



HAL
open science

Theoretical contribution to the U-processes in Markov and dependent setting: asymptotic and bootstraps

Inass Soukariéh

► **To cite this version:**

Inass Soukariéh. Theoretical contribution to the U-processes in Markov and dependent setting: asymptotic and bootstraps. Statistics [math.ST]. Université de Technologie de Compiègne, 2022. English. NNT : 2022COMP2709 . tel-03975688

HAL Id: tel-03975688

<https://theses.hal.science/tel-03975688>

Submitted on 6 Feb 2023

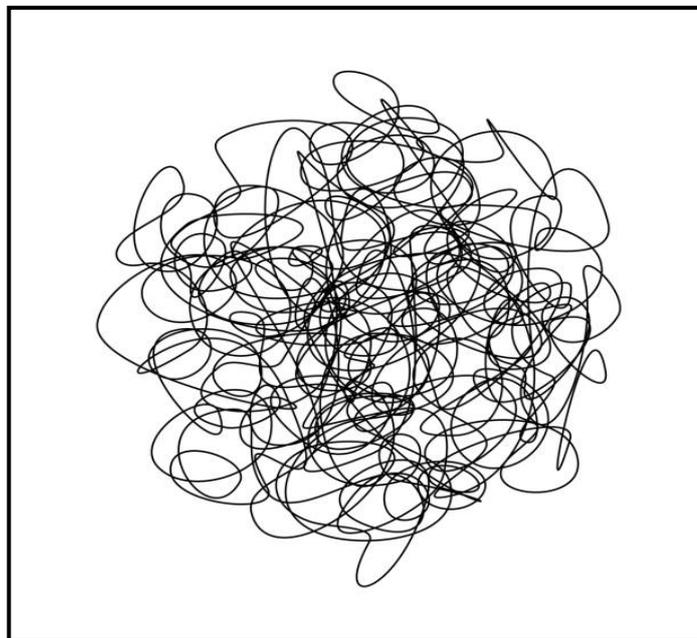
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

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 Inass SOUKARIEH

Theoretical contribution to the U-processes in Markov and dependent setting: asymptotic and bootstraps

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



Soutenue le 6 décembre 2022

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

D2709

THÈSE DE DOCTORAT

Université de Technologie de Compiègne

Theoretical contribution to the U -processes in Markov and dependent setting: asymptotic and bootstraps

Soutenue par

Inass SOUKARIEH

Le 06 Décembre 2022

École doctorale n°71

**École doctorale des
Sciences pour les
ingénieries**

Spécialité

Mathématiques appliquées

Préparée au

Laboratoire de mathématiques
appliquées de Compiègne(LMAC)

Composition du jury :

Michel BRONIATOWSKI Prof. Sorbonne Université	<i>Rapporteur</i>
Mustapha RACHDI Prof. Université Grenoble Alpes	<i>Rapporteur</i>
Mokhtar Z. ALAYA MCF. Université de Technologie de Compiègne	<i>Examineur</i>
Ghislaine GAYRAUD Prof. Université de Technologie de Compiègne	<i>Présidente du jury</i>
Amor KEZIOU MCF. HDR. Université de Reims Champagne- Ardenne	<i>Examineur</i>
Nikolaos LIMNIOS Prof. Université de Technologie de Compiègne	<i>Co-Directeur</i>
Salim BOUZEBDA Prof. Université de Technologie de Compiègne	<i>Directeur</i>

To my parents.

For always loving and supporting me.

And

To the memory of May.

Whose support, dedication to academic excellence
and perseverance inspired me throughout the duration of my studies.

Acknowledgements

I would like to thank my Ph.D. supervisors, Professors Salim BOUZEBDA and Nikolaos LIMNIOS, for supporting me during these past three years. My sincerest thanks to my supervisor Salim BOUZEBDA, for his invaluable supervision and tutelage during my Ph.D. degree. I have been lucky to have a supervisor who cared so much about my work and responded promptly to my questions and queries.

My gratitude also goes to the jury members who agreed to examine and evaluate my work. Prof. Michel BRONIATOWSKI and Prof. Mustapha RACHDI have been very generous with their time, ideas, and insightful comments. Thank you so much. My thanks also go to the jury members: the president Prof. Ghislaine GAYRAUD, Dr. Amor KEZIOU, and Dr. Mokhtar Z. ALAYA.

Completing this work would have been all the more difficult were it not for the support and friendship provided by the members of LMAC. I am indebted to them for their help. In particular, I would like to thank Thouria, Nour El-Houda, and my officemate Djaouad.

My parents deserve endless thanks for their support and love. Without them, nothing would have been possible.

My lovely husband, no words can describe my thanks and appreciation to you. You are always my source of inspiration, security and advice. You taught me to believe in myself, be unique, and always be persistent.

My sisters, Katia and Hiba, and my aunt, who experienced all of my research's ups and downs, offered invaluable support and humor over the years. Thanks aren't enough for you.

To my many friends and family in France and Lebanon, you should know that your support and encouragement were worth more than I can express on paper.

To my friends Hassan, Bouchra and Zeinab – thank you for your continued friendship, support, chats, and laughs.

Publications

Papers in Peer-Reviewed Journals

- [P.1] Inass SOUKARIEH, Salim BOUZEBDA. "*Exchangeably weighted bootstraps of General Markov U-process*". **Mathematics**, 2022, vol. 10, no 20, p.3745. 42pp. DOI: <https://doi.org/10.3390/math10203745>.
- [P.2] Inass SOUKARIEH, Salim BOUZEBDA. "*Renewal Type Bootstrap for U-process Markov Chains*". **Markov Processes and Related Fields**.
- [P.3] Inass SOUKARIEH, Salim BOUZEBDA. "*Renewal Type Bootstrap for Increasing degree U-process of Markov Chain*". **Journal of Multivariate Analysis**, vol. 195. DOI :<https://doi.org/10.1016/j.jmva.2022.105143>.
- [P.4] Inass SOUKARIEH, Salim BOUZEBDA. "*Nonparametric Conditional U-processes for Locally Stationary Random Fields under Stochastic Sampling Design*". **Mathematics**, vol. 11, no 16. DOI: <https://doi.org/10.3390/math11010016>.
- [P.5] Inass SOUKARIEH, Salim BOUZEBDA. "*Conditional U-processes for Locally Stationary Functional Time Series*". Submitted.

Conferences

- [C.1] Inass SOUKARIEH, Salim BOUZEBDA. "*Renewal Type Bootstrap for U- process Markov chains*". The 14th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2021), King's College London, UK, 18-20 December 2021.
- [C.2] Inass SOUKARIEH, Salim BOUZEBDA. "*Regenerative Bootstrap of Harris atomic Markov U -process*". Congrès Jeunes Chercheurs en Mathématiques et leurs Applications(CJC-MA 2022), Calais, France, 21-23 September 2022.

Résumé

Le monde produit 2,5 quintillions d'octets par jour¹, appelés mégadonnées. Le volume, la valeur, la variété, la vitesse et la véracité définissent les cinq caractéristiques du Big Data qui représentent une complexité fondamentale pour de nombreux algorithmes d'apprentissage automatique, tels que le clustering, la reconnaissance d'images et d'autres techniques d'apprentissage modernes. Avec ces données volumineuses, les estimations d'hyperparamètres ne prennent pas la forme de la moyenne de l'échantillon (non linéaire), mais celle de la forme de la moyenne sur m -tuples, appelée l'estimateur U -statistique. Nous considérons dans cette thèse la collection de U -statistiques, connue sous le nom de U -processus, pour deux types de variables dépendantes, les données Markoviennes et les variables aléatoires localement stationnaires. Ainsi, nous avons divisé notre travail en deux parties pour aborder chaque type indépendamment.

Dans la première partie, nous considérons les variables Markoviennes. Nous nous concentrons particulièrement sur les développements de U -processus bootstrappés dans un cadre de Harris. L'idée fondamentale utilisée repose sur les méthodes régénératives consistant essentiellement à diviser l'échantillon en blocs de données indépendants et identiquement distribués (i.i.d.), où chaque bloc correspond à des segments de chemin entre deux visites à un atome appelé A formant une séquence de renouvellement. Nous caractérisons les propriétés limites pour les U -processus indexés par des classes de fonctions uniformément bornées et non bornées. Nous montrons la consistance du bootstrap dans ce cadre. L'approche du bootstrap permet de contourner les problèmes fréquemment rencontrés pour l'évaluation des lois limites dépendants d'une manière complexe de paramètres inconnus. La technique de bootstrap que nous utilisons dans cette thèse est le bootstrap de renouvellement, où l'échantillon bootstrap est formé par rééchantillonnage à partir des blocs. Comme les blocs non bootstrappés sont indépendants, une partie des preuves se réduit au cas i.i.d. Les principales difficultés sont liées à la taille aléatoire des blocs rééchantillonnés, ce qui crée un problème non trivial de temps d'arrêts aléatoires, constituant un des grands obstacles de la généralisation de la théorie dans notre contexte. Pour contourner cette difficulté, nous avons utilisé comme étape intermédiaire la substitution du temps d'arrêt aléatoire par son espérance. La convergence faible des U -processus bootstrappés est très délicate dans notre cadre, en particulier l'équicontinuité

¹<https://financesonline.com/how-much-data-is-created-every-day/>

en utilisant la comparaison avec le U -processus initial. Nous avons étendu les résultats susmentionnés au cas où le degré du U -processus croît avec la taille de l'échantillon n , le noyau variant dans une classe de fonctions. Nous avons caractérisé la convergence faible pour le bootstrap de renouvellement pour le U -processus à degré infini en faisant usage de la technique de découplage combinée avec des techniques de symétrisation. Enfin, nous considérons un bootstrap pondéré échangeable des U -processus empiriques.

Dans la deuxième partie de cette thèse, les données dépendantes sont représentées par des fonctions aléatoires localement stationnaires. Propulsés par la représentation croissante des séries temporelles par des données fonctionnelles ou courbes, et le comportement non stationnaire de ces dernières, nous nous sommes intéressés au U -processus conditionnel des séries temporelles fonctionnelles localement stationnaires. Plus précisément, nous avons étudié la convergence faible des U -processus conditionnels, indexée par des classes de fonctions, dans le cadre de données fonctionnelles localement stationnaires. Nous avons caractérisé la convergence faible dans les deux cas lorsque la classe de fonctions est bornée ou non bornée satisfaisant certaines conditions de moment. Enfin, nous étendons la théorie asymptotique du U -processus conditionnel au champ aléatoire fonctionnel localement stationnaire $\{\mathbf{X}_{s,A_n} : s \in R_n\}$ observés à des emplacements espacés irréguliers dans $R_n = [0, A_n]^d \in \mathbb{R}^d$, et incluant à la fois le domaine croissant pur et le domaine croissant mixte. Nous avons obtenu la convergence faible dans les deux cas lorsque la classe de fonctions est bornée ou non bornée. Ces résultats sont établis sous des conditions structurelles assez générales sur les classes de fonctions et les modèles sous-jacents. La majeure partie de la thèse est motivée par les problèmes d'apprentissage statistiques, y compris parmi beaucoup d'autres, les problèmes de discrimination, l'apprentissage métrique et le classement multipartite.

Mots clés : U -processus, Bootstrap, Classe de fonctions, variables dépendantes, Chain de Markov, Processus régénératif, séries temporelles, donnée fonctionnelle, Stationnarité, Stationnarité locale, Champ aléatoire.

Abstract

The world is producing 2.5 quintillion bytes daily², known as big data. Volume, value, variety, velocity, and veracity define the five characteristics of big data that represent a fundamental complexity for many machine learning algorithms, such as clustering, image recognition, and other modern learning techniques. With this large data, hyperparameter estimations do not take the form of the sample mean (not linear). Instead, they take the form of average over m -tuples, known as the U -statistic estimator in probability and statistics. In this work, we treat the collection of U -statistics, known as the U -process, for two types of dependent variables, the Markovian data, and locally stationary random variables. Thus, we have divided our work into two parts to address each type independently.

In the first part, we deal with Markovian data. The approach relies on regenerative methods, which essentially involve dividing the sample into independent and identically distributed (i.i.d.) blocks of data, where each block corresponds to the path segments between two visits of an atom called A , forming a renewal sequence. We derive the limiting theory for Harris recurrent Markov chain over uniformly bounded and unbounded classes of functions. We show that the results can be generalized also to the bootstrapped U statistics. The bootstrap approach bypasses the problems faced with the asymptotic behavior due to the unknown parameters of limiting distribution. Furthermore, the bootstrap technique we use in this thesis is the renewal bootstrap, where the bootstrap sample is formed by resampling the blocks. Since the non-bootstrapped blocks are independent, most proofs reduce to the i.i.d. case. The main difficulties are related to the random size of the resampled blocks, which creates a problem with random stopping times. This problem is degraded by replacing the random stopping time with their expectation. Also, since we resample from a random number of blocks, and the bootstrap equicontinuity can be verified by comparing with the initial process, the weak convergence of the bootstrap U -process must be treated very carefully. We successfully derive the results in the case of the k -Harris Markov chain. We extend all the above results to the case where the degree of U -statistic grows with the sample size n , with the kernel varying in a class of functions. We provide the uniform limit theory for the renewal bootstrap for the infinite-degree U -process with the help of the decoupling technique combined with symmetrization tech-

²<https://financesonline.com/how-much-data-is-created-every-day/>

niques in addition to the chaining inequality. Remaining in the Markovian setting, we extend the weighted bootstrap empirical processes to a high-dimensional estimation. We consider an exchangeably weighted bootstrap of the general function-indexed empirical U -processes.

In the second part of this thesis, dependent data are represented by locally stationary random variables. Propelled by the increasing representation of the data by functional or curves time series and the non-stationary behavior of the latter, we are interested in the conditional U -process of locally stationary functional time series. More precisely, we investigate the weak convergence of the conditional U -processes in the locally stationary functional mixing data framework. We treat the weak convergence in both cases when the class of functions is bounded or unbounded, satisfying some moment conditions. Finally, we extend the asymptotic theory of conditional U -process to the locally stationary functional random field $\{\mathbf{X}_{s,A_n} : s \in R_n\}$ observed at irregular spaced locations in $R_n = [0, A_n]^d \in \mathbb{R}^d$, and include both pure increasing domain and mixed increasing domain. We treat the weak convergence in both cases when the class of functions is bounded or unbounded, satisfying some moment conditions. These results are established under fairly general structural conditions on the classes of functions and the underlying models. The major part of the thesis is motivated by machine learning problems, including, among many others, discrimination problems, metric learning, and multipartite ranking.

Keywords : U -processes, Bootstrap, Class of functions, dependent variables, Markov chain, Regenerative process, Time series, Functional data, Stationarity, Local stationary time series, Random fields.

Contents

Acknowledgements	ii
Publications	iii
Résumé	iv
Abstract	vi
Contents	vii
1 Introduction	1
1 U -statistics and U -processes	1
1.1 Background	5
1.1.1 Neural network ensemble based U -statistics	5
1.1.2 Maximum Mean Discrepancy	7
1.1.3 Hilbert Schmidt Independence Criterion	8
1.1.4 Kernel Stein Discrepancy	10
2 Markov chain	11
2.1 Motivated examples	12
3 Bootstrap	13
3.1 Bootstrap for Markov chain	15
4 Weak dependence measure, stationarity and locally stationarity	16
5 Functional data	20
6 Examples of applications	23
6.1 Clustering	23
6.2 Testing stochastic monotonicity	24
6.3 Testing curvature and monotonicity of nonparametric regression	25
6.4 Lévy-driven MA random field	26
7 Organisation of the dissertation	27
2 Mathematical Background	40
1 Mathematical Background	41
1.1 Weak convergence	41
1.2 Classes of functions	42
2 Useful notes for studying U -processes	45
2.1 Bootstrapped Empirical processes	46

2.2	The mixing notion	49
3	Locally stationary random field	51
3.1	Random field	51
3	Renewal type bootstrap for increasing degree U-process of a Markov chain	53
1	Introduction	55
2	Preliminaries and notation	58
3	Infinite degree U -statistics and Hoeffding decomposition	60
4	Asymptotic theory: first-order projection	62
4.1	Strong law of large numbers	63
4.2	Weak convergence theorem	63
5	Increasing degree U -processes: asymptotic theory	65
6	Bootstrap of the increasing degree U -processes	67
6.1	Asymptotic validity of bootstrapped Markovian U-statistics	68
7	Examples and Applications	71
8	Mathematical developments	77
8.1	Proof of Theorem 5.1:	83
8.2	Proof of Theorem 5.2:	88
8.3	Proof of Theorem 6.1:	89
4	Renewal type bootstrap for U- process Markov chains	93
1	Introduction	95
2	Preliminaries and notation	98
2.1	Markov chain Assumptions	98
2.2	U -process	101
3	Hoeffding decomposition	103
3.1	Technical Assumptions:	105
3.2	General case	107
3.3	Regenerative approximation	108
4	Asymptotic theory for Markovian U -statistics	108
5	Uniform bootstrap central limit theorem	110
5.1	RBB method	110
5.2	Splitting technique	112
5.3	Bootstrap of U -statistic	112
6	Generalized U -statistics	114
6.1	Decomposition of generalized U -statistics	115
6.2	Regenerative Chains	116
6.3	Asymptotic theory for Markovian generalized U -statistics	117
7	Bootstrapping generalized U -statistics	118
8	Examples	120
8.1	Classes of functions	120
8.2	U -statistics	122
9	Mathematical developments	123
10	Appendix	147
5	Exchangeably weighted bootstraps of the General Markov U-process	151

1	Introduction	153
2	Notation and definitions	156
2.1	Markov chain	156
2.2	Exchangeable weights	159
2.3	The U -process framework	160
2.4	Gaussian chaos Process	161
2.5	Technical Assumptions:	162
3	Preliminary results	163
3.1	The bootstrapped U - Processes	165
4	Weighted Bootstrap weak convergence	168
4.1	Bootstrap weights examples:	171
5	Applications	172
5.1	Symmetry test	172
5.2	Kendall's tau	173
5.3	Test of independence	174
6	Mathematical development	174
6	Weak Convergence of the Conditional U-processes for Locally Stationary Functional Time Series	198
1	Introduction	200
2	Background and preliminaries	204
2.1	Notations	204
2.2	Model	204
2.3	Local stationarity	205
2.4	Mixing conditions	206
2.5	Kernel estimation	207
2.6	Small ball probability	208
2.7	VC-type classes of functions	208
2.8	Assumptions	210
2.9	Comments	213
3	Uniform convergence rates for kernel estimators	215
3.1	Hoeffding's decomposition	215
3.2	Uniform convergence rate	217
4	Weak convergence for kernel estimators	217
5	Applications	219
5.1	Discrimination	219
5.2	Metric learning	220
6	Mathematical Developments	221
7	Nonparametric conditional U-processes for locally stationary random fields under stochastic sampling design	270
1	introduction	272
2	The functional framework	276
2.1	Notation	276
2.2	Generality on the model	277
2.3	Local stationarity	277
2.4	Sampling design	278

2.5	Mixing condition	279
2.6	Generality on the model	281
2.6.1	Small ball probability	282
2.6.2	VC-type classes of functions	282
2.7	Conditions and comments	285
2.7.1	Comments:	288
3	Uniform convergence rates for kernel estimators	290
3.1	Hoeffding's decomposition	290
3.2	Strong uniform convergence rate	292
4	Weak convergence for kernel estimators	293
5	Applications	295
5.1	Metric learning	295
5.2	Multipartite Ranking	296
5.3	Set Indexed Conditional U -Statistics	297
5.4	Discrimination	298
6	Extension to the censored case	299
7	The bandwidth selection criterion	303
8	Mathematical developments	305
8.1	A.1. Preliminaries	306
8	Conclusion and perspectives	352
1	Concluding remarks : Chapter 3	352
2	Concluding remarks: Chapter 4	353
3	Concluding remarks: Chapter 5	354
4	Concluding remarks: Chapter 6	355
5	Concluding remarks: chapter 7	356
	Bibliography	359

Chapter 1

Introduction

1 U -statistics and U -processes

U -statistics arose with Halmos, 1946 where he searched for a class of “*best-unbiased estimate[s]*”, and then formally introduced by Hoeffding, 1948. The idea of Halmos is as follows: Let \mathcal{P} be a class of probability measures $\{P_\theta : \theta \in \Theta\}$, he aimed to construct an optimal, unbiased estimator for θ using $\{X_i\}_{i=1}^n$ a sample of i.i.d real valued random variables with measure \mathbb{P}_θ , for each θ . He found that if there exist a function $h : \mathbb{R}^m \rightarrow \mathbb{R}$ such that \mathbb{P}_θ^k , the k -fold product measure for \mathbb{P}_θ , is equal to θ for every $\theta \in \Theta$, then the minimum-variance unbiased estimator for θ can be written as follows:

$$U_n(h) := \frac{(n-m)!}{n!} \sum_{\mathbf{i} \in I_n^m} h(X_{i_1}, \dots, X_{i_m}), \quad (1.1)$$

where

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

and the function h is called the kernel of the U -statistic $U_n(h)$, and it is said to have degree m . If the kernel h is symmetric (permutation-invariant), the U -statistic formula will be:

$$U_n(h) := \frac{(n-m)!m!}{n!} \sum_{\mathbf{i} \in I_n^m} h(X_{i_1}, \dots, X_{i_m}), \quad (1.2)$$

where

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

in other word, if h is symmetric, $\mathbf{i} \in I_n^m$ is a m -tuple (i_1, \dots, i_m) of m **ordered** and distinct elements of \mathbb{N}^n .

Recently, the explosive data enrichment motivated the researchers to be attracted by the regularized estimation and dimension reduction of high-dimensional data problems, which include the graphical models (e.g. M. Yuan et al., 2007, Bühlmann et al., 2011),

discriminant analysis (see Mai et al., 2012), covariance matrix estimation (e.g. Bickel et al., 2008, X. Chen et al., 2013) and others problems. These issues require the consistent estimation of an expectation of U -statistics of order two. This gives us a sight of the importance of this estimator in modern probability. Let us cite some examples.

Example 1. Consider $g(x, y) = \frac{1}{2}(x - y)^2$. A short calculation shows that the related U -statistic is the well-known variance estimator

$$U_n(g) = \frac{1}{n-1} \sum_{1 \leq i \leq n} (X_i - \bar{X})^2.$$

Example 2. Let $g(x, y) = |x - y|$. Then the corresponding U -statistic is

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

known as Gini's mean difference.

Example 3. Let $d = p \times p$. 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 of form (1.1) with the quadratic kernel, for $x_1, x_2 \in \mathbb{R}^p$,

$$h(x_1, x_2) = (x_1 - x_2)(x_1 - x_2)^\top / 2.$$

Example 4. The covariance matrix quantifies the linear dependency in a random vector. The rank correlation is another measure of the nonlinear dependency in a random vector. 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 $T = \{\tau_{mk}\}_{m,k=1}^p$ is a matrix-valued U -statistic with a bounded kernel. It is clear 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. *Hoeffding's D From the symmetric kernel,*

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

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

Example 6 (Blum-Kiefer-Rosenblatt's R). *The symmetric kernel*

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

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

Example 7. *Bergsma-Dassios-Yanagimoto's τ^* Bergsma et al., 2014 introduced a rank correlation statistic as a U -statistic of order 4 with the symmetric kernel*

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

Here

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

The development of U -statistics and their theory received great importance. It became one of the central places in statistical problems due to their interest characteristic as unbiased estimators of parameters of interest and as components of higher order terms in expansions of smooth statistics like von Mises expansion or delta-method. It is convenient also to consider U -processes indexed by class of kernels \mathcal{H} which is a collection of U -statistics $\{U_n(h) : h \in \mathcal{H}\}$. In general and similar to empirical process theory, the theory of U -processes aims to comprehend the process's behavior uniformly over the index set by combining the pointwise properties of the U -process with attributes of the

index set. Hoeffding, 1961 continues with the theory of U , and he solved one of the main mysteries of the nature of U -statistics, where he offered us what we called the Hoeffding's decomposition, or the H-decomposition, which allows the U -statistic to be written as a sum of uncorrelated terms. We use the H-decomposition heavily in deriving properties of the different order U -statistics and processes generated in this manuscript. This decomposition is based on a sequence of conditional expectations and kernel projections, and it will be presented in detail in the sequel. Up to 1986, only the U -statistics with kernel h belonging to \mathbb{R} is studied. For instance, we have Berk, 1966 who found a reverse martingale structure for U -statistic and after P. K. Sen, 1974a 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 H. Rubin et al., 1980. All these papers, among others like the pioneered work of Serfling, 1980, who has shown that $(nU_n)_{n \in \mathbb{N}^*}$ converges in distribution to a random variable which is written, in terms of the eigenvalues of an operator associated with the kernel $h(\cdot)$, as a weighted sum of Chi-squared independent variables, also he improves the rate of convergence when certain conditional variances are zero, are for $h \in \mathbb{R}$. The first works for h belong to the Hilbert space is devoted to Dehling et al., 1984 and Borovskikh, 1986. We refer to A. J. Lee, 1990 and Koroljuk et al., 1994 books for a detailed review and major historical developments in this field.

Besides, de la Peña, 1992 and de la Peña et al., 1995 mainly discovered the decoupling inequality for U -processes which play an important role in the asymptotic theory of U -statistics and U -processes, followed by others like Arcones et al., 1993, Giné et al., 1994, Arcones et al., 1995. Decoupling is a tool where the decoupled U -statistics can be treated as a sum of independent random variables conditionally on one of the different independent sequences, and it allows us to use maximal inequalities and randomization by Rademacher variables. This makes the analysis of U -processes deal more or less with that of empirical processes. Other important results can be viewed in Arcones and Yu, 1994 where he developed the functional central limit theorems for U -processes under absolute regular observations. Hoffmann-Jorgensen type moment inequalities for entirely degenerate U -statistics of any order m were found by Giné et al., 1992. Klass et al., 1997 provided moment inequalities for non-negative generalised U -statistics of order $m = 2$ as well. There is also established more recent research on both U -statistics and U -processes, such as Adamczak, 2006 who extended the results of Giné et al., 2000 to canonical U -statistics of arbitrary order. Using an estimator given by Frees, 1994, Giné et al., 2007a proposed a local U -statistic process and demonstrated its central limit theorems for several norms, including the sup and the L_p -norms (for $p \geq 1$). Others investigated the fields of application of U -statistics; for example, Cléménçon et al., 2008 used the U -statistics in the ranking problems and clustering Cléménçon, 2014. A natural occurrence of U -statistics also took place in learning on graphs Biau et al., 2006 or in expansions of smooth statistics

Robins et al., 2009 as components of higher-order terms. In 2016, Gábor and Joly published an article Joly et al., 2016 named Robust estimation of U -statistics, in which they investigate the estimation of the mean of multivariate functions in some specific cases.

1.1 Background

In this part, we will give some examples of the application of U -statistics that explain our interest in this estimator.

1.1.1 Neural network ensemble based U -statistics

This example is considered by Schupbach et al., 2020. In their work, they aimed to estimate uncertainty in ensemble regressors to the case of neural network ensembles using the U -statistics and to prove that neural networks fit within this U -statistics framework, especially infinite-degree U -statistics framework. In order to understand the situation, we will briefly define each term. Neural network ensembles are used to improve predictive power and specifically to train many networks with different hyperparameter combinations while only using the network that performs the best on the validation set. An ensemble neural network method is Negative correlation learning, where an ensemble of neural networks is trained simultaneously with a loss function that contains a penalty for the correlation between the networks. The authors used the U -statistics to describe the resampling technique used in this chapter which is subsample aggregating (subbagging). Consider a random sample $(\mathbf{X}, \mathbf{Y}) \stackrel{\text{iid}}{\sim} F$ of size n , a neural network N from a subbagged sample size m taken from the dataset is built. The same strategy is done for all $\binom{n}{m}$ subsamples. Then, the predictions average for some x^* from these neural networks can be taken as an estimate of our predicted value. This average is written as follows

$$b(\mathbf{x}^*) = \frac{1}{\binom{n}{m}} \sum_{(i)} N_{\mathbf{x}^*}((X_{i_1}, Y_{i_1}), \dots, (X_{i_m}, Y_{i_m})).$$

Given some regularity conditions-unbiased and permutation symmetric, we have a procedure that results in a U -statistic for these predicted values. N has been shown to be an unbiased estimator that is asymptotically consistent and, if trained by the batch update, is permutation symmetric. N can be used as an estimator, resulting in a U -statistic for the predicted values. Unfortunately, it is generally computationally infeasible to build neural networks for all $\binom{n}{m}$ subsamples of the data. It has been shown that taking

$b \leq \binom{n}{m}$ subsamples of size m results in

$$b_b(\mathbf{x}^*) = \frac{1}{b} \sum_{(i)} N_{\mathbf{x}^*}((X_{i_1}, Y_{i_1}), \dots, (X_{i_m}, Y_{i_m})),$$

which is an incomplete U -statistic. Even so, this has been shown to be asymptotically normal and unbiased by Janson, 1984, assuming the variance of the estimator converges to zero at a rate faster than \sqrt{n} . Neural Network N has been shown to be a mean integrated squared error (MISE) consistent, as

$$MISE = O\left(\frac{C_f^2}{s}\right) + O\left(\frac{sd}{n} \log n\right),$$

where C_f is a constant related to smoothing from the training process, s is the number of nodes in the network, d is the number of covariates, and n is the sample size. It may also make sense to scale b with n . Specifically, considering subsamples of size $b_n = \binom{n}{m_n}$ give

$$b_{b_n, m_n}(\mathbf{x}^*) = \frac{1}{b_n} \sum_{(i)} N_{\mathbf{x}^*}((X_{i_1}, Y_{i_1}), \dots, (X_{i_{m_n}}, Y_{i_{m_n}})),$$

which is an infinite order U -statistic or a resampled statistic when $b_n \neq \binom{n}{m_n}$. Frees, 1994 developed necessary and sufficient conditions for asymptotic normality when b_n grows faster than n , and Mentch et al., 2016 developed conditions for individual means $\mathbb{E}[b(\mathbf{x}^*)]$ for all growth rates of b_n with respect to n . So as long as the estimates for a bounded regression function are bounded, the variance of the kernel function ϕ is bounded, $\lim \frac{n}{b_n} = \alpha$, $\lim \frac{m_n}{\sqrt{n}} = 0$, and $\lim \sigma_{1, m_n} \neq 0$, then the infinite order U -statistic will be asymptotically normal with the following distributions given in Mentch et al., 2016:

- if $\alpha = 0$, then

$$\frac{\sqrt{n}(U_{n, b_n, m_n} - \theta_{b_n})}{\sqrt{b_n^2 \zeta_{1, m_n}}} \xrightarrow{d} \mathcal{N}(0, 1);$$

- if $0 < \alpha < \infty$, then

$$\frac{\sqrt{b_n}(U_{n, b_n, m_n} - \theta_{m_n})}{\sqrt{\frac{m_n^2}{\alpha} \zeta_{1, b_n} + \zeta_{m_n, b_n}}} \xrightarrow{d} \mathcal{N}(0, 1);$$

- if $\alpha = \infty$, then

$$\frac{\sqrt{b_n}(U_{n, b_n, m_n} - \theta_{m_n})}{\sqrt{\zeta_{b_n, m_n}}} \xrightarrow{d} \mathcal{N}(0, 1).$$

m_n is chosen approximately on the order of \sqrt{n} . This choice of m_n replaces the require-

ment of exponential tails on the error distribution with the requirement that $n\mathbb{P}(|\varepsilon| > \sqrt{n}) \rightarrow 0$. It also assures that the $\lim_{n \rightarrow \infty} \frac{m_n}{\sqrt{n}} = 0$. Finally, by choosing a small m_n , the time complexity is similar to a bootstrap method while generating large ensembles. Note that it is not required to choose m_n on the order of \sqrt{n} . The subbagging approach presupposes that the estimator is constructed in the same way. Outside of the weak regularity constraints, the distributional findings do not rely on the way of generating the neural network; nonetheless, the subbagging approach requires that each neural network be formed using the same method. This would rule out the use of dropout since each estimator would be constructed from randomly selected samples of nodes on each training period. The estimator would therefore require a new justification, similar to the extension of U -statistics to random forests.

1.1.2 Maximum Mean Discrepancy

The following example is treated in Schrab et al., 2022 and Kim et al., 2022. Let $\{X_i\}_{i=1}^{n_1}$ be a sequence of i.i.d random variables with probability density function \mathbb{P} and $\{Y_j\}_{j=1}^{n_2}$ be a sequence of i.i.d random variables with probability density function \mathbb{Q} , such that $\{X_i\}_{i=1}^{n_1}$ is independent from $\{Y_j\}_{j=1}^{n_2}$ and the sample sizes n_1 and n_2 are balanced, i.e., there exists a constant C such that $\max(n_1, n_2) \leq C \min(n_1, n_2)$. The goal is to test if the two samples have the same distribution or not, then the hypothesis is as follows:

$$\begin{cases} H_0 : \mathbb{P} = \mathbb{Q}, \\ H_1 : \mathbb{P} \neq \mathbb{Q}. \end{cases}$$

In order to test this hypothesis, a non-parametric kernel test has been introduced by Gretton et al., 2012, based on a measure between probability densities \mathbb{P} and \mathbb{Q} on \mathbb{R}^d called the Maximum Mean Discrepancy (MMD). This measure is the integral probability metric (IPM; Müller, 1997) over a reproducing kernel Hilbert space \mathcal{H}_K (RKHS; Aronszajn, 1950) with associated kernel K . The MMD is defined as the \mathcal{H}_K -norm of the difference between the mean embeddings $\mu_{\mathbb{P}}(u) := \mathbb{E}_{X \sim \mathbb{P}}[K(X, u)]$ and $\mu_{\mathbb{Q}}(u) := \mathbb{E}_{Y \sim \mathbb{Q}}[K(Y, u)]$ for $u \in \mathbb{R}^d$ (see Gretton et al., 2012, Lemma 4). For X' and Y' are independent copies of X and Y respectively, Define

$$\begin{aligned} \text{MMD}_K^2(\mathbb{P}, \mathbb{Q}) &:= \left(\sup_{f \in \mathcal{H}_K: \|f\|_{\mathcal{H}_K} \leq 1} |\mathbb{E}_{\mathbb{P}}[f(X)] - \mathbb{E}_{\mathbb{Q}}[f(Y)]| \right)^2 \\ &= \|\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}\|_{\mathcal{H}_K}^2 \\ &= \mathbb{E}_{\mathbb{P}, \mathbb{P}}[K(X, X')] - 2\mathbb{E}_{\mathbb{P}, \mathbb{Q}}[K(X, Y)] + \mathbb{E}_{\mathbb{Q}, \mathbb{Q}}[K(Y, Y')]. \end{aligned}$$

In Fukumizu et al., 2007; Sriperumbudur et al., 2011, there is a essential property that

guarantees that $\text{MMD}_K^2(\mathbb{P}, \mathbb{Q}) = 0$ if and only if $\mathbb{P} = \mathbb{Q}$. This property provides the use of MMD as a two-sample test and leads Gretton et al., 2012 (Lemma 6) to estimate the parameter of interest MMD, using the following unbiased quadratic-time MMD estimator:

$$\begin{aligned} \widehat{\text{MMD}}_K^2(X_{n_1}, Y_{n_2}) &= \frac{1}{n_1(n_1 - 1)} \sum_{(i, i') \in I_{n_1}^2} K(X_i, X_{i'}) - \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} K(X_i, Y_j) \\ &\quad + \frac{1}{n_2(n_2 - 1)} \sum_{(j, j') \in I_{n_2}^2} K(Y_j, Y_{j'}) \\ &= \frac{\mathbf{1}^\top \widetilde{\mathbf{K}}_{XX} \mathbf{1}}{n_1(n_1 - 1)} - 2 \frac{\mathbf{1}^\top \mathbf{K}_{XY} \mathbf{1}}{n_1 n_2} + \frac{\mathbf{1}^\top \widetilde{\mathbf{K}}_{YY} \mathbf{1}}{n_2(n_2 - 1)}, \end{aligned}$$

where $\widetilde{\mathbf{K}}_{XX}$ and $\widetilde{\mathbf{K}}_{YY}$ are the kernel matrices $\mathbf{K}_{XX} := (K(X_i, X_{i'}))_{1 \leq i, i' \leq n_1}$ and $\mathbf{K}_{YY} := (K(Y_j, Y_{j'}))_{1 \leq j, j' \leq n_2}$ with diagonal entries set to 0, where $\mathbf{K}_{XY} := (K(X_i, Y_j))_{1 \leq i \leq n_1, 1 \leq j \leq n_2}$, and where $\mathbf{1}$ is a one-dimensional vector with all entries equal to 1 of variable length determined by the context. Based on the bivariate symmetric kernel K , define the following kernel:

$$h_K^{\text{MMD}}(x_1, x_2; y_1, y_2) := K(x_1, x_2) - K(x_1, y_2) - K(x_2, y_1) + K(y_1, y_2), \quad (1.3)$$

and write the MMD estimator $\widehat{\text{MMD}}_K^2(X_{n_1}, Y_{n_2})$ as a two-sample U -statistic (h_K^{MMD} is not symmetric) as follows:

$$\widehat{\text{MMD}}_K^2(X_{n_1}, Y_{n_2}) := \frac{(n_1 - 2)! (n_2 - 2)!}{n_1! n_2!} \sum_{(i, i') \in I_{n_1}^2} \sum_{(j, j') \in I_{n_2}^2} h_K^{\text{MMD}}(X_i, X_{i'}; Y_j, Y_{j'}).$$

By the unbiased property of U -statistics, the symmetric form of $\widehat{\text{MMD}}_K^2(X_{n_1}, Y_{n_2})$, defined by symmetrization of the kernel $h_K^{\text{MMD}}(x_1, x_2; y_1, y_2)$, is an unbiased estimator of $\mathbb{E}(h_K^{\text{MMD}}(x_1, x_2; y_1, y_2))$. Kim et al., 2022 studied the properties of this estimator based on the treatment of error Type *I* and error Type *II*. In contrast, Schrab et al., 2022 replace the quadratic-time U -statistic $\widehat{\text{MMD}}_K^2(X_{n_1}, Y_{n_2})$ with a second order incomplete U -statistic and then they provide quantile and variance bounds using a wild bootstrap.

1.1.3 Hilbert Schmidt Independence Criterion

Let \mathbb{P}_{XY} denote the joint distribution of i.i.d pairs of samples $\{X_i\}_{i \in N}$ and $\{Y_i\}_{i \in N}$, and let $\mathbb{P}_X \mathbb{P}_Y$ the product of their marginal distributions. If we observe the pairs $Z_i := (X_i, Y_i) \stackrel{i.i.d}{\sim}$

\mathbb{P}_{XY} , the hypotheses for testing independence are:

$$\begin{cases} H_0 : \mathbb{P}_{XY} = \mathbb{P}_X \mathbb{P}_Y, \\ H_1 : \mathbb{P}_{XY} \neq \mathbb{P}_X \mathbb{P}_Y. \end{cases}$$

To test these hypotheses, a non-parametric kernel test based on the Hilbert Schmidt Independence Criterion (HSIC) is given by Gretton et al., 2007; Gretton et al., 2005. We will explain this test in the following. Define the HSIC as:

$$\begin{aligned} \text{HSIC}_{K,L}(p_{xy}) &:= \text{MMD}_K^2(\mathbb{P}_{XY}, \mathbb{P}_X \mathbb{P}_Y) \\ &= \mathbb{E}_{\mathbb{P}_{XY}, \mathbb{P}_{XY}} [K(X, X') L(Y, Y')] \\ &\quad - 2\mathbb{E}_{\mathbb{P}_{XY}} [\mathbb{E}_{\mathbb{P}_X} [K(X, X')] \mathbb{E}_{\mathbb{P}_Y} [L(Y, Y')]] \\ &\quad + \mathbb{E}_{\mathbb{P}_X, \mathbb{P}_X} [K(X, X')] \mathbb{E}_{\mathbb{P}_Y, \mathbb{P}_Y} [L(Y, Y')], \end{aligned}$$

with K and L are two bivariate kernels, symmetric in their arguments, and define the product kernel associated with $K(x_1, x_2)$ and $L(y_1, y_2)$ by

$$\begin{aligned} h_{K,L}^{\text{HSIC}} \{ &(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4) \} \\ &:= \{K(x_1, x_2) + K(x_3, x_4) - K(x_1, x_3) - K(x_2, x_4)\} \\ &\quad \times \{L(y_1, y_2) + L(y_3, y_4) - L(y_1, y_3) - L(y_2, y_4)\}. \end{aligned}$$

Clearly, using the definition of h_K^{MMD} given in Equation (1.3), we have

$$h_{K,L}^{\text{HSIC}}(z_1, z_2, z_3, z_4) := \frac{1}{4} h_K^{\text{MMD}}(x_1, x_2; x_3, x_4) h_L^{\text{MMD}}(y_1, y_2; y_3, y_4).$$

Based on this definition, the estimator of HSIC is a fourth-order U -statistic defined by

$$\begin{aligned} \widehat{\text{HSIC}}_{K,L}(Z) &:= \frac{(N-4)!}{N!} \sum_{(i,j,r,s) \in I_N^4} K(X_i, X_j) L(Y_r, Y_s) \\ &= \frac{(N-4)!}{N!} \sum_{(i,j,r,s) \in I_N^4} h_{K,L}^{\text{HSIC}}(Z_i, Z_j, Z_r, Z_s). \end{aligned}$$

This estimator is treated in Gretton et al., 2007, and in L. Song et al., 2012 where they provided the following closed-form expression

$$\widehat{\text{HSIC}}_{K,L}(Z_N) = \frac{1}{N(N-3)} \left(\text{tr}(\widetilde{\mathbf{K}}\widetilde{\mathbf{L}}) + \frac{\mathbf{1}^\top \widetilde{\mathbf{K}} \mathbf{1} \mathbf{1}^\top \widetilde{\mathbf{L}} \mathbf{1}}{(N-1)(N-2)} - \frac{2}{N-2} \mathbf{1}^\top \widetilde{\mathbf{K}} \widetilde{\mathbf{L}} \mathbf{1} \right),$$

where $\widetilde{\mathbf{K}}$ and $\widetilde{\mathbf{L}}$ are the kernel matrices $\mathbf{K} := (K(X_i, X_j))_{1 \leq i, j \leq N}$ and $\mathbf{L} := (L(Y_i, Y_j))_{1 \leq i, j \leq N}$ with diagonal entries set to 0, and $\mathbf{1}$ is a one-dimensional vector with all entries equal to

1 of variable length determined by the context. After symmetrizing the kernel h_k^{HISC} with its symmetrized variant

$$\bar{h}_{K,L}^{\text{HISC}}(z_1, z_2, z_3, z_4) := \frac{1}{4!} \sum_{(i_1, i_2, i_3, i_4) \in I_4^4} h_{K,L}^{\text{HISC}}(z_{i_1}, z_{i_2}, z_{i_3}, z_{i_4}),$$

Kim et al., 2022 give the type II error of the permutation test based on the unbiased estimator $\widehat{\text{HSIC}}_{K,L}$.

1.1.4 Kernel Stein Discrepancy

Remaining in the kernels test settings, we will give in this example a non-parametric goodness-of-fit test extracted from the *Kernel Stein Discrepancy* (KSD). It is good to mention here that a review of classical goodness-of-fit tests can be found in Lehmann et al., 2005, where most methods have computational difficulty for unnormalized distributions. The studied test in this example was introduced by Chwialkowski et al., 2016; Q. Liu et al., 2016 where they searched for writing the maximum discrepancy between the target distribution \mathbb{P} and the observed sample distribution \mathbb{Q} in a modified Reproducing kernel Hilbert space RKHS. The technique of combining Stein's identity with RKHS was first developed by Oates et al., 2017 for variance reduction. The test can be described as follows. Let \mathbb{P} be a given model density on \mathbb{R}^d , and let $\{X_i\}_{i \in n}$ an i.i.d sample drawn from a density \mathbb{Q} on \mathbb{R}^d , the goal is to test the hypotheses:

$$\begin{cases} H_0 : \mathbb{P} = \mathbb{Q}, \\ H_1 : \mathbb{P} \neq \mathbb{Q}. \end{cases}$$

The Kernel Stein Discrepancy (KSD) can be defined as

$$\begin{aligned} \text{KSD}_{\mathbb{P},K}^2(\mathbb{Q}) &:= \text{MMD}_{h_{K,\mathbb{P}}^{\text{KSD}}}^2(\mathbb{Q}, \mathbb{P}) \\ &= \mathbb{E}_{\mathbb{Q},\mathbb{Q}} \left[h_{K,\mathbb{P}}^{\text{KSD}}(Z, Z') \right] - 2\mathbb{E}_{\mathbb{Q},\mathbb{P}} \left[h_{K,\mathbb{P}}^{\text{KSD}}(Z, X) \right] + \mathbb{E}_{\mathbb{P},\mathbb{P}} \left[h_{K,\mathbb{P}}^{\text{KSD}}(X, X') \right] \\ &= \mathbb{E}_{\mathbb{Q},\mathbb{Q}} \left[h_{K,\mathbb{P}}^{\text{KSD}}(Z, Z') \right], \end{aligned}$$

where X is a random variable distributed according to the target measure \mathbb{P} , and, for all $x, y \in \mathbb{R}^d$,

$$\begin{aligned} h_{K,\mathbb{P}}^{\text{KSD}}(x, y) &:= \left(\nabla \log \mathbb{P}(x)^\top \nabla \log \mathbb{P}(y) \right) K(x, y) + \nabla \log \mathbb{P}(y)^\top \nabla_x K(x, y) \\ &\quad + \nabla \log \mathbb{P}(x)^\top \nabla_y K(x, y) + \sum_{i=1}^d \frac{\partial}{\partial x_i \partial y_i} K(x, y). \end{aligned}$$

For X from the target measure, the Stein kernel satisfies the Stein identity

$$\mathbb{E}_{\mathbb{P}} \left[h_{K,\mathbb{P}}^{\text{KSD}}(X, \cdot) \right] = 0,$$

for instance, see Chwialkowski et al., 2016, Lemma 5.1, which is a desirable property. It is straightforward to estimate the squared Stein discrepancy $\text{KSD}_{\mathbb{P},K}^2(\mathbb{Q})$ from samples $\{X_i\}_{i \in n}$, Chwialkowski et al., 2016 used a V -Statistic as a quadratic time estimator, while it can also be computed as the second-order U -statistic:

$$\widehat{\text{KSD}}_{\mathbb{P},K}^2(X) := \frac{(n-2)!}{n!} \sum_{(i,j) \in I_n^2} h_{K,\mathbb{P}}^{\text{KSD}}(X_i, X_j) = \frac{\mathbf{1}^\top \bar{\mathbf{H}} \mathbf{1}}{n(n-1)},$$

where $\bar{\mathbf{H}}$ is the kernel matrix $\mathbf{H} := \left(h_{K,\mathbb{P}}^{\text{KSD}}(X_i, X_j) \right)_{1 \leq i, j \leq N}$ with diagonal entries set to 0? and $\mathbf{1}$ is a one-dimensional vector with all entries equal to 1 of variable length determined by the context. The Stein kernel h_K^{KSD} is already symmetric.

2 Markov chain

Stochastic processes are mathematical models of random phenomena that change over time or space. Due to this notion of randomness, this type of process has a diverse type of applications (telecommunication, finance, internet, supply chains, medicine, energy, etc...), and has different classes like Poisson processes, Brownian motion, regenerative processes, etc... The focus of our work is on Markov chain, one of the principal classes of stochastic process, commonly used to model random dynamical systems, storage, queuing models, signal processing, control theory, and much more applications. A Markov chain is a sequence of random variables X_0, X_1, \dots with values in a countable set S if at any time n , the future states X_{n+1}, X_{n+2}, \dots depend on the history X_0, X_1, \dots, X_n only through the present state X_n . So, we can formally define a Markov chain as follows:

Definition 2.1. A stochastic process $X = \{X_n : n \geq 0\}$ on a countable set S is a Markov Chain if, for any $i, j \in S$ and $n \geq 0$,

$$\mathbb{P}\{X_{n+1} = j \mid X_0, \dots, X_n\} = \mathbb{P}\{X_{n+1} = j \mid X_n\}, \quad (2.1)$$

and

$$\mathbb{P}\{X_{n+1} = j \mid X_n = i\} = p_{ij}. \quad (2.2)$$

where p_{ij} is the probability that the Markov chain passes from state i to state j . These transition probabilities satisfy

$$\sum_{j \in S} p_{ij} = 1, i \in S,$$

and the matrix $P = (p_{ij})$ is the transition matrix of the chain.

Equation (2.1) is called the Markov property, where we can understand the dependence structure of the Markov model, which says that the next state is in the past and present only through the present state. Equation (2.2) shows that the Markov chain is time-homogeneous since the transition probabilities do not depend on the time parameter n . If not, the process X_n will be called a non-time-homogeneous Markov chain.

2.1 Motivated examples

Example 8 (Random walks). Let $\{Z_n, n \in \mathbb{N}^*\}$ be a sequence of i.i.d. random variables with values in $X = \mathbb{R}^d$ and distribution μ . Let X_0 be a random variable in \mathbb{R}^d independent of $\{Z_n, n \in \mathbb{N}^*\}$. A random walk with jump or increment distribution μ is a process $\{X_k, k \in \mathbb{N}\}$ defined by X_0 and the recurrence

$$X_k = X_{k-1} + Z_k, \quad k \geq 1.$$

This model follows a recurrence with $f(x, z) = x + z$, and thus the process $\{X_k, k \in \mathbb{N}\}$ is a Markov chain with kernel given for $x \in \mathbb{R}^d$ and $A \in \mathcal{B}(\mathbb{R}^d)$ by $P(x, A) = \mu(A - x)$; that is, P is entirely determined by the increment distribution μ .

Example 9 (ARMA(p,q)). A generalization of the AR(p) model is obtained by adding a moving average part to the autoregression:

$$X_k = \mu + \alpha_1 X_{k-1} + \dots + \alpha_p X_{k-p} + Z_k + \beta_1 Z_{k-1} + \dots + \beta_q Z_{k-q},$$

where $\{Z_k, k \in \mathbb{Z}\}$ is a sequence of i.i.d. random variables with $\mathbb{E}[Z_0] = 0$. This yields a

Markov chain of order $r = p \vee q$. Indeed, setting $\alpha_j = 0$ if $j > p$ and $\beta_j = 0$ if $j > q$ yields

$$\begin{pmatrix} X_{k+1} \\ \vdots \\ X_{k+r} \end{pmatrix} = \begin{pmatrix} 0 & 1 & \cdots & & \\ \vdots & 0 & 1 & \cdots & \\ \vdots & \ddots & \ddots & & \\ 0 & \cdots & 0 & 1 & \\ \alpha_r & & \cdots & \alpha_1 & \end{pmatrix} \begin{pmatrix} X_k \\ \vdots \\ X_{k+r-1} \end{pmatrix} + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \mu + Z_k + \beta_1 Z_{k-1} + \cdots + \beta_r Z_r \end{pmatrix}.$$

Example 10 (ARCH (p)). *Many financial time series, such as log-returns of share prices, stock indexes, and exchange rates, are commonly employed in econometrics and applied financial literature to illustrate stochastic volatility and heavy-tailedness. A linear time series model cannot accurately catch these characteristics simultaneously. To capture these latters, nonlinear models were devised, because a linear time series model requires heavy-tailed marginal distributions. The input noise sequence must also be heavy-tailed. Heavy-tailed marginals can be constructed for nonlinear models even when the system is injected with a light-tailed input, such as normal noise. Taking the autoregressive conditional heteroscedastic model of order p into account, the ARCH(p) model is defined as a solution to the recurrence problem:*

$$\begin{aligned} X_k &= \sigma_k Z_k, \\ \sigma_k^2 &= \alpha_0 + \alpha_1 X_{k-1}^2 + \cdots + \alpha_p X_{k-p}^2, \end{aligned}$$

where the coefficients $\alpha_j \geq 0, j \in \{0, \dots, p\}$, are nonnegative, and $\{Z_k, k \in \mathbb{Z}\}$ is a sequence of i.i.d. random variable with zero mean (often assumed to be standard Gaussian). The ARCH(p) process is a Markov chain of order p . Assume that Z_1 has a density g with respect to Lebesgue measure on \mathbb{R} .

3 Bootstrap

Let X_1, \dots, X_n be a sequence of random variables with joint probability \mathbb{P} , and let θ be our parameter of interest depending on the unknown joint probability of the underlying sequence. Statistical inference constantly searches to find an estimator to this parameter, and many standard methods were introduced to getting this target depending on the form of parameter θ , like maximum or quasi-Likelihood, M-estimators, kernels estimators, etc. For $\hat{\theta}_n$ an estimator of the parameter θ ; it is crucial to test the estimator and verify the accuracy of the estimator. As the joint probability of X_1, \dots, X_n is unknown, as well as the sampling distribution of the centred estimator $\hat{\theta}_n - \theta$, many quantities related to the inference of the estimator, like the quantiles and the mean square error MSE are

unknown¹. The resampling techniques and bootstrap were general methods to find the latest quantities.

Bootstrapping techniques have been extensively used to handle a wide range of issues. It tries to restore the relationship between the “**population**” and the “**sample**”, by treating the sample as the podium of the underlying population and appropriately resampling from it to produce the “bootstrap sample”, which acts as an analog of the original sample. If the resampling technique is properly designed, the “resample” and the current sample are supposed to represent the initial relationship between the population and the sample. Thus, to address the statistical inference problems related to the unknown distribution, statisticians use the “**sample**” and the “**resamples**”, instead of the “population” and the “sample”, which are either known or have known distributions.

Following such extensive attention, various methodological improvements have arisen to enhance the original bootstrap approach proposed by Efron, 1979, even if the core concept remains intact (e.g., see the methodological discussion on the classical bootstrap methods in D. A. Freedman, 1984; Freedman et al., 1984; Efron et al., 1986; Efron et al., 1993). The bootstrap involves resampling observations to earn a decent estimate of the statistical features of the original population. The constraint that the observations in the sample are realizations of independent and identically distributed random variables is a significant limitation of Efron’s bootstrap approaches. However, in the case of real-life time series, this condition needs to be more accurate. When such a hypothesis is rejected, a theoretical model for the data is required, and the bootstrap method is used for the model errors. Other bootstrap techniques have been presented to overcome this weakness of Efron’s bootstrap and avoid model misspecification. We can cite the *block*, *sieve*, and *local* methods of bootstrapping. The *block* bootstrap was introduced by Künsch, 1989, where he tried to extend the standard Jackknife and bootstrap for a general stationary process without resorting to reduction to the i.i.d case, while the *sieve* bootstrap is an alternative of block bootstrap procedure, for the stationary categorical time series. It is worth mentioning here that there are several varieties of the block bootstrap method; we can cite Politis et al., 1992, 1994b for the blocks-of-blocks bootstrap and the stationary bootstrap respectively, and Paparoditis et al., 2001b, 2002b for trapped block bootstrap. Nevertheless, they only partially achieved their objective (keeping the dependence structure of the data) because they faced the loss of dependency among blocks. Indeed, Bühlmann, 2002 gives us a good comparison between the previously mentioned methods. The *local* bootstrap procedure follows a general autoregressive structure and involves resampling from a neighborhood of each data point. The advantage of these methods is that they are fully data-driven, in other words, the dependence structure of

¹Parameters like θ are called level-1 parameters, while parameters like the mean square error are level-2 parameters. Bootstrap, in general, was introduced to fix the problems of higher-level parameters (2-level and higher)

data can be easily captured, controlling the risk of incorrectly describing the model and the problems of estimating its parameters.

3.1 Bootstrap for Markov chain

Different contributions to resampling approaches based on Markov chain theory can be grouped. Earlier methods for bootstrapping Markov chains were developed by Athreya et al., 1992b; Basawa et al., 1990; Datta et al., 1992; Kulperger et al., 1989 and Kulperger, 1999, where they focused on the transition probabilities of a stationary Markov chain with a distinction between the different strategies studied. Furthermore, Bühlmann, 2002 developed the idea of a sieve bootstrap, introduced by himself in Bühlmann, 1997, where he searches to fit Markovian models to a data series and resampling randomly from the residuals. He proves that this bootstrap technique has a faster convergence rate for variance estimation than the more general block bootstrap. This strategy searches for relevant pathways in a hierarchical manner, which might be a significant drawback when time dependency is not monotonically decreasing.

As well, Rajarshi, 1990 and Horowitz, 2003 embrace the so-called *local bootstrap* techniques for Markov processes invented in Paparoditis et al., 2001a, 2002a, where they estimate the conditional density by kernel methods and generate bootstrap data by successively sampling observations according to the estimated density.

Another collection of works is included in the works of Andrews, 2005; Csiszár et al., 2000; Finesso, 1992; Kieffer, 1993; C.-C. Liu et al., 1994; Merhav et al., 1989. This collection investigates the problem of estimating the order of a Markov chain, assuming that all states are relevant for delays up to the estimated order at all times. However, in certain situations, a reasonable estimate of the relevant states is more significant than an exact estimate of the process’s “memory”. For instance, in economics and finance, we refer to the bootstrapping of series with regimes that characterize the dynamics of various processes (such as traded volumes in stock markets and prices in commodity markets).

Furthermore, *Regenerative block bootstrap* approach was introduced by Athreya et al., 1992b and Datta et al., 1993 and have been developed by Radulović, 2004 and then in Bertail et al., 2011a; Bertail et al., 2006b. We adopted this method in our work, and it will be more developed in the framework of U -processes in the following of this manuscript.

On the other hand, we explore in Chapter 5 the new possibility brought by considering bootstrap weights other than Efron’s multinomial weights for U -processes Markov chain. The non-parametric bootstrap technique has been extended to estimate the posterior distribution for some statistics. The idea is to explore the possibility of considering bootstrap weights other than Efron’s multinomial weights. This general resampling scheme was first introduced by D. B. Rubin, 1981, and extensively studied by Barbe et al., 1995, who suggested the name “*Weighted bootstrap*” and in Mason et al., 1992 and Præstgaard et al.,

1993, who showed that, for a large class of exchangeable weights, the bootstrap empirical processes are asymptotically validated both in probability and almost surely sense. Other versions of Efron’s bootstrap are also studied in Chatterjee et al., 2005 using the term “*Generalized bootstrap*”.

4 Weak dependence measure, stationarity and locally stationarity

This section is devoted to introducing other mathematical concepts used in this thesis. Preferably, we will introduce different notions of weak dependence, stationarity, and local stationarity. We proceed with a basic introduction to functional data and finish with a short recap of important results on weak convergence in metric spaces.

Historically, the independent random variables were frequently studied, and many results in probability theory and statistics (like the central limit theorem) were first shown for such variables, see Liapounoff, 1900; Lindeberg, 1922, among others. Then, a generalization of the central limit theorem to time series was attempted, i.e, a generalization to dependent random variables indexed in discrete time (see, e.g., Bernstein, 1927). In this setting, the central limit theorem is based on the so-called strong mixing condition (α -mixing condition) by Rosenblatt, 1956 was born, and it represents a significant breakthrough in the analysis of dependent variables. Other mixing conditions, such as β -mixing by Volkonskiui et al., 1959 or φ -mixing by Blum et al., 1963, were later suggested and frequently utilized for the analysis of dependent data; see, for example, Doukhan, 2012 for a review of different mixing approaches.

Despite the attractive theoretical properties of the strong mixing condition, two significant downsides appear. First, it is not easy to be verified in reality (see, for instance, Vogt, 2012), and second, there are significant groups of linear processes that do not meet the strong mixing condition (see, e. g., Andrews, 1984). Wu, 2005 provided another technique for weak dependence, the so-called physical dependency measure, which is simple to compute and proves to be a useful tool. Both techniques, strong mixing and physical dependency, suggest that, as the distance between two random variables becomes bigger, their covariance disappears. From this idea, Brillinger, 1981 utilized constraints on cumulants, an extension of covariance, to ensure that a given time series is weakly dependent. Strong mixing implies Brillinger’s cumulant condition under particular moment assumptions (see, for example, Statulevicius et al., 1988 and Doukhan et al., 1989). Following that, we define the dependent measures.

Definition 4.1 (Strong Mixing Condition). Let \mathcal{F} and \mathcal{G} be σ -fields on a probability space (Ω, \mathcal{A}) and define their mixing coefficient as

$$\alpha(\mathcal{F}, \mathcal{G}) = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{F}, B \in \mathcal{G}\}.$$

A stationary time series $(X_t)_{t \in \mathbb{Z}}$ satisfies the strong mixing condition or is simply called strongly mixing if the mixing coefficients:

$$\alpha(k) = \sup_{t \in \mathbb{Z}} \alpha\left(\sigma\left\{(X_s)_{s=-\infty}^t\right\}, \sigma\left\{(X_s)_{s=t+k}^{\infty}\right\}\right),$$

vanish as k tends to infinity, where $\sigma(\mathcal{B})$ denotes the σ -field generated by a family of random variables \mathcal{B} .

Definition 4.2 (Cumulant Condition). A stationary time series $(X_t)_{t \in \mathbb{Z}}$ satisfies the cumulant^a condition for some $k \in \mathbb{N}$, if

$$\sum_{t_1, \dots, t_{k-1} = -\infty}^{\infty} |\text{cum}(X_{t_1}, \dots, X_{t_k})| \leq C < \infty,$$

for some constant $C \in \mathbb{R}$.

^a **Cumulants:**[Doukhan et al., 1989] Let $\{X_1, \dots, X_k\}$ be centered real-valued random variables. Let $\mathbf{X} = (X_1, \dots, X_k)$ and $m_{\mathbf{v}} = \mathbb{E}X_1^{v_1} \dots X_k^{v_k}$ if $\mathbf{v} = (v_1, \dots, v_k) \in \mathbb{N}^k$. If $\varphi(t) = \mathbb{E}e^{it\mathbf{X}}$, $t \in \mathbb{R}^k$, is the characteristic function of \mathbf{X} , the Taylor expansions of $\varphi(t)$, and $\log \varphi(t)$ if \mathbf{X} admits n -th order moments is:

$$\begin{aligned} \varphi(t) &= \sum_{|\mathbf{v}| < n} \frac{i^{|\mathbf{v}|}}{\mathbf{v}!} m_{\mathbf{v}} t^{\mathbf{v}} + O(|t|^n), \\ \log \varphi(t) &= \sum_{|\mathbf{v}| < n} \frac{i^{|\mathbf{v}|}}{\mathbf{v}!} c_{\mathbf{v}} t^{\mathbf{v}} + O(|t|^n) \quad \text{for } t \rightarrow 0. \end{aligned}$$

Here $\mathbf{v} \in \mathbb{N}^k$, $t = (t_1, \dots, t_k) \in \mathbb{R}^k$ and $|\mathbf{v}| = v_1 + \dots + v_k \leq n$, $\mathbf{v}! = v_1! \dots v_k!$, $t = t_1^{v_1} \dots t_k^{v_k}$. The coefficients $c_{\mathbf{v}}$ are called *cumulants* of \mathbf{X} . Leonov et al., 1959 write

$$\begin{aligned} m_{\mathbf{v}} &= \sum_{\lambda_1 + \dots + \lambda_q = \mathbf{v}} \frac{1}{q!} \frac{\mathbf{v}!}{\lambda_1! \dots \lambda_q!} \prod_{j=1}^q c_{\lambda_j}, \\ c_{\mathbf{v}} &= \sum_{\lambda_1 + \dots + \lambda_q = \mathbf{v}} \frac{(-1)^{q-1}}{q} \frac{\mathbf{v}!}{\lambda_1! \dots \lambda_q!} \prod_{j=1}^q m_{\lambda_j}, \end{aligned}$$

where the sums are taken for every integer q and $\lambda_1, \dots, \lambda_q \in \mathbb{N}^k$ such that $\lambda_1 + \dots + \lambda_q = \mathbf{v}$. In the above definition, we used $\text{cum}(X_1, \dots, X_k) = c_{\mathbf{v}}$

Definition 4.3 (Physical Dependence Measure). *Let $(\eta_t)_{t \in \mathbb{Z}}$ be a sequence of independent identically distributed random variables and let η'_0 be an independent copy of η_0 . Further, define the filtrations:*

$$\mathcal{F}_t = (\dots, \eta_{-2}, \eta_{-1}, \eta_0, \eta_1, \dots, \eta_t) \quad \text{and} \quad \mathcal{F}_t^* = (\dots, \eta_{-2}, \eta_{-1}, \eta'_0, \eta_1, \dots, \eta_t).$$

Finally, let $G : \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}$ denote a possibly nonlinear filter such that $G(\mathcal{F}_t)$ and $G(\mathcal{F}_t^)$ are properly defined random variables. The physical dependence measure of G , for some $k \in \mathbb{N}$, is defined as*

$$\delta_k(G, t) = \left(\mathbb{E} |G(\mathcal{F}_t) - G(\mathcal{F}_t^*)|^k \right)^{1/k}.$$

In classical probability theory and statistics, random variables are commonly assumed to be identically distributed. Similarly to the generalization of the independence assumption, another path of generalization is to allow for a varying distribution. An analogy of this assumption and its generalization to the time series analysis is considered by stationarity and nonstationarity. Stationarity is a trivial indicator that plays a big role in time series modeling. Various models, techniques, and methodologies have been made for stochastic processes based on this indicator; furthermore, an important asymptotic theory was well developed, like weak convergence, limit distributions theorems, or ergodic theorem. However, a weak stationarity assumption is not always advantageous for modeling Spatio-temporal data, even with detrending and deseasonalization, and not all type of time series shows a stationarity behavior, frequently seen in many physical phenomena and in economic data, which makes all these methodologies unfitted and unsuitable. A more realistic environment is based on the idea that many processes in real applications behave locally like stationary processes but evolve continuously over time. This realistic concept is called *locally stationarity*, and it is explicitly introduced first by Dahlhaus, 1996a. Historically, R. A. Silverman, 1957 introduced the *locally stationary random process*. As its name indicates, this type of process approximates a non-stationary process by a stationary one locally over short stretches of time. The intuitive idea of local stationarity was also discussed in the works of Priestley, 1965, Dahlhaus, 1997, Neumann et al., 1997, Sakiyama et al., 2004, Dahlhaus and Polonik, 2006, and many others. It is worth mentioning here that the pioneering works of Dahlhaus, 1997 represent a solid basis for the inference of locally stationary processes. Besides the generalization of stationary processes, this new approach removes time-varying parameters. Other important notions of local stationarity were introduced by Birr et al., 2017; Vogt, 2012 and Zhou et al., 2009.

Definition 4.4 (Local Stationarity, Vogt, 2012). A sequence of time series $(X_{t,T})_{t \in \mathbb{Z}}$ in a normed vector space $(V, \|\cdot\|)$, indexed by $T \in \mathbb{N}$, is called locally stationary (of order ρ) if there exists a family of strictly stationary time series $\{X_t^{(u)} : t \in \mathbb{Z}\}$, indexed by $u \in [0, 1]$, and an array of real-valued random variables $\{P_{t,T}^{(u)} : 1 \leq t \leq T, T \in \mathbb{N}, u \in [0, 1]\}$ with $\mathbb{E} |P_{t,T}^{(u)}|^\rho < \infty$, uniformly in $1 \leq t \leq T, T \in \mathbb{N}$ and $u \in [0, 1]$, such that

$$\|X_{t,T} - X_t^{(u)}\| \leq \left(\left| \frac{t}{T} - u \right| + \frac{1}{T} \right) P_{t,T}^{(u)}.$$

Even though $\{(X_{t,T})_{t \in \mathbb{Z}} : T \in \mathbb{N}\}$ is a sequence of time series, we will occasionally call $(X_{t,T})_{t \in \mathbb{Z}}$ a locally stationary time series.

Example 11. Let $(\varepsilon_t)_{t \in \mathbb{Z}}$ be a sequence of independent, standard normally distributed random variables and consider the extension:

$$X_{t,T} = \sum_{i=0}^{\infty} a_i(t/T) \varepsilon_{t-i}$$

of the MA(∞)-process $X_t = \sum_{i=0}^{\infty} a^i \varepsilon_{t-i}$, where $a_i : [0, 1] \rightarrow \mathbb{R}, i \in \mathbb{N}_0$ are continuous functions. Under certain regularity assumptions, $X_{t,T}$ is locally stationary in the sense of Vogt, 2012 and Zhou et al., 2009, for example, if $a_i(u) = a^i f(u)$ for some Lipschitz continuous function $f : [0, 1] \rightarrow \mathbb{R}$, a constant $a \in (0, 1)$ and any $i \in \mathbb{N}$. For the approximating family of stationary processes defined by

$$X_t^{(u)} = \sum_{i=0}^{\infty} a^i f(u) \varepsilon_{t-i},$$

it follows

$$\begin{aligned} |X_{t,T} - X_t^{(u)}| &= \left| \sum_{i=0}^{\infty} a^i (f(t/T) - f(u)) \varepsilon_{t-i} \right| \\ &\leq \sum_{i=0}^{\infty} a^i |f(t/T) - f(u)| |\varepsilon_{t-i}| \\ &\leq \left| \frac{t}{T} - u \right| C \sum_{i=0}^{\infty} a^i |\varepsilon_{t-i}|. \end{aligned}$$

Thus, $X_{t,T}$ is locally stationary of order 2 according to Vogt, 2012, as for

$$P_{t,T}^{(u)} = \sum_{i=0}^{\infty} a^i |\varepsilon_{t-i}|,$$

it holds

$$\mathbb{E} |P_{t,T}^{(u)}|^2 = \sum_{i_1, i_2=0}^{\infty} a^{i_1+i_2} \mathbb{E} [|\varepsilon_{t-i_1} \varepsilon_{t-i_2}|] \leq \sum_{i_1, i_2=0}^{\infty} a^{i_1+i_2} < \infty.$$

Local stationarity according to the second concept follows from Proposition 3 of Zhou et al., 2009.

5 Functional data

Modern statistics and statistical learning systems often have an interest in complex data analysis from both a methodological and a practical standpoint, particularly in the setting of big data. The goal is to create statistical models and algorithms with solid foundations that can extract knowledge from unstructured data while addressing issues with data complexity, such as heterogeneity, high dimensionality, dynamical behaviors, and missing data. Most statistical methods use data whose constituent elements are typically unstructured finite-dimensional vectors called $x_i \in \mathbb{R}^d$. However, rather than being defined as finite-dimensional vectors in many application fields, the individual data units are better expressed as functions, curves, or surfaces. This case is referred to as a “functional” aspect of the data. To clarify this aspect, a functional random variable is defined as a random variable with values in an infinite dimensional space \mathcal{F} . For example, this space \mathcal{F} can be a space of functions and linear operators... According to the terminology used in the literature, we speak both of functional random variables and of functional data, which includes, in particular, everything concerning the statistical analysis of curves. The difficulties, from a theoretical and practical perspective, comes from the fact that the observations of this type of variable are supposed to belong to an infinite dimensional space. Frédéric Ferraty proposed a general theoretical framework capable of systematically overriding the challenges caused by this functional context. Furthermore, applications and practical developments take an important place too in his works. He was inspired and motivated by the following guidelines. The first in Bosq, 1991, where he stated in a study dealing with autoregressive Hilbertian processes: “*These being nonparametric by themselves, it seems rather heavy to introduce a nonparametric model for observation lying in functional space ...*”. The second can be found in the section “Challenges for the Future” of Ramsay et al., 2005 book “Functional Data Analysis” where he reclaims that: “*theoretical aspects of Functional Data analysis have not been researched in sufficient depth, and it is hoped that appropriate theoretical developments will feed back into advances in practical methodology*”. For this purpose, we can find the following definitions in Ferraty et al., 2006.

Definition 5.1. *A random variable \mathcal{X} is called a functional variable (f.v.) if it takes values in an infinite dimensional space (or functional space). An observation χ of \mathcal{X} is called functional data.*

Definition 5.2. Let \mathcal{Z} be a random variable valued in some infinite-dimensional space F and let ϕ be a mapping defined on F and depending on the distribution of \mathcal{Z} . A model for the estimation of ϕ consists in introducing some constraint of the form $\phi \in \mathcal{C}$. The model is called a functional parametric model for estimating ϕ if \mathcal{C} is indexed by a finite number of elements of F . Otherwise, the model is called a functional nonparametric model.

Definition 5.3. Let X be a random vector-valued in \mathbb{R}^p and let ϕ be a function defined on \mathbb{R}^p and depending on the distribution of \mathbf{X} . A model for estimating ϕ consists of introducing some constraint of the form $\phi \in \mathcal{C}$. The model is called a parametric model for the estimation of ϕ if \mathcal{C} is indexed by a finite number of elements of \mathbb{R} . Otherwise, the model is called a nonparametric model.

Definition 5.4. Let \mathcal{Z} be a random variable valued in some infinite-dimensional space F and let ϕ be a mapping defined on F and depending on the distribution of \mathcal{Z} . A model for the estimation of ϕ consists in introducing some constraint of the form $\phi \in \mathcal{C}$. The model is called a functional parametric model for the estimation of ϕ if \mathcal{C} is indexed by a finite number of elements of F . Otherwise, the model is called a functional nonparametric model.

To better understand the importance of functional data analysis, readers can refer to Bosq, 2000; Ferraty et al., 2006; Horváth et al., 2012; Levitin et al., 2007; Ramsay et al., 2002 among others.

Another challenge related to the functional data is the proximity measures. In finite-dimensional, any classical norm can be used to measure the closeness between two mathematical objects due to the equivalence between all norms in Euclidean space, which is not the case for infinite dimensional space; that is why the choice of the preliminary norm in the functional context is crucial and restrictive. Researchers found that the **semi-metric** are more adapted than the metric space in this case. In fact, a semi metric is a metric where the condition that $d(x, y) = 0 \Rightarrow x = y$ is not satisfied. In the functional data, the family of semi-metric must be adapted for each dataset and each statistical problem. Another concept has been extended to the functional nonparametric framework, the local weighting techniques or local smoothing. In the univariate non-functional case, the idea of the local weighting around a fixed real number x is to attribute at each real random variable X_i a weight taking into consideration the distance between x and X_i , the greater the distance between x and X_i , the smaller the weighting.

In our work, we will use the Kernel estimation method, which is widely used in the nonparametric setting and is such a valuable method to do local weighting, and having the celebrity from the Nadaraya-Watson kernel estimates introduced by Nadaraja, 1964 and Watson, 1964. This method is based on a kernel function and a parameter h , called the smoothing parameter, window width, or bandwidth, that governs the amount of smoothing applied to the sample. The effect of the smoothing parameter h is important since it

influences the level of smoothing. A little value of h may lead the estimator to display unimportant details. Whereas a big value of h causes oversmoothing of the information contained in the sample, which may disguise certain important characteristics, such as multimodality. Many kernel functions can be found in the literature for symmetrical and asymmetrical functions. Below we will cite some of them:

Symmetrical kernel function K :

$$\text{Epanechnikov: } K(t) = \begin{cases} \frac{3}{4\sqrt{5}} \left(1 - \frac{1}{5}t\right)^2 & \text{for } |t| < \sqrt{5}, \\ 0 & \text{for } |t| \geq \sqrt{5}, \end{cases}$$

$$\text{Biweight: } K(t) = \begin{cases} \frac{15}{16} (1 - t^2)^2 & \text{for } |t| < 1, \\ 0 & \text{for } |t| \geq 1. \end{cases}$$

Asymmetrical kernel functions K :

$$\text{Reciprocal Inverse: } K_{IG}(x, b; t) = \frac{1}{\sqrt{2\pi b t^3}} e^{-\frac{1}{2bx} \left(\frac{t}{x} - 2 + \frac{x}{t}\right)},$$

$$\text{Gamma 2: } K_{GAM2}(\rho_b(x), b; t) = \frac{t^{\rho_b(x)-1} e^{-t/b}}{b^{\rho_b(x)} \Gamma(\rho_b(x))}$$

For the functional case, the kernel method can be seen as follows: Let $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_n$ be n functional random variable valued in E and let χ be a fixed element of E . The functional extension of multivariate kernel local weighting ideas would be to transform the n functional random variable $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_n$ into the n quantities

$$\frac{1}{\phi(h)} K\left(\frac{d(\chi, \mathcal{X}_i)}{h}\right),$$

where d is a semi-metric on E , K is a real (asymmetrical) kernel. In this expression $\phi(h)$ would be defined as the probability that \mathcal{X} belong to $B(\chi, h)$ where

$$B(\chi, h) = \{\chi' \in E, d(\chi, \chi') \leq h\},$$

is the ball centred at χ and of radius h , with respect to the topology induced by the semi-metric d . This definition shows us the link between local weighting and the notion of small ball probabilities, which play an essential role in the analysis of functional kernel estimates due to their relation with the semi-metric d , and this will be shown with the rate of convergence of our estimator where the small ball probability function $\phi(h)$ appears.

6 Examples of applications

6.1 Clustering

Clustering refers to the unsupervised learning task that consists in partitioning a set of data points X_1, \dots, X_n in a feature space \mathcal{X} into a finite collection of subgroups depending on their similarity (in a sense that must be specified). Roughly, data points in the same subgroup should be more similar to each other than to those lying in other subgroups. One may refer to Chapter 14 in Hastie et al., 2009 for an account of state-of-the-art clustering techniques. Formally, let $M \geq 2$ be the number of desired clusters and consider $D : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_+$ such that $D(x, x) = 0$ for any $x \in \mathcal{X}$. D measures the dissimilarity between pairs of observations $(x, x') \in \mathcal{X}^2$: the larger $D(x, x')$, the less similar x and x' . For instance, if $\mathcal{X} \subset \mathbb{R}^d$, D could take the form $D(x, x') = \Psi(\|x - x'\|_q)$, where $q \geq 1$, $\|a\|_q = \left(\sum_{i=1}^d |a_i|^q\right)^{1/q}$ for all $a \in \mathbb{R}^d$ and $\Psi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is any borelian non-decreasing function such that $\Psi(0) = 0$. In this context, the goal of clustering methods is to find a partition \mathcal{P} of the feature space \mathcal{X} in a class Π of partition candidates that minimizes the following empirical clustering risk:

$$\widehat{W}_n(\mathcal{P}) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} D(X_i, X_j) \cdot \Phi_{\mathcal{P}}(X_i, X_j),$$

where

$$\Phi_{\mathcal{P}}(x, x') = \sum_{\mathcal{C} \in \mathcal{P}} \mathbb{1}\{(x, x') \in \mathcal{C}^2\}.$$

Assuming that the data X_1, \dots, X_n are i.i.d. realizations of a generic random variable X drawn from an unknown probability distribution $F(dx)$ on \mathcal{X} , the quantity $\widehat{W}_n(\mathcal{P})$, also known as the intra-cluster similarity or within cluster point scatter, is a one sample U -statistic of degree two ($K = 1$ and $d_1 = 2$) with kernel given by:

$$\forall (x, x') \in \mathcal{X}^2, \quad H_{\mathcal{P}}(x, x') = D(x, x') \cdot \Phi_{\mathcal{P}}(x, x'),$$

provided that

$$\iint_{(x, x') \in \mathcal{X}^2} D^2(x, x') \cdot \Phi_{\mathcal{P}}(x, x') F(dx) F(dx') < +\infty.$$

The expectation of the empirical clustering risk $\widehat{W}_n(\mathcal{P})$ is given by

$$W(\mathcal{P}) = \mathbb{E}[D(X, X') \cdot \Phi_{\mathcal{P}}(X, X')],$$

where X' is an independent copy of the r.v. X , and is named the clustering risk of the partition \mathcal{P} . The statistical analysis of the clustering performance of minimizers $\widehat{\mathcal{P}}_n$ of the empirical risk (5) over a class Π of appropriate complexity can be found in Cléménçon, 2014. Based on the theory of U -processes, it is shown in particular how to establish rate bounds for the excess of clustering risk of any empirical minimizer, $W(\widehat{\mathcal{P}}_n) - \inf_{\mathcal{P} \in \Pi} W(\mathcal{P})$ namely, under appropriate complexity assumptions on the cells forming the partition candidates. More detail on this example can be found in Cléménçon et al., 2016.

6.2 Testing stochastic monotonicity

Let us recall the following example from X. Chen et al., 2020. 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:

$$H_0 : F_{Y|X}(y | x) \leq F_{Y|X}(y | x') \forall y \in \mathbb{R} \text{ whenever } x \geq x'.$$

Testing for stochastic monotonicity is an important topic in various applied fields such as economics Blundell et al., 2007; Ellison et al., 2011; Solon, 1992. For this problem, S. Lee et al., 2009 consider a test for H_0 based on a local Kendall's tau statistic, inspired by Ghosal et al., 2000. Let $(X_i, Y_i), i = 1, \dots, n$ be i.i.d. copies of (X, Y) . S. Lee et al., 2009 consider the U -process

$$U_n(x, y) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \{1(Y_i \leq y) - 1(Y_j \leq y)\} \text{sign}(X_i - X_j) L_{b_n}(x - X_i) L_{b_n}(x - X_j),$$

where $b_n \rightarrow 0$ is a sequence of bandwidths, $\text{sign}(x) = 1(x > 0) - 1(x < 0)$ is the sign function, and L_{b_n} is a one dimensional kernel function with bandwidth b_n . 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. S. Lee et al., 2009 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 X, Y are independent, under which $U_n(x, y)$ is centered. Under independence between X and Y and regularity conditions, they derive a Gumbel limiting distribution for a properly scaled version of S_n using techniques from (Piterbarg, 1996). However, they do not consider bootstrap approximations to S_n . It should be noted that S. Lee et al., 2009 considered a slightly more general setup than that described above. They allow X_i not to be directly observed but assume that estimated X_i are available and cover the case where X is multi-dimensional.

6.3 Testing curvature and monotonicity of nonparametric regression

Let us recall the following example from X. Chen et al., 2020. Consider the nonparametric regression model $Y = f(X) + \varepsilon$ with $\mathbb{E}[\varepsilon | X] = 0$, where Y is a scalar outcome variable, X is an m -dimensional vector of regressors, ε is an error term, and f is the conditional mean function $f(x) = \mathbb{E}[Y | X = x]$. We observe i.i.d. copies $V_i = (X_i, Y_i), i = 1, \dots, n$ of $V = (X, Y)$. We are interested in testing for qualitative features (e.g., curvature, monotonicity) of the regression function f .

Abrevaya et al., 2005 consider a simplex statistic to test linearity, concavity, and convexity of f under the assumption that the conditional distribution of ε given X is symmetric. To define their test statistics, for $x_1, \dots, x_{m+1} \in \mathbb{R}^m$, let

$$\Delta^\circ(x_1, \dots, x_{m+1}) = \left\{ \sum_{i=1}^{m+1} a_i x_i : 0 < a_j < 1, j = 1, \dots, m+1, \sum_{i=1}^{m+1} a_i = 1 \right\}$$

denote the interior of the simplex spanned by x_1, \dots, x_{m+1} , and define $\mathcal{D} = \bigcup_{j=1}^{m+2} \mathcal{D}_j$, where

$$\mathcal{D}_j = \left\{ (x_1, \dots, x_{m+2}) \in \mathbb{R}^{m \times (m+2)} : \begin{array}{l} x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_{m+2} \text{ are affinely independent} \\ \text{and } x_j \in \Delta^\circ(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_{m+2}) \end{array} \right\}.$$

The sets $\mathcal{D}_1, \dots, \mathcal{D}_{m+2}$ are disjoint. For given $v_i = (x_i, y_i) \in \mathbb{R}^m \times \mathbb{R}, i = 1, \dots, m+2$, if $(x_1, \dots, x_{m+2}) \in \mathcal{D}$ 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 $x_j = \sum_{i \neq j} a_i x_i$; 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 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 et al., 2005 consider the following *localized simplex statistic*

$$U_n(x) = \frac{1}{|I_n^{m+2}|} \sum_{(i_1, \dots, i_{m+2}) \in I_n, m+2} \varphi(V_{i_1}, \dots, V_{i_{m+2}}) \prod_{k=1}^{m+2} L_{b_n}(x - X_{i_k}), \quad (6.1)$$

where

$$\varphi(v_1, \dots, v_{m+2}) = 1\{(x_1, \dots, x_{m+2}) \in \mathcal{D}\} \text{sign}(w(v_1, \dots, v_{m+2})),$$

which is a U -process of order $(m+2)$. To test concavity and convexity of f , Abrevaya et al., 2005 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.

6.4 Lévy-driven MA random field

This example is an example of a wide class of random fields and represents an application to our theoretical results; for instance, see Chapter 7. Lévy process is a stochastic process whose increments in overlapping time intervals are independent and whose increments are stationary on time; for details, see Sato, 1999 and Bertoin, 1996.

Define $L = \{L(A) : A \in \mathcal{B}(\mathbb{R}^d)\}$ as an infinitely divisible random measure on some probability space (Ω, \mathcal{A}, P) , in the sense that

1. For each sequence $(E_m)_{m \in \mathbb{N}}$ of disjoint sets in $\mathcal{B}(\mathbb{R}^d)$,
 - i) $L(\cup_{m=1}^{\infty} E_m) = \sum_{m=1}^{\infty} L(E_m)$ a.s. whenever $\cup_{m=1}^{\infty} E_m \in \mathcal{B}(\mathbb{R}^d)$,
 - ii) $(L(E_m))_{m \in \mathbb{N}}$ is a sequence of independent random variables.
2. The random variable $L(A)$ has an infinitely divisible distribution for any $A \in \mathcal{B}(\mathbb{R}^d)$.

A Lévy-Khintchine representation of the form $\varphi_{L(A)}(t) = \exp(|A|\psi(t))$ is the characteristic function $\varphi_{L(A)}(t)$ of $L(A)$ where

$$\psi(t) = it\gamma_0 - \frac{1}{2}t^2\sigma_0 + \int_{\mathbb{R}} \{e^{itx} - 1 - itxI(x \in [-1, 1])\} \nu_0(x)dx$$

and $i = \sqrt{-1}$, $\gamma_0 \in \mathbb{R}$, $0 \leq \sigma_0 < \infty$, ν_0 is a Lévy density with $\int_{\mathbb{R}} \min\{1, x^2\} \nu_0(x)dx < \infty$, and $|A|$ is the Lebesgue measure of A . The triplet $(\gamma_0, \sigma_0, \nu_0)$ determines the distribution of random measure L , which is called the Lévy characteristic of L ; Sato, 1999 is a standard reference on Lévy processes.

A Lévy-driven MA random field driven by an infinitely divisible random measure L , which we call Lévy random measure, is defined by

$$X(s) = \int_{\mathbb{R}^d} g(s-v)L(dv)$$

for every $s \in \mathbb{R}^d$. In particular, when $g(\cdot)$ is a kernel function of the form

$$g(\mathbf{s}) = \sum_{i=1}^p \frac{b(\lambda_i)}{a'(\lambda_i)} e^{\lambda_i \|\mathbf{s}\|}$$

where a' is the the derivative of the polynomial $a(z) = \prod_{i=1}^p (z^2 - \lambda_i^2)$, with real coefficients and distinct negative zeros $\lambda_1, \dots, \lambda_p$, and $b(z) = \prod_{i=1}^q (z^2 - \xi_i^2)$, with real coefficients and real zeros ξ_1, \dots, ξ_q . $X(s)$ is a univariate (isotropic) CARMA(p,q) random field, which is a special case of Lévy-driven MA random fields. Brockwell et al., 2017 generalized CARMA(p,q) from \mathbb{R} to \mathbb{R}^d .

For, $g : [0, 1]^d \times [0, \infty) \rightarrow \mathbb{R}$ is a bounded function such that $|g(\mathbf{u}, \cdot) - g(\mathbf{v}, \cdot)| \leq C\|\mathbf{u} - \mathbf{v}\|\bar{g}(\cdot)$ with $C < \infty$, if we define:

$$X_{s, A_n} = \int_{\mathbb{R}^d} g\left(\frac{s}{A_n}, \|\mathbf{s} - \mathbf{v}\|\right) L(d\mathbf{v}) \text{ and } X_u(\mathbf{s}) = \int_{\mathbb{R}^d} g(\mathbf{u}, \|\mathbf{s} - \mathbf{v}\|) L(d\mathbf{v}),$$

then, X_{s, A_n} is a locally stationary Lévy-driven MA random fields, and $X_u(\mathbf{s})$ is a strictly stationary Lévy-driven MA random fields.

7 Organisation 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 understand the forthcoming chapters. We also review some of the standard facts concerning U -processes and their weak convergence, class of functions, mixing condition, and random field, with particular attention given to the basic tools needed to treat the U -processes. Readers not interested in the asymptotic theory may merely go over this chapter to become familiar with the concepts and definitions central to the work. It should be mentioned here that, for the clarity of each chapter, other definitions and notions are also defined in each of them.

Chapter 3. Renewal type bootstrap for increasing degree U -process of a Markov chain

Let $X = (X_n)_{n \in \mathbb{N}}$ be an homogeneous Markov chain defined on a measurable space (E, \mathcal{E}) . Let $\pi(x, dy)$ the transition probability, $\nu = \nu(i)_{i > 0}$ initial probability. Therefore, we will denote by \mathbb{P}_ν or just \mathbb{P} the probability measure determined by $\mathbb{P} = (\pi, \nu)$. Consider a parametric function $\Theta(h) = \Theta$, which is a unique invariant measure for the Harris-recurrent Markov chain, and a sequence of parameters Θ_m such that Θ_m converges to θ , these two parameters are defined as follows, for \mathbb{P} a probability measure:

$$\Theta(h) = \mathbb{P}h(X_1, X_2, \dots),$$

and, for $h_m : E^m \rightarrow \mathbb{R}$,

$$\Theta_m(h_m) = \mathbb{P}^m h_m(X_1, \dots, X_m) = \int_E \dots \int_E h_m(x_1, \dots, x_m) \boldsymbol{\mu}(dx_1) \dots \boldsymbol{\mu}(dx_m),$$

There exists an infinite-argument unbiased estimator for this parameter and it can be

defined as an infinite-degree U -statistic:

$$U_n(h_m) = \binom{n}{m}^{-1} \sum_{\mathbf{i} \in I_n^m} h_m(X_{i_1}, \dots, X_{i_m}). \quad (7.1)$$

Throughout the chapter, \mathcal{H}_m will denote the class of kernels h_m , of degree $m = m_n$ that increases with sample size n , and where the kernel is a function $h_m : E^m \rightarrow \mathbb{R}$.

In this chapter, we investigate the uniform limit theory for a U -statistic of increasing degree, also called an infinite-degree U -statistic defined in (7.1). Infinite-degree U -statistics (IDUS) are useful tools for constructing simultaneous prediction intervals that quantify the uncertainty of several methods such as subbagging and random forests. The stochastic process based on collections of U -statistics is referred to as a U -process, and if the U -statistic is infinite-degree, we have an infinite-degree U -process. C. Heilig et al., 2001 provided conditions for the pointwise asymptotic theory for the infinite-degree U -processes. The main purpose here is to extend their findings to the Markovian setting. The second aim is to provide the uniform limit theory for the renewal bootstrap for the infinite-degree U -process, which is of its own interest. The main ingredients are the decoupling technique combined with symmetrization techniques of C. Heilig et al., 2001 to obtain uniform weak law of large numbers and functional central limit theorem for the infinite-degree U -process. To the best of our knowledge, we are the first to present successful results on the infinite-degree U -process in the Markovian context. The primary purpose of the present chapter is to generalize the work of U -statistic and U -process for the case where the degree m converges to infinity or, in other words, increases with the size of the sample in the framework of Markov chains. We generalize the work of C. Heilig et al., 2001 to the Harris recurrent Markov chain by using the renewal properties of this chain. In addition, we study the limit theorems of the U -statistic estimators in a class of functions that keep some properties to find the desired results. We also consider the bootstrapped version of the U -process, which is in its own interest. The context of the present chapter has, as far as we know, yet to be considered so far in the literature, which is substantially more complicated than the independence framework. We commonly use the abbreviations IDUS or IOUS for increasing (or infinite) degree/order U -statistics.

In this chapter, we have established the weak law of large numbers and the uniform central limit theorems given by the following. The appropriate notation and definitions are given in Chapter 3.

Theorem 7.1

Let $\{X_i\}_{i=1}^n$ be a sequence of Harris-recurrent Markov chain random variables and assume that the following conditions hold.

(A.1) The class \mathcal{G}_m is a uniformly bounded permissible such that for the envelope \bar{G}_m , for all m , $V > 1$ and $\varepsilon \in (0, 1]$:

$$\mathcal{N} \left(\varepsilon \|\bar{G}_m\|_{L_2(\mathbb{P})}, \mathcal{G}_m, \|\cdot\|_{L_2(\mathbb{P})} \right) \leq A\varepsilon^{-V}.$$

(A.2) $\sigma_1(g_m, \check{g}_m) = \lim_{n \rightarrow \infty} \mathbb{P} g_m(X) \check{g}_m(X)$ exists for every g_m, \check{g}_m in \mathcal{G}_m , where, for h_m and $\check{h}_m \in \mathcal{H}_m$, $\check{g}_m(x) = \sum_{j=1}^m j \check{h}_{j|1}(x)$ and $g_m(x) = \sum_{j=1}^m j h_{j|1}(x)$. For

$$\rho_n^2(g_m, \check{g}_m) = \sum_{j=1}^m \sum_{k=1}^m jk \mathbb{P} \left[h_{j|1}(X_1) - \check{h}_{j|1}(X_1) \right] \left[h_{k|1}(X_1) - \check{h}_{k|1}(X_1) \right],$$

the limit $\rho(\cdot, \cdot) = \lim_{n \rightarrow \infty} \rho_n(\cdot, \cdot)$ is well-defined, and for all sequences of functions $\{\bar{g}_n\}$ and $\{\check{g}_n\}$ in \mathcal{G}_m , if $\rho(\bar{g}_n, \check{g}_n) \rightarrow 0$, then $\rho_n(\bar{g}_n, \check{g}_n) \rightarrow 0$.

(A.3) $\sum_{i=1}^n \mathbb{P} \bar{G}_m(X_i)^2 \left\{ \bar{G}_m(X_i)^2 > \varepsilon \right\} \rightarrow 0$, for every $\varepsilon > 0$.

(A.4) We suppose that $\mathbb{P}_A \left[\left(\sum_{j=1}^{\mathcal{J}_0} g_m(X_j) \right)^2 \right] < \infty$.

(A.5) $\mathbb{E}(\tau)^{2+\alpha} < \infty$ ($\alpha > 0$ fixed).

Then $\sqrt{n}(\mathbb{P}_n - \mathbb{P})g_m$ converges in distribution to a mean-zero Gaussian process which is uniformly ρ -continuous and has covariance given by

$$\sigma(g, \check{g}) = \lim_{n \rightarrow \infty} \sum_{j=1}^m \sum_{k=1}^m jk \mathbb{P} \left[h_{j|1}(X_1) - \check{h}_{j|1}(X_1) \right] \left[h_{k|1}(X_1) - \check{h}_{k|1}(X_1) \right].$$

Theorem 7.2: Weak convergence theorem for IDUS

Let $(X_n)_{n \geq 0}$ be a Harris-recurrent Markov chain with atom \mathbf{A} , and $\tau(j)$ the renewal time such that $\mathbb{E}(\tau)^{2+\alpha} < \infty$. Suppose also that all conditions of Theorem 7.1 hold. Then, if the class \mathcal{H}_m has \mathbb{P}^m -square integrable envelope \bar{H}_m , i.e., $\mathbb{P}^m(\bar{H}_m^2) < \infty$, we have $\sqrt{n}[U_n(h_m) - \mathbb{P}^m(h_m)]$ converges in distribution to a mean-zero Gaussian process \mathbb{G} which is uniformly ρ -continuous.

Theorem 7.3

Let $\{X_i\}_{i=1}^n$ be a sequence of r. v. of Harris Markov chain, and \mathcal{H}_m an euclidean permissible class of function. Under the assumptions of Theorem 7.1 and Theorem 7.2, we have

$$\sup_{\Upsilon \in \text{BL}_1} \left| \mathbb{E}^* \left(\Upsilon \left(\sqrt{n^*} \left(U_n^*(\tilde{h}_m) - U_n(\tilde{h}_m) \right) \right) \right) - \mathbb{E}(\Upsilon(\mathbb{G})) \right| \rightarrow 0 \text{ in probability, } \quad (7.2)$$

where BL_1 is the set of all functions $\Upsilon : \ell^\infty(\mathcal{H}_m) \mapsto [0, 1]$ such that $|\Upsilon(z_1) - \Upsilon(z_2)| \leq \|z_1 - z_2\|_{\mathcal{H}_m}$ for every z_1, z_2 .

To better understand the idea of infinite degree U -statistics, we will give some examples extracted from different references. Although only four examples will be given here, they stand as archetypes for various examples that can be similarly investigated. This includes Simultaneous prediction intervals for random forests, the Renewal estimator, the Kaplan-Meier estimator, and the Subsampling Distribution.

Chapter 4. Renewal type bootstrap for U - process Markov chains

The primary purpose of the present chapter is to establish bootstrap uniform functional central limit theorems U -processes for Harris recurrent Markov chains over uniformly classes of functions satisfying some entropy conditions. To simplify our approach, we will employ the well-known regenerative properties of Markov chains, avoiding some complicated mixing conditions. The most related work of the present chapter is (2011a). In the last reference, the authors provided an approach to the study of U -statistics in the Markovian setup based on the (pseudo-) regenerative properties of Harris Markov chains. Exploiting the fact that any sample path X_1, \dots, X_n of a general Harris chain X may be divided into asymptotically independent and identically distributed (i.i.d.) data blocks. This kind of regeneration in Markov chains has been detailed in the paper of Nummelin, 1991. A similar approach will be used in our work without restriction to the case $m = 2$. The present chapter considers the general framework of the U -processes in the Markov chain setting and their bootstrapped versions. This investigation is far from trivial, and it is harder to control equicontinuity, which forms an unsolved open problem in the literature. The main difficulties in proving our results are related to the random size of the resampled blocks. This fact creates problems with random stopping times (in real and Banach space settings). In fact, delicate mathematical derivations will be required to cope with U -processes in our context. Our result is obtained under the minimal condition of the envelope function. We also consider an extension to the k Markov chain setting and prove the bootstrap consistency. The main theorems are the following.

Theorem 7.4: Weak convergence

Let $(X_n)_n$ be an Harris-recurrent Markov chain with an atom \mathbf{A} satisfying Assumption 2.1 and such that, for a fixed $\gamma > 0$,

$$\sup_{x \in \mathbf{A}} \mathbb{E}(\tau)^{2+\gamma} < \infty \quad (7.3)$$

and satisfying also the technical assumptions (C.1), (C.2), (C.3), (C.4) and (C.5). Let \mathcal{F} a class of measurable functions uniformly bounded on E and

$$\int_0^\infty \sqrt{\log N(\varepsilon, \mathcal{F}, e_{n,2})} d\varepsilon < \infty,$$

^a with expectation zero, i.e.,

$$\mathbb{E} \int_0^\infty n^{-1/2} \log N(\varepsilon, \mathcal{F}, e_{n,2}) d\varepsilon \rightarrow 0.$$

Then the U -statistic converges weakly in $l^\infty(\mathcal{F})$ to a Gaussian process G indexed by \mathcal{F} whose sample paths are bounded and uniformly continuous with respect to the metric $L_2(\mathbb{P})$.

^auniform entropy integral

Theorem 7.5

Let $(X_n)_n$ be a positive recurrent Harris Markov chain, with an accessible atom \mathbf{A} , X_n satisfies the conditions (C.1) and (C.2) (moments assumptions). Let \mathcal{F} be a uniform bounded class of functions with an envelope H square integrable such that:

$$\int_0^\infty (\log N(\varepsilon, \mathcal{F}, e_{n,2}))^{m/2} d\varepsilon < \infty.$$

Then the process

$$(n^*)^{1/2}(U_{n^*}^*(h) - U_n(h)),$$

converges weakly in probability under \mathbb{P}_ν to a Gaussian process G indexed by \mathcal{F} whose sample paths are bounded and uniformly continuous with respect to the metric $L_2(\mathbb{P})$.

Theorem 7.6

Let $U_{(n_1, n_2)}(h)$ be a generalized U -statistic based on two Markov chains trajectories $(X_1^{(1)}, \dots, X_{n_1}^{(1)})$ and $(X_1^{(2)}, \dots, X_{n_2}^{(2)})$, and $R_{(l_{n_1-1}, l_{n_2-1})}(h)$ the regenerative U -statistics generated by two samples of blocks independents. Let $\tau_{\mathbf{A}_1}$ and $\tau_{\mathbf{A}_2}$ represent the renewal time for the first and second chains with atom \mathbf{A}_1 and \mathbf{A}_2 respectively. Then

$$n^{1/2}(U_{(n_1, n_2)} - \Theta) \xrightarrow{d} N\left(0, \frac{l}{n}(\mathbb{E}_{\mathbf{A}_1}(\tau_{\mathbf{A}_1})^{-2})(\mathbb{E}_{\mathbf{A}_2}(\tau_{\mathbf{A}_2})^{-2}\left(\frac{m_1^2 \delta_{1,0}^2}{p} + \frac{m_2^2 \delta_{0,1}^2}{1-p}\right))\right),$$

and

$$l^{1/2}R_{(l_{n_1-1}, l_{n_2-1})} \xrightarrow{d} N(0, l(p^{-1}m_1^2\delta_{1,0}^2 + (1-p)^{-1}m_2^2\delta_{0,1}^2)),$$

where $l = l_{n_1} + l_{n_2}$, $n = n_1 + n_2$ and $p = l_{n_1}/l \rightarrow p, 0 < p < 1$. We have also

$$\delta_{(c,d)} = \text{Var } h^{(c,d)}(B_1^{(1)}, B_2^{(1)}; B_1^{(2)}, B_2^{(2)}),$$

such that $\delta_{1,0}, \delta_{0,1} > 0$.

Chapter 5. Exchangeably weighted bootstraps of the General Markov U -process

We consider an exchangeably weighted bootstrap of the general function-indexed empirical U -processes in the Markov setting, a natural higher-order generalization of the weighted bootstrap empirical processes extending the previous chapter. Many bootstrap resampling schemes emerge as special cases of our results. To the best of our knowledge, this general context has yet to be considered so far in the literature. We will combine, in a non-trivial way, the techniques of the renewal bootstrap with the randomly weighted bootstrap. At this point, we mention a connection between moving blocks bootstrap and its modification, matched block bootstrap. Instead of artificially breaking a sample into blocks of a fixed size and then resampling from them, the latter tries to match the blocks to make a smoother transition; refer to Radulović, 2004 for the explanation. This chapter aims to investigate the exchangeable bootstrap for the U -processes in a similar fashion as in Q. Han, 2022. As in the previous chapter, the main difficulties in proving Theorems 4.2 related to the random size of the resampled blocks, which generates problems with the random stopping times; and can not be removed by replacing a random stopping time with its expectation. In the present setting, the random bootstrap variables are formed by resampling from a random number of blocks. The conditioning arguments can overcome the problem, but the answer is negative. Our proof uses some arguments from Radulović, 2004 and Q. Han, 2022 by verifying bootstrap stochastic equicontinuity by comparing it

to the original process in a similar way as in Giné et al., 1990. However, as will be seen later, the problem requires much more than ‘simply’ combining ideas from these papers. The following theorems are established under fairly general structural conditions on the classes of functions (eventually unbounded) and the underlying distributions. This chapter provides the first general theoretical study of the bootstrap of the empirical U -processes in the Markov setting.

Theorem 7.7

Let $(X_n)_n$ be a positive recurrent Harris Markov chain, with an accessible atom \mathbf{A} , X_n satisfies the conditions (C.1) and (C.2) (moments assumptions), (C.3), (C.4), (C.5). Let \mathcal{F} be a uniform bounded class of functions with an envelope H square integrable such that:

$$\int_0^\infty (\log N(\varepsilon, \mathcal{F}, e_{n,2}))^{m/2} d\varepsilon < \infty.$$

Then the process Z_n converges weakly in probability under \mathbb{P}_ν to a Gaussian process $\mathbb{G}_\mathbb{P}$ indexed by \mathcal{F} whose sample paths are bounded and uniformly continuous with respect to the metric $L_2(\mathbb{P}_\nu)$.

Theorem 7.8

Suppose Assumptions (A1) to (A4), and Conditions(C.1)-(C.5) hold. Let $\mathcal{F} \subset L_2^{c,m}(P)$ permissible², admit a P^m -square integrable envelope F such that

$$\int_0^1 \left(\sup_Q \log \mathcal{N}(\varepsilon \|F\|_{L_2(Q)}, \mathcal{F}, L_2(Q)) \right)^{m/2} d\varepsilon < \infty,$$

where the supremum is taken over all discrete probability measures. Then

$$\sup_{\psi \in \text{BL}} \left| \mathbb{E}_\xi \psi \left(Z_n^*(h) \right) - \mathbb{E} \psi(c \cdot \mathbb{K}_\mathbb{P}) \right| \rightarrow_{\mathbb{P}_\nu} 0,$$

where c is the constant in (A3), and the convergence in probability $\rightarrow_{\mathbb{P}_\nu}$ is with respect to the outer probability of \mathbb{P}^∞ defined on $(E^\infty, \mathcal{E}^\infty)$.

An application to testing symmetry, the test of independence, and Kendall’s Tau are provided.

Chapter 6. Weak convergence of the Conditional U -processes for Locally Stationary Functional Time Series

U -statistics are a fundamental type of statistics that arise from modeling quantities of interest defined by multi-subject responses. The empirical mean of a random variable X is generalized to sums over every m -tuple of distinct observations of X by U -statistics. W. Stute [Ann. Probab. 19 (1991) 812–825] introduced a class of U -statistics that can be thought of as a generalization of the Nadaraya-Watson regression function estimates. Stute proved their strong pointwise consistency to :

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

Stationarity is frequently assumed in time series modeling, resulting in various models, techniques, research, and methodologies. Even with detrending and deseasonalization, the stationarity assumption is only sometimes advantageous for modeling Spatio-temporal data. It is worth noting that some important time series models are not stationary, which is common in many physical phenomena and economic data. As a result, the stationarity assumption is violated, rendering classical methods inapplicable. To address this issue, R. A. Silverman, 1957 extended the concept of the stationary process to the so-called *locally stationary* random process. As the name implies, this type of process approximates a non-stationary process by a stationary one locally over short periods.

The intuitive concept of local stationarity is also discussed in the works of Priestley, 1965, Dahlhaus, 1997, Neumann et al., 1997, Sakiyama et al., 2004, Dahlhaus and Polonik, 2006, to name a few. It is worth noting that Dahlhaus's pioneering work provides a solid foundation for inferring locally stationary processes. In addition to generalizing stationary processes, this new approach eliminates time-varying parameters. Since the empirical processes theory rocketed in the resolution of statistical problems and spread in time series analysis and regression estimation, Dahlhaus et al., 2009, Vogt, 2012 and, more recently, Mayer et al., 2020 and Phandoidaen et al., 2022. The extension of the preceding investigation to conditional empirical U -processes is of great practical and theoretical interest. We are specifically interested in the conditional U -process, which is indexed by a class of functions in the functional data setting. The main goal of this chapter is to establish weak convergence of conditional U -processes in the locally stationary functional mixing data framework. The kernel estimator considered is defined by

$$\tilde{r}_n^{(m)}(\varphi, \mathbf{x}, \mathbf{u}; h_n) = \frac{\sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d(x_k, X_{i_k, n})}{h_n} \right) \right\} \varphi(Y_{\mathbf{i}, n})}{\sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d(x_k, X_{i_k, n})}{h_n} \right) \right\}},$$

where $K_1(\cdot)$ and $K_2(\cdot)$ denote one-dimensional kernel functions. Here, $h = h_n$ is a bandwidth satisfying $h \rightarrow 0$ as $n \rightarrow \infty$ and $\varphi : \mathcal{Y}^m \rightarrow \mathbb{R}$ is a symmetric measurable function belonging to some class of functions \mathcal{F}_m . The primary goal of this chapter is to provide the first comprehensive theoretical investigation in this context. This necessitates applying large sample theory techniques for empirical and U -empirical processes. We look specifically at the weak convergence of the conditional empirical process indexed by a suitable class of functions and conditional U -processes when the explicative variable is functional. We consider weak convergence in both cases where the class of functions is bounded or unbounded and some moment conditions are satisfied. The precise statement of the following theorems is given in the corresponding chapter.

Theorem 7.9

Let $\mathcal{F}_m \mathcal{H}^m$ be a measurable VC-subgraph class of functions such as Assumption 6 is satisfied. Assume also that Assumptions 1, 2, 3, and 4 are satisfied. Then we have

$$\begin{aligned} & \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in [C_1 h, 1 - C_1 h]^m} \left| \tilde{r}_n^{(m)}(\varphi, \mathbf{x}, \mathbf{u}; h_n) - r^{(m)}(\varphi, \mathbf{x}, \mathbf{u}) \right| \\ &= O_{\mathbb{P}} \left(\sqrt{\frac{\log n}{nh^m \phi(h)}} + h^{2m \wedge \alpha} \right). \end{aligned} \quad (7.4)$$

Theorem 7.10

Let $\mathcal{F}_m \mathcal{H}^m$ be a measurable VC-subgraph class of functions, and assume that all the assumptions of Section 2.8 are satisfied. Then as $n \rightarrow \infty$, for any $\mathbf{x} \in \mathcal{H}^m$ and $\mathbf{u} \in [0, 1]^m$:

$$\sqrt{nh^m \phi(h)} \left(\tilde{r}_n^{(m)}(\varphi, \mathbf{x}, \mathbf{u}; h_n) - r^{(m)}(\varphi, \mathbf{x}, \mathbf{u}) \right)$$

converges in law to a Gaussian process $\{\mathbb{G}_n(\psi) : \psi \in \mathcal{F}_m \mathcal{H}^m\}$ that admits a version with uniformly bounded and uniformly continuous paths with respect to $\|\cdot\|_2$ -norm with covariance function given in (4.2).

To motivate our work, we considered discrimination and metric learning as applications. We believe that the local stationary process provided in this chapter could be applied in various fields, such as economics and medicine. Therefore, relaxing the assumption of stationary behavior of the data-generating processes has received increased interest from the statistical community in recent years. It is encouraged by the rise of possible applications that contradict the stationary assumption, such as the work of Anderson et al., 2019; Bardet et al., 2018; Cléménçon et al., 2004, among others.

Chapter 7. Nonparametric conditional U -processes for locally stationary random fields under stochastic sampling design

For many years, statisticians and probabilists have worked hard to solve the regression problem, which greatly impacts the development of a wide range of methods. Indeed, various topics, such as modeling, estimation method purposes, tests, and other related issues, have been addressed. Aside from the parametric framework, in which a finite number of parameters must be estimated based on an a priori specified model structure, the nonparametric setting is dedicated to data with no a priori structural information. Nonparametric procedures have natural drawbacks in estimation biases and rate of convergence losses compared to parametric methods. Kernel nonparametric function estimation methods have long piqued the interest of researchers looking for good references to the research literature in this area, along with statistical applications consult B. W. Silverman, 1986, Nadaraya, 1989, Härdle, 1990, Wand et al., 1995, Eggermont et al., 2001, Devroye et al., 2001 and the references therein. By extending the previous chapter to the spatial data setting, we will focus on constructing consistent kernel-type estimators for conditional U -statistics in this article. Spatial data collected at measurement sites and statistically treated typically arise in a variety of fields of research, including econometrics, epidemiology, environmental science, image analysis, oceanography, meteorology, geostatistics, and many others. Refer to Ripley, 1981, Rosenblatt, 1985, Guyon, 1995 and Cressie, 2015 and the references therein for good sources of references to the research literature in this area as well as statistical applications. We cite some key references in the context of nonparametric estimation for spatial data, primarily concerned with the estimation of probability density and regression functions. Tran, 1990, Tran et al., 1993, Biau et al., 2004, Dabo-Niang et al., 2013 and the references therein.

We use these studies to present a more general and abstract context by focusing on conditional U -processes for locally stationary random fields $\{\mathbf{X}_{s,A_n} : s \in R_n\}$ in \mathbb{R}^p observed at irregularly spaced locations in $R_n = [0, A_n]^d \subset \mathbb{R}^d$. We adopt a stochastic sampling scheme that can generate irregularly spaced sampling sites in a flexible manner and include both pure increasing domain and mixed increasing domain frameworks. Motivated by numerous applications, the theory of U -statistics and U -processes (introduced in a seminal work by Hoeffding, 1948) has received considerable attention in recent decades. The primary goal of this chapter is to consider a general framework and characterize the weak convergence of the conditional U -processes based regular sequence of spatial random functions. This investigation is far from trivial, and it is even more difficult to control the asymptotic equi-continuity under minimal conditions in this general context, which constitutes a largely unsolved open problem in the literature. We hope to fill this gap in the literature by combining results from Arcones and Yu, 1994, and Bouzebda et al., 2019b, 2022 with techniques for dealing with functional data from Masry, 2005. However,

as will be seen later, the problem necessitates far more than simply combining ideas from previous results. To deal with the regular functional data in our context, delicate mathematical derivations will be required. This necessitates the effective application of large sample theory techniques developed for empirical processes, where we used results from Arcones and Yu, 1994 and Bouzebda et al., 2019b, 2022.

The purpose of this work is to establish the weak convergence of the conditional U -process based on the following U -statistic

$$\begin{aligned}
 \widehat{r}_n^{(m)}(\mathbf{x}, \mathbf{u}; h_n) &:= \widehat{r}_n^{(m)}(\varphi, \mathbf{x}, \mathbf{u}; h_n) \\
 &= \frac{\sum_{\mathbf{i} \in I_n^m} \varphi(\mathbf{Y}_{\mathbf{s}_{i_1, A_n}}, \dots, \mathbf{Y}_{\mathbf{s}_{i_m, A_n}}) \prod_{j=1}^m \left\{ \overline{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j, A_n}}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j, A_n}})}{h_n} \right) \right\}}{\sum_{\mathbf{i} \in I_n^m} \prod_{j=1}^m \left\{ \overline{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j, A_n}}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j, A_n}})}{h_n} \right) \right\}} \\
 &= \frac{\sum_{\mathbf{i} \in I_n^m} \varphi(\mathbf{Y}_{\mathbf{s}_{i_1, A_n}}, \dots, \mathbf{Y}_{\mathbf{s}_{i_m, A_n}}) \prod_{j=1}^m \left\{ \prod_{\ell=1}^d K_1 \left(\frac{u_{j, \ell} - \frac{\mathbf{s}_{i_j, \ell}}{A_n}}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j, A_n}})}{h_n} \right) \right\}}{\sum_{\mathbf{i} \in I_n^m} \prod_{j=1}^m \left\{ \prod_{\ell=1}^d K_1 \left(\frac{u_{j, \ell} - \frac{\mathbf{s}_{i_j, \ell}}{A_n}}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j, A_n}})}{h_n} \right) \right\}},
 \end{aligned}$$

where

$$I_n^m := \{\mathbf{i} = (i_1, \dots, i_m) : 1 \leq i_j \leq n \text{ and } i_r \neq i_j \text{ if } r \neq j\}, \quad \overline{K}(\mathbf{u}) = \prod_{\ell=1}^d K_1(u_\ell),$$

and $\varphi : \mathcal{Y}^m \rightarrow \mathbb{R}$ is a symmetric measurable function belonging to some class of functions \mathcal{F}_m , and $\{h_n\}_{n \in \mathbb{N}^*}$ a sequence of positive real numbers satisfying $h_n \rightarrow 0$ as $n \rightarrow \infty$.

The following theorem provides uniform rates of strong convergence.

Theorem 7.11

Let $\mathcal{F}_m \mathcal{K}^m$ be a measurable VC-subgraph class of functions satisfying Assumption 13. Let $I_h = [C_1 h, 1 - C_1 h]^{dm}$ and let S_c be a compact subset of \mathbb{R}^m . Suppose that

$$\inf_{\mathbf{u} \in [0,1]^{dm}, \mathbf{x} \in S_c} f(\mathbf{u}, \mathbf{x}) > 0.$$

Then, under Assumptions 7, 8, 9, Condition (B1) in 10, 11 and 12 (with $W_{s_i, A_n} = 1$ and ε_{s_i, A_n}), the following result holds for P_S almost surely:

$$\begin{aligned} & \sup_{\mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{x} \in \mathcal{K}^m} \sup_{\mathbf{u} \in I_h, \mathbf{x} \in S_c} \left| \hat{r}_n^{(m)}(\mathbf{x}, \mathbf{u}; h_n) - r^{(m)}(\mathbf{x}, \mathbf{u}; h_n) \right| \\ &= O_{P, |S} \left(\sqrt{\log n / n h^{md} \phi^m(h)} + h^{2\wedge\alpha} + \frac{1}{A_n^{dp} \phi(h)} \right), \end{aligned}$$

where $p = \min\{1, \rho\}$.

The following theorem gives the main results of the chapter concerning the uniform CLT for the conditional U -processes.

Theorem 7.12

Let $\mathcal{F}_m \mathcal{K}^m$ be a measurable VC-subgraph class of functions satisfying Assumption 13. Suppose that $f_S(\mathbf{u}) > 0$ and $\varepsilon_{s_{i_j}, A_n} = \sigma(s_{i_j}/A_n, \mathbf{x}) \varepsilon_{i_j}$, where $\sigma(\cdot, \cdot)$ is continuous and $\{\varepsilon_i\}_{i=1}^n$ is a sequence of i.i.d. random variables with mean zero and variance 1. Moreover, suppose $n h^{m(d+1)+4} \rightarrow c_0$ for a constant c_0 . If all assumptions assumed in Theorem 3.1 hold in addition of Conditions (B2), (B3) and (B4), then the following result holds for P_S almost surely:

$$\sqrt{n h^{md} \phi^m(h)} \left[\hat{r}_n^{(m)}(\varphi, \mathbf{x}, \mathbf{u}; h_n) - r^{(m)}(\varphi, \mathbf{x}, \mathbf{u}) - B_{\mathbf{u}, \mathbf{x}} \right]$$

converges to a Gaussian process \mathbb{G}_n over $\mathcal{F}_m \mathcal{K}^m$, whose simple paths are bounded and informally continuous with respect to $\|\cdot\|_2$ -norm with covariance function given by (4.1) and where the bias term $B_{\mathbf{u}, \mathbf{x}} = O_{P, |S}(h^{2\wedge\alpha})$.

This chapter provides detailed applications, including metric learning and the multipartite ranking adapted to the spatial framework. The sampling design discussed is sufficiently applicable for numerous applications such as geostatistical and environmental monitoring applications (cf. Lahiri, 2003a). Likewise, the results can be applied to different scenarios such as $CARMA(p, q)$ random fields, for instance, see Brockwell et al., 2017. Furthermore, the obtained results can also be used to study simultaneous confidence bands for the mean of multivariate Spatio-temporal data, and investigate the inference of

the time-varying mean of univariate Spatio-temporal data.

Chapter 8. Conclusions and perspectives

In this chapter, we conclude this thesis by giving some remarks and future developments.

Remark 7.1. *The thesis is written in the form of research articles and certainly does not avoid a certain number of repetitions, but this choice of writing allows each chapter to be read as an autonomous entity.*

Chapter 2

Mathematical Background

Objective

This chapter presents some basic tools and concepts that will be used in the remainder of this thesis.

Contents

1	Mathematical Background	41
1.1	Weak convergence	41
1.2	Classes of functions	42
2	Useful notes for studying U -processes	45
2.1	Bootstrapped Empirical processes	46
2.2	The mixing notion	49
3	Locally stationary random field	51
3.1	Random field	51

1 Mathematical Background

1.1 Weak convergence

In this section, we give some notions for the weak convergence of a stochastic process. It should be noted that the weak convergence of a stochastic process is a generalization of the convergence in law from random vectors to sample paths of the stochastic process. Let $(\mathcal{X}, \mathcal{A}, \mathbb{P})$ be a probability space on which we define the sequence X_1, \dots, X_n and a collection of random variables $X = \{X(t) = X(t, \omega), \omega \in \mathcal{X}, t \in T\}$, T is an arbitrary index set. Suppose that the set T is equipped with a semi-metric ρ and (\mathbb{D}, d) is a metric space.

Definition 1.1. • *The collection $X = \{X(t) = X(t, \omega), \omega \in \mathcal{X}, t \in T\}$, is a stochastic process.*

- *An empirical process is a stochastic process based on random observations X_1, \dots, X_n .*
- *For a fixed point $\omega \in \mathcal{X}$, the map:*

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

is called the sample path of the stochastic process X .

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

Now we say that the process X_n converges weakly to a Borel measurable process X , and we write $X_n \rightsquigarrow X$ if the sample paths of X_n behave in distribution like X when $n \rightarrow \infty$. This is reflected in

$$X_n \rightsquigarrow X \iff \forall f \in C_b(\mathbb{D}) : \mathbb{E}^*(f(X_n)) \longrightarrow \mathbb{E}(f(X)), \quad (1.1)$$

where

$$C_b(\mathbb{D}) := \{f : \mathbb{D} \rightarrow \mathbb{R}; \text{ with } f \text{ continuous and bounded}\}.$$

If \mathbb{P} is the law of X then the last expression can be rewritten as

$$\mathbb{E}^* f(X_n) \rightarrow \int f(x) d\mathbb{P}(x), \text{ for every } f \in C_b(\mathbb{D}).$$

However, in practice, the latter formulation is not easy to handle. An equivalent theorem is given in Theorem 2.1 in Kosorok, 2008.

Theorem 1.2 (Kosorok, 2008). *The stochastic process X_n converges weakly to a tight stochastic process X in $\ell^\infty(T)$, if and only if:*

(i) *For all finite $\{t_1, \dots, t_k\} \subset T$, the finite-dimensional distribution of $\{X_n(t_1), \dots, X_n(t_k)\}$ converges to that of $\{X(t_1), \dots, X(t_k)\}$;*

(ii) *There exists a semi-metric ρ for which T is totally bounded such that for all $\varepsilon > 0$:*

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P}^* \left\{ \sup_{s, t \in T: \rho(s, t) < \delta} |X_n(t) - X_n(s)| > \varepsilon \right\} = 0. \quad (1.2)$$

Beneficial results are the continuous mapping theorem and Slutsky's Theorem:

Theorem 1.3. (Continuous mapping) *Let $g : \mathbb{D} \mapsto \mathbb{Y}$ be continuous at all points in $\mathbb{D}_0 \subset \mathbb{D}$, where \mathbb{D} and \mathbb{Y} are metric spaces. Then if $X_n \rightsquigarrow X$ in \mathbb{D} , with $\mathbb{P}_*(X \in \mathbb{D}_0) = 1$, then $g(X_n) \rightsquigarrow g(X)$.*

Theorem 1.4. (Slutsky's theorem) *Suppose $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow c$, where X is separable and c is a fixed constant. Then the following are true:*

(i) $(X_n, Y_n) \rightsquigarrow (X, c)$.

(ii) *If X_n and Y_n are in the same metric space, then $X_n + Y_n \rightsquigarrow X + c$.*

(iii) *Assume in addition that the Y_n are scalars. Then whenever $c \in \mathbb{R}$, $Y_n X_n \rightsquigarrow cX$. Also, whenever $c \neq 0$, $X_n/Y_n \rightsquigarrow X/c$.*

Generally, when dealing with empirical processes, the index set $T = \mathcal{F}$ is a class of measurable functions. That is what, in the following section, we give some definitions and examples concerning these classes.

1.2 Classes of functions

This section is devoted to entropy, a fundamental tool for the empirical process. The primary use of such entropy calculus in this thesis is for establishing the rate of convergence U -estimators as discussed in Chapter (1) and evaluating whether the class of functions \mathcal{F} is Glivenko-Cantelli and/or Donsker or neither. There are several additional uses of entropy bounds, we refer the interested reader to the monographs of A. W. van der Vaart et al., 1996 and Kosorok, 2008; see also Pakes et al., 1989.

Definition 1.5. *An envelope function of a class \mathcal{F} is any function $x \mapsto F(x)$ such that*

$$|f(x)| \leq F(x),$$

for every x and f .

Definition 1.6. 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 1.7. The subgraph of a function $f : \mathcal{X} \mapsto \mathbb{R}$ is the subset of $\mathcal{X} \times \mathbb{R}$ given by

$$\{(x, t) : t < f(x)\}.$$

Definition 1.8. A class of functions \mathcal{F} is called a VC-subgraph class if the collections of all subgraphs of the functions in \mathcal{F} form a VC-class of sets in $\mathcal{X} \times \mathbb{R}$.

Example 12. let $\mathcal{C} = \{C \subset \mathcal{X}\}$ and $\mathcal{F}(\mathcal{C}) = \{\mathbb{1}_{\{X \in C\}, C \in \mathcal{C}}\}$. Then $\mathcal{F}(\mathcal{C})$ is a VC-subgraph class if and only if \mathcal{C} is a VC class of sets.

Definition 1.9. A class \mathcal{F} of measurable functions is \mathbb{P} -measurable if the map

$$(x_1, \dots, x_n) \mapsto \sup_{f \in \mathcal{F}} \left\| \sum_{i=1}^n e_i f(x_i) \right\|$$

is measurable for all $(e_1, \dots, e_n) \in \mathbb{R}^n$.

A stronger but easier-to-verify measurability assumption is pointwise measurability, defined as:

Definition 1.10. 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 .

Definition 1.11. (Covering number). Let $(\mathcal{F}, \|\cdot\|)$ be a subset of a normed space of real functions f on some set. The covering number $N(\varepsilon, \mathcal{F}, \|\cdot\|)$ is the minimal number of balls $\{g : \|g - f\| < \varepsilon\}$ of radius ε needed to cover the set \mathcal{F} . The entropy (without bracketing) is the logarithm of the covering number. Define

$$J(\delta, \mathcal{F}) = \sup_Q \int_0^\delta \sqrt{1 + \log N(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q))} d\varepsilon,$$

where the supremum is taken over all finitely discrete probability measures Q with $\|F\|_{Q,2} > 0$.

Definition 1.12. (Bracketing number). Given two functions l and u , the bracket $[l, u]$ is the set of all functions f with $l \leq f \leq u$. An ε bracket is a bracket $[l, u]$ with $\|l - u\| < \varepsilon$. The bracketing number $N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|)$ is the minimum number of ε brackets needed to cover \mathcal{F} . The entropy with bracketing is the logarithm of the bracketing number. For a given norm $\|\cdot\|$, define bracketing integral of a class of functions \mathcal{F} as

$$J_{[]}(\delta, \mathcal{F}, \|\cdot\|) = \int_0^\delta \sqrt{1 + \log N_{[]}(\varepsilon \|F\|, \mathcal{F}, \|\cdot\|)} d\varepsilon.$$

The next lemma presents a link between the covering and the packing numbers of a functions class \mathcal{F} .

Lemma 1.1. *For a class of functions \mathcal{F} we have:*

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

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

Example 13. *The set \mathcal{F} of all indicator functions $\mathbb{1}_{\{(-\infty, t]\}}$ of cells in \mathbb{R} satisfies :*

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

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

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

For more details and discussion on this example, refer to Example 2.5.4 of A. W. van der Vaart et al., 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 $(1/\varepsilon)$, see Theorem 9.19 of Kosorok, 2008.

Example 14. *(Classes of functions that are Lipschitz in a parameter, Section 2.7.4 in A. W. van der Vaart et al., 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 A. W. van der Vaart et al., 1996 and Lemma 9.18 of Kosorok, 2008, it follows, for any norm $\|\cdot\|_{\mathcal{F}}$ on \mathcal{F} , that :

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

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

Example 15. *Let us consider as an example the classes of functions that are smooth up to order α defined as follows, see Section 2 of A. W. van der Vaart et al., 1996. For $0 < \alpha < \infty$ let $\lfloor \alpha \rfloor$ 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 :

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

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 :

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

Note that for $\alpha \leq 1$, this class consists of bounded functions $\varphi(\cdot)$ that satisfy a Lipschitz condition. Kolmogorov et al., 1959 computed the entropy of the classes of $C_F^\alpha(\mathcal{X})$ for the uniform norm. As a consequence of their results A. W. van der Vaart et al., 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 $\varepsilon > 0$:

$$\log N_{[\]}(\varepsilon F \gamma(\mathcal{X}), C_F^\alpha(\mathcal{X}), L_2(\gamma)) \leq K \left(\frac{1}{\varepsilon}\right)^{d/\alpha},$$

$N_{[\]}$ is the bracketing number, refer to Definition 2.1.6 of A. W. van der Vaart et al., 1996 and we refer to Theorem 2.7.1 of A. W. van der Vaart et al., 1996 for a variant of the last inequality. By Lemma 9.18 of Kosorok, 2008, we have :

$$\log N(\varepsilon F \gamma(\mathcal{X}), C_F^\alpha(\mathcal{X}), L_2(\gamma)) \leq K \left(\frac{1}{\varepsilon}\right)^{d/\alpha}.$$

2 Useful notes for studying U -processes

First, we begin by introducing some notation needed in this thesis. Let X_1, \dots, X_n are i.i.d. \mathbb{P} on \mathcal{X} . Then the empirical measure \mathbb{P}_n is defined by

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

where δ_x denotes the Dirac measure at x . For each $n \geq 1$, \mathbb{P}_n denotes the random discrete probability measure which puts mass $1/n$ at each of the n points X_1, \dots, X_n . For a real valued function f on \mathcal{X} , we write

$$\mathbb{P}_n(f) := \int f d\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^n f(X_i).$$

If \mathcal{F} is a class of functions defined on \mathcal{X} , then $\{\mathbb{P}_n(f) : f \in \mathcal{F}\}$ is the empirical measure indexed by \mathcal{F} . Let us assume that

$$\mathbb{P}f := \int f d\mathbb{P},$$

exists for each $f \in \mathcal{F}$. The empirical process \mathbb{G}_n is defined by

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

and the collection of random variables $\{\mathbb{G}_n(f) : f \in \mathcal{F}\}$ as f varies over \mathcal{F} is called the empirical process indexed by \mathcal{F} . We define the following quantity :

$$\|\mathbb{G}_n\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |\mathbb{G}_n(f)|.$$

Definition 2.1. A class \mathcal{F} of measurable functions $f : \mathcal{X} \rightarrow \mathbb{R}$ with $\mathbb{P}|f| < \infty$ for every $f \in \mathcal{F}$ is called *Glivenko-Cantelli (GC)* if

$$\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |\mathbb{P}_n f - \mathbb{P}f| \rightarrow 0, \quad \text{in probability (or almost surely).}$$

Definition 2.2. A class \mathcal{F} of measurable functions $f : \mathcal{X} \rightarrow \mathbb{R}$ is *Donsker* if the empirical process $\{\mathbb{G}_n f : f \in \mathcal{F}\}$ indexed by \mathcal{F} converges in distribution in the space $\ell^\infty(\mathcal{F})$ to a tight random element.

Theorem 2.3 (Entropy control with bracketing number, A. W. van der Vaart et al., 1996). Let \mathcal{F} be a class of measurable functions with envelope F . Then

$$E[\|\mathbb{G}_n(f)\|_{\mathcal{F}}^*] \leq K J_{[]} (1, \mathbb{F}, L_2(p)) \|F\|_{\mathbb{P}, 2},$$

where K does not depend on \mathcal{F} or F .

2.1 Bootstrapped Empirical processes

Let \mathbb{P}_n be the empirical measure of an i.i.d. sample X_1, \dots, X_n from a probability measure \mathbb{P} . Given the sample values, let X_1^*, \dots, X_n^* be an i.i.d. sample from \mathbb{P}_n . The bootstrap¹ empirical measure and process are, respectively, defined by

$$\hat{\mathbb{P}}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i^*},$$

¹The bootstrap is the statistical procedure that models sampling from a population by the process of resampling from the sample.

and

$$\widehat{\mathbb{G}}_n = \sqrt{n} \left(\widehat{\mathbb{P}}_n - \mathbb{P}_n \right).$$

Giné et al., 1990 proved the following result, for a class of function \mathcal{F} with its envelope F . Here we will consider that the class \mathcal{F} is the collection of indicator functions of sets of the form $[0, c]$, $0 < c \leq 1$. Then under measurability restriction on \mathcal{F} , we have:

$$\mathbb{G}_n \rightsquigarrow \mathbb{G} \quad \text{and} \quad \mathbb{P}F^2 < \infty,$$

is equivalent to

$$\widehat{\mathbb{G}}_n \rightsquigarrow \mathbb{G} \quad \text{for almost all data sequences} \quad X_1, X_2, \dots,$$

where \mathbb{G} is some tight Brownian bridge, and the weak convergence is in $\ell^\infty(\mathcal{F})$. This result is proved "in probability" by the same authors, who settled questions about the validity of Efron's bootstrap in many situations. We can remark that the bootstrap empirical measure given before can be expressed as

$$\widehat{\mathbb{P}}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i^*} = \frac{1}{n} \sum_{i=1}^n \xi_{ni} \delta_{X_i},$$

where ξ_{ni} is the number of times that X_i is "redrawn" from the original sample. As observed by Efron, 1982, Section 2.9, pages 17-72, this suggests that there are, in fact, not just one way but several ways to bootstrap; and this is the idea of the exchangeable-weighted bootstrap. Let $\mathbf{W} = \{W_{ni}, i = 1, 2, \dots, n, n = 1, 2, \dots\}$ are a triangular array defined on the probability space $(\mathcal{Z}, \mathcal{E}, \widehat{\mathbb{P}})$. Let $W_n \equiv (W_{n1}, \dots, W_{nn})$ be an exchangeable vector of nonnegative weights which sum to n . Then the exchangeably weighted bootstrap empirical measure is defined by

$$\widehat{\mathbb{P}}_n = \frac{1}{n} \sum_{i=1}^n W_{ni} \delta_{X_i},$$

with corresponding bootstrap empirical process

$$\widehat{\mathbb{G}}_n = \sqrt{n} \left(\widehat{\mathbb{P}}_n - \mathbb{P}_n \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (W_{ni} - 1) \delta_{X_i}. \quad (2.1)$$

The formulation of the weighted bootstrap was originally initiated by Lo, 1993. Mason et al., 1992 and Præstgaard et al., 1993 established sufficient conditions on the weights \mathbf{W} for the exchangeable weighted bootstrap to work asymptotically, where they suggested the following general conditions on \mathbf{W} .

(B.1) The vectors $W_n = (W_{n1}, W_{n2}, \dots, W_{nn})^T$ are exchangeable for all $n = 1, 2, \dots$, i.e.,

for any permutation $\pi = (\pi_1, \dots, \pi_n)$ of $(1, 2, \dots, n)$, the joint distribution of

$$\pi(W_n) = (W_{n\pi_1}, W_{n\pi_2}, \dots, W_{n\pi_n})^T$$

is the same as that of W_n .

(B.2) $W_{ni} \geq 0$ for all n, i and $\sum_{i=1}^n W_{ni} = n$ for all n .

(B.3) The following $L_{2,1}$ norm of W_{n1} is uniformly bounded:

$$R_n = \int_0^\infty \sqrt{\widehat{\mathbb{P}}(W_{n1} \geq u)} du \leq K < \infty.$$

(B.4) $\lim_{\lambda \rightarrow \infty} \limsup_{n \rightarrow \infty} \sup_{t \geq \lambda} t^2 \widehat{\mathbb{P}}\{W_{n1} \geq t\} = 0$.

(B.5) $\frac{1}{n} \sum_{i=1}^n (W_{ni} - 1)^2 \rightarrow c^2 > 0$ in $\widehat{\mathbb{P}}$ -probability.

We note that Efron's nonparametric bootstrap (or multinomial bootstrap) corresponds to the choice of the weights

$$W_n \sim \text{Multinomial}(n; n^{-1}, \dots, n^{-1})$$

for which conditions **(B.1)**–**(B.5)** are satisfied. In general, in order to satisfy the conditions **(B.3)**–**(B.5)** we have to impose some moment conditions on W_{ni} , see their Lemma 3.1. The other sampling schemes that satisfy conditions **(B.1)**–**(B.5)**, include *Bayesian bootstrap*, *Multiplier bootstrap*, *Double bootstrap*, and *Urn bootstrap*. These examples are sufficient to show that conditions **(B.1)**–**(B.5)** are very general. It is worth noticing that the value of c in **(B.5)** is independent of the sample at hand and depends only on the chosen resampling method, e.g., $c = 1$ for the nonparametric bootstrap and Bayesian bootstrap, whereas $c = \sqrt{2}$ for the double bootstrap. A more precise discussion of this general formulation of the bootstrap and further details can be found in Mason et al., 1992, Præstgaard et al., 1993, Barbe et al., 1995, A. W. van der Vaart et al., 1996, §3.6.2., p. 353, Kosorok, 2008, §10. p. 179, Cheng et al., 2010. The interested reader may refer to Billingsley, 1968, Aldous, 1985, and Kallenberg, 2002 for excellent general coverage of the theory of exchangeability. One could claim that the general first-order limit theory for the bootstrap was known to Laplace by about 1810 (since Laplace developed one of the earliest general central limit theorems); and that second-order property were developed by Chebyshev at the end of the 19th Century, as mentioned by Peter Hall ². In 1923, Hubback began a series of crop trials in the Indian states of Bihar and Orissa, where he developed spatial sampling schemes. In 1927 he published an account of his work

²<http://www.cms.zju.edu.cn/conference/2005/zlx/peter.pdf>

in Bulletin No. 166 of the Indian Agricultural Research Institute. Notice that the idea of bootstrap appeared in different forms in Mahalanobis, 1946, Quenouille, 1949, 1956, Tukey, 1958, Simon, 1969, Chapters 23-25 and Maritz et al., 1978.

We assume further that the collection \mathcal{F} possesses enough measurability for randomization with i.i.d. multipliers to be possible and the usual Fubini's theorem can be used freely; such a set of conditions is $\mathcal{F} \in \text{NLDM}(\mathbb{P})$ (Nearly Linearly Deviation Measurable), and $\mathcal{F}^2, \mathcal{F}'^2 \in \text{NLSM}(\mathbb{P})$ (Nearly Linearly Supremum Measurable) in the terminology of Giné et al., 1990. Here \mathcal{F}^2 and \mathcal{F}'^2 denote the classes of squared functions and squared differences of functions from \mathcal{F} , respectively. When all of these conditions hold, we write $\mathcal{F} \in \text{M}(\mathbb{P})$. It is known that $\mathcal{F} \in \text{M}(\mathbb{P})$ if \mathcal{F} is countable, or if the empirical processes \mathbb{G}_n are stochastically separable, or if \mathcal{F} is image admissible Suslin (see Giné et al., 1990, p. 853 and 854). The following Præstgaard et al., 1993's result concerns a central limit theorem in probability for the bootstrap empirical process as given in (2.1) indexed by the class \mathcal{F} .

Theorem 2.4. *Let $\mathcal{F} \in \text{M}(\mathbb{P})$ be a class of $L_2(\mathbb{P})$ functions, and let \mathbf{W} be a triangular array of bootstrap weights satisfying assumptions 1.-5.. Then*

$$\mathcal{F} \text{ is } \mathbb{P} - \text{Donsker}$$

implies that

$$\widehat{\mathbb{G}}_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n (W_{nj} - 1) \delta_{X_j} \rightsquigarrow c\mathbb{G} \text{ in } l^\infty(\mathcal{F}) \text{ in probability,}$$

where c is given by assumption A5.

If, in addition, the envelope function F is square integrable, then the result holds almost everywhere.

These results for bootstrapped empirical processes can be applied to many kinds of bootstrapped estimators since most estimators can be expressed as functional empirical processes.

2.2 The mixing notion

To estimate the dependence between two σ -algebras \mathcal{A} and \mathcal{B} defined on a probability space (Ω, \mathcal{D}, P) , we are going to use some classical measures of dependence (c.f. Bradley, 2005) :

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup |P(A \cap B) - P(A)P(B)|, \quad A \in \mathcal{A}, \quad B \in \mathcal{B}, \quad (2.2)$$

$$\beta(\mathcal{A}, \mathcal{B}) := \sup \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J |P(A_i \cap B_j) - P(A_i)P(B_j)|, \quad (2.3)$$

$$\phi(\mathcal{A}, \mathcal{B}) := \sup |P(B|A) - P(A)P(B)|, \quad A \in \mathcal{A}, B \in \mathcal{B}, P(A) > 0, \quad (2.4)$$

$$\psi(\mathcal{A}, \mathcal{B}) := \sup |P(A \cap B) - P(A)P(B)| / P(A)P(B), \quad A \in \mathcal{A}, B \in \mathcal{B}, \quad (2.5)$$

$$\rho(\mathcal{A}, \mathcal{B}) := \sup \left| \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} \right|, \quad X \in L_2(\mathcal{A}), Y \in L_2(\mathcal{B}), \quad (2.6)$$

where for $\beta(\mathcal{A}, \mathcal{B})$ the sup is taken over all pairs of finite partitions $\{A_1, \dots, A_I\}$ and $\{B_1, \dots, B_J\}$ of Ω such that $A_i \in \mathcal{A}$ for all $1 \leq i \leq I$ and $B_j \in \mathcal{B}$ for all $1 \leq j \leq J$. The coefficient in (2.2) is called strong mixing or α -mixing coefficient, that in equation (2.3) is called absolute regularity or β -mixing. In equation (2.4), the coefficient is the uniform mixing coefficient or ϕ -mixing coefficient, and in equation (2.5), the coefficient is ψ -mixing. Let $\mathbf{Z}_1, \mathbf{Z}_2, \dots$ be a stationary sequence of random variables defined on the probability space (Ω, \mathcal{D}, P) and let:

$$\sigma_J^L := \sigma(\mathbf{Z}_i, J \leq i \leq L).$$

Assume that Ξ denotes any of the dependence coefficients $\alpha, \beta, \phi, \psi, \rho$. A c -mixing sequence is defined by requiring the Ξ -mixing coefficient Ξ_k to satisfy :

$$\Xi_k := \sup_{J \in \mathbb{Z}} c(\sigma_{-\infty}^J, \sigma_{J+k}^\infty) \xrightarrow{k \rightarrow \infty} 0.$$

The following proposition gives the relations between the different mixing.

Proposition 2.5. *we have:*

$$\begin{aligned} 2\alpha(\mathcal{A}, \mathcal{B}) &\leq \beta(\mathcal{A}, \mathcal{B}) \leq \phi(\mathcal{A}, \mathcal{B}) \leq \frac{1}{2}\psi(\mathcal{A}, \mathcal{B}), \\ 4\alpha(\mathcal{A}, \mathcal{B}) &\leq \rho(\mathcal{A}, \mathcal{B}) \leq 2\phi^{1/2}(\mathcal{A}, \mathcal{B})\phi^{1/2}(\mathcal{B}, \mathcal{A}), \\ \rho(\mathcal{A}, \mathcal{B}) &\leq \phi(\mathcal{A}, \mathcal{B}). \end{aligned}$$

From the previous proposition, we can deduce the following schema:

$$\psi - \text{mixing} \begin{matrix} \Rightarrow \\ \neq \end{matrix} \phi - \text{mixing} \Rightarrow \left\{ \begin{array}{l} \beta - \text{mixing} \begin{matrix} \Rightarrow \\ \neq \end{matrix} \alpha - \text{mixing} \\ \uparrow \Downarrow \\ \rho - \text{mixing} \begin{matrix} \Rightarrow \\ \neq \end{matrix} \alpha - \text{mixing} \end{array} \right. \quad (2.7)$$

In this thesis, we are mainly interested in establishing some properties of β -mixing conditional U -processes.

