Limit Theorems Beyond Sums of I.I.D Observations

Morgane Austern

Submitted in partial fulfillment of the

requirements for the degree

of Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2019

ABSTRACT

Limit Theorems Beyond Sums of I.I.D Observations

Morgane Austern

We consider second and third order limit theorems–namely central-limit theorems, Berry-Esseen bounds and concentration inequalities– and extend them for "symmetric" random objects, and general estimators of exchangeable structures.

At first, we consider random processes whose distribution satisfies a symmetry property. Examples include exchangeability, stationarity, and various others. We show that, under a suitable mixing condition, estimates computed as ergodic averages of such processes satisfy a central limit theorem, a Berry-Esseen bound, and a concentration inequality. These are generalized further to triangular arrays, to a class of generalized U-statistics, and to a form of random censoring. As applications, we obtain new results on exchangeability, and on estimation in random fields and certain network model; extend results on graphon models; give a simpler proof of a recent central limit theorem for marked point processes; and establish asymptotic normality of the empirical entropy of a large class of processes. In certain special cases, we recover well-known properties, which can hence be interpreted as a direct consequence of symmetry. The proofs adapt Stein's method.

Subsequently, we consider a sequence of–potentially random–functions (f_n) and a sequence of exchangeable structures (X_n). We show that, under general stability conditions, the random variables $f_n(X_n)$ are asymptotically normal. Those conditions are vaguely reminiscent of those familiar from concentration results, however not identical. We require that the output of the function f_n does not vary significantly when an entry is disturbed;

and the size of this variation should not depend markedly on the other entries. Our result generalizes a number of known results, and as corollaries, we obtain new results for several applications: For randomly sub-sampled subgraphs; for risk estimates obtained by K-fold cross validation; and for the empirical risk of double bagging algorithms. The proof adapts the martingale central-limit theorem.

Table of Contents

List of Figures

List of Tables

Acknowledgments

This dissertation owes much to excellent mentors and colleagues at Columbia University. I would like to express my sincere gratitude to my advisors Peter Orbanz and Arian Maleki for their continuous support and guidance. I am particularly thankful for their patience and reassuring words, which helped me greatly to conquer my insecurities. I would like to extend a special thanks to Peter Orbanz for introducing me to the beautiful world of amenable group actions; and to Arian Maleki for giving me my first research project. I am also extremely grateful to Wenda Zhou and Victor Veitch who enriched immensely my time at Columbia with stimulating conversations and collaborations.

I would like to thank the committee members, Sumit Mukherjee, Cynthia Rush, and Ryan P. Adams, for their willingness to serve on the committee and their thoughtful feedbacks. I am particularly thankful to Ryan P. Adams for his generosity and his insightful ideas drawing fascinating connections between machine learning and the work presented in this thesis.

I am grateful for the support from the Department of Statistics throughout my doctoral studies. It was a true privilege to be part of this institution and to be given the resources and time to write this dissertation.

On a personal note, I am forever indebted to my family for their incredible support and their unconditional love. I am blessed to have as parents Sylvie and Patrick, who gave me more than I could ever express. The never ending string of phone calls and messages, helped me tremendously to overcome the most lonely and stressful times of my studies. I could never have done it without them. I am also extremely thankful to my brother Madec for believing in me and supporting me for as long as I can remember. I would like to also acknowledge my grand parents for their constant love and happiness to see me. Neuchatel will always be for me synonym with happiness. Finally, my partner Nicholas deserves my deepest gratitude. He is my strongest advocate, the one I turn to whenever in doubt, and his words of love and support helped me immensely through sleepless nights. Thank you again for helping me manage the tumult of daily life whenever I feel overwhelmed.

To Catherine Rudaz who by her love marked forever my life;

To Robert Austern who I wished would have known me as an adult,
I will always regret our lost conversations.

Chapter 1

Introduction

Limit theorems are the theoretical foundation of statistical inference. For example, the law of large numbers guarantees the convergence of estimators, and the central-limit theorem allows one to build confidence intervals. These are standard tools for studying sums of independent and identically distributed observations. However, many quantities of interest do not fall in this category: for instance in estimation problems for networks and graphs. In addition many interesting objects in machine learning, such as the cross-validated risk, depend in a complex fashion on the observations. One would like to have universal theorems for distributionally "symmetric" dependent data and for general statistics beyond empirical averages. This is the object of this dissertation: to generalize the classical limit theorems to those general settings.

A random element X of a space $\mathbf{X}$ is *invariant* or *symmetric* if there is a group $\mathbb{G}$ of transformations of $\mathbf{X}$ such that $\phi X \stackrel{\mathrm{d}}{=} X$ for all $\phi \in \mathbb{G}$. Models characterized by transformation invariance have recently garnered considerable attention in statistics. They include graphon models [1,3,7,37,43], their relatives such as graphex models [8,14,56] and edge-exchangeable graphs [13, 19, 33], and various models for relational data and preference prediction used in machine learning e.g [46]. More classical examples are stationary time series, the exchangeable random partitions that underpin much of Bayesian nonparametrics e.g [32, 49], and rotation- and shift-invariant random fields [5,34]. All of these admit some form of canonical sample average that plays a role analogous to the empirical measure of an i.i.d. sample. In general for a given countable symmetry group $\mathbb{G}$, we choose a suitable sequence $\mathbf{A}_1, \mathbf{A}_2, \ldots$ finite subsets of $\mathbb{G}$, and define empirical averages as

$$\mathbb{F}_n(f, X) = \frac{1}{|\mathbf{A}_n|} \sum_{\phi \in \mathbf{A}_n} f(\phi X) \qquad \text{for any } f \in \mathbf{L}_1(X) \; . \tag{1.1}$$

These averages admit a common representation, and a result of ergodic theory (the pointwise theorem of [41]) implies that the estimators they define are generically consistent. However, higher order convergence results, characterizing the speed of convergence, were lacking. We prove, under mild moment and mixing conditions, a series of universal second and third order limit theorems: central-limit theorems, concentration inequalities and Berry-Esseen bounds. We then generalize those results to triangular arrays, a form of generalized U-statistics and random censoring. We apply this new set of tools to obtain several new results.

Many quantities in machine learning cannot be approximated by simple empirical averages —the cross-validated risk in a prediction framework or the empirical risk of double bagging algorithms are two such examples. Although concentration inequalities [9] [59] have been used in many cases to study the speed of convergence of different algorithms, they suffer from the major drawback that they are rarely tight. Indeed, they do not answer questions such as: does algorithm A converge β times faster than algorithm B? What is the asymptotic shape of the confidence region? To answer to those questions we would need to know the limiting distributions of the quantities of interest. Let Y be a random exchangeable structure and let $(F_n)_n$ be a sequence of—possibly random—functions. Our goal is to study the asymptotic distribution of $F_n(Y)$. Under some moment and stability conditions we prove that $F_n(Y)/\sqrt{n}$ is asymptotically Gaussian. In the case where $Y = (Y_i)$ is a sequence of independent observations the stability conditions formalize the notion that

$$\Delta_i(Y_1, \ldots, Y_n) := F_n(Y_1, \ldots, Y_n) - F_n(Y_1, \ldots, Y_{i-1}, Y_i', Y_{i+1}, \ldots, Y_n)$$

should not be too big if Y_i' is an independent copy of Y_i; and that

$$\Delta_i(Y_1, \ldots, Y_n) - \Delta_i(Y_1, \ldots, Y_{j-1}, Y_j', Y_{j+1}, \ldots, Y_n)$$

should be small for most j. We apply this tool to prove several new results, notably on the cross-validated risk.

1.1 Organization

Chapter 2 reviews quickly the classical limit theorems—that we propose to extend—as well as the literature upon which we build. We also present some background and preliminaries that will be used in later chapters. Notably some useful facts about group theory, group actions and invariant random objects are presented.

Chapter 3 centers around the limiting behavior of empirical averages of symmetric random objects. We present in this chapter the main results–universal central limit theorems, Berry Esseen bounds and concentration inequalities–as well as some extensions to triangular arrays, generalized U-statistics and random censoring. Then we present some applications to stationary random fields, exchangeable random structures, graphex and marked point processes. In addition, we approximate an uncountable symmetry group by discretization, and study the effect this approximation has on estimation.

Chapter 4 generalizes the limit theorems–notably the central-limit theorem– for functions of exchangeable structures. After presenting some corollaries for smooth functions of i.i.d sequences, we use these new tools to study estimation on randomly sampled subgraphs, the empirical risk of double bagging algorithms and the cross-validated risk of predictors.

Chapter 5 studies examples coming from information theory. We present a generalized notion of entropy for general invariant objects and prove that the empirical entropy is asymptotically Gaussian. We also study the limiting behavior of the Kolmogorov complexity, another important information theoretic quantity that has been proven in the case of stationary ergodic processes–to share a lot of similarities with the Shannon entropy.

Chapter 2

Background

In this chapter we present some of the necessary background. First we go over the classical limit theorems–that we propose to extend–and explain why they are fundamental. Then we present some useful facts about group theory, group actions and invariant random objects.

2.1 Classical limit theorems and some applications

Before extending the classical limit theorems to new settings, it is interesting to remember what they are and why they are fundamental to statistical inference.
Let $\boldsymbol{Y}$ be a standard Borel space, and let Y be a random element. The goal of statistics is to learn properties of the distribution of Y, as for example the expected value $\mathbb{E}(f(Y))$ for a specific function $f \in \mathbf{L}_1(Y)$. We observe (Y_i) independent and identically distributed (i.i.d) observations that have the same distribution as Y; and we use those observations to estimate the quantities of interest. An estimator $\hat{\theta}_n$ will be called an empirical average if there is a function $g : \boldsymbol{Y} \to \mathbb{R}$ such that $\hat{\theta}_n := \frac{1}{n}\sum_{i\leq n} g(Y_i)$. The law of large numbers guarantees convergence of those estimators.

Theorem 2.1. *If $g \in \mathbf{L}_1(Y)$ then the following holds:*

$$\frac{1}{n}\sum_{i\leq n} g(Y_i) \xrightarrow{a.s,\mathbf{L}_1} \mathbb{E}(g(Y))$$

The central-limit theorem gives the exact speed of convergence of $\hat{\theta}_n$ allowing us to draw asymptotic confidence intervals and rejection regions.

Theorem 2.2. *If $g \in \mathbf{L}_2(Y)$ then the following holds:*

$$\frac{1}{\sqrt{n}}\sum_{i\leq n}[g(Y_i) - \mathbb{E}(g(Y))] \xrightarrow{d} N(0,\sigma^2),$$

where $\sigma^2 := \mathrm{var(g(Y))}$.

To know the rate at which $\hat{\theta}_n$ become normal, we use Berry-Esseen bounds. This notably evaluate the precision of the asymptotic confidence intervals.

THEOREM 2.3. *If $g \in \mathbf{L}_3(Y)$ and d_W is the Wasserstein distance then there is a universal constant $C < \infty$ such that :*

$$d_W\Big(\frac{1}{\sqrt{n}\sigma}\sum_{i\leq n}[g(Y_i) - \mathbb{E}(g(Y))], N(0,\sigma^2)\Big) \leq \frac{C\mathbb{E}(g(Y)^3)}{\sigma^3\sqrt{n}}$$

where $\sigma^2 := \mathrm{var(g(Y))}$.

Finally, concentration inequalities give finite sample guarantees:

THEOREM 2.4. *If $g \in \mathbf{L}_\infty(Y)$ then the following holds :*

$$P\Big(\big|\frac{1}{n}\sum_{i\leq n} g(Y_i) - \mathbb{E}(g(Y))\big| \geq t\Big) \leq 2e^{-\frac{nt^2}{2\|g(Y)\|_{\mathbf{L}_\infty}}}, \quad \forall t > 0.$$

These theorems are **fundamental** to statistical inference theory. Because of their importance, different extensions have been proposed: for non i.i.d observations and for more general estimators than empirical averages. For the former, the most successful generalization proposed is for the law of large numbers, which has been extended to empirical averages of any random invariant objects. This theory is due to E.Linderstrauss and the result of a cumulation of work by Ornstein, Weiss, Furstenberg, and others [63]. We present it in Section 2.3.2. Central-limit theorems have also been proposed for certain specific invariant random objects: for stationary random fields [5], exchangeable sequences [12], jointly exchangeable arrays [1] etc. However there is no general theory giving conditions that guarantee asymptotic normality of those empirical averages. The same can be said about concentration inequalities: some extensions exist for stationary processes cite but no general theory exists. Limit theorems have been studied for different types of statistics beyond empirical averages. U-statistics are arguably very similar to empirical averages, and are very important in estimation theory. They have been therefore extensively studied and many limit theorems have been proposed e.g [55], [39], [20], or [29]. Another well studied class of statistics are Lipschitz functions of i.i.d observations $f_n(X_1, \ldots, X_n)$. Many quantities of interest in machine learning and high-dimensional statistics are of this form, making this an especially important case to study. Concentration inequalities are widely adapted to this setting, see [9] for an extensive treatment. As this draft was nearing its completion it was pointed out to us that [16] proposed a very elegant central-limit theorem for a subclass of those statistics satisfying some additional stability conditions. However limit theorems for those more complex statistics have not been extensively studied beyond the case of i.i.d observations and mixing stationary sequences.

2.2 Basic background in group theory

A group $(\mathbb{G}, \cdot)$ is a set coupled with an operation rule—that we will call group multiplication–respecting four key propreties:

- The set $\mathbb{G}$ is closed under multiplication: $a \cdot b \in \mathbb{G}$ for all $a, b \in \mathbb{G}$.
- The multiplication operation is associative: $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for all $a, b, c \in \mathbb{G}$.
- There exists an element $e \in \mathbb{G}$–called the neutral element–such that $e \cdot a = a \cdot e = a$ for all $a \in \mathbb{G}$.
- For any element $a \in \mathbb{G}$ there is an element $a^{-1} \in \mathbb{G}$–called inverse–such that $aa^{-1} = a^{-1}a = e$.

Groups play an important role in many fields of mathematics from geometry to ergodic theory. We present some examples of well-known groups.

Examples.

- $(\mathbb{Z}, +)$ is a group with addition as group multiplication.
- $(\mathbb{R}, +)$ is a group with addition as group multiplication.
- Denote $\mathbb{S}(\mathbb{N})$ the set of permutations of $\mathbb{N}$ that changes at most a finite number of integers; then $(\mathbb{S}(\mathbb{N}), \circ)$ is a group with the composition operator as group multiplication.
- $(GL(n), \cdot)$, where $GL(n)$ is the set of $n \times n$ invertible matrices, is a group with the matrix multiplication operator as group multiplication.
- Real Lie groups are groups having the property of also being finite-dimensional real manifolds whose group multiplication and inverse functions are smooth maps.

A topological group is a group with a topology with respect to which the multiplication $\mathbb{G} \times \mathbb{G} \to \mathbb{G}$ and inverse map $\mathbb{G} \to \mathbb{G}$ are continuous. Many groups encountered in mathematics are topological. We will want to restrict ourselves to a subset of topological groups called locally compact, second-countable (lcsc) groups. Those will have desirable properties for reasoning about statistics and probability. A group is locally compact if every element $g \in \mathbb{G}$ admits a compact neighborhood. We define $\mathcal{B}(\mathbb{G})$ the sigma-field generated by all the open sets of $\mathbb{G}$ and we call it the Borel sigma-field of $\mathbb{G}$. The following proprety will be key throughout.

Proposition 2.5. *All locally compact and Hausdroff groups admits a left-invariant measure* $|\cdot|$ *called* Haar-measure*:*

$$|gG| = |G|, \qquad \forall g \in \mathbb{G},\ G \in \mathcal{B}(\mathbb{G}).$$

Such a measure is unique up to a multiplicative factor.

If $\mathbb{G}$ is countable, we always assume the topology to be discrete, and Haar measure is the counting measure: Compact subsets are finite, and $|\bullet|$ is the cardinality. A topological group will in addition be second countable if it admit a countable subset that is dense for the group topology.

PROPOSITION 2.6. *All locally compact and second countable groups are complete and admit a left-invariant metric* $d : \mathbb{G} \times \mathbb{G} \to \mathbb{R}^+$ *such that*

$$d(g \cdot g_1, g \cdot g_2) = d(g_1, g_2), \quad \forall g, g_1, g_2 \in \mathbb{G}.$$

EXAMPLES. Many groups are lcsc. Some examples are:

- The group of reals and the group of integers.
- Connected Lie groups.
- All discrete (countable) groups.

From now on, we will will ommit the notation $\cdot$ when referring to the group multiplication and abbreviate $a \cdot b$ into ab.

2.3 Group actions and invariant distributions

Throughout, $\mathbf{X}$ denotes a standard Borel space and $\mathcal{B}(\mathbf{X})$ its Borel sets. For a random element X of $\mathbf{X}$ and $p > 0$, we denote by $\mathbf{L}_p(X)$ the set of measurable functions $f : \mathbf{X} \to \mathbb{R}$ satisfying $\mathbb{E}[|f(X)|^p] < \infty$.

2.3.1 Invariance

A **measurable action** of $\mathbb{G}$ on $\mathbf{X}$ is a jointly measurable map

$$(\phi, x) \mapsto \phi(x) \quad \text{with} \quad (\phi\phi')(x) = \phi(\phi'(x)) \;\; \text{and} \;\; ex = x \;\; \text{for all } \phi, \phi' \in \mathbb{G}\,,$$

where e denotes the identity element of $\mathbb{G}$. Given a measurable action of $\mathbb{G}$, a probability measure P on $\mathcal{B}(\mathbf{X})$ is **$\mathbb{G}$-invariant** if $P \circ \phi^{-1} = P$ for all $\phi \in \mathbb{G}$. Similarly, a random element X of $\mathbf{X}$ is $\mathbb{G}$-invariant if its law is, that is, if $\phi(X) \overset{\mathrm{d}}{=} X$ for all $\phi \in \mathbb{G}$. As we will see, certain statistical properties of $\mathbb{G}$-invariant random elements depend only on the group, rather than the space $\mathbf{X}$ or the chosen action.

EXAMPLES. (i) For $n \in \mathbb{N}$, denote by $\mathbb{S}_n$ the group of permutations of n elements, i.e. of all bijections $\mathbb{N} \to \mathbb{N}$ that leave all but the first n numbers invariant. The set $\mathbb{S}_\infty := \cup_{n\in\mathbb{N}} \mathbb{S}_n$ of all permutations of $\mathbb{N}$ with finite support is an lcsc group, the **finitary symmetric group**. Let $\mathbf{X}$ be the space $\mathbb{R}^\mathbb{N}$ of real-valued sequences. For any such sequence $x = (x_1, x_2, \ldots)$, define a measurable action as $\phi(x) := (x_{\phi(1)}, x_{\phi(2)}, \ldots)$ for every $\phi \in \mathbb{S}_\infty$. An $\mathbb{S}_\infty$-invariant random sequence $X = (X_1, X_2, \ldots)$ is then called an **exchangeable sequence**. The qualifier *infinitely exchangeable* is sometimes used, especially in Bayesian statistics, to distinguish from invariance under finite sets of permutations.

(ii) Let $\mathbf{X} = \mathbb{R}^{\mathbb{Z}^d}$, for some dimension $d \in \mathbb{N}$. A random element $X = (X_i)_{i\in\mathbb{Z}^d}$ of $\mathbf{X}$ is hence a real-valued random sequence (if $d = 1$) or a random field on a grid (if $d \geq 2$). We choose the lcsc group $\mathbb{G} = \mathbb{Z}^d$, and the action

$$\phi(x) := (x_{i+\phi})_{i\in\mathbb{Z}^d} \qquad \text{for any } x = (x_i) \in \mathbf{X}, \phi \in \mathbb{G} .$$

A $\mathbb{G}$-invariant random element X of $\mathbf{X}$ is called **stationary**.

(iii) Let Ω be a standard Borel space and $\mathbf{X} := \Omega^{\mathbb{N}^d}$, for some $d \in \mathbb{N}$, endowed with the product topology and its Borel sets. Elements $x = (x_{i_1,\ldots,i_d})_{i_1,\ldots,i_d\in\mathbb{N}}$ of $\mathbf{X}$ are called a ***d*-arrays**. An action of $\mathbb{S}_\infty$ on d-arrays can be defined as

$$(\phi, x) = (x_{\phi(i_1),\ldots,\phi(i_d)})_{i_1,\ldots,i_d\in\mathbb{N}} \quad \text{for all } x \in \mathbf{X}, \phi \in \mathbb{S}_\infty . \tag{2.1}$$

If $d = 1$, x is a sequence, and the action coincides with that in Example (i). For $d = 2$, x is an infinite matrix, and the action permutes rows and columns (both by the same permutation). A random d-array X invariant under this action is called **jointly exchangeable**. If X is instead invariant under the action of $\mathbb{S}_\infty^d$ defined by

$$((\phi_1, \ldots, \phi_d), x) = (x_{\phi_1(i_1),\ldots,\phi_d(i_d)}) \quad \text{for all } x \in \mathbf{X}, \phi_1, \ldots, \phi_d \in \mathbb{S}_\infty \tag{2.2}$$

it is called **separately exchangeable**.

2.3.2 Laws of large numbers

Let $\mathbf{A}_1, \mathbf{A}_2, \ldots \subset \mathbb{G}$ be a sequence of compact subsets of $\mathbb{G}$. We define

$$\mathbb{F}_n(f, x) := \frac{1}{|\mathbf{A}_n|} \int_{\mathbf{A}_n} f(\phi x)|d\phi| ,$$

and call $(\mathbb{F}_n)$ the **empirical measure** defined by $(\mathbf{A}_n)$. If $\mathbb{G}$ is discrete, $|\bullet|$ is the set size, the sets $\mathbf{A}_n$ are finite, and $\mathbb{F}_n$ is the sum (1.1). We first ask for a law of large numbers: Under what conditions can we assume $\mathbb{F}_n(f, X) \to \mathbb{E}[f(X)]$ a.s. as $n \to \infty$, for some function f? To give a general answer, the expectation must be replaced by conditional expectation, defined as follows. For a given measurable action of $\mathbb{G}$ on $\mathbf{X}$, we denote the set of all $\mathbb{G}$-invariant probability measures on $\mathbf{X}$ by $\mathbf{I}_\mathbb{G}$. A Borel set $A \in \mathcal{B}(\mathbf{X})$ is **almost invariant** if $P(A \triangle \phi A) = 0$ for all $P \in \mathbf{I}_\mathbb{G}$ and all $\phi \in \mathbb{G}$, where $\triangle$ denotes symmetric difference. The collection $\sigma(\mathbb{G})$ of all almost invariant sets is a σ-algebra. We abbreviate conditioning on

$\sigma(\mathbb{G})$ as

$$\mathbb{E}[\bullet|\mathbb{G}] := \mathbb{E}[\bullet|\sigma(\mathbb{G})] \quad \text{and} \quad P(\bullet|\mathbb{G}) := P(\bullet|\sigma(\mathbb{G})) .$$

The relevant law of large numbers for a $\mathbb{G}$-invariant variable X is then

$$\mathbb{F}_n(f, X) \xrightarrow{n\to\infty} \mathbb{E}[f(X)|\mathbb{G}] \quad \text{a.s. for all } f \in \mathbf{L}_1(X) . \tag{2.3}$$

For (2.3) to hold, one must certainly require

$$\frac{|\mathbf{A}_n \cap K\mathbf{A}_n|}{|\mathbf{A}_n|} \quad \xrightarrow{n\to\infty} \quad 1 \qquad \text{for all compact subset } K \subset \mathbb{G} . \tag{2.4}$$

A sequence $(\mathbf{A}_n)$ of compact sets satisfying (2.4) is called a **Følner sequence**, and $\mathbb{G}$ is **amenable** if such a sequence exists. A Følner sequence is **tempered** if

$$\Big|\bigcup_{k<n} \mathbf{A}_k^{-1}\mathbf{A}_n\Big| \le c|\mathbf{A}_n| \qquad \text{for some } c > 0 \text{ and all } n \in \mathbb{N} . \tag{2.5}$$

Not every Følner sequence is tempered, but every group containing a Følner sequence also contains one that is tempered [41]. The pointwise theorem for amenable groups, the culmination of a long line of work by Ornstein, Weiss, Furstenberg, and others [63], states that the necessary condition (2.4) is essentially sufficient:

Theorem 2.7 (E. Lindenstrauss [41]). *If X is invariant under a measurable action of an amenable lcsc group, and $f \in \mathbf{L}_1(X)$, the empirical measure defined by a tempered Følner sequence satisfies* (2.3)*.*

Examples. (iv) The group $\mathbb{S}_\infty$ is amenable, and choosing $\mathbf{A}_n := \mathbb{S}_n$ defines a tempered Følner sequence. Its empirical measure is $\frac{1}{n!}\sum_{\phi\in\mathbb{S}_n} f(\phi x)$.

(v) Consider a stationary random field on the grid $\mathbb{Z}^2$, as in Example (ii). Stationarity is invariance under the group $\mathbb{G} = \mathbb{Z}^2$, acting by addition. The sets $\mathbf{A}_n = \{(i,j)\,|\,|i|,|j| \le n\}$ form a tempered Følner sequence for $\mathbb{G}$. The image under $\mathbf{A}_n$ of a fixed point in the index set $\mathbb{Z}^2$, say $(0,0)$, is the subgrid $\Omega_n := \mathbf{A}_n(0,0) = \{-n,\ldots,n\}^2$, see Figure 2.1. Condition (2.4) hence implies

$$|\partial\Omega_n|\,/\,|\Omega_n| \xrightarrow{n\to\infty} 0 \quad \text{where} \quad \partial\Omega_n = \Omega_n \backslash \Omega_{n-1} , \tag{2.6}$$

which is the standard condition used in limit theorems for random fields to control dependence between disjoint regions of the grid [5].

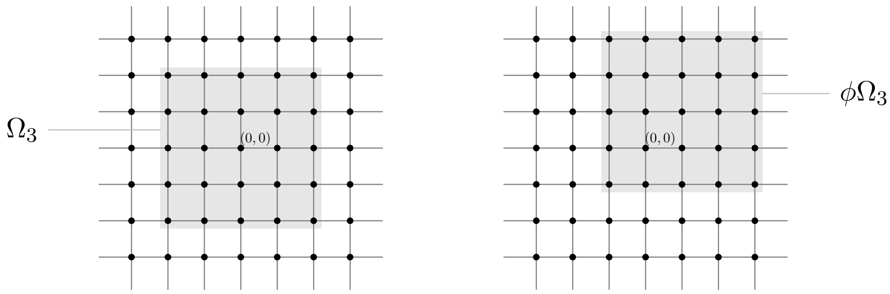

Figure 2.1: The amenability condition (2.4) illustrated on the grid $\Omega := \mathbb{Z}^2$. The shift group on Ω is $\mathbb{G} = \mathbb{Z}^2$, acting by addition. The centered subgrid Ω_n of radius n can be represented as $\Omega_n = \mathbf{A}_n(0,0)$, and $|\partial\Omega_n| \approx |\mathbf{A}_n \triangle \phi\mathbf{A}_n|$ for $\phi = (+1,+1)$.

2.3.3 Ergodicity

Theorem 2.7 establishes convergence to conditional expectations $\mathbb{E}[\bullet|\mathbb{G}]$, and hence to expectations under the conditional probability $P(\bullet|\mathbb{G})$. To clarify the meaning of this result, we must characterize the conditionals, and that requires the concept of ergodic measures. Ergodicity is perhaps most familiar in the context of discrete-time stochastic processes: Informally, an ergodic process "eventually forgets where it came from". More formally, a stationary process is ergodic if every shift-invariant event occurs with probability either 0 or 1 [36, 51]. To generalize this idea, we substitute shift-invariance by $\mathbb{G}$-invariance: A probability measure P is **ergodic** if

$$\text{(i) } P \text{ is } \mathbb{G}\text{-invariant} \quad \text{ and } \quad \text{(ii) } P(A) \in \{0,1\} \text{ whenever } A \in \sigma(\mathbb{G}) ,$$

see [24, 44]. The relevant characterization of the conditionals is the ergodic decomposition theorem, which states that each realization of $P(\bullet|\mathbb{G})$ is ergodic. One part of that statement is trivial: By the elementary properties of regular conditional probabilities, $P(\bullet|\mathbb{G})$ satisfies (ii) almost surely, since it is $\sigma(\mathbb{G})$-measurable. Whether (i) holds is a much harder question, however, and need not be the case for arbitrary groups, but does hold if $\mathbb{G}$ is lcsc.

THEOREM 2.8 (Ergodic decomposition [36, 44]). *Let $\mathbb{G}$ be an lcsc group acting measurably on a standard Borel space $\mathbf{X}$. The set $\mathbf{I}_{\mathbb{G}}$ of invariant measures is convex, and an invariant measure is an extreme point of $\mathbf{I}_{\mathbb{G}}$ if and only if it is ergodic. A probability measure P on $\mathbf{X}$ is $\mathbb{G}$-invariant if and only if there exists a random ergodic measure ξ_P, i.e. a random variable with values in the set of extreme points of $\mathbf{I}_{\mathbb{G}}$, such that*

$$P[\bullet|\mathbb{G}] = \xi_P(\bullet) \qquad a.s. \tag{2.7}$$

Any invariant random element X can, according to (2.7), be generated in two steps: By first selecting an ergodic measure ξ_{P} at random, followed by a draw $X|\xi_{\mathrm{P}} \sim \xi_{\mathrm{P}}$ from this measure. If f is a measurable function, then

$$\mathbb{E}[f(X)|\mathbb{G}] =_{\text{a.s.}} \xi_{\text{P}}(f) \quad \text{ and, if } X \text{ is ergodic,} \quad \mathbb{E}[f(X)|\mathbb{G}] =_{\text{a.s.}} \mathbb{E}[f(X)] \ .$$

The example most common in statistics is de Finetti's theorem:

Example. (vi) Let $X = (X_1, X_2, \ldots)$ be an exchangeable sequence, with entries in some standard Borel space $\mathbf{X}_0$. By the Hewitt-Savage 0–1 law [35], X is ergodic if and only if it is i.i.d. There is hence a random probability measure μ_{P} on $\mathbf{X}_0$ such that (2.7) takes the form $P(\bullet|\mathbb{G}) = \mu_{\text{P}}^{\infty}(\bullet)$ a.s. That implies the well-known de Finetti integral identity, namely

$$P(X \in \bullet) = \mathbb{E}[P(\bullet|\mathbb{G})] = \int_M m^{\infty}(\bullet) Q_{\text{P}}(dm) \ , \tag{2.8}$$

where M is the set of probability measures on $\mathbf{X}_0$, and Q_{P} the law of μ_{P}.

The case of de Finetti's theorem helps to illustrate implications for statistics: Suppose the invariant random element X is an "infinitely large" random object, and that the function f depends only on a "small part" of X. Such a small part may be a single entry (if X is an infinite sequence), a single vertex (for infinite graphs), or a finite-size patch (if X is a continuous random field). Applying different elements ϕ of the group $\mathbb{G}$ to X moves different parts of X within the domain of f, and hence $\mathbf{F}_n(f, X)$ averages over different parts of X.

Examples. (vii) Let X be a stationary Markov chain with a countable set S of states, indexed by $\mathbb{Z}$. (That is, we define symmetry as shift-invariance as in Example (ii), where $\mathbf{X} = S^{\mathbb{Z}}$ and $d = 1$, and consider only those invariant sequences X that additionally satisfy the Markov property). Then X is ergodic if it is irreducible and aperiodic [35].

(viii) Let X be a jointly exchangeable array, as in Example (iii). Such arrays are characterized by the Aldous-Hoover theorem [36] as follows: For simplicity, let $d = 2$. Let $\mathbf{H}$ be the set of measurable functions $[0,1]^3 \to \Omega$, and let U_i and U_{ij}, for $i, j \in \mathbb{N}$, be independent, uniform random variables in $[0,1]$. For any $h \in \mathbf{H}$, the random array $X_h := (h(U_i, U_j, U_{ij}))_{i,j \in \mathbb{N}}$ is jointly exchangeable. The Aldous-Hoover theorem shows that a jointly exchangeable array X is ergodic if and only if $X \stackrel{\text{d}}{=} X_h$ for some $h \in \mathbf{H}$. For $d > 2$, the function h is more generally of the form $[0,1]^{2^d - 1} \to \Omega$, see [36]. For applications of exchangeable arrays in statistics and machine learning, see [46].

(ix) Let $\mathbf{G}$ be the set of all undirected, simple graphs with vertex set $\mathbb{N}$. An **exchangeable graph** is a random element X of $\mathbf{G}$ whose distribution is invariant under arbitrary permutations of the vertex set by elements of $\mathbb{S}_{\infty}$ [21]. A **graphon** is a measurable function $w : [0,1]^2 \to [0,1]$ that satisfies $w(u, u') = w(u', u)$, for all $u, u' \in [0,1]$ [7,43]. Let $\mathbf{W}$ be the set of all such functions. Any graphon w defines an exchangeable graph X_w as

$$X_w := (X_{ij})_{i<j\in\mathbb{N}} \quad \text{where} \quad X_{ij} := \mathbb{I}\{U_{ij} \leq w(U_i, U_j)\} ,$$

for a family of uniform variables U_i, U_{ij} as in the previous example. An exchangeable graph X is ergodic if and only if it is generated by a graphon, that is, if $X \stackrel{\mathrm{d}}{=} X_w$, for some $w \in \mathbf{W}$. Informally, that follows from Example (viii): A graph can be identified with its adjency matrix, which is a 2-array with $\Omega = \{0, 1\}$, and X is exchangeable iff its adjency matrix is jointly exchangeable. It is therefore ergodic iff $X \stackrel{\mathrm{d}}{=} X_h$ for some $h \in \mathbf{H}$. Since the adjacency matrix is symmetric, h is symmetric in its first two arguments. The function $w(u, u') := \int_{[0,1]} h(u, u', v)dv$ is hence a graphon, and $X_w \stackrel{\mathrm{d}}{=} X_h$. See [21] for a rigorous argument. In statistics, the term *graphon model* refers to a family of distributions $\{P_w | w \in \mathcal{W}\}$, where P_w is the law of X_w, and $\mathcal{W}$ a suitable subset of $\mathbf{W}$. In [25, 37], for example, $\mathcal{W}$ contains only Hölder-smooth elements of $\mathbf{W}$.

Chapter 3

Limit theorems for invariant distributions

3.1 Assumptions

Throughout, we impose the following assumptions: $\mathbb{G}$ always denotes a lcsc group that is amenable (as defined in Section 2.3.2) and acts measurably on a standard Borel space $\mathbf{X}$. A metric d on $\mathbb{G}$ is **left-invariant** if $d \circ (\phi \otimes \phi)^{-1} = d$ for all $\phi \in \mathbb{G}$. For a subset G of $\mathbb{G}$, we denote by $\mathbf{B}_t(G) = \{\phi \in \mathbb{G} | \min_{\phi' \in G} d(\phi, \phi') \leq t\}$ the d-ball of radius t around G. For the identity element e of $\mathbb{G}$, we abbreviate $\mathbf{B}_t := \mathbf{B}_t(e)$. Every amenable lcsc group admits a left-invariant metric satisfying

$$\tfrac{|\mathbf{B}_{n+1} \setminus \mathbf{B}_n|}{|\mathbf{B}_n \setminus \mathbf{B}_{n-1}|} = O(1) , \tag{3.1}$$

and we always assume d to have this property.

Table 3.1: Examples of invariant objects

invariant objects X	ergodic measures	eq. (2.8) specializes to	eq. (2.3) specializes to
exchangeable sequences	i.i.d. sequences	de Finetti's theorem	law of large numbers
stationary time series	ergodic processes		Birkhoff's theorem
exchangeable graphs	graphon models	Aldous-Hoover theorem	graph limit convergence
graphs generated by inv. point processes	graphex models	Kallenberg's representation theorem [36]	empirical graphex
exchangeable arrays	dissociated arrays	Aldous-Hoover theorem	Kallenberg's LLN

3.2 Mixing

Lindenstrauss' theorem requires existence of a first moment, as one would expect for a law of large numbers. To formulate central limit theorems, we must strengthen the hypothesis to (i) a second-moment condition on $f(X)$, and (ii) a mixing condition. In the context of mixing, it can be helpful to think of $(f(\phi X))_{\phi\in\mathbb{G}}$ as a real-valued stochastic process indexed by $\mathbb{G}$. Informally, mixing means that, if two subsets G and G' of $\mathbb{G}$ are "far away" from each other, the segments $(f(\phi X))_{\phi\in G}$ and $(f(\phi X))_{\phi\in G'}$ of the process are approximately independent.

To make this notion precise, consider some subset G of $\mathbb{G}$. The set of events in $\mathbf{X}$ that can be formulated in terms of the segment $(f(\phi X))_{\phi\in G}$ is the σ-algebra $\sigma_f(G) = \sigma(f\circ\phi, \phi\in G)$. Since distance in $\mathbb{G}$ is measured by a metric d as defined above, the set of group elements with distance larger than t from G is $\mathbb{G}\setminus\mathbf{B}_t(G)$. For $t>0$, we hence define

$$\mathcal{C}(t) := \big\{(A,B)\in\sigma_f(\phi_1,\phi_2)\otimes\sigma_f(G)\big|G\subset\mathbb{G},\phi_1,\phi_2\in\mathbb{G}\setminus\mathbf{B}_t(G)\big\}\ . \tag{3.2}$$

We then call the function $\alpha:\mathbb{R}_{\geq 0}\to(0,\infty)$ defined by

$$\alpha(t) := \sup_{(A,B)\in\mathcal{C}(t)} |P(A)P(B)-P(A\cap B)|$$

the **mixing coefficient** for f and P, and P is **mixing** with respect to f if $\alpha(t)\to 0$ as $t\to\infty$. The function $\alpha(\bullet|\mathbb{G})$ given by

$$\alpha(t|\mathbb{G}) := \sup_{(A,B)\in\mathcal{C}(t)} \mathbb{E}[|P(A|\mathbb{G})P(B|\mathbb{G})-P(A\cap B|\mathbb{G})|]$$

is the **conditional mixing coefficient**, and P is **conditionally mixing** if $\alpha(t|\mathbb{G})\to 0$ as t diverges. The next lemma shows that $\alpha(t|\mathbb{G})=0$ is never a stronger condition than $\alpha(t)=0$. Example (xii) below shows it is indeed strictly weaker. We measure the "spread" of a subset $G\subset\mathbb{G}$ within the group by defining $N(G) := \min\{d(\phi,\phi')|\phi,\phi'\in G \text{ distinct}\}$.

Lemma 3.1. *(i) The mixing coefficients satisfy $\alpha(n|\mathbb{G})\leq 2\alpha(n)$ for all $n\in\mathbb{N}$. (ii) Let $\mathbb{G}$ be discrete. Then $\alpha(n|\mathbb{G})=0$ holds if and only if, for all subsets $G\subset\mathbb{G}$ with $N(G)\geq n$, the law of the sequence $(f(\phi_i X))_{\phi_i\in G}$ does not depend on the enumeration $\phi_1,\phi_2,\dots$ of G.*

Examples. (x) Arguably the most commonplace notion of mixing is that for stationary discrete-time processes [10]. Let $(\zeta_n)_{n\in\mathbb{Z}}$ be a stationary, real-valued stochastic process. Mixing requires that any initial segment $(\zeta_i)_{i\leq k}$ and the tail $(\zeta_i)_{i\geq k+t}$ become approximately independent as t grows large. That is formalized by requiring

$$|P(A\cap B)-P(A)P(B)| \xrightarrow{t\to\infty} 0 \ \text{ if } A\in\sigma(\zeta_1,\dots,\zeta_k), B\in\sigma((\zeta_i)_{i\geq k+t}).$$

The sequence $(\alpha_t)_{t\in\mathbb{N}}$ with

$$\alpha_t := \sup_{k\in\mathbb{N},A,B} |P(A\cap\phi_t^{-1}B) - P(A)P(B)| \tag{3.3}$$

is called the alpha-mixing coefficient.
(xi) A stationary Markov random field X is a stationary random field $X = (X_i)_{i\in\mathbb{Z}^d}$, as in Example (ii), that has the Markov property

$$P(X_G \in \bullet\,|X_{\mathbb{Z}^d\setminus G}) = P(X_G \in \bullet\,|X_{\{i\in\mathbb{Z}^d|\, \min_{j\in G} d(i,j)=1\}}) \qquad \text{for all } G \subset \mathbb{Z}^d\,.$$

To average over a stationary field, f is typically chosen as the coordinate function $f : (x_i)_{i\in\mathbb{Z}^d} \mapsto x_0$ at the origin $0 \in \mathbb{Z}^d$. Assume that X satisfies

$$a_\phi := \sup_{i|d(i,0)=1} \sup_{A,B\in\mathcal{B}(\mathbf{X})} |P(X_0 \in A|X_i \in B) - P(X_0 \in A)| \leq \frac{1}{2d}\,, \tag{3.4}$$

also known as the Dobrushin condition [26]. Then X is mixing with respect to f. That is straightforward to verify for $d = 1$: Condition (3.4) implies

$$\begin{aligned}
&|P(X_0\in A|X_{-t}\in B) - P(X_0\in A)| \\
&\leq |\mathbb{E}[P(X_0\in A|X_{-1}) - P(X_0\in A)|X_{-t} \in B] - \mathbb{E}[P(X_0\in A|X_{-1}) - P(X_0\in A)]| \\
&\leq 2\alpha(t-1)a_\phi \qquad \text{for all } t \in \mathbb{N},
\end{aligned}$$

where expectations are taken with respect to X_{-1}. There are hence positive constants c_1 and c_2 such that $\alpha(t) \leq c_1 k e^{-c_2 t}$ for all $t \in \mathbb{N}$, i.e. the mixing coefficient decays exponentially. The same holds for $d \geq 2$, by a more technical argument [26].

(xii) Let $X = (X_1, X_2, \ldots)$ be a real-valued random sequence, and f the first coordinate function $f : (x_1, x_2, \ldots) \mapsto x_1$. If X is exchangeable, it need not be mixing with respect to f, but it is conditionally mixing. To illustrate the difference, generate a sequence by drawing a random value X_1 from some distribution on $\mathbb{R}$, and set $X_2 = X_3 = \ldots = X_1$. Then X is exchangeable; in terms of de Finetti's theorem, its entries are drawn conditionally independently from the point mass δ_{X_1}, given X_1. Clearly, X_1 and X_t do not become less dependent as t grows. However any exchangeable sequence is conditionally mixing with respect to f: For any subsets $F, G \subset \mathbb{Z}$, conditional independence of the entries implies

$$(X_i)_{i\in F} \perp\!\!\!\perp_{\mathbb{G}} (X_j)_{j\in G} \qquad \text{whenever } F \cap G = \emptyset\,,$$

and hence $\alpha(t|\mathbb{G}) = 0$ for all $t > 0$.

3.3 Limit theorems

Two hypotheses will repeatedly appear in our results: We either require

$$\text{(i)}\;\; \mathbb{E}[f(X)^2] < \infty \qquad\qquad \text{(ii)}\;\; \alpha(n|\mathbb{G}) = 0 \quad \text{for some } n \in \mathbb{N}\,, \tag{3.5}$$

or that there exists an $\varepsilon > 0$ such that

$$\text{(i)}\;\; \mathbb{E}[f(X)^{2+\varepsilon}] < \infty \qquad\qquad \text{(ii)}\;\; \int_{\mathbb{G}} \alpha(d(e,\phi)|\mathbb{G})^{\frac{\varepsilon}{2+\varepsilon}}|d\phi| < \infty\,, \tag{3.6}$$

where e is the identity element of $\mathbb{G}$. If $\mathbb{G}$ is countable, (3.6ii) becomes

$$\sum_{n\in\mathbb{N}} |\mathbf{B}_{n+1}\backslash\mathbf{B}_n|\,\alpha(n)^{\varepsilon/(2+\varepsilon)} < \infty\,.$$

3.3.1 Results

To keep expressions simple, we assume that f is centered, in the sense that $\mathbb{E}[f(X)|\mathbb{G}] = 0$, and write $\overline{\mathbf{L}}_1(X)$ for the set of all such centered $f \in \mathbf{L}_1(X)$. Since conditions (3.5) and (3.6) both imply that $f \in \mathbf{L}_1(X)$, and $\mathbb{E}[f(X)|\mathbb{G}] = 0$ can be centered for any $\mathbf{L}_1$-function, this constitutes no loss of generality. We define an element-wise and a total variance as

$$\eta^2(\phi) := \mathbb{E}[f(X)f(\phi X)|\mathbb{G}] \qquad \text{for } \phi \in \mathbb{G} \text{ and} \qquad \eta^2 := \int_{\mathbb{G}} \eta^2(\phi)|d\phi|\,.$$

THEOREM 3.2. *Let $\mathbb{F}_n$ be the empirical measure defined by a tempered Følner sequence in $\mathbb{G}$, X a $\mathbb{G}$-invariant random element, and $f \in \overline{\mathbf{L}}_1(X)$. If either* (3.5) *or* (3.6) *is satisfied, $\mathbb{F}_n(f,X)$ has a normal limit*

$$\sqrt{|\mathbf{A}_n|}\;\mathbb{F}_n(f,X) \xrightarrow{d} \eta Z \qquad \textit{for } Z \sim N(0,1)\,. \tag{3.7}$$

The asymptotic variance η^2 is a random variable that is almost surely finite, and can be chosen independent of Z.

The rate of convergence in Lindenstrauss' theorem is thus $|\mathbf{A}_n|^{-\frac{1}{2}}$, and depends only on the choice of Følner sequence and hence on the group; it does not depend on the action, or on the space $\mathbf{X}$. The central limit theorem is complemented by a Berry-Esseen type bound, which quantifies how closely the nth sample average in the central limit theorem resembles the limiting normal law. The Wasserstein metric d_{W} of order 1 is used to compare laws e.g [50]. We decompose the integral in (3.6ii) as

$$\tau(r) := \int_{\mathbb{G}\backslash\mathbf{B}_r} \alpha_2(d(e,\phi)|\mathbb{G})^{\frac{\varepsilon}{2+\varepsilon}}|d\phi|\,,$$

so that condition (3.6ii) takes the form $\tau(0) < \infty$.

THEOREM 3.3. *Let $\mathbb{F}_n$ be the empirical measure defined by a tempered Følner sequence in $\mathbb{G}$, X $\mathbb{G}$-invariant, $f \in \overline{\mathbf{L}}_1(X)$, and Z an independent standard normal variable. Abbreviate $s_p := \|f(X)/\eta\|_p$. If (3.5) holds,*

$$d_{\mathrm{W}}\Big(\frac{\sqrt{|\mathbf{A}_n|}}{\eta}\mathbb{F}_n(f,X),Z\Big) \;\leq\; \frac{\kappa_1}{\sqrt{|\mathbf{A}_n|}} + \kappa_2\frac{|\mathbf{A}_n \triangle \mathbf{B}_k\mathbf{A}_n|}{|\mathbf{A}_n|}$$

for constants κ_1 of order $O(s_4^3|\mathbf{B}_k|^2)$ and $\kappa_2 = O(s_2^2)$. If (3.6) holds instead for some $\varepsilon > 0$, then for any sequence $0 < b_1 < b_2 < \ldots$,

$$d_{\mathrm{W}}\Big(\frac{\sqrt{|\mathbf{A}_n|}}{\eta}\mathbb{F}(f,X),Z\Big) \leq \kappa_3\tau(b_n) + \frac{\kappa_4|\mathbf{B}_{b_n}|}{\sqrt{|\mathbf{A}_n|}} + \kappa_3\frac{|\mathbf{A}_n| - |\mathbf{A}_n \cap \mathbf{B}_{b_n}\mathbf{A}_n|}{|\mathbf{A}_n|}$$

for constants $\kappa_3 = O(s_{2+\varepsilon}^2)$ and $\kappa_4 = O(s_{4+2\varepsilon}^3\tau(0))$. In either case, the asymptotic variance η^2 is finite almost surely.

Observe that, if $\alpha_2(k|\mathbb{G}) = 0$ for some k and $\tau(\bullet)$ is hence bounded, the remaining quantity is of order $|\mathbf{A}_n|^{-1/2}$, which matches the Berry-Esseen bound $n^{-1/2}$ for n i.i.d. variables [50]. Condition (3.5) is the stronger one and results in a simpler bound. Why (3.6) complicates matters is explained for the discrete case by Lemma 3.1: The laws of sequences $(f(\phi_i X)_i)$ are not generally independent of the enumeration of the elements ϕ_i. The final term in either bound is a variance correction, required since the empirical measure is scaled by the overall standard deviation η, rather than that given by $\mathbf{A}_n$. The choice of the sequence (b_n) is subject to a trade-off, since $\tau(b)$ decreases with b, whereas $|\mathbf{B}_b|$ increases.

The bound may simplify if $\mathbb{G}$ has additional properties. For example:

COROLLARY 3.4. *Let $\mathbb{G}$ be a finitely generated, nilpotent group of rank r, and d the word metric with respect to a finite generator. Then $\mathbf{A}_n := B_n$ defines a tempered Følner sequence, and hence an empirical measure $\mathbb{F}_n$. Let X be an invariant random element, and $f \in \overline{\mathbf{L}}_1(X)$. If there exist $\varepsilon, \delta > 0$ such that $\alpha(n|\mathbb{G}) = O(n^{r+\delta})$ and $f(X)/\eta \in \mathbf{L}_{4+2\varepsilon}(X)$, then*

$$d_{\mathrm{W}}\Big(\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta}\mathbb{F}_n(f,X),Z\Big) = O(n^{-r/(2(r+\delta))}) \qquad \text{for } Z \sim N(0,1)\,.$$

The right-hand side decreases both in r, since the sets $\mathbf{A}_n$ grow with r, and in δ, since a larger value of δ implies stronger mixing. Informally, X is closer to being exchangeable for a larger value of δ, and as we will see in Section 3.6.2, the mixing condition vanishes entirely in the exchangeable case.

3.4 Generalizations

This section generalizes Theorems 3.2 and 3.3 in three ways: To triangular arrays of random variables, to generalized U-statistics, and to randomly subsampled averages. In more detail:

(1) Triangular arrays. The elementary central limit theorem of Lindeberg concerns the partial sums $\sum_{j\leq n}\zeta_j$ of an i.i.d. sequence (ζ_j) of random variables. It can be generalized to **triangular arrays** of the form (ζ_{ij}), where the sequence $(\zeta_{ij})_i$ is i.i.d. for each i: Under suitable conditions, central limit theorems hold for a diagonal sequence of partial sums $\sum_{j\leq n}\zeta_{nj}$. [35] treats i.i.d. central limit theorems comprehensively in terms of triangular arrays. In the invariant world, a single invariant random variable X_i assumes the role of the entire sequence (ζ_{ij}) above. We consider a sequence (X_i) of such variables—which all satisfy the same invariance, but need not have the same distribution—and extend our results to diagonal sequences of the form $f_n(X_n)$.

(2) Generalized U-statistics. If $\zeta := (\zeta_i)$ is an i.i.d. sequence and g a symmetric function with k arguments, the quantity

$$\frac{1}{n^k}\sum_{|i_1|,\ldots,|i_k|\leq n} g(\zeta_{i_1},\ldots,\zeta_{i_k})$$

is called a **U-statistic**, and it is well-known that such statistics generically satisfy central limit theorems e.g [55]. Define $\zeta^k := (\zeta_{i_1},\ldots,\zeta_{i_k})$ we can view U-statistics as empirical averages of ζ^k over $[-n,n]^k$. However the process ζ^k is not stationary on $\mathbb{Z}^d$, only invariant under joint translation: $t,\zeta^k \to (\zeta_{i_1+t},\ldots,\zeta_{i_k+t})$. Similarly in the general case, we consider standard Borel spaces $\mathbf{X}$ that admit an action of $\mathbb{G}^k$ and a random element X_k whose distribution is invariant under a strict subgroup of $\mathbb{G}^k$. We will be interested in studying empirical averages of X_k over $\mathbf{A}_n^k$.

(3) Randomized averages. Rather than averaging over all transformations in a Følner set $\mathbf{A}_n$, one might consider using only a random subset $\widehat{\mathbf{A}}_n$. In the discrete case, the empirical measure $\mathbb{F}_n$ then becomes

$$\widehat{\mathbb{F}}_n(f,x) \;=\; |\widehat{\mathbf{A}}_n|^{-1}\sum_{\phi\in\widehat{\mathbf{A}}_n} f(\phi x)\;, \tag{3.8}$$

and we also consider the non-discrete case below. Limit theorems hold under suitable conditions on the random sets $\widehat{\mathbf{A}}_n$.

3.4.1 Definitions

Throughout, we consider random element X_n of a standard Borel space $\mathbf{X}_n$, for $n\in\mathbb{N}$. On each space $\mathbf{X}_n$, we consider a function $g_n\in\mathbf{L}_1(\mathbf{X}_n)$. We ask how the sequence $g_n(X_n)$ converges as $n\to\infty$. To define symmetry, we again choose an amenable lcsc group $\mathbb{G}$, and now distinguish two types of actions: We fix a sequence (k_n) of integers, typically chosen such that $0 < k_1 \leq k_2 \leq \ldots$. For any such integer, the k_n-fold product $\mathbb{G}^{k_n} := \mathbb{G} \times \ldots \times \mathbb{G}$, endowed with the product topology, is again an lcsc group. If $(\mathbf{A}_m)$ is a tempered Følner sequence in $\mathbb{G}$, then $(\mathbf{A}_m^{k_n})$ is a tempered Følner sequence in $\mathbb{G}^{k_n}$. For each n, we choose a measureable action of $\mathbb{G}^{k_n}$ on $\mathbf{X}_n$. We use boldface to denote elements of $\mathbb{G}^{k_n}$, so that the action defines transformations $\boldsymbol{\phi} x_n = (\phi_1, \ldots, \phi_{k_n}) x_n$ for $x_n \in \mathbf{X}_n$, and this we refer to as the **full action** of $\mathbb{G}^{k_n}$. The full action defines the **joint action** of $\mathbb{G}$ on $\mathbf{X}_n$, given by

$$\phi x_n := (\phi, \ldots, \phi) x_n \quad \text{for } \phi \in \mathbb{G}, x_n \in \mathbf{X}_n .$$

The relevant notion of distributional invariance in the following is that every image $\boldsymbol{\phi} X_n$ of X_n under an element of $\mathbb{G}^{k_n}$ remains invariant under the joint action,

$$\psi \boldsymbol{\phi} X_n = (\psi\phi_1, \ldots, \phi_{k_n}) X_n \stackrel{\mathrm{d}}{=} \boldsymbol{\phi} X_n \tag{3.9}$$

for all $\psi \in \mathbb{G}$ and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_{k_n}) \in \mathbb{G}^{k_n}$. This property is considerably weaker than invariance under the full action, but stronger than invariance $\psi X_n \stackrel{\mathrm{d}}{=} X_n$ under the joint action. To define conditioning, it suffices to consider only the joint action, i.e. a Borel set A in $\mathbf{X}_n$ is almost invariant if $P(\psi A \triangle A) = 0$ for all $\psi \in \mathbb{G}$, and the σ-algebra of such sets is again denoted $\sigma(\mathbb{G})$. Where convenient, we center $g_n(X_n)$, and write

$$h_n(\bullet) := g_n(\bullet) - \mathbb{E}[g_n(\bullet)|\mathbb{G}] .$$

A lot of statistics of interest are of the form $h_n(X_n)$, U statistics are an example.

EXAMPLE 3.1. *For general invariance structures, U-statistics are statistics of the form*

$$\sum_{\phi_1,\ldots,\phi_{k_n} \in \mathbf{A}_n} F_n(f(\phi_1 X), \cdots, f(\phi_{k_n} X))$$

where f and (F_n) are functions $f : \mathbf{X} \to \mathbb{R}$ and $F_n : \mathbb{R}^{k_n} \to \mathbb{R}$. We embed it in this framework.
Let $\mathbf{X}_n := \left[\mathbb{R}^{k_n}\right]^{\mathbb{G}^{k_n}}$ be the space of arrays indexed by $\mathbb{G}^{k_n}$ with entries being k_n-dimensional vectors. Make $\mathbb{G}^{k_n}$ act on $\mathbf{X}_n$ by translating the indices of the array:

$$\tilde{\boldsymbol{\phi}} x \longrightarrow \left(x_{\tilde{\phi}_1\phi_1,\ldots,\tilde{\phi}_{k_n}\phi_{k_n}}\right)_{\phi_1,\ldots,\phi_{k_n} \in \mathbb{G}^{k_n}} \quad \forall x \in \mathbf{X}_n,\ \forall (\tilde{\phi}_1, \ldots, \tilde{\phi}_{k_n}) \in \mathbb{G}^{k_n}.$$

We define X_n to be $X_n := \Big(f(\phi_1 X), \ldots f(\phi_{k_n} X)\Big)_{\phi_1,\ldots,\phi_{k_n} \in \mathbb{G}^{k_n}}$ and notice that X_n is invariant under joint action of $\mathbb{G}$—but not necessarily under the action $\mathbb{G}^{k_n}$. Therefore the only thing left to define is the function $g_n : \mathbf{X}_n \to \mathbb{R}$ which we choose to be $g_n : x \longrightarrow F_n(x_{e,\ldots,e})$

for all $x \in \mathbf{X}_n$. We successfully embedded U-statistics in this framework and have

$$\sum_{\boldsymbol{\phi} \in \mathbf{A}_n^{k_n}} g_n(\boldsymbol{\phi} X_n) = \sum_{\phi_1, \ldots, \phi_{k_n} \in \mathbf{A}_n} F_n(f(\phi_1 X), \cdots, f(\phi_{k_n} X)).$$

To obtain second-order convergence result, we have to impose smoothness and mixing conditions on the sequence (g_n) that are expressed in terms of the full actions of the groups $\mathbb{G}^{k_n}$, and measure how sensitive g_n is to the effect of any single coordinate in $\mathbb{G}^{k_n}$. To measure smoothness, we consider $\boldsymbol{\phi} \in \mathbb{G}^{k_n}$ and manipulate a single entry using a transformation $(e, \ldots, e, \psi_i, e, \ldots, e)$ that is identity in all but its ith coordinate: The function g_n is p-**Lipschitz** in its ith argument, with Lipschitz coefficient $c_{ip}(g_n)$, if

$$\sup_{\psi_i \in \mathbb{G}, \boldsymbol{\phi} \in \mathbb{G}^{k_n}} \frac{1}{2} \left\| g_n \circ \boldsymbol{\phi} - g_n \circ (e, \ldots, e, \psi_i, e, \ldots, e) \boldsymbol{\phi} \right\|_p \; \leq \; c_{ip}(g_n) \; .$$

To formulate mixing, we need a suitable analogue of the sets $\mathcal{C}(t)$ in Section 3.2. To this end, let F and G be Borel sets in $\mathbb{G}^{k_n}$, and denote by $\mathrm{pr}_i F := \{\phi_i | \boldsymbol{\phi} \in F\}$ the projection of F onto the ith coordinate. Informally, we have to express that all ith coordinate entries of elements of F have distance at least t to all other coordinate entries in F, and all entries of vectors in G. The set of all these other entries is

$$[F, G]_i := \{\phi_j | \phi \in F, j \neq i\} \cup \{\phi_j | \phi \in G, j \leq k_n\} \; ,$$

and we define the dissimilarity measure

$$\delta_i(F, G) := \inf \{ d(\phi, \psi) \, | \, \phi \in \mathrm{pr}_i F, \psi \in [F, G]_i \} \; ,$$

where d is the metric on $\mathbb{G}$. Define the set system

$$\mathcal{C}_i(t) := \{(A, B) \in \sigma_{g_n}(\boldsymbol{\phi}, \boldsymbol{\phi}') \otimes \sigma_{g_n}(G) \, | \, \boldsymbol{\phi}, \boldsymbol{\phi}' \in \mathbb{G}^{k_n}, G \in \mathcal{B}(\mathbb{G}^{k_n}), \delta_i(\{\boldsymbol{\phi}, \boldsymbol{\phi}'\}, G) \geq t\} \; .$$

Compared to the previous definition of the conditional mixing coefficient $\alpha(\bullet | \mathbb{G})$ in Section 3.2, we substitute one of the conditional probabilities $P(\bullet | \mathbb{G})$ by a quantity that measures only the effect of a single coordinate of the full action: Lindenstrauss' theorem implies that, if X is $\mathbb{G}$-invariant, the conditional probability of a Borel set A in $\mathbf{X}$ can be written as

$$P(A | \mathbb{G}) \; = \; \mathbb{E}[\, \mathbb{I}\{X \in A\} | \mathbb{G}] \; = \; \lim_{m \to \infty} \frac{1}{|\mathbf{A}_m|} \int_{\mathbf{A}_m} \mathbb{I}\{\phi X \in A\} |d\phi| \; .$$

For a random element X_n of $\mathbf{X}_n$, we substitue this by

$$P_i(B) \ := \ \lim_{m\to\infty} \frac{1}{|\mathbf{A}_m|} \int_{\mathbf{A}_m} \mathbb{I}\{(e,\ldots,e,\phi_i,e,\ldots,e)X_n \in B\}|d\phi_i| \ ,$$

for $B \in \mathcal{B}(\mathbf{X}_n)$. Like $P(\bullet|\mathbb{G})$, the quantity $P_i(\bullet)$ is random. We then define the **marginal mixing coefficient**

$$\alpha_n(t|\mathbb{G}) := \sup_{i\le k_n} \sup_{(A,B)\in\mathcal{C}_i(t)} |P(A,B|\mathbb{G}) - \mathbb{E}[P_i(A)\mathbb{I}\{X_n \in B\}|\mathbb{G}]|$$

EXAMPLE 3.2. *If $k_n = 1$ and $X_n = X$ are constant, and if we denote $\alpha_{g_n}(\cdot|\mathbb{G})$ the conditional mixing coefficient for the process $\big(g_n(\phi X)\big)_{\phi\in\mathbb{G}}$ then we have $\alpha_n(t|\mathbb{G}) \le \alpha_{g_n}(t|\mathbb{G})$ for all $t \in \mathbb{R}^+$.*
In Example 3.1 we had $X_n = \Big(f(\phi_1 X), \ldots f(\phi_{k_n} X)\Big)_{\phi_1,\ldots,\phi_{k_n}\in\mathbb{G}^{k_n}}$ $\mathbb{G}^{k_n}$ acting by translating the indices and $g_n(\boldsymbol{\phi} X_n) = F_n(f(\phi_1 X), \ldots f(\phi_{k_n} X))$ for all $\boldsymbol{\phi} \in \mathbb{G}^{k_n}$. Then if $\alpha(\cdot|\mathbb{G})$ is the conditional mixing coefficient of $(f(\phi X))_{\phi\in\mathbb{G}}$ we can prove that

$$\alpha_n(\cdot|\mathbb{G}) \le \alpha(\cdot|\mathbb{G}).$$

Hypotheses (3.5) and (3.6) are then replaced by one of the following conditions, which use uniform integrability (UI). There either is a $k \in \mathbb{N}$ such that

$$\text{(i)}\ \sup_n \alpha_n(k|\mathbb{G}) = 0 \qquad \text{(ii)}\ \sup_n \sum_{i\le k_n} c_{i,2}(g_n) < \infty \qquad \text{(iii)}\ \big(h_n(\phi X_n)^2\big)_{\phi\in\mathbb{G}^{k_n}} \text{ is UI} \tag{3.10}$$

or an $\varepsilon > 0$ such that

$$\text{(i)}\ \sup_n \int_{\mathbb{G}} \alpha_n^{\frac{\epsilon}{2+\epsilon}}(d(e,g))d|g| < \infty \ \ \text{(ii)}\ \sup_n \sum_{i\le k_n} c_{i,2+\varepsilon}(g_n) < \infty \ \ \text{(iii)}\ \big(h_n(\phi X_n)^{2+\varepsilon}\big)_{\phi\in\mathbb{G}^{k_n}} \text{ is UI.} \tag{3.11}$$

3.4.2 Subsampling of Følner sets

For a probability measure P, denote by $\mathbb{E}_P$ the expectation taken with respect to P. More generally, if m is a σ-finite measure and A a Borel set with $0 < m(A) < \infty$, we write

$$\mathbb{E}_m[f(\Phi)|\Phi \in A] \ := \ \frac{1}{m(A)} \int_A f(\phi)m(d\phi) \ .$$

The empirical measure $\mathbb{F}_n$ can then be written as

$$\mathbb{F}_n(f,X) = \mathbb{E}_{|\bullet|}[f(\Phi X)|\Phi \in \mathbf{A}_n] \ .$$

To randomize these averages, we substitute Haar measure by a random measure μ, or more generally—in the context of triangular arrays—by a sequence (μ_n) of random measures satisfying

$$\mu_n \text{ is } \sigma\text{-finite on } \mathbb{G}^{k_n} \qquad \text{and} \qquad \mu_n(\mathbf{A}_n^{k_n}) > 0 \qquad \text{a.s. for all } n \in \mathbb{N}\,. \tag{3.12}$$

For a sequence (X_n) of $\mathbb{G}$-invariant random variables, we then replace $\mathbb{F}_n$ by the randomized empirical measure

$$\widehat{\mathbb{F}}_n(h_n, X_n) \;:=\; \mathbb{E}_{\mu_n}[h_n(\Phi X_n)|\Phi \in \mathbf{A}_n^{k_n}]\,. \tag{3.13}$$

The subsample represented by the random measure μ_n must cover $\mathbf{A}_n^{k_n}$ sufficiently well—for example, if the group in question is $(\mathbb{R},+)$, it must not concentrate on a hyperplane in $\mathbb{R}^{k_n}$. Formally, that can be ensured as follows: For all subsets $F, G \subset \mathbb{G}^{k_n}$ and all subsets $\mathcal{S}_1, \mathcal{S}_2 \subset \mathbb{N}$ denote

$$[F,G]_{\mathcal{S}_1}^{\mathcal{S}_2} := \Big([F,G]_i \cap pr_k(F \cup G)\Big)_{(i,k)\in\mathcal{S}_1\times\mathcal{S}_2}$$

the array with the (i,k)th-entry the sets of transformations ϕ that are in $pr_k(G)$ or $pr_k(F) \setminus pr_i(F)$.

