2,n} \exp \{ -(1 - \alpha) k_{1,n} / 4 \} \\ &\leq m N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}}) n^{1 - \{(1 - \alpha) / 4\} \frac{k_{1,n}}{\ln n}} < \infty \end{aligned} \quad (3.9.101)$$

and

$$\begin{aligned} \sum_{\ell=1}^{N_{\varepsilon}(\mathcal{S}_{\mathcal{X}})} \sum_{k=k_{1,n}}^{k_{2,n}} \sum_{j=1}^m \mathbb{P} \left(H_{n,k}(t_{\ell}(t_j)) \geq \Phi_{t_{\ell}(t_j)}^{-1} \left(\frac{k_{2,n}}{\sqrt{\beta_{n,j}} n} \right) \right) &\leq N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}}) k_{2,n} \exp \{ -(1 - \alpha) k_{1,n} / 2\alpha \} \\ &\leq N_{\varepsilon_n}(\mathcal{S}_{\mathcal{X}}) n^{1 - \{(1 - \alpha) / 2\alpha\} \frac{k_{2,n}}{\ln n}} < \infty \end{aligned} \quad (3.9.102)$$

Thus (L₂) is checked.

Checking the condition (L₃)

Notice That

$$\begin{aligned}
 & |\mathfrak{M}_{n-}^{(m)}(\varphi, \mathbf{t}; \beta_n) - \beta_n \mathfrak{M}^{(m)}(\varphi, \mathbf{t})| \\
 &= \left| \frac{\sum_{i \in I(m,n)} \varphi(Y_{i_1}, \dots, Y_{i_m}) \prod_{j=1}^m \mathbb{K}\left(\frac{d(X_{i_j}, t_j)}{h_j^-(t_j)}\right)}{\sum_{i \in I(m,n)} \prod_{j=1}^m \mathbb{K}\left(\frac{d(X_{i_j}, t_j)}{h_j^+(t_j)}\right)} - \frac{\phi_{\mathbf{t}}(\mathbf{h}^-(\mathbf{t}))}{\phi_{\mathbf{t}}(\mathbf{h}^+(\mathbf{t}))} r^{(m)}(\varphi, \mathbf{t}) \right| \\
 &= \frac{\phi_{\mathbf{t}}(\mathbf{h}^-(\mathbf{t}))}{\phi_{\mathbf{t}}(\mathbf{h}^+(\mathbf{t}))} \left| \frac{\phi_{\mathbf{t}}(\mathbf{h}^+(\mathbf{t}))}{\phi_{\mathbf{t}}(\mathbf{h}^-(\mathbf{t}))} \frac{\sum_{i \in I(m,n)} \varphi(Y_{i_1}, \dots, Y_{i_m}) \prod_{j=1}^m \mathbb{K}\left(\frac{d(X_{i_j}, t_j)}{h_j^-(t_j)}\right)}{\sum_{i \in I(m,n)} \prod_{j=1}^m \mathbb{K}\left(\frac{d(X_{i_j}, t_j)}{h_j^+(t_j)}\right)} - r^{(m)}(\varphi, \mathbf{t}) \right| \\
 &= \beta_n \left| \frac{u_n(\varphi, \mathbf{t}; \mathbf{h}^-(\mathbf{t}))}{u_n(1, \mathbf{t}; \mathbf{h}^+(\mathbf{t}))} - r^{(m)}(\varphi, \mathbf{t}) \right| \tag{3.9.103}
 \end{aligned}$$

The study of (3.9.103) is similar to the proofs of Theorem 3.4.1.9 and Theorem 3.4.1.10, as we can clearly see that

$$\left| \frac{u_n(\varphi, \mathbf{t}; \mathbf{h}^-(\mathbf{t}))}{u_n(1, \mathbf{t}; \mathbf{h}^+(\mathbf{t}))} - r^{(m)}(\varphi, \mathbf{t}) \right| \leq \left| \frac{u_n(\varphi, \mathbf{t}; \mathbf{h}^-(\mathbf{t}))}{u_n(1, \mathbf{t}; \mathbf{h}^+(\mathbf{t}))} - \frac{\mathbb{E}[u_n(\varphi, \mathbf{t}; \mathbf{h}^-(\mathbf{t}))]}{\mathbb{E}[u_n(1, \mathbf{t}; \mathbf{h}^+(\mathbf{t}))]} \right| \tag{3.9.104}$$

$$+ \left| \frac{\mathbb{E}[u_n(\varphi, \mathbf{t}; \mathbf{h}^-(\mathbf{t}))]}{\mathbb{E}[u_n(1, \mathbf{t}; \mathbf{h}^+(\mathbf{t}))]} - r^{(m)}(\varphi, \mathbf{t}) \right|. \tag{3.9.105}$$

Let us start with (3.9.104), we have

$$\begin{aligned}
 & \left| \frac{u_n(\varphi, \mathbf{t}; \mathbf{h}^-(\mathbf{t}))}{u_n(1, \mathbf{t}; \mathbf{h}^+(\mathbf{t}))} - \frac{\mathbb{E}[u_n(\varphi, \mathbf{t}; \mathbf{h}^-(\mathbf{t}))]}{\mathbb{E}[u_n(1, \mathbf{t}; \mathbf{h}^+(\mathbf{t}))]} \right| \\
 & \leq \frac{|u_n(\varphi, \mathbf{t}; \mathbf{h}^-(\mathbf{t})) - \mathbb{E}(u_n(\varphi, \mathbf{t}; \mathbf{h}^-(\mathbf{t})))|}{|u_n(1, \mathbf{t}; \mathbf{h}^+(\mathbf{t}))|} \\
 & \quad + \frac{|\mathbb{E}(u_n(\varphi, \mathbf{t}; \mathbf{h}^-(\mathbf{t})))| \cdot |u_n(1, \mathbf{t}; \mathbf{h}^+(\mathbf{t})) - \mathbb{E}(u_n(1, \mathbf{t}; \mathbf{h}^+(\mathbf{t})))|}{|u_n(1, \mathbf{t}; \mathbf{h}^+(\mathbf{t}))| \cdot |\mathbb{E}(u_n(1, \mathbf{t}; \mathbf{h}^+(\mathbf{t})))|} \\
 & =: \text{I} + \text{II}.
 \end{aligned}$$

Applying the same calculation as in the proof of Theorem 3.4.1.9, it follows that:

$$\begin{aligned}
 & \sup_{\varphi \in \mathcal{F}_m} \sup_{\mathcal{K}^m(\mathbf{h}^-(\mathbf{t}), \mathbf{h}^+(\mathbf{t})) \in \mathcal{H}_0^{(m)} \times \mathcal{H}_0^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \left| \frac{u_n(\varphi, \mathbf{t}; \mathbf{h}^-(\mathbf{t}))}{u_n(1, \mathbf{t}; \mathbf{h}^+(\mathbf{t}))} - r^{(m)}(\varphi, \mathbf{t}) \right| \\
 & = O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}{n \phi(\tilde{a}_n)}} \right).
 \end{aligned}$$

Now, for (3.9.105), using the fact that

$$\mathbb{E}[u_n(1, \mathbf{t}; \mathbf{h}^+(\mathbf{t}))] \geq \kappa_1^m \frac{\phi_{\mathbf{t}}(\mathbf{h}^+(\mathbf{t})/2)}{\phi_{\mathbf{t}}(\mathbf{h}^+(\mathbf{t}))} \geq \kappa_1^m C_3,$$

therefore,

$$\left| \frac{u_n(\varphi, \mathbf{t}; \mathbf{h}^-(\mathbf{t}))}{u_n(\mathbf{1}, \mathbf{t}; \mathbf{h}^+(\mathbf{t}))} - r^{(m)}(\varphi, \mathbf{t}) \right| \leq \kappa_1^{-m} C_3^{-1} |u_n(\varphi, \mathbf{t}; \mathbf{h}^-(\mathbf{t})) - r^{(m)}(\varphi, \mathbf{t})|,$$

following the same steps as in Theorem 3.4.1.10, we can easily conclude that

$$\sup_{\varphi \tilde{\mathbf{K}} \in \mathcal{F}_m \mathcal{K}^m} \sup_{(\mathbf{h}^-(\mathbf{t}), \mathbf{h}^+(\mathbf{t})) \in \mathcal{H}_n^{(m)} \times \mathcal{H}_n^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \left| \frac{\mathbb{E}[u_n(\varphi, \mathbf{t}; \mathbf{h}^-(\mathbf{t}))]}{\mathbb{E}[u_n(\mathbf{1}, \mathbf{t}; \mathbf{h}^+(\mathbf{t}))]} - r^{(m)}(\varphi, \mathbf{t}) \right| = O(\tilde{b}_n^Y).$$

Consequently,

$$\begin{aligned} & \sup_{\varphi \tilde{\mathbf{K}} \in \mathcal{F}_m \mathcal{K}^m} \sup_{(\mathbf{h}^-(\mathbf{t}), \mathbf{h}^+(\mathbf{t})) \in \mathcal{H}_n^{(m)} \times \mathcal{H}_n^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \left| \frac{u_n(\varphi, \mathbf{t}; \mathbf{h}^-(\mathbf{t}))}{u_n(\mathbf{1}, \mathbf{t}; \mathbf{h}^+(\mathbf{t}))} - r^{(m)}(\varphi, \mathbf{t}) \right| \\ &= O(\tilde{b}_n^Y) + O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}{n\phi(\tilde{a}_n)}} \right). \end{aligned}$$

Since $\beta_n \rightarrow 1$, we can also conclude that

$$\begin{aligned} & \sup_{\varphi \tilde{\mathbf{K}} \in \mathcal{F}_m \mathcal{K}^m} \sup_{(\mathbf{h}^-(\mathbf{t}), \mathbf{h}^+(\mathbf{t})) \in \mathcal{H}_0^{(m)} \times \mathcal{H}_0^{(m)}} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \left| \frac{\sum_{i \in \mathbb{I}(m,n)} \varphi(Y_{i_1}, \dots, Y_{i_m}) \prod_{j=1}^m \mathbf{K} \left(\frac{d(X_{i_j}, t_j)}{h_j^-(t_j)} \right)}{\sum_{i \in \mathbb{I}(m,n)} \prod_{j=1}^m \mathbf{K} \left(\frac{d(X_{i_j}, t_j)}{h_j^+(t_j)} \right)} - \beta_n r^{(m)}(\varphi, \mathbf{t}) \right| \\ &= O(\tilde{b}_n^Y) + O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}{n\phi(\tilde{a}_n)}} \right), \end{aligned}$$

which implies that

$$\begin{aligned} & \sup_{\varphi \tilde{\mathbf{K}} \in \mathcal{F}_m \mathcal{K}^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \left| \frac{\sum_{i \in \mathbb{I}(m,n)} \varphi(Y_{i_1}, \dots, Y_{i_m}) \prod_{j=1}^m \mathbf{K} \left(\frac{d(X_{i_j}, t_j)}{h_j^-(t_j)} \right)}{\sum_{i \in \mathbb{I}(m,n)} \prod_{j=1}^m \mathbf{K} \left(\frac{d(X_{i_j}, t_j)}{h_j^+(t_j)} \right)} - \beta_n r^{(m)}(\varphi, \mathbf{t}) \right| \\ &= O \left(\phi^{-1} \left(\frac{k_{2,n}}{\rho_n^* n} \right)^Y \right) + O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}{n\phi \left(\mu^* \phi_t^{-1} \left(\frac{\rho_n^* k_{1,n}}{n} \right) \right)}} \right). \end{aligned} \quad (3.9.106)$$

Finally, by putting $\varphi \equiv 1$ in (3.9.106), we get

$$\begin{aligned} & \sup_{\tilde{\mathbf{K}} \in \mathcal{K}^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{\mathbf{t} \in \mathcal{S}_{\mathcal{X}}^m} \left| \frac{\sum_{i \in I(m,n)} \prod_{j=1}^m \mathbb{K} \left(\frac{d(X_{i_j}, t_j)}{h_j^-(t_j)} \right)}{\sum_{i \in I(m,n)} \prod_{j=1}^m \mathbb{K} \left(\frac{d(X_{i_j}, t_j)}{h_j^+(t_j)} \right)} - \beta_n \right| \\ &= O \left(\phi^{-1} \left(\frac{k_{2,n}}{\rho_n^* n} \right)^\gamma \right) + O_{a.co} \left(\sqrt{\frac{\Psi_{\mathcal{S}_{\mathcal{X}}} \left(\frac{\log n}{n} \right)}{n \phi \left(\mu^* \phi_t^{-1} \left(\frac{\rho_n^* k_{1,n}}{n} \right) \right)}} \right). \end{aligned} \quad (3.9.107)$$

Hence, (L₃) is checked. Now, with all the conditions of Lemma 3.9.1 satisfied, it follows that

$$\sup_{\tilde{\mathbf{G}} \in \mathcal{F}_m \mathcal{G}^m} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{\mathbf{t} \in \mathcal{S}_{\Omega}^m} |\mathfrak{M}_n^{(m)}(\varphi, \mathbf{t}; \mathbf{D}_{n,k}(\mathbf{t})) - \mathfrak{M}^{(m)}(\varphi, \mathbf{t})| = O_{a.co}(\mathcal{U}_n),$$

which is exactly the desired result; hence, the proof is completed. \square

Appendix

In the sequel, we define X, X_1, \dots, X_n to be i.i.d random variables defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and taking values in some measurable space (Ψ, \mathcal{B}) , and \mathcal{F} to be a \mathbb{P} -measurable class of measurable functions with envelope function F , such that :

$$\mathbb{E}(F^2(X)) < \infty.$$

We further assume that \mathcal{F} has the following property:

- For any sequence of i.i.d. \mathcal{X} -valued random variables Z_1, Z_2, \dots it holds that

$$\mathbb{E} \left\| \sum_{i=1}^k \{g(Z_i) - \mathbb{E}g(Z_1)\} \right\|_{\mathcal{G}} \leq C_1 \sqrt{k} \|G(Z_1)\|_2, \quad 1 \leq k \leq n,$$

where $C_1 > 1$ is a constant depending on \mathcal{G} only.

Lemma 3.9.1.6 (Theorem 2.14.1 van der Vaart and Wellner [1996]) *For an empirical process $\alpha_n(f)$ indexed by the class of functions \mathcal{F} with the notation :*

$$\|\alpha_n(f)\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |\alpha_n(f)|$$

and $J(\delta, \mathcal{F})$ meaning $\sup_Q \int_0^\delta \sqrt{1 + \log N(\epsilon, \mathcal{F}, \|\cdot\|_{L_2(Q)})} d\epsilon$, we have, for $p > 1$,

$$\|\|\alpha_n(f)\|_{\mathcal{F}}\|_p \leq CJ(\delta, \mathcal{F}) \|F\|_{p \vee 2}.$$

Lemma 3.9.1.7 (Theorem 3.1 Dony and Einmahl [2009]) Let \mathcal{F} be a pointwise measurable function class satisfying the above assumptions. If we suppose that the empirical process $\alpha_n(f)$ satisfies :

$$\mathbb{E} \|\alpha_n(f)\|_{\mathcal{F}} \leq C \|F\|_2, \quad (3.9.108)$$

then for any measurable subset $B \in \mathcal{B}$:

$$\mathbb{E} \|\alpha_n(f \mathbf{1}_B)\|_{\mathcal{F}} \leq 2C \|F \mathbf{1}_B\|_2.$$

From Theorem 3.2 of Dony and Einmahl [2009], it follows that a VC-type class of functions satisfies, always, the condition (3.9.108).

Lemma 3.9.1.8 (Bernstein type inequality Fact 4.2 Dony and Einmahl [2009]) Assume that for some $H > 0$ and $p \geq 2$ the r.v.s X, X_1, \dots, X_n satisfy :

$$\mathbb{E}(F^p(X)) \leq (p!/2)\sigma^2 H^{p-2}, \quad \text{where} \quad \sigma^2 \geq \mathbb{E}(F^2(X)),$$

then for $\mu'_n = \mathbb{E}(\|\sqrt{n}\alpha_n(f)\|_{\mathcal{F}})$, we have for any $z > 0$:

$$\begin{aligned} \mathbb{P} \left\{ \max_{1 \leq k \leq n} \|\sqrt{k}\alpha_k(f)\|_{\mathcal{F}} \geq \mu'_n + z \right\} &\leq \exp\left(\frac{-z^2}{2n\sigma^2 + 2zH}\right) \\ &\leq \exp\left(\frac{-z^2}{4n\sigma^2}\right) \vee \exp\left(\frac{-z}{4H}\right). \end{aligned}$$

Lemma 3.9.1.9 (Theorem A. page 201 Serfling [1980]) Let f be a symmetric function taking its variables from \mathcal{X}^m satisfying $\|f\|_{\infty} \leq c$,

$$\mathbb{E}f(X_1, \dots, X_m) = \theta,$$

and

$$\sigma^2 = \text{Var}(f(X_1, \dots, X_m)),$$

then for $t > 0$ and $n \geq m$, we have :

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

Lemma 3.9.1.10 (Proposition 4. Arcones [1995]) Let X_1, \dots, X_m be i.i.d. random variables with values in some measurable space (S, \mathcal{S}) . Let \mathcal{F} be a class of symmetric functions f from S^m satisfying some measurability conditions. Suppose that there exists a finite constant c_0 such that for each $x_j \in S$ we have :

$$\mathbb{E} \left\{ \left\| n^{1-m} \sum_{I_n^m} \zeta_{i_1} \zeta_{i_2} f(x_{i_1}, \dots, x_{i_m}) \right\|_{\mathcal{F}} \right\} \leq c_0,$$

and that there is a finite constant b such that $\sup_{f \in \mathcal{F}} |f(X)| \leq b$, a.s. Then for each $u > 0$:

$$\mathbb{P} \left\{ n^{1/2} \left\| \sum_{j=2}^m \binom{m}{j} u_n(\pi_{j,m} f) \right\|_{\mathcal{F}} \geq u \right\} \leq 2 \exp \left(-\frac{un^{1/2}}{2^{m+5} m^{m+1} b c_0} \right),$$

where the variables ζ_i are Rademacher variables.

Lemma 3.9.1.11 (Lemma 6.1 in Kara *et al.* [2017], p.186) Let X_1, \dots, X_m be independent Bernoulli random variables with $\mathbb{P}(X_i = 1) = p$, for all $i = 1, \dots, n$. Set $U = X_1 + \dots + X_m$ and $\mu = pn$. Then, for any $\omega \geq 1$, we have

$$\mathbb{P}(U \geq (1 + \omega)\mu) \leq \exp \left\{ \frac{-\mu \min(\omega^2, \omega)}{4} \right\},$$

and if $\omega \in (0, 1)$, we have

$$\mathbb{P}\{U \leq (1 - \omega)\mu\} \leq \exp \{-\mu(\omega^2/2)\}.$$

3.10 References

- Abrevaya, J. and Jiang, W. (2005). A nonparametric approach to measuring and testing curvature. *J. Bus. Econom. Statist.*, **23**(1), 1–19. [36](#)
- Almanjahie, I. M., Bouzebda, S., Chikr Elmezouar, Z., and Laksaci, A. (2022). The functional k NN estimator of the conditional expectile: uniform consistency in number of neighbors. *Stat. Risk Model.*, **38**(3-4), 47–63. [39](#)
- Aneiros, G., Cao, R., Fraiman, R., Genest, C., and Vieu, P. (2019). Recent advances in functional data analysis and high-dimensional statistics. *J. Multivariate Anal.*, **170**, 3–9. [38](#)
- Arcones, M. A. (1995). A Bernstein-type inequality for U-statistics and U-processes. *Statist. Probab. Lett.*, **22**(3), 239–247. [116](#)
- Arcones, M. A. and Giné, E. (1993). Limit theorems for U-processes. *Ann. Probab.*, **21**(3), 1494–1542. [36](#), [49](#), [97](#), [104](#)
- Arcones, M. A. and Wang, Y. (2006). Some new tests for normality based on U-processes. *Statist. Probab. Lett.*, **76**(1), 69–82. [36](#)
- Attouch, M., Laksaci, A., and Rafea, F. (2019). On the local linear estimate for functional regression: uniform in bandwidth consistency. *Comm. Statist. Theory Methods*, **48**(8), 1836–1853. [39](#)
- Bellet, A. and Habrard, A. (2015). Robustness and generalization for metric learning. *Neuro-computing*, **151**, 259–267. [62](#)
- Bellet, A., Habrard, A., and Sebban, M. (2013). A survey on metric learning for feature vectors and structured data. *CoRR*, **abs/1306.6709**. [61](#)
- Biau, G. and Devroye, L. (2015). *Lectures on the nearest neighbor method*. Springer Series in the Data Sciences. Springer, Cham. [41](#)

- Blum, J. R., Kiefer, J., and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.*, **32**, 485–498. [64](#)
- Bogachev, V. I. (1998). *Gaussian measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI. [54](#)
- Bosq, D. (2000). *Linear processes in function spaces*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York. Theory and applications. [38](#)
- Bosq, D. and Blanke, D. (2007). *Inference and prediction in large dimensions*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester; Dunod, Paris. [38](#)
- Bouzebda, S. (2012). On the strong approximation of bootstrapped empirical copula processes with applications. *Math. Methods Statist.*, **21**(3), 153–188. [39](#), [46](#)
- Bouzebda, S. and Chaouch, M. (2022). Uniform limit theorems for a class of conditional Z-estimators when covariates are functions. *J. Multivariate Anal.*, **189**, Paper No. 104872, 21. [55](#)
- 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**(4), 864–914. [39](#), [57](#)
- Bouzebda, S. and Elhatab, I. (2011). Uniform-in-bandwidth consistency for kernel-type estimators of Shannon’s entropy. *Electron. J. Stat.*, **5**, 440–459. [39](#)
- 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**(2), 452–509. [38](#), [39](#), [41](#), [43](#), [45](#), [50](#), [66](#), [70](#), [89](#), [90](#)
- Bouzebda, S., Elhatab, I., and Seck, C. T. (2018). Uniform in bandwidth consistency of nonparametric regression based on copula representation. *Statist. Probab. Lett.*, **137**, 173–182. [39](#)
- Bouzebda, S., Elhatab, I., and Nemouchi, B. (2021). On the uniform-in-bandwidth consistency of the general conditional u-statistics based on the copula representation. *Journal of Nonparametric Statistics*, **0**(0), 1–38. [38](#), [39](#)
- Brunel, E. and Comte, F. (2006). Adaptive nonparametric regression estimation in presence of right censoring. *Math. Methods Statist.*, **15**(3), 233–255. [56](#)
- Burba, F., Ferraty, F., and Vieu, P. (2009). k -nearest neighbour method in functional nonparametric regression. *J. Nonparametr. Stat.*, **21**(4), 453–469. [41](#), [50](#), [66](#), [67](#)
- Cao, Q., Guo, Z.-C., and Ying, Y. (2016). Generalization bounds for metric and similarity learning. *Mach. Learn.*, **102**(1), 115–132. [62](#)

- Carbonez, A., Györfi, L., and van der Meulen, E. C. (1995). Partitioning-estimates of a regression function under random censoring. *Statist. Decisions*, **13**(1), 21–37. [55](#)
- Chen, Y. and Datta, S. (2019). Adjustments of multi-sample U-statistics to right censored data and confounding covariates. *Comput. Statist. Data Anal.*, **135**, 1–14. [57](#)
- Cléménçon, S. (2011). On u-processes and clustering performance. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc. [36](#)
- Cléménçon, S., Colin, I., and Bellet, A. (2016). Scaling-up empirical risk minimization: optimization of incomplete U-statistics. *J. Mach. Learn. Res.*, **17**, Paper No. 76, 36. [61](#)
- Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *J. Statist. Plann. Inference*, **147**, 1–23. [38](#)
- 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**(4), 680–700. [57](#)
- Dauxois, J., Pousse, A., and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *J. Multivariate Anal.*, **12**(1), 136–154. [38](#)
- de la Peña, V. H. and Giné, E. (1999). *Decoupling*. Probability and its Applications (New York). Springer-Verlag, New York. From dependence to independence, Randomly stopped processes. U-statistics and processes. Martingales and beyond. [36](#)
- Deheuvels, P. (2011). One bootstrap suffices to generate sharp uniform bounds in functional estimation. *Kybernetika (Prague)*, **47**(6), 855–865. [46](#), [47](#)
- Deheuvels, P. and Mason, D. M. (2004). General asymptotic confidence bands based on kernel-type function estimators. *Stat. Inference Stoch. Process.*, **7**(3), 225–277. [39](#)
- Dony, J. and Einmahl, U. (2009). Uniform in bandwidth consistency of kernel regression estimators at a fixed point. In *High dimensional probability V: the Luminy volume*, volume 5 of *Inst. Math. Stat. (IMS) Collect.*, pages 308–325. Inst. Math. Statist., Beachwood, OH. [39](#), [116](#)
- Dony, J. and Mason, D. M. (2008). Uniform in bandwidth consistency of conditional U-statistics. *Bernoulli*, **14**(4), 1108–1133. [38](#), [64](#)
- Dudley, R. M. (1967). The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Functional Analysis*, **1**, 290–330. [46](#)

- Dudley, R. M. (1984). A course on empirical processes. In *École d'été de probabilités de Saint-Flour, XII—1982*, volume 1097 of *Lecture Notes in Math.*, pages 1–142. Springer, Berlin. [59](#)
- Dudley, R. M. (2014). *Uniform central limit theorems*, volume 142 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, New York, second edition. [46](#)
- Dwass, M. (1956). The large-sample power of rank order tests in the two-sample problem. *Ann. Math. Statist.*, **27**, 352–374. [60](#)
- Einmahl, U. and Mason, D. M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoret. Probab.*, **13**(1), 1–37. [39](#)
- Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.*, **33**(3), 1380–1403. [39](#), [46](#)
- Ezzahrioui, M. and Ould-Saïd, E. (2008). Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data. *J. Nonparametr. Stat.*, **20**(1), 3–18. [89](#)
- Faivishevsky, L. and Goldberger, J. (2008). Ica based on a smooth estimation of the differential entropy. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems*, volume 21. Curran Associates, Inc. [36](#)
- Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis*. Springer Series in Statistics. Springer, New York. Theory and practice. [38](#), [42](#), [47](#), [74](#), [82](#)
- Ferraty, F., Mas, A., and Vieu, P. (2007). Nonparametric regression on functional data: inference and practical aspects. *Aust. N. Z. J. Stat.*, **49**(3), 267–286. [54](#)
- Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2010a). Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference*, **140**(2), 335–352. [39](#), [46](#)
- Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2010b). Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference*, **140**(2), 335–352. [46](#)
- Fix, E. and Hodges, J. L. (1951). Discriminatory analysis. nonparametric discrimination: Consistency properties. *USAF School of Aviation Medicine*, **57**(3), 238–247. [41](#)
- Fix, E. and Hodges, J. L. (1989). Discriminatory analysis. nonparametric discrimination: Consistency properties. *Int. Stat. Rev.*, **57**(3), 238–247. [41](#)
- Földes, A. and Rejtő, L. (1981). A LIL type result for the product limit estimator. *Z. Wahrsch. Verw. Gebiete*, **56**(1), 75–86. [58](#)

- Fu, K.-A. (2012). An application of U-statistics to nonparametric functional data analysis. *Comm. Statist. Theory Methods*, **41**(9), 1532–1542. [39](#)
- Gasser, T., Hall, P., and Presnell, B. (1998). Nonparametric estimation of the mode of a distribution of random curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **60**(4), 681–691. [42](#)
- Geenens, G. (2011). Curse of dimensionality and related issues in nonparametric functional regression. *Stat. Surv.*, **5**, 30–43. [38](#)
- Ghosal, S., Sen, A., and van der Vaart, A. W. (2000). Testing monotonicity of regression. *Ann. Statist.*, **28**(4), 1054–1082. [36](#)
- Giné, E. and Mason, D. M. (2007a). Laws of the iterated logarithm for the local U-statistic process. *J. Theoret. Probab.*, **20**(3), 457–485. [36](#), [38](#)
- Giné, E. and Mason, D. M. (2007b). On local U-statistic processes and the estimation of densities of functions of several sample variables. *Ann. Statist.*, **35**(3), 1105–1145. [36](#)
- Goia, A. and Vieu, P. (2016). An introduction to recent advances in high/infinite dimensional statistics [Editorial]. *J. Multivariate Anal.*, **146**, 1–6. [38](#)
- Hall, P. (1984). Asymptotic properties of integrated square error and cross-validation for kernel estimation of a regression function. *Z. Wahrsch. Verw. Gebiete*, **67**(2), 175–196. [64](#)
- Halmos, P. R. (1946). The theory of unbiased estimation. *Ann. Math. Statistics*, **17**, 34–43. [36](#), [37](#)
- Härdle, W. and Marron, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.*, **13**(4), 1465–1481. [64](#)
- Harel, M. and Puri, M. L. (1996). Conditional U-statistics for dependent random variables. *J. Multivariate Anal.*, **57**(1), 84–100. [37](#)
- Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics*, **19**, 293–325. [36](#), [37](#), [60](#), [63](#), [95](#)
- Hollander, M. and Proschan, F. (1972). Testing whether new is better than used. *Ann. Math. Statist.*, **43**, 1136–1146. [63](#)
- Horváth, L. and Kokoszka, P. (2012). *Inference for functional data with applications*. Springer Series in Statistics. Springer, New York. [38](#)
- Horváth, L. and Rice, G. (2015). An introduction to functional data analysis and a principal component approach for testing the equality of mean curves. *Rev. Mat. Complut.*, **28**(3), 505–548. [38](#)

- Hsing, T. and Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester. 38
- Jadhav, S. and Ma, S. (2019). Kendall's tau for functional data analysis. 63
- Jin, R., Wang, S., and Zhou, Y. (2009). Regularized distance metric learning: theory and algorithm. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 22. Curran Associates, Inc. 62
- Joly, E. and Lugosi, G. (2016). Robust estimation of U-statistics. *Stochastic Process. Appl.*, **126**(12), 3760–3773. 36
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53**, 457–481. 56
- Kara, L.-Z., Laksaci, A., Rachdi, M., and Vieu, P. (2017). Data-driven k NN estimation in nonparametric functional data analysis. *J. Multivariate Anal.*, **153**, 176–188. 39, 76, 86, 117
- Kara-Zaitri, L., Laksaci, A., Rachdi, M., and Vieu, P. (2017). Uniform in bandwidth consistency for various kernel estimators involving functional data. *J. Nonparametr. Stat.*, **29**(1), 85–107. 39
- Kleffe, J. (1973). Principal components of random variables with values in a separable Hilbert space. *Math. Operationsforsch. Statist.*, **4**(5), 391–406. 38
- Kohler, M., Máthé, K., and Pintér, M. (2002). Prediction from randomly right censored data. *J. Multivariate Anal.*, **80**(1), 73–100. 55, 56
- Kolmogorov, A. N. and Tihomirov, V. M. (1959). ϵ -entropy and ϵ -capacity of sets in function spaces. *Uspehi Mat. Nauk*, **14**(2 (86)), 3–86. 46
- Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York. 46
- Kudraszow, N. L. and Vieu, P. (2013). Uniform consistency of k NN regressors for functional variables. *Statist. Probab. Lett.*, **83**(8), 1863–1870. 46, 50, 67, 89
- Lee, A. J. (1990). *U-statistics*, volume 110 of *Statistics: Textbooks and Monographs*. Marcel Dekker, Inc., New York. Theory and practice. 63
- Lee, S., Linton, O., and Whang, Y.-J. (2009). Testing for stochastic monotonicity. *Econometrica*, **77**(2), 585–602. 36
- Lehmann, E. L. (1951). A general concept of unbiasedness. *Ann. Math. Statistics*, **22**, 587–592. 60

- Li, W. V. and Shao, Q.-M. (2001). Gaussian processes: inequalities, small ball probabilities and applications. In *Stochastic processes: theory and methods*, volume 19 of *Handbook of Statist.*, pages 533–597. North-Holland, Amsterdam. [54](#)
- Ling, N. and Vieu, P. (2018). Nonparametric modelling for functional data: selected survey and tracks for future. *Statistics*, **52**(4), 934–949. [38](#)
- Ling, N., Liu, Y., and Vieu, P. (2017). On asymptotic properties of functional conditional mode estimation with both stationary ergodic and responses MAR. In *Functional statistics and related fields*, *Contrib. Stat.*, pages 173–178. Springer, Cham. [64](#)
- Ling, N., Meng, S., and Vieu, P. (2019). Uniform consistency rate of *k*NN regression estimation for functional time series data. *J. Nonparametr. Stat.*, **31**(2), 451–468. [39](#)
- Ling, N., Wang, L., and Vieu, P. (2020). Convergence rate of kernel regression estimation for time series data when both response and covariate are functional. *Metrika*, **83**(6), 713–732. [64](#)
- Liu, Q., Lee, J., and Jordan, M. (2016). A kernelized stein discrepancy for goodness-of-fit tests. In M. F. Balcan and K. Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 276–284, New York, New York, USA. PMLR. [36](#)
- Loftsgaarden, D. O. and Quesenberry, C. P. (1965). A nonparametric estimate of a multivariate density function. *Ann. Math. Statist.*, **36**, 1049–1051. [41](#)
- 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 Statist.*, **18**(2), 159–184. [55](#), [56](#)
- Masry, E. (2005). Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Process. Appl.*, **115**(1), 155–177. [42](#)
- Mayer-Wolf, E. and Zeitouni, O. (1993). The probability of small Gaussian ellipsoids and associated conditional moments. *Ann. Probab.*, **21**(1), 14–24. [54](#)
- Müller, H.-G. (2016). Peter Hall, functional data analysis and random objects. *Ann. Statist.*, **44**(5), 1867–1887. [38](#)
- Müller, H.-G. and Yao, F. (2008). Functional additive models. *J. Amer. Statist. Assoc.*, **103**(484), 1534–1544. [64](#)
- Nadaraja, E. A. (1964). On a regression estimate. *Teor. Veroyatnost. i Primenen.*, **9**, 157–159. [37](#)

- Nagy, S. (2017). An overview of consistency results for depth functionals. In *Functional statistics and related fields*, Contrib. Stat., pages 189–196. Springer, Cham. 38
- Nolan, D. and Pollard, D. (1987). U-processes: rates of convergence. *Ann. Statist.*, **15**(2), 780–799. 36, 46
- Novo, S., Aneiros, G., and Vieu, P. (2019). Automatic and location-adaptive estimation in functional single-index regression. *J. Nonparametr. Stat.*, **31**(2), 364–392. 39
- Peng, W., Coleman, T., and Mentch, L. (2019). Asymptotic distributions and rates of convergence for random forests via generalized u-statistics. 36
- Polonik, W. and Yao, Q. (2002). Set-indexed conditional empirical and quantile processes based on dependent data. *J. Multivariate Anal.*, **80**(2), 234–255. 59
- Prakasa Rao, B. L. S. and Sen, A. (1995). Limit distributions of conditional U-statistics. *J. Theoret. Probab.*, **8**(2), 261–301. 37
- Rachdi, M. and Vieu, P. (2007). Nonparametric regression for functional data: automatic smoothing parameter selection. *J. Statist. Plann. Inference*, **137**(9), 2784–2801. 64, 65
- Ramsay, J. O. and Dalzell, C. J. (1991). Some tools for functional data analysis. *J. Roy. Statist. Soc. Ser. B*, **53**(3), 539–572. With discussion and a reply by the authors. 64
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition. 38
- Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics*, **14**(1), 1–17. 38
- Schick, A., Wang, Y., and Wefelmeyer, W. (2011). Tests for normality based on density estimators of convolutions. *Statist. Probab. Lett.*, **81**(2), 337–343. 36
- Sen, A. (1994). Uniform strong consistency rates for conditional U-statistics. *Sankhyā Ser. A*, **56**(2), 179–194. 37
- Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. John Wiley & Sons, Inc., New York. Wiley Series in Probability and Mathematical Statistics. 62, 116
- Shang, H. L. (2014). A survey of functional principal component analysis. *AStA Adv. Stat. Anal.*, **98**(2), 121–142. 38
- Sherman, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica*, **61**(1), 123–137. 36
- Sherman, R. P. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Ann. Statist.*, **22**(1), 439–459. 36

- Shi, J. Q. and Choi, T. (2011). *Gaussian process regression analysis for functional data*. CRC Press, Boca Raton, FL. 38
- Stute, W. (1986). Conditional empirical processes. *Ann. Statist.*, **14**(2), 638–647. 62
- Stute, W. (1991). Conditional U-statistics. *Ann. Probab.*, **19**(2), 812–825. 37
- Stute, W. (1993). Almost sure representations of the product-limit estimator for truncated data. *Ann. Statist.*, **21**(1), 146–156. 36
- Stute, W. (1994a). L^p -convergence of conditional U-statistics. *J. Multivariate Anal.*, **51**(1), 71–82. 60
- Stute, W. (1994b). Universally consistent conditional U-statistics. *Ann. Statist.*, **22**(1), 460–473. 60
- Stute, W. (1996). Symmetrized NN-conditional U-statistics. In *Research developments in probability and statistics*, pages 231–237. VSP, Utrecht. 37
- Stute, W. and Wang, J.-L. (1993). Multi-sample U-statistics for censored data. *Scand. J. Statist.*, **20**(4), 369–374. 57
- v. Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statistics*, **18**, 309–348. 37
- van der Vaart, A. (2002). The statistical work of Lucien Le Cam. *Ann. Statist.*, **30**(3), 631–682. Dedicated to the memory of Lucien Le Cam. 46
- van der Vaart, A. and van Zanten, H. (2007). Bayesian inference with rescaled Gaussian process priors. *Electron. J. Stat.*, **1**, 433–448. 46
- van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics. 46, 115
- Vieu, P. (2018). On dimension reduction models for functional data. *Statist. Probab. Lett.*, **136**, 134–138. 38
- Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A*, **26**, 359–372. 37
- Yuan, A., Giurcanu, M., Luta, G., and Tan, M. T. (2017). U-statistics with conditional kernels for incomplete data models. *Ann. Inst. Statist. Math.*, **69**(2), 271–302. 57

Chapter 4

Asymptotic properties of conditional U-statistics using delta sequences

Ce chapitre développe le contenu d'un article publié [Communications in Statistics-Theory and Methods (2023). <https://doi.org/10.1080/03610926.2023.2179887>], mis en forme pour être inséré dans le présent manuscrit de thèse.

Title : *Asymptotic properties of conditional U-statistics using delta sequences*

Abstract

W. Stute [Ann. Probab. 19 (1991) 812–825] 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 :

$$r^{(k)}(\varphi, \mathbf{t}) := \mathbb{E}[\varphi(Y_1, \dots, Y_k) | (X_1, \dots, X_k) = \mathbf{t}], \text{ for } \mathbf{t} \in \mathbb{R}^{dk}.$$

This paper deals with a quite general nonparametric statistical curve estimation setting, including the Stute estimator as a particular case. The class of “delta sequence estimators” is defined and treated here. This class also includes the orthogonal series and histogram methods. The theoretical results concerning the exponential inequalities and the asymptotic normality established in this paper are (or will be) key tools for further functional estimation developments. As a by-product of our proofs, we state consistency results for the delta sequences conditional U-statistics estimator under the random censoring. Potential applications include discrimination problems, metric learning, multipartite ranking, Kendall rank correlation coefficient, generalized U-Statistics, and set indexed conditional U-statistics.

Keywords : Nonparametric estimation; U-statistics; conditional distribution; delta sequences; kernel estimation; machine learning problems.

Mathematics Subject Classification : 60F05 ; 60G15 ; 60G10 ; 62G05 ; 62G07; 62H 12.

4.1 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-statistics are useful for solving complex statistical problems. Examples are density estimation, non-parametric regression tests, and goodness-of-fit tests. In particular, U-statistics play a role in analyzing 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 analyze 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 \[2007a,b\]](#), [Schick et al. \[2011\]](#) introduced new tests for normality which use 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. \[2009a\]](#), [Ghosal et al. \[2000a\]](#), [Abrevaya and Jiang \[2005a\]](#)], cross-validation for density estimation [40], 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 \[2013\]](#). For excellent resource of references on the U-statistics and U-processes, the interested reader may refer to [Borovskikh \[1996\]](#), [Koroljuk and Borovskikh \[1994\]](#), [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 \[1999\]](#). For more recent references, we refer to [Bouzebda and Soukariéh \[2023b\]](#); [Soukariéh and Bouzebda \[2022, 2023\]](#). 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 \[1964\]](#) and [Watson \[1964\]](#)) estimates of a regression function.