3 Locally stationary random field

3.1 Random field

The notation “random field” can be heard in different domains. The most straightforward clarification in the statistical field, a random field is a stochastic process usually taking value in Euclidean space and defined over a parameter space of dimensionality at least 1. The literature on random fields is not rich like that of stochastic process; we can cite Dougherty, 1999, Yaglom, 1987, Adler et al., 2007 and recently Christakos, 2012.

Random’s fields concept can be defined similarly to that of stochastic processes. In this case, each random variable x_1, x_2, \dots, x_n is associated with the points $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n$ in the space \mathbb{R}^n . A random spatial field can be considered as a function of events $\omega \in \Omega$, where Ω is the sample space, and also as a function of the spatial position $\mathbf{s} \in R^n$, that is, $x(\mathbf{s}) = x(\omega, \mathbf{s})$. Let $\mathbf{x}(\mathbf{s})$ be the vector random field which represents the set of random spatial fields $x_1(\mathbf{s}), x_2(\mathbf{s}), \dots, x_m(\mathbf{s})$, that is,

$$\mathbf{x}(\mathbf{r}) = [x_1(\mathbf{s}), x_2(\mathbf{s}), \dots, x_m(\mathbf{s})]^T.$$

The distribution function of a vector random spatial field is then defined as:

$$F_{\mathbf{x}}(\mathbf{x}, \mathbf{s}) = P(\{\omega : \mathbf{x}(\mathbf{s}) \leq \mathbf{x}; \mathbf{s} \in \mathbb{R}^n\}).$$

Remember again that random fields extend the concept of a stochastic process. A stochastic process is a random field for which the spatial argument $\mathbf{r} \in R^n$, is introduced for $n = 1$ and $\mathbf{s} \rightarrow s \rightarrow t$ so that the random variable becomes $\mathbf{x}(t)$, as before.

The distribution function is related to the probability density through the expression

$$p_{\mathbf{x}}(\mathbf{x}, \mathbf{s}) = \frac{\partial^n F_{\mathbf{x}}(\mathbf{x}, \mathbf{s})}{\partial x^n},$$

and consequently

$$F_{\mathbf{x}}(\mathbf{x}, \mathbf{r}) = \int_{-\infty}^x p_{\mathbf{x}}(\mathbf{x}', \mathbf{s}) d\mathbf{x}',$$

and in fact, such a random field is said to be a white field (analogously to the white process seen previously).

Basically, all the concepts defined for random processes can be generalized for spatial random fields:

- Markovian process \rightarrow Markovian field;
- Gaussian process \rightarrow Gaussian field;
- White process \rightarrow white field.

We start with defining a random field more formally using random variables.

Definition 3.1 (Random field). *Let (Ω, \mathcal{F}, P) be a probability space. Given $d, \delta \in \mathbb{N}$, we say that a function X :*

$$X : \Omega \times \mathbb{Z}^d \rightarrow \mathbb{R}^\delta$$

is a Random Field if and only if $\forall t \in \mathbb{Z}^d$, the function:

$$\begin{aligned} X_t : \Omega &\rightarrow \mathbb{R}^\delta \\ \omega &\rightarrow X(\omega, t) \end{aligned}$$

is $\mathcal{F} - \mathcal{B}(\mathbb{R}^\delta)$ -measurable; where \mathcal{B} denotes the Borel σ -algebra.

Definition 3.2. *A random variable \mathcal{X} is called a functional variable (f.v.) if it takes values in an infinite dimensional space (or functional space). An observation χ of \mathcal{X} is called functional data.*

Definition 3.3 (Strict stationarity). *A random sequence $(X_n)_{n \in \mathbb{Z}}$ is said to be strictly stationary if, for each $k \geq 0$, the distribution of the vector (X_l, \dots, X_{l+k}) does not depend on $l \in \mathbb{Z}$.*

Definition 3.4 (Weak stationarity). *A random sequence $(X_n)_{n \in \mathbb{Z}}$ is second order stationary if $\mathbb{E}X_l^2 < \infty$ and if only:*

$$\mathbb{E}X_l = \mathbb{E}X_0, \quad \text{and} \quad \text{Cov}(X_l, X_{k+l}) = \text{Cov}(X_0, X_k), \quad \text{for each } l, k \in \mathbb{Z}.$$

Definition 3.5. *A function K from \mathbb{R} into \mathbb{R}^+ such that $\int K = 1$ is called a kernel of type I if there exist two real constants $0 < C_1 < C_2 < \infty$ such that:*

$$C_1 1_{[0,1]} \leq K \leq C_2 1_{[0,1]}.$$

ii) A function K from \mathbb{R} into \mathbb{R}^+ such that $\int K = 1$ is called a kernel of type II if its support is $[0, 1]$ and if its derivative K' exists on $[0, 1]$ and satisfies for two real constants $-\infty < C_2 < C_1 < 0$:

$$C_2 \leq K' \leq C_1.$$

Chapter 3

Renewal type bootstrap for increasing degree U -process of a Markov chain

This chapter develops the content of an accepted article published in *Journal of Multivariate Analysis* with the required modifications to fit this thesis manuscript.

Objective

In this chapter, we investigate the uniform limit theory for a U -statistic of increasing degree, also called an infinite-degree U -statistic. Infinite-degree U -statistics (IDUS) (or infinite-order U -statistics (IOUS)) are useful tool for constructing simultaneous prediction intervals that quantify the uncertainty of several methods such as subbagging and random forests. The stochastic process based on collections of U -statistics is referred to as a U -process, and if the U -statistic is of infinite-degree, we have an infinite-degree U -process. C. Heilig et al., 2001 provided conditions for the pointwise asymptotic theory for the infinite-degree U -processes. The main purpose here is to extend their findings to the Markovian setting. The second aim is to provide the uniform limit theory for the renewal bootstrap for the infinite-degree U -process, which is of its own interest. The main ingredients are the decoupling technique combined with symmetrization techniques of C. Heilig et al., 2001 to obtain uniform weak law of large numbers and functional central limit theorem for the infinite-degree U -process. The results obtained in this chapter are, to our knowledge, the first known results on the infinite-degree U -process in the Markovian setting.

Contents

1	Introduction	55
2	Preliminaries and notation	58
3	Infinite degree U -statistics and Hoeffding decomposition	60
4	Asymptotic theory: first-order projection	62

4.1	Strong law of large numbers	63
4.2	Weak convergence theorem	63
5	Increasing degree U -processes: asymptotic theory	65
6	Bootstrap of the increasing degree U -processes	67
6.1	Asymptotic validity of bootstrapped Markovian U-statistics	68
7	Examples and Applications	71
8	Mathematical developments	77
8.1	Proof of Theorem 5.1:	83
8.2	Proof of Theorem 5.2:	88
8.3	Proof of Theorem 6.1:	89

1 Introduction

Let $X = (X_n)_{n \in \mathbb{N}}$ be an homogeneous Markov chain, defined on a measurable space (E, \mathcal{E}) , where \mathcal{E} is a separable σ -algebra. The infinite or increasing degree U -statistics can be defined as :

$$U_n(h_m) = \frac{(n - m_n)!}{n!} \sum_{\mathbf{i} \in I(n, m_n)} h_m(X_{i_1}, \dots, X_{i_{m_n}}), \quad (1.1)$$

where $I(n, m_n) = \{\mathbf{i} = (i_1, \dots, i_{m_n}) : 1 \leq i_j \leq n \text{ and } i_j \neq i_r \text{ if } j \neq r\}$, is the set of all m_n -tuples of different integers between 1 and n . Common examples of estimators for a finite degree are the empirical variance, Gini's mean difference, or Kendall's rank correlation coefficient, while a classical test based on a U -statistic is Wilcoxon's signed rank test for the hypothesis of the location at zero (see, e.g., A. W. van der Vaart, 1998, Example 12.4). Halmos, 1946; Hoeffding, 1948; v. Mises, 1947, who provided (amongst others) the first asymptotic results for the case that the underlying random variables are independent and identically distributed. U -statistics of infinite-order (IOUS). Frees uses the term infinite order U -statistics (IOUS), and C. Heilig et al. used the term infinite degree U -statistics (IDUS) in keeping with Hoeffding's original use of the terms degree and order, so the terms IOUS and IDUS indicate the same object. have attracted renewed interest in the recent statistics and machine learning literature in relation to uncertainty quantification for Breiman's bagging Breiman, 1996 and random forests Breiman, 2001. In such applications, the tree-based prediction rules can be thought of as U -statistics with deterministic and random kernels, respectively, and their order corresponds to the sub-sample size of the training data Mentch et al., 2016. Statistically, the subsample size m used to build each tree needs to increase with the total sample size n to produce reliable predictions. As a leading example, we mention the construction of simultaneous prediction intervals for a version of random forests discussed in Mentch et al., 2016; Y. Song et al., 2019. We mention that several classes of statistics share commonalities with IDUS, including infinite-degree V -statistics, partial-sum U -processes [there are two perspectives on U -processes: 1) they are infinite-dimensional versions of U -statistics (with one kernel); 2) they are stochastic processes that are nonlinear generalizations of empirical processes] symmetric statistics, and elementary symmetric polynomials. The computational intractability of the IDUS is a significant constraint when the sample size and/or order are large. Extensive literature have treated the theory of U -statistics of fixed order, for instance, see among many others Arcones et al., 1993; Bouzebda et al., 2019a; Bouzebda and Soukariéh, 2022a; Bouzebda et al., 2021; Bouzebda and Nemouchi, 2020, 2022; Bouzebda and Nezzal, 2022; de la Peña et al., 1999; Hoeffding, 1948; A. J. Lee, 1990; Serfling, 1980; Soukariéh et al., 2022. However, the U -statistics of infinite degree

have been studied for the first time by Frees, 1989, he called it the infinite order U -statistics. In his work, he aimed to show if U_n is still an unbiased estimator, even more, a desirable estimator of a parameter of interest, say θ , and what conditions are needed to be satisfied for the asymptotic convergences. Besides, he presented many examples to illustrate the usefulness of the IDUS. The authors of Bickel et al., 1997; Bouzebda, Elhattab, and Ferfache, 2022; Politis et al., 1994a consider statistics based on resampling procedures. Among the procedures summarized in Bickel et al., 1997, the m out of n bootstrap is an infinite-degree V -statistic and the n choose m bootstrap is an infinite-degree U -statistic. Historically speaking, Kohatsu-Higa, 1991 generalized the study of Frees, 1989 by representing the weak convergence of infinite U -process in the degenerate case. Precisely, he used a technique from his previous work Kohatsu-Higa, 1990. Kohatsu-Higa in Rempala et al., 1999 discussed the projection's conditions of the asymptotic behavior of U -statistics from the finite degree to the infinite degree. Kohatsu-Higa Kohatsu-Higa, 1991 revisited the results related to those of Dynkin et al., 1983 for symmetric statistics. C. Heilig et al. C. Heilig et al., 2001 treated the subject of infinite degree, as well as many previous works for both, like C. M. Heilig, 1997; C. M. Heilig et al., 1998. On the other hand, Shieh in Shieh, 1994 has proposed an analogy between infinite-degree V -statistics and the IDUS, which may have more efficient calculation techniques than U -statistics, even though they are not typically unbiased. A class of deterministic weighted U -statistics was proposed in Pinheiro et al., 2009, called quasi U -statistics. The kernels of this U -statistics are degenerate of degree two, and then they generalized their investigation to the kernels of degree m in Pinheiro et al., 2011 and to an infinite order degenerate kernels in Pinheiro et al., 2014. The last cited references' major concern is the limiting laws' characterization. But the problem in both cases, finite and infinite order, resides in the fact that the limiting distributions of the U -processes, or their functionals, are rather complicated, which does not permit explicit computation in practice. To overcome that difficulty, many authors suggested and studied solutions like Arcones et al., 1992; Q. Han, 2022; Korolyuk et al., 1990 by investigating what is called bootstrapped U -statistics. Then using the bootstrap approach, we can simulate the limiting distribution of U -statistics. Those last references used Efron's bootstrap developed in Efron, 1979, which is an important non-parametric technique to solve the problem of complex limiting distribution. Nevertheless, as it is well known, Efron's technique is assigned to the independent and identically distributed random variable. As a result, different types of processes emerge in response to an expanded number of statistical criteria while adhering to the core principles of the i.i.d. (independent and identically distributed) bootstrap. D. A. Freedman, 1984 gives for the first time an extension of Efron's technique to the dependent framework, followed by Bose, 1988; Carlstein, 1986a; Rajarshi, 1990 who offer different non-parametric and semi-parametric methods or models of bootstrap concerning the dependent data. Over and above, as the Markov chain model is an important tool for statisticians, engineers, data scientists, and

econometricians, the investigation of bootstrap in this field is relevant. This investigation can be found in many papers like in Athreya et al., 1992a, 1992b; Datta et al., 1993. In the dependent setup, the most results of U -statistics are proved under the mixing conditions; see the work of Borovkova et al., 2002; Borovkova et al., 2001; Dehling et al., 1989, 1991; Denker et al., 1983; Hsing et al., 2004. However, in the stationary setup, it can be referred to Hsing et al., 2004; Yoshihara, 1976. Note that Y. Song et al., 2019 have investigated an incomplete version of the infinite order U -statistics with a random kernel. They derived sub-sampling procedures for making inferences on the parameter of interest with a class of high-dimensional random kernels of diverging orders. In our approach, we will use the same strategy as our previous work Bouzebda and Soukarieh, 2022a, i.e., no mixing conditions are involved. We are based on the strong property of Markov chains, which allows us to divide the chain into i.i.d. blocks, based on Bertail et al., 2011a. In addition to the previously mentioned work, Bertail et al., 2011a, the authors in Duchemin et al., 2020; Fort et al., 2012; Harel et al., 1989 remain within the framework of Markov chain U -statistics while showing other assumptions and properties concerning this estimator. The primary purpose of the present chapter is to generalize the work of U -statistics and U -process for the case where the degree m converges to infinity or, in other words, increases with the size of the sample in the framework of Markov chains. We generalize the work of C. Heilig et al., 2001 to the Harris recurrent Markov chain by using the renewal properties of this chain. In addition, we study the limit theorems of the U -statistic estimators in a class of functions that keep some properties to find the desired results. We also consider the bootstrapped version of the U -process, which is of its own interest. The context of the present chapter has, to the best of our knowledge, not been considered so far in the literature, which is substantially more complicated than the independence framework. Recall that we use the abbreviations IDUS or IOUS for increasing (or infinite) degree/order U -statistics.

This chapter is organized as follows. In Section 2, we summarize notation, definitions, and properties related to the Markov chain and U -statistics with a kernel of many arbitrary arguments belonging to a class of functions. Section 3 reviews U -statistics with infinite degree, including the Hoeffding decomposition. In Section 4, we apply the empirical process technique to the first-order projection for the U -statistics based on the Harris Markov chain. The core of this chapter is the asymptotic theory of IDUS represented in Section 5 and the asymptotic theory of bootstrapped IDUS in Section 6. Then, Section 7 collects some examples and potential applications. Finally, to avoid interrupting the flow of the presentation, all mathematical proofs are presented in Section 8.

2 Preliminaries and notation

In this part and to facilitate the notation and indices of U -statistics and Markov chain, we will properly define each term used in this chapter. Let $X = (X_n)_{n \in \mathbb{N}}$ be an homogeneous Markov chain, i.e., the chain has a stationary transition probabilities, defined on a measurable space (E, \mathcal{E}) , where \mathcal{E} is a separable σ -algebra. Let $\pi(x, dy)$ be the transition probability, $\nu = \nu(i)_{i > 0}$ denotes the initial probability. Therefore, we will denote by \mathbb{P}_ν or just \mathbb{P} the probability measure determined by $\mathbb{P} = (\pi, \nu)$. Likewise, \mathbb{E} will denote the integration with respect to \mathbb{P} . Under the same assumptions in Bouzebda and Soukarieh, 2022a, we will assume that $(X_n)_{n \geq 0}$ be an Harris-recurrent Markov chain with a unique invariant measure μ . We will also suppose that the chain is regenerative and there exists a set \mathbf{A} called an atom. We point out that the atomic setting includes the whole class of Harris recurrent Markov chains with a countable state space (for which any recurrent state is an accessible atom) and many other specific Markovian models widely used by practitioners for modeling storage and queuing systems, for instance, refer to Sect. 2.4 in S. P. Meyn et al., 1993 and Asmussen, 1987 for an overview. visited infinite times almost surely. See Douc et al., 2018; Revuz, 1984, for an exhaustive treaty of the basic concepts of the Markov chain theory. The existence of recurrent atoms \mathbf{A} gives an immediate consequence for constructing a regenerative extension of this chain. That is why we have to define:

- The hitting times

$$\mathcal{T}_0 := \inf\{n \geq 0 : X_n \in \mathbf{A}\}, \quad \mathcal{T}_j := \inf\{n > \mathcal{T}_{j-1} : X_n \in \mathbf{A}\};$$

- The renewal times

$$\tau(0) := \mathcal{T}_0 + 1, \quad \tau(j) := \mathcal{T}_j - \mathcal{T}_{j-1}.$$

Similar to the regenerative process, the sequence of renewal times $\{\tau(j)\}_{j=1}^\infty$ is i.i.d. and independent of the choice of the initial probability. In this work, we set $\tau = \tau(1)$. Note that the hitting time \mathcal{T}_j is finite for any starting probability and for all $j \in \mathbb{N}$. Let $B_j : E \rightarrow \mathbb{T}$ a sequence of the following random variables: $B_0 = \{X_0, \dots, X_{\mathcal{T}_0}\}$ and $B_j = \{X_{\mathcal{T}_{j-1}+1}, \dots, X_{\mathcal{T}_j}\}$ in $\mathbb{T} = \bigcup_{n=1}^\infty E^n$, for all $j \in \{1, \dots, \ell_n\}$ and $B_{\ell_n}^{(n)} = \{X_{\mathcal{T}_{\ell_n+1}}, \dots, X_n\}$ defined by

$$\ell_n = \sum_{1 \leq i \leq n} \mathbb{1}\{X_i \in \mathbf{A}\}.$$

For more details on the preceding objects, we refer to Radulović, 2004. In the modern theory of the empirical processes, it is customary to identify \mathbb{P} and the empirical measure \mathbb{P}_n with the mappings given by

$$f \rightarrow \mathbb{P}f = \int_E f d\mathbb{P}, \quad \text{and} \quad f \rightarrow \mathbb{P}_n f = \int_E f d\mathbb{P}_n = \frac{1}{n} \sum_{k=1}^n f(X_k).$$

Given a measure \mathbb{Q} on (E, \mathcal{E}) , for $1 \leq p < \infty$, we define

$$\|f\|_{L_p(\mathbb{Q})} = \left[\int_E |f|^p d\mathbb{Q} \right]^{1/p}.$$

Throughout the article \mathcal{H}_m will denote the class of kernels h_m , for each $m = m_n$, where the kernel is a function $h_m : E^m \rightarrow \mathbb{R}$. Consider the parameter $\Theta = \Theta(h)$ for which there is an unbiased estimator and a sequence of parameters $\Theta_m(h_m)$ such that $\Theta_m(h_m)$ converges to $\Theta(h)$, these two parameters are defined as follows, for \mathbb{P} a probability measure:

$$\Theta(h) = \mathbb{P}h(X_1, X_2, \dots),$$

and, for $h_m : E^m \rightarrow \mathbb{R}$,

$$\Theta_m(h_m) = \mathbb{P}^m h_m(X_1, \dots, X_m) = \int_E \cdots \int_E h_m(x_1, \dots, x_m) \boldsymbol{\mu}(dx_1) \cdots \boldsymbol{\mu}(dx_m).$$

The kernel h_m may be seen as a “grand” kernel and it can be written as a sum of sub-kernel $\{f_\iota\}_{\iota=1}^m$:

$$h_m(x_1, \dots, x_m) = f_1(x_1) + f_2(x_1, x_2) + \cdots + f_m(x_1, \dots, x_m) = \sum_{\iota=1}^m f_\iota(x_1, \dots, x_\iota).$$

Construct now the symmetrized kernel. The permutation-symmetrized grand kernel is:

$$\tilde{h}_m(x_1, \dots, x_m) = \frac{1}{m!} \sum_{\mathbf{i} \in I(m,m)} h_m(x_{i_1}, \dots, x_{i_m}),$$

where

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

with cardinal $\binom{n}{m} = \frac{n!}{m!(n-m)!}$. Finally, we define the IDUS as in (1.1) in the following way

$$U_n(h_m) = \frac{(n-m)!}{n!} \sum_{\mathbf{i} \in I(n,m)} h_m(X_{i_1}, \dots, X_{i_m}),$$

and for the symmetric kernel by :

$$U_n(\tilde{h}_m) = \binom{n}{m}^{-1} \sum_{\mathbf{i} \in I(n,m)} \tilde{h}_m(X_{i_1}, \dots, X_{i_m}).$$

We highlight that $U_n(h_m) = U_n(\tilde{h}_m)$. Let us recall the following usual definitions. The function H is an envelope for the class \mathcal{H} if $|h(x)| \leq H(x)$ for all $x \in E^m$ and $h \in \mathcal{H}$. For a metric space (\mathcal{H}, d) , the covering number $\mathcal{N}(\varepsilon, \mathcal{H}, d)$ is the minimal number of balls of size ε needed to cover \mathcal{H} . The metric that we use here is the $L_2(\mathbb{Q})$ -norm denoted by $\|\cdot\|_{L_2(\mathbb{Q})}$.

Definition 2.1. A class \mathcal{H} of measurable functions $E^m \rightarrow \mathbb{R}$ is called Euclidean (C, V) for an envelope $H(\cdot)$ and admissible characteristic (C, V) (positive constants) such that $C \geq (3\sqrt{e})^V$ and $V \geq 1$, if for all probability measure \mathbb{Q} on (E^m, \mathcal{E}) with $0 < \|H\|_{L_2(\mathbb{Q})} < \infty$ and every $0 < \varepsilon < 1$,

$$\mathcal{N}(\varepsilon \|H\|_{L_2(\mathbb{Q})}, \mathcal{H}, \|\cdot\|_{L_2(\mathbb{Q})}) \leq C \varepsilon^{-V}.$$

3 Infinite degree U -statistics and Hoeffding decomposition

As previously stated, we are interested in the estimation of

$$\Theta(h_m) = \mathbb{E}h_m(X_1, \dots, X_m) = \int_E \dots \int_E h_m(x_1, \dots, x_m) \boldsymbol{\mu}(dx_1) \dots \boldsymbol{\mu}(dx_m),$$

where $h_m : E^m \rightarrow \mathbb{R}$ is a kernel function. The estimation of this parameter should be possible using the U -statistics of infinite degree. Hence we make use of the symmetric form of this estimator in the following such that for $m = m_n$ and \tilde{h}_m a symmetric kernel

$$U_n(\tilde{h}_m) = \binom{n}{m}^{-1} \sum_{\mathbf{i} \in I(n,m)} \tilde{h}_m(X_{i_1}, \dots, X_{i_m}), \quad (3.1)$$

where

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

Hoeffding decomposition is a valuable method for achieving asymptotic limits of U -statistics, as in C. Heilig et al., 2001. Here, we will present this decomposition for the infinite degree U -statistics. The kernel is supposed to be symmetric because it is required for the Hoeffding decomposition. Remind that, if $\mathbb{P}^m |h_m| < \infty$, then $U_n(\tilde{h}_m)$ is an unbiased estimator. Using this projection of the kernel, we obtain the Hoeffding

decomposition of the symmetric statistic, for $c \leq m$, as follows

$$U_n(\tilde{h}_m) = \sum_{c=0}^m \binom{m}{c} U_{n,c}(\tilde{h}_{m(c)}), \quad (3.2)$$

where

$$U_{n,c}(\tilde{h}_{m(c)}) = \binom{n}{c}^{-1} \sum_{\mathbf{i} \in I(n,c)} \tilde{h}_{m(c)}(X_{i_1}, \dots, X_{i_c}),$$

such that

$$\begin{aligned} \tilde{h}_{m(c)}(x_1, \dots, x_c) &= \tilde{h}_{m|c}(x_1, \dots, x_c) - \Theta(h_m) - \sum_{k=1}^{c-1} \sum_{\boldsymbol{\pi} \in I(c,k)} \tilde{h}_{m(k)}(x_{\pi_1}, \dots, x_{\pi_k}) \\ &= \sum_{k=0}^c (-1)^{c-k} \sum_{\boldsymbol{\pi} \in I(c,k)} [\tilde{h}_{m|c}(x_{\pi_1}, \dots, x_{\pi_k}) - \Theta(h_m)] \\ &= \sum_{k=0}^c (-1)^{c-k} \sum_{\boldsymbol{\pi} \in I(c,k)} \tilde{h}_{m|c}(x_{\pi_1}, \dots, x_{\pi_k}), \end{aligned} \quad (3.3)$$

where $I(c, k) := \{\boldsymbol{\pi} = (\pi_1, \dots, \pi_k) : 1 \leq \pi_j \leq c \text{ and } \pi_j \neq \pi_r \text{ if } j \neq r\}$ and

$$\tilde{h}_{m|c}(x_1, \dots, x_c) = \int \tilde{h}_m(x_1, \dots, x_m) d\mathbb{P}(x_{c+1}) \dots d\mathbb{P}(x_m).$$

We have used the fact

$$\sum_{k=0}^c (-1)^{c-k} \sum_{\boldsymbol{\pi} \in I(c,k)} \Theta_m(h_m) = \Theta_m(h_m) \sum_{k=0}^c (-1)^{c-k} \binom{c}{k} = \Theta_m(h_m)(1-1)^c = 0.$$

The first order projection, as is widely known, is the leader term for the central limit theorem in the finite degree U -statistics. However, due to its dependency on m in the infinite degree, this term has far more sensitive characteristics. To be more precise, notice that

$$mU_{n,1}(\tilde{h}_{m(1)}) := \frac{1}{n} \sum_{i=1}^n m\tilde{h}_{m(1)}(X_i),$$

where $m\tilde{h}_{m(1)}$ can be viewed as:

$$m\tilde{h}_{m(1)}(x) = m \left[\tilde{H}_{m|1}(x) - \mathbb{P}^m h_m \right] = \sum_{j=1}^m j \left[\tilde{h}_{j|1}(x) - \mathbb{P}^j h_j \right].$$

where by applying (3.3),

$$\begin{aligned} m\widetilde{H}_{m|1}(x) &= m \sum_{j=1}^m \binom{m}{j}^{-1} \binom{m-1}{j} \widetilde{h}_{j|0} + m \sum_{j=1}^m \binom{m}{j}^{-1} \binom{m-1}{j-1} \widetilde{h}_{j|1}(x) \\ &= \sum_{j=1}^m (m-j) \mathbb{P}^j h_j + j \widetilde{h}_{j|1}(x). \end{aligned}$$

Subtract $m\mathbb{P}^m h_m = \sum_{j=1}^m m\mathbb{P}^j h_j$ from the previous expression to find $m\widetilde{h}_{m(1)}(x)$. Taking

$$g_m(x) = \sum_{j=1}^m j \widetilde{h}_{j|1}(x),$$

we can see the first projection as follows:

$$mU_{n,1}(\widetilde{h}_{m(1)}) = (\mathbb{P}_n - \mathbb{P}) g_m. \quad (3.4)$$

The definition of the new function $g_m(x)$ gives a convenient way to analyze the behavior of $m\widetilde{h}_{m|1}(x)$.

4 Asymptotic theory: first-order projection

In the case of increasing degree U -statistic, the first-order projection is used in the same way as with finite-degree U -statistics, except that it has considerably more delicate characteristics, especially in our Markov chain framework. That is why we start by studying the asymptotic behavior of this empirical process as an intermediate step to investigate the U -process. We adapt in a non-trivial way the techniques from both papers C. Heilig et al., 2001; Levental, 1988 . However, as will be seen later, the problem requires much more than 'simply' combining ideas from these two papers. In fact, delicate mathematical derivations will be required to cope with the IDUS in the Markovian framework. From now on, \mathcal{H}_m denotes the class of symmetric functions h_m .

4.1 Strong law of large numbers

Theorem 4.1

Let $(X_n)_{n \geq 0}$ be a Harris-recurrent Markov chain, and $\tau(j)$ the renewal time such that $\tau(0) < \infty$ a.s and $\mathbb{E}(\tau) < \infty$. For all m , let $\mathcal{G}_m = \{g_m = \sum_{j=1}^m j h_{j|1} : h_m \in \mathcal{H}_m\}$ a permissible class of functions such that

$$\mathcal{N}(\varepsilon \|\bar{G}_m\|_{L_1(\mathbb{P})}, \mathcal{G}_m, \|\cdot\|_{L_1(\mathbb{P})}) \leq A\varepsilon^{-V_m},$$

with an envelope \bar{G}_m and $\limsup \mathbb{P}^m \bar{G}_m^2 < \infty$. Then as $n \rightarrow \infty$, we have

$$\sup_{g_m \in \mathcal{G}_m} |\mathbb{P}_n g_m - \mathbb{P} g_m| \rightarrow 0, \text{ a.s.}$$

Remark 4.1. *C. Heilig et al. proved the weak law of large numbers under the conditions that $K_m m V_m^{1/2} = o(n^{1/2})$ and $m V_m^{1/2} = o(n^{1/2})$. This condition is crucial to investigate all terms in the Hoeffding decomposition. In Theorem 4.1, we treated only the first term of the Hoeffding decomposition by providing the asymptotic convergence in a general way where we make use of Pollard, 1984, Theorem 37. Another subtle difference is that C. Heilig et al. use that the class of functions is bounded by some constant depending on m , while in this chapter, we use the truncation arguments to obtain the law of large numbers.*

4.2 Weak convergence theorem

Recall that controlling the rate with increasing $m = m_n$ is a difficult aspect that distinguishes the infinite degree U -processes (IDUP). C. Heilig et al., 2001 investigated the central limit theorem of the first order projection, which is treated as an empirical process. We will modify the proof for the case of Markovian data by keeping similar conditions as in the i.i.d. setting with particular attention to the factor of m in front of $U_{n,1}(h_{m(1)})$ in equation (3.4).

Theorem 4.2

Let $\{X_i\}_{i=1}^n$ be a sequence of Harris-recurrent Markov chain random variables and assume that the following conditions hold.

(A.1) The class \mathcal{G}_m is a uniformly bounded permissible such that for the envelope \bar{G}_m , for all m , $V > 1$ and $\varepsilon \in (0, 1]$:

$$\mathcal{N} \left(\varepsilon \|\bar{G}_m\|_{L_2(\mathbb{P})}, \mathcal{G}_m, \|\cdot\|_{L_2(\mathbb{P})} \right) \leq A\varepsilon^{-V}.$$

(A.2) $\sigma_1(g_m, \check{g}_m) = \lim_{n \rightarrow \infty} \mathbb{P} g_m(X) \check{g}_m(X)$ exists for every g_m, \check{g}_m in \mathcal{G}_m , where, for h_m and $\check{h}_m \in \mathcal{H}_m$, $\check{g}_m(x) = \sum_{j=1}^m j \check{h}_{j|1}(x)$ and $g_m(x) = \sum_{j=1}^m j h_{j|1}(x)$. For

$$\rho_n^2(g_m, \check{g}_m) = \sum_{j=1}^m \sum_{k=1}^m jk \mathbb{P} \left[h_{j|1}(X_1) - \check{h}_{j|1}(X_1) \right] \left[h_{k|1}(X_1) - \check{h}_{k|1}(X_1) \right],$$

the limit $\rho(\cdot, \cdot) = \lim_{n \rightarrow \infty} \rho_n(\cdot, \cdot)$ is well-defined, and for all sequences of functions $\{\bar{g}_n\}$ and $\{\check{g}_n\}$ in \mathcal{G}_m , if $\rho(\bar{g}_n, \check{g}_n) \rightarrow 0$, then $\rho_n(\bar{g}_n, \check{g}_n) \rightarrow 0$.

(A.3) $\sum_{i=1}^n \mathbb{P} \bar{G}_m(X_i)^2 \left\{ \bar{G}_m(X_i)^2 > \varepsilon \right\} \rightarrow 0$, for every $\varepsilon > 0$.

(A.4) We suppose that $\mathbb{P}_A \left[\left(\sum_{j=1}^{\mathcal{J}_0} g_m(X_j) \right)^2 \right] < \infty$.

(A.5) $\mathbb{E}(\tau)^{2+\alpha} < \infty$ ($\alpha > 0$ fixed).

Then $\sqrt{n}(\mathbb{P}_n - \mathbb{P})g_m$ converges in distribution to a mean-zero Gaussian process which is uniformly ρ -continuous and has co-variance given by

$$\sigma(g, \check{g}) = \lim_{n \rightarrow \infty} \sum_{j=1}^m \sum_{k=1}^m jk \mathbb{P} \left[h_{j|1}(X_1) - \check{h}_{j|1}(X_1) \right] \left[h_{k|1}(X_1) - \check{h}_{k|1}(X_1) \right].$$

Remark 4.2. *At first glance, one can not see the connection between $g(\cdot)$ and $\check{g}(\cdot)$ in the variance given in the last theorem. One can consider $g(\cdot) = h(\cdot, t)$ and in a similar way $\check{g}(\cdot) = h(\cdot, s)$, for $s, t \in T$ (T is some index set) that is in agreement with the notation given in C. Heilig et al., 2001, then we have*

$$\sigma(s, t) = \lim_{n \rightarrow \infty} \sum_{j=1}^m \sum_{k=1}^m jk \mathbb{P} \left[h_{j|1}(X_1; s) - \mathbb{P} h_{j|1}(X_1; t) \right] \left[h_{k|1}(X_1; s) - \mathbb{P} h_{k|1}(X_1; t) \right],$$

and

$$\rho_n^2(s, t) = \sum_{j=1}^m \sum_{k=1}^m jk \mathbb{P} \left[h_{j|1}(X_1; s) - h_{j|1}(X_1; t) \right] \left[h_{k|1}(X_1; s) - h_{k|1}(X_1; t) \right].$$

5 Increasing degree U-processes: asymptotic theory

We now consider the IDUPs. We will extend and complement the work of C. Heilig et al., 2001 by considering the theory of the IDUP for the Markovian setting. In the next theorem, we seek to establish the law of large numbers, which follows from the almost sure convergence in the fixed degree settings. The main problem is that the increasing degree or infinite degree U -statistics is not a reverse martingale given any filtration, as well as the kernel changes with the sample size n .

Let $X_1, \dots, X_n, X'_1, \dots, X'_n$ be two independent samples from the distribution \mathbb{P} . Also, let $W_i = X_i$ and $W_{n+i} = X'_i, i = 1, \dots, n$. For $\zeta \geq 1$, suppose that \mathcal{H}_ζ is a class of symmetric kernels $\{h_\zeta\}$ of degree ζ , with envelop H_ζ and for h_ζ and $\check{h}_\zeta \in \mathcal{H}_\zeta$, define the pseudometric

$$d_{2n,\zeta}(h_\zeta, \check{h}_\zeta) = \left[\frac{U_{2n,\zeta}(h_\zeta - \check{h}_\zeta)^2}{U_{2n,\zeta}(H_\zeta^2)} \right]^{1/2}$$

where $U_{2n,\zeta}(h_\zeta)$ is a U -statistic based on the double sample W_1, \dots, W_{2n} and h_ζ is a symmetric kernel of degree ζ .

Theorem 5.1: Law of Large Numbers

Let $(X_n)_{n \geq 0}$ be a Harris-recurrent Markov chain with atom \mathbf{A} and $\mathbb{E}(\tau) < \infty$. For $m = 1, \dots, \infty$, let $h_m : E^m \rightarrow \mathbb{R}$ be the associated kernel belonging to the class of functions \mathcal{H}_m . Assume that the kernels have an envelope \bar{H}_m for each m and

$$\lim_{m \rightarrow \infty} \mathbb{P}^m \bar{H}_m < \infty.$$

Let r_m be a sequence of integers, and

$$\phi(m) = n^{-m/2} \Gamma_m \mathbb{P} \left[\left[U_{2n,m} \bar{H}_m^2 \right]^{1/2} \int_0^{\delta_{n,m}} N(x, d_{2n,m}, \mathcal{H}_m)^{1/2r_m} dx \right], \quad (5.1)$$

where $\Gamma_m = [64m!(16r_m - 8)^m]^{1/2}$. If

$$\phi(m) = o(1), \quad (5.2)$$

then, as $n \rightarrow \infty$, we have

$$\sup_{h_m \in \mathcal{H}_m} |U_n(h_m) - \mathbb{P}^m(h_m)| \xrightarrow{\mathbb{P}} 0.$$

Remark 5.1. *It would be possible to use the Hoeffding decomposition of the infinite degree U -process to prove the law of large numbers given in Theorem 5.1 and apply the maximal inequality to terms of this decomposition. For $c = 1, \dots, m$, let r_c be a sequence of positive integers and $\mathcal{H}_{m(c)}$ be the classes of functions associated with the kernel projections $h_{m(c)}$, each with the envelope function $H_{m(c)}$. It will be required to assume that the classes $\mathcal{H}_{m(c)}$ are totally bounded under pseudometric $d_{2n,c}$ for each c, m , and $m < n/3$ and*

$$\sum_{c=1}^m \phi(c) = o(1),$$

where

$$\phi(c) = n^{-c/2} \binom{m}{c} \Gamma_c \mathbb{P} \left[\left[U_{2n,c} H_{m(c)}^2 \right]^{1/2} \int_0^{\delta_{n,c}} N(x, d_{2n,c}, \mathcal{H}(h_{m(c)}))^{1/2r_c} dx \right], \quad (5.3)$$

and $\Gamma_c = [64c!(16r_c - 8)^c]^{1/2}$.

Remark 5.2. *Suppose the kernel classes are Euclidean (A_m, V_m) Condition (5.1) becomes :*

$$\phi(m) = n^{-m/2} 8m!^{1/2} (8V_m)^{m/2} \left(\frac{A_m}{4} \right)^{\frac{1}{V_m+1}} (V_m + 1) (\mathbb{P}^m \bar{H}_m^2)^{1/2}, \quad (5.4)$$

for $r_m \geq (V_m + 1)/2$, Sherman Sherman, 1994 studied this case and gives us the results. Note that by evaluating the integral we have $\int_0^{1/4} (Ax^{-V})^{1/2r} dx = A^{1/2r} 4^{V/2r-1} \left(1 - \frac{V}{2r}\right)^{-1}$ and by applying (concave) Jensen's inequality we have

$$\mathbb{P} \left(U_{2n,m} \bar{H}_m^2 \right)^{1/2} \leq \left(\mathbb{P} U_{2n,m} \bar{H}_m^2 \right)^{1/2} = \left(\mathbb{P}^m \bar{H}_m^2 \right)^{1/2}.$$

Also, we write $\phi(c)$ defined in Equation (5.3) with a general form for the case of class of functions associated with the kernel projections, if the classes $\mathcal{H}_{m(c)}$ are euclidean with parameters (A_c, V_c) for the envelopes $H_{m(c)}$ and if $r_c \geq (V_c + 1)/2$, then $\phi(c)$ can be viewed as:

$$\phi(c) = n^{-c/2} \binom{m}{c} 8c!^{1/2} (8V_c)^{c/2} \left(\frac{A_c}{4} \right)^{\frac{1}{V_c+1}} (V_c + 1) (\mathbb{P}^c H_{m(c)}^2)^{1/2}, \quad (5.5)$$

if in addition $|H_{m(c)}| \leq M_m$ for each c , and for M_m a finite constant, then $(\mathbb{P}^c \widetilde{H}_{m(c)}^2)^{1/2} \leq 2^i M_m$ and the results are still valid.

The following theorem aims to prove the weak convergence of the infinite degree U -processes for the Harris Markov chains. This convergence is investigated over a class of functions \mathcal{H}_m , where certain conditions will be preserved against any complications caused by the non-fixed kernels.

Theorem 5.2: Weak convergence theorem for IDUS

Let $(X_n)_{n \geq 0}$ be a Harris-recurrent Markov chain with atom \mathbf{A} , and $\tau(j)$ the renewal time such that $\mathbb{E}(\tau)^{2+\alpha} < \infty$. Suppose also that all conditions of Theorem 4.2 hold. Then, if the class \mathcal{H}_m has \mathbb{P}^m -square integrable envelope \bar{H}_m , i.e., $\mathbb{P}^m(\bar{H}_m^2) < \infty$, we have $\sqrt{n}[U_n(h_m) - \mathbb{P}^m(h_m)]$ converges in distribution to a mean-zero Gaussian process \mathbb{G} which is uniformly ρ -continuous.

Remark 5.3. Let us recall that the strong or α -mixing coefficient between two σ -fields, \mathcal{A} and \mathcal{B} in (Ω, \mathcal{C}, P) , was defined by Rosenblatt, 1956 as

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup_{(A, B) \in \mathcal{A} \times \mathcal{B}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

The alpha mixing coefficient associated to a sequence of random variables is defined as

$$\alpha_k = \sup \{ \alpha(\mathcal{F}_j, \mathcal{G}_{j+k}) : j \in \mathbb{N} \},$$

where $\mathcal{F}_j = \sigma(X_0, X_1, \dots, X_j)$ and $\mathcal{G}_n = \sigma(X_n, X_{n+1}, \dots)$. Theorem 2.2 in Bolthausen, 1982 provides a connection between α -mixing coefficients and renewal times for countable Markov chains, given by

$$\sum_{i=1}^{\infty} i^\lambda \alpha_i < \infty \Leftrightarrow \mathbb{E}(\tau^{2+\lambda}) < \infty, \text{ for every } \lambda \geq 0.$$

6 Bootstrap of the increasing degree U-processes

We describe here the resampling algorithm, and then we will establish the bootstrap version of the uniform central limit theorem of IDUS for the Harris Markov chains. As we have mentioned in the introduction, as the data is not i.i.d., we will use the block bootstrap to resample a new sequence. There is a connection with moving blocks bootstrap and its modification, matched block bootstrap. Instead of artificially breaking a sample into blocks of a fixed size and then resampling from them, the latter tries to match the blocks to make a smoother transition. Recall that the regeneration-based bootstrap procedure introduced by Datta et al., 1993 to deal with atomic Markov chains and further explored by Bertail et al., 2006c; Radulović, 2004. Keep in mind that ℓ_n is the number of visits to the atom \mathbf{A} .

As we explained in Section 2, we divide the observed sample $X^{(n)} = (X_1, \dots, X_n^{(n)})$ into $(\ell_n + 1)$ regenerative blocks $B_0, \dots, B_{\ell_n-1}, B_{\ell_n}^{(n)} \in \mathbb{T}$, where $B_0 = \{X_0, \dots, X_{\mathcal{T}_0}\}$ and $B_j = \{X_{\mathcal{T}_{j-1}+1}, \dots, X_{\mathcal{T}_j}\}$ in $\mathbb{T} = \bigcup_{n=1}^{\infty} E^n$, for all $j \in \{1, \dots, \ell_n\}$ where ℓ_n total number of blocks. We drop the first and the last blocks, non-regenerative blocks, to avoid the bias

problem, see Remark 6.3. We draw sequentially bootstrap data blocks $B_{1,n}^*, \dots, B_{k,n}^*$ independently. We generate the bootstrap blocks until the joint length of the bootstrap blocks

$$l^*(k) = \sum_{j=1}^k l(B_{j,n}^*)$$

of the bootstrap data series is larger than n . Let

$$\ell_n^* = \inf\{k > 1, l^*(k) > n\}.$$

In this case we can see, for $j \in \{1, \dots, \ell_n^* - 1\}$, that

$$\mathbb{P}^* [B_j^* = B_k] = \frac{1}{\ell_n^* - 1}, \quad (6.1)$$

where \mathbb{P}^* is the bootstrap distribution given $(B_n)_n$, and we will denote by \mathbb{E}^* and Var^* respectively the expectation and the variance of the bootstrapped sample conditionally on the original sample. We generate a trajectory $X^{*(n)} = (X_1^*, \dots, X_{n^*}^*)$ of length $n^* = l^*(N_n^*)$ by connecting the bootstrap data blocks created by the previous mechanism, then plugging this trajectory sample to build the bootstrapped U -statistic:

$$U_n^*(\tilde{h}_m) = \binom{n^*}{m} \sum_{i \in I(n^*, m)} \tilde{h}_m(X_{i_1}^*, \dots, X_{i_m}^*) = \mathbb{E}^*(U_n^*) + \frac{1}{n^*} \sum_{i=1}^{n^*} m \tilde{h}_{m(1)}(X_i^*) + \sum_{c=2}^m \binom{m}{c} U_{n,c}^*(\tilde{h}_{m(c)}) \quad (6.2)$$

$$:= \mathbb{E}^*(U_n^*) + m U_{n,1}^*(\tilde{h}_{m(1)}) + U_{n,R}^*(\tilde{h}_m), \quad (6.3)$$

where $h_{m(c)}$ is the projected kernels defined in (3.3).

6.1 Asymptotic validity of bootstrapped Markovian U -statistics

In the following theorem we give the main result concerning the bootstrap of the infinite degree U -processes for the Harris Markov chains.

Theorem 6.1

Let $\{X_i\}_{i=1}^n$ be a sequence of r. v. of Harris Markov chain, and \mathcal{H}_m an euclidean permissible class of function. Under the assumptions of Theorem 4.2 and Theorem 5.2, we have

$$\sup_{\Upsilon \in \text{BL}_1} \left| \mathbb{E}^* \left(\Upsilon \left(\sqrt{n^*} \left(U_n^*(\tilde{h}_m) - U_n(\tilde{h}_m) \right) \right) \right) - \mathbb{E}(\Upsilon(\mathbb{G})) \right| \rightarrow 0 \text{ in probability,} \quad (6.4)$$

where BL_1 is the set of all functions $\Upsilon : \ell^\infty(\mathcal{H}_m) \mapsto [0, 1]$ such that $|\Upsilon(z_1) - \Upsilon(z_2)| \leq \|z_1 - z_2\|_{\mathcal{H}_m}$ for every z_1, z_2 .

Remark 6.1. *The main difficulties in proving Theorem 6.1 related to the random size of the resampled blocks. This generates problems with the random stopping times, and this can not be removed by replacing a random stopping time with its expectation. In the present setting, the random bootstrap variables are formed by resampling from a random number of blocks. One can think that the use of the conditioning arguments can overcome the problem, but the answer is negative. To circumvent this, we will follow the reasoning of Radulović, 2004.*

Remark 6.2. *In the present chapter, we have considered a renewal type of bootstrap for atomic Markov chains under minimal moment conditions on renewal times. The atomic Markov chains assumption can be dropped by mimicking the ideas of Athreya et al., 1978; Nummelin, 1978 by introducing an artificial atom and deriving the bootstrap procedure that applies to non-atomic Markov chains. Precisely, in the case of a general irreducible chain X with a transition kernel $\Pi(x, dy)$ satisfying a minorization condition: $\forall x \in S, \Pi(x, dy) \geq \delta \psi(dy)$, for an accessible measurable set S , a probability measure ψ and $\delta \in]0, 1[$ (note that such a minorization condition always holds for Π or an iterate when the chain is irreducible), an atomic extension (X, Y) of the chain may be explicitly constructed by the Nummelin splitting technique (see Nummelin, 1984) from the parameters (S, δ, ψ) and the transition probability Π , see for instance Bertail et al., 2006c; Radulović, 2004. From a practical viewpoint, the size of the first block may be large compared to the size n of the whole trajectory, for instance, in the case where the expected return time to the (pseudo-)atom when starting with the initial probability distribution is large. The effective sample size for constructing the data blocks and the corresponding statistic is then dramatically reduced. However, in Bertail et al., 2008 some simulations are given together with examples including the content-dependent storage systems and the general AR models supporting the method discussed in this work.*

Remark 6.3. *The first data block B_0 may produce a bias of order $O(n^{-1})$ when the initial distribution ν is different from the limiting distribution μ (non-stationary Markov chain), for instance, refer to Bertail et al., 2004, Proposition 3.1. That is why it is advised not to employ estimators based on the entire trajectory. This fact is well known in the Bayesian literature when the matter is to control the convergence of Markov chain Monte Carlo algorithms. It is closely related to the important problem of burn-in (time), that is, the time that one should wait before the marginal of a (simulated) chain is close enough to the limit distribution (see, for instance, Hobert et al., 2004). When the statistic is produced using exclusively regeneration blocks, the first block does not affect the regenerative block bootstrap estimate's second-order properties. However, from a practical standpoint, the initial block size may be greater than the size n of the entire trajectory, such as when the estimated return time to the atom is long when starting with ν . The effective sample size for building the data blocks and accompanying statistics is consequently substantially decreased. We mention that eliminating the first block allows deleting the only quantity depending on the initial distribution ν in the first order term of the bias, which may be interesting for estimation purpose and is crucial when the matter is to deal with an estimator of which variance or sampling distribution may be approximated by a resampling procedure in a nonstationary setting. On the other hand, the last block could be empty if the last block ("full" block) ends with n . In such a case, for mimicking the distribution of the original statistic, it is preferable, heuristically speaking, to draw sequentially the bootstrap blocks B_1^*, \dots, B_k^* independently from the empirical distribution \mathbb{P}_n , until $l(B_0) + \sum_{j=1}^k l(B_j^*)$ is larger than n , taking into account in an effective way the size $l(B_0)$ this way. Although it does not play any role asymptotically, since $l(B_0)/n = O_{\mathbb{P}_\nu}(n^{-1})$ as $n \rightarrow \infty$, simulation studies show that this may improve the finite sample properties of the bootstrapped distribution, see for instance Bertail et al., 2008 for more details and recommendations about this problem.*

Remark 6.4. To better understand the role of the block number, let us describe the Moving Block Bootstrap (MBB). Suppose that $\{X_t\}_{t \in \mathbb{Z}}$ is a stationary weakly dependent time series and that $\{X_1, \dots, X_n\}$ are observed. Let ℓ be an integer satisfying $1 \leq \ell < n$. Define the blocks $\mathcal{B}_1 = (X_1, \dots, X_\ell)$, $\mathcal{B}_2 = (X_2, \dots, X_{\ell+1})$, \dots , $\mathcal{B}_N = (X_N, \dots, X_n)$, where $N = n - \ell + 1$. For simplicity, suppose that ℓ divides n . Let $b = n/\ell$. To generate the MBB samples, we select b blocks at random with replacement from the collection $\{\mathcal{B}_1, \dots, \mathcal{B}_N\}$. Since each resampled block has ℓ elements, concatenating the elements of the b resampled blocks serially yields $b \cdot \ell = n$ bootstrap observations X_1^*, \dots, X_n^* . Note that if we set $\ell = 1$, then the MBB reduces to the standard bootstrap method of Efron, 1979 for i.i.d. data. However, for a valid approximation in the dependent case, it is typically required that $\ell^{-1} + n^{-1}\ell = o(1)$ as $n \rightarrow \infty$. It is well known (see Chung, 1967, Chapter I.14) that for any starting probability ν and all $j \in \mathbb{N}_0$, $\mathcal{T}_j < \infty$, \mathbb{P} a.s. In our setting, ℓ is not fixed and is replaced by $\mathcal{T}_j - \mathcal{T}_{j-1}$. Therefore the number of blocks $k \rightarrow \infty$ as $n \rightarrow \infty$. At this point, we mention that the conditions on k are imposed implicitly by the fact that, by definition, k is related to l^* .

7 Examples and Applications

For a better understanding of the idea of infinite degree U -statistics, we will give some examples extracted from different references. Although only four examples will be given here, they stand as archetypes for various examples that can be investigated similarly.

Example 16 (Simultaneous prediction intervals for random forests (Y. Song et al., 2019)). Consider a training dataset of size n , $\{(Y_1, Z_1), \dots, (Y_n, Z_n)\} = \{X_1, \dots, X_n\} = X_1^n$, where $Y_i \in \mathcal{Y}$ is a vector of features and $Z_i \in \mathbb{R}$ is a response. Let h be a deterministic prediction rule that takes as input a sub-sample $\{X_{i_1}, \dots, X_{i_m}\}$ with $1 \leq m \leq n$ and outputs predictions on d testing points (y_1^*, \dots, y_d^*) in the feature space \mathcal{Y} . For random forests, the tree-based prediction rule is constructed on each sub-sample with additional randomness. Specifically, let $\{W_i : i \in I(m, n)\}$ be a collection of random variables taking value in a measurable space (S', \mathcal{S}') that are independent of the data X_1^n . Let $H : E^m \times S' \rightarrow \mathbb{R}^d$ be an $\mathcal{E}^m \otimes \mathcal{S}'$ -measurable function such that $\mathbb{E}[H(x_1, \dots, x_m, W)] = h(x_1, \dots, x_m)$. Then predictions of random forests are given by a d -dimensional U -statistic with random kernel H :

$$\hat{U}_n := \frac{(n-m)!}{n!} \sum_{i \in I(m, n)} H(X_{i_1}, \dots, X_{i_m}, W_i). \quad (7.1)$$

where the random kernel H varies with m . For forests based on subsamples, (Mentch et al., 2016) first show the asymptotic normality of its estimator under a U -statistic framework with growing tree (kernel) size $m = o(\sqrt{n})$, where m is the subsampling size and n is the training sample size. Unfortunately, the conditions in Mentch et al., 2016 for asymptotic

normality cannot hold simultaneously. Further conditions and proofs are given in DiCiccio et al., 2022; Hediger et al., 2022; Peng et al., 2022; Qiu et al., 2022; Zhou et al., 2021.

Example 17 (Renewal estimator). *In this example, we will treat the famous estimator of Frees, 1989; the renewal estimator. This estimator introduced by Frees, 1989 was completed by C. Heilig et al., 2001. Let (E, \mathcal{E}) be a probability space, and let $\{X_i\}_{i=1}^n$ be a sequence of random variables with probability measure \mathbb{P} (which is assumed to have finite second moment). Then, for some index set T , and for each $t \in T$*

$$\left\{ \omega \in E : \sum_{i=1}^k X_i(\omega) \in (-\infty, t] \right\},$$

denotes the event that the k th renewal occurs by time t . The k -fold convolution measure is given by

$$\mathbb{M}^{[k]}(t) = \mathbb{P} \left\{ \sum_{i=1}^k X_i \leq t \right\}.$$

The renewal measure is

$$N(t) = \sum_{k=1}^{\infty} \mathbb{M}^{[k]}(t).$$

To estimate this function, we can use the U -statistic estimator for the convolution measure and define the sub-kernel and the U -statistic respectively by:

$$h_k(x_1, \dots, x_k; t) = \mathbb{1} \{x_1 + \dots + x_k \leq t\},$$

$$\mathbb{M}_n^{[k]}(t) = U_{n,k} h_k(\dots; t) = \binom{n}{k}^{-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{1} \{X_{i_1} + \dots + X_{i_k} \leq t\}. \quad (7.2)$$

Frees' non-parametric renewal estimator is then the following grand kernel, averaged over the sample:

$$H_m(x_1, \dots, x_m; t) = \sum_{k=1}^m \mathbb{1} \{x_1 + \dots + x_k \leq t\},$$

and

$$\widehat{N}_m(t) = U_{n,m} H_m(\dots; t) = \frac{(n-m)!}{n!} \sum_{1 \leq i_1 \dots i_m \leq n} \sum_{k=1}^m \mathbb{1} \{X_{i_1} + \dots + X_{i_k} \leq t\} = \sum_{k=1}^m U_{n,k} h_k(\dots; t),$$

where $m = m_n$ grows with n and this estimator is seen as a U -statistic of degree m or a sum of U -statistics with kernels of degree $k \in \{1, \dots, m\}$, and its expectation converges to the renewal function N . The following theorems extend Lemma 4 and 5 of C. Heilig et al., 2001 to the Markov setting.

Theorem 7.1. *Let $\{X_i\}_{i=1}^n$ be a sequence of Harris-recurrent Markov chain random variables with positive mean, finite variance, with $\mathbb{E}(\tau) < \infty$. Let Γ be a finite constant. The associated renewal estimator with grand kernel defined in (7.3) satisfies the uniform weak law of large numbers over $[0, \Gamma]$, for any $m < n$ and $m \rightarrow \infty$,*

$$\sup_{t \in [0, \Gamma]} \left| \widehat{N}_m(t) - \mathbb{P}_m H_m \right| \xrightarrow{\mathbb{P}} 0.$$

Proof of Theorem 7.1.

This theorem is an immediate consequence of Theorem 5.1 and, as in C. Heilig et al., 2001, the use of the fact that if the class \mathcal{H}_m is euclidean $(C, 2)$ for the constant envelope m and that $\mathcal{H}_{m(i)}$ is Euclidean $(C, 4)$ for the envelope $2^i m$, at this point refer to C. Heilig et al., 2001 and Dudley, 1985, Example 5.4. We can see also the convergence of $|\mathbb{P}^m h_m - N(t)|$ to 0, as well as $m \rightarrow \infty$, we have

$$\sup_{t \in [0, \Gamma]} |\mathbb{P}^m H_m - N(t)| = \sum_{k > m} \mathbb{M}^{[k]}(\Gamma) \rightarrow 0.$$

The rest of the proof follows similar lines of the proof of Lemma 4 in C. Heilig et al., 2001.

Remark 7.2. *The preceding theorem is an empirical law of large numbers. No additional assumptions relative to the i.i.d. case are needed to get the result except the assumption that guarantees the existence of a steady state distribution for the process, namely $\mathbb{E}(\tau) < \infty$.*

Theorem 7.3. *Let $\{X_i\}_{i=1}^n$ be a sequence of Harris-recurrent Markov chain random variables with positive mean and finite variance, and that for some $\eta > 0$, $\mathbb{P}|X_1^-|^{5+\eta} > 0$, where $X_1^- = \min(X, 0)$, $\mathbb{E}(\tau)^{2+\alpha} < \infty$ ($\alpha > 0$ fixed). Let Γ be a finite constant. Then for any $m \geq n^{1/(6+2\eta)}$, we have the functional central limit theorem, over $[0, \Gamma]$, for the IDUS $n^{1/2}(\widehat{N}_m(t) - \mathbb{P}_m H_m)$ associated with the renewal estimator.*

Proof of Theorem 7.3.

We can apply Theorem 5.2 or we can prove that

$$\mathbb{P} \left(\sup_{t \in [0, \Gamma]} \left| \widehat{N}_m(t) - \mathbb{P}_m H_m - (\mathbb{P}_n - \mathbb{P}) g_m \right| > \varepsilon \right) \rightarrow 0,$$

under the condition $\sum_{j=2}^m \phi(j) = o(n^{-1/2})$, where ϕ is defined in (5.3) and then we consider the functional central limit theorem for the first order projection. We have

$$g_m(x; t) = \sum_{j=1}^m j \tilde{h}_{j|1}(x; t) = \sum_{j=1}^m j \mathbb{M}^{[j-1]}(t - x).$$

Now, we consider the first-order projection $n^{1/2} (\mathbb{P}_n - \mathbb{P}) g_m$:

$$\begin{aligned} \sigma_{m|1}(s, t) &= \mathbb{P} [g_m(X_1; s) - \mathbb{P} g_m(s)] [g_m(X_1; t) - \mathbb{P} g_m(t)] \\ &= \sum_{j=1}^m \sum_{k=1}^m jk \left[\mathbb{P} \mathbb{M}^{[j-1]}(s - X_1) \mathbb{M}^{[k-1]}(t - X_1) - \mathbb{M}^{[j]}(s) \mathbb{M}^{[k]}(t) \right]. \end{aligned}$$

Observe that

$$\mathbb{P} \mathbb{M}^{[j-1]}(s - X_1) \mathbb{M}^{[k-1]}(t - X_1) - \mathbb{M}^{[j]}(s) \mathbb{M}^{[k]}(t) \leq \mathbb{P} \mathbb{M}^{[j-1]}(s - X_1) = \mathbb{M}^{[j]}(s).$$

The roles of j and k are symmetric, so

$$\mathbb{P} \mathbb{M}^{[j-1]}(s - X_1) \mathbb{M}^{[k-1]}(t - X_1) - \mathbb{M}^{[j]}(s) \mathbb{M}^{[k]}(t) \leq \mathbb{M}^{[j]}(s) \wedge \mathbb{M}^{[k]}(t). \quad (7.3)$$

For $r \geq 2$, condition $\mathbb{P}|X^-|^r > 0$ implies

$$\sum_{k \geq 1} k^{r-2} \mathbb{M}^{[k]}(t) < \infty,$$

for instance, see Gut, 1974b, which in combination with (7.3) gives that the limit of $\sigma_{m|1}(s, t)$ is finite, say $\sigma_1(s, t)$. We also have g_m^2 is uniformly integrable over the interval $[0, \mathbf{\Gamma}]$. Taking $\bar{G}_m(x) = g_m(x, \mathbf{\Gamma})$, and the fact that $g_m(x, t)$ is nondecreasing in t , all conditions of Theorem 4.2 are fulfilled. Hence the proof is complete.

Remark 7.4. *In the proof of the preceding theorem, we have followed the lines of Lemma 5 in C. Heilig et al., 2001. The main difference is that the last reference used the paper Theorem 2.1 of Gut, 1974a as a key step in the proof that is tailored for the independent setting. The key step in this paper uses Theorem 2.1 of Gut, 1974b, that is adapted to the dependent framework. In addition, we suppose that $\mathbb{E}(\tau)^{2+\alpha} < \infty$, which is a fundamental condition to prove the uniform CLT in the Markov framework. The key step in the proof of the uniform CLT in Levental, 1988 is Lemma 4.2, which shows that the main condition concerning the entropy for the uniform CLT in the i.i.d. case still holds for the block structure that we have here even though the length of each block is random, provided that $\mathbb{E}(\tau)^{2+\alpha} < \infty$.*

Example 18 (C. Heilig et al., 2001). *In the right censorship model, the data set is given by $\{(Y_i, \delta_i) : 1 \leq i \leq n\}$, where $Y_i = \min(T_i, C_i)$ and $\delta_i = \mathbb{1}\{T_i \leq C_i\}$ for $i \geq 1$. Here, $\{T_i : i \geq 1\}$ is a sequence of Markov nonnegative lifetimes, and $\{C_i : i \geq 1\}$ is an independent sequence of independent and identically distributed nonnegative censoring times, defined on the same probability space. Set $T = T_1, C = C_1, Y = Y_1, \delta = \delta_1$. For example, survival data in clinical trials or failure time data in reliability studies are often subject to such censoring. More specifically, many statistical experiments result in incomplete samples, even under well-controlled conditions. For example, clinical data for surviving most types of disease are usually censored by other competing risks to life that result in death. Liebscher et al., 1997 studied the strong convergences and the weak convergences of the Kaplan Meier Kaplan et al., 1958 estimator in the Harris recurrent Markov chain framework. For example, in reliability statistics, failure of one element of a system can be censored by failure of another element of the same system, provided both elements are connected in series. However, failure times of elements may strongly depend on the load the system is subject too. Consequently, failure time T and censoring time C may become dependent; for instance, refer to Liebscher et al., 1997. In this example, assume that T and C are independent and $Y = T \wedge C$, we have $\mathbb{P}(T > t) = \mathbb{P}(Y > t)/\mathbb{P}(C > t)$. The censoring distribution has a cumulative hazard similar to the failure hazard:*

$$Q(t) = \int_0^t \frac{\mathbb{P}(Y \in du, \delta = 0)}{\mathbb{P}(Y > u)}, \quad \mathbb{P}(C > t) = \exp(-Q(t)).$$

So $\mathbb{P}(T > t) = \mathbb{P}(Y > t) \exp(Q(t))$. Let $\bar{F}_Y(t) = \mathbb{P}(Y > t)$ and $F_{Y_0}(t) = \mathbb{P}(Y \leq t, \delta = 0)$. By expanding the exponential and rewriting each Q as $\int dQ$:

$$\mathbb{P}(T > t) = \mathbb{P}(Y > t) \left[1 + \int_0^t \frac{F_{Y_0}(du_1)}{\bar{F}_Y(u_1)} + \int_0^t \frac{F_{Y_0}(du_1)}{\bar{F}_Y(u_1)} \int_0^{u_1} \frac{F_{Y_0}(du_2)}{\bar{F}_Y(u_2)} + \dots \right].$$

The first-order term can be written as follows, for $k \geq 1$:

$$\int_0^t \frac{F_{Y_0}(du_1)}{\bar{F}_Y(u_1)} = \int_0^t \sum_{k_1=1}^{\infty} F_Y(u_1)^{k_1} F_{Y_0}(du_1) = \int_0^t \sum_{k_1=1}^{\infty} \gamma_{k_1}(du_1).$$

The empirical version of $\gamma_k(du)$ is

$$q_k(x_1, \dots, x_k; du) = \mathbb{1}\{y_1 \in du, \delta_1 = 0\} \mathbb{1}\{y_2 \leq y_1\} \dots \mathbb{1}\{y_k \leq y_1\},$$

where $x_i = (y_i, \delta_i)$, so that $\mathbb{P}q_k(X_1, \dots, X_k; t) = \int_0^t \gamma_k(du)$. The first three terms of the grand kernel have respectively the form:

$$\mathbb{P}(Y > t) \leftrightarrow \mathbb{1}\{Y_n > t\}, \tag{7.4}$$

$$\begin{aligned}
 \mathbb{P}(Y > t)Q(t) &= \mathbb{P}(Y > t) \int_0^t \sum_{k_1=1}^{\infty} \gamma_{k_1}(du_1) \leftrightarrow \mathbb{1}\{Y_n > t\} \int_0^t \sum_{k_1=1}^{n-1} q_{k_1}(X_1, \dots, X_{k_1}; du_1) \\
 &= \mathbb{1}\{Y_n > t\} \sum_{k_1=1}^{n-1} q_{k_1}(X_1, \dots, X_{k_1}; t) \\
 &= \mathbb{1}\{Y_n > t\} \mathbb{1}\{Y_1 \leq t, \delta_1 = 0\} \sum_{k_1=1}^{n-1} \prod_{j=1}^k \mathbb{1}\{Y_j \leq Y_1\}, \\
 \mathbb{P}(Y > t)Q(t)^2/2 &\leftrightarrow \mathbb{1}\{Y_n > t\} \int_0^t \sum_{k_1=1}^{n-2} q_{k_1}(X_1, \dots, X_{k_1}; du_1) \\
 &\quad \times \int_0^{u_1} \sum_{k_2=1}^{n-k_1-1} q_{k_2}(X_{k_1+1}, \dots, X_{k_1+k_2}; du_2).
 \end{aligned}$$

We now extend the general expansion of the grand kernel. So let $S_m(t) = \{(u_1, \dots, u_m) \in \mathbb{R}^m : 0 \leq u_1 \leq \dots \leq u_m \leq t\}$, so that

$$\int_{S_2(t)} f(du_1) g(du_2) = \int_0^t f(du_1) \int_0^{u_1} g(du_2).$$

In the parameter expansion, we have infinite sums over index sets of the form $\{1, 2, \dots\}^m$. Let

$$K_m = \{(k_1, \dots, k_m) \in \mathbb{N}^m\}.$$

Now, by the the Péano series expansion, see for instance Gill et al., 1990, we can write the fully expanded parameter:

$$\mathbb{P}(T > t) = \mathbb{P}(Y > t) \sum_{i=0}^{\infty} \int_{\mathbf{u} \in S_i(t)} \sum_{\mathbf{k} \in K_i} \prod_{j=1}^i \gamma_{k_j}(du_j).$$

Let $\mathbf{X}(k_1 + (1, \dots, k_2)) \stackrel{\text{def}}{=} (X_{k_1+1}, \dots, X_{k_1+k_2})$ and $J_{m,n} = \{(k_1, \dots, k_m) \in \mathbb{N}^m : \sum_j k_j < n\}$, the grand kernel is:

$$H_n(X_1, \dots, X_n) = \mathbb{1}\{Y_n > t\} \sum_{i=0}^{n-1} \int_{\mathbf{u} \in S_i(t)} \sum_{\mathbf{k} \in J_{i,n}} \prod_{j=1}^i q_{k_j} \left(\mathbf{X} \left(\sum_{\ell < j} k_\ell + (1, \dots, k_j) \right); du_j \right).$$

The grand kernel has the following expectation

$$\mathbb{P}H_n(t) = \mathbb{P}(Y > t) \sum_{i=0}^{n-1} \int_{\mathbf{u} \in S_i(t)} \sum_{\mathbf{k} \in J_{i,n}} \prod_{j=1}^i \gamma_{k_j}(du_j).$$

C. M. Heilig, 1997, Proposition 5.6 showed that this IDUP is equivalent to the Kaplan-Meier estimator. The IDUS offers an alternative way to investigate the Kaplan-Meier estimator that is entirely different from the view of Alvarez-Andrade et al., 2013; Bouzebda

and El-hadjali, 2020a; Liebscher et al., 1997. In the last mentioned references, the authors heavily use the notion of the Hadamard differentiability in connection with the weak convergence of the bivariate empirical processes that involve more intricate steps of the proof. Finally, in Chapter 5 of C. M. Heilig, 1997, IDUP theory is applied to several example estimators in renewal theory and survival analysis. He also compares the IDUP method with other techniques for obtaining similar results for the examples.

In DiCiccio et al., 2022, some motivating examples of the IOUS are presented, including a class of Hodges-Lehmann estimators, subsampling estimators, maximin tests, and combining p -values through data splitting. Let us recall the following.

Example 19 (Subsampling Distribution). *Let $\{X_i\}_{i=1}^n$ be a sequence of Harris-recurrent Markov chain random variables with distribution \mathbb{P} , where interest focuses on a real-valued parameter $\xi(\mathbb{P})$. Assume $\hat{\xi}_n = \hat{\xi}_n(X_1, \dots, X_n)$ is an estimator of $\xi(\mathbb{P})$. Fix $1 < k < n$ and let S_1, \dots, S_N be the $N = \binom{n}{k}$ subsets of size k taken without replacement from the data, ordered in any fashion. For a given hypothesized value of ξ , say ξ_0 , let $J_n(t, \mathbb{P})$ be the true c.d.f. of $\tau_n(\hat{\xi}_n - \xi_0)$, evaluated at some generic t . Typically, $\tau_n = \sqrt{n}$. Then, a subsampling estimator of $J_n(t, \mathbb{P})$ is given by*

$$U_n(t) = \frac{1}{N} \sum_{i=1}^N I \left\{ \tau_k \left(\hat{\xi}_k(S_i) - \xi_0 \right) \leq t \right\}.$$

Evidently, for each t , $U_n(t)$ is a U -statistic of degree k . In order to consistently estimate the true distribution $J_n(t, \mathbb{P})$, it is generally required that $k \rightarrow \infty$.

8 Mathematical developments

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

Proof of Theorem 4.1.

Note that the envelope \bar{G}_m is derived from the envelope \bar{H}_m in a similar way \mathcal{G}_m is derived from \mathcal{H}_m . We can assume, without loss of generality, that the envelope \bar{G}_m is bounded by K_m , a finite constant for each m , and

$$\lim_{m \rightarrow \infty} \mathbb{P}^m \bar{G}_m < \infty.$$

We have

$$\begin{aligned} \|\mathbb{P}_n g_m - \mathbb{P} g_m\| &\leq \left\| \mathbb{P}_n \left(g_m \mathbf{1}_{\{\bar{G}_m \leq K_m\}} \right) - \mathbb{P} \left(g_m \mathbf{1}_{\{\bar{G}_m \leq K_m\}} \right) \right\| + \left\| \mathbb{P}_n \left(g_m \mathbf{1}_{\{\bar{G}_m > K_m\}} \right) \right\| \\ &\quad + \left\| \mathbb{P} \left(g_m \mathbf{1}_{\{\bar{G}_m > K_m\}} \right) \right\|. \end{aligned}$$

We observe that:

$$\left\| \mathbb{P} \left(g_m \mathbf{1}_{\{\bar{G}_m > K_m\}} \right) \right\| \leq \left\| \mathbb{P} \left(\bar{G}_m \mathbf{1}_{\{\bar{G}_m > K_m\}} \right) \right\|, \quad (8.1)$$

and

$$\left\| \mathbb{P}_n \left(g_m \mathbf{1}_{\{\bar{G}_m > K_m\}} \right) \right\| \leq \left\| \mathbb{P}_n \left(\bar{G}_m \mathbf{1}_{\{\bar{G}_m > K_m\}} \right) \right\|. \quad (8.2)$$

Using Lemma 2.11 in Levental, 1988, we can see that :

$$\mathbb{P}_n \left(\bar{G}_m \mathbf{1}_{\{\bar{G}_m > K_m\}} \right) \rightarrow \left\| \mathbb{P} \left(\bar{G}_m \mathbf{1}_{\{\bar{G}_m > K_m\}} \right) \right\|. \quad (8.3)$$

By combining (8.1), (8.2) and (8.3), we obtain

$$\limsup_{n \rightarrow \infty} \left\{ \left\| \mathbb{P}_n \left(g_m \mathbf{1}_{\{\bar{G}_m > K_m\}} \right) \right\| \right\} + \left\| \mathbb{P} \left(g_m \mathbf{1}_{\{\bar{G}_m > K_m\}} \right) \right\| \leq 2 \left\| \mathbb{P} \left(\bar{G}_m \mathbf{1}_{\{\bar{G}_m > K_m\}} \right) \right\|.$$

Choose a K_m large enough to make the right hand size of the last inequality as small as we want. Now, we will see, as in Levental, 1988, if we can assume that $g_m \geq 0$, for every $g_m \in \mathcal{G}_m$. In order to prove this hypothesis, it is sufficient to show that for

$$\mathcal{G}_m^+ = \{g_m^+ = g_m \vee 0 \mid g_m \in \mathcal{G}_m\}, \quad \mathcal{G}_m^- = \{g_m^- = g_m \wedge 0 \mid g_m \in \mathcal{G}_m\},$$

if the condition of Euclidean class of functions holds then it holds for \mathcal{G}_m , i.e., if

$$\mathcal{N} \left(\varepsilon \|\bar{G}_m^+\|_{L_2(\mathbb{Q})}, \mathcal{G}_m^+, \|\cdot\|_{L_2(\mathbb{Q})} \right) \leq A\varepsilon^{-V_m}, \quad \mathcal{N} \left(\varepsilon \|\bar{G}_m^-\|_{L_2(\mathbb{Q})}, \mathcal{G}_m^-, \|\cdot\|_{L_2(\mathbb{Q})} \right) \leq A\varepsilon^{-V_m}.$$

Then

$$\mathcal{N} \left(\varepsilon \|\bar{G}_m\|_{L_2(\mathbb{Q})}, \mathcal{G}_m, \|\cdot\|_{L_2(\mathbb{Q})} \right) \leq A\varepsilon^{-V_m}.$$

This assumption permits to work with $g_m \geq 0$, for all $g_m \in \mathcal{G}_m$. Let $(\ell_n - 1)$ be the number of blocks of our Markov chain with elimination of the first and the last one. An explanation of this Markov chain drawing can be found in our previous paper Bouzebda and Soukarieh, 2022a, in the same manner as in Bertail et al., 2011a. For

$$\ell_n = \sum_{i=1}^n \mathbf{1}_{\{X_i \in A\}} = \min\{j, \tau(j) > n\},$$

be the number of successive visit to the atom A . It is obvious to see that

$$\sum_{0 \leq j \leq (\ell_n - 1)} \sum_{k=\mathcal{T}_j+1}^{\mathcal{T}_{j+1}} g_m(X_k) \leq n \mathbb{P}g_m \leq \sum_{0 \leq j \leq \ell_n} \sum_{k=\mathcal{T}_j+1}^{\mathcal{T}_{j+1}} g_m(X_k) + \sum_{k=1}^{\mathcal{T}_0} g_m(X_k).$$

This implies that

$$\begin{aligned} & \left[\frac{(\ell_n - 1)}{n} \right] \left[\frac{1}{(\ell_n - 1)} \sum_{0 \leq j \leq (\ell_n - 1)} \sum_{k=\mathcal{T}_j+1}^{\mathcal{T}_{j+1}} g_m(X_k) - \frac{n}{(\ell_n - 1)} \mathbb{P}g_m \right] \\ & \leq \mathbb{P}_n g_m - \mathbb{P}g_m \leq \left[\frac{\ell_n}{n} \right] \left[\frac{1}{\ell_n} \sum_{0 \leq j \leq \ell_n} \sum_{k=\mathcal{T}_j+1}^{\mathcal{T}_{j+1}} g_m(X_k) + \frac{1}{\ell_n} \sum_{j=1}^{\mathcal{T}_0} g_m(X_j) - \frac{n}{\ell_n} \mathbb{P}g_m \right]. \end{aligned} \quad (8.4)$$

To prove the SLLN of our process, it is sufficient to prove the convergence of each bound to 0 a.s. We have, as $\ell_n \rightarrow \infty$,

$$\left\| \frac{1}{\ell_n} \sum_{j=1}^{\mathcal{T}_0} g_m(X_j) \right\| = \frac{1}{\ell_n} \left\| \frac{\boldsymbol{\tau}(0)}{\boldsymbol{\tau}(0)} \sum_{j=1}^{\mathcal{T}_0} g_m(X_j) \right\| \leq \frac{1}{\ell_n} \times \|\boldsymbol{\tau}(0)\| \times \mathbb{P}_A \left(\sum_{j=1}^{\mathcal{T}_0} g_m(X_j) \right) \rightarrow 0 \quad \text{a.s.}$$

As well as

$$\boldsymbol{\tau}(0) \mathbb{P}_A \left(\sum_{j=1}^{\mathcal{T}_0} g_m(X_j) \right) < \infty.$$

Keep in mind that according to Radulović, 2004, Lemma 3.2 we have:

$$\mathbb{P}_\nu \left(\lim_{n \rightarrow +\infty} \frac{\ell_n}{n} \rightarrow \alpha^{-1} = (\mathbb{E}(\boldsymbol{\tau}))^{-1} \right) = 1.$$

We now treat the following term

$$\left\| \frac{1}{\ell_n} \sum_{0 \leq j \leq \ell_n} \sum_{k=\mathcal{T}_j+1}^{\mathcal{T}_{j+1}} g_m(X_k) - \mathbb{E}(\boldsymbol{\tau}) \mathbb{P}g_m \right\|.$$

To this end, we will apply the following theorem.

Theorem 8.1 (PollardPollard, 1984). *For each n , let \mathcal{F}_n be a permissible class of functions whose covering numbers satisfy*

$$\sup_Q \mathcal{N} \left(\varepsilon^2/2, \mathcal{F}_n, \|\cdot\|_{L_1(\mathbb{Q})} \right) \leq A(\varepsilon^2/2)^{-V}, \quad \text{for } 0 < \varepsilon < 1,$$

with constants A and V not depending on n and a measure \mathbb{Q} . Let $\{\alpha_n\}$ be a non-increasing sequence of positive numbers for which $n\delta_n^2\alpha_n^2 \gg \log n$. If $|f| \leq 1$ and $(\mathbb{P}f^2)^{1/2} \leq \delta_n$ for

each f in \mathcal{F}_n , then, we have almost surely

$$\sup_{\mathcal{F}_n} |\mathbb{P}_n f - \mathbb{P} f| \ll \delta_n^2 \alpha_n.$$

Without loss of generality, we will consider that the envelop function is less than 1. Using the proof of the last theorem, found in Pollard, 1984, Theorem 37, and the fact that

$$\mathcal{N}(\varepsilon \|\bar{G}_m\|, \mathcal{G}_n, \|\cdot\|_{L_1(\mathbb{P})}) \leq \mathcal{N}(\varepsilon^2/2, \mathcal{G}_n, \|\cdot\|_{L_1(\mathbb{P})}) \leq A(\varepsilon^2/2)^{-V_m},$$

the left and the remaining of right hand sizes of (8.4) must converge to 0, this completes the proof.

Proof of Theorem 4.2.

In order to prove the weak convergence for an empirical process, as it is well known, it is sufficient and necessary to prove finite dimensional convergence and stochastic equicontinuity of the U -process. The finite dimensional convergence will be considered if, for every collection $\{g_m^{(1)}, \dots, g_m^{(k)}\}$,

$$\left(\sqrt{n}(\mathbb{P}_n - \mathbb{P})g_m^{(1)}, \dots, \sqrt{n}(\mathbb{P}_n - \mathbb{P})g_m^{(k)}\right), \quad (8.5)$$

converges in probability to $(\mathbb{G}_p(g_m^{(1)}), \dots, \mathbb{G}_p(g_m^{(k)}))$, using the second and the third conditions, the Lindeberg central limit theorem implies that $\sqrt{n}(\mathbb{P}_n - \mathbb{P})g_m$ converges in probability to the Gaussian process $\mathbb{G}_p(g_m)$. Combining this theorem with the classical proof for the multivariate CLT in (8.5), whose essential idea is to demonstrate that for every fixed $a_1, \dots, a_k \in \mathbb{R}$,

$$\sum_{i=1}^k a_i \sqrt{n} (\mathbb{P}_n g_m^{(i)} - \mathbb{P} g_m^{(i)}) \rightarrow N(0, \sigma^2), \quad (8.6)$$

where

$$\sigma^2 = \sum_{i=1}^k a_i^2 \text{Var} \left(\sqrt{n} (\mathbb{P}_n g_m^{(i)} - \mathbb{P} g_m^{(i)}) \right) + \sum_{i \neq j} a_i a_j \text{Cov} \left(\sqrt{n} (\mathbb{P}_n g_m^{(i)} - \mathbb{P} g_m^{(i)}), \sqrt{n} (\mathbb{P}_n g_m^{(j)} - \mathbb{P} g_m^{(j)}) \right).$$

Take $q_m = \sum_{i=1}^k a_i g_m^{(i)}$, and by linearity, we have, in probability,

$$\sqrt{n} (\mathbb{P}_n q_m - \mathbb{P} q_m) \rightarrow \mathbb{G}_p(q_m).$$

Consequently, the finite dimensional convergence is proved.

Remark 8.2. We can refer to Chung, 1967; Radulović, 2004 to prove the CLT of an empirical process for Markov chain, so under condition (A.5), which is a necessary condition for the CLT, we can see that $\sqrt{n}(\mathbb{P}_n - \mathbb{P})g_m$ converges in probability to the Gaussian process $\mathbb{G}_p(g_m)$, and then we use equation (8.6) with the property of linearity to establish the finite dimensional convergence. The use of these references does not imposed the two conditions (A.2) and (A.3).

We consider now the equicontinuity, in the same way as Levental, 1988, we can see that

$$\begin{aligned}
\sqrt{n} |\mathbb{E}(g_m - \check{g}_m)| &\leq n^{-\frac{1}{2}} \left| \sum_{j=1}^{\mathcal{T}_0} (g_m - \check{g}_m)(X_j) - \mathbb{P}(g_m - \check{g}_m) \cdot \boldsymbol{\tau}(0) \right| \\
&+ n^{-\frac{1}{2}} \left| \sum_{1 \leq j \leq \ell_n - 1} \sum_{k=\mathcal{T}_j+1}^{\mathcal{T}_{j+1}} (g_m - \check{g}_m)(X_k) - \mathbb{E}(\boldsymbol{\tau}) \mathbb{P}(g_m - \check{g}_m) \right| \\
&+ n^{-\frac{1}{2}} \left| \sum_{\mathcal{T}_{\ell_n+1} \leq j \leq n} [(g_m - \check{g}_m)(X_j) - \mathbb{P}(g_m - \check{g}_m)] \right| \\
&+ n^{-\frac{1}{2}} |[\mathbb{E}(\boldsymbol{\tau}(1) - \boldsymbol{\tau}(0)) \cdot (\ell_n - 1) - \boldsymbol{\tau}(\ell_n - 1) + \boldsymbol{\tau}(0)] \mathbb{P}(g_m - \check{g}_m)| \\
&= A + B + C + D. \tag{8.7}
\end{aligned}$$

For $g_m - \check{g}_m \in \mathcal{G}_m$, observe that

$$A = n^{-\frac{1}{2}} \left| \sum_{j=1}^{\mathcal{T}_0} (g_m - \check{g}_m)(X_j) - \mathbb{P}(g_m - \check{g}_m) \cdot \boldsymbol{\tau}(0) \right| \leq n^{-\frac{1}{2}} \times \boldsymbol{\tau}(0) \times \mathbb{E}_A \left(\sum_{j=1}^{\mathcal{T}_0} (g_m - \check{g}_m)(X_j) \right) \rightarrow 0.$$

For the second term in (8.7), we infer that

$$\begin{aligned}
B &= \left(n^{-\frac{1}{2}} \sum_{\mathcal{T}_{\ell_n+1} \leq j \leq n} [(g_m - \check{g}_m)(X_j) - \mathbb{P}(g_m - \check{g}_m)] \right)^2 \\
&\leq \left(4 \mathbb{E}_A \left(\sum_{j=1}^{\mathcal{T}_0} (g_m - \check{g}_m)(X_j) \right) \cdot (n - \mathcal{T}(\ell_n)) / n^{\frac{1}{2}} \right)^2 \\
&\leq 16 \mathbb{E}_A \left(\sum_{j=1}^{\mathcal{T}_0} (g_m - \check{g}_m)(X_j) \right)^2 \cdot \max_{1 \leq j \leq n} (\mathcal{T}_{j+1} - \mathcal{T}_j)^2 / n.
\end{aligned}$$

We can see that by using condition (A.5)

$$\mathbb{P} \left\{ \max_{1 \leq j \leq n} (\mathcal{T}_{j+1} - \mathcal{T}_j)^2 / n > \varepsilon \right\} \leq n^{-1} \cdot \mathbb{P} \{ \boldsymbol{\tau}^2 / n > \varepsilon \} \rightarrow 0.$$

Note that

$$\boldsymbol{\tau}([\ell_n]) = \boldsymbol{\tau}(0) + \sum_{j=1}^{\lfloor n/\alpha \rfloor - 1} \{ \mathcal{T}_{j+1} - \mathcal{T}_j \}.$$

Let us consider the term D in (8.7). If we assume that

$$(g_m - \check{g}_m) \in [\delta]_m = \{(g_m - \check{g}_m) : g_m - \check{g}_m \in \mathcal{G}_m \text{ and } \|g_m - \check{g}_m\| < \delta_m\},$$

then we infer that

$$\begin{aligned} & |[\mathbb{E}(\boldsymbol{\tau}(1) - \boldsymbol{\tau}(0))(\ell_n - 1) - \boldsymbol{\tau}(\ell_n - 1) + \boldsymbol{\tau}(0)] \mathbb{P}(g_m - \check{g}_m)| / n^{\frac{1}{2}} \\ & \leq \delta_m \left| [\mathbb{E}(\boldsymbol{\tau}(1) - \boldsymbol{\tau}(0))(\ell_n - 1) - \boldsymbol{\tau}(\ell_n - 1) + \boldsymbol{\tau}(0)] / n^{\frac{1}{2}} \right|. \end{aligned} \quad (8.8)$$

By choosing appropriately δ_m , we have the convergence of D to 0 in probability. It remains the study of the term B . We have

$$\begin{aligned} & \left| \sum_{j=1}^{\ell_n-1} \sum_{k=\mathcal{T}_j+1}^{\mathcal{T}_{j+1}} [(g_m - \check{g}_m)(X_k) - \mathbb{E}(\boldsymbol{\tau})\mathbb{P}(g_m - \check{g}_m)] \right| \\ & \leq \left| \sum_{a \leq j \leq b} \sum_{k=\mathcal{T}_j+1}^{\mathcal{T}_{j+1}} [(g_m - \check{g}_m)(X_k) - \mathbb{E}(\boldsymbol{\tau})\mathbb{P}(g_m - \check{g}_m)] \right| \\ & \quad + \left| \sum_{1 \leq j \leq \lfloor n/\mathbb{E}(\boldsymbol{\tau}) \rfloor} \sum_{k=\mathcal{T}_j+1}^{\mathcal{T}_{j+1}} [(g_m - \check{g}_m)(X_k) - \mathbb{E}(\boldsymbol{\tau})\mathbb{P}(g_m - \check{g}_m)] \right|, \end{aligned}$$

where $a = \min(\ell_n - 1, \lfloor n/\mathbb{E}(\boldsymbol{\tau}) \rfloor)$ and $b = \max(\ell_n - 1, \lfloor n/\mathbb{E}(\boldsymbol{\tau}) \rfloor)$. Let us introduce $H = \{\text{integer } s : |s - \lfloor n/\mathbb{E}(\boldsymbol{\tau}) \rfloor| \leq cn^{1/2}\}$. For the first term in the last inequality, we have

$$\begin{aligned} & \left| \sum_{a \leq j \leq b} \sum_{k=\mathcal{T}_j+1}^{\mathcal{T}_{j+1}} [(g_m - \check{g}_m)(X_k) - \mathbb{E}(\boldsymbol{\tau})\mathbb{P}(g_m - \check{g}_m)] \right| \\ & \leq \max_{s \in H} 2 \left| \sum_{\lfloor n/\mathbb{E}(\boldsymbol{\tau}) \rfloor - cn^{1/2} \leq j \leq s} \sum_{k=\mathcal{T}_j+1}^{\mathcal{T}_{j+1}} [(g_m - \check{g}_m)(X_k) - \mathbb{E}(\boldsymbol{\tau})\mathbb{P}(g_m - \check{g}_m)] \right| \\ & \leq \max_{1 \leq s \leq cn^{1/2}} 2 \left| \sum_{1 \leq j \leq s} \sum_{k=\mathcal{T}_j+1}^{\mathcal{T}_{j+1}} [(g_m - \check{g}_m)(X_k) - \mathbb{E}(\boldsymbol{\tau})\mathbb{P}(g_m - \check{g}_m)] \right| \\ & \leq \max_{1 \leq s \leq cn^{1/2}} 2 \left| \sum_{1 \leq j \leq s} \sum_{k=\mathcal{T}_j+1}^{\mathcal{T}_{j+1}} [g_m(X_k) - \mathbb{E}(\boldsymbol{\tau})\mathbb{P}g_m] \right| \\ & \quad + \max_{1 \leq s \leq cn^{1/2}} 2 \left| \sum_{1 \leq j \leq s} \sum_{k=\mathcal{T}_j+1}^{\mathcal{T}_{j+1}} (\check{g}_m(X_k) - \mathbb{E}(\boldsymbol{\tau})\mathbb{P}\check{g}_m) \right| \\ & \leq 2 \sup_{g_m \in \mathcal{G}_m} \max_{1 \leq s \leq cn^{1/2}} 2 \left| \sum_{1 \leq j \leq s} \sum_{k=\mathcal{T}_j+1}^{\mathcal{T}_{j+1}} (g_m(X_k) - \mathbb{E}(\boldsymbol{\tau})\mathbb{P}g_m) \right| \\ & \leq 2 \sup_{g_m \in \mathcal{G}_m} \max_{1 \leq s \leq n} \left\{ 2 \left| \sum_{1 \leq j \leq s} \sum_{k=\mathcal{T}_j+1}^{\mathcal{T}_{j+1}} (g_m(X_k) - \mathbb{E}(\boldsymbol{\tau})\mathbb{P}g_m) \right| \right\}, \end{aligned}$$

where $2n^{1/2}c \leq n$. By dividing the last inequality by \sqrt{n} and using Lemma 3.2 of Levental, 1988 which prove the convergence

$$\left| \frac{1}{n} \sum_{j=1}^n \sum_{k=\mathcal{F}_{j+1}}^{\mathcal{F}_{j+1}} g_m(X_k) - \mathbb{E}(\boldsymbol{\tau}) \mathbb{P} g_m \right| \rightarrow 0.$$

Hence the proof is complete.

8.1 Proof of Theorem 5.1:

] To prove our results, we make use two main steps. Before start the proof, we will recall all necessary material.

Definition 8.3 (de la Peña et al., 1999). *Let $\{e_i\}$ and $\{d_i\}$ be two sequences of random variables adapted to the σ -fields $\{\mathcal{F}_i\}$. Then $\{e_i\}$ and $\{d_i\}$ are tangent with respect to $\{\mathcal{F}_i\}$ if, for all i $\mathcal{L}(d_i | \mathcal{F}_{i-1}) = \mathcal{L}(e_i | \mathcal{F}_{i-1})$, where $\mathcal{L}(d_i | \mathcal{F}_{i-1})$ denotes the conditional probability law of d_i given \mathcal{F}_{i-1} .*

Proposition 8.4 (de la Peña et al., 1999). *For any sequence of random variables $\{d_i\}$ adapted to an increasing sequence \mathcal{F}_i of σ -algebras, there always exists a decoupled sequence $\{e_i\}$ (on a possibly enlarged probability space) which is tangent to the original sequence and in addition conditionally independent given a master σ -field g . Frequently $g = \sigma(\{d_i\})$.*

Remark 8.5. *It is possible to obtain more than one tangent decoupled to any sequence d_i . For example, if we want two tangents, it is sufficient to take two \mathcal{F}_{i-1} independent repetitions of d_i , e_i and \tilde{e}_i , independent from each other and having the same distribution.*

Definition 8.3 and Proposition 8.4 will be used for the purpose of decoupling our Markov random variables to independent random variables, which is the important advantage behind decoupling. de la Peña et al., 1999 studied the idea of decoupling and proposed the idea of tangent sequence for dependent random variables. Recently, many authors use the tangent sequence as McDonald et al., 2017; Rakhlin et al., 2011.

The use of tangent sequence represents the first step in our proof, the second is to studied the symmetrization technique, we will adopt the results of C. Heilig et al., 2001 for the symmetrization of IDUS, extended nicely from Arcones et al., 1993. Let us recall the following lemma.

Lemma 8.1 (Bonami's inequality). *Let $u(i_1, \dots, i_m)$ be a mapping that is permutation symmetric in its m indices. For each integer $s \geq 1$, we have*

$$\mathbb{P} \left[\sum_{\mathbf{i} \in (n, m)} \xi_{i_1} \dots \xi_{i_m} u(i_1, \dots, i_m) \right]^{2s} \leq (2s - 1)^{ms} \|u\|_2^{2s},$$

where $\|u\|_2^2 \stackrel{\text{def}}{=} \sum_{\mathbf{i} \in (n, m)} u(i_1, \dots, i_m)^2$, and $\{\xi_i\}_{i=1}^n$ is a sequence of independent sign variables placing mass $\frac{1}{2}$ at $+1$ and -1 .

Lemma 8.2 (Chaining inequality). (Nolan et al., 1987) *Let $\Psi(\cdot)$ be a convex, strictly increasing function on $[0, \infty)$ with $0 \leq \Psi(0) \leq 1$. Let p be a positive integer. Suppose that the set T is endowed with pseudometric d , and the stochastic process $\{Z(t) : t \in T\}$ satisfies the following conditions.*

(i) *If $d(s, t) = 0$, then $Z(s) = Z(t)$ almost surely.*

(ii) *If $d(s, t) > 0$, then*

$$\mathbb{P} [\Psi (|Z(s) - Z(t)|^p / d(s, t)^p)] \leq 1.$$

(iii) *There exists a point t_0 in T for which $\delta = \sup_T d(t, t_0) < \infty$.*

(iv) *The sample paths of Z are continuous.*

Then

$$\left[\mathbb{P} \left(\sup_T |Z(t) - Z(t_0)|^p \right) \right]^{1/p} \leq 8 \int_0^{\delta/4} [\Psi^{-1}(\mathcal{N}(x, d, T))]^{1/p} dx, \quad (8.9)$$

where $\mathcal{N}(x, d, T)$ is the covering number defined in the previous section.

Theorem 8.6. (Maximal inequality) [Sherman, 1994] *Let \mathcal{H} be a class of degenerate functions $\{h\}$ of degree j with envelope H . Assume $\mathbb{P}^j H^2 < \infty$. Let p and r be positive integers, and let*

$$\Gamma_j = [64j!(16pr - 8)^j]^{p/2}, \quad \gamma_{n,j} = [U_{2n,j}(H^2)]^{1/2}, \quad \delta_{n,j} = \sup_{\mathcal{H}} [U_{2n,j}(h^2)]^{1/2} / 4 [U_{2n,j}(H^2)]^{1/2}$$

Then

$$\mathbb{P} \left[\sup_{\mathcal{H}} |n^{j/2} U_{n,j}(h)|^p \right] \leq \Gamma_j \mathbb{P} \left[\gamma_{n,j} \int_0^{\delta_{n,j}} \mathcal{N}(x, d_j, \mathcal{H})^{1/2pr} dx \right]^p.$$

Lemma 8.3 (C. M. Heilig, 1997). *Let $\mathbf{Y}_i = \{Y_1^{(i)}, \dots, Y_n^{(i)}\}$ and $\mathbf{Y}'_i = \{Y_1'^{(i)}, \dots, Y_n'^{(i)}\}$. Define $W_i = \mathbf{Y}_i$ and $W_{n+i} = \mathbf{Y}'_i$, for $i = 1, \dots, n$.*

$$\sum_{\mathbf{i} \in I(n, m)} [h^\circ(\mathbf{Y}_i, \mathbf{Y}'_i)]^2 \leq 2^m \sum_{\mathbf{j} \in I((2n), m)} h(W_{j_1}, \dots, W_{j_m})^2.$$

It is worth noting that Bonami's inequality is used to prove the second condition of the chaining lemma. The interested reader can find the details for the proofs of the lemmas, proposition and theorem above in C. Heilig et al., 2001. We are now equipped to start the proof.

Let $\mathbf{X} = \{X_1, \dots, X_n\}$ be a sequence of Markov chain. Let $\mathbf{Y} = \{Y_1^{(i)}, \dots, Y_n^{(i)}\}$, for $i = 1, \dots, n$, be n independent copies of the original random variables. Moreover, we can see that $Y_i^{(i)}$ have the same distribution as X_i . We have:

$$\begin{aligned} \mathbb{P}\left(\sup_{h_m \in \mathcal{H}} U_n(\tilde{h}_m)\right) &= \mathbb{P}_{\mathbf{X}} \left(\sup_{h_m \in \mathcal{H}} \binom{n}{m}^{-1} \sum_{\mathbf{i} \in I(n,m)} \tilde{h}_m(X_{i_1}, \dots, X_{i_m}) \right) \\ &= \mathbb{P}_{\mathbf{X}} \left(\sup_{h_m \in \mathcal{H}} \mathbb{P}_{\mathbf{Y}} \left[\binom{n}{m}^{-1} \sum_{\mathbf{i} \in I(n,m)} \tilde{h}_m(Y_{i_1}^{(1)}, \dots, Y_{i_m}^{(m)}) \right] \right) \\ &\leq \mathbb{P}_{\mathbf{X}} \mathbb{P}_{\mathbf{Y}} \left(\sup_{h_m \in \mathcal{H}} \binom{n}{m}^{-1} \sum_{\mathbf{i} \in I(n,m)} \tilde{h}_m(Y_{i_1}^{(1)}, \dots, Y_{i_m}^{(m)}) \right) \text{ (By Jensen's inequality).} \end{aligned}$$

The last inequality are valid under condition of existence of

$$\mathbb{P}_{\mathbf{Y}} \left(\sup_{h_m \in \mathcal{H}} \binom{n}{m}^{-1} \sum_{\mathbf{i} \in I(n,m)} \tilde{h}_m(Y_{i_1}^{(1)}, \dots, Y_{i_m}^{(m)}) \right).$$

Then we succeed to switch from dependence variables to the independence setting. We will pass to treat the last inequality, which have the same distribution of $\mathbb{P}\left(\sup_{h_m \in \mathcal{H}} U_n(\tilde{h}_m)\right)$, as IDUS of i.i.d.random variables. We will use the complete sign-symmetrization technique. This technique can be summarized as follows:

- Introduce a second i.i.d sample $\{\mathbf{Y}'\}$ independent of $\{\mathbf{Y}\}$, with the same distribution.
- Introduce an i.i.d sign-variables $\{\varepsilon_i\}_{i=1}^n$ independent of both samples $\{\mathbf{Y}'\}$ and $\{\mathbf{Y}\}$, taking values ± 1 such that $\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = -1) = 1/2$.
- Investigate some approximations and moment inequality using chaining lemma and Bonami's inequality mentioned above to get a maximal inequality.
- Applied results of the last step to the U -process generated from the i.i.d sequence.

Let $\{\mathbf{Y}, \mathbf{Y}'\}$ be independent variables defined above. Let \mathcal{H} be a class of functions related to the kernel h . for $j \in \{1, \dots, m\}$, define:

$$h^\circ(\mathbf{Y}_i, \mathbf{Y}'_i) := \sum_{b=0}^{2^m-1} (-1)^{\sum b_j} h\left(Y_{i_1}^{b_1}, \dots, Y_{i_m}^{b_m}\right),$$

for (b_1, \dots, b_k) is a binary expansion of an integer b taking value in $[0, 2^m - 1]$. This kernel is defined in a way such that:

$$\mathbb{P}_{\mathbf{Y}} \left[\sum_{b=0}^{2^m-1} (-1)^{\sum b_j} h_m \left(Y_{i_1}^{b_1}, \dots, Y_{i_m}^{b_m} \right) \mid Y_1, \dots, Y_n \right] = h \left(Y_{i_1}^{(1)}, \dots, Y_{i_m}^{(m)} \right).$$

Return to the main equation, we have:

$$\begin{aligned} & \mathbb{P}_{\mathbf{Y}} \sup_{h_m \in \mathcal{H}} \left| n^{m/2} \binom{n}{m}^{-1} \sum_{\mathbf{i} \in I(n,m)} h_m(Y_{i_1}^{(1)}, \dots, Y_{i_m}^{(m)}) \right|^p \\ &= \mathbb{P}_{\mathbf{Y}} \sup_{h_m \in \mathcal{H}_m} \left| n^{m/2} \binom{n}{m}^{-1} \sum_{\mathbf{i} \in I(n,m)} \left[\mathbb{P}_{\mathbf{Y}} \left(\sum_{b=0}^{2^m-1} (-1)^{\sum b_j} h_m \left(Y_{i_1}^{b_1}, \dots, Y_{i_m}^{b_m} \right) \mid Y_1, \dots, Y_n \right) \right] \right|^p \\ &\leq \mathbb{P}_{\mathbf{Y}} \sup_{h_m \in \mathcal{H}} \left| n^{m/2} \binom{n}{m}^{-1} \sum_{\mathbf{i} \in I(n,m)} h^\circ(\mathbf{Y}_i, \mathbf{Y}'_i) \right|^p. \end{aligned} \quad (8.10)$$

We will now move on to the sign variables. As we will mention before, we introduce an i.i.d sign-variables $\{\varepsilon_i\}_{i=1}^n$ independent of both samples $\{\mathbf{Y}'\}$ and $\{\mathbf{Y}\}$, taking values ± 1 such that $\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = -1) = 1/2$. We will define a new double samples $\{Z_i, Z'_i\}$ in a way that:

$$(Z_i, Z'_i) \stackrel{\text{def}}{=} \{\varepsilon_i = 1\} (Y_i, Y'_i) + \{\varepsilon_i = -1\} (Y'_i, Y_i).$$

These sample are constructed by conserving the same joint distribution of $\{\mathbf{Y}'\}$ and $\{\mathbf{Y}\}$, that is why we can replace $\sum_{\mathbf{i} \in I(n,m)} h^\circ(\mathbf{Y}_i, \mathbf{Y}'_i)$ by $\sum_{\mathbf{i} \in I(n,m)} h^\circ(\mathbf{Z}_i, \mathbf{Z}'_i)$ in (8.10). So we have, for $j \in \{1, \dots, m\}$:

$$\begin{aligned} & \mathbb{P}_{\mathbf{Z}} \sup_{h_m \in \mathcal{H}_m} \left[\sum_{\mathbf{i} \in I(n,m)} h^\circ(\mathbf{Z}_i, \mathbf{Z}'_i) \right] \\ &= \mathbb{P}_{\mathbf{Z}} \sup_{h_m \in \mathcal{H}_m} \left[\sum_{\mathbf{i} \in I(n,m)} \varepsilon_{i_1} \sum_{b=0}^{2^m-1} (-1)^{\sum b_j} h_m \left(Y_{i_1}^{b_1}, Z_{i_2}^{b_2}, \dots, Z_{i_m}^{b_m} \right) \right] \\ &= \mathbb{P}_{\mathbf{Z}} \sup_{h_m \in \mathcal{H}_m} \left[\sum_{\mathbf{i} \in I(n,m)} \varepsilon_{i_1} \varepsilon_{i_2} \sum_{b=0}^{2^m-1} (-1)^{\sum b_j} h_m \left(Y_{i_1}^{b_1}, Y_{i_2}^{b_2}, Z_{i_3}^{b_3}, \dots, Z_{i_m}^{b_m} \right) \right] \\ &= \mathbb{P}_{\mathbf{Z}} \sup_{h_m \in \mathcal{H}_m} \left[\sum_{\mathbf{i} \in I(n,m)} \varepsilon_{i_1} \dots \varepsilon_{i_m} \sum_{b=0}^{2^m-1} (-1)^{\sum b_j} h_m \left(Y_{i_1}^{b_1}, \dots, Y_{i_m}^{b_m} \right) \right] \\ &= \mathbb{P}_{\mathbf{Z}} \sup_{h_m \in \mathcal{H}_m} \left[\sum_{\mathbf{i} \in I(n,m)} \varepsilon_{i_1} \dots \varepsilon_{i_m} h^\circ(\mathbf{Y}_i, \mathbf{Y}'_i) \right]. \end{aligned}$$

Let

$$U_{n,m}^\circ(h) = \binom{n}{m}^{-1} \sum_{\mathbf{i} \in I(n,m)} \varepsilon_{i_1} \dots \varepsilon_{i_k} h^\circ(\mathbf{Y}_i, \mathbf{Y}'_i), \quad (8.11)$$

adjust the inequality (8.10) to consider the complete sign symmetrized process $U_{n,k}^\circ h$ in place of

$$\binom{n}{m}^{-1} \sum_{\mathbf{i} \in I(n,m)} h_m(Y_{i_1}^{(i_1)}, \dots, Y_{i_k}^{(i_k)})$$

and we obtain the following:

$$\mathbb{P}_{\mathbf{Y}} \sup_{h_m \in \mathcal{H}_m} \left| n^{m/2} \binom{n}{m}^{-1} \sum_{\mathbf{i} \in I(n,m)} h_m(Y_{i_1}^{(i_1)}, \dots, Y_{i_m}^{(i_m)}) \right|^p \leq \mathbb{P}_{\mathbf{Z}} \sup_{h_m \in \mathcal{H}_m} \left| n^{m/2} U_{n,m}^\circ(h) \right|^p. \quad (8.12)$$

At that time, the IDUS represented in the right hand size in (8.12) is prepared to be used in the chaining inequality. If $U_{2n,m} \bar{H}_m^2$ is finite, the pseudo-metric related to the inequality is:

$$d_{2n,m}(h_m, \check{h}_m) = \left[\frac{U_{2n,m} (h_m - \check{h}_m)^2}{U_{2n,m} \bar{H}_m^2} \right]^{1/2}.$$

The stochastic process $\mathbb{Z}(t)$ defined in the inequality will be defined as $n^{m/2} U_{n,m}^\circ h / [U_{2n,m} \bar{H}_m^2]^{1/2}$, in this case inequality (8.9) will be:

$$\left[\mathbb{P}_W \left(\sup_{h_m \in \mathcal{H}} \left| n^{m/2} U_{n,m}^\circ(h_m - \check{h}_m) / [U_{2n,m} \bar{H}_m^2]^{1/2} \right|^p \right) \right]^{1/p} \leq 8 \int_0^{\delta/4} \left[\Psi^{-1}(\mathcal{N}(x, d_{2n,m}, \mathcal{H}_m)) \right]^{1/p} dx, \quad (8.13)$$

where the function $\Psi(x)$ is convex and it can be equal to x^{2r}/γ and $\gamma = [m!(16pr - 8)^m]^{pr}$, chosen in this way to fulfill the second condition of Lemma 8.2. We have:

$$\begin{aligned} \mathbb{P}_W \Psi \left(\frac{n^{m/2} |U_{n,m}^\circ(h_m - \check{h}_m)|^p}{d_{2n,m}(h_m, \check{h}_m)^p} \right) &= \gamma^{-1} \mathbb{P}_W \left[\frac{n^{m/2} U_{n,m}^\circ(h_m, \check{h}_m)}{[U_{2n,m}(h_m, \check{h}_m)^2]^{1/2}} \right]^{2rp} \\ &= \gamma^{-1} \mathbb{P}_W \left[\frac{n^{m/2} \binom{2n}{m}^{1/2} \sum_{\mathbf{i} \in I(n,m)} \varepsilon_{i_1} \dots \varepsilon_{i_m} (h_m^\circ - \check{h}_m^\circ)(\mathbf{Y}_i, \mathbf{Y}'_i)}{\binom{n}{m} [\sum_{\mathbf{j} \in I(2n,m)} (h_m - \check{h}_m)(W_{j_1}, \dots, W_{j_m})^2]^{1/2}} \right]^{2rp} \\ &\quad (\text{Take } u(i_1, \dots, i_m) = \frac{(h_m^\circ - \check{h}_m^\circ)(\mathbf{Y}_i, \mathbf{Y}'_i)}{\sum_{\mathbf{j} \in I(2n,m)} (h_m - \check{h}_m)(W_{j_1}, \dots, W_{j_m})} \text{ and } n^{m/2} \binom{2n}{m}^{1/2} \binom{n}{m}^{-1} < 2^m) \\ &\leq \frac{4^{kpr}}{\gamma} [m!^{pr} (2pr - 1)^{mpr}] \mathbb{P}_W \left[\frac{(h_m^\circ - \check{h}_m^\circ)(\mathbf{Y}_i, \mathbf{Y}'_i)^2}{\sum_{\mathbf{j} \in I(2n,m)} (h_m - \check{h}_m)(W_{j_1}, \dots, W_{j_m})^2} \right]^{pr} \quad (\text{by Bonami's inequality}) \\ &\leq \frac{4^{mpr}}{\gamma} [m!^{pr} (2pr - 1)^{mpr}] \times (2^m)^{pr} \quad (\text{by Lemma 8.3}) \\ &= \gamma^{-1} [m!(16pr - 8)^m]^{pr} \leq 1, \quad (\text{as well as } \gamma = [m!(16pr - 8)^m]^{pr}). \end{aligned} \quad (8.14)$$

Here the second condition are achieved. To met the third one, it sufficient to take the

g_m a zero function, the first and the last condition related to the continuity can be seen directly from (8.14) as well as:

$$\left| U_{n,m}^\circ(h_m - \check{h}_m) \right| < 2^m d_{2n,m}(h_m, \check{h}_m).$$

Then, with $\Psi^{-1}(x) = (\gamma x)^{1/2r}$, we obtain from Lemma 8.9 the result (8.13):

$$\mathbb{P}_W \left(\sup_{h_m \in \mathcal{H}_m} \left| U_{n,m}^\circ(h_m - \check{h}_m) / [U_{2n,m} \bar{H}^2]^{1/2} \right|^p \right) \leq \left[8 \int_0^{\delta_{n,m}} [\gamma(\mathcal{N}(x, d_{2n,m}, \mathcal{H}_m))]^{1/2pr} dx \right]^p,$$

with $\Gamma_m = 8^p \gamma^{1/2r} = [64m!(16pr - 8)^m]^{p/2}$, this expectation satisfies

$$\mathbb{P} \sup_{h_m \in \mathcal{H}_m} \left| n^{m/2} U_{n,m}^\circ(h) \right|^p \leq \Gamma_m \mathbb{P} \left[[U_{2n,m} \bar{H}^2]^{1/2} \int_0^{\delta_{n,m}} \mathcal{N}(x, d_{2n,k}, \mathcal{H}_m)^{1/2pr} dx \right]^p. \quad (8.15)$$

Under Condition (5.2), the desired result will be obtained.

8.2 Proof of Theorem 5.2:

] Using the Hoeffding decomposition, we can see that

$$\sqrt{n} [U_n(\tilde{h}_m) - \Theta(h_m)] = \sqrt{n} m U_{n,1}(\tilde{h}_{m(1)}) + \sqrt{n} U_{n,R}(\tilde{h}_m) + \sqrt{n} [\Theta_m(h_m) - \Theta(h)], \quad (8.16)$$

where $U_{n,R}(\tilde{h}_m)$ denotes the remainder term of the Hoeffding decomposition in (3.2) (for $c \geq 2$). Using Chebyshev's inequality, we obtain, if $m \leq n - 2$,

$$\begin{aligned} \mathbb{P} \left(\sqrt{n} |U_{n,R}(\tilde{h}_m)| > \varepsilon \right) &\leq \frac{n}{\varepsilon^2} \sum_{c=2}^m \binom{m}{c}^2 \binom{n}{c}^{-1} \mathbb{P}^m [\tilde{h}_{m(c)}]^2 \leq \frac{n}{\varepsilon^2} \binom{n}{2}^{-1} \sum_{c=2}^m \mathbb{P}^m \left[\binom{m}{c} \tilde{h}_{m(c)} \right]^2 \\ &= \frac{1}{n(1-n^{-1})\varepsilon^2} \mathbb{P}^m h_m^{\star 2}, \end{aligned}$$

for $h_m^* := \binom{m}{c} \tilde{h}_{m(c)}$. If $\mathbb{P}^m (h_m^*)^2 < \infty$ a.s., it is clear that the remainder term of Hoeffding decomposition vanishes when $n \rightarrow \infty$. Now, let us show that the last term in (8.16) converges. Frees, 1989 proposed a sufficient condition that the parameter $\Theta(h)$ exists and finite, and they used the reverse martingale property to show that

$$\Theta_m(h_m) - \Theta(h) = o(n^{-1}).$$

Furthermore, C. M. Heilig, 1997 mentioned that the proof of this term varies according to each example. Finally, the first term of (8.16) can be viewed directly from the Theorem 4.2.

Remark 8.7. *As the degree $m = m_n$ of kernel increases with the sample size, the IDUS can not be assumed reverse martingales, but it can be approximated by a reverse one, with a necessary requirement that the infinite degree kernel exists and is finite. Readers can refer to C. Heilig et al., 2001; C. M. Heilig, 1997, who have represented the approximated martingale in their work.*

Remark 8.8. *The proof of Theorem 5.2 imposed some strong conditions. In C. Heilig et al., 2001, different conditions are imposed with different proof for the weak convergence theorem of infinite degree U -statistics. The main ingredients are the use of the chaining inequality from Nolan et al., 1987 and the maximal inequality in Sherman, 1994 combined with some geometric properties of function classes.*

Remark 8.9. *In this remark, we highlight an alternative method to prove Theorem 5.2. In order to investigate the weak convergence $\sqrt{n}(U_n(h_m) - \mathbb{P}^m h_m)$ to a Gaussian limit distribution \mathbb{G}_p and without the strong condition on the remainder term of the Hoeffding decomposition. This method has the same main steps of the proof of the law of large number (Theorem 5.1) applied to the Hoeffding decomposition of the U -process defined in (8.11) and it can be achieved immediately, it is sufficient to suppose that $\sum_{c=2}^m \phi(c) = o(n^{-1/2})$, where $\phi(c)$ and Γ_c is defined as in Remark 5.1. Then*

$$\sup_{h_m \in \mathcal{H}_m} \sqrt{n} \left| U_n(h_m) - \mathbb{P}^m(h_m) - mU_{n,1}(\tilde{h}_{m(1)}) \right| \rightarrow 0 \quad \text{in probability } \mathbb{P}.$$

8.3 Proof of Theorem 6.1:

] A main step towards the Gaussian approximation and the validity of the bootstrap method is related to the Hoeffding decomposition of the U -statistics. As we have mentioned above, this decomposition can be written as follows

$$U_n^*(\tilde{h}_m) = \mathbb{E}^*(U_n^*) + mU_{n,1}^*(\tilde{h}_{m(1)}) + U_{n,R}^*(\tilde{h}_m). \quad (8.17)$$

To get equation (6.4), the key step is to bound the nonlinear terms of the Hoeffding decomposition. To be more specific, we will prove the convergence of the variance of the last term in (8.17) to zero. After, we will prove that the difference of the variance of the bootstrapped linear term and the principal one converges to 0 also in probability. In the following, we will rigorously prove these steps. Notice that

$$\begin{aligned} \text{Var}^* \left[\sqrt{n^*} U_{n,R}^*(\tilde{h}_m) \right] &= \text{Var}^* \left[\sqrt{n^*} \sum_{c=2}^m \binom{m}{c} \binom{n^*}{c}^{-1} \tilde{h}_{m(c)} \right] \\ &= n^* \sum_{c=2}^m \binom{m}{c}^2 \binom{n^*}{c}^{-2} \text{Var}^*(\tilde{h}_{m(c)}) + 2\sqrt{n^*} \sum_{c,j=2}^m \binom{m}{c} \binom{n^*}{c}^{-1} \binom{m}{j} \binom{n^*}{j}^{-1} \text{Cov}^*(\tilde{h}_{m(c)}, \tilde{h}_{m(j)}) \end{aligned}$$

$$= n^* \sum_{c=2}^m \binom{m}{c}^2 \binom{n^*}{c}^{-2} \mathbb{P}^*(\tilde{h}_{m(c)})^2 + A.$$

To treat the term A in the last inequality, observe that

$$\begin{aligned} A &= 2\sqrt{n^*} \sum_{c,j=2}^m \binom{m}{c} \binom{n^*}{c}^{-1} \binom{m}{j} \binom{n^*}{j}^{-1} \text{Cov}^*(\tilde{h}_{m(c)}, \tilde{h}_{m(j)}) \\ &= 2\sqrt{n^*} \sum_{c,j=2}^m \binom{m}{c} \binom{n^*}{c}^{-1} \binom{m}{j} \binom{n^*}{j}^{-1} \left[\mathbb{P}^*(\tilde{h}_{m(c)}\tilde{h}_{m(j)}) - \mathbb{P}^*(\tilde{h}_{m(c)})\mathbb{P}^*(\tilde{h}_{m(j)}) \right]. \end{aligned}$$

Now, for the bootstrap expectation, we must distinguish in our bootstrap method of two types of variables, X_i^* drawn from the same blocks \mathbf{B}_i and X_i^* drawn from different blocks. We will treat both cases in the following.

- For $a_1, \dots, a_i, b_1, \dots, b_j$ generated from different block:

$$\begin{aligned} & \left| \mathbb{P} \left[\mathbb{P}^* \left(\tilde{h}_{m(c)} \left(X_{a_1}^*, \dots, X_{a_c}^* \right) \tilde{h}_{m(j)} \left(X_{b_1}^*, \dots, X_{b_j}^* \right) \right) \right] \right| \\ &= \left| \mathbb{P} \left[\frac{1}{n^{*m}} \sum_{a_1, \dots, a_c, b_1, \dots, b_j=1}^{n^*} \left(\tilde{h}_{m(c)} \left(X_{a_1}^*, \dots, X_{a_c}^* \right) \tilde{h}_{m(j)} \left(X_{b_1}^*, \dots, X_{b_j}^* \right) \right) \right] \right| \\ &\leq \frac{1}{n^{*m}} \sum_{a_1, \dots, a_c, b_1, \dots, b_j=1}^{n^*} \mathbb{P} \left| \tilde{h}_{m(c)} \left(X_{a_1}^*, \dots, X_{a_c}^* \right) \tilde{h}_{m(j)} \left(X_{b_1}^*, \dots, X_{b_j}^* \right) \right|. \end{aligned}$$

- For a_{s_1}, \dots, a_{k_1} lie in the same block, then they are dependent between them and independent compared with $\{a_1, \dots, a_c\} \setminus \{a_{s_1}, \dots, a_{k_1}\}$.

$$\begin{aligned} & \left| \mathbb{P} \left[\mathbb{P}^* \left(\tilde{h}_{m(c)} \left(X_{a_1}^*, \dots, X_{a_c}^* \right) \tilde{h}_{m(j)} \left(X_{b_1}^*, \dots, X_{b_j}^* \right) \right) \right] \right| \\ &\leq \frac{1}{n^{*m-k}} \sum_{a_1, \dots, a_i, b_1, \dots, b_j=1} \mathbb{P} \left| \tilde{h}_{m(c)} \left(X_{a_1}^*, \dots, X_{a_s}^*, X_{a_{s+1}}^*, \dots, X_{a_{s+k}}^*, \dots, X_{a_c}^* \right) \times \tilde{h}_{m(j)} \left(X_{b_1}^*, \dots, X_{b_j}^* \right) \right| \\ &\leq \frac{1}{n^{*m-k}} \sum_{a_1, \dots, a_c, b_1, \dots, b_j=1}^{n^*} \mathbb{P} \left| \tilde{h}_{m(c)} \left(X_{a_1}^*, \dots, X_{a_i}^* \right) \tilde{h}_{m(j)} \left(X_{b_1}^*, \dots, X_{b_j}^* \right) \right|. \end{aligned}$$

According to Radulović, 2004, Lemma 3.2 $n^*/n \rightarrow 1$ in probability $\mathbb{P} \times \mathbb{P}^*$. We can see that that this expectation converges with a rate $O(n^{-(m-k)})$. Finally, let $\mathcal{H}_{m(j)}$ be a class of function associated with the kernel projections each with envelop $H_{m(j)}$, for $j \in \{1, \dots, m\}$. We have

$$\begin{aligned} \text{Var}^* \left[\sqrt{n^*} U_{n,R}^* \left(\tilde{h}_m \right) \right] &\leq n^* \binom{n^*}{2}^{-2} \sum_{c=2}^m \binom{m}{c}^2 \mathbb{P}^*(\tilde{h}_{m(c)})^2 \\ &+ 2\sqrt{n^*} \sum_{c,j=2}^m \binom{m}{c} \binom{n^*}{c}^{-1} \binom{m}{j} \binom{n^*}{j}^{-1} \frac{1}{n^{*m-k}} \sum_{a_1, \dots, a_c, b_1, \dots, b_j=1}^{n^*} \end{aligned}$$

$$\begin{aligned}
 & \mathbb{P} \left| \tilde{h}_{m(c)} \left(X_{a_1}^*, \dots, X_{a_i}^* \right) \tilde{h}_{m(j)} \left(X_{b_1}^*, \dots, X_{b_j}^* \right) \right| \\
 \leq & \frac{1}{n^*(n^* - 1)^2} \sum_{c=2}^m \binom{m}{c}^2 \mathbb{P}^* (H_{m(c)})^2 \\
 & + 2\sqrt{n^*} \binom{n^*}{2}^{-2} \sum_{c,j=2}^m \binom{m}{c} \binom{m}{j} \frac{1}{n^{*m-k}} \sum_{a_1, \dots, a_c, b_1, \dots, b_j=1}^{n^*} \\
 & \mathbb{P} \left(H_{m(c)} \left(X_{a_1}^*, \dots, X_{a_i}^* \right) H_{m(j)} \left(X_{b_1}^*, \dots, X_{b_j}^* \right) \right) \\
 \leq & \frac{1}{n(n-1)^2} \sum_{c=2}^m \binom{m}{c}^2 \mathbb{P}^* (H_{m(c)})^2 \\
 & + \frac{2}{n^{1/2}(n-1)^2} \sum_{c,j=2}^m \binom{m}{c} \binom{m}{j} \frac{1}{n^{m-k}} \\
 & \sum_{a_1, \dots, a_c, b_1, \dots, b_j=1}^{n^*} \mathbb{P} \left(H_{m(c)} \left(X_{a_1}^*, \dots, X_{a_c}^* \right) H_{m(j)} \left(X_{b_1}^*, \dots, X_{b_j}^* \right) \right) \\
 \leq & \frac{1}{n(n-1)^2} \sum_{c=2}^m \binom{m}{c}^2 \mathbb{P}^* (H_{m(c)})^2 \\
 & + \frac{2}{n^{1/2}(n-1)^2} \sum_{c,j=2}^m \binom{m}{c} \binom{m(n)}{j} \left[\mathbb{P} (H_{m(c)})^2 \right]^{1/2} \left[\mathbb{P} (H_{m(j)})^2 \right]^{1/2}.
 \end{aligned}$$

It is worth be cognizant that if the class of function \mathcal{G}_m is euclidean (A_m, V_m) and bounded above by K_m , then $\mathcal{H}_{m(j)}$ is euclidean $(A_m, 2V_m)$ for the envelop $\mathbf{H}_{m(j)} = 2^j M_m$.

Hence :

$$\text{Var} \left[\sqrt{n} U_{n,R} \left(\tilde{h}_m \right) \right] \xrightarrow{n \rightarrow \infty} 0 \quad \text{and} \quad \text{Var}^* \left[\sqrt{n^*} U_{n,R}^* \left(\tilde{h}_m \right) \right] \xrightarrow{\mathbb{P}^*} 0. \quad (8.18)$$

Now, as it is well known, a random sequence $(X_n : n \in \mathbb{N})$ converges to a random variable X in probability if and only if each subsequence $(X_{n_k} : k \in \mathbb{N})$ contains a further subsequence $(X_{n_{k_j}} : j \in \mathbb{N})$ converges almost surely to X . We can see that under equation (8.18), we can find an arbitrary subsequence n_k such that

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}^* \left[\sqrt{n^*} \left(U_{n_k}^* (h) - \mathbb{E}^* \left[U_{n_k}^* (h) \right] \right) \leq x \right] - \mathbb{P}^* \left[m U_{n_k,1}^* \left(\tilde{h}_{m(1)} \right) \leq x \right] \right| \xrightarrow{\text{a.s.}} 0. \quad (8.19)$$

In that case, the convergence in probability is achieved:

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}^* \left[\sqrt{n^*} \left(U_n^* (h) - \mathbb{E}^* \left[U_n^* (h) \right] \right) \leq x \right] - \mathbb{P}^* \left[m U_{n,1}^* \left(\tilde{h}_{m(1)} \right) \leq x \right] \right| \xrightarrow{n \rightarrow \infty} 0. \quad (8.20)$$

Finally and based on the paper Datta et al., 1993, we can see clearly that:

$$\left| \text{Var}^* \left[\sqrt{n^*} m U_{n,1}^* \left(\tilde{h}_{m(1)} \right) \right] - \text{Var} \left[\sqrt{n} m U_{n,1} \left(\tilde{h}_{m(1)} \right) \right] \right| \xrightarrow{\text{a.s.}} 0. \quad (8.21)$$

Pass now to the proof of stochastic asymptotic equicontinuity in probability. Let

$$\mathcal{H}'_\delta = \left\{ h_m - \check{h}_m : h_m, \check{h}_m \in \mathcal{H}_m, \|h_m - \check{h}_m\|_{L^2(\mathbb{P}^m)} \leq \delta \right\},$$

it is enough to show that a.s.

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P}^* \left\| \sqrt{n^*} (U_n^*(h_m) - U_n(h_m)) \right\|_{\mathcal{H}'_\delta} = 0, \quad (8.22)$$

where for any functional Φ and a class of function \mathcal{H} , we assume $\|\Phi(h)\|_{\mathcal{H}} := \sup_{h \in \mathcal{H}} |\Phi(h)|$

and $0 \in \mathcal{H}'_\delta$. Keep in mind that we have

$$\begin{aligned} \sqrt{n^*} U_n^*(h_m) &= \sqrt{n^*} \mathbb{E}^*(U_n^*) + m U_{n,1}^*(h_{m(1)}) + \sum_{c=2}^m \binom{m}{c} U_{n,c}^*(h_{m(c)}), \\ \sqrt{n} U_n(h_m) &= \sqrt{n} \mathbb{E}(U_n) + \sqrt{n} m U_{n,1}(h_{m(1)}) + \sqrt{n} \sum_{c=2}^m \binom{m}{c} U_{n,c}(h_{m(c)}). \end{aligned}$$

An application of the triangle inequality gives

$$\begin{aligned} \left\| \sqrt{n^*} (U_n^*(h_m) - U_n(h_m)) \right\|_{\mathcal{H}'_\delta} &= \left\| \sqrt{n^*} \mathbb{E}^*(U_n^*) + \sqrt{n^*} m U_{n,1}^*(h_{m(1)}) + \sqrt{n^*} \sum_{c=2}^m \binom{m}{c} U_{n,c}^*(h_{m(c)}) \right. \\ &\quad \left. - \sqrt{n^*} \mathbb{E}(U_n) - \sqrt{n^*} m U_{n,1}(h_{m(1)}) - \sqrt{n^*} \sum_{c=2}^m \binom{m}{c} U_{n,c}(h_{m(c)}) \right\|_{\mathcal{H}'_\delta} \\ &\leq \left\| \sqrt{n} (m U_{n,1}^* - m U_{n,1})(h_{m(1)}) \right\|_{\mathcal{H}'_\delta} + \left\| \sqrt{n} \sum_{c=2}^m \binom{m}{c} U_{n,c}^*(h_{m(c)}) \right\|_{\mathcal{H}'_\delta} + \left\| \sqrt{n} \sum_{c=2}^m \binom{m}{c} U_{n,c}(h_{m(c)}) \right\|_{\mathcal{H}'_\delta} \end{aligned}$$

The first term in the inequality is the bootstrapped Markovian empirical process investigated among other by S. P. Meyn et al., 1993; Radulović, 2004, and under the same assumptions as Theorem 4.2, tends to zero. The second and the third terms are the Hoeffding nonlinear terms converging to zero as n tends to infinity.

Chapter 4

Renewal type bootstrap for U - process Markov chains

This chapter develops the content of an accepted article published in *Markov Processes and Related Fields* with the required modifications to fit this thesis manuscript.

Objective

The primary purpose of the present chapter is to establish bootstrap uniform functional central limit theorems U -processes for Harris recurrent Markov chains over uniform classes of functions satisfying some entropy conditions. To simplify our approach, we will employ the well-known regenerative properties of Markov chains, avoiding some complicated mixing conditions. Our result is obtained under minimal conditions on the envelope function. In addition, we consider an extension to the k Markov chain setting and prove the bootstrap consistency. The theoretical uniform central limit theorems set out below are (or will be) key tools for further developments in Markovian data analysis.

Contents

1	Introduction	95
2	Preliminaries and notation	98
	2.1 Markov chain Assumptions	98
	2.2 U -process	101
3	Hoeffding decomposition	103
	3.1 Technical Assumptions:	105
	3.2 General case	107
	3.3 Regenerative approximation	108
4	Asymptotic theory for Markovian U -statistics	108
5	Uniform bootstrap central limit theorem	110
	5.1 RBB method	110

5.2	Splitting technique	112
5.3	Bootstrap of U -statistic	112
6	Generalized U -statistics	114
6.1	Decomposition of generalized U -statistics	115
6.2	Regenerative Chains	116
6.3	Asymptotic theory for Markovian generalized U -statistics	117
7	Bootstrapping generalized U -statistics	118
8	Examples	120
8.1	Classes of functions	120
8.2	U -statistics	122
9	Mathematical developments	123
10	Appendix	147

1 Introduction

Motivated by a wide variety of applications, a substantial focus has been paid, over the last decades, to the theory of the U -Statistics (initiated in the pioneering work by Hoeffding, 1948) and the U -Processes. U -processes help address complicated statistical issues. Examples include density approximation, non-parametric regression tests, and goodness-of-fit tests. Mainly, in several different ways, U -processes occur in statistics, e.g., as components of higher order terms in von Mises expansions. In the study of estimators (including function estimators) with various degrees of smoothness, U -statistics were used in particular. In this context, for an analysis of a product limit estimator of truncated data, Stute, 1993 applied almost sure uniform bounds for the \mathbb{P} -canonical U -Processes. In addition, Arcones et al., 2006 provided two new normality tests based on the U -processes. Based on standardized findings of Giné et al., 2007b, new measures for normality that use L_1 -distances between regular normal density and local U -statistics based on structured observations as test statistics were implemented by Schick et al., 2011. Joly et al., 2016 addressed the calculation of the mean of multivariate functions in the case of likely heavy-tailed distributions and adopted a median-of-means based on U -statistics. In a wide variety of methodological applications, for instance, to establish the limiting distributions of M -estimators [see, for example, Arcones et al., 1993, Sherman, 1993, Sherman, 1993 and Sherman, 1993], to check the qualitative function characteristics in non-parametric statistics [S. Lee et al., 2009, Ghosal et al., 2000, Abrevaya et al., 2005], or to set cross-validation for density estimation, U -processes are valuable methods. An extension of the finite case of U -statistics to the infinite case is called infinite-order U -statistic. Briefly, the case of an infinite quantity of variables was explained that would later represent a parameter that can be estimated by a finite sum of terms wherein each term can be approximated unbiasedly. This extension, introduced by Frees, 1989, is a useful tool to build simultaneous prediction intervals that measure the uncertainty of existing approaches such as sub-bagging and random forests. The following relevant example is given by Y. Song et al., 2019:

Example 20 (*Simultaneous prediction intervals for random forests* (Y. Song et al., 2019)). Let us consider a training dataset of size n ,

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

where $Y_i \in \mathcal{Y}$ is a vector of features and $Z_i \in \mathbb{R}$ is a response. Let h be a deterministic prediction rule that takes as input a sub-sample $\{X_{i_1}, \dots, X_{i_m}\}$ with $1 \leq m \leq n$ and outputs predictions on d testing points (y_1^*, \dots, y_d^*) in the feature space \mathcal{Y} . For random forests, the tree-based prediction rule is constructed on each sub-sample with additional randomness. Specifically, let $\{W_{\mathbf{i}} : \mathbf{i} \in I_n^m\}$ be a collection of i.i.d. random variables

taking value in a measurable space (S', \mathcal{S}') that are independent of the data X_1^n , where

$$I_n^m = \{\mathbf{i} = (i_1, \dots, i_m) : i_j \in \mathbb{N}, 1 \leq i_j \leq n, i_j \neq i_k \text{ if } j \neq k\}.$$

is the set of all m -tuples of different integers between 1 and n . Let $H : \mathcal{X}^m \times S' \rightarrow \mathbb{R}^d$ be an $\mathcal{S}^m \otimes \mathcal{S}'$ -measurable function such that

$$\mathbb{E}[H(x_1, \dots, x_m, W)] = h(x_1, \dots, x_m).$$

Then predictions of random forests are given by a d -dimensional U -statistic with random kernel H :

$$\hat{U}_n := \frac{(n-m)!}{n!} \sum_{\mathbf{i} \in I_n^m} H(X_{i_1}, \dots, X_{i_m}, W_{\mathbf{i}}), \quad (1.1)$$

where the random kernel H varies with m .

On the other hand, Peng et al., 2022 developed in great depth the notion of generalized U -statistics random forest predictions. The problem of rating instances has gained particular attention in machine learning because of its great significance. In particular rating issues, it is important to compare two separate observations, depending on their observed features, and to determine which is better instead of merely classifying them. Ordering problems have many uses in diverse fields of banking (Data Mining Tool for Direct Marketing Data Extraction), document type classification, and so on . . . In order to solve these frequent problems, U -statistics play an instrumental role, refer to Cléménçon et al., 2008. Historically, the first asymptotic results of U -statistics for the case of independent and identically random variables were provided by Halmos, 1946, v. Mises, 1947 and Hoeffding, 1948, among others. While under weak dependency assumptions, asymptotic results have seen, for example, in Borovkova et al., 2001, Denker et al., 1983 or more recently in Leucht, 2012 and more broadly in Leucht et al., 2013, Bouzebda et al., 2019b and Bouzebda and Nemouchi, 2020. For an excellent reference resources for U -statistics and U -processes, the reader may refer to A. J. Lee, 1990, Arcones et al., 1993, Koroljuk et al., 1994, Arcones, Chen, et al., 1994, Arcones et al., 1995 and Borovskikh, 1996. A valuable contribution in the theory of U -processes is given by de la Peña et al., 1999. They also have a pivotal role in developing the theory of U -processes by following patterns from the theory of empirical process and adding new techniques such as decoupling inequality and randomization. The key request for an estimator is to approximate the confidence intervals. By far, the most preferred confidence interval is a standard confidence interval based on a normal or student t -distribution. These basic intervals are considered useful methods, but they are based on an estimate that can be very misleading in practice. Aiming to set the problem of the estimation of confidence intervals, Efron, 1979 discussed the "Bootstrap" method, which was and still is an attractive solution. In this groundbreak-

ing article, a broad variety of articles in the literature have introduced bootstrap techniques, discussed, researched, and applied them. As one of the most significant concepts in the practice of statistics, the bootstrap has presented a wealth of creative probability problems, which, in essence, provided the foundation for the development of modern mathematical theories. The bootstrap can be briefly described as follows. Let $T(\mathbb{F})$ be a functional of an unknown df $\mathbb{F}(\cdot)$, X_1, \dots, X_n a sample from $\mathbb{F}(\cdot)$, and X_1^*, \dots, X_n^* an independent and identically distributed [i.i.d.] sample with common distribution given by the empirical distribution $\mathbb{F}_n(\cdot)$ of the original sample. The distribution of $\{T(\mathbb{F}_n) - T(\mathbb{F})\}$ is then approximated by that of $\{T(\mathbb{F}_n^*) - T(\mathbb{F}_n)\}$ conditionally on X_1, \dots, X_n , with $\mathbb{F}_n^*(\cdot)$ being the empirical distribution of X_1^*, \dots, X_n^* . The main idea behind the bootstrap is: if the sample is representative of the underlying population, inferences on the population characteristics can be made by resampling the current sample, and an approximation of the distribution of a function of the observations and the underlying distribution will be made by replacing the unknown distribution by the empirical distribution of the sample, for more detail, see Efron et al., 1993 and Shao et al., 1995. Generally, it is well known that the bootstrap functions in the i.i.d. case if and only if the central limit theorem holds for the random variable under consideration. Further, we refer the reader to the Giné et al., 1990 seminal article. There is a huge amount of literature on bootstrap and its applications. It is not the purpose of this chapter to survey this extensive literature. In the present work, we are interested in the bootstrap for the U -processes based on Markov chains. Results on the bootstrap for the empirical processes based on Markov chains are reported in Radulović, 2004 and Ciołek, 2016. Bootstrap of U -statistics in the independent setting are investigated in Csörgő et al., 2013, Arcones and Giné, 1994 and X. Chen, 2018, and in dependent setting in Dehling et al., 2010, Leucht et al., 2013 and Sharipov et al., 2016. The most related work of the present chapter is (2011a). In the last reference, the authors provided an approach to the study of U -statistics in the Markovian setup based on the (pseudo-) regenerative properties of Harris Markov chains. Exploiting the fact that any sample path X_1, \dots, X_n of a general Harris chain X may be divided into asymptotically independent and identically distributed (i.i.d.) data blocks. This kind of regeneration in Markov chains has been detailed in the paper of Nummelin, 1991. A similar approach will be used in our work without restriction to the case $m = 2$. The present chapter considers the general framework of the U -processes in the Markov chain setting and their bootstrapped versions. This investigation is far from trivial and harder to control equicontinuity, constituting a fundamentally unsolved open problem in the literature. Our work aims at filling this gap in the literature. The main difficulties in proving our results are related to the random size of the resampled blocks. This creates problems with random stopping times (in real and Banach space settings). However, as will be proven later, the problem requires much more than “simply” combining ideas from the existing results. In fact, delicate mathematical derivations will be required to cope

with U -processes in our context. This requires the effective application of large sample theory techniques developed for the empirical processes.

The layout of the present chapter is as follows. Section 2 is devoted to the introduction of the Markov framework and the definitions needed in our work. In Section 3, we recall the necessary ingredient for the U -statistics in the Markov setting. We provide some asymptotic results in Section 4, including the consistency, the weak convergence of U -processes, and the law of the iterated logarithm. In Section 5, we will come up with the main results concerning the bootstrap of the U -statistics. In Section 6, we extend our work to the k -Markov chains. The bootstrap results are given in Section 7. In Section 8, we collect some examples of the class of functions and examples of conditional U -statistics. To prevent from interrupting the flow of the presentation, all proofs are gathered in Section 9. A few relevant technical results are given in the Appendix.

2 Preliminaries and notation

Before properly presenting our results and providing a strong base for our discussion, we will need to introduce some notation and fundamental definitions.

2.1 Markov chain Assumptions

Let $X = (X_n)_{n \in \mathbb{N}}$ be a homogeneous ψ -irreducible Markov chain, which means this chain has a stationary transition probabilities, defined on a measurable space (E, \mathcal{E}) , where \mathcal{E} is a separable σ -algebra. Let $\pi(x, dy)$ the transition probability, $\nu = \nu(i)_{i>0}$ initial probability. Therefore, we will denote by \mathbb{P}_ν or just \mathbb{P} the probability measure for $P = (\pi, \nu)$. Likewise, \mathbb{E}_ν will denote the integration with respect to \mathbb{P}_ν . We will further assume that the Markov chain is Harris positive recurrent with an atom \mathbf{A} . Let us recall the following definitions.

Definition 2.1 (Harris-recurrent). *A Markov chain $X = (X_n)_{n \in \mathbb{N}}$ is said to be Harris-recurrent if there exists a σ -finite measure such that, for ψ a positive measure on a countable generated measurable space (E, \mathcal{E}) , $\psi(E) > 0$ and if for all $B \in \mathcal{E}$ with $\psi(B) > 0$ then*

$$\mathbb{P}_\nu(x, \cup_{i=1}^{\infty} (X_i \in B)) = 1 \text{ for any } x \in E.$$

Definition 2.2 (irreducibility). *The chain is ψ -irreducible if there exists a σ -finite measure ψ such that, for all set $B \in \mathcal{E}$, when $\psi(B) > 0$, for any $x \in E$ there exists $n > 0$ such that $\mathbb{P}^n(x, B) > 0$.*

Definition 2.3 (aperiodicity). *Assuming ψ -irreducibility, there exists $d' \in \mathbb{N}^*$ and disjoint sets $D_1, \dots, D_{d'}$ (set $D_{d'+1} = D_1$) positively weighted by ψ such that $\psi(E \setminus \cup_{1 \leq i \leq d'} D_i) = 0$ and $\forall x \in D_i, \mathbb{P}(x, D_{i+1}) = 1$. The period of the chain is the greatest common divisor d of such integers, it is said to be aperiodic if $d = 1$.*

Recall that a chain is positive Harris recurrent and aperiodic if and only if it is ergodic Nummelin, 1984, Proposition 6.3, i.e., there exists a probability measure π , called the stationary distribution, such that, in total variation distance,

$$\lim_{n \rightarrow +\infty} \|\mathbb{P}^n(x, \cdot) - \pi\|_{\text{tv}} = 0.$$

Definition 2.4 (Strong Markov property). *Let $(X_n)_{n \geq 0}$ be a Markov chain and let T be a stopping time of $(X_n)_{n \geq 0}$. Then conditionally on $T < \infty$ and $X_T = i$, $(X_{T+n})_{n \geq 0}$ is a Markov chain and independent of X_0, \dots, X_T .*

Definition 2.5 (small sets). *A set $S \in \mathcal{E}$ is said to be Ψ -small if there exists $\delta > 0$, a positive probability measure Ψ supported by S and an integer $m \in \mathbb{N}^*$ such that*

$$\forall x \in S, B \in \mathcal{E}, \mathbb{P}^m(x, B) \geq \delta \Psi(B). \quad (2.1)$$

Definition 2.6. *Let $(X_n)_{n \geq 0}$ be a Markov Chain taking value in (E, \mathcal{E}) . We say that $(X_n)_{n \geq 0}$ is positive recurrent if*

1. $(X_n)_{n \geq 0}$ is (A, p, ν, m) recurrent (or Harris-recurrent if \mathcal{E} is countably generated), where $A \in \mathcal{E}$ a set, $0 < p < 1$, m an integer and ν a probability measure.
2. $\sup_{x \in \mathbf{A}} \mathbb{E}_x(\mathcal{T}_0) < \infty$, where \mathcal{T}_0 is the hitting time of \mathbf{A} by the m step chain, roughly speaking,

$$\mathcal{T}_0 = \min\{i \geq 1 : X_{i,m} \in \mathbf{A}\}.$$

Definition 2.7. *A ψ -irreducible aperiodic chain X is called regenerative or atomic if there exists a measurable set \mathbf{A} called an atom, in such a way that $\psi(\mathbf{A}) > 0$ and for all $(x, y) \in \mathbf{A}^2$ we have $\mathbb{P}(x, \cdot) = \mathbb{P}(y, \cdot)$. Roughly speaking, an atom is a set on which the transition probabilities are the same. If the chain visits a finite number of states, then any state or any subset of the states is an atom.*

One of the essential properties of Harris-recurrent Markov chains is the existence of invariant distribution, which we will be called μ (a limiting probability distribution, also called occupation measure). Also, Harris-recurrent Markov chains can permanently be embedded in a certain Markov chain on an extended sample space possessing a recurrent atom. The fact of the existence of recurrent atom \mathbf{A} gives an immediate consequence for the construction of a regenerative extension of this chain. The time the chain hits a

given atom (recurrent state) is considered the regenerative time. Athreya et al., 1978 and Nummelin, 1978 give the construction of such a regenerative extension. This technique is discussed in Section 5.2. The development of a regenerative extension makes the use of regenerative technique possible to study this type of Markov chain. As mentioned above, we will assume in this work that the Harris-recurrent chain is atomic, i.e., the set which is infinitely almost sure recurrent is well-defined and accessible. This set \mathbf{A} is called an atom. By definition, an atom \mathbf{A} is a measurable set, in \mathcal{E} , such that $\mu(\mathbf{A}) > 0$, and for all $x, y \in \mathbf{A}$, $\pi(x, \cdot) = \pi(y, \cdot)$. We will indicate by $\mathbb{P}_{\mathbf{A}}$ (respectively by $\mathbb{E}_{\mathbf{A}}$) the probability measure on the underlying space such that $x \in \mathbf{A}$ (respectively the $\mathbb{P}_{\mathbf{A}}$ -expectation).