Let $\mathcal{A}_k := \{A \in \mathcal{B}(\mathbb{G}^k)\,|\,|\mathrm{pr}_i A| \geq 1 \text{ for } i \leq k\}$. The sequence (μ_n) is **well-spread** if the family

$$\Big(\frac{|\mathbf{A}_n|^{\mathrm{card}(\mathcal{S}_1\times\mathcal{S}_2)}}{|A|}\mathbb{E}_{\mu_n\otimes\mu_n}\big(\mathbb{I}\big(\big[\{\mathbf{\Phi}\},\{\mathbf{\Phi'}\}\big]_{\mathcal{S}_1}^{\mathcal{S}_2} \in A]\big)\big|\mathbf{\Phi},\mathbf{\Phi'} \in A_n^{k_n}\big)\Big)_{\substack{\mathcal{S}_1,\mathcal{S}_2\subset\mathbb{N}\\ A\in\mathcal{A}_{\mathrm{card}(\mathcal{S}_1\times\mathcal{S}_2)}\\ n\in\mathbb{N}}}$$

is uniformly integrable. If only a single random variable X is considered (rather than (X_n)), we similarly call a random measure μ on $\mathbb{G}^k$ well-spread if the above condition holds for the sequences (k_n) and (μ_n) defined by $k_n := k$ and $\mu_n := \mu$.

Well-spreadness is a second-order property—note the product measure $\mu_n \otimes \mu_n$ in the definition—which suffices for asymptotic normality. Generalizing the Berry-Esseen bound requires a fourth-order condition. Define

$$\mathcal{S}^n := \sup_{\substack{\mathcal{S}_1,\mathcal{S}_2\subset\mathbb{N}\\ A\in\mathcal{A}_{\mathrm{card}(\mathcal{S}_1\times\mathcal{S}_2)}}} \mathbb{E}\Big[\frac{|\mathbf{A}_n|^{\mathrm{card}(\mathcal{S}_1\times\mathcal{S}_2)}}{|A|}\mathbb{E}_{\mu_n^4}\Big(\mathbb{I}\big(\big[\{\mathbf{\Phi}_1,\mathbf{\Phi}_2\},\{\mathbf{\Phi}_3,\mathbf{\Phi}_4\}\big]_{\mathcal{S}_1}^{\mathcal{S}_2} \in A\big)\big|\mathbf{\Phi}_{1:4} \in A_n^{k_n}\Big)\Big]\,,$$

and $\mathcal{S} := \sup_n \mathcal{S}^n$. We call (μ_n) **strongly well-spread** with spreading coefficient $\mathcal{S}$, if $\mathcal{S} < \infty$. If a sequence is strongly well-spread, it is well-spread.

Examples. (1) Let Π be a Poisson point process on $\mathbb{G}^k$. Then the random measure $\mu(\,\bullet\,) := |\Pi \cap \bullet\,|$ is strongly well-spread if

$$\sup_{A\subset\mathbb{G}^k,\,|A|<\infty} \frac{\mathbb{E}\big[|\Pi \cap A|\big]}{|A|} \;<\; \infty\,.$$

(2) Let $\mathbb{G}$ be discrete. For each n, let Π_n be a point process on $\mathbb{G}^{k_n}$ with

$$\Pi_n \cap \mathbf{A}_n^{k_n} \Big| \big(|\Pi_n \cap \mathbf{A}_n^{k_n}| = m\big) \quad \overset{\mathrm{d}}{=} \quad (\Phi_1, \dots, \Phi_n) \;,$$

where the variables Φ_i are uniformly drawn from $\mathbf{A}_n^{k_n}$ either with or without replacement. Then the random measures $\mu_n(\bullet) := |\Pi_n \cap \bullet|$ form a strongly well-spread sequence.

3.4.3 Generalized limit theorems

As previously, we associate variance contributions with each component $\widehat{\mathbb{F}}_n$. To this end, we consider a restricted average, in which one coordinate j in the group $\mathbb{G}^{k_n}$ is fixed to $\psi \in \mathbb{G}$: For a random element Φ of $\mathbb{G}^{k_n}$, write $\Phi^{j:\psi} = (\Phi_1, \dots, \Phi_{j-1}, \psi, \Phi_{j+1}, \dots, \Phi_{k_n})$, and let μ_{nj} be the marginal of μ_n with jth coordinate integrated out. Define

$$\widehat{\mathbb{F}}_{nj}(f, x, \psi) := \mathbb{E}_{\mu_{nj}}[f(\Phi^{j:\psi}x)|\Phi \in \mathbf{A}_n^{k_n}] \text{ and } \widehat{\mathbb{F}}_{\infty,j}(f, x, \psi) := \lim_{n\to\infty} \widehat{\mathbb{F}}_{nj}(f, x, \psi) \;.$$

For a random element Ψ of $\mathbb{G}$ with law

$$\mu_n^{i-j} := |\mathbf{A}_n| \, \mathbb{E}_{\mu_n \otimes \mu_n}\big[\mathbb{1}\{\Phi_j^{-1}\Phi_i' \in \bullet\}\big|\Phi, \Phi' \in \mathbf{A}_n^{k_n-1}\big] \;,$$

define the random variance contribution

$$\widehat{\eta}_{nm}^2 := \sum_{ij} \mathbb{E}_{\mu_n^{i-j}}\big[\mathbb{E}\big[\widehat{\mathbb{F}}_{\infty,i}(h_n, X_n, e)\widehat{\mathbb{F}}_{\infty,j}(h_n, X_n, \Psi)|\mathbb{G}\big] \,\big|\, d(e, \Psi) \leq m\big] \;.$$

Theorems 3.2 and 3.3 can then be restated more generally as follows:

THEOREM 3.5. *Let $(\mathbf{A}_n)$ be a tempered Følner sequence in $\mathbb{G}$, and (X_n) a sequence of $\mathbb{G}$-invariant random elements of $\mathbf{X}$. Define (k_n) and (h_n) as in Section 3.4.1. For each n, let μ_n be an almost surely σ-finite random measure on $\mathbb{G}^{k_n}$, and require: (1) The sequence (μ_n) is well-spread and independent of (X_n). (2) The constants k_n satisfy $k_n = o(|\mathbf{A}_n|^{\frac{1}{4}})$, and $\alpha(n|\mathbb{G}) \xrightarrow{n\to\infty} 0$ for all $k \leq \limsup k_n$. (3) Either condition* (3.10) *or* (3.11) *holds. If the limits*

$$\widehat{\eta}_{nm} \xrightarrow{p} \eta_m \quad \text{as } n \to \infty \qquad \text{and} \qquad \eta_m \xrightarrow{\mathbf{L}_2} \eta \quad \text{as } m \to \infty$$

exist, then $\sqrt{|\mathbf{A}_n|}\,\widehat{\mathbb{F}}_n(h_n, X_n) \xrightarrow{d} \eta Z$ holds as $n \to \infty$, for $Z \sim N(0,1)$.

THEOREM 3.6. *Assume the conditions of Theorem 3.5 hold, and require that (μ_n) is strongly well-spread, with spreading coefficients $\mathcal{S}_1, \mathcal{S}_2$. If condition* (3.10) *holds for some $k \in \mathbb{N}$, there is a constant κ_1 such that*

$$d_{\mathrm{W}}\Big(\frac{\sqrt{|\mathbf{A}_n|}}{\eta}\widehat{\mathbb{F}}_n(h_n, X_n), Z\Big) \;\leq\; \kappa_1 \frac{k_n^2}{\sqrt{|\mathbf{A}_n|}} + \Big\|\frac{\widehat{\eta}_{n,k_n}^2 - \eta^2}{\eta^2}\Big\| \;,$$

where κ_1 *is of order* $O\big((\mathcal{S}\wedge 1)\big[(\sum_i c_{i,4})^3\wedge 1\big]|\mathbf{B}_K|^2\big)$. *If* (3.11) *holds instead, set*

$$\mathcal{R}(b) := \sum_{t\geq b} |\mathbf{B}_{t+1}\setminus\mathbf{B}_t|\alpha_n(t|\mathbb{G})^{\frac{\varepsilon}{2+\varepsilon}} \qquad \textit{for } b\in\mathbb{N}$$

Then for any sequence $0<b_1<b_2<\ldots$ *of integers,*

$$d_{\mathrm{W}}\Big(\frac{\sqrt{|\mathbf{A}_n|}}{\eta}\widehat{\mathbb{F}}_n(h_n,X_n),Z\Big) \leq \kappa_2\mathcal{R}(b_n) + \kappa_3\frac{k_n^2|\mathbf{B}_{b_n}|}{\sqrt{|\mathbf{A}_n|}} + \Big\|\frac{\widehat{\eta}^2_{n,k_n}-\eta^2}{\eta^2}\Big\|\,,$$

where $\kappa_2 = O\big((\sum_i c_{i,2+\epsilon})^2(\mathcal{S}\wedge 1)\big)$ *and* $\kappa_3 = O\big(\big[(\sum_i c_{i,4+2\epsilon})^3\wedge 1\big][\mathcal{S}\wedge 1]\mathcal{R}(0)\big)$.

To recover Theorems 3.2 and 3.3, choose $k_n = 1$ and $\mu_n(\bullet) = |\bullet|$ for all $n\in\mathbb{N}$.

3.5 Concentration

3.5.1 Context

We next give a concentration result, which we formulate for a specific subcase of the setting presented in Section 3.4. We take $\mathbf{X}$ and $\boldsymbol{Y}$ to be standard Borel spaces and (X_n) to be a sequence of $\mathbb{G}$-invariant random elements of $\mathbf{X}$. Let (k_n) be a non decreasing sequence; and let $(f_n:\mathbf{X}\to\boldsymbol{Y})$ and $(g_n:\boldsymbol{Y}^{k_n}\to\mathbb{R})$ be measurable functions. We take (μ_n) to be a sequence of random measure with μ_n almost-surely σ-finite on $\mathbb{G}^{k_n}$ and define

$$h_n(\boldsymbol{\phi}X_n) := g_n(f_n(\boldsymbol{\phi}_1X_n),\ldots,\boldsymbol{\phi}_{k_n}X_n) - \mathbb{E}[g_n(f_n(\boldsymbol{\phi}_1X_n),\ldots,\boldsymbol{\phi}_{k_n}X_n)|\mathbb{G}], \quad \forall\boldsymbol{\phi}\in\mathbb{G}^{k_n}.$$

We want to establish concentration results for $\widehat{\mathbb{F}}_n(h_n,X_n) := \mathbb{E}_{\boldsymbol{\phi}\sim\mu_n}\big(h_n(\boldsymbol{\phi}X_n)\big|A_n^{k_n}\big)$.

3.5.2 Definitions and results

The result requires two types of conditions. One is widely used in the concentration literature: A function $f:\mathbf{X}^k\to\mathbb{R}$ is **self-bounded** if there are constants $\delta_1,\ldots,\delta_k$, the **self-bounding coefficients**, such that

$$\frac{1}{2}|f(\mathbf{x})-f(\mathbf{x}')| \;\leq\; \sum_{i\leq k}\delta_i\mathbf{I}\{x_i\neq x_i'\} \qquad \text{for all } \mathbf{x},\mathbf{x}'\in\mathbf{X}^k\,,$$

see e.g [9]. The function is **uniformly L_1-continuous** in $\mathbb{G}$ if

$$\sup_{\phi\in\mathbb{G}^k,\,\phi'\in B_\varepsilon(\phi)} \|h(\boldsymbol{\phi}x)-h(\boldsymbol{\phi}'x)\|_1 \;\xrightarrow{\varepsilon\to 0}\; 0\,,$$

where $\mathbf{B}^k_\varepsilon(\phi) := \{\phi'\in\mathbb{G}^k\,|\,d(\phi_i,\phi_i')\leq\varepsilon \text{ for } i\leq k\}$.

The second condition controls interactions between function values under different transformations of X. Suppose first $\mathbb{G}$ is discrete. For a stochastic process $Y = (Y_\phi)_{\phi\in\mathbb{G}}$ on the group, we measure how the full conditional at the identity element e,

$$\mathbf{p}_e(\bullet|\mathbf{y}) := P(Y_e \in \bullet|Y_\phi = y_\phi, \phi \neq e) ,$$

changes if we modify the remaining path at a single point. Define

$$\Lambda[Y] := \sum_{\phi\in\mathbb{G}\setminus\{e\}} \sup_{\substack{\mathbf{y},\mathbf{y}'\in\mathbf{X}^{\mathbb{G}} \\ y_\psi=y'_\psi \text{ if } \psi\neq\phi}} \|\mathbf{p}_e(\bullet|\mathbf{y}) - \mathbf{p}_e(\bullet|\mathbf{y}')\|_{\mathrm{TV}} .$$

For $\mathbb{G} = \mathbb{Z}$, this is the **Dobrushin interdependence coefficient** e.g [54], and we use the same terminology for a general discrete group $\mathbb{G}$.

For an uncountable group, we generalize the definition as follows: Let (ε_n) be a sequence of positive scalars with $\epsilon_n \to 0$, and for each n, let C_n be an "ε_n-grid" in $\mathbb{G}$, that is, a countable subset $C_n \subset \mathbb{G}$ satisfying

$$\text{(i) } e \in C_n \quad \text{(ii) } d(\phi, \phi') \geq \varepsilon_n \text{ for } \phi, \phi' \in C_n \text{ distinct} \quad \text{(iii) } \bigcup_{\phi\in C_n} B_{\varepsilon_n}(\phi) = \mathbb{G} .$$

Let (X_ϕ) be a process on $\mathbb{G}$. If the limit

$$\lim_{n\to\infty} \frac{1 - \Lambda[(X_\phi)_{\phi\in C_n}]}{|\mathbf{B}_{\varepsilon_n}|} = 1 - \rho[X]$$

exists, we call ρ the **Dobrushin curvature** of (X_ϕ) for (ε_n) and (C_n). For a discrete group, we recover $\rho[X] = \Lambda[X]$. A continuous example would be $\mathbb{G} = \mathbb{R}$ and a continuous-time Markov process $X = (X_t)_{t\in\mathbb{R}}$, in which case

$$\rho[X] = \lim_{t\to\infty} \frac{1}{t} \sup_{x,y\in\mathbb{R}} \|\mathcal{L}(X_0|X_t = x) - \mathcal{L}(X_0|X_t = y)\|_{\mathrm{TV}} .$$

Denote by $\mathcal{S}(n) := \{\mathbf{a} \in [0, 1]^n | \sum_i a_i = 1\}$ the n-dimensional simplex, and by $\mathfrak{P}(A)$ the set all of partitions of a set A into finitely many measurable subsets.

THEOREM 3.7. *Let $(\mathbf{A}_n)$ be a tempered Følner sequence in $\mathbb{G}$, and (X_n) a sequence of $\mathbb{G}$-invariant random elements of $\mathbf{X}$. Define (k_n) and (h_n) as in Section 3.5.1 and let (c_i) be the self bounding coefficients of h_n. For each n, let μ_n be an almost surely σ-finite random measure on $\mathbb{G}^{k_n}$, and define*

$$\tau_n := \sup_{\substack{\mathbf{a}\in\mathcal{S}(n) \\ \pi\in\mathfrak{P}(\mathbf{A}_n)}} \sum_j \big(\sum_i a_i \mu_n\big(\mathbf{A}_n^{j-1} \times \pi_j \times \mathbf{A}_n^{k_n-j} \mid \mathbf{A}_n^{k_n}\big)\big)^2 \frac{|\mathbf{A}_n|}{\sup_j |\pi_j|} .$$

To compute Dobrushin curvatures, fix a sequence (C_m) *of subsets in* $\mathbb{G}$ *as above, and define* $\rho_n := \rho[f_n(\phi X_n)_{\phi\in\mathbb{G}}]$. *If* $\tau_n < \infty$ *for all* n, *then*

$$P\bigl(\widehat{\mathbb{F}}_n(h_n, X_n) \geq t\bigr) \;\leq\; 2\mathbb{E}\Bigl(\exp\Bigl(-\frac{(1-\rho_n)|\mathbf{A}_n|}{(\sum_{i\leq k_n} c_i)^2\tau_n}\,t^2\Bigr)\Bigr) \qquad \text{for all } t > 0\;.$$

Although the result does not impose an explicit mixing assumption, its conditions imply mixing properties that are considerably stronger than those required by Theorems 3.2–3.6. Its statement is complicated by the random measures μ_n. If we fix μ_n to Haar measure on the group $\mathbb{G}^{k_n}$, it simplifies:

Corollary 3.8. *If* μ_n *in Theorem 3.7 is Haar measure on* $\mathbb{G}^{k_n}$, *then*

$$P\bigl(\widehat{\mathbb{F}}_n(h_n, X_n) \geq t\bigr) \;\leq\; 2\exp\Bigl(-\frac{(1-\rho_n)|\mathbf{A}_n|}{(\sum_{i\leq k_n} c_i)^2}t^2\Bigr) \qquad \text{for any } t > 0, n \in \mathbb{N}\;.$$

3.6 Applications

This section considers applications of our results to random fields, various exchangeable random structures, stochastic block models, "graphex" models, and marked point processes. Most of these results are novel, but we also obtain known results as special cases.

3.6.1 Random fields

A random field $(X_t)_{t\in\mathbb{G}}$ indexed by either $\mathbb{G} = \mathbb{Z}^d$ or $\mathbb{G} = \mathbb{R}^d$ is **stationary** if it is invariant under $\mathbb{G}$ acting on the index set by addition. In the discrete case $\mathbb{G} = \mathbb{Z}^d$, amenability specializes to the boundary condition (2.6), and Theorem 3.2 to Bolthausen's theorem [5]. For $\mathbb{G} = \mathbb{R}^d$, Theorem 3.2 implies a continuous counterpart to Bolthausen's result:

Corollary 3.9. *Let* $X = (X_t)_{t\in\mathbb{R}^d}$ *be a stationary random field, and* $f \in \mathbf{L}_1(X)$ *a real-valued function with mixing coefficient* α. *If* (3.6) *holds,*

$$\frac{\sqrt{s}}{s^d}\int_{[0,s]^d} f(X_t)|dt| \quad \xrightarrow{s\to\infty} \quad \eta Z \qquad a.s.$$

for $\eta^2 := \int_{\mathbb{R}^d} \mathbb{E}\bigl[f(X_0)f(X_t)\big|\mathbb{G}\bigr]|dt|$.

3.6.2 Exchangeability

A random element X of a standard Borel space is **exchangeable** if it is invariant under a measurable action of the finitary symmetric group $\mathbb{S}_\infty$. This group is discrete and amenable, and $\mathbf{A}_n := \mathbb{S}_n$ defines a tempered Følner sequence. We use the notation

$$\mathbb{F}^i(X,\phi) := \lim_{n\to\infty} \tfrac{1}{|\mathbb{S}_n^i|}\textstyle\sum_{\phi'\in\mathbb{S}_n^i} f(\phi'\phi X) \qquad \text{where } \mathbb{S}_n^i := \{\phi\in\mathbb{S}_n\,|\,\phi(i)=i\}$$

random structure X	ergodic structures	CLT (3.14) due to
exchangeable sequence	i.i.d. sequences	H. Bühlmann [12]
edge-exch. graphs [13, 19, 33]	i.i.d. seq. of edges	-
exchangeable partition [49]	paint-box dist.	-
homogeneous fragmentation [49]	i.i.d. seq. of paint-boxes	-
exchangeable graph [36]	graphon distributions	[3]
		[1]
jointly exch. array [36]	dissociated array	Eagleson/Weber [23]
separately exch. array [36]	dissociated array	-

Table 3.2: Examples of exchangeable structures

to specify the requisite asymptotic variance.

COROLLARY 3.10. *Let X be an exchangeable random element of a standard Borel space* $\mathbf{X}$. *Assume that* $\mathbb{E}[f(X)^2] < \infty$, *and require*

$$\textstyle\sum_{i\in\mathbb{N}} \limsup_j \|f(X) - f(\tau_{ij}X)\|_{\mathbf{L}_2} \quad < \quad \infty\,,$$

where τ_{ij} denotes the transposition of i and j. Then

$$\sqrt{n}\,\mathbb{F}_n(f, X) \;=\; \tfrac{\sqrt{n}}{n!}\textstyle\sum_{\phi\in\mathbb{S}_n} f(\phi X) \;\xrightarrow{d}\; \eta Z \qquad \text{as } n\to\infty \tag{3.14}$$

for a standard normal variable Z. The asymptotic variance η^2 is given by

$$\eta^2 = \textstyle\sum_{i,j\in\mathbb{N}} \mathrm{Cov}\big[\mathbb{F}^i(X,\tau_{ij}), \mathbb{F}^j(X,e)\big|\mathbb{S}_\infty\big] < \infty \quad a.s.$$

If in addition $\mathbb{E}(\frac{f(X)^4}{\eta^4}) < \infty$ *and,*

$$\textstyle\sum_{i\in\mathbb{N}} \limsup_j \Big\|\dfrac{f(X) - f(\tau_{ij}X)}{\eta}\Big\|_{\mathbf{L}_2} \quad < \quad \infty\,,$$

then the Wasserstein distance to the normal limit is

$$d_{\mathrm{W}}\big(\tfrac{\sqrt{n}}{\eta}\mathbb{F}_n(f,X), Z\big) = O\Big(\min_{k\in\mathbb{N}}\big[\tfrac{k^2}{\sqrt{n}} + \textstyle\sum_{i>k}\limsup_j \Big\|\dfrac{f(X) - f(\tau_{ij}X)}{\eta}\Big\|_{\mathbf{L}_2}\big]\Big)\,.$$

Table 3.2 lists examples of exchangeable structures to which the corollary applies. In some cases, (3.14) is known.

Separate and jointly exchangeable arrays are different in nature see Equation (2.1) and Equation (2.2). A separately exchangeable k-array is invariant under an action of $\Pi_{i=1}^k\mathbb{S}_\infty$— instead of $\mathbb{S}_\infty$. This is reflected by a faster rate of convergence of statistics on those arrays.

We use the notations

$$\mathbb{F}^{i_1,\ldots,i_k}(X,\boldsymbol{\phi}) := \lim_{n\to\infty} \frac{1}{|\mathbb{S}_n^{i_1,\ldots,i_k}|} \sum\nolimits_{\phi' \in \mathbb{S}_n^{i_1,\ldots,i_k}} f(\boldsymbol{\phi'}\boldsymbol{\phi}X),$$

where $\mathbb{S}_n^{i_1,\ldots,i_k} := \{\boldsymbol{\phi} \in \Pi_{i=1}^k \mathbb{S}_n \,|\, \phi(i_l) = i_l \;\forall l \leq k\}$; and denote the permutation:

$$\boldsymbol{\phi}_{j_1,\ldots,j_k}^{i_1,\ldots,i_k} := \big(\tau_{i_1,j_1},\ldots,\tau_{i_k,j_k}\big) \in \Pi_{i=1}^k \mathbb{S}_\infty.$$

We obtain:

Corollary 3.11. *Let X be an separately exchangeable k-array. Assume that $\mathbb{E}[f(X)^2] < \infty$. Then*

$$n^{\frac{k}{2}}\,\mathbb{F}_n(f,X) \;=\; \frac{\sqrt{n^k}}{n!} \sum\nolimits_{\phi\in\mathbb{S}_n} f(\phi X) \;\xrightarrow{d}\; \eta Z \qquad \textit{as } n\to\infty \tag{3.15}$$

for a standard normal variable Z. The asymptotic variance η^2 is given by

$$\eta^2 = \sum\nolimits_{\substack{i_1,\ldots,i_k\in\mathbb{N}^k \\ j_1,\ldots,j_k\in\mathbb{N}^k}} \operatorname{Cov}\big[\mathbb{F}^{i_1,\ldots,i_k}(X,\boldsymbol{\phi}_{j_1,\ldots,j_k}^{i_1,\ldots,i_k}), \mathbb{F}^{j_1,\ldots,j_k}(X,\boldsymbol{e})\big|\mathbb{S}_\infty\big] < \infty \quad \textit{a.s.}$$

Remark (Erdös-Rényi graphs). Statistics of exchangeable Erdös-Rényi graphs converge at a faster rate: The first n vertices are represented by an $n \times n$ submatrix of the adjacency matrix. Since the graphs are undirected, one only has to consider its upper triangle. For exchangeable graphs in general, the submatrix is invariant under $\mathbb{S}_n$ acting on rows and columns, which guarantees rate $\sqrt{n}$. In the Erdös-Rényi case, the upper triangle entries are i.i.d. Bernoulli variables, and hence distributionally invariant under any permutation of the entries. Since there are $n(n-1)/2$ entries, that is invariance under the larger group $\mathbb{S}_{n(n-1)/2}$, and the rate increases to n. ◁

Remark (Subsampling sample averages). We proved that the estimator $\frac{1}{n!}\sum_{\pi\in\mathbb{S}_n} f(\pi X)$ approximates $\mathbb{E}[f(X)|\mathbb{S}_\infty]$ and characterized its speed of convergence. However from a computational point of view, this estimator might be hard to work with. Indeed for a structure of size n, one has to average over $n!$ permutations. In the case of exchangeable sequences, benign cancellation reduces the number of terms to n, but the same is not true for exchangeable structures in general. One can hence ask whether S_n may be approximated by averaging over a smaller subset of permutations, possibly generated at random. In general, this is possible if the random subset "covers" the group $\mathbb{S}_n$ sufficiently well. A way to build such a subset is to choose a sequence (k_n) in $\mathbb{N}$ that grows at least quadratically, $O(k_n) = n^2$, and then to choose k_n elements drawn uniformly with replacement from $\mathbb{S}_n$ to form the random subset G_n. Using Theorem 3.6 we can prove that: $\frac{1}{k_n}\sum_{\pi\in G_n} f(\pi\cdot X)$ approximates $\mathbb{E}[f(X)|\mathbb{S}_\infty]$ at the same rate than $\frac{1}{n!}\sum_{\pi\in\mathbb{S}_n} f(\pi X)$. ◁

3.6.3 Stochastic block models with growing number of classes

A class of exchangeable graphs popular in statistics are stochastic block models, and a nonparametric (but no longer exchangeable) variant can be defined whose number of classes increases with sample size e.g [17]. Theorems 3.5 and 3.6 imply asymptotic normality results for such models.

Specify a probability distribution $\pi := (\pi_1, \ldots, \pi_r)$ on the finite set $[r]$, and a symmetric matrix $P := (P_{ij})_{i,j\leq r}$ with $P_{ij} \in [0,1]$. Generate a random undirected graph X with vertex set $\mathbb{N}$, by generating its adjacency matrix $(X_{ij})_{i,j\in\mathbb{N}}$ as follows: Draw a sequence $I_1, I_2, \ldots$ of indices in $[r]$ i.i.d. from π, and

$$X_{ij}|I_i, I_j \sim \text{Bernoulli}(P_{I_iI_j}) \qquad \text{for each } i < j \in \mathbb{N} \;.$$

Since clearly $(X_{ij}) \stackrel{\text{d}}{=} (X_{\phi(i)\phi(j)})$ for every permutation ϕ, the graph is exchangeable. The distribution on graphs so defined is called a **stochastic block model** with r communities. To apply the model to data, an observed graph with n vertices is explained as the submatrix $(X_{ij})_{i,j\leq n}$. A nonparametric extension can be defined by allowing r to grow with sample size: Define a monotonically increasing function $r : \mathbb{N} \to \mathbb{N}$, and a sequence of graphons w_n given by two sequences $\pi^n = (\pi^n_1, \ldots, \pi^n_{r(n)})$ and $P^n = (P^n_{ij})_{i,j\leq r(n)}$. For each n, let X^n be a random graph (with vertex set $\mathbb{N}$ generated by the stochastic block model with parameters π^n and P^n. An observed graph of size n is then explained as $(X^n_{ij})_{i,j\leq n}$. Since $(X^n)_n$ is a triangular array of invariant processes, Theorems 3.5 and 3.6 apply. To keep notation reasonable, we do not state a general results, but specifically choose f as the triangle density,

$$f(x) = \lim_{n\to\infty} \frac{\#\text{ triangles in } (x_{ij})_{i,j\leq n}}{\#\text{ triangles in complete graph of size } n}$$

Denote $g(x) := \mathbb{I}\{\,(x_{ij})_{i,j\leq 3} \text{ is triangle}\,\}$.

Corollary 3.12. *Let (X_n) be a sequence of random graphs, where X_n is generated by a stochastic block model w_n with $r(n)$ classes, and define*

$$E_i(n) := \textstyle\sum_{j\leq r(n)} \pi^n_j \big(P_{ij} \sum_{k\leq r(n)} \pi_k P_{ik} P_{jk}\big)$$

and $\eta(n)^2 := \sum_{i\leq r(n)} \pi^n_i E_i(n)(E_i(n) - \sum_{j\leq r(n)} \pi^n_j E_j(n))$. Then for $Z \sim N(0,1)$,

$$\tfrac{\sqrt{n}}{n(n-1)(n-2)\eta(n)}\big(\textstyle\sum_{\phi\in\mathbb{S}_n} g(\phi X_n) - \sum_{i\leq r(n)} \pi^n_i E_i(n)\big) \quad \xrightarrow{d} \quad Z \;,$$

and the Wasserstein distance between the left- and right-hand side scales as

$$O\Big(\tfrac{P((1,2,3)=\text{triangle},(1,2,4)\neq\text{triangle})^{\frac{3}{4}}}{\eta(n)^3} \frac{1}{\sqrt{n}}\Big) \;.$$

The terms $E_i(n)$ and $\eta(n)^2$, and the scaling by $n(n-1)(n-2)$, are specific to the triangle density, and change for other statistics.

Remark (Sparsified graphons). Another nonparametric extension of exchangeable graphs are sparsified graphon models e.g [3, 37], originally introduced in [4]. An ergodic exchangeable graph X can be generated by "graphon" function $w : [0,1]^2 \to [0,1]$ [7]. A sparsified graphon model generates sequence of exchangeable graphs X^n, each with vertex set $\mathbb{N}$, from the graphons $\rho(n)w$, for a monotonically decreasing function $\rho : \mathbb{N} \to (0,1]$. An observed graph of size n is then explained as $(X^n_{ij})_{i,j \leq n}$. This is again a triangular array, and Theorems 3.5 and 3.6 apply. ◁

3.6.4 Graphex models

We next consider an invariance which, although referred to as "exchangeability", is in fact an action of a group considerably more complicated than $\mathbb{S}_\infty$. It first arose in H. Bühlmann's generalization of de Finetti's theorem to continuous-time processes: If a càdlàg process has exchangeable increments, it is a mixture of Lévy processes [35]. Since increments are defined on intervals in $\mathbb{R}_+$, exchangeability is defined by swapping intervals: Fix $\delta > 0$ and two points $a, b \in \mathbb{R}_+$ with $b > a + \delta$, and let $\phi_{a,b,\delta}$ be the transformation of $\mathbb{R}_+$ that swaps $[a, a+\delta)$ and $[b, b+\delta)$, and is identity otherwise. Let $\mathbb{T}_1$ be the group generated by all such maps, where we exclude interval pairs that overlap. A càdlàg process $(X_t)_{t \in \mathbb{R}_+}$ then has **exchangeable increments** if the process, regarded as a sum of increments, is invariant under action of $\mathbb{T}_1$ on the index set $\mathbb{R}_+$. Lévy processes are closely related to Poisson processes and random measures, and Bühlmann's result was generalized to exchangeable random measures on the quadrant $\mathbb{R}^2_+$ by Kallenberg [36]; here, one similarly defines a group $\mathbb{T}_2$ generated by swapping rectangles. [14] invoke this representation to define a class of invariant random graphs, which has been extended to a generalization of graphon models in [8, 56]. In this generalization, graphons, which are functions on $[0,1]^2$, are replaced by functions on $\mathbb{R}^2_+$.

Let $\omega : \mathbb{R}^2_+ \to [0,1]$ be a symmetric measurable function, and let $\Pi = \big((U_1, V_1), (U_2, V_2), \ldots\big)$ be a unit-rate Poisson process on $\mathbb{R}^2_+$. Generate a random countable set $X_\omega \subset \mathbb{R}^2_+$, by drawing its indicator function conditionally independently as

$$\mathbb{I}_X\big((V_i, V_j)\big)\big|(U_i, U_j) \sim \text{Bernoulli}(\omega(U_i, U_j)) \qquad \text{for } i < j \;,$$

with $\mathbb{I}_X = 0$ otherwise. The random variable X_ω is $\mathbb{T}_2$-invariant, and Kallenberg's representation shows it is indeed ergodic. It defines an undirected random graph on the vertex set $\mathbb{N}$, in which the edge (i,j) is present if $(V_i, V_j) \in \Lambda$: Fix $s \in (0, \infty]$, and define a graph $g_s(X_\omega)$ as

$$(i,j) \in g_s(X_\omega) \quad :\Leftrightarrow \quad (V_i, V_j) \in X_\omega \cap [0,s)^2 \;.$$

The expected size of $g_s(X_\omega)$ increases with s, and finite (infinite) values of s yield finite (infinite) graphs a.s. The function ω (combined with additional information not relevant to our purposes) is called a *graphex* in [56].

If an instance $G \stackrel{\text{d}}{=} g_s(X_\omega)$ is observed, one can ask whether it is possible to obtain an estimate of ω. Assuming s is given, [57] propose to estimate the restricted function $\omega|_{[0,s]^2}$ as follows: Let N be the number of vertices in G. Subdivide $[0,s)^2$ into quadratic patches I_{ij}, and define a piece-wise constant function $\hat{\omega}$ on $[0,s)^2$ as

$$\hat{\omega}[G,s]\big|_{I_{ij}} := G_{ij} \qquad \text{where } I_{ij} := \big[\tfrac{i-1}{N}s, \tfrac{i}{N}s\big) \times \big[\tfrac{j-1}{N}s, \tfrac{j}{N}s\big) \ .$$

A main result of [57] is that this estimator is consistent on bounded domains $[0,t)^2$, in the sense that

$$g_t(X_{\hat{\omega}[g_s(X_\omega),s]}) \ \xrightarrow{\text{d}}\ g_t(X_\omega) \qquad \text{for every } t \in (0,\infty) \text{ as } s \to \infty \ .$$

The next result shows that it is possible to estimate a statistic of the unobserved random set $X_{\mathcal{W}}$ from the observed graph $g_t(X_{\mathcal{W}})$.

Proposition 3.13. *Let $\mathcal{W}$ be a random measurable symmetric function $\mathbb{R}_+^2 \to [0,1]$. Fix $t > 0$, and for a countable subset $\pi \in \mathbb{R}_+^2$, define the function $f_t(\pi) := \big|\pi \cap [0,t)^2\big|$. As $s \to \infty$,*

$$\sqrt{s}\Big(\mathbb{E}\big[f_t(X_{\hat{\omega}[g_s(X_{\mathcal{W}}),s]})\big|g_s(X_{\mathcal{W}})\big] - \mathbb{E}\big[f_t(X_{\mathcal{W}})\big|\mathbb{G}\big]\Big) \ \xrightarrow{d}\ \eta Z$$

for $Z \sim N(0,1)$, where $\eta^2 = 4t^4\,\mathrm{Cov}\big[|\Pi_{\mathcal{W}} \cap [0,1]^2|, |\Pi_{\mathcal{W}} \cap [0,1] \times [0,2]|\big|\mathbb{G}\big]$.

This is not a direct consequence of Theorem 3.2, since the groups $\mathbb{T}_1$ and $\mathbb{T}_2$ are much larger than $\mathbb{S}_\infty$, and not locally compact. Countability of the point process is used to sidestep the problem.

3.6.5 Marked point processes

As a final example, we consider the problem of estimating a function h of type of marked point process known as a random geometric measure e.g [48]. The processes have applications in physics and to k-nearest neighbor graphs, among others, and substantial recent work addresses asymptotic normality of estimates e.g. [30, 42, 48]. These typically require lengthy proofs. The high vantage point provided by Lindenstrauss' theorem allows us to simplify these considerably, and we derive a prototypical result as a direct corollary of Theorem 3.5. To this end, let $\mathbf{X}$ be a Polish space, and $K \subset \mathbf{X}$ a compact subset such that

$$\textstyle\bigcup_{\phi\in\mathbb{G}}\phi(K) = \mathbf{X} \qquad \text{and} \qquad \big|\{\phi \in \mathbb{G} \,|\, \phi(K) \cap K = \emptyset\}\big| \ < \ \infty \ . \tag{3.16}$$

Let Π be a locally finite, marked point process on $\mathbf{X}' := \mathbf{X} \times \mathcal{M}$, where $\mathcal{M}$ is a Borel space of marks [35]. $\mathbb{G}$ acts on the process by transforming points, but not marks: If $(X, M) \in \Pi$, then $(\phi(X), M) \in \phi(\Pi)$.

We define a conditional measure p on $\mathbf{X}'$ that conditions on a marked point (x, m), or, more generally, on a marked point and an additional countable set $\mathcal{Q} \subset \mathbf{X}'$ of such points. This is a measurable function $p : \mathbf{X}' \times \mathcal{F} \to \mathbf{M}$, where $\mathcal{F}$ is the set of countable subsets of $\mathbf{X}'$, and $\mathbf{M}$ the space of σ-finite measures on $\mathbf{X}'$. For a fixed set $\mathcal{Q} \in \mathcal{F}$,

$$\xi_{\mathcal{Q}}(\bullet) := \sum_{(x,m)\in\mathcal{Q}} p((x, m), \mathcal{Q}, \bullet)$$

then defines a σ-finite measure on $\mathbf{X}'$. If $\mathbf{A}_1, \mathbf{A}_2, \ldots \subset \mathbb{G}$ are sets of transformations, one can generate translates $\phi(K)$ of the fixed set K by elements $\phi \in \mathbf{A}_n$, and randomly generate $\mathcal{Q}$ by intersecting these with Π to obtain $\Pi_n := \Pi \cap \mathbf{A}_n K \times \mathcal{M}$. The *random* measure ξ_{Π_n} is called a **random geometric measure** e.g [48]. Of interest is the limiting behavior of

$$\mathbb{E}_{\xi_{\Pi_n}}[h(X, M)] = \int h(x, m)\xi_{\Pi_n}(dx, dm) \qquad \text{for } h \text{ bounded measurable.}$$

To state a result, let $\alpha(\bullet|\mathbb{G})$ be the conditional mixing coefficient for the function $f(\mathcal{Q}) := \mathbb{E}_{\xi_{\mathcal{Q}\cap K}}[h(X, M)]$.

Corollary 3.14. *Let Π be $\mathbb{G}$-invariant, $\phi(\Pi) \stackrel{\text{d}}{=} \Pi$ for all $\phi \in \mathbb{G}$, and $(\mathbf{A}_n)$ a Følner sequence. Require $\alpha(m|\mathbb{G}) \to \infty$ for $k \in \mathbb{N}$ and $m \to \infty$, and*

$$\int_{\mathbb{G}} \alpha_2(d(e, \phi)|\mathbb{G})^{\frac{\epsilon}{2+\epsilon}} |d\phi| < \infty \qquad \textit{and} \qquad \|f(\Pi)^{2+\epsilon}\| < \infty \ \textit{ is U.I.}$$

for some $\epsilon \geq 0$. Then, for an independent variable $Z \sim N(0, 1)$,

$$\frac{\mathbb{E}_{\mu_\xi(\Pi_n)}\big[h(X, M)] - \mathbb{E}[\mathbb{E}_{\mu_\xi(\Pi_n)}[h(X, M))|\mathbb{G}]\big]}{\sqrt{|\mathbf{A}_n|}} \xrightarrow{d} \eta Z \qquad \textit{as } n \to \infty\ ,$$

where $\eta^2 := \int_{\mathbb{G}} \mathrm{Cov}(f(\Pi_n), f(\phi\Pi_n)|\mathbb{G}) d|\phi|$. Moreover,

$$d_{\mathrm{W}}\Big(\tfrac{\mathbb{E}_{\mu_\xi(\Pi_n)}[h(X,M) - \mathbb{E}[\mathbb{E}_{\mu_\xi(\Pi_n)}[h(X,M)]|\mathbb{G}]}{\eta\sqrt{|\mathbf{A}_n|}}, Z\Big) = O\Big(\tfrac{\max\{1, \|f(\Pi)\|_4^3\}}{\sqrt{|\mathbf{A}_n|}}\Big)$$

A standard hypothesis in the relevant literature is a "stabilization condition" [48]. If it holds, and Π is a Poisson process, Corollary 3.14 holds. It implies, for example, that the intrinsic volumes of so-called *germ-grain models* [42] are asymptotically normal. It can also be used to recover the results of [30]: Choose $\mathbf{X}$ as the (Polish) space of compact subsets of $\mathbb{R}^d$, endowed with the Fell topology. Let $\mathbb{G} = \mathbb{R}^d$ act on $\mathbf{X}$ by shifts, $\phi(C) := \{x + \phi | x \in C\}$ for a compact set $C \in \mathbf{X}$ and $\phi \in \mathbb{R}^d$. One can then define a Poisson point process Π on $\mathbf{X}$, each point of which represents a compact set in $\mathbb{R}^d$ [**?**,]Chapter 3]aleph. Abbreviate $Z := \cup_{C\in\Pi} C$, and let $V_0, \ldots, V_d$ denote the intrinsic volumes in $\mathbb{R}^d$. Corollary 3.14 then states

$$n^{-\frac{d}{2}}\big(V_i(Z \cap [-n; n]^d) - \mathbb{E}[V_i(Z \cap [-1; 1]^d)]\big) \xrightarrow{d} \sigma_i Z \quad \text{as } n \to \infty\ ,$$

where $\sigma_i^2 = \mathrm{Var}[V_i(Z \cap [-1;1]^d)]$ is finite almost surely.

3.7 Approximation by subsets of transformations

Suppose X is invariant under a group $\mathbb{G}$, but part of this invariance is neglected, in the sense that sample averages are computed only with respect to a (non-compact) subgroup $\mathbb{H}$—for example, because one only has reason to assume invariance under $\mathbb{H}$, or because $\mathbb{G} \backslash \mathbb{H}$ contains computationally intractable elements. Since $\mathbb{G}$ is lcsc and amenable, so is $\mathbb{H}$, and Theorems 3.2 and 3.3 explain convergence $|\mathbf{A}_n^{\mathbb{H}}|^{-1} \int_{\mathbf{A}_n^{\mathbb{H}}} f(\phi X)|d\phi| \xrightarrow{n\to\infty} \mathbb{E}[f(X)|\mathbb{H}]$ almost surely, for a Følner sequence $\mathbf{A}_n^{\mathbb{H}}$ in $\mathbb{H}$. The next results shows that a stronger statement is possible: Provided the relevant mixing condition for f is satisfied for the entire group $\mathbb{G}$, the sequence $\mathbb{F}_n^{\mathbb{H}}$ even converges to $\mathbb{E}[f(X)|\mathbb{G}]$, although at a slower rate. If the group $\mathbb{H}$ is obtained from $\mathbb{G}$ by "factoring out" a compact subgroup—that is, if there is compact group $\mathbb{K} \subset \mathbb{G}$ such that $\mathbb{H}$ and $\mathbb{K}$ generate $\mathbb{G}$—the rate is reduced only by a constant.

Proposition 3.15. *Let $\mathbb{G}$ be generated by the union of a non-compact group $\mathbb{H}$ and a compact group $\mathbb{K}$, and let X be $\mathbb{G}$-invariant. Let $f \in \mathbf{L}_1(X)$ satisfy* (3.6). *If $(\mathbf{A}_n)$ is a Følner sequence in $\mathbb{G}$, then $(\mathbf{A}_n \cap \mathbb{H})$ is a Følner sequence in $\mathbb{H}$, and there exist random variables $\eta, \eta_{\mathbb{H}} \in \mathbf{L}_2(X)$ such that*

$$\frac{1}{\sqrt{|\mathbf{A}_n \cap \mathbb{H}|}} \int_{\mathbf{A}_n \cap \mathbb{H}} \big(f(\phi X) - \mathbb{E}[f(X)|\mathbb{G}]\big)|d\phi| \xrightarrow{d} \eta_{\mathbb{H}} Z$$

and

$$\frac{1}{\sqrt{|\mathbf{A}_n|}} \int_{\mathbf{A}_n} \big(f(\phi X) - \mathbb{E}[f(X)|\mathbb{G}]\big)|d\phi| \xrightarrow{d} \eta Z \;.$$

The ratio $\beta := \sqrt{|\mathbb{K}|}\frac{\eta_{\mathbb{H}}}{\eta}$ is given by

$$\beta^2 - 1 = \frac{1}{\eta^2} \int_{\mathbb{H}} \int_{\mathbb{K}} \mathbb{E}[f(X)(f(\phi X) - f(\psi\phi X))|\mathbb{G}]|d\psi||d\phi| \quad a.s.$$

For example, suppose $X = (X_t)_{t\in\mathbb{R}}$ is a random field invariant under the Euclidean group—the group of rotations and translations of Euclidean space—and one averages only with respect to translations. Then $\mathbb{H} = \mathbb{R}^d$, $\mathbb{K}$ is the orthogonal group $\mathbb{O}_d$, and convergence slows by a factor of

$$\beta^2 - 1 = \frac{1}{\eta^2}\mathbb{E}\big[f(X) \int_{\mathbb{R}^d} \int_{\mathbb{O}_d} (f(X + \phi) - f(\theta X + \phi))|d\theta||d\phi|\big] \;. \tag{3.17}$$

Both this idea, and the subsampling of Følner sets in Theorem 3.5, can be used to implement computationally tractable approximations. To illustrate the concept, consider again the random field $X = (X_t)_{t\in\mathbb{R}}$, assuming invariance under both the translation group $\mathbb{R}^d$ and the rotation group $\mathbb{O}_d$. If the objective is to estimate $\mathbb{E}[f(X)]$, one can avoid integration over the groups by discretization. We approximate the group $\mathbb{R}^d$ by a finite, growing grid $\{-n, \ldots, n\}$, and $\mathbb{O}_d$ by random subsample.

Corollary 3.16. *Let $X = (X_t)_{t \in \mathbb{R}^d}$ be a random field invariant under rotations and translations of $\mathbb{R}^d$, and fix $m \in \mathbb{N}$. For each $z \in \mathbb{Z}^d$, let $\Theta_1^z, \ldots, \Theta_m^z$ be independent, uniform random elements of the orthogonal group O_d. Then if X satisfies* (3.6) *we obtain that*

$$\frac{1}{m\sqrt{(2n)^d}} \sum_{\substack{z \in \{-n,\ldots,n\}^d \\ j \leq m}} \big(f(\Theta_j^z(X+z)) - \mathbb{E}[f(X)]\big) \xrightarrow{d} \eta_m Z$$

as $n \to \infty$, for an almost surely finite random variable η_m. Relative to the empirical measure of the full isometry group, convergence slows by a coefficient $\beta_m^2 - 1 = \frac{1}{(m\eta_m^2)} \mathbb{E}\big[f(X) \int_{\mathbb{O}_d} (f(X) - f(\theta X))|d\theta|\big]$. (3.17).

This is true only if the random rotations Θ_j^z are regenerated for each shift z. If one generates m random rotations only once, the rate slows.

3.8 Proof overview

The main results, Theorem 3.2–3.6, concern asymptotic normality, and are proven using adaptations of Stein's method e.g [50]: For the function class

$$\mathcal{F} := \big\{ t \in \mathcal{C}^2(\mathbb{R}) \,\big|\, \|t\|_\infty \leq 1, \|t'\|_\infty \leq \sqrt{2/\pi}, \|t''\|_\infty \leq 2 \big\} \tag{3.18}$$

and a real-valued random variable W, Stein's inequality guarantees

$$d_{\mathrm{W}}(W, Z) \;\leq\; \sup_{t \in \mathcal{F}} \big|\mathbb{E}[W t(W) - t'(W)]\big| \qquad \text{for } Z \sim N(0,1) \,. \tag{3.19}$$

We substitute W by an (suitably scaled) term of the form $\frac{\sqrt{|\mathbf{A}_n|}}{\eta(n)} \mathbb{F}_n(f, X)$, or $\frac{\sqrt{|\mathbf{A}_n|}}{\eta(n)} \widehat{\mathbb{F}}_n(h_n, X)$ for the generalized results in Section 3.4. Central limit theorems are then established by showing the right-hand side vanishes as $n \to \infty$, and Berry-Esseen bounds by bounding it as a function of n. Proofs for Theorems 3.2 and 3.3, which are considerably less complicated then in the general case, are given in Chapter 6.

3.8.1 Proofs of the general results

The generalized limit theorems in Section 3.4 consider a sequence (h_n) of functions, and W_n is now given by the randomized empirical measure $\widehat{\mathbb{F}}_n(h_n, X)$ defined in (3.13). The proofs still resemble Stein's method, but require a number of modifications:
(1) For triangular arrays, the dimension k_n of the group may grow with n. We hence define

a surrogate that depends only on the first i elements of ϕ as

$$\bar{g}^i_n(\phi x) := \lim_{m\to\infty} \tfrac{1}{|\mathbf{A}_m|^{k_n-i}} \int_{\mathbf{A}_m^{k_n-i}} h_n(\phi_1 x,\ldots,\phi_i x,\psi_{i+1}x,\ldots,\psi_{k_n}x)|d\psi_{i+1:k_n}|\;.$$

That makes h_n a telescopic sum $h_n(\phi x)=\sum_{i\leq k_n}\bar{h}^i_n(\phi x)$ where $\bar{h}^i_n(\phi x):=\bar{g}^i_n(\phi x)-\bar{g}^{i-1}_n(\phi x)$, and

$$\mathbb{E}[t(W)W-t'(W)]=\mathbb{E}\Big[\frac{t(W)\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_{i\in\mathbb{N}}\mathbb{E}_{\mu_n}[\bar{h}^i_n(\phi x_n)|A^{k_n}_n]-t'(W)\Big]$$

(2) In the i.i.d. case, Stein's method considers averages $w_{i,n}=\sum_{|j-i|\leq m}\frac{x_j}{\eta\sqrt{n}}$ over a "dependency neighborhood" around an index i [50]. Regarded as a special case of a group action, the neighborhood $\{j\,|\,|j-i|\leq m\}$ is a neighborhood in the group. For general amenable actions, one must construct a suitable neighborhood in $\mathbb{G}^{k_n}$, without relying on a total order. Given the sequence (b_n) in Theorem 3.6 and an element $\phi\in\mathbb{G}^{k_n}$, we define $I_{b_n}(\phi_i,\phi'):=\{j|d(\phi'_j,\phi_i)\leq b_n\}$ and

$$\tilde{h}^{\phi_i,b_n}_n(\phi' x):=\frac{1}{|\mathbf{A}_m|^{I_{b_n}(\phi_i,\phi')}}\int_{\mathbf{A}_m^{k_n}\cap\{\theta|\theta_j=\phi'_j\text{ for }j\notin I_{b_n}(\phi_i,\phi')\}} h_n(\theta x)|d\theta|\;.$$

The average $w_{i,n}$ is then substituted by an expression of the form

$$\tilde{w}_{i,\phi,b_n}=\mathbb{E}_{\mu_n}\Big[\frac{\sqrt{|\mathbf{A}_n|}}{\eta}\big(h_n(\phi' X_n)-\tilde{h}^{\phi_i,b_n}_n(\phi' X_n)\big)\Big|\phi,\phi'\in\mathbf{A}^{k_n}_n\Big]\;.$$

(3) Substitution of the empirical measure $\mathbb{F}_n$ by the randomized averages $\widehat{\mathbb{F}}_n$ introduces additional randomness. One must hence additionally control the probability of selecting transformations in the dependency neighborhood.

(4) To obtain a central limit theorem in cases where no fourth moment is available, standard techniques—such as truncating the function h_n outside a compact set to obtain a function in $\mathbf{L}_4(X)$—are not applicable for general group actions, and additional arguments are required in this case.

3.8.2 Comments on other proof techniques

The choice of Stein's method is not arbitrary: Except for certain benign groups, other standard techniques fail. These include Lindeberg's replacement trick and martingale methods for central limit theorems; the replacement trick and Fourier techniques for Berry-Esseen bounds; and the Efron-Stein inequality and other standard techniques for concentration proofs. There are a several obstacles:

(i) Topology of the group. An integral step of many martingale proofs, and of the Efron-Stein approach to concentration, is to group observations into blocks. Dependence between blocks is then controlled using an isoperimetric argument (i.e. the block boundaries are of

negligible size). Such arguments apply to some groups, such as $\mathbb{G} = \mathbb{Z}$, but fail even for $\mathbb{G} = \mathbb{Z}^2$, which occurs for example in the random field in Figure 2.1. Bolthausen's use of Stein's method, in [5], addresses an instance of this problem.
(ii) Lack of a total order. Replacement arguments, such Lindeberg's method or the Efron-Stein inequality, rely on the left-invariant total order of $\mathbb{Z}$ to replace random variables sequentially. That makes them inapplicable, for example, to groups with torsion.
(iii) Uncountable groups, since replacement arguments require discreteness.

Martingales warrant an additional remark: They yield simple and elegant proofs of asymptotic normality if X is an exchangeable sequence or array, and are invoked e.g. by [43] for convergence, and by [23] for asymptotic normality. The related results of [1,3] are proven differently, but could similarly be obtained using martingales. Martingales are applicable if $\mathbb{G}$ contains a sequence $\mathbb{G}_1 \subset \mathbb{G}_2 \subset \ldots$ of finite subgroups such that $\mathbb{G} = \cup_n \mathbb{G}_n$ (choose $\mathbb{G}_n = \mathbb{S}_n$ for $\mathbb{G} = \mathbb{S}_\infty$). If so, $\mathbf{A}_n := \mathbb{G}_n$ defines a Følner sequence. Then $(\mathbb{F}_n, \sigma(\mathbb{G}_n))_n$ is a reverse martingale, which implies (2.3), and Theorem 3.2 follows from the reverse martingale central limit theorem. However, the method has limitations even for $\mathbb{G} = \mathbb{S}_\infty$. For example: If (X_i) is an exchangeable sequence and h real-valued, $(h(X_i, X_j))_{ij}$ is an exchangeable array, but even with proper normalization, $\sum_{i<j} h(X_i, X_j)$ is not a reverse martingale unless h is symmetric.

Chapter 4

Limit theorems for stable functions of exchangeable structures

Throughout this chapter, $\mathbf{X}$ designates a standard Borel space and (X_n) a sequence of exchangeable random variables with values in $\mathbf{X}$. For any subset $G \subset \mathbb{N}$ we denote by $\mathbb{S}^{\setminus G}$ the subgroup of permutations that leave G invariant:

$$\mathbb{S}^{\setminus G} := \{\pi \in \mathbb{S}_\infty | \pi(i) = i, \ \forall i \in G\}$$

and $\mathbb{S}(G)$ designates the subgroup of permutations that have G as a support:

$$\mathbb{S}(G) := \{\pi \in \mathbb{S}_\infty | \pi(i) = i \ \forall i \notin G\}.$$

A function $f : \mathbf{X} \to \mathbb{R}$ has n **degrees of freedom** if

$$f(X) =_{\text{a.s.}} f(\pi(X)) \qquad \text{for all } \pi \in \mathbb{S}^{\setminus [n]} .$$

Choose a sequence of–potentially random–functions $(F_n : \mathbf{X} \to \mathbb{R})$ where F_n is of order n. The goal of this chapter is to answer to the following question:

Under what conditions is $\frac{1}{\sqrt{n}} F_n(X_n)$ asymptotically Gaussian?

4.1 Definitions

We define the relation $\underset{n}{\equiv}$, an equivalence relation on $\mathbf{X}$:

$$x \underset{n}{\equiv} x' \iff \exists \pi \in \mathbb{S}^{\setminus [n]} \text{ such that } \mathrm{x} = \pi \mathrm{x}'.$$

We write $[X]_n := \{x' | x' \underset{n}{\equiv} x\}$ for the equivalence class of $x \in \mathbf{X}$, and $E_n := X/ \underset{n}{\equiv}$ the quotient space. If $\mathbf{X} = \mathbb{R}^{\mathbb{N}}$ is the space of sequences then $x \underset{n}{\equiv} x' \iff (x_1, \ldots, x_n) =$

$(x'_1, \dots, x'_n)$. A function $f : \mathbf{X} \to \mathbb{R}$ has n degrees of freedom if $f(\cdot)$ is constant on $[X]_n$. Similarly for all subsets $A \subset \mathbb{N}$ we define the equivalence relation $\underset{A}{\equiv}$ as:

$$x \underset{A}{\equiv} x' \iff x = \pi x' \text{ for } \pi \in \mathbb{S}^{\setminus \mathrm{A}}.$$

We denote $[x]_A := \{x' | x \underset{A}{\equiv} x'\}$ the equivalence claxs of x.