To be more precise, let us consider the sequence of independent and identically distributed random elements $\{(\mathbf{X}_i, \mathbf{Y}_i), i \in \mathbb{N}^*\}$ with $\mathbf{X}_i \in \mathbb{R}^d$ and $\mathbf{Y}_i \in \mathbb{R}^{d'}$, $d, d' \geq 1$. Let $\varphi : \mathbb{R}^{d+k} \rightarrow \mathbb{R}$ be a measurable function. In this paper, we are primarily concerned with the estimation of the conditional expectation or regression function :

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

whenever it exists, i.e., $\mathbb{E}(|\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k)|) < \infty$. We now introduce a kernel function $K : \mathbb{R}^d \rightarrow \mathbb{R}$

with support contained in $[-B, B]^d$, $B > 0$, satisfying :

$$\sup_{\mathbf{x} \in \mathbb{R}^d} |\mathbf{K}(\mathbf{x})| =: \kappa < \infty \quad \text{and} \quad \int \mathbf{K}(\mathbf{x}) d\mathbf{x} = 1. \quad (4.1.1)$$

Stute [1991] introduced a class of estimators for $r^{(k)}(\varphi, \mathbf{t})$, called conditional U-statistics, which is defined for each $\mathbf{t} \in \mathbb{R}^{dk}$ to be :

$$\widehat{r}_n^{(k)}(\varphi, \mathbf{t}; h_n) = \frac{\sum_{(i_1, \dots, i_k) \in I(k, n)} \varphi(Y_{i_1}, \dots, Y_{i_k}) \mathbf{K}\left(\frac{\mathbf{t}_1 - \mathbf{X}_{i_1}}{h_n}\right) \dots \mathbf{K}\left(\frac{\mathbf{t}_k - \mathbf{X}_{i_k}}{h_n}\right)}{\sum_{(i_1, \dots, i_k) \in I(k, n)} \mathbf{K}\left(\frac{\mathbf{x}_1 - \mathbf{X}_{i_1}}{h_n}\right) \dots \mathbf{K}\left(\frac{\mathbf{x}_k - \mathbf{X}_{i_k}}{h_n}\right)}, \quad (4.1.2)$$

where :

$$I(k, n) = \{\mathbf{i} = (i_1, \dots, i_k) : 0 \leq i_j \leq n \quad \text{and} \quad i_j \neq i_r \quad \text{if} \quad j \neq r\},$$

is the set of all k -tuples of different integers between 1 and n and $\{h_n\}_{n \geq 1}$ is a sequence of positive constants converging to zero at the rate $nh_n^{dk} \rightarrow \infty$. In the particular case $k = 1$, the $r^{(k)}(\varphi, \mathbf{t})$ reduces to $r^{(1)}(\varphi, \mathbf{t}) = \mathbb{E}(\varphi(\mathbf{Y})|\mathbf{X} = \mathbf{t})$ and Stute's estimator becomes the Nadaraya-Watson (**Nadaraja [1965]** and **Watson [1964]**) estimator of $r^{(1)}(\varphi, \mathbf{t})$ given by :

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

The work of **Sen [1994]** was devoted to estimating the rate of the uniform convergence in \mathbf{t} of $\widehat{r}_n^{(k)}(\varphi, \mathbf{t}; h_n)$ to $r^{(k)}(\varphi, \mathbf{t})$. In the paper of **Prakasa Rao and Sen [1995]**, the limit distributions of $\widehat{r}_n^{(k)}(\varphi, \mathbf{t}; h_n)$ 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 neighbor 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., $h_n, h_n \in [a_n, b_n]$ where $a_n < b_n \rightarrow 0$ at some specific rate) of $\widehat{r}_n^{(k)}(\varphi, \mathbf{t}; h_n)$. In addition, uniform consistency is also established over $\varphi \in \mathcal{F}$ for a suitably restricted class of functions \mathcal{F} , extended in **Bouzebda and Nemouchi [2019]**, **Bouzebda and Nemouchi [2020]**, **Bouzebda et al. [2021]**, **Bouzebda and Nezzal [2022]** and **Bouzebda et al. [2022]**. The main tool in their result is the use of the local conditional U-process investigated in **Giné and Mason [2007a]**. In the present paper, we will consider a general method for functional estimation by using the delta sequences. Delta sequences (also called “approximate identities” or “summability kernels”) appear in many branches of mathematics, but probably the most important applications are those in the theory of generalized functions. The basic use of delta sequences is the regularization of generalized functions. The proposed methods generalize several non-parametric estimation methods, including the kernel estimators given in (4.1.2) of **Stute [1991]**. To be more precise, the broad class of delta-sequence estimators includes the histogram estimators, Chentsov's projection estimators **Chentsov [1962]**,

and nearest-neighbor estimators, among others. Certain types of these sequences were already studied by [Watson and Leadbetter \[1964\]](#), who called them “ δ -function sequences”. They established, among other things, the asymptotic unbiasedness and the asymptotic variance of estimators based on them but did not consider convergence rates. [Winter \[1975\]](#) obtained the rate of strong consistency as well as the rate of asymptotic bias for estimators associated with delta sequences arising from the Fejér kernel of the Fourier series. The delta sequence method of density estimation of [Susarla and Walter \[1981\]](#) is extended to certain non-i.i.d. cases in [Prakasa Rao \[1978\]](#), where the observations are assumed to be sampled from a stationary Markov process. [Ahmad \[1981\]](#) considered the delta-sequence estimator for the marginal distribution of a strictly stationary stochastic process satisfying some mixing conditions. [Basu and Sahoo \[1985\]](#) studied the rate of local and global convergence rates of delta-sequence-type estimators of the density function, its derivative, and its mode. [Nolan and Marron \[1989\]](#) proved the uniform strong consistency of delta-sequence estimators. [Isogai \[1990\]](#) partially generalized the usual non-parametric estimators of a regression function using an estimator based on quasi-positive delta sequences. [Marron and Härdle \[1986\]](#) considered a general nonparametric statistical curve estimation setting called the class of “fractional delta sequence estimators”. [Cristóbal Cristóbal *et al.* \[1987\]](#) used the delta method to investigate the correlation model. [Prakasa Rao \[2010\]](#) considered the problem of estimation of density function by the method of delta sequences for functional data with values in an infinite dimensional separable Banach space; we may also refer to [Ouassou and Rachdi \[2012\]](#). The copula estimation using the delta sequences methods is considered in [Bouzebda \[2012\]](#). The problem of the nonparametric minimax estimation of a multivariate density at a given point by the delta sequences was investigated in [Belitser \[2000\]](#). [Walter \[1983\]](#) used the delta sequence to propose an important application to the classification problem of the value of the discrete random variable. In this functional context, the “delta sequence estimators” class is defined and discussed. [Bouzebda *et al.* \[2023\]](#) established the almost complete uniform convergence with the rates of these estimators under certain broad conditions.

The present study aims to introduce and investigate the delta sequence estimators for the conditional U-statistics. To our knowledge, this problem was open up to the present, and it gives the primary motivation for our paper.

The layout of the present article is structured as follows. Section [4.2](#) is devoted to introducing the delta sequences and the definitions we need in our work, where we introduce the new family of estimators. Section [4.3](#) gives the paper’s main results concerning the exponential inequalities and the limiting law. In Section [4.4](#), we provide an important application for the censored data setting that is of its own interest. In Section [4.5](#), we provide some applications including the discrimination in [§4.5.1](#), the metric learning in [§4.5.2](#), the multipartite ranking in [§4.5.3](#), the Ranking problems in [§4.5.4](#), Kendall rank correlation coefficient [§4.5.5](#), generalized U-Statistics [§4.5.6](#), set indexed conditional U-statistics [§4.5.7](#) and testing stochastic monotonicity problems in [§4.5.8](#). In Section [4.5.10](#), we collect some examples of classes of functions and examples of conditional U-statistics. In Section [4.6](#), we briefly discussed the bandwidth se-

lection criterion. Some concluding remarks and possible future developments are relegated to 6.2. To prevent interrupting the presentation flow, all proofs are gathered in Section 4.7. A few relevant technical results are given in Appendix 4.8.

4.2 Preliminaries and estimation procedure

Let $(\mathbf{X}, \mathbf{Y}), (\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_n, \mathbf{Y}_n)$ be independent and identically distributed random vectors with common joint density function $f_{\mathbf{X}, \mathbf{Y}} : \mathbb{R}^d \times \mathbb{R}^{d'} \rightarrow [0, \infty[$. Let $\varphi : \mathbb{R}^{d+k} \rightarrow \mathbb{R}$ be a measurable function. We are interested in the estimation of the conditional expectation or the regression function

$$r^{(k)}(\varphi, \tilde{\mathbf{x}}) = \mathbb{E}(\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \tilde{\mathbf{x}}), \text{ for } \tilde{\mathbf{x}} \in \mathbb{R}^{dk}, \quad (4.2.1)$$

whenever it exists, i.e. $\mathbb{E}(|\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k)|) < \infty$. To be more precise, we want to extend the results of conditional U-statistics presented in Stute [1991] using the delta sequences method, known to unify many estimation approaches. To make sense of the results and before we present a new estimate of $r^{(k)}(\varphi, \tilde{\mathbf{x}})$, the following definitions and notation are needed. Throughout this paper, any multivariate point will be written in bold. To avoid confusion, we note $\mathbf{x} = (x_1, \dots, x_d)$ for a generic point in the Euclidean space \mathbb{R}^d and we denote $\tilde{\mathbf{x}} := (\mathbf{x}_1, \dots, \mathbf{x}_k)$ a k -tuple of multivariate points $\mathbf{x}_i \in \mathbb{R}^d, 1 \leq i \leq k$. We will also be considering \mathcal{X} , a compact subset of \mathbb{R}^d .

4.2.1 Delta-sequences

For an open set $\mathcal{X} \subset \mathbb{R}^d$, let $\mathcal{C}_0^\infty(\mathcal{X})$ denote the space of infinitely differentiable functions on \mathcal{X} having compact support.

Definition 4.2.1.1 A sequence of functions $\{\delta_m\}$, in $L^\infty(\mathcal{X} \times \mathcal{X})$, is said to be a delta-sequence on \mathcal{X} if for each $\phi \in \mathcal{C}_0^\infty(\mathcal{X})$, and $\mathbf{x} \in \mathcal{X}$, as $m \rightarrow \infty$, we have

$$\int_{\mathcal{X}} \delta_m(\mathbf{x}, \mathbf{y}) \phi(\mathbf{y}) d\mathbf{y} \rightarrow \phi(\mathbf{x}). \quad (4.2.2)$$

Definition 4.2.1.2 A delta-sequence $\{\delta_m\}$ on \mathbb{R}^d is said to be of positive type if $\delta_m \geq 0$ and for any $\mathbf{x} \in \mathbb{R}^d$, we have,

- (i) $\int_{\mathbb{R}^d} \delta_m(\mathbf{x}, \mathbf{y}) d\mathbf{y} = 1$;
- (ii) $\sup_{r>0} r \left\{ \int_{|\mathbf{x}-\mathbf{y}|>r} \delta_m(\mathbf{x}, \mathbf{y}) d\mathbf{y} \right\} = O(m^{-1})$;
- (iii) $\sup_{\mathbf{x} \in \mathbb{R}^d} |\delta_m(\mathbf{x}, \cdot)| = O(m^d)$;

and for each $\eta > 0$,

- (iv) $\sup\{\delta_m(\mathbf{x}, \mathbf{y}) : |\mathbf{x} - \mathbf{y}| > \eta\} \rightarrow 0$.

Let $\delta_{m,j}(x_j, y_j)$ be a delta-sequence of positive type, for $j = 1, \dots, d$, then, from Proposition 2.2 [Susarla and Walter \[1981\]](#),

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

is a delta-sequence of positive type. The delta-sequence technique is a device to simultaneously study properties of estimates pertaining to several usual methods in nonparametric function estimation. Introduced by [Watson and Leadbetter \[1964\]](#), the delta-sequence method relative to the density estimation was studied among others by [Földes and Révész \[1974\]](#) and [Walter and Blum \[1979\]](#) whereas it was extended to the multivariate case by [Susarla and Walter \[1981\]](#).

Remark 4.2.1.3 For $(x, y) \in \mathbb{R}^2$, set

$$\delta_m(x, y) = mK(m(x - y)),$$

where $K(\cdot)$ is a positive bounded function, such that,

$$\int_{\mathbb{R}} K(x) dx = 1 \quad \text{and} \quad \lim_{x \rightarrow \infty} |x|K(x) = 0.$$

[Parzen \[1962\]](#) proved that $\delta_m(\cdot, \cdot)$ constitutes a delta-sequence.

For convenience, we present two examples of positive delta sequences from [Susarla and Walter \[1981\]](#).

Example 4.2.1.4 The multivariate Fejer kernel of Fourier series

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

where

$$\delta_m(x_i, y_i) = \frac{(\sin\{(m+1)(x_i - y_i)/2\})^2}{2\pi(\sin\{(x_i - y_i)/2\})^2} \mathbb{I}\{|y_i - x_i| \leq \pi\},$$

defines a positive delta-sequence, refer to [[Susarla and Walter, 1981, Example 2.2.](#)].

Example 4.2.1.5 Let us denote by \mathfrak{B} a fixed compact subset of \mathbb{R}^d . For each m , partition \mathfrak{B} into $\mathfrak{B}_{m,1}, \dots, \mathfrak{B}_{m,m^d}$ disjoint sets each of which has diameter $\leq b/m$, for some constant b depending on \mathfrak{B} . Then

$$\delta_m(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^{m^d} \mathbb{I}\{\mathbf{x} \in \mathfrak{B}_{m,j}\} \mathbb{I}\{\mathbf{y} \in \mathfrak{B}_{m,j}\},$$

is a positive delta-sequence fulfilling [Definition 4.2.1.2](#).

Here and elsewhere, $m_n \rightarrow \infty$, $m_n/n \rightarrow 0$ as $n \rightarrow \infty$.

4.2.2 The main estimate

To estimate the regression function $r^{(k)}(\varphi, \cdot)$ given in (4.2.1), we propose the following general family of estimates for each $\tilde{\mathbf{x}} \in \mathbb{R}^{dk}$, defined by

$$\hat{r}_n^{(k)}(\varphi, \tilde{\mathbf{x}}; m_n) = \begin{cases} \frac{\sum_{(i_1, \dots, i_k) \in I(k, n)} \varphi(\mathbf{Y}_{i_1}, \dots, \mathbf{Y}_{i_k}) \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k})}{\sum_{(i_1, \dots, i_k) \in I(k, n)} \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k})}, \\ \text{if } \sum_{(i_1, \dots, i_k) \in I(k, n)} \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k}) \neq 0, \\ \\ \frac{n!}{(n-k)!k!} \sum_{(i_1, \dots, i_k) \in I(k, n)} \varphi(\mathbf{Y}_{i_1}, \dots, \mathbf{Y}_{i_k}) \\ \text{if } \sum_{(i_1, \dots, i_k) \in I(k, n)} \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k}) = 0. \end{cases} \quad (4.2.3)$$

From now on, we shall use the following notation:

$$\mathbf{X}^{(k)} := (\mathbf{X}_1, \dots, \mathbf{X}_k) \in \mathbb{R}^{dk}, \quad \text{and } \mathbf{X}_i^{(k)} := (\mathbf{X}_{i_1}, \dots, \mathbf{X}_{i_k}) \in \mathbb{R}^{dk}, \quad \mathbf{i} \in I(k, n),$$

$$\mathbf{Y}^{(k)} := (\mathbf{Y}_1, \dots, \mathbf{Y}_k) \in \mathbb{R}^{d'k}, \quad \text{and } \mathbf{Y}_i^{(k)} := (\mathbf{Y}_{i_1}, \dots, \mathbf{Y}_{i_k}) \in \mathbb{R}^{d'k}, \quad \mathbf{i} \in I(k, n),$$

we also define

$$G_{\varphi, \tilde{\mathbf{x}}, m_n}(\mathbf{X}^{(k)}, \mathbf{Y}^{(k)}) = \varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \prod_{i=1}^k \delta_{m_n}(\mathbf{x}_i, \mathbf{X}_i),$$

and

$$u_n(\varphi, \tilde{\mathbf{x}}, m_n) := u_n^{(k)}(G_{\varphi, \tilde{\mathbf{x}}, m_n}) = \frac{(n-k)!}{n!} \sum_{\mathbf{i} \in I(k, n)} G_{\varphi, \tilde{\mathbf{x}}, m_n}(\mathbf{X}_i^{(k)}, \mathbf{Y}_i^{(k)}).$$

We can see that

$$\hat{r}_n^{(k)}(\varphi, \tilde{\mathbf{x}}, m_n) = \frac{u_n(\varphi, \tilde{\mathbf{x}}, m_n)}{u_n(1, \tilde{\mathbf{x}}, m_n)}, \quad (4.2.4)$$

where $u_n(1, \tilde{\mathbf{x}}, m_n)$ denotes the U-statistic with $\varphi \equiv 1$. Note that $u_n(\varphi, \tilde{\mathbf{x}}, m_n)$ is a classical U-statistic generated by the kernel $G_{\varphi, \tilde{\mathbf{x}}, m_n}$.

Remark 4.2.2.1 *Since the kernel $G_{\varphi, \tilde{\mathbf{x}}, m_n}(\cdot, \cdot)$ is not a symmetric function, we construct its symmetrized version*

$$\tilde{G}_{\varphi, \tilde{\mathbf{x}}, m_n}(\mathbf{t}, \mathbf{y}) = \frac{1}{k!} \sum_{\sigma \in I(k, k)} G_{\varphi, \tilde{\mathbf{x}}, m_n}(\mathbf{t}_\sigma, \mathbf{y}_\sigma) = \frac{1}{k!} \sum_{\sigma \in I(k, k)} \varphi(\mathbf{y}_\sigma) \tilde{\delta}_{m_n}(\tilde{\mathbf{x}}, \mathbf{t}_\sigma), \quad (4.2.5)$$

where $\mathbf{t}_\sigma = (\mathbf{t}_{\sigma_1}, \dots, \mathbf{t}_{\sigma_k})$, $\mathbf{y}_\sigma = (\mathbf{y}_{\sigma_1}, \dots, \mathbf{y}_{\sigma_k})$ and

$$\tilde{\delta}_{m_n}(\tilde{\mathbf{x}}, \mathbf{t}) = \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j).$$

Notice that, after the symmetrization, the expectation

$$\mathbb{E}[\tilde{G}_{\varphi, \tilde{\mathbf{x}}, m_n}(\mathbf{t}, \mathbf{y})] = \mathbb{E}[G_{\varphi, \mathbf{x}, m_n}(\mathbf{t}, \mathbf{y})].$$

This remark will be useful when applying Lemma 4.8.0.2 later in the proofs.

4.3 Main results

4.3.1 Conditions and comments

Our results are stated under some assumptions we gather hereafter for easy reference.

(C.1.) Let $\alpha = (\alpha_1, \dots, \alpha_k)$ be a multi-index of non-negative integers α_i , such that $|\alpha| = \sum_{i=1}^k \alpha_i$, we assume that the density function $f_{\mathbf{X}}(\cdot)$ is $|\alpha|$ -times continuously differentiable on \mathcal{X} and there exists a finite constant C_α such that, for all $\mathbf{x}_1, \dots, \mathbf{x}_k$ and $j = 1, \dots, k$

$$\sup_{h_{\mathbf{x}, \mathbf{u}} \in [0, 1]} \left| \frac{\partial^{\alpha_j}}{\partial x_{i_1} \cdots \partial x_{i_{\alpha_j}}} f_{\mathbf{X}}(\mathbf{x}_j + h_{\mathbf{x}, \mathbf{u}} \mathbf{u}_j) \right| \leq C_\alpha;$$

(C.2.) We also suppose that the density function $f_{\mathbf{X}}(\cdot)$ is such that, for all $\mathbf{x} \in \mathcal{X}$,

(C.2.a) For some finite constant $f_{\mathbf{X}, \max}$, we have $f_{\mathbf{X}}(\mathbf{x}) \leq f_{\mathbf{X}, \max}$,

(C.2.b) There exists a constant $f_{\mathbf{X}, \min} > 0$ such that $f_{\mathbf{X}}(\mathbf{x}) > f_{\mathbf{X}, \min}$;

(C.3.) There exists a measure μ on $(\mathbb{R}^{d'}, \mathcal{A})$ such that $\mathbb{P}_{\mathbf{X}, \mathbf{Y}}$ is absolutely continuous with respect to $\text{Leb}_d \otimes \mu$, where Leb_d is the Lebesgue measure on \mathbb{R}^d ;

(C.4.) For every $\mathbf{y} \in \mathbb{R}^{d'}$, $\mathbf{x} \mapsto f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y})$ is differentiable almost everywhere up to the order α . Moreover, there exists positive finite constants $C_{\alpha, \psi}$ and $C_{\varphi, f, \alpha}$, such that, for every $0 \leq i_1, \dots, i_{\alpha_j} \leq d$

$$\sup_{h_{\mathbf{x}, \mathbf{u}} \in [0, 1]} \left| \frac{\partial^{\alpha_j}}{\partial x_{i_1} \cdots \partial x_{i_{\alpha_j}}} f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}_j + h_{\mathbf{x}, \mathbf{u}} \mathbf{u}_j, \mathbf{y}_j) \right| \leq C_{\alpha, \psi},$$

and

$$\begin{aligned} & \sum_{l=1}^{\alpha-1} \frac{1}{l!} \sum_{\alpha_1 + \dots + \alpha_k = l} \binom{\alpha}{\alpha_{1:k}} \int \left| \prod_{j=1}^k (\varphi(\mathbf{y}_1, \dots, \mathbf{y}_k) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) | \mathbf{X}_j = \mathbf{x}_j, \forall j = 1, \dots, k]) \right| \\ & \quad \times \left\{ \sum_{i_1, \dots, i_{\alpha_j} = 1}^d u_{j, i_1} \cdots u_{j, i_{\alpha_j}} \right\} \left| \frac{\partial^{\alpha_j}}{\partial x_{i_1} \cdots \partial x_{i_{\alpha_j}}} f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}_j + h_{\mathbf{x}, \mathbf{u}} \mathbf{u}_j, \mathbf{y}_j) \right| d\mu(\mathbf{y}_1) \cdots d\mu(\mathbf{y}_k) \\ & \leq C_{\varphi, f, \alpha} \prod_{j=1}^k \|\mathbf{u}_j\|^d. \end{aligned} \quad (4.3.2)$$

For every choice of $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathcal{X}$ and $\mathbf{y}_1, \dots, \mathbf{y}_k \in \mathbb{R}^{d'}$, $\mathbf{u}_1, \dots, \mathbf{u}_k \in \mathbb{R}^d$ such that $\mathbf{x}_i + \mathbf{u}_i \in \mathcal{X}$.

(C.5) The function $\varphi(\cdot)$ is bounded and there exists a constant C_φ such that $\|\varphi(\cdot)\|_\infty := C_\varphi < +\infty$.

(C.5') The function $\varphi(\cdot)$ is unbounded and there exists a positive function B_φ such that for all $\ell \geq 1$ and $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathbb{R}^d$,

$$\mathbb{E} \left[|\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k)|^\ell | \mathbf{X}_1 = \mathbf{x}_1, \dots, \mathbf{X}_k = \mathbf{x}_k \right] \leq B_\varphi(\mathbf{x}_1, \dots, \mathbf{x}_k)^\ell \ell!,$$

such that $B_\varphi(\mathbf{X}_1, \dots, \mathbf{X}_k) \leq \tilde{B}_\varphi$ almost surely, for some finite positive constant \tilde{B}_φ .

Comments

We impose the conditions (C.1)-(C.4) for technical reasons. The conditions (C.1) and (C.2) on the marginal density $f_{\mathbf{X}}(\cdot)$ of \mathbf{X} which means that $f_{\mathbf{X}}(\cdot)$ is bounded and has finite partial derivatives is necessary to establish finite bounds. These conditions are standard in literature, especially when some approximation theorems, such as the Taylor expansion formulas, are used, which is the case in this paper. For the same reasons, we impose the conditions (C.4) on the joint density function of the vector (\mathbf{X}, \mathbf{Y}) . The conditions (C.5) deals with a bounded function $\varphi(\cdot)$, while the conditional Bernstein assumption in (C.5') helps to control the tail behavior of $\varphi(\cdot)$ so that the exponential concentration bounds are available. Notice that (C.5') is important to the truncation method. Note that the condition (C.5') may be replaced by more general hypotheses upon moments of \mathbf{Y} as in Deheuvels [2011] and Bouzebda [2023]. That is

(C.5)'' We denote by $\{\mathcal{M}(x) : x \geq 0\}$ a nonnegative continuous function, increasing on $[0, \infty)$, and such that, for some $s > 2$, ultimately as $x \uparrow \infty$,

$$(i) \ x^{-s} \mathcal{M}(x) \downarrow; (ii) \ x^{-1} \mathcal{M}(x) \uparrow. \quad (4.3.3)$$

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

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

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

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

4.3.2 Exponential bounds

The following result concerns the case when the function $\varphi(\cdot)$ is bounded in the sense of the condition (C.5).

Theorem 4.3.2.1 (Exponential bound for the estimator $\hat{r}_n^{(k)}$ with fixed $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathcal{X}^k$) Let $\delta_{m_n}(\cdot, \cdot)$ be a positive sequence belonging to $L^\infty(\mathbb{R}^d \times \mathbb{R}^d)$, if the conditions (C.1)-(C.5) are satisfied then there exist positive constants C_1, C_2 , and C_3 such that

$$\kappa_1 := C_1 m^{2dk} \quad \text{and} \quad \kappa_2 = C_2 m^{dk},$$

and for every $t, t' > 0$ such that,

$$C_3 m^{-\lambda} + t < \frac{f_{\mathbf{X}, \min}^k}{2},$$

we have

$$\begin{aligned} \mathbb{P} \left(\left| \hat{r}_n^{(k)}(\varphi, \tilde{\mathbf{x}}, m_n) - r^{(k)}(\varphi, \tilde{\mathbf{x}}) \right| < \left(1 + C_4 m^{-\lambda} + C_5 t \right) \times \left(C_6 m^{-\lambda} + t' \right) \right) \\ \geq 1 - 2 \exp \left(- \frac{[n/k] t^2}{2\kappa_1 + \frac{2}{3} \{C_8^k - \kappa_2\} t} \right) - 2 \exp \left(- \frac{[n/k] t^2}{2C_7 + C_8 t'} \right), \end{aligned} \quad (4.3.4)$$

where

$$\begin{aligned} C_\delta &:= \sup_{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^d} |\boldsymbol{\delta}_{m_n}(\mathbf{x}, \mathbf{y})|, & C_4 &:= 4C_3 f_{X, \max}^k f_{X, \min}^{-2k}, \\ C_5 &:= 4f_{X, \max}^k f_{X, \min}^{-2k}, & C_6 &:= C_{\delta, \alpha, \varphi} m^{-\lambda} f_{X, \min}^{-k}, \\ C_7 &:= C_\varphi^2 f_{X, \min}^{-2k} C_\delta^{2k}, & C_8 &:= \frac{8}{3} C_\varphi C_\delta^k f_{X, \min}^{-k}. \end{aligned}$$

The proof of Theorem 4.3.2.1 is postponed until Sect. 4.7.

Now, we present a more general result concerning the case when the function φ is unbounded in the sense of the condition (C.5'). That being said, the preceding theorem constitutes an important step in the truncation method used in the proof of the following one.

Theorem 4.3.2.2 *Assume that the conditions (C.1)-(C.4) and (C.5') are satisfied, for every $t, t' > 0$ such that*

$$C_3 m^{-\lambda} + t < \frac{f_{X, \min}^k}{2},$$

we have

$$\begin{aligned} \mathbb{P} \left(\left| \widehat{r}_n^{(k)}(\varphi, \tilde{\mathbf{x}}, m_n) - r^{(k)}(\varphi, \tilde{\mathbf{x}}) \right| < \left(1 + C_4 m^{-\lambda} + C_5 t\right) \times \left(C_6 m^{-\lambda} + t'\right) \right) \\ \geq 1 - 2 \exp \left(-\frac{[n/k] t^2}{2\kappa_1 + \frac{2}{3} \{C_\delta^k - \kappa_2\} t} \right) - 2 \exp \left(-\frac{[n/k] t^2}{2\tilde{C}_7 + \tilde{C}_8 t'} \right), \end{aligned} \quad (4.3.5)$$

where

$$\begin{aligned} \tilde{C}_7 &= 128 (B_{\varphi, \mathbf{x}} + \tilde{B}_\varphi)^2 f_{X, \min}^{-k} \kappa'_2, \\ \tilde{C}_8 &= 2 (B_{\varphi, \mathbf{x}} + \tilde{B}_\varphi) f_{X, \min}^{-k} \kappa'_1, \end{aligned}$$

and for some positive constants C'_1 and C'_2

$$B_{\varphi, \mathbf{x}} := B_\varphi(\mathbf{x}_1, \dots, \mathbf{x}_k), \quad \kappa'_1 := C'_1 m^{dk}, \quad \text{and} \quad \kappa'_2 := C'_2 m^{d(2k-1)}.$$

The proof of Theorem 4.3.2.2 is postponed until Sect. 4.7.

We have the following immediate result by combining the last theorem and the Borel-Cantelli lemma.

Corollary 4.3.2.3 *Assume that the conditions (C.1)-(C.4) and (C.5') are satisfied. We have, as $n \rightarrow \infty$,*

$$\left| \widehat{r}_n^{(k)}(\varphi, \tilde{\mathbf{x}}, m_n) - r^{(k)}(\varphi, \tilde{\mathbf{x}}) \right| = O \left(\sqrt{\frac{m_n^{dk} \log m_n}{n}} \right), \quad a.s.$$

The proof of the preceding theorems will be based on the following technical lemmas.

Lemma 4.3.2.4 *Let $\boldsymbol{\delta}_{m_n}(\cdot, \cdot)$ be a positive sequence, and $f_{\mathbf{X}} \in L^p(\mathbb{R}^d)$, with $1 \leq p \leq \infty$, the density function of the random vector \mathbf{X} satisfying condition (C.1), then we have*

$$\left| \mathbb{E}[u_n(1, \tilde{\mathbf{x}}; m_n)] - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right| = O(m^{-\lambda}), \quad (4.3.6)$$

where $\lambda := \min\{k, kq^{-1}\}$, with $q^{-1} = 1 - p^{-1}$.

The proof of Lemma 4.3.2.4 is postponed until Sect. 4.7.

Lemma 4.3.2.5 *Let $\delta_{m_n}(\cdot, \cdot)$ be a positive sequence belonging to $L^\infty(\mathbb{R}^d \times \mathbb{R}^d)$, if the conditions of Lemma 4.3.2.4 are satisfied then, for some $t > 0$, we have*

$$\mathbb{P}\left(\left|u_n(1, \tilde{\mathbf{x}}; m_n) - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j)\right| \leq C_3 m^{-\lambda} + t\right) \geq 1 - \exp\left(-\frac{[n/k]t^2}{2\kappa_1 + \frac{2}{3}\{C_\delta^k - \kappa_2\}t}\right). \quad (4.3.7)$$

The proof of Lemma 4.3.2.5 is postponed until Sect. 4.7.

Lemma 4.3.2.6 *If the conditions of Lemma 4.3.2.5 are satisfied, and if for some $t > 0$,*

$$C_3 m^{-\lambda} + t < \frac{f_{\mathbf{X}, \min}^k}{2},$$

we have

$$\mathbb{P}\left(\left|\frac{1}{U_n(1, \tilde{\mathbf{x}}, m_n)} - \frac{1}{\prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j)}\right| \leq \frac{4}{f_{\mathbf{X}, \min}^{2k}} (C_3 m^{-\lambda} + t)\right) \geq 1 - \exp\left(-\frac{[n/k]t^2}{2\kappa_1 + \frac{2}{3}\{C_\delta^k - \kappa_2\}t}\right), \quad (4.3.8)$$

and

$$\frac{\prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j)}{\sum_{\mathbf{i} \in I(k, n)} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{X}_{i_j})} \leq 1 + \frac{4f_{\mathbf{X}, \max}^k}{f_{\mathbf{X}, \min}^{2k}} (C_3 m^{-\lambda} + t). \quad (4.3.9)$$

The proof of Lemma 4.3.2.6 is postponed until Sect. 4.7.

Lemma 4.3.2.7 *Under Assumptions (C.1)-(C.4), there exists a constant $C_{\delta, \alpha, \varphi} > 0$ such that We have for all $\mathbf{i} \in I(k, n)$,*

$$|\mathbb{E}[S_{\mathbf{i}}]| \leq \frac{C_{\delta, \alpha, \varphi} m^{-\lambda}}{f_{\mathbf{X}, \min}^k},$$

where

$$\mathbb{E}[S_{\mathbf{i}}] = \prod_{j=1}^k \frac{\delta_{m_n}(\mathbf{x}_j, \mathbf{X}_{i_j})}{f_{\mathbf{X}}(\mathbf{x}_j)} \{\varphi(\mathbf{Y}_{i_1}, \dots, \mathbf{Y}_{i_k}) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) | (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}]\}.$$

The proof of Lemma 4.3.2.7 is postponed until Sect. 4.7.

Lemma 4.3.2.8 *Under Assumptions (C.2.b) and (C.5') for every $t > 0$, we have*

$$\mathbb{P}\left(\sum_{\mathbf{i} \in I(k, n)} (S_{\mathbf{i}} - \mathbb{E}[S_{\mathbf{i}}]) > t\right) \leq \exp\left(-\frac{t^2 f_{\mathbf{Z}, \min}^{2k} [n/k]}{128 (B_{g, \mathbf{z}} + \tilde{B}_g)^2 C_K^{2k-1} + 2t (B_{g, \mathbf{z}} + \tilde{B}_g) C_\delta^k f_{\mathbf{Z}, \min}^k}\right).$$

The proof of Lemma 4.3.2.8 is postponed until Sect. 4.7.

4.3.3 Limiting distribution

In this section, we establish the central limit theorem for the estimate defined in (4.2.3). In order to do that, we will need the following assumptions:

4.3.4 Conditions

(A.1) Let $m = O\left(n^{(\frac{1}{d})-\epsilon}\right)$ for some $0 < d\epsilon < 1$;

(A.2) The delta sequence δ_m has a compact support;

(A.3) Let $\tilde{\mathbf{x}} = (\mathbf{x}_1, \dots, \mathbf{x}_k)$ be a point of continuity for each

$$r_{jl}(\tilde{\mathbf{x}}) = \begin{cases} 0 & \text{if } \mathbf{x}_j \neq \mathbf{x}_l, \\ \mathbb{E}_{j,l}(\tilde{\mathbf{x}}) & \text{if } \mathbf{x}_j = \mathbf{x}_l, \end{cases}$$

where

$$\mathbb{E}_{j,l}(\tilde{\mathbf{x}}) = \mathbb{E} \left[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_{j-1}, \mathbf{Y}, \mathbf{Y}_{j+1}, \dots, \mathbf{Y}_k) \varphi(\mathbf{Y}_{k+1}, \dots, \mathbf{Y}_{k+j-1}, \mathbf{Y}, \mathbf{Y}_{k+j+1}, \dots, \mathbf{Y}_{2k}) \right. \\ \left. | \mathbf{X}_i = \mathbf{x}_i \text{ for } i \neq j, \mathbf{X}_{k+r} = \mathbf{x}_r \text{ for } r \neq l \text{ and } \mathbf{X} = \mathbf{x}_j = \mathbf{x}_l \right];$$

(A.4) The density function $f_{\mathbf{X}}$ is continuous at each \mathbf{x}_j , $1 \leq j \leq k$, with $f_{\mathbf{X}}(\mathbf{x}_j) > 0$;

(A.5) $r_{j,l,s}(\cdot, \cdot, \cdot)$ is bounded in a neighborhood of $(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}, \tilde{\mathbf{x}}) \in \mathcal{X}^{3k}$, for all $1 \leq j, l, s \leq k$, where :

$$r_{j,l,s}(\tilde{\mathbf{z}}_k, \tilde{\mathbf{z}}_{2k}, \tilde{\mathbf{z}}_{3k}) = \mathbb{E} \left[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_{j-1}, \mathbf{Y}, \mathbf{Y}_{j+1}, \dots, \mathbf{Y}_k) \right. \\ \times \varphi(\mathbf{Y}_{k+1}, \dots, \mathbf{Y}_{k+j-1}, \mathbf{Y}, \mathbf{Y}_{k+j+1}, \dots, \mathbf{Y}_{2k}), \\ \times \varphi(\mathbf{Y}_{2k+1}, \dots, \mathbf{Y}_{2k+j-1}, \mathbf{Y}, \mathbf{Y}_{2k+j+1}, \dots, \mathbf{Y}_{3k}) \\ \left. | \mathbf{X}_i = \mathbf{z}_i; 1 \leq i \leq 3k, i \neq j, k+1, 2k+s, \mathbf{X} = \mathbf{z} \right],$$

and for $1 \leq s \leq 3$, $\tilde{\mathbf{z}}_{sk} = (\mathbf{z}_{(s-1)k+1}, \dots, \mathbf{z}_{(s-1)k+j-1}, \mathbf{z}, \mathbf{z}_{(s-1)k+j+1}, \dots, \mathbf{z}_{sk})$;

(A.6) $r_{1,2}^{(k)}(\cdot, \cdot)$ is bounded in a neighborhood of $(\tilde{\mathbf{x}}, \tilde{\mathbf{x}})$, where

$$r_{1,2}^{(k)}(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2) = \mathbb{E} \left[\varphi(\mathbf{Y}_{i_1}, \dots, \mathbf{Y}_{i_k}) \varphi(\mathbf{Y}_{j_1}, \dots, \mathbf{Y}_{j_k}) | (\mathbf{X}_{i_1}, \dots, \mathbf{X}_{i_k}) = \tilde{\mathbf{x}}_1, (\mathbf{X}_{j_1}, \dots, \mathbf{X}_{j_k}) = \tilde{\mathbf{x}}_2 \right].$$

(A.7) Let $r^{(k)}(\varphi, \cdot)$ admit an expansion

$$r^{(k)}(\varphi, \mathbf{t} + \Delta) = r^{(k)}(\varphi, \mathbf{t}) + \left\{ \frac{\partial}{\partial \mathbf{t}} r^{(k)}(\varphi, \mathbf{t}) \right\}^t \Delta + \frac{1}{2} \Delta^t \left\{ \frac{\partial^2}{\partial \mathbf{t}^2} r^{(k)}(\varphi, \mathbf{t}) \right\} \Delta + o(\Delta^t \Delta),$$

as $\Delta \rightarrow 0$, for all \mathbf{t} in a neighborhood of $\tilde{\mathbf{x}}$.

(A.8) $f_{\mathbf{X}}(\cdot)$ is twice differentiable in neighborhoods of \mathbf{x}_j , $1 \leq j \leq k$,

(A.9) $\{\delta_{m_n}\}$ is symmetric at zero.

Below, we write $Z \stackrel{\mathcal{D}}{=} \mathcal{N}(\mu, \sigma^2)$ whenever the random variable Z follows a normal law with expectation μ and variance σ^2 , $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution. We also denote

$$U_n(\varphi, \tilde{\mathbf{x}}, m_n) = \frac{u_n(\varphi, \tilde{\mathbf{x}}, m_n)}{N},$$

where

$$N = \prod_{j=1}^k \mathbb{E}[\delta_{m_n}(\mathbf{x}_j, \mathbf{X}_1)].$$

Our main result in this section is summarized as follows.

Theorem 4.3.4.1 *Under assumptions (A.1)-(A.6) and if $r^{(k)}(\cdot, \cdot)$ is continuous at $\tilde{\mathbf{x}}$, then :*

$$\sqrt{nm_n^{-d}} \left(\hat{r}^{(k)}(\varphi, \mathbf{x}, m_n) - \mathbb{E}[U_n(\varphi, \tilde{\mathbf{x}}, m_n)] \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \rho^2),$$

where

$$\rho^2 := \sum_{i=1}^k \sum_{j=1}^k \mathbb{1}_{\{\mathbf{x}_i = \mathbf{x}_j\}} \left[r_{jl}(\tilde{\mathbf{x}}) - (r^{(k)}(\tilde{\mathbf{x}}))^2 \right] \int \delta^2(\mathbf{x}, \mathbf{t}) d\mathbf{t} / f_{\mathbf{X}}(\mathbf{t}_i). \quad (4.3.10)$$

The proof of Theorem 4.3.4.1 is postponed until Sect. 4.7.

The following corollary is more or less straightforward, given Theorem 4.3.4.1.

Corollary 4.3.4.2 *If in addition to Theorem 4.3.4.1, (A.7)-(A.9) hold, then*

$$m_n^{2d} \left[\mathbb{E}U_n(\varphi, \tilde{\mathbf{x}}, m_n) - r^{(k)}(\varphi, \tilde{\mathbf{x}}) \right] = \frac{1}{2} \left[\prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j) \mathbf{t}^t \{r''(\tilde{\mathbf{x}})\} \mathbf{t} \cdot d\mathbf{t} / \tilde{f}(\tilde{\mathbf{x}}) \right. \\ \left. - \int \prod_{i=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j) \mathbf{t}^t \{\tilde{f}''(\tilde{\mathbf{x}})\} \mathbf{t} \cdot d\mathbf{t} \frac{r^{(k)}(\varphi, \tilde{\mathbf{x}})}{\tilde{f}(\tilde{\mathbf{x}})} \right] + o(1),$$

where

$$r(\tilde{\mathbf{x}}) = r^{(k)}(\varphi, \tilde{\mathbf{x}}) \tilde{f}(\tilde{\mathbf{x}}), \quad \tilde{f}(\tilde{\mathbf{x}}) = \prod_{i=j}^k f_{\mathbf{X}}(\mathbf{x}_j)$$

and we have

$$\left(nm_n^{-d} \right)^{1/2} \left[\hat{r}_n^{(k)}(\varphi, \tilde{\mathbf{x}}, m_n) - r^{(k)}(\varphi, \tilde{\mathbf{x}}) \right] \rightarrow \mathcal{N}(0, \rho^2),$$

provided that $nm_n^{-5d} \rightarrow 0$.

Remark 4.3.4.3 *Under squared loss the optimal m_n^{-d} satisfies $nm_n^{-5d} \rightarrow c$, some finite c depending on $r^{(k)}(\varphi, \cdot)$ and $f_{\mathbf{X}}(\cdot)$. In this case the conclusion of Corollary 4.3.4.2 also holds, but with $\mathcal{N}(0, \rho^2)$ replaced by $\mathcal{N}(a, \rho^2)$ for some $a \neq 0$.*

Remark 4.3.4.4 *When \mathbf{X} is independent of \mathbf{Y} , $\mathbb{E}(\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \tilde{\mathbf{x}}) = \mathbb{E}(\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k))$. Hence, up to a scaling factor, the limiting distribution is equal to the limit of the unconditional U-statistic pertaining to the \mathbf{Y} sequence, as should be expected. Observe, however, that the standardizing factor is $(nm_n^{-d})^{1/2}$ with $m_n \rightarrow \infty$, indicating a lower rate of convergence. This is the cost incurred when drawing conclusions regarding conditional (local) quantities.*

Remark 4.3.4.5 *It is well known that the estimation problems of a regression function (conditional model) are especially hard 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. The idea is to consider the additive models to simplify the problem of regression estimation by fitting only functions to the data that 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} onto 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 together with the interpretability of the parametric ones, for details on these methods for functional data.*

Remark 4.3.4.6 *Let us recall the general kernel-type estimator of the regression function defined, for $\mathbf{x} \in \mathbb{R}^d$, by*

$$\widehat{m}_{\varphi,n}(\mathbf{x}, h_n) := \frac{\sum_{i=1}^n \varphi(Y_i) K((\mathbf{x} - \mathbf{X}_i) / h_n)}{\sum_{i=1}^n K((\mathbf{x} - \mathbf{X}_i) / h_n)}. \quad (4.3.11)$$

By setting $\varphi(y) = y$ into (4.3.11) we get the classical Nadaraya-Watson kernel regression function estimator of $m(\mathbf{x}) := \mathbb{E}(Y | \mathbf{X} = \mathbf{x})$ given by

$$\widehat{m}_{n;h_n}(\mathbf{x}) := \frac{\sum_{i=1}^n Y_i K((\mathbf{x} - \mathbf{X}_i) / h_n)}{\sum_{i=1}^n K((\mathbf{x} - \mathbf{X}_i) / h_n)}. \quad (4.3.12)$$

We define the internal estimator at some predefined point \mathbf{x} by

$$\widehat{m}_{n;h_n}^{\text{Int}}(\mathbf{x}) := \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right) \frac{Y_i}{\widehat{q}_n(\mathbf{X}_i)}. \quad (4.3.13)$$

For more details on the estimators (4.3.12) and (4.3.13), the interested reader may refer, e.g., to Wand and Jones [1995]. Linton and Jacho-Chávez [2010] pointed out that Mack and Müller [1989] were the first to propose $\widehat{m}_{n;h_n}^{\text{Int}}(\mathbf{x})$, for $d = 1$, with a view to the estimation of derivatives by computing the derivative of the regression, which has a simpler form than the derivative of the Nadaraya-Watson smoother. The term “internal” stands for the fact that the factor $f_n^{-1}(\mathbf{X}_i)$ is internal to the summation, while the estimator $\widehat{m}_{n;h_n}(\mathbf{x})$ has the factor

$$\widehat{q}_n^{-1}(\mathbf{x}) = \left(\frac{1}{na_n^d} \sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_i}{a_n}\right) \right)^{-1}$$

externally to the summation. *Jones et al. [1994]* considered various versions of kernel-type regression estimators, especially the Nadaraya-Watson estimator and the local linear estimator. They established the equivalence between the local linear estimator and the internal estimator. *Linton and Jacho-Chávez [2010]* and *Shen and Xie [2013]* indicated that the internal estimators are particularly adequate for the additive nonparametric regression model. Notice that the estimator in (4.3.13) has also been used by *Hengartner and Sperlich [2005]* in the context of estimating additive models. The authors showed that it has some additional theoretical advantages over using the Nadaraya-Watson estimator. They also mention the computational attractiveness and better performance of the internal estimator (compared to its classical counterpart), particularly when the covariates are correlated and nonuniformly distributed. It would be of interest to consider our estimator by using the internal estimator idea.