The imposed conditions on the Markov chain ensure that the defined atom \mathbf{A} (or the constructed one in the case of a non-atomic chain) is one recurrent class, and let us define the following terms.

Hitting times: Define $\mathcal{T}_j : E \rightarrow \mathbb{N} \cup \{\infty\}$ by

$$\begin{aligned}\mathcal{T}_0 &:= \inf\{n \geq 0 : X_n \in \mathbf{A}\}, \\ \mathcal{T}_j &:= \inf\{n \geq \mathcal{T}_{j-1} : X_n \in \mathbf{A}\}.\end{aligned}\tag{2.2}$$

A well-known property of the hitting time is that for all $k \in \mathbb{N}$, $\mathcal{T}_j < \infty$, $\mathbb{P}_\nu - a.s$ Chung, 1967, chap. I14.

Renewal times: Using the Hitting times, we can define the renewal times as

$$\begin{aligned}\tau_0 &:= \mathcal{T}_0 + 1, \\ \tau(j) &:= \mathcal{T}_j - \mathcal{T}_{j-1}.\end{aligned}\tag{2.3}$$

Similar to the regenerative process, the sequence of renewal times $\{\tau(j)\}_{j=1}^\infty$ is i.i.d and independent of the choice of the initial probability. Throughout this work, we set $\tau = \tau(1)$. We will denote throughout the chapter $\alpha = \mathbb{E}_{\mathbf{A}}(\tau)$.

Regenerative Blocks: Let $l_n := \max\{j : \sum_{i=0}^j \tau(i) \leq n\}$ be the number of visits to the atom \mathbf{A} . Using the Strong property of the Markov chain and under the Markov chain assumption in 2.1, it is possible to divide the given sample (X_1, \dots, X_n) into a sequence of blocks $\{B_j\}_{j=0}^{l_n}$ such that, for all $j = 1, \dots, l_n$ = number total of blocks:

$$\begin{aligned}\mathbf{B}_0 &= \{X_1, \dots, X_{\mathcal{T}_0}\} \\ \mathbf{B}_j &= \{X_{\mathcal{T}_{j-1}+1}, \dots, X_{\mathcal{T}_j}\} \quad \text{in } \mathbb{T} = \bigcup_{n=1}^{\infty} E^n,\end{aligned}\tag{2.4}$$

$$\mathbf{B}_{l_n}^{(n)} = \{X_{\mathcal{T}_{l_n-1}+1}, \dots, X_n\}$$

The length of each block will be denoted by $l(B_j) := \mathcal{T}_j - \mathcal{T}_{j-1}$.

2.2 U -process

Let $(X_n)_{n \in \mathbb{N}}$ a sequence of random variables with values in a measurable space (E, \mathcal{E}) , and let $h : E^m \rightarrow \mathbb{R}$ a measurable function symmetric in its arguments, a U -statistic of order (or degree) m and kernel $h(\cdot)$ is defined as:

$$U_n(h) = \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_n^m} h(X_{i_1}, \dots, X_{i_m}), \text{ for } n \geq m, \quad (2.5)$$

where

$$I_n^m = \{(i_1, \dots, i_m) : i_j \in \mathbb{N}, 0 \leq i_j \leq n, i_j \neq i_k \text{ if } j \neq k\}.$$

Accordingly, a U -process is the collection $\{U_n(h) : h \in \mathcal{F}\}$, with \mathcal{F} is the class of kernels $h(\cdot)$ of m variables. Decoupling inequality of U -statistics and U -process plays a significant role in the latest developments of its asymptotic theory. As a result, decoupling inequality can give a relation between the quantities

$$\mathbb{E}\Phi\left(\left|\sum_{I_n^m} h(X_{i_1}, \dots, X_{i_m})\right|\right) \text{ and } \mathbb{E}\Phi\left(\left|\sum_{I_n^m} h(X_{i_1}^1, \dots, X_{i_m}^m)\right|\right),$$

where $\Phi(\cdot)$ is a non-negative function, and $\{X_i^k\}, k = 1, \dots, m$ are independent copies of the original sequence $\{X_i\}$. One of the good reasons for decoupling is randomization, which is frequently used in the study of the asymptotic theory of U -statistics, and studied by de la Peña et al., 1999, and used in the work of Arcones et al., 1993. The main idea of randomization is to compare the tail probabilities or moments of the original U -statistic or process, $\sum_{I_n^m} h(X_{i_1}, \dots, X_{i_m})$, with the tail probabilities or moments of the statistic

$$\sum_{I_n^m} \varepsilon_{i_1} \dots \varepsilon_{i_r} h(X_{i_1}, \dots, X_{i_m}),$$

where ε_i are independent Rademacher variables independent from X_i , and $1 \leq r \leq m$ depend on the degree of degeneracy (centering) of the kernel $h(\cdot)$.

Definition 2.8. (de la Peña et al., 1999) A symmetric \mathbb{P}^m -integrable kernel $h : E^m \rightarrow \mathbb{R}$ is \mathbb{P} -degenerate of order $r - 1$ if and only if

$$\int h(x_1, \dots, x_m) d\mathbb{P}^{m-r+1}(x_r, \dots, x_m) = \int h d\mathbb{P}^m$$

holds for any $x_1, \dots, x_{r-1} \in E$, whereas

$$\int h(x_1, \dots, x_m) d\mathbb{P}^{m-r}(x_{r+1}, \dots, x_m),$$

is not a constant function.

Besides, $h(\cdot)$ is said to be canonical or completely degenerated if the integral with respect to one variable is equal to zero, i.e.,

$$\int h(x_1, \dots, x_m) d\mathbb{P}(x_1) = 0 \quad \text{for all } x_2, \dots, x_m \in E.$$

The fact that the kernel is completely degenerate with the condition $\mathbb{P}^m h^2 < \infty$, is used for the orthogonality of the different elements of the Hoeffding decomposition of U -statistics. This decomposition is examined in Section 3.

Definition 2.9 (V-statistics). Let $\{X_n\}_{n \in \mathbb{N}}$ a sequence of random variables with values in a measurable space (E, \mathcal{E}) , and let $h : E^m \rightarrow \mathbb{R}$ a measurable function symmetric in its arguments, a V-statistics of degree m and kernel $h(\cdot)$ is defined as:

$$V_n(h) = n^{-m} \sum_{i_1=0}^n \dots \sum_{i_m=0}^n h(X_{i_1}, \dots, X_{i_m}), \quad n \geq m. \quad (2.6)$$

We can associate some distances $e_{n,p}$ to the covering numbers, where

$$e_{n,p} = (U_n(|f - g|^p))^{1/p}.$$

In this work, we use the two distances defined afterward

$$e_{n,2}(f, g) = \left(\frac{(n-m)!}{n!} \sum_{0 \leq i_1 < \dots < i_m \leq n} (f - g)(X_{i_1}, \dots, X_{i_m})^2 \right)^{1/2}.$$

For decoupled statistics, we also associated covering numbers, well-known as $\tilde{N}(\varepsilon, \mathcal{F}, \tilde{e}_{n,p})$ and a distance, which can be defined for $p = 2$ as follows:

$$\tilde{e}_{n,2}(f, g) = n^{1/2} \frac{(n-m)!}{n!} \left[\mathbb{E}_\varepsilon \left(\sum_{0 \leq i_1 < \dots < i_m \leq n} \varepsilon_{i_1} (f - g)(X_{i_1}, \dots, X_{i_m}) \right)^2 \right]^{1/2}.$$

Definition 2.10 (Permissible classes of function). *Let (Ω, Σ, P) be a complete probability space, (E, \mathcal{E}) a measurable space (\mathcal{E} a Borel σ -field on E). \mathcal{F} is a class of function indexed by a parameter x that belongs to a set E . \mathcal{F} is called permissible if it can be indexed by a E such that:*

- *There exists a function $g(x, f) = f(x)$ defined from $S \times \mathcal{F}$ to \mathbb{R} in such a way that this function is $\mathcal{L} \otimes \mathcal{B}(\mathcal{F})$ measurable function, where $\mathcal{B}(\mathcal{F})$ is the Borel σ -algebra generated by the metric on \mathcal{F} .*
- *E is a Suslin measurable space that is mean E is an analytic subset of a compact metric space E from which it inherits its metric and Borel σ -field.*

Definition 2.11. *A class \mathcal{F} of measurable functions $E \rightarrow \mathbb{R}$ is said to be of VC-type (or Vapnik-Chervonenkis type) for an envelope F and admissible characteristic (C, v) (positive constants) such that $C \geq (3\sqrt{e})^v$ and $v \geq 1$, if for all probability measure \mathbb{Q} on (E, \mathcal{E}) with $0 < \|F\|_{L_2(\mathbb{Q})} < \infty$ and every $0 < \varepsilon < 1$,*

$$\mathcal{N}(\varepsilon \|F\|_{L_2(\mathbb{Q})}, \mathcal{F}, \|\cdot\|_{L_2(\mathbb{Q})}) \leq C\varepsilon^{-v}.$$

Assuming that the class is countable to avoid measurability issues (but the non-countable case may be handled similarly using outer probability and additional measurability assumptions, see A. W. van der Vaart et al., 1996).

Definition 2.12 (Stochastic equicontinuity). (Pollard, 1984) *Let $\{Z_n\}$ be a sequence of stochastic processes. Call $\{Z_n\}$ stochastically equicontinuous at t_0 if for each $\delta > 0$ there exists a neighborhood D of t_0 such that*

$$\limsup \mathbb{P} \left\{ \sup_D |Z_n(t) - Z_n(t_0)| \right\} < \varepsilon. \quad (2.7)$$

In the context of U -process $\{U_n\}$, stochastic equicontinuity at a function $g \in \mathcal{F}$ (where this class of functions needs to be permissible) generally implies that $|U_n(h) - U_n(g)|$ should be uniformly small for all $h(\cdot)$ close enough to $g(\cdot)$, with high probability and for all n large enough.

3 Hoeffding decomposition

A significant issue was detected in recovering the estimation of our parameter of interest using the U -process. The given shape of this parameter is as follows:

$$\Theta(h) = \int_{x_1 \in E} \dots \int_{x_k \in E} h(x_1, \dots, x_k) \boldsymbol{\mu}(dx_1) \dots \boldsymbol{\mu}(dx_k),$$

where $h : E^m \rightarrow \mathbb{R}$ is a kernel function. The estimation of this parameter should be possible using the U -statistics of the form:

$$U_n(h) = \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_n^m} h(X_{i_1}, \dots, X_{i_m}), \text{ for } n \geq m, \quad (3.1)$$

As the parameter of interest will be defined and based on Kac's theorem for occupation measure, $\boldsymbol{\mu}(h)$ in the regeneration setup can be written as follows:

$$\Theta(h) = \frac{1}{(\mathbb{E}_A(\boldsymbol{\tau}))^m} \mathbb{E}_A \left(\sum_{i_1=\mathcal{T}_0+1}^{\mathcal{T}_1} \sum_{i_2=\mathcal{T}_1+1}^{\mathcal{T}_2} \dots \sum_{i_m=\mathcal{T}_{(m-1)}+1}^{\mathcal{T}_m} h(X_{i_1}, \dots, X_{i_m}) \right). \quad (3.2)$$

In the Markovian context and since the variables are not independent, the approximation related to the i.i.d. blocks and the regenerative case will be introduced below:

Definition 3.1. (*Regenerative Kernel*) Let $h : E^m \rightarrow \mathbb{R}$ a kernel. We define the regenerative kernel $\boldsymbol{\omega}_h : \mathbb{T}^m \rightarrow \mathbb{R}$ as follows:

$$\boldsymbol{\omega}_h((x_{11}, \dots, x_{1n_1}), \dots, (x_{k1}, \dots, x_{kn_k})) = \sum_{i_1=1}^{n_1} \dots \sum_{i_k=1}^{n_k} h(x_{1i_1}, \dots, x_{ki_k}). \quad (3.3)$$

It is not necessary that the kernel $\boldsymbol{\omega}_h(\cdot)$ to be symmetric, as soon as $h(\cdot)$. In fact, we can use the symmetrization of $S_m \boldsymbol{\omega}_h$ in the following way

$$(S_m \boldsymbol{\omega}_h) = (m!)^{-1} \sum_{\sigma(1)=1}^{n_1} \dots \sum_{\sigma(m)=1}^{n_k} h(x_{\sigma(1)}, \dots, x_{\sigma(m)}), \quad (3.4)$$

where the sum is over all permutations $\sigma = \{i_1, \dots, i_m\}$ of $\{1, \dots, m\}$. Next, we consider the U -statistic formed by the regenerative data.

Definition 3.2. (*Regenerative U -statistic*) Let $h : E^m \rightarrow \mathbb{R}$ a kernel such that $\Theta(|h|) < \infty$ and set $\tilde{h}(\cdot) = h(\cdot) - \Theta(h)$. The regenerative U -statistic associated to the sequence of regenerative blocks $\{B_j\}_{j=1}^L$, for $L = l_n - 1 \geq 1$, generated by the Markov chain is given by

$$R_L(h) = \binom{L}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_L^m} \boldsymbol{\omega}_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}). \quad (3.5)$$

We mention that $R_L(h)$ is a standard U -statistic with zero mean. Hence we can write its Hoeffding decomposition according to the following

$$R_L(h) = S_L(h) + D_L(h), \quad (3.6)$$

with

$$S_L(h) = \frac{m}{L} \sum_{i=1}^L \tilde{h}^{(1)}(\mathbf{B}_i), \quad (3.7)$$

$$D_L(h) = \sum_{j=2}^m \binom{m}{j} \binom{L}{j}^{-1} \sum_{1 \leq i_1 < \dots < i_j \leq L} \tilde{h}^{(j)}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_j}). \quad (3.8)$$

We define the associated functions:

$$\begin{aligned} \tilde{h}^{(1)}(\mathbf{B}_1) &= \mathbb{E}(\omega_{\tilde{h}}(\mathbf{B}_1, \dots, \mathbf{B}_m \mid \mathbf{B}_1)) = \omega_{\tilde{h}}^{(1)}(\mathbf{B}_1), \\ \tilde{h}^{(c)}(\mathbf{B}_1, \dots, \mathbf{B}_c) &= \mathbb{E}(\omega_{\tilde{h}}(\mathbf{B}_1, \dots, \mathbf{B}_m \mid \mathbf{B}_1, \dots, \mathbf{B}_c)) \\ &= \omega_{\tilde{h}}^{(c)}(\mathbf{B}_1, \dots, \mathbf{B}_c) - \sum_{i_1=1}^m \tilde{h}^{(1)}(\mathbf{B}_{i_1}) - \sum_{1 \leq i_1 < i_2 \leq m} \tilde{h}^{(2)}(\mathbf{B}_{i_1}, \mathbf{B}_{i_2}) \\ &\quad - \dots - \sum_{1 \leq i_1 < \dots < i_{c-1} \leq m} \tilde{h}^{(c-1)}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_{c-1}}), \end{aligned}$$

where $\tilde{h}^{(c)}(\cdot)$ represents the conditional expectation of $\omega_{\tilde{h}}(\cdot)$ given c of the coordinates, for all $\mathbf{B}_c \in \mathbb{T}$. The U -statistics $D_L(h)$, obtained by truncating the Hoeffding decomposition after the first term, is degenerate, which means that its variance converges to a constant of order $1/L^2$ (see Bose et al., 2018 Theorem 1.1). The leading term is $S_L(h)$ that converges to the normal distribution with mean zero and variance equal to $\mathbb{E}(\tilde{h}^{(1)^2}(\mathbf{B}_1))$. Then, only the first term of the Hoeffding decomposition of $R_L(h)$ asserts their limit.

3.1 Technical Assumptions:

For our results, we need the following assumptions.

(C.1) (Block-length assumption) For all $q \geq 1$, and $l \geq 1$,

$$\mathbb{E}_{\nu}[\tau^l] < \infty, \quad \mathbb{E}_A[\tau^q] < \infty;$$

(C.2) (Non-regenerative blocks) For $l \geq 1$, we have

$$\mathbb{E}_{\nu} \left[\left(\sum_{i_1=1}^{\mathcal{T}_0} \sum_{i_2=\mathcal{T}_0+1}^{\mathcal{T}_1} \sum_{i_3=\mathcal{T}_1+1}^{\mathcal{T}_2} \dots \sum_{i_m=\mathcal{T}_{m-1}+1}^{\mathcal{T}_m} |h(X_{i_1}, \dots, X_{i_m})| \right)^l \right] < \infty,$$

also

$$\mathbb{E}_{\nu} \left[\left(\sum_{i_1=\mathcal{T}_0+1}^{\mathcal{T}_1} \sum_{i_2=\mathcal{T}_1+1}^{\mathcal{T}_2} \dots \sum_{i_{m-1}=\mathcal{T}_{m-1}+1}^{\mathcal{T}_m} \sum_{i_m=\mathcal{T}(l_n)+1}^n |h(X_{i_1}, \dots, X_{i_{m-1}}, X_{i_m})| \right)^l \right] < \infty$$

(C.3) (Block-sum : Moment assumptions) For $l \geq 1$, we have

$$\mathbb{E}_\nu \left[\left(\sum_{i_1=\mathcal{T}_0+1}^{\mathcal{T}_1} \sum_{i_2=\mathcal{T}_1+1}^{\mathcal{T}_2} \dots \sum_{i_m=\mathcal{T}_{m-1}+1}^{\mathcal{T}_m} |h(X_{i_1}, \dots, X_{i_m})| \right)^l \right] < \infty,$$

and

$$\mathbb{E}_A \left[\left(\sum_{\mathcal{T}_0+1 \leq i_1 \leq \dots \leq i_m \leq \mathcal{T}_1} h(X_{i_1}, \dots, X_{i_m}) \right)^l \right] < \infty;$$

(C.4) For $l \geq 1$, we have

$$\mathbb{E}_\nu \left[\left(\sum_{i_1=\mathcal{T}_0+1}^{\mathcal{T}_1} \sum_{i_2=\mathcal{T}_1+1}^{\mathcal{T}_2} \underbrace{\sum_{i_k=\mathcal{T}_k+1}^{\mathcal{T}_{k+1}} \dots \sum_{i_k=\mathcal{T}_k+1}^{\mathcal{T}_{k+1}}}_{u \text{ times}} \sum_{i_{k+u}=\mathcal{T}_{k+u}+1}^{\mathcal{T}_{k+u+1}} \dots \sum_{i_m=\mathcal{T}_{m-1}+1}^{\mathcal{T}_m} |h(X_{i_1}, \underbrace{X_{i_k}, \dots, X_{i_k}}_{u \text{ times}}, X_{i_{k+u}}, \dots, X_{i_m})| \right)^l \right] < \infty;$$

(C.5) (Non-degeneracy.) We suppose also that

$$\mathbb{E}_A \left[\left(\sum_{i=\mathcal{T}_0+1}^{\mathcal{T}_1} h_1(X_i) \right)^2 \right] > 0.$$

Remark 3.3. [Moment assumptions] In practice, we recall that block-moment assumptions for the split Markov chain can be generally checked by establishing drift conditions of Lyapounov's type for the original chain, see Chapter 11 in S. Meyn et al., 2009b and Douc et al., 2008, as well as the references therein. All these moment conditions are discussed in detail in the book of S. P. Meyn et al., 1993, Chapters 11 & 17. There is a key condition in the proof of ergodic theorems in the Markovian context, which is the fact that $\mathbb{E}_{\mathbf{A}}(\tau_0) < \infty$, for all \mathbf{A} a set in \mathcal{E} , such that $\psi(\mathbf{A}) > 0$. In fact, when there is a finite invariant measure and an atom \mathbf{A} , then this condition is right-founded. We also refer to Bertail et al., 2006a for an explicit check of such conditions on several important examples and to Section 4.1.2 of Bertail et al., 2011b for sufficient conditions expressed in terms of uniform return rate to small sets. Finally, as discussed in Chapter 8 of Revuz, 1984, similar conditions can be expressed in potential kernels. Observe that, in the positive recurrent case, the assumptions of (C.1) are not independent when $\nu = \mu$: from basic renewal theory, one has $\mathbb{P}_\mu(\tau = k) = (\mathbb{E}_A[\tau])^{-1} \mathbb{P}_A(\tau \geq k)$ for all $k \geq 1$. Hence, conditions $\mathbb{E}_\mu[\tau^l] < \infty$ and $\mathbb{E}_A[\tau^{l+1}] < \infty$ are equivalent.

3.2 General case

Let $(X_n)_{n \geq 0}$ be a sequence of random variables. We can write the projection mentioned above and define the degenerate U -statistic in terms of integral. More precisely, we have

$$\Theta := \mathbb{P}^m h = \int \dots \int h(x_1, \dots, x_m) d\boldsymbol{\mu}(x_1, \dots, x_m), \quad (3.9)$$

with

$$h^{(1)}(x) := \mathbb{P}^{m-1} h = \int \dots \int h(x, x_2, \dots, x_m) d\boldsymbol{\mu}(x_2, \dots, x_m). \quad (3.10)$$

It is known that if

$$\mathbb{E}|h(X_1, \dots, X_n)| < \infty,$$

the the U -statistic converges almost sure and in L_1 to $\mathbb{E}h(X_1, \dots, X_n)$, with $h(\cdot)$ is a real measurable symmetric function. To ensure that $\mathbb{E}|h(X_1, \dots, X_m)| < \infty$, we must have $h > 0$ and

$$\left\{ \frac{1}{n^m} \sum h(X_1, \dots, X_m) \right\}_{n=1}^{\infty}$$

is bounded. We can also define the Hoeffding projection of a kernel $h : E^m \rightarrow \mathbb{R}$ symmetric as:

$$\boldsymbol{\pi}_k h(X_1, \dots, X_k) := \boldsymbol{\pi}_{k,m}^{\mathbb{P}} h(X_1, \dots, X_k) := (\delta_{x_1} - \mathbb{P}) \dots (\delta_{x_k} - \mathbb{P}) \mathbb{P}^{m-k} h.$$

For $x_i \in E$ and $0 \leq k \leq m$, such that $\boldsymbol{\pi}_0 h = \mathbb{P}^m h$ bring a decomposition of the U -statistic into a sum of U -statistics of orders $k \leq m$, and $\boldsymbol{\pi}_k(h)$ is a degenerate function of k -variables, for $k > 0$, the Hoeffding decomposition for any U -statistics $U_n(h)$ is:

$$U_n(h) = \sum_{k=0}^m \binom{m}{k} U_n^k(\boldsymbol{\pi}_k(h)).$$

The kernel $h(\cdot)$ is degenerated of order $r - 1$ if and only if the Hoeffding expansion except the constant term start at term r , i.e.,

$$U_n(h) - \mathbb{P}^m(h) = \sum_{k=r}^m \binom{m}{k} U_n^k(\boldsymbol{\pi}_k(h)).$$

We are primarily interested in the investigation of the behavior of

$$\|U_n(h) - \mathbb{P}^m(h)\|_{\mathcal{F}} = \sup_{h \in \mathcal{F}} |U_n(h) - \mathbb{P}^m h|,$$

with \mathcal{F} is an uncountable family of symmetric functions $h : E^m \rightarrow \mathbb{R}$. The first term in this decomposition is a linear function written as follows:

$$S_n(h) = \frac{m}{n} \sum_{i=1}^n (\mathbb{P}^{m-1}h(X_i) - \mathbb{P}^m h) = \frac{m}{n} \sum_{i=1}^n \boldsymbol{\pi}_1 h(X_i).$$

We have in terms of regenerative blocks:

$$S_L(h) = \frac{m}{L} \sum_{k=1}^L \tilde{h}_1(\mathbf{B}_k) := \frac{m}{L} \sum_{k=1}^L (\mathbb{P}^{m-1}h(\mathbf{B}_k) - \mathbb{P}^m h).$$

3.3 Regenerative approximation

The following proposition plays an instrumental role in the present chapter. Actually, the U -statistics of order $m = 2$ tends to be more widely used in literature, and this proposition can be seen in the work of Bertail et al., 2011a. However, when $m \geq 3$, our proofs shall contain additional terms which do not exist when $m = 2$. Such situations require delicate treatment and place additional conditions.

Proposition 3.4. *Suppose that conditions (C.1), (C.2), (C.3), and (C.4) are fulfilled. Let*

$$W_n(h) = U_n(h) - \boldsymbol{\Theta}(h) - \binom{l_n - 1}{m} \binom{n}{m}^{-1} R_L(h). \quad (3.11)$$

Then, we have the following stochastic convergences:

$$W_n(h) \rightarrow 0, \quad \mathbb{P} - a.s.$$

4 Asymptotic theory for Markovian U -statistics

We point out that the U -statistic $U_n^m(h)$, given in (3.1), can be approximated by the regenerative U -statistic $R_L(h)$ (3.5) based on the observed regenerative blocks. The following theorems provide the behavior of Markovian U -statistics over a uniformly bounded class of functions that satisfy some entropy conditions. We begin the study of limit theorems by proving the strong law of large numbers in the following theorem.

Theorem 4.1: Strong Law of Large Numbers

Let $(X_n)_n$ be a Markov chain satisfying the Assumption 2.1. Under the assumptions (C.1), (C.2), (C.3) and (C.4) mentioned in Section 3 and if $\boldsymbol{\Theta}(|h|) < \infty$, we have, as $n \rightarrow \infty$:

$$U_n(h) \rightarrow \boldsymbol{\Theta}(h), \quad \mathbb{P}_\nu \text{ a.s.}$$

Theorem 4.2 represents a fundamental result in this chapter. A result on the CLT for the U -statistics will be presented in this theorem. Assume that the class of measurable function \mathcal{F} satisfies the CLT, or the U -statistic satisfies the CLT, if the Gaussian process G_{Θ} indexed by \mathcal{F} is sample continuous on \mathcal{F} with covariance

$$\mathbb{E}[G_{\Theta}(f) - G_{\Theta}(g)] = \Theta(fg) - (\Theta f)(\Theta g); \quad f, g \in \mathcal{F},$$

and

$$n^{1/2}(U_n - P^m)(h) \xrightarrow{\mathcal{L}} mG_{\Theta} \circ \mathbb{P}^{m-1} \circ h; \quad \text{in } l^{\infty}(\mathcal{F}).$$

The convergence of $n^{1/2}(U_n - P^m)$ is in the sense Hoffmann-Jørgensen, which is defined as, for all $F : l^{\infty}(\mathcal{F}) \rightarrow \mathbb{R}$ bounded and continuous,

$$\mathbb{E}F\left(n^{1/2}(U_n - P^m)(h)\right)^{\circ} \rightarrow \mathbb{E}F(mG_{\Theta} \circ \mathbb{P}^{m-1} \circ h),$$

where $F\left(n^{1/2}(U_n - P^m)\right)^{\circ}$ is the measurable envelope of $F\left(n^{1/2}(U_n - P^m)\right)$. Note that $h(\cdot)$ must be symmetric for the convergence to the Gaussian process; if not, we must use the symmetrization form of $h(\cdot)$, $S_m(h)$, in the limit $mG_{\Theta} \circ \mathbb{P}^{m-1} \circ S_m(h)$ instead of $h(\cdot)$.

Theorem 4.2: Weak convergence

Let $(X_n)_n$ be an Harris-recurrent Markov chain with an atom \mathbf{A} satisfying Assumption 2.1 and such that, for a fixed $\gamma > 0$,

$$\sup_{x \in \mathbf{A}} \mathbb{E}(\tau)^{2+\gamma} < \infty \quad (4.1)$$

and satisfying also the technical assumptions (C.1), (C.2), (C.3), (C.4), and (C.5). Let \mathcal{F} a class of measurable functions uniformly bounded on E and

$$\int_0^{\infty} \sqrt{\log N(\varepsilon, \mathcal{F}, e_{n,2})} d\varepsilon < \infty,$$

^a with expectation zero, i.e.,

$$\mathbb{E} \int_0^{\infty} n^{-1/2} \log N(\varepsilon, \mathcal{F}, e_{n,2}) d\varepsilon \rightarrow 0.$$

Then the U -statistic converges weakly in $l^{\infty}(\mathcal{F})$ to a Gaussian process G indexed by \mathcal{F} whose sample paths are bounded and uniformly continuous with respect to the metric $L_2(\mathbb{P})$.

^auniform entropy integral

Remark 4.1. *Note that condition 4.1 is related to the validity of strong mixing conditions for dependent data and the renewal times of stationary sequences. This question is discussed in Bolthausen, 1982 and Bolthausen, 1980. A brief and clear explanation of this relation is also used by Radulović, 2004.*

In addition, a rate of convergence for the SLLN can be given by the law of iterated logarithm by proving that the distance between $U_n(h)$ and the parameter to estimate $\Theta(h)$ is of order $\sqrt{\log \log n}$. It is noteworthy that the law of iterated logarithm has been proved by Dehling, 1989 and Dehling et al., 1986.

Theorem 4.3: Law of Iterated logarithm

Let $(X_n)_n$ be a Markov chain satisfying the Assumption 2.1. Suppose that the Assumptions (C.1), (C.2), (C.3), (C.4), and (C.5) are fulfilled. Then we have

$$\limsup_{n \rightarrow \infty} \frac{\sqrt{n}(U_n(h) - \Theta(h))}{\sqrt{2\sigma^2 \log \log n}} = 1, \quad \mathbb{P}_\nu - a.s.$$

5 Uniform bootstrap central limit theorem

We now establish the bootstrap central limit theorem over a uniformly bounded class of functions that satisfy the uniform entropy condition. Starting by describing the resampling algorithm and establishing our bootstrap version of the uniform central limit theorem for Harris Markov chains.

5.1 RBB method

The Regenerative blocks bootstrap (RBB) method steps are as follows:

- Recall that l_n the number of visits to the atom \mathbf{A} . Divided the observed sample $X^{(n)} = (X_1, \dots, X_n^{(n)})$ into $(l_n + 1)$ regenerative blocks $\mathbf{B}_0, \dots, \mathbf{B}_{l_n-1}, \mathbf{B}_{l_n}^{(n)} \in \mathbb{T}$. Drop the first and the last blocks (to avoid bias) (non-regenerative blocks).
- Draw sequentially bootstrap data blocks $\mathbf{B}_{1,n}^*, \dots, \mathbf{B}_{k,n}^*$ independently from the empirical distribution function

$$\mathbb{P}_n = \frac{1}{l_n - 1} \sum_{i=1}^{l_n-1} \delta_{\mathbf{B}_i},$$

the generated of blocks bootstrap will be until the joint length of bootstrap blocks

$$l^*(k) = \sum_{j=1}^k l(\mathbf{B}_{j,n}^*)$$

exceeds n . Let

$$l_n^* = \inf\{k > 1, l^*(k) > n\} \quad \text{and} \quad L^* = l_n^* - 1$$

- Reconstruct, by binding the blocks together $X^* = (\mathbf{B}_{1,n}^*, \dots, \mathbf{B}_{l_n^*-1,n}^*)$, a new trajectory of size $l^*(l_n^* - 1)$, with $\mathbf{B}_{0,n}^* = \mathbf{B}_0$ and $\mathbf{B}_{l_n,n}^* = \mathbf{B}_{l_n}^{(n)}$.

Denoted by n^* length of bootstrapped sample, which is given by

$$n^* = \sum_{i=1}^{L^*} l(\mathbf{B}_i^*).$$

Remark 5.1. *The fact that some observations may be sampled more than once while others are not sampled at all is one of the potential disadvantages of the original bootstrap developed by Efron, 1979. The weighted (or smooth) bootstrap is an alternative that has been proven to be computationally more efficient in numerous situations. This has been demonstrated both theoretically and empirically. The use of subsampling, a type of resampling method, is another approach that can be taken. This technique is utilized to estimate statistical estimator properties such as the variance, distribution function, and so on. The subsampling method is less effective than the bootstrapping method in general, but it can be useful in many circumstances where other approaches are ineffective. Let's recall some facts about the repeats in the resampled sample from Hall, 1987, Appendix 1, Fisher et al., 1991, pp. 160-161, Hall, 1992, Appendix I. With probability one, all values in the sample $\mathcal{X} = X_1, \dots, X_n$ are distinct. The number of different unordered resamples $\mathcal{X}^* = X_1^*, \dots, X_n^*$ that can be drawn from \mathcal{X} with replacement, equals the number of ways of choosing non-negative integers k_1, \dots, k_n satisfying $k_1 + \dots + k_n = n$. This is given by*

$$N(n) = \binom{2n-1}{n},$$

which is asymptotic to $(n\pi)^{-\frac{1}{2}}2^{2n-1}$ as $n \rightarrow \infty$. The qualification "with probability one" here and below refers to realizations of \mathcal{X} . It means that if \mathcal{C} is the collection of realizations for which the qualified statement is valid, then $\mathbb{P}(\mathcal{X} \in \mathcal{C}) = 1$. Not all atoms have equal probability. The atom with the greatest mass is the one corresponding to $\mathcal{X}^ = \mathcal{X}$, and has probability $p_n = n!/n^n \sim (2n\pi)^{\frac{1}{2}}e^{-n}$. If B bootstrap simulations are conducted, the chance that each results in a different value is not less than*

$$(1 - p_n)(1 - 2p_n) \dots \{1 - (B - 1)p_n\} \geq 1 - \frac{1}{2}B^2p_n,$$

provided $(B - 1)p_n < 1$. Therefore if B increases at a slower rate than $n^{-1/4}e^{\frac{1}{2}n}$, in particular, if $B = O(n^c)$ for some $c > 0$, then the chance that an atom is sampled twice converges to zero.

5.2 Splitting technique

Geared toward conserving the entire power of renewal theory in Markov chain analysis on a general state space, a strategy is provided for "splitting" the state space of a Harris recurrent Markov chain in such a way that an "atom" is introduced into the split state space while preserving the chain's recurrent character. This strategy is called the splitting technique, and it was introduced by Nummelin, 1978. The basic assumption in this technique is that for all $x \in E$ and A a set in \mathcal{E} , there exists $\delta \in \mathcal{E}^+$ with $\psi(h) > 0$ such that the k -step transition probability \mathbb{P}^k is bounded by $h \times \phi(A)$, where ϕ is a probability measure. This assumption can be viewed as the existence of a small set S for the Harris Markov chain treated (see definition 2.5). This Minorization assumption is automatically set up in a countably generated σ -algebra. For the sake of simplicity, most of the literature presents the case with $k = 1$ and then the idea is to build a Markov chain X^* that somehow the original Markov chain X is "embedded" in the chain X^* and has an atom that is visited infinitely often with probability one by extending the sample space \mathcal{E} in a new one \mathcal{E}^* . That is why a new sequence of Bernoulli $Y_n \in \{0, 1\}$ should be defined and the new chain will be $X^* = \{X_n, Y_n\}$ in (E^*, \mathcal{E}^*) with a transition probability \mathbb{P}^* presented above. For all arbitrary $x \in S$ and $A \in \mathcal{E}$:

- If $Y_n = 1$ (with probability δ) then we draw X_{n+1} according to $\Psi(A)$,
- If $Y_n = 0$ (with probability $1-\delta$) then
$$\begin{cases} X_{n+1} = (1 - \delta)^{-1}(\pi(X_n, A) - \delta\Psi(A)). & \delta < 1 \\ X_{n+1} = \Psi(A) & \delta = 1 \end{cases}$$

As a result of this method, the atom aimed to build is $S \times \{1\}$ and Nummelin, 1978 has proved that the constructed Markov chain and the original one have the same marginal distribution and they are both Harris recurrent Markov chains.

In this setting, we note $A_{\mathcal{A}} = S \times \{1\}$ the atom of the *split* chain X^* , and $\tau_{A_{\mathcal{A}}}(i)$ where $i = 1, \dots, \hat{l}_n$ the successive hitting time, where \hat{l}_n is the total number of visits of the split chain to $A_{\mathcal{A}}$ up to time n . The approximated blocks are of the form $(\widehat{\mathcal{B}}_0, \dots, \widehat{\mathcal{B}}_j, \dots, \widehat{\mathcal{B}}_{\hat{l}_n-1}, \widehat{\mathcal{B}}_{\hat{l}_n}^{(n)})$. Then we could apply the RBB to the sample path of X^* as in Bertail et al., 2011a.

5.3 Bootstrap of U -statistic

We can write the bootstrap U -statistics as follows

$$U_n^*(h) = \binom{n^*}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_{n^*}^m} h(X_{i_1}^*, \dots, X_{i_m}^*). \quad (5.1)$$

The bootstrapped regenerative U -statistics can be written as follows

$$R_{L^*}^*(h) = \binom{L^*}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_{L^*}^m} \omega_{\tilde{h}}(\mathbf{B}_{i_1, n}^*, \dots, \mathbf{B}_{i_m, n}^*). \quad (5.2)$$

Let \mathcal{L}^* denotes the conditional law given the sample $\{\mathbf{B}_1, \dots, \mathbf{B}_L\}$ and \mathbb{E}^* denotes conditional expectation given the same sample. In order to prove the weak convergence of $U_n^*(h)$, it will be sufficient to prove the convergence of $R_{L^*}^*(h)$. As $\mathbf{B}_{i, n}^*$ are i.i.d., we can use the same proof of Arcones and Giné, 1994 for bootstrap U -process, Theorem 2.1. therein, we will impose the following conditions to prove the convergence of $R_{L^*}^*(h)$:

(C.6) Let \mathcal{F} denotes a measurable class of symmetric functions $h : E^m \rightarrow \mathbb{R}$. There is a function $\lambda : (0, \infty) \rightarrow [0, \infty)$ with $\int_0^\infty \lambda(\eta) d(\eta) < \infty$ such that for each probability measure Θ with $\Theta F^2 < \infty$

$$\left[\log N_2(\eta(\Theta F^2)^{1/2}, \mathcal{F}, \Theta) \right]^{1/2} \leq \lambda(\eta), \quad \eta > 0, \quad (5.3)$$

(C.7) For each i_1, \dots, i_m ,

$$\mathbb{E} |F(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m})|^{2 \text{card}\{i_1, \dots, i_m\}/m} < \infty. \quad (5.4)$$

The last conditions are essential in the proof of the bootstrap CLT, as in Arcones and Giné, 1994, to ensure the convergence of V -statistics, and represent the law of large numbers for V -statistics.

$$V_m^{l_n} = (l_n - 1)^{-m} \sum_{i_1, \dots, i_m=1}^{l_n-1} \omega_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \rightarrow \mathbb{E} \omega_{\tilde{h}}(\mathbf{B}_1, \dots, \mathbf{B}_m), \quad a.s.$$

For more details, we can see Arcones et al., 1992 and the law of large numbers for U -statistics can be bootstrapped under some integrability's conditions.

Theorem 5.1

Assume that for every possible combination of integers i_1, \dots, i_m , condition (C.7) is fulfilled. Then

$$U_n^*(h) \xrightarrow{\mathbb{P}^*} \mathbb{E} h(X_1, \dots, X_m) \quad a.s. \quad (5.5)$$

The theorem below establishes the weak convergence of bootstrapped U -statistic. We show that if the class of functions satisfies the entropy condition (C6), the CLT for the U -process indexed by this class of functions can be bootstrapped. Recall that $\mathcal{L}(\cdot)$ reflects convergence in the sense of Hoffmann-Jørgensen, 1991.

Theorem 5.2

Assume that the conditions (C.6)-(C.7) are satisfied. Then if $l_n^* \rightarrow \infty$,

$$\begin{aligned} & \lim_{n \rightarrow +\infty} \mathcal{L}^* \left\{ L^{*1/2} (R_{L^*}^*(h) - V_m^{l_n^*}(h)) : h \in \mathcal{F} \right\} \\ & \stackrel{a.s.}{=} \lim_{n \rightarrow +\infty} \mathcal{L} \left\{ L^{1/2} (R_L(h) - \mathbb{P}^m h) : h \in \mathcal{F} \right\}. \end{aligned} \quad (5.6)$$

In the following, using other conditions for the class of functions is a must to set the bootstrap uniform weak convergence.

Theorem 5.3

Let $(X_n)_n$ be a positive recurrent Harris Markov chain, with an accessible atom \mathbf{A} , X_n satisfies the conditions (C.1) and (C.2) (moments assumptions). Let \mathcal{F} be a uniform bounded class of functions with an envelope H square integrable such that:

$$\int_0^\infty (\log N(\varepsilon, \mathcal{F}, e_{n,2}))^{m/2} d\varepsilon < \infty.$$

Then the process

$$(n^*)^{1/2} (U_{n^*}^*(h) - U_n(h)),$$

converges weakly in probability under \mathbb{P}_ν to a Gaussian process G indexed by \mathcal{F} whose sample paths are bounded and uniformly continuous with respect to the metric $L_2(\mathbb{P})$.

6 Generalized U -statistics

In this section, we will generalize the study to the case of many Markov chain trajectories. We consider k Markov chains, each of which is described in a measurable space (E_i, \mathcal{E}_i) , with a transition probability π_i , initial probability ν_i and invariant probability measure μ_i for $1 \leq i \leq k$. Then, the k trajectories of Markov chains will be represented by $X_{n1}^{(1)} = \{X_1^{(1)}, \dots, X_{n_1}^{(1)}\}, \dots, X_{n_k}^{(k)} = \{X_1^{(k)}, \dots, X_{n_k}^{(k)}\}$. Throughout this section, we assume that all these chains are regenerative, each of them with an accessible atom A_i , and denote by $\mathbf{B}_j^{(i)}$ the corresponding regenerative blocks. In the case of a Harris Markov chain, all effects can be preserved using the Nummelin technique, Section 5 § 5.2. Let Θ defined by

$$\begin{aligned} \Theta &= \mathbb{E} h \left(X_1^{(1)}, \dots, X_{m_1}^{(1)}; \dots; X_1^{(k)}, \dots, X_{m_k}^{(k)} \right) \\ &= \int \dots \int h \left(u_1^{(1)}, \dots, u_{m_1}^{(1)}, \dots, u_1^{(k)}, \dots, u_{m_k}^{(k)} \right) \prod_{j=1}^k d\mu_{n_j} \left(u_1^{(j)}, \dots, u_{m_j}^{(j)} \right), \end{aligned} \quad (6.1)$$

where $h(\cdot)$ is assumed, without loss of generality, to be a symmetric kernel defined on $E_1^{m_1} \times \dots \times E_k^{m_k}$ with values in \mathbb{R} . To approximate this parameter, the U -statistic can be written as follows:

$$U_{(n_1, \dots, n_m)}(h) = \left\{ \prod_{j=1}^k \binom{n_j}{m_j}^{-1} \right\} \sum_c h \left(X_{i_{11}}^{(1)}, \dots, X_{i_{1m_1}}^{(1)}; \dots; X_{i_{k1}}^{(k)}, \dots, X_{i_{km_k}}^{(k)} \right). \quad (6.2)$$

Denote by $\{i_{j1}, \dots, i_{jm_j}\}$ a set of m_j distinct elements of the set $\{1, 2, \dots, n_j\}$, where $1 \leq j \leq k$ and \sum_c denotes summation over all combination and $n_j \geq m_j$ whatever $j = 1, \dots, k$.

The study of multi-sample U -statistics began with Lehmann, 1951 and Dwass, 1956. P. K. Sen, 1977 provided a convergence almost sure for generalized U -statistics for integrable kernels. Weak convergence has also been studied by P. K. Sen, 1974b. Zhang et al., 1996 established the asymptotic distribution for bootstrapping generalized U -statistics and bootstrapping generalized V -statistics, and their work represents a good reference in this framework. All these studies require an independent and identical distribution between sequences and between observations of each sequence.

6.1 Decomposition of generalized U -statistics

As the sample U -statistics, we can write a decomposition of the generalized U -statistics as follows

$$U_{(n_1, \dots, n_m)} = \widehat{U} + \widehat{D},$$

in such a way that

$$\widehat{U} - \Theta = \sum_{j=1}^k \sum_{i=1}^{n_j} \frac{m_j}{n_j} h_j^* \left(X_i^{(j)} \right), \quad (6.3)$$

where

$$h_j^*(x) = \mathbb{E} h \left(X_1^{(1)}, \dots, X_{m_1}^{(1)}; \dots; X_1^{(k)}, \dots, X_{m_k}^{(k)} \mid X_1^{(j)} = x \right) - \Theta.$$

We define

$$\begin{aligned} & h^{(c_1, \dots, c_k)}(x_1^{(1)}, \dots, x_{c_1}^{(1)}; \dots; \dots; x_1^{(k)}, \dots, x_{c_k}^{(k)}) \\ &= \int \dots \int h(u_1^{(1)}, \dots, u_{m_1}^{(1)}, \dots, u_1^{(k)}, \dots, u_{m_k}^{(k)}) \\ & \quad \prod_{j=1}^k \left[\prod_{i=1}^{c_j} d \left(\delta_{x_i^{(j)}} \left(u_i^{(j)} \right) - \mu_j \left(u_i^{(j)} \right) \right) \prod_{i=c_j+1}^{m_j} d\mu_j \left(u_i^{(j)} \right) \right]. \end{aligned}$$

Such that $\delta_{x_i^{(j)}}(\cdot)$ represents the d.f. of a single point mass at x , and μ_i is the probability measure of each sequence. Note that sequences must be i.i.d to get this writing. We can

write the principal kernel as follows:

$$\begin{aligned} & h \left(X_{i_{11}}^{(1)}, \dots, X_{i_{1m_1}}^{(1)}; \dots; X_{i_{k1}}^{(k)}, \dots, X_{i_{km_k}}^{(k)} \right) \\ &= \sum_{c_1=0}^{m_1} \dots \sum_{c_k=0}^{m_k} \sum_{(m_1, c_1)} \dots \sum_{(m_k, c_k)} h^{(c_1, \dots, c_k)} \left(x_1^{(1)}, \dots, x_{c_1}^{(1)}; \dots; \dots; x_1^{(k)}, \dots, x_{c_k}^{(k)} \right). \end{aligned}$$

Hence, we can write now the H -decomposition of generalized U -statistics:

$$\begin{aligned} U_n(h) &= \sum_{c_1=0}^{m_1} \dots \sum_{c_k=0}^{m_k} \prod_{j=1}^k \left[\binom{n_j}{c_j}^{-1} \binom{m_j}{c_j} \right] \\ &\quad \sum_{(n_1, c_1)} \dots \sum_{(n_k, c_k)} h^{(c_1, \dots, c_k)} \left(x_1^{(1)}, \dots, x_{c_1}^{(1)}; \dots; \dots; x_1^{(k)}, \dots, x_{c_k}^{(k)} \right). \end{aligned}$$

Note that the function $h^{(c_1, \dots, c_k)}(\cdot)$ satisfy:

$$\mathbb{E} \left\{ h^{(c_1, \dots, c_k)} \left(x_1^{(1)}, \dots, x_{c_1}^{(1)}; \dots; \dots; x_1^{(k)}, \dots, x_{c_k}^{(k)} \right) \right\} = 0.$$

6.2 Regenerative Chains

As mentioned above, we have k - regenerative Markov chains, so by applying the strong property of the Markov chain to each chain that we have, k - sequences of independent identically distributed regenerative blocks can be obtained, $\{\mathbf{B}_0^{(1)}, \dots, \mathbf{B}_{l_{n_1}}^{(1)}\}, \dots, \{\mathbf{B}_0^{(k)}, \dots, \mathbf{B}_{l_{n_k}}^{(k)}\}$, resp. generated from the chains $(X_{n_1}^1)_{n_1 > 0}, \dots, (X_{n_k}^k)_{n_k > 0}$. Note that each chain meets all the properties, conditions, and results studied in the preceding sections of the Markov chain sample. As shown in the previous work, we presume that $\mathbf{B}_0^{(j)}$ and $\mathbf{B}_{l_{n_j}}^{(j)}$ are the non-regenerative blocks (possibly empty), to prevent bias. In the case of non-stationary Markov chains, i.e., the initial distribution is different from the stationary distribution, the bias of the first block is noteworthy. For simplicity's sake, we shall limit the study to the case where $k = 2$ and $m_j = 2$. All results can be extended to the general case.

Definition 6.1. (*Regenerative Kernel*) Let $h(\cdot)$ be a kernel of $m_1 + m_2$ arguments, wish is symmetric in each argument, the regenerative kernel related to $h(\cdot)$ is given by:

$$\begin{aligned} \omega_h \left(\left(x_1^{(1)}, \dots, x_{k_1}^{(1)} \right); \left(x_1^{(1)}; \dots, x_{k_2}^{(1)} \right); \left(x_1^{(2)}, \dots, x_{k_3}^{(2)} \right); \left(x_1^{(2)}, \dots, x_{k_4}^{(2)} \right) \right) \\ = \sum_{i_1=1}^{k_1} \sum_{i_2=1}^{k_2} \sum_{i_3=1}^{k_3} \sum_{i_4=1}^{k_4} h(x_{i_1}, x_{i_2}, x_{i_3}, x_{i_4}). \end{aligned}$$

Definition 6.2. (*Regenerative Generalized U -Statistic*) Let $h : E^{m_1} \times \dots \times E^{m_k} \rightarrow \mathbb{R}$ be a symmetric kernel, such that $\boldsymbol{\mu}(|h|) < \infty$. The regenerative U -statistic can be written as follows:

$$R_{(l_{n_1-1}, \dots, l_{n_m-1})}(h) = \prod_{j=1}^k \binom{l_j-1}{m_j}^{-1} \sum_c \boldsymbol{\omega}_{\tilde{h}}(\mathbf{B}_{i_{11}}^{(1)}, \dots, \mathbf{B}_{i_{1m_1}}^{(1)}, \dots, \mathbf{B}_{i_{k1}}^{(k)}, \dots, \mathbf{B}_{i_{km_k}}^{(k)}). \quad (6.4)$$

Then, for $k = 2$ and $m_j = 2$, (6.4) can be given by

$$\begin{aligned} R_{(l_{n_1-1}, l_{n_2-1})}(h)(h) &= \binom{l_{n_1}-1}{2}^{-1} \binom{l_{n_2}-1}{2}^{-1} \sum_{1 \leq i_{11} < i_{12} \leq l_{n_1}-1} \sum_{1 \leq i_{21} < i_{22} \leq l_{n_2}-1} \\ &\quad \times \boldsymbol{\omega}_{\tilde{h}}(\mathbf{B}_{i_{11}}^{(1)}, \mathbf{B}_{i_{12}}^{(1)}, \mathbf{B}_{i_{21}}^{(2)}, \mathbf{B}_{i_{22}}^{(2)}). \end{aligned} \quad (6.5)$$

Therefore, its Hoeffding decomposition is given by :

$$\begin{aligned} R_{(l_{n_1-1}, l_{n_2-1})}(h) &= \sum_{c_1=0}^{m_1=2} \sum_{c_2=0}^2 \binom{l_{n_1}-1}{c_1}^{-1} \binom{l_{n_2}-1}{c_2}^{-1} \binom{2}{c_1} \binom{2}{c_2} \\ &\quad \times \sum_{(l_{n_1}-1, c_1)} \sum_{(l_{n_2}-1, c_2)} h^{(c_1, c_2)}(\mathbf{B}_{i_{11}}^{(1)}, \mathbf{B}_{i_{12}}^{(1)}, \mathbf{B}_{i_{21}}^{(2)}, \mathbf{B}_{i_{22}}^{(2)}) \\ &= S(h) + D(h). \end{aligned} \quad (6.6)$$

Let

$$W_n(h) = U_{(n_1, n_2)}(h) - \boldsymbol{\mu}(h) - \binom{l_{n_1}-1}{2} \binom{l_{n_2}-1}{2} \binom{n_1}{2}^{-1} \binom{n_2}{2}^{-1} R_{(l_{n_1-1}, l_{n_2-1})}(h). \quad (6.7)$$

Proposition 6.3. Assume that each one of the k -Markov chains satisfies the same conditions of Proposition 3.4. We have $W_n(h)$ converges to 0 as $n \rightarrow \infty$.

6.3 Asymptotic theory for Markovian generalized U -statistics

Let us recall

$$U_{(n_1, n_2)}(h) = \binom{n_1}{2} \binom{n_2}{2} \sum_{1 \leq i_{11} < i_{12} \leq n_1} \sum_{1 \leq i_{21} < i_{22} \leq n_2} h(X_{i_{11}}^{(1)}, X_{i_{12}}^{(1)}; X_{i_{21}}^{(2)}, X_{i_{22}}^{(2)}),$$

and

$$\begin{aligned} &R_{(l_{n_1-1}, l_{n_2-1})}(h) \\ &= \binom{l_{n_1}-1}{2}^{-1} \binom{l_{n_2}-1}{2}^{-1} \sum_{1 \leq i_{11} < i_{12} \leq l_{n_1}-1} \sum_{1 \leq i_{21} < i_{22} \leq l_{n_2}-1} \boldsymbol{\omega}_{\tilde{h}}(\mathbf{B}_{i_{11}}^{(1)}, \mathbf{B}_{i_{12}}^{(1)}, \mathbf{B}_{i_{21}}^{(2)}, \mathbf{B}_{i_{22}}^{(2)}). \end{aligned}$$

Theorem 6.1: Strong Law of Large Numbers

Assume that

$$\mathbb{E}\{|h|(\log^+ |h|)^{k-1}\} < \infty,$$

and under the conditions used in the SLLN for the simple chain in Theorem 4.1, we have, as $n \rightarrow \infty$,

$$U_{(n_1, n_2)}(h) \rightarrow \Theta(h).$$

Theorem 6.2

Let $U_{(n_1, n_2)}(h)$ be a generalized U -statistic based on two Markov chains trajectories $(X_1^{(1)}, \dots, X_{n_1}^{(1)})$ and $(X_1^{(2)}, \dots, X_{n_2}^{(2)})$, and $R_{(l_{n_1-1}, l_{n_2-1})}(h)$ the regenerative U -statistics generated by two samples of blocks independents. Let $\tau_{\mathbf{A}_1}$ and $\tau_{\mathbf{A}_2}$ represent the renewal time for the first and second chains with atom \mathbf{A}_1 and \mathbf{A}_2 respectively. Then

$$n^{1/2}(U_{(n_1, n_2)} - \Theta) \xrightarrow{d} N\left(0, \frac{l}{n} (\mathbb{E}_{\mathbf{A}_1}(\tau_{\mathbf{A}_1})^{-2} (\mathbb{E}_{\mathbf{A}_2}(\tau_{\mathbf{A}_2})^{-2} \left(\frac{m_1^2 \delta_{1,0}^2}{p} + \frac{m_2^2 \delta_{0,1}^2}{1-p}\right))\right),$$

and

$$l^{1/2} R_{(l_{n_1-1}, l_{n_2-1})} \xrightarrow{d} N(0, l(p^{-1} m_1^2 \delta_{1,0}^2 + (1-p)^{-1} m_2^2 \delta_{0,1}^2)),$$

where $l = l_{n_1} + l_{n_2}$, $n = n_1 + n_2$ and $p = l_{n_1}/l \rightarrow p$, $0 < p < 1$. We have also

$$\delta_{(c,d)} = \text{Var } h^{(c,d)}(B_1^{(1)}, B_2^{(1)}; B_1^{(2)}, B_2^{(2)}),$$

such that $\delta_{1,0}, \delta_{0,1} > 0$.

7 Bootstrapping generalized U -statistics

In this section, we apply the bootstrap method to generalized U -statistics and study the asymptotic properties of the bootstrapped U -statistics indexed by a class of functions \mathcal{H} . We can draw bootstrap samples from $\mathbb{P}_{1, l_{n_1}}$ and $\mathbb{P}_{1, l_{n_2}}$ correspondingly, see Section 5, and obtain two mutually independent random sequences $X_{n_1}^* = (\mathbf{B}_{1, n_1}^*, \dots, \mathbf{B}_{l_{n_1}^* - 1, n_1}^*)$ and $X_{n_2}^* = (\mathbf{B}_{1, n_2}^*, \dots, \mathbf{B}_{l_{n_2}^* - 1, n_2}^*)$. Let $n^* = n_1^* + n_2^*$ where n_j^* is the length of bootstrapped sample of each chain

$$n_j^* = \sum_{i_j=1}^{l_{n_j}^*} l(\mathbf{B}_{j, i_j}^*),$$

and $l^* = l_{n_1}^* + l_{n_2}^*$. The generalized bootstrap of the U -statistic and the regenerative bootstrapped generalized U -statistic can be written as follow:

$$U_{(n_1^*, n_2^*)}^*(h) = \binom{n_1^*}{2} \binom{n_2^*}{2} \sum_{1 \leq i_{11} < i_{12} \leq n_1} \sum_{1 \leq i_{21} < i_{22} \leq n_2} h(X_{i_{11}}^*, X_{i_{12}}^*; X_{i_{21}}^*, X_{i_{22}}^*), \quad (7.1)$$

and

$$\begin{aligned} & R_{(l_{n_1}^* - 1, l_{n_2}^* - 1)}^*(h) \\ &= \binom{l_{n_1}^* - 1}{2}^{-1} \binom{l_{n_2}^* - 1}{2}^{-1} \sum_{1 \leq i_{11} < i_{12} \leq l_{n_1}^* - 1} \sum_{1 \leq i_{21} < i_{22} \leq l_{n_2}^* - 1} \omega_{\tilde{h}}(\mathbf{B}_{i_{11}}^*, \mathbf{B}_{i_{12}}^*; \mathbf{B}_{i_{21}}^*, \mathbf{B}_{i_{22}}^*) \end{aligned} \quad (7.2)$$

Let us introduce the following conditions.

(C8) \mathcal{H} is a class of measurable function, and for any $h \in \mathcal{H}$, h is symmetric in its arguments.

(C9) \mathcal{H} is permissible, there exists $H > 0$, $|h| \leq H$ for any $h \in \mathcal{H}$ and

$$\mathbb{E}(H^2(X_{i_{11}}, X_{i_{12}}, X_{i_{21}}, X_{i_{22}})(\log^+ H(X_{i_{11}}, X_{i_{12}}, X_{i_{21}}, X_{i_{22}}))^2) < \infty,$$

for any $1 \leq i_{jk} \leq n_k$, and where $\log^+(\cdot) = \max\{0, \log(\cdot)\}$.

Theorem 7.1

If \mathcal{H} satisfies the conditions (C8) and (C9), in addition if

$$\sum_{j=1}^2 (\boldsymbol{\mu}_j h_{j,1}^2 - (\boldsymbol{\mu}_j h_{j,1})^2) > 0, \quad \text{for any } h \in \mathcal{H};$$

where $h_{j,1} = \mathbb{E}(h(X_{11}, X_{12}; X_{21}, X_{22}) | X_{j1} = x)$, $j = \{1, 2\}$, and let $n_j/n \rightarrow p_j$, $0 < p_j < 1$, then the U -process

$$T_n^* = \sqrt{n} \left(U_{(n_1^*, n_2^*)}^*(h) - U_{(n_1, n_2)}(h) \right)$$

converges to a Gaussian process $G_{\boldsymbol{\mu}}$ over \mathcal{H} , whose simple paths are bounded and informally continuous with respect to the $\mathcal{L}_2(\mathbb{P})$, with mean zero and covariance

$$\text{Cov}(G_{\boldsymbol{\mu}}h, G_{\boldsymbol{\mu}}g) = \sum_{j=1}^2 \frac{m_j^2}{p_j} \left[\boldsymbol{\mu}_j(h_{j,1}g_{j,1}) - (\boldsymbol{\mu}_j h_{j,1})(\boldsymbol{\mu}_j g_{j,1}) \right],$$

for any $h, g \in \mathcal{H}$.

Remark 7.1. *Ciotek, 2016 established a bootstrap uniform functional central limit theorem over uniformly bounded classes of functions for Harris recurrent Markov chains. As in the present chapter, the bootstrap in question is the approximate block bootstrap introduced by Bertail et al., 2006b. The proof of Ciotek, 2016 relies on the bootstrap central limit theorem of Giné et al., 1990. In this chapter, we develop theory and tools for studying bootstraps of U -processes, a natural higher-order generalization of the bootstraps of the empirical processes investigated in Ciotek, 2016. The random size of the resampled blocks, which is more complex in our context than for empirical processes, poses the greatest difficulty in establishing our conclusions. In the present work, we have investigated the case of k Markov chains in their full generality, which has been studied elsewhere in a general framework. Finally, when we investigate U -processes, the nonlinear terms appearing in the Hoeffding decomposition are difficult to control in our setting and do not appear in the study of the empirical process, which is the greatest challenge in generalizing to our setting.*

8 Examples

8.1 Classes of functions

Example 21. *The set \mathcal{F} of all indicator functions $\mathbb{I}_{\{(-\infty, t]\}}$ of cells in \mathbb{R} satisfies*

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

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

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

For more details and discussion on this example, refer to Example 2.5.4 of A. W. van der Vaart et al., 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 $(1/\varepsilon)$, see Theorem 9.19 of Kosorok, 2008.

Example 22. *(Classes of functions that are Lipschitz in a parameter, Section 2.7.4 in A. W. van der Vaart et al., 1996). Let \mathcal{F} be the class of functions $x \mapsto f(t, x)$ that are Lipschitz in the index parameter $t \in T$. Suppose that*

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

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 A. W. van der Vaart et al., 1996 and Lemma

9.18 of Kosorok, 2008, it follows, for any norm $\|\cdot\|_{\mathcal{F}}$ on \mathcal{F} , that

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

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

Example 23. As an example, the classes of functions that are smooth up to order α are defined as follows; see Section 2.7.1 of A. W. van der Vaart et al., 1996 and Section 2 of A. van der Vaart, 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

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

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

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

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

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

Note that for $\alpha \leq 1$ this class consists of bounded functions f that satisfy a Lipschitz condition. Kolmogorov et al., 1959 computed the entropy of the classes of $C_M^\alpha(\mathcal{X})$ for the uniform norm. As a consequence of their results A. van der Vaart, 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 $\varepsilon > 0$,

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

$\mathcal{N}_{[\cdot]}$ is the bracketing number, refer to Definition 2.1.6 of A. W. van der Vaart et al., 1996 and we refer to Theorem 2.7.1 of A. W. van der Vaart et al., 1996 for a variant of the last inequality. By Lemma 9.18 of Kosorok, 2008, we have

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

8.2 U-statistics

Example 24. 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 :

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

B. W. Silverman, 1978 has used this kernel in order to propose the U -process to test uniformity on the circle.

Example 25. *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 *A. J. 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

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

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

Example 26. 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 et al., 1972)*.

Example 27. For :

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

we obtain the variance of Y .

Example 28. Many machine learning models like deep neural networks, support vector machines and K -means use the method of gradient descent to find a minimizer to their

risks. In Cléménçon et al., 2016, an example of the application of generalized U -statistics in this field is provided. A representation of the gradient descent iterations:

$$\theta_{t+1} = \theta_t - \eta_t \nabla_{\theta} L(\theta_t),$$

where η_t is called the learning rate, $\theta \in \Theta \subset \mathbb{R}^d$ be a parameter space and ∇_{θ} the gradient operator with respect to θ . Let $H : \prod_{k=1}^K \mathcal{X}_k^{a_k} \times \Theta \rightarrow \mathbb{R}$ the loss function and $\mathbf{X}_{\{1, \dots, n_k\}}^{(k)} = (X_1^{(k)}, \dots, X_{n_k}^{(k)})$ with $1 \leq k \leq K$ a random variables. The risk minimization problem consists of finding the minimum of the gradient

$$L(\theta) = \mathbb{E} \left[H \left(X_1^{(1)}, \dots, X_{d_1}^{(1)}, \dots, X_1^{(K)}, \dots, X_{d_K}^{(K)}; \theta \right) \right] = \mu(H(\cdot; \theta)).$$

An estimator of this parameter can be written by generalized U -statistics, and then:

$$\theta_{t+1} = \theta_t - \eta_t \nabla_{\theta} \hat{L}_{\mathbf{n}}(\theta_t),$$

where

$$U_n(H) = \nabla_{\theta} \hat{L}_{\mathbf{n}}(\theta) = \frac{1}{\prod_{k=1}^K \binom{n_k}{d_k}} \sum_{I_1} \dots \sum_{I_K} \nabla_{\theta} H \left(\mathbf{X}_{I_1}^{(1)}; \mathbf{X}_{I_2}^{(2)}; \dots; \mathbf{X}_{I_K}^{(K)}; \theta \right).$$

9 Mathematical developments

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

Proof of Proposition 3.4

Let $\mathbf{B}_0 = \{X_1, \dots, X_{\mathcal{I}_0}\}$ and $\mathbf{B}_{l_n}^{(n)} = \{X_{\mathcal{I}_{l_n-1}+1}, \dots, X_n\}$ the possibly empty non-regenerative blocks of observations. Note that, for $l_n \leq 2$, the demonstration can be viewed directly in Bertail et al., 2011a, under the assumptions (C1), (C2) and (C3) where we can see that obtain $\mathbb{P}_{\nu}(l_n \leq 2) = O(n^{-2})$. Otherwise, for $l_n > 2$, we can write $W_n(h)$ as follow:

$$W_n(h) = (I) + (II),$$

where

$$(I) = \frac{1}{\binom{n}{m}} \left[\sum_{1 \leq i_1 < \dots < i_{m-1} \leq l_n - 1} \omega_h(\mathbf{B}_0, \mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_{m-1}}) \right]$$

$$\begin{aligned}
 & + \sum_{1 \leq i_1 < \dots < i_{m-1} \leq l_n - 1} \omega_h^{\sim}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_{m-1}}, \mathbf{B}_{l_n}) \Big], \\
 (II) & = \frac{1}{\binom{n}{m}} \left[\sum_{j=2}^m \sum_{0 \leq k < i_1 < \dots < i_{m-1-j} \leq l_n} \omega_h^{\sim}(\mathbf{B}_k, \dots, \mathbf{B}_k, \mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_{m-1-j}}) \right. \\
 & \quad \left. - \sum_{j=2}^m \sum_{1 \leq k < i_1 < \dots < i_{m-1-j} \leq n} \tilde{h}(X_k, \dots, X_k, X_{i_1}, \dots, X_{i_m}) \right] \\
 & = \frac{1}{\binom{n}{m}} \left[\sum_{(I_m^{l_n-1})^c} \omega_h^{\sim}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) - \sum_{(I_m^n)^c} \tilde{h}(X_{i_1}, \dots, X_{i_m}) \right],
 \end{aligned}$$

where

$$(I_m^s)^c = \{(i_1, \dots, i_m) : i_j \in \mathbb{N}, 1 \leq i_j \leq n; \text{ at least there is } j \text{ and } k \text{ such that } i_j = i_k\},$$

the complement of index set, with cardinal equals to $\binom{s+m-1}{m} - \binom{s}{m} := \overline{\binom{s}{m}}$. To prove the convergence of $W_n(h)$ to zero in probability, we must fulfill the convergence of (I) and (II) to zero in probability.

$$\begin{aligned}
 \mathfrak{A} & = \overline{\binom{n}{m}}^{-1} \sum_{(I_m^{l_n-1})^c} \omega_h^{\sim}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \\
 & = \overline{\binom{n}{m}}^{-1} \sum_{(I_m^{l_n-1})^c} \left\{ \omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) - \sum_{i_1=\mathcal{I}_0+1}^{\mathcal{I}} \sum_{i_2=\mathcal{I}+1}^{\mathcal{I}_2} \dots \sum_{i_m=\mathcal{I}_{m+1}}^{\mathcal{I}_{m+1}} \Theta(h) \right\} \\
 & = \overline{\binom{n}{m}}^{-1} \left\{ \sum_{(I_m^{l_n-1})^c} \omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) - \sum_{(I_m^{l_n-1})^c} (l(\mathbf{B}_{i_k})^k) * (l(\mathbf{B}_{i_{(k+1)}})) * \dots * (l(\mathbf{B}_{i_{(m)}})) \Theta(h) \right\} \\
 & = \overline{\binom{n}{m}}^{-1} \left\{ \frac{\binom{l_n-1}{m}}{\binom{l_n-1}{m}} \sum_{(I_m^{l_n-1})^c} \omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \right. \\
 & \quad \left. - \frac{\binom{l_n-1}{m}}{\binom{l_n-1}{m}} \sum_{(I_m^{l_n-1})^c} (l(\mathbf{B}_{i_k})^k) * (l(\mathbf{B}_{i_{(k+1)}})) * \dots * (l(\mathbf{B}_{i_{(m)}})) \Theta(h) \right\} \\
 & = \overline{\binom{n}{m}}^{-1} \left\{ \frac{\binom{l_n-1}{m}}{\binom{n}{m}} \sum_{(I_m^{l_n-1})^c} \omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \sum_{(I_m^{l_n-1})^c} \omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \right. \\
 & \quad \left. - \frac{1}{\binom{l_n-1}{m}} \sum_{(I_m^{l_n-1})^c} (l(\mathbf{B}_{i_k})^k) * (l(\mathbf{B}_{i_{(k+1)}})) * \dots * (l(\mathbf{B}_{i_{(m)}})) \Theta(h) \right\}
 \end{aligned}$$

$$\begin{aligned}
 &\leq \frac{l_n^m}{n^m} \left\{ \frac{1}{\left[\binom{l_n+m}{m} - \binom{l_n-1}{m} \right] (I_m^{l_n-1})^c} \sum \omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \right. \\
 &\quad \left. - \frac{1}{\left[\binom{l_n+m}{m} - \binom{l_n-1}{m} \right] (I_m^{l_n-1})^c} \sum (l(\mathbf{B}_{i_k})^k) * (l(\mathbf{B}_{i_{(k+1)}})) * \dots * (l(\mathbf{B}_{i_m})) \Theta(h) \right\} \\
 &\xrightarrow{n \rightarrow \infty} \alpha^{-m} [\mathbb{E}(\omega_h(\mathbf{B}_1, \underbrace{\mathbf{B}_k, \dots, \mathbf{B}_k}_{u \text{ times}}, \mathbf{B}_{k+u}, \dots, \mathbf{B}_m)) - \mathbb{E}_A((\tau)^u) (\mathbb{E}_A(\tau))^{m-u} \Theta(h)],
 \end{aligned}$$

where $1 \leq k \leq m$ and $1 \leq u \leq k$. Applying the SLLN for Harris Markov chains to find the convergence of

$$\mathfrak{B} = \overline{\binom{n}{m}}^{-1} \sum_{(I_m^n)^c} \tilde{h}(X_{i_1}, \dots, X_{i_m}),$$

to

$$\int \dots \int h(x_1, \underbrace{x_k, \dots, x_k}_{u \text{ times}}, x_{k+u}, \dots, x_m) d\Theta(x_1) d\Theta^u(x_1) d\Theta(x_{k+u}) \dots d\Theta(x_m) - \Theta(h).$$

Using the conditions, all terms in \mathfrak{A} and \mathfrak{B} are finite, and we can prove the convergence of (II) to zero. Now for (I), applying the SLLN and always by Lemma 10.2 part i) where we can see that

$$\mathbb{P}_\nu \left(\lim_{n \rightarrow +\infty} \frac{l_n}{n} \rightarrow \alpha^{-1} \right) = 1. \quad (9.1)$$

We have

$$\begin{aligned}
 &n^{-2m} \mathbb{E}_\nu \left[\left(\sum_{1 \leq i_1 < \dots < i_{m-1} \leq l_n-1} \omega_{\tilde{h}}(\mathbf{B}_0, \mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_{m-1}}) \right)^2 \right] \\
 &= n^{-2m} \mathbb{E}_\nu \left[\left(\frac{\binom{l_n-1}{m-1}}{\binom{l_n-1}{m-1}} \sum_{1 \leq i_1 < \dots < i_{m-1} \leq l_n} \omega_{\tilde{h}}(\mathbf{B}_0, \mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_{m-1}}) \right)^2 \right] \\
 &\leq \mathbb{E}_\nu \left[\left(\frac{\binom{l_n-1}{m-1}}{n^m} \frac{1}{\binom{l_n-1}{m-1}} \sum_{1 \leq i_1 < \dots < i_{m-1} \leq l_n} \omega_{|\tilde{h}|}(\mathbf{B}_0, \mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_{m-1}}) \right)^2 \right] \\
 &\leq \alpha^{-2m} \mathbb{E}_\nu \left[\left(\omega_{|\tilde{h}|}(\mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_{m-1}) \right)^2 \right] \\
 &\leq \alpha^{-2m} \mathbb{E}_\nu \left[\left(\omega_{|\tilde{h}|}(\mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_{m-1}) - \sum_{i_0=0}^{\mathcal{I}_0} \sum_{i_1=\mathcal{I}_0+1}^{\mathcal{I}_1} \sum_{i_2=\mathcal{I}_1+1}^{\mathcal{I}_2} \dots \sum_{i_{m-1}=\mathcal{I}_{m-1}+1}^{\mathcal{I}_m} \Theta(h) \right)^2 \right] \\
 &\leq \alpha^{-2m} \mathbb{E}_\nu \left[\left(\omega_{|\tilde{h}|}(\mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_{m-1}) - (\tau_0)(\tau) \dots (\tau(m)) \Theta(h) \right)^2 \right] \\
 &\leq 2\alpha^{-2m} \left\{ \mathbb{E}_\nu \left[\left(\omega_{|\tilde{h}|}(\mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_{m-1}) \right)^2 \right] + \left[\mathbb{E}_\nu(\tau_0) \mathbb{E}_\nu[(\tau) \dots (\tau(m))] \Theta(h)^2 \right] \right\}
 \end{aligned}$$

$$\begin{aligned}
 &\leq 2\alpha^{-2m} \left\{ \mathbb{E}_\nu \left[\left(\omega_{|h|}(\mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_{m-1}) \right)^2 \right] + \left[\mathbb{E}_\nu(\tau_0) \mathbb{E}_\nu(l(\mathbf{B}_1)) \dots \mathbb{E}_\nu(l(\mathbf{B}_{m-1})) \Theta(h)^2 \right] \right\} \\
 &\leq 2\alpha^{-2m} \left\{ \mathbb{E}_\nu \left[\left(\omega_{|h|}(\mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_{m-1}) \right)^2 \right] + \left[\mathbb{E}_\nu(\tau_0) \mathbb{E}_A(\tau) \dots \mathbb{E}_A(\tau) \Theta(h)^2 \right] \right\} \\
 &\leq 2\alpha^{-2m} \left\{ \mathbb{E}_\nu \left[\left(\omega_{|h|}(\mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_{m-1}) \right)^2 \right] + \mathbb{E}_\nu(\tau_0) (\mathbb{E}_A(\tau))^{m-1} \Theta(h)^2 \right\} < \infty.
 \end{aligned}$$

We obtain, in turn, that

$$\begin{aligned}
 &n^{-2m} \mathbb{E}_\nu \left[\left(\sum_{1 \leq i_1 < \dots < i_{m-1} \leq l_n} \omega_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_{m-1}}, \mathbf{B}_{l_n}) \right)^2 \right] \\
 &\leq \alpha^{-2m} \mathbb{E}_\nu \left[\left(\omega_{\tilde{h}}(\mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_{m-1}, \mathbf{B}_{l_n}) \right)^2 \right] \\
 &\leq \alpha^{-2m} \mathbb{E}_\nu \left[\left(\omega_{|h|}(\mathbf{B}_1, \dots, \mathbf{B}_{m-1}, \mathbf{B}_{l_n}) - \sum_{i_1=\mathcal{I}_0+1}^{\mathcal{I}} \sum_{i_2=\mathcal{I}_1+1}^{\mathcal{I}_2} \dots \sum_{i_{m-1}=\mathcal{I}_{m-1}+1}^{\mathcal{I}_m} \sum_{i_m=\mathcal{I}(l_n)+1}^n \Theta(|h|) \right)^2 \right] \\
 &\leq 2\alpha^{-2m} \left\{ \mathbb{E}_\nu \left[\left(\omega_{|h|}(\mathbf{B}_1, \dots, \mathbf{B}_{m-1}, \mathbf{B}_{l_n}) \right)^2 \right] + \left[\mathbb{E}_\nu(l(\mathbf{B}_1)) \dots \mathbb{E}_\nu(l(\mathbf{B}_{l_n})) \Theta(|h|)^2 \right] \right\} \\
 &\leq 2\alpha^{-2m} \left\{ \mathbb{E}_\nu \left[\left(\omega_{|h|}(\mathbf{B}_1, \dots, \mathbf{B}_{m-1}, \mathbf{B}_{l_n}) \right)^2 \right] + (\mathbb{E}_A(\tau))^m \Theta(|h|)^2 \right\} < \infty.
 \end{aligned}$$

Hence, (I) also converges to zero a.s under \mathbb{P}_ν as $n \rightarrow \infty$. □

Proof of theorem 4.1

According to Hoeffding, 1948, these results can be found directly from Serfling, 1980 and A. J. Lee, 1990, the SLLN for U -statistics provides the convergence of $R_L(h)$ to 0, by the fact that $R_L(h)$ is centered. Using Proposition 3.4 combined with Lemma 10.2 part i), the result of the theorem follows. □

Proof of Theorem 4.2

In what follows, let L denote the number of Blocks observed. We find that

$$R_L(h) = S_L(h) + D_L(h), \tag{9.2}$$

where $S_L(h)$ and $D_L(h)$ are also well defined. Keep in mind that the blocks \mathbf{B}_i are i.i.d., and the kernel can be non-symmetric. We must now prove that $\|R_L(h) - \mathbb{P}^m h\|$ converges weakly to $G \circ \mathbb{P}^{m-1} \circ h$, $h \in \mathcal{F}$. Then we need to show that:

1.

$L^{1/2}S_L(h)$ converges weakly to G on $l^\infty(\mathcal{F})$,

2.

$$\|L^{-m+1/2}D_L(h)\|_{\mathcal{F}} \rightarrow 0.$$

Let

$$\mathbb{P}_L(h_1) = \frac{1}{L} \sum_{i=1}^L \tilde{h}_1(\mathbf{B}_i),$$

and introduce

$$Z_L(h) = \sqrt{L}(\mathbb{P}_L(h_1) - \mathbb{P}(h_1)) = \frac{1}{\sqrt{L}} \sum_{i=1}^L (\mathbb{P}_L(\mathbb{P}^{m-1}(h)) - \mathbb{P}^m(h)).$$

By using the fact (9.1), we can replace the random variable $L = l_n - 1$ with the deterministic quantity \check{L} and we write

$$Z_{\check{L}}(h) = \frac{1}{\sqrt{\check{L}}} \sum_{i=1}^{\check{L}} (\mathbb{P}_{\check{L}}(\mathbb{P}^{m-1}(h)) - \mathbb{P}^m(h)) + o_{\mathbb{P}},$$

where

$$\check{L} = 1 + \left\lfloor \frac{n}{\mathbb{E}_A(\tau)} \right\rfloor.$$

Clearly, $Z_{\check{L}}$ is an empirical process for which we want to prove its weak convergence to a Gaussian process G . In order to establish weak convergence for an empirical process, it is sufficient and necessary to prove finite-dimensional convergence and stochastic equicontinuity. For the finite multidimensional convergence, we have to prove that $(Z_{\check{L}}(h_{i_1}), \dots, Z_{\check{L}}(h_{i_k}))$ converges weakly to $(G(h_{i_1}), \dots, G(h_{i_k}))$ for every fixed finite collection of functions

$$\{h_{i_1}, \dots, h_{i_k}\} \subset \mathcal{F}.$$

In order to fix this, it is enough to show that for every fixed $a_1, \dots, a_k \in \mathbb{R}$,

$$\sum_{j=1}^k a_j Z_{\check{L}}(h_{i_j}) \rightarrow N(0, \sigma^2), \text{ in distribution,} \quad (9.3)$$

where

$$\sigma^2 = \sum_{j=1}^k a_j^2 \text{Var}(Z_{\check{L}}(h_{i_j})) + \sum_{s \neq r} a_j a_i \text{Cov}(Z_{\check{L}}(h_{i_s}), Z_{\check{L}}(h_{i_r})).$$

By linearity, we have

$$\sum_{j=1}^k a_j Z_{\check{L}}(h_{i_j}) = Z_{\check{L}} \left(\sum_{j=1}^k a_j h_{i_j} \right) = Z_{\check{L}}(f).$$

So the finite multidimensional convergence is proved if and only if

$$Z_{\check{L}}(f) \rightarrow G(f), \quad \text{for } f \in \mathcal{F}.$$

According to S. Meyn et al., 2009a and in the same footsteps of the arguments of Chapter 17, we can prove that

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{\check{L}} \tilde{h}_1(\mathbf{B}_j) \rightarrow N(0, \gamma_{h_1}^2),$$

where, under Condition (C5),

$$\gamma_{h_1}^2 = \alpha \mathbb{E}_A \left(\tilde{h}_1^2(\mathbf{B}_1) \right).$$

We readily infer that we have

$$\sqrt{L} S_L(h) \rightarrow N \left(0, m^2 \mathbb{E}_A \left(\tilde{h}_1^2(\mathbf{B}_1) \right) \right).$$

Due to this demonstration, we can find that $Z_L(f) \rightarrow G(f)$, $f \in \mathcal{F}$, and finally, the finite-dimensional convergence is proved.

Now, to verify the equicontinuity, we need to check that for every $\varepsilon > 0$,

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \mathbb{P} \left(\sup_{d(f,g) \leq \delta} |Z_L(f) - Z_L(g)| > \varepsilon \right) = 0, \quad (9.4)$$

where $d(\cdot, \cdot)$ is a pseudo distance for which the class \mathcal{F} is totally bounded, and f, g belong to \mathcal{F} . According to Levental, 1988, we have

$$\begin{aligned} |Z_L(f - g)| &= \left| \frac{1}{\sqrt{L}} \left[\sum_{k=1}^L (f - g)(\mathbf{B}_k) - \mathbb{P}^m(f - g) \right] \right| \\ &\leq \left| \frac{1}{\sqrt{L}} \sum_{a \leq k \leq b} ((f - g)(\mathbf{B}_k) - \mathbb{P}^m(f - g)) \right| \\ &\quad + \left| \frac{1}{\sqrt{L}} \sum_{1 \leq k \leq \lfloor n/\mathbb{E}(\tau) \rfloor} ((f - g)(\mathbf{B}_k) - \mathbb{P}^m(f - g)) \right|, \end{aligned} \quad (9.5)$$

where

$$a = \min(L, \lfloor n/\mathbb{E}(\tau) \rfloor) \text{ and } b = \max(L, \lfloor n/\mathbb{E}(\tau) \rfloor).$$

For the left-hand part of the last inequality, we have

$$\begin{aligned}
 & \left| \sum_{a \leq k \leq b} ((f - g)(\mathbf{B}_k) - \mathbb{P}^m(f - g)) \right| \\
 & \leq \max_{s \in H} 2 \left| \sum_{\lfloor n/\mathbb{E}(\tau) \rfloor - cn^{1/2} \leq k \leq s} ((f - g)(\mathbf{B}_k) - \mathbb{P}^m(f - g)) \right|, \\
 & \quad (\text{for } H = \{ \text{integer } s : |s - \lfloor n/\mathbb{E}(\tau) \rfloor| \leq cn^{1/2} \}), \\
 & \leq \max_{1 \leq s \leq cn^{1/2}} 2 \left| \sum_{1 \leq k \leq s} ((f - g)(\mathbf{B}_k) - \mathbb{P}^m(f - g)) \right| \\
 & \leq \max_{1 \leq s \leq cn^{1/2}} 2 \left| \sum_{1 \leq k \leq s} (f(\mathbf{B}_k) - \mathbb{P}^m(f)) \right| + \max_{1 \leq s \leq cn^{1/2}} 2 \left| \sum_{1 \leq k \leq s} (g(\mathbf{B}_k) - \mathbb{P}^m(g)) \right| \\
 & \leq 2 \sup_{f \in \mathcal{F}} \max_{1 \leq s \leq cn^{1/2}} 2 \left| \sum_{1 \leq k \leq s} (f(\mathbf{B}_k) - \mathbb{P}^m(f)) \right| \\
 & \leq 2 \sup_{f \in \mathcal{F}} \max_{1 \leq s \leq n} \left\{ 2 \left| \sum_{1 \leq k \leq s} (f(\mathbf{B}_k) - \mathbb{P}^m(f)) \right| \right\}.
 \end{aligned}$$

Divide the last inequality by $L^{1/2}$ and using the convergence result in Lemma 2.11 in Levental, 1988 with Condition (C1), we obtain the desired result. The right-hand part in the inequality will be treated using the Lemma 4.2 in Levental, 1988 providing that $\mathbb{E}_A(\tau)^{2+\alpha} < \infty$, for $\alpha > 0$ and the hypothesis of finite uniform entropy integral.

So as mentioned above, the first term in the Hoeffding decomposition is just the sum of i.i.d random variables, which are asymptotically normal when multiplied by \sqrt{L} . Now, to complete the convergence of the regenerative U -statistic, we must treat the remaining terms of its Hoeffding decomposition. For $\zeta \in \mathcal{F}$, let us introduce

$$\zeta := \omega_{\tilde{h}}(\mathbf{B}_1, \dots, \mathbf{B}_m) - \mathbb{P}^m(h) - \sum_{i=1}^m \pi_1 h(\mathbf{B}_i), \quad \mathbf{B}_i \in \mathbb{T}. \quad (9.6)$$

One can see that ζ is centred, that is

$$\int \zeta(\mathbf{B}_1, \dots, \mathbf{B}_m) dP(\mathbf{B}_1) \dots dP(\mathbf{B}_i) \dots dP(\mathbf{B}_m) = 0. \quad (9.7)$$

By randomization theorem, according to Arcones et al., 1993 (for $r = 2$):

$$\mathbb{E} \left\| \sum_{1 \leq i_1 < \dots < i_m \leq \check{L}} \zeta(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \right\|$$

$$= \mathbb{E} \left\| \sum_{1 \leq i_1 < \dots < i_m \leq \check{L}} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} \zeta(\mathbf{B}_{i_1}^{(1)}, \dots, \mathbf{B}_{i_m}^{(1)}) \right\|.$$

Hence, for \mathfrak{C} a constant:

$$\begin{aligned} & \mathbb{E} \left\| \check{L}^{-1/2} \sum_{1 \leq i_1 < \dots < i_m \leq \check{L}} \zeta(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \right\|_{\mathcal{F}} \\ & \leq \mathbb{E} \left\| \check{L}^{-1/2} \sum_{1 \leq i_1 < \dots < i_m \leq \check{L}} \varepsilon_{i_1} \varepsilon_{i_2} \zeta(\mathbf{B}_{i_1}^{(1)}, \dots, \mathbf{B}_{i_m}^{(1)}) \right\|_{\mathcal{F}} \\ & \leq \mathfrak{C} \mathbb{E} \left\| \check{L}^{-1/2} \sum_{1 \leq i_1 < \dots < i_m \leq \check{L}} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} h(\mathbf{B}_{i_1}^{(1)}, \dots, \mathbf{B}_{i_m}^{(1)}) \right\|_{\mathcal{F}} \\ & \leq \mathfrak{C} \mathbb{E} \left\| \check{L}^{-1/2} \sum_{1 \leq i_1 < \dots < i_m \leq \check{L}} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} h(\mathbf{B}_{i_1}^{(1)}, \dots, \mathbf{B}_{i_m}^{(1)}) \right\|_{\mathcal{F}} \\ & \leq \mathfrak{C} \mathbb{E} \int_0^\infty \check{L}^{-1/2} \log N_{n,2}(\varepsilon, \mathcal{F}) d\varepsilon. \end{aligned}$$

It is sufficient now to use the theorem hypothesis of uniform entropy integral to complete the proof of the theorem. \square

Proof of Theorem 4.3

We can see the detailed proof of the law of iterated algorithm (LIL) in A. J. Lee, 1990 Chapter 3 Theorem 1 of §3.5 in the i.i.d. framework. Briefly, we have:

$$\frac{\sqrt{L} D_L}{\sqrt{L \log \log L}} \rightarrow 0 \quad a.s.,$$

where $Var(D_L) = \Theta(n^{-2})$. Then

$$\limsup_n \frac{\sqrt{L}(R_L(h) - P^m(h))}{\sqrt{2m^2 \mathbb{E}_A(\tilde{h}_1^2(\mathbf{B}_1)) \log \log(L)}} = 1.$$

This, in turn, implies that

$$\limsup_n \frac{\sqrt{n}(U_n(h) - \Theta(h))}{\sqrt{8m^2 \sigma_{h_1}^2 \log \log n}} = 1,$$

as we have $W_n(h) \rightarrow 0$ with mean equal to zero. \square

Proof of Theorem 5.1

We must begin the proof by the bootstrap of the law of large number for regenerative U -statistic $R_{L^*}^*(h)$, the i.i.d case, studied and proved in Arcones et al., 1992. This convergence is based on the law of large numbers of the regenerative V -statistic

$$V_L(h) = L^{-m} \sum_{i_1, \dots, i_m=1}^n \omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}),$$

which converges to $\mathbb{P}^m(h) = \mathbb{E}\omega_h(\mathbf{B}_1, \dots, \mathbf{B}_m)$ almost sure in probability, under the condition

$$\mathbb{E}\omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m})^{2\text{card}\{i_1, \dots, i_m\}/m} < \infty.$$

Hence, the proof of the theorem is found immediately using the Proposition 3.4, which is still true if we replace the U -statistic, and the regenerative one by the bootstrapped U -statistic, and the bootstrapped regenerative U -statistic. \square

Proof of Theorem 5.2

Step 1:

Before we state the proof of this theorem, we will replace the random variable represented by the number of Blocks $L = l_n - 1$ with its deterministic quantity $\mathbb{E}(l_n - 1) = \mathbb{E}(L)$. For

$$\mathcal{Z}_L = L^{1/2} \left[\frac{1}{\binom{L}{m}} \sum_{(i_1, \dots, i_m) \in I_L^m} (\omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) - \mathbb{E}(\omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}))) \right]_{h \in \mathcal{F}},$$

and

$$\widetilde{\mathcal{Z}}_L = \mathbb{E}(L)^{1/2} \left[\frac{1}{\binom{\mathbb{E}(L)}{m}} \sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(L)}^m} (\omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) - \mathbb{E}(\omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}))) \right]_{h \in \mathcal{F}}.$$

The goal is to show that, for

$$\mathcal{F}_\delta = \left\{ f - g, f, g \in \mathcal{F}, \|f - g\|_{L_2(\mathbb{P}^m)} \leq \delta \right\}.$$

$$\mathbb{P} \left[\left\| \widetilde{\mathcal{Z}}_L - \mathcal{Z}_L \right\|_{\mathcal{F}_\delta} > \varepsilon \right] < \varepsilon. \quad (9.8)$$

This result signifies that the stochastic equicontinuity of the process \mathcal{Z}_L implies stochastic equicontinuity of the process $\widetilde{\mathcal{Z}}_L$. Without loss of generality, we will assume that $L \leq$

$\mathbb{E}(L)$, and we will define the following set:

$$\mathfrak{J}(V) := \{(i_1, \dots, i_m) : 1 \leq i_1 < \dots < i_m \leq V : i_j \neq i_k \text{ for } j \neq k, \\ \text{such that } \exists \ell \in \{1, \dots, m\} : L \leq i_\ell \leq \mathbb{E}(L)\}.$$

We have:

$$\begin{aligned} & \mathbb{P} \left[\left\| \widetilde{\mathcal{Z}}_L - \mathcal{Z}_L \right\|_{\mathcal{F}_\delta} > \varepsilon \right] \\ = & \mathbb{P} \left[\left\| \frac{\check{L}^{1/2}}{\binom{\check{L}}{m}} \sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(L)}^m} (\omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) - \mathbb{E}(\omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}))) \right. \right. \\ & \left. \left. - \frac{L^{1/2}}{\binom{L}{m}} \sum_{(i_1, \dots, i_m) \in I_L^m} (\omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) - \mathbb{E}(\omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}))) \right\|_{\mathcal{F}_\delta} > \varepsilon \right] \\ \text{Define } \mathfrak{J} := & \{(i_1, \dots, i_m) : 1 \leq i_1 < \dots < i_m \leq \mathbb{E}(L) : i_j \neq i_k \text{ for } j \neq k, \\ & \text{such that } \exists \ell \in \{1, \dots, m\} : L \leq i_\ell \leq \mathbb{E}(L)\}. \\ \leq & \mathbb{P} \left[\left\| \frac{\check{L}^{1/2}}{\binom{\check{L}}{m}} \sum_{(i_1, \dots, i_m) \in \mathfrak{J}} (\omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) - \mathbb{E}(\omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}))) \right\|_{\mathcal{F}_\delta} > \varepsilon \right] \\ = & \mathbb{P} \left[\left\| \frac{\check{L}^{1/2}}{\binom{\check{L}}{m}} \sum_{(i_1, \dots, i_m) \in \mathfrak{J}} (\omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) - \mathbb{E}(\omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}))) \right\|_{\mathcal{F}_\delta} > \varepsilon \cap |\mathbb{E}(L) - L| \leq k\sqrt{n} \right] \\ & + \mathbb{P}(|\mathbb{E}(L) - L| > k\sqrt{n}), \end{aligned}$$

According to Lemma 10.2 part i), we can see that $|\mathbb{E}(L) - L| = O_{\mathbb{P}}(\sqrt{n})$, i.e., there exists $k > 0$ such that $\mathbb{P}(|\mathbb{E}(L) - L| > k\sqrt{n})$ is bounded by ε for every $\varepsilon > 0$, and the first expression in the last inequality is bounded by

$$\begin{aligned} & \mathbb{P} \left[\max_{M \leq K_\varepsilon \sqrt{n} + \mathbb{E}L} \left\| \frac{\check{L}^{1/2}}{\binom{\check{L}}{m}} \sum_{(i_1, \dots, i_m) \in \mathfrak{J}'} (\omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) - \mathbb{E}(\omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}))) \right\|_{\mathcal{F}_\delta} > \varepsilon \right] \\ \text{where } \mathfrak{J}' := & \{(i_1, \dots, i_m) : 1 \leq i_1 < \dots < i_m \leq \mathbb{E}L, \\ & \exists \ell = 1, \dots, m, \mathbb{E}L < i_\ell \leq M, i_j \neq i_k \text{ for } j \neq k\} \tag{9.9} \\ \leq & C_1 \mathbb{P} \left[\left\| \frac{\check{L}^{1/2}}{\binom{\check{L}}{m}} \sum_{(i_1, \dots, i_m) \in \mathfrak{J}''} (\omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) - \mathbb{E}(\omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}))) \right\|_{\mathcal{F}_\delta} > C_2 \varepsilon \right], \end{aligned}$$

where

$$\mathcal{J}'_m := \{(i_1, \dots, i_m) : 1 \leq i_1 < \dots < i_m \leq \mathbb{E}L, \exists \ell = 1, \dots, m, \\ \mathbb{E}L < i_\ell \leq \mathbb{E}L + K_\varepsilon \sqrt{n}, i_j \neq i_k \text{ for } j \neq k\}.$$

This last equation converges to 0 by stochastic equicontinuity of the regenerative U -statistic. Therefore, this proves Expression (9.8).

Step 2:

We proved in Section 4 the convergence of

$$L^{1/2}(R_L(h) - \mathbb{P}^m h).$$

So we need only to prove bootstrap convergence. For the finite-dimensional convergence, we can refer to de la Peña et al., 1999 to prove the linearity of U -process Section 4.2, and Arcones et al., 1992 (Remark 2.5, 2.10 and Corollary 2.6) to see the convergence of the finite-dimensional distributions of the bootstrap. Hence, we just need to prove the equicontinuity of

$$\{L^{*1/2}(R_{L^*}^*(h, \mathbb{P}_n) - V_L(h, \mathbb{P})) : h \in \mathcal{F}\}.$$

We should comment that decoupling and randomization techniques of de la Peña et al., 1999 and Arcones et al., 1993 cannot be directly applied here since we have a random number of blocks l_n as well as a random number of $l(\mathbf{B}_j)$ from which the bootstrapped length $l(\mathbf{B}_{j,n}^*)$ are re-sampled. However, in Step 1, we were able to prove that we have an equivalent asymptotic equicontinuity if we replace the number of blocks L by the deterministic quantity $\mathbb{E}(L)$, and since the bootstrap quantities L^* and n^* are equivalents to n and L , one can use without hesitation their deterministic forms and the decoupling technique. Hence, it is sufficient to show that, for $h \in \mathcal{F}_\delta$,

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow +\infty} \sup \mathbb{E}^* \left\| \mathbb{E}(L)^{1/2} (R_{\mathbb{E}(L)}^*(h, \mathbb{P}_n) - V_{\mathbb{E}(L)}(h, \mathbb{P})) \right\|_{\mathcal{F}_\delta} = 0, \quad (9.10)$$

where

$$\mathcal{F}_\delta = \{f - g, f, g \in \mathcal{F}, \|f - g\|_{L_2(\mathbb{P}^m)} \leq \delta\}.$$

We have

$$\begin{aligned} (*) &:= \mathbb{E}^* \left\| \mathbb{E}(L)^{1/2} (R_{\mathbb{E}(L)}^*(h, \mathbb{P}_n) - V_{\mathbb{E}(L)}(h, \mathbb{P})) \right\|_{\mathcal{F}_\delta} \\ &\leq \mathbb{E}^* \left\| \mathbb{E}(L)^{1/2} (R_{\mathbb{E}(L)}^*(h, \mathbb{P}_n) - \mathbb{P}^m h + \mathbb{P}^m h - V_{\mathbb{E}(L)}(h, \mathbb{P})) \right\|_{\mathcal{F}_\delta} \\ &\leq \mathbb{E}^* \left\| \mathbb{E}(L)^{1/2} (R_{\mathbb{E}(L)}^*(h, \mathbb{P}_n) - \mathbb{P}^m h) \right\|_{\mathcal{F}_\delta} + \mathbb{E}(L)^{1/2} \left\| \mathbb{P}^m h - V_{\mathbb{E}(L)}(h, \mathbb{P}) \right\|_{\mathcal{F}_\delta}. \end{aligned}$$

Using the fact that the V -statistics converges to $\mathbb{P}^m h$, we have, for some positive constant c ,

$$\begin{aligned} &\leq \mathbb{E}^* \left\| \mathbb{E}(L)^{1/2} (R_{\mathbb{E}(L)}^*(h, \mathbb{P}_n) \right\|_{\mathcal{F}_\delta} \\ &\leq \mathbb{E}^* \left\| \mathbb{E}(L)^{1/2} \binom{\mathbb{E}(L)}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(L)}^m} \omega_{\tilde{h}}(\mathbf{B}_{i_1}^*, \dots, \mathbf{B}_{i_m}^*) \right\|_{\mathcal{F}_\delta} \\ &\leq c \mathbb{E}^* \left\| \mathbb{E}(L)^{1/2} \binom{\mathbb{E}(L)}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(L)}^m} \omega_{\tilde{h}}(\mathbf{B}_{i_1}^{*(1)}, \dots, \mathbf{B}_{i_m}^{*(m)}) \right\|_{\mathcal{F}_\delta}. \end{aligned}$$

By decoupling

$$\leq c \mathbb{E}^* \left\| \mathbb{E}(L)^{1/2} \binom{\mathbb{E}(L)}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(L)}^m} \omega_{\tilde{h}}(\mathbf{B}_{i_1}^{*(1)}, \dots, \mathbf{B}_{i_m}^{*(m)}) \right\|_{\mathcal{F}_\delta}.$$

By symmetrisation, where $\{\varepsilon_i^{(1)}\}$ is a Rademacher sequence independent of $\{\mathbf{B}_i^{*(1)}\}$. We have

$$\leq c \mathbb{E}^* \left\| \mathbb{E}(L)^{1/2} \binom{\mathbb{E}(L)}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(L)}^m} \sum_{1 \leq j_1 \leq m} \varepsilon_{i_{j_1}}^{(j_1)} \omega_{\tilde{h}}(\mathbf{B}_{i_1}^{*(1)}, \dots, \mathbf{B}_{i_m}^{*(m)}) \right\|_{\mathcal{F}_\delta}.$$

If we add to the last inequality, within the norm signs, the term

$$\sum_{1 \leq j_1 \leq m} \varepsilon_{i_{j_1}}^{(j_1)},$$

then the p th moment of the norm does not decrease, see de la Peña et al., 1999 Section 3.5.1.

$$\leq c \mathbb{E}^* \left\| \mathbb{E}(L)^{1/2} \binom{\mathbb{E}(L)}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(L)}^m} (\varepsilon_{i_1}^{(1)} + \dots + \varepsilon_{i_m}^{(m)}) \omega_{\tilde{h}}(\mathbf{B}_{i_1}^{*(1)}, \dots, \mathbf{B}_{i_m}^{*(m)}) \right\|_{\mathcal{F}_\delta}.$$

By symmetrisation, where $\{\varepsilon_i^{(1)}\}$ is a Rademacher sequence independent of $\{\mathbf{B}_i^{*(1)}\}$

$$\leq c \mathbb{E}^* \left\| \mathbb{E}(L)^{1/2} \binom{\mathbb{E}(L)}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(L)}^m} (\varepsilon_{i_1} + \dots + \varepsilon_{i_m}) \omega_{\tilde{h}}(\mathbf{B}_{i_1}^{*(1)}, \dots, \mathbf{B}_{i_m}^{*(m)}) \right\|_{\mathcal{F}_\delta}.$$

By coupling (Theorem 3.1.2 de la Peña et al., 1999) or Proposition 2.1 Arcones and Giné, 1994, we obtain

$$\leq c\mathbb{E}^* \left\| \mathbb{E}(L)^{1/2} \binom{\mathbb{E}(L)}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(L)}^m} \varepsilon_{i_1} \boldsymbol{\omega}_h^{\sim}(\mathbf{B}_{i_1}^*, \dots, \mathbf{B}_{i_m}^*) \right\|_{\mathcal{F}_\delta}.$$

By Proposition 2.6 Arcones and Giné, 1994, we have

$$\begin{aligned} &\leq c\mathbb{E}^* \int_0^{D^*} (\log N(\eta, \mathcal{F}_\delta, \sigma))^{1/2} d\eta \\ &\leq c\mathbb{E}^* \int_0^{D^*} \boldsymbol{\lambda}(\eta/2(\mathbb{P}_n^m F^2)^{1/2}) d\eta \\ &\leq c \left(\mathbb{E}^* \mathbb{P}_n^m F^2 \right)^{1/2} \left(\mathbb{E}^* \left(\int_0^{D^*/2(\mathbb{P}_n^m F^2)^{1/2}} \boldsymbol{\lambda}(\eta) d\eta \right)^2 \right)^{1/2}. \end{aligned}$$

Using the fact that $N(\varepsilon, \mathcal{F}, \sigma) \leq (N(\varepsilon/2, \mathcal{F}, \sigma))^2$ and the first hypothesis. Note that $0 \in \mathcal{F}_\delta$ and \mathbb{E}_ε denotes integration with respect to the Rademacher variables ε_i . The quantity σ is defined by:

$$\sigma(f, g) = \left(\mathbb{E}_\varepsilon \left| \mathbb{E}(L)^{1/2} \binom{\mathbb{E}(L)}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(L)}^m} \varepsilon_{i_1} \boldsymbol{\omega}_{(f-g)}^{\sim}(\mathbf{B}_{i_1}^*, \dots, \mathbf{B}_{i_m}^*) \right|^2 \right)^{1/2}.$$

The diameter of \mathbb{T} , D^* is given by

$$D^* = \sup_{h \in \mathcal{F}_\delta} \mathbb{E}_\varepsilon \left(\left| \mathbb{E}(L)^{1/2} \binom{\mathbb{E}(L)}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(L)}^m} \varepsilon_{i_1} \boldsymbol{\omega}_h^{\sim}(\mathbf{B}_{i_1}^*, \dots, \mathbf{B}_{i_m}^*) \right|^2 \right)^{1/2}.$$

Arcones et al., 1992 proved that

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow +\infty} \sup_{n \rightarrow \infty} D = 0$$

in probability \mathbb{P}^* , and then we have, for all $\varepsilon > 0$,

$$\begin{aligned} &\lim_{\delta \rightarrow 0} \lim_{n \rightarrow +\infty} \sup_{n \rightarrow \infty} \mathbb{E}^* \left(\int_0^{D^*/2(\mathbb{P}_n^m F^2)^{1/2}} \boldsymbol{\lambda}(\eta) d\eta \right)^2 \\ &= \lim_{\delta \rightarrow 0} \lim_{n \rightarrow +\infty} \sup_{n \rightarrow \infty} \mathbb{E}^* \left(\int_0^{D'} (\mathbb{I}_{(D' > \varepsilon)} + \mathbb{I}_{(D' < \varepsilon)}) \boldsymbol{\lambda}(\eta) d\eta \right)^2 \\ &= \lim_{\delta \rightarrow 0} \lim_{n \rightarrow +\infty} \sup_{n \rightarrow \infty} \mathbb{E}^* \left(\int_0^{D'} (\mathbb{I}_{(D' > \varepsilon)} + \mathbb{I}_{(D' < \varepsilon)}) \boldsymbol{\lambda}(\eta) d\eta \right)^2 \end{aligned}$$

$$\begin{aligned}
 &\leq \lim_{\delta \rightarrow 0} \lim_{n \rightarrow +\infty} \sup_{n \rightarrow \infty} \left[\left(\int_0^\infty \lambda(\eta) d\eta \right)^2 \times \mathbb{P}^*(D' > \varepsilon) + \left(\int_0^\varepsilon \lambda(\eta) d\eta \right)^2 \right] \\
 &\quad \text{as } \int_0^\infty \lambda(\eta) d\eta < \infty \\
 &\leq \left(\int_0^\varepsilon \lambda(\eta) d\eta \right)^2 \text{ a.s.}
 \end{aligned}$$

Hence the proof is complete. \square

Proof of Theorem 5.3

As mentioned above, we have $n^* = \sum_{j=1}^{l_n^*-1} l(\mathbf{B}_j^*)$ the size of bootstrapped sample X^* . For the Bootstrapped U -statistics in (5.1), Hoeffding decomposition, as in Section 3, can be written as follow:

$$U_n^*(h) = \Theta + \frac{m}{n^*} \sum_{i=1}^{n^*} h^1(X_i^*) + \sum_{j=2}^m \binom{m}{j} \binom{n^*}{j}^{-1} \sum_{1 \leq i_1 < \dots < i_j \leq n^*} h^{(j)}(X_{i_1}^*, \dots, X_{i_j}^*), \quad (9.11)$$

with

$$\begin{aligned}
 h^1(x) &= \mathbb{E}(h(X_1^*, \dots, X_m^* | X_1^*)) \\
 h^{(c)}(x_1, \dots, x_c) &= \mathbb{E}(h(X_1^*, \dots, X_m^* | X_1^*, \dots, X_c^*)).
 \end{aligned}$$

In order to prove the consistency of $U_n^*(h)$, usually, it is enough to prove the convergence of the second summation to zero. Hence we can also write the Hoeffding decomposition of regenerative U -statistic block bootstrap in 5.2 and treat its convergence.

$$\begin{aligned}
 R_{L^*}^*(h) &= \frac{m}{L^*} \sum_{i=1}^{L^*} h^1(B_i^*) \\
 &\quad + \sum_{j=2}^m \binom{m}{j} \binom{L^*}{j}^{-1} \sum_{1 \leq i_1 < \dots < i_j \leq L^*-1} h^{(j)}(B_{i_1}^*, \dots, B_{i_j}^*). \quad (9.12)
 \end{aligned}$$

Let

$$Z_{L^*}^* = \sqrt{L^*} \left(\frac{1}{L^*} \sum_{i=1}^{L^*} h^1(B_i^*) - \frac{1}{L} \sum_{j=1}^L \tilde{h}_1(\mathbf{B}_j) \right).$$

We must begin the demonstration of convergence of regenerative bootstrapped U -statistic by proving the weak convergence of Z_n^* , i.e., the convergence of this process in probability under \mathbb{P}_ν to a Gaussian process G indexed by \mathcal{F} whose simple paths are bounded and uniformly continuous with respect to the metric $L_2(\mathbb{P}^m)$. Note that Giné et al., 1990

studied the weak convergence and the central limit theorem for Bootstrapped sample. To prove the weak convergence of the first summation, we do the same footstep used in Theorem 4.2 Section 4. We want to prove first the finite-dimensional convergence of distributions of $Z_{l_n}^*$ to G . So, we only need to prove that for every fixed finite collection of functions $\{f_{i_1}, \dots, f_{i_k}\} \subset \mathcal{F}$, $(Z_{l_n}^*(f_{i_1}), \dots, Z_{l_n}^*(f_{i_k}))$ converges weakly in probability to $(G(f_{i_1}), \dots, G(f_{i_k}))$. It is sufficient to show that for every fixed collection $(a_1, \dots, a_k) \in \mathbb{R}$ we have;

$$\sum_{j=1}^k a_j Z_{L^*}^*(f_{i_j}) \rightarrow N(0, \sigma^2) \text{ in probability under } \mathbb{P}_\nu,$$

where

$$\sigma^2 = \sum_{j=1}^k a_j^2 \text{Var}\left(Z_{L^*}^*(f_{i_j})\right) + \sum_{s \neq r} a_s a_r \text{Cov}\left(Z_{L^*}^*(f_{i_s}), Z_{L^*}^*(f_{i_r})\right). \quad (9.13)$$

Take

$$h(\cdot) = \sum_{j=1}^k a_j f_{i_j}(\cdot).$$

By linearity of $h(\cdot)$, and theorem 1, we have $Z_{L^*}^*(h) \rightarrow G(h)$. The convergences of Z_{L^*} to $G(h)$ in probability under \mathbb{P}_ν was guaranteed by Bertail et al., 2006b. Hence, finite-dimensional convergence is proved. To verify stochastic asymptotic equicontinuity, we are going to check if for every $\varepsilon > 0$

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \mathbb{P}^* \left(\sup_{d(f,g) \leq \delta} |Z_{L^*}^*(f) - Z_{L^*}^*(g)| > \varepsilon \right) = 0 \text{ in probability under } \mathbb{P}_\nu.$$

Notice that we have

$$\begin{aligned} Z_{L^*}^* &= \sqrt{L^*} \left(\frac{1}{L^*} \sum_{i=1}^{L^*} h^1(B_i^*) - \frac{1}{L} \sum_{j=1}^L \tilde{h}_1(\mathbf{B}_j) \right) \\ &= \frac{\sqrt{L^*}}{L^*} \left(\sum_{i=1}^{L^*} h^1(B_i^*) - L^* \mathbb{P}_n h \right) \\ &= (\sqrt{L^*})^{-1} \left(\sum_{i=1}^{L^*} h^1(B_i^*) - \sum_{i=1}^{L^*} \mathbb{P}_n h \right) \\ &= (\sqrt{L^*})^{-1} \left(\sum_{i=1}^{L^*} (h^1(B_i^*) - \mathbb{P}_n h) \right) \\ &= \left(\sqrt{1 + \left\lfloor \frac{n}{\mathbb{E}_A(\boldsymbol{\tau})} \right\rfloor} \right)^{-1} \left(\sum_{i=1}^{1 + \left\lfloor \frac{n}{\mathbb{E}_A(\boldsymbol{\tau})} \right\rfloor} (h^1(B_i^*) - \mathbb{P}_n h) \right) + o_{\mathbb{P}^*}(1). \end{aligned}$$

Let us introduce

$$T_n^* = \left(1 + \left\lfloor \frac{n}{\mathbb{E}_A(\tau)} \right\rfloor\right)^{-1/2} \left(\sum_{i=1}^{1 + \left\lfloor \frac{n}{\mathbb{E}_A(\tau)} \right\rfloor} (h^1(B_i^*) - \mathbb{P}_n h) \right),$$

now we switch to treat this process. As we study the equicontinuity here, we take $\Gamma = f - g$. For simplicity, let

$$\Psi(n) = 1 + \left\lfloor \frac{n}{\mathbb{E}_A(\tau)} \right\rfloor \quad \text{and} \quad Y_i = \Gamma^1(B_i^*) - \mathbb{P}_n \Gamma.$$

We shall now evaluate the following probability

$$\begin{aligned} \mathbb{P}^* \left(\left\| \sum_{i=1}^{\Psi(n)} Y_i \right\|_{\mathcal{F}_\delta} > \varepsilon \Psi(n) \right) &= \mathbb{P}^* \left(\left\| Y_1 + \dots + Y_{\Psi(n)} \right\|_{\mathcal{F}_\delta} > \varepsilon \sqrt{\Psi(n)} \right) \\ &\leq \mathbb{P}^* \left(\left\| \Gamma^1(B_1^*) + \dots + \Gamma^1(B_{\Psi(n)}^*) \right\|_{\mathcal{F}_\delta} > \frac{\varepsilon \sqrt{\Psi(n)}}{2} \right) \\ &\quad + \mathbb{P}^* \left(\left\| \sum_{i=1}^{\Psi(n)} \mathbb{P}_n \Gamma \right\|_{\mathcal{F}_\delta} > \frac{\varepsilon \sqrt{\Psi(n)}}{2} \right) \\ &= I + II. \end{aligned}$$

Using Markov's inequality and the fact that $\Gamma^1(B_i^*)$ are i.i.d, for every $i \geq 1$, we have:

$$\begin{aligned} I &= \mathbb{P}^* \left(\left\| \Gamma^1(B_1^*) + \dots + \Gamma^1(B_{\Psi(n)}^*) \right\|_{\mathcal{F}_\delta} > \frac{\varepsilon \sqrt{\Psi(n)}}{2} \right) \\ &\leq \frac{4}{\Psi(n)} \mathbb{E}^* \left(\left\| \Gamma^1(B_1^*) + \dots + \Gamma^1(B_{\Psi(n)}^*) \right\|_{\mathcal{F}_\delta} \right)^2 \\ &= \frac{4}{\Psi(n)} \Psi(n) \mathbb{E}^* \left(\left\| \Gamma^1(B_1^*) \right\|_{\mathcal{F}_\delta} \right)^2. \end{aligned}$$

Remark that we have

$$\mathbb{E}^* \left(\left\| \Gamma^1(B_1^*) \right\|_{\mathcal{F}_\delta} \right)^2 = \frac{1}{\Psi(n)} \sum_{i=1}^{\Psi(n)} \left\| \Gamma^1(B_1^*) \right\|_{\mathcal{F}_\delta}^2 \rightarrow \mathbb{E}_A \left(\left\| \Gamma^1(\mathbf{B}_1) \right\|_{\mathcal{F}_\delta} \right)^2 \quad \text{a.s.}$$

We have

$$\mathbb{E}_A \left(\left\| \Gamma^1(\mathbf{B}_1) \right\|_{\mathcal{F}_\delta} \right)^2 = \mathbb{E}_A \left(\left\| \sum_{i=\mathcal{T}_0+1}^{\mathcal{T}_1} \Gamma^1(X_i) \right\|_{\mathcal{F}_\delta} \right)^2$$

$$\begin{aligned}
 &= \mathbb{E}_A \left(\left\| \sum_{i=\mathcal{F}_0+1}^{\mathcal{F}_1} \Gamma^{1^2}(X_i) \right\|_{\mathcal{F}_\delta} \right) + \mathbb{E}_A \left(\left\| \sum_{i=\mathcal{F}_0+1}^{\mathcal{F}_1} \sum_{i \neq j} \Gamma^1(X_i) \Gamma^1(X_j) \right\|_{\mathcal{F}_\delta} \right) \\
 &\leq \delta^2 \mathbb{E}_A(\boldsymbol{\tau}) + 2\delta^2 (\mathbb{E}_A(\boldsymbol{\tau}))^2 \\
 &\xrightarrow{\delta \rightarrow 0} 0 \text{ in } \mathbb{P}_\nu \text{ in probability.}
 \end{aligned}$$

This in turn implies that

$$I = \mathbb{P}^* \left(\left\| \Gamma^1(B_1^*) + \dots + \Gamma^1(B_{\Psi(n)}^*) \right\|_{\mathcal{F}_\delta} > \frac{\varepsilon \sqrt{\Psi(n)}}{2} \right),$$

converges to zero in \mathbb{P}_ν probability. For the asymptotic behavior of II , and $\|\mathbb{P}_n \Gamma\|_{\mathcal{F}_\delta} \rightarrow 0$ in \mathbb{P}_ν probability, by using the stochastic equicontinuity of Z_n proved in Section 4, so :

$$\begin{aligned}
 \mathbb{P}^* \left(\left\| \sum_{i=1}^{\Psi(n)} \mathbb{P}_n \Gamma \right\|_{\mathcal{F}_\delta} > \frac{\varepsilon \sqrt{\Psi(n)}}{2} \right) &= \mathbb{P}^* \left(\|(\Psi(n) - 1) \mathbb{P}_n \Gamma\|_{\mathcal{F}_\delta} > \frac{\varepsilon \sqrt{\Psi(n)}}{2} \right) \\
 &\leq \frac{4}{\Psi(n)} \mathbb{E}^* \left(|\Psi(n) - 1| \|\mathbb{P}_n \Gamma\|_{\mathcal{F}_\delta}^2 \right) \\
 &\xrightarrow{n \rightarrow \infty} 0 \text{ in } \mathbb{P}_\nu \text{ probability.}
 \end{aligned}$$

These two convergences imply the equicontinuity of the linear part of regenerative bootstrapped U -statistics. Then finite multidimensional convergence and equicontinuity are satisfied by Z_n^* . The second summation of (9.12) will be proved in the same footsteps used in the proof of Theorem 1. We have, for $C > 0$,

$$\begin{aligned}
 &\mathbb{E}^* \left(\left\| \mathbb{E}(L)^{1/2} \sum_{j=2}^m \binom{m}{j} \binom{l_n^* - 1}{j}^{-1} \sum_{1 \leq i_1 < \dots < i_j \leq l_n^* - 1} h^{(j)}(B_{i_1}^*, \dots, B_{i_j}^*) \right\| \right) \\
 &\leq \sum_{j=2}^m \binom{m}{j} \binom{\Psi(n)}{j}^{-1} \mathbb{E}^* \left((\Psi(n))^{1/2} \left\| \sum_{1 \leq i_1 < \dots < i_j \leq \Psi(n)} h^{(j)}(B_{i_1}^*, \dots, B_{i_j}^*) \right\| \right) + o_{\mathbb{P}^*} \\
 &\leq C \sqrt{\Psi(n)} \sum_{j=2}^m \binom{m}{j} \binom{\Psi(n)}{j}^{-1} \mathbb{E}^* \left(\left\| \sum_{1 \leq i_1 < \dots < i_j \leq \Psi(n)} h^{(j)}(B_{i_1}^{*(1)}, \dots, B_{i_j}^{*(j)}) \right\| \right) + o_{\mathbb{P}^*} \\
 &\leq C \sqrt{\Psi(n)} \sum_{j=2}^m \binom{m}{j} \binom{\Psi(n)}{j}^{-1} \mathbb{E}^* \left(\left\| \sum_{1 \leq i_1 < \dots < i_j \leq \Psi(n)} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} h^{(j)}(B_{i_1}^{*(1)}, \dots, B_{i_j}^{*(j)}) \right\| \right) + o_{\mathbb{P}^*} \\
 &\leq C \sqrt{\Psi(n)} \sum_{j=2}^m \binom{m}{j} \binom{\Psi(n)}{j}^{-1} \mathbb{E}^* \left(\left\| \sum_{1 \leq i_1 < \dots < i_j \leq \Psi(n)} \sum_{1 \leq j_1 < j_2 \leq j} \varepsilon_{i_1}^{(j_1)} \varepsilon_{i_2}^{(j_2)} h^{(j)}(B_{i_1}^{*(1)}, \dots, B_{i_j}^{*(j)}) \right\| \right) + o_{\mathbb{P}^*} \\
 &\leq C \sqrt{\Psi(n)} \sum_{j=2}^m \binom{m}{j} \binom{\Psi(n)}{j}^{-1} \mathbb{E}^* \left(\left\| \sum_{1 \leq i_1 < \dots < i_j \leq \Psi(n)} \sum_{1 \leq j_1 < j_2 \leq j} \varepsilon_{i_{j_1}} \varepsilon_{i_{j_2}} h^{(j)}(B_{i_1}^*, \dots, B_{i_j}^*) \right\| \right) + o_{\mathbb{P}^*}
 \end{aligned}$$

$$\begin{aligned}
 &\leq C\sqrt{\Psi(n)} \sum_{j=2}^m \binom{m}{j} \left(\Psi(n)\right)^{-1} \mathbb{E}^* \left(\left\| \sum_{1 \leq i_1 < \dots < i_j \leq \Psi(n)} \varepsilon_{i_1} \varepsilon_{i_2} h^{(j)}(B_{i_1}^*, \dots, B_{i_j}^*) \right\| \right) \\
 &\leq C\sqrt{\Psi(n)} \binom{\Psi(n)}{m}^{-1} \mathbb{E}^* \left(\left\| \sum_{1 \leq i_1 < \dots < i_m \leq \Psi(n)} \varepsilon_{i_1} \varepsilon_{i_2} \omega_h(B_{i_1}^*, \dots, B_{i_m}^*) \right\|_{\mathcal{F}} \right) + o_{\mathbb{P}^*} \\
 &\leq C\sqrt{\Psi(n)} \binom{\Psi(n)}{m}^{-1} \mathbb{E}^* \left(\left\| \sum_{1 \leq i_1 < \dots < i_m \leq \Psi(n)} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} \omega_h(B_{i_1}^*, \dots, B_{i_m}^*) \right\|_{\mathcal{F}} \right) + o_{\mathbb{P}^*} \\
 &\leq C\mathbb{E}^* \left(\left\| (\Psi(n))^{-m+1/2} \sum_{1 \leq i_1 < \dots < i_m \leq \Psi(n)} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} \omega_h(B_{i_1}^*, \dots, B_{i_m}^*) \right\|_{\mathcal{F}} \right) + o_{\mathbb{P}^*} \text{ (By randomization)} \\
 &\leq C\mathbb{E}^* \left(\left\| (\Psi(n))^{-m+1/2} \sum_{1 \leq i_1 < \dots < i_m \leq \Psi(n)} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} (\omega_h \mathbf{1}_{\{H \leq M\}})(B_{i_1}^*, \dots, B_{i_m}^*) \right\|_{\mathcal{F}} \right) \\
 &\quad + C\mathbb{P}^m(H \mathbf{1}_{\{H > M\}}) + o_{\mathbb{P}^*},
 \end{aligned}$$

where H is the envelope of \mathcal{F} , and by integrability of H , $\mathbb{P}^m(H \mathbf{1}_{\{H > M\}}) \rightarrow 0$ as $M \rightarrow \infty$. Let

$$\mathcal{F}_M = \{h \mathbf{1}_{\{H \leq M\}} : h \in \mathcal{F}\}$$

be a subset of \mathcal{F} of cardinality $N(\mathcal{F}_M, e_{n,1}, \delta)$, δ dense in \mathcal{F}_M for the distance $e_{n,1}$, and let \mathbb{E}_ε the integration with respect to the Rademacher variables only. We have

$$\begin{aligned}
 &\leq C\mathbb{E}^* \left(\left\| \Psi(n)^{-m+1/2} \sum_{1 \leq i_1 < \dots < i_m \leq \Psi(n)} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} (\omega_h \mathbf{1}_{\{H \leq M\}})(B_{i_1}^*, \dots, B_{i_m}^*) \right\|_{\mathcal{F}} \right) \\
 &\leq C\delta + C\mathbb{E}_\varepsilon \left(\left\| \Psi(n)^{-m+1/2} \sum_{1 \leq i_1 < \dots < i_m \leq \Psi(n)} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} (\omega_h \mathbf{1}_{\{H \leq M\}})(B_{i_1}^*, \dots, B_{i_m}^*) \right\|_{\mathcal{F}_{\delta, M}} \right),
 \end{aligned}$$

where $\mathcal{F}_{\delta, M}$ is a subset of \mathcal{F}_M . We have

$$\begin{aligned}
 &\leq C\delta + C\mathbb{E}_\varepsilon \left(\left\| (\Psi(n))^{-m+1/2} \sum_{1 \leq i_1 < \dots < i_m \leq \Psi(n)} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} (\omega_h(B_{i_1}^*, \dots, B_{i_m}^*)) \right\|_{\mathcal{F}_{\delta, M}} \right) \\
 &\leq C\delta + KM\mathbb{E}_\varepsilon \left[\left(\log 2 + \log \tilde{N}(\mathcal{F}, \tilde{e}_{n,1}, \delta n^{-1/2}) \right); \text{ for } K < \infty \right. \\
 &\quad \left. \times \max_{h \in \mathcal{F}_M} \left(\mathbb{E}_\varepsilon \left| (\Psi(n))^{-m+1/2} \sum_{1 \leq i_1 < \dots < i_m \leq \Psi(n)} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} (\omega_h(B_{i_1}^*, \dots, B_{i_m}^*)) \right|^2 \right)^{1/2} \right].
 \end{aligned}$$

But,

$$\begin{aligned} & \mathbb{E} \left(\max_{h \in \mathcal{F}_M} \mathbb{E}_\varepsilon \left| (\Psi(n))^{-m+1/2} \sum_{1 \leq i_1 < \dots < i_m \leq \Psi(n)} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} (\omega_h(B_{i_1}^*, \dots, B_{i_m}^*)) \right|^2 \right) \\ & \simeq 4\mathbb{E}((\Psi(n))^{2(m-1)})\mathbb{E}H^2(\mathbf{B}_1^*, \dots, \mathbf{B}_m^*). \end{aligned}$$

Then, we infer that

$$\begin{aligned} & \mathbb{E}^* \left(\left\| (\Psi(n))^{1/2} \sum_{j=2}^m \binom{m}{j} \binom{\Psi(n)}{j}^{-1} \sum_{1 \leq i_1 < \dots < i_j \leq \Psi(n)} h^{(j)}(B_{i_1}^*, \dots, B_{i_j}^*) \right\| \right) \\ & \leq C\delta + KM\Psi(n)^{-1/2} \mathbb{E}_\varepsilon \left[(\log 2 + \log \widetilde{N}(\mathcal{F}, \tilde{e}_{n,1}, \delta n^{-1/2}))^2 \right]^{1/2} \left[\mathbb{E}H^2(\mathbf{B}_1^*, \dots, \mathbf{B}_m^*) \right]^{1/2}. \end{aligned} \quad (9.14)$$

We must show $\log \widetilde{N}(\mathcal{F}, \tilde{e}_{n,1}, \delta n^{-1/2})$ is finite, which can be noticed by using the property of covering numbers as it follows:

$$\widetilde{N}(\mathcal{F}, \tilde{e}_{n,1}, \delta n^{-1/2}) \leq (\widetilde{N}(\mathcal{F}, \tilde{e}_{n,2}, \delta n^{-1/2}/2))^2 < \infty. \quad (9.15)$$

So, for $c < \infty$, (9.14) and (9.15) imply

$$\lim_{n \rightarrow \infty} \mathbb{E}^* \left(\left\| (\Psi(n))^{1/2} \sum_{j=2}^m \binom{m}{j} \binom{\Psi(n)}{j}^{-1} \sum_{1 \leq i_1 < \dots < i_j \leq \Psi(n)} h^{(j)}(B_{i_1}^*, \dots, B_{i_j}^*) \right\|_{\mathcal{F}} \right) = 0. \quad (9.16)$$

Hence the proof is complete. \square

Proof of Proposition 6.3

We can write $W_n(h)$ as follow:

$$W_n(h) = (I) + (II),$$

where

$$\begin{aligned} (I) &= \binom{n_1-1}{2}^{-1} \binom{n_2-1}{2}^{-1} \\ & \times \left[\sum_{(I_{(n_1, n_2)})^c} \omega_h(\mathbf{B}_0^{(1)}, \mathbf{B}_{i_{11}}^{(1)}; \mathbf{B}_0^{(2)}, \mathbf{B}_{i_{21}}^{(2)}) + \sum_{i_{11}=1}^{l_{n_1}} \sum_{i_{21}=1}^{l_{n_2}} \omega_h(\mathbf{B}_{i_{11}}^{(1)}, \mathbf{B}_{l_{n_1}}^{(1)}; \mathbf{B}_{i_{21}}^{(2)}, \mathbf{B}_{l_{n_2}}^{(2)}) \right] \\ (II) &= \binom{n_1-1}{2}^{-1} \binom{n_2-1}{2}^{-1} \end{aligned}$$

$$\times \left[\sum_{i_{11}=1}^{l_{n_1}-1} \sum_{i_{21}=1}^{l_{n_2}-1} \omega_{\tilde{h}}(\mathbf{B}_{i_{11}}^{(1)}, \mathbf{B}_{i_{11}}^{(1)}; \mathbf{B}_{i_{21}}^{(2)}, \mathbf{B}_{i_{21}}^{(2)}) - \sum_{i_{11}=1}^{n_1} \sum_{i_{21}=1}^{n_2} \tilde{h}(X_{i_{11}}^{(1)}, X_{i_{11}}^{(1)}; X_{i_{21}}^{(2)}, X_{i_{21}}^{(2)}) \right].$$

Note that , $l_{n_i}/n_i \rightarrow \alpha_i^{-1}$ \mathbb{P}_{ν_i} -a.s, $\forall 1 \leq i \leq n_k$. Let us treat the convergence of (I) and (II) to prove our results. Note that

$$\begin{aligned} A &= n_1^{-1} n_2^{-1} \sum_{i_{11}=1}^{l_{n_1}-1} \sum_{i_{21}=1}^{l_{n_2}-1} \omega_{\tilde{h}}(\mathbf{B}_{i_{11}}^{(1)}, \mathbf{B}_{i_{11}}^{(1)}; \mathbf{B}_{i_{21}}^{(2)}, \mathbf{B}_{i_{21}}^{(2)}) \\ &= n_1^{-1} n_2^{-1} \sum_{i_{11}=1}^{l_{n_1}-1} \sum_{i_{21}=1}^{l_{n_2}-1} \left[\omega_h(\mathbf{B}_{i_{11}}^{(1)}, \mathbf{B}_{i_{11}}^{(1)}; \mathbf{B}_{i_{21}}^{(2)}, \mathbf{B}_{i_{21}}^{(2)}) \right. \\ &\quad \left. - \sum_{i_{11}=\tau_{A_1}+1}^{\tau_{A_1}(2)} \sum_{i_{11}=\tau_{A_1}+1}^{\tau_{A_1}(2)} \sum_{i_{21}=\tau_{A_2}+1}^{\tau_{A_2}(2)} \sum_{i_{21}=\tau_{A_2}+1}^{\tau_{A_2}(2)} \Theta(h) \right] \\ &= n_1^{-1} n_2^{-1} \left[\sum_{i_{11}=1}^{l_{n_1}-1} \sum_{i_{21}=1}^{l_{n_2}-1} \omega_h(\mathbf{B}_{i_{11}}^{(1)}, \mathbf{B}_{i_{11}}^{(1)}; \mathbf{B}_{i_{21}}^{(2)}, \mathbf{B}_{i_{21}}^{(2)}) \right. \\ &\quad \left. - \sum_{i_{11}=1}^{l_{n_1}-1} \sum_{i_{21}=1}^{l_{n_2}-1} \sum_{i_{11}=\mathcal{F}_{A_1}(0)+1}^{\mathcal{F}_{A_1}(1)} \sum_{i_{21}=\mathcal{F}_{A_2}(0)+1}^{\mathcal{F}_{A_2}(1)} \Theta(h) \right] \\ &= \frac{(l_{n_1}-1) \times (l_{n_2}-1)}{n_1 \times n_2} \left[\frac{1}{(l_{n_1}-1) \times (l_{n_2}-1)} \sum_{i_{11}=1}^{l_{n_1}-1} \sum_{i_{21}=1}^{l_{n_2}-1} \omega_h(\mathbf{B}_{i_{11}}^{(1)}, \mathbf{B}_{i_{11}}^{(1)}; \mathbf{B}_{i_{21}}^{(2)}, \mathbf{B}_{i_{21}}^{(2)}) \right. \\ &\quad \left. - \frac{1}{(l_{n_1}-1) \times (l_{n_2}-1)} \sum_{i_{11}=1}^{l_{n_1}-1} \sum_{i_{21}=1}^{l_{n_2}-1} (\tau_{A_1})(\tau_{A_1})(\tau_{A_2})(\tau_{A_2}) \Theta(h) \right] \\ &= \frac{(l_{n_1}-1) \times (l_{n_2}-1)}{n_1 \times n_2} \left[\frac{1}{(l_{n_1}-1) \times (l_{n_2}-1)} \sum_{i_{11}=1}^{l_{n_1}-1} \sum_{i_{21}=1}^{l_{n_2}-1} \omega_h(\mathbf{B}_{i_{11}}^{(1)}, \mathbf{B}_{i_{11}}^{(1)}; \mathbf{B}_{i_{21}}^{(2)}, \mathbf{B}_{i_{21}}^{(2)}) \right. \\ &\quad \left. - \frac{1}{(l_{n_1}-1) \times (l_{n_2}-1)} \sum_{i_{11}=1}^{l_{n_1}-1} \sum_{i_{21}=1}^{l_{n_2}-1} (l(\mathbf{B}_{i_{11}}^1))(l(\mathbf{B}_{i_{11}}^1))(l(\mathbf{B}_{i_{21}}^2))(l(\mathbf{B}_{i_{21}}^2)) \Theta(h) \right] \\ &\rightarrow \alpha_1^{-1} \alpha_2^{-1} \left\{ \mathbb{E}(\omega_h(\mathbf{B}_1^{(1)}, \mathbf{B}_1^{(1)}; \mathbf{B}_1^{(2)}, \mathbf{B}_1^{(2)}) - \mathbb{E}_{A_1}((\tau_{A_1})^2) \mathbb{E}_{A_2}((\tau_{A_2})^2) \Theta(h) \right\}. \end{aligned}$$

Now for I . We have

$$\begin{aligned} &n_1^{-2} n_2^{-2} \mathbb{E} \left[\left(\sum_{i_{11}=1}^{l_{n_1}} \sum_{i_{21}=1}^{l_{n_2}} \omega_{\tilde{h}}(\mathbf{B}_0^{(1)}, \mathbf{B}_{i_{11}}^{(1)}; \mathbf{B}_0^{(2)}, \mathbf{B}_{i_{21}}^{(2)}) \right)^2 \right] \\ &= n_1^{-2} n_2^{-2} \mathbb{E} \left[\left(\frac{(l_{n_1})(l_{n_2})}{(l_{n_1})(l_{n_2})} \sum_{i_{11}=1}^{l_{n_1}} \sum_{i_{21}=1}^{l_{n_2}} \omega_{\tilde{h}}(\mathbf{B}_0^{(1)}, \mathbf{B}_{i_{11}}^{(1)}; \mathbf{B}_0^{(2)}, \mathbf{B}_{i_{21}}^{(2)}) \right)^2 \right] \end{aligned}$$

$$\begin{aligned}
 &= \mathbb{E} \left[\left(\frac{(l_{n_1})(l_{n_2})}{n_1 \times n_2} \frac{1}{l_{n_1} \times l_{n_2}} \sum_{i_{11}=1}^{l_{n_1}} \sum_{i_{21}=1}^{l_{n_2}} \omega_{\tilde{h}}(\mathbf{B}_0^{(1)}, \mathbf{B}_{i_{11}}^{(1)}; \mathbf{B}_0^{(2)}, \mathbf{B}_{i_{21}}^{(2)}) \right)^2 \right] \\
 &\leq \mathbb{E} \left[\left(\omega_{|\tilde{h}|}(\mathbf{B}_0^{(1)}, \mathbf{B}_1^{(1)}; \mathbf{B}_0^{(2)}, \mathbf{B}_1^{(2)}) \right)^2 \right] \\
 &\leq \mathbb{E} \left[\left(\omega_{|h|}(\mathbf{B}_0^{(1)}, \mathbf{B}_1^{(1)}; \mathbf{B}_0^{(2)}, \mathbf{B}_1^{(2)}) - \sum_{i_{10}=0}^{\mathcal{T}_{A_1}(0)} \sum_{i_{11}=\mathcal{T}_{A_1}(0)+1}^{\mathcal{T}_{A_1}(1)} \sum_{i_{20}=0}^{\mathcal{T}_{A_2}(0)} \sum_{i_{21}=\mathcal{T}_{A_2}(0)+1}^{\mathcal{T}_{A_2}(1)} \Theta(h) \right)^2 \right] \\
 &\leq 2\mathbb{E} \left[\left(\omega_{|h|}(\mathbf{B}_0^{(1)}, \mathbf{B}_1^{(1)}; \mathbf{B}_0^{(2)}, \mathbf{B}_1^{(2)}) \right)^2 \right] + \left[\sum_{i_{10}=0}^{\mathcal{T}_{A_1}(0)} \sum_{i_{11}=\mathcal{T}_{A_1}(0)+1}^{\mathcal{T}_{A_1}(1)} \sum_{i_{20}=0}^{\mathcal{T}_{A_2}(0)} \sum_{i_{21}=\mathcal{T}_{A_2}(0)+1}^{\mathcal{T}_{A_2}(1)} \Theta(h)^2 \right] \\
 &\leq 2\mathbb{E} \left[\left(\omega_{|h|}(\mathbf{B}_0^{(1)}, \mathbf{B}_1^{(1)}; \mathbf{B}_0^{(2)}, \mathbf{B}_1^{(2)}) \right)^2 \right] \\
 &\quad + \left[\mathbb{E}_{\nu_1}(\tau_{A_1}(0)) \mathbb{E}_{A_1}(\tau_{A_1}(1)) \mathbb{E}_{\nu_2}(\tau_{A_2}(0)) \mathbb{E}_{A_2}(\tau_{A_2}(1)) \Theta(h)^2 \right] \\
 &< \infty.
 \end{aligned}$$

The last one, and in the same footsteps, we have

$$\begin{aligned}
 &n_1^{-2} n_2^{-2} \mathbb{E} \left[\left(\omega_{\tilde{h}}(\mathbf{B}_{i_{11}}^{(1)}, \mathbf{B}_{l_{n_1}}^{(1)}; \mathbf{B}_{i_{21}}^{(2)}, \mathbf{B}_{l_{n_2}}^{(2)}) \right)^2 \right] \\
 &\leq 2 \left\{ \mathbb{E} \left[\left(\omega_{|h|}(\mathbf{B}_1^{(1)}, \mathbf{B}_2^{(1)}; \mathbf{B}_1^{(2)}, \mathbf{B}_2^{(2)}) \right)^2 \right] \right. \\
 &\quad \left. + \left[\sum_{i_{11}=\mathcal{T}_{A_1}(0)+1}^{\mathcal{T}_{A_1}(1)} \sum_{i_{1m_1}=\mathcal{T}_{A_1}(l_{n_1})+1}^{n_1} \sum_{i_{21}=\mathcal{T}_{A_2}(0)+1}^{\mathcal{T}_{A_2}(1)} \sum_{i_{2m_2}=\mathcal{T}_{A_2}(l_{n_2})+1}^{n_2} \Theta(|h|)^2 \right] \right\} \\
 &\leq 2 \left\{ \mathbb{E} \left[\left(\omega_{|h|}(\mathbf{B}_1^{(1)}, \mathbf{B}_2^{(1)}; \mathbf{B}_1^{(2)}, \mathbf{B}_2^{(2)}) \right)^2 \right] \right. \\
 &\quad \left. + \left[\mathbb{E}(l(\mathbf{B}_1^{(1)})) \mathbb{E}(l(\mathbf{B}_{l_{n_1}}^{(1)})) \mathbb{E}(l(\mathbf{B}_1^{(2)})) \mathbb{E}(l(\mathbf{B}_{l_{n_2}}^{(2)})) \Theta(|h|)^2 \right] \right\} \\
 &\leq 2 \left\{ \mathbb{E} \left[\left(\omega_{|h|}(\mathbf{B}_1^{(1)}, \mathbf{B}_2^{(1)}; \mathbf{B}_1^{(2)}, \mathbf{B}_2^{(2)}) \right)^2 \right] + (\mathbb{E}_{A_1}(\tau_{A_1}))^2 (\mathbb{E}_{A_2}(\tau_{A_2}))^2 \Theta(|h|)^2 \right\} \\
 &< \infty.
 \end{aligned}$$

Therefore the proof is complete. \square

Proof of theorem 6.1

The proof is due to P. K. Sen, 1977. In our case, $k = 2$, the condition will be

$$\mathbb{E}\{|h|(\log^+ |h|)\} < \infty,$$

then the regenerative U -statistics converges almost sure to 0 (as it is centered). Therefore, Theorem 6.1 holds from Proposition 6.3. \square

Proof of Theorem 6.2

The H -decomposition of $R_{(l_{n_1}, l_{n_2})}$ can be written as

$$R_{(l_{n_1}, l_{n_2})}(h) = m_1 H_{l_{n_1}, l_{n_2}}^{(1,0)}(h) + m_2 H_{l_{n_1}, l_{n_2}}^{(0,1)}(h) + D_{l_{n_1}, l_{n_2}}(h), \quad (9.17)$$

such that

$$H_{l_{n_1}, l_{n_2}}^{(1,0)}(h) = (l_{n_1} - 1)^{-1} \sum_{i=1}^{l_{n_1}-1} \omega_{\tilde{h}}^{(1,0)}(\mathbf{B}_i^{(1)}),$$

and

$$H_{l_{n_1}, l_{n_2}}^{(0,1)}(h) = (l_{n_2} - 1)^{-1} \sum_{j=1}^{l_{n_2}-1} \omega_{\tilde{h}}^{(0,1)}(\mathbf{B}_j^{(2)}).$$

We know that $H_{l_{n_1}, l_{n_2}}^{(1,0)}(h)$ and $H_{l_{n_1}, l_{n_2}}^{(0,1)}(h)$ converge to a normal distributions with mean zero and variance $\delta_{1,0}$ and $\delta_{0,1}$ respectively. We know also that

$$\text{Var}(D_{l_{n_1}, l_{n_2}}) = O(l^{-1}).$$

To complete the proof, it is necessary to take $l = l_{n_1} + l_{n_2}$ and suppose that $\min(l_{n_1}, l_{n_2}) \rightarrow \infty$, $p_n = l_{n_1}/l \rightarrow p$. The proof is attributed to A. J. Lee, 1990. Lehmann, 1951 and P. K. Sen, 1974b have both demonstrated weak convergences in various methods. \square

Proof of Theorem 7.1

In our Markovian context, we will set the weak convergence of Markovian U -statistic via the regenerative method, where the samples of the blocks are i.i.d. So, we will handle the weak convergence of the bootstrapped regenerative U -statistic (7.2) by using

$$R_L(h) = \binom{l_{n_1}^* - 1}{2}^{-1} \binom{l_{n_2}^* - 1}{2}^{-1} \sum_{1 \leq i_{11} < i_{12} \leq l_{n_1} - 1} \sum_{1 \leq i_{21} < i_{22} \leq l_{n_2} - 1} \omega_{\tilde{h}}(\mathbf{B}_{i_{11}}^*, \mathbf{B}_{i_{12}}^*; \mathbf{B}_{i_{21}}^*, \mathbf{B}_{i_{22}}^*).$$

For the first step, we write the H -decomposition for this U -statistic :

$$\begin{aligned} R_{(l_{n_1}, l_{n_2})}(h) &= \sum_{c_1=0}^{m_1=2} \sum_{c_2=0}^{m_2=2} \binom{l_{n_1} - 1}{c_1}^{-1} \binom{l_{n_2} - 1}{c_2}^{-1} \binom{m_1}{c_1} \binom{m_2}{c_2} \\ &\quad \sum_{(l_{n_1}-1, c_1)} \sum_{(l_{n_2}-1, c_2)} h^{(c_1, c_2)}(\mathbf{B}_{i_{11}}^*, \mathbf{B}_{i_{12}}^*; \mathbf{B}_{i_{21}}^*, \mathbf{B}_{i_{22}}^*) \end{aligned}$$

$$\begin{aligned}
 &= \frac{m_1}{l_{n_1^*}^* - 1} \sum_{i_1=1}^{l_{n_1^*}^* - 1} \tilde{h}^{(1,0)}(\mathbf{B}_{i_1}^*) + \frac{m_2}{l_{n_2^*}^* - 1} \sum_{i_2=1}^{l_{n_2^*}^* - 1} \tilde{h}^{(1,0)}(\mathbf{B}_{i_2}^*) + D^*(h) \\
 &= S^*(h) + D^*(h),
 \end{aligned}$$

where

$$\begin{aligned}
 \tilde{h}^{(1,0)}(b^*) &= \mathbb{E}(\omega_{\tilde{h}}(\mathbf{B}_{11}^*, \mathbf{B}_{12}^*; \mathbf{B}_{21}^*, \mathbf{B}_{22}^* | \mathbf{B}_{11}^* = b)) = \omega_{\tilde{h}}^{(1)}(\mathbf{B}_{11}^*), \\
 \tilde{h}^{(0,1)}(b^*) &= \mathbb{E}(\omega_{\tilde{h}}(\mathbf{B}_{11}^*, \mathbf{B}_{12}^*; \mathbf{B}_{21}^*, \mathbf{B}_{22}^* | \mathbf{B}_{21}^* = b)) = \omega_{\tilde{h}}^{(1)}(\mathbf{B}_{21}^*).
 \end{aligned}$$

Using the projection method, based on the Hoeffding decomposition of this U -statistic, we begin with the weak convergence of the linear part of this decomposition. Let

$$Z_{(l_{n_1}, l_{n_2})}^* = \sqrt{l^*} (S^*(h) - S(h)),$$

where

$$l^* = (l_{n_1}^* - 1) + (l_{n_2}^* - 1).$$

Commonly, we begin to prove first the finite-dimensional convergence of distributions of $Z_{(l_{n_1}, l_{n_2})}^*$ to G_{Θ} . So, we have to prove that for every fixed finite collection of functions $\{h_1, \dots, h_k\} \subset \mathcal{H}$, $(Z_{l_n^*}^*(h_{i_1}), \dots, Z_{(l_{n_1}, l_{n_2})}^*(h_{i_k}))$ converges weakly in probability to $(G_{\Theta}(h_{i_1}), \dots, G_{\Theta}(h_{i_k}))$. It is sufficient to show that for every fixed collection $(a_1, \dots, a_k) \in \mathbb{R}$, we have

$$\sum_{j=1}^k a_j Z_{l_n^*}^*(f_{i_j}) \rightarrow G_{\Theta}(h) \text{ in probability under } \mathbb{P}_{\nu},$$

where

$$\sigma^2 = \sum_{j=1}^k a_j^2 \text{Var}\left(Z_n(h_{i_j})\right) + \sum_{s \neq r} a_s a_r \text{Cov}\left(Z_n(h_{i_s}), Z_n(h_{i_r})\right) \quad (9.18)$$

and

$$h(\cdot) = \sum_{j=1}^k a_j h_{i_j}(\cdot).$$

We just need to show the linearity of $h(\cdot)$, to ensure this convergence. So, we have

$$\begin{aligned}
 Z_{(l_{n_1}, l_{n_2})}^*(h) &= \sqrt{l^*} (S^*(h) - S(h)) \\
 &= \sqrt{l^*} \left(\frac{m_1}{l_{n_1^*}^* - 1} \sum_{i_1=1}^{l_{n_1^*}^* - 1} \tilde{h}^{(1,0)}(\mathbf{B}_{i_1}^*) + \frac{m_2}{l_{n_2^*}^* - 1} \sum_{i_2=1}^{l_{n_2^*}^* - 1} \tilde{h}^{(1,0)}(\mathbf{B}_{i_2}^*) \right. \\
 &\quad \left. - \frac{m_1}{l_{n_1} - 1} \sum_{i_1=1}^{l_{n_1} - 1} \tilde{h}^{(1,0)}(\mathbf{B}_{i_1}) - \frac{m_2}{l_{n_2} - 1} \sum_{i_2=1}^{l_{n_2} - 1} \tilde{h}^{(1,0)}(\mathbf{B}_{i_2}) \right)
 \end{aligned}$$

$$= \sqrt{l^*} \left(\frac{m_1}{\sqrt{l_{n_1^*-1}^*}} Z_1^*(h) + \frac{m_2}{\sqrt{l_{n_2^*-1}^*}} Z_2^*(h) \right).$$

We have

$$\begin{aligned} \sum_{j=1}^k a_j Z_{(l_{n_1}, l_{n_2})}^*(h_{i_j}) &= \sqrt{l^*} \left(\frac{m_1}{\sqrt{l_{n_1^*-1}^*}} \sum_{j=1}^k a_j Z_1^*(h_{i_j}) + \frac{m_2}{\sqrt{l_{n_2^*-1}^*}} \sum_{j=1}^k a_j Z_2^*(h_{i_j}) \right) \\ &= \sqrt{l^*} \left(\frac{m_1}{\sqrt{l_{n_1^*-1}^*}} Z_1^* \left(\sum_{j=1}^k a_j h_{i_j} \right) + \frac{m_2}{\sqrt{l_{n_2^*-1}^*}} Z_2^* \left(\sum_{j=1}^k a_j h_{i_j} \right) \right) \\ &= Z_{(l_{n_1}, l_{n_2})}^* \left(\sum_{j=1}^k a_j h_{i_j} \right). \end{aligned}$$

Therefore, by linearity of Z_1^* and Z_2^* , we can see the linearity of $Z_{(l_{n_1}, l_{n_2})}^*$, which is sufficient to be used with Theorem 6.2, to prove the finite multidimensional convergence for this linear term. For the second part, stochastic equicontinuity of this empirical process, we can check that:

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \mathbb{P}^* \left(\sup_{d(f, g) \leq \delta} |Z_{(l_{n_1}, l_{n_2})}^*(f - g)| > \varepsilon \right) = 0 \text{ in probability under } \mathbb{P}_\nu. \quad (9.19)$$

for all $f, g \in \mathcal{H}$ and $\delta > 0$. Note that

$$\begin{aligned} Z_{(l_{n_1}, l_{n_2})}^* &= \sqrt{l^*} (S^*(h) - S(h)) \\ &= \sqrt{l^*} \left(\frac{m_1}{l_{n_1^*-1}^*} \sum_{i_1=1}^{l_{n_1^*-1}^*} \tilde{h}^{(1,0)}(\mathbf{B}_{i_1}^*) + \frac{m_2}{l_{n_2^*-1}^*} \sum_{i_2=1}^{l_{n_2^*-1}^*} \tilde{h}^{(1,0)}(\mathbf{B}_{i_2}^*) \right. \\ &\quad \left. - \frac{m_1}{l_{n_1-1}} \sum_{i_1=1}^{l_{n_1-1}} \tilde{h}^{(1,0)}(\mathbf{B}_{i_1}) - \frac{m_2}{l_{n_2-1}} \sum_{i_2=1}^{l_{n_2-1}} \tilde{h}^{(1,0)}(\mathbf{B}_{i_2}) \right) \\ &\leq \sqrt{l^*} \left[\left(\frac{m_1}{l_{n_1^*-1}^*} \sum_{i_1=1}^{l_{n_1^*-1}^*} \tilde{h}^{(1,0)}(\mathbf{B}_{i_1}^*) - \frac{m_1}{l_{n_1-1}} \sum_{i_1=1}^{l_{n_1-1}} \tilde{h}^{(1,0)}(\mathbf{B}_{i_1}) \right) \right. \\ &\quad \left. + \left(\frac{m_2}{l_{n_2^*-1}^*} \sum_{i_2=1}^{l_{n_2^*-1}^*} \tilde{h}^{(1,0)}(\mathbf{B}_{i_2}^*) - \frac{m_2}{l_{n_2-1}} \sum_{i_2=1}^{l_{n_2-1}} \tilde{h}^{(1,0)}(\mathbf{B}_{i_2}) \right) \right] \\ &\leq \sqrt{l^*} \left(\frac{m_1}{l_{n_1^*-1}^*} \sum_{i_1=1}^{l_{n_1^*-1}^*} \tilde{h}^{(1,0)}(\mathbf{B}_{i_1}^*) - \frac{m_1}{l_{n_1-1}} \sum_{i_1=1}^{l_{n_1-1}} \tilde{h}^{(1,0)}(\mathbf{B}_{i_1}) \right) \\ &\quad + \sqrt{l^*} \left(\frac{m_2}{l_{n_2^*-1}^*} \sum_{i_2=1}^{l_{n_2^*-1}^*} \tilde{h}^{(1,0)}(\mathbf{B}_{i_2}^*) - \frac{m_2}{l_{n_2-1}} \sum_{i_2=1}^{l_{n_2-1}} \tilde{h}^{(1,0)}(\mathbf{B}_{i_2}) \right) \\ &\leq \frac{\sqrt{l^*}}{\sqrt{l_{n_1^*-1}^*}} (Z_1^*(h)) + \frac{\sqrt{l^*}}{\sqrt{l_{n_2^*-1}^*}} (Z_2^*(h)). \end{aligned}$$

We can complete the demonstration if we prove that $\frac{\sqrt{l^*}}{\sqrt{l_{n_i}^* - 1}}$ is finite, while $Z_1^*(h)$ and $Z_2^*(h)$ will be treated as the same way as the proof of Theorem 5.3, and then the stochastic equicontinuity is established. According to A. J. Lee, 1990 and Bertail et al., 2006b, we have to see that the number of blocks and the number of observations satisfies the following

$$\frac{l}{l_{n_i} - 1} \rightarrow p_i^{-1}, \quad n^*/n \rightarrow 1, \quad (l_{n_i}^* - 1)/n_i - \mathbb{E}_{A_i}(\tau_{A_i})^{-1} \rightarrow 0.$$

As $V(\sqrt{l^*}D^*(h)) = O(\sum_{i=1}^2 n/n_i) = O(l^{-1})$, the proof of this theorem is completed. Note that Zhang et al., 1996 treat the convergence of bootstrapped generalized U -process by proving $\sqrt{n}(U_n - U_n^*)$, so they did not use or treat those terms. Also, we can use this reference as a second method for the demonstration where the projection method is not used. \square

10 Appendix

Lemma 10.1 (Levental, 1988). *Let $(X_n)_n$ be a regenerative process, then*

- *If $\mathbb{E}_A(\tau) < \infty$, then*

$$\boldsymbol{\mu}(B) = \frac{1}{\mathbb{E}_A(\tau)} \mathbb{E}_A \left(\sum_{\mathcal{T}_0 < k \leq \mathcal{T}_1} \mathbb{1}_B(X_k) \right),$$

here $\boldsymbol{\mu}(\cdot)$ is called a steady state distribution, for any set $B \in \mathcal{E}$.