A Borel set A is **invariant** under $\mathbb{S}(G)$ if $\pi(A) = A$ for all $\pi \in \mathbb{S}(G)$. The collection of invariant sets forms a σ-algebra, denoted $\sigma(\mathbb{S}(G))$, and we similarly define the σ-algebras $\sigma(\mathbb{S}^{\setminus G})$. When conditioning on any of these σ-algebras, we abbreviate

$$\mathbb{E}[\bullet | \mathbb{S}(G)] := \mathbb{E}[\bullet | \sigma(\mathbb{S}(G))] \quad \text{and} \quad \mathbb{E}[\bullet | \mathbb{S}^{\setminus G}] := \mathbb{E}[\bullet | \sigma(\mathbb{S}^{\setminus G})] .$$

Informally, an action of $\mathbb{S}_\infty$ defines how X is broken down into parts, which are then exchanged by a given permutation—in the examples above, these parts would respectively be entries of a sequence, row-column pairs of a 2-array, vertices of a graph, and elements of $\mathbb{N}$ in a partition. To measure the influence of the ith part on f, we define

$$\mathbb{A}_i(f, x) := \lim_{n \to \infty} \frac{1}{n} \sum_{k \leq n} f(\tau_{in} x) ,$$

where τ_{in} is the permutation that only swaps i and n. For an exchangeable sequence $X = (X_n)$ as in Example (i),

$$\mathbb{A}_i(f, X) \ =_{\text{a.s.}} \ \mathbb{E}[f(X) | X_k, k \neq i] \quad \text{for } f \in \mathbf{L}_1(X) .$$

If X is a general exchangeable random variable and f is of order $k \in \mathbb{N}$, the pointwise ergodic theorem [41] implies

$$\mathbb{A}_i(f, X) \ =_{\text{a.s.}} \ \mathbb{E}[f(X) | \mathbb{S}(i \cup \{k, k+1, \dots\})] \quad \text{for } f \in \mathbf{L}_1(X) .$$

$\mathbb{A}_i$ thus "averages out" the contribution of X_i on $f(X)$. For any function $f \in \mathbf{L}_2(X)$, we define the difference operators

$$\Delta_i(f, X) := f(X) - \mathbb{A}_i(f, X) \ \text{ and } \ \Delta_{ij}(f, X) := \Delta_i(f, X) - \mathbb{A}_j(\Delta_i(f, X)) .$$

For ease of presentation, we introduce the following sigma-fields that are useful in the rest of the exposition: for all $i, n \in \mathbb{N}$ we write

$$\mathbb{S}_n(i) := \mathbb{S}(i \bigcup \{n, n+1, \dots\})$$

and for all $j \in \mathbb{N}$ we write

$$\mathbb{S}_n(i,j) := \mathbb{S}(\{i,j\} \bigcup \{n, n+1, \dots\}).$$

4.2 Limit theorems

We consider a sequence of **random** functions F_n, each satisfying $F_n \in \mathbf{L}_2(X)$ almost surely. Recall a sequence (Y_n) of real-valued random variables is uniformly integrable (UI) if $\lim_{t\to\infty} \sup_n \mathbb{E}[|Y_n|; |Y_n| > t] = 0$. The first question we want to answer is: What conditions guarantee that $1/n\big[F_n(X_n) - \mathbb{E}[F_n(X_n)|\mathbb{S}_\infty]\big]$ converges to 0? We introduce the following hypothesis:

(B1) The collection $(\Delta_i(F_n, X_n))_{i,n\in\mathbb{N}}$ is UI.

THEOREM 4.1. *Let (X_n) be a sequence of exchangeable random elements of $\mathbf{X}$, and (F_n) a sequence of random elements of $\mathbf{L}_2(X)$, independent of (X_n), such that F_n has n degrees of freedom. If (B1) holds then the following also does:*

$$\frac{1}{n}\Big[F_n(X_n) - \mathbb{E}[F_n(X_n)|\mathbb{S}_\infty]\Big] \to 0.$$

In the case where $\mathbf{X} = \mathbb{R}^{\mathbb{N}}$ and $X = (X_i)$ is an i.i.d process, if the functions (f_n) are uniformly-Lipchitz then $\frac{f_n(X_1,\dots,X_n)}{n}$ is self averaging.
If we impose stronger conditions on the family $\big(\Delta_i(F_n, X_n)\big)$ then $F_n(X_n)$ concentrates:
(B1') For each integer $i \in \mathbb{N}$ there is c_i an $\mathbb{S}_\infty$-measurable random variable such that

$$|\Delta_i(F_n, X_n)| \overset{a.s}{\leq} c_i.$$

THEOREM 4.2. *Let (X_n) be a sequence of exchangeable random elements of $\mathbf{X}$, and (F_n) a sequence of random elements of $\mathbf{L}_2(X)$, independent of (X_n), such that F_n has n degrees of freedom. If (B1') holds then the following also does:*

$$P\Big(\frac{1}{n}\Big|F_n(X_n) - \mathbb{E}[F_n(X_n)|\mathbb{S}_\infty]\Big| \geq \epsilon) \leq 2\mathbb{E}\Big(e^{-\frac{n}{2\sum_i c_i^2}}\Big), \quad \forall \epsilon > 0.$$

Imposing conditions on the family $\big(\Delta_i(F_n, X_n)\big)$ is however not enough to obtain a central-limit theorem.

EXAMPLE (Counter-example). Let $(X_i)_{i\in\mathbb{N}}$ be an i.i.d sequence with $X_i \sim \text{unif}([-1,1])$ and let (h_n) be a sequence of functions:

$$h_n(X_{-n}, \dots, X_n) := \max\big(0, \sum_{-n\leq i\leq n} X_i\big).$$

The family $\big(\Delta_i(h_n, X_n)\big)$ is bounded: $\sup_i \Big|\Delta_i(h_n, X_n)\Big| \leq 2$. However $\frac{\max\big(0, \sum_{-n\leq i\leq n} X_i\big)}{\sqrt{n}}$ is not asymptotically Gaussian:

$$\frac{1}{\sqrt{n}} \max\big(0, \sum_{-n\leq i\leq n} X_i\big) \xrightarrow{d} \max(0, Z), \quad Z \sim N(0, 2).$$

To obtain a central-limit theorem we need to impose additional conditions:
(H1)The collection $(\Delta_i^2(F_n, X_n))_{i,n\in\mathbb{N}}$ is UI.

(H2) The sequence $(F_n(X_n))$ is stable on average,

$$\tfrac{1}{n^2} \textstyle\sum_{j\in\mathbb{N}} \big(\sum_{i\in\mathbb{N}} \|\Delta_{ij}(F_n, X_n)\|_2\big)^2 \;\to\; 0 \qquad \text{as } n \to \infty\ .$$

(H3) The asymptotic variance exists, that is,

$$\tfrac{1}{n}\mathrm{Var}[F_n(X_n)|\mathbb{S}_\infty] \;\xrightarrow{\mathrm{p}}\; \eta^2 \qquad \text{as } n \to \infty\ ,$$

for some $\sigma(\mathbb{S}_\infty)$-measurable function η.

(H2) guarantees that the empirical variance converges to $\mathrm{var}(\mathrm{F_n(X_n)}|\mathbb{S}_\infty)$, and (H3) guarantees the convergence of $\mathrm{var}(\mathrm{F_n(X_n)}|\mathbb{S}_\infty)$ to a random variable η^2.

Theorem 4.3. *Let (X_n) be a sequence of exchangeable random elements of $\mathbf{X}$, and (F_n) a sequence of random elements of $\mathbf{L}_2(X)$, independent of (X_n), such that F_n has n degrees of freedom. If (H1)–(H3) hold then,*

$$\sqrt{n}\big(F_n(X) - \mathbb{E}[F_n(X)|\mathbb{S}_\infty]\big) \;\xrightarrow{\mathrm{d}}\; \eta Z \qquad \text{as } n \to \infty\ ,$$

for a standard normal variable Z.

Remark. Shortly before the completion of this draft, it was pointed out to us that in [16] the authors had studied a similar problem for independent and identically distributed data. They proposed conditions very similar to ours under which $f_n(X_1, \ldots, X_n)/n$ would be asymptotically normal. Two differences with this work are worth pointing out. Firstly we propose to study this problem in the more general setting of arbitrary exchangeable objects. This allows us to move beyond the case of sequences and gives us a single theorem to study a variety of objects from exchangeable graphs to exchangeable partitions. Secondly it is interesting to point out that the proof techniques are different. In [16] the authors proposed as adaptation of the Stein method to prove his results, providing him with a Berry-Esseen bound. Our proof is based on the martingale central-limit theorem, which allow us to derive a simpler proof, and make fewer assumptions on the moments of $(\Delta_i(f_n, X_n))$. $\triangleleft$

4.3 Smooth functions of an exchangeable sequence

If the functions in Theorem 4.3 are sufficiently smooth, the difference operators Δ_i and Δ_{ij} can be substituted by derivatives, as the next result illustrates. For simplicity, we assume X is an exchangeable sequence in a compact set of $\mathbb{R}$ and the functions F_n are non-random, but neither assumption is essential. To state the result, we denote by ρ_i^z the substitution function

$$\rho_i^z(x) := (x_1, \ldots, x_{i-1}, z, x_{i+1}, \ldots)$$

for any real-valued sequence $x = (x_1, x_2, \ldots)$ and $z \in \mathbb{R}$.

Corollary 4.4. *Let X be an exchangeable sequence in a compact set $A \subset \mathbb{R}$, and (f_n) a sequence of functions with $f_n \in \mathcal{C}^2(\mathbb{R}^n)$. Consider the following properties:*

(a) The sequence $\big(\sup_{z\in A} \frac{\partial f_n}{\partial x_i}(\rho_i^z(X))^2\big)_{i,n\in\mathbb{N}}$ is uniformly integrable.

(b) $\frac{1}{n^2}\sum_{j\in\mathbb{N}}\big(\sum_{i\in\mathbb{N}} \sup_{y,z\in A}\big\|\frac{\partial^2 f_n}{\partial x_i \partial x_j}(\rho_j^y(\rho_i^z(X)))\big\|_2\big)^2 \to 0$ as $n \to \infty$.

Then (a) implies (H1), and (b) implies (H2). Thus, if (H3) also holds,

$$\sqrt{n}\big(f_n(X) - \mathbb{E}[f_n(X)|\mathbb{S}_\infty]\big) \xrightarrow{\text{d}} \eta Z \quad \textit{for} \quad Z \sim N(0,1)\ .$$

One can compare the conditions to those familiar from concentration results, where a typical way to control influence of arguments on the functions would be a uniform Lipschitz condition

$$|f_n(x) - f_n(x')| < c_n|x - x'| \quad \text{where} \quad \sup_n c_n < \infty\ .$$

This condition implies (a) above, but not (b). Indeed, it does not suffice to guarantee a central limit theorem, as we mentioned earlier.

4.4 Estimation from random sampled subgraphs

Let $\mathbf{G}$ be the set of undirected, simple graphs with vertex set $\mathbb{N}$, and let X be an exchangeable random graph. For $x \in \mathbf{G}$, denote by $x|_n$ the induced subgraph of x on the vertex set $\{1, \ldots, n\}$, and by $\mathbf{G}_n := \mathbf{G}|_n$ the set of graphs of size n. Fix $k \in \mathbb{N}$. By a **sampling scheme**, we mean a sequence $(S_n)_{n>k}$, such that

(a) each S_n is a random measurable function $\mathbf{G}_n \to \mathbf{G}_k$

(b) for each $x_n \in \mathbf{G}_n$ and $n \in \mathbb{N}$, $S_n(x_n)$ is a.s. a subgraph of x_n.

Examples. *(i)* Uniform vertex sampling: Given the input graph x_n, select k vertices uniformly without replacement, and report the induced subgraph S_n on these vertices.
(ii) Random walk sampling: Select a vertex of x_n uniformly at random, and start a simple random walk of length k at this vertex. Report the path $\mathcal{P}_n$ of the random walk and the

induced subgraph S_n on this path.

See [6, 47] for more on sampling schemes. When learning from a sequence of independent and identically distributed observations, most estimators can be interpreted as minimizing the expected loss of a randomly selected observation: $\min_\theta \frac{1}{n}\sum_{i\leq n}\mathcal{L}(X_i,\theta)$. When working with graph data, we define the empirical risk as the average loss over a randomly sampled subgraph [**?**]:$\mathbb{E}[f(S_n(X|_{[n]}))|X]$. However while it is trivial to show that the empirical loss for i.i.d data follows a central limit theorem, this is less straightforward for a random graph. This is due to the potential dependence between the graph and the sampling scheme: For example, in sampling scheme (ii), the probability that two vertices are sampled depends on the graph in a complex way. To be able to prove the desired central limit theorem we will need to impose some conditions on the sampling scheme. To do so we introduce some notations. We abbreviate

$$F_{nk}(x_k,X) := f(x_k)P[S_n(X|_{[n]}) = x_k|X] \quad \text{for all } x_k\in\mathbf{G}_k,$$

and for all subsets $A\subset\mathbb{N}$ we denote $\mathcal{V}_k(A) := \{(v_1,\ldots,v_k)|v_i\in A\ v_i\neq v_j,\ \forall i\neq j\}$. Moreover we

$$\Delta_i(F_{nk}(X|_{v_k},X)) := F_{nk}(X|_{v_k},X) - \mathbb{E}\big(F_{nk}(X|_{v_k},X)\big|\mathbb{S}_n(i)\big),$$

and

$$\Delta_{i,j}(F_{nk}(X|_{v_k},X)) := \Delta_i(F_{nk}(X|_{v_k},X)) - \mathbb{E}\big(\Delta_i(F_{nk}(X|_{v_k},X))\big|\mathbb{S}_n(j)\big).$$

To get asymptotic normality we impose the following conditions :

(a1) The family $\big(n^{2k}F^2_{nk}(X|_{v_k},X)\big)_{n,v_k\in\mathcal{V}_k([n])}$ is U.I.

(a2) The family $\big(n^{2k+2}\Delta_i(F_{nk}(X|_{v_k},X))^2\big)_{n,v_k\in\mathcal{V}_k([n]\backslash i)}$ is U.I.

(a3) $n^{2k}\sum_{i\leq n}\Big[\sum_{j\leq n}\sup_{v_k\in\mathcal{V}_k([n]\backslash\{i,j\})}\big\|\Delta_{ij}(F_{nk}(X|_{v_k},X))\big\|_{\mathbf{L}_2}\Big]^2\to 0.$

(a4) There is some $\sigma(\mathbb{S}_\infty)$-measurable random variable η such that

$$\sum_{i,j\leq k} n^{2k}\mathrm{Cov}\big[F_{nk}(X|_{[k]},X),F_{nk}(X|_{(k+1,\ldots,k+i-1,j,k+i+1,\ldots,2k)},X)|\mathbb{S}_\infty\big] \xrightarrow{\text{p}} \eta^2.$$

If we impose those conditions we have:

Corollary 4.5. *Let X be an exchangeable random graph, $k\in\mathbb{N}$ a constant, and $f:\mathbf{G}_k\to\mathbb{R}$ a function. Let $(S_n)_{n>k}$ be a sampling scheme, that satisfies (a1)–(a4). Then*

$$\frac{1}{\sqrt{n}}\big(\mathbb{E}[f(S_n(X|_{[n]}))|X] - \mathbb{E}[f(S_n(X_{[n]}))|\mathbb{S}_\infty]\big) \xrightarrow{\text{d}} \eta Z$$

for $Z \sim N(0,1)$ and $n \to \infty$.

PROPOSITION 4.6. *The sampling schemes defined in Examples (i) and (ii) both satisfy conditions (a1)–(a4) of Corollary 4.5.*

4.5 Cross validation

Throughout this section $\mathbf{X}$ and $\boldsymbol{Y}$ will designate two standard Borel spaces and X and Y will be two exchangeable elements of respectively $\mathbf{X}$ and $\boldsymbol{Y}$ chosen such that $Z := (X, Y)$ is jointly exchangeable:

$$\pi Z = (\pi X, \pi Y) \overset{d}{=} Z, \quad \forall \pi \in \mathbb{S}_\infty.$$

We consider prediction problems on $\mathbf{X} \times \boldsymbol{Y}$. A predictor of size k is a function $f_n : E_n(\mathbf{X} \times \boldsymbol{Y}) \times E_k(\mathbf{X}) \to E_k(\boldsymbol{Y})$ that takes as input a training sample $[Z]_n$ and observations $[X']_k$ and outputs a prediction $f_n([Z]_n, [X']_k)$ for the equivalence class $[Y']_k$.

EXAMPLE. If $\mathbf{X} = \boldsymbol{Y} = \mathbb{R}^{\mathbb{N}}$ is the space of sequences. Then a predictor of size 1 is a function that take as input a training sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ and a new observation X' and predicts the value of Y' by $f_n\Big((X_1, Y_1), \ldots, (X_n, Y_n),\ X'\Big)$.

EXAMPLE. If $\mathbf{X} = \boldsymbol{G}$ is the space of undirected simple graphs and $\boldsymbol{Y} = \mathbb{R}^{\mathbb{N}}$ is the space of sequences. Then a predictor of size k is a function that take as input a graph G_n of size n and a sequence of corresponding labels $Y_1, \ldots, Y_n$, and for a new subgraph G'_k of size k predicts the labels $Y'_1, \ldots, Y'_k$.

The loss is measured by a function $\mathcal{L} : E_k(\boldsymbol{Y}) \times E_k(\boldsymbol{Y}) \to \mathbb{R}_+$, which defines the risk on "new observations"

$$R_n(f_n, [Z]_n) := \mathbb{E}\big[\mathcal{L}\big(f_n([Z]_n, [X]_{n+1:n+k}), [Y]_{n+1:n+k}\big)\big|\mathbb{S}^{\setminus[n]}\big],$$

and the expected risk:

$$R(f_n) := \mathbb{E}\big[\mathcal{L}\big(f_n([Z]_n, [X]_{n+1:n+k}), [Y]_{n+1:n+k}\big)\big|\mathbb{S}_\infty\big].$$

When $\mathbf{X} = \boldsymbol{Y} = \mathbb{R}^{\mathbb{N}}$ to estimate this risk function we can either split the sample $(Z_1, \ldots, Z_n)$ into a training set and a testing set, or either proceed to K-fold cross-validation. We generalize this concept for general objects Z. In this goal, we define a sequence (K_n) and write $(B_1^n, \ldots, B_{K_n}^n)$ a partition of the set $\{1, \ldots, n\}$ into K_n sets of almost equal size:

$$\big|\mathrm{card}(\mathrm{B_j^n}) - \mathrm{card}(\mathrm{B_i^n})\big| \leq 1, \quad \forall \mathrm{i}, \mathrm{j} \leq \mathrm{K_n}.$$

This partition will correspond to the different folds. We define $m_i^n := n - \mathrm{card}(\mathrm{B_i^n})$ the number of integers in $n \setminus B_n^i$, and for any subset $A \subset \mathbb{N}$ we write

$$\mathcal{S}_k(A) := \{(i_1, \dots, i_k) | (i_1, \dots, i_k) \in A^k \text{ s.t } \mathrm{i_j} \neq \mathrm{i_l} \text{ if } \mathrm{j} \neq \mathrm{l}\}.$$

The empirical risk of the predictor $f_{m_i^n}\big([Z]_{[n]\setminus B_i^n}, \cdot\big)$ on B_i^n is

$$\hat{R}_{i,n} := \frac{1}{|\mathcal{S}_k(B_i^n)|} \sum_{(i_1,\dots,i_k)\in\mathcal{S}_k(B_i^n)} \mathcal{L}\Big(f_{m_i^n}\big([Z]_{[n]\setminus B_i^n}, [X]_{i_1,\dots,i_k}\big), [Y]_{i_1,\dots,i_k}\big)\Big).$$

Note that the predictor $f_{m_i^n}\big([Z]_{[n]\setminus B_i^n}, \cdot\big)$ does <u>not</u> depend on $[Z]_{B_i^n}$. For example if $\mathbf{X} = \boldsymbol{Y} = \mathbb{R}^{\mathbb{N}}$, it depends only on the training sample $\{Z_j | j \notin B_i^n\}$. The K_n-fold cross validated risk is defined as: $\hat{R}_{\mathrm{cross,n}} := \frac{1}{K_n}\sum_{i\leq K_n} \hat{R}_{i,n}$.
If $\mathbf{X} = \mathbb{R}^{\mathbb{N}}$ is the space of sequences then the same observation X_i is used $K_n - 1$ times to compute $K_n - 1$ different estimators. Therefore the dependence between the empirical risks $(\hat{R}_{i,n})$ is not trivial. It is sensible to ask the following questions: How fast does $\hat{R}_{\mathrm{cross,n}}$ converges to $R(f_n)$ as n goes to infinity? How does that compare with the speed at which $\hat{R}_{1,n}$ converges to $R(f_n)$?

We can answer those questions using Theorem 4.3 if we impose certain conditions on the functions (f_n). To do so we need to introduce some notations. We define

$$\Delta_{i,n}(\mathcal{L}, Z) := \mathcal{L}\big(f_n([Z]_n, [X]_{n+1:n+k}), [Y]_{n+1:n+k}\big) - \mathbb{E}\big[\mathcal{L}\big(f_n([Z]_n, [X]_{n+1:n+k}), [Y]_{n+1:n+k}\big)\big|S_n(i)\big],$$

this measures the effect on the loss of a "small change of the training set". Secondly we write

$$\Delta_{i,n}(R_n, Z) := \mathbb{E}\Big[\Delta_{i,n}(\mathcal{L}, Z)\Big|\mathbb{S}^{\setminus[n]}\Big],$$

this measures the effect on the risk function of a "small change of the training set". The conditions that we impose are:

(H0) $\Big(\mathcal{L}\big(f_m\big([Z]_{[m]}, [X]_{m+1,\dots,m+k}\big), [Y]_{m+1,\dots,m+k}\big)^2\Big)$ is uniformly integrable

(H1) $\sqrt{nK_n}\sup_i \Big\|\Delta_{i,n}(\mathcal{L}, Z)\Big\|_{\mathbf{L_2}} = o(1)$, and $(n\Delta_{i,n}(R_n, Z))$ is uniformly integrable.

(H2) $\sum_{i\leq n}\Big[\sum_{j\leq n}\Big\|\Delta_{i,n}(R_n, Z) - \mathbb{E}\big[\Delta_{i,n}(R_n, Z)\big|\mathbb{S}_n(j)\big]\Big\|_{\mathbf{L_2}}\Big]^2 = o(1)$.

(H3) $\frac{K_n}{n}\sum_{i\leq n}\Big[\sum_{j\leq n}\Big\|\Delta_{i,n}(\mathcal{L}, Z) - \mathbb{E}\big[\Delta_{i,n}(\mathcal{L}, Z)\big|\mathbb{S}_n(j)\big]\big)\Big\|_{\mathbf{L_2}}\Big]^2 = o(1)$.

(H4) There is $\mathbb{S}_\infty$ exchangeable random variable σ such that

$$\mathrm{var}\Big(\mathcal{L}\big(\mathrm{f_n}([\mathrm{Z}]_\mathrm{n}, [\mathrm{X}]_{\mathrm{n+1:n+k}}), [\mathrm{Y}]_{\mathrm{n+1:n+k}}\big)\Big|\mathbb{S}_\infty\Big) \xrightarrow{\mathrm{P}} \sigma^2$$

Corollary 4.7. *Let X and Y be two exchangeable elements of respectively $\mathbf{X}$ and $\boldsymbol{Y}$ chosen such that $Z := (X, Y)$ is jointly exchangeable. If hypothesis (H0)–(H4) hold and $K_n = o(n^{-\frac{1}{4}})$ then the following holds:*

$$\sqrt{\frac{K_n}{n}}\Big[\hat{R}_{1,n} - R(f_{m_1^n})\Big] \xrightarrow{d} \sigma N$$

where $N \sim N(0, 1)$ is independent from Z. Moreover the following also holds

$$\frac{1}{\sqrt{n}}\Big[\hat{R}_{\text{cross,n}} - \frac{1}{K_n}\sum_{i \le K_n} R(f_{m_i^n})\Big] \xrightarrow{d} \sigma N$$

where $N \sim N(0, 1)$ is independent from Z.

Therefore if (H1)–(H4) holds and the expected risk satisfies $R(f_n) - R(f_{n-1}) = o(\frac{1}{\sqrt{n}})$ a simple split between train and test sets will be $\frac{1}{\sqrt{K_n}}$ times slower than K_n fold cross validation.

One can wonder how important are conditions (H2) and (H3). In the following we give an example of an estimation framework that does not satisfy those and show that not only the cross validated risk is not asymptotically Gaussian but also ill-behaved.

Example (Counter-example). Let $\mathbf{X} = \boldsymbol{Y} = \mathbb{R}^{\mathbb{N}}$ be the space of sequences and let X be an i.i.d sequence of uniform random variables with $X_1 \sim \text{unif}[0, 1]$. Let Y be another i.i.d sequence defined as $Y_i := \mathbb{I}(X_i \le \frac{1}{2})$ and set $Z = (X, Y)$. Define the prediction functions (f_n) to be the nearest neighbor predictor

$$f_n\big(Z_1, \ldots, Z_n, X'\big) := Y_{c(Z,X')},$$

where $c(Z, X') = \mathrm{argmin}_{i \le n}|X_i - X'|$. We set the loss functions $(\mathcal{L}_n)$ to be

$$\mathcal{L}_n(f_n\big(Z_1, \ldots, Z_n, X'), Y') = n\mathbb{I}\ (Y' \ne f_n\big(Z_1, \ldots, Z_n, X')).$$

Set $K_n = 2$. Then we can prove that the hypothesis (H2) and (H3) are not respected and that the 2-fold cross-validated risk is not asymptotically normal. Moreover as n goes to infinity

$$P\Big(\hat{R}_{\text{cross,n}} > \hat{R}_{1,n}\Big) \to \frac{1}{2}.$$

4.6 Double bagging

Throughout this section $\mathbf{X} = \boldsymbol{Y} = \mathbb{R}^{\mathbb{N}}$ will be the space of sequences and X and Y will be sequences of independent and identically distributed observations. We denote $Z := (X, Y)$, and we consider prediction problems on $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}}$. In the previous section, we wanted to understand how the cross validated risk behaved asymptotically. In this section we will consider M different predictors that we combine by a procedure called double bagging [28]; and we study the asymptotic behavior of this algorithm. Our goal is to study the empirical

---

1. Take as input n observations $(Z_1, \dots, Z_n)$ and M different predictors $(f_n^k : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R})_{k \leq M}$.
2. Partition $[n]$ into M into different sets $(C_1, \dots, C_M)$ of approximately equal size:
$$\max_{i \neq j} |C_i - C_j| \leq 1.$$
3. Define the random sets $(B_k)_{k \leq M}$ as: $B_k := \{Z_j |\ j \in C_k\}$.
4. Our predictors will be $\big(f_{|B_k|}^k(B_k, \cdot)\big)_{k \leq M}$
5. Evaluate for each data point X_i the Mth different estimators,
$$\hat{V}_i := \big(f_{|B_1|}^1(B_1, X_i), \dots, f_{|B_M|}^M(B_M, X_i)\big)$$
6. Regress Y on $\hat{V} := (\hat{v}_1, \dots, \hat{v}_M)$: we choose weights $(\hat{\theta}_k(Z))_{k \leq M}$ to minimize the following,
$$(\theta_1, \dots, \theta_M) \to \sum_{i=1}^n \Big(Y_i - \sum_{m \leq M} \theta_m f_{|B_m|}^m(B_m, X_i)\Big)^2.$$
7. Our global predictor is then,
$$\hat{f}_n^{2,M}(Z, x) := \sum_{i \leq M} \hat{\theta}_m(Z) f_{|B_m|}^m(B_m, x).$$

---

risk of this estimator:
$$\frac{1}{n} \sum_{k \leq n} \big[Y_k - \hat{f}_n^{2,M}(Z, X_k)\big]^2.$$

To state our theorems we require the following notations. We write

$$\Delta_{i,k}^{m,n}(Z) := f_n^m((Z_1, \dots, Z_n), X_k) - \mathbb{E}\big[f_n^m((Z_1, \dots, Z_n), X_k) | \mathbb{S}_n(i)\big].$$

We take $F(n, Z_1, \dots, Z_n)$ to be the following matrix

$$F(n, Z_1, \dots, Z_n) := \big(f_{|B_m|}^m(B_m, X_i)\big)_{m \leq M, i \leq n}.$$

Moreover we denote the loss of observation z as

$$\hat{R}_{Z_1, \dots, Z_n}(z) := \Big(y - \sum_{m \leq M} \hat{\theta}_m(Z) f_{|B_m|}^m(B_m, x)\Big)^2.$$

With those notations in hand we consider the following hypotheses:

(H1) The following is finite,

$$\sup_{\substack{n\in\mathbb{N}\\ m\leq M}} \sup_{i,k\leq n} n\Big\| f_n^m((Z_1,\ldots,Z_n),X_k) - \mathbb{E}\big[f_n^m((Z_1,\ldots,Z_n),X_k)|\mathbb{S}_n(i)\big]\Big\|_{L_\infty} < \infty$$

(H2)The following converges to zero,

$$\sup_{\substack{k\in\mathbb{N}\\ m\leq M}} \frac{1}{n^2}\sum_{j\in\mathbb{N}}\Big[\sum_{i\in\mathbb{N}} n\Big\|\Delta_{i,k}^{m,n}(Z) - \mathbb{E}[\Delta_{i,k}^{m,n}(Z)|\mathbb{S}_n(j)]\Big\|_{L_2}\Big]^2 \to 0.$$

(H3) The estimators (f_n^m) are symmetric:

$$f_n^m((Z_1,\ldots,Z_n),\cdot) = f_n^m((Z_{\pi(1)},\ldots,Z_{\pi(n)}),\cdot), \quad \forall n\in\mathbb{N}, m\leq M, \pi\in\mathbb{S}_\infty.$$

(H4)There is a real $\epsilon>0$ such that

$$P\Big(\lambda_{(1)}\big(F(n,Z_1,\ldots,Z_n)^T F(n,Z_1,\ldots,Z_n)\big) < n\epsilon\Big) = o(\frac{1}{\sqrt{n}}),$$

where for a matrix A we denote $\lambda_{(1)}(A)$ the smallest eigenvalue.

(H5) There are $\sigma_1,\sigma_2\in\mathbb{R}$ such that the following holds,

$$n\mathrm{var}\Big(\mathbb{E}\big[\hat{\mathrm{R}}_{\mathrm{Z_1,\ldots,Z_n}}(\mathrm{Z}_1')\big|\hat{\mathrm{f}}^{2,\mathrm{M}}(\mathrm{Z}_1,\ldots,\mathrm{Z}_\mathrm{n},\cdot)\big]\Big) \to \sigma_1^2,$$

and

$$\mathbb{E}\Big(\mathrm{var}\Big(\hat{\mathrm{R}}_{\mathrm{Z_1,\ldots,Z_n}}(\mathrm{Z}_1')\Big|\hat{\mathrm{f}}^{2,\mathrm{M}}(\mathrm{Z}_1,\ldots,\mathrm{Z}_\mathrm{n},\cdot)\Big)\Big) \to \sigma_2^2,$$

where Z_1' is an independent copy of Z_1.

Theorem 4.8. *Let X and Y be independent and identically distributed sequences with Y taking value in a compact A. If (H1)–(H5) holds then the following holds,*

$$\frac{1}{\sqrt{n}}\Big[\sum_{k\leq n}\big[Y_k - f_n^{2,M}((Z_1,\ldots,Z_n),X_k)\big]^2 - \mathbb{E}\big(\big[Y_k - f_n^{2,M}((Z_1,\ldots,Z_n),X_k)\big]^2\big)\Big] \xrightarrow{d} N(0,\sigma^2)$$

where $\sigma^2 := 2M^2\sigma_1^2 + M\sigma_2^2$

Certain assumptions were made stronger than needed, this was done to simplify the statement of the proof. A similar theorem could have been obtained if the blocks B_i where created randomly by sampling indexes in $[n]$ with or without remplacement and for general exchangeable structures.

4.7 Remarks on the proof of Theorem 4.3

The key idea is to transform $f_n(X_n)$ into a sum of martingale differences.

We write

$$\frac{1}{\sqrt{n}}\big[f_n(X_n) - \mathbb{E}\big(f_n(X_n)\big|\mathbb{S}_\infty\big)\big] = \frac{1}{\sqrt{n}}\sum_{i\in\mathbb{N}} \mathbb{E}\big(f_n(X)\big|\mathbb{S}^{\setminus[|1,i|]}\big) - \mathbb{E}\big(f_n(X)\big|\mathbb{S}^{\setminus[|1,i-1|]}\big).$$

The idea is to exploit the martingale central-limit-theorem to prove the desired result.

THEOREM 4.9. *Let $(S_{i,n}, \mathbb{F}_i)_{i,n\in\mathbb{N}}$ be a triangular array of martingales with martingale differences $Y_{i,n}$. Suppose that,*

- *For all positive $\epsilon > 0$ we have $\sum_{i\leq n} \mathbb{E}\big(Y_{i,n}^2 \mathbb{I}(|Y_{i,n}| \geq \epsilon)|\mathbb{F}_{i-1}\big) \to 0$*
- *There is σ an $\mathbb{F}_\infty$-measurable such that $\sum_{i\leq n} Y_{i,n}^2 \xrightarrow{P} \sigma^2$*

Then we will have $S_{i,n} \xrightarrow{d} \sigma Z$, where $Z \sim N(0,1)$.

Therefore to prove Theorem 4.3 it is enough to prove the following:

- For all positive $\epsilon > 0$ we have $\frac{1}{n}\sum_{i\leq n} \mathbb{E}\big(X_{i,n}^2 \mathbb{I}(|X_{i,n}| \geq \epsilon)|\mathbb{F}_{i-1}\big) \to 0$. And for this we use hypothesis $(H1)$.
- We have the following convergence $\frac{1}{n}\sum_{i\leq n} X_{i,n}^2 - \mathbb{E}\Big(X_{i,n}^2\Big|\mathbb{F}_\infty\Big) \xrightarrow{P} 0$ and to prove this we use hypothesis $(H2)$.
- The following converges, $\frac{1}{n}\sum_{i\leq n} \mathbb{E}\Big(\sum_{i\leq n} X_{i,n}^2\Big|\mathbb{S}_\infty\Big) \xrightarrow{P} \sigma^2$ and for this we can use $(H3)$.

Chapter 5

Examples in information theory

In this chapter we are interested in examples coming from information theory. The Shannon entropy is a key quantity for discrete valued stochastic processes that quantifies the rate at which information is generated [18]. We will first present a generalized notion of entropy for general invariant objects and prove that the empirical entropy is asymptotically Gaussian. The Kolmogorov complexity is another important information theoretic quantity that has been proven–in the case of stationary ergodic processes–to share many similarities with the Shannon entropy. We will also prove a central limit theorem for this quantity.

5.1 Entropy estimation

For a discrete random variable X with law P, the entropy is defined as $H[X] := -\mathbb{E}[\log P(X)]$. There is no immediate generalization of H to random elements of uncountable spaces; rather, additional structure is required to obtain a useful definition. The canonical example, ubiquitous in information theory, are discrete-time stochastic processes [51]. Ergodic theory generalizes entropy beyond stationary processes using group actions; the work of [41] makes this definition possible for all discrete groups admitting a tempered Følner sequence that does not grow "too slowly". For a subclass of these groups, our methods yields a central limit theorem for such generalized entropies.

Consider first a discrete-time stochastic process $X = (X_n)_{n\in\mathbb{Z}}$ with values in a finite alphabet $[k]$. The sequence X takes values in the uncountable space $\mathbf{X} := [k]^{\mathbb{Z}}$. There is hence no notion of a mass function, but if X is stationary, one can still define entropy as a rate of entropies of joint distributions,

$$h[X] := \lim_{n} \tfrac{1}{n} H(X_1, \ldots, X_n) \; . \tag{5.1}$$

The Shannon-McMillan-Breiman theorem [51] shows that

$$-\tfrac{1}{n}\log P(X_1,\ldots,X_n) \xrightarrow{n\to\infty} h[X] \qquad \text{a.s.} \tag{5.2}$$

if X is stationary and ergodic. The term on the left is also known as the **empirical entropy**. That the limit exists almost surely is a consequence of stationarity; that it is constant, of ergodicity. In the ergodic stationary case, one can hence define $h[X]$ alternatively as a limit of empirical entropies.

In ergodic theory, this definition is generalized as follows [24,63]: let a discrete group $\mathbb{G}$ acts measurably on a standard Borel space $\mathbf{X}$. Choose a finite partition $\mathcal{S} = (\mathcal{S}_1,\ldots,\mathcal{S}_k)$ of $\mathbf{X}$ into Borel sets, and write $\mathcal{S}(x) = i$ if $x \in \mathcal{S}_i$. An invariant random element X of $\mathbf{X}$ defines a process

$$(S_\phi)_{\phi\in\mathbb{G}} \quad \text{where} \quad S_\phi = \mathcal{S}(\phi X)\,, \quad \text{and we abbreviate } S_A := (S_\phi)_{\phi\in A}\,.$$

Choose a sequence of finite subsets $\mathbf{A}_1, \mathbf{A}_2, \ldots$ of $\mathbb{G}$, and define entropy as

$$h_{\mathcal{S}}[X] := \lim_n h_n(\mathcal{S}, X) \quad \text{where} \quad h_n(\mathcal{S}, x) := -\tfrac{1}{|\mathbf{A}_n|}\log P\big((\mathcal{S}(\phi x))_{\phi\in\mathbf{A}_n}\big)\,.$$

For illustration, consider the stationary ergodic sequence X above, where $\mathbf{X} = [k]^{\mathbb{Z}}$ and $\mathbb{G} = \mathbb{Z}$, and choose $\mathbf{A}_n = \{-n,\ldots,n\}$. Suppose we do not have access to X itself, but only to the coarsened information $\mathcal{S}(X)$. To read off the entry X_n, note that, since each element $\phi \in \mathbb{G}$ can be read as a map $\phi(n) = n + \phi$, we have $(X_n)_{n\in\mathbb{Z}} = (X_{\phi(0)})_{\phi\in\mathbb{Z}}$. By choosing $\mathcal{S}$ as the set of equivalence classes $\mathcal{S}_i = \{x \in \mathbf{X} | x_0 = i\}$, we can hence extract X_n as $\mathcal{S}(\phi X)$, where ϕ is the shift by $-n$, and $h_{\mathcal{S}}[X]$ specializes to (5.2).

For the definition of $h_{\mathcal{S}}$ to be meaningful, the sequence (h_n) must converge and derandomize in the limit, which is not at all trivial, but indeed true under remarkably general conditions: If $\mathbb{G}$ is amenable, $(\mathbf{A}_n)$ a tempered Følner sequence, and $|\mathbf{A}_n|/\log(n) \to \infty$, then $h_{\mathcal{S}}[X]$ exists almost surely, and is constant if X is ergodic [41]. The function $h_n(\mathcal{S}, \bullet)$ is thus a strongly consistent estimator of the entropy. To show it is asymptotically normal, we impose some mixing conditions. Define the function $f : \mathbf{X} \to \mathbb{R}$ to be $f(X) = \mathcal{S}(X)$ and define the set system

$$\mathcal{C}_m(t) := \big\{(A,B) \in \sigma_f(F) \otimes \sigma_f(G) \big| F, G \subset \mathbb{G}, |F| \le m, F \in \mathbb{G}\backslash \mathbf{B}_t(G)\big\}\,.$$

We then define the function $\alpha(\cdot, m) : \mathbb{R}_{\ge 0} \to (0,\infty)$ to be

$$\alpha(t,m) := \sup_{(A,B)\in\mathcal{C}_m(t)} |P(A)P(B) - P(A\cap B)|.$$

Note that when we defined the mixing coefficient in (3.2) we fixed m to be equal to 2. To prove the asymptotically Gaussianity of the empirical entropy we require the following to hold:

$$\textstyle\sum_{i\in\mathbb{N}} |\mathbf{B}_i| \min_{m\leq i}\big(\rho_m + \alpha^{\frac{\varepsilon}{2+\varepsilon}}(i-2m, |\mathbf{B}_m|)\big) < \infty \; , \tag{5.3}$$

where (ρ_m) are coefficients defined as

$$\rho_m := \sup_{A\subset\mathbb{G}} \big\|\log P(S_e|S_A) - \log P(S_e|S_{A\cap\mathbf{B}_m})\big\|_2 \; .$$

THEOREM 5.1 (Central limit theorem for empirical entropy). *Let $\mathbb{G}$ be finitely generated, with tempered Følner sequence $(\mathbf{A}_n)$ and a left-invariant total order $\preceq$. Fix a finite partition $\mathcal{S}$ of $\mathbf{X}$ into Borel sets. Let X be a $\mathbb{G}$-ergodic random element for which S satisfies the moment condition $\sup_{A\subset\mathbb{G}} \|\log(P(S_e|S_A))\|_{2+\varepsilon} < \infty$ and the mixing condition* (5.3)*, both for some $\varepsilon > 0$. Then*

$$\frac{h_n(\mathcal{S}, X) - h_{\mathcal{S}}[X]}{\sqrt{\mathbf{A}_n}} \quad \xrightarrow{d} \quad \eta Z \qquad \textit{as } n \to \infty$$

with asymptotic variance

$$\eta^2 = \textstyle\sum_{\phi\in\mathbb{G}} \mathrm{Cov}\big[\log(P(S_e|(S_\psi)_{\psi\preceq e})), \log(P(S_\phi|(S_\psi)_{\psi\preceq\phi}))\big] < \infty \qquad a.s.$$

A few remarks: *(i)* A Berry-Esseen bound can be obtained similarly. *(ii)* Left-invariance of the order $\preceq$ is not required for asymptotic normality, but rather to obtain a simple expression for the asymptotic variance. *(iii)* The asymptotic variance does not depend on the choice of $\preceq$. *(iv)* The condition that $\mathbb{G}$ supports a total order implies that it is *torsion-free*, which means that $\phi^m \neq e$ for all $m \in \mathbb{N}$ and all $\phi \in \mathbb{G}\backslash\{e\}$. Examples of discrete torsion-free groups include the additive groups $(\mathbb{Z}^d, +)$ and discrete Heisenberg groups (both of which also satisfy the other conditions of Theorem 5.1). Symmetric groups and rotation groups are not torsion-free. *(v)* The theorem remains valid if ergodicity of X is weakened to symmetry, in which case all relevant quantities must be conditioned on $\sigma(\mathbb{G})$.

5.2 Kolmogorov complexity of stationary sequences.

The Kolmogorov complexity of a binary sequence is defined as the length of the shortest program fed to a universal Turing machine that would print the sequence and halt. More formally, let U denote a Universal Turing machine. Given a program p the sequence printed by U is denoted with $U(p)$.

Definition 5.2. Let $\mathcal{P}_X$ denote the set of all binary programs that can generate a finite length binary sequence X and halt. Then, the Kolmogorov complexity of X is denoted with

$K(X)$ and is defined as

$$K(X) \triangleq \inf_{p \in \mathcal{P}_X} \text{length}(p),$$

where length(p) denotes the length of the sequence. Furthermore, the Kolmogorov complexity of any finite-length finite-alphabet sequence is the Kolmogorov complexity of its binary representation.

Apart from its mathematical elegance, Kolmogorov complexity has exhibited promising *theoretical* results in other areas of research including inductive inference [52], denoising [22, 58], linear regression [31], density estimation [2], etc. See [60] for more applications.

The intuitive similarity between the Kolmogorov complexity and Shannon's entropy rate has motivated researchers to pursue and establish formal connections between these two fundamental quantities [11, 27, 38, 40, 45, 53, 61, 62, 64, 65]. A celebrated result in this line of research is the following theorem due to Levin:

THEOREM 5.3. *[65] Let (X_i) be a binary stationary and ergodic process, whose law is computable. Then*

$$\frac{K(X_1, X_2, \ldots, X_n)}{n} \overset{a.s.}{\to} H(X_1|X_0, \ldots, X_{-\infty}).$$

According to this theorem Shannon's entropy rate can be seen as an approximation of the Kolmogorov complexity of the process. The goal of this section is to obtain a more refined connection between the Kolmogorov complexity and Shannon entropy. In particular, we would like to achieve the following two goals:

1. Obtain the rate at which $\frac{K(X_1, X_2, \ldots, X_n)}{n}$ converges to $H(X_1|X_0, \ldots, X_{-\infty})$.
2. Obtain finite sample upper bounds for $\left|\frac{K(X_1, X_2, \ldots, X_n)}{n} - H(X_1|X_0, \ldots, X_{-n})\right|$.

Note that, unlike all the existing results, the theorems we prove in this section offer information on the connection of Kolmogorov complexity and Shannon entropy for finite length sequences.

THEOREM 5.4. *Let (X_i) be a stationary and ergodic process, and let $(\alpha(i))$ be the mixing coefficients as defined in Equation (3.3). We assume that $X_1 \in A$, where $A = \{a_1, .., a_l\}$ with $l < \infty$. Furthermore, we suppose that*

C_1. The Kolmogorov complexity of all a_js is finite, i.e., $\max_{i \in \{1,\ldots,l\}} K(a_i) < \infty$.

C_2. We assume that there are fixed numbers K, $\beta > 1$, and $C > 1$, such that

- $\alpha(n) \leq K n^{-\beta}$.
- $\sum_{i=1}^{\infty} \rho_i < \infty$
- $\left|H(X_0 \middle| X_{-n:-1}) - H(X_0 \middle| X_{-\infty:-1})\right| \leq K 2^{-Cn \log(l)}$.

If we define

$$\sigma^2 \triangleq \text{var}(\log(P(X_0|X_{-1:-\infty}))) + 2\sum_{k=1}^{\infty} \text{cov}(\log(P(X_0|X_{-1:-\infty})), \log(P(X_k|X_{k-1:-\infty}))),$$

then $\sigma^2 < \infty$, *and*

$$\sqrt{n}\left(\frac{K(X_{1:n})}{n} - H(X_0|X_{-1}, \ldots, X_{-\infty})\right) \xrightarrow{d} N(0, \sigma^2).$$

Note that Theorem 5.4 implies Theorem 5.3. However, this result provides the rate of convergence as well.

Both Theorem 5.3 and Theorem 5.4 are concerned with the asymptotic behavior of the Kolmogorov complexity, and do not provide any information on the finite sample behavior of this quantity. Our next goal is to derive probabilistic upper bounds on the discrepancy of the Kolmogorov complexity and Shannon entropy in finite sample sizes. Our next theorem shows that such bounds can be obtained with stronger mixing conditions than those in Theorem 5.4. Before we state our result we review a notion of stability for the likelihood of a process, which is required in our next theorem.

Definition 5.5. The Hamming distance between two sequences $x_{1:n} := (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $y_{1:n} := (y_1, \ldots, y_n) \in \mathbb{R}^n$ is defined as

$$d_n(x_{1:n}, y_{1:n}) \triangleq \sum_{i=1}^{n} \mathbb{I}(x_i \neq y_i).$$

The Hamming distance enables us to define the notion of M-stability.

Definition 5.6. Let (X_i) being a stationary m-Markov process with X_1 taking value in the finite set A. The M-stability coefficient of (X_i) is defined as

$$M \triangleq \sup_n \sup_{\substack{x, x' \in A^{\mathbb{N}} \\ s.t\ d_n(x_{1:n}, x'_{1:n}) \leq 1}} |\log(P(X_1 = x_1, \ldots, X_n = x_n) - \log(P(X_1 = x'_1, \ldots, X_n = x'_n))|.$$

We will say that (X_i) is M-stable if its M-stability coefficient is finite.

The following example clarifies the notion of M-stability coefficient.

Remark. Consider $\{X_i\}_{i=-\infty}^{\infty}$ a finite-state Markov chain of order m. If

$$\rho \triangleq \min_{x_1, \ldots, x_{-m} \in A} P(x_1|x_0, \ldots, X_{-m}) > 0,$$

then the M-stability coefficient satisfies

$$M \leq (m+1)\log\Big(\frac{1}{\rho}\Big).$$

◁

THEOREM 5.7. *Let $X = (X_i)$ denote a stationary Markov process of order m. We assume that $X_1 \in A$, where $A = \{a_1, .., a_l\}$ with $l < \infty$. Furthermore, we assume that*

1. *The Kolmogorov complexity of all a_js is finite, i.e., $\max_{i\in\{1,\dots,l\}} K(a_i) < \infty$.*
2. *The M-stability coefficient of the process, M, is finite.*
3. *The Dobrushin interdependence coefficient (defined in Section 3.5.2) satisfies: $\Lambda(X) < 1$.*

Then there are two sequences $(\gamma_n)_n$ and $(K_n)_n$ such that (i) $\gamma_n \underset{n\to\infty}{\sim} l^{m+1}\frac{\log^*(n)}{n}$, (ii) $K_n \underset{n\to\infty}{\sim} \log^*(n)$, and (iii) for any $t > \gamma_n$

$$\mathbb{P}\Big(\frac{K(X_{1:n})}{n} - H(X_1|X_0,\dots,X_{-m+1}) \geq t\Big) \leq e^{-\frac{2n(t-\gamma_n)^2(1-\Lambda(X))}{M^2}}, \tag{5.4}$$

And,

$$P\Big(\frac{K(X_{1:n})}{n} - H(X_1|X_0,\dots X_{-m+1}) < -t\Big) \leq e^{-\frac{2n(t-\gamma(n))^2(1-\Lambda(X))}{K_n^2}}. \tag{5.5}$$

Chapter 6

Proofs of Limit Theorems for invariant distributions

Recall that we have to establish two results in the basic case (Theorems 3.2 and 3.3), and two results in the general case (Theorems 3.5 and 3.6). Each result holds under two alternative hypothesis ((3.5) and (3.6), and (3.10) and (3.11), respectively). In principle, only the general results require proof—Theorems 3.2 and 3.3 then follows as special cases. Since the general case requires more sophisticated arguments, and complicates notation, this appendix is structured as follows:

- In this section (Chapter 6), we prove the basic theorems in the (more difficult) case of hypothesis (3.6) to clarify the argument. The general proofs follow the same outline.
- The concentration result, Propositions 3.13 and 3.15, and the entropy central limit theorem are proven in Section 6.2.
- The proofs of the general Theorems 3.5 and 3.6 follow in Section 6.1.

6.0.1 An auxiliary result

We first establish that the conditional (resp. marginal) mixing coefficients provide valid upper bounds on certain terms involving multiple group elements, provided these elements differ sufficiently. Those facts are used throughout the proofs, whenever terms are upper-bounded using α. We will present them into two different lemmas.

To make the exposition of the first lemma clearer we define some notations. We have X a $\mathbb{G}$-invariant random element and Y a real-valued random variable with $Y \perp\!\!\!\perp_{\mathbb{G}} X$. For all finite subset $G := \{\phi_1, \ldots, \phi_{|G|}\} \subset \mathbb{G}$ we write $GZ := (f(\phi_1 X), \ldots, f(\phi_{|G|} X))$ and choose a function $h_{|G|} : \mathbb{R}^{|G|} \times \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$. The following lemma holds.

Lemma 6.1. *Fix $l \in \mathbb{N}$. Select any element $\phi, \phi_1, \phi_2 \in \mathbb{G}$, and subset $G \subset \mathbb{G}$ such that $\bar{d}(G, \{\phi_1, \phi_2\}) \geq l$, $\bar{d}(G, \phi^{-1}\{\phi_1, \phi_2\}) \geq l$. Then the following holds:*

$$\|\mathbb{E}(h_{|G|}(GZ, \{\phi_1, \phi_2\}Z, Y)|\mathbb{G}, Y) - \mathbb{E}(h_{|G|}(GZ, \phi^{-1}\{\phi_1, \phi_2\}Z, Y)|\mathbb{G}, Y)\|_1 \leq 4C\ \alpha^{\frac{\varepsilon}{2+\varepsilon}}(l|\mathbb{G}),$$

where $C = \|h_{|G|}(GZ, \{\phi_1, \phi_2\}Z, Y) - h_{|G|}(GZ, \phi^{-1}\{\phi_1, \phi_2\}Z, Y)\|_{L_{1+\frac{\varepsilon}{2}}}$.

Proof. Abbreviate $G_2 := \{\phi_1, \phi_2\}$, $G_3 := \phi^{-1}G_2$, and

$$\Delta h(X, Y) := h_{|G|}(GZ, G_2Z, Y) - h_{|G|}(GZ, G_3Z, Y)\ .$$

We first consider the case $\|\Delta h\|_\infty < \infty$, and then the general case.
Case 1: $\|\Delta h\|_\infty < \infty$. Fix $\delta > 0$. Then there is $N_\delta \in \mathbb{N}$, sets $(A_i, B_i, C_i)_{i \leq N_\delta}$, and coefficients $c_1, \ldots, c_{N_\delta}$ with $|c_i| \leq \|\Delta h\|_\infty$ such that the approximation

$$\Delta h^*(X, Y) := \sum_{i=1}^{N_\delta} c_i \mathbb{I}\{GZ \in A_i, Y \in C_i\}\big(\mathbb{I}\{G_2Z \in B_i\} - \mathbb{I}\{G_3Z \in B_i\}\big)$$

satisfies $\|\Delta h(X, Y) - \Delta h^*(X, Y)\|_1 \leq \delta$. Moreover we have,

$$\begin{aligned}\|\mathbb{E}\Big(\Delta h^*(X, Y)\Big|\mathbb{G}, Y\Big)\|_1 &\leq \sum_{i=1}^{N_\delta} |c_i| \big\|\mathbb{E}\big[\mathbb{I}\{GZ \in A_i, Y \in C_i\}\big(\mathbb{I}\{G_2Z \in B_i\} - \mathbb{I}\{G_3Z \in B_i\}\big)\big|\mathbb{G}\big]\big\|_1 \\ &\leq 2\|\Delta h\|_\infty\ \alpha(l|\mathbb{G})\ ,\end{aligned}$$

where the second inequality follows from the definition of the conditional mixing coefficient and by the triangle inequality. Since δ may be arbitrarily small,

$$\big\|\mathbb{E}[h_{|G|}(GZ, G_2Z, Y) - h_{|G|}(GZ, G_3Z, Y]|\mathbb{G}, Y)\big\|_1 \leq 2\|\Delta h\|_\infty\ \alpha(l|\mathbb{G}).$$

Case 2: $\|\Delta h\|_\infty$ *not bounded*. With no loss of generality, we can suppose that $\|\Delta h\|_1 \leq 1$. For $r \in \mathbb{R}$, define $\Delta h_r := \Delta h\ \mathbb{I}\{\Delta h \leq r\}$ and $\overline{\Delta h_r} := \Delta h - \Delta h_r$. By Hölder's inequality,

$$\begin{aligned}\big\|\mathbb{E}[h_{|G|}(GZ, G_2Z, Y) - h_{|G|}(GZ, G_3Z, Y]|\mathbb{G}, Y)\big\|_1 &\leq \big\|\Delta h_r\big\|_1 + \big\|\overline{\Delta h_r}\big\|_1 \\ &\leq 2r\alpha(l|\mathbb{G}) + 2r^{-\frac{\varepsilon}{2}}\ .\end{aligned}$$

The result follows for $r = \alpha(l|\mathbb{G})^{\frac{-2}{2+\varepsilon}}$.

□

We know establish how the marginal mixing coefficients will be useful to bound some key quantities.

Once again to make expositions clearer we need to introduce some notations. Those will be very similar to the one introduced for the previous lemma. Take **X** to be a Borel space that admits an action from $\mathbb{G}^k$. Choose X a random element of $\mathbf{X}$ that is invariant under joint action of $\mathbb{G}$; and Y a real-valued random variable with $Y \perp\!\!\!\perp_{\mathbb{G}} X$. For all finite subset $G := \{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_{|G|}\} \subset \mathbb{G}^k$ we write $GZ := (g(\boldsymbol{\phi}_1 X), \ldots, g(\boldsymbol{\phi}_{|G|} X))$ and choose a function $h_{|G|} : \mathbb{R}^{|G|} \times \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$. The following lemma holds.

LEMMA 6.2. *Fix $l \in \mathbb{N}$. Select any element $\boldsymbol{\phi}, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2 \in \mathbb{G}^k$, and subsets $G \subset \mathbb{G}^k$ such that $\delta_k(\{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2\}, G_1) \geq l$ and $\delta_k(e^{k,\phi}\{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2\}, G) \geq l$.*

Then the following holds:

$$\|\mathbb{E}(h_{|G|}(GZ, \{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2\}Z, Y)|\mathbb{G}, Y) - \mathbb{E}(h_{|G|}(GZ, \{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2\}Z, Y)|\mathbb{G}, Y)\|_1 \leq 4C \ \alpha^{*\frac{\varepsilon}{2+\varepsilon}}(l|\mathbb{G}),$$

where $C = \|h_{|G|}(GZ, \{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2\}Z, Y) - h_{|G|}(GZ, \{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2\}Z, Y)\|_{L_{1+\frac{\varepsilon}{2}}}$ where $\alpha^*(\cdot|\mathbb{G})$ denotes the marginal mixing coefficient of g .

PROOF. The proof is very similar to the previous one. Abbreviate $G_2 := \{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2\}$ and $\tilde{G}_2 = e^{k,\phi} G_2$ and write

$$\Delta h(X, Y) := h_{|G|}(GZ, G_2 Z, Y) - h_{|G|}(GZ, \tilde{G}_2 Z, Y) \ .$$

We first consider the case $\|\Delta h\|_\infty < \infty$, and then the general case.

Case 1: $\|\Delta h\|_\infty < \infty$. Fix $\delta > 0$. Then there is $N_\delta \in \mathbb{N}$, sets $(A_i, B_i, C_i)_{i \leq N_\delta}$, and coefficients $c_1, \ldots, c_{N_\delta}$ with $|c_i| \leq \|\Delta h\|_\infty$ such that the approximation

$$\Delta h^*(X, Y) := \sum_{i=1}^{N_\delta} c_i \mathbb{I}\{GZ \in A_i, Y \in C_i\}\big(\mathbb{I}\{G_2 Z \in B_i\} - \mathbb{I}\{\tilde{G}_2 Z \in B_i\}\big)$$

satisfies $\|\Delta h(X, Y) - \Delta h^*(X, Y)\|_1 \leq \delta$. Moreover we have,

$$\begin{aligned}
&\|\mathbb{E}\Big(\Delta h^*(X, Y)\Big|\mathbb{G}, Y\Big)\|_1 \\
&\leq \sum_{i=1}^{N_\delta} |c_i| \big\|\mathbb{E}\big[\mathbb{I}\{GZ \in A_i, Y \in C_i\}\big(\mathbb{I}\{G_2 Z \in B_i\} - \mathbb{I}\{\tilde{G}_2 Z \in B_i\}\big)\big|\mathbb{G}\big]\big\|_1 \\
&\leq 2\|\Delta h\|_\infty \ \alpha^*(l|\mathbb{G}) \ ,
\end{aligned}$$

where the second inequality follows from the definition of the conditional mixing coefficient by the triangle inequality. Since δ may be arbitrarily small,

$$\big\|\mathbb{E}[h(GZ, G_2 Z, Y) - h(GZ, \tilde{G}_2 Z, Y]|\mathbb{G}, Y)\big\|_1 \leq 2\|\Delta h\|_\infty \ \alpha^*(l|\mathbb{G}).$$

Case 2: $\|\Delta h\|_\infty$ *not bounded*. With no loss of generality, we can suppose that $\|\Delta h\|_1 \leq 1$.

For $r \in \mathbb{R}$, define $\Delta h_r := \Delta h \, \mathbb{I}\{\Delta h \leq r\}$ and $\overline{\Delta h_r} := \Delta h - \Delta h_r$. By Hölder's inequality,

$$\big\|\mathbb{E}[h(GZ, G_2 Z, Y) - h(GZ, \tilde{G}_2 Z, Y]|\mathbb{G}, Y)\big\|_1 \leq \big\|\Delta h_r\big\|_1 + \big\|\overline{\Delta h_r}\big\|_1$$
$$\leq 2r\alpha^*(l|\mathbb{G}) + 2r^{-\frac{\varepsilon}{2}} \; .$$

The result follows for $r = \alpha^{*\frac{-2}{2+\varepsilon}}$.