4.4 Conditional U-statistics for censored data

Consider a triple (Y, C, \mathbf{X}) of random variables defined in $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^d$. 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

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

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 $\alpha > 0$,

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

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

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

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

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

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

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

when Y is right-censored. To estimate $r^{(1)}(\psi, \cdot)$, we make use of the Inverse Probability of Censoring Weighted (I.P.C.W.) estimators that have recently gained popularity in the censored data literature (see [Kohler et al. \[2002\]](#), [Carbonez et al. \[1995\]](#)). The key ideas of I.P.C.W. estimators are as follows. Introduce the real-valued function $\Phi_\psi(\cdot, \cdot)$ defined on \mathbb{R}^2 by

$$\Phi_\psi(y, c) = \frac{\mathbb{1}_{\{y \leq c\}} \psi(y \wedge c)}{1 - G(y \wedge c)}. \quad (4.4.1)$$

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

(\mathcal{S}) C and (Y, \mathbf{X}) are independent.

We have

$$\begin{aligned} r^{(1)}(\Phi_\psi, \mathbf{x}) &:= \mathbb{E}(\Phi_\psi(Y, C) | \mathbf{X} = \mathbf{x}) \\ &= \mathbb{E} \left\{ \frac{\mathbb{1}_{\{Y \leq C\}} \psi(Z)}{1 - G(Z)} | \mathbf{X} = \mathbf{x} \right\} \\ &= \mathbb{E} \left\{ \frac{\psi(Y)}{1 - G(Y)} \mathbb{E}(\mathbb{1}_{\{Y \leq C\}} | \mathbf{X}, Y) | \mathbf{X} = \mathbf{x} \right\} \\ &= r^{(1)}(\psi, \mathbf{x}). \end{aligned} \quad (4.4.2)$$

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 $\mathbf{x} \in \mathcal{I}$, $h \geq 0$, $1 \leq i \leq n$,

$$\bar{\omega}_{n,K,h,i}^{(1)}(\mathbf{x}) := K \left(\frac{\mathbf{x} - \mathbf{X}_i}{h} \right) / \sum_{j=1}^n K \left(\frac{\mathbf{x} - \mathbf{X}_j}{h} \right). \quad (4.4.3)$$

In view of (4.4.1), (4.4.2), and (4.4.3), whenever $G(\cdot)$ is known, a kernel estimator of $r^{(1)}(\psi, \cdot)$ is given by

$$\check{r}_n^{(1)}(\psi, \mathbf{x}; h_n) = \sum_{i=1}^n \bar{\omega}_{n,K,h,i}^{(1)}(\mathbf{x}) \frac{\Delta_i \psi(Z_i)}{1 - G(Z_i)}. \quad (4.4.4)$$

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

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

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

$$\check{r}_n^{(1)*}(\psi, \mathbf{x}; h_n) = \sum_{i=1}^n \bar{\omega}_{n,K,h,i}^{(1)}(\mathbf{x}) \frac{\Delta_i \Psi(Z_i)}{1 - G_n^*(Z_i)}, \quad (4.4.5)$$

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 k th ordered statistic associated with the sample (Z_1, \dots, Z_n) for $k = 1, \dots, n$ and $\Delta_{(k)}$ is the Δ_j corresponding to $Z_k = Z_j$. When the variable of interest is a right-censored function of the (conditional) law, it can generally not be estimated on the complete support (see [Brunel and Comte \[2006\]](#)). 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. We refer also to [Bouzebda and El-hadjali \[2020\]](#) and [Bouzebda and Taachouche \[2022\]](#). To our best knowledge, the problem of the estimation of the conditional U-statistics in the censored setting was open up to present, and it gives the main motivation to the study of this section. A natural extension of the function defined in (4.4.1) is given by

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

From this, we have an analogous relation to (4.4.2) given by

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

An analog estimator to (4.1.2) in the censored case is given by

$$\check{r}_n^{(k)}(\psi, \mathbf{x}; m_n) = \sum_{(i_1, \dots, i_k) \in \mathcal{I}(k, n)} \frac{\Delta_{i_1} \cdots \Delta_{i_k} \Psi(Z_{i_1}, \dots, Z_{i_k})}{(1 - G(Z_{i_1})) \cdots (1 - G(Z_{i_k}))} \bar{\omega}_{n, \delta, m_n, \mathbf{i}}^{(k)}(\mathbf{x}), \quad (4.4.6)$$

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

$$\bar{\omega}_{n, \delta, m_n, \mathbf{i}}^{(k)}(\mathbf{x}) = \frac{\delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k})}{\sum_{(i_1, \dots, i_k) \in I(k, n)} \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k})}. \quad (4.4.7)$$

The estimator that we will investigate is given by

$$\check{r}_n^{(k)*}(\psi, \mathbf{x}; m_n) = \sum_{(i_1, \dots, i_k) \in I(k, n)} \frac{\Delta_{i_1} \cdots \Delta_{i_k} \Psi(Z_{i_1}, \dots, Z_{i_k})}{(1 - G_n^*(Z_{i_1}) \cdots (1 - G_n^*(Z_{i_k})))} \bar{\omega}_{n, \delta, m_n, \mathbf{i}}^{(k)}(\mathbf{x}). \quad (4.4.8)$$

Corollary 4.4.0.1 *Assume that the conditions (C.1)-(C.4) and (C.5') are satisfied. We have, as $n \rightarrow \infty$,*

$$\left| \check{r}_n^{(k)*}(\psi, \mathbf{x}; m_n) - r^{(k)}(\varphi, \tilde{\mathbf{x}}) \right| = O\left(\sqrt{\frac{m_n^{dk} \log m_n}{n}} \right), \quad a.s.$$

This last result is a direct consequence of Corollary (4.3.2.3) and the law of iterated logarithm for $G_n^*(\cdot)$ established in Földes and Rejtő [1981] ensures that

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

For more details refer to Bouzebda *et al.* [2022, 2023] and Bouzebda and Soukarieh [2023a].

4.5 Applications

4.5.1 Discrimination problems

Now, we apply the results of 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 $1, \dots, M$. The sets

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

then yield a partition of the feature space. Predicting the value of $\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k)$ is tantamount to predicting the set in the partition to which $(\mathbf{Y}_1, \dots, \mathbf{Y}_k)$ belongs. For any discrimination rule g , we have

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

where

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

The above inequality becomes equality if

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

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

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

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

$$\mathfrak{M}_n^j(\tilde{\mathbf{x}}) = \frac{\sum_{(i_1, \dots, i_k) \in \mathcal{I}(k, n)} \mathbb{1}\{\varphi(Y_{i_1}, \dots, Y_{i_k}) = j\} \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k})}{\sum_{(i_1, \dots, i_k) \in \mathcal{I}(k, n)} \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k})}, \quad (4.5.1)$$

Set

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

Let us introduce

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

The discrimination rule $g_{0,n}(\cdot)$ is asymptotically Bayes' risk consistent

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

This follows from the apparent relation

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

4.5.2 Metric learning

Metric learning aims at adapting the metric to the data and has attracted much interest in recent years; for instance, see [Bellet *et al.* \[2013\]](#) and [Cl  men  on *et al.* \[2016\]](#) for an account of metric learning and its applications. This is motivated by a variety of applications ranging from computer vision to information retrieval through bioinformatics. To illustrate the usefulness of this concept, we present the metric learning problem for supervised classification as in [Cl  men  on *et al.* \[2016\]](#). Let us consider independent copies $(X_1, Y_1), \dots, (X_n, Y_n)$ of a $\mathcal{X} \times \mathcal{Y}$ valued random couple (X, Y) , where \mathcal{X} is some feature space and $\mathcal{Y} = \{1, \dots, C\}$, with $C \geq 2$ say, a finite set of labels. Let \mathcal{D} be a set distance measures $D: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_+$. Intuitively, the goal of metric learning in this setting is to find a metric under which pairs of points with the same label are close to each other, and those with different labels are far away. The natural way to define the risk of a metric D is

$$R(D) = \mathbb{E} \left[\phi \left((1 - D(X, X')) \cdot (2\mathbb{1}\{Y = Y'\} - 1) \right) \right], \quad (4.5.2)$$

where $\phi(u)$ is a convex loss function upper bounding the indicator function $\mathbb{1}\{u \geq 0\}$, for instance, the hinge loss $\phi(u) = \max(0, 1 - u)$. To estimate $R(D)$, we consider the natural empirical estimator

$$R_n(D) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \phi \left((D(X_i, X_j) - 1) \cdot (2\mathbb{1}\{Y_i = Y_j\} - 1) \right), \quad (4.5.3)$$

which is a one sample U-statistic of degree two with kernel given by:

$$\varphi_D((x, y), (x', y')) = \phi((D(x, x') - 1) \cdot (2\mathbb{1}\{y = y'\} - 1)).$$

The convergence to (4.5.2) of a minimizer of (4.5.3) has been studied in the frameworks of algorithmic stability (Jin *et al.* [2009]), algorithmic robustness (Bellet and Habrard [2015]) and based on the theory of U-processes under appropriate regularization (Cao *et al.* [2016]).

4.5.3 Multipartite ranking

Let us recall the problem from Cl emen on *et al.* [2016]. Let $X \in \mathcal{X}$ be a random vector of attributes/features and the (temporarily hidden) ordinal labels $Y \in \{1, \dots, K\}$ assigned to it. Based on a training set of labeled examples, the goal of multipartite ranking is to rank the attributes/features in the same order as that induced by the labels. This statistical learning problem is applied in a wide range of fields (e.g., medicine, finance, search engines, e-commerce). Rankings are generally defined utilizing a scoring function $s : \mathcal{X} \rightarrow \mathbb{R}$, transporting the natural order on the real line onto the feature space and the gold standard for evaluating the ranking performance of $s(x)$ is the ROC manifold or its usual summary the VUS criterion (VUS standing for Volume Under the ROC Surface), see Cl emen on and Robbiano [2015] and the references therein. Cl emen on *et al.* [2013] characterized the optimal scoring functions as those that are optimal for all bipartite subproblems. More precisely, they are increasing transforms of the likelihood ratio dF_{k+1}/dF_k , where F_k denotes the class-conditional distribution for the k -th class. The authors proved when the set of optimal scoring functions is non-empty, it corresponds to the functions that maximize the volume under the ROC surface

$$\text{VUS}(s) = \mathbb{P}\{s(X_1) < \dots < s(X_K) \mid Y_1 = 1, \dots, Y_K = K\}$$

Given K independent samples $(X_1^{(k)}, \dots, X_{n_k}^{(k)}) \stackrel{\text{i.i.d.}}{\sim} F_k(dx)$ for $k = 1, \dots, K$, the empirical counterpart of the VUS can be written in the following way:

$$\widehat{\text{VUS}}(s) = \frac{1}{\prod_{k=1}^K n_k} \sum_{i_1=1}^{n_1} \dots \sum_{i_K=1}^{n_K} \mathbb{1}\{s(X_{i_1}^{(1)}) < \dots < s(X_{i_K}^{(K)})\} \quad (4.5.4)$$

The empirical VUS (4.5.4) is a K -sample U-statistic of degree $(1, \dots, K)$ with kernel given by:

$$\varphi_s(x_1, \dots, x_K) = \mathbb{1}\{s(x_1) < \dots < s(x_K)\}.$$

4.5.4 Ranking problems

For its great importance, the problem of ranking instances has received particular attention in machine learning. In some specific ranking problems, it is necessary to compare two different observations based on their observed characteristics and decide which one is better instead of simply classifying them. The ordering problems have many applications in other areas of banking (Data mining process for direct marketing data extraction), document type classification,

and so on. The problems of ordering/ranking are frequent problems in which U-statistics come into play. In this kind of challenge, the aim is to establish a universal and consistent ordering method. Suppose that we want to establish an order between the first components of the two pairs $(X, Y), (X', Y')$ of independent and identically distributed observations in $\mathcal{X} \times \mathbb{R}$. The variables Y and Y' are respective labels of the variables X and X' that we want to order by observing them (and not their labels). Usually, we decide that X is better than X' if $Y > Y'$. To see things more clearly, we introduce the new variable:

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

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

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

The performance measure of r is given by the following ranking risk:

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

A natural estimate for $L(\cdot)$ according to [Cl  men  on *et al.* \[2008\]](#) is:

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

where $(X_1, Y_1), \dots, (X_n, Y_n)$ are n independent, identically distributed copies of (X, Y) , and $Z_{i,j} = \frac{Y_i - Y_j}{2}$. One can easily see that L_n is a U-statistic with $m = 2$. For more details the reader is invited to consult [Cl  men  on *et al.* \[2008\]](#) and [Rejchel \[2012\]](#).

4.5.5 Kendall rank correlation coefficient

To test the independence of one-dimensional random variables Y_1 and Y_2 [Kendall \[1938\]](#) proposed a method based on the U-statistic K_n with the kernel function :

$$\varphi((s_1, t_1), (s_2, t_2)) = \mathbb{1}_{\{(s_2 - s_1)(t_2 - t_1) > 0\}} - \mathbb{1}_{\{(s_2 - s_1)(t_2 - t_1) \leq 0\}}. \quad (4.5.5)$$

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

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

where $\mathbf{t} = (t_1, t_2) \in \mathbb{I} \subset \mathbb{R}^2$ and $\varphi(\cdot)$ is Kendall's kernel (4.5.5). Suppose that $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are d_1 and d_2 -dimensional random vectors respectively and $d_1 + d_2 = d$. Furthermore, suppose that Y_1, \dots, Y_n are observations of $(\boldsymbol{\xi}, \boldsymbol{\eta})$, we are interested in testing :

$$H_0 : \boldsymbol{\xi} \text{ and } \boldsymbol{\eta} \text{ are conditionally independent given } X. \text{ vs } H_a : H_0 \text{ is not true.} \quad (4.5.6)$$

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

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

An application of Corollary 4.3.2.3 gives

$$|\widehat{r}_n^{(2)}(\varphi^a, \mathbf{x}; m_n) - r^{(2)}(\varphi^a, \mathbf{x})| = \left(\sqrt{\frac{m_n^{2d} \log m_n}{n}} \right). \quad (4.5.7)$$

4.5.6 Generalized U-Statistics

The extension to the case of several samples is straightforward. Consider ℓ independent collections, of independent observations

$$\left\{ (\mathbf{X}_1^{(1)}, \mathbf{Y}_1^{(1)}), (\mathbf{X}_2^{(1)}, \mathbf{Y}_2^{(1)}), \dots \right\}, \dots, \left\{ (\mathbf{X}_1^{(\ell)}, \mathbf{Y}_1^{(\ell)}), (\mathbf{X}_2^{(\ell)}, \mathbf{Y}_2^{(\ell)}), \dots \right\}.$$

Let, for $\mathbf{t} \in \mathbb{R}^{d(k_1 + \dots + k_\ell)}$,

$$\begin{aligned} r^{(k, \ell)}(\varphi, \mathbf{t}) &= r^{(k, \ell)}(\varphi, \mathbf{t}_1, \dots, \mathbf{t}_\ell) \\ &= \mathbb{E} \left(\varphi \left(\mathbf{Y}_1^{(1)}, \dots, \mathbf{Y}_{k_1}^{(1)}; \dots; \mathbf{Y}_1^{(\ell)}, \dots, \mathbf{Y}_{k_\ell}^{(\ell)} \right) \mid \left(\mathbf{X}_1^{(j)}, \dots, \mathbf{X}_{m_j}^{(j)} \right) = \mathbf{t}_j, j = 1, \dots, \ell \right), \end{aligned}$$

where φ is assumed, without loss of generality, to be symmetric within each of its ℓ blocks of arguments. Corresponding to the "kernel" φ and assuming $n_1 \geq k_1, \dots, n_\ell \geq k_\ell$, the the conditional U-statistic for estimation of $r^{(k, \ell)}(\varphi, \mathbf{t})$ is defined as

$$\begin{aligned} \widehat{r}_n^{(k, \ell)}(\varphi, \mathbf{t}) &= \frac{\sum_c \varphi \left(\mathbf{Y}_{i_{11}}^{(1)}, \dots, \mathbf{Y}_{i_{1k_1}}^{(1)}; \dots; \mathbf{Y}_{i_{k_1}}^{(\ell)}, \dots, \mathbf{Y}_{i_{k_\ell k_\ell}}^{(\ell)} \right) \mathbf{K} \left(\mathbf{X}_{i_{11}}^{(1)}, \dots, \mathbf{X}_{i_{1k_1}}^{(1)}; \dots; \mathbf{X}_{i_{k_1}}^{(\ell)}, \dots, \mathbf{X}_{i_{k_\ell k_\ell}}^{(\ell)} \right)}{\sum_c \mathbf{K} \left(\mathbf{X}_{i_{11}}^{(1)}, \dots, \mathbf{X}_{i_{1k_1}}^{(1)}; \dots; \mathbf{X}_{i_{k_1}}^{(\ell)}, \dots, \mathbf{X}_{i_{k_\ell k_\ell}}^{(\ell)} \right)}, \end{aligned}$$

where

$$\mathbf{K} \left(\mathbf{X}_{i_{11}}^{(1)}, \dots, \mathbf{X}_{i_{1k_1}}^{(1)}; \dots; \mathbf{X}_{i_{k_1}}^{(\ell)}, \dots, \mathbf{X}_{i_{k_\ell k_\ell}}^{(\ell)} \right) = \prod_{j=1}^{\ell} \delta_{m_n} \left(\mathbf{x}_1^{(j)}, \mathbf{X}_{i_1}^{(j)} \right) \cdots \delta_{m_n} \left(\mathbf{x}_{k_j}^{(j)}, \mathbf{X}_{i_{k_j}}^{(j)} \right).$$

Here $\{i_{j1}, \dots, i_{jm_j}\}$ denotes a set of k_j distinct elements of the set $\{1, 2, \dots, n_j\}$ $1 \leq j \leq \ell$, and \sum_c denotes summation over all such combinations. The extension of **Hoeffding** [1948] treatment

of one-sample U -statistics to the ℓ sample case is due to [Lehmann \[1951\]](#) and [Dwass \[1956\]](#). One can use Corollary 4.3.2.3 to infer that

$$\left| \widehat{r}_n^{(k,\ell)}(\varphi, \mathbf{t}) - r^{(k,\ell)}(\varphi, \mathbf{t}) \right| \longrightarrow 0. \quad (4.5.8)$$

4.5.7 Set indexed conditional U-statistics

We aim to study the links between X and Y, by estimating functional operators associated to the conditional distribution of Y given X such as the regression operator, for $C_1 \times \cdots \times C_m := \widetilde{\mathcal{C}} \in \mathbb{R}^{d^k}$

$$\mathbb{G}^{(m)}(C_1 \times \cdots \times C_m | \mathbf{t}) = \mathbb{E} \left(\prod_{i=1}^m \mathbb{1}_{\{Y_i \in C_i\}} | (X_1, \dots, X_m) = \mathbf{t} \right) \text{ for } \mathbf{t} \in \mathbb{R}^{dm}.$$

As a particular case of (4.2.3), we estimate $\mathbb{G}^{(m)}(C_1 \times \cdots \times C_m | \mathbf{t})$

$$\widehat{\mathbb{G}}_n^{(m)}(\widetilde{\mathcal{C}}, \mathbf{t}) = \frac{\sum_{(i_1, \dots, i_k) \in \mathcal{I}(k,n)} \mathbb{1}_{\{Y_{i_j} \in C_j\}} \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k})}{\sum_{(i_1, \dots, i_k) \in \mathcal{I}(k,n)} \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k})} \quad (4.5.9)$$

One can apply Corollary 4.3.2.3 to infer that

$$\left| \widehat{\mathbb{G}}_n^{(m)}(\widetilde{\mathcal{C}}, \mathbf{t}) - \mathbb{G}^{(m)}(\widetilde{\mathcal{C}} | \mathbf{t}) \right| \longrightarrow 0. \quad (4.5.10)$$

4.5.8 Testing stochastic monotonicity

This application is given in [Chen and Kato \[2020\]](#). Let us recall their setting. Let X, Y be real-valued random variables and denote by $F_{Y|X}(y | x)$ the conditional distribution function of Y given X. Consider the problem of testing the stochastic monotonicity

$$H_0 : F_{Y|X}(y | x) \leq F_{Y|X}(y | x') \quad \forall y \in \mathbb{R} \text{ whenever } x \geq x'.$$

Testing for stochastic monotonicity is an important topic in a variety of applied fields such as economics [Blundell et al. \[2007\]](#); [Ellison and Ellison \[2011\]](#); [Solon \[1992\]](#). For this problem, [Lee et al. \[2009b\]](#) consider a test for H_0 based on a local Kendall's tau statistic, inspired by [Ghosal et al. \[2000b\]](#). Let $(X_i, Y_i), i = 1, \dots, n$ be i.i.d. copies of (X, Y) . [Lee et al. \[2009b\]](#) consider the U-process

$$U_n(x, y) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \{\mathbb{1}(Y_i \leq y) - \mathbb{1}(Y_j \leq y)\} \text{sign}(X_i - X_j) \mathbb{K} \left(\frac{x - X_i}{h_n} \right) \mathbb{K} \left(\frac{y - Y_j}{h_n} \right),$$

where $\text{sign}(x) = \mathbb{1}(x > 0) - \mathbb{1}(x < 0)$ is the sign function. They propose to reject the null hypothesis if $S_n = \sup_{(x,y) \in \mathcal{X} \times \mathcal{Y}} U_n(x, y) / c_n(x)$ is large, where \mathcal{X}, \mathcal{Y} are subsets of the supports of X, Y, respectively and $c_n(x) > 0$ is a suitable normalizing constant. Under independence between X and Y, and under regularity conditions, they derive a Gumbel limiting distribution for a properly scaled version of S_n using techniques from [Piterbarg, 1996](#). Using the procedure of the present paper, we can give the following generalization

$$U_n(x, y) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \{\mathbb{1}(Y_i \leq y) - \mathbb{1}(Y_j \leq y)\} \text{sign}(X_i - X_j) \delta_{m_n}(x, X_i) \delta_{m_n}(y, Y_j).$$

4.5.9 Testing curvature and monotonicity of nonparametric regression

This application is given in [Chen and Kato \[2020\]](#). Let us recall their setting. Consider the nonparametric regression model

$$Y = f(\mathbf{X}) + \varepsilon \quad \text{with} \quad \mathbb{E}[\varepsilon | \mathbf{X}] = 0,$$

where Y is a scalar outcome variable, \mathbf{X} is an m -dimensional vector of regressors, ε is an error term, and f is the conditional mean function $f(\mathbf{x}) = \mathbb{E}[Y | \mathbf{X} = \mathbf{x}]$. We observe i.i.d. copies $\mathbf{V}_i = (\mathbf{X}_i, Y_i)$, $i = 1, \dots, n$ of $\mathbf{V} = (\mathbf{X}, Y)$. We are interested in testing for qualitative features (e.g., curvature, monotonicity) of the regression function f .

[Abrevaya and Jiang \[2005b\]](#) consider a simplex statistic to test linearity, concavity, and convexity of f under the assumption that the conditional distribution of ε given \mathbf{X} is symmetric. To define their test statistics, for $\mathbf{x}_1, \dots, \mathbf{x}_{m+1} \in \mathbb{R}^m$, let

$$\Delta^\circ(\mathbf{x}_1, \dots, \mathbf{x}_{m+1}) = \left\{ \sum_{i=1}^{m+1} a_i \mathbf{x}_i : 0 < a_j < 1, j = 1, \dots, m+1, \sum_{i=1}^{m+1} a_i = 1 \right\}$$

denote the interior of the simplex spanned by $\mathbf{x}_1, \dots, \mathbf{x}_{m+1}$, and define

$$\mathcal{D} = \bigcup_{j=1}^{m+2} \mathcal{D}_j,$$

where

$$\mathcal{D}_j = \left\{ (\mathbf{x}_1, \dots, \mathbf{x}_{m+2}) \in \mathbb{R}^{m \times (m+2)} : \begin{array}{l} \mathbf{x}_1, \dots, \mathbf{x}_{j-1}, \mathbf{x}_{j+1}, \dots, \mathbf{x}_{m+2} \text{ are affinely independent} \\ \text{and } \mathbf{x}_j \in \Delta^\circ(\mathbf{x}_1, \dots, \mathbf{x}_{j-1}, \mathbf{x}_{j+1}, \dots, \mathbf{x}_{m+2}) \end{array} \right\}.$$

The sets $\mathcal{D}_1, \dots, \mathcal{D}_{m+2}$ are disjoint. For given $\mathbf{v}_i = (\mathbf{x}_i, y_i) \in \mathbb{R}^m \times \mathbb{R}$, $i = 1, \dots, m+2$, if $(\mathbf{x}_1, \dots, \mathbf{x}_{m+2}) \in \mathcal{D}$ then there exist a unique index $j = 1, \dots, m+2$ and a unique vector $(a_i)_{1 \leq i \leq m+2, i \neq j}$ such that $0 < a_i < 1$ for all $i \neq j$, $\sum_{i \neq j} a_i = 1$, and $\mathbf{x}_j = \sum_{i \neq j} a_i \mathbf{x}_i$; then, define

$$w(v_1, \dots, v_{m+2}) = \sum_{i \neq j} a_i y_i - y_j.$$

The index j and vector $(a_i)_{1 \leq i \leq m+2, i \neq j}$ are functions of \mathbf{x}_i 's. The set \mathcal{D} is symmetric (i.e., its indicator function is symmetric), and $w(v_1, \dots, v_{m+2})$ is symmetric in its arguments. Under this notation, [Abrevaya and Jiang \[2005b\]](#) consider the following *localized simplex statistic*

$$U_n(x) = \frac{(n-m-2)!}{n!} \sum_{(i_1, \dots, i_{m+2}) \in I_{n, m+2}} \varphi(\mathbf{V}_{i_1}, \dots, \mathbf{V}_{i_{m+2}}) \prod_{k=1}^{m+2} K\left(\frac{\mathbf{x} - \mathbf{X}_{ik}}{h_n}\right), \quad (4.5.11)$$

where $\varphi(\mathbf{v}_1, \dots, \mathbf{v}_{m+2}) = 1\{(\mathbf{x}_1, \dots, \mathbf{x}_{m+2}) \in \mathcal{D}\} \text{sign}(w(\mathbf{v}_1, \dots, \mathbf{v}_{m+2}))$, which is a U-statistic of order $(m+2)$. To test concavity and convexity of f , [Abrevaya and Jiang \[2005b\]](#) propose to reject the hypotheses if $\bar{S}_n = \sup_{x \in \mathcal{X}} U_n(x)/c_n(x)$ and $\underline{S}_n = \inf_{x \in \mathcal{X}} U_n(x)/c_n(x)$ are large and small, respectively, where \mathcal{X} is a subset of the support of X and $c_n(x) > 0$ is a suitable normalizing constant.

On the other hand, Ghosal *et al.* [2000b] study testing monotonicity of f when $m = 1$ and ε is independent of X . To be specific, they test whether f is increasing, and propose to reject the hypothesis if $S_n = \sup_{x \in \mathcal{X}} \check{U}_n(x)/c_n(x)$ is large, where \mathcal{X} is a subset of the support of X ,

$$\check{U}_n(x) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \text{sign}(Y_j - Y_i) \text{sign}(X_i - X_j) K\left(\frac{x - X_i}{h_n}\right) K\left(\frac{x - X_j}{h_n}\right), \quad (4.5.12)$$

and $c_n(x) > 0$ is a suitable normalizing constant. Ghosal *et al.* [2000b] argue that as far as the size control is concerned, it is enough to choose, as a critical value, the $(1 - \alpha)$ -quantile of S_n when $f \equiv 0$, under which $U_n(x)$ is centered. Under $f \equiv 0$ and under regularity conditions, Ghosal *et al.* [2000b] derive a Gumbel limiting distribution for a properly scaled version of S_n . Using the methods proposed in the present paper, alternative procedures to (4.5.11) and (4.5.12) are given by

$$U_n(x) = \frac{(n-m-2)!}{n!} \sum_{(i_1, \dots, i_{m+2}) \in I_{n, m+2}} \varphi(V_{i_1}, \dots, V_{i_{m+2}}) \prod_{k=1}^{m+2} \delta_{m_n}(x, X_{i_k}), \quad (4.5.13)$$

and

$$\check{U}_n(x) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \text{sign}(Y_j - Y_i) \text{sign}(X_i - X_j) \delta_{m_n}(x, X_i) \delta_{m_n}(x, X_j). \quad (4.5.14)$$

4.5.10 Examples of U -statistics

Example 4.5.10.1 Let $\widehat{Y_1 Y_2}$ 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 :

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

Silverman [1978] has used this kernel in order to propose a U -process to test uniformity on the circle.

Example 4.5.10.2 Hoeffding [1948] introduced the parameter

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

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 Δ has the property that $\Delta = 0$ if and only if Y_1 and Y_2 are independent. From Lee [1990], an alternative expression for Δ can be developed by introducing the functions

$$\Psi(y_1, y_2, y_3) = \begin{cases} 1 & \text{if } y_2 \leq y_1 < y_3 \\ 0 & \text{if } y_1 < y_2, y_3 \text{ or } y_1 \geq y_2, y_3 \\ -1 & \text{if } y_3 \leq y_1 < y_2 \end{cases}$$

and

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

We have

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

We have

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

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

Example 4.5.10.3 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 4.5.10.4 (Hoeffding's D). From the symmetric kernel,

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

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

Example 4.5.10.5 (Blum-Kiefer-Rosenblatt's R). The symmetric kernel

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

yields Blum-Kiefer-Rosenblatt's R statistic ([Blum et al. \[1961\]](#)), which is a rank-based U-statistic of order 6. At this point, we refer to [Bouzebda and Zari \[2013\]](#), [Bouzebda \[2011, 2014\]](#), [Bouzebda and Keziou \[2010a,b\]](#).

Example 4.5.10.6 (Bergsma-Dassios-Yanagimoto's τ^*). [Bergsma and Dassios \[2014\]](#) introduced a rank correlation statistic as a U-statistic of order 4 with the symmetric kernel

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

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

Example 4.5.10.7 The sample covariance matrix $\hat{S}_n = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)(X_i - \bar{X}_n)^\top$ is an unbiased estimator of the covariance matrix $\Sigma = \text{Cov}(X_1)$. Here, \hat{S}_n is a matrix-valued U-statistic with the quadratic kernel $h(x_1, x_2) = (x_1 - x_2)(x_1 - x_2)^\top / 2$ for $x_1, x_2 \in \mathbb{R}^p$.

Example 4.5.10.8 Two generic vectors $y = (y_1, y_2)$ and $z = (z_1, z_2)$ in \mathbb{R}^2 are said to be concordant if $(y_1 - z_1)(y_2 - z_2) > 0$. For $m, k = 1, \dots, p$, define

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

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

Example 4.5.10.9 The Gini mean difference. The Gini index provides another popular measure of dispersion. It corresponds to the case where $E \subset \mathbb{R}$ and $h(x, y) = |x - y|$:

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

Example 4.5.10.10 The Wilcoxon Statistic. Suppose that $E \subset \mathbb{R}$ is symmetric around zero. As an estimate of the quantity $\int_{(x,y) \in E^2} \{2\mathbb{1}_{\{x+y>0\}} - 1\} dF(x)dF(y)$, it is pertinent to consider the statistic

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

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

Example 4.5.10.11 The Takens estimator. Suppose that $E \subset \mathbb{R}^d$, $d \geq 1$. Denote by $\|\cdot\|$ the usual Euclidean norm on \mathbb{R}^d . In [Borovkova et al. \[1999\]](#), the following estimate of the correlation integral,

$$C_F(r) = \int \mathbb{1}_{\{\|x-x'\| \leq r\}} dF(x)dF(x'), \quad r > 0,$$

is considered:

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

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

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

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

4.6 The bandwidth selection criterion

Following [Dony and Mason \[2008\]](#) and [Bouzebda and Nemouchi \[2022\]](#), the leave-one-out cross validation procedure allows to define, for any fixed $\mathbf{i} = (i_1, \dots, i_k) \in I(k, n)$:

$$\hat{r}_{n,\mathbf{i}}^{(k)}(\varphi, \mathbf{x}; m_n) = \frac{\sum_{(j_1, \dots, j_k) \in I(k, n)(\mathbf{i})} \varphi(Y_{j_1}, \dots, Y_{j_k}) \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{j_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{j_k})}{\sum_{(i_1, \dots, i_k) \in I(k, n)} \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k})}, \quad (4.6.1)$$

where

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

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

$$\text{CV}(\varphi, h) := \frac{(n-k)!}{n!} \sum_{\mathbf{i} \in I(k, n)} \left(\varphi(\mathbf{Y}_{\mathbf{i}}) - \hat{r}_{n,\mathbf{i}}^{(k)}(\varphi, \mathbf{X}_{\mathbf{i}}; m_n) \right)^2 \widetilde{\mathcal{W}}(\mathbf{X}_{\mathbf{i}}), \quad (4.6.2)$$

where

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

A natural way for choosing the bandwidth is to minimize the precedent criterion, so let's choose \hat{m}_n minimizing

$$\sup_{\varphi \in \mathcal{F}} \text{CV}(\varphi, h).$$

One can replace (4.6.2) by

$$\text{CV}(\varphi, h) := \frac{(n-k)!}{n!} \sum_{\mathbf{i} \in I(k, n)} \left(\varphi(\mathbf{Y}_{\mathbf{i}}) - \hat{r}_{n,\mathbf{i}}^{(k)}(\varphi, \mathbf{X}_{\mathbf{i}}; m_n) \right)^2 \widehat{\mathcal{W}}(\mathbf{X}_{\mathbf{i}}, \mathbf{t}), \quad (4.6.3)$$

where

$$\widehat{\mathcal{W}}(\mathbf{s}, \mathbf{x}) := \prod_{i=1}^k \widehat{W}(\mathbf{s}_i, \mathbf{t}_i).$$

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

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

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

Remark 4.6.0.1 *We can use a different bandwidth criterion suggested by [Silverman \[1986\]](#), the rule of thumb. Strictly speaking, since the cross-validated bandwidth m_n is random, the asymptotic theory can only be justified with this random bandwidth via a specific stochastic equicontinuity argument. For testing a parametric model for conditional mean function against*

a nonparametric alternative, [Horowitz and Spokoiny \[2001\]](#) proposed an adaptive-rate-optimal rule. [Gao and Gijbels \[2008\]](#) presents the other method for selecting a proper bandwidth. [Gao and Gijbels \[2008\]](#) propose, utilizing the Edgeworth expansion of the asymptotic distribution of the test, to select the bandwidth such that the power function of the test problem is maximized while the size function is controlled. Although any choice of bandwidth m_n that satisfies the assumption will produce the asymptotic distribution in [Theorem 4.3.4.1](#), we need guidance on choosing m_n in practice. Idealistically, we should choose a m_n that provides the greatest power (e.g., test based on the Kendall tau) or small MSE for a given sample size, but deriving this procedure is complicated enough to warrant a separate study. Future investigation will focus on the three approaches mentioned above.

4.7 Mathematical development

This section is devoted to the proof of our results. The aforementioned notation is also used in what follows.

4.7.1 Proofs of the results of [Section 4.3.2](#)

In this section, we follow the steps of [Derumigny \[2019\]](#) while making the necessary changes to fit our extended framework. Let us start with the proof of [Lemma 4.3.2.4](#).