- *If $\mathbb{E}(\tau) < \infty$, then*

$$\frac{1}{n} \sum_{1 \leq i \leq n} f(X_i) \rightarrow \boldsymbol{\mu}(f) \quad \text{a.s. for every } f \in L_1(\boldsymbol{\mu}).$$

- *The $\tau(j)$ are i.i.d. random variables, also the $f(\mathbf{B}_j), j = 1, 2, \dots$ are i.i.d. for any measurable function $f(\cdot)$.*

Lemma 10.2 (Radulović, 2004). *Let $(X_n)_n$ be a Markov Chain satisfying the Assumption 2.1 with an initial probability ν and (C1) and let $\alpha = \mathbb{E}_A(\tau)$ defined in (2.1), we have:*

i) $\left| \frac{\mathbb{E}_\nu(l_n)}{\alpha^{-1}n} - 1 \right| \leq \frac{C}{n}$, for $C > 0$ and for $\mu > 0$, $\sqrt{n} \left(\frac{\mathbb{E}_\nu(l_n)}{\alpha^{-1}n} - 1 \right) \rightarrow N(0, \mu^2)$ in the sense of Hoffman-Jørgensen.

ii) For n large enough and for $\varepsilon > 0$ there exists $M > 0$

$$\mathbb{P} \left(\left| n - \sum_{i=0}^{l_n} l(\mathbf{B}_i) \right| > Mn^{1/4} \right) < \varepsilon$$

iii) For n^* denotes the size of the bootstrapped sample and n denotes the size of initial one, we have

$$\frac{n^*}{n} \xrightarrow{\mathbb{P}_\nu \times \mathbb{P}^*} 0.$$

iv) If X_i is a sequence of random variables such that $\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i \rightarrow C$ a.s and if t_n is an integer valued sequence of random variables, then

$$\frac{1}{t_n} \sum_{i=1}^{t_n} X_i \rightarrow C.$$

Lemma 10.3 (Montgomery-Smith, 1993). *Let $(X_i)_{i=0}^\infty$ be a sequence of independent identically distributed random variables, then there is a universally positive and finite constants C_1 and C_2 such that*

$$\mathbb{P} \left(\max_{k \leq n} \left\| \sum_{i=1}^k X_i \right\| > t \right) \leq C_1 \mathbb{P} \left(\left\| \sum_{i=1}^n X_i \right\| > C_2 t \right). \quad (10.1)$$

Theorem 10.1 (de la Peña et al., 1999). *If \mathcal{H} is a measurable class of real-valued functions of m variables with envelope H satisfying the integrability conditions*

$$E |H(X_{i_1}, \dots, X_{i_m})|^{\sharp(i_1, \dots, i_m)/m} < \infty$$

for all $1 \leq i_1, \dots, i_m \leq m$, then the V -processes $\{V_n^n(h) : h \in \mathcal{H}\}$ satisfy the strong law of large numbers, that is

$$\|V_n(h) - P^m h\|_{\mathcal{H}} \rightarrow 0 \quad \text{a.s.},$$

if and only if the U -processes $\{U_n(h) : h \in \mathcal{H}\}$ do.

Explicit details for the Hoeffding decomposition:

To examine the oscillations of U -process, we can dissect the various components of its Hoeffding decomposition. Hoeffding decomposition or the H-decomposition is the representation of U -statistics of degree m into a sum of m uncorrelated U -statistics of degree $1, \dots, m$ respectively, often used to study the asymptotic behavior of U -statistic, which is directly related to its order of degeneracy. We have

$$R_L(h) = \binom{L}{m}^{-1} \sum_{1 \leq i_1 < \dots < i_m \leq L} \omega_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) = \binom{L}{m}^{-1} H_n.$$

Let

$$H_{nj} = \sum_{1 \leq i_1 < \dots < i_j \leq L} \tilde{h}^{(j)}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_j}).$$

For $c = m$;

$$\begin{aligned} R_L(h) &= \binom{L}{m}^{-1} \sum_{1 \leq i_1 < \dots < i_m \leq L} \omega_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \\ &= \binom{L}{m}^{-1} \sum_{1 \leq i_1 < \dots < i_m \leq L} \sum_{k=1}^m \sum_{1 \leq j_1 < \dots < j_k \leq m} \tilde{h}^{(k)}(\mathbf{B}_{i_{j_1}}, \dots, \mathbf{B}_{i_{j_k}}) \\ &= \binom{L}{m}^{-1} \sum_{k=1}^m \sum_{1 \leq i_1 < \dots < i_m \leq L} \sum_{1 \leq j_1 < \dots < j_k \leq m} \tilde{h}^{(k)}(\mathbf{B}_{i_{j_1}}, \dots, \mathbf{B}_{i_{j_k}}) \\ &= \binom{L}{m}^{-1} \sum_{k=1}^m \binom{L-k}{m-k} \sum_{1 \leq i_1 < \dots < i_j \leq L} \tilde{h}^{(j)}(\mathbf{B}_{i_{j_1}}, \dots, \mathbf{B}_{i_{j_k}}) \\ &= \sum_{k=1}^m \binom{L}{m}^{-1} \binom{L-k}{m-k} \sum_{1 \leq i_1 < \dots < i_j \leq L} \tilde{h}^{(j)}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_j}) \\ &= \sum_{k=1}^m \binom{L}{k}^{-1} \binom{m}{k} \sum_{1 \leq i_1 < \dots < i_j \leq L} \tilde{h}^{(j)}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_j}). \end{aligned}$$

According to Hoeffding, the representation of U -statistics can be being as follow:

$$R_L(h) = \sum_{j=1}^m \binom{m}{j} \binom{L}{j}^{-1} H_{nj}.$$

This decomposition asserts the martingale property for the sequence $\{H_{nj}\}_{n>j}$ for each $j = 1, \dots, m$. Note that the terms of H_{nj} are non-correlated with increasing variance of order n . Then,

$$R_L(h) = \binom{m}{1} \binom{L}{1} H_{n1} + \sum_{j=2}^m \binom{m}{j} \binom{L}{j}^{-1} H_{nj}.$$

The interested reader can refer to Serfling, 1980 and A. J. Lee, 1990 for more details.

Remark 10.2. *In general, statistics of interest can be treated as a vector space projected in the sub-space. We are interested in the projection to evaluate this statistic's variance and the rate of convergence. The space of U -statistics is L_2 , the Hilbert space constructed by square-integrable random variables, given by the usual dot product. Suppose $L_2^{(1)}$ represents a sub-space of random variables of form $\sum h(X_i)$ where X_i are i.i.d. with the mean square-integrable. In that case, the linear term is a projection in this sub-space, and the other term of the Hoeffding decomposition is orthogonal to $L_2^{(1)}$.*

Order of convergence of the regenerative U -statistics

To show the moment rate of convergence of regenerative U -statistics, we first introduce the following lemma

Lemma 10.4 (Moment inequality of U -statistics). *(Zhang et al., 1996) Let U_n a U -statistics of degree 2 ($m = 2$). Suppose that $\mathbb{E}h(X_1, X_2) = 0$ and $\mathbb{E}h(X_1, X_2)^q < \infty$ for $q \geq 2$. Then there exists constants C_{1q} and C_{2q} depend only on q such that*

$$\mathbb{E}|U_n|^q \leq C_{1q}n^{-q}(\mathbb{E}|h(X_1, X_2)|^2)^{q/2} + C_{2q}n^{2-2q}(\mathbb{E}|h(X_1, X_2)|^q) \quad (10.2)$$

Looking at the second-order moment of regenerative U -statistics to show the rate of convergence, then we have:

$$\begin{aligned} \mathbb{E}|R_L(h)|^q &= \mathbb{E} \left| \binom{l_n - 1}{m}^{-1} \sum_{1 \leq i_1 < \dots < i_m \leq L} \omega_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \right|^q \\ &\leq C_{1q}(l_n - 1)^{-q} \left(\mathbb{E} \left| \omega_{\tilde{h}}(\mathbf{B}_1, \dots, \mathbf{B}_m) \right|^2 \right)^{q/2} + C_{2q}(l_n - 1)^{2-2q} \mathbb{E} \left| \omega_{\tilde{h}}(\mathbf{B}_1, \dots, \mathbf{B}_m) \right|^q \\ &\leq C_{1q}(l_n - 1)^{-q} \mathbb{E} \left| \omega_{\tilde{h}}(\mathbf{B}_1, \dots, \mathbf{B}_m) \right|^{2q/2} + C_{1q}(l_n - 1)^{-q} \mathbb{E}_A(\boldsymbol{\tau})^m (|\boldsymbol{\Theta}(h)|^2)^{q/2} \\ &\quad + C_{2q}(l_n - 1)^{2-2q} \mathbb{E} \left| \omega_{|h|}(\mathbf{B}_1, \dots, \mathbf{B}_m) \right|^q + C_{2q}(l_n - 1)^{2-2q} \mathbb{E}_A(\boldsymbol{\tau})^m |\boldsymbol{\Theta}(h)|^q. \end{aligned}$$

For $q = 2$ and using the previous lemma with the fact that all terms in the last inequality are finite, we can find that $R_L(h)$ is of order $O((l_n - 1)^{-2})$.

Blocks assumptions

Regarding the first and the last block, which represents the sum respectively from the initial state of the Markov chain up to the first regeneration time and the initial segment of the last block, which could be empty if the last block ("full" block) ends with n . Adamczak in Adamczak, 2008 gives us a bound for the first and last blocks to control them.

Chapter 5

Exchangeably weighted bootstraps of the General Markov U -process

This chapter develops the content of an accepted article published in Mathematics with the required modifications to fit this thesis manuscript.

Objective

We explore an exchangeably weighted bootstrap of the general function-indexed empirical U -processes in the Markov setting, which is a natural higher-order generalization of the weighted bootstrap empirical processes. As a result of our findings, a considerable variety of bootstrap resampling strategies arise. This chapter aims to provide theoretical justifications for the exchangeably weighted bootstrap consistency in the Markov setup. General structural conditions on the classes of functions (possibly unbounded) and the underlying distributions are required to establish our results. This chapter provides the first general theoretical study of the bootstrap of the empirical U -processes in the Markov setting. Potential applications include the symmetry test, Kendall's tau and the test of independence.

Contents

1	Introduction	153
2	Notation and definitions	156
2.1	Markov chain	156
2.2	Exchangeable weights	159
2.3	The U -process framework	160
2.4	Gaussian chaos Process	161
2.5	Technical Assumptions:	162
3	Preliminary results	163
3.1	The bootstrapped U - Processes	165
4	Weighted Bootstrap weak convergence	168

4.1	Bootstrap weights examples:	171
5	Applications	172
5.1	Symmetry test	172
5.2	Kendall's tau	173
5.3	Test of independence	174
6	Mathematical development	174

1 Introduction

U -statistics are a class of estimators, initially explored in association with unbiased estimators by Halmos, 1946 and officially introduced by Hoeffding, 1948, and are defined as follows: let $\{X_i\}_{i=1}^\infty$ be a sequence of random variables defined on a measurable space (E, \mathcal{E}) , and let $h : E^m \rightarrow \mathbb{R}$ be a measurable function, the U -statistics of order m and kernel h based on the sequence $\{X_i\}$ are

$$U_n(h) = \frac{(n-m)!}{n!} \sum_{(i_1, \dots, i_m) \in I_n^m} h(X_{i_1}, \dots, X_{i_m}), \quad n \geq m,$$

where

$$I_n^m = \{(i_1, \dots, i_m) : i_j \in \mathbb{N}, 1 \leq i_j \leq n, i_j \neq i_k \text{ if } j \neq k\}.$$

The empirical variance, Gini's mean difference or Kendall's rank correlation coefficient are common examples of U -estimators, while a classical test based on a U -statistic is Wilcoxon's signed rank test for the hypothesis of the location at zero (see, e.g., A. W. van der Vaart, 1998, Example 12.4). Halmos, 1946, v. Mises, 1947 and Hoeffding, 1948 provided, amongst others, the first asymptotic results for the case in which the underlying random variables have independent and identical distributions. Extensive literature have treated the theory of U -statistics, for instance, see Serfling, 1980, de la Peña et al., 1999, Arcones et al., 1993, A. J. Lee, 1990, etc. Complex statistical issues are also amenable to being solved using U -processes. Examples include tests for goodness-of-fit, nonparametric regression, and density estimation. U -processes are a set of U -statistics that are indexed by a family of kernels. U -processes might be viewed as infinite-dimensional variants of U -statistics with a single kernel function or as nonlinear stochastic extensions of empirical processes. Both thoughts have the following advantages: first, considering a large group of statistics rather than a single statistic is more statistically interesting. Second, We may use ideas from the theory of empirical processes to construct limit or approximation theorems for U -processes. Nevertheless, achieving results in U -processes is not easy. Extending U -statistics to U -processes necessitates significant effort and distinct methodologies; generalizing empirical processes to U -processes is quite challenging, especially when U -processes are presented in the stationary setting. We highlight that the U -processes are used often in statistics, such as when higher order terms are a part of von Mises expansions. Particularly, the study of estimators (including function estimators) with various smoothness degrees involves U -statistics. For instance, Stute, 1993 applied almost sure uniform bounds for \mathbb{P} -canonical U -processes to analyse the product limit estimator for truncated data. Two new tests for normality based on U -processes are also presented in Arcones et al., 2006. Inspired by Giné et al., 2007a, 2007b, Schick et al., 2011 developed another tests for normality that employ weighted L_1 -distances between the standard nor-

mal density and local U -statistics based on standardized observations as test statistics. Estimating the mean of multivariate functions in case of possibly heavy-tailed distributions was explored by Joly et al., 2016, they presented the median-of-means too, and both explorations were based on U -statistics. Besides, other researchers emphasize the importance of U -processes, Ghosal et al., 2000; S. Lee et al., 2009 and Abrevaya et al., 2005 used it for testing qualitative features of functions in nonparametric statistics, Nolan et al., 1987 represent the cross-validation for density estimation using U -statistics, in addition to Arcones et al., 1993, Sherman, 1994 and de la Peña et al., 1999 where they established limiting distributions of M -estimators. Since then, this discipline has made significant advancements, and the results have been broadly interpreted. Asymptotic behavior are demonstrated under weak dependence assumptions, for example, in the works of Yoshihara, 1976, Borovkova et al., 2001, and in Denker et al., 1983 or more recently in Leucht, 2012 as well as more generally in Leucht et al., 2013 and Bouzebda and Nemouchi, 2022. But, in practice, explicit computation is not always possible due to the complexity of the U -processes' limiting distributions or their functionals. We suggest a general bootstrap of the U -processes in the Markov setting to solve this issue, which is a challenging problem. The concept of the bootstrap, given by Efron, 1979, in the case of independent and identically distributed (iid) random variables, is to re-sample from an original sample of observations of an unknown marginal distribution function $F(x)$, X_1, \dots, X_n , a new i.i.d sample X_1^*, \dots, X_n^* with the marginal distribution function $F_n(x)$, which represents the empirical distribution function constructed from the original sample. Moreover, it is commonly known that the bootstrap approach gives a better approximation to the statistic's distribution, mainly when the sample size is small, Hall, 1992. Bootstrap for U -statistics of independent observations were studied by Bickel et al., 1981, Arcones et al., 1992, Dehling et al., 1994 and Leucht et al., 2009. However, the Bootstrap technique is different and not the same for dependent variables because the dependence structure cannot be conserved in the new sample. For this reason, other *Blockwise Bootstrap methods* were introduced, aiming to keep the structure of dependence. Among those methods, we can cite the *circular Block Bootstrap* introduced by Politis et al., 1992 and the *non-overlapping block bootstrap* introduced by Carlstein, 1986b. Politis et al., 1994b proposed a bootstrap method related to the weakly dependent stationary observation, the *stationary bootstrap*. This latter can be seen as an expansion of the circular block bootstrap, where a random variable, such as a geometric random variable, can be used for the block length. It is important to note that Efron's initial bootstrap formulation (see Efron, 1979) had a few flaws. To be more precise, certain observations might be sampled several times while others might not be at all. A more generalized version of the bootstrap, the *weighted bootstrap*, has been developed to get around this issue and has also been demonstrated to be computationally more appealing in some applications. This resampling strategy was initially described in D. B. Rubin, 1981 and thoroughly investigated by Bickel et al.,

1981, who coined the name "weighted bootstrap". For example, bayesian bootstrap when the weighted vector

$$(\xi_{n1}, \dots, \xi_{nn}) = (M_{n1}, \dots, M_{nn}),$$

is equal to the vector of n spacings of $n - 1$ ordered uniform $(0, 1)$ random variables in distributions, that is,

$$(M_{n1}, \dots, M_{nn}) \sim \text{Dirichlet}(n; 1, \dots, 1).$$

For more details see Lo, 1993. This diversity of resampling approaches necessitates the use of a uniform approach, commonly known as general weighted resampling, and was first described by Mason et al., 1992 and has since been developed by Præstgaard et al., 1993 and A. van der Vaart, 1996. Alvarez-Andrade et al., 2013 investigated the almost sure rate of convergence of strong approximation for the weighted bootstrap process by a sequence of Brownian bridge, refer to Bouzebda, 2012 for the multivariate setting and Bouzebda, Elhattab, and Ferfache, 2022 for recent references. The concept of the generalized bootstrap, introduced by Mason et al., 1992 is extended to the class of nondegenerate U -statistics of degree 2 and the corresponding Studentized U -statistics by Hůsková et al., 1993, refer to Janssen, 1994 and Alvarez-Andrade et al., 2020. Q. Han, 2022 generalized in his article this theory for a higher order. In his work, he developed a multiplier inequality of U -process for i.i.d random variables. We mention that the multiplier processes' theory is directly and strongly related to the symmetrization inequalities investigated by Arcones et al., 1993 and de la Peña et al., 1999.

This chapter aims to investigate the exchangeable bootstrap for U -processes in the same way that Q. Han, 2022 did but without the restriction of independence setting. The previous reference focused on U -processes in the independent framework, whereas this chapter considers U -processes in the dependent setting of Markov chains. We believe we are the first to present a successful consideration in this general context. We will combine the techniques of the renewal bootstrap with the randomly weighted bootstrap in a non-trivial way. We mention a connection between moving blocks bootstrap and its modification, matched block bootstrap, at this point. Instead of artificially splitting a sample into fixed-size blocks and then resampling them, the latter seeks to match the blocks to create a smoother transition; for more information, see Radulović, 2004. The main difficulties in proving Theorem 4.2 are due to the random size of the resampled blocks. This randomness generates a problem with the random stopping times, which can not be removed by replacing a random stopping time with its expectation. In the present setting, the bootstrap random variables are generated by resampling from a random number of blocks. One can think that using the conditioning arguments can overcome the problem,

but the answer is negative. Our proof uses some arguments from Radulović, 2004 and Q. Han, 2022 by verifying bootstrap stochastic equicontinuity by comparing it to the original process in a similar way as in Giné et al., 1990. However, as we shall see later, integrating concepts from these papers is not enough to solve the problem. To deal with U -processes in the Markov framework, sophisticated mathematical derivations will be necessary. We present the first complete theoretical justification of the bootstrap consistency. This justification requires the efficient use of large sample theoretical approaches established for U -empirical processes.

The rest of this chapter is organized as follows. Section 2 is devoted to the introduction of the Markov framework, the U -process, the bootstrap weights, and the definitions needed in our work. In Section 3, we recall the necessary ingredient for the U -statistics and the U -processes in the Markov setting. Also, we provide some asymptotic results including the weak convergence of U -processes in Theorem 3.1. In Section 4, we will come up with the main results concerning the bootstrap of the U -processes. In Section 5, we collect some examples of weighted U -statistics. To prevent interrupting the flow of the presentation, all proofs are gathered in Section 6. The Appendix contains a few pertinent technical findings and proofs.

2 Notation and definitions

In what follows, we aim to define properly our settings. For this reason, we have collected the definitions and notation needed.

2.1 Markov chain

Let $X = (X_n)_{n \in \mathbb{N}}$ be an homogeneous ψ -irreducible Markov chain, that means that the chain has a stationary transition probabilities, defined on a measurable space (E, \mathcal{E}) , where \mathcal{E} is a separable σ -algebra. Let $\pi(x, dy)$ be the transition probability, $\nu = \nu(i)_{i > 0}$ the initial probability. Therefore, we will denote by \mathbb{P}_ν or just \mathbb{P} the probability measure for $\mathbb{P} = (\pi, \nu)$. Likewise, \mathbb{E}_ν will denote the integration with respect to \mathbb{P}_ν . In our framework, let \mathbb{P}_x be a probability measure such that $X_0 = x$ and $X_0 \in E$ and $\mathbb{E}_x(\cdot)$ is the \mathbb{P}_x -expectation. We will further assume that the Markov chain is Harris positive recurrent¹ with an atom \mathbf{A} .

1

Definition 2.1 (Harris-recurrent). *A Markov chain $X = (X_n)_{n \in \mathbb{N}}$ is said to be Harris-recurrent if there exists a σ -finite measure such that, for ψ a positive measure on a countable generated measurable space (E, \mathcal{E}) , $\psi(E) > 0$ and if for all $B \in \mathcal{E}$ with $\psi(B) > 0$ then*

$$P_x(\cup_{i=1}^{\infty} (X_i \in B)) = 1 \quad \text{for any } x \in E.$$

Recall that a chain is positive Harris recurrent and aperiodic if and only if it is ergodic Nummelin, 1984, Proposition 6.3, i.e., there exists a probability measure π , called the stationary distribution, such that, in total variation distance,

$$\lim_{n \rightarrow +\infty} \|\mathbb{P}^n(x, \cdot) - \pi\|_{\text{tv}} = 0.$$

Definition 2.2 (small sets). *A set $S \in \mathcal{E}$ is said to be Ψ -small if there exists $\delta > 0$, a positive probability measure Ψ supported by S and an integer $m \in \mathbb{N}^*$ such that*

$$\forall x \in S, B \in \mathcal{E}, \mathbb{P}^m(x, B) \geq \delta \Psi(B). \quad (2.1)$$

Definition 2.3. *Let $(X_n)_{n \geq 1}$ be a Markov Chain taking value in (E, \mathcal{E}) . We say that $(X_n)_{n \geq 1}$ is positive recurrent if*

1. $(X_n)_{n \geq 1}$ is (A, p, ν, m) recurrent (or Harris-recurrent if \mathcal{E} is countably generated), where $A \in \mathcal{E}$ a set, $0 < p < 1$, m an integer and ν a probability measure.
2. $\sup_{x \in A} \mathbb{E}_x(\mathcal{T}_0) < \infty$, where \mathcal{T}_0 is the hitting time of A by the m step chain, roughly speaking, $\mathcal{T}_0 = \min\{i \geq 1 : X_{i,m} \in A\}$.

Definition 2.4. *A ψ -irreducible aperiodic^a chain X is called regenerative or atomic if there exists a measurable set \mathbf{A} called an atom, in such a way that $\psi(\mathbf{A}) > 0$ and for all $(x, y) \in \mathbf{A}^2$ we have $\mathbb{P}(x, \cdot) = \mathbb{P}(y, \cdot)$. Roughly speaking, an atom is a set on which the transition probabilities are the same. If a finite number of states or subsets are visited from the chain, then any state or any subset of the states is actually an atom.*

^a

Definition 2.5 (aperiodicity). *Assuming ψ -irreducibility^a, there exists $d' \in \mathbb{N}^*$ and disjoint sets $D_1, \dots, D_{d'}$ (set $D_{d'+1} = D_1$) positively weighted by ψ such that*

$$\psi(E \setminus \cup_{1 \leq i \leq d'} D_i) = 0$$

and

$$\forall x \in D_i, \mathbb{P}(x, D_{i+1}) = 1.$$

The period of the chain is the greatest common divisor d of such integers, it is said to be aperiodic if $d = 1$.

^a

Definition 2.6 (irreducibility). *The chain is ψ -irreducible if there exists a σ -finite measure ψ such that, for all set $B \in \mathcal{E}$, when $\psi(B) > 0$, for any $x \in E$ there exists $n > 0$ such that $\mathbb{P}^n(x, B) > 0$.*

One of the most important properties of Harris-recurrent Markov chains is the existence of invariant distribution which we will be called μ (a limiting probability distribution, also called occupation measure). Also, Harris-recurrent Markov chains can always

be embedded in a certain Markov chain on an extended sample space with a recurrent atom. The existence of recurrent atom \mathbf{A} gives an immediate consequence for the construction of a regenerative extension of this chain. The time that the chain hits a given atom (recurrent state) is seen as the regenerative time. Athreya et al., 1978 and Nummelin, 1978 give the construction of such a regenerative extension. The development of a regenerative extension makes the use of regenerative technique possible in order to study this type of Markov chain. As we mentioned above, we will assume in this work that the Harris-recurrent chain is atomic, i.e., the set which is infinitely almost sure is well defined and accessible, this set \mathbf{A} is called an atom. By definition, an atom \mathbf{A} is a set, in \mathcal{E} , where $\mu(\mathbf{A}) > 0$, and for all $x, y \in \mathbf{A}$, $\pi(x, \cdot) = \pi(y, \cdot)$. Let $\mathbb{P}_{\mathbf{A}}$ (respectively by $\mathbb{E}_{\mathbf{A}}$) be the probability measure on the underlying space such that $x \in \mathbf{A}$ (respectively the $\mathbb{P}_{\mathbf{A}}$ -expectation).

The imposed conditions on the Markov chain insure that the defined atom \mathbf{A} (or the constructed one in the case of a non-atomic chain) is one recurrent class, and let us define the following terms.

Hitting times: Define $\mathcal{T}_j : E \rightarrow \mathbb{N} \cup \{\infty\}$ by

$$\begin{aligned}\mathcal{T}_0 &:= \inf\{n \geq 0 : X_n \in \mathbf{A}\}, \\ \mathcal{T}_j &:= \inf\{n \geq \mathcal{T}_{j-1} : X_n \in \mathbf{A}\}.\end{aligned}\tag{2.2}$$

A well known property of the hitting time is that for all $j \in \mathbb{N}$, $\mathcal{T}_j < \infty$, \mathbb{P}_{ν} - *a.s* Chung, 1967, chap. I14.

Renewal times: Using the hitting times, we can define the renewal times as

$$\begin{aligned}\tau_0 &:= \mathcal{T}_0 + 1, \\ \tau(j) &:= \mathcal{T}_j - \mathcal{T}_{j-1}.\end{aligned}\tag{2.3}$$

Similar to the regenerative process, the sequence of renewal times $\{\tau(j)\}_{j=1}^{\infty}$ is i.i.d. and it is independent of the choice of the initial probability. All over this work, we set $\tau = \tau(1)$ and $\alpha = \mathbb{E}_{\mathbf{A}}(\tau)$.

Regenerative Blocks: Let $l_n := \max\{j : \sum_{i=0}^j \tau(j) \leq n\}$ be the number of visits to the atom \mathbf{A} . Using the strong property² of Markov chain it is possible to divide the given sample (X_1, \dots, X_n) into a sequence of blocks $\{\mathbf{B}_j\}_{j=0}^{l_n}$ such that:

$$\begin{aligned} \mathbf{B}_0 &= \{X_1, \dots, X_{\mathcal{T}_0}\}, \\ \mathbf{B}_j &= \{X_{\mathcal{T}_{j-1}+1}, \dots, X_{\mathcal{T}_j}\} \text{ in } \mathbb{T} = \bigcup_{n=1}^{\infty} E^n, \text{ for all } j = 1, \dots, l_n, \\ \mathbf{B}_{l_n}^{(n)} &= \{X_{\mathcal{T}_{l_n-1}+1}, \dots, X_n\}, \end{aligned} \quad (2.4)$$

here l_n is the total number of blocks. The length of each block will be denoted by

$$l(\mathbf{B}_j) := \mathcal{T}_j - \mathcal{T}_{j-1}.$$

2.2 Exchangeable weights

In what follows, $\boldsymbol{\xi}$ represents a real-valued random variable, $\boldsymbol{\xi}_i$ are independent from (X_i) . For $1 \leq p < \infty$, we denote the p -norm by

$$\|\boldsymbol{\xi}\|_p = (\mathbb{E}(|\boldsymbol{\xi}|^p))^{1/p}.$$

Assuming the following:

(A1) $(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n)$ are exchangeable³, non-negative, symmetric and for all n

$$\sum_{i=1}^n \boldsymbol{\xi}_i = n.$$

(A2) $\frac{1}{n} \max_{1 \leq i \leq n} (\boldsymbol{\xi}_i - 1)^2 \rightarrow 0$ in $\mathbb{P}_{\boldsymbol{\xi}}$ -probability which satisfied by the assumption of the moment

$$\sup_n \|\boldsymbol{\xi}_1\|_{2m,1} < \infty.$$

2

Definition 2.7 (Strong Markov property). *Let $(X_n)_{n \geq 0}$ be a Markov chain, with T the stopping time of $(X_n)_{n \geq 0}$. Then conditionally on $T < \infty$ and $X_T = i$, $(X_{T+n})_{n \geq 0}$ is a sequence of Markov chain and is independent of X_0, \dots, X_T .*

3

Definition 2.8. (Exchangeability). *Let $\boldsymbol{\xi}_{n1}, \dots, \boldsymbol{\xi}_{nn}$ be a sequence of random variables with joint distribution $\mathbb{P}_{\boldsymbol{\xi}}$ and let $\Sigma(n)$ be the group of all permutations acting on $\{1, \dots, n\}$. We say that $\boldsymbol{\xi}_{n1}, \dots, \boldsymbol{\xi}_{nn}$ is exchangeable if, for all $\sigma(i) \in \Sigma(n)$,*

$$\mathbb{P}_{\boldsymbol{\xi}}(\boldsymbol{\xi}_{n1}, \dots, \boldsymbol{\xi}_{nn}) = \mathbb{P}_{\boldsymbol{\xi}}(\boldsymbol{\xi}_{n\sigma(1)}, \dots, \boldsymbol{\xi}_{n\sigma(n)}).$$

(A3) There exists $c > 0$ such that, in \mathbb{P}_ξ -probability,

$$\frac{1}{n} \sum_{i=1}^n (\xi_i - 1)^2 \rightarrow c^2 > 0.$$

(A4) Assume

$$\lim_{\lambda \rightarrow \infty} \overline{\lim}_{t \geq \lambda} t^2 \mathbb{P}_\xi(\xi_1 \geq t) = 0.$$

2.3 The U -process framework

Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of random variables with values in a measurable space (E, \mathcal{E}) . Let $h : E^m \rightarrow \mathbb{R}$ be a measurable function symmetric in its arguments. The U -statistic of order (or degree) m and kernel $h(\cdot)$ is defined as:

$$U_n(h) = \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_n^m} h(X_{i_1}, \dots, X_{i_m}), \text{ for } n \geq m. \quad (2.5)$$

The fact the the kernel is completely degenerated with the condition $\mathbb{P}^m h^2 < \infty$, are used for the orthogonality of the different elements of the Hoeffding decomposition of the U -statistics. We can associate some distances $e_{n,p}$ to the covering numbers, where

$$e_{n,p} = (U_n(|f - g|^p))^{1/p}.$$

In this work, we use the two distances defined afterward

$$e_{n,2}(f, g) = \left(\frac{(n-m)!}{n!} \sum_{0 \leq i_1 < \dots < i_m \leq n} (f - g)(X_{i_1}, \dots, X_{i_m})^2 \right)^{1/2}.$$

For decoupled statistics, we also associated covering numbers, well-known as $\widetilde{N}(\varepsilon, \mathcal{F}, \tilde{e}_{n,p})$ and a distance, which can be defined for $p = 2$ as follows:

$$\tilde{e}_{n,2}(f, g) = n^{1/2} \frac{(n-m)!}{n!} \left[\mathbb{E}_\varepsilon \left(\sum_{0 \leq i_1 < \dots < i_m \leq n} \varepsilon_{i_1} (f - g)(X_{i_1}, \dots, X_{i_m}) \right)^2 \right]^{1/2}.$$

In the context of U -process $\{U_n\}$, stochastic equicontinuity at a function $g \in \mathcal{F}$ implies generally that $|U_n(h) - U_n(g)|$ should be uniformly small for all $h(\cdot)$ close enough to $g(\cdot)$, with high probability and for all n large enough.

2.4 Gaussian chaos Process

Definition 2.9. Let H denotes a real separable Hilbert space with scalar product $\langle \cdot, \cdot \rangle_H$. We say that a stochastic process $\mathbb{G} = \{\mathbb{G}_{\mathbb{P}}(h), h \in H\}$ defined in a complete probability space $(E, \mathcal{E}, \mathbb{P})$ is an *isonormal Gaussian process* (or a *Gaussian process on H*) if $\mathbb{G}_{\mathbb{P}}$ is a centered Gaussian family of random variables such that $\mathbb{E}(\mathbb{G}_{\mathbb{P}}(h)\mathbb{G}_{\mathbb{P}}(g)) = \langle h, g \rangle_H$ for all $h, g \in H$.

Define the mapping $h \rightarrow \mathbb{G}_{\mathbb{P}}(h)$. Under the assumption mentioned above, this map is linear and it provides a linear isometry of H onto a closed subspace $L_2(E, \mathcal{E}, \mathbb{P})$ which contains a zero mean Gaussian random variables as elements. Let $\mathbb{K}_{\mathbb{P}}$ be the isonormal Gaussian chaos process associated with $\mathbb{G}_{\mathbb{P}}$ determined by:

$$\mathbb{K}_{\mathbb{P}}(h_m^\psi) = (m!)^{\frac{1}{2}} \mathfrak{R}_m(\mathbb{G}_{\mathbb{P}}(\psi), \mathbb{E}\psi^2, 0, \dots, 0),$$

where $h_m^\psi(x_1, \dots, x_m) = \psi(x_1) \cdots \psi(x_m)$, $\psi \in L_2(\mathbb{P})$, \mathfrak{R}_m is a polynomial defined as a sum of monomials of degree m , de la Peña et al., 1999 give us a simple expression of this polynomial, extracted from Newton's identity given by

$$\sum_{1 \leq i_1 < \dots < i_m \leq n} t_{i_1} \cdots t_{i_m} = \mathfrak{R}_m\left(\sum_{i=1}^n t_i, \sum_{i=1}^n t_i^2, \dots, \sum_{i=1}^n t_i^m\right).$$

Therefore,

$$\sum_{1 \leq i_1 < \dots < i_m \leq n} \psi(x_{i_1}) \cdots \psi(x_{i_m}) = \mathfrak{R}_m\left(\sum_{i=1}^n \psi(x_i), \sum_{i=1}^n \psi(x_i)^2, \dots, \sum_{i=1}^n \psi(x_i)^m\right).$$

Hence, by continuous of mapping theorem, we can see that CLT and LLN give:

$$\begin{aligned} & \left(\binom{n}{m_1}^{\frac{1}{2}} U_n(h_{m_1}^{\psi_1}), \dots, \binom{n}{m_r}^{\frac{1}{2}} U_n(h_{m_r}^{\psi_r}) \right) \\ & \rightarrow \left((m_1!)^{\frac{1}{2}} R_{k_1}(\mathbb{G}_{\mathbb{P}}(\psi_1), \mathbb{E}\psi_1^2, 0, \dots, 0), \dots, (m_r!)^{\frac{1}{2}} \mathfrak{R}_{k_r}(\mathbb{G}_{\mathbb{P}}(\psi_r), \mathbb{E}\psi_r^2, 0, \dots, 0) \right). \end{aligned}$$

Under linearity of kernel, we only need to show that:

$$\begin{aligned} & \left\{ \binom{n}{m}^{\frac{1}{2}} U_n(f) : f \in \mathcal{F} \right\} \\ & \xrightarrow{d} \left\{ \mathbb{K}_{\mathbb{P}}(f_k) = m! \mathfrak{R}_m(\mathbb{G}_{\mathbb{P}}(\psi), \mathbb{E}\psi^2, 0, \dots, 0) : f_k \in \mathcal{F} \right\} \text{ in } \ell^\infty(\mathcal{F}), \end{aligned}$$

to hold the weak convergence. The limit $\mathbb{K}_{\mathbb{P}}$ is useful in the case of degenerate U -statistics and it provides a convergence of all moments, which in turn plays a crucial role because it

is due to the hypercontractivity which makes the uniform integrability better. For a good explanation of \mathbb{K}_P , readers are invited to see de la Peña et al., 1999, Chapter 4, Section 4.2.

2.5 Technical Assumptions:

For our results, we need the following assumptions.

(C.1) (Block-length assumption) For all $q \geq 1$, and $l \geq 1$,

$$\mathbb{E}_\nu[\tau^l] < \infty, \quad \mathbb{E}_A[\tau^q] < \infty;$$

(C.2) (Non-regenerative blocks) For $l \geq 1$, we have

$$\mathbb{E}_\nu \left[\left(\sum_{i_1=1}^{\mathcal{T}_0} \sum_{i_2=\mathcal{T}_0+1}^{\mathcal{T}_1} \sum_{i_3=\mathcal{T}_1+1}^{\mathcal{T}_2} \cdots \sum_{i_m=\mathcal{T}_{m-1}+1}^{\mathcal{T}_m} |h(X_{i_1}, \dots, X_{i_m})| \right)^l \right] < \infty,$$

also

$$\mathbb{E}_\nu \left[\left(\sum_{i_1=\mathcal{T}_0+1}^{\mathcal{T}_1} \sum_{i_2=\mathcal{T}_1+1}^{\mathcal{T}_2} \cdots \sum_{i_{m-1}=\mathcal{T}_{m-1}+1}^{\mathcal{T}_m} \sum_{i_m=\mathcal{T}(l_n)+1}^n |h(X_{i_1}, \dots, X_{i_{m-1}}, X_{i_m})| \right)^l \right] < \infty$$

(C.3) (Block-sum : Moment assumptions) For $l \geq 1$, we have

$$\mathbb{E}_\nu \left[\left(\sum_{i_1=\mathcal{T}_0+1}^{\mathcal{T}_1} \sum_{i_2=\mathcal{T}_1+1}^{\mathcal{T}_2} \cdots \sum_{i_m=\mathcal{T}_{m-1}+1}^{\mathcal{T}_m} |h(X_{i_1}, \dots, X_{i_m})| \right)^l \right] < \infty,$$

and

$$\mathbb{E}_A \left[\left(\sum_{\mathcal{T}_0+1 \leq i_1 \leq \dots \leq i_m \leq \mathcal{T}_1} h(X_{i_1}, \dots, X_{i_m}) \right)^l \right] < \infty;$$

(C.4) For $l \geq 1$, we have

$$\mathbb{E}_\nu \left[\left(\sum_{i_1=\mathcal{T}_0+1}^{\mathcal{T}_1} \sum_{i_2=\mathcal{T}_1+1}^{\mathcal{T}_2} \underbrace{\sum_{i_k=\mathcal{T}_k+1}^{\mathcal{T}_{k+1}} \cdots \sum_{i_k=\mathcal{T}_k+1}^{\mathcal{T}_{k+1}}}_{u \text{ times}} \sum_{i_{k+u}=\mathcal{T}_{k+u}+1}^{\mathcal{T}_{k+u+1}} \cdots \sum_{i_m=\mathcal{T}_{m-1}+1}^{\mathcal{T}_m} |h(X_{i_1}, \underbrace{X_{i_k}, \dots, X_{i_k}}_{u \text{ times}}, X_{i_{k+u}}, \dots, X_{i_m})| \right)^l \right] < \infty;$$

(C.5) (Non-degeneracy.) We suppose also that

$$\mathbb{E}_A \left[\left(\sum_{i=\mathcal{T}_0+1}^{\mathcal{T}_1} h_1(X_i) \right)^2 \right] > 0.$$

Remark 2.10. *[Moment assumptions] In practice, we recall that block-moment assumptions for the split Markov chain can be generally checked by establishing drift conditions of Lyapounov's type for the original chain, see Chapter 11 in S. Meyn et al., 2009b and Douc et al., 2008, as well as the references therein. All these moment conditions are discussed in detail in the book of S. P. Meyn et al., 1993, Chapters 11 & 17. There is a key condition in the proof of ergodic theorems in the Markovian context, which is the fact that $\mathbb{E}_{\mathbf{A}}(\tau_0) < \infty$, for all \mathbf{A} a set in \mathcal{E} , such that $\psi(\mathbf{A}) > 0$. In fact, when there is a finite invariant measure and an atom \mathbf{A} then this condition is right founded. We also refer to Bertail et al., 2006a for an explicit check of such conditions on several important examples and to §4.1.2 of Bertail et al., 2011b for sufficient conditions expressed in terms of uniform return rate to small sets. Finally, as discussed in Chapter 8 of Revuz, 1984, similar conditions can be expressed in potential kernels. Observe that, in the positive recurrent case, the assumptions of (C.1) are not independent when $\nu = \mu$: from basic renewal theory, one has $\mathbb{P}_\mu(\tau = k) = (\mathbb{E}_A[\tau])^{-1} \mathbb{P}_A(\tau \geq k)$ for all $k \geq 1$. Hence, conditions $\mathbb{E}_\mu[\tau^l] < \infty$ and $\mathbb{E}_A[\tau^{l+1}] < \infty$ are equivalent.*

3 Preliminary results

A significant issue was detected in recovering the estimation of our parameter of interest using the U-process. The given shape of this parameter is as follow:

$$\Theta(h) = \int_{x_1 \in E} \dots \int_{x_k \in E} h(x_1, \dots, x_k) \boldsymbol{\mu}(dx_1) \dots \boldsymbol{\mu}(dx_k),$$

where $h : E^m \rightarrow \mathbb{R}$ is a kernel function. The estimation of this parameter should be possible using the U -statistics of the form:

$$U_n(h) = \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_n^m} h(X_{i_1}, \dots, X_{i_m}), \text{ for } n \geq m, \quad (3.1)$$

As the parameter of interest will be defined, and based on Kac's theorem for occupation measure, $\boldsymbol{\mu}(h)$ in the regeneration setup can be written as follows:

$$\boldsymbol{\mu}(h) = \frac{1}{(\mathbb{E}_A(\boldsymbol{\tau}))^m} \mathbb{E}_A \left(\sum_{i_1=\mathcal{T}_0+1}^{\mathcal{T}_1} \sum_{i_2=\mathcal{T}_1+1}^{\mathcal{T}_2} \dots \sum_{i_m=\mathcal{T}_{(m-1)}+1}^{\mathcal{T}_m} h(X_{i_1}, \dots, X_{i_m}) \right). \quad (3.2)$$

In the Markovian context and since the variables are not independent, the approximation related to the i.i.d. blocks and the regenerative case will be introduced below:

Definition 3.1. (*Regenerative Kernel*) Let $h : E^m \rightarrow \mathbb{R}$ a kernel. We define the regenerative kernel $\boldsymbol{\omega}_h : \mathbb{T}^m \rightarrow \mathbb{R}$ as follows:

$$\boldsymbol{\omega}_h((x_{11}, \dots, x_{1n_1}), \dots, (x_{k1}, \dots, x_{kn_k})) = \sum_{i_1=1}^{n_1} \dots \sum_{i_k=1}^{n_k} h(x_{1i_1}, \dots, x_{ki_k}). \quad (3.3)$$

It is not necessary that the kernel $\boldsymbol{\omega}_h(\cdot)$ to be symmetric, as soon as $h(\cdot)$. In fact, we can use the symmetrization of $S_m \boldsymbol{\omega}_h$ in the following way

$$(S_m \boldsymbol{\omega}_h) = (m!)^{-1} \sum_{\sigma(1)=1}^{n_1} \dots \sum_{\sigma(m)=1}^{n_k} h(x_{\sigma(1)}, \dots, x_{\sigma(m)}), \quad (3.4)$$

where the sum is over all permutations $\sigma = \{i_1, \dots, i_m\}$ of $\{1, \dots, m\}$. Next, we consider the U -statistic formed by the regenerative data.

Definition 3.2. (*Regenerative U -statistic*) Let $h : E^m \rightarrow \mathbb{R}$ a kernel such that $\Theta(|h|) < \infty$ and set $\tilde{h}(\cdot) = h(\cdot) - \Theta(h)$. The regenerative U -statistic associated with the sequence of regenerative blocks $\{\mathbf{B}_j\}_{j=1}^L$, generated by the Markov chain is given by

$$R_{l_n}(h) = \binom{l_n - 1}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_{l_n-1}^m} \boldsymbol{\omega}_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}). \quad (3.5)$$

Hence, $R_{l_n}(h)$ is a standard U -statistics with mean zero.

Proposition 3.3. Let us define

$$W_n(h) = U_n(h) - \Theta(h) - \binom{l_n - 1}{m} \binom{n}{m}^{-1} R_{l_n}(h). \quad (3.6)$$

Then, under conditions (C.1), (C.2), (C.3) and (C.4), we have the following stochastic convergences:

$$W_n(h) \rightarrow 0, \quad \mathbb{P}_\nu - a.s.$$

Before stating the weak convergence in the next theorem, we will define the corresponding U -processes related to the U -statistic U_n and the regenerative U -statistics R_L respectively:

$$Z_n := \binom{n}{m}^{1/2} [U_n - \Theta(h)], \quad (3.7)$$

$$T_{l_n} := \binom{l_n}{m}^{1/2} [R_{l_n} - \mathbb{E}(R_{l_n})]. \quad (3.8)$$

Theorem 3.1

Let $(X_n)_n$ be a positive recurrent Harris Markov chain, with an accessible atom \mathbf{A} , X_n satisfies the conditions (C.1) and (C.2) (moments assumptions), (C.3), (C.4), (C.5). Let \mathcal{F} be a uniform bounded class of functions with an envelope H square integrable such that:

$$\int_0^\infty (\log N(\varepsilon, \mathcal{F}, e_{n,2}))^{m/2} d\varepsilon < \infty.$$

Then the process Z_n converges weakly in probability under \mathbb{P}_ν to a Gaussian process $\mathbb{G}_\mathbb{P}$ indexed by \mathcal{F} whose sample paths are bounded and uniformly continuous with respect to the metric $L_2(\mathbb{P}_\nu)$.

3.1 The bootstrapped U - Processes

Trying to facilitate the bootstrap technique, we write the detailed steps of regenerative block construction and the weighted bootstrap method in the following algorithm:

Algorithm 1 Regenerative Block and weighted bootstrap construction:

1. Identify the number of visits $l_n = \sum_{i=0}^n \mathbb{1}_{X_i \in \mathbf{A}}$ to the atom \mathbf{A} .
2. Divide the observed sample $X^{(n)} = (X_1, \dots, X_n^{(n)})$ into $(l_n + 1)$ regenerative blocks $\mathbf{B}_0, \dots, \mathbf{B}_{l_n-1}, \mathbf{B}_{l_n}^{(n)} \in \mathbb{T}$, each block B_i with a length $l(\mathbf{B}_i) \equiv \tau_i$.
3. Drop the first and the last blocks if $\tau_{l_n} < n$ to avoid bias.
4. Let $\boldsymbol{\xi} = (\xi_{i,l_n}, i = 1, \dots, n)$ be a triangular array of random variables. Define the weighted bootstrap empirical measure from the data:

$$\mathbb{P}_n^* = \frac{1}{l_n} \sum_{i=1}^n \xi_{i,l_n} \delta_{\mathbf{B}_i}.$$

In what follow, we will denote by \mathbb{P}^* and \mathbb{E}^* respectively the conditional probability and the conditional expectation given the sample $\{X_1, \dots, X_n\}$. The same notation will

be used for the sample $\{\mathbf{B}_1, \dots, \mathbf{B}_{L_n}\}$. Define the bootstrapped U -statistic as

$$U_n^*(h) = \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_n^m} \xi_{i_1, n} \dots \xi_{i_m, n} h(X_{i_1}, \dots, X_{i_m}) \quad (3.9)$$

and the regenerative bootstrapping

$$R_{l_n}^*(h) = \binom{l_n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_{l_n}^m} \xi_{i_1, l_n} \dots \xi_{i_m, l_n} \omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}). \quad (3.10)$$

and the U -processes are:

$$\begin{aligned} Z_n^* &:= \binom{n}{m}^{1/2} [U_n^*(h) - U_n(h)]_{h \in \mathcal{F}} \\ &= \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in I_n^m} (\xi_{i_1, n} - 1) \dots (\xi_{i_m, n} - 1) h(X_{i_1}, \dots, X_{i_m}). \end{aligned} \quad (3.11)$$

and

$$\begin{aligned} T_{l_n}^* &:= \binom{l_n}{m}^{1/2} [R_{l_n}^*(h) - \mathbb{E}(R_{l_n}^*(h))]_{h \in \mathcal{F}} \\ &= \binom{l_n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in I_{l_n}^m} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}). \end{aligned} \quad (3.12)$$

Given a real-valued function Δ_n defined on the product probability space. We say that Δ_n is of an order $o_{\mathbb{P}_\xi^\circ}(1)$ in \mathbb{P}_ν° -probability if for any $\varepsilon, \delta > 0$

$$\mathbb{P}_\nu^\circ \left\{ \mathbb{P}_{\xi|X}^\circ (|\Delta_n| > \varepsilon) > \delta \right\} \longrightarrow 0 \quad \text{as } n \rightarrow \infty$$

and that Δ_n is of an order $O_{\mathbb{P}_\xi^\circ}(1)$ in \mathbb{P}_ν° -probability if for any $\delta > 0$, there exists a $0 < M < \infty$ such that

$$\mathbb{P}_\nu^\circ \left\{ \mathbb{P}_{\xi|X}^\circ (|\Delta_n| \geq M) > \delta \right\} \longrightarrow 0 \quad \text{as } n \rightarrow \infty$$

We must comment here that the bootstrap works in probability if

$$d_{BL}(T_{l_n}^*, T_{l_n}) \rightarrow 0 \quad \text{in probability,}$$

where

$$d_{BL}(T_{l_n}^*, T_{l_n}) = \sup_{g \in BL(l^\infty(\mathcal{F}))} \left| \mathbb{E}g(T_{l_n}^*)^\circ - \mathbb{E}g(T_{l_n}) \right|,$$

and

$$BL(l^\infty(\mathcal{F})) := \{g : l^\infty(\mathcal{F}) \rightarrow \mathbb{R}, |g(x) - g(y)| \leq \|x - y\|_{\mathcal{F}}, \|g\|_\infty \leq 1\},$$

and $g(T_{l_n}^*)^\circ$ the measurable envelop of $g(T_{l_n}^*)$. In addition, for any measurable random elements Y_n and Y , the convergence in law of Y_n to Y will be in the sense of Hoffman-Jorgensen, which is defined as

$$\mathbb{E}g(Y_n)^\circ \rightarrow \mathbb{E}g(Y),$$

for g bounded and continuous. This weak convergence is metrizable by Theorem 6.3 in the Appendix.

Proposition 3.4. *Suppose that the bootstrap weights (ξ_1, \dots, ξ_n) satisfy Assumptions (A1)-(A4). Let*

$$W_n^*(h) := U_n^*(h) - \binom{l_n - 1}{m} \binom{n}{m}^{-1} R_{l_n}^*(h). \quad (3.13)$$

Then we have

$$W_n^*(h) \rightarrow 0, \quad \mathbb{P}_\nu \times \mathbb{P}_\xi - a.s.$$

The proof of Proposition 3.4 is postponed until Section 6 Now, in the following lemma there is some instrumental results needed later.

Lemma 3.1. *Let $(X_n)_n$ be a Markov chain defined in 2.1. Define $p := \mathbb{P}(X_0 \in \mathbf{A}) = \alpha^{-1}$. Then, for any initial probability ν , we have:*

i) *For some $\eta > 0$ and $C > 0$:*

$$\left| \frac{\mathbb{E}_\nu(l_n)}{np} - 1 \right| \leq \frac{C}{n} \quad \text{and} \quad \sqrt{n} \left(\frac{l_n}{np} - 1 \right) \rightarrow N(0, \eta^2). \quad (3.14)$$

ii) $\frac{n^*}{n} \rightarrow 1$ *in $\mathbb{P}_\nu \times \mathbb{P}_\xi$ -probability .*

iii) *Let X_i be a sequence of random variables. If*

$$T_n = \frac{1}{n} \sum_{i=1}^n X_i \rightarrow C \quad a.s.$$

Then, for any integer t_n valued sequence of random variables,

$$\frac{1}{t_n} \sum_{i=1}^{t_n} X_i \rightarrow C \quad \text{in } \mathbb{P}_\nu\text{-probability.}$$

4 Weighted Bootstrap weak convergence

In this section, we will extend some existing results concerning the multiplier U -process to prove the bootstrap uniform weak convergence. Most of these results can be founded in Q. Han, 2022 generalizing the empirical process work of Præstgaard et al., 1993 in the iid setting. The weak convergence will be proved for degenerate U -processes, as we mentioned before, and under the weighted regenerative bootstrap schemes described in Algorithm 1. Before stating the weak convergence theorem, we recall the following important results. The next theorem, proved in Q. Han, 2022, is a sharp multiplier inequality, which is essential in the study of the multiplier U -process. These results are based on the decoupling symmetrized U -process, a basic framework of U -statistics. Radulović, 2004 solved these problems for the empirical process settings in the Markov setting (multinomial bootstrap) that we generalise to the U -process by considering more general weights, i.e., the exchangeable weighted bootstrap.

Theorem 4.1: Q. Han, 2022

Let (ξ_1, \dots, ξ_n) be a random vector independent of (Y_1, \dots, Y_n) . Then there exists some measurable function $\psi_n : \mathbb{R}_{\geq 0}^m \rightarrow \mathbb{R}_{\geq 0}$ such that the expected supremum of the decoupled ^a U -processes

$$\mathbb{E} \left\| \sum_{1 \leq i_k \leq \ell_k, 1 \leq k \leq m} \varepsilon_{i_1}^{(1)} \cdots \varepsilon_{i_m}^{(m)} f(Y_{i_1}^{(1)}, \dots, Y_{i_m}^{(m)}) \right\|_{\mathcal{F}} \leq \psi_n(\ell_1, \dots, \ell_m),$$

for all $1 \leq \ell_1, \dots, \ell_m \leq n$, consequently,

$$\begin{aligned} & \mathbb{E} \left\| \sum_{1 \leq i_1, \dots, i_m \leq \ell_{n-1}} \xi_{i_1} \cdots \xi_{i_m} f(Y_{i_1}^{(1)}, \dots, Y_{i_m}^{(m)}) \right\|_{\mathcal{F}} \\ & \leq K \int_{\mathbb{R}_{\geq 0}^m} \mathbb{E} \psi_n \left(\sum_{i=1}^{\ell_{n-1}} \mathbb{1}_{|\xi_i| > t_1}, \dots, \sum_{i=1}^{\ell_{n-1}} \mathbb{1}_{|\xi_i| > t_m} \right) dt_1 \cdots dt_m. \end{aligned}$$

Furthermore, if there exists a concave and non-decreasing function $\bar{\psi}_n : \mathbb{R} \rightarrow \mathbb{R}$ such that $\psi_n(\ell_1, \dots, \ell_m) = \bar{\psi}_n(\prod_{k=1}^m \ell_k)$, then

$$\begin{aligned} & \mathbb{E} \left\| \sum_{1 \leq i_1, \dots, i_m \leq n} \xi_{i_1} \cdots \xi_{i_m} f(Y_{i_1}^{(1)}, \dots, Y_{i_m}^{(m)}) \right\|_{\mathcal{F}} \\ & \leq K \int_{\mathbb{R}_{\geq 0}^m} \bar{\psi}_n \left(\sum_{1 \leq i_1, \dots, i_m \leq n} \prod_{k=1}^m \mathbb{P}(|\xi_{i_k}| > t_k)^{1/m} \right) dt_1 \cdots dt_m. \end{aligned}$$

Here $K > 0$ is a constant depending on m only, and can be taken as $K = 2^{2m} \prod_{k=2}^m (k^k - 1)$ for $m \geq 2$.

^aHere "decoupled" refers to fact that $\{Y_i^{(k)}\}, k \in \mathbb{N}$ are independent copies of $\{Y_i\}$, and $\{\varepsilon_i^{(k)}\}, k \in \mathbb{N}$ are independent copies of the Rademacher sequence $\{\varepsilon_i\}$.

Lemma 4.1 (Q. Han, 2022). Let $\{\mathcal{F}_{(\ell_1, \dots, \ell_m), n} : 1 \leq \ell_1, \dots, \ell_m \leq n, n \in \mathbb{N}\}$ be function classes such that $\mathcal{F}_{(\ell_1, \dots, \ell_m), n} \supset \mathcal{F}_{(n, \dots, n), n}$ for all $1 \leq \ell_1, \dots, \ell_m \leq n$. Suppose that ξ_i 's have the same marginal distributions with $\|\xi_1\|_{2m,1} < \infty$. Suppose that there exists some bounded measurable function $a : \mathbb{R}_{\geq 0}^{m(n)} \rightarrow \mathbb{R}_{\geq 0}$ with $a(\ell_1, \dots, \ell_m) \rightarrow 0$ as $\ell_1 \wedge \dots \wedge \ell_m \rightarrow \infty$, such that the expected supremum of the decoupled U -processes satisfies

$$\mathbb{E} \left\| \sum_{1 \leq i_k \leq \ell_k, 1 \leq k \leq m} \varepsilon_{i_1}^{(1)} \cdots \varepsilon_{i_m}^{(m)} f(Y_{i_1}^{(1)}, \dots, Y_{i_m}^{(m)}) \right\|_{\mathcal{F}_{(\ell_1, \dots, \ell_m), n}} \leq a(\ell_1, \dots, \ell_m) \left(\prod_{k=1}^m \ell_k \right)^{1/2},$$

for all $1 \leq \ell_1, \dots, \ell_m \leq n$. Then

$$n^{-m/2} \mathbb{E} \left\| \sum_{1 \leq i_1, \dots, i_m \leq n} \xi_{i_1} \cdots \xi_{i_m(n)} f(Y_{i_1}^{(1)}, \dots, Y_{i_m}^{(m)}) \right\|_{\mathcal{F}_{(n, \dots, n), n}} \rightarrow 0, \quad n \rightarrow \infty.$$

The main result of this chapter is represented in the following theorem. It is worth noting here that it is not easy to prove the stochastic equicontinuity in the present setting as explained in the introduction.

Theorem 4.2

Suppose Assumptions (A1) to (A4), and Conditions (C.1)-(C.5) hold. Let $\mathcal{F} \subset L_2^{c,m}(P)$ permissible⁴, admit a P^m -square integrable envelope F such that

$$\int_0^1 \left(\sup_Q \log \mathcal{N}(\varepsilon \|F\|_{L_2(Q)}, \mathcal{F}, L_2(Q)) \right)^{m/2} d\varepsilon < \infty,$$

where the supremum is taken over all discrete probability measures. Then

$$\sup_{\psi \in \text{BL}} \left| \mathbb{E}_\xi \psi \left(Z_n^*(h) \right) - \mathbb{E} \psi(c \cdot \mathbb{K}_P) \right| \rightarrow_{\mathbb{P}_\nu} 0,$$

where c is the constant in (A3), and the convergence in probability $\rightarrow_{\mathbb{P}_\nu}$ is with respect to the outer probability of \mathbb{P}^∞ defined on $(E^\infty, \mathcal{E}^\infty)$.

⁴[Permissible classes of function] Let $(E, \mathcal{E}, \mathbb{P})$ be a measurable space (\mathcal{E} a Borel σ -field on E). Let \mathcal{F} be a class of functions indexed by a parameter x that belongs to a set E . \mathcal{F} is called permissible if it can be indexed by a E such that:

- There exists a function $g(x, f) = f(x)$ defined from $S \times \mathcal{F}$ to \mathbb{R} in such a way that this function is $\mathcal{L} \otimes \mathcal{B}(\mathcal{F})$ measurable function, where $\mathcal{B}(\mathcal{F})$ is the Borel σ -algebra generated by the metric on \mathcal{F} .
- E is a Suslin measurable space that is mean E is an analytic subset of a compact metric space E from which it inherits its metric and Borel σ -field.

4.1 Bootstrap weights examples:

Let (ξ_1, \dots, ξ_n) be a class of real random variables satisfy Assumptions (A1)- (A4). We give some examples of bootstrap weights, for instance, refer Præstgaard et al., 1993 and Cheng, 2015 for more explanation.

Bayesian resampling scheme. In this case, for (ξ_1, \dots, ξ_n) are positive i.i.d. random variables with mean μ and finite variance σ^2 . The weights satisfy $\|\xi_1\|_{2,1} < \infty$, and define

$$\bar{\xi}_n = \sum_{i=1}^n \xi_i.$$

The Bayesian Bootstrapped weight can be defined as:

$$\xi_{ni} = \xi_i / \bar{\xi}_n,$$

satisfying

$$\|\xi_{n1}\|_{2,1} = \int_0^\infty \sqrt{\mathbb{P}_\xi(\xi_{n1} \geq u)} du.$$

For $\xi_{ni} \sim Exponential(1)$ or $\xi_{ni} \sim Gamma(4, 1)$, the Bayesian weights are distributional equivalent with Dirichlet weights. For the value of c^2 , we have:

$$\frac{1}{n} \sum_{i=1}^n (\xi_{ni} - 1)^2 \rightarrow \frac{\sigma^2}{\mu^2} := c^2, \quad n \rightarrow \infty.$$

Efron's resampling scheme. For Efron's bootstrap, we have

$$(\xi_1, \dots, \xi_n) \sim Multinomial(n; n^{-1}, \dots, n^{-1}).$$

Condition (A1) follows directly. Condition (A3) follows usind Mason et al., 1992, Lemma 4.1, and Condition (A2) is detailed in Hůsková et al., 1993.

The delete h -Jackknife . Shao et al., 1987 permute deterministic weights w_n , where,

$$w_n = \left\{ \frac{n}{n-h}, \dots, \frac{n}{n-h}, 0, \dots, 0 \right\} \quad \text{with} \quad \sum_{i=1}^n w_{ni} = n.$$

in order to build a new bootstrap weights, they defined the new weights $\xi_{nj} := w_{nR_n(j)}$ where $R_n(\cdot)$ is a random permutation uniformly distributed over $\{1, \dots, n\}$. These weights are called The delete h -Jackknife. In order to achieve Assumption (A3), we must assume that $h/n \rightarrow \alpha \in (0, 1)$, as $c^2 = h/(n-h)$ and $c > 0$.

The multivariate hypergeometric resampling scheme As its name indicates, the bootstrap weights of this scheme follow the multivariate hypergeometric distribution with density:

$$P(\xi_{n1} = \varepsilon_1, \dots, \xi_{nn} = \varepsilon_n) = \frac{\binom{K}{\varepsilon_1} \cdots \binom{K}{\varepsilon_n}}{\binom{nK}{n}},$$

where K is a positive integer. Assumption (A3) is satisfied with $c^2 = (K - 1)/K$.

Remark 4.1. As was pointed out in Præstgaard et al., 1993, the preceding mentioned bootstraps are "smoother" in some way than the multinomial bootstrap because they place some (random) weight on all elements in the sample, whereas the multinomial bootstrap applies the positive weight at about $1 - (1 - n^{-1})^n \rightarrow 1 - e^{-1} = 0.6322$ proportion of each element of the sample, on the average. Notice that when $\omega_i \sim \text{Gamma}(4, 1)$ so that the ξ_{ni}/n are equivalent to four-spacings from a sample of $4n - 1$ Uniform $(0, 1)$ random variables. In Weng, 1989 and van Zwet, 1979, it was noticed that, in addition to being four times more expensive to implement, the choice of four-spacings depends on the functional of interest and is not universal.

Remark 4.2. It is worth noticing that an appropriate choice of the bootstrap weights ξ_{ni} 's implies a smaller limit variance, that is, c^2 is smaller than 1. A typical example is the multivariate hypergeometric bootstrap, refer to Præstgaard et al., 1993, Example 3.4 and the Subsample Bootstrap, Pauly, 2012, Remark 2.2-(3). A detailed discussion about the choice of the weights is certainly out of the scope of the present chapter, we refer for review to Shao et al., 1995.

5 Applications

5.1 Symmetry test

This example gives application for the bootstrap U -statistics, inspired by the goodness of fit tests Fan et al., 1999. They have considered the symmetry test for the distribution of X_t . Let $(X_t)_{t \in \mathbb{N}}$ a stationary mixing process with $f_X(\cdot)$ the Lebesgue density. Testing the hypothesis:

$$\begin{cases} H_0 : f_X(u) = f_X(-u), & \text{almost every where,} \\ H_1 : f_X(u) \neq f_X(-u) & \text{on a set of positive measure.} \end{cases} \quad (5.1)$$

The estimator of $f_X(u)$ is:

$$\hat{f}_X(u) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{u - X_i}{h_n}\right),$$

where $K(\cdot)$ is a kernel function and $h_n > 0$ is a smoothing parameter or the bandwidth. An appropriate estimator of the integrated squared difference represent the symmetry test:

$$I = \int_{\mathbb{R}} (f_X(u) - f_X(-u))^2 du.$$

According to Fan et al., 1999, I can be estimated by

$$\hat{I}_n := \frac{4}{n^2 h_n} \sum_{1 \leq i < j \leq n} \Phi_n(X_i, X_j),$$

where

$$\Phi_n(X_i, X_j) = K_{X_i, X_j} - K_{X_i, -X_j}$$

with $K_{X_i, Y_j} = K\left(\frac{X_i - Y_j}{h_n}\right)$, for $Y_j \in \{X_j, -X_j\}$. Clearly, \hat{I}_n is a degenerate U -statistic with kernel varying with the sample size n . Thus, the stationary bootstrap test,

$$\hat{I}_n^* := \frac{4}{n^2 h_n} \sum_{1 \leq i < j \leq n} \Phi_n(X_i^*, X_j^*),$$

can be shown to have the same limiting as \hat{I}_n .

5.2 Kendall's tau

The covariance matrix quantifies the linear dependency in a random vector. The rank correlation is another measure of the nonlinear dependency in a random vector. 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 $T = \{\tau_{mk}\}_{m,k=1}^p$ is a matrix-valued U -statistic with a bounded kernel. It is clear 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.

5.3 Test of independence

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 A. J. 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

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

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

6 Mathematical development

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

Proof of Proposition 3.4

We have

$$\begin{aligned} U_n^*(h) &= \binom{l_n - 1}{m} \binom{n}{m}^{-1} R_{l_n}^*(h) \\ &= \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_n^m} \xi_{i_1, n^*} \dots \xi_{i_m, n^*} h(X_{i_1}, \dots, X_{i_m}) \\ &\quad - \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_n^m} \xi_{i_1, l_n} \dots \xi_{i_m, l_n} \omega_h^{\sim}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \end{aligned}$$

$$\begin{aligned}
 &= \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_n^m} \xi_{i_1, n^*} \dots \xi_{i_m, n^*} h(X_{i_1}, \dots, X_{i_m}) \\
 &\quad - \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_n^m} \xi_{i_1, l_n} \dots \xi_{i_m, l_n} \omega_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \\
 &\quad + \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_n^m} h(X_{i_1}, \dots, X_{i_m}) - \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_n^m} h(X_{i_1}, \dots, X_{i_m}) \\
 &\quad + \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_n^m} \{\xi_{i_1, l_n} \dots \xi_{i_m, l_n} - 1\} \omega_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \\
 &\quad - \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_n^m} \{\xi_{i_1, l_n} \dots \xi_{i_m, l_n} - 1\} \omega_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \\
 &= \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_n^m} \{\xi_{i_1, n^*} \dots \xi_{i_m, n^*} - 1\} h(X_{i_1}, \dots, X_{i_m}) \\
 &\quad - \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_n^m} \{\xi_{i_1, l_n} \dots \xi_{i_m, l_n} - 1\} \omega_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \\
 &\quad + \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_n^m} \omega_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \\
 &\quad - \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_n^m} h(X_{i_1}, \dots, X_{i_m}).
 \end{aligned}$$

Given $J \subseteq \{1, \dots, m\}$ ($J = \emptyset$ is not excluded), and $\mathbf{i} = (i_1, \dots, i_m) \in \{1, \dots, n\}^m$, we set \mathbf{i}_J to be the point of $\{1, \dots, n\}^{|J|}$ obtained from \mathbf{i} by deleting the coordinates in the places not in J (e.g., if $\mathbf{i} = (3, 4, 2, 1)$, then $\mathbf{i}_{\{1,3\}} = (3, 2)$). Also, \sum_{i_j} indicates the sum over $1 \leq i_j \leq n, j \in J$ (for instance, if $m = 4$ and $J = \{1, 3\}$, then

$$\sum_{\mathbf{i}_J} h_{\mathbf{i}} = \sum_{\mathbf{i}_{\{1,3\}}} h_{i_1, i_2, i_3, i_4} = \sum_{1 \leq i_1, i_3 \leq n} h_{i_1, i_2, i_3, i_4} (X_{i_1}^{(1)}, \dots, X_{i_4}^{(4)}).$$

By convention, $\sum_{\mathbf{i}} a = a$. Notice that

$$\begin{aligned}
 &\mathbb{E} \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_n^m} \{\xi_{i_1, n} \dots \xi_{i_m, n} - 1\} h(X_{i_1}, \dots, X_{i_m}) \\
 &= \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_n^m} \mathbb{E} \{\xi_{i_1, n} \dots \xi_{i_m, n} - 1\} \mathbb{E} h(X_{i_1}, \dots, X_{i_m}) \\
 &= \binom{n}{m}^{-1} \sum_{\mathbf{i}_{\{1, \dots, m-1\}}} \sum_{j=1}^m \mathbb{E} \left\{ \prod_{k=1, k \neq j}^m \xi_{i_k, n} \mathbb{E} \left\{ \sum_{i_j=1}^n (\xi_{i_j, n} - 1) \mid \prod_{k=1, k \neq j}^m \xi_{i_k, n} \right\} \right\} \\
 &\quad \times \mathbb{E} h(X_{i_1}, \dots, X_{i_m})
 \end{aligned}$$

$$= 0.$$

In a similar way, we have

$$\mathbb{E} \binom{n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_n^m} \{\xi_{i_1, l_n} \cdots \xi_{i_m, l_n} - 1\} \omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) = 0.$$

Making use of the Proposition 3.3 and the law of large numbers, we infer that

$$U_n^*(h) - \binom{l_n - 1}{m} \binom{n}{m}^{-1} R_{l_n}^*(h) = 0, \quad a.s.$$

Hence the proof is completed. □

Proof of Lemma 3.1

The proof of the part *i*) and part *iii*) follows from Radulović, 2004, Lemma 3.1 & Lemma 3.2. In order to prove *ii*), we need to show that, for every $\varepsilon > 0$:

$$\mathbb{P}_\nu \times \mathbb{P}_{\xi|X} \left(\left| \frac{n^*}{n} - 1 \right| > \varepsilon \right) \rightarrow 0, \quad (6.1)$$

which follows if, conditioned on the sample,

$$\mathbb{P}_{\xi|X} \left(\left| \frac{n^*}{n} - 1 \right| > \varepsilon \right) \rightarrow 0. \quad (6.2)$$

We have:

$$\begin{aligned} \frac{n^*}{n} - 1 &= \frac{\sum_{i=1}^{l_n} \xi_i \tau_i}{n} - 1 = \frac{l_n}{n} \left[\frac{\sum_{i=1}^{l_n} (\xi_i \tau_i + \xi_i \mathbb{E}^*(\tau) - \xi_i \mathbb{E}^*(\tau)) - n}{l_n} \right] \\ &= \frac{l_n}{n} \left[\frac{\sum_{i=1}^{l_n} \xi_i (\tau_i - \mathbb{E}^*(\tau))}{l_n} \right] + \frac{l_n}{n} \left[\frac{\sum_{i=1}^{l_n} \xi_i \mathbb{E}^*(\tau)}{l_n} - \frac{n}{l_n} \right] \\ &= I + II. \end{aligned} \quad (6.3)$$

We will denote the \mathbb{E}^* for the expectation conditionally on X_1, \dots, X_n . By the fact that τ_i are i.i.d. and using Chebyshev's inequality, we have:

$$\begin{aligned} \mathbb{P}_{\xi|X} (|I| > \varepsilon) &\leq \varepsilon^{-2} \left(\frac{l_n}{n} \right)^2 \frac{1}{l_n} \mathbb{E}_{\xi|X} (\xi_{1, l_n} (\tau_1 - \mathbb{E}^*(\tau)))^2 \\ &\leq 2\varepsilon^{-2} \left(\frac{l_n}{n} \right) \frac{1}{n} \mathbb{E}(\xi_{1, l_n}^2) \frac{1}{l_n} \sum_{i=1}^{l_n} \tau_i^2 \rightarrow 0 \quad \text{in probability.} \end{aligned}$$

The last inequality follows using *i*) which imply that $\frac{l_n}{n} \rightarrow p$, *iii*) where

$$\frac{1}{l_n} \sum_{i=1}^{l_n} \tau_i^2 \rightarrow \mathbb{E}(\tau^2),$$

for $\mathbb{E}(\xi_{1,l_n}^2) < \infty$. For *II* we have:

$$\begin{aligned} II &= \frac{l_n}{n} \left[\frac{\sum_{i=1}^{l_n} \xi_i \mathbb{E}^*(\tau)}{l_n} - \frac{n}{l_n} \right] = \frac{l_n}{n} \left[\mathbb{E}^*(\tau) - \frac{n}{l_n} \right] \quad \text{by (A1)} \\ &= \frac{l_n}{n} \left[\frac{1}{l_n} \sum_{i=1}^{l_n} (\tau_i - \mathbb{E}(\tau) + \mathbb{E}(\tau)) - \frac{n}{l_n} \right] \\ &= \frac{l_n}{n} \left[\frac{1}{l_n} \sum_{i=1}^{l_n} (\tau_i - \mathbb{E}(\tau)) \right] + \frac{l_n}{n} \left[\mathbb{E}(\tau) - \frac{n}{l_n} \right]. \end{aligned}$$

The last equality converges to zero by the fact that $n/l_n \rightarrow \alpha = \mathbb{E}(\tau)$ and by *iii*)

$$\frac{1}{l_n} \sum_{i=1}^{l_n} (\tau_i - \mathbb{E}(\tau)) \rightarrow 0.$$

This proves Lemma 3.1. □

Proof of Theorem 4.2

For the weak convergence, we need to show the finite-dimensional convergence and the asymptotic equicontinuity. According to Proposition 3.4 and de la Peña et al., 1999, the finite-dimensional convergence will be considered if, for every fixed finite collection of functions $\{f_1, \dots, f_k\} \subset \mathcal{F}$,

$$\left(\binom{n}{m_1}^{-1/2} R_{l_n}^*(f_1), \dots, \binom{n}{m_k}^{-1/2} R_{l_n}^*(f_k) \right) \rightarrow \left(\mathbb{K}_{\mathbb{P}}(f_1), \dots, \mathbb{K}_{\mathbb{P}}(f_k) \right),$$

where $\mathbb{K}_{\mathbb{P}}$ is the Gaussian chaos process. According to Cramér-Wold and countability of \mathcal{F} , we only need to show that for any $f \in L_2^{c,m}(\mathbb{P})$,

$$\sup_{\psi \in \text{BL}} \left| \mathbb{E} \left[\psi \left(\binom{n}{m}^{-1/2} R_{l_n}^*(f) \right) \middle| \{\mathbf{B}_i\} \right] - \mathbb{E} \psi(c \cdot \mathbb{K}_{\mathbb{P}}(f)) \right| \rightarrow 0 \text{ a.s.} \quad (6.4)$$

By de la Peña et al., 1999, Section 4.2 and Arcones et al., 1992, Section 2A, any $f \in L_2^{c,m}(\mathbb{P})$ can be expanded in $L_2(\mathbb{P}^m)$ by $f = \sum_{q=1}^{\infty} c_q h_m^{\psi_q}$, where $\{c_q\}$ is a sequence of real numbers, and

$$h_m^{\psi_q}(x_1, \dots, x_m) \equiv \psi_q(x_1) \cdots \psi_q(x_m)$$

for some bounded $\psi_q \in L_2^{c,1}(\mathbb{P})$. Fix $\varepsilon > 0$. Then there exists $Q_\varepsilon \in \mathbb{N}$ such that with $f_n^\varepsilon \equiv \sum_{q=1}^{Q_\varepsilon} c_q h_m^{\psi_q}$,

$$\|f - f_\varepsilon\|_{L_2(P^m)} \leq \varepsilon.$$

The left hand side of (6.4) can be further bounded by

$$\begin{aligned} & \sup_{\psi \in \text{BL}} \left| \mathbb{E} \left[\psi \left(\binom{n}{m}^{-1/2} R_{l_n}^*(f) \right) \middle| \{\mathbf{B}_i\} \right] - \mathbb{E} \psi(c \cdot \mathbb{K}_{\mathbb{P}}(f)) \right| \\ & \leq \sup_{\psi \in \text{BL}} \left| \mathbb{E} \left[\psi \left(\binom{n}{m}^{-1/2} R_{l_n}^*(f) \right) \middle| \{\mathbf{B}_i\} \right] - \mathbb{E} \left[\psi \left(\binom{n}{m}^{-1/2} R_{l_n}^*(f^\varepsilon) \right) \middle| \{\mathbf{B}_i\} \right] \right| \\ & \quad + \sup_{\psi \in \text{BL}} \left| \mathbb{E} \left[\psi \left(\binom{n}{m}^{-1/2} R_{l_n}^*(f^\varepsilon) \right) \middle| \{\mathbf{B}_i\} \right] - \mathbb{E} \psi(c \cdot \mathbb{K}_{\mathbb{P}}(f^\varepsilon)) \right| \\ & \quad + \sup_{\psi \in \text{BL}} \left| \mathbb{E} \psi(c \cdot \mathbb{K}_{\mathbb{P}}(f^\varepsilon)) - \mathbb{E} \psi(c \cdot \mathbb{K}_{\mathbb{P}}(f)) \right| \\ & \equiv (I) + (II) + (III). \end{aligned} \tag{6.5}$$

Let $\bar{f}^\varepsilon \equiv f - f^\varepsilon$, and noting that that ψ is bounded by 1, and using Lemma 3.1, we can replace l_n by the $\varphi(n) = \lfloor \frac{n}{\mathbb{E}_A(\tau)} \rfloor$ which is deterministic. In the following, we will denote by π a random permutation uniformly distributed over $\Sigma(n)$, the set of all permutations over $1, \dots, n$. We have

$$\begin{aligned} (I)^2 & \leq \mathbb{E}^* \left| 2 \wedge \binom{n}{m}^{1/2} R_{l_n}^*(\bar{f}^\varepsilon) \right|^2 \\ & \lesssim \mathbb{E}_{\xi|X} \mathbb{E}_R \left(1 \wedge n^{-m/2} \sum_{1 \leq i_1 \neq \dots \neq i_m \leq \varphi(n)} (\xi_{\pi_{i_1}} - 1) \cdots (\xi_{\pi_{i_m}} - 1) \bar{f}^\varepsilon(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \right)^2 \\ & \lesssim \sum_{\alpha_i \in \{1,2\}; \sum_{i=1}^l \alpha_i = 2m, \alpha_1 \geq \dots \geq \alpha_l, 1 \leq l \leq m} \mathbb{E}_{\xi}^* \left[1 \wedge n^{-m/2} \mathbb{E}_R \left[\prod_{i=1}^l (\xi_{\pi_i} - 1)^{\alpha_i} \right] \right. \\ & \quad \left. \times \sum_{\substack{i_1 \neq \dots \neq i_m, \\ i'_1 \neq \dots \neq i'_m, \\ i_j = i'_j, 1 \leq j \leq \max\{j: \alpha_j = 2\}}} \bar{f}^\varepsilon(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m(n)}) \bar{f}^\varepsilon(\mathbf{B}_{i'_1}, \dots, \mathbf{B}_{i'_m(n)}) \right] \\ & \lesssim \sum_{\substack{\alpha_i \in \{1,2\}; \sum_{i=1}^l \alpha_i = 2m, \\ \alpha_1 \geq \dots \geq \alpha_l, 1 \leq l \leq m}} \mathbb{E} \left[1 \wedge \frac{1}{n} \sum_{i=1}^n (\xi_i - 1)^2 \right]^m \\ & \quad \times n^{-\ell} \sum_{\substack{i_1 \neq \dots \neq i_m, \\ i'_1 \neq \dots \neq i'_m, \\ i_j = i'_j, 1 \leq j \leq \max\{j: \alpha_j = 2\}}} \bar{f}^\varepsilon(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \bar{f}^\varepsilon(\mathbf{B}_{i'_1}, \dots, \mathbf{B}_{i'_m}). \end{aligned}$$

We have, according to Hůsková et al., 1993, for (ξ_1, \dots, ξ_n) a non-negative sequence of variables such that $\sum_{i=1}^n \xi_i = n$. For $\pi = (\pi_1, \dots, \pi_n)$ be a random permutation of

$\{1, \dots, n\}$. Then for any $\ell \in \mathbb{N}$ and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_\ell) \in \mathbb{N}^\ell$,

$$\left| \mathbb{E}_\pi \left[\prod_{i=1}^{\ell} (\xi_{\pi_i} - 1)^{\alpha_i} \right] \right| \leq C_{l,\boldsymbol{\alpha}} n^{-\ell} \left[\sum_{i=1}^{l_n} (\xi_i - 1)^2 \right]^{\sum_i \alpha_i / 2}.$$

And according to Frees, 1989 and Rempala et al., 1999, we have :

$$n^{-\ell} \sum_{\substack{i_1 \neq \dots \neq i_m, \\ i'_1 \neq \dots \neq i'_m, \\ i_j = i'_j, 1 \leq j \leq \max\{j : \alpha_j = 2\}}} \bar{f}^\varepsilon(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \bar{f}^\varepsilon(\mathbf{B}_{i'_1}, \dots, \mathbf{B}_{i'_m})$$

$$\rightarrow_{\text{a.s.}} \mathbb{E}_A(\boldsymbol{\tau})^{-\ell} \mathbb{E} \bar{f}^\varepsilon(\mathbf{B}_1, \dots, \mathbf{B}_m) \bar{f}^\varepsilon(\mathbf{B}'_1, \dots, \mathbf{B}'_m)$$

$$\text{(where } \mathbf{B}_j = \mathbf{B}'_j \text{ for } 1 \leq j \leq \max\{j : \alpha_j = 2\} \text{ and for } l_n/n \rightarrow \mathbb{E}_A(\boldsymbol{\tau})^{-1})$$

$$\leq \mathbb{E}_A(\boldsymbol{\tau})^{-\ell} \mathbb{P}^m \bar{f}^{\varepsilon^2} \leq \varepsilon^2. \quad \text{(Under Condition (C.1) and (C.3).)}$$

Hence we have

$$\limsup_{n \rightarrow \infty} (I) \lesssim_{m,\xi} \varepsilon, \quad \text{a.s.} \quad (6.6)$$

Now for the second term, we have:

$$\begin{aligned} & \binom{n}{m}^{-1/2} R_{l_n}^*(f^\varepsilon) \\ &= \frac{1}{\binom{n}{m}^{1/2}} \sum_{q=1}^{Q_\varepsilon} c_q \sum_{1 \leq i_1 < \dots < i_m \leq \varphi(n)} (\xi_{\pi_{i_1}} - 1) \cdots (\xi_{\pi_{i_m}} - 1) \psi_q(\mathbf{B}_{i_1}) \cdots \psi_q(\mathbf{B}_{i_m}) \\ &= \frac{\varphi(n)^{m/2}}{\binom{n}{m}^{1/2}} \sum_{q=1}^{Q_\varepsilon} c_q \mathfrak{R}_m \left(\frac{1}{\varphi(n)^{1/2}} \sum_{i=1}^{\varphi(n)} (\xi_{\pi_i} - 1) \psi_q(\mathbf{B}_i), \dots, \frac{1}{\varphi(n)^{m/2}} \sum_{i=1}^{\varphi(n)} (\xi_{\pi_i} - 1)^m \psi_q^m(\mathbf{B}_i) \right) \\ &\equiv (1 + o(1))(m!)^{1/2} \mathbb{E}_A(\boldsymbol{\tau})^{-m/2} \sum_{q=1}^{Q_\varepsilon} c_q \mathfrak{R}_m(A_{\varphi(n),q}^{(1)}, \dots, A_{\varphi(n),q}^{(m)}), \end{aligned}$$

where \mathfrak{R}_m is the polynomial of degree m (cf. de la Peña et al., 1999, pp. 175):

$$\sum_{1 \leq i_1 < \dots < i_m \leq \varphi(n)} t_{i_1} \cdots t_{i_m} = \mathfrak{R}_m \left(\sum_{i=1}^{\varphi(n)} t_i, \sum_{i=1}^{\varphi(n)} t_i^2, \dots, \sum_{i=1}^{\varphi(n)} t_i^m \right). \quad (6.7)$$

As we mentioned before, this polynomial follows from Newton's inequality and allows us to show a polynomial function as a sum of monomials. All we need now is to check each argument of this polynomial function.

For $\ell = 1$: We first recall the following lemma from A. W. van der Vaart et al., 1996.

Lemma 6.1 (A. W. van der Vaart et al., 1996). *Let (a_1, \dots, a_n) be a vector and (ξ_1, \dots, ξ_n) be a vector of exchangeable random variables. Suppose that*

$$\bar{a}_n = \frac{1}{n} \sum_{i=1}^n a_i = 0, \quad \frac{1}{n} \sum_{i=1}^n a_i^2 \rightarrow \sigma^2, \quad \lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n a_i^2 \mathbf{1}_{\{|a_i| > M\}} = 0,$$

and

$$\bar{\xi}_n = \frac{1}{n} \sum_{i=1}^n \xi_i = 0, \quad \frac{1}{n} \sum_{i=1}^n \xi_i^2 \rightarrow_{\mathbb{P}_\xi} \alpha^2, \quad \frac{1}{n} \max_{1 \leq i \leq n} \xi_i^2 \rightarrow_{\mathbb{P}_\xi} 0.$$

Then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n a_i \xi_i \rightarrow \mathcal{N}(0, \sigma^2 \alpha^2).$$

Applying Lemma 6.1 with $a_i \equiv \psi_q(\mathbf{B}_i) - \mathbb{P}_n \psi_q$ and ξ_i replaced by $\xi_{R_i} - 1$, we can see that

$$A_{\varphi(n), q}^{(1)} \rightarrow c \cdot \mathbb{G}_{\mathbb{P}}(\psi_q), \quad \text{a.s.},$$

where $\mathbb{G}_{\mathbb{P}}$ is a Gaussian process defined on $L_2^{c,1}(\mathbb{P})$ with co-variance

$$\mathbb{E} \mathbb{G}_{\mathbb{P}}(f) \mathbb{G}_{\mathbb{P}}(g) = \mathbb{P}(fg), \quad \text{for } f, g \in L_2^{c,1}(\mathbb{P}).$$

For $\ell = 2$: Note that

$$\begin{aligned} \mathbb{E}_\pi^{*, \xi}(A_{\varphi(n), q}^{(2)}) &= \frac{1}{\varphi(n)} \sum_{i=1}^{\varphi(n)} (\xi_i - 1)^2 \cdot \frac{1}{\varphi(n)} \sum_{i=1}^{\varphi(n)} \psi_q^2(\mathbf{B}_i) \\ &\rightarrow_{\mathbb{P}_{\nu, \xi}} c^2 \mathbb{E} \psi_q^2(\mathbf{B}_1) = c^2 \mathbb{E}_A \left[\left(\sum_{i=\mathcal{T}_0+1}^{\mathcal{T}_1} h_1(X_i) \right)^2 \right], \quad \text{a.s.} \end{aligned}$$

Furthermore,

$$\begin{aligned} &\text{Var}^{*, \xi}(A_{\varphi(n), q}^{(2)}) \\ &= \mathbb{E}^{*, \xi} \left(A_{\varphi(n), q}^{(2)} \right)^2 - \left(\mathbb{E}_\pi^{*, \xi} \left(A_{\varphi(n), q}^{(2)} \right) \right)^2 \\ &= \mathbb{E}_\pi^{*, \xi} \left[\frac{1}{\varphi(n)} \sum_{i=1}^{\varphi(n)} (\xi_i - 1)^2 \psi_q^2(\mathbf{B}_{\pi_i}) \right]^2 - \left[\frac{1}{\varphi(n)} \sum_{i=1}^{\varphi(n)} (\xi_i - 1)^2 \mathbb{P}_n \psi_q^2 \right]^2 \\ &= \frac{1}{\varphi(n)^2} \sum_{i, j} (\xi_i - 1)^2 (\xi_j - 1)^2 \left[\mathbb{E}_\pi^* \psi_q^2(\mathbf{B}_{\pi_i}) \psi_q^2(\mathbf{B}_{\pi_j}) - (\mathbb{P}_n \psi_q^2)^2 \right] \\ &= \frac{1}{\varphi(n)^2} \sum_i (\xi_i - 1)^4 \left[\mathbb{E}_\pi^* \psi_q^4(\mathbf{B}_{\pi_i}) - (\mathbb{P}_n \psi_q^2)^2 \right] \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{\varphi(n)^2} \sum_{i \neq j} (\xi_i - 1)^2 (\xi_j - 1)^2 \left[\mathbb{E}_\pi^* \psi_q^2(\mathbf{B}_{\pi_i}) \psi_q^2(\mathbf{B}_{\pi_j}) - (\mathbb{P}_n \psi_q^2)^2 \right] \\
 & \leq \frac{1}{\varphi(n)^2} \sum_i (\xi_i - 1)^4 \cdot \mathbb{P}_n \psi_q^4 + \frac{1}{\varphi(n)^2} \left(\sum_i (\xi_i - 1)^2 \right)^2 \cdot \frac{1}{\varphi(n) - 1} \mathbb{P}_n \psi_q^4 \\
 & \leq C \frac{1}{\varphi(n)^2} \sum_{i=1}^n (\xi_i - 1)^4 \cdot \mathbb{P}_n \psi_q^4 \\
 & \leq C \|\psi_q\|_\infty^4 \frac{\max_i (\xi_i - 1)^2}{\varphi(n)} \cdot \frac{1}{\varphi(n)} \sum_{i=1}^{\varphi(n)} (\xi_i - 1)^2 \xrightarrow{\mathbb{P}_\xi} 0, \text{ a.s.}
 \end{aligned}$$

The first inequality in the above display follows since

$$\begin{aligned}
 & \mathbb{E}_\pi^* \psi_q^2(\mathbf{B}_{\pi_i} \mathbf{B}_{\pi_i}) \psi_q^2(\mathbf{B}_{\pi_j}) - (\mathbb{P}_n \psi_q^2)^2 \\
 & = \frac{1}{\varphi(n)(\varphi(n) - 1)} \left[\sum_{i \neq j} \psi_q^2(\mathbf{B}_{\pi_i}) \psi_q^2(\mathbf{B}_{\pi_j}) \right] - (\mathbb{P}_n \psi_q^2)^2 \\
 & \leq \frac{1}{\varphi(n) - 1} (\mathbb{P}_n \psi_q^2)^2 \leq \frac{1}{\varphi(n) - 1} \mathbb{P}_n \psi_q^4.
 \end{aligned}$$

This shows that

$$A_{\varphi(n),q}^{(2)} \xrightarrow{\mathbb{P}_\xi} c^2 \mathbb{E} \psi_q^2 \text{ a.s.}$$

For $\ell \geq 3$:

$$\begin{aligned}
 \mathbb{E}_\pi^{*,\xi} |A_{\varphi(n),q}^{(\ell)}| & \leq \frac{1}{\varphi(n)^{\ell/2}} \sum_{i=1}^{\varphi(n)} |\xi_i - 1|^\ell \cdot \frac{1}{\varphi(n)} \sum_{i=1}^{\varphi(n)} |\psi_q(\mathbf{B}_i)|^\ell \\
 & \leq \left(\frac{\max_i |\xi_i - 1|^2}{\varphi(n)} \right)^{\frac{\ell-2}{2}} \cdot \frac{1}{\varphi(n)} \sum_{i=1}^{\varphi(n)} |\xi_i - 1|^2 \cdot \|\psi_q\|_\infty \\
 & \xrightarrow{\mathbb{P}_\xi} 0, \text{ a.s.}
 \end{aligned}$$

This shows that

$$A_{\varphi(n),q}^{(\ell)} \xrightarrow{\mathbb{P}_\xi} 0, \text{ a.s.}$$

Then we have

$$R_m(A_{\varphi(n),q}^{(1)}, \dots, A_{\varphi(n),q}^{(m)}) \rightarrow R_m(\mathbb{G}_\mathbb{P}(c\psi_q), \mathbb{E}(c\psi_q)^2, 0, \dots, 0) = c(m!)^{-1/2} \cdot \mathbb{K}_\mathbb{P}(\psi_q) \text{ a.s.},$$

where $\mathbb{K}_\mathbb{P}$ be the Gaussian chaos process defined on⁵

$$\mathbb{R} \oplus L_2^{c,\mathbb{N}}(\mathbb{P}) \equiv \mathbb{R} \oplus \left(\bigoplus_{m=1}^{\infty} L_2^{c,m}(\mathbb{P}) \right).$$

⁵ \oplus is the orthogonal sum in $L_2(E^\infty, \mathcal{E}^\infty, P^\infty)$

Hence, it follows that, by linearity of $\mathbb{K}_{\mathbb{P}}$,

$$\binom{n}{m}^{-1/2} R_{l_n}^*(\bar{f}^\varepsilon) \rightarrow c \cdot \mathbb{K}_{\mathbb{P}}(f^\varepsilon), \quad \text{a.s.}$$

The last term in (6.5) follows from the definition of $\mathbb{K}_{\mathbb{P}}$

$$(III) \leq c\sqrt{\mathbb{E}\mathbb{K}_{\mathbb{P}}^2(\bar{f}^\varepsilon)} \rightarrow 0 \quad (\varepsilon \rightarrow 0). \quad (6.8)$$

All these final results give finite-dimensional convergence.

Now, we will take a step-by-step approach to establish stochastic equicontinuity. We assume that the class of functions must be bounded, so we will suppose that $h \leq H$, for H an envelop. Throughout the sequel, we will denote by

$$\mathcal{F}_\delta := \{f, g \in \mathcal{F} : d(f, g) \leq \delta\}.$$

Step 1

Let

$$Z_n^* := \binom{n^*}{m}^{1/2} [U_n^*(h) - \mathbb{E}^*(U_n^*(h))], \quad (6.9)$$

and

$$\check{T}_{l_n}^* := \binom{n}{m}^{-1/2} \binom{l_n}{m} [R_{l_n}^* - \mathbb{E}^*(R_{l_n}^*)]. \quad (6.10)$$

In this step, we must prove that stochastic equicontinuity of U -process implies the one of regenerative U -process. This is a consequence of 3.3, and for the weighted bootstrap the Proposition 3.4 and part ii) of Lemma 3.1.

Step 2

Define

$$\check{T}_{l_n}^* := \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in I_{l_n}^m} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m})$$

and

$$\tilde{T}_{l_n}^* = \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(l_n)}^m} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}).$$

Hypothesis: Stochastic equicontinuity of $\check{T}_{l_n}^*$ implies stochastic equicontinuity of $\tilde{T}_{l_n}^*$.

Proof. In order to prove the previous implication, we only need to show that:

$$\begin{aligned} & \mathbb{P}^* \left(\left\| \tilde{T}_{l_n}^* - \check{T}_{l_n}^* \right\|_{\mathcal{F}_\delta} > \varepsilon \right) \\ & \leq \mathbb{P}^* \left(\left\| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(l_n)}^m} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \right\|_{\mathcal{F}_\delta} > \varepsilon \right). \end{aligned}$$

Suppose that $l_n \leq \mathbb{E}(l_n)$, the opposite case can be treated in a similar way. We have

$$\begin{aligned} & \mathbb{P}^* \left(\left\| \tilde{T}_{l_n}^* - \check{T}_{l_n}^* \right\|_{\mathcal{F}_\delta} > \varepsilon \right) \\ & = \mathbb{P}^* \left(\left\| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(l_n)}^m} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \right. \right. \\ & \quad \left. \left. - \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in I_{l_n}^m} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \right\|_{\mathcal{F}_\delta} > \varepsilon \right) \end{aligned}$$

Define $\mathfrak{J} := \{(i_1, \dots, i_m) : 1 \leq i_1 < \dots < i_m \leq \mathbb{E}(l_n) : i_j \neq i_k \text{ for } j \neq k, \text{ such that } \exists \ell \in \{1, \dots, m\} : l_n \leq i_\ell \leq \mathbb{E}(l_n)\}$

$$\begin{aligned} & \leq \mathbb{P}^* \left(\left\| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in \mathfrak{J}} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \right\|_{\mathcal{F}_\delta} > \varepsilon \right) \\ & = \mathbb{P}^* \left(\left\| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in \mathfrak{J}} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) > \varepsilon \right. \right. \\ & \quad \left. \left. \cap (|\mathbb{E}(l_n) - l_n| \leq n/4) \right\|_{\mathcal{F}_\delta} \right) + \mathbb{P}^* (|\mathbb{E}(l_n) - l_n| > n/4). \end{aligned}$$

However, $|\mathbb{E}(l_n) - l_n| = O_{\mathbb{P}}(\sqrt{n})$ by Lemma 3.1 part i). Then there exists a constant $K > 0$, such that for every $\varepsilon > 0$,

$$\mathbb{P}^* (|\mathbb{E}(l_n) - l_n| > n/4) < \varepsilon,$$

and the first expression in the previous expression will be bounded by:

$$\mathbb{P}^* \left(\max_{M \leq n/2 + \mathbb{E}(l_n)} \left\| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in \mathfrak{J}'} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \right\|_{\mathcal{F}_\delta} > \varepsilon \right)$$

where $\mathfrak{J}' := \{(i_1, \dots, i_m) : 1 \leq i_1 < \dots < i_m \leq \mathbb{E}(l_n), \exists \ell = 1, \dots, m, \mathbb{E}(l_n) < i_\ell \leq M, i_j \neq i_k \text{ for } j \neq k\}$

$$\leq C_1 \mathbb{P}^* \left(\left\| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in \mathcal{J}''} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \right\|_{\mathcal{F}_\delta} > C_2 \varepsilon \right)$$

where $\mathcal{J}''_m := \{(i_1, \dots, i_m) : 1 \leq i_1 < \dots < i_m \leq \mathbb{E}(l_n), \exists \ell = 1, \dots, m, \mathbb{E}(l_n) < i_\ell \leq \mathbb{E}(l_n) + n/2, i_j \neq i_k \text{ for } j \neq k\}$.

The last expression follows by Montgomery-Smith inequality. Since

$$\mathbb{E}(l_n)/n \rightarrow \alpha^{-1},$$

the last expression matches the stochastic equicontinuity condition for $\tilde{T}_{l_n}^*$. This proves this step. \square

Before passing to the next step, we will introduce a new bootstrap sample. Define $\hat{\mathbf{B}}_i := \{X_{\mathcal{J}_{i-1+1}}, \dots, X_{\mathcal{J}_i}\}$ for $i = 1, \dots, \mathbb{E}(l_n)$. Now, apply the weighted bootstrap procedure on the sample $\{\hat{\mathbf{B}}_i\}_{i=1}^{\mathbb{E}(l_n)}$. This new procedure is the same as the old one for \mathbf{B}_i , but we aim here to replace the random quantity l_n with a deterministic one which is $\mathbb{E}(l_n)$.

Step 3

Define:

$$\hat{T}_{l_n}^* = \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(l_n)}^m} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_{\tilde{h}}(\hat{\mathbf{B}}_{i_1}, \dots, \hat{\mathbf{B}}_{i_m})$$

Hypothesis: Stochastic equicontinuity of $\tilde{T}_{l_n}^*$ implies stochastic equicontinuity of $\hat{T}_{l_n}^*$.

Proof. First case: $l_n \leq \mathbb{E}(l_n)$:

In this case, all of the terms in the following computation should be multiplied with $\mathbb{1}_{(l_n \leq \mathbb{E}(l_n))}$. We'll leave it out to keep the already complex notation simple. Define

$$\begin{aligned} \mathcal{A}_n &:= \{\mathbf{B}_1, \dots, \mathbf{B}_{l_n}\} \\ \vec{T}_{l_n}^* &:= \hat{T}_{l_n}^* \mathbb{1}_{(\hat{\mathbf{B}}_{i_1}, \dots, \hat{\mathbf{B}}_{i_m}) \in \mathcal{A}_n} + T_{l_n}^* \mathbb{1}_{(\hat{\mathbf{B}}_{i_1}, \dots, \hat{\mathbf{B}}_{i_m}) \in \mathcal{A}_n^c}. \end{aligned}$$

$\vec{T}_{l_n}^*$ is well defined, i.i.d, and have the same distribution as $T_{l_n}^*$ and $(i_1, \dots, i_m) \in I_{\mathbb{E}(l_n)}^m$. Hence, if we show that:

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P}^* \left(\left\| \vec{T}_{l_n}^* \right\|_{\mathcal{F}_\delta} > \varepsilon \right) = 0 \text{ in probability,}$$

then the stochastic equicontinuity of $\tilde{T}_{l_n}^*$ is established. But we aim to approximate the one of $\hat{T}_{l_n}^*$. In order to achieve our goal, it is sufficient to estimate:

$$\begin{aligned}
 & \left\| \vec{T}_{l_n}^* - \hat{T}_{l_n}^* \right\|_{\mathcal{F}_\delta} \\
 &= \left\| \hat{T}_{l_n}^* \mathbf{1}_{(\hat{\mathbf{B}}_{i_1}, \dots, \hat{\mathbf{B}}_{i_m}) \in \mathcal{A}_n} + T_{l_n}^* \mathbf{1}_{(\hat{\mathbf{B}}_{i_1}, \dots, \hat{\mathbf{B}}_{i_m}) \in \mathcal{A}_n^c} - \left[\hat{T}_{l_n}^* \mathbf{1}_{(\hat{\mathbf{B}}_{i_1}, \dots, \hat{\mathbf{B}}_{i_m}) \in \mathcal{A}_n} + \hat{T}_{l_n}^* \mathbf{1}_{(\hat{\mathbf{B}}_{i_1}, \dots, \hat{\mathbf{B}}_{i_m}) \in \mathcal{A}_n^c} \right] \right\|_{\mathcal{F}_\delta} \\
 &\leq \left\| T_{l_n}^* \mathbf{1}_{(\hat{\mathbf{B}}_{i_1}, \dots, \hat{\mathbf{B}}_{i_m}) \in \mathcal{A}_n^c} \right\|_{\mathcal{F}_\delta} + \left\| \hat{T}_{l_n}^* \mathbf{1}_{(\hat{\mathbf{B}}_{i_1}, \dots, \hat{\mathbf{B}}_{i_m}) \in \mathcal{A}_n^c} \right\|_{\mathcal{F}_\delta} \\
 &:= I_n + II_n.
 \end{aligned} \tag{6.11}$$

For I_n : Let

$$S_n^* := \sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(l_n)}^m} \mathbf{1}_{(\hat{\mathbf{B}}_{i_1}, \dots, \hat{\mathbf{B}}_{i_m}) \in \mathcal{A}_n^c},$$

conditioned on the sample, we have:

$$\begin{aligned}
 & \mathcal{L} \left(\sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(l_n)}^m} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \mathbf{1}_{(\hat{\mathbf{B}}_{i_1}, \dots, \hat{\mathbf{B}}_{i_m}) \in \mathcal{A}_n^c} \right) \\
 &= \mathcal{L} \left(\sum_{(i_1, \dots, i_m) \in I_{S_n^*}^m} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \right).
 \end{aligned}$$

Hence,

$$\begin{aligned}
 & \mathbb{P}^*(I_n > \varepsilon) \\
 &= \mathbb{P}^* \left(\binom{n}{m}^{-1/2} \left\| \sum_{(i_1, \dots, i_m) \in I_{S_n^*}^m} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \right\|_{\mathcal{F}_\delta} > \varepsilon \right) \\
 &= \mathbb{P}^* \left(\left\| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in I_{S_n^*}^m} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) > \varepsilon \right. \right. \\
 & \quad \left. \left. \cap (S_n^* \leq K\sqrt{n}) \right\|_{\mathcal{F}_\delta} \right) + \mathbb{P}^*(S_n^* > K\sqrt{n}) \\
 &\leq \mathbb{P}^* \left(\max_{M \leq K\sqrt{n}} \left\| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in \mathcal{J}'} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \right\|_{\mathcal{F}_\delta} > \varepsilon \right) \\
 & \quad + \mathbb{P}^*(S_n^* > K\sqrt{n}), \text{ for any } K > 0.
 \end{aligned}$$

where $\mathcal{J}' := \{(i_1, \dots, i_m) : 1 \leq i_1 < \dots < i_m \leq S_n^*, \exists \ell = 1, \dots, m, S_n^* < i_\ell \leq M, i_j \neq i_k \text{ for } j \neq k\}$.

$$\leq C_1 \mathbb{P}^* \left(\left\| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in \mathfrak{J}_m''} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_h(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \right\|_{\mathcal{F}_\delta} > C_2 \varepsilon \right) + \mathbb{P}^* \left(S_n^* > K\sqrt{n} \right), \text{ for any } K > 0, \quad (6.12)$$

where

$$\mathfrak{J}_m'' := \{(i_1, \dots, i_m) : 1 \leq i_1 < \dots < i_m \leq S_n^*, \exists \ell = 1, \dots, m, S_n^* < i_\ell \leq K\sqrt{n}, i_j \neq i_k \text{ for } j \neq k\}.$$

For n large enough, we need to show that there exists a $K > 0$ such that

$$\mathbb{P}^* \left(S_n^* > K\sqrt{n} \right) \rightarrow 0.$$

As $\mathbb{1}_{\widehat{\mathbf{B}}_i \in \mathcal{A}_n^c}$ are i.i.d and bounded,

$$\frac{S_n^* - \mathbb{E}(S_n^*)}{\sqrt{\mathbb{E}(l_n)}} \rightarrow N(0, \eta^2) \text{ in probability.}$$

therefore, we can find $\mathcal{M} > 0$ such that

$$\mathbb{P}^* \left(S_n^* > \mathbb{E}(S_n^*) + \mathcal{M}\sqrt{n} \right) < \varepsilon.$$

But,

$$\mathbb{E}(S_n^*) = \mathbb{E}(l_n) \mathbb{P}^* \left(\widehat{\mathbf{B}}_i^* \in \mathcal{A}_n^c \right) = \mathbb{E}(l_n) - l_n = O_{\mathbb{P}}(\sqrt{n}),$$

by Lemma 3.1 i), then

$$\mathbb{P}^* \left(S_n^* > K\sqrt{n} \right) \rightarrow 0.$$

Then, we only need to estimate the first part in (6.12). Define the following bootstrap procedure: Let $\vec{\mathbf{B}}_i := \{X_{\mathcal{I}_{i-1}+1}, \dots, X_{\mathcal{I}_i}, 0, 0, \dots\}$ and let $\vec{\mathcal{F}}$ be a class of function, related to the class of functions \mathcal{F} , such that, for every $\vec{\omega}_h \in \vec{\mathcal{F}}$:

$$\begin{cases} \vec{\omega}_h(\vec{\mathbf{B}}_1, \vec{\mathbf{B}}_2, \dots, \vec{\mathbf{B}}_k) = \sum_{i_1=1}^{\infty} \dots \sum_{i_k=1}^{\infty} h(x_{i_1}, \dots, x_{i_k}) \mathbb{1}_{x_k \neq 0} & \text{if defined,} \\ \infty & \text{otherwise.} \end{cases} \quad (6.13)$$

It is classical that $\{\vec{\mathbf{B}}_i\}$ are i.i.d. Applying the same bootstrap method of Algorithm 1.

This new sample allows us to enlarge and bound (6.12) by

$$\mathbb{P}^* \left(\sup_{\vec{h} \in \vec{\mathcal{H}}} \left| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in \mathcal{J}_m''} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \vec{\omega}_{\vec{h}}(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) \right| > \varepsilon \right), \quad (6.14)$$

where $\vec{h} \in \vec{\mathcal{H}} = \{\vec{\omega}_{\vec{f}} - \vec{\omega}_{\vec{g}}, \vec{f}, \vec{g} \in \vec{\mathcal{F}}\}$ the corresponding class $\mathcal{H} = \{\omega_f - \omega_g, f, g \in \mathcal{F}\}$, each one with envelop \tilde{F} and F respectively. To estimate the last expression, we will use the *bracketing*. Define the *Bracket* $[f^\ell, f^u]$ by:

$$[f^\ell, f^u] := \{f \in \mathcal{F} : f^\ell \leq f \leq f^u\},$$

and the *bracketing entropy number* by $N_1(\gamma, \mathcal{F}, \mathbb{P})$ which is denote the smallest the minimal number $N \geq 1$ for which there exist functions $f_1^\ell, \dots, f_N^\ell$ and f_1^u, \dots, f_N^u such that:

$$\begin{cases} \mathcal{F} \subset \bigcup_{k=1}^N [f_k^\ell, f_k^u], \\ \int_S (f_k^u - f_k^\ell) \mathbb{P} \leq \gamma. \end{cases} \quad (6.15)$$

For the class of functions \mathcal{H} , consider the brackets $[h^\ell, h^u]$, such that $\mathbb{E}^*(h^\ell, h^u) \leq \gamma$, where $\gamma > 0$ and it will be determined later. In this framework, the bracketing entropy number will be $N^*(\gamma) := N_1(\gamma, \mathcal{H}, \mathbb{D}^*)$, for

$$\mathbb{D}^* = \binom{l_n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_{l_n}^m} \xi_{i_1, l_n} \dots \xi_{i_m, l_n} \delta_{(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m})}.$$

Hence, we have the following inequalities

$$\begin{aligned} & \sup_{\vec{h} \in \vec{\mathcal{H}}} \left| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in \mathcal{J}_m''} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \vec{\omega}_{\vec{h}}(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) \right| \\ & \leq \max_{k \leq N^*(\gamma)} \left| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in \mathcal{J}_m''} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) (h_k^u - h_k^\ell) (\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) \right| \\ & \leq \max_{k \leq N^*(\gamma)} \left| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in \mathcal{J}_m''} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \right. \\ & \quad \left. [h_k^u(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) - \mathbb{E}^*(h_k^u(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}))] \right| \\ & + \max_{k \leq N^*(\gamma)} \left| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in \mathcal{J}_m''} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \right. \end{aligned}$$

$$\begin{aligned}
 & \left| h_k^\ell(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) - \mathbb{E}^* \left(h_k^\ell(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) \right) \right| \\
 & + \max_{k \leq N^*(\gamma)} \left| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in \mathcal{J}_m''} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \right. \\
 & \quad \left. \left[\mathbb{E}^* \left(h_k^u(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) \right) - \mathbb{E}^* \left(h_k^\ell(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) \right) \right] \right| \\
 := & I_A + I_B + I_C. \tag{6.16}
 \end{aligned}$$

Treating each term, keeping in mind the condition (A.1), i.e., $\sum_{i=1}^n \xi_i = n$, we have

$$\begin{aligned}
 I_C & := \max_{1 \leq k \leq N^*(\gamma)} \left| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in \mathcal{J}_m''} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \right. \\
 & \quad \left. \left[\mathbb{E}^* \left(h_k^u(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) \right) - \mathbb{E}^* \left(h_k^\ell(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) \right) \right] \right| \\
 & \leq \gamma \left| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in \mathcal{J}_m''} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \right| \\
 & = \gamma \left| \binom{n}{m}^{-1} \sum_{\mathbf{i}_{\{1, \dots, m-1\}}} \sum_{j=1}^m \prod_{k=1, k \neq j}^m \xi_{i_k, n} \left\{ \sum_{i_j=1}^n (\xi_{i_j, n} - 1) \right\} \right| = 0,
 \end{aligned}$$

and

$$\begin{aligned}
 \mathbb{P}(I_B > \varepsilon) & := \mathbb{P} \left(\max_{k \leq N^*(\gamma)} \left| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in \mathcal{J}_m''} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \right. \right. \\
 & \quad \left. \left. \left[h_k^\ell(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) - \mathbb{E}^* \left(h_k^\ell(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) \right) \right] \right| > \varepsilon \right) \\
 & \leq N^*(\gamma) \max_{k \leq N^*(\gamma)} \mathbb{P} \left(\left| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in \mathcal{J}_m''} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \right. \right. \\
 & \quad \left. \left. \left[h_k^\ell(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) - \mathbb{E}^* \left(h_k^\ell(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) \right) \right] \right| > \varepsilon \right) \\
 & \quad \text{(For } h_k^\ell = h_k^\ell \mathbf{1}_{h_k^\ell \leq M_n} + h_k^\ell \mathbf{1}_{h_k^\ell > M_n} \text{)} \\
 & \leq N^*(\gamma) \max_{k \leq N^*(\gamma)} \mathbb{P} \left(\left| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in \mathcal{J}_m''} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \right. \right. \\
 & \quad \left. \left. \left[h_k^\ell \mathbf{1}_{h_k^\ell \leq M_n}(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) - \mathbb{E}^* \left(h_k^\ell \mathbf{1}_{h_k^\ell \leq M_n}(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) \right) \right] \right| > \varepsilon \right) \\
 & + N^*(\gamma) \max_{k \leq N^*(\gamma)} \mathbb{P} \left(\left| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in \mathcal{J}_m''} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \right. \right.
 \end{aligned}$$

$$\begin{aligned}
 & \left| h_k^\ell \mathbf{1}_{h_k^\ell > M_n}(\vec{B}_{i_1}, \dots, \vec{B}_{i_m}) - \mathbb{E}^* \left(h_k^\ell \mathbf{1}_{h_k^\ell > M_n}(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) \right) \right| > \varepsilon \\
 \leq & \frac{N^*(\gamma)}{\varepsilon^2 \binom{n}{m}^{1/2}} \mathbb{E}^* \left((\xi_{1,l_n} - 1) \dots (\xi_{m,l_n} - 1) \left[h_k^\ell \mathbf{1}_{h_k^\ell \leq M_n} - \mathbb{E}^* \left(h_k^\ell \mathbf{1}_{h_k^\ell \leq M_n} \right) \right] \right)^2 \\
 & + \frac{N^*(\gamma)}{\varepsilon} \mathbb{E}^* \left| (\xi_{1,l_n} - 1) \dots (\xi_{m,l_n} - 1) \left[h_k^\ell \mathbf{1}_{h_k^\ell > M_n} - \mathbb{E}^* \left(h_k^\ell \mathbf{1}_{h_k^\ell > M_n} \right) \right] \right| \\
 \leq & \frac{N^*(\gamma)}{\varepsilon^2 \binom{n}{m}^{1/2}} \mathbb{E}_\xi \left(\prod_{i=1}^m (\xi_{i,l_n} - 1)^2 \right) \mathbb{E}^* \left(\left[h_k^\ell \mathbf{1}_{h_k^\ell \leq M_n} - \mathbb{E}^* \left(h_k^\ell \mathbf{1}_{h_k^\ell \leq M_n} \right) \right] \right)^2 \\
 & + \frac{N^*(\gamma)}{\varepsilon} \mathbb{E}^* \left| (\xi_{1,l_n} - 1) \dots (\xi_{m,l_n} - 1) \left[h_k^\ell \mathbf{1}_{h_k^\ell > M_n} - \mathbb{E}^* \left(h_k^\ell \mathbf{1}_{h_k^\ell > M_n} \right) \right] \right| \\
 \leq & \frac{N^*(\gamma)}{\varepsilon^2 \binom{n}{m}^{1/2}} c^2 \times 4M_n^2 + \frac{N^*(\gamma)}{\varepsilon} \mathbb{E}^* \left| (\xi_{1,l_n} - 1) \dots (\xi_{m,l_n} - 1) h_k^\ell \mathbf{1}_{h_k^\ell > M_n} \right| \\
 \leq & \frac{N^*(\gamma)}{\varepsilon^2 \binom{n}{m}^{1/2}} c^2 \times 4M_n^2 + \frac{2N^*(\gamma)}{\varepsilon} \mathbb{E}^* \left| (\xi_{1,l_n} - 1) \dots (\xi_{m,l_n} - 1) \tilde{F} \mathbf{1}_{\tilde{F} > M_n} \right| \\
 \leq & \frac{N^*(\gamma)}{\varepsilon^2 \binom{n}{m}^{1/2}} c^2 \times 4M_n^2 \\
 & + \frac{4N^*(\gamma)}{\varepsilon} \binom{l_n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in I_{l_n}^m} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \left| \tilde{F} \mathbf{1}_{\tilde{F} > M_n}(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) \right|,
 \end{aligned} \tag{6.17}$$

yet, $\vec{\mathbf{B}}_i$ are i.i.d and and $\mathbb{E}(\tilde{F}) = \mathbb{E}(\tau)^m \mathbb{E}F < \infty$, for any $M_n \nearrow \infty$, we have

$$\left(\frac{\mathbb{E}(l_n)}{m} \right)^{-1} \sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(l_n)}^m} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \tilde{F} \mathbf{1}_{\tilde{F} > M_n}(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) \rightarrow 0 \quad a.s.$$

Using the same argument as in part iii) of Lemma 3.1, we can prove that

$$\left(\frac{l_n}{m} \right)^{-1} \sum_{(i_1, \dots, i_m) \in I_{l_n}^m} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \left| \tilde{F} \mathbf{1}_{\tilde{F} > M_n}(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) \right| \rightarrow 0 \quad \text{in probability.}$$

Then it remains just to find that, for every fixed $\gamma > 0$, $N^*(\gamma)$ is bounded in probability, as the last expression in (6.17) does not depend on k . Good to be noting that $N_1(\gamma, \vec{\mathcal{H}}, \mathbb{P})$ is finite, due to the boundness of $\vec{\mathcal{H}}$ by $2F$ with $\mathbb{E}\tilde{F}(\vec{\mathbf{B}}) < \infty$ and the fact that $\vec{\mathbf{B}}_i$ are i.i.d and discrete random variables. Under the norm $L_1(\mathbb{P})$, define a $\gamma/2$ Brackets, $h_1^\ell, \dots, h_{N(\gamma/2)}^\ell$ and $h_1^u, \dots, h_{N(\gamma/2)}^u$. Observe that

$$\max_{j \leq N(\gamma/2)} \left| \left(\frac{l_n}{m} \right)^{-1/2} \sum_{(i_1, \dots, i_m) \in I_{l_n}^m} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) (h_j^u - h_j^\ell) (\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) \right|, \tag{6.18}$$

converge to zero in probability, with $N(\gamma/2)$ does not depend on n . That implies that $N^*(\gamma) \leq N(\gamma/2) < \infty$ in probability. Replacing h^ℓ by h^u , I_A will be identical to I_B , i.e. I_A also converges to 0 in probability. This proves the convergence of I_n to 0 in probability.

For II_n : In the same manner, let

$$S_n^* := \sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(l_n)}^m} \mathbf{1}_{(\widehat{\mathbf{B}}_{i_1}, \dots, \widehat{\mathbf{B}}_{i_m}) \in \mathcal{A}_n^c}.$$

Define a new bootstrap sample $\{\mathbf{B}_i^{**}\}$ in $i = l_n + 1, \dots, \mathbb{E}(l_n)$. Clearly, the new sample is well defined since we assumed at the beginning that $l_n \leq \mathbb{E}(l_n)$, and it is defined independently from \mathbf{B}_i^* and $\widehat{\mathbf{B}}_i^*$. In this case:

$$\begin{aligned} & \mathcal{L} \left(\sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(l_n)}^m} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_{\tilde{h}}(\widehat{\mathbf{B}}_{i_1}, \dots, \widehat{\mathbf{B}}_{i_m}) \mathbf{1}_{(\widehat{\mathbf{B}}_{i_1}, \dots, \widehat{\mathbf{B}}_{i_m}) \in \mathcal{A}_n^c} \right) \\ &= \mathcal{L} \left(\sum_{(i_1, \dots, i_m) \in I_{S_n^*}^m} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_{\tilde{h}}(\mathbf{B}_{i_1}^{**}, \dots, \mathbf{B}_{i_m}^{**}) \right). \end{aligned}$$

Hence, as same as in (6.12), we have:

$$\begin{aligned} & \mathbb{P}^*(II_n > \varepsilon) \\ &= \mathbb{P}^{**} \left(\left\| \sum_{(i_1, \dots, i_m) \in I_{S_n^*}^m} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \right\|_{\mathcal{F}_\delta} > \varepsilon \right) \\ &\leq C_1 \mathbb{P}^{**} \left(\left\| \sum_{(i_1, \dots, i_m) \in \mathcal{J}_m''} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \omega_{\tilde{h}}(\mathbf{B}_{i_1}, \dots, \mathbf{B}_{i_m}) \right\|_{\mathcal{F}_\delta} > C_2 \varepsilon \right) \\ &\quad + \mathbb{P}^*(S_n^* > K\sqrt{n}), \text{ for any } K > 0, \end{aligned} \tag{6.19}$$

where

$$\begin{aligned} \mathcal{J}_m'' := \{ & (i_1, \dots, i_m) : 1 \leq i_1 < \dots < i_m \leq S_n^*, \exists \ell = 1, \dots, m, \\ & S_n^* < i_\ell \leq K\sqrt{n}, i_j \neq i_k \text{ for } j \neq k \}. \end{aligned}$$

Using the same bootstrap procedure defined previously for I_n , let

$$\vec{\mathbf{B}}_i := \{X_{\mathcal{T}_{i-1}+1}, \dots, X_{\mathcal{T}_i}, 0, 0, \dots\},$$

for $i = l_n + 1, \dots, \mathbb{E}(l_n)$, and let $\vec{\mathcal{F}}$ be a class of function such that, for every $\vec{\omega}_h \in \vec{\mathcal{F}}$:

$$\begin{cases} \vec{\omega}_h(\vec{\mathbf{B}}_1, \vec{\mathbf{B}}_2, \dots, \vec{\mathbf{B}}_k) = \sum_{i_1=1}^{\infty} \dots \sum_{i_k=1}^{\infty} h(x_{i_1}, \dots, x_{i_k}) \mathbb{1}_{x_k \neq 0} & \text{if defined} \\ \infty & \text{otherwise.} \end{cases} \quad (6.20)$$

It is classical that $\{\vec{\mathbf{B}}_i\}$ are i.i.d. Applying the same bootstrap method of Algorithm 1. This new sample allows us to enlarge and bound (6.12) by

$$\mathbb{P}^{**} \left(\sup_{\vec{h} \in \vec{\mathcal{H}}} \left| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in \mathcal{J}_m''} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \vec{\omega}_{\vec{h}}(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) \right| > \varepsilon \right), \quad (6.21)$$

where

$$\vec{h} \in \vec{\mathcal{H}} = \{ \vec{\omega}_{\vec{f}} - \vec{\omega}_{\vec{g}}, \vec{f}, \vec{g} \in \vec{\mathcal{F}} \}$$

corresponding to the class

$$\mathcal{H} = \{ \omega_f - \omega_g, f, g \in \mathcal{F} \},$$

each one with envelop \vec{F} and F respectively. As before, for the class of functions \mathcal{H} , consider the brackets $[h^\ell, h^u]$, such that

$$\mathbb{E}^{**}(h^\ell, h^u) \leq \gamma,$$

where $\gamma > 0$ and it will be determined later. In this framework, the bracketing entropy number will be $N^{**}(\gamma) := N_1(\gamma, \mathcal{H}, \mathbb{D}^{**})$, for

$$\mathbb{D}^{**} = \binom{\mathbb{E}(l_n) - l_n}{m}^{-1} \sum_{(i_1, \dots, i_m) \in \mathcal{J}_{\mathbb{E}(l_n) - l_n}^m} \xi_{i_1, l_n} \dots \xi_{i_m, l_n} \delta_{(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m})}.$$

Following the same arguments from displays (6.16) through (6.17), we can find that (6.21) will be

$$\begin{aligned} &\leq N^{**}(\gamma) \max_{k \leq N^{**}(\gamma)} \mathbb{P} \left(\left| \binom{n}{m}^{-1/2} \sum_{(i_1, \dots, i_m) \in \mathcal{J}_m''} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \right. \right. \\ &\quad \left. \left. [h_k^\ell(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) - \mathbb{E}^{**}(h_k^\ell(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}))] \right| > \varepsilon \right) \\ &\leq \frac{N^{**}(\gamma)}{\varepsilon^2 \binom{n}{m}^{1/2}} c^2 \times 4M_n^2 + \frac{2N^{**}(\gamma)}{\varepsilon} \mathbb{E}^{**} |(\xi_{1, l_n} - 1) \dots (\xi_{m, l_n} - 1) \tilde{F} \mathbb{1}_{\tilde{F} > M_n}| \\ &\leq \frac{N^{**}(\gamma)}{\varepsilon^2 \binom{n}{m}^{1/2}} c^2 \times 4M_n^2 + \frac{4N^{**}(\gamma)}{\varepsilon} \end{aligned}$$

$$\left(\binom{\mathbb{E}(l_n) - l_n}{m} \right)^{-1} \sum_{(i_1, \dots, i_m) \in I_{\mathbb{E}(l_n) - l_n}^m} (\xi_{i_1, l_n} - 1) \dots (\xi_{i_m, l_n} - 1) \left| \tilde{F} \mathbf{1}_{\tilde{F} > M_n}(\vec{\mathbf{B}}_{i_1}, \dots, \vec{\mathbf{B}}_{i_m}) \right|. \quad (6.22)$$

Here, we must pay attention to the randomness of N^{**} which depends on n . According to Lemma 3.1 i), we can see that $|\mathbb{E}(l_n) - l_n| \rightarrow \infty$ in probability, under the assumption that $l_n < \mathbb{E}(l_n)$. Now, using the same treatment of I_n , and for $M_n := n^{1/3}$ (to provide the convergence of M_n to ∞), as same as in Radulović, 2004, this allows the convergence of (6.22) to 0 in probability. Estimating now N^{**} by considering the same $\gamma/2$ brackets $h_1^\ell, \dots, h_{N(\gamma/2)}^\ell$ and $h_1^u, \dots, h_{N(\gamma/2)}^u$. We have $N^{**}(\gamma) < N(\gamma/2)$, which does not depend on n . Then, II_n is proved. Following the same footsteps, we can prove the case where $l_n > \mathbb{E}(l_n)$. This proves Step 3. \square

The end of the previous step provides that we only need to show the stochastic equicontinuity of $\widehat{T}_{l_n}^*$, where the number of blocks is replaced by the deterministic quantity $\mathbb{E}(l_n)$. In order to achieve the equicontinuity of this statistic, Lemma 4.1 shows that it is sufficient to prove that:

$$\begin{aligned} & \mathbb{E} \left\| \sum_{1 \leq i_k \leq \ell_k, 1 \leq k \leq m(n)} \varepsilon_{i_1}^{(1)} \dots \varepsilon_{i_m}^{(m)} \omega_{\vec{h}} \left(\widehat{\mathbf{B}}_{i_1}^{(1)}, \dots, \widehat{\mathbf{B}}_{i_m}^{(m)} \right) \right\|_{\overrightarrow{\mathcal{H}}} \\ & \leq a(\ell_1, \dots, \ell_m) \left(\prod_{k=1}^m \ell_k \right)^{1/2} \end{aligned}$$

for all $1 \leq \ell_1, \dots, \ell_m \leq n$. We begin to define the distance:

$$e_\ell^2(f, g) \equiv \frac{1}{\prod_{k=1}^m \ell_k} \sum_{1 \leq i_k \leq \ell_k, 1 \leq k \leq m} \omega_{\vec{h}}^2 \left(\widehat{\mathbf{B}}_{i_1}^{(1)}, \dots, \widehat{\mathbf{B}}_{i_m}^{(m)} \right),$$

defined in L_2 , associated with the Rademacher process

$$\left\{ \frac{1}{(\prod_{k=1}^m \ell_k)^{1/2}} \sum_{1 \leq i_k \leq \ell_k, 1 \leq k \leq m} \varepsilon_{i_1}^{(1)} \dots \varepsilon_{i_m}^{(m)} \omega_{\vec{h}} \left(\widehat{\mathbf{B}}_{i_1}^{(1)}, \dots, \widehat{\mathbf{B}}_{i_m}^{(m)} \right) : \vec{h} \in \overrightarrow{\mathcal{H}} \mid \widehat{\mathbf{B}}_1, \dots, \widehat{\mathbf{B}}_m \right\}.$$

Take $\|f\|_\ell^2 \equiv e_\ell^2(f, 0)$ and

$$r_\ell(\delta) \equiv \sup_{f \in \overrightarrow{\mathcal{F}}_\delta} \|f\|_\ell^2.$$

Using Corollary 6.2, we have

$$\mathbb{E} \left\| \frac{1}{(\prod_{k=1}^m \ell_k)^{1/2}} \sum_{1 \leq i_k \leq \ell_k, 1 \leq k \leq m} \varepsilon_{i_1}^{(1)} \dots \varepsilon_{i_m}^{(m)} \omega_{\vec{h}} \left(\widehat{\mathbf{B}}_{i_1}^{(1)}, \dots, \widehat{\mathbf{B}}_{i_m}^{(m)} \right) \right\|_{\overrightarrow{\mathcal{H}}}^2$$

$$\begin{aligned}
 &\leq C \int_0^{r_\ell(\delta)} (\log \mathcal{N}(\varepsilon, \mathcal{F}, e_\ell))^{m/2} d\varepsilon \\
 &= C \|F\|_\ell \cdot \int_0^{r_\ell(\delta)/\|F\|_\ell} (\log \mathcal{N}(\varepsilon \|F\|_\ell, \mathcal{F}, e_\ell))^{m/2} d\varepsilon \\
 &\leq C \|F\|_\ell \cdot \int_0^{r_\ell(\delta)/\|F\|_\ell} \left(\sup_Q \log \mathcal{N}(\varepsilon \|F\|_{L_2(Q)}, \mathcal{F}, L_2(Q)) \right)^{m/2} d\varepsilon. \quad (6.23)
 \end{aligned}$$

Assuming that $F \geq 1$, the upper bound in the integral can be replaced by $r_\ell(\delta)$. The following proposition is necessary for the following

Proposition 6.1 (Q. Han, 2022). *Let $\{X_i\}$ be i.i.d. random variables with law \mathbb{P} . Let \mathcal{H} be a class of measurable real-valued functions defined on $(\mathcal{X}^m, \mathcal{A}^m)$ with an \mathbb{P}^m -integrable envelope such that the following holds: for any fixed $\delta > 0, M > 0, 1 \leq k \leq m$,*

$$\max_{1 \leq j' \leq k} \mathbb{E} \left(\frac{\log \mathcal{N}(\delta, (\pi_k \mathcal{H})_M, e_{\ell, j'})}{\ell_{j'}} \right)^{1/2} \rightarrow 0 \quad (6.24)$$

holds for any $\ell_1 \wedge \dots \wedge \ell_k \rightarrow \infty$. Here for $\ell = (\ell_1, \dots, \ell_k)$ and $\{X_i\}_{i=1}^\infty$,

$$e_{\ell, j'}(f, g) \equiv \frac{1}{\ell_{j'}} \sum_{i_{j'}=1}^{\ell_{j'}} \left| \frac{1}{\prod_{j \neq j'} \ell_j} \sum_{1 \leq i_j \leq \ell_j: j \neq j'} (f - g)(X_{i_1}, \dots, X_{i_k}) \right|$$

and

$$(\pi_k \mathcal{H})_M \equiv \{h 1_{H_k \leq M} : h \in \pi_k \mathcal{H}\},$$

where H_k is an envelope for $\pi_k \mathcal{H}$. Then

$$\sup_{h \in \mathcal{H}} \left| \frac{1}{\prod_{k=1}^m \ell_k} \sum_{1 \leq i_k \leq \ell_k, 1 \leq k \leq m} (h(X_{i_1}, \dots, X_{i_m}) - \mathbb{P}^m h) \right| \rightarrow 0$$

in L_1 as $\ell_1 \wedge \dots \wedge \ell_m \rightarrow \infty$. The above display can be replaced by the decoupled version.