□

6.0.2 Proof of Theorems 3.2 and 3.3 under hypothesis (3.6)

Step 1: Bounding the right-hand side of (3.19). Let (b_n) be a non-decreasing sequence in $\mathbb{N}$. The values b_n will serve as radii of metric balls $\mathbf{B}_{b_n}$ in $\mathbb{G}$. In Theorem 3.3, (b_n) is given by hypothesis; for Theorem 3.2, we will have to choose a suitable sequence further on. Recall that f is centered, in the sense that $\mathbb{E}[f(X)|\mathbb{G}] = 0$. For each $n \in \mathbb{N}$, we approximate the asymptotic variance η by a random variable $\eta(n)$, which will be constructed explicitly at the end of the proof. For each n, we then apply (3.19) to

$$W \;:=\; \tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{F}_n(f, X) \;=\; \tfrac{1}{\eta(n)\sqrt{|\mathbf{A}_n|}} \int_{\mathbf{A}_n} f(\phi X)|d\phi| \; ,$$

where we abbreviate the (random) normalizer of the integral as

$$Z_n := \eta(n)\sqrt{|\mathbf{A}_n|} \; .$$

Now consider the right-hand side of (3.19). For a given $\phi \in \mathbb{G}$ and a constant $b > 0$, define the truncated integral

$$W_b^{\phi} := \frac{1}{Z_n} \int_{\mathbf{A}_n} \mathbb{I}\{d(\phi, \phi') \geq b\} f(\phi' X)|d\phi'| \; .$$

An application of the triangle inequality then yields

$$|\mathbb{E}[Wt(W) - t'(W)]| \;\leq\; \mathbb{E}\big[\int_{\mathbf{A}_n} \frac{f(\phi X)}{Z_n}(t(W) - t(W_{b_n}^{\phi})) - t'(W)|d\phi|\big]$$
$$+ \, \mathbb{E}\big[\tfrac{1}{Z_n} \int_{\mathbf{A}_n} f(\phi X) t(W_{b_n}^{\phi})|d\phi|\big] \; . \qquad (6.1)$$

If we denote the truncation error of W_b^{ϕ} by

$$\Delta_b^{\phi} := W - W_b^{\phi} \; ,$$

we can bound the first term of (6.1) further as

$$\mathbb{E}\Big[\int_{\mathbf{A}_n} \tfrac{f(\phi X)(t(W)-t(W^\phi_{b_n}))}{Z_n} - t'(W)|d\phi|\Big] \leq \Big|\mathbb{E}\Big[\int_{\mathbf{A}_n} \tfrac{f(\phi X)(t(W)-t(W^\phi_{b_n}))-\Delta^\phi_{b_n} t'(W)}{Z_n}|d\phi|\Big]\Big|$$
$$+ \Big|\mathbb{E}\big[t'(W)\big(1-\int_{\mathbf{A}_n} \tfrac{f(\phi X)}{Z_n}\Delta^\phi_{b_n}|d\phi|\big)\big]\Big|$$
$$\leq \Big|\mathbb{E}\Big[\int_{\mathbf{A}_n} \tfrac{f(\phi X)(t(W)-t(W^\phi_{b_n}))-\Delta^\phi_{b_n} t'(W)}{Z_n}|d\phi|\Big]\Big| + \sqrt{\tfrac{2}{\pi}}\big\|1-\tfrac{1}{Z_n}\mathbb{E}\big[\int_{\mathbf{A}_n} f(\phi X)\Delta^\phi_{b_n}|d\phi|\big|\mathbb{G}\big]\big\|$$
$$+ \sqrt{\tfrac{2}{\pi}}\big\|\tfrac{1}{Z_n}\int_{\mathbf{A}_n} f(\phi X)\Delta^\phi_{b_n} - \mathbb{E}[f(\phi X)\Delta^\phi_{b_n}|\mathbb{G}]|d\phi|\big\| \;.$$

For our purposes, Stein's inequality (3.19) hence takes the specific form

$$\begin{aligned} d_{\mathrm{W}}(W,Z^*) \;&\leq\; \sup_{t\in\mathcal{F}} \mathbb{E}\big[\tfrac{1}{Z_n}\int_{A_n} f(\phi X)t(W^\phi_{b_n})d|\phi|\big] \\ &+\; \sup_{t\in\mathcal{F}} \Big|\mathbb{E}\big[\tfrac{1}{Z_n}\int_{A_n} f(\phi X)(t(W)-t(W^\phi_{b_n})-\Delta^\phi_{b_n}t'(W))d|\phi|\big]\Big| \\ &+\; \sqrt{\tfrac{2}{\pi}}\Big\|1-\tfrac{1}{Z_n}\mathbb{E}\big[\int_{A_n} f(\phi X)\Delta^\phi_{b_n}d|\phi|\big|\mathbb{G}\big]\Big\| \\ &+\; \sqrt{\tfrac{2}{\pi}}\Big\|\tfrac{1}{Z_n}\int_{A_n}\big[f(\phi X)\Delta^\phi_{b_n} - \mathbb{E}[f(\phi X)\Delta^\phi_{b_n}|\mathbb{G}]\big]|d\phi|\Big\|_1 \\ &=: \text{(a)} + \text{(b)} + \text{(c)} + \text{(d)} \;. \end{aligned} \tag{6.2}$$

The leg work of the proof is to control these four terms.

Step 2: Bounding the terms (a)–(d). To bound (a), we observe

$$\Big|\mathbb{E}\big[\tfrac{1}{Z_n}f(\phi X)t(W_{\phi,b_n})\big]\Big| \leq \sum_{j\geq\lfloor|\mathbf{B}_{b_n}|/\delta\rfloor}\Big|\mathbb{E}\big[\tfrac{f(\phi X)(t(W_{\phi,j\delta})-t(W_{\phi,(j+1)\delta}))}{Z_n}\big]\Big|$$
$$\leq 4\sqrt{\tfrac{2}{\pi|\mathbf{A}_n|}}\textstyle\sum_{j\geq\lfloor|\mathbf{B}_{b_n}|/\delta\rfloor}\big\|\tfrac{f(X)}{\eta(n)}\big\|_{2+\varepsilon}\big\|W_{\phi,j\delta}-W_{\phi,(j+1)\delta}\big\|_{L_{2+\varepsilon}}\alpha^{\frac{\varepsilon}{2+\varepsilon}}(j\delta|\mathbb{G}) \;.$$

Since the above holds for any $\phi \in \mathbb{G}$ and $\delta > 0$, we conclude that

$$\text{(a)} \;\leq\; 4\big\|\tfrac{f(X)}{\eta(n)}\big\|^2_{2+\varepsilon}\int_{\mathbb{G}\setminus\mathbf{B}_{b_n}}\alpha^{\frac{\varepsilon}{2+\varepsilon}}(d(e,\phi)|\mathbb{G})d|\phi| \;.$$

To bound (b), we apply the triangle inequality to obtain

$$\Big|\mathbb{E}\big[\int_{A_n}\tfrac{f(\phi X)(t(W)-t(W^\phi_{b_n})-\Delta^\phi_{b_n}t'(W))}{Z_n}d|\phi|\big]\Big|$$
$$\leq \Big|\mathbb{E}\big[\int_{A_n}\mathbb{I}\{f(\phi X)\leq\gamma_n\}\tfrac{f(\phi X)(t(W)-t(W^\phi_{b_n})-\Delta^\phi_{b_n}t'(W))}{Z_n}d|\phi|\big]\Big|$$

$$+ \Big|\mathbb{E}\Big[\int_{A_n} \mathbb{I}\{f(\phi X) > \gamma_n\} \frac{f(\phi X)(t(W) - t(W_{b_n}^{\phi}) - \Delta_{b_n}^{\phi} t'(W))}{Z_n} d|\phi|\Big]\Big| =: (\text{b1}) + (\text{b2}) .$$

Since $t \in \mathcal{F}$ implies in particular

$$|t(x+h) - t(x) - ht'(x)| \le 2|h| \sup_{y \in [x, x+h]} |t'(y)| , \tag{6.3}$$

we have

$$\begin{aligned}
(\text{b1}) &\le 2\Big|\mathbb{E}\big[\tfrac{1}{Z_n} \int_{\mathbf{A}_n} |f(\phi X)| \mathbb{I}\{|f(\phi X)| > \gamma_n\} |\Delta_{b_n}^{\phi}| d|\phi|\big]\Big| \\
&\le 2\big\| \tfrac{f(X)\mathbb{I}\{|f(X)| \ge \gamma_n\}}{\eta(n)} \big\|_{2+\varepsilon} \big\| \tfrac{f(X)}{\eta(n)} \big\|_{2+\varepsilon} \frac{\int_{\mathbf{A}_n^2} \mathbb{I}\{d(\phi, \phi') \le b_n\} |d\phi||d\phi'|}{|\mathbf{A}_n|} \\
&\le 2|\mathbf{B}_{b_n}| \big\| \tfrac{f(X)\mathbb{I}\{|f(X)| \ge \gamma_n\}}{\eta(n)} \big\|_{2+\varepsilon} \big\| \tfrac{f(X)}{\eta(n)} \big\|_{2+\varepsilon}
\end{aligned}$$

The second term (b2) can be bounded using a Taylor expansion, where we have to consider triples of transformations $\phi_1, \phi_2, \phi_3 \in \mathbb{G}$. For $p, q > 0$ with $\frac{1}{p} + \frac{1}{q} = 1$, we obtain

$$\begin{aligned}
(\text{b2}) &\le \Big| \int_{\mathbf{A}_n^3} \frac{\mathbb{E}\big[f(\phi_1 X)\mathbb{I}\{|f(\phi_1 X)| \le \gamma_n\}\mathbb{I}\{d(\phi_1, \phi_2), d(\phi_1, \phi_3) \le b_n\} f(\phi_2 X) f(\phi_3 X)\big]}{Z_n^3} |d\phi_{1:3}| \Big| \\
&\le \frac{8|\mathbf{B}_{b_n}|}{\sqrt{|\mathbf{A}_n|}} \big\| \tfrac{f(X)}{\eta(n)} \big\|_{2q(1+\frac{\varepsilon}{2})}^2 \big\| f(X)\mathbb{I}(f(X) \le \gamma_n) \big\|_{p(1+\frac{\varepsilon}{2})} \int_{\mathbb{G}} \alpha^{\frac{\varepsilon}{2+\varepsilon}}(d(e, \phi)|\mathbb{G}) d|\phi| .
\end{aligned}$$

For (c), we again apply the triangle inequality, which yields

$$\begin{aligned}
(\text{c}) \cdot \sqrt{\tfrac{\pi}{2}} &= \Big\| \frac{\eta(n)^2 - \int_{\mathbf{A}_n^2} \frac{1}{|\mathbf{A}_n|} \mathbb{E}[\mathbb{I}\{d(\phi, \phi') \le b_n\} f(\phi X) f(\phi' X)|\mathbb{G}] |d\phi||d\phi'|}{\eta(n)^2} \Big\| \\
&\le \mathbb{E}\big[\big| \tfrac{\eta(n)^2 - \eta_{b_n}^2}{\eta(n)^2} \big|\big] + \Big\| \frac{\eta_{b_n}^2 - \int_{\mathbf{A}_n^2} |\mathbf{A}_n|^{-1} \mathbb{E}[\mathbb{I}\{d(\phi, \phi') \le b_n\} f(\phi X) f(\phi' X)|\mathbb{G}] |d\phi||d\phi'|}{\eta(n)^2} \Big\| \\
&\le \mathbb{E}\big[\big| \tfrac{\eta(n)^2 - \eta_{b_n}^2}{\eta(n)^2} \big|\big] + \big\| \tfrac{f(X)}{\eta(n)} \big\|_2^2 \tfrac{|\mathbf{A}_n \triangle \mathbf{B}_{b_n} \mathbf{A}_n|}{|\mathbf{A}_n|} .
\end{aligned}$$

It remains to bound (d). To this end, we denote

$$f^{\ge \gamma_n}(x) := f(x)\mathbb{I}\{|f(x)| \ge \gamma_n\} ,$$

and $f^{<\gamma_n}$ is defined analogously. We then have

$$\Big\| \int_{\mathbf{A}_n \times \mathbf{A}_n \mathbf{B}_{b_n}} \frac{f^{\ge \gamma_n}(\phi X) f(\phi' X) - \mathbb{E}[f^{\ge \gamma_n}(\phi X) f(\phi' X)|\mathbb{G}]}{\eta(n)^2 |\mathbf{A}_n|} |d\phi||d\phi'| \Big\|_1 \le |\mathbf{B}_{b_n}| \big\| \tfrac{f^{\ge \gamma_n}(X)}{\eta(n)} \big\|_2 \big\| \tfrac{f(X)}{\eta(n)} \big\|_2 .$$

Abbreviate $F_{\phi\phi'} := \frac{1}{\eta(n)^2}\big(f^{<\gamma_n}(\phi X)f^{<\gamma_n}(\phi' X) - \mathbb{E}[f^{<\gamma_n}(\phi X)f^{<\gamma_n}(\phi' X)|\mathbb{G}]\big)$. We then observe that, for any quadruple $\phi_1, \ldots, \phi_4 \in \mathbb{G}$,

$$\big\|\mathrm{Cov}[F_{\phi_1,\phi_2}, F_{\phi_3,\phi_4}|\mathbb{G}]\big\|_1 \leq 4\big\|\tfrac{f(X)\mathbb{I}\{|f(X)|\leq\gamma_n\}}{\eta(n)}\|^4_{4+2\varepsilon} \\ \alpha^{\frac{\varepsilon}{2+\varepsilon}}\big(d((\phi_1,\phi_2),(\phi_3,\phi_4))|\mathbb{G}\big).$$

We hence have

$$\|\int_{\mathbf{A}_n\times\mathbf{A}_n\mathbf{B}_{b_n}} \frac{1}{|\mathbf{A}_n|}F_{\phi\phi'}|d\phi||d\phi'|\|_1 \\ \leq 4\frac{|\mathbf{B}_{b_n}|}{\sqrt{|\mathbf{A}_n|}}\|\tfrac{f(X)\mathbb{I}\{|f(X)|\leq\gamma_n\}}{\eta(n)}\|^2_{4+2\varepsilon}(\int_{\mathbb{G}}\alpha^{\frac{\varepsilon}{2+\varepsilon}}(d(e,\phi)|\mathbb{G})|d\phi|)^{\frac{1}{2}}\;,$$

which in turn yields

$$(\mathrm{d})\cdot\sqrt{\tfrac{\pi}{2}} \leq 4\frac{|\mathbf{B}_{b_n}|}{\sqrt{|\mathbf{A}_n|}}\big\|\tfrac{f(X)}{\eta(n)}\mathbb{I}\{|f(X)|\leq\gamma_n\}\big\|^2_{4+2\varepsilon}(\int_{\mathbb{G}}\alpha^{\frac{\varepsilon}{2+\varepsilon}}(d(e,\phi)|\mathbb{G})|d\phi|)^{\frac{1}{2}} \\ + 2|\mathbf{B}_{b_n}|\big\|\tfrac{f^{\geq\gamma_n}(X)}{\eta(n)}\big\|_2\big\|\tfrac{f(X)}{\eta(n)}\big\|_2\;.$$

Step 3: Deriving the central limit theorem. To deduce Theorem 3.2, we need to prove that $d_W(S_n, \eta Z) \to 0$, where $Z \sim N(0,1)$ is chosen to be an independent normal. For this we must choose suitable sequences (b_n) and (γ_n). To this end, we collect terms from the upper bounds above, and write

$$r^1_n := \|f(X)\mathbb{I}\{|f(X)| \geq \gamma_n\}\|_2 \quad r^2_n := \frac{|\mathbf{B}_{b_n}|\gamma_n^2}{\sqrt{|\mathbf{A}_n|}} + |\mathbf{B}_{b_n}|r^1_n \quad r^3_n := \frac{|\mathbf{A}_n\triangle\mathbf{B}_{b_n}\mathbf{A}_n|}{|\mathbf{A}_n|}\;.$$

The Følner sets satisfy $|\mathbf{A}_n| \to \infty$, and it is hence always possible to choose a sequence (b_n) growing slowly enough to guarantee $r^3_n \to 0$. Similarly, given (b_n), we can choose a divergent sequence (γ_n) such that $r^1_n \to 0$ and $r^2_n \to 0$. With these in place, we can choose two further divergent sequnces (δ_n) and (ε_n) such that

$$\varepsilon_n < \delta_n \text{ for all } n \quad\text{ and }\quad \frac{\delta_n}{\varepsilon_n^3}\big(r^2_n + r^3_n\big) \to 0 \quad\text{ as } n \to \infty\;,$$

and construct random variables $\eta(n)$ from the asymptotic variance η as

$$\eta(n) := \eta\mathbb{I}\{\eta \in [\varepsilon_n, \delta_n]\} + \varepsilon_n\mathbb{I}\{\eta \notin [\varepsilon_n, \delta_n]\}\;. \tag{6.4}$$

For $S_n := \sqrt{|\mathbf{A}_n|}\mathbb{F}_n(X)$, we then have

$$d_{\mathrm{W}}(S_n, \eta Z) \;\leq\; d_{\mathrm{W}}(S_n, \eta(n)Z) + d_{\mathrm{W}}(\eta(n)Z, \eta Z) \\ \leq\; d_{\mathrm{W}}(S_n, \eta(n)Z) + \|Z\|_1\|(\eta - \varepsilon_n)\mathbb{I}\{\eta \notin [\varepsilon_n, \delta_n]\}\|_1\;.$$

Since $\|\eta\|_1 < \infty$, the last term satisfies

$$\|Z\|_1 \|(\eta - \varepsilon_n)\mathbb{I}\{\eta \notin [\varepsilon_n, \delta_n]\}\|_1 \to 0 \qquad \text{as } \delta_n \to \infty \text{ and } \varepsilon_n \to 0 \ .$$

It hence suffices to show $d_{\mathrm{W}}(S_n, \eta(n)Z) \to 0$. To do this we first notice that the following lemma holds.

Lemma 6.3. *Let X, Y, and Z be random variables in $\mathbf{L}_2(\mathbb{R})$, such that $Y \geq a$ a.s. for some $a > 0$ and (X, Y) and Z are independent. Then*

$$d_{\mathrm{W}}(XY, Z) \leq d_{\mathrm{W}}(X, Z/Y)/a \ .$$

Proof. For all $\epsilon > 0$ exists, by definition of the Wasserstein distance, a coupling (X', Y') and Z' of (X, Y) and Z such that

$$\mathbb{E}[X'Y' - Z'] \leq d_{\mathrm{W}}(XY, Z) + \epsilon \ .$$

As Y is lower-bounded, this coupling satisfies

$$\mathbb{E}[|X'Y'/Y' - Z'/Y'|] \leq \mathbb{E}[|X'Y' - Z'|]/a \leq \big(d_{\mathrm{W}}(XY, Z) + \epsilon\big)/a$$

Since $X'Y'/Y' \stackrel{\mathrm{d}}{=} X'$, the variables $X'Y'/Y'$ and Z'/Y' are a valid coupling of X and Z/Y, which implies

$$d_{\mathrm{W}}(X, Y/Z) \leq \mathbb{E}[|X' - Z'/Y'|] \leq \big(d_{\mathrm{W}}(XY, Z) + \epsilon\big)/a$$

for all $\epsilon > 0$. □

We have hence shown that

$$d_{\mathrm{W}}(S_n, \eta(n)Z) \ \leq \ \delta_n d_{\mathrm{W}}\big(\tfrac{S_n}{\eta(n)}, \, Z\big) \ . \qquad (6.5)$$

Application of (6.2) to the right-hand side of (6.5) then yields Theorem 3.2.

Step 4: Deriving the Berry-Esseen bound. Theorem 3.3 follows almost immediately from (6.2). In this case, the sequence (b_n) is given by hypothesis, and we set $\eta(n) = \eta$ for all n. In the bound on (b2) above, we choose $p = \frac{3}{2}$ and $q = 3$. For any choice of $\gamma_n > 0$, we then have

$$\|f(X)\mathbb{I}\{|f(X) \leq \gamma_n|\}\|_{\frac{3}{2}} \ \leq \ \|f(X)\|_3 \ .$$

We can hence substitute the bounds on (a), (b), (c), and (d) above into (6.2). Each of these bounds depends on a constant γ_n, and we choose a sequence (γ_n) with $\gamma_n \to \infty$. Theorem 3.3 then follows from (6.2) for $n \to \infty$.

6.1 Proofs of the general limit theorems

6.1.1 Notation

Since the proofs use Stein's method, they are inevitably notation-heavy, and some additional abbreviations will proof useful. The concatenation of two vectors u and v is denoted $[u, v]$. Throughout, (k_n), and (b_n) are the non-decreasing integer sequences as used in the theorems. For $\boldsymbol{\phi} \in \mathbb{G}^k$, we write

$$\partial(\boldsymbol{\phi}) := \min_{i \neq j} d(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j) \ .$$

In addition to (k_n), we require a second non-decreasing sequence (k'_n) with $k'_n \leq k_n$. We abbreviate

$$\mu_n^*(A|A_n^{k_n}) = \mathbb{E}_{\mu_n}\big[\mathbb{I}\{\boldsymbol{\phi} \in A, \partial(\boldsymbol{\phi}) \geq b_n\}\big|A_n^{k_n}\big] \quad \text{for } A \subset \mathbb{G} \text{ measurable,} \tag{6.6}$$

which is a random measure on $\mathbb{G}^{k_n}$. For the functions g_n in Section 3.4, we write, with slight abuse of notation,

$$M_p(g_n) := \sup_{\boldsymbol{\phi} \in \mathbb{G}^{k_n}} \big\|\tfrac{g_n(\boldsymbol{\phi} X_n)}{\eta_n}\big\|_{L_p} \ , \qquad \text{and} \qquad C_p := \sum_{i=1}^{\infty} c_{i,p} \ ,$$

where $c_{i,p}$ are the Lipschitz coefficients. In the functions h_n, we repeatedly have to average out the effect of transformation vector $\boldsymbol{\phi}' \in \mathbb{G}^{k_n}$ whose coordinates are "close" to a given transformation ϕ. To this end, the set of indices of such coordinates

$$\mathcal{I}_{b,k}(\phi, \boldsymbol{\phi}') := \{i \leq k : \ d(\phi, \boldsymbol{\phi}'_i) \leq b\} \quad \text{for } k \leq k_n, b > 0 \ .$$

We then average, in a Følner sense, by defining

$$\bar{h}_n^{\phi,b,k}(\boldsymbol{\phi}' X_n) := \lim_p \frac{1}{|\mathbf{A}_p|^{|\mathcal{I}_{b,k}(\phi,\boldsymbol{\phi}')|}} \int_{\{\boldsymbol{\theta}|\boldsymbol{\theta}_i = \boldsymbol{\phi}'_i, i \notin \mathcal{I}_{b,k}(\phi,\boldsymbol{\phi}')\}} h_n(\boldsymbol{\theta} X_n)|d\boldsymbol{\theta}| \ .$$

Coordinate entries far away from ϕ are held fixed. Rouhgly, this corresponds to integrating out the marked area in Figure 2.1. For the central limit theorem, which assumes the random measures μ_n are well-spread (but not necessarily strongly well-spread), the moments of (μ_n) are controlled using a sequence (β_n) with $\beta_n \to \infty$, and the subsets of $\mathbb{G}^{k_n}$ defined as

$$V_{i,\beta_n}(n) := \Big\{\boldsymbol{\phi}' \in \mathbb{G}^{k_n} \ \Big| \ \sup_{j \leq k'_n} \frac{|\mathbf{A}_n|}{|\mathbf{B}_{b_n}|} P_{\mu_n^*}(d(\boldsymbol{\phi}_i, \boldsymbol{\phi}'_j) \leq b_n|\mathbf{A}_n^{k_n}, \boldsymbol{\phi}') \leq k'_n \beta_n\Big\} \ .$$

Similarly, a sequence (γ_n) with $\gamma_n \to \infty$ is used to control higher moments of $\frac{h_n}{\eta(n)}$. Given (γ_n), we write

$$\Gamma_{i,p}(\gamma_n) := \sup_{\phi \in \mathbb{G}^{k_n}} \Big\| \frac{\bar{h}_n^i(\boldsymbol{\phi} X_n)\mathbb{I}\{|\bar{h}_n^i(\boldsymbol{\phi} X_n)| \leq \gamma_n c_{i,2}(g_n)\}}{\eta(n)} \Big\|_p \qquad \text{for } i \leq k_n \; .$$

For a strongly well-spread sequence, the spreading coefficient $\mathcal{S}$ was defined in Section 3.4. A similar coefficients in the well-spread case is

$$\mathcal{S}_w \; := \sup_{\substack{\mathcal{S}_1, \mathcal{S}_2 \subset \mathbb{N} \\ A \in \mathcal{A}_{\text{card}(\mathcal{S}_1 \times \mathcal{S}_2)} \\ n \in \mathbb{N}}} \mathbb{E}\Big[\frac{|\mathbf{A}_n|^{\text{card}(\mathcal{S}_1 \times \mathcal{S}_2)}}{|A|} \mathbb{E}_{\mu_n \otimes \mu_n}\big(\mathbb{I}\big([\{\boldsymbol{\Phi}\}, \{\boldsymbol{\Phi'}\}]_{\mathcal{S}_1}^{\mathcal{S}_2} \in A]\big) \big| \boldsymbol{\Phi}, \boldsymbol{\Phi'} \in A_n^{k_n}\big) \Big]$$

6.1.2 Main lemmas

Recall that, in the proof of the basic case, we specialized Stein's inequality to (6.2), bounded the constituent terms individually, and then deduced both limit theorems from the resulting bound. In the general case, Lemmas 6.5 and 6.6 below substitute for (6.2). The main work of the proof is then once again to upper-bound individual terms. We first state an auxiliary result:

LEMMA 6.4. *Under the conditions of Theorem 3.6, for all $m \in \mathbb{N}$,*

$$\frac{\sum_{i \geq m} |\mathbf{B}_{i+1} \setminus \mathbf{B}_i| \alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(i|\mathbb{G})}{\int_{\mathbb{G} \setminus \mathbf{B}_{m-1}} \alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(d(e,\phi)|\mathbb{G}) d|\phi|} < \infty$$

PROOF. Abbreviate $r := \sup_i \frac{|\mathbf{B}_{i+1} \setminus \mathbf{B}_i|}{|\mathbf{B}_i \setminus \mathbf{B}_{i-1}|}$. Then

$$\begin{aligned} \textstyle\sum_{i \geq m} |\mathbf{B}_{i+1} \setminus \mathbf{B}_i| \alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(i|\mathbb{G}) &\leq r \textstyle\sum_{i \geq m} |\mathbf{B}_i \setminus \mathbf{B}_{i-1}| \alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(i|\mathbb{G}) \\ &\leq r \int_{\mathbb{G} \setminus \mathbf{B}_{m-1}} \alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(d(e,\phi)|\mathbb{G}) d|\phi| \; , \end{aligned}$$

where we have used assumption (3.1). □

The two main lemmas for the proof of Theorems 3.5 and 3.6 are the following:

LEMMA 6.5. *For a positive random variable $\eta(n)$ with $\eta(n) \perp\!\!\!\perp_{\mathbb{G}} X_n$,*

$$\begin{aligned} \Big| d_W\Big(\tfrac{1}{\eta(n)} \mathbb{E}_{\mu_n}[h_n(\boldsymbol{\phi} X_n) | \mathbf{A}_n^{k_n}], Z^*\Big) - d_W\Big(\tfrac{1}{\eta(n)} \mathbb{E}_{\mu_n^*}[h_n(\boldsymbol{\phi} X_n) | \mathbf{A}_n^{k_n}], Z^*\Big) \Big| \\ \leq \frac{k_n^2 C_1(\frac{g_n}{\eta(n)}) |\mathbf{B}_{b_n}| S_w}{\sqrt{|\mathbf{A}_n|}} \; . \end{aligned}$$

PROOF. By definition of the spreading coefficient $\mathcal{S}_w$,

$$
\begin{aligned}
&\Big\|\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n}[\mathbb{I}\{\partial(\boldsymbol{\phi})\leq b_n\}h_n(\boldsymbol{\phi}X_n)|\mathbf{A}_n^{k_n}]\Big\|_1\\
&\leq M_1\big(\tfrac{g_n(\cdot)}{\eta(n)}\big)\mathbb{E}[\sqrt{|\mathbf{A}_n|}\mathbb{P}_{\mu_n}(\partial(\boldsymbol{\phi})\leq b_n|\mathbf{A}_n^{k_n})]\\
&\leq \frac{k_n^2 M_1\big(\tfrac{g_n(\cdot)}{\eta(n)}\big)|\mathbf{B}_{b_n}|}{\sqrt{|\mathbf{A}_n|}}\sup_{i\neq j}\mathbb{E}\big[\tfrac{|\mathbf{A}_n|}{|\mathbf{B}_{b_n}|}\mathbb{P}_{\mu_n^*}(\mathbb{I}\{\boldsymbol{\phi}_i^{-1}\boldsymbol{\phi}_j\in B_{b_n}\}|\mathbf{A}_n^{k_n})\big]\\
&\leq \frac{k_n^2 M_1\big(\tfrac{g_n(\cdot)}{\eta(n)}\big)|\mathbf{B}_{b_n}|\mathcal{S}_w}{\sqrt{|\mathbf{A}_n|}}.
\end{aligned}
$$

That yields the desired result using,

$$
\begin{aligned}
&\Big|d_W\Big(\tfrac{1}{\eta(n)}\mathbb{E}_{\mu_n}[h_n(\boldsymbol{\phi}X_n)|\mathbf{A}_n^{k_n}],Z^*\Big)-d_W\Big(\tfrac{1}{\eta(n)}\mathbb{E}_{\mu_n^*}[h_n(\boldsymbol{\phi}X_n)|\mathbf{A}_n^{k_n}],Z^*\Big)\Big|\\
\leq\ & d_W\Big(\tfrac{1}{\eta(n)}\mathbb{E}_{\mu_n}[h_n(\boldsymbol{\phi}X_n)|\mathbf{A}_n^{k_n}],\tfrac{1}{\eta(n)}\mathbb{E}_{\mu_n^*}[h_n(\boldsymbol{\phi}X_n)|\mathbf{A}_n^{k_n}]\Big)\\
\leq\ & \Big\|\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}[\mathbb{I}\{\partial(\boldsymbol{\phi})\leq b_n\}h_n(\boldsymbol{\phi}X_n)|\mathbf{A}_n^{k_n}]\Big\|_1.
\end{aligned}
$$

□

LEMMA 6.6. *Let $\eta(n)$ be a positive random variable with $\eta(n)\perp\!\!\!\perp_{\mathbb{G}}X_n$, and $\mathcal{F}$ the function class* (3.18). *Let*

$$W^*:=\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i\mathbb{E}_{\mu_n^*}[\bar{h}_n^i(\boldsymbol{\phi}X_n)|\mathbf{A}_n^{k_n}]\ ,$$

and abbreviate, for given sequences (b_n) and (k_n'),

$$W_{in}^{\boldsymbol{\phi}}:=\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}[\bar{h}_n^{\boldsymbol{\phi}_i,b_n,k_n}(\boldsymbol{\phi}'X_n)|\mathbf{A}_n^{k_n}]\qquad\textit{and}\qquad\Delta_{in}^{\boldsymbol{\phi}}=W^*-W_{in}^{\boldsymbol{\phi}}\ .$$

Then, for an independent variable $Z^*\sim N(0,1)$,

$$
\begin{aligned}
&d_W(W^*,Z^*)\\
&\leq\sup_{t\in\mathcal{F}}\Big|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi}X_n)t(W_{in}^{\boldsymbol{\phi}})|\mathbf{A}_n^{k_n}\big]\big]\Big|\\
&+\sup_{t\in\mathcal{F}}\Big|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi}X_n)(t(W^*)-t(W_{in}^{\boldsymbol{\phi}})-\Delta_{in}^{\boldsymbol{\phi}}t'(W^*))|\mathbf{A}_n^{k_n}\big]\big]\Big| \qquad (6.7)\\
&+\sqrt{\tfrac{2}{\pi}}\Big\|1-\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i\mathbb{E}[\mathbb{E}_{\mu_n^*}[\bar{h}_n^i(\boldsymbol{\phi}X_n)\Delta_{in}^{\boldsymbol{\phi}}|\mathbf{A}_n^{k_n}]|\mathbb{G}]\Big\|\\
&+\sqrt{\tfrac{2}{\pi}}\textstyle\sum_i\Big\|\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi}X_n)\Delta_{in}^{\boldsymbol{\phi}}-\mathbb{E}[\bar{h}_n^i(\boldsymbol{\phi}X_n)\Delta_{in}^{\boldsymbol{\phi}}|\mathbb{G}]\big|\mathbf{A}_n^{k_n}\big]\Big\|_1\ .
\end{aligned}
$$

PROOF. By the Stein inequality,

$$d_W(W^*,Z^*)\leq\sup_{t\in\mathcal{F}}|\mathbb{E}[W^*t(W^*)-t'(W^*)]|\ .$$

We decompose the term on the right-hand side. Since $h_n = \sum_i \bar{h}_n^i$,

$$|\mathbb{E}[W^* t(W^*) - t'(W^*)]| \leq \mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)} \sum_i \mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi} X_n) t(W_{in}^{\boldsymbol{\phi}}) | \mathbf{A}_n^{k_n}\big]\big]$$
$$+ \mathbb{E}\Big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)} \sum_i \mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi} X_n)(t(W^*) - t(W_{in}^{\boldsymbol{\phi}}))\big|\mathbf{A}_n^{k_n}\big] - t'(W^*)\Big]$$

The final term can be bounded further as

$$\begin{aligned}
&\mathbb{E}\Big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)} \textstyle\sum_i \mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi} X_n)\big(t(W^*) - t(W_{in}^{\boldsymbol{\phi}})\big)\big|\mathbf{A}_n^{k_n}\big] - t'(W^*)\Big] \\
&\leq \Big|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)} \textstyle\sum_i \mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi} X_n)(t(W^*) - t(W_{in}^{\boldsymbol{\phi}}) - \Delta_{in}^{\boldsymbol{\phi}} t'(W^*))|\mathbf{A}_n^{k_n}\big]\big]\Big| \\
&+ \Big|\mathbb{E}\big[\textstyle\sum_i t'(W^*)\big(1 - \mathbb{E}_{\mu_n^*}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)} \bar{h}_n^i(\boldsymbol{\phi} X_n)\Delta_{in}^{\boldsymbol{\phi}}|\mathbf{A}_n^{k_n}\big]\big)\big]\Big| \\
&\leq \Big|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)} \textstyle\sum_i \mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi} X_n)(t(W^*) - t(W_{in}^{\boldsymbol{\phi}}) - \Delta_{in}^{\boldsymbol{\phi}} t'(W^*))|\mathbf{A}_n^{k_n}\big]\big]\Big| \\
&+ \sqrt{\tfrac{2}{\pi}}\big\|1 - \tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)} \textstyle\sum_i \mathbb{E}[\mathbb{E}_{\mu_n^*}[\bar{h}_n^i(\boldsymbol{\phi} X_n)\Delta_{in}^{\boldsymbol{\phi}}|\mathbf{A}_n^{k_n}]|\mathbb{G}]\big\| \\
&+ \sqrt{\tfrac{2}{\pi}} \textstyle\sum_i \Big\|\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)} \mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi} X_n)\Delta_{in}^{\boldsymbol{\phi}} - \mathbb{E}[\bar{h}_n^i(\boldsymbol{\phi} X_n)\Delta_{in}^{\boldsymbol{\phi}}|\mathbb{G}]\big|\mathbf{A}_n^{k_n}\big]\Big\|_1.
\end{aligned}$$

□

6.1.3 Bounding the first term in Lemma 6.6

We will now proceed to upper-bound each of the four terms on the right-hand side of 6.6 separately. To bound the first term, we observe:

Lemma 6.7. *Under the conditions of Theorem 3.6, the inequalities*

$$\|\widehat{\mathbb{F}}_{\infty,i}(h_n, X_n, e)\|_{L_p} \leq c_{i,p}(g_n) \quad \text{and} \quad \mathbb{E}(\mathbb{E}_{\mu_n^{i-j}}[\mathbb{I}_{\mathbf{B}_b}]) \leq \mathcal{S}_w|\mathbf{B}_b|$$

hold for all $i, n, b \in \mathbb{N}$ and all $p \in \mathbb{R}$.

Proof. From the definition of $\widehat{\mathbb{F}}$, we obtain the first inequality as

$$\begin{aligned}
\|\widehat{\mathbb{F}}_{\infty,i}(h_n, X_n, e)\| &= \Big\|\lim_p \frac{1}{|\mathbf{A}_p|^{k_n}} \int_{\mathbf{A}_p^{k_n}} h_n(\boldsymbol{\phi}_{1:i-1} e \boldsymbol{\phi}_{i+1:k_n} X_n) - h_n(\boldsymbol{\phi} X_n) d|\boldsymbol{\phi}|\Big\|_{L_p} \\
&\leq \lim_p \frac{1}{|\mathbf{A}_p|^{k_n}} \int_{\mathbf{A}_p^{k_n}} \|h_n(\boldsymbol{\phi}_{1:i-1} e \boldsymbol{\phi}_{i+1:k_n} X_n) - h_n(\boldsymbol{\phi} X_n)\| d|\boldsymbol{\phi}| \leq c_{i,p}(g_n)\ ,
\end{aligned}$$

and $\mathbb{E}[\mathbb{E}_{\mu_n^{i-j}}[\mathbb{I}_{B_b}]] = |A_n|\mathbb{E}[\mathbb{E}_{\mu_n^{\otimes 2}}[\mathbb{I}_{\boldsymbol{\phi}_j^{-1}\boldsymbol{\phi}_i' \in B_b}|\mathbf{A}_n^{2k_n}]] \leq \mathcal{S}_w|\mathbf{B}_b|$ yields the second statement. □

Lemma 6.8. *Assume hypothesis* (3.10). *Then*

$$\sup_{t \in \mathcal{F}} \mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)} \textstyle\sum_i \mathbb{E}_{\mu_n^*}[\bar{h}_n^i(\boldsymbol{\phi} X_n) t(W_{in}^{\boldsymbol{\phi}})|\mathbf{A}_n^{k_n}]\big] \leq K_1 C_2\big(\tfrac{g_n}{\eta(n)}\big) \textstyle\sum_{k_n' < i} c_{i,2}\big(\tfrac{g_n}{\eta(n)}\big),$$

where $K_1 = O(|B_K|\mathcal{S}_w)$. *If hypothesis* (3.11) *holds instead,*

$$\sup_{t\in\mathcal{F}} \mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)} \textstyle\sum_i \mathbb{E}_{\mu_n^*}[\bar{h}_n^i(\boldsymbol{\phi}X_n)t(W_{in}^{\boldsymbol{\phi}})|\mathbf{A}_n^{k_n}]\big]$$
$$\leq K_1 C_{2+\varepsilon}\big(\tfrac{g_n}{\eta(n)}\big)\big[\tfrac{k_n}{\sqrt{|\mathbf{A}_n|}} + C_{2+\varepsilon}\big(\tfrac{g_n}{\eta(n)}\big)\big]\mathcal{R}_{b_n} + K_2|\mathbf{B}_{b_n}|C_2\big(\tfrac{g_n}{\eta(n)}\big)\textstyle\sum_{k_n'<i} c_{i,2}\big(\tfrac{g_n}{\eta(n)}\big)\ ,$$

where $K_1 = O(S_w)$ *and* $K_2 = O(\mathcal{S}_w)$.

Proof. We proof the (harder) case of hypothesis (3.11) first, and then highlight changes required for (3.10). Similar to $W_{in}^{\boldsymbol{\phi}}$, we abbreviate

$$W_{ibk}^{\boldsymbol{\phi}} := \tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}[\bar{h}_n^{\boldsymbol{\phi}_i,b,k}(\boldsymbol{\phi}'X_n)|\mathbf{A}_n^{k_n}]\ ,$$

so that in particular $W_{in}^{\boldsymbol{\phi}} = W_{ib_nk_n'}^{\boldsymbol{\phi}}$. For all $t \in \mathcal{F}$,

$$\begin{aligned} &\textstyle\sum_i \Big|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi}X_n)t(W_{in}^{\boldsymbol{\phi}})\big|\mathbf{A}_n^{k_n}\big]\big]\Big| \\ \overset{(*)}{\leq}\ &\textstyle\sum_i \mathbb{E}\Big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big[|\bar{h}_n^i(\boldsymbol{\phi}X_n)||W_{in}^{\boldsymbol{\phi}} - W_{ib_nk_n}^{\boldsymbol{\phi}}|\big|\mathbf{A}_n^{k_n}\big]\Big] \\ +\ &\textstyle\sum_i \Big|\mathbb{E}\Big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi}X_n)t(W_{ib_nk_n}^{\boldsymbol{\phi}})\big|\mathbf{A}_n^{k_n}\big]\Big]\Big| \end{aligned} \tag{6.8}$$

where $(*)$ holds since t is 1-Lipschitz. To bound the first item on the right-hand side, we use the definition the Lipschitz coefficients of g_n, to obtain

$$\begin{aligned} &\textstyle\sum_i \mathbb{E}\Big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big[\big|\bar{h}_n^i(\boldsymbol{\phi}X_n)\big|\big|W_{in}^{\boldsymbol{\phi}} - W_{ib_nk_n}^{\boldsymbol{\phi}}\big|\big|\mathbf{A}_n^{k_n}\big]\Big] \\ \leq\ &\textstyle\sum_i \mathbb{E}\Big[|\mathbf{A}_n|\mathbb{E}_{\mu_n^{\otimes 2}}\big[\sum_{j\in\mathcal{J}_n} c_{i,2}\big(\tfrac{g_n}{\eta(n)}\big)c_{j,2}\big(\tfrac{g_n}{\eta(n)}\big)\big|\mathbf{A}_n^{2k_n}\big]\Big] \\ \leq\ &|\mathbf{B}_{b_n}|\textstyle\sum_i\sum_{k_n'<j\leq k_n} c_{i,2}\big(\tfrac{g_n}{\eta(n)}\big)c_{j,2}\big(\tfrac{g_n}{\eta(n)}\big)\mathcal{S}_w\ , \end{aligned}$$

where we have abbreviated $\mathcal{J}_n = \mathcal{I}_{b_n,k_n}(\boldsymbol{\phi}_i, \boldsymbol{\phi}') \setminus \mathcal{I}_{b_n,k_n'}(\boldsymbol{\phi}_i, \boldsymbol{\phi}')$. To bound the second term, consider the element $\boldsymbol{\phi} \in \mathbb{G}^{k_n}$ in (6.8). We can choose a sequence $(\boldsymbol{\phi}^{i,j})_{j\in\mathbb{N}}$ in $\mathbb{G}^{k_n}$, whose coordinates differ more and more from $\boldsymbol{\phi}_i$ as j increases, as follows: Set $\boldsymbol{\phi}^{i,0} = \boldsymbol{\phi}$. For $j \geq 1$, choose

$$\boldsymbol{\phi}_k^{i,j} := \begin{cases} \boldsymbol{\phi}_k^{i,j-1} & \text{if } d(\boldsymbol{\phi}_k, \boldsymbol{\phi}_i) \notin [j, j+1) \\ \text{any } \boldsymbol{\phi}_k^{i,j} \text{ with } d(\boldsymbol{\phi}_k^{i,j}, \boldsymbol{\phi}_i) > \operatorname{diam}(\mathbf{A}_n) & \text{if } d(\boldsymbol{\phi}_k, \boldsymbol{\phi}_i) \in [j, j+1) \end{cases}$$

for each $k \leq k_n$. Then

$$\textstyle\sum_i \Big|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\bar{h}_n^i(\boldsymbol{\phi}X_n)t(W_{\boldsymbol{\phi},b_n,k_n,i})\big]\Big|$$

$$
\begin{aligned}
&\le \textstyle\sum_i \sum_{j\ge b_n} \Big|\mathbb{E}\big[\frac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\bar{h}^i_n(\boldsymbol{\phi}^{i,j+1}X_n)[t(W^{\boldsymbol{\phi}}_{ijk_n}) - t(W^{\boldsymbol{\phi}}_{i(j+1)k_n})]\big]\Big| \\
&+ \textstyle\sum_i \sum_{j\ge b_n} \Big|\mathbb{E}\big[\frac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}[\bar{h}^i_n(\boldsymbol{\phi}^{i,j+1}X_n) - \bar{h}^i_n(\boldsymbol{\phi}^{i,j}X_n)]t(W^{\boldsymbol{\phi}}_{ijk_n})\big]\Big| \\
&\le 4\sqrt{\tfrac{2}{\pi}} \textstyle\sum_{l,\ l\ge b_n} \sum_i c_{i,2+\varepsilon}\big(\frac{g_n}{\eta(n)}\big)|\mathbf{A}_n|\big\|W^{\boldsymbol{\phi}}_{ilk_n} - W^{\boldsymbol{\phi}}_{i(l+1)k_n}\big\|_{2+\varepsilon}\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(l|\mathbb{G}) \\
&+ 4\textstyle\sum_i c_{i,2+\varepsilon}\big(\frac{g_n}{\eta(n)}\big)\sum_{l\ge b_n} \alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(l|\mathbb{G})\sqrt{|\mathbf{A}_n|}\mathbb{I}\{d(\boldsymbol{\phi}_i, \boldsymbol{\phi}_{\setminus i}) \in [l, l+1]\}\ .
\end{aligned}
$$

Since that is true for any $\boldsymbol{\phi} \in \mathbb{G}^{k_n}$, we conclude

$$
\begin{aligned}
&\textstyle\sum_i \Big|\mathbb{E}\big[\frac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big(\bar{h}^i_n(\boldsymbol{\phi}X_n)t(W^{\boldsymbol{\phi}}_{in})\big|\mathbf{A}_n^{k_n}\big)\big]\Big| \\
&\le 4\sqrt{\tfrac{2}{\pi}}\textstyle\sum_i c_{i,2+\varepsilon}\big(\frac{g_n}{\eta(n)}\big) \\
&\textstyle\sum_j c_{j,2+\varepsilon}\big(\frac{g_n}{\eta(n)}\big) \quad \mathbb{E}\big[\mathbb{E}_{\mu_n^{\otimes 2}}\big(\mathbb{I}\{j \notin \mathcal{I}_{b_n,k_n}(\boldsymbol{\phi}_i, \boldsymbol{\phi}')\}|\mathbf{A}_n|\alpha^{\frac{\varepsilon}{2+\varepsilon}}(d(\boldsymbol{\phi}_i, \boldsymbol{\phi}'_j)|\mathbb{G})|\mathbf{A}_n^{2k_n}\big)\big] \\
&+ 4\textstyle\sum_i c_{i,2+\varepsilon}\big(\frac{g_n}{\eta(n)}\big)\sum_j \mathbb{E}\big[\mathbb{E}_{\mu_n^*}\big(\sqrt{|\mathbf{A}_n|}\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(d(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j)|\mathbb{G})|\mathbf{A}_n^{k_n}\big)\big] \\
&\le 4\textstyle\sum_i c_{i,2+\varepsilon}\big(\frac{g_n}{\eta(n)}\big)\mathcal{S}_w \\
&\quad \Big(\frac{k_n}{\sqrt{|\mathbf{A}_n|}} + \sqrt{\tfrac{2}{\pi}}\textstyle\sum_i c_{i,2+\varepsilon}\big(\frac{g_n}{\eta(n)}\big)\Big)\sum_{i\ge b_n}\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(i|\mathbb{G})|\mathbf{B}_{i+1}\setminus\mathbf{B}_i|\ .
\end{aligned}
$$

That establishes the result under (3.11). If (3.10) is assumed instead, the second term of Equation (6.8) vanishes. We hence have

$$
\sup_{t\in\mathcal{F}}\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i \mathbb{E}_{\mu_n^*}\big[\bar{h}^i_n(\boldsymbol{\phi}X_n)t(W^{\boldsymbol{\phi}}_{in})|\mathbf{A}_n^{k_n}\big]\big] \le\ |\mathbf{B}_K|\mathcal{S}_w\sum_i\sum_{k'_n<j\le k_n} c_{i,2}\big(\tfrac{g_n}{\eta(n)}\big)c_{j,2}\big(\tfrac{g_n}{\eta(n)}\big)\ ,
$$

and the result also holds under (3.10). □

6.1.4 The second term in Lemma 6.6

For the second term, we have to control interactions of random triples $\boldsymbol{\phi_1}, \boldsymbol{\phi_3}, \boldsymbol{\phi_3} \in \mathbb{G}^{k_n}$, that satisfy the condition

$$
d(\boldsymbol{\phi_1}_i, \boldsymbol{\phi_2}_j), d(\boldsymbol{\phi_1}_i, \boldsymbol{\phi_3}_l) \le b_n \quad \text{and} \quad \boldsymbol{\phi_2} \in V_{i,\beta_n}(n) \tag{6.9}
$$

for all $i, j, l \le k_n$, and either

$$
\text{(i)}\ \min_{l\le k_n} d(\boldsymbol{\phi_2}_j, \boldsymbol{\phi_3}_l) \in [k, k+1] \ \text{ or } \ \text{(ii)}\ \min_{\substack{l\le k_n\\ l\ne i}} d(\boldsymbol{\phi_2}_j, \boldsymbol{\phi_1}_l) \in [k, k+1]\ . \tag{6.10}
$$

The upper bound on the term in Lemma 6.6 must be established for fixed values of n and β_n. Given such values, we quantify the condition by choosing a constant $S_2^*(k_n)$ that satisfies

$$\frac{|\mathbf{A}_n|^2 \, \big\|\mathbb{E}_{\mu_n^{\otimes 3}}\big[\mathbb{I}\{\boldsymbol{\phi_1}, \boldsymbol{\phi_3}, \boldsymbol{\phi_3} \models (6.9) \text{ and } (6.10i)\}\big|\mathbf{A}_n^{3k_n}\big]\big\|}{|\mathbf{B}_{k+1} \setminus \mathbf{B}_k||\mathbf{B}_{b_n}|k_n} \leq S_2^*(k_n)$$

$$\text{and} \qquad \frac{|\mathbf{A}_n|^2 \, \big\|\mathbb{E}_{\mu_n^{\otimes 3}}\big[\mathbb{I}\{\boldsymbol{\phi_1}, \boldsymbol{\phi_3}, \boldsymbol{\phi_3} \models (6.9) \text{ and } (6.10ii)\}\big|\mathbf{A}_n^{3k_n}\big]\big\|}{|\mathbf{B}_{k+1} \setminus \mathbf{B}_k||\mathbf{B}_{b_n}|k_n} \leq S_2^*(k_n)\ .$$

Similarly, we choose a constant S_0^* such that

$$\frac{|\mathbf{A}_n|}{|\mathbf{B}_m|}\big\|\mathbb{E}_{\mu_n^{\otimes 2}}[\mathbb{I}\{d(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j) \leq m \text{ and } \phi' \notin V_{i,\beta_n}(n)\}|\mathbf{A}_n^{2k_n}]\big\| \leq S_0^*$$

for all $n, m \in \mathbb{N}$ and $i, j \leq k_n$.

Lemma 6.9. *Assume* (3.10) *holds. Then for* $t \in \mathcal{F}$*, and any* $p, q > 0$ *satisfying* $\frac{1}{p} + \frac{1}{q} = 1$*,*

$$\begin{aligned}
&\sup_{H\in\mathcal{F}}\Big|\mathbb{E}\Big(\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big(\bar{h}_n^i(\boldsymbol{\phi}X_n)(t(W^*) - t(W_{in}^{\boldsymbol{\phi}}) - \Delta_{in}^{\boldsymbol{\phi}}t'(W^*))|\mathbf{A}_n^{k_n}\big)\Big)\Big| \\
&\leq K_1\tfrac{k_n k_n'}{\sqrt{|\mathbf{A}_n|}}C_{2q}\big(\tfrac{g_n}{\eta(n)}\big)^2 \ S_2^*(k_n')\textstyle\sum_i \Gamma_{i,p}(\gamma_n) \\
&+ K_2 S_0^* C_2(\tfrac{g_n}{\eta(n)})^2 + K_3 C_2(\tfrac{g_n}{\eta(n)})\textstyle\sum_i c_{i,2}\big(\tfrac{\bar{h}_n^i(\boldsymbol{\phi}X_n)}{\eta(n)}\mathbb{I}\{\tfrac{\bar{h}_n^i(\boldsymbol{\phi}X_n)}{\eta(n)} \geq \gamma_n\}\big)\ ,
\end{aligned}$$

where $K_1 = O(|\mathbf{B}_k|^2)$ *and* $K_2 = O(|\mathbf{B}_K|)$ *and* $K_3 = O(\mathcal{S}_w|\mathbf{B}_K|)$*. If* (3.11) *holds instead,*

$$\begin{aligned}
&\sup_{H\in\mathcal{F}}\big|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i \mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi}X_n)(t(W^*) - t(W_{in}^{\boldsymbol{\phi}}) - \Delta_{in}^{\boldsymbol{\phi}}t'(W^*))|\mathbf{A}_n^{k_n}\big]\big]\big| \\
&\leq K_1\frac{k_n k_n'|\mathbf{B}_{b_n}|S_2^*(k_n')}{\sqrt{|\mathbf{A}_n|}}C_{(2+\varepsilon)q}\big(\tfrac{g_n}{\eta(n)}\big)^2 \ \sum_i \Gamma_{i,p(1+\frac{\varepsilon}{2})}(\gamma_n) \\
&+ K_2|\mathbf{B}_{b_n}|S_0^* C_2\big(\tfrac{g_n}{\eta(n)}\big)^2 + K_3|\mathbf{B}_{b_n}|C_2\big(\tfrac{g_n}{\eta(n)}\big)\textstyle\sum_i c_{i,2}\big(\tfrac{\bar{h}_n^i(\boldsymbol{\phi}X_n)}{\eta(n)}\mathbb{I}\{\tfrac{\bar{h}_n^i(\boldsymbol{\phi}X_n)}{\eta(n)} \geq \gamma_n\}\big)
\end{aligned}$$

where $K_1 = O(\mathcal{R}_0)$ *and* $K_2 = O(1)$ *and* $K_3 = O(\mathcal{S}_w)$*.*

Proof. Suppose first (3.10) holds. By the triangle inequality,

$$\begin{aligned}
&\big|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i \mathbb{E}_{\mu_n^*}\big[\underbrace{\bar{h}_n^i(\boldsymbol{\phi}X_n)(t(W^*) - t(W_{in}^{\boldsymbol{\phi}}) - \Delta_{in}^{\boldsymbol{\phi}}t'(W^*))}_{:=T}|\mathbf{A}_n^{k_n}\big]\big]\big| \\
&\qquad \leq \big|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i \mathbb{E}_{\mu_n^*}\big[T\,\mathbb{I}\{|\tfrac{\bar{h}_n^i(\boldsymbol{\phi}X_n)}{c_{i,2}(g_n)}| > \gamma_n\}\,|\,\mathbf{A}_n^{k_n}\big]\big]\big| \\
&\qquad + \big|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i \mathbb{E}_{\mu_n^*}\big[T\,\mathbb{I}\{|\tfrac{\bar{h}_n^i(\boldsymbol{\phi}X_n)}{c_{i,2}(g_n)}| \leq \gamma_n\}\,|\,\mathbf{A}_n^{k_n}\big]\big]\big|\ .
\end{aligned}$$

We again bound each term separately. Since $t \in \mathcal{F}$, it satisfies (6.3), hence

$$\begin{aligned}
&\big|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i \mathbb{E}_{\mu_n^*}\big[T\,\mathbb{I}\{|\tfrac{\bar{h}_n^i(\phi X_n)}{c_{i,2}(g_n)}| > \gamma_n\}\,|\,\mathbf{A}_n^{k_n}\big]\big]\big| \\
&\leq 2\textstyle\sum_i \big|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big[|\bar{h}_n^i(\phi X_n)|\mathbb{I}\{\tfrac{|\bar{h}_n^i(\phi X_n)|}{c_{i,2}(g_n)} > \gamma_n\}|\Delta_{in}^{\phi}||\mathbf{A}_n^{k_n}\big]\big]\big| \\
&\leq 2|\mathbf{A}_n|\sum_{j\leq k_n'} c_{j,2}(\frac{g_n}{\eta(n)}) \\
&\quad\cdot\textstyle\sum_i c_{i,2}\big(\tfrac{\bar{h}_n^i}{\eta(n)}\mathbb{I}\{\tfrac{|\bar{h}_n^i(\phi X_n)|}{c_{i,2}(g_n)} > \gamma_n\}\big)\mathbb{E}\big[\mathbb{E}_{\mu_n^{\otimes 2}}[\mathbb{I}\{d(\phi_i,\phi_j') \leq b_n\}|\mathbf{A}_n^{2k_n}]\big] \\
&\leq \mathcal{S}_w|\mathbf{B}_{b_n}|C_2\big(\tfrac{g_n}{\eta(n)}\big)\textstyle\sum_i c_{i,2}\big(\tfrac{\bar{h}_n^i(\phi X_n)}{\eta(n)}\mathbb{I}\{\tfrac{|\bar{h}_n^i(\phi X_n)|}{c_{i,2}(g_n)} > \gamma_n\}\big)\;.
\end{aligned}$$

To bound the second term, we abbreviate

$$\begin{aligned}
&\tilde{W}_{in}^{\phi} := \mathbb{E}_{\mu_n^*}\big[\mathbb{I}\{\phi' \in V_i(\beta_n)\}\bar{h}_n^{\phi_i,b_n,k_n'}(\phi' X_n)\big|\mathbf{A}_n^{k_n},\phi\big] \\
\text{and}\quad &\tilde{\Delta}_{in}^{\phi} := \mathbb{E}_{\mu_n^*}\big[\mathbb{I}\{\phi' \in V_i(\beta_n)\}\big(h_n(\phi' X_n) - \bar{h}_n^{\phi_i,b_n,k_n'}(\phi' X_n)\big)\big|\mathbf{A}_n^{k_n},\phi\big]\;.
\end{aligned}$$

Again using the triangle inequality, we have

$$\begin{aligned}
&\big|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i \mathbb{E}_{\mu_n^*}\big[T\,\mathbb{I}\{|\tfrac{\bar{h}_n^i(\phi X_n)}{c_{i,2}(g_n)}| \leq \gamma_n\}\,|\,\mathbf{A}_n^{k_n}\big]\big]\big| \\
\leq&\big|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\phi X_n)\mathbb{I}\{|\tfrac{\bar{h}_n^i(\phi X_n)}{c_{i,2}(g_n)}|\leq\gamma_n\}\big(t(\tilde{W}_{in}^{\phi}) - t(W_{in}^{\phi})\big)\big|\mathbf{A}_n^{k_n}\big]\big]\big| \\
+&\big|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\phi X_n)\mathbb{I}\{|\tfrac{\bar{h}_n^i(\phi X_n)}{c_{i,2}(g_n)}|\leq\gamma_n\}(\Delta_{in}^{\phi} - \tilde{\Delta}_{in}^{\phi})t'(W_{in}^{\phi})|\mathbf{A}_n^{k_n}\big]\big]\big| \\
+&\big|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\phi X_n)\mathbb{I}\{|\tfrac{\bar{h}_n^i(\phi X_n)}{c_{i,2}(g_n)}|\leq\gamma_n\}(t(\tilde{W}_{in}^{\phi}) - t(\tilde{W}_{in}^{\phi}) - \tilde{\Delta}_{in}^{\phi}t'(W_{in}^{\phi}))\big|\mathbf{A}_n^{k_n}\big]\big]\big| \\
&=: \text{(a)} + \text{(b)} + \text{(c)}\;,
\end{aligned}$$

and we further have to bound the terms (a), (b), and (c). Since t is Lipschitz,

$$\begin{aligned}
\text{(a)} &\leq 2\textstyle\sum_i c_{i,2}\big(\tfrac{g_n}{\eta(n)}\big)\displaystyle\sum_{j\leq k_n'} c_{j,2}\big(\tfrac{g_n}{\eta(n)}\big) \\
&\qquad\textstyle\sum_{j\leq k_n'}\mathbb{E}[|\mathbf{A}_n|\mathbb{E}_{\mu_n^{\otimes 2}}[\mathbb{I}\{d(\phi_i,\phi_j)\leq b_n, \phi' \notin V_{i\beta_n}\}|\mathbf{A}_n^{2k_n}]] \\
&\leq 2C_2\big(\tfrac{g_n}{\eta(n)}\big)^2\sup_{i,j}\mathbb{E}[|\mathbf{A}_n|\mathbb{E}_{\mu_n^{\otimes 2}}\big[\mathbb{I}\{d(\phi_i,\phi_j)\leq b_n, \phi' \notin V_{i\beta_n}\}|\mathbf{A}_n^{2k_n}\big]\!\big].
\end{aligned}$$

Analogously, we have

$$\begin{aligned}
\text{(b)} \leq 2|\mathbf{B}_{b_n}|&\textstyle\sum_i c_{i,2}\big(\tfrac{g_n}{\eta(n)}\big)\displaystyle\sum_{j\leq k_n} cj,2\big(\tfrac{g_n}{\eta(n)}\big) \\
&\sup_{i,j}\tfrac{1}{|\mathbf{B}_{b_n}|}\mathbb{E}[|\mathbf{A}_n|\mathbb{E}_{\mu_n^{\otimes 2}}[\mathbb{I}\{d(\phi_i,\phi_j)\leq b_n, \phi' \notin V_{i\beta_n}\}|\mathbf{A}_n^{2k_n}]]\;.
\end{aligned}$$

To bound (c), we again have to control interactions betweens elements of $\mathbb{G}^{k_n}$. In addition to the element $\boldsymbol{\phi}$ in (c), fix two further elements $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$, and a list $\boldsymbol{\psi}^0, \ldots, \boldsymbol{\psi}^{b_n}$ constructed for $b = 0, \ldots, b_n$ as follows:

- Set $\boldsymbol{\psi}^0 = \boldsymbol{\phi}_2$,
- If either $\min\{\bar{d}(\boldsymbol{\psi}_k^{b-1}, \boldsymbol{\phi}), \bar{d}(\boldsymbol{\psi}_k^{b-1}, \boldsymbol{\phi}_1)\} \notin [b, b+1)$ or $k \notin \mathcal{I}_{b_n,k'_n}(\boldsymbol{\phi}_i, \boldsymbol{\phi}_2)$, choose $\boldsymbol{\psi}_k^b := \boldsymbol{\psi}_k^{b-1}$.
- Otherwise, choose $\boldsymbol{\psi}_k^b$ such that $\bar{d}(\boldsymbol{\psi}_k^b, \boldsymbol{\phi}) > b_n$ and $\bar{d}(\boldsymbol{\psi}_k^b, \boldsymbol{\phi}_1) > b_n$.