Proof of [Lemma 4.3.2.4](#)

We first remark that

$$\begin{aligned}
 & \left| \mathbb{E}[u_n(1, \tilde{\mathbf{x}}; m_n)] - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right| \\
 &= \left| \frac{(n-k)!}{n!} \mathbb{E} \left[\sum_{\mathbf{i} \in \mathbb{I}(k, n)} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{X}_{i_j}) \right] - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right| \\
 &= \left| \mathbb{E} \left[\prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{X}_j) \right] - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right| \\
 &= \left| \int_{\mathbb{R}^{dk}} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j) \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{t}_j) \prod_{j=1}^k d\mathbf{t}_j - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right| \tag{4.7.1}
 \end{aligned}$$

$$= \left| \int_{\mathbb{R}^{dk}} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j) \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{t}_j) \prod_{j=1}^k d\mathbf{t}_j - \int_{\mathbb{R}^{dk}} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j) \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \prod_{j=1}^k d\mathbf{t}_j \right| \tag{4.7.2}$$

From (4.7.1) to (4.7.2), we have used the condition (i) of Definition 4.2.1.2. Consequently, we infer that

$$\begin{aligned}
 & \left| \mathbb{E}[u_n(1, \tilde{\mathbf{x}}; m_n)] - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right| \\
 &= \left| \int_{\mathbb{R}^{dk}} \left\{ \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{t}_j) - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right\} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j) \prod_{j=1}^k d\mathbf{t}_j \right| \\
 &= \left| \int_{\|\mathbf{t}_1 - \mathbf{x}_1\| < \eta} \cdots \int_{\|\mathbf{t}_k - \mathbf{x}_k\| < \eta} \left\{ \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{t}_j) - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right\} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j) \prod_{j=1}^k d\mathbf{t}_j \right. \\
 &\quad \left. + \int_{\|\mathbf{t}_1 - \mathbf{x}_1\| \geq \eta} \cdots \int_{\|\mathbf{t}_k - \mathbf{x}_k\| \geq \eta} \left\{ \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{t}_j) - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right\} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j) \prod_{j=1}^k d\mathbf{t}_j \right| \\
 &\leq \left| \int_{\|\mathbf{t}_1 - \mathbf{x}_1\| < \eta} \cdots \int_{\|\mathbf{t}_k - \mathbf{x}_k\| < \eta} \left\{ \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{t}_j) - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right\} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j) \prod_{j=1}^k d\mathbf{t}_j \right| \\
 &\quad + \left| \int_{\|\mathbf{t}_1 - \mathbf{x}_1\| \geq \eta} \cdots \int_{\|\mathbf{t}_k - \mathbf{x}_k\| \geq \eta} \left\{ \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{t}_j) - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right\} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j) \prod_{j=1}^k d\mathbf{t}_j \right| \\
 &= \mathcal{B}_1 + \mathcal{B}_2.
 \end{aligned}$$

By the change of variables, we readily obtain for $\mathbf{u}_j = \mathbf{t}_j - \mathbf{x}_j$, $j = 1, \dots, k$:

$$\begin{aligned}
 \mathcal{B}_1 &= \left| \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \left\{ \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{t}_j) - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right\} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j) \prod_{j=1}^k d\mathbf{t}_j \right| \\
 &= \left| \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \left\{ \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j + \mathbf{u}_j) - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right\} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j) \prod_{j=1}^k d\mathbf{u}_j \right| \\
 &= \left| \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \left\{ \Phi_{\mathbf{x}, \mathbf{u}}(1) - \Phi_{\mathbf{x}, \mathbf{u}}(0) \right\} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j) \prod_{j=1}^k d\mathbf{u}_j \right|,
 \end{aligned}$$

where, for $h \in [-1, 1]$,

$$\Phi_{\mathbf{x}, \mathbf{u}}(h) := \prod_{i=1}^k f_{\mathbf{X}}(\mathbf{x}_i + h\mathbf{u}_i).$$

Observe that the function $\Phi_{\mathbf{x}, \mathbf{u}}(\cdot)$ has at least the same regularity as $f_{\mathbf{X}}(\cdot)$, which means that it is α -differentiable. By a Taylor-Lagrange expansion, we obtain

$$\mathcal{B}_1 = \left| \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \left(\sum_{l=1}^{\alpha-1} \frac{1}{l!} \Phi_{\mathbf{x}, \mathbf{u}}^{(l)}(0) + \frac{1}{\alpha!} \Phi_{\mathbf{x}, \mathbf{u}}^{(\alpha)}(h_{\mathbf{x}, \mathbf{u}}) \right) \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j) \prod_{j=1}^k d\mathbf{u}_j \right| \leq \mathcal{B}'_1 + \mathcal{B}''_1,$$

where

$$\mathcal{B}'_1 = \left| \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \left(\sum_{l=1}^{\alpha-1} \frac{1}{l!} \Phi_{\mathbf{x}, \mathbf{u}}^{(l)}(0) \right) \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j) \prod_{j=1}^k d\mathbf{u}_j \right|,$$

and

$$\mathcal{B}''_1 = \left| \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \left(\frac{1}{\alpha!} \Phi_{\mathbf{x}, \mathbf{u}}^{(\alpha)}(h_{\mathbf{x}, \mathbf{u}}) \right) \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j) \prod_{j=1}^k d\mathbf{u}_j \right|.$$

Let us start by studying the term \mathcal{B}'_1 . For $l > 0$, we observe that

$$\begin{aligned}\Phi_{\mathbf{x}, \mathbf{u}}^{(l)}(0) &= \sum_{\alpha_1 + \dots + \alpha_k = l} \binom{\alpha}{\alpha_{1:k}} \prod_{j=1}^k \frac{\partial^{\alpha_j} (f_X(\mathbf{x}_j + h\mathbf{u}_j))}{\partial h^{\alpha_j}}(0) \\ &= \sum_{\alpha_1 + \dots + \alpha_k = l} \binom{\alpha}{\alpha_{1:k}} \prod_{j=1}^k \sum_{i_1, \dots, i_{\alpha_j} = 1}^d u_{j, i_1} \dots u_{j, i_{\alpha_j}} \frac{\partial^{\alpha_j} f_X}{\partial x_{i_1} \dots \partial x_{i_{\alpha_j}}}(\mathbf{x}_j + h_{\mathbf{x}, \mathbf{u}} \mathbf{u}_j),\end{aligned}$$

where

$$\binom{\alpha}{\alpha_{1:k}} := \alpha! / \left(\prod_{j=1}^k (\alpha_j!) \right),$$

is the multinomial coefficient. Thus, we readily infer

$$\begin{aligned}\mathcal{B}'_1 &= \left| \int_{\|\mathbf{u}_1\| < \eta} \dots \int_{\|\mathbf{u}_k\| < \eta} \left(\sum_{l=1}^{\alpha-1} \frac{1}{l!} \Phi_{\mathbf{x}, \mathbf{u}}^{(l)}(0) \right) \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j) \prod_{j=1}^k d\mathbf{u}_j \right| \\ &= \left| \int_{\|\mathbf{u}_1\| < \eta} \dots \int_{\|\mathbf{u}_k\| < \eta} \sum_{l=1}^{\alpha-1} \frac{1}{l!} \left\{ \sum_{\alpha_1 + \dots + \alpha_k = l} \binom{\alpha}{\alpha_{1:k}} \prod_{j=1}^k \right. \right. \\ &\quad \times \left. \sum_{i_1, \dots, i_{\alpha_j} = 1}^d u_{j, i_1} \dots u_{j, i_{\alpha_j}} \frac{\partial^{\alpha_j} f_X}{\partial x_{i_1} \dots \partial x_{i_{\alpha_j}}}(\mathbf{x}_j + h_{\mathbf{x}, \mathbf{u}} \mathbf{u}_j) \right\} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j) \prod_{j=1}^k d\mathbf{u}_j \Big| \\ &\leq \int_{\|\mathbf{u}_1\| < \eta} \dots \int_{\|\mathbf{u}_k\| < \eta} \sum_{l=1}^{\alpha-1} \frac{1}{l!} \left\{ \sum_{\alpha_1 + \dots + \alpha_k = l} \binom{\alpha}{\alpha_{1:k}} \prod_{j=1}^k \right. \\ &\quad \times \left. \sum_{i_1, \dots, i_{\alpha_j} = 1}^d \left| u_{j, i_1} \dots u_{j, i_{\alpha_j}} \right| \left| \frac{\partial^{\alpha_j} f_X}{\partial x_{i_1} \dots \partial x_{i_{\alpha_j}}}(\mathbf{x}_j + h_{\mathbf{x}, \mathbf{u}} \mathbf{u}_j) \right| \right\} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j) \prod_{j=1}^k d\mathbf{u}_j \\ &\leq \int_{\|\mathbf{u}_1\| < \eta} \dots \int_{\|\mathbf{u}_k\| < \eta} \sum_{l=1}^{\alpha-1} \frac{1}{l!} \left\{ \sum_{\alpha_1 + \dots + \alpha_k = l} \binom{\alpha}{\alpha_{1:k}} \prod_{j=1}^k \right. \\ &\quad \times \left. \sum_{i_1, \dots, i_{\alpha_j} = 1}^d \left| u_{j, i_1} \dots u_{j, i_{\alpha_j}} \right| \sup_{h \in [0, 1]} \left| \frac{\partial^{\alpha_j} f_X}{\partial x_{i_1} \dots \partial x_{i_{\alpha_j}}}(\mathbf{x}_j + h_{\mathbf{x}, \mathbf{u}} \mathbf{u}_j) \right| \right\} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j) \prod_{j=1}^k d\mathbf{u}_j \\ &\leq C_\alpha \int_{\|\mathbf{u}_1\| < \eta} \dots \int_{\|\mathbf{u}_k\| < \eta} \sum_{l=1}^{\alpha-1} \frac{1}{l!} \left\{ \sum_{\alpha_1 + \dots + \alpha_k = l} \binom{\alpha}{\alpha_{1:k}} \prod_{j=1}^k \sum_{i_1, \dots, i_{\alpha_j} = 1}^d \left| u_{j, i_1} \dots u_{j, i_{\alpha_j}} \right| \right\} \\ &\quad \times \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j) \prod_{j=1}^k d\mathbf{u}_j, \tag{4.7.4}\end{aligned}$$

where for the passage from (4.7.3) to (4.7.4) we have used the condition (C.1). In addition, we can find a positive constant $C_{d, \alpha}$ such that

$$\sum_{l=1}^{\alpha-1} \frac{1}{l!} \sum_{\alpha_1 + \dots + \alpha_k = l} \binom{\alpha}{\alpha_{1:k}} \prod_{j=1}^k \sum_{i_1, \dots, i_{\alpha_j} = 1}^d \left| u_{j, i_1} \dots u_{j, i_{\alpha_j}} \right| \leq C_{d, \alpha} \prod_{j=1}^k \|\mathbf{u}_j\|^d,$$

implying that

$$\begin{aligned}
 \mathcal{B}'_1 &\leq C_\alpha C_{d,\alpha} \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \prod_{j=1}^k \|\mathbf{u}_j\|^d \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j) \prod_{j=1}^k d\mathbf{u}_j \\
 &\leq C_\alpha C_{d,\alpha} \int_{\|\mathbf{u}_2\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \prod_{j=2}^k \|\mathbf{u}_j\|^d \prod_{j=2}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j) \\
 &\quad \times \left\{ \int_{\|\mathbf{u}_1\| < \eta} \|\mathbf{u}_1\|^d \delta_{m_n}(\mathbf{x}_1, \mathbf{x}_1 + \mathbf{u}_1) d\mathbf{u}_1 \right\} d\mathbf{u}_2 \cdots d\mathbf{u}_k.
 \end{aligned}$$

Let us denote

$$\mathcal{J}_1 := \int_{\|\mathbf{u}_1\| < \eta} \|\mathbf{u}_1\|^d \delta_{m_n}(\mathbf{x}_1, \mathbf{x}_1 + \mathbf{u}_1) d\mathbf{u}_1.$$

Now, following the same method of [Susarla and Walter \[1981\]](#) in our extended setting, by a change of variables to the polar coordinates, we obtain

$$\mathcal{J}_1 \leq \int_0^\eta \epsilon_m(\mathbf{x}_1, r) r^d r^{d-1} dr = \left\{ \int_0^{1/m} + \int_{1/m}^\eta \right\} = \mathcal{J}'_1 + \mathcal{J}''_1,$$

for sufficiently small η and $1 \leq m\eta$ and

$$\epsilon_m(\mathbf{x}_1, r) = \int_{\mathcal{S}_{\mathbf{x}_1}^{d-1}} \delta_{m_n}(\mathbf{x}_1, \mathbf{x}_1 + r\mathbf{t}'_1) d\delta(\mathbf{t}'_1). \quad (4.7.5)$$

Here $\mathcal{S}_{\mathbf{x}_1}^{d-1}$ denotes the surface of the unit sphere with center at \mathbf{x}_1 and $d\delta(\mathbf{t}'_1)$ represents the elements of the surface area on $\mathcal{S}_{\mathbf{x}_1}^{d-1}$. Using the condition (ii) of Definition 4.2.1.2 yields to

$$\int_r^\infty \epsilon_m(\mathbf{x}_1, s) s^{d-1} ds = O(m^{-1} r^{-1}) \text{ uniformly in } r > 0. \quad (4.7.6)$$

By condition (iii) of Definition 4.2.1.2, we have

$$\mathcal{J}'_1 = O(m_n^{-d}).$$

By (4.7.5), an integration by parts gives

$$\begin{aligned}
 \mathcal{J}''_1 &= \left\{ r^d \int_r^\infty \epsilon_m(\mathbf{x}_1, s) s^{d-1} ds \right\}_{1/m}^\eta + d \int_{1/m}^\eta r^{d-1} \left(\int_r^\infty \epsilon_m(\mathbf{x}_1, s) s^{d-1} ds \right) dr \\
 &\leq O(m^{-1}) + O(m_n^{-d}) + C_1 m^{-1} \int_{1/m}^\eta r^{d-2} dr \\
 &= O(m^{-1}) + O(m_n^{-d}).
 \end{aligned}$$

Consequently we have

$$\mathcal{J}_1 = O(m^{-1}).$$

In the same way, we will calculate

$$\mathcal{J}_2 := \int_{\|\mathbf{u}_2\| < \eta} \mathcal{J}_1 \|\mathbf{u}_2\|^d \delta_{m_n}(\mathbf{x}_2, \mathbf{x}_2 + \mathbf{u}_2) d\mathbf{u}_2.$$

By the fact that the term \mathcal{I}_1 does not depend on \mathbf{u}_2 , then \mathcal{I}_2 becomes

$$\mathcal{I}_2 = \mathcal{I}_1 \int_{\|\mathbf{u}_2\| < \eta} \|\mathbf{u}_2\|^d \delta_{m_n}(\mathbf{x}_2, \mathbf{x}_2 + \mathbf{u}_2) d\mathbf{u}_2,$$

implying that

$$\mathcal{I}_2 = O(m^{-2}).$$

By repeating the same reasoning for the remaining integrals, we readily obtain

$$\mathcal{B}'_1 = C_\alpha C_{d,\alpha} \prod_{j=1}^k O(m^{-1}) = O(m^{-k}).$$

We now consider the term \mathcal{B}''_1 . We have

$$\begin{aligned} \mathcal{B}''_1 &= \left| \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \frac{1}{\alpha!} \phi_{\mathbf{x}, \mathbf{u}}^{(\alpha)}(h_{\mathbf{x}, \mathbf{u}}) \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j) \prod_{j=1}^k d\mathbf{u}_j \right| \\ &\leq \frac{1}{\alpha!} \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \left\{ \sum_{\alpha_1 + \cdots + \alpha_k = \alpha} \binom{n}{\alpha_{1:k}} \prod_{j=1}^k \sum_{i_1, \dots, i_{\alpha_j}=1}^d |u_{j,i_1} \cdots u_{j,i_{\alpha_j}}| \right. \\ &\quad \times \sup_{h \in [0,1]} \left| \frac{\partial^{\alpha_j} f_{\mathbf{x}}}{\partial x_{i_1} \cdots \partial x_{i_{\alpha_j}}}(\mathbf{x}_j + h_{\mathbf{x}, \mathbf{u}} \mathbf{u}_j) \right| \left. \right\} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j) \prod_{j=1}^k d\mathbf{u}_j \\ &\leq \frac{C_\alpha}{\alpha!} \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \left\{ \sum_{\alpha_1 + \cdots + \alpha_k = \alpha} \binom{n}{\alpha_{1:k}} \prod_{j=1}^k \sum_{i_1, \dots, i_{\alpha_j}=1}^d |u_{j,i_1} \cdots u_{j,i_{\alpha_j}}| \right\} \\ &\quad \times \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j) \prod_{j=1}^k d\mathbf{u}_j. \end{aligned}$$

Similarly, we can find a positive constant $C'_{d,\alpha}$ in such a way that

$$\sum_{\alpha_1 + \cdots + \alpha_k = \alpha} \binom{n}{\alpha_{1:k}} \prod_{j=1}^k \sum_{i_1, \dots, i_{\alpha_j}=1}^d |u_{j,i_1} \cdots u_{j,i_{\alpha_j}}| \leq C'_{d,\alpha} \prod_{j=1}^k \|\mathbf{u}_j\|^d,$$

which, by following the same steps to investigate the term \mathcal{B}'_1 , readily implies that

$$\mathcal{B}''_1 \leq \frac{C_\alpha C'_{d,\alpha}}{\alpha!} \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \prod_{j=1}^k \|\mathbf{u}_j\|^d \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j) d\mathbf{u}_k = O(m^{-k}).$$

Consequently, we get

$$\mathcal{B}_1 = O(m^{-k}).$$

Now, let us study \mathcal{B}_2 . We have

$$\begin{aligned}
 \mathcal{B}_2 &= \left| \int_{\|\mathbf{t}_1 - \mathbf{x}_1\| \geq \eta} \cdots \int_{\|\mathbf{t}_k - \mathbf{x}_k\| \geq \eta} \left\{ \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{t}_j) - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right\} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j) \prod_{j=1}^k d\mathbf{t}_j \right| \\
 &\leq \int_{\|\mathbf{t}_1 - \mathbf{x}_1\| \geq \eta} \cdots \int_{\|\mathbf{t}_k - \mathbf{x}_k\| \geq \eta} \left| \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{t}_j) \right| \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j) \prod_{j=1}^k d\mathbf{t}_j \\
 &\quad + \int_{\|\mathbf{t}_1 - \mathbf{x}_1\| \geq \eta} \cdots \int_{\|\mathbf{t}_k - \mathbf{x}_k\| \geq \eta} \left| \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right| \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j) \prod_{j=1}^k d\mathbf{t}_j \\
 &\leq \int_{\|\mathbf{t}_1\| \geq \eta} \cdots \int_{\|\mathbf{t}_k\| \geq \eta} \left| \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j + \mathbf{t}_j) \right| \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j + \mathbf{x}_j) \prod_{j=1}^k d\mathbf{t}_j \\
 &\quad + \int_{\|\mathbf{t}_1\| \geq \eta} \cdots \int_{\|\mathbf{t}_k\| \geq \eta} \left| \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right| \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{t}_j) \prod_{j=1}^k d\mathbf{t}_j.
 \end{aligned}$$

By applying Hölder's inequality, we readily get

$$\begin{aligned}
 &\int_{\|\mathbf{t}_1\| \geq \eta} \cdots \int_{\|\mathbf{t}_k\| \geq \eta} \left| \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j + \mathbf{t}_j) \right| \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{t}_j) \prod_{j=1}^k d\mathbf{t}_j \\
 &\quad + \int_{\|\mathbf{t}_1\| \geq \eta} \cdots \int_{\|\mathbf{t}_k\| \geq \eta} \left| \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right| \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{t}_j) \prod_{j=1}^k d\mathbf{t}_j \\
 &\leq \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \left| \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j + \mathbf{t}_j) \right| \prod_{j=1}^k \mathbb{1}_{\{\mathbf{t}_j \in \mathbb{R}^d; \|\mathbf{t}_j\| > \eta\}} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{t}_j) \prod_{j=1}^k d\mathbf{t}_j \\
 &\quad + \left| \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right| \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \prod_{j=1}^k \mathbb{1}_{\{\mathbf{t}_j \in \mathbb{R}^d; \|\mathbf{t}_j\| > \eta\}} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{t}_j) \prod_{j=1}^k d\mathbf{t}_j \\
 &\leq \left\| \prod_{j=1}^k f_{\mathbf{X}} \right\|_p \left\| \prod_{j=1}^k \Psi_{\eta, j} \delta_{m_n} \right\|_q + \left| \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right| \left\| \prod_{j=1}^k \Psi_{\eta, j} \delta_{m_n} \right\|_1 \\
 &\leq \prod_{j=1}^k \|f_{\mathbf{X}}\|_p \|\Psi_{\eta, j} \delta_{m_n}\|_q + \left| \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right| \left\| \prod_{j=1}^k \Psi_{\eta, j} \delta_{m_n} \right\|_1 \\
 &\leq \prod_{j=1}^k \|f_{\mathbf{X}}\|_p \|\Psi_{\eta, j} \delta_{m_n}\|_{\infty}^{\frac{1}{p}} \|\Psi_{\eta, j} \delta_{m_n}\|_1^{\frac{1}{q}} + \left| \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right| \left\| \prod_{j=1}^k \Psi_{\eta, j} \delta_{m_n} \right\|_1,
 \end{aligned}$$

where

$$\Psi_{\eta, j} := \mathbb{1}_{\{\mathbf{t}_j \in \mathbb{R}^d; \|\mathbf{t}_j\| > \eta\}}.$$

Making use of the fact that $f_{\mathbf{X}} \in L^p(\mathbb{R}^d)$ and by the condition (iv) Definition 4.2.1.2, for $j = 1, \dots, k$, we have

$$\|\Psi_{\eta, j} \delta_{m_n}\|_{\infty} = O(1),$$

and

$$\|\Psi_{\eta, j} \delta_{m_n}\|_1 = \int_{\|\mathbf{t}_j\| > \eta} \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{t}_j) d\mathbf{t}_j = O(m^{-1} \eta^{-1}).$$

Hence we conclude that

$$\mathcal{B}_2 = O(m^{-k/q}).$$

Finally, we have proved that

$$\left| \mathbb{E}[u_n(1, \tilde{\mathbf{x}}; m_n)] - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right| = \mathcal{B}_1 + \mathcal{B}_2 = O(m^{-k}) + O(m^{-k/q}).$$

By denoting $\lambda := \min\{k, kq^{-1}\}$, the last equation is rewritten as

$$\left| \mathbb{E}[u_n(1, \tilde{\mathbf{x}}; m_n)] - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right| = O(m^{-\lambda}). \quad (4.7.7)$$

Hence, the proof of the lemma is complete. \square

Proof of Lemma 4.3.2.5

To obtain our result, one usually writes the difference between the estimator $u_n(1, \tilde{\mathbf{x}}; m_n)$ and $\prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j)$, as the sum of a stochastic term and the so-called bias. We have

$$u_n(1, \tilde{\mathbf{x}}; m_n) - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) = (u_n(1, \tilde{\mathbf{x}}; m_n) - \mathbb{E}[u_n(1, \tilde{\mathbf{x}}; m_n)]) + \left(\mathbb{E}[u_n(1, \tilde{\mathbf{x}}; m_n)] - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right).$$

In order to prove Lemma 4.3.2.5, we have to study the two quantities in the last equation. Since the bias part is given in Lemma 4.3.2.4, we only need to bound the stochastic part. We have

$$|u_n(1, \tilde{\mathbf{x}}; m_n) - \mathbb{E}[u_n(1, \tilde{\mathbf{x}}; m_n)]| = \frac{(n-k)!}{n!} \sum_{\mathbf{i} \in I(k, n)} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{X}_{i_j}) - \mathbb{E} \left[\prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{X}_j) \right].$$

Then, we can apply Lemma 4.8.0.2 to the function $\tilde{G}_{1, \tilde{\mathbf{x}}, m_n}(\mathbf{X}^{(k)}, \mathbf{Y}^{(k)})$ defined in (4.2.5).

We have

$$\begin{aligned} \mathbb{E} \left[\tilde{G}_{1, \tilde{\mathbf{x}}, m_n}(\mathbf{X}^{(k)}, \mathbf{Y}^{(k)}) \right] &= \prod_{j=1}^k \mathbb{E} [\delta_{m_n}(\mathbf{x}_j, \mathbf{X}_j)] \\ &= \prod_{j=1}^k \int_{\mathbb{R}^d} \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j) f_{\mathbf{X}}(\mathbf{t}_j) d\mathbf{t}_j \\ &\leq \prod_{j=1}^k \int_{\mathbb{R}^d} \left(\sup_{\mathbf{t}_j \in \mathbb{R}^d} \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j) \right) f_{\mathbf{X}}(\mathbf{t}_j) d\mathbf{t}_j \end{aligned} \quad (4.7.8)$$

$$\leq \prod_{j=1}^k O(m^d) \int_{\mathbb{R}^d} f_{\mathbf{X}}(\mathbf{t}_j) d\mathbf{t}_j \quad (4.7.9)$$

$$\leq O(m^{dk}). \quad (4.7.10)$$

The passage from (4.7.8) to (4.7.9) is by the condition (iii) of the Definition 4.2.1.2 and we easily obtain (4.7.10) by taking into account that that $f_{\mathbf{X}}(\cdot)$ is a density function. For the variance, we have

$$\begin{aligned}
 \text{Var} \left[\bar{G}_{1, \tilde{\mathbf{x}}, m_n}(\mathbf{X}^{(k)}, \mathbf{Y}^{(k)}) \right] &\leq \mathbb{E} \left[\bar{G}_{1, \tilde{\mathbf{x}}, m_n}(\mathbf{X}^{(k)}, \mathbf{Y}^{(k)})^2 \right] = \prod_{j=1}^k \mathbb{E} \left[\delta_{m_n}(\mathbf{x}_j, \mathbf{X}_j)^2 \right] \\
 &= \prod_{j=1}^k \int_{\mathbb{R}^d} \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j)^2 f_{\mathbf{X}}(\mathbf{t}_j) d\mathbf{t}_j \\
 &\leq \prod_{j=1}^k \int_{\mathbb{R}^d} \left(\sup_{\mathbf{t}_j \in \mathbb{R}^d} \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j) \right)^2 f_{\mathbf{X}}(\mathbf{t}_j) d\mathbf{t}_j \\
 &= \prod_{j=1}^k \left(O(m^d) \right)^2 \int_{\mathbb{R}^d} f_{\mathbf{X}}(\mathbf{t}_j) d\mathbf{t}_j \\
 &\leq \prod_{j=1}^k O(m^{2d}) = O(m^{2dk}). \tag{4.7.11}
 \end{aligned}$$

Using the fact that $\delta_m \in L^\infty(\mathbb{R}^d \times \mathbb{R}^d)$, then there exists a constant $C_\delta > 0$ in such a way that

$$0 \leq \prod_{j=1}^k \sup_{\substack{\mathbf{x}_j \in \mathbb{R}^d \\ \mathbf{y}_j \in \mathbb{R}^d}} \delta_{m_n}(\mathbf{x}_j, \mathbf{y}_j) \leq C_\delta^k < \infty.$$

This when combined with (4.7.10) and (4.7.11) respectively, implies that there exist constants $C_1, C_2 > 0$ in such a way that

$$\mathbb{E} \left[\bar{G}_{1, \tilde{\mathbf{x}}, m_n}(\mathbf{X}^{(k)}, \mathbf{Y}^{(k)}) \right] \leq C_1 m^{dk} =: \kappa_1,$$

and

$$\text{Var} \left[\bar{G}_{1, \tilde{\mathbf{x}}, m_n}(\mathbf{X}^{(k)}, \mathbf{Y}^{(k)})^2 \right] \leq C_2 m^{2dk} =: \kappa_2.$$

An application of Lemma 4.8.0.2 gives

$$\mathbb{P} \{ u_n(1, \tilde{\mathbf{x}}; m_n) - \mathbb{E}[u_n(1, \tilde{\mathbf{x}}; m_n)] \geq t \} \leq \exp \left(- \frac{[n/k] t^2}{2\kappa_2 + \frac{2}{3} \{ C_\delta^k - \kappa_1 \} t} \right). \tag{4.7.12}$$

Finally, the proof is complete by combining the Lemma 4.3.2.4 with (4.7.12). \square

Proof of Theorem 4.3.2.1

Notice that

$$\begin{aligned}
 & \left| \widehat{r}_n^{(k)}(\varphi, \tilde{\mathbf{x}}, m_n) - r^{(k)}(\varphi, \tilde{\mathbf{x}}) \right| \\
 &= \left| \frac{\sum_{\mathbf{i} \in I(k, n)} \varphi(\mathbf{Y}_{i_1}, \dots, \mathbf{Y}_{i_k}) \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{X}_{i_j})}{\sum_{\mathbf{i} \in I(k, n)} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{X}_{i_j})} - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \right| \\
 &= \frac{1}{\sum_{\mathbf{i} \in I(k, n)} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{X}_{i_j})} \left| \sum_{\mathbf{i} \in I(k, n)} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{X}_{i_j}) \{ \varphi(\mathbf{Y}_{i_1}, \dots, \mathbf{Y}_{i_k}) \right. \\
 &\quad \left. - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \right| \\
 &= \frac{\prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j)}{\sum_{\mathbf{i} \in I(k, n)} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{X}_{i_j})} \left| \sum_{\mathbf{i} \in I(k, n)} \prod_{j=1}^k \frac{\delta_{m_n}(\mathbf{x}_j, \mathbf{X}_{i_j})}{f_{\mathbf{X}}(\mathbf{x}_j)} \{ \varphi(\mathbf{Y}_{i_1}, \dots, \mathbf{Y}_{i_k}) \right. \\
 &\quad \left. - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \right| \\
 &= \frac{\prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j)}{\sum_{\mathbf{i} \in I(k, n)} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{X}_{i_j})} \left| \sum_{\mathbf{i} \in I(k, n)} S_{\mathbf{i}} \right| \\
 &= \frac{\prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j)}{u_n(1, \tilde{\mathbf{x}}; m_n)} \left| \sum_{\mathbf{i} \in I(k, n)} S_{\mathbf{i}} \right|.
 \end{aligned}$$

In order to obtain the desired result, we need to bound separately $\prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) / u_n(1, \tilde{\mathbf{x}}, m_n)$, the bias

$$\text{term } \left| \sum_{\mathbf{i} \in I(k, n)} \mathbb{E}[S_{\mathbf{i}}] \right| \text{ and the stochastic part } \left| \sum_{\mathbf{i} \in I(k, n)} (S_{\mathbf{i}} - \mathbb{E}[S_{\mathbf{i}}]) \right|.$$

Proof of Lemma 4.3.2.6

Making of the condition (C.2.b), we have

$$\left| \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right| \geq |u_n(1, \tilde{\mathbf{x}}, m_n)| - \left| u_n(1, \tilde{\mathbf{x}}, m_n) - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right| > f_{\mathbf{X}, \min}^k.$$

By applying Lemma 4.3.2.5, it follows that, there exists a constant $C_{\delta, \min} > 0$ such that

$$|u_n(1, \tilde{\mathbf{x}}, m_n)| > C_{\delta, \min},$$

We now apply the mean value inequality for the function $x \mapsto 1/x$, we get

$$\left| \frac{1}{u_n(1, \tilde{\mathbf{x}}, m_n)} - \frac{1}{\prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j)} \right| \leq \frac{1}{N_*^2} \left| u_n(1, \tilde{\mathbf{x}}, m_n) - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right|,$$

where N_* lies between $u_n(1, \tilde{\mathbf{x}}, m_n)$ and $\prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j)$. By applying Lemma 4.3.2.5, for some $t > 0$, we have

$$\mathbb{P} \left(\left| u_n(1, \tilde{\mathbf{x}}; m_n) - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right| \leq t + C_3 m^{-\lambda} \right) \geq 1 - \exp \left(- \frac{[n/k] t^2}{2\kappa_1 + \frac{2}{3} \{C_\delta^k - \kappa_2\} t} \right).$$

By choosing

$$t + C_3 m^{-\lambda} < \frac{f_{\mathbf{X}, \min}^k}{2},$$

it follows that

$$\left| u_n(1, \tilde{\mathbf{x}}, m_n) - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right| \leq \frac{1}{2} \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j).$$

We have

$$\frac{f_{\mathbf{X}, \min}^k}{2} \leq u_n(1, \tilde{\mathbf{x}}, m_n).$$

Notice that

$$\frac{f_{\mathbf{X}, \min}^k}{2} \leq \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j).$$

Thus, we have

$$\frac{f_{\mathbf{X}, \min}^k}{2} \leq N_*.$$

Combining the previous inequalities, we finally get

$$\left| \frac{1}{u_n(1, \tilde{\mathbf{x}}, m_n)} - \frac{1}{\prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j)} \right| \leq \frac{1}{N_*^2} \left| u_n(1, \tilde{\mathbf{x}}, m_n) - \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{x}_j) \right| \leq \frac{4}{f_{\mathbf{X}, \min}^{2k}} (C_3 m^{-\lambda} + t).$$

Hence, the proof is complete. \square

Proof of Lemma 4.3.2.7

Let us first remark that

$$\begin{aligned} 0 &= \int \{ \varphi(\mathbf{y}_1, \dots, \mathbf{y}_k) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \} \prod_{j=1}^k f_{\mathbf{Y}|\mathbf{X}_j=\mathbf{x}_j}(\mathbf{y}_j) \prod_{j=1}^k d\mu(\mathbf{y}_j) \\ &= \int \{ \varphi(\mathbf{y}_1, \dots, \mathbf{y}_k) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \} \prod_{j=1}^k \frac{f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}_j, \mathbf{y}_j)}{f_{\mathbf{X}}(\mathbf{x}_j)} \prod_{j=1}^k d\mu(\mathbf{y}_j). \end{aligned} \quad (7.13)$$

For all $\mathbf{i} \in I(k, n)$, we have

$$\begin{aligned}
 |\mathbb{E}[\mathbf{S}_i]| &= \left| \mathbb{E} \left[\prod_{j=1}^k \frac{\delta_{m_n}(\mathbf{x}_j, \mathbf{X}_{i_j})}{f_{\mathbf{X}}(\mathbf{x}_j)} \{ \varphi(\mathbf{Y}_{i_1}, \dots, \mathbf{Y}_{i_k}) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \} \right] \right| \\
 &= \left| \int_{\mathbb{R}^{dk} \times \mathbb{R}^{d'k}} \{ \varphi(\mathbf{y}_1, \dots, \mathbf{y}_k) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \} \right. \\
 &\quad \left. \times \prod_{j=1}^k \frac{\delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j)}{f_{\mathbf{X}}(\mathbf{x}_j)} f_{\mathbf{X}, \mathbf{Y}}(\mathbf{t}_j, \mathbf{y}_j) \prod_{j=1}^k d\mu(\mathbf{y}_j) \prod_{j=1}^k d\mathbf{t}_j \right| \\
 &= \left| \int_{\mathbb{R}^{dk} \times \mathbb{R}^{d'k}} \{ \varphi(\mathbf{y}_1, \dots, \mathbf{y}_k) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \} \right. \\
 &\quad \left. \times \left(\prod_{j=1}^k f_{\mathbf{X}, \mathbf{Y}}(\mathbf{t}_j, \mathbf{y}_j) - \prod_{j=1}^k f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}_j, \mathbf{y}_j) \right) \prod_{j=1}^k \frac{\delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j)}{f_{\mathbf{X}}(\mathbf{x}_j)} \prod_{j=1}^k d\mu(\mathbf{y}_j) \prod_{j=1}^k d\mathbf{t}_j \right|. \quad (7.14)
 \end{aligned}$$

Notice that (4.7.14) is concluded from (4.7.13) and the condition (i) of Definition 4.2.1.2. We have

$$\begin{aligned}
 |\mathbb{E}[\mathbf{S}_i]| &\leq \left| \int_{\mathbb{R}^{d'k}} \int_{\|\mathbf{u}_1\| < \eta} \dots \int_{\|\mathbf{u}_k\| < \eta} \{ \varphi(\mathbf{y}_1, \dots, \mathbf{y}_k) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \} \right. \\
 &\quad \left. \times \left(\prod_{j=1}^k f_{\mathbf{X}, \mathbf{Y}}(\mathbf{t}_j, \mathbf{y}_j) - \prod_{j=1}^k f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}_j, \mathbf{y}_j) \right) \prod_{j=1}^k \frac{\delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j)}{f_{\mathbf{X}}(\mathbf{x}_j)} \prod_{j=1}^k d\mu(\mathbf{y}_j) \prod_{j=1}^k d\mathbf{t}_j \right| \\
 &\quad + \left| \int_{\mathbb{R}^{d'k}} \int_{\|\mathbf{u}_1\| \geq \eta} \dots \int_{\|\mathbf{u}_k\| \geq \eta} \{ \varphi(\mathbf{y}_1, \dots, \mathbf{y}_k) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \} \right. \\
 &\quad \left. \times \left(\prod_{j=1}^k f_{\mathbf{X}, \mathbf{Y}}(\mathbf{t}_j, \mathbf{y}_j) - \prod_{j=1}^k f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}_j, \mathbf{y}_j) \right) \prod_{j=1}^k \frac{\delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j)}{f_{\mathbf{X}}(\mathbf{x}_j)} \prod_{j=1}^k d\mu(\mathbf{y}_j) \prod_{j=1}^k d\mathbf{t}_j \right| \\
 &= \mathcal{S}_{\varphi,1} + \mathcal{S}_{\varphi,2}.
 \end{aligned}$$

Let us start with $\mathcal{S}_{\varphi,1}$. We now apply the Taylor-Lagrange formula to the function

$$\Psi_{\mathbf{x}, \mathbf{u}}(h) := \prod_{i=1}^k f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}_i + h\mathbf{u}_i, \mathbf{y}_i)$$

to infer that

$$\begin{aligned}
 & \mathcal{S}_{\varphi,1} \\
 &= \left| \int_{\mathbb{R}^{d'k}} \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \left\{ \varphi(\mathbf{y}_1, \dots, \mathbf{y}_k) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \right\} \right. \\
 & \quad \times \left. \left(\Psi_{\mathbf{x}, \mathbf{u}}(1) - \Psi_{\mathbf{x}, \mathbf{u}}(0) \right) \prod_{j=1}^k \frac{\delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j)}{f_{\mathbf{X}}(\mathbf{x}_j)} d\mu(\mathbf{y}_j) \prod_{j=1}^k d\mathbf{u}_j \right| \\
 &= \left| \int_{\mathbb{R}^{d'k}} \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \left\{ \varphi(\mathbf{y}_1, \dots, \mathbf{y}_k) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \right\} \right. \\
 & \quad \times \left. \left(\sum_{i=1}^{\alpha-1} \frac{1}{i!} \Psi_{\mathbf{x}, \mathbf{u}}^{(i)}(0) + \frac{1}{\alpha!} \Psi_{\mathbf{x}, \mathbf{u}}^{(\alpha)}(h_{\mathbf{x}, \mathbf{u}}) \right) \prod_{j=1}^k \frac{\delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j)}{f_{\mathbf{X}}(\mathbf{x}_j)} \prod_{j=1}^k d\mu(\mathbf{y}_j) \prod_{j=1}^k d\mathbf{u}_j \right| \\
 &= \left| \int_{\mathbb{R}^{d'k}} \left\{ \varphi(\mathbf{y}_1, \dots, \mathbf{y}_k) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \right\} \right. \\
 & \quad \times \left. \left\{ \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \left(\sum_{i=1}^{\alpha-1} \frac{1}{i!} \Psi_{\mathbf{x}, \mathbf{u}}^{(i)}(0) + \frac{1}{\alpha!} \Psi_{\mathbf{x}, \mathbf{u}}^{(\alpha)}(h_{\mathbf{x}, \mathbf{u}}) \right) \prod_{j=1}^k \frac{\delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j)}{f_{\mathbf{X}}(\mathbf{x}_j)} \prod_{j=1}^k d\mathbf{u}_j \right\} \right. \\
 & \quad \times \left. \prod_{j=1}^k d\mu(\mathbf{y}_j) \right| \\
 &\leq \left| \int_{\mathbb{R}^{d'k}} \left\{ \varphi(\mathbf{y}_1, \dots, \mathbf{y}_k) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \right\} \right. \\
 & \quad \times \left. \left\{ \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \sum_{i=1}^{\alpha-1} \frac{1}{i!} \Psi_{\mathbf{x}, \mathbf{u}}^{(i)}(0) \prod_{j=1}^k \frac{\delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j)}{f_{\mathbf{X}}(\mathbf{x}_j)} \prod_{j=1}^k d\mathbf{u}_j \right\} \prod_{j=1}^k d\mu(\mathbf{y}_j) \right| \\
 & \quad + \left| \int_{\mathbb{R}^{d'k}} \left\{ \varphi(\mathbf{y}_1, \dots, \mathbf{y}_k) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \right\} \right. \\
 & \quad \times \left. \left\{ \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \frac{1}{\alpha!} \Psi_{\mathbf{x}, \mathbf{u}}^{(\alpha)}(h_{\mathbf{x}, \mathbf{u}}) \prod_{j=1}^k \frac{\delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j)}{f_{\mathbf{X}}(\mathbf{x}_j)} \prod_{j=1}^k d\mathbf{u}_j \right\} \prod_{j=1}^k d\mu(\mathbf{y}_j) \right| \\
 &= \mathcal{S}'_{\varphi,1} + \mathcal{S}''_{\varphi,1}.
 \end{aligned}$$

Making use of the condition **(C.2.b)**, we have

$$\begin{aligned}
 \mathcal{S}'_{\varphi,1} &\leq \frac{1}{f_{\mathbf{X}, \min}^k} \int_{\mathbb{R}^{d'k}} \left| \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \left\{ \varphi(\mathbf{y}_1, \dots, \mathbf{y}_k) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \right\} \right. \\
 & \quad \times \left. \sum_{i=1}^{\alpha-1} \frac{1}{i!} \Psi_{\mathbf{x}, \mathbf{u}}^{(i)}(0) \right\} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j) \prod_{j=1}^k d\mathbf{u}_j \prod_{j=1}^k d\mu(\mathbf{y}_j) \Big|.
 \end{aligned}$$

Similar to the proof of Lemma 4.3.2.4, we have that for $l > 0$

$$\begin{aligned}
 \psi^{(l)}(h) &= \sum_{\alpha_1 + \dots + \alpha_k = l} \binom{n}{\alpha_{1:k}} \prod_{j=1}^k \frac{\partial^{\alpha_j} (f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}_j + h\mathbf{u}_j, \mathbf{y}_j))}{\partial h^{\alpha_j}} \\
 &= \sum_{\alpha_1 + \dots + \alpha_k = l} \binom{\alpha}{\alpha_{1:k}} \prod_{j=1}^k \sum_{i_1, \dots, i_{\alpha_j} = 1}^d u_{j, i_1} \cdots u_{j, i_{\alpha_j}} \frac{\partial^{\alpha_j} f_{\mathbf{X}, \mathbf{Y}}}{\partial x_{i_1} \cdots \partial x_{i_{\alpha_j}}}(\mathbf{x}_j + h\mathbf{u}_j, \mathbf{y}_j),
 \end{aligned}$$

provided that, for all $j = 1, \dots, k$,

$$\sup_{h \in [0,1]} \left| \frac{\partial^{\alpha_j} f_{\mathbf{X}, \mathbf{Y}}}{\partial x_{i_1} \cdots \partial x_{i_{\alpha_j}}} (\mathbf{x}_j + h \mathbf{u}_j, \mathbf{y}_j) \right| \leq C_{\alpha, \psi}.$$

It is possible to find a constant $C_{d, \psi}$ in such a way that

$$\sum_{l=1}^{\alpha-1} \frac{1}{l!} \sum_{\alpha_1 + \dots + \alpha_k = l} \binom{\alpha}{\alpha_1:k} \prod_{j=1}^k \sum_{i_1, \dots, i_{\alpha_j}=1}^d u_{j, i_1} \cdots u_{j, i_{\alpha_j}} \leq C_{d, \psi} \prod_{j=1}^k \|\mathbf{u}_j\|^d.$$

Then we have

$$\begin{aligned} \mathcal{S}'_{\varphi, 1} &\leq \frac{1}{f_{\mathbf{X}, \min}^k} \int_{\mathbb{R}^{d'k}} \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \left\{ \varphi(\mathbf{y}_1, \dots, \mathbf{y}_k) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \right. \\ &\quad \times \sum_{i=1}^{\alpha-1} \left(\sum_{\alpha_1 + \dots + \alpha_k = l} \frac{1}{l!} \binom{\alpha}{\alpha_1:k} \prod_{j=1}^k \sum_{i_1, \dots, i_{\alpha_j}=1}^d u_{j, i_1} \cdots u_{j, i_{\alpha_j}} \frac{\partial^{\alpha_j} f_{\mathbf{X}, \mathbf{Y}}}{\partial x_{i_1} \cdots \partial x_{i_{\alpha_j}}} (\mathbf{x}_j + h \mathbf{u}_j, \mathbf{y}_j) \right) \left. \right\} \\ &\quad \times \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j) \prod_{j=1}^k d\mathbf{u}_j \prod_{j=1}^k d\mu(\mathbf{y}_j) \\ &\leq \frac{1}{f_{\mathbf{X}, \min}^k} \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j) \left| \int_{\mathbb{R}^{d'k}} \varphi(\mathbf{y}_1, \dots, \mathbf{y}_k) \right. \\ &\quad \left. - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \right. \\ &\quad \times \sum_{i=1}^{\alpha-1} \left(\sum_{\alpha_1 + \dots + \alpha_k = l} \frac{1}{l!} \binom{\alpha}{\alpha_1:k} \prod_{j=1}^k \sum_{i_1, \dots, i_{\alpha_j}=1}^d u_{j, i_1} \cdots u_{j, i_{\alpha_j}} \frac{\partial^{\alpha_j} f_{\mathbf{X}, \mathbf{Y}}}{\partial x_{i_1} \cdots \partial x_{i_{\alpha_j}}} (\mathbf{x}_j + h \mathbf{u}_j, \mathbf{y}_j) \right) \\ &\quad \left. \prod_{j=1}^k d\mu(\mathbf{y}_j) \right| \prod_{j=1}^k d\mathbf{u}_j \\ &\leq \frac{C_{\alpha, \psi} C_{d, \psi} C_{\varphi, \alpha}}{f_{\mathbf{X}, \min}^k} \left\{ \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \prod_{j=1}^k \|\mathbf{u}_j\|^d \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j) \prod_{j=1}^k d\mathbf{u}_j \right\} \\ &\leq \frac{C_{\alpha, \psi} C_{d, \psi} C_{\varphi, \alpha}}{f_{\mathbf{X}, \min}^k} O(m^{-k}). \end{aligned}$$

Now, using the same reasoning to calculate $\mathcal{S}''_{\varphi, 1}$, we readily obtain that

$$\begin{aligned} &\int_{\mathbb{R}^{d'k}} \left\{ \varphi(\mathbf{y}_1, \dots, \mathbf{y}_k) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \right\} \\ &\quad \times \left\{ \int_{\|\mathbf{u}_1\| < \eta} \cdots \int_{\|\mathbf{u}_k\| < \eta} \frac{1}{\alpha!} \Psi_{\mathbf{x}, \mathbf{u}}^{(\alpha)}(h_{\mathbf{x}, \mathbf{u}}) \prod_{j=1}^k \frac{\delta_{m_n}(\mathbf{x}_j, \mathbf{x}_j + \mathbf{u}_j)}{f_{\mathbf{X}}(\mathbf{x}_j)} \prod_{j=1}^k d\mathbf{u}_j \right\} \prod_{j=1}^k d\mu(\mathbf{y}_j) \\ &\leq \frac{C_{\alpha, \psi} C_{d, \psi} C_{\varphi, \alpha}}{\alpha! f_{\mathbf{X}, \min}^k} O(m^{-k}). \end{aligned}$$