By this proposition, $\|F\|_\ell \rightarrow_{\mathbb{P}} \|F\|_{L_2(P)}$ as $\ell_1 \wedge \dots \wedge \ell_m \rightarrow \infty$, therefore, it suffices to get $r_\ell(\delta) \rightarrow_p 0$ as $\ell_1 \wedge \dots \wedge \ell_m \rightarrow \infty$ and $\delta \rightarrow 0$. It is obvious that all that is left to do now is to demonstrate that

$$\sup_{f \in \mathcal{F}_\delta} \left| \frac{1}{\prod_{k=1}^m \ell_k} \sum_{1 \leq i_k \leq \ell_k, 1 \leq k \leq m} \left(\omega_h^2(\widehat{\mathbf{B}}_{i_1}^{(1)}, \dots, \widehat{\mathbf{B}}_{i_m}^{(m)}) - \mathbb{P}^m \omega_h^2 \right) \right| \rightarrow_p 0. \quad (6.25)$$

Verifying condition (6.24)

$$\max_{1 \leq j' \leq k} \mathbb{E} \left(\frac{\log \mathcal{N}(\delta, \mathcal{F}_M^2, e_{\ell, j'})}{\ell_{j'}} \right)^{1/2}$$

$$\begin{aligned}
 &\leq (\ell_1 \wedge \cdots \wedge \ell_k)^{-1/2} \mathbb{E} \left[\int_0^\delta \left(\log \mathcal{N} \left(\varepsilon, \mathcal{F}_M^2, e_{\ell, j'} \right) \right)^{m/2} d\varepsilon \right] \\
 &\leq \left(\frac{\delta}{\sqrt{2M}} \right)^{-1} (\ell_1 \wedge \cdots \wedge \ell_k)^{-1/2} \mathbb{E} \left[\int_0^{\delta/\sqrt{2M}} \left(\log \mathcal{N} \left(\varepsilon, \mathcal{F}_M^2, e_{\ell, j'} \right) \right)^{m/2} d\varepsilon \right] \\
 &\leq (\delta/2M)^{-1} (\ell_1 \wedge \cdots \wedge \ell_k)^{-1/2} \int_0^1 \left(\sup_Q \log \mathcal{N} \left(\varepsilon \|F\|_{L_2(Q)}, \mathcal{F}, L_2(Q) \right) \right)^{m/2} d\varepsilon \\
 &\times \|F\|_{L_2(\mathbb{P}^m)} \rightarrow 0. \tag{6.26}
 \end{aligned}$$

The shift from the second to the third line is true because

$$\mathcal{N} \left(\delta, \mathcal{F}_M^2, L_2(Q) \right) \leq \mathcal{N} \left(\delta/\sqrt{2M}, \mathcal{F}_M, L_2(Q) \right).$$

As the condition is verified, as well as $\ell_1 \wedge \cdots \wedge \ell_m \rightarrow \infty$, (6.25) follows directly using the previous proposition. Hence, there exists some sequence $\{a_\ell\}$, in a way that $a_\ell \rightarrow 0$ for any sequence $\{\delta_\ell\}$ with $\delta_\ell \rightarrow 0$ both under $\ell_1 \wedge \cdots \wedge \ell_m \rightarrow \infty$, such that:

$$\mathbb{E} \left\| \sum_{1 \leq i_k \leq \ell_k, 1 \leq k \leq m} \varepsilon_{i_1}^{(1)} \cdots \varepsilon_{i_m}^{(m)} \omega_{\tilde{h}} \left(\widehat{\mathbf{B}}_{i_1}^{(1)}, \dots, \widehat{\mathbf{B}}_{i_m}^{(m)} \right) \right\|_{\tilde{\mathcal{H}}} \leq a_\ell \left(\prod_{k=1}^m \ell_k \right)^{1/2}. \tag{6.27}$$

An application of Lemma 4.1 proves that

$$n^{-m/2} \mathbb{E} \left\| \sum_{1 \leq i_1, \dots, i_m \leq n} (\xi_{i_1} - 1) \cdots (\xi_{i_m} - 1) \omega_{\tilde{h}} \left(\widehat{\mathbf{B}}_{i_1}, \dots, \widehat{\mathbf{B}}_{i_m} \right) \right\|_{\mathcal{F}_{\delta_n}} \rightarrow 0, \quad n \rightarrow \infty.$$

This completes the proof for the asymptotic equicontinuity. \square

Appendix

This appendix contains supplementary information that is an essential part in providing a more comprehensive understanding of the chapter.

Proof of Theorem 4.1:

We have

$$\mathbb{E} \left\| \sum_{1 \leq i_1 < \dots < i_m \leq n} \xi_{i_1} \cdots \xi_{i_m} f \left(Y_{i_1}^{(1)}, \dots, Y_{i_m}^{(m)} \right) \right\|_{\mathcal{F}}.$$

By decoupling of U -process, due to de la Peña et al., 1999

$$\leq \mathbf{C}_m \mathbb{E} \left\| \sum_{1 \leq i_1 < \dots < i_m \leq n} \xi_{i_1} \dots \xi_{i_m} f(Y_{i_1}^{(1)}, \dots, Y_{i_m}^{(m)}) \right\|_{\mathcal{F}}.$$

By symmetrization, due to de la Peña et al., 1999, we have

$$\leq 2^m \mathbf{C}_m \mathbb{E} \left\| \sum_{1 \leq i_1 < \dots < i_m \leq n} \xi_{i_1} \dots \xi_{i_m} \varepsilon_{i_1}^{(1)} \dots \varepsilon_{i_m}^{(m(n))} f(Y_{i_1}^{(1)}, \dots, Y_{i_m}^{(m)}) \right\|_{\mathcal{F}}$$

for $(\text{sgn}(\xi_1)\varepsilon_1^*, \dots, \text{sgn}(\xi_n)\varepsilon_n^*)$ a sequence independent and with the same distribution of (ξ_1, \dots, ξ_n) , By the invariance of $(\mathbb{P}_\varepsilon \otimes \mathbb{P})^{mn}$ and the fact that ξ is independent of X, ε , we have that

$$\begin{aligned} &= 2^m \mathbf{C}_m \mathbb{E}_{x, \varepsilon} \left\| \sum_{1 \leq i_1 < \dots < i_m \leq n} |\xi_{i_1}| \dots |\xi_{i_m}| \text{sgn}(\xi_{i_1})\varepsilon_{i_1}^{(1)} \dots \text{sgn}(\xi_{i_m})\varepsilon_{i_m}^{(m(n))} f(Y_{i_1}^{(1)}, \dots, Y_{i_m}^{(m)}) \right\|_{\mathcal{F}} \\ &= 2^m \mathbf{C}_m \mathbb{E}_{x, \varepsilon} \left\| \sum_{1 \leq i_1 < \dots < i_m \leq l_n - 1} |\xi_{i_1}| \dots |\xi_{i_m}| \varepsilon_{i_1}^{(1)} \dots \varepsilon_{i_m}^{(m(n))} f(Y_{i_1}^{(1)}, \dots, Y_{i_m}^{(m)}) \right\|_{\mathcal{F}}, \end{aligned}$$

using the reversed order statistics of $\{|\xi_i|\}_{i=1}^n$, $|\xi_{(1)}| \geq \dots \geq |\xi_{(n)}|$, and the permutations between the different sequences of random variable, and in the same footsteps as Q. Han, 2022,

$$= 2^m \mathbf{C}_m \mathbb{E} \left\| \sum_{1 \leq i_1 < \dots < i_m \leq n} |\xi_{(i_1)}| \dots |\xi_{(i_m)}| \varepsilon_{i_1}^{(1)} \dots \varepsilon_{i_m}^{(m(n))} f(Y_{i_1}^{(1)}, \dots, Y_{i_m}^{(m)}) \right\|_{\mathcal{F}}$$

substitute $\xi_{(i)}$ by $\sum_{k=i}^n \xi_{(k)} - \xi_{(k+1)}$, with $|\xi_{(n+1)}| = 0$,

$$\begin{aligned} &\leq 2^m \mathbf{C}_m \mathbb{E} \left\| \sum_{1 \leq i_1 < \dots < i_m \leq n} \sum_{k_j \geq i_j, 1 \leq j \leq m} (|\xi_{(l_1)}| - |\xi_{(l_1+1)}|) \dots (|\xi_{(l_m)}| - |\xi_{(l_m+1)}|) \right. \\ &\quad \left. \times \varepsilon_{i_1}^{(1)} \dots \varepsilon_{i_m}^{(m(n))} f(Y_{i_1}^{(1)}, \dots, Y_{i_m}^{(m)}) \right\|_{\mathcal{F}} \\ &\leq 2^m \mathbf{C}_m \mathbb{E} \left\| \sum_{1 \leq i_1, \dots, i_m \leq n} \sum_{l_k \geq i_k, 1 \leq k \leq m} (|\xi_{(l_1)}| - |\xi_{(l_1+1)}|) \dots (|\xi_{(l_m)}| - |\xi_{(l_m+1)}|) \right. \\ &\quad \left. \times \varepsilon_{i_1}^{(1)} \dots \varepsilon_{i_m}^{(m(n))} f(Y_{i_1}^{(1)}, \dots, Y_{i_m}^{(m)}) \right\|_{\mathcal{F}} \\ &\leq 2^m \mathbf{C}_m \mathbb{E} \left[\sum_{1 \leq \ell_1, \dots, \ell_m \leq n} (|\xi_{(\ell_1)}| - |\xi_{(\ell_1+1)}|) \dots (|\xi_{(\ell_m)}| - |\xi_{(\ell_m+1)}|) \right. \\ &\quad \left. \times \mathbb{E} \left\| \sum_{1 \leq i_k \leq \ell_k, 1 \leq k \leq m} \varepsilon_{i_1}^{(1)} \dots \varepsilon_{i_m}^{(m(n))} f(Y_{i_1}^{(1)}, \dots, Y_{i_m}^{(m)}) \right\|_{\mathcal{F}} \right] \end{aligned}$$

$$\begin{aligned}
 &\leq 2^m \mathbf{C}_m \mathbb{E} \left[\sum_{1 \leq \ell_1, \dots, \ell_m \leq n} \int_{|\xi_{(\ell_1+1)}|}^{|\xi_{(\ell_1)}|} \cdots \int_{|\xi_{(\ell_m+1)}|}^{|\xi_{(\ell_m)}|} \psi_n(\ell_1, \dots, \ell_m) dt_m \cdots dt_1 \right] \\
 &\leq 2^m \mathbf{C}_m \mathbb{E} \left[\sum_{1 \leq \ell_1, \dots, \ell_m \leq n} \int_{|\xi_{(\ell_1+1)}|}^{|\xi_{(\ell_1)}|} \cdots \int_{|\xi_{(\ell_m+1)}|}^{|\xi_{(\ell_m)}|} \right. \\
 &\quad \left. \psi_n(\{|i : |\xi_i| > t_1\}, \dots, \{|i : |\xi_i| > t_m\}) dt_m \cdots dt_1 \right] \\
 &\leq 2^m \mathbf{C}_m \mathbb{E} \left[\int_{\mathbb{R}_{\geq 0}^m} \psi_n(\{|i : |\xi_i| > t_1\}, \dots, \{|i : |\xi_i| > t_m\}) dt_1 \cdots dt_m \right] \\
 &\leq 2^m C_m \int_{\mathbb{R}_{\geq 0}^m} \mathbb{E} \psi_n \left(\sum_{i=1}^n \mathbf{1}_{|\xi_i| > t_1}, \dots, \sum_{i=1}^n \mathbf{1}_{|\xi_i| > t_m} \right) dt_1 \cdots dt_m. \quad (\text{By Fubini theorem.})
 \end{aligned}$$

Now suppose that $\psi_n(\ell_1, \dots, \ell_m) = \bar{\psi}_n(\prod_{k=1}^m \ell_k)$. Then we may further bound the above display by

$$\begin{aligned}
 &\int_{\mathbb{R}_{\geq 0}^m} \mathbb{E} \bar{\psi}_n \left(\prod_{k=1}^m \sum_{i=1}^n \mathbf{1}_{|\xi_i| > t_k} \right) dt_1 \cdots dt_m \\
 &= \int_{\mathbb{R}_{\geq 0}^m} \mathbb{E} \bar{\psi}_n \left(\sum_{1 \leq i_1, \dots, i_m \leq n} \prod_{k=1}^m \mathbf{1}_{|\xi_{i_k}| > t_k} \right) dt_1 \cdots dt_m \\
 &\leq \int_{\mathbb{R}_{\geq 0}^m} \bar{\psi}_n \left(\sum_{1 \leq i_1, \dots, i_m \leq n} \mathbb{E} \prod_{k=1}^m \mathbf{1}_{|\xi_{i_k}| > t_k} \right) dt_1 \cdots dt_m \quad (\text{by Jensen's inequality}) \\
 &\leq \int_{\mathbb{R}_{\geq 0}^m} \bar{\psi}_n \left(\sum_{1 \leq i_1, \dots, i_m \leq n} \prod_{k=1}^m \mathbb{P}(|\xi_{i_k}| > t_k)^{1/m} \right) dt_1 \cdots dt_m,
 \end{aligned}$$

where the last inequality follows from generalized Hölder's inequality and the assumption that $\bar{\psi}_n$ is non-decreasing.

Proof of Lemma 4.1

For $\psi_n(\ell_1, \dots, \ell_m) \equiv a(\ell_1, \dots, \ell_m) (\prod_{k=1}^m \ell_k)^{1/2}$. Theorem 4.1 implies that:

$$\begin{aligned}
 &\mathbb{E} \left\| \sum_{1 \leq i_1, \dots, i_m \leq n} \xi_{i_1} \cdots \xi_{i_m} f(Y_{i_1}, \dots, Y_{i_m}) \right\|_{\mathcal{F}_{(n, \dots, n), n}} \\
 &\leq K_m \int_{\mathbb{R}_{\geq 0}^m} \mathbb{E} \left[a \left(\sum_{i=1}^n \mathbf{1}_{|\xi_i| > t_1}, \dots, \sum_{i=1}^n \mathbf{1}_{|\xi_i| > t_m} \right) \prod_{k=1}^m \left(\sum_{i=1}^n \mathbf{1}_{|\xi_i| > t_k} \right)^{1/2} \right] dt_1 \cdots dt_m \\
 &\leq K_m \int_{\mathbb{R}_{\geq 0}^m} A_{2,n}(t_1, \dots, t_m) \left\{ \mathbb{E} \prod_{k=1}^m \sum_{i=1}^n \mathbf{1}_{|\xi_i| > t_k} \right\}^{1/2} dt_1 \cdots dt_m \\
 &\leq K_m \int_{\mathbb{R}_{\geq 0}^m} A_{2,n}(t_1, \dots, t_m) \left(\sum_{1 \leq i_1, \dots, i_m \leq n} \prod_{k=1}^m \mathbb{P}(|\xi_{i_k}| > t_k)^{1/m} \right)^{1/2} dt_1 \cdots dt_m
 \end{aligned}$$

$$= n^{m/2} K_m \int_{\mathbb{R}_{\geq 0}^m} A_{2,n}(t_1, \dots, t_m) \prod_{k=1}^m \mathbb{P}(|\xi_1| > t_k)^{1/2m} dt_1 \cdots dt_m.$$

Here

$$A_{2,n}(t_1, \dots, t_m) \equiv \left\{ \mathbb{E} \left[a^2 \left(\sum_{i=1}^n \mathbf{1}_{|\xi_i| > t_1}, \dots, \sum_{i=1}^n \mathbf{1}_{|\xi_i| > t_m} \right) \right] \right\}^{1/2} \rightarrow 0$$

as long as none of $\{\mathbb{P}(|\xi_1| > t_k) : 1 \leq k \leq m\}$ vanishes. The claim now follows from dominated convergence theorem.

Corollary 6.2 (de la Peña et al., 1999). *Let $X(t), t \in T$, be a (weak) Gaussian or Rademacher chaos process of degree m and let*

$$d_X(s, t) := [\mathbb{E}(X(t) - X(s))^2]^{1/2}, s, t \in T.$$

If

$$\int_0^D (\log N(T, d_X, \varepsilon))^{m/2} d\varepsilon < \infty.$$

Then there is a version of X , which we keep denoting as X , with almost all of its sample paths in $C_u(T, d_X)$ and such that

$$\left\| \sup_{t \in T} |X(t)| \right\|_{\psi_{2/m}} \leq \|X(t_0)\|_{\psi_2} + K \int_0^D (\log(N(T, d_X, \varepsilon)))^{m/2} d\varepsilon,$$

and

$$\left\| \sup_{\substack{d_X(s,t) \leq \delta \\ s,t \in T}} |X(t) - X(s)| \right\|_{\psi_{2/m}} \leq K \int_0^\delta (\log(N(T, d_X, \varepsilon)))^{m/2} d\varepsilon,$$

for all $0 < \delta \leq D$, where K is a universal constant and D is the diameter of T for the pseudodistance d_X and ψ_2 defined in the last reference. In fact, every separable version of X satisfies these properties.

Theorem 6.3 (Dudley, 1990). *For any random elements Y_n with values in a metric space (S, d) , where Y is measurable and has separable range, the following are equivalent:*

1. Y_n converge in law to Y .
2. $d_{\text{BL}}(Y_n, Y) \rightarrow 0$ as $n \rightarrow \infty$.

Chapter 6

Weak Convergence of the Conditional U -processes for Locally Stationary Functional Time Series

This chapter develops the content of a submitted article with the required modifications to fit this thesis manuscript.

Objective

In recent years, the direction has turned to non-stationary time series. Here the situation is more complicated: it is often unclear how to set down a meaningful asymptotic for non-stationary processes. For this reason, the theory of locally stationary processes arose, and it is based on infill asymptotics created from non-parametric statistics. The current chapter aims to develop a framework for inference of locally stationary functional time series based on the so-called conditional U -statistics introduced by W. Stute [Ann. Probab. 19 (1991) 812–825], and may be viewed as a generalization of the Nadaraya-Watson regression function estimates. In this chapter, we are mainly interested in establishing weak convergence of conditional U -processes in the locally stationary functional mixing data framework. More precisely, we investigate the weak convergence of conditional U -processes when the explicative variable is functional. We treat the weak convergence when the class of functions is bounded or unbounded, satisfying some moment conditions. These results are established under fairly general structural conditions on the classes of functions and the underlying models. The theoretical results established in this chapter are (or will be) critical tools for further functional data analysis developments.

Contents

1	Introduction	200
---	------------------------	-----

2	Background and preliminaries	204
2.1	Notations	204
2.2	Model	204
2.3	Local stationarity	205
2.4	Mixing conditions	206
2.5	Kernel estimation	207
2.6	Small ball probability	208
2.7	VC-type classes of functions	208
2.8	Assumptions	210
2.9	Comments	213
3	Uniform convergence rates for kernel estimators	215
3.1	Hoeffding's decomposition	215
3.2	Uniform convergence rate	217
4	Weak convergence for kernel estimators	217
5	Applications	219
5.1	Discrimination	219
5.2	Metric learning	220
6	Mathematical Developments	221

1 Introduction

Historically, the first asymptotic results of U -statistics for the case of independent and identically distributed random variables were provided by Halmos, 1946, v. Mises, 1947 and Hoeffding, 1948, among others. While for the asymptotic results under the weak dependency assumptions, we may refer to Borovkova et al., 2001, in Denker et al., 1983 or more recently in Leucht, 2012 and more broadly in Leucht et al., 2013, and Bouzebda et al., 2019b, 2020. For an excellent reference resources for U -statistics and U -processes, the reader concerned may refer to A. J. Lee, 1990, Koroljuk et al., 1994, Borovskikh, 1996, Arcones et al., 1993, Arcones, Chen, et al., 1994 and Arcones et al., 1995. A profound insight into the theory of U -processes is given by de la Peña et al., 1999. They have an even more pivotal role in developing the theory of U -processes by following patterns from the theory of empirical process and adding new techniques such as decoupling inequality and randomization. U -processes are powerful tools for a broad range of statistical applications, such as testing for qualitative features of functions in nonparametric statistics (see Abrevaya et al., 2005, Ghosal et al., 2000, S. Lee et al., 2009), cross-validation for density estimation (see Nolan et al., 1987), and establishing limiting distributions of M-estimators (see, e.g., Arcones et al., 1993, de la Peña et al., 1999, Sherman, 1993, Sherman, 1994). In a wide variety of problems that received a good deal of attention in the machine learning literature and ranging from clustering to image recognition through ranking or learning on graphs, natural estimates of the risk are not basic sample means but take the form U -statistics, refer to Cléménçon et al., 2016. The ranking problem can be considered a pairwise classification and the empirical ranking error of any given prediction rule is a U -statistic of order 2, refer to Cléménçon et al., 2008, just like the within-cluster point scatter in cluster analysis, considered in Cléménçon, 2014, or empirical performance measures in metric learning, for instance, refer to Cao et al., 2016. For the U -statistics with random kernels of diverging orders we refer to Frees, 1989, Rempala et al., 1999, C. Heilig et al., 2001 and Y. Song et al., 2019. Infinite-order U -statistics are useful tools for constructing simultaneous prediction intervals that quantify the uncertainty of ensemble methods such as subbagging and random forests; for more details on the subject, refer to Peng et al., 2022. The MeanNN approach estimation for differential entropy introduced by Faivishevsky et al., 2009 is a particular of the U -statistic. Using U -statistics, Q. Liu et al., 2016 proposed a new test statistic for goodness-of-fit tests. Cléménçon, 2014 defined a measure by U -statistics to quantify the clustering quality of a partition. Cybis et al., 2018 have explored a model-free approach for clustering and classification of genetic data based on U -statistics that leads to alternative ways of looking at these problems. Cybis et al., 2018 were motivated by the fact that the U -statistics are versatile enough to be applied to a wide variety of genetic problems and adaptable enough to consider the specificities of different data types. Lim et al., 2013 proposed the use of

the U -statistics, in a natural way, for analyzing random compressed sensing matrices in the non-asymptotic regime.

Stute, 1991 introduced a class of estimators for $r^{(m)}(\varphi, \mathbf{t})$, called conditional U -statistics, which may be viewed as a generalization of the Nadaraya-Watson (Nadaraya, 1964-Watson, 1964) estimates of a regression function. Nonparametric density (Akaike, 1954, Rosenblatt, 1956 and Parzen, 1962) and regression function estimation has been the subject of intense investigation by both statisticians and probabilistic for many years, and this has guided the development of a large variety of methods. Kernel nonparametric function estimation methods have long attracted a great deal of attention. For good sources of references to the research literature in this area, along with statistical applications, consult B. W. Silverman, 1986, Nadaraya, 1989, Härdle, 1990, Wand et al., 1995, Eggermont et al., 2001, Devroye et al., 2001 and the references therein. In the present work, we will investigate problems related to the nonparametric conditional U -statistics. For the reader's convenience, we first introduce Stute's estimators. Let us consider a regular 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 chapter, we are primarily concerned with the estimation of the conditional expectation or regression function :

$$r^{(m)}(\varphi, \mathbf{t}) = \mathbb{E}(\varphi(Y_1, \dots, Y_m) \mid (\mathbf{X}_1, \dots, \mathbf{X}_m) = \mathbf{t}), \quad \text{for } \mathbf{t} \in \mathbb{R}^{dm}, \quad (1.1)$$

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 :

$$\sup_{\mathbf{x} \in \mathbb{R}^d} |K(\mathbf{x})| =: \kappa < \infty \quad \text{and} \quad \int_{\mathbb{R}^d} K(\mathbf{x}) d\mathbf{x} = 1. \quad (1.2)$$

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 :

$$\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n) = \frac{\sum_{(i_1, \dots, i_m) \in I_n^m} \varphi(Y_{i_1}, \dots, Y_{i_m}) K\left(\frac{\mathbf{t}_1 - \mathbf{X}_{i_1}}{h_n}\right) \dots K\left(\frac{\mathbf{t}_m - \mathbf{X}_{i_m}}{h_n}\right)}{\sum_{(i_1, \dots, i_m) \in I_n^m} K\left(\frac{\mathbf{t}_1 - \mathbf{X}_{i_1}}{h_n}\right) \dots K\left(\frac{\mathbf{t}_m - \mathbf{X}_{i_m}}{h_n}\right)}, \quad (1.3)$$

where

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

is the set of all m -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 $nh_n^{dm} \rightarrow \infty$. In the particular case $m = 1$, the $r^{(m)}(\varphi, t)$ is reduced to

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

and Stute's estimator becomes the Nadaraya-Watson estimator of $r^{(1)}(\varphi, t)$. The work of A. Sen, 1994 was devoted to estimating 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 et al., 1995, the limit distributions of $\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$ are discussed and compared with those obtained by Stute. Harel et al., 1996 extend the results of Stute, 1991, under appropriate mixing conditions, to weakly dependent data (see also Basu et al., 2002) and have applied their findings to verify the Bayes risk consistency of the corresponding discrimination rules in the same spirit as in Stute, 1994a and Section 5.1. Stute, 1996 proposed symmetrized nearest neighbor conditional U -statistics as alternatives to the usual kernel-type estimators; we may also refer to Bouzebda et al., 2021. Fu, 2012 had considered the functional conditional U -statistic and had established its finite-dimensional asymptotic normality. Despite the importance of the subject, nonparametric estimation of the conditional U -statistics in a functional data framework has received relatively scant attention. Some recent advances are given in Bouzebda et al., 2021; Bouzebda and Nemouchi, 2020, where the authors consider the problems related to the uniform bandwidth consistency in a general setting. Jadhav et al., 2019 considered the test of independence in the functional framework based on the Kendall statistics, which can be considered as particular cases of the U -statistics. Extending the above exploration to conditional empirical U -processes in the functional setting is practically useful and technically more challenging. There are two perspectives on conditional U -processes: 1) they are infinite-dimensional versions of conditional U -statistics (with one kernel); 2) they are stochastic processes that are nonlinear generalizations of conditional empirical processes. Both views are useful in that: 1) statistically, it is of greater interest to consider a rich class of statistics rather than a single statistic; 2) mathematically, we can borrow the insights from empirical process theory to derive limit or approximation theorems for U -processes. Significantly, 1) extending U -statistics to U -processes requires substantial efforts and different techniques; and 2) generalization from conditional empirical processes to conditional U -processes is highly nontrivial.

Stationarity condition is frequently assumed in time series modeling producing various models, techniques, research, and methodologies. However, the stationarity assumption is not always advantageous for modeling Spatio-temporal data, even with detrending and deseasonalization. Some important time series models are not stationary, frequently seen in many physical phenomena and economic data. Therefore, the stationarity assumption is violated, making the classical methods unsuitable. To overcome this problem, R. A. Silverman, 1957 generalized the notion of the stationary process to the so-called *locally*

stationary random process. As its name indicates, this type of process approximates a nonstationary process by a stationary one locally over short stretches of time. The intuitive idea of local stationarity is also discussed in the works of Priestley, 1965, Dahlhaus, 1997, Neumann et al., 1997, Sakiyama et al., 2004, Dahlhaus and Polonik, 2006, among many others. It is worth mentioning here that the pioneering work of Dahlhaus represents a solid basis for the inference of locally stationary processes. Besides the generalization of stationary processes, this new approach removes time-varying parameters. During the last decade, the theory of empirical processes for locally stationary times series has received much attention since the empirical processes theory rocketed in the resolution of statistical problems and propagated in the time series analysis and regression estimation, to mention just a few references Dahlhaus et al., 2009, Vogt, 2012, and recently Mayer et al., 2020 and Phandoidaen et al., 2022. The extension of the above exploration to conditional empirical U -processes is of great interest from both practical and theoretical points of view. More precisely, we consider the conditional U -process, indexed by a class of functions in the functional data setting. We present an excerpt from Aneiros et al., 2019: *"Functional data analysis (FDA) is a branch of statistics concerned with the analysis of infinite-dimensional variables such as curves, sets, and images. It has undergone phenomenal growth over the past 20 years, stimulated in part by major advances in data collection technology 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 FDA, the reader is referred to the books of Ramsay et al., 2002, Ferraty et al., 2006. These references provide basic analysis methods and various case studies in several areas, including criminology, economics, archaeology, and neurophysiology. It should be noted that the extension of probability theory to random variables taking values in normed spaces (e.g., Banach and Hilbert spaces), including extensions of certain classical asymptotic limit theorems, predates the recent literature on functional data; the reader is referred to the books Araujo et al., 1980. Gasser et al., 1998 have considered density and mode estimation for data-taking values in a normed vector space. In the context of regression estimation, nonparametric models were considered by Ferraty et al., 2006; for recent references, see Mohammedi et al., 2021 and Almanjahie, Bouzebda, Kaid, et al., 2022a. We may refer also to Bosq, 2000, Horváth et al., 2012, Ling et al., 2018. Recently, modern empirical process theory was used to treat functional data. Ferraty et al., 2010 provided the uniform consistency rates of some functionals of the conditional distribution, including the regression function, the conditional cumulative distribution, and the conditional density. Kara-Zaitri et al., 2017 also gave consistency rates for some functional nonparametric models, including the regression function, the conditional distribution, the conditional density, and the conditional hazard function, uniformly in bandwidth (UIB consistency). Bouzebda, 2016 provide several limiting law results for the conditional

model in the functional setting for ergodic data. One can refer to recent advances to Bouzebda and Chaouch, 2022 and Almanjahie, Bouzebda, Chikr Elmezouar, et al., 2022.

The main focus of this chapter is to consider a general framework of the conditional U -process of arbitrary fixed order indexed by a class of functions in a nonparametric framework. More precisely, we consider the conditional U -process in the setting of functional covariate by taking into account a possible nonstationary behavior of the functional time series. The main aim of the present work is to provide a first full theoretical investigation in this setting. This requires the effective application of large sample theory techniques, which were developed for the empirical processes and U -empirical processes.

The organization of this chapter is as follows. In Section 2, we describe our framework and provide the necessary definitions and some explanations together with technical assumptions. Sections 3 and 4 are devoted to the main results. In particular, Section 3 gives the results concerning the rate of convergence where we recall the instrumental tool of the Hoeffding decomposition. Section 4 contains weak convergence results. Some applications are given in Section 5. We provide all the proofs in Section 6. Finally, some technical properties and lemmas are given in Appendix 6.

2 Background and preliminaries

2.1 Notations

Throughout the article, we write $a_n \lesssim b_n$ if there exists a constant C independent of n whose numeric value may change from line to line, unless otherwise specified, such that $a_n \leq Cb_n$ for all n , and $a_n \ll b_n$ if $a_n/b_n \rightarrow 0$ as $n \rightarrow \infty$. For all n , if $a_n \lesssim b_n$ and $b_n \lesssim a_n$, we will write $a_n \sim b_n$. We will also denote (i_1, \dots, i_m) by \mathbf{i} and $(i_1/n, \dots, i_m/n)$ by \mathbf{i}/n . For any $c, d \in \mathbb{R}$, we write $c \vee d = \max\{c, d\}$ and $c \wedge d = \min\{c, d\}$. $[a]$ will denote the integer part of a number. For $m < n$ two positive integer, let $C_m^n = \frac{n!}{(n-m)!m!}$. Let $I_n^m := \{\mathbf{i} = (i_1, \dots, i_m) : 1 \leq i_j \leq n \text{ and } i_j \neq i_r \text{ if } j \neq r\}$, is the set of all m -tuples of different integers between 1 and n .

2.2 Model

Let $\{Y_{i,n}, X_{i,n}\}_{i=1}^n$ be stochastic processes, where $Y_{i,n}$ is in some space \mathcal{Y} and $X_{i,n}$ takes values in some abstract space \mathcal{H} . We will assume that \mathcal{H} a semi-metric vector space with semi-metric $d(\cdot, \cdot)$ ¹, while in most applications, this space would be Hilbert or Banach space with $d(u, v) = \|u - v\|$, for all u and $v \in \mathcal{H}$. Let $\varphi(\cdot)$ be any function of k -variables (the U -kernel) such that $\varphi(Y_1, \dots, Y_m)$ is integrable. For $\mathbf{x} = (x_1, \dots, x_m) \in \mathcal{H}^m$, define

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

the regression functional parameter

$$r^{(m)}(\varphi, \frac{\mathbf{i}}{n}, \mathbf{x}) := \mathbb{E}(\varphi(Y_1, \dots, Y_m) \mid X_1 = x_1, \dots, X_m = x_m); \quad \mathbf{i} = (1, \dots, m) \quad (2.1)$$

In this study, we consider the following model

$$\varphi(Y_{\mathbf{i},n}) = r^{(m)}\left(\varphi, \frac{\mathbf{i}}{n}, X_{\mathbf{i},n}\right) + \sigma\left(\frac{\mathbf{i}}{n}, X_{\mathbf{i},n}\right) \varepsilon_{\mathbf{i}}, \mathbf{i} = (i_1, \dots, i_m); 1 \leq i_j \leq n, \quad (2.2)$$

where $\{\varepsilon_i\}_{i \in \mathbb{Z}}$ is a sequence of independent and identically distributed random variables, independent of $\{X_{i,n}\}_{i=1}^n$. We denote $\sigma\left(\frac{\mathbf{i}}{n}, X_{\mathbf{i},n}\right) \varepsilon_{\mathbf{i}}$ by $\varepsilon_{\mathbf{i},n}$. We also assume that $\{X_{i,n}\}$ is a locally stationary functional time series. The regression function $r^{(m)}(\varphi, \cdot)$ is allowed to change smoothly over time, depending on a rescaled quantity \mathbf{i}/n not on the point \mathbf{i} (\mathbf{i} usually represent the time in time series framework).

2.3 Local stationarity

We were considering non-stationary processes with dynamics that change slowly over time and may thus be behaved as stationarity at a local level. So the process $\{X_{i,n}\}$ can be approximated in a stochastic sense by a stationary process $\{X_{i,n}^{(u)}\}$ around each rescaled time point u , that is for those i where $i/n - u$ is small. As we are interested in the functional data, we call a functional time series locally stationary if a stationary functional time series can locally approximate it. A standard definition of local stationarity will be given in the following.

Definition 2.1 (Local stationarity). *A sequence of stochastic processes, indexed by $n \in \mathbb{N}$ and taking values in \mathcal{H} , $\{X_{i,n}\}$ is locally stationary if for all rescaled times $u \in [0, 1]$, there exists an associated \mathcal{H} -valued process $\{X_i^{(u)}\}$ strictly stationary such that*

$$d\left(X_{i,n}, X_i^{(u)}\right) \leq \left(\left|\frac{i}{n} - u\right| + \frac{1}{n}\right) U_{i,n}^{(u)} \text{ a.s.}, \quad (2.3)$$

for all $1 \leq i \leq n$, where $\{U_{i,n}^{(u)}\}$ is a positive valued process satisfying $\mathbb{E}[(U_{i,n}^{(u)})^\rho] < C$ for some $\rho > 0$, $C < \infty$ that is independent of u, i , and n .

van Delft et al., 2018 give this definition with \mathcal{H} is a Hilbert space $L_{\mathbb{R}}^2[0, 1]$, and all real-valued functions are square-integrable with respect to the Lebesgue measure on the interval $[0, 1]$ with the inner product L_2 -norm:

$$\|f\|_2 = \sqrt{\langle f, f \rangle}, \quad \langle f, g \rangle = \int_0^1 f(t)g(t)dt,$$

where $f, g \in L_{\mathbb{R}}^2([0, 1])$. The authors also give sufficient conditions so that an $L_{\mathbb{R}}^2([0, 1])$ -valued stochastic process $\{X_{i,n}\}$ satisfies (2.3) with $d(f, g) = \|f - g\|_2$ and $\rho = 2$.

Remark 2.2. *van Delft et al., 2018* generalize the definition of local stationary processes, given by *Dahlhaus, 1996b*, to the functional setting in the frequency domain. Under the following assumptions:

(A1) (i) $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is a weakly stationary white noise process taking values in H with spectral representation $\varepsilon_t = \int_{-\pi}^{\pi} e^{i\omega t} dZ_{\omega}$, where Z_{ω} is a 2π -periodic orthogonal increment process taking values in $H_{\mathbb{C}}$; (ii) the functional process $X_{t,T}$ with $t = 1, \dots, T$ and $T \in \mathbb{N}$ is given by

$$X_{t,T} = \int_{-\pi}^{\pi} e^{i\omega t} \mathcal{A}_{t,\omega}^{(T)} dZ_{\omega} \quad \text{a.e. in } H$$

with transfer operator $\mathcal{A}_{t,\omega}^{(T)} \in \mathcal{B}_p$ and an orthogonal increment process Z_{ω} .

(A2) There exists $\mathcal{A} : [0, 1] \times [-\pi, \pi] \rightarrow S_p(H_{\mathbb{C}})$ with $\mathcal{A}_{u,\cdot} \in \mathcal{B}_p$ and $\mathcal{A}_{u,\omega}$ being continuous in u such that for all $T \in \mathbb{N}$

$$\sup_{\omega, t} \left\| \mathcal{A}_{t,\omega}^{(T)} - \mathcal{A}_{\frac{t}{T}, \omega} \right\|_p = O\left(\frac{1}{T}\right).$$

They had proved in *van Delft et al., 2018*, Proposition 2.2 that:

Proposition 2.3. *Suppose that assumptions (A1) and (A2) hold. Then $\{X_{i,n}\}$ is a locally stationary process in \mathcal{H} .*

2.4 Mixing conditions

Statistical observations are frequently not independent but close to being so. Dependence can have powerful consequences for statistical inference. *Mixing* is a term that illustrates how close a sequence of random variables is to being independent. It may be used to extend standard findings for independent sequences to weakly dependent or mixing sequences. The development of *the theory of mixing conditions* has been motivated by the fact that a time series seems to have some "asymptotic independence" properties and is used to handle the analysis and statistical inference of the latter. Let us define the type of dependency used in this work. Let (Ω, \mathcal{F}, P) be a probability space. Let $\mathbf{Z}_1, \mathbf{Z}_2, \dots$ be a stationary sequence of random variables on some probability space $(\Omega, \mathcal{D}, \mathbb{P})$ and let σ_i^j be the σ -field generated by $\mathbf{Z}_i, \dots, \mathbf{Z}_j$, for $i, j \geq 1$. Moreover, for an array $\{Z_{i,n} : 1 \leq i \leq n\}$, define the coefficients

$$\beta(k) = \sup_{i, n: 1 \leq i \leq n-k} \beta(\sigma(Z_{s,n}, 1 \leq s \leq i), \sigma(Z_{s,n}, i+k \leq s \leq n)),$$

where $\sigma(Z)$ is the σ -field generated by Z . The array $\{Z_{i,n}\}$ is said to be β -mixing if $\beta(k) \rightarrow 0$. Keep in mind that β -mixing implies α -mixing.

Throughout the sequel, we assume that the sequence of random elements $\{(X_{i,n}, Y_{i,n}), i = 1, \dots, n; n \geq 1\}$ is absolutely regular. The Markov chains, for instance, are β -mixing under the milder Harris recurrence condition if the underlying space is finite [Davydov, 1973]. In the concluding remarks, see Section 4, we give more motivation to the choice of the regular process rather than the strong mixing processes.

2.5 Kernel estimation

We are interested in estimating the regression function represented in (2.1). The kernel estimator is defined to be

$$\tilde{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}; h_n) = \frac{\sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d(x_k, X_{i_k, n})}{h_n} \right) \right\} \varphi(Y_{\mathbf{i}, n})}{\sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d(x_k, X_{i_k, n})}{h_n} \right) \right\}}, \quad (2.4)$$

where $K_1(\cdot)$ and $K_2(\cdot)$ denote one-dimensional kernel functions. Here, $h = h_n$ is a bandwidth satisfying $h \rightarrow 0$ as $n \rightarrow \infty$ and $\varphi : \mathcal{Y}^m \rightarrow \mathbb{R}$ is a symmetric measurable function belonging to some class of functions \mathcal{F}_m . This estimator is a conditional U -statistic based on the sequence of random variables $\{Y_{i,n}, X_{i,n}\}_{i=1}^n$ and of kernel $\varphi \times K_1 \times K_2$. It is well known that this type of statistics has been introduced by Stute, 1991. We need to introduce some notation to investigate the weak convergence of the conditional empirical process and the conditional U -process in the functional data setting. Let

$$\mathcal{F}_m = \{\varphi : \mathcal{Y}^m \rightarrow \mathbb{R}\},$$

denote a point-wise measurable class of real-valued symmetric measurable functions on \mathcal{Y}^m with a measurable envelope function :

$$F(\mathbf{y}) \geq \sup_{\varphi \in \mathcal{F}_m} |\varphi(\mathbf{y})|, \text{ for } \mathbf{y} \in \mathcal{Y}^m. \quad (2.5)$$

For a kernel function $K(\cdot)$, we define the point-wise measurable class of functions, for $1 \leq m \leq n$

$$\mathcal{K}^m := \left\{ \mathbf{x} \mapsto \prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{k, n})}{h} \right), h > 0 \text{ and } \mathbf{x} \in \mathcal{K}^m \right\}.$$

The conditional U -process indexed by $\mathcal{F}_m \mathcal{K}^m$:

$$\left\{ \mathbb{G}_n(\varphi, \mathbf{x}) := \sqrt{nh^m \phi^m(h_n)} \left(\tilde{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}; h_n) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \right) \right\}_{\mathcal{F}_m \mathcal{K}^m}. \quad (2.6)$$

Notice that point-wise measurability is crucial in our setting. In fact, it allows us to express our claimed results in their usual form under the classical definition of the probability notion avoiding the abstract notion of outer probability or outer expectation A. W. van der Vaart et al., 1996.

Remark 2.4. *The bandwidth h is the same in each direction to make things easier while dealing with product kernels. However, the findings may be easily adjusted for the scenario of non-product kernels and varying bandwidths.*

Remark 2.5. *Our estimator differs from the classical one in the conditional U -statistics not only in the type of the sequence $\{X_i\}_i$ but also in the addition of a kernel in the time direction, so we have smoothness from the direction of the covariate $(X_{i,n})$ and from the time direction, which allows us to cover the characteristic that a regression model varies over time.*

2.6 Small ball probability

One of the technical difficulties in infinite-dimensional spaces is the lack of a universal reference measure, such as the Lebesgue measure, which causes the non-existence of the density function of the functional variable. We use the usual “small-ball probability” notion to overcome this difficulty. Specifically, controlling the concentration of probability measure of the functional variable on a small ball is done by the function $\phi_x(\cdot)$, which is defined, for a fixed $x \in \mathcal{H}$ for all $r > 0$, by

$$\mathbb{P}(X \in B(x, r)) =: \phi_x(r) > 0, \quad (2.7)$$

where the abstract space \mathcal{H} is equipped with the semi-metric $d(\cdot, \cdot)$ and

$$B(x, r) = \{y \in \mathcal{H} : d(x, y) \leq r\}$$

is a ball in \mathcal{H} with the center $x \in \mathcal{H}$ and radius r .

2.7 VC-type classes of functions

The asymptotic analysis of functional data is related to some concentration properties expressed in terms of the small-ball probability notion; when considering a process indexed by a class of functions, one needs to take into account other topological concepts: metric entropy, VC-subgraph classes (“VC” for Vapnik and Červonenkis).

Definition 2.6. Let $\mathcal{S}_\mathcal{E}$ be a subset of a semi-metric space \mathcal{E} , a finite set of points $\{e_1, \dots, e_N\} \subset \mathcal{E}$ is called, for a given $\varepsilon > 0$, a ε -net of $\mathcal{S}_\mathcal{E}$ if :

$$\mathcal{S}_\mathcal{E} \subseteq \cup_{j=1}^N B(e_j, \varepsilon).$$

If $N_\varepsilon(\mathcal{S}_\mathcal{E})$ is the cardinality of the smallest ε -net (the minimal number of open balls of radius ε) in \mathcal{E} , needed to cover $\mathcal{S}_\mathcal{H}$, then we call Kolmogorov's entropy (metric entropy) of the set $\mathcal{S}_\mathcal{E}$, the quantity

$$\psi_{\mathcal{S}_\mathcal{E}}(\varepsilon) := \log N_\varepsilon(\mathcal{S}_\mathcal{E}).$$

From its name, one can figure that Kolmogorov (cf. Kolmogorov et al., 1959) introduced this concept of metric entropy, which was subsequently studied for numerous metric spaces. This concept was used by Dudley, 1967 to give sufficient conditions for the continuity of Gaussian processes. It was the basis for striking generalizations of Donsker's theorem on the weak convergence of the empirical process. Suppose that $\mathcal{B}_\mathcal{H}$ and $\mathcal{I}_\mathcal{H}$ are two subsets of the semi-metric space \mathcal{H} with Kolmogorov's entropy (for the radius ε) $\psi_{\mathcal{B}_\mathcal{H}}(\varepsilon)$ and $\psi_{\mathcal{I}_\mathcal{H}}(\varepsilon)$ respectively, then the Kolmogorov entropy for the subset $\mathcal{B}_\mathcal{H} \times \mathcal{I}_\mathcal{H}$ of the semi-metric space \mathcal{H}^2 by :

$$\psi_{\mathcal{B}_\mathcal{H} \times \mathcal{I}_\mathcal{H}}(\varepsilon) = \psi_{\mathcal{B}_\mathcal{H}}(\varepsilon) + \psi_{\mathcal{I}_\mathcal{H}}(\varepsilon).$$

Hence, $m\psi_{\mathcal{I}_\mathcal{H}}(\varepsilon)$ is the Kolmogorov entropy of the subset $\mathcal{I}_\mathcal{H}^m$ of the semi-metric space \mathcal{H}^m . Noting that if we designate by d the semi-metric on \mathcal{H} then, we can define the semi metric on \mathcal{H}^m by :

$$d_{\mathcal{H}^m}(\mathbf{x}, \mathbf{z}) := \frac{1}{m}d(x_1, z_1) + \dots + \frac{1}{m}d(x_m, z_m) \quad (2.8)$$

for

$$\mathbf{x} = (x_1, \dots, x_m), \mathbf{z} = (z_1, \dots, z_m) \in \mathcal{H}^m.$$

Notice that the semi-metric plays an essential role in this kind of study. The reader will find useful discussions about how to choose the semi-metric in Ferraty et al., 2006, Chapters 3 and 11.

We also need to deal with another topological concept which is VC-subgraph classes.

Definition 2.7. 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 2.8. 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).

A VC-class of functions \mathcal{F} with envelope function F have the following entropy property, for a given $1 \leq q < \infty$, there are constants a and b such as :

$$N(\varepsilon, \mathcal{F}, \|\cdot\|_{L_q(Q)}) \leq a \left(\frac{(QF^q)^{1/q}}{\varepsilon} \right)^b \quad (2.9)$$

for any $\varepsilon > 0$ and each probability measure such that $QF^q < \infty$. For instance, see Nolan et al., 1987, Lemma 22, Dudley, 2014, §4.7. A. W. van der Vaart et al., 1996, Theorem 2.6.7, Kosorok, 2008, §9.1 provide a number of sufficient conditions under which (2.9) holds, we may refer also to Deheuvels, 2011, §3.2 for further discussions.

2.8 Assumptions

For the reader's convenience, we have gathered the essential assumptions in the following.

Assumption 1. [Model and distribution assumptions]

- i) The process $\{X_{i,n}\}$ is locally stationary satisfy that for each time point $u \in [0, 1]$, there exists a stationary process $\{X_i^{(u)}\}$ such that

$$d(X_{i,n}, X_i^{(u)}) \leq \left(\left| \frac{i}{n} - u \right| + \frac{1}{n} \right) U_{i,n}^{(u)} \text{ a.s.},$$

with $\mathbb{E}[(U_{i,n}^{(u)})^\rho] < C$ for some $\rho > 0$, $C < \infty$

- ii) Let $B(x, h)$ be a ball centered at $x \in \mathcal{H}$ with radius h , defined in Section 2.6, and let $c_d < C_d$ be positive constants. For all $\mathbf{u} \in [0, 1]^m$,

$$0 < c_d \phi^m(h) f_1(\mathbf{x}) \leq \mathbb{P} \left((X_{i_1}^{(u_1)}, \dots, X_{i_m}^{(u_m)}) \in \mathbf{B}(\mathbf{x}, h) \right) =: F_u(h; \mathbf{x}) \leq C_d \phi^m(h) f_1(\mathbf{x}), \quad (2.10)$$

where $\phi(0) = 0$ and $\phi(u)$ is absolutely continuous in a neighborhood of the origin,

$f_1(x)$ is a non-negative functional in $x \in \mathcal{H}$ and

$$\mathbf{B}(\mathbf{x}, h) = \prod_{i=1}^m B(x_i, h).$$

iii) There exist constants $C_\phi > 0$ and $\varepsilon_0 > 0$ such that for any $0 < \varepsilon < \varepsilon_0$,

$$\int_0^\varepsilon \phi(u) du > C_\phi \varepsilon \phi(\varepsilon). \quad (2.11)$$

iv) Let $\psi(h) \rightarrow 0$ as $h \rightarrow 0$, and $f_2(\mathbf{x})$ is a non-negative functional in $\mathbf{x} := (x_1, \dots, x_m) \in \mathcal{H}^m$

$$\sup_{\mathbf{i} \in I_n^m} \mathbb{P} \left(\left((X_{i_1, n}, \dots, X_{i_m, n}), (X_{i'_1, n}, \dots, X_{i'_m, n}) \right) \in \mathbf{B}(\mathbf{x}, h) \times \mathbf{B}(\mathbf{x}, h) \right) \leq \psi^m(h) f_2(\mathbf{x}).$$

We will also assume that the ratio $\psi(h)/\phi^2(h)$ is bounded.

Assumption 2. [Kernel assumptions]

i) $K_1(\cdot)$ is a symmetric kernel around zero, bounded, and has a compact support, that is, $K_1(v) = 0$ for all $|v| > C_1$ for some $C_1 < \infty$. Moreover,

$$\int K_1(z) dz = 1$$

and $K_1(\cdot)$ is Lipschitz continuous, that is,

$$|K_1(v_1) - K_1(v_2)| \leq C_2 |v_1 - v_2|$$

for some $C_2 < \infty$ and all $v_1, v_2 \in \mathbb{R}$.

ii) The kernel $K_2(\cdot)$ is non-negative, bounded, and has a compact support in $[0, 1]$ such that $0 < K_2(0)$ and $K_2(1) = 0$. We can also see $K_2(\cdot)$ as an asymmetrical triangle kernel, that is, $K_2(x) = (1 - x)\mathbb{1}_{(x \in [0, 1])}$, and $K_2(\cdot)$ is Lipschitz continuous, that is,

$$|K_2(v_1) - K_2(v_2)| \leq C_2 |v_1 - v_2|.$$

Moreover, $K'_2(v) = dK_2(v)/dv$ exists on $[0, 1]$ and for two real constants $-\infty < C'_1 < C'_2 < 0$, we have:

$$C'_2 \leq K'_2(v) \leq C'_1.$$

Assumption 3. [Smoothness]

i) $r^{(m)}(\mathbf{u}, \mathbf{x})$ is twice continuously partially differentiable with respect to \mathbf{u} . We also assume that

$$\sup_{\mathbf{u}_1, \mathbf{u}_2 \in [0,1]^m} |r^{(m)}(\mathbf{u}_1, \mathbf{x}) - r^{(m)}(\mathbf{u}_2, \mathbf{z})| \leq c_m (d_{\mathcal{H}^m}(\mathbf{x}, \mathbf{z})^\alpha + \|\mathbf{u}_1 - \mathbf{u}_2\|^\alpha) \quad (2.12)$$

for some $c_m > 0$ and $\alpha > 0$ and the semi-metric $d_{\mathcal{H}^m}(\mathbf{x}, \mathbf{z})$ is defined on \mathcal{H}^m by:

$$d_{\mathcal{H}^m}(\mathbf{x}, \mathbf{z}) := \frac{1}{m}d(x_1, z_1) + \cdots + \frac{1}{m}d(x_m, z_m)$$

for $\mathbf{x} = (x_1, \dots, x_m), \mathbf{z} = (z_1, \dots, z_m) \in \mathcal{H}^m$.

ii) $\sigma : [0, 1] \times \mathcal{H}^m \rightarrow \mathbb{R}$ is bounded by some constant $C_\sigma < \infty$ from above and by some constant $c_\sigma > 0$ from below, that is, $0 < c_\sigma \leq \sigma(\mathbf{u}, \mathbf{x}) \leq C_\sigma < \infty$ for all \mathbf{u} and \mathbf{x} .

iii) $\sigma(\cdot, \cdot)$ is Lipschitz continuous with respect to \mathbf{u} .

iv) $\sup_{\mathbf{u} \in [0,1]^m} \sup_{z: d(\mathbf{x}, \mathbf{z}) \leq h} |\sigma(\mathbf{u}, \mathbf{x}) - \sigma(\mathbf{u}, \mathbf{z})| = o(1)$ as $h \rightarrow 0$.

Let $W_{i,\varphi,n}$ be an array of one-dimensional random variables. In this study, this array will be equal to $W_{i,\varphi,n} = 1$ and $W_{i,\varphi,n} = \varepsilon_{i,n}$.

Assumption 4. [Mixing].

i) It holds that $\sup_{\mathbf{x} \in \mathcal{H}^m} \mathbb{E}|W_{i,n}|^\zeta \leq C$ and $\sup_{\mathbf{x} \in \mathcal{H}^m} \mathbb{E}[|W_{i,n}|^\zeta | X_{i,n} = \mathbf{x}] \leq C$ for $\zeta > 2$ and $C < \infty$.

ii) The β -mixing coefficients of the array $\{X_{i,n}, W_{i,n}\}$ satisfy $\beta(k) \leq Ak^{-\gamma}$ for some $A > 0$ and $\gamma > 2$. We also assume that $\delta + 1 < \gamma(1 - \frac{2}{\nu})$ for some $\nu > 2$ and $\delta > 1 - \frac{2}{\nu}$, and

$$h^{2(1 \wedge \alpha) - 1} \left(\phi(h)a_n + \sum_{k=a_n}^{\infty} k^\delta (\beta(k))^{1 - \frac{2}{\nu}} \right) \rightarrow 0, \quad (2.13)$$

as $n \rightarrow \infty$, where $a_n = [(\phi(h))^{-(1 - \frac{2}{\nu})/\delta}]$ and for all $\alpha > 0$.

iii) Let $\alpha_n = \sqrt{\frac{\log n}{nh^m \phi(h)}}$. As $n \rightarrow \infty$, $\frac{(\log n)^{\frac{-m+\gamma+1}{2} + \zeta_0(\gamma+1)}}{n^{\frac{-m+\gamma+1}{2} - 1 - \frac{\gamma+1}{\zeta}} h^{\frac{m+\gamma+1}{2}} \phi(h)^{\frac{-m+\gamma+1}{2}}} \rightarrow 0$ for some $\zeta_0 > 0$.

iv) $nh^2, nh\phi(h) \rightarrow \infty$,

Assumption 5 (Blocking assumptions). There exists a sequence of positive integers $\{v_n\}$ satisfying $v_n \rightarrow \infty$, $v_n = o(\sqrt{nh\phi(h)})$ and $\sqrt{\frac{n}{h\phi(h)}}\beta(v_n) \rightarrow \infty$ as $n \rightarrow \infty$.

Assumption 6. [Class of functions assumptions]

The classes of functions \mathcal{K}^m and \mathcal{F}_m are such that :

- i) The class of functions \mathcal{F}_m is bounded and its envelope function satisfies for some $0 < M < \infty$:

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

- ii) 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 ν such that:

$$N(\varepsilon, \mathcal{F}_m \mathcal{K}^m, \|\cdot\|_{L_2(Q)}) \leq \left(\frac{b \|F \kappa^m\|_{L_2(Q)}}{\varepsilon} \right)^\nu$$

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

- iii) The class of functions \mathcal{F}_m is unbounded and its envelope function satisfies for some $p > 2$:

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

- iv) The metric entropy of the class $\mathcal{F} \mathcal{K}$ satisfies, for some $2 < p < \infty$:

$$\begin{aligned} \int_0^\infty (\log N(u, \mathcal{F} \mathcal{K}, \|\cdot\|_1))^{\frac{1}{2}} du &< \infty, \\ \int_0^\infty (\log N(u, \mathcal{F} \mathcal{K}, \|\cdot\|_2))^{\frac{1}{2}} du &< \infty, \\ \int_0^\infty (\log N(u, \mathcal{F} \mathcal{K}, \|\cdot\|_p))^{\frac{1}{2}} du &< \infty. \end{aligned}$$

2.9 Comments

Some comments on the assumptions are in order. Most of these assumptions are inspired by Gasser et al., 1998, Masry, 2005, Ferraty et al., 2006, Vogt, 2012 and Kurisu, 2022a. In Assumption 1, we begin by formalizing the property of X_i to be locally stationary, and we continue by some conditions on the distribution behavior of the variables. Equation (2.10) controls the behavior of the small ball probability around zero and is the usual condition on the small ball probability. This approximately shows that the small ball probability can be written approximately as the product of two independent functions $\phi^m(\cdot)$ and $f_1(\cdot)$, for instance, for $m = 1$, refer to Mayer-Wolf et al., 1993 for the diffusion process, Bogachev, 1998 for a Gaussian measure, W. V. Li et al., 2001 for a general Gaussian process and these assumptions have been employed by Masry, 2005 for strongly mixing processes. For example, the function $\phi(\cdot)$ can be expressed as $\phi(\varepsilon) = \varepsilon^\delta \exp(-C/\varepsilon^a)$ with $\delta \geq 0$ and $a \geq 0$, and it corresponds to the Ornstein–Uhlenbeck and general diffusion processes (for

such processes, $a = 2$ and $\delta = 0$) and the fractal processes (for such processes, $\delta > 0$ and $a = 0$). We refer to the paper of Ferraty et al., 2007 for other examples. Assumptions 2 are the kernel assumptions. We can see that the kernel $K_2(\cdot)$ is a symmetric one belonging to the kernel's family of type II , which contains continuous kernels (triangle, quadratic...). It also assumes that the kernels are compactly supported on $[0, 1]$ to obtain an expression for the asymptotic variance. The *Lipschitz type* assumptions on $K_2(\cdot)$ and $\sigma(\cdot, \cdot)$ (Assumption 2ii) and Assumption 3iii)) are essential to obtain the rate of convergence. Assumption 3 is an essential one, where it requires limiting the growth of $r^{(m)}(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$. The latter is a restriction on the function $r^{(m)}(\cdot, \cdot)$, $\sigma(\cdot, \cdot)$ and their derivations by preventing them from growing too fast outside a large bound. Assumption 4 ii) is a standard assumption of mixing condition. Assumption 4 iii) and iv) are technical conditions to obtain the desired results. Assumption 6 asserts that we deal with a class of functions satisfying some entropy conditions. Part ii) and iii) of these assumptions are both connected to the class of functions, but the first one declares that this class is bounded. Nevertheless, as we are interested in demonstrating the functional central limit theorem for conditional U -processes indexed by unbounded class of function, part iii) substitutes the first one in this case.

All these general assumptions are sufficiently weak relative to the different objects involved in the statement of our main results. They cover and exploit the principal axes of this contribution, which are the topological structure of the functional variables, the probability measure in this functional space, the measurability concept on the class of function, and the uniformity controlled by the entropy properties.

Remark 2.9. *Note that the Assumption 6 iii) may be replaced by more general hypotheses upon moments of \mathbf{Y} as in Deheuvels, 2011. That is*

iii)" 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 $x \uparrow \infty$,

$$x^{-s} \mathcal{M}(x) \downarrow; \quad x^{-1} \mathcal{M}(x) \uparrow. \quad (2.14)$$

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:

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

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 Uniform convergence rates for kernel estimators

Before expressing the asymptotic behavior of our estimator represented in (2.4), we will generalize the study to a U -statistic estimator defined by:

$$\widehat{\psi}(\mathbf{u}, \mathbf{x}, \varphi) = \frac{(n-m)!}{n!h^m\phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h} \right) K_2 \left(\frac{d(x_k, X_{i_k,n})}{h} \right) \right\} W_{\mathbf{i}, \varphi, n}, \quad (3.1)$$

where $W_{\mathbf{i}, \varphi, n}$ is an array of one-dimensional random variables. In this study, we use the results with $W_{\mathbf{i}, \varphi, n} = 1$ and $W_{\mathbf{i}, \varphi, n} = \varepsilon_{\mathbf{i}, n}$.

3.1 Hoeffding's decomposition

Note $\widehat{\psi}(\mathbf{u}, \mathbf{x}, \varphi)$ is a classical U -statistic with a kernel depending on n . We define

$$\begin{aligned} \xi_k &:= \frac{1}{h} K_1 \left(\frac{u_k - k/n}{h} \right), \\ H(Z_1, \dots, Z_m) &:= \prod_{k=1}^m \frac{1}{\phi(h)} K_2 \left(\frac{d(x_k, X_{k,n})}{h} \right) W_{\mathbf{i}, \varphi, n}, \end{aligned}$$

thus, the U -statistic in (3.1) can be viewed as a weighted U -statistic of degree m :

$$\widehat{\psi}(\mathbf{u}, \mathbf{x}, \varphi) = \frac{(n-m)!}{n!} \sum_{\mathbf{i} \in I_n^m} \xi_{i_1} \dots \xi_{i_m} H(Z_{i_1}, \dots, Z_{i_m}). \quad (3.2)$$

We can write Hoeffding's decomposition in this case. If we will not assume symmetry for $W_{\mathbf{i}, \varphi, n}$ or H , we must define:

- The expectation of $H(Z_{i_1}, \dots, Z_{i_m})$:

$$\theta(\mathbf{i}) := \mathbb{E}H(Z_{i_1}, \dots, Z_{i_m}) = \int W_{\mathbf{i}, \varphi, n} \prod_{k=1}^m \frac{1}{\phi(h)} K_2 \left(\frac{d(x_k, \nu_{k,n})}{h} \right) d\mathbb{P}_{\mathbf{i}}(z_{\mathbf{i}}). \quad (3.3)$$

- For all $\ell \in \{1, \dots, m\}$ the position of the argument, construct the function π_ℓ such that:

$$\pi_\ell(z; z_1, \dots, z_{m-1}) := (z_1, \dots, z_{\ell-1}, z, z_\ell, \dots, z_{m-1}). \quad (3.4)$$

- Define:

$$H^{(\ell)}(z; z_1, \dots, z_{m-1}) := H\{\pi_\ell(z; z_1, \dots, z_{m-1})\} \quad (3.5)$$

$$\theta^{(\ell)}(i; i_1, i_2, \dots, i_{m-1}) := \theta\{\pi_\ell(i; i_1, i_2, \dots, i_{m-1})\}. \quad (3.6)$$

Hence, the first order expansion of $H(\cdot)$ will be seen as:

$$\begin{aligned}
 \widetilde{H}^{(\ell)}(z) &:= \mathbb{E} \left(H^{(\ell)}(z, Z_1, \dots, Z_{m-1}) \right) \\
 &= \int W_{(1, \dots, \ell-1, i, \ell, \dots, m-1)} \prod_{\substack{k=1 \\ k \neq i}}^{m-1} \frac{1}{\phi(h)} K_2 \left(\frac{d(x_k, \nu_k)}{h} \right) \times \frac{1}{\phi(h)} K_2 \left(\frac{d(x_i, \nu_i)}{h} \right) \\
 &\quad \mathbb{P}(d\nu_1, \dots, d\nu_{\ell-1}, d\nu_{\ell}, \dots, d\nu_{m-1}) \\
 &:= \frac{1}{\phi(h)} K_2 \left(\frac{d(x_i, x)}{h} \right) \times w \times \int W_{(1, \dots, \ell-1, \ell, \dots, m-1)} \prod_{\substack{k=1 \\ k \neq i}}^{m-1} \frac{1}{\phi(h)} K_2 \left(\frac{d(x_k, \nu_k)}{h} \right) \\
 &\quad \mathbb{P}(d\nu_1, \dots, d\nu_{\ell-1}, d\nu_{\ell}, \dots, d\nu_{m-1}),
 \end{aligned} \tag{3.7}$$

with \mathbb{P} is the underlying probability measure, and define

$$f_{i, i_1, \dots, i_{m-1}}^{(\ell)} := \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_i \xi_{i_{\ell}} \cdots \xi_{i_{m-1}} \left(\widetilde{H}^{(\ell)}(z) - \theta^{(\ell)}(i; i_1, \dots, i_{m-1}) \right). \tag{3.8}$$

Then, the first-order projection can be defined as:

$$\widehat{H}_{1,i}(\mathbf{u}, \mathbf{x}, \varphi) := \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} f_{i, i_1, \dots, i_{m-1}}^{(\ell)}, \tag{3.9}$$

where

$$I_{n-1}^{m-1}(-i) := \{1 \leq i_1 < \dots < i_{m-1} \leq n \text{ and } i_j \neq i \text{ for all } j \in \{1, \dots, m-1\}\}.$$

For the remainder terms, we denote by $\mathbf{i} \setminus i_{\ell} := (i_1, \dots, i_{\ell-1}, i_{\ell+1}, \dots, i_m)$ and for $\ell \in \{1, \dots, m\}$, let

$$H_{2,\mathbf{i}}(\mathbf{z}) := H(\mathbf{z}) - \sum_{\ell=1}^m \widetilde{H}_{\mathbf{i} \setminus i_{\ell}}^{(\ell)}(z_{\ell}) + (m-1)\theta(\mathbf{i}), \tag{3.10}$$

where

$$\widetilde{H}_{\mathbf{i} \setminus i_{\ell}}^{(\ell)}(z_{\ell}) = \mathbb{E} \left(H(Z_1, \dots, Z_{\ell-1}, z, Z_{\ell+1}, \dots, Z_{m-1}) \right),$$

defined in (3.7), this projection derive us to the following remainder term:

$$\widehat{\psi}_{2,\mathbf{i}}(\mathbf{u}, \mathbf{x}, \varphi) := \frac{(n-m)!}{(n)!} \sum_{\mathbf{i} \in I_n^m} \xi_{i_1} \cdots \xi_{i_m} H_{2,\mathbf{i}}(\mathbf{z}). \tag{3.11}$$

Finally, using Equation (3.9) and Equation (3.11), and under conditions that :

$$\mathbb{E} \left(\widehat{H}_{1,i}(\mathbf{u}, X, \varphi) \right) = 0, \tag{3.12}$$

$$\mathbb{E} (H_{2,\mathbf{i}}(\mathbf{Z} \mid Z_k)) = 0 \text{ a.s.}, \tag{3.13}$$

we get the Hoeffding, 1948 decomposition:

$$\begin{aligned}\widehat{\psi}(\mathbf{u}, \mathbf{x}, \varphi) - \mathbb{E}(\widehat{\psi}(\mathbf{u}, \mathbf{x}, \varphi)) &= \frac{1}{n} \sum_{i=1}^n \widehat{H}_{1,i}(\mathbf{u}, \mathbf{x}, \varphi) + \widehat{\psi}_{2,i}(\mathbf{u}, \mathbf{x}, \varphi) \\ &:= \widehat{\psi}_{1,i}(\mathbf{u}, \mathbf{x}, \varphi) + \widehat{\psi}_{2,i}(\mathbf{u}, \mathbf{x}, \varphi).\end{aligned}$$

For more details, the interested reader can refer to F. Han et al., 2018, Lemma 2.2.

3.2 Uniform convergence rate

We start by giving the following general result.

Proposition 3.1. *Let $\mathcal{F}_m \mathcal{K}^m$ be a measurable VC-subgraph class of functions such as Assumption 6 is satisfied and assume that Assumptions 1, 2, 3, 4 are also satisfied. Then the following result holds*

$$\sup_{\mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in [0,1]^m} |\widehat{\psi}(\mathbf{u}, \mathbf{x}, \varphi) - \mathbb{E}[\widehat{\psi}(\mathbf{u}, \mathbf{x}, \varphi)]| = O_{\mathbb{P}} \left(\sqrt{\frac{\log n}{nh^m \phi^m(h)}} \right).$$

Based on the previous proposition, the uniform convergence rate of the kernel estimator $\widehat{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}; h_n)$ can be investigated. It should be essential to mention that when $m = 1$ and fixed function φ , we obtain the same results as the pointwise convergence rate of the regression function of a strictly stationary functional time series in Ferraty et al., 2006. The following theorem (2.4) will generalize the uniform convergence rate to the estimator.

Theorem 3.1

Let $\mathcal{F}_m \mathcal{K}^m$ be a measurable VC-subgraph class of functions such as Assumption 6 is satisfied and assume that Assumptions 1, 2, 3 and 4 are satisfied. Then we have

$$\begin{aligned}\sup_{\mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in [C_1 h, 1 - C_1 h]^m} |\widehat{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}; h_n) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x})| \\ = O_{\mathbb{P}} \left(\sqrt{\frac{\log n}{nh^m \phi^m(h)}} + h^{2m \wedge \alpha} \right).\end{aligned}\tag{3.14}$$

4 Weak convergence for kernel estimators

In this section, we are interested in studying the weak convergence of the conditional U -processes under absolute regular observations. Observe that

$$\begin{aligned}\widehat{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}; h_n) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) &= \frac{1}{\widehat{r}_1(\varphi, \mathbf{u}, \mathbf{x})} \left(\widehat{g}_1(\mathbf{u}, \mathbf{x}) + \widehat{g}_2(\mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{x}, \mathbf{u}) \widehat{r}_1(\varphi, \mathbf{u}, \mathbf{x}) \right) \\ &= \frac{1}{\widehat{r}_1(\varphi, \mathbf{u}, \mathbf{x})} \left(\widehat{g}_1(\mathbf{u}, \mathbf{x}) + \widehat{g}^B(\mathbf{u}, \mathbf{x}) \right),\end{aligned}\tag{4.1}$$

where

$$\begin{aligned}\tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}) &= \frac{(n-m)!}{n!h^m\phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1\left(\frac{u_k - i_k/n}{h}\right) K_2\left(\frac{d(x_k, X_{i_k, n})}{h}\right) \right\}, \\ \hat{g}_1(\mathbf{u}, \mathbf{x}) &= \frac{(n-m)!}{n!h^m\phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1\left(\frac{u_k - i_k/n}{h}\right) K_2\left(\frac{d(x_k, X_{i_k, n})}{h}\right) \right\} W_{\mathbf{i}, \varphi, n}, \\ \hat{g}_2(\mathbf{u}, \mathbf{x}) &= \frac{(n-m)!}{n!h^m\phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1\left(\frac{u_k - i_k/n}{h}\right) K_2\left(\frac{d(x_k, X_{i_k, n})}{h}\right) \right\} r^{(m)}\left(\frac{\mathbf{i}}{n}, X_{i, n}\right).\end{aligned}$$

Under the same assumption in Theorem 3.1, we will show in the next theorem that $\text{Var}(\hat{g}^B(\mathbf{u}, \mathbf{x})) = o\left(\frac{1}{nh^m\phi^m(h)}\right)$ and $1/\tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}) = O_{\mathbb{P}}(1)$. Then, we have

$$\tilde{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}; h_n) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) = \frac{\hat{g}_1(\mathbf{u}, \mathbf{x})}{\tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x})} + B_n(\mathbf{u}, \mathbf{x}) + o_{\mathbb{P}}\left(\sqrt{\frac{1}{nh^m\phi^m(h)}}\right),$$

where $B_n(\mathbf{u}, \mathbf{x}) = \mathbb{E}[\hat{g}^B(\mathbf{u}, \mathbf{x})]/\mathbb{E}[\tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x})]$ is the ‘‘bias’’ term and $\frac{\hat{g}_1(\mathbf{u}, \mathbf{x})}{\tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x})}$ is the ‘‘variance’’ term. Let us define, for $\varphi_1, \varphi_2 \in \mathcal{F}_m$

$$\begin{aligned}\sigma(\varphi_1, \varphi_2) &= \lim_{n \rightarrow \infty} \mathbb{E}\left(\sqrt{nh^m\phi^m(h)}(\tilde{r}_n^{(m)}(\varphi_1, \mathbf{u}, \mathbf{x}; h_n) - r^{(m)}(\varphi_1, \mathbf{u}, \mathbf{x}))\right. \\ &\quad \left. \times \sqrt{nh^m\phi^m(h)}(\tilde{r}_n^{(m)}(\varphi_2, \mathbf{u}, \mathbf{x}; h_n) - r^{(m)}(\varphi_2, \mathbf{u}, \mathbf{x}))\right).\end{aligned}\quad (4.2)$$

In the following, we would set $K_2(\cdot)$ as the asymmetrical triangle kernel, that is, $K_2(x) = (1-x)\mathbf{1}_{(x \in [0,1])}$ to simplify the proof. The main results of this section are given in the following theorems.

Theorem 4.1

Let $\mathcal{F}_m \mathcal{H}^m$ be a measurable VC-subgraph class of functions, and assume that all the assumptions of Section 2.8 are satisfied for both case $W_{\mathbf{i}, \varphi, n} = 1$ and $W_{\mathbf{i}, \varphi, n} = \varepsilon_{i, n}$. Then as $n \rightarrow \infty$, for any $\mathbf{x} \in \mathcal{H}^m$ and $\mathbf{u} \in [0, 1]^m$, the U-process

$$\sqrt{nh^m\phi^m(h)}(\tilde{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}; h_n) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) - B_n(\mathbf{u}, \mathbf{x}))$$

converges to a Gaussian process G_n over $\mathcal{F}_m \mathcal{H}^m$, whose simple paths are bounded and informally continuous with respect to $\|\cdot\|_2$ -norm with covariance function given in (4.2).

If we want to investigate the weak convergence of our estimator in the standard way, i.e., using the well-known steps of Hoeffding decomposition and then finite-dimensional convergence and the equicontinuity, we can follow the next theorem. In the proof of the

following theorem, we will represent the conditional U -process in terms of the U -process based on a stationary sequence that will be proved to converge to some Gaussian process.

Theorem 4.2

Let $\mathcal{F}_m \mathcal{K}^m$ be a measurable VC-subgraph class of functions, and assume that all the assumptions of Section 2.8 are satisfied. Then as $n \rightarrow \infty$, for any $\mathbf{x} \in \mathcal{H}^m$ and $\mathbf{u} \in [0, 1]^m$:

$$\sqrt{nh^m \phi^m(h)} \left(\hat{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}; h_n) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \right)$$

converges in law to a Gaussian process $\{\mathbb{G}_n(\psi) : \psi \in \mathcal{F}_m \mathcal{K}^m\}$ that admits a version with uniformly bounded and uniformly continuous paths with respect to $\|\cdot\|_2$ -norm with covariance function given in (4.2).

5 Applications

Although only the following examples will be given here, they stand as archetypes for a variety of problems that can be investigated in a similar way.

5.1 Discrimination

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

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

where

$$\mathfrak{M}^j(\mathbf{x}) = \mathbb{P}(\varphi(\mathbf{Y}) = j \mid \mathbf{X} = \mathbf{x}), \quad \mathbf{x} \in \mathcal{H}^m.$$

The above inequality becomes equality if

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

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

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

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 $1 \leq j \leq M$,

$$\mathfrak{M}_n^j(\mathbf{x}) = \frac{\sum_{\mathbf{i} \in I_n^m} \mathbf{1}\{\varphi(Y_{i_1}, \dots, Y_{i_m}) = j\} \prod_{k=1}^m \left\{ K_1 \left(\frac{u - i_k/n}{h} \right) K_2 \left(\frac{d(x, X_{i_k, n})}{h} \right) \right\}}{\sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u - i_k/n}{h} \right) K_2 \left(\frac{d(x, X_{i_k, n})}{h} \right) \right\}},$$

Set

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

Let us introduce

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

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

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

This follows from the apparent relation:

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

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 $\mathcal{H} \times \mathcal{Y}$ valued random couple (X, Y) , where \mathcal{H} 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 $D : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}_+$. Intuitively, the aim 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

$$R(D) = \mathbb{E}[\phi((1 - D(X, X')) \cdot (2\mathbf{1}\{Y = Y'\} - 1))], \quad (5.1)$$

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

$$R_n(D) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \phi((D(X_i, X_j) - 1) \cdot (2\pi \{Y_i = Y_j\} - 1)), \quad (5.2)$$

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 (5.1) of a minimizer of (5.2) has been studied in the frameworks of algorithmic stability (Jin et al., 2009), algorithmic robustness (Bellet et al., 2015) and based on the theory of U -processes under appropriate regularization (Cao et al., 2016).

6 Mathematical Developments

This section is devoted to the proof of our results. The previously presented notation continues to be used in the following. We begin this section with the following lemma before providing the proofs of the main results.

Lemma 6.1. *Let $K_2(\cdot)$ denote one dimensional kernel function satisfying Assumption 2 part i), if Assumption 1, then:*

$$i) \quad \mathbb{E} \left| \prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) - \prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \right| \lesssim \frac{m\phi^{m-1}(h)}{nh}. \quad (6.1)$$

$$ii) \quad \mathbb{E} \left[\prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right] \lesssim \frac{m\phi^{m-1}(h)}{nh} + \phi^m(h). \quad (6.2)$$

$$iii) \quad \mathbb{E} \left[\prod_{k=1}^m K_2^2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \right] \sim \phi^m(h). \quad (6.3)$$

The proof of this lemma is given below.

Proof of Lemma 6.1 For the first inequality $i)$, by assuming that the kernel function $K_2(\cdot)$ is an asymmetrical triangle kernel, that is, $K_2(x) = (1 - x)\mathbf{1}_{(x \in [0, 1])}$, we have

$$\mathbb{E} \left| \prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) - \prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \right|$$

$$\begin{aligned}
 &= \mathbb{E} \left[\left| \sum_{k=1}^m \left(K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) - K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \right) \right| \right. \\
 &\quad \times \left. \left| \prod_{i=1}^{k-1} K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right| \times \left| \prod_{j=k+1}^m K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \right| \right] \\
 &\hspace{15em} \text{(Using a telescoping argument)} \\
 &\leq \left\{ \mathbb{E} \left[\left| \sum_{k=1}^m \left(K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) - K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \right) \right| \right]^3 \right\}^{1/3} \\
 &\quad \times \left\{ \mathbb{E} \left[\left| \prod_{i=1}^{k-1} K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right|^3 \right] \right\}^{1/3} \times \left\{ \mathbb{E} \left[\left| \prod_{j=k+1}^m K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \right|^3 \right] \right\}^{1/3} \\
 &\hspace{15em} \text{(By Hölder's inequality)} \\
 &\leq \left\{ \mathbb{E} \left[\sum_{k=1}^m \left| \left(K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) - K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \right) \right|^3 \right] \right\}^{1/3} \\
 &\quad \times \left\{ \prod_{i=1}^{k-1} \left(\mathbb{E} \left| K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right|^{3p_i} \right)^{1/p_i} \right\}^{1/3} \\
 &\quad \times \left\{ \prod_{j=k+1}^m \mathbb{E} \left(\left| K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \right|^{3q_j} \right)^{1/q_j} \right\}^{1/3} \hspace{5em} \text{(By Hölder's inequality)} \\
 &\lesssim \left\{ \sum_{k=1}^m \mathbb{E} \left| \frac{1}{nh} U_{i_k, n}^{(i_k/n)} \right|^3 \right\}^{1/3} \times \left\{ \prod_{i=1}^{k-1} \left(\mathbb{E} \left| \mathbb{1}_{\{d(x_k, X_{i_k, n}) \leq h\}} \right|^{3p_i} \right)^{1/p_i} \right. \\
 &\quad \times \left. \left. \prod_{j=k+1}^m \left(\mathbb{E} \left| \mathbb{1}_{\{d(x_k, X_{i_k, n}^{(i_k/n)}) \leq h\}} \right|^{3q_j} \right)^{1/q_j} \right\}^{1/3} \hspace{5em} \text{(By Assumption 1)} \\
 &\lesssim \left\{ \sum_{k=1}^m \frac{1}{n^3 h^3} \mathbb{E} \left| U_{i_k, n}^{(i_k/n)} \right|^3 \right\}^{1/3} \times \left\{ \prod_{i=1}^{k-1} \left(F^{3p_i}(h, x_k) \right)^{1/p_i} \prod_{j=k+1}^m \left(F^{3q_j}(h, x_k) \right)^{1/q_j} \right\}^{1/3} \\
 &\lesssim \frac{1}{nh} \left\{ \sum_{k=1}^m \mathbb{E} \left| U_{i_k, n}^{(i_k/n)} \right|^3 \right\}^{1/3} \times \left\{ \prod_{i=1}^{k-1} C_d \phi^3(h) f_1^3(x_k) \times \prod_{j=k+1}^m C_d \phi(h)^3 f_1^3(x_k) \right\}^{1/3} \\
 &\lesssim \frac{m \phi^{m-1}(h)}{nh}. \hspace{15em} (6.4)
 \end{aligned}$$

For the second inequality *ii*), we have:

$$\mathbb{E} \left[\prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right]$$

$$= \mathbb{E} \left[\prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) - \prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) + \prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \right].$$

By linearity of the expectation, inequality *i*) and for

$$\mathbb{E} \left[\prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \right] \lesssim \phi^{m-1}(h),$$

using Assumption 1 part iv), the proof of this inequality holds. Now, we consider the last one. Set $\widetilde{K}_2^2(x_k) := K_2^2(d(x_k, X_{i_k, n}^{(i_k/n)}))$:

$$\begin{aligned} \mathbb{E} \left[\left(\prod_{k=1}^m \widetilde{K}_2^2 \left(\frac{x_k}{h} \right) \right) \right] &= \int_0^h \cdots \int_0^h \prod_{k=1}^m \widetilde{K}_2^2 \left(\frac{y_k}{h} \right) \mathbb{P}(dy_1, \dots, dy_k) \\ &= -\frac{2}{h} \int_0^h \cdots \int_0^h \prod_{j=1, j \neq k}^m \widetilde{K}_2(y_j) \\ &\quad \times \int_0^h \widetilde{K}_2(y_\ell) \widetilde{K}_2'(y_\ell) \mathbb{P}(dy_1, \dots, dy_{\ell-1}, y_\ell, dy_{\ell+1}, \dots, y_k) dy_\ell \quad (\text{Integration by parts}) \\ &= \frac{(-2)^m}{h^m} \int_0^h \cdots \int_0^h \prod_{k=1}^m \widetilde{K}_2(y_k) \widetilde{K}_2'(y_k) \mathbb{P}(y_1, \dots, y_k) dy_1 \dots dy_k \\ &\sim \frac{2^m}{h^m} \int_0^h \cdots \int_0^h \prod_{k=1}^m \widetilde{K}_2(y_k) \widetilde{K}_2'(y_k) \phi_k(y_k) d\mathbb{P}(y_1, \dots, y_k) \\ &= \frac{2^m}{h^m} \int_0^h \cdots \int_0^h \prod_{k=1}^m \left(1 - \frac{y_k}{h} \right) \phi_k(y_k) d\mathbb{P}(y_1, \dots, y_k) \\ &\quad (\text{Using Assumption 2 ii) and } K_2(x) = (1-x)I(x \in [0, 1])) \\ &= \frac{2^m}{h^{2m}} \int_0^h \cdots \int_0^h \left(\int_0^y \cdots \int_0^y \prod_{k=1}^m \phi_k(\varepsilon_k) d(\varepsilon_1, \dots, \varepsilon_k) \right) d\mathbb{P}(y_1, \dots, y_k) \\ &\quad (\text{By an integration by parts}) \\ &\sim \frac{2^m}{h^{2m}} \int_0^h \cdots \int_0^h \prod_{k=1}^m y_k \phi_k(y_k) d\mathbb{P}(y_1, \dots, y_k) \sim \frac{1}{h^{2m}} \phi^m(h) h^{2m} \sim \phi^m(h). \end{aligned}$$

The final result holds by using the small ball lower bound given in (2.11). Hence, inequality (6.3) follows. \square

Proof of Proposition 3.1

As previously mentioned, our statistic is a weighted U -statistic that can be decomposed to a sum of U -statistics using the Hoeffding decomposition. We will treat this decomposition detailed in the Sub-section 3.1 to achieve the desired results. In the mentioned section,

we have seen that

$$\widehat{\psi}(\mathbf{u}, \mathbf{x}, \varphi) - \mathbb{E}(\widehat{\psi}(\mathbf{u}, \mathbf{x}, \varphi)) = \widehat{\psi}_{1,\mathbf{i}}(\mathbf{u}, \mathbf{x}, \varphi) + \widehat{\psi}_{2,\mathbf{i}}(\mathbf{u}, \mathbf{x}, \varphi),$$

where the linear term $\widehat{\psi}_{1,\mathbf{i}}(\mathbf{u}, \mathbf{x}, \varphi)$ and the remainder term $\widehat{\psi}_{2,\mathbf{i}}(\mathbf{u}, \mathbf{x}, \varphi)$ are well defined in (3.9) and (3.11) respectively. We aim to prove that the linear term leads this statistic's convergence rate while the remaining one converges to zero, almost sure, as $n \rightarrow \infty$. We will begin treating the first term in the decomposition.

For $B = [0, 1]$, $\alpha_n = \sqrt{\log n/nh^m\phi^m(h)}$ and $\tau_n = \rho_n n^{1/\zeta}$, where ζ is a positive constant that appears in Assumption 4 part i), with $\rho_n = (\log n)^{\zeta_0}$ for some $\zeta_0 > 0$. Define

$$\widetilde{H}_1^{(\ell)}(z) := \widetilde{H}^{(\ell)}(z) \mathbb{1}_{\{|W_{i,n}| \leq \tau_n\}}, \quad (6.5)$$

$$\widetilde{H}_2(z) := \widetilde{H}^{(\ell)}(z) \mathbb{1}_{\{|W_{i,n}| > \tau_n\}}, \quad (6.6)$$

and

$$\begin{aligned} \widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) - \theta(\mathbf{i}) &= \frac{1}{n} \sum_{i=1}^n \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_{\ell} \xi_{i_{\ell}} \cdots \xi_{i_{m-1}} \widetilde{H}_1^{(\ell)}(z), \\ \widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}, \varphi) - \theta(\mathbf{i}) &= \frac{1}{n} \sum_{i=1}^n \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_{\ell} \xi_{i_{\ell}} \cdots \xi_{i_{m-1}} \widetilde{H}_2^{(\ell)}(z). \end{aligned}$$

Clearly, we have

$$\begin{aligned} &\widehat{\psi}_{1,\mathbf{i}}(\mathbf{u}, \mathbf{x}, \varphi) - \mathbb{E}\widehat{\psi}_{1,\mathbf{i}}(\mathbf{u}, \mathbf{x}, \varphi) \\ &= \left[\widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) - \mathbb{E}\widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) \right] + \left[\widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}, \varphi) - \mathbb{E}\widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}, \varphi) \right]. \end{aligned} \quad (6.7)$$

First, we can see that

$$\begin{aligned} &\mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}, \varphi) - \theta(\mathbf{i}) \right| > \alpha_n \right) \\ &= \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}, \varphi) - \theta(\mathbf{i}) \right| > \alpha_n \right) \\ &\quad \cap \left\{ \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \bigcup_{i=1}^n |W_{i,n}| > \tau_n \right\} \cup \left\{ \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \left\{ \bigcup_{i=1}^n |W_{i,n}| > \tau_n \right\}^c \right\} \\ &\leq \mathbb{P} \left\{ \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}, \varphi) - \theta(\mathbf{i}) \right| > \alpha_n \right. \\ &\quad \left. \cap \left\{ \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \bigcup_{i=1}^n |W_{i,n}| > \tau_n \right\} \right\} \\ &\quad + \mathbb{P} \left\{ \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}, \varphi) \right| > \alpha_n \right. \end{aligned}$$

$$\begin{aligned}
 & \cap \left\{ \left\{ \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \bigcup_{i=1}^n |W_{i,n}| > \tau_n \right\}^c \right\} \\
 & \leq \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |W_{i,n}| > \tau_n \text{ for some } i = 1, \dots, n \right) + \mathbb{P}(\emptyset) \\
 & \leq \tau_n^{-\zeta} \sum_{i=1}^n \mathbb{E} \left[\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |W_{i,n}|^\zeta \right] \leq n \tau_n^{-\zeta} = \rho_n^{-\zeta} \rightarrow 0.
 \end{aligned}$$

We infer that

$$\mathbb{E} \left[\left| \widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}, \varphi) \right| \right] \leq \frac{1}{n} \sum_{i=1}^n \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_i \xi_{i_\ell} \cdots \xi_{i_{m-1}} \mathbb{E} \left(\left| \widetilde{H}_2^{(\ell)}(z) \right| \right), \quad (6.8)$$

where

$$\begin{aligned}
 \mathbb{E} \left(\left| \widetilde{H}_2^{(\ell)}(z) \right| \right) &= \mathbb{E} \left[\frac{1}{\phi(h)} K_2 \left(\frac{d(x_i, X_{i,n})}{h} \right) W_{i,n} \times \int W_{(1, \dots, \ell-1, \ell, \dots, m)} \right. \\
 & \quad \left. \prod_{\substack{k=1 \\ k \neq i}}^m \frac{1}{\phi(h)} K_2 \left(\frac{d(x_k, \nu_k)}{h} \right) \mathbb{P}(d\nu_1, \dots, d\nu_{\ell-1}, d\nu_\ell, \dots, d\nu_{m-1}) \mathbb{1}_{\{|W_{i,n}| > \tau_n\}} \right] \\
 & \lesssim \tau_n^{-(\zeta-1)} \mathbb{E} \left[\frac{1}{\phi(h)} K_2 \left(\frac{d(x_i, X_{i,n})}{h} \right) |W_{i,n}|^\zeta \right] \\
 & \lesssim \frac{\tau_n^{-(\zeta-1)}}{\phi(h)} \mathbb{E} \left[K_2 \left(\frac{d(x_i, X_{i,n})}{h} \right) \right] \\
 & \lesssim \frac{\tau_n^{-(\zeta-1)}}{\phi(h)} \times \left[\frac{1}{nh} + \phi(h) \right] \\
 & \lesssim \frac{\tau_n^{-(\zeta-1)}}{nh\phi(h)} + \tau_n^{-(\zeta-1)}, \quad (6.9)
 \end{aligned}$$

where

$$\begin{aligned}
 & \mathbb{E} \left(K_2 \left(\frac{d(x_i, X_{i,n})}{h} \right) \right) \\
 &= \mathbb{E} \left(K_2 \left(\frac{d(x_i, X_{i,n})}{h} \right) + K_2 \left(\frac{d(x_i, X_i^{i/n})}{h} \right) - K_2 \left(\frac{d(x_i, X_i^{i/n})}{h} \right) \right) \\
 &\leq \mathbb{E} \left| K_2 \left(\frac{d(x_i, X_{i,n})}{h} \right) - K_2 \left(\frac{d(x_i, X_i^{i/n})}{h} \right) \right| + \mathbb{E} \left| K_2 \left(\frac{d(x_i, X_i^{i/n})}{h} \right) \right| \\
 &\lesssim Ch^{-1} \mathbb{E} \left| d(x_i, X_{i,n}) - d(x_i, X_i^{i/n}) \right| + \mathbb{E} \left[\mathbb{1}_{(d(x, X_{i,n}^{i/n}) \leq h)} \right] \quad (K_2 \text{ is Lipschitz})
 \end{aligned}$$

$$\begin{aligned} &\lesssim \frac{1}{nh} \mathbb{E} \left| U_i^{(i/n)} \right| + F_{i/n}(h; x_i) \text{ (using Assumption 1 i)} \\ &\lesssim \frac{1}{nh} + \phi(h). \end{aligned}$$

Hence we obtain

$$\begin{aligned} \mathbb{E} \left[\left| \widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}, \varphi) \right| \right] &\leq \frac{1}{n} \sum_{i=1}^n \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_i \xi_{i_\ell} \cdots \xi_{i_{m-1}} \mathbb{E} \left(\left| H_2^{(\ell)}(z) \right| \right) \\ &\lesssim \underbrace{\frac{1}{n} \sum_{i=1}^n \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_i \xi_{i_\ell} \cdots \xi_{i_{m-1}}}_{\leq C \text{ uniformly in } \mathbf{u}} \times \left[\frac{\tau_n^{-(\zeta-1)}}{nh\phi(h)} + \tau_n^{-(\zeta-1)} \right] \\ &\lesssim \left[\frac{\tau_n^{-(\zeta-1)}}{nh\phi(h)} + \tau_n^{-(\zeta-1)} \right] \lesssim \tau_n^{-(\zeta-1)} = (\rho_n n^{1/\zeta})^{-(\zeta-1)} \lesssim \alpha_n. \end{aligned}$$

As a result, we infer that

$$\sup_{\mathcal{F}_m} \sup_{\mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}, \varphi) - \mathbb{E} \widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}, \varphi) \right| = O_{\mathbb{P}}(\alpha_n). \quad (6.10)$$

Second, let us treat

$$\sup_{\mathcal{F}_m} \sup_{\mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) - \mathbb{E} \widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) \right|.$$

To achieve the aimed result, we will cover the region $B^m = [0, 1]^m$ by

$$\bigcup_{k_1, \dots, k_m=1}^{N(\mathbf{u})} \prod_{j=1}^m \mathbf{B}(\mathbf{u}_{k_j}, r),$$

for some radius r . Hence, for each $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_m) \in [0, 1]^m$, there exists $\mathbf{l}(\mathbf{u}) = (l(\mathbf{u}_1), \dots, l(\mathbf{u}_m))$, where $\forall 1 \leq i \leq m, 1 \leq l(\mathbf{u}_i) \leq N(\mathbf{u})$ and such that

$$\mathbf{u} \in \prod_{i=1}^m \mathbf{B}(\mathbf{u}_{l(\mathbf{u}_i)}, r) \text{ and } |\mathbf{u}_i - \mathbf{u}_{l(\mathbf{u}_i)}| \leq r, \text{ for } 1 \leq i \leq m,$$

then for each $\mathbf{u} \in [0, 1]^m$, the closest center will be $\mathbf{u}_l(\mathbf{u})$, and the ball with the closest centre will be defined by

$$\mathcal{B}(\mathbf{u}, \mathbf{l}(\mathbf{u}), r) := \prod_{j=1}^m \mathbf{B}(\mathbf{u}_{k_j}, r).$$

In the same way \mathcal{H}^m should be covered by

$$\bigcup_{k_1, \dots, k_m=1}^{N(x)} \prod_{j=1}^m \mathbf{B}(x_{k_j}, r),$$

for some radius r . Hence, for each $\mathbf{x} = (x_1, \dots, x_m) \in \mathcal{H}^m$, there exists $\mathbf{l}(\mathbf{x}) = (l(x_1), \dots, l(x_m))$, where $\forall 1 \leq i \leq m, 1 \leq l(x_i) \leq N(x)$ and such that

$$\mathbf{x} \in \prod_{i=1}^m B(u_{l(x_i)}, r) \text{ and } d(x_i, x_{l(x_i)}) \leq r, \text{ for } 1 \leq i \leq m,$$

then for each $\mathbf{x} \in \mathcal{H}^m$, the closest center will be $\mathbf{x}_1(\mathbf{x})$, and the ball with the closest centre will be defined by

$$\mathbf{B}(\mathbf{x}, \mathbf{l}(\mathbf{x}), r) := \prod_{i=1}^m B(x_{l(x_i)}, r).$$

We define:

$$K^*(\boldsymbol{\omega}, \mathbf{v}) = C \prod_{k=1}^m \mathbb{1}_{(|\omega_k| \leq 2C_1)} \prod_{k=1}^m K_2(v_k) \text{ for } (\boldsymbol{\omega}, \mathbf{v}) \in \mathbb{R}^2.$$

We can show that, for $(u, x) \in B_{j,n}$ and n large enough,

$$\begin{aligned} & \left| \prod_{k=1}^m K_1\left(\frac{u_k - \frac{i_k}{n}}{h}\right) - \prod_{k=1}^m K_1\left(\frac{u_{j,k} - \frac{i_k}{n}}{h}\right) \right| K_2\left(\frac{d(x_i, X_{i,n})}{h}\right) \\ & \leq \alpha_n K^*\left(\frac{u_n - \frac{i}{n}, d(x_i, X_{i,n})}{h}\right). \end{aligned}$$

Let

$$\begin{aligned} \bar{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) &= \frac{1}{nh^m \phi(h)} \sum_{i=1}^n \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} \prod_{k=1}^m K^*\left(\frac{u_k - \frac{i_k}{n}, d(x_{i_k}, X_{i_k,n})}{h}\right) W_{i,n} \\ & \int W_{(1, \dots, \ell-1, \ell, \dots, m)} \prod_{\substack{k=1 \\ k \neq i}}^m \frac{1}{\phi(h)} K_2\left(\frac{d(x_k, \nu_k)}{h}\right) \mathbb{P}(d\nu_1, \dots, d\nu_{\ell-1}, d\nu_{\ell}, \dots, d\nu_{m-1}) \mathbb{1}_{\{|W_{i,n}| \leq \tau_n\}}. \end{aligned}$$

Note that $\mathbb{E} \left[\left| \bar{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) \right| \right] \leq M < \infty$ for some sufficiently large M . Then we obtain

$$\begin{aligned} & \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \hat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) - \mathbb{E} \left[\hat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) \right] \right| \tag{6.11} \\ & \leq \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \left| \hat{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}) - \mathbb{E} \left[\hat{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}) \right] \right| \\ & \quad + \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \alpha_n \left(\left| \bar{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}) \right| + \mathbb{E} \left[\left| \bar{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}) \right| \right] \right) \\ & \leq \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \left| \hat{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}) - \mathbb{E} \left[\hat{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}) \right] \right| \\ & \quad + \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \left| \bar{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}) - \mathbb{E} \left[\bar{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}) \right] \right| + 2MF(\mathbf{y})\alpha_n. \end{aligned}$$

Therefore

$$\begin{aligned}
 & \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) - \mathbb{E} \left[\widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) \right] \right| > 4M\alpha_n \right) \\
 & \leq N_{\mathcal{F}_m \mathcal{H}^m} N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i(x)}, r)} \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i(\mathbf{u})}, r)} \\
 & \quad \mathbb{P} \left(\left| \widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) - \mathbb{E} \left[\widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) \right] \right| > 4M\alpha_n \right) \\
 & \leq Q_{1,n} + Q_{2,n}, \tag{6.12}
 \end{aligned}$$

where, for $N_{\mathcal{F}_m \mathcal{H}^m} N_{(x)}^m N_{(u)}^m$ denotes the covering number related respectively to the class of functions $\mathcal{F}_m \mathcal{H}^m$, the balls that cover $[0, 1]^m$ and the balls that cover \mathcal{H}^m .

$$\begin{aligned}
 Q_{1,n} &= N_{\mathcal{F}_m \mathcal{H}^m} N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i(x)}, r)} \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i(\mathbf{u})}, r)} \\
 & \quad \mathbb{P} \left(\left| \widehat{\psi}_1(\mathbf{u}_j, \mathbf{x}) - \mathbb{E} \left[\widehat{\psi}_1(\mathbf{u}_j, \mathbf{x}) \right] \right| > M\alpha_n \right), \\
 Q_{2,n} &= N_{\mathcal{F}_m \mathcal{H}^m} N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i(x)}, r)} \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i(\mathbf{u})}, r)} \\
 & \quad \mathbb{P} \left(\left| \bar{\psi}_1(\mathbf{u}_j, \mathbf{x}) - \mathbb{E} \left[\bar{\psi}_1(\mathbf{u}_j, \mathbf{x}) \right] \right| > M\alpha_n \right).
 \end{aligned}$$

Notice that $Q_{1,n}$ and $Q_{2,n}$ might be treated in the same way, so, we restrict our attention to $Q_{1,n}$. Write:

$$\begin{aligned}
 & \mathbb{P} \left(\left| \widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) - \mathbb{E} \widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) \right|_{\mathcal{F}_m \mathcal{H}^m} > M\alpha_n \right) \\
 &= \mathbb{P} \left[\left| h^m \phi^m(h) \sum_{i=1}^n \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_i \xi_{i_\ell} \cdots \xi_{i_{m-1}} H_1^{(\ell)}(z) \right. \right. \\
 & \quad \left. \left. - \mathbb{E} \left(h^m \phi^m(h) \sum_{i=1}^n \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_i \xi_{i_\ell} \cdots \xi_{i_{m-1}} H_1^{(\ell)}(z) \right) \right|_{\mathcal{F}_m \mathcal{H}^m} \right. \\
 & \quad \left. > Mn \frac{(n-1)!}{(n-m)!} \alpha_n h^m \phi^m(h) \right] \\
 &= \mathbb{P} \left(\left| \sum_{i=1}^n \Phi_{i,n}(\mathbf{u}, \mathbf{x}, \varphi) \right|_{\mathcal{F}_m \mathcal{H}^m} > Mn \frac{(n-1)!}{(n-m)!} \alpha_n h^m \phi^m(h) \right).
 \end{aligned}$$

Note that the array $\{\Phi_{i,n}(u, x)\}$ is α -mixing for each fixed (u, x) with mixing coefficients $\beta_{\Phi,n}$ such that $\beta_{\Phi,n}(k) \leq \beta(k)$. We apply Lemma 6.6 with

$$\varepsilon := Mn \frac{(n-1)!}{(n-m)!} h^m \phi^m(h) \alpha_n,$$

and $b_n = C\tau_n$ for sufficiently large $C > 0$ and $S_n = \alpha_n^{-1} \tau_n^{-1}$. As same as Masry, 2005,

Theorem 2, we can see that $\sigma_{S_n, n}^2 \leq C' S_n h^m \phi^m(h)$, and we obtain:

$$\begin{aligned}
 & \mathbb{P} \left(\left| \sum_{i=1}^n Z_{i,n}(u, x) \right| \geq \varepsilon \right) \\
 & \leq 4 \exp \left(-\frac{\varepsilon^2}{64 \sigma_{S_n, n}^2 \frac{n}{S_n} + \frac{8}{3} \varepsilon b_n S_n} \right) + 4 \frac{n}{S_n} \beta(S_n) \\
 & \leq 4 \exp \left(-\frac{M^2 \alpha_n^2 n^2 \left(\frac{(n-1)!}{(n-m)!} \right)^2 h^{2m} \phi^{2m}(h)}{64 C' S_n h \phi(h) \frac{n}{S_n} + \frac{8}{3} M n \frac{(n-1)!}{(n-m)!} h^m \phi^m(h) \alpha_n b_n S_n} \right) + 4 \frac{n}{S_n} \beta(S_n) \\
 & \leq 4 \exp \left(-\frac{M \left(\sqrt{\log n / n h^m \phi^m(h)} \right)^2 n \left(\frac{(n-1)!}{(n-m)!} \right)}{64 C' h^m \phi^m(h) \frac{(n-m)!}{M(n-1)!} + \frac{8}{3} C h^m \phi^m(h)} \right) + 4 \frac{n}{S_n} \beta(S_n) \\
 & \lesssim \exp \left(-\frac{M \frac{(n-1)!}{(n-m)!} \log n}{64 \frac{(n-m)!}{(n-1)!} \frac{C'}{M} + \frac{8}{3} C} \right) + n S_n^{-\gamma-1}.
 \end{aligned}$$

To get the last inequality, we must choose $M > C'$. Since $N \leq C h^{-m} \phi(h) \alpha_n^{-m}$, it follows that

$$\widehat{Q}_n \leq O(R_{1n}) + O(R_{2n}),$$

with

$$\begin{aligned}
 R_{1T} &= h^{-m} \alpha_n^{-m} n^{-\frac{M \frac{(n-1)!}{(n-m)!}}{64 \frac{(n-m)!}{(n-1)!} + 3C}} \\
 R_{2T} &= h^{-m} \alpha_n^{-m} n S_n^{-\gamma-1}.
 \end{aligned}$$

For M sufficiently large, we can see that $R_{1n} \leq n^{-\varsigma}$ for some small $\varsigma > 0$. As $\frac{\phi_n \log T}{T^\theta h^{d+1}} = o(1)$ by assumption, we further get that

$$\begin{aligned}
 R_{2n} &= h^{-m} \alpha_n^{-m} n S_n^{-\gamma-1} \\
 &= h^{-m} n \left(\sqrt{\frac{\log n}{n h^m \phi^m(h)}} \right)^{-m} (\alpha_n^{-1} \tau_n^{-1})^{-\gamma-1} \\
 &= h^{-m} \left(\sqrt{\frac{\log n}{n h^m \phi^m(h)}} \right)^{-m+\gamma+1} ((\log n)^{\zeta_0} n^{1/\zeta})^{\gamma+1} \\
 &= \frac{(\log n)^{-\frac{m+\gamma+1}{2} + \zeta_0(\gamma+1)}}{n^{\frac{-m+\gamma+1}{2} - 1 - \frac{\gamma+1}{\zeta}} h^{\frac{m+\gamma+1}{2}} \phi(h)^{\frac{-m+\gamma+1}{2}}}.
 \end{aligned}$$

By our Assumptions 4 part ii), it holds that $R_{2n} = o(1)$. This shows the result. Now let's move on to the nonlinear part of the Hoeffding decomposition. Accordingly, the goal is to prove that

$$\mathbb{P} \left[\sup_{\mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \widehat{\psi}_{2,i}(\mathbf{u}, \mathbf{x}, \varphi) \right| > \lambda \right] \rightarrow 0 \text{ as } n \rightarrow \infty.$$

In the following, we will give a lemma which can be viewed as a technical result in the proof of our proposition.

Lemma 6.2. *Let $\mathcal{F}_m \mathcal{K}^m$ be a uniformly bounded class of measurable canonical functions, $m \geq 2$. Suppose that there are finite constants \mathbf{a} and \mathbf{b} such that the $\mathcal{F}_m \mathcal{K}^m$ covering number satisfies :*

$$N(\varepsilon, \mathcal{F}_m \mathcal{K}^m, \|\cdot\|_{L_2(Q)}) \leq \mathbf{a} \varepsilon^{-\mathbf{b}}, \quad (6.13)$$

for every $\varepsilon > 0$ and every probability measure Q . If the mixing coefficients β of the local stationary sequence $\{Z_i = (X_{i,n}, W_{i,n})\}_{i \in \mathbb{N}^*}$ achieve:

$$\beta(k) k^r \rightarrow 0, \text{ as } k \rightarrow \infty, \quad (6.14)$$

for some $r > 1$, then :

$$\sup_{\mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \mathbb{P} \left[h^{m/2} \phi^{m/2}(h) n^{-m+1/2} \sum_{\mathbf{i} \in I_n^m} \xi_{i_1} \cdots \xi_{i_m} H(Z_{i_1}, \dots, Z_{i_m}) \right] \rightarrow 0. \quad (6.15)$$

Remark 6.1. *As mentioned before, $W_{i,n}$ will be equal to 1 or $\varepsilon_{i,n} = \sigma\left(\frac{i}{n}, X_{i,n}\right) \varepsilon_i$. In the proof of the previous Lemma, $W_{i,n}$ will be equal $\varepsilon_{i,n} = \sigma\left(\frac{i}{n}, X_{i,n}\right) \varepsilon_i$, and we will use the notation $W_{i,n}^{(u)}$ to indicate $\sigma(\mathbf{u}, \mathbf{x}) \varepsilon_i$*

Proof of Lemma 6.2 The proof of this lemma is based on the blocking method and, in particular, on the techniques employed by Arcones and Yu, 1994: their fundamental concept is partitioning the strictly stationary sequence (Z_1, \dots, Z_n) into $2n$ blocks, each of which is of length a_n and a remainder block of length $n - 2v_n a_n$. This concept is called *Bernstein's method*, referred to Bernstein, 1927, in which we are enabled to apply the symmetrization and the many other techniques available for the i.i.d random variables. To establish this independence between the blocks, the little ones should be placed between two consecutive big blocks, and it should be asymptotically negligible. In this proof, for clarity of exposition, we present the case of $m = 2$, the sizes a_n and b_n are different where b_n denotes the size of the alternative blocks. Both a_n and b_n satisfy

$$b_n \ll a_n, \quad (v_n - 1)(a_n + b_n) < n \leq v_n(a_n + b_n), \quad (6.16)$$

and set, for $1 \leq j \leq v_n - 1$:

$$H_j^{(U)} = \{i : (j-1)(a_n + b_n) + 1 \leq i \leq (j-1)(a_n + b_n) + a_n\},$$

$$\begin{aligned} \mathbb{T}_j^{(U)} &= \{i : (j-1)(a_n + b_n) + a_n + 1 \leq i \leq (j-1)(a_n + b_n) + a_n + b_n\}, \\ \mathbb{H}_{v_n}^{(U)} &= \{i : (v_n - 1)(a_n + b_n) + 1 \leq i \leq n \wedge (v_n - 1)(a_n + b_n) + a_n\}, \\ \mathbb{T}_{v_n}^{(U)} &= \{i : (v_n - 1)(a_n + b_n) + a_n + 1 \leq i \leq n\}. \end{aligned}$$

Then introduce the sequence of independent blocks (η_1, \dots, η_n) such as :

$$\mathcal{L}(\eta_1, \dots, \eta_n) = \mathcal{L}(Z_1, \dots, Z_{a_n}) \times \mathcal{L}(Z_{a_n+1}, \dots, Z_{2a_n}) \times \dots$$

An application of the result of Eberlein, 1984 implies that for any measurable set A:

$$\begin{aligned} & \left| \mathbb{P}\left\{ \left(\eta_1, \dots, \eta_{a_n}, \eta_{2a_n+1}, \dots, \eta_{3a_n}, \dots, \eta_{2(v_n-1)a_n+1}, \dots, \eta_{2v_n a_n} \right) \in A \right\} \right. \\ & \quad \left. - \mathbb{P}\left\{ \left(Z_1, \dots, Z_{a_n}, Z_{2a_n+1}, \dots, Z_{3a_n}, \dots, Z_{2(v_n-1)a_n+1}, \dots, Z_{2v_n a_n} \right) \in A \right\} \right| \\ & \leq 2(v_n - 1)\beta(a_n). \end{aligned} \tag{6.17}$$

As we are dealing with a locally stationary sequence (X_1, \dots, X_n) , then the sequence of the independent blocks that we use in the following is given by $\{\eta_i\}_{i \in \mathbb{N}^*}$. We decompose the process according to this distribution of the blocks:

$$\begin{aligned} & \sum_{i_1 \neq i_2}^n \frac{1}{h^2 \phi^2(h)} \prod_{k=1}^2 \left\{ K_1 \left(\frac{u_k - i_k/n}{h} \right) K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right\} W_{i_1, i_2, \varphi, n} \\ &= \sum_{p \neq q}^{v_n} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \frac{W_{i_1, i_2, \varphi, n}}{h^2 \phi^2(h)} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h} \right) K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right\} \\ &+ \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in \mathbb{H}_p^{(U)}} \frac{W_{i_1, i_2, \varphi, n}}{h^2 \phi^2(h)} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h} \right) K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right\} \\ &+ 2 \sum_{p=1}^{v_n} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{q: |q-p| \geq 2} \sum_{i_2 \in \mathbb{T}_q^{(U)}} \frac{W_{i_1, i_2, \varphi, n}}{h^2 \phi^2(h)} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h} \right) K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right\} \\ &+ 2 \sum_{p=1}^{v_n} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{q: |q-p| \leq 1} \sum_{i_2 \in \mathbb{T}_q^{(U)}} \frac{W_{i_1, i_2, \varphi, n}}{h^2 \phi^2(h)} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h} \right) K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right\} \\ &+ \sum_{p \neq q}^{v_n} \sum_{i_1 \in \mathbb{T}_p^{(U)}} \sum_{i_2 \in \mathbb{T}_q^{(U)}} \frac{W_{i_1, i_2, \varphi, n}}{h^2 \phi^2(h)} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h} \right) K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right\} \\ &+ \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2} \sum_{i_1, i_2 \in \mathbb{T}_p^{(U)}} \frac{W_{i_1, i_2, \varphi, n}}{h^2 \phi^2(h)} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h} \right) K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right\} \\ & := \text{I} + \text{II} + \text{III} + \text{IV} + \text{V} + \text{VI}. \end{aligned} \tag{6.18}$$

(I): The same type of block but not the same block: Assume that the sequence of independent blocks $\{\eta_i\}_{i \in \mathbb{N}^*}$ is of size a_n . An application of (6.17) shows that:

$$\begin{aligned}
 & \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi^{-1}(h) \sum_{p \neq q}^{v_n} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} \right. \right. \\
 & \quad \left. \left. \prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) W_{i_1, i_2, \varphi, n} \right| > \delta \right) \\
 & \leq \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi^{-1}(h) \sum_{p \neq q}^{v_n} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} \right. \right. \\
 & \quad \left. \left[\prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) - \prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i/n)})}{h} \right) \right] W_{i_1, i_2, \varphi, n} \right| > \delta \right) \\
 & + \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi^{-1}(h) \sum_{p \neq q}^{v_n} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} \right. \right. \\
 & \quad \left. \prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i/n)})}{h} \right) [W_{i_1, i_2, \varphi, n} - W_{i_1, i_2, \varphi, n}^{(u)}] \right| > \delta \right) \\
 & + \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi^{-1}(h) \sum_{p \neq q}^{v_n} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} \right. \right. \\
 & \quad \left. \prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i/n)})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right| > \delta \right) \\
 & \leq o_{\mathbb{P}}(1) + o_{\mathbb{P}}(1) + \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi^{-1}(h) \sum_{p \neq q}^{v_n} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \right. \\
 & \quad \left. \left. K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, \varphi, n}^{(u)} \right| > \delta \right) + 2v_n \beta(b_n)
 \end{aligned}$$

Because:

$$\begin{aligned}
 & \mathbb{E} \left| n^{-3/2} h \phi^{-1}(h) \sum_{p \neq q}^{v_n} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} \right. \\
 & \quad \left. \left[\prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) - \prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i/n)})}{h} \right) \right] W_{i_1, i_2, \varphi, n} \right| \\
 & = n^{-3/2} h \phi^{-1}(h) \sum_{p \neq q}^{v_n} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} \\
 & \quad \left| \mathbb{E} \left[\prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) - \prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i/n)})}{h} \right) \right] W_{i_1, i_2, \varphi, n} \right|
 \end{aligned}$$

$$\begin{aligned}
 &= n^{-3/2} h \phi^{-1}(h) \sum_{p \neq q} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} \\
 &\quad \mathbb{E} \left| \left[\prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) - \prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i, n}^{(i/n)})}{h} \right) \right] \sigma \left(\frac{\mathbf{i}}{n}, X_{i, n} \right) \varepsilon_{\mathbf{i}} \right| \\
 &= n^{-3/2} h \phi^{-1}(h) \sum_{p \neq q} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} \mathbb{E}(\varepsilon_{\mathbf{i}}) \mathbb{E} \left| \left[\prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right. \right. \\
 &\quad \left. \left. - \prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i, n}^{(i/n)})}{h} \right) \right] \left[\sigma \left(\frac{\mathbf{i}}{n}, X_{i, n} \right) - \sigma(\mathbf{u}, X_{i, n}) + \sigma(\mathbf{u}, X_{i, n}) \right] \right| \\
 &\lesssim n^{-3/2} h \phi^{-1}(h) \sum_{p \neq q} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} \mathbb{E}(\varepsilon_{\mathbf{i}}) (\sigma(\mathbf{u}, \mathbf{x}) + o_{\mathbb{P}}(1)) \\
 &\quad \mathbb{E} \left| \left[\prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) - \prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i, n}^{(i/n)})}{h} \right) \right] \right| \\
 &\lesssim n^{-3/2} h \phi^{-1}(h) \sum_{p \neq q} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} \mathbb{E}(\varepsilon_{\mathbf{i}}) (\sigma(\mathbf{u}, \mathbf{x}) + o_{\mathbb{P}}(1)) \left[\frac{m \phi^{m-1}(h)}{nh} \right] \\
 &\quad \text{(where } m=2 \text{ and using Lemma 6.1 Equation (6.1))} \\
 &\sim o_{\mathbb{P}}(1),
 \end{aligned}$$

and

$$\begin{aligned}
 &\mathbb{E} \left| n^{-3/2} h \phi^{-1}(h) \sum_{p \neq q} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} \right. \\
 &\quad \left. \prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i/n)})}{h} \right) [W_{i_1, i_2, \varphi, n} - W_{i_1, i_2, \varphi, n}^{(u)}] \right| \\
 &= n^{-3/2} h \phi^{-1}(h) \sum_{p \neq q} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} \\
 &\quad \mathbb{E} \left| \prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i/n)})}{h} \right) \left[\sigma \left(\frac{\mathbf{i}}{n}, X_{i, n} \right) \varepsilon_{\mathbf{i}} - \sigma(\mathbf{u}, \mathbf{x}) \varepsilon_{\mathbf{i}} \right] \right| \\
 &= n^{-3/2} h \phi^{-1}(h) \sum_{p \neq q} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} \mathbb{E}(\varepsilon_{\mathbf{i}}) \\
 &\quad \mathbb{E} \left| \prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i/n)})}{h} \right) \left[\sigma \left(\frac{\mathbf{i}}{n}, X_{i, n} \right) - \sigma(\mathbf{u}, \mathbf{x}) \right] \right| \\
 &\lesssim n^{-3/2} h \phi^{-1}(h) \sum_{p \neq q} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \xi_{i_1} \xi_{i_2} \mathbb{E}(\varepsilon_{\mathbf{i}}) (o_{\mathbb{P}}(1)) \int_0^h \prod_{k=1}^m K_2 \left(\frac{y_k}{h} \right) dF_{i_k/n}(y_k, x_k)
 \end{aligned}$$

$$\lesssim n^{-3/2}h\phi^{-1}(h) \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \xi_{i_1} \xi_{i_2} \mathbb{E}(\varepsilon_i)(o_{\mathbb{P}}(1))(\phi^2(h)) \sim o_{\mathbb{P}}(1). \quad (6.19)$$

We keep the choice of b_n and v_n such that

$$v_n b_n^r \leq 1, \quad (6.20)$$

which implies that $2v_n \alpha_{b_n} \rightarrow 0$ as $n \rightarrow \infty$, so the term to consider is the second summand. For the second part of the inequality, we will use the work of Arcones et al., 1993 in the non-fixed kernels settings, precisely, we will define $f_{i_1, \dots, i_m} = \prod_{k=1}^m \xi_{i_k} \times H$ and $\mathcal{F}_{i_1, \dots, i_m}$ respectively as a collections of kernels and the class of functions related to this kernel, then we will use de la Peña et al., 1999, Theorem 3.1.1 and Remarks 3.5.4 part 2 for decoupling and randomization. As we mentioned above, we will suppose that $m = 2$. Then we can see that

$$\begin{aligned} & \mathbb{E} \left\| n^{-3/2}h\phi^{-1}(h) \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \xi_{i_1} \xi_{i_2} \prod_{k=1}^2 K_2 \left(\frac{d(x_k, \eta_{i_k})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2, \mathcal{K}^2} \\ &= \mathbb{E} \left\| n^{-3/2}h\phi(h) \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} f_{i_1, i_2}(\mathbf{u}, \boldsymbol{\eta}) \right\|_{\mathcal{F}_{i_1, i_2}} \\ &\leq c_2 \mathbb{E} \left\| n^{-3/2}h\phi(h) \sum_{p \neq q} \varepsilon_p \varepsilon_q \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} f_{i_1, i_2}(\mathbf{u}, \boldsymbol{\eta}) \right\|_{\mathcal{F}_{i_1, i_2}} \\ &\leq c_2 \mathbb{E} \int_0^{D_{nh}^{(U_1)}} N(t, \mathcal{F}_{i_1, i_2}, \tilde{d}_{nh, 2}^{(1)}) dt, \quad (\text{By Lemma 6.7 and Proposition 6.2.}) \quad (6.21) \end{aligned}$$

where $D_{nh}^{(U_1)}$ is the diameter of \mathcal{F}_{i_1, i_2} according to the distance $\tilde{d}_{nh, 2}^{(1)}$, respectively defined as

$$\begin{aligned} D_{nh}^{(U_1)} &:= \left\| \mathbb{E}_{\varepsilon} \left\| n^{-3/2}h\phi(h) \sum_{p \neq q} \varepsilon_p \varepsilon_q \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} f_{i_1, i_2}(\mathbf{u}, \boldsymbol{\eta}) \right\|_{\mathcal{F}_{i_1, i_2}} \right\| \\ &= \left\| \mathbb{E}_{\varepsilon} \left\| n^{-3/2}h\phi^{-1}(h) \sum_{p \neq q} \varepsilon_p \varepsilon_q \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \xi_{i_1} \xi_{i_2} \prod_{k=1}^2 K_2 \left(\frac{d(x_k, \eta_{i_k})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2, \mathcal{K}^2} \right\|, \end{aligned}$$

and :

$$\tilde{d}_{nh, 2}^{(1)} \left(\xi_1. K_{2,1} W'^{(u)}, \xi_2. K_{2,2} W''^{(u)} \right)$$

$$\begin{aligned}
 & := \mathbb{E}_\varepsilon \left| n^{-3/2} h \phi^{-1}(h_n) \sum_{p \neq q}^{\nu_n} \varepsilon_p \varepsilon_q \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \left[\xi_{1i_1} \xi_{1i_2} \prod_{k=1}^2 K_{1,2} \left(\frac{d(x_k, \eta_{i_k})}{h} \right) W'_{i_1, i_2, \varphi, n}(u) \right. \right. \\
 & \quad \left. \left. - \xi_{2i_1} \xi_{2i_2} \prod_{k=1}^2 K_{2,2} \left(\frac{d(x_k, \eta_{i_k})}{h} \right) W''_{i_1, i_2, \varphi, n}(u) \right] \right|.
 \end{aligned}$$

Let consider another semi-norm $\tilde{d}_{nh,2}^{(2)}$:

$$\begin{aligned}
 & \tilde{d}_{nh,2}^{(2)} \left(\xi_1.K_{2,1}W'(u) , \xi_2.K_{2,2}W''(u) \right) \\
 & = \frac{1}{nh^2\phi^2(h)} \left[\sum_{i \neq j}^{\nu_n} \left(\xi_{1i_1} \xi_{1i_2} \prod_{k=1}^2 K_{1,2} \left(\frac{d(x_k, \eta_{i_k})}{h} \right) W'_{i_1, i_2, \varphi, n}(u) \right. \right. \\
 & \quad \left. \left. - \xi_{2i_1} \xi_{2i_2} \prod_{k=1}^2 K_{2,2} \left(\frac{d(x_k, \eta_{i_k})}{h} \right) W''_{i_1, i_2, \varphi, n}(u) \right)^2 \right]^{1/2}.
 \end{aligned}$$

One can see that

$$\tilde{d}_{nh,2}^{(1)} \left(\xi_1.K_{2,1}W'(u) , \xi_2.K_{2,2}W''(u) \right) \leq a_n n^{-1/2} h \phi(h) \tilde{d}_{nh,2}^{(2)} \left(\xi_1.K_{2,1}W'(u) , \xi_2.K_{2,2}W''(u) \right).$$

We readily infer that

$$\begin{aligned}
 & \mathbb{E} \left\| n^{-3/2} h \phi^{-1}(h) \sum_{p \neq q}^{\nu_n} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \xi_{i_1} \xi_{i_2} \prod_{k=1}^2 K_2 \left(\frac{d(x_k, \eta_{i_k})}{h} \right) W_{i_1, i_2, \varphi, n}(u) \right\|_{\mathcal{F}_2 \mathcal{H}^2} \\
 & \leq c_2 \mathbb{E} \int_0^{D_{nh}^{(U_1)}} N \left(t a_n^{-1} n^{1/2}, \mathcal{F}_{i,j}, \tilde{d}_{nh,2}^{(2)} \right) dt \\
 & \leq c_2 a_n n^{-1/2} \mathbb{P} \left\{ D_{nh}^{(U_1)} a_n^{-1} n^{1/2} \geq \lambda_n \right\} + c_m a_n n^{-1/2} \int_0^{\lambda_n} \log t^{-1} dt, \quad (6.22)
 \end{aligned}$$

where $\lambda_n \rightarrow 0$. We have

$$\frac{\left(\int_0^{\lambda_n} \log t^{-1} dt \right)}{\left(\lambda_n \log \lambda_n^{-1} \right)} \rightarrow 0,$$

where a_n and λ_n must be chosen in such a way that the following relation will be achieved

$$a_n \lambda_n n^{-1/2} \log \lambda_n^{-1} \rightarrow 0. \quad (6.23)$$

Making use of the triangle inequality, in combination with Hoeffding's trick, we obtain readily that

$$a_n n^{-1/2} \mathbb{P} \left\{ D_{nh}^{(U_1)} \geq \lambda_n a_n n^{-1/2} \right\}$$

$$\begin{aligned}
 &\leq \lambda_n^{-2} a_n^{-1} n^{-5/2} h \phi^{-1}(h) \mathbb{E} \left\| \sum_{p \neq q}^{v_n} \left[\sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \right. \\
 &\quad \left. \left. K_2 \left(\frac{d(x_2, \eta'_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right]^2 \right\|_{\mathcal{F}_2 \mathcal{K}^2} \\
 &\leq c_2 v_n \lambda_n^{-2} a_n^{-1} n^{-5/2} h \phi^{-1}(h) \mathbb{E} \left\| \sum_{p=1}^{v_n} \left[\sum_{i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \right. \\
 &\quad \left. \left. K_2 \left(\frac{d(x_2, \eta'_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right]^2 \right\|_{\mathcal{F}_2 \mathcal{K}^2}, \tag{6.24}
 \end{aligned}$$

where $\{\eta'_i\}_{i \in \mathbb{N}^*}$ are independent copies of $(\eta_i)_{i \in \mathbb{N}^*}$. By imposing :

$$\lambda_n^{-2} a_n^{1-r} n^{-1/2} \rightarrow 0, \tag{6.25}$$

we readily infer that

$$\begin{aligned}
 &\left\| v_n \lambda_n^{-2} a_n^{-1} n^{-5/2} h \phi^{-1}(h) \mathbb{E} \sum_{p=1}^{v_n} \left[\sum_{i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} \prod_{k=1}^2 K_2 \left(\frac{d(x_k, \eta_{i_k})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right]^2 \right\|_{\mathcal{F}_2 \mathcal{K}^2} \\
 &\leq O \left(\lambda_n^{-2} a_n^{1-r} n^{-1/2} \right).
 \end{aligned}$$

A symmetrization of the last inequality in (6.24) succeeded by an application of the Proposition 6.2 in the Appendix, gives

$$\begin{aligned}
 &v_n \lambda_n^{-2} a_n^{-1} n^{-5/2} h \phi^{-1}(h) \mathbb{E} \left\| \sum_{p=1}^{v_n} \left[\sum_{i_1, i_2 \in H_p^{(U)}} \varepsilon_p \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \right. \\
 &\quad \left. \left. K_2 \left(\frac{d(x_2, \eta'_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right]^2 \right\|_{\mathcal{F}_2 \mathcal{K}^2} \\
 &\leq c_2 \mathbb{E} \left(\int_0^{D_{nh}^{(U_2)}} (\log N(u, \mathcal{F}_{i,j}, \tilde{d}_{nh,2}^l)^{1/2}) \right), \tag{6.26}
 \end{aligned}$$

where

$$\begin{aligned}
 D_{nh}^{(U_2)} &= \left\| \mathbb{E}_\varepsilon \left| v_n \lambda_n^{-2} a_n^{-1} n^{-5/2} \phi^{-1}(h) \right. \right. \\
 &\quad \left. \left. \sum_{p=1}^{v_n} \varepsilon_p \left[\sum_{i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, \eta'_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right]^2 \right\|_{\mathcal{F}_2 \mathcal{K}^2}.
 \end{aligned}$$

and for $\xi_1.K_{2,1}W'$, $\xi_2.K_{2,2}W'' \in \mathcal{F}_{ij}$:

$$\begin{aligned} & \tilde{d}_{nh,2} \left(\xi_1.K_{2,1}W'^{(u)} , \xi_2.K_{2,2}W''^{(u)} \right) \\ := & \mathbb{E}_\varepsilon \left| v_n \lambda_n^{-2} a_n^{-1} n^{-5/2} \phi^{-1}(h_n) \sum_{p=1}^{v_n} \varepsilon_p \left[\left(\sum_{i_1, i_2 \in H_p^{(U)}} \xi_{1i_1} \xi_{1i_2} K_{2,1} \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_{2,1} \left(\frac{d(x_2, \eta'_{i_2})}{h} \right) \right. \right. \right. \\ & \left. \left. \left. W'_{i_1, i_2, \varphi, n}^{(u)} \right)^2 - \left(\sum_{i_1, i_2 \in H_p^{(U)}} \xi_{2i_1} \xi_{2i_2} K_{2,2} \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_{2,2} \left(\frac{d(x_2, \eta'_{i_2})}{h} \right) W''_{i_1, i_2, \varphi, n}^{(u)} \right)^2 \right] \right|. \end{aligned}$$

By the fact that :

$$\begin{aligned} & \mathbb{E}_\varepsilon \left| v_n \lambda_n^{-2} a_n^{-1} n^{-5/2} \phi^{-1}(h_n) \sum_{p=1}^{v_n} \varepsilon_p \left(\sum_{i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, \eta'_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right)^2 \right| \\ & \leq a_n^{3/2} \lambda_n^{-2} n^{-1} \left[v_n^{-1} a_n^{-2} \phi^{-2}(h_n) \sum_{p=1}^{v_n} \sum_{i_1, i_2 \in H_p^{(U)}} \left(\xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, \eta'_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right)^4 \right]^{1/2}, \end{aligned}$$

so:

$$a_n^{3/2} \lambda_n^{-2} n^{-1} \rightarrow 0, \quad (6.27)$$

we have the convergence of (6.26) to zero. For the choice of a_n , b_n and v_n , it should be noted that all the values satisfying (6.16), (6.20), (6.23), (6.25) and (6.26) are suitable.

(II): The same block:

$$\begin{aligned} & \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi(h) \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \frac{1}{\phi^2(h)} \xi_{i_1} \xi_{i_2} \right. \right. \\ & \quad \left. \left. K_2 \left(\frac{d(x_1, X_{i_1, n})}{h} \right) K_2 \left(\frac{d(x_2, X_{i_2, n})}{h} \right) W_{i_1, i_2, \varphi, n} \right| > \delta \right) \\ & \leq \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left\| n^{-3/2} h \phi(h) \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \frac{1}{\phi^2(h)} \xi_{i_1} \xi_{i_2} \right. \right. \\ & \quad \left. \left[\prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) - \prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i/n)})}{h} \right) \right] W_{i_1, i_2, \varphi, n} \right| > \delta \right) \\ & + \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi(h) \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} \right. \right. \\ & \quad \left. \left. \prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i/n)})}{h} \right) \left[W_{i_1, i_2, \varphi, n} - W_{i_1, i_2, \varphi, n}^{(u)} \right] \right| > \delta \right) \end{aligned}$$

$$\begin{aligned}
 & + \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi(h) \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} \right. \right. \\
 & \quad \left. \left. \prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i,n}^{(i/n)})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right| > \delta \right) \\
 & \leq \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi^{-1}(h) \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_i, \eta_{i_1})}{h} \right) \right. \right. \\
 & \quad \left. \left. K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right| > \delta \right) + 2v_n \alpha_{b_n} \tag{6.28}
 \end{aligned}$$

In the same manner of I, we can show that the first and the second term in the previous inequality is of order $o_{\mathbb{P}}(1)$. So, as the preceding proof, it suffices to prove that

$$\begin{aligned}
 \mathbb{E} \left(\left\| n^{-3/2} h \phi^{-1}(h) \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \right. \\
 \left. \left. K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{K}^2} \right) \rightarrow 0.
 \end{aligned}$$

Notice that we treat a uniformly bounded classes functions in, we obtain uniformly in $B^m \times \mathcal{F}_2 \mathcal{K}^2$

$$\mathbb{E} \left(\sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n} \right) = O(a_n).$$

This implies that we have to prove that, for $\mathbf{u} \in B^m$

$$\begin{aligned}
 \mathbb{E} \left(\left\| n^{-3/2} h \phi^{-1}(h) \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \left[\xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_i, \eta_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right. \right. \right. \\
 \left. \left. \left. - \mathbb{E} \left(\xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n} \right) \right] \right\|_{\mathcal{F}_2 \mathcal{K}^2} \right) \rightarrow 0. \tag{6.29}
 \end{aligned}$$

As for empirical processes, to prove (6.29), it suffices to symmetrize and show that

$$\begin{aligned}
 \mathbb{E} \left(\left\| n^{-3/2} h \phi^{-1}(h) \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \varepsilon_p \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \right. \\
 \left. \left. K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{K}^2} \right) \rightarrow 0.
 \end{aligned}$$

In a similar way as in (6.21), we infer that :

$$\mathbb{E} \left(\left\| n^{-3/2} h \phi^{-1}(h) \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \varepsilon_p \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2, \mathcal{X}^2} \right) \leq \mathbb{E} \left(\int_0^{D_{nh}^{(U_3)}} (\log N(u, \mathcal{F}_{i_1, i_2}, \tilde{d}_{nh, 2}^{(3)}))^{1/2} du \right),$$

where

$$D_{nh}^{(U_3)} = \left\| \mathbb{E}_\varepsilon \left| n^{-3/2} h \phi^{-1}(h) \sum_{p=1}^{v_n} \varepsilon_p \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2, \mathcal{X}^2} \right\|, \quad (6.30)$$

and the semi-metric $\tilde{d}_{nh, 2}^{(3)}$ is defined by

$$\begin{aligned} & \tilde{d}_{nh, 2}^{(3)} \left(\xi_1 K_{2,1} W^{(u)}, \xi_2 K_{2,2} W''^{(u)} \right) \\ &= \mathbb{E}_\varepsilon \left| n^{-3/2} h \phi^{-1}(h) \sum_{p=1}^{v_n} \varepsilon_p \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \left(\xi_{1i_1} \xi_{1j} K_{2,1} \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_{2,1} \left(\frac{d(x_2, \eta_{i_2})}{h} \right) \right. \right. \\ & \quad \left. \left. W_{i_1, i_2, \varphi, n}^{(u)} - \xi_{2i_1} \xi_{2j} K_{2,2} \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_{2,2} \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}''^{(u)} \right) \right|. \end{aligned}$$

Since we are trading uniformly bounded classes of functions, we infer that

$$\begin{aligned} & \mathbb{E}_\varepsilon \left| n^{-3/2} h \phi^{-1}(h_n) \sum_{p=1}^{v_n} \varepsilon_p \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right| \\ & \leq a_n^{3/2} (n)^{-1} h \phi^{-1}(h_n) \left[\frac{1}{v_n a_n^2} \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2; i_1, i_2 \in H_p^{(U)}} \left(\xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \right. \\ & \quad \left. \left. K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right)^2 \right]^{1/2} \leq O \left(a_n^{3/2} (n)^{-1} \phi^{-1}(h_n) \right). \end{aligned}$$

Since $a_n^{3/2} (n)^{-1} \phi^{-1}(h) \rightarrow 0$, $D_{nh}^{(U_3)} \rightarrow 0$, we obtain $\text{II} \rightarrow 0$ as $n \rightarrow \infty$.

(III): Different types of blocks:

$$\begin{aligned}
 & \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \frac{1}{\phi^2(h)} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, X_{i_1})}{h} \right) \right. \right. \\
 & \quad \left. \left. K_2 \left(\frac{d(x_2, X_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n} \right| > \delta \right) \\
 & \leq \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi^{-1}(h) \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \frac{1}{\phi^2(h)} \xi_{i_1} \xi_{i_2} \right. \right. \\
 & \quad \left. \left[\prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) - \prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i/n)})}{h} \right) \right] W_{i_1, i_2, \varphi, n} \right| > \delta \right) \\
 & + \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi^{-1}(h) \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} \right. \right. \\
 & \quad \left. \prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i/n)})}{h} \right) \left[W_{i_1, i_2, \varphi, n} - W_{i_1, i_2, \varphi, n}^{(u)} \right] \right| > \delta \right) \\
 & + \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} h \phi^{-1}(h) \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} \right. \right. \\
 & \quad \left. \prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i/n)})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right| > \delta \right)
 \end{aligned} \tag{6.31}$$

As we claimed before, the first and the second summands of the previous inequality have been treated. We have left with the last summation where an application of (6.17) shows that

$$\begin{aligned}
 & \sum_{p=1}^{v_n} \mathbb{E} \left\| n^{-3/2} h \phi^{-1}(h) \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} \prod_{k=1}^2 K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i/n)})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2} \\
 & \leq \sum_{p=1}^{v_n} \mathbb{E} \left\| n^{-3/2} h \phi^{-1}(h) \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} \right. \\
 & \quad \left. K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2} + n^{-3/2} h \phi^{-1}(h_n) v_n^2 a_n b_n \beta(a_n),
 \end{aligned}$$

we have

$$n^{-3/2} \phi^{-1}(h) v_n^2 a_n b_n \beta(a_n) \rightarrow 0,$$

using Condition (6.14) and the choice of a_n , b_n and v_n . For $p = 1$ and $p = v_n$:

$$\begin{aligned} & \mathbb{E} \left\| n^{-3/2} h \phi^{-1}(h) \sum_{i_1 \in H_1^{(U)}} \sum_{q:|q-p| \geq 2}^{v_n} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, X_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, X_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2} \\ &= \mathbb{E} \left\| n^{-3/2} h \phi^{-1}(h) \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{v_n} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, X_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, X_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2}. \end{aligned}$$

For $2 \leq p \leq v_n - 1$, we obtain :

$$\begin{aligned} & \mathbb{E} \left\| n^{-3/2} h \phi^{-1}(h) \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2}^{v_n} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, X_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, X_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2} \\ &= \mathbb{E} \left\| n^{-3/2} h \phi^{-1}(h) \sum_{i_1 \in H_1^{(U)}} \sum_{q=4}^{v_n} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, X_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, X_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2} \\ &\leq \mathbb{E} \left\| n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{v_n} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, X_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, X_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2}, \end{aligned}$$

therefore it suffices to treat the convergence:

$$\begin{aligned} & \mathbb{E} \left\| v_n n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{v_n} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \\ & \quad \left. K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2} \longrightarrow 0. \quad (6.32) \end{aligned}$$

By similar arguments as in Arcones and Yu, 1994, the usual symmetrization is applied and:

$$\begin{aligned} & \mathbb{E} \left\| v_n n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{v_n} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \\ & \quad \left. K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2} \\ &\leq 2 \mathbb{E} \left\| v_n n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{v_n} \sum_{i_2 \in T_q^{(U)}} \varepsilon_q \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \\ & \quad \left. K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2} \end{aligned}$$

$$\begin{aligned}
 &= 2\mathbb{E} \left\{ \left\| v_n n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{v_n} \sum_{i_2 \in T_q^{(U)}} \varepsilon_q \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \right. \\
 &\quad \left. \left. K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2, \mathcal{H}^2} \mathbb{1}_{\{D_{nh}^{(U_4)} \leq \gamma_n\}} \right\} \\
 &+ 2\mathbb{E} \left\{ \left\| v_n n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{v_n} \sum_{i_2 \in T_q^{(U)}} \varepsilon_q \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \right. \\
 &\quad \left. \left. K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2, \mathcal{H}^2} \mathbb{1}_{\{D_{nh}^{(U_4)} > \gamma_n\}} \right\} \\
 &= 2\text{III}_1 + 2\text{III}_2, \tag{6.33}
 \end{aligned}$$

where

$$D_{nh}^{(U_4)} = \left\| v_n n^{-3/2} h \phi^{-1}(h_n) \left[\sum_{q=3}^{v_n} \left(\sum_{i_2 \in T_q^{(U)}} \sum_{i_1 \in H_1^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \right. \right. \\
 \left. \left. \left. K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right) \right]^2 \right\|_{\mathcal{F}_2, \mathcal{H}^2}^{1/2}. \tag{6.34}$$

In a similar way as in (6.21), we infer that

$$\text{III}_1 \leq c_2 \int_0^{\gamma_n} (\log N(t, \mathcal{F}_{i_1, i_2}, \tilde{d}_{nh, 2}^{(4)}))^{1/2} dt, \tag{6.35}$$

where

$$\begin{aligned}
 &\tilde{d}_{nh, 2}^{(4)} \left(\xi_{1 \cdot} K_{2,1} W^{(u)}, \xi_{2 \cdot} K_{2,2} W''^{(u)} \right) \\
 &:= \mathbb{E}_\varepsilon \left| v_n n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{v_n} \sum_{i_2 \in T_q^{(U)}} \varepsilon_q \left[\xi_{1 i_1} \xi_{1 i_2} K_{2,1} \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \right. \\
 &\quad \left. \left. K_{2,1} \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} - \xi_{2 i_1} \xi_{2 i_2} K_{2,2} \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_{2,2} \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}''^{(u)} \right] \right|.
 \end{aligned}$$

Since we have

$$\mathbb{E}_\varepsilon \left| v_n n^{-3/2} h \phi^{-1}(h) \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{v_n} \sum_{i_2 \in T_q^{(U)}} \varepsilon_q \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \\
 \left. K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right|$$

$$\leq a_n^{-1/2} b_n h^2 \phi(h) \left(\frac{1}{a_n b_n v_n h^2 \phi^4(h_n)} \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{v_n} \sum_{i_2 \in T_q^{(U)}} \left[\xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right]^2 \right)^{1/2},$$

and considering the semi-metric

$$\begin{aligned} & \tilde{d}_{nh,2}^{(5)} \left(\xi_1 K_{2,1} W'^{(u)}, \xi_2 K_{2,2} W''^{(u)} \right) \\ := & \left(\frac{1}{a_n b_n v_n h^2 \phi^4(h)} \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{v_n} \sum_{i_2 \in T_q^{(U)}} \left[\xi_{1i_1} \xi_{1i_2} K_{2,1} \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_{2,1} \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}'^{(u)} - \xi_{2i_1} \xi_{2i_2} K_{2,2} \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_{2,2} \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}''^{(u)} \right]^2 \right)^{1/2}. \end{aligned}$$

We show that the expression in (6.35) is bounded as follows

$$v_n^{1/2} b_n n^{-1/2} h^2 \phi(h) \int_0^{v_n^{-1/2} b_n^{-1} n^{1/2} h^2 \gamma_n} (\log N(t, \mathcal{F}_{i_1, i_2}, \tilde{d}_{nh,2}^{(5)}))^{1/2} dt,$$

by choosing $\gamma_n = n^{-\alpha}$ for some $\alpha > (17r - 26)/60r$, we get the convergence to zero of the previous quantity. To bound the second term on the right hand side of (6.33), we can notice that

$$\begin{aligned} \text{III}_2 &= \mathbb{E} \left\{ \left\| v_n n^{-3/2} h \phi^{-1}(h) \sum_{i_1 \in H_1^{(U)}} \sum_{q=3}^{v_n} \sum_{i_2 \in T_q^{(U)}} \varepsilon_q \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{K}^2} \mathbb{1}_{\{D_{nh}^{(U_4)} > \gamma_n\}} \right\} \\ &\leq a_n^{-1} b_n n^{1/2} h \phi^{-1}(h) \mathbb{P} \left\{ \left\| v_n^2 n^{-3} h^2 \phi^{-2}(h_n) \sum_{q=3}^{v_n} \left(\sum_{i_2 \in T_q^{(U)}} \sum_{i_1 \in H_1^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right)^2 \right\|_{\mathcal{F}_2 \mathcal{K}^2} \geq \gamma_n^2 \right\}. \quad (6.36) \end{aligned}$$

We now apply the square root trick to the last expression conditionally on H_1^U . We denote by \mathbb{E}_T the expectation with respect to $\sigma\{i_2 : i_2 \in T_q, q \geq 3\}$ and we will suppose that any

class of functions \mathcal{F}_m is unbounded and its envelope function satisfies for some $p > 2$:

$$\theta_p := \sup_{\mathbf{t} \in \mathcal{H}_m} \mathbb{E}(F^p(\mathbf{Y}) | \mathbf{X} = \mathbf{t}) < \infty, \quad (6.37)$$

for $2r/(r-1) < s < \infty$, (in the notation in of Giné et al., 1984, Lemma 5.2).

$$M_n = v_n^{1/2} \mathbb{E}_T \left(\sum_{j \in T_q^{(U)}} \sum_{i \in H_1^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, X_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, X_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right)^2$$

where

$$t = \gamma_n^2 a_n^{5/2} n^{1/2} h \phi^{-1}(h_n), \quad \rho = \lambda = 2^{-4} \gamma_n a_n^{5/4} n^{1/4} h^{1/2} \phi^{-1/2}(h_n),$$

and

$$m = \exp \left(\gamma_n^2 n h^2 \phi^{-2}(h_n) b_n^{-2} \right).$$

However, since we need $t > 8M_n$, and $m \rightarrow \infty$, by similar arguments as in Arcones and Yu, 1994, page 69, we get the convergence of (6.35) and (6.36) to zero.