Such a list always exists. Abbreviate $G(\boldsymbol{\phi}) := h_n(\boldsymbol{\phi} X_n) - \bar{h}_n^{\boldsymbol{\phi}_i, b_n, k'_n}(\boldsymbol{\phi} X_n)$. An application of the triangle inequality yields

$$\begin{aligned}
&\big|\mathbb{E}\big[\tfrac{|\bar{h}_n^i(\boldsymbol{\phi} X_n)|}{\eta(n)^3}\mathbb{I}\{\tfrac{|\bar{h}_n^i(\boldsymbol{\phi} X_n)|}{c_{i,2}(g_n)} \le \gamma_n\}G(\boldsymbol{\phi}_1)G(\boldsymbol{\phi}_2)\big]\big| \\
&\le \textstyle\sum_l \Big(\mathbb{E}\big[\tfrac{|\bar{h}_n^i(\boldsymbol{\phi} X_n)|}{\eta(n)^3}\mathbb{I}\{\tfrac{|\bar{h}_n^i(\boldsymbol{\phi} X_n)|}{c_{i,2}(g_n)} \le \gamma_n\}G(\boldsymbol{\phi}_1)G(\boldsymbol{\psi}^l)\big] \\
&\qquad - \mathbb{E}\big[\tfrac{|\bar{h}_n^i(\boldsymbol{\phi} X_n)|}{\eta(n)^3}\mathbb{I}\{\tfrac{|\bar{h}_n^i(\boldsymbol{\phi} X_n)|}{c_{i,2}(g_n)} \le \gamma_n\}G(\boldsymbol{\phi}_1)G(\boldsymbol{\psi}^{l-1})\big]\Big) \\
&\le \Gamma_{i,q(1+\frac{\varepsilon}{2})}(\gamma_n)\textstyle\sum_{j,l} c_{l,2p(1+\frac{\varepsilon}{2})}\big(\tfrac{g_n}{\eta(n)}\big)c_{j,2p(1+\frac{\varepsilon}{2})}\big(\tfrac{g_n}{\eta(n)}\big) \\
&\quad \alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(\min\{\bar{d}(\boldsymbol{\phi}_{2,l}, \boldsymbol{\phi}), \bar{d}(\boldsymbol{\phi}_{2,l}, \boldsymbol{\phi}_1 s)\}|\mathbb{G})
\end{aligned}$$

where the sum in the final term runs over the index range $j \in \mathcal{I}_{b_n,k'_n}(\boldsymbol{\phi}_i, \boldsymbol{\phi}_1)$ and $l \in \mathcal{I}_{b_n,k'_n}(\boldsymbol{\phi}_i, \boldsymbol{\phi}_2)$. By Taylor expansion, we hence obtain

$$\begin{aligned}
\text{(c)} &\le \sqrt{\tfrac{2}{\pi}}\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\textstyle\sum_i \mathbb{E}_{\mu_n^*}\big[|\bar{h}_n^i(\boldsymbol{\phi} X_n)|(\tilde{\Delta}_{in}^{\boldsymbol{\phi}})^2 \,\big|\, \mathbf{A}_n^{k_n}\big]\big] \\
&\le \frac{16k'_n k_n|\mathbf{B}_{b_n}|\big(\sum_i c_{i,q(1+\frac{\varepsilon}{2})}(\frac{g_n}{\eta(n)})\big)^2\big(\sum_i \Gamma_{i,p(1+\frac{\varepsilon}{2})}(\gamma_n)\big)S_2^*(k'_n)\mathcal{R}_0}{\sqrt{|\mathbf{A}_n|}},
\end{aligned}$$

which establishes the result under hypothesis (3.11). If (3.10) holds instead, we follow the same proof outline, with the difference that there is some $K \in \mathbb{N}$ such that $b_n = K$ for all n, and that any two elements separated by a distance of at least K are conditionally independent. In this case,

$$\begin{aligned}
&\big|\mathbb{E}\big[\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi} X_n)(t(W^*) - t(W_{in}^{\boldsymbol{\phi}}) - \Delta_{in}^{\boldsymbol{\phi}}t'(W^*))|\mathbf{A}_n^{k_n}\big]\big]\big| \\
&\le \frac{4k'_n k_n|\mathbf{B}_K|^2}{\sqrt{|\mathbf{A}_n|}}\big(\textstyle\sum_i c_{i,2q}\big(\tfrac{g_n}{\eta(n)}\big)\big)^2\big(\sum_i \Gamma_{i,p(1+\frac{\varepsilon}{2})}(\gamma_n)\big)S_2^*(k'_n) \\
&\quad + 2\textstyle\sum_i c_{i,2}\big(\tfrac{g_n}{\eta(n)}\big)\displaystyle\sum_j c_{j,2}\big(\tfrac{g_n}{\eta(n)}\big)\mathbb{E}[|\mathbf{A}_n|\mathbb{E}_{\mu_n^{\otimes 2}}[\mathbb{I}\{\bar{d}(\boldsymbol{\phi}_i, \boldsymbol{\phi}'_{1:j}) \le K, \boldsymbol{\phi} \notin V_{i\beta_n}\}|\mathbf{A}_n^{2k_n}]] \\
&\quad + 2\mathcal{S}_w|\mathbf{B}_K|k'_n \textstyle\sum_i c_{i,2}\big(\tfrac{\bar{h}_n^i(\boldsymbol{\phi} X_n)}{\eta(n)}\mathbb{I}\{|\tfrac{\bar{h}_n^i(\boldsymbol{\phi} X_n)}{c_{i,2}(g_n)}| > \gamma_n\}\big)M_2\big(\tfrac{g_n}{\eta(n)}\big)\,,
\end{aligned}$$

and the result holds under (3.10). □

6.1.5 The third term in Lemma 6.6

LEMMA 6.10. *Fix $p, q > 0$ such that $\frac{1}{p} + \frac{1}{q} = 1$. If (3.10) holds,*

$$\Big\|1 - \tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}\big[\mathbb{E}_{\mu_n^*}\big[h_n(\boldsymbol{\phi}X_n)\Delta_{in}^{\boldsymbol{\phi}}\big|\mathbf{A}_n^{k_n}\big]\big|\mathbb{G}\big]\Big\|$$
$$\leq \mathbb{E}\big[\big|\tfrac{\eta(n)^2-\hat{\eta}_{n,K}^2}{\eta(n)^2}\big|\big] + K_1 C_2\big(\tfrac{g_n}{\eta(n)}\big)\textstyle\sum_{j>k_n'} c_{j,2}\big(\tfrac{g_n}{\eta(n)}\big) + \tfrac{K_2 k_n^4}{|\mathbf{A}_n|}C_2\big(\tfrac{g_n}{\eta(n)}\big)^2 ,$$

where $K_1 = O(\mathcal{S}_w|\mathbf{B}_K|)$ and $K_2 = O(\mathcal{S}_w|\mathbf{B}_K|^2)$. If (3.11) holds instead,

$$\Big\|1 - \tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}\big[\mathbb{E}_{\mu_n^*}\big[h_n(\boldsymbol{\phi}X_n)\Delta_{in}^{\boldsymbol{\phi}}\big|\mathbf{A}_n^{k_n}\big]\big|\mathbb{G}\big]\Big\|$$
$$\leq K_2|\mathbf{B}_{b_n}|C_2\big(\tfrac{g_n}{\eta(n)}\big)\textstyle\sum_{j>k_n'} c_{j,2}\big(\tfrac{g_n}{\eta(n)}\big) + K_1\mathcal{S}_w\dfrac{k_n^4|\mathbf{B}_{b_n}|^2}{|\mathbf{A}_n|}C_2(\dfrac{g_n}{\eta(n)})^2$$
$$+ K_3 C_{2+\varepsilon}\big(\tfrac{g_n}{\eta(n)}\big)^2\tfrac{k_n^2|\mathbf{B}_{b_n}|}{|\mathbf{A}_n|}\mathcal{S}_w\mathcal{R}_{b_n} + \mathbb{E}\big[\big|\tfrac{\eta(n)^2-\hat{\eta}_{n,b_n}^2}{\eta(n)^2}\big|\big] ,$$

where $K_1 = O(1)$ and $K_2 = O(\mathcal{S}_w)$ and $K_3 = O(1)$.

PROOF. Assume first that (3.11) holds. As above, we use the abbreviation $G^k(\boldsymbol{\phi}') := h_n(\boldsymbol{\phi}'X_n) - \bar{h}_n^{\boldsymbol{\phi}_i,b_n,k}(\boldsymbol{\phi}'X_n)$. By the triangle inequality,

$$\begin{aligned}
&\Big\|\frac{\eta(n)^2 - |\mathbf{A}_n|\sum_i \mathbb{E}_{\mu_n^{\otimes 2}}[\mathbb{E}[\bar{h}_n^i(\boldsymbol{\phi}X_n)G^{k_n'}(\boldsymbol{\phi}')|\mathbb{G}]|\mathbf{A}_n^{2k_n}]}{\eta(n)^2}\Big\| \\
&\leq \Big\|\frac{|\mathbf{A}_n|\sum_i \mathbb{E}_{\mu_n^{\otimes 2}}[\mathbb{E}[\bar{h}_n^i(\boldsymbol{\phi}X_n)(\bar{h}_n^{\boldsymbol{\phi}_i,b_n,k_n'}(\boldsymbol{\phi}'X_n) - \bar{h}_n^{\boldsymbol{\phi}_i,b_n,k_n}(\boldsymbol{\phi}'X_n))|\mathbb{G}]\big|\mathbf{A}_n^{2k_n}]}{\eta(n)^2}\Big\| \\
&+ \Big\|\frac{\hat{\eta}_{n,b_n}^2 - |\mathbf{A}_n|\sum_i \mathbb{E}_{\mu_n^{\otimes 2}}[\mathbb{E}[\bar{h}_n^i(\boldsymbol{\phi}X_n)G^{k_n}(\boldsymbol{\phi}')|\mathbb{G}]|\mathbf{A}_n^{2k_n}]}{\eta(n)^2}\Big\| + \mathbb{E}\big[\big|\tfrac{\eta(n)^2-\hat{\eta}_{n,b_n}^2}{\eta(n)^2}\big|\big] \\
&=: \text{(a)} + \text{(b)} + \text{(c)} .
\end{aligned} \tag{6.11}$$

We can further bound terms (a) and (b). By definition of the Lipschitz coefficients,

$$\begin{aligned}
\text{(a)} &\leq \sum_i \Big\|\frac{|\mathbf{A}_n|\mathbb{E}_{\mu_n^{\otimes 2}}[\sum_{j\in\mathcal{I}_{b_n,k_n}(\boldsymbol{\phi}_i,\boldsymbol{\phi}')\setminus\mathcal{I}_{b_n,k_n'}(\boldsymbol{\phi}_i,\boldsymbol{\phi}')} c_{i,2}\big(\tfrac{g_n}{\eta(n)}\big)c_{j,2}\big(\tfrac{g_n}{\eta(n)}\big)|\mathbf{A}_n^{2k_n}]}{\eta(n)^2}\Big\| \\
&\leq \mathcal{S}_w|\mathbf{B}_{b_n}|\textstyle\sum_i\sum_{j>k_n'} c_{i,2}\big(\tfrac{g_n}{\eta(n)}\big)c_{j,2}\big(\tfrac{g_n}{\eta(n)}\big) .
\end{aligned}$$

To bound (b), abbreviate $H(\boldsymbol{\phi}, \boldsymbol{\phi}') := \bar{h}_n^i(\boldsymbol{\phi}X_n)\big(h_n(\boldsymbol{\phi}'X_n) - \bar{h}_n^{\boldsymbol{\phi}_i,b_n,k_n}(\boldsymbol{\phi}'X_n)\big)$, and consider the index set

$$\mathcal{J}(\boldsymbol{\phi}, \boldsymbol{\phi}') := \{i, j | d(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j') \leq b_n\} . \tag{6.12}$$

Then let $\boldsymbol{\psi}, \boldsymbol{\psi}'$ be two elements of $\mathbb{G}^{k_n}$ such that, for the same index pair (i, j),

$$\boldsymbol{\psi}_i = \boldsymbol{\phi}_i \quad \text{and} \quad \boldsymbol{\psi}_j' = \boldsymbol{\phi}_j' . \tag{6.13}$$

Using a telescopic sum, we have

$$
\begin{aligned}
&\big\|\mathbb{E}\big[\tfrac{1}{\eta(n)^2}H(\boldsymbol{\phi},\boldsymbol{\phi}')\big|\mathbb{G}\big]-\mathbb{E}\big[\tfrac{1}{\eta(n)^2}H(\boldsymbol{\psi},\boldsymbol{\psi}')\big|\mathbb{G}\big]\big\|_1\\
&\leq \textstyle\sum_{l=0}^{k_n-1}\big\|\mathbb{E}\big[\tfrac{1}{\eta(n)^2}\big(H([\boldsymbol{\psi}_{1:l},\boldsymbol{\phi}_{l+1:k_n}],\boldsymbol{\phi}')-H([\boldsymbol{\psi}_{1:l+1},\boldsymbol{\phi}_{l+2:k_n}],\boldsymbol{\phi}')\big)\big|\mathbb{G}\big]\big\|_1\\
&+\textstyle\sum_{l=0}^{k_n-1}\big\|\mathbb{E}\big[\tfrac{1}{\eta(n)^2}\big(H(\boldsymbol{\psi},[\boldsymbol{\psi}'_{1:l},\boldsymbol{\phi}'_{l+1:k_n}])-H(\boldsymbol{\psi},[\boldsymbol{\psi}'_{1:l+1},\boldsymbol{\phi}'_{l+2:k_n}])\big)\big|\mathbb{G}\big]\big\|_1\\
&\leq 16\textstyle\sum_{l\neq i}c_{l,2+\varepsilon}\big(\tfrac{g_n}{\eta(n)}\big)c_{j,2+\varepsilon}\big(\tfrac{g_n}{\eta(n)}\big)\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}\big(\bar{d}([\boldsymbol{\psi}_l,\boldsymbol{\phi}_l],\ [\boldsymbol{\phi}',\boldsymbol{\phi}_{l+1:k_n},\boldsymbol{\psi}_{1:l-1}])\big|\mathbb{G}\big)\\
&+16\textstyle\sum_{l\neq j}c_{l,2+\varepsilon}\big(\tfrac{g_n}{\eta(n)}\big)c_{i,2+\varepsilon}\big(\tfrac{g_n}{\eta(n)}\big)\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}\big(\bar{d}([\boldsymbol{\psi}'_l,\boldsymbol{\phi}'_l],\ [\boldsymbol{\phi},\boldsymbol{\phi}'_{l+1:k_n},\boldsymbol{\psi}'_{1:l-1}])||\mathbb{G}\big)\;.
\end{aligned}
$$

By definition, $\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,\boldsymbol{\phi}_i)\widehat{\mathbb{F}}_{\infty,j}(h_n,X_n,\boldsymbol{\phi}'_j)$ is the average of $H(\boldsymbol{\psi},\boldsymbol{\psi}')$ over the set of pairs $(\boldsymbol{\psi},\boldsymbol{\psi}')$ satisfying (6.13).

Therefore for $(i,j)\in\mathcal{J}(\boldsymbol{\phi},\boldsymbol{\phi}')$,

$$
\begin{aligned}
&\Big\|\mathbb{E}\big[\tfrac{1}{\eta(n)^2}H(\boldsymbol{\phi},\boldsymbol{\phi}')\big|\mathbb{G}\big]-\mathbb{E}\big[\tfrac{1}{\eta(n)^2}\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,\boldsymbol{\phi}_i)\widehat{\mathbb{F}}_{\infty,j}(h_n,X_n,\boldsymbol{\phi}'_j)|\mathbb{G}\big]\Big\|_1\\
&\leq 16\textstyle\sum_{l\neq i}c_{l,2+\varepsilon}\big(\tfrac{g_n}{\eta(n)}\big)c_{j,2+\varepsilon}\big(\tfrac{g_n}{\eta(n)}\big)\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}\big(\bar{d}(\boldsymbol{\phi}_l,[\boldsymbol{\phi}',\boldsymbol{\phi}_{l+1:k_n}])\big|\mathbb{G}\big)\\
&+16\textstyle\sum_{l\neq j}c_{l,2+\varepsilon}\big(\tfrac{g_n}{\eta(n)}\big)c_{i,2+\varepsilon}\big(\tfrac{g_n}{\eta(n)}\big)\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}\big(\bar{d}(\boldsymbol{\phi}'_l,[\boldsymbol{\phi}'_{l+1:k_n},\boldsymbol{\phi}_i])|\mathbb{G}\big)\;.
\end{aligned}
$$

For all $i,j\leq k_n$, we hence obtain

$$
\begin{aligned}
&\Big\|\mathbb{E}_{\mu_n^{\otimes 2}}\big[\tfrac{\mathbb{I}\{\mathcal{J}(\boldsymbol{\phi},\boldsymbol{\phi}')=\{i,j\},\boldsymbol{\phi}'\in V_{i,\beta_n}\}(H(\boldsymbol{\phi},\boldsymbol{\phi}')-\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,\boldsymbol{\phi}_i)\widehat{\mathbb{F}}_{\infty,j}(h_n,X_n,\boldsymbol{\phi}'_j))}{\eta(n)^2}\big|\mathbf{A}_n^{2k_n}\big]\Big\|\\
&\leq 256\big(\textstyle\sum_l c_{l,2+\varepsilon}\big(\tfrac{g_n}{\eta(n)}\big)\big)^2\tfrac{k_n^2|\mathbf{B}_{b_n}|}{|\mathbf{A}_n|}\textstyle\sum_{m\geq b_n}|\mathbf{B}_{m+1}\backslash\mathbf{B}_m|\mathcal{S}_w\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(m|\mathbb{G})\;.
\end{aligned}
$$

We can then upper-bound (b) as

$$
\begin{aligned}
&\big\|\mathbb{E}_{\mu_n^{\otimes 2}}\big[\tfrac{\mathbb{I}\{\mathcal{J}(\boldsymbol{\phi},\boldsymbol{\phi}')=\{i,j\}\}(H(\boldsymbol{\phi},\boldsymbol{\phi}')-\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,\boldsymbol{\phi}_i)\widehat{\mathbb{F}}_{\infty,j}(h_n,X_n,\boldsymbol{\phi}'_j))}{\eta(n)^2}\big|\mathbf{A}_n^{2k_n}\big]\big\|\\
&+\big\|\mathbb{E}_{\mu_n^{\otimes 2}}\big[\tfrac{\mathbb{I}\{\mathcal{J}(\boldsymbol{\phi},\boldsymbol{\phi}')\subsetneq\{i,j\}\}(H(\boldsymbol{\phi},\boldsymbol{\phi}')-\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,\boldsymbol{\phi}_i)\widehat{\mathbb{F}}_{\infty,j}(h_n,X_n,\boldsymbol{\phi}'_j))}{\eta(n)^2}\big|\mathbf{A}_n^{2k_n}\big]\big\|\\
&=: b_{ij}^1+b_{ij}^2\ \geq\text{(b)}\;.
\end{aligned}
$$

We have already obtained a bound for b_{ij}^1 above.For b_{ij}^2

$$
\textstyle\sum_{ij}b_{ij}^2\leq 4\frac{\mathcal{S}_w|\mathbf{B}_{b_n}|^2k_n^4}{|\mathbf{A}_n|}M_2\big(\tfrac{g_n}{\eta(n)}\big)^2\;.
$$

Substituting the bounds for (a) and (b) so obtained back into (6.11) then completes the proof under hypothesis (3.11). If (3.10) holds instead, correlations between elements separated by a distance exceeding some constant K have no effect. In this case,

$$\begin{aligned}
&\big\|t'(W^*)\big(1-\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}\big[\mathbb{E}_{\mu_n^*}\big[h_n(\boldsymbol{\phi}X_n)\Delta_{in}^{\boldsymbol{\phi}}\big|\mathbf{A}_n^{k_n}\big]\big|\mathbb{G}\big]\big\|\\
&\leq\sqrt{\tfrac{2}{\pi}}\Big(\mathbb{E}\big[\big|\tfrac{\eta(n)^2-\hat{\eta}_{n,K}^2}{\eta(n)^2}\big|\big]+\mathcal{S}_w|\mathbf{B}_K|\textstyle\sum_i\sum_{k_n'<j\leq k_n}c_{i,2}\big(\tfrac{g_n}{\eta(n)}\big)c_{j,2}\big(\tfrac{g_n}{\eta(n)}\big)\\
&+4\tfrac{\mathcal{S}_w|\mathbf{B}_K|^2k_n^4}{|\mathbf{A}_n|}M_2\big(\tfrac{g_n}{\eta(n)}\big)^2\Big)\;,
\end{aligned}$$

which completes the proof. □

6.1.6 The fourth term in Lemma 6.6

The final term in Lemma 6.6 corresponds to a fourth moment, and we have to consider interactions between quadruples $\boldsymbol{\phi}_1,\ldots,\boldsymbol{\phi}_4$ of random elements of $\mathbb{G}^{k_n}$. Once again, n, b_n, β_n and k_n are fixed. For a quadruple of indices i,j,l,m, we are interested in whether the random elements satisfy

$$d(\boldsymbol{\phi}_{1,i},\boldsymbol{\phi}_{2,j})\leq b_n \qquad d(\boldsymbol{\phi}_{3,l},\boldsymbol{\phi}_{4,m})\leq b_n \tag{6.14}$$

$$\text{and}\quad \boldsymbol{\phi}_1\in V_{i,\beta_n}\quad \boldsymbol{\phi}_2\in V_{j,\beta_n}\quad \boldsymbol{\phi}_3\in V_{l,\beta_n}\quad \boldsymbol{\phi}_4\in V_{m,\beta_n}\;. \tag{6.15}$$

We then choose a constant S_4^* such that

$$\tfrac{|\mathbf{A}_n|^3}{|A||\mathbf{B}_{b_n}|^2}\big\|\mathbb{E}_{\mu_n^{\otimes 4}}\big[\mathbb{I}\{\boldsymbol{\phi}_1,\ldots,\boldsymbol{\phi}_4\models(6.14),(6.15)\text{ and }\boldsymbol{\phi}_{2,j}^{-1}\boldsymbol{\phi}_{3,m}\in A\}\big|\mathbf{A}_n^{4k_n}\big]\big\|\leq S_4^*$$

holds for every Borel set $A\subset\mathbb{G}^{k_n}$ with $|A|\geq 1$.

Lemma 6.11. *Fix $p,q>0$ with $\frac{1}{p}+\frac{1}{q}=1$. Assume (3.10) holds. Then*

$$\begin{aligned}
&\textstyle\sum_i\big\|\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi}X_n)\Delta_{in}^{\boldsymbol{\phi}}-\mathbb{E}[\bar{h}_n^i(\boldsymbol{\phi}X_n)\Delta_{in}^{\boldsymbol{\phi}}|\mathbb{G}]\big|\mathbf{A}_n^{k_n}\big]\big\|_1\\
&\leq K_1\tfrac{{k_n'}^2}{\sqrt{|\mathbf{A}_n|}}C^4_{4(1+\frac{\varepsilon}{2})}\big(\tfrac{g_n}{\eta(n)}\mathbb{I}\{|\tfrac{g_n}{\eta(n)}|\leq\gamma_n\}\big)\sqrt{S_4^*}+\tfrac{K_2k_n^4}{|\mathbf{A}_n|}C_2^2\big(\tfrac{g_n}{\eta(n)}\big)\\
&+K_3C_2\big(\tfrac{g_n}{\eta(n)}\big)^2|\mathbf{B}_k|S_0^*\;,
\end{aligned}$$

where $K_1=O(|\mathbf{B}_K|^{\frac{3}{2}})$ and $K_2=O(\mathcal{S}_w|\mathbf{B}_K|^2)$ and $K_3=O(\mathcal{S}_w)$. If (3.11) holds instead, then

$$\begin{aligned}
&\textstyle\sum_i\big\|\tfrac{\sqrt{|\mathbf{A}_n|}}{\eta(n)}\mathbb{E}_{\mu_n^*}\big[\bar{h}_n^i(\boldsymbol{\phi}X_n)\Delta_{in}^{\boldsymbol{\phi}}-\mathbb{E}[\bar{h}_n^i(\boldsymbol{\phi}X_n)\Delta_{in}^{\boldsymbol{\phi}}|\mathbb{G}]\big|\mathbf{A}_n^{k_n}\big]\big\|_1\\
&\leq K_1\Big(|\mathbf{B}_{b_n}|S_0^*C^2_{2+\varepsilon}\big(\tfrac{g_n}{\eta(n)}\big)+\tfrac{|\mathbf{B}_{b_n}|^2k_n^4}{|\mathbf{A}_n|}C_{2+\varepsilon}\big(\tfrac{g_n}{\eta(n)}\big)^2\Big)\\
&+K_2\mathcal{S}_w\frac{k_n^2|\mathbf{B}_{b_n}|\mathcal{R}_{b_n}C^2_{2+\varepsilon}\big(\frac{g_n}{\eta(n)}\big)}{|\mathbf{A}_n|}\\
&+K_3|\mathbf{B}_{b_n}|C_{2+\varepsilon}\big(\tfrac{g_n}{\eta(n)}\big)\textstyle\sum_i\big(\mathbb{E}[|\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)|^2\mathbb{I}\{|\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)|>\gamma_nc_{i,2}(g_n)\}]\big)^{\frac{1}{2}}
\end{aligned}$$

$$+ K_4 \frac{|\mathbf{B}_{b_n}| {k'_n}^2 C^2_{4(1+\frac{\varepsilon}{2})} \big(\frac{g_n}{\eta(n)} \mathbb{I}\{ |\frac{g_n}{\eta(n)}| \leq \gamma_n \} \big)}{\sqrt{|\mathbf{A}_n|}} \sqrt{S^*_4} \,,$$

for $K_1 = O(\mathcal{S}_w)$ *and* $K_2 = O(1)$ *and* $K_3 = O(\mathcal{S}_w)$ *and* $K_4 = O(\mathcal{R}_0^{\frac{1}{2}})$.

PROOF. First suppose (3.11) holds. As previously, we use the abbreviation:

$$H(\boldsymbol{\phi}, \boldsymbol{\phi}', i) = \bar{h}^i_n(\boldsymbol{\phi} X_n) \big(h_n(\boldsymbol{\phi}' X_n) - \bar{h}^{\boldsymbol{\phi}_i, b_n, k'_n}_n (\boldsymbol{\phi}' X_n) \big),$$

where we now additionally keep track of the index i. This term will now occur in its conditionally centered form,

$$\overline{H}(\boldsymbol{\phi}, \boldsymbol{\phi}', i) = H(\boldsymbol{\phi}, \boldsymbol{\phi}', i) - \mathbb{E}[H(\boldsymbol{\phi}, \boldsymbol{\phi}', i) | \mathbb{G}] \;.$$

We also must consider interactions between the random measures $\overline{\mathbb{F}}_{\infty,i}$ for different values of i, and hence terms of the form

$$\begin{aligned} F_{ij}(\phi, \phi', \tau) = \widehat{\mathbb{F}}_{\infty,i}(h_n, X_n, \phi) \mathbb{I}\{ \widehat{\mathbb{F}}_{\infty,i}(h_n, X_n, \phi) \leq \tau \} \\ \cdot \widehat{\mathbb{F}}_{\infty,j}(h_n, X_n, \phi') \mathbb{I}\{ \widehat{\mathbb{F}}_{\infty,j}(h_n, X_n, \phi') \leq \tau \} \end{aligned}$$

for any threshold $\tau \in (0, \infty]$. These terms again occur in centered form,

$$\overline{F}_{ij}(\phi, \phi', \tau) = F_{ij}(\phi, \phi', \tau) - \mathbb{E}[F_{ij}(\phi, \phi', \tau) | \mathbb{G}]$$

Using the triangle inequality, we obtain:

$$\begin{aligned}
&\textstyle\sum_i \big\| \frac{\sqrt{|\mathbf{A}_n|}}{\eta(n)} \mathbb{E}_{\mu^*_n} \big[\bar{h}^i_n(\boldsymbol{\phi} X_n) \Delta^{\boldsymbol{\phi}}_{in} - \mathbb{E}[\bar{h}^i_n(\boldsymbol{\phi} X_n) \Delta^{\boldsymbol{\phi}}_{in} | \mathbb{G}] \big| \mathbf{A}^{k_n}_n \big] \big\|_1 \\
&\leq \big\| \textstyle\sum_i \frac{|\mathbf{A}_n|}{\eta(n)^2} \mathbb{E}_{\mu^{\otimes 2}_n} \big[\mathbb{I}\{ (\boldsymbol{\phi}, \boldsymbol{\phi}') \in V_{j,\beta_n} \times V_{i,\beta_n} \} \overline{H}(\boldsymbol{\phi}, \boldsymbol{\phi}', i) \big| \mathbf{A}^{2k_n}_n \big] \big\| \\
&+ \big\| \textstyle\sum_i \frac{|\mathbf{A}_n|}{\eta(n)^2} \mathbb{E}_{\mu^{\otimes 2}_n} \big[\mathbb{I}\{ (\boldsymbol{\phi}, \boldsymbol{\phi}') \notin V_{j,\beta_n} \times V_{i,\beta_n} \} \overline{H}(\boldsymbol{\phi}, \boldsymbol{\phi}', i) \big| \mathbf{A}^{2k_n}_n \big] \big\| \\
&\leq \mathbb{E} \big[\textstyle\sum_i \frac{|\mathbf{A}_n|}{\eta(n)^2} \mathbb{E}_{\mu^{\otimes 2}_n} \big[\big\| \mathbb{E}\big[|(\overline{H}(\boldsymbol{\phi}, \boldsymbol{\phi}', i)| \big| \mathbb{G} \big] - \mathbb{E}\big[|\overline{F}_{ij}(\boldsymbol{\phi}_i, \boldsymbol{\phi}'_j, \infty)| \big| \mathbb{G} \big] \big\| \big| \mathbf{A}^{2k_n}_n \big] \big] \\
&+ |\mathbf{A}_n| \textstyle\sum_{i \leq k_n, j \leq k'_n} \\
&\qquad \big\| \mathbb{E}_{\mu^{\otimes 2}_n} \big[\mathbb{I}\{ (\boldsymbol{\phi}, \boldsymbol{\phi}') \in V_{j,\beta_n} \times V_{i,\beta_n}, d(\boldsymbol{\phi}_i, \boldsymbol{\phi}'_j) \leq b_n \} \frac{\overline{F}_{ij}(\boldsymbol{\phi}_i, \boldsymbol{\phi}'_j, \infty)}{\eta(n)^2} | \mathbf{A}^{2k_n}_n \big] \big\| \\
&+ \big\| \textstyle\sum_i \frac{|\mathbf{A}_n|}{\eta(n)^2} \mathbb{E}_{\mu^{\otimes 2}_n} \big[\mathbb{I}\{ (\boldsymbol{\phi}, \boldsymbol{\phi}') \notin V_{j,\beta_n} \times V_{i,\beta_n} \} \overline{H}(\boldsymbol{\phi}, \boldsymbol{\phi}', i) \big| \mathbf{A}^{2k_n}_n \big] \big\| \\
&=: \text{(a)} + \text{(b)} + \text{(c)} \;.
\end{aligned}$$

To bound (a), we proceed similarly as in the proof of Lemma 6.10. We again use the index set $\mathcal{J}(\boldsymbol{\phi}, \boldsymbol{\phi}')$ defined in (6.12). Then

$$\big\| \tfrac{1}{\eta(n)^2} \big(\mathbb{E}\big[|\overline{H}(\boldsymbol{\phi}, \boldsymbol{\phi}', i)| | \mathbb{G} \big] - \mathbb{E}\big[|\overline{F}_{ij}(\boldsymbol{\phi}_i, \boldsymbol{\phi}'_j, \infty)| | \mathbb{G} \big) \big\|_1$$

$$\le 64 \sum_{l\neq i} c_{l,2+\varepsilon}\big(\tfrac{g_n}{\eta(n)}\big) c_{j,2+\varepsilon}\big(\tfrac{g_n}{\eta(n)}\big) \alpha_n^{\frac{\varepsilon}{2+\varepsilon}}\big(d(\boldsymbol{\phi}_l,[\boldsymbol{\phi}',\boldsymbol{\phi}_{l+1:k_n}])\big|\mathbb{G}\big)$$
$$+ 64 \sum_{l\neq j} c_{l,2+\varepsilon}\big(\tfrac{g_n}{\eta(n)}\big) c_{i,2+\varepsilon}\big(\tfrac{g_n}{\eta(n)}\big) \alpha_n^{\frac{\varepsilon}{2+\varepsilon}}\big(d(\boldsymbol{\phi}'_l,[\boldsymbol{\phi}'_{l+1:k_n},\boldsymbol{\phi}_i])|\mathbb{G}\big)$$

The smaller constants, compared to Lemma 6.10, are due to the fact that the terms $\overline{H}(\boldsymbol{\phi},\boldsymbol{\phi}',i)$ have smaller Lipschitz coefficients than the similar terms $H(\boldsymbol{\phi},\boldsymbol{\phi}')$ (which involve k_n rather than k'_n, and are not centered). Therefore

$$\mathbb{E}\big[\sum_i \tfrac{|\mathbf{A}_n|}{\eta(n)^2}\mathbb{E}_{\mu_n^{\otimes 2}}[\mathbb{I}\{\mathcal{J}(\boldsymbol{\phi},\boldsymbol{\phi}')=\{i,j\}\}\big\|\mathbb{E}[\big|\overline{H}(\boldsymbol{\phi},\boldsymbol{\phi}',i)\big||\mathbb{G}]-\mathbb{E}[\big|\overline{F}_{ij}(\boldsymbol{\phi}_i,\boldsymbol{\phi}'_j,\infty)\big||\mathbb{G}]\big\||\mathbf{A}_n^{2k_n}]\big]$$
$$\le 64\mathcal{S}_w \tfrac{k_n^2|\mathbf{B}_{b_n}|}{|\mathbf{A}_n|}\big(\sum_l c_{l,2+\varepsilon}\big(\tfrac{g_n}{\eta(n)}\big)\big)^2 \sum_{i\ge b_n}|\mathbf{B}_{i+1}\setminus\mathbf{B}_i|\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(i|\mathbb{G})$$

and similarly

$$\mathbb{E}\big[\sum_i \tfrac{|\mathbf{A}_n|}{\eta(n)^2}\mathbb{E}_{\mu_n^{\otimes 2}}[\mathbb{I}\{\mathcal{J}(\boldsymbol{\phi},\boldsymbol{\phi}')\not\subset\{i,j\}\}\big\|\mathbb{E}[\big|\overline{H}(\boldsymbol{\phi},\boldsymbol{\phi}',i)\big||\mathbb{G}]-\mathbb{E}[\big|\overline{F}_{ij}(\boldsymbol{\phi}_i,\boldsymbol{\phi}'_j,\infty)\big||\mathbb{G}]\big\||\mathbf{A}_n^{2k_n}]\big]$$
$$\le \tfrac{8\mathcal{S}_w|\mathbf{B}_{b_n}|^2k_n^4}{|\mathbf{A}_n|}M_2\big(\tfrac{g_n}{\eta(n)}\big)^2\;,$$

we have a bound for term (a). To bound (b), we first note

$$\big\|\sum_{i,j}|\mathbf{A}_n|\mathbb{E}_{\mu_n^{\otimes 2}}\big[\mathbb{I}\{\boldsymbol{\phi}'\in V_{i,\beta_n},d(\boldsymbol{\phi}_i,\boldsymbol{\phi}'_j)\le b_n\}\tfrac{\overline{F}_{ij}(\boldsymbol{\phi}_i,\boldsymbol{\phi}'_j,\infty)-\overline{F}_{ij}(\boldsymbol{\phi}_i,\boldsymbol{\phi}'_j,\gamma_n)}{\eta(n)^2}\big|\mathbf{A}_n^{2k_n}\big]\big\|_1$$
$$\le 4\sum_{\min\{i,j\}\le k'_n}\Big(\mathbb{E}[|\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)|^2\mathbb{I}\{|\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)|>\gamma_n c_{i,2}(g_n)\}]\Big)^{\frac{1}{2}}$$
$$c_{j,2}\big(\tfrac{g_n}{\eta(n)}\big)\mathbb{E}[\mathbb{E}_{\mu_n^{\otimes 2}}\mathbb{I}\{\boldsymbol{\phi}_i^{-1}\boldsymbol{\phi}'_j\in\mathbf{B}_{b_n}\}|\mathbf{A}_n^{2k_n}]]$$
$$\le 8\mathcal{S}_w|\mathbf{B}_{b_n}|\big(\sum_j c_{j,2}\big(\tfrac{g_n}{\eta(n)}\big)\big)$$
$$\sum_i\big(\mathbb{E}[|\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)|^2\mathbb{I}\{|\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)|>\gamma_n c_{i,2}(g_n)\}]\big)^{\frac{1}{2}}$$

Write $\zeta_i := \|\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)\mathbb{I}\{|\widehat{\mathbb{F}}_{\infty,i}(h_n,X_n,e)|\le\gamma_n c_{i,2}(g_n)\}\|_{L_{4+2\varepsilon}}$, and upper-bound $\widehat{\mathbb{F}}$ as $\widehat{\mathbb{F}}_{\infty,i}^{\gamma_n} := \min\{\widehat{\mathbb{F}}_{\infty,i},\gamma_n\}$. Then for elements $\phi_1,\ldots,\phi_4\in\mathbb{G}$ and indices i,j,l,m, we have

$$\big\|\mathrm{Cov}[\widehat{\mathbb{F}}_{\infty,i}^{\gamma_n}(h_n,X_n,\phi_1)\widehat{\mathbb{F}}_{\infty,l}^{\gamma_n}(h_n,X_n,\phi_2),\;\widehat{\mathbb{F}}_{\infty,j}^{\gamma_n}(h_n,X_n,\phi_3)\widehat{\mathbb{F}}_{\infty,m}^{\gamma_n}(h_n,X_n,\phi_4)]\big\|$$
$$\le 4\,\zeta_i\,\zeta_j\,\zeta_l\,\zeta_m\,\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}\big(\bar{d}((\phi_1,\phi_2),(\phi_3,\phi_4))\big|\mathbb{G}\big)$$

By definition of S_4^*, we then have

$$\text{(b)} \le 8\tfrac{|\mathbf{B}_{b_n}|{k'_n}^2}{\sqrt{|\mathbf{A}_n|}}\Big(S_4^*\sum_i|\mathbf{B}_{i+1}\setminus\mathbf{B}_i|\alpha_n^{\frac{\varepsilon}{2+\varepsilon}}(i|\mathbb{G})\Big)^{\frac{1}{2}}\sum_{i\le k_n,j\le k'_n}\zeta_i\,\zeta_j\;.$$

The final term (c) is upper-bounded by

$$2\mathcal{S}_w\big(\sum_i c_{i,2}\big(\tfrac{g_n}{\eta(n)}\big)\big)^2\sup_{i,j}\mathbb{E}[\mathbb{E}_{\mu_n^{\otimes 2}}[|\mathbf{A}_n|\mathbb{I}\{\boldsymbol{\phi}'\notin V_i(\beta_n),d(\boldsymbol{\phi}_i,\boldsymbol{\phi}_{\boldsymbol{j}})\le b_n\}|\mathbf{A}_n^{2k_n}]]\;,$$

which concludes the proof under hypothesis (3.11). If (3.10) holds instead, there is again a constant distance K beyond which correlations vanish, and

$$\text{(a)} \leq \tfrac{8\mathcal{S}_w|\mathbf{B}_K|^2 k_n^4}{|\mathbf{A}_n|} M_2^2\big(\tfrac{g_n}{\eta(n)}\big) \qquad \text{(b)} \leq |\mathbf{B}_{b_n}|S_0^*\mathcal{S}_w\big(\textstyle\sum_i c_{i,2}\big(\tfrac{g_n}{\eta(n)}\big)\big)^2$$

$$\text{(c)} \leq 2\tfrac{|\mathbf{B}_K|^{\frac{3}{2}} {k'_n}^2}{\sqrt{|\mathbf{A}_n|}} \sqrt{S_4^*} \textstyle\sum_{i\leq k_n, j\leq k'_n} \zeta_i\zeta_j \ ,$$

which completes the proof of the lemma. □

6.1.7 Proof of the Berry-Esseen theorem

To proof Theorem 3.6, let μ_n^* be the random measure defiend in Equation (6.6). We consider the variable

$$W := \tfrac{1}{\eta(n)}\mathbb{E}_{\mu_n}[h_n(\boldsymbol{\phi}X_n)|\mathbf{A}_n^{k_n}] = \tfrac{1}{\eta(n)}\textstyle\sum_i \mathbb{E}_{\mu_n}[\bar{h}_n^i(\boldsymbol{\phi}X_n)|\mathbf{A}_n^{k_n}] \ ,$$

and similarly define W^* by substituting μ_n^* for μ_n, as in Lemma 6.6. If (b_n) is the increasing sequence chosen in the theorem, Lemma 6.5 shows

$$\big|d_{\mathrm{W}}(W, Z^*) - d_{\mathrm{W}}(W^*, Z^*)\big| \ \leq \ \frac{k_n^2 C_1(\frac{g_n}{\eta(n)})|\mathbf{B}_{b_n}|S_w}{\sqrt{|\mathbf{A}_n|}} \ .$$

(If hypothesis Equation (3.10) is assumed, we can in particular choose $b_n = K$ for all n and some K.) We can apply Lemma 6.6, where we choose $\eta(n) := \eta$ and $k'_n := k_n$ for all n. In Lemma 6.8–6.11, we can set $p = \frac{3}{2}$ and $q = \frac{1}{3}$. The constants S_2^*, S_3^*, S_4^* and the weak spreading coefficient $\mathcal{S}_w$ can then be bounded in terms of the (strong) spreading coefficients as

$$S_2^* \leq \mathcal{S} \qquad S_4^* \leq \mathcal{S} \qquad \mathcal{S}_w \leq \mathcal{S} \ ,$$

and substitute these into the bounds in Lemma 6.8–6.11. The sequences (β_n), which controls the moments of (μ_n), and (γ_n), which controls moments of $\frac{h_n}{\eta(n)}$, are relevant in the proof of the central limit theorem; for present purposes, we can set $\beta_n = \gamma_n = \infty$ for all n, and note that

$$\|\bar{h}_n^i(\boldsymbol{\phi}X_n)\mathbb{I}\{|\bar{h}_n^i(\boldsymbol{\phi}X_n)|\leq\gamma_n c_{i,2}\big(\tfrac{g_n}{\eta(n)}\big)\}\|_3 = \|\bar{h}_n^i(\boldsymbol{\phi}X_n)\|_3 \leq c_{i,3}\big(\tfrac{g_n}{\eta}\big)$$

and $\zeta_i \leq c_{4+2\epsilon,i}\big(\frac{g_n}{\eta}\big)$. Substituting all terms into Lemma 6.6 completes the proof.

6.1.8 Proof of the central limit theorem

To obtain Theorem 3.5 we want to prove that $d_W\big(\sqrt{|\mathbf{A}_n|}\hat{\mathbb{F}}_n(h_n, X_n), \eta Z\big) \to 0$ where $Z \sim N(0,1)$ is chosen to be an independent normal. We first note that

$$\|\widehat{\eta}^2_{m,n} - \eta^2_m\|_1 \xrightarrow{n\to\infty} 0 \qquad \text{for all } m \in \mathbb{N} \ . \tag{6.16}$$

That is the case since, for every $\varepsilon > 0$, we have

$$\begin{aligned}\mathbb{E}[|\widehat{\eta}^2_{m,n} - \eta^2_m|] &\leq \varepsilon + \mathbb{E}[\eta^2_m \mathbb{I}\{|\widehat{\eta}^2_{m,n} - \eta^2_m| > \varepsilon\}) + \mathbb{E}\big[\widehat{\eta}^2_{m,n}\mathbb{I}\{|\widehat{\eta}^2_{m,n} - \eta^2_m| > \varepsilon\}\big] \\ &\leq \varepsilon + \mathbb{E}[\eta^2_m \mathbb{I}\{|\widehat{\eta}^2_{m,n} - \eta^2_m| > \varepsilon\}] + |\mathbf{B}_m|\mathcal{S}_w\Big(\sum_i c_{i,2}(g_n \mathbb{I}\{|\widehat{\eta}^2_{m,n} - \eta^2_m| > \varepsilon\})\Big)^2 ,\end{aligned}$$

and (6.16) follows by uniform integrability of $(g_n(\boldsymbol{\phi}X_n)^2)_{\boldsymbol{\phi},n}$.

We next must define suitable sequences of coefficients γ_n, β_n, k_n, k'_n, and b_n as they appear in the bounds given by Lemma 6.5 and 6.6. These must be chosen to ensure the relevant terms in the bounds converge to 0 as $n \to \infty$. We first choose (γ_n) and (β_n) to satisfy $r^1_n := \beta_n\gamma^2_n k^2_n/\sqrt{|\mathbf{A}_n|} \to 0$. Such sequences exist, since $k^2_n/\sqrt{|\mathbf{A}_n|} \to 0$. We can then find sequences (k'_n) and (b_n) that satisfy $r^2_n := |\mathbf{B}_{b_n}|k'_n S^*_0 \to 0$ and

$$r^3_n := |\mathbf{B}_{b_n}|k'_n\Big(\textstyle\sum_i c_{i,2}\big(\bar{h}^i_n(\boldsymbol{\phi}X_n)\mathbb{I}\{\frac{|\bar{h}^i_n(\boldsymbol{\phi}X_n)|}{c_{i,2}(g_n)} > \gamma_n\}\big)\Big) \to 0 \ ,$$

which is possible $S^*_0 \to 0$ as $\beta_n \to \infty$, and

$$r^4_n := |\mathbf{B}_{b_n}|\Big(\textstyle\sum_{k'_n<i} c_{i,2+\varepsilon}(g_n)\Big) \to 0 \quad \text{and} \quad r^5_n := |\mathbf{B}_{b_n}|\frac{k^2_n\gamma^2_n}{\sqrt{|\mathbf{A}_n|}} \to 0 \ .$$

Consequently, we can choose sequences (δ_n) and (ε_n), with $\delta_n \to \infty$ and $\varepsilon_n \to \infty$ such that

$$\delta_n/\varepsilon^3_n \to 0 \quad \text{and} \quad \delta_n r^j_n/\varepsilon^3_n \xrightarrow{n\to\infty} 0 \quad \text{for } j = 1,\ldots,5 \ .$$

Because of (6.16), the sequence can additionally be chosen to satisfy

$$\frac{\|\widehat{\eta}^2_{m,n} - \eta^2_m\|_1\delta_n}{\varepsilon^2_n} \xrightarrow{n\to\infty} 0 \ .$$

Let η be the asymptotic variance, as in the hypothesis of the theorem. Given (ε_n) and (δ_n), we construct the sequence $(\eta(n))_n$ as in (6.4). The Lemma 6.3 then applies, and as in (6.5), we obtain

$$d_{\mathrm{W}}(S_n, \eta(n)Z) \leq \delta_n d_{\mathrm{W}}\big(\tfrac{S_n}{\eta(n)}, Z\big) \quad \text{for} \quad S_n := \sqrt{|\mathbf{A}_n|}\,\widehat{\mathbb{F}}_n(h_n, X_n) \ .$$

To apply Lemma 6.5 and Lemma 6.6, we note that

$$\sup_n \sum_i c_{i,2}\big(\bar{h}_n^i(\boldsymbol{\phi} X_n)\mathbb{I}\{|\tfrac{\bar{h}_n^i(\boldsymbol{\phi} X_n)}{c_{i,2}(g_n)}| > \gamma_n\}\big) \to 0 \qquad \text{as } \gamma_n \to \infty \;.$$

Recall that the constants S_0^*, S_2^*, etc by definition depend on the specific choice of the sequence (k_n') and (β_n). With both sequences given,

$$S_2^* \leq k_n'\beta_n\mathcal{S}_w \qquad S_4^* \leq {k_n'}^2\beta_n^2\mathcal{S}_w \qquad S_0^* \to 0 \;.$$

Moreover, we have $\sum_{i\leq k_n, j\leq k_n'} \zeta_i\,\zeta_j \leq \frac{\gamma_n}{\varepsilon_n^2}\sum_i c_{i,2+\varepsilon}(g_n)$ and

$$\sum_i \big\|\bar{h}_n^i(\boldsymbol{\phi} X_n)\mathbb{I}\{|\bar{h}_n^i(\boldsymbol{\phi} X_n)| \leq \gamma_n c_{i,2}\big(\tfrac{g_n}{\eta(n)}\big)\}\big\|_{L_\infty} \leq \gamma_n \sum_i c_{2,i}(g_n) \;.$$

Substituting into Lemma 6.5 and 6.6, we then obtain an upper bound on $d_{\mathrm{W}}(S_n/\eta, Z)$ and hence, as shown above, on $d_{\mathrm{W}}(S_n, Z)$ as claimed.

6.2 Other proofs

This section collects proofs remaining once the main limit theorems are established: The concentration inequality (Theorem 3.7), Propositions 3.13 and 3.15, and Theorem 5.1, the entropy central limit theorem.

Proof of Theorem 3.7. The proof strategy is to approximate by sums the integral $\mathbb{E}_{\mu_n}[h_n(\boldsymbol{\phi}\cdot X_n)|\mathbf{A}_n^{k_n}]$, and first establish concentration of each sum. To this end, let $\varepsilon_m := \frac{1}{m}$ for $m \in \mathbb{N}$, and let (C_m) and ε_m-grid of subsets of $\mathbb{G}$, as defined in Section 3.5. Let $\mathbf{B}_\varepsilon(\phi)$ denote the ball $\{\phi' \in \mathbb{G} | d(\phi, \phi') \leq \varepsilon\}$ of radius ε around ϕ. For each m, we can choose a partition $\mathcal{P}_m$ of $\mathbb{G}$ such that

$$\text{each } \phi \in C_m \text{ is in a separate block of } \mathcal{P}_m \qquad \text{and} \qquad \mathcal{P}_m(\phi) \subset \mathbf{B}_{\frac{1}{m}}(\phi) \;,$$

where $\mathcal{P}_m(\phi)$ is the block of the partition containing ϕ. Since $\mathcal{P}_m$ partitions $\mathbb{G}$, the product partition $\mathcal{P}_m^{k_n} := \mathcal{P}_m \times \ldots \times \mathcal{P}_m$ partitions $\mathbb{G}^{k_n}$, and we discretize the integral as

$$\Sigma_{nm} := \sum_{\boldsymbol{\phi}\in C_m^{k_n}} \mathbb{E}_{\mu_n}\big[\mathcal{P}_m^{k_n}(\boldsymbol{\phi})|\mathbf{A}_n^{k_n}\big] h_n(\boldsymbol{\phi} X_n) \;.$$

For each fixed $n \in \mathbb{N}$, the sum Σ_{nm} satisfies

$$\big\|\Sigma_{nm} - \mathbb{E}_{\mu_n^*}[h_n(\boldsymbol{\phi} X_n)\big|\mathbf{A}_n^{k_n}]\big\|_1 \leq \sup_{\substack{\boldsymbol{\phi}\in\mathbb{G}^{k_n} \\ \boldsymbol{\phi}'\in\mathbf{B}_{\varepsilon_m}(\phi)^{k_n}}} \|h_n(\boldsymbol{\phi} X_n) - h_n(\boldsymbol{\phi}' X_n)\|_1 \xrightarrow{m} 0.$$

By hypothesis, h_n is $\mathbf{L}_1$ uniformly contintuous in $\boldsymbol{\phi}$, hence

$$P\big(|\mathbb{E}_{\mu_n}[h_n(\boldsymbol{\phi}X_n)|\mathbf{A}_n^{k_n}]| \geq \lambda \,\big|\, \mu_n\big) \;\leq\; \limsup_m P\big(|\Sigma_{nm}| \geq \lambda \,\big|\, \mu_n\big)$$

for all $\lambda > 0$. By hypothesis, Σ_{nm} is self-bounded, with self-bounding constants given by $\sum_i c_i \mathbb{E}_{\mu_n}[\mathcal{P}(\frac{1}{m},\phi)|\mathbf{A}_n^{k_n}]$. We can hence use [**?**,]Theorem 4.3]Chatterjee:2005 to obtain

$$\begin{aligned} P(|\Sigma_{mn}| \geq \lambda|\mu_n) \quad &\leq \quad 2\mathbb{E}\Big(\exp\Big(-\frac{\big(1-\Lambda((X_\phi)_{\phi\in C_m})\big)\lambda^2}{\sum_{\phi\in C_m}(\sum_i c_i \mathbb{E}_{\mu_n}[\mathcal{P}_m(\phi)|\mathbf{A}_n^{k_n}])^2}\Big)\Big) \\ &\leq \quad 2\mathbb{E}\Big(\exp\Big(-|\mathbf{A}_n|\frac{(1-\Lambda((X_\phi)_{\phi\in C_m}))\lambda^2}{\tau_n|\mathbf{B}_{\frac{1}{m}}|\big(\sum_i c_i\big)^2}\Big)\Big)\;. \end{aligned}$$

where the second inequality is obtained using the definition of τ_n. Since the above holds for any m, we let $m \to \infty$, and use the definition of ρ_n to obtain

$$P(|\mathbb{E}_{\mu_n}(h_n(\boldsymbol{\phi}X_n)|\mathbf{A}_n^{k_n})| \geq \lambda|\mu_n) \leq 2\mathbb{E}\Big(\exp\big(-|\mathbf{A}_n|\tfrac{(1-\rho_n)\lambda^2}{[\sum_i c_i]^2\tau_n}\big)\Big)\;,$$

which completes the proof. □

Proof of Corollary 3.10. We want to use theorems Theorem 3.5 and Theorem 3.6. In this goal we approximate the quantity of interest by a generalized U-statistics. First, for all index $i \in \mathbb{N}$ we denote

$$d_i := \limsup_j \|f(X) - f(\tau_{ij}X)\|_{\mathbf{L}_2},$$

and

$$d_i(\eta) := \limsup_j \Big\|\frac{f(X) - f(\tau_{ij}X)}{\eta}\Big\|_{\mathbf{L}_2}.$$

Then we define $\mathbb{S}_m^{[i]} = \{\phi \in \mathbb{S}_m || \phi([i]) = [i]\}$ the set of permutations that leave the segment $[i]$ invariant.Then we define a surrogate of $f(\phi X)$ that depends only on the image $\phi[i]$ as

$$\bar{f}^i(\phi x) := \lim_{m\to\infty} \tfrac{1}{|\mathbb{S}_m^{[i]}|} \sum_{\mathbb{S}_m^{[i]}} f(\phi x)\;.$$

In addition we average out the kth coordinate of $\bar{f}^i$ and write it as:

$$\bar{f}^{i,k}(\phi x) := \lim_{m\to\infty} \frac{1}{m} \sum_{l\leq m} \bar{f}^i(\tau_{l,k}\phi x)\;,$$

We are now ready to state the proof. First, we want to prove that for any increasing sequence (k_n) that diverges $k_n \to \infty$ the following holds:

$$\frac{\sqrt{n}}{|\mathbb{S}_n|}\sum_{\mathbb{S}_n}\big[f(\phi X)-\bar{f}^{k_n}(\phi X)\big]\xrightarrow{L_1}0.$$

To prove this we notice that

$$\begin{aligned}&\Big\|\frac{\sqrt{n}}{|\mathbb{S}_n|}\sum_{\mathbb{S}_n}\big[f(\phi X)-\bar{f}^{k_n}(\phi X)\big]\Big\|_{\mathbb{L}_1}^2\\&\leq\frac{n}{|\mathbb{S}_n|^2}\sum_{k\geq k_n}\sum_{\phi_1,\phi_2\in\mathbb{S}_n}\mathbb{E}\Big[\big[\bar{f}^{k+1}(\phi_1 X)-\bar{f}^{k}(\phi_1 X)\big]\big[f(\phi_2 X)-\bar{f}^{k_n}(\phi_2 X)\big]\Big]\end{aligned}$$

However for all permutations $\phi_1,\phi_2\in\mathbb{S}_n$ and integers $k\leq m\in\mathbb{N}$ such that $\phi_2(m)=\phi_2(k)$ we have

$$\mathbb{E}\Big[\big[\bar{f}^{k+1}(\phi_1 X)-\bar{f}^{k}(\phi_1 X)\big]\bar{f}^{\infty,m}(\phi_2 X)\Big]=0,$$

and moreover

$$\mathbb{E}\Big[\big[\bar{f}^{k+1}(\phi_1 X)-\bar{f}^{k}(\phi_1 X)\big]\bar{f}^{k_n,m}(\phi_2 X)\Big]=0,$$

Therefore we obtain that

$$\begin{aligned}&\Big|\mathbb{E}\Big[\big[\bar{f}^{k+1}(\phi_1 X)-\bar{f}^{k}(\phi_1 X)\big]\big[f(\phi_2 X)-\bar{f}^{k_n}(\phi_2 X)\big]\Big]\Big|\\&=\Big|\mathbb{E}\Big[\big[\bar{f}^{k+1}(\phi_1 X)-\bar{f}^{k}(\phi_1 X)\big]\big[f(\phi_2,X)-\bar{f}^{\infty,m}(\phi_2 X)-\bar{f}^{k_n,m}(\phi_2 X)+\bar{f}^{k_n,m}(\phi_2 X)\big]\Big]\Big|\\&\leq\Big\|\bar{f}^{k+1}(\phi_1 X)-\bar{f}^{k}(\phi_1 X)\Big\|_{\mathbb{L}_2}\Big\|f(\phi_2,X)-\bar{f}^{\infty,m}(\phi_2 X)-\bar{f}^{k_n,m}(\phi_2 X)+\bar{f}^{k_n,m}(\phi_2 X)\Big\|_{\mathbb{L}_2}\\&\leq 2d_kd_m.\end{aligned}$$

This implies that,

$$\begin{aligned}&\Big\|\frac{\sqrt{n}}{|\mathbb{S}_n|}\sum_{\mathbb{S}_n}\big[f(\phi X)-\bar{f}^{k_n}(\phi X)\big]\Big\|_{\mathbb{L}_1}^2\\&\leq\frac{n}{|\mathbb{S}_n|^2}\sum_{k\geq k_n}\sum_{m\in\mathbb{N}}\sum_{\phi_1,\phi_2\in\mathbb{S}_n}\mathbb{I}(\phi_1(k)=\phi_2(m))d_kd_m\\&\leq[\sum_{k\geq k_n}d_k][\sum_{m\in\mathbb{N}}d_m]\to 0.\end{aligned}$$

Therefore given any choice of a sequence (k_n) that grows as $k_n=o(n^{\frac{1}{4}})$ it is enough to prove that $\frac{\sqrt{n}}{|\mathbb{S}_n|}\sum_{\mathbb{S}_n}\bar{f}^{k_n}(\phi X)$ is asymptotically gaussian. In this objective, for all integers $j_1,\ldots,j_k$ we will associate a permutation: $\phi_{j_1,\ldots,j_k}=\Pi_{l=1}^{k}\tau_{l,j_l}$ that is such that $\phi([k])=$

$(j_1, \ldots, j_k)$. Then we let $\mathbb{G}^{'}$ be the group of integers $\mathbb{Z}$, $(\mathbf{X}_n)$ be for all n the space of vectors $\mathbf{X}_n = \mathbb{R}^{k_n}$ and X_n to a random element of $\mathbf{X}_n$ defined as

$$X_n = \Big(\bar{f}^{k_n}(\phi_{z_1,\ldots,z_{k_n}} X)\Big)_{z_1,\ldots,z_{k_n}\in\mathbb{Z}}$$

We will take $\mathbb{Z}^{k_n}$ to act on $\mathbf{X}_n$ in the following way:

$$(z_1, \ldots, z_{k_n})\Big(x_{j_1,\ldots,j_{k_n}}\Big)_{j_1,\ldots,j_{k_n}\in\mathbb{Z}^{k_n}} \longrightarrow \Big(x_{z_1+j_1,\ldots,z_{k_n}+j_{k_n}}\Big)_{j_1,\ldots,j_{k_n}\in\mathbb{Z}^{k_n}}$$

We can note that X_n is invariant under joint action of $\mathbb{Z}$ and for all n the marginal mixing coefficients are such that

$$\alpha_k^n(t|\mathbb{G}) = 0, \ \forall t > 0.$$

Therefore we can use Theorem 3.5 to get the desired result.