Finally, we need to study $\mathcal{S}_{\varphi,2}$ to conclude this proof. We have

$$\begin{aligned}
 \mathcal{S}_{\varphi,2} &= \left| \int_{\mathbb{R}^{d'k}} \int_{\|\mathbf{u}_1\| \geq \eta} \cdots \int_{\|\mathbf{u}_k\| \geq \eta} \left\{ \varphi(\mathbf{y}_1, \dots, \mathbf{y}_k) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \right\} \right. \\
 &\quad \times \left(\prod_{j=1}^k f_{\mathbf{X},\mathbf{Y}}(\mathbf{t}_j, \mathbf{y}_j) - \prod_{j=1}^k f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}_j, \mathbf{y}_j) \right) \prod_{j=1}^k \frac{\delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j)}{f_{\mathbf{X}}(\mathbf{x}_j)} \prod_{j=1}^k d\mu(\mathbf{y}_j) \prod_{j=1}^k d\mathbf{t}_j \left. \right| \\
 &\leq \frac{1}{f_{\mathbf{X},\min}^k} \left| \int_{\mathbb{R}^{d'k}} \int_{\|\mathbf{u}_1\| \geq \eta} \cdots \int_{\|\mathbf{u}_k\| \geq \eta} \left\{ \varphi(\mathbf{y}_1, \dots, \mathbf{y}_k) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \right\} \right. \\
 &\quad \times \left(\prod_{j=1}^k f_{\mathbf{X},\mathbf{Y}}(\mathbf{t}_j, \mathbf{y}_j) - \prod_{j=1}^k f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}_j, \mathbf{y}_j) \right) \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j) \prod_{j=1}^k d\mu(\mathbf{y}_j) \prod_{j=1}^k d\mathbf{t}_j \left. \right| \\
 &\leq \frac{C_\varphi}{f_{\mathbf{X},\min}^k} \left| \int_{\|\mathbf{u}_1\| \geq \eta} \cdots \int_{\|\mathbf{u}_k\| \geq \eta} \left(\prod_{j=1}^k f_{\mathbf{X},\mathbf{Y}}(\mathbf{t}_j, \mathbf{y}_j) - \prod_{j=1}^k f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}_j, \mathbf{y}_j) \right) \prod_{j=1}^k \delta_{m_n}(\mathbf{x}_j, \mathbf{t}_j) \prod_{j=1}^k d\mu(\mathbf{y}_j) \prod_{j=1}^k d\mathbf{t}_j \right| \\
 &\leq \frac{C_\varphi}{f_{\mathbf{X},\min}^k} O(m^{-k/q}).
 \end{aligned}$$

Consequently, we can find a constant $C_{\delta,\alpha,\varphi} > 0$ in such a way that we have for all $\mathbf{i} \in I(k, n)$,

$$|\mathbb{E}[S_{\mathbf{i}}]| \leq \frac{C_{\delta,\alpha,\varphi} m^{-\lambda}}{f_{\mathbf{X},\min}^k}.$$

Now, we bound the stochastic part. We have the following equality:

$$\left| \sum_{\mathbf{i} \in I(k,n)} (S_{\mathbf{i}} - \mathbb{E}[S_{\mathbf{i}}]) \right| = \left| \sum_{\mathbf{i} \in I(k,n)} \tilde{\varphi}((\mathbf{X}_{i_1}, \mathbf{Y}_{i_1}), \dots, (\mathbf{X}_{i_k}, \mathbf{Y}_{i_k})) \right|,$$

with $\tilde{\varphi}$ defined by

$$\begin{aligned}
 &\tilde{\varphi}((\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_k, \mathbf{Y}_k)) \\
 &= \prod_{j=1}^k \frac{\delta_{m_n}(\mathbf{x}_j, \mathbf{X}_j)}{f_{\mathbf{X}}(\mathbf{x}_j)} \left\{ \varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \right\} \\
 &\quad - \mathbb{E} \left[\prod_{j=1}^k \frac{\delta_{m_n}(\mathbf{x}_j, \mathbf{X}_j)}{f_{\mathbf{X}}(\mathbf{x}_j)} \left\{ \varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) - \mathbb{E}[\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \right\} \right].
 \end{aligned}$$

Notice that $\mathbb{E}[\tilde{\varphi}((\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_k, \mathbf{Y}_k))] = 0$ by construction. Considering condition **(C.5)**, if φ is bounded, we can immediately derive the bound for this stochastic component. Indeed, we can easily conclude that

$$\|\tilde{\varphi}\|_\infty \leq 4C_\varphi f_{\mathbf{X},\min}^{-k} \kappa_1.$$

Moreover, we have

$$\begin{aligned}
 & \text{Var} [\tilde{\varphi}((\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_k, \mathbf{Y}_k))] \\
 & \leq \mathbb{E} \left[\prod_{j=1}^k \frac{\delta_{m_n}^2(\mathbf{x}_j, \mathbf{X}_j)}{f_{\mathbf{X}}^2(\mathbf{x}_j)} \varphi^2(\mathbf{Y}_1, \dots, \mathbf{Y}_k) \right] \\
 & \leq \int_{\mathbb{R}^{dk}} \int_{\mathbb{R}^{d'k}} \prod_{j=1}^k \frac{\delta_{m_n}^2(\mathbf{x}_j, \mathbf{t}_j)}{f_{\mathbf{X}}^2(\mathbf{x}_j)} \varphi^2(\mathbf{y}_1, \dots, \mathbf{y}_k) \prod_{j=1}^k f_{\mathbf{X}, \mathbf{Y}}(\mathbf{t}_j, \mathbf{y}_j) \prod_{j=1}^k d\mathbf{t}_j \prod_{j=1}^k d\mu(\mathbf{y}_j) \\
 & \leq C_{\varphi}^2 f_{\mathbf{X}, \min}^{-2k} \int_{\mathbb{R}^{dk}} \prod_{j=1}^k \delta_{m_n}^2(\mathbf{x}_j, \mathbf{t}_j) \left\{ \int_{\mathbb{R}^{d'k}} \prod_{j=1}^k f_{\mathbf{X}, \mathbf{Y}}(\mathbf{t}_j, \mathbf{y}_j) \prod_{j=1}^k d\mu(\mathbf{y}_j) \right\} \prod_{j=1}^k d\mathbf{t}_j \\
 & \leq C_{\varphi}^2 f_{\mathbf{X}, \min}^{-2k} \int_{\mathbb{R}^{dk}} \prod_{j=1}^k \delta_{m_n}^2(\mathbf{x}_j, \mathbf{t}_j) \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{t}_j) \prod_{j=1}^k d\mathbf{t}_j \\
 & \leq C_{\varphi}^2 f_{\mathbf{X}, \min}^{-2k} \|\delta_{m_n}(\mathbf{x}_j, \cdot)\|_{\infty}^{2k} \\
 & \leq C_{\varphi}^2 f_{\mathbf{X}, \min}^{-2k} \kappa_2.
 \end{aligned}$$

Now, by applying Lemma 4.8.0.2, we readily get

$$\mathbb{P} \left(\left| \sum_{\mathbf{i} \in I(k, n)} (S_{\mathbf{i}} - \mathbb{E}[S_{\mathbf{i}}]) \right| > t \right) \leq 2 \exp \left(- \frac{[n/k] t^2}{2C_{\varphi}^2 f_{\mathbf{X}, \min}^{-2k} \kappa_2 + \frac{8}{3} C_{\varphi} \kappa_1 f_{\mathbf{X}, \min}^{-k} t} \right).$$

Hence, the proof is complete. \square

The following proof gives the result in the case of unbounded function $\varphi(\cdot)$.

Proof of lemma 4.3.2.8

The same decomposition for U-statistics as in Hoeffding [1963] gives us

$$\sum_{\mathbf{i} \in I(k, n)} S_{\mathbf{i}} - \mathbb{E}[S_{\mathbf{i}}] = \frac{1}{n!} \sum_{\mathbf{i} \in I(n, n)} \frac{1}{[n/k]} \sum_{j=1}^{[n/k]} V_{n, j, i},$$

where

$$V_{n, j, i} := \tilde{\varphi} \left((\mathbf{X}_{i_{(1+(j-1)k)}, \mathbf{Y}_{i_{(2+(j-1)k)}}), \dots, (\mathbf{X}_{i_{(jk)}, \mathbf{Y}_{i_{(jk+1)}}} \right).$$

For any $\lambda > 0$, we have

$$\begin{aligned}
 \mathbb{P} \left(\sum_{\mathbf{i} \in I(k, n)} S_{\mathbf{i}} - \mathbb{E}[S_{\mathbf{i}}] > t \right) & \leq e^{-\lambda e} \mathbb{E} \left[\exp \left(\lambda \sum_{\mathbf{i} \in I(k, n)} S_{\mathbf{i}} - \mathbb{E}[S_{\mathbf{i}}] \right) \right] \\
 & \leq e^{-\lambda e} \mathbb{R} \left[\exp \left(\lambda \frac{1}{n!} \sum_{\mathbf{i} \in I(n, n)} \frac{1}{[n/k]} \sum_{j=1}^{[n/k]} V_{n, j, i} \right) \right] \\
 & \leq e^{-\lambda e} \frac{1}{n!} \sum_{\mathbf{i} \in I(n, n)} \mathbb{E} \left[\exp \left(\lambda \frac{1}{[n/k]} \sum_{j=1}^{[n/k]} V_{n, j, i} \right) \right] \\
 & \leq e^{-\lambda e} \frac{1}{n!} \sum_{\mathbf{i} \in I(n, n)} \prod_{j=1}^{[n/k]} \mathbb{E} \left[\exp \left(\lambda \frac{1}{[n/k]} V_{n, j, i} \right) \right] \\
 & \leq e^{-\lambda} \left(\sup_{\mathbf{i} \in I(n, n), j=1, \dots, [n/k]} \mathbb{E} \left[\exp \left(\lambda [n/k]^{-1} V_{n, j, i} \right) \right] \right)^{[n/k]}. \quad (4.7.15)
 \end{aligned}$$

For $\ell \geq 2$, using the inequality

$$(a + b + c + d)^\ell \leq 4^\ell (a^\ell + b^\ell + c^\ell + d^\ell),$$

we obtain

$$\begin{aligned} \mathbb{E} \left[|V_{n,j,i}|^\ell \right] &= \mathbb{E} \left[|V_{n,1,i}|^\ell \right] \\ &\leq 4^\ell \mathbb{E} \left[\left| \varphi(\mathbf{Y}_{i_1}, \dots, \mathbf{Y}_{i_k}) \right|^\ell \prod_{j=1}^k \frac{\delta_{m_n}^\ell(\mathbf{x}_j, \mathbf{X}_{i_j})}{f_{\mathbf{X}}^\ell(\mathbf{x}_j)} \right] \\ &\quad + 4^\ell \mathbb{E} \left[\left| \mathbb{E}[\varphi(\mathbf{Y}_{i_1}, \dots, \mathbf{Y}_{i_k}) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \right|^\ell \prod_{j=1}^k \frac{\delta_{m_n}^\ell(\mathbf{x}_j, \mathbf{X}_{i_j})}{f_{\mathbf{X}}^\ell(\mathbf{x}_j)} \right] \\ &\quad + 4^\ell \left| \mathbb{E} \left[\varphi(\mathbf{Y}_{i_1}, \dots, \mathbf{Y}_{i_k}) \prod_{j=1}^k \frac{\delta_{m_n}^\ell(\mathbf{x}_j, \mathbf{X}_{i_j})}{f_{\mathbf{X}}^\ell(\mathbf{x}_j)} \right] \right|^\ell \\ &\quad + 4^4 \left| \mathbb{E} \left[\left| \mathbb{E}[\varphi(\mathbf{Y}_{i_1}, \dots, \mathbf{Y}_{i_k}) \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x}] \right| \prod_{j=1}^k \frac{\delta_{m_n}^\ell(\mathbf{x}_j, \mathbf{X}_{i_j})}{f_{\mathbf{X}}^\ell(\mathbf{x}_j)} \right] \right|^\ell. \end{aligned}$$

Using Jensen's inequality for the function $x \mapsto |x|^\ell$ with the second, third, and fourth terms and the law of iterated expectations for the first and the third terms, we get

$$\begin{aligned} \mathbb{E} \left[|V_{n,j,i}|^\ell \right] &\leq 4^\ell \cdot 2 \mathbb{E} \left[\mathbb{E} \left[\left| \varphi(\mathbf{Y}_{i_1}, \dots, \mathbf{Y}_{i_k}) \right|^\ell \mid \mathbf{X}_{i_1}, \dots, \mathbf{X}_{i_k} \right] \prod_{j=1}^k \frac{\delta_{m_n}^\ell(\mathbf{x}_j, \mathbf{X}_{i_j})}{f_{\mathbf{X}}^\ell(\mathbf{x}_j)} \right] \\ &\quad + 4^\ell \cdot 2 \mathbb{E} \left[\mathbb{E} \left[\left| \varphi(\mathbf{Y}_{i_1}, \dots, \mathbf{Y}_{i_k}) \right|^\ell \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = \mathbf{x} \right] \prod_{j=1}^k \frac{\delta_{m_n}^\ell(\mathbf{x}_j, \mathbf{X}_{i_j})}{f_{\mathbf{X}}^\ell(\mathbf{x}_j)} \right] \\ &\leq 4^\ell \cdot 2 \mathbb{E} \left[\left(B_\varphi^\ell(\mathbf{X}_1, \dots, \mathbf{X}_k) + B_\varphi^\ell(\mathbf{x}_1, \dots, \mathbf{x}_k) \right)^\ell \ell! \prod_{j=1}^k \frac{\delta_{m_n}^\ell(\mathbf{x}_j, \mathbf{X}_{i_j})}{f_{\mathbf{X}}^\ell(\mathbf{x}_j)} \right] \\ &\leq 4^\ell \cdot 2 \left(\tilde{B}_\varphi^\ell + B_\varphi^\ell(\mathbf{x}_1, \dots, \mathbf{x}_k) \right) \ell! \left(\|\delta_{m_n}(\mathbf{x}, \cdot)\|_{\infty}^k f_{\mathbf{X},\min}^{-k} \right)^{\ell-1} f_{\mathbf{X},\min}^{-k} \\ &\leq 2 \left(4(\tilde{B}_\varphi + B_{\varphi,\mathbf{x}}) \|\delta_{m_n}(\mathbf{x}, \cdot)\|_{\infty}^k f_{\mathbf{X},\min}^{-k} \right)^\ell \ell! \|\delta_{m_n}(\mathbf{x}, \cdot)\|_{\infty}^{-1}, \end{aligned}$$

where $B_{\varphi, \mathbf{x}} := B_{\varphi}(\mathbf{x}_1, \dots, \mathbf{x}_k)$. Observe that $\mathbb{E}[V_{n,j,i}] = 0$ by construction of $\tilde{\varphi}(\cdot)$, which means that

$$\begin{aligned} & \mathbb{E}[\exp(\lambda[n/k]^{-1}V_{n,j,i})] \\ &= 1 + \sum_{\ell=2}^{\infty} \frac{\mathbb{E}[(\lambda[n/k]^{-1}V_{n,j,i})^{\ell}]}{\ell!} \end{aligned} \quad (4.7.16)$$

$$\begin{aligned} &\leq 1 + 2\|\delta_{m_n}(\mathbf{x}, \cdot)\|_{\infty}^{-1} \sum_{\ell=2}^{\infty} \left(4\lambda[n/k]^{-1}(\tilde{B}_{\varphi} + B_{\varphi, \mathbf{x}})\|\delta_{m_n}(\mathbf{x}, \cdot)\|_{\infty}^k f_{\mathbf{X}, \min}^{-k}\right)^{\ell} \\ &\leq 1 + 2\|\delta_{m_n}(\mathbf{x}, \cdot)\|_{\infty}^{-1} \cdot \frac{\left(4\lambda[n/k]^{-1}(\tilde{B}_{\varphi} + B_{\varphi, \mathbf{x}})\|\delta_{m_n}(\mathbf{x}, \cdot)\|_{\infty}^k f_{\mathbf{X}, \min}^{-k}\right)^2}{1 - 4\lambda[n/k]^{-1}(\tilde{B}_{\varphi} + B_{\varphi, \mathbf{x}})\|\delta_{m_n}(\mathbf{x}, \cdot)\|_{\infty}^k f_{\mathbf{X}, \min}^{-k}} \\ &\leq \exp\left(\frac{32\lambda^2(\tilde{B}_{\varphi} + B_{\varphi, \mathbf{x}})^2\|\delta_{m_n}(\mathbf{x}, \cdot)\|_{\infty}^{2k-1}}{1 - 4\lambda[n/k]^{-1}(\tilde{B}_{\varphi} + B_{\varphi, \mathbf{x}})\|\delta_{m_n}(\mathbf{x}, \cdot)\|_{\infty}^k f_{\mathbf{X}, \min}^{-k}}\right), \end{aligned} \quad (4.7.17)$$

where (4.7.17) follows from the inequality

$$1 + x \leq \exp(x).$$

Combining the latter bound with (4.7.15), we get

$$\begin{aligned} & \mathbb{P}\left(\sum_{i \in J_{k,a}} S_i - \mathbb{E}[S_i] > t\right) \\ &\leq \exp\left(-\lambda t + \frac{32\lambda^2[n/k]^{-2}(\tilde{B}_{\varphi} + B_{\varphi, \mathbf{x}})^2\|\delta_{m_n}(\mathbf{x}, \cdot)\|_{\infty}^{2k-1} f_{\mathbf{X}, \min}^{-2k}}{f_{\mathbf{X}, \min}^{2k}[n/k] - 4\lambda(\tilde{B}_{\varphi} + B_{\varphi, \mathbf{x}})\|\delta_{m_n}(\mathbf{x}, \cdot)\|_{\infty}^k f_{\mathbf{X}, \min}^k}\right). \end{aligned} \quad (4.7.18)$$

Noting that the right-hand side term within the exponential has the form $-\lambda t + \frac{a\lambda^2}{b-c\lambda}$, We select the value

$$\lambda_* = \frac{tb}{2a + tc} = \frac{t f_{\mathbf{X}, \min}^{2k}[n/k]}{64(\tilde{B}_{\varphi} + B_{\varphi, \mathbf{x}})^2\|\delta_{m_n}(\mathbf{x}, \cdot)\|_{\infty}^{2k-1} + t(\tilde{B}_{\varphi} + B_{\varphi, \mathbf{x}})\|\delta_{m_n}(\mathbf{x}, \cdot)\|_{\infty}^k f_{\mathbf{X}, \min}^k} \quad (4.7.19)$$

such that

$$-\lambda_* t + \frac{a\lambda_*^2}{b - c\lambda_*} = -\frac{t^2 b}{4a + 2ct} = -\frac{t}{2}\lambda_*.$$

As a result, the right-hand side term of (4.7.18) may be simplified, and when combined with (4.7.19), we get

$$\begin{aligned} & \mathbb{P}\left(\sum_{i \in I(k,n)} S_i - \mathbb{E}[S_i] > t\right) \\ &\leq \exp\left(-\frac{t^2 f_{\mathbf{X}, \min}^{2k}[n/k]}{128(\tilde{B}_{\varphi} + B_{\varphi, \mathbf{x}})^2\|\delta_{m_n}(\mathbf{x}, \cdot)\|_{\infty}^{2k-1} + t(\tilde{B}_{\varphi} + B_{\varphi, \mathbf{x}})\|\delta_{m_n}(\mathbf{x}, \cdot)\|_{\infty}^k f_{\mathbf{X}, \min}^k}\right). \end{aligned}$$

Now, using the condition (iii) of Definition 4.2.1.2, there exist positive constants C'_1, C'_2 such that

$$\kappa'_1 =: C'_1 m^{dk} \quad \text{and} \quad \kappa'_2 =: C'_2 m^{d(2k-1)},$$

that implies that

$$\mathbb{P}\left(\sum_{\mathbf{i} \in \mathcal{I}(k,n)} S_{\mathbf{i}} - \mathbb{E}[S_{\mathbf{i}}] > t\right) \leq \exp\left(-\frac{t^2 f_{\mathbf{X},\min}^{2k} [n/k]}{128 (\tilde{B}_{\varphi} + B_{\varphi,\mathbf{x}})^2 \kappa_2' + t (\tilde{B}_{\varphi} + B_{\varphi,\mathbf{x}}) \kappa_1' f_{\mathbf{X},\min}^k}\right).$$

Hence, we obtain the desired result. \square

4.7.2 Proofs of the results of Section 4.3.3

Before we start the proofs of this section, we will state some lemmas that are necessary to obtain the desired results. It is worth mentioning that we will follow the steps of [Stute \[1991\]](#) while making the appropriate changes to fit our general setting.

Lemma 4.7.2.1 *Under assumptions (A.1)-(A.4), and if $\mathbb{E}\varphi^2 < \infty$, the Hájek projection \hat{U}_n of U_n satisfies :*

(i)

$$\lim_{n \rightarrow \infty} \mathbb{E}\left[(nm_n^{-d})^{1/2} (\hat{U}_n - \theta_n)\right]^2 = \sigma^2(\varphi),$$

where

$$\sigma^2(\varphi) := \sum_{i=1}^k \sum_{j=1}^k \mathbb{1}_{\{x_i=x_j\}} r_{ij}(\tilde{\mathbf{x}}) \int_{\mathbb{R}^d} \boldsymbol{\delta}^2(\mathbf{x}_i, \mathbf{t}) d\mathbf{t} / f_{\mathbf{X}}(\mathbf{x}_i). \quad (4.7.20)$$

(ii) *and if, in addition, assumption (A.5) is verified, we have*

$$(nm_n^{-d})^{1/2} (\hat{U}_n - \theta_n) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(\varphi)). \quad (4.7.21)$$

In the following lemma, we show that U_n has the same asymptotic distribution as \hat{U}_n .

Lemma 4.7.2.2 *Under assumption (A.1)-(A.6), we have*

$$(nm_n^{-d})^{1/2} (U_n - \theta_n) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(\varphi)). \quad (4.7.22)$$

Proof of Lemma 4.7.2.1

Let us start with (i), we write

$$\begin{aligned} \theta_n &= \mathbb{E}[U_n(\varphi, \tilde{\mathbf{x}}, m_n)] \\ &= \mathbb{N}^{-1} \int_{\mathbb{R}^{dk}} \hat{r}_n^{(k)}(\varphi, \mathbf{t}; m_n) \prod_{j=1}^k \boldsymbol{\delta}_{m_n}(\mathbf{x}_j, \mathbf{t}_j) \prod_{j=1}^k f_{\mathbf{X}}(\mathbf{t}_j) \prod_{j=1}^k d\mathbf{t}_j, \end{aligned}$$

The Hájek projection $\widehat{U}_n(\varphi, \mathbf{x}, m_n)$ of $U_n(\varphi, \tilde{\mathbf{x}}, m_n)$ satisfies

$$\widehat{U}_n - \theta_n = n^{-1} \sum_{i=1}^n \bar{\varphi}_n(\mathbf{X}_i, \mathbf{Y}_i),$$

with

$$\bar{\varphi}_n(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^k [\varphi_{n,j}(\mathbf{x}, \mathbf{y}) - \theta_n],$$

and $\varphi_{n,j}(\mathbf{x}, \mathbf{y})$ is defined by

$$\varphi_{n,j}(\mathbf{x}, \mathbf{y}) = N^{-1} \int_{\mathbb{R}^{dk} \times \mathbb{R}^{d'k}} \varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_{j-1}, \mathbf{y}, \mathbf{Y}_{j+1}, \dots, \mathbf{Y}_k) \prod_{\substack{r=1 \\ r \neq j}}^k \delta_{m_n, r}(\mathbf{x}_r, \mathbf{X}_r) \delta_{m_n}(\mathbf{x}_j, \mathbf{x}) d\mathbb{P},$$

where \mathbb{P} represents the underlying probability measure. By independence, we get

$$\begin{aligned} n\mathbb{E}(\widehat{U}_n - \theta_n)^2 &= \mathbb{E}[\bar{\varphi}_n^2(\mathbf{X}, \mathbf{Y})] \\ &= \sum_{j=1}^k \sum_{l=1}^k \mathbb{E}[\varphi_{n,j}(\mathbf{X}, \mathbf{Y}) - \theta_n] [\varphi_{n,l}(\mathbf{X}, \mathbf{Y}) - \theta_n]. \end{aligned}$$

Using the fact that (\mathbf{X}, \mathbf{Y}) , $(\mathbf{X}_i, \mathbf{Y}_i)_{1 \leq i \leq 2k}$ are i.i.d, We get

$$\begin{aligned} &\mathbb{E}[\varphi_{n,j}(\mathbf{X}, \mathbf{Y})\varphi_{n,l}(\mathbf{X}, \mathbf{Y})] \\ &= N^{-2} \int_{\mathbb{R}^{2dk} \times \mathbb{R}^{2d'k}} \varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_{j-1}, \mathbf{Y}, \mathbf{Y}_{j+1}, \dots, \mathbf{Y}_k) \varphi(\mathbf{Y}_{k+1}, \dots, \mathbf{Y}_{k+l-1}, \mathbf{Y}, \mathbf{Y}_{k+l+1}, \dots, \mathbf{Y}_{2k}) \\ &\quad \times \prod_{\substack{r=1 \\ r \neq j}}^k \delta_{m_n, r}(\mathbf{x}_r, \mathbf{X}_r) \prod_{\substack{s=1 \\ s \neq l}}^k \delta_{m_n, s}(\mathbf{x}_s, \mathbf{X}_{k+s}) \delta_{m_n}(\mathbf{x}_j, \mathbf{X}) \delta_{m_n, l}(\mathbf{x}_l, \mathbf{X}) d\mathbb{P}. \end{aligned}$$

By condition (A.2), we have, for $\mathbf{x}_j \neq \mathbf{x}_l$

$$\delta_{m_n}(\mathbf{x}_j, \mathbf{X})\delta_{m_n}(\mathbf{x}_l, \mathbf{X}) \rightarrow 0,$$

Consequently,

$$\mathbb{E}[\varphi_{n,j}(\mathbf{X}, \mathbf{Y})\varphi_{n,l}(\mathbf{X}, \mathbf{Y})] = 0 \text{ for all } \mathbf{x}_j \neq \mathbf{x}_l.$$

In the case of $\mathbf{x}_j = \mathbf{x}_l$ and $\tilde{\mathbf{x}} = (\mathbf{x}_1, \dots, \mathbf{x}_k)$ is a point of continuity for $r_{j,l}^{(k)}$ we have

$$\frac{\mathbb{E}^2 \left[\frac{1}{m_n^d} \delta_{m_n}(\mathbf{x}_j, \mathbf{X}_1) \right]}{\mathbb{E} \left[\frac{1}{m_n^d} \delta_{m_n}^2(\mathbf{x}_j, \mathbf{X}_1) \right]} \mathbb{E}[\varphi_{n,j}(\mathbf{X}, \mathbf{Y})\varphi_{n,l}(\mathbf{X}, \mathbf{Y})] \rightarrow r_{j,l}(\tilde{\mathbf{x}}), \quad (4.7.23)$$

by a differentiation argument. Now, provided that the density function $f_{\mathbf{X}}(\cdot)$ is continuous and $f_{\mathbf{X}}(\mathbf{x}_j) = f_{\mathbf{X}}(\mathbf{x}_l) > 0$, by the fact that

$$\mathbb{E} \left[\frac{1}{m_n^d} \delta_{m_n}(\mathbf{x}_j, \mathbf{X}_1) \right] \sim \frac{1}{m_n^d} f_{\mathbf{X}}(\mathbf{x}_j),$$

and

$$\mathbb{E} \left[\frac{1}{m_n^d} \delta_{m_n}^2(\mathbf{x}_j, \mathbf{X}_1) \right] \sim \frac{1}{m_n^d} f_{\mathbf{X}}(\mathbf{x}_j) \int \delta_{m_n}^2(\mathbf{x}_j, \mathbf{t}) dt,$$

we readily infer that

$$\frac{\mathbb{E}^2 \left[\frac{1}{m_n^d} \delta_{m_n}(\mathbf{x}_j, \mathbf{X}_1) \right]}{\mathbb{E} \left[\frac{1}{m_n^d} \delta_{m_n}^2(\mathbf{x}_j, \mathbf{X}_1) \right]} \sim \frac{1}{m_n^d} \frac{f_{\mathbf{X}}(\mathbf{x}_j)}{\int \delta_{m_n}^2(\mathbf{x}_j, \mathbf{t}) dt}.$$

Hence,

$$\frac{1}{m_n^d} \mathbb{E} [\varphi_{n,j}(\mathbf{X}, \mathbf{Y}) \varphi_{n,l}(\mathbf{X}, \mathbf{Y})] \rightarrow r_{j,l}(\tilde{\mathbf{x}}) \frac{\int \delta_{m_n}^2(\mathbf{x}_j, \mathbf{t}) d\mathbf{t}}{f_{\mathbf{X}}(\mathbf{x}_j)}.$$

Using the fact that $r^{(k)}(\varphi, \tilde{\mathbf{x}})$ is bounded in the neighborhood of $\tilde{\mathbf{x}}$, then $\theta_n, n \geq 1$, is bounded, therefore we have $m_n^{-d} \theta_n^2 \rightarrow 0$, from which we have

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[(nm_n^{-d})^{1/2} (\hat{U}_n - \theta_n) \right]^2 = \sigma^2(\varphi), \quad (4.7.24)$$

where $\sigma^2(\varphi)$ is defined in (4.7.20).

Now, in order to prove (4.7.21), we only need to verify Lyapunov's condition for third moments. This implies showing that

$$n^{-1/2} (m_n^{-d})^{3/2} \mathbb{E} |\bar{\varphi}_n(\mathbf{X}, \mathbf{Y})|^3 \rightarrow 0. \quad (4.7.25)$$

Using the fact that

$$|a - b|^3 \leq 3(|a|^3 + |a|^2|b| + |a||b|^2 + |b|^3),$$

we conclude that an upper bound for the absolute third moment of $\bar{\varphi}_n(\mathbf{X}, \mathbf{Y})$ is dominated by sums of the form

$$\mathbb{E} |\varphi_{n,i}(\mathbf{X}, \mathbf{Y}) \varphi_{n,j}(\mathbf{X}, \mathbf{Y}) \varphi_{n,l}(\mathbf{X}, \mathbf{Y})|.$$

Following the same steps as before, we may restrict ourselves to triples (i, j, l) such that $\mathbf{x}_i = \mathbf{x}_j = \mathbf{x}_l$. Under assumption (A.5), we have

$$\mathbb{E} |\varphi_{n,i}(\mathbf{X}, \mathbf{Y}) \varphi_{n,j}(\mathbf{X}, \mathbf{Y}) \varphi_{n,l}(\mathbf{X}, \mathbf{Y})| = O(nm_n^{-2d}),$$

taking into account condition (A.1), gives us the desired result. This concludes the proof of the lemma. \square

Proof of Lemma 4.7.2.2

In order to study the asymptotic distribution of U_n , we need to bound the variance of $U_n - \hat{U}_n$. To do that, it is sufficient to show that

$$(nm_n^{-d})^{1/2} [U_n - \hat{U}_n] \rightarrow 0 \text{ in } L^2.$$

Using the centered variance formula for a centered, or zero mean U-statistic of degree k , for $\mathbf{Z}_i, i \geq 1$ i.i.d., we have

$$\mathcal{V}_n = \frac{(n-k)!}{n!} \sum_{\mathbf{i} \in I(k,n)} \frac{\tilde{G}(\mathbf{Z}_{i_1}, \dots, \mathbf{Z}_{i_k})}{N},$$

with a not necessary symmetric U Kernel \tilde{G} , that is square-integrable, which gives us

$$\text{Var}(\mathcal{V}_n) \left[\frac{(n-k)!}{n!} \right]^2 \sum_{r=1}^k \frac{(n-r)!}{(n-2k+r)!} \sum_{\binom{r}{N^2}} \frac{I(\Delta_1, \Delta_2)}{N^2},$$

where Δ_1 and Δ_2 represents positions of some length $1 \leq r \leq k$, and

$$I(\tilde{\Delta}_1, \tilde{\Delta}_2) = \int \tilde{G}(z_1, \dots, z_k) \tilde{G}(y_1, \dots, y_k) F(dz_1) \cdots F(dz_{2k-r}),$$

with the y 's in position $\tilde{\Delta}_2$ coincide with the z 's in position $\tilde{\Delta}_1$ and are taken from z_{k+1}, \dots, z_{2k-r} otherwise. Furthermore, $\Sigma^{(r)}$ denotes summation over all positions $\tilde{\Delta}_1, \tilde{\Delta}_2$ with cardinality r and F is the common d.f. of the Z 's. When applied to $V_n = U_n - \hat{U}_n$, upon recalling \tilde{G} from [Serfling \[1980\]](#) (in the symmetric case), we get

$$\Sigma^{(1)} I(\tilde{\Delta}_1, \tilde{\Delta}_2) = 0.$$

Furthermore, by [\(A.6\)](#),

$$N^{-2} I(\tilde{\Delta}_1, \tilde{\Delta}_2) = O\left(m_n^{dr}\right) \quad \text{for each } 2 \leq r \leq k.$$

In conclusion, we have

$$\begin{aligned} nm_n^{-d} \text{Var}(U_n - \hat{U}_n) &= O\left[nm_n^{-d} \sum_{r=2}^k \begin{bmatrix} n \\ k \end{bmatrix}^{-1} \begin{bmatrix} k \\ r \end{bmatrix} \begin{bmatrix} n-k \\ k-r \end{bmatrix} (m_n^{-d})^{-r} \right] \\ &= O\left[\sum_{r=2}^k (nm_n^{-d})^{1-r} \right] = O\left[(nm_n^{-d})^{-1} \right] = o(1) \end{aligned}$$

Hence, the proof is complete. □

Proof of Theorem [4.3.4.1](#)

In order to obtain the desired result, we shall first apply the Cramér-Wold device to investigate the asymptotic behavior of the two-dimensional vector

$$(U_n(\varphi_1, \tilde{\mathbf{x}}, m_n) - \theta_n(\varphi_1), U_n(\varphi_2, \tilde{\mathbf{x}}, m_n) - \theta_n(\varphi_2)),$$

where $U_n(\varphi_1, \tilde{\mathbf{x}}, m_n)$ and $U_n(\varphi_2, \tilde{\mathbf{x}}, m_n)$ are U-statistics with U kernels $G_{\varphi_1, \tilde{\mathbf{x}}, m_n}$ and $G_{\varphi_2, \tilde{\mathbf{x}}, m_n}$ respectively, satisfying the smoothness assumptions of [Lemma 4.7.2.2](#).

Let c_1 and c_2 be any two real numbers. We can clearly see that

$$c_1 U_n(\varphi_1, \tilde{\mathbf{x}}, m_n) + c_2 U_n(\varphi_2, \tilde{\mathbf{x}}, m_n) = U_n(c_1 \varphi_1 + c_2 \varphi_2, \tilde{\mathbf{x}}, m_n) \equiv U_n(\varphi, \tilde{\mathbf{x}}, m_n),$$

which means we can apply [Lemma 4.7.2.2](#). Specification of $\sigma^2(\varphi)$ leads to the following lemma.

Lemma 4.7.2.3 *Under assumptions of [Lemma 4.7.2.2](#), we have*

$$(nm_n^{-d})^{1/2} [U_n(\varphi_1, \tilde{\mathbf{x}}, m_n) - \theta_n(\varphi_1), U_n(\varphi_2, \tilde{\mathbf{x}}, m_n) - \theta_n(\varphi_2)] \rightarrow \mathcal{N}(\mathbf{0}, \Sigma)$$

in distribution, with

$$\Sigma = \begin{bmatrix} \sigma^2(\varphi_1, \varphi_1) & \sigma^2(\varphi_1, \varphi_2) \\ \sigma^2(\varphi_1, \varphi_2) & \sigma^2(\varphi_2, \varphi_2) \end{bmatrix}$$

and where for two functions g and h ,

$$\sigma^2(g, h) = \sum_{j=1}^k \sum_{l=1}^k \mathbb{1}_{\{x_j \neq x_l\}} r_{jl}^{gh}(\tilde{\mathbf{x}}) \int \delta_{m_n}^2(\mathbf{x}_j, \mathbf{t}) d\mathbf{t} / f_{\mathbf{X}}(x_j),$$

and

$$r_{jl}^{gh}(\tilde{\mathbf{x}}) = \mathbb{E} [g(\mathbf{Y}_1, \dots, \mathbf{Y}, \dots, \mathbf{Y}_k) h(\mathbf{Y}_{k+1}, \dots, \mathbf{Y}, \dots, \mathbf{Y}_{2k}) \mid \dots],$$

with \mathbf{Y} entering in the j th and l th positions.

The limit distribution of $\hat{r}_n^{(k)}(\varphi, \tilde{\mathbf{x}}, m_n)$ may now be easily deduced from the prior lemma.

We have

$$\hat{r}_n^{(k)}(\varphi, \tilde{\mathbf{x}}, m_n) = \frac{U_n(\varphi, \tilde{\mathbf{x}}, m_n)}{U_n(1, \tilde{\mathbf{x}}, m_n)},$$

Define

$$g(x_1, x_2) = x_1/x_2 \text{ for } x_2 \neq 0$$

which means

$$D = \left[\frac{\partial g}{\partial x_1}, \frac{\partial g}{\partial x_2} \right] = [x_2^{-1}, -x_1 x_2^{-2}]$$

Since, by continuity of $r^{(k)}(\varphi, \tilde{\mathbf{x}})$, we have

$$\mathbb{E} U_n(\varphi, \tilde{\mathbf{x}}, m_n) \rightarrow r^{(k)}(\varphi, \tilde{\mathbf{x}}),$$

and

$$\mathbb{E} U_n(1, \tilde{\mathbf{x}}, m_n) = 1$$

we may deduce from Lemma (4.7.2.3)

$$(nm_n^{-d})^{1/2} \left[\hat{r}_n^{(k)}(\varphi, \tilde{\mathbf{x}}, m_n) - \mathbb{E} U_n(\varphi, \tilde{\mathbf{x}}, m_n) \right] \rightarrow \mathcal{N}(0, \rho^2),$$

where

$$\rho^2 = (1, -r^{(k)}(\varphi, \tilde{\mathbf{x}})) \Sigma \begin{bmatrix} 1 \\ -r^{(k)}(\varphi, \tilde{\mathbf{x}}) \end{bmatrix},$$

where ρ^2 is defined in (4.3.10). This concludes the proof of Theorem 4.3.4.1. \square

4.8 Appendix

Lemma 4.8.0.1 (Lemma 2.1 Susarla and Walter [1981]) Let $f \in L^p(\mathbb{R}^d)$, $1 \leq p \leq \infty$ ¹ and let $\{\delta_m\}$ be a delta sequence of positive type.

¹The vector space $L_p(\mathbb{R}^d)$, $1 \leq p < \infty$ is the set of all measurable functions such that $\int_{\mathbb{R}^d} |f(x)|^p dx < \infty$. The norm is defined by

$$\|f\|_{L^p} = \left(\int_{\mathbb{R}^d} |f(\mathbf{x})|^p d\mathbf{x} \right)^{1/p}.$$

(i) Then

$$\int \delta_m(\mathbf{x}, \mathbf{t}) f(\mathbf{t}) d\mathbf{t} \rightarrow f(\mathbf{x}) \quad \text{a.e. Lebesgue}(\mathbf{x}).$$

(ii) If $p > 1$, and $|f(\mathbf{x} + \mathbf{t}) - f(\mathbf{x})| \leq c|\mathbf{t}|^\gamma$, for $a < \gamma < 1$, $a > 0$, then

$$\left| \int \delta_m(\mathbf{x}, \mathbf{t}) f(\mathbf{t}) d\mathbf{t} - f(\mathbf{x}) \right| = O\left(m^{-\delta}\right),$$

where $\delta = \min\{q^{-1}, \gamma\}$, with $q^{-1} = 1 - p^{-1}$.

Lemma 4.8.0.2 (Theorem A, Serfling [1980] (10, p.201)) Let $k > 0, n \geq k, \mathbf{X}_1, \dots, \mathbf{X}_n$ i.i.d. random vectors with values in a measurable space \mathcal{X} and $g : \mathcal{X}^k \rightarrow [a, b]$ be a real bounded function. Set

$$\theta := \mathbb{E}[g(\mathbf{X}_{1:k})] \quad \text{and} \quad \sigma^2 := \text{Var}[g(\mathbf{X}_{1:k})].$$

Then, for any $t > 0$

$$\mathbb{P}\left(\left(\binom{n}{k}\right)^{-1} \sum_{i \in \mathfrak{S}_{k,n}^\uparrow} g(\mathbf{X}_{i(1:k)}) - \theta \geq t\right) \leq \exp\left(-\frac{[n/k]t^2}{2\sigma^2 + (2/3)(b-\theta)t}\right),$$

where $\mathfrak{J}_{k,n}$ is the set of bijective functions from $\{1, \dots, k\}$ to $\{1, \dots, n\}$ and $\mathfrak{S}_{k,n}^\uparrow$ is the subset of $\mathfrak{J}_{k,n}$ made of increasing functions.

Remark 4.8.0.3 Note that $g(\cdot)$ does not need to be symmetric for this bound to hold. Indeed, if $g(\cdot)$ is not symmetric, we can nonetheless apply this lemma to the symmetrized version \tilde{g} defined as

$$\tilde{g}(\mathbf{x}_{1:k}) := (k!)^{-1} \sum_{i \in \mathfrak{J}_{k,k}} g(\mathbf{x}_{i(1:k)}),$$

and we get the result.