(IV): Different types of blocks: The target here is to prove that:

$$\mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q: |q-p| \leq 1} \sum_{i_2 \in H_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, X_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, X_{i_2})}{h} \right) W_{i_1, i_2, \varphi, n} \right| > \delta \right) \rightarrow 0.$$

We have

$$\begin{aligned} & \left\| n^{-3/2} h \phi^{-1}(h) \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q: |q-p| \leq 1} \sum_{i_2 \in T_q^{(U)}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, X_{i_1}^{(i_1/n)})}{h} \right) K_2 \left(\frac{d(x_2, X_{i_2}^{(i_2/n)})}{h} \right) W_{i_1, i_2, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{K}^2} \\ & \leq c_2 v_n a_n b_n n^{-3/2} h \phi^{-1}(h) \rightarrow 0. \end{aligned}$$

Hence the proof of the lemma is complete. \square

The final step in the proof of Proposition 3.1 lies in the use of Lemma 6.2 to prove that the nonlinear term converges to zero. \square

Proof of Theorem 3.1

Equation (4.1) in Section 4 shows that

$$\begin{aligned} & \tilde{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}; h_n) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \\ &= \frac{1}{\tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x})} \left(\hat{g}_1(\mathbf{u}, \mathbf{x}) + \hat{g}_2(\mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{x}, \mathbf{u}) \tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}) \right), \end{aligned}$$

where

$$\begin{aligned} \tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}) &= \frac{(n-m)!}{n! h^m \phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h} \right) K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right\}, \\ \hat{g}_1(\mathbf{u}, \mathbf{x}) &= \frac{(n-m)!}{n! h^m \phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h} \right) K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right\} W_{\mathbf{i}, \varphi, n}, \\ \hat{g}_2(\mathbf{u}, \mathbf{x}) &= \frac{(n-m)!}{n! h^m \phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h} \right) K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right\} r^{(m)} \left(\frac{\mathbf{i}}{n}, X_{i, n} \right). \end{aligned}$$

The proof of this theorem is involved and divided into the following 4 steps, where in each one, we aim to show that

Step 1.

$$\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |\hat{g}_1(\mathbf{u}, \mathbf{x})| = O_{\mathbb{P}} \left(\sqrt{\log n / n h^m \phi(h)} \right).$$

Step 2.

$$\begin{aligned} & \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |\hat{g}_2(\mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}; h_n)| \\ & - \mathbb{E}(\hat{g}_2(\mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}; h_n)) = O_{\mathbb{P}} \left(\sqrt{\log n / n h^m \phi^m(h)} \right). \end{aligned}$$

Step 3.

$$\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \mathbb{E}(\hat{g}_2(\mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}; h_n)) \right| = O(h^2) + O(h^\alpha).$$

Step 4

$$\frac{1}{\inf_{\mathcal{F}_m \mathcal{H}^m} \inf_{\mathbf{x} \in \mathcal{H}^m} \inf_{\mathbf{u} \in [C_1 h, 1 - C_1 h]^m} |\tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x})|} = O_{\mathbb{P}}(1).$$

Step (1) follows directly from Proposition 3.1. The second one holds by replacing $\varphi(Y_{i_1}, \dots, Y_{i_m})$ with $\hat{g}_2(\mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}; h_n)$ and applying Proposition 3.1. We will pass now to the proof of Step (4). Let

$$\tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}) = \hat{r}_1(\varphi, \mathbf{u}, \mathbf{x}) + \bar{r}_1(\varphi, \mathbf{u}, \mathbf{x}), \quad (6.38)$$

where

$$\begin{aligned}\widehat{r}_1(\varphi, \mathbf{u}, \mathbf{x}) &= \frac{(n-m)!}{n!h^m\phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1\left(\frac{u_k - i_k/n}{h}\right) K_2\left(\frac{d(x_k, X_{i_k,n}^{(i/n)})}{h}\right) \right\} \\ \bar{r}_1(\varphi, \mathbf{u}, \mathbf{x}) &= \frac{(n-m)!}{n!h^m\phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1\left(\frac{u_k - i_k/n}{h}\right) \\ &\quad \prod_{k=1}^m \left[K_2\left(\frac{d(x_k, X_{i_k,n})}{h}\right) - K_2\left(\frac{d(x_k, X_{i_k,n}^{(i/n)})}{h}\right) \right].\end{aligned}$$

For $W \equiv 1$, the previous proposition proved that

$$\sup_{\mathcal{F}_m} \sup_{\mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |\tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}) - \mathbb{E}(\tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}))| = o_{\mathbb{P}}(1).$$

So we can see that

$$\begin{aligned}\tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}) &= \tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}) + \mathbb{E}(\tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x})) - \mathbb{E}(\tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x})) \\ &= o_{\mathbb{P}}(1) + \mathbb{E}[\widehat{r}_1(\varphi, \mathbf{u}, \mathbf{x})] + \mathbb{E}[\bar{r}_1(\varphi, \mathbf{u}, \mathbf{x})].\end{aligned}\tag{6.39}$$

Furthermore, we have

$$\begin{aligned}\mathbb{E}(\bar{r}_1(\varphi, \mathbf{u}, \mathbf{x})) &= \mathbb{E}\left(\frac{(n-m)!}{n!h^m\phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1\left(\frac{u_k - i_k/n}{h}\right) \right. \\ &\quad \left. \prod_{k=1}^m \left[K_2\left(\frac{d(x_k, X_{i_k,n})}{h}\right) - K_2\left(\frac{d(x_k, X_{i_k,n}^{(i/n)})}{h}\right) \right] \right) \\ &\lesssim \frac{(n-m)!}{n!h^m\phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1\left(\frac{u_k - i_k/n}{h}\right) \left(\frac{m\phi^{m-1}(h)}{nh}\right) = o(1).\end{aligned}\tag{6.40}$$

$$\tag{6.41}$$

The final result follows as $K_2(\cdot)$ is Lipschitz (Assumption 2, i)) and by applying Assumption 1i), by Lemma 6.4), uniformly in \mathbf{u} . We also have

$$\begin{aligned}\mathbb{E}[\widehat{r}_1(\varphi, \mathbf{u}, \mathbf{x})] &= \frac{(n-m)!}{n!h^m\phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1\left(\frac{u_k - i_k/n}{h}\right) \mathbb{E}\left[\prod_{k=1}^m K_2\left(\frac{d(x_k, X_{i_k,n}^{(i/n)})}{h}\right)\right] \\ &= \frac{(n-m)!}{n!h^m\phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1\left(\frac{u_k - i_k/n}{h}\right) \int_0^h \prod_{k=1}^m K_2\left(\frac{y_k}{h}\right) dF_{i_k/n}(y_k, x_k) \\ &\gtrsim \frac{(n-m)!}{n!h^m\phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1\left(\frac{u_k - i_k/n}{h}\right) \phi^m(h) f_1(\mathbf{x}) \sim f_1(\mathbf{x}) > 0,\end{aligned}$$

uniformly in \mathbf{u} . Then, we obtain

$$\begin{aligned} & \frac{1}{\inf_{\mathcal{F}_m \mathcal{H}^m \mathbf{x} \in \mathcal{H}^m \mathbf{u} \in [C_1 h, 1 - C_1 h]^m} |\tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x})|} \\ &= \frac{1}{\inf_{\mathcal{F}_m \mathcal{H}^m \mathbf{x} \in \mathcal{H}^m \mathbf{u} \in [C_1 h, 1 - C_1 h]^m} o(1) + o_{\mathbb{P}}(1) + \mathbb{E} [\hat{r}_1(\varphi, \mathbf{u}, \mathbf{x})]} = O_{\mathbb{P}}(1). \end{aligned} \quad (6.42)$$

We should define $K_0 : [0, 1] \rightarrow \mathbb{R}$ a Lipschitz continuous function with support $[0, q]$ for some $q > 1$ and such that $K_0(x) = 1, \forall x \in [0, 1]$. Notice that

$$\mathbb{E} \left[\hat{g}_2(\mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \tilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}; h_n) \right] = \sum_{i=1}^4 Q_i(\mathbf{u}, \mathbf{x}), \quad (6.43)$$

where Q_i can be defined as follows

$$Q_i(\mathbf{u}, \mathbf{x}) = \frac{(n-m)!}{n! h^m \phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \left\{ \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h} \right) \right\} q_i(\mathbf{u}, \mathbf{x}), \quad (6.44)$$

such that

$$\begin{aligned} q_1(\mathbf{u}, \mathbf{x}) &= \mathbb{E} \left[\prod_{k=1}^m K_0 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \left\{ \prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right. \right. \\ &\quad \left. \left. - \prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \right\} \times \left\{ r^{(m)}(\varphi, \frac{\mathbf{i}}{n}, X_{\mathbf{i}, n}) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \right\} \right], \\ q_2(\mathbf{u}, \mathbf{x}) &= \mathbb{E} \left[\prod_{k=1}^m \left\{ K_0 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \right\} \right. \\ &\quad \left. \left\{ r^{(m)}(\varphi, \frac{i_k}{n}, X_{i_k, n}) - r^{(m)}(\varphi, \frac{i_k}{n}, X_{i_k, n}^{(i_k/n)}) \right\} \right], \\ q_3(\mathbf{u}, \mathbf{x}) &= \mathbb{E} \left[\left\{ \prod_{k=1}^m K_0 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) - \prod_{k=1}^m K_0 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \right\} \right. \\ &\quad \left. \prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \times \left\{ r^{(m)}(\varphi, \frac{i_k}{n}, X_{i_k, n}^{(i_k/n)}) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \right\} \right], \\ q_4(\mathbf{u}, \mathbf{x}) &= \mathbb{E} \left[\prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \left\{ r^{(m)}(\varphi, \frac{i_k}{n}, X_{i_k, n}^{(i_k/n)}) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \right\} \right]. \end{aligned}$$

Observe that

$$\begin{aligned}
 Q_1(\mathbf{u}, \mathbf{x}) &\lesssim \frac{(n-m)!}{n!h^m\phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \left\{ \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h} \right) \mathbb{E} \left[\prod_{k=1}^m K_0 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right. \right. \\
 &\quad \left. \left| \prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) - \prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \right| \right. \\
 &\quad \left. \left. \times \left| r^{(m)} \left(\varphi, \frac{\mathbf{i}}{n}, \mathbf{X}_{i, n} \right) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \right| \right] \right\},
 \end{aligned}$$

using Assumption 3 (i), these properties of $r(\mathbf{u}, \mathbf{x})$ allow us to show that

$$\begin{aligned}
 &\prod_{k=1}^m K_0 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \left| r^{(m)} \left(\varphi, \frac{\mathbf{i}}{n}, \mathbf{X}_{i, n} \right) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \right| \\
 &\lesssim \prod_{k=1}^m K_0 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \left(d_{\mathcal{H}^m}(\mathbf{X}_{i, n}, \mathbf{x}) + \|\mathbf{u} - \frac{\mathbf{i}}{n}\| \right)^\alpha \lesssim h^{m\wedge\alpha}
 \end{aligned}$$

and under Assumption 2, part ii), we will use Lemma 6.1, Equation (6.1) to show that:

$$\begin{aligned}
 Q_1(\mathbf{u}, \mathbf{x}) &\lesssim \frac{(n-m)!}{n!h^m\phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \left\{ \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h} \right) \right\} \times h^{m\wedge\alpha} \times \frac{m\phi^{m-1}(h)}{nh}. \\
 &\lesssim \frac{1}{n\phi(h)h^{m-(m\wedge\alpha)}} \quad \text{uniformly in } \mathbf{u}. \tag{6.45}
 \end{aligned}$$

In a similar way, we can see that

$$\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in [C_1 h, 1 - C_1 h]^m} Q_2(\mathbf{u}, \mathbf{x}) \lesssim \frac{1}{n\phi(h)h^{m-(m\wedge\alpha)}}, \tag{6.46}$$

and

$$\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in [C_1 h, 1 - C_1 h]^m} Q_3(\mathbf{u}, \mathbf{x}) \lesssim \frac{1}{n\phi(h)h^{m-(1\wedge\alpha)}}. \tag{6.47}$$

For the last term, we have

$$\begin{aligned}
 Q_4(\mathbf{u}, \mathbf{x}) &= \frac{(n-m)!}{n!h^m\phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \left\{ \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h} \right) \right\} \\
 &\quad \mathbb{E} \left[\prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \left\{ r^{(m)} \left(\varphi, \frac{\mathbf{i}}{n}, X_{i, n}^{(i/n)} \right) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \right\} \right].
 \end{aligned}$$

Using Lemma 6.3 and inequality (2.12) and Assumption 1, it becomes apparent that

$$\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in [C_1 h, 1 - C_1 h]^m} |Q_4(\mathbf{u}, \mathbf{x})|$$

$$\begin{aligned}
 &\leq \frac{(n-m)!}{n!h^m\phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \left| \left\{ \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h} \right) \right\} \right| \\
 &\quad \mathbb{E} \left[\left| \prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \right| \left| \left\{ r^{(m)} \left(\varphi, \frac{\mathbf{i}}{n}, X_{\mathbf{i}, n}^{(\mathbf{i}/n)} \right) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \right\} \right| \right] \\
 &\lesssim \frac{(n-m)!}{n!h^m\phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \left| \left\{ \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h} \right) \right\} \right| \\
 &\quad \mathbb{E} \left[\left| \prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \right| \left(d_{\mathcal{H}^m} (X_{\mathbf{i}, n}^{(\mathbf{i}/n)}, \mathbf{x}) + \|\mathbf{u} - \frac{\mathbf{i}}{n}\| \right)^\alpha \right] \\
 &\lesssim \frac{(n-m)!}{n!h^m\phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \left| \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h} \right) \right. \\
 &\quad \left. - \int_0^1 \cdots \int_0^1 \frac{1}{h^m} \prod_{k=1}^m K_1 \left(\frac{(u_k - v_k)}{h} \right) dv_k \right| \mathbb{E} \left[\left| \prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \right| h^\alpha \right] \\
 &\quad + \frac{(n-m)!}{n!h^m\phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \int_0^1 \cdots \int_0^1 \frac{1}{h^m} \prod_{k=1}^m K_1 \left(\frac{u_k - v_k}{h} \right) dv_k \\
 &\quad \times \mathbb{E} \left[\left| \prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n}^{(i_k/n)})}{h} \right) \right| h^\alpha \right] \\
 &\lesssim O \left(\frac{1}{n^m h^{2m}} \right) h^\alpha + h^\alpha.
 \end{aligned} \tag{6.48}$$

But

$$(n)^{-m} h^{\alpha-2m} \lesssim h^{2m} \phi^m(h) \ll h^{2m},$$

we can see that

$$\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in [C_1 h, 1 - C_1 h]^m} |Q_4(\mathbf{u}, \mathbf{x})| \ll h^{2m} + h^\alpha.$$

Indeed, under our assumptions, the approximation error can be viewed as

$$O \left(\frac{1}{n^m \phi(h) h^{m-(1 \wedge \alpha)}} \right) \ll h^{2m \wedge \alpha}. \tag{6.49}$$

This inequality completes the proof. \square

Proof of Theorem 4.1

The subject here is to prove the weak convergence, so the finite-dimensional convergence and the asymptotic equicontinuity for the stochastic U -process

$$\sqrt{nh^m \phi(h)} D_n(f) = \sqrt{nh^m \phi(h)} \left(\tilde{r}_n^{(m)}(\varphi, \mathbf{i}, \mathbf{u}; h_n) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}) - B_T(u, x) \right)$$

must be proved over all the functions classes mentioned in the framework. By de la Peña et al., 1999, Section 4.2, the finite-dimensional convergence simply asserts that every finite set of functions f_1, \dots, f_q in L_2

$$\left(\sqrt{nh^m \phi(h)} D_n(f_1), \dots, \sqrt{nh^m \phi(h)} D_n(f_q) \right) \quad (6.50)$$

converges to the corresponding finite-dimensional distributions of the process G_p . Cramér-Wold and the countability of the different classes allow us to reduce the situation from weak convergence of U -process to weak convergence of U -statistics with kernel f_r , for all $r \in \{1, \dots, q\}$, as well as the U -process is a linear operator. Then, we only need to show that $\sqrt{nh^m \phi(h)} D_n(f_r)$ converge to a Gaussian distribution. So, for a fixed kernel, we have

$$\begin{aligned} & \tilde{r}_n^{(m)}(\varphi, \mathbf{i}, \mathbf{u}; h_n) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}) \\ &= \frac{1}{\tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u})} \left(\hat{g}_1(\mathbf{u}, \mathbf{x}) + \hat{g}_2(\mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}) \tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u}; h_n) \right) \\ &= \frac{1}{\tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u})} \left(\hat{g}_1(\mathbf{u}, \mathbf{x}) + \hat{G}(\mathbf{u}, \mathbf{x}) \right), \end{aligned} \quad (6.51)$$

where

$$\hat{G}(\mathbf{u}, \mathbf{x}) = \hat{g}_2(\mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}) \tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u}; h_n).$$

We begin treating each term. For this sake, we will calculate the variance of $\hat{G}(\mathbf{u}, \mathbf{x})$. Take

$$\Delta_{i,n}(\mathbf{u}, \mathbf{x}) = \prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \left[r^{(m)}\left(\varphi, \frac{\mathbf{i}}{n}, \mathbf{u}, X_{i,n}\right) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}) \right].$$

Observe that

$$\begin{aligned} \text{Var}(\hat{G}(\mathbf{u}, \mathbf{x})) &= \text{Var} \left(\hat{g}_2(\mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}) \tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u}; h_n) \right) \\ &= \text{Var} \left(\frac{(n-m)!}{n! h^m \phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \left\{ \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h} \right) \right\} \Delta_{i,n}(\mathbf{u}, \mathbf{x}) \right) \\ &= \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2 \left(\frac{u_k - i_k/n}{h} \right) \text{Var}(\Delta_{i,n}(\mathbf{u}, \mathbf{x})) \\ &\quad + \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\substack{i_1 \neq \dots \neq i_m, \\ i'_1 \neq \dots \neq i'_m, \\ \exists j/i_j \neq i'_j}} \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h} \right) \prod_{k=1}^m K_1^2 \left(\frac{u_k - i'_k/n}{h} \right) \\ &\quad \text{Cov}(\Delta_{i,n}(\mathbf{u}, \mathbf{x}), \Delta_{i',n}(\mathbf{u}, \mathbf{x})) \\ &:= V_1 + V_2. \end{aligned} \quad (6.52)$$

Looking on V_1 ,

$$\begin{aligned}
 |V_1| &= \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2 \left(\frac{u_k - i_k/n}{h} \right) \text{Var}(\Delta_{\mathbf{i},n}(\mathbf{u}, \mathbf{x})) \\
 &= \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2 \left(\frac{u_k - i_k/n}{h} \right) \left[\mathbb{E}(\Delta_{\mathbf{i},n}^2(\mathbf{u}, \mathbf{x})) - (\mathbb{E}(\Delta_{\mathbf{i},n}(\mathbf{u}, \mathbf{x})))^2 \right] \\
 &\leq \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2 \left(\frac{u_k - i_k/n}{h} \right) \mathbb{E}(\Delta_{\mathbf{i},n}^2(\mathbf{u}, \mathbf{x})) \\
 &\leq \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2 \left(\frac{u_k - i_k/n}{h} \right) \mathbb{E} \left(\prod_{k=1}^m K_2^2 \left(\frac{d(x_k, X_{i_k,n})}{h} \right) \right. \\
 &\quad \left. \left[r^{(m)}\left(\varphi, \frac{\mathbf{i}}{n}, \mathbf{u}, X_{\mathbf{i},n}\right) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}) \right]^2 \right) \\
 &\leq \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2 \left(\frac{u_k - i_k/n}{h} \right) \mathbb{E} \left(\prod_{k=1}^m K_2^2 \left(\frac{d(x_k, X_{i_k,n})}{h} \right) \right. \\
 &\quad \left. - \prod_{k=1}^m K_2^2 \left(\frac{d(x_k, X_{i_k,n}^{(i_k/n)})}{h} \right) + \prod_{k=1}^m K_2^2 \left(\frac{d(x_k, X_{i_k,n}^{(i_k/n)})}{h} \right) \right. \\
 &\quad \left. \times \left[r^{(m)}\left(\varphi, \frac{\mathbf{i}}{n}, \mathbf{u}, X_{\mathbf{i},n}\right) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}) \right]^2 \right) \\
 &\leq \frac{((n-m)!)^2 h^{2\alpha}}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2 \left(\frac{u_k - i_k/n}{h} \right) \mathbb{E} \left[\prod_{k=1}^m K_2^2 \left(\frac{d(x_k, X_{i_k,n})}{h} \right) \right. \\
 &\quad \left. - \prod_{k=1}^m K_2^2 \left(\frac{d(x_k, X_{i_k,n}^{(i_k/n)})}{h} \right) \right] + \mathbb{E} \left(\prod_{k=1}^m K_2^2 \left(\frac{d(x_k, X_{i_k,n}^{(i_k/n)})}{h} \right) \right). \tag{6.53}
 \end{aligned}$$

The last part of (6.53) follows from the smoothness assumption on $r^{(m)}$ in 3 i)

$$\left| r^{(m)}\left(\varphi, \frac{\mathbf{i}}{n}, \mathbf{u}, X_{\mathbf{i},n}\right) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}) \right|^2 \lesssim h^{2\alpha}. \tag{6.54}$$

Combining the latter inequalities with Equation (6.3) and Equation (6.1) from Lemma 6.1 where:

$$\begin{aligned}
 &\mathbb{E} \left(\prod_{k=1}^m K_2^2 \left(\frac{d(x_k, X_{i_k,n})}{h} \right) - \prod_{k=1}^m K_2^2 \left(\frac{d(x_k, X_{i_k,n}^{(i_k/n)})}{h} \right) \right) \\
 &\lesssim \mathbb{E} \left(\prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k,n})}{h} \right) - \prod_{k=1}^m K_2 \left(\frac{d(x_k, X_{i_k,n}^{(i_k/n)})}{h} \right) \right),
 \end{aligned}$$

to get

$$\begin{aligned}
 |V_1| &\lesssim \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\mathbf{i} \in J_n^m} \prod_{k=1}^m K_1^2 \left(\frac{u_k - i_k/n}{h} \right) \\
 &\lesssim \frac{h^{2\alpha}(h) ((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \left[\frac{m\phi^{m-1}(h)}{nh} + \phi^m(h) \right] \ll \frac{1}{nh^m \phi(h)}. \quad (6.55)
 \end{aligned}$$

A deep sight into the work of Arcones, 1998, specially Lemma 2, makes us see V_2 as follows:

$$\begin{aligned}
 |V_2| &= \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\substack{1 \leq i_1 \leq \dots \leq i_{2m} \leq n \\ j_1 \geq j_2, \dots, j_m}} \prod_{k=1}^{2m} K_1 \left(\frac{u_k - i_k/n}{h} \right) \\
 &\quad \times \text{Cov} \left(\Delta_{i_{\sigma(1)}, \dots, i_{\sigma(2m)}, n}(\mathbf{u}, \mathbf{x}), \Delta_{i_{\sigma(m+1)}, \dots, \sigma(2m), n}(\mathbf{u}, \mathbf{x}) \right) \\
 &\leq \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\substack{1 \leq i_1 \leq \dots \leq i_{2m} \leq n \\ j_1 \geq j_2, \dots, j_m}} \prod_{k=1}^{2m} K_1 \left(\frac{u_k - i_k/n}{h} \right) \\
 &\quad \left| \mathbb{E} \left[\Delta_{i_{\sigma(1)}, \dots, i_{\sigma(2m)}, n}(\mathbf{u}, \mathbf{x}) \Delta_{i_{\sigma(m+1)}, \dots, \sigma(2m), n}(\mathbf{u}, \mathbf{x}) \right] \right| \\
 &\leq \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\substack{1 \leq i_1 \leq \dots \leq i_{2m} \leq n \\ j_1 \geq j_2, \dots, j_m}} \prod_{k=1}^{2m} K_1 \left(\frac{u_k - i_k/n}{h} \right) \times cM^2 \left(1 + \sum_{k=1}^{n-1} k^{m-1} \beta_k^{(p-2)/p} \right) \\
 &\lesssim \frac{((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\substack{1 \leq i_1 \leq \dots \leq i_{2m} \leq n \\ j_1 \geq j_2, \dots, j_m}} \prod_{k=1}^{2m} K_1 \left(\frac{u_k - i_k/n}{h} \right) \times M^2 \ll \frac{1}{nh^m \phi(h)}, \quad (6.56)
 \end{aligned}$$

where

$$M := \sup_{1 \leq i_1 < \dots < i_m < \infty} \mathbb{E} [|\Delta_{i,n}|^p]^{1/p},$$

and $j_1 = i_2 - i_1$, $j_l = \min(i_{2l-1} - i_{2l-2}, i_{2l} - i_{2l-1})$ for $2 \leq l \leq m-1$, and $j_m = i_{2m} - i_{2m-1}$. If $j_1 = \max(j_1, \dots, j_m)$, we compare the initial sequence $\{X_1, \dots, X_n\}$ with the one having the independent blocks $\{i_1\}$, $\{i_2, \dots, i_{2m}\}$ and the same block distribution. Then it is easy now to show that

$$\text{Var}(\widehat{G}(\mathbf{u}, \mathbf{x})) \leq |V_1| + |V_2| = o\left(\frac{1}{nh^m \phi(h)}\right). \quad (6.57)$$

This result implies the quadratic-mean convergence of $\widehat{G}(\mathbf{u}, \mathbf{x})$ with the rate specified as follows

$$\widehat{G}(\mathbf{u}, \mathbf{x}) - \mathbb{E}(\widehat{G}(\mathbf{u}, \mathbf{x})) = o\left(\frac{1}{nh\phi^m(h)}\right) \text{ in probability.} \quad (6.58)$$

Let's remember that

$$\tilde{r}_n^{(m)}(\varphi, \mathbf{i}, \mathbf{u}; h_n) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}) = \frac{1}{\tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u})} \left(\hat{g}_1(\mathbf{u}, \mathbf{x}) + \hat{G}(\mathbf{u}, \mathbf{x}) \right),$$

$$B_T(\mathbf{u}, \mathbf{x}) = \mathbb{E}[\hat{g}^B(u, x)] / \mathbb{E}[\tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u})],$$

$$\tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u}) = \mathbb{E}[\tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u})] + o_{\mathbb{P}}(1),$$

and

$$\lim_{n \rightarrow \infty} \mathbb{E}[\tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u})] > 0,$$

then

$$\tilde{r}_n^{(m)}(\varphi, \mathbf{i}, \mathbf{u}; h_n) - r^{(m)}(\varphi, \mathbf{i}, \mathbf{u}) = \frac{\hat{g}_1(\mathbf{u}, \mathbf{x})}{\tilde{r}_1(\varphi, \mathbf{i}, \mathbf{u})} + B_n(\mathbf{u}, \mathbf{x}) + o_{\mathbb{P}}\left(\frac{1}{nh^m\phi(h)}\right).$$

In the next, we will consider the first part of the last equation.

$$\begin{aligned} & \sqrt{nh^m\phi(h)} \text{Var}(\hat{g}_1(x)) \\ &= \frac{\sqrt{nh^m\phi(h)}((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2\left(\frac{u_k - i_k/n}{h}\right) \text{Var} \left[\prod_{k=1}^m K_2\left(\frac{d(x_k, X_{i_k, n})}{h}\right) \varepsilon_{\mathbf{i}, n} \right] \\ &= \frac{\sqrt{nh^m\phi(h)}((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2\left(\frac{u_k - i_k/n}{h}\right) \mathbb{E} \left[\prod_{k=1}^m K_2^2\left(\frac{d(x_k, X_{i_k, n})}{h}\right) \varepsilon_{\mathbf{i}, n}^2 \right] \\ &= \frac{\sqrt{nh^m\phi(h)}((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2\left(\frac{u_k - i_k/n}{h}\right) \\ & \quad \mathbb{E} \left[\prod_{k=1}^m K_2^2\left(\frac{d(x_k, X_{i_k, n})}{h}\right) \sigma^2\left(\frac{\mathbf{i}}{n}, X_{\mathbf{i}, n}\right) \varepsilon_{\mathbf{i}}^2 \right] \\ & \quad (\{\varepsilon_{\mathbf{i}}\}_{\mathbf{i} \in \mathbb{Z}} \text{ is a sequence i.i.d r.v.'s, independent of } \{X_{i, n}\}_{i=1}^n) \\ &= \frac{\mathbb{E}(\varepsilon_{\mathbf{i}}^2) \sqrt{nh^m\phi(h)}((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2\left(\frac{u_k - i_k/n}{h}\right) \\ & \quad \mathbb{E} \left[\prod_{k=1}^m K_2^2\left(\frac{d(x_k, X_{i_k, n})}{h}\right) \right] \left[\sigma^2\left(\frac{\mathbf{i}}{n}, X_{\mathbf{i}, n}\right) \right] \\ &= \frac{\mathbb{E}(\varepsilon_{\mathbf{i}}^2) \sqrt{nh^m\phi(h)}((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2\left(\frac{u_k - i_k/n}{h}\right) \\ & \quad \mathbb{E} \left[\prod_{k=1}^m K_2^2\left(\frac{d(x_k, X_{i_k, n})}{h}\right) \right] \left[\sigma^2(\mathbf{u}, \mathbf{x}) + o(1) \right] \\ & \quad (\text{According to Assumption 3 [ii]- iii)- iv]) \end{aligned}$$

$$\begin{aligned}
 & + \frac{\mathbb{E}(\varepsilon_{\mathbf{i}}^2) o(\phi^m(h)) \sqrt{nh^m \phi(h)} ((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2 \left(\frac{u_k - i_k/n}{h} \right) [\sigma^2(\mathbf{u}, \mathbf{x}) + o(1)] \\
 = & \frac{\mathbb{E}(\varepsilon_{\mathbf{i}}^2) o(\phi^m(h)) \sqrt{nh^m \phi(h)} ((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2 \left(\frac{u_k - i_k/n}{h} \right) \\
 & \mathbb{E} \left[\prod_{k=1}^m K_2^2 \left(\frac{d(x_k, X_{i_k, n}^{i_k/n})}{h} \right) \right] [\sigma^2(\mathbf{u}, \mathbf{x}) + o(1)] + o(1) \\
 \sim & \frac{\mathbb{E}(\varepsilon_{\mathbf{i}}^2) o(\phi^m(h)) (\sigma^2(\mathbf{u}, \mathbf{x}) + o(1)) \sqrt{nh^m \phi(h)} ((n-m)!)^2}{(n!)^2 h^{2m} \phi^{2m}(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1^2 \left(\frac{u_k - i_k/n}{h} \right) \\
 \sim & \frac{1}{\sqrt{nh^m \phi(h)}} \mathbb{E}(\varepsilon_{\mathbf{i}}^2) \sigma^2(\mathbf{u}, \mathbf{x}) \int_{[0, h]^m} \prod_{k=1}^m K_1^2(\mathbf{z}) d\mathbb{P}(z_1, \dots, z_m).
 \end{aligned} \tag{6.59}$$

We can then see that the weak convergence is somewhat linked to the weak convergence of \hat{g}_1 . In order to show the last one, we will: 1) truncate the function \hat{g}_1 as we work with an unbounded class of functions, 2) verify that the remainder term of truncation converges to zero, 3) apply the Hoeffding's decomposition to the truncated part, 4) prove that the non-linear term of this decomposition converges to zero, 5) prove the weak convergence to the linear term, by proving the finite-dimensional convergence and the asymptotic equicontinuity. All these steps follow in the same way as in the proof of Theorem 4.2. Hence the proof is complete. \square

Proof of Theorem 4.2

We have

$$\tilde{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}; h_n) = \frac{\sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h} \right) K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right\} \varphi(Y_{\mathbf{i}, n})}{\sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h} \right) K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right\}}, \tag{6.60}$$

Define:

$$G_{\varphi, \mathbf{i}}(\mathbf{x}, \mathbf{y}) := \frac{\prod_{k=1}^m \left\{ K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right\} \varphi(Y_{\mathbf{i}, n})}{\mathbb{E} \prod_{k=1}^m \left\{ K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right\}} \quad \text{for } \mathbf{x} \in \mathcal{H}^m, \mathbf{y} \in \mathcal{Y}^m;$$

$$\begin{aligned}
 \mathcal{G} & := \{G_{\varphi, \mathbf{i}}(\cdot, \cdot) \mid \varphi \in \mathcal{F}_m, \mathbf{i} = (i_1, \dots, i_m)\}; \\
 \mathcal{G}^{(k)} & := \{\pi_{k, m} G_{\varphi, \mathbf{i}}(\cdot, \cdot), \varphi \in \mathcal{F}_m, \};
 \end{aligned}$$

$$\mathbb{U}_n(\varphi, \mathbf{i}) = \mathbb{U}_n^{(m)}(G_{\varphi, \mathbf{i}}) := \frac{(n-m)!}{n!} \sum_{i \in I_n^m} \prod_{k=1}^m \xi_{i_k} G_{\varphi, \mathbf{i}}(\mathbf{X}_i, \mathbf{Y}_i);$$

and the U -empirical process is defined to be

$$\mu_n(\varphi, \mathbf{i}) := \sqrt{nh^m \phi(h)} \{ \mathbb{U}_n(\varphi, \mathbf{i}) - \mathbb{E}(\mathbb{U}_n(\varphi, \mathbf{i})) \}.$$

Then ,

$$\tilde{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}; h_n) = \frac{\mathbb{U}_n(\varphi, \mathbf{i})}{\mathbb{U}_n(\mathbf{1}, \mathbf{i})}$$

In order to establish the weak convergence of our estimator, it must be established first for $\mu_n(\varphi, \mathbf{i})$. We have mentioned before that we deal with unbounded classes of functions, that is why we should truncate the function $G_{\varphi, \mathbf{i}}(\mathbf{x}, \mathbf{y})$, indeed, for $\lambda_n = n^{1/p}$, with $p > 0$, we have:

$$\begin{aligned} G_{\varphi, \mathbf{i}}(\mathbf{x}, \mathbf{y}) &= G_{\varphi, \mathbf{i}}(\mathbf{x}, \mathbf{y}) \mathbb{1}_{\{F(\mathbf{y}) \leq \lambda_n\}} + G_{\varphi, \mathbf{i}}(\mathbf{x}, \mathbf{y}) \mathbb{1}_{\{F(\mathbf{y}) > \lambda_n\}} \\ &:= G_{\varphi, \mathbf{i}}^{(T)}(\mathbf{x}, \mathbf{y}) + G_{\varphi, \mathbf{i}}^{(R)}(\mathbf{x}, \mathbf{y}). \end{aligned}$$

We can write the U -statistic as follows :

$$\begin{aligned} \mu_n(\varphi, \mathbf{i}) &= \sqrt{nh^m \phi(h)} \left\{ \mathbb{U}_n^{(m)}(G_{\varphi, \mathbf{i}}^{(T)}) - \mathbb{E}(\mathbb{U}_n^{(m)}(G_{\varphi, \mathbf{i}}^{(T)})) \right\} \\ &\quad + \sqrt{nh^m \phi(h)} \left\{ \mathbb{U}_n^{(m)}(G_{\varphi, \mathbf{i}}^{(R)}) - \mathbb{E}(\mathbb{U}_n^{(m)}(G_{\varphi, \mathbf{i}}^{(R)})) \right\} \\ &:= \sqrt{nh^m \phi(h)} \left\{ \mathbb{U}_n^{(T)}(\varphi, \mathbf{i}) - \mathbb{E}(\mathbb{U}_n^{(T)}(\varphi, \mathbf{i})) \right\} \\ &\quad + \sqrt{nh^m \phi(h)} \left\{ \mathbb{U}_n^{(R)}(\varphi, \mathbf{i}) - \mathbb{E}(\mathbb{U}_n^{(R)}(\varphi, \mathbf{i})) \right\} \\ &:= \mu_n^{(T)}(\varphi, \mathbf{i}) + \mu_n^{(R)}(\varphi, \mathbf{i}). \end{aligned} \tag{6.61}$$

The first term is the truncated part and the second is the remaining one. We have to prove that:

1. $\mu_n^{(T)}(\varphi, \mathbf{i})$ converges to a Gaussian process.
2. The remainder part is negligible, in the sense that

$$\left\| \sqrt{nh^m \phi(h)} \left\{ \mathbb{U}_n^{(R)}(\varphi, \mathbf{i}) - \mathbb{E}(\mathbb{U}_n^{(R)}(\varphi, \mathbf{i})) \right\} \right\|_{\mathcal{F}_m \mathcal{H}^m} \xrightarrow{\mathbb{P}} 0.$$

For the first point, we will use the decomposition of Hoeffding, which would be the same as the previous decomposition in Subsection 3.1 except that we replace $W_{\mathbf{i}, n}$ by $\varphi(Y_{\mathbf{i}, n})$

$$\mathbb{U}_n^{(T)}(\varphi, \mathbf{i}) - \mathbb{E}(\mathbb{U}_n^{(T)}(\varphi, \mathbf{i})) := \mathbb{U}_{1, n}(\varphi, \mathbf{i}) + \mathbb{U}_{2, n}(\varphi, \mathbf{i}),$$

where,

$$\mathbb{U}_{1,n}(\varphi, \mathbf{i}) := \frac{1}{n} \sum_{i=1}^n \widehat{H}_{1,i}(\mathbf{u}, \mathbf{x}, \varphi) \quad (6.62)$$

$$\mathbb{U}_{2,n}(\varphi, \mathbf{i}) := \frac{(n-m)!}{(n)!} \sum_{\mathbf{i} \in I_n^m} \xi_{i_1} \cdots \xi_{i_m} H_{2,\mathbf{i}}(\mathbf{z}). \quad (6.63)$$

The convergence of $\mathbb{U}_{2,n}(\varphi, \mathbf{i})$ to zero in probability follows from Lemma 6.2. Hence, it is enough to show that $\mathbb{U}_{1,n}(\varphi, \mathbf{i})$ converges weakly to a Gaussian process called $\mathbb{G}(\varphi)$. In order to achieve our goal, we will go through finite-dimensional convergence and equicontinuity.

The finite-dimensional convergence simply asserts that every finite set of functions f_1, \dots, f_q in L_2 , for $\widetilde{\mathbb{U}}$ the centred form of \mathbb{U} :

$$\left(\sqrt{nh^m \phi(h)} \widetilde{\mathbb{U}}_{1,n}(f_1, \mathbf{i}), \dots, \sqrt{nh^m \phi(h)} \widetilde{\mathbb{U}}_{1,n}(f_q, \mathbf{i}) \right) \quad (6.64)$$

converges to the corresponding finite-dimensional distributions of the process $\mathbb{G}(\varphi)$. It is sufficient to show that for every fixed collection $(a_1, \dots, a_k) \in \mathbb{R}$ we have;

$$\sum_{j=1}^q a_j \widetilde{\mathbb{U}}_{1,n}(f_j, \mathbf{i}) \rightarrow N(0, \sigma^2),$$

where

$$\sigma^2 = \sum_{j=1}^q a_j^2 \text{Var} \left(\widetilde{\mathbb{U}}_{1,n}(f_j, \mathbf{i}) \right) + \sum_{s \neq r} a_s a_r \text{Cov} \left(\widetilde{\mathbb{U}}_{1,n}(f_s, \mathbf{i}), \widetilde{\mathbb{U}}_{1,n}(f_r, \mathbf{i}) \right). \quad (6.65)$$

Take

$$h(\cdot) = \sum_{j=1}^q a_j f_j(\cdot).$$

By linearity of $h(\cdot)$, we have to see that $\widetilde{\mathbb{U}}_{1,n}(h, \mathbf{i}) \rightarrow \mathbb{G}(h)$. Let

$$N = \mathbb{E} \prod_{k=1}^m \left\{ K_2 \left(\frac{d(x_k, X_{i_k, n})}{h} \right) \right\}.$$

We have:

$$\begin{aligned} \widetilde{\mathbb{U}}_{1,n}(h_n, \mathbf{i}) &= N^{-1} \times \frac{1}{n} \sum_{i=1}^n \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_i \xi_{i_\ell} \cdots \xi_{i_{m-1}} \frac{1}{\phi(h)} K_2 \left(\frac{d(x_i, X_i)}{h} \right) \\ &\quad \times \int h(y_1, \dots, y_{\ell-1}, Y_i, y_\ell, \dots, y_{m-1}) \prod_{\substack{k=1 \\ k \neq i}}^{m-1} \frac{1}{\phi(h)} K_2 \left(\frac{d(x_k, \nu_k)}{h} \right) \\ &\quad \mathbb{P}(d(\nu_1, y_1), \dots, d(\nu_{\ell-1}, y_{\ell-1}), d(\nu_\ell, y_\ell), \dots, d(\nu_{m-1}, y_{m-1})), \end{aligned}$$

$$= N^{-1} \frac{1}{n} \sum_{i=1}^n \xi_i \frac{1}{\phi(h)} K_2 \left(\frac{d(x_i, X_i)}{h} \right) \tilde{h}(Y_i).$$

Now, we will use the blocking procedure for this empirical process. We will decompose the set $\{1, \dots, n\}$ into $2\nu_n + 1$ subsets contains small and big blocks. To keep the same notation as in Lemma 6.2, the size of large blocks is a_n and the small one is b_n , such that :

$$\nu_n := \left\lfloor \frac{n}{a_n + b_n} \right\rfloor, \quad \frac{b_n}{a_n} \rightarrow 0, \quad \frac{a_n}{n} \rightarrow 0, \quad \frac{n}{a_n} \beta(b_n) \rightarrow 0. \quad (6.66)$$

In this case we can see that :

$$\begin{aligned} \tilde{U}_{1,n}(h, \mathbf{j}) &= \sum_{j=1}^{\nu_n-1} \hat{U}_{j,n}^{(1)} + \sum_{j=1}^{\nu_n-1} \hat{U}_{j,n}^{(2)} + \hat{U}_{j,n}^{(3)} \\ &:= \tilde{U}_{1,n}^{(1)} + \tilde{U}_{1,n}^{(2)} + \tilde{U}_{1,n}^{(3)}, \end{aligned} \quad (6.67)$$

where

$$\hat{U}_{j,n}^{(1)} = N^{-1} \sum_{i=j(a_n+b_n)+1}^{j(a_n+b_n)+a_n} \xi_i \frac{1}{\phi(h)} K_2 \left(\frac{d(x_i, X_i)}{h} \right) \tilde{h}(Y_i), \quad (6.68)$$

$$\hat{U}_{j,n}^{(2)} = N^{-1} \sum_{i=j(a_n+b_n)+a_n+1}^{(j+1)(a_n+b_n)} \frac{1}{\xi_i \phi(h)} K_2 \left(\frac{d(x_i, X_i)}{h} \right) \tilde{h}(Y_i), \quad (6.69)$$

$$\hat{U}_{j,n}^{(3)} = N^{-1} \sum_{i=\nu(a_n+b_n)+1}^n \xi_i \frac{1}{\phi(h)} K_2 \left(\frac{d(x_i, X_i)}{h} \right) \tilde{h}(Y_i). \quad (6.70)$$

First, we aim to prove that $\frac{1}{n} \mathbb{E}(\tilde{U}_{1,n}^{(2)})^2 \rightarrow 0$ and $\frac{1}{n} \mathbb{E}(\tilde{U}_{1,n}^{(3)})^2 \rightarrow 0$ to show that the case of summation over the small blocks and the summation over the last one are asymptotically negligible. Hence,

$$\mathbb{E}(\tilde{U}_{1,n}^{(2)})^2 = \text{Var} \left(\sum_{j=1}^{\nu_n-1} \hat{U}_{j,n}^{(2)} \right) = \sum_{j=1}^{\nu_n-1} \text{Var} \left(\hat{U}_{j,n}^{(2)} \right) + \sum_{j=1}^{\nu_n-1} \sum_{\substack{k=1 \\ j \neq k}}^{\nu_n-1} \text{Cov} \left(\hat{U}_{j,n}^{(2)}, \hat{U}_{k,n}^{(2)} \right),$$

We have :

$$\begin{aligned} \text{Var} \left(\hat{U}_{j,n}^{(2)} \right) &= \text{Var} \left(N^{-1} \sum_{i=j(a_n+b_n)+a_n+1}^{(j+1)(a_n+b_n)} \xi_i \frac{1}{\phi(h)} K_2 \left(\frac{d(x_i, X_i)}{h} \right) \tilde{h}(Y_i) \right) \\ &= N^{-2} \frac{1}{\phi^2(h)} \frac{a_n}{a_n} \sum_{i=j(a_n+b_n)+a_n+1}^{(j+1)(a_n+b_n)} \xi_i^2 \text{Var} \left(K_2 \left(\frac{d(x_i, X_i)}{h} \right) \tilde{h}(Y_i) \right) \\ &\lesssim \frac{b_n}{\phi^2(h) (m\phi^{m-1}(h)/nh + \phi^m(h))^2} \times \left[\int_{[0,h]} K_1^2 \left(\frac{\omega}{h} \right) d\omega \right] \end{aligned}$$

$$\times \text{Var} \left(K_2 \left(\frac{d(x_i, X_i)}{h} \right) \tilde{h}(Y_i) \right). \quad (\text{Using Lemma 6.1ii})$$

Thus,

$$\begin{aligned} \sum_{j=1}^{\nu_n-1} \text{Var} \left(\hat{\mathbb{U}}_{j,n}^{(2)} \right) &\lesssim \nu_n b_n \frac{1}{\phi^{2(m+1)}(h)} \left[\int_{[0,h]} K_1^2 \left(\frac{\omega}{h} \right) d\omega \right] \times \text{Var} \left(K_2 \left(\frac{d(x_i, X_i)}{h} \right) \tilde{h}(Y_i) \right) \\ &\sim \frac{n}{a_n + b_n} b_n \sim \frac{n b_n}{a_n} = o_{\mathbb{P}}(n), \quad (\text{by (6.66)}) . \end{aligned} \quad (6.71)$$

and

$$\begin{aligned} &\sum_{j=1}^{\nu_n-1} \sum_{\substack{k=1 \\ j \neq k}}^{\nu_n-1} \text{Cov} \left(\hat{\mathbb{U}}_{j,n}^{(2)}, \hat{\mathbb{U}}_{k,n}^{(2)} \right) \\ &= \sum_{\substack{j=1 \\ j \neq k}}^{\nu_n-1} \sum_{k=1}^{\nu_n-1} \sum_{i=j(a_n+b_n)+a_n+1}^{(j+1)(a_n+b_n)} \sum_{i'=k(a_n+b_n)+a_n+1}^{(k+1)(a_n+b_n)} \frac{N-2}{\phi(h)} \xi_i \xi_{i'} \\ &\quad \text{Cov} \left(K_2 \left(\frac{d(x_i, X_i)}{h} \right) \tilde{h}(Y_i), K_2 \left(\frac{d(x_{i'}, X_{i'})}{h} \right) \tilde{h}(Y_{i'}) \right) \\ &= \sum_{\substack{j=1 \\ j \neq k}}^{\nu_n-1} \sum_{k=1}^{\nu_n-1} \sum_{l_1=1}^{b_n} \sum_{l_2=1}^{b_n} \frac{1}{N^2 \phi^2(h)} \xi_{\lambda_i+l_1} \xi_{\lambda_{i'}+l_2} \\ &\quad \text{Cov} \left(K_2 \left(\frac{d(x_{\lambda_i+l_1}, X_{\lambda_i+l_1})}{h} \right) \tilde{h}(Y_{\lambda_i+l_1}), K_2 \left(\frac{d(x_{\lambda_{i'}+l_2}, X_{\lambda_{i'}+l_2})}{h} \right) \tilde{h}(Y_{\lambda_{i'}+l_2}) \right), \end{aligned}$$

where $\lambda_i = j(a_n + b_n) + a_n$, but for $j \neq k$, $|\lambda_i - \lambda_{i'} + l_1 - l_2| \geq b_n$, then

$$\begin{aligned} &\left| \sum_{\substack{j=1 \\ j \neq k}}^{\nu_n-1} \sum_{k=1}^{\nu_n-1} \text{Cov} \left(\hat{\mathbb{U}}_{j,n}^{(2)}, \hat{\mathbb{U}}_{k,n}^{(2)} \right) \right| \\ &\leq \sum_{\substack{j=1 \\ |j-k| \geq b_n}}^{\nu_n-1} \sum_{k=1}^{\nu_n-1} \frac{1}{N^2 \phi^2(h)} \xi_j \xi_k \left| \text{Cov} \left(K_2 \left(\frac{d(x_j, X_j)}{h} \right) \tilde{h}(Y_j), K_2 \left(\frac{d(x_k, X_k)}{h} \right) \tilde{h}(Y_k) \right) \right|, \end{aligned}$$

here, the use of Davydov's lemma (Lemma 6.8) is necessary, we have

$$\begin{aligned} &\left| \text{Cov} \left(K_2 \left(\frac{d(x_j, X_j)}{h} \right) \tilde{h}(Y_j), K_2 \left(\frac{d(x_k, X_k)}{h} \right) \tilde{h}(Y_k) \right) \right| \\ &\leq 8 \left(\mathbb{E} \left| K_2 \left(\frac{d(x_j, X_j)}{h} \right) \right|^p \right)^{1/p} \mathbb{E}(|\tilde{h}(Y_j)|^p)^{1/p} \beta(|i-j|)^{1-2/p} \\ &\lesssim \phi(h) \mathbb{E}(|\tilde{h}(Y_j)|^p)^{1/p} \beta(|i-j|)^{1-2/p}, \end{aligned}$$

By (6.66), we find that

$$\frac{1}{n} \text{Var} \left(\tilde{\mathbb{U}}_{i,n}^{(3)} \right) \rightarrow 0.$$

We should now demonstrate that the summation of variables in $\tilde{\mathbb{U}}_{1,n}^{(1)}$ are asymptotically independent so that we can use it after the conditions of Lindeberg-Feller for asymptotic finite normality. An application of Lemma 6.9, where $\hat{\mathbb{U}}_{a,n}^{(1)}$ is $\mathcal{F}_{i_a}^{j_a}$ -measurable with $i_a = a(a_n + b_n) + 1$ and $j_a = a(a_n + b_n) + a_n$, gives us

$$\left| \mathbb{E} \left(\exp(itn^{-1/2} \tilde{\mathbb{U}}_{1,n}^{(1)}) - \prod_{i=0}^{\nu_n-1} \mathbb{E} \left(\exp(itn^{-1/2} \hat{\mathbb{U}}_{i,n}^{(1)}) \right) \right| \leq 16\nu_n \beta(b_n + 1), \quad (6.74)$$

which tends to zero using (6.66), then the asymptotic independence is achieved. We can see also that

$$\begin{aligned} \frac{1}{n} \text{var} \left(\tilde{\mathbb{U}}_{1,n}^{(1)} \right) &\lesssim \frac{\nu_n a_n}{n \phi^2(h) N^2} \times \left[\int_{[0,h]} K_1^2 \left(\frac{\omega}{h} \right) d\omega \right] \times \text{Var} \left(K_2 \left(\frac{d(x_i, X_i)}{h} \right) \tilde{h}(Y_i) \right) \\ &\rightarrow \frac{1}{\phi^2(h) N^2} \times \left[\int_{[0,h]} K_1^2 \left(\frac{\omega}{h} \right) d\omega \right] \times \text{Var} \left(K_2 \left(\frac{d(x_i, X_i)}{h} \right) \tilde{h}(Y_i) \right) \text{ (since } \nu_n a_n / n \rightarrow 1 \text{)}. \\ &:= \mathbb{V}(X, Y). \end{aligned} \quad (6.75)$$

So far, we have the last condition of finite dimensional convergence to establish. Notice that, for n is sufficiently large, $\{|\hat{\mathbb{U}}_{i,n}^{(1)}| > \varepsilon \mathbb{V}(X, Y) \sqrt{n}\}$ is empty, so

$$\frac{1}{n} \sum_{i=0}^{\nu_n-1} \mathbb{E} \left(\hat{\mathbb{U}}_{i,n}^{(1)2} \mathbb{1}_{\{|\hat{\mathbb{U}}_{i,n}^{(1)}| > \varepsilon \mathbb{V}(X, Y) \sqrt{n}\}} \right) \rightarrow 0. \quad (6.76)$$

Hence, the demonstration of finite-dimensional convergences is complete. We end up with asymptotic equicontinuity. We have to prove that:

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \sqrt{nh^m \phi(h)} \left\| \tilde{\mathbb{U}}_{1,n}(h_n, \mathbf{i}) \right\|_{\mathcal{F}\mathcal{H}(\delta, \|\cdot\|_p)} > \varepsilon \right\} = 0, \quad (6.77)$$

where,

$$\begin{aligned} \mathcal{F}\mathcal{H}(\delta, \|\cdot\|_p) &:= \left\{ \tilde{\mathbb{U}}'_{1,n}(h_n, \mathbf{i}) - \tilde{\mathbb{U}}''_{1,n}(h_n, \mathbf{i}) : \right. \\ &\quad \left. \left\| \tilde{\mathbb{U}}'_{1,n}(h_n, \mathbf{i}) - \tilde{\mathbb{U}}''_{1,n}(h_n, \mathbf{i}) \right\| < \delta, \tilde{\mathbb{U}}'_{1,n}(h_n, \mathbf{i}), \tilde{\mathbb{U}}''_{1,n}(h_n, \mathbf{i}) \in \mathcal{F}\mathcal{H} \right\} \end{aligned} \quad (6.78)$$

for

$$\tilde{\mathbb{U}}'_{1,n}(h_n, \mathbf{i}) = N^{-1} \frac{1}{n} \sum_{i=1}^n \xi_i \frac{1}{\phi(h)} K_{2,1} \left(\frac{d(x_i, X_i)}{h} \right) \tilde{h}_1(Y_i) - \mathbb{E} \left(\mathbb{U}'_{1,n}(h_n, \mathbf{i}) \right)$$

$$\tilde{\mathbb{U}}''_{1,n}(h_n, \mathbf{i}) = N^{-1} \frac{1}{n} \sum_{i=1}^n \xi_i \frac{1}{\phi(h)} K_{2,2} \left(\frac{d(x_i, X_i)}{h} \right) \tilde{h}_2(Y_i) - \mathbb{E} \left(\mathbb{U}''_{1,n}(h_n, \mathbf{i}) \right) \quad (6.79)$$

At this point, we will use also the chaining technique of Arcones and Yu, 1994, and by Bouzebda et al., 2019b for the conditional setting. The main idea is to break down a sequence (X_1, \dots, X_n) , into $2\nu_n$, equal-sized blocks, that each one is of length equal to a_n and a remainder block of length $n - 2\nu_n a_n$ that is (for $1 \leq j \leq \nu_n$):

$$\begin{aligned} H_j &= \{i : 2(j-1)a_n + 1 \leq i \leq (2j-1)a_n\}, \\ T_j &= \{i : (2j-1)a_n + 1 \leq i \leq 2ja_n\}, \\ R &= \{i : (2\nu_n a_n + 1 \leq i \leq n)\}. \end{aligned}$$

The values of ν_n, a_n are given in the following. Another ingredient is essential, in this proof, that is a sequence of independent blocks $(\zeta_1, \dots, \zeta_n)$ such as:

$$\mathcal{L}(\zeta_1, \dots, \zeta_n) = \mathcal{L}(X_1, \dots, X_{a_n}) \times \mathcal{L}(X_{a_n+1}, \dots, X_{2a_n}) \times \dots$$

In the same line as Bouzebda et al., 2019b, the results of the work of Eberlein, 1984 on β -mixing are applied, and get, for any measurable set A :

$$\begin{aligned} & \left| \mathbb{P} \left\{ \left(\zeta_1, \dots, \zeta_{a_n}, \zeta_{2a_n+1}, \dots, \zeta_{3a_n}, \dots, \zeta_{2(\nu_n-1)a_n+1}, \dots, \zeta_{2\nu_n a_n} \right) \in A \right\} \right. \\ & \quad \left. - \mathbb{P} \left\{ \left(X_1, \dots, X_{a_n}, X_{2a_n+1}, \dots, X_{3a_n}, \dots, X_{2(\nu_n-1)a_n+1}, \dots, X_{2\nu_n a_n} \right) \in A \right\} \right| \\ & \leq 2(\nu_n - 1)\beta_{a_n}. \end{aligned} \quad (6.80)$$

And we will work only with the independent blocks called $\zeta_i = (\eta_i, \varsigma_i)$ sequences instead of working with the dependence variables, we will use the same strategy as in Lemma 6.2 to pass from the sequence of locally stationary random variables to the stationary one:

$$\begin{aligned} & \mathbb{P} \left\{ \left\| (n\phi(h))^{-1/2} h^{m/2} N^{-1} \sum_{j=1}^{\nu_n} \left(\xi_j K_2 \left(\frac{d(x_j, X_j)}{h} \right) \tilde{h}(Y_j) - \mathbb{E}(\mathbb{U}_{1,n}(h_n, \mathbf{i})) \right) \right\|_{\mathcal{F}\mathcal{H}(b, \|\cdot\|_p)} > \varepsilon \right\} \\ & \leq 2\mathbb{P} \left\{ \left\| (n\phi(h))^{-1/2} h^{m/2} N^{-1} \sum_{j=1}^{\nu_n} \sum_{i \in H_j} \left(\xi_i K_2 \left(\frac{d(x_i, \eta_i)}{h} \right) \tilde{h}(\varsigma_i) - \mathbb{E}(\mathbb{U}_{1,n}(h_n, \mathbf{i})) \right) \right\|_{\mathcal{F}\mathcal{H}(b, \|\cdot\|_p)} > \varepsilon' \right\} \\ & \quad + 2(\nu_n - 1)\beta_{a_n} + o_{\mathbb{P}}(1). \end{aligned} \quad (6.81)$$

We choose

$$a_n = [(\log n)^{-1} (n^{p-2} \phi^p(h_K))^{1/2(p-1)}] \text{ and } \nu_n = \left\lfloor \frac{n}{2a_n} \right\rfloor - 1.$$

Making use of the condition (v) of Assumption 4, we get $(\nu_n - 1)\beta_{a_n} \rightarrow 0$ as $n \rightarrow 0$, then it's just a matter of the first term in the right-hand side of (6.81). The blocks being

independent, we symmetrize using a sequence $\{\varepsilon_j\}_{j \in \mathbb{N}^*}$ of i.i.d. Rademacher variables, i.e., r.v.'s with

$$\mathbb{P}(\varepsilon_j = 1) = \mathbb{P}(\varepsilon_j = -1) = 1/2.$$

It should be noted that the sequence $\{\varepsilon_j\}_{j \in \mathbb{N}^*}$ is independent of the sequence $\{\boldsymbol{\xi}_i = (\varsigma_i, \zeta_i)\}_{i \in \mathbb{N}^*}$, thus it remains to establish, for all $\varepsilon > 0$,

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \left\| (n\phi(h))^{-1/2} h^{m/2} N^{-1} \sum_{j=1}^{\nu_n} \sum_{i \in H_j} \left(\xi_i K_2 \left(\frac{d(x_i, \eta_i)}{h} \right) \tilde{h}_{(\varsigma_i)} - \mathbb{E}(\mathbb{U}_{1,n}(h_n, \mathbf{i})) \right) \right\|_{\mathcal{F}\mathcal{H}^{(b, \|\cdot\|_p)}} > \varepsilon \right\} \quad (6.82)$$

Define the semi-norm:

$$\begin{aligned} \tilde{d}_{n\phi,2} := & \left((n\phi(h))^{-1/2} h^{m/2} N^{-1} \sum_{j=1}^{\nu_n} \sum_{i \in H_j} \left| \left(\xi_i K_{2,1} \left(\frac{d(x_i, \eta_i)}{h} \right) \tilde{h}_{1(\varsigma_i)} - \mathbb{E}(\mathbb{U}'_{1,n}(h_n, \mathbf{i})) \right) \right. \right. \\ & \left. \left. - \left(\xi_i K_{2,2} \left(\frac{d(x_i, \eta_i)}{h} \right) \tilde{h}_{2(\varsigma_i)} - \mathbb{E}(\mathbb{U}''_{1,n}(h_n, \mathbf{i})) \right) \right|^2 \right)^{1/2} \end{aligned} \quad (6.83)$$

and the covering number defined for any class of functions \mathcal{E} by :

$$\tilde{N}_{n\phi,2}(u, \mathcal{E}) := N_{n\phi,2}(u, \mathcal{E}, \tilde{d}_{n\phi,2}).$$

By the latter, we can bound (6.82), (more details are in Bouzebda et al., 2019b). In the same way, as in Bouzebda et al., 2019b and back in Arcones and Yu, 1994, as a result of the independence between the blocks and Assumption 6 ii), and by applying Giné et al., 1984, Lemma 5.2, the equicontinuity is achieved, and then the weak convergence too.

Now, we need to show that :

$$\mathbb{P} \left\{ \left\| \mu_n^{(R)}(\varphi, \mathbf{t}) \right\|_{\mathcal{F}_m \mathcal{H}^m} > \lambda \right\} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

For clarity purposes, we restrict ourselves to $m = 2$. We have:

$$\begin{aligned} \mu_n^{(R)}(\varphi, \mathbf{i}) &= \sqrt{n\phi(h_n)} \left\{ \mathbb{U}_n^{(R)}(\varphi, \mathbf{i}) - \mathbb{E} \left(\mathbb{U}_n^{(R)}(\varphi, \mathbf{i}) \right) \right\} \\ &= \frac{\sqrt{n\phi(h_n)}}{n(n-1)} \sum_{i_1 \neq i_2}^n \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{t}}^{(R)} \left(((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right) \right. \\ &\quad \left. - \mathbb{E} \left[G_{\varphi, \mathbf{i}}^{(R)} \left(((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right) \right] \right\} \\ &\leq \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p \neq q}^{\nu_n} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{i}}^{(R)} \left(((X_i, X_j), (Y_i, Y_j)) \right) \right. \\ &\quad \left. - \mathbb{E} \left[G_{\varphi, \mathbf{i}}^{(R)} \left(((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right) \right] \right\} \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2, i_1, i_2 \in H_p^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{i}}^{(R)}((X_i, X_j), (Y_i, Y_j)) \right. \\
 & \qquad \qquad \qquad \left. - \mathbb{E} \left[G_{\varphi, \mathbf{i}}^{(R)}((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right] \right\} \\
 & + 2 \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q: |q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{i}}^{(R)}((X_i, X_j), (Y_i, Y_j)) \right. \\
 & \qquad \qquad \qquad \left. - \mathbb{E} \left[G_{\varphi, \mathbf{i}}^{(R)}((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right] \right\} \\
 & + 2 \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q: |q-p| \leq 1} \sum_{i_2 \in T_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{t}}^{(R)}((X_i, X_j), (Y_i, Y_j)) \right. \\
 & \qquad \qquad \qquad \left. - \mathbb{E} \left[G_{\varphi, \mathbf{t}}^{(R)}((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right] \right\} \\
 & + \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p \neq q} \sum_{i_1 \in T_p^{(U)}} \sum_{i_2 \in T_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{t}}^{(R)}((X_i, X_j), (Y_i, Y_j)) \right. \\
 & \qquad \qquad \qquad \left. - \mathbb{E} \left[G_{\varphi, \mathbf{t}}^{(R)}((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right] \right\} \\
 & + \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p=1}^{v_n} \sum_{i_1 \neq i_2} \sum_{i_1, i_2 \in T_p^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{t}}^{(R)}((X_i, X_j), (Y_i, Y_j)) \right. \\
 & \qquad \qquad \qquad \left. - \mathbb{E} \left[G_{\varphi, \mathbf{t}}^{(R)}((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right] \right\} \\
 & =: \text{I}' + \text{II}' + \text{III}' + \text{IV}' + \text{V}' + \text{VI}' .
 \end{aligned}$$

We will use blocking arguments and treat the resulting terms. We start by considering the first I'. We have

$$\begin{aligned}
 & \mathbb{P} \left\{ \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{i}}^{(R)}((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right. \right. \right. \\
 & \qquad \qquad \qquad \left. \left. - \mathbb{E} \left[G_{\varphi, \mathbf{t}}^{(R)}((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right] \right\} \right\|_{\mathcal{F}_2, \mathcal{H}^2} > \delta \right\} \\
 & \leq \mathbb{P} \left\{ \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p \neq q} \sum_{i_1 \in H_p^{(U)}} \sum_{i_2 \in H_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{i}}^{(R)}((\zeta_{i_1}, \zeta_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) \right. \right. \right. \\
 & \qquad \qquad \qquad \left. \left. - \mathbb{E} \left[G_{\varphi, \mathbf{i}}^{(R)}((\zeta_{i_1}, \zeta_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) \right] \right\} \right\|_{\mathcal{F}_2, \mathcal{H}^2} > \delta \right\} + 2v_n \beta(b_n) .
 \end{aligned}$$

Notice that $v_n \beta_{b_n} \rightarrow 0$ and recall that for all $\varphi \in \mathcal{F}_m$, and :

$$\mathbf{x} \in \mathcal{H}^2, \mathbf{y} \in \mathcal{Y}^2 : \mathbf{1}_{\{d(\mathbf{x}, X_{i,n}) \leq h\}} F(\mathbf{y}) \geq \varphi(\mathbf{y}) K_2 \left(\frac{d(\mathbf{x}, X_{i,n})}{h} \right) .$$

Hence, by the symmetry of $F(\cdot)$:

$$\begin{aligned}
 & \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p \neq q} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{i}}^{(R)}((\zeta_{i_1}, \zeta_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) \right. \right. \\
 & \quad \left. \left. - \mathbb{E} \left[G_{\varphi, \mathbf{t}}^{(R)}((\zeta_{i_1}, \zeta_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) \right] \right\} \right\|_{\mathcal{F}_2 \mathcal{X}^2} \\
 & \lesssim \left| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p \neq q} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ F(\zeta_i, \zeta_j) \mathbb{1}_{\{F > \lambda_n\}} \right. \right. \\
 & \quad \left. \left. - \mathbb{E} \left[F(\zeta_i, \zeta_j) \mathbb{1}_{\{F > \lambda_n\}} \right] \right\} \right|. \tag{6.84}
 \end{aligned}$$

We are going to use Chebyshev's inequality, Hoeffding's trick and inequality, respectively to obtain:

$$\begin{aligned}
 & \mathbb{P} \left\{ \left| \frac{1}{\sqrt{n\phi(h)}} \sum_{p \neq q} \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ F(\zeta_i, \zeta_j) \mathbb{1}_{\{F > \lambda_n\}} \right. \right. \right. \\
 & \quad \left. \left. - \mathbb{E} \left[F(\zeta_i, \zeta_j) \mathbb{1}_{\{F > \lambda_n\}} \right] \right\} \right| > \delta \right\} \\
 & \lesssim \delta^{-2} n^{-1} \phi^{-1}(h) \text{Var} \left(\sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} F(\zeta_i, \zeta_j) \mathbb{1}_{\{F > \lambda_n\}} \right) \\
 & \lesssim c_2 v_n \delta^{-2} n^{-1} \phi^{-1}(h) \text{Var} \left(\sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} F(\zeta_i, \zeta_j) \mathbb{1}_{\{F > \lambda_n\}} \right) \\
 & \lesssim 2c_2 v_n \delta^{-2} n^{-2} \phi^{-1}(h) \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \mathbb{E} \left[(F(\zeta_1, \zeta_2))^2 \mathbb{1}_{\{F > \lambda_n\}} \right]. \tag{6.85}
 \end{aligned}$$

Under Assumption 6 iii), we have for each $\lambda > 0$:

$$\begin{aligned}
 & c_2 v_n \delta^{-2} n^{-2} \phi^{-1}(h_n) \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \mathbb{E} \left[(F(\zeta_1, \zeta_2))^2 \mathbb{1}_{\{F > \lambda_n\}} \right] \\
 & = c_2 v_n \delta^{-2} n^{-2} \phi^{-1}(h_n) \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \int_0^\infty \mathbb{P} \left\{ (F(\zeta_1, \zeta_2))^2 \mathbb{1}_{\{F > \lambda_n\}} \geq t \right\} dt \\
 & = c_2 v_n \delta^{-2} n^{-2} \phi^{-1}(h_n) \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \int_0^{\lambda_n} \mathbb{P} \{ F > \lambda_n \} dt \\
 & + c_2 v_n \delta^{-2} n^{-2} \phi^{-1}(h_n) \sum_{i_1 \in \mathbb{H}_p^{(U)}} \sum_{i_2 \in \mathbb{H}_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \int_{\lambda_n}^\infty \mathbb{P} \{ (F)^2 > t \} dt,
 \end{aligned}$$

which tends to 0 as $n \rightarrow \infty$. Terms II', V' and VI' will be treated in the same way as the previous term. The terms II', VI' do not follow the same line because the variables $\{\zeta_i, \zeta_j\}_{i,j \in H_p^{(U)}}$ (or $\{\zeta_i, \zeta_j\}_{i,j \in T_p^{(U)}}$ for VI') belong to the same blocks. Term IV' can be deduced from the study of Terms I' and III'. Considering the term III', we have

$$\begin{aligned}
 & \mathbb{P} \left\{ \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{i}}^{(R)}((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right. \right. \right. \\
 & \quad \left. \left. \left. - \mathbb{E} \left[G_{\varphi, \mathbf{i}}^{(R)}((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right] \right\} \right\|_{\mathcal{F}_2 \mathcal{K}^2} > \delta \right\} \\
 & \leq \mathbb{P} \left\{ \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{i}}^{(R)}((\varsigma_{i_1}, \varsigma_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) \right. \right. \right. \\
 & \quad \left. \left. \left. - \mathbb{E} \left[G_{\varphi, \mathbf{i}}^{(R)}((\varsigma_{i_1}, \varsigma_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) \right] \right\} \right\|_{\mathcal{F}_2 \mathcal{K}^2} > \delta \right\} + \frac{v_n^2 a_n b_n \beta(a_n)}{\sqrt{n\phi(h_n)}}. \tag{6.86}
 \end{aligned}$$

We have also

$$\begin{aligned}
 & \mathbb{P} \left\{ \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{i}}^{(R)}((\varsigma_{i_1}, \varsigma_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) \right. \right. \right. \\
 & \quad \left. \left. \left. - \mathbb{E} \left[G_{\varphi, \mathbf{i}}^{(R)}((\varsigma_{i_1}, \varsigma_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) \right] \right\} \right\|_{\mathcal{F}_2 \mathcal{K}^2} > \delta \right\} \\
 & \leq \mathbb{P} \left\{ \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{i}}^{(R)}((\varsigma_{i_1}, \varsigma_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) \right. \right. \right. \\
 & \quad \left. \left. \left. - \mathbb{E} \left[G_{\varphi, \mathbf{i}}^{(R)}((\varsigma_{i_1}, \varsigma_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) \right] \right\} \right\|_{\mathcal{F}_2 \mathcal{K}^2} > \delta \right\}.
 \end{aligned}$$

Since the equation (6.84) is still satisfied, the problem is reduced to

$$\begin{aligned}
 & \mathbb{P} \left\{ \left| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ F(\zeta_i, \zeta_j) \mathbf{1}_{\{F > \lambda_n\}} \right. \right. \right. \\
 & \quad \left. \left. \left. - \mathbb{E} \left[F(\zeta_i, \zeta_j) \mathbf{1}_{\{F > \lambda_n\}} \right] \right\} \right| > \delta \right\} \\
 & \lesssim \delta^{-2} n^{-1} \phi(h_n) \text{Var} \left(\sum_{p=1}^{v_n} \sum_{i_1 \in H_p^{(U)}} \sum_{q:|q-p| \geq 2} \sum_{i_2 \in T_q^{(U)}} \phi(h_n) \xi_{i_1} \xi_{i_2} F(\zeta_i, \zeta_j) \mathbf{1}_{\{F > \lambda_n\}} \right),
 \end{aligned}$$

we follow the same procedure as in (6.85). The rest has just been shown to be asymptotically negligible. Finally, with $r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \rightarrow \mathbb{E}(\mathbb{U}_n(\varphi, \mathbf{i}))$, and for $(\mathbb{U}_n(\mathbf{1}, \mathbf{i})) \xrightarrow{\mathbb{P}} 1$, the weak convergence of our estimator is accomplished. \square

Appendix

This appendix contains supplementary information that is an essential part of providing a more comprehensive understanding of the chapter.

Lemma 6.3. *Let $I_h = [C_1h, 1 - C_1h]$. Suppose that kernel K_1 satisfies Assumption 2 part(i). Then for $q = 0, 1, 2$ and $m > 1$:*

$$\sup_{u \in I_h} \left| \frac{1}{n^m h^m} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h} \right) \left(\frac{u_k - i_k/n}{h} \right)^q - \int_0^1 \cdots \int_0^1 \frac{1}{h^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{(u_k - v_k)}{h} \right) \left(\frac{(u_k - v_k)}{h} \right)^q \right\} \prod_{k=1}^m dv_k \right| = O \left(\frac{1}{nh^{m+1}} \right).$$

Proof of Lemma 6.3 Notice that

$$\begin{aligned} & \sup_{u \in I_h} \left| \frac{1}{n^m h^m} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h} \right) \left(\frac{u_k - i_k/n}{h} \right)^q - \int_0^1 \cdots \int_0^1 \frac{1}{h^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{(u_k - v_k)}{h} \right) \left(\frac{(u_k - v_k)}{h} \right)^q \right\} \prod_{k=1}^m dv_k \right| \\ &= \sup_{u \in I_h} \left| \frac{1}{h^m} \sum_{\mathbf{i} \in I_n^m} \left\{ \prod_{k=1}^m \frac{1}{n} K_1 \left(\frac{u_k - i_k/n}{h} \right) \left(\frac{u_k - i_k/n}{h} \right)^q - \prod_{k=1}^m \int_{(i_{k-1})/n}^{i_k/n} \left\{ K_1 \left(\frac{(u_k - v_k)}{h} \right) \left(\frac{(u_k - v_k)}{h} \right)^q \right\} dv_k \right\} \right| \\ &= \sup_{u \in I_h} \left| \frac{1}{h^m} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \int_{(i_{k-1})/n}^{i_k/n} K_1 \left(\frac{u_k - i_k/n}{h} \right) \left(\frac{u_k - i_k/n}{h} \right)^q dv_k - \prod_{k=1}^m \int_{(i_{k-1})/n}^{i_k/n} \left\{ K_1 \left(\frac{(u_k - v_k)}{h} \right) \left(\frac{(u_k - v_k)}{h} \right)^q \right\} dv_k \right| \\ &\leq \frac{C}{nh^{m+1}}. \end{aligned}$$

The last inequality can be founded by applying some additional treatment with the help of the mean value theorem and the telescoping arguments. \square

Lemma 6.4. *Suppose that kernel K_1 satisfies Assumption 2 part (i) and let $g : [0, 1] \times \mathcal{H} \rightarrow \mathbb{R}$, $(\mathbf{u}, \mathbf{x}) \mapsto g(\mathbf{u}, \mathbf{x})$ be continuously differentiable with respect to \mathbf{u} . Then,*

$$\begin{aligned} & \sup_{u \in I_h} \left| \frac{1}{n^m h^m} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h} \right) g \left(\frac{i_k}{n}, x_k \right) - \prod_{k=1}^m g(u_k, x_k) \right| \\ &= O \left(\frac{1}{nh^{m+1}} \right) + o(h). \end{aligned} \quad (6.87)$$

Proof of Lemma 6.4 Remark that

$$\begin{aligned} & \sup_{u \in I_h} \left| \frac{1}{n^m h^m} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h} \right) g \left(\frac{i_k}{n}, x_k \right) - \prod_{k=1}^m g(u_k, x_k) \right| \\ &\leq \sup_{u \in I_h} \left| \frac{1}{n^m h^m} \sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m K_1 \left(\frac{u_k - i_k/n}{h} \right) g \left(\frac{i_k}{n}, x_k \right) \right. \\ &\quad \left. - \prod_{k=1}^m \int_0^1 \frac{1}{h^m} K_1 \left(\frac{u_k - v_k}{h} \right) g(v_k, x_k) dv_k \right| \\ &\quad + \sup_{u \in I_h} \left| \prod_{k=1}^m \int_0^1 \frac{1}{h^m} K_1 \left(\frac{u_k - v_k}{h} \right) g(v_k, x_k) dv_k - \prod_{k=1}^m g(u_k, x_k) \right| \\ &\leq \frac{C}{nh^{m+1}} + o(h), \end{aligned}$$

where

$$\begin{aligned} & \sup_{u \in I_h} \left| \prod_{k=1}^m \int_0^1 \frac{1}{h} K_1 \left(\frac{u_k - v_k}{h} \right) g(v_k, x_k) dv_k - \prod_{k=1}^m g(u_k, x_k) \right| \\ &= \sup_{u \in I_h} \left| \sum_{k=1}^m \int_0^1 \frac{1}{h} K_1 \left(\frac{u_k - v_k}{h} \right) g(v_k, x_k) dv_k - g(u_k, x_k) \right. \\ &\quad \left. \times \prod_{j=1}^{k-1} \int_0^1 \frac{1}{h} K_1 \left(\frac{u_j - v_j}{h} \right) g(v_j, x_j) dv_j \prod_{i=k+1}^m g(u_i, x_i) \right| \\ &\leq C \sup_{u \in I_h} \left| \sum_{k=1}^m \int_0^1 \frac{1}{h} K_1 \left(\frac{u_j - v_j}{h} \right) [g(v_k, x_k) - g(u_k, x_k)] dv_k \right| \\ &\leq C \sup_{u \in I_h} \left| \sum_{k=1}^m \int_{-u_k/h}^{(1-u_k)/h} K_1(t_k) [g(u_k + ht_k, x_k) - g(u_k, x_k)] dt_k \right| \\ &\leq C \sup_{u \in I_h} \left| \sum_{k=1}^m \int_{-u_k/h}^{(1-u_k)/h} K_1(t_k) [g'(u_k + \theta_k, x_k) ht_k + o(h)] dt_k \right| \leq o(h). \end{aligned}$$

This completes the proof. \square

The following result is an exponential inequality for strongly mixing sequences given in Liebscher, 1996.

Lemma 6.5. (Liebscher, 1996, Theorem 2.1). Let $\{Z_{i,n}\}$ be a zero-mean triangular array such that $|Z_{i,n}| \leq b_n$ with α -mixing coefficients $\alpha(k)$. Then for any $\varepsilon > 0$ and $S_n \leq n$ with $\varepsilon > 4S_nb_n$,

$$\mathbb{P} \left(\left| \sum_{i=1}^n Z_{i,i}(u, x) \right| \geq \varepsilon \right) \leq 4 \exp \left(-\frac{\varepsilon^2}{64\sigma_{S_n,n}^2 \frac{n}{S_n} + \frac{8}{3}\varepsilon b_n S_n} \right) + 4 \frac{n}{S_n} \alpha(S_n). \quad (6.88)$$

Lemma 6.6. Let $\{Z_{i,n}\}$ be a zero-mean triangular array such that $|Z_{i,n}| \leq b_n$ with β -mixing coefficients $\beta(k)$. Then for any $\varepsilon > 0$ and $S_n \leq n$ with $\varepsilon > 4S_nb_n$,

$$\mathbb{P} \left(\left| \sum_{i=1}^n Z_{i,i}(u, x) \right| \geq \varepsilon \right) \leq 4 \exp \left(-\frac{\varepsilon^2}{64\sigma_{S_n,n}^2 \frac{n}{S_n} + \frac{8}{3}\varepsilon b_n S_n} \right) + 4 \frac{n}{S_n} \beta(S_n). \quad (6.89)$$

Proof of Lemma 6.6 Using Lemma 6.5 and the fact that for any σ -algebra \mathcal{A} and \mathcal{B} , $\alpha(\mathcal{A}, \mathcal{B}) \subseteq \beta(\mathcal{A}, \mathcal{B})$, Lemma 6.8 holds.

Lemma 6.7. (de la Peña, 1992) Let X_1, \dots, X_n be a sequence of independent random elements taking values in a Banach space $(B, \|\cdot\|)$ with $\mathbb{E}X_i = 0$ for all i . Let $\{\varepsilon_i\}$ be a sequence of independent Bernoulli r.v.'s independent of $\{X_i\}$. Then, for any convex increasing function Φ ,

$$\mathbb{E}\Phi \left(\frac{1}{2} \left\| \sum_{i=1}^n X_i \varepsilon_i \right\| \right) \leq \mathbb{E}\Phi \left(\left\| \sum_{i=1}^n X_i \right\| \right) \leq \mathbb{E}\Phi \left(2 \left\| \sum_{i=1}^n X_i \varepsilon_i \right\| \right).$$

Proposition 6.2. (Arcones et al., 1993, Proposition 3.6) Let $\{X_i : i \in n\}$ be a process satisfying, for $m \geq 1$:

$$(\mathbb{E} \|X_i - X_j\|^p)^{1/p} \leq \left(\frac{p-1}{q-1} \right)^{m/2} (\mathbb{E} \|X_i - X_j\|^q)^{1/q}, \quad 1 < q < p < \infty,$$

and the semi-metric : $\rho(j, i) = (\mathbb{E} \|X_i - X_j\|^2)^{1/2}$. There exists a constant $K = K(m)$ such that :

$$\mathbb{E} \sup_{i,j \in n} \|X_i - X_j\| \leq K \int_0^D [\log N(\varepsilon, n, \rho)]^{m/2} d\varepsilon,$$

where D is the ρ -diameter of n .

Corollary 6.3. (Hall et al., 1980) Suppose that X and Y are random variables which are \mathcal{G} and \mathcal{H} -measurable, respectively, and that $\mathbb{E}|X|^p < \infty, \mathbb{E}|Y|^q < \infty$, where $p, q > 1, p^{-1} + q^{-1} < 1$. Then

$$|\mathbb{E}XY - \mathbb{E}X\mathbb{E}Y| \leq 8 \|X\|_p \|Y\|_q [\alpha(\mathcal{G}, \mathcal{H})]^{1-p^{-1}-q^{-1}}.$$

Lemma 6.8. *Suppose that X and Y are random variables which are \mathcal{G} and \mathcal{H} -measurable, respectively, and that $E|X|^p < \infty$, $E|Y|^q < \infty$, where p ,*

$$q > 1, p^{-1} + q^{-1} < 1.$$

Then

$$|\mathbb{E}XY - \mathbb{E}X\mathbb{E}Y| \leq 8\|X\|_p\|Y\|_q[\beta(\mathcal{G}, \mathcal{H})]^{1-p^{-1}-q^{-1}}.$$

Proof of Lemma 6.8: This Lemma follows directly using Corollary 6.3 and the fact that for any σ -algebra \mathcal{A} and \mathcal{B} , $\alpha(\mathcal{A}, \mathcal{B}) \subseteq \beta(\mathcal{A}, \mathcal{B})$. \square

Lemma 6.9. *Let V_1, \dots, V_L be strongly mixing random variables measurable with respect to the σ -algebras $\mathcal{F}_{i_1}^{j_1}, \dots, \mathcal{F}_{i_L}^{j_L}$ respectively with $1 \leq i_1 < j_1 < i_2 < \dots < j_L \leq n$, $i_{l+1} - j_l \geq w \geq 1$ and $|V_j| \leq 1$ for $j = 1, \dots, L$. Then*

$$\left| \mathbb{E} \left(\prod_{j=1}^L V_j \right) - \prod_{j=1}^L \mathbb{E}(V_j) \right| \leq 16(L-1)\alpha(w),$$

where $\alpha(w)$ is the strongly mixing coefficient.

Chapter 7

Nonparametric conditional U -processes for locally stationary random fields under stochastic sampling design

This chapter develops the content of an accepted article published in Mathematics with the required modifications to fit this thesis manuscript.

Objective

W. Stute (Ann. Probab. 19, No. 2 (1991), 812-825) presented the so-called conditional U -statistics generalizing the Nadaraya-Watson estimates of the regression function. Stute demonstrated their pointwise consistency and the asymptotic normality. In this chapter, we extend the results to a more abstract setting. We develop an asymptotic theory of conditional U -statistics for locally stationary random fields $\{\mathbf{X}_{s,A_n} : s \text{ in } R_n\}$ observed at irregularly spaced locations in $R_n = [0, A_n]^d$ a subset of \mathbb{R}^d . We employ a stochastic sampling scheme that may create irregularly spaced sampling sites in a flexible manner and includes both pure and mixed increasing domain frameworks. We specifically examine the rate of the strong uniform convergence and the weak convergence of conditional U -processes when the explicative variable is functional. We examine the weak convergence where the class of functions is either bounded or unbounded and satisfies specific moment conditions. These results are achieved under somewhat general structural conditions pertaining to the classes of functions and the underlying models. The theoretical results developed in this chapter are (or will be) essential building blocks for several future breakthroughs in functional data analysis.

Contents

1	introduction	272
---	------------------------	-----

2	The functional framework	276
2.1	Notation	276
2.2	Generality on the model	277
2.3	Local stationarity	277
2.4	Sampling design	278
2.5	Mixing condition	279
2.6	Generality on the model	281
2.7	Conditions and comments	285
3	Uniform convergence rates for kernel estimators	290
3.1	Hoeffding's decomposition	290
3.2	Strong uniform convergence rate	292
4	Weak convergence for kernel estimators	293
5	Applications	295
5.1	Metric learning	295
5.2	Multipartite Ranking	296
5.3	Set Indexed Conditional U -Statistics	297
5.4	Discrimination	298
6	Extension to the censored case	299
7	The bandwidth selection criterion	303
8	Mathematical developments	305
8.1	A.1. Preliminaries	306

1 introduction

The regression problem has been studied by statisticians and probability theorists for many years, resulting in a vast array of approaches. Various themes have been covered, such as modeling, estimate method applications, tests, and other related topics. In addition to the parametric framework, in which one must estimate a finite number of parameters based on an a priori specified model structure, the nonparametric framework is devoted to data that lacks a priori structural information. As inherent disadvantages, nonparametric processes are susceptible to estimation biases and reductions in convergence rates compared to parametric methods. Kernel nonparametric function estimation techniques have long been of great interest; for good references to research literature and statistical applications in this area, see B. W. Silverman, 1986, Nadaraya, 1989, Härdle, 1990, Wand et al., 1995, Eggermont et al., 2001, Devroye et al., 2001 and the references therein. Even though they are widely used, they are just one of several possible approaches to building reliable function estimators. Despite their popularity, methods such as nearest-neighbor, spline, neural network, and wavelet analysis are examples of these approaches. These techniques have been utilized on a vast range of different types of data. In this article, our focus will be narrowed to the development of consistent kernel-type estimators for the conditional U -statistics in the context of spatial data. Spatial data are typically generated in numerous research fields, such as econometrics, epidemiology, environmental science, image analysis, oceanography, meteorology, geostatistics, etc. These data are typically collected in various fields and treated statistically on measurement sites. Consult Ripley, 1981, Rosenblatt, 1985, Guyon, 1995 and Cressie, 2015 as well as the references contained in these works, to find reliable sources of references to the research literature in this area and discover some statistical applications. In the context of nonparametric estimation for spatial data, the existing papers are mostly concerned with estimating probability density and regression functions. Hence, we will cite some important references Tran, 1990, Tran et al., 1993, Biau et al., 2004, Dabo-Niang et al., 2013, Ndiaye et al., 2022 and the references that they are included. By considering the conditional U -processes, we give a more generic and abstract context based on this research. With many possible applications, the idea of U -statistics (introduced in a landmark work by Hoeffding, 1948), and U -processes have attracted a great deal of interest over the past few decades.

U -processes are effective for resolving intricate statistical issues: density estimation, nonparametric regression tests, and goodness-of-fit tests are among the examples. Specifically, U -processes emerge in statistics in a variety of contexts, such as the terms of higher order in von Mises expansions. In particular, U -statistics assist in the analysis of estimators, and function estimators, with varying degrees of smoothness. For example, Stute, 1993 aimed to analyze the product limit estimator for shortened data, so he

employs almost sure uniform bounds for \mathbb{P} -canonical U -processes. Also, Arcones et al., 2006 introduced two novel normality tests based on U -processes. Likewise, new tests for normality that use as test statistics weighted L_1 -distances between the standard normal density and local U -statistics based on standardized data were introduced by Giné et al., 2007a, 2007b, Schick et al., 2011. In addition, Joly et al., 2016 challenged the estimate of the mean of multivariate functions under the assumption of possibly heavy-tailed distributions and presented the median-of-means based on U -statistics. The applications of U -processes in various statistical applications may also include tests for functions' qualitative features in nonparametric statistics [c.f. S. Lee et al., 2009, Ghosal et al., 2000, Abrevaya et al., 2005], cross-validation for density estimation [Nolan et al., 1987], and establishing limiting distributions of M -estimators [see, e.g., Arcones et al., 1993, Sherman, 1994, de la Peña et al., 1999]. Historically, Arcones et al., 1993 furnishes the necessary and sufficient criteria for the law of large numbers and the sufficient conditions for the central limit theorem for U -processes, equipped by Halmos, 1946, v. Mises, 1947 and Hoffding, 1948, who provided (amongst others) the first asymptotic results for the case that the underlying random variables are independent and identically distributed. However, under weak dependency assumptions, asymptotic outcomes are illustrated in Yoshihara, 1976, Borovkova et al., 2001, in Denker et al., 1983 or just lately in Leucht, 2012 and in more general setting in Leucht et al., 2013 and Bouzebda and Nemouchi, 2022, Bouzebda, Nezzal, and Zari, 2022, Bouzebda and Soukariéh, 2022a; Soukariéh et al., 2022, 2023. The applicability of U -statistics in estimation and machine learning applications is comprehensive. We refer to the U -statistics with random kernels of divergent orders to Frees, 1989, Rempala et al., 1999, C. Heilig et al., 2001, Y. Song et al., 2019 and Soukariéh et al., 2023. Infinite-order U -statistics are helpful tools for creating simultaneous prediction intervals. These constructed intervals are important to quantify ensemble methods' uncertainty such as subbagging and random forests. For additional information on the topic, c.f Peng et al., 2019. The MeanNN method estimation for differential entropy, which was first described by Faivishevsky et al., 2009, is a remarkable instance of the U -statistic. A novel test statistic for goodness-of-fit tests was proposed by Q. Liu et al., 2016 using U -statistics. Using U -statistics, the conference Cléménçon, n.d. proposed a measure to quantify the level of clustering quality exhibited by a partition. The interested reader may refer to quote Borovskikh, 1996, Koroljuk et al., 1994 and A. J. Lee, 1990 for outstanding resources of references on the U -statistics. The book of de la Peña et al., 1999 provides a profound and in-depth view of the notion of U -processes.

In this work, our primary focus is on the scenario, including spatial-functional data. We give an excerpt from Aneiros et al., 2019: *"Functional data analysis (FDA) is a branch of statistics concerned with the analysis of infinite-dimensional variables such as curves, sets, and images. It has undergone phenomenal growth over the past 20 years, stimulated in part by major advances in data collection technology 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." The reader is directed to the works of reference Ramsay et al., 2002 and Ferraty et al., 2006 for an overview of this subject area. These references include fundamental approaches to functional data analysis and a wide range of case studies from diverse disciplines, such as criminology, economics, archaeology, and neurophysiology. It is important to note that the extension of probability theory to random variables taking values in normed vector spaces (for example, Banach and Hilbert spaces), including extensions of certain classical asymptotic limit theorems, predates the recent literature on functional data; the reader is referred to the book Araujo et al., 1980 for more information on this topic. Considering density and mode estimates for data with values in a normed vector space is the focus of the work done by Gasser et al., 1998. The problem of the curse of dimensionality, which occurs when functional data have too many dimensions, is discussed in this study, along with potential solutions to the issue. According to Ferraty et al., 2006, nonparametric models were deemed useful in regression estimation. We could also refer to Bosq, 2000, Horváth et al., 2012, Ling et al., 2018.

Modern empirical process theory has recently been applied to processing functional data. Ferraty et al., 2010 provided the consistency rates of several conditional models, such as the regression function, the conditional cumulative distribution, the conditional density, and others, uniformly over a subset of the explanatory variable. Bouzebda and Chaouch, 2022 extended Kara-Zaitri et al., 2017's UIB consistency to the ergodic setting. Attouch et al., 2019 considered the problem of local linear estimation of the regression function when the regressor is functional and showed strong convergence, with specified rates, uniformly in bandwidth parameters. Ling et al., 2019 examined the k -nearest neighbors (kNN) estimate of the nonparametric regression model for strong mixing functional time series data and determined the uniform, almost complete convergence rate of the k NN estimator under some mild conditions. Bouzebda, 2016 treated the ergodic data and offered a variety of results related to the limiting distribution for the conditional mode in the functional setting, for recent references, c.f Mohammedi et al., 2021, Bouzebda, Mohammedi, et al., 2022, Bouzebda and Nezzal, 2022; Bouzebda, Nezzal, and Zari, 2022, Didi et al., 2022 and Almanjahie, Bouzebda, Chikr Elmezouar, et al., 2022; Almanjahie, Bouzebda, Kaid, et al., 2022b.

Stute, 1991 raised a class of estimators for $r^{(m)}(\varphi, \mathbf{t})$, known as conditional U -statistics, attempted to generalize the Nadaraya-Watson regression function estimations. Foremost, we present Stute's estimators. Consider regular 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 study, the estimation of the conditional expectation, or

regression function, is our primary concern:

$$r^{(m)}(\varphi, \mathbf{t}) = \mathbb{E}(\varphi(Y_1, \dots, Y_m) \mid (\mathbf{X}_1, \dots, \mathbf{X}_m) = \mathbf{t}), \quad \text{for } \mathbf{t} \in \mathbb{R}^{dm}, \quad (1.1)$$

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 :

$$\sup_{\mathbf{x} \in \mathbb{R}^d} |K(\mathbf{x})| =: \kappa < \infty \quad \text{and} \quad \int K(\mathbf{x}) d\mathbf{x} = 1. \quad (1.2)$$

Hence, the class of estimators for $r^{(m)}(\varphi, \mathbf{t})$, given by Stute, 1991, is defined, for each $\mathbf{t} \in \mathbb{R}^{dm}$, as follows:

$$\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n) = \frac{\sum_{\mathbf{i} \in I_n^m} \varphi(Y_{i_1}, \dots, Y_{i_m}) K\left(\frac{\mathbf{t}_1 - \mathbf{X}_{i_1}}{h_n}\right) \dots K\left(\frac{\mathbf{t}_m - \mathbf{X}_{i_m}}{h_n}\right)}{\sum_{\mathbf{i} \in I_n^m} K\left(\frac{\mathbf{t}_1 - \mathbf{X}_{i_1}}{h_n}\right) \dots K\left(\frac{\mathbf{t}_m - \mathbf{X}_{i_m}}{h_n}\right)}, \quad (1.3)$$

where

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

denotes the set of all m -tuples of different integers i_j between 1 and n and $\{h_n := h_n\}_{n \geq 1}$ is a sequence of positive constants that converge to zero with rate $nh_n^m \rightarrow \infty$.

For $m = 1$, the $r^{(m)}(\varphi, t)$ becomes

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

and the estimate of Stute will be transformed to the Nadaraya-Watson estimator of $r^{(1)}(\varphi, t)$.

Behind, A. Sen, 1994 aimed to estimate the rate of uniform convergence in \mathbf{t} of $\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$ to $r^{(m)}(\varphi, \mathbf{t})$. While, the study of Prakasa Rao et al., 1995 developed the limit distributions of $\hat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$, by discussing and contrasting the findings of Stute. Correspondingly, under appropriate mixing conditions, Harel et al., 1996 spread the results of Stute, 1991 to weakly dependent data and employed their findings to validate the Bayes risk consistency of the relevant discrimination rules. Stute, 1996 suggested symmetrized nearest neighbor conditional U -statistics as alternatives to conventional kernel estimators. Fu, 2012 taken into consideration the functional conditional U -statistic and established the finite-dimensional asymptotic normality. Nevertheless, the nonparametric estimate of the conditional U -statistics in the functional data framework had not

received significant attention, despite the subject's relevance. Some recent developments are discussed in references Bouzebda et al., 2021; Bouzebda and Nemouchi, 2020, in which the authors examine the challenges associated with maintaining a uniform bandwidth in a general framework. The test of independence in the functional framework was based on the Kendall statistics, which may be thought of as examples of the U -statistics, for instance, see Jadhav et al., 2019. The extension of the investigation described above to conditional empirical U -processes is theoretically attractive, practically helpful, and technically challenging.

The primary objective of this study is to examine a general framework and the weak convergence's characterization of the regular sequence of random spatial functions based on conditional U -processes. This inquiry is simple, as it is difficult to hold the asymptotic equicontinuity under minimal conditions in this general setting, which constitutes a fundamentally unresolved open subject in the literature. We intend to fill this gap in the literature by merging the findings of Arcones and Yu, 1994 and Bouzebda et al., 2019a; Bouzebda and Nemouchi, 2022 with techniques handling the functional data given in Masry, 2005 and Kurisu, 2022a, 2022b; Kurisu et al., 2021. However, as demonstrated in the following section, the challenge requires much more than "just" merging concepts from the current outcomes. In fact, complex mathematical derivations are necessary to deal with the typical functional data in our context. This requires the successful application of large-sample theoretical tools, which have been established for empirical processes.

The structure of the present article is as follows. Section 2 introduces the functional framework and the definitions requested in our work. The assumptions used in our asymptotic analysis go along with a brief discussion. Section 3 gives the uniform rates of the strong convergence. Section 4 includes the chapter's main results concerning the uniform weak convergence for the conditional U -processes. In Section 5, we provide some potential applications. In Section 6, we consider the conditional U -statistics in the right censored data framework. In Section 7, we present how to select the bandwidth through the cross-validation procedures. All proofs are gathered in Section 8 to prevent interrupting the presentation flow. Finally, some relevant technical results are given in the Appendix.

2 The functional framework

2.1 Notation

For any set $A \subset \mathbb{R}^d$, $|A|$ represents the Lebesgue measure of A and $\llbracket A \rrbracket$ denotes the number of elements in A . For any positive sequence a_n, b_n , we write $a_n \lesssim b_n$ if a constant $C > 0$ independent of n exists such that $a_n \leq Cb_n$ for all n , $a_n \sim b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$, and $a_n \ll b_n$ si $a_n/b_n \rightarrow 0$ as $n \rightarrow \infty$. We use the notation \xrightarrow{d} to indicate convergence in the distribution. We write $X \stackrel{d}{=} Y$ if the random variables X and Y have

the same distribution. $P_{\mathcal{S}}$ will denote the joint probability distribution of the sequence of independent and identically distributed (i.i.d.) random vectors $\{\mathbf{S}_{0,j}\}_{j \geq 1}$, and $P_{\cdot|\mathcal{S}}$ is the conditional probability distribution for $\{\mathbf{S}_{0,j}\}_{j \geq 1}$. Let $E_{\cdot|\mathcal{S}}$ represent the conditional expectation and $\text{Var}_{\cdot|\mathcal{S}}$ represent the variance for $\{\mathbf{S}_{0,j}\}_{j \geq 1}$.

2.2 Generality on the model

In this investigation, we examine the following model:

$$\begin{aligned} & \varphi(\mathbf{Y}_{s_{i_1}, A_n}, \dots, \mathbf{Y}_{s_{i_m}, A_n}) \\ &= r^{(m)} \left(\varphi, \mathbf{X}_{s_{i_1}, A_n}, \dots, \mathbf{X}_{s_{i_m}, A_n}, \frac{\mathbf{s}_{i_1}}{A_n}, \dots, \frac{\mathbf{s}_{i_m}}{A_n} \right) + \prod_{j=1}^m \sigma \left(\frac{\mathbf{s}_{i_j}}{A_n}, \mathbf{x} \right) \varepsilon_{i_j} \\ &= r^{(m)} \left(\varphi, \mathbf{X}_{s_{i_1}, A_n}, \dots, \mathbf{X}_{s_{i_m}, A_n}, \frac{\mathbf{s}_{i_1}}{A_n}, \dots, \frac{\mathbf{s}_{i_m}}{A_n} \right) + \prod_{j=1}^m \varepsilon_{s_{i_j}, A_n}, \mathbf{s}_{i_j} \in R_n, j = 1, \dots, (m, 1) \end{aligned}$$

where $\mathbb{E}[\varepsilon_{s, A_n} | \mathbf{X}_{s, A_n}] = 0$ and $R_n = [0, A_n]^d \subset \mathbb{R}^d$ denotes a sampling region with $A_n \rightarrow \infty$ as $n \rightarrow \infty$. Here, \mathbf{Y}_{s_j, A_n} and \mathbf{X}_{s, A_n} denote random functions in \mathcal{H} and \mathcal{Y} . We consider $\{\mathbf{X}_{s, A_n} : \mathbf{s} \in R_n\}$ as a locally stationary random function field on $R_n \subset \mathbb{R}^d$ ($d \geq 2$). As suggested by Dahlhaus, 1997, locally stationary processes are nonstationary time series in which the parameters of the time series can change over time. Locally in time, they can be modeled by a stationary time series, which makes it possible to use asymptotic theories to estimate the parameters of models that depend on time. Time series analyses mostly look at locally stationary models in a parametric framework with coefficients that change over time.

2.3 Local stationarity

A random function field $\{\mathbf{X}_{s_j, A_n} : \mathbf{s} \in R_n\}$ ($A_n \rightarrow \infty$ as $n \rightarrow \infty$) is considered to be locally stationary if it exhibits behavior that is approximately stationary in the local space. To guaranteed that it is locally stationary around each rescaled space point \mathbf{u} , a process $\{\mathbf{X}_{s, A_n}\}$ can be approximated by a stationary random function field $\{\mathbf{X}_{\mathbf{u}}(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$ stochastically, for instance, see Dahlhaus and Subba Rao, 2006. The following is one possible way to define this idea.

Definition 2.1. The \mathcal{H} -valued stochastic process $\{\mathbf{X}_{s,A_n} : \mathbf{s} \in R_n\}$ denotes locally stationary if for each rescaled time point $\mathbf{u} \in [0, 1]^d$, there exists an associated \mathcal{H} -valued process $\{\mathbf{X}_{\mathbf{u}}(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$ with the following properties:

(i) $\{\mathbf{X}_{\mathbf{u}}(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$ denotes strictly stationary.

(ii) It holds that

$$d(\mathbf{X}_{s,A_n}, \mathbf{X}_{\mathbf{u}}(\mathbf{s})) \leq \left(\left\| \frac{\mathbf{s}}{A_n} - \mathbf{u} \right\|_2 + \frac{1}{A_n^d} \right) U_{s,A_n}(\mathbf{u}) \text{ a.s.}, \quad (2.2)$$

where $\{U_{s,A_n}(\mathbf{u})\}$ denotes a process of positive variables satisfying $\mathbb{E}[(U_{s,A_n}(\mathbf{u}))^\rho] < C$ for some $\rho > 0$, $C < \infty$; C is independent of \mathbf{u} , \mathbf{s} , and A_n . $\|\cdot\|_2$ is arbitrary norms of \mathbb{R}^d .

The concept of local stationarity for real-valued time series was first presented by Dahlhaus, 1997, and Definition 2.1 is a natural extension of that idea.

In addition, the definition we offer is the same as that of van Delft et al., 2018 (Definition 2.1) when \mathcal{H} is the Hilbert space $L^2_{\mathbb{R}}([0, 1])$ of all real-valued functions that are square integrable with respect to the Lebesgue measure on the unit interval $[0, 1]$ with the L_2 -norm given by

$$\|f\|_2 = \sqrt{\langle f, f \rangle}, \quad \langle f, g \rangle = \int_0^1 f(t)g(t)dt,$$

where $f, g \in L^2_{\mathbb{R}}([0, 1])$. In addition to this, the authors provide necessary conditions so that an $L^2_{\mathbb{R}}([0, 1])$ -valued stochastic process $\{X_{t,T}\}$ satisfies (2.2) with $d(f, g) = \|f - g\|_2$ and $\rho = 2$.

2.4 Sampling design

We are going to look at the stochastic sampling strategy in order to accommodate the data that are irregularly spaced. First, defining R_n , the sampling region. Let $\{A_n\}_{n \geq 1}$ be a sequence of positive numbers such that $A_n \rightarrow \infty$ as $n \rightarrow \infty$. We consider the sampling region as follows:

$$R_n = [0, A_n]^d. \quad (2.3)$$

We will discuss the (random) sample designs we will use. Let $f_{\mathbf{s}}(\mathbf{s}_0)$ be a continuous, everywhere positive probability density function on $R_0 = [0, 1]^d$, and let $\{\mathbf{S}_{0,j}\}_{j \geq 1}$ be a sequence of i.i.d. random vectors with probability density $f_{\mathbf{s}}(\mathbf{s}_0)$ such that $\{\mathbf{S}_{0,j}\}_{j \geq 1}$ and $\{\mathbf{X}_{s,A} : \mathbf{s} \in R_n\}$ share a common probability space (Ω, \mathcal{F}, P) and are independent. The

realizations $\mathbf{s}_{0,1}, \dots, \mathbf{s}_{0,n}$ of random vectors $\mathbf{S}_{0,1}, \dots, \mathbf{S}_{0,n}$ by the following relation:

$$\mathbf{s}_j = A_n \mathbf{s}_{0,j}, \quad j = 1, \dots, n.$$

gives the sampling sites $\mathbf{s}_1, \dots, \mathbf{s}_n$

Herein, we assume that $nA_n^{-d} \rightarrow \infty$ as $n \rightarrow \infty$.

Remark 2.2. *In practice, A_n can be derived by taking the sampling region's diameter. We can extend the applicability of the assumption (2.3) to R_n to a broader range of situations, i.e.,*

$$R_n = \prod_{j=1}^d [0, A_{j,n}],$$

where $A_{j,n}$ are sequences of positive constants with $A_{j,n} \rightarrow \infty$ as $n \rightarrow \infty$. To avoid more challenging outcomes, we assumed (2.3). For additional discussion, please refer to Hall et al., 1994, Matsuda et al., 2009, Lahiri, 2003a, Lahiri, 2003b, Chapter 12, Kurisu et al., 2021 and Kurisu, 2022b.

2.5 Mixing condition

The sequence $\mathbf{Z}_1, \mathbf{Z}_2$, is said β -mixing or absolute regular, refer to Volkonskiui et al., 1959 and Rosenblatt, 1956, if :

$$\beta(k) := \mathbb{E} \sup_{l \geq 1} \left\{ \left| \mathbb{P} \left(A | \sigma_l^t \right) - \mathbb{P} (A) \right| : A \in \sigma_{l+k}^\infty \right\} \longrightarrow 0 \text{ as } k \rightarrow \infty.$$

Notably, Ibragimov et al., 1969 produced a comprehensive description of stationary Gaussian processes matching the last condition. Now we define β -mixing coefficients for a random function field $\widetilde{\mathbf{X}}$. Let $\sigma_{\widetilde{\mathbf{X}}}(T) = \sigma(\{\widetilde{\mathbf{X}}(\mathbf{s}) : \mathbf{s} \in T\})$ be the σ -field generated by variables $\{\widetilde{\mathbf{X}}(\mathbf{s}) : \mathbf{s} \in T\}$, $T \subset \mathbb{R}^d$. For subsets T_1 and T_2 of \mathbb{R}^d , let

$$\bar{\beta}(T_1, T_2) = \sup \frac{1}{2} \sum_{j=1}^J \sum_{k=1}^K |P(A_j \cap B_k) - P(A_j)P(B_k)|,$$

where the supremum is taken over all pairs of (finite) partitions $\{A_1, \dots, A_J\}$ and $\{B_1, \dots, B_K\}$ of \mathbb{R}^d such that $A_j \in \sigma_{\widetilde{\mathbf{X}}}(T_1)$ and $B_k \in \sigma_{\widetilde{\mathbf{X}}}(T_2)$. Furthermore, let

$$d(T_1, T_2) = \inf \{ |\mathbf{x} - \mathbf{y}| : \mathbf{x} \in T_1, \mathbf{y} \in T_2 \},$$

where $|\mathbf{x}| = \sum_{j=1}^d |x_j|$ for $\mathbf{x} \in \mathbb{R}^d$, and let $\mathcal{R}(b)$ be the collection of all finite disjoint unions of cubes in \mathbb{R}^d with a volume total not exceeding b . Subsequently, the β -mixing

coefficients for the random field $\widetilde{\mathbf{X}}$ can be defined as

$$\beta(a; b) = \sup\{\bar{\beta}(T_1, T_2) : d(T_1, T_2) \geq a, T_1, T_2 \in \mathcal{R}(b)\}. \quad (2.4)$$

We assume that a non-increasing function β_1 with $\lim_{a \rightarrow \infty} \beta_1(a) = 0$ and a non-decreasing function g_1 exist such that the β -mixing coefficient $\beta(a; b)$ satisfies the following inequality:

$$\beta(a; b) \leq \beta_1(a)g_1(b), \quad a > 0, b > 0, \quad (2.5)$$

where g_1 may be unbounded for $d \geq 2$.

Remark 2.3 (Some remarks about mixing conditions). *The size of index sets T_1 and T_2 in the definition of $\beta(a; b)$ must be restricted. Let us explain this point. If the β -mixing coefficients of a random field \mathbf{X} are defined similarly to the β -mixing coefficients for the time series as follows: Let \mathcal{O}_1 and \mathcal{O}_2 be half-planes with boundaries L_1 and L_2 , respectively. For each real number $a > 0$, define*

$$\beta(a) = \sup\{\bar{\beta}(\mathcal{O}_1, \mathcal{O}_2) : d(\mathcal{O}_1, \mathcal{O}_2) \geq a\},$$

where sup is taken over all pairs of parallel lines L_1 and L_2 such that $d(L_1, L_2) \geq a$. Subsequently, Bradley, 1989, Theorem 1 shows that if $\{\mathbf{X}(s) : s \in \mathbb{R}^2\}$ is a strictly stationary mixing random field, and $a > 0$ is a real number. Then $\beta(a) = 1$ or 0. This means that if a random field \mathbf{X} is β -mixing ($(\lim_{a \rightarrow \infty} \beta(a) = 0)$), then, for η a positive constant and for some $a > \eta$, the random field \mathbf{X} is "m-dependent", i.e., $\beta(a) = 0$. But, this is highly restricted in practice. In order to loosen these results and make them more flexible for practical purposes, it will be necessary to restrict the size of T_1 and T_2 and adopt Definition 2.4 for the β -mixing. We refer to Kurisu, 2022b, Kurisu et al., 2021, Bradley, 1993, Doukhan, 1994 and Dedecker et al., 2007 for additional information on mixing coefficients for random fields.

Lahiri, 2003a writes the form of mixing condition given in Equation (2.5) for the α -mixing condition and it was considered also in the works of Lahiri et al., 2006 and Bandyopadhyay et al., 2015. We have considered the β -mixing case, and it is well known that the β -mixing implies the α -mixing. In general, in the expression (2.5) β_1 is a function defined in a way that it could be dependent on n as the random field \mathbf{X}_{s, A_n} depends on n , yet, g does not, just for the simplicity sake, despite that the general cases where g changes with n are not difficult. We note that the random field \mathbf{Y}_{s, A_n} (or $\varphi(\mathbf{Y}_{s, A_n})$) not necessary satisfies the mixing condition (2.5), since the mixing condition is assumed for \mathbf{X}_{s, A_n} , but with the regression form represented by the model in (2.1), \mathbf{Y}_{s, A_n} (or $\varphi(\mathbf{Y}_{s, A_n})$) may have a flexible dependence structure.

2.6 Generality on the model

Let $\{\mathbf{X}_{s,A_n}, \mathbf{Y}_{s,A_n} : \mathbf{s} \in R_n\}$ be random variables where \mathbf{Y}_{s,A_n} is in \mathcal{Y} and \mathbf{X}_{s,A_n} takes values in some semi-metric space \mathcal{H} with a semi-metric $d(\cdot, \cdot)$ ¹ defining a topology to measure the proximity between two elements of \mathcal{H} and which is dissociated from the definition of X in order to prevent concerns with measurability. This study aims to establish the weak convergence of the conditional U -process using the following U -statistic.

$$\begin{aligned}
 \widehat{r}_n^{(m)}(\mathbf{x}, \mathbf{u}; h_n) &:= \widehat{r}_n^{(m)}(\varphi, \mathbf{x}, \mathbf{u}; h_n) \\
 &= \frac{\sum_{\mathbf{i} \in I_n^m} \varphi(\mathbf{Y}_{s_{i_1, A_n}}, \dots, \mathbf{Y}_{s_{i_m, A_n}}) \prod_{j=1}^m \left\{ \overline{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j, A_n}}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j, A_n}})}{h_n} \right) \right\}}{\sum_{\mathbf{i} \in I_n^m} \prod_{j=1}^m \left\{ \overline{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j, A_n}}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j, A_n}})}{h_n} \right) \right\}} \\
 &= \frac{\sum_{\mathbf{i} \in I_n^m} \varphi(\mathbf{Y}_{s_{i_1, A_n}}, \dots, \mathbf{Y}_{s_{i_m, A_n}}) \prod_{j=1}^m \left\{ \prod_{\ell=1}^d K_1 \left(\frac{u_{j,\ell} - \frac{s_{i_j, \ell}}{A_n}}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j, A_n}})}{h_n} \right) \right\}}{\sum_{\mathbf{i} \in I_n^m} \prod_{j=1}^m \left\{ \prod_{\ell=1}^d K_1 \left(\frac{u_{j,\ell} - \frac{s_{i_j, \ell}}{A_n}}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j, A_n}})}{h_n} \right) \right\}} \quad (2.6)
 \end{aligned}$$

where

$$\begin{aligned}
 I_n^m &:= \{ \mathbf{i} = (i_1, \dots, i_m) : 1 \leq i_j \leq n \text{ and } i_r \neq i_j \text{ if } r \neq j \}, \\
 \overline{K}(\mathbf{u}) &= \prod_{\ell=1}^d K_1(u_\ell), \quad (2.7)
 \end{aligned}$$

and $\varphi : \mathcal{Y}^m \rightarrow \mathbb{R}$ is a symmetric, measurable function that belongs to some class of functions \mathcal{F}_m , and $\{h_n\}_{n \in \mathbb{N}^*}$ a sequence of positive real numbers satisfying $h_n \rightarrow 0$ as $n \rightarrow \infty$. In order to examine the weak convergence of the conditional empirical process and the conditional U -process under functional data, we must introduce new notations, let

$$\mathcal{F}_m = \{ \varphi : \mathcal{Y}^m \rightarrow \mathbb{R} \},$$

is a point-wise measurable class of real-valued symmetric measurable functions on \mathcal{Y}^m with a measurable envelope function :

$$F(\mathbf{y}) \geq \sup_{\varphi \in \mathcal{F}_m} |\varphi(\mathbf{y})|, \text{ for } \mathbf{y} \in \mathcal{Y}^m. \quad (2.8)$$

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

For a kernel function $K(\cdot)$, we define the point-wise measurable class of functions, for $1 \leq m \leq n$

$$\mathcal{H}^m := \left\{ (x_1, \dots, x_m) \mapsto \prod_{i=1}^m K \left(\frac{d(x_i, \cdot)}{h_n} \right), 0 < h_n < 1 \text{ and } (x_1, \dots, x_m) \in \mathcal{H}^m \right\}.$$

We use the notation

$$\psi(\cdot, \cdot) \in \mathcal{F}_m \mathcal{H}^m := \{ \varphi_1(\cdot) \varphi_2(\cdot) : \varphi_1 \in \mathcal{F}_m, \varphi_2(\cdot) \in \mathcal{H}^m \},$$

and

$$\psi(\cdot, \cdot) \in \mathcal{F}_1 \mathcal{H}^1 := \mathcal{F} \mathcal{H} = \{ \varphi_1(\cdot) \varphi_2(\cdot) : \varphi_1 \in \mathcal{F}_1, \varphi_2(\cdot) \in \mathcal{H}^1 \}.$$

2.6.1 Small ball probability

In the absence of a universal reference measure, such as the Lebesgue measure, the density function of the functional variable does not exist, which is one of the technical challenges in infinite-dimensional spaces. To circumvent this obstacle, we employ the concept of "small-ball probability." The function $\phi_x(\cdot)$ precisely controls the concentration of the probability measure of the functional variable on a small ball, which is defined, for a fixed $x \in \mathcal{H}$ for all $r > 0$, by :

$$\mathbb{P}(X \in B(x, r)) =: \phi_x(r) > 0, \tag{2.9}$$

where the space \mathcal{H} is equipped with the semi-metric $d(\cdot, \cdot)$, and :

$$B(x, r) = \{z \in \mathcal{H} : d(z, x) \leq r\}$$

is a ball in \mathcal{H} with the center x and radius r .

2.6.2 VC-type classes of functions

The asymptotic analysis of functional data is related to concentration properties expressed in terms of the small-ball probability concept. When considering a process indexed by a class of functions, one must also account for other topological concepts, including metric entropy and VC-subgraph classes ("VC" for Vapnik and Červonenkis).

Definition 2.4. Let \mathcal{S}_ε denote a subset of a semi-metric space \mathcal{E} and N_ε a positive integer, a finite set of points $\{e_1, \dots, e_{N_\varepsilon}\} \subset \mathcal{E}$ is called, for a given $\varepsilon > 0$, a ε -net of \mathcal{S}_ε if :

$$\mathcal{S}_\varepsilon \subseteq \bigcup_{j=1}^{N_\varepsilon} B(e_j, \varepsilon).$$

If $N_\varepsilon(\mathcal{S}_\varepsilon)$ denotes the cardinality of the smallest ε -net (the minimal number of open balls of radius ε) in \mathcal{E} , needed to cover \mathcal{S}_ε , then we call Kolmogorov's entropy (metric entropy) of the set \mathcal{S}_ε , the quantity

$$\psi_{\mathcal{S}_\varepsilon}(\varepsilon) := \log N_\varepsilon(\mathcal{S}_\varepsilon).$$

From its name, one can deduce that Kolmogorov invented this idea of metric entropy (cf. Kolmogorov et al., 1959), which was then explored for different metric spaces. This concept was utilized by Dudley, 1967 to provide necessary conditions for the continuity of Gaussian processes. It served as the foundation for remarkable expansions of Donsker's theorem on the weak convergence of empirical processes. $\mathcal{B}_\mathcal{H}$ and $\mathcal{S}_\mathcal{H}$ represent two subsets of the space \mathcal{H} with Kolmogorov's entropy (for the radius ε) $\psi_{\mathcal{B}_\mathcal{H}}(\varepsilon)$ and $\psi_{\mathcal{S}_\mathcal{H}}(\varepsilon)$ respectively, then the Kolmogorov entropy for the subset $\mathcal{B}_\mathcal{H} \times \mathcal{S}_\mathcal{H}$ of the semi-metric space \mathcal{H}^2 by :

$$\psi_{\mathcal{B}_\mathcal{H} \times \mathcal{S}_\mathcal{H}}(\varepsilon) = \psi_{\mathcal{B}_\mathcal{H}}(\varepsilon) + \psi_{\mathcal{S}_\mathcal{H}}(\varepsilon).$$

Hence, $m\psi_{\mathcal{S}_\mathcal{H}}(\varepsilon)$ is the Kolmogorov entropy of the subset $\mathcal{S}_\mathcal{H}^m$ of the semi-metric space \mathcal{H}^m . We specify by d the semi-metric on \mathcal{H} then, this semi-metric defined on \mathcal{H}^m by :

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

for

$$\mathbf{x} = (x_1, \dots, x_m), \mathbf{y} = (y_1, \dots, y_m) \in \mathcal{H}^m.$$

In this type of study, the semi-metric plays a crucial role. The reader can discover helpful discussions on how to select the semi-metric in Ferraty et al., 2006 (see Chapter 3 and Chapter 13). We must additionally consider another topological term; namely, VC-subgraph classes ("VC" for Vapnik and Červonenkis).

Definition 2.5. We call a class of subsets \mathcal{C} on a set C a VC-class if there exists a polynomial $\mathfrak{P}(\cdot)$ such that, for every set of N_ε points in C , the class \mathcal{C} picks out at most $\mathfrak{P}(N_\varepsilon)$ distinct subsets.

Definition 2.6. 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, i.e., 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, VC-class functions are identified by their polynomial covering number (the minimal number of required functions to make a covering on the entire class of functions).

A VC-class of functions \mathcal{F} with envelope function F have the following entropy property, for a given $1 \leq q < \infty$, there are constants a and b such as :

$$N(\varepsilon, \mathcal{F}, \|\cdot\|_{L_q(Q)}) \leq a \left(\frac{(QF^q)^{1/q}}{\varepsilon} \right)^b \quad (2.10)$$

for any $\varepsilon > 0$ and each probability measure such that $QF^q < \infty$. For instance, see Nolan et al., 1987, Lemma 22, Dudley, 2014, §4.7. A. W. van der Vaart et al., 1996, Theorem 2.6.7, Kosorok, 2008, §9.1 provide a number of sufficient conditions under which (2.10) holds, we may refer also to Deheuvels, 2011, §3.2 for further discussions.

2.7 Conditions and comments

Assumption 7. [Model and distribution assumptions]

(M1) The \mathcal{H} -valued stochastic process $\{\mathbf{X}_{\mathbf{s}, A_n} : \mathbf{s} \in R_n\}$ is locally stationary. Hence, for each time point $\mathbf{u} \in [0, 1]^d$, a strictly stationary process $\{\mathbf{X}_{\mathbf{u}}(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$ exists such that, for $\|\cdot\|$ an arbitrary norm on \mathbb{R}^d ,

$$d(\mathbf{X}_{\mathbf{s}, A_n}, \mathbf{X}_{\mathbf{u}}(\mathbf{s})) \leq \left(\left\| \frac{\mathbf{s}}{A_n} - \mathbf{u} \right\|_2 + \frac{1}{A_n^d} \right) U_{\mathbf{s}, A_n}(\mathbf{u}) \text{ a.s.}, \quad (2.11)$$

with $\mathbb{E}[(U_{\mathbf{s}, A_n}(\mathbf{u}))^\rho] < C$ for some $\rho \geq 1$ and $C < \infty$ that is independent of \mathbf{u}, \mathbf{s} and A_n .

(M2) For $i = 1, \dots, m$, let $B(x_i, h) = \{y \in \mathcal{H} : d(x_i, y) \leq h\}$ be a ball centered at $x_i \in \mathcal{H}$ with radius h , and let $c_d < C_d$ be positive constants. For all $\mathbf{u} \in [0, 1]^d$,

$$\phi_{\mathbf{x}}(h_n) := \mathbb{P}(\mathbf{X}_{\mathbf{u}}(\mathbf{s}_1) \in B(x_1, h_n), \dots, \mathbf{X}_{\mathbf{u}}(\mathbf{s}_m) \in B(x_m, h_n)) = F_{\mathbf{u}}(h_n, x_1, \dots, x_m)$$

satisfies :

$$0 < c_d \phi(h) f_1(\mathbf{x}) \leq \phi_{\mathbf{x}}(h) \leq C_d \phi(h) f_1(\mathbf{x}), \quad (2.12)$$

where $\phi(h) \rightarrow 0$ as $h \rightarrow \infty$, and $f_1(\mathbf{x})$ is a nonnegative functional in $\mathbf{x} \in \mathcal{H}^m$. Moreover, there exist constants $C_\phi > 0$ and $\varepsilon_0 > 0$ such that for any $0 < \varepsilon < \varepsilon_0$,

$$\int_0^\varepsilon \phi(u) du > C_\phi \varepsilon \phi(\varepsilon). \quad (2.13)$$

(M3) Let $\mathcal{X}_{\mathbf{s}, A_n} = (\mathbf{X}_{\mathbf{s}_m, A_n}, \dots, \mathbf{X}_{\mathbf{s}_1, A_n})$ and $\mathcal{X}_{\mathbf{v}, A_n} = (\mathbf{X}_{\mathbf{v}_1, A_n}, \dots, \mathbf{X}_{\mathbf{v}_m, A_n})$ and $B(\mathbf{x}, h) = \prod_{i=1}^m B(x_i, h)$. Assume

$$\sup_{\mathbf{s}, \mathbf{x}, A_n} \sup_{\mathbf{s} \neq \mathbf{v}} \mathbb{P}((\mathcal{X}_{\mathbf{s}, A_n}, \mathcal{X}_{\mathbf{v}, A_n}) \in B(\mathbf{x}, h) \times B(\mathbf{x}, h)) \leq \psi(h) f_2(\mathbf{x}),$$

where $\psi(h) \rightarrow 0$ as $h \rightarrow 0$, and $f_2(\mathbf{x})$ is a nonnegative functional in $\mathbf{x} \in \mathcal{H}^m$. We assume that the ratio $\psi(h)/\phi^2(h)$ is bounded.

(M4) $\sigma : [0, 1] \times \mathcal{H}^m \rightarrow \mathbb{R}$ is bounded by some constant $C_\sigma < \infty$ from above and by some constant $c_\sigma > 0$ from below, that is, $0 < c_\sigma \leq \sigma(\mathbf{u}, \mathbf{x}) \leq C_\sigma < \infty$ for all \mathbf{u} and \mathbf{x} .

(M5) $\sigma(\cdot, \cdot)$ is Lipschitz continuous with respect to \mathbf{u} .

(M6) $\sup_{\mathbf{u} \in [0,1]^m} \sup_{z: d(\mathbf{x}, \mathbf{z}) \leq h} |\sigma(\mathbf{u}, \mathbf{x}) - \sigma(\mathbf{u}, \mathbf{z})| = o(1)$ as $h \rightarrow 0$.

(M7) $r^{(m)}(\mathbf{u}, \mathbf{x})$ is Lipschitz, and it satisfy

$$\sup_{\mathbf{u} \in [0,1]^m} |r^{(m)}(\mathbf{u}_1, \mathbf{x}) - r^{(m)}(\mathbf{u}_2, \mathbf{z})| \leq c_m (d_{\mathcal{H}^m}(\mathbf{x}, \mathbf{z})^\alpha + \|\mathbf{u}_1 - \mathbf{u}_2\|^\alpha) \quad (2.14)$$

for some $c_m > 0$ and $\alpha > 0$ and the semi-metric $d_{\mathcal{H}^m}(\mathbf{x}, \mathbf{z})$ is defined on \mathcal{H}^m by:

$$d_{\mathcal{H}^m}(\mathbf{x}, \mathbf{z}) := \frac{1}{m}d(x_1, z_1) + \dots + \frac{1}{m}d(x_m, z_m)$$

for $\mathbf{x} = (x_1, \dots, x_m), \mathbf{z} = (z_1, \dots, z_m) \in \mathcal{H}^m$, and it is twice continuously partially differentiable with first derivatives

$$\partial_{u_i} r^{(m)}(\mathbf{u}, \mathbf{x}) = \frac{\partial}{\partial u_i} r^{(m)}(\mathbf{u}, \mathbf{x}),$$

and second derivatives

$$\partial_{u_i u_j}^2 r^{(m)}(\mathbf{u}, \mathbf{x}) = \frac{\partial^2}{\partial u_i \partial u_j} r^{(m)}(\mathbf{u}, \mathbf{x}).$$

Assumption 8. [Kernel assumptions]

(KB1) The kernel $K_2(\cdot)$ is nonnegative, bounded by $\tilde{\kappa}$, and has support in $[0, 1]$ such that $0 < K_2(0)$ and $K_2(1) = 0$. Moreover, $K_2'(v) = dK_2(v)/dv$ exists on $[0, 1]$ and satisfies $C_1' \leq K_2'(v) \leq C_2'$ for two real constants $-\infty < C_1' < C_2' < 0$.

(KB2) The kernel $\bar{K} : \mathbb{R}^d \rightarrow [0, \infty)$ is bounded and has compact support $[-C, C]^d$. Moreover,

$$\int_{[-C, C]^d} \bar{K}(\mathbf{x}) dx = 1, \int_{[-C, C]^d} \mathbf{x}^\alpha \bar{K}(\mathbf{x}) dx = 0, \text{ for any } \boldsymbol{\alpha} \in \mathbb{Z}^d \text{ with } |\boldsymbol{\alpha}| = 1,$$

$$\text{and } |\bar{K}(\mathbf{u}) - \bar{K}(\mathbf{v})| \leq C \|\mathbf{u} - \mathbf{v}\|.$$

(KB3) The bandwidth h converges to zero at least at a polynomial rate; that is, there exists a small $\xi_1 > 0$ such that $h \leq Cn^{-\xi_1}$ for some constant $0 < C < \infty$.

Assumption 9. [Sampling design assumptions]

(S1) For any $\boldsymbol{\alpha} \in \mathbb{N}^d$ with $|\boldsymbol{\alpha}| = 1, 2$, $\partial^\alpha f_{\mathcal{S}}(\mathbf{s})$ exists and is continuous on $(0, 1)^d$.

(S2) $C_0 \leq nA_n^{-d} \leq C_1 n^\eta$ for some $C_0, C_1 > 0$ and small $\eta \in (0, 1)$.

Assumption 10. [Block decomposition assumptions]

(B1) Let $\{A_{1,n}\}_{n \geq 1}$ and $\{A_{2,n}\}_{n \geq 1}$ be two sequences of positive numbers such that $A_{1,n}, A_{2,n} \rightarrow \infty$, $A_{2,n} = o(A_{1,n})$, and $A_{1,n} = o(A_n)$, or $\frac{A_{1,n}}{A_n} + \frac{A_{2,n}}{A_{1,n}} \leq C_0^{-1} n^{-\eta} \rightarrow 0$ for some $C_0 > 0$ and $\eta > 0$.

(B2) We have $\lim_{n \rightarrow \infty} n A_n^{-d} = \kappa \in (0, \infty]$ with $A_n \geq n^{\bar{\kappa}}$ for some $\bar{\kappa} > 0$.

(B3) We have

$$\left(\frac{1}{n h^{md} \phi(h)} \right)^{1/3} \left(\frac{A_{1,n}}{A_n} \right)^{2d/3} \left(\frac{A_{2,n}}{A_{1,n}} \right)^{2/3} g_1^{1/3} \left(A_{1,n}^d \right) \sum_{k=1}^{A_n/A_{1,n}} k^{d-1} \beta_1^{1/3} (k A_{1,n} + A_{2,n}) \rightarrow 0.$$

(B4) We have $\lim_{n \rightarrow \infty} A_n^d A_{1,n}^{-d} \beta(A_{2,n}; A_n^d) = 0$.

Assumption 11. [Regularity conditions] Let $\alpha_n = \sqrt{\log n / n h^{md} \phi(h)}$. As $n \rightarrow \infty$,

(R1) $h^{-(md)} \phi(h)^{-1} \alpha_n^{md} A_n^d A_{1,n}^{-d} \beta(A_{2,n}; A_n^d) \rightarrow 0$ and $A_{1,n}^d A_n^{-d} n h^{md} \phi(h) (\log n) \rightarrow 0$,

(R2) $n^{1/2} h^{(md)/2} \phi(h)^{1/2} / A_{1,n}^d n^{1/\zeta} \geq C_0 n^\eta$ for some $0 < C_0 < \infty$ and $\eta > 0$ and $\zeta > 2$.

(R3) $A_n^{dp} \phi(h) \rightarrow \infty$, where p is defined in the sequel.

Assumption 12. (E1) For $W_{\mathbf{s}, A_n} = \prod_{j=1}^m \varepsilon_{s_{i_j}, A_n}$, it holds that $\sup_{\mathbf{x} \in \mathcal{H}^m} \mathbb{E} |W_{\mathbf{s}, A_n}|^\zeta \leq C$ and

$$\sup_{\mathbf{x} \in \mathcal{H}^m} \mathbb{E} \left[|W_{\mathbf{s}, A_n}|^\zeta \mid X_{\mathbf{i}, n} = \mathbf{x} \right] \leq C$$

for $\zeta > 2$ and $C < \infty$.

(E2) The β -mixing coefficients of the array $\{\mathbf{X}_{\mathbf{s}, A_n}, W_{\mathbf{s}, A_n}\}$ satisfy $\beta(a; b) \leq \beta_1(a) g_1(b)$ with $\beta_1(a) \rightarrow 0$ as $a \rightarrow \infty$.

Assumption 13. [Class of functions assumptions]

The classes of functions \mathcal{H}^m and \mathcal{F}_m are such that :

(C1) The class of functions \mathcal{F}_m is bounded, and its envelope function satisfies for some $0 < M < \infty$:

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

(C2) The class of functions \mathcal{F}_m is unbounded and its envelope function satisfies for some $p > 2$:

$$\theta_p := \sup_{\mathbf{t} \in \mathcal{H}^m} \mathbb{E} (F^p(\mathbf{Y}) \mid \mathbf{x} = \mathbf{x}) < \infty.$$

(C3) The metric entropy of the class $\mathcal{F}_m \mathcal{H}^m$ satisfies, for some $2 < p < \infty$:

$$\int_0^\infty (\log N(u, \mathcal{F}_m \mathcal{H}^m, L_1(\mathbb{P}^m)))^{\frac{1}{2}} du < \infty,$$

$$\int_0^\infty (\log N(u, \mathcal{F}_m \mathcal{H}^m, L_2(\mathbb{P}^m)))^{\frac{1}{2}} du < \infty,$$

$$\int_0^\infty (\log N(u, \mathcal{F}_m \mathcal{H}^m, L_p(\mathbb{P}^m)))^{\frac{1}{2}} du < \infty.$$

2.7.1 Comments:

When it comes to functional data, traditional statistical methods are entirely ineffective. In our non-parametric functional regression model, we took on the complex theoretical challenge of establishing functional central limit theorems for the conditional U -process, under functional absolute regular data and in the context of a two-fold situation. This was accomplished by adopting a two-fold framework. Despite this, the imposed assumptions coincide with some properties of the infinite-dimensional space. These properties include the topological structure on \mathcal{H}^m , the probability distribution of \mathbf{X} , and the measurability concept for the classes \mathcal{F}_m and \mathcal{H}^m , consequently, a discussion regarding the aforementioned assumptions is necessary. The majority of these assumptions were motivated by Gasser et al., 1998, Masry, 2005, Ferraty et al., 2006, Vogt, 2012, Bouzebda and Nemouchi, 2022, Kurisu, 2022a and Kurisu, 2022b. The assumption 7 is the beginning of a formalization of the property of X_i to be locally stationary, and we continue by placing certain restrictions on the distribution behavior of the variables. This allows us to formalize the property in a more precise manner. The condition (M1) refers to the idea of a locally stationary time series, and various random fields can fulfill this requirement. Kurisu, 2022b gave us some examples, and particularly he proved that this condition is satisfied for locally stationary versions of Lévy-driven moving average random fields. Condition (M2) has been adopted by Masry, 2005, who in turn was inspired by Gasser et al., 1998 in his non-parametric density estimation under functional observations. Masry, 2005 clarifies that if $\mathcal{H}^m = \mathbb{R}^m$, then the condition overlaps with the fundamental axioms of probability calculus, furthermore if \mathcal{H}^m is an infinitely dimensional Hilbert space, then $\phi(h_n)$ can drop toward 0 by an exponential speed as $n \rightarrow \infty$. Equation (2.12) controls the behavior of the small ball probability around zero and is the quite usual condition on the small ball probability. This approximately shows that the small ball probability can be written approximately as the product of two independent functions $\phi^m(\cdot)$ and $f_1(\cdot)$, for instance, for $m = 1$, refer to Mayer-Wolf et al., 1993 for the diffusion process, Bogachev, 1998 for a Gaussian measure, W. V. Li et al., 2001 for a general Gaussian process and Masry, 2005 employed these assumptions for strongly mixing processes. For example, the function $\phi(\cdot)$ can be expressed as $\phi(\varepsilon) = \varepsilon^\delta \exp(-C/\varepsilon^a)$ with $\delta \geq 0$ and $a \geq 0$, and it corresponds to the Ornstein-Uhlenbeck and general diffusion processes (for such processes, $a = 2$ and $\delta = 0$) and the fractal processes (for such processes, $\delta > 0$ and $a = 0$). We refer to the paper of Ferraty et al., 2007 for other examples. Conditions (M4), (M5), (M6) and Assumption 8 represent the regularity conditions, and they are

the umbrella that covers the limiting theorems of such a process. Besides, due to the sampling design strategy employed in section 2.4, a non-uniform density is possible across the sampling region, by which the number of sampling sites is enabled to increase at different rates with respect to the region's volume $O(A_n^d)$. This sampling design allows the pure increasing domain case ($\lim_{n \rightarrow \infty} nA_n^{-d} = \kappa \in (0, \infty)$) and the mixed increasing domain case ($\lim_{n \rightarrow \infty} nA_n^{-d} = \infty$). Assumption 9 is assumed to address this sampling design and the infill sampling criteria in the stochastic design case, which can also be seen in Lahiri et al., 1999 and Lahiri, 2003b. Besides the non-uniform possibility of the sampling density, an approach for irregularly spaced sampling sites based on a homogeneous Poisson point process was discussed in Cressie, 1993, Chapter 8, where the sampling sites must have a uniform distribution over the sampling region. This makes the sampling design used in this work more flexible than the homogeneous Poisson point process and more useful for practical applications. Condition (B1) in 10 is related to the Blocking technique used to decompose the sampling region R_n into big and small blocks. The sequences $\{A_{1,n}\}$ and $\{A_{2,n}\}$ are related to the large-block-small-block argument respectively, which is commonly used in proving CLTs for sums of mixing random variables; see Lahiri, 2003b. Precisely, $A_{1,n}$ corresponds to the side length of large blocks, while $A_{2,n}$ corresponds to the side length of small blocks. Further, Assumptions 12 help for deriving the weak convergence of the conditional U -statistic $\hat{\psi}$ defined in Section 3. Condition (C1) says that we are dealing with bounded functions, but we are also interested in establishing the functional central limit theorem for conditional U -processes indexed by an unbounded class of functions; therefore, this condition will be replaced by (C2). Each of these generic assumptions is sufficiently weak in connection to the many objects described in our preliminary results. They discuss and utilize the four key axes of this work, which are the topological structure of the functional variables, the probability measure in this functional space, the idea of measurability on the class of functions, and the uniformity governed by the entropy characteristics.

Remark 2.7. Note that the Assumption (C4) in 13 might be substituted by more general hypotheses upon moments of \mathbf{Y} as in Deheuvels, 2011. That is

(C4)' We denote by $\{\mathcal{M}(x) : x \geq 0\}$ a non-negative continuous function, increasing on $[0, \infty)$, and such that, for some $s > 2$, ultimately as $x \uparrow \infty$,

$$x^{-s} \mathcal{M}(x) \downarrow; \quad x^{-1} \mathcal{M}(x) \uparrow. \quad (2.15)$$

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

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

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 Uniform convergence rates for kernel estimators

Before expressing the asymptotic behaviour of our estimator represented in (2.6), we will generalize the study to a U -statistic estimator defined by:

$$\hat{\psi}(\mathbf{u}, \mathbf{x}) = \frac{(n-m)!}{n! h^{md} \phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{j=1}^m \left\{ \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j}, A_n})}{h_n} \right) \right\} W_{\mathbf{s}_{\mathbf{i}}, A_n}, \quad (3.1)$$

where $W_{\mathbf{s}_{\mathbf{i}}, A_n}$ is an array of one-dimensional random variables. In this study, we use the results with $W_{\mathbf{s}_{\mathbf{i}}, A_n} = 1$ and $W_{\mathbf{s}_{\mathbf{i}}, A_n} = \prod_{j=1}^m \varepsilon_{\mathbf{s}_{i_j}, A_n}$.