To prove the Berry-Esseen bound, the reasoning is very similar. Given a choice $k \in \mathbb{N}$ we have:

$$d_W\Big(\frac{\sqrt{n}}{\eta|\mathbb{S}_n|}\sum_{\mathbb{S}_n} f(\phi X), \frac{\sqrt{n}}{\eta|\mathbb{S}_n|}\sum_{\mathbb{S}_n} \bar{f}^k(\phi X)\Big) \leq [\sum_{l\geq k} d_l(\eta)][\sum_{m\in\mathbb{N}} d_m(\eta)].$$

We denote $\eta^2(n) := \sum_{i,j\leq k} \operatorname{cov}\Big(\mathbb{F}^{\mathrm{i}}(\mathrm{X},\mathrm{e})\mathbb{F}^{\mathrm{j}}(\mathrm{X},\phi)\Big|\mathbb{G}\Big)$ and note that

$$\begin{aligned}\Big\|\frac{\eta^2(n) - \eta^2}{\eta^2}\Big\| &\leq \Big\|\frac{\sum_{l\geq k}\sum_{m\in\mathbb{N}} \operatorname{cov}\Big(\mathbb{F}^{\mathrm{l}}(\mathrm{X},\mathrm{e})\mathbb{F}^{\mathrm{m}}(\mathrm{X},\phi)\Big|\mathbb{G}\Big)}{\eta^2}\Big\| \\ &\leq \Big[\sum_{m\in\mathbb{N}} d_m(\eta)\Big]\sum_{l\geq k} d_l(\eta).\end{aligned}$$

Therefore by a simple application of Theorem 3.6 we get that

$$\begin{aligned}&d_W\Big(\frac{\sqrt{n}}{\eta|\mathbb{S}_n|}\sum_{\mathbb{S}_n} f(\phi X), Z\Big) \\ &\leq C\Big[\frac{k^2}{\sqrt{n}} + \sum_{l\geq k} d_l(\eta)\Big],\end{aligned}$$

for a specific real $C \in \mathbb{R}$. The desired result is proven. □

Proof of Proposition 3.13. Consider the random sets

$$\mathcal{V}_{lm} := \{(V_i, V_j) \in X_{\mathcal{W}} \cap [l, l+1] \times [m, m+1]\}\ .$$

Then $(|\mathcal{V}_{lm})|)_{l,m}$ is an exchangeable array. Since

$$\mathbb{E}[f_t(X_{\hat{w}(g_s(X_{\mathcal{W}}),s)})|g_s(X_{\mathcal{W}})] = \sum_{(i,j)\in g_s(X_{\mathcal{W}})} \mathbb{P}\big(\{(i,j)\in g_t(X_{\hat{w}(g_s(X_{\mathcal{W}}),s)})\}|g_s(X_{\mathcal{W}})\big)$$
$$= \frac{t^2}{s^2}|\{(i,j)\in g_s(X_{\mathcal{W}})\}| = \frac{t^2}{s^2}|\{(V_i,V_j)\in X_{\mathcal{W}}\cap[0,s]^2\}| = \frac{t^2}{s^2}\sum_{l,m\le s-1}|\mathcal{V}_{lm}| \; ,$$

the result follows from Theorem 3.5. □

Proof of Proposition 3.15. We first note that

$$\mathbb{E}[f(X)f(\phi X)|\mathbb{H}] = \lim_n \frac{1}{|\mathbf{A}_n\cap\mathbb{H}|}\int_{\mathbf{A}_n\cap\mathbb{H}} f(\theta\phi X)f(\theta X)d|\theta|$$

for all $\phi\in\mathbb{H}$, and hence $\mathbb{E}[f(X)f(\phi X)|\mathbb{H}] = \mathbb{E}[f(X)f(\phi X)|\mathbb{G}]$. Using Theorem 3.5, we obtain

$$\int_{\mathbf{A}_n\cap\mathbb{H}} \frac{f(\phi X)}{\sqrt{\mathbf{A}_n\cap\mathbb{H}}}|d\phi| \xrightarrow{d} \eta_H Z \quad \text{and} \quad \int_{\mathbf{A}_n}\frac{f(\phi X)}{\sqrt{\mathbf{A}_n}}|d\phi| \xrightarrow{d} \eta Z \; .$$

Since the random variables η and η_H satisfy

$$\begin{aligned}|\mathbb{K}|\eta_H^2 - \eta^2 &= |\mathbb{K}|\int_{\mathbb{H}}\mathbb{E}[f(X)f(\phi X)|\mathbb{H}]|d\phi| - \eta^2\\ &= \int_{\mathbb{H}}\int_{\mathbb{K}} f(X)[f(\phi X) - f(\phi\theta X)]|d\theta||d\phi| \; ,\end{aligned}$$

the result follows.

□

Sketch of the poof of Corollary 3.16. We can once use Theorem 3.5 that allows us to handle the randomness of the rotations $(\mathbf{\Theta}_j^Z)$. For this we need to prove the convergence of:

$$\frac{1}{m^2(2n)^d}\sum_{\boldsymbol{z}_1,\boldsymbol{z}_2\in[-n,n]^d}\sum_{i,j\le m}\big[\mathrm{cov}\big[\mathrm{f}(\mathbf{\Theta}_{\mathrm{i}}^{\boldsymbol{z_1}}\mathrm{X}+\boldsymbol{z_1}),\mathrm{f}(\mathbf{\Theta}_{\mathrm{j}}^{\boldsymbol{z_2}}\mathrm{X}+\boldsymbol{z_2})|(\mathbf{\Theta}_{\mathrm{i}}^{\boldsymbol{z}})_{\mathrm{z,i}}\big] \; ,$$

to its mean. This can be done with Markov inequality. □

Proof of Corollary 3.14. We can easily see that,

$$\begin{aligned}&\frac{\mathbb{E}_{\mu_\xi(\Pi_n)}\big[h(X,M)\big] - \mathbb{E}[\mathbb{E}_{\mu_\xi(\Pi_n)}[h(X,M)]|\mathbb{G}]}{\sqrt{|\mathbf{A}_n|}}\\ &= \frac{1}{|\mathbf{A}_n|}\int_{A_n}\big[f(\phi(\Pi)) - \mathbb{E}\big[f(\Pi)\big|\mathbb{G}\big]\big]d|\phi|\end{aligned}$$

Therefore as the conditions Corollary 3.14 implies the ones in Theorem 3.2, then the results stand.

In the literature, we usually impose conditions on $p(\cdot,\cdot,\cdot)$ that imply the hypothesis of Corollary 3.14. To give those, we first define random variables called radius of stabilization for all points $(x,m) \in \mathbf{X} \times \mathcal{M}$ as,

$$R(x,m,\Pi) := \inf_{r>0} \Big\{ r > 0 \Big| p((x,m),\Pi,\cdot) = p((x,m),\Pi \bigcap B(x,r),\cdot) \Big\}.$$

If there is no finite r such that $p((x,m),\Pi,\cdot) = p((x,m),\Pi \bigcap B(x,r),\cdot)$ then we will have $R(x,m,\Pi) = \infty$.

A standard condition is to ask that the tails of $\big(R(x,m,\Pi)\big)_{(x,m)\in\mathbf{X}\times\mathcal{M}}$ are decreasing polynomial fast. Indeed, we will say that p satisfies the polynomially stabilization condition with index $q > 1$ if

$$\sup_{s>0} s^q \sup_{(x,m)\in K\times\mathcal{M}} P\Big(R(x,m,\Pi) > s\Big) < \infty.$$

If the group $\mathbb{G}$ has a polynomial growth rate r, i.e $\sup_{i>0} i^{-r}\big|\mathbb{B}_i\big| < \infty$, and the function p is polynomially stabilizating with index $q > \frac{(2+2\epsilon)r}{\epsilon}$ and Π is a Poisson process then

$$\int_{\mathbb{G}} \alpha_2^{\frac{\epsilon}{2+\epsilon}}(d(e,g)|\mathbb{G})d|g| < \infty.$$

To prove this we will want to exploit the fact that events $A \in \sigma\big(\Pi \bigcap F\big)$ and $B \in \sigma\big(\Pi \bigcap G\big)$ are independent if $F \bigcap G = \emptyset$. To do so we will use the stabilization conditions. We define the functions $(f_s)_{s>0}$ to be such that for all sets $\mathcal{Q} \in \mathcal{F}$ we have

$$f_s(\mathcal{Q}) = f_s(\mathcal{Q} \bigcap B_s(0)).$$

Then we easily see that for all measurable sets A, subset $F \subset \mathbb{G}$ and all integers $s, n \in \mathbb{N}$ we have that

$$\begin{aligned}
&\Big|P\Big(\big(f(\phi(\Pi))\big)_{\phi\in F} \in A\Big) - P\Big(\big(f_s(\phi(\Pi))\big)_{\phi\in F} \in A\Big)\Big| \\
&\leq P\Big(\big(f(\phi(\Pi))\big)_{\phi\in F} \neq \big(f_s(\phi(\Pi))\big)_{\phi\in F}\Big) \\
&\leq P\Big(\big(f(\phi(\Pi))\big)_{\phi\in F} \neq \big(f_s(\phi(\Pi))\big)_{\phi\in F}, \mathrm{card}((\phi(\mathrm{K}))_{\phi\in\mathrm{F}} \textstyle\bigcap \Pi) \leq |\mathrm{F}|\mu(\mathrm{K})\mathrm{r}\Big) \\
&\quad + P\Big(\mathrm{card}((\phi(\mathrm{K}))_{\phi\in\mathrm{F}} \textstyle\bigcap \Pi) \geq |\mathrm{F}|\mu(\mathrm{K})\mathrm{r}\Big) \\
&\overset{(a)}{\leq} P\Big(\mathrm{card}((\phi(\mathrm{K}))_{\phi\in\mathrm{F}} \textstyle\bigcap \Pi) \geq |\mathrm{F}|\mu(\mathrm{K})\mathrm{r}\Big) + |\mathrm{F}|\mathrm{r}\mu(\mathrm{K}) \sup_{(\mathrm{x},\mathrm{m})\in\mathrm{K}} \mathrm{P}\Big(\mathrm{R}(\mathrm{x},\mathrm{m},\Pi) > \mathrm{s}\Big),
\end{aligned}$$

where to get (a) we exploited the fact that Π was a poisson process.

Therefore if $|F| \leq 2$, by choosing $r = s^{\frac{q}{2} - \frac{(2+2\epsilon)r}{2\epsilon}}$ we can see that there is a constant $C < \infty$ such that,

$$\Big|P\Big(\big(f(\phi(\Pi))\big)_{\phi\in F} \in A\Big) - P\Big(\big(f_s(\phi(\Pi))\big)_{\phi\in F} \in A\Big)\Big| \leq C s^{-\frac{(2+2\epsilon)r}{2\epsilon} - \frac{q}{2}}.$$

Similarly for all set $G \subset \mathbb{G}$ such that $\bar{d}(F,G) \geq s$ we have,

$$\begin{aligned}
&\Big|P\Big(\big(f(\phi(\Pi))\big)_{\phi\in G} \in A\Big) - P\Big(\big(f_{\bar{d}(F,\phi)-\frac{s}{2}}(\phi(\Pi))\big)_{\phi\in G} \in A\Big)\Big| \\
&\leq P\Big(\big(f(\phi(\Pi))\big)_{\phi\in G} \neq \big(f_{\bar{d}(F,\phi)-\frac{s}{2}}(\phi(\Pi))\big)_{\phi\in G}\Big) \\
&\overset{(a)}{\leq} \sum_{j\geq s} P\Big(\big(f(\phi(\Pi))\big)_{\phi\in G\bigcap\{\phi|\bar{d}(\phi,F)\in[j,j+1)\}} \neq \big(f_{\frac{2j-s}{2}}(\phi(\Pi))\big)_{\phi\in G\bigcap\{\phi|\bar{d}(\phi,F)\in[j,j+1)\}}\Big) \\
&\overset{(b)}{\leq} 2^{q+1}C\sum_{j\geq 0}[j+s]^{r-1-\frac{(2+2\epsilon)r}{2\epsilon}-\frac{q}{2}}
\end{aligned}$$

where (a) is a simple union bound and to get (b) we exploited the assumption on the growth rate of the group.

Therefore for all subsets $F, G \subset \mathbb{G}$ such that $|F| \leq 2$ and all measurable sets A, B we have that,

$$\begin{aligned}
&\Big|P\Big(\big(f(\phi(\Pi))\big)_{\phi\in F} \in A, \big(f(\phi(\Pi))\big)_{\phi\in G} \in B\Big) - P\Big(\big(f(\phi(\Pi))\big)_{\phi\in F} \in A\Big)P\Big(\big(f(\phi(\Pi))\big)_{\phi\in G} \in B\Big)\Big| \\
&\leq \Big|P\Big(\big(f_b(\phi(\Pi))\big)_{\phi\in F} \in A, \big(f(\phi(\Pi))\big)_{\phi\in G} \in B\Big) - P\Big(\big(f(\phi(\Pi))\big)_{\phi\in F} \in A\Big)P\Big(\big(f(\phi(\Pi))\big)_{\phi\in G} \in B\Big)\Big| \\
&\leq 4C\Big[\frac{\bar{d}(F,G)}{2}\Big]^{r-\frac{(2+2\epsilon)r}{2\epsilon}-\frac{q}{2}}.
\end{aligned}$$

This implies that $\alpha(s|\mathbb{G}) \leq 4C\big[\frac{s}{2}\big]^{-\frac{r}{\epsilon}-\frac{q}{2}}$. Hence as $\frac{r}{\epsilon} + \frac{q}{2} > \frac{(2+\epsilon)r}{\epsilon}$ we obtain the desired result. If the group grows faster than polynomial the proof can be adapted by requiring stronger stabilization propreties. □

Chapter 7

Proof for limit theorems for stable functions of exchangeable structures.

7.1 Useful lemmas

LEMMA 7.1. *Let $\mathbf{X}$ be a Borel space and let X be a random element of $\mathbf{X}$ invariant under $\mathbb{S}_\infty$. For all functions $g \in \mathbb{L}_1(X)$ and all $i \in \mathbb{N}$ the following holds:*

$$\mathbb{E}\big(g(X)\big|\mathbb{S}^{\setminus[i]}\big) - \mathbb{E}\big(g(X)\big|\mathbb{S}^{\setminus[i-1]}\big) = \mathbb{E}\Big(\Delta_i(g,X)\Big|\mathbb{S}^{\setminus[i]}\Big). \tag{7.1}$$

PROOF. This is a consequence of the tower property and Theorem 2.7. Indeed we have:

$$\begin{aligned}
&\mathbb{E}[g(X)|\mathbb{S}^{\setminus[i]}] - \mathbb{E}[g(X)|\mathbb{S}^{\setminus[i-1]}] \\
&= \mathbb{E}[g(X)|\mathbb{S}^{\setminus[i]}] - \mathbb{E}\big[\mathbb{E}(g(X)|\mathbb{S}^{\setminus[i]})\big|\mathbb{S}^{\setminus[i-1]}\big] \\
&= \mathbb{E}(g(X)|\mathbb{S}^{\setminus[i]}) - \lim_{l\to\infty} \frac{1}{|\mathbb{S}([l]\setminus[i-1])|} \sum_{\phi\in\mathbb{S}([l]\setminus[i-1])} \mathbb{E}(g(\phi X)|\mathbb{S}^{\setminus[i]}) \\
&\overset{(a)}{=} \lim_{l\to\infty} \frac{\big|\{\phi|\phi(i)\neq i\} \bigcap \mathbb{S}([l]\setminus[i-1])\big|}{|\mathbb{S}([l]\setminus[i-1])|} \mathbb{E}(\Delta_i(g,X)|\mathbb{S}^{\setminus[i]}) \\
&= \mathbb{E}(\Delta_i(g,X)|\mathbb{S}^{\setminus[i]}),
\end{aligned}$$

where (a) is a consequence of the following fact:

$$\mathbb{E}(g(\phi X)|\mathbb{S}^{\setminus[i]}) = \begin{cases} \mathbb{E}(g(X)|\mathbb{S}^{\setminus[i]}), \text{ if } \phi(\mathrm{i}) = \mathrm{i} \\ \mathbb{E}\big(\mathbb{A}_i(g,X)|\mathbb{S}^{\setminus[i]}\big) \text{ otherwise} \end{cases} \qquad \forall \phi \in \mathbb{S}(\mathbb{N}\setminus[i-1]).$$

□

Lemma 7.2. *Fix $k \in \mathbb{N}$. Let $\mathbf{X}$ be a Borel space and let X be a random element of $\mathbf{X}$ invariant under $\mathbb{S}_\infty$. For all functions $g \in \mathbb{L}_2(X)$ of order k we have:*

$$\Big\|g(X) - \mathbb{A}_i(g, X)\Big\|_{L_2} = \frac{1}{2} \lim_{l\to\infty} \Big\|g(X) - g(\tau_{i,l}X)\Big\|_{L_2}, \quad \forall i \in \mathbb{N}.$$

Proof. The idea of the proof is very similar to how classically the equality $\mathrm{Var(Y)} = \frac{1}{2}\mathbb{E}((\mathrm{Y} - \mathrm{Y}')^2)$ is proven. Indeed for all $l > k$ we have:

$$\begin{aligned}
&\Big\|g(X) - g(\tau_{i,l}X)\Big\|^2_{\mathbf{L}_2} \\
&= \Big\|g(X) - \mathbb{A}_i(g, X) + \mathbb{A}_i(g, X) - g(\tau_{i,l}X)\Big\|^2_{\mathbf{L}_2} \\
&= \Big\|g(X) - \mathbb{A}_i(g, X)\Big\|^2_{\mathbf{L}_2} + \Big\|\mathbb{A}_i(g, X) - g(\tau_{i,l}X)\Big\|^2_{\mathbf{L}_2} \\
&\quad + 2\mathbb{E}\Big(\Big[\mathbb{A}_i(g, X) - g(\tau_{i,l}X)\Big]\Big[g(X) - \mathbb{A}_i(g, X)\Big]\Big).
\end{aligned}$$

As g is invariant under $\mathbb{S}^{\setminus[k]}$ we get that $\mathbb{A}_i(g, X) = \mathbb{A}_i(g, \tau_{i,l}X)$, and we obtain:

$$\begin{aligned}
&\Big\|\mathbb{A}_i(g, X)) - g(\tau_{i,l}X)\Big\|^2_{L_2}, \\
&= \Big\|\mathbb{A}_i(g, \tau_{i,l}X) - g(X)\Big\|^2_{L_2} \\
&= \Big\|g(X) - \mathbb{A}_i(g, X)\Big\|^2_{L_2}
\end{aligned}$$

This implies that

$$\begin{aligned}
&\Big\|g(X) - g(\tau_{i,l}X)\Big\|^2_{L_2} \\
&= 2\Big\|g(X) - \mathbb{A}_i(g, X)\Big\|^2_{L_2} + 2\mathbb{E}\Big[\mathbb{A}_i(g, X) - g(\tau_{i,l}X)\Big]\Big[g(X) - \mathbb{A}_i(g, X)\Big].
\end{aligned}$$

As g is of order k we get that

$$g(\tau_{i,l}X) = \frac{1}{m-k} \sum_{j\in[|k+1,m|]} g(\tau_{l,j}\tau_{i,l}X), \quad \forall m > k,$$

which allows us to write:

$$\begin{aligned}
&\mathbb{E}\Big(\Big[\mathbb{A}_i(g,X)-g(\tau_{i,l}X)\Big]\Big[g(X)-\mathbb{A}_i(g,X)\Big]\Big)\\
&=\mathbb{E}\Big(\Big[\mathbb{A}_i(g,X)-\frac{1}{m-k}\sum_{j\in[|k+1,m|]}g(\tau_{l,j}\tau_{i,l}X)\Big]\Big[g(X)\mathbb{A}_i(g,X)\Big]\Big)\\
&=\lim_{m\to\infty}\mathbb{E}\Big(\Big[\mathbb{A}_i(g,X)-\frac{1}{m-k}\sum_{j\in[|k+1,m|]}g(\tau_{l,j}\tau_{i,l}X)\Big]\Big[g(X)-\mathbb{A}_i(g,X)\Big]\Big)\\
&=0
\end{aligned}$$

□

If $\mathbf{X}=\mathbb{R}^{\mathbb{N}}$ is the space of sequences and X is an exchangeable element of $\mathbf{X}$ we have an alternative expression for $\|\Delta_{i,j}(g,x)\|$. Using Theorem 2.8, we decompose the distribution p_X of X into a mixture of ergodic distributions $p_X(\cdot)=\int_{\mathbb{I}_{\mathbb{S}_\infty}}\nu(\cdot)dm(\nu)$; and we choose X' to be another process such that $(X,X')\sim\int_{\mathbb{I}_{\mathbb{S}_\infty}}\nu(\cdot)\otimes\nu(\cdot)dm(\nu)$. Note that such a coupling always exists.

For any pair of integers $i,j\in\mathbb{N}$ we define X^i and $X^{i,j}$ to be two processes that interpolate with X with in the following way,

- X and X^i agree everywhere except on the ith coordinate,

$$X^i_l:=\begin{cases}X_l \text{ if } \mathrm{l}\neq \mathrm{i}\\ X'_i \text{ if } \mathrm{l}=\mathrm{i}\end{cases}$$

- $X^{i,j}$ and X^i agree everywhere exept on the jth coordinate

$$X^{i,j}_l:=\begin{cases}X^i_l \text{ if } \mathrm{l}\neq \mathrm{j}\\ X'_j \text{ if } \mathrm{l}=\mathrm{j}\end{cases}$$

With those notations in hand we can introduce the following lemma:

LEMMA 7.3. *Fix $k\in\mathbb{N}$. Let $\mathbf{X}$ be a Borel space and let X be a random element of $\mathbf{X}$ invariant under $\mathbb{S}_\infty$. For all functions $g\in\mathbb{L}_2(X)$ of order k we have:*

$$\|\Delta_{i,j}(g,X)\|_{L_2}=\frac{1}{2}\Big\|g(X)-g(X^i)-\big[g(X^j)-g(X^{i,j})\big]\Big\|_{L_2}$$

PROOF. The idea of the proof is very similar to the classical proof of the following equality: $\mathrm{Var}(\mathrm{Y})=\frac{1}{2}\mathbb{E}[(\mathrm{Y}-\mathrm{Y}')^2]$.

We abbreviate $\Delta_i(X) := g(X) - g(X^i)$ and $\Delta_i(X^j) = g(X^j) - g(X^{i,j})$; and then write:

$$\begin{aligned}
&\Big\|g(X) - g(X^i) - \big[g(X^j) - g(X^{i,j})\big]\Big\|_{L_2}^2 \\
&= \mathbb{E}\Big(\Big[\Delta_i(X) - \mathbb{E}\big(\Delta_i(X)|\mathbb{S}_k(j)\big)\Big]^2\Big) + \mathbb{E}\Big(\Big[\Delta_i(X^j) - \mathbb{E}\big(\Delta_i(X^j)|\mathbb{S}_k(j)\big)\Big]^2\Big) \\
&\quad + 2\mathbb{E}\Big(\Big[\Delta_i(X^j) - \mathbb{E}\big(\Delta_i(X^j)|\mathbb{S}_k(j)\big)\Big]\Big[\Delta_i(X) - \mathbb{E}\big(\Delta_i(X)|\mathbb{S}_k(j)\big)\Big]\Big).
\end{aligned}$$

The next step is to note that $X_j \perp\!\!\!\perp_{\mathbb{S}_\infty} X_j'$, which implies that

$$\begin{aligned}
&\mathbb{E}\Big(\Big[\Delta_i(X^j) - \mathbb{E}\big(\Delta_i(X^j)|\mathbb{S}_k(j)\big)\Big]\Big[\Delta_i(X) - \mathbb{E}\big(\Delta_i(X)|\mathbb{S}_k(j)\big)\Big]\Big) \\
&= \mathbb{E}\Big[\mathbb{E}\Big(\big[\Delta_i(X^j) - \mathbb{E}(\Delta_i(X^j)|\mathbb{S}_k(j))\big]\big[\Delta_i(X) - \mathbb{E}(\Delta_i(X)|\mathbb{S}_k(j))\big]\Big|\mathbb{S}_k(j)\Big)\Big] \\
&= \mathbb{E}\Big(\mathbb{E}\Big[\Delta_i(X^j) - \mathbb{E}(\Delta_i(X^j)|\mathbb{S}_k(j))\Big|\mathbb{S}_k(j)\Big]\ \mathbb{E}\Big[\Delta_i(X) - \mathbb{E}\big(\Delta_i(X)|\mathbb{S}_k(j)\big)\Big|\mathbb{S}_k(j)\big)\Big]\Big) = 0,
\end{aligned} \tag{7.2}$$

Therefore Equation (7.2) implies that

$$\begin{aligned}
&\Big\|g(X) - g(X^i) - \big[g(X^j) - g(X^{i,j})\big]\Big\|_{L_2}^2 \\
&= \mathbb{E}\Big(\Big[\Delta_i(X) - \mathbb{E}(\Delta_i(X)|\mathbb{S}_k(j))\Big]^2\Big) + \mathbb{E}\Big(\Big[\Delta_i(X^j) - \mathbb{E}(\Delta_i(X^j)|\mathbb{S}_k(j))\Big]^2\Big),
\end{aligned}$$

and as $X \overset{d}{=} X^j$ we have

$$\Big\|g(X) - g(X^i) - \big[g(X^j) - g(X^{i,j})\big]\Big\|_{L_2}^2 = 2\mathbb{E}\Big(\Big[\Delta_i(X) - \mathbb{E}(\Delta_i(X)|\mathbb{S}_k(j))\Big]^2\Big).$$

The next step will be to separate $\mathbb{E}\Big(\Big[\Delta_i(X) - \mathbb{E}(\Delta_i(X)|\mathbb{S}_k(j))\Big]^2\Big)$ in three terms by introducing $\mathbb{A}_i(g,X) - \mathbb{A}_j(\mathbb{A}_i(g,X))$. Using the fact that

$$\mathbb{A}_j(g,X) = \mathbb{E}(g(X)|\mathbb{S}_k(j)), \text{ and } \mathrm{\mathbb{A}_j(g, X^i) = \mathbb{E}(g(X^i)|\mathbb{S}_k(j))}.$$

We write:

$$\begin{aligned}
&\mathbb{E}\Big(\Big[\Delta_i(X) - \mathbb{E}(\Delta_i(X)|\mathbb{S}_k(j))\Big]^2\Big) \\
&= \mathbb{E}\Big(\Big[g(X) - \mathbb{A}_i(g,X) - \mathbb{A}_j(g,X) + \mathbb{A}_j(\mathbb{A}_i(g,X))\Big]^2\Big) \\
&+ \mathbb{E}\Big(\Big[g(X^i) - \mathbb{A}_i(g,x) - \mathbb{A}_j(g,X^i) + \mathbb{A}_j(\mathbb{A}_i(g,X))\Big]^2\Big) \\
&+ 2\mathbb{E}\Big[(\Big[g(X^i) - \mathbb{A}_i(g,x) - \mathbb{A}_j(g,X^i) + \mathbb{A}_j(\mathbb{A}_i(g,X))\Big]\Big[g(X) - \mathbb{A}_i(g,X) - \mathbb{A}_j(g,X) + \mathbb{A}_j(\mathbb{A}_i(g,X))\Big]\Big].
\end{aligned}$$

But similarly as for Equation (7.2) we can prove that

$$\mathbb{E}\Big[\big(\Big[g(X^i)-\mathbb{A}_i(g,x)-\mathbb{A}_j(g,X^i)+\mathbb{A}_j(\mathbb{A}_i(g,X))\Big]\Big[g(X)-\mathbb{A}_i(g,X)-\mathbb{A}_j(g,X)+\mathbb{A}_j(\mathbb{A}_i(g,X))\Big]\Big]=0 \tag{7.3}$$

and also that

$$\begin{aligned}&\mathbb{E}\Big(\Big[g(X)-\mathbb{A}_i(g,X)-\mathbb{A}_j(g,X)+\mathbb{A}_j(\mathbb{A}_i(g,X))\Big]^2\Big)\\&=\mathbb{E}\Big(\Big[g(X^i)-\mathbb{A}_i(g,x)-\mathbb{A}_j(g,X^i)+\mathbb{A}_j(\mathbb{A}_i(g,X))\Big]^2\Big).\end{aligned}$$

Therefore we get that,

$$\Big\|g(X)-g(X^i)-\big[g(X^j)-g(X^{i,j})\big]\Big\|_{L_2}^2=4\|\Delta_{i,j}(g(X))\|_{L_2}^2.$$

□

7.2 Proof of the main theorem

Proof. This proof uses Theorem 4.9, and to do so we re-express $f_n(X_n)$ in the following way:

$$f_n(X_n)-\mathbb{E}\big(f_n(X_n)\big|\mathbb{S}_\infty\big)=\sum_{l\leq n}X_{i,n},$$

where $(X_{i,n})$ is a triangular array of martingale differences:

$$X_{i,n}:=\mathbb{E}(f_n(X_n)|\mathbb{S}^{\setminus[i]})-\mathbb{E}(f_n(X_n)|\mathbb{S}^{\setminus[i-1]})\quad\forall i,n\in\mathbb{N}.$$

As $(\mathbb{S}^{\setminus[i]})_i$ is a filtration, $(\sum_{i=0}^k X_{i,n},\mathbb{S}^{\setminus[k]})_{k,n\in\mathbb{N}}$ is a triangular array of martingales.Therefore to prove Theorem 4.3 it is enough to prove the following two points:

1. For all $\epsilon>0$, $\frac{1}{n}\sum_{i=1}^n\mathbb{E}(X_{i,n}^2\mathbb{I}(|X_{n,i}|>\sqrt{n}\epsilon)|\mathbb{S}^{\setminus[i-1]})\xrightarrow{P}0$.

2. $\frac{1}{n}\sum_{i=1}^n X_{i,n}^2\xrightarrow{P}\sigma^2$.

- Proof of the first point: To prove this we will first note that $(X_{i,n}^2)_{i,n}$ is uniformly-integrable (uniformly integrable). Indeed let $M>0$ be a positive real, then using Lemma 7.1 we have

$$\begin{aligned}\mathbb{E}\Big(X_{i,n}^2\mathbb{I}(|X_{i,n}|\geq M)\Big)&\leq\mathbb{E}\Big(\mathbb{E}\big(\Delta_i(f_n,X_n)\big|\mathcal{C}_{i+1}\big)^2\mathbb{I}(|X_{i,n}|\geq M)\Big)\\&\leq\Big\|\Delta_i(f_n,X_n)\mathbb{I}(|X_{i,n}|\geq M)\Big\|_{L_2}^2,\end{aligned}$$

And therefore as $(\Delta_i^2(f_n, X_n))_{i,n\in\mathbb{N}}$ is uniformly integerable we have that for all positive $\epsilon > 0$,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_{i,n}^2\mathbb{I}(|X_{i,n}| > \epsilon\sqrt{n})) \leq \sup_{i\in\mathbb{N}}\mathbb{E}(X_{i,n}^2\mathbb{I}(|X_{i,n}| > \epsilon\sqrt{n})) \to 0.$$

- Proof of the second point. The second point consists of proving a law of large number for the empirical variance. By hypothesis we know that the following converges $\frac{\mathrm{var}(\mathrm{f_n}(\mathrm{X_n})|\mathbb{S}_\infty)}{n} \xrightarrow{P} \sigma^2$; therefore it is enough to prove that

$$\frac{1}{n}\sum_{i\in\mathbb{N}} X_{i,n}^2 - \frac{\mathrm{var}(\mathrm{f_n}(\mathrm{X_n})|\mathbb{S}_\infty)}{n} \xrightarrow{P} 0.$$

It is useful to note that as $(X_{i,n})$ is a sequence of martingale differences we have

$$\mathrm{var}(\mathrm{f_n}(\mathrm{X_n})|\mathbb{S}_\infty) = \sum_{\mathrm{i}\leq\mathrm{n}}\mathbb{E}\Big(\Delta_{\mathrm{i}}^2(\mathrm{f_n},\mathrm{X_n})\Big|\mathbb{S}_\infty\Big).$$

Therefore to prove the desired result we need to prove that $\frac{1}{n}\sum_{i\leq n}X_{i,n}^2$ is close to its conditional expectation. To do this we will exploit hypothesis $(H2)$ through a variance argument. As we made no assumptions that f_n was in $\mathbb{L}_4$ we first need to bound each $X_{i,n}$, before passing at a higher moment. For this we introduce the following function $g : \mathbb{N}\times\mathbb{R}\to\mathbb{R}$ such that,

$$g(M,x) := \begin{cases} x, \text{ if } |\mathrm{x}| \leq \mathrm{M} \\ M(M+1) - Mx \text{ if } \mathrm{x} \in (\mathrm{M}, \mathrm{M}+1] \\ Mx - M(M+1) \text{ if } \mathrm{x} \in [-(\mathrm{M}+1), -\mathrm{M}) \\ 0 \text{ otherwise} \end{cases} \qquad \forall M\in\mathbb{N}, x\in\mathbb{R}.$$

We denote for all integers $i, n \in \mathbb{N}$:

$$X_{i,n}(M) := g(M, X_{i,n}), \quad \forall M > 0;$$

And we introduce the sequence (β_n) chosen such that:

- $\beta_n \to \infty$
- $\frac{\beta_n^3}{n^2}\sum_{j\in\mathbb{N}}\Big[\sum_{i\in\mathbb{N}}\|\Delta_{i,j}(f_n, X_n)\|_{L_2}\Big]^2 \to 0.$

This sequence will be our proposed upper bound for $X_{i,n}$. Then using a triangular inequality we have:

$$\begin{aligned}
&\Big\|\frac{1}{n}\sum_{i\le n}[X_{i,n}^2-\mathbb{E}(X_{i,n}^2|\mathcal{S}_\infty)\Big\|_{L_1}\\
&\le \Big\|\frac{1}{n}\sum_{i\le n}[X_{i,n}^2-X_{i,n}^2(\beta_n)-\mathbb{E}(X_{i,n}^2|\mathcal{S}_\infty)+\mathbb{E}(X_{i,n}^2(\beta_n)|\mathcal{S}_\infty)\Big\|_{L_1}\\
&+\Big\|\frac{1}{n}\sum_{i\le n}[X_{i,n}^2(\beta_n)-\mathbb{E}(X_{i,n}^2(\beta_n)|\mathcal{S}_\infty)\Big\|_{L_1} \qquad (7.4)\\
&\overset{(a)}{\le} \Big\|\frac{2}{n}\sum_{i\le n}\mathbb{I}(|X_{i,n}|>\beta_n)\Big[\Big|X_{i,n}^2\Big|+\Big|\mathbb{E}(X_{i,n}^2|\mathbb{S}_\infty)\Big|\Big]\Big\|_{L_1}\\
&+\Big\|\frac{1}{n}\sum_{i\le n}\big[X_{i,n}^2(\beta_n)-\mathbb{E}(X_{i,n}^2(\beta_n)|\mathbb{S}_\infty)\big]\Big\|_{L_1},
\end{aligned}$$

where to get (a) we used the fact that the following inequality is true,

$$|g(M,x)|\le|x|,\quad \forall M\in\mathbb{N},\forall x\in\mathbb{R}.$$

The rest of the proof will then consist on bounding sucesscively the first and second term of the right hand side of Equation (7.4). For the first term we will exploit the uniform integrability of $(X_{i,n})_{i,n\in\mathbb{N}}$. Indeed this implies that $\sup_{i,n}\|X_{i,n}^2\mathbb{I}(|X_{i,n}|>\beta_n)\|\to 0$ converges to zero. Therefore using Jensen inequality we can see that:

$$\begin{aligned}
&\Big\|\frac{1}{n}\sum_{i\in\mathbb{N}}\mathbb{I}(|X_{i,n}|>\beta_n)\big[\Big|X_{i,n}^2\Big|+\Big|\mathbb{E}(X_{i,n}^2|\mathbb{S}_\infty)\Big|\big]\Big\|_{L_1}\\
&\le 2\sup_{i,n}\|X_{i,n}^2\mathbb{I}(|X_{i,n}|>\beta_n)\|\to 0.
\end{aligned}$$

After bounding the first term we now want to bound the second term of Equation (7.4). In this goal we introduce

$$Y_{i,j}(n):=\mathbb{E}\Big(X_{i,n}^2(\beta_n)|\mathbb{S}^{\setminus[j]}\Big)-\mathbb{E}\Big(X_{i,n}^2(\beta_n)|\mathbb{S}^{\setminus[j-1]}\Big),\quad \forall i,j,n\in\mathbb{N}$$

Then note that

$$\begin{aligned}
&\mathrm{var}\Big(\frac{1}{\mathrm{n}}\sum_{\mathrm{i}\le\mathrm{n}}\mathrm{X}^2_{\mathrm{i,n}}(\beta_\mathrm{n})\Big)\\
&=\mathrm{var}\Big(\sum_{\mathrm{j}\in\mathbb{N}}\mathbb{E}\Big(\frac{1}{\mathrm{n}}\sum_{\mathrm{i}\le\mathrm{n}}\mathrm{X}^2_{\mathrm{i,n}}(\beta_\mathrm{n})|\mathbb{S}^{\setminus[\mathrm{j}]}\Big)-\mathbb{E}\Big(\frac{1}{\mathrm{n}}\sum_{\mathrm{i}\le\mathrm{n}}\mathrm{X}^2_{\mathrm{i,n}}(\beta_\mathrm{n})|\mathbb{S}^{\setminus[\mathrm{j}-1]}\Big)\Big)\\
&=\sum_{j\in\mathbb{N}}\mathbb{E}\Big(\Big[\mathbb{E}\Big(\frac{1}{n}\sum_{i\le n}X^2_{i,n}(\beta_n)|\mathbb{S}^{\setminus[j]}\Big)-\mathbb{E}\Big(\frac{1}{n}\sum_{i\le n}X^2_{i,n}(\beta_n)|\mathbb{S}^{\setminus[j-1]}\Big)\Big]^2\Big)\\
&=\frac{1}{n^2}\sum_{j\in\mathbb{N}}\sum_{i,k\in\mathbb{N}}\mathbb{E}\Big[Y_{i,j}(n)Y_{k,j}(n)\Big]\\
&\le\frac{1}{n^2}\sum_{j\in\mathbb{N}}\sum_{i,k\in\mathbb{N}}\Big\|Y_{i,j}(n)\Big\|_{L_2}\Big\|Y_{k,j}(n)\Big\|_{L_2}\\
&\le\frac{1}{n^2}\sum_{j\in\mathbb{N}}\Big[\sum_{i\in\mathbb{N}}\Big\|Y_{i,j}(n)\Big\|_{L_2}\Big]^2
\end{aligned}$$

The remaining of the proof will then consist on proposing a good bound to the norms $\Big\|Y_{i,j}(n)\Big\|._{L_2}$ To do this we abbreviate

$$X_{i,n}(\phi):=\mathbb{E}\big(f_n(\phi X_n)\big|\mathbb{S}^{\setminus[i]}\big)-\mathbb{E}\big(f_n(\phi X_n)\big|\mathbb{S}^{\setminus[i-1]}\big);$$

and

$$X_{i,n}(\beta_n,\phi):=g(\beta_n,X_{i,n}(\phi)).$$

Then for all integers $i,j\le n$ thanks to Lemma 7.1 we have,

$$\begin{aligned}
&\Big\|Y_{i,j}(n)\Big\|^2_{L_2}\\
&=\Big\|\mathbb{E}\Big(X^2_{i,n}(\beta_n)\Big|\mathbb{S}^{\setminus[j]}\Big)-\mathbb{E}\Big(X^2_{i,n}(\beta_n)\Big|\mathbb{S}^{\setminus[j-1]}\Big)\Big\|^2_{L_2}\\
&\overset{(a)}{=}\Big\|\mathbb{E}\Big(X^2_{i,n}(\beta_n)-\mathbb{E}\big(X^2_{i,n}(\beta_n)|\mathbb{S}(\{j\})\big)\Big|\mathbb{S}^{\setminus[j]}\Big)\Big\|^2_{L_2}\\
&\le\Big\|X^2_{i,n}(\beta_n)-\mathbb{E}\big(X^2_{i,n}(\beta_n)|\mathbb{S}(\{j\})\big)\Big\|^2_{L_2}\\
&\overset{(b)}{\le}\frac{1}{2}\Big\|X_{i,n}(\beta_n,e)^2-X_{i,n}(\beta_n,\tau_{j,l})^2\Big\|^2_{L_2}\\
&\le\frac{1}{2}\Big\|\Big[X_{i,n}(\beta_n,e)-X_{i,n}(\beta_n,\tau_{j,l})\Big]\Big[X_{i,n}(\beta_n,e)+X_{i,n}(\beta_n,\tau_{j,l})\Big]\Big\|^2_{L_2}\\
&\overset{(c)}{\le}4\beta_n^2\lim_{l\to\infty}\Big\|X_{i,n}(\beta_n,e)-X_{i,n}(\beta_n,\tau_{j,l})\Big\|^2_{L_2}
\end{aligned}\tag{7.5}$$

where to get (a) we used Lemma 7.1, for (b) we used Lemma 7.2 and for (c) the fact that $|X_{i,n}|\le\beta_n$ is upper-bounded by β_n.

We want to bound $\Big\|X_{i,n}(\beta_n, e) - X_{i,n}(\beta_n, \tau_{j,l})\Big\|_{L_2}^2$ using $\Delta_{i,j}$. However there is a slight complication coming from the fact that we modified $X_{i,n}$ by upper bounding it by β_n. This will make things a bit more complex.

We can note that for any real $\beta \in \mathbb{R}$ the function $x \to g(\beta, x)$ is $\beta - Lipschitz$. Therefore we have,

$$\begin{aligned}\Big|X_{i,n}(\beta_n, e) - X_{i,n}(\beta_n, \tau_{j,l})\Big| &= \Big|g(\beta_n, X_{i,n}) - g(\beta_n, X_{i,n}(\tau_{j,l})\Big| \\ &\leq \beta_n \Big|X_{i,n} - X_{i,n}(\tau_{j,l})\Big|.\end{aligned}$$

This implies that

$$\begin{aligned}\Big\|Y_{i,j}(n)\Big\|_{L_2} &\leq \beta_n^{\frac{3}{2}} \Big\|X_{i,n} - X_{i,n}(\tau_{j,l})\Big\| \\ &\leq \beta_n^{\frac{3}{2}} \Big\|\Delta_i(f_n, X_n) - \Delta_i(f_n, \tau_{j,l}X)\Big\|_{L_2} \\ &\leq \beta_n^{\frac{3}{2}} \Big\|\Delta_{i,j}(f_n, X_n)\Big\|_{L_2},\end{aligned}$$

Therefore we have that,

$$\begin{aligned}&\mathrm{var}\Big(\frac{1}{\mathrm{n}} \sum_{\mathrm{i} \leq \mathrm{n}} \mathrm{X}_{\mathrm{i,n}}^2(\beta_{\mathrm{n}})\Big) \\ &\leq \frac{1}{n^2} \sum_{j \in \mathbb{N}} \Big[\sum_{i \in \mathbb{N}} \Big\|Y_{i,j}(n)\Big\|_{L_2}\Big]^2 \\ &\leq \frac{\beta_n^3}{n^2} \sum_{j \in \mathbb{N}} \Big[\sum_{i \in \mathbb{N}} \Big\|\Delta_{i,j}(f_n, X_n)\Big\|_{L_2}\Big]^2 \to 0.\end{aligned}$$

□

7.3 Proof of Corollary 4.4

PROOF. According to Theorem 4.3 we know that we only need to prove the following two items, to obtain the desired central-limit theorem,

- The following is uniformly-integrable $(\Delta_i^2(f_n, X_n))_{i,n \in \mathbb{N}}$
- The following is true asymptotically $\frac{1}{n^2} \sum_{j \in \mathbb{N}} \Big[\sum_{i \in \mathbb{N}} \|\Delta_{i,j}(f_n, X_n)\|_{L_2}\Big]^2 \to 0.$

The first thing that will be useful to note is that by hypothesis there is $C < \infty$ such that, $|X_1| \overset{a.s}{\leq} C$.

- <u>Proof of the first point</u> Let $B \in \mathbb{R}$ be any real using the mean-value theorem we get that,

$$\begin{aligned}
&\Big\|\mathbb{I}(|\Delta_i(f_n,X_n)|>B)\ \Delta_i(f_n,X_n)\Big\|_{L_2}\\
&\overset{(a)}{\leq} \lim_{l\to\infty}\Big\|\mathbb{I}(|\Delta_i(f_n,X_n)|>B)\ \big(f_n(X_n)-f_n(\tau_{i,l}X)\big)\Big\|_{L_2}\\
&\overset{(b)}{\leq} \lim_{l\to\infty}\Big\|\mathbb{I}(|\Delta_i(f_n,X_n)|>B)\ \sup_{x\in A}\Big|\frac{\delta f_n}{\delta x_i}(\rho_i^x(X))\Big|\big|X_i-X_l\big|\Big\|_{L_2}\\
&\leq 2C\Big\|\mathbb{I}(|\Delta_i(f_n,X_n)|>B)\ \sup_{x\in A}\Big|\frac{\delta f_n}{\delta x_i}(\rho_i^x(X))\Big|\Big\|_{L_2},
\end{aligned}$$

  where to get (a) we used Lemma 7.2 and (b) is a consequence of the mean value theorem.

  Hence using the exchangeability of X we deduce that $(\Delta_i^2(f_n,X_n))_{i,n\in\mathbb{N}}$ is uniformly integrable as desired.

- <u>Proof of the second point</u> The second point will also be a consequence of sucessive uses of the mean-value theorem.Using Theorem 2.8, we decompose the distribution p_X of X into a mixture of ergodic distributions $p_X(\cdot)=\int_{\mathbb{I}_{\mathbb{S}_\infty}}\nu(\cdot)dm(\nu)$; and we choose X' to be another process such that $(X,X')\sim\int_{\mathbb{I}_{\mathbb{S}_\infty}}\nu(\cdot)\otimes\nu(\cdot)dm(\nu)$.

  The idea is to exploit Lemma 7.3 and the mean value theorem. Indeed we know that there are $\tilde{X}^1,\tilde{X}^2$ random elements of $\mathbb{R}^{\mathbb{N}}$ satisfying

$$\tilde{X}_l^1=\tilde{X}_l^2 \quad \forall l\neq j;$$

  and

$$\begin{aligned}
&\Big\|\big[f_n(X_n)-f_n(X^i)-[f_n(X^j)-f_n(X^{i,j})]\Big\|\\
&\leq\Big\|\frac{\delta f_n}{\delta x_i}(\tilde{X}^1)(X_i-X_i')-\frac{\delta f_n}{\delta x_i}(\tilde{X}^2)(X_i-X_i')\Big\|\\
&\leq 2C\Big\|\frac{\delta f_n}{\delta x_i}(\tilde{X}^1)-\frac{\delta f_n}{\delta x_i}(\tilde{X}^2)\Big\|
\end{aligned}$$

  But by an addditional use of the mean-value theorem we get that,

$$\Big\|\frac{\delta f_n}{\delta x_i}(\tilde{X}^1)-\frac{\delta f_n}{\delta x_i}(\tilde{X}^2)\Big\|\leq 2C\sup_{x,y\in A}\Big\|\frac{\delta^2 f_n}{\delta x_i\delta x_j}(\rho_i^x(\rho_j^y(X))\Big\|_{L_2}$$

  Therefore we get by Lemma 7.3 that,

$$\frac{1}{n^2}\sum_{j\in\mathbb{N}}\Big[\sum_{i\in\mathbb{N}}\Big\|\Delta_{i,j}(f_n,X_n)\Big\|_{L_2}\Big]^2\leq\frac{16C^4}{n^2}\sum_{j\in\mathbb{N}}\Big[\sum_{i\leq n}\sup_{x,y\in A}\Big\|\frac{\delta^2 f_n}{\delta x_i\delta x_j}(\rho_i^x(\rho_j^y(X))\Big\|_{L_2}\Big]^2\to 0$$

□

7.4 Proof of Theorem 4.8

Choose functions (f_n) to be such that $f_n(Z) = \sum_{k\leq n} \left[Y_k - \hat{f}_n^{2,M}((Z_1,\ldots,Z_n),X_k)\right]^2$. According to Theorem 4.3 we know that we only need to prove the following three items, to obtain the desired cental-limit theorem,

- The following is uniformely-integrable $(\Delta_i^2(f_n,Z))_{i,n\in\mathbb{N}}$
- The following is true assymptotically $\frac{1}{n^2}\sum_{i\leq n}\left[\sum_{j\leq n}\|\Delta_{i,j}(f_n,Z)\|_{L_2}\right]^2 \to 0$.
- The following is true assymptotically $\frac{1}{n}\mathrm{var}\big(\mathrm{f_n(Z)}|\mathbb{F}_\infty\big) \xrightarrow{\mathrm{P}} \sigma^2$

The key point to prove each one of those points is to exploit the stability proprety of the linear regression and of the predictors (f_n^m).

First of all let us introduce some notations. We take Z' and Z^* to be two independent process that have the same distribution than Z. For all integers $i,j\in\mathbb{N}$ we choose Z^i, $Z^{i,j}$ and $Z^{*(i,j)}$ to indicate the following processes,

$$Z_l^i = \begin{cases} Z_l \text{ if } \mathrm{l}\neq \mathrm{i} \\ Z_i' \text{ otherwise}\end{cases} \qquad Z_l^{i,j} = \begin{cases} Z_l \text{ if } \mathrm{l}\neq\mathrm{i,j} \\ Z_i' \text{ if } \mathrm{l}=\mathrm{i} \\ Z_j'\mathrm{if l = j}\end{cases} \qquad Z_l^{*(i,j)} = \begin{cases} Z_l \text{ if } \mathrm{l}\neq\mathrm{i,j} \\ Z_i^* \text{ if } \mathrm{l}=\mathrm{i} \\ Z_j^*\mathrm{if l = j}\end{cases}.$$

For ease of notations, for a process X we denote $X_{i:k}$ the random vector $(X_i,\ldots,X_k)$, and we write $B_m(X) := \{X_j | j\leq n\ K_j = m\}$.

Before diving in the main part of the proof it is useful to derive for all integers $j\in\mathbb{N}$ bounds for

$$f_n(Z) - f_n(Z^j) := \sum_{i\leq n}\Big(Y_i - \hat{f}_n^{2,M}(Z_{1:n},X_i)\Big)^2 - \Big(Y_i^j - \hat{f}_n^{2,M}(Z_{1:n}^j,X_i^j)\Big)^2,$$

and to do this we will exploit the definition of $\hat{f}_n^{M,2}$ as a minimimum.

Indeed as the weights $(\hat{\theta}_m(Z_{1:n}))_{m\leq M}$ are chosen to minimize

$$(\theta_1,\ldots,\theta_M) \to \sum_{i=1}^n \Big(Y_i - \sum_{m\leq M}\theta_m f_{|B_m|}^m(B_m(Z),X_i)\Big)^2,$$

we have,

$$
\begin{aligned}
&f_n(Z) - f_n(Z^j)\\
&= \sum_{i \le n} \Big(Y_i - \hat{f}_n^{M,2}(Z_{1:n}, X_i)\Big)^2 - \Big(Y_i^j - \hat{f}_n^{M,2}(Z_{1:n}^j, X_i^j)\Big)^2\\
&\le \sum_{i \le n} \Big(Y_i - \sum_{m \le M} \hat{\theta}_{n,m}(Z_{1:n}^j) f_{|B_m|}^m(B_m(Z), X_i)\Big)^2 - \Big(Y_i^j - \sum_{m \le M} \hat{\theta}_{n,m}(Z_{1:n}^j) f_{|B_m|}^m(B_m(Z^j), X_i^j)\Big)^2
\end{aligned} \tag{7.6}
$$

And reversely we also have,

$$
\begin{aligned}
&f_n(Z) - f_n(Z^j)\\
&= \sum_{i \le n} \Big(Y_i - \hat{f}_n^{M,2}(Z_{1:n}, X_i)\Big)^2 - \Big(Y_i^j - \hat{f}_n^{M,2}(Z_{1:n}^j, X_i^j)\Big)^2\\
&\ge \sum_{i \le n} \Big(Y_i - \sum_{m \le M} \hat{\theta}_{n,m}(Z_{1:n}) f_{|B_m|}^m(B_m(Z), X_i)\Big)^2 - \Big(Y_i^j - \sum_{m \le M} \hat{\theta}_{n,m}(Z_{1:n}) f_{|B_m|}^m(B_m(Z^j), X_i^j)\Big)^2
\end{aligned}
$$

Therefore if for all integers $i, j \in \mathbb{N}$ we denote $\bar{R}_n^{i,j}(Z_{1:n})$ and $\underline{R}_n^{i,j}(Z_{1:n})$ the following quantities,

$$
\begin{aligned}
\bar{R}_n^{i,j}(Z_{1:n}) &:= \Big(Y_i - \sum_{m \le M} \hat{\theta}_{n,m}(Z_{1:n}^j) f_{|B_m|}^m(B_m(Z), X_i)\Big)^2\\
&\quad - \Big(Y_i^j - \sum_{m \le M} \hat{\theta}_{n,m}(Z_{1:n}^j) f_{|B_m|}^m(B_m(Z^j), X_i^j)\Big)^2\\
\underline{R}_n^{i,j}(Z_{1:n}) &:= \Big(Y_i - \sum_{m \le M} \hat{\theta}_{n,m}(Z_{1:n}) f_{|B_m|}^m(B_m(Z), X_i)\Big)^2\\
&\quad - \Big(Y_i^j - \sum_{m \le M} \hat{\theta}_{n,m}(Z_{1:n}) f_{|B_m|}^m(B_m(Z^j), X_i^j)\Big)^2
\end{aligned}
$$

Then we clearly have,

$$
\sum_{i \le n} \underline{R}_n^{i,j}(Z_{1:n}) \le f_n(Z) - f_n(Z^j) \le \sum_{i \le n} \bar{R}_n^{i,j}(Z_{1:n}). \tag{7.7}
$$

The rest of the proof will consist on working on $\bar{R}_n^{i,j}$ and $\underline{R}_n^{i,j}$.

- Proof of the first point Using Equation (7.7) we clearly see that if we can upper-bound $\Big\| \sum_{i \le n} \bar{R}_n^{i,j} \Big\|_{\mathbf{L}_3}$ and $\Big\| \sum_{i \le n} \underline{R}_n^{i,j} \Big\|_{\mathbf{L}_3}$ then we have the desired result.

  To do this note that we supposed that (i) the observations Y take value in a compact i.e $\Big\|Y_1\Big\|_{L_\infty} < \infty$ and (ii) the predictors are bounded, i.e $\sup_{m,n} \Big\|f_n^m\Big\|_{L_\infty} < \infty$. This implies that the optimal weights will also be bounded

$$\sup_{m\le M}\Big\|\hat{\theta}_m(Z)\Big\|_{L_\infty}.$$

Therefore we can easily see that,

$$\begin{aligned}
&\Big\|\sum_{i\le n}\bar{R}_n^{i,j}(Z_{1:n})\Big\|_{\mathbf{L}_3}\\
&\le 2\Big\|\sum_{i\le n}\sum_{m\le M}\hat{\theta}_{n,m}(Z_{1:n}^j)\big[Y_i^j f_{|B_m|}^m(B_m(Z^j),X_i^j)-Y_i f_{|B_m|}^m(B_m(Z),X_i)\big]\Big\|_{\mathbf{L}_\infty}+\Big\|Y_j-Y_j'\Big\|_{\mathbf{L}_\infty}\\
&+\Big\|\sum_{i\le n}\sum_{m,m'\le M}\hat{\theta}_{n,m}(Z_{1:n}^j)\hat{\theta}_{n,m'}(Z_{1:n}^j)\big[f_{|B_m|}^m(B_m(Z),X_i)f_{|B_{m'}|}^{m'}(B_{m'}(Z_{1:n}),X_i)\\
&\qquad\qquad\qquad -f_{|B_m|}^m(B_m(Z^j),X_i^j)f_{|B_{m'}|}^{m'}(B_{m'}(Z_{1:n}^j),X_i^j)\big]\Big\|_{\mathbf{L}_\infty}.
\end{aligned}\tag{7.8}$$

We will upper bound each term on the right hand side of Equation (7.8) separately. By triangular inequality we re-express the first term as:

$$\begin{aligned}
&\Big\|\sum_{i\le n}\sum_{m\le M}\hat{\theta}_{n,m}(Z_{1:n}^j)\big[Y_i^j f_{|B_m|}^m(B_m(Z^j),X_i^j)-Y_i f_{|B_m|}^m(B_m(Z),X_i)\big]\Big\|_{\mathbf{L}_3}\\
&\le\Big\|\sum_{\substack{i\le n\\ j\ne i}}\sum_{m\le M}\hat{\theta}_{n,m}(Z_{1:n}^j)Y_i\big[f_{|B_m|}^m(B_m(Z^j),X_i)-f_{|B_m|}^m(B_m(Z),X_i)\big]\Big\|_{\mathbf{L}_3}\\
&+\Big\|\sum_{m\le M}\hat{\theta}_{n,m}(Z_{1:n}^j)\big[Y_j' f_{|B_m|}^m(B_m(Z^j),X_j')-Y_j^j f_{|B_m|}^m(B_m(Z),X_j)\big]\Big\|_{\mathbf{L}_3}\\
&\le M\Big\|Y_1\Big\|_{L_\infty}\sup_{m\le M}\Big\|\hat{\theta}_{n,m}\Big\|_{L_\infty}\sup_k k\Big\|f_k^m(Z_{1:k},X_1')-f_k^m(Z_{1:k}^1,X_1')\Big\|_{L_3}\\
&\quad+2M\Big\|Y_1\Big\|_{L_\infty}\sup_{m\le M}\Big\|\hat{\theta}_{n,m}\Big\|_{L_\infty}\Big\|f_{|B_m|}^m\Big\|_{L_\infty}.
\end{aligned}$$

The second term is bounded using the hypothesis of the corollary:

$$\Big\|Y_j-Y_j'\Big\|_{\mathbf{L}_\infty}\le 2\big\|Y_1\big\|_{\mathbf{L}_\infty}.$$

Finally, for the last term of Equation (7.8) we can write:

$$\Big\|\sum_{i\leq n}\sum_{m,m'\leq M}\hat{\theta}_{n,m}(Z^j_{1:n})\hat{\theta}_{n,m'}(Z^j_{1:n})\big[f^m_{|B_m|}(B_m(Z),X_i)f^{m'}_{|B_{m'}|}(B_{m'}(Z_{1:n}),X_i)$$

$$-f^m_{|B_m|}(B_m(Z^j),X^j_i)f^{m'}_{|B_{m'}|}(B_{m'}(Z^j_{1:n}),X^j_i)\big]\Big\|_{L_\infty}$$

$$\leq\Big\|\sum_{i\leq n}\sum_{m,m'\leq M}\hat{\theta}_{n,m}(Z^j_{1:n})\hat{\theta}_{n,m'}(Z^j_{1:n})f^{m'}_{|B_{m'}|}(B_m(Z_{1:n}),X_i)$$

$$\big[f^m_{|B_m|}(B_m(Z),X_i)-f^m_{|B_m|}(B_m(Z^j),X^j_i)\big]\Big\|_{L_\infty}$$

$$+\Big\|\sum_{i\leq n}\sum_{m,m'\leq M}\hat{\theta}_{n,m}(Z^j_{1:n})\hat{\theta}_{n,m'}(Z^j_{1:n})f^m_{|B_m|}(B_m(Z),X_i)$$

$$\big[f^{m'}_{|B_{m'}|}(B_{m'}(Z_{1:n}),X_i)-f^{m'}_{|B_{m'}|}(B_{m'}(Z^j_{1:n}),X^j_i)\big]\Big\|_{L_\infty}$$

$$\leq 2M^2\Big\|Y_1\Big\|_{L_\infty}\sup_{m,n,i,j}\Big\|\hat{\theta}_{n,m}\Big\|_{L_\infty}\Big\|f^m_{|B_m|}(B_m(Z),X_i)\Big|_{L_\infty}\Big\|f^m_{|B_m|}(B_m(Z^j),X^j_i)-f^m_{|B_m|}(B_m(Z),X_i)\Big\|_{L_\infty}$$

Therefore using the fact that we assumed the estimators to be stable we have that

$$\sup_{j,n\in\mathbb{N}}\Big\|\sum_{i\leq n}\bar{R}^{i,j}_n(Z_{1:n})\Big\|_{L_\infty}<\infty$$

In the exact same fashion we can also obtain that

$$\sup_{j,n\in\mathbb{N}}\Big\|\sum_{i\leq n}\underline{R}^{i,j}_n(Z_{1:n})\Big\|_{L_\infty}<\infty$$

Therefore this implies that $(\Delta^2_i(f_n,Z))_{i,n\in\mathbb{N}}$ is bounded and therefore uniformely integrable.