4.9 References

- Abrevaya, J. and Jiang, W. (2005a). A nonparametric approach to measuring and testing curvature. *J. Bus. Econom. Statist.*, **23**(1), 1–19. [128](#)
- Abrevaya, J. and Jiang, W. (2005b). A nonparametric approach to measuring and testing curvature. *J. Bus. Econom. Statist.*, **23**(1), 1–19. [150](#)
- Ahmad, I. A. (1981). A note on nonparametric density estimation for dependent variables using a delta sequence. *Ann. Inst. Statist. Math.*, **33**(2), 247–254. [130](#)
- Arcones, M. A. and Giné, E. (1993). Limit theorems for U-processes. *Ann. Probab.*, **21**(3), 1494–1542. [128](#)

- Arcones, M. A. and Giné, E. (1995). On the law of the iterated logarithm for canonical U-statistics and processes. *Stochastic Process. Appl.*, **58**(2), 217–245. [128](#)
- Arcones, M. A. and Wang, Y. (2006). Some new tests for normality based on U-processes. *Statist. Probab. Lett.*, **76**(1), 69–82. [128](#)
- Arcones, M. A., Chen, Z., and Giné, E. (1994). Estimators related to U-processes with applications to multivariate medians: asymptotic normality. *Ann. Statist.*, **22**(3), 1460–1477. [128](#)
- Basu, A. K. and Sahoo, D. K. (1985). Speed of convergence in density estimation by delta sequence and estimation of mode. *Calcutta Statist. Assoc. Bull.*, **34**(133-134), 1–13. [130](#)
- Belitser, E. (2000). Local minimax pointwise estimation of a multivariate density. *Statist. Neerlandica*, **54**(3), 351–365. [130](#)
- Bellet, A. and Habrard, A. (2015). Robustness and generalization for metric learning. *Neuro-computing*, **151**, 259–267. [146](#)
- Bellet, A., Habrard, A., and Sebban, M. (2013). A survey on metric learning for feature vectors and structured data. *CoRR*, **abs/1306.6709**. [145](#)
- Bergsma, W. and Dassios, A. (2014). A consistent test of independence based on a sign covariance related to Kendall’s tau. *Bernoulli*, **20**(2), 1006–1028. [152](#)
- Blum, J. R., Kiefer, J., and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.*, **32**, 485–498. [152](#)
- Blundell, R., Gosling, A., Ichimura, H., and Meghir, C. (2007). Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica*, **75**(2), 323–363. [149](#)
- Borovkova, S., Burton, R., and Dehling, H. (1999). Consistency of the Takens estimator for the correlation dimension. *Ann. Appl. Probab.*, **9**(2), 376–390. [153](#)
- Borovkova, S., Burton, R., and Dehling, H. (2001). Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. *Trans. Amer. Math. Soc.*, **353**(11), 4261–4318. [128](#)
- Borovskikh, Y. V. (1996). *U-statistics in Banach spaces*. VSP, Utrecht. [128](#)
- Bouzebda, S. (2011). Some new multivariate tests of independence. *Math. Methods Statist.*, **20**(3), 192–205. [152](#)
- Bouzebda, S. (2012). On the strong approximation of bootstrapped empirical copula processes with applications. *Math. Methods Statist.*, **21**(3), 153–188. [130](#)

- Bouzebda, S. (2014). General tests of independence based on empirical processes indexed by functions. *Stat. Methodol.*, **21**, 59–87. [152](#)
- Bouzebda, S. (2023). General tests of conditional independence based on empirical processes indexed by functions. *Jpn. J. Stat. Data Sci.*, pages 1–65. [135](#)
- 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**(4), 864–914. [143](#)
- Bouzebda, S. and Keziou, A. (2010a). New estimates and tests of independence in semiparametric copula models. *Kybernetika (Prague)*, **46**(1), 178–201. [152](#)
- Bouzebda, S. and Keziou, A. (2010b). A new test procedure of independence in copula models via χ^2 -divergence. *Comm. Statist. Theory Methods*, **39**(1-2), 1–20. [152](#)
- Bouzebda, S. and Nemouchi, B. (2019). Central limit theorems for conditional empirical and conditional U-processes of stationary mixing sequences. *Math. Methods Statist.*, **28**(3), 169–207. [129](#)
- 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**(2), 452–509. [129](#)
- 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*, pages 1–56. [154](#)
- 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. [129](#)
- Bouzebda, S. and Soukariéh, I. (2023a). Non-parametric conditional U-processes for locally stationary functional random fields under stochastic sampling design. *Mathematics*, **11**(1), 1–70. [144](#)
- Bouzebda, S. and Soukariéh, I. (2023b). Renewal type bootstrap for U-process Markov chains. *Markov Process. Related Fields*, pages 1–52. [128](#)
- Bouzebda, S. and Taachouche, N. (2022). Rates of the strong uniform consistency for the kernel-type regression function estimators with general kernels on manifolds. *Math. Methods Statist.*, pages 1–62. [143](#)
- Bouzebda, S. and Zari, T. (2013). Asymptotic behavior of weighted multivariate Cramér-von Mises-type statistics under contiguous alternatives. *Math. Methods Statist.*, **22**(3), 226–252. [152](#)

- 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**(2), 321–358. [129](#)
- Bouzebda, S., El-hadjali, T., and Ferfache, A. A. (2022). Uniform in bandwidth consistency of conditional U-statistics adaptive to intrinsic dimension in presence of censored data. *Sankhya A*, pages 1–59. [129](#), [144](#)
- Bouzebda, S., Nezzal, A., and Zari, T. (2023). Uniform consistency for functional conditional U-statistics using delta-sequences. *Mathematics*, **11**(1), 1–39. [130](#), [144](#)
- Brunel, E. and Comte, F. (2006). Adaptive nonparametric regression estimation in presence of right censoring. *Math. Methods Statist.*, **15**(3), 233–255. [143](#)
- Cao, Q., Guo, Z.-C., and Ying, Y. (2016). Generalization bounds for metric and similarity learning. *Mach. Learn.*, **102**(1), 115–132. [146](#)
- Carbonez, A., Györfi, L., and van der Meulen, E. C. (1995). Partitioning-estimates of a regression function under random censoring. *Statist. Decisions*, **13**(1), 21–37. [142](#)
- Chen, X. and Kato, K. (2020). Jackknife multiplier bootstrap: finite sample approximations to the U-process supremum with applications. *Probab. Theory Related Fields*, **176**(3-4), 1097–1163. [149](#), [150](#)
- Chen, Y. and Datta, S. (2019). Adjustments of multi-sample U-statistics to right censored data and confounding covariates. *Comput. Statist. Data Anal.*, **135**, 1–14. [143](#)
- Chentsov, N. N. (1962). Evaluation of an unknown distribution density from observations. *Sov. Math., Dokl.*, **3**, 1559–1562. [129](#)
- Cléménçon, S. and Robbiano, S. (2015). The TreeRank Tournament algorithm for multipartite ranking. *J. Nonparametr. Stat.*, **27**(1), 107–126. [146](#)
- Cléménçon, S., Lugosi, G., and Vayatis, N. (2008). Ranking and empirical minimization of U-statistics. *Ann. Statist.*, **36**(2), 844–874. [147](#)
- Cléménçon, S., Robbiano, S., and Vayatis, N. (2013). Ranking data with ordinal labels: optimality and pairwise aggregation. *Mach. Learn.*, **91**(1), 67–104. [146](#)
- Cléménçon, S., Colin, I., and Bellet, A. (2016). Scaling-up empirical risk minimization: optimization of incomplete U-statistics. *J. Mach. Learn. Res.*, **17**, Paper No. 76, 36. [145](#), [146](#)
- Cristóbal Cristóbal, J. A., Faraldo Roca, P., and González Manteiga, W. (1987). A class of linear regression parameter estimators constructed by nonparametric estimation. *Ann. Statist.*, **15**(2), 603–609. [130](#)

- 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**(4), 680–700. [143](#)
- de la Peña, V. H. and Giné, E. (1999). *Decoupling*. Probability and its Applications (New York). Springer-Verlag, New York. From dependence to independence, Randomly stopped processes. U-statistics and processes. Martingales and beyond. [128](#)
- Deheuvels, P. (2011). One bootstrap suffices to generate sharp uniform bounds in functional estimation. *Kybernetika (Prague)*, **47**(6), 855–865. [135](#)
- Denker, M. and Keller, G. (1983). On U-statistics and v. Mises' statistics for weakly dependent processes. *Z. Wahrsch. Verw. Gebiete*, **64**(4), 505–522. [128](#)
- Derumigny, A. (2019). Estimation of a regular conditional functional by conditional U-statistics regression. *arXiv preprint arXiv:1903.10914*. [155](#)
- Dony, J. and Mason, D. M. (2008). Uniform in bandwidth consistency of conditional U-statistics. *Bernoulli*, **14**(4), 1108–1133. [129](#), [154](#)
- Dwass, M. (1956). The large-sample power of rank order tests in the two-sample problem. *Ann. Math. Statist.*, **27**, 352–374. [149](#)
- Ellison, G. and Ellison, S. F. (2011). Strategic entry deterrence and the behavior of pharmaceutical incumbents prior to patent expiration. *American Economic Journal: Microeconomics*, **3**(1), 1–36. [149](#)
- Földes, A. and Rejtő, L. (1981). A LIL type result for the product limit estimator. *Z. Wahrsch. Verw. Gebiete*, **56**(1), 75–86. [144](#)
- Földes, A. and Révész, P. (1974). A general method for density estimation. *Stud. Sci. Math. Hungar.*, **9**, 81–92 (1975). [132](#)
- Gao, J. and Gijbels, I. (2008). Bandwidth selection in nonparametric kernel testing. *J. Amer. Statist. Assoc.*, **103**(484), 1584–1594. [155](#)
- Ghosal, S., Sen, A., and van der Vaart, A. W. (2000a). Testing monotonicity of regression. *Ann. Statist.*, **28**(4), 1054–1082. [128](#)
- Ghosal, S., Sen, A., and van der Vaart, A. W. (2000b). Testing monotonicity of regression. *Ann. Statist.*, **28**(4), 1054–1082. [149](#), [151](#)
- Giné, E. and Mason, D. M. (2007a). Laws of the iterated logarithm for the local U-statistic process. *J. Theoret. Probab.*, **20**(3), 457–485. [128](#), [129](#)
- Giné, E. and Mason, D. M. (2007b). On local U-statistic processes and the estimation of densities of functions of several sample variables. *Ann. Statist.*, **35**(3), 1105–1145. [128](#)

- Györfi, L., Kohler, M., Krzyżak, A., and Walk, H. (2002). *A distribution-free theory of non-parametric regression*. Springer Series in Statistics. Springer-Verlag, New York. [140](#)
- Halmos, P. R. (1946). The theory of unbiased estimation. *Ann. Math. Statistics*, **17**, 34–43. [128](#)
- Harel, M. and Puri, M. L. (1996). Conditional U-statistics for dependent random variables. *J. Multivariate Anal.*, **57**(1), 84–100. [129](#)
- Hengartner, N. W. and Sperlich, S. (2005). Rate optimal estimation with the integration method in the presence of many covariates. *J. Multivariate Anal.*, **95**(2), 246–272. [141](#)
- Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics*, **19**, 293–325. [128](#), [148](#), [151](#)
- Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, **58**, 13–30. [169](#)
- Hollander, M. and Proschan, F. (1972). Testing whether new is better than used. *Ann. Math. Statist.*, **43**, 1136–1146. [152](#)
- Horowitz, J. L. and Spokoiny, V. G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, **69**(3), 599–631. [155](#)
- Isogai, E. (1990). Nonparametric estimation of a regression function by delta sequences. *Ann. Inst. Statist. Math.*, **42**(4), 699–708. [130](#)
- Jin, R., Wang, S., and Zhou, Y. (2009). Regularized distance metric learning: theory and algorithm. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 22. Curran Associates, Inc. [146](#)
- Joly, E. and Lugosi, G. (2016). Robust estimation of U-statistics. *Stochastic Process. Appl.*, **126**(12), 3760–3773. [128](#)
- Jones, M. C., Davies, S. J., and Park, B. U. (1994). Versions of kernel-type regression estimators. *J. Amer. Statist. Assoc.*, **89**(427), 825–832. [141](#)
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53**, 457–481. [142](#)
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**(1/2), 81–93. [147](#)
- Kohler, M., Máthé, K., and Pintér, M. (2002). Prediction from randomly right censored data. *J. Multivariate Anal.*, **80**(1), 73–100. [142](#), [143](#)

- 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. 128
- Lee, A. J. (1990). *U-statistics*, volume 110 of *Statistics: Textbooks and Monographs*. Marcel Dekker Inc., New York. Theory and practice. 128, 151
- Lee, S., Linton, O., and Whang, Y.-J. (2009a). Testing for stochastic monotonicity. *Econometrica*, **77**(2), 585–602. 128
- Lee, S., Linton, O., and Whang, Y.-J. (2009b). Testing for stochastic monotonicity. *Econometrica*, **77**(2), 585–602. 149
- Lehmann, E. L. (1951). A general concept of unbiasedness. *Ann. Math. Statistics*, **22**, 587–592. 149
- Leucht, A. (2012). Degenerate U- and V-statistics under weak dependence: asymptotic theory and bootstrap consistency. *Bernoulli*, **18**(2), 552–585. 128
- Leucht, A. and Neumann, M. H. (2013). Degenerate U- and V-statistics under ergodicity: asymptotics, bootstrap and applications in statistics. *Ann. Inst. Statist. Math.*, **65**(2), 349–386. 128
- Linton, O. B. and Jacho-Chávez, D. T. (2010). On internally corrected and symmetrized kernel estimators for nonparametric regression. *TEST*, **19**(1), 166–186. 140, 141
- Mack, Y. P. and Müller, H.-G. (1989). Derivative estimation in nonparametric regression with random predictor variable. *Sankhyā Ser. A*, **51**(1), 59–72. 140
- 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 Statist.*, **18**(2), 159–184. 141, 143
- Marron, J. S. and Härdle, W. (1986). Random approximations to some measures of accuracy in nonparametric curve estimation. *J. Multivariate Anal.*, **20**(1), 91–113. 130
- Nadaraja, E. A. (1964). On a regression estimate. *Teor. Veroyatnost. i Primenen.*, **9**, 157–159. 128
- Nadaraja, E. A. (1965). On non-parametric estimates of density functions and regression. *Teor. Veroyatnost. i Primenen.*, **10**, 199–203. 129
- Nolan, D. and Marron, J. S. (1989). Uniform consistency of automatic and location-adaptive delta-sequence estimators. *Probab. Theory Related Fields*, **80**(4), 619–632. 130

- Ouassou, I. and Rachdi, M. (2012). Regression operator estimation by delta-sequences method for functional data and its applications. *AStA Adv. Stat. Anal.*, **96**(4), 451–465. [130](#)
- Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, **33**, 1065–1076. [132](#)
- Piterbarg, V. I. (1996). *Asymptotic methods in the theory of Gaussian processes and fields*, volume 148 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI. Translated from the Russian by V. V. Piterbarg, Revised by the author. [149](#)
- Prakasa Rao, B. L. S. (1978). Density estimation for Markov processes using delta-sequences. *Ann. Inst. Statist. Math.*, **30**(2), 321–328. [130](#)
- Prakasa Rao, B. L. S. (2010). Nonparametric density estimation for functional data by delta sequences. *Braz. J. Probab. Stat.*, **24**(3), 468–478. [130](#)
- Prakasa Rao, B. L. S. and Sen, A. (1995). Limit distributions of conditional U-statistics. *J. Theoret. Probab.*, **8**(2), 261–301. [129](#)
- Rejchel, W. (2012). On ranking and generalization bounds. *J. Mach. Learn. Res.*, **13**, 1373–1392. [147](#)
- Schick, A., Wang, Y., and Wefelmeyer, W. (2011). Tests for normality based on density estimators of convolutions. *Statist. Probab. Lett.*, **81**(2), 337–343. [128](#)
- Sen, A. (1994). Uniform strong consistency rates for conditional U-statistics. *Sankhyā Ser. A*, **56**(2), 179–194. [129](#)
- Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. John Wiley & Sons, Inc., New York. Wiley Series in Probability and Mathematical Statistics. [175](#), [177](#)
- 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**(3), 599–615. [154](#)
- Shen, J. and Xie, Y. (2013). Strong consistency of the internal estimator of nonparametric regression with dependent data. *Statist. Probab. Lett.*, **83**(8), 1915–1925. [141](#)
- Sherman, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica*, **61**(1), 123–137. [128](#)
- Sherman, R. P. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Ann. Statist.*, **22**(1), 439–459. [128](#)
- Silverman, B. W. (1978). Distances on circles, toruses and spheres. *J. Appl. Probability*, **15**(1), 136–143. [151](#)

- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London. 154
- Solon, G. (1992). Intergenerational income mobility in the united states. *The American Economic Review*, **82**(3), 393–408. 149
- Soukarieh, I. and Bouzebda, S. (2022). Exchangeably weighted bootstraps of general Markov U-process. *Mathematics*, **10**(20), 1–42. 128
- Soukarieh, I. and Bouzebda, S. (2023). Renewal type bootstrap for increasing degree U-process of a Markov chain. *J. Multivariate Anal.*, **195**, 105143. 128
- Stute, W. (1991). Conditional U-statistics. *Ann. Probab.*, **19**(2), 812–825. 128, 129, 131, 172
- Stute, W. (1993). Almost sure representations of the product-limit estimator for truncated data. *Ann. Statist.*, **21**(1), 146–156. 128
- Stute, W. (1994a). L^p -convergence of conditional U-statistics. *J. Multivariate Anal.*, **51**(1), 71–82. 144
- Stute, W. (1994b). Universally consistent conditional U-statistics. *Ann. Statist.*, **22**(1), 460–473. 144
- Stute, W. (1996). Symmetrized NN-conditional U-statistics. In *Research developments in probability and statistics*, pages 231–237. VSP, Utrecht. 129
- Stute, W. and Wang, J.-L. (1993). Multi-sample U-statistics for censored data. *Scand. J. Statist.*, **20**(4), 369–374. 143
- Susarla, V. and Walter, G. (1981). Estimation of a multivariate density function using delta sequences. *Ann. Statist.*, **9**(2), 347–355. 130, 132, 158, 176
- von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statistics*, **18**, 309–348. 128
- Walter, G. and Blum, J. (1979). Probability density estimation using delta sequences. *Ann. Statist.*, **7**(2), 328–340. 132
- Walter, G. G. (1983). A general approach to classification problems. *Inform. Sci.*, **30**(1), 67–77. 130
- Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*, volume 60 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, Ltd., London. 140
- Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A*, **26**, 359–372. 128, 129
- Watson, G. S. and Leadbetter, M. R. (1964). Hazard analysis. II. *Sankhyā Ser. A*, **26**, 101–116. 130, 132

Winter, B. B. (1975). Rate of strong consistency of two nonparametric density estimators. *Ann. Statist.*, **3**, 759–766. [130](#)

Yuan, A., Giurcanu, M., Luta, G., and Tan, M. T. (2017). U-statistics with conditional kernels for incomplete data models. *Ann. Inst. Statist. Math.*, **69**(2), 271–302. [143](#)

Chapter 5

Uniform consistency for functional conditional U-statistics using delta-sequences

Ce chapitre développe le contenu d'un article publié [Mathematics (2023), 11(1), 161. <https://doi.org/10.3390/math11010161>], mis en forme pour être inséré dans le présent manuscrit de thèse.

Title : *Uniform consistency for functional conditional U-statistics using delta-sequences.*

abstract

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

Keywords : Nonparametric estimation; regression-type models; U-statistics; conditional distribution; delta sequences; kernel estimation; machine learning problems.

Mathematics Subject Classification : 60F05 ; 60G15 ; 60G10 ; 62G05 ; 62G07; 62H 12.

5.1 Introduction

The regression problem has been studied by statisticians and probability theorists for many years to keep up with the various problems and topics brought up by technological and computational advances, resulting in the creation of many advanced and complex techniques. Among the problems addressed are modeling, estimation method applications, and tests. In this paper, we are interested in nonparametric regression estimation. Unlike the parametric framework, where one must estimate a finite number of parameters based on a specified structural model a priori, nonparametric estimation does not require any specific structure; instead, it allows the data to speak for themselves. However, as natural drawbacks, nonparametric procedures are more susceptible to estimation biases and losses in convergence rates than parametric methods. Since their introduction, kernel nonparametric function estimation approaches have garnered a significant amount of attention; for references to research literature and statistical applications in this area, consult [Silverman \[1986\]](#), [Nadaraya \[1989\]](#), [Härdle \[1990\]](#), [Wand and Jones \[1995\]](#), [Eggermont and LaRiccia \[2001\]](#), [Devroye and Lugosi \[2001\]](#), [Jiang \[2022\]](#) and the references therein. Popular as they may be, they represent only one of the numerous methods for developing accurate function estimators. These include nearest-neighbor, spline, neural network, and wavelet approaches. In addition, these techniques have been applied to a vast array of data types. This article will focus on constructing consistent estimators for the conditional U-statistics for functional data based on the delta sequence. The theory of U-statistics and U-processes, which was initially introduced in the seminal work of [Hoeffding \[1948\]](#), has garnered a significant amount of interest over the course of the last few decades as a result of the diverse applications to which it has been applied. U-statistics can be utilized to solve complex statistical problems. Among the examples are nonparametric regression, density estimation, and goodness-of-fit tests. Furthermore, U-statistics contribute to the study of estimators with various degrees of smoothness (including function estimators). [Stute \[1993\]](#), for instance, analyzes the product limit estimator for truncated data applying a.s. uniform bounds for \mathbb{P} -canonical U-processes. [Arcones and Wang \[2006\]](#) introduces two new normality tests based on U-processes. Using the findings of [Giné and Mason \[2007a,b\]](#), [Schick et al. \[2011\]](#) provided new normality tests that used as test statistics weighted L_1 -distances between the standard normal density and local U-statistics that were based on standardized observations. These tests were used to determine whether or not the data was normally distributed. [Joly and Lugosi \[2016\]](#) worked on the estimate of the mean of multivariate functions under the premise of possibly heavy-tailed distributions and presented the U-based median-of-means. U-processes are also necessary for a wide variety of statistical applications, such as the examination of qualitative aspects of functions in nonparametric statistics [[Lee et al. \[2009\]](#), [Ghosal et al. \[2000\]](#)] as well as establishing limiting distributions of M-estimators [see, for example, [Arcones and Giné \[1993\]](#), [Sherman \[1994\]](#), [de la Peña and Giné \[1999\]](#)]. In [Boniece et al. \[2022\]](#) the authors consider the problem of detecting distributional changes in a sequence of high dimensional data by using the weighted cumulative sums of U-statistics stemming from L_p norms. In [Kulathinal and Dewan \[2022\]](#), the authors proposed

tests based on U-statistics for testing the equality of marginal density functions. In the paper [Minsker \[2022\]](#), the following problem is considered : Is it possible, given a sample of random variables that are independent, identically distributed, and have a finite variance, to build an estimator of the unknown mean that performs almost as well as if the data were normally distributed? The argument that was presented in the previous work is based on a new deviation inequality for the U-statistics of order that is permitted to rise with sample size. This inequality is the most important part of the argument. The first asymptotic results for the scenario in which the underlying random variables are assumed to be independent and distributed in an identical fashion were presented by [Halmos \[1946\]](#), [von Mises \[1947\]](#) and [Hoeffding \[1948\]](#) (amongst others). In contrast, the asymptotic results under weak dependency assumption were demonstrated in [Borovkova et al. \[2001\]](#), in [Denker and Keller \[1983\]](#) or more recently in [Leucht \[2012\]](#) and in more general setting in [Bouzebda and Soukariéh \[2023\]](#); [Leucht and Neumann \[2013\]](#); [Soukariéh and Bouzebda \[2022, 2023\]](#). The interested reader may refer to [Lee \[1990\]](#) and [Arcones and Giné \[1993\]](#) for an excellent collection of references on U-statistics and U-processes. We also refer to [de la Peña and Giné \[1999\]](#) for a profound understanding of the theory of U-processes.

In the present work, we consider the conditional U-statistics introduced by [Stute \[1991\]](#), that can be seen as generalizations of the Nadaraya-Watson ([Nadaraja \[1964\]](#), and [Watson \[1964\]](#)) regression function estimates. To be more precise, let us consider the sequence of independent and identically distributed random vectors $\{\mathbf{X}_i, \mathbf{Y}_i, i \in \mathbb{N}^*\}$ with $\mathbf{X}_i \in \mathbb{R}^d$ and $\mathbf{Y}_i \in \mathbb{R}^{d'}$, $d, d' \geq 1$. Let $\varphi : \mathbb{R}^{d'k} \rightarrow \mathbb{R}$ denote a measurable function. Within the scope of this work, our primary focus is on the estimation of the conditional expectation or regression function, as follows:

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

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

$$\sup_{\mathbf{x} \in \mathbb{R}^d} |K(\mathbf{x})| =: \kappa < \infty \text{ and } \int_{\mathbb{R}^d} K(\mathbf{x}) d\mathbf{x} = 1. \quad (5.1.1)$$

[Stute \[1991\]](#) established a new category of estimators for $r^{(k)}(\varphi, \mathbf{t})$, called conditional U-statistics, that is defined for each $\mathbf{t} \in \mathbb{R}^{dk}$ to be :

$$\widehat{r}_n^{(k)}(\varphi, \mathbf{t}; h_n) = \frac{\sum_{(i_1, \dots, i_k) \in I(k, n)} \varphi(Y_{i_1}, \dots, Y_{i_k}) K\left(\frac{\mathbf{t}_1 - \mathbf{X}_{i_1}}{h_n}\right) \dots K\left(\frac{\mathbf{t}_k - \mathbf{X}_{i_k}}{h_n}\right)}{\sum_{(i_1, \dots, i_k) \in I(k, n)} K\left(\frac{\mathbf{x}_1 - \mathbf{X}_{i_1}}{h_n}\right) \dots K\left(\frac{\mathbf{x}_k - \mathbf{X}_{i_k}}{h_n}\right)}, \quad (5.1.2)$$

where :

$$I(k, n) = \{\mathbf{i} = (i_1, \dots, i_k) : 0 \leq i_j \leq n \text{ and } i_j \neq i_r \text{ if } j \neq r\},$$

is the set of all k -tuples of different integers between 1 and n and $\{h_n\}_{n \geq 1}$ is a sequence of positive constants that, at a certain rate, converge to the value zero, $nh_n^{dk} \rightarrow \infty$. In the particular

case $k = 1$, the $r^{(k)}(\varphi, \mathbf{t})$ reduces to $r^{(1)}(\varphi, \mathbf{t}) = \mathbb{E}(\varphi(\mathbf{Y})|\mathbf{X} = \mathbf{t})$ and the estimator developed by Stute is now known as the Nadaraya-Watson estimator $r^{(1)}(\varphi, \mathbf{t})$ given by :

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

The work of Sen [1994] focused on the estimation of the rate of the uniform convergence in \mathbf{t} of $\widehat{r}_n^{(k)}(\varphi, \mathbf{t}; h_n)$ to $r^{(k)}(\varphi, \mathbf{t})$. In Prakasa Rao and Sen [1995], the limit distributions of $\widehat{r}_n^{(k)}(\varphi, \mathbf{t}; h_n)$ are analyzed and compared to those produced by Stute. Under suitable mixing conditions, Harel and Puri [1996] extend the results of Stute [1991] to weakly dependent data and uses their findings to evaluate the Bayes risk consistency of the corresponding discriminating rules.

As alternatives to the standard kernel-type estimators, Stute [1996] presented symmetrized nearest neighbor conditional U-statistics. The work of has seen a major advancement because to the contributions of Dony and Mason [2008] where a far more strong version of consistency can be found, to be specific, uniform in \mathbf{t} and in bandwidth consistency (i.e., $h_n, h_n \in [a_n, b_n]$ where $a_n < b_n \rightarrow 0$ at some specific rate) of $\widehat{r}_n^{(k)}(\varphi, \mathbf{t}; h_n)$. Additionally, uniform consistency is achieved across $\varphi \in \mathcal{F}$ for a suitably restricted class of functions \mathcal{F} , extended in Bouzebda and Nemouchi [2019], Bouzebda and Nemouchi [2020], Bouzebda *et al.* [2021], Bouzebda *et al.* [2022b] and Bouzebda and Nezzal [2022]. The key component of their findings is the local conditional U-process studied in Giné and Mason [2007a].

The case of functional data is the primary focus of this research. We present an excerpt from Aneiros *et al.* [2019]: "Functional data analysis (FDA) is a branch of statistics concerned with analyzing infinite-dimensional variables such as curves, sets, and images. It has undergone phenomenal growth over the past 20 years, stimulated partly by significant data collection technology advances that have brought about the "Big Data" revolution. Often perceived as a somewhat arcane area of research at the turn of the century, FDA is now one of the most active and relevant fields of investigation in data science." For an introduction to the subject of FDA, we refer to the books of Ramsay and Silverman [2002a,b], Ferraty and Vieu [2006a], which contain different case studies in economics, archaeology, criminology, and neurophysiology, as well as fundamental analysis techniques. It is important to note that the extension of probability theory to random variables with values in normed vector spaces (such as Banach and Hilbert spaces), in conjunction with extensions of certain classical asymptotic limit theorems, predates the recent literature on functional data. This fact can be demonstrated by tracing back the history of the subject (see for instance, Araujo and Giné [1980]). Gasser *et al.* [1998] investigated density and mode estimation for data with values in a normed vector space. At the same time, he brought attention to the issue of the curse of dimensionality, which affects functional data, and he suggested potential remedies to the problem. In the context of regression estimation, Ferraty and Vieu [2006a] considered the nonparametric models. We may refer also to Bosq [2000], Horváth and Kokoszka [2012] and Ling and Vieu [2018]. Recently, Modern theory has been applied to the treatment of functional data. For instance, Ferraty *et al.* [2010] provided the consistency rates of several functionals of the conditional distribution, such as the regression function, the conditional cumulative distribution, and the conditional density, amongst others,

uniformly over a subset of the explanatory variable. Other examples include conditional density, which is a measure of the density of the conditional distribution, and conditional cumulative distribution, which is a measure of the conditional cumulative distribution. [Kara-Zaitri et al. \[2017\]](#) also investigated the consistency rates for some functionals nonparametric models, such as the regression function, the conditional distribution, the conditional density, and the conditional hazard function, uniformly in bandwidth (UIB consistency) extended to the ergodic setting by [Bouzebda and Chaouch \[2022\]](#). In the paper [Attouch et al. \[2019\]](#), the topic of local linear estimation of the regression function in the case when the regressor is functional was investigated, and the results indicated robust convergence (with rates) consistently across bandwidth parameters. In the work of [Ling et al. \[2019\]](#), the k -nearest neighbors (kNN) estimate of the nonparametric regression model for heavy mixing of functional time series data was explored. Under some mild conditions, a uniform and practically perfect convergence rate of the k -nearest neighbors estimator was established. In the work [Bouzebda et al. \[2016\]](#), the authors offer a variety of solutions for limiting laws for the conditional mode in the functional setting for ergodic data; for some current references, see the following: [Mohammedi et al. \[2021\]](#), [Bouzebda et al. \[2022a\]](#), [Bouzebda and Nezzal \[2022\]](#), [Bouzebda and Nemouchi \[2022\]](#), [Almanjahie et al. \[2022a,b\]](#); [Litimein et al. \[2023\]](#).

We will consider a general method for functional estimation by using the delta sequences. Delta sequences (also called “approximate identities” or “summability kernels”) arise in a wide variety of subfields within mathematics. Still, the applications that pertain to the theory of generalized functions are likely the most significant ones. The regularization of generalized functions is the major application for delta sequences. The proposed methods generalize several nonparametric estimation methods, including the kernel estimators given in (5.1.2) of [Stute \[1991\]](#). To be more precise, the broad class of delta-sequence estimators includes the histogram estimators, Chentsov’s projection estimators [Chentsov \[1962\]](#), and nearest-neighbor estimators, among others. Certain types of these sequences were already studied by [Watson and Leadbetter \[1964\]](#), who called them “ δ -function sequences”. They established, among other things, the asymptotic unbiasedness and the asymptotic variance of estimators based on them but did not consider convergence rates. [Winter \[1975\]](#) obtained the rate of strong consistency and the rate of asymptotic bias for estimators associated with delta sequences arising from the Fejér kernel of the Fourier series. The delta sequence method of density estimation of [Susarla and Walter \[1981\]](#) is extended to certain non-i.i.d. cases in [Prakasa Rao \[1978\]](#), where it is assumed that the observations are taken from a stationary Markov process. [Ahmad \[1981\]](#) considered the delta-sequence estimator for the marginal distribution of a strictly stationary stochastic process satisfying some mixing conditions. In [Basu and Sahoo \[1985\]](#), the author investigated the local and global convergence rates of delta-sequence type estimators of the density function, its derivative, and its mode. [Nolan and Marron \[1989\]](#) proved the uniform strong consistency of delta-sequence estimators. [Isogai \[1990\]](#) partially generalized the usual nonparametric estimators of a regression function by using an estimator based on quasi-positive delta sequences. [Marron and Härdle \[1986\]](#) considered a general nonparametric statistical curve estimation setting

called the class of “fractional delta sequence estimators”. [Cristóbal Cristóbal *et al.* \[1987\]](#) used the delta method to investigate the correlation model. [Prakasa Rao \[2010a\]](#) looked at the problem of estimating the density function of functional data with values in an infinite-dimensional separable Banach space using the method of delta sequences; for further information, we can also look into [Ouassou and Rachdi \[2012\]](#). The copula estimation using the delta sequences methods is considered in [Bouzebda \[2012\]](#). The problem of the nonparametric minimax estimation of a multivariate density at a given point by the delta sequences was investigated in [Belitser \[2000\]](#). [Walter \[1983\]](#) used the delta sequence to propose an essential application to the classification problem of the value of the discrete random variable.

The goal of the current study is to present and investigate the delta sequences estimators for the conditional U-statistics for functional data, more specifically for random elements taking values in an infinite-dimensional separable Banach space, such as the space of continuous functions on the interval $[0, 1]$ endowed with the supremum norm. This will allow the delta sequences estimators for the conditional U-statistics for functional data to be utilized for functional data analysis. Examples of functional data that can appear in these spaces include stochastic processes with continuous sample paths on a finite interval associated with the supremum norm and stochastic processes whose sample paths are square-integrable on the real line. Both of these types of stochastic processes can occur on the real line. The dimensionality problem must be addressed in nonparametric functional data analysis in two ways: first, by working with data that have an infinite number of dimensions, and second, by making universal assumptions about the infinite number of dimensions for the probability distribution of variables in nonparametric modeling. This structure’s twofold infinity of dimensions is the basis for all subsequent developments in the discipline. Our previous work, delivered in the multivariate setting and cited as [Bouzebda and Nezzal \[2023\]](#), is extended here in the present study. Although the concept behind our estimation approach is similar to that presented in [Bouzebda and Nezzal \[2023\]](#) (containing the Stute estimator), we make allowances for the infinite dimensionality of the covariate. We have determined the asymptotic characteristics of the multivariate delta sequence estimators, [Bouzebda and Nezzal \[2023\]](#). Their findings do not directly apply to the current situation since we are working with a covariate with an unlimited number of dimensions. As a result, we must utilize other reasoning in our proofs to deal with the broader framework. These findings are beneficial in their own right, but they are also necessary for the inquiry being conducted in this work. To “simply” combine ideas from other publications would not be sufficient to solve the issue, as will be demonstrated in the following paragraphs. To be able to deal with delta sequence U-statistic estimators for functional data, you will need to resort to intricate mathematical derivations. Compared to the previous studies written on delta sequence estimators, the current paper considers the setting of an unbounded function φ , which adds a significant amount of complexity to the proof. The general assumptions that are required for the derivations of the asymptotic results for the conditional U-statistics delta sequence estimators are presented in this study.

The format of this article is structured as follows. Section [4.2](#) is devoted to introducing the

delta sequences and the definitions we need in our work, where we introduce the new family of estimators. Section 5.3 gives the paper's main results concerning the uniform convergences. In Section 5.4, we present a significant application for the censored data context of its interest. In Section 5.5, we provide some applications, including the Kendall rank correlation coefficient in §5.5.1, the discrimination in §5.5.2, the the metric learning in §5.5.3 and the time series prediction from a continuous set of past values in §5.5.4. Some final observations and possible developments in the future are moved to Section 6.3. To maintain a smooth flow throughout the presentation, all proofs have been compiled in Section 5.6. A selection of significant technical findings is presented in Appendix 5.6.

5.2 Preliminaries and estimation procedure

Let $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a probability space, $(\mathcal{X}, d(\cdot, \cdot))$ denotes an infinite-dimensional separable Banach space equipped with a norm $\|\cdot\|$ such that $d(u, v) = \|u - v\|$ and \mathcal{B} be the σ -algebra of Borel subsets of \mathcal{X} . Let us consider a sequence $\{X_i, Y_i : i \geq 1\}$ of independent identically distributed random copies of the random element (X, Y) , where X is a random element defined on $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in $(\mathcal{X}, \mathcal{B})$ and Y takes values in some abstract space $(\mathcal{Y}, \mathcal{B}')$. Fu [2012] introduced the functional conditional U-statistics when $\mathbf{x} \in \mathcal{X}^m$ some semi-metric space as a generalization of Stute's estimator by :

$$\hat{r}_n^{(m)}(\varphi, \mathbf{x}; h_K) = \frac{\sum_{(i_1, \dots, i_m) \in I(m, n)} \varphi(Y_{i_1}, \dots, Y_{i_m}) K\left(\frac{d(x_1, X_{i_1})}{h_K}\right) \dots K\left(\frac{d(x_m, X_{i_m})}{h_K}\right)}{\sum_{(i_1, \dots, i_m) \in I(m, n)} K\left(\frac{d(x_1, X_{i_1})}{h_K}\right) \dots K\left(\frac{d(x_m, X_{i_m})}{h_K}\right)}. \quad (5.2.1)$$

As we mentioned, the delta-sequences procedures can be considered a more general class, including kernel estimation techniques. Therefore we can naturally obtain a more general class of functional conditional U-statistics by replacing the kernel $K(\cdot)$ in equation (5.2.1) with positive delta sequences $\delta_m(\cdot, \cdot)$ (see Definition 5.2.0.2), which allows us to introduce the following conditional U-statistic for each $\mathbf{x} = (x_1, \dots, x_k) \in \mathcal{X}^k$ and $\varphi : \mathcal{Y}^k \rightarrow \mathbb{R}$ a measurable function, by

$$\hat{r}_n^{(k)}(\varphi, \mathbf{x}; m_n) = \begin{cases} \frac{\sum_{(i_1, \dots, i_k) \in I(k, n)} \varphi(Y_{i_1}, \dots, Y_{i_k}) \delta_{m_n}(x_1, X_{i_1}) \dots \delta_{m_n}(x_k, X_{i_k})}{\sum_{(i_1, \dots, i_k) \in I(k, n)} \delta_{m_n}(x_1, X_{i_1}) \dots \delta_{m_n}(x_k, X_{i_k})}, \\ \text{if } \sum_{(i_1, \dots, i_k) \in I(k, n)} \delta_{m_n}(x_1, X_{i_1}) \dots \delta_{m_n}(x_k, X_{i_k}) \neq 0, \\ \frac{n!}{(n-k)!k!} \sum_{(i_1, \dots, i_k) \in I(k, n)} \varphi(Y_{i_1}, \dots, Y_{i_k}) \\ \text{if } \sum_{(i_1, \dots, i_k) \in I(k, n)} \delta_{m_n}(x_1, X_{i_1}) \dots \delta_{m_n}(x_k, X_{i_k}) = 0, \end{cases} \quad (5.2.2)$$

which we consider estimating the regression function

$$r^{(k)}(\varphi, \mathbf{x}) = \mathbb{E}(\varphi(Y_1, \dots, Y_k) \mid (X_1, \dots, X_k) = \mathbf{x}), \text{ for } \mathbf{x} \in \mathcal{X}^k, \quad (5.2.3)$$

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

Remark 5.2.0.1 *It is worth noting that X may admit a probability density function $f(\cdot)$ in relation to a σ -finite measure μ on $(\mathcal{X}, \mathcal{B})$ in such a way that (for instance, refer [Dabo-Niang \[2004\]](#), [Prakasa Rao \[2010a\]](#) and [Ouassou and Rachdi \[2012\]](#)):*

$$\mathbb{P}(X \in A) = \int_A f(x) \mu(dx), \text{ for every } A \in \mathcal{B} \text{ such that } 0 < \mu(A) < \infty. \quad (5.2.4)$$

The concept of this remark is elaborated upon in [Prakasa Rao \[2010b\]](#) and its references. We denote by $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space and a nondecreasing family of sub- σ -algebras of \mathcal{F} is denoted by $\{\mathcal{F}_t\}_{t \geq 0}$. Let $\{W(t)\}_{t \geq 0}$ denote a standard Wiener process defined on $(\Omega, \mathcal{F}, \mathbb{P})$ in such a way that W_t is \mathcal{F}_t -measurable. We highlight that the probability measure μ_W on the space $\mathcal{C}^0(0, T)$ is connected with a Borel σ -algebra generated by the supremum norm topology is induced by the standard Wiener process. Let $\{X(t)\}_{0 \leq t \leq T}$ be a diffusion process defined the stochastic differential equation:

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

where $X(0) = x_0$ for $0 \leq t \leq T$. By imposing some assumptions on the functions $a(\cdot, \cdot)$ and $b(\cdot, \cdot)$, we can establish that the probability measure μ_X on the space $\mathcal{C}^0(0, T)$ induced by the process X is absolutely continuous with respect to the probability measure μ_W . In addition, applying Girsanov's Theorem permits the computation of the Radon-Nikodym derivative of μ_X with respect to μ_W . The probability density of X on the space $\mathcal{C}^0(0, T)$ is the μ_W derivative, for instance, see [Prakasa Rao \[1999b\]](#). From this point of view, the main motivation leading to the analysis of functional data is the inference of stochastic processes; the reader is referred to [Prakasa Rao \[1999a,b\]](#). For the purposes of drawing conclusions, we make the assumption that the entirety of the process can be observed. However, if the process can only be seen at discrete times, either on a tiny grid or when the data are sparse, then other approaches, such as parametric inference for discrete data, need to be devised. For instance, for the diffusion processes, these new methods are necessary (cf. [Rachdi and Monsan \[1999\]](#), [Prakasa Rao \[1983, 1999b\]](#)). Observe that if $A = B(x, \kappa)$ for $(x, \kappa) \in \mathcal{X} \times \mathbb{R}_+^*$, then (5.2.4) allows for the small ball probability to be considered.

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

$$S_{\mathcal{X}} \subset \mathcal{S}_n := \bigcup_{\ell=1}^{d_n} B(t_\ell, \epsilon), \quad (5.2.5)$$

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

It is worth mentioning that the hypothesis (5.2.5) is essential for assuming a geometrical link between the number d_n of balls and their radius ϵ . In addition, this condition is fulfilled in usual nonparametric problems when $\mathcal{X} = \mathbb{R}^p$ is endowed with the Euclidean metric on \mathbb{R}^p (because $\kappa = p$ suffices). However, this topological characteristic does not hold for any abstract semi-metric space, as Ferraty and Vieu [2008] explains. Before we can use the delta-sequences approach to estimate the value of the regression operator $r^{(k)}(\cdot)$ in the model (5.2.3), we must first have the following definition.