3.1 Hoeffding's decomposition

Note $\hat{\psi}(\mathbf{u}, \mathbf{x})$ is a standard U -statistic with a kernel depending on n . We define

$$\begin{aligned} \xi_j &:= \frac{1}{h^d} \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right), \\ H(Z_1, \dots, Z_m) &:= \prod_{j=1}^m \frac{1}{\phi(h)} K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j}, A_n})}{h_n} \right) W_{\mathbf{s}_{\mathbf{i}}, A_n}, \end{aligned}$$

thus, the U -statistic in (3.1) can be viewed as a weighted U -statistic of degree m :

$$\hat{\psi}(\mathbf{u}, \mathbf{x}) = \frac{(n-m)!}{n!} \sum_{\mathbf{i} \in I_n^m} \xi_{i_1} \dots \xi_{i_m} H(Z_{i_1}, \dots, Z_{i_m}). \quad (3.2)$$

We can write Hoeffding's decomposition in this case. If we will not assume symmetry for W_{s_i, A_n} or H , we must define:

- The expectation of $H(Z_{i_1}, \dots, Z_{i_m})$:

$$\theta(\mathbf{i}) := \mathbb{E}H(Z_{i_1}, \dots, Z_{i_m}) = \int W_{s_i, A_n} \prod_{j=1}^m \frac{1}{\phi(h)} K_2 \left(\frac{d(x_j, \nu_{s_{i_j}, A_n})}{h_n} \right) d\mathbb{P}_{\mathbf{i}}(z_{\mathbf{i}}). \quad (3.3)$$

- For all $\ell \in \{1, \dots, m\}$ the position of the argument, construct the function π_ℓ such that:

$$\pi_\ell(z; z_1, \dots, z_{m-1}) := (z_1, \dots, z_{\ell-1}, z, z_\ell, \dots, z_{m-1}).$$

- Define:

$$\begin{aligned} H^{(\ell)}(z; z_1, \dots, z_{m-1}) &:= H\{\pi_\ell(z; z_1, \dots, z_{m-1})\} \\ \theta^{(\ell)}(i; i_1, i_2, \dots, i_{m-1}) &:= \theta\{\pi_\ell(i; i_1, i_2, \dots, i_{m-1})\}. \end{aligned}$$

Hence, the first order expansion of $H(\cdot)$ will be seen as:

$$\begin{aligned} \widetilde{H}^{(\ell)}(z) &:= \mathbb{E}\left(H^{(\ell)}(z, Z_1, \dots, Z_{m-1})\right) \quad (3.4) \\ &= \int W_{s_{(1, \dots, \ell-1, i, \ell, \dots, m-1)}, A_n} \prod_{\substack{j=1 \\ j \neq i}}^{m-1} \frac{1}{\phi(h)} K_2 \left(\frac{d(x_j, \nu_{s_j, A_n})}{h} \right) \times \frac{1}{\phi(h)} K_2 \left(\frac{d(x_i, \nu_{s_i, A_n})}{h} \right) \\ &\quad \times \mathbb{P}(d\nu_1, \dots, d\nu_{\ell-1}, d\nu_\ell, \dots, d\nu_{m-1}) \\ &:= \frac{1}{\phi(h)} K_2 \left(\frac{d(x_i, X_{s_i, A_n})}{h} \right) W_{s_i, A_n} \times \int W_{s_{(1, \dots, \ell-1, \ell, \dots, m-1)}, A_n} \prod_{\substack{j=1 \\ j \neq i}}^{m-1} \frac{1}{\phi(h)} K_2 \left(\frac{d(x_j, \nu_{s_j, A_n})}{h} \right) \\ &\quad \times \mathbb{P}(d\nu_1, \dots, d\nu_{\ell-1}, d\nu_\ell, \dots, d\nu_{m-1}), \end{aligned}$$

with \mathbb{P} is the underlying probability measure, and define

$$f_{i, i_1, \dots, i_{m-1}}^{(\ell)} := \sum_{\ell=1}^m \xi_{i_1} \dots \xi_{i_{\ell-1}} \xi_i \xi_{i_\ell} \dots \xi_{i_{m-1}} \left(\widetilde{H}^{(\ell)}(z) - \theta^{(\ell)}(i; i_1, \dots, i_{m-1}) \right). \quad (3.5)$$

Then, the first order projection can be defined as:

$$\widehat{H}_{1,i}(\mathbf{u}, \mathbf{x}) := \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} f_{i, i_1, \dots, i_{m-1}}^{(\ell)}, \quad (3.6)$$

where

$$I_{n-1}^{m-1}(-i) := \{1 \leq i_1 < \dots < i_{m-1} \leq n \text{ and } i_j \neq i \text{ for all } j \in \{1, \dots, m-1\}\}.$$

For the remainder terms, we denote by $\mathbf{i} \setminus i_\ell := (i_1, \dots, i_{\ell-1}, i_{\ell+1}, \dots, i_m)$ and for $\ell \in \{1, \dots, m\}$, let

$$H_{2,\mathbf{i}}(\mathbf{z}) := H(\mathbf{z}) - \sum_{\ell=1}^m \widetilde{H}_{\mathbf{i} \setminus i_\ell}^{(\ell)}(z_\ell) + (m-1)\theta(\mathbf{i}), \quad (3.7)$$

where

$$\widetilde{H}_{\mathbf{i} \setminus i_\ell}^{(\ell)}(z_\ell) = \mathbb{E}(H(Z_1, \dots, Z_{\ell-1}, z, Z_{\ell+1}, \dots, Z_{m-1})),$$

defined in (3.4), this projection derive us to the following remainder term:

$$\widehat{\psi}_2(\mathbf{u}, \mathbf{x}) := \frac{(n-m)!}{(n)!} \sum_{\mathbf{i} \in I_n^m} \xi_{i_1} \dots \xi_{i_m} H_{2,\mathbf{i}}(\mathbf{z}). \quad (3.8)$$

Finally, using Equation (3.6) and Equation (3.8), and under conditions that :

$$\mathbb{E}(\widehat{H}_{1,i}(\mathbf{u}, X)) = 0, \quad (3.9)$$

$$\mathbb{E}(H_{2,\mathbf{i}}(\mathbf{Z} \mid Z_k)) = 0 \text{ a.s.}, \quad (3.10)$$

we get the Hoeffding, 1948 decomposition:

$$\begin{aligned} \widehat{\psi}(\mathbf{u}, \mathbf{x}) - \mathbb{E}(\widehat{\psi}(\mathbf{u}, \mathbf{x})) &= \frac{1}{n} \sum_{i=1}^n \widehat{H}_{1,i}(\mathbf{u}, \mathbf{x}) + \widehat{\psi}_2(\mathbf{u}, \mathbf{x}) \\ &:= \widehat{\psi}_1(\mathbf{u}, \mathbf{x}) + \widehat{\psi}_2(\mathbf{u}, \mathbf{x}). \end{aligned}$$

3.2 Strong uniform convergence rate

We start by giving the following general result concerning the rate of convergence of the U -process presented in (3.1).

Proposition 3.1. *Let $\mathcal{F}_m \mathcal{K}^m$ be a measurable VC-subgraph class of functions satisfying Assumption 13 and assume that Assumptions 8, 9, Condition (B1) in 10, 11 and 12 are also satisfied. Then the following result holds*

$$\sup_{\mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in [0,1]^m} |\widehat{\psi}(\mathbf{u}, \mathbf{x}) - \mathbb{E}[\widehat{\psi}(\mathbf{u}, \mathbf{x})]| = O_{P_{\cdot|S}} \left(\sqrt{\frac{\log n}{nh^{md}\phi^m(h)}} \right) \quad P_S - a.s.$$

Next, the uniform rate of convergence of the estimator (2.6) of the mean function $r^{(m)}$ in the model (2.1) will be given, using the results of the last proposition.

Theorem 3.1

Let $\mathcal{F}_m \mathcal{K}^m$ be a measurable VC-subgraph class of functions satisfying Assumption 13. Let $I_h = [C_1 h, 1 - C_1 h]^{dm}$ and let S_c be a compact subset of \mathcal{K}^m . Suppose that

$$\inf_{\mathbf{u} \in [0,1]^d} f(\mathbf{u}) > 0.$$

Then, under Assumptions 7, 8, 9, Condition (B1) in 10, 11 and 12 (with $W_{s_i, A_n} = 1$ and ε_{s_i, A_n}), the following result holds for P_S almost surely:

$$\begin{aligned} & \sup_{\mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{x} \in S_c} \sup_{\mathbf{u} \in I_h} \left| \hat{r}_n^{(m)}(\mathbf{x}, \mathbf{u}; h_n) - r^{(m)}(\mathbf{x}, \mathbf{u}; h_n) \right| \\ &= O_{P_{\cdot|S}} \left(\sqrt{\log n / n h^{md} \phi^m(h)} + h^{2 \wedge \alpha} + \frac{1}{A_n^{dp} \phi(h)} \right), \end{aligned}$$

where $p = \min\{1, \rho\}$, and $\rho > 0$ given in Definition 2.1 .

It is worth to note here that the approximation of the functional time series \mathbf{X}_{s, A_n} by a functional stationary random field $\mathbf{X}_{\mathbf{u}}(s)$ provides the error term $A_n^{-dp} \phi^{-1}(h)$.

4 Weak convergence for kernel estimators

In this section, we are interested in studying the weak convergence of the conditional U -process, defined by Equation (2.6), under absolute regular observations. The following theorem represents the main result in this work concerning the weak convergence of the functional locally stationary random field estimator. Let us define, for $\varphi_1, \varphi_2 \in \mathcal{F}_m$

$$\begin{aligned} \sigma(\varphi_1, \varphi_2) &= E_{\cdot|S} \left(\sqrt{nh^{md} \phi^m(h)} (\hat{r}_n^{(m)}(\varphi_1, \mathbf{x}, \mathbf{u}; h_n) - r^{(m)}(\varphi_1, \mathbf{x}, \mathbf{u})) \right. \\ &\quad \left. \times \sqrt{nh^{md} \phi^m(h)} (\hat{r}_n^{(m)}(\varphi_2, \mathbf{x}, \mathbf{u}; h_n) - r^{(m)}(\varphi_2, \mathbf{x}, \mathbf{u})) \right). \end{aligned} \tag{4.1}$$

Theorem 4.1

Let $\mathcal{F}_m \mathcal{K}^m$ be a measurable VC-subgraph class of functions satisfying Assumption 13. Suppose that $f_S(\mathbf{u}) > 0$ and $\varepsilon_{\mathbf{s}_{i_j}, A_n} = \sigma(\mathbf{s}_{i_j}/A_n, \mathbf{x}) \varepsilon_{i_j}$, where $\sigma(\cdot, \cdot)$ is continuous and $\{\varepsilon_i\}_{i=1}^n$ is a sequence of i.i.d. random variables with mean zero and variance 1. Moreover, suppose $nh^{m(d+1)+4} \rightarrow c_0$ for a constant c_0 . If all assumptions assumed in Theorem 3.1 hold in addition of Conditions (B2), (B3) and (B4), then the following result holds for P_S almost surely:

$$\sqrt{nh^{md}\phi^m(h)} \left[\hat{r}_n^{(m)}(\varphi, \mathbf{x}, \mathbf{u}; h) - r^{(m)}(\varphi, \mathbf{x}, \mathbf{u}) - B_{\mathbf{u}, \mathbf{x}} \right]$$

converges to a Gaussian process \mathbb{G}_n over $\mathcal{F}_m \mathcal{K}^m$, whose simple paths are bounded and informally continuous with respect to $\|\cdot\|_2$ -norm with co-variance function given by (4.1) and where the bias term $B_{\mathbf{u}, \mathbf{x}} = O_{P, |S}(h^{2\wedge\alpha})$.

Remark 4.1. Set $A_n^d = O(n^{1-\bar{\eta}_1})$ for some $\bar{\eta}_1 \in [0, 1)$, $A_{1,n} = O(A_n^{\gamma_{A_1}})$, $A_{2,n} = O(A_n^{\gamma_{A_2}})$ with $0 < \gamma_{A_2} < \gamma_{A_1} < 1/3$ and $p = \min\{1, \rho\} = 1$. Assume that we can take a sufficiently large $\zeta > 2$ such that $\frac{2}{\zeta} < (1 - \bar{\eta}_1)(1 - 3\gamma_{A_1})$. Then, Assumption 10 is satisfied for $d \geq 1$.

Remark 4.2. It is simple to modify the proofs of our results to demonstrate that they still hold when the entropy condition is replaced by the bracketing condition:

$$\int_0^\infty (\log N_{[\]}(u, \mathcal{F} \mathcal{K}, L_p(\mathbb{P}^m)))^{\frac{1}{2}} du < \infty$$

Refer to p. 270 of A. W. van der Vaart, 1998 for the definition of $N_{[\]}(u, \mathcal{F} \mathcal{K}, L_p(\mathbb{P}^m))$.

Remark 4.3. *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 bandwidth, however, is more problematic. It is worth noticing that the choice of the bandwidth is crucial to obtain a good rate of consistency; for example, it has a big influence on the size of the estimate's bias. In general, we are interested in the selection of bandwidth that produces an estimator which has a good balance between the bias and the variance of the considered estimators. It is then more appropriate to consider the bandwidth varying according to the criteria applied and 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 the subject. It would be of interest to establish uniform-in-bandwidth central limit theorems in our setting; i.e., we will let $h > 0$ vary in such a way that $h'_n \leq h \leq h''_n$, where $\{h'_n\}_{n \geq 1}$ and $\{h''_n\}_{n \geq 1}$ are two sequences of positive constants such that $0 < h'_n \leq h''_n < \infty$ and, for either choice of $h_n = h'_n$ or $h_n = h''_n$, fulfills our conditions. It will be of interest to show that*

$$\sup_{h'_n \leq h \leq h''_n} \sqrt{nh^{md}\phi^m(h)} \left[\hat{r}_n^{(m)}(\varphi, \mathbf{x}, \mathbf{u}; h) - r^{(m)}(\varphi, \mathbf{x}, \mathbf{u}) - B_{\mathbf{u}, \mathbf{x}} \right]$$

converges to a Gaussian process \mathbb{G}_n over $\mathcal{F}_m \mathcal{H}^m$.

5 Applications

5.1 Metric learning

Metric learning tries to adapt the metric to the data and has garnered significant interest in recent years; for an overview of metric learning and its applications, see Bellet et al., 2013; Cléménçon et al., 2016. This is prompted by applications ranging from computer vision to bioinformatics-based information retrieval. As an example of the utility of this notion, we give the metric learning issue for supervised classification as described in Cléménçon et al., 2016. Let us consider dependent copies $(\mathbf{X}_{s_1, A_n}, Y_1), \dots, (\mathbf{X}_{s_n, A_n}, Y_n)$ of a $\mathcal{H} \times \mathcal{Y}$ valued random couple (X, Y) , where \mathcal{H} 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 $D : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}_+$. The intuitive objective of metric learning in this context is to identify a measure under which points with the same label are close together and those with different labels are far apart. The standard way to define the risk of a metric D is as follows:

$$R(D) = \mathbb{E} [\phi((1 - D(X, X')) \cdot (2 \mathbf{1}\{Y = Y'\} - 1))], \quad (5.1)$$

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

empirical estimator

$$\begin{aligned}
 R_n(D) &= \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \overline{K} \left(\frac{\mathbf{u}_i - \mathbf{s}_i/A_n}{h_n} \right) \overline{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_j/A_n}{h_n} \right) \\
 &\quad \times \phi \left(\left(D \left(\mathbf{X}_{\mathbf{s}_i, A_n}, \mathbf{X}_{\mathbf{s}_j, A_n} \right) - 1 \right) \cdot (2\pi \{Y_i = Y_j\} - 1) \right), \quad (5.2)
 \end{aligned}$$

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

$$\varphi_D((x, y), (x', y')) = \phi \left((D(x, x') - 1) \cdot (2 \mathbf{1}\{y = y'\} - 1) \right).$$

The convergence to (5.1) of a minimizer of (5.2), in the non-spatial setting, has been studied in the frameworks of algorithmic stability Jin et al., 2009, algorithmic robustness Bellet et al., 2015 and based on the theory of U -processes under appropriate regularization Cao et al., 2016.

5.2 Multipartite Ranking

Let us recall the problem from Cl emen on et al., 2016. Let $X \in \mathcal{H}$ be a random vector of attributes/features and the (temporarily hidden) ordinal labels $Y \in \{1, \dots, K\}$ assigned to it. The goal of multipartite ranking, which uses a training set of labeled examples, is to put the attributes or features in the same order as the labels. Many different fields use this statistical learning problem (e.g., medicine, finance, search engines, e-commerce). Rankings are usually defined by a scoring function, $s : \mathcal{H} \rightarrow \mathbb{R}$, which moves the natural order on the real line to the feature space. The ROC manifold, or its usual summary, the VUS criterion (VUS stands for Volume Under the ROC Surface), is the gold standard for evaluating the ranking performance of $s(x)$; see Cl emen on et al., 2015 and the references therein. The best scoring functions, according to Cl emen on et al., 2013, are those that are best for all bipartite subproblems. More specifically, they are increasing transformations of the likelihood ratio dF_{k+1}/dF_k , where F_k is the class-conditional distribution for the k th class. When the set of optimal scoring functions is not empty, the authors showed that it is the same as the set of functions that maximizes the amount of space under the ROC surface

$$\text{VUS}(s) = \mathbb{P} \{s(\mathbf{X}_{s_1}) < \dots < s(\mathbf{X}_{s_K}) \mid Y_1 = 1, \dots, Y_K = K\}.$$

Given K independent samples $(\mathbf{X}_{s_1, A_{n_k}}^{(k)}, \dots, \mathbf{X}_{s_{n_k}, A_{n_k}}^{(k)})$ with distribution $F_k(dx)$ for $k = 1, \dots, K$, the empirical counterpart of the VUS can be written in the following way:

$$\widehat{\text{VUS}}(s)$$

$$= \frac{1}{\prod_{k=1}^K n_k} \sum_{i_1=1}^{n_1} \cdots \sum_{i_k=1}^{n_k} \prod_{j=1}^K \overline{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \mathbf{1} \left\{ s \left(\mathbf{X}_{\mathbf{s}_{i_1}, A_{n_1}}^{(1)} \right) < \cdots < s \left(\mathbf{s}_{i_K}, A_{n_K}^{(K)} \right) \right\}. \quad (5.3)$$

The empirical VUS (5.3) is a K -sample U -statistic of degree $(1, \dots, K)$ with kernel given by:

$$\varphi_s(x_1, \dots, x_K) = \mathbf{1} \{s(x_1) < \cdots < s(x_K)\}.$$

5.3 Set Indexed Conditional U -Statistics

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

$$\mathbb{G}^{(m)}(C_1 \times \cdots \times C_m \mid \mathbf{t}, \mathbf{u}) = \mathbb{E} \left(\prod_{i=1}^m \mathbf{1} \{Y_i \in C_i\} \mid (\mathbf{X}_1, \dots, \mathbf{X}_k) = (t_1, \dots, t_m) = \mathbf{t} \right) \text{ for } \mathbf{t} \in S_c,$$

where $\mathbf{u} = (u_1, \dots, u_d)$. We define metric entropy with the inclusion of the class of sets \mathcal{C} . For each $\varepsilon > 0$, the covering number is defined as:

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

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

$$\log \mathcal{N}(\varepsilon, \mathcal{C}, \mathbb{G}^{(1)}(\cdot \mid x)) \leq H_\gamma(\varepsilon), \text{ for all } \varepsilon > 0, \quad (5.4)$$

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 et al., 2002, it is worth noticing that the condition (5.4), $\gamma = 0$, 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 (5.4), $\gamma = (d-1)/2$. This and other classes of sets satisfying (5.4) with $\gamma > 0$ can

be found in Dudley, 1984. As a particular case of (2.6), we estimate $\mathbb{G}^{(m)}(C_1 \times \cdots \times C_m \mid \mathbf{t})$

$$\widehat{\mathbb{G}}_n^{(m)}(\tilde{\mathbb{C}}, \mathbf{t}, \mathbf{u}) = \frac{\sum_{i \in I_n^m} \prod_{j=1}^m \mathbf{1}\{\mathbf{Y}_{s_{i_j, A_n}} \in C_j\} \left\{ \prod_{\ell=1}^d K_1 \left(\frac{u_{j, \ell} - \frac{s_{i_j, \ell}}{A_n}}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j, A_n}})}{h_n} \right) \right\}}{\sum_{i \in I_n^m} \prod_{j=1}^m \left\{ \prod_{\ell=1}^d K_1 \left(\frac{u_{j, \ell} - \frac{s_{i_j, \ell}}{A_n}}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j, A_n}})}{h_n} \right) \right\}}. \quad (5.5)$$

One can apply Theorem 3.1 to infer that, in probability,

$$\sup_{\tilde{\mathbb{C}} \in \mathcal{C}^m} \sup_{\mathbf{t} \in S_c, \mathbf{u} \in I_h} \left| \widehat{\mathbb{G}}_n^{(m)}(\tilde{\mathbb{C}}, \mathbf{t}) - \mathbb{G}^{(m)}(\tilde{\mathbb{C}} \mid \mathbf{t}) \right| \longrightarrow 0. \quad (5.6)$$

Remark 5.1. Another point of view is to consider the following situation, for a compact $\mathbf{J} \subset \mathbb{R}^{dm}$,

$$\mathbf{G}^{(m)}(\mathbf{y}_1, \dots, \mathbf{y}_k \mid \mathbf{t}, \mathbf{u}) = \mathbb{E} \left(\prod_{i=1}^m \mathbf{1}\{\mathbf{Y}_i \leq \mathbf{y}_i\} \mid (\mathbf{X}_1, \dots, \mathbf{X}_m) = \mathbf{t} \right) \text{ for } \mathbf{t} \in S_c, (\mathbf{y}_1, \dots, \mathbf{y}_m) \in \mathbf{J}.$$

Let $L(\cdot)$ be a distribution in \mathbb{R}^d and h_n is a sequence of positive real numbers. One can estimate $\mathbf{G}^{(m)}(\mathbf{y}_1, \dots, \mathbf{y}_m \mid \mathbf{t}, \mathbf{u}) = \mathbf{G}^{(m)}(\mathbf{y} \mid \mathbf{t}, \mathbf{u})$ by

$$\widehat{\mathbf{G}}_n^{(m)}(\mathbf{y}, \mathbf{t}, \mathbf{u}) := \frac{\sum_{i \in I_n^m} \prod_{j=1}^m L \left(\frac{\mathbf{Y}_{s_{i_j, A_n}} - t_j}{h_n} \right) \left\{ \prod_{\ell=1}^d K_1 \left(\frac{u_{j, \ell} - \frac{s_{i_j, \ell}}{A_n}}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j, A_n}})}{h_n} \right) \right\}}{\sum_{i \in I_n^m} \prod_{j=1}^m \left\{ \prod_{\ell=1}^d K_1 \left(\frac{u_{j, \ell} - \frac{s_{i_j, \ell}}{A_n}}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j, A_n}})}{h_n} \right) \right\}}.$$

One can use Theorem 3.1 to infer that, in probability,

$$\sup_{\mathbf{t} \in S_c, \mathbf{u} \in I_h} \sup_{\mathbf{y} \in \mathbf{J}} \left| \widehat{\mathbf{G}}_n^{(m)}(\mathbf{y}, \mathbf{t}, \mathbf{u}) - \mathbf{G}^{(m)}(\mathbf{y} \mid \mathbf{t}, \mathbf{u}) \right| \longrightarrow 0. \quad (5.7)$$

5.4 Discrimination

Now, we apply the results on the problem of discrimination described in Section 3 of Stute, 1994c, 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

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

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 $g(\cdot)$, we have

$$\mathbb{P}(g(\mathbf{X}_1, \dots, \mathbf{X}_m) = \varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_m)) \leq \sum_{j=1}^M \int_{\{\mathbf{x}: g(\mathbf{t})=j\}} \max m^j(\mathbf{t}) d\mathbb{P}(\mathbf{t}),$$

where

$$m^j(\mathbf{t}) = \mathbb{P}(\varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_m) = j \mid \mathbf{X}_1, \dots, \mathbf{X}_m = \mathbf{t}), \quad \mathbf{t} \in S_c.$$

The above inequality becomes equality if

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

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

$$\mathbf{L}^* = 1 - \mathbb{P}(g_0(\mathbf{X}_1, \dots, \mathbf{X}_m) = \varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_m)) = 1 - \mathbb{E} \left\{ \max_{1 \leq j \leq M} m^j(\mathbf{t}) \right\}$$

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

$$m_n^j(\mathbf{x}, \mathbf{u}) = \frac{\sum_{i \in I_n^m} \mathbf{1}\{\varphi(\mathbf{Y}_{s_{i_1, A_n}}, \dots, \mathbf{Y}_{s_{i_m, A_n}}) = j\} \prod_{j=1}^m \left\{ \prod_{\ell=1}^d K_1 \left(\frac{u_{j,\ell} - \frac{s_{i_j, \ell}}{A_n}}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j, A_n}})}{h_n} \right) \right\}}{\sum_{i \in I_n^m} \prod_{j=1}^m \left\{ \prod_{\ell=1}^d K_1 \left(\frac{u_{j,\ell} - \frac{s_{i_j, \ell}}{A_n}}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j, A_n}})}{h_n} \right) \right\}}.$$

Set

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

Let us introduce

$$\mathbf{L}_n^* = \mathbb{P}(g_{0,n}(\mathbf{X}_1, \dots, \mathbf{X}_m) \neq \varphi(\mathbf{Y}_1, \dots, \mathbf{Y}_m)).$$

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

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

This follows from the obvious relation

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

6 Extension to the censored case

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

use Maillot et al., 2009 notation and we work with a sample $\{(Y_i, C_i, \mathbf{X}_{s_i, A_n})\}$ of 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 $\delta_i := \mathbf{1}\{Y_i \leq C_i\}$, $1 \leq i \leq n$. Accordingly, the observed sample is

$$\mathcal{D}_n = \{(Z_i, \delta_i, \mathbf{X}_{s_i, A_n}), i = 1, \dots, n\}.$$

Survival data in clinical trials or failure time data in reliability studies, for example, are often subject to such censoring. More specifically, many statistical experiments result in incomplete samples, even under well-controlled conditions. For example, clinical data for surviving most types of disease are usually censored by other competing risks to life which result in death. In the sequel, we impose the following assumptions upon the distribution of (\mathbf{X}, Y) . 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 $\mathbf{x} \in \mathcal{H}$, given by

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

when Y is right-censored. To estimate $r^{(1)}(\psi, \cdot)$, we make use of the Inverse Probability of Censoring Weighted (I.P.C.W.) estimators have recently gained popularity in the censored data literature (see Kohler et al., 2002, Carbonez et al., 1995, Brunel et al., 2006). The key idea of I.P.C.W. estimators is as follows. Introduce the real-valued function $\Phi_\psi(\cdot, \cdot)$ defined on \mathbb{R}^2 by

$$\Phi_\psi(y, c) = \frac{\mathbf{1}\{y \leq c\}\psi(y \wedge c)}{1 - G(y \wedge c)}. \quad (6.1)$$

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, \mathbf{X}) are independent.

We have

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

$$\begin{aligned}
&= \mathbb{E} \left\{ \frac{\mathbf{1}\{Y \leq C\} \psi(Z)}{1 - G(Z)} \mid \mathcal{X} = \mathbf{x} \right\} \\
&= \mathbb{E} \left\{ \frac{\psi(Y)}{1 - G(Y)} \mathbb{E}(\mathbf{1}\{Y \leq C\} \mid \mathbf{X}, Y) \mid \mathbf{X} = \mathbf{x} \right\} \\
&= r^{(1)}(\psi, \mathbf{x}).
\end{aligned} \tag{6.2}$$

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 $\mathbf{x} \in \mathcal{H}$, $h \geq l_n$, $1 \leq i \leq n$,

$$\bar{\omega}_{n, K_{1,2}, h_n, i}^{(1)}(x, u) := \frac{\prod_{\ell=1}^d K_1 \left(\frac{u_\ell - \frac{s_{j,\ell}}{A_n}}{h_n} \right) K_2 \left(\frac{d(x, \mathbf{X}_{s_{j,A_n}})}{h_n} \right)}{\sum_{j=1}^n \prod_{\ell=1}^d K_1 \left(\frac{u_\ell - \frac{s_{j,\ell}}{A_n}}{h_n} \right) K_2 \left(\frac{d(x, \mathbf{X}_{s_{j,A_n}})}{h_n} \right)}. \tag{6.3}$$

We assume that h satisfies **(H.1)**. In view of (6.1), (6.2), and (6.3), whenever $G(\cdot)$ is known, a kernel estimator of $r^{(1)}(\psi, \cdot)$ is given by

$$\check{r}_n^{(1)}(\psi, x, u; h_n) = \sum_{i=1}^n \bar{\omega}_{n, K_{1,2}, h_n, i}^{(1)}(x, u) \frac{\delta_i \psi(Z_i)}{1 - G(Z_i)}. \tag{6.4}$$

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 et al., 1958. Namely, adopting the conventions

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

and $0^0 = 1$ and setting

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

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 $r^{(1)}(\psi, \cdot)$

$$\check{r}_n^{(1)*}(\psi, x, u; h_n) = \sum_{i=1}^n \bar{\omega}_{n, K_{1,2}, h_n, i}^{(1)}(x, u) \frac{\delta_i \psi(Z_i)}{1 - G_n^*(Z_i)}, \tag{6.5}$$

refer to Kohler et al., 2002 and Maillot et al., 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 et al., 2006). To obtain our results, we will work under the following assumptions.

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

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

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

We now have all the ingredients to state the result corresponding to the censored case. By combining the results of Proposition 9.6 and Lemma 9.7 of Bouzebda and El-hadjali, 2020b, Theorem 3.1, we have, in probability,

$$\sup_{x, u} \left| \check{r}_n^{(1)*}(\psi, x, u; h_n) - \widehat{\mathbb{E}}(\check{r}_n^{(1)*}(\psi, x, u; h_n)) \right| \rightarrow 0. \quad (6.6)$$

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 et al., 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, Y. Chen et al., 2019 proposed adjustments to the classical U -statistics. A. Yuan et al., 2017 proposed a different way in the estimation procedure of the U -statistic by using a substitution estimator of the conditional kernel given the observed data. To our best knowledge, the problem of the estimation of the conditional U -statistics was opened up to the present, and it gives and main motivation to the study of this section. A natural extension of the function defined in (6.1) is given by

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

From this, we have an analogous relation to (6.2) given by

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

$$\begin{aligned}
 &= \mathbb{E} \left(\frac{\psi(Y_1, \dots, Y_m)}{\prod_{i=1}^m \{1 - G(Y_i)\}} \mathbb{E} \left(\prod_{i=1}^m \{\mathbf{1}\{Y_i \leq C_i\}\} \mid (Y_1, \mathbf{X}_1), \dots, (Y_m, \mathbf{X}_m) \right) \mid (\mathbf{X}_1, \dots, \mathbf{X}_m) = \mathbf{t} \right) \\
 &= \mathbb{E} (\psi(Y_1, \dots, Y_m) \mid (\mathbf{X}_1, \dots, \mathbf{X}_m) = \mathbf{t}) = m_\psi(\mathbf{t}).
 \end{aligned}$$

An analogue estimator to (2.6) in the censored case is given by

$$\check{r}_n^{(m)}(\psi, \mathbf{t}, \mathbf{u}; h_n) = \sum_{(i_1, \dots, i_m) \in I(m, n)} \frac{\delta_{i_1} \cdots \delta_{i_m} \psi(Z_{i_1}, \dots, Z_{i_m})}{(1 - G(Z_{i_1})) \cdots (1 - G(Z_{i_m}))} \bar{\omega}_{n, K_{1,2}, h_n, \mathbf{i}}^{(m)}(\mathbf{t}, \mathbf{u}), \quad (6.7)$$

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

$$\bar{\omega}_{n, K_{1,2}, h_n, \mathbf{i}}^{(k)}(\mathbf{x}, \mathbf{u}) = \frac{\prod_{j=1}^m \left\{ \prod_{\ell=1}^d K_1 \left(\frac{u_{j,\ell} - \frac{s_{i_j, \ell}}{A_n}}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j, A_n}})}{h_n} \right) \right\}}{\sum_{i \in I_n^m} \prod_{j=1}^m \left\{ \prod_{\ell=1}^d K_1 \left(\frac{u_{j,\ell} - \frac{s_{i_j, \ell}}{A_n}}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j, A_n}})}{h_n} \right) \right\}}. \quad (6.8)$$

The estimator that we will investigate is given by

$$\check{r}_n^{(m)*}(\psi, \mathbf{t}, \mathbf{u}; h_n) = \sum_{(i_1, \dots, i_k) \in I(m, n)} \frac{\delta_{i_1} \cdots \delta_{i_k} \psi(Z_{i_1}, \dots, Z_{i_m})}{(1 - G_n^*(Z_{i_1})) \cdots (1 - G_n^*(Z_{i_m}))} \bar{\omega}_{n, K_{1,2}, h_n, \mathbf{i}}^{(k)}(\mathbf{t}, \mathbf{u}). \quad (6.9)$$

Corollary 6.1. *Under the assumptions (A.1)-(A.2) and the conditions of Theorem 3.1, we have*

$$\begin{aligned}
 &\sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in I_h, \mathbf{x} \in S_c} \left| \check{r}_n^{(m)*}(\psi, \mathbf{t}, \mathbf{u}; h_n) - \mathbb{E} \check{r}_n^{(m)*}(\psi, \mathbf{t}, \mathbf{u}; h_n) \right| \\
 &= O_{P, |S} \left(\sqrt{\log n / n h^{md} \phi^m(h)} + \frac{1}{A_n^{dp} \phi(h)} \right),
 \end{aligned}$$

In the last corollary, we use the law of iterated logarithm for $G_n^*(\cdot)$ established in Földes et al., 1981 ensuring that

$$\sup_{t \leq \tau} |G_n^* - G(t)| = O \left(\sqrt{\frac{\log \log n}{n}} \right) \quad \text{almost surely as } n \rightarrow \infty. \quad (6.10)$$

At this point, we may refer to Bouzebda and El-hadjali, 2020b, Bouzebda, El-hadjali, and Ferfache, 2022 and Bouzebda and Nezzal, 2022.

7 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 et al., 1985 and Rachdi et al., 2007. This parameter has to be selected suitably, either in the standard finite dimensional case, or in the infinite dimensional framework for insuring good practical performances. However, according to our knowledge, such studies do not presently exist for treating a such general functional conditional U -statistic. Nevertheless an extension of the leave-one-out cross validation procedure allows to define, for any fixed $\mathbf{j} = (j_1, \dots, j_m) \in I_n^m$:

$$\begin{aligned} & \hat{r}_{n,\mathbf{j}}^{(m)}(\mathbf{x}, \mathbf{u}; h_n) \\ &= \frac{\sum_{\mathbf{i} \in I_n^m(\mathbf{j})} \varphi(\mathbf{Y}_{s_{i_1, A_n}}, \dots, \mathbf{Y}_{s_{i_m, A_n}}) \prod_{j=1}^m \left\{ \prod_{\ell=1}^d K_1 \left(\frac{u_{j,\ell} - \frac{s_{i_j, \ell}}{A_n}}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j, A_n}})}{h_n} \right) \right\}}{\sum_{\mathbf{i} \in I_n^m} \prod_{j=1}^m \left\{ \prod_{\ell=1}^d K_1 \left(\frac{u_{j,\ell} - \frac{s_{i_j, \ell}}{A_n}}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j, A_n}})}{h_n} \right) \right\}} \end{aligned} \quad (7.1)$$

where :

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

The equation (7.1) represents the leave-out- $(\mathbf{X}_{\mathbf{j}}, \mathbf{Y}_{\mathbf{j}})$ estimator of the functional regression and also could be considered as a predictor of $\varphi(\mathbf{Y}_{s_{j_1, A_n}}, \dots, \mathbf{Y}_{s_{j_m, A_n}}) := \varphi(\mathbf{Y}_{\mathbf{j}})$. In order to minimize the quadratic loss function, we introduce the following criterion, we have for some (known) non-negative weight function $\mathscr{W}(\cdot)$:

$$CV(\varphi, h_n) := \frac{(n-m)!}{n!} \sum_{\mathbf{j} \in I_n^m} \left(\varphi(\mathbf{Y}_{\mathbf{j}}) - \hat{r}_{n,\mathbf{j}}^{(m)}(\mathbf{X}_{\mathbf{j}}, \mathbf{u}; h_n) \right)^2 \widetilde{\mathscr{W}}(\mathbf{X}_{\mathbf{j}}), \quad (7.2)$$

where $\mathbf{X}_{\mathbf{j}} = (\mathbf{X}_{s_{j_1, A_n}}, \dots, \mathbf{X}_{s_{j_m, A_n}})$. Following the ideas developed by Rachdi et al., 2007, a natural way for choosing the bandwidth is to minimize the precedent criterion, so let's choose $\hat{h}_n \in [a_n, b_n]$ minimizing among $h \in [a_n, b_n]$:

$$CV(\varphi, h_n).$$

The main interest of our results is the possibility to derive the asymptotic properties of our estimate even if the bandwidth parameter is a random variable, like in the last equation. Following Benhenni et al., 2007 where the bandwidths are locally chosen by a data-driven method based on the minimization of a functional version of a cross-validated criterion, one can replace (7.2) by

$$CV(\varphi, h_n) := \frac{(n-m)!}{n!} \sum_{\mathbf{j} \in I_n^m} \left(\varphi(\mathbf{Y}_{\mathbf{j}}) - \hat{r}_{n,\mathbf{j}}^{(m)}(\mathbf{X}_{\mathbf{j}}, \mathbf{u}; h_n) \right)^2 \widehat{\mathscr{W}}(\mathbf{X}_{\mathbf{j}}, \mathbf{x}), \quad (7.3)$$

where

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

In practice, one takes for $\mathbf{i} \in I_n^m$, the uniform global weights $\widehat{\mathcal{W}}(\mathbf{X}_i) = 1$, and the local weights

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

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

Remark 7.1. *For notational convenience, we have chosen the same bandwidth sequence for each margin. This assumption can be dropped easily. If one wants to use the vector bandwidths (see, in particular, Chapter 12 of Devroye et al., 2001). With obvious changes of notation, our results and their proofs remain true when h_n is replaced by a vector bandwidth $\mathbf{h}_n = (h_n^{(1)}, \dots, h_n^{(d)})$, where $\min h_n^{(i)} > 0$. In this situation we set $h_n = \prod_{i=1}^d h_n^{(i)}$, and for any vector $\mathbf{v} = (v_1, \dots, v_d)$ we replace \mathbf{v}/h_n by $(v_1/h_n^{(1)}, \dots, v_d/h_n^{(d)})$. For ease of presentation, we chose to use real-valued bandwidths throughout.*

Remark 7.2. *We mention that a different bandwidth criterion suggested by B. W. Silverman, 1986 is the rule of thumb. Strictly speaking, since the cross-validated bandwidth is random, the asymptotic theory can only be justified with this random bandwidth via a specific stochastic equicontinuity argument. Cross-validation is employed by Q. Li et al., 2009 to examine the equality of two unconditional and conditional functions in the context of mixed categorical and continuous data. However, this approach, which is optimal for estimation, loses its optimality when applied to non-parametric kernel testing. For testing a parametric model for conditional mean function against a non-parametric alternative, Horowitz et al., 2001 proposed an adaptive-rate-optimal rule. Gao et al., 2008 present the other method for selecting a proper bandwidth. Ref. Gao et al., 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 is maximized while the size function is controlled. Future investigation will focus on the aforementioned three approaches.*

8 Mathematical developments

The proofs for our results are covered in this section. The following continues to use the notations that were previously presented.

To avoid the repetition of the Blocking technique and the notation used, we will devote the following subsection to introducing all notations needed for this decomposition.

8.1 A.1. Preliminaries

This treatment requires an extension of the Blocking techniques of Bernstein to the spacial process, refer to Kurisu, 2022b. Let us introduce some notations related to this technique. Recall that $\{A_{1,n}\}$ and $\{A_{2,n}\}$ are sequences of positive numbers such that

$$A_{1,n}/A_n + A_{2,n}/A_{1,n} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Let

$$A_{3,n} = A_{1,n} + A_{2,n}.$$

We consider a partition of \mathbb{R}^d by hypercubes of the form $\Gamma_n(\boldsymbol{\ell}; \mathbf{0}) = (\boldsymbol{\ell} + (0, 1]^d) A_{3,n}$, $\boldsymbol{\ell} = (\ell_1, \dots, \ell_d)' \in \mathbb{Z}^d$ and divide $\Gamma_n(\boldsymbol{\ell}; \mathbf{0})$ into 2^d hypercubes as follows:

$$\Gamma_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}) = \prod_{j=1}^d I_j(\varepsilon_j), \boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_d)' \in \{1, 2\}^d, \quad (8.1)$$

where for $j = 1, \dots, d$,

$$I_j(\varepsilon_j) = \begin{cases} (\ell_j A_{3,n}, \ell_j A_{3,n} + A_{1,n}] & \text{if } \varepsilon_j = 1, \\ (\ell_j A_{3,n} + A_{1,n}, (\ell_j + 1) A_{3,n}] & \text{if } \varepsilon_j = 2. \end{cases} \quad (8.2)$$

We note that

$$|\Gamma_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon})| = A_{1,n}^{q(\boldsymbol{\varepsilon})} A_{2,n}^{d-q(\boldsymbol{\varepsilon})} \quad (8.3)$$

for any $\boldsymbol{\ell} \in \mathbb{Z}^d$ and $\boldsymbol{\varepsilon} \in \{1, 2\}^d$, where

$$q(\boldsymbol{\varepsilon}) = [\{1 \leq j \leq d : \varepsilon_j = 1\}].$$

Let $\boldsymbol{\varepsilon}_0 = (1, \dots, 1)'$. The partitions $\Gamma_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}_0)$ correspond to “large blocks” and the partitions $\Gamma_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon})$ for $\boldsymbol{\varepsilon} \neq \boldsymbol{\varepsilon}_0$ correspond to “small blocks”. Let

$$L_{1,n} = \{\boldsymbol{\ell} \in \mathbb{Z}^d : \Gamma_n(\boldsymbol{\ell}, 0) \subset R_n\}$$

be the index set of all hypercubes $\Gamma_n(\boldsymbol{\ell}, 0)$ that are contained in R_n , and let

$$L_{2,n} = \{\boldsymbol{\ell} \in \mathbb{Z}^d : \Gamma_n(\boldsymbol{\ell}, 0) \cap R_n \neq \emptyset, \Gamma_n(\boldsymbol{\ell}, 0) \cap R_n^c \neq \emptyset\}$$

denote the boundary hypercubes index set. Define $L_n = L_{1,n} \cup L_{2,n}$.

Proof of Proposition 3.1

As we mentioned, our statistic is a weighted U -statistic that can be decomposed into a sum of U -statistics using the Hoeffding decomposition. We will treat this decomposition detailed in the Sub-section 3.1 to achieve the desired results. In the mentioned section, we have seen that

$$\widehat{\psi}(\mathbf{u}, \mathbf{x}) - \mathbb{E}(\widehat{\psi}(\mathbf{u}, \mathbf{x})) = \widehat{\psi}_1(\mathbf{u}, \mathbf{x}) + \widehat{\psi}_2(\mathbf{u}, \mathbf{x}),$$

where the linear term $\widehat{\psi}_1(\mathbf{u}, \mathbf{x})$ and the remainder term $\widehat{\psi}_2(\mathbf{u}, \mathbf{x})$ are well defined in (3.6) and (3.8) respectively. We aim to prove that the linear term leads the rate of convergence of this statistic while the remaining one converges to zero almost surely as $n \rightarrow \infty$. We will deal with the first term in the decomposition. For $B = [0, 1]$, $\alpha_n = \sqrt{\log n / nh^{md} \phi^m(h)}$ and $\tau_n = \rho_n n^{1/\zeta}$, where ζ is a positive constant given in Assumption 12 part i), with $\rho_n = (\log n)^{\zeta_0}$ for some $\zeta_0 > 0$. Define

$$\widetilde{H}_1^{(\ell)}(z) := \widetilde{H}^{(\ell)}(z) \mathbf{1}_{\{|W_{s_i, A_n}| \leq \tau_n\}}, \quad (8.4)$$

$$\widetilde{H}_2(z) := \widetilde{H}^{(\ell)}(z) \mathbf{1}_{\{|W_{s_i, A_n}| > \tau_n\}}, \quad (8.5)$$

and

$$\widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}) - \theta(\mathbf{i}) = \frac{1}{n} \sum_{i=1}^n \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_i \xi_{i_\ell} \cdots \xi_{i_{m-1}} \widetilde{H}_1^{(\ell)}(z),$$

$$\widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}) - \theta(\mathbf{i}) = \frac{1}{n} \sum_{i=1}^n \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_i \xi_{i_\ell} \cdots \xi_{i_{m-1}} \widetilde{H}_2^{(\ell)}(z).$$

Clearly, we have

$$\begin{aligned} & \widehat{\psi}_1(\mathbf{u}, \mathbf{x}) - \mathbb{E}\widehat{\psi}_1(\mathbf{u}, \mathbf{x}) \\ &= \left[\widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}) - \mathbb{E}\widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}) \right] + \left[\widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}) - \mathbb{E}\widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}) \right]. \end{aligned} \quad (8.6)$$

To begin, it is plain to see that

$$\begin{aligned} & P_{|\mathcal{S}} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}) - \theta(\mathbf{i}) \right| > \alpha_n \right) \\ &= P_{|\mathcal{S}} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}) - \theta(\mathbf{i}) \right| > \alpha_n \right) \\ & \cap \left\{ \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \bigcup_{i=1}^n |W_{s_i, A_n}| > \tau_n \right\} \cup \left\{ \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \left\{ \bigcup_{i=1}^n |W_{s_i, A_n}| > \tau_n \right\}^c \right\} \end{aligned}$$

$$\begin{aligned}
 &\leq P_{\cdot|\mathcal{S}} \left\{ \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}, \varphi) - \theta(\mathbf{i}) \right| > \alpha_n \right. \\
 &\quad \left. \cap \left\{ \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \bigcup_{i=1}^n |W_{s_i, A_n}| > \tau_n \right\} \right\} \\
 &+ P_{\cdot|\mathcal{S}} \left\{ \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \widehat{\psi}_2^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) - \theta(\mathbf{i}) \right| > \alpha_n \right. \\
 &\quad \left. \cap \left\{ \left\{ \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \bigcup_{i=1}^n |W_{s_i, A_n}| > \tau_n \right\}^c \right\} \right\} \\
 &\leq P_{\cdot|\mathcal{S}} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |W_{s_i, A_n}| > \tau_n \text{ for some } i = 1, \dots, n \right) + P_{\cdot|\mathcal{S}}(\emptyset) \\
 &\leq \tau_n^{-\zeta} \sum_{i=1}^n E_{\cdot|\mathcal{S}} \left[\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |W_{s_i, A_n}|^\zeta \right] \leq n \tau_n^{-\zeta} = \rho_n^{-\zeta} \rightarrow 0.
 \end{aligned}$$

We infer that

$$E_{\cdot|\mathcal{S}} \left[\left| \widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}) \right| \right] \leq \frac{1}{n} \sum_{i=1}^n \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_i \xi_{i_\ell} \cdots \xi_{i_{m-1}} E_{\cdot|\mathcal{S}} \left(\left| \widetilde{H}_2^{(\ell)}(z) \right| \right),$$

where

$$\begin{aligned}
 &E_{\cdot|\mathcal{S}} \left(\left| \widetilde{H}_2^{(\ell)}(z) \right| \right) \\
 &= E_{\cdot|\mathcal{S}} \left[\left| \frac{1}{\phi(h)} K_2 \left(\frac{d(x_i, X_{s_i, A_n})}{h} \right) W_{s_i, A_n} \times \int W_{s_{(1, \dots, \ell-1, \ell, \dots, m-1)}, A_n} \prod_{\substack{j=1 \\ j \neq i}}^{m-1} \frac{1}{\phi(h)} \right. \right. \\
 &\quad \left. \left. K_2 \left(\frac{d(x_j, \nu_{s_j, A_n})}{h} \right) \mathbb{P}(d\nu_1, \dots, d\nu_{\ell-1}, d\nu_\ell, \dots, d\nu_{m-1}) \right| \right] \\
 &\leq E_{\cdot|\mathcal{S}} \left[\left| \frac{1}{\phi(h)} \left(K_2 \left(\frac{d(x_i, X_{s_i, A_n})}{h} \right) + K_2 \left(\frac{d(x_i, X_{u_i}(s_i))}{h} \right) \right. \right. \right. \\
 &\quad \left. \left. - K_2 \left(\frac{d(x_i, X_{u_i}(s_i))}{h} \right) \right) W_{s_i, A_n} \mathbf{1}_{\{|W_{s_i, A_n}| > \tau_n\}} \right| \right] \\
 &\leq \frac{\tau_n^{-(\zeta-1)}}{\phi(h)} E_{\cdot|\mathcal{S}} \left[\left| K_2 \left(\frac{d(x_i, X_{s_i, A_n})}{h} \right) - K_2 \left(\frac{d(x_i, X_{u_i}(s_i))}{h} \right) \right| |W_{s_i, A_n}|^\zeta \right. \\
 &\quad \left. + \left| K_2 \left(\frac{d(x_i, X_{u_i}(s_i))}{h} \right) \right| |W_{s_i, A_n}|^\zeta \right] \\
 &\leq \frac{\tau_n^{-(\zeta-1)}}{\phi(h)} E_{\cdot|\mathcal{S}} \left[h^{-1} |d(x_i, X_{s_i, A_n}) - d(x_i, X_{u_i}(s_i))| |W_{s_i, A_n}|^\zeta \right. \\
 &\quad \left. + E_{\cdot|\mathcal{S}} \left[\left| K_2 \left(\frac{d(x_i, X_{u_i}(s_i))}{h} \right) \right| |W_{s_i, A_n}|^\zeta \right] \right]
 \end{aligned}$$

$$\begin{aligned}
&\lesssim \frac{\tau_n^{-(\zeta-1)}}{\phi(h)} \times \left[\frac{1}{nh} + \phi(h) \right] \\
&\lesssim \frac{\tau_n^{-(\zeta-1)}}{nh\phi(h)} + \tau_n^{-(\zeta-1)}. \tag{8.7}
\end{aligned}$$

Hence we have

$$\begin{aligned}
E_{\cdot|\mathbf{S}} \left[\left| \widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}) \right| \right] &\lesssim \frac{1}{n} \sum_{i=1}^n \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_i \xi_{i_\ell} \cdots \xi_{i_{m-1}} \left[\tau_n^{-(\zeta-1)} \right] \\
&\lesssim \tau_n^{-(\zeta-1)} \frac{1}{n^m} \sum_{\mathbf{i} \in I_n^m} \prod_{j=1}^m \frac{1}{h^d} \overline{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \\
&= \frac{C}{\tau_n^{(\zeta-1)}} \left(f_{\mathbf{S}}(\mathbf{u}) + O \left(\sqrt{\frac{\log n}{nh^{md}}} + h^2 \right) \right) \text{ (Using Lemma 8.2)} \\
&\leq \frac{C}{\tau_n^{(\zeta-1)}} = C \rho_n^{-(\zeta-1)} n^{-(\zeta-1)/\zeta} \leq C \alpha_n \quad P_{\mathbf{S}} - a.s.
\end{aligned}$$

As a result, we obtain that

$$\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}) - E_{\cdot|\mathbf{S}} \widehat{\psi}_1^{(2)}(\mathbf{u}, \mathbf{x}) \right| = O_{P_{\cdot|\mathbf{S}}}(\alpha_n). \tag{8.8}$$

Second, let us treat

$$\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) - \mathbb{E} \widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) \right|.$$

Recall the large blocks and small blocks and the notation given in Section 8.1, and define

$$\begin{aligned}
S_{\mathbf{s}, A_n}(\mathbf{u}, \mathbf{x}) &:= \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_i \xi_{i_\ell} \cdots \xi_{i_{m-1}} \widetilde{H}_1^{(\ell)}(z), \\
S_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}) &= \sum_{i: \mathbf{s}_i \in \Gamma_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}) \cap R_n} S_{\mathbf{s}, A_n}(\mathbf{u}, \mathbf{x}) = \left(S_n^{(1)}(\boldsymbol{\ell}; \boldsymbol{\varepsilon}), \dots, S_n^{(p)}(\boldsymbol{\ell}; \boldsymbol{\varepsilon}) \right)'.
\end{aligned}$$

Then we have

$$\begin{aligned}
S_n &= \left(S_n^{(1)}, \dots, S_n^{(m)} \right)' \\
&= \sum_{i=1}^n S_{\mathbf{s}, A_n}(\mathbf{u}, \mathbf{x}) \\
&= \sum_{\boldsymbol{\ell} \in L_n} S_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}_0) + \underbrace{\sum_{\boldsymbol{\varepsilon} \neq \boldsymbol{\varepsilon}_0} \sum_{\boldsymbol{\ell} \in L_{1,n}} S_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon})}_{=: S_{2,n}(\boldsymbol{\varepsilon})} + \underbrace{\sum_{\boldsymbol{\varepsilon} \neq \boldsymbol{\varepsilon}_0} \sum_{\boldsymbol{\ell} \in L_{2,n}} S_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon})}_{=: S_{3,n}(\boldsymbol{\varepsilon})} \\
&=: S_{1,n} + \sum_{\boldsymbol{\varepsilon} \neq \boldsymbol{\varepsilon}_0} S_{2,n}(\boldsymbol{\varepsilon}) + \sum_{\boldsymbol{\varepsilon} \neq \boldsymbol{\varepsilon}_0} S_{3,n}(\boldsymbol{\varepsilon}). \tag{8.9}
\end{aligned}$$

In order to achieve our result, we will pass by the following two steps.

Step 1 (Reduction to independence). Recall

$$S_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}) = \sum_{i: s_i \in \Gamma_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}) \cap R_n} S_{\mathbf{s}, A_n}(\mathbf{u}, \mathbf{x}).$$

For each $\boldsymbol{\varepsilon} \in \{1, 2\}^d$, let $\{\check{S}_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}) : \boldsymbol{\ell} \in L_n\}$ be a sequence of independent random variable in \mathbb{R} under $P_{\cdot|\mathbf{S}}$ such that

$$\check{S}_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}) \stackrel{d}{=} S_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}), \text{ under } P_{\cdot|\mathbf{S}}, \boldsymbol{\ell} \in L_n.$$

Define

$$\check{S}_{1,n} = \sum_{\boldsymbol{\ell} \in L_n} \check{S}_n(\boldsymbol{\ell}; \varepsilon_0) = (\check{S}_{1,n}^{(1)}, \dots, \check{S}_{1,n}^{(m)})'$$

and for $\boldsymbol{\varepsilon} \neq \varepsilon_0$, define

$$\check{S}_{2,n}(\boldsymbol{\varepsilon}) = \sum_{\boldsymbol{\ell} \in L_{1,n}} \check{S}_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon})$$

and

$$\check{S}_{3,n}(\boldsymbol{\varepsilon}) = \sum_{\boldsymbol{\ell} \in L_{2,n}} \check{S}_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}).$$

We start by confirming the following results:

$$\sup_{t>0} |P_{\cdot|\mathbf{S}}(S_{1,n} > t) - P_{\cdot|\mathbf{S}}(\check{S}_{1,n} > t)| \leq C \left(\frac{A_n}{A_{1,n}} \right)^d \beta(A_{2,n}; A_n^d) \quad (8.10)$$

$$\sup_{t>0} |P_{\cdot|\mathbf{S}}(\|S_{2,n}(\boldsymbol{\varepsilon})\|_\infty > t) - P_{\cdot|\mathbf{S}}(\|\check{S}_{2,n}(\boldsymbol{\varepsilon})\|_\infty > t)| \leq C \left(\frac{A_n}{A_{1,n}} \right)^d \beta(A_{2,n}; A_n^d) \quad (8.11)$$

$$\sup_{t>0} |P_{\cdot|\mathbf{S}}(\|S_{3,n}(\boldsymbol{\varepsilon})\|_\infty > t) - P_{\cdot|\mathbf{S}}(\|\check{S}_{3,n}(\boldsymbol{\varepsilon})\|_\infty > t)| \leq C \left(\frac{A_n}{A_{1,n}} \right)^d \beta(A_{2,n}; A_n^d) \quad (8.12)$$

Keep in mind that

$$\llbracket L_n \rrbracket = O\left((A_n/A_{3,n})^d\right) \lesssim (A_n/A_{1,n})^d.$$

For $\boldsymbol{\varepsilon} \in \{1, 2\}^d$ and $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2 \in L_n$ with $\boldsymbol{\ell}_1 \neq \boldsymbol{\ell}_2$, let

$$\mathcal{J}_1(\boldsymbol{\varepsilon}) = \{1 \leq i_1 \leq n : \mathbf{s}_{i_1} \in \Gamma_n(\boldsymbol{\ell}_1; \boldsymbol{\varepsilon})\}, \mathcal{J}_2(\boldsymbol{\varepsilon}) = \{1 \leq i_2 \leq n : \mathbf{s}_{i_2} \in \Gamma_n(\boldsymbol{\ell}_2; \boldsymbol{\varepsilon})\}.$$

For any $\mathbf{s}_{i_k} = (s_{1,i_k}, \dots, s_{d,i_k})$, $k = 1, 2$ in such a way that $i_1 \in \mathcal{J}_1(\boldsymbol{\varepsilon})$ and $i_2 \in \mathcal{J}_2(\boldsymbol{\varepsilon})$, we get $\max_{1 \leq u \leq d} |s_{u,i_1} - s_{u,i_2}| \geq A_{2,n}$ using the definition of $\Gamma(\boldsymbol{\ell}; \boldsymbol{\varepsilon})$. This gives

$$|\mathbf{s}_{i_1} - \mathbf{s}_{i_2}| \geq A_{2,n}.$$

For any $\boldsymbol{\varepsilon} \in \{1, 2\}^d$, let $S_n(\boldsymbol{\ell}_1; \boldsymbol{\varepsilon}), \dots, S_n(\boldsymbol{\ell}_{\llbracket L_n \rrbracket}; \boldsymbol{\varepsilon})$ be an arrangement of $\{S_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}) : \boldsymbol{\ell} \in L_n\}$.

Let $P_{\cdot|\mathbf{S}}^{(a)}$ be the marginal distribution of $S_n(\boldsymbol{\ell}_a; \boldsymbol{\varepsilon})$ and let $P_{\cdot|\mathbf{S}}^{(a:b)}$ be the joint distribution of $\{S_n(\boldsymbol{\ell}_k; \boldsymbol{\varepsilon}) : a \leq k \leq b\}$. The β -mixing property of \mathbf{X} gives that for $1 \leq k \leq \llbracket L_n \rrbracket - 1$,

$$\|P_{\cdot|\mathbf{S}} - P_{\cdot|\mathbf{S}}^{(1:k)} \times P_{\cdot|\mathbf{S}}^{(k+1:\llbracket L_n \rrbracket)}\|_{\text{TV}} \lesssim \beta(A_{2,n}; A_n^d).$$

The inequality is independent of the arrangement of $\{S_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}) : \boldsymbol{\ell} \in L_n\}$. Therefore, the assumption 8.66 in Lemma 8.5 is fulfilled for $\{S_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}) : \boldsymbol{\ell} \in L_n\}$ with $\tau \sim \beta(A_{2,n}; A_n^d)$ and $m \lesssim (A_n/A_{1,n})^d$. Combining the boundary condition on R_n and Lemma 8.5, we get (8.10)- (8.12).

Remark 8.1. *Since*

$$\llbracket \{\boldsymbol{\varepsilon} \in \{1, 2\}^d : \boldsymbol{\varepsilon} \neq \boldsymbol{\varepsilon}_0\} \rrbracket = 2^d - 1, \llbracket L_{1,n} \rrbracket \sim (A_n/A_{3,n})^d \sim (A_n/A_{1,n})^d$$

and

$$\llbracket L_{2,n} \rrbracket \sim (A_n/A_{3,n})^{d-1} \sim (A_n/A_{1,n})^{d-1} \ll \llbracket L_{1,n} \rrbracket,$$

Lemma 8.6 and Equation (8.3) give for n sufficiently large, the summands numbers of $S_{2,n}$ and $S_{3,n}$ are at most

$$O\left(A_{1,n}^{d-1} A_{2,n} n A_n^{-d} (A_n/A_{1,n})^d\right) = O\left(\frac{A_{2,n}}{A_{1,n}} n\right)$$

and

$$O\left(A_{1,n}^{d-1} A_{2,n} n A_n^{-d} (A_n/A_{1,n})^{d-1}\right) = O\left(\frac{A_{2,n}}{A_n} n\right),$$

respectively.

Step 2: Recall that we aim to treat

$$\sup_{\mathcal{F}_m} \sup_{\mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) - E_{\cdot|\mathbf{S}} \widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) \right|.$$

To achieve the aimed result, we will cover the region $B^m = [0, 1]^{dm}$ by

$$\bigcup_{k_1, \dots, k_m=1}^{N(\mathbf{u})} \prod_{j=1}^m \mathbf{B}(\mathbf{u}_{k_j}, r),$$

for some radius r . Hence, for each $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_m) \in [0, 1]^{dm}$, there exists $\mathbf{l}(\mathbf{u}) = (l(\mathbf{u}_1), \dots, l(\mathbf{u}_m))$, where $\forall 1 \leq i \leq m, 1 \leq l(\mathbf{u}_i) \leq N(\mathbf{u})$ in such a way that

$$\mathbf{u} \in \prod_{i=1}^m \mathbf{B}(\mathbf{u}_{l(\mathbf{u}_i)}, r) \quad \text{and} \quad |\mathbf{u}_i - \mathbf{u}_{l(\mathbf{u}_i)}| \leq r, \quad \text{for } 1 \leq i \leq m,$$

then for each $\mathbf{u} \in [0, 1]^{dm}$, the closest center will be $\mathbf{u}_1(\mathbf{u})$, and the ball with the closest

center will be defined by

$$\mathcal{B}(\mathbf{u}, \mathbf{l}(\mathbf{u}), r) := \prod_{j=1}^m \mathbf{B}(\mathbf{u}_{k_j}, r).$$

In the same way \mathcal{H}^m should be covered by

$$\bigcup_{k_1, \dots, k_m=1}^{N(x)} \prod_{j=1}^m \mathbf{B}(x_{k_j}, r),$$

for some radius r . Hence, for each $\mathbf{x} = (x_1, \dots, x_m) \in \mathcal{H}^m$, there exists $\mathbf{l}(\mathbf{x}) = (l(x_1), \dots, l(x_m))$, where $\forall 1 \leq i \leq m, 1 \leq l(x_i) \leq N(x)$ in such a way that

$$\mathbf{x} \in \prod_{i=1}^m B(u_{l(x_i)}, r) \text{ and } d(x_i, x_{l(x_i)}) \leq r, \text{ for } 1 \leq i \leq m,$$

then for each $\mathbf{x} \in \mathcal{H}^m$, the closest center will be $\mathbf{x}_1(\mathbf{x})$, and the ball with the closest centre will be defined by

$$\mathbf{B}(\mathbf{x}, \mathbf{l}(\mathbf{x}), r) := \prod_{i=1}^m B(x_{l(x_i)}, r).$$

We define:

$$K^*(\boldsymbol{\omega}, \mathbf{v}) \geq C_0 \prod_{j=1}^m \prod_{\ell=1}^d \mathbf{1}_{(|\omega_{j,\ell}| \leq 2C_1)} \prod_{j=1}^m K_2(v_k) \text{ for } (\boldsymbol{\omega}, \mathbf{v}) \in \mathbb{R}^2.$$

We can show that, for $(u, x) \in B_{j,n}$ and n large enough,

$$\begin{aligned} & \left| \overline{K} \left(\frac{\mathbf{u} - \mathbf{s}/A_n}{h_n} \right) K_2 \left(\frac{d(x_i, X_{\mathbf{s}_i, A_n})}{h} \right) - \overline{K} \left(\frac{\mathbf{u}_n - \mathbf{s}/A_n}{h_n} \right) K_2 \left(\frac{d(x_{n,i}, X_{\mathbf{s}_i, A_n})}{h} \right) \right| \\ & \leq \alpha_n K^* \left(\frac{\mathbf{u}_n - \mathbf{s}/A_n, d(x_{n,i}, X_{\mathbf{s}_i, A_n})}{h_n} \right). \end{aligned}$$

Then, for

$$\begin{aligned} \widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}) &= \frac{1}{n} \sum_{i=1}^n \xi_i \frac{1}{\phi(h)} K_2 \left(\frac{d(x_i, X_{\mathbf{s}_i, A_n})}{h} \right) W_{\mathbf{s}_i, A_n} \mathbf{1}_{\{|W_{\mathbf{s}_i, A_n}| \leq \tau_n\}} \\ & \times \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_{i_\ell} \cdots \xi_{i_{m-1}} \\ & \times \int W_{\mathbf{s}_{(1, \dots, \ell-1, \ell, \dots, m-1)}, A_n} \prod_{\substack{j=1 \\ j \neq i}}^{m-1} \frac{1}{\phi(h)} K_2 \left(\frac{d(x_j, \nu_{\mathbf{s}_j, A_n})}{h} \right) \mathbb{P}(d\nu_1, \dots, d\nu_{\ell-1}, d\nu_\ell, \dots, d\nu_{m-1}). \end{aligned}$$

Let us define

$$\begin{aligned}
 \bar{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}) &= \frac{1}{nh^d \phi(h)} \sum_{i=1}^n K^* \left(\frac{\mathbf{u}_n - \mathbf{s}_i / A_n, d(x_{n,i}, X_{\mathbf{s}_i, A_n})}{h_n} \right) W_{\mathbf{s}_i, A_n} \mathbf{1}_{\{|W_{\mathbf{s}_i, A_n}| \leq \tau_n\}} \\
 &\quad \times \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_{i_\ell} \cdots \xi_{i_{m-1}} \\
 &\quad \times \int W_{\mathbf{s}_{(1, \dots, \ell-1, \ell, \dots, m-1)}, A_n} \prod_{\substack{j=1 \\ j \neq i}}^{m-1} \frac{1}{\phi(h)} K_2 \left(\frac{d(x_j, \nu_{\mathbf{s}_j, A_n})}{h} \right) \mathbb{P}(d\nu_1, \dots, d\nu_{\ell-1}, d\nu_\ell, \dots, d\nu_{m-1}) \\
 &:= \frac{1}{nh^d \phi(h)} \sum_{i=1}^n S'_{\mathbf{s}, A_n}(\mathbf{u}, \mathbf{x}). \tag{8.13}
 \end{aligned}$$

We mention that

$$E_{\cdot|\mathbf{S}} \left[\left| \bar{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) \right| \right] \leq M < \infty,$$

for some M large enough. Let $N_{\mathcal{F}_m \mathcal{H}^m} N_{(x)}^m N_{(u)}^m$ denotes the covering number related respectively to the class of functions $\mathcal{F}_m \mathcal{H}^m$, the balls that cover $[0, 1]^m$ and the balls that cover \mathcal{H}^m . Then we obtain

$$\begin{aligned}
 &\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in \mathcal{B}} \left| \widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) - E_{\cdot|\mathbf{S}} \left[\widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) \right] \right| \\
 &\leq N_{\mathcal{F}_m \mathcal{H}^m} N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i_1(x)}, r)} \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i_1(u)}, r)} \left| \widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) - E_{\cdot|\mathbf{S}} \left[\widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}, \varphi) \right] \right| \\
 &\leq N_{\mathcal{F}_m \mathcal{H}^m} N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i_1(x)}, r)} \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i_1(u)}, r)} \left| \widehat{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}) - E_{\cdot|\mathbf{S}} \left[\widehat{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}) \right] \right| \\
 &\quad + N_{\mathcal{F}_m \mathcal{H}^m} N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i_1(x)}, r)} \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i_1(u)}, r)} \alpha_n \left(\left| \bar{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}) \right| + E_{\cdot|\mathbf{S}} \left[\left| \bar{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}) \right| \right] \right) \\
 &\leq N_{\mathcal{F}_m \mathcal{H}^m} N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i_1(x)}, r)} \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i_1(u)}, r)} \left| \widehat{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}) - E_{\cdot|\mathbf{S}} \left[\widehat{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}) \right] \right| \\
 &\quad + N_{\mathcal{F}_m \mathcal{H}^m} N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i_1(x)}, r)} \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i_1(u)}, r)} \left| \bar{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}) - E_{\cdot|\mathbf{S}} \left[\bar{\psi}_1^{(1)}(\mathbf{u}_n, \mathbf{x}) \right] \right| \\
 &\quad + 2MF(\mathbf{y})\alpha_n \\
 &\leq N_{\mathcal{F}_m \mathcal{H}^m} N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i_1(x)}, r)} \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i_1(u)}, r)} \left| \sum_{\ell \in L_n} S_n(\ell; \varepsilon_0) \right| \\
 &\quad + N_{\mathcal{F}_m \mathcal{H}^m} N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i_1(x)}, r)} \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i_1(u)}, r)} \left| \sum_{\varepsilon \neq \varepsilon_0} \sum_{\ell \in L_{1,n}} S_n(\ell; \varepsilon) \right| \\
 &\quad + N_{\mathcal{F}_m \mathcal{H}^m} N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i_1(x)}, r)} \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i_1(u)}, r)} \left| \sum_{\varepsilon \neq \varepsilon_0} \sum_{\ell \in L_{2,n}} S_n(\ell; \varepsilon) \right| \\
 &\quad + N_{\mathcal{F}_m \mathcal{H}^m} N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i_1(x)}, r)} \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i_1(u)}, r)} \left| \sum_{\ell \in L_n} S'_n(\ell; \varepsilon_0) \right| \\
 &\quad + N_{\mathcal{F}_m \mathcal{H}^m} N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i_1(x)}, r)} \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i_1(u)}, r)} \left| \sum_{\varepsilon \neq \varepsilon_0} \sum_{\ell \in L_{1,n}} S'_n(\ell; \varepsilon) \right|
 \end{aligned}$$

$$\begin{aligned}
 & + N_{\mathcal{F}_m} \mathcal{K}^m N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i(x),r})} \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i(u),r})} \left| \sum_{\varepsilon \neq \varepsilon_0} \sum_{\ell \in L_{2,n}} S'_n(\ell; \varepsilon) \right| + 2MF(\mathbf{y})\alpha_n.
 \end{aligned} \tag{8.14}$$

Even more, for each $\varepsilon \in \{1, 2\}^d$, let $\{\check{S}'_n(\ell; \varepsilon) : \ell \in L_n\}$ denote a sequence of independent random vectors in \mathbb{R}^m under $P_{\cdot|\mathcal{S}}$ such that

$$\check{S}'_n(\ell; \varepsilon) \stackrel{d}{=} S'_n(\ell; \varepsilon), \text{ under } P_{\cdot|\mathcal{S}}, \ell \in L_n.$$

Show that

$$\begin{aligned}
 & P_{\cdot|\mathcal{S}} \left(N_{\mathcal{F}_m} \mathcal{K}^m N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i(x),r})} \right. \\
 & \quad \left. \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i(u),r})} \left| \widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}) - E_{\cdot|\mathcal{S}} \left[\widehat{\psi}_1^{(1)}(\mathbf{u}, \mathbf{x}) \right] \right| > 2^{md+1} M a_n \right) \\
 & \leq N_{\mathcal{F}_m} \mathcal{K}^m N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i(x),r})} \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i(u),r})} \\
 & \quad P_{\cdot|\mathcal{S}} \left(\sup_{(\mathbf{u}, \mathbf{x}) \in B_k} \left| \widehat{\psi}_1(\mathbf{u}, \mathbf{x}) - E_{\cdot|\mathcal{S}} \left[\widehat{\psi}_1(\mathbf{u}, \mathbf{x}) \right] \right| > 2^{md+1} M a_n \right) \\
 & \leq \sum_{\varepsilon \in \{1,2\}^d} \widehat{Q}_n(\varepsilon) + \sum_{\varepsilon \in \{1,2\}^d} \bar{Q}_n(\varepsilon) + 2^{md+1} N_{\mathcal{F}_m} \mathcal{K}^m N_{(x)}^m N_{(u)}^m \left(\frac{A_n}{A_{1,n}} \right)^d \beta(A_{2,n}; A_n^d), \tag{8.16}
 \end{aligned}$$

where

$$\begin{aligned}
 \widehat{Q}_n(\varepsilon_0) & = N_{\mathcal{F}_m} \mathcal{K}^m N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i(x),r})} \\
 & \quad \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i(u),r})} P_{\cdot|\mathcal{S}} \left(\left| \sum_{\ell \in L_n} \check{S}_n(\ell; \varepsilon_0) \right| > M a_n n^m h^{md} \phi(h) \right),
 \end{aligned}$$

$$\begin{aligned}
 \bar{Q}_n(\varepsilon_0) & = N_{\mathcal{F}_m} \mathcal{K}^m N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i(x),r})} \\
 & \quad \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i(u),r})} P_{\cdot|\mathcal{S}} \left(\left| \sum_{\ell \in L_n} \check{S}'_n(\ell; \varepsilon_0) \right| > M a_n n^m h^{md} \phi(h) \right),
 \end{aligned}$$

and for $\varepsilon \neq \varepsilon_0$

$$\begin{aligned}
 \widehat{Q}_n(\varepsilon) & = N_{\mathcal{F}_m} \mathcal{K}^m N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i(x),r})} \\
 & \quad \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i(u),r})} P_{\cdot|\mathcal{S}} \left(\left| \sum_{\ell \in L_n} \check{S}_n(\ell; \varepsilon) \right| > M a_n n^m h^{md} \phi(h) \right),
 \end{aligned}$$

$$\bar{Q}_n(\boldsymbol{\varepsilon}) = N_{\mathcal{F}_m} \mathcal{N}^m N_{(x)}^m N_{(u)}^m \max_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(x_{i_1(x)}, r)} \sup_{1 \leq i_1 < \dots < i_m \leq m} \sup_{B(\mathbf{u}_{i_1(\mathbf{u})}, r)} P_{|\mathcal{S}} \left(\left| \sum_{\ell \in L_n} \check{S}'_n(\ell; \boldsymbol{\varepsilon}) \right| > M a_n n^m h^{md} \phi(h) \right).$$

Due to the similarity between the two case: $\boldsymbol{\varepsilon} \neq \boldsymbol{\varepsilon}_0$ and $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}_0$, we are going to treat \hat{Q}_n only for $\boldsymbol{\varepsilon} \neq \boldsymbol{\varepsilon}_0$. An application of Lemma 8.6, with the fact that $\check{S}_n(\ell; \boldsymbol{\varepsilon})$ are zero-mean random variables, shows us that:

$$P_{|\mathcal{S}} \left(\left| \sum_{\ell \in L_n} \check{S}_n(\ell; \boldsymbol{\varepsilon}) \right| > M a_n n h^{md} \phi(h) \right) \leq 2 P_{|\mathcal{S}} \left(\sum_{\ell \in L_n} \check{S}_n(\ell; \boldsymbol{\varepsilon}) > M a_n n h^{md} \phi(h) \right)$$

and

$$\begin{aligned} |\check{S}_n(\ell; \boldsymbol{\varepsilon})| &\leq C A_{1,n}^{d-1} A_{2,n} (\log n) \tau_n, P_{\mathcal{S}} - \text{ a.s.} \quad (\text{from Lemma 8.6}) \\ E_{|\mathcal{S}} \left[\left(\check{S}_n(\ell; \boldsymbol{\varepsilon}) \right)^2 \right] &\leq C h^{md} \phi(h) A_{1,n}^{d-1} A_{2,n} (\log n), P_{\mathcal{S}} - \text{ a.s.} \quad (\text{By Lemma 8.7}) \end{aligned} \quad (8.17)$$

Using Bernstein's inequality represented in Lemma 8.8, we have

$$\begin{aligned} &P_{|\mathcal{S}} \left(\sum_{\ell \in L_n} \check{S}_n(\ell; \boldsymbol{\varepsilon}) > M a_n n h^{md} \phi(h) \right) \\ &\leq \exp \left(- \frac{\frac{1}{2} \times M n h^{md} \phi(h) \log n}{\left(\frac{A_n}{A_{1,n}} \right)^d A_{1,n}^{d-1} A_{2,n} h^{md} \phi(h) (\log n) + \frac{1}{3} \times M^{1/2} n^{1/2} h^{md/2} \phi(h)^{1/2} (\log n)^{1/2} A_{1,n}^{d-1} A_{2,n} \tau_n} \right). \end{aligned}$$

Observe that

$$\begin{aligned} \frac{n h^{md} \log n}{\left(\frac{A_n}{A_{1,n}} \right)^d A_{1,n}^{d-1} A_{2,n} h^{md} (\log n)} &= n A_n^{-d} \left(\frac{A_{1,n}}{A_{2,n}} \right) \gtrsim \frac{A_{1,n}}{A_{2,n}} \gtrsim n^\eta, \quad (8.18) \\ \frac{n h^{md} \phi(h) \log n}{n^{1/2} h^{md/2} \phi(h)^{1/2} (\log n)^{1/2} A_{1,n}^{d-1} A_{2,n} \tau_n} &= \frac{n^{1/2} h^{md/2} \phi(h)^{1/2} (\log n)^{1/2}}{A_{1,n}^{md} \left(\frac{A_{2,n}}{A_{1,n}} \right) \rho_n n^{1/\zeta}} \geq C_0 n^{\eta/2} \quad (8.19) \end{aligned}$$

Taking $M > 0$ sufficiently large, and for $N \leq C h^{-md} \phi(h) \alpha_n^{-m}$, this shows the desired result.

We must move on to the nonlinear part of the Hoeffding decomposition. Accordingly, the goal is to prove that:

$$P_{|\mathcal{S}} \left[\sup_{\mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |\hat{\psi}_2(\mathbf{u}, \mathbf{x})| > \lambda \right] \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (8.20)$$

In the following, we will give a lemma that can be viewed as a technical result in the proof of our proposition, and it helps us to achieve our goal in Expression (8.20). The proof of this lemma used the blocking technique defined before but for the U -statistic, making the block treatment more complicated.

Lemma 8.1. *Let $\mathcal{F}_m \mathcal{K}^m$ be a uniformly bounded class of measurable canonical functions, $m \geq 2$. Assume that there are finite constants \mathbf{a} and \mathbf{b} in such a way that the $\mathcal{F}_m \mathcal{K}^m$ covering number fulfils :*

$$N(\varepsilon, \mathcal{F}_m \mathcal{K}^m, \|\cdot\|_{L_2(Q)}) \leq \mathbf{a}\varepsilon^{-\mathbf{b}}, \quad (8.21)$$

for all $\varepsilon > 0$ and all probability measure Q . If the mixing coefficients β of the local stationary sequence $\{Z_i = (X_{\mathbf{s}_i, A_n}, W_{\mathbf{s}_i, A_n})_{i \in \mathbb{N}^*}$ satisfy Condition (E2) in Assumption 12, then, for some $r > 1$, we have:

$$\sup_{\mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \mathbb{P} \left[h^{md/2} \phi^{m/2}(h) n^{-m+1/2} \sum_{\mathbf{i} \in I_n^m} \xi_{i_1} \cdots \xi_{i_m} H(Z_{i_1}, \dots, Z_{i_m}) \right] \rightarrow 0. \quad (8.22)$$

Remark 8.2. *As mentioned before, $W_{\mathbf{s}_i, A_n}$ will be equal to 1 or $\varepsilon_{\mathbf{s}_i, A_n} = \sigma\left(\frac{\mathbf{s}_i}{A_n}, X_{\mathbf{s}_i, A_n}\right) \varepsilon_{\mathbf{s}_i}$. In the proof of the previous Lemma, $W_{\mathbf{s}_i, A_n}$ will be equal $\varepsilon_{\mathbf{i}, n} = \sigma\left(\frac{\mathbf{i}}{n}, X_{\mathbf{i}, n}\right) \varepsilon_{\mathbf{i}}$, and we will use the notation $W_{\mathbf{s}_i, A_n}^{(u)}$ to indicate $\sigma(\mathbf{u}, \mathbf{x}) \varepsilon_{\mathbf{i}}$*

Proof of Lemma 8.1

This lemma's proof is based on the blocking technique employed by Arcones and Yu, 1994, and called *Bernstein's method*, referred to Bernstein, 1927, in which we are enabled to apply the symmetrization and the many other techniques available for the i.i.d random variables. We will extend this technique to the spacial processes in the U -statistics setting, in the same line as in Lahiri, 2003a. In addition to the notation in Section 8.1, define

$$L_n := L_{1,n} \cup L_{2,n},$$

$$\Delta_1 = \{\ell_2 : \min_{1 \leq i \leq d} |\ell_{1i} - \ell_{2i}| \leq 1\}$$

$$\Delta_2 = \{\ell_2 : \min_{1 \leq i \leq d} |\ell_{1i} - \ell_{2i}| \geq 2\}$$

With the notation introduced above, it is easy to show that, for $m = 2$,

$$\frac{1}{h^{2d} \phi^2(h)} \sum_{\mathbf{i} \in I_n^2} \prod_{j=1}^2 \left\{ \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j}, A_n})}{h_n} \right) \right\} W_{\mathbf{s}_i, A_n}$$

$$\begin{aligned}
&= \frac{1}{h^{2d}\phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \\
&\quad \times \prod_{j=1}^2 \left\{ \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j}, A_n})}{h_n} \right) \right\} W_{\mathbf{s}_i, A_n} \\
&+ \frac{1}{h^{2d}\phi^2(h)} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{i_2: s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n \\ i_1 \neq i_2}} \\
&\quad \times \prod_{j=1}^2 \left\{ \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j}, A_n})}{h_n} \right) \right\} W_{\mathbf{s}_i, A_n} \\
&+ 2 \frac{1}{h^{2d}\phi^2(h)} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\Delta_2} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon) \cap R_n} \\
&\quad \times \prod_{j=1}^2 \left\{ \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j}, A_n})}{h_n} \right) \right\} W_{\mathbf{s}_i, A_n} \\
&+ 2 \frac{1}{h^{2d}\phi^2(h)} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\Delta_1} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon) \cap R_n} \\
&\quad \times \prod_{j=1}^2 \left\{ \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j}, A_n})}{h_n} \right) \right\} W_{\mathbf{s}_i, A_n} \\
&+ \frac{1}{h^{2d}\phi^2(h)} \sum_{\substack{\ell_1 \neq \ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon) \cap R_n} \\
&\quad \times \prod_{j=1}^2 \left\{ \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j}, A_n})}{h_n} \right) \right\} W_{\mathbf{s}_i, A_n} \\
&+ \frac{1}{h^{2d}\phi^2(h)} \sum_{\substack{\ell_1 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_1 < i_2 : s_{i_1}, s_{i_2} \in \Gamma_n(\ell_1; \varepsilon) \cap R_n} \\
&\quad \times \prod_{j=1}^2 \left\{ \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j}, A_n})}{h_n} \right) \right\} W_{\mathbf{s}_i, A_n} \\
&:= \text{I} + \text{II} + \text{III} + \text{IV} + \text{V} + \text{VI}. \tag{8.23}
\end{aligned}$$

(I): The same type of blocks but not the same block: Let $\{\eta_i\}_{i \in \mathbb{N}^*}$ be a sequence of independent blocks. An application of Lemma 8.5 shows that:

$$\begin{aligned}
&\mathbb{P} \left(\sup_{\mathcal{F}_m} \sup_{\mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| n^{-3/2} \frac{1}{h^{2d}\phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \right. \right. \\
&\quad \left. \left. \times \prod_{j=1}^2 \left\{ \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j}, A_n})}{h_n} \right) \right\} W_{\mathbf{s}_i, A_n} \right| > \delta \right)
\end{aligned}$$

$$\begin{aligned}
 &\leq \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \right. \right. \\
 &\quad \left. \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j} / A_n}{h_n} \right) \left[\prod_{j=1}^2 K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j}, A_n})}{h_n} \right) - \prod_{j=1}^2 K_2 \left(\frac{d(x_j, X_{\mathbf{u}_j}(\mathbf{s}_{i_j}))}{h} \right) \right] W_{\mathbf{s}_{i_1, A_n}} \right| > \delta \Big) \\
 &+ \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \right. \right. \\
 &\quad \left. \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j} / A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, X_{\mathbf{u}_j}(\mathbf{s}_{i_j}))}{h} \right) [W_{\mathbf{s}_{i_1, A_n}} - W_{\mathbf{s}_{i_1, A_n}}^{(u)}] \right| > \delta \Big) \\
 &+ \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \right. \right. \\
 &\quad \left. \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j} / A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, X_{\mathbf{u}_j}(\mathbf{s}_{i_j}))}{h} \right) W_{\mathbf{s}_{i_1, A_n}}^{(u)} \right| > \delta \Big) \\
 &\leq \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \right. \right. \\
 &\quad \left. \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j} / A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_{i_1, A_n}}^{(u)} \right| > \delta \Big) \\
 &\quad + C \left(\frac{A_n}{A_{1,n}} \right)^d \beta(A_{2,n}; A_n^d) + o_{\mathbb{P}}(1) + o_{\mathbb{P}}(1),
 \end{aligned}$$

Because:

$$\begin{aligned}
 &E_{\cdot | \mathbf{S}} \left| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j} / A_n}{h_n} \right) \right. \\
 &\quad \left. \left[\prod_{j=1}^2 K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j}, A_n})}{h_n} \right) - \prod_{j=1}^2 K_2 \left(\frac{d(x_j, X_{\mathbf{u}_j}(\mathbf{s}_{i_j}))}{h} \right) \right] W_{\mathbf{s}_{i_1, A_n}} \right| \\
 &= \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j} / A_n}{h_n} \right) \\
 &\quad E_{\cdot | \mathbf{S}} \left| \left[\prod_{j=1}^2 K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j}, A_n})}{h_n} \right) - \prod_{j=1}^2 K_2 \left(\frac{d(x_j, X_{\mathbf{u}_j}(\mathbf{s}_{i_j}))}{h} \right) \right] W_{\mathbf{s}_{i_1, A_n}} \right| \\
 &= \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j} / A_n}{h_n} \right) \\
 &\quad E_{\cdot | \mathbf{S}} \left| \left[\prod_{j=1}^2 K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j}, A_n})}{h_n} \right) - \prod_{j=1}^2 K_2 \left(\frac{d(x_j, X_{\mathbf{u}_j}(\mathbf{s}_{i_j}))}{h} \right) \right] \prod_{j=1}^m \varepsilon_{\mathbf{s}_{i_j}, A_n} \right|
 \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n^{3/2}h^{2d}\phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \\
&\quad E_{\cdot|\mathbf{S}} \left| \left[\prod_{j=1}^2 K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j}, A_n})}{h_n} \right) - \prod_{j=1}^2 K_2 \left(\frac{d(x_j, X_{u_j}(\mathbf{s}_{i_j}))}{h} \right) \right] \prod_{j=1}^m \sigma \left(\frac{\mathbf{s}_{i_j}}{A_n}, X_{s_{i_j}, A_n} \right) \varepsilon_{s_{i_j}} \right| \\
&= \frac{1}{n^{3/2}h^{2d}\phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^m E_{\cdot|\mathbf{S}}(\varepsilon_{s_{i_j}}) \\
&\quad E_{\cdot|\mathbf{S}} \left| \left[\prod_{j=1}^2 K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j}, A_n})}{h_n} \right) - \prod_{j=1}^2 K_2 \left(\frac{d(x_j, X_{u_j}(\mathbf{s}_{i_j}))}{h} \right) \right] \right. \\
&\quad \left. \left[\prod_{j=1}^m \sigma \left(\frac{\mathbf{s}_{i_j}}{A_n}, X_{s_{i_j}, A_n} \right) - \prod_{j=1}^m \sigma(\mathbf{u}_j, \mathbf{x}_j) + \prod_{j=1}^m \sigma(\mathbf{u}_j, \mathbf{x}_j) \right] \right| \\
&\leq \frac{1}{n^{3/2}h^{2d}\phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^m E_{\cdot|\mathbf{S}}(\varepsilon_{s_{i_j}}) \\
&\quad E_{\cdot|\mathbf{S}} \left[C \sum_{j=1}^m \left| K_2 \left(\frac{d(x_j, X_{s_{i_j}, A_n})}{h_n} \right) - K_2 \left(\frac{d(x_j, X_{\frac{\mathbf{s}_{i_j}}{A_n}}(\mathbf{s}_{i_j}))}{h} \right) \right|^p \right] \prod_{j=1}^m [\sigma(\mathbf{u}_j, \mathbf{x}_j) + o_{\mathbb{P}}(1)] \\
&\quad \text{(Using a telescoping argument, and the boundedness of } K_2 \text{ for } p = \min(\rho, 1) \text{ and } C < \infty) \\
&\lesssim \frac{1}{n^{3/2}h^{2d}\phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^m E_{\cdot|\mathbf{S}}(\varepsilon_{s_{i_j}}) \\
&\quad E_{\cdot|\mathbf{S}} \left[\phi^{m-1}(h) \left| \frac{C}{A_n^d} U_{s_{i_j}, A_n} \left(\frac{\mathbf{s}_{i_j}}{A_n} \right) \right|^p \right] \prod_{j=1}^m [\sigma(\mathbf{u}_j, \mathbf{x}_j) + o_{\mathbb{P}}(1)] \\
&\sim o_{\mathbb{P}}(1), \tag{8.24}
\end{aligned}$$

and

$$\begin{aligned}
&E_{\cdot|\mathbf{S}} \left| \frac{1}{n^{3/2}h^{2d}\phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \right. \\
&\quad \left. \prod_{j=1}^2 K_2 \left(\frac{d(x_i, X_{u_j}(\mathbf{s}_{i_j}))}{h} \right) [W_{s_{i_1}, A_n} - W_{s_{i_2}, A_n}^{(u)}] \right| \\
&= \frac{1}{n^{3/2}h^{2d}\phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^m E_{\cdot|\mathbf{S}}(\varepsilon_{s_{i_j}}) \\
&\quad E_{\cdot|\mathbf{S}} \left| \prod_{j=1}^2 K_2 \left(\frac{d(x_i, X_{u_j}(\mathbf{s}_{i_j}))}{h} \right) \left[\prod_{j=1}^m \sigma \left(\frac{\mathbf{s}_{i_j}}{A_n}, X_{s_{i_j}, A_n} \right) - \prod_{j=1}^m \sigma(\mathbf{u}_j, \mathbf{x}_j) \right] \right| \\
&\lesssim \frac{1}{n^{3/2}h^{2d}\phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^m E_{\cdot|\mathbf{S}}(\varepsilon_{s_{i_j}})
\end{aligned}$$

$$\begin{aligned}
 & \times (o_{\mathbb{P}}(1)) \int_0^h \prod_{k=1}^m K_2 \left(\frac{y_k}{h} \right) dF_{i_k/n}(y_k, x_k) \\
 \lesssim & \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^m E_{\cdot|\mathbf{S}}(\varepsilon_{s_{i_j}}) \\
 & \times (o_{\mathbb{P}}(1)) (\phi^2(h)) \\
 \sim & o_{\mathbb{P}}(1). \tag{8.25}
 \end{aligned}$$

Under the assumptions of the lemma, we have $\beta(a; b) \leq \beta_1(a)g_1(b)$ with $\beta_1(a) \rightarrow 0$ as $a \rightarrow \infty$ and $n \rightarrow \infty$, so the term to consider is the first summand. For the second part of the inequality, we will use the work of Y. Song et al., 2019 in the non-fixed kernels settings, precisely, we will define $f_{i_1, \dots, i_m} = \prod_{k=1}^m \xi_{i_k} \times H$ and $\mathcal{F}_{i_1, \dots, i_m}$ respectively as a collections of kernels and the class of functions related to this kernel, then we will use de la Peña et al., 1999, Theorem 3.1.1 and Remarks 3.5.4 part 2 for decoupling and randomization. As we mentioned above, we will suppose that $m = 2$. Then we can see that

$$\begin{aligned}
 & E_{\cdot|\mathbf{S}} \left\| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \right. \\
 & \quad \left. \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{i, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2} \\
 & = E_{\cdot|\mathbf{S}} \left\| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} f_{i_1, i_2}(\mathbf{u}, \boldsymbol{\eta}) \right\|_{\mathcal{F}_{i_1, i_2}} \\
 & \leq c_2 E_{\cdot|\mathbf{S}} \left\| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \varepsilon_p \varepsilon_q \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} f_{i_1, i_2}(\mathbf{u}, \boldsymbol{\eta}) \right\|_{\mathcal{F}_{i_1, i_2}} \\
 & \leq c_2 E_{\cdot|\mathbf{S}} \int_0^{D_{nh}^{(U_1)}} N(t, \mathcal{F}_{i_1, i_2}, \tilde{d}_{nh, 2}^{(1)}) dt, \quad (\text{By Lemma 8.10 and Proposition 8.6.}) \tag{8.26}
 \end{aligned}$$

where $D_{nh}^{(U_1)}$ is the diameter of \mathcal{F}_{i_1, i_2} according to the distance $\tilde{d}_{nh, 2}^{(1)}$, respectively defined as

$$\begin{aligned}
 D_{nh}^{(U_1)} & := \left\| \mathbb{E}_{\varepsilon} \left\| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \varepsilon_p \varepsilon_q \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} f_{i_1, i_2}(\mathbf{u}, \boldsymbol{\eta}) \right\|_{\mathcal{F}_{i_1, i_2}} \right\| \\
 & = \left\| \mathbb{E}_{\varepsilon} \left\| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \varepsilon_p \varepsilon_q \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \right. \right. \\
 & \quad \left. \left. \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{s_i, A_n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2} \right\|,
 \end{aligned}$$

and :

$$\begin{aligned} & \tilde{d}_{nh,2}^{(1)} \left(\xi_1.K_{2,1}W'^{(u)} , \xi_2.K_{2,2}W''^{(u)} \right) \\ & := \mathbb{E}_\varepsilon \left| \frac{1}{n^{3/2}h^d\phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \varepsilon_p \varepsilon_q \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} [\xi_{1i_1} \xi_{1i_2} \right. \\ & \quad \left. \prod_{k=1}^2 K_{1,2} \left(\frac{d(x_k, \eta_{i_k})}{h} \right) W'_{s_i, A_n}{}^{(u)} - \xi_{2i_1} \xi_{2i_2} \prod_{k=1}^2 K_{2,2} \left(\frac{d(x_k, \eta_{i_k})}{h} \right) W''_{s_i, A_n}{}^{(u)} \right]. \end{aligned}$$

Let consider another semi-norm $\tilde{d}_{nh,2}^{(2)}$:

$$\begin{aligned} & \tilde{d}_{nh,2}^{(2)} \left(\xi_1.K_{2,1}W'^{(u)} , \xi_2.K_{2,2}W''^{(u)} \right) \\ & = \frac{1}{nh^d\phi^2(h)} \left[\sum_{\ell_1 \neq \ell_2 \in L_n} \left(\xi_{1i_1} \xi_{1i_2} \prod_{k=1}^2 K_{1,2} \left(\frac{d(x_k, \eta_{i_k})}{h} \right) W'_{s_i, A_n}{}^{(u)} \right. \right. \\ & \quad \left. \left. - \xi_{2i_1} \xi_{2i_2} \prod_{k=1}^2 K_{2,2} \left(\frac{d(x_k, \eta_{i_k})}{h} \right) W''_{s_i, A_n}{}^{(u)} \right)^2 \right]^{1/2}. \end{aligned}$$

One can see that

$$\tilde{d}_{nh,2}^{(1)} \left(\xi_1.K_{2,1}W'^{(u)} , \xi_2.K_{2,2}W''^{(u)} \right) \leq A_{1,n} n^{-1/2} h^d \phi(h) \tilde{d}_{nh,2}^{(2)} \left(\xi_1.K_{2,1}W'^{(u)} , \xi_2.K_{2,2}W''^{(u)} \right).$$

We readily infer that

$$\begin{aligned} & E_{\cdot|\mathbf{S}} \left\| \frac{1}{n^{3/2}h^{2d}\phi^2(h)} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \right. \\ & \quad \left. \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{i}, \varphi, n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2} \\ & \leq c_2 E_{\cdot|\mathbf{S}} \int_0^{D_{nh}^{(U_1)}} N \left(t A_{1,n}^{-d} n^{1/2}, \mathcal{F}_{i,j}, \tilde{d}_{nh,2}^{(2)} \right) dt \\ & \leq c_2 A_{1,n}^d n^{-1/2} \mathbb{P} \left\{ D_{nh}^{(U_1)} A_{1,n}^{-d} n^{1/2} \geq \lambda_n \right\} + c_m A_{1,n}^d n^{-1/2} \int_0^{\lambda_n} \log t^{-1} dt, \quad (8.27) \end{aligned}$$

where $\lambda_n \rightarrow 0$. We have

$$\frac{\left(\int_0^{\lambda_n} \log t^{-1} dt \right)}{\left(\lambda_n \log \lambda_n^{-1} \right)} \rightarrow 0,$$

where λ_n must be chosen in such a way that the following relation will be achieved

$$A_{1,n}^d \lambda_n n^{-1/2} \log \lambda_n^{-1} \rightarrow 0. \quad (8.28)$$

By utilizing the triangle inequality in conjunction with Hoeffding's trick, we are easily able to derive that

$$\begin{aligned} & A_{1,n}^d n^{-1/2} \mathbb{P} \left\{ D_{nh}^{(U_1)} \geq \lambda_n A_{1,n}^d n^{-1/2} \right\} \\ & \leq \lambda_n^{-2} A_{1,n}^{-d} n^{-5/2} h \phi^{-1}(h) E_{\cdot|\mathbf{S}} \left\| \sum_{\ell_1 \neq \ell_2 \in L_n} \left[\sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \right. \\ & \quad \left. \left. K_2 \left(\frac{d(x_2, \eta'_{i_2})}{h} \right) W_{s_{i_1}, A_n}^{(u)} \right]^2 \right\|_{\mathcal{F}_2 \mathcal{K}^2} \\ & \leq c_2 \llbracket L_n \rrbracket \lambda_n^{-2} A_{1,n}^{-d} n^{-5/2} h \phi^{-1}(h) E_{\cdot|\mathbf{S}} \left\| \sum_{\ell_1 \in L_n} \left[\sum_{i_1, i_2: s_{i_1}, s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \right. \\ & \quad \left. \left. K_2 \left(\frac{d(x_2, \eta'_{i_2})}{h} \right) W_{s_{i_1}, A_n}^{(u)} \right]^2 \right\|_{\mathcal{F}_2 \mathcal{K}^2}, \end{aligned} \quad (8.29)$$

where $\{\eta'_i\}_{i \in \mathbb{N}^*}$ are independent copies of $(\eta_i)_{i \in \mathbb{N}^*}$. By imposing :

$$\lambda_n^{-2} A_{1,n}^{d-r} n^{-1/2} \rightarrow 0, \quad (8.30)$$

we readily infer that

$$\begin{aligned} & \left\| \llbracket L_n \rrbracket \lambda_n^{-2} A_{1,n}^{-d} n^{-5/2} h \phi^{-1}(h) E_{\cdot|\mathbf{S}} \sum_{\ell_1 \in L_n} \left[\sum_{i_1, i_2: s_{i_1}, s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \xi_{i_1} \xi_{i_2} \prod_{k=1}^2 K_2 \left(\frac{d(x_k, \eta_{i_k})}{h} \right) W_{s_{i_1}, A_n}^{(u)} \right]^2 \right\|_{\mathcal{F}_2 \mathcal{K}^2} \\ & \leq O(\lambda_n^{-2} A_{1,n}^{d-r} n^{-1/2}). \end{aligned}$$

A symmetrization of the last inequality in (8.29) succeeded by an application of the Proposition 8.6 in the Appendix, gives

$$\begin{aligned} & \llbracket L_n \rrbracket \lambda_n^{-2} A_{1,n}^{-d} n^{-5/2} h \phi^{-1}(h) E_{\cdot|\mathbf{S}} \left\| \sum_{\ell_1 \in L_n} \left[\sum_{i_1, i_2: s_{i_1}, s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \varepsilon_p \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \right. \\ & \quad \left. \left. K_2 \left(\frac{d(x_2, \eta'_{i_2})}{h} \right) W_{s_{i_1}, A_n}^{(u)} \right]^2 \right\|_{\mathcal{F}_2 \mathcal{K}^2} \\ & \leq c_2 E_{\cdot|\mathbf{S}} \left(\int_0^{D_{nh}^{(U_2)}} \left(\log N(u, \mathcal{F}_{i,j}, \tilde{d}'_{nh,2}) \right)^{1/2} \right), \end{aligned} \quad (8.31)$$

where

$$D_{nh}^{(U_2)} = \left\| \mathbb{E}_\varepsilon \left[\left[L_n \right] \lambda_n^{-2} A_{1,n}^{-d} n^{-5/2} \phi^{-1}(h) \sum_{\ell_1 \in L_n} \left[\sum_{i_1, i_2: s_{i_1}, s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, \eta'_{i_2})}{h} \right) W_{s_i, A_n}^{(u)} \right]^2 \right] \right\|_{\mathcal{F}_2 \mathcal{H}^2}.$$

and for $\xi_1, K_{2,1} W'$, $\xi_2, K_{2,2} W'' \in \mathcal{F}_{ij}$:

$$\begin{aligned} & \tilde{d}'_{nh,2} \left(\xi_1, K_{2,1} W'^{(u)}, \xi_2, K_{2,2} W''^{(u)} \right) \\ & := \mathbb{E}_\varepsilon \left[\left[L_n \right] \lambda_n^{-2} A_{1,n}^{-d} n^{-5/2} \phi^{-1}(h) \sum_{\ell_1 \in L_n} \varepsilon_p \left[\left(\sum_{i_1, i_2: s_{i_1}, s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \xi_{i_1} \xi_{i_2} K_{2,1} \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_{2,1} \left(\frac{d(x_2, \eta'_{i_2})}{h} \right) W_{s_i, A_n}^{(u)} \right)^2 - \left(\sum_{i_1, i_2 \in H_p^{(U)}} \xi_{2i} \xi_{2j} K_{2,2} \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_{2,2} \left(\frac{d(x_2, \eta'_{i_2})}{h} \right) W_{s_i, A_n}^{(u)} \right)^2 \right] \right] \right|. \end{aligned}$$

By the fact that :

$$\begin{aligned} & \mathbb{E}_\varepsilon \left[\left[L_n \right] \lambda_n^{-2} A_{1,n}^{-d} n^{-5/2} \phi^{-1}(h) \sum_{\ell_1 \in L_n} \varepsilon_p \left(\sum_{i_1, i_2: s_{i_1}, s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, \eta'_{i_2})}{h} \right) W_{s_i, A_n}^{(u)} \right)^2 \right] \right] \\ & \leq A_{1,n}^{3d/2} \lambda_n^{-2} n^{-1} \left[\left[L_n \right]^{-1} A_{1,n}^{-2d} \phi^{-2}(h_n) \sum_{\ell_1 \in L_n} \sum_{i_1, i_2: s_{i_1}, s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \left(\xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_i, \eta_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, \eta'_j)}{h} \right) W_{s_i, A_n}^{(u)} \right)^4 \right]^{1/2}, \end{aligned}$$

so:

$$A_{1,n}^{d3/2} \lambda_n^{-2} n^{-1} \rightarrow 0, \quad (8.32)$$

we have the convergence of (8.31) to zero. Recall that

$$\left[L_n \right] = O \left((A_n / A_{3,n})^d \right) \lesssim (A_n / A_{1,n})^d.$$

(II): The same block:

$$\mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \frac{1}{h^{2d} \phi^2(h)} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{i_2: s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n \\ i_1 \neq i_2}} \right. \right.$$

$$\begin{aligned}
 & \times \prod_{j=1}^2 \left\{ \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j}, A_n})}{h_n} \right) \right\} W_{\mathbf{s}_{i_j}, A_n} \Big| > \delta \Big) \\
 \leq & \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{i_2: s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n \\ i_1 \neq i_2}} \right. \right. \\
 & \left. \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \left[\prod_{j=1}^2 K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j}, A_n})}{h_n} \right) - \prod_{j=1}^2 K_2 \left(\frac{d(x_i, X_{\mathbf{u}_j}(\mathbf{s}_{i_j}))}{h} \right) \right] W_{\mathbf{s}_{i_j}, A_n} \right| > \delta \Big) \\
 & + \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{i_2: s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n \\ i_1 \neq i_2}} \right. \right. \\
 & \left. \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_i, X_{\mathbf{u}_j}(\mathbf{s}_{i_j}))}{h} \right) [W_{\mathbf{s}_{i_j}, A_n} - W_{\mathbf{s}_{i_j}, A_n}^{(u)}] \right| > \delta \Big) \\
 & + \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{i_2: s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n \\ i_1 \neq i_2}} \right. \right. \\
 & \left. \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_i, X_{\mathbf{u}_j}(\mathbf{s}_{i_j}))}{h} \right) W_{\mathbf{s}_{i_j}, A_n}^{(u)} \right| > \delta \Big) \\
 \leq & \mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{i_2: s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n \\ i_1 \neq i_2}} \right. \right. \\
 & \left. \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_{i_j}, A_n}^{(u)} \right| > \delta \Big) \\
 & + C \left(\frac{A_n}{A_{1,n}} \right)^d \beta(A_{2,n}; A_n^d) + o_{\mathbb{P}}(1) + o_{\mathbb{P}}(1),
 \end{aligned}$$

In the same manner, as I, we can show that the first and the second term in the previous inequality is of order $o_{\mathbb{P}}(1)$. So, as the preceding proof, it suffices to prove that

$$\begin{aligned}
 E_{\cdot|\mathbf{S}} \left(\left\| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{i_2: s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n \\ i_1 \neq i_2}} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \right. \right. \\
 \left. \left. \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_{i_j}, A_n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{H}^2} \right) \rightarrow 0.
 \end{aligned}$$

Notice that we treat a uniformly bounded classes functions in, we obtain uniformly in

$B^m \times \mathcal{F}_2 \mathcal{K}^2$

$$E_{\cdot|\mathbf{S}} \left(\sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{i_2: s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n \\ i_1 \neq i_2}} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_i, A_n}^{(u)} \right) = O(a_n).$$

This implies that we have to prove that, for $\mathbf{u} \in B^m$

$$\begin{aligned} \mathbb{E} E_{\cdot|\mathbf{S}} \left(\left\| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{i_1 \neq i_2 \\ i_2: s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n}} \right. \right. \\ \left. \left. \left[\prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_i, A_n}^{(u)} \right. \right. \right. \\ \left. \left. \left. - E_{\cdot|\mathbf{S}} \left(\prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_i, A_n}^{(u)} \right) \right] \right\|_{\mathcal{F}_2 \mathcal{K}^2} \right) \rightarrow 0. \end{aligned} \quad (8.33)$$

As for empirical processes, to prove (8.33), it's enough to symmetrize and show that

$$\begin{aligned} E_{\cdot|\mathbf{S}} \left\| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{i_2: s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n \\ i_1 \neq i_2}} \varepsilon_p \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \right. \\ \left. \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_i, A_n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{K}^2} \rightarrow 0. \end{aligned}$$

Similarly to how in (8.26), we have

$$\begin{aligned} E_{\cdot|\mathbf{S}} \left\| \frac{1}{n^{3/2} h^{d+1} \phi(h)} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{i_2: s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n \\ i_1 \neq i_2}} \varepsilon_p \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \right. \\ \left. \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_i, A_n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{K}^2} \leq \mathbb{E} \left(\int_0^{D_{nh}^{(U_3)}} (\log N(u, \mathcal{F}_{i_1, i_2}, \tilde{d}_{nh, 2}^{(3)}))^{1/2} du \right), \end{aligned}$$

where

$$D_{nh}^{(U_3)} = \left\| \mathbb{E}_\varepsilon \left[\frac{1}{n^{3/2}h^d\phi(h)} \sum_{\ell_1 \in L_n} \varepsilon_p \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{i_2: s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n \\ i_1 \neq i_2}} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \right. \right. \\ \left. \left. \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_i, A_n}^{(u)} \right] \right\|_{\mathcal{F}_2 \mathcal{H}^2}, \quad (8.34)$$

and the semi-metric $\tilde{d}_{nh,2}^{(3)}$ is defined by

$$\tilde{d}_{nh,2}^{(3)} \left(\xi_1 K_{2,1} W'^{(u)}, \xi_2 K_{2,2} W''^{(u)} \right) \\ = \mathbb{E}_\varepsilon \left| \frac{1}{n^{3/2}h^d\phi(h)} \sum_{\ell_1 \in L_n} \varepsilon_p \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{i_2: s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n \\ i_1 \neq i_2}} \left(\xi_{1i} \xi_{1j} K_{2,1} \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \right. \\ \left. \left. K_{2,1} \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W'_{\mathbf{s}_i, A_n}{}^{(u)} - \xi_{2i} \xi_{2j} K_{2,2} \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_{2,2} \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W''_{\mathbf{s}_i, A_n}{}^{(u)} \right) \right|.$$

Since we are considering uniformly bounded classes of functions, we obtain

$$\mathbb{E}_\varepsilon \left| n^{-3/2}h\phi^{-1}(h_n) \sum_{\ell_1 \in L_n} \varepsilon_p \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{i_2: s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n \\ i_1 \neq i_2}} \xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \\ \left. K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{\mathbf{s}_i, A_n}^{(u)} \right| \\ \leq A_{1,n}^{3d/2} n^{-1} h \phi^{-1}(h_n) \left[\frac{1}{\llbracket L_n \rrbracket A_{1,n}^2} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{i_2: s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n \\ i_1 \neq i_2}} \left(\xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) \right. \right. \\ \left. \left. K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{\mathbf{s}_i, A_n}^{(u)} \right)^2 \right]^{1/2} \leq O \left(A_{1,n}^{3d/2} n^{-1} h \phi^{-1}(h_n) \right).$$

Since $A_{1,n}^{3d/2} n^{-1} h \phi^{-1}(h) \rightarrow 0$, $D_{nh}^{(U_3)} \rightarrow 0$, we obtain $\text{II} \rightarrow 0$ as $n \rightarrow \infty$.

(III): Different types of blocks: Avoiding the repetition, we can directly see that:

$$\mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \frac{1}{h^{2d}\phi^2(h)} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\Delta_2} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \right. \right.$$

$$\begin{aligned}
& \sum_{i_2:s_{i_2} \in \Gamma_n(\ell_2;\varepsilon) \cap R_n} \prod_{j=1}^2 \left\{ \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j}, A_n})}{h_n} \right) \right\} W_{\mathbf{s}_{i_1}, A_n} \Big| > \delta \Big) \\
\leq & \mathbb{P} \left(\sup_{\mathcal{F}_m, \mathcal{K}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{\ell_1 \in L_n} \sum_{i_1:s_{i_1} \in \Gamma_n(\ell_1;\varepsilon_0) \cap R_n} \sum_{\Delta_2} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \right. \right. \\
& \left. \left. \sum_{i_2:s_{i_2} \in \Gamma_n(\ell_2;\varepsilon) \cap R_n} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_{i_1}, A_n}^{(u)} \right| > \delta \right) \\
& + C \left(\frac{A_n}{A_{1,n}} \right)^d \beta(A_{2,n}; A_n^d) + o_{\mathbb{P}}(1) + o_{\mathbb{P}}(1). \tag{8.35}
\end{aligned}$$

For $p = 1$ and $p = \nu_n$:

$$\begin{aligned}
& E_{\cdot|\mathbf{S}} \left\| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{i_1:s_{i_1} \in \Gamma_n(\ell_1;\varepsilon_0) \cap R_n} \sum_{\Delta_2} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2:s_{i_2} \in \Gamma_n(\ell_2;\varepsilon) \cap R_n} \right. \\
& \quad \left. \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_{i_1}, A_n}^{(u)} \right\|_{\mathcal{F}_2, \mathcal{K}^2} \\
& = E_{\cdot|\mathbf{S}} \left\| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{i_1:s_{i_1} \in \Gamma_n(\ell_1;\varepsilon_0) \cap R_n} \sum_{\substack{\ell_2: \min_{1 \leq i \leq d} \ell_{2i} = 3}} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2:s_{i_2} \in \Gamma_n(\ell_2;\varepsilon) \cap R_n} \right. \\
& \quad \left. \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_{i_1}, A_n}^{(u)} \right\|_{\mathcal{F}_2, \mathcal{K}^2}.
\end{aligned}$$

For $2 \leq p \leq \nu_n - 1$, we obtain :

$$\begin{aligned}
& E_{\cdot|\mathbf{S}} \left\| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{i_1:s_{i_1} \in \Gamma_n(\ell_1;\varepsilon_0) \cap R_n} \sum_{\Delta_2} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2:s_{i_2} \in \Gamma_n(\ell_2;\varepsilon) \cap R_n} \right. \\
& \quad \left. \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_{i_1}, A_n}^{(u)} \right\|_{\mathcal{F}_2, \mathcal{K}^2} \\
& = E_{\cdot|\mathbf{S}} \left\| \frac{1}{n^{3/2} h^{2d} \phi^2(h)} \sum_{i_1:s_{i_1} \in \Gamma_n(\ell_1;\varepsilon_0) \cap R_n} \sum_{\substack{\ell_2: \min_{1 \leq i \leq d} \ell_{2i} = 4}} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2:s_{i_2} \in \Gamma_n(\ell_2;\varepsilon) \cap R_n} \right. \\
& \quad \left. \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_{i_1}, A_n}^{(u)} \right\|_{\mathcal{F}_2, \mathcal{K}^2}
\end{aligned}$$

$$\leq E_{\cdot|\mathbf{S}} \left\| \frac{1}{n^{3/2}h^{2d}\phi^2(h)} \sum_{i_1:s_{i_1} \in \Gamma_n(\boldsymbol{\ell}_1;\varepsilon_0) \cap R_n} \sum_{\substack{\boldsymbol{\ell}_2: \min_{1 \leq i \leq d} \ell_{2i}=3 \\ \boldsymbol{\ell}_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2:s_{i_2} \in \Gamma_n(\boldsymbol{\ell}_2;\varepsilon) \cap R_n} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_i, A_n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{K}^2},$$

therefore it suffices to show that:

$$E_{\cdot|\mathbf{S}} \left\| \frac{1}{n^{3/2}h^{2d}\phi^2(h)} \sum_{i_1:s_{i_1} \in \Gamma_n(\boldsymbol{\ell}_1;\varepsilon_0) \cap R_n} \sum_{\substack{\boldsymbol{\ell}_2: \min_{1 \leq i \leq d} \ell_{2i}=3 \\ \boldsymbol{\ell}_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2:s_{i_2} \in \Gamma_n(\boldsymbol{\ell}_2;\varepsilon) \cap R_n} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_i, A_n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{K}^2}. \quad (8.36)$$

By similar arguments as in Arcones and Yu, 1994, the usual symmetrization is applied and:

$$\begin{aligned} & E_{\cdot|\mathbf{S}} \left\| \frac{\llbracket L_n \rrbracket}{n^{3/2}h^{2d}\phi^2(h)} \sum_{i_1:s_{i_1} \in \Gamma_n(\boldsymbol{\ell}_1;\varepsilon_0) \cap R_n} \sum_{\substack{\boldsymbol{\ell}_2: \min_{1 \leq i \leq d} \ell_{2i}=3 \\ \boldsymbol{\ell}_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2:s_{i_2} \in \Gamma_n(\boldsymbol{\ell}_2;\varepsilon) \cap R_n} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_i, A_n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{K}^2} \\ & \leq 2E_{\cdot|\mathbf{S}} \left\| \frac{\llbracket L_n \rrbracket}{n^{3/2}h^{2d}\phi^2(h)} \sum_{i_1:s_{i_1} \in \Gamma_n(\boldsymbol{\ell}_1;\varepsilon_0) \cap R_n} \sum_{\substack{\boldsymbol{\ell}_2: \min_{1 \leq i \leq d} \ell_{2i}=3 \\ \boldsymbol{\ell}_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2:s_{i_2} \in \Gamma_n(\boldsymbol{\ell}_2;\varepsilon) \cap R_n} \varepsilon_q \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_i, A_n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{K}^2} \\ & = 2E_{\cdot|\mathbf{S}} \left\{ \left\| \frac{\llbracket L_n \rrbracket}{n^{3/2}h^{2d}\phi^2(h)} \sum_{i_1:s_{i_1} \in \Gamma_n(\boldsymbol{\ell}_1;\varepsilon_0) \cap R_n} \sum_{\substack{\boldsymbol{\ell}_2: \min_{1 \leq i \leq d} \ell_{2i}=3 \\ \boldsymbol{\ell}_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2:s_{i_2} \in \Gamma_n(\boldsymbol{\ell}_2;\varepsilon) \cap R_n} \varepsilon_q \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_i, A_n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{K}^2} \mathbf{1}_{\left\{ D_{nh}^{(U_4)} \leq \gamma_n \right\}} \right\} \end{aligned}$$

$$\begin{aligned}
 & +2E_{\mathbf{S}} \left\| \left\| \frac{\llbracket L_n \rrbracket}{n^{3/2} h^{2d} \phi^2(h)} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{\ell_2: \min_{1 \leq i \leq d} \ell_{2i} = 3}} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon) \cap R_n} \right. \\
 & \quad \left. \varepsilon_q \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j} / A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_{i_1, A_n}}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{K}^2} \mathbf{1}_{\left\{ D_{nh}^{(U_4)} > \gamma_n \right\}} \right\| \\
 & = 2\text{III}_1 + 2\text{III}_2, \tag{8.37}
 \end{aligned}$$

where

$$\begin{aligned}
 & D_{nh}^{(U_4)} \\
 & = \left\| \frac{\llbracket L_n \rrbracket}{n^{3/2} h^{2d} \phi^2(h)} \left[\sum_{\substack{\ell_2: \min_{1 \leq i \leq d} \ell_{2i} = 3 \\ \varepsilon \neq \varepsilon_0}} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \left(\sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon) \cap R_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \right. \right. \right. \\
 & \quad \left. \left. \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j} / A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_{i_1, A_n}}^{(u)} \right)^2 \right]^{1/2} \right\|_{\mathcal{F}_2 \mathcal{K}^2} \tag{8.38}
 \end{aligned}$$

In a similar way as in (8.26), we infer that

$$\text{III}_1 \leq c_2 \int_0^{\gamma_n} \left(\log N \left(t, \mathcal{F}_{i_1, i_2}, \tilde{d}_{nh, 2}^{(4)} \right) \right)^{1/2} dt, \tag{8.39}$$

where

$$\begin{aligned}
 & \tilde{d}_{nh, 2}^{(4)} \left(\xi_1. K_{2,1} W'^{(u)}, \xi_2. K_{2,2} W''^{(u)} \right) \\
 & := \mathbb{E}_\varepsilon \left| \llbracket L_n \rrbracket n^{-3/2} h \phi^{-1}(h_n) \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{\ell_2: \min_{1 \leq i \leq d} \ell_{2i} = 3}} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon) \cap R_n} \right. \\
 & \quad \left. \varepsilon_q \left[\xi_{1i_1} \xi_{1i_2} K_{2,1} \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_{2,1} \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{\mathbf{s}_{i_1, A_n}}'^{(u)} \right. \right. \\
 & \quad \left. \left. - \xi_{2i_1} \xi_{2i_2} K_{2,2} \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_{2,2} \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{\mathbf{s}_{i_1, A_n}}''^{(u)} \right] \right|.
 \end{aligned}$$

Since we have

$$\mathbb{E}_\varepsilon \left| \llbracket L_n \rrbracket n^{3/2} h^{2d} \phi^2(h) \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{\ell_2: \min_{1 \leq i \leq d} \ell_{2i} = 3}} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon) \cap R_n} \varepsilon_q \xi_{i_1} \xi_{i_2} \right|$$

$$\begin{aligned}
 & \left| K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{\mathbf{s}_i, A_n}^{(u)} \right| \\
 \leq & A_{1,n}^{-d/2} A_{2,n}^d h^{-d+1} \phi(h) \left(\frac{1}{A_{1,n}^d A_{2,n}^d \llbracket L_n \rrbracket h^{d-1} \phi^4(h_n)} \sum_{i_1: \mathbf{s}_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{\ell_2: \min_{1 \leq i \leq d} \ell_{2i} = 3 \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2: \mathbf{s}_{i_2} \in \Gamma_n(\ell_2; \varepsilon) \cap R_n} \left[\xi_{i_1} \xi_{i_2} K_2 \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_2 \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{\mathbf{s}_i, A_n}^{(u)} \right]^2 \right)^{1/2},
 \end{aligned}$$

and considering the semi-metric

$$\begin{aligned}
 & \tilde{d}_{nh,2}^{(5)}(\xi_1 K_{2,1} W'^{(u)}, \xi_2 K_{2,2} W''^{(u)}) \\
 := & \left(\frac{1}{A_{1,n}^d A_{2,n}^d \llbracket L_n \rrbracket h^{d-1} \phi^4(h_n)} \sum_{i_1: \mathbf{s}_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{\ell_2: \min_{1 \leq i \leq d} \ell_{2i} = 3 \\ \varepsilon \neq \varepsilon_0}} \sum_{\ell_2 \in L_{1,n} \cup L_{2,n}} \sum_{i_2: \mathbf{s}_{i_2} \in \Gamma_n(\ell_2; \varepsilon) \cap R_n} [\xi_{i_1} \xi_{i_2} K_{2,1} \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_{2,1} \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{\mathbf{s}_i, A_n}'^{(u)} - \xi_{2i_1} \xi_{2i_2} K_{2,2} \left(\frac{d(x_1, \eta_{i_1})}{h} \right) K_{2,2} \left(\frac{d(x_2, \eta_{i_2})}{h} \right) W_{\mathbf{s}_i, A_n}''^{(u)}]^2 \right)^{1/2}.
 \end{aligned}$$

We demonstrate that the statement in (8.39) is bounded as follows

$$\llbracket L_n \rrbracket^{1/2} A_{2,n}^d n^{-1/2} h^2 \phi(h) \int_0^{\llbracket L_n \rrbracket^{-1/2} A_{2,n}^{-d} n^{1/2} h^{2d} \gamma_n} (\log N(t, \mathcal{F}_{i_1, i_2}, \tilde{d}_{nh,2}^{(5)}))^{1/2} dt,$$

by choosing $\gamma_n = n^{-\alpha}$ for some $\alpha > (17r - 26)/60r$, we obtain the convergence of the preceding quantity to zero. In order to bound the second term on the right-hand side of (8.37), we can mention that

$$\begin{aligned}
 \text{III}_2 &= \mathbb{E} \left\{ \left\| \llbracket L_n \rrbracket n^{-3/2} h \phi^{-2}(h) \sum_{i_1: \mathbf{s}_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{\ell_2: \min_{1 \leq i \leq d} \ell_{2i} = 3 \\ \varepsilon \neq \varepsilon_0}} \sum_{\ell_2 \in L_{1,n} \cup L_{2,n}} \sum_{i_2: \mathbf{s}_{i_2} \in \Gamma_n(\ell_2; \varepsilon) \cap R_n} \varepsilon_q \right. \right. \\
 & \quad \left. \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j} / A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_i, A_n}^{(u)} \right\|_{\mathcal{F}_2 \mathcal{K}^2} \mathbf{1}_{\{D_{nh}^{(U_4)} > \gamma_n\}} \right\} \\
 & \leq A_{1,n}^{-1} A_{2,n} n^{1/2} h^d \phi^{-1}(h) \mathbb{P} \left\{ \left\| \llbracket L_n \rrbracket^2 n^{-3} h^2 \phi^{-2}(h_n) \sum_{\substack{\ell_2: \min_{1 \leq i \leq d} \ell_{2i} = 3 \\ \varepsilon \neq \varepsilon_0}} \sum_{\ell_2 \in L_{1,n} \cup L_{2,n}} \left(\sum_{i_2: \mathbf{s}_{i_2} \in \Gamma_n(\ell_2; \varepsilon) \cap R_n} \right. \right. \right. \\
 & \quad \left. \left. \left. \sum_{i_1: \mathbf{s}_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \prod_{j=1}^2 \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j} / A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_i, A_n}^{(u)} \right)^2 \right\|_{\mathcal{F}_2 \mathcal{K}^2} \geq \gamma_n^2 \right\}. \quad (8.40)
 \end{aligned}$$

We are going to use the square root method on the last expression conditionally on $\Gamma_n(\boldsymbol{\ell}_1; \varepsilon_0) \cap R_n$. We denote by $E_{\varepsilon \neq \varepsilon_0}$ the expectation with respect to $\sigma\{\eta_{i_2}, \varepsilon \neq \varepsilon_0\}$ and we will suppose that any class of functions \mathcal{F}_m is unbounded and its envelope function satisfies for some $p > 2$:

$$\theta_p := \sup_{\mathbf{x} \in \mathcal{H}^m} \mathbb{E}(F^p(\mathbf{Y}) | \mathbf{X} = \mathbf{x}) < \infty, \quad (8.41)$$

for $2r/(r-1) < s < \infty$, (in the notation in of Giné et al., 1984, Lemma 5.2).

$$M_n = \llbracket L_n \rrbracket^{1/2} E_{\varepsilon \neq \varepsilon_0} \left(\sum_{i_2: s_{i_2} \in \Gamma_n(\boldsymbol{\ell}_2; \varepsilon) \cap R_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\boldsymbol{\ell}_1; \varepsilon_0) \cap R_n} \prod_{j=1}^2 \overline{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \prod_{j=1}^2 K_2 \left(\frac{d(x_j, \eta_{i_j})}{h} \right) W_{\mathbf{s}_{i_1}, A_n}^{(u)} \right)^2,$$

where

$$x = \gamma_n^2 A_{1,n}^{5d/2} n^{1/2} h^{md/2} \phi^{-m/2}(h), \quad \rho = \lambda = 2^{-4} \gamma_n A_{1,n}^{5d/4} n^{1/4} h^{md/4} \phi^{-m/4}(h),$$

and

$$m = \exp \left(\gamma_n^2 n h^{2d} \phi^{-2}(h_n) A_{2,n}^{-2d} \right).$$

However, since we need $t > 8M_n$, and $m \rightarrow \infty$, by similar arguments as in Arcones and Yu, 1994, page 69, We reach the convergence of (8.39) and (8.40) to zero.

(IV): Blocks of different types: The target here is to prove that:

$$\mathbb{P} \left(\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \frac{1}{h^{2d} \phi^2(h)} \sum_{\boldsymbol{\ell}_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\boldsymbol{\ell}_1; \varepsilon_0) \cap R_n} \sum_{\Delta_1} \sum_{\substack{\boldsymbol{\ell}_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2: s_{i_2} \in \Gamma_n(\boldsymbol{\ell}_2; \varepsilon) \cap R_n} \right. \right. \\ \left. \left. \times \prod_{j=1}^2 \left\{ \overline{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j}, A_n})}{h_n} \right) \right\} W_{\mathbf{s}_{i_1}, A_n} \right| > \delta \right) \rightarrow 0.$$

We have

$$\left\| n^{-3/2} \frac{1}{h^{2d} \phi^2(h)} \sum_{\boldsymbol{\ell}_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\boldsymbol{\ell}_1; \varepsilon_0) \cap R_n} \sum_{\Delta_1} \sum_{\substack{\boldsymbol{\ell}_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2: s_{i_2} \in \Gamma_n(\boldsymbol{\ell}_2; \varepsilon) \cap R_n} \right\|$$

$$\begin{aligned} & \times \prod_{j=1}^2 \left\{ \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j}, A_n})}{h_n} \right) \right\} W_{\mathbf{s}_{i_1}, A_n} \Big\|_{\mathcal{F}_2, \mathcal{H}^2} \\ & \leq c_2 \llbracket L_n \rrbracket A_{1,n}^d A_{2,n}^d n^{-3/2} h^{-d} \phi^{-1}(h) \rightarrow 0. \end{aligned}$$

Hence the proof of the lemma is complete. \square

The final step in the proof of Proposition 3.1 lies in the use of Lemma 8.1 to prove that the nonlinear term converges to zero. \square

Proof of Theorem 3.1

We have

$$\begin{aligned} & \widehat{r}_n^{(m)}(\varphi, \mathbf{x}, \mathbf{u}; h_n) - r^{(m)}(\varphi, \mathbf{x}, \mathbf{u}) \\ & = \frac{1}{\widetilde{r}_1(\varphi, \mathbf{x}, \mathbf{u})} \left(\widehat{g}_1(\mathbf{u}, \mathbf{x}) + \widehat{g}_2(\mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{x}, \mathbf{u}) \widetilde{r}_1(\varphi, \mathbf{x}, \mathbf{u}) \right), \end{aligned}$$

where

$$\begin{aligned} \widetilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}) & = \frac{(n-m)!}{n! h^{md} \phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{j=1}^m \left\{ \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j}, A_n})}{h_n} \right) \right\}, \\ \widehat{g}_1(\mathbf{u}, \mathbf{x}) & = \frac{(n-m)!}{n! h^{md} \phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{j=1}^m \left\{ \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j}, A_n})}{h_n} \right) \right\} \prod_{j=1}^m \varepsilon_{\mathbf{s}_{i_j}, A_n}, \\ \widehat{g}_2(\mathbf{u}, \mathbf{x}) & = \frac{(n-m)!}{n! h^{md} \phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{j=1}^m \left\{ \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j}, A_n})}{h_n} \right) \right\} \\ & \quad \times r^{(m)} \left(\varphi, \mathbf{X}_{\mathbf{s}_{i_1}, A_n}, \dots, \mathbf{X}_{\mathbf{s}_{i_m}, A_n}, \frac{\mathbf{s}_{i_1}}{A_n}, \dots, \frac{\mathbf{s}_{i_m}}{A_n} \right). \end{aligned}$$

The proof of this theorem is involved and divided into the following 4 steps, where in each one, we aim to show that

Step 1.

$$\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |\widehat{g}_1(\mathbf{u}, \mathbf{x})| = O_{\mathbb{P}} \left(\sqrt{\log n / n h^{md} \phi^m(h)} \right).$$

Step 2.

$$\begin{aligned} & \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} |\widehat{g}_2(\mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \widetilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}; h_n) \\ & - E_{\cdot|\mathbf{S}}(\widehat{g}_2(\mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \widetilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}; h_n))| = O_{\mathbb{P}} \left(\sqrt{\log n / n h^{md} \phi^m(h)} \right). \end{aligned}$$

Step 3. Let $\kappa_2 = \int_{\mathbb{R}} x^2 K(x) dx$.

$$\begin{aligned} & \sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| E_{\cdot|\mathbf{S}} \left(\widehat{g}_2(\mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \widetilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}; h_n) \right) \right| \\ &= O \left(\frac{1}{A_n^{dp} \phi(h)} \right) + o(h^2), P_S - a.s. \end{aligned}$$

Step 4.

$$\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \widetilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}) - E_{\cdot|\mathbf{S}}(\widetilde{r}_1(\varphi, \mathbf{u}, \mathbf{x})) \right| = o_{P_{\cdot|\mathbf{S}}}(1).$$

It is clear that **Step 1.** follows directly from Proposition 3.1 for $W_{\mathbf{s}_i, A_n} = \prod_{j=1}^m \varepsilon_{\mathbf{s}_{i_j}, A_n}$. The second one (**Step 2.**) holds also if we replace $W_{\mathbf{s}_i, A_n}$ with $\widehat{g}_2(\mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \widetilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}; h_n)$ then applying Proposition 3.1.

We will pass now to **Step 4.** Observe that

for $W_{\mathbf{s}_i, A_n} \equiv 1$, the previous mentioned proposition proved that

$$\sup_{\mathcal{F}_m \mathcal{H}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in B^m} \left| \widetilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}) - E_{\cdot|\mathbf{S}}(\widetilde{r}_1(\varphi, \mathbf{u}, \mathbf{x})) \right| = o_{P_{\cdot|\mathbf{S}}}(1).$$

Step 3. will be treated in what follows:

Let $K_0 : [0, 1] \rightarrow \mathbb{R}$ be a Lipschitz continuous function compactly support on $[-qC_1, qC_1]$ for some $q > 1$ and such that $K_0(x) = 1, \forall x \in [-C_1, C_1]$. Show that

$$E_{\cdot|\mathbf{S}} \left[\widehat{g}_2(\mathbf{u}, \mathbf{x}) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \widetilde{r}_1(\varphi, \mathbf{u}, \mathbf{x}; h_n) \right] = \sum_{i=1}^4 Q_i(\mathbf{u}, \mathbf{x}), \quad (8.42)$$

where Q_i can be defined as follows

$$Q_i(\mathbf{u}, \mathbf{x}) = \frac{(n-m)!}{n! h^{md} \phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \left\{ \prod_{j=1}^m \overline{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \right\} q_i(\mathbf{u}, \mathbf{x}), \quad (8.43)$$

such that

$$\begin{aligned} q_1(\mathbf{u}, \mathbf{x}) &= E_{\cdot|\mathbf{S}} \left[\prod_{j=1}^m K_0 \left(\frac{d(x_j, X_{\mathbf{s}_{i_j}, A_n})}{h_n} \right) \left\{ \prod_{j=1}^m K_2 \left(\frac{d(x_j, X_{\mathbf{s}_{i_j}, A_n})}{h_n} \right) \right. \right. \\ &\quad \left. \left. - \prod_{j=1}^m K_2 \left(\frac{d \left(x_i, X_{\frac{\mathbf{s}_{i_j}}{A_n}}(s_{i_j}) \right)}{h} \right) \right\} \times \left\{ r^{(m)} \left(\varphi, \frac{\mathbf{s}_i}{A_n}, \mathbf{X}_{\mathbf{s}_i, A_n} \right) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \right\} \right], \\ q_2(\mathbf{u}, \mathbf{x}) &= E_{\cdot|\mathbf{S}} \left[\prod_{j=1}^m \left\{ K_0 \left(\frac{d(x_j, X_{\mathbf{s}_{i_j}, A_n})}{h_n} \right) K_2 \left(\frac{d \left(x_i, X_{\frac{\mathbf{s}_{i_j}}{A_n}}(s_{i_j}) \right)}{h} \right) \right\} \right] \end{aligned}$$

$$\begin{aligned}
 q_3(\mathbf{u}, \mathbf{x}) &= E_{|\mathbf{S}} \left[\left\{ r^{(m)} \left(\varphi, \frac{\mathbf{s}_i}{A_n}, \mathbf{X}_{s_i, A_n} \right) - r^{(m)} \left(\varphi, \frac{\mathbf{s}_i}{A_n}, X_{s_i/A_n}(s_i) \right) \right\} \right], \\
 q_4(\mathbf{u}, \mathbf{x}) &= E_{|\mathbf{S}} \left[\left\{ \prod_{j=1}^m K_0 \left(\frac{d(x_j, X_{s_{i_j}, A_n})}{h_n} \right) - \prod_{j=1}^m K_0 \left(\frac{d(x_i, X_{\frac{s_{i_j}}{A_n}}(s_{i_j}))}{h} \right) \right\} \right. \\
 &\quad \left. \prod_{j=1}^m K_2 \left(\frac{d(x_i, X_{\frac{s_{i_j}}{A_n}}(s_{i_j}))}{h} \right) \times \left\{ r^{(m)} \left(\varphi, \frac{\mathbf{s}_i}{A_n}, \mathbf{X}_{s_i/A_n}(s_i) \right) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \right\} \right], \\
 q_4(\mathbf{u}, \mathbf{x}) &= E_{|\mathbf{S}} \left[\prod_{j=1}^m K_2 \left(\frac{d(x_i, X_{\frac{s_{i_j}}{A_n}}(s_{i_j}))}{h} \right) \left\{ r^{(m)} \left(\varphi, \frac{\mathbf{s}_i}{A_n}, \mathbf{X}_{s_i/A_n}(s_i) \right) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \right\} \right].
 \end{aligned}$$

Observe that

$$\begin{aligned}
 Q_1(\mathbf{u}, \mathbf{x}) &\lesssim \frac{(n-m)!}{n! h^{md} \phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \left\{ \prod_{j=1}^m \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) E_{|\mathbf{S}} \left[\prod_{j=1}^m K_0 \left(\frac{d(x_j, X_{s_{i_j}, A_n})}{h_n} \right) \right. \right. \\
 &\quad \left. \left| \prod_{j=1}^m K_2 \left(\frac{d(x_j, X_{s_{i_j}, A_n})}{h_n} \right) - \prod_{j=1}^m K_2 \left(\frac{d(x_i, X_{\frac{s_{i_j}}{A_n}}(s_{i_j}))}{h} \right) \right| \right. \\
 &\quad \left. \left. \times \left| r^{(m)} \left(\varphi, \frac{\mathbf{s}_i}{A_n}, \mathbf{X}_{s_i, A_n} \right) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \right| \right\},
 \end{aligned}$$

using the properties of $r^{(m)}(\mathbf{u}, \mathbf{x})$ allow us to show that

$$\prod_{j=1}^m K_0 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j}, A_n})}{h_n} \right) \left| r^{(m)} \left(\varphi, \frac{\mathbf{s}_i}{A_n}, \mathbf{X}_{s_i, A_n} \right) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \right| \leq Ch^m$$

$$\begin{aligned}
 Q_1(\mathbf{u}, \mathbf{x}) &\leq \frac{(n-m)!}{n! h^{md} \phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \left\{ \prod_{j=1}^m \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \right\} E_{|\mathbf{S}} [Ch^m \\
 &\quad \times C \sum_{j=1}^m \left| K_2 \left(\frac{d(x_j, X_{s_{i_j}, A_n})}{h_n} \right) - K_2 \left(\frac{d(x_i, X_{\frac{s_{i_j}}{A_n}}(s_{i_j}))}{h} \right) \right|^p] \quad (\text{Using the telescoping}
 \end{aligned}$$

argument, and the boundness of K_2 for $p = \min(\rho, 1)$ and $C < \infty$)

$$\leq \frac{(n-m)!}{n! h^{md} \phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \left\{ \prod_{j=1}^m \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \right\} E_{|\mathbf{S}} \left[Ch^m \sum_{j=1}^m \left| \frac{C}{A_n^d h} U_{s_{i_j}, A_n} \left(\frac{\mathbf{s}_{i_j}}{A_n} \right) \right|^p \right]$$

$$\leq \frac{C}{A_n^{pd} \phi^m(h) h^{p-m}} \quad \text{uniformly in } \mathbf{u}.$$

In a similar way, and for

$$\mathbb{E} \left[\prod_{j=1}^m K_2 \left(\frac{d \left(x_i, X_{\frac{\mathbf{s}_{i_j}}{A_n}}(s_{i_j}) \right)}{h} \right) \right] \leq C \phi^{m-1}(h),$$

and since $r^{(m)}(\cdot)$ is Lipschitz and $d \left(\mathbf{X}_{\mathbf{s}_{i_j}, A_n}, \mathbf{X}_{\frac{\mathbf{s}_{i_j}}{A_n}}(s_{i_j}) \right) \leq \frac{C}{A_n^d} U_{\mathbf{s}_{i_j}, A_n} \left(\frac{\mathbf{s}_{i_j}}{A_n} \right)$ and the variable $U_{\mathbf{s}_{i_j}, A_n} \left(\frac{\mathbf{s}_{i_j}}{A_n} \right)$ have finite p -th moment, we can see that

$$\begin{aligned} Q_2(\mathbf{u}, \mathbf{x}) &= \frac{(n-m)!}{n! h^{md} \phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{j=1}^m \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \\ &\quad E_{\cdot|\mathbf{S}} \left[\prod_{j=1}^m \left\{ K_0 \left(\frac{d(x_j, X_{\mathbf{s}_{i_j}, A_n})}{h_n} \right) K_2 \left(\frac{d \left(x_i, X_{\frac{\mathbf{s}_{i_j}}{A_n}}(s_{i_j}) \right)}{h} \right) \right\} \right. \\ &\quad \left. \left\{ r^{(m)} \left(\varphi, \frac{\mathbf{s}_i}{A_n}, \mathbf{X}_{\mathbf{s}_i, A_n} \right) - r^{(m)} \left(\varphi, \frac{\mathbf{s}_i}{A_n}, X_{\mathbf{s}_i/A_n}(s_i) \right) \right\} \right] \\ &\leq \frac{(n-m)!}{n! h^{md} \phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \left\{ \prod_{j=1}^m \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \right\} E_{\cdot|\mathbf{S}} \left[\phi^{m-1}(h) \left| \frac{C}{A_n^d} U_{\mathbf{s}_{i_j}, A_n} \left(\frac{\mathbf{s}_{i_j}}{A_n} \right) \right|^p \right] \\ &\leq \frac{C}{A_n^{pd} \phi(h)}, \end{aligned} \tag{8.44}$$

and

$$\sup_{\mathcal{F}_m \mathcal{K}^m} \sup_{\mathbf{x} \in \mathcal{H}^m} \sup_{\mathbf{u} \in I_h^m} Q_3(\mathbf{u}, \mathbf{x}) \lesssim \frac{1}{A_n^{pd} \phi^m(h) h^{p-m}}. \tag{8.45}$$

For the last term, we have

$$\begin{aligned} Q_4(\mathbf{u}, \mathbf{x}) &= \frac{(n-m)!}{n! h^{md} \phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{j=1}^m \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \\ &\quad E_{\cdot|\mathbf{S}} \left[\prod_{j=1}^m K_2 \left(\frac{d \left(x_i, X_{\frac{\mathbf{s}_{i_j}}{A_n}}(s_{i_j}) \right)}{h} \right) \left\{ r^{(m)} \left(\varphi, \frac{\mathbf{s}_i}{A_n}, \mathbf{X}_{\mathbf{s}_i/A_n}(s_i) \right) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \right\} \right]. \end{aligned}$$

Using Lemma 8.2 and inequality (2.14) and under Assumption 7, it follows that

$$|Q_4(\mathbf{u}, \mathbf{x})| \leq \frac{(n-m)!}{n! h^{md} \phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{j=1}^m \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right)$$

$$\begin{aligned}
 & E_{\cdot|\mathbf{s}} \left[\prod_{j=1}^m K_2 \left(\frac{d \left(x_i, X_{\frac{\mathbf{s}_{i_j}}{A_n}}(s_{i_j}) \right)}{h} \right) \left| r^{(m)} \left(\varphi, \frac{\mathbf{s}_i}{A_n}, \mathbf{X}_{\mathbf{s}_i/A_n}(\mathbf{s}_i) \right) - r^{(m)}(\varphi, \mathbf{u}, \mathbf{x}) \right| \right] \\
 & \lesssim \frac{(n-m)!}{n! h^{md} \phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \prod_{j=1}^m \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \\
 & E_{\cdot|\mathbf{s}} \left[\prod_{j=1}^m K_2 \left(\frac{d \left(x_i, X_{\frac{\mathbf{s}_{i_j}}{A_n}}(s_{i_j}) \right)}{h} \right) \left| d_{\mathcal{H}^m} \left(\mathbf{X}_{\mathbf{s}_i/A_n}(\mathbf{s}_i), \mathbf{x} \right) + \left\| \mathbf{u} - \frac{\mathbf{s}_i}{A_n} \right\|^\alpha \right| \right] \\
 & \lesssim \frac{(n-m)!}{n! h^{md} \phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \left| \prod_{j=1}^m \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) \right. \\
 & \quad \left. - \int_0^1 \cdots \int_0^1 \frac{1}{h^m} \prod_{j=1}^m \bar{K} \left(\frac{(u_j - v_j)}{h} \right) dv_j \right| E_{\cdot|\mathbf{s}} \left[\left| \prod_{j=1}^m K_2 \left(\frac{d \left(x_i, X_{\frac{\mathbf{s}_{i_j}}{A_n}}(s_{i_j}) \right)}{h} \right) \right| \times h^\alpha \right] \\
 & \quad + \frac{(n-m)!}{n! h^{md} \phi^m(h)} \sum_{\mathbf{i} \in I_n^m} \int_0^1 \cdots \int_0^1 \frac{1}{h^{md}} \prod_{j=1}^m \bar{K} \left(\frac{u_j - v_j}{h} \right) dv_j \times E_{\cdot|\mathbf{s}} \left[\phi^{m-1}(h) h^\alpha \right] \\
 & \lesssim O_{P_{\cdot|\mathbf{s}}} \left(h^{2\wedge\alpha} \right).
 \end{aligned} \tag{8.46}$$

Adding the obtained results of Q_i , $1 \leq i \leq 4$, **Step 3.** yields, so as the rate of convergence of the estimator. \square

Proof of Theorem 4.1

Recall that

$$\hat{r}_n^{(m)}(\varphi, \mathbf{x}, \mathbf{u}; h_n) = \frac{\sum_{\mathbf{i} \in I_n^m} \varphi(\mathbf{Y}_{\mathbf{s}_{i_1, A_n}}, \dots, \mathbf{Y}_{\mathbf{s}_{i_m, A_n}}) \prod_{j=1}^m \left\{ \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j, A_n}})}{h_n} \right) \right\}}{\sum_{\mathbf{i} \in I_n^m} \prod_{j=1}^m \left\{ \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{s}_{i_j}/A_n}{h_n} \right) K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j, A_n}})}{h_n} \right) \right\}}.$$

For $\mathbf{x} \in \mathcal{H}^m$, $\mathbf{y} \in \mathcal{Y}^m$, define

$$G_{\varphi, \mathbf{i}}(\mathbf{x}, \mathbf{y}) := \frac{\prod_{j=1}^m \left\{ K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j, A_n}})}{h_n} \right) \right\} \varphi(\mathbf{Y}_{\mathbf{s}_{i_1, A_n}}, \dots, \mathbf{Y}_{\mathbf{s}_{i_m, A_n}})}{\mathbb{E} \prod_{j=1}^m \left\{ K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j, A_n}})}{h_n} \right) \right\}};$$

$$\begin{aligned}
 \mathcal{G} & := \{G_{\varphi, \mathbf{i}}(\cdot, \cdot) \mid \varphi \in \mathcal{F}_m, \mathbf{i} = (i_1, \dots, i_m)\}; \\
 \mathcal{G}^{(k)} & := \{\pi_{k, m} G_{\varphi, \mathbf{i}}(\cdot, \cdot), \varphi \in \mathcal{F}_m, \};
 \end{aligned}$$

$$\mathbb{U}_n(\varphi) = \mathbb{U}_n^{(m)}(G_{\varphi, \mathbf{i}}) := \frac{(n-m)!}{n!} \sum_{\mathbf{i} \in I_n^m} \prod_{j=1}^m \xi_{i_j} G_{\varphi, \mathbf{i}}(\mathbf{X}_i, \mathbf{Y}_i);$$

and the U -empirical process is defined to be

$$\mu_n(\varphi) := \sqrt{nh^m \phi(h)} \{ \mathbb{U}_n(\varphi) - \mathbb{E}(\mathbb{U}_n(\varphi)) \}.$$

Then

$$\tilde{r}_n^{(m)}(\varphi, \mathbf{x}, \mathbf{u}; h_n) = \frac{\mathbb{U}_n(\varphi)}{\mathbb{U}_n(1)}.$$

In order to establish the weak convergence of our estimator, it must be established first for $\mu_n(\varphi)$. We have mentioned before that we deal with unbounded classes of functions, that is why we should truncate the function $G_{\varphi, \mathbf{i}}(\mathbf{x}, \mathbf{y})$, indeed, for $\lambda_n = n^{1/p}$, with $p > 0$, we have:

$$\begin{aligned} G_{\varphi, \mathbf{i}}(\mathbf{x}, \mathbf{y}) &= G_{\varphi, \mathbf{i}}(\mathbf{x}, \mathbf{y}) \mathbf{1}_{\{F(\mathbf{y}) \leq \lambda_n\}} + G_{\varphi, \mathbf{i}}(\mathbf{x}, \mathbf{y}) \mathbf{1}_{\{F(\mathbf{y}) > \lambda_n\}} \\ &:= G_{\varphi, \mathbf{i}}^{(T)}(\mathbf{x}, \mathbf{y}) + G_{\varphi, \mathbf{i}}^{(R)}(\mathbf{x}, \mathbf{y}). \end{aligned}$$

We can write the U -statistic as follows :

$$\begin{aligned} \mu_n(\varphi) &= \sqrt{nh^m \phi(h)} \{ \mathbb{U}_n^{(m)}(G_{\varphi, \mathbf{i}}^{(T)}) - \mathbb{E}(\mathbb{U}_n^{(m)}(G_{\varphi, \mathbf{i}}^{(T)})) \} \\ &\quad + \sqrt{nh^m \phi(h)} \{ \mathbb{U}_n^{(m)}(G_{\varphi, \mathbf{i}}^{(R)}) - \mathbb{E}(\mathbb{U}_n^{(m)}(G_{\varphi, \mathbf{i}}^{(R)})) \} \\ &:= \sqrt{nh^m \phi(h)} \{ \mathbb{U}_n^{(T)}(\varphi, \mathbf{i}) - \mathbb{E}(\mathbb{U}_n^{(T)}(\varphi)) \} \\ &\quad + \sqrt{nh^m \phi(h)} \{ \mathbb{U}_n^{(R)}(\varphi) - \mathbb{E}(\mathbb{U}_n^{(R)}(\varphi)) \} \\ &:= \mu_n^{(T)}(\varphi) + \mu_n^{(R)}(\varphi). \end{aligned} \tag{8.47}$$

The first term is the truncated part and the second is the remaining one. We have to prove that:

1. $\mu_n^{(T)}(\varphi)$ converges to a Gaussian process.
2. The remainder part doesn't matter much, in the sense that

$$\left\| \sqrt{nh^m \phi(h)} \{ \mathbb{U}_n^{(R)}(\varphi) - \mathbb{E}(\mathbb{U}_n^{(R)}(\varphi)) \} \right\|_{\mathcal{F}_m \mathcal{X}^m} \xrightarrow{\mathbb{P}} 0.$$

For the first point, we will use the decomposition of Hoeffding, which would be the same as the previous decomposition in Subsection 3.1 except that we replace $W_{\mathbf{i}, n}$ by $\varphi(Y_{\mathbf{i}, n})$

$$\mathbb{U}_n^{(T)}(\varphi) - \mathbb{E}(\mathbb{U}_n^{(T)}(\varphi)) := \mathbb{U}_{1, n}(\varphi) + \mathbb{U}_{2, n}(\varphi),$$

where

$$\mathbb{U}_{1,n}(\varphi) := \frac{1}{n} \sum_{i=1}^n \widehat{H}_{1,i}(\mathbf{u}, \mathbf{x}, \varphi), \quad (8.48)$$

$$\mathbb{U}_{2,n}(\varphi) := \frac{(n-m)!}{(n)!} \sum_{\mathbf{i} \in I_n^m} \xi_{i_1} \cdots \xi_{i_m} H_{2,\mathbf{i}}(\mathbf{z}). \quad (8.49)$$

The convergence of $\mathbb{U}_{2,n}(\varphi)$ to zero in probability follows from Lemma 8.1. Hence, it is enough to show that $\mathbb{U}_{1,n}(\varphi)$ converges weakly to a Gaussian process called $\mathbb{G}(\varphi)$. In order to achieve our goal, we will go through finite-dimensional convergence and equicontinuity.