- Proof of the second point The key element of the second point will be to bound for all integers $k,j\in\mathbb{N}$ the following quantities

$$\Big|\sum_{i\leq n}\underline{R}^{i,k}_n(Z)-\bar{R}^{i,k}_n(Z^j)\Big|,\quad\text{and}\ \Big|\sum_{\mathrm{i\leq n}}\bar{\mathrm{R}}^{\mathrm{i,k}}_{\mathrm{n}}(\mathrm{Z})-\underline{\mathrm{R}}^{\mathrm{i,k}}_{\mathrm{n}}(\mathrm{Z^j})\Big|.$$

And to do this we will need to give an upper bound to $n\Big\|\underline{\theta}_n(Z_{1:n})-\underline{\theta}_n(Z^k_{1:n})\Big\|$ for all integer $k\in\mathbb{N}$ where $\underline{\theta}_n(Z_{1:n})$ designates the vector of optimal weights,

$$\underline{\theta}_n(Z_{1:n}):=\big(\hat{\theta}_1(Z_{1:n}),\ldots,\hat{\theta}_M(Z_{1:n})\big).$$

This will be possible by exploiting the propreties of the weights $\underline{\theta}_n(Z_{1:n})$. Indeed under the event that the smallest eigenvalue of the matrix $F(n,Z_{1:n})$ is bigger than

$n\epsilon$ we have that the vector of weights is equal to the following,

$$\underline{\theta}_n(Z_{1:n}) = \Big(F(n, Z_{1:n})^T F(n, Z_{1:n})\Big)^{-1} F(n, Z_{1:n}) Y_{1:n}.$$

For shortness of notation we will denote for all integers $k, n \in \mathbb{N}$ the following event, $E_n^k := \{\min\Big(\lambda_{(1)}\big(F(n, Z_{1:n})^T F(n, Z_{1:n})\big), \lambda_{(1)}\big(F(n, Z_{1:n}^k)^T F(n, Z_{1:n})\big)\Big) > n\epsilon\}$. Therefore under $E_n^k(Z_{1:n})$ by triangular inequality we get the following,

$$\begin{aligned}
&\Big\|\mathbb{I}\big(E_n^k\big)\left[\underline{\theta}_n(Z_{1:n}) - \underline{\theta}_n(Z_{1:n}^k)\right]\Big\|_{L_2} \\
&\leq \Big\|\mathbb{I}\big(E_n^k\big)\Big(F(n, Z_{1:n})^T F(n, Z_{1:n})\Big)^{-1} F(n, Z_{1:n})\Big[Y_{1:n} - Y_{1:n}^k\Big]\Big\|_{L_2} \\
&+ \Big\|\mathbb{I}\big(E_n^k\big)\Big(F(n, Z_{1:n})^T F(n, Z_{1:n})\Big)^{-1}\Big[F(n, Z_{1:n}) - F(n, Z_{1:n}^k)\Big] Y_{1:n}^k\Big\|_{L_2} \\
&+ \Big\|\mathbb{I}\big(E_n^k\big)\Big[\Big(F(n, Z_{1:n})^T F(n, Z_{1:n})\Big)^{-1} - F(n, Z_{1:n}^k)^T F(n, Z_{1:n}^k)\Big)^{-1}\Big] F(n, Z_{1:n}^k) Y_{1:n}^k\Big\|_{L_2}.
\end{aligned} \tag{7.9}$$

And we are going to bound each term successively. Indeed firstly we have,

$$\begin{aligned}
&\Big\|\mathbb{I}\big(E_n^k\big)\Big(F(n, Z_{1:n})^T F(n, Z_{1:n})\Big)^{-1} F(n, Z_{1:n})\Big[Y_{1:n} - Y_{1:n}^k\Big]\Big\|_{L_2} \\
&\overset{(a)}{\leq} \frac{1}{n\epsilon}\Big\|\mathbb{I}\big(E_n^k\big) F(n, Z_{1:n})\Big[Y_{1:n} - Y_{1:n}^k\Big]\Big\|_{L_2} \\
&\overset{(b)}{\leq} \frac{2M}{n\epsilon} \sup_m \Big\|f_{|B_m|}^m\Big\|_{L_\infty} \Big\|Y_1\Big\|_{L_\infty}
\end{aligned}$$

where to get (a) we used the fact that under $E_n(Z_{1:n})$ we have $\lambda_{(1)}(F(n, Z_{1:n})^T F(n, Z_{1:n})) > n\epsilon$ and (b) comes from the following equality,

$$F(n, Z_{1:n})\Big[Y_{1:n}^k - Y_{1:n}^k\Big] = \Big(f_{|B_m|}^m(B_m(Z), X_k)\Big[Y_k - Y_k'\Big]\Big)_{m \leq M}.$$

Similarly we are get that,

$$\begin{aligned}
&\Big\|\mathbb{I}\big(E_n^k\big)\Big(F(n, Z_{1:n})^T F(n, Z_{1:n})\Big)^{-1}\Big[F(n, Z_{1:n}) - F(n, Z_{1:n}^k)\Big] Y_{1:n}^k\Big\|_{L_2} \\
&\overset{(a)}{\leq} \frac{1}{n\epsilon}\Big\|\Big[F(n, Z_{1:n}) - F(n, Z_{1:n}^k)\Big] Y_{1:n}^k\Big\|_{L_2} \\
&\leq \frac{M}{\epsilon} \sup_m \Big\|\Delta_k(f_{|B_m|}^m, Z)\Big\|_{L_2} \Big\|Y_1\Big\|_{L_\infty}
\end{aligned}$$

where to get (a) we used the fact that under $E_n(Z_{1:n})$ we have

$$\lambda_{(1)}(F(n, Z_{1:n})^T F(n, Z_{1:n})) > n\epsilon.$$

And finally for the last term of Equation (7.11) we get that,

$$
\begin{aligned}
&\Big\|\mathbb{I}(E_n^k)\Big[\Big(F(n,Z_{1:n})^T F(n,Z_{1:n})\Big)^{-1} - F(n,Z_{1:n}^k)^T F(n,Z_{1:n}^k)\Big)^{-1}\Big]F(n,Z_{1:n}^k)Y_{1:n}^k\Big\|_{L_2}\\
&\le nM\Big\|Y\Big\|_{L_\infty}\sup_{m\le M}\Big\|f^m_{|B_m|}\Big\|_{L_\infty}\Big\|\mathbb{I}(E_n^k)\Big(F(n,Z_{1:n})^T F(n,Z_{1:n})\Big)^{-1} - \Big(F(n,Z_{1:n}^k)^T F(n,Z_{1:n}^k)\Big)^{-1}\Big\|_{L_2}\\
&\overset{(a)}{\le} \frac{M}{n\epsilon^2}\Big\|Y\Big\|_{L_\infty}\sup_{m\le M}\Big\|f^m_{|B_m|}\Big\|_{L_\infty}\Big\|F(n,Z_{1:n})^T F(n,Z_{1:n}) - F(n,Z_{1:n}^k)^T F(n,Z_{1:n}^k)\Big\|_{L_2}\\
&\overset{(b)}{\le} \frac{M}{\epsilon^2}\Big\|Y\Big\|_{L_\infty}\sup_{m\le M}\Big\|f^m_{|B_m|}\Big\|^2_{L_\infty}\Big\|\Delta_k(\hat{f}_{m,n},Z)\Big\|_{L_2}.
\end{aligned}
\tag{7.10}
$$

where (a) comes from the fact for any two invertible matrices A and B we have that,

$$
\Big\|A^{-1}-B^{-1}\Big\| \le \Big\|A^{-1}\Big\|\;\Big\|B^{-1}\Big\|\;\Big\|A-B\Big\|
$$

and where to (b) we used the following inequality,

$$
\begin{aligned}
&\Big\|F(n,Z_{1:n})^T F(n,Z_{1:n}) - F(n,Z_{1:n}^k)^T F(n,Z_{1:n}^k)\Big\|_{L_2}\\
&\le \sum_{m,m'\le M}\sum_{i\le n}\Big\|\hat{f}_{n,m}(B_m(Z_{1:n}),X_i)f^{m'}_{|B_{m'}|}(B_{m'}(Z_{1:n}),X_i)\\
&\qquad\qquad - \hat{f}_{n,m}(B_m(Z_{1:n}^k),X_i^k)f^{m'}_{|B_{m'}|}(B_{m'}(Z_{1:n}),X_i^k)\Big\|_{L_2}\\
&\le \sup_{m\le M}\Big\|f^m_{|B_m|}\Big\|_{L_\infty}\Big\|\Delta_k(\hat{f}_{m,n},Z)\Big\|_{L_2}
\end{aligned}
$$

Therefore combining Section 7.4, Section 7.4, Equation (7.10) and the stability of the estimators we easily obtain that,

$$
\Big\|\underline{\theta}_n(Z_{1:n}) - \underline{\theta}_n(Z_{1:n}^k)\Big\|_{L_2} = O(\frac{1}{n}). \tag{7.11}
$$

Now that we proved this we are ready to bound $\Big\|\sum_{i\le n}\underline{R}_n^{i,k}(Z) - \bar{R}_n^{i,k}(Z^j)\Big\|$. And for this we will first write

$$
\begin{aligned}
&\Big\|\sum_{i\le n}\underline{R}_n^{i,k}(Z) - \bar{R}_n^{i,k}(Z^j)\Big\|\\
&\le \Big\|\mathbb{I}\Big(E_n^k(Z_{1:n}),E_n^k(Z_{1:n}^j)\Big)\sum_{i\le n}\underline{R}_n^{i,k}(Z) - \bar{R}_n^{i,k}(Z^j)\Big\|\\
&+ \Big\|\mathbb{I}\Big(E_n^k(Z_{1:n})^c\bigcup E_n^k(Z_{1:n}^j)^c\Big)\sum_{i\le n}\underline{R}_n^{i,k}(Z) - \bar{R}_n^{i,k}(Z^j)\Big\|
\end{aligned}
\tag{7.12}
$$

Bounding the first term on the right hand side of Equation (7.12) will be done using Equation (7.11) and the second one simply by noting that the probability of $E_n^k(Z_{1:n})^c \bigcup E_n^k(Z_{1:n}^j)^c$ is small. Indeed we have,

$$\begin{aligned}
&\Big\|\mathbb{I}\Big(E_n^k(Z_{1:n})^c \bigcup E_n^k(Z_{1:n}^j)^c\Big) \sum_{i\leq n} \underline{R}_n^{i,k}(Z) - \bar{R}_n^{i,k}(Z^j)\Big\| \\
&\leq \Big[\Big\|\sum_{i\leq n} \underline{R}_n^{i,k}\Big\|_{L_\infty} + \Big\|\sum_{i\leq n} \bar{R}_n^{i,k}\Big\|_{L_\infty}\Big] P(E_n^k(Z_{1:n})^c \bigcup E_n^k(Z_{1:n}^j)^c).
\end{aligned}$$

But we already previously proved that

$$\sup_{k,n} \Big[\Big\|\sum_{i\leq n} \underline{R}_n^{i,k}\Big\|_{L_\infty} + \Big\|\sum_{i\leq n} \bar{R}_n^{i,k}\Big\|_{L_\infty}\Big] < \infty$$

and that we assumed that,

$$P(E_n^k(Z_{1:n})^c \bigcup E_n^k(Z_{1:n}^j)^c) = o(\frac{1}{\sqrt{n}})$$

Therefore we have that,

$$\Big\|\mathbb{I}\Big(E_n^k(Z_{1:n})^c \bigcup E_n^k(Z_{1:n}^j)^c\Big) \sum_{i\leq n} \underline{R}_n^{i,k}(Z) - \bar{R}_n^{i,k}(Z^j)\Big\| = o(\frac{1}{\sqrt{n}}) \tag{7.13}$$

Now that this has been proven we will move on to bounding the first term of Equation (7.12). For this first we will remember that

$$\begin{aligned}
&\sum_{i\leq n} \bar{R}_n^{i,j}(Z_{1:n}) \\
&= 2\sum_{i\leq n}\sum_{m\leq M} \hat{\theta}_{n,m}(Z_{1:n}^j)\big[Y_i^j f_{|B_m|}^m(B_m(Z^j), X_i^j) - Y_i f_{|B_m|}^m(B_m(Z), X_i)\big] + [Y_j - Y_j'] \\
&+ \sum_{i\leq n}\sum_{m,m'\leq M} \hat{\theta}_{n,m}(Z_{1:n}^j)\hat{\theta}_{n,m'}(Z_{1:n}^j)\big[f_{|B_m|}^m(B_m(Z), X_i) f_{|B_{m'}|}^{m'}(B_{m'}(Z_{1:n}), X_i) \\
&\qquad - f_{|B_m|}^m(B_m(Z^j), X_i^j) f_{|B_{m'}|}^{m'}(B_{m'}(Z_{1:n}^j), X_i^j)\big]
\end{aligned}$$

We will want to break $\sum_{i\leq n} \bar{R}_n^{i,j}(Z_{1:n}) - R_n^{i,j}(Z_{1:n})$ into several parts and then bound each one of them seperately.

For simplicity of notation for all integers $i, j, m \in \mathbb{N}$ we will write

$$\mathcal{D}_{i,j}^{m,\hat{\theta}}(Z_{1:n}) := \hat{\theta}_{n,m}(Z_{1:n})\big[Y_i^j f_{|B_m|}^m(B_m(Z^j), X_i^j) - Y_i f_{|B_m|}^m(B_m(Z), X_i)\big],$$

and

$$\mathcal{D}_{i,j}^{m}(Z_{1:n}) := Y_i^j f_{|B_m|}^m(B_m(Z^j), X_i^j) - Y_i f_{|B_m|}^m(B_m(Z), X_i),$$

We can note that we have already proven that $\sup_{j,n\in\mathbb{N}}\Big\|\sum_{i\le n}\mathcal{D}^m_{i,j}(Z_{1:n})\Big\|_{L_\infty}<\infty$. Using triangular inequality we then get that,

$$\begin{aligned}
&\Big\|\mathbb{I}\Big(E^k_n(Z_{1:n}),E^k_n(Z^j_{1:n})\Big)\sum_{i\le n}\sum_{m\le M}\mathcal{D}^{m,\hat\theta}_{i,j}(Z_{1:n})-\mathcal{D}^{m,\hat\theta}_{i,j}(Z^k_{1:n})\Big\|_{L_2}\\
&\le\Big\|\mathbb{I}\Big(E^k_n(Z_{1:n}),E^k_n(Z^j_{1:n})\Big)\sum_{i\le n}\sum_{m\le M}\Big[\hat\theta_{n,m}(Z_{1:n})-\hat\theta_{n,m}(Z^{j,k}_{1:n})\Big]\mathcal{D}^m_{i,j}(Z_{1:n})\Big\|_{L_2}\\
&+\Big\|\sum_{i\le n}\sum_{m\le M}\hat\theta_{n,m}(Z^{j,K}_{1:n})\big[\mathcal{D}^m_{i,j}(Z_{1:n})-\mathcal{D}^m_{i,j}(Z^k_{1:n})\big]\Big\|_{L_2}\\
&\le\Big\|\mathbb{I}\Big(E^k_n(Z_{1:n}),E^k_n(Z^j_{1:n})\Big)\big[\underline{\hat\theta}_{n,m}(Z_{1:n})-\underline{\hat\theta_{n,m}}(Z^{j,k}_{1:n})\big]\Big\|_{L_2}\Big\|\sum_{i\le n}\mathcal{D}^m_{i,j}(Z_{1:n})\Big\|_{L_\infty} \qquad (7.14)\\
&\quad+\sup_m\Big\|\hat\theta_{n,m}\Big\|_{L_\infty}\Big\|\sum_{i\le n}\sum_{m\le M}\mathcal{D}^m_{i,j}(Z_{1:n})-\mathcal{D}^m_{i,j}(Z^k_{1:n})\Big\|_{L_2}\\
&\le\Big\|\mathbb{I}\Big(E^k_n(Z_{1:n}),E^k_n(Z^j_{1:n})\Big)\big[\underline{\hat\theta}_{n,m}(Z_{1:n})-\underline{\hat\theta_{n,m}}(Z^{j,k}_{1:n})\big]\Big\|_{L_2}\Big\|\sum_{i\le n}\mathcal{D}^m_{i,j}(Z_{1:n})\Big\|_{L_\infty}\\
&+\sup_m\Big\|\hat\theta_{n,m}\Big\|_{L_\infty}\Big\|Y\Big\|_{L_\infty}\Big[\sup_{k\in\mathbb{N}}\Big\|\Delta_k(f^m_{|B_m|},Z)\Big\|_{L_2}+n\Big\|\Delta_{k,j}(f^m_{|B_m|},Z)\Big\|_{L_2}\Big].
\end{aligned}$$

Therefore we get that,

$$\begin{aligned}
&\Big\|\mathbb{I}\Big(E^k_n(Z_{1:n}),E^k_n(Z^j_{1:n})\Big)\sum_{i\le n}\sum_{m\le M}\mathcal{D}^{m,\hat\theta}_{i,j}(Z_{1:n})-\mathcal{D}^{m,\hat\theta}_{i,j}(Z^k_{1:n})\Big\|_{L_2}\\
&=O(\frac{1}{n}+n\sup_{m\le M}\Big\|\Delta_{k,j}(f^m_{|B_m|},Z)\Big\|_{L_2}.
\end{aligned} \qquad (7.15)$$

Therefore the only term that we have not bounded yet is

$$\Big\|\mathbb{I}\Big(E^k_n(Z_{1:n}),E^k_n(Z^j_{1:n})\Big)\sum_{i\le n}\sum_{m,m'\le M}\mathcal{E}^{m,m',\hat\theta}_{i,j}(Z_{1:n})-\mathcal{E}^{m,m',\hat\theta}_{i,j}(Z^k_{1:n})\Big\|_{L_2}), \qquad (7.16)$$

where for all integers $m,m',i,j\in\mathbb{N}$ we set,

$$\begin{aligned}
\mathcal{E}^{m,m',\hat\theta}_{i,j}(Z_{1:n}):=&\hat\theta_{n,m}(Z^j_{1:n})\hat\theta_{n,m'}(Z^j_{1:n})\big[f^m_{|B_m|}(B_m(Z),X_i)f^{m'}_{|B_{m'}|}(B_{m'}(Z_{1:n}),X_i)\\
&-f^m_{|B_m|}(B_m(Z^j),X^j_i)f^{m'}_{|B_{m'}|}(B_{m'}(Z^j_{1:n}),X^j_i)\big].
\end{aligned} \qquad (7.17)$$

The idea to bound this term will be exactly the same, and as previously to simplify matters we will introduce a new term,

$$\begin{aligned}
\mathcal{E}^{m,m'}_{i,j}(Z_{1:n}):=&f^m_{|B_m|}(B_m(Z),X_i)f^{m'}_{|B_{m'}|}(B_{m'}(Z_{1:n}),X_i)\\
&-f^m_{|B_m|}(B_m(Z^j),X^j_i)f^{m'}_{|B_{m'}|}(B_{m'}(Z^j_{1:n}),X^j_i).
\end{aligned} \qquad (7.18)$$

As previously we have already established that, $\sup_{m,m',j,n\in\mathbb{N}} |\Big\| \sum_{i\leq n}\sum_{m,m'\leq M} \mathcal{E}_{i,j}^{m,m'}(Z_{1:n})\Big\|_{L_\infty} < \infty$.

Therefore we get that

$$\begin{aligned}
&\Big\|\mathbb{I}\Big(E_n^k(Z_{1:n}), E_n^k(Z_{1:n}^j)\Big)\sum_{i\leq n}\sum_{m,m'\leq M}\mathcal{E}_{i,j}^{m,m',\hat{\theta}}(Z_{1:n}) - \mathcal{E}_{i,j}^{m,m',\hat{\theta}}(Z_{1:n}^k)\Big\|_{L_2} \\
&\leq \Big\|\mathbb{I}\Big(E_n^k(Z_{1:n})\big[\underline{\theta}(Z_{1:n}^k) - \underline{\theta}(Z_{1:n})\big]\Big\|_{L_2}\Big\|\sum_{i\leq n}\sum_{m,m'\leq M}\mathcal{E}_{i,j}^{m,m'}(Z_{1:n})\Big\|_{L_\infty} \\
&+ M\sup_{m\leq M}\Big\|\hat{\theta}_m\Big\|_{L_\infty}\Big\|\sum_{i\leq n}\sum_{m,m'\leq M}\mathcal{E}_{i,j}^{m,m'}(Z_{1:n}) - \mathcal{E}_{i,j}^{m,m'}(Z_{1:n}^k)\Big\|_{L_2} \\
&\leq \Big\|\mathbb{I}\Big(E_n^k(Z_{1:n})\big[\underline{\theta}(Z_{1:n}^k) - \underline{\theta}(Z_{1:n})\big]\Big\|_{L_2}\Big\|\sum_{i\leq n}\sum_{m,m'\leq M}\mathcal{E}_{i,j}^{m,m'}(Z_{1:n})\Big\|_{L_\infty} \\
&+ M\sup_{m\leq M}\Big\|\hat{\theta}_m\Big\|_{L_\infty}\Big\|f_{|B_m|}^m\Big\|\sup_{m\leq M} n\Big\|\Delta_{j,k}(f_{|B_m|}^m, Z)\Big\|_{L_2}
\end{aligned}$$

Therefore we get that similarly to Equation (7.14) that,

$$\begin{aligned}
&\Big\|\mathbb{I}\Big(E_n^k(Z_{1:n}), E_n^k(Z_{1:n}^j)\Big)\sum_{i\leq n}\sum_{m,m'\leq M}\mathcal{E}_{i,j}^{m,m',\hat{\theta}}(Z_{1:n}) - \mathcal{E}_{i,j}^{m,m',\hat{\theta}}(Z_{1:n}^k)\Big\|_{L_2} \\
&= O(\frac{1}{n} + \sup_{m\leq M} n\Big\|\Delta_{j,k}(f_{|B_m|}^m, Z)\Big\|_{L_2}).
\end{aligned} \tag{7.19}$$

By combining Equation (7.13),Equation (7.15) and Equation (7.19) we proved the desired point.

- Proof of the third pointTo prove the desired result we will re-express $\mathrm{var}(\mathrm{f_n(Z)})$ in the following way,

$$\begin{aligned}
\frac{1}{n}\mathrm{var}(\mathrm{f_n(Z)}) &= \frac{1}{n}\mathrm{var}\Big(\sum_{\mathrm{i\leq n}}\hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_i})\Big) \\
&= \frac{1}{n}\sum_{i\leq n}\mathrm{var}\Big(\hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_i})\Big) + \frac{2}{\mathrm{n}}\sum_{\mathrm{i<j\leq n}}\mathrm{cov}\Big(\hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_i}), \hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_j})\Big)
\end{aligned} \tag{7.20}$$

We will then seperately re-express each term on the righ-hand side of Equation (7.20). The key idea to do so is to exploit the stability assumptions on the estimators.

First we can notice that for all integers $i \in \mathbb{N}$ we have that ,

$$\begin{aligned}
&\Big|\mathrm{var}\Big(\hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_i})\Big) - \mathrm{var}\Big(\hat{\mathrm{R}}_{\mathrm{Z^i_{1:n}}}(\mathrm{Z_i})\Big)\Big| \\
&\leq \Big|\mathbb{E}\Big(\hat{R}_{Z_{1:n}}(Z_i)^2 - \hat{R}_{Z^i_{1:n}}(Z_i)^2\Big)\Big| + \Big|\mathbb{E}\Big(\hat{R}_{Z_{1:n}}(Z_i)\Big)^2 - \mathbb{E}\Big(\hat{R}_{Z^i_{1:n}}(Z_i)\Big)^2\Big| \\
&\leq 4\Big\|\hat{R}_{Z_{1:n}}\Big\|_{L_\infty}\Big\|\hat{R}_{Z_{1:n}}(Z_i) - \hat{R}_{Z^i_{1:n}}(Z_i)\Big)\Big\|_{L_2}
\end{aligned}$$

But we had already proved that $\sup_{n\in\mathbb{N}}\Big\|\hat{R}_{Z_{1:n}}\Big\|_{L_\infty} < \infty$. Moreover we can easily see that

$$\sup_{i,n\in\mathbb{N}}\Big\|\hat{R}_{Z_{1:n}}(Z_i) - \hat{R}_{Z^i_{1:n}}(Z_i)\Big)\Big\|_{L_2} = O(\frac{1}{n}) \tag{7.21}$$

Indeed by triangular inequality we have,

$$\begin{aligned}
&\Big\|\hat{R}_{Z_{1:n}}(Z_i) - \hat{R}_{Z^i_{1:n}}(Z_i)\Big)\Big\|_{L_2} \\
&\leq \Big\|Y_i\Big[\sum_{m\leq M}\hat{\theta}_{n,n}(Z_{1:n})f^m_{|B_m|}(B_m(Z),Z_i) - \sum_{m\leq M}\hat{\theta}_{n,n}(Z^i_{1:n})f^m_{|B_m|}(B_m(Z^i_{1:n}),Z^i_i)\Big]\Big\|_{L_2} \\
&+ \Big\|\Big[\sum_{m\leq M}\hat{\theta}_{n,n}(Z_{1:n})f^m_{|B_m|}(B_m(Z),Z_i)\Big]^2 - \Big[\sum_{m\leq M}\hat{\theta}_{n,n}(Z^i_{1:n})f^m_{|B_m|}(B_m(Z^i_{1:n}),Z^i_i)\Big]^2\Big\|_{L_2} \\
&\leq M\Big\|Y_1\Big\|_{L_\infty}\Big[\sup_{m\leq M}\Big\|f^m_{|B_m|}\Big\|_{L_\infty}\Big\|\underline{\theta}_{m,n}(Z_{1:n}) - \underline{\theta}_{m,n}(Z^i_{1:n})\Big\|_{L_2} + \sup_{m\leq M}\Big\|\hat{\theta}_{m,n}\Big\|_{L_\infty}\Big\|\Delta_i(\hat{f}_{m,n},Z)\Big\|_{L_2}\Big] \\
&\quad + M^2\sup_{m\leq M}\Big\|f^m_{|B_m|}\Big\|_{L_\infty}\sup_{m\leq M}\Big\|\hat{f}_{m,n}\Big\|_{L_\infty}\Big\|\underline{\theta}_{m,n}(Z_{1:n}) - \underline{\theta}_{m,n}(Z^i_{1:n})\Big\|_{L_2} \\
&\quad + M^2\sup_{m\leq M}\Big\|f^m_{|B_m|}\Big\|_{L_\infty}\sup_{m\leq M}\Big\|\hat{\theta}_{m,n}\Big\|_{L_\infty}\Big\|\Delta_i(f^m_{|B_m|},Z)\Big\|_{L_2}\Big] = O(\frac{1}{n})
\end{aligned}$$

Therefore we can easily see that,

$$\frac{1}{n}\sum_{i\leq n}\big[\mathrm{var}\Big(\hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_i})\Big) - \mathrm{var}\Big(\hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z'_1})\Big)\big] \to 0. \tag{7.22}$$

Similarly we can phrase a similar argument for the covariances. Indeed exploiting the fact that the covariance is a quadratic form we can see that for all integers $i, j \in \mathbb{N}$ we have,

$$\begin{aligned}
&\mathrm{cov}\Big(\hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_i}),\hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_j})\Big) \\
&= \mathrm{cov}\Big(\hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_i}) - \hat{\mathrm{R}}_{\mathrm{Z^{i,j}_{1:n}}}(\mathrm{Z_i}),\hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_j})\Big) + \mathrm{cov}\Big(\hat{\mathrm{R}}_{\mathrm{Z^{i,j}_{1:n}}}(\mathrm{Z_i}),\hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_j})\Big) \\
&= \mathrm{cov}\Big(\hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_i}) - \hat{\mathrm{R}}_{\mathrm{Z^{i,j}_{1:n}}}(\mathrm{Z_i}),\hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_j}) - \hat{\mathrm{R}}_{\mathrm{Z^{*(i,j)}_{1:n}}}(\mathrm{Z_j})\Big) + \mathrm{cov}\Big(\hat{\mathrm{R}}_{\mathrm{Z^{i,j}_{1:n}}}(\mathrm{Z_i}),\hat{\mathrm{R}}_{\mathrm{Z^{*(i,j)}_{1:n}}}(\mathrm{Z_j})\Big) \\
&\quad - \mathrm{cov}\Big(\hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_i}) - \hat{\mathrm{R}}_{\mathrm{Z^{i,j}_{1:n}}}(\mathrm{Z_i}),\hat{\mathrm{R}}_{\mathrm{Z^{*(i,j)}_{1:n}}}(\mathrm{Z_j})\Big) - \mathrm{cov}\Big(\hat{\mathrm{R}}_{\mathrm{Z^{i,j}_{1:n}}}(\mathrm{Z_i}),\hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_j}) - \hat{\mathrm{R}}_{\mathrm{Z^{*(i,j)}_{1:n}}}(\mathrm{Z_j})\Big)
\end{aligned}$$

And we will show now that at the exception of $\mathrm{cov}\Big(\hat{\mathrm{R}}_{\mathrm{Z}^{\mathrm{i,j}}_{1:n}}(\mathrm{Z_i}), \hat{\mathrm{R}}_{\mathrm{Z}^{\mathrm{i,j}}_{1:n}}(\mathrm{Z_j})\Big)$ all the other terms are of the order of $\frac{1}{n^2}$ or smaller. And the reason for this will be different according to the different terms. First we can see by independence of respetivelly (Z_i, Z_j), (Z'_i, Z'_j) and (Z^*_i, Z^*_j) that we have,

$$\mathrm{cov}\Big(\hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_i}) - \hat{\mathrm{R}}_{\mathrm{Z}^{\mathrm{i,j}}_{1:n}}(\mathrm{Z_i}), \hat{\mathrm{R}}_{\mathrm{Z}^{*(\mathrm{i,j})}_{1:n}}(\mathrm{Z_j})\Big) + \mathrm{cov}\Big(\hat{\mathrm{R}}_{\mathrm{Z}^{\mathrm{i,j}}_{1:n}}(\mathrm{Z_i}), \hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_j}) - \hat{\mathrm{R}}_{\mathrm{Z}^{*(\mathrm{i,j})}_{1:n}}(\mathrm{Z_j})\Big) = 0$$

Therefore there is only one term left to bound, but by Cauchy-Swartz and Equation (7.21) we can easily see that

$$\Big|\mathrm{cov}\Big(\hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_i}) - \hat{\mathrm{R}}_{\mathrm{Z}^{\mathrm{i,j}}_{1:n}}(\mathrm{Z_i}), \hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_j}) - \hat{\mathrm{R}}_{\mathrm{Z}^{*(\mathrm{i,j})}_{1:n}}(\mathrm{Z_j})\Big)\Big| = \mathrm{O}(\frac{1}{\mathrm{n}^2})$$

Therefore this implies that,

$$\Big|\frac{1}{n}\sum_{i<j\leq n} \mathrm{cov}\Big(\hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_i}), \hat{\mathrm{R}}_{\mathrm{Z_{1:n}}}(\mathrm{Z_j})\Big) - \mathrm{cov}\Big(\hat{\mathrm{R}}_{\mathrm{Z}^{\mathrm{i,j}}_{1:n}}(\mathrm{Z_i}), \hat{\mathrm{R}}_{\mathrm{Z}^{*(\mathrm{i,j})}_{1:n}}(\mathrm{Z_j})\Big)\Big| = \mathrm{O}(\frac{1}{\mathrm{n}}) \tag{7.23}$$

And combining Equation (7.22) and Equation (7.23) we can easily see that,

$$\Big|\frac{1}{n}\mathrm{var}(\mathrm{f_n}(\mathrm{Z})) - \mathrm{M}\mathrm{var}\Big(\hat{\mathrm{R}}_{\mathrm{Z}^{1}_{1:n}}(\mathrm{Z_1})\Big) - 2\mathrm{M}^2\mathrm{n}\ \mathrm{cov}\Big(\hat{\mathrm{R}}_{\mathrm{Z}^{1,2}_{1:n}}(\mathrm{Z_1}), \hat{\mathrm{R}}_{\mathrm{Z}^{*(1,2)}_{1:n}}(\mathrm{Z_2})\Big)\Big| = \mathrm{o}(1) \tag{7.24}$$

Therefore we will want to prove that

$$M\mathrm{var}\Big(\hat{\mathrm{R}}_{\mathrm{Z}^{1}_{1:n}}(\mathrm{Z_1})\Big) - 2\mathrm{M}^2\mathrm{n}\ \mathrm{cov}\Big(\hat{\mathrm{R}}_{\mathrm{Z}^{1,2}_{1:n}}(\mathrm{Z_1}), \hat{\mathrm{R}}_{\mathrm{Z}^{*(1,2)}_{1:n}}(\mathrm{Z_2})\Big) \to \sigma^2$$

Exploiting the classical decomposition of the covariance function with respect to the conditional covariance we get that,

$$\begin{aligned}&\mathrm{cov}\Big(\hat{\mathrm{R}}_{\mathrm{Z}^{1,2}_{1:n}}(\mathrm{Z_1}), \hat{\mathrm{R}}_{\mathrm{Z}^{*(1,2)}_{1:n}}(\mathrm{Z_2})\Big)\\ &= \mathbb{E}\Big(\mathrm{cov}\Big(\hat{\mathrm{R}}_{\mathrm{Z}^{1,2}_{1:n}}(\mathrm{Z_1}), \hat{\mathrm{R}}_{\mathrm{Z}^{*(1,2)}_{1:n}}(\mathrm{Z_2})\Big|\hat{\mathrm{f}}^{2,\mathrm{M}}_{\mathrm{Z_{1:n}}}\Big)\Big) + \mathrm{cov}\Big(\mathbb{E}\big(\hat{\mathrm{R}}_{\mathrm{Z}^{1,2}_{1:n}}(\mathrm{Z_1})\big|\hat{\mathrm{f}}^{2,\mathrm{M}}_{\mathrm{Z_{1:n}}}\big), \mathbb{E}\big(\hat{\mathrm{R}}_{\mathrm{Z}^{1,2}_{1:n}}(\mathrm{Z_2})\big|\hat{\mathrm{f}}^{2,\mathrm{M}}_{\mathrm{Z_{1:n}}}\big)\Big)\end{aligned}$$

But by independence of Z_1 and Z_2 we can easily see that,

$$\mathrm{cov}\Big(\hat{\mathrm{R}}_{\mathrm{Z}^{1,2}_{1:n}}(\mathrm{Z_1}), \hat{\mathrm{R}}_{\mathrm{Z}^{*(1,2)}_{1:n}}(\mathrm{Z_2})\Big|\hat{\mathrm{f}}^{2,\mathrm{M}}_{\mathrm{Z_{1:n}}}\Big) \overset{\mathrm{a.s}}{=} 0.$$

And therefore we have that,

$$\begin{aligned}&\mathrm{cov}\Big(\hat{\mathrm{R}}_{\mathrm{Z}^{1,2}_{1:n}}(\mathrm{Z}_1), \hat{\mathrm{R}}_{\mathrm{Z}^{*(1,2)}_{1:n}}(\mathrm{Z}_2)\Big)\\ &= \mathrm{cov}\Big(\mathbb{E}\big(\hat{\mathrm{R}}_{\mathrm{Z}^{1,2}_{1:n}}(\mathrm{Z}_1)\big|\hat{\mathrm{f}}^{2,\mathrm{M}}_{\mathrm{Z}_{1:n}}\big), \mathbb{E}\big(\hat{\mathrm{R}}_{\mathrm{Z}^{1,2}_{1:n}}(\mathrm{Z}_2)\big|\hat{\mathrm{f}}^{2,\mathrm{M}}_{\mathrm{Z}_{1:n}}\big)\Big) \overset{(a)}{=} \mathrm{var}\Big(\mathbb{E}\big(\mathrm{R}_{\mathrm{Z}_{1:n}}(\mathrm{Z}'_1)\big|\hat{\mathrm{f}}^{2,\mathrm{M}}_{\mathrm{Z}_{1:n}}\big)\Big)\end{aligned}$$

where (a) is obtained by noting that $\mathbb{E}\big(\hat{R}_{Z^{1,2}_{1:n}}(Z_1)\big|\hat{f}^{2,M}_{Z_{1:n}}\big) = \mathbb{E}\big(\hat{R}_{Z^{1,2}_{1:n}}(Z_2)\big|\hat{f}^{2,M}_{Z_{1:n}}\big)$.
Similarly we can also exploit the classical decomposition of the variance in terms of conditional variances to get,

$$\mathrm{var}\Big(\hat{\mathrm{R}}_{\mathrm{Z}_{1:n}}(\mathrm{Z}'_1)\Big) = \mathrm{var}\Big(\mathbb{E}\Big(\hat{\mathrm{R}}_{\mathrm{Z}_{1:n}}(\mathrm{Z}'_1)\Big|\mathrm{f}^{\mathrm{M}}_{\mathrm{n}}\Big)\Big) + \mathbb{E}\Big(\mathrm{var}\Big(\hat{\mathrm{R}}_{\mathrm{Z}_{1:n}}(\mathrm{Z}'_1)\Big|\hat{\mathrm{f}}^{2,\mathrm{M}}_{\mathrm{Z}_{1:n}}\Big)\Big)$$

Therefore as we have assumed that,

$$n\mathrm{var}\Big(\mathbb{E}\big(\hat{\mathrm{R}}_{\mathrm{Z}_{1:n}}(\mathrm{Z}'_1)\big|\hat{\mathrm{f}}^{2,\mathrm{M}}_{\mathrm{Z}_{1:n}}\big)\Big) \to \sigma_1^2$$

and

$$\mathbb{E}\Big(\mathrm{var}\Big(\hat{\mathrm{R}}_{\mathrm{Z}_{1:n}}(\mathrm{Z}'_1)\Big|\hat{\mathrm{f}}^{2,\mathrm{M}}_{\mathrm{Z}_{1:n}}\Big)\Big) \to \sigma_2^2$$

we easily can see that,

$$\frac{1}{n}\mathrm{var}(\mathrm{f_n}(\mathrm{Z})) \to \sigma^2$$

7.5 Proof of Corollary 4.5

Proof. Let (f_n) be the sequence of functions such that $f_n(X) = n\big[\mathbb{E}[f(S_k(X|_{[n]}))|X] - \mathbb{E}[f(S_k(X|_{[n]}))|\mathbb{S}_\infty]\big]$. According to Theorem 4.3 we know that we only need to prove the following three items, to obtain the desired central-limit theorem,

- The following is uniformly-integrable $(\Delta_i^2(f_n, X_n))_{i,n\in\mathbb{N}}$
- The following is true asymptotically $\frac{1}{n^2}\sum_{j\in\mathbb{N}}\Big[\sum_{i\in\mathbb{N}}\|\Delta_{i,j}(f_n, X_n)\|_{L_2}\Big]^2 \to 0.$
- The following is true asymptotically $\frac{1}{n}\mathrm{var}\big(\mathrm{f_n}(\mathrm{X_n})|\mathbb{S}_\infty\big) \xrightarrow{\mathrm{P}} \sigma^2$
- Proof of the first point

Note that for any positive $A \in \mathbb{R}_+$ the following holds

$$\begin{aligned}
&\Big\|\Delta_i(f_n, X)\mathbb{I}(|\Delta_i(f_n, X)| > A)\Big\|_{L_2} \\
&= n\Big\| \sum_{v_k \in \mathcal{V}_k([n])} \mathbb{I}(|\Delta_i(f_n, X)| > A)\ \Big[F_{nk}(X|_{v_k}, X) - \mathbb{E}\big[F_{nk}(X|_{v_k}, X)\big|\mathbb{S}_k(i)\big]\Big]\Big\|_{L_2} \\
&\leq 2n \sum_{\substack{v_k \in \mathcal{V}_k([n]) \\ i \in v_k}} \Big\|\mathbb{I}(|\Delta_i(f_n, X)| > A)\ F_{nk}(X|_{v_k}, X)\Big\|_{L_2} \\
&\quad + n \sum_{\substack{v_k \in \mathcal{V}_k([n]) \\ i \notin v_k}} \Big\|\mathbb{I}(|\Delta_i(f_n, X)| > A)\ \Delta_i(F_{nk}(X|_{v_k}, X)\Big\|_{L_2} \\
&\leq 2kn^k \sup_{v_k \in \mathcal{V}_k([x]_n)} \Big\|\mathbb{I}(|\Delta_i(f_n, X)| > A)F_{nk}(X|_{v_k}, X)\Big\|_{L_2} \\
&+ n^{k+1} \sup_{\substack{v_k \in \mathcal{V}_k([n]) \\ i \notin v_k}} \Big\|\mathbb{I}(|\Delta_i(f_n, X)| > A)\ \Delta_i(F_{nk}(X|_{v_k}, X))\Big\|_{L_2}
\end{aligned}$$

Therefore exploiting hypothesis (a_1) and (a_2) we get the desired result.

- Proof of the second point For all $i, j \in \mathbb{N}$ we have

$$\begin{aligned}
&\Big\|\Delta_{i,j}(f_n)\Big\|_{L_2} \\
&\leq n \sum_{\substack{v_k \in \mathcal{V}_k([n]) \\ i,j \in v_k}} \big\|\Delta_{i,j}(F_{nk}(X|_{v_k}, X)\big\|_{\mathbf{L_2}} + 2n \sum_{\substack{v_k \in \mathcal{V}_k([n]) \\ i \notin v_k, j \notin v_k}} \big\|\Delta_{i,j}(F_{nk}(X|_{v_k}, X)\big\|_{\mathbf{L_2}} \\
&\quad + n \sum_{\substack{v_k \in \mathcal{V}_k([n]) \\ i \notin v_k, j \in v_k}} \big\|\Delta_{i,j}(F_{nk}(X|_{v_k}, X)\big\|_{\mathbf{L_2}} \\
&\leq k^2 n^{k-1} \sup_{\substack{v_k \in \mathcal{V}_k([n]) \\ i,j \in v_k}} \big\|\Delta_{i,j}(F_{nk}(X|_{v_k}, X)\big\|_{\mathbf{L_2}} + 2kn^k \sup_{\substack{v_k \in \mathcal{V}_k([n]) \\ i \in v_k, j \notin v_k}} \big\|\Delta_{i,j}(F_{nk}(X|_{v_k}, X)\big\|_{\mathbf{L_2}} \\
&\quad + n^{k+1} \sup_{\substack{v_k \in \mathcal{V}_k([n]) \\ i,j \notin v_k}} \big\|\Delta_{i,j}(F_{nk}(X|_{v_k}, X)\big\|_{\mathbf{L_2}}
\end{aligned}$$

Therefore using hypothesis (b) we get the desired result.

- Proof of the third point Before presenting the key argument we note that

$$\frac{\mathrm{var}(\mathrm{f_n}(\mathrm{X_n})|\mathbb{S}_\infty)}{n} = n \sum_{v_k, v'_k \in \mathcal{V}_k([x]_n)} \mathrm{cov}\big(\mathrm{F_{nk}}(\mathrm{X}|_{\mathrm{v_k}}, \mathrm{X}), \mathrm{F_{nk}}(\mathrm{X}|_{\mathrm{v'_k}}, \mathrm{X})\big|\mathbb{S}_\infty\big).$$

The proof will be centered around two successive arguments. The first one will allow us to neglect the covariance $\mathrm{cov}\big(\mathrm{F_{nk}}(\mathrm{X}|_{\mathrm{v_k}}, \mathrm{X}), \mathrm{F_{nk}}(\mathrm{X}|_{\mathrm{v'_k}}, \mathrm{X})\big|\mathbb{S}_\infty\big)$ when $\mathrm{card}(\mathrm{v_k} \bigcap \mathrm{v'_k}) \neq 1$;

the second one will use the exchangeability of X to simplify the remaining covariances. First note that: $\mathrm{card}\{\mathrm{v_k}, \mathrm{v'_k} \in \mathcal{V}_\mathrm{k}([\mathrm{n}]) | \mathrm{card}(\mathrm{v_k} \bigcap \mathrm{v'_k}) > 1\} \leq \mathrm{k}^4 \mathrm{n}^{\mathrm{k}-2}$. Therefore we have

$$\Big| n \sum_{\substack{x_k, x'_k \in \mathcal{V}_k([n]) \\ \mathrm{card}(\mathrm{v_k} \bigcap \mathrm{v'_k}) > 1}} \mathrm{cov}\big(\mathrm{F_{nk}}(\mathrm{X}|_{\mathrm{v_k}}, \mathrm{X}), \mathrm{F_{nk}}(\mathrm{X}|_{\mathrm{v'_k}}, \mathrm{X}) | \mathbb{S}_\infty\big)\Big\|^2_{\mathrm{L_2}}$$
$$\leq k^4 n^{2k-1} \sup_{v_k \in \mathcal{V}_k([x]_n)} \big\| F_{nk}(X|_{v_k}, X)\big\|^2_{\mathbf{L_2}}$$

Moreover if we write:

$$Y_i^{v_k}(X) := \mathbb{E}[F_{nk}(X|_{v_k}, X) | \mathbb{S}^{\setminus i}] - \mathbb{E}[F_{nk}(X|_{v_k}, X) | \mathbb{S}^{\setminus i-1}]$$

then we have

$$\Big\| n \sum_{\substack{x_k, x'_k \in \mathcal{V}_k([n]) \\ \mathrm{card}(\mathrm{v_k} \bigcap \mathrm{v'_k}) = 0}} \mathrm{cov}\big(\mathrm{F_{nk}}(\mathrm{X}|_{\mathrm{v_k}}, \mathrm{X}), \mathrm{F_{nk}}(\mathrm{X}|_{\mathrm{v'_k}}, \mathrm{X}) | \mathbb{S}_\infty\big)\Big\|^2_{\mathrm{L_2}}$$
$$\leq n^{2k+1} \sup_{\substack{x_k, x'_k \in \mathcal{V}_k([n]) \\ \mathrm{card}(\mathrm{v_k} \bigcap \mathrm{v'_k}) = 0}} \big\|\mathrm{cov}\big(\mathrm{F_{nk}}(\mathrm{X}|_{\mathrm{v_k}}, \mathrm{X}), \mathrm{F_{nk}}(\mathrm{X}|_{\mathrm{v'_k}}, \mathrm{X}) | \mathbb{S}_\infty\big)\big\|_{\mathbf{L_2}}$$
$$\leq n^{2k+1} \sup_{\substack{x_k, x'_k \in \mathcal{V}_k([n]) \\ \mathrm{card}(\mathrm{v_k} \bigcap \mathrm{v'_k}) = 0}} \sum_{i,j \leq n} \big\|\mathrm{cov}\big(\mathrm{Y_i^{v_k}}(\mathrm{X}), \mathrm{Y_j^{v'_k}}(\mathrm{X}) | \mathbb{S}_\infty\big)\big\|_{\mathbf{L_2}}$$
$$\leq n^{2k+1} \sup_{\substack{x_k, x'_k \in \mathcal{V}_k([n]) \\ \mathrm{card}(\mathrm{v_k} \bigcap \mathrm{v'_k}) = 0}} \sum_{i,j \leq n} \big\|\mathrm{cov}\big(\mathrm{Y_i^{v_k}}(\mathrm{X}) - \mathbb{E}[\mathrm{Y_i^{v_k}}(\mathrm{X}) | \mathbb{S}_\mathrm{n}(\mathrm{j})], \mathrm{Y_j^{v'_k}}(\mathrm{X}) - \mathbb{E}[\mathrm{Y_j^{v'_k}}(\mathrm{X}) | \mathbb{S}_\mathrm{n}(\mathrm{i})] | \mathbb{S}_\infty\big)\big\|_{\mathbf{L_2}}$$
$$\leq k^2 n^{2k+1} \sum_{i,j \leq n} \sup_{\substack{v_k \in \mathcal{V}_k([n]) \\ i,j \notin v_k}} \|\Delta_{i,j}(F_{nk}(X|_{v_k}, X)\big\|^2_{\mathbf{L_2}} \to 0.$$

Therefore we know that:

$$\Big| \frac{1}{n} \mathrm{var}(\mathrm{f_n}(\mathrm{X_n}) | \mathbb{S}_\infty) - \mathrm{n} \sum_{\substack{x_k, x'_k \in \mathcal{V}_k([n]) \\ \mathrm{card}(\mathrm{v_k} \bigcap \mathrm{v'_k}) = 1}} \mathrm{cov}\big(\mathrm{F_{nk}}(\mathrm{X}|_{\mathrm{v_k}}, \mathrm{X}), \mathrm{F_{nk}}(\mathrm{X}|_{\mathrm{v'_k}}, \mathrm{X}) | \mathbb{S}_\infty\big)\Big| = \mathrm{o}(1).$$

Moreover as X is a exchangeable for all $v_k = (v_1, \dots, v_k), v'_k = (v'_1, \dots, v'_k) \in \mathcal{V}_k([n])$ such that $\{v_1, \dots, v_k\} \bigcap \{v'_1, \dots, v'_k\} = \{i\} = \{v'_m\} = \{v_{m'}\}$ we get that

$$\mathrm{cov}\big(\mathrm{F_{nk}}(\mathrm{X}|_{\mathrm{v_k}}, \mathrm{X}), \mathrm{F_{nk}}(\mathrm{X}|_{\mathrm{v'_k}}, \mathrm{X}) | \mathbb{S}_\infty\big)$$
$$= \mathrm{Cov}\big[F_{nk}(X|_{[k]}, X), F_{nk}(X|_{(k+1, \dots, k+m-1, m', k+m+1, \dots, 2k)}, X) | \mathbb{S}_\infty\big]$$

Therefore we have

$$n \sum_{m,m' \leq k} \sum_{i \in \mathbb{N}} \sum_{\substack{v_k, v'_k \in \mathcal{V}_k([n]) \\ \{v_1,\dots,v_k\} \bigcap \{v'_1,\dots,v'_k\} = \{i\} \\ v_m = u_{m'} = i}} \mathrm{cov}\big(\mathrm{F_{nk}}(\mathrm{X}|_{\mathrm{v_k}}, \mathrm{X}), \mathrm{F_{nk}}(\mathrm{X}|_{\mathrm{v'_k}}, \mathrm{X}) | \mathbb{S}_\infty\big)$$
$$= \sum_{i,j \leq k} n^{2k} \mathrm{Cov}\big[F_{nk}(X|_{[k]}, X), F_{nk}(X|_{(k+1,\dots,k+i-1,j,k+i+1,\dots,2k)}, X) | \mathbb{S}_\infty\big] \to \eta^2.$$

Therefore we obtain that $\frac{\mathrm{var}(\mathrm{f_n}(\mathrm{X_n})|\mathbb{S}_\infty)}{n} \to \sigma^2$.

□

7.6 Proof for cross-validation

Proof. The proof for the K_n-fold cross validated risk and for $\hat{R}_{1,n}$ are very similar. However it is slightly more complex for the former and we will present this case. We choose: $f_n(X) = n\hat{R}_{\mathrm{cross,n}}(X_n)$. According to Theorem 4.3 we know that we only need to prove the following two items, to obtain the desired central-limit theorem,

- The following is uniformly-integrable $(\Delta_i^2(f_n, Z))_{i,n \in \mathbb{N}}$
- The following is true asymptotically $\frac{1}{n^2} \sum_i \Big[\sum_j \|\Delta_{i,j}(f_n, Z)\|_{\mathbf{L}_2} \Big]^2 \to 0$.

The key point to prove each one of those points will be to exploit the stability property of the predictors. To make the presentation simpler we write: $\mathcal{B}(i) \in [K_n]$ the index such that $i \in B^n_{\mathcal{B}(i)}$. Moreover throughout we will abbreviate $i_1, \dots, i_k$ by $i_{1:k}$.

- Proof of the first point We will break $|\Delta_i(f_n, Z)|$ into more manageable parts and study each term successively. In this goal, we introduce the following notation:

$$\Delta_i^{i_{1:k}}(\mathcal{L}, Z) := \mathcal{L}\big(f_{m^n_{\mathcal{B}(i)}}\big([Z]_{[n]\setminus B^n_{\mathcal{B}(i)}}, [X]_{i_{1:k}}\big), [Y]_{i_{1:k}}\big)\big)$$
$$- \mathbb{E}\Big(\mathcal{L}\big(f_{m^n_{\mathcal{B}(i)}}\big([Z]_{[n]\setminus B^n_{\mathcal{B}(i)}}, [X]_{i_{1:k}}\big), [Y]_{i_{1:k}}\big)\big) | \mathbb{S}_n(i)\Big).$$

Using the triangle inequality we obtain that:

$$
\begin{aligned}
&\Big|\Delta_i(f_n, Z)\Big| \\
&\leq \frac{n}{K_n|\mathcal{S}_k(B^n_{\mathcal{B}(i)})|} \sum_{\substack{(i_1,\dots,i_k)\in\mathcal{S}_k(B^n_{\mathcal{B}(i)}) \\ \exists l\leq k \text{ s.t } \mathrm{i_l=i}}} \Big|\Delta_i^{i_{1:k}}(\mathcal{L}, Z)\Big| \\
&+ \sum_{j\neq\mathcal{B}(i)} \Big|\frac{n}{K_n|\mathcal{S}_k(B^n_j)|} \sum_{(i_1,\dots,i_k)\in\mathcal{S}_k(B^n_j)} \mathcal{L}\big(f_{m^n_j}\big([Z]_{[n]\setminus B^n_j}, [X]_{i_{1:k}}\big), [Y]_{i_{1:k}}\big)\big) \\
&\qquad\qquad - \mathbb{E}\Big(\mathcal{L}\big(f_{m^n_j}\big([Z]_{[n]\setminus B^n_j}, [X]_{i_{1:k}}\big), [Y]_{i_{1:k}}\big)\big)|\mathbb{S}_n(i)\Big)\Big| \\
&\leq a_{i,n} + b_{i,n}
\end{aligned}
\tag{7.25}
$$

To prove that $(\Delta_i(f_n, Z)^2)$ is uniformly integrable it is enough to prove that the families $(a^2_{i,n})$ and $(b^2_{i,n})$ are uniformly integrable. We therefore study each family separately.

Using (H0) and the joint exchangeability of Z we obtain that the family $\Big(\mathcal{L}\big(f_{m^n_{\mathcal{B}(i)}}\big([Z]_{[n]\setminus B^n_{\mathcal{B}(i)}}, [X]_{i_{1:k}}\big), [Y]_{i_{1:k}}\big)^2\Big)$ is uniformly integrable. Moreover note that as $\sup_{i,j}\big||B^n_i| - |B^n_j|\big| \leq 1$ we have

$$
\sup_{i\leq K_n, n\in\mathbb{N}} \frac{n}{K_n|\mathcal{S}_k(B^n_i)|} \mathrm{card}\Big(\{(\mathrm{i_1},\dots,\mathrm{i_k})\in\mathcal{S}_\mathrm{k}(\mathrm{B^n_i})|\exists \mathrm{l}\leq \mathrm{k \ s.t\ i_l = i}\}\Big) \leq 2\mathrm{k}.
$$

Therefore the family $(a^2_{i,n})$ is uniformly integrable.

To prove that the family $(b^2_{i,n})$ is uniformly integrable we separate it into more manageable pieces. In this goal we introduce the following notation:

$$
H^j_{i_{1:k}} := \mathcal{L}\big(f_{m^n_j}\big([Z]_{[n]\setminus B^n_j}, [X]_{i_{1:k}}\big), [Y]_{i_{1:k}}\big)\big) - R_{m^n_j}(f_{m^n_j}, [Z]_{[n]\setminus B^n_j}),
$$

this allows us to write

$$
\begin{aligned}
b_i, n &\leq \frac{n}{K_n} \sum_{j\leq K_n} \Big|R_{m^n_j}(f_{m^n_j}, [Z]_{[n]\setminus B^n_j}) - \mathbb{E}\big(R_{m^n_j}(f_{m^n_j}, Z_{[n]\setminus B^n_j})|\mathbb{S}_n(i)\big)\Big| \\
&+ \sum_{j\neq\mathcal{B}(i)} \Big|\frac{n}{K_n|\mathcal{S}_k(B^n_j)|} \sum_{(i_1,\dots,i_k)\in\mathcal{S}_k(B^n_j)} H^j_{i_{1:k}} - \mathbb{E}\big(H^j_{i_{1:k}}\big|\mathbb{S}_n(i)\big)\Big| \\
&\leq c_{i,n} + d_{i,n}.
\end{aligned}
$$

Hypothesis (H1) guarantees that $(c^2_{i,n})$ is uniformly integrable.

We need to prove that the family $(d^2_{i,n})$ is uniformly integrable, which is true if the family $\big(\big[\frac{n}{|\mathcal{S}_k(B^n_j)|}\sum_{(i_1,\dots,i_k)\in\mathcal{S}_k(B^n_j)} H^j_{i_{1:k}} - \mathbb{E}(H^j_{i_{1:k}}|\mathbb{S}_n(i))\big]^2\big)_{j\neq\mathcal{B}_i, n\in\mathbb{N}}$ is uniformly integrable.

A key point to prove this is to notice that

$$H^j_{i_{1:k}} \perp\!\!\!\perp_{[Z]_{[n]\setminus B^n_j}} H^j_{j_1,\ldots,j_k}, \qquad \text{if } \{\mathrm{i}_{1:\mathrm{k}}\} \bigcap \{\mathrm{j}_1,\ldots,\mathrm{j}_\mathrm{k}\} = \emptyset. \tag{7.26}$$

Introduce for all $K \in \mathbb{R}_+$ the following truncated random variable:

$$\bar{H}^{i,K}_{i_{1:k}} := \mathbb{I}(\sqrt{nK_n}|H^j_{i_{1:k}} - \mathbb{E}(H^j_{i_{1:k}}|\mathbb{S}_n(i))| \leq K)\ \big[H^j_{i_{1:k}} - \mathbb{E}(H^j_{i_{1:k}}|\mathbb{S}_n(i))\big],$$

and

$$\bar{H}^{i,\geq K}_{i_{1:k}} := \mathbb{I}(\sqrt{nK_n}|H^j_{i_{1:k}} - \mathbb{E}(H^j_{i_{1:k}}|\mathbb{S}_n(i))| \geq K)\ \big[H^j_{i_{1:k}} - \mathbb{E}(H^j_{i_{1:k}}|\mathbb{S}_n(i))\big].$$

Then using Equation (7.26) we obtain that there is a constant $C \in \mathbb{R}$ such that

$$\begin{aligned}
&\Big\|\frac{n}{|\mathcal{S}_k(B^n_j)|}\sum_{(i_1,\ldots,i_k)\in\mathcal{S}_k(B^n_j)} \bar{H}^{i,K}_{i_{1:k}} - \mathbb{E}(\bar{H}^{i,K}_{i_{1:k}}|Z_{[n]\setminus B^n_j})\Big\|^4_{\mathbf{L}_4}\\
&\leq \frac{k^4n^4}{[|B^n_j|-k]^2}\Big\|\bar{H}^{i,K}_{i_{1:k}} - \mathbb{E}(\bar{H}^{i,K}_{i_{1:k}}|Z_{[n]\setminus B^n_j})\Big\|^4_{\mathbf{L}_4}\\
&\leq \frac{Ck^4n^2K^4}{K_n^2[|B^n_j|-k]^2}
\end{aligned}$$

Therefore

$$\sup_{j,n}\Big\|\frac{n}{|\mathcal{S}_k(B^n_j)|}\sum_{(i_1,\ldots,i_k)\in\mathcal{S}_k(B^n_j)} \bar{H}^{i,K}_{i_{1:k}} - \mathbb{E}(\bar{H}^{i,K}_{i_{1:k}}|Z_{[n]\setminus B^n_j})\Big\|_{\mathbf{L}_4} < \infty.$$

This implies that $(b^2_{i,n})$ is uniformly integrable if

$$\lim_{K\to\infty}\sup_{\substack{n\in\mathbb{N}\\ j\leq K_n}}\Big\|\frac{n}{|\mathcal{S}_k(B^n_j)|}\sum_{(i_1,\ldots,i_k)\in\mathcal{S}_k(B^n_j)} \bar{H}^{i,\geq K}_{i_{1:k}} - \mathbb{E}(\bar{H}^{i,\geq K}_{i_{1:k}}|Z_{[n]\setminus B^n_j})\Big\|^2_{\mathbf{L}_2} \to 0.$$

Using Equation (7.26) we write:

$$\begin{aligned}
&\Big\|\frac{n}{|\mathcal{S}_k(B^n_j)|}\sum_{(i_1,\ldots,i_k)\in\mathcal{S}_k(B^n_j)} \bar{H}^{i,\geq K}_{i_{1:k}} - \mathbb{E}(\bar{H}^{i,\geq K}_{i_{1:k}}|Z_{[n]\setminus B^n_j})\Big\|^2_{\mathbf{L}_2}\\
&\leq \frac{k^2n}{K_n|m^n_j-k|}\sup_{\substack{i\leq n\\ i_1,\ldots,i_k}} nK_n\Big\|\bar{H}^{i,\geq K}_{i_{1:k}} - \mathbb{E}(\bar{H}^{i,\geq K}_{i_{1:k}}|Z_{[n]\setminus B^n_j})\Big\|^2_{\mathbf{L}_2}\\
&\leq \frac{2k^2n}{K_n|m^n_j-k|}\sup_{i\leq n} nK_n\Big\|\Delta_{i,n}(\mathcal{L},Z)\Big\|^2_{\mathbf{L}_2}.
\end{aligned}$$

This implies the desired result.