Definition 5.2.0.2 A sequence of non-negative functions $\{\delta_{m_n}(x, y) : m_n \geq 1\} = \{\delta_m(x, y), m \geq 1\}$ defined on $\mathcal{X}^k \times \mathcal{X}^k$ is called a delta-sequence with respect to the measure μ if the following properties are satisfied:

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

$$\lim_{m \rightarrow +\infty} \sup_{\mathbf{x} \in \mathcal{S}_{\mathcal{X}}^k} \left| \int_{\mathbb{B}(\mathbf{x}, \gamma)} \delta_m(\mathbf{x}, \mathbf{y}) \mu(d\mathbf{y}) - 1 \right| = 0, \quad (5.2.6)$$

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

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

$$\sup_{(\mathbf{x}, \mathbf{y}) \in \mathcal{S}_{\mathcal{X}}^k \times \mathcal{X}^k} \delta_m(\mathbf{x}, \mathbf{y}) \leq C_1 s_m < \infty, \quad (5.2.7)$$

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

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

$$|\delta_m(\mathbf{x}_1, \mathbf{y}) - \delta_m(\mathbf{x}_2, \mathbf{y})| \leq C s_m^{\beta_2} d(\mathbf{x}_1, \mathbf{x}_2)^{\beta_1} \quad \text{for all } \mathbf{x}_1, \mathbf{x}_2, \mathbf{y} \in \mathcal{X}^k, \quad (5.2.8)$$

where

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

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

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

$$\lim_{m \rightarrow \infty} \sup_{\substack{\mathbf{x} \in \mathcal{S}_{\mathcal{X}}^k \\ \mathbf{y} \in \bar{\mathbb{B}}(\mathbf{x}, \gamma)}} \delta_m(\mathbf{x}, \mathbf{y}) d(\mathbf{x}, \mathbf{y}) = 0, \quad (5.2.9)$$

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

Notice that the conditions (C.1)-(C.4) of Definition 5.2.0.2 are modelled after a similar set of conditions for kernel-type estimators. The condition (C.2) corresponds to the bound of δ_m over $\mathcal{S}_{\mathcal{X}}^k \times \mathcal{X}^k$ whereas the condition (C.3) is pertains to the uniform Lipschitz property of $\delta_m(\mathbf{x}, \mathbf{y})$. Contrarily, the condition (C.4) is and it is not an assumption on the bound of $d(\mathbf{x}, \mathbf{y})$ over $\mathcal{S}_{\mathcal{X}}^k \times \mathcal{S}_{\mathcal{X}}^k$ but an assumption on the limiting behaviour of $\delta_m(\mathbf{x}, \mathbf{y}) d(\mathbf{x}, \mathbf{y})$ as $m \rightarrow +\infty$.

Proposition 5.2.0.3 *Let $\{\delta_{m,1}(x_1, y_1)\}, \dots, \{\delta_{m,k}(x_k, y_k)\}$ each be non-negative delta-sequence with respect to the measure μ , then*

$$\delta_m(\mathbf{x}, \mathbf{y}) := \prod_{j=1}^k \delta_{m,j}(x_j, y_j), \quad (5.2.10)$$

is also a non-negative delta sequence.

This proposition is similar to Proposition 2.2 [Susarla and Walter \[1981\]](#) when $\mathcal{X} = \mathbb{R}^d$, which means that the product of non-negative delta-sequences is also a positive delta sequence. The proposition provides a flexible way to construct delta sequences in high dimensions in a similar way to kernel type estimation. Unless otherwise specified we will set

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

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

5.2.1 Examples of delta sequence

In this section, following the notation of [Ouassou and Rachdi \[2012\]](#), we provide guidelines for constructing and recovering some well-known estimators in literature.

Example 5.2.1.1 Kernel estimator

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

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

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

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

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

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

then, by Proposition 5.2.0.3, $\delta_m(\mathbf{x}, \mathbf{y})$ is a non-negative delta sequence, and the conditional

U-statistic is defined in this case by

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

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

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

this quantity tends to zero when m is sufficiently large.

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

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

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

Example 5.2.1.2 Histogram estimator

Let $\mathcal{P}_n = \{A_{n,j}, j \in J_n\}$ be a partition of the set F (cf. [Maouloud \[2007\]](#)), such that

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

Denote

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

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

Example 5.2.1.3 *Orthogonal series estimator*

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

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

As stated in [Walter \[1965\]](#), $\delta_m(\cdot, \cdot)$ in this case are delta sequences. Now using [Proposition 5.2.0.3](#), we can see that

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

are also a positive delta sequences.

For more examples of delta sequences we refer to [Walter and Blum \[1979\]](#) and [Susarla and Walter \[1981\]](#).

5.2.2 Conditions and comments

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

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

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

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

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

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

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

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

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

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

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

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

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

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

5.2.3 Comments on the assumptions

Similar to conditions (C.1)-(C.4), assumption (C.5) is also modelled after some kernel-type conditions, and it allows us to select β_1 and β_2 in condition (C.3). Due to the infinite nature of the problem, additional constraints are required to achieve uniform consistency across the pseudo-compact set. Ferraty and Vieu [2008], discussed the assumption (5.2.5). This condition holds trivially for any finite-dimensional Euclidean space and remains valid for projection-based metric spaces with infinite dimensions. Condition (C.7) concerning the boundedness of the function $\varphi(\cdot)$ is essential to establish exponential bounds, this, coupled with the technical condition (C.6), allows us to obtain the almost complete convergence later in the proofs. Note that we can replace condition (C.7) with a more general one, that is, condition (C.7)', to obtain the results when the function $\varphi(\cdot)$ is unbounded. Finally, to establish precise rates of almost complete convergence in the functional context, additional conditions related to the topological nature of the problem are required. Mainly the assumption (C.8) and (C.9), where the latter condition concerning the Lipschitz property of the operator $r^{(k)}(\cdot)$ is standard when studying with uniform consistency.

Remark 5.2.3.1 *Note that the condition (C.7) can be replaced by more broad hypotheses at specific times of \mathbf{Y} , as shown in Deheuvels [2011]. That is*

(C.7)' *We denote by $\{\mathcal{M}(x) : x \geq 0\}$ a nonnegative continuous function, increasing on $[0, \infty)$, and such that, for some $s > 2$, ultimately as $x \uparrow \infty$,*

$$(i) \ x^{-s} \mathcal{M}(x) \downarrow; (ii) \ x^{-1} \mathcal{M}(x) \uparrow. \quad (5.2.11)$$

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

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

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

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

The boundedness assumption on $\varphi(\cdot)$ can be substituted by a finite moment assumption (C.7)', but doing so will add a significant amount of additional complexity to the proofs; for further information, check the most recent reference Bouzebda and Taachouche [2022], Chokri and Bouzebda [2022], Bouzebda and Nezzal [2022], Bouzebda and Nemouchi [2020, 2022] for more details.

5.3 Some asymptotic results

In this subsection, we will discuss the uniform consistency of the functional conditional U-statistic, which is defined by (5.2.2). First, let us provide basic notation

$$\begin{aligned}\mathbf{X} &:= (X_1, \dots, X_k) \in \mathcal{X}^k, & \mathbf{Y} &:= (Y_1, \dots, Y_k) \in \mathcal{Y}^k, \\ \mathbf{X}_i &:= (X_{i_1}, \dots, X_{i_k}), & \mathbf{Y}_i &:= (Y_{i_1}, \dots, Y_{i_k}), \\ G_{\varphi, \mathbf{x}}(\mathbf{X}, \mathbf{Y}) &:= \varphi(\mathbf{Y}) \delta_m(\mathbf{x}, \mathbf{X}) \quad \text{for } \mathbf{x} \in S_{\mathcal{X}}^k,\end{aligned}$$

$$u_n(\varphi, \mathbf{x}, m_n) = u_n^{(k)}(G_{\varphi, \mathbf{x}}) := \frac{(n-k)!}{n!} \sum_{i \in I(k, n)} G_{\varphi, \mathbf{x}}(\mathbf{X}_i, \mathbf{Y}_i).$$

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

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

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

$$\hat{\mathbb{E}}(\hat{r}_n^{(k)}(\varphi, \mathbf{x}; m_n)) = \frac{\mathbb{E}(u_n(\varphi, \mathbf{x}, m_n))}{\mathbb{E}(u_n(1, \mathbf{x}, m_n))}. \quad (5.3.1)$$

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

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

Suppose that L is a function of $\ell \geq 1$ variables, symmetric in its entries. Then, the Hoeffding projections (see [Hoeffding \[1948\]](#) and [de la Peña and Giné \[1999\]](#)) with respect to \mathbb{P} , for $1 \leq k \leq \ell$, are defined as

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

and

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

for some measures Q_i on S we denote

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

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

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

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

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

For example,

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

Remark 5.3.0.1 *The function $G_{\varphi, \mathbf{x}, m_n}$ is not necessarily symmetric; when we need to symmetrize them, we have :*

$$\bar{G}_{\varphi, \mathbf{x}}(\mathbf{x}, \mathbf{y}) := \frac{1}{k!} \sum_{\sigma \in \mathcal{I}_k^k} G_{\varphi, \mathbf{x}, m_n}(\mathbf{x}_\sigma, \mathbf{y}_\sigma) = \frac{1}{k!} \sum_{\sigma \in \mathcal{I}_k^k} \varphi(\mathbf{y}_\sigma) \delta_{m_n}(\mathbf{x}_\sigma, \mathbf{y}_\sigma),$$

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

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

and the U-statistic

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

do not change.

5.3.1 Uniform consistency of functional conditional U-statistics

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

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

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

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

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

Theorem 5.3.1.1 *Under the conditions (C.1)-(C.4), and if conditions (C.5), (C.7) are satisfied, then we have:*

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

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

Theorem 5.3.1.2 *Under the conditions (C.1)-(C.4), and if conditions (C.5), (C.6) and (C.7') are satisfied, then we have:*

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

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

Theorem 5.3.1.3 *Under the conditions (C.1)-(C.4), and if conditions (C.5), (C.6) and condition (C.7) (or (C.7')) are satisfied, then we have:*

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

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

Theorem 5.3.1.4 *Under the conditions (C.1)-(C.4) and (C.9), we have:*

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

The following corollary is more or less straightforward, given Theorem 5.3.1.3 and Theorem 5.3.1.4.

Corollary 5.3.1.5 *Under the conditions of Theorem 5.3.1.3 and Theorem 5.3.1.4 it follows that, as m tends to infinity:*

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

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

5.3.2 Uniform strong consistency with rates

This section is devoted to the uniform version with the rate of Theorem 5.3.1.5's result. More specifically, our objective is to obtain the uniform almost-complete convergence of $\hat{r}_n^{(k)}(\cdot)$ on some subset $S_{\mathcal{X}}^k$ of \mathcal{X}^k satisfying condition (5.2.5). In the following theorem, we establish the bias order.

Theorem 5.3.2.1 *Under the conditions (C.1)-(C.4), and if conditions (C.8) and (C.9) are satisfied, then we have:*

$$\sup_{\mathbf{x} \in \mathcal{S}_{\mathcal{X}}^k} \left| \widehat{\mathbb{E}} \left(\widehat{r}_n^{(k)}(\varphi, \mathbf{x}; m_n) \right) - r^{(k)}(\varphi, \mathbf{x}) \right| = O(D_m). \quad (5.3.2)$$

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

Corollary 5.3.2.2 *Under the conditions of Theorems 5.3.1.3 and 5.3.2.1 it follows that :*

$$\sup_{\mathbf{x} \in \mathcal{S}_{\mathcal{X}}^k} \left| \widehat{r}_n^{(k)}(\varphi, \mathbf{x}; m_n) - r^{(k)}(\varphi, \mathbf{x}) \right| = O(D_m) + O_{a.co} \left(\sqrt{\frac{s_m \log(m)}{m}} \right). \quad (5.3.3)$$

5.4 Conditional U-statistics for censored data

Consider a triple (Y, C, \mathbf{X}) of random variables defined in $\mathbb{R} \times \mathbb{R} \times \mathcal{X}$. 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

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

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

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

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 \mathcal{F} of real measurable functions defined on \mathbb{R} , and assume that \mathcal{F} is of VC-type. We recall the regression function of $\psi(Y)$ evaluated at $\mathbf{X} = \mathbf{x}$, for $\psi \in \mathcal{F}$ and $\mathbf{x} \in \mathcal{I}_\alpha$, given by

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

when Y is right-censored. To estimate $r^{(1)}(\psi, \cdot)$, we make use of the Inverse Probability of Censoring Weighted (I.P.C.W.) estimators that have recently gained popularity in the censored data literature (see Kohler *et al.* [2002], Carbonez *et al.* [1995]). The key idea of I.P.C.W. estimators is as follows. Introduce the real-valued function $\Phi_\psi(\cdot, \cdot)$ defined on \mathbb{R}^2 by

$$\Phi_\psi(y, c) = \frac{\mathbb{1}\{y \leq c\}\psi(y \wedge c)}{1 - G(y \wedge c)}. \quad (5.4.1)$$

Assuming the function $G(\cdot)$ to be known, first note that

$$\Phi_\psi(Y_i, C_i) = \Delta_i \psi(Z_i) / (1 - G(Z_i))$$

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

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

We have

$$\begin{aligned} r^{(1)}(\Phi_\psi, \mathbf{x}) &:= \mathbb{E}(\Phi_\psi(Y, C) | \mathbf{X} = \mathbf{x}) \\ &= \mathbb{E}\left\{ \frac{\mathbb{1}\{Y \leq C\}\psi(Z)}{1 - G(Z)} \mid \mathbf{X} = \mathbf{x} \right\} \\ &= \mathbb{E}\left\{ \frac{\psi(Y)}{1 - G(Y)} \mathbb{E}(\mathbb{1}\{Y \leq C\} | \mathbf{X}, Y) \mid \mathbf{X} = \mathbf{x} \right\} \\ &= r^{(1)}(\psi, \mathbf{x}). \end{aligned} \quad (5.4.2)$$

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

$$\bar{\omega}_{n,K,h,i}^{(1)}(\mathbf{x}) := \delta_{m_n}(\mathbf{x}, \mathbf{X}_i) / \sum_{j=1}^n \delta_{m_n}(\mathbf{x}, \mathbf{X}_j). \quad (5.4.3)$$

Making use of the equations (5.4.1), (5.4.2), and (5.4.3), whenever $G(\cdot)$ is known, we define the kernel estimator of $r^{(1)}(\psi, \cdot)$ by

$$\check{r}_n^{(1)}(\psi, \mathbf{x}; h_n) = \sum_{i=1}^n \bar{\omega}_{n,K,h,i}^{(1)}(\mathbf{x}) \frac{\Delta_i \psi(Z_i)}{1 - G(Z_i)}. \quad (5.4.4)$$

Since the function $G(\cdot)$ is unknown, it is to be estimated. Let $G_n^*(\cdot)$ denote the Kaplan-Meier estimator of the function $G(\cdot)$ [Kaplan and Meier \[1958\]](#). To be precise, adopting the conventions

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

and $0^0 = 1$ and setting

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

we have

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

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

$$\check{r}_n^{(1)*}(\psi, \mathbf{x}; h_n) = \sum_{i=1}^n \bar{\omega}_{n,K,h,i}^{(1)}(\mathbf{x}) \frac{\Delta_i \Psi(Z_i)}{1 - G_n^*(Z_i)}, \quad (5.4.5)$$

the reader is invited to see the papers of [Kohler *et al.* \[2002\]](#) and [Maillot and Viallon \[2009\]](#). The convention $0/0 = 0$ is used, this quantity is well defined, since $G_n^*(Z_i) = 1$ if and only if $Z_i = Z_{(n)}$ and $\Delta_{(n)} = 0$, where $Z_{(k)}$ is the k th ordered statistic related with the sample (Z_1, \dots, Z_n) for $k = 1, \dots, n$ and $\Delta_{(k)}$ is the Δ_j corresponding to $Z_k = Z_j$. When the variable of interest is right-censored, it is often impossible to estimate the function of the (conditional) law on the whole support (see [Brunel and Comte \[2006\]](#)). [Datta *et al.* \[2010\]](#) introduces a right-censored version of an unconditional U-statistic with a kernel of degree $m \geq 1$ based on the notion of a mean-preserving reweighting technique. [Stute and Wang \[1993\]](#) have demonstrated the almost sure convergence of multi-sample U-statistics under random censorship and presented an application by analyzing the consistency of a novel class of tests meant to evaluate distribution equality. [Chen and Datta \[2019\]](#) presented improvements to the traditional U-statistics to counteract potential biases caused by right-censoring of the outcomes and the existence of confounding factors. [Yuan *et al.* \[2017\]](#) suggested an alternative method for estimating the U-statistic by employing a substitution estimator of the conditional kernel given observed data. We also refer to [Bouzebda and El-hadjali \[2020\]](#), [Bouzebda *et al.* \[2022b\]](#) and [Bouzebda and Nezzal \[2022\]](#). To our best knowledge, estimating the conditional U-statistics in the censored data setting is a current open problem, and it gives the main motivation for the study of this section.

The function described by (5.4.1) has a natural expansion given by

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

We have an analogous relationship to (5.4.2) based on the formula:

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

An analogue estimator to (5.2.2) in the censored situation is given by

$$\check{r}_n^{(k)}(\psi, \mathbf{x}; m_n) = \sum_{(i_1, \dots, i_k) \in I(k, n)} \frac{\Delta_{i_1} \cdots \Delta_{i_k} \Psi(Z_{i_1}, \dots, Z_{i_k})}{(1 - G(Z_{i_1})) \cdots (1 - G(Z_{i_k}))} \bar{\omega}_{n, \delta, m_n, \mathbf{i}}^{(k)}(\mathbf{x}), \quad (5.4.6)$$

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

$$\bar{\omega}_{n, \delta, m_n, \mathbf{i}}^{(k)}(\mathbf{x}) = \frac{\delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k})}{\sum_{(i_1, \dots, i_k) \in I(k, n)} \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k})}. \quad (5.4.7)$$

The estimator we shall examine is provided by

$$\check{r}_n^{(k)*}(\psi, \mathbf{x}; m_n) = \sum_{(i_1, \dots, i_k) \in I(k, n)} \frac{\Delta_{i_1} \cdots \Delta_{i_k} \Psi(Z_{i_1}, \dots, Z_{i_k})}{(1 - G_n^*(Z_{i_1})) \cdots (1 - G_n^*(Z_{i_k}))} \bar{\omega}_{n, \delta, m_n, \mathbf{i}}^{(k)}(\mathbf{x}). \quad (5.4.8)$$

In a similar way as in Bouzebda *et al.* [2022b], we arrive to the following conclusion.

Corollary 5.4.0.1 *Assume that the condition (I) and the assumptions of Theorems 5.3.1.3 and 5.3.2.1 are satisfied. Then, we have*

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

This last result is a direct consequence of Corollary (5.3.2.2) and the law of iterated logarithm for $G_n^*(\cdot)$ obtained in Földes and Rejtő [1981] gives

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

At this point, we may refer to Bouzebda and El-hadjali [2020], Bouzebda *et al.* [2022b] and Bouzebda and Nezzal [2022].

5.5 Applications

5.5.1 Kendall rank correlation coefficient

To test the independence of one-dimensional random variables Y_1 and Y_2 Kendall [1938] proposed a method based on the U-statistic K_n with the kernel function :

$$\varphi((s_1, t_1), (s_2, t_2)) = \mathbb{1}_{\{(s_2 - s_1)(t_2 - t_1) > 0\}} - \mathbb{1}_{\{(s_2 - s_1)(t_2 - t_1) \leq 0\}}. \quad (5.5.1)$$

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

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

where $\mathbf{t} = (t_1, t_2) \in \mathbb{I} \subset \mathbb{R}^2$ and $\varphi(\cdot)$ is Kendall's kernel (5.5.1). Suppose that $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are d_1 and d_2 -dimensional random vectors respectively and $d_1 + d_2 = d$. Furthermore, suppose that Y_1, \dots, Y_n are observations of $(\boldsymbol{\xi}, \boldsymbol{\eta})$, we are interested in testing :

$$H_0 : \boldsymbol{\xi} \text{ and } \boldsymbol{\eta} \text{ are conditionally independent given } X. \text{ vs } H_a : H_0 \text{ is not true.} \quad (5.5.2)$$

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

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

An application of Corollary 5.3.2.2 gives

$$\sup_{\mathbf{x} \in \mathcal{S}_{\mathcal{X}}^2} |\widehat{r}_n^{(2)}(\varphi^a, \mathbf{x}; m_n) - r^{(2)}(\varphi^a, \mathbf{x})| = O(D_m) + O_{a.co} \left(\sqrt{\frac{s_m \log(m)}{m}} \right). \quad (5.5.3)$$

5.5.2 Discrimination problems

Now, we apply these findings to the discrimination problem outlined in Section 3 of [Stute \[1994b\]](#), refer to also to [Stute \[1994a\]](#). We will employ a similar setup and notation. Let $\varphi(\cdot)$ be any function taking at most finitely many values, say $1, \dots, M$. The sets

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

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

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

where

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

The inequality described above becomes equality if

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

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

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

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

$$\mathfrak{M}_n^j(\tilde{\mathbf{x}}) = \frac{\sum_{(i_1, \dots, i_k) \in \mathcal{I}(k, n)} \mathbb{1}\{\varphi(Y_{i_1}, \dots, Y_{i_k}) = j\} \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k})}{\sum_{(i_1, \dots, i_k) \in \mathcal{I}(k, n)} \delta_{m_n}(\mathbf{x}_1, \mathbf{X}_{i_1}) \cdots \delta_{m_n}(\mathbf{x}_k, \mathbf{X}_{i_k})}, \quad (5.5.4)$$

Set

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

Let us define

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

The discrimination rule $g_{0,n}(\cdot)$ is asymptotically Bayes' risk consistent

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

This is a consequence of the relation

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

5.5.3 Metric learning

Metric learning seeks to adapt the metric to the data and has garnered a great deal of attention in recent years; for instance, see [Bellet *et al.* \[2013\]](#) and [Cl  men  on *et al.* \[2016\]](#) for an account of metric learning and its applications. This is driven by a wide variety of applications, spanning from information retrieval via bioinformatics to computer vision as the primary source of inspiration. For the purpose of demonstrating the applicability of this idea, we will now discuss the metric learning problem for supervised classification as shown in [Cl  men  on *et al.* \[2016\]](#). Let us consider independent copies $(X_1, Y_1), \dots, (X_n, Y_n)$ of a $\mathcal{X} \times \mathcal{Y}$ valued random couple (X, Y) , where \mathcal{X} is some feature space and $\mathcal{Y} = \{1, \dots, C\}$, with $C \geq 2$ say, a finite set of labels. Let \mathcal{D} be a set distance measures $D: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_+$. The purpose of metric learning in this context is, intuitively speaking, to identify a metric under which pairs of points with the same label are close to each other, while those with different labels are far away from each other. A natural way to characterize the risk associated with a metric D is as follows

$$R(D) = \mathbb{E} [\phi((1 - D(X, X')) \cdot (2\mathbb{1}\{Y = Y'\} - 1))], \quad (5.5.5)$$

where $\phi(u)$ is a convex loss function upper bounding the indicator function $\mathbb{1}\{u \geq 0\}$, for instance, the hinge loss $\phi(u) = \max(0, 1 - u)$. To estimate $R(D)$, we consider the usual empirical estimator

$$R_n(D) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \phi((D(X_i, X_j) - 1) \cdot (2\mathbb{1}\{Y_i = Y_j\} - 1)), \quad (5.5.6)$$

which is one sample U-statistic of degree two with kernel given by:

$$\varphi_D((x, y), (x', y')) = \phi((D(x, x') - 1) \cdot (2\mathbf{1}\{y = y'\} - 1)).$$

The convergence to (5.5.5) of a minimizer of (5.5.6) has been studied in the frameworks of algorithmic stability (Jin *et al.* [2009]), algorithmic robustness (Bellet and Habrard [2015]) and based on the theory of U-processes under appropriate regularization (Cao *et al.* [2016]).

5.5.4 Time series prediction from continuous set of past values

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

$$X_i = \{Z(t), (i-1)T \leq t < iT\}.$$

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

$$\varphi(Z_1(\tau + s), \dots, Z_k(\tau + s)) = r^{(k)}(Z_1(t), \dots, Z_k(t)), \text{ for } \tau - T \leq t < \tau. \quad (5.5.7)$$

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

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

where $\mathbf{z} = (z_1, \dots, z_k) = \{(Z_1(t), \dots, Z_k(t)), \text{ for } \tau - T \leq t < \tau\}$. Corollary 5.3.2.2 provides mathematical support for this nonparametric functional predictor and extends previous results in numerous ways in Ferraty and Vieu [2006a] and Ouassou and Rachdi [2012]. Notice that this modelization encompasses a wide variety of practical applications, as this procedure allows for the consideration of a large number of past process values without being affected by the curse of dimensionality.

5.5.5 Example of U-kernels

Example 5.5.5.1 *Hoeffding's D From the symmetric kernel,*

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

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

Example 5.5.5.2 (Blum-Kiefer-Rosenblatt's R) The symmetric kernel

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

gives Blum-Kiefer-Rosenblatt's R statistic (see [Blum et al. \[1961\]](#)), which is a rank-based U-statistic of order 6, refer also to [Bouzebda \[2011, 2014, 2023\]](#); [Bouzebda and Zari \[2013\]](#); [Bouzebda et al. \[2011\]](#).

Example 5.5.5.3 Bergsma-Dassios-Yanagimoto's τ^* [Bergsma and Dassios \[2014\]](#) introduced a rank correlation statistic as a U-statistic of order 4 with the symmetric kernel

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

Here

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

Example 5.5.5.4 Two generic vectors $y = (y_1, y_2)$ and $z = (z_1, z_2)$ in \mathbb{R}^2 are said to be concordant if $(y_1 - z_1)(y_2 - z_2) > 0$. For $m, k = 1, \dots, p$, define

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

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

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

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

Example 5.5.5.5 The Gini mean difference. The Gini index provides another usual measure of dispersion. It corresponds to the case where $E \subset \mathbb{R}$ and $h(x, y) = |x - y|$:

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

5.6 Mathematical development

This section contains the proof of our results. The preceding notation is also used in the subsequent text. Keeping in mind the relation (5.2.5), we can conclude that, for each $\mathbf{x} = (x_1, \dots, x_k) \in S_{\mathcal{X}}^k$, there exists $\boldsymbol{\ell}(\mathbf{x}) = (\ell(x_1), \dots, \ell(x_k))$ where $\forall 1 \leq i \leq k, 1 \leq \ell(x_i) \leq d_n$ and such that

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

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

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

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

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

Proof of Theorem 5.3.1.1

We need to establish that there exists some $\eta > 0$, in such a way that

$$\sum_{n \geq 1} \mathbb{P} \left\{ \sup_{\mathbf{x} \in S_{\mathcal{X}}^k} \sqrt{\frac{m}{s_m \log(m)}} |u_n(\varphi, \mathbf{x}, m_n) - \mathbb{E}(u_n(\varphi, \mathbf{x}, m_n))| \geq \eta \right\} < \infty. \quad (5.6.1)$$

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

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

We first remark that we have

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

where

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

In order to get the desired result, we apply Lemma 5.6.0.1 on the function $H(\cdot, \cdot)$. Throughout the rest of the proof, we suppose the function $G_{\varphi, \mathbf{x}}$ is symmetric. Moreover, it is clear that the function $H(\cdot, \cdot)$ is bounded by $2MC_1 s_m$ by condition (C.2) and the fact that the function $\varphi(\cdot)$ is bounded by the condition (C.7). We obviously remark that,

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

by design, and

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

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

$$\begin{aligned} & \mathbb{P} \left(\left| u_n(\varphi, \mathbf{x}, m_n) - \mathbb{E}[u_n(\varphi, \mathbf{x}, m_n)] \right| > \eta \sqrt{\frac{s_m \log(m)}{m}} \right) \\ & \leq 2 \exp \left[-\frac{n((s_m \log(m))/m)\eta^2}{4(MC_1 s_m)^2 + \frac{4}{3}MC_1 s_m \eta \sqrt{(s_m \log(m))/m}} \right]. \end{aligned} \quad (5.6.2)$$

We can write

$$\begin{aligned} & \mathbb{P} \left(\sup_{\mathbf{x} \in S_{\mathcal{X}}^k} |u_n(\varphi, \mathbf{x}; m_n) - \mathbb{E}[u_n(\varphi, \mathbf{x}; m_n)]| > 2\eta \sqrt{\frac{s_m \log(m)}{m}} \right) \\ & \leq \mathbb{P} \left(\sup_{\mathbf{x} \in S_{\mathcal{X}}^k} |u_n(\varphi, \mathbf{x}; m_n) - u_n(\varphi, \mathbf{x}_{\ell(\mathbf{x})}; m_n) \right. \\ & \quad \left. + \mathbb{E}[u_n(\varphi, \mathbf{x}_{\ell(\mathbf{x})}; m_n)] - \mathbb{E}[u_n(\varphi, \mathbf{x}; m_n)] \right| > \eta \sqrt{\frac{s_m \log(m)}{m}} \right) \end{aligned} \quad (5.6.3)$$

$$+ \mathbb{P} \left(\sup_{\mathbf{x} \in S_{\mathcal{X}}^k} |u_n(\varphi, \mathbf{x}_{\ell(\mathbf{x})}; m_n) - \mathbb{E}[u_n(\varphi, \mathbf{x}_{\ell(\mathbf{x})}; m_n)]| > \eta \sqrt{\frac{s_m \log(m)}{m}} \right). \quad (5.6.4)$$

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

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

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

$$\sup_{\mathbf{x} \in \mathcal{S}_{\mathcal{X}}^k} |u_n(\varphi, \mathbf{x}; m_n) - u_n(\varphi, \mathbf{x}_{\ell(\mathbf{x})}; m_n)| \leq O(s_m^{\beta_2} \epsilon_n^{\beta_1}) = O\left(\sqrt{\frac{s_m \log(m)}{m}}\right), \quad (5.6.5)$$

by condition (C.5). We deduce from (5.6.5) that:

$$\begin{aligned} & |\mathbb{E}[u_n(\varphi, \mathbf{x}_{\ell(\mathbf{x})}; m_n)] - \mathbb{E}[u_n(\varphi, \mathbf{x}; m_n)]| \\ &= |\mathbb{E}[u_n(\varphi, \mathbf{x}_{\ell(\mathbf{x})}; m_n) - u_n(\varphi, \mathbf{x}; m_n)]| \\ &\leq \mathbb{E}[|u_n(\varphi, \mathbf{x}_{\ell(\mathbf{x})}; m_n) - u_n(\varphi, \mathbf{x}; m_n)|]. \end{aligned} \quad (5.6.6)$$

The passage from (5.6.6) to (5.6.6) follows by applying Jensen's inequality further to some properties of the absolute value function. Now using the fact that the function $\varphi(\cdot)$ is bounded and that the function δ_m is Lipschitz in addition for any constant a , $\mathbb{E}[a] = a$, we can directly conclude that

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

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

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

Continue, now, with (5.6.4), supposing that the kernel function $G_{\varphi, \mathbf{x}_{\ell}}(\cdot)$ is symmetric, we have to decompose the U-statistic by making use of the **Hoeffding [1948]** decomposition, we infer that

$$\begin{aligned} & u_n(\varphi, \mathbf{x}_{\ell}; m_n) - \mathbb{E}[u_n(\varphi, \mathbf{x}_{\ell}; m_n)] \\ &= \sum_{p=1}^k \frac{k!}{(k-p)!} u_n^{(p)}(\pi_{p,k}(G_{\varphi, \mathbf{x}_{\ell}}, m_n)) \\ &= k u_n^{(1)}(\pi_{1,k}(G_{\varphi, \mathbf{x}_{\ell}})) + \sum_{p=2}^k \frac{k!}{(k-p)!} u_n^{(p)}(\pi_{p,k}(G_{\varphi, \mathbf{x}_{\ell}})). \end{aligned} \quad (5.6.7)$$

Let us first start with the linear term. We have

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

From Hoeffding's projection, we have

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

Set

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

We can see that Z_i are independent and identically distributed random variables bounded by $2kMC_1 s_m$ with zero mean and

$$\sigma^2 \leq (MC_1 s_m)^2.$$

An application of Bernstein's inequality yields

$$\begin{aligned} & \mathbb{P} \left(\sup_{\mathbf{x} \in S_{\mathcal{X}}^k} |u_n^{(1)}(\pi_{1,k}(G_{\varphi, \mathbf{x}_\ell}))| > \eta \sqrt{\frac{s_m \log(m)}{m}} \right) \\ & \leq \sum_{i=1}^{d_n} \mathbb{P} \left(\max_{1 \leq \ell_i \leq d_n} |u_n^{(1)}(\pi_{1,k}(G_{\varphi, \mathbf{x}_\ell}))| > \eta \sqrt{\frac{s_m \log(m)}{m}} \right) \\ & \leq 2d_n \exp \left[-\frac{n((s_m \log(m))/m)\eta^2}{4(MC_1 s_m)^2 + \frac{4}{3}MC_1 s_m \eta \sqrt{(s_m \log(m))/m}} \right] \\ & \leq n^{\alpha - \tau \eta^2 / C_4}, \end{aligned}$$

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

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

Consequently, we get the following:

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

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

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

which implies that, for $1 \leq i \leq k$ and $\ell = (\ell_1, \dots, \ell_k)$:

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

To prove the above-mentioned equation, we need to apply Proposition 1 of Arcones [1995] (see Lemma 5.6.0.2). We can see that $G_{\varphi, \mathbf{x}_\ell}$ is bounded by $MC_1 s_m$, hence for $\eta > 0$ we have

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

where $\varepsilon_0 = \frac{\eta}{\sqrt{n}}$. Now for $t = \eta \sqrt{\frac{s_m \log(m)}{m}}$, Lemma 5.6.0.2 gives us :

$$\begin{aligned} & \mathbb{P} \left(\left| \sum_{p=2}^k \frac{k!}{(k-p)!} u_n^{(p)}(\pi_{p,k}(G_{\varphi, \mathbf{x}_\ell})) \right| > \varepsilon_0 \sqrt{\frac{s_m \log(m)}{m}} \right) \\ & \leq 2 \exp \left(-\frac{t(n-1)^{1/2}}{2^{k+2} k^{k+1} MC_1 s_m} \right) \\ & \leq 2 \exp \left(-\frac{\eta \sqrt{s_m \log(m)/m} (n-1)^{1/2}}{2^{k+2} k^{k+1} MC_1 s_m} \right) \\ & \leq 2 \exp \left(-\frac{\eta \sqrt{\log(m)/m} (n-1)^{1/2}}{2^{k+2} k^{k+1} MC_1 \sqrt{s_m}} \right). \end{aligned}$$

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

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

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

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

Consequently, we have

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

Hence the proof is achieved. □

Proof of Theorem 5.3.1.2

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

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

we have

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

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

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

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

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

The first term of the right side $u_n^{(T)}(\varphi, \mathbf{x}, m_n)$, as usual, is called the truncated part and the second one $u_n^{(R)}(\varphi, \mathbf{x}, m_n)$ is the remainder part. Let's investigate the term $u_n^{(T)}(\varphi, \mathbf{x}, m_n)$.

Truncated Part :

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

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

where

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

Similar to the proof of Theorem 5.3.1.1, we apply Lemma 5.6.0.1 on the function $H^{(\text{T})}(\cdot, \cdot)$. Throughout the rest of the proof, we suppose that the function $G_{\varphi, \mathbf{x}}^{(\text{T})}$ is symmetric. Moreover, it is clear that the function $H^{(\text{T})}(\cdot, \cdot)$ is bounded by $2\lambda\xi_n^{1/q}C_1s_m$ by condition (C.2). We obviously remark that,

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

by design, and

$$\sigma^2 = \text{Var}(H^{(\text{T})}(\mathbf{X}, \mathbf{Y})) \leq 2(\lambda\xi_n^{1/q}C_1s_m)^2.$$

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

$$\begin{aligned} & \mathbb{P} \left(\left| u_n^{(\text{T})}(\varphi, \mathbf{x}, m_n) - \mathbb{E} \left(u_n^{(\text{T})}(\varphi, \mathbf{x}, m_n) \right) \right| > \eta \sqrt{\frac{s_m \log(m)}{m}} \right) \\ & \leq 2 \exp \left[- \frac{n((s_m \log(m))/m)\eta^2}{4(\lambda\xi_n^{1/q}C_1s_m)^2 + \frac{4}{3}\lambda\xi_n^{1/q}C_1s_m\eta\sqrt{(s_m \log(m))/m}} \right]. \end{aligned}$$

We can write

$$\begin{aligned} & \mathbb{P} \left(\sup_{\mathbf{x} \in \mathcal{S}_{\mathcal{X}}^k} \left| u_n^{(\text{T})}(\varphi, \mathbf{x}, m_n) - \mathbb{E} \left(u_n^{(\text{T})}(\varphi, \mathbf{x}, m_n) \right) \right| > 2\eta \sqrt{\frac{s_m \log(m)}{m}} \right) \\ & \leq \mathbb{P} \left(\sup_{\mathbf{x} \in \mathcal{S}_{\mathcal{X}}^k} \left| u_n^{(\text{T})}(\varphi, \mathbf{x}; m_n) - u_n^{(\text{T})}(\varphi, \mathbf{x}_{\ell(\mathbf{x})}; m_n) \right. \right. \\ & \quad \left. \left. + \mathbb{E} \left[u_n^{(\text{T})}(\varphi, \mathbf{x}_{\ell(\mathbf{x})}; m_n) \right] - \mathbb{E} \left[u_n^{(\text{T})}(\varphi, \mathbf{x}; m_n) \right] \right| > \eta \sqrt{\frac{s_m \log(m)}{m}} \right) \\ & \quad + \mathbb{P} \left(\sup_{\mathbf{x} \in \mathcal{S}_{\mathcal{X}}^k} \left| u_n^{(\text{T})}(\varphi, \mathbf{x}_{\ell(\mathbf{x})}; m_n) - \mathbb{E} \left[u_n^{(\text{T})}(\varphi, \mathbf{x}_{\ell(\mathbf{x})}; m_n) \right] \right| > \eta \sqrt{\frac{s_m \log(m)}{m}} \right). \quad (5.6.8) \end{aligned}$$

Notice that

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

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

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

and we can write

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

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

$$\sup_{\mathbf{x} \in S_{\mathcal{X}}^k} \mathbb{E}[W_{j, T}]^v = \sup_{\mathbf{x} \in S_{\mathcal{X}}^k} \left(C_2 s_m^{\beta_2} \epsilon_n^{\beta_1} \right)^v \mathbb{E}[\mathbb{E}[|\varphi^{(T)}(Y_{i_1}, \dots, Y_{i_k})| | \mathbf{X} = \mathbf{x}]]^v \quad (5.6.9)$$

$$\leq \left(C_2 s_m^{\beta_2} \epsilon_n^{\beta_1} \right)^v (\lambda \xi_n)^{v/q} \mu_q^{v/q} \quad (5.6.10)$$

$$\leq C_2^v \left(\frac{s_m \log m}{m} \right)^{v/2} (\lambda \xi_n)^{v/q} \mu_q^{v/q} \quad (5.6.11)$$

$$\begin{aligned}
 & \leq C_2^v \lambda^{v/q} (\xi_n)^{v/q - v/2} (s_m)^{1 - v/2} \mu_q^{v/q} (s_m)^{v-1} \\
 & \leq \mathfrak{C} s_m^{(v-1)}.
 \end{aligned}$$

The passage from (5.6.9) to (5.6.10) is possible by the use of the Jensen's inequality for the concave function z^a , for $0 < a \leq 1$, while (5.6.11) is by condition (C.5). Then for $v \geq 2$

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

where $\mathfrak{C} > 0$. Then, an application of classical inequality (see Corollary A.8-(ii) [Ferraty and Vieu \[2006b\]](#)) with $Z_i := W_{j, T}$ and $a_n^2 = s_m$, which gives us

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

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

$$\sup_{\mathbf{x} \in S_{\mathcal{X}}^k} |u_n^{(\text{T})}(\varphi, \mathbf{x}; m_n) - u_n^{(\text{T})}(\varphi, \mathbf{x}_{\ell(\mathbf{x})}; m_n)| = O_{a.co} \left(\sqrt{\frac{s_m \log(m)}{m}} \right). \quad (5.6.12)$$

Now, we get from (5.6.12) that:

$$\begin{aligned} & |\mathbb{E}[u_n^{(\text{T})}(\varphi, \mathbf{x}_{\ell(\mathbf{x})}; m_n)] - \mathbb{E}[u_n^{(\text{T})}(\varphi, \mathbf{x}; m_n)]| \\ &= |\mathbb{E}[u_n^{(\text{T})}(\varphi, \mathbf{x}_{\ell(\mathbf{x})}; m_n) - u_n^{(\text{T})}(\varphi, \mathbf{x}; m_n)]| \end{aligned} \quad (5.6.13)$$

$$\leq \mathbb{E}[|u_n^{(\text{T})}(\varphi, \mathbf{x}_{\ell(\mathbf{x})}; m_n) - u_n^{(\text{T})}(\varphi, \mathbf{x}; m_n)|]. \quad (5.6.14)$$

Similar to the bounded case, the transition from (5.6.13) to (5.6.14) is due to Jensen's inequality and some properties of the absolute value function. Furthermore, using the fact that for any constant a , $\mathbb{E}[a] = a$, we can directly conclude that

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

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

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

Continue, now, with (5.6.8), by imposing that the kernel function $G_{\varphi, \mathbf{x}_{\ell}}^{(\text{T})}(\cdot)$ is symmetric, the U-statistic is decomposed according to [Hoeffding \[1948\]](#) decomposition, that is

$$\begin{aligned} & u_n^{(\text{T})}(\varphi, \mathbf{x}_{\ell}; m_n) - \mathbb{E}[u_n^{(\text{T})}(\varphi, \mathbf{x}_{\ell}; m_n)] \\ &= \sum_{p=1}^k \frac{k!}{(k-p)!} u_n^{(p)} \left(\pi_{p,k}(G_{\varphi, \mathbf{x}_{\ell}}^{(\text{T})}) \right) \\ &= k u_n^{(1)} \left(\pi_{1,k}(G_{\varphi, \mathbf{x}_{\ell}}^{(\text{T})}) \right) + \sum_{p=2}^k \frac{k!}{(k-p)!} u_n^{(p)} \left(\pi_{p,k}(G_{\varphi, \mathbf{x}_{\ell}}^{(\text{T})}) \right). \end{aligned} \quad (5.6.15)$$