The finite-dimensional convergence simply asserts that every finite set of functions f_1, \dots, f_q in L_2 , for $\widetilde{\mathbb{U}}$ the centred form of \mathbb{U} :

$$\left(\sqrt{nh^m \phi(h)} \widetilde{\mathbb{U}}_{1,n}(f_1), \dots, \sqrt{nh^m \phi(h)} \widetilde{\mathbb{U}}_{1,n}(f_q) \right) \quad (8.50)$$

converges to the corresponding finite-dimensional distributions of the process $\mathbb{G}(\varphi)$. It is sufficient to show that for every fixed collection $(a_1, \dots, a_q) \in \mathbb{R}^q$ we have

$$\sum_{j=1}^q a_j \widetilde{\mathbb{U}}_{1,n}(f_j) \rightarrow N(0, v^2),$$

where

$$v^2 = \sum_{j=1}^q a_j^2 \text{Var}\left(\widetilde{\mathbb{U}}_{1,n}(f_j)\right) + \sum_{s \neq r} a_s a_r \text{Cov}\left(\widetilde{\mathbb{U}}_{1,n}(f_s), \widetilde{\mathbb{U}}_{1,n}(f_r)\right). \quad (8.51)$$

Take

$$h(\cdot) = \sum_{j=1}^q a_j f_j(\cdot).$$

By linearity of $h(\cdot)$, we have to see that

$$\widetilde{\mathbb{U}}_{1,n}(h, \mathbf{i}) \rightarrow \mathbb{G}(h).$$

Let

$$N = \mathbb{E} \prod_{j=1}^m \left\{ K_2 \left(\frac{d(x_j, \mathbf{X}_{\mathbf{s}_{i_j}, A_n})}{h_n} \right) \right\}.$$

We have:

$$\begin{aligned} & \widetilde{\mathbb{U}}_{1,n}(h_n) \\ &= N^{-1} \times \frac{1}{n} \sum_{i=1}^n \frac{(n-m)!}{(n-1)!} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_i \xi_{i_\ell} \cdots \xi_{i_{m-1}} \frac{1}{\phi(h)} K_2 \left(\frac{d(x_i, X_{\mathbf{s}_i, A_n})}{h_n} \right) \end{aligned}$$

$$\begin{aligned}
 & \times \int h(y_1, \dots, y_{\ell-1}, Y_i, y_\ell, \dots, y_{m-1}) \prod_{\substack{j=1 \\ j \neq i}}^{m-1} \frac{1}{\phi(h)} K_2 \left(\frac{d(x_j, \mathbf{X}_{s_{i_j, A_n}})}{h_n} \right) \\
 & \quad \mathbb{P}(d(\nu_1, y_1), \dots, d(\nu_{\ell-1}, y_{\ell-1}), d(\nu_\ell, y_\ell), \dots, d(\nu_{m-1}, y_{m-1})), \\
 := & N^{-1} \frac{1}{n} \sum_{i=1}^n \xi_i \frac{1}{\phi(h)} K_2 \left(\frac{d(x_i, X_{s_i, A_n})}{h_n} \right) \tilde{h}(Y_i).
 \end{aligned}$$

The next step requires an extension of the Blocking techniques of Bernstein to the spacial process where all notions are defined in Section 8.1.

Recall that $L_n = L_{1,n} \cup L_{2,n}$ and define:

$$\mathbb{Z}_{s, A_n}(\mathbf{u}, \mathbf{x}) := \xi_i \frac{1}{\phi(h)} K_2 \left(\frac{d(x_i, X_{s_i, A_n})}{h_n} \right) \tilde{h}(Y_i), \quad (8.52)$$

and

$$\mathbb{Z}_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}) = \sum_{i: s_i \in \Gamma_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}) \cap R_n} \mathbb{Z}_{s, A_n}(\mathbf{u}, \mathbf{x}) = \left(\mathbb{Z}_n^{(1)}(\boldsymbol{\ell}; \boldsymbol{\varepsilon}), \dots, \mathbb{Z}_n^{(p)}(\boldsymbol{\ell}; \boldsymbol{\varepsilon}) \right)'. \quad (8.53)$$

Then we have

$$\begin{aligned}
 \tilde{\mathbb{U}}_{1,n}(h_n) &= \sum_{i=1}^n \mathbb{Z}_{s, A_n}(\mathbf{u}, \mathbf{x}) \\
 &= \sum_{\boldsymbol{\ell} \in L_n} \mathbb{Z}_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}_0) + \sum_{\boldsymbol{\varepsilon} \neq \boldsymbol{\varepsilon}_0} \underbrace{\sum_{\boldsymbol{\ell} \in L_{1,n}} \mathbb{Z}_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon})}_{=: \mathbb{Z}_{2,n}(\boldsymbol{\varepsilon})} + \sum_{\boldsymbol{\varepsilon} \neq \boldsymbol{\varepsilon}_0} \underbrace{\sum_{\boldsymbol{\ell} \in L_{2,n}} \mathbb{Z}_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon})}_{=: \mathbb{Z}_{3,n}(\boldsymbol{\varepsilon})} \\
 &=: \mathbb{Z}_{1,n} + \sum_{\boldsymbol{\varepsilon} \neq \boldsymbol{\varepsilon}_0} \mathbb{Z}_{2,n}(\boldsymbol{\varepsilon}) + \sum_{\boldsymbol{\varepsilon} \neq \boldsymbol{\varepsilon}_0} \mathbb{Z}_{3,n}(\boldsymbol{\varepsilon}). \quad (8.54)
 \end{aligned}$$

Lemma 8.9 proves that $\mathbb{Z}_{2,n}$ and $\mathbb{Z}_{3,n}$, for $\boldsymbol{\varepsilon} \neq \boldsymbol{\varepsilon}_0$, are asymptotically negligible. Treating now the variance of $\mathbb{Z}_{1,n}$ is clear, first, mixing conditions are used to replace large blocks with independent random variables, and then Lyapunov's condition for the central limit theorem is applied to the sum of independent random variables. Similar to the proof of Proposition 3.1 using Lemma 8.5, as in Equation (8.10), observe that

$$\sup_{t>0} \left| P_{\cdot|\mathbf{S}}(\mathbb{Z}_{1,n} > t) - P_{\cdot|\mathbf{S}}(\check{\mathbb{Z}}_{1,n} > t) \right| \leq C \left(\frac{A_n}{A_{1,n}} \right)^d \beta(A_{2,n}; A_n^d), \quad (8.55)$$

where $\{\check{\mathbb{Z}}_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}) : \boldsymbol{\ell} \in L_n\}$ denotes a sequence of independent random vectors in \mathbb{R}^p under $P_{\cdot|\mathbf{S}}$ such that

$$\check{\mathbb{Z}}_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}) \stackrel{d}{=} \mathbb{Z}_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}), \text{ under } P_{\cdot|\mathbf{S}}, \boldsymbol{\ell} \in L_n.$$

Applying Lyapunov's condition for the central limit theorem for the sum of independent random variables, the remaining condition of finite-dimensional convergence must be established.

We end up with the asymptotic equicontinuity. We have to prove that:

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \sqrt{nh^m \phi(h)} \left\| \tilde{\mathbb{U}}_{1,n}(h_n, \mathbf{i}) \right\|_{\mathcal{F}\mathcal{H}(\delta, \|\cdot\|_p)} > \varepsilon \right\} = 0, \quad (8.56)$$

where,

$$\begin{aligned} \mathcal{F}\mathcal{H}(\delta, \|\cdot\|_p) := & \left\{ \tilde{\mathbb{U}}'_{1,n}(h_n) - \tilde{\mathbb{U}}''_{1,n}(h_n) : \right. \\ & \left. \left\| \tilde{\mathbb{U}}'_{1,n}(h_n) - \tilde{\mathbb{U}}''_{1,n}(h_n) \right\| < \delta, \tilde{\mathbb{U}}'_{1,n}(h_n), \tilde{\mathbb{U}}''_{1,n}(h_n) \in \mathcal{F}\mathcal{H} \right\}, \end{aligned} \quad (8.57)$$

for

$$\begin{aligned} \tilde{\mathbb{U}}'_{1,n}(h_n) &= N^{-1} \frac{1}{n} \sum_{i=1}^n \xi_i \frac{1}{\phi(h)} K_{2,1} \left(\frac{d(x_i, X_{s_i, A_n})}{h_n} \right) \tilde{h}_1(Y_i) - \mathbb{E} \left(\mathbb{U}'_{1,n}(h_n) \right) \\ \tilde{\mathbb{U}}''_{1,n}(h_n) &= N^{-1} \frac{1}{n} \sum_{i=1}^n \xi_i \frac{1}{\phi(h)} K_{2,2} \left(\frac{d(x_i, X_{s_i, A_n})}{h_n} \right) \tilde{h}_2(Y_i) - \mathbb{E} \left(\mathbb{U}''_{1,n}(h_n) \right) \end{aligned} \quad (8.58)$$

At this point, we will adapt the chaining technique found in Arcones and Yu, 1994, and use it for the conditional setting with the locally stationary process in Bouzebda and Soukarieh, 2022b but for random fields, as in Lemma 8.1.

Using the same strategy also as in Lemma 8.1 to pass from the sequence of locally stationary random variables to the stationary one and find that, for $\zeta_i = (\eta_i, \varsigma_i)$ the independent blocks sequences:

$$\begin{aligned} & \mathbb{P} \left\{ \left\| (n\phi(h))^{-1/2} h^{m/2} N^{-1} \sum_{i=1}^n \left(\xi_i K_2 \left(\frac{d(x_i, X_i)}{h} \right) \tilde{h}(Y_i) - \mathbb{E}(\mathbb{U}_{1,n}(h_n)) \right) \right\|_{\mathcal{F}\mathcal{H}(b, \|\cdot\|_p)} > \varepsilon \right\} \\ & \leq 2\mathbb{P} \left\{ \left\| (n\phi(h))^{-1/2} h^{m/2} N^{-1} \sum_{\ell \in L_n} \sum_{i: s_i \in \Gamma_n(\ell; \varepsilon_0) \cap R_n} \left(\xi_i K_2 \left(\frac{d(x_i, \eta_i)}{h} \right) \tilde{h}(\varsigma_i) \right. \right. \right. \\ & \quad \left. \left. \left. - \mathbb{E}(\mathbb{U}_{1,n}(h_n)) \right) \right\|_{\mathcal{F}\mathcal{H}(b, \|\cdot\|_p)} > \varepsilon' \right\} + C \left(\frac{A_n}{A_{1,n}} \right)^d \beta(A_{2,n}; A_n^d) + o_{\mathbb{P}}(1). \end{aligned} \quad (8.59)$$

Taking Advantage of the Condition (E2) in Assumption 12, we obtain $\beta(A_{2,n}; A_n^d) \rightarrow 0$ as $n \rightarrow 0$, then it's simply a matter of placing the first phrase in the right-hand sight of (8.59). Due to the fact that the blocks are independent, we symmetricize using a sequence $\{\varepsilon_j\}_{j \in \mathbb{N}^*}$ of i.i.d. Rademacher variables, i.e., r.v's with

$$\mathbb{P}(\varepsilon_j = 1) = \mathbb{P}(\varepsilon_j = -1) = 1/2.$$

It is important to notice that the sequence $\{\varepsilon_j\}_{j \in \mathbb{N}^*}$ is independent of the sequence

$\{\xi_i = (\varsigma_i, \zeta_i)\}_{i \in \mathbb{N}^*}$, therefore it remains to establish, for all $\varepsilon > 0$ and $\delta \rightarrow 0$,

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \left\| (n\phi(h))^{-1/2} h^{m/2} N^{-1} \sum_{\ell \in L_n} \sum_{i: s_i \in \Gamma_n(\ell; \varepsilon_0) \cap R_n} \left(\xi_i K_2 \left(\frac{d(x_i, \eta_i)}{h} \right) \tilde{h}(\varsigma_i) - \mathbb{E}(\mathbb{U}_{1,n}(h_n, \mathbf{i})) \right) \right\|_{\mathcal{F}\mathcal{X}(b, \|\cdot\|_p)} > \varepsilon \right\} < \delta. \quad (8.60)$$

Define the semi-norm:

$$\begin{aligned} \tilde{d}_{n\phi,2} &:= \left((n\phi(h))^{-1/2} h^{m/2} N^{-1} \sum_{\ell \in L_n} \sum_{i: s_i \in \Gamma_n(\ell; \varepsilon_0) \cap R_n} \left| \left(\xi_i K_{2,1} \left(\frac{d(x_i, \eta_i)}{h} \right) \tilde{h}_1(\varsigma_i) - \mathbb{E}(\mathbb{U}'_{1,n}(h_n, \mathbf{i})) \right) - \left(\xi_i K_{2,2} \left(\frac{d(x_i, \eta_i)}{h} \right) \tilde{h}_2(\varsigma_i) - \mathbb{E}(\mathbb{U}''_{1,n}(h_n, \mathbf{i})) \right) \right|^2 \right)^{1/2} \end{aligned} \quad (8.61)$$

and the covering number defined for any class of functions \mathcal{E} by :

$$\tilde{N}_{n\phi,2}(u, \mathcal{E}) := N_{n\phi,2}(u, \mathcal{E}, \tilde{d}_{n\phi,2}).$$

Because of the latter, we are able to bound (8.60), (more details are in Bouzebda et al., 2019b). In the same way, as in Bouzebda et al., 2019b and before in Arcones and Yu, 1994, as a result of the independence between the blocks and Assumption 13 (C3), and by applying Giné et al., 1984, Lemma 5.2, the equicontinuity is achieved, and then the weak convergence too.

Now, we need to show that :

$$\mathbb{P} \left\{ \left\| \mu_n^{(R)}(\varphi, \mathbf{t}) \right\|_{\mathcal{F}_m \mathcal{X}^m} > \lambda \right\} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

For clarity purposes, we restrict ourselves to $m = 2$. Using the same notation as in Lemma 8.1, we have the following decomposition:

$$\begin{aligned} \mu_n^{(R)}(\varphi, \mathbf{i}) &= \sqrt{nh^{m+d}\phi(h_n)} \left\{ \mathbb{U}_n^{(R)}(\varphi, \mathbf{i}) - \mathbb{E} \left(\mathbb{U}_n^{(R)}(\varphi, \mathbf{i}) \right) \right\} \\ &= \frac{\sqrt{nh^{m+d}\phi(h_n)}}{n(n-1)} \sum_{i_1 \neq i_2}^n \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{t}}^{(R)} \left(((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right) - \mathbb{E} \left[G_{\varphi, \mathbf{i}}^{(R)} \left(((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right) \right] \right\} \\ &\leq \frac{1}{\sqrt{nh^{m+d}\phi(h_n)}} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \left\{ \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{i}}^{(R)} \left((X_i, X_j), (Y_i, Y_j) \right) - \mathbb{E} \left[G_{\varphi, \mathbf{i}}^{(R)} \left((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2}) \right) \right] \right\} \right\} \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{\sqrt{nh^{m+d}\phi(h_n)}} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\substack{i_2: s_{i_2} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n \\ i_1 \neq i_2}} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{i}}^{(R)}((X_i, X_j), (Y_i, Y_j)) - \mathbb{E} \left[G_{\varphi, \mathbf{i}}^{(R)}((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right] \right\} \\
 +2 & \frac{1}{\sqrt{nh^{m+d}\phi(h_n)}} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\Delta_2} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{i}}^{(R)}((X_i, X_j), (Y_i, Y_j)) - \mathbb{E} \left[G_{\varphi, \mathbf{i}}^{(R)}((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right] \right\} \\
 +2 & \frac{1}{\sqrt{nh^{m+d}\phi(h_n)}} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\Delta_1} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{t}}^{(R)}((X_i, X_j), (Y_i, Y_j)) - \mathbb{E} \left[G_{\varphi, \mathbf{t}}^{(R)}((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right] \right\} \\
 & + \frac{1}{\sqrt{nh^{m+d}\phi(h_n)}} \sum_{\substack{\ell_1 \neq \ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{t}}^{(R)}((X_i, X_j), (Y_i, Y_j)) - \mathbb{E} \left[G_{\varphi, \mathbf{t}}^{(R)}((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right] \right\} \\
 & + \frac{1}{\sqrt{nh^{m+d}\phi(h_n)}} \sum_{\substack{\ell_1 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_1 < i_2 : s_{i_1}, s_{i_2} \in \Gamma_n(\ell_1; \varepsilon) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{t}}^{(R)}((X_i, X_j), (Y_i, Y_j)) - \mathbb{E} \left[G_{\varphi, \mathbf{t}}^{(R)}((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right] \right\} \\
 =: & \text{I}' + \text{II}' + \text{III}' + \text{IV}' + \text{V}' + \text{VI}' .
 \end{aligned}$$

We shall employ blocking arguments and evaluate the terms that result. We begin by examining the first I' . We obtain

$$\begin{aligned}
 & \mathbb{P} \left\{ \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{i}}^{(R)}((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right. \right. \right. \\
 & \quad \left. \left. \left. - \mathbb{E} \left[G_{\varphi, \mathbf{t}}^{(R)}((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right] \right\} \right\|_{\mathcal{F}_2, \mathcal{X}^2} > \delta \right\} \\
 \leq & \mathbb{P} \left\{ \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} \right. \right. \\
 & \quad \left. \left. \left\{ G_{\varphi, \mathbf{i}}^{(R)}((\zeta_{i_1}, \zeta_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) - \mathbb{E} \left[G_{\varphi, \mathbf{i}}^{(R)}((\zeta_{i_1}, \zeta_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) \right] \right\} \right\|_{\mathcal{F}_2, \mathcal{X}^2} > \delta \right\} \\
 & + C \left(\frac{A_n}{A_{1,n}} \right)^d \beta(A_{2,n}; A_n^d) .
 \end{aligned}$$

Recall that for all $\varphi \in \mathcal{F}_m$, and :

$$\mathbf{x} \in \mathcal{X}^2, \mathbf{y} \in \mathcal{Y}^2 : \mathbf{1}_{\{d(\mathbf{x}, X_{i,n}) \leq h\}} F(\mathbf{y}) \geq \varphi(\mathbf{y}) K_2 \left(\frac{d(x_i, X_{s_i, A_n})}{h_n} \right).$$

Hence, by the symmetry of $F(\cdot)$:

$$\begin{aligned} & \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ G_{\varphi, \mathbf{i}}^{(R)} \left((\zeta_{i_1}, \zeta_{i_2}), (\zeta_{i_1}, \zeta_{i_2}) \right) \right. \right. \\ & \quad \left. \left. - \mathbb{E} \left[G_{\varphi, \mathbf{t}}^{(R)} \left((\zeta_{i_1}, \zeta_{i_2}), (\zeta_{i_1}, \zeta_{i_2}) \right) \right] \right\} \right\|_{\mathcal{F}_2, \mathcal{X}^2} \\ & \lesssim \left| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ F(\zeta_i, \zeta_j) \mathbf{1}_{\{F > \lambda_n\}} \right. \right. \\ & \quad \left. \left. - \mathbb{E} \left[F(\zeta_i, \zeta_j) \mathbf{1}_{\{F > \lambda_n\}} \right] \right\} \right|. \end{aligned} \quad (8.62)$$

We are going to use Chebyshev's inequality, Hoeffding's trick and inequality, respectively to obtain:

$$\begin{aligned} & \mathbb{P} \left\{ \left| \frac{1}{\sqrt{n\phi(h)}} \sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} \left\{ F(\zeta_i, \zeta_j) \mathbf{1}_{\{F > \lambda_n\}} \right. \right. \right. \\ & \quad \left. \left. - \mathbb{E} \left[F(\zeta_i, \zeta_j) \mathbf{1}_{\{F > \lambda_n\}} \right] \right\} \right| > \delta \right\} \\ & \lesssim \delta^{-2} n^{-1} \phi^{-1}(h) \text{Var} \left(\sum_{\ell_1 \neq \ell_2 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} F(\zeta_i, \zeta_j) \mathbf{1}_{\{F > \lambda_n\}} \right) \\ & \lesssim c_2 \llbracket L_n \rrbracket \delta^{-2} n^{-1} \phi^{-1}(h) \text{Var} \left(\sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} F(\zeta_i, \zeta'_j) \mathbf{1}_{\{F > \lambda_n\}} \right) \\ & \lesssim 2c_2 \llbracket L_n \rrbracket \delta^{-2} n^{-2} \phi^{-1}(h) \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} \mathbb{E} \left[(F(\zeta_1, \zeta_2))^2 \mathbf{1}_{\{F > \lambda_n\}} \right]. \end{aligned} \quad (8.63)$$

Under Assumption 13 iii), we have for each $\lambda > 0$:

$$\begin{aligned} & c_2 \llbracket L_n \rrbracket \delta^{-2} n^{-2} \phi^{-1}(h_n) \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} \mathbb{E} \left[(F(\zeta_1, \zeta_2))^2 \mathbf{1}_{\{F > \lambda_n\}} \right] \\ & = c_2 \llbracket L_n \rrbracket \delta^{-2} n^{-2} \phi^{-1}(h_n) \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} \\ & \quad \times \int_0^\infty \mathbb{P} \left\{ (F(\zeta_1, \zeta_2))^2 \mathbf{1}_{\{F > \lambda_n\}} \geq t \right\} dt \end{aligned}$$

$$\begin{aligned}
 &= c_2 \llbracket L_n \rrbracket \delta^{-2} n^{-2} \phi^{-1}(h_n) \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} \int_0^{\lambda_n} \mathbb{P} \{F > \lambda_n\} dt \\
 &+ c_2 \llbracket L_n \rrbracket \delta^{-2} n^{-2} \phi^{-1}(h_n) \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon_0) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} \int_{\lambda_n}^{\infty} \mathbb{P} \{(F)^2 > t\} dt,
 \end{aligned}$$

converging to 0 as $n \rightarrow \infty$. Terms II', V' and VI' will be handled the same way as the last term was. The terms II', VI' do not follow the same line because the variables $\{\zeta_i, \zeta_j\}_{\varepsilon=\varepsilon_0}$ (or $\{\zeta_i, \zeta_j\}_{\varepsilon \neq \varepsilon_0}$ for VI') belong to the same blocks. Term IV' can be deduced from the study of Terms I' and III'. Considering the term III', we have

$$\begin{aligned}
 &\mathbb{P} \left\{ \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\Delta_2} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} \right. \right. \\
 &\quad \left. \left. \left\{ G_{\varphi, \mathbf{i}}^{(R)}((X_i, X_j), (Y_i, Y_j)) - \mathbb{E} \left[G_{\varphi, \mathbf{i}}^{(R)}((X_{i_1}, X_{i_2}), (Y_{i_1}, Y_{i_2})) \right] \right\} \right\|_{\mathcal{F}_2 \mathcal{H}^2} > \delta \right\} \\
 &\leq \mathbb{P} \left\{ \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\Delta_2} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} \right. \right. \\
 &\quad \left. \left. \left\{ G_{\varphi, \mathbf{i}}^{(R)}((\varsigma_{i_1}, \varsigma_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) - \mathbb{E} \left[G_{\varphi, \mathbf{i}}^{(R)}((\varsigma_{i_1}, \varsigma_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) \right] \right\} \right\|_{\mathcal{F}_2 \mathcal{H}^2} > \delta \right\} \\
 &+ \frac{\llbracket L_n \rrbracket A_{1,n}^d A_{2,n}^d \beta(A_{2,n}; A_n^d)}{\sqrt{n\phi(h_n)}}.
 \end{aligned} \tag{8.64}$$

We have also

$$\begin{aligned}
 &\mathbb{P} \left\{ \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\Delta_2} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} \right. \right. \\
 &\quad \left. \left. \left\{ G_{\varphi, \mathbf{i}}^{(R)}((\varsigma_{i_1}, \varsigma_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) - \mathbb{E} \left[G_{\varphi, \mathbf{i}}^{(R)}((\varsigma_{i_1}, \varsigma_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) \right] \right\} \right\|_{\mathcal{F}_2 \mathcal{H}^2} > \delta \right\} \\
 &\leq \mathbb{P} \left\{ \left\| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\Delta_2} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} \right. \right.
 \end{aligned}$$

$$\left\{ G_{\varphi, \mathbf{i}}^{(R)}((\zeta_{i_1}, \zeta_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) - \mathbb{E} \left[G_{\varphi, \mathbf{i}}^{(R)}((\zeta_{i_1}, \zeta_{i_2}), (\zeta_{i_1}, \zeta_{i_2})) \right] \right\} \Bigg|_{\mathcal{F}_2, \mathcal{K}^2} > \delta \Bigg\}.$$

Since (8.62) is still true, the problem can be reduced to

$$\begin{aligned} & \mathbb{P} \left\{ \left| \frac{1}{\sqrt{n\phi(h_n)}} \sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\Delta_2} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} \right. \right. \\ & \quad \left. \left. \left\{ F(\zeta_i, \zeta_j) \mathbf{1}_{\{F > \lambda_n\}} - \mathbb{E} \left[F(\zeta_i, \zeta_j) \mathbf{1}_{\{F > \lambda_n\}} \right] \right\} \right| > \delta \right\} \\ & \lesssim \delta^{-2} n^{-1} \phi(h_n) \text{Var} \left(\sum_{\ell_1 \in L_n} \sum_{i_1: s_{i_1} \in \Gamma_n(\ell_1; \varepsilon_0) \cap R_n} \sum_{\Delta_2} \sum_{\substack{\ell_2 \in L_{1,n} \cup L_{2,n} \\ \varepsilon \neq \varepsilon_0}} \sum_{i_2: s_{i_2} \in \Gamma_n(\ell_2; \varepsilon) \cap R_n} \phi(h_n) \xi_{i_1} \xi_{i_2} \right. \\ & \quad \left. \times F(\zeta_i, \zeta_j) \mathbf{1}_{\{F > \lambda_n\}} \right), \end{aligned}$$

the identical technique is followed as in (8.63). The remainder has just been demonstrated to be asymptotically negligible. Finally, with $\hat{r}^{(m)}(\varphi, \mathbf{x}, \mathbf{u}) \rightarrow \mathbb{E}(\mathbb{U}_n(\varphi, \mathbf{i}))$, and for $(\mathbb{U}_n(1, \mathbf{i})) \xrightarrow{\mathbb{P}} 1$, the weak convergence of our estimator is accomplished. \square

Appendix

This appendix contains supplementary information that is an essential part of providing a more comprehensive understanding of the chapter.

Assumption 14.

(KD1) (KB2) in Assumption 8 holds.

(KD2) For any $\boldsymbol{\alpha} \in \mathbb{Z}^d$ with $|\boldsymbol{\alpha}| = 1, 2$, $\partial^{\boldsymbol{\alpha}} f_{\mathbf{S}}(\mathbf{s})$ exist and continuous on $(0, 1)^d$.

Define

$$\hat{f}_{\mathbf{S}}(\mathbf{u}) = \frac{1}{nh^d} \sum_{j=1}^n \bar{K}_h(\mathbf{u} - \mathbf{S}_{0,j}).$$

Lemma 8.2. Masry, 1996, Theorem 2 . Under Assumption 14 and $h \rightarrow 0$ such that $nh^d/(\log n) \rightarrow \infty$ as $n \rightarrow \infty$, we have that

$$\sup_{\mathbf{u} \in [0,1]^d} |\hat{f}_{\mathbf{S}}(\mathbf{u}) - f_{\mathbf{S}}(\mathbf{u})| = O \left(\sqrt{\frac{\log n}{nh^d}} + h^2 \right) P_{\mathbf{S}} - a.s.$$

Lemma 8.3. Let $I_h = [C_1 h, 1 - C_1 h]$. Suppose that kernel K_1 satisfies Assumption 14 part(i). Then for $q = 0, 1, 2$ and $m > 1$:

$$\sup_{u \in I_h} \left| \frac{1}{n^m h^{md}} \sum_{\mathbf{i} \in I_h^m} \prod_{j=1}^m \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{S}_{0,i_j}}{h_n} \right) \left(\frac{\mathbf{u}_j - \mathbf{S}_{0,i_j}}{h} \right)^q - \int_{\mathbb{R}^{md}} \frac{1}{h^{md}} \prod_{j=1}^m \left\{ \bar{K} \left(\frac{\mathbf{u}_j - \boldsymbol{\omega}_j}{h_n} \right) \left(\frac{\mathbf{u}_j - \boldsymbol{\omega}_j}{h} \right)^q \right\} f_{\mathbf{S}}(\boldsymbol{\omega}_j) \prod_{j=1}^m d\boldsymbol{\omega}_j \right| = O \left(\sqrt{\frac{\log n}{nh^{dm}}} \right) P_{\mathbf{S}} - a.s.$$

Lemma 8.4. Suppose that kernel \bar{K} satisfies Assumption 14. Let $g : [0, 1]^{md} \times \mathcal{H}^m \rightarrow \mathbb{R}$, $(\mathbf{u}, \mathbf{x}) \mapsto g(\mathbf{u}, \mathbf{x})$ be a function continuously partially differentiable with respect to \mathbf{u}_j . For $k = 1, 2$, we have

$$\begin{aligned} \sup_{u \in I_h} \left| \frac{1}{n^m h^{md}} \sum_{\mathbf{i} \in I_h^m} \prod_{j=1}^m \bar{K} \left(\frac{\mathbf{u}_j - \mathbf{S}_{0,i_j}}{h_n} \right)^k g(\mathbf{S}_{0,i_j}, \mathbf{x}_j) - \prod_{j=1}^m \kappa_k f_{\mathbf{S}}(\mathbf{u}_j) g(\mathbf{u}_j, \mathbf{x}_j) \right| \\ = O \left(\sqrt{\frac{\log n}{nh^{md}}} \right) + o(h), \quad P_{\mathbf{S}} - a.s. \end{aligned} \quad (8.65)$$

where

$$\kappa_k = \int_{\mathbb{R}^d} \bar{K}^k(\mathbf{x}) d\mathbf{x}.$$

For any probability measure Q on a product measure space $(\Omega_1 \times \Omega_2, \Sigma_1 \times \Sigma_2)$, we may define the β -mixing coefficients as follows:

Definition 8.3. Yu, 1994, Definition 2.5 Let Q_1 and Q_2 be the marginal probability measures of Q on (Ω_1, Σ_1) and (Ω_2, Σ_2) , respectively. We set

$$\beta(\Sigma_1, \Sigma_2, Q) = P \sup \{ |Q(B | \Sigma_1) - Q_2(B)| : B \in \Sigma_2 \}.$$

The following lemma holds true for every finite n and is essential for the generation of independent blocks for β -mixing sequences.

Lemma 8.5. *Yu, 1994, Corollary 2.7* Let $m \in \mathbb{N}$ and let Q denote a probability measure on a product space $(\prod_{i=1}^m \Omega_i, \prod_{i=1}^m \Sigma_i)$ with the associated marginal measures Q_i on (Ω_i, Σ_i) . Assume that h is a bounded measurable function on the product probability space in such a way that $|h| \leq M_h < \infty$. For $1 \leq a \leq b \leq m$, let Q_a^b be the marginal measure on $(\prod_{i=a}^b \Omega_i, \prod_{i=a}^b \Sigma_i)$. For a given $\tau > 0$, suppose that, for all $1 \leq k \leq m-1$,

$$\|Q - Q_1^k \times Q_{k+1}^m\|_{TV} \leq 2\tau \quad (8.66)$$

where $Q_1^k \times Q_{k+1}^m$ is the product measure and $\|\cdot\|_{TV}$ is the total variation. Then

$$|Qh - Ph| \leq 2M_h(m-1)\tau,$$

where $P = \prod_{i=1}^m Q_i$, $Qh = \int h dQ$ and $Ph = \int h dP$.

Lemma 8.6. *Let*

$$\mathcal{I}_n = \{\mathbf{i} \in \mathbb{Z}^d : (\mathbf{i} + (0, 1]^d) \cap R_n \neq \emptyset\}.$$

Then we have

$$P_S \left(\sum_{j=1}^n \mathbf{1} \{A_n \mathbf{S}_{0,j} \in (\mathbf{i} + (0, 1]^d) \cap R_n\} > 2(\log n + nA_n^{-d}) \text{ for some } \mathbf{i} \in \mathcal{I}_n, \text{ i.o.} \right) = 0$$

and

$$P_S \left(\sum_{j=1}^n \mathbf{1} \{A_n \mathbf{S}_{0,j} \in \Gamma_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon})\} > CA_{1,n}^{q(\boldsymbol{\varepsilon})} A_{2,n}^{d-q(\boldsymbol{\varepsilon})} nA_n^{-d} \text{ for some } \boldsymbol{\ell} \in L_{1,n}, \text{ i.o.} \right) = 0$$

for any $\boldsymbol{\varepsilon} \in \{1, 2\}^d$, where "i.o." stands for infinitely often.

Proof. See the proof in Lahiri, 2003a, Lemma A.1 for each statement. \square

Remark 8.4. *Lemma 8.6 implies that each $\Gamma_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon})$ contains at most $CA_{1,n}^{q(\boldsymbol{\varepsilon})} A_{2,n}^{d-q(\boldsymbol{\varepsilon})} nA_n^{-d}$ samples P_S -almost surely.*

Lemma 8.7. *Under Assumptions 8, 9, Condition (B1) in 10, 11, 12, and 14, we have :*

$$E_{|\mathbf{S}} \left[\left(\bar{S}_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}) \right)^2 \right] \leq CA_{1,n}^{d-1} A_{2,n} (nA_n^{-d} + \log n) h^{md} \phi(h).$$

Proof of Lemma 8.7:

We have

$$E_{|\mathbf{S}} \left[\left(\bar{S}_n(\boldsymbol{\ell}; \boldsymbol{\varepsilon}) \right)^2 \right]$$

$$= \sum_{i:s_i \in \Gamma_n(\ell; \varepsilon) \cap R_n} E_{\cdot|\mathbf{S}} \left[\bar{S}_{s_i, A_n}^2(\mathbf{u}, \mathbf{x}) \right] + \sum_{i \neq j: s_i, s_j \in \Gamma_n(\ell; \varepsilon) \cap R_n} E_{\cdot|\mathbf{S}} \left[\bar{S}_{s_i, A_n}(\mathbf{u}, \mathbf{x}) \bar{S}_{s_j, A_n}(\mathbf{u}, \mathbf{x}) \right]$$

where

$$\begin{aligned} & \sum_{i:s_i \in \Gamma_n(\ell; \varepsilon) \cap R_n} E_{\cdot|\mathbf{S}} \left[\bar{S}_{s_i, A_n}^2(\mathbf{u}, \mathbf{x}) \right] \\ & \leq (n-1)^{-2m} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1}^2 \cdots \xi_{i_{\ell-1}}^2 \xi_{i_\ell}^2 \cdots \xi_{i_{m-1}}^2 \left(\int W_{s_{(1, \dots, \ell-1, \ell, \dots, m-1)}, A_n} \prod_{\substack{j=1 \\ j \neq i}}^{m-1} \frac{1}{\phi(h)} \right. \\ & \quad \left. K_2 \left(\frac{d(x_j, \nu_{s_j, A_n})}{h} \right) \mathbb{P}(d\nu_1, \dots, d\nu_{\ell-1}, d\nu_\ell, \dots, d\nu_{m-1}) \right)^2 \\ & \quad \left\{ E_{\cdot|\mathbf{S}} \left(\frac{1}{\phi^2(h)} K_2^2 \left(\frac{d(x_i, X_{s_i, A_n})}{h} \right) W_{s_i, A_n}^2 \right) \right. \\ & \quad \left. + \left[E_{\cdot|\mathbf{S}} \left(\frac{1}{\phi^2(h)} K_2 \left(\frac{d(x_i, X_{s_i, A_n})}{h} \right) W_{s_i, A_n} \right) \right]^2 \right\} \end{aligned} \quad (8.67)$$

$$\leq C \phi^2(h) (n-1)^{-2m} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1}^2 \cdots \xi_{i_{\ell-1}}^2 \xi_{i_\ell}^2 \cdots \xi_{i_{m-1}}^2 \quad P_{\mathbf{S}} - a.s. \quad (8.68)$$

Likewise, we can see that

$$\begin{aligned} & E_{\cdot|\mathbf{S}} \left[\bar{S}_{s_i, A_n}(\mathbf{u}, \mathbf{x}) \bar{S}_{s_j, A_n}(\mathbf{u}, \mathbf{x}) \right] \\ & \leq C \phi^2(h) (n-1)^{-2m} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1}^2 \cdots \xi_{i_{\ell-1}}^2 \xi_{i_\ell}^2 \cdots \xi_{i_{m-1}}^2 \quad P_{\mathbf{S}} - a.s. \end{aligned} \quad (8.69)$$

Applying Lemma 8.6 and Lemma 8.2 to find that

$$\begin{aligned} & \sum_{i:s_i \in \Gamma_n(\ell; \varepsilon) \cap R_n} \bar{K}_h^2 \left(u - \frac{s_j}{A_n} \right) \times (n-1)^{-2m} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1}^2 \cdots \xi_{i_{\ell-1}}^2 \xi_{i_\ell}^2 \cdots \xi_{i_{m-1}}^2 \\ & \leq C \sum_{i:s_i \in \Gamma_n(\ell; \varepsilon) \cap R_n} \bar{K}_h \left(u - \frac{s_i}{A_n} \right) \times (n-1)^{-2m} \sum_{I_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_{i_\ell} \cdots \xi_{i_{m-1}} \\ & \leq C h^{md} \mathbb{I}[\{i : s_i \in \Gamma_n(\ell; \varepsilon) \cap R_n\}] \\ & \leq C h^{md} A_{1,n}^{d-1} A_{2,n} \left(n A^{-d} + \log n \right), \quad P_{\mathbf{S}} - a.s., \end{aligned}$$

and

$$\sum_{i \neq j: s_i, s_j \in \Gamma_n(\ell; \varepsilon) \cap R_n} \bar{K}_h \left(u - \frac{s_i}{A_n} \right) \bar{K}_h \left(u - \frac{s_j}{A_n} \right)$$

$$\begin{aligned}
 & \times (n-1)^{-m} \left(\sum_{\Gamma_{n-1}^{m-1}(-i)} \sum_{\ell=1}^m \xi_{i_1} \cdots \xi_{i_{\ell-1}} \xi_{i_\ell} \cdots \xi_{i_{m-1}} \sum_{\Gamma_{n-1}^{m-1}(-j)} \sum_{\ell=1}^m \xi_{j_1} \cdots \xi_{j_{\ell-1}} \xi_{j_\ell} \cdots \xi_{j_{m-1}} \right) \\
 & \leq \left(\sum_{j: s_j \in \Gamma_n(\ell; \varepsilon) \cap R_n} \bar{K}_h \left(\mathbf{u} - \frac{s_j}{A_n} \right) (n-1)^{-m} \sum_{\Gamma_{n-1}^{m-1}(-j)} \sum_{\ell=1}^m \xi_{j_1} \cdots \xi_{j_{\ell-1}} \xi_{j_\ell} \cdots \xi_{j_{m-1}} \right)^2 \\
 & \leq Ch^{2md} [\{j : s_j \in \Gamma_n(\ell; \varepsilon) \cap R_n\}]^2 \\
 & \leq Ch^{2md} A_{1,n}^{2(d-1)} A_{2,n}^2 (nA^{-d} + \log n)^2, P_S - a.s.
 \end{aligned}$$

Since

$$A_{1,n}^{d-1} A_{2,n} (nA_n^{-d} + \log n) h^{md} \phi(h) \leq A_{1,n}^d (nA_n^{-d} + \log n) h^{md} \phi(h) = o(1),$$

we have

$$\begin{aligned}
 E_{\cdot|S} \left[\left(\bar{S}_n(\ell; \varepsilon) \right)^2 \right] & \leq C \left\{ A_{1,n}^{d-1} A_{2,n} (nA_n^{-d} + \log n) h^{md} \phi(h) \right. \\
 & \quad \left. + A_{1,n}^{2(d-1)} A_{2,n}^2 (n^2 A_n^{-2d} + \log^2 n) h^{2(d)} \phi^2(h) \right\} \\
 & \leq CA_{1,n}^{d-1} A_{2,n} (nA_n^{-d} + \log n) h^{md} \phi(h), P_S - a.s.
 \end{aligned}$$

Lemma 8.8 (Bernstein's inequality). *Let X_1, \dots, X_n be zero-mean independent random variables. Assume that*

$$\max_{1 \leq i \leq n} |X_i| \leq M < \infty, \quad a.s.$$

For all $t > 0$, we have

$$\mathbb{P} \left(\sum_{i=1}^n X_i \geq t \right) \leq \exp \left(- \frac{\frac{t^2}{2}}{\sum_{i=1}^n \mathbb{E} [X_i^2] + \frac{Mt}{3}} \right)$$

Lemma 8.9. *Under Assumptions 8,9, 10, and 12, we have*

$$\frac{1}{nh^{md} \phi^m(h)} \text{Var}_{\cdot|S} \left(\sum_{\ell \in L_{1,n}} Z_n(\ell; \varepsilon) \right) = o(1), P_S - a.s. \quad (8.70)$$

$$\frac{1}{nh^{md} \phi^m(h)} \text{Var}_{\cdot|S} \left(\sum_{\ell \in L_{2,n}} Z_n(\ell; \varepsilon) \right) = o(1), P_S - a.s. \quad (8.71)$$

Proof of Lemma 8.9

We have

$$\begin{aligned}
 \frac{1}{nh^{md}\phi(h)} \text{Var}_{|\mathcal{S}} \left(\sum_{\ell \in L_{1,n}} \mathbb{Z}_n(\ell; \varepsilon) \right) &= \frac{1}{nh^{md}\phi(h)} \sum_{\ell \in L_{1,n}} E_{|\mathcal{S}} \left[(\mathbb{Z}_n(\ell; \varepsilon))^2 \right] \\
 &\quad + \frac{1}{nh^{md}\phi(h)} \sum_{\ell_1 \neq \ell_2 \in L_{1,n}} E_{|\mathcal{S}} [\mathbb{Z}_n(\ell_1; \varepsilon) \mathbb{Z}_n(\ell_2; \varepsilon)] \\
 &:= I_1 + I_2.
 \end{aligned} \tag{8.72}$$

Using Lemma 8.7 and Assumption 10, it is easy to see that

$$\begin{aligned}
 I_1 &\leq C \frac{1}{nh^{md}\phi(h)} \left(\frac{A_n}{A_{1,n}} \right)^d A_{1,n}^{d-1} A_{2,n} (nA_n^{-d} + \log n) h^{md} \phi(h) \\
 &\leq C \frac{A_{2,n}}{A_{1,n}} (\log n) = o(1).
 \end{aligned} \tag{8.73}$$

For I_2 , using Rio, 2013, Theorem 1.1, we have:

$$\begin{aligned}
 &E_{|\mathcal{S}} [\mathbb{Z}_n(\ell_1; \varepsilon) \mathbb{Z}_n(\ell_2; \varepsilon)] \\
 &\leq E_{|\mathcal{S}} [|\mathbb{Z}_n(\ell_1; \varepsilon)|^3]^{1/3} E_{|\mathcal{S}} [|\mathbb{Z}_n(\ell_2; \varepsilon)|^3]^{1/3} \beta^{1/3} (d(\ell_1, \ell_2) A_{2,n}, A_{1,n}^{md}) \\
 &\leq E_{|\mathcal{S}} [|\mathbb{Z}_n(\ell_1; \varepsilon)|^3]^{1/3} E_{|\mathcal{S}} [|\mathbb{Z}_n(\ell_2; \varepsilon)|^3]^{1/3} \beta_1^{1/3} (d(\ell_1, \ell_2) A_{2,n}) g_1^{1/3} (A_{1,n}^{md}).
 \end{aligned}$$

The first inequality holds using Equation (2.5), and for $d(\ell_1, \ell_2) = \max_{1 \leq j \leq d} |\ell_1^j - \ell_2^j|$. Using the same strategy as Lemma 8.7, we have

$$E_{|\mathcal{S}} [|\mathbb{Z}_n(\ell_1; \varepsilon)|^3] \leq C A_{1,n}^{d-1} A_{2,n} (nA_n^{-d} + \log n) h^{md},$$

and

$$E_{|\mathcal{S}} [|\mathbb{Z}_n(\ell_2; \varepsilon)|^3] \leq C A_{1,n}^{d-1} A_{2,n} (nA_n^{-d} + \log n) h^{md}.$$

Note that for $\ell_1, \ell_2 \in L_{1,n}$, $\Gamma(\ell_1; \varepsilon_0)$ and $\Gamma(\ell_2; \varepsilon_0)$ in R_n are separated by the (ℓ^1-) distance

$$d(\Gamma(\ell_1; \varepsilon_0), \Gamma(\ell_2; \varepsilon_0)) \geq ((|\ell_1 - \ell_2| - d) + A_{3n}) + A_{2n}.$$

$$\begin{aligned}
 I_2 &\leq C \frac{\left(A_{1,n}^{d-1} A_{2,n} (nA_n^{-d} + \log n) h^{p+d} \right)^{2/3}}{nh^{d+p}} \\
 &\quad \times \sum_{\ell_1, \ell_2 \in L_{1,n}, \ell_1 \neq \ell_2} \beta_1^{1/3} ((|\ell_1 - \ell_2| - d) + A_{3n} + A_{2n}) g_1^{1/3} (A_{1,n}^d) \\
 &\leq C \left\{ \left(\frac{1}{nh^{d+p}} \right)^{1/3} \left(\frac{A_{1,n}}{A_n} \right)^{2d/3} \left(\frac{A_{2,n}}{A_{1,n}} \right)^{2/3} + \frac{A_{1,n}^{(d-1)/3} A_{2,n}^{1/3} (\log n)^{1/3}}{nh^{(d+p)/3}} \right\}
 \end{aligned}$$

$$\times g_1^{1/3} \left(A_{1,n}^d \right) \left\{ \beta_1^{1/3} (A_{2,n}) + \sum_{k=1}^{A_n/A_{1,n}} k^{d-1} \beta_1^{1/3} (kA_{3,n} + A_{2,n}) \right\} = o(1) \quad (8.74)$$

The last inequality follows using Assumption 10 and for

$$|\ell_1 - \ell_2| = \sum_{j=1}^d |\ell_{1,j} - \ell_{2,j}|.$$

Equation (8.71) could be treated similarly to (8.70).

Remark 8.5. *In order to prove that the summation over the small block is asymptotically negligible, we can use the method of Kurisu et al., 2021 where they used to pass from the dependence structure of variables to the independence as a first step, then they proved the convergence of second order expectation to zero using a maximal inequality. This method avoids the treatment of covariance, and it is based on the use of maximal inequality.*

Proposition 8.6. *Y. Song et al., 2019, Proposition 3.6 Let $\{X_i : i \in n\}$ be a process satisfying, for $m \geq 1$:*

$$\left(\mathbb{E} \|X_i - X_j\|^p \right)^{1/p} \leq \left(\frac{p-1}{q-1} \right)^{m/2} \left(\mathbb{E} \|X_i - X_j\|^q \right)^{1/q}, \quad 1 < q < p < \infty,$$

and the semi-metric :

$$\rho(j, i) = \left(\mathbb{E} \|X_i - X_j\|^2 \right)^{1/2}.$$

There exists a constant $K = K(m)$ such that :

$$\mathbb{E} \sup_{i,j \in n} \|X_i - X_j\| \leq K \int_0^D [\log N(\varepsilon, n, \rho)]^{m/2} d\varepsilon,$$

where D is the ρ -diameter of n .

Lemma 8.10. *(de la Peña, 1992) Let X_1, \dots, X_n be a sequence of independent random elements taking values in a Banach space $(B, \|\cdot\|)$ with $\mathbb{E}X_i = 0$ for all i . Let $\{\varepsilon_i\}$ be a sequence of independent Bernoulli r.v.'s independent of $\{X_i\}$. Then, for any convex increasing function Φ ,*

$$\mathbb{E} \Phi \left(\frac{1}{2} \left\| \sum_{i=1}^n X_i \varepsilon_i \right\| \right) \leq \mathbb{E} \Phi \left(\left\| \sum_{i=1}^n X_i \right\| \right) \leq \mathbb{E} \Phi \left(2 \left\| \sum_{i=1}^n X_i \varepsilon_i \right\| \right).$$

Chapter 8

Conclusion and perspectives

1 Concluding remarks : Chapter 3

In chapter 3, we have presented in this paper the concept of increasing degree U -statistics in order to treat parameters in high dimensional U -statistics. We attempted to demonstrate, via some examples, that this expansion may be relevant in a variety of scenarios. Unlike the literature, we studied this estimator within the framework of a Markov Harris chain. This chain is presented as a regenerative process due to its explicit atom or the artificial one using the splitting techniques. Infinite degree V -statistics, partial sum U -processes, symmetric statistics, and elementary symmetric polynomials share a common property with infinite-degree U -statistics. We used the empirical process theory to build a method to handle the increasing degree U -processes uniformly over kernel classes. More precisely, we develop theory and tools for studying U -processes, a natural higher-order generalization of the empirical processes. We also use decoupling theory to enable the symmetrization of dependent variables and U -statistics indexed by a class of functions. From different types of decoupling, we used the decoupling of the tangent sequence, where we adapted two sequences generated from the same filtration and with the same joint distribution. This technique allows us to use a new sequence with a more controllable structure than the original one, highlighting the advantage of decoupling by making relations between two processes where one of them has a more straightforward dependence structure. Arcones et al., 1993 used the decoupling results for U -process developed in de la Peña, 1992 as a critical tool for exploring the general theory of the U -processes and to obtain exponential and Bernstein type inequalities for the last statistics. It may be remarked that Hitczenko, 1988; Zinn, 1985 provided a solid groundwork for the idea of decoupling by presenting reasonably generic, easily applicable outcomes. Their findings include decoupling inequalities demonstrating that the L_p norms of two tangent processes are similar in some cases. Accessing other tools, we decided to use the com-

plete sign symmetrization techniques, provided by C. Heilig et al., 2001, adapted from Sherman, 1994, among the two techniques generated in their works: partial sign symmetrization and complete sign symmetrization technique. He introduced this technique for completely degenerate kernels of arbitrary finite degrees. The two previous techniques impose conditions on the growing rate of m_n . For reader convenience, we explain the difference between these techniques. To obtain the uniform weak law of large numbers in the partial sign symmetrization technique is satisfied by the condition

$$n^{-1/2}M_m \sum_{i=1}^m \mathbb{P}(\log \mathcal{N}(\varepsilon/m, d_{2n,i}, \mathcal{H}(h_m(i))))^{1/2} = o(1).$$

Using the complete sign-symmetrization technique, the condition is

$$n^{-1/2}mM_m \sum_{i=1}^m (2r_i - 1)^{i/2} 2^i \mathbb{P} \int_0^{1/4} \mathcal{N}(x, d_{2n,i}, \mathcal{H}(h_m(i)))^{1/2r_i} dx = o(1),$$

where $\{M_m\}$ is a deterministic sequence as our sequence of envelopes for the grand kernels h_m . However, suppose the grand permutation symmetrized grand kernel class is Euclidean. In that case, the partial sign symmetrization result is still valid, but the complete sign-symmetrization assumption assumes all the Hoeffding projection classes are Euclidean. In this case, the parameter V_m of the Euclidean class increases faster than V_i in the projected kernel classes. That is why m grows more quickly in the partial sign symmetrization technique than the complete sign-symmetrization does. In a future investigation, it will be of interest to present some sufficient conditions for the bootstrap of U -statistics (processes) of infinite degree over a stationary sequence of random variables satisfying an α -mixing condition of the weak dependence concept, which requires non trivial mathematics, that goes well beyond the scope of the present paper. In particular, we need to develop some new maximal moment inequalities. As we have mentioned in the introduction, a significant obstacle in using the IOUS is their computational intractability when the sample size and/or order are large. It will be interesting to derive non-asymptotic Gaussian approximation error bounds for an incomplete version of the IOUS with a random kernel in the Markov framework.

2 Concluding remarks: Chapter 4

Chapter 4 is concerned with the U -processes and the renewal type of bootstrap for recurrent Harris Markov chain under minimal moment conditions and entropy conditions for a uniformly class of functions. Three primary findings have been obtained. First, to avoid the use of mixing conditions, as we deal with markovian variables, we use the regenerative property of this chain to end up with independent blocks and then write a regenerative

estimator of our parameter of interest, which is directly connected to the key estimator, as provided in Proposition 3.4. We can notice that the convergence of the regenerative estimator in (3.5), or the regenerative U -statistic is of order $O((l_n - 1)^{-1/2})$ (see appendix in Chapter 4 for detail). The second main result concerns the weak convergence of the U -process in the Markov settings. The third main result considers the weak convergence of the renewal bootstrap of the U -process under minimum conditions over the class of function. We generalize our result to the non-trivial setting of the k -Markov chains. Considering the subsampling procedures for the U -processes in the Markov chains would be interesting. In contrast to the bootstrap, the subsampling procedures need minimal conditions for asymptotic validity. Another problem to be studied in the future is the characterization of the asymptotic properties of the censored U -process for the Markov chains that require non-trivial mathematics.

3 Concluding remarks: Chapter 5

Chapter 5 is concerned with the randomly weighted bootstrap of the U -process in a Markov framework. A large number of bootstrap resampling schemes emerge as special cases of our setting, in particular, the multinomial bootstrap, which is the most known bootstrap scheme introduced by Efron, 1979. One of the main tools is approximating the Markov U -process by the corresponding regenerative one. We looked to mimic this result in Proposition 3.4, in order to approximate the weighted bootstrap U -process U_n^* to the regenerative weighted bootstrap U -process $R_{l_n}^*$. Other technical arguments are given in Lemma 3.1 extended from the work of Radulović, 2004. These intricate tools are used to reach full independence of regenerative block variables by proving that a deterministic one can substitute the random size of blocks, which is the main problem for the extension of the bootstrap results to the Markov framework. After the lengthy proof to arrive at independence, we use the results of Q. Han, 2022. All the above steps conduct us to prove the weak convergence of regenerative block weighted bootstrap U -process, which implies the weak convergence of weighted bootstrap U -process. We would like to consider the extension of the work to the semi-Markov setting. The more delicate problem is considering the setting of incomplete data, such as censored cases or missing data. To the best of our knowledge, this problem is not considered even for the original sample (without bootstrap) in the Markov framework. It would be interesting to extend our work to the case of the local stationary process, which requires non-trivial mathematics; this would go well beyond the scope of the present chapter.

4 Concluding remarks: Chapter 6

In this chapter, we developed the theory of U -processes for locally stationary variables with a functional data framework. The primary objective is to use functional local stationary approximations to perform an asymptotic analysis for statistical inference of non-stationary time series. We highlight that using absolutely regular conditions or the β -mixing conditions is essential because it does not depend on the entropy dimension of the class as the other type of mixing. Also, β -mixing is much easier to use than strong mixing because it allows decoupling and covers interesting examples. By the way, we note that Ibragimov et al., 1969 obtained a complete description of stationary Gaussian processes satisfying the β -mixing condition. In addition, the β -mixing coincides with the $L_2(\mathbb{P})$ -norm, which plays an important role, while the use of strong mixing, the class of functions with a finite dimension of entropy, requires some polynomial rate of decay of strong mixing coefficients, and this rate depends on the entropy dimension of the class of functions. Also in the strong mixing, the L_1 -norm is involved, where the metric entropy function $H(\cdot, T, d)$ will be defined with respect to the pseudo-metric $d(s, t) = \sqrt{\text{Var}(\mathbb{G}(s) - \mathbb{G}(t))}$, for a Gaussian process $\mathbb{G}(\cdot)$, satisfies the integrability condition

$$\int_0^1 \sqrt{H(u, T, d)} du < +\infty.$$

As a result, we have stated the rate of convergence, where we show that, under suitable conditions, the kernel estimator $\tilde{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}; h_n)$ constructed with the bandwidth h converges to the regression operator $r^{(m)}(\varphi, \mathbf{u}, \mathbf{x})$ with a rate:

$$O_{\mathbb{P}} \left(\sqrt{\frac{\log n}{nh^m \phi^m(h)}} + h^{2m \wedge \alpha} \right),$$

this rate effectively shows the role played by the small ball probability function, which appeared in the first term, and is directly related to the concentration of the functional variables X_i . The second term is related to the bias of the estimate. It depends on the smoothness of the operator $r^{(m)}(\varphi, \mathbf{u}, \mathbf{x})$, represented by the *Lipschitz* condition and its parameter α . It is important to point out that the concentration of functional variables X , the small ball probability, and the convergence rate are proportional, i.e., the estimator will be more efficient if the variables are less dispersed and the small ball probability will be higher. As the empirical process settings, the rate of convergence is valid over a subset $[Ch, Ch - 1]^m \times \{\mathbf{x}\}^m$, but it can be extended for a subset $[Ch - 1, 1]^m \times \{\mathbf{x}\}^m$, for forecasting purposes. It is possible to use one-sided kernels or boundary-corrected kernels to reach the latter goal, but both under conditions that kernels have compact support and are Lipschitz.

The weak convergence follows the classical steps: the finite-dimensional convergence and the equicontinuity of the conditional U -processes. The finite-dimensional convergence led using Hoeffding decomposition, followed by approaching independence via a block decomposition strategy and then proving a central limit theorem for independent variables. The equicontinuity requires complete oversight, and the contents are very technical and lengthy due to the general and comprehensive framework we have considered, see the next section. Extending non-parametric functional principles to a generic dependent structure is a relatively new area of research. Shedding light on the importance that mixing is a type of asymptotic independence assumption currently adopted in the pursuit of simplicity but might be unsuitable in scenarios when the data is highly dependent. It is worth noting that the ergodic framework eliminates the frequently used strong mixing condition, its variations for measuring dependence, and the more complicated probabilistic computations they imply. It would be of interest to investigate the following kNN estimator:

$$\hat{r}_n^{(m)}(\varphi, \mathbf{u}, \mathbf{x}; h_n) = \frac{\sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d(x_k, X_{i_k, n})}{H_{n,k}(x_k)} \right) \right\} \varphi(Y_{\mathbf{i}, n})}{\sum_{\mathbf{i} \in I_n^m} \prod_{k=1}^m \left\{ K_1 \left(\frac{u_k - i_k/n}{h_n} \right) K_2 \left(\frac{d(x_k, X_{i_k, n})}{H_{n,k}(x_k)} \right) \right\}},$$

where

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

where

$$B(t, h) = \{z \in \mathcal{H} : d(z, t) \leq h\}$$

is a ball in \mathcal{H} with the center $t \in \mathcal{H}$ and radius h , and $\mathbb{1}_A$ is the indicator function of the set A .

5 Concluding remarks: chapter 7

In Chapter 7, we considered the kernel type estimator for conditional U -statistics, including a particular case, the Nadaraya-Watson estimator, in a functional setting with random fields. To obtain our results, we ought to make assumptions requiring some regularity on the conditional U -statistics and conditional moments, some decay rates on the probability of the variables belonging to shrinking open balls, and suitable decreasing rates on the mixing coefficients. Mainly, the conditional moment assumption enables the consideration of unbounded classes of functions. The proof of the weak convergence respects a typical technique: finite dimensional convergence and equicontinuity of the conditional U -processes.

Both results, the uniform rate of convergence and the weak convergence, are grounded on a general blocking technique adjusted for irregularly spaced sampling sites, where we need to pay attention to the effect of the non-equidistant sampling sites. We intricately reduce the work to the independent setting to address this issue. Indeed, as there is no practical guidance for introducing order to spatial points as opposed to time series, not asymptotically but exactly independent blocks of observations have been constructed by (Yu, 1994, Corollary 2.7) (Lemma 8.5) and then results of independent data could be applied directly to the independent blocks. Here, Ref. Yu, 1994 declares that the uniform convergence result requires the β -mixing condition to connect the original sequence with the sequence of the independent blocks, and this connection still holds under the ϕ -mixing condition but is not necessary under the α -mixing conditions. Therefore, we use the β -mixing sequence as we aim to derive the weak convergence for processes indexed by classes of functions.

Kurusu, 2022b in his work gives us a possible extension of the sampling region inspired by Lahiri, 2003a. This extension can be explained as follows. It is feasible to generalize the definition of the sample region R_n to include non-standard forms. For instance, we may use the sample region concept Lahiri, 2003a as follows: First, let R_n be the sampling region. Define R_0^* as an open connected subset of $(-2, 2]^d$ containing $[-1, 1]^d$ and R_0 as a Borel set such that $R_0^* \subset R_0 \subset \bar{R}_0^*$, and where for any set $S \subset \mathbb{R}^d$, \bar{S} signifies its closure. Let $\{A_n\}_{n \geq 1}$ be a sequence of positive numbers such that $A_n \rightarrow \infty$ as $n \rightarrow \infty$ and define $R_n = A_n R_0$ as a sampling region. In addition, for any sequence of positive numbers $\{a_n\}_{n \geq 1}$ with $a_n \rightarrow 0$ as $n \rightarrow \infty$, let $O(a_n^{-d+1})$, as $n \rightarrow \infty$, be the number of cubes of the form $a_n(\mathbf{i} + [0, 1]^d)$, $\mathbf{i} \in \mathbb{Z}^d$ with their lower left corner $a_n \mathbf{i}$ on the lattice $a_n \mathbb{Z}^d$ that intersects both R_0 and R_0^c (see Condition B in Lahiri, 2003a, Chapter 12, Section 12.2) (This condition is the prototype R_0 boundary's condition; it must always be assumed on the region R_n to prevent pathological situations, and it is satisfied by the majority of areas of practical significance. This condition is satisfied in the plane ($d = 2$), for instance, if the boundary ∂R_0 of R_0 is defined by a simple rectifiable curve of limited length. When sample sites are defined on the integer grid \mathbb{Z}^d , this condition means that the effect of data points toward the boundary of R_n is small compared to the overall number of data points). In addition, define f as a continuous, everywhere positive probability density function on R_0 , and let $\{S_{0,i}\}_{i \geq 1}$ be a sequence of i.i.d. random vectors with density f . Assume that $\{S_{0,i}\}_{i \geq 1}$ and \mathbf{X}_{s,A_n} are independent. Replacing our setting in Section 2.4 with this new one, our results still hold, and it will be possible to show uniform convergence and weak convergence under the same assumptions and identical proofs. For future investigation, it will be interesting to relax the mixing conditions to the weak dependence (or the ergodicity framework). This generalization is nontrivial, since we need some maximal moment inequalities in our asymptotic results that are not

available in this setting. Another interesting direction is to consider the incomplete data setting (missing at random, censored in different schemes) for locally spatial-functional data. A natural question is how to adapt our results to the wavelet-based estimators, the delta sequence estimators, the k NN estimators, and the local linear estimators.

Bibliography

- Abrevaya, J., & Jiang, W. (2005). A nonparametric approach to measuring and testing curvature. *J. Bus. Econom. Statist.*, *23*(1), 1–19 (cit. on pp. [25](#), [95](#), [154](#), [200](#), [273](#)).
- Adamczak, R. (2006). Moment inequalities for U -statistics. *Ann. Probab.*, *34*(6), 2288–2314 (cit. on p. [4](#)).
- Adamczak, R. (2008). A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.*, *13*, no. 34, 1000–1034 (cit. on p. [150](#)).
- Adler, R. J., & Taylor, J. E. (2007). *Random fields and geometry*. Springer, New York. (Cit. on p. [51](#)).
- Akaike, H. (1954). An approximation to the density function. *Ann. Inst. Statist. Math., Tokyo*, *6*, 127–132 (cit. on p. [201](#)).
- Aldous, D. J. (1985). Exchangeability and related topics. *École d'été de probabilités de Saint-Flour, XIII—1983* (pp. 1–198). Springer, Berlin. (Cit. on p. [48](#)).
- Almanjahie, I. M., Bouzebda, S., Chikr Elmezouar, Z., & 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 (cit. on pp. [204](#), [274](#)).
- Almanjahie, I. M., Bouzebda, S., Kaid, Z., & Laksaci, A. (2022a). Nonparametric estimation of expectile regression in functional dependent data. *J. Nonparametr. Stat.*, *34*(1), 250–281 (cit. on p. [203](#)).
- Almanjahie, I. M., Bouzebda, S., Kaid, Z., & Laksaci, A. (2022b). Nonparametric estimation of expectile regression in functional dependent data. *J. Nonparametr. Stat.*, *34*(1), 250–281 (cit. on p. [274](#)).
- Alvarez-Andrade, S., & Bouzebda, S. (2013). Strong approximations for weighted bootstrap of empirical and quantile processes with applications. *Stat. Methodol.*, *11*, 36–52 (cit. on pp. [76](#), [155](#)).
- Alvarez-Andrade, S., & Bouzebda, S. (2020). Cramér's type results for some bootstrapped U -statistics. *Statist. Papers*, *61*(4), 1685–1699 (cit. on p. [155](#)).
- Anderson, R., & Sandsten, M. (2019). Inference for time-varying signals using locally stationary processes. *J. Comput. Appl. Math.*, *347*, 24–35 (cit. on p. [35](#)).
- Andrews, D. W. K. (1984). Nonstrong mixing autoregressive processes. *J. Appl. Probab.*, *21*(4), 930–934 (cit. on p. [16](#)).
- Andrews, D. W. K. (2005). Higher-order improvements of the parametric bootstrap for Markov processes. *Identification and inference for econometric models* (pp. 171–215). Cambridge Univ. Press, Cambridge. (Cit. on p. [15](#)).
- Aneiros, G., Cao, R., Fraiman, R., Genest, C., & Vieu, P. (2019). Recent advances in functional data analysis and high-dimensional statistics. *J. Multivariate Anal.*, *170*, 3–9 (cit. on pp. [203](#), [273](#)).

- Araujo, A., & Giné, E. (1980). *The central limit theorem for real and Banach valued random variables*. John Wiley & Sons, New York-Chichester-Brisbane. (Cit. on pp. 203, 274).
- Arcones, M. A., & Yu, B. (1994). Central limit theorems for empirical and U -processes of stationary mixing sequences. *J. Theoret. Probab.*, 7(1), 47–71 (cit. on pp. 4, 36, 37, 230, 241, 244, 261, 262, 276, 316, 328, 331, 340, 341).
- Arcones, M. A. (1998). The law of large numbers for U -statistics under absolute regularity. *Electron. Comm. Probab.*, 3, 13–19 (cit. on p. 252).
- Arcones, M. A., Chen, Z., & Giné, E. (1994). Estimators related to U -processes with applications to multivariate medians: Asymptotic normality. *Ann. Statist.*, 22(3), 1460–1477 (cit. on pp. 96, 200).
- Arcones, M. A., & Giné, E. (1992). On the bootstrap of U and V statistics. *Ann. Statist.*, 20(2), 655–674 (cit. on pp. 56, 113, 131, 133, 135, 154, 177).
- Arcones, M. A., & Giné, E. (1993). Limit theorems for U -processes. *Ann. Probab.*, 21(3), 1494–1542 (cit. on pp. 4, 55, 83, 95, 96, 101, 129, 133, 153–155, 200, 234, 268, 273, 352).
- Arcones, M. A., & Giné, E. (1994). U -processes indexed by Vapnik-červonenkis classes of functions with applications to asymptotics and bootstrap of U -statistics with estimated parameters. *Stochastic Process. Appl.*, 52(1), 17–38 (cit. on pp. 97, 113, 135).
- Arcones, M. A., & Giné, E. (1995). On the law of the iterated logarithm for canonical U -statistics and processes. *Stochastic Process. Appl.*, 58(2), 217–245 (cit. on pp. 4, 96, 200).
- Arcones, M. A., & Wang, Y. (2006). Some new tests for normality based on U -processes. *Statist. Probab. Lett.*, 76(1), 69–82 (cit. on pp. 95, 153, 273).
- Aronszajn, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68, 337–404 (cit. on p. 7).
- Asmussen, S. (1987). *Applied probability and queues*. John Wiley & Sons, Ltd., Chichester. (Cit. on p. 58).
- Athreya, K. B., & Fuh, C. D. (1992a). Bootstrapping Markov chains. *Exploring the limits of bootstrap (East Lansing, MI, 1990)* (pp. 49–64). Wiley, New York. (Cit. on p. 57).
- Athreya, K. B., & Fuh, C. D. (1992b). Bootstrapping Markov chains: Countable case. *J. Statist. Plann. Inference*, 33(3), 311–331 (cit. on pp. 15, 57).
- Athreya, K. B., & Ney, P. (1978). A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.*, 245, 493–501 (cit. on pp. 69, 100, 158).
- Attouch, M., Laksaci, A., & Ráfaa, F. (2019). On the local linear estimate for functional regression: Uniform in bandwidth consistency. *Comm. Statist. Theory Methods*, 48(8), 1836–1853 (cit. on p. 274).
- Bandyopadhyay, S., Lahiri, S. N., & Nordman, D. J. (2015). A frequency domain empirical likelihood method for irregularly spaced spatial data. *Ann. Statist.*, 43(2), 519–545 (cit. on p. 280).
- Barbe, P., & Bertail, P. (1995). *The weighted bootstrap* (Vol. 98). Springer-Verlag, New York. (Cit. on pp. 15, 48).
- Bardet, J.-M., & Doukhan, P. (2018). Non-parametric estimation of time varying AR(1)-processes with local stationarity and periodicity. *Electron. J. Stat.*, 12(2), 2323–2354 (cit. on p. 35).
- Basawa, I. V., Green, T. A., McCormick, W. P., & Taylor, R. L. (1990). Asymptotic bootstrap validity for finite Markov chains. *Comm. Statist. Theory Methods*, 19(4), 1493–1510 (cit. on p. 15).

- Basu, A. K., & Kundu, A. (2002). Limit distribution for conditional U -statistics for dependent processes. *Proceedings of the Fourth International Triennial Calcutta Symposium on Probability and Statistics (2000)*, 52(205-208), 381–407 (cit. on p. 202).
- Bellet, A., & Habrard, A. (2015). Robustness and generalization for metric learning. *Neurocomputing*, 151, 259–267 (cit. on pp. 221, 296).
- Bellet, A., Habrard, A., & Sebban, M. (2013). A survey on metric learning for feature vectors and structured data. *CoRR*, abs/1306.6709 (cit. on pp. 220, 295).
- Benhenni, K., Ferraty, F., Rachdi, M., & Vieu, P. (2007). Local smoothing regression with functional data. *Comput. Statist.*, 22(3), 353–369 (cit. on p. 304).
- Bergsma, W., & Dassios, A. (2014). A consistent test of independence based on a sign covariance related to Kendall’s tau. *Bernoulli*, 20(2), 1006–1028 (cit. on p. 3).
- Berk, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *Ann. Math. Statist.*, 37, 51–58, correction, *ibid.* 745–746 (cit. on p. 4).
- Bernstein, S. (1927). Sur l’extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Math. Ann.*, 97(1), 1–59 (cit. on pp. 16, 230, 316).
- Bertail, P., & Cléménçon, S. (2011a). A renewal approach to Markovian U -statistics. *Math. Methods Statist.*, 20(2), 79–105 (cit. on pp. 15, 30, 57, 78, 97, 108, 112, 123).
- Bertail, P., & Cléménçon, S. (2011b). A renewal approach to Markovian U -statistics. *Math. Methods Statist.*, 20(2), 79–105 (cit. on pp. 106, 163).
- Bertail, P., & Cléménçon, S. (2004). Edgeworth expansions of suitably normalized sample mean statistics for atomic Markov chains. *Probab. Theory Related Fields*, 130(3), 388–414 (cit. on p. 70).
- Bertail, P., & Cléménçon, S. (2006a). Regeneration-based statistics for Harris recurrent Markov chains. *Dependence in probability and statistics* (pp. 3–54). Springer, New York. (Cit. on pp. 106, 163).
- Bertail, P., & Cléménçon, S. (2006b). Regenerative block bootstrap for Markov chains. *Bernoulli*, 12(4), 689–712 (cit. on pp. 15, 120, 137, 147).
- Bertail, P., & Cléménçon, S. (2006c). Regenerative block bootstrap for Markov chains. *Bernoulli*, 12(4), 689–712 (cit. on pp. 67, 69).
- Bertail, P., & Cléménçon, S. (2008). Approximate regenerative-block bootstrap for Markov chains. *Comput. Statist. Data Anal.*, 52(5), 2739–2756 (cit. on pp. 69, 70).
- Bertoin, J. (1996). *Lévy processes* (Vol. 121). Cambridge University Press, Cambridge. (Cit. on p. 26).
- Biau, G., & Bleakley, K. (2006). Statistical inference on graphs. *Statist. Decisions*, 24(2), 209–232 (cit. on p. 4).
- Biau, G., & Cadre, B. (2004). Nonparametric spatial prediction. *Stat. Inference Stoch. Process.*, 7(3), 327–349 (cit. on pp. 36, 272).
- Bickel, P. J., Götze, F., & van Zwet, W. R. (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statist. Sinica*, 7(1), 1–31 (cit. on p. 56).
- Bickel, P. J., & Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.*, 9(6), 1196–1217 (cit. on p. 154).
- Bickel, P. J., & Levina, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.*, 36(1), 199–227 (cit. on p. 2).

- Billingsley, P. (1968). *Convergence of probability measures*. John Wiley & Sons, Inc., New York-London-Sydney. (Cit. on p. 48).
- Birr, S., Volgushev, S., Kley, T., Dette, H., & Hallin, M. (2017). Quantile spectral analysis for locally stationary time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79(5), 1619–1643 (cit. on p. 18).
- Blum, J. R., Kiefer, J., & Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.*, 32, 485–498 (cit. on p. 3).
- Blum, J., Hanson, D. L., & Koopmans, L. H. (1963). *On the strong law of large numbers for a class of stochastic processes*. Sandia Corporation. (Cit. on p. 16).
- Blundell, R., Gosling, A., Ichimura, H., & Meghir, C. (2007). Changes in the distribution of male and female wage accounting for employment composition using bounds. *Econometrica*, 75(2), 323–363 (cit. on p. 24).
- Bogachev, V. I. (1998). *Gaussian measures* (Vol. 62). American Mathematical Society, Providence, RI. (Cit. on pp. 213, 288).
- Bolthausen, E. (1980). The Berry-Esseen theorem for functionals of discrete Markov chains. *Z. Wahrsch. Verw. Gebiete*, 54(1), 59–73 (cit. on p. 110).
- Bolthausen, E. (1982). The Berry-Esseen theorem for strongly mixing Harris recurrent Markov chains. *Z. Wahrsch. Verw. Gebiete*, 60(3), 283–289 (cit. on pp. 67, 110).
- Borovkova, S., Burton, R., & Dehling, H. (2002). From dimension estimation to asymptotics of dependent U -statistics. *Limit theorems in probability and statistics, Vol. I (Balatonlelle, 1999)* (pp. 201–234). János Bolyai Math. Soc., Budapest. (Cit. on p. 57).
- Borovkova, S., Burton, R., & 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 (cit. on pp. 57, 96, 154, 200, 273).
- Borovskikh, Y. V. (1986). The theory of U -statistics in a Hilbert space. *Akad. Nauk Ukrain. SSR Inst. Mat. Preprint*, (78), 56 (cit. on p. 4).
- Borovskikh, Y. V. (1996). *U -statistics in Banach spaces*. VSP, Utrecht. (Cit. on pp. 96, 200, 273).
- Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions. *Ann. Statist.*, 16(4), 1709–1722 (cit. on p. 56).
- Bose, A., & Chatterjee, S. (2018). *U -statistics, mm-estimators and resampling*. Springer. (Cit. on p. 105).
- Bosq, D. (2000). *Linear processes in function spaces* (Vol. 149). Springer-Verlag, New York. (Cit. on pp. 21, 203, 274).
- Bosq, D. (1991). Modelization, nonparametric estimation and prediction for continuous time processes. *Nonparametric functional estimation and related topics (Spetses, 1990)* (pp. 509–529). Kluwer Acad. Publ., Dordrecht. (Cit. on p. 20).
- Bouzebda, S. (2012). On the strong approximation of bootstrapped empirical copula processes with applications. *Math. Methods Statist.*, 21(3), 153–188 (cit. on p. 155).
- Bouzebda, S. (2016). Some applications of the strong approximation of the integrated empirical copula processes. *Math. Methods Statist.*, 25(4), 281–303 (cit. on pp. 203, 274).
- Bouzebda, S., & Nemouchi, B. (2019a). Central limit theorems for conditional empirical and conditional U -processes of stationary mixing sequences. *Math. Methods Statist.*, 28(3), 169–207 (cit. on pp. 55, 276).

- Bouzebda, S., & Soukarieh, I. (2022a). Renewal type bootstrap for U -process markov chains. *Markov Process. Related Fields*, 1–52 (cit. on pp. 55, 57, 58, 78, 273).
- Bouzebda, S., & Soukarieh, I. (2022b). Weak convergence of the conditional U -statistics for locally stationary functional time series. *Submitted* (cit. on p. 340).
- Bouzebda, S., & 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 (cit. on pp. 204, 274).
- Bouzebda, S., & El-hadjali, T. (2020a). Uniform convergence rate of the kernel regression estimator adaptive to intrinsic dimension in presence of censored data. *J. Nonparametr. Stat.*, 32(4), 864–914 (cit. on p. 76).
- Bouzebda, S., & El-hadjali, T. (2020b). Uniform convergence rate of the kernel regression estimator adaptive to intrinsic dimension in presence of censored data. *J. Nonparametr. Stat.*, 32(4), 864–914 (cit. on pp. 302, 303).
- Bouzebda, S., El-hadjali, T., & Ferfache, A. A. (2022). Uniform in bandwidth consistency of conditional U -statistics adaptive to intrinsic dimension in presence of censored data. *Sankhya A*, 1–59 (cit. on p. 303).
- Bouzebda, S., Elhattab, I., & Ferfache, A. A. (2022). General m -estimator processes and their m out of n bootstrap with functional nuisance parameters. *Methodol. Comput. Appl. Probab.*, 1–45 (cit. on pp. 56, 155).
- Bouzebda, S., Elhattab, I., & 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 (cit. on pp. 55, 202, 276).
- Bouzebda, S., Mohammedi, M., & Laksaci, A. (2022). The k -nearest neighbors method in single index regression model for functional quasi-associated time series data. *Rev. Mat. Complut.*, 1–30 (cit. on p. 274).
- Bouzebda, S., & Nemouchi, B. (2019b). Central limit theorems for conditional empirical and conditional U -processes of stationary mixing sequences. *Math. Methods Statist.*, 28(3), 169–207 (cit. on pp. 36, 37, 96, 200, 261, 262, 341).
- Bouzebda, S., & 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 (cit. on pp. 55, 96, 200, 202, 276).
- Bouzebda, S., & Nemouchi, B. (2022). Weak-convergence of empirical conditional processes and conditional U -processes involving functional mixing data. *Stat. Inference Stoch. Process. To appear*, 1–60 (cit. on pp. 36, 37, 55, 154, 273, 276, 288).
- Bouzebda, S., & 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.*, To appear, 1–103 (cit. on pp. 55, 274, 303).
- Bouzebda, S., Nezzal, A., & Zari, T. (2022). Uniform consistency for functional conditional U -statistics using delta-sequences. *Mathematics*, 10(24), 1–38 (cit. on pp. 273, 274).
- Bradley, R. C. (1989). A caution on mixing conditions for random fields. *Statist. Probab. Lett.*, 8(5), 489–491 (cit. on p. 280).

- Bradley, R. C. (1993). Some examples of mixing random fields. *Rocky Mountain J. Math.*, *23*(2), 495–519 (cit. on p. 280).
- Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.*, *2*, 107–144 (cit. on p. 49).
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140 (cit. on p. 55).
- Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32 (cit. on p. 55).
- Brillinger, D. R. (1981). *Time series* (Second). Holden-Day, Inc., Oakland, Calif. (Cit. on p. 16).
- Brockwell, P. J., & Matsuda, Y. (2017). Continuous auto-regressive moving average random fields on \mathbb{R}^n . *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *79*(3), 833–857 (cit. on pp. 26, 38).
- Brunel, E., & Comte, F. (2006). Adaptive nonparametric regression estimation in presence of right censoring. *Math. Methods Statist.*, *15*(3), 233–255 (cit. on pp. 300, 302).
- Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli*, *3*(2), 123–148 (cit. on p. 15).
- Bühlmann, P. (2002). Sieve bootstrap with variable-length Markov chains for stationary categorical time series. *J. Amer. Statist. Assoc.*, *97*(458), 443–471 (cit. on pp. 14, 15).
- Bühlmann, P., & van de Geer, S. (2011). *Statistics for high-dimensional data*. Springer, Heidelberg. (Cit. on p. 1).
- Cao, Q., Guo, Z.-C., & Ying, Y. (2016). Generalization bounds for metric and similarity learning. *Mach. Learn.*, *102*(1), 115–132 (cit. on pp. 200, 221, 296).
- Carbonez, A., Györfi, L., & van der Meulen, E. C. (1995). Partitioning-estimates of a regression function under random censoring. *Statist. Decisions*, *13*(1), 21–37 (cit. on p. 300).
- Carlstein, E. (1986a). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.*, *14*(3), 1171–1179 (cit. on p. 56).
- Carlstein, E. (1986b). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.*, *14*(3), 1171–1179 (cit. on p. 154).
- Chatterjee, S., & Bose, A. (2005). Generalized bootstrap for estimating equations. *Ann. Statist.*, *33*(1), 414–436 (cit. on p. 16).
- Chen, X. (2018). Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications. *Ann. Statist.*, *46*(2), 642–678 (cit. on p. 97).
- Chen, X., & 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 (cit. on pp. 24, 25).
- Chen, X., Xu, M., & Wu, W. B. (2013). Covariance and precision matrix estimation for high-dimensional time series. *Ann. Statist.*, *41*(6), 2994–3021 (cit. on p. 2).
- Chen, Y., & Datta, S. (2019). Adjustments of multi-sample U -statistics to right censored data and confounding covariates. *Comput. Statist. Data Anal.*, *135*, 1–14 (cit. on p. 302).
- Cheng, G. (2015). Moment consistency of the exchangeably weighted bootstrap for semiparametric M -estimation. *Scand. J. Stat.*, *42*(3), 665–684 (cit. on p. 171).
- Cheng, G., & Huang, J. Z. (2010). Bootstrap consistency for general semiparametric M -estimation. *Ann. Statist.*, *38*(5), 2884–2915 (cit. on p. 48).

- Christakos, G. (2012). *Random field models in earth sciences*. Courier Corporation. (Cit. on p. 51).
- Chung, K. L. (1967). *Markov chains with stationary transition probabilities* (Second). Springer-Verlag New York, Inc., New York. (Cit. on pp. 71, 81, 100, 158).
- Chwialkowski, K., Strathmann, H., & Gretton, A. (2016). A kernel test of goodness of fit. *International conference on machine learning*, 2606–2615 (cit. on pp. 10, 11).
- Ciolek, G. (2016). Bootstrap uniform central limit theorems for Harris recurrent Markov chains. *Electron. J. Stat.*, 10(2), 2157–2178 (cit. on pp. 97, 120).
- Cléménçon, S. (2014). A statistical view of clustering performance through the theory of U -processes. *J. Multivariate Anal.*, 124, 42–56 (cit. on pp. 4, 24, 200).
- Cléménçon, S. (n.d.). On U -processes and clustering performance. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems*. Curran Associates, Inc. (Cit. on p. 273).
- Cléménçon, S., Colin, I., & Bellet, A. (2016). Scaling-up empirical risk minimization: Optimization of incomplete U -statistics. *J. Mach. Learn. Res.*, 17, Paper No. 76, 36 (cit. on pp. 24, 123, 200, 220, 295, 296).
- Cléménçon, S., Lugosi, G., & Vayatis, N. (2008). Ranking and empirical minimization of U -statistics. *Ann. Statist.*, 36(2), 844–874 (cit. on pp. 4, 96, 200).
- Cléménçon, S., & Robbiano, S. (2015). The TreeRank Tournament algorithm for multipartite ranking. *J. Nonparametr. Stat.*, 27(1), 107–126 (cit. on p. 296).
- Cléménçon, S., Robbiano, S., & Vayatis, N. (2013). Ranking data with ordinal labels: Optimality and pairwise aggregation. *Mach. Learn.*, 91(1), 67–104 (cit. on p. 296).
- Cléménçon, S., & Slim, S. (2004). Statistical analysis of financial time series under the assumption of local stationarity. *Quant. Finance*, 4(2), 208–220 (cit. on p. 35).
- Cressie, N. A. C. (1993). *Statistics for spatial data*. John Wiley & Sons, Inc., New York. (Cit. on p. 289).
- Cressie, N. A. C. (2015). *Statistics for spatial data* (Revised). John Wiley & Sons, Inc., New York. (Cit. on pp. 36, 272).
- Csiszár, I., & Shields, P. C. (2000). The consistency of the BIC Markov order estimator. *Ann. Statist.*, 28(6), 1601–1619 (cit. on p. 15).
- Csörgő, M., & Nasari, M. M. (2013). Asymptotics of randomly weighted u - and v -statistics: Application to bootstrap. *J. Multivariate Anal.*, 121, 176–192 (cit. on p. 97).
- Cybis, G. B., Valk, M., & Lopes, S. R. C. (2018). Clustering and classification problems in genetics through U -statistics. *J. Stat. Comput. Simul.*, 88(10), 1882–1902 (cit. on p. 200).
- Dabo-Niang, S., & Yao, A.-F. (2013). Kernel spatial density estimation in infinite dimension space. *Metrika*, 76(1), 19–52 (cit. on pp. 36, 272).
- Dahlhaus, R. (1996a). On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Process. Appl.*, 62(1), 139–168 (cit. on p. 18).
- Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.*, 25(1), 1–37 (cit. on pp. 18, 34, 203, 277, 278).
- Dahlhaus, R. (1996b). On the kullback-leibler information divergence of locally stationary processes. *Stochastic processes and their applications*, 62(1), 139–168 (cit. on p. 206).

- Dahlhaus, R., & Polonik, W. (2006). Nonparametric quasi-maximum likelihood estimation for Gaussian locally stationary processes. *Ann. Statist.*, *34*(6), 2790–2824 (cit. on pp. 18, 34, 203).
- Dahlhaus, R., & Polonik, W. (2009). Empirical spectral processes for locally stationary time series. *Bernoulli*, *15*(1), 1–39 (cit. on pp. 34, 203).
- Dahlhaus, R., & Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes. *Ann. Statist.*, *34*(3), 1075–1114 (cit. on p. 277).
- Datta, S., Bandyopadhyay, D., & 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 (cit. on p. 302).
- Datta, S., & McCormick, W. P. (1992). Bootstrap for a finite state Markov chain based on i.i.d. resampling. *Exploring the limits of bootstrap (East Lansing, MI, 1990)* (pp. 77–97). Wiley, New York. (Cit. on p. 15).
- Datta, S., & McCormick, W. P. (1993). Regeneration-based bootstrap for Markov chains. *Canad. J. Statist.*, *21*(2), 181–193 (cit. on pp. 15, 57, 67, 91).
- Davydov, J. A. (1973). Mixing conditions for Markov chains. *Teor. Veroyatnost. i Primenen.*, *18*, 321–338 (cit. on p. 207).
- de la Peña, V. H. (1992). Decoupling and Khintchine’s inequalities for U -statistics. *Ann. Probab.*, *20*(4), 1877–1892 (cit. on pp. 4, 268, 351, 352).
- de la Peña, V. H., & Montgomery-Smith, S. J. (1995). Decoupling inequalities for the tail probabilities of multivariate U -statistics. *Ann. Probab.*, *23*(2), 806–816 (cit. on p. 4).
- de la Peña, V. H., & Giné, E. (1999). *Decoupling*. Springer-Verlag, New York. (Cit. on pp. 55, 83, 96, 101, 102, 133–135, 148, 153–155, 161, 162, 177, 179, 195, 197, 200, 234, 250, 273, 320).
- Dedecker, J., Doukhan, P., Lang, G., León R., J. R., Louhichi, S., & Prieur, C. (2007). *Weak dependence: With examples and applications* (Vol. 190). Springer, New York. (Cit. on p. 280).
- Deheuvels, P. (2011). One bootstrap suffices to generate sharp uniform bounds in functional estimation. *Kybernetika (Prague)*, *47*(6), 855–865 (cit. on pp. 210, 214, 284, 290).
- Dehling, H. (1989). The functional law of the iterated logarithm for von Mises functionals and multiple Wiener integrals. *J. Multivariate Anal.*, *28*(2), 177–189 (cit. on p. 110).
- Dehling, H., Denker, M., & Philipp, W. (1984). Invariance principles for von Mises and U -statistics. *Z. Wahrsch. Verw. Gebiete*, *67*(2), 139–167 (cit. on p. 4).
- Dehling, H., Denker, M., & Philipp, W. (1986). A bounded law of the iterated logarithm for Hilbert space valued martingales and its application to U -statistics. *Probab. Theory Relat. Fields*, *72*(1), 111–131 (cit. on p. 110).
- Dehling, H., & Mikosch, T. (1994). Random quadratic forms and the bootstrap for U -statistics. *J. Multivariate Anal.*, *51*(2), 392–413 (cit. on p. 154).
- Dehling, H., & Taqqu, M. S. (1989). The empirical process of some long-range dependent sequences with an application to U -statistics. *Ann. Statist.*, *17*(4), 1767–1783 (cit. on p. 57).
- Dehling, H., & Taqqu, M. S. (1991). Bivariate symmetric statistics of long-range dependent observations. *J. Statist. Plann. Inference*, *28*(2), 153–165 (cit. on p. 57).
- Dehling, H., & Wendler, M. (2010). Central limit theorem and the bootstrap for U -statistics of strongly mixing data. *J. Multivariate Anal.*, *101*(1), 126–137 (cit. on p. 97).

- Denker, M., & Keller, G. (1983). On U -statistics and v. Mises' statistics for weakly dependent processes. *Z. Wahrsch. Verw. Gebiete*, *64*(4), 505–522 (cit. on pp. 57, 96, 154, 200, 273).
- Devroye, L., & Lugosi, G. (2001). *Combinatorial methods in density estimation*. Springer-Verlag, New York. (Cit. on pp. 36, 201, 272, 305).
- DiCiccio, C., & Romano, J. (2022). CLT for U -statistics with growing dimension. *Statist. Sinica*, *32*(1), 323–344 (cit. on pp. 72, 77).
- Didi, S., Al Harby, A., & Bouzebda, S. (2022). Wavelet density and regression estimators for functional stationary and ergodic data: Discrete time. *Mathematics*, *10*(19) (cit. on p. 274).
- Douc, R., Guillin, A., & Moulines, E. (2008). Bounds on regeneration times and limit theorems for subgeometric markov chains. *Annales de l'IHP Probabilités et statistiques*, *44*(2), 239–257 (cit. on pp. 106, 163).
- Douc, R., Moulines, E., Priouret, P., & Soulier, P. (2018). *Markov chains*. Springer, Cham. (Cit. on p. 58).
- Dougherty, E. R. (1999). *Random processes for image and signal processing*. SPIE Optical Engineering Press, Bellingham, WA; IEEE Press, Piscataway, NJ. (Cit. on p. 51).
- Doukhan, P., & León, J. R. (1989). Cumulants for stationary mixing random sequences and applications to empirical spectral density. *Probab. Math. Statist.*, *10*(1), 11–26 (cit. on pp. 16, 17).
- Doukhan, P. (1994). *Mixing* (Vol. 85). Springer-Verlag, New York. (Cit. on p. 280).
- Doukhan, P. (2012). *Mixing: Properties and examples* (Vol. 85). Springer Science & Business Media. (Cit. on p. 16).
- Duchemin, Q., De Castro, Y., & Lacour, C. (2020). *Concentration inequality for U -statistics of order two for uniformly ergodic Markov chains, and applications*. (Cit. on p. 57).
- Dudley, R. M. (1967). The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Functional Analysis*, *1*, 290–330 (cit. on pp. 209, 283).
- Dudley, R. M. (1984). A course on empirical processes. *École d'été de probabilités de Saint-Flour, XII—1982* (pp. 1–142). Springer, Berlin. (Cit. on pp. 297, 298).
- Dudley, R. M. (1985). The structure of some Vapnik-červonenkis classes. *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983)*, 495–508 (cit. on p. 73).
- Dudley, R. M. (1990). Nonlinear functionals of empirical measures and the bootstrap. *Probability in Banach spaces, 7 (Oberwolfach, 1988)* (pp. 63–82). Birkhäuser Boston, Boston, MA. (Cit. on p. 197).
- Dudley, R. M. (2014). *Uniform central limit theorems* (Second, Vol. 142). Cambridge University Press, New York. (Cit. on pp. 210, 284).
- Dwass, M. (1956). The large-sample power of rank order tests in the two-sample problem. *Ann. Math. Statist.*, *27*, 352–374 (cit. on p. 115).
- Dynkin, E. B., & Mandelbaum, A. (1983). Symmetric statistics, Poisson point processes, and multiple Wiener integrals. *Ann. Statist.*, *11*(3), 739–745 (cit. on p. 56).
- Eberlein, E. (1984). Weak convergence of partial sums of absolutely regular sequences. *Statist. Probab. Lett.*, *2*(5), 291–293 (cit. on pp. 231, 261).
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.*, *7*(1), 1–26 (cit. on pp. 14, 56, 71, 96, 111, 154, 354).

- Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci.*, 1(1), 54–77 (cit. on p. 14).
- Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans* (Vol. 38). Society for Industrial; Applied Mathematics (SIAM), Philadelphia, Pa. (Cit. on p. 47).
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap* (Vol. 57). Chapman; Hall, New York. (Cit. on pp. 14, 97).
- Eggermont, P. P. B., & LaRiccia, V. N. (2001). *Maximum penalized likelihood estimation. Vol. I*. Springer-Verlag, New York. (Cit. on pp. 36, 201, 272).
- Ellison, G., & 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 (cit. on p. 24).
- Faivishevsky, L., & Goldberger, J. (2009). Ica based on a smooth estimation of the differential entropy. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in neural information processing systems*. Curran Associates, Inc. (Cit. on pp. 200, 273).
- Fan, Y., & Ullah, A. (1999). On goodness-of-fit tests for weakly dependent processes using kernel method. *J. Nonparametr. Statist.*, 11(1-3), 337–360 (cit. on pp. 172, 173).
- Ferraty, F., Mas, A., & Vieu, P. (2007). Nonparametric regression on functional data: Inference and practical aspects. *Australian & New Zealand Journal of Statistics*, 49, 267–286 (cit. on pp. 214, 288).
- Ferraty, F., Laksaci, A., Tadj, A., & Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference*, 140(2), 335–352 (cit. on pp. 203, 274).
- Ferraty, F., & Vieu, P. (2006). *Nonparametric functional data analysis*. Springer, New York. (Cit. on pp. 20, 21, 203, 209, 213, 217, 274, 283, 288).
- Finesso, L. (1992). Estimation of the order of a finite Markov chain. *Recent advances in mathematical theory of systems, control, networks and signal processing, I (Kobe, 1991)* (pp. 643–645). Mita, Tokyo. (Cit. on p. 15).
- Fisher, N. I., & Hall, P. (1991). Bootstrap algorithms for small samples. *J. Statist. Plann. Inference*, 27(2), 157–169 (cit. on p. 111).
- Földes, A., & Rejtő, L. (1981). A LIL type result for the product limit estimator. *Z. Wahrsch. Verw. Gebiete*, 56(1), 75–86 (cit. on p. 303).
- Fort, G., Moulines, E., Priouret, P., & Vandekerkhove, P. (2012). A simple variance inequality for U -statistics of a Markov chain with applications. *Statist. Probab. Lett.*, 82(6), 1193–1201 (cit. on p. 57).
- Freedman, D. A. (1984). On bootstrapping two-stage least-squares estimates in stationary linear models. *Ann. Statist.*, 12(3), 827–842 (cit. on pp. 14, 56).
- Freedman, & Peters, C. (1984). Bootstrapping a regression equation: Some empirical results. *J. Amer. Statist. Assoc.*, 79(385), 97–106 (cit. on p. 14).
- Frees, E. W. (1989). Infinite order U -statistics. *Scand. J. Statist.*, 16(1), 29–45 (cit. on pp. 55, 56, 72, 88, 95, 179, 200, 273).

- Frees, E. W. (1994). Estimating densities of functions of observations. *J. Amer. Statist. Assoc.*, 89(426), 517–525 (cit. on pp. 4, 6).
- Fu, K.-A. (2012). An application of U -statistics to nonparametric functional data analysis. *Comm. Statist. Theory Methods*, 41(9), 1532–1542 (cit. on pp. 202, 275).
- Fukumizu, K., Gretton, A., Sun, X., & Schölkopf, B. (2007). Kernel measures of conditional dependence. *Advances in neural information processing systems*, 20 (cit. on p. 7).
- Gao, J., & Gijbels, I. (2008). Bandwidth selection in nonparametric kernel testing. *J. Amer. Statist. Assoc.*, 103(484), 1584–1594 (cit. on p. 305).
- Gasser, T., Hall, P., & 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 (cit. on pp. 203, 213, 274, 288).
- Ghosal, S., Sen, A., & van der Vaart, A. W. (2000). Testing monotonicity of regression. *Ann. Statist.*, 28(4), 1054–1082 (cit. on pp. 24, 95, 154, 200, 273).
- Gill, R. D., & Johansen, S. (1990). A survey of product-integration with a view toward application in survival analysis. *Ann. Statist.*, 18(4), 1501–1555 (cit. on p. 76).
- Giné, E., & Latała, J., Rafałand Zinn. (2000). Exponential and moment inequalities for U -statistics. *High dimensional probability, II (Seattle, WA, 1999)* (pp. 13–38). Birkhäuser Boston, Boston, MA. (Cit. on p. 4).
- Giné, E., & Mason, D. M. (2007a). Laws of the iterated logarithm for the local U -statistic process. *J. Theoret. Probab.*, 20(3), 457–485 (cit. on pp. 4, 153, 273).
- Giné, E., & 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 (cit. on pp. 95, 153, 273).
- Giné, E., & Zinn, J. (1984). Some limit theorems for empirical processes. *Ann. Probab.*, 12(4), 929–998 (cit. on pp. 244, 262, 331, 341).
- Giné, E., & Zinn, J. (1990). Bootstrapping general empirical measures. *Ann. Probab.*, 18(2), 851–869 (cit. on pp. 33, 47, 49, 97, 120, 136, 156).
- Giné, E., & Zinn, J. (1992). On Hoffmann-Jørgensen’s inequality for U -processes. *Probability in Banach spaces, 8 (Brunswick, ME, 1991)* (pp. 80–91). Birkhäuser Boston, Boston, MA. (Cit. on p. 4).
- Giné, E., & Zinn, J. (1994). A remark on convergence in distribution of U -statistics. *Ann. Probab.*, 22(1), 117–125 (cit. on p. 4).
- Gregory, G. G. (1977). Large sample theory for U -statistics and tests of fit. *Ann. Statist.*, 5(1), 110–123 (cit. on p. 4).
- Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., & Smola, A. (2012). A kernel two-sample test. *J. Mach. Learn. Res.*, 13, 723–773 (cit. on pp. 7, 8).
- Gretton, A., Fukumizu, K., Teo, C., Song, L., Schölkopf, B., & Smola, A. (2007). A kernel statistical test of independence. *Advances in neural information processing systems*, 20 (cit. on p. 9).
- Gretton, A., Herbrich, R., Smola, A., Bousquet, O., & Schölkopf, B. (2005). Kernel methods for measuring independence. *J. Mach. Learn. Res.*, 6, 2075–2129 (cit. on p. 9).
- Gut, A. (1974a). On the moments and limit distributions of some first passage times. *Ann. Probability*, 2, 277–308 (cit. on p. 74).

- Gut, A. (1974b). On the moments of some first passage times for sums of dependent random variables. *Stochastic Process. Appl.*, 2, 115–126 (cit. on p. 74).
- Guyon, X. (1995). *Random fields on a network*. Springer-Verlag, New York. (Cit. on pp. 36, 272).
- Hall, P., & Heyde, C. C. (1980). *Martingale limit theory and its application*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London. (Cit. on p. 268).
- 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 (cit. on p. 304).
- Hall, P. (1987). On the bootstrap and likelihood-based confidence regions. *Biometrika*, 74(3), 481–493 (cit. on p. 111).
- Hall, P. (1992). *The bootstrap and Edgeworth expansion*. Springer-Verlag, New York. (Cit. on pp. 111, 154).
- Hall, P., & Patil, P. (1994). Properties of nonparametric estimators of autocovariance for stationary random fields. *Probab. Theory Related Fields*, 99(3), 399–424 (cit. on p. 279).
- Halmos, P. R. (1946). The theory of unbiased estimation. *Ann. Math. Statistics*, 17, 34–43 (cit. on pp. 1, 55, 96, 153, 200, 273).
- Han, F., & Qian, T. (2018). On inference validity of weighted U -statistics under data heterogeneity. *Electron. J. Stat.*, 12(2), 2637–2708 (cit. on p. 217).
- Han, Q. (2022). Multiplier U -processes: Sharp bounds and applications. *Bernoulli*, 28(1), 87–124 (cit. on pp. 32, 56, 155, 156, 168–170, 193, 195, 354).
- Härdle, W. (1990). *Applied nonparametric regression* (Vol. 19). Cambridge University Press, Cambridge. (Cit. on pp. 36, 201, 272).
- Härdle, W., & Marron, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.*, 13(4), 1465–1481 (cit. on p. 304).
- Harel, M., & Puri, M. L. (1989). Weak convergence of the U -statistic and weak invariance of the one-sample rank order statistic for Markov processes and ARMA models. *J. Multivariate Anal.*, 31(2), 258–265 (cit. on p. 57).
- Harel, M., & Puri, M. L. (1996). Conditional U -statistics for dependent random variables. *J. Multivariate Anal.*, 57(1), 84–100 (cit. on pp. 202, 275).
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (Second). Springer, New York. (Cit. on p. 23).
- Hediger, S., Michel, L., & Näf, J. (2022). On the use of random forest for two-sample testing. *Comput. Statist. Data Anal.*, 170, Paper No. 107435, 34 (cit. on p. 72).
- Heilig, C., & Nolan, D. (2001). Limit theorems for the infinite-degree U -process. *Statist. Sinica*, 11(1), 289–302 (cit. on pp. 28, 53, 55–57, 60, 62–65, 72–75, 83, 85, 89, 200, 273, 353).
- Heilig, C. M. (1997). *An empirical process approach to U -processes of increasing degree*. ProQuest LLC, Ann Arbor, MI. (Cit. on pp. 56, 76, 77, 84, 88, 89).
- Heilig, C. M., & Nolan, D. (1998). Applications of the infinite-degree U -process. *Manuscript in preparation* (cit. on p. 56).
- Hitczenko, P. (1988). Comparison of moments for tangent sequences of random variables. *Probab. Theory Related Fields*, 78(2), 223–230 (cit. on p. 352).

- Hobert, J. P., & Robert, C. P. (2004). A mixture representation of π with applications in Markov chain Monte Carlo and perfect sampling. *Ann. Appl. Probab.*, *14*(3), 1295–1305 (cit. on p. 70).
- Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics*, *19*, 293–325 (cit. on pp. 1, 36, 55, 95, 96, 122, 126, 153, 174, 200, 217, 272, 273, 292).
- Hoeffding, W. (1961). *The strong law of large numbers for u -statistics*. (tech. rep.). North Carolina State University. Dept. of Statistics. (Cit. on p. 4).
- Hoffmann-Jørgensen, J. (1991). *Stochastic processes on Polish spaces* (Vol. 39). Aarhus Universitet, Matematisk Institut, Aarhus. (Cit. on p. 113).
- Hollander, M., & Proschan, F. (1972). Testing whether new is better than used. *Ann. Math. Statist.*, *43*, 1136–1146 (cit. on p. 122).
- Horowitz, J. L. (2003). Bootstrap methods for Markov processes. *Econometrica*, *71*(4), 1049–1082 (cit. on p. 15).
- Horowitz, J. L., & Spokoiny, V. G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, *69*(3), 599–631 (cit. on p. 305).
- Horváth, L., & Kokoszka, P. (2012). *Inference for functional data with applications*. Springer, New York. (Cit. on pp. 21, 203, 274).
- Hsing, T., & Wu, W. B. (2004). On weighted U -statistics for stationary processes. *Ann. Probab.*, *32*(2), 1600–1631 (cit. on p. 57).
- Hůsková, M., & Janssen, P. (1993). Consistency of the generalized bootstrap for degenerate U -statistics. *Ann. Statist.*, *21*(4), 1811–1823 (cit. on pp. 155, 171, 178).
- Ibragimov, I. A., & Slev, V. N. (1969). A certain condition for the regularity of Gaussian stationary sequence. *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, *12*, 113–125 (cit. on pp. 279, 355).
- Jadhav, S., & Ma, S. (2019). Kendall’s tau for functional data analysis. (Cit. on pp. 202, 276).
- Janson, S. (1979). *The asymptotic distribution of degenerate u -statistics*. [Uppsala] Department of Mathematics, Uppsala University. (Cit. on p. 4).
- Janson, S. (1984). The asymptotic distributions of incomplete U -statistics. *Z. Wahrsch. Verw. Gebiete*, *66*(4), 495–505 (cit. on p. 6).
- Janssen, P. (1994). Weighted bootstrapping of U -statistics. *Journal of statistical planning and inference*, *38*(1), 31–41 (cit. on p. 155).
- Jin, R., Wang, S., & Zhou, Y. (2009). Regularized distance metric learning: theory and algorithm. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, & A. Culotta (Eds.), *Advances in neural information processing systems*. Curran Associates, Inc. (Cit. on pp. 221, 296).
- Joly, E., & Lugosi, G. (2016). Robust estimation of U -statistics. *Stochastic Process. Appl.*, *126*(12), 3760–3773 (cit. on pp. 5, 95, 154, 273).
- Kallenberg, O. (2002). *Foundations of modern probability* (Second). Springer-Verlag, New York. (Cit. on p. 48).
- Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, *53*, 457–481 (cit. on pp. 75, 301).

- Kara-Zaitri, L., Laksaci, A., Rachdi, M., & Vieu, P. (2017). Uniform in bandwidth consistency for various kernel estimators involving functional data. *J. Nonparametr. Stat.*, *29*(1), 85–107 (cit. on pp. 203, 274).
- Kieffer, J. C. (1993). Strongly consistent code-based identification and order estimation for constrained finite-state model classes. *IEEE Trans. Inform. Theory*, *39*(3), 893–902 (cit. on p. 15).
- Kim, I., Balakrishnan, S., & Wasserman, L. (2022). Minimax optimality of permutation tests. *The Annals of Statistics*, *50*(1), 225–251 (cit. on pp. 7, 8, 10).
- Klass, M. J., & 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 (cit. on p. 4).
- Kohatsu-Higa, A. (1990). Weak convergence of a sequence of stochastic processes related with U -statistics. *Osaka J. Math.*, *27*(2), 361–371 (cit. on p. 56).
- Kohatsu-Higa, A. (1991). Weak convergence of infinite order U -processes. *Statist. Probab. Lett.*, *12*(2), 145–150 (cit. on p. 56).
- Kohler, M., Máthé, K., & Pintér, M. (2002). Prediction from randomly right censored data. *J. Multivariate Anal.*, *80*(1), 73–100 (cit. on pp. 300, 301).
- Kolmogorov, A. N., & Tihomirov, V. M. (1959). ε -entropy and ε -capacity of sets in function spaces. *Uspehi Mat. Nauk*, *14*(2 (86)), 3–86 (cit. on pp. 45, 121, 209, 283).
- Koroljuk, V. S., & Borovskich, Y. V. (1994). *Theory of U -statistics* (Vol. 273). Kluwer Academic Publishers Group, Dordrecht. (Cit. on pp. 4, 96, 200, 273).
- Korolyuk, V. S., & Borovskikh, Y. V. (1990). U -statistics from bootstrap sampling. *Ukrain. Mat. Zh.*, *42*(10), 1403–1406 (cit. on p. 56).
- Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer, New York. (Cit. on pp. 41, 42, 44, 45, 48, 120, 121, 210, 284).
- Kulperger, R. J. (1999). Countable state Markov process bootstrap. *J. Statist. Plann. Inference*, *76*(1-2), 19–29 (cit. on p. 15).
- Kulperger, R. J., & Prakasa Rao, B. L. S. (1989). Bootstrapping a finite state Markov chain. *Sankhyā Ser. A*, *51*(2), 178–191 (cit. on p. 15).
- Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.*, *17*(3), 1217–1241 (cit. on p. 14).
- Kurusu, D. (2022a). Nonparametric regression for locally stationary functional time series. *Electron. J. Stat.*, *16*(2), 3973–3995 (cit. on pp. 213, 276, 288).
- Kurusu, D. (2022b). Nonparametric regression for locally stationary random fields under stochastic sampling design. *Bernoulli*, *28*(2), 1250–1275 (cit. on pp. 276, 279, 280, 288, 306, 357).
- Kurusu, D., Kato, K., & Shao, X. (2021). Gaussian approximation and spatially dependent wild bootstrap for high-dimensional spatial data. *arXiv preprint arXiv:2103.10720* (cit. on pp. 276, 279, 280, 351).
- Lahiri, S. N. (2003a). Central limit theorems for weighted sums of a spatial process under a class of stochastic and fixed designs. *Sankhyā*, *65*(2), 356–388 (cit. on pp. 38, 279, 280, 316, 347, 357).
- Lahiri, S. N. (2003b). *Resampling methods for dependent data*. Springer-Verlag, New York. (Cit. on pp. 279, 289).

- Lahiri, S. N., & Zhu, J. (2006). Resampling methods for spatial regression models under a class of stochastic designs. *Ann. Statist.*, *34*(4), 1774–1813 (cit. on p. 280).
- Lahiri, S. N., Kaiser, M. S., Cressie, N., & Hsu, N.-J. (1999). Prediction of spatial cumulative distribution functions using subsampling. *J. Amer. Statist. Assoc.*, *94*(445), 86–110 (cit. on p. 289).
- Lee, A. J. (1990). *U-statistics* (Vol. 110). Marcel Dekker, Inc., New York. (Cit. on pp. 4, 55, 96, 122, 126, 130, 144, 147, 149, 153, 174, 200, 273).
- Lee, S., Linton, O., & Whang, Y.-J. (2009). Testing for stochastic monotonicity. *Econometrica*, *77*(2), 585–602 (cit. on pp. 24, 95, 154, 200, 273).
- Lehmann, E. L. (1951). Consistency and unbiasedness of certain nonparametric tests. *Ann. Math. Statistics*, *22*, 165–179 (cit. on pp. 115, 144).
- Lehmann, E. L., & Romano, J. P. (2005). *Testing statistical hypotheses* (Third). Springer, New York. (Cit. on p. 10).
- Leonov, V. P., & Shiryaev, A. N. (1959). On a method of calculation of semi-invariants. *Theory of Probability & its applications*, *4*(3), 319–329 (cit. on p. 17).
- Leucht, A. (2012). Degenerate U - and V -statistics under weak dependence: Asymptotic theory and bootstrap consistency. *Bernoulli*, *18*(2), 552–585 (cit. on pp. 96, 154, 200, 273).
- Leucht, A., & Neumann, M. H. (2009). Consistency of general bootstrap methods for degenerate U -type and V -type statistics. *J. Multivariate Anal.*, *100*(8), 1622–1633 (cit. on p. 154).
- Leucht, A., & 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 (cit. on pp. 96, 97, 154, 200, 273).
- Levental, S. (1988). Uniform limit theorems for Harris recurrent Markov chains. *Probab. Theory Related Fields*, *80*(1), 101–118 (cit. on pp. 62, 74, 78, 81, 83, 128, 129, 147).
- Levitin, D. J., Nuzzo, R. L., Vines, B. W., & Ramsay, J. (2007). Introduction to functional data analysis. *Canadian Psychology/Psychologie canadienne*, *48*(3), 135 (cit. on p. 21).
- Li, Q., Maasoumi, E., & Racine, J. S. (2009). A nonparametric test for equality of distributions with mixed categorical and continuous data. *J. Econometrics*, *148*(2), 186–200 (cit. on p. 305).
- Li, W. V., & Shao, Q.-M. (2001). Gaussian processes: Inequalities, small ball probabilities and applications. *Stochastic processes: Theory and methods* (pp. 533–597). North-Holland, Amsterdam. (Cit. on pp. 213, 288).
- Liapounoff, A. (1900). Sur une proposition de la théorie des probabilités. *13*(4), 359–386 (cit. on p. 16).
- Liebscher, E., & Schäbe, H. (1997). A generalization of the Kaplan-Meier estimator to Harris-recurrent Markov chains. *Statist. Papers*, *38*(1), 63–75 (cit. on pp. 75, 77).
- Liebscher, E. (1996). Strong convergence of sums of α -mixing random variables with applications to density estimation. *Stochastic Process. Appl.*, *65*(1), 69–80 (cit. on pp. 267, 268).
- Lim, F., & Stojanovic, V. M. (2013). On u -statistics and compressed sensing i: Non-asymptotic average-case analysis. *IEEE Transactions on Signal Processing*, *61*(10), 2473–2485 (cit. on p. 200).
- Lindeberg, J. W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.*, *15*(1), 211–225 (cit. on p. 16).
- Ling, N., Meng, S., & Vieu, P. (2019). Uniform consistency rate of k NN regression estimation for functional time series data. *J. Nonparametr. Stat.*, *31*(2), 451–468 (cit. on p. 274).

- Ling, N., & Vieu, P. (2018). Nonparametric modelling for functional data: Selected survey and tracks for future. *Statistics*, *52*(4), 934–949 (cit. on pp. 203, 274).
- Liu, C.-C., & Narayan, P. (1994). Order estimation and sequential universal data compression of a hidden markov source by the method of mixtures. *IEEE Transactions on Information Theory*, *40*(4), 1167–1180 (cit. on p. 15).
- Liu, Q., Lee, J., & Jordan, M. (2016). A kernelized stein discrepancy for goodness-of-fit tests. *International conference on machine learning*, 276–284 (cit. on pp. 10, 200, 273).
- Lo, A. Y. (1993). A Bayesian method for weighted sampling. *Ann. Statist.*, *21*(4), 2138–2148 (cit. on pp. 47, 155).
- Mahalanobis, P. C. (1946). Sample surveys of crop yields in india. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, *7*(3), 269–280 (cit. on p. 49).
- Mai, Q., Zou, H., & Yuan, M. (2012). A direct approach to sparse discriminant analysis in ultra-high dimensions. *Biometrika*, *99*(1), 29–42 (cit. on p. 2).
- Maillot, B., & 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 (cit. on pp. 300, 301).
- Maritz, J. S., & Jarrett, R. G. (1978). A note on estimating the variance of the sample median. *Journal of the American Statistical Association*, *73*(361), 194–196 (cit. on p. 49).
- Mason, D. M. (2012). Proving consistency of non-standard kernel estimators. *Stat. Inference Stoch. Process.*, *15*(2), 151–176 (cit. on p. 295).
- Mason, D. M., & Newton, M. A. (1992). A rank statistics approach to the consistency of a general bootstrap. *Ann. Statist.*, *20*(3), 1611–1624 (cit. on pp. 15, 47, 48, 155, 171).
- Masry, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *J. Time Ser. Anal.*, *17*(6), 571–599 (cit. on p. 345).
- Masry, E. (2005). Nonparametric regression estimation for dependent functional data: Asymptotic normality. *Stochastic Process. Appl.*, *115*(1), 155–177 (cit. on pp. 36, 213, 228, 276, 288).
- Matsuda, Y., & Yajima, Y. (2009). Fourier analysis of irregularly spaced data on \mathbb{R}^d . *J. R. Stat. Soc. Ser. B Stat. Methodol.*, *71*(1), 191–217 (cit. on p. 279).
- Mayer, U., Zähle, H., & Zhou, Z. (2020). Functional weak limit theorem for a local empirical process of non-stationary time series and its application. *Bernoulli*, *26*(3), 1891–1911 (cit. on pp. 34, 203).
- Mayer-Wolf, E., & Zeitouni, O. (1993). The probability of small Gaussian ellipsoids and associated conditional moments. *Ann. Probab.*, *21*(1), 14–24 (cit. on pp. 213, 288).
- McDonald, D. J., & Shalizi, C. R. (2017). Rademacher complexity of stationary sequences. (Cit. on p. 83).
- Mentch, L., & Hooker, G. (2016). Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. *J. Mach. Learn. Res.*, *17*, Paper No. 26, 41 (cit. on pp. 6, 55, 71).
- Merhav, N., Gutman, M., & Ziv, J. (1989). On the estimation of the order of a Markov chain and universal data compression. *IEEE Trans. Inform. Theory*, *35*(5), 1014–1019 (cit. on p. 15).
- Meyn, S. P., & Tweedie, R. L. (1993). *Markov chains and stochastic stability*. Springer-Verlag London, Ltd., London. (Cit. on pp. 58, 92, 106, 163).
- Meyn, S., & Tweedie, R. L. (2009a). *Markov chains and stochastic stability* (2nd). Cambridge University Press. (Cit. on p. 128).

- Meyn, S., & Tweedie, R. L. (2009b). *Markov chains and stochastic stability* (Second). Cambridge University Press, Cambridge. (Cit. on pp. 106, 163).
- Mohammedi, M., Bouzebda, S., & 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 (cit. on pp. 203, 274).
- Montgomery-Smith, S. J. (1993). Comparison of sums of independent identically distributed random vectors. *Probab. Math. Statist.*, 14(2), 281–285 (1994) (cit. on p. 148).
- Müller, A. (1997). Integral probability metrics and their generating classes of functions. *Adv. in Appl. Probab.*, 29(2), 429–443 (cit. on p. 7).
- Nadaraja, È. A. (1964). On a regression estimate. *Teor. Verojatnost. i Primenen.*, 9, 157–159 (cit. on pp. 21, 201).
- Nadaraya, È. A. (1989). *Nonparametric estimation of probability densities and regression curves* (Vol. 20). Kluwer Academic Publishers Group, Dordrecht. (Cit. on pp. 36, 201, 272).
- Ndiaye, M., Dabo-Niang, S., & Ngom, P. (2022). Nonparametric prediction for spatial dependent functional data under fixed sampling design. *Rev. Colombiana Estadíst.*, 45(2), 391–428 (cit. on p. 272).
- Neumann, M. H., & von Sachs, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Ann. Statist.*, 25(1), 38–76 (cit. on pp. 18, 34, 203).
- Nolan, D., & Pollard, D. (1987). *U*-processes: Rates of convergence. *Ann. Statist.*, 15(2), 780–799 (cit. on pp. 84, 89, 154, 200, 210, 273, 284).
- Nummelin, E. (1978). A splitting technique for Harris recurrent Markov chains. *Z. Wahrsch. Verw. Gebiete*, 43(4), 309–318 (cit. on pp. 69, 100, 112, 158).
- Nummelin, E. (1984). *General irreducible Markov chains and nonnegative operators* (Vol. 83). Cambridge University Press, Cambridge. (Cit. on pp. 69, 99, 157).
- Nummelin, E. (1991). Renewal representations for Markov operators. *Adv. Math.*, 90(1), 15–46 (cit. on pp. 30, 97).
- Oates, C. J., Girolami, M., & Chopin, N. (2017). Control functionals for Monte Carlo integration. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79(3), 695–718 (cit. on p. 10).
- Pakes, A., & Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, 57(5), 1027–1057 (cit. on p. 42).
- Paparoditis, E., & Politis, D. N. (2001a). A Markovian local resampling scheme for nonparametric estimators in time series analysis. *Econometric Theory*, 17(3), 540–566 (cit. on p. 15).
- Paparoditis, E., & Politis, D. N. (2001b). Tapered block bootstrap. *Biometrika*, 88(4), 1105–1119 (cit. on p. 14).
- Paparoditis, E., & Politis, D. N. (2002a). The local bootstrap for Markov processes. (Cit. on p. 15).
- Paparoditis, E., & Politis, D. N. (2002b). The tapered block bootstrap for general statistics from stationary sequences. *Econom. J.*, 5(1), 131–148 (cit. on p. 14).
- Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33, 1065–1076 (cit. on p. 201).

- Pauly, M. (2012). Consistency of the subsample bootstrap empirical process. *Statistics*, 46(5), 621–626 (cit. on p. 172).
- Peng, W., Coleman, T., & Mentch, L. (2019). Asymptotic distributions and rates of convergence for random forests via generalized u -statistics. (Cit. on p. 273).
- Peng, W., Coleman, T., & Mentch, L. (2022). Rates of convergence for random forests via generalized U -statistics. *Electron. J. Stat.*, 16(1), 232–292 (cit. on pp. 72, 96, 200).
- Phandoidaen, N., & Richter, S. (2022). Empirical process theory for locally stationary processes. *Bernoulli*, 28(1), 453–480 (cit. on pp. 34, 203).
- Pinheiro, A., & Sen, P. K. (2014). Quasi U -statistics of infinite order and applications to the subgroup decomposition of some diversity measures. *São Paulo J. Math. Sci.*, 8(2), 285–309 (cit. on p. 56).
- Pinheiro, A., Sen, P. K., & Pinheiro, H. P. (2011). A class of asymptotically normal degenerate quasi U -statistics. *Ann. Inst. Statist. Math.*, 63(6), 1165–1182 (cit. on p. 56).
- Pinheiro, A., Sen, P. K., & Pinheiro, H. P. (2009). Decomposability of high-dimensional diversity measures: Quasi- U -statistics, martingales and nonstandard asymptotics. *J. Multivariate Anal.*, 100(8), 1645–1656 (cit. on p. 56).
- Piterbarg, V. I. (1996). *Asymptotic methods in the theory of Gaussian processes and fields* (Vol. 148). American Mathematical Society, Providence, RI. (Cit. on p. 24).
- Politis, D. N., & Romano, J. P. (1992). A circular block-resampling procedure for stationary data. *Exploring the limits of bootstrap (East Lansing, MI, 1990)* (pp. 263–270). Wiley, New York. (Cit. on pp. 14, 154).
- Politis, D. N., & Romano, J. P. (1994a). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.*, 22(4), 2031–2050 (cit. on p. 56).
- Politis, D. N., & Romano, J. P. (1994b). The stationary bootstrap. *J. Amer. Statist. Assoc.*, 89(428), 1303–1313 (cit. on pp. 14, 154).
- Pollard, D. (1984). *Convergence of stochastic processes*. Springer-Verlag, New York. (Cit. on pp. 63, 79, 80, 103).
- Polonik, W., & Yao, Q. (2002). Set-indexed conditional empirical and quantile processes based on dependent data. *J. Multivariate Anal.*, 80(2), 234–255 (cit. on p. 297).
- Præstgaard, J., & Wellner, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.*, 21(4), 2053–2086 (cit. on pp. 15, 47–49, 155, 168, 171, 172).
- Prakasa Rao, B. L. S., & Sen, A. (1995). Limit distributions of conditional U -statistics. *J. Theoret. Probab.*, 8(2), 261–301 (cit. on pp. 202, 275).
- Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes.(With discussion). *J. Roy. Statist. Soc. Ser. B*, 27, 204–237 (cit. on pp. 18, 34, 203).
- Qiu, R., Yu, Z., & Zhu, R. (2022). Random forests weighted local fréchet regression with theoretical guarantee. *CoRR*, abs/2202.04912 (cit. on p. 72).
- Quenouille, M. H. (1949). Approximate tests of correlation in time-series. *J. Roy. Statist. Soc. Ser. B*, 11, 68–84 (cit. on p. 49).
- Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika*, 43, 353–360 (cit. on p. 49).
- Rachdi, M., & Vieu, P. (2007). Nonparametric regression for functional data: Automatic smoothing parameter selection. *J. Statist. Plann. Inference*, 137(9), 2784–2801 (cit. on p. 304).

- Radulović, D. (2004). Renewal type bootstrap for Markov chains. *Test*, 13(1), 147–192 (cit. on pp. 15, 32, 59, 67, 69, 79, 81, 90, 92, 97, 110, 148, 155, 156, 168, 176, 192, 354).
- Rajarshi, M. B. (1990). Bootstrap in Markov-sequences based on estimates of transition density. *Ann. Inst. Statist. Math.*, 42(2), 253–268 (cit. on pp. 15, 56).
- Rakhlin, A., Sridharan, K., & Tewari, A. (2011). Online learning: Stochastic, constrained, and smoothed adversaries (J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, & K. Q. Weinberger, Eds.). (Cit. on p. 83).
- Ramsay, J. O., & Silverman, B. W. (2002). *Applied functional data analysis*. Springer-Verlag, New York. (Cit. on pp. 21, 203, 274).
- Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis* (Second). Springer, New York. (Cit. on p. 20).
- Rempala, G., & Gupta, A. (1999). Weak limits of U -statistics of infinite order. *Random Oper. Stochastic Equations*, 7(1), 39–52 (cit. on pp. 56, 179, 200, 273).
- Revuz, D. (1984). *Markov chains* (Second, Vol. 11). North-Holland Publishing Co., Amsterdam. (Cit. on pp. 58, 106, 163).
- Rio, E. (2013). *Inequalities and limit theorems for weakly dependent sequences* (3rd cycle). (Cit. on p. 350).
- Ripley, B. D. (1981). *Spatial statistics*. John Wiley & Sons, Inc., New York. (Cit. on pp. 36, 272).
- Robins, J., Li, L., Tchetgen, E., & van der Vaart, A. W. (2009). Quadratic semiparametric von Mises calculus. *Metrika*, 69(2-3), 227–247 (cit. on p. 5).
- Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U.S.A.*, 42, 43–47 (cit. on pp. 16, 67, 201, 279).
- Rosenblatt, M. (1985). Stationary sequences and random fields. (Cit. on pp. 36, 272).
- Rubin, D. B. (1981). The Bayesian bootstrap. *Ann. Statist.*, 9(1), 130–134 (cit. on pp. 15, 154).
- Rubin, H., & Vitale, R. A. (1980). Asymptotic distribution of symmetric statistics. *Ann. Statist.*, 8(1), 165–170 (cit. on p. 4).
- Sakiyama, K., & Taniguchi, M. (2004). Discriminant analysis for locally stationary processes. *J. Multivariate Anal.*, 90(2), 282–300 (cit. on pp. 18, 34, 203).
- Sato, K.-i. (1999). *Lévy processes and infinitely divisible distributions* (Vol. 68). Cambridge University Press, Cambridge. (Cit. on p. 26).
- Schick, A., Wang, Y., & Wefelmeyer, W. (2011). Tests for normality based on density estimators of convolutions. *Statist. Probab. Lett.*, 81(2), 337–343 (cit. on pp. 95, 153, 273).
- Schrab, A., Kim, I., Guedj, B., & Gretton, A. (2022). Efficient aggregated kernel tests using incomplete U -statistics. *arXiv preprint arXiv:2206.09194* (cit. on pp. 7, 8).
- Schupbach, J., Sheppard, J. W., & Forrester, T. (2020). Quantifying uncertainty in neural network ensembles using u -statistics. *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–8 (cit. on p. 5).
- Sen, A. (1994). Uniform strong consistency rates for conditional U -statistics. *Sankhyā Ser. A*, 56(2), 179–194 (cit. on pp. 202, 275).
- Sen, P. K. (1974a). Almost sure behaviour of U -statistics and von Mises’ differentiable statistical functions. *Ann. Statist.*, 2, 387–395 (cit. on p. 4).

- Sen, P. K. (1974b). Weak convergence of generalized U -statistics. *Ann. Probability*, 2(1), 90–102 (cit. on pp. 115, 144).
- Sen, P. K. (1977). Almost sure convergence of generalized U -statistics. *Ann. Probability*, 5(2), 287–290 (cit. on pp. 115, 143).
- Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. John Wiley & Sons, Inc., New York. (Cit. on pp. 4, 55, 126, 149, 153).
- 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 (cit. on p. 305).
- Shao, J., & Tu, D. S. (1995). *The jackknife and bootstrap*. Springer-Verlag, New York. (Cit. on pp. 97, 172).
- Shao, J., & Wu, C. (1987). Heteroscedasticity-robustness of jackknife variance estimators in linear models. *The Annals of Statistics*, 1563–1579 (cit. on p. 171).
- Sharipov, O. S., Tewes, J., & Wendler, M. (2016). Bootstrap for U -statistics: A new approach. *J. Nonparametr. Stat.*, 28(3), 576–594 (cit. on p. 97).
- Sherman, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica*, 61(1), 123–137 (cit. on pp. 95, 200).
- Sherman, R. P. (1994). Maximal inequalities for degenerate U -processes with applications to optimization estimators. *Ann. Statist.*, 22(1), 439–459 (cit. on pp. 66, 84, 89, 154, 200, 273, 353).
- Shieh, G. S. (1994). Infinite order V -statistics. *Statist. Probab. Lett.*, 20(1), 75–80 (cit. on p. 56).
- Silverman, B. W. (1978). Distances on circles, toruses and spheres. *J. Appl. Probability*, 15(1), 136–143 (cit. on p. 122).
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman & Hall, London. (Cit. on pp. 36, 201, 272, 305).
- Silverman, R. A. (1957). *Locally stationary random processes*. New York University, Institute of Mathematical Sciences, Division of Electromagnetic Research. (Cit. on pp. 18, 34, 202).
- Simon, J. (1969). *Basic research methods in social science*. Random House, New York. (Cit. on p. 49).
- Solon, G. (1992). Intergenerational income mobility in the united states. *The American Economic Review*, 82(3), 393–408 (cit. on p. 24).
- Song, L., Smola, A., Gretton, A., Bedo, J., & Borgwardt, K. (2012). Feature selection via dependence maximization. *J. Mach. Learn. Res.*, 13, 1393–1434 (cit. on p. 9).
- Song, Y., Chen, X., & Kato, K. (2019). Approximating high-dimensional infinite-order U -statistics: Statistical and computational guarantees. *Electron. J. Stat.*, 13(2), 4794–4848 (cit. on pp. 55, 57, 71, 95, 200, 273, 320, 351).
- Soukariéh, I., & Bouzebda, S. (2022). Exchangeably weighted bootstraps of general Markov U -process. *Mathematics*, 10(20), 1–42 (cit. on pp. 55, 273).
- Soukariéh, I., & Bouzebda, S. (2023). Renewal type bootstrap for increasing degree u -process of a markov chain. *Journal of Multivariate Analysis*, 195, 105143 (cit. on p. 273).
- Sriperumbudur, B. K., Fukumizu, K., & Lanckriet, G. R. G. (2011). Universality, characteristic kernels and RKHS embedding of measures. *J. Mach. Learn. Res.*, 12, 2389–2410 (cit. on p. 7).

- Statulevicius, V., & Jakimavicius, D. (1988). Estimates of semiinvariants and centered moments of stochastic processes with mixing. i. *Lithuanian Mathematical Journal*, 28(1), 67–80 (cit. on p. 16).
- Stute, W. (1991). Conditional U -statistics. *Ann. Probab.*, 19(2), 812–825 (cit. on pp. 201, 202, 207, 274, 275).
- Stute, W. (1993). Almost sure representations of the product-limit estimator for truncated data. *Ann. Statist.*, 21(1), 146–156 (cit. on pp. 95, 153, 272).
- Stute, W. (1994a). L^p -convergence of conditional U -statistics. *J. Multivariate Anal.*, 51(1), 71–82 (cit. on pp. 202, 219, 298).
- Stute, W. (1994b). Universally consistent conditional U -statistics. *Ann. Statist.*, 22(1), 460–473 (cit. on p. 219).
- Stute, W. (1994c). Universally consistent conditional U -statistics. *Ann. Statist.*, 22(1), 460–473 (cit. on p. 298).
- Stute, W. (1996). Symmetrized NN-conditional U -statistics. *Research developments in probability and statistics* (pp. 231–237). VSP, Utrecht. (Cit. on pp. 202, 275).
- Stute, W., & Wang, J.-L. (1993). Multi-sample U -statistics for censored data. *Scand. J. Statist.*, 20(4), 369–374 (cit. on p. 302).
- Tran, L. T. (1990). Kernel density estimation on random fields. *J. Multivariate Anal.*, 34(1), 37–53 (cit. on pp. 36, 272).
- Tran, L. T., & Yakowitz, S. (1993). Nearest neighbor estimators for random fields. *J. Multivariate Anal.*, 44(1), 23–46 (cit. on pp. 36, 272).
- Tukey, J. W. (1958). A problem of Berkson, and minimum variance orderly estimators. *Ann. Math. Statist.*, 29, 614 (cit. on p. 49).
- van der Vaart, A. W. (1998). *Asymptotic statistics* (Vol. 3). Cambridge University Press, Cambridge. (Cit. on pp. 55, 153, 294).
- van der Vaart, A. (1996). New Donsker classes. *Ann. Probab.*, 24(4), 2128–2140 (cit. on pp. 121, 155).
- van der Vaart, A. W., & Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer-Verlag, New York. (Cit. on pp. 42, 44–46, 48, 103, 120, 121, 180, 208, 210, 284).
- van Delft, A., & Eichler, M. (2018). Locally stationary functional time series. *Electron. J. Stat.*, 12(1), 107–170 (cit. on pp. 205, 206, 278).
- van Zwet, W. R. (1979). The Edgeworth expansion for linear combinations of uniform order statistics. *Proceedings of the Second Prague Symposium on Asymptotic Statistics (Hradec Králové, 1978)*, 93–101 (cit. on p. 172).
- v. Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statistics*, 18, 309–348 (cit. on pp. 55, 96, 153, 200, 273).
- Vogt, M. (2012). Nonparametric regression for locally stationary time series. *Ann. Statist.*, 40(5), 2601–2633 (cit. on pp. 16, 18, 19, 34, 203, 213, 288).
- Volkonskiui, V. A., & Rozanov, Y. A. (1959). Some limit theorems for random functions. I. *Theor. Probability Appl.*, 4, 178–197 (cit. on pp. 16, 279).
- Wand, M. P., & Jones, M. C. (1995). *Kernel smoothing* (Vol. 60). Chapman; Hall, Ltd., London. (Cit. on pp. 36, 201, 272).

- Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A*, *26*, 359–372 (cit. on pp. [21](#), [201](#)).
- Weng, C.-S. (1989). On a second-order asymptotic property of the Bayesian bootstrap mean. *Ann. Statist.*, *17*(2), 705–710 (cit. on p. [172](#)).
- Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA*, *102*(40), 14150–14154 (cit. on p. [16](#)).
- Yaglom, A. M. (1987). *Correlation theory of stationary and related random functions. Vol. I*. Springer-Verlag, New York. (Cit. on p. [51](#)).
- Yoshihara, K.-i. (1976). Limiting behavior of U -statistics for stationary, absolutely regular processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, *35*(3), 237–252 (cit. on pp. [57](#), [154](#), [273](#)).
- Yu, B. (1994). Rates of convergence for empirical processes of stationary mixing sequences. *Ann. Probab.*, *22*(1), 94–116 (cit. on pp. [346](#), [347](#), [357](#)).
- Yuan, A., Giurcanu, M., Luta, G., & Tan, M. T. (2017). U -statistics with conditional kernels for incomplete data models. *Ann. Inst. Statist. Math.*, *69*(2), 271–302 (cit. on p. [302](#)).
- Yuan, M., & Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika*, *94*(1), 19–35 (cit. on p. [1](#)).
- Zhang, D., & Li, G. (1996). Bootstrapping generalized U -processes and V -processes and their applications in projection pursuit. *Acta Math. Appl. Sinica (English Ser.)*, *12*(4), 354–362 (cit. on pp. [115](#), [147](#), [150](#)).
- Zhou, Z., Mentch, L., & Hooker, G. (2021). V -statistics and variance estimation. *J. Mach. Learn. Res.*, *22*, Paper No. [287], 48 (cit. on p. [72](#)).
- Zhou, Z., & Wu, W. B. (2009). Local linear quantile estimation for nonstationary time series. *Ann. Statist.*, *37*(5B), 2696–2729 (cit. on pp. [18–20](#)).
- Zinn, J. (1985). Comparison of martingale difference sequences. *Probability in Banach spaces, V (Medford, Mass., 1984)* (pp. 453–457). Springer, Berlin. (Cit. on p. [352](#)).