- Proof of the second point To prove this point we will decompose $\Delta_{i,j}(f_n, X_n)$ into more manageable parts. To so we will use the decomposition of $\Delta_i(f_n, X_n)$ introduced in Equation (7.25) and obtain:

$$\begin{aligned}
&\Big|\Delta_{i,j}(f_n, Z)\Big| \\
&\leq \frac{n}{K_n|\mathcal{S}_k(B^n_{\mathcal{B}(i)})|}\mathbb{I}(\mathcal{B}_j = \mathcal{B} - i)\sum_{\substack{(i_1,\dots,i_k)\in\mathcal{S}_k(B^n_{\mathcal{B}(i)})\\ \exists l\leq k \text{ s.t } \mathrm{i_l=i}}}\Big|\Delta_i^{i_{1:k}}(\mathcal{L}, Z) - \mathbb{E}\big(\Delta_i^{i_{1:k}}(\mathcal{L}, Z)\big|\mathbb{S}_n(j)\big)\Big| \\
&+ \frac{n}{K_n}\sum_{l\leq K_n}\Big|R_{m_l^n}(f_{m_l^n}, [Z]_{[n]\backslash B_l^n}) - \mathbb{E}\big(R_{m_l^n}(f_{m_l^n}, Z_{[n]\backslash B_l^n})|\mathbb{S}_n(i)\big) \\
&\qquad - \mathbb{E}\big(R_{m_l^n}(f_{m_l^n}, Z_{[n]\backslash B_l^n})|\mathbb{S}_n(j)\big) + \mathbb{E}\big(R_{m_l^n}(f_{m_l^n}, Z_{[n]\backslash B_l^n})|\mathbb{S}_n(i,j)\big)\Big| \\
&+ \sum_{\substack{l\neq\mathcal{B}_i\, l\neq\mathcal{B}_i\\ l\neq\mathcal{B}_j l\neq\mathcal{B}_j}}\Big|\frac{n}{K_n|\mathcal{S}_k(B^n_l)|}\sum_{(i_1,\dots,i_k)\in\mathcal{S}_k(B^n_l)} H^l_{i_{1:k}} - \mathbb{E}\big(H^l_{i_{1:k}}|\mathbb{S}_n(i)\big)\big) \\
&\qquad - \mathbb{E}\big(H^l_{i_{1:k}}|\mathbb{S}_n(j)\big) + \mathbb{E}\big(H^l_{i_{1:k}}|\mathbb{S}_n(i,j)\big)\Big| \\
&\leq a_{i,j,n} + b_{i,j,n} + c_{i,j,n}
\end{aligned} \tag{7.27}$$

We will bound the norm each term successively. Firstly we have

$$\begin{aligned}
&\Big\|a_{i,j,n}\Big\|_{\mathbf{L}_2} \\
&\leq \frac{n\mathbb{I}(\mathcal{B}_j = \mathcal{B}_i)}{K_n|\mathcal{S}_k(B^n_{\mathcal{B}(i)})|}\sum_{\substack{(i_1,\dots,i_k)\in\mathcal{S}_k(B^n_{\mathcal{B}(i)})\\ \exists l\leq k \text{ s.t } \mathrm{i_l=i}}}\Big\|\Delta_i^{i_{1:k}}(\mathcal{L}, Z) - \mathbb{E}\big(\Delta_i^{i_{1:k}}(\mathcal{L}, Z)\big|\mathbb{S}_n(j)\big)\Big\|_{\mathbf{L}_2} \\
&\leq \frac{2n\ \mathrm{card}\{(\mathrm{i_1},\dots,\mathrm{i_k})\in\mathcal{S}_\mathrm{k}(\mathrm{B^n_{\mathcal{B}(i)}})|\exists \mathrm{l}\leq\mathrm{k \text{ s.t } i_l = i}\}}{K_n|\mathcal{S}_k(B^n_{\mathcal{B}(i)})|}\mathbb{I}(\mathcal{B}_j = \mathcal{B}_i)\max_{i_{1:k}}\Big\|\Delta_i^{i_{1:k}}(\mathcal{L}, Z)\Big\|_{\mathbf{L}_2} \\
&\leq 2k\mathbb{I}(\mathcal{B}_j = \mathcal{B}_i)\frac{n}{n-k}\big\|\Delta_{i,n}(\mathcal{L}, Z)\big\|_{\mathbf{L}_2}.
\end{aligned}$$

Moreover it is straightforward to see that

$$\|b_{i,j,n}\|_{\mathbf{L}_2} \leq n\big\|\Delta_{i,j,n}(R_n, Z)\big\|_{\mathbf{L}_2}.$$

Now that we successfully bounded $\Big\|a_{i,j,n}\Big\|_{\mathbf{L}_2}$ and $\|b_{i,j,n}\|_{\mathbf{L}_2}$ we move on to upper-bounding the last term. We define

$$\mathcal{C}_l^{\neq} = \mathrm{card}\{(\mathrm{i_1},\dots,\mathrm{i_k}),(\mathrm{j_1},\dots,\mathrm{j_k})\in\mathcal{S}_\mathrm{k}(\mathrm{B^n_l})|\exists \mathrm{l_1},\mathrm{l_2}\leq \mathrm{k \text{ s.t } i_{l_1} = i_{l_2}}\}$$

then using Equation (7.26) we write

$$
\begin{aligned}
\|c_{i,j,n}\|_{\mathbf{L}_2} &\le \frac{1}{K_n}\sum_{\substack{l\neq\mathcal{B}_i\\ l\neq\mathcal{B}_j}}\Big\|\frac{n}{|\mathcal{S}_k(B_l^n)|}\sum_{(i_1,\dots,i_k)\in\mathcal{S}_k(B_l^n)} H^l_{i_{1:k}}-\mathbb{E}\big(H^l_{i_{1:k}}|\mathbb{S}_n(i)\big)\big) \\
&\qquad\qquad - \mathbb{E}\big(H^l_{i_{1:k}}|\mathbb{S}_n(j)\big)+\mathbb{E}\big(H^l_{i_{1:k}}|\mathbb{S}_n(i,j)\big)\Big\|_{\mathbf{L}_2} \\
&\le \sup_{\substack{l\neq\mathcal{B}_i\\ l\neq\mathcal{B}_j}}\frac{n\sqrt{\mathcal{C}_l^{\neq}}}{|\mathcal{S}_k(B_l^n)|}\Big\|\Delta_{i,j,n}(\mathcal{L},Z)\Big\|_{\mathbf{L}_2} \\
&\le \frac{knK_n}{\sqrt{n-k}}\Big\|\Delta_{i,j,n}(\mathcal{L},Z)\Big\|_{\mathbf{L}_2}
\end{aligned}
$$

Therefore we have

$$
\begin{aligned}
&\frac{1}{n^2}\sum_{i\le n}\Big[\sum_{j\le n}\|\Delta_{i,j}(f_n,X_n)\|_{\mathbf{L}_2}\Big]^2 \\
&\le 3\Big[\frac{k^2K_n^2}{n-k}\sum_{i\le n}\Big[\sum_{j\le n}\Big\|\Delta_{i,j,n}(\mathcal{L},Z)\Big\|_{\mathbf{L}_2}\Big]^2+\sum_{i\le n}\Big[\sum_{j\le n}\big\|\Delta_{i,j,n}(R_n,Z)\big\|_{\mathbf{L}_2}\Big]^2 \\
&\quad + 4k^2\frac{n^3}{K_n^2(n-k)^2}\sup_i\big\|\Delta_{i,n}(\mathcal{L},Z)\big\|^2_{\mathbf{L}_2}\Big]\to 0.
\end{aligned}
$$

□

Chapter 8

Proofs examples in information theory

8.1 Proof of Theorem 5.1

PROOF. Since the group is countable, we can define an order $\prec$ on $\mathbb{G}$ by enumerating the elements of $\mathbf{A}_n$ as $\phi_1^n, \phi_2^n \ldots$, and choosing the order such that $\phi_{i-1}^n \prec \phi_i^n$. For the process (S_ϕ), define the σ-algebras

$$\mathcal{T}_n(\phi) := \sigma\{S_{\phi'} \,|\, \phi' \in \mathbf{A}_n, \phi' \prec \phi\} \quad \text{and} \quad \mathcal{T}(\phi) := \sigma\{S_{\phi'} \,|\, \phi' \prec \phi\} \;.$$

With these in hand, we define functions

$$\begin{aligned} f_n(S,\phi) :=& \; \log P(S_\phi | \mathcal{T}_n(\phi)) - \mathbb{E}[\log P(S_\phi | \mathcal{T}_n(\phi))] \\ g_m(S,\phi) :=& \; \log P(S_\phi | \mathcal{T}(\phi) \cap \mathbf{B}_m) - \mathbb{E}[\log P(S_\phi | \mathcal{T}(\phi) \cap \mathbf{B}_m)\big] \end{aligned}$$

An application of the chain rule then yields

$$\tfrac{1}{\sqrt{|\mathbf{A}_n|}}\big(\log P(S_{\mathbf{A}_n}) - \mathbb{E}[\log P(S_{\mathbf{A}_n})]\big) = \tfrac{1}{\sqrt{|\mathbf{A}_n|}} \textstyle\sum_{\phi \in \mathbf{A}_n} f(S,\phi) \;.$$

Now consider a ϕ such that $\mathcal{T}_n(\phi) \cap \mathbf{B}_m = \mathcal{T}(\phi) \cap \mathbf{B}_m$. Then

$$\|f_n(S,\phi) - g_m(S,\phi)\|_{L_2} \leq \rho_m \;.$$

The number of $\phi \in \mathbf{A}_n$ for which that is *not* the case is

$$|\{\phi \in \mathbf{A}_n \,|\, \mathcal{T}_n(\phi) \cap \mathbf{B}_m \neq \mathcal{T}(\phi) \cap \mathbf{B}_m\}| \;\leq\; |\mathbf{A}_n \triangle \mathbf{B}_m \mathbf{A}_n|$$

Denote $M_p := \sup_{\phi \in \mathbb{G}, A \subset \mathbb{G}} \|\log P(X_\phi | X_A)\|_{\mathbf{L}_p}$. Then for any $k \in \mathbb{N}$, and any $\phi, \phi' \in \mathbb{G}$ that satisfy $d(\phi, \phi') \geq i$, we have

$$\mathrm{Cov}\big[f_n(S,\phi) - g_m(S,\phi), f_n(S,\phi') - g_m(S,\phi')\big] \leq 4\rho_m \big[\rho_k + 4M_{2+\varepsilon}^2 \alpha^{\frac{\varepsilon}{2+\varepsilon}}(i-k, |\mathbf{B}_m|)\big] \ .$$

Therefore for any sequence (b_n) satisfying $\frac{|\mathbf{A}_n \triangle \mathbf{B}_{b_n} \mathbf{A}_n|}{|\mathbf{A}_n|} \to 0$ and $b_n \to \infty$ we have

$$\frac{1}{\sqrt{|\mathbf{A}_n|}} \sum_{\phi \in \mathbf{A}_n} f_n(S,\phi) - g_{b_n}(S,\phi) \xrightarrow{L_2} 0 \ .$$

Moreover, if α^m denotes the mixing coefficient of g_m, then $\alpha^m(i) \leq \rho_k + \alpha(i - 2m, |\mathbf{B}_m|)$. Theorem 3.5 hence implies

$$\frac{\sum_{\phi \in \mathbf{A}_n} g_m(\phi X)}{\sqrt{|\mathbf{A}_n|}} \xrightarrow{d} \eta_m Z \quad \text{ for } \quad \eta_m^2 := \sum_{\phi} \mathrm{Cov}[g_m(X), g_m(\phi X)] \ ,$$

and since $\eta_m \xrightarrow{m \to \infty} \eta$, the result follows. □

8.2 Proofs for Kolmogorov complexity.

For notational convenience throughout we will denote $X_{i:k}$ the random vector $(X_1, \ldots, X_k)$.

8.2.1 Useful lemmas

There are two simple results on the Kolmogorov complexity that we employ in our proofs. We mention these two as simple lemmas that we can refer to later in the proofs of our main results. For the proof of these results a reader may refer to [18], Chapter 14 (Example 14.2.7 and Theorem 14.2.4)

Lemma 8.1. *Let n denote an integer number. Then, we have the following upper-bound on the Kolmogorov complexity of n:*

$$K(n) \leq \log^*(n) + c,$$

where

$$\log^*(x) = \begin{cases} 0 & x \leq 1 \\ 1 + \log^*(\log(x)) & x > 1, \end{cases} \tag{8.1}$$

and where c is a constant that depends only on the universal machine.

It is straightforward to show that $\forall n \geq 1, \log^*(n) < 2\log(n) + 2$. Another result that will be used about the Kolmogorov complexity in our section is the following:

Lemma 8.2. *If* $C_v \triangleq \{x \in \bigcup_{i=1}^{\infty}\{0,1\}^i \mid K(x) < v\}$, *then* $|C_v| \leq 2^v$.

The following lemma is a consequence of Theorem 3.3 and will be useful in the proof of the central limit theorem.

Lemma 8.3. *Let* (X^n) *be a sequence of stationary ergodic processes. Let* $(\alpha_n(i))$ *be the alpha-mixing coefficients of* X^n. *Suppose there is* $\beta > 1$ *such that* $\sup_n \left\|X_1^n\right\|_{\frac{8\beta}{\beta-1}} < \infty$*; and such that there is* $K > 0$ *such that*

$$\alpha_n(i) \leq \min(K(i - K\log(n))^{-\beta}, 2).$$

Then if there is $\sigma \in \mathbb{R}$ *such that* $\sigma_n^2 := \mathrm{var(X_1^n)} \to \sigma^2$ *we have*

$$\frac{1}{\sqrt{n}}\sum_{i\leq n} X_i^n \xrightarrow{d} N(0,\sigma^2).$$

Proof. If $\sigma^2 > 0$, we want to use Theorem 3.3. Write $Z \sim N(0,1)$ to be a standard normal and set $\epsilon := \frac{2(\beta+1)}{\beta-1}$ then there is $\tilde{K} < \infty$ such that for all $b > 0$

$$\sum_{i\geq b} \alpha_n^{\frac{\epsilon}{2+\epsilon}}(i) \leq \begin{cases} \tilde{K}(i-b)^{\frac{\beta-1}{2}} + K\log(n) \text{ if } \mathrm{b \leq K\log(n)} \\ \tilde{K}(i-b)^{\frac{\beta-1}{2}} \text{ otherwise} \end{cases},$$

and $\sup_n \|X_1^n\|_{4+2\epsilon} < \infty$. Set the sequence (b_n) in Theorem 3.3 to be $b_n = n^{-\frac{1}{4}}$ we obtain

$$d_W\Big(\frac{1}{\sqrt{n}}\sum_{i\leq n} X_i^n, \sigma^2 Z\Big) = O(\max(n^{-\frac{\beta-1}{4}}, n^{-\frac{1}{4}}\log(n), |\sigma_n^2 - \sigma^2|) \to 0.$$

If $\sigma^2 = 0$ then we obtain the desired result by a simple Markov inequality. □

8.2.2 Proof of Remark

For $n \in \mathbb{N}$, consider the two vectors $x, x' \in A^n$ such that $d_n(x,x') \leq 1$. If $d_n(x,x') = 0$, then we can easily see that $|\log(P(x) - \log(P(x'))| = 0$. Hence, we assume that $d_n(x,x') = 1$. Suppose that $x_i \neq x_i'$. If $i \in [2, |n-m-1|]$, then

$$\begin{aligned}
&|\log(P(x_{1:n}) - \log(P(x'_{1:n}))| \\
&\leq |\log(P(x_{1:i-1}) - \log(P(x'_{1:i-1}))| + \sum_{j=i}^{m+1+i} |\log(P(x_j|x_{j-1:j-m})) - \log(P(x'_j|x'_{j-1:j-m}))| \\
&\quad + |\log(P(x_{m+1+i:n}|x_{i+1:m+i})) - \log(P(x'_{m+1+i:n}|x'_{i+1:m+i}))| \\
&= \sum_{j=i}^{m+1+i} |\log(P(x_j|x_{j-1:j-m})) - \log(P(x'_j|x'_{j-1:j-m}))| \leq -(m+1)\log(\rho).
\end{aligned}$$

This comes from the following facts: (i) For every $j < i$, $x_{1:j} = x'_{1:j}$. Hence, $|\log(P(x_{1:i-1}) - \log(P(x'_{1:i-1}))| = 0$, (ii) For every $j > m+1+i$, $x_{j:n} = x'_{j:n}$. Hence, $|\log(P(x_{m+1+i:n}|x_{i+1:m+i})) - \log(P(x'_{m+1+i:n}|x'_{i+1:m+i}))| = 0$. (iii) Finally, $\forall i \in [|i, m+1+i|]$ $|\log(P(x_j|x_{j-1:j-m})) - \log(P(x'_j|x'_{j-1:j-m}))| \leq -\log(\rho)$. The proof for $i \notin [2, |n-m-1|]$ is similar and is hence skipped.

8.2.3 Proof of Theorem 5.4

8.2.3.1 Lower bound

PROOF. Before we discuss the details of the proof, we give a brief overview of the proof strategy to help the reader navigate through the proof more easily. Consider the sequence $X_1, X_2, \ldots, X_n$ with $X_i \in A$, for all i. We assume that $|A| = l$. In this section, we first present a simple program that a universal computer can use to generate this sequence.

Define $m_n \triangleq \frac{\frac{1}{2}-\epsilon}{\log(l)} \log(n)$, where $\frac{1}{2} - \frac{1}{2C} > \epsilon > 0$. Note that C is the same constant as the one used in Condition 3 in the statement of the theorem. The program first tells the universal computer the first m_n bits in the sequence. Then, counts the number of times each $(m_n+1)-$tuple is present in the remaining sequence and reports it.[1] In other words, if we define

$$f_j^{m_n,n} \triangleq \frac{\sum_{k=m_n+1}^{n} \mathbb{I}_{X_{k-m_n:k}=a_j^{m_n}}}{n-m_n}, \tag{8.2}$$

where $a_j^{m_n}$ is the j^{th} element (in a specific order that is described to the universal computer) of A^{m_n+1}, then the numbers $f_j^{m_n,n}$ are described to the universal computer. Let $\mathbf{f}^{m_n,n}$ denote the vector of all the empirical counts, i.e.,

$$\mathbf{f}^{m_n,n} \triangleq (f_1^{m_n,n}, f_2^{m,n}, \ldots, f_{l^{m_n+1}}^{m_n,n}).$$

Define an operator $O_f : A^n \to [0,1]^{l^{m_n+1}}$ that takes $X_1, X_2, \ldots, X_n$ as input and returns $\mathbf{f}^{m_n,n}$ as its output. Then, define the type of a sequence $X_{1:n}$ as the following set:

$$\mathcal{T}_{X_{1:n}} \triangleq \{Z_{1:n} \ : \ O_f(X_{1:n}) = O_f(Z_{1:n}) \text{ and } Z_{1:m_n} = X_{1:m_n}\}.$$

Given the information known to the universal computer so far, it has already access to $\mathcal{T}_{X_{1:n}}$. The only remaining piece of information that the universal computer should have to reconstruct the entire sequence is the index of the sequence $X_1, X_2, \ldots, X_n$ among all the sequences in its type. Let's count the number of bits we have used so far to describe the sequence.

Our description requires bits to specify the following quantities: (i)m_n, (ii)each a_j, (iii) the first m_n bits, (iv)the frequency of observing each possible block of length (m_n+1) in

[1]For instance, if $m_n = 1$, then for the sequence 01001 the couple $(0,1)$ is present twice, the couple $(1,0)$ once and $(0,0)$ once.

$X_{1:n}$, (v) a systematic way to build all the sequences of length n in $\mathcal{T}_{X_{1:n}}$, (vi) the index of $X_{1:n}$ in $\mathcal{T}_{X_{1:n}}$.

(i) $K(m_n) \leq \log^*(m_n) + c$.

(ii) To describe each a_j at most $l \max_{j \leq l} K(a_j)$ are required.

(iii) To describe the first m_n symbols we require $m_n(\log^*(l) + c)$.

(iv) To describe the frequency of each block we require $l^{m_n+1} \log^*(n)$ bits. The reason is clear, there are l^{m_n+1} different l-ary blocks of length $m_n + 1$. Each of them can have at most n elements in them.

(iv) So far the universal computer has detected $\mathcal{T}_{X_{1:n}}$. Now we should describe which element of $\mathcal{T}_{X_{1:n}}$ $X_{1:n}$ is. As the first step we write a constant size program so that the universal computer realizes what ordering of sequences we are using. The next step is to specify the index of our sequence in this list. To evaluate the number of bits required for describing the index we count the number of elements in $\mathcal{T}_{X_{1:n}}$.

Define $\tilde{P}^{m_n}$ as a new measure on $X_1, X_2, \ldots, X_n$ that has the following properties:

1. $\tilde{P}^{m_n}$ has the m_n-Markov property, i.e.,

$$\tilde{P}^{m_n}(X_{1:n}) = \tilde{P}^{m_n}(X_{1:m_n}) \prod_{i=m_n+1}^{n} \tilde{P}^{m_n}(X_i \mid X_{i-m_n:i-1}).$$

2. The $m_n + 1^{\text{th}}$-dimension transition probabilities are the same as those of the original distribution P, i.e.,

$$\tilde{P}^{m_n}(X_i \mid X_{i-1}, \ldots, X_{i-m_n}) = P(X_i \mid X_{i-1}, \ldots, X_{i-m_n}).$$

For notational simplicity we consider the notation

$$Q_j^{m_n} \triangleq \tilde{P}^{m_n}(X_{m_n} = a_{j,m_n}^{m_n} | X_0 = a_{j,0}^{m_n}, \ldots, X_{m_n-1} = a_{j,m_n-1}^{m_n}), \tag{8.3}$$

where $(a_{j,0}^{m_n}, \ldots, a_{j,m_n}^{m_n})$ is the j^{th} element of A^{m_n+1}. With this new notation we count the number of elements in $\mathcal{T}_{X_1:n}$. Note that the first m_n symbols are already known. Let's call them $x_1, x_2, \ldots, x_{m_n}$. Since,

$$\sum_{X_{1:n} \in \mathcal{T}_{X_{1:n}}} \tilde{P}^{m_n}(X_{m_n+1}, \ldots, X_n \mid X_1 = x_1, X_2 = x_2, \ldots, X_{m_n} = x_{m_n}) \leq 1,$$

we have

$$\sum_{X_{1:n}\in\mathcal{T}_{X_{1:n}}} \prod_{i=m_n+1}^{n} \tilde{P}^{m_n}(X_i|X_{i-1},\ldots,X_{i-m_n}) = |\mathcal{T}_{X_{1:n}}| \prod_{j=1}^{\ell^{m_n+1}} (Q_j^{m_n})^{(n-m_n)f_j^{m_n,n}}$$

Hence,

$$|\mathcal{T}_{X_{1:n}}| < 2^{-(n-m_n)\sum_{j=1}^{l^{(m_n+1)}} f_j^{m_n,n} \log Q_j^{m_n}}.$$

This implies that coding the index of an element of $\mathcal{T}_{X_{1:n}}$ requires less than $-(n-m_n)\sum_{j=1}^{l^{m_n+1}} f_j^{m_n,n} \log Q_j^{m_n}$ bits. Combining all the above pieces we obtain the following upper bound for the length of our program:

$$K(X_{1:n}) \leq C' + \log^*(m_n) + l\max_{j\leq l} K(a_j) + l^{(m_n+1)}\log^* n + m_n \log^* l - (n-m_n)\sum_{j=1}^{l^{m_n+1}} f_j^{m_n,n} \log Q_j^{m_n}. \tag{8.4}$$

Our goal is to show that $\frac{K(X_1,\ldots,X_n)}{\sqrt{n}} - \sqrt{n}H(X_1|X_{0:-\infty})$ converges in distribution to a normal random variable. Note that the first five terms in (8.4) are deterministic and when divided by $\sqrt{n}$, they converge to zero. Hence, we focus on the only remaining term, i.e., $(n-m_n)\sum_{j=1}^{l^{m_n+1}} f_j^{m,n} \log Q_j^{m_n}$. We have

$$\begin{aligned}
&\sqrt{n}\Big(\frac{1}{n}(n-m_n)\sum_{j=1}^{l^{m_n+1}} f_j^{m_n,n}\log Q_j^{m_n} + H(X_0|X_{-1},\ldots,X_{-\infty})\Big) \\
&= \sqrt{n}\Big(\frac{n-m_n}{n}\sum_{j=1}^{l^{m_n+1}} f_j^{m_n,n}\log Q_j^{m_n} + H(X_0|X_{-1},\ldots,X_{-m_n})\Big) \\
&+ \sqrt{n}(H(X_0|X_{-1},\ldots,X_{-\infty}) - H(X_0|X_{-1},\ldots,X_{-m_n})).
\end{aligned} \tag{8.5}$$

Our first claim is that

$$\sqrt{n}(H(X_0|X_{-1},\ldots,X_{-\infty}) - H(X_0|X_{-1},\ldots,X_{-m_n})) \to 0, \tag{8.6}$$

as $n \to 0$. To see why this holds, note that

$$\begin{aligned}
&\sqrt{n}|H(X_0|X_{-1},\ldots,X_{-\infty}) - H(X_0|X_{-1},\ldots,X_{-m_n}))| \\
&\leq \sqrt{n}K2^{-Cm_n\log(l)} \\
&\leq Kn^{\frac{1}{2}-C(\frac{1}{2}-\epsilon)} \to 0,
\end{aligned}$$

as $n\to\infty$. Here we should remind the reader that we have picked $m_n = \frac{\frac{1}{2}-\epsilon}{\log l}\log n$ with ϵ satisfying $\frac{1}{2} - \frac{1}{2C} > \epsilon > 0$. Combining (8.5) and (8.6) we conclude that the only remaining step is to show that $\sqrt{n}(\frac{n-m_n}{n}\sum_{j=1}^{l^{m_n+1}} f_j^{m_n,n}\log Q_j^{m_n} + H(X_0|X_{-1},\ldots,X_{-m_n}))$ is Gaussian.

Toward this goal we first define

$$Y_j^{m_n} \triangleq \log P(X_j|X_{j-1}, X_{j-2}, \ldots, X_{j-m_n}).$$

Note that $\sum_{j=1}^{l^{m_n+1}} f_j^{m_n,n} \log Q_j^{m_n} = \frac{1}{n-m_n} \sum_{j=m_n+1}^{n} Y_j^{m_n}$. Define

$$S_n^{m_n} \triangleq \sum_{i=m_n+1}^{n} Y_i^{m_n}.$$

To prove the Gaussianity of $S_n^{m_n}$ we employ Corollary 8.3. First let us check the conditions of this theorem for $Y_j^{m_n}$:

1. Boundedness of $\mathbb{E}|Y_j^{m_n}|^{\frac{8(\beta+1)}{\beta-1}}$: First note that

$$\begin{aligned}
&\mathbb{E}|Y_j^{m_n}|^{\frac{8(\beta+1)}{\beta-1}} \\
&= \mathbb{E} \sum_{x_j \in A} P(x_j|X_{j-1}, X_{j-2}, \ldots, X_{j-m_n})|\log P(x_j|X_{j-1}, X_{j-2}, \ldots, X_{j-m_n})|^{\frac{8(\beta+1)}{\beta-1}} \\
&= \mathbb{E} \sum_{x_j \in A} g(P(x_j|X_{j-1}, X_{j-2}, \ldots, X_{j-m_n})),
\end{aligned}$$

   where the function g is defined in the following way: $g : [0,1] \to \mathbb{R}$ and $g(t) = t|\log(t)|^{\frac{8(\beta+1)}{\beta-1}}$ for $t \neq 0$, and also $g(0) = 0$. It is straightforward to check the following properties of g:

   (i) g(t) is continuous at zero.

   (ii) There exists $C_{\frac{2(\beta+1)}{\beta-1}} \in (0,1)$ such that $g'(C_{\frac{8(\beta+1)}{\beta-1}}) = 0$

   (iii) $g'(t) > 0$ for $t < C_{\frac{8(\beta+1)}{\beta-1}}$

   (iv) $g'(t) \leq 0$ for $t > C_{\frac{8(\beta+1)}{\beta-1}}$.

   This automatically implies that $g(t) \leq g(C_{\frac{8(\beta+1)}{\beta-1}})$ for all $t \in [0,1]$. Combing this fact with (1) implies

$$\mathbb{E}|Y_j^{m_n}|^{\frac{8(\beta+1)}{\beta-1}} = \mathbb{E} \sum_{X_j \in A} g(P(X_j|X_{j-1}, X_{j-2}, \ldots, X_{j-m_n}) \leq lg(C_{\frac{8(\beta+1)}{\beta-1}}). \tag{8.7}$$

   Note that the upper bound does not depend on either m_n, n or j.

2. The mixing coefficient α: First let $\alpha^{Y^{m_n}}(i)$ denote the α-mixing coefficient for the Y^{m_n} sequence, and let $\alpha(i)$ denote the α mixing coefficient for the original process

$X_1, \dots, X_n$. It is straightforward to check that for every $i > m_n$

$$\alpha^{Y^{m_n}}(i) \leq \alpha(i - m_n) \leq \begin{cases} K(i-m_n)^{-\beta}, \ i > m_n \\ 1, \text{ otherwise.} \end{cases}$$

where the last step is due to Condition 2 in the statement of the theorem. As a reminder we have $m_n = O(\log(n))$.

3. For notational simplicity in the rest of the proof we use the notation $\sum_{j=1}^{n-m_n} Y_j^{m_n}$ instead of $\sum_{j=m_n+1}^{n} Y_j^{m_n}$. Define $\tilde{\sigma}_{k,n}^2 = \frac{1}{n}\mathrm{var}(\sum_{j=1}^{k} Y_j^{m_n})$. We will later prove that $\frac{\tilde{\sigma}_{n,n}^2}{n} \to \sigma^2$, where

$$\begin{aligned} \sigma^2 \triangleq\ & \mathrm{var}(\log(P(X_0|X_{-1}, \dots, X_{-\infty}))) \\ & +2\sum_k \mathrm{cov}(\log(P(X_0|X_{-1}, \dots, X_{-\infty})), \log(P(X_k|X_{k-1}, \dots, X_{-\infty}))). \end{aligned} \tag{8.8}$$

We have

$$\begin{aligned} \frac{\mathrm{var}(\sum_{j=1}^{n-m_n} Y_j^{m_n})}{n-m_n} &= \mathrm{var}(Y_1^{m_n}) + \frac{2}{n}\sum_{i=1}^{n}\sum_{k=i+1}^{n} \mathrm{cov}(Y_i^{m_n}, Y_k^{m_n}) \\ &= \mathrm{var}(Y_1^{m_n}) + \frac{2}{n}\sum_{i=1}^{n}\sum_{k=2}^{i} \mathrm{cov}(Y_1^{m_n}, Y_k^{m_n}), \end{aligned} \tag{8.9}$$

where to obtain the last equality we used the stationarity of the process $Y_1^{m_n}, Y_2^{m_n}, \dots$. Our goal is to show that this quantity converges to σ^2. We simplify the expression of (8.9) in the following two steps:

1. Simplifying $\mathrm{var}(Y_1^{m_n})$: First note that

$$\begin{aligned} & |\mathbb{E}(\log(P(X_1|X_{0:-m_n+1}))) - \mathbb{E}(\log(P(X_1|X_{0:-\infty})))| \\ & \leq (\mathbb{E}|(\log(P(X_1|X_{0:-m_n+1}))) - \log(P(X_1|X_{0:-\infty}))|^2)^{\frac{1}{2}} = \rho_{m_n}^{\frac{1}{2}} \to 0. \end{aligned} \tag{8.10}$$

To obtain the last inequality we used Holder's and to obtain the last convergence we used Condition 2 in the statement of the theorem. Furthermore, note that

$$\begin{aligned} & |\mathbb{E}(\log^2(P(X_1|X_{0:-m_n}))) - \mathbb{E}(\log^2(P(X_1|X_{0:-\infty})))| \\ \leq\ & (\mathbb{E}|(\log(P(X_1|X_{0:-m_n+1}))) - \log(P(X_1|X_{0:-\infty}))|^2)^{\frac{1}{2}} \\ & \times(\mathbb{E}|(\log(P(X_1|X_{0:-m_n+1}))) + \log(P(X_1|X_{0:-\infty}))|^2)^{\frac{1}{2}} \to 0. \end{aligned} \tag{8.11}$$

To prove the last convergence we should note that the first term goes to zero according to Condition 2 in the statement of the theorem. Furthermore, similar to the proof of

(8.7) we can show that the last expectation is bounded. Hence, it is straightforward to combine the above two equations and obtain

$$\mathrm{var}(Y_1^{m_n}) = \mathrm{var}(\log(P(X_1|X_{0:-m_n+1}))) \to \mathrm{var}(\log(P(X_1|X_{0:-\infty}))). \quad (8.12)$$

2. Our second step is to discuss the covariance terms in (8.9). Define

$$\begin{aligned} s_{i,n} &\triangleq \sum_{k=2}^{i} \mathrm{cov}(Y_1^{m_n}, Y_k^{m_n}), \\ s &\triangleq \sum_{j} \mathrm{cov}(\log(P(X_1|X_{-1:-\infty})), \log(P(X_j|X_{j-1:-\infty}))). \end{aligned}$$

Note that our goal is to bound

$$\frac{1}{n}|\sum_{i=1}^{n}(s_{i,n} - s)| \leq \frac{1}{n}\sum_{i=1}^{2m_n}|s_{i,n} - s| + \frac{1}{n}\sum_{i=2m_n+1}^{n}|s_{i,n} - s|. \quad (8.13)$$

We will prove later that $\sup_i |s_{i,n} - s|$ is bounded. Hence, since $m_n/n \to 0$, we conclude that the first term goes to zero. Hence, we focus on the second term. Define $Z_j \triangleq \log(P(X_j|X_{j-1:-\infty}))$. Then we have

$$\begin{aligned} \frac{1}{n}\sum_{i=2m_n+1}^{n}|s_{i,n} - s| &\leq \frac{1}{n}\sum_{i=2m_n+1}^{n}\sum_{j=2}^{2m_n}|\mathrm{cov}(Y_1^{m_n}, Y_j^{m_n}) - \mathrm{cov}(Z_1, Z_j)| \\ &+ \frac{1}{n}\sum_{i=2m_n+1}^{n}\sum_{j=2m_n+1}^{i}|\mathrm{cov}(Y_1^{m_n}, Y_j^{m_n}) - \mathrm{cov}(Z_1, Z_j)| \\ &+ \frac{1}{n}\sum_{i=2m_n+1}^{n}\sum_{j=i}^{\infty}|\mathrm{cov}(Z_1, Z_j)|. \end{aligned} \quad (8.14)$$

We will show that the each of the three terms on the right converge to zero. Before we proceed further, note that

$$\mathbb{E}(|Y_1^{m_n} - Z_1|^2) = \mathbb{E}|\log(P(X_1|X_{0:-m_n+1})) - \log(P(X_1|X_{0:-\infty}))|^2 = \rho_{m_n}^2. \quad (8.15)$$

Furthermore, similar to the proof of (8.7) it is straightforward to show that

$$\begin{aligned} \mathbb{E}(|Z_j|) \leq (\mathbb{E}|Z_j|^2)^{\frac{1}{2}} < M, \\ \mathbb{E}|Y_j^{m_n}| \leq (\mathbb{E}|Y_j^{m_n}|^2)^{\frac{1}{2}} < M, \end{aligned} \quad (8.16)$$

where $M = l \sup_{t\in[0,1]} |g_2(t)|$ with $g_2(t) = t|\log(t)|^2$. Now we turn our attention to bounding the terms in (8.14).

$$\begin{aligned}
&|\text{cov}(Y_1^{m_n}, Y_j^{m_n}) - \text{cov}(Z_1, Z_j)| \le |\text{cov}(Y_1^{m_n} - Z_1, Z_j)| + |\text{cov}(Y_1^{m_n}, Z_j - Y_j^{m_n})| \\
&\le \mathbb{E}|(Y_1^{m_n} - Z_1)Z_j| + |\mathbb{E}(Y_1^{m_n} - Z_1)\mathbb{E}Z_j| + \mathbb{E}|Y_1^{m_n}(Z_j - Y_j^{m_n})| + |\mathbb{E}(Y_1^{m_n})\mathbb{E}(Z_j - Y_j^{m_n})| \\
&\le (\mathbb{E}|Y_1^{m_n} - Z_1|^2)^{\frac{1}{2}}(\mathbb{E}|Z_j|^2)^{\frac{1}{2}} + (\mathbb{E}|Y_j^{m_n} - Z_j|^2)^{\frac{1}{2}}\mathbb{E}|Z_j| \\
&\quad + (\mathbb{E}|Y_j^{m_n} - Z_j|^2)^{\frac{1}{2}}(\mathbb{E}|Y_1^{m_n}|^2)^{\frac{1}{2}} + (\mathbb{E}|Y_j^{m_n} - Z_j|^2)^{\frac{1}{2}}\mathbb{E}|Y_1^{m_n}| \\
&\le 4M\rho_{m_n}. \qquad (8.17)
\end{aligned}$$

Hence, we conclude that

$$\frac{1}{n}\sum_{i=2m_n+1}^{n}\sum_{j=1}^{2m_n}|\text{cov}(Y_1^{m_n}, Y_j^{m_n}) - \text{cov}(Z_1, Z_j)| \le \frac{n - 2m_n}{n}2m_n 4M\rho_{m_n} \to 0,$$

as $n \to \infty$. Note that the last convergence in the theorem is derived from Condition 2 in the statement of the theorem. Now we find a bound on the second term in (8.14). Define

$$W_j \triangleq \log(P(X_j|X_{j/2:j-1})).$$

Then, we have

$$\begin{aligned}
&|\text{cov}(Y_1^{m_n}, Y_j^{m_n}) - \text{cov}(Z_1, Z_j)| \le |\text{cov}(Y_1^{m_n} - Z_1, Z_j)| + |\text{cov}(Y_1^{m_n}, Z_j - Y_j^{m_n})| \\
&\le |\text{cov}(Y_1^{m_n} - Z_1, Z_j - W_j)| + |\text{cov}(Y_1^{m_n} - Z_1, W_j)| + |\text{cov}(Y_1^{m_n}, Z_j - Y_j^{m_n})| \\
&\le |\text{cov}(Y_1^{m_n} - Z_1, Z_j - W_j)| + |\text{cov}(Y_1^{m_n} - Z_1, W_j)| + |\text{cov}(Y_1^{m_n}, Z_j - W_j)| \\
&\quad + |\text{cov}(Y_1^{m_n}, W_j)| + |\text{cov}(Y_1^{m_n}, Y_j^{m_n})| \qquad (8.18)
\end{aligned}$$

The strategy that we use to bound the terms $|\text{cov}(Y_1^{m_n} - Z_1, Z_j - W_j)|$ and $|\text{cov}(Y_1^{m_n}, Z_j - W_j)|$ is the same. Also, the strategy we use to bound $|\text{cov}(Y_1^{m_n} - Z_1, W_j)|$ and $|\text{cov}(Y_1^{m_n}, W_j)|$ is the same. Hence, we only derive the bounds for the following three terms: (i) $|\text{cov}(Y_1^{m_n}, Z_j - W_j)|$, (ii) $|\text{cov}(Y_1^{m_n}, W_j)|$, and (iii) $|\text{cov}(Y_1^{m_n}, Y_j^{m_n})|$.

(a) $|\text{cov}(Y_1^{m_n}, Z_j - W_j)|$: By using Holder's inequality we conclude that

$$\begin{aligned}
|\text{cov}(Y_1^{m_n}, Z_j - W_j)| &\le \mathbb{E}|Y_1^{m_n}(Z_j - W_j)| + \mathbb{E}|Y_1^{m_n}|\mathbb{E}|Z_j - W_j| \\
&\le 2(\mathbb{E}|Z_j - W_j|^2)^{\frac{1}{2}}\mathbb{E}(|Y_1^{m_n}|^2)^{\frac{1}{2}} \\
&\le 2\rho_{j/2}^{\frac{1}{2}}M. \qquad (8.19)
\end{aligned}$$

(b) $|\text{cov}(Y_1^{m_n}, W_j)|$: Note that W_j is measurable with respect to $\mathcal{F}_{j/2}^{j}$ and $Y_1^{m_n}$ is measurable with respect to $\mathcal{F}_{-\infty}^{1}$. Hence, by employing Lemma 7 we conclude that

$$|\mathrm{cov}(Y_1^{m_n}, W_j)| \leq \alpha(j/2)^{\frac{\beta+1}{2\beta}}(4 + 2\tilde{M}),$$

where $\tilde{M} = lg(C_{\frac{2(\beta+1)}{\beta-1}})$. Note that to obtain the last inequality we have used (8.7).

(c) $|\mathrm{cov}(Y_1^{m_n}, Y_j^{m_n})|$: Similar to the argument of the previous case we conclude that

$$|\mathrm{cov}(Y_1^{m_n}, Y_j^{m_n})|| \leq \alpha(j - m_n)^{\frac{\beta+1}{2\beta}}(4 + 2\tilde{M}).$$

Combining (8.18) and the above three cases, we conclude that

$$\begin{aligned} &\frac{1}{n}\sum_{i=2m_n+1}^{n}\sum_{j=2m_n+1}^{i}|\mathrm{cov}(Y_1^{m_n}, Y_j^{m_n}) - \mathrm{cov}(Z_1, Z_j)| \\ &\leq \frac{1}{n}\sum_{i=2m_n+1}^{n}\sum_{j=2m_n+1}^{\infty} 4\rho_{j/2}M + 2\alpha(j/2)^{\frac{\beta+1}{2\beta}}(4+2M) + \alpha(j-m_n)^{\frac{\beta+1}{2\beta}}(4+2M) \\ &\leq \sum_{j=2m_n+1}^{\infty} 4\rho_{j/2}M + 2\alpha(j/2)^{\frac{\beta+1}{2\beta}}(4+2M) + \alpha(j-m_n)^{\frac{\beta+1}{2\beta}}(4+2M) \to 0, \qquad (8.20) \end{aligned}$$

as $n \to \infty$. The last term of (8.14) can be bounded in exactly similar fashion, i.e., we use the upper bound $|\mathrm{cov}(Z_1, Z_j)| \leq |\mathrm{cov}(Z_1, Z_j - W_j)| + |\mathrm{cov}(Z_1, W_j)|$, and then employ Lemma 7 and the definition of ρ to bound the error. Since the proof is similar we skip it.

Combining all these steps we conclude that

$$\frac{1}{n}|\sum_{i=1}^{n}(s_{i,n} - s)| \to 0. \qquad (8.21)$$

Equations (8.9), (8.12), and (8.21) together prove that

$$\frac{\mathrm{var}(\sum_{j=1}^{n-m_n} Y_j^{m_n})}{n - m_n} \to \sigma^2.$$

Now we can apply Corollary 8.3 and conclude that

$$\lim\inf_{n\to\infty} P(\sqrt{n}(\frac{1}{n}K(X_{1:n}) - H(X_0|X_{-\infty:-1})) \leq t) \geq \Phi(t\sigma),$$

which is one side of what we had to prove.

□

8.2.3.2 Upper bound

PROOF. Define $\delta_n \triangleq n^{-\frac{2}{3}}$.

$$P\Big(\frac{K(X_{1:n})}{n} < -\frac{\log(P(X_{1:n}))}{n} - \delta_n\Big)$$
$$\leq P\left(\frac{K(X_{1:n})}{n} < -\frac{\log(P(X_{1:n}))}{n} - \delta_n, \frac{K(X_{1:n})}{n} < x\right) + P\left(\frac{K(X_{1:n})}{n} > x\right). \quad (8.22)$$

Our goal is to show that under a proper choice of x, both probabilities on the right converge to zero as $n \to \infty$. First note that

$$P\left(\frac{K(X_{1:n})}{n} < -\frac{\log(P(X_{1:n}))}{n} - \delta_n, \frac{K(X_{1:n})}{n} < x\right)$$
$$\leq \sum_{i=1}^{nx} \sum_{\substack{v \text{ as } K(v)=i,\\ P(v)<2^{-(i+n\delta_n)}}} P(X_{1:n} = v) \leq \sum_{i=1}^{nx} \sum_{\substack{v \text{ as } K(v)=i,\\ P(v)<2^{-(i+n\delta_n)}}} 2^{\log P(v)}$$
$$\leq \sum_{i=1}^{nx} \sum_{\substack{v \text{ as } K(v)=i,\\ P(v)<2^{-(i+n\delta_n)}}} 2^{-(i+n\delta_n)} \leq \sum_{i=1}^{nx} 2^i 2^{-(i+n\delta_n)} \leq nx2^{-n\delta_n} \to 0. \quad (8.23)$$

Furthermore, if we choose $x = \frac{3}{2}\max_i K(a_i)$, we have

$$P\left(\frac{K(X_{1:n})}{n} > \frac{3}{2}\max_i K(a_i)\right) \to 0. \quad (8.24)$$

as $n \to \infty$. Hence, by combining (8.22), (8.23), and (8.24), we have

$$P\Big(\frac{K(X_{1:n})}{n} < -\frac{\log(P(X_{1:n}))}{n} - \delta_n\Big) \to 0. \quad (8.25)$$

On the other hand, $\forall t$,

$$P\Big(\sqrt{n}(\frac{1}{n}K(X_{1:n}) - H(X_0|X_{-\infty:-1})) \leq t\Big)$$
$$\leq P\Big(\frac{K(X_{1:n})}{n} < -\frac{\log(P(X_{1:n}))}{n} - \delta_n\Big) + P\Big(\sqrt{n}(-\frac{\log(P(X_{1:n}))}{n} - \delta_n - H(X_0|X_{-\infty:-1})) \leq t\Big).$$

Note two main points about our last expression: (i) According to (8.25) the first term goes to zero as $n \to \infty$. (ii) According to Theorem 5.1the term is asymptotically normal $(-\frac{\log(P(X_{1:n}))}{\sqrt{n}} - \sqrt{n}\delta_n - \sqrt{n}H(X_0|X_{-\infty:-1}))$.

Therefore we proved the desired result.

□

8.2.4 Proof of Theorem 5.7

Before we go to the details of the proof we will review the main ideas.

To get inequality 5.4, we are going to use the upper and lower bounds on the Kolmogorov complexity, derived in the proof of Theorem 5.4. We will obtain concentration-inequalities and combine them to obtain a concentration result for the Kolmogorov complexity. Note that we use the notations defined in (8.2) and (8.3). Define

$$g(X_{1:n}) \triangleq (n-m) \sum_{j=1}^{l^{m+1}} f_j^{m,n} \log Q_j^m.$$

We would like to use Theorem 3.7 to show that $g(X_{1:n})$ concentrates. Toward this goal we need to compute the self bounding coefficients of $g(X_{1:n})$.

To show inequality 5.5 we will also use Theorem 4.3 [15]. Toward this goal we need to do the following two steps: (i)Compute self bounding coefficients for $X_{1:n} \to \frac{K(X_{1:n})}{n}$. (ii) Calculate an upper-bound for $\mathbb{E}(\frac{K(X_{1:n})}{n}) - H(X_1|X_{0:-m+1})$. With this summary we now discuss the details of the proof.

First, according to the proof of Theorem 5.4, using the notations introduced in (8.4), we have

$$K(X_{1:n}) \leq C' + \log^*(m) + l \max_{j \leq l} K(a_j) + l^{(m+1)} \log^* n - m \log^* l - (n-m) \sum_{j=1}^{l^{m+1}} f_j^{m,n} \log Q_j^m.$$

Let $C_1(n) \triangleq C' + \log^*(m) + l \max_{j \leq l} K(a_j) + l^{(m+1)} \log^* n + m \log^* l$. We prove that the function g is M-Lipschitz and therefore that its self bounding coefficients respect $c_i \leq M$. Then we use Theorem 3.7. Note that

$$(n-m) \sum_{j=1}^{l^{m+1}} f_j^{m,n} \log Q_j^m = \sum_{j=m+1}^{n} \sum_{k=1}^{l^{m+1}} I_{(X_{j-m:j}=a_k^m)} \log(Q_k^m),$$

where a_k^m is the k^{th} element of A^m. Let $x, x' \in A^n$ denote two vectors that only differ at the j^{th}-coordinate (i.e. $x_i = x'_i, \ \forall i \neq j$). Then, by the M-stability assumption of the theorem $|g(x) - g(x')| \leq M$ (note that g is the log-likelihood of $X_{m+1:n}$). Hence, g is M-Lipschitz for the Hamming metric. Theorem 4.3 [15] implies that for every $t > 0$

$$P\Big((n-m) \sum_{j=1}^{l^{m+1}} f_j^{m,n} \log Q_j^m + (n-m) H(X_1|X_{0:-m+1}) \geq t\Big) \leq e^{-\frac{2t^2(1-\Lambda(X))}{nM^2}}.$$

It is straightforward to confirm that for every t, if $t - \frac{C_1(n)}{n} + \frac{m}{n}H(X_1|X_{0:-m+1}) > 0$, then

$$\begin{aligned}
&P\Big(\frac{K(X_{1:n})}{n} - H(X_1|X_{0:-m+1}) \geq t\Big) \\
&\leq P\Big(\frac{1}{n}((n-m)\sum_{j=1}^{l^{m+1}} f_j^{m,n} \log Q_j^m - [(n-m)H(X_1|X_{0:-m+1}) + mH(X_1|X_{0:-m+1})]) \geq t - \frac{C_1(n)}{n}\Big) \\
&\leq P\Big(\frac{1}{n}((n-m)\sum_{j=1}^{l^{m+1}} f_j^{m,n} \log Q_j^m - (n-m)H(X_1|X_{0:-m+1})) \geq t - \frac{C_1(n)}{n} + \frac{m}{n}H(X_1|X_{0:-m+1})\Big) \\
&\leq e^{-\frac{n\left(t - \frac{C_1(n)}{n} + \frac{m}{n}H(X_1|X_{0:-m+1})\right)^2 (1-\Lambda(X))}{2M^2}}.
\end{aligned} \tag{8.26}$$

We now want to discuss the details of the proof of inequality 5.5. For $n \in \mathbb{N}$, consider the two vectors $x, x' \in A^{n^2}$ such that $d_n(x, x') \leq 1$. If $d_n(x, x') = 0$, then we can easily see that $|K(x_{1:n}) - K(x'_{1:n})| = 0$. Hence, we assume that $d_n(x, x') = 1$. Suppose that $x_i \neq x'_i$. Then we can note that if the universal machine knows $x_{1:n}$ to know $x'_{1:n}$ it only need to know i and x'_i. Therefore

$$K(x'_{1:n}) \leq K(x_{1:n}) + C' + \log^*(n) + \max_i K(a_i),$$

where C' is a constant that depends only on the universal Turing machine. Since the previous inequality is symmetric in x, x' we obtain that the function $x_{1:n} \to \frac{1}{n}K(x_{1:n})$ is $\frac{C' + \log^*(n) + \max_i K(a_i)}{n}$-Lipschitz. Hence, Theorem 4.3 [15] implies that for every $t > 0$

$$P\Big(\frac{1}{n}K(X_{1:n}) - \mathbb{E}(\frac{1}{n}K(X_{1:n})) < -t\Big) \leq e^{-\frac{2nt^2(1-\Lambda(X))}{(C' + \log^*(n) + \max_i K(a_i))^2}}.$$

Moreover thanks to Kraft inequality and the positivity of the Kullback-Leibler divergence we have that $\mathbb{E}(\log(\frac{P(X_{1:n})}{2^{-K(X_{1:n})}})) \geq 0$, hence $H(X_{1:n}) \geq \mathbb{E}(K(X_{1:n}))$. Furthermore, according to (8.4) we have that for all $m \in \mathbb{N}$

$$\mathbb{E}(K(X_{1:n})) \leq C' + \log^*(m) + l \max_{j \leq l} K(a_j) + l^{m+1} \log^* n + m \log^* l + (n-m)H(X_1|X_{0:-m+1}).$$

Hence

$$\begin{aligned}
&|\frac{1}{n}\mathbb{E}(K(X_{1:n})) - H(X_1|X_{0:-m+1})| \\
&\qquad \leq \frac{C' + \log^*(m) + l \max_{j \leq l} K(a_j) + l^{m+1} \log^* n + m \log^* l + mH(X_1)}{n}.
\end{aligned} \tag{8.27}$$

Therefore, we have that $\forall t > -\frac{C_1(n)}{n} + \frac{m}{n}H(X_1|X_{0:-m+1})$

$$P\Big(\frac{1}{n}K(X_{1:n}) - H(X_1|X_{0:-m+1}) \leq -t\Big) \leq e^{-\frac{2n(t-\frac{C_1(n)}{n}-\frac{m}{n}H(X_1|X_{0:-m+1}))^2(1-\Lambda(X))}{(C'+\log^*(n)+\max_i K(a_i))^2}}.$$

The theorem has been proven with taking $\gamma_n := -\frac{C_1(n)}{n} + \frac{m}{n}H(X_1|X_{0:-m+1})$ and $K_n := C' + \log^*(n) + \max_i K(a_i)$.

Bibliography

[1] Christophe Ambroise and Catherine Matias. New consistent and asymptotically normal parameter estimates for random-graph mixture models. *J. R. Stat. Soc. Ser. B.*, 74:3–35, 2012.

[2] A. R. Barron and T. M. Cover. Minimum complexity density estimation. *IEEE Trans. Information Theory*, 37(4):1034–1054, 1991.

[3] P. J. Bickel, A. Chen, and E. Levina. The method of moments and degree distributions for network models. *Ann. Statist.*, 39(5):2280–2301, 2011.

[4] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures Algorithms*, 31(1):3–122, 2007.

[5] E. Bolthausen. On the central limit theorem for stationary mixing random fields. *Ann. Probab.*, 10(4):1047–1050, 1982.

[6] C. Borgs, J. T. Chayes, H. Cohn, and V. Veitch. Sampling perspectives on sparse exchangeable graphs. 2017. arxiv 1708.03237.

[7] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs. I. *Adv. Math.*, 219(6):1801–1851, 2008.

[8] Christian Borgs, Jennifer T. Chayes, Henry Cohn, and Nina Holden. Sparse exchangeable graphs and their limits via graphon processes. 2016. arxiv 1601.07134.

[9] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013.

[10] Richard C Bradley et al. Basic properties of strong mixing conditions. a survey and some open questions. *Probability surveys*, 2:107–144, 2005.

[11] A. Brudno. Entropy and the complexity of the trajectories of a dynamic system. *Trudy Moskovskogo Matematicheskogo Obshchestva*, 44:124–149, 1982.

[12] Hans Bühlmann. Le problème "limite central" pour les variables aléatoires échangeables. *C. R. Acad. Sci. Paris*, 246:534–536, 1958.

[13] Diana Cai, Trevor Campbell, and Tamara Broderick. Edge-exchangeable graphs and sparsity. In *Advances in Neural Information Processing Systems 29*, pages 4249–4257. 2016.

[14] Francois Caron and Emily B. Fox. Sparse graphs using exchangeable random measures. *J. Roy. Statist. Soc. Ser. B*, 2017.

[15] Sourav Chatterjee. *Concentration inequalities with exchangeable pairs*. PhD thesis, 2005.

[16] Sourav Chatterjee et al. A new method of normal approximation. *The Annals of Probability*, 36(4):1584–1610, 2008.

[17] D. S. Choi, P. J. Wolfe, and E. M. Airoldi. Stochastic blockmodels with a growing number of classes. *Biometrika*, 99(2):273–284, 2012.

[18] T. Cover and J. Thomas. *Elements of Information Theory.* John Wiley & Sons, 2012.

[19] Harry Crane and Walter Dempsey. Edge exchangeable models for interaction networks. *J. Amer. Statist. Assoc.*, 2017.

[20] Manfred Denker and Gerhard Keller. On u-statistics and v. mise'statistics for weakly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 64(4):505–522, 1983.

[21] Persi Diaconis and Svante Janson. Graph limits and exchangeable random graphs. *Rendiconti di Matematica, Serie VII*, 28:33–61, 2008.

[22] D. L. Donoho. *The Kolmogorov sampler.* Department of Statistics, Stanford University, 2002.

[23] G. K. Eagleson and N. C. Weber. Limit theorems for weakly exchangeable arrays. *Math. Proc. Cambridge Philos. Soc.*, 84(1):123–130, 1978.

[24] Manfred Einsiedler and Thomas Ward. *Ergodic theory.* Springer, 2011.

[25] C. Gao, Y. Lu, and H. H. Zhou. Rate-optimal graphon estimation. *Ann. Statist.*, 43:2624–2652, 2015.

[26] H.-O. Georgii. *Gibbs measures and phase transitions.* de Gruyter, 2nd edition, 2011.

[27] Daniel Hammer, Andrei E Romashchenko, Alexander Shen, and Nikolai K Vereshchagin. Inequalities for shannon entropies and Kolmogorov complexities. In *Proc.Twelfth Annual IEEE Conference on Computational Complexity*, pages 13–23, 1997.

[28] Torsten Hothorn and Berthold Lausen. Double-bagging: Combining classifiers by bootstrap aggregation. *Pattern Recognition*, 36(6):1303–1309, 2003.

[29] Tailen Hsing, Wei Biao Wu, et al. On weighted u-statistics for stationary processes. *The Annals of Probability*, 32(2):1600–1631, 2004.

[30] Daniel Hug, Gunter Last, and Matthias Schulte. Second order properties and central limit theorems for geometric functionals of Boolean models. *Ann. Appl. Probab.*, 2016.

[31] S. Jalali and A. Maleki. Minimum complexity pursuit. In *Proc. Allerton Conference on Communication, Control, and Computing*, pages 1764–1770, 2011.

[32] L. F. James, A. Lijoi, and I. Prünster. Posterior analysis for normalized random measures with independent increments. *Scand. J. Stat.*, 36:76–97, 2009.

[33] Svante Janson. On edge exchangeable random graphs. *J. Stat. Phys.*, 2017.

[34] Jens Ledet Jensen and Hans Künsch. On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. *Ann. Inst. Statist. Math.*, 46(3):475–486, 1994.

[35] O. Kallenberg. *Foundations of Modern Probability.* Springer, 2nd edition, 2001.

[36] O. Kallenberg. *Probabilistic Symmetries and Invariance Principles.* Springer, 2005.

[37] O. Klopp, A. B. Tsybakov, and N. Verzelen. Oracle inequalities for network models and sparse graphon estimation. *Ann. Statist.*, 45(1):316–354, 2017.

[38] A. N. Kolmogorov. Three approaches to the quantitative definition of information. *information transmission*, 1(1):1–7, 1965.

[39] A J Lee. *U-statistics: Theory and Practice.* Routledge, 2019.

[40] S Leung-Yan-Cheong and T Cover. Some equivalences between Shannon entropy and Kolmogorov complexity. *IEEE Transactions on Information Theory*, 24(3):331–338, 1978.

[41] Elon Lindenstrauss. Pointwise theorems for amenable groups. *Invent. Math.*, 146(2):259–295, 2001.

[42] Heinrich Lothar and Molchanov Ilya. Central limit theorem for a class of random measures associated with germ-grain models. *Adv. Appl. Probab.*, 1999.

[43] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96:933–957, 2006.

[44] A. Maitra. Integral representations of invariant measures. *Trans. Amer. Math. Soc.*, 229:209–225, 1977.

[45] An Muchnik and N Vereshchagin. Shannon entropy vs. Kolmogorov complexity. In *International Computer Science Symposium in Russia*, pages 281–291. Springer, 2006.

[46] P. Orbanz and D. M. Roy. Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Trans. on PAMI*, 37:437–461, 2015.

[47] Peter Orbanz. Subsampling large graphs and invariance in networks. 2017. Preprint. arxiv 1710.04217.

[48] Matthew Penrose. Gaussian limits for random geometric measures. *Electron. J. Probab.*, 12:989–10235.

[49] J. Pitman. *Combinatorial stochastic processes*. Springer, 2006.

[50] Nathan Ross. Fundamentals of Stein's method. *Probab. Surv.*, 8:210–293, 2011.

[51] P. C. Shields. *The Ergodic Theory of Discrete Sample Paths*. AMS, 1996.

[52] R. J. Solomonoff. Complexity-based induction systems: comparisons and convergence theorems. *IEEE Trans. Information Theory*, 24(4):422–432, 1978.

[53] Daby M Sow and Alexandros Eleftheriadis. Complexity distortion theory. *IEEE Transactions on Information Theory*, 49(3):604–608, 2003.

[54] Daniel W Stroock and Boguslaw Zegarlinski. The logarithmic Sobolev inequality for discrete spin systems on a lattice. *Comm. Math. Phys.*, 149(1):175–193, 1992.

[55] A. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

[56] Victor Veitch and Daniel M. Roy. The class of random graphs arising from exchangeable random measures. 2015. arxiv 1512.03099.

[57] Victor Veitch and Daniel M. Roy. Sampling and estimation for (sparse) exchangeable graphs. 2016. arxiv 1611.00843.

[58] Nikolai K Vereshchagin and Paul MB Vitányi. Rate distortion and denoising of individual data using Kolmogorov complexity. *IEEE Transactions on Information Theory*, 56(7):3438–3454, 2010.

[59] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge University Press, 2018.

[60] Paul MB Vitanyi and Ming Li. *An introduction to Kolmogorov complexity and its applications*, volume 34. Springer Heidelberg, 1997.

[61] Paul MB Vitányi and Ming Li. Minimum description length induction, Bayesianism, and Kolmogorov complexity. *IEEE Transactions on information theory*, 46(2):446–464, 2000.

[62] V. V'yugin. Ergodic theorems for individual random sequences. *Theoretical Computer Science*, 207(2):343–361, 1998.

[63] Benjamin Weiss. Actions of amenable groups. In *Topics in dynamics and ergodic theory*, volume 310 of *London Math. Soc. Lecture Note Ser.* Cambridge Univ. Press, 2003.

[64] Wojciech H Zurek. *Complexity, Entropy and the Physics of Information.* CRC Press, 2018.

[65] A. K. Zvonkin and L. A. Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Mathematical Surveys*, 25(6):83–124, 1970.