Let us first start with the linear term. We have

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

From Hoeffding's projection, we have

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

Set

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

We can clearly see that $Z_i^{(T)}$ are independent and identically distributed random variables bounded by $2k\lambda\xi_n^{1/q}C_1s_m$ with zero mean and

$$\sigma^2 \leq (\lambda\xi_n^{1/q}C_1s_m)^2.$$

An application of Bernstein's inequality yields

$$\begin{aligned} & \mathbb{P} \left(\sup_{\mathbf{x} \in \mathcal{S}_{\mathcal{X}}^k} \left| u_n^{(1)} \left(\pi_{1,k}(G_{\varphi, \mathbf{x}_\ell}^{(T)}) \right) \right| > \eta \sqrt{\frac{s_m \log(m)}{m}} \right) \\ & \leq \sum_{i=1}^{d_n} \mathbb{P} \left(\max_{1 \leq \ell_i \leq d_n} \left| u_n^{(1)} \left(\pi_{1,k}(G_{\varphi, \mathbf{x}_\ell}^{(T)}) \right) \right| > \eta \sqrt{\frac{s_m \log(m)}{m}} \right) \\ & \leq 2d_n \exp \left[- \frac{n((s_m \log(m))/m)\eta^2}{4(\lambda\xi_n^{1/q}C_1s_m)^2 + \frac{4}{3}\lambda\xi_n^{1/q}C_1s_m\eta\sqrt{(s_m \log(m))/m}} \right] \\ & \leq n^{\alpha - \tau\eta^2/C'_4}, \end{aligned}$$

for some positive constant C'_4 , resulting from the fact $m \leq n$ and $\log(m) \geq \tau \log(n)$. This implies that

$$\begin{aligned} & \sum_{n \geq 1} \mathbb{P} \left(\sup_{\mathbf{x} \in \mathcal{S}_{\mathcal{X}}^k} \left| u_n^{(1)} \left(\pi_{1,k}(G_{\varphi, \mathbf{x}_\ell}^{(T)}) \right) \right| > \eta \sqrt{\frac{s_m \log(m)}{m}} \right) \\ & \leq \sum_{n \geq 1} n^{\alpha - \tau\eta^2/C'_4} < \infty. \end{aligned}$$

Consequently, we get :

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

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

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

which implies that, for $1 \leq i \leq k$ and $\ell = (\ell_1, \dots, \ell_k)$:

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

To prove the above-mentioned equation, we need to apply Proposition 1 of Arcones [1995] (see Lemma 5.6.0.2). We can see that $G_{\varphi, \mathbf{x}_\ell}^{(T)}$ is bounded by $C_1 s_m \lambda \xi_n^{1/q}$, hence for $\eta > 0$ we have

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

where $\varepsilon_0 = \frac{\eta}{\sqrt{n}}$. Now for $t = \eta \sqrt{\frac{s_m \log(m)}{m}}$, Lemma 5.6.0.2 gives us :

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

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

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

where

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

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

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

Consequently, we have

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

Remainder Part

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

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

We have establish that the process is negligible, meaning that

$$\sup_{\mathbf{x} \in S_{\mathcal{X}}^k} \frac{\sqrt{m} \left| u_n^{(k)}(G_{\varphi, \mathbf{x}}^{(R)}) - \mathbb{E} \left(u_n^{(k)}(G_{\varphi, \mathbf{x}}^{(R)}) \right) \right|}{\sqrt{s_m \log(m)}} = o_{a.co}(1). \quad (5.6.16)$$

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

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

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

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

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

$$\sup_{\mathbf{x} \in S_{\mathcal{X}}^k} \frac{\sqrt{m} \left| u_n^{(k)}(G_{\varphi, \mathbf{x}}^{(R)}) \right|}{\sqrt{s_m \log(m)}} \leq (s_m^{-1} \xi_n)^{1/2} u_n^{(k)}(\tilde{F} \mathbb{1}_{\{\tilde{F} > \lambda \xi_n^{1/q}\}}) \quad (5.6.17)$$

$$\leq C_7 \xi_n u_n^{(k)}(\tilde{F} \mathbb{1}_{\{\tilde{F} > \lambda \xi_n^{1/q}\}}), \quad (5.6.18)$$

and

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

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

$$\sup_{\mathbf{x} \in S_{\mathcal{X}}^k} \frac{\sqrt{m} \left| \mathbb{E} \left(u_n^{(k)}(G_{\varphi, \mathbf{x}}^{(R)}) \right) \right|}{\sqrt{s_m \log(m)}} = o(1). \quad (5.6.19)$$

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

$$u_n^{(k)}(\tilde{F} \mathbb{1}_{\{\tilde{F} > \lambda \xi_n^{1/q}\}}) = o_{a.co} \left((s_m^{-1} \xi_n)^{-1/2} \right). \quad (5.6.20)$$

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

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

so by using the fact that

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

we deduce that

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

Finally, note that (5.6.17) implies

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

The preceding results of the arbitrary choice of $\lambda > 0$ gives that (5.6.20) holds, which, by combining with (5.6.19) and (5.6.17), completes the proof of (5.6.16). We finally obtain

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

Hence the proof is complete. □

Proof of Theorem 5.3.1.3

The conclusion of Theorem 5.3.1.3 can be obtained from the results of Theorem 5.3.1.1 and Theorem 5.3.1.2. We have

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

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

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

Hence now, depending on whether the function $\varphi(\cdot)$ is bounded or unbounded, we can apply Theorem 5.3.1.1 or Theorem 5.3.1.2 (respectively) to handle both \mathcal{S}_1 and \mathcal{S}_2 , and get for some $c'' > 0$ with probability 1 :

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

Hence the proof is complete. □

Proof of Theorem 5.3.1.4

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

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

Notice that

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

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

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

Which means that

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

where

$$I_1(\mathbf{x}) := \frac{1}{\mathbb{E}[\boldsymbol{\delta}_m(\mathbf{x}, \mathbf{X})]} \int_{\mathbb{B}(\mathbf{x}, \gamma)} \left(r^{(k)}(\varphi, \mathbf{t}) - r^{(k)}(\varphi, \mathbf{x}) \right) \boldsymbol{\delta}_m(\mathbf{x}, \mathbf{t}) \tilde{f}(\mathbf{t}) \mu(d\mathbf{t}), \quad (5.6.21)$$

and

$$I_2(\mathbf{x}) := \frac{1}{\mathbb{E}[\boldsymbol{\delta}_m(\mathbf{x}, \mathbf{X})]} \int_{\bar{\mathbb{B}}(\mathbf{x}, \gamma)} \left(r^{(k)}(\varphi, \mathbf{t}) - r^{(k)}(\varphi, \mathbf{x}) \right) \boldsymbol{\delta}_m(\mathbf{x}, \mathbf{t}) \tilde{f}(\mathbf{t}) \mu(d\mathbf{t}). \quad (5.6.22)$$

Therefore, we need to study the asymptotic behavior of both $\sup_{\mathbf{x} \in S_{\mathcal{X}}^k} (I_1(\mathbf{x}))$ and $\sup_{\mathbf{x} \in S_{\mathcal{X}}^k} (I_2(\mathbf{x}))$ to get

the desired result.

Let us start with the term $\sup_{\mathbf{x} \in S_{\mathcal{X}}^k} (I_1(\mathbf{x}))$, we have

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

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

$$\begin{aligned} & \sup_{\mathbf{x} \in S_{\mathcal{X}}^k} |I_1(\mathbf{x})| \\ &\leq \frac{C_f}{\mathbb{E}[\boldsymbol{\delta}_m(\mathbf{x}, \mathbf{X})]} \sup_{\mathbf{x} \in S_{\mathcal{X}}^k} \int_{\mathbb{B}(\mathbf{x}, \gamma)} C_3 d(\mathbf{x}, \mathbf{t}) \boldsymbol{\delta}_m(\mathbf{x}, \mathbf{t}) \mu(d\mathbf{t}) \\ &\leq \frac{C_f C_3 \gamma}{\mathbb{E}[\boldsymbol{\delta}_m(\mathbf{x}, \mathbf{X})]} \sup_{\mathbf{x} \in S_{\mathcal{X}}^k} \int_{\mathbb{B}(\mathbf{x}, \gamma)} \boldsymbol{\delta}_m(\mathbf{x}, \mathbf{t}) \mu(d\mathbf{t}), \end{aligned} \quad (5.6.23)$$

hence the term in (5.6.23) can be chosen smaller than 2ε as $m \rightarrow \infty$ by using the condition (5.2.6).

To investigate the term $\sup_{\mathbf{x} \in S_{\mathcal{X}}^k} (I_2(\mathbf{x}))$, notice that

$$\begin{aligned} & \sup_{\mathbf{x} \in S_{\mathcal{X}}^k} |I_2(\mathbf{x})| \\ &= \sup_{\mathbf{x} \in S_{\mathcal{X}}^k} \left| \frac{1}{\mathbb{E}[\boldsymbol{\delta}_m(\mathbf{x}, \mathbf{X})]} \int_{\bar{\mathbb{B}}(\mathbf{x}, \gamma)} \left(r^{(k)}(\varphi, \mathbf{t}) - r^{(k)}(\varphi, \mathbf{x}) \right) \boldsymbol{\delta}_m(\mathbf{x}, \mathbf{t}) \tilde{f}(\mathbf{t}) \mu(d\mathbf{t}) \right| \\ &\leq \frac{C_f C_3}{\mathbb{E}[\boldsymbol{\delta}_m(\mathbf{x}, \mathbf{X})]} \sup_{\mathbf{x} \in S_{\mathcal{X}}^k} \int_{\bar{\mathbb{B}}(\mathbf{x}, \gamma)} d(\mathbf{x}, \mathbf{t}) \boldsymbol{\delta}_m(\mathbf{x}, \mathbf{t}) \mu(d\mathbf{t}). \end{aligned} \quad (5.6.24)$$

By condition (5.2.9), we conclude that

$$\sup_{\mathbf{x} \in S_{\mathcal{X}}^k} |I_2(\mathbf{x})| \rightarrow 0 \text{ as } m \rightarrow \infty.$$

This concludes the proof of the Theorem. \square

Proof of Theorem 5.3.2.1

Following the same steps as the proof of Theorem 5.3.1.4, we can write directly :

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

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

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

where $I_1(\mathbf{x})$ and $I_2(\mathbf{x})$ are defined in (5.6.21) and (5.6.22) respectively. Now equation (5.6.23) gives us

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

by conditions (C.8) and (C.1). On the other hand, equation (5.6.24) gives us :

$$\begin{aligned} & \sup_{\mathbf{x} \in \mathcal{S}_{\mathcal{X}}^k} |I_2(\mathbf{x})| \\ & \leq \frac{C_f C_3}{\mathbb{E}[\boldsymbol{\delta}_m(\mathbf{x}, \mathbf{X})]} \sup_{\mathbf{x} \in \mathcal{S}_{\mathcal{X}}^k} \int_{\mathbb{B}(\mathbf{x}, \gamma)} d(\mathbf{x}, \mathbf{t}) \boldsymbol{\delta}_m(\mathbf{x}, \mathbf{t}) \mu(d\mathbf{t}) \\ & \leq \frac{C_f C_3 D_m}{\mathbb{E}[\boldsymbol{\delta}_m(\mathbf{x}, \mathbf{X})]} \sup_{\mathbf{x} \in \mathcal{S}_{\mathcal{X}}^k} \int_{\mathbb{B}(\mathbf{x}, \gamma)} \boldsymbol{\delta}_m(\mathbf{x}, \mathbf{t}) \mu(d\mathbf{t}) \\ & \leq O(D_m), \end{aligned}$$

by condition (C.8). This completes the proof of the theorem. \square

Appendix

Lemma 5.6.0.1 (Theorem A. page 201 Serfling [1980]) *Let f denote a symmetric \mathcal{X}^k -valued function fullfilling $\|f\|_{\infty} \leq c$,*

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

and

$$\sigma^2 = \text{Var} (f (X_1, \dots, X_k)),$$

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

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

Lemma 5.6.0.2 (Proposition.1 Arcones [1995]) *If $G : S^k \rightarrow \mathbb{R}$ is a measurable symmetric function with $\|G\|_\infty = b$ then*

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

Definition 5.6.0.3 *A symmetric and \mathbb{P}^m -integrable kernel $f : \mathcal{X}^k \rightarrow \mathbb{R}$ is \mathbb{P} -degenerate of order $r - 1$, notationally $f \in L_2^r(\mathbb{P}^k)$, if and only if*

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

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

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

is not a constant function.

5.7 References

- Ahmad, I. A. (1981). A note on nonparametric density estimation for dependent variables using a delta sequence. *Ann. Inst. Statist. Math.*, **33**(2), 247–254. [191](#)
- Almanjahie, I. M., Bouzebda, S., Chikr Elmezouar, Z., and Laksaci, A. (2022a). The functional k NN estimator of the conditional expectile: uniform consistency in number of neighbors. *Stat. Risk Model.*, **38**(3-4), 47–63. [191](#)
- Almanjahie, I. M., Bouzebda, S., Kaid, Z., and Laksaci, A. (2022b). Nonparametric estimation of expectile regression in functional dependent data. *J. Nonparametr. Stat.*, **34**(1), 250–281. [191](#)
- Aneiros, G., Cao, R., Fraiman, R., Genest, C., and Vieu, P. (2019). Recent advances in functional data analysis and high-dimensional statistics. *J. Multivariate Anal.*, **170**, 3–9. [190](#)
- Araujo, A. and Giné, E. (1980). *The central limit theorem for real and Banach valued random variables*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York-Chichester-Brisbane. [190](#)

- Arcones, M. A. (1995). A bernstein-type inequality for u-statistics and u-processes. *Statistics & probability letters*, **22**(3), 239–247. [215](#), [221](#), [227](#)
- Arcones, M. A. and Giné, E. (1993). Limit theorems for U-processes. *Ann. Probab.*, **21**(3), 1494–1542. [188](#), [189](#)
- Arcones, M. A. and Wang, Y. (2006). Some new tests for normality based on U-processes. *Statist. Probab. Lett.*, **76**(1), 69–82. [188](#)
- Attouch, M., Laksaci, A., and Rafea, F. (2019). On the local linear estimate for functional regression: uniform in bandwidth consistency. *Comm. Statist. Theory Methods*, **48**(8), 1836–1853. [191](#)
- Basu, A. K. and Sahoo, D. K. (1985). Speed of convergence in density estimation by delta sequence and estimation of mode. *Calcutta Statist. Assoc. Bull.*, **34**(133-134), 1–13. [191](#)
- Belitser, E. (2000). Local minimax pointwise estimation of a multivariate density. *Statist. Neerlandica*, **54**(3), 351–365. [192](#)
- Bellet, A. and Habrard, A. (2015). Robustness and generalization for metric learning. *Neuro-computing*, **151**, 259–267. [209](#)
- Bellet, A., Habrard, A., and Sebban, M. (2013). A survey on metric learning for feature vectors and structured data. *CoRR*, **abs/1306.6709**. [208](#)
- Bergsma, W. and Dassios, A. (2014). A consistent test of independence based on a sign covariance related to Kendall’s tau. *Bernoulli*, **20**(2), 1006–1028. [210](#)
- Blum, J. R., Kiefer, J., and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.*, **32**, 485–498. [210](#)
- Boniece, B. C., Horváth, L., and Jacobs, P. (2022). Change point detection in high dimensional data with u-statistics. [188](#)
- Borovkova, S., Burton, R., and Dehling, H. (2001). Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. *Trans. Amer. Math. Soc.*, **353**(11), 4261–4318. [189](#)
- Bosq, D. (2000). *Linear processes in function spaces*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York. Theory and applications. [190](#)
- Bouzebda and Taachouche (2022). Rates of the strong uniform consistency for the kernel-type regression function estimators with general kernels on manifolds. *Math. Methods Statist.*, pages 1–62. [199](#)
- Bouzebda, S. (2011). Some new multivariate tests of independence. *Math. Methods Statist.*, **20**(3), 192–205. [210](#)

- Bouzebda, S. (2012). On the strong approximation of bootstrapped empirical copula processes with applications. *Math. Methods Statist.*, **21**(3), 153–188. [192](#)
- Bouzebda, S. (2014). General tests of independence based on empirical processes indexed by functions. *Stat. Methodol.*, **21**, 59–87. [210](#)
- Bouzebda, S. (2023). General tests of conditional independence based on empirical processes indexed by functions. *Jpn. J. Stat. Data Sci. To appear.*, pages 1–60. [210](#)
- Bouzebda, S. and Chaouch, M. (2022). Uniform limit theorems for a class of conditional Z-estimators when covariates are functions. *J. Multivariate Anal.*, **189**, Paper No. 104872, 21. [191](#)
- 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**(4), 864–914. [205](#), [206](#)
- Bouzebda, S. and Nemouchi, B. (2019). Central limit theorems for conditional empirical and conditional U-processes of stationary mixing sequences. *Math. Methods Statist.*, **28**(3), 169–207. [190](#)
- 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**(2), 452–509. [190](#), [199](#)
- 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*, pages 1–56. [191](#), [199](#)
- 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. [190](#), [191](#), [199](#), [205](#), [206](#)
- Bouzebda, S. and Nezzal, A. (2023). Asymptotic properties of conditional U-statistics using delta sequences. *Comm. Statist. Theory Methods*, pages 1–45. [192](#)
- Bouzebda, S. and Soukarieh, I. (2023). Renewal type bootstrap for U-process Markov chains. *Markov Process. Related Fields*, pages 1–52. [189](#)
- Bouzebda, S. and Zari, T. (2013). Asymptotic behavior of weighted multivariate Cramér-von Mises-type statistics under contiguous alternatives. *Math. Methods Statist.*, **22**(3), 226–252. [210](#)
- Bouzebda, S., El Faouzi, N.-E., and Zari, T. (2011). On the multivariate two-sample problem using strong approximations of empirical copula processes. *Comm. Statist. Theory Methods*, **40**(8), 1490–1509. [210](#)

- Bouzebda, S., Chaouch, M., and Laïb, N. (2016). Limiting law results for a class of conditional mode estimates for functional stationary ergodic data. *Math. Methods Statist.*, **25**(3), 168–195. [191](#)
- 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**(2), 321–358. [190](#)
- Bouzebda, S., Mohammedi, M., and Laksaci, A. (2022a). The k-nearest neighbors method in single index regression model for functional quasi-associated time series data. *Rev. Mat. Complut.*, pages 1–30. [191](#)
- Bouzebda, S., El-hadjali, T., and Ferfache, A. A. (2022b). Uniform in bandwidth consistency of conditional U-statistics adaptive to intrinsic dimension in presence of censored data. *Sankhya A*, pages 1–59. [190](#), [205](#), [206](#)
- Brunel, E. and Comte, F. (2006). Adaptive nonparametric regression estimation in presence of right censoring. *Math. Methods Statist.*, **15**(3), 233–255. [205](#)
- Cao, Q., Guo, Z.-C., and Ying, Y. (2016). Generalization bounds for metric and similarity learning. *Mach. Learn.*, **102**(1), 115–132. [209](#)
- Carbonez, A., Györfi, L., and van der Meulen, E. C. (1995). Partitioning-estimates of a regression function under random censoring. *Statist. Decisions*, **13**(1), 21–37. [204](#)
- Chen, Y. and Datta, S. (2019). Adjustments of multi-sample U-statistics to right censored data and confounding covariates. *Comput. Statist. Data Anal.*, **135**, 1–14. [205](#)
- Chentsov, N. N. (1962). Evaluation of an unknown distribution density from observations. *Sov. Math., Dokl.*, **3**, 1559–1562. [191](#)
- Chokri, K. and Bouzebda, S. (2022). Uniform-in-bandwidth consistency results in the partially linear additive model components estimation. *Communications in Statistics-Theory and Methods*, pages 1–42. [199](#)
- Cléménçon, S., Colin, I., and Bellet, A. (2016). Scaling-up empirical risk minimization: optimization of incomplete U-statistics. *J. Mach. Learn. Res.*, **17**, Paper No. 76, 36. [208](#)
- Cristóbal Cristóbal, J. A., Faraldo Roca, P., and González Manteiga, W. (1987). A class of linear regression parameter estimators constructed by nonparametric estimation. *Ann. Statist.*, **15**(2), 603–609. [192](#)
- Dabo-Niang, S. (2004). Kernel density estimator in an infinite-dimensional space with a rate of convergence in the case of diffusion process. *Appl. Math. Lett.*, **17**(4), 381–386. [194](#)

- 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**(4), 680–700. [205](#)
- de la Peña, V. H. and Giné, E. (1999). *Decoupling*. Probability and its Applications (New York). Springer-Verlag, New York. From dependence to independence, Randomly stopped processes. U-statistics and processes. Martingales and beyond. [188](#), [189](#), [200](#)
- Deheuvels, P. (2011). One bootstrap suffices to generate sharp uniform bounds in functional estimation. *Kybernetika (Prague)*, **47**(6), 855–865. [199](#)
- Denker, M. and Keller, G. (1983). On U-statistics and v. Mises' statistics for weakly dependent processes. *Z. Wahrsch. Verw. Gebiete*, **64**(4), 505–522. [189](#)
- Devroye, L. and Lugosi, G. (2001). *Combinatorial methods in density estimation*. Springer Series in Statistics. Springer-Verlag, New York. [188](#)
- Dony, J. and Mason, D. M. (2008). Uniform in bandwidth consistency of conditional U-statistics. *Bernoulli*, **14**(4), 1108–1133. [190](#)
- Eggermont, P. P. B. and LaRiccia, V. N. (2001). *Maximum penalized likelihood estimation*. Vol. I. Springer Series in Statistics. Springer-Verlag, New York. Density estimation. [188](#)
- Ferraty, F. and Vieu, P. (2006a). *Nonparametric functional data analysis*. Springer Series in Statistics. Springer, New York. Theory and practice. [190](#), [209](#)
- Ferraty, F. and Vieu, P. (2006b). *Nonparametric functional data analysis*. Springer Series in Statistics. Springer, New York. Theory and practice. [218](#)
- Ferraty, F. and Vieu, P. (2008). Erratum of: non-parametric models for functional data, with application in regression, time-series prediction and curve discrimination. *Journal of non-parametric statistics*, **20**(2), 187–189. [195](#), [199](#)
- Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2010). Rate of uniform consistency for non-parametric estimates with functional variables. *J. Statist. Plann. Inference*, **140**(2), 335–352. [190](#)
- Földes, A. and Rejtő, L. (1981). A LIL type result for the product limit estimator. *Z. Wahrsch. Verw. Gebiete*, **56**(1), 75–86. [206](#)
- Fu, K.-A. (2012). An application of u-statistics to nonparametric functional data analysis. *Communications in Statistics-Theory and Methods*, **41**(9), 1532–1542. [193](#)
- Gasser, T., Hall, P., and Presnell, B. (1998). Nonparametric estimation of the mode of a distribution of random curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **60**(4), 681–691. [190](#)

- Ghosal, S., Sen, A., and van der Vaart, A. W. (2000). Testing monotonicity of regression. *Ann. Statist.*, **28**(4), 1054–1082. [188](#)
- Giné, E. and Mason, D. M. (2007a). Laws of the iterated logarithm for the local U-statistic process. *J. Theoret. Probab.*, **20**(3), 457–485. [188](#), [190](#)
- Giné, E. and Mason, D. M. (2007b). On local U-statistic processes and the estimation of densities of functions of several sample variables. *Ann. Statist.*, **35**(3), 1105–1145. [188](#)
- Halmos, P. R. (1946). The theory of unbiased estimation. *Ann. Math. Statistics*, **17**, 34–43. [189](#)
- Härdle, W. (1990). *Applied nonparametric regression*, volume 19 of *Econometric Society Monographs*. Cambridge University Press, Cambridge. [188](#)
- Harel, M. and Puri, M. L. (1996). Conditional U-statistics for dependent random variables. *J. Multivariate Anal.*, **57**(1), 84–100. [190](#)
- Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics*, **19**, 293–325. [188](#), [189](#), [200](#), [213](#), [219](#)
- Horváth, L. and Kokoszka, P. (2012). *Inference for functional data with applications*. Springer Series in Statistics. Springer, New York. [190](#)
- Isogai, E. (1990). Nonparametric estimation of a regression function by delta sequences. *Ann. Inst. Statist. Math.*, **42**(4), 699–708. [191](#)
- Jiang, J. (2022). *Nonparametric Statistics*, pages 379–415. Springer International Publishing, Cham. [188](#)
- Jin, R., Wang, S., and Zhou, Y. (2009). Regularized distance metric learning: theory and algorithm. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 22. Curran Associates, Inc. [209](#)
- Joly, E. and Lugosi, G. (2016). Robust estimation of U-statistics. *Stochastic Process. Appl.*, **126**(12), 3760–3773. [188](#)
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53**, 457–481. [205](#)
- Kara-Zaitri, L., Laksaci, A., Rachdi, M., and Vieu, P. (2017). Uniform in bandwidth consistency for various kernel estimators involving functional data. *J. Nonparametr. Stat.*, **29**(1), 85–107. [191](#)
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**(1/2), 81–93. [206](#)

- Kohler, M., Máthé, K., and Pintér, M. (2002). Prediction from randomly right censored data. *J. Multivariate Anal.*, **80**(1), 73–100. [204](#), [205](#)
- Kulathinal, S. and Dewan, I. (2022). Weighted u-statistics for likelihood-ratio ordering of bivariate data. *Statist. Papers.* [188](#)
- Lee, A. J. (1990). *U-statistics*, volume 110 of *Statistics: Textbooks and Monographs*. Marcel Dekker Inc., New York. Theory and practice. [189](#)
- Lee, S., Linton, O., and Whang, Y.-J. (2009). Testing for stochastic monotonicity. *Econometrica*, **77**(2), 585–602. [188](#)
- Leucht, A. (2012). Degenerate U- and V-statistics under weak dependence: asymptotic theory and bootstrap consistency. *Bernoulli*, **18**(2), 552–585. [189](#)
- Leucht, A. and Neumann, M. H. (2013). Degenerate U- and V-statistics under ergodicity: asymptotics, bootstrap and applications in statistics. *Ann. Inst. Statist. Math.*, **65**(2), 349–386. [189](#)
- Ling, N. and Vieu, P. (2018). Nonparametric modelling for functional data: selected survey and tracks for future. *Statistics*, **52**(4), 934–949. [190](#)
- Ling, N., Meng, S., and Vieu, P. (2019). Uniform consistency rate of k NN regression estimation for functional time series data. *J. Nonparametr. Stat.*, **31**(2), 451–468. [191](#)
- Litimein, O., Laksaci, A., Mechab, B., and Bouzebda, S. (2023). Local linear estimate of the functional expectile regression. *Statistics & Probability Letters*, **192**, 109682. [191](#)
- 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 Statist.*, **18**(2), 159–184. [203](#), [205](#)
- Maouloud, S. M. O. (2007). *Quelques aspects fonctionnels et non fonctionnels des grandes déviations et des déviations modérées en estimation non-paramétrique*. Ph.D. thesis, Université de Reims-Champagne Ardenne. [197](#)
- Marron, J. S. and Härdle, W. (1986). Random approximations to some measures of accuracy in nonparametric curve estimation. *J. Multivariate Anal.*, **20**(1), 91–113. [191](#)
- Minsker, S. (2022). U-statistics of growing order and sub-gaussian mean estimators with sharp constants. [189](#)
- Mohammedi, M., Bouzebda, S., and Laksaci, A. (2021). The consistency and asymptotic normality of the kernel type expectile regression estimator for functional data. *J. Multivariate Anal.*, **181**, Paper No. 104673, 24. [191](#)

- Nadaraja, E. A. (1964). On a regression estimate. *Teor. Verojatnost. i Primenen.*, **9**, 157–159. [189](#)
- Nadaraya, E. A. (1989). *Nonparametric estimation of probability densities and regression curves*, volume 20 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht. Translated from the Russian by Samuel Kotz. [188](#)
- Nolan, D. and Marron, J. S. (1989). Uniform consistency of automatic and location-adaptive delta-sequence estimators. *Probab. Theory Related Fields*, **80**(4), 619–632. [191](#)
- Ouassou, I. and Rachdi, M. (2012). Regression operator estimation by delta-sequences method for functional data and its applications. *AStA Adv. Stat. Anal.*, **96**(4), 451–465. [192](#), [194](#), [196](#), [209](#)
- Prakasa Rao, B. L. S. (1978). Density estimation for Markov processes using delta-sequences. *Ann. Inst. Statist. Math.*, **30**(2), 321–328. [191](#)
- Prakasa Rao, B. L. S. (1983). *Nonparametric functional estimation. Probability and Mathematical Statistics*. Orlando, Florida, etc.: Academic Press (Harcourt Brace Jovanovich, Publishers). XIV, 522 p. \$ 70.00 (1983). [194](#)
- Prakasa Rao, B. L. S. (1999a). *Semimartingales and their statistical inference*, volume 83 of *Monogr. Stat. Appl. Probab.* Boca Raton, FL: Chapman & Hall/ CRC. [194](#)
- Prakasa Rao, B. L. S. (1999b). *Statistical inference for diffusion type processes*, volume 8 of *Kendall's Libr. Stat.* London: Arnold. [194](#)
- Prakasa Rao, B. L. S. (2010a). Nonparametric density estimation for functional data by delta sequences. *Braz. J. Probab. Stat.*, **24**(3), 468–478. [192](#), [194](#)
- Prakasa Rao, B. L. S. (2010b). Nonparametric density estimation for functional data by delta sequences. *Braz. J. Probab. Stat.*, **24**(3), 468–478. [194](#)
- Prakasa Rao, B. L. S. and Sen, A. (1995). Limit distributions of conditional U-statistics. *J. Theoret. Probab.*, **8**(2), 261–301. [190](#)
- Rachdi, M. and Monsan, V. (1999). Asymptotic properties of p -adic spectral estimates of second order. *J. Combin. Inform. System Sci.*, **24**(2-4), 113–142. [194](#)
- Ramsay, J. O. and Silverman, B. W. (2002a). *Applied functional data analysis*. Springer Series in Statistics. Springer-Verlag, New York. Methods and case studies. [190](#)
- Ramsay, J. O. and Silverman, B. W. (2002b). *Applied functional data analysis*. Springer Series in Statistics. Springer-Verlag, New York. Methods and case studies. [190](#)
- Schick, A., Wang, Y., and Wefelmeyer, W. (2011). Tests for normality based on density estimators of convolutions. *Statist. Probab. Lett.*, **81**(2), 337–343. [188](#)

- Sen, A. (1994). Uniform strong consistency rates for conditional U-statistics. *Sankhyā Ser. A*, **56**(2), 179–194. [190](#)
- Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. John Wiley & Sons, Inc., New York. Wiley Series in Probability and Mathematical Statistics. [226](#)
- Sherman, R. P. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Ann. Statist.*, **22**(1), 439–459. [188](#)
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London. [188](#)
- Soukariéh, I. and Bouzebda, S. (2022). Exchangeably weighted bootstraps of general Markov U-process. *Mathematics*, **10**(20), 1–42. [189](#)
- Soukariéh, I. and Bouzebda, S. (2023). Renewal type bootstrap for increasing degree u -process of a Markov chain. *J. Multivariate Anal.*, **195**, 105143. [189](#)
- Stute, W. (1991). Conditional U-statistics. *Ann. Probab.*, **19**(2), 812–825. [189](#), [190](#), [191](#)
- Stute, W. (1993). Almost sure representations of the product-limit estimator for truncated data. *Ann. Statist.*, **21**(1), 146–156. [188](#)
- Stute, W. (1994a). L^p -convergence of conditional U-statistics. *J. Multivariate Anal.*, **51**(1), 71–82. [207](#)
- Stute, W. (1994b). Universally consistent conditional U-statistics. *Ann. Statist.*, **22**(1), 460–473. [207](#)
- Stute, W. (1996). Symmetrized NN-conditional U-statistics. In *Research developments in probability and statistics*, pages 231–237. VSP, Utrecht. [190](#)
- Stute, W. and Wang, J.-L. (1993). Multi-sample U-statistics for censored data. *Scand. J. Statist.*, **20**(4), 369–374. [205](#)
- Susarla, V. and Walter, G. (1981). Estimation of a multivariate density function using delta sequences. *Ann. Statist.*, **9**(2), 347–355. [191](#), [196](#), [198](#)
- von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statistics*, **18**, 309–348. [189](#)
- Walter, G. and Blum, J. (1979). Probability density estimation using delta sequences. *the Annals of Statistics*, pages 328–340. [198](#)
- Walter, G. G. (1965). Expansions of distributions. *Transactions of the American Mathematical Society*, **116**, 492–510. [198](#)

- Walter, G. G. (1983). A general approach to classification problems. *Inform. Sci.*, **30**(1), 67–77. [192](#)
- Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*, volume 60 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, Ltd., London. [188](#)
- Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A*, **26**, 359–372. [189](#)
- Watson, G. S. and Leadbetter, M. R. (1964). Hazard analysis. II. *Sankhyā Ser. A*, **26**, 101–116. [191](#)
- Winter, B. B. (1975). Rate of strong consistency of two nonparametric density estimators. *Ann. Statist.*, **3**, 759–766. [191](#)
- Yuan, A., Giurcanu, M., Luta, G., and Tan, M. T. (2017). U-statistics with conditional kernels for incomplete data models. *Ann. Inst. Statist. Math.*, **69**(2), 271–302. [205](#)

Chapter 6

Conclusions and perspectives

6.1 Concluding remarks : Chapter 3

In Chapter 3, we have considered the problem of the nonparametric estimation of the conditional U-statistics when the covariates take values in some abstract space. We have introduced the k -NN estimator of the conditional U-statistics. More precisely, we have considered a statistical procedure combining the local adaptivity and the easiness of implementation of k -NN estimates together with the conditional U-statistics flexibility in the regression problems involving the functional data. There are basically no restrictions on the choice of the kernel function in our setup, apart from satisfying some mild conditions. The selection of the NN, however, is more problematic. It is worth noticing that the choice of the NN 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 NN 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 NN varying according to the criteria applied and to the available data and location, which cannot be achieved by using the classical methods. The interested reader may refer to [Mason \[2012\]](#) for more details and discussion on a similar subject for the selection of the bandwidth parameter in the finite-dimensional kernel estimation. The UINN consistency result given in Corollary 3.4.2.5 shows that any choice of k between $k_{1,n}$ and $k_{2,n}$ ensures the consistency of $\hat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h}_{n,k}(\mathbf{t}))$. Namely, the fluctuation of k in some chosen set does not affect the consistency of the nonparametric estimator $\hat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h}_{n,k}(\mathbf{t}))$. In our future study, we will investigate the weak convergence of $\hat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h}_{n,k}(\mathbf{t}))$. A future research direction would be to establish some asymptotic properties of $\hat{r}_n^{*(m)}(\varphi, \mathbf{t}; \mathbf{h}_{n,k}(\mathbf{t}))$ as such investigated in this work in the setting of serially dependent observations, which requires nontrivial mathematics, this would go well beyond the scope of the present chapter, one can refer to [Arcones and Yu \[1994\]](#) and [Bouzebda and Nemouchi \[2019\]](#).

6.2 Concluding remarks : Chapter 4

In Chapter 4, we deal with a quite general nonparametric statistical curve estimation setting, including the Stute estimator as a particular case. The class of “delta sequence estimators” is defined and treated here. This class also includes the orthogonal series and histogram methods. To obtain our results, we need assumptions requiring some regularity on the conditional U-statistics and conditional moments. In particular, the conditional moment assumption permits to consider the unbounded setting. To stress the importance of this study, we have presented some potential applications, including discrimination problems, metric learning, and multipartite ranking. Extending non-parametric functional ideas to general dependence structure is a rather underdeveloped field. Note that the ergodic framework avoids the widely used strong mixing condition and its variants to measure the dependency and the very involved probabilistic calculations that it implies. It would be interesting to extend our work to the case of the functional ergodic data Bouzebda *et al.* [2015, 2022a], Bouzebda and Didi [2017a,b, 2021, 2022] and Bouzebda and Chaouch [2022a] which requires non-trivial mathematics, this would go well beyond the scope of the present work. Another research direction is to consider the projection pursuit regression and projection pursuit conditional distribution, which need an extension and generalization of the methods used in the present work. If we assume that the regression function $r^{(k)}(\varphi, \cdot)$ is smooth enough that is differentiable at a fixed \mathbf{t}_0 , it will be better to use the local polynomial regression techniques, refer to Fan and Gijbels [1996], to obtain a more appropriate estimate at \mathbf{t}_0 than that given by the Nadaraya-Watson type estimator. We will not treat the uniform consistency of such estimators in the present paper and leave it for future investigation.

6.3 Concluding remarks : Chapter 5

In Chapter 5, the conditional U-statistics regression operator estimation methods for random elements taking values in an infinite-dimensional separable Banach space are generalized to the delta-sequences techniques. The space of continuous functions on the interval $(0, 1)$ with the supremum norm illustrates a separable Banach space. Notably, the method of delta-sequences unifies the kernel method of the probability density function estimation, the histogram method, and a few other methods, including the method of orthogonal series for appropriate choices of orthonormal bases in the one-dimensional and finite-dimensional cases. We have obtained strong uniform consistency results in abstract settings under some conditions on the model. The general framework that we consider extends the existing methods to higher-order statistics; this has a significant impact both from a theoretical and practical point of view. In a future investigation, considering the limiting law of the conditional U-statistics regression estimators based on the delta sequence will be of interest. A natural extension of the present investigation is to consider the serial-dependent setting such as the mixing (see Bouzebda and Soukariéh [2022a,b], Bouzebda and Nemouchi [2022], Bouzebda *et al.* [2022b]) or the ergodic processes

(see [Bouzebda and Chaouch \[2022b\]](#), [Didi and Bouzebda \[2022\]](#)). In a future investigation of the functional delta sequence local linear approach estimators, it will be natural to think about the possibility of obtaining an alternative estimator that benefits from the advantages of both methods, the local linear method and the delta sequence approach. This is because both methods have their own distinct advantages. Many methods have been developed and established to construct, in asymptotically optimal ways, bandwidth selection rules for nonparametric kernel estimators, particularly for the Nadaraya-Watson regression estimator. We quote several of these methods, including [Rachdi and Vieu \[2007\]](#), [Bouzebda *et al.* \[2022c\]](#), [Bouzebda and Nezzal \[2022\]](#). This parameter needs to have an appropriate value chosen for it to ensure that satisfactory practical performances are achieved, either in the typical situation of finite dimensions or in the framework of infinite dimensions. On the other hand, to the best of our knowledge, no such studies are currently conducted to treat generic functional conditional U-statistics. This exemplifies a potential new avenue for research in the future.

6.4 References

- Arcones, M. A. and Yu, B. (1994). Central limit theorems for empirical and U-processes of stationary mixing sequences. *J. Theoret. Probab.*, **7**(1), 47–71. [237](#)
- Bouzebda, S. and Chaouch, M. (2022a). Uniform limit theorems for a class of conditional Z-estimators when covariates are functions. *J. Multivariate Anal.*, **189**, Paper No. 104872, 21. [238](#)
- Bouzebda, S. and Chaouch, M. (2022b). Uniform limit theorems for a class of conditional z-estimators when covariates are functions. *Journal of Multivariate Analysis*, **189**, 104872. [239](#)
- Bouzebda, S. and Didi, S. (2017a). Additive regression model for stationary and ergodic continuous time processes. *Comm. Statist. Theory Methods*, **46**(5), 2454–2493. [238](#)
- Bouzebda, S. and Didi, S. (2017b). Multivariate wavelet density and regression estimators for stationary and ergodic discrete time processes: asymptotic results. *Comm. Statist. Theory Methods*, **46**(3), 1367–1406. [238](#)
- Bouzebda, S. and Didi, S. (2021). Some asymptotic properties of kernel regression estimators of the mode for stationary and ergodic continuous time processes. *Rev. Mat. Complut.*, **34**(3), 811–852. [238](#)
- Bouzebda, S. and Didi, S. (2022). Some results about kernel estimators for function derivatives based on stationary and ergodic continuous time processes with applications. *Comm. Statist. Theory Methods*, **51**(12), 3886–3933. [238](#)

- Bouzebda, S. and Nemouchi, B. (2019). Central limit theorems for conditional empirical and conditional U-processes of stationary mixing sequences. *Math. Methods Statist.*, **28**(3), 169–207. [237](#)
- 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*, pages 1–56. [238](#)
- 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. [239](#)
- Bouzebda, S. and Soukarieh, I. (2022a). Non-parametric conditional u-processes for locally stationary functional random fields under stochastic sampling design. *Mathematics*, **11**(1), 16. [238](#)
- Bouzebda, S. and Soukarieh, I. (2022b). Nonparametric conditional u -processes for locally stationary functional random fields under stochastic sampling design. *Mathematics*, **10**(24), 1–70. [238](#)
- Bouzebda, S., Didi, S., and El Hajj, L. (2015). Multivariate wavelet density and regression estimators for stationary and ergodic continuous time processes: asymptotic results. *Math. Methods Statist.*, **24**(3), 163–199. [238](#)
- Bouzebda, S., Chaouch, M., and Didi Biha, S. (2022a). Asymptotics for function derivatives estimators based on stationary and ergodic discrete time processes. *Ann. Inst. Statist. Math.*, **74**(4), 737–771. [238](#)
- Bouzebda, S., Mohammedi, M., and Laksaci, A. (2022b). The k -nearest neighbors method in single index regression model for functional quasi-associated time series data. *Rev. Mat. Complut.*, pages 1–30. [238](#)
- Bouzebda, S., El-hadjali, T., and Ferfache, A. A. (2022c). Uniform in bandwidth consistency of conditional U-statistics adaptive to intrinsic dimension in presence of censored data. *Sankhya A*, pages 1–59. [239](#)
- Didi, S. and Bouzebda, S. (2022). Wavelet density and regression estimators for continuous time functional stationary and ergodic processes. *Mathematics*, **10**(22), 4356. [239](#)
- Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*, volume 66 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London. [238](#)
- Mason, D. M. (2012). Proving consistency of non-standard kernel estimators. *Stat. Inference Stoch. Process.*, **15**(2), 151–176. [237](#)
- Rachdi, M. and Vieu, P. (2007). Nonparametric regression for functional data: automatic smoothing parameter selection. *J. Statist. Plann. Inference*, **137**(9), 2784–2801. [239](